ORIGINAL ARTICLE

Actinorhodopsin genes discovered in diverse freshwater habitats and among cultivated freshwater *Actinobacteria*

Adrian K Sharma[1], Katrin Sommerfeld[1], George S Bullerjahn[2], Audrey R Matteson[3], Steven W Wilhelm[3], Jan Jezbera[4,5], Ulrike Brandt[4], W Ford Doolittle[1] and Martin W Hahn[4]

[1]*Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada;* [2]*Department of Biological Sciences, Bowling Green State University, Bowling Green, OH, USA;* [3]*Department of Microbiology, The University of Tennessee, Knoxville, TN, USA;* [4]*Institute for Limnology, Austrian Academy of Sciences, Mondsee, Austria and* [5]*Institute of Hydrobiology, Biology Centre of the AS CR, v. v. i., České Budějovice, Czech Republic*

Microbial rhodopsins are membrane proteins that utilize a retinal chromophore to harvest sunlight for energetic and photosensory functions. Recently, a group of novel rhodopsin sequences named 'actinorhodopsins' (ActRs) was hypothesized to exist among uncultured planktonic *Actinobacteria*. ActRs were discovered by mining metagenomic data obtained during the Venter Institute's Global Ocean Sampling expedition, from a hypersaline lagoon, two estuaries and a freshwater lake. On the basis of these findings, and many studies that show *Actinobacteria* are common inhabitants of lakes, we predicted that ActR genes would likely be present in other freshwater habitats and among the genomes of cultivated *Actinobacteria*. Using degenerate polymerase chain reaction primers, we discovered an ActR gene present in an actinobacterial isolate of the family Microbacteriaceae. Isolate MWH-Uga1 was cultivated prior to this study from a freshwater pond in Uganda and belongs to a group of *Actinobacteria* previously identified in freshwater ecosystems. ActR genes were also discovered present in numerous mixed cultures containing freshwater *Actinobacteria* and among environmental DNA samples obtained from three freshwater sources; a small woodland pond and the Laurentian Great Lakes Superior and Erie. An analysis of small subunit ribosomal RNA genes from metagenomic DNA samples harboring ActR genes suggests that organisms belonging to the acI lineage, an uncultured group of *Actinobacteria* commonly present in fresh waters, may utilize rhodopsins. The co-occurrence of an acI organism with a specific ActR variant in a mixed culture supports our hypothesis.

Introduction

Microbial rhodopsins are membrane proteins that use a retinal chromophore to harvest solar energy for ion pumping and photoreception. In 2000, Beja *et al.* discovered rhodopsin-encoding genes in the surface waters of the earth's oceans and designated this novel group of proteins 'proteorhodopsin'. Their designation was based on the discovery of metagenomic fragments that linked the first discovered of these proteorhodopsin genes to a small subunit ribosomal RNA (SSU rRNA) gene defining the

Correspondence: AK Sharma, Department of Biochemistry and Molecular Biology, Dalhousie University, 5850 College Street, Halifax, Nova Scotia, Canada B3H1X5.
E-mail: aksharma@dal.ca

uncultivated marine SAR86 group of the γ-*Proteobacteria* (Beja *et al.*, 2000). Since then, proteorhodopsin genes have been found in high abundance in the photic zone of the ocean, where they are believed to confer proteorhodopsin-based phototrophy and light-sensing abilities on diverse groups of marine microorganisms (Rusch *et al.*, 2007; Campbell *et al.*, 2008; Fuhrman *et al.*, 2008).

Recently, proteorhodopsin-related genes were discovered through metagenomic analysis of several non-marine aquatic environments sampled during the Venter Institute's Global Ocean Sampling (GOS) expedition (Rusch *et al.*, 2007). Rusch *et al.* described the presence of these genes among microbes inhabiting the freshwater Lake Gatun (part of the Panama Canal), the estuaries Delaware Bay and Chesapeake Bay, and the hypersaline Punta Cormorant lagoon in Ecuador. In a separate analysis of rhodopsin

sequences from these non-marine environments in Sharma et al. (2008), we described three novel groups of sequences (LG1, LG2 and PCL1, Figure 1) comprising the majority of rhodopsin diversity. Our earlier analyses also consisted of a bioinformatic assessment of GOS-derived (Rusch et al., 2007) metagenomic assemblies containing rhodopsin genes of these three novel clades and found they were almost exclusively linked to open reading frames most closely related to homologues encoded in sequenced actinobacterial genomes. In Sharma et al. (2008) we designated these novel clades of rhodopsins collectively as 'actinorhodopsins (ActRs)' based on (1) the aforementioned pattern of gene linkage on GOS metagenomic assemblies, (2) the observation by Rusch et al. (2007) that actinobacterial SSU rRNA sequences were abundant in non-marine samples and (3) the knowledge that Actinobacteria are common inhabitants of freshwater ecosystems (Warnecke et al., 2004). However in our previous study, we were unable to identify the specific actinobacterial carriers of ActRs from GOS samples. On the basis of our findings in Sharma et al. (2008), we predicted that ActR gene sequences would likely be found in other freshwater environments and in the genomes of cultivated Actinobacteria. To test these predictions, we designed degenerate primers and used the polymerase chain reaction (PCR) to amplify ActR genes from freshwater metagenomic DNA samples and Actinobacteria cultivated from various freshwater sources.

Methods

Sampling and DNA preparation

Point Pleasant Park pond in Halifax, NS was sampled in late summer (early September) 2006. Pond water was collected in a sterile 1 l glass bottle and prefiltered through Whatman paper. Biomass was concentrated onto a 0.2 μm filter using a disposable Nalgene bottle top filter apparatus and the filter was frozen at −20 °C overnight, then the following day DNA was isolated by bead beating a slice (∼1/8) of the 0.2 μm filter using the FastDNA SPIN for Soil kit (MP Biomedicals, LLC., Solon, OH, USA). Lake Superior biomass was obtained during a research cruise on the R/V Blue Heron in June 2005 during isothermal mixing of the water column. Water was collected from a depth of 25 m at station Castle Danger 1 (CD1; 47.30°N, 91.67°W; see Ivanikova et al. (2007) for sampling details). Biomass was collected from the central basin of Lake Erie (Environment Canada Station 84, 41° 56′ 46″ N, 81° 38′ 46″ W) during a July 2005 cruise aboard the CCGS Limnos when the water column was thermally stratified (for sampling details see Wilhelm et al., 2006). DNA for rhodopsin PCR was integrated from
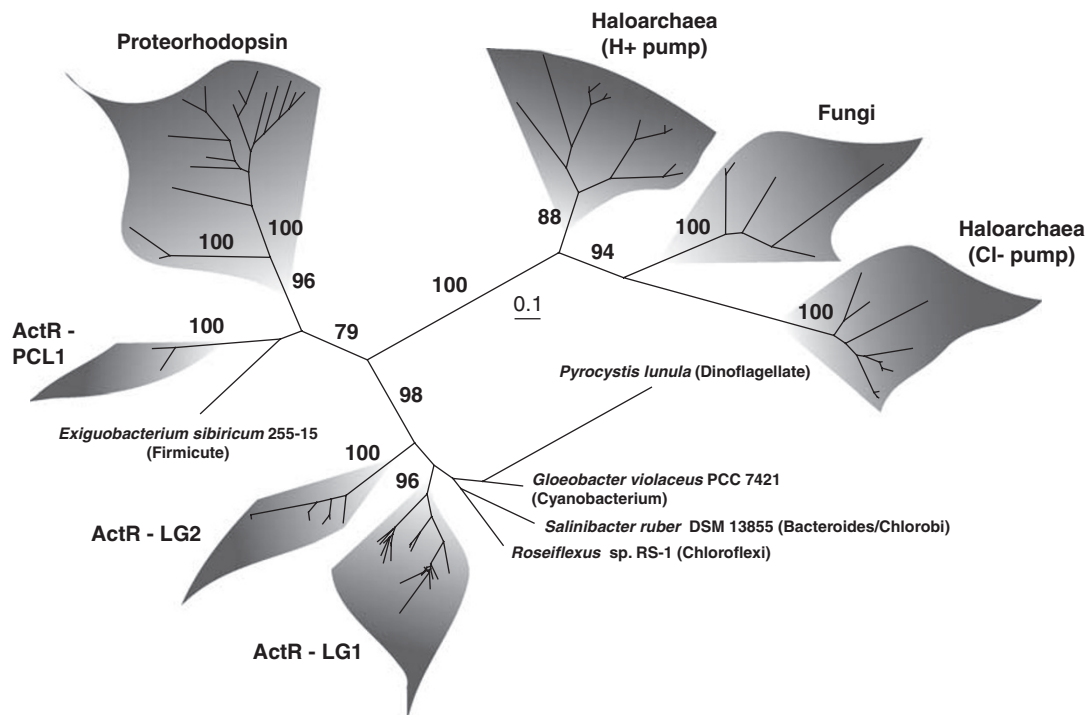

Figure 1 Best maximum likelihood tree of 81 selected microbial rhodopsin proteins (177 positions). Only selected bootstrap values are displayed. Major clades of rhodopsin sequences are outlined. In Sharma et al. (2008), we described three clades of actinorhodopsins (ActR) identified from the Global Ocean Sampling (GOS) expedition's metagenomic samples of non-marine sites (Rusch et al., 2007). In addition to ActR sequences from our earlier study, this tree also contains five full-length LG1 ActR sequences obtained here, from actinobacterial cultures MWH-Uga1, MWH-Dar1, MWH-Dar4, MWH-Ta8, MWH-EgelM2-3. Other sequences were obtained from GenBank searches and the accession numbers for sequences in each group are given in Supplementary file 1.

different sample depths in the Lake Erie water column at 5 m (24 °C), 13.5 m (16.5 °C) and 16.5 m (9 °C). Filters containing Great Lakes biomass were rinsed with 3 ml of TE buffer to remove cells and the resulting suspension was spun down at maximum speed in a table top microcentrifuge at 4 °C. Following centrifugation, the pellet was treated using the bacterial genomic DNA isolation using CTAB protocol of Feil WS and Feil H available on the Joint Genome Institute's recommended protocols page (http://my.jgi.doe.gov/general/).

Rhodopsin PCR

Degenerate PCR primers (Table 1) were designed to amplify partial gene sequences from ActR sequences clades (LG1, LG2 and PCL1) uncovered from our previous study (Sharma et al., 2008). Rhodopsins were amplified from metagenomic DNA samples using either the combined set of three primers targeting ActR clades LG1 and LG2 (Table 1) or a set of three degenerate primers previously designed by Sabehi et al. (2005) to target proteorhodopsin. PCR reactions were performed in 20 μl volumes using Eppendorf MasterMix ($2.5 \times$) containing 1 μl or ∼20 ng of DNA, 8 μl of master mix, 1 μl of a 10 μM working solution of each of the three primers (3 μl total) and 8 μl dH$_2$O. The amplification steps were: initial denaturation at 95 °C for 4 min, then 40 cycles of 94 °C for 1 min, annealing temperature (46.6 °C for actR primers or 44 °C for pR primers) for 1 min, 68 °C for 30 s, then a final elongation step of 68 °C for 2 min. PCR products of ∼330 bp were gel-purified (MiniElute; Qiagen, Valencia, CA, USA), and cloned by using the Invitrogen TOPO-TA kit. Clones were screened by colony PCR using primers flanking the insert in the TOPO vector and amplicons were cleaned using the High Pure PCR Product Purification kit (Roche, Indianapolis, IN, USA) and sequenced using the GenomeLab Dye Terminator Cycle Sequencing with Quick Start kit on a CEQ 8000 genetic analysis capillary sequencer both from Beckman Coulter (Fullerton, CA, USA). Sequences were edited using Sequencher 4.8 from Gene Codes Corporation (Ann Arbor, MI, USA).

ActR gene amplification from cultivated Actinobacteria

All investigated cultures (Supplementary Table S1) were obtained using the filtration-acclimatization method and maintained as outlined in Hahn et al. (2004). Genomic DNA was extracted from the cultures using the FastDNA kit and the FastPrep instrument (MP Biomedicals, LLC.) as recommended by the manufacturer. Partial ActR gene sequences were obtained as indicated above and multiple clones were sequenced for each ActR gene fragment amplified and cloned from a single culture. Only culture MWH-EgelM2-3 was discovered to have three ActR sequence variants and the seven remaining cultures were found to harbor only a single actR allele each, which was subsequently verified by direct sequencing of actR PCR products from each culture. A long walk PCR technique designed by Katz et al. (2000) was used to obtain full-length ActR gene sequences.

SSU rRNA PCR

With the exception of the acI strain, SSU rRNA sequences of all other actinobacterial strains were directly sequenced with universal primers (Hahn et al., 2003) or by using the more specific LEAD1 primer (Hahn and Pöckl, 2005). The SSU rRNA gene of the acI actinobacterium from MWH-EgelM2-3 was initially identified with an acI-specific SSU rRNA PCR screen developed by Newton et al. (2007) (using AcI.856R 5'-TCGCASAAACCGTGGGAAG-3') and by Warnecke et al. (2005). PCR reactions were carried out as above and the amplification cycles were: initial denaturation at 94 °C for 2 min, followed by 30 cycles of 94 °C for 40 s, annealing at 55.5 °C for 40 s, elongation at 68 °C for 30 s and then a single cycle of 68 °C for 2 min. Owing to the low abundance of the acI organism (<0.1% of total bacterial cells) in the culture as revealed by fluorescence in situ hybridization (FISH; data not shown), the full-length SSU rRNA gene could not be obtained with universal bacterial primers. The AcI.856.R primer was paired with the Bact.SSU.8F primer (annealing temperature 56.5 °C) in one reaction and the complement of the AcI.856.R primer,

Table 1 Primers designed to target Actinorhodopsin sequence clades

ActR clade	Recovered by PCR	Forward primer(s)	Reverse primer(s)	Additional sequence diversity targeted
LG1	Yes	YRYVDW (5'-TAYMGNTAYGTNGAYTGG-3') RYIDWL[a] (5'-MGNTAYATHGAYTGGYT-3')	WGVYPI (5'-ATNGGRTANACNCCCCA-3')	Gloeobacter violaceus PCC7421
LG2	No	YRYADW (5'-TAYMGNTAYGCNGAYTGG-3')	WGVYPI (5'-AT NGG RTA NAC NCC CCA-3')	Roseiflexus sp. RS-1
PCL1[b]	No	RYIDWL[a] (5'-MGNTAYATHGAYTGGYT-3') (L/Y/F/Y) RYVDW[a] (5'-WWNMGNTAYGTNGAYTGG-3')	NKVGFG (5'-CC RAA NCC NAC YTT RTT-3') GWMIYP (5'-GG RTA DAT CAT CCA NCC-3')	Exiguobacterium sibiricum 255-15[c]

[a]Primer sequences that were the same as those from Sabehi et al. (2005).
[b]PCL1 primers did not amplify rhodopsin genes from any of the DNA samples obtained in this study.
[c]The rhodopsin sequence of E. sibiricum would only be targeted from the top set of PCL1 forward and reverse primers.

AcI.856.F (5′-CTTCCACGGTTTSTGCGA-3′) was paired with Bact.SSU.1492R (annealing temperature 49.5 °C) in another reaction to generate two sequence fragments. Concatenation of the two parts into a single longer sequence and analysis by BLAST recover a sequence identical to a previously deposited acI SSU rRNA sequence (EU117876) from Mirror Lake, Wisconsin (Newton et al., 2007). The sequence of the full-length acI SSU rRNA gene was verified by amplification of the entire sequence with match-specific PCR primers and direct sequencing.

Rhodopsin phylogenetic methods
The tree in Figure 1 was constructed using only full-length sequences. ActR sequences obtained in this study, as well additional ones from the NCBI database representing haloarchaea and fungi, were added to alignments generated in our previous study of ActRs as outlined in Sharma et al. (2008), using the program Hmmer (Eddy, 1998). Partial ActR sequences amplified from freshwater metagenomic DNA samples and from cultures are represented in a separate tree in Figure 3. Full-length and partial LG1 sequences identified in Sharma et al. (2008) that completely overlapped with those amplified here were also aligned with Hmmer. The site-specific clustering analysis of partial and full-length GOS rhodopsin sequences mapped onto the phylogeny in Figure 3 was performed in Sharma et al. (2008). All rhodopsin phylogenetic trees were calculated in the PHYML program version 2.4.5 (Guindon and Gascuel, 2003; Guindon et al., 2005) under the WAG substitution matrix (Whelan and Goldman, 2001), with a discrete Γ distribution (Yang, 1993) divided into eight rate categories and the proportion of invariable sites estimated. A total of 100 PHYML bootstrap replicates were calculated for each tree.

SSU rRNA phylogenetic analyses
SSU rRNA sequences representing different lineages of freshwater Actinobacteria (Hahn et al., 2003; Warnecke et al., 2004) were used as queries in BLAST searches of unassembled metagenomic data sets of non-marine GOS samples at the CAMERA website (http://camera.calit2.net/) (Seshadri et al., 2007). SSU rRNA sequences obtained from BLAST searches and the Actinobacteria cultures were aligned in ARB (Ludwig et al., 2004), edited manually and a tree was constructed using MEGA4 (Tamura et al., 2007). Note, the number of nucleotide positions used for the tree construction (745) was restricted by the length of some GOS sequences. Analysis of 121 actinobacterial SSU rRNA sequences from assembled GOS metagenomic data (Seshadri et al., 2007) was performed as follows: sequences were aligned as above, then a Neighbor-joining backbone tree was calculated first with reference sequences representing the phylum Actinobacteria, which included longer GOS sequences from assembled data and our own searches

of unassembled data (sequence positions 47–1037, Escherichia coli numbering). Second, shorter sequences were added to the tree using a parsimony algorithm (within ARB) and this method almost completely conserved the overall tree topology. Sequences were assigned to previously defined SSU rRNA actinobacterial clusters based on their position in the tree.

Nucleotide sequence accession numbers
Rhodopsin sequences obtained in this study were deposited in GenBank under the following accession numbers: FJ545138−FJ545171 (Lake Superior), FJ545172−FJ545208 (Lake Erie) and FJ545213−FJ545222 (Luna cultures).

Results and discussion

To test our hypotheses regarding the environmental and organismal distributions of ActR, we attempted to PCR amplify ActR-encoding gene fragments from both environmental DNA samples and cultivated Actinobacteria. Degenerate PCR primers were designed for this study (Table 1) based on the three ActR sequence clades (Figure 1) that we uncovered previously from non-marine GOS samples in Sharma et al. (2008). In addition, each primer pair encodes the capability of targeting rhodopsin sequences from additional bacterial phyla (Table 1).

Isolate MWH-Uga1 encodes an LG1-ActR
In Sharma et al. (2008), we used the term 'actinorhodopsin' to designate restricted clades of environmental rhodopsin sequences, which appeared on many GOS-derived metagenomic assemblies (Rusch et al., 2007) linked to genes whose closest relatives were actinobacterial homologues. Although our earlier analysis could not identity the specific organism(s) responsible for ActR sequences from non-marine GOS samples, it predicted that such genes should be present in planktonic Actinobacteria. In the current study we investigate a number of cultures containing Actinobacteria (Supplementary Table S1) obtained in previous works from a variety of freshwater sources (Hahn et al., 2003, 2004; Hahn and Pöckl, 2005). Our attention was initially drawn to these specific cultures because of the diversity of pigmentation represented by these organisms (Figure 2). Rhodopsin proteins are documented to contribute to similar pigmentation in other rhodopsin-expressing organisms (Bolhuis et al., 2006; Gomez-Consarnau et al., 2007). Alternatively, such coloration is often caused by the presence of carotenoid pigments (Trutko et al., 2005), which increases the probability that these organisms carry rhodopsin genes, because the retinal chromophore of these photoproteins is a by-product of carotenoid biosynthesis (Martinez et al., 2007).
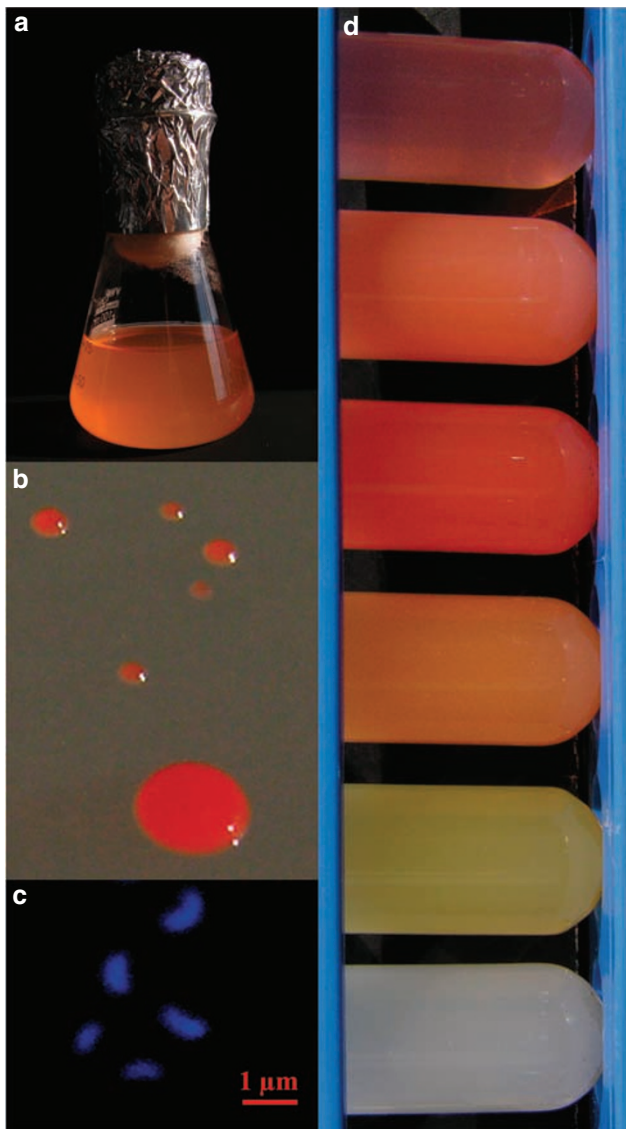
730

Figure 2 Phenotypic characteristics of actinobacterial cultures harboring ActR genes. Left column; (**a**) culture MWH-Dar1 grown in NSY medium, (**b**) colonies of MWH-Dar1 on NSY agar, (**c**) microphotograph of MWH-Dar1 cells stained with 4,6-diamidino-2-phenyl indole (DAPI). Right column; (**d**) 10-fold concentrations of cultures grown in NSY medium. From the top: culture MWH-BeijW16, MWH-EgelM2-3, MWH-Dar1, MWH-Uga1, MWH-Mo1 (actR not detected), *Polynucleobacter* sp. QLW-P1DMWA-1 (*Betaproteobacteria*, rhodopsin-negative based on genome sequence).

Here we validate the prediction that ActR genes are encoded among actinobacterial genomes, through the discovery of a freshwater isolate, strain MWH-Uga1, encoding an ActR gene belonging to the LG1 group outlined in Figure 1. Figure 3 displays the phylogenetic and ecological relationships of LG1 ActRs and shows that the MWH-Uga1 sequence groups among others from the GOS sample of the Punta Cormorant lagoon (Rusch *et al.*, 2007), previously described by us in Sharma *et al.* (2008). Isolate MWH-Uga1 was obtained by Hahn and Pöckl (2005) from a eutrophic pond in Uganda. Phenoty-

pically, this organism is characterized by an extremely small cell size of $<0.1\,\mu m^3$ (similar to the cells shown in Figure 2) and orange pigmentation (also see Figure 2). Figure 4 displays the taxonomic placement of this organism as defined by its SSU rRNA gene. Strain MWH-Uga1 is a member of the family Microbacteriaceae and belongs to a group of *Actinobacteria* previously identified from freshwater habitats, named the Luna SSU rRNA cluster. The Luna cluster (Hahn *et al.*, 2003) is named from the Latin translation of moon (luna), which is the name of the lake (Mondsee = moon lake) where the first member of this SSU rRNA cluster was cultivated. For ease of reference, the large monophyletic clade in Figure 4 containing organisms of this SSU rRNA cluster, as well as numerous sequences from cultivated and environmental sources will be defined here as 'the Luna lineage'.

LG1 ActRs in other Luna lineage Actinobacteria
LG1 ActR gene fragments were amplified from seven mixed cultures whose primary constituents are also *Actinobacteria* of the Luna lineage (Table 2). Each of these cultures contains a Luna actinobacterium and a non-actinobacterial strain, the latter of which can differ in taxonomic affiliation and proportion of total cell numbers (<1 to $>20\%$) between cultures (Hahn, 2009). The Luna *Actinobacteria* present in mixed cultures share general phenotypic similarities to isolate MWH-Uga1, such as small cell sizes, and pigmentation (Figure 2). Each culture except one (see below) exhibits a single actinobacterial ribotype and a single amplified LG1 ActR, both of which fall into the monophyletic clusters represented by the Luna lineage (Figure 4) and the MWH-Uga1 ActR sequence (Figure 3), respectively. On the basis of these collective observations, we infer that the ActRs recovered from the mixed cultures originate from the Luna strain present in each. These results provide further evidence that ActR genes are encoded in freshwater *Actinobacteria* and suggest that such genes are common among organisms belonging to the Luna lineage. However, of the 20 cultures screened by PCR (Supplementary Table S1) only 8 were positive for ActR genes, even though nearly all contained pigmented Luna *Actinobacteria*. In some instances organisms lacking detectable ActR genes were nearly identical at the SSU rRNA level to rhodopsin-encoding *Actinobacteria* (Figure 4). The absence of gene products for some cultures may result from mismatches to degenerate primers, whereas others may simply lack ActR.

ActR gene fragments recovered from environmental DNA samples
To show that ActR genes were present in other freshwater environments, besides those sampled during the GOS by Rusch *et al.* (2007), we used our degenerate ActR primers (Table 1) to amplify
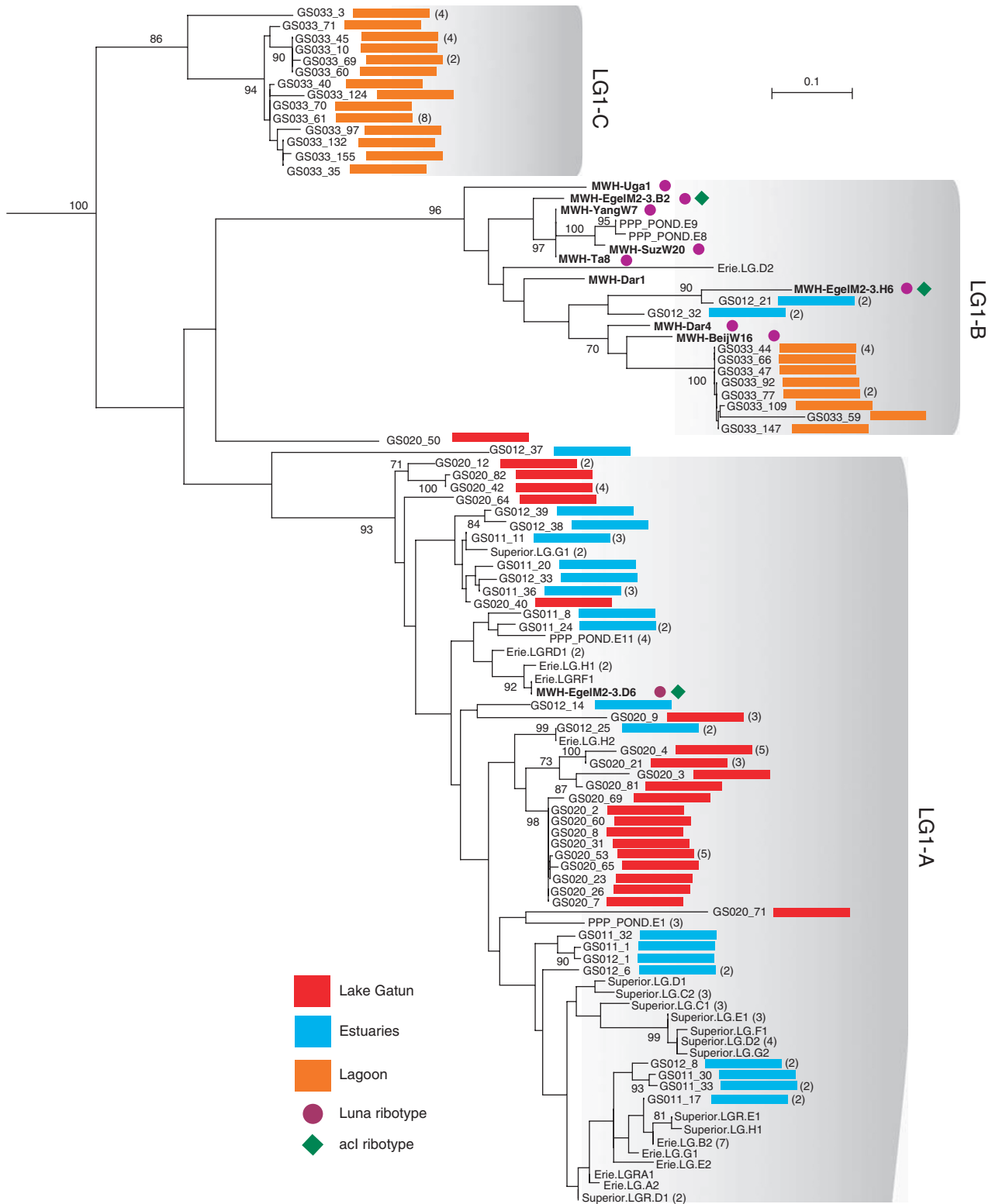
Figure 3 Best maximum likelihood tree of LG1 ActRs proteins (109 taxa, 96 positions). Only bootstrap values ⩾70 are displayed. The LG1 group is broken down into three subclades and LG2-type ActRs (outlined in Figure 1) are used as an outgroup. Sequences originally described in Sharma *et al.* (2008), retrieved from the Global Ocean Sampling (GOS) expedition's non-marine metagenomic samples are color coded according to sample origin. ActR sequences amplified in this study from actinobacterial cultures are indicated in bold, and the actinobacterial ribotype(s) associated with each are coded by symbols. Sequences amplified from environmental samples are labeled according to sample origin, and the number of times a particular sequence was detected in our library is indicated in brackets. We also show the number of times a sequence was detected in each non-marine GOS metagenomic sample, by including the results of an analysis from our previous study in Sharma *et al.* (2008), where partial and full-length rhodopsin protein sequences were clustered at a phylogenetic distance of ⩽0.1. From this analysis, most LG1 sequence clusters formed from each GOS sample have a representative in the tree above.
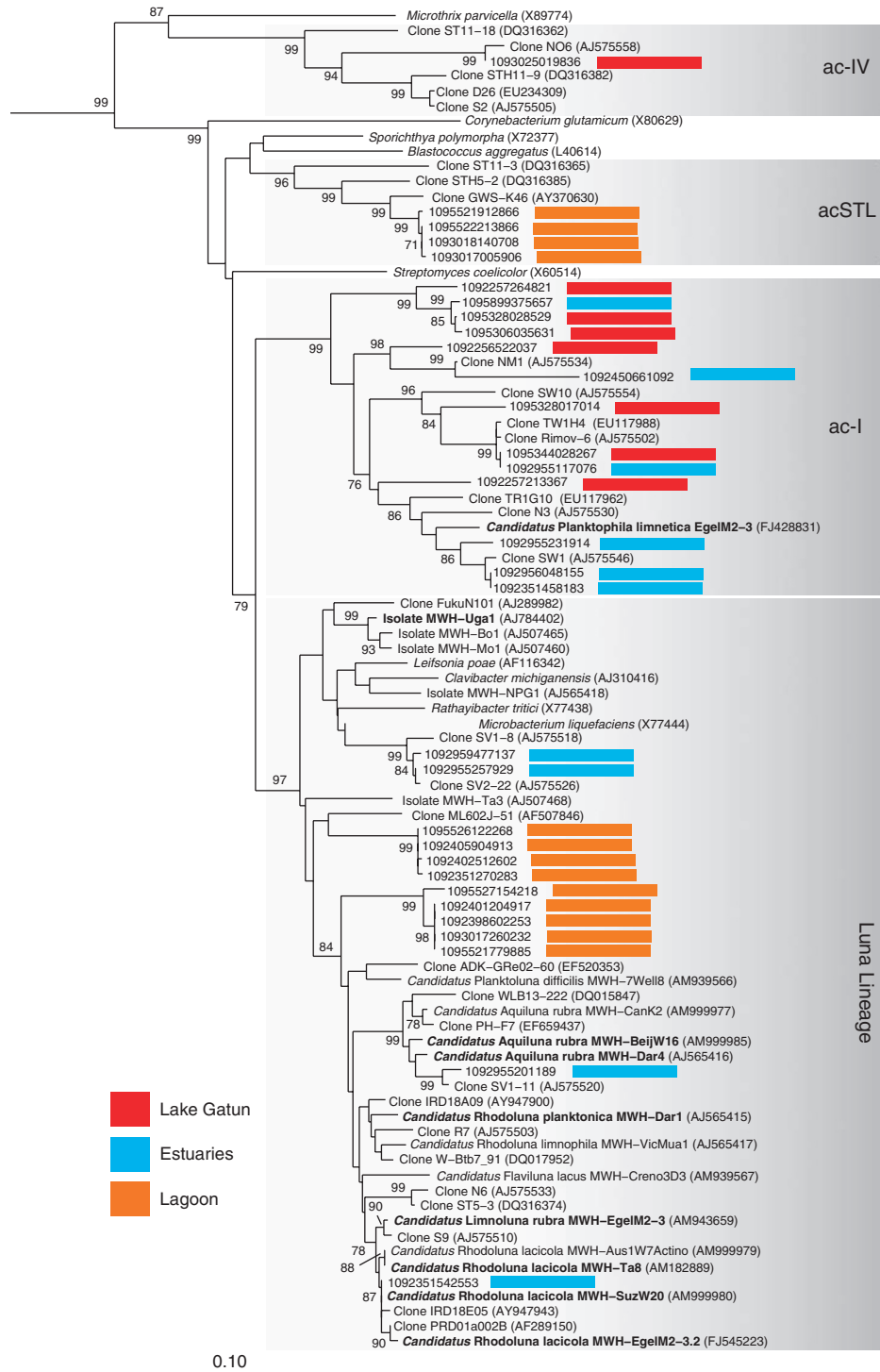
Figure 4 Neighbor-joining tree (with Jukes/Cantor correction) of 92 partial actinobacterial small subunit ribosomal RNA (SSU rRNA) sequences (*E. coli* positions 95–840) from the Global Ocean Sampling (GOS) expedition's non-marine metagenomic samples and cultivated freshwater *Actinobacteria*, within the context of selected sequences from the phylum *Actinobacteria*. Bootstrap values (MEGA4) represent percentage support of the nodes based on 1000 resamplings (only values ⩾ 70% are shown). Actinobacterial SSU rRNA sequences from pure and mixed cultures containing ActR genes are indicated in bold. Groups of freshwater *Actinobacteria* defined by past environmental SSU rRNA studies are outlined. Sequences from non-marine GOS samples are color coded according to their sample origin (as in Figure 3) and labeled with CAMERA database accession numbers.

environmental DNA samples obtained from the surface of three freshwater habitats (Table 2). The first was a small (<1 Ha) shallow (depth <1 m)

pond, named 'Point Pleasant Pond' located in a wooded area in Halifax, Nova Scotia, Canada. The pond was selected as an initial site to test our actR

Table 2 Freshwater isolates, mixed cultures or metagenomic samples harboring ActR gene(s)

Sample type (isolate, mixed culture or metagenomic)	Geographical origin	Habitat	Trophic status	Climate zone	Conserved acidic aa at positions 97 and 108[a,b] for proton pumping
Isolate MWH-Uga1	Jinja, Uganda	Pond	Eutrophic	Tropical	Yes (1/1)[c]
Culture MWH-Dar1[d]	Dar es Salaam, Tanzania	Pond	Eutrophic	Tropical	Yes (1/1)[c]
Culture MWH-Dar4[e]	Dar es Salaam, Tanzania	Pond	Eutrophic	Tropical	Yes (1/1)[c]
Culture MWH-Ta8[f]	Lake Taihu, China	Lake	Hypertrophic	Subtropical	Yes (1/1)[c]
Culture MWH-YangW7	Yangtze River, Nanjing, China	River	Eutrophic	Subtropical	Yes (1/1)
Culture MWH-SuzW20[f]	Suzhou, China	Pond	Eutrophic	Subtropical	Yes (1/1)
Culture MWH-BeijW16[e]	Beijing, China	Pond	Not determined	Temperate	Yes (1/1)
Culture MWH-EgelM2-3[g]	Lake Egelsee, Austria	Lake	Eutrophic	Temperate	Yes (3/3)[c]
Lake Superior	North America	Lake	Ultra-oligotrophic	Temperate	100% (34/34)[h]
Lake Erie	North America	Lake	Meso/Oligotrophic	Temperate	100% (37/37)[h]
Point Pleasant Park Pond	Halifax, NS, Canada	Pond	Not determined	Temperate	100% (9/9)
Lake Gatun	Panama Canal	Lake	Meso/Eutrophic	Tropical	97% (75/77)[h]
Delaware Bay	USA	Estuary	Eutrophic	Temperate	95% (41/43)[h]
Chesapeake Bay	USA	Estuary	Supertrophic	Temperate	97% (37/38)[h]
Punta Cormorant lagoon	Galapagos Islands	Lagoon	Unknown	Tropical	98% (126/128)[h]

[a]Conserved acidic amino acids (D97 and E108) in PR critical for proton pumping function in rhodopsins.
[b]Note that for partial ActR gene sequences amplified in this study, the forward primer encodes the D97 position one codon in from the 3′ end (see Table 1).
[c]Cultures from which full-length ActR genes were obtained.
[d]Proposed 'Candidatus Rhodoluna planktonica' name accepted in Hahn (2009) as are all others shown here.
[e]Candidatus Aquiluna rubra.
[f]Candidatus Rhodoluna lacicola.
[g]One Luna-type cell in this mixed culture associated with an MWH-EgelM2-3 SSU rRNA sequence in Figure 4 named Candidatus Limnoluna rubra.
[h]In addition to ActR, percentages include all microbial rhodopsin sequence diversity obtained from these environmental samples that encode the amino-acid positions of interest.

primers and was sampled in late summer 2006, when the water level was severely decreased due to evaporation and was rich in particulate matter. The other two samples originated from the North American Great Lakes. Here we concentrate our efforts on samples from Lakes Superior and Erie previously obtained for diversity studies of cyanobacterial picoplankton (Wilhelm et al., 2006; Ivanikova et al., 2007). Lake Superior represents a unique environment because its waters are extremely phosphate limited, contributing to low rates of primary production similar to non-bloom conditions at the Bermuda Atlantic Time Series Site (RW Sterner, personal communication). Such low rates in turn yield levels of dissolved organic carbon $<2\,mg\,l^{-1}$ (Hassler et al., 2009). In contrast, Lake Erie has the highest rates of primary productivity among the Great Lakes, and high growth rates of phytoplankton in the summer lead to seasonal hypoxia in the bottom waters of the shallow (24 m) central basin (Wilhelm et al., 2006).

A total of 52 partial rhodopsin sequences were obtained by the combined use of the LG1 and 2 primer sets shown in Table 1. Of these, 43 were obtained from Lake Superior (3.1 °C in June) and Lake Erie (22 °C in July) and 9 from the initial test site, Point Pleasant Pond. Nearly all sequences recovered with the LG-type primers (51 of 52) encoded proteins that phylogenetically cluster within the LG1 clade of ActRs (Figure 3). Our prior analysis of LG1 sequences from GOS data in Sharma et al. (2008) identified a subclade within this grouping that represented nearly all of the LG1-type ActRs described from the freshwater Lake Gatun sample (44 of 45) and those from Delaware and Chesapeake Bays (32 of 38), which are estuaries that are subject to large inputs from freshwater runoffs. The sequences obtained from freshwater sites sampled in this current study also display a similar pattern (Figure 3), where 48 of 51 LG1-type ActRs fall within that same subclade, which we now define here as LG1-A. Collectively, these environmental sequence data in Figure 3 show that similar ActR genes are present among diverse freshwater habitats. Such observations verify our initial prediction based on our previous findings in Sharma et al. (2008) and the knowledge that many environmental SSU rRNA studies show that members of the Actinobacteria are widespread in fresh waters (Glockner et al., 2000; Warnecke et al., 2004; Allgaier and Grossart, 2006).

Although the cultures containing Luna-type Actinobacteria outlined above were all established from freshwater environments (Table 2), these organisms did not appear to carry genes of the LG1-A subclade that represent most of the ActR sequence diversity recovered from freshwater metagenomic DNA samples (Figure 3). Rather, ActRs from the Luna Actinobacteria cluster with sequences recovered from the GOS sample of the hypersaline Punta Cormorant lagoon, in a second subclade we refer to here as LG1-B. These observations suggest that

734

either Luna-type organisms do not belong to the group of *Actinobacteria* responsible for the LG1-A phylotypes or that a paralog representing that clade is undetectable with our primers due to nucleotide mismatches or gene absence. Therefore, in an attempt to understand the taxonomic distribution of *Actinobacteria* present in metagenomic samples containing ActRs, we performed a comparative analysis of SSU rRNA genes.

Analysis of SSU rRNA genes from samples containing ActR

Assessment of dominant ribotypes in non-marine GOS metagenomic samples by Rusch *et al.* (2007) revealed a large contribution of sequences from *Actinobacteria*, and this observation helped inform our prediction in Sharma *et al.* (2008) that planktonic *Actinobacteria* encode rhodopsins. Our current findings from cultivated *Actinobacteria* substantiate the value of linking strong signals from phylogenetic markers, to inferences based on high-frequency patterns in metagenomic data. Here, signals from SSU rRNA genes are used from metagenomic samples and cultivated *Actinobacteria* to gain insight into the distribution of ActR genes among members of this phylum.

Three SSU rRNA data sets were used for this exercise and the first comes from previous work by Jensen and Lauro (2008), who thoroughly analyzed actinobacterial SSU rRNA sequence diversity from assembled GOS metagenomic data (Rusch *et al.*, 2007) and kindly provided us their data set from non-marine sites. The phylogenetic affinity, as determined by us, for each of the 121 partial and full-length sequences from this data set is summarized in Figure 5. As a second independent assessment, phylogenetic analysis was performed on sequences recovered from BLAST searches of the unassembled GOS metagenomic data (Seshadri *et al.*, 2007) from the same non-marine sample sites using SSU rRNA sequence representatives for each of the known lineages of freshwater *Actinobacteria* as queries (Figure 4). The third data set consists of actinobacterial sequences recovered from a clone library of bacterial SSU rRNA gene fragments amplified from Lake Erie DNA samples (Supplementary Figure S1). Intra and intersite comparisons of ActRs and rRNAs from environmental samples were interpreted in the context of sequences obtained from cultures.

In Sharma *et al.* (2008) , we described a set of ActR sequences that appears to originate from Luna-related *Actinobacteria*. Above we outlined the LG1-B clade consisting of Luna-affiliated ActRs and sequences previously described from the hypersaline lagoon GOS sample (Figure 3). Our analysis of GOS lagoon SSU rRNA sequences retrieved from CAMERA (Seshadri *et al.*, 2007) shows two distinct groups that branch within the Luna lineage (Figure 4). It is likely that the
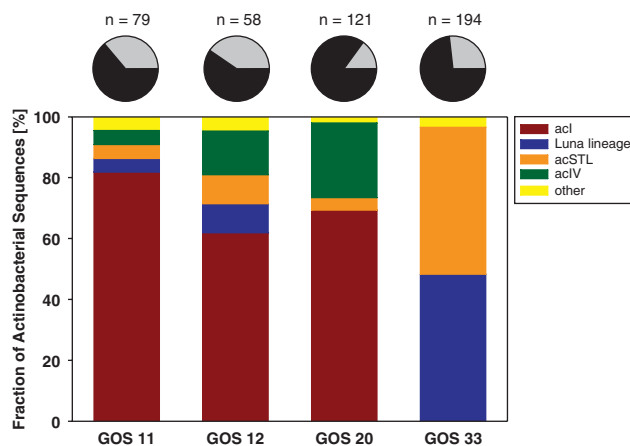

Figure 5 Group-specific phylogenetic affiliation for 121 actinobacterial small subunit ribosomal RNA (SSU rRNA) sequences identified from assembled metagenomic data from the Global Ocean Sampling (GOS) expedition's non-marine samples. These samples are indicated on the *x* axis by their GOS identifiers: the estuaries Delaware Bay (GOS11) and Chesapeake Bay (GOS12), the freshwater Lake Gatun (GOS20) and the hypersaline Punta Cormorant Lagoon (GOS33). Only SSU rRNA sequences ⩾400 nucleotides in length were analyzed. Groups of freshwater *Actinobacteria* are color coded according to key and the number of sequences belonging to a specific group in each sample is shown by the proportional representation of a individually colored bar in the graph. Pie charts at the top of each bar depict the total percentage of actinobacterial SSU rRNA sequences (gray) among the total number of SSU rRNA sequences (*n*) contained in the respective metagenomic sample. The sequences used in this analysis were provided by Jensen and Lauro (see text for further details).

organisms represented by these lagoon SSU rRNA sequences are responsible for at least the ActRs that branch in the LG1-B clade with cultivated Luna. In addition, the presence of a few Luna-related ActRs in GOS estuarine samples can also be accounted for by related SSU sequences (Figures 4 and 5).

The lack of Luna lineage rRNAs detected from the GOS Lake Gatun sample (Figure 5) and in the Lake Erie data obtained by us (Supplementary Figure S1) suggests it is unlikely that such organisms account for the LG1-A subclade of sequences in Figure 3 that represents most of ActR gene diversity recovered from freshwater metagenomic DNA samples. Interestingly, SSU rRNA sequences from the uncultured acI lineage comprise over 60% of actinobacterial SSU rRNA genes recovered from the non-marine GOS samples where LG1-A ActRs are found abundant (Figure 5). Organisms belonging to the acI lineage are found widespread and abundant among diverse types of freshwater environments (Allgaier and Grossart, 2006; Newton *et al.*, 2007). The distribution of the acI SSU rRNA genes follows that of the LG1-A ActRs; both are frequently observed in Lake Gatun, Lake Erie (Supplementary Figure S1) and estuarine samples but absent from the Punta Cormorant lagoon (Figure 5). Although these correlations provide a strong lead for future studies, the ability to link acI cell types to LG1-A ActRs is hampered by our current knowledge of this group, which is limited to environmental SSU rRNA

sequences and FISH studies (Warnecke et al., 2005; Newton et al., 2007). In the following section we provide additional evidence to support the hypothesis that acI Actinobacteria are the bearers of the LG1-A ActRs in the form of a mixed culture containing an acI cell type.

Culture MWH-EgelM2-3 contains an acI cell type

MWH-EgelM2-3 was the only mixed culture found to harbor two phylogenetically distinct ActR variants, suggesting that more than one actinobacterial strain was present. This culture, established from a small eutrophic lake in Austria (Table 2), carried ActRs of both the LG1-A and LG-1 B subclades (Figure 3). On the basis of the recovery of LG1-B ActRs from the multiple Luna cultures discussed above, we infer that ActR variants of this type from culture MWH-EgelM2-3 can be accounted for by the presence of Luna organisms (Figure 4). Given the high frequency of acI SSU rRNA genes in metagenomic samples containing LG1-A ActRs (Figure 5), we suspected an acI organism accounted for the LG1-A sequence in MWH-EgelM2-3. Using the acI-specific FISH probe acI-852 (Warnecke et al., 2005), we confirmed that an acI organism was present at an abundance of <0.1% of the total cell population in MWH-EgelM2-3. Subsequently, we obtained the SSU rRNA gene for this organism and demonstrated its affiliation with the acI lineage (Figure 4). Meanwhile, we further analyzed the phylogeny of this acI cell type and proposed to establish the Candidatus species 'Candidatus Planktophila limnetica' for this strain (Jezbera et al., 2009).

Strikingly, we found that the SSU rRNA sequence of 'Candidatus Planktophila limnetica' was identical to a partial acI sequence (Erie.Aughypo31) from Lake Erie (Supplementary Figure S1), which could explain the recovery of identical LG1-A ActR gene fragments obtained from each of these sources (Figure 3). In combination with the high incidence of acI SSU rRNA genes in metagenomic samples harboring LG1-A ActRs, these findings further support the hypothesis that the acI cell type present in this mixed culture is the bearer of an ActR variant absent from all other cultures containing Luna Actinobacteria. Although absolute evidence is required to validate this hypothesis, the presence of the acI organism in culture will provide a future means to assess if acI Actinobacteria encode ActRs.

Functional homology and sequence characteristics of ActRs

Of the ActRs we described from non-marine GOS metagenomic data in Sharma et al. (2008), nearly all conserved two acidic amino acids, D97 and E108, critical for a light-driven proton pumping function (Table 2). Five full-length ActR genes obtained from actinobacterial cultures here using a walking PCR technique (Katz et al., 2000) also conserve these residues (Table 2), suggesting that ActRs may function in photoheterotrophy. ActR gene fragments obtained from environmental samples and cultures in this study were amplified with a forward primer that encodes the D97 residue close to the 3′ end (Table 1). Therefore, it cannot be stated with certainty that these partial sequences also conserve both acidic amino acids, although each of them conserves the E108 position. Further analysis of ActRs sequence characteristics reveals that genes of the LG1-A and LG-B type exhibit a much lower GC content (~40–50%, Supplementary Figure S2) than commonly perceived for organisms belonging to the phylum Actinobacteria, who are commonly referred to as 'High GC Gram-positive bacteria' (Ventura et al., 2007).

Taxonomic and ecological distributions of rhodopsins in nature

Together these environmental and genomic sequence data reflect previous observations that planktonic Actinobacteria have a global distribution in diverse freshwater environments and suggest that at least some of them use ActRs. In addition to ActR, proteorhodopsin-encoding gene fragments affiliated with proteobacterial and bacteroidetes sequences (Supplementary Table S2) were also amplified from the Great Lakes Superior and Erie. Several recent studies highlight the broad taxonomic and ecological distribution of microbial rhodopsin genes in diverse aquatic environments (Rusch et al., 2007; Atamna-Ismaeel et al., 2008; Sharma et al., 2008), suggesting they are important in the adaptation of microbes to life on the earth's surface. However, exploratory studies such as ours can only provide insight into sequence diversity, generating fundamental questions regarding rhodopsin function and its impact at both cellular and ecosystem levels. A recent analysis by Gonzalez et al. (2008) of the genome sequence of a proteorhodopsin-encoding marine bacterium proposed this organism may be capable of both carbon fixation and rhodopsin-based phototrophy. Such ideas highlight the vast potential for novel metabolic strategies involving rhodopsins. Freshwater Actinobacteria are an extremely successful group of organisms of unknown metabolic potential that also encode ActRs. The discovery of cultivated actinobacterial representatives encoding these genes represents an exciting opportunity to understand the significance of rhodopsins to their biology and to freshwater ecosystems.

736

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.

Atamna-Ismaeel N, Sabehi G, Sharon I, Witzel KP, Labrenz M, Jürgens K *et al.* (2008). Widespread distribution of proteorhodopsins in freshwater and brackish ecosystems. *ISME J* **2**: 656–662.

Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP *et al.* (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902–1906.

Bolhuis H, Palm P, Wende A, Falb M, Rampp M, Rodriguez-Valera F *et al.* (2006). The genome of the square archaeon *Haloquadratum walsbyi*: life at the limits of water activity. *BMC Genomics* **7**: 169.

Campbell BJ, Waidner LA, Cottrell MT, Kirchman DL. (2008). Abundant proteorhodopsin genes in the North Atlantic Ocean. *Environ Microbiol* **10**: 99–109.

Eddy SR. (1998). Profile hidden Markov models. *Bioinformatics* **14**: 755–763.

Fuhrman JA, Schwalbach MS, Stingl U. (2008). Proteorhodopsins: an array of physiological roles? *Nat Rev Microbiol* **6**: 488–494.

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.

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.

Gonzalez JM, Fernandez-Gomez B, Fernandez-Guerra A, Gomez-Consarnau L, Sanchez O, Coll-Llado M *et al.* (2008). Genome analysis of the proteorhodopsin-containing marine bacterium Polaribacter sp. MED152 (Flavobacteria). *Proc Natl Acad Sci USA* **105**: 8724–8729.

Guindon S, Gascuel O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696–704.

Guindon S, Lethiec F, Duroux P, Gascuel O. (2005). PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* **33**: W557–W559.

Hahn MW. (2009). Description of seven *Candidatus* species affiliated with the phylum Actinobacteria, representing planktonic freshwater bacteria. *Int J Syst Evol Microbiol* **59**: 112–121.

Hahn MW, Lünsdorf H, Wu Q, Schauer M, Höfle MG, Boenigk J *et al.* (2003). Isolation of novel ultramicrobacteria classified as Actinobacteria from five freshwater habitats in Europe and Asia. *Appl Environ Microbiol* **69**: 1442–1451.

Hahn MW, Pöckl M. (2005). Ecotypes of planktonic Actinobacteria with identical 16S rRNA genes adapted to thermal niches in temperate, subtropical, and tropical freshwater habitats. *Appl Environ Microbiol* **71**: 766–773.

Hahn MW, Stadler P, Wu QL, Pöckl M. (2004). The filtration-acclimatization method for isolation of an important fraction of the not readily cultivable bacteria. *J Microbiol Methods* **57**: 379–390.

Hassler CS, Havens SM, Bullerjahn GS, McKay RML, Twiss MR. (2009). Iron speciation in western Lake Superior measured using combined physical, chemical and biological assessment. *Limnol Oceanogr* (in press).

Ivanikova NV, Popels LC, McKay RM, Bullerjahn GS. (2007). Lake Superior supports novel clusters of cyanobacterial picoplankton. *Appl Environ Microbiol* **73**: 4055–4065.

Jensen PR, Lauro FM. (2008). An assessment of actinobacterial diversity in the marine environment. *Antonie Van Leeuwenhoek* **94**: 51–62.

Jezbera J, Sharma AK, Brandt U, Doolittle WF, Hahn MW. (2009). 'Candidatus Planktophila limnetica', an Actinobacterium representing one of the most numerically important taxa in freshwater bacterioplankton. *Int J Syst Evol Microbiol* (in press).

Katz LA, Curtis EA, Pfunder M, Landweber LF. (2000). Characterization of novel sequences from distantly related taxa by walking PCR. *Mol Phylogenet Evol* **14**: 318–321.

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.

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.

Newton RJ, Jones SE, Helmus MR, McMahon KD. (2007). Phylogenetic ecology of the freshwater Actinobacteria acI lineage. *Appl Environ Microbiol* **73**: 7169–7176.

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S *et al.* (2007). The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* **5**: e77.

Sabehi G, Loy A, Jung KH, Partha R, Spudich JL, Isaacson T *et al.* (2005). New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biol* **3**: e273.

Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M. (2007). CAMERA: a community resource for metagenomics. *PLoS Biol* **5**: e75.

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.

Tamura K, Dudley J, Nei M, Kumar S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.

Trutko SM, Dorofeeva LV, Evtushenko LI, Ostrovskii DN, Hintz M, Wiesner J *et al.* (2005). [Isoprenoid pigments

in representatives of the family Microbacteriaceae]. *Mikrobiologiia* **74**: 335–341.

Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF, Chater KF *et al.* (2007). Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* **71**: 495–548.

Warnecke F, Amann R, Pernthaler J. (2004). Actinobacterial 16S rRNA genes from freshwater habitats cluster in four distinct lineages. *Environ Microbiol* **6**: 242–253.

Warnecke F, Sommaruga R, Sekar R, Hofer JS, Pernthaler J. (2005). Abundances, identity, and growth state of actinobacteria in mountain lakes of different UV transparency. *Appl Environ Microbiol* **71**: 5551–5559.

Whelan S, Goldman N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* **18**: 691–699.

Wilhelm SW, Bullerjahn GS, Rinta-Kanto JM, Eldridge ML, Bourbonniere RA. (2006). Seasonal hypoxia and the genetic diversity of prokaryote populations in the central basin hypolimnion of Lake Erie: evidence for abundant picocyanobacteria and photosynthesis. *J Great Lakes Res* **32**: 657–671.

Yang Z. (1993). Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol Biol Evol* **10**: 1396–1401.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)