GENETICS AND ECOLOGY OF THE ACI LINEAGE OF FRESHWATER ACTINOBACTERIA

BY TREVOR GHYLIN

A Dissertation submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

(CIVIL AND ENVIRONMENTAL ENGINEERING)

at the

UNIVERSITY OF WISCONSIN – MADISON

2014

ABSTRACT

The acI lineage of Actinobacteria are frequently the most abundant bacterial group in widely varied freshwater systems worldwide. As the most abundant freshwater bacteria, they are likely important in water quality, freshwater ecology and carbon and nutrient cycling in freshwater systems. However, the acI lineage remains to be isolated and cultured and therefore has been difficult to study with traditional laboratory experiments.

We utilized DNA-based methods to explore the acI lineage without the need for isolation and cultivation. The dissertation presented here is comprised of three chapters, which were written with the intent of publishing in scientific journals and are formatted as such.

The first chapter presents our results from genomic investigations of the acI lineage. We obtained mostly complete genomes of 11 acI bacteria utilizing single-cell genomics techniques. These are the first acI genomes ever analyzed. Analysis of these genomes revealed some interesting differences between acI and their Actinomycetales relatives as well as other freshwater bacteria. Most notably, we found enrichment of genes focused on the uptake and breakdown of complex organic nitrogen compounds such as polyamines, cyanophycin, dipeptides, and branched-chain amino acids. Some of these genes also indicate a potential relationship between acI and phytoplankton. AcI's cyanophycinase gene may be used to breakdown cyanophycin produced by cyanobacteria, while acI's genes for polyamine uptake may indicate that acI consumes decay products of diatoms, which have a high concentration of polyamines in their unique cell walls. AcI also contain genes for the uptake of lipopolysaccharides, a cell-wall component of gram-negative bacteria. The genes for cyanophycin, polyamine and lipopolysaccharide share a common trait that they are likely used to breakdown and consume decay products from phytoplankton or other bacteria. This is in line with what is known about other Actinomycetales, which are known to degrade complex organics and are important in the decay processes.

The second chapter presents the results of an in-situ analysis of acI's population dynamics over ten years in Lake Mendota. This study represents the longest and most phylogenetically resolved study of acI population dynamics ever performed. Samples were collected bi-weekly or monthly from Lake Mendota from 2000 to 2010. These samples were analyzed with qPCR and 16S Illumina Miseq sequencing to quantify the relative abundance of acI and other bacteria on each sampling date. These data were used in conjunction with long-term environmental data collected by the North Temperate Lakes Long Term Ecological Research team, which includes physical, chemical and biological data

from Lake Mendota. The acI lineage was consistently the most abundant bacterial group in the lake with acI-B1 and acI-A6 comprising at least 75% of the acI lineage bacteria detected. The lack of variation in acI's abundance made it somewhat challenging to deduce drivers for its population dynamics. Fortunately, the long-term dataset revealed one very unique year (2006) in which acI relative abundance declined substantially to near extinction from the lake. This type of decline in acI has not been documented previously in literature. Surprisingly, total actinobacteria reached its peak in 2006 due to the rise of non-acI freshwater actinobacteria such as acTH1-A1, Luna1-A1, acSTL-A1, acIV-D. The primary unique feature of 2006 was the complete lack of a usually predictable diatom (*Stephanodiscus parvus*) bloom in April. This finding led us to further investigate the relationship between acI and phytoplankton. We found that acI-A6 correlates with diatom blooms while acI-B1 does not. We found that acI-A6 correlates with certain cyanobacteria (*Aphanocapsa* and *Merismopedia*) while acI-B1 has negative correlations with these cyanobacteria. AcI-B1 had positive correlations with *Aphanizomenon* and *Microcystis*. Correlations between acI-A6, acI-B1 and phytoplankton tend to become stronger in the weeks after phytoplankton blooms, indicating that they benefit from the senescence and decay of these blooms. AcI-C2, had surprisingly strong correlations with cyanobacteria with no delay, indicating that this lineage is associated with live, active cyanobacterial blooms. We also found that acI-B1 and acI-A6 had strong negative correlations with the zooplankton *Copepods* and *Daphnia*, while acTH1-A1, Luna1-A1, acSTL-A1, acIV-D had strong positive correlations with *Copepods* and *Daphnia*. This may also help explain why acI declined and was replaced by non-acI Actinobacteria in 2006, which had high numbers of zooplankton.

The third chapter presents the results of a mesocosm experiment. We collected 24 liters of lake water and filtered through 2um to remove bacterivorous grazers. We then added 11 different substrates (cyanobacteria, diatoms and substances associated with cyanobacteria and diatoms) and monitored the impact on the bacterial population using qPCR and 16S rRNA gene Illumina Miseq "tag" sequencing. We found that putrescine and aspartic acid significantly enriched acI populations while spermidine, arginine and N-acetylglucosamine did not. AcI relative abundance increased from day 3 to day 17 in live and killed diatom mesoscoms but not live cyanobacteria mesocosms. We also found that acI relative abundance was greater when phytoplankton was killed and lysed as compared to live phytoplankton, indicating a focus on decay and lysis products.

ACKNOWLEDGEMENTS

(www.microbedetectives.com) that helps me apply my PhD work to solving real world issues in Environmental Engineering. I have learned an incredible amount about the business world since launching the company. I feel this was the perfect fit for my required BTP internship.

Thank you to my committee (Katherine (Trina) McMahon, Joel Pedersen, Garret Suen, James (Jim) Hurley, and Stefan Bertilsson). The depth and breadth of their knowledge and expertise is truly mind-boggling. I'm blessed to be able to tap into their wisdom.

Thanks also to the Water Environment Federation Canham Graduate Studies Fellowship and the Biotechnology Training Program of the National Institutes of Health for financing my PhD work. I could not have done of this work without the financial support.

My PhD journey has been an interdisciplinary endeavor melding biotechnology, environmental engineering, limnology, sustainability (including social science), and even business. It has been quite a journey and the ultimate end-cap to nine years of university life. I've been blessed with financial, intellectual and emotional support from a variety of organizations and individuals throughout the entire time. Thanks to everyone that has had faith in me and helped me along the way. I'm excited to apply what I've learned to create a better, smarter world.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DNA	Dioxyribonucleic acid
DOE	Department of Energy
g	Grams
ISME	International Society of Microbial Ecology
JGI	Joint Genome Institute
L	Liters
mg	Milligrams
N	Nitrogen
NH_3-N	Ammonia
NO_3^--N	Nitrate
NO_2^--N	Nitrite
NO_x^--N	Nitrate + Nitrite
PCR	Polymerase Chain Reaction
qPCR	Quantitative Polymerase Chain Reaction
RNA	Ribonucleic Acid
s	Standard Deviation
TKN	Total Kjeldahl Nitrogen
TN	Total Nitrogen
P	Phosphorus
TP	Total Phosphorus

1. LITERATURE REVIEW

1.1. INTRODUCTION

Actinobacteria are often the numerically dominant phylum in lakes, sometimes representing more than 50% of the bacterial population (Newton, 2011). However, these bacteria have been difficult to study in a controlled lab setting due to difficulty in cultivation (Jezbera, 2009). As a result, many studies have utilized molecular tools such as FISH, PCR, and metagenomics to study acI populations in situ (Newton, 2007; Sharma, 2009).

Temporal studies of freshwater bacteria have mostly agreed that the freshwater actinobacteria relative population (as % of total bacteria) peaks in spring and fall through winter with a low during summer months. The spring peak generally coincides with a peak in heterotrophic nanoflagellates (grazers) that lags a spring phytoplankton bloom by a few weeks. The actinobacteria population correlates with oxygen and has a negative correlation with dissolved phosphorus (Glockner, 2000; Burkert, 2003; Allgaier, 2006; Zeder, 2009; Salcher, 2010).

Several of these studies utilized FISH (fluorescent in-situ hybridization), which has limited phylogenetic resolution. As a result many studies simply group all Actinobacteria together. However, Allgaier et al (2006) used FISH probes to distinguish between acI-A and acI-B and found acI-A had little seasonality while acI-B had more distinct spring and fall peaks in abundance. This FISH study was somewhat limited as the FISH probes were unable to detect 9-40% of acI cells and could not resolve the phylogeny to detect tribes (i.e. acI-B1, acI-A1, acI-A6, etc.). It is possible that the acI-A tribes had seasonal fluctuations that weren't apparent when grouped together.

1.2. DISCOVERY

Glockner et al (1999, 2000) were the first to report a newly discovered group of Actinobacteria that reside in and dominate freshwater systems. Newton et al (2011) provides a synoptic analysis of

studies that have since confirmed acI's abundance in varied freshwater systems worldwide. Warnecke et al (2005) and Martinez-Garcia (2011) have shown that acI is more active than the average freshwater bacteria and is not simply dormant, as some had previously speculated.

1.3. ACI CULTURING AND ISOLATION

The acI lineage of Actinobacteria has proven difficult to isolate and culture. In 2003, Hahn et al (2003) isolated 9 Actinobacteria from various lakes and identified them as the Luna lineage. This lineage is also known as the acII lineage (Newton, 2011). This lineage falls in a different family (Microbacteriaceae) than acI (Actinomycetales).

Jezbera et al (2009) were able to enrich acI but were not able to isolate these bacteria. Garcia et al (2013) published another successful enrichment of acI (acI-B2) bacteria but the enrichment contained other non-acI bacteria as well, most notably *Polynucleobacter*. The authors cited potential difficulties in cultivation due to slow growth, low cell densities, the need for unknown or unstable growth factors and the dependence on other bacteria to produce certain substances or remove toxic materials.

1.4. MORPHOLOGY

The enrichment of acI-B2 (Garcia, 2013) grew to low cell densitites indicating an oligotrophic nature. Cells were slightly curved rods with an average length of 0.66 μm and a maximum length of 1 μm. The diameter varied between 0.2 and 0.4 μm. This size is comparable with that found in nature (Glöckner et al., 2000).

Hahn et al (2003) found that their acII isolates were Ultramicrobacteria <0.2 um diameter. And were "Selenoid" or vibrio shape. All isolates were yellow or red indicating they may produce secondary metabolites. The pigmentation may also indicate presence of rhodopsin and potential UV resistance.

1.5. ACI ECOLOGY

1.5.1. SEASONALITY

Several studies have investigated the seasonality of freshwater bacterial communities. The results from these studies occasionally disagree but they mostly suggest that the acI lineage reaches peak abundance in spring, declines somewhat during summer and peaks again during fall (Glockner, 2000; Burkert, 2003; Boucher, 2005; Allgaier, 2006). The spring and fall peaks typically correspond either with grazing events (Salcher, 2010) or phytoplankton blooms (Glockner, 2000; Salcher, 2010). Studies generally report that acI relative abundance declines during phytoplankton blooms but increases afterward, coinciding with grazing (Zeder, 2009; Salcher, 2010). Thus, many have studied acI's resistance to grazing and this attribute is well documented (Tarao, 2009). However, some studies have found acI abundance increasing associated with diatom blooms (Salcher, 2010) in the absence of grazing. A possible explanation for acI's dynamics was proposed by Glockner et al (2000) when they hypothesized that acI are able to more efficiently consume lower levels of carbon at lower temperatures.

Some studies have also demonstrated the difficulty with in situ studies of acI due to the lack of seasonal reproducibility (Boucher, 2005) or the lack of consistent correlations between acI and environmental variables across different lakes.

Allgaier et al (2006) found that Actinobacteria population peaked in June and July in all studied lakes in Germany, except for one basin. All of their studied lakes also showed a peak in Actinobacteria abundance from October to December. The lowest Actinobacteria abundances were observed in summer (August to September). This study observed little seasonal variability in acI-A except for slight variations in spring and fall. Stronger seasonal dynamics were seen for acI-B, which mimicked the spring/fall pattern seen with the Actinobacteria. This study was somewhat limited as their FISH probes were unable to detect 9-40% of acI cells and could not detect any acI cells at a more resolved

phylogeny (i.e. tribe level (acI-B1 or acI-A1)). It is possible that acI-A tribes have some seasonality but when they are lumped together the gross effect is not readily visible.

1.5.2. UV RESISTANCE

Warnecke et al (2005) found that the acI lineage represented >90% of all Actinobacteria and the abundance of the acI lineage correlates positively with UV Transparency in mountain lakes, possibly indicating a competitive advantage due to UV stress resistance or an ability to beneficially harness solar radiation. This study also found that the acI bacteria were actively synthesizing DNA, indicating that they are native freshwater bacteria not simply soil bacteria washed in from the watershed (Warnecke, 2005). This study found few or no acII or acIV bacteria in the oligotrophic and mesotrophic mountain lakes.

The correlation does not prove a causal link but the authors speculated that acI may have resistance to UV due to pigmentation, DNA repair mechanisms and high GC content. However, recent genomic studies (Ghai, 2012; Garcia, 2012) have found that acI have low GC content.

Debroas et al (2009) discovered an abundance of genes involved in DNA replication and repair in freshwater Actinobacteria and postulated that these genes may help explain some of the apparent resistance to UV.

1.5.3. RESISTANCE TO PREDATION

Several phenotypic features of aquatic bacteria have been interpreted as adaptations to escape protistan grazing pressure, and have been cited to explain the dominance of certain bacterial taxa (Pernthaler, 2005). Protection from protistan grazing could, moreover, be mediated by properties intrinsic to the bacterial cell wall. Gram-positive bacteria are consumed by protists at significantly lower rates than Gram-negative strains (Iriberri et al., 1994). Thus, Actinobacteria may benefit from a relatively fast growth rate and limited vulnerability to protistan grazing when cohabiting with other

ecologically important bacterial groups in environments characterized by strong grazing pressure (Jezbera et al., 2005).

Some studies have shown that acI is better protected from protistan grazing than other bacterioplankton heterotrophs (Pernthaler, 2001; Jezbera, 2005). Hahn et al (2003) found that some of the acII (aka Luna; related to acI) bacteria were also resistant to grazing by heterotrophic nanoflagellates.

One of the first studies on acI grazing resistance (Pernthaler, 2001) found that acI abundance increased when a size-selective grazer was introduced but did not increase when a different grazer was introduced that was not size selective. Their results indicate acI's small size may provide grazing resistance but only if the dominant grazer is a size-selective predator.

Jezbera et al (2005) further investigated acI grazing resistance by analyzing food vacuoles from protists in a reservoir and found that Actinobacteria were negatively selected by these protists. Jezbera et al (2006) also found that heterotrophic nanoflagellates preferentially consumed larger bacteria and avoided the small Actinobacteria. They hypothesized that the small size of freshwater Actinobacteria was the primary reason for its resistance to grazing.

A more recent study demonstrated a potential mechanism for acI's grazing resistance may be its unique cell wall structure. Tarao et al (2009) found that predation resulted in an 86% decline in *Polynucleobacter* while no significant decrease was observed in nine actinobacterial strains (acII/Luna-2). These results demonstrated that acII's grazing resistance was not due solely to its small size as *Polynucleobacter* is a similar size. When they removed the S-layer or cell surface proteins, the grazing rate on acII increased about 5-fold, indicting that cell surface structures have some impact on grazing resistance.

Though many studies have demonstrated acI's resistance to grazing, one study showed that acI is susceptible to viral lysis (Simek, 2006).

1.5.4. ENERGY AND NUTRIENT SOURCES

Garcia et al (2013) found that putrescine, pyruvate, triethylamine, n-acetylglucosamine and pentoses (xylose and ribose) enriched acI-B2 whereas glucose did not. However, this was an enrichment study that cannot prove acI actually consumed these substances. Other bacteria in the enrichment may have consumed the substances and produced byproducts that stimulated acI.

Several studies have been performed using CARD-FISH MAR (Catalyzed reporter deposition Fluorescent In-situ Hybridization Microautoradiography) to investigate substrate consumption in freshwater Actinobacteria (Table 1). The primary findings of these studies are that acI has been observed to consume glucose, leucine, thymidine, di-NAG and an amino acid mixture. AcI-A did not consume acetate in one study (Buck, 2009) and did not consume NAG in another study (Beier, 2011).

1.5.4.1. CARD-FISH MAR Substrate Incorporation Studies

Table 1. Results of FISH-MAR studies on acI

Substrate	Active cells (% of total acI)	Reference
Glucose	15-25%	(Buck, Grossart et al. 2009)
Leucine	10-20% (3 studies) 70-80% (Perez study)	(Buck, Grossart et al. 2009; Michaela M. Salcher 2010; Perez, Hortnagl et al. 2010; Eckert, Salcher et al. 2012)
Acetate	0% (acI-A) 5-20% (acI-B2 – oxic and anoxic) 15-20% (acI-B3 – anoxic only)	(Buck, Grossart et al. 2009)
Amino Acid Mixture	25%	(Michaela M. Salcher 2010)
Thymidine	40-60%	(Perez, Hortnagl et al. 2010)
N-acetyl Glucosamine	0-3% (Eckert) 10-30% (Beier)	(Beier and Bertilsson 2011; Eckert, Salcher et al. 2012)
di-N-acetyl Glucosamine	10-15%	(Beier and Bertilsson 2011)

1.5.4.2. Actinorhodopsin

Sharma et al (2008, 2009) first reported finding a rhodopsin gene in freshwater Actinobacteria of the acI lineage. Further genomic investigation has confirmed that Actinorhodopsin is likely part of the

core genome of acI bacteria (Garcia, 2012; Martinez-Garcia, 2012; Ghylin, 2014). These studies did

not find a Rubisco gene for carbon fixation but Garcia et al (2012) and Ghylin et al (2014) did report

carbonic anhydrase and PEP carboxylase that could indicate the capability of anaplerotic carbon

fixation. Wurzbacher et al (2012) performed an in situ experiment to investigate the transcription of

actinorhodopsin. They found a circadian expression cycle of actinorhodopsin that surprisingly was not

linked to sunlight. They found minimum expression of actinorhodopsin at dusk and maximum

expression just before sunrise. Wurzbacher et al (2012) suggested that actinorhodopsin may be used for

light-driven transport of substances across the thick gram positive cell wall of acI. They indicate that

this pattern points to a circadian-triggered expression, which might be coupled to protein turnover rates

and accumulation of damaged proteins due to photooxidation during daytime.

1.5.5. pH

The acidity or alkalinity of a lake seems to determine which acI clades thrive in the lake

(Newton, 2007). Newton et al (2007) found that some acI clades (acI-A1, acI-B2, acI-B3) correlated

with acidic lakes while others correlated with alkaline lakes (acI-B1, acI-A2, acI-A6). Lindstrom et al

(2005) found that pH was one of the strongest drivers of bacterial community composition in 15

northern European lakes.

1.5.6. TEMPERATURE

Temperature is also thought to be an important driver of lacustrine bacterial communities

(Lindstrom, 2005) however, studies have shown little correlation between acI and temperature

(Boucher, 2005).

A couple studies demonstrated more specific relationships between acI and Temperature.

Dziallas et al (2011) investigated bacteria associated with microcystis blooms. They found that acI

was present but only when temperature was below 20C while Hahn et al (2005) showed that Luna2 (a

relative of acI) clades that had nearly identical 16S genes from temperate, sub-tropical and tropical regions displayed much difference temperature preferences despite their high phylogenetic similarity.

1.5.7. OXYGEN

The acI lineage is most common in the oxygenated epilimnion. However, acI has been found in hypolimnetic waters (Allgaier, 2006) where the population tends to correlate with oxygen concentration (Boucher, 2005; Taipale, 2009).

1.5.8. NUTRIENTS, CARBON SOURCE, LAKE TYPE (HUMIC, OLIGOTROPHIC, EUTROPHIC)

Studies have indicated that acI is typically an oligotrophic bacteria thriving in low nutrient environments. Newton et al (2011) found that acI-B1 decreased in population when carbon, nitrogen and phosphorus were added to lake mesocosms. Haukka et al (2006) also found that acI correlated with less eutrophic conditions. Other studies have described difficulty in culturing acI, indicating a potentially obligate oligotrophic lifestyle (Jezbera, 2009; Zeder, 2009).

One study seemed to indicate that acI-B dominates humic lakes (Taipale, 2009) while a more comprehensive study (Jones, 2009) found that acI-A was also abundant in some humic lakes.

AcI is frequently abundant in the epilimnion and has proven to be consistent across lake types, as the Actinobacteria are common among oligotrophic (Humbert, 2009), mesotrophic (Debroas, 2009; Humbert, 2009; Zeder, 2009), eutrophic (Wu, 2007), and dystrophic (Newton, 2006) lakes. Actinobacteria have been found to dominate freshwater systems globally. Studies have shown high Actinobacteria abundances in lakes in North America (Newton, 2007), Europe (Glockner, 2000), Africa (Wever, 2005), Asia (Wu, 2006), Australia (Hahn, 2005), South America (Corno, 2009), and Antarctica (Pearce, 2003).

1.5.9. EPILIMNION VS HYPOLIMNION

Most acI studies have demonstrated that acI is an aerobic bacteria commonly associated with the epilimnion of lakes (Salcher, 2010) and usually insignificant in the hypolimnion.

1.5.10. AUTOCHTHONOUS VS ALLOCHTHONOUS

Jones et al (2009) found some evidence that members of the Actinobacteria tribes may partition by broad substrate source categories, such as the ratio of allochthonous to autochthonous carbon produced in a lake. They found that the majority of acI-A OTUs appeared in lakes with low to intermediate CtCh (Colour:Chlorophyll ratio) while two acI-B OTUs occurred in lakes with high CtCh. They hypothesized that OTUs within the acI-A lineage occur in autochthonous lakes and OTUs within the acIB lineage prefer allochthonous systems. The acII and acIV lineages tended to occur in lakes with lower CtCh, suggesting they may be associated with autochthonous lakes as well. This resource partitioning between clades and tribes has been postulated as a reason for shifts in the growth rates and biomass changes observed pre- to post-phytoplankton bloom in Lake Zurich (Zeder, 2009).

Debroas et al (2009) published a metagenomic study of Lac du Bourget. They hypothesized that Actinobacteria in freshwater ecosystems may consume the low- molecular-weight products (< 700 Da), derived from algal excretion. These substances have been shown to be an important carbon source for heterotrophic microorganisms (Maurin et al., 1997; Richardot et al., 2001). Adding support to this hypothesis, Allgaier et al (2007) found that phytoplankton-derived DOM was a determinant factor in Actinobacteria community dynamics.

Salcher et al (2010) found that acI correlated with a spring Cyrptomonas bloom and nanoflagellate grazers in spring. They also found that acI correlated with a fall diatom bloom and they hypothesized that acI may have a competitive advantage at P-limiting conditions in the presence of low molecular weight compounds released by diatoms.

Pernthaler (2001) performed a grazing experiment with cryptomonas and found acI abundance was high when a size selective grazer was there but not when another grazer was present.

Zeder et al (2009) found that Flavobacteria were the most important group for consuming phytoplankton exudates before and during a phytoplankton bloom. They also found that Actinobacteria population peaked about 1 week after the phytoplankton bloom.

Dziallas et al (2011) found that bacterial community composition associated with Microcystis sp. differed significantly with temperature, bacterial source community and number of incubated Cyanobacterial strains. They found that AcI were only present in the 20°C treatments and disappeared at higher incubation temperatures.

Eiler et al (2004) found acI bacteria associated with Cyanobacterial blooms though acI bacteria have not been found in physical association with members of the Cyanobacteria or other phytoplankton (Kolmonen, 2004) but may become a greater part of the bacterioplankton community during phytoplankton blooms (Allgaier, 2006; Salcher, 2010; Zeder, 2009).

1.5.11. TOP DOWN VS BOTTOM UP CONTROL

Salcher et al (2010) found that acI abundance increased during grazing by heterotrophic nanoflagellates in spring. Other studies have also indicated acI's success may be due to its grazing resistance (Pernthaler, 2001; Jezbera, 2005; Tarao, 2009). However, Salcher et al (2010) observed an increase in acI abundance coinciding with a fall diatom bloom in the absence of grazers. They hypothesized that at very low phosphorus concentrations, grazing resistance may not be the sole factor for acI's success in freshwater systems.

1.5.12. IN SITU STUDIES

Allgaier et al (2007) studied Actinobacteria in the epi, meta and hypolinnion in four limnologically different lakes. They found that the Actinobacteria communities were correlated with

conductivity, total phosphorus, alkalinity and primary production. However, the relationships were not uniform across water layers or between lakes.

Buck et al (2009) found that acI-A comprised 60-90% of all Actinobacteria in the epilimnion of two studied lakes but were not detectable in the hypolimnion. They found acI-B2 (FISH probe includes most acI-B) in all samples (epi and hypolimnion) but at lower concentrations and they found acI-B3 (small fraction of acI-B not targeted by previous probe) only in the hypolimnion. They observed glucose incorporation in all three clades. They had similar results with leucine, however they found that acI-A did not take up acetate while acI-B2 and acI-B3 took up acetate in oxic and anoxic conditions.

Eckert et al (2012) found that opportunistic Flavobacteria and Betaproteobacteria dominated the initial period after N-acetylglucosamine (NAG) was added to mesocosms. However, these populations gave way to the rise of Actinobacteria during a subsequent phase of high predation pressure. N-acetylglucosamine is a component of cell walls and the authors hypothesize that the Actinobacteria may be consuming NAG and cell remnants from protist grazing of their competitors.

Salcher et al (2010) published one of the most comprehensive in situ studies of freshwater bacteria to date. They observed bacterial community composition as well as phytoplankton and nanoflagellates over time and at various depths. They observed acI abundance increase following a spring Cryptomonas blooms and corresponding with proliferation of heterotrophic nanoflagellates. They found the acI population again increase in fall corresponding with a diatom bloom in the absence of nanoflagellate grazers. They hypothesized that acI may have a competitive advantage at P-limiting conditions in the presence of low molecular weight compounds released by diatoms. They also found that acI correlated positively with oxygen and heterotrophic nanoflagellates and negatively with phosphorus. AcI incorporated amino acids at a rate greater than average for all bacteria in the lake.

1.6. ACI PHYLOGENY

Zwart et al (2002) published a 16S study including freshwater Actinobacteria and were the first to define freshwater Actinobacterial clusters. They identified the acI lineage as ACK-M1 at that time.

Warnecke et al (2004) published a comprehensive phylogenetic analysis of 63 full length 16S sequences obtained from seven different freshwater habitats. This paper identified four main lineages of freshwater Actinobacteria (acI, acII, acIII, and acIV) with several sublineages. The acI lineage was found mostly in lakes and rivers but also estuaries, and limited detection in a hot spring and marine sample. The acII lineage was found mostly in lakes with some in a hot spring and a couple found in marine environments. The acIII lineage was found exclusively in estuaries and a hypersaline soda lake, where no acI or acII were detected. The acIV lineage was found in nearly every studied water habitat including activated sludge and soil systems. This study did not find any of these lineages in the hypolimnion of lakes. The acI-acIII lineages are associated with actinomycetales while the acIV lineage is associated with acidomicrobiales (Newton, 2011).

1.7. ACI GENOMICS

Ghai et al (2011, 2012) published the first metagenomic studies on acI. They found that acI was the dominant bacteria in their metagenomic data from the Amazon. They also found that acI has low GC (40-50%) content, which is unusual for the typically high-GC Actinobacteria.

Martinez-Garcia (2012) published a large study on 712 single call genomes focused on photoheterotophy. They found that the majority of rhodopsin containing bacteria could be assigned to the Actinobacteria phylum. Furthermore, most Actinobateria analyzed (including acI, acSTL and Luna) contained rhodopsin genes. This study found that freshwater Actinobacteria likelyplay a major role in freshwater photoheterotrophy.

Garcia et al (2012) published the first genome from acI. This genome was obtained using single cell isolation, amplification and DNA sequencing, also called single-cell genomics. The single

amplified genome (SAG) was estimated to be nearly complete (>95%). The genome from the acI-B1

tribe was one of the smallest genomes found in free-living bacteria (1.16 MB) and had low GC (42%).

The genome contained an actinorhodopsin gene as well as transporters for pentoses such as xylose and

ribose.

REFERENCES

Allgaier, M., & Grossart, H. (2006). Diversity and Seasonal Dynamics of Actinobacteria Populations in Four Lakes in Northeastern Germany Diversity and Seasonal Dynamics of Actinobacteria Populations in Four Lakes in Northeastern Germany, 72(5). doi:10.1128/AEM.72.5.3489

Allgaier, M., Brückner, S., Jaspers, E., & Grossart, H.-P. (2007). Intra- and inter-lake variability of free-living and particle-associated Actinobacteria communities. Environmental Microbiology, 9(11), 2728–41. doi:10.1111/j.1462-2920.2007.01385.x

Beier, S., & Bertilsson, S. (2011). Uncoupling of chitinase activity and uptake of hydrolysis products in freshwater bacterioplankton. Limnology and Oceanography, 56(4), 1179–1188. doi:10.4319/lo.2011.56.4.1179

Boucher, D., Jardillier, L., Debroas, D., & Pascal, B. (2005). Succession of bacterial community composition over two consecutive years in two aquatic systems : a natural lake and a lake-reservoir. doi:10.1111/j.1574.6941.2005.00011.x

Buck, U., Grossart, H., Amann, R., & Pernthaler, J. (2009). Substrate incorporation patterns of bacterioplankton populations in stratified and mixed waters of a humic lake, 11, 1854–1865. doi:10.1111/j.1462-2920.2009.01910.x

Burkert, U., Warnecke, F., Babenzien, D., Pernthaler, J., & Zwirnmann, E. (2003). Members of a Readily Enriched β -Proteobacterial Clade Are Common in Surface Waters of a Humic Lake Members of a Readily Enriched ɴʟ -Proteobacterial Clade Are Common in Surface Waters of a Humic Lake. doi:10.1128/AEM.69.11.6550

Corno, G., Modenutti, B. E., Callieri, C., Balseiro, E. G., Bertoni, R., & Caravati, E. (2009). Bacterial diversity and morphology in deep ultraoligotrophic Andean lakes: The role of UVR on vertical distribution. Limnology and Oceanography, 54(4), 1098–1112. doi:10.4319/lo.2009.54.4.1098

Debroas, D., Humbert, J.-F., Enault, F., Bronner, G., Faubladier, M., & Cornillot, E. (2009). Metagenomic approach studying the taxonomic and functional diversity of the bacterial community in a mesotrophic lake (Lac du Bourget--France). Environmental Microbiology, 11(9), 2412–24. doi:10.1111/j.1462-2920.2009.01969.x

Dziallas, C., & Grossart, H.-P. (2011). Temperature and biotic factors influence bacterial communities associated with the cyanobacterium Microcystis sp. Environmental Microbiology, 13(6), 1632–41. doi:10.1111/j.1462-2920.2011.02479.x

Eckert, E. M., Salcher, M. M., Posch, T., Eugster, B., & Pernthaler, J. (2012). Rapid successions affect microbial N-acetyl-glucosamine uptake patterns during a lacustrine spring phytoplankton bloom, 14, 794–806. doi:10.1111/j.1462-2920.2011.02639.x

Eiler, A., & Bertilsson, S. (2004). Composition of freshwater bacterial communities associated with cyanobacterial blooms in four Swedish lakes. Environmental Microbiology, 6(12), 1228–43. doi:10.1111/j.1462-2920.2004.00657.x

Garcia, S. L., McMahon, K. D., Martinez-Garcia, M., Srivastava, A., Sczyrba, A., Stepanauskas, R., Grossart, H.-P., et al. (2012). Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. The ISME journal, 1–11. doi:10.1038/ismej.2012.86

Garcia, S. L., McMahon, K. D., Grossart, H.-P., & Warnecke, F. (2013). Successful enrichment of the ubiquitous freshwater acI A ctinobacteria. Environmental Microbiology Reports, n/a–n/a. doi:10.1111/1758-2229.12104

Ghai, R., Rodriguez-Valera, F., McMahon, K. D., Toyama, D., Rinke, R., Cristina Souza de Oliveira, T., … Henrique-Silva, F. (2011). Metagenomics of the water column in the pristine upper course of the Amazon river. PloS One, 6(8), e23785. doi:10.1371/journal.pone.0023785

Ghai, R., K. D. McMahon, et al. (2012). "Breaking a paradigm: cosmopolitan and abundant freshwater actinobacteria are low GC." Environmental Microbiology Reports 4(1): 29-35.

Glöckner, F. O., Fuchs, B. M., Glo, F. O., & Amann, R. (1999). Bacterioplankton Compositions of Lakes and Oceans : a First Comparison Based on Fluorescence In Situ Hybridization Bacterioplankton Compositions of Lakes and Oceans : a First Comparison Based on Fluorescence In Situ Hybridization.

Glöckner, F. O., Zaichikov, E., Belkova, N., Denissova, L., Pernthaler, J., Pernthaler, A., Amann, R., et al. (2000). Comparative 16S rRNA Analysis of Lake Bacterioplankton Reveals Globally Distributed Phylogenetic Clusters Including an Abundant Group of Actinobacteria Comparative 16S rRNA Analysis of Lake Bacterioplankton Reveals Globally Distributed Phylogenetic Cluster. doi:10.1128/AEM.66.11.5053-5065.2000.Updated

Hahn, M. W., Lünsdorf, H., Wu, Q., Höfle, M. G., Boenigk, J., Stadler, P., … Ho, M. G. (2003). Isolation of Novel Ultramicrobacteria Classified as Actinobacteria from Five Freshwater Habitats in Europe and Asia Isolation of Novel Ultramicrobacteria Classified as Actinobacteria from Five Freshwater Habitats in Europe and Asia. doi:10.1128/AEM.69.3.1442

Hahn, M. W., Pöckl, M., & Po, M. (2005). Ecotypes of Planktonic Actinobacteria with Identical 16S rRNA Genes Adapted to Thermal Niches in Temperate , Subtropical , and Tropical Freshwater Habitats Ecotypes of Planktonic Actinobacteria with Identical 16S rRNA Genes Adapted to Thermal Niches in Te. doi:10.1128/AEM.71.2.766

Haukka, K., Kolmonen, E., Hyder, R., Hietala, J., Vakkilainen, K., Kairesalo, T., … Sivonen, K. (2006). Effect of nutrient loading on bacterioplankton community composition in lake mesocosms. Microbial Ecology, 51(2), 137–46. doi:10.1007/s00248-005-0049-7

Humbert, J.-F., Dorigo, U., Cecchi, P., Le Berre, B., Debroas, D., & Bouvy, M. (2009). Comparison of the structure and composition of bacterial communities from temperate and tropical freshwater ecosystems. Environmental Microbiology, 11(9), 2339–50. doi:10.1111/j.1462-2920.2009.01960.x

Iriberri, J., Azúa, I., Labirua-Iturburu, a, Artolozaga, I., & Barcina, I. (1994). Differential elimination of enteric bacteria by protists in a freshwater system. The Journal of Applied Bacteriology, 77(5), 476–83. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8002473

Jezbera, J., Hornák, K., & Simek, K. (2005). Food selection by bacterivorous protists: insight from the analysis of the food vacuole content by means of fluorescence in situ hybridization. FEMS microbiology ecology, 52(3), 351–63. doi:10.1016/j.femsec.2004.12.001

Jezbera, J., Hornák, K., & Simek, K. (2006). Prey selectivity of bacterivorous protists in different size fractions of reservoir water amended with nutrients. Environmental microbiology, 8(8), 1330–9. doi:10.1111/j.1462-2920.2006.01026.x

Jezbera, J., Sharma, A. K., Brandt, U., Doolittle, W. F., & Hahn, M. W. (2009). "Candidatus Planktophila limnetica", an actinobacterium representing one of the most numerically important taxa in freshwater bacterioplankton. International Journal of Systematic and Evolutionary Microbiology, 59(Pt 11), 2864–9. doi:10.1099/ijs.0.010199-0

Jones, S. E., Newton, R. J., & McMahon, K. D. (2009). Evidence for structuring of bacterial community composition by organic carbon source in temperate lakes. Environmental Microbiology, 11(9), 2463–72. doi:10.1111/j.1462-2920.2009.01977.x

Kolmonen, E., Sivonen, K., Rapala, J., & Haukka, K. (2004). Diversity of cyanobacteria and heterotrophic bacteria in cyanobacterial blooms in Lake Joutikas, Finland. Aquatic Microbial Ecology, 36, 201–211. doi:10.3354/ame036201

Lindström, E. S., Agterveld, M. P. K., Lindstro, E. S., & Zwart, G. (2005). Distribution of Typical Freshwater Bacterial Groups Is Associated with pH , Temperature , and Lake Water Retention Time Distribution of Typical Freshwater Bacterial Groups Is Associated with pH , Temperature , and Lake Water Retention Time. doi:10.1128/AEM.71.12.8201

Martinez-Garcia, M., Swan, B. K., Poulton, N. J., Gomez, M. L., Masland, D., Sieracki, M. E., & Stepanauskas, R. (2011). High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. The ISME Journal, 6(1), 113–23. doi:10.1038/ismej.2011.84

Maurin, N., Amblard, C., & Bourdier, G. (1997). Phytoplanktonic excretion and bacterial reassimilation in an oligomesotrophic lake: molecular weight fractionation. Journal of Plankton Research, 19(8), 1045–1068. doi:10.1093/plankt/19.8.1045

Newton, R. J., Kent, A. D., Triplett, E. W., & McMahon, K. D. (2006). Microbial community dynamics in a humic lake: differential persistence of common freshwater phylotypes. Environmental Microbiology, 8(6), 956–70. doi:10.1111/j.1462-2920.2005.00979.x

Newton, R. J., Jones, S. E., Helmus, M. R., & McMahon, K. D. (2007). Phylogenetic ecology of the freshwater Actinobacteria acI lineage. Applied and environmental microbiology, 73(22), 7169–76. doi:10.1128/AEM.00794-07

Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., & Bertilsson, S. (2011). A guide to the natural history of freshwater lake bacteria. Microbiology and molecular biology reviews : MMBR (Vol. 75, pp. 14–49). doi:10.1128/MMBR.00028-10

Paerl, H. W. (2009). Paerl 2009 Estuaries Controlling Eutrophication.pdf. Estuaries and Coasts, 32, 593–601.

Pearce, D. a, Gast, C. J., Lawley, B., & Ellis-Evans, J. C. (2003). Bacterioplankton community diversity in a maritime Antarctic lake, determined by culture-dependent and culture-independent techniques. FEMS Microbiology Ecology, 45(1), 59–70. doi:10.1016/S0168-6496(03)00110-7

Pérez, M. T., Hörtnagl, P., & Sommaruga, R. (2010). Contrasting ability to take up leucine and thymidine among freshwater bacterial groups: implications for bacterial production measurements. Environmental Microbiology, 12(1), 74–82. doi:10.1111/j.1462-2920.2009.02043.x

Pernthaler, J., Posch, T., Karel, S., Pernthaler, A., Glöckner, F. O., Psenner, R., … Vrba, J. (2001). Predator-Specific Enrichment of Actinobacteria from a Cosmopolitan Freshwater Clade in Mixed Continuous Culture Predator-Specific Enrichment of Actinobacteria from a Cosmopolitan Freshwater Clade in Mixed Continuous Culture. doi:10.1128/AEM.67.5.2145

Pernthaler, J., & Amann, R. (2005). Fate of Heterotrophic Microbes in Pelagic Habitats : Focus on Populations Fate of Heterotrophic Microbes in Pelagic Habitats : Focus on Populations, 69(3). doi:10.1128/MMBR.69.3.440

Richardot, M., Debroas, D., Thouvenot, A., Sargos, D., Berthon, J. L., & Dévaux, J. (2001). Influence of cladoceran grazing activity on dissolved organic matter , enzymatic hydrolysis and bacterial growth.

Salcher, M. M., Pernthaler, J., & Posch, T. (2010). Spatiotemporal distribution and activity patterns of bacteria from three phylogenetic groups in an oligomesotrophic lake. Limnology and Oceanography, 55(2), 846–856. doi:10.4319/lo.2009.55.2.0846

Shade, A., Read, J. S., Youngblut, N. D., Fierer, N., Knight, R., Kratz, T. K., … McMahon, K. D. (2012). Lake microbial communities are resilient after a whole-ecosystem disturbance. The ISME Journal, 6(12), 2153–67. doi:10.1038/ismej.2012.56

Sharma, A. K., Zhaxybayeva, O., Papke, R. T., & Doolittle, W. F. (2008). Actinorhodopsins: proteorhodopsin-like gene sequences found predominantly in non-marine environments. Environmental Microbiology, 10(4), 1039–56. doi:10.1111/j.1462-2920.2007.01525.x

Sharma, A. K., Sommerfeld, K., Bullerjahn, G. S., Matteson, A. R., Wilhelm, S. W., Jezbera, J., Brandt, U., et al. (2009). Actinorhodopsin genes discovered in diverse freshwater habitats and among cultivated freshwater Actinobacteria. The ISME journal, 3(6), 726–37. doi:10.1038/ismej.2009.13

Simek, K., Weinbauer, M. G., Hornák, K., Jezbera, J., Nedoma, J., & Dolan, J. R. (2006). Grazer and virus-induced mortality of bacterioplankton accelerates development of Flectobacillus populations in a freshwater community. Environmental Microbiology, 9(3), 789–800. doi:10.1111/j.1462-2920.2006.01201.x

Taipale, S., Jones, R., & Tiirola, M. (2009). Vertical diversity of bacteria in an oxygen-stratified humic lake, evaluated using DNA and phospholipid analyses. Aquatic Microbial Ecology, 55(April), 1–16. doi:10.3354/ame01277

Tarao, M., Jezbera, J., & Hahn, M. W. (2009). Involvement of cell surface structures in size-independent grazing resistance of freshwater Actinobacteria. Applied and environmental microbiology, 75(14), 4720–6. doi:10.1128/AEM.00251-09

Warnecke, F., Amann, R., & Pernthaler, J. (2004). Actinobacterial 16S rRNA genes from freshwater habitats cluster in four distinct lineages. Environmental Microbiology, 6(3), 242–253. doi:10.1111/j.1462-2920.2004.00561.x

Warnecke, F., Sommaruga, R., Sekar, R., Julia, S., Pernthaler, J., & Hofer, J. S. (2005). Abundances , Identity , and Growth State of Actinobacteria in Mountain Lakes of Different UV Transparency Abundances , Identity , and Growth State of Actinobacteria in Mountain Lakes of Different UV Transparency. doi:10.1128/AEM.71.9.5551

Wever, A. De, Muylaert, K., Der, K. Van, Pirlot, S., Cocquyt, C., Descy, J., … Plisnier, P. (2005). Bacterial Community Composition in Lake Tanganyika : Vertical and Horizontal Heterogeneity Bacterial Community Composition in Lake Tanganyika : Vertical and Horizontal Heterogeneity. doi:10.1128/AEM.71.9.5029

Wu, Q. L., Zwart, G., Schauer, M., Kamst-van Agterveld, M. P., & Hahn, M. W. (2006). Bacterioplankton community composition along a salinity gradient of sixteen high-mountain lakes located on the Tibetan Plateau, China. Applied and Environmental Microbiology, 72(8), 5478–85. doi:10.1128/AEM.00767-06

Wu, X., Xi, W., Ye, W., & Yang, H. (2007). Bacterial community composition of a shallow hypertrophic freshwater lake in China, revealed by 16S rRNA gene sequences. FEMS Microbiology Ecology, 61(1), 85–96. doi:10.1111/j.1574-6941.2007.00326.x

Wurzbacher, C., Salka, I., & Grossart, H.-P. (2012). Environmental actinorhodopsin expression revealed by a new in situ filtration and fixation sampler. Environmental Microbiology Reports, 4(5), 491–497. doi:10.1111/j.1758-2229.2012.00350.x

Zeder, M., Peter, S., Shabarova, T., & Pernthaler, J. (2009). A small population of planktonic Flavobacteria with disproportionally high growth during the spring phytoplankton bloom in a prealpine lake. Environmental microbiology, 11(10), 2676–86. doi:10.1111/j.1462-2920.2009.01994.x

Zwart, G., Crump, B., Kamst-van Agterveld, M., Hagen, F., & Han, S. (2002). Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. Aquatic Microbial Ecology, 28, 141–155. doi:10.3354/ame028141

2. INTRODUCTION

The acI are an intriguing group of tiny free-living bacteria that dominate the fresh waters of the planet. However, little is known about these bacteria and the role they play in water quality and the global carbon cycle (Warnecke, 2004; Lindstrom, 2005). They are one of the smallest known free-living bacteria and they have the capability to harvest sunlight with a novel rhodopsin protein (Sharma, 2009), even though they are heterotrophs. These bacteria are surprisingly resilient to seasonal environmental changes, remaining abundant in summer months as well as winter. They have proven to be successful in widely varying habitats from high nutrient eutrophic settings to low nutrient oligotrophic lakes and even estuaries and glacial meltwater. However, the reasons for the success of acI remain a mystery. The acI bacteria likely play a critical role in the health of our freshwater systems, yet little is known about them because they have been difficult to grow in cultures that can be studied in a lab setting (Newton, 2011).

Freshwater systems and the commerce they support are threatened by pollution and landscape change that have resulted in public health hazards (harmful algal blooms), reduced aesthetics (decreased water clarity and odors from decaying algae and plant matter) and dead zones that are inhospitable to fish (Carpenter, 1998; Paerl, 2009). In order to better understand these issues, we must understand the bacteria that mediate nutrient cycling and algae growth in lakes (Lindeman, 1941; F. Azam, 1983 ; Biddanda, 2002). The most abundant of these bacteria is the acI lineage (Newton, 2007).

Additionally, carbon emissions in freshwater systems play a significant role in climate change through the oxidation, storage, and release of terrestrially derived and locally produced carbon. About 1.9 Pg C y_{-1} is delivered to inland waters from the terrestrial landscape, of which at least 0.75 is released to the atmosphere as gas exchange and 0.23 is buried in sediments. The total amount of organic matter stored in the sediments of lakes and reservoirs is estimated to exceed that stored in the sediments of the world's oceans (Cole, 2007). Microbial communities drive the flow of carbon through

lakes by processing dissolved organic carbon (DOC), breaking down particulate organic carbon (POC), fixing and respiring large quantities of CO2, quenching methane produced in sediments, and sequestering carbon in biomass. As the most abundant freshwater bacteria, the acI bacteria likely play an important role in carbon cycling.

The McMahon lab has studied the acI lineage for several years and has observed interesting patterns in population dynamics and metabolic potential (Newton, 2011; Garcia, 2012; Ghai, 2012). Based on preliminary analysis of qPCR and genetic data, we hypothesize that the acI have unique traits that provide a competitive advantage when scavenging for carbon and nitrogen during nutrient poor conditions. Our further analyses in this dissertation shed light on the mechanisms of acI's success and its role in the freshwater environment.

REFERENCES

Azam, F., & Graf, J. S. (1983). The Ecological Role of Water-Column Microbes in the Sea *, *10*, 257 263.

Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley AN, & VH, S. (1998). Nonpoint Pollution of Surface Waters with Phosphorus and Nitrogen. Ecological Society of America, 8(3), 559–568.

Cole, J. J., Caraco, W. H. McDowell, L. J. Tranvik, R. G. Striegl,C. M. Duarte, P. Kortelainen, J. A. Downing, J. J. Middelburg and J. MelackReviewed (2007). "Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon." Ecosystems, 10(1): 171-184.

Cotner, J. B., & Biddanda, B. A. (2002). Players , Large Role : Biogeochem Pelagic Aquatic Ecosystems, 5(2), 105–121.

Garcia, S. L., McMahon, K. D., Martinez-Garcia, M., Srivastava, A., Sczyrba, A., Stepanauskas, R., Grossart, H.-P., et al. (2012). Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. The ISME journal, 1–11. doi:10.1038/ismej.2012.86

Ghai, R., K. D. McMahon, et al. (2012). "Breaking a paradigm: cosmopolitan and abundant freshwater actinobacteria are low GC." Environmental Microbiology Reports 4(1): 29-35.

Lindeman, R. (1941). Seasonal Food-Cycle Dynamics in a Senescent Lake. American Midland Naturalist, 26(3), 636–673.

Lindström, E. S., Agterveld, M. P. K., Lindstro, E. S., & Zwart, G. (2005). Distribution of Typical Freshwater Bacterial Groups Is Associated with pH , Temperature , and Lake Water Retention Time Distribution of Typical Freshwater Bacterial Groups Is Associated with pH , Temperature , and Lake Water Retention Time. doi:10.1128/AEM.71.12.8201

Newton, R. J., Jones, S. E., Helmus, M. R., & McMahon, K. D. (2007). Phylogenetic ecology of the freshwater Actinobacteria acI lineage. Applied and environmental microbiology, 73(22), 7169–76. doi:10.1128/AEM.00794-07

Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., & Bertilsson, S. (2011). A guide to the natural history of freshwater lake bacteria. Microbiology and molecular biology reviews : MMBR (Vol. 75, pp. 14–49). doi:10.1128/MMBR.00028-10

Paerl, H. W. (2009). Paerl 2009 Estuaries Controlling Eutrophication.pdf. Estuaries and Coasts, 32, 593–601.

Sharma, A. K., Sommerfeld, K., Bullerjahn, G. S., Matteson, A. R., Wilhelm, S. W., Jezbera, J., Brandt, U., et al. (2009). Actinorhodopsin genes discovered in diverse freshwater habitats and among cultivated freshwater Actinobacteria. The ISME journal, 3(6), 726–37. doi:10.1038/ismej.2009.13

Warnecke, F., Sommaruga, R., Sekar, R., Julia, S., Pernthaler, J., & Hofer, J. S. (2005). Abundances , Identity , and Growth State of Actinobacteria in Mountain Lakes of Different UV Transparency Abundances , Identity , and Growth State of Actinobacteria in Mountain Lakes of Different UV Transparency. doi:10.1128/AEM.71.9.5551

3. FRESHWATER ACTINOBACTERIA (ACI LINEAGE) SINGLE-CELL GENOMICS

Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage

Trevor W Ghylin[1], Sarahi L Garcia[2,4], Francisco Moya[1], Ben O Oyserman[1], Patrick Schwientek[3],

Katrina T. Forest[2], James Mutschler[2], Leong-Keat Chan[3], Manuel Martinez-Garcia[5,9], Alexander

Sczyrba[3,7], Ramunas Stepanauskas[5], Hans-Peter Grossart[6,8], Tanja Woyke[3], Falk Warnecke[4], Rex

Malmstrom[3], Stefan Bertilsson[10], Katherine D McMahon[1,2]

1 Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, USA
2 Department of Bacteriology, University of Wisconsin-Madison Madison, WI, USA
3 DOE Joint Genome Institute, Walnut Creek, CA, USA
4 Jena School for Microbial Communication (JSMC) and Microbial Ecology Group at Friedrich Schiller University Jena, Jena, Germany
5 Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, USA
6 Department of Limnology of Stratified Lakes, Leibniz-Institute for Freshwater Ecology and Inland Fisheries, Stechlin, Germany
7 Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany
8 Institute for Biochemistry and Biology, Postdam University, Postdam, Germany
9 Current address: University of Alicante, Alicante, Spain.
10 Department of Ecology and Genetics, Limnology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden

* correspondence: Katherine McMahon, Departments of Civil and Environmental Engineering, and

Bacteriology, University of Wisconsin 1550 Linden Drive Room 5552 Madison, WI 53706, Email:

kdmcmahon@wisc.edu

Final Submission June 2014 to ISME Journal for Publication

Full citation:

Ghylin, T. W.; Garcia, S. L.; Moya, F.; Oyserman, B. O.; Schwientek, P.; Forest, K. T.; Mutschler, J.; Chan, L. K.; Martinez-Garcia, M.; Sczyrba, A.; Stepanauskas, R.; Grossart, H. P.; Woyke, T.; Warnecke, F.; Malmstrom, R.; Bertilsson, S.; McMahon, K. D. (2014). Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage. ISME. 10.1038/ismej.2014.135

Contributions: Trevor Ghylin performed most of the genomic analysis and drafted the manuscript with major review and editing provided by Sarahi Garcia and Katherine McMahon. Francisco Moya constructed the pathways diagram. Ben Oyserman performed analysis of codon adaptive index and C:N ratio in the genome, much of which was removed from the final manuscript. Patrick Schwientek performed analysis of genome completeness. Katrine Forest analyzed the genes and coded proteins associated with and adjacent to actinorhodopsin in the genomes. James Mutschler provided some metabolic analysis on sulfur and phosphorus uptake. Leong-Keat Chan coordinated others and the Joint Genome Institute for their involvement in this paper. Manuel-Martinez Garcia, Alexander Sczyrba, and Ramunas Stepanauskas were involved in the lab methods for isolating, amplifying and sequencing the single cells in this study. Hans-Peter Grossart provided an additional acI genome to add to this study. Tanja Woyke and Rex Malmstrom were the lead managers at the Joint Genome Institute managing the sequencing, annotation and some genomic analysis. Falk Warnecke and Stefan Bertilsson provided expert lake microbiology input and guidance. Katherine D McMahon was the primary investigator and advisor of Trevor Ghylin.

3.1. ABSTRACT

Members of the acI lineage of Actinobacteria are the most abundant microorganisms in most freshwater lakes, however, our understanding of the keys to their success and their role in carbon and nutrient cycling in freshwater systems has been hampered by the lack of pure cultures and genomes. We obtained draft genome assemblies from 11 single cells representing three acI tribes (acI-A1, acI-A7, acI-B1) from four temperate lakes in the United States and Europe. Comparative analysis among tribes showed that members of the cosmopolitan acI-B1 tribe have the smallest estimated genome sizes as well as the lowest GC% and average codon usage bias across coding sequences. The acI genomes contain actinorhodopsin as well as some genes involved in anaplerotic carbon fixation indicating the capacity to supplement their known heterotrophic lifestyle. Although nitrogen can be scarce in freshwater systems, the acI genomes do not display strong evolutionary selection for nitrogen conservation as the N:C ratio in the protein coding genes is average as compared to other sequenced bacterial genomes. This suggests that their mechanisms for uptake of amino acids, polyamines and cyanophycin fulfill their nitrogen requirements. Genome-level differences between the acI-A and acI-B clades suggest specialization at the clade level for carbon substrate acquisition. Overall, the acI genomes appear to be highly streamlined versions of Actinobacteria that include some genes allowing it to take advantage of the light and N-rich organic compounds. This work significantly expands the known metabolic potential of the cosmopolitan freshwater acI lineage and its ecological and genetic traits.

3.2. INTRODUCTION

Members of the acI lineage within the phylum Actinobacteria are an intriguing group of free-living ultramicrobacteria that dominate many freshwater ecosystems (Newton et al. 2011), including high nutrient eutrophic (Wu et al. 2007), low nutrient oligotrophic (Humbert et al. 2009) and dystrophic lakes (Newton et al. 2006). They are also abundant in some marine estuaries (Glockner et al. 2000).

Moreover, as a lineage, acI shows a smaller seasonal abundance variation when compared to other major freshwater bacteria (Newton et al. 2011; Salcher et al. 2010; Eckert et al. 2012; Rösel & Grossart 2012; Allgaier & Grossart 2006). Due to their high abundance and high metabolic activity (Salcher et al. 2010; Warnecke et al. 2005; Allgaier & Grossart 2006), the acI bacteria likely play a critical role in carbon and other nutrient cycling in freshwater systems, yet their genomic features remain elusive. This is in part due to the historical lack of isolated representatives and limited genomic data.

Based on 16S rRNA gene sequence analysis, the acI lineage comprises three distinct clades (A, B, and C) (Newton et al. 2011). Each of these clades has been further subdivided into "tribes," members of which share ≥97% 16S rRNA gene sequence identity (in total 13 such tribes). The most abundant and prevalent tribes are acI-A1, acI-A6, acI-A7, and acI-B1 (Newton et al. 2011). The tribes seem to niche-partition based on pH, with some clades (e.g. acI-A1 and acI-B2) preferring slightly more acidic environments (Newton et al. 2007).

Several single-cell targeted studies have attempted to elucidate acI's capability to take up and consume specific carbon sources, using fluorescent in situ hybridization (FISH) combined with catalyzed reporter deposition (CARD) and microautoradiography (MAR). These studies have shown that acI members can take up glucose (Buck et al. 2009), leucine (Buck et al. 2009; Eckert et al. 2012; Perez et al. 2010; Salcher et al. 2010), acetate (Buck et al. 2009), thymidine (Perez et al. 2010), N-acetyl-glucosamine (NAG) (Beier & Bertilsson 2011; Eckert et al. 2012) and di-NAG (Beier & Bertilsson 2011; Eckert et al. 2012; Tada & Grossart 2013). A recent MAR-FISH study performed on fall samples from Lake Zürich in Switzerland showed that acI bacteria consumed an amino acid mixture as well (Salcher et al. 2013). The acI in this study did not take up acetate, fructose, arginine, aspartate, glutamate, glutamine, serine, glycine, or alanine. However, due to the inherent limit of phylogenetic resolution associated with rRNA-targeted FISH probes, many of these studies generally

cannot attribute the ability to uptake key substrates to clades or tribes within the lineage. Only a few studies have used FISH probes that differentiate between members of the acI-A and acI-B clades. One such study, performed by Buck and colleagues, showed evidence of substrate-based niche partitioning, where acI-B-positive cells could consume acetate but acI-A-positive cells could not (Buck et al. 2009).

A few recent studies have used (meta)genomics-based analysis to study acI Actinobacteria. Ghai and colleagues analyzed metagenomic data from lakes, estuaries (Ghai et al. 2012), and rivers (Ghai et al. 2011), and found that acI members have a lower than expected GC genomic content. Martinez-Garcia *et al* (Martinez-Garcia et al. 2012) used single-cell genomics and found that the origins of more than 80% of rhodopsins found among lake bacteria belong to the actinobacteria. Another study reported the first nearly complete acI genome from the acI-B1 tribe obtained using single-cell genomics (Garcia et al. 2013). In short, the genome was small (estimated to be just over 1 Mbp) and low GC (~42%). Metabolic reconstruction indicated that members of acI-B1 are facultative aerobes with a capability for taking up and metabolizing pentoses such as xylose. The authors also confirmed an actinorhodopsin gene in the genome.

Here we greatly expand on the single-cell genome based analysis (Garcia et al. 2013) by analyzing ten additional acI single amplified genomes (SAGs) from three different acI tribes and four different lakes. We also compare the gene content of acI genomes to that of cultivated members of their parent order, Actinomycetales, and to that of other abundant freshwater bacteria. Our data show that acI members have unique traits that likely provide a competitive advantage when scavenging for energy, carbon, nitrogen and phosphorus in freshwater habitats. The presence of several genes involved in the uptake and metabolism of organic nitrogen compounds suggests that N-rich organic matter may be a significant source of carbon, nitrogen, and energy for acI biosynthesis.

3.3. MATERIALS AND METHODS

3.3.1. SINGLE AMPLIFIED GENOME (SAG) GENERATION AND SELECTION

Water samples (1-ml) were collected from the upper 0.5m to 1m of each of four lakes (Mendota, Sparkling, Damariscotta, Stechlin) and cryopreserved (Table S1), as previously described (Martinez-Garcia et al. 2012; Garcia et al. 2013) (Supplementary Online Material). Bacterial single amplified genomes (SAGs) were generated and identified at the Bigelow Laboratory Single Cell Genomics Center (SCGC; http://www.bigelow.org/scgc), as detailed in (Martinez-Garcia et al. 2012) (Table 1).

Ten SAGs from lakes Mendota, Sparkling and Damariscotta were selected during the SAG library screening step described in (Martinez-Garcia et al. 2012). Partial 16S rRNA genes amplified previously (Martinez-Garcia et al. 2012) were phylogenetically classified to the freshwater "tribe" level by insertion into reference trees in the ARB software package (Newton et al. 2011; Ludwig et al. 2004). SAGs were selected in order to compare among acI tribes and source lakes. The one SAG from Lake Stechlin was selected from a separate library constructed at the SCGC as described in (Garcia et al. 2013) and its 16S rRNA gene was 100% identical to the acI-B1 SAG previously analyzed (AAA027-L06). Phylogenetic reconstruction was conducted by maximum likelihood (RAxML) (Stamatakis et al, 2008) with 1000 bootstrap runs on the CIPRES web portal (www.phylo.org) using near full length reference 16S rRNA gene sequences from a manually curated alignment (Newton et al, 2011) and a 50% base frequency filter (total 1402 positions) (Figure 1). For four SAGs only short fragments (~400 bp) were available. Bootstrap values are indicated above nodes with greater than 50% support and the scale bar represents 10 base substitutions per 100 nt positions.

3.3.2. GENOME SEQUENCING, ASSEMBLY, CONTAMINATION DETECTION AND ANNOTATION

The draft assemblies of the single-cell genomes were generated at the DOE Joint genome

Institute (JGI) using the Illumina technology. An Illumina standard shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform by pooling libraries for approximately ten SAGs per lane. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov/. Raw Illumina sequence data were filtered for known Illumina sequencing and library preparation artifacts and then screened and trimmed according to the k-mers present in the dataset. Reads representing highly abundant k-mers were removed such that no k-mers with a coverage of more than 30x were present after filtering. Contigs with an average k-mer depth of less than 2x were removed. The following steps were then performed for assembly: (1) filtered Illumina reads were assembled using Velvet version 1.1.04 (Zerbino & Birney 2008). The VelvetOptimiser script (version 2.1.7) was used with default optimization functions (n50 for k-mer choice, total number of base pairs in large contigs for cov_cutoff optimization). (2) 1 to 3 kbp simulated paired end reads were created from Velvet contigs using the wgsim software. (3) the normalized Illumina reads were assembled together with simulated read pairs using Allpaths-LG (version 41043) (Gnerre et al. 2011).

We employed a combination of tetramer principal component analysis and blast searches against reference databases to identify contigs that may originate from DNA contaminants (Woyke et al. 2009). No putative contaminants were found in any of the assemblies.

Genes were identified using Prodigal (Hyatt et al. 2010). The predicted CDSs (coding DNA sequences) were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database (nr), UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAScan-SE tool (Hacker & Kaper 2000) was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA (Pruesse et al. 2007). Other non–coding RNAs such as the RNA components of the protein secretion complex and the RNase P were identified by searching genomes for the corresponding Rfam profiles using INFERNAL (Makarova et al. 1999). Additional gene prediction analysis and manual functional

annotation was performed within the Integrated Microbial Genomes (IMG) (Markowitz et al. 2012)

platform developed by the Joint Genome Institute, Walnut Creek, CA, USA (http://img.jgi.doe.gov).

3.3.3. GENOME COMPLETENESS AND SIZE ESTIMATES

Genome size and completeness were estimated using a conserved single copy gene (CSCG) set

that has been determined from all finished actinobacterial genome sequences (n=151) in the IMG

database (Markowitz et al. 2012). The set consists of 158 CSCGs that were found to occur only once in

at least 95% of all genomes by analysis of an abundance matrix based on hits to the protein family

(Pfam) database (Punta et al. 2012). Hidden Markov models of the identified Pfams (Table S2) were

used to search all SAG assemblies by means of the HMMER3 software (Eddy 2011). Resulting best

hits above the trusted cutoff (TC field as provided in the HMM files from Pfam) were counted and the

completeness was estimated as the ratio of found CSCG to total CSCGs in the set after normalization to

95%. The estimated complete genome size was then calculated by dividing the estimated genome

coverage by the total assembly size.

3.3.4. COMPARATIVE ANALYSIS

Annotations with the Clusters of Orthologous Groups (COGs) of proteins for each SAG and

comparison genome from the Actinomycetales, *Polynucleobacter, Limnohabitans,* and LD12 were

downloaded from the IMG website (http://img.jgi.doe.gov/) (Table S3). A presence/absence COG list

was generated and used to determine COG prevalence (% of genomes containing a certain COG). The

data were used to determine COG prevalence (% of genomes containing a certain COG) in acI as well

as the other organisms of interest. The COG list was then sorted based on difference in prevalence (%

prevalence in acI minus % prevalence in other group) between acI and a group of interest. COGs at the

top of this sorted list had the greatest difference in prevalence between acI and the group of interest and

were considered over-represented in acI.

Average protein identity (API) between SAG pairs was calculated using all-versus-all blastp from predicted coding sequences in all 11 SAGs to identify best hits for each SAG pair. Results were parsed using custom perl scripts to identify reciprocal blast best hits that had alignments over at least 50% of both the query and the subject sequence lengths. Averages were calculated for each pair based on the reciprocal blast best hit results.

3.3.5. CODON BIAS

For each genome we calculated the frequency of each amino acid codon out of the total codons encoding the corresponding amino acid, once in the ribosomal protein genes and once in the rest of the genes. We then calculated the average difference between codon frequencies in the ribosomal proteins and in the rest of the proteins in the genomes. In order to calculate the genome CAI_{ave} (Codon Adaptation Index) for each organism we computed the average of CAI values over all genes in the genome (Rocha 2004). The CAI of a gene was computed as described by (Sharp & Li 1987).

3.4. RESULTS AND DISCUSSION

3.4.1. GENOME STATISTICS AND PHYLOGENETIC AFFILIATION

We sequenced ten acI single amplified genomes (SAGs) and analyzed them along with the one previously sequenced SAG (Garcia et al. 2013). These eleven SAGs belong to three different acI tribes and four different lakes. Our genome completeness estimates ranged from 34% to a nearly complete genome. Back calculating the estimated genome size of the SAGs it could be observed that all eleven of them were small (1 - 2 Mbp) and had relatively low GC% (40 to 48%). This low GC% had been observed already in metagenomic assemblies (Ghai et al. 2012). The previously analyzed acI-B1 genome also determined low GC content (~42%) and small genome size (~1.16 Mbp) (Garcia et al. 2013). Our comparative genomic analysis confirms that these features are broadly representative of the acI lineage and further distinguishes among the acI tribes, with acI-B1 having the lowest GC and

smallest genome (Table 1 and Table S4). In general, the average genome completeness for the acI SAGs was 68% (based on the fraction of CSCGs). Therefore, we would expect to encounter a core, single-copy gene in about seven of the 11 SAGs.

3.5. ACI METABOLISM: HOW DOES THE ACI LINEAGE MAKE A LIVING IN FRESHWATER?

3.5.1. CARBON AND ENERGY

The previously published AAA027-L06 acI-B1 genome indicated a facultative aerobic lifestyle with mostly complete archetypal central metabolism (glycolysis, pentose phosphate pathway, and citrate cycle), oxidative phosphorylation machinery, and the ability to ferment pyruvate (Garcia et al. 2013). These and other features were largely conserved among our eleven sequenced acI SAGs (Figure 2), with exceptions attributable either to incomplete genome recovery or niche diversification among clades or tribes (discussed further below). As noted for AAA027-L06, the SAGs seem capable of metabolizing glucose but lack an obvious glucose transport system. This is curious in light of FISH-MAR based studies clearly demonstrating the ability of acI to incorporate glucose (Salcher et al. 2013; Buck et al. 2009). However, most SAGs did contain ABC-type sugar transport components (e.g. COG1653) (Table 2 and Table S5) and using the Transporter Classification Database (Saier et al. 2009), a periplasmic component with closest match to the glucose-binding protein in *Thermus thermophilus* (TC#3.A.1.1.24) was found in eight SAGs but was missing in AAA027-L06. Thus, we cannot confirm nor refute the ability of all acI members to take up glucose based on genomic evidence alone. Other transporters that the eleven acI SAGs have in common, include ABC-type transporters for ribose/xylose/arabinose/galactoside, polyamines, dipeptides, and branched chain amino acids (Table 2). The polyamine transporters are likely being used for putrescine uptake, since the SAGs harbor downstream pathways for its eventual conversion to succinate via the transamination pathway (Dasu et al. 2006; Chou et al. 2008) (Figure 2). Most SAGs also contained a putative cyanophycinase

(COG4242), allowing acI to access this C- and N-storage compound synthesized by cyanobacteria (discussed further below). Genes for the uptake of carboxylic acids in acI-A1, acI-A7 and acI-B1 were not found in any of the sequenced genomes, consistent with FISH-MAR based studies (Buck et al. 2009; Salcher et al. 2013). We could not identify machinery to support the ability to take up and incorporate N-acetyl glucosamine (NAG) (Beier & Bertilsson 2011), but this could mainly be because the pathways involved have not been identified. It is probable that one of the many COGs involved in the uptake of carbohydrates enables NAG uptake (Table 2).

Others have found that acI abundances change through the seasons with maxima in spring and fall (Allgaier & Grossart 2006) and that acI abundance positively correlates with solar radiation (Warnecke et al. 2005), suggesting they gain some benefit from higher light intensities either through phototrophy or consumption of photochemically produced labile dissolved organic carbon (DOC). The latter possibility is contradicted by the apparent inability of acI to utilize carboxylic acids (see above), a major product of photochemical DOC reactions. Instead, previous work suggested that actinorhodopsins are broadly distributed in freshwater Actinobacteria, providing the potential for phototrophy (Garcia et al. 2013; Sharma et al. 2008, 2009; Martinez-Garcia et al. 2012; Salka et al. 2014). We confirmed the presence of actinorhodopsin in eight of the eleven acI SAGs (barring AAA023-J06, AAA028-A23, and AB141-P03) making this gene a likely part of the broadly-shared genes of the acI lineage. Moreover, enzymes required for synthesis of the presumed actinorhodopsin chromophore, retinal were identified in ten of the 11 acI SAGs, suggesting the likely assembly of functional rhodopsin in vivo. Most commonly the four enzymes that lead from the ubiquitous sterol precursor farnesyl pyrophosphate to ß–carotene (Martinez et al. 2007) are encoded in a cluster. The final enzyme in the pathway, which catalyzes cleavage of ß–carotene to the retinal chromophore, was somewhat surprisingly encoded only in two of the 11 SAGs, and not co-located with the gene for

rhodopsin itself, as it is in many marine proteobacteria (Riedel et al. 2013; Martinez et al. 2007; Vollmers et al. 2013). This puzzle remains to be explored. Rhodopsin in *Pelagibacter ubique* and other marine bacteria promotes survival during nutrient starvation periods (DeLong & Beja 2010; Steindler et al. 2011; Gomez-Consarnau et al. 2007, 2010), thus actinorhodopsin may serve a similar function in acI. Interestingly, actinorhodopsin expression in a German lake was not linked directly to sunlight but rather to a circadian schedule with a maximum expression rate just before dawn (Wurzbacher et al. 2012). Comparative genomic analysis also revealed the presence of genes involved in anaplerotic carbon fixation (carbonic anhydrase and phosphoenolpyruvate (PEP) carboxylase). The acI SAGs lack the RuBisCO enzyme and other pathways for carbon fixation, but PEP carboxylase may provide them with the ability to synthesize oxaloacetate, an intermediate in the TCA cycle that can be used to replenish precursors needed for growth (Figure 2). This may provide acI with the ability to grow photoheterotrophically using actinorhodopsin, though we can only speculate in the absence of experimental evidence.

3.5.2. NITROGEN, PHOSPHORUS AND SULFUR

As reported previously for the acI-B1 SAG AAA027-L06, the lineage appears to lack identifiable genes involved in sulfate, sulfite, nitrate, and nitrite assimilation. Seven SAGs did harbor ammonia permease (COG0004), suggesting free ammonia is an N-source for some or all acI members. We found no evidence for urea transport or catabolism. The SAGs do not harbor a complete urea cycle among them. However, genome content across the lineage shows the capability to consume N-rich carbon sources, including polyamines, cyanophycin, di- and oligopeptides, and branched chain amino acids such as leucine, isoleucine, and valine. This has intriguing implications for acI's success in freshwater, since it seems to obtain both carbon and nitrogen from the same substrate compounds. This observation would build on previous observations based on MAR-FISH using probes targeting the

entire Actinobacteria phylum suggesting that acI contributes more than would have been expected based on its biomass to total amino acid turnover in lakes (Salcher et al. 2010).

Genes involved in acquisition of these N-rich compounds seem to be unique to acI compared with other freshwater bacteria (Table 2) as well as the Actinomycetales (Table 3). These genes also generally have high codon bias (Table S6), indicating possibly high expression rates. Additionally, an analysis of the N content of the acI protein sequences (Supplementary Material) indicates that acI has not evolved to minimize N usage in protein synthesis, providing evidence that N is not typically a limiting nutrient for acI. We also found no evidence of genes for nitrite/nitrate uptake and assimilation. This may be related to acI's specialized ability to take up N-rich organic compounds (such as putrescine (32% N), cyanophycin (23% N), oligo and di-peptides, and amino acids) as energy and nutrient sources. However, very little prior research has focused on organic-N concentrations or sources for plankton.

We found no evidence of genes for sulfite/sulfate assimilation or reduction in the acI SAGs, as was previously reported for AAA027-L06, suggesting the proposed dependence on cysteine synthase for sulfur incorporation holds for the entire lineage (Garcia et al. 2013). Cysteine synthase could be found in eight of the eleven SAGs.

We searched for evidence of phosphorus acquisition strategies that may allow for high-affinity transport (Pst) or low-affinity high rate of transport (Pit). Six of the acI genomes have a PstSCAB transport system. Three other genomes had 75% of these genes; the partial gene absence was likely due to incomplete genome recovery. Only the two most incomplete SAGs were missing the majority of the Pst complex. The associated regulatory protein PhoU and phosphorus starvation activated protein PhoH were also found in most of the genomes. The presence of the Pst complex (in lieu of the absence of low-affinity Pit phosphorus transport complex) provides acI with the ability to acquire phosphorus in low nutrient settings, further indicating that these organisms are well adapted to phosphorus limitation,

which is a common feature of these and most freshwater lakes (Magnuson et al. 2006). We could not find any evidence for the ability to metabolize phosphonates in the acI SAGs.

3.5.3. SHARED AND DIFFERENTIAL GENE CONTENT INVOLVED IN HETEROTROPHIC GROWTH AMONG FRESHWATER BACTERIA

We asked whether the differential gene content among sequenced freshwater bacterial genomes might also explain the ecological success of acI during growth. First we examined DOC uptake COGs for acI versus other Actinomycetales (424 genomes) and freshwater bacteria such as LD12-alpha-proteobacteria relative of the marine SAR11 (10 genomes), and *Polynucleobacter sp.* (1 genome) (Table 2). We found that several COGs involved in polyamine uptake and metabolism were more common in acI than in the other analyzed genomes. In fact the ten LD12 genomes did not contain any of these COGs. Other interesting COGs that were more common in acI were nucleotide transporter and several sugar transporters and two ribose/xylose transporters. In general, acI had more COGs for general carbohydrate transport than LD12 and *Polynucleobacter*.

As described above, the acI genomes lack carboxylic acid and dicarboxylic acid transporters while they were present in *Polynucleobacter* and LD12. Compared to the other freshwater bacteria, general amino acid uptake COGs were under-represented in acI SAGs and COGs for di- and oligo-peptides were over-represented. The di- and oligo-peptide transporters may have some connection with the cyanophycinase genes found in acI (and not in LD12 or *Polynucleobacter*). Homologs of cyanophycinase are typically found in cyanobacteria where they are used to break down an intracellular granular C and N storage molecule, cyanophycin (Richter et al. 1999). Cyanophycin synthetase and/or cyanophycinase homologs have been detected in about 10% of heterotrophic genomes analyzed (Fueser & Steinbuechel 2007; Krehenbrink et al. 2002). The acI SAGs analyzed here have the genes required to break down cyanophycin granules and take up the resulting di-peptides and amino acids as a source of energy and N. However, it is unclear whether this cyanophycinase can be secreted, since we were

unable to identify an obvious secretion signal peptide. Secreted cyanophycinases have been reported in both Gram negative and Gram positive soil bacteria (Obst et al. 2004, 2002; Sallam et al. 2011), but in these cases signal sequence cleavage was demonstrated or inferred. In the case of the acI cyanophycinases, we could not identify an obvious secretion mechanism. Although the SAG genomes encode canonical Sec pathway components as well as twin arginine translocation genes, the predicted cyanophycinase polypeptides do not have these signal translocation sequences. Similarly, sortases that would anchor secreted proteins to the cell wall were predicted in the acI genomes, but the sortase motif was not found in the cyanophycinase polypeptides. We did not identify homologs of genes involved cyanophycin synthesis. The potential ability of acI to break down cyanophycin hints at a potential interaction between cyanobacteria (or other microbes making cyanophycin), whereby acI members acquire energy, C and N from this polymer synthesized by others. Curiously, we found two tandem copies of cyanophycin synthetase in the *Polynucleobacter* genome (locus tags Pnuc_670 and Pnuc_671) but not in acI. This has implications for potential interactions between acI and *Polynucleobacter* when they co-occur (i.e. *Polynucleobacter* synthesizing cyanophycin and acI breaking it down). Co-occurrence of acI and *Polynucleobacter sp.* has already been observed in laboratory enrichments (Garcia et al. in press).

COGs involved in lipid transport were under-represented in acI, with only the glycerol transporter, COG1133 (long chain fatty acid transport) and COG2867 (Oligoketide cyclase/lipid transport protein) present. Taken altogether, these findings provide genome-level confirmation of experimental observations (Salcher et al. 2013): the differences in types of DOC uptake genes found in each lineage indicate potential specialization in substrate preferences and ecological niches.

3.5.4. COMPARATIVE GENOMICS IDENTIFIES THE BASIS FOR ECOLOGICAL DIFFERENTIATION WITHIN THE ACI LINEAGE

We sought to determine whether differences in gene content might show potential niche

separation between members of the acI lineage (e.g. separation of acI-A vs acI-B) (Figure 3, Table S7). The alpha and beta-galactosidases were common to acI-A1 and acI-A7, but were absent in all acI-B1 SAGs (Table S7). Interestingly, an over-representation of glycosidases such as alpha and beta-galactosidases was found in freshwater systems compared to marine systems (Eiler et al. 2013). This effect may be largely due to the presence of the ubiquitous acI-A clade in freshwater. The acI-A clade also had two glycerol uptake COGs that were not found in acI-B1. The acI-A7 tribe had three COGs involved in amino acid uptake as well as an N-acetylglucosamine kinase and oligopeptide transporter not found in acI-A1 or acI-B1. AcI-A have functional ABC transport systems for xylose, while acI-B has functional ABC transport systems for fructose. Both share an ABC transport system for ribose. The acI-B1 genomes carried a ribose transporter (COG1869) that was not found in acI-A1 or acI-A7 genomes (Table S8). The acI-B1 SAGs also carried a ribonuclease, homospermidine synthase, and aromatic ring-cleaving dioxygenase that were not found in acI-A. Overall, the acI-A tribes seem to have more metabolic versatility and a wider array of COGs involved in the uptake of potential C substrates than does acI-B1. This is also reflected in the smaller estimated genome size of acI-B1 (Table 1).

We note that based on the genome recoveries (Table 1), the probability of an acI "shared COG set" gene missing from all the genomes due to incomplete recovery is 0.00003% in acI, 0.01% in acI-B1, 5.2% in acI-A1 and 5.7% in acI-A7, for any single COG. These probabilities indicate that the "shared COG set" based on the SAGs for acI and acI-B1 are essentially complete, but that the "shared COG set" for acI-A and corresponding tribes is likely to expand as much as 5% with the sequencing of more genomes.

The average protein similarity between the acI SAGs indicated surprising diversity within clades and a range in variation within tribes, at the protein-coding gene level (Table S9). Notably, the protein similarity across clades (i.e. acI-A vs. acI-B) was 68-74%, which was similar to the protein

similarity across tribes in the acI-A clade (i.e. acI-A1 vs acI-A7), which was about 70-78%. AcI-A1 was 71-78% similar to acI-A7 and was 71-74% similar to acI-B1, while SAGs within acI-B1 shared 82-97% similarity.

3.5.5. SHARED GENE CONTENT WITHIN acI DISTINGUISHES THE LINEAGE FROM OTHER ACTINOMYCETALES

Any genes that distinguish acI clades from their relatives within the Actinomycetales are good candidates for further inquiry into their role in acI ecophysiology, as they may explain why acI is such an omnipresent and successful resident of freshwater systems (Newton et al. 2011). However, we identified only two COGs that were found in acI to the exclusion of all other sequenced Actinomycetales: an aromatic ring-cleaving dioxygenase (COG3805) and an uncharacterized conserved protein (COG2859). The aromatic ring-cleaving dioxygenase (COG3805) was found only in acI-B1 and may provide the ability to consume difficult aromatic compounds for energy and C. Its best match by blastp was to Pfam PF08883 (Dopa 4,5-dioxygenase family protein) from the cyanobacterium *Scytonema hofmanni* (59% similarity). Homologs are usually involved in the synthesis of chromophores (betalamic acid), but we were unable to identify other genes in that pathway in acI-B1, therefore its function in acI-B1 is unknown. Overall, the phylogenetic distribution of best blast hits identified in IMG is mostly restricted to genes within other Actinobacteria. Only four of the 158 Pfams used for genome completion estimates were absent in all 11 SAGs (Table S2).

We identified the top 25 COGs that are over-represented in the acI SAGs as compared to the 424 Actinomycetales reference genomes available in the IMG database in March 2013 (Table 3). Bacteriorhodopsin (COG5524) was the most over-represented COG in this analysis, and represents the putative light-harvesting protein previously identified as actinorhodopsin (Sharma et al. 2009; Martinez-Garcia et al. 2012; Sharma et al. 2008; Wurzbacher et al. 2012). The aforementioned cyanophycinase (COG4242) was another unique COG over-represented in the acI lineage. Other

notable over-represented COGs were those involved in the uptake and metabolism of polyamines (spermidine/putrescine) and amino acids as well as glycosyl hydrolases (carbohydrate breakdown), nicotinamide mononucleotide transporter (source of pyridine), and inorganic pyrophosphatase (breaks down pyrophosphate into two molecules of phosphate). Thus, acI genomes look like highly streamlined versions of "typical" Actinobacteria that include some genes (with best matches to other phyla) that allow it to specialize by taking advantage of the light and N-rich organic compounds.

3.6. CONCLUSIONS

Members of the acI lineage are clearly specialized relative to their parent order (Actinomycetales) and other sequenced freshwater bacteria. Their highly streamlined genomes and small cell size suggest they share broad niche dimensions with ultramicrobacteria such as freshwater members of the SAR11 clade. While many characteristics of the SAGs analyzed here were consistent with the only other previously published acI genome (AAA027-L06) (Garcia et al. 2013), investigating three different tribes showed ecological differentiation among them. We identified features that hint at a preference for N-rich compounds as well as the potential for some level of photoheterotrophic metabolism. Our findings form a rich foundation for further study of the acI lineage using techniques such as metatranscriptomics combined with experimentation.

3.7. ACKNOWLEDGEMENTS

3.8. REFERENCES

Allgaier M, Grossart HP. (2006). Diversity and seasonal dynamics of Actinobacteria populations in four lakes in northeastern Germany. Appl. Environ. Microbiol. 72:3489–3497.

Beier S, Bertilsson S. (2011). Uncoupling of chitinase activity and uptake of hydrolysis products in freshwater bacterioplankton. Limnol. Oceanogr. 56:1179–1188.

Buck U, Grossart HP, Amann R, Pernthaler J. (2009). Substrate incorporation patterns of bacterioplankton populations in stratified and mixed waters of a humic lake. Environ. Microbiol. 11:1854–1865.

Chou HT, Kwon D-H, Hegazy M, Lu C-D. (2008). Transcriptome analysis of agmatine and putrescine catabolism in Pseudomonas aeruginosa PAO1. J. Bacteriol. 190:1966–75.

Dasu VV, Nakada Y, Ohnishi-Kameyama M, Kimura K, Itoh Y. (2006). Characterization and a role of Pseudomonas aeruginosa spermidine dehydrogenase in polyamine catabolism. Microbiology 152:2265–72.

DeLong EF, Beja O. (2010). The Light-Driven Proton Pump Proteorhodopsin Enhances Bacterial Survival during Tough Times. PLoS Biol. 8.

Eckert EM, Salcher MM, Posch T, Eugster B, Pernthaler J. (2012). Rapid successions affect microbial N-acetyl-glucosamine uptake patterns during a lacustrine spring phytoplankton bloom. Environ. Microbiol. 14:794–806.

Eddy SR. (2011). Accelerated Profile HMM Searches. PLoS Comput. Biol. 7:e1002195.

Eiler A, Zaremba-Niedzwiedzka K, Andersson SGE, Martinez-Garcia M, McMahon KD, Stepanauskas R, et al. (2013). Productivity and salinity structuring of the microplankton revealed by comparative freshwater metagenomics. Environ. Microbiol. In press.

Fueser G, Steinbuechel A. (2007). Analysis of genome sequences for genes of cyanophycin metabolism: Identifying putative cyanophycin metabolizing prokaryotes. Macromol. Biosci. 7:278–296.

Garcia SL, McMahon KD, Grossart H-P, Warnecke F. Successful enrichment of the ubiquitous freshwater acI Actinobacteria. Environ. Microbiol. Rep.

Garcia SL, McMahon KD, Martinez-Garcia M, Srivastava A, Sczyrba A, Stepanauskas R, et al. (2013). The metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. ISME J. 7:137–147.

Ghai R, McMahon KD, Henrique-Silva F. (2011). Metagenomics of the water column in the pristine upper course of the Amazon River. PLoS One 6:e23785.

Ghai R, McMahon KD, Rodriguez-Valera F. (2012). Breaking a paradigm: cosmopolitan and abundant freshwater Actinobacteria are low GC. Environ. Microbiol. Rep. 4:29–35.

Glockner FO, Zaichikov E, Belkova N, Denissova L, Pernthaler J, Pernthaler A, et al. (2000). Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of actinobacteria. Appl. Environ. Microbiol. 66:5053–5065.

Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc. Natl. Acad. Sci. U. S. A. 108:1513–1518.

Gomez-Consarnau L, Akram N, Lindell K, Pedersen A, Neutze R, Milton DL, et al. (2010). Proteorhodopsin Phototrophy Promotes Survival of Marine Bacteria during Starvation. PLoS Biol. 8.

Gomez-Consarnau L, Gonzalez JM, Coll-Llado M, Gourdon P, Pascher T, Neutze R, et al. (2007). Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. Nature 445:210–213.

Hacker J, Kaper JB. (2000). Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. 54:641–679.

Humbert JF, Dorigo U, Cecchi P, Le Berre B, Debroas D, Bouvy M. (2009). Comparison of the structure and composition of bacterial communities from temperate and tropical freshwater ecosystems. Environ. Microbiol. 11:2339–2350.

Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11.

Krehenbrink M, Oppermann-Sanio FB, Steinbuchel A. (2002). Evaluation of non-cyanobacterial genome sequences for occurrence of genes encoding proteins homologous to cyanophycin synthetase and cloning of an active cyanophycin synthetase from Acinetobacter sp strain DSM 587. Arch. Microbiol. 177:371–380.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, et al. (2004). ARB: a software environment for sequence data. Nucleic Acids Res. 32:1363–1371.

Magnuson JJ, Kratz TK, Benson BJ. (2006). Long Term Dynamics of Lakes in the Landscape.

Makarova KS, Aravind L, Galperin MY, Grishin N V, Tatusov RL, Wolf YI, et al. (1999). Comparative genomics of the archaea (Euryarchaeota): Evolution of conserved protein families, the stable core, and the variable shell. Genome Res. 9:608–628.

Markowitz VM, Chen I-M a, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. (2012). IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res. 40:D115–22.

Martinez A, Bradley AS, Waldbauer JR, Summons RE, Delong EF. (2007). Proteorhodopsin photosystem gene expression enables photophosphorylation in a heterologous host. Proc. Natl. Acad. Sci. 104:5590–5595.

Martinez-Garcia M, Swan BK, Poulton NJ, Gomez ML, Masland D, Sieracki ME, et al. (2012). High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. ISME J 6:113–123.

Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. (2011). A guide to the natural history of freshwater lake bacteria. Microbiol. Mol. Biol. Rev. 75:14–49.

Newton RJ, Jones SE, Helmus MR, McMahon KD. (2007). Phylogenetic ecology of the freshwater Actinobacteria acI lineage. Appl. Environ. Microbiol. 73:7169–76.

Newton RJ, Kent AD, Triplett EW, McMahon KD. (2006). Microbial community dynamics in a humic lake: differential persistence of common freshwater phylotypes. Environ. Microbiol. 8:956–970.

Obst M, Oppermann-Sanio FB, Luftmann H, Steinbuchel A. (2002). Isolation of cyanophycin-degrading bacteria, cloning and characterization of an extracellular cyanophycinase gene (cphE) from Pseudomonas anguilliseptica strain BI - The cphE gene from P-anguilliseptica BI encodes a cyanophycin-hydrolyzing enzyme. J. Biol. Chem. 277:25096–25105.

Obst M, Sallam A, Luftmann H, Steinbuchel A. (2004). Isolation and characterization of gram-positive cyanophycin-degrading bacteria - Kinetic studies on cyanophycin depolymerase activity in aerobic bacteria. Biomacromolecules 5:153–161.

Perez MT, Hortnagl P, Sommaruga R. (2010). Contrasting ability to take up leucine and thymidine among freshwater bacterial groups: implications for bacterial production measurements. Environ. Microbiol. 12:74–82.

Poretsky RS, Sun S, Mou XZ, Moran MA. (2010). Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. Environ. Microbiol. 12:616.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35:7188–7196.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. (2012). The Pfam protein families database. Nucleic Acids Res. 40:D290–301.

Richter R, Hejazi M, Kraft R, Ziegler K, Lockau W. (1999). Cyanophycinase, a peptidase degrading the cyanobacterial reserve material multi-L-arginyl-poly-L-aspartic acid (cyanophycin) - Molecular cloning of the gene of Synechocystis sp PCC 6803, expression in Escherichia coli, and biochemical characterization of. Eur. J. Biochem. 263:163–169.

Riedel T, Gomez-Consarnau L, Tomasch J, Martin M, Jarek M, Gonzlez JM, et al. (2013). Genomics and Physiology of a Marine Flavobacterium Encoding a Proteorhodopsin and a Xanthorhodopsin- Like Protein. PLoS One 8:e57487.

Rocha EP. (2004). Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Res. 14:2279–2286.

Rösel S, Grossart H. (2012). Contrasting dynamics in activity and community composition of free-living and particle-associated bacteria in spring. Aquat. Microb. Ecol. 66:169–181.

Saier MH, Yen MR, Noto K, Tamang DG, Elkan C. (2009). The Transporter Classification Database: recent advances. Nucleic Acids Res. 37 :D274–D278.

Salcher MM, Pernthaler J, Posch T. (2010). Spatiotemporal distribution and activity patterns of bacteria from three phylogenetic groups in an oligomesotrophic lake. Limnol. Oceanogr. 55:846–856.

Salcher MM, Posch T, Pernthaler J. (2013). In situ substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. ISME J.

Salka I, Wurzbacher C, Garcia SL, Labrenz M, Jürgens K, Grossart H-P. (2014). Distribution of acI-Actinorhodopsin genes in Baltic Sea salinity gradients indicates adaptation of facultative freshwater photoheterotrophs to brackish waters. Environ. Microbiol. In press.

Sallam A, Kalkandzhiev D, Steinbüchel A. (2011). Production optimization of cyanophycinase ChpE al from Pseudomonas alcaligenes DIP1. AMB Express 1:38.

Sharma AK, Sommerfeld K, Bullerjahn GS, Matteson AR, Wilhelm SW, Jezbera J, et al. (2009). Actinorhodopsin genes discovered in diverse freshwater habitats and among cultivated freshwater Actinobacteria. ISME J. 3:726–737.

Sharma AK, Zhaxybayeva O, Papke RT, Doolittle WF. (2008). Actinorhodopsins: proteorhodopsin-like gene sequences found predominantly in non-marine environments. Environ. Microbiol. 10:1039–1056.

Sharp PM, Li WH. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15:1281–1295.

Steindler L, Schwalbach MS, Smith DP, Chan F, Giovannoni SJ. (2011). Energy Starved Candidatus Pelagibacter Ubique Substitutes Light-Mediated ATP Production for Endogenous Carbon Respiration. PLoS One 6.

Tada Y, Grossart H-P. (2013). Community shifts of actively growing lake bacteria after N-acetyl-glucosamine addition: improving the BrdU-FACS method. ISME J. In press.

Vollmers J, Voget S, Dietrich S, Gollnow K, Smits M, Meyer K, et al. (2013). Poles Apart : Arctic and Antarctic Octadecabacter strains Share High Genome Plasticity and a New Type of Xanthorhodopsin. PLoS One 8:e63422.

Warnecke F, Sommaruga R, Sekar R, Hofer JS, Pernthaler J. (2005). Abundances, identity, and growth state of Acintobacteria in mountain lakes of different UV transparency. Appl. Environ. Microbiol. 71:5607–5609.

Woyke T, Xie G, Copeland A, Gonzalez JM, Han C, Kiss H, et al. (2009). Assembling the marine metagenome, one cell at a time. PLoS One 4:e5299.

Wu X, Xi W, Ye W, Yang H. (2007). Bacterial community composition of a shallow hypertrophic freshwater lake in China, revealed by 16S rRNA gene sequences. FEMS Microbiol. Ecol. 61:85–96.

Wurzbacher C, Salka I, Grossart H-P. (2012). Environmental actinorhodopsin expression revealed by a new in situ filtration and fixation sampler. Environ. Microbiol. Rep. 4:491–497.

Zerbino DR, Birney E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829.

3.9. FIGURES

Figure 1. Phylogenetic placement of the SAGs within the acI lineage and relative to other sequenced actinobacterial genomes, and to the previously sequenced AAA027-L06 SAG based on nearly full-length 16S rRNA gene sequences. When only short (~400 bp) amplified gene fragments were available (Martinez-Garcia et al. 2012), sequences were added after tree construction within the ARB Software using the maximum parsimony criterion (Ludwig et al. 2004). These shorter sequences are noted along with their corresponding accession number. Shaded sequences are from SAGs.

Figure 1

0.1 changes per site
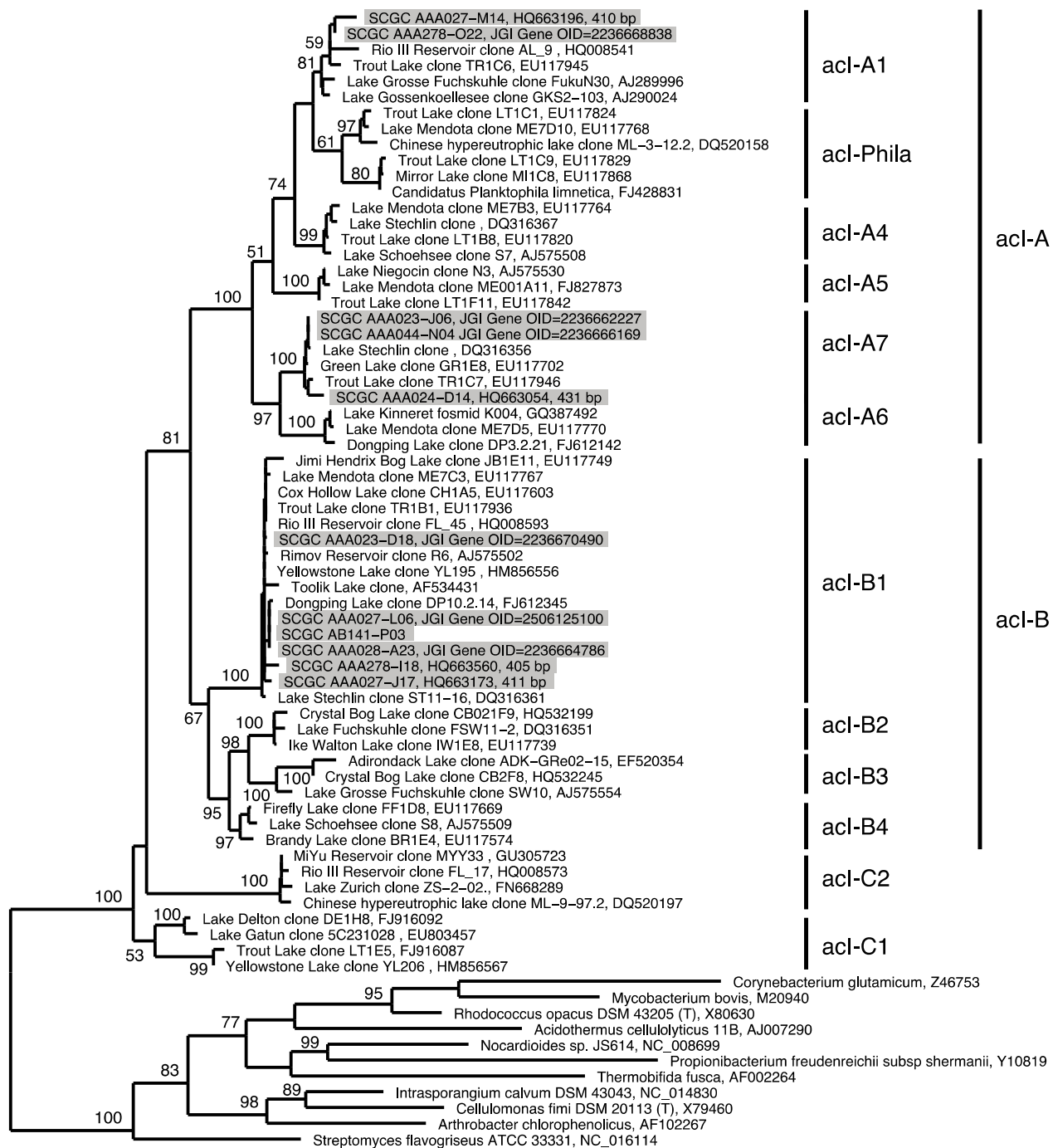
Figure 2. Central carbon metabolism and other relevant metabolic pathways identified in acI SAGs. Circles denote genes encoding the necessary enzymes are present in that clade (blue for acI-A and green for acI-B), with the size of the circle being proportional to the percentage of SAGs within that clade that were found to contain that gene. Presence and absence of genes was determined within the IMG environment based on KEGG annotations.

GLYCOLYSIS

PENTOSE PHOSPHATE PATHWAY

Glucose → Glucono-1,5-lactone-6P → 6-Phospho-gluconate

Xylose

Glucose-6P

Xylulose

Fructose-6P → Arabino-Hexulose-6P → Ribulose-5P ← Xylulose-5P

Fructose-6P → Xylulose-5P

Ribose-5P → **Ribose**

Fructose-1,6P2

PRPP

Glyceraldehide-3P

Phenylalanine Tryptophan

1,3-Biphosphoglycerate

Isoleucine Valine Leucine

3-Biphosphoglycerate

Spermidine

2-Biphosphoglycerate

Extracellular Polyamine

CO2 (atmosphere)

Phosphoenolpyruvate

Homo-spermidine ← **Putrescine**

Pyruvate

1,3-Diaminopropane

4-aminobutyraldehyde

S-Acetyldihydro-lipoamide-E

Glutamine

Acetate

3-Amino propanaldehyde

4-aminobutyrate

NH3

Acetyl-CoA

β-alanine

Succinate semialdehyde

Glutamate

Proline

Acetaldehyde

Aspartate

Ornithine

Ethanol

Citrate

L-arginino succinase UREA CYCLE

Oxaloacetate

cis-Aconitate

Fumarate

Malate

Isocitrate

Arginine

Fumarate

Oxalosuccinate

TCA CYCLE

Citruline

Succinate

2-Oxoglutarate

n(Asp-4-Arg)dipeptide

Succinil-CoA

3-Carboxy-1-hydroxypropyl-ThPP

Cyanophycin

S-Succinyl-dihydrolipoamide-E

Dihydro-lipoamide-E → Lipoamide E

75-100% 50-75% 25-50% 0-25%   Percentage

● ● ● · acl-A (1 & 7)

● ● ● · acl-B (1)

Figure 2

Figure 3. Shared and differential COG content among acI tribes within the clades based on comparison of two acI-A1 SAGs, three acI-A7 SAGs, and six acI-B1 SAGs (1,193 COGs total).

Figure 3

Markedly more COGs are shared among acI tribes than are unique for any individual acI tribe
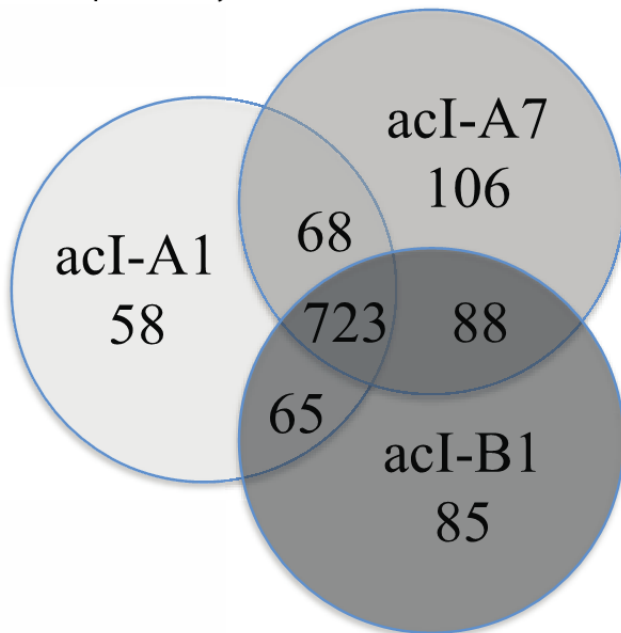
Note: The probability of an acI "shared COG set" gene missing from all the genomes due to incomplete recovery is 0.000003% in acI, 0.003% in acI-B1, 2.7% in acI-A1 and 4.1% in acI-A7, for any single COG. Venn circles include all COGs found in acI genomes, similar to pangenomes.

3.10. TABLES

Table 2. AcI single-cell genome metadata and overall features

Full SAG ID	IMG Taxon OID	Tribe	Estimated Genome Completeness[1]	Lake[2]	Date Collected	Mbp	Estimated Complete Genome Size (Mbp)	%GC	Scaffolds	Genes
SCGC AAA027-M14	2236661003	acI-A1	41%	MEN	5/19/09	0.8	2.0	47%	22	863
SCGC AAA278-O22	2236661007	acI-A1	96%	DAM	8/19/09	1.1	1.1	48%	43	1238
SCGC AAA044-N04	2236661005	acI-A7	78%	DAM	4/28/09	1.3	1.7	46%	23	1341
SCGC AAA023-J06	2236661001	acI-A7	36%	SPA	5/28/09	0.7	1.9	45%	98	818
SCGC AAA024-D14	2264265190	acI-A7	71%	SPA	5/28/09	0.8	1.1	45%	82	892
SCGC AAA027-J17	2236661002	acI-B1	84%	MEN	5/19/09	1.0	1.2	42%	81	1094
SCGC AAA028-A23	2236661004	acI-B1	80%	MEN	5/19/09	0.8	1.0	42%	64	913
SCGC AAA278-I18	2236661006	acI-B1	74%	DAM	8/19/09	0.9	1.2	41%	54	1037
SCGC AAA023-D18	2236661009	acI-B1	51%	SPA	5/28/09	0.8	1.6	40%	67	827
SCGC AAA027-L06	2505679121	acI-B1	>99%	MEN	5/19/09	1.2	1.2	42%	75	1226
SCGC AB141-P03	2236876028	acI-B1	34%	STE	5/25/10	0.7	2.0	41%	66	736

[1]Estimated based on presence of single copy genes found in other Actinomycetales, as described in Methods

[2]MEN – Lake Mendota (Eutrophic); DAM – Damariscotta Lake (Mesotrophic); SPA – Sparkling Lake (Oligotrophic); STE – Lake Stechlin (Oligotrophic)

Table 3. Select DOC Uptake COGs in acI, LD12, PnecC and Actinomycetales[1]. Cells are shaded according to percent of genomes that contain that COG, to aid in visual interpretation of numbers. A full list of DOC uptake COGs is presented in Table S5.

Category	Clusters of Orthologous Groups (COGs)	Percent of genomes containing that COG[2]			
		acI (11)	LD12 (10)	Pnec C (1)	Actm (424)
Amino Acids, general	COG0531 Amino acid transporters	100%	0%	100%	0%
	COG0834 ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	27%	50%	100%	97%
	COG1126 ABC-type polar amino acid transport system, ATPase component	27%	60%	100%	78%
	COG0765 ABC-type amino acid transport system, permease component	27%	50%	100%	78%
Amino Acids, branched chain	COG0683 ABC-type branched-chain amino acid transport systems, periplasmic component	91%	60%	100%	51%
	COG0410 ABC-type branched-chain amino acid transport systems, ATPase component	91%	60%	100%	48%
	COG0559 Branched-chain amino acid ABC-type transport system, permease components	91%	60%	100%	47%
	COG4177 ABC-type branched-chain amino acid transport system, permease component	91%	60%	100%	46%
	COG0411 ABC-type branched-chain amino acid transport systems, ATPase component	91%	60%	100%	45%
	COG1296 Predicted branched-chain amino acid permease (azaleucine resistance)	45%	80%	100%	62%
di- & oligo-peptides	COG1173 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	73%	0%	0%	100%
	COG0601 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	64%	0%	0%	100%
	COG0747 ABC-type dipeptide transport system, periplasmic component	64%	10%	0%	99%
	COG0444 ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	55%	0%	0%	89%
	COG4166 ABC-type oligopeptide transport system, periplasmic component	27%	0%	100%	96%
	COG4166 ABC-type oligopeptide transport system, periplasmic component	27%	0%	100%	96%
	COG4608 ABC-type oligopeptide transport system, ATPase component	27%	0%	0%	100%
Polyamines	COG1176 ABC-type spermidine/putrescine transport system, permease component I	91%	0%	0%	29%
	COG3842 ABC-type spermidine/putrescine transport systems, ATPase components	82%	0%	100%	63%
	COG1177 ABC-type spermidine/putrescine transport system, permease component II	82%	0%	0%	35%
	COG0687 Spermidine/putrescine-binding periplasmic protein	82%	0%	0%	28%
Lipids	COG0580 Glycerol uptake facilitator and related permeases (Major Intrinsic Protein Family)	64%	50%	0%	63%
Nucleotides &	COG3201 Nicotinamide mononucleotide transporter	100%	0%	0%	71%

Category	COG				
Coenzymes	COG2233 Xanthine/uracil permeases	27%	0%	0%	67%
Carbs, general	COG1129 ABC-type sugar transport system, ATPase component	100%	0%	0%	65%
	COG1682 ABC-type polysaccharide/polyol phosphate export systems, permease component	91%	10%	100%	97%
	COG1879 ABC-type sugar transport system, periplasmic component	91%	0%	0%	63%
	COG1134 ABC-type polysaccharide/polyol phosphate transport system, ATPase component	73%	20%	100%	98%
	COG1653 ABC-type sugar transport system, periplasmic component	73%	0%	0%	95%
	COG1175 ABC-type sugar transport systems, permease components	73%	0%	0%	95%
	COG0395 ABC-type sugar transport system, permease component	73%	0%	0%	95%
	COG0738 Fucose permease	64%	0%	0%	47%
	COG3839 ABC-type sugar transport systems, ATPase components	36%	0%	0%	97%
	COG2211 Na+/melibiose symporter and related transporters	36%	0%	0%	57%
	COG2271 Sugar phosphate permease	27%	50%	100%	100%
	COG3822 ABC-type sugar transport system, auxiliary component	9%	0%	0%	0%
Carbs, pentoses	COG1172 Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components	91%	0%	0%	63%
	COG4214 ABC-type xylose transport system, permease component	55%	0%	0%	36%
	COG2182 Maltose-binding periplasmic proteins/domains	18%	0%	0%	42%
	COG3833 ABC-type maltose transport systems, permease component	9%	0%	0%	41%

[1]DOC COG list based on (Poretsky et al. 2010)

[2]Actm = Actinomycetales; PnecC = *Polynucleobacter necessarius asymbioticus.* Number of genomes included in the analysis is provided in parentheses.

Table 4. COGs over-represented in acI as compared to Actinomycetales (Actm). Cells are shaded according to magnitude of over-representation factor, to aid in interpretation of numbers.

	% of genomes containing COG		Over-representation factor acI/Actm
	Actm	acI	
COG5524 Bacteriorhodopsin	0.5%	73%	154
COG3858 Predicted glycosyl hydrolase	5%	73%	14
COG4242 Cyanophycinase and related exopeptidases	7%	73%	11
COG2956 Predicted N-acetylglucosaminyl transferase	5%	45%	8.8
COG0826 Collagenase and related proteases	7%	55%	8.0
COG4401 Chorismate mutase	13%	55%	4.4
COG5496 Predicted thioesterase	13%	55%	4.1
COG1792 Cell shape-determining protein	17%	64%	3.7
COG1077 Actin-like ATPase involved in cell morphogenesis	19%	64%	3.3
COG1176 ABC-type spermidine/putrescine transport system, permease component I	28%	91%	3.3
COG3938 Proline racemase	17%	55%	3.2
COG0687 Spermidine/putrescine-binding periplasmic protein	27%	82%	3.0
COG1054 Predicted sulfurtransferase	25%	73%	2.9
COG1748 Saccharopine dehydrogenase and related proteins	24%	64%	2.6
COG1089 GDP-D-mannose dehydratase	30%	73%	2.4
COG3808 Inorganic pyrophosphatase	39%	91%	2.3
COG1177 ABC-type spermidine/putrescine transport system, permease component II	37%	82%	2.2
COG3201 Nicotinamide mononucleotide transporter	46%	100%	2.2
COG4177 ABC-type branched-chain amino acid transport system, permease component	46%	91%	2.0
COG0411 ABC-type branched-chain amino acid transport systems, ATPase component	46%	91%	2.0

3.11. SUPPLEMENTARY ONLINE MATERIAL

Comparative single-cell genomics reveals potential ecological niches for freshwater ultramicrobacteria of the acI Actinobacteria lineage

Trevor W Ghylin[1], Francisco Moya[1], Ben O Oyserman[1], Sarahi L Garcia[2,4], James Mutschler[2], Patrick Schwientek[3], Katrina T. Forest[2], Leong-Keat Chan[3], Manuel Martinez-Garcia[5,9], Alexander Sczyrba[3,7], Ramunas Stepanauskas[5], Hans-Peter Grossart[6,8], Tanja Woyke[3], Falk Warnecke[4], Rex Malmstrom[3], Stefan Bertilsson[10], Katherine D McMahon[1,2]

1 Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, USA
2 Department of Bacteriology, University of Wisconsin-Madison Madison, WI, USA
3 DOE Joint Genome Institute, Walnut Creek, CA, USA
4 Jena School for Microbial Communication (JSMC) and Microbial Ecology Group at Friedrich Schiller University Jena, Jena, Germany
5 Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, USA
6 Department of Limnology of Stratified Lakes, Leibniz-Institute for Freshwater Ecology and Inland Fisheries, Stechlin, Germany
7 Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany
8 Institute for Biochemistry and Biology, Postdam University, Postdam, Germany
9 Current address: University of Alicante, Alicante, Spain.
10 Department of Ecology and Genetics, Limnology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden

Supplementary Materials and Methods

Water samples were collected from 0.5-1m depth in the US lakes (Mendota, Sparkling, and Damariscotta) and cryopreserved with 6% glycine betaine (Sigma) at $-80^{o}C$ until used. Water was collected from the surface from Lake Stechlin (Germany) at the deepest point of the lake, prefiltered through 12 um and fixed in 4% betaine and stored at -

$80^{o}C$ until used. Before cell sorting, the samples were diluted 10 x with sterile-filtered Lake Mendota water and pre-screened through a 70 um Q2 mesh-size cell strainer (Becton Dickinson). Diluted subsamples (1–3 ml) were incubated for 10–120 min with SYTO-9 DNA stain (5 uM final concentration; Invitrogen) for cell detection. Cell sorting was performed with a MoFlo (Beckman Coulter) flow cytometer using a 488-nm argon laser for excitation, a 70-um nozzle orifice and a CyClone robotic arm for droplet deposition into microplates. The cytometer was triggered on side scatter. The 'single 1 drop' mode was used for maximal sort purity, which ensures the absence of non-target particles within the target cell drop and the drops immediately surrounding the cell. High nucleic acid content prokaryote cells were deposited into 384-well plates containing 0.6 ul 1 x TE buffer per well and stored at $-80^{o}C$ until further processing. Of the 384 wells, 315 were dedicated for single cells, 66 were used as negative controls (no droplet deposition) and 3 received 10 cells each (positive controls). The accuracy of 10 um fluorescent bead deposition into the 384-well plates was verified by microscopically examining the presence of beads in the plate wells. Of the 2–3 plates examined each sort day, <2% wells were found to not contain a bead and <0.5% wells were found to contain more than one bead. The latter is most likely caused by co-deposition of two beads attached to each other, which at certain orientation may have similar optical properties to a single bead.

The cells were lysed and their DNA was denatured using cold KOH and then amplified using multiple displacement amplification (MDA) (Dean et al. 2002; Raghunathan et al. 2005). The 10-ul MDA reactions contained 2 U-ul^{-1} Repliphi polymerase (Epicentre), 1 x reaction buffer (Epicentre), 0.4mM each dNTP (Epicentre), 2mM DTT (Epicentre), 50mM phosphorylated random hexamers (IDT) and 1 uM SYTO-9 (Invitrogen) (all final

concentration). The MDA reactions were run at 30°C for 12–16 h, and then inactivated by 15 min incubation at 65°C. The amplified genomic DNA was stored at -80°C until further processing. We refer to the MDA products originating from individual cells as SAGs.

The instruments and the reagents were decontaminated for DNA before sorting and MDA setup, as previously described (Woyke et al. 2011; Stepanauskas & Sieracki 2007). Cell sorting and MDA setup were performed in a HEPA (high-efficiency particulate air)-filtered environment. As a quality control, the kinetics of all MDA reactions was monitored by measuring the SYTO-9 fluorescence using FLUOstar Omega (BMG). The critical point (Cp) was determined for each MDA reaction as the time required to produce half of the maximal fluorescence. The Cp is inversely correlated to the amount of DNA template (Zhang et al. 2006). The Cp values were significantly lower in 1-cell wells compared with 0-cell wells ($p<0.05$; Wilcoxon two sample test). Previous studies demonstrate the reliability of our methodology with insignificant levels of DNA contamination (Martinez-Garcia et al. 2012; Stepanauskas & Sieracki 2007; Woyke et al. 2009; Fleming et al. 2011; Heywood et al. 2011; Swan et al. 2011; Woyke et al. 2011).

The MDA products were diluted 50-fold in sterile TE buffer. Then, 0.5 ul aliquots of the dilute MDA products served as templates in 5 ul real-time PCR. The small subunit rRNA and rhodopsin genes were targeted in these PCR using primers and thermal cycling conditions specified in Martinez-Garcia et al (Martinez-Garcia et al. 2012) and sequenced from both ends using Sanger technology at Beckman Coulter Genomics. To obtain sufficient quantity of genomic DNA for shotgun sequencing, the original MDA products were re-amplified using similar MDA conditions as above: eight replicate 125 ul reactions were performed and then pooled together, resulting in 100 ug of genomic dsDNA.

Supplementary Results

Tribe specific functionality

We next sought to determine how many COGs were shared and unique within the acI lineage when comparing among clades, tribes, and individual SAGs. Each of the acI tribes appeared to harbor about 60-100 unique COGs that were not present in the other acI tribes (**Figure 3**). The acI-A7 tribe had the largest number of unique COGs (106) even though we analyzed only three SAGs from that tribe compared with six SAGs from acI-B1. The three acI-B1 SAGs from Lake Mendota were all very similar in terms of COG representation (**Figure S1**), which is consistent with their phylogenetic relationships (**Figure 1**). The most complete SAG (AAA027-L06 from Lake Mendota) had nearly all COGs found in the other two acI-B1 SAGs (AAA027-J17and AAA028-A23).

acI vs. freshwater bacteria

We explored the differences in acI gene content as compared to the only other genome publicly available from broadly distributed freshwater bacteria (*Polynucleobacter necessarius asymbioticus* QLW-P1DMWA-1)(Jezberova et al. 2010; Hahn et al. 2009; Meincke et al. 2012) (**Table S10**) and ten unpublished LD12 SAGs (**Figure S2**). **Figure S2** demonstrates the niche partitioning in freshwater systems in which acI and LD12 have significantly different genetic capabilities and likely consume different materials and utilize different life strategies.

Table S10 shows several COGs involved in the uptake of polyamines (spermidine/putrescine), di- and oligo-peptides, sugars, glycerol uptake and metabolism found in acI but not in *Polynucleobacter*. Other over-represented acI COGs were two drug

exporters (0842 and 2409), a carbon fixation gene (2352), several COGs involved in nucleotide metabolism, a phosphate starvation protein (1702), rhodopsin (light harvesting protein), cyanophycinase (enzyme used to break down a nitrogen and energy storage molecule produced by cyanobacteria), and a photorepair protein (1119). The table shows many COGs involved in nucleotide uptake and metabolism as well as COGs involved in carbohydrate and amino acid/polyamine uptake and metabolism and cell wall biosynthesis. The table also includes COGs that may be involved with metabolism of cell wall components, including N-acetyl glucosamine (NAG), which is also a chitin degradation byproduct. These COGs include lysozyme (3757) and glycerophosphoryl diester phosphodiesterase (0584), which is involved in the breakdown of glycerophospholipids, the main component of biological membranes.

Table S11 shows that the one *Polynucleobacter* genome had more COGs in every category than the combined 11 SAGs from the acI lineage. This demonstrates that the acI are streamlined genomes as compared with the cultured *Polynucleobacter* genome. The acI have the highest relative proportion of COGs in amino acid, carbohydrate, and nucleotide transport and metabolism. The acI genomes had the lowest relative proportion of COGs in cell wall biogenesis, inorganic ion transport, lipid transport, posttranslational modification and secondary metabolite production.

Codon Bias

In order to generate hypotheses about which genes in the acI lineage might be highly expressed under particular environmental conditions, we calculated the Codon Adaptation Index (CAI) (Sharp & Li 1987) for the protein coding genes in each SAG (**Figure S3, Table**

S6, S12, S13). A biased codon usage is thought to be predictive of efficient translation that results in a higher rate of gene expression (Sharp et al. 2010) or more accurate translation of critical proteins (Stoletzki & Eyre-Walker 2007). A genome-scale measure is obtained by computing CAI_{ave}, the average of the CAI values of all genes in a genome (Rocha 2004), and has been found to range from 0.35 (most biased) to 0.82 (least biased) with an average of 0.59 and a threshold of 0.59 as a cutoff for designating a genome as biased overall (Botzman & Margalit 2011). In comparison, the CAI_{ave} of the acI SAGs is 0.48 for acI-A1 (47% GC), 0.54 for acI-A7 (45% GC) and 0.63 for acI-B1 (41% GC), implying that only the acI-A1 genome is markedly biased overall. However, the CAI_{ave} might be masked by the relatively low GC content and small genome (Botzman & Margalit 2011); individual genes may still have biased CAI. **Table S6** shows COGs found in the acI SAGs with high codon bias (CAI > 0.7). Interestingly, many of the COGs found in the comparative genomics analysis with the Actinomycetales and *Polynucleobacter* are also highly biased genes. The biased genes include xylose and sugar transport, phosphate transporter and inorganic pyrophosphatase, cold shock proteins, dipeptide transport, amino acid synthesis, and an oligopeptide transporter and an amino acid transporter specific to acI-A7. However, one interesting omission is cyanophycinase. This COG had a CAI of 0.47, 0.51 and 0.62 in acI-A1, A7 and B1 respectively. This may indicate that this COG is most highly expressed in acI-B1. It also seems to indicate that it is not one of the highly expressed genes in acI in general. Another explanation could be that this gene was acquired via horizontal gene transfer and hasn't had time to fully evolve a strong codon bias. Spermidine/putrescine, bacteriorhodopsin, and branched-chain amino acid transport are also biased COGs. The cold shock proteins may enable acI to survive the cold Wisconsin winters while its competitors die out, allowing it to

maintain an active community throughout the year. Peroxiredoxin is further down the list with a CAI of 0.65 but this COG may provide acI with protection against oxidative stress due to UV.

Table S12 shows COGs with CAI greater than 0.6 that are unique to a specific acI tribe. These genes may be highly transcribed and may represent key differences between the acI tribes and the ecological niches they inhabit.

The leading categories containing highly biased COGs (CAI > 0.6) were energy production, amino acid transport and carbohydrate transport (Table S13). There are few secondary metabolite COGs and only one COG involved in motility and that was only found in one SAG (acI-B1). Also, there were few COGs involved in the transport of lipids, nucleotides or inorganic ions.

Homospermidine synthase is also included in this list (shown in yellow) but is unique to acI-B1. Homospermidine is a common polyamine and is also the first step in the synthesis of pyrrolizidine alkaloid, which is an insecticide and potential anti-microbial known to be produced by plants (Joosten & Van Veen 2011).

Figure S3 shows the codon bias distribution of genes in the most biased acI tribe (acI-A1) and the most complete acI-A1 SAG (91%). This figure shows that most of the ribosomal genes had a CAI greater than 0.65, indicating a high level of codon bias, while many of the other genes had CAI of 0.4-0.55, indicating less bias. The figure also shows that 4% of the non-ribosomal genes had a CAI greater than 0.65, 8% were greater than 0.6.

Nitrogen Content of Proteome

Some organisms preferentially select amino acids to reduce their nitrogen requirements, especially in oceans where nitrogen is the limiting nutrient. (Grzymski &

Dussaq 2012) found that nitrogen costs, on average, decrease as GC content, cell size and genome size decrease. Based on the low GC%, small cell size and small genome of the acI lineage, we would expect to see a low N:C content in the encoded proteome.

Surprisingly, the acI SAGs have an average N:C content (~0.272) in their genomes when compared with 1,187 other bacterial genomes (0.25 − 0.3), indicating low selective pressure from nitrogen starvation. However, there are specific genes that appear to have a low N:C ratio (data not shown). These genes include phosphate, amino acid, dipeptide/oligopeptide, glycerol, sugar, xylose/ribose, and polyamine transporters as well as bacteriorhodopsin. Cyanophycinase is again missing from this list as it has an average N:C ratio.

If nitrogen starvation was a selective pressure on the acI genomes, we would expect to see a correlation between highly expressed genes and low N:C content. However, we found an absence of ribosomal proteins (highly expressed genes) with low N:C ratio (data not shown), indicating that nitrogen conservation may not be a strong selective pressure. Indeed, the average N:C ratio of the ribosomal proteins is about 0.274, about average for the proteome. Based on the average N:C content of the entire proteome and the lack of N conservation in the ribosomal proteins, it appears that there is little or no selection for nitrogen conservation in these genomes.

The lack of nitrogen conservation in the acI proteome may be an indication that nitrogen is not usually a limiting nutrient for these bacteria. Typically, phosphorus (not nitrogen) is considered to be the limiting nutrient in freshwater systems as nitrogen is generally abundant during spring and fall turnover in many lakes. However, nitrogen concentrations (especially ammonia) generally approach zero after spring when

phytoplankton grow and consume any available nitrogen. The acI have a plethora of genes involved in taking up nitrogenous compounds (amino acids, di- and oligo-peptides, branched chain amino acids, polyamines, cyanophycin), some of which are not found in competing freshwater bacteria (i.e. polyamine, di- and oligo-peptide uptake and cyanophycinase). This wide variety of nitrogen uptake functionality likely allows them to obtain nitrogen from unique sources (i.e. polyamines, di- and oligo-peptide, cyanophycin) that are not available to other bacteria as they lack the machinery.

Supplementary Figures and Tables

Figure S1. (A): Comparison of acI-B1 SAGs from Lake Mendota and Sparkling Lake. (B): Comparison of acI-B1 SAGs across four lakes.

A

JGI ID	Genome Completeness	Mbp
acI-B1 MEN J17	84%	1.0
acI-B1 MEN A23	80%	0.8
acI-B1 MEN L06	>99%	1.2
acI-B1 SPA D18	51%	0.8

B

ID	Genome Completeness	Mbp
acI-B1 MEN L06	>99%	1.2
acI-B1 SPA D18	51%	0.8
acI-B1 DAM I18	74%	0.9
acI-B1 STE P03	34%	0.7

MEN – Lake Mendota; DAM – Lake Damariscotta; SPA – Sparkling Lake; STE – Lake Stechlin
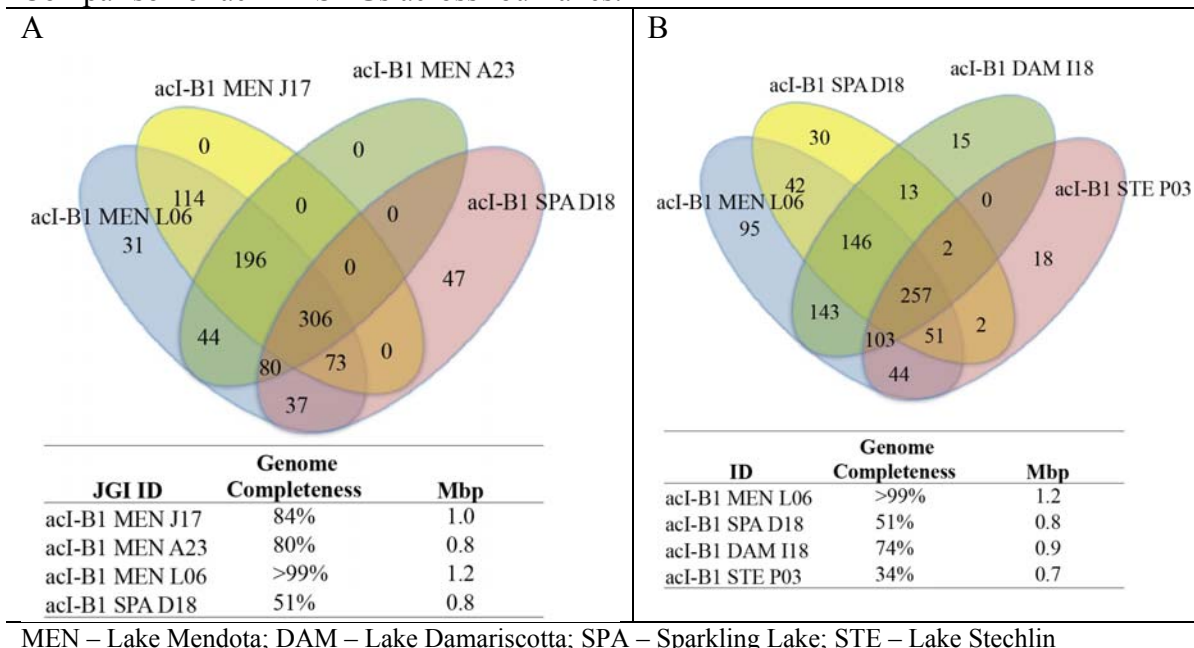
Figure S2. Venn Diagram: Comparison of freshwater bacteria: acI, *Polynucleobacter* (freshwater beta-proteobacteria) and LD12 (freshwater sister group to marine alpha-proteobacteria *Pelagibacter ubique* (SAR11))
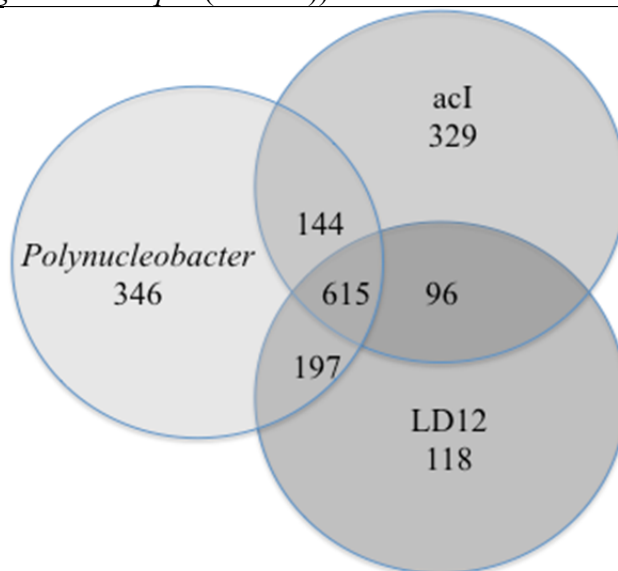
Figure S3. Codon Bias Distribution in Genes of SAG SCGC AAA278-O22 (acI-A1 from Lake Damariscotta)
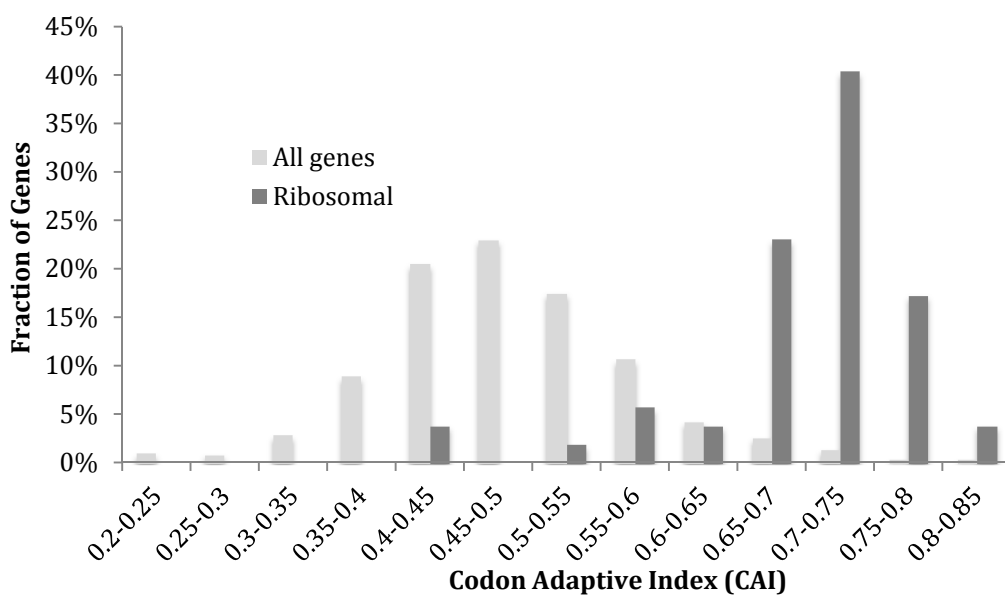
Supplementary Tables

Table S1. Metadata for lakes from which SAGs were obtained

Lake	Surface Area (km^2)	Maximum Depth (m)	Trophic State	Location	Region
Mendota (MEN)	40	25	Eutrophic	43°6'19.58" 89°24'28.71"	Southern WI, USA
Sparkling (SPA)	0.65	20	Oligotrophic	46°0'34.13" 89°42'2.24"	Northern WI, USA
Damariscotta (DAM)	19	35	Mesotrophic	44°10'38.31" 69°29'12.42"	Maine, USA
Stechlin (STE)	4.5	70	Oligotrophic	53°9'5.59" 13°1'34.22"	Germany

Table S2. List of 158 PFAMs used to estimate genome completeness and HMM cutoff used to declare a query gene as a match to the reference PFAM. Shaded rows are PFAMs that were missing in all acI-SAGs.

Name	PFAM accession	HMM length	Cutoff	Description
dsrm	PF00035.20	67	21.40	Double-stranded RNA binding motif
Tubulin	PF00091.20	216	51.90	Tubulin/FtsZ family, GTPase domain
ATP-synt_A	PF00119.15	215	64.10	ATP synthase A chain
ATP-synt_C	PF00137.16	66	21.85	ATP synthase subunit C
RecA	PF00154.16	323	275.00	recA bacterial DNA recombination protein
PGK	PF00162.14	384	235.00	Phosphoglycerate kinase
Ribosomal_S12	PF00164.20	122	85.35	Ribosomal protein S12
Ribosomal_S7	PF00177.16	148	89.40	Ribosomal protein S7p/S5e
Ribosomal_L2	PF00181.18	77	52.00	Ribosomal Proteins L2, RNA binding domain
Ribosomal_S3_C	PF00189.15	85	47.10	Ribosomal protein S3, C-terminal domain
Ribosomal_S19	PF00203.16	81	54.45	Ribosomal protein S19
OSCP	PF00213.13	172	44.35	ATP synthase delta (OSCP) subunit
ATP-synt	PF00231.14	290	155.65	ATP synthase
Ribosomal_L22	PF00237.14	105	56.10	Ribosomal protein L22p/L17e
Ribosomal_L14	PF00238.14	122	76.60	Ribosomal protein L14p/L23e
Ribosomal_L16	PF00252.13	133	74.05	Ribosomal protein L16p/L10e
Ribosomal_L23	PF00276.15	92	36.35	Ribosomal protein L23
Ribosomal_L5	PF00281.14	56	31.85	Ribosomal protein L5
Ribosomal_L3	PF00297.17	263	66.05	Ribosomal protein L3
Ribosomal_L11	PF00298.14	69	44.70	Ribosomal protein L11, RNA binding domain
Bac_DnaA	PF00308.13	219	148.10	Bacterial dnaA protein

Ribosomal_S15	PF00312.17	83	39.30	Ribosomal protein S15
Ribosomal_S2	PF00318.15	211	131.95	Ribosomal protein S2
Ribosomal_L30	PF00327.15	52	25.90	Ribosomal protein L30p/L7e
Ribosomal_S5	PF00333.15	67	43.05	Ribosomal protein S5, N-terminal domain
Ribosomal_S10	PF00338.17	97	56.40	Ribosomal protein S10p/S20e
SecY	PF00344.15	342	174.70	SecY translocase
Ribosomal_L6	PF00347.18	77	57.50	Ribosomal protein L6
Ribosomal_S17	PF00366.15	69	40.00	Ribosomal protein S17
Ribosomal_S9	PF00380.14	121	66.45	Ribosomal protein S9/S16
Ribosomal_S8	PF00410.14	129	70.50	Ribosomal protein S8
Ribosomal_S11	PF00411.14	110	73.35	Ribosomal protein S11
Ribosomal_L20	PF00453.13	108	62.25	Ribosomal protein L20
Ribosomal_L10	PF00466.15	100	37.85	Ribosomal protein L10
RNA_pol_Rpb2_6	PF00562.23	386	252.45	RNA polymerase Rpb2, domain 6
Ribosomal_L13	PF00572.13	128	77.30	Ribosomal protein L13
Ribosomal_L4	PF00573.17	192	97.15	Ribosomal protein L4/L1 family
SecE	PF00584.15	57	25.15	SecE/Sec61-gamma subunits of protein translocation complex
RNA_pol_Rpb1_2	PF00623.15	166	79.90	RNA polymerase Rpb1, domain 2
Guanylate_kin	PF00625.16	183	67.05	Guanylate kinase
UBA	PF00627.26	37	9.95	UBA/TS-N domain
Ribonuclease_3	PF00636.21	114	27.05	Ribonuclease III domain
Ribosomal_L5_C	PF00673.16	95	57.90	ribosomal L5P family C-terminus
Ribosomal_L1	PF00687.16	221	71.05	Ribosomal protein L1p/L10e family
IF3_C	PF00707.17	88	48.00	Translation initiation factor IF-3, C-terminal domain
Ribosomal_L18e	PF00828.14	129	37.30	Ribosomal protein L18e/L15
Ribosomal_L21p	PF00829.16	96	43.30	Ribosomal prokaryotic L21 protein
Ribosomal_L29	PF00831.18	58	35.05	Ribosomal L29 protein
Ribosomal_L18p	PF00861.17	119	56.10	Ribosomal L18p/L5e family
Ribosomal_S16	PF00886.14	62	33.35	Ribosomal protein S16
EF_TS	PF00889.14	221	102.00	Elongation factor TS
GMP_synt_C	PF00958.17	93	56.60	GMP synthase C terminal domain
CPSase_sm_chain	PF00988.17	131	79.40	Carbamoyl-phosphate synthase small chain, CPSase domain
RNA_pol_A_bac	PF01000.21	111	35.50	RNA polymerase Rpb3/RpoA insert domain
Ribosomal_L27	PF01016.14	81	52.40	Ribosomal L27 protein
GTP1_OBG	PF01018.17	156	92.35	GTP1/OBG
CoaE	PF01121.15	180	89.40	Dephospho-CoA kinase
EFP	PF01132.15	55	31.50	Elongation factor P (EF-P) OB domain
CTP_transf_1	PF01148.15	259	78.40	Cytidylyltransferase family
ATP_bind_3	PF01171.15	182	57.55	PP-loop family
RNA_pol_L	PF01193.19	82	31.85	RNA polymerase Rpb3/Rpb11 dimerisation domain

Ribosomal_L17	PF01196.14	97	56.50	Ribosomal protein L17
Ribosomal_L19	PF01245.15	113	70.80	Ribosomal protein L19
Ribosomal_S6	PF01250.12	92	42.15	Ribosomal protein S6
SAICAR_synt	PF01259.13	249	125.40	SAICAR synthetase
Chorismate_synt	PF01264.16	346	206.00	Chorismate synthase
Ribosomal_L9_N	PF01281.14	48	29.30	Ribosomal protein L9, N-terminal domain
RuvA_N	PF01330.16	61	22.10	RuvA N terminal domain
RNase_HII	PF01351.13	198	59.70	Ribonuclease HII
Ribosomal_L25p	PF01386.14	88	35.80	Ribosomal L25p family
tRNA-synt_2d	PF01409.15	247	153.15	tRNA synthetases class II core domain (F)
PseudoU_synth_1	PF01416.15	105	34.30	tRNA pseudouridine synthase
TruB_N	PF01509.13	149	76.35	TruB family pseudouridylate synthase (N terminal domain)
Ribosomal_L35p	PF01632.14	61	25.85	Ribosomal protein L35
Ribosomal_S20p	PF01649.13	84	37.00	Ribosomal protein S20
SmpB	PF01668.13	68	43.45	SmpB protein
Flavokinase	PF01687.12	125	54.75	Riboflavin kinase
Transcrip_reg	PF01709.15	234	146.65	Transcriptional regulator
IPPT	PF01715.12	253	130.00	IPP transferase
Ham1p_like	PF01725.11	189	89.05	Ham1 family
FtsJ	PF01728.14	181	26.80	FtsJ-like methyltransferase
PolyA_pol	PF01743.15	126	46.55	Poly A polymerase head domain
tRNA_m1G_MT	PF01746.16	186	80.30	tRNA (Guanine-1)-methyltransferase
RRF	PF01765.14	165	97.25	Ribosome recycling factor
RimM	PF01782.13	84	31.15	RimM N-terminal domain
Methyltransf_5	PF01795.14	310	163.65	MraW methylase family
zf-CHC2	PF01807.15	98	48.65	CHC2 zinc finger
UPF0102	PF02021.12	93	37.80	Uncharacterised protein family UPF0102
RBFA	PF02033.13	104	38.55	Ribosome-binding factor A
RuvC	PF02075.12	149	80.95	Crossover junction endodeoxyribonuclease RuvC
UPF0054	PF02130.12	145	56.50	Uncharacterized protein family UPF0054
RecR	PF02132.10	41	22.20	RecR protein
Cytidylate_kin	PF02224.13	158	66.95	Cytidylate kinase
NusG	PF02357.14	92	34.15	Transcription termination factor nusG
UPF0079	PF02367.12	123	45.80	Uncharacterised P-loop hydrolase UPF0079
MraZ	PF02381.13	72	50.55	MraZ protein
Seryl_tRNA_N	PF02403.17	108	36.35	Seryl-tRNA synthetase N-terminal domain
Oligomerisation	PF02410.10	100	41.45	Oligomerisation domain
GidB	PF02527.10	184	72.30	rRNA small subunit methyltransferase G
YgbB	PF02542.11	157	77.05	YgbB family
RecO_C	PF02565.10	118	25.30	Recombination protein O C terminal
Exonuc_VII_L	PF02601.10	319	102.30	Exonuclease VII, large subunit
DUF177	PF02620.12	119	38.60	Uncharacterized ACR, COG1399

RecX	PF02631.11	122	36.95	RecX family
GatB_Yqey	PF02637.13	148	60.85	GatB domain
HTH_WhiA	PF02650.9	191	86.35	Sporulation Regulator WhiA C terminal domain
DXP_reductoisom	PF02670.11	129	58.15	1-deoxy-D-xylulose 5-phosphate reductoisomerase
CPSase_L_D3	PF02787.14	121	57.65	Carbamoyl-phosphate synthetase large chain, oligomerisation domain
ATP-synt_DE_N	PF02823.11	80	32.20	ATP synthase, Delta/Epsilon chain, beta-sandwich domain
GARS_C	PF02843.11	93	37.15	Phosphoribosylglycinamide synthetase, C domain
GARS_N	PF02844.10	100	51.30	Phosphoribosylglycinamide synthetase, N domain
Phe_tRNA-synt_N	PF02912.13	73	22.25	Aminoacyl tRNA synthetase class II, N-terminal domain
GatB_N	PF02934.10	289	181.95	GatB/GatE catalytic domain
GAD	PF02938.9	95	30.80	GAD domain
SRP_SPB	PF02978.14	104	51.15	Signal peptide binding domain
DNA_ligase_ZBD	PF03119.11	28	15.95	NAD-dependent DNA ligase C4 zinc finger domain
FDX-ACB	PF03147.9	94	38.10	Ferredoxin-fold anticodon binding domain
GreA_GreB_N	PF03449.10	74	35.30	Transcription elongation factor, N-terminal
TRCF	PF03461.10	101	38.30	TRCF domain
B5	PF03484.10	70	19.75	tRNA synthetase B5 domain
UPF0081	PF03652.10	135	60.20	Uncharacterised protein family (UPF0081)
Ribosomal_S5_C	PF03719.10	74	44.45	Ribosomal protein S5, C-terminal domain
PNPase	PF03726.9	83	28.95	Polyribonucleotide nucleotidyltransferase, RNA binding domain
SecG	PF03840.9	74	26.05	Preprotein translocase SecG subunit
Ribosomal_L11_N	PF03946.9	60	44.50	Ribosomal protein L11, N-terminal domain
Ribosomal_L2_C	PF03947.13	130	86.45	Ribosomal Proteins L2, C-terminal domain
Ribosomal_L9_C	PF03948.9	87	36.45	Ribosomal protein L9, C-terminal domain
RNA_pol_Rpb2_7	PF04560.15	83	45.70	RNA polymerase Rpb2, domain 7
RNA_pol_Rpb2_2	PF04561.9	190	41.35	RNA polymerase Rpb2, domain 2
RNA_pol_Rpb2_1	PF04563.10	203	42.75	RNA polymerase beta subunit
RNA_pol_Rpb2_3	PF04565.11	68	42.70	RNA polymerase Rpb2, domain 3
RNA_pol_Rpb1_3	PF04983.13	157	35.85	RNA polymerase Rpb1, domain 3
RNA_pol_Rpb1_1	PF04997.7	338	146.25	RNA polymerase Rpb1, domain 1
RNA_pol_Rpb1_5	PF04998.12	277	111.40	RNA polymerase Rpb1, domain 5
IF3_N	PF05198.11	76	47.60	Translation initiation factor IF-3, N-terminal domain
RuvB_C	PF05491.8	76	44.10	Holliday junction DNA helicase ruvB C-terminus
Trigger_N	PF05697.8	145	62.60	Bacterial trigger factor protein (TF)
YchF-GTPase_C	PF06071.8	84	60.70	Protein of unknown function (DUF933)
LepA_C	PF06421.7	108	78.05	GTP-binding protein LepA C-terminus
FAD_syn	PF06574.7	158	75.40	FAD synthetase
EFP_N	PF08207.7	58	39.25	Elongation factor P (EF-P) KOW-like domain

Bac_DnaA_C	PF08299.6	70	44.30	Bacterial dnaA protein helix-turn-helix
DXP_redisom_C	PF08436.7	84	53.95	1-deoxy-D-xylulose 5-phosphate reductoisomerase C-terminal
UvrC_HhH_N	PF08459.6	155	84.35	UvrC Helix-hairpin-helix N-terminal
NusA_N	PF08529.6	122	31.30	NusA N-terminal domain
DALR_2	PF09190.6	63	12.90	DALR domain
DUF1967	PF09269.6	69	38.55	Domain of unknown function (DUF1967)
Elong-fact-P_C	PF09285.6	56	37.30	Elongation factor P, C-terminal
Ftsk_gamma	PF09397.5	67	35.15	Ftsk gamma domain
RNA_pol_Rpb2_45	PF10385.4	66	39.05	RNA polymerase beta subunit external 1 domain
DnaB_bind	PF10410.4	59	21.80	DnaB-helicase binding domain of primase
ParA	PF10609.4	81	35.80	ParA/MinD ATPase like
RecO_N	PF11967.3	80	39.35	Recombination protein O N terminal
IF-2	PF11987.3	109	54.35	Translation-initiation factor 2
MgsA_C	PF12002.3	168	108.85	MgsA AAA+ ATPase C terminal
FtsZ_C	PF12327.3	95	52.65	FtsZ family, C-terminal domain
UvrB	PF12344.3	44	33.85	Ultra-violet resistance protein B
DUF3662	PF12401.3	116	56.85	Protein of unknown function (DUF2662)

Table S3 Metadata for JGI genomes

Taxon OID	Status	Genome Name / Sample Name	Genome Size	Gene Count
2236661000	Permanent Draft	alpha proteobacterium SCGC AAA023-L09	774923	921
2236876027	Permanent Draft	alpha proteobacterium SCGC AAA024-N17	328144	397
2264265094	Permanent Draft	alpha proteobacterium SCGC AAA027-C06	775384	936
2236876030	Permanent Draft	alpha proteobacterium SCGC AAA027-J10	792980	952
2236876031	Permanent Draft	alpha proteobacterium SCGC AAA027-L15	719587	840
2236661008	Permanent Draft	alpha proteobacterium SCGC AAA028-C07	846566	974
2236347069	Permanent Draft	alpha proteobacterium SCGC AAA028-D10	925141	1091
2236876032	Permanent Draft	alpha proteobacterium SCGC AAA280-B11	674250	815
2236876029	Permanent Draft	alpha proteobacterium SCGC AAA280-P20	720523	838
2236347068	Permanent Draft	alpha proteobacterium SCGC AAA487-M09	627365	800
640427123	Finished	Mycobacterium tuberculosis F11 (ExPEC)	4424435	4019
2511231114	Finished	Corynebacterium variabile DSM 44702	3433007	3131
650716061	Finished	Mycobacterium sp. JDM601	4643668	4398
2519103194	Permanent Draft	Salinispora arenicola CNR107	5767406	5209
2512047052	Draft	Actinomyces sp. F0384	3133330	2704
648861015	Draft	Rothia dentocariosa M567	2532787	2185
2508501012	Permanent Draft	Saccharomonospora marina XMU15, DSM 45390	5965593	5784
2511231151	Finished	Actinoplanes sp. SE50/110	9239851	8385
2511231111	Finished	Corynebacterium diphtheriae PW8	2530683	2412
2517572163	Permanent Draft	Salinispora pacifica CNY237	5713643	5257
648028043	Finished	Mobiluncus curtisii ATCC 43063	2146480	1965
2515154087	Permanent Draft	Frankia sp. BCU110501	7891711	6839
646311968	Finished	Xylanimonas cellulosilytica DSM 15894	3831380	3549
648276697	Draft	Mycobacterium tuberculosis SUMu005	4331659	4258
2518285526	Permanent Draft	Streptomyces viridosporus T7A, ATCC 39115	8292505	7648
643692033	Finished	Rhodococcus erythropolis PR4	6895538	6505
2507525000	Permanent Draft	Promicromonospora kroppenstedtii RS16, DSM 19349	5802150	6746
2517572001	Permanent Draft	Rothia mucilaginosa Lim (High quality draft)	2278618	1796
2510065062	Permanent Draft	Actinospica robiniae DSM 44927	9918887	8726
2516493018	Permanent Draft	Mycobacterium sp. 155	4609894	4427
2516143102	Draft	Mycobacterium abscessus M93	5078984	4955
647000278	Draft	Mycobacterium parascrofulaceum ATCC BAA-614	6295508	6510
2515154156	Permanent Draft	Corynebacterium propinquum DSM 44285	2458635	2285
642979363	Draft	Mycobacterium tuberculosis EAS054	4366920	4150
2517572166	Permanent Draft	Streptomyces sp. CNS335	8591972	7389
649989929	Draft	Dietzia cinnamea P4	3555295	3593
2516493032	Permanent Draft	Salinispora pacifica DSM 45547	5447211	4981

649989992	Draft	Pseudonocardia sp. P1	6388771	6674
2517287023	Permanent Draft	Salinispora pacifica DSM 45548	5097790	4888
2502376845	Draft	Amycolatopsis sp. ATCC 39116	8342184	8675
646564505	Finished	Arcanobacterium haemolyticum CCM, DSM 20595	1986154	1885
2514752031	Draft	Streptomyces sp. W007	9057348	7969
647000297	Draft	Propionibacterium acnes J165	2500083	2453
2515154141	Permanent Draft	Nonomuraea coxensis DSM 45129	8996452	8548
645951870	Draft	Corynebacterium aurimucosum ATCC 700975	2730325	2672
648276692	Draft	Mycobacterium tuberculosis KZN R506	4367618	4219
648276704	Draft	Mycobacterium tuberculosis SUMu012	4325814	4322
651324106	Draft	Streptomyces griseoaurantiacus M045	7712377	6910
646564564	Finished	Rhodococcus opacus B4	8834939	8259
637000116	Finished	Frankia sp. CcI3	5433628	4621
2515154011	Permanent Draft	Longispora albida DSM 44784	6821241	6439
649989902	Draft	Actinomyces sp. oral taxon 171 F0337	3002669	3202
2517572158	Permanent Draft	Salinispora pacifica CNT045	5771968	5399
2519899568	Draft	Arthrobacter globiformis NBRC 12137	4954410	4582
2511231140	Finished	Streptomyces hygroscopicus jinggangensis 5008	10145833	8935
2506783014	Finished	Mycobacterium chubuense NBB4	6342624	6069
2517572149	Permanent Draft	Micromonospora sp. CNB394	6344798	5836
637000085	Finished	Corynebacterium jeikeium K411	2476822	2186
651324023	Draft	Corynebacterium bovis DSM 20582	2522962	2392
643692028	Finished	Mycobacterium bovis BCG Tokyo 172	4371711	3996
2517093016	Permanent Draft	Saccharomonospora saliphila YIM 90502	4030889	4272
2518645626	Permanent Draft	Salinispora pacifica CNY330	5498479	5176
637000319	Finished	Thermobifida fusca YX	3642249	3195
645951826	Draft	Mycobacterium tuberculosis CPHL_A	4376881	4187
2518645550	Finished	Mycobacterium sp. MOTT36Y	5613626	5177
2513237284	Draft	Candidatus Aquiluna sp. IMCC13023	1359862	1410
645951800	Draft	Streptomyces sp. AA4	9152172	8642
647000216	Draft	Brevibacterium mcbrellneri ATCC 49030	2561804	2489
2517572155	Permanent Draft	Salinispora pacifica CNQ768	5480836	5100
649633057	Finished	Intrasporangium calvum 7KIP, DSM 43043	4024382	3710
649989903	Draft	Actinomyces sp. oral taxon 178 F0338	2733433	2579
2503754045	Draft	Streptomyces viridosporus T7A, ATCC 39115	7868260	8053
644736339	Finished	Catenulispora acidiphila ID139908, DSM 44928	10467782	9125
643886207	Draft	Corynebacterium matruchotii ATCC 33806	2967145	3195
637000215	Finished	Propionibacterium acnes KPA171202	2560265	2385
2509276054	Permanent Draft	Saccharomonospora xinjiangensis XJ-54	4776006	4419
2517572162	Permanent Draft	Salinispora pacifica CNT851	5263032	4974
646564581	Finished	Thermobispora bispora R51, DSM 43833	4189976	3661

2517572159	Permanent Draft	Salinispora pacifica CNT124	5164130	4796
637000330	Finished	Tropheryma whipplei TW08/27	925938	840
2517287019	Permanent Draft	Salinispora pacifica DSM 45544	5464752	5064
643886084	Draft	Corynebacterium pseudogenitalium ATCC 33035	2587776	2660
2517434006	Draft	Brevibacterium casei S18	3663872	3290
645951849	Draft	Streptomyces sp. SPB78	6897976	6544
650716029	Finished	Corynebacterium resistens DSM 45100	2601311	2230
2505679068	Finished	Isoptericola variabilis 225	3307740	3080
639633042	Finished	Mycobacterium sp. KMS	6256079	6089
2517093035	Finished	Corynebacterium ulcerans 0102	2579188	2417
651324091	Draft	Propionibacterium sp. 409-HC1	2528533	2551
645058853	Draft	Streptomyces sp. C	7916041	7768
2513237320	Draft	Gordonia terrae NBRC 100016	5669149	5181
2518285554	Permanent Draft	Salinispora arenicola CNS744	5741573	5262
2518645564	Finished	Corynebacterium pseudotuberculosis 258	2314404	2195
2517093034	Finished	Amycolatopsis mediterranei S699	10246920	9291
641522641	Finished	Mycobacterium abscessus CIP 104536	5090491	4991
2511231083	Finished	Corynebacterium diphtheriae HC01	2427149	2321
637000149	Finished	Leifsonia xyli xyli CTCB07	2584158	2358
2513237387	Draft	Rhodococcus pyridinivorans AK37	5244611	4875
644736391	Finished	Mycobacterium tuberculosis KZN 1435 (MDR)	4398250	4107
2505679089	Finished	Cellulomonas fimi NRS 133, ATCC 484	4266344	3875
2519103109	Finished	Mycobacterium intracellulare ATCC 13950	5402402	5193
2512564010	Permanent Draft	Amycolatopsis thermoflava N1165, DSM 44574	8693468	8633
2512564050	Finished	Mycobacterium intracellulare MOTT-02	5409696	5198
645058700	Draft	Propionibacterium acidifaciens F0233	2064725	1997
2513237290	Draft	Rhodococcus imtechensis RKJ300	8231340	7733
637000115	Finished	Frankia alni ACN14a	7497934	6795
2512047063	Draft	Mycobacterium colombiense CECT 3035	5576382	5279
651053059	Finished	Propionibacterium acnes 6609	2560282	2402
2518285603	Permanent Draft	Actinokineospora enzanensis DSM 44649	8119858	7409
649633006	Finished	Arthrobacter arilaitensis re117, CIP 108037	3918192	3518
648276703	Draft	Mycobacterium tuberculosis SUMu011	4340737	4267
2509601042	Permanent Draft	Jiangella gansuensis YIM 002, DSM 44835	5585780	5250
648276632	Draft	Corynebacterium accolens ATCC 49726	2368466	2416
643692018	Finished	Corynebacterium aurimucosum CN-1, ATCC 700975	2819226	2617
2513237375	Draft	Actinomyces graevenitzii C83	2196917	1910
651324037	Draft	Gordonia neofelifaecis NRRL B-59395	4257286	4034
2513237335	Draft	Nesterenkonia sp. F	2812133	2545
648276698	Draft	Mycobacterium tuberculosis SUMu006	4328768	4241
647000277	Draft	Mobiluncus mulieris 28-1	2452380	2370

2513237078	Draft	Amycolatopsis sp. 75iv2, ATCC 39116	8394307	8314
2515154029	Permanent Draft	Corynebacterium doosanense DSM 45436	2647343	2657
643348509	Finished	Arthrobacter chlorophenolicus A6	4980870	4744
642979350	Draft	Mycobacterium tuberculosis GM 1503	4261330	4163
642555140	Finished	Mycobacterium ulcerans Agy99	5805761	4306
2508501116	Permanent Draft	Saccharomonospora halophila 8	4373567	3713
2512564036	Finished	Corynebacterium pseudotuberculosis 316	2310415	2234
646564561	Finished	Propionibacterium acnes SK137	2495334	2408
642979306	Draft	Corynebacterium amycolatum SK46	2513912	2154
2501939612	Draft	Candidatus Frankia datiscae Dg1	106126	393
643886015	Draft	Actinomyces urogenitalis DSM 15434	2614023	2453
637000170	Finished	Mycobacterium leprae TN	3268203	2750
2511231095	Finished	Propionibacterium acnes TypeIA2 P.acn31	2498766	2298
646311931	Finished	Geodermatophilus obscurus DSM 43160	5322497	5155
647533233	Draft	Streptomyces clavuligerus ATCC 27064	8556892	7381
643348566	Finished	Mycobacterium leprae Br4923	3268071	1654
641736240	Draft	Mycobacterium tuberculosis H37Ra	4326442	4489
645058800	Draft	Rothia mucilaginosa ATCC 25296	2255158	1791
650716060	Finished	Mycobacterium canettii CIPT 140010059	4482059	3909
645951813	Draft	Micrococcus luteus NCTC 2665	2320592	2298
2508501106	Finished	Mycobacterium rhodesiae NBB3	6415739	6342
647000206	Draft	Actinomyces odontolyticus F0309	2422895	2427
2518285536	Permanent Draft	Streptomyces sp. CNS615	7349627	6283
640069320	Finished	Mycobacterium sp. JLS	6048425	5855
2517093017	Permanent Draft	Saccharomonospora halophila 8	3685023	4035
640427122	Finished	Mycobacterium gilvum PYR-GCK	5982829	5683
2518645610	Permanent Draft	Streptomyces sulphureus DSM 40104	7104952	6252
648276750	Draft	Streptomyces violaceusniger Tu 4113	10988130	9557
642979349	Draft	Mycobacterium tuberculosis T17	4277414	4300
648276634	Draft	Corynebacterium matruchotii ATCC 14266	2855988	2679
2511231086	Finished	Kitasatospora setae KM-6054, NBRC 14216	8783278	7669
2515154142	Permanent Draft	Streptomyces vitaminophilus DSM 41686	6429887	5575
2511231115	Finished	Corynebacterium diphtheriae CDCE 8392	2433326	2330
648276689	Draft	Mobiluncus curtisii curtisii ATCC 35241	2136873	1946
648861016	Draft	Streptomyces pristinaespiralis ATCC 25486	8133379	6937
2515154197	Permanent Draft	Streptomyces scabrisporus DSM 41855	11392659	10155
641522620	Finished	Corynebacterium urealyticum DSM 7109	2369219	2084
2511231219	Finished	Propionibacterium acnes ATCC 11828	2488626	2310
2515154155	Permanent Draft	Actinopolymorpha alba DSM 45243	7983713	7545
651053061	Finished	Pseudonocardia dioxanivorans CB1190	7440794	7071
649989974	Draft	Mobiluncus curtisii holmesii ATCC 35242	2087529	1881
645951854	Draft	Mycobacterium tuberculosis K85	4399672	4249

643886002	Draft	Corynebacterium lipophiloflavum DSM 44291	2293743	2427
2516653049	Permanent Draft	Sporichthya polymorpha DSM 43042	5500153	5279
2516143011	Draft	Propionibacterium acnes PRP-38	2507426	2287
651053043	Finished	Mycobacterium tuberculosis CCDC5079	4398812	3695
2517572154	Permanent Draft	Salinispora arenicola CNY280	6066372	5597
2517287026	Permanent Draft	Actinopolyspora halophila DSM 43834	5353291	4973
2519899569	Draft	Gordonia araii NBRC 100433	3906554	3845
641522642	Finished	Mycobacterium marinum M, ATCC BAA-535	6660144	5501
643886147	Draft	Corynebacterium glucuronolyticum ATCC 51867	2784713	2701
646564557	Finished	Nocardiopsis dassonvillei dassonvillei DSM 43111	6543312	5647
2517572122	Permanent Draft	Arthrobacter sp. 131MFCol6.1	4432383	4031
649990018	Draft	Streptomyces sp. SA3_actG	7443083	6612
649989901	Draft	Actinomyces oris K20	2872429	2992
2517572157	Permanent Draft	Salinispora pacifica CNS996	5661283	5349
650377927	Finished	Corynebacterium pseudotuberculosis I19	2337730	2127
643886058	Draft	Corynebacterium accolens ATCC 49725	2413333	2387
2516143052	Draft	Mycobacterium abscessus M94	5095496	5063
2507262030	Permanent Draft	Amycolatopsis sp. 75iv2, ATCC 39116	8442518	8328
2518645556	Finished	Nocardiopsis alba ATCC BAA-2165	5848211	5609
2511231059	Finished	Corynebacterium diphtheriae HC03	2478364	2330
644736323	Finished	Actinosynnema mirum 101, DSM 43827	8248144	7176
2515154018	Permanent Draft	Corynebacterium lubricantis DSM 45231	2944170	2885
2507262046	Permanent Draft	Saccharomonospora saliphila YIM 90502	5038217	6663
2519103058	Draft	Saccharomonospora azurea SZMC 14600	4973727	4604
645058855	Draft	Streptomyces viridochromogenes DSM 40736	8548109	7831
646862346	Finished	Streptomyces bingchenggensis BCW-1	11936683	10106
637000331	Finished	Tropheryma whipplei Twist	927303	864
641228492	Finished	Frankia sp. EAN1pec	8982042	7250
641736194	Draft	Mycobacterium tuberculosis Haarlem	4347292	4376
2518285561	Permanent Draft	Salinispora pacifica CNR942	5473304	5086
2517572123	Permanent Draft	Arthrobacter sp. 135MFCol5.1	4453574	4148
2502376847	Draft	Streptomyces viridosporus T7A	7811542	8295
646564566	Finished	Segniliparus rotundus CDC 1076, DSM 44985	3157527	3126
647000298	Draft	Propionibacterium acnes SK187	2510934	2431
2517287020	Permanent Draft	Amycolatopsis benzoatilytica DSM 43387	8704271	8155
2505679016	Finished	Cellulomonas gilvus ATCC 13127	3526441	3262
637000171	Finished	Mycobacterium sp. MCS	5920523	5704
639633044	Finished	Mycobacterium vanbaalenii PYR-1	6491865	6047
2511231142	Finished	Corynebacterium pseudotuberculosis 1/06-A	2279118	2127
2512047068	Draft	Propionibacterium sp. CC003-HC2	2550549	2455
2517434005	Draft	Brevibacterium casei S18	3663872	3290

2502171150	Draft	Actinomyces naeslundii MG1	3042856	2552
644736380	Finished	Kytococcus sedentarius 541, DSM 20547	2785024	2692
2517093007	Permanent Draft	Frankia sp. BMG5.12	7589313	6342
651053058	Finished	Propionibacterium acnes 266	2494578	2402
2511231091	Finished	Corynebacterium diphtheriae HC02	2468612	2306
2515154143	Permanent Draft	Sciscionella marina DSM 45152	8536451	8212
2511231126	Finished	Corynebacterium diphtheriae 241	2426551	2318
649633084	Finished	Propionibacterium freudenreichii shermanii CIRM-BIA1	2616384	2426
639633006	Finished	Arthrobacter sp. FB24	5070478	4622
645058861	Draft	Mycobacterium tuberculosis KZN 605 (XDR)	4237019	4019
644736390	Finished	Micrococcus luteus Fleming NCTC 2665	2501097	2342
2513237117	Permanent Draft	Propionibacterium acidipropionici DSM 4900	3589364	3371
642791623	Draft	Streptomyces sp. Mg1	7105723	6715
2515154015	Permanent Draft	Actinomycetospora chiangmaiensis DSM 45062	5861079	5756
645058728	Draft	Streptomyces griseoflavus Tu4000	7364052	7207
639279307	Finished	Corynebacterium glutamicum Nakagawa ATCC 13032	3309401	3032
637000173	Finished	Mycobacterium tuberculosis H37Rv (lab strain)	4411532	4062
648231703	Finished	Corynebacterium pseudotuberculosis C231	2328208	2113
647000281	Draft	Mycobacterium tuberculosis KZN V2475	4356301	4211
2517572167	Permanent Draft	Streptomyces sp. CNT372	6456450	5674
644736404	Finished	Saccharomonospora viridis P101, DSM 43017	4308349	3962
643886201	Draft	Rhodococcus erythropolis SK121	6785398	6767
2512564058	Finished	Corynebacterium pseudotuberculosis P54B96	2337657	2207
2519103195	Permanent Draft	Salinispora arenicola CNY234	5569418	5027
2513237185	Finished	Mycobacterium tuberculosis RGTB423	4406587	3670
2515154059	Permanent Draft	Corynebacterium ulceribovis DSM 45146	2300451	2158
2519899837	Draft	Mycobacterium abscessus 47J26	4874047	4872
647000310	Draft	Saccharopolyspora erythraea NRRL 2338	8079083	7356
2518285562	Permanent Draft	Salinispora pacifica CNS055	5263298	4883
2518285553	Permanent Draft	Salinispora arenicola CNS051	5853040	5312
2513237386	Draft	Actinomyces sp. F0330	3385844	3077
2506783060	Finished	Mycobacterium smegmatis JS623	7221766	7311
639633041	Finished	Mycobacterium smegmatis MC2 155	6988209	6941
2502171183	Draft	Pseudonocardia dioxanivorans CB1190	7305198	7060
2511231075	Finished	Corynebacterium diphtheriae 31A	2535346	2458
2511231152	Finished	Mycobacterium bovis BCG Mexico	4350386	4030
648028006	Finished	Amycolatopsis mediterranei U32	10236715	9292
640427140	Finished	Salinispora tropica CNB-440	5183331	4664
650716009	Finished	Amycolicicoccus subflavus DQS3-9A1	4863490	4759
648276691	Draft	Mobiluncus mulieris FB024-16	2384880	2164

2513237121	Permanent Draft	Corynebacterium sputi DSM 45148	2917684	2853
647000231	Draft	Corynebacterium jeikeium ATCC 43734	2426461	2279
2517572214	Permanent Draft	Arthrobacter sp. 162MFSha1.1	4399171	4183
641228502	Finished	Renibacterium salmoninarum ATCC 33209	3155250	3558
650377905	Finished	Arthrobacter phenanthrenivorans Sphe3	4535320	4273
2519103087	Draft	Mycobacterium thermoresistibile ATCC 19527	4870742	4662
651324105	Draft	Streptomyces clavuligerus ATCC 27064	9134976	7680
2510065084	Permanent Draft	Cryptosporangium arvum YU 629-21, DSM 44712	9195993	8650
637000305	Finished	Streptomyces coelicolor A3(2)	9054847	8325
2515075013	Permanent Draft	Amycolatopsis nigrescens CSC17Ta-90, DSM 44992	9111009	8660
647000328	Draft	Streptomyces griseus XyelbKG-1 1	8566464	7450
649633070	Finished	Mycobacterium gilvum Spyr1	5783292	5434
649633089	Finished	Rhodococcus equi 103S	5043170	4570
2517434013	Permanent Draft	Amycolatopsis balhimycina DSM 44591	10858503	9919
648276701	Draft	Mycobacterium tuberculosis SUMu009	4320544	4248
2506783011	Finished	Candidatus Frankia datiscae Dg1	5323186	4579
2511231063	Finished	Propionibacterium acnes TypeIA2 P.acn33	2489623	2287
2517572153	Permanent Draft	Salinispora arenicola CNY011	5722126	5192
648276633	Draft	Corynebacterium genitalium ATCC 33030	2349653	2289
651324006	Draft	Actinomyces sp. oral taxon 448 str. F0400	2801370	2573
645058856	Draft	Streptomyces lividans TK24	8190887	7636
649633045	Finished	Frankia sp. EuI1c	8815781	7262
2513237262	Draft	Corynebacterium casei UCMA 3821	3112736	2925
2513237389	Draft	Propionibacterium avidum ATCC 25577	2533496	2428
646311932	Finished	Gordonia bronchialis DSM 43247	5290012	5002
2517572146	Permanent Draft	Acaricomes phytoseiuli DSM 14247	2419519	2387
2511231113	Finished	Streptomyces cattleya NRRL 8057	8092553	7585
2515154214	Permanent Draft	Propionibacterium acnes DSM 1897	2478994	2329
647533234	Draft	Streptomyces sp. SPB74	6970553	5808
640753031	Finished	Kineococcus radiotolerans SRS30216	4956672	4785
645058823	Draft	Streptomyces albus J1074	6619469	6114
639633040	Finished	Mycobacterium bovis BCG Pasteur 1173P2	4374522	4048
2510065021	Permanent Draft	Streptomyces viridosporus T7A, ATCC 39115	8293298	7648
2517572222	Permanent Draft	Corynebacterium capitovis DSM 44611	1960866	1940
640427108	Finished	Clavibacter michiganensis michiganensis NCPPB 382	3395237	3169
2513237321	Draft	Gordonia sputi NBRC 100414	4952979	4642
2515154055	Permanent Draft	Corynebacterium caspium DSM 44850	1843677	1699
650716057	Finished	Microbacterium testaceum StLB037	3982034	3727
643692008	Finished	Beutenbergia cavernae HKI 0122, DSM 12333	4669183	4278

651053044	Finished	Mycobacterium tuberculosis CCDC5180	4405981	3639
643692019	Finished	Corynebacterium kroppenstedtii DSM 44385	2446804	2073
638341107	Draft	Janibacter sp. HTCC2649	4228723	4156
650716030	Finished	Corynebacterium ulcerans BR-AD22	2606374	2398
642979311	Draft	Mycobacterium tuberculosis T92	4260358	4300
2517572164	Permanent Draft	Salinispora tropica CNS416	5180643	4862
2517572160	Permanent Draft	Salinispora pacifica CNT584	5217717	4912
637000234	Finished	Rhodococcus jostii RHA1	9702737	9242
2506381019	Finished	Frankia sp. EUN1f	9322173	7833
637000169	Finished	Mycobacterium bovis AF2122/97	4345492	4014
2517287016	Permanent Draft	Actinopolyspora iraqiensis IQ-H1	3895973	4294
2519899585	Draft	Rhodococcus opacus M213	9194165	8680
648276695	Draft	Mycobacterium tuberculosis SUMu003	4345124	4290
2511231125	Finished	Propionibacterium acnes TypeIA2 P.acn17	2522885	2317
2513237283	Draft	Rothia aeria F0474	2584293	2333
2504756003	Draft	Isoptericola variabilis 225	3309747	3067
2511231107	Finished	Corynebacterium diphtheriae VA01	2395441	2258
643886057	Draft	Corynebacterium striatum ATCC 6940	2724288	2727
2511231181	Finished	Amycolatopsis mediterranei S699, ATCC 13685	10236779	9639
640427110	Finished	Corynebacterium glutamicum R	3363299	3174
2503283002	Draft	Isoptericola variabilis 225 (Isoptericola variabilis strain 225)	3283612	3239
2508501044	Permanent Draft	Saccharomonospora azurea NA-128, DSM 44631	4763852	4530
645951824	Draft	Actinomyces sp. oral taxon 848 F0332	2518918	2644
2517093032	Finished	Mycobacterium massiliense GO 06	5068807	2677
649989904	Draft	Actinomyces sp. oral taxon 180 F0310	2348120	2124
2518645623	Permanent Draft	Actinomadura flavalba DSM 45200	6167782	5740
644736376	Finished	Jonesia denitrificans 55134, DSM 20603	2749646	2635
648861017	Draft	Streptomyces sviceus ATCC 29083	9313494	8275
649989973	Draft	Mobiluncus curtisii ATCC 51333	2111654	1907
644736345	Finished	Corynebacterium efficiens YS-314	3219505	3064
639633001	Finished	Acidothermus cellulolyticus 11B	2443540	2229
648276694	Draft	Mycobacterium tuberculosis SUMu002	4321025	4250
644736331	Finished	Brachybacterium faecium 6-10, DSM 4810	3614992	3189
645951864	Draft	Mycobacterium tuberculosis T46	4347699	4183
645058827	Draft	Streptomyces roseosporus NRRL 11379	7763119	7141
647000230	Draft	Corynebacterium ammoniagenes DSM 20306	2763612	2704
648276699	Draft	Mycobacterium tuberculosis SUMu007	4321517	4232
2517434008	Permanent Draft	Salinispora pacifica DSM 45549	5198196	5060
642979310	Draft	Mycobacterium tuberculosis 94_M4241A	4410654	4282
639633005	Finished	Arthrobacter aurescens TC1	5226648	4793
2518285558	Permanent Draft	Salinispora arenicola CNY231	5832889	5262

2508501119	Permanent Draft	Saccharomonospora glauca K62, DSM 43769	4564108	4386
643886137	Draft	Corynebacterium glucuronolyticum ATCC 51866	2845674	2787
649989926	Draft	Dermacoccus sp. Ellin185	3115364	2976
2510461000	Permanent Draft	Saccharomonospora paurometabolica YIM 90007	4669397	5115
2511231099	Finished	Corynebacterium diphtheriae BH8	2485519	2436
2511231067	Finished	Corynebacterium pseudotuberculosis 3/99-5	2337938	2239
2509887025	Permanent Draft	Frankia sp. QA3	7590853	6546
2508501039	Permanent Draft	Frankia sp. CN3	9978592	8412
650716103	Finished	Verrucosispora maris AB-18-032	6732271	6069
2513237120	Permanent Draft	Glycomyces tenuis DSM 44171	5735294	5399
649633069	Finished	Micromonospora sp. L5	6962533	6326
2515154014	Permanent Draft	Corynebacterium ciconiae DSM 44920	2545566	2274
2515154017	Permanent Draft	Actinoplanes globisporus DSM 43857	10988507	10116
2511231153	Finished	Corynebacterium pseudotuberculosis CIP52.97	2320595	2194
2512564041	Finished	Mycobacterium intracellulare MOTT-64	5501090	5297
2518285563	Permanent Draft	Salinispora pacifica CNS860	5357469	5040
639279306	Finished	Corynebacterium glutamicum Kalinowski ATCC 13032	3282708	3154
2512564033	Finished	Blastococcus saxobsidens DD2	4875340	4910
2512564056	Finished	Mycobacterium tuberculosis UT205	4418088	3852
646311963	Finished	Thermomonospora curvata DSM 43183	5639016	5061
645058719	Draft	Corynebacterium tuberculostearicum SK141	2372261	2265
2518285552	Permanent Draft	Salinispora arenicola CNP193	5766351	5357
2504756063	Draft	Isoptericola variabilis J5	3247456	3055
649989905	Draft	Actinomyces viscosus C505	3133750	2703
2511231089	Finished	Corynebacterium diphtheriae C7 (beta)	2499189	2414
2511231172	Finished	Nocardia cyriacigeorgica GUH-2	6194645	5560
2518285551	Permanent Draft	Salinispora arenicola CNP105	5919675	5478
2511231070	Finished	Mycobacterium tuberculosis KZN 4207 (DS)	4394985	4043
645058725	Draft	Mycobacterium avium avium ATCC 25291	4857995	4684
2516653079	Permanent Draft	Haloglycomyces albus DSM 45210	3535419	3266
646564565	Finished	Sanguibacter keddieii ST-74, DSM 10542	4253413	3800
2511231130	Finished	Mycobacterium tuberculosis CTRI-2	4398525	3996
2517572124	Permanent Draft	Arthrobacter sp. 161MFSha2.1	4572124	4363
2503754044	Draft	Thermobifida cellulosilytica TB100	4974248	4915
639633039	Finished	Mycobacterium avium 104	5475491	5305
648276696	Draft	Mycobacterium tuberculosis SUMu004	4338184	4283
640963058	Draft	Actinomyces odontolyticus ATCC 17982	2393758	2219
648231702	Finished	Corynebacterium pseudotuberculosis 1002	2335112	2117
647000329	Draft	Streptomyces sp. SirexAA-E	7372760	6593
647000280	Draft	Mycobacterium tuberculosis KZN 4207	4383975	4192
640069329	Finished	Saccharopolyspora erythraea NRRL 2338	8212805	7290

2513237176	Finished	Actinoplanes missouriensis NBRC 102363	8773466	8202
642555124	Finished	Clavibacter michiganensis sepedonicus ATCC 33113	3403786	3168
2515154012	Permanent Draft	Mycobacterium hassiacum DSM 44199	5082325	4948
2519103193	Permanent Draft	Salinispora arenicola CNH962	5434989	5012
651324005	Draft	Actinomyces sp. oral taxon 170 str. F0386	3135160	3093
2518285555	Permanent Draft	Salinispora arenicola CNX481	5570975	5088
2508501052	Permanent Draft	Mycobacterium tusciae JS617	7306413	7168
644736393	Finished	Nakamurella multipartita Y-104, DSM 44233	6060298	5471
2513237388	Draft	Propionibacterium sp. 5_U_42AFAA	2530254	2425
2512564011	Permanent Draft	Amycolatopsis halophila YIM 93223, DSM 45216	5551297	5187
2508501013	Permanent Draft	Saccharomonospora cyanea NA-134	5408301	5196
648276693	Draft	Mycobacterium tuberculosis SUMu001	4356118	3942
647000305	Draft	Rhodococcus equi ATCC 33707	5229298	5116
637000198	Finished	Nocardia farcinica IFM 10152	6292344	6011
2511231212	Finished	Corynebacterium pseudotuberculosis 42/02-A	2337606	2164
2517572156	Permanent Draft	Salinispora pacifica CNS863	5438612	5125
647533235	Draft	Streptomyces sp. e14	7928946	6270
649633093	Finished	Rothia dentocariosa ATCC 17931	2506025	2281
2515154153	Permanent Draft	Corynebacterium pilosum DSM 20521	2531632	2451
2518645624	Permanent Draft	Salinispora tropica CNR699	5608799	5261
646311958	Finished	Streptosporangium roseum DSM 43021	10369518	9510
2519899744	Draft	Mycobacterium fortuitum fortuitum DSM 46621	6300050	6299
638341130	Draft	Mycobacterium tuberculosis C	4276200	4087
640427124	Finished	Mycobacterium tuberculosis H37Ra	4419977	4099
645058822	Draft	Streptomyces roseosporus NRRL 15998	7560086	6986
651053019	Finished	Corynebacterium ulcerans 809	2502095	2246
643886108	Draft	Mobiluncus mulieris ATCC 35243	2398290	2348
638341022	Draft	Brevibacterium linens BL2	4366969	3833
2516493031	Permanent Draft	Actinopolyspora mortivallis HS-1, DSM 44261	4233350	3981
642979328	Draft	Streptomyces clavuligerus ATCC 27064	6729086	6060
647000296	Draft	Propionibacterium acnes J139	2481963	2414
637000304	Finished	Streptomyces avermitilis MA-4680	9119895	7792
648028042	Finished	Micromonospora aurantiaca ATCC 27029	7025559	6360
648276700	Draft	Mycobacterium tuberculosis SUMu008	4327514	4261
2517572165	Permanent Draft	Streptomyces sp. CNQ766	8474577	7248
2517572194	Permanent Draft	Salinispora pacifica DSM 45543	5458289	5120
2515154131	Permanent Draft	Corynebacterium mastitidis DSM 44356	2370005	2302
645058721	Draft	Mycobacterium tuberculosis 98-R604 INH-RIF-EM	4286999	4159
2518285560	Permanent Draft	Salinispora arenicola CNY260	5775769	5256

639633046	Finished	Nocardioides sp. JS614	5293685	4975
649989995	Draft	Segniliparus rugosus ATCC BAA-974	3567567	3565
641522653	Finished	Streptomyces griseus griseus NBRC 13350	8545929	7222
2518645537	Finished	Mycobacterium smegmatis MC2 155	6988208	6745
645058739	Draft	Mycobacterium intracellulare ATCC 13950	5328562	5282
2517287032	Permanent Draft	Arthrobacter sp. MA-N2	4833792	4680
2517572101	Permanent Draft	Frankia sp. DC12	6884336	5933
651324077	Draft	Nocardioidaceae bacterium Broad-1	5892013	5686
2513237122	Permanent Draft	Thermocrispum agreste DSM 44070	4196086	3841
2513237258	Draft	Mobilicoccus pelagius NBRC 104925	3544917	3162
2506783048	Permanent Draft	Mycobacterium rhodesiae JS60	7294533	7116
651324092	Draft	Propionibacterium sp. 434-HC2	2568635	2565
2513237372	Draft	Nocardia brasiliensis ATCC 700358	9489024	8548
637000172	Finished	Mycobacterium tuberculosis CDC1551	4403837	4300
2519103184	Permanent Draft	Smaragdicoccus niigatensis DSM 44881	5320466	5200
2517572152	Permanent Draft	Salinispora arenicola CNX814	5688460	5188
645058857	Draft	Streptomyces hygroscopicus ATCC 53653	10466286	9574
2511231154	Finished	Corynebacterium pseudotuberculosis PAT10	2335323	2200
650716059	Finished	Mycobacterium africanum GM041182	4389314	3880
2518285535	Permanent Draft	Streptomyces sp. CNB091	8228950	7272
643886017	Draft	Actinomyces coleocanis DSM 15436	1719346	1598
2513237119	Permanent Draft	Thermocrispum municipale DSM 44069	4526423	4256
2517572242	Permanent Draft	Amycolatopsis methanolica 239, DSM 44096	7196860	7236
2517572184	Permanent Draft	Streptomyces sp. CNY228	6959154	6082
2512564042	Finished	Gordonia polyisoprenivorans VH2, DSM 44266	5844299	5188
645058824	Draft	Streptomyces ghanaensis ATCC 14672	8223278	7840
647533185	Draft	Micromonospora carbonacea var. africana. ATCC 39149	6819904	5692
637000168	Finished	Mycobacterium avium paratuberculosis K-10	4829781	4415
2518285559	Permanent Draft	Salinispora arenicola CNY256	5781119	5301
648028019	Finished	Corynebacterium pseudotuberculosis FRC41	2337913	2171
642979309	Draft	Mycobacterium tuberculosis 02_1987	4443138	4318
2501939613	Draft	Candidatus Frankia datiscae Dg1	140817	387
2512047058	Draft	Gordonia alkanivorans NBRC 16433	5071550	4709
2515154016	Permanent Draft	Catelliglobosispora koreensis DSM 44566	7688589	7606
2511231084	Finished	Corynebacterium diphtheriae HC04	2484332	2342
2516653042	Permanent Draft	Salinispora pacifica DSM 45546	5450338	5005
2516143008	Permanent Draft	Amycolatopsis alba DSM 44262	9811274	9036
648276690	Draft	Mobiluncus mulieris ATCC 35239	2464242	2430
2518645565	Finished	Corynebacterium pseudotuberculosis Cp162	2293464	2150
2513237214	Finished	Corynebacterium pseudotuberculosis 267	2337628	2249
642979364	Draft	Mycobacterium tuberculosis T85	4299331	4251

2501779502	Draft	Micromonospora aurantiaca ATCC 27029	6975051	6405
2512564073	Finished	Mycobacterium tuberculosis RGTB327	4380119	3739
647000274	Draft	Micrococcus luteus SK58	2622687	2549
650716058	Finished	Microlunatus phosphovorus NM-1	5683123	5391
648276702	Draft	Mycobacterium tuberculosis SUMu010	4352987	4280
641228504	Finished	Salinispora arenicola CNS-205	5786361	5169
2511231088	Finished	Corynebacterium diphtheriae INCA 402	2449071	2292
645951842	Draft	Corynebacterium efficiens YS-314, DSM 44549	3113100	3089
2515154032	Permanent Draft	Amycolatopsis taiwanensis DSM 45107	8777509	8504
2517572161	Permanent Draft	Salinispora pacifica CNT609	5328928	5099
646564571	Finished	Stackebrandtia nassauensis LLR-40K-21, DSM 44728	6841557	6541
645058849	Draft	Mycobacterium kansasii ATCC 12478	6400522	5962
646311938	Finished	Kribbella flavida DSM 17836	7579488	7149
646564576	Finished	Streptomyces scabiei 87.22	10148695	8841
2517572190	Permanent Draft	Streptomyces sp. CNT302	7174271	6237
646564587	Finished	Tsukamurella paurometabola 33, DSM 20162	4479724	4391
2518645627	Permanent Draft	Salinispora pacifica CNY331	5740370	5391
2508501117	Permanent Draft	Actinopolyspora iraqiensis IQ-H1	4276250	3631
651285011	Draft	Streptomyces sp. Tu6071	7506727	6733
637000082	Finished	Corynebacterium diphtheriae gravis NCTC 13129	2488635	2395
645951844	Draft	Aeromicrobium marinum DSM 15272	2585197	3625
2517572233	Permanent Draft	Salinispora arenicola CNT859	5822487	5320
2518285564	Permanent Draft	Streptomyces sp. CNY243	8481985	7235
2515154140	Permanent Draft	Kribbella catacumbae DSM 19601	9627810	9069
2519103192	Permanent Draft	Salinispora arenicola CNH877	5795920	5299
646564520	Finished	Cellulomonas flavigena 134, DSM 20109	4123179	3783
646311951	Finished	Rothia mucilaginosa DY-18	2264603	1965
2519103185	Permanent Draft	Salinispora arenicola CNS673	5875699	5421
647000279	Draft	Mycobacterium tuberculosis 210	4395332	4182
2511231200	Finished	Streptomyces cattleya DSM 46488	8095515	7650
642555133	Finished	Kocuria rhizophila DC2201	2697540	2413

Table S4. AcI SAG overall features averaged within tribes

	Average Estimated Genome Size (Mbp)	Average GC (%)	Total SAG Sequence (Mbp)	Unique COGs[*]	Unique COGs per Mbp	Codon Adaptive Index (CAI)	N:C Content of Proteome
acI-A1	1.6	47%	2.0	58	30	0.48	0.274
acI-A7	1.5	45%	2.8	106	38	0.54	0.274
acI-B1	1.4	41%	5.3	85	16	0.63	0.272

[*]COGs not found in the other two tribes

Table S5. DOC Uptake COGs in acI, LD12, PnecC and Actinomycetales

		acI	LD12	PnecC	Actm
Amino Acids, general	COG0531 Amino acid transporters	100%	0%	100%	0%
	COG0834 ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	27%	50%	100%	97%
	COG1126 ABC-type polar amino acid transport system, ATPase component	27%	60%	100%	78%
	COG0765 ABC-type amino acid transport system, permease component	27%	50%	100%	78%
	COG4597 ABC-type amino acid transport system, permease component	0%	50%	0%	49%
	COG4215 ABC-type arginine transport system, permease component	0%	0%	0%	0%
	COG4160 ABC-type arginine/histidine transport system, permease component	0%	0%	0%	0%
	COG1115 Na+/alanine symporter	0%	0%	0%	97%
	COG0814 Amino acid permeases	0%	0%	0%	12%
	COG4598 ABC-type histidine transport system, ATPase component	0%	0%	0%	0%
	COG4525 ABC-type taurine transport system, ATPase component	0%	0%	0%	0%
	COG3633 Na+/serine symporter	0%	0%	0%	0%
Amino Acids, branched chain	COG0683 ABC-type branched-chain amino acid transport systems, periplasmic component	91%	60%	100%	51%
	COG0410 ABC-type branched-chain amino acid transport systems, ATPase component	91%	60%	100%	48%
	COG0559 Branched-chain amino acid ABC-type transport system, permease components	91%	60%	100%	47%
	COG4177 ABC-type branched-chain amino acid transport system, permease component	91%	60%	100%	46%
	COG0411 ABC-type branched-chain amino acid transport systems, ATPase component	91%	60%	100%	45%
	COG1296 Predicted branched-chain amino acid permease (azaleucine resistance)	45%	80%	100%	62%
	COG1687 Predicted branched-chain amino acid permeases (azaleucine resistance)	0%	0%	0%	30%
	COG1114 Branched-chain amino acid permeases	0%	0%	0%	26%
Amino Acids, di- & oligo-peptides	COG1173 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	73%	0%	0%	100%
	COG0601 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	64%	0%	0%	100%
	COG0747 ABC-type dipeptide transport system, periplasmic component	64%	10%	0%	99%
	COG0444 ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	55%	0%	0%	89%
	COG4166 ABC-type oligopeptide transport system, periplasmic component	27%	0%	100%	96%
	COG4166 ABC-type oligopeptide transport system, periplasmic component	27%	0%	100%	96%
	COG4608 ABC-type oligopeptide transport system, ATPase component	27%	0%	0%	100%
	COG3104 Dipeptide/tripeptide permease	0%	0%	0%	49%
	COG1124 ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	0%	0%	0%	22%
Poly amines	COG1176 ABC-type spermidine/putrescine transport system, permease component I	91%	0%	0%	29%
	COG3842 ABC-type spermidine/putrescine transport systems, ATPase components	82%	0%	100%	63%
	COG1177 ABC-type spermidine/putrescine transport system, permease component II	82%	0%	0%	35%
	COG0687 Spermidine/putrescine-binding periplasmic protein	82%	0%	0%	28%
Lipids	COG0580 Glycerol uptake facilitator and related permeases (Major Intrinsic Protein Family)	64%	50%	0%	63%
	COG2867 Oligoketide cyclase/lipid transport protein	0%	60%	100%	98%
	COG1133 ABC-type long-chain fatty acid transport system, fused permease and ATPase components	0%	60%	0%	0%
	COG2067 Long-chain fatty acid transport protein	0%	0%	0%	0%
tides & Coenz	COG3201 Nicotinamide mononucleotide transporter	100%	0%	0%	71%
	COG2233 Xanthine/uracil permeases	27%	0%	0%	67%
	COG1953 Cytosine/uracil/thiamine/allantoin permeases	0%	0%	0%	27%

COG4143 ABC-type thiamine transport system, periplasmic component	0%	0%	0%	30%
COG1972 Nucleoside permease	0%	0%	0%	7%
COG4145 Na+/panthothenate symporter	0%	0%	0%	2%
COG3840 ABC-type thiamine transport system, ATPase component	0%	0%	0%	0%
COG 5042 Purine nucleoside permease	0%	0%	0%	0%

		acI	LD12	PnecC	Actm
Carbohydrates, general	COG1129 ABC-type sugar transport system, ATPase component	100%	0%	0%	65%
	COG1682 ABC-type polysaccharide/polyol phosphate export systems, permease component	91%	10%	100%	97%
	COG1879 ABC-type sugar transport system, periplasmic component	91%	0%	0%	63%
	COG1134 ABC-type polysaccharide/polyol phosphate transport system, ATPase component	73%	20%	100%	98%
	COG1653 ABC-type sugar transport system, periplasmic component	73%	0%	0%	95%
	COG1175 ABC-type sugar transport systems, permease components	73%	0%	0%	95%
	COG0395 ABC-type sugar transport system, permease component	73%	0%	0%	95%
	COG0738 Fucose permease	64%	0%	0%	47%
	COG3839 ABC-type sugar transport systems, ATPase components	36%	0%	0%	97%
	COG2211 Na+/melibiose symporter and related transporters	36%	0%	0%	57%
	COG2271 Sugar phosphate permease	27%	50%	100%	100%
	COG3822 ABC-type sugar transport system, auxiliary component	9%	0%	0%	0%
	COG1925 Phosphotransferase system, HPr-related proteins	0%	0%	100%	57%
	COG1762 Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)	0%	0%	100%	29%
	COG2893 Phosphotransferase system, mannose/fructose-specific component IIA	0%	0%	100%	4%
	COG2213 Phosphotransferase system, mannitol-specific IIBC component	0%	0%	0%	11%
	COG3730 Phosphotransferase system sorbitol-specific component IIC	0%	0%	0%	3%
	COG3732 Phosphotransferase system sorbitol-specific component IIBC	0%	0%	0%	3%
	COG4158 Predicted ABC-type sugar transport system, permease component	0%	0%	0%	2%
	COG5037 Gluconate transport-inducing protein	0%	0%	0%	0%
Carbs, pentoses	COG1172 Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components	91%	0%	0%	63%
	COG4214 ABC-type xylose transport system, permease component	55%	0%	0%	36%
	COG2182 Maltose-binding periplasmic proteins/domains	18%	0%	0%	42%
	COG3833 ABC-type maltose transport systems, permease component	9%	0%	0%	41%
	COG1455 Phosphotransferase system cellobiose-specific component IIC	0%	0%	0%	2%
Carboxylic Acids	COG1593 TRAP-type C4-dicarboxylate transport system, large permease component	0%	100%	100%	15%
	COG1638 TRAP-type C4-dicarboxylate transport system, periplasmic component	0%	90%	100%	11%
	COG3090 TRAP-type C4-dicarboxylate transport system, small permease component	0%	90%	100%	12%
	COG2358 TRAP-type uncharacterized transport system, periplasmic component	0%	80%	100%	28%
	COG4666 TRAP-type uncharacterized transport system, fused permease components	0%	80%	100%	7%
	COG4664 TRAP-type mannitol/chloroaromatic compound transport system, large permease component	0%	60%	100%	1%
	COG4665 TRAP-type mannitol/chloroaromatic compound transport system, small permease component	0%	60%	100%	1%
	COG4663 TRAP-type mannitol/chloroaromatic compound transport system, periplasmic component	0%	60%	100%	0%
	COG1301 Na+/H+-dicarboxylate symporters	0%	0%	100%	84%
	COG0651 Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	0%	0%	0%	77%
	COG0471 Di- and tricarboxylate transporters	0%	0%	0%	42%
	COG1620 L-lactate permease	0%	0%	0%	40%
	COG1823 Predicted Na+/dicarboxylate symporter	0%	0%	0%	0%
	COG5037 Gluconate transport-inducing protein	0%	0%	0%	0%
Compatible Solutes	COG2113 ABC-type proline/glycine betaine transport systems, periplasmic components	9%	0%	0%	28%
	COG4175 ABC-type proline/glycine betaine transport system, ATPase component	9%	0%	0%	28%
	COG4176 ABC-type proline/glycine betaine transport system, permease	9%	0%	0%	29%

component				
COG1174 ABC-type proline/glycine betaine transport systems, permease component	9%	0%	0%	83%
COG1125 ABC-type proline/glycine betaine transport systems, ATPase components	9%	0%	0%	74%
COG0591 Na+/proline symporter	0%	70%	100%	66%
COG1292 Choline-glycine betaine transporter	0%	0%	0%	83%

Notes: Actm = Actinomycetales; PnecC = *Polynucleobacter necessarius asymbioticus*.
DOC COG list based on (Poretsky et al. 2010)

Table S6. AcI COGs with elevated codon adaptation index (CAI >0.7).

Category	COG Description	COG	CAI for tribe or lineage				Count by tribe		
			A1	A7	B1	acI	A1	A7	B1
Replication, recombination and repair	Bacterial nucleoid DNA-binding protein	776	0.80	0.77	0.77	0.78	1	3	3
Carbohydrate transport and metabolism	**ABC-type xylose transport system, periplasmic component**	4213	0.81	0.78	0.74	0.78	1	3	1
Translation, ribosomal structure and biogenesis	GTPases - translation elongation factors	50	0.80	0.78	0.73	0.77	1	2	2
Carbohydrate transport and metabolism	**ABC-type sugar transport system, periplasmic component**	1653	0.78	0.77	0.76	0.77	2	3	3
Amino acid transport and metabolism	**ABC-type oligopeptide transport system, periplasmic component**	4166		0.77		0.77		3	
Energy production and conversion	Fumarase	114			0.77	0.77			1
General function prediction only	DNA uptake lipoprotein	4105			0.77	0.77			1
Transcription	DNA-directed RNA polymerase, beta' subunit/160 kD subunit	86	0.75	0.75	0.77	0.76	1	2	4
Inorganic ion transport and metabolism	**ABC-type phosphate transport system, periplasmic component**	226	0.77	0.75	0.74	0.75	2	3	4
Posttranslational modification, protein turnover, chaperones	Co-chaperonin GroES (HSP10)	234	0.68	0.76	0.82	0.75	2	1	2
Transcription	DNA-directed RNA polymerase, beta subunit/140 kD subunit	85	0.71	0.78	0.76	0.75	1	2	4
Transcription	**Cold shock proteins**	1278	0.77	0.74	0.73	0.75	1	3	4
Cell cycle control, cell division, chromosome partitioning	Actin-like ATPase involved in cell division	849			0.75	0.75			1
Carbohydrate transport and metabolism	Maltose-binding periplasmic proteins/domains	2182		0.75	0.74	0.75		2	1
Amino acid transport and metabolism	**ABC-type dipeptide transport system, periplasmic component**	747	0.75	0.73	0.73	0.74	1	3	2
Secondary metabolites biosynthesis, transport and catabolism	**Homospermidine synthase**	5310			0.74	0.74	0		3
Posttranslational modification, protein turnover, chaperones	**Peroxiredoxin**	450			0.73	0.73			1
Cell motility	Tfp pilus assembly protein PilF	3063			0.73	0.73			1
Function unknown	Uncharacterized protein conserved in bacteria	3181		0.73		0.73		2	

Function unknown	Uncharacterized protein conserved in bacteria	2318			0.73	0.73			1
Function unknown	Uncharacterized conserved protein	3347			0.73	0.73			1
Coenzyme transport and metabolism	S-adenosylhomocysteine hydrolase	499	0.71	0.74	0.73	0.73	1	3	2
Amino acid transport and metabolism	**ABC-type amino acid transport/signal transduction systems, periplasmic component/domain**	834		0.73		0.73		3	
Translation, ribosomal structure and biogenesis	Polyribonucleotide nucleotidyltransferase (polynucleotide phosphorylase)	1185	0.71	0.73	0.74	0.72	1	1	2
Replication, recombination and repair	ATP-dependent nuclease, subunit B	3857			0.72	0.72			2
Carbohydrate transport and metabolism	Fructose/tagatose bisphosphate aldolase	191	0.71	0.69	0.76	0.72	2	2	1
Posttranslational modification, protein turnover, chaperones	Chaperonin GroEL (HSP60 family)	459	0.69	0.73	0.74	0.72	2	3	4
General function prediction only	Uncharacterized C-terminal domain of topoisomerase IA	1754			0.72	0.72			1
Secondary metabolites biosynthesis, transport and catabolism	Aromatic ring-cleaving dioxygenase	3805			0.72	0.72			1
Transcription	DNA-directed RNA polymerase, alpha subunit/40 kD subunit	202	0.74	0.68	0.72	0.72	1	1	3
Amino acid transport and metabolism	**Glutamine synthetase**	174	0.72	0.71	0.71	0.72	1	3	4
Inorganic ion transport and metabolism	Rhodanese-related sulfurtransferase	2897	0.71			0.71	2		
Amino acid transport and metabolism	Ketol-acid reductoisomerase	59	0.69	0.70	0.75	0.71	2	3	4
General function prediction only	Uncharacterized ABC-type transport system, periplasmic component/surface lipoprotein	1744	0.62	0.78	0.74	0.71	1	3	4
Carbohydrate transport and metabolism	**ABC-type sugar transport system, periplasmic component**	1879	0.69	0.70	0.74	0.71	2	1	4
Posttranslational modification, protein turnover, chaperones	ATPases with chaperone activity, ATP-binding subunit	542	0.67	0.72	0.74	0.71	2	3	4
Translation, ribosomal structure and biogenesis	Translation elongation factor Ts	264	0.71	0.72	0.69	0.71	2	3	4
Amino acid transport and metabolism	**ABC-type branched-chain amino acid transport systems, periplasmic**	683	0.69	0.71	0.72	0.71	2	3	4

	component								
Carbohydrate transport and metabolism	DhnA-type fructose-1,6-bisphosphate aldolase and related enzymes	1830			0.71	0.71			1
Carbohydrate transport and metabolism	**ABC-type sugar transport systems, ATPase components**	3839	0.69	0.74	0.69	0.71	1	3	1
Nucleotide transport and metabolism	Ribonucleotide reductase, alpha subunit	209	0.72	0.69	0.71	0.71	1	1	3
Replication, recombination and repair	RecA/RadA recombinase	468	0.66	0.73	0.73	0.71	1	3	3
General function prediction only	**Bacteriorhodopsin**	5524		0.71	0.70	0.71	0	3	3
Transcription	Transcription antiterminator	250	0.69	0.70	0.73	0.71	1	2	4
General function prediction only	Predicted alternative tryptophan synthase beta-subunit (paralog of TrpB)	1350			0.70	0.70			2
Posttranslational modification, protein turnover, chaperones	Protease subunit of ATP-dependent Clp proteases	740	0.67	0.71	0.73	0.70	1	3	4
General function prediction only	Dioxygenases related to 2-nitropropane dioxygenase	2070			0.70	0.70			1
Amino acid transport and metabolism	**Spermidine/putrescine-binding periplasmic protein**	687	0.67	0.71	0.72	0.70	2	3	3
Energy production and conversion	**Inorganic pyrophosphatase**	3808	0.70	0.70	0.70	0.70	2	3	3
Transcription	DNA-directed RNA polymerase, subunit K/omega	1758	0.64	0.72	0.74	0.70	2	3	3
Nucleotide transport and metabolism	5'-nucleotidase/2',3'-cyclic phosphodiesterase and related esterases	737		0.70		0.70		3	
Posttranslational modification, protein turnover, chaperones	Molecular chaperone	443	0.68	0.70	0.72	0.70	2	1	4
General function prediction only	Predicted RNA-binding protein (contains KH domain)	1837		0.63	0.76	0.70		3	2

Gene count is highlighted in gray where a gene is found in only one tribe.

Table S7. COGs found in acI-A but not acI-B

		Number of SAGs containing the COG		
		acI-A1	acI-A7	acI-A
COG2968	Uncharacterized conserved protein	2	3	5
COG0041	Phosphoribosylcarboxyaminoimidazole (NCAIR) mutase	1	3	4
COG0239	Integral membrane protein possibly involved in chromosome condensation	1	3	4
COG1982	**Arginine/lysine/ornithine decarboxylases**	1	3	4
COG2211	Na+/melibiose symporter and related transporters	1	3	4
COG2385	**Sporulation protein and related proteins**	1	3	4
COG3022	Uncharacterized protein conserved in bacteria	1	3	4
COG3548	Predicted integral membrane protein	1	3	4
COG3889	Predicted solute binding protein	1	3	4
COG0034	Glutamine phosphoribosylpyrophosphate amidotransferase	2	2	4
COG0519	GMP synthase, PP-ATPase domain/subunit	2	2	4
COG0578	**Glycerol-3-phosphate dehydrogenase**	2	2	4
COG1194	A/G-specific DNA glycosylase	2	2	4
COG1544	Ribosome-associated protein Y (PSrp-1)	2	2	4
COG1609	Transcriptional regulators	2	2	4
COG1846	Transcriptional regulators	2	2	4
COG1977	Molybdopterin converting factor, small subunit	2	2	4
COG3250	**Beta-galactosidase/beta-glucuronidase**	2	2	4
COG3483	**Tryptophan 2,3-dioxygenase (vermilion)**	2	2	4
COG3844	**Kynureninase**	2	2	4
COG4799	Acetyl-CoA carboxylase, carboxyltransferase component (subunits alpha and beta)	2	2	4
COG0646	Methionine synthase I (cobalamin-dependent), methyltransferase domain	0	3	3
COG0737	5'-nucleotidase/2',3'-cyclic phosphodiesterase and related esterases	0	3	3
COG0765	**ABC-type amino acid transport system, permease component**	0	3	3
COG0834	**ABC-type amino acid transport/signal transduction systems, periplasmic component/domain**	0	3	3
COG1062	Zn-dependent alcohol dehydrogenases, class III	0	3	3
COG1126	**ABC-type polar amino acid transport system, ATPase component**	0	3	3
COG1574	Predicted metal-dependent hydrolase with the TIM-barrel fold	0	3	3
COG1802	Transcriptional regulators	0	3	3
COG2039	Pyrrolidone-carboxylate peptidase (N-terminal pyroglutamyl peptidase)	0	3	3
COG2258	Uncharacterized protein conserved in bacteria	0	3	3

COG				
COG2971	**Predicted N-acetylglucosamine kinase**	0	3	3
COG4132	ABC-type uncharacterized transport system, permease component	0	3	3
COG4166	**ABC-type oligopeptide transport system, periplasmic component**	0	3	3
COG0138	AICAR transformylase/IMP cyclohydrolase PurH (only IMP cyclohydrolase domain in Aful)	1	2	3
COG0236	Acyl carrier protein	1	2	3
COG0299	Folate-dependent phosphoribosylglycinamide formyltransferase PurN	1	2	3
COG0304	3-oxoacyl-(acyl-carrier-protein) synthase	1	2	3
COG0340	Biotin-(acetyl-CoA carboxylase) ligase	1	2	3
COG0554	**Glycerol kinase**	1	2	3
COG0623	Enoyl-[acyl-carrier-protein] reductase (NADH)	1	2	3
COG1364	N-acetylglutamate synthase (N-acetylornithine aminotransferase)	1	2	3
COG1570	**Exonuclease VII, large subunit**	1	2	3
COG2030	Acyl dehydratase	1	2	3
COG2084	3-hydroxyisobutyrate dehydrogenase and related beta-hydroxyacid dehydrogenases	1	2	3
COG2115	**Xylose isomerase**	1	2	3
COG2313	Uncharacterized enzyme involved in pigment biosynthesis	1	2	3
COG2723	**Beta-glucosidase/6-phospho-beta-glucosidase/beta-galactosidase**	1	2	3
COG3345	**Alpha-galactosidase**	1	2	3
COG3391	Uncharacterized conserved protein	1	2	3
COG4768	Uncharacterized protein containing a divergent version of the methyl-accepting chemotaxis-like domain	1	2	3
COG4992	**Ornithine/acetylornithine aminotransferase**	1	2	3
COG1183	Phosphatidylserine synthase	1	2	3
COG2271	**Sugar phosphate permease**	1	2	3
COG0046	Phosphoribosylformylglycinamidine (FGAM) synthase, synthetase domain	2	1	3
COG0047	Phosphoribosylformylglycinamidine (FGAM) synthase, glutamine amidotransferase domain	2	1	3
COG0140	Phosphoribosyl-ATP pyrophosphohydrolase	2	1	3
COG0150	Phosphoribosylaminoimidazole (AIR) synthetase	2	1	3
COG1268	Uncharacterized conserved protein	2	1	3
COG2230	Cyclopropane fatty acid synthase and related methyltransferases	2	1	3
COG0063	Predicted sugar kinase	0	2	2
COG0257	Ribosomal protein L36	0	2	2
COG0314	Molybdopterin converting factor, large subunit	0	2	2
COG0315	Molybdenum cofactor biosynthesis enzyme	0	2	2
COG0405	Gamma-glutamyltransferase	0	2	2

COG				
COG0521	Molybdopterin biosynthesis enzymes	0	2	2
COG0746	Molybdopterin-guanine dinucleotide biosynthesis protein A	0	2	2
COG1285	Uncharacterized membrane protein	0	2	2
COG1714	Predicted membrane protein/domain	0	2	2
COG1957	Inosine-uridine nucleoside N-ribohydrolase	0	2	2
COG2055	Malate/L-lactate dehydrogenases	0	2	2
COG2407	L-fucose isomerase and related proteins	0	2	2
COG2759	Formyltetrahydrofolate synthetase	0	2	2
COG2807	Cyanate permease	0	2	2
COG2855	Predicted membrane protein	0	2	2
COG2885	Outer membrane protein and related peptidoglycan-associated (lipo)proteins	0	2	2
COG2896	Molybdenum cofactor biosynthesis enzyme	0	2	2
COG3684	Tagatose-1,6-bisphosphate aldolase	0	2	2
COG3894	Uncharacterized metal-binding protein	0	2	2
COG4805	Uncharacterized protein conserved in bacteria	0	2	2
COG2204	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains	0	2	2
COG0153	Galactokinase	1	1	2
COG0331	(acyl-carrier-protein) S-malonyltransferase	1	1	2
COG0332	3-oxoacyl-[acyl-carrier-protein] synthase III	1	1	2
COG0376	Catalase (peroxidase I)	1	1	2
COG0448	ADP-glucose pyrophosphorylase	1	1	2
COG0476	Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 2	1	1	2
COG0736	Phosphopantetheinyl transferase (holo-ACP synthase)	1	1	2
COG1277	ABC-type transport system involved in multi-copper enzyme maturation, permease component	1	1	2
COG1523	Type II secretory pathway, pullulanase PulA and related glycosidases	1	1	2
COG1722	Exonuclease VII small subunit	1	1	2
COG2047	Uncharacterized protein (ATP-grasp superfamily)	1	1	2
COG2887	RecB family exonuclease	1	1	2
COG3064	Membrane protein involved in colicin uptake	1	1	2
COG3448	CBS-domain-containing membrane protein	1	1	2
COG3836	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	1	1	2
COG4770	Acetyl/propionyl-CoA carboxylase, alpha subunit	1	1	2
COG1209	dTDP-glucose pyrophosphorylase	1	1	2
COG1898	dTDP-4-dehydrorhamnose 3,5-epimerase and related enzymes	1	1	2
COG3467	Predicted flavin-nucleotide-binding protein	1	1	2
COG3958	Transketolase, C-terminal subunit	1	1	2
COG3959	Transketolase, N-terminal subunit	1	1	2

COG0067	Glutamate synthase domain 1	2	0	2
COG0079	Histidinol-phosphate/aromatic aminotransferase and cobyric acid decarboxylase	2	0	2
COG0107	Imidazoleglycerol-phosphate synthase	2	0	2
COG0118	Glutamine amidotransferase	2	0	2
COG0131	Imidazoleglycerol-phosphate dehydratase	2	0	2
COG0139	Phosphoribosyl-AMP cyclohydrolase	2	0	2
COG0151	Phosphoribosylamine-glycine ligase	2	0	2
COG0152	Phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) synthase	2	0	2
COG0788	Formyltetrahydrofolate hydrolase	2	0	2
COG1310	Predicted metal-dependent protease of the PAD1/JAB1 superfamily	2	0	2
COG1573	Uracil-DNA glycosylase	2	0	2
COG1828	Phosphoribosylformylglycinamidine (FGAM) synthase, PurS component	2	0	2
COG2227	2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1,4-benzoquinol methylase	2	0	2
COG2897	Rhodanese-related sulfurtransferase	2	0	2

Note: Genes of particular interest and/or discussed in the text are shown in bold.

Table S8. COGs found in acI-B but not acI-A

		Number of acI-B SAGs containing this COG
COG0328	Ribonuclease HI	6
COG1869	ABC-type ribose transport system, auxiliary component	5
COG3375	Uncharacterized conserved protein	5
COG0388	Predicted amidohydrolase	4
COG0671	Membrane-associated phospholipid phosphatase	4
COG1072	Panthothenate kinase	4
COG5310	Homospermidine synthase	4
COG1350	Predicted alternative tryptophan synthase beta-subunit (paralog of TrpB)	3
COG2040	Homocysteine/selenocysteine methylase (S-methylmethionine-dependent)	3
COG2740	Predicted nucleic-acid-binding protein implicated in transcription termination	3
COG3588	Fructose-1,6-bisphosphate aldolase	3
COG3805	Aromatic ring-cleaving dioxygenase	3
COG4757	Predicted alpha/beta hydrolase	3

Table S9. Average percent protein identity between SAGs

	acIA1 027-M14 MEN	acIA1 278-O22 DAM	acIA7 023-J06 SPA	acIA7 024-D14 SPA	acIA7 044-N04 DAM	acIB1 027-L06 MEN	acIB1 028-A23 MEN	acIB1 027-J17 MEN	acIB1 023-D18 SPA
acIA1 027-M14 MEN									
acIA1 278-O22 DAM	80.2								
acIA7 023-J06 SPA	65.4	67.4							
acIA7 024-D14 SPA	68.5	70.4	90.5						
acIA7 044-N04 DAM	68.7	71.0	91.3	90.3					
acIB1 027-L06 MEN	60.6	62.0	61.6	63.4	62.6				
acIB1 028-A23 MEN	61.6	63.9	61.6	63.7	63.0	99.3			
acIB1 027-J17 MEN	59.2	62.2	60.6	63.1	62.6	84.2	85.4		
acIB1 023-D18 SPA	60.4	62.1	60.3	61.8	63.5	79.4	79.4	79.7	
acIB1 278-I18 DAM	59.8	61.7	60.7	63.0	62.6	80.1	80.4	80.2	80.0

Table S10. COGs over-represented in acI as compared to *Polynucleobacter necessarius asymbioticus*. Data represents percentage of genomes containing each COG.

COG Category	Description	acI	Pnec
Amino Acids	**COG1176 ABC-type spermidine/putrescine transport system, permease component I**	**91%**	**0%**
Amino Acids	COG0010 Arginase/agmatinase/formimionoglutamate hydrolase, arginase family	91%	0%
Amino Acids	COG0160 4-aminobutyrate aminotransferase and related aminotransferases	82%	0%
Amino Acids	COG2008 Threonine aldolase	82%	0%
Amino Acids	**COG0687 Spermidine/putrescine-binding periplasmic protein**	**82%**	**0%**
Amino Acids	**COG1177 ABC-type spermidine/putrescine transport system, permease component II**	**82%**	**0%**
Amino Acids	COG3200 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase	82%	0%
Amino Acids	COG0083 Homoserine kinase	73%	0%
Amino Acids	**COG1173 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components**	**73%**	**0%**
Amino Acids	COG0404 Glycine cleavage system T protein (aminomethyltransferase)	73%	0%
Amino Acids	COG1003 Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain	73%	0%
Amino Acids	COG1748 Saccharopine dehydrogenase and related proteins	64%	0%
Amino Acids	**COG0601 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components**	**64%**	**0%**
Amino Acids	**COG0747 ABC-type dipeptide transport system, periplasmic component**	**64%**	**0%**
Amino Acids	COG4448 L-asparaginase II	64%	0%
Amino Acids	COG2873 O-acetylhomoserine sulfhydrylase	64%	0%
Carbohydrates	**COG1129 ABC-type sugar transport system, ATPase component**	**100%**	**0%**
Carbohydrates	**COG1172 Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components**	**91%**	**0%**
Carbohydrates	**COG1879 ABC-type sugar transport system, periplasmic component**	**91%**	**0%**
Carbohydrates	**COG2017 Galactose mutarotase and related enzymes**	**82%**	**0%**
Carbohydrates	COG0205 6-phosphofructokinase	82%	0%
Carbohydrates	COG0698 Ribose 5-phosphate isomerase RpiB	82%	0%
Carbohydrates	COG3386 Gluconolactonase	73%	0%
Carbohydrates	COG1082 Sugar phosphate isomerases/epimerases	73%	0%
Carbohydrates	**COG0395 ABC-type sugar transport system, permease component**	**73%**	**0%**
Carbohydrates	**COG1175 ABC-type sugar transport systems, permease components**	**73%**	**0%**
Carbohydrates	**COG1653 ABC-type sugar transport system, periplasmic component**	**73%**	**0%**
Carbohydrates	COG1070 Sugar (pentulose and hexulose) kinases	64%	0%

Carbohydrates	**COG0738 Fucose permease**	**64%**	**0%**
Carbohydrates	**COG0580 Glycerol uptake facilitator and related permeases (Major Intrinsic Protein Family)**	**64%**	**0%**
Cell Division	COG2177 Cell division protein	100%	0%
Cell Division	COG2884 Predicted ATPase involved in cell division	100%	0%
Cell wall	COG1247 Sortase and related acyltransferases	64%	0%
Cell wall	**COG3757 Lyzozyme M1 (1,4-beta-N-acetylmuramidase)**	**64%**	**0%**
Coenzyme	COG3201 Nicotinamide mononucleotide transporter	100%	0%
Coenzyme	COG0111 Phosphoglycerate dehydrogenase and related dehydrogenases	91%	0%
Coenzyme	COG0447 Dihydroxynaphthoic acid synthase	82%	0%
Coenzyme	COG1165 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase	82%	0%
Coenzyme	COG1169 Isochorismate synthase	82%	0%
Coenzyme	COG0214 Pyridoxine biosynthesis enzyme	64%	0%
Coenzyme	COG0311 Predicted glutamine amidotransferase involved in pyridoxine biosynthesis	64%	0%
Defense	COG0842 ABC-type multidrug transport system, permease component	73%	0%
Energy	COG0584 Glycerophosphoryl diester phosphodiesterase	91%	0%
Energy	COG0667 Predicted oxidoreductases (related to aryl-alcohol dehydrogenases)	91%	0%
Energy	**COG2352 Phosphoenolpyruvate carboxylase**	**73%**	**0%**
Energy	COG1048 Aconitase A	73%	0%
Energy	COG0538 Isocitrate dehydrogenases	64%	0%
General	COG1847 Predicted RNA-binding protein	100%	0%
General	COG2372 Uncharacterized protein, homolog of Cu resistance protein CopC	100%	0%
General	COG2409 Predicted drug exporters of the RND superfamily	100%	0%
General	COG5512 Zn-ribbon-containing, possibly RNA-binding protein and truncated derivatives	100%	0%
General	**COG5524 Bacteriorhodopsin**	**73%**	**0%**
General	COG1090 Predicted nucleoside-diphosphate sugar epimerase	64%	0%
General	COG0121 Predicted glutamine amidotransferase	64%	0%
General	COG0637 Predicted phosphatase/phosphohexomutase	64%	0%
Inorganic ion	COG1119 ABC-type molybdenum transport system, ATPase component/**photorepair** protein PhrA	73%	0%
Lipids	COG1597 Sphingosine kinase and enzymes related to eukaryotic diacylglycerol kinase	64%	0%
Nucleotides	COG0213 Thymidine phosphorylase	91%	0%
Nucleotides	COG1457 Purine-cytosine permease and related proteins	91%	0%
Nucleotides	**COG1816 Adenosine deaminase**	**91%**	**0%**
Nucleotides	**COG0572 Uridine kinase**	**91%**	**0%**
Nucleotides	**COG0295 Cytidine deaminase**	**82%**	**0%**

Nucleotides	COG0274 Deoxyribose-phosphate aldolase	73%	0%
Nucleotides	COG0005 Purine nucleoside phosphorylase	73%	0%
Nucleotides	COG0634 Hypoxanthine-guanine phosphoribosyltransferase	73%	0%
Nucleotides	COG0044 Dihydroorotase and related cyclic amidohydrolases	73%	0%
Nucleotides	COG0035 Uracil phosphoribosyltransferase	64%	0%
Nucleotides	COG1435 Thymidine kinase	64%	0%
Nucleotides	COG0232 dGTP triphosphohydrolase	64%	0%
Protein mod	COG1333 ResB protein required for cytochrome c biosynthesis	73%	0%
Protein mod	COG1928 Dolichyl-phosphate-mannose--protein O-mannosyl transferase	73%	0%
Protein mod	COG0785 Cytochrome c biogenesis protein	73%	0%
DNA Processes	COG0648 Endonuclease IV	91%	0%
DNA Processes	COG2452 Predicted site-specific integrase-resolvase	73%	0%
DNA Processes	COG2816 NTP pyrophosphohydrolases containing a Zn-finger, probably nucleic-acid-binding	64%	0%
DNA Processes	COG4581 Superfamily II RNA helicase	64%	0%
Sec Metabolites	**COG1233 Phytoene dehydrogenase and related proteins**	**91%**	**0%**
Sec Metabolites	**COG4242 Cyanophycinase and related exopeptidases**	**73%**	**0%**
Sec Metabolites	COG2124 Cytochrome P450	64%	0%
Signals	COG0631 Serine/threonine protein phosphatase	91%	0%
Signals	COG3920 Signal transduction histidine kinase	73%	0%
Signals	COG4585 Signal transduction histidine kinase	73%	0%
Signals	COG3707 Response regulator with putative antiterminator output domain	64%	0%
Signals	**COG1702 Phosphate starvation-inducible protein PhoH, predicted ATPase**	**64%**	**0%**
Transcription	COG1167 Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family) and their eukaryotic orthologs	100%	0%
Transcription	COG1316 Transcriptional regulator	91%	0%
Transcription	COG2378 Predicted transcriptional regulator	91%	0%
Transcription	COG1061 DNA or RNA helicases of superfamily II	82%	0%
Transcription	COG1940 Transcriptional regulator/sugar kinase	73%	0%
Transcription	COG1329 Transcriptional regulators, similar to M. xanthus CarD	73%	0%
Translation	COG0349 Ribonuclease D	73%	0%
Translation	COG0423 Glycyl-tRNA synthetase (class II)	64%	0%
Translation	COG2519 tRNA(1-methyladenosine) methyltransferase and related methyltransferases	64%	0%

Notes: Genes of particular interest are shown in **bold**

Table S11. COG Category Comparison; acI vs *Polynucleobacter* (Pnec)

	acI	Pnec	acI/Pnec
Amino acid transport and metabolism	92	117	0.8
Carbohydrate transport and metabolism	38	46	0.8
Cell cycle control, cell division, chromosome partitioning	11	16	0.7
Cell motility	0	0	n/a
Cell wall/membrane/envelope biogenesis	39	93	0.4
Coenzyme transport and metabolism	40	86	0.5
Defense mechanisms	6	12	0.5
Energy production and conversion	61	105	0.6
Function unknown	37	149	0.2
General function prediction only	68	129	0.5
Inorganic ion transport and metabolism	17	61	0.3
Intracellular trafficking, secretion, and vesicular transport	13	21	0.6
Lipid transport and metabolism	22	52	0.4
Nucleotide transport and metabolism	38	45	0.8
Posttranslational modification, protein turnover, chaperones	31	73	0.4
Replication, recombination and repair	45	75	0.6
Secondary metabolites biosynthesis, transport and catabolism	6	20	0.3
Signal transduction mechanisms	18	30	0.6
Transcription	29	39	0.7
Translation, ribosomal structure and biogenesis	99	124	0.8

Table S12. COGs unique to specific acI tribes having CAI > 0.6

acI-A1				
	COG	**acI-A1 CAI**	**acI-A1 Count**	**acI-A7 Count**	**acI-B1 Count**
Rhodanese-related sulfurtransferase	2897	0.71	2	0	0
Periplasmic glycine betaine/choline-binding (lipo)protein of an ABC-type transport system (osmoprote	1732	0.69	1	0	0
Predicted 6-phosphogluconate dehydrogenase	1023	0.63	1	0	0
Predicted sugar nucleotidyltransferases	1213	0.63	1	0	0
acI-A7					
	COG	**acI-A7 CAI**	**acI-A1 Count**	**acI-A7 Count**	**acI-B1 Count**
Ribosomal protein L36	257	0.81	0	3	0
ABC-type oligopeptide transport system, periplasmic component	4166	0.77	0	3	0
Uncharacterized protein conserved in bacteria	3181	0.73	0	2	0
ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	834	0.73	0	3	0
5'-nucleotidase/2',3'-cyclic phosphodiesterase and related esterases	737	0.70	0	3	0
Uncharacterized metal-binding protein	3894	0.63	0	1	0
Predicted peptidase	4099	0.63	0	1	0
Methionine synthase I (cobalamin-dependent), methyltransferase domain	646	0.63	0	3	0
Uncharacterized protein conserved in bacteria	4805	0.62	0	3	0
Formyltetrahydrofolate synthetase	2759	0.61	0	1	0
Subtilisin-like serine proteases	1404	0.60	0	2	0
acI-B1					
	COG	**acI-B1 CAI**	**acI-A1 Count**	**acI-A7 Count**	**acI-B1 Count**
Homospermidine synthase	5310	0.74	0	0	3
ATP-dependent nuclease, subunit B	3857	0.72	0	0	2
Predicted alternative tryptophan synthase beta-subunit (paralog of TrpB)	1350	0.7	0	0	2
Fructose-1,6-bisphosphate aldolase	3588	0.67	0	0	2
3-hexulose-6-phosphate synthase and related proteins	269	0.65	0	0	2
Predicted nucleic-acid-binding protein implicated in transcription termination	2740	0.65	0	0	2
Uncharacterized conserved protein	3375	0.65	0	0	3
ABC-type ribose transport system, auxiliary component	1869	0.64	0	0	3
Panthothenate kinase	1072	0.64	0	0	2
Ribonuclease HI	328	0.62	0	0	4
Predicted membrane protein	2246	0.62	0	0	2
Homocysteine/selenocysteine methylase (S-methylmethionine-dependent)	2040	0.62	0	0	2
Membrane-associated phospholipid phosphatase	671	0.61	0	0	3

Table S13. Categories of acI COGs with CAI >0.6

COG Category %	COG Category Count	Category #	Category Description
14%	42	3	Energy production and conversion
12%	36	5	Amino acid transport and metabolism
9%	27	18	General function prediction only
9%	25	7	Carbohydrate transport and metabolism
9%	25	10	Translation, ribosomal structure and biogenesis
7%	19	15	Posttranslational modification, protein turnover, chaperones
5%	16	11	Transcription
4%	11	8	Coenzyme transport and metabolism
4%	11	12	Replication, recombination and repair
3%	10	6	Nucleotide transport and metabolism
3%	10	16	Inorganic ion transport and metabolism
3%	9	9	Lipid transport and metabolism
3%	8	13	Cell wall/membrane/envelope biogenesis
2%	5	4	Cell cycle control, cell division, chromosome partitioning
1%	3	17	Secondary metabolites biosynthesis, transport and catabolism
0%	1	14	Cell motility

4. FRESHWATER ACTINOBACTERIA LINKED TO GRAZED AND DECAYING PHYTOPLANKTON BLOOMS

Trevor W Ghylin[1], Ben Crary[1], Diana Medina[2], Jack Gilbert[3], Katherine D McMahon[1,2]

1 Department of Civil and Environmental Engineering, University of Wisconsin, Madison, WI, USA

2 Department of Biology, University of Puerto Rico, Mayaguez, Puerto Rico, USA

3 Earth Microbiome Project, Argonne National Lab, Chicago, IL, USA

*correspondence: Katherine McMahon, Department of Civil and Environmental Engineering, University of Wisconsin 1550 Linden Drive Room 5552 Madison, WI 53706, Email: kdmcmahon@cae.wisc.edu

Key words: 16S metagenomics, qPCR, acI, freshwater Actinobacteria, diatoms, cyanobacteria, phytoplankton

4.1. ABSTRACT

Actinobacteria are frequently the most abundant bacterial phylum in freshwater lakes with the majority of these in the acI lineage. Due to their abundance, these bacteria are likely important for water quality, lake ecology, and carbon and nitrogen cycling. However, due to the lack of a viable monophyletic cell culture, these organisms have been difficult to study. In order to study these bacteria in situ, we utilized molecular methods such as quantitative polymerase chain reaction (qPCR) and Illumina 16S sequencing (iTag) to monitor bacterial populations in Lake Mendota (Madison, WI) from 2000 to 2010 in conjunction with environmental data collected by the North Temperate Lakes Long Term Ecological Research Team (NTL-LTER). We observed a significant decline in acI during 2006, corresponding to the lack of a usually predictable spring diatom bloom. Our results demonstrate an interesting shift in Actinobacteria community in early spring (March) and also during the 2006 period in which acI decreases and non-acI (acIV, acSTL, acTH, Luna) clades increase, resulting in a net increase in total Actinobacteria. The lack of a diatom bloom in 2006 corresponding to a substantial decrease in acI suggests a relationship between these two groups. We found a significant positive correlation between acI-A6 and diatoms and a significant positive correlation between acI-B1 and *Aphanizomenon* (cyanobacteria). Both of these correlations were strongest with a lag time of 4 weeks, indicating acI abundance increases weeks after phytoplankton blooms. This may be evidence that acI consumes decay products from phytoplankton blooms rather than exudate from live blooms. Surprisingly, one acI tribe, acI-C2, had very strong correlations with several species of cyanobacteria with no lag, indicating that acI-C2 associates with active phytoplankton blooms, unlike the other acI tribes.

Our previous genomic study found genes for oligopeptide and polyamine uptake and metabolism in acI, which are key cell-wall components of diatoms and are likely released during grazing, lysis and decay of diatom blooms. AcI likely obtains a significant source of energy and nutrients from long chain polyamines and polypeptides (i.e. silaffins) known to occur in diatoms. These compounds may be difficult for most bacteria to consume, providing acI with a source of nutrients and energy not accessible to its competitors. A similar relationship may be true between acI and Cyanobacteria in which acI consumes cyanophycin and cell wall/membrane components of post-bloom cyanobacteria. The results of this study indicate that acI may obtain some of its energy and nutrient requirements during the decay of phytoplankton blooms.

4.2. INTRODUCTION

The acI lineage of Actinobacteria are often the numerically dominant microbial group in lakes, sometimes representing more than 50% of the bacterial population (Newton, 2011; Glockner, 2000; Warnecke, 2005). However, these bacteria have been difficult to study in a controlled lab setting due to difficulty in cultivation (Jezbera, 2009; Garcia, 2013). As a result, many studies have utilized molecular tools such as diversity studies using the 16S rRNA gene to study acI populations in situ (Warnecke, 2004; Newton, 2007; Sharma, 2009) and fluorescent in situ hybridization (FISH) (Warnecke, 2005; Allgaier, 2006; Salcher, 2010; Beier, 2011; Salcher, 2012). The FISH studies typically have limited phylogenetic resolution. As a result many studies simply group all Actinobacteria together. However, Allgaier et al (Allgaier, 2006) used FISH probes to distinguish between acI-A and acI-B and found acI-A had less seasonality than acI-B, which had more distinct spring and fall peaks in abundance.

This FISH study was somewhat limited as the FISH probes were unable to detect 9-40% of acI cells and could not resolve the phylogeny to detect tribes (i.e. acI-B1, acI-A1, acI-A6, etc.). It is possible that the acI-A tribes had seasonal fluctuations that weren't apparent when grouped together.

Temporal studies of freshwater bacteria have mostly shown that the freshwater Actinobacteria abundance (as % of total bacteria) peaks in spring (April/May) and fall (August-November) with slightly lower relative abundance during summer months (Allgaier, 2006; Zeder, 2009; Salcher, 2010). The spring peak generally coincides with a peak in heterotrophic nanoflagellates (grazers) that lags a spring phytoplankton bloom by a few weeks. The spring phytoplankton bloom is frequently dominated by diatoms that are able grow quickly in cold water. The Actinobacteria population typically correlates with oxygen and has a negative correlation with dissolved phosphorus (Glockner, 2000; Burkert, 2003; Allgaier, 2006; Zeder, 2009; Salcher, 2010; Garcia, 2013a). Most studies have been limited to 1 or 2 seasons/years and have limited phylogenetic resolution (i.e. may identify phylum only). Our study expands on the previously published studies by utilizing two independent methods (qPCR and 16S amplicon analysis) over a longer period of time (10 years of samples) and with much greater phylogenetic resolution (i.e. distinguish between tribes such as acI-A1 and acI-B1) than previous studies (Allgaier, 2007; Newton, 2007). We utilized qPCR and iTag (Illumina based 16S rRNA gene sequencing) methods to track bacterial abundance in Lake Mendota (Madison, WI) from 2000 to 2010. QPCR was used to track the abundance of several tribes (acI-A1, acI-A2, acI-A4, acI-A6, and acI-B1) of freshwater Actinobacteria, while the iTag data allowed us to track the abundance of all bacterial OTUs.

The qPCR data is an important complement to the iTag data because the 16S sequencing is performed after a 16S PCR step and is prone to PCR bias. The 16S metagenomic data also allows us to correlate the abundance of acI clades with various other bacteria in the iTag data. The qPCR and iTag data were also coupled with LTER (long term ecological research) data, which includes extensive long-term physical and chemical data for Lake Mendota (Magnuson, 2005; Carpenter, 2006; Shade, 2007) to investigate chemical, biological and physical drivers of acI population change.

4.3. MATERIALS AND METHODS

4.3.1. LAKE CHARACTERISTICS

Lake Mendota (Madison, Wisconsin; 43°06'N, 89°24'W), is one of the most well-studied lakes in the world, and it is a Long Term Ecological Research Site affiliated with the Center for Limnology at the University of Wisconsin (Carpenter, 2006). It is dimictic and eutrophic with an average depth of 12.8 m, maximum depth of 25.3 m, and total surface area of 39.38 km2. It typically mixes in March-April, and again in October-November, but remains thermally stratified throughout the summer and winter months. Ice-on averages 119 d in winter, though this number appears to be gradually decreasing as a result of climate change (Magnuson, 2005).

4.3.2. SAMPLE COLLECTION

Lake water was collected from the top 12m of the epilimnion over the deepest part of the pelagic zone (-24 m) every two weeks during the open-water phase, as described originally (Yannarell, 2003) but with minor modifications (Yannarell, 2005). Bacteria were recovered by filtration on 0.2-um polyethersulfone filters (Pall-Supor200, Gelman), without

prefiltration. Filters were frozen at -80°C and stored prior to deoxyribonucleic acid (DNA)

extraction. Physical and environmental parameters of Lake Mendota were measured by the

North Temperate Lakes Long Term Ecological Research (NTL-LTER) project at the

University of Wisconsin (UW) - Madison Center for Limnology every two weeks. These data

are available through the NTL-LTER website at http://www.limnology.wisc.edu. Figure 1

and Figure 2 provide information about the time distribution of the samples used in this study.

QPCR was performed on 142 samples and iTag was performed on 95 samples.

4.3.3. SAMPLE PROCESSING

Total community DNA was extracted from filters using the QBiogene Bio101 FastDNA kit,

using manufacturer's instruction with modification as described previously (Yannarell, 2003;

Yannarell, 2005).

4.3.4. ILLUMINA ITAG

Extracted DNA samples from the Lake Mendota samples were sent to the Earth Microbiome

Project (http://www.earthmicrobiome.org/emp-standard-protocols/) for 16S analysis on the

Illumina platform (Caporaso, 2012). Samples were PCR amplified targeting the v4 region of

the 16S rRNA gene using bacteria/archaeal 515f/806r primers (Caporaso, 2012).

4.3.5. BIOINFORMATICS

All sequences were processed with Mothur, using the Schloss SOP with minor modifications

(Schloss, 2009). Chimeras were removed with the uchime algorithm (Edgar, 2011). Reads

were aligned to the Greengenes database and subsampled to 12,904 reads per sample. OTUs

were clustered at 97% using the average neighbor linkage algorithm.

OTUs were first classified against a custom curated taxonomic database that includes common freshwater taxa (Newton, 2011). OTUs that could not be assigned with 70% confidence to the 5th taxonomic level with this database were reclassified against the Greengenes. Classifications were kept if they were made with at least 60% confidence.

4.3.6. QPCR

QPCR reactions were performed in a 96 well optical Icycler plate with clear optical film cover. Each well was filled with 7.5 ul of IQ Sybrgreen supermix, 0.9 ul forward primer, 0.9 ul reverse primer, 4.7 ul de-ionized water, 1 ul template (extracted DNA from lake sample).

Quantitative polymerase chain reaction (qPCR) was performed on sections of the 16S rRNA gene with primers and annealing temperatures shown in Table 5. Standards and negative controls were created from a clone library obtained from Lake Mendota. Eight-point calibration curves for qPCR were produced by 10-fold serial dilution of positive controls in duplicate within each assay, at 10^1 to 10^8 target copies per reaction. QPCR was conducted on a Biorad Icycler, and conditions were as follows: 2 min at 50°C, 3 min initial denaturation at 95°C, 40 cycles of 40 s denaturation at 94°C, 30 s annealing at temperature shown in Table 5, 30s elongation at 72°C. The qPCR run was immediately followed by a melt curve step starting at 55°C for 10s and increasing setpoint temperature by 0.5°C every 10s until 95°C (about 80 steps). The machine finished with a final 55°C for 1 min, then infinite hold at 4°C.

Positive and negative controls were generated from appropriate clones from 16S plus internal transcribed spacer rRNA clone libraries created from Lake Mendota samples. Briefly, clones were obtained from -80°C clone libraries. Clones were re-grown in LB with 50 ug/ml kanamycin for 16 hours at 37°C. Plasmid was extracted from the liquid cultures using a QIAGEN plasmid mini kit (QIAGEN, Valencia, CA). The plasmid was purified with a QIAquick PCR purification kit (QIAGEN), and mass concentration was determined by nanodrop. Standards were made via 16S PCR amplification of the plasmids using Promega goTaq hot start mix (25 ul) with M13 primers (M13f (20 uM; 0.1 ug/ul): 1 ul; M13r (20 uM; 0.1 ug/ul): 1ul); Template (plasmid prep): 1 ul. PCR cycle: $94°C$ - 2 min, 25 cycles of ($94°C$ – 35s, $55°C$ – 45 s, $72°C$ – 2 min) $72°C$ – 5 min, $4°C$ hold. Mass concentration of the standards was determined using nanodrop. Diluted with TE buffer to obtain 10^8 copies/ul standards.

Copy number was calculated based on the mass concentration and the average molecular weight of the plasmid. For all unknown samples, 1 ul of community-derived genomic DNA was added as the template in qPCR reactions. In each assay, a no-template control was included to check for contamination and primer-dimer formation. To avoid nonspecific cross-detection, a negative control (a plasmid containing a nontarget 16S fragment with the fewest mismatches to the primer set) was also included and applied at 108 copies per reaction.

4.3.7. iTag Cyanobacteria Normalization

It should be noted that both the cyanobacteria and acI iTag data are relative abundance data, which makes correlations difficult. A cyanobacterial bloom would naturally

reduce the relative abundance of acI in the dataset even if the absolute abundance of acI

remained the same or increased somewhat. This is due to the fact that we have 12,904 reads

per sample date, therefore, more cyanobacteria reads means less reads of other bacteria. In

order to correct for this effect, the bacterial iTag data were normalized to estimate

Actinobacteria read abundance if Cyanobacteria reads were not captured in the dataset. The

normalization equation is as follows: bacterial OTU raw reads (#) + (OTU raw reads/total

non-cyanobacterial reads) * cyanobacterial reads. This normalization generally increased

correlation coefficients between acI and cyanobacteria by 0.01 to 0.1 as compared to non-

normalized data.

4.3.8. iTag Cyanobacteria Taxonomy

It should be noted that the taxonomic assignments for the 6 cyanobacteria are only

accurate to the class or family level. These classifications are difficult due to the similarities

in the V4 region of Cyanobacterial 16S genes across different genera. However, the

Nostocophycideae OTU (#42) is likely *Aphanizomenon* (or possibly *Anabaena*) based on

blast results and LTER phytoplankton data. Also, the LTER data shows that the

Synechococcaceae OTUs (#122, 424, 392, 403) are likely *Aphanocapsa* or *Aphanothece*,

while the Oscillatoriales/Phormidiaceae OTU (#242) is likely *Planktothrix*.

4.4. RESULTS AND DISCUSSION

4.4.1. acI Abundance

Actinobacteria is the dominant phylum (38% of bacteria) in the 10-year time series of

iTag data, followed by beta-proteobacteria (17%), bacteroidetes (16%) and alpha-

proteobacteria (8%) (Figure 3). AcI is the dominant lineage within the Actinobacteria (58%

of Actinobacteria) with acI-B1 and acI-A6 as the most dominant tribes (36% and 33% of acI,

respectively). These results are consistent with a meta-analysis performed by Newton et al

(2011) of 47 lakes that indicated Actinobacteria are frequently the dominant phylum in lakes

with acI-B1 and acI-A6 as the most abundant tribes. One notable difference is that acI-A1 is

not abundant in Lake Mendota (<1% of bacteria), while it tends to be one of the most

abundant tribes in other lakes (~6% of bacteria).

4.4.2. AcI ABUNDANCE OCCASIONALLY VERY LOW IN EARLY SPRING AND RISES FOLLOWING DIATOM BLOOM

AcI is consistently 20-25% of the bacterial population throughout most of the year.

However, March appears to be a unique time period when the consistently dominant acI

lineage decreases significantly in relative abundance while other Actinobacteria increase

(Figure 4). Figure 5 shows that total Actinobacteria reach their peak in spring due to a rise in

acIV and acTH lineages even though the acI lineage is at its low point. The acSTL and Luna

lineages peak in late spring/early summer. The decrease in acI during early spring may be

due to lake mixing which could bring hypolimnetic bacteria into the epilimnion and also

entrain particles and nutrients. It should be noted that we have only 2 sample points in March

due to difficulty in sampling during this ice off period. More sampling is required to confirm

our findings.

Salcher et al (2010) observed peaks in Actinobacteria relative abundance in spring

and fall, which coincided with grazing of a Cryptomonas bloom (spring) and a fall diatom

bloom. Unfortunately this study was limited to a phylum-level analysis. Our study also

observed spring and fall peaks in Actinobacteria lagging a diatom bloom (spring) and

cyanobacteria bloom (late summer) (Figure 4, Figure 5, Figure 6). However, the March peak

in Actinobacteria in our study was mostly due to increases in non-acI Actinobacteria (acIV,

acTH, acSTL and Luna), while in April, acI becomes the dominant Actinobacteria. These findings demonstrate that it may be difficult to draw conclusions about acI based on abundance data at the phylum level. QPCR data from the 10-year time series generally corroborates the trends seen in the iTag data (Figure 6).

Data from 3 years (2000, 2007 and 2009) indicate that the acI relative abundance is occasionally very low in late winter/early spring and rises after the spring diatom bloom (Figure 7). In 2000, the combined acI (acI-B1 + acI-A6) population increased from 0.004% (3/30/00) to 17% (4/27/00) following a *Stephanodiscus* bloom in April. In 2007, the combined acI (acI-B1 + acI-A6) population increased from 0.5% (11/17/06) to 2.3% (5/11/07) following a *Stephanodiscus* bloom in April. In 2009, the combined acI (acI-B1 + acI-A6) population increased from 0.5% (4/22/09) to 5.3% (6/9/09) following a *Stephanodiscus* bloom in May. This seasonal pattern is similar to other studies that have observed acI absolute abundance increase associated with spring phytoplankton blooms (Salcher, 2010; Zeder, 2009; Allgaier, 2006; Parveen, 2011; Rosel, 2012; Zakharova, 2013; Eckert, 2012a; Eiler, 2006).

4.4.3. LACK OF SPRING DIATOM BLOOM CORRESPONDS TO DRAMATIC YEAR-LONG REDUCTION IN ACI

Figure 8 (iTAG) and Figure 9 (qPCR) show that acI is consistently the most abundant bacteria (as a fraction of bacterial 16S) throughout the decade but declines substantially in 2006. Although acI is at its minimum in 2006, Figure 8 shows total Actinobacteria reaching its maximum due to an increase in the acIV, acTH, acSTL and Luna lineages. This data presents a similar picture to that seen in the seasonal data, which showed total Actinobacteria reaching a maximum in March when acI was at its minimum due to the abundance of other

Actinobacteria lineages. This data again demonstrates the difficulty in using broad FISH

probes that target entire phyla, in lieu of more phylogenetically resolved methods such as

qPCR or iTag. If we had used only phylum specific primers, we would have seen little

change in Actinobacteria over the time series even though there is a drastic shift in lineages

in 2006.

Figure 10 shows the annual average relative abundance of Actinobacteria obtained

from iTag data along with long-term data for diatoms and cyanobacteria. This figure shows

that the decline in acI corresponds to a minimum in the spring diatom bloom and also a low

year for cyanobacteria. In fact, the *Stephanodiscus parvus* diatom bloom which is the most

abundant diatom in Lake Mendota and which occurs very predictably every April was

completely absent in 2006.

The 2006 data show small blooms of other diatoms (*Cyclotella* and *Aulacoseira*), but

total diatoms were five times less than in 2009, the second smallest diatom bloom in the

dataset. The 2006 decrease in acI corresponds with the following observations from the

LTER, climate, iTag, and qPCR datasets:

- Lack of a spring diatom (*Stephanodiscus parvus*) bloom in 2006. Diatoms returned in 2007.
- Total Actinobacteria reached 10 year high due to non-acI Actinobacteria lineages (acIV, acTH, acSTl, Luna)
- Copepod nauplii (zooplankton) reached its highest 10 year abundance in May 2006
- Green algae blooms (*Monomastix astigmata* 3/6/06-4/4/06; several species on 4/14/06: *Rhodomonas minuta, Mallomonas akrokomas, Cryptomonas erosa, Sphaerocystis schroeteri, Chlamydomonas*)
- 2002-2005 was an extended drought with 2006 having normal precipitation and the following years having record precipitation events
- 2005 was the driest year with 24.5" precipitation, however, 2002 had 26" and no significant changes in acI population were observed in 2002 or 2003
- Relatively low pH, DOC, NO_3NO_2, and relatively high ammonia, dissolved reactive phosphorus and dissolved silica (as a yearly average)

4.4.4. AcI CORRELATION WITH PHYTOPLANKTON

4.4.4.3. Diatoms

We found that acI typically reaches its minimum relative abundance just prior to the spring diatom bloom and acI tend to increase in relative abundance after spring diatom blooms, corresponding with a bloom of zooplankton grazers (Figure 4).

Figure 11, Figure 12, and Figure 13 show correlations between acI, cyanobacteria and diatoms. We found a significant positive correlation between acI-A6 and diatoms. This correlation increases with greater lag time with the highest correlation at 4 weeks lag time (R=0.4, p=0.00), indicating acI-A6 increases weeks after diatom blooms. AcI-B1 had a negative correlation with diatoms but a significant positive correlation with Nostocophycideae, a class of cyanobacteria. This correlation was strongest at 28 days lag (R=0.21, p=0.04), again indicating that acI-B1 increases weeks after the Nostocophycideae bloom. The Nostocophycideae OTU may represent *Aphanizomenon* as this was the closest blast hit and acI-B1 also had a significant positive correlation with *Aphanizomenon* in the LTER dataset, which was strongest at 14 days lag (R=0.25, p =0.02). This data can be found in supplementary results that display the full correlation results between the acI lineages and phytoplankton and zooplankton.

4.4.4.4. Cyanobacteria

Figure 12 is a network analysis graphic that shows correlations between the most abundant bacteria in the iTag data and the most abundant cyanobacteria, diatoms and zooplankton in the LTER data.

AcI-C2 is the only acI tribe to have a significant strong direct correlation with cyanobacteria. AcI-C2 shows a strong correlation with Synechococcophycieae but no

correlation with the Nostocophycideae OTU. The acI-A tribes have strong negative correlations with Synechococcophycideae and the Oscillatoriophycideae OTUs, but a weakly positive correlation with the Nostocophycideae OTU. Three other freshwater Actinobacteria tribes, Luna-1, acIV-lamia and acSTL-A1, are also negatively correlated with cyanobacteria while acTH1-A1 has a slight positive correlation with two of the Synechococcophycideae OTUs. If the correlation is shifted to look for a lag correlation (4 week lag) (Figure 11) we see a positive correlation between acI-B1, acI-A6 and the Nostocophycideae OTU (r = 0.21 (acI-B1), 0.03 (acI-A6)) and Oscillatoriophycideae OTUs (r = 0.12 (acI-B1)), however the correlation with the Synechococcophycideae remains strongly negative. The data also show that the Synechococcophycideae OTUs correlate strongly with each other, indicating that they may be simply different genotypes within a single species or they may have some other correlative relationship.

4.4.5. AcI LAG CORRELATION WITH PHYTOPLANKTON INDICATES SAPROPHYTIC LIFESTYLE

We found that acI-A6 has a lag correlation with diatom blooms whereas acI-B1 has a lag correlation with Nostocophycideae (cyanobacteria) (Figure 11). Figure 11 indicates that acI correlates most strongly with phytoplankton four weeks after phytoplankton blooms rather than during blooms. This evidence may suggest that acI consumes intracellular phytoplankton material released due to grazing or lysis of phytoplankton during bloom decay rather than consuming algal exudate from active phytoplankton blooms.

Several studies have linked acI to phytoplankton blooms (Salcher, 2010; Zeder, 2009; Allgaier, 2006; Parveen, 2011; Rosel, 2012; Eiler, 2006; Zakharova, 2013; Eckert, 2012a). Parveen et al (2011) found that acI comprised a large fraction of both the free-living and

particle-attached bacteria associated with phytoplankton blooms, similar to Allgaier et al (2007), suggesting that acI may play an active role in the hydrolysis and degradation of phytoplankton. However, other studies (Eiler, 2006; Salomon, 2003) found no acI bacteria attached to cyanobacteria, though they have been found in the free-living fraction during cyanobacterial blooms (Eiler, 2006; Kolmonen, 2004).

Most studies have suggested that acI benefits from exudate produced from living algal blooms. Few studies have investigated the impact of organic matter released from phytoplankton lysis and degradation on acI. Lysis of phytoplankton blooms can be a significant source of organic carbon. Brussard et al (1995) found that 75% of the decline in a marine green algal bloom was due to cell lysis rather than zooplankton grazing. They also found that bacterial production was positively correlated with phytoplankton cell lysis. Van Hannen et al (1999) found that acI-A (known as ACK-M1 at the time) increased following viral lysis of a cyanobacterial bloom. More recent studies have found that acI likely benefit from n-acetyl-glucosamine (NAG) from cell-wall derived organic carbon and also released by hydrolysis of chitin mediated by *Flavobacteria* (Eckert, 2013; Eckert, 2012a; Beier, 2011). The Actinomycetales, the order containing acI, are known as saprophytes (consume nutrients from dead organic matter) and are known for breaking down complex organic polymers (Goodfellow, 1983). Therefore, it may not be surprising that acI appear to proliferate from lysis products following phytoplankton blooms rather than exudate during blooms.

Other studies have also observed a relationship between Actinobacteria and diatoms. Salcher et al (2010) observed a rise in Actinobacteria relative abundance in spring and fall using fluorescent in-situ hybridization (FISH) targeting the phylum. The spring rise occurred during grazing of a *Cryptomonas* (green algae) bloom by heterotrophic nanoflagellates while

the fall maximum occurred during a diatom bloom. The authors hypothesized that the spring maximum was due to acI's grazing resistance and the fall maximum was due to acI's ability to thrive in low phosphorus conditions and consume low molecular weight compounds from the diatom bloom. Eckert et al (2012) also observed a rise in acI shortly after a spring diatom bloom.

4.4.6. AcI GENOMIC EVIDENCE LINKS acI TO PHYTOPLANKTON

Interestingly, our prior work showed that acI genomes are enriched in genes for consuming nitrogen rich compounds (i.e. polyamines, oligo- and di-peptides, amino acids, cyanophycin) (Ghylin, 2014) that may be related to diatoms and cyanobacteria.

4.4.7. POLYAMINES AND POLYPEPTIDES IN DIATOM CELL-WALL MAY BE SUBSTRATES FOR acI

Several studies have shown that diatoms use long-chain polyamines and polypeptides known as silaffins to structure their silica cell walls (Kroger, 2000; Sumper, 2004; Sumper, 2005; Bridoux, 2012). Bidle et al (2014) found that the diatom silica organic matrix was efficiently dissolved and recycled by bacterial activity, primarily ectoprotease from *Flavobacteria*. Polyamines are also associated with all living matter and are present in heterotrophic bacteria, cyanobacteria and green algae as well (Nishibori, 2003; Nishibori, 2004). Based on these results as well as our prior work showing that acI genomes are enriched in genes for consuming nitrogen rich compounds (i.e. polyamines, oligo- and di-peptides, amino acids, cyanophycin) (Ghylin, 2014), we hypothesize that decaying diatoms provide a significant source of energy, carbon and nitrogen for the acI bacteria in the form of organic nitrogen compounds such as long chain polyamines and polypeptides.

4.4.8. CYANOPHYCIN AND CYANOBACTERIAL CELL WALLS MAY BE SUBSTRATES FOR ACI

Our previous genome analysis (Ghylin, 2014) revealed genes in acI for the breakdown of cyanophycin (storage molecule typically found in Cyanobacteria) and uptake of lipopolysaccharides (primary cell wall component of gram-negative bacteria such as Cyanobacteria). The genomes also included genes such as lysozyme (COG 3757) and glycerophosphoryl diester phosphodiesterase (COG 0584) that can degrade cell membrane materials. Additionally, Eckert et al (2012a) found that acI was very active in uptake of N-acetylglucosamine, a byproduct of cell-wall degradation. Therefore, we hypothesize that decaying Cyanobacteria (and potentially heterotrophic bacteria) provide a significant source of energy, carbon and nitrogen for the acI bacteria in the form of organic compounds such as cyanophycin, lipopolysaccharides and other cell wall and membrane components.

AcI correlation with green algae

AcI-A6 has a significant positive correlation with the green algae, Cryptomonas, with the strongest correlation at 28 days lag (see data in supplementary results). This is the only significant positive correlation between any acI tribe and any green algae in Lake Mendota. This finding is similar to that of Salcher et al (2010) who found (using FISH) that Actinobacteria abundance rose during grazing of a Cryptomonas bloom in spring.

AcI correlation with zooplankton

AcI has significant negative correlations with Daphnia and Copepod and no correlation with Diaptomid. Interestingly, the non-acI Actinobacteria (acTH1-A1, Luna1-A1, acSTL-A1, and acIV-D) have strong positive correlations with Daphnia and Copepod but not Diaptomid (see supplementary data). These correlations are evident in 2006 when acI was at its 10-year minimum and non-acI Actinobacteria, *Daphnia* and *Copepod* were at 10-year maximums.

The acI data is based on iTag data while the zooplankton data were obtained from LTER counts.

4.4.9. CORRELATIONS BETWEEN ACI AND OTHER BACTERIA

Figure 14 shows correlations between the top 20 bacteria in the iTag data. This data shows that acI-B1, acI-A4 and acI-A6 correlate positively, while acTH1-A1, Luna1, acIV-lamia, and acSTL-A1 negatively correlate with the acI tribes. AcI-B1 also correlates with bacI-A1, LD12, LD28 and Lhab-A1, which are all planktonic organisms like acI.

AcI-A6 has a few different correlations from acI-B1. AcI-A6 does not correlate with the LD12 (alpha-proteobacteria) and LD28 (beta-proteobacteria) but does correlate with PnecC (*Polynucleobacter*) and betI-A and has a neutral correlation with betIII-A, while acI-B1 has a strong negative correlation with this tribe.

Studies have found that PnecC (*Polynucleobacter*) is readily enriched upon exposure to elevated levels of DOC. BetI-A is known as a fast growing organism that responds to nutrient pulses and is associated with low molecular weight algal exudates and is susceptible to grazing (Newton, 2011). The correlation between acI-A6 and PnecC and BetI-A as well as its rapid growth after the spring diatom bloom (Figure 5) may provide further evidence that acI-A6 is linked to diatoms.

4.4.10. TOP-DOWN VERSUS BOTTOM-UP FACTORS CONTRIBUTING TO ACI'S SUCCESS

Several studies have demonstrated grazing resistance in acI and have focused on this characteristic as a primary factor in its success (Jezbera, 2006; Jezbera, 2005; Tarao, 2009). Other studies have indicated that acI may benefit from organic carbon released from lysed

cells during grazing events, indicating that acI benefit not only from grazing resistance but also from an ability to compete for organic matter released during grazing events.

Eckert et al (2013) suggested top-down (grazing resistance) and bottom-up factors (physiological capability of consuming organic substances released during grazing/viral lysis of phytoplankton) as being both important for acI's success. They focused on n-acteyl glucosamine (NAG) released from chitin and peptidoglycan. Salcher et al (2010) also suggested both top-down and bottom-up factors contribute to acI's success as they observed acI abundance increase during grazing in spring and also following a diatom bloom in fall.

Although grazing resistance is an important feature of acI, our results, indicating lag correlations (Figure 11) between acI and phytoplankton add more evidence of acI's relationship with phytoplankton as an important bottom-up driver for its success.

4.4.1. LIMITATIONS OF THIS STUDY

This study utilized qPCR and 16s rRNA gene iTag sequencing to determine relative abundances of freshwater bacteria and to correlate changes in relative abundance with environmental drivers. Due to the nature of relative abundance data, it is possible that a bacterial population can have no change in absolute abundance over time while the relative abundance increases or decreases due to proliferation or death of other bacteria. It is possible that the apparent correlation between acI-A6 and diatoms is due simply to a decrease in other bacteria after diatom blooms while acI-A6 absolute abundance remains stable.

Additionally, this study employed tribe-level assignments of iTag OTUs. Subsequent analysis suggest that the V4 region of the 16S rRNA gene does not provide sufficient resolution to reliably distinguish among tribes within a clade, for some freshwater lineages. It is therefore possible that the identifications at the tribe levels (i.e. acI-B1, acI-A6) were

inaccurate and thus conclusions about specific tribes should be taken with caution. The identification at the clade level (i.e. acI-B vs acI-A) is generally reliable and the conclusions made for acI-B1 and acI-A6 likely hold for acI-B and acI-A since acI-B1 and acI-A6 are the most dominant tribes in their respective clades.

The results of this study, namely the correlations between acI and phytoplankton should be confirmed with absolute abundance data using techniques such as fluorescent in situ hybridization (FISH).

4.5. CONCLUSIONS

This study expands on previous work and represents the longest and most phylogenetically resolved time series analysis of acI population dynamics ever published. This data confirms previous findings that acI-B1 and acI-A6 dominate Lake Mendota. The data also show that acI-A6 rises more quickly than acI-B1 in the spring, following the annual diatom bloom. The qPCR data agreed well with the iTag data, which lends credibility to the use of the newer iTag sequencing methods to obtain comprehensive bacterial community data.

Previous studies have focused on the grazing resistance of acI as its relative abundance typically peaks during grazing events (Tarao, 2009; Eckert, 2013; Jezbera, 2006). However, the apparent link between acI and decaying phytoplankton (diatoms and cyanobacteria) indicates that acI likely benefits from grazing events due to the release of specific compounds from phytoplankton such as polyamines and cyanophycin. Thus, acI's resistance to grazing is probably not the only factor for its success during these events.

This study found a significant positive correlation between acI-A6 and diatoms and a correlation between acI-B1 and *Aphanizomenon,* indicating that different acI tribes may feed on decay products of different phytoplankton. Also, our data show that the strongest correlations between acI and phytoplankton occur four weeks after phytoplankton blooms, indicating that acI relative abundance increases after these blooms and may be feeding on bloom decay products rather than exudate from active blooms.

This study revealed a rare regime shift in 2006 from an acI dominated lake to a lake dominated by non-acI Actinobacteria which corresponded with the absence of the usually predictable diatom bloom (*Stephanodiscus parvus*) in April. This shift also corresponded with the end of a 4-year drought.

This study found that acI-C2 has a strongly positive correlation with cyanobacteria with no lag, indicating it is associated with active cyanobacteria blooms, unlike all other acI tribes. The data indicate a possible association between acI and decaying cyanobacteria as acI-B1 tends to increase after *Aphanizomenon* blooms. Additionally, our previous study of acI genomes found a unique cyanophycinase gene that may allow acI to break down a storage polymer (cyanophycin) produced in cyanobacteria. Additionally, the genomic analysis found several genes for the degradation and uptake of cell wall and membrane components of gram-negative bacteria such as cyanobacteria.

The data presented here seem to indicate a relationship between acI (especially acI-A6) and decaying diatoms, with acI increasing in abundance after the annual spring diatom bloom. Also, the reduced relative abundance of the acI population in 2006 coincided with an absence of a spring diatom bloom. Based on previous studies (Garcia, 2012; Sumper, 2004), it seems that the acI bacteria may gain a significant amount of their energy, carbon and

nitrogen requirements by consuming polypeptides and polyamines released from decaying diatoms. A recent study was able to demonstrate enrichment of acI-B2 in the presence of a polyamine (putrescine) (Garcia, 2013), a likely decay product of diatoms.

Previous studies have shown that diatoms contain long chain polyamines and polypeptides that help structure their unique crystalline cell walls. Our previous studies of acI genomes indicated a unique abundance of genes involved in the uptake and metabolism of polyamines, oligo- and di-peptides and other nitrogenous organic compounds. Based on these findings, we propose that diatoms are an important source of nitrogenous organic compounds that acI are uniquely able to consume.

This study also demonstrates the importance of phylogenetic resolution as acI tribes had varying seasonal patterns and the acI as a lineage had a fairly steady population while total Actinobacteria peaked in spring and fall due to non-acI Actinobacteria populations.

The data presented in this study validate genomic data from our previous study and indicate that the acI lifestyle is similar to that of other Actinomycetes in that it appears to rely on decaying biomass (possibly released during grazing or lysis of phytoplankton) for its energy and nutrient needs. This data demonstrates that acI likely relies primarily on autochthonous (phytoplankton-derived) carbon rather than allochthonous (terrestrial-derived) carbon.

4.6. ACKNOWLEDGEMENTS

4.7. FIGURES

Figure 1. Sample Distribution (qPCR)

Sample Distribution

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	x	xx			x	x	xx	x	x		x	x
2009		x		xx		xxx	xx	xx	x	xx	x	x
2008				xx		xx	xx	xx	xx	xx		
2007				xx		xx	xxxx	xx	x	x	x	
2006			x		x	x	xx	xx	x	xx	xx	
2005			x		xx	x	x	xx		x	x	
2004				xx		xx	xx	x	xxx	x		
2003				xx		xx	x	x	xx	x	x	
2002		x		xx	xx	xx	xx	xx	xx	xx	xx	
2001			x	xx	xx	xxx	xxx	xx	xx	xx	xx	
2000			xx	x	xxx	xx	x	xx	xx	xx	x	

Year (vertical axis, 1999–2011) vs. Month (horizontal axis)

Figure 2. Sample Distribution (iTag)

Sample Distribution

Figure 3. Actinobacteria average abundance in Lake Mendota (2000-2010) based on Illumina 16S Tag Data
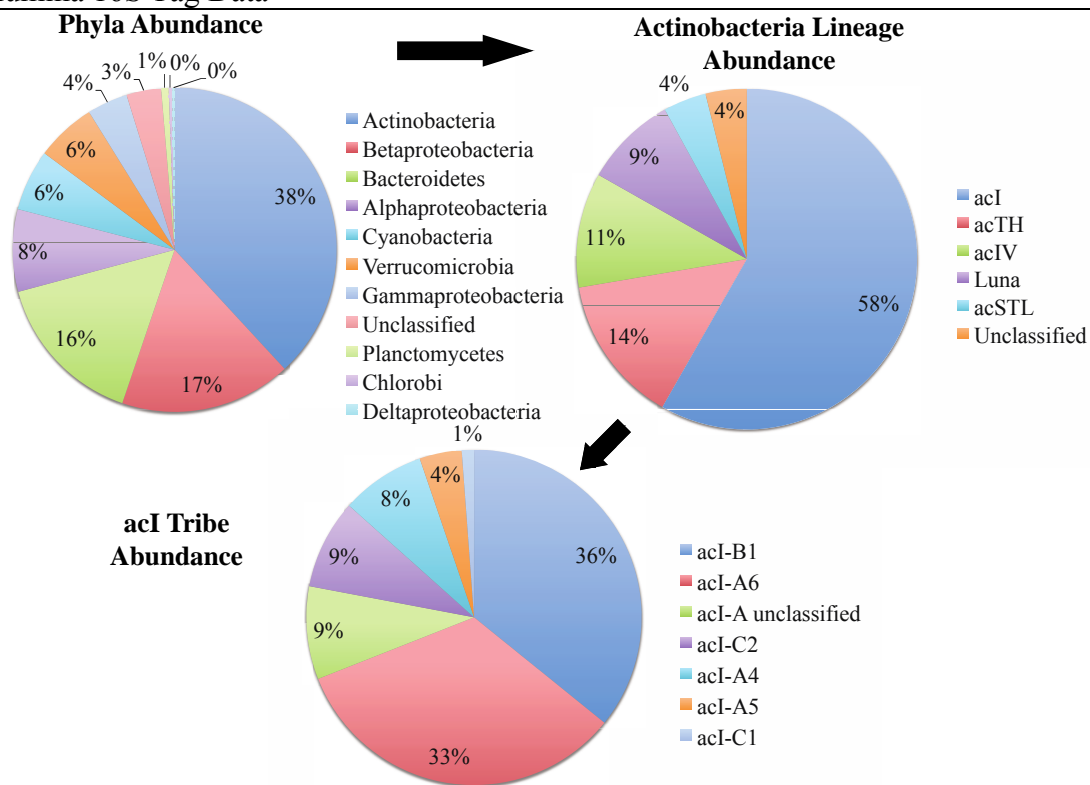
Figure 4. Average Seasonal Abundance of Actinobacteria, Cyanobacteria, Diatoms and Zooplankton in Lake Mendota (2000-2010)
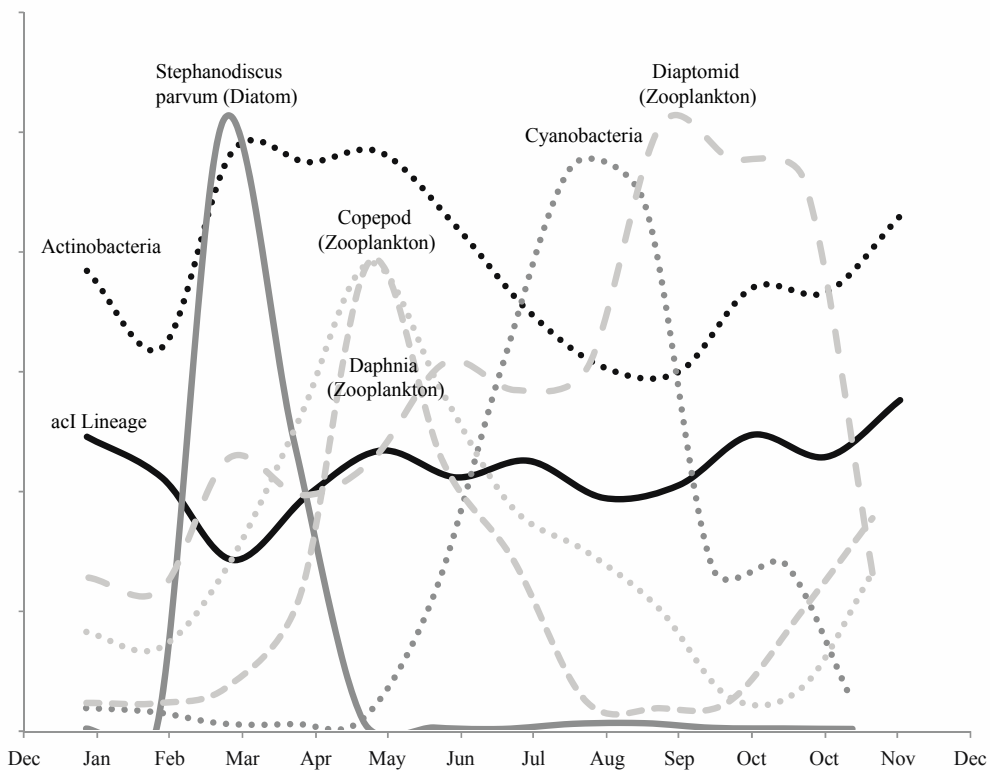
Note: Cyanobacteria, Actinobacteria and acI trends are constructed from relative abundance data from 16S tags. Cyanobacteria represents the sum of the 6 most abundant OTUs, which account for the vast majority of 16S cyanobacteria reads. Diatom and Zooplankton trends are constructed from absolute abundance data from the LTER database. Y-axis labels intentionally removed for simplicity as several plots are super-imposed. Figure intends to display general trends rather than quantitative data.

Figure 5. Actinobacteria Abundance by lineage in Lake Mendota averaged by month (2000-2010) (iTag data)
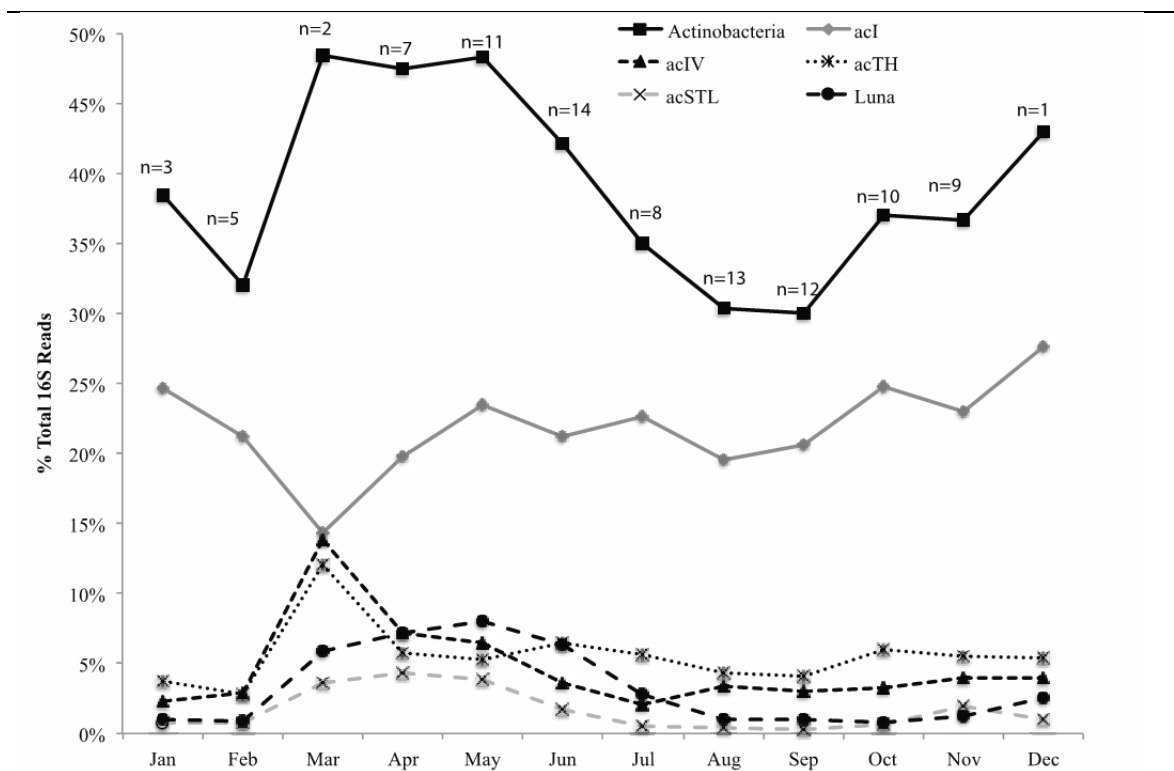
Figure 6. Actinobacteria Abundance by lineage in Lake Mendota averaged by month (2000-2010) (qPCR data)
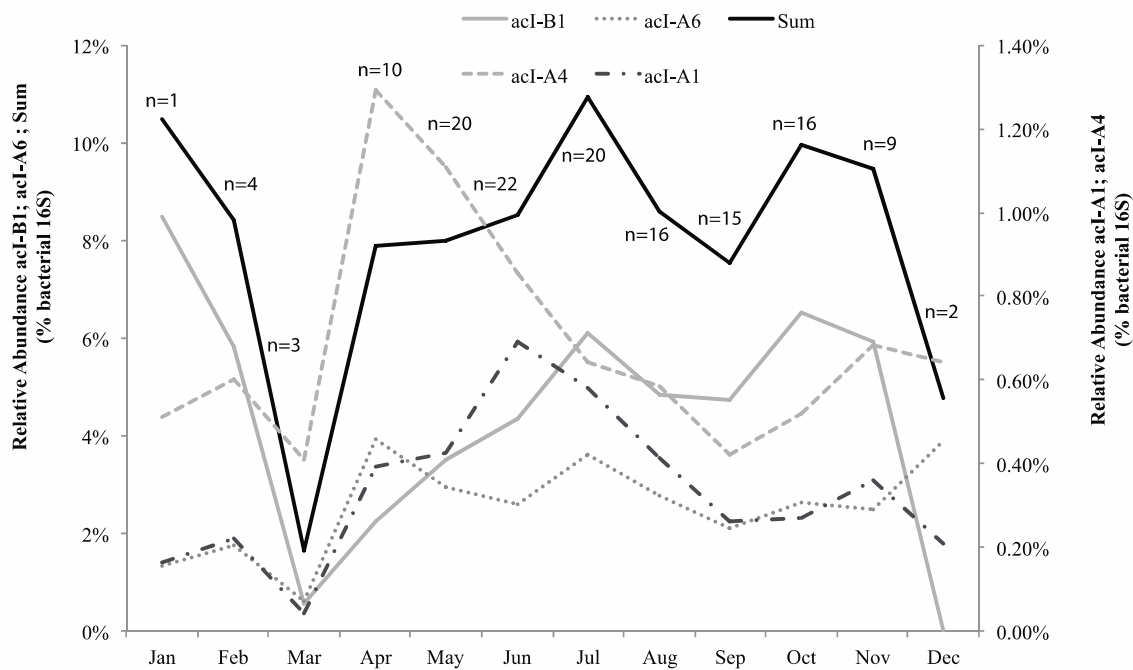
Figure 7. Actinobacteria and Diatom abundance during spring 2000, 2007, 2009 (qPCR data)
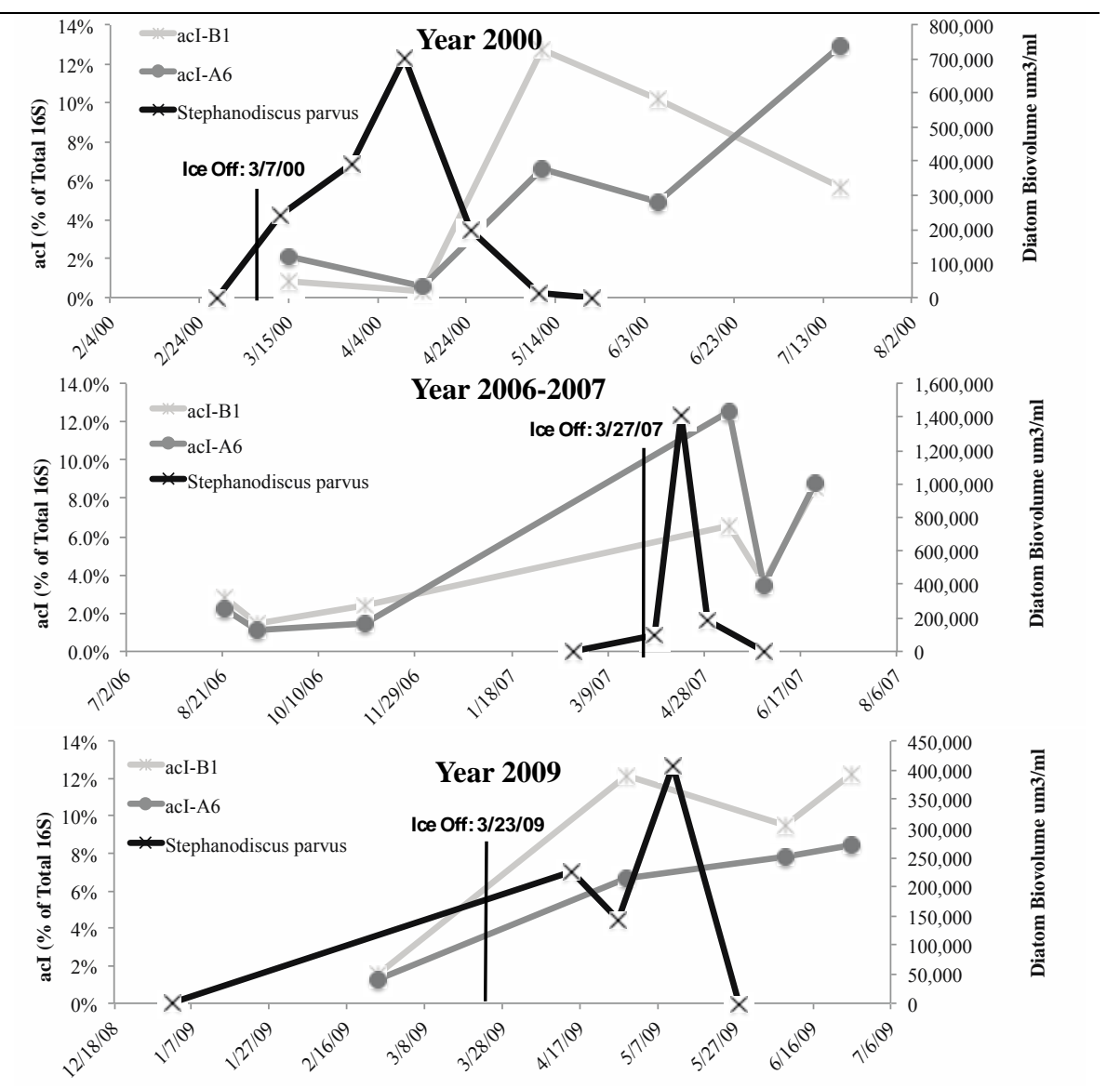
Figure 8. Actinobacteria Lineage Abundance in Lake Mendota averaged by year (iTag data 2000-2010)
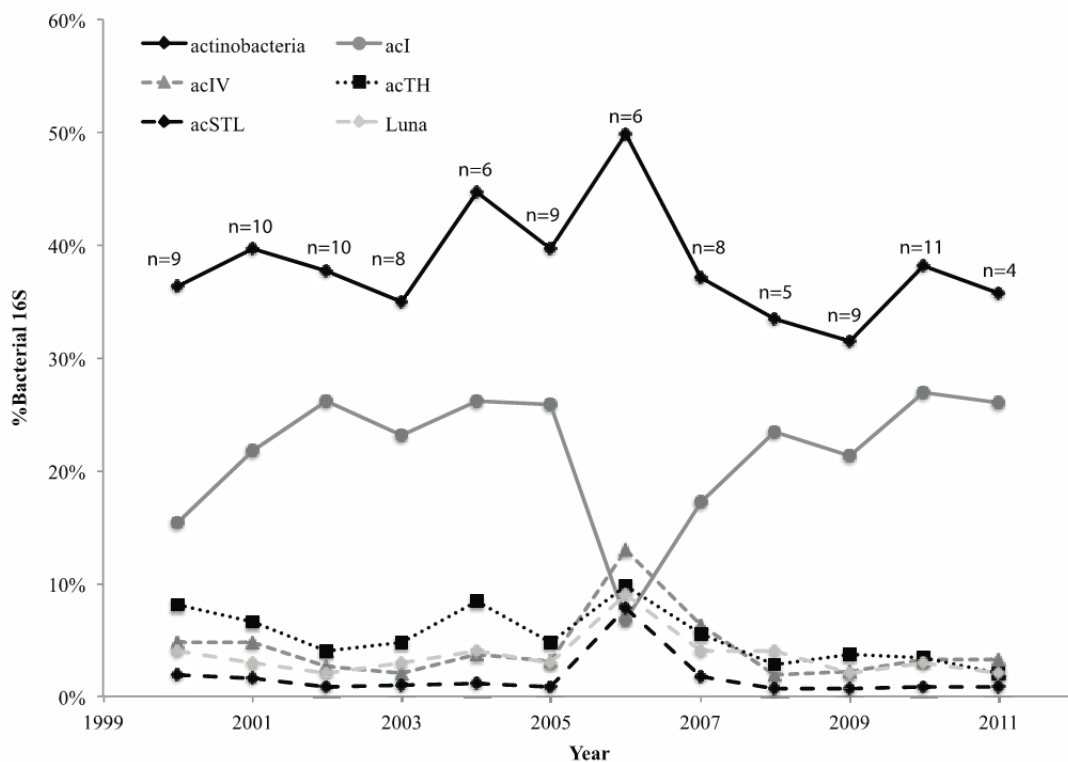
Figure 9. Actinobacteria Abundance by lineage in Lake Mendota averaged by year (2000-2010) (qPCR data)
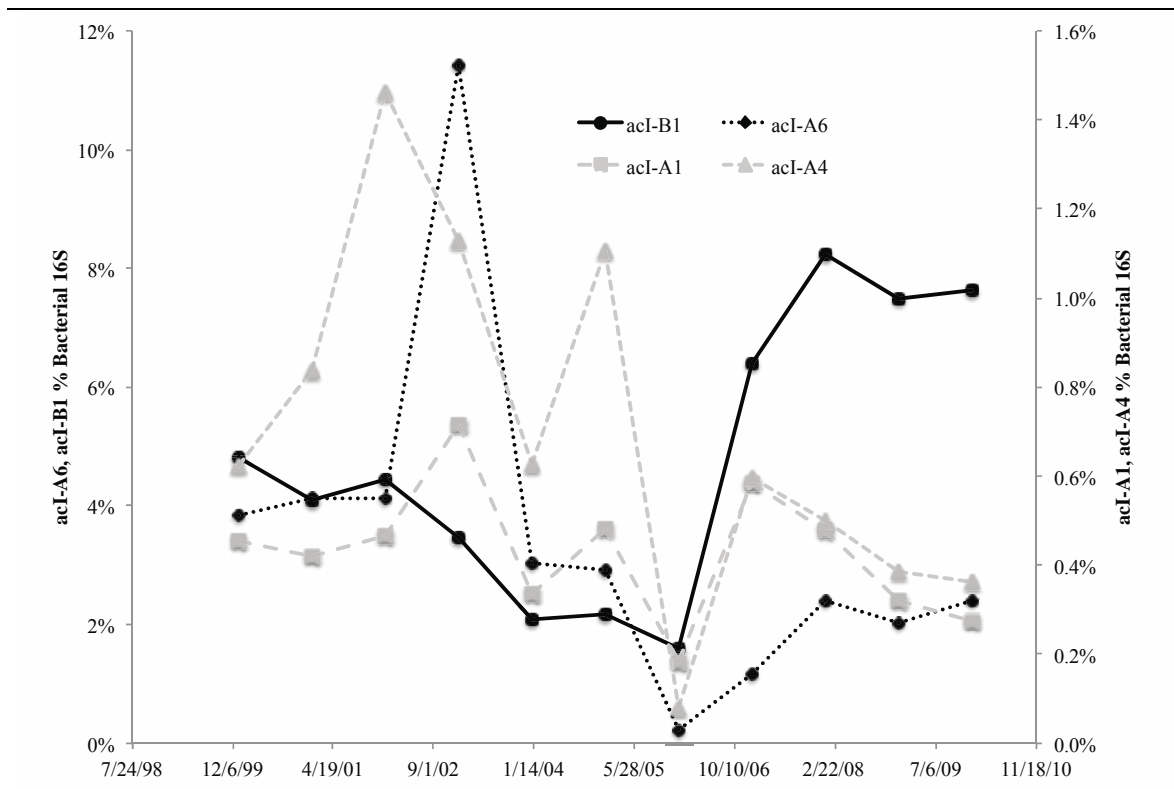
Figure 10. Actinobacteria and Phytoplankton (Diatoms and Cyanobacteria) Annual Abundance in Lake Mendota (16S iTag data and LTER data 2000-2010)
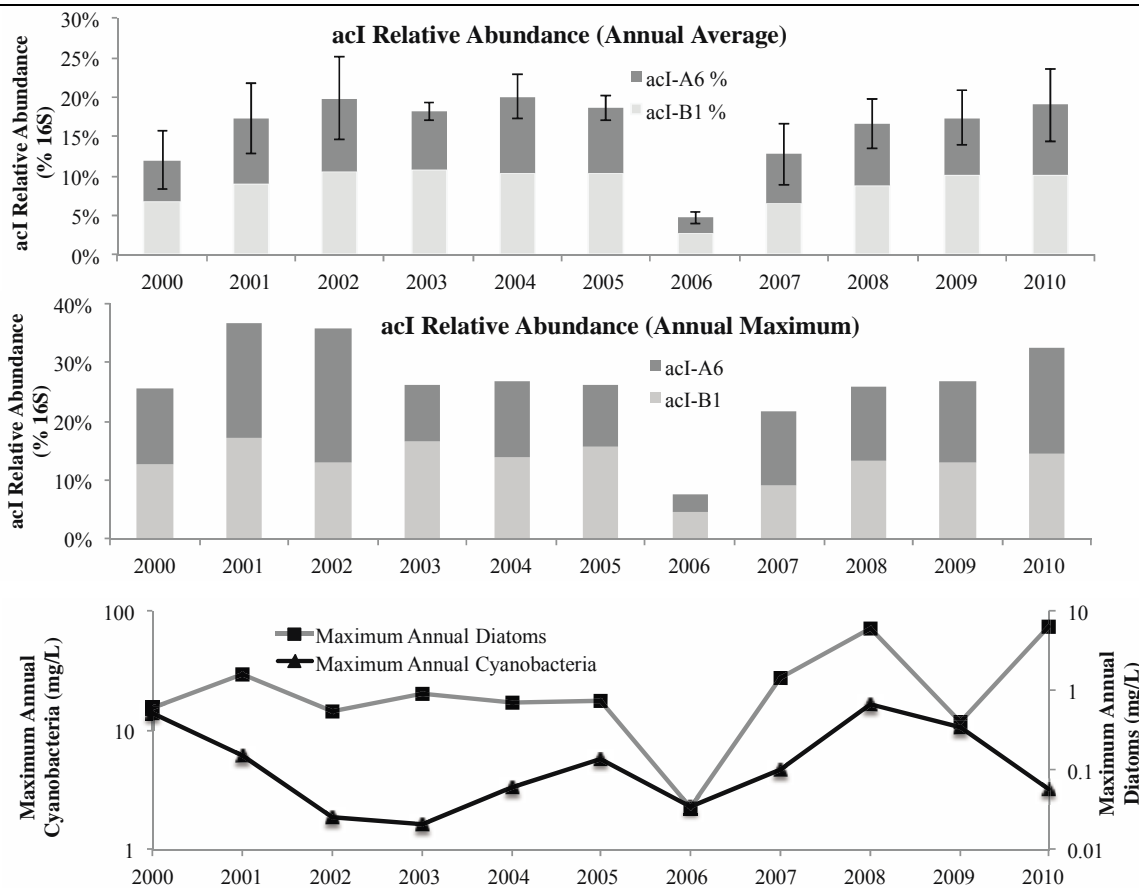
Note: Non-acI Actinobacteria are not shown for simplicity. These lineages include acIV, acTH, acSTL and Luna. They comprise the minority of Actinobacteria on average but represented over 90% of Actinobacteria in 2006. Maximum phytoplankton abundances are shown instead of total annual abundance because the number of sampling dates varies per year. Actinobacteria data are based on iTag while Cyanobacteria and diatoms are based on LTER data.

Figure 11. Actinobacteria correlation with Diatoms and Cyanobacteria (iTag data)
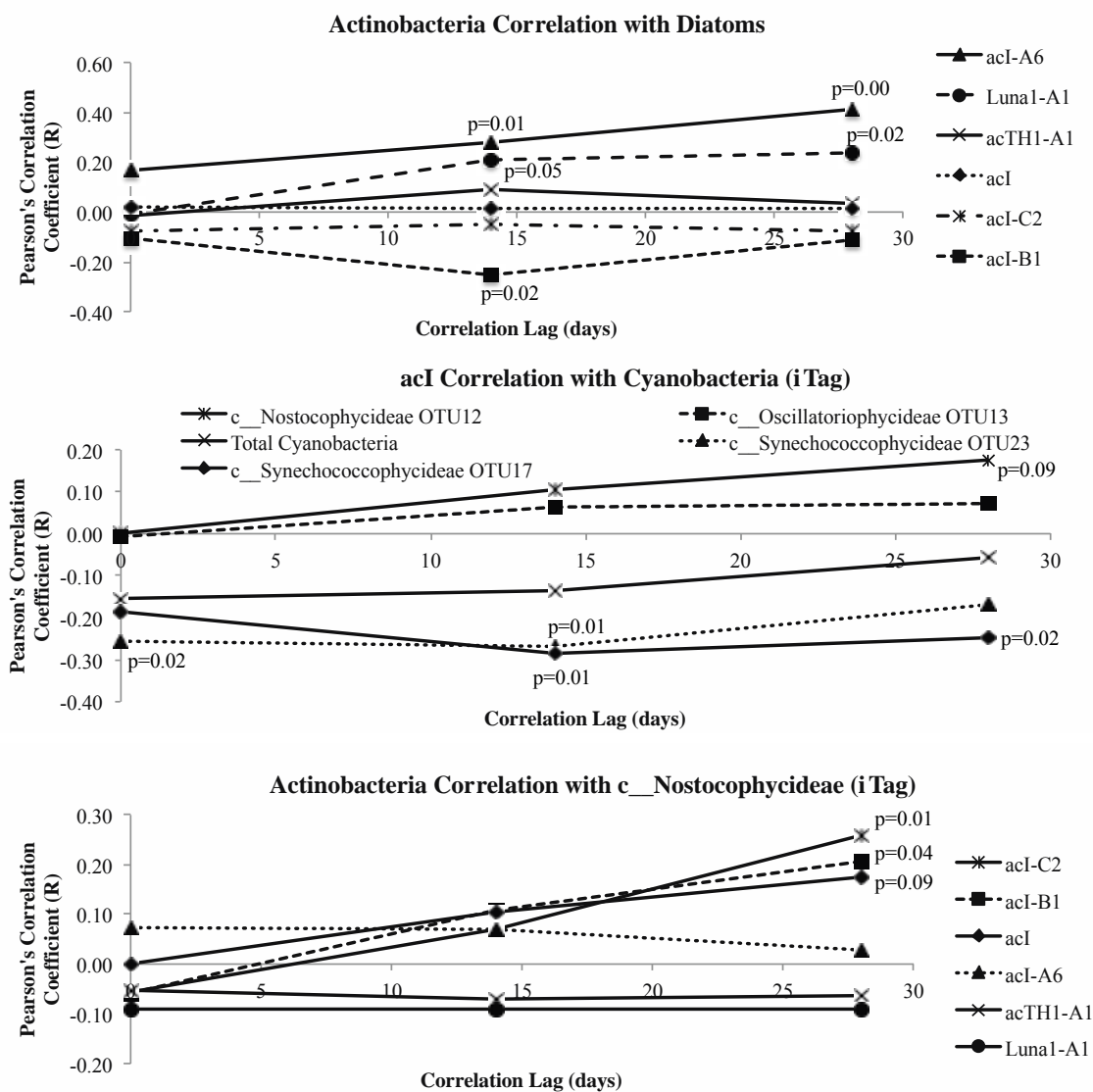
Note: Lag times of 0 days, 14 days and 28 days are shown in figure. Actinobacteria data were shifted 14 days and 28 days to test for a lag correlation with diatoms and cyanobacteria to test the hypothesis that some acI consume decay products released after diatom and cyanobacteria blooms rather than exudate from live blooms.

Figure 12. Correlation Network Analysis: Actinobacteria, Phytoplankton and Zooplankton (LTER and iTag data)
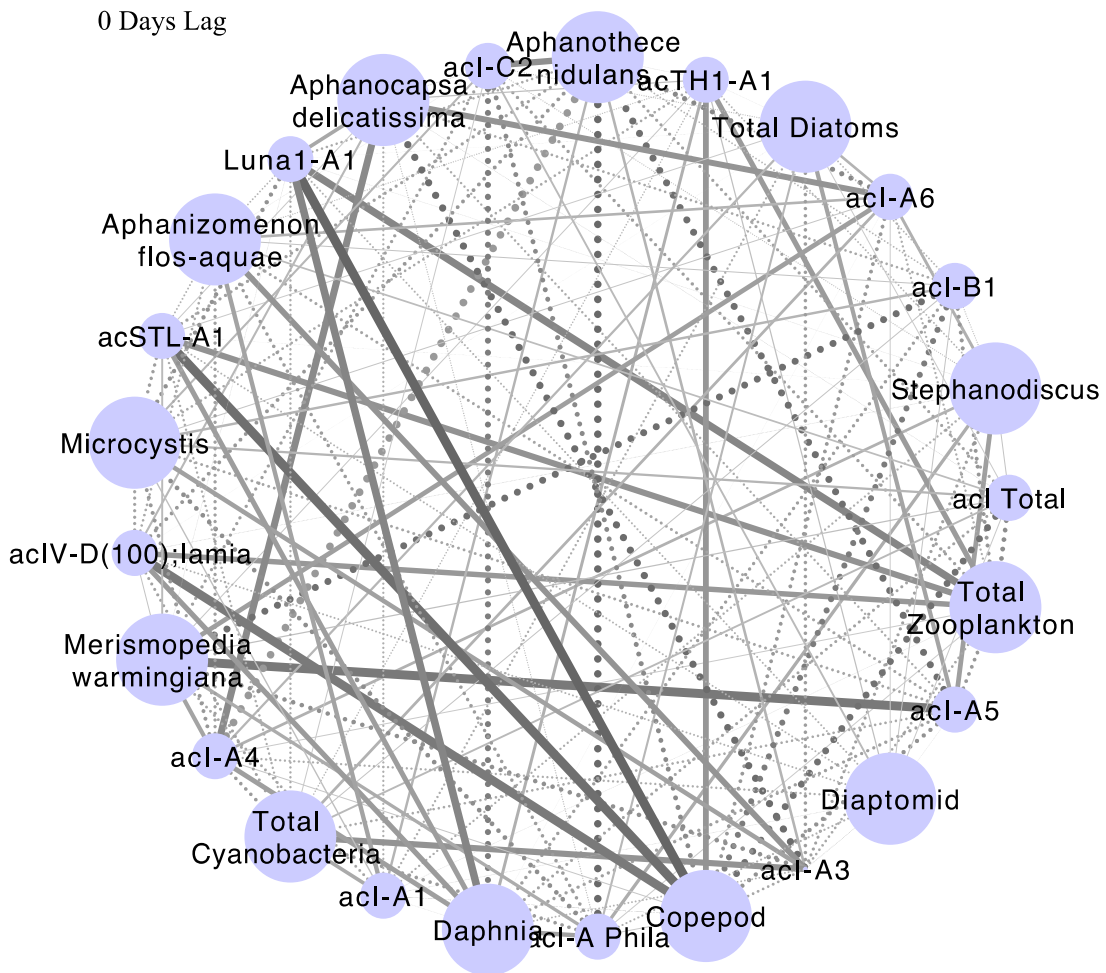
Note: Solid lines indicate positive correlation, dashed lines are negative. Line width indicates strength of correlation.

Figure 13. Correlation Network Analysis: Actinobacteria, Cyanobacteria 14 Day Lag (LTER and iTag data)

Note: Solid lines indicate positive correlation, dashed lines are negative. Line width indicates strength of correlation.
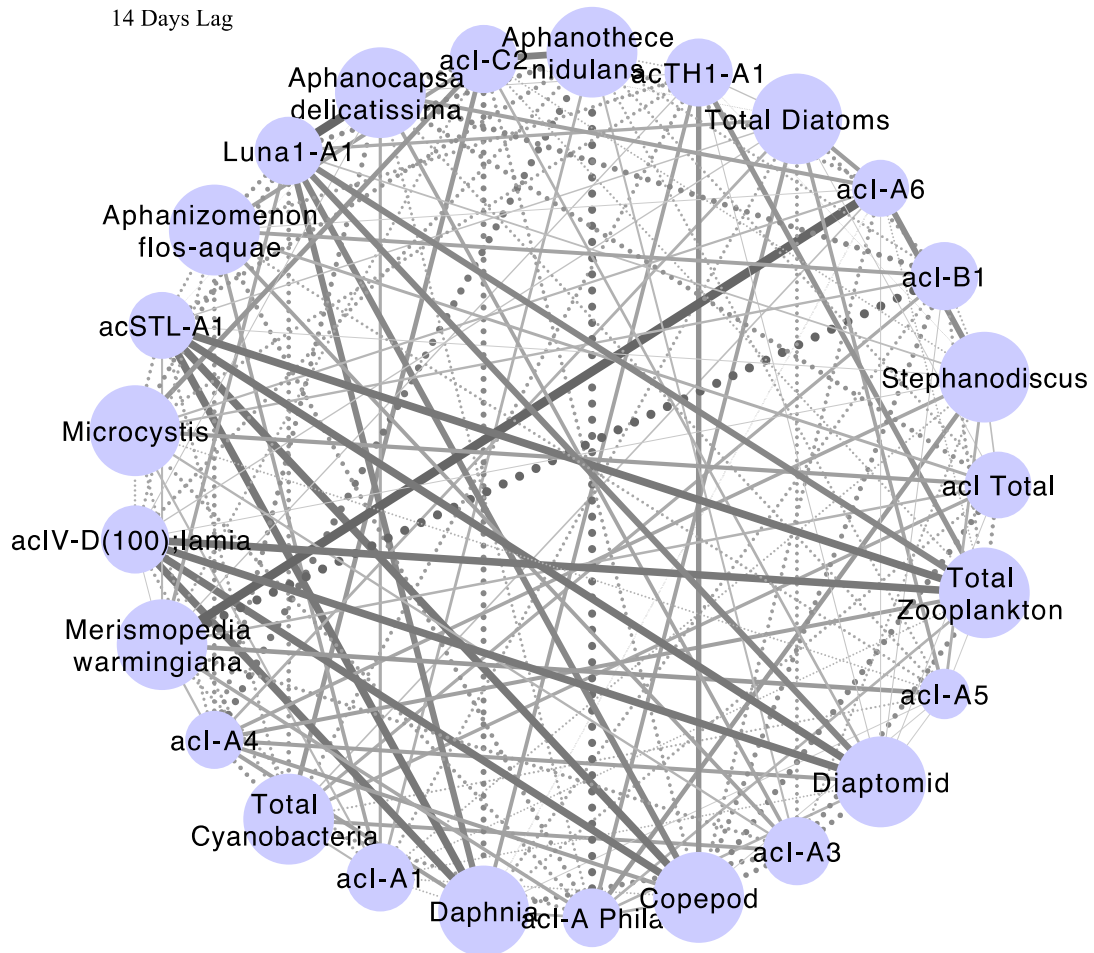
Figure 14. Correlation Network Analysis between Top 10 bacteria and Top 7 Actinobacteria (16S iTag data)
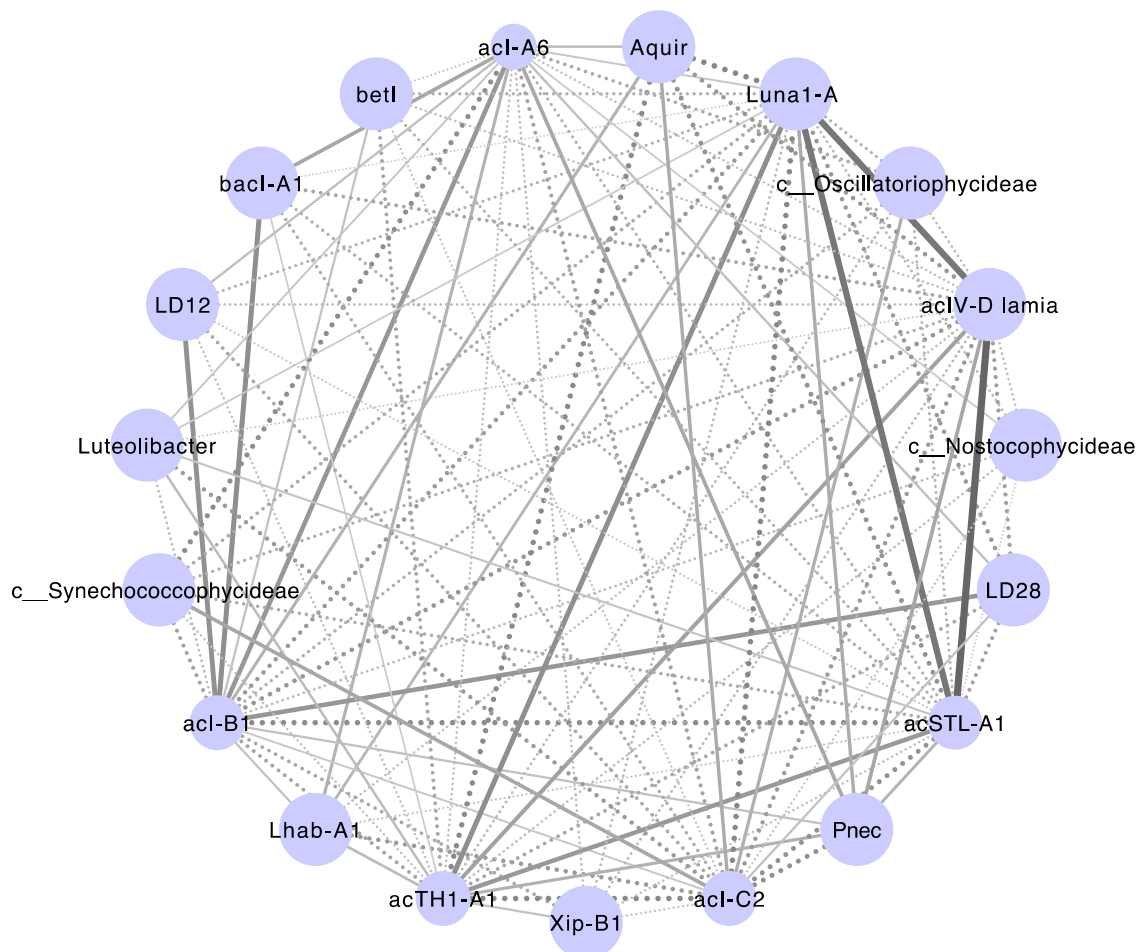
Note: Solid lines indicate positive correlation, dashed lines are negative. Line width indicates strength of correlation.

4.8. TABLES

Table 5. qPCR Details

Primer	Sequence (5'-3')	Target	Amplicon Size (bp)	Annealing Temp(°C)	Negative Control Cluster	Specific Fractioning Limit[1]
acIB1-615f	AGGCTCAACCTCAGGCC	acI-B1	222	61	acI-B4	0.01%
acIB1-836r	CCCACAACTAGTGCCCAC					
acIA1-78f	ACCTTCGGGTGTGAATTAGCG	acI-A1	131	67	acI-A3	1.0%
acIA1-198r	CTTTCCAGCCCCGATCATGC					
acIA4-70f	GAGATCACTTCGGTGAAGGA	acI-A4	109	63.5	acI-A5	1.0%
acIA4-184r	GCTCATATCCGGTATTAGCC					
acIA6-641f	TACGGGCGAGCTTGAGTATGG	acI-A6	215	67	acI-A5	0.1%
acIA6-855r	CGCAGAGACCGTGGAAGGT					
EUB-340f	TCCTACGGGAGGCAGCAG	All Bacteria	~194	51	N/A	n/a
EUB-533r	TTACCGCGGCTGCTGGCAC					

1 He, 2007

4.9. REFERENCES

Allgaier, M., & Grossart, H. (2006). Diversity and Seasonal Dynamics of Actinobacteria Populations in Four Lakes in Northeastern Germany Diversity and Seasonal Dynamics of Actinobacteria Populations in Four Lakes in Northeastern Germany, *72*(5). doi:10.1128/AEM.72.5.3489

Allgaier, M., Brückner, S., Jaspers, E., & Grossart, H.-P. (2007). Intra- and inter-lake variability of free-living and particle-associated Actinobacteria communities. *Environmental microbiology*, *9*(11), 2728–41. doi:10.1111/j.1462-2920.2007.01385.x

Beier, S., & Bertilsson, S. (2011). Uncoupling of chitinase activity and uptake of hydrolysis products in freshwater bacterioplankton. *Limnology and Oceanography*, *56*(4), 1179–1188. doi:10.4319/lo.2011.56.4.1179

Bidle, K. D., & Azam, F. (2014). Bacterial control of silicon regeneration from diatom detritus : Significance of bacterial ectohydrolases and species identity, *46*(7), 1606–1623.

Bridoux, M. C., Keil, R. G., & Ingalls, A. E. (2012). Analysis of natural diatom communities reveals novel insights into the diversity of long chain polyamine (LCPA) structures involved in silica precipitation. *Organic Geochemistry, 47*, 9–21. doi:10.1016/j.orggeochem.2012.02.010

Brussaard, C. P. D., Riegman, R., Noordeloos, A. A. M., Cadee, G. C., Witte, H., Kop, A. J., Nieuwland, G., et al. (1995). Effects of grazing , sedimentation and phytoplankton cell lysis on the structure of a coastal pelagic food web, *123*, 259–271.

Burkert, U., Warnecke, F., Babenzien, D., Pernthaler, J., & Zwirnmann, E. (2003). Members of a Readily Enriched β -Proteobacterial Clade Are Common in Surface Waters of a Humic Lake Members of a Readily Enriched ɴʟ -Proteobacterial Clade Are Common in Surface Waters of a Humic Lake. doi:10.1128/AEM.69.11.6550

Caporaso, J. G., Lauber, C. L., Walters, W. a, Berg-Lyons, D., Huntley, J., Fierer, N., … Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, *6*(8), 1621–4. doi:10.1038/ismej.2012.8

Carpenter, S. R., Lathrop, R. C., Nowak, P. J., Bennett, E. M., Reed, T., & Soranno, P. A. (2006). The ongoing experiment : the restoration of Lake Mendota and its watershed. (B. J. Benson, Ed.)*Long-Term Dynamics of Lakes in the Landscape: Long-Term Ecological Research on North Temperate Lakes*. Oxford University Press.

Eckert, E. M., Salcher, M. M., Posch, T., Eugster, B., & Pernthaler, J. (2012). Rapid successions affect microbial N-acetyl-glucosamine uptake patterns during a lacustrine

spring phytoplankton bloom. *Environmental microbiology*, *14*(3), 794–806. doi:10.1111/j.1462-2920.2011.02639.x

Eckert, E. M., Baumgartner, M., Huber, I. M., & Pernthaler, J. (2013). Grazing resistant freshwater bacteria profit from chitin and cell-wall-derived organic carbon. *Environmental microbiology*, *1*. doi:10.1111/1462-2920.12083

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, *27*(16), 2194–2200. doi:10.1093/bioinformatics/btr381

Eiler, A., Olsson, J. a., & Bertilsson, S. (2006). Diurnal variations in the auto- and heterotrophic activity of cyanobacterial phycospheres (Gloeotrichia echinulata) and the identity of attached bacteria. *Freshwater Biology*, *51*(2), 298–311. doi:10.1111/j.1365-2427.2005.01493.x

Garcia, S. L., McMahon, K. D., Martinez-Garcia, M., Srivastava, A., Sczyrba, A., Stepanauskas, R., Grossart, H.-P., et al. (2012). Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *The ISME journal*, 1–11. doi:10.1038/ismej.2012.86

Garcia, S. L., McMahon, K. D., Grossart, H.-P., & Warnecke, F. (2013). Successful enrichment of the ubiquitous freshwater acI A ctinobacteria. *Environmental Microbiology Reports*, n/a–n/a. doi:10.1111/1758-2229.12104

Glöckner, F. O., Zaichikov, E., Belkova, N., Denissova, L., Pernthaler, J., Pernthaler, A., Amann, R., et al. (2000). Comparative 16S rRNA Analysis of Lake Bacterioplankton Reveals Globally Distributed Phylogenetic Clusters Including an Abundant Group of Actinobacteria Comparative 16S rRNA Analysis of Lake Bacterioplankton Reveals Globally Distributed Phylogenetic Cluste. doi:10.1128/AEM.66.11.5053-5065.2000.Updated

Goodfellow, M; Williams, S. T. (1983). Ecology of actinomycetes. *Annual review of microbiology*, *37*(41), 189–216. doi:10.1146/annurev.mi.37.100183.001201

Hannen, E. J. Van, Zwart, G., Van, M. P., Gons, H. J., Ebert, J., & Laanbroek, J. (1999). Changes in Bacterial and Eukaryotic Community Structure after Mass Lysis of Filamentous Cyanobacteria Associated with Viruses Changes in Bacterial and Eukaryotic Community Structure after Mass Lysis of Filamentous Cyanobacteria Associated with Viruses †.

He, S., Gall, D. L., & McMahon, K. D. (2007). "Candidatus Accumulibacter" population structure in enhanced biological phosphorus removal sludges as revealed by polyphosphate kinase genes. *Applied and environmental microbiology*, *73*(18), 5865–74. doi:10.1128/AEM.01207-07

Jezbera, J., Hornák, K., & Simek, K. (2005). Food selection by bacterivorous protists: insight from the analysis of the food vacuole content by means of fluorescence in situ hybridization. *FEMS microbiology ecology*, *52*(3), 351–63. doi:10.1016/j.femsec.2004.12.001

Jezbera, J., Hornák, K., & Simek, K. (2006). Prey selectivity of bacterivorous protists in different size fractions of reservoir water amended with nutrients. *Environmental microbiology*, *8*(8), 1330–9. doi:10.1111/j.1462-2920.2006.01026.x

Kolmonen, E., Sivonen, K., Rapala, J., & Haukka, K. (2004). Diversity of cyanobacteria and heterotrophic bacteria in cyanobacterial blooms in Lake Joutikas, Finland. *Aquatic Microbial Ecology*, *36*, 201–211. doi:10.3354/ame036201

Kröger, N., Deutzmann, R., Bergsdorf, C., & Sumper, M. (2000). Species-specific polyamines from diatoms control silica morphology. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(26), 14133–8. doi:10.1073/pnas.260496497

Magnuson, J. J. . T. K. K. and B. J. B. (2005). Long term dynamics of lakes in the landscape. *Oxford Univ. Press*.

Newton, R. J., Jones, S. E., Helmus, M. R., & McMahon, K. D. (2007). Phylogenetic ecology of the freshwater Actinobacteria acI lineage. *Applied and environmental microbiology*, *73*(22), 7169–76. doi:10.1128/AEM.00794-07

Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., & Bertilsson, S. (2011). *A guide to the natural history of freshwater lake bacteria. Microbiology and molecular biology reviews : MMBR* (Vol. 75, pp. 14–49). doi:10.1128/MMBR.00028-10

Newton, R. J., & McMahon, K. D. (2011a). Seasonal differences in bacterial community composition following nutrient additions in a eutrophic lake. *Environmental microbiology*, *13*(4), 887–99. doi:10.1111/j.1462-2920.2010.02387.x

Nishibori, N., Matuyama, Y., Uchida, T., Moriyama, T., Ogita, Y., Oda, M., & Hirota, H. (2003). Spatial and temporal variations in free polyamine distributions in Uranouchi Inlet, Japan. *Marine Chemistry*, *82*(3-4), 307–314. doi:10.1016/S0304-4203(03)00076-8

Nishibori, N., & Nishijima, T. (2004). Changes in polyamine levels during growth of a red-tide causing phytoplankton Chattonella antiqua (Raphidophyceae). *European Journal of Phycology*, *39*(1), 51–55. doi:10.1080/0967026031001636677

Parveen, B., Reveilliez, J.-P., Mary, I., Ravet, V., Bronner, G., Mangot, J.-F., Domaizon, I., et al. (2011). Diversity and dynamics of free-living and particle-associated Betaproteobacteria and Actinobacteria in relation to phytoplankton and zooplankton

communities. *FEMS microbiology ecology*, *77*(3), 461–76. doi:10.1111/j.1574-6941.2011.01130.x

Rösel, S., & Grossart, H. (2012). Contrasting dynamics in activity and community composition of free-living and particle-associated bacteria in spring. *Aquatic Microbial Ecology*, *66*(2), 169–181. doi:10.3354/ame01568

Saba, G. K., Steinberg, D. K., & Bronk, D. a. (2011). The relative importance of sloppy feeding, excretion, and fecal pellet leaching in the release of dissolved carbon and nitrogen by Acartia tonsa copepods. *Journal of Experimental Marine Biology and Ecology*, *404*(1-2), 47–56. doi:10.1016/j.jembe.2011.04.013

Salcher, M. M., Pernthaler, J., & Posch, T. (2010). Spatiotemporal distribution and activity patterns of bacteria from three phylogenetic groups in an oligomesotrophic lake. *Limnology and Oceanography*, *55*(2), 846–856. doi:10.4319/lo.2009.55.2.0846

Salcher, M. M., Pernthaler, J., Frater, N., & Posch, T. (2011). Vertical and longitudinal distribution patterns of different bacterioplankton populations in a canyon-shaped, deep prealpine lake. *Limnology and Oceanography*, *56*(6), 2027–2039. doi:10.4319/lo.2011.56.6.2027

Salcher, M. M., Posch, T., & Pernthaler, J. (2012). In situ substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. *The ISME journal*, 1–12. doi:10.1038/ismej.2012.162

Salomon, P. S., Janson, S., & Granéli, E. (2003). Molecular identification of bacteria associated with filaments of Nodularia spumigena and their effect on the cyanobacterial growth. *Harmful Algae*, *2*(4), 261–272. doi:10.1016/S1568-9883(03)00045-3

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., … Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537–41. doi:10.1128/AEM.01541-09

Shade, A., Kent, A., & Jones, S. (2007). Interannual dynamics and phenology of bacterial communities in a eutrophic lake. *Limnology and …*, *52*(2), 487–494. Retrieved from http://www.jstor.org/stable/10.2307/40006100

Sharma, A. K., Sommerfeld, K., Bullerjahn, G. S., Matteson, A. R., Wilhelm, S. W., Jezbera, J., Brandt, U., et al. (2009). Actinorhodopsin genes discovered in diverse freshwater habitats and among cultivated freshwater Actinobacteria. *The ISME journal*, *3*(6), 726–37. doi:10.1038/ismej.2009.13

Sumper, M., Kro, N., & Regensburg, D.-. (2004). Silica formation in diatoms : the function of long-chain polyamines and silaffins, 2059–2065.

Sumper, M., Brunner, E., & Lehmann, G. (2005). Biomineralization in diatoms: characterization of novel polyamines associated with silica. *FEBS letters*, *579*(17), 3765–9. doi:10.1016/j.febslet.2005.06.001

Tarao, M., Jezbera, J., & Hahn, M. W. (2009). Involvement of cell surface structures in size-independent grazing resistance of freshwater Actinobacteria. *Applied and environmental microbiology*, *75*(14), 4720–6. doi:10.1128/AEM.00251-09

Warnecke, F., Amann, R., & Pernthaler, J. (2004). Actinobacterial 16S rRNA genes from freshwater habitats cluster in four distinct lineages. *Environmental Microbiology*, *6*(3), 242–253. doi:10.1111/j.1462-2920.2004.00561.x

Warnecke, F., Sommaruga, R., Sekar, R., Julia, S., Pernthaler, J., & Hofer, J. S. (2005). Abundances , Identity , and Growth State of Actinobacteria in Mountain Lakes of Different UV Transparency Abundances , Identity , and Growth State of Actinobacteria in Mountain Lakes of Different UV Transparency. doi:10.1128/AEM.71.9.5551

Yannarell, a C., Kent, a D., Lauster, G. H., Kratz, T. K., & Triplett, E. W. (2003). Temporal patterns in bacterial communities in three temperate lakes of different trophic status. *Microbial ecology*, *46*(4), 391–405. doi:10.1007/s00248-003-1008-9

Yannarell, A. C., & Triplett, E. W. (2005). Geographic and Environmental Sources of Variation in Lake Bacterial Community Composition †. *Applied and environmental microbiology*, *71*(1). doi:10.1128/AEM.71.1.227

Zakharova, Y. R., Galachyants, Y. P., Kurilkina, M. I., Likhoshvay, A. V, Petrova, D. P., Shishlyannikov, S. M., Ravin, N. V, et al. (2013). The structure of microbial community and degradation of diatoms in the deep near-bottom layer of Lake Baikal. *PloS one*, *8*(4), e59977. doi:10.1371/journal.pone.0059977

Zeder, M., Peter, S., Shabarova, T., & Pernthaler, J. (2009). A small population of planktonic Flavobacteria with disproportionally high growth during the spring phytoplankton bloom in a prealpine lake. *Environmental microbiology*, *11*(10), 2676–86. doi:10.1111/j.1462-2920.2009.01994.x

5. FRESHWATER ACTINOBACTERIA RESPONSE TO PHYTOPLANKTON AND PHYTOPLANKTON-ASSOCIATED SUBSTANCES

Trevor W Ghylin[1], Yujin Lee[1], Ben Crary[1], Wen-Chi Tseng[2], Katherine D McMahon[1]

1 Department of Civil and Environmental Engineering, University of Wisconsin, Madison, WI, USA

2 Department of Chemical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan

*correspondence: Katherine McMahon, Department of Civil and Environmental Engineering, University of Wisconsin 1550 Linden Drive Room 5552 Madison, WI 53706, Email: kdmcmahon@cae.wisc.edu

Key words: acI lineage, freshwater bacteria, mesocosm, diatoms, cyanobacteria, phytoplankton, putrescine, polyamines, cyanophycin

To be submitted to Microbial Ecology

Contributions: Trevor Ghylin designed and executed all aspects of the mesocosm experiment with assistance from Yujin Lee. Mr. Ghylin performed most of the analysis and drafted the manuscript. Ben Crary processed the raw DNA data to assign taxonomy using Mothur. Wen-Chi Tseng provided cyanophycin for the experiment. Katherine McMahon was the primary investigator and advisor of Mr. Ghylin.

5.1. ABSTRACT

Our previous genome-based studies indicated that members of the freshwater acI lineage of Actinobacteria contain genes provide these organisms with the ability to take up and metabolize nitrogen-rich organic compounds such as polyamines and cyanophycin, which are present in phytoplankton. In order to further elucidate the role of these genes in acI's

ecological fitness, we set up lake mesocosms and added various organic nitrogen and carbon compounds in attempts to enrich acI. We used quantitative polymerase chain reaction (qPCR) and 16S rRNA gene tag sequencing to track acI abundance over time. The qPCR and tag sequencing results demonstrate enrichment of acI in the presence of putrescine and aspartic acid after a period of 17 days. AcI relative abundance was lowest in the mesocosms containing N-acetylglucosamine, arginine, or spermidine. AcI relative abundance was greater in mesocosms amended with killed phytoplankton than in those amended with live phytoplankton and acI abundance increased over time even though acI decreased in abundance over time in the no-amendment control. This experiment provides evidence that acI can consume putrescine while revealing possible hints at a relationship between acI and phytoplankton.

5.2. INTRODUCTION

Members of the acI lineage of Actinobacteria are often numerically dominant in freshwater lakes, sometimes representing more than 50% of the bacterial community (Newton, 2011). However, these bacteria have been difficult to study in a controlled lab setting due to difficulty in cultivation (Hahn, 2009). As a result, many studies have utilized molecular tools such as FISH, PCR, and metagenomics to study the acI lineage (Newton, 2007; Sharma, 2009). Our previous studies utilized single cell genomics to reveal acI genes focused on consuming organic nitrogen compounds (i.e. polyamines and cyanophycin) (Garcia, 2013; Ghylin2014) and 16S tag sequencing to reveal a potential link between acI and phytoplankton (diatoms and cyanobacteria) (unpublished). Previous studies have shown that polyamines are important in the structuring of the diatom cell wall (Kroger, 2000; Sumper, 2004; Sumper, 2005; Mou, 2011), and cyanophycin is an important storage polymer in cyanobacteria (Lawry, 1982). Surprisingly, previous studies have shown a negative correlation between acI and DOC, nutrients and phytoplankton blooms (Burkert, 2003; Newton, 2011a; Eckert, 2012). Eckert et al (2013) found that Actinobacteria relative abundance decreased during a phytoplankton bloom but increased later during the grazing of the phytoplankton bloom. The authors indicated this behavior may be due to the documented grazing resistance of acI (Jezbera, 2005; Jezbera, 2006; Tarao, 2009). Other studies have also correlated acI with grazing periods, especially heterotrophic nanoflagellate grazing (Pernthaler, 2001; Salcher, 2010). However, top down control (i.e. grazing) may not be the only factor (or even the primary factor) contributing to acI's relative success during grazing events. Instead, the release of complex organics during phytoplankton grazing and lysis may stimulate acI populations. Studies have demonstrated that acI consumes cell wall-derived

organic carbon such as N-acetylglucosamine (NAG is a subunit of peptidoglycan in bacterial cell walls) and chitin (polymer of NAG found in cell wall of diatoms and the exoskeleton of crustaceans such as daphnia and copepods) (Beier, 2011; Eckert, 2012; Eckert, 2013).

5.2.1. GENOMIC LINKS BETWEEN acI AND CYANOBACTERIA

The acI genomes include putative cyanophycinase genes indicating the potential ability to degrade storage polymers produced by cyanobacteria (Ghylin, 2014). The genomes of acI also encode for genes to take up lipopolysaccharides, which make up the outer membranes of gram-negative bacteria, such as cyanobacteria and most other aquatic bacteria (Ghylin, 2014). The acI (Actinobacteria) are the predominant gram-positive bacteria in aquatic systems as the only other gram-positive phylum (Firmicutes) are not typically abundant. The gram-positive bacteria lack the outer membrane and lipopolysaccharides of the gram-negative bacteria that constitute most other aquatic bacteria (i.e. Proteobacteria, Cyanobacteria, Bacteroidetes). Furthermore, the acI genomes include putative genes for colicin production, which is an antibiotic that attacks bacterial outer membranes and causes cell lysis. The acI genomes also contain lysozyme M1 (Garcia, 2012; Ghylin, 2014) which attacks the bond between NAG and NAM (N-acetyl muramic acid) subunits of peptidoglycan in bacterial cell walls. These genes indicate a potential role for acI in the degradation (and possibly attack) of cyanobacteria and other freshwater bacteria.

Polyamines are also a component of cyanobacteria and other freshwater bacteria as they are essential in all living cells. The putative polyamine uptake genes found in acI genomes could potentially be used to consume polyamines released from cyanobacterial lysis.

5.2.2. L**INK BETWEEN AC**I **AND** P**HYTOPLANKTON**

Bacteria are thought to utilize about half of the organic matter produced by diatoms using hydrolytic enzymes (alpha- and beta-glucosidases, lipase, protease, alkaline phosphatase and chitinase) to attack live and killed diatoms (Polarizing, 1999). The acI genomes contain some of these enzymes including lipase and the glucosidases (found in acI-A only) (Ghylin, 2014).

This information, along with the genomic evidence discussed previously, indicates a potential reliance of acI on autochthonous organic nitrogen and carbon produced by phytoplankton (polyamines from diatoms and cyanophycin from cyanobacteria) and cell wall components of diatoms (polyamines) and gram-negative bacteria (lipopolysaccharides in Cyanobacteria, Proteobacteria, Bacteroidetes). AcI likely gains access to these materials when they are released from sloppy feeding of grazers on diatoms (Moller, 2004) and bacteria (cyanobacteria and heterotrophs), cell autolysis (Brussaard, 1995; Sarthou, 2005; Saba, 2011), or direct attack from acI antibiotics and lyzozyme on cyanobacteria and other gram-negative bacteria. Previous studies seem to indicate that acI relative abundance declines during phytoplankton blooms when other bacteria proliferate on exudates from live phytoplankton (Burkert, 2003; Eckert, 2012). The acI instead thrive during grazing periods after blooms when phytoplankton and other bacterial cells are lysed (Eckert, 2012), a process which acI may actively facilitate with antibiotics, lysozyme and hydrolytic enzymes. Part of acI's success may be due to grazing resistance (top down control), however it also seems that the acI are well suited to facilitate cell lysis and compete for complex organics released in this process (bottom up control).

In order to determine whether acI is able to compete well for lysis products from phytoplankton and bacteria, we designed a mesocosm experiment to remove grazers from lake water and amend with various substrates related to phytoplankton and bacteria. We tracked the resulting acI population using qPCR and 16S DNA sequencing (Illumina Miseq).

5.3. RESULTS AND DISCUSSION

This mesocosm experiment was designed to test the hypotheses that acI consumes cyanophycin (and degradation products: arginine, aspartic acid), phytoplankton-derived (diatoms and cyanobacteria) organic matter, polyamines (putrescine and spermidine) and N-acetylglucosamine, a breakdown product of chitin and peptidoglycan from bacterial cell walls.

Twenty-four liters of water were collected from the deep hole (43.098N, 89.405W) in Lake Mendota (Madison, WI) on July 29, 2013. The water was collected in a 12m long, ¾ inch (2cm) diameter clear flexible plastic tube. The tube was lowered through the water column, corked and retrieved to obtain a 12m integrated sample of the epilimnion. Filling the tube produces about 3.5L, thus the tube was filled 7 times to obtain at least 24 liters. Water was stored in the lab at room temperature. The water was taken to the lab and vacuum filtered through Whatman 595 (4-7um) filter paper on Tuesday July 30, 2013. The filtered water was then filtered again onto a 2um polycarbonate Millipore (47mm) filter on Thursday August 1, 2013, similar to the method use by Mou et al (2011). This pore size was chosen to allow passage of bacteria but exclude most grazers including heterotrophic nanoflagellates which are known to graze on bacteria and are mostly 2-10um in length (Callieri, 2002). This filtration likely excluded other organisms as well including colonial and filamentous (i.e. Aphanizomenon, Anabaena, and Oscillatoria) cyanobacteria. However, some cyanobacteria in Lake Mendota are small unicellular organisms (Aphanocapsa, Aphanothece, Microcystis) that could potentially pass through this filter. A small pore size is important as Jezbera et al (2006) showed that filtering through 5um actually enriched heterotrophic nanoflagellate (HNF) grazers as it removed their primary predator, the ciliates but did not remove the HNF.

Filtration would likely also remove most ciliates and crustaceans such as daphnia and copepods.

A bacterial sample was collected from the twice-filtered water by filtering 250ml through a 0.2um filter. This sample represents the "Day 0" bacterial community for the mesocosms.

The twice-filtered water was mixed in five gallon buckets on Friday August 2, 2013 and 1L was added to each of 24 two-liter clear polycarbonate bottles. Substrate amendments were made to the bottles as shown in Table 6. Caps were screwed securely onto the bottles. More details regarding the diatom and cyanobacteria additions can be found in the following section.

The mesocosms were deployed in Lake Mendota on Friday August 2, 2013 (Day 0) at 2pm (CST), 5 meters from shore at the Center for Limnology. The mesocosm bottles were contained in a cargo net affixed to a pier with a rope. The bottles floated approximately 2 meters away from the pier and were exposed to sunlight and wave action.

5.3.1. CYANOBACTERIA AND DIATOMS

Gloeotrichia (cyanobacteria, University of Texas Algae Culture Collection ID:LB941) was grown in 1 liter of BG11 culture media (Life Technologies Part#: A1379901) for approximately 3 weeks at room temperature to a density of 1,200 filaments/ml (Figure 15). Cyclotella meneghiniana was obtained from Linda Graham at UW-Madison (Graham, 1982, 1995, 2012). Cyclotella was grown in 2 liters of modified SD11 media as per Graham et al (2012) for approximately 4 weeks to a density of 20,000 cells/ml (Figure 15).

Half of the Cyclotella culture (1 Liter) was centrifuged at 4,000 rpm for 10 minutes in 250mL bottles to concentrate the cells and remove culture media. Supernatant was poured off

and the resulting final concentrated volume was 40mL. The 40mL was centrifuged again at 4,000 rpm for 10 minutes. Supernatant was poured off resulting in a final volume of 8mL of concentrated Cyclotella. Approximately 1ml of the concentrated culture was removed and archived in a -80C freezer for future analysis, leaving 7mL of concentrated culture.

Half of the Gloeotrichia culture (500mL) was centrifuged at 4,000 rpm for 10 minutes in 250mL bottles to concentrate the cells and remove culture media. Supernatant was poured off and the resulting final concentrated volume was 40mL. The 40mL was centrifuged again at 4,000 rpm for 10 minutes. Supernatant was poured off resulting in a final volume of 8mL of concentrated Gloeotrichia. Approximately 1ml of the concentrated culture was removed and archived in a -80C freezer for future analysis, leaving 7mL of concentrated culture.

The Cyclotella and Gloeotrichia concentrated cultures were frozen in a -80C freezer then thawed in warm water to kill and lyse the cells. The cultures were frozen a second time in an ethanol/dry ice bath and thawed again in warm water.

The other half of the cultures (1L of Cyclotella and 500mL of Gloeotrichia) received the same centrifuge treatment as described above to result in a concentrated end product of about 7mL. These cultures were not frozen and remained alive when added to the mesocosms.

Table 6. Substrates tested in microcosms

Microcosm #	Content
1-2	Control
3-4	Cyanophycin Soluble (~30 mg L_{-1})
5-6	Cyanophycin Insoluble (~30 mg L_{-1})
7-8	Cyanobacteria Live
9-10	Cyanobacteria Killed
11-12	Arginine (~30 mg L_{-1})
13-14	Aspartic acid (~30 mg L_{-1})
15-16	Spermidine (~30 mg L_{-1})
17-18	Putrescine (~30 mg L_{-1})

19-20	Diatoms Live
21-22	Diatoms Killed
23-24	NAG (N-acetylglucosamine) (~30 mg L$_{-1}$)

Notes: Soluble and Insoluble cyanophycin were obtained from Wen-Chi Tseng (Tseng, 2012)

5.3.2. SAMPLE COLLECTION

Samples were collected on day 3 (August 5, 2014), day 7 (August 9, 2014), and day 17 (August 19, 2014) as follows:

- All bottles were removed from the lake

- 125mL was removed from each of the mesocosms; the mesocosms were placed back in the lake; the 125mL samples were brought to the lab

- 50ml of sample was filtered onto a 0.2um polyethersulfone filters (Pall-Supor200, Gelman). Filters were frozen at -80°C and stored prior to deoxyribonucleic acid (DNA) extraction.

- The remainder (75mL) was preserved for future analysis

5.3.3. DNA EXTRACTION

Total community DNA was extracted from filters using the QBiogene Bio101 FastDNA kit, using manufacturer's instruction with modification as described previously (Yannarell, 2003; 2005).

5.3.4. REAL-TIME POLYMERASE CHAIN REACTION (QPCR)

QPCR reactions were performed in a 96 well optical Icycler plate with clear optical film cover. Each well was filled with 7.5 ul of IQ Sybrgreen supermix, 0.9 ul forward primer (20uM), 0.9 ul reverse primer (20uM), 4.7 ul de-ionized water, 1 ul template (extracted DNA from lake sample).

Quantitative polymerase chain reaction (qPCR) was performed on sections of the 16S rRNA gene with acI primer and annealing temperature as per Garcia et al (Garcia, 2013).

Standards and negative controls were created from a clone library obtained from Lake Mendota. Eight-point calibration curves for qPCR were produced by 10-fold serial dilution of positive controls in triplicate within each assay, at 10^1 to 10^8 target copies per reaction. QPCR was conducted on a Biorad Icycler, and conditions were as follows: 2 min at 50°C, 3 min initial denaturation at 95°C, 40 cycles of 40 s denaturation at 94°C, 30 s annealing at temperature shown in Garcia et al (Garcia, 2013), 30s elongation at 72°C. The qPCR run was immediately followed by a melt curve step starting at 55°C for 10s and increasing setpoint temperature by 0.5C every 10s until 95°C (about 80 steps). The machine finished with a final 55°C for 1 min, then infinite hold at 4°C.

Standards and negative controls were generated from appropriate clones (acI-B1 for standards and acIV-C for negative control) from 16S plus internal transcribed spacer rRNA clone libraries created from Lake Mendota samples. Briefly, clones were obtained from -80C clone libraries. Clones were re-grown in LB with 50 ug/ml kanamycin for 16 hours at 37C. Plasmid was extracted from the liquid cultures using a QIAGEN plasmid mini kit (QIAGEN, Valencia, CA). The plasmid was purified with a QIAquick PCR purification kit (QIAGEN), and mass concentration was determined by nanodrop. Standards were made via 16S PCR amplification of the plasmids using Promega goTaq hot start mix (25 ul) with M13 primers (M13f (20 uM; 0.1 ug/ul): 1 ul; M13r (20 uM;

0.1 ug/ul): 1ul); Template (plasmid prep): 1 ul. PCR cycle: 94 C - 2 min, 25 cycles of (94 C

– 35s, 55 C – 45 s, 72 C – 2 min) 72 C – 5 min, 4 C hold. Mass concentration of the

standards was determined using nanodrop and diluted with TE buffer to obtain 10^8 copies/ul

standards.

Copy number was calculated based on the mass concentration and the average

molecular weight of the plasmid. For all unknown samples, 1 ul of community-derived

genomic DNA was added as the template in qPCR reactions. In each assay, a no-template

control was included to check for contamination and primer-dimer formation. To avoid

nonspecific cross-detection, a negative control (acIV-C) (a plasmid containing a nontarget

16S fragment with the fewest mismatches to the primer set) was also included and applied at

10^8 copies per reaction.

5.3.5. 16S PCR AND DNA SEQUENCING

10ul of extracted DNA from samples from day 0, day 3 and day 17 were sent to MR

DNA (www.mrdnalab.com, Shallowater, TX, USA) for 16S rRNA gene amplification (V4

region: 515/806) and DNA sequencing with Illumina Miseq.

The 16S rRNA gene V4 variable region PCR primers 515/806 with barcode on the

forward primer were used in a 30 cycle PCR using the HotStarTaq Plus Master Mix Kit

(Qiagen, USA) under the following conditions: 94°C for 3 minutes, followed by 28 cycles of

94°C for 30 seconds, 53°C for 40 seconds and 72°C for 1 minute, after which a final

elongation step at 72°C for 5 minutes was performed. After amplification, PCR products

were checked in 2% agarose gel to determine the success of amplification and the relative

intensity of bands. Multiple samples were pooled together (e.g., 100 samples) in equal

proportions based on their molecular weight and DNA concentrations. Pooled samples were

purified using calibrated Ampure XP beads. Then the pooled and purified PCR product was used to prepare DNA library by following Illumina TruSeq DNA library preparation protocol. Sequencing was performed at MR DNA (www.mrdnalab.com, Shallowater, TX, USA) on a MiSeq following the manufacturer's guidelines.

Bioinformatics

All sequences were processed with Mothur, using the Schloss SOP with minor modifications (Schloss, 2009). Chimeras were removed with the uchime algorithm (Edgar, 2011). Reads were aligned to the Greengenes database and subsampled to 3,800 reads per sample. OTUs were clustered at 97% using the average neighbor linkage algorithm.

OTUs were first classified against a custom curated taxonomic database that includes common freshwater taxa (Newton, 2011). OTUs that could not be assigned with 70% confidence to the 5th taxonomic level with this database were reclassified against the Greengenes. Classifications were kept if they were made with at least 60% confidence.

5.4. RESULTS

Figure 16 and **Table 7** shows the qPCR data from the mesocosm experiment. These data indicate that acI relative abundance decreased in the control over time. This may be due to bottle effects or because grazers were excluded from the water, reducing one of acI's competitive advantages.

Figure 16 (qPCR data) and Figure 17 (16S itags data) clearly demonstrates that acI becomes enriched in the mesocosms augmented with putrescine or aspartic acid. However, it is interesting to note that acI relative abundance declines in the first three days (Figure 17) in these mesocosms then increases in later sampling dates (day 7 and 17). DNA sequencing data

resulted in the same findings (Figure 17). Interestingly, acI was not enriched in mesocosms with spermidine, a polyamine similar to putrescine. These findings correspond to those of Garcia et al (2013) who also found that putrescine but not spermidine enriched acI. AcI was also not enriched in mesocosms with arginine, an amino acid similar to aspartic acid. A previous CARD-FISH MAR study (Salcher, 2012) indicated that acI took up leucine, thymidine, glucose and a mixture of 15 amino acids. That study found that acI did not take up aspartic acid. Surprisingly, our results indicate that the acI from Lake Mendota benefit from aspartic acid whether they directly consume it or benefit from other bacteria consuming this substance.

The enrichment of acI in the presence of aspartic acid indicates a potential relationship with cyanophycin degradation. Cyanophycin is a granular storage polymer produced by cyanobacteria and some other bacteria (Krehenbrink, 2002). The acI genomes contain putative cyanophycinase genes that seem to encode the enzyme responsible for cyanophycin degradation (Ghylin, 2014). Cyanophycin is degraded to arginine/aspartic acid dipeptides by cyanophycinase (Richter, 1999). These dipeptides are further degraded to release arginine and aspartic acid by dipeptidase, which is also putatively encoded in the acI genomes (Ghylin, 2014).

Although the acI genomes have putative cyanophycinase genes (Ghylin, 2014), our mesocosm experiment did not enrich acI (compared to the control group) when soluble or insoluble cyanophycin was added. However, the acI relative abundance was increasing over time in two of the cyanophycin mesocosms while the control mesocosms were decreasing. Further investigation should be done with a longer experiment (60 days) to test the effect of cyanophycin on acI.

The Actinomycetales, the order containing acI, are known as saprophytes (consume nutrients from dead organic matter) and are known for breaking down complex organic polymers (Goodfellow, 1983). Our previous in situ study of acI demonstrated a lag correlation between acI and phytoplankton, indicating that acI may proliferate from lysis products following phytoplankton blooms rather than exudate during blooms. We aimed to test this hypothesis in this mesocosm study by adding live and killed phytoplankton to different mesocosms and monitoring acI.

The qPCR data (relative abundance) in Figure 16 shows that acI was not enriched (i.e. greater than the control) in the live or killed phytoplankton mesocosms as compared to the control. However, this figure indicates that acI relative abundance and copies/ul are increasing in the diatom mesocosms over time while decreasing in the control. If the experiment had run longer we may have seen further enrichment of acI. In contrast, acI relative abundance decreases over time in the mesocosms containing live cyanobacteria, arginine, spermidine and N-acetylglucosamine.

The 16S itag data in Figure 17 shows that acI relative abundance was relatively high in the killed diatom mesocosms as compared with other substrates at day 17. This figure also shows that acI relative abundance was higher in one of the killed cyanobacteria mesocosms at day 17 as compared with the control.

AcI relative abundance was higher in the killed diatom and cyanobacteria mesocoms as compared with the live diatom and cyanobacteria mesocosms (Figure 16, Figure 17), providing support for the hypothesis that acI benefits from decay products of phytoplankton blooms more than exudate from live phytoplankton blooms.

Although some research (Eckert, 2012; Eckert, 2013) has demonstrated that acI consume N-acetylglucosamine, our mesocosm experiment did not enrich acI when this substance was added (Figure 16, Figure 17).

The soluble cyanophycin and killed cyanobacteria mesocosms had high standard deviations at day 17 (s > 100% of mean). All other mesocosms had relatively low standard deviation (s < 25% of mean). Standard deviation is shown with error bars on Figure 16. The standard deviation between the duplicate mesocosms tended to increase over time. The cause for the high variation in duplicates is unknown. The high standard deviation was consistent across both the qPCR and 16S itag datasets, indicating the variation is likely an actual difference in bacterial community in the mesocosms rather than variation in the analytical methods. The bottles may have had slightly different conditions for sunlight and wave action that may have resulted in diverging communities over time.

As all mesocosm water was pre-filtered through 2um filters, we excluded grazers that have been implicated in acI's success (Tarao, 2009). Therefore, enrichment of acI should be due solely to bottom-up factors (i.e. acI competing for substrate). Though top-down factors such as grazing resistance are likely important for acI's success, our data provides further evidence that bottom-up factors are likely also important.

5.4.1. OTHER BACTERIA

BetI-A (betaproteobacteria) was the most abundant bacteria on average in all mesocosms (Figure 18, Figure 19) followed by Thioclava (alphaproteobacteria) (Figure 20), Novo-A1 (alphaproteobacteria) (Figure 21), Pseudo-A1 (Gammaproteobacteria) (Figure 22), and f__Pirellulaceae (Planctomycetes) (**Figure 23**). BetI-A was not enriched in the controls and was most enriched in aspartic acid and putrescine mesocosms, indicating a potential

relationship with acI which was also abundant in those mesocosms. This data agrees with the correlation between acI and BetI-A we found in the ten-year Lake Mendota time series.

Polynucleobacter was also abundant in the aspartic acid mesoscosms while alfII Brev, LD12 and Thioclava were abundant in the putrescine mesocosms. The ten-year time series data also indicated a correlation between LD12 and acI.

5.4.2. LIMITATIONS OF THIS STUDY

This study utilized qPCR and 16s rRNA gene iTag sequencing to determine relative abundances of freshwater bacteria and the effect of various substrates on acI populations. Due to the nature of relative abundance data, it is possible that a bacterial population can have no change in absolute abundance over time while the relative abundance increases or decreases due to proliferation or death of other bacteria. It is possible that the apparent correlation between acI and putrescine and aspartic acid was due simply to a decrease in the absolute abundance of other bacteria while acI absolute abundance remained stable. It is also possible that putrescine and aspartic acid had some toxic effects on freshwater bacteria at the dosed concentrations and acI was simply more resistant to this toxicity. However, the EUB qPCR data seem to indicate that total bacteria counts increased in these mesocosms even though this method is prone to DNA extraction and PCR bias.

Additionally, though we attempted to exclude grazers with filtration, it is possible that some grazers passed through the prefiltration step and into the mesocosms. Thus grazing could have been a driver of acI increase in relative abundance as the grazers consumed non-acI bacteria. This is consistent with acI's known resistance to grazing.

The results of this study, namely the correlations between acI, putrescine and aspartic acid should be confirmed with absolute abundance data using techniques such as fluorescent in situ hybridization (FISH).

5.5. CONCLUSIONS

This mesocosm experiment further demonstrated the connection between acI and putrescine as seen in prior work (Garcia, 2013; Ghylin, 2014). Our mesocosm work also demonstrated enrichment of acI in the presence of aspartic acid, which has not been demonstrated previously. The data from this study indicate that acI is more enriched in killed phytoplankton mesocosms than in live phytoplankton mesocosms. We also found that acI relative abundance increased over time in the live and killed diatom mesocosms and one of the killed cyanobacteria and soluble cyanophycin mesocosms. A longer duration experiment may show long-term enrichment of acI with these substrates.

Although we have learned about acI from genomic investigation (Garcia, 2012; Ghylin, 2014), this lineage remains somewhat elusive to study in an experimental setting. The genomes include putative genes for polyamine uptake, which corresponds well with our finding that putrescine enriches acI. However, the genomes also include putative cyanophycinase genes. Surprisingly, we were not able to demonstrate enrichment of acI with cyanophycin (soluble or insoluble), although acI relative abundance was increasing in one soluble cyanophycin mesocosms from day 3 to day 17.

Genomic evidence seems to indicate a relationship between acI and phytoplankton (Ghylin, 2014). The data show acI relative abundance increasing in the diatom mesocosms from day 3 to day 17 while decreasing in the control and live cyanobacteria mesocosms. AcI

relative abundance was greater in the killed phytoplankton mesocosms (diatom and cyanobacteria) as compared with the live phytoplankton mesocosms.

Evidence from this experiment hints at a relationship between acI and phytoplankton as acI relative abundance increased over time in three of the four sets of phytoplankton-amended mesocosms (not live cyanobacteria).
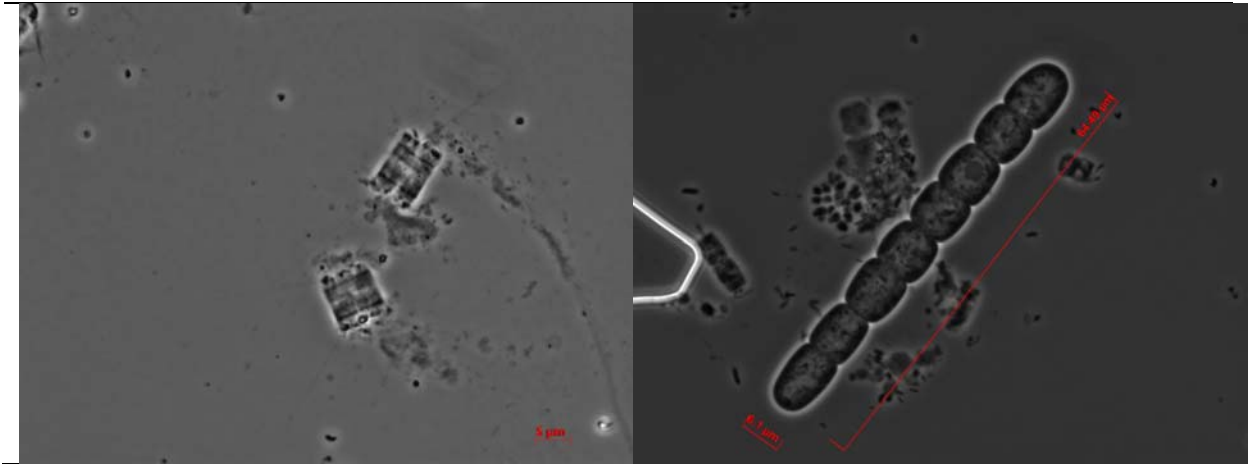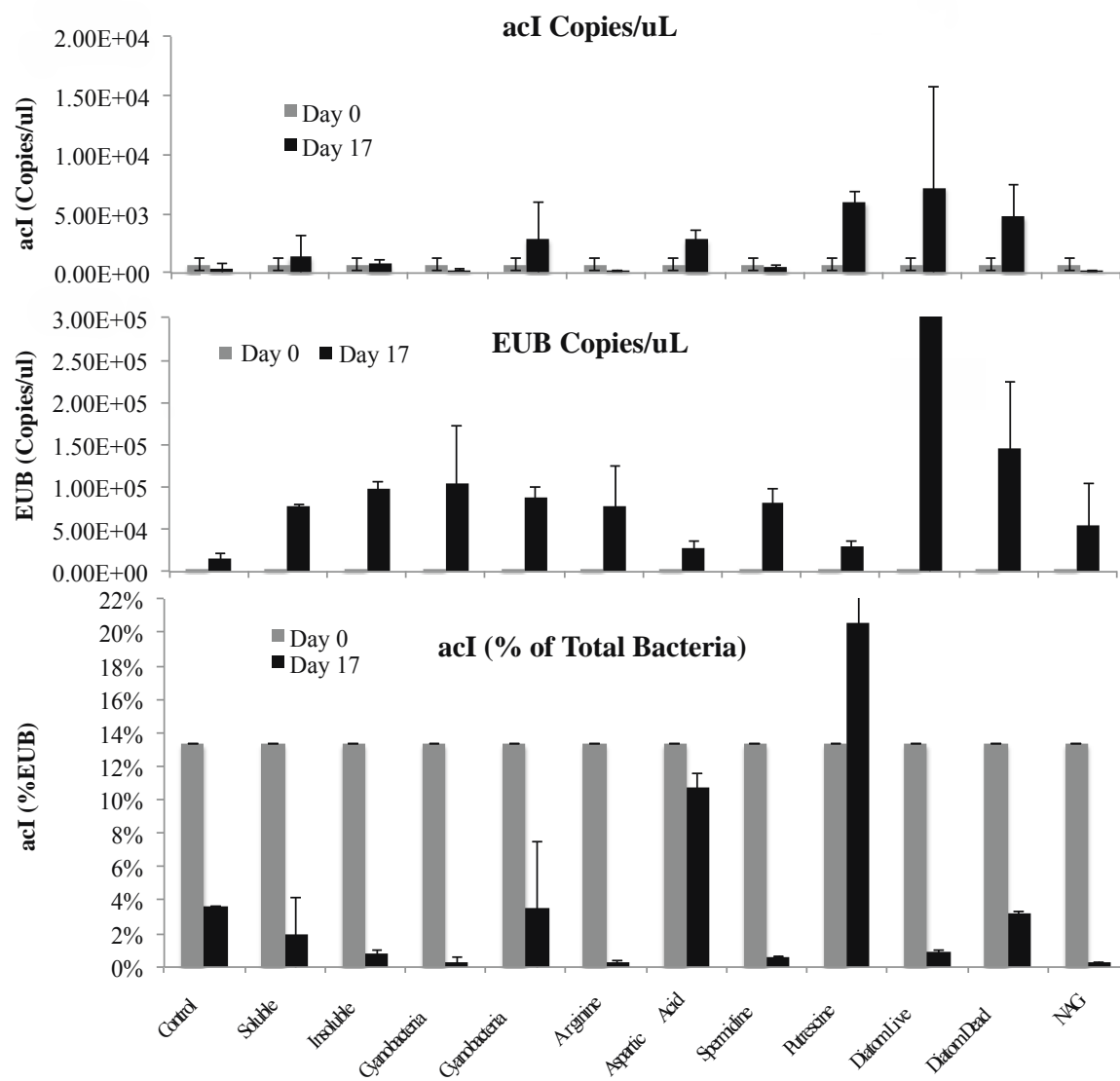
5.6. ACKNOWLEDGEMENTS

5.7. FIGURES

Figure 15. Microscope Images of Cyclotella (left) and Gloeotrichia (right) Cultures
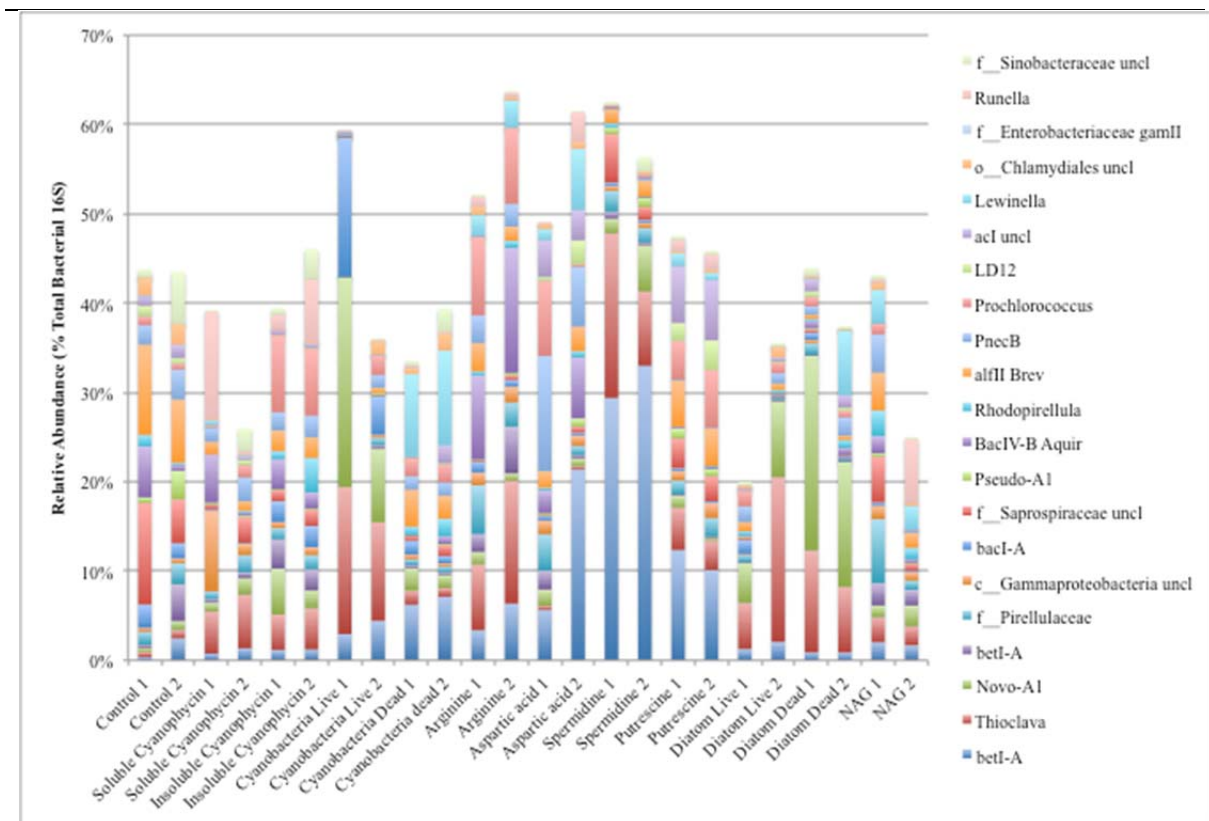
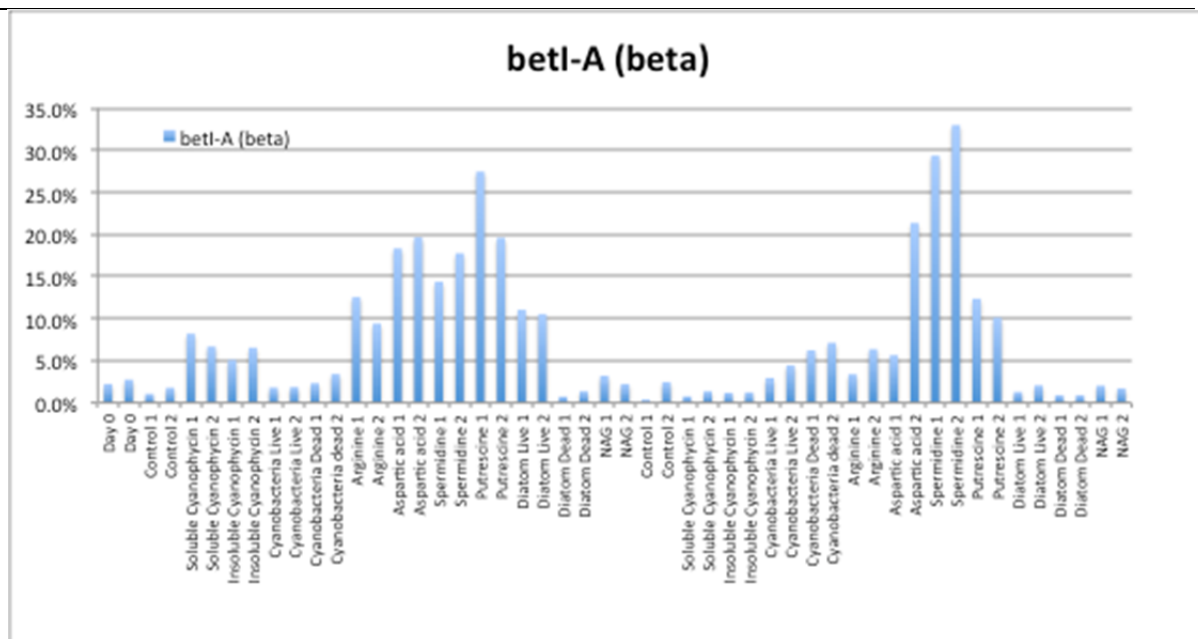**Figure 16.** acI Abundance qPCR data

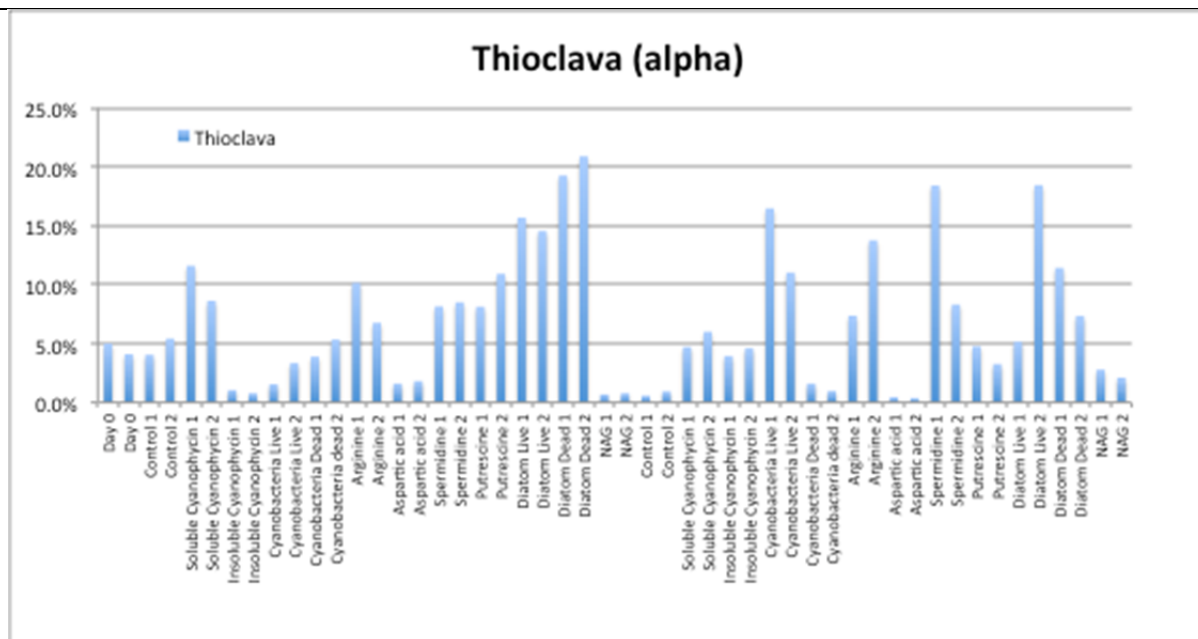**Figure 17.** acI Relative Abundance 16S itag data (% of Total Bacteria)
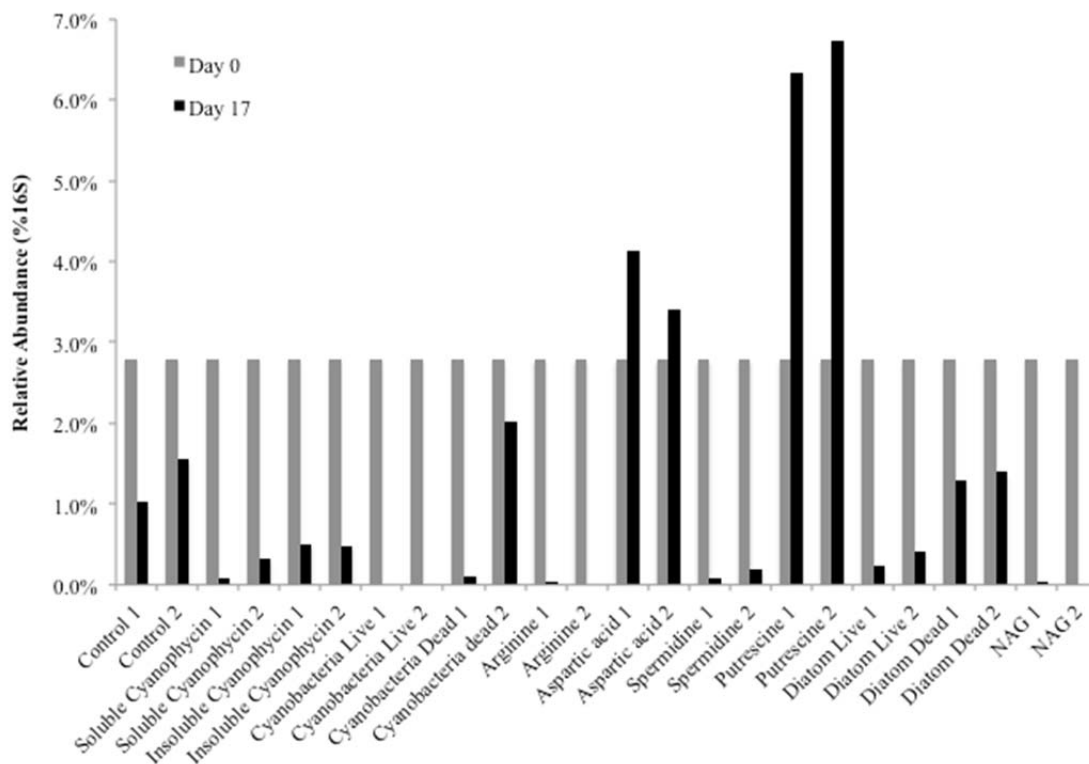
Note: Day 7 samples were not sequenced due to budget constraints.

Figure 18. Other Bacteria Relative Abundance 16S itag data (% of Total Bacteria)
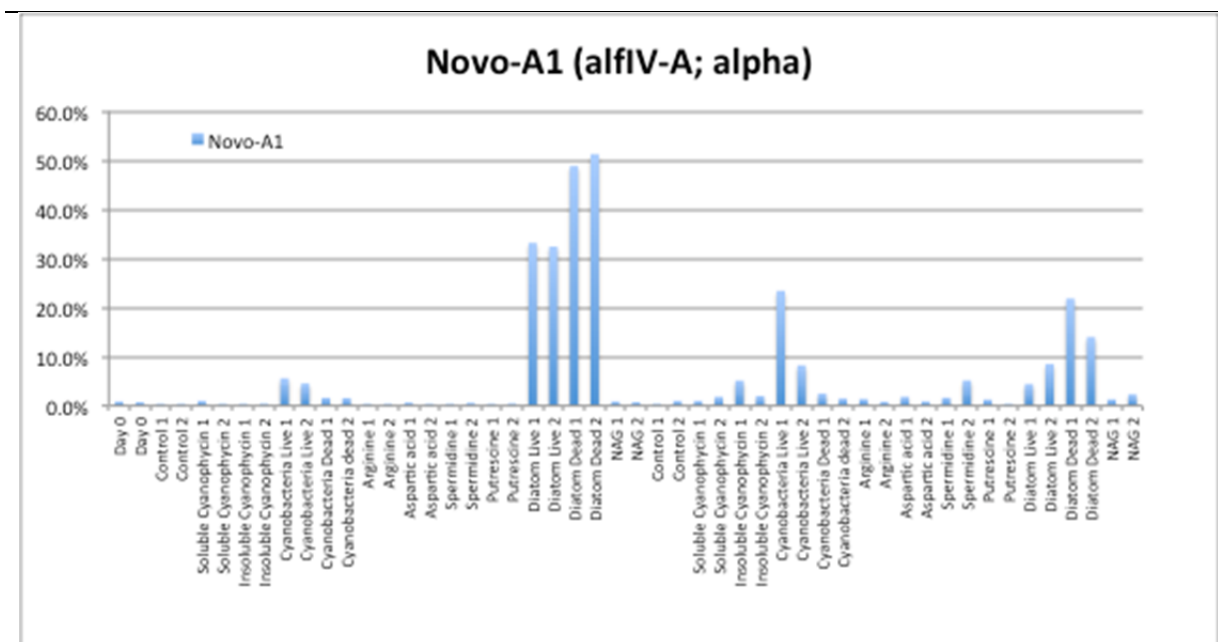
Figure 19. BetI-A; 1 of 5 most abundant mesocosm bacteria; Relative Abundance 16S Itag data (% of Total Bacteria)
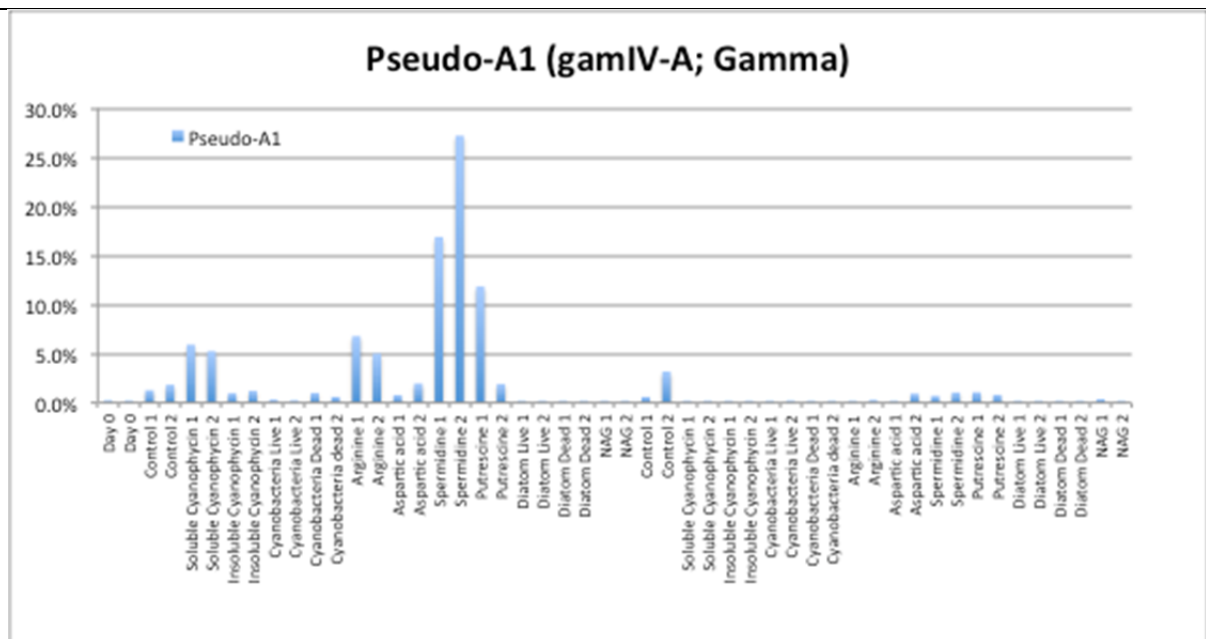

Figure 20. Thioclava; 1 of 5 most abundant mesocosm bacteria; Relative Abundance 16S Itag data (% of Total Bacteria)
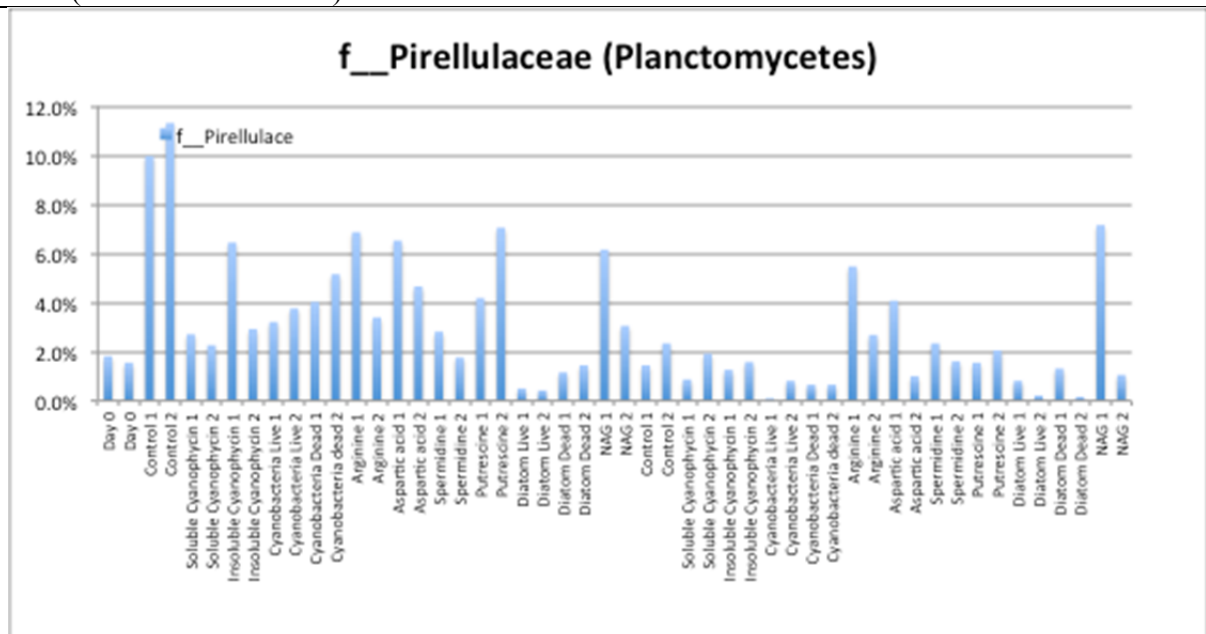
Figure 21. Novo-A1; 1 of 5 most abundant mesocosm bacteria; Relative Abundance 16S Itag data (% of Total Bacteria)

Figure 22. Pseudo-A1; 1 of 5 most abundant mesocosm bacteria; Relative Abundance 16S Itag data (% of Total Bacteria)

Figure 23. BetI-A; 1 of 5 most abundant mesocosm bacteria; Relative Abundance 16S Itag data (% of Total Bacteria)

5.8. TABLES

Table 7. acI Abundance qPCR data

		acI (%EUB)
(Day 3) Aug.5th,2013	Control 1	
	Control 2	9.4%
	Soluble Cyanophycin 1	0.3%
	Soluble Cyanophycin 2	0.4%
	Insoluble Cyanophycin 1	0.9%
	Insoluble Cyanophycin 2	0.6%
	Cyanobacteria Live 1	0.9%
	Cyanobacteria Live 2	2.4%
	Cyanobacteria Dead 1	1.0%
	Cyanobacteria dead 2	2.4%
	Arginine 1	0.0%
	Arginine 2	0.9%
	Aspartic acid 1	
	Aspartic acid 2	4.6%
	Spermidine 1	4.1%
	Spermidine 2	2.4%
	Putrescine 1	2.9%
	Putrescine 2	3.1%
	Diatom Live 1	0.1%
	Diatom Live 2	0.1%
	Diatom Dead 1	0.5%
	Diatom Dead 2	0.2%
	NAG 1	0.7%
	NAG 2	0.4%
(Day 7) Aug.9th,2013	Control 1	
	Control 2	4.7%
	Soluble Cyanophycin 1	1.4%
	Soluble Cyanophycin 2	1.1%
	Insoluble Cyanophycin 1	1.5%
	Insoluble Cyanophycin 2	1.0%
	Cyanobacteria Live 1	0.6%
	Cyanobacteria Live 2	0.8%
	Cyanobacteria Dead 1	3.5%
	Cyanobacteria dead 2	6.8%
	Arginine 1	0.6%
	Arginine 2	0.5%
	Aspartic acid 1	15.8%
	Aspartic acid 2	7.7%
	Spermidine 1	1.0%
	Spermidine 2	0.9%
	Putrescine 1	10.1%
	Putrescine 2	15.9%
	Diatom Live 1	0.7%
	Diatom Live 2	0.2%
	Diatom Dead 1	1.0%
	Diatom Dead 2	2.0%

	NAG 1	0.9%
	NAG 2	0.5%
(Day 17) Aug.19th,2013	Control 1	
	Control 2	3.6%
	Soluble Cyanophycin 1	0.3%
	Soluble Cyanophycin 2	3.5%
	Insoluble Cyanophycin 1	0.9%
	Insoluble Cyanophycin 2	0.6%
	Cyanobacteria Live 1	0.1%
	Cyanobacteria Live 2	0.5%
	Cyanobacteria Dead 1	0.8%
	Cyanobacteria dead 2	6.3%
	Arginine 1	0.1%
	Arginine 2	0.4%
	Aspartic acid 1	10.2%
	Aspartic acid 2	11.3%
	Spermidine 1	0.6%
	Spermidine 2	0.6%
	Putrescine 1	25.6%
	Putrescine 2	15.6%
	Diatom Live 1	0.8%
	Diatom Live 2	1.0%
	Diatom Dead 1	3.1%
	Diatom Dead 2	3.3%
	NAG 1	0.3%
	NAG 2	0.2%

5.9. REFERENCES

Beier, S., & Bertilsson, S. (2011). Uncoupling of chitinase activity and uptake of hydrolysis products in freshwater bacterioplankton. *Limnology and Oceanography*, *56*(4), 1179–1188. doi:10.4319/lo.2011.56.4.1179

Brussaard, C. P. D., Riegman, R., Noordeloos, A. A. M., Cadee, G. C., Witte, H., Kop, A. J., Nieuwland, G., et al. (1995). Effects of grazing , sedimentation and phytoplankton cell lysis on the structure of a coastal pelagic food web, *123*, 259–271.

Burkert, U., Warnecke, F., Babenzien, D., Pernthaler, J., & Zwirnmann, E. (2003). Members of a Readily Enriched β -Proteobacterial Clade Are Common in Surface Waters of a Humic Lake. *Applied and Environmental Microbiology*, *69*(11), 6550–6559. doi:10.1128/AEM.69.11.6550

Callieri, C., Karjalainen, S. M., & Passoni, S. (2002). Grazing by ciliates and heterotrophic nanoflagellates on picocyanobacteria in Lago Maggiore , Italy. *Journal of Plankton Research*, *24*(8), 785–796.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology*, *72*(7), 5069–72. doi:10.1128/AEM.03006-05

Edgar,RC, Haas,BJ, Clemente,JC, Quince,C, Knight,R (2011) UCHIME improves sensitivity and speed of chimera detection, *Bioinformatics* doi: 10.1093/bioinformatics/btr381 [PMID 21700674].

Eckert, E. M., Salcher, M. M., Posch, T., Eugster, B., & Pernthaler, J. (2012). Rapid successions affect microbial N-acetyl-glucosamine uptake patterns during a lacustrine spring phytoplankton bloom. *Environmental microbiology*, *14*(3), 794–806. doi:10.1111/j.1462-2920.2011.02639.x

Eckert, E. M., Baumgartner, M., Huber, I. M., & Pernthaler, J. (2013). Grazing resistant freshwater bacteria profit from chitin and cell-wall-derived organic carbon. *Environmental microbiology*, *1*. doi:10.1111/1462-2920.12083

Garcia, S. L., McMahon, K. D., Martinez-Garcia, M., Srivastava, A., Sczyrba, A., Stepanauskas, R., Grossart, H.-P., et al. (2012). Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *The ISME journal*, 1–11. doi:10.1038/ismej.2012.86

Garcia, S. L., McMahon, K. D., Grossart, H.-P., & Warnecke, F. (2013). Successful enrichment of the ubiquitous freshwater acI A ctinobacteria. *Environmental Microbiology Reports*, n/a–n/a. doi:10.1111/1758-2229.12104

Goodfellow, M; Williams, S. T. (1983). Ecology of actinomycetes. *Annual review of microbiology*, *37*(41), 189–216. doi:10.1146/annurev.mi.37.100183.001201

Graham, J. M., Auer, M. T., Canale, R. P., & Hoffmann, J. P. (1982). Ecological Studies and Mathematical Modeling of Cladophora in Lake Huron: 4. Photosynthesis and Respiration as Functions of Light and Temperature. *Journal of Great Lakes Research*, *8*(1), 100–111. doi:10.1016/S0380-1330(82)71948-3

Graham, J. M., Lembi, C. A., Adrian, H. L., & Spencer, D. F. (1995). Physiological responses to temperature and irradiance in Spirogyra (Zygnematales, Charophyceae). *Journal of Phycology*, (31), 531–540.

Graham, J. M., Graham, L. E., Zulkifly, S. B., Pfleger, B. F., Hoover, S. W., & Yoshitani, J. (2012). Freshwater diatoms as a source of lipids for biofuels. *Journal of industrial microbiology & biotechnology*, *39*(3), 419–28. doi:10.1007/s10295-011-1041-5

Hahn, M. W. (2009). Description of seven candidate species affiliated with the phylum Actinobacteria, representing planktonic freshwater bacteria. *International journal of systematic and evolutionary microbiology*, *59*(Pt 1), 112–7. doi:10.1099/ijs.0.001743-0

Jezbera, J., Hornák, K., & Simek, K. (2005). Food selection by bacterivorous protists: insight from the analysis of the food vacuole content by means of fluorescence in situ hybridization. *FEMS microbiology ecology*, *52*(3), 351–63. doi:10.1016/j.femsec.2004.12.001

Jezbera, J., Hornák, K., & Simek, K. (2006). Prey selectivity of bacterivorous protists in different size fractions of reservoir water amended with nutrients. *Environmental microbiology*, *8*(8), 1330–9. doi:10.1111/j.1462-2920.2006.01026.x

Krehenbrink, M., Oppermann-Sanio, F.-B., & Steinbüchel, A. (2002). Evaluation of non-cyanobacterial genome sequences for occurrence of genes encoding proteins homologous to cyanophycin synthetase and cloning of an active cyanophycin synthetase from Acinetobacter sp. strain DSM 587. *Archives of microbiology*, *177*(5), 371–80. doi:10.1007/s00203-001-0396-9

Kröger, N., Deutzmann, R., Bergsdorf, C., & Sumper, M. (2000). Species-specific polyamines from diatoms control silica morphology. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(26), 14133–8. doi:10.1073/pnas.260496497

Lawry, N. H; Simon, R. D. (1982). The Normal and Induced Occurrence of Cyanophycin Inclusion Bodies in Several Blue-Green Algae. *Journal of Phycology*, *18*, 391–399.

Moller, E. F. (2004). Sloppy feeding in marine copepods: prey-size-dependent production of dissolved organic carbon. *Journal of Plankton Research*, *27*(1), 27–35. doi:10.1093/plankt/fbh147

Mou, X., Vila-Costa, M., Sun, S., Zhao, W., Sharma, S., & Moran, M. A. (2011). Metatranscriptomic signature of exogenous polyamine utilization by coastal bacterioplankton. *Environmental Microbiology Reports*, *3*(6), 798–806. doi:10.1111/j.1758-2229.2011.00289.x

Newton, R. J., Jones, S. E., Helmus, M. R., & McMahon, K. D. (2007). Phylogenetic ecology of the freshwater Actinobacteria acI lineage. *Applied and environmental microbiology*, *73*(22), 7169–76. doi:10.1128/AEM.00794-07

Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., & Bertilsson, S. (2011). *A guide to the natural history of freshwater lake bacteria. Microbiology and molecular biology reviews : MMBR* (Vol. 75, pp. 14–49). doi:10.1128/MMBR.00028-10

Newton, R. J., & McMahon, K. D. (2011a). Seasonal differences in bacterial community composition following nutrient additions in a eutrophic lake. *Environmental microbiology*, *13*(4), 887–99. doi:10.1111/j.1462-2920.2010.02387.x

Pernthaler, J., Posch, T., Karel, S., Pernthaler, A., Glöckner, F. O., Psenner, R., Vrba, J. (2001). Predator-Specific Enrichment of Actinobacteria from a Cosmopolitan Freshwater Clade in Mixed Continuous Culture. *Applied and Environmental Microbiology*, *67*(5), 2145–2155. doi:10.1128/AEM.67.5.2145

Polarizing, C., Bidle, K. D., & Azam, F. (1999). Accelerated dissolution of diatom silica by marine bacterial assemblages. *Letters to Nature*, *397*(February), 508–512.

Saba, G. K., Steinberg, D. K., & Bronk, D. a. (2011). The relative importance of sloppy feeding, excretion, and fecal pellet leaching in the release of dissolved carbon and nitrogen by Acartia tonsa copepods. *Journal of Experimental Marine Biology and Ecology*, *404*(1-2), 47–56. doi:10.1016/j.jembe.2011.04.013

Salcher, M. M., Pernthaler, J., & Posch, T. (2010). Spatiotemporal distribution and activity patterns of bacteria from three phylogenetic groups in an oligomesotrophic lake. *Limnology and Oceanography*, *55*(2), 846–856. doi:10.4319/lo.2009.55.2.0846

Salcher, M. M., Pernthaler, J., Frater, N., & Posch, T. (2011). Vertical and longitudinal distribution patterns of different bacterioplankton populations in a canyon-shaped, deep prealpine lake. *Limnology and Oceanography*, *56*(6), 2027–2039. doi:10.4319/lo.2011.56.6.2027

Salcher, M. M., Posch, T., & Pernthaler, J. (2012). In situ substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. *The ISME journal*, 1–12. doi:10.1038/ismej.2012.162

Sarthou, G., Timmermans, K. R., Blain, S., & Tréguer, P. (2005). Growth physiology and fate of diatoms in the ocean: a review. *Journal of Sea Research*, *53*(1-2), 25–42. doi:10.1016/j.seares.2004.01.007

Schloss, P.D. et al., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, 75(23), pp.7537–41. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2786419&tool=pmcentrez&rendertype=abstract [Accessed March 1, 2013]

Sharma, A. K., Sommerfeld, K., Bullerjahn, G. S., Matteson, A. R., Wilhelm, S. W., Jezbera, J., Brandt, U., et al. (2009). Actinorhodopsin genes discovered in diverse freshwater habitats and among cultivated freshwater Actinobacteria. *The ISME journal*, *3*(6), 726–37. doi:10.1038/ismej.2009.13

Sumper, M., Kro, N., & Regensburg, D.-. (2004). Silica formation in diatoms : the function of long-chain polyamines and silaffins. *Journal of Materials Chemistry*, *14*, 2059–2065.

Sumper, M., Brunner, E., & Lehmann, G. (2005). Biomineralization in diatoms: characterization of novel polyamines associated with silica. *FEBS letters*, *579*(17), 3765–9. doi:10.1016/j.febslet.2005.06.001

Tarao, M., Jezbera, J., & Hahn, M. W. (2009). Involvement of cell surface structures in size-independent grazing resistance of freshwater Actinobacteria. *Applied and environmental microbiology*, *75*(14), 4720–6. doi:10.1128/AEM.00251-09

Tseng, W.-C., Fang, T.-Y., Cho, C.-Y., Chen, P.-S., & Tsai, C.-S. (2012). Assessments of growth conditions on the production of cyanophycin by recombinant Escherichia coli strains expressing cyanophycin synthetase gene. *Biotechnology progress*, *28*(2), 358–63. doi:10.1002/btpr.1513

Yannarell, a C., Kent, a D., Lauster, G. H., Kratz, T. K., & Triplett, E. W. (2003). Temporal patterns in bacterial communities in three temperate lakes of different trophic status. *Microbial ecology*, *46*(4), 391–405. doi:10.1007/s00248-003-1008-9

Yannarell, A. C., & Triplett, E. W. (2005). Geographic and Environmental Sources of Variation in Lake Bacterial Community Composition †. *Applied and environmental microbiology*, *71*(1). doi:10.1128/AEM.71.1.227

6. CONCLUSIONS

The analysis of 11 acI single amplified genomes indicated the presence of polyamine uptake genes, a cyanophycinase gene for the breakdown of cyanophycin and uptake genes for the resulting dipeptide breakdown products from cyanophycin. These genes indicated a potential relationship between acI and phytoplankton such as diatoms (rich in polyamines in the silica cell wall) and cyanobacteria (producers of cyanophycin).

Lu et al (2014) indicate that phytoplankton are the major source of polyamines in surface waters. The question that arises is whether it is possible for acI to obtain all of its nitrogen needs from polyamines. A brief analysis of stoichiometry indicates that the acI bacteria would have an excess of nitrogen available if they were to consume putrescine or cyanophycin as a sole source of energy and nitrogen as the C:N:P ratio of freshwater bacteria has been found to be about 75:15:1 (Makino & Cotner, 2004) while the C:N ratio of putrescine and cyanophycin are both about 1.7. This simple analysis assumes that putrescine or cyanophycin are available in abundance and they are the only compounds consumed by acI. A more in depth mass balance analysis is required to determine what percentage of acI's energy and nitrogen needs could be met by the mass of putrescine and cyanophycin produced during phytoplankton blooms. The data in Lu et al (2014) provides the best reference for in situ concentrations of polyamines in natural waters to date and provides a good basis for this analysis.

The findings from the ten-year study of acI in Lake Mendota indicated a lag correlation between acI-A6 and diatoms and between acI-B1 and the Nostocophycideae cyanobacteria. These findings added more weight to the genomic findings that indicated a potential relationship between acI and intracellular components of diatoms and cyanobacteria

which are likely released from dying phytoplankton blooms. However, the 11 genomes indicate that acI-A and acI-B are genetically very similar and share genes for polyamine uptake and cyanophycinase. Based on this genetic evidence we would not necessarily expect acI-A to associate with diatoms while acI-B associated with cyanobacteria since they have similar genes.

Finally, the results of the mesocosm study demonstrated a relationship between putrescine and aspartic acid and acI which added further weight to the link between acI and phytoplankton as putrescine is a polyamine abundant in diatom cell walls and aspartic acid is a breakdown product of cyanophycin, a storage granule produced by cyanobacteria. A brief look at the DNA data shows that both acI-B1 and acI-A6 had relatively higher abundances in the putrescine and aspartic acid mesocosms. This data needs to be reviewed further as the tribe level identifications may not be reliable. However, if the data holds, it would indicate that acI-B1 and acI-A6 may not have niches focused on specific phytoplankton but rather they both benefit from aspartic acid and putrescine that may be released from cyanobacteria and diatoms.

6.1. REFERENCES

Lu, X., Zou, L., Clevinger, C., Liu, Q., Hollibaugh, J., Mou, X. (2014). Temporal dynamics and depth variations of dissolved free amino acids and polyamines in coastal seawater determined by high-performance liquid chromatography. *Marine Chemistry*. Vol 163, pp: 36-44.

Makino, W., Cotner, J. B. (2004). Elemental stoichiometry of a heterotrophic bacterial community in a freshwater lake: implications for growth- and resource-dependent variations. *Aquatic Microbial Ecology*. Vol: 34, pp: 33-41.

7. FUTURE DIRECTIONS

The dissertation work presented here has revealed some promising clues about the ecology and role of acI in freshwater systems such as an apparent link with decaying phytoplankton. However, the limitations of the methods used in this work leave some opportunities for further research on acI.

A metatranscriptomics study could be incredibly powerful and help to answer many questions about acI. This study could include bi-weekly samples from an autochthonous lake (i.e. Mendota) and an allochthonous lake (i.e. Crystal Bog) from February (before the first phytoplankton bloom) through November (after the final phytoplankton bloom). We might hypothesize that acI would express genes for the uptake of polyamines after the spring diatom bloom and cyanophycinase after the summer cyanobacteria bloom. We might also hypothesize that actinorhodopsin expression would peak in the middle of summer between the major phytoplankton blooms when organic carbon is more scarce. Some other questions we may answer with metatranscriptomics are:

- Which carbon substances are consumed by acI?

- How much of acI's energy needs are provided by decaying phytoplankton?

- Does acI go dormant between blooms? If not, why does it's population have such little variation in the absence of its food source?

- Does acI switch to using actinorhodopsin to harness solar energy? How exactly does it use solar energy?

The link between acI and phytoplankton should be explored more experimentally to definitively demonstrate that acI does in fact consume carbon released from decaying phytoplankton. This work should also explore the response of acI from various species of phytoplankton (i.e. *Stephaniscus parvum, Aphanizomenon, Cyclotella, Microcystis*). We might hypothesize that acI benefits from most of the dominant freshwater phytoplankton since acI is so abundant in most freshwater systems. However, it is possible that certain tribes of acI (i.e. acI-B1) benefit more from specific phytoplankton (i.e. *Aphanizomenon*). This might explain why we see acI tribes partition based on lake pH. For example, maybe a higher lake pH results in *Aphanizomenon* as the dominant Cyanobacteria which specifically benefits acI-B1. This work could be done by incubating several distinct phytoplankton cultures with radiolabelled or isotopic carbon that is incorporated into the organic carbon molecules (i.e. putrescine, cyanophycin, etc) during growth. A portion of the labeled-carbon phytoplankton cultures could then be killed and dosed as substrate to lake water similar to the mesocosm study performed in this dissertation. The filtered lake water should be screened to verify the absence of grazing organisms if we intend to study the bottom-up drivers (i.e. substrate niche competitiveness) of acI rather than the top-down drivers (i.e. grazing resistance). The acI populations should be tracked using fluorescent in-situ hybridization (FISH) for quantitative results rather than the relative abundance produced by 16S sequencing, though 16S sequencing could be a helpful compliment as it provides very detailed data on the entire bacterial community.

Further research should also investigate acI bacteria in lakes that are dominated by allochthonous carbon. Do we find acI-B1 and acI-A6 in abundance in these lakes? Based on the findings in this dissertation it seems likely that these tribes would not be abundant in

humic lakes. Are there other acI bacteria that are abundant in humic lakes? If so, it would be interesting to see how the genomes compare with acI-B1 and acI-A6, which are known to be abundant in autochthonous lakes. If acI are abundant in humic lakes, what are the primary substances they are likely consuming? Are polyamines and oligopeptides abundant in these systems?

We have learned much about acI in the past few years but much about these important bacteria remains a mystery. Fortunately, the genetic tools available to study these bacteria have become vastly more powerful in the past couple years due primarily to advances in next-generation DNA sequencing, bioinformatics and transcriptomics techniques. The next few years of research may result in the development of a draft metabolic/ecological model of acI that estimates the energy and nutrient inputs to acI across various seasons and identifies the sources of these substrates. This result would be truly impressive considering the acI bacteria remains to be isolated which has limited the application of traditional microbiolocial research techniques.