npg

ORIGINAL ARTICLE

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

Trevor W Ghylin[1], Sarahi L Garcia[2,3], Francisco Moya[1], Ben O Oyserman[1], Patrick Schwientek[4], Katrina T Forest[2], James Mutschler[2], Jeffrey Dwulit-Smith[2], Leong-Keat Chan[4], Manuel Martinez-Garcia[5,10], Alexander Sczyrba[4,6], Ramunas Stepanauskas[5], Hans-Peter Grossart[7,8], Tanja Woyke[4], Falk Warnecke[3,*], Rex Malmstrom[4], Stefan Bertilsson[9] and 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]*Jena School for Microbial Communication (JSMC) and Microbial Ecology Group at Friedrich Schiller University Jena, Jena, Germany;* [4]*DOE Joint Genome Institute, Walnut Creek, CA, USA;* [5]*Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, USA;* [6]*Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany;* [7]*Department of Limnology of Stratified Lakes, Leibniz-Institute for Freshwater Ecology and Inland Fisheries, Stechlin, Germany;* [8]*Institute for Biochemistry and Biology, Potsdam University, Potsdam, Germany and* [9]*Department of Ecology and Genetics, Limnology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden*

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 of acI SAGs and other available freshwater bacterial genomes showed that acI has more gene content directed toward carbohydrate acquisition as compared to *Polynucleobacter* and LD12 *Alphaproteobacteria*, which seem to specialize more on carboxylic acids. The acI genomes contain actinorhodopsin as well as some genes involved in anaplerotic carbon fixation indicating the capacity to supplement their known heterotrophic lifestyle. 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 sunlight and N-rich organic compounds such as polyamines, di- and oligopeptides, branched-chain amino acids and cyanophycin. This work significantly expands the known metabolic potential of the cosmopolitan freshwater acI lineage and its ecological and genetic traits.
The ISME Journal (2014) **8,** 2503–2516; doi:10.1038/ismej.2014.135; published online 5 August 2014

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 (Allgaier and Grossart, 2006; Salcher *et al.*, 2010; Newton *et al.*, 2011; Eckert *et al.*, 2012; Rösel and Grossart, 2012). Due to their high abundance and high metabolic activity (Warnecke *et al.*, 2005; Allgaier and Grossart, 2006; Salcher *et al.*, 2010), the acI bacteria likely have a critical role in carbon and other nutrient cycling in freshwater systems, yet their genomic features

Correspondence: KD McMahon, Department of Civil and Environmental Engineering, and Bacteriology, University of Wisconsin 1550, Linden Drive Room 5552, Madison, WI 53706, USA.
E-mail: kdmcmahon@wisc.edu
[10]Current address: University of Alicante, Alicante, Spain.
*Deceased.

remain elusive. This is in part due to the historical lack of isolated representatives and limited genomic data.

On the basis of 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 (for example, 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 and microautoradiography (MAR). These studies have shown that acI members can take up glucose (Buck *et al.*, 2009), leucine (Buck *et al.*, 2009; Perez *et al.*, 2010; Salcher *et al.*, 2010; Eckert *et al.*, 2012), acetate (Buck *et al.*, 2009), thymidine (Perez *et al.*, 2010), *N*-acetylglucosamine (NAG) (Beier and Bertilsson, 2011; Eckert *et al.*, 2012) and di-NAG (Beier and Bertilsson, 2011; Eckert *et al.*, 2012; Tada and Grossart, 2014). 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 *et al.*, 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 *et al.* 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.* (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 10 additional acI single-amplified genomes (SAGs) from three different acI tribes and four different lakes. The additional genomes allowed us to confirm key gene functions and genomic features (for example, GC content) that are conserved within the lineage while also revealing clade-specific differentiation. 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. This kind of analysis was not possible with only one draft genome (the previously published acI-B1 SAG called SCGC AAA027-L06). 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.

Materials and methods

SAG generation and selection
Water samples (1 ml) were collected from the upper 0.5 to 1 m of each of four lakes (Mendota, Sparkling, Damariscotta, Stechlin) and cryopreserved (Supplementary Table S1), as previously described (Martinez-Garcia *et al.*, 2012; Garcia *et al.*, 2013) (Supplementary Online Material). Bacterial 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 (Ludwig *et al.*, 2004; Newton *et al.*, 2011). 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%

Table 1 AcI SAG metadata and overall features

Full SAG ID	IMG taxon OID	Tribe	Estimated genome completeness (%)[a]	Lake	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	1282
SCGC AB141-P03	2236876028	acI-B1	34	STE	5/25/10	0.7	2.0	41	66	736

Abbreviations: DAM, Damariscotta Lake (Mesotrophic); MEN, Lake Mendota (Eutrophic); SAG, single-amplified genome; SPA, Sparkling Lake (Oligotrophic); STE, Lake Stechlin (Oligotrophic).
[a]Estimated based on the presence of single-copy genes found in other Actinomycetales, as described in Materials and methods.

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.

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 10 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 ×30 were present after filtering. Contigs with an average k-mer depth of less than ×2 were removed. The following steps were then performed for assembly: (1) filtered Illumina reads were assembled using Velvet version 1.1.04 (Zerbino and 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) non-redundant database, UniProt, TIGRFam, Pfam, KEGG, Clusters of Orthologous Group (COG) and InterPro databases. The tRNAScan-SE tool (Hacker and 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).

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 (Supplementary 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
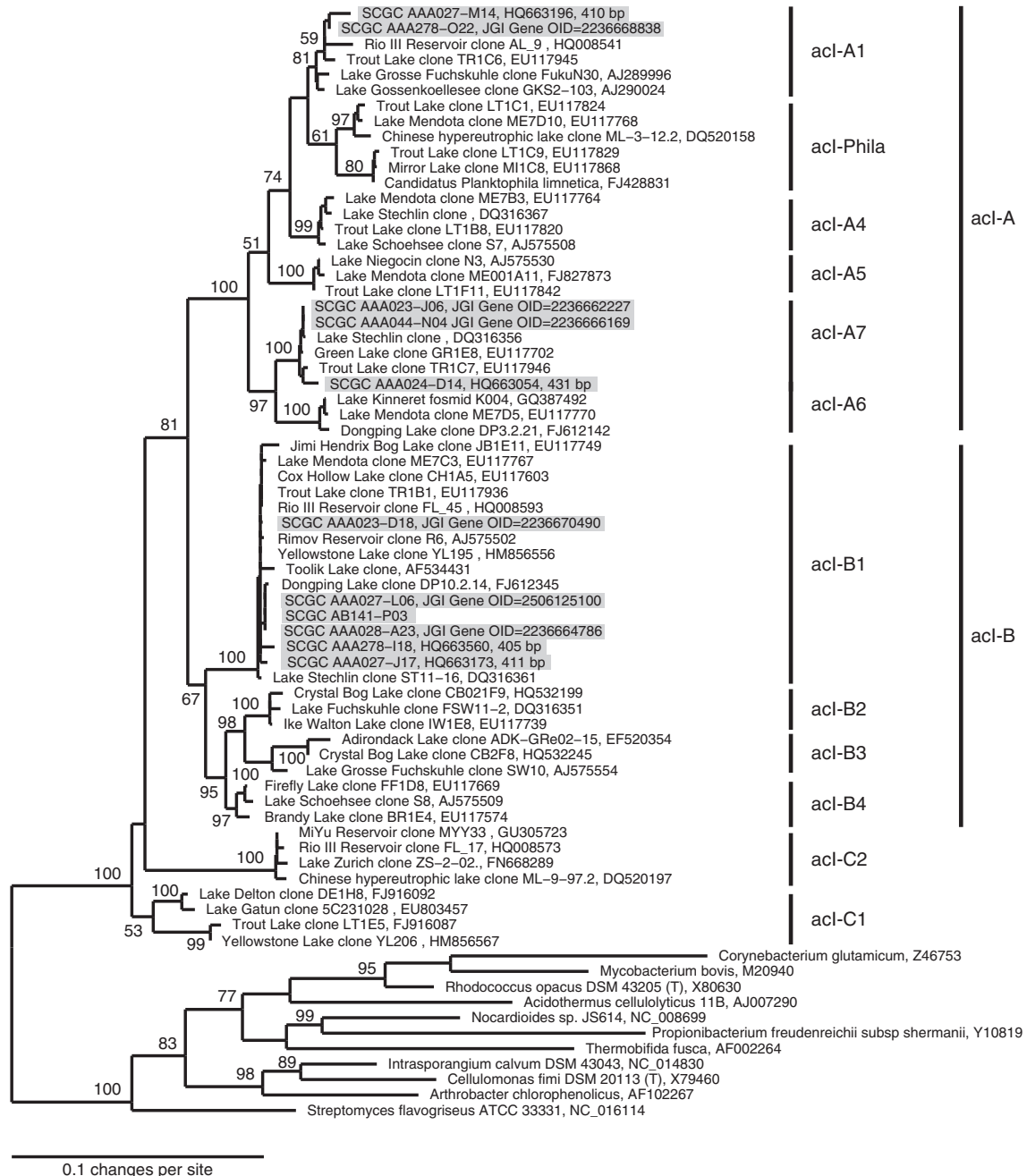
Figure 1 Phylogenetic placement of the SAGs within the acI lineage and relative to other sequenced actinobacterial genomes, and to the previously sequenced AAA027-L06 SAG based on nearly full-length 16S rRNA gene sequences. When only short (~400 bp) amplified gene fragments were available (Martinez-Garcia *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.

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

Comparative analysis
Annotations with the COGs of proteins for each SAG and comparison genome from the Actinomycetales order, *Polynucleobacter* genus, and freshwater *Alphaproteobacteria* LD12

tribe were downloaded from the IMG website (http://img.jgi.doe.gov/; Supplementary 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 divided by % 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 overrepresented in acI.

Average protein identity 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.

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 overall genes in the genome (Rocha, 2004). The CAI of a gene was computed as described by (Sharp and Li, 1987).

Results and discussion

Genome statistics and phylogenetic affiliation
We sequenced ten acI SAGs and analyzed them along with the one previously sequenced SAG (Garcia et al., 2013). These 11 SAGs belong to three different acI tribes and originate from four different lakes. A phylogenetic tree of these SAGs is presented in Figure 1. Our genome completeness estimates ranged from 34% to a nearly complete genome. Only four of the 158 Pfams used for genome completion estimates were absent in all 11 SAGs (Supplementary Table S2). Back calculating the estimated genome size of the SAGs, it could be observed that all 11 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 ($\sim$42%) and small genome size ($\sim$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 (Table 1 and Supplementary Table S4). The average genome completeness 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.

acI metabolism: how does the acI lineage make a living in freshwater?
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 ten newly 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 MAR-FISH-based studies clearly demonstrating the ability of acI to incorporate glucose (Buck et al., 2009; Salcher et al., 2013). However, most SAGs did contain ABC-type sugar transport components (for example, COG1653) (Table 2 and Supplementary 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 11 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, as 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). 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 MAR-FISH-based studies (Buck et al., 2009; Salcher et al., 2013).

Most SAGs also contained a putative cyanophycinase (COG4242), allowing acI to access this C- and N-storage compound synthesized by cyanobacteria. 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 (Krehenbrink et al., 2002; Fueser and Steinbuechel, 2007). The acI SAGs analyzed here have the genes required to break down cyanophycin granules and take up the resulting dipeptides and amino acids as a source of energy and N. However, it is unclear whether this cyanophycinase can be secreted, as we were unable to identify an obvious secretion signal peptide. Secreted cyanophycinases have been reported in both Gram negative and Gram positive soil bacteria
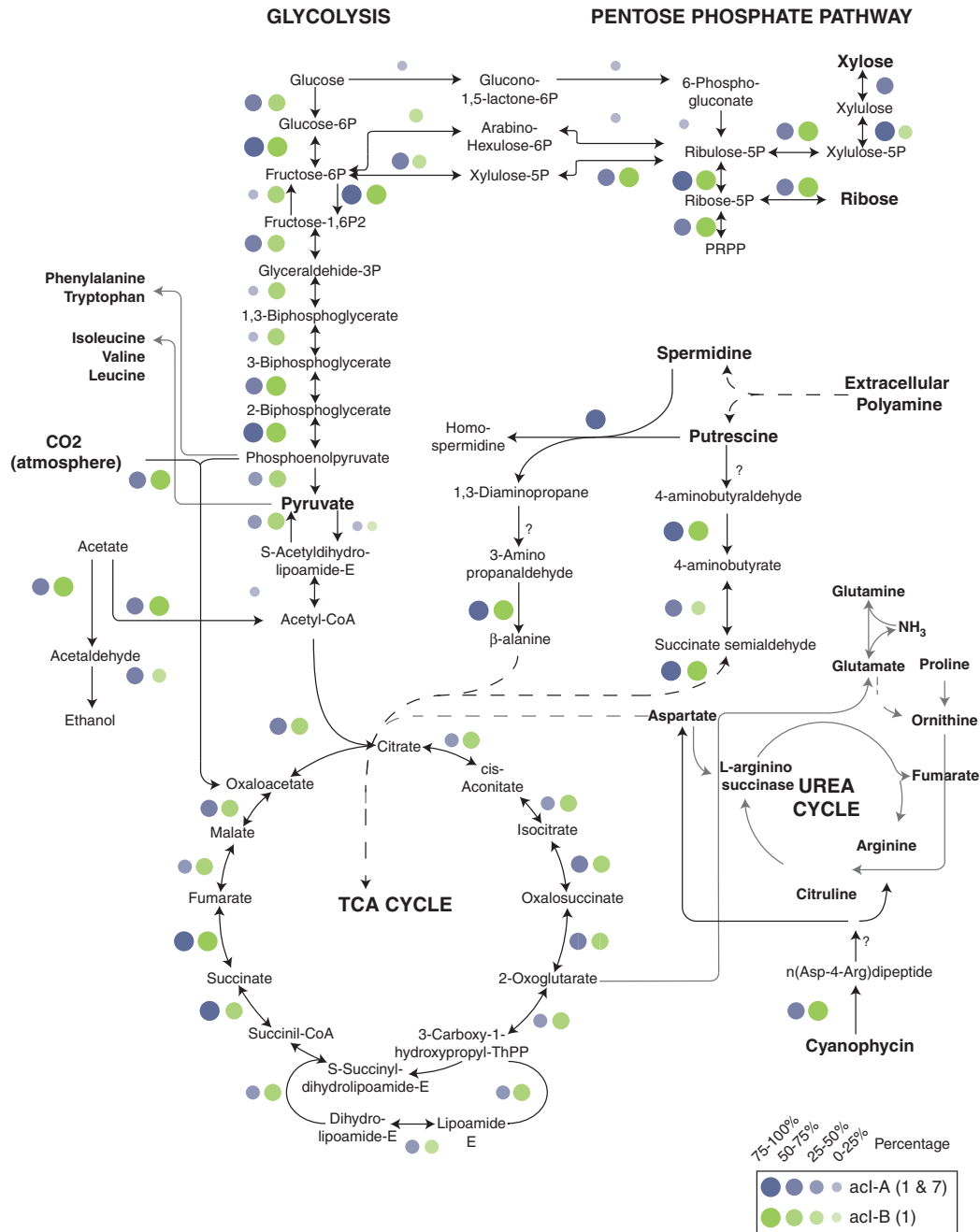
Figure 2 Central carbon metabolism and other relevant metabolic pathways identified in acI SAGs. Circles denote genes encoding the necessary enzymes are present in that clade (blue for acI-A and green for acI-B), with the size of the circle being proportional to the percentage of SAGs within that clade that were found to contain that gene. The presence and absence of genes was determined within the IMG environment based on KEGG annotations.

(Obst et al., 2002, 2004; Sallam et al., 2011), but in these cases signal sequence cleavage was demonstrated or inferred. 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) and acI, whereby acI members acquire energy, C and N from this polymer synthesized by others. Curiously, we found two tandem copies of cyanophycin synthetase in each of the two *Polynucleobacter* genomes but not in acI. This has implications for

Table 2 Select DOC uptake COGs in acI, LD12, PnecC and Actinomycetales[a]

Category	Clusters of Orthologous Groups (COGs)	Percent of genomes containing that COG[b]			
		acI (11)	LD12 (10)	PnecC (2)	Actm (424)
Amino Acids, general	COG0531 Amino acid transporters	100%	0%	100%	90%
	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%	100%	28%
Nucleotides & Coenzymes	COG3201 Nicotinamide mononucleotide transporter	100%	0%	0%	71%
	COG2233 Xanthine/uracil permeases	27%	0%	0%	67%
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%	50%	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%
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%
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%
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%	100%	36%
	COG2182 Maltose-binding periplasmic proteins/domains	18%	0%	0%	42%
	COG3833 ABC-type maltose transport systems, permease component	9%	0%	100%	41%

Abbreviations: Actm, non-acI Actinomycetales; Carbs, carbohydrates; COG, Clusters of Orthologous Group; DOC, dissolved organic carbon; PnecC, *Polynucleobacter necessarius asymbioticus*.
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 Supplementary Table S5.
[a]DOC COG list based on (Poretsky *et al.* 2010). [b]LD12, a cosmopolitan freshwater *Alphaproteobacteria* tribe that is a sister group to *Pelagibacter ubique* and SAR11. Number of genomes included in the analysis is provided in parentheses.

potential interactions between acI and *Polynucleobacter* when they co-occur (that is, *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.*, 2014). However, this kind of interaction is only speculative until experimental confirmation can be made.

Others have found that acI abundances change through the seasons with maxima in spring and fall (Allgaier and 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 and provide the potential for phototrophy (Sharma *et al.*, 2008, 2009; Martinez-Garcia *et al.*, 2012; Garcia *et al.*, 2013; Salka *et al.*, 2014). We confirmed the presence of actinorhodopsin in all but three of the eleven acI SAGs (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 or its precursors were identified in 10 of the 11 acI SAGs, suggesting the likely assembly of functional rhodopsin *in vivo*. In the acI SAGs, the four enzymes that lead from farnesyl pyrophosphate, the ubiquitous sterol precursor, to β-carotene, the bicyclic conjugated chromophore, (Martinez *et al.*, 2007) are most commonly encoded in a cluster. The final pathway enzyme, a β-carotene cleaving oxygenase, was somewhat surprisingly sequenced only in two of the 11 SAGs, and not co-located with the rhodopsin gene as it is in many marine proteobacteria (Martinez *et al.*, 2007; Riedel *et al.*, 2013; Vollmers *et al.*, 2013). However, one of these SAGs places the oxygenase near a β-carotene pathway gene, likely indicating that it once composed part of a full pathway operon. The genes necessary for synthesis of a carotenoid glycoside ester sit just upstream of the cluster, suggesting a secondary chromophore for actinorhodopsin similar to the salinixanthin discovered hydrophobically clinging to xantho- and *Gleobacter* rhodopsins (Luecke *et al.*, 2008; Imasheva *et al.*, 2009). At the protein level, a change of bacteriorhodopsin's tryptophan 138 to a glycine provides space for the four-keto ring of such a putative secondary chromophore, and six of the hydrophobic amino acid side chains that would interact with this putative rhodopsin-hugging photon-funnel are shared between the actinorhodopsins and xanthorhodopsin (Balashov *et al.*, 2010). This has exciting implications for the role of a chromophore other than retinal in members of the

acI lineage, though further experimental work is needed to confirm whether such a chromophore is actually synthesized by acI.

Rhodopsin in *Pelagibacter ubique* and other marine bacteria promotes survival during nutrient starvation periods (Gomez-Consarnau *et al.*, 2007; DeLong and Beja, 2010; Gómez-Consarnau *et al.*, 2010; Steindler *et al.*, 2011); thus actinorhodopsin may serve a similar function in acI. In such an auxiliary system, it remains to be explored why there is not co-localization of the rhodopsin gene, the oxygenase gene and the core carotenoid synthesis operon, in acI genomes. 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 carboxylase). The acI SAGs lack the RuBisCO enzyme and other pathways for carbon fixation, but phosphoenolpyruvate carboxylase may provide them with the ability to synthesize oxaloacetate, an intermediate in the tricarboxylic acid cycle 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. Notably, a marine Bacteroidetes (*Polaribacter*) lacking RuBisCO but harboring proteorhodopsin has been shown to increase CO_2 fixation rates in the light (González *et al.*, 2008).

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 11 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, as it seems to obtain both C and N 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). Indeed, genes involved in acquisition of these N-rich compounds also generally have high codon bias in the SAGs (Supplementary Table S6), indicating possibly high expression rates. Some prior research has focused on organic-N concentrations

or sources for freshwater bacterioplankton. Buck et al. (2009) used MAR-FISH to test glucose, leucine and acetate assimilation in freshwater. They found that Actinobacteria had the highest percentage of cells incorporating leucine but not glucose or acetate, indicating a preference for nitrogen-rich organics. Salcher et al. (2013) also found that acI was active in leucine uptake. Some MAR-FISH studies have suggested that another nitrogen-rich compound, NAG may be an important acI substrate (Beier and Bertilsson, 2011; Eckert et al., 2012, 2013). However, we could not identify machinery to support the ability to take up and incorporate this substance. This may well simply be due to the fact that the biochemical 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).

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 was found in eight of the eleven SAGs. AcI's apparent reliance on reduced sulfur compounds is an interesting similarity to that of the ubiquitous marine SAR11 clade, which was the first free-living bacteria recognized to rely on reduced sulfur for growth (Grote et al., 2012; Carini et al., 2013).

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 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.

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 Clusters of Orthologous Groups (COGs) for acI versus other Actinomycetales (424 genomes) and freshwater bacteria such as LD12 *Alphaproteobacteria*

relative of the marine SAR11 (10 genomes) and *Polynucleobacter sp.* (two genomes; Table 2 and Supplementary Table S5). 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 10 LD12 genomes did not contain any of these COGs. Other interesting COGs that were more common in acI were nucleotide transporters, several sugar transporters and two ribose/xylose transporters. Overall, acI had more COGs for general carbohydrate transport than LD12 and *Polynucleobacter*.

As described above, the acI genomes lacked carboxylic acid and dicarboxylic acid transporters while they were present in *Polynucleobacter* and LD12. Compared with the other freshwater bacteria, general amino acid uptake COGs were underrepresented in acI SAGs and COGs for di- and oligopeptides were overrepresented. The di- and oligo-peptide transporters may have some connection with the cyanophycinase genes found in acI (and not in LD12 or *Polynucleobacter*). COGs involved in lipid transport were underrepresented in acI (as compared with LD12 and *Polynucleobacter*), 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. AcI genomes indicate specialization in polyamines, di- and oligopeptides and carbohydrates (including pentoses). LD12 genomes indicate specialization in lipids and carboxylic acids, whereas *Polynucleobacter* specializes in general carbohydrates and carboxylic acids.

Comparative genomics hints at 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 (for example, separation of acI-A vs acI-B) (Figure 3, Supplementary Table S7). It should be noted that this analysis is preliminary due to the incompleteness of the SAGs. Little is known about the ecology of the acI tribes as most previous studies lacked phylogenetic resolution to the tribe level. However, Newton et al. (2007) found that acI tribes differentiated based on pH with acI-B2 and acI-A1 correlating with acidic lakes and acI-B1 and acI-A6 correlating with alkaline lakes. The ratio of terrestrially sourced carbon to in-lake-derived carbon also seems to predict tribe distribution, with acI-B1, -A6, and -A7 generally preferring lakes with lower color-to-chlorophyll (CtCh) ratios as compared with other tribes (Jones et al., 2009). This ratio is calculated using measurements of chromophoric dissolved
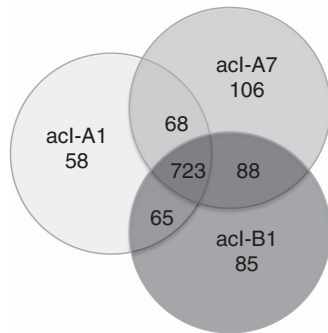
Figure 3 Distribution of 1193 COGs found in the acI lineage based on the analysis of 11 SAGs. Shared and differential COG content among acI tribes within the clades based on the comparison of two acI-A1 SAGs, three acI-A7 SAGs and six acI-B1 SAGs. Markedly, more COGs are shared among acI tribes than are unique to any particular tribe. We note that the probability of an acI 'shared COG set' gene being missing from all the genomes due to incomplete genome recovery is 0.000003% for all of acI, 0.003% in acI-B1, 2.7% in acI-A1 and 4.1% in acI-A7, for any single COG.

organic matter (by light absorbance at 440 nm) and chlorophyll-a concentrations, making it a useful proxy for the extent of allochthony of a lake ecosystem. Interestingly, acI-A1 is found in lakes with both low and high CtCh. As the tribes included in the current study are usually found in lakes with low CtCh, it remains to be learned whether other tribes (for example, acI-B2, acI-A4, acI-A5) have different gene content that relates to life at high CtCh.

Members of the acI-A clade contained markedly more unique COGs than members of the acI-B clade (compare Supplementary Table S7 with Supplementary Table S8), suggesting they are more metabolically versatile. However, this might be explained by the fact that two acI-A tribes are represented in our SAG collection as compared with one acI-B tribe. Nevertheless, we found that the alpha and beta-galactosidases were common to acI-A1 and acI-A7, but were absent in all acI-B1 SAGs (Supplementary Table S7). The presence of these genes in acI-A SAGs suggests a specialization in breaking down poly- or oligo-saccharides as compared with acI-B. Interestingly, an over-representation of glycosidases such as alpha and beta-galactosidases was found in freshwater systems compared with marine systems (Eiler et al., 2013). This effect may be largely due to the presence of the ubiquitous acI-A clade in freshwater. Zakharova et al. (2013) studied bacteria associated with diatom degradation in the near bottom water layer of Lake Baikal and found high beta-galactosidase activity. They also found that Actinobacteria were a significant portion of the bacterial community. 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 NAG kinase and oligopeptide transporter not found in acI-A1 or acI-B1. Both clades share what appears

to be an ABC transport system for ribose (COG1172), but the acI-B1 genomes uniquely carried an auxillary component (ribose pyranase - COG1869) that may give this tribe access to certain pentose isomers (Supplementary Table S8). The acI-B1 SAGs did carry an aromatic ring-cleaving dioxygenase (DOPA 4,5-dioxygenase) that was not found in acI-A or any other Actinomycetales (COG3805) but is commonly found in fungal genomes. Homologs of this protein are key for the biosynthesis of pigments such as betalains (red and yellow pigments) in plants, suggesting the role of a unique pigment in acI-B1.

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 0.000003% in acI, 0.003% in acI-B1, 2.7% in acI-A1 and 4.1% in acI-A7, for any single COG. These probabilities were calculated by multiplying the probability of a gene missing in each genome (that is, same as percent of genome incompleteness). For example, the acI-A1 genome SCGC AAA027-M14 is only 41% complete, therefore the probability of a 'shared COG set' gene missing from the genome is 59%. The probability of the same gene missing from the acI-A1 genome SCGC AAA278-O22 is 4% and the probability of the gene missing from both genomes due to incompleteness is 2.7% (that is, 59 × 4%). 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 4.1% with the sequencing of more genomes.

The average protein similarity between the acI SAGs indicated diversity within clades and a range in variation within tribes, at the protein-coding gene level (Supplementary Table S9). The average amino acid identity across tribes (65–70% acI-A1 vs acI-A7) indicate that the acI-A clade is similar to a traditional taxonomic genus, which has been estimated to have 65%–70% average protein identity (Konstantinidis and Tiedje, 2007). The comparison of amino acid identity across clades (60–65% amino acid identity acI-B1 vs acI-A1 and acI-A7) indicates that the acI lineage is comparable to a traditional taxonomic family.

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.

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

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. Although 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. Confirmation of the characteristics discovered

Table 3 COGs overrepresented in acI as compared withwith non-acI Actinomycetales (Actm)

	% 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

Abbreviation: COG, Clusters of Orthologous Group.
Cells are shaded according to magnitude of overrepresentation factor, to aid in interpretation of numbers.

in AAA027-L06 within multiple SAGs representing three tribes provided more conclusive evidence of features conserved within the lineage. We identified features that hint at a preference for N-rich compounds (polyamines, di- and oligopeptides, branched-chain amino acids, cyanophycin) 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.

Notes

This article is dedicated to the memory of one of our co-authors, Falk Warnecke, who passed away just before its publication.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

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.

Balashov SP, Imasheva ES, Choi AR, Jung K-H, Liaaen-Jensen S, Lanyi JK. (2010). Reconstitution of gloeobacter rhodopsin with echinenone: role of the 4-keto group. *Biochemistry* **49**: 9792–9799.

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.

Carini P, Steindler L, Beszteri S, Giovannoni SJ. (2013). Nutrient requirements for growth of the extreme oligotroph "Candidatus Pelagibacter ubique" HTCC1062 on a defined medium. *ISME J* **7**: 592–602.

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**: 1966–1975.

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**: 2265–2272.

DeLong EF, Beja O. (2010). The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. *PLoS Biol* **8**: e1000359.

Eckert EM, Baumgartner M, Huber IM, Pernthaler J. (2013). Grazing resistant freshwater bacteria profit from chitin and cell-wall derived organic carbon. *Environ Microbiol* **15**: 2019–2030.

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. (2014). Successful enrichment of the ubiquitous freshwater acI Actinobacteria. *Environ Microbiol Rep* **6**: 21–27.

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 USA* **108**: 1513–1518.

Gómez-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**: e1000358.

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.

González JM, Fernández-Gómez B, Fernàndez-Guerra A, Gómez-Consarnau L, Sánchez O, Coll-Lladó M et al. (2008). Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria). *Proc Natl Acad Sci USA* **105**: 8724–8729.

Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ et al. (2012). Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *MBio* **3**: e00252–12.

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**: 119.

Imasheva ES, Balashov SP, Choi AR, Jung K. (2009). Reconstitution of *Gloeobacter violaceus* rhodopsin with a light-harvesting carotenoid antenna. *Biochemistry* **48**: 10948–10955.

Jones SE, Newton RJ, McMahon KD. (2009). Evidence for structuring of bacterial community composition by organic carbon source in temperate lakes. *Environ Microbiol* **11**: 2463–2472.

Konstantinidis KT, Tiedje JM. (2007). Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr Opin Microbiol* **10**: 504–509.

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.

Luecke H, Schobert B, Stagno J, Imasheva ES, Wang JM, Balashov SP et al. (2008). Crystallographic structure of xanthorhodopsin, the light-driven proton pump with a dual chromophore. *Proc Natl Acad Sci USA* **105**: 16561–16565.

Magnuson JJ, Kratz TK, Benson BJ. (2006). Long term dynamics of lakes in the landscape, Oxford University Press: New York, USA.

Makarova KS, Aravind L, Galperin MY, Grishin NV, 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–D122.

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 USA* **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**: 7169–7176.

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–D301.

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* **7**: 896–907.

Salka I, Wurzbacher C, Garcia SL, Labrenz M, Jurgens K, Grossart HP. (2014). Distribution of acI-Actinorhodopsin genes in Baltic Sea salinity gradients indicates adaptation of facultative freshwater photoheterotrophs to brackish waters. *Environ Microbiol* **16**: 586–597.

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.

Stamatakis A, Hoover P, Rougemont J. (2008). A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Syst Biol* **57**: 758–771.

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**: e19725.

Tada Y, Grossart H-P. (2014). Community shifts of actively growing lake bacteria after N-acetyl-glucosamine addition: improving the BrdU-FACS method. *ISME J* **8**: 441–454.

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.

Zakharova YR, Kurilkina MI, Likhoshvay AV, Shishlyannikov SM, Kalyuzhnaya OV, Petrova DP *et al.* (2013). Effect of bacteria from the bottom water layer of lake Baikal on degradation of diatoms. *Paleontol J* **47**: 1030–1034.

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