Environmental Microbiology (2005) **7**(12), 1896–1908 doi:10.1111/j.1462-2920.2005.00883.x

Aerobic anoxygenic photosynthesis genes and operons in uncultured bacteria in the Delaware River

Lisa A. Waidner and David L. Kirchman*
University of Delaware, College of Marine Studies, 700 Pilottown Road, Lewes, DE 19958, USA.

Summary

Photosynthesis genes and operons of aerobic anoxygenic photosynthetic (AAP) bacteria have been examined in a variety of marine habitats, but genomic information about freshwater AAP bacteria is lacking. The goal of this study was to examine photosynthesis genes of AAP bacteria in the Delaware River. In a fosmid library, we found two clones bearing photosynthesis gene clusters with unique gene content and organization. Both clones contained 37 open reading frames, with most of those genes encoding known AAP bacterial proteins. The genes in one fosmid were most closely related to those of AAP bacteria in the *Rhodobacter* genus. The genes of the other clone were related to those of freshwater beta-proteobacteria. Both clones contained the *acs*F gene, which is required for aerobic bacteriochlorophyll synthesis, suggesting that these bacteria are not anaerobes. The beta-proteobacterial fosmid has the *puf* operon B-A-L-M-C and is the first example of an uncultured bacterium with this operon structure. The alpha-3-proteobacterial fosmid has a rare gene order (Q-B-A-L-M-X), previously observed only in the *Rhodobacter* genus. Phylogenetic analyses of photosynthesis genes revealed a possible freshwater cluster of AAP beta-proteobacteria. The data from both Delaware River clones suggest there are groups of freshwater or estuarine AAP bacteria distinct from those found in marine environments.

Introduction

Aerobic anoxygenic photosynthetic (AAP) bacteria appear to be common in oxic aquatic ecosystems (Kolber *et al*., 2000; Beja *et al*., 2002; Schwalbach and Fuhrman, 2005) and may alter current models of the carbon cycle and other biogeochemical processes (Karl, 2002). Cultured AAP bacteria photosynthesize with the use of bacteriochlorophyll a (bchl a), but do not evolve oxygen (Yurkov and Beatty, 1998). A more thorough understanding of AAP bacteria is needed to elucidate their role in aquatic environments.

The diversity of marine AAP bacteria is often explored with the gene encoding the alpha subunit of the photosynthetic reaction centre, *puf*M (Beja *et al*., 2002; Oz *et al*., 2005; Schwalbach and Fuhrman, 2005). On the basis of *puf*M, AAP bacteria have been classified into two clusters of proteobacteria (Nagashima *et al*., 1997). One consists of a mixture of alpha-1 and alpha-2, beta- and gamma-proteobacteria, referred to here as the 'mixed cluster' (Nagashima *et al*., 1997). The other cluster contains the alpha-3-proteobacterial group (*Roseobacter*- and *Rhodobacter*-like) and alpha-4 (*Erythrobacter*-like) (Beja *et al*., 2002; Koblizek *et al*., 2003; Oz *et al*., 2005). Few studies have examined the diversity of freshwater AAP bacteria. Karr *et al*. (2003) sequenced the *puf*M of aerobic and anaerobic AAP bacteria in a frozen lake in Antarctica. Bacteria isolated from Crater Lake, an ultraoligotrophic freshwater system, were also examined for the presence of *puf*M genes (Page *et al*., 2004).

In addition to *puf*M, AAP bacteria require approximately 40 genes to construct functional photosynthetic complexes (Yurkov and Beatty, 1998). The bchl a and carotenoid genes (*bch*, *crt*) and those encoding proteins of light harvesting (*puf*B-A) and reaction centres (*puf*L-M and *puh*A) are organized in two major superoperons together in a 45- to 50-kb gene cluster (Bauer *et al*., 1991). The gene order within each of the superoperons and the relative orientation of the operons vary among genera of proteobacteria (Young *et al*., 1989). Furthermore, some genes in the anoxygenic photosynthesis gene clusters are specific to aerobic environments. For instance, the *acs*F gene encodes an enzyme required for aerobic bchl a synthesis (Pinta *et al*., 2003). This gene is also present in four BAC clones from uncultured AAP bacteria in Monterey Bay (Beja *et al*., 2002). Another example is the *ppa*A gene. It encodes a protein that activates expression of photosynthesis proteins in aerobic conditions (Gomelsky *et al*., 2003). These genes and others may reveal information about the function and taxonomy of uncultured AAP bacteria.

Two studies have used a metagenomic approach (Rondon *et al*., 2000) to examine the photosynthetic gene clus-

Received 10 March, 2005; accepted 2 June, 2005. *For correspondence. E-mail kirchman@cms.udel.edu; Tel. (+1) 302 645 4375; Fax (+1) 302 645 4028.

ters of uncultured AAP bacteria from marine waters (Beja *et al*., 2002; Oz *et al*., 2005). To date, however, an environmental genomic approach has not been used to examine freshwater AAP bacteria. The *puf* operon of only one freshwater AAP bacterium, *Roseateles (Rts.) depolymerans*, has been sequenced (Suyama *et al*., 2002).

The goal of this study was to examine freshwater AAP bacteria in the Delaware River and to compare them to their marine counterparts. Sequences of photosynthesis and other genes allow us to explore issues about the potential for anoxygenic photosynthesis in freshwater and the phylogeny of freshwater AAP bacteria. In addition, the metagenomic approach can be used to examine whether freshwater anoxygenic photosynthetic bacteria are aerobic or anaerobic. Both types are possible in freshwaters like the Delaware River, in which anaerobic bacteria from sediments are potentially introduced into an aerobic water column. We found AAP photosynthesis operons with gene content and organization previously uncharacterized among marine AAP bacteria.

Results and discussion

We isolated two clones containing the *puf*L-M gene cluster from a metagenomic fosmid library constructed with bacterial DNA from the Delaware River, the freshwater end of an estuary (Cottrell *et al*., 2005). One abundant group of 16S rDNA genes in this library was from beta-proteobacteria, which comprised 17% of clones bearing 16S rRNA genes. Microscopic enumeration by fluorescence *in situ* hybridization revealed that beta-proteobacteria were 27% of prokaryotes in the river sample (Cottrell *et al*., 2005). To determine what types of bacteria comprise the freshwater AAP bacterial community and to explore other questions about freshwater AAP bacteria, we completely sequenced the *puf*L-M fosmid clones.

Beta-proteobacterial fosmid clone

One clone, DelRiverFos06H03, contains an insert of 32 765 bp of genomic DNA (Fig. 1). The superoperon, *crt*EF-*bch*CXYZ-*puf*BALM (Young *et al*., 1989; Igarashi *et al*., 2001), was found in this fosmid (Fig. 1). This clone also contains a portion of the conserved *bch*FNBHLM-*lha*A-*puh*A operon (Bauer *et al*., 1991; Igarashi *et al*., 2001). The relative positions and transcriptional directions of the superoperons are similar to that of the beta-proteobacterium *Rubrivivax* (*Rvi*.) *gelatinosus* (Fig. 1). The gene order in these superoperons, along with the presence of all these genes, suggests that the photosynthetic complex from this uncultured organism is functional (Lawrence and Roth, 1996; Dandekar *et al*., 1998; Overbeek *et al*., 1999).

Most of the genes of this clone are most closely related to those of beta-proteobacteria. Of the 37 predicted protein coding regions, 23 encode anoxygenic photosynthesis proteins (Table 1). Of these, 22 are 40–79% identical to proteins of beta-proteobacteria. Seven of the non-photosynthesis genes encode proteins 53–86% similar to proteins of beta-proteobacteria (Table 1). Genes of the *hem* operon (Fig. 1), which encode enzymes involved in biosynthesis of cytochrome haems (Panek and O'Brian, 2002), provide further evidence that this fosmid was from a beta-proteobacterium. Of the four *hem* genes, three are closely related (64–72% identical amino acids) to those of *Polaromonas* sp. JS666, a beta-proteobacterium in the freshwater Comamonadaceae subgroup (Table S1). The phylogeny of the *hem*F gene follows that of the 16S rRNA gene in proteobacteria, and the *hem*F gene of this clone clades with other Comamonadaceae beta-proteobacteria (Fig. 2). This clone was designated the beta-proteobacterial fosmid.

This clone was likely from an aerobic bacterium capable of anoxygenic photosynthesis. Just downstream of the *puh* operon is the *acs*F gene (Fig. 1), which encodes

Table 1. Analysis of open reading frames (ORFs) of beta-proteobacterial clone DelRiverFos06H03.

Number of ORFs	Most similar organism	Phylogenetic group[a]	Amino acid identities (%)[b]	Amino acid similarities (%)[b]
Photosynthesis genes (23 total)				
20	*Rubrivivax gelatinosus*	Beta-proteobacteria	40–79	51–89
2	*Roseateles depolymerans*	Beta-proteobacteria	65–67	76
1	*Rhodobacter sulfidophilus*	Alpha-3-proteobacteria	57	77
Non-photosynthesis genes (9 total)				
4	*Polaromonas* sp. JS666	Beta-proteobacteria	63–80	74–86
3	*Rubrivivax gelatinosus*	Beta-proteobacteria	42–46	53–60
1	*Mesorhizobium loti*	Alpha-2-proteobacteria	38	57
1	*Archaeoglobus fulgidus*	*Euryarchaeota*	34	54
Unknown				
5	None	Unknown	–	–

a. Phylogenetic affiliations are based on 16S rRNA.
b. Ranges of per cent amino acid identities and similarities of fosmid clone ORFs to homologues.
Open reading frames were analysed by BLAST-X, and the organism bearing the most similar gene was noted. For complete annotation, see *Supplementary material* Table S1.

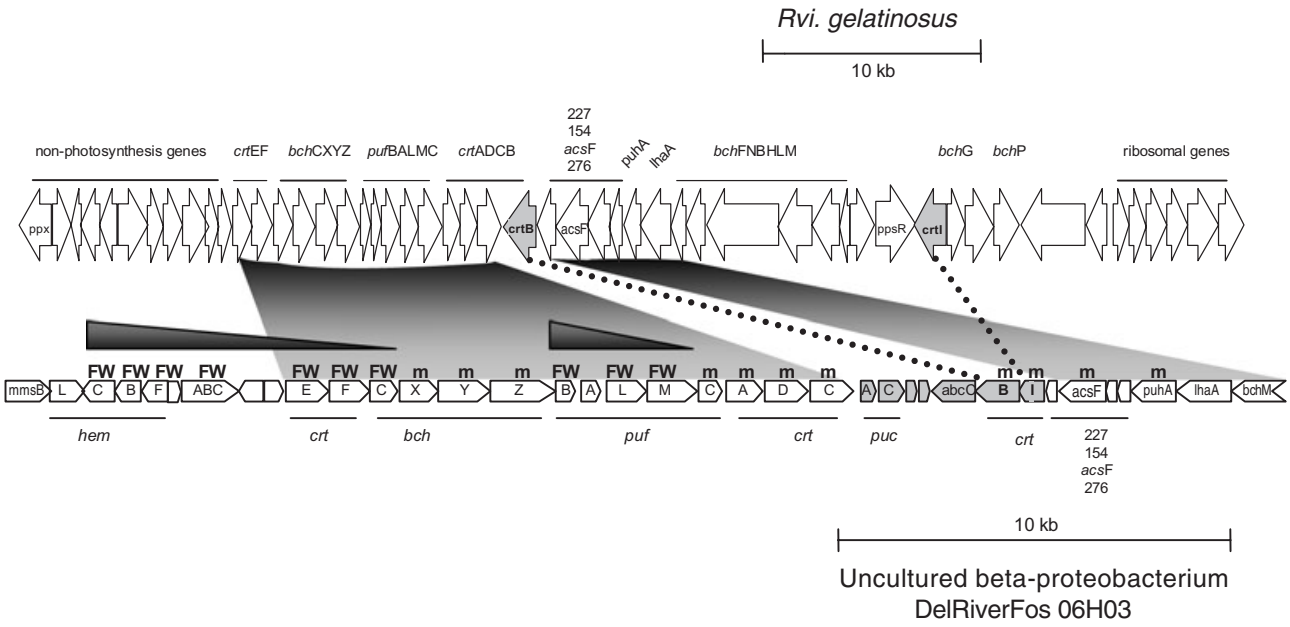

Fig. 1. Uncultured Delaware River beta-proteobacterium DelRiverFos06H03 and full photosynthesis gene cluster of *Rvi. gelatinosus* (accession number AB034704). 'FW' above open reading frame indicates it was in a separate freshwater cluster in phylogenetic analyses, while 'm' indicates it clustered with the mixed cluster of AAP bacteria. Gene symbols: *crt*, carotenoid; *bch*, bacteriochlorophyll; *puf* and *puh*, reaction centre; *puc*, light harvesting II; *lha*, light harvesting I; *hem*, coenzyme metabolism; *mms*, lipid metabolism. Open reading frame numbers 227, 154 and 276 for *Rvi. gelatinosus* hypothetical proteins correspond to accession numbers BAA94052, BAA94051 and BAA94049 respectively.

a protein required for aerobic cyclization of magnesium-protoporphyrins in bchl a synthesis (Ouchane *et al*., 2004). The reaction catalysed by this enzyme is oxygen-dependent, and this gene is not present in the strictly anaerobic anoxygenic green sulfur bacterium *Chlorobium tepidum* (Ouchane *et al*., 2004). The anaerobic anoxygenic phototroph, *Rhodobacter* (*Rba*.) *sphaeroides*, also contains the *acs*F gene (Zsebo and Hearst, 1984; Choudhary and Kaplan, 2000; Pinta *et al*., 2002), but this organism can form functional reaction centres in aerobic cultures (Koblizek *et al*., 2005). The oxygen-independent counterpart of *acs*F (*bch*E) was not found in this fosmid clone (Fig. 1). In addition, this clone contains the oxygen-dependent oxidase *hem*F, but not the oxygen-independent dehydrogenase counterpart *hem*N (Layer *et al*., 2002; Panek and O'Brian, 2002). The presence of *hem*F and *acs*F, along with the absence of their anaerobic counterparts, suggest that the organism was an AAP bacterium and not an anaerobe from anoxic sediments.

Alpha-3-proteobacterial fosmid clone

The other clone, DelRiverFos13D03, appears to be related to alpha-3-proteobacteria. This clone contains a genomic fragment of 34 908 bp, with 37 predicted open reading frames (ORFs) (Fig. 3). Most (35) of the 37 genes in the clone are similar to those of alpha-3 proteobacteria in the *Rhodobacter* genus (Table 2). Both conserved superoperons, *crt*EF-*bch*CXYZ-*puf*BALM and *bch*FN-BHLM-*lha*A-*puh*A, were found in this clone, suggesting that the DNA originated from a photosynthetically competent organism. The photosynthesis gene cluster is similar to that of *Rba. sphaeroides* (Fig. 3). Of the 37 predicted ORFs, 26 encode proteins that are 41% to 85% identical to known anoxygenic photosynthesis proteins (Table 2).

There are 10 ORFs encoding non-photosynthesis proteins, some of which can be used as phylogenetic markers. These include *rpm*E, *rpl*S, *riml* and *trm*D, which encode large subunit ribosomal proteins (Fig. 3). The phylogeny of the *rpm*E gene follows 16S rRNA phylogeny among proteobacteria (Makarova *et al*., 2001; Lecompte *et al*., 2002). The *rpm*E gene of this fosmid is most closely related to *Rba. sphaeroides* in the alpha-3-proteobacterial clade, but is distinct from *Silicibacter* (*Roseobacter*-like) (Fig. 4). In addition, phylogenetic analyses of the *trm*D and *rpl*S proteins place this fosmid clone in the *Rhodobacter* clade of alpha-3-proteobacteria (data not shown). This clone was therefore referred to as the alpha-3-proteobacterial fosmid.

This fosmid, like the beta-proteobacterial clone, contains the *acs*F gene (Fig. 3). The alpha-3-proteobacterial fosmid also contains the *ppa*A gene, which activates photopigment production and *puc* operon expression under aerobic conditions in *Rba. sphaeroides* (Gomelsky *et al*.,

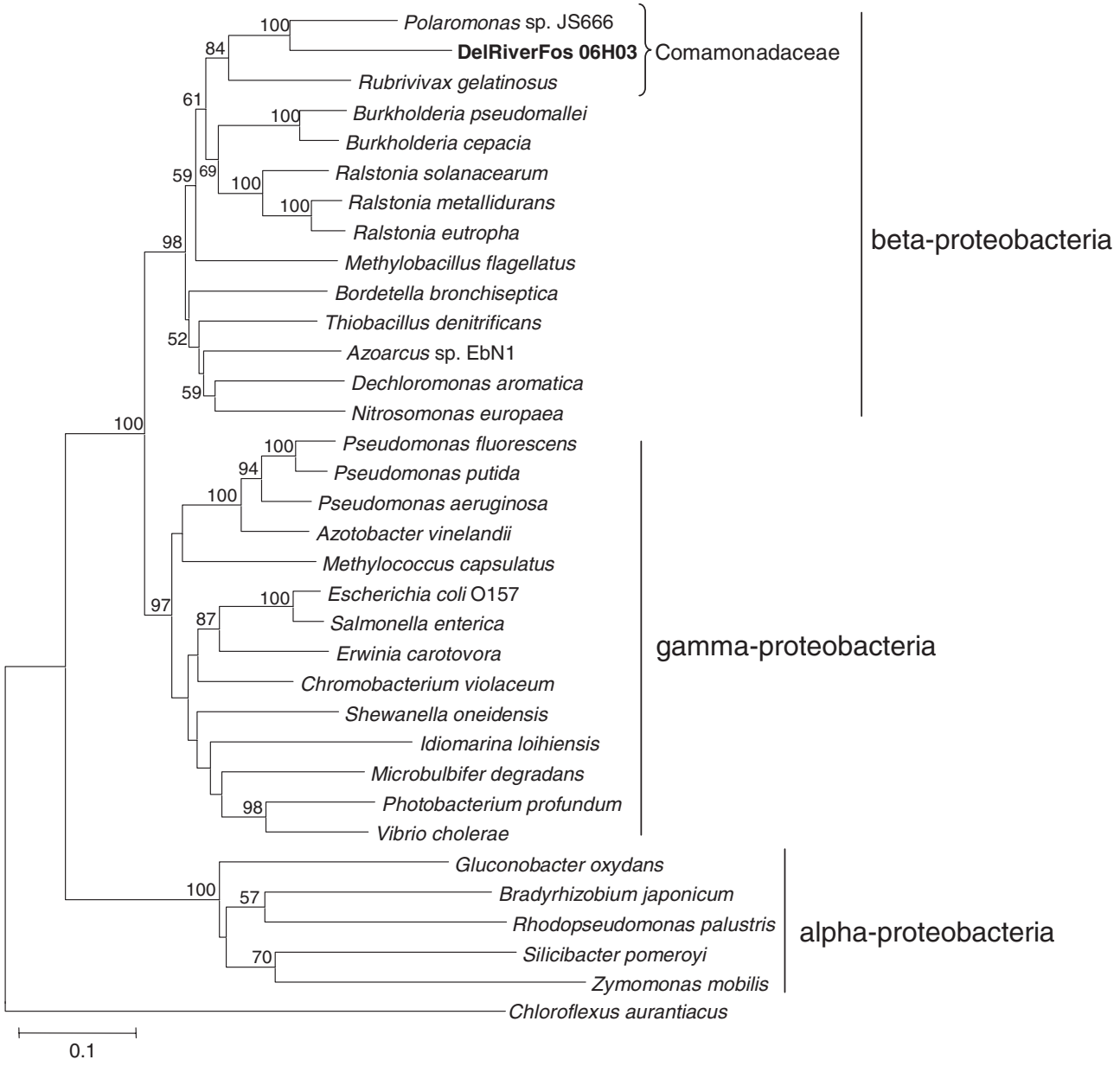

Fig. 2. Phylogenetic analysis of the *hem*F gene. The neighbour-joining tree was constructed on the basis of 780 nucleotide positions. Alpha-, beta- and gamma-proteobacterial clusters are designated on the basis of the 16S rRNA gene.

2003). The presence of these genes indicates that this photosynthesis operon was from an AAP bacterium.

The *puf*M gene from our clone clustered with those of alpha-3-proteobacteria and is most closely related to *Rhodobacter puf*M genes (Fig. 5). Similarly, the *puf*L gene of this fosmid fell in a cluster with *Rba. capsulatus* and *Rba. sphaeroides*, as well as the alpha-3-proteobacterium HTCC515 isolated from Crater Lake (Page *et al.*, 2004) (Fig. S1D). Additionally, phylogenetic analyses of proteins encoded by *crt*E, *crt*F, *bch*C, *bch*X, *crt*D and *acs*F of the alpha-3-proteobacterial fosmid indicate they were most similar to proteins in *Rhodobacter* isolates (Figs S1 and S2).

These phylogenetic analyses, as well as of the *puf* operon organization (see below), suggest that this clone was from a *Rhodobacter*-like bacterium. This was surprising, given that this type of bacteria was not abundant in the fosmid library (based on 16S rDNA data) or in the river (Cottrell *et al.*, 2005). However, *Rhodobacter* are typical freshwater or estuarine bacteria (Hansen and Imhoff, 1985; Mullins *et al.*, 1995; Suzuki *et al.*, 1997; Crump *et al.*, 1999, 2004), while *Roseobacter*, another group

Table 2. Analysis of open reading frames (ORFs) of alpha-3-proteobacterial clone DelRiverFos13D03.

Number of ORFs	Homologue organism	Phylogenetic group[a]	Amino acid identities (%)[b]	Amino acid similarities (%)[b]
Photosynthesis genes (26 total)				
17	*Rhodobacter sphaeroides*	Alpha-3-proteobacteria	41–85	62–92
3	*Rhodobacter capsulatus*	Alpha-3-proteobacteria	56–77	65–92
5	Monterey Bay BAC 60D04	Alpha-3-proteobacteria	52–77	66–83
1	*Rhodobacter blasticus*	Alpha-3-proteobacteria	44	55
Non-photosynthesis genes (10 total)				
7	*Rhodobacter sphaeroides*	Alpha-3-proteobacteria	49–83	65–84
1	*Rhodobacter capsulatus*	Alpha-3-proteobacteria	48	61
1	Monterey Bay BAC 60D04	Alpha-3-proteobacteria	68	80
1	*Aspergillus nidulans*	*Ascomycota*, Fungi	35	49
Unknown				
1	None	Unknown	–	–

Open reading frames were analysed by BLAST-X, and the organism bearing the most similar gene was noted. For complete annotation, see *Supplementary material* Table S2.
a. Phylogenetic affiliations are based on 16S rRNA for cultured and on *puf*M for Monterey Bay BAC 60D04.
b. Ranges of per cent amino acid identities and similarities of fosmid clone ORFs to homologues.

of alpha-3-proteobacteria, dominate marine habitats (Gonzalez and Moran, 1997). The alpha-3-proteobacterium represented by this clone may be related to estuarine bacteria capable of success in a wide range of salinities.

Freshwater cluster of AAP beta-proteobacteria

Analyses of individual photosynthesis genes of the beta-proteobacterial clone revealed a freshwater cluster of AAP beta-proteobacteria, to date uncharacterized. The *pufM* gene of this fosmid is in this freshwater cluster with several other freshwater representatives (Fig. 6). The *puf*L clustered with those of the freshwater beta-proteobacteria *Rhodoferax (Rfx.) fermentans* and *Rts. depolymerans* (Fig. S1D). Other genes, such as *hem*F (Fig. 2) *hem*C, *hem*B, ABC-ATPase (data not shown) and *puf*B (Fig. S1E), are also in this freshwater clade of beta-proteobacteria, containing cultured members of the

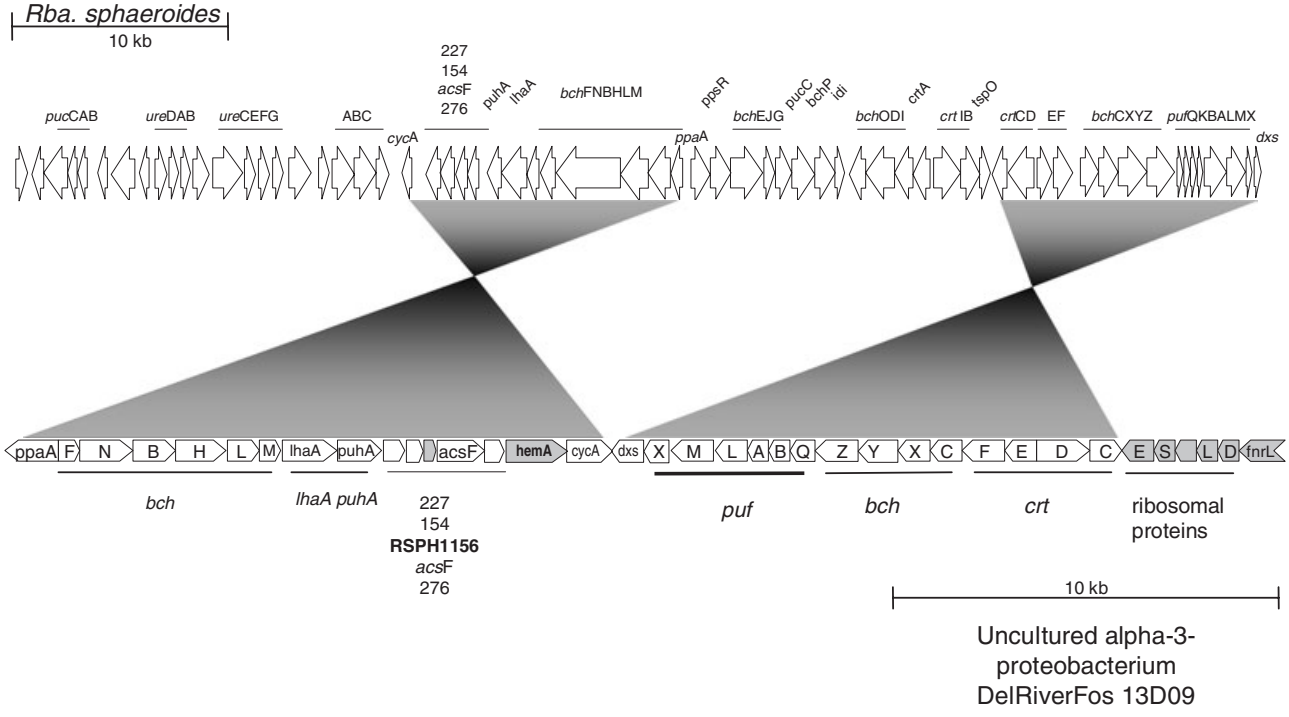

Fig. 3. Uncultured Delaware River alpha-3-proteobacterium DelRiverFos13D03 and photosynthesis gene cluster of *Rba. sphaeroides* (Accession number AF195122). Gene symbols: *crt*, carotenoid; *bch*, bacteriochlorophyll; *puf* and *puh*, reaction centre; *lha*, light harvesting I; *ppa*, photosynthesis regulation. *Rba. sphaeroides* Rsph03001156 Accession number ZP_00005240, was abbreviated RSPH 1156.

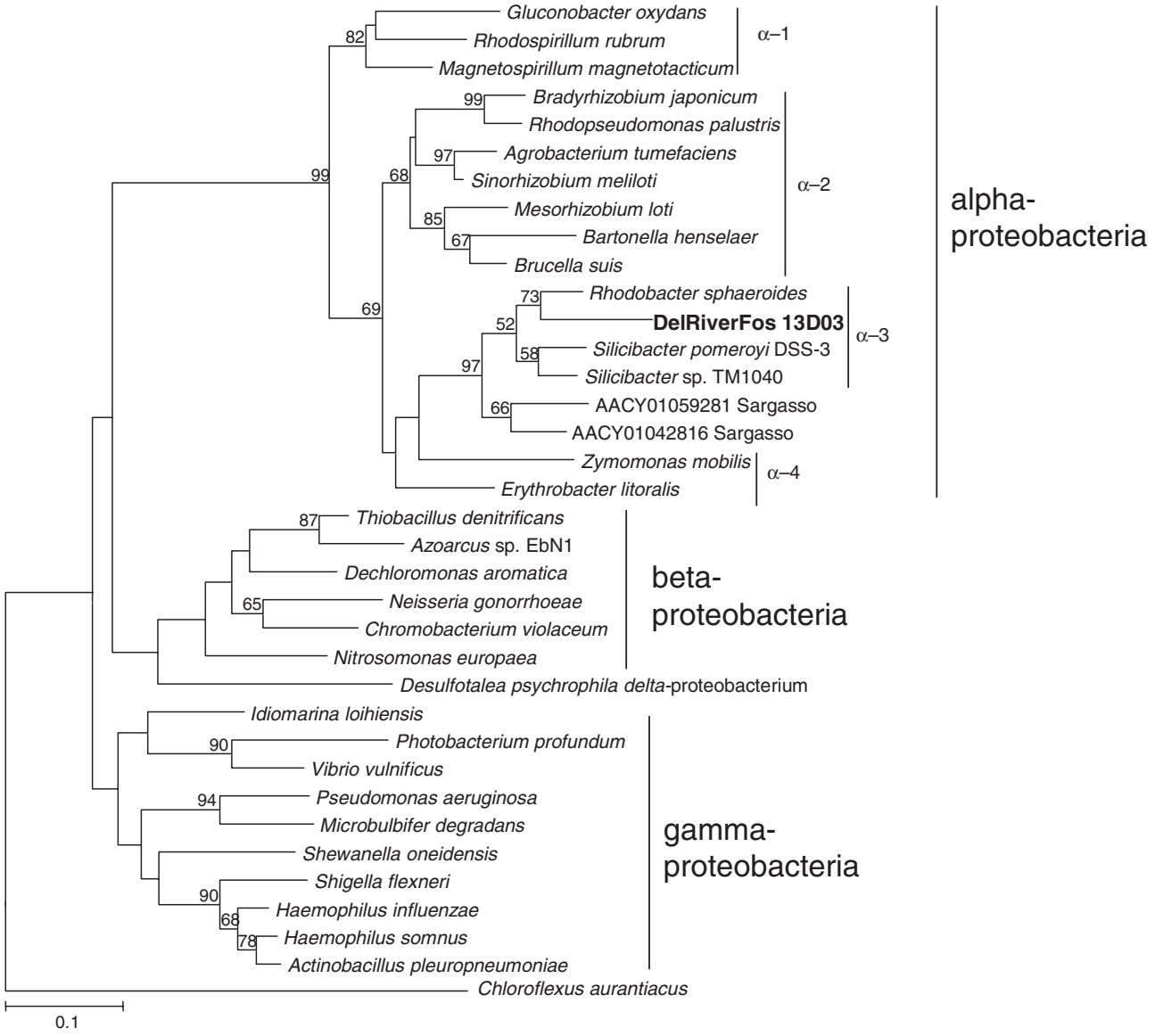

Fig. 4. Phylogenetic analysis of the *rpm*E gene. The neighbour-joining tree of the proteobacterial (C-) form of *rpm*E was constructed on the basis of 146 nucleotide positions. Alpha-, beta- and gamma-proteobacterial clusters are designated on the basis of the 16S rRNA gene. α-1, α-2, α-3 and α-4 designations refer to the subclasses of the alpha-proteobacteria.

Comamonadaceae group. The *crt*E, *crt*F and *bch*C genes of the beta-proteobacterial fosmid also provide support for this freshwater cluster, but more sequences of these genes from freshwater are needed (Fig. S1A–C). These results are summarized schematically in Fig. 1, with 'FW' marking the genes that are in a separate freshwater cluster.

In contrast, the *acs*F gene of the beta-proteobacterial clone is in the mixed cluster of cultured and uncultured AAP bacteria (Fig. S2D). Proteins encoded by *bch*X, *puf*C and *crt*D (Fig. S1A–C) and *bch*Y, *bch*Z, *crt*A, *crt*C, *crt*I, *crt*B and *puh*A (data not shown) of the beta-proteobacterial fosmid are in the mixed cluster. These data are summarized in Fig. 1, where genes marked 'm' are from the mixed cluster.

The pattern of genes in this fosmid (Fig. 1) suggests a lateral gene exchange between freshwater beta-proteobacteria and mixed cluster AAP proteobacteria, with one of the recombination sites being between the *puf*M and *puf*C genes. The freshwater AAP genes are present in the *crt*EF and *puf*BALM portions of the *crt-bch-puf* superoperon. In addition to large horizontal transfer events (Nagashima *et al*., 1997; Igarashi *et al*., 2001; Raymond *et al*., 2002), portions of photosynthesis operons appear to have been rearranged or subjected to multiple horizontal transfer rearrangements (Beja *et al*.,

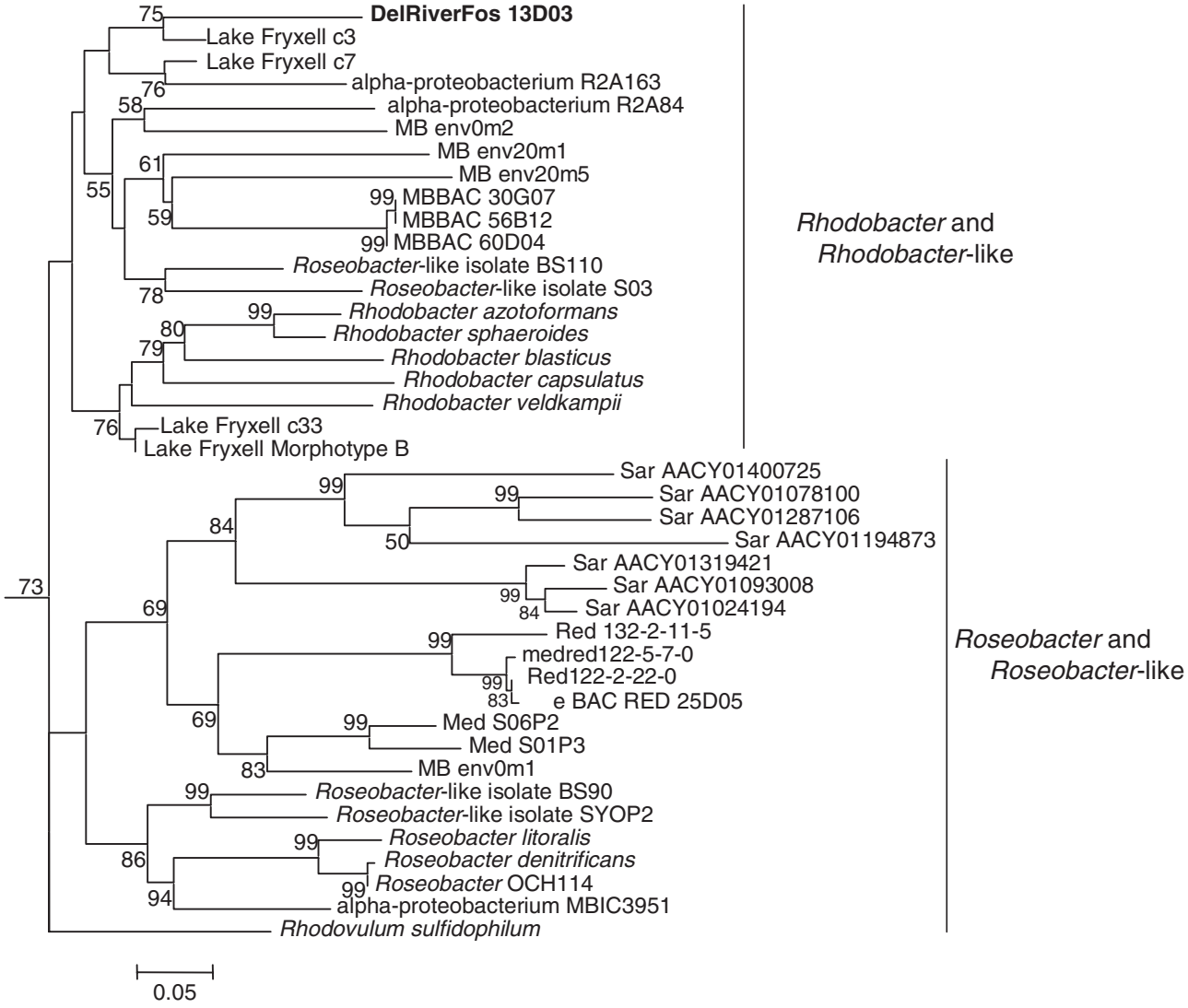

Fig. 5. Phylogenetic analysis of *puf*M in the alpha-3-proteobacterial cluster. Sargasso Sea clones are abbreviated 'Sar' followed by the nucleotide accession number. Lake Fryxell sequences are from Karr *et al.* (2003). The scale bar represents five nucleotide substitutions per 100 positions.

2002). Reaction centre genes do not form the same clusters as pigment biosynthesis genes in phylogenetic analyses (Xiong *et al.*, 2000). The data summarized in Fig. 1 provide further evidence of these rearrangements.

A complication in this analysis is the lack of genomic data for AAP beta-proteobacteria from freshwater environments. The sequences of most of these genes from freshwater or estuarine AAP bacteria are not available. Other than *puf*M, no photosynthesis genes of uncultured freshwater AAP representatives have been sequenced to date.

Overall, the genes in the beta-proteobacterial fosmid are most similar to bacteria of the beta-I-proteobacterial cluster (Glöckner *et al.*, 2000). This cluster is also known as the strictly freshwater *Rhodoferax* cluster, in the Comamonadaceae clade of beta-proteobacteria (Zwart *et al.*, 2002). Of the beta-proteobacterial 16S rRNA sequences in the Delaware River fosmid library, most belonged to the Comamonadaceae clade (Cottrell *et al.*, 2005). Representative AAP Comamonadaceae bacteria include *Rts. depolymerans*, *Rvi. gelatinosus, Rfx. fermentans* and the Crater Lake beta-proteobacterium HTCC528. This fosmid provides clues to a strictly freshwater cluster of AAP beta-proteobacteria. Analyses of a *puf*M polymerase chain reaction (PCR) library place 88% of uncultured AAP bacteria in the Delaware River in this cluster (L. A. Waidner and D. L. Kirchman, in preparation).

Gene order of the puf *operon*

We analysed gene order (synteny) to gain further insights into the AAP genes of the fosmid clones. The *puf* operon of the beta-proteobacterial fosmid consists of *puf* B-A-L-

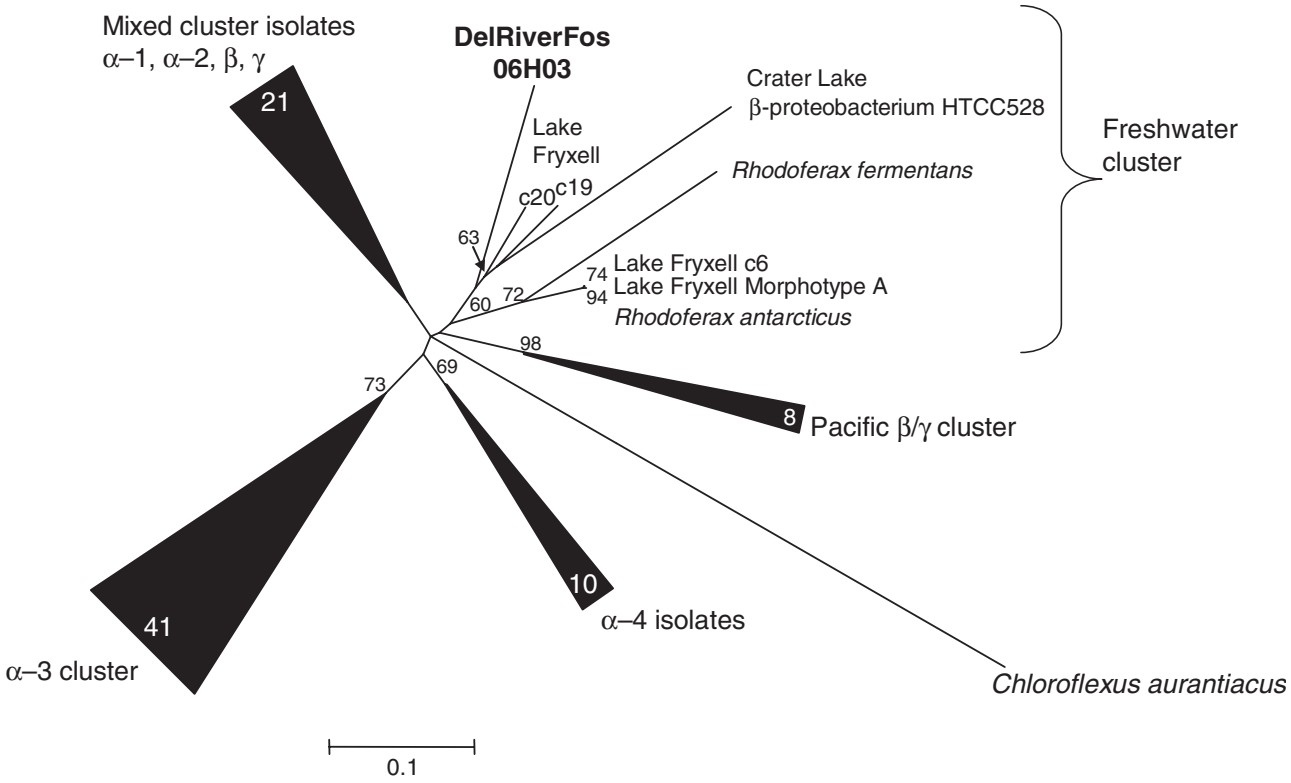

Fig. 6. Phylogenetic relationships of *puf*M genes. Clone numbers of Lake Fryxell uncultured bacteria *puf*M are indicated 'Lake Fryxell *c*'. The scale bar represents 10 nucleotide substitutions per 100 positions. The alpha-3-proteobacterial cluster containing 41 taxa is expanded in Fig. 5. *Chloroflexus aurantiacus* was the outgroup.

M-C (Table 3). The six representatives of cultured AAP bacteria with this operon are in the mixed proteobacterial cluster based on 16S rRNA phylogeny (Table 3). While B-A-L-M-C is common in cultured AAP bacteria (Tuschak, 2003), the Delaware River beta-proteobacterial fosmid is the only known uncultured representative to have this operon organization. Like the beta-proteobacterial clone, the alpha-3-proteobacterial fosmid contains a *puf* gene order also unique among uncultured AAP bacteria (Table 3). This clone contains *puf* Q-B-A-L-M-X, an organization seen so far only in *Rba. capsulatus* and *Rba. sphaeroides* (Tuschak, 2003).

To further investigate the relationships among AAP bacterial *puf* operons, we examined gene order using a clustering analysis (Fig. 7). The same approach has been applied to ribosomal protein operons (Coenye and Vandamme, 2005). In this analysis, there are two major groups of AAP bacteria. The *puf* operon of the beta-proteobacterial fosmid (B-A-L-M-C) clusters with those of mixed proteobacterial AAP bacteria, containing alpha-1, alpha-2, beta- and gamma-proteobacteria.

The second major cluster of *puf* operons contains alpha-3 and alpha-4-proteobacteria and uncultured proteobacteria with *puf*M genes most closely related to those of alpha-3 proteobacteria (Fig. 7). The organization of the *puf* genes in a Red Sea BAC clone is the same as that of a *Roseobacter*-like isolate, *Rhodovulum sulfidophilum*, which is the only cultured representative with that gene order (Tsukatani *et al*., 2004; Oz *et al*., 2005). The operon of our alpha-3-proteobacterial fosmid, however, is distinct from those in marine *Roseobacter*-like bacteria. Its gene order is unique to the *Rhodobacter* genus.

A third minor cluster is composed of the *puf* operons in which the B-A (light-harvesting I) genes are downstream of *puf*L-M (Fig. 7). This type of *puf* operon has been observed to date only in cultured gamma-proteobacteria and in two Pacific coastal water BAC clones (Beja *et al*., 2002).

This analysis places the beta-proteobacterial fosmid in the mixed cluster of AAP bacteria, which is supported by phylogenetic analyses of some genes (Fig. 1). In contrast, analyses of other genes, such as *crt*E-F and *puf*L-M, suggest that this clone is in a separate freshwater cluster, distinct from the mixed cluster of AAP bacteria (Fig. 6 and Fig. S1). Furthermore, this operon sequence (B-A-L-M-C) is conserved among the AAP representatives in the Comamonadaceae group of beta-proteobacteria. Two other beta-proteobacteria, *Rts. depolymerans* and *Rvi. gelatinosus*, also have the B-A-L-M-C gene order (Fig. 7).

Table 3. Gene orders in *puf* operons of cultured and uncultured proteobacteria.

Operon	Cultured representative	Accession	Uncultured representative	Accession	Phylogeny[a]
B-A-L-M-C-B-A-B-A	*Allochromatium vinosum*	AB050620/AB011811	–	–	Gamma
	Lamprocystis purpurea	AY177752	–	–	Gamma
L-M-C-B-A	–	–	MB BAC29C02 [c]	AE008920	Beta
	–	–	MB BAC65D09 [c]	AE008919	Beta
B-A-L-M	*Rhodospirillum rubrum*	JGI drafts contig 129[b]	–	–	Alpha-1
	Bradyrhizobium sp. ORS278	AF182374	–	–	Alpha-2
	Rhodopseudomonas palustris	NC_005296	–	–	Alpha-2
B-A-L-M-C	–	–	DelawareRiver 06H03	AY912082	Beta
	Rvi. gelatinosus	AB034704	–	–	Beta
	Rts. depolymerans	AB028938	–	–	Beta
	Rhodospirillum molischianum	D50654	–	–	Alpha-1
	Acidiphilium angustum	AB005219	–	–	Alpha-1
	Roseospirillum parvum	AY242845	–	–	Alpha-2
	Ectothiorhodospira shaposhnikovii	AF018955	–	–	Gamma
Q-K-B-A-L-M-X	*Rba. sphaeroides*	AJ010302			Alpha-3
Q-B-A-L-M-X	–	–	DelawareRiver 13D03	AY912081	Alpha-3
	Rba. capsulatus	Z11165	–	–	Alpha-3
Q-B-A-L-M	–	–	MB BAC60D04[c]	AE008921	Alpha-3
	Erb. longus	AB035570	–	–	Alpha-4
Q-B-A-L-M-C	–	–	eBACred25D05[c]	AY671989	Alpha-3
	Rhodovulum sulfidophilum	AB020784	–	–	Alpha-3

a. Phylogenetic affiliations are based on 16S rRNA for cultured and on *puf*M for uncultured organisms. Alpha-1, alpha-2, alpha-3, alpha-4, beta and gamma designations refer to the subclasses of proteobacteria.
b. *Rhodospirillum rubrum* genome is available at http://genome.jgi-psf.org/microbial/index.html
c. MB BAC represents Monterey Bay BAC clones (Beja *et al*., 2002), and eBACred is from the Red Sea uncultured proteobacterium (Oz *et al*., 2005).

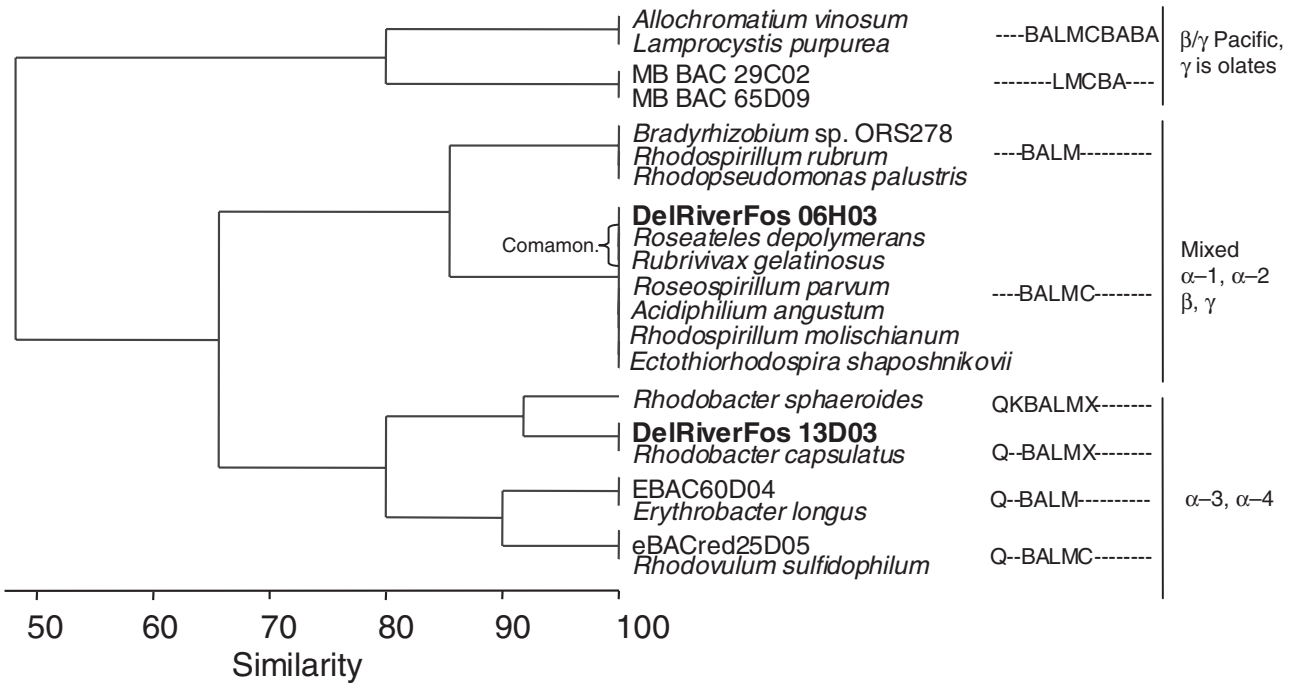

Fig. 7. Similarity dendrogram of gene orders and content of *puf* operons. Gene orders of the *puf* operon are indicated to the right of each group. Phylogenetic assignments were based on 16S rRNA for cultured organisms and on the *puf*M gene for uncultured proteobacteria. Comamon., Comamonadaceae group of the beta-proteobacteria.

There are several explanations for the apparent discrepancies in these results. First, genomic data from freshwater AAP bacteria are lacking. The only freshwater AAP bacterium whose *puf* operon has been sequenced is *Rts. depolymerans*, which was isolated from a river (Suyama *et al*., 2002). In addition, the phylogenetic resolution of the synteny analysis is not as great as those of individual gene or protein analyses. For example, the *puf*M tree is based on 633 nucleotide positions, with four possibilities per position (G/A/T/C). The relationships among these *puf* operons are based on 12 positions, with only two possibilities (1/0) at each position. Finally, these data could be explained by horizontal gene transfers of the photosynthesis operons between AAP bacteria in the freshwater and mixed clusters as discussed above.

Conclusions

These Delaware River clones may be representative of AAP bacteria in oxic freshwaters. The metagenomic data indicate that these AAP bacteria are aerobic, and that one type is a *Rhodobacter*-like bacterium, while the other is a beta-proteobacterium in the *Rhodoferax* clade. Both clones likely are from photosynthetically competent bacteria because they contained most of the photosynthesis genes needed for AAP metabolism. The *puf*M of the beta-proteobacterial clone and of other freshwater AAP bacteria form a freshwater cluster. The genomic information from the beta-proteobacterial clone provides hints of lateral transfers of photosynthesis genes between AAP bacteria in the freshwater and mixed clusters. Overall, the data from these gene clusters suggest the existence of groups of AAP bacteria different from what have been observed so far in marine habitats.

Experimental procedures

Screening the fosmid library for photosynthesis operons

The construction of the library was outlined in detail previously (Cottrell *et al*., 2005). In brief, we constructed a fosmid library in the pCC1-Fos vector (Epicentre, Madison, WI) from a December 2001 sample of picoplankton (< 0.8 μm) DNA from surface water in the Delaware River near Trenton, NJ (40°7.69′N, 74°49.05′W), 198 km from the mouth of the estuary.

Pools of 96 clones from each microtitre plate were used as templates for PCR after induction with arabinose. Clones were grown individually in microtitre plates in 100 μl of Luria–Bertani (LB) with 12.5 μg ml^{-1} chloramphenicol at 37°C for 16 h. The cultures were then pooled and diluted in LB with chloramphenicol and induction solution (Epicentre) and grown further for 5 h at 37°C. The DNA of the pooled clones was isolated manually by alkaline lysis, precipitated with ethanol, and resuspended in 50 μl of purified water. Yields ranged from 100 to 1000 ng μl^{-1}. We screened the pooled DNA for the *pufM* gene by PCR amplification using primers *pufM*R and *puf*LF (Achenbach *et al*., 2001). The pooled DNA (100 ng) was used in 25 μl PCR reactions, with 0.5 U *Taq* DNA polymerase and 1X buffer (Promega, Madison, WI), 0.2 mM dNTPs, 2.5 mM $MgCl_2$, 0.2 mg ml^{-1} bovine serum albumin (Sigma A-7030, Sigma-Aldrich, St. Louis, MO) and 0.2 μM each primer. The reactions were subjected to 30 rounds of amplification (94°C 1 min, 58°C 0.5 min, 72°C 2 min).

Two pools resulted in positive 1.5-kb PCR products, the predicted size of the *puf*L-M gene amplicon. To confirm the presence of the *puf*M genes, the two 1.5-kb PCR products were subcloned into vector pCR2.1 (Invitrogen, Carlsbad CA) as per the manufacturer's instructions and partially sequenced using vector primers M13F and M13R.

Sequencing and sequence analyses

The inserts of the two *puf*M-bearing clones (33 and 35 kb) were fully sequenced by the Joint Genome Institute of the Department of Energy by shotgun cloning and sequencing. The insert sequences were automatically annotated using GLIMMER (Delcher *et al*., 1999). The annotations were also manually edited.

To determine the protein most closely related to each ORF in the fosmid insert sequences, the nucleotide sequences were compared with sequences in the protein database using BLAST-X version 2.2.9 (Altschul *et al*., 1997). To confirm the start codons of each ORF computed by GLIMMER's automated annotation, we downloaded all related protein sequences of each gene and manually aligned with the fosmid protein sequences using CLUSTAL-W (Thompson *et al*., 1994).

Phylogenetic and molecular evolutionary analyses of protein sequences were conducted using MEGA version 3.0 (Kumar *et al*., 2004). Phylogenetic trees were formed using the consensus of 1050 replications of Minimum Evolution calculations, using the p-distance model.

For nucleotide analyses of *puf*M, *puf*L and other genes, sequences were converted to protein sequences and aligned with CLUSTAL-W using MegAlign version 5.08 (DNAStar, Madison, WI). The alignment of *puf*M was based on 633 nucleotide positions of 74 publicly available sequences and the two Delaware River fosmids. Twelve short sequences (227 bp of C-terminal sequence) from the Lake Fryxell study (Karr *et al*., 2003) were added to the alignment. The *puf*L alignment was based on 486 nucleotide sites of 58 taxa, including the two fosmid *puf*L genes. Alignments were converted nucleotide sequences, converted to MEGA format, keeping predefined alignment gaps, and imported into the MEGA package (Kumar *et al*., 2004). Evolutionary distances and phylogenetic trees were constructed in MEGA, using the Kimura 2-parameter nucleotide model, with pairwise deletion of gaps. Bootstrap values were based on 1050 replications.

Gene order analysis of puf operon

The *puf* gene orders among 21 cultured and uncultured anoxygenic photosynthetic bacteria were compared using a method described by Coenye and Vandamme, 2005). In brief, genes were considered as either present or absent (1 or 0),

arrayed into three separate gene orders, and then subjected to cluster analysis by using the Ward clustering parameters (Ward, 1963) using the PAST program (Hammer *et al*., 2001). This method applies equal weight to each gene (binary character), regardless of repeat number, and regardless of the character at any position. The accession numbers for the operons in this analysis are provided in Table 3. The presence/absence matrices were produced for each taxon, with a '1' for gene presence and '0' for absence using three gene orders: *puf* Q-K-B-A-L-M-C-X-B-A-B-A, *puf* A-B-A-B-C-X-M-L-A-B-K-Q, or *puf* Q-K-B-A-L-M-X-C-B-A-B-A. The three resulting dendrograms results were identical. Phylogenetic designations of Monterey Bay BAC clones 60D04, 65D09, 29C02 and eBACred25D05 were based on *pufM* phylogenetic analyses (Beja *et al*., 2002; Oz *et al*., 2005).

Accession numbers

The nucleotide sequences of the fosmid inserts were deposited in GenBank with Accession numbers AY912081 and AY912082. Accession numbers of protein and nucleotide sequences used in phylogenetic analyses are provided in *Supplementary material* Table S3.

Acknowledgements

The sequencing and assembly of the fosmid clones was performed by the production sequencing group at the DOE/Joint Genome Institute through the Sequence-For-Others Program under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program and the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48, Lawrence Berkeley National Laboratory under contract No. DE-AC03-76SF00098 and Los Alamos National Laboratory under contract No. W-7405-ENG-36. Preliminary annotations were performed by The Institute for Genomic Research through funding by the Department of Energy. We thank Tom Hanson for several helpful suggestions on synteny and phylogenetic analyses. Muktak Aklujkar and Tom Beatty graciously provided insight into photosynthesis operon organization. This work was supported by DOE Grant DE-FG02-01ER63142 and by NSF OCE 9908808.

References

Achenbach, L.A., Carey, J., and Madigan, M.T. (2001) Photosynthetic and phylogenetic primers for detection of anoxygenic phototrophs in natural environments. *Appl Environ Microbiol* **67:** 2922–2926.

Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

Bauer, C.E., Buggy, J.J., Yang, Z.M., and Marrs, B.L. (1991) The superoperonal organization of genes for pigment biosynthesis and reaction center proteins is a conserved feature in *Rhodobacter capsulatus* – analysis of overlapping bchB and puhA transcripts. *Mol Gen Genet* **228:** 433–444.

Beja, O., Suzuki, M.T., Heidelberg, J.F., Nelson, W.C., Preston, C.M., Hamada, T., *et al.* (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415:** 630–633.

Choudhary, M., and Kaplan, S. (2000) DNA sequence analysis of the photosynthesis region of *Rhodobacter sphaeroides* 2.4.1. *Nucleic Acids Res* **28:** 862–867.

Coenye, T., and Vandamme, P. (2005) Organisation of the S10, spc and alpha ribosomal protein gene clusters in prokaryotic genomes. *FEMS Microbiol Lett* **242:** 117–126.

Cottrell, M.T., Waidner, L.A., Yu, L., and Kirchman, D.L. (2005) Bacterial diversity of metagenomic and PCR libraries from the Delaware River. *Environ Microbiol* (in press). 10.1111/j.1462-2920.2005.00762.x.

Crump, B.C., Armbrust, E.V., and Baross, J.A. (1999) Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia River, its estuary, and the adjacent coastal ocean. *Appl Environ Microbiol* **65:** 3192–3204.

Crump, B.C., Hopkinson, C.S., Sogin, M.L., and Hobbie, J.E. (2004) Microbial biogeography along an estuarine salinity gradient: combined influences of bacterial growth and residence time. *Appl Environ Microbiol* **70:** 1494–1505.

Dandekar, T., Snel, B., Huynen, M., and Bork, P. (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci* **23:** 324–328.

Delcher, A.L., Harmon, D., Kasif, S., White, O., and Salzberg, S.L. (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27:** 4636–4641.

Glöckner, F.O., Zaichikov, E., Belkova, N., Denissova, L., Pernthaler, J., Pernthaler, A., and Amann, R. (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.

Gomelsky, L., Sram, J., Moskvin, O.V., Horne, I.M., Dodd, H.N., Pemberton, J.M., *et al.* (2003) Identification and *in vivo* characterization of PpaA, a regulator of photosystem formation in *Rhodobacter sphaeroides*. *Microbiology (UK)* **149:** 377–388.

Gonzalez, J.M., and Moran, M.A. (1997) Numerical dominance of a group of marine bacteria in the alpha-subclass of the class Proteobacteria in coastal seawater. *Appl Environ Microbiol* **63:** 4237–4242.

Hammer, Ø., Harper, D.A.T., and Ryan, P.D. (2001) PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica* **4:** 9pp.

Hansen, T.A., and Imhoff, J.F. (1985) *Rhodobacter veldkampii*, a new species of phototrophic purple nonsulfur bacteria. *Int J Syst Bacteriol* **35:** 115–116.

Igarashi, N., Harada, J., Nagashima, S., Matsuura, K., Shimada, K., and Nagashima, K.V.P. (2001) Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria. *J Mol Evol* **52:** 333–341.

Karl, D.M. (2002) Microbiological oceanography – hidden in a sea of microbes. *Nature* **415:** 590–591.

Karr, E.A., Sattley, W.M., Jung, D.O., Madigan, M.T., and Achenbach, L.A. (2003) Remarkable diversity of phototrophic purple bacteria in a permanently frozen Antarctic lake. *Appl Environ Microbiol* **69:** 4910–4914.

Koblizek, M., Beja, O., Bidigare, R.R., Christensen, S., Benitez-Nelson, B., Vetriani, C., *et al.* (2003) Isolation and

characterization of *Erythrobacter* sp strains from the upper ocean. *Arch Microbiol* **180:** 327–338.

Koblizek, M., Shih, J.D., Breitbart, S.I., Ratcliffe, E.C., Kolber, Z.S., Hunter, C.N., and Niederman, R.A. (2005) Sequential assembly of photosynthetic units in *Rhodobacter sphaeroides* as revealed by fast repetition rate analysis of variable bacteriochlorophyll a fluorescence. *Biochim Biophys Acta* **1706:** 220–231.

Kolber, Z.S., Van Dover, C.L., Niederman, R.A., and Falkowski, P.G. (2000) Bacterial photosynthesis in surface waters of the open ocean. *Nature* **407:** 177–179.

Kumar, S., Tamura, K., and Nei, M. (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings Bioinformatics* **5:** 150–163.

Lawrence, J.G., and Roth, J.R. (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* **143:** 1843–1860.

Layer, G., Verfurth, K., Mahlitz, E., and Jahn, D. (2002) Oxygen-independent coproporphyrinogen-III oxidase *hem*N from *Escherichia coli*. *J Biol Chem* **277:** 34136–34142.

Lecompte, O., Ripp, R., Thierry, J.-C., Moras, D., and Poch, O. (2002) Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res* **30:** 5382–5390.

Makarova, K., Ponomarev, V., and Koonin, E. (2001) Two C or not two C: recurrent disruption of Zn-ribbons, gene duplication, lineage-specific gene loss, and horizontal gene transfer in evolution of bacterial ribosomal proteins. *Genome Biology*. doi: 10.1186/gb-2001-2-9-research0033.

Mullins, T.D., Britschgi, T.B., Krest, R.L., and Giovannoni, S.J. (1995) Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. *Limnol Oceanogr* **40:** 148–158.

Nagashima, K.V.P., Hiraishi, A., Shimada, K., and Matsuura, K. (1997) Horizontal transfer of genes coding for the photosynthetic reaction centers of purple bacteria. *J Mol Evol* **45:** 131–136.

Ouchane, S., Steunou, A.S., Picaud, M., and Astier, C. (2004) Aerobic and anaerobic Mg-protoporphyrin monomethyl ester cyclases in purple bacteria – a strategy adopted to bypass the repressive oxygen control system. *J Biol Chem* **279:** 6385–6394.

Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G.D., and Maltsev, N. (1999) The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci USA* **96:** 2896–2901.

Oz, A., Sabehi, G., Koblizek, M., Massana, R., and Beja, O. (2005) *Roseobacter*-like bacteria in Red and Mediterranean Sea aerobic anoxygenic photosynthetic populations. *Appl Environ Microbiol* **71:** 344–353.

Page, K.A., Connon, S.A., and Giovannoni, S.J. (2004) Representative freshwater bacterioplankton isolated from Crater Lake, Oregon. *Appl Environ Microbiol* **70:** 6542–6550.

Panek, H., and O'Brian, M.R. (2002) A whole genome view of prokaryotic haem biosynthesis. *Microbiology (UK)* **148:** 2273–2282.

Pinta, V., Picaud, M., Reiss-Husson, F., and Astier, C. (2002) *Rubrivivax gelatinosus acs*F (previously orf358) codes for a conserved, putative binuclear-iron-cluster-containing protein involved in aerobic oxidative cyclization of Mg-protoporphyrin IX monomethylester. *J Bacteriol* **184:** 746–753.

Pinta, V., Ouchane, S., Picaud, M., Takaichi, S., Astier, C., and Reiss-Husson, F. (2003) Characterization of unusual hydroxy- and ketocarotenoids in *Rubrivivax gelatinosus*: involvement of enzyme crtF or crtA. *Arch Microbiol* **179:** 354–362.

Raymond, J., Zhaxybayeva, O., Gogarten, J.P., Gerdes, S.Y., and Blankenship, R.E. (2002) Whole-genome analysis of photosynthetic prokaryotes. *Science* **298:** 1616–1620.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., *et al.* (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66:** 2541–2547.

Schwalbach, M.S., and Fuhrman, J.A. (2005) Wide-ranging abundances of aerobic anoxygenic phototrophic bacteria in the world ocean revealed by epifluorescence microscopy and quantitative PCR. *Limnol Oceanogr* **50:** 620–628.

Suyama, T., Shigematsu, T., Suzuki, T., Tokiwa, Y., Kanagawa, T., Nagashima, K.V.P., and Hanada, S. (2002) Photosynthetic apparatus in *Roseateles depolymerans* 61A is transcriptionally induced by carbon limitation. *Appl Environ Microbiol* **68:** 1665–1673.

Suzuki, M.T., Rappe, M.S., Haimberger, Z.W., Winfield, H., Adair, N., Strobel, J., and Giovannoni, S.J. (1997) Bacterial diversity among small-subunit rRNA gene clones and cellular isolates from the same seawater sample. *Appl Environ Microbiol* **63:** 983–989.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL-W – improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22:** 4673–4680.

Tsukatani, Y., Matsuura, K., Masuda, S., Shimada, K., Hiraishi, A., and Nagashima, K.V.P. (2004) Phylogenetic distribution of unusual triheme to tetraheme cytochrome subunit in the reaction center complex of purple photosynthetic bacteria. *Photosynth Res* **79:** 83–91.

Tuschak, C. (2003) Variability of photosynthesis genes in purple bacteria and molecular determinants of their spectral properties. In *Univ Oldenburg Dissertations*. Oldenburg, Germany: Univ. of Oldenburg, pp. 36–38.

Ward, J.H. (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* **58:** 236–244.

Xiong, J., Fischer, W.M., Inoue, K., Nakahara, M., and Bauer, C.E. (2000) Molecular evidence for the early evolution of photosynthesis. *Science* **289:** 1724–1730.

Young, D.A., Bauer, C.E., Williams, J.C., and Marrs, B.L. (1989) Genetic evidence for superoperonal organization of genes for photosynthetic pigments and pigment-binding proteins in *Rhodobacter capsulatus*. *Mol Gen Genet* **218:** 1–12.

Yurkov, V.V., and Beatty, J.T. (1998) Aerobic anoxygenic phototrophic bacteria. *Microbiol Mol Biol Rev* **62:** 695–724.

Zsebo, K.M., and Hearst, J.E. (1984) Genetic physical mapping of a photosynthetic gene-cluster from r-capsulata. *Cell* **37:** 937–947.

Zwart, G., Crump, B.C., Agterveld, M., Hagen, F., and Han, S.K. (2002) Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat Microb Ecol* **28:** 141–155.

Supplementary material

The following supplementary material is available for this article online:

Table S1. Annotation of beta-proteobacterial fosmid clone DelRiverFos06H03.

Table S2. Annotation of alpha-3-proteobacterial clone DelRiverFos13D03.

Table S3. Accession numbers of sequences used in phylogenetic analyses.

Fig. S1. Phylogenetic analyses of *crt*E, *crt*F, *bch*C, *puf* L and *puf* B.

Fig. S2. Phylogenetic analyses of *bch*X, *puf* C, *crt*D and *acs*F.