

Table S1. Analysis of open reading frames of beta-proteobacterial fosmid clone DelRiverFos06H03. Most similar homologues of each open reading frame and percent identical and similar amino acids was determined by BLASTX. Start codon of each gene is in location column, with complement DNA sequences indicated by c(xxx).

ORF	Location	Length (bp)	Gene name	Gene function	Accession #	% Identical	% Similar	Organism ^a
<u>photosynthesis genes</u>								
1	3	159	bchM (partial)	Mg protoporphyrin IX monomethyl ester oxidative cyclase subunit	BAA94055	54	66	<i>Rvi. gelatinosus</i>
2	176	1443	lhaA	COG0477: Permeases of the major facilitator superfamily; Light-harvesting I (B870) complex assembly protein	T50901	53	66	<i>Rvi. gelatinosus</i>
3	1650	765	puhA	H subunit of photosynthetic reaction center complex	BAA94053	63	73	<i>Rvi. gelatinosus</i>
6	3609	1083	acsF	Mg-protoporphyrin IX monomethylester aerobic cyclization system	AAL25840	69	83	<i>Rvi. gelatinosus</i>
8	5476	1578	crtI	Phytoene dehydrogenase	AAO93135	64	77	<i>Rvi. gelatinosus</i>
9	7133	1059	crtB	Prephytoene pyrophosphate synthase	T50895	54	65	<i>Rvi. gelatinosus</i>
10	8209	444	bluF, abcC	ABC-type transport, ATPase component	ZP_00242636	43	55	<i>Rvi. gelatinosus</i>
13	c(10439)	1398	pucC	Chlorophyll Major Facilitator Superfamily (MFS) exporter	CAB52412	57	70	<i>Rvi. gelatinosus</i>
14	c(10690)	198	pucA	Light-harvesting complex II protein b-800/850, alpha chain	S66225	57	77	<i>Rba. sulfidophilus</i> ^b
15	c(11823)	864	crtC	Hydroxyneurosporene synthase	T50894	56	73	<i>Rvi. gelatinosus</i>
16	c(13352)	1533	crtD	Methoxyneurosporene dehydrogenase	T50893	51	65	<i>Rvi. gelatinosus</i>
17	c(14172)	828	crtA	Spheroidene/spirilloxanthin monooxygenase	T50892	45	54	<i>Rvi. gelatinosus</i>
18	c(15265)	1062	pufC	Photosynthetic reaction center cytochrome subunit	BAB19670	65	76	<i>Rts. depolymerans</i>
19	c(16239)	978	pufM	Photosynthetic reaction center M subunit	BAB19668	67	76	<i>Rts. depolymerans</i>
20	c(17124)	837	pufL	Photosynthetic reaction center L subunit	P51760	69	75	<i>Rvi. gelatinosus</i>
21	c(17417)	210	pufA	Light-harvesting LHI alpha subunit	AAO93121	66	77	<i>Rvi. gelatinosus</i>
22	c(17652)	225	pufB	pufB homolog; light-harvesting LHI beta subunit	C49964	40	51	<i>Rvi. gelatinosus</i>
23	c(19507)	1527	bchZ	Chlorophyllide reductase subunit Z	T50884	79	89	<i>Rvi. gelatinosus</i>
24	c(20988)	1485	bchY	Chlorophyllide reductase, Y subunit	T50883	70	81	<i>Rvi. gelatinosus</i>
25	c(22107)	1017	bchX	Chlorophyllide reductase, X subunit	T50882	77	84	<i>Rvi. gelatinosus</i>
26	c(23057)	954	bchC	2-alpha-hydroxy ethyl bacteriochlorophyllide oxidase	T50881	57	72	<i>Rvi. gelatinosus</i>
27	c(24315)	1221	crtF	Hydroxyneurosporene methyltransferase	T50880	40	54	<i>Rvi. gelatinosus</i>
28	c(25241)	915	crtE	Geranylgeranyl pyrophosphate synthetase	AAO93113	55	66	<i>Rvi. gelatinosus</i>

Table S1, continued

<u>non-photosynthesis genes</u>								
4	2411	693	4	Hypothetical membrane protein similar to <i>Rvi. gelatinosus</i> ORF227	AAO93127	43	60	<i>Rvi. gelatinosus</i>
5	3100	480	5	Conserved hypothetical protein similar to <i>Rvi. gelatinosus</i> ORF154	T50898	46	53	<i>Rvi. gelatinosus</i>
7	4685	783	7	Hypothetical membrane protein similar to <i>Rvi. gelatinosus</i> ORF276	AAO93125	42	58	<i>Rvi. gelatinosus</i>
11	c(8838)	129	11	Hypothetical protein	none	none	none	none
12	c(8981)	114	12	Hypothetical protein	none	none	none	none
29	c(25504)	180	29	Hypothetical protein	none	none	none	none
30	25421	180	30	Hypothetical protein	none	none	none	none
31	c(27348)	1752	ABC_ATPase	COG1132: ABC-type multidrug transport, ATPase and permease components	ZP_00361603	63	74	<i>Polaromonas</i> sp. JS666
32	c(27474)	120	32	Hypothetical protein	none	none	none	none
33	27591	909	hemF	COG0408: Coproporphyrinogen III oxidase, aerobic oxidoreductase protein	ZP_00362870	72	83	<i>Polaromonas</i> sp. JS666
34	28525	1002	hemB	COG0113: Delta-aminolevulinic acid dehydrogenase	ZP_00360714	80	86	<i>Polaromonas</i> sp. JS666
35	29543	945	hemC	COG0181: Porphobilinogen deaminase	ZP_00364864	64	74	<i>Polaromonas</i> sp. JS666
36	c(31806)	1308	hemL	Glutamate-1-semialdehyde 2,1-aminomutase	NP_070069	34	54	<i>Archaeoglobus fulgidus</i> ^c
37	c(32702)	900	mmsB, partial	3-hydroxyisobutyrate dehydrogenase	BAB49696	38	57	<i>Mesorhizobium loti</i> ^d

(a) All are beta-proteobacteria, except where noted: (b) alpha-3-proteobacteria (*Rhodobacteraceae*) (c) Archaea (*Euryarchaeota*) and (d) alpha-2-proteobacteria (*Rhizobiales*). Gene abbreviations: light harvesting (*pufB-A*), reaction centers (*pufL-M* and *puhA*), carotenoid synthesis (*crt*), bacteriochlorophyll synthesis (*bch*) Genera abbreviations: *Rubrivivax* (*Rvi.*); *Rhodobacter* (*Rba.*); *Roseateles* (*Rts.*).

Table S2: BLAST-X analysis of open reading frames of the alpha-3-proteobacterial fosmid clone DelRiverFos13D03. Start codon of each gene is in location column, with complement DNA sequences indicated by c(xxx).

ORF	Location	Length (bp)	Gene name	Gene function	Accession #	% Identical	% Similar	Organism ^a
<u>photosynthesis genes</u>								
1	c(556)	556	ppaA, partial	Palindromic regulatory element; photosynthesis gene regulator	CAB38727	41	62	<i>Rba. sphaeroides</i>
2	878	4989	bchF	2-vinyl bacteriochlorophyllide hydratase	CAA77527	68	75	<i>Rba. capsulatus</i>
3	1384	1272	bchN	Light-independent protochlorophyllide reductase, N subunit	AAF24275	78	83	<i>Rba. sphaeroides</i>
4	2732	1587	bchB	Light-independent protochlorophyllide reductase, B subunit	CAB38724	72	83	<i>Rba. sphaeroides</i>
5	4308	3564	bchH	Mg-protoporphyrin IX chelatase subunit H	AAF24273	71	83	<i>Rba. sphaeroides</i>
6	7954	900	bchL	Light-independent protochlorophyllide reductase, iron sulfur ATP-binding	AAM48678	77	83	MB BAC60D04 ^b
7	8853	666	bchM	Magnesium-protoporphyrin IX methyltransferase C-terminus	CAB38721	74	81	<i>Rba. sphaeroides</i>
8	9548	1410	lhaA	COG0477: Permeases of the major facilitator superfamily; Light-harvesting I (B870) complex assembly protein	AAF24270	63	76	<i>Rba. sphaeroides</i>
9	10970	774	puhA	Subunit H of reaction centre	CAA44976	65	77	<i>Rba. sphaeroides</i>
13	13149	1098	acsF	Aerobic Cyclase System, Fe-containing subunit	ZP_00005239	78	88	<i>Rba. sphaeroides</i>
16	16394	423	cycA	Cytochrome c2	AAM48665	52	66	MB BAC60D04 ^b
18	c(19171)	228	pufX	Putative intrinsic membrane protein	BAD44690	44	55	<i>Rba. blasticus</i>
19	c(20103)	921	pufM	Photosynthetic reaction center M subunit	AAM48658	68	78	MB BAC60D04 ^b
20	c(20944)	849	pufL	Photosynthetic reaction center L subunit	15826552	61	72	<i>Rba. sphaeroides</i>
21	c(21260)	192	pufA	Light-harvesting protein B-870 (LH-1), alpha chain	CAA77553	77	92	<i>Rba. capsulatus</i>
22	c(21423)	150	pufB	Light-harvesting protein B-870 (LH1) beta chain	LBRFAS	81	93	<i>Rba. sphaeroides</i>
23	c(21762)	222	pufQ	pufQ protein	P16069	42	56	<i>Rba. sphaeroides</i>
24	c(23237)	1479	bchZ	Chlorophyllide reductase subunit Z	T50755	85	92	<i>Rba. sphaeroides</i>
25	c(24805)	1569	bchY	Protochlorophyllide reductase Y subunit	T50754	75	83	<i>Rba. sphaeroides</i>
26	c(25870)	1005	bchX	Chlorophyllide reductase, X subunit	AAM48653	74	79	MB BAC60D04 ^b
27	c(26820)	954	bchC	2-alpha-hydroxy ethyl bacteriochlorophyllide oxidase	CAA77547	56	65	<i>Rba. capsulatus</i>
28	c(28097)	1155	crtF	Hydroxyneurosporene methyltransferase	ZP_00207380	57	67	<i>Rba. sphaeroides</i>
29	c(28959)	882	crtE	Geranylgeranyl pyrophosphate synthetase	P54976	75	86	<i>Rba. sphaeroides</i>
30	29098	1437	crtD	Methoxyneurosporene dehydrogenase	ZP_00207382	55	66	<i>Rba. sphaeroides</i>
31	30567	876	crtC	Hydroxyneurosporene dehydrogenase	AAM48648	62	74	MB BAC60D04 ^b
37	c(34906)	321	fnrL, partial	Transcriptional activator protein fnrL	P51007	76	87	<i>Rba. sphaeroides</i>

Table S2, Continued

<u>non-photosynthesis genes</u>								
10	11740	633	10	Conserved hypothetical membrane protein similar to <i>Rba. sphaeroides</i> ORF03001158 and <i>Rvi. gelatinosus</i> ORF 227	ZP_00207585	49	65	<i>Rba. sphaeroides</i>
11	12377	456	11	Conserved hypothetical membrane protein similar to <i>Rba. sphaeroides</i> ORF128 and <i>Rvi. gelatinosus</i> ORF154	AAF24267	51	69	<i>Rba. sphaeroides</i>
12	12856	297	12	Conserved hypothetical protein similar to <i>Rba. sphaeroides</i> ORF03001156	ZP_00005240	78	93	<i>Rba. sphaeroides</i>
14	14316	807	14	Hypothetical protein similar to <i>Rba. capsulatus</i> ORF274 and <i>Rvi. gelatinosus</i> ORF276	CAA77516	48	61	<i>Rba. capsulatus</i>
15	15112	1212	hemA	5-aminolevulinic acid synthase (ALAS)	AAM48669	68	80	MB BAC60D04 ^b
17	c(18775)	1896	dxs	Deoxyxylulose-5-phosphate synthase	ZP_00006273	72	83	<i>Rba. sphaeroides</i>
32	c(31698)	222	rpmE	Ribosomal protein L31	ZP_00006766	78	94	<i>Rba. sphaeroides</i>
33	c(32084)	375	rplS	Ribosomal protein L19	ZP_00006765	83	84	<i>Rba. sphaeroides</i>
34	c(32357)	144	34	Hypothetical protein	none	none	none	none
35	c(32990)	477	rimL	Acetyltransferases, including N-acetylases of ribosomal proteins	EAA66964	35	49	<i>Aspergillus nidulans</i> _c
36	c(34487)	774	trmD	tRNA-(guanine-N1)-methyltransferase	ZP_00006764	68	83	<i>Rba. sphaeroides</i>

(a) The 16S rRNA phylogenetic identities of all organisms listed are α -3 proteobacteria (based on 16S rRNA gene), except as noted (b) Uncultured Monterey Bay proteobacterium MB BAC60D04 (α -3-proteobacterium, based on *pufM* gene) (c) *Ascomycota*, Fungi. Gene abbreviations: See Table S1. Genus abbreviation: *Rhodobacter* (*Rba.*)

Table S3. Accession numbers of sequences used in analyses. Protein or nucleotide accession numbers are listed, depending on which sequences were used for the analysis. Locus tag refers to locus name, ORF identification number, or base positions in genomic DNA encoding gene of interest.

<u>Organism</u>	<u>Figure</u>	<u>Gene analyzed</u>	<u>Accession#</u>	<u>Locus tag</u>
Acidiphilium acidophilum	Figures 5-6	pufM nucleotide	AB013379	/coded by 1308..2076
Acidiphilium angustum	Figure S1E	pufB protein	BAA25556.1	n/a
Acidiphilium angustum	Figure S2B	pufC protein	BAA25560.1	n/a
Acidiphilium angustum	Figures 5-6	pufM nucleotide	AB005219	/coded by 1835..2815
Acidiphilium multivorum	Figures 5-6	pufM nucleotide	AB005221	/coded by 824..1584
Acidiphilium organovorum	Figures 5-6	pufM nucleotide	AB005222	/coded by 823..>1578
Acidiphilium rubrum	Figure S2B	pufC protein	BAA25553	n/a
Actinobacillus pleuropneumoniae serovar	Figure 4	rpmE nucleotide	NZ_AAACK01000032	Aple02001617
Agrobacterium sanguineum	Figures 5-6	pufM nucleotide	AB011074	/coded by 753..1505
Agrobacterium tumefaciens str. C58	Figure 4	rpmE nucleotide	NC_003063	AGR_L_2181
Allochroamatium vinosum	Figure S1E	pufB protein	BAA32738.1	n/a
Allochroamatium vinosum	Figure S2B	pufC protein	O82947	n/a
Alpha proteobacterium MBIC3951	Figures 5-6	pufM nucleotide	AB018690	/coded by 777..>1538
Alpha proteobacterium R2A163	Figures 5-6	pufM nucleotide	AF393993	/coded by <1..>609
Alpha proteobacterium R2A84	Figures 5-6	pufM nucleotide	AF393991	/coded by <1..>609
Azoarcus sp. EbN1	Fig 2	hemF nucleotide	NC_006513.1	ebA1156
Azotobacter vinelandii	Fig 2	hemF nucleotide	NZ_AAAU02000008.1	Avin02002391
Bartonella henselae	Figure 4	rpmE nucleotide	NC_005956	BH15050
Blastochloris sulfoviridis	Figures 5-6	pufM nucleotide	AB017041	/coded by 751..>1500
Blastochloris viridis	Figure S1E	pufB protein	P04124	n/a
Blastochloris viridis	Figure S2B	pufC protein	CAA29223	n/a
Bordetella bronchiseptica	Fig 2	hemF nucleotide	NC_002927	BB1869
Bradyrhizobium japonicum	Fig 2	hemF nucleotide	NC_004463	blI2481
Bradyrhizobium japonicum	Figure 4	rpmE nucleotide	NC_004463	bsI1507
Bradyrhizobium sp. ORS278	Figures 5-6	pufM nucleotide	AF182374.6	/coded_by="AF182374.6:15391..16317"
Bradyrhizobium sp. ORS278	Figure S1A	crtE protein	AAR98495.1	n/a
Bradyrhizobium sp. ORS278	Figure S1B	crtF nucleotide	AF182374.6	/coded_by="AF182374.6:7262..8407"
Bradyrhizobium sp. ORS278	Figure S1C	bchC nucleotide	AF182374.6	/coded_by="AF182374.6:8498..9439"
Bradyrhizobium sp. ORS278	Fig S1D	pufL nucleotide	AF182374.6	/coded_by="AF182374.6:14530..15372"
Bradyrhizobium sp. ORS278	Figure S1E	pufB protein	AAG42642.1	n/a
Bradyrhizobium sp. ORS278	Figure S2A	bchX protein	AAL68696.1	n/a
Bradyrhizobium sp. ORS278	Figure S2C	crtD protein	AAR98494.1	n/a
Brucella suis 1330	Figure 4	rpmE nucleotide	AE014291	BR1716
Burkholderia cepacia	Fig 2	hemF nucleotide	NZ_AAEI01000031	Bcepa03006760
Burkholderia pseudomallei	Fig 2	hemF nucleotide	NC_006350	BPSL1163
Chloroflexus aurantiacus	Fig 2	hemF nucleotide	NZ_AAAH01000961.2	Chlo02002478
Chloroflexus aurantiacus	Figures 5-6	pufM nucleotide	NZ_AAAH01001023.2	/coded_by="9348..10271"
Chloroflexus aurantiacus	Figure S1B	crtF nucleotide	AF288602.1	/coded_by="AF288602.1:1912..2976"
Chloroflexus aurantiacus	Fig S1D	pufL nucleotide	X14979.1	/coded by 11..946
Chloroflexus aurantiacus	Figure S2B	pufC protein	P33325	n/a
Chromatium purpuratum	Figure S1E	pufB protein	AAB30675.1	n/a
Chromatium vinosum	Figures 5-6	pufM nucleotide	D50647	/coded by 791..>1560
Chromobacterium violaceum	Fig 2	hemF nucleotide	NC_005085	CV0757
Chromobacterium violaceum	Figure 4	rpmE nucleotide	AE016916.1	CV1747
Crater Lake beta-proteo. HTCC528	Figures 5-6	pufM nucleotide	AY584589	/coded by complement(<1..>540)
Dechloromonas aromatica	Fig 2	hemF nucleotide	NZ_AADF01000002.1	Daro03001415

<u>Organism</u>	<u>Figure</u>	<u>Gene analyzed</u>	<u>Accession#</u>	<u>Locus tag</u>
Dechloromonas aromatica RCB	Figure 4	rpmE nucleotide	NZ_AADF01000005.1	Daro03002480
Desulfotalea psychrophila LSv54	Figure 4	rpmE nucleotide	CR522870.1	DP2727
eBACred25D05	Figures 5-6	pufM nucleotide	AY671989	/coded by complement(16879..17811)
eBACred25D05	Figure S1A	crtE protein	AAT90315.1	n/a
eBACred25D05	Figure S1B	crtF nucleotide	AY671989	/coded by complement(17868..18707)
eBACred25D05	Figure S1C	bchC nucleotide	AY671989	/coded by complement(23597..24628)
eBacRed25D05	Fig S1D	pufL nucleotide	AY671989	/coded_by="complement(17868..18707)"
eBACred25D05	Figure S1E	pufB protein	AY671989	n/a
eBACred25D05	Figure S2A	bchX protein	AAT90312.1	n/a
eBACred25D05	Figure S2B	pufC protein	AAT90304	n/a
eBACred25D05	Figure S2C	crtD protein	AAT90316.1	n/a
Ectothiorhodospira halophila	Figure S1E	pufB protein	P80104	n/a
Ectothiorhodospira shaposhnikovii	Fig S1D	pufL nucleotide	AF018955.1	/coded_by="AF018955.1:1309..2133
Ectothiorhodospira shaposhnikovii	Figure S1E	pufB protein	AAB67244.1	n/a
Ectothiorhodospira shaposhnikovii	Figure S2B	pufC protein	AAB67248.1	n/a
Ectothiorhodospira shaposhnikovii	Figures 5-6	pufM nucleotide	AF018955	/coded by 2150..3124
Erwinia carotovora	Fig 2	hemF nucleotide	NC_004547.2	ECA0874
Erythrobacter (Roseobacter) OCH114	Figures 5-6	pufM nucleotide	X57597	/coded by 1942..2937
Erythrobacter litoralis	Figures 5-6	pufM nucleotide	AB010981	/coded by 753..>1505
Erythrobacter litoralis HTCC2594	Figure 4	rpmE nucleotide	NZ_AAGG01000008.1	ELI2259
Erythrobacter longus	Figure S1E	pufB protein	BAA96514.1	n/a
Erythrobacter longus	Figures 5-6	pufM nucleotide	D50648	/coded by 774..>1546
Erythrobacter sp. MBIC3960	Figures 5-6	pufM nucleotide	AB027515	/coded by 2127..>2879
Erythrobacter sp. MBIC3960	Figure S1E	pufB protein	BAA78670.1	n/a
Erythrobacter sp. NAP1	Figures 5-6	pufM nucleotide	AY326260	/coded by 1..>753
Erythrobacter sp. OCH114	Figure S1E	pufB protein	CAA40815.1	n/a
Erythrobacter sp. OCH114	Figure S2B	pufC protein	CAA40820.1	n/a
Escherichia coli O157	Fig 2	hemF nucleotide	NC_002695.1	ECs3307
Gluconobacter oxydans	Fig 2	hemF nucleotide	NC_006677	GOX1896
Gluconobacter_oxydans	Figure 4	rpmE nucleotide	NC_006677.1	GOX2398
Haemophilus influenzae R2866	Figure 4	rpmE nucleotide	NZ_AADP01000001	Hflu203000223
Haemophilus somnus 2336	Figure 4	rpmE nucleotide	NZ_AACJ01000002	Haso02000167
Idiomarina loihiensis	Fig 2	hemF nucleotide	NC_006512	IL0023
Idiomarina loihiensis L2TR	Figure 4	rpmE nucleotide	NC_006512.1	IL2462
Lake Fryxell Morphotype A	Figures 5-6	pufM nucleotide	AY178005.1	/coded by <1..>229
Lake Fryxell Morphotype B	Figures 5-6	pufM nucleotide	AY178006	/coded by <1..>229
Lamprocystis purpurea	Figures 5-6	pufM nucleotide	AY177752.1	/coded_by="AY177752.1:2424..3398"
Lamprocystis purpurea	Figure S1E	pufB protein	AAO47393.1	n/a
Lamprocystis purpurea	Figure S2B	pufC protein	AAO47392.1	n/a
Magnetospirillum magnetotacticum	Figure 4	rpmE nucleotide	NZ_AAAP01003684	Magn03007540
Magnetospirillum magnetotacticum	Figure S2C	crtD protein	ZP_00050928.1	n/a
Magnetospirillum magnetotacticum	Figure S1C	bchC nucleotide	NZ_AAAP01001457	Magn03002192
marine Roseobacter BS110	Figures 5-6	pufM nucleotide	AY675568	/coded by <1..>575
Mesorhizobium loti MAFF303099	Figure 4	rpmE nucleotide	NC_002678	msr3935
Methylobacillus flagellatus	Fig 2	hemF nucleotide	NZ_AADX01000006	Mflag03001216
Methylococcus capsulatus	Fig 2	hemF nucleotide	AE017282.2	MCA2791
Microbulbifer degradans	Fig 2	hemF nucleotide	NZ_AABI03000001	Mdeg02000931
Microbulbifer degradans	Figure 4	rpmE nucleotide	NZ_AABI03000007	Mdeg02002840
Neisseria gonorrhoeae	Figure 4	rpmE nucleotide	AE004969.1	NGO2126
Nitrosomonas europaea	Fig 2	hemF nucleotide	NC_004757.1	NE1876
Nitrosomonas europaea	Figure 4	rpmE nucleotide	NC_004757.1	NE1036
Phaeospirillum molischianum	Figure S1E	pufB protein	BAA22795.1	n/a

<u>Organism</u>	<u>Figure</u>	<u>Gene analyzed</u>	<u>Accession#</u>	<u>Locus tag</u>
Phaeospirillum molischianum	Figure S2B	pufC protein	BAA22799.1	n/a
Photobacterium profundum	Fig 2	hemF nucleotide	CR378674	PBPRA3572
Photobacterium profundum	Figure 4	rpmE nucleotide	NC_006370.1	PBPRA0258
Polaromonas sp. JS666	Fig 2	hemF nucleotide	NZ_AAFO01000006	PJS6w01002386
Porphyrobacter neustonensis	Figures 5-6	pufM nucleotide	AB011073	/coded by 756..>1508
Porphyrobacter tepidarius MBIC3363	Figures 5-6	pufM nucleotide	AB020599	/coded by 753..>1505
Pseudomonas aeruginosa	Fig 2	hemF nucleotide	NZ_AABO07000003.1	Paer03004425
Pseudomonas aeruginosa	Figure 4	rpmE nucleotide	NC_002516.1	PA5049
Pseudomonas fluorescens	Fig 2	hemF nucleotide	NZ_AAAT03000003.1	Pflu02002414
Pseudomonas putida	Fig 2	hemF nucleotide	NC_002947.3	PP0073
Ralstonia eutropha	Fig 2	hemF nucleotide	NZ_AADY01000002	Raeut03003388
Ralstonia metallidurans	Fig 2	hemF nucleotide	NZ_AAAI02000002	Reut02002293
Ralstonia solanacearum	Fig 2	hemF nucleotide	AL646068	RS2192
Rhodobacter azotoformans	Figures 5-6	pufM nucleotide	AB062783	/coded by 761..>1636
Rhodobacter blasticus	Figures 5-6	pufM nucleotide	D50649	/coded by 779..>1551
Rhodobacter capsulatus	Figures 5-6	pufM nucleotide	Z11165.1	/coded by 42576..43499
Rhodobacter capsulatus	Figure S1A	crtE protein	P17060	n/a
Rhodobacter capsulatus	Figure S1B	crtF nucleotide	Z11165.1	/coded_by="Z11165.1:34580..35761"
Rhodobacter capsulatus	Figure S1C	bchC nucleotide	Z11165.1	/coded_by="Z11165.1:35924..36868"
Rhodobacter capsulatus	Fig S1D	pufL nucleotide	Z11165.1	/coded by 41735..42583
Rhodobacter capsulatus	Figure S1E	pufB protein	1002280A	n/a
Rhodobacter capsulatus	Figure S2A	bchX protein	CAA77548.1	n/a
Rhodobacter capsulatus	Figure S2C	crtD protein	1613414F	n/a
Rhodobacter sphaeroides	Figure S1A	crtE protein	AAF24294.1	n/a
Rhodobacter sphaeroides	Figure S1B	crtF nucleotide	AF195122.1	/coded_by="AF195122.1:56639..57778"
Rhodobacter sphaeroides	Figure S1E	pufB protein	CAB38752.1	n/a
Rhodobacter sphaeroides	Figure S2A	bchX protein	AAF24297.1	n/a
Rhodobacter sphaeroides	Figure S2D	acsF nucleotide	NZ_AA AE01000100.1	Rsph03001155
Rhodobacter sphaeroides	Figures 5-6	pufM nucleotide	AF195122	/coded by 64605..65531
Rhodobacter sphaeroides	Figure S1C	bchC nucleotide	NZ_AA AE01000130	Rsph03002233
Rhodobacter sphaeroides 2.4.1	Fig S1D	pufL nucleotide	NZ_AA AE01000130.1	Rsph03002227
Rhodobacter sphaeroides 2.4.1	Figure S2C	crtD protein	ZP_00207382.1	n/a
Rhodobacter sphaeroides 2.4.1	Figure 4	rpmE nucleotide	NZ_AA AE01000140	Rsph03002732
Rhodobacter veldkampii	Figures 5-6	pufM nucleotide	AB062784	/coded by 760..>1606
Rhodobium marinum	Figure S1E	pufB protein	P80260	n/a
Rhodocyclus tenuis	Figures 5-6	pufM nucleotide	D50651	/coded by 778..>1547
Rhodoferax antarcticus	Figures 5-6	pufM nucleotide	AF333236	/coded by <1..>229
Rhodoferax fermentans	Fig S1D	pufL nucleotide	D50650.1	/coded_by="D50650.1:<1..762"
Rhodoferax fermentans	Figures 5-6	pufM nucleotide	D50650	/coded by 780..>1549
Rhodomicrobium vannielii	Figures 5-6	pufM nucleotide	D50652	/coded by 777..1546
Rhodopseudomonas palustris	Fig 2	hemF nucleotide	BX572597	RPA1514
Rhodopseudomonas palustris	Figure 4	rpmE nucleotide	BX572595	RPA0918
Rhodopseudomonas palustris	Figure S1A	crtE protein	CAE26961.1	n/a
Rhodopseudomonas palustris	Figure S1B	crtF nucleotide	NC_005296.1	RPA1520
Rhodopseudomonas palustris	Fig S1D	pufL nucleotide	AB015977.1	/coded_by="AB015977.1:1..834"
Rhodopseudomonas palustris	Figure S2D	acsF nucleotide	NC_005296	RPA1552
Rhodopseudomonas palustris	Figure S1C	bchC nucleotide	NC_005296	RPA1521
Rhodopseudomonas palustris	Figure S1E	pufB protein	CAE26967.1	n/a
Rhodopseudomonas palustris	Figure S2A	bchX protein	CAE26964.1	n/a
Rhodopseudomonas palustris	Figure S2C	crtD protein	CAE26960.1	n/a
Rhodopseudomonas palustris	Figures 5-6	pufM nucleotide	BX572597	RPA1528
Rhodospirillum molischianum	Figures 5-6	pufM nucleotide	D50654	/coded by 2307..3281
Rhodospirillum photometricum	Figures 5-6	pufM nucleotide	D50681	/coded by 788..>1557

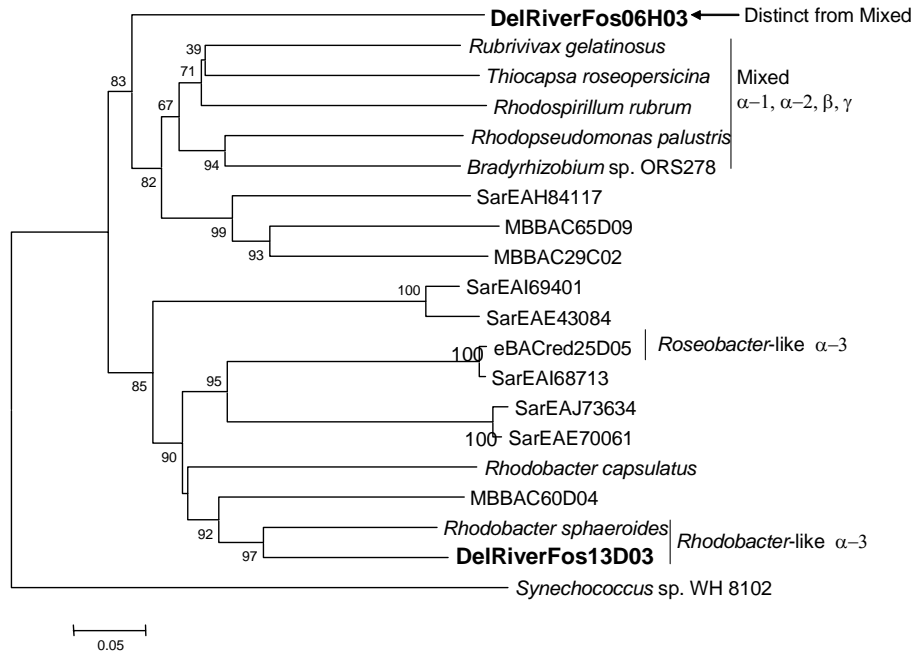
<u>Organism</u>	<u>Figure</u>	<u>Gene analyzed</u>	<u>Accession#</u>	<u>Locus tag</u>
Rhodospirillum rubrum	Figure S1A	crtE protein	ZP_00270578.1	n/a
Rhodospirillum rubrum	Figure S1B	crtF nucleotide	NZ_AAAG02000001.1	Rrub02000620
Rhodospirillum rubrum	Figure S1C	bchC nucleotide	NZ_AAAG02000001.1	/codedby"complement(695532..696485)
Rhodospirillum rubrum	Figure S2A	bchX protein	ZP_00270575.1	n/a
Rhodospirillum rubrum	Figure S2C	crtD protein	AAN75037.1	n/a
Rhodospirillum rubrum	Figure 4	rpmE nucleotide	NZ_AAAG02000002.1	Rrub02001869
Rhodovulum sulfidophilum	Fig S1D	pufL nucleotide	AB020784.1	/coded_by="AB020784.1:1964..2818
Rhodovulum sulfidophilum	Figure S1E	pufB protein	BAA76945.1	n/a
Rhodovulum sulfidophilum	Figure S2B	pufC protein	BAA76949	n/a
Rhodovulum sulfidophilum	Figures 5-6	pufM nucleotide	AB020784	/coded by 2821..3786
Roseateles depolymerans	Figures 5-6	pufM nucleotide	AB028938	/coded by 2457..3437
Roseateles depolymerans	Figure S1E	pufB protein	BAB19665.1	n/a
Roseateles depolymerans	Figure S2B	pufC protein	BAB19670.1	n/a
Roseobacter denitrificans	Figure S1E	pufB protein	P26274	n/a
Roseobacter denitrificans	Figure S2B	pufC protein	CAA58310.1	n/a
Roseobacter denitrificans MBIC2684	Figures 5-6	pufM nucleotide	AB016989	/coded by 770..>1531
Roseobacter litoralis MBIC2685	Figures 5-6	pufM nucleotide	AB016990	/coded by 770..>1531
Roseobacter sp. BS90	Figures 5-6	pufM nucleotide	AY675565	/coded by <1..>778
Roseobacter sp. S03	Figures 5-6	pufM nucleotide	AY675566	/coded by 1..>776
Roseobacter sp. SYOP2	Figures 5-6	pufM nucleotide	AY675567	/coded by <1..>632
Roseococcus thiosulfatophilus	Figures 5-6	pufM nucleotide	AY064410	/coded by 775..>1541
Roseospirillum parvum	Figures 5-6	pufM nucleotide	AY242845.1	/coded_by="AY242845.1:2440..3411"
Roseospirillum parvum	Figure S1E	pufB protein	AAP70699.1	n/a
Roseospirillum parvum	Figure S2B	pufC protein	AAP70703.1	n/a
Rubritepida flocculans	Figures 5-6	pufM nucleotide	AY064409	/coded by 777..>1543
Rubrivivax gelatinosus	Fig 2	hemF nucleotide	NZ_AAEM01000001	Rgel02000224
Rubrivivax gelatinosus	Figure S1A	crtE protein	AAO93113.1	n/a
Rubrivivax gelatinosus	Figure S1B	crtF nucleotide	AB034704.1	/coded_by="AB034704.1:10837..12009"
Rubrivivax gelatinosus	Figure S1C	bchC nucleotide	AB034704.1	/coded_by="AB034704.1:12161..13114"
Rubrivivax gelatinosus	Fig S1D	pufL nucleotide	AB034704.1	/coded_by="AB034704.1:17990..18829
Rubrivivax gelatinosus	Figure S1E	pufB protein	AAO93120.1	n/a
Rubrivivax gelatinosus	Figure S2A	bchX protein	AAO93116.1	n/a
Rubrivivax gelatinosus	Figure S2B	pufC protein	T50891	n/a
Rubrivivax gelatinosus	Figure S2C	crtD protein	AAC44798.2	n/a
Rubrivivax gelatinosus	Figure S2D	acsF nucleotide	AY234384.1	/coded_by="complement(17078..18154)"
Rubrivivax gelatinosus	Figures 5-6	pufM nucleotide	D16822	/coded by 2069..3046
Salmonella enterica	Fig 2	hemF nucleotide	NC_006511.1	SPA0415
Sargasso uncultured	Figure 4	rpmE nucleotide	AACY01059281	
Sargasso uncultured	Figure 4	rpmE nucleotide	AACY01042816	
Shewanella oneidensis	Fig 2	hemF nucleotide	NC_004347	SO0038
Shewanella oneidensis MR-1	Figure 4	rpmE nucleotide	NC_004347.1	SO4120
Shigella flexneri	Figure 4	rpmE nucleotide	NC_004337	SF4014
Silicibacter pomeroyi	Fig 2	hemF nucleotide	NC_003911	SPO3653
Silicibacter pomeroyi DSS-3	Figure 4	rpmE nucleotide	CP000031.1	SPO3256
Silicibacter sp. TM1040	Figure 4	rpmE nucleotide	NZ_AAFG01000015.1	STM1w01003714
Sinorhizobium melliloti 1021	Figure 4	rpmE nucleotide	NC_003047	SMc03990
Sphingomonas natatoria	Figures 5-6	pufM nucleotide	AB012060	/coded by 754..>1506
Sphingomonas sp. NT12	Figures 5-6	pufM nucleotide	AB026182	/coded by 753..>1505
Sphingomonas ursincola	Figures 5-6	pufM nucleotide	AB031016	/coded by 1304..>2075
Synechococcus sp. WH 8102	Figure S1A	crtE protein	NP_896835.1	n/a
Thermochromatium Tepidum	Figure S2B	pufC protein	1EYSC	n/a

<u>Organism</u>	<u>Figure</u>	<u>Gene analyzed</u>	<u>Accession#</u>	<u>Locus tag</u>
Thiobacillus denitrificans	Fig 2	hemF nucleotide	NZ_AAFH01000005.1	TdenA01002508
Thiobacillus denitrificans	Figure 4	rpmE nucleotide	NZ_AAFH01000002.1	TdenA01001259
Thiocapsa roseopersicina	Figure S1A	crtE protein	AAP59037.1	n/a
Thiocapsa roseopersicina	Figure S1B	crtF nucleotide	AF528191.1	/coded_by="AF528191.1:18865..19980"
Thiocapsa roseopersicina	Figure S1C	bchC nucleotide	AF528191.1	/coded_by="AF528191.1:20089..21042"
Thiocapsa roseopersicina	Figure S2C	crtD protein	AAP59036.1	n/a
Thiocapsa roseopersicina	Figures 5-6	pufM nucleotide	AJ544223	/coded by 792..>1561
Thiocystis gelatinosa	Figures 5-6	pufM nucleotide	D50653	/coded by 791..>1560
Uncultured Lake Fryxell clone LFc1	Figures 5-6	pufM nucleotide	AY177971.1	/coded by <1..>229
Uncultured Lake Fryxell clone LFc12	Figures 5-6	pufM nucleotide	AY177981	/coded by <1..>229
Uncultured Lake Fryxell clone LFc19	Figures 5-6	pufM nucleotide	AY177988.1	/coded by <1..>229
Uncultured Lake Fryxell clone LFc20	Figures 5-6	pufM nucleotide	AY177989.1	/coded by <1..>229
Uncultured Lake Fryxell clone LFc33	Figures 5-6	pufM nucleotide	AY178002	/coded by <1..>229
Uncultured Lake Fryxell clone LFc35	Figures 5-6	pufM nucleotide	AY178004	/coded by <1..>229
Uncultured Lake Fryxell clone LFc6	Figures 5-6	pufM nucleotide	AY177976.1	/coded by <1..>229
Uncultured Lake Fryxell clone LFc25	Figures 5-6	pufM nucleotide	AY177994	/coded by <1..>229
Uncultured Lake Fryxell clone LFc3	Figures 5-6	pufM nucleotide	AY177973	/coded by <1..>229
Uncultured Lake Fryxell clone LFc7	Figures 5-6	pufM nucleotide	AY177977	/coded by <1..>229
Uncultured Med/Red clone 122-2_22-0	Figures 5-6	pufM nucleotide	AY675571	/coded by 1..>800
Uncultured Med/Red clone 122-4_27-0	Figures 5-6	pufM nucleotide	AY675572	/coded by <1..>690
Uncultured Med/Red clone E100P3	Figures 5-6	pufM nucleotide	AY675570	/coded by 1..>771
Uncultured Med/Red clone S01P12	Figures 5-6	pufM nucleotide	AY675574.1	/coded by <1..>758
Uncultured Med/Red clone S01P3	Figures 5-6	pufM nucleotide	AY675576	/coded by 1..>776
Uncultured Med/Red clone S06P2	Figures 5-6	pufM nucleotide	AY675575	/coded by 1..>776
Uncultured Med/Red clone S06P3	Figures 5-6	pufM nucleotide	AY675573	/coded by 1..>770
Uncultured Monterey Bay BAC29C02	Figures 5-6	pufM nucleotide	AE008920	/coded by 9159..10133
Uncultured Monterey Bay BAC29C02	Figure S1A	crtE protein	AAM48595.1	n/a
Uncultured Monterey Bay BAC29C02	Figure S1B	crtF nucleotide	AE008920	/coded by 8318..9142
Uncultured Monterey Bay BAC29C02	Figure S1C	bchC nucleotide	AE008920	/coded by 3112..4062
Uncultured Monterey Bay BAC29C02	Figure S2A	bchX protein	AAM48598.1	n/a
Uncultured Monterey Bay BAC29C02	Figure S2B	pufC protein	AAM48604.1	n/a
Uncultured Monterey Bay BAC29C02	Figure S2D	acsF nucleotide	AE008920	/coded by complement(32458..33495)
Uncultured Monterey Bay BAC29C02	Fig S1D	pufL nucleotide	AE008920.1	coded_by="AE008920.1:8318..9142
Uncultured Monterey Bay BAC60D04	Figures 5-6	pufM nucleotide	AE008921	/coded by 24445..25368
Uncultured Monterey Bay BAC60D04	Figure S1A	crtE protein	AAM48650.1	n/a
Uncultured Monterey Bay BAC60D04	Figure S1B	crtF nucleotide	AE008921	/coded by 23601..24431
Uncultured Monterey Bay BAC60D04	Figure S1C	bchC nucleotide	AE008921	/coded by 17814..18755
Uncultured Monterey Bay BAC60D04	Figure S1E	pufB protein	AAM48656.1	n/a
Uncultured Monterey Bay BAC60D04	Figure S2A	bchX protein	AAM48653.1	n/a
Uncultured Monterey Bay BAC60D04	Figure S2C	crtD protein	AAM48649.1	n/a
Uncultured Monterey Bay BAC60D04	Figure S2D	acsF nucleotide	AE008921	/coded by complement(37290..38417)
Uncultured Monterey Bay BAC60D04	Fig S1D	pufL nucleotide	AE008921	/coded by 23601..24431
Uncultured Monterey Bay BAC63A02	Figure S1B	crtF nucleotide	AY458639.1	/coded_by="AY458639.1:89034..90164
Uncultured Monterey Bay BAC63A02	Figure S2C	crtD protein	AAR37797.1	n/a
Uncultured Monterey Bay BAC65D02	Figure S2C	crtD protein	AAR37850.1	n/a
Uncultured Monterey Bay BAC65D09	Figures 5-6	pufM nucleotide	AE008919.1	/coded by 12081..13055
Uncultured Monterey Bay BAC65D09	Figure S1A	crtE protein	AAL76349.1	n/a
Uncultured Monterey Bay BAC65D09	Figure S1B	crtF nucleotide	AE008919	/coded by 11242..12066
Uncultured Monterey Bay BAC65D09	Figure S1C	bchC nucleotide	AE008919.1	/coded_by="AE008919.1:6037..6984"
Uncultured Monterey Bay BAC65D09	Figure S2A	bchX protein	AAL76352.1	n/a
Uncultured Monterey Bay BAC65D09	Figure S2B	pufC protein	AAL76357.1	n/a
Uncultured Monterey Bay BAC65D09	Figure S2D	acsF nucleotide	AE008919.1	/coded by complement(39307..40362)
Uncultured Monterey Bay BAC65D09	Fig S1D	pufL nucleotide	AE008919.1	/coded_by="AE008919.1:11242..12066

<u>Organism</u>	<u>Figure</u>	<u>Gene analyzed</u>	<u>Accession#</u>	<u>Locus tag</u>
Uncultured Monterey Bay BAC69B03	Figure S2D	acsF nucleotide	AY458648	/coded by complement(20320..21375)
Uncultured Monterey Bay EBAC24D02	Figures 5-6	pufM nucleotide	AF393983	/coded by <1..>609
Uncultured Monterey Bay EBAC30G07	Figures 5-6	pufM nucleotide	AF393985	/coded by <1..>609
Uncultured Monterey Bay EBAC39B11	Figures 5-6	pufM nucleotide	AF393986	/coded by <1..>609
Uncultured Monterey Bay EBAC52B02	Figures 5-6	pufM nucleotide	AF393987	/coded by <1..>609
Uncultured Monterey Bay EBAC56B12	Figures 5-6	pufM nucleotide	AF393988	/coded by <1..>609
uncultured Sargasso	Figures 5-6	pufM nucleotide	AACY01024194	/coded by 818..1741
uncultured Sargasso	Figures 5-6	pufM nucleotide	AACY01078100	/coded by complement(980..1927)
uncultured Sargasso	Figures 5-6	pufM nucleotide	AACY01093008	/coded by complement(1374..>2149)
uncultured Sargasso	Figures 5-6	pufM nucleotide	AACY01194873	/coded by 75..974
uncultured Sargasso	Figures 5-6	pufM nucleotide	AACY01287106	/coded by complement(<1..>644)
uncultured Sargasso	Figures 5-6	pufM nucleotide	AACY01319421	/coded by <1..>807
uncultured Sargasso	Figures 5-6	pufM nucleotide	AACY01400725	/coded by <1..>747
uncultured Sargasso	Figure S1A	crtE protein	EAH84117	n/a
uncultured Sargasso	Figure S1A	crtE protein	EAI69401	n/a
uncultured Sargasso	Figure S1A	crtE protein	EAE43084	n/a
uncultured Sargasso	Figure S1A	crtE protein	EAI68713	n/a
uncultured Sargasso	Figure S1A	crtE protein	EAJ73634	n/a
uncultured Sargasso	Figure S1A	crtE protein	EAE70061	n/a
uncultured Sargasso	Figure S1B	crtF nucleotide	AACY01744385	/coded by <1..770
uncultured Sargasso	Figure S1B	crtF nucleotide	AACY01085688	/coded by 973..2082
uncultured Sargasso	Fig S1D	pufL nucleotide	AACY01228868	/coded by 153..>963
uncultured Sargasso	Fig S1D	pufL nucleotide	AACY01080678	/coded by complement(568..>1453)
uncultured Sargasso	Fig S1D	pufL nucleotide	AACY01024194	/coded by <1..825
uncultured Sargasso	Figure S2A	bchX protein	EAB45090	n/a
uncultured Sargasso	Figure S2A	bchX protein	EAB18610	n/a
uncultured Sargasso	Figure S2D	acsF nucleotide	AACY01294631	/coded by <1..>765
Vibrio cholerae	Fig 2	hemF nucleotide	AE004096	VC0055
Vibrio vulnificus	Figure 4	rpmE nucleotide	AE016801.1	VV11360
Zymomonas mobilis	Fig 2	hemF nucleotide	AE008692.1	ZMO0951
Zymomonas mobilis	Figure 4	rpmE nucleotide	AE008692.1	ZMO1145

Figure S1. Examples of freshwater cluster genes. (A) *crfE* protein analysis, based on 346 amino acids. (B) *crfF* nucleotide analysis, 880 base pairs. (C) *bchC* nucleotide, 1278 base pairs. (D) *pufL* gene, based on 606 nucleotide positions (E) *pufB* protein, 110 amino acid positions.

S1A



S1B

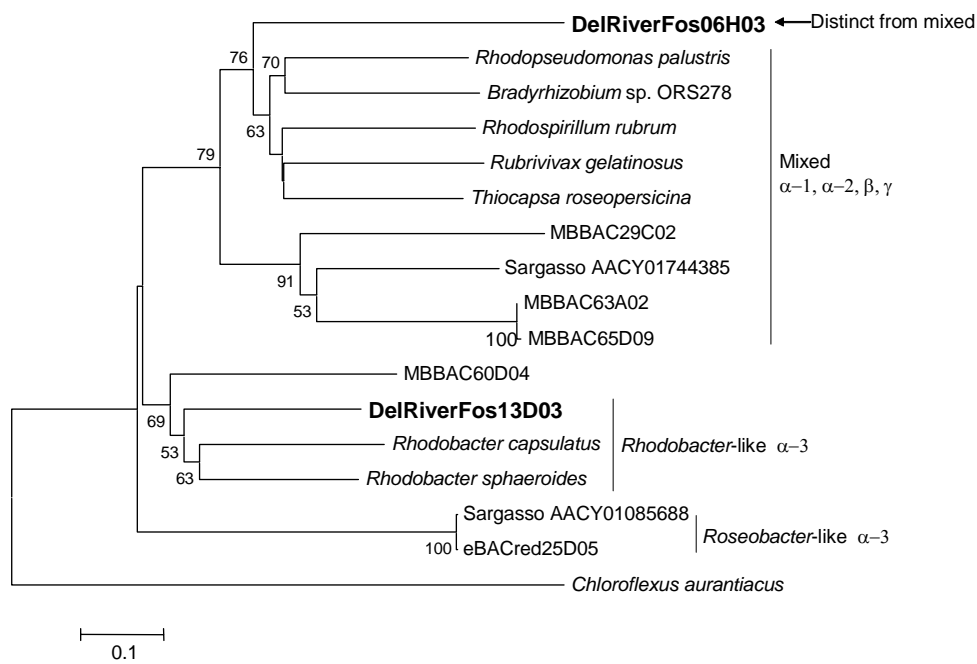


Figure S1 continued.

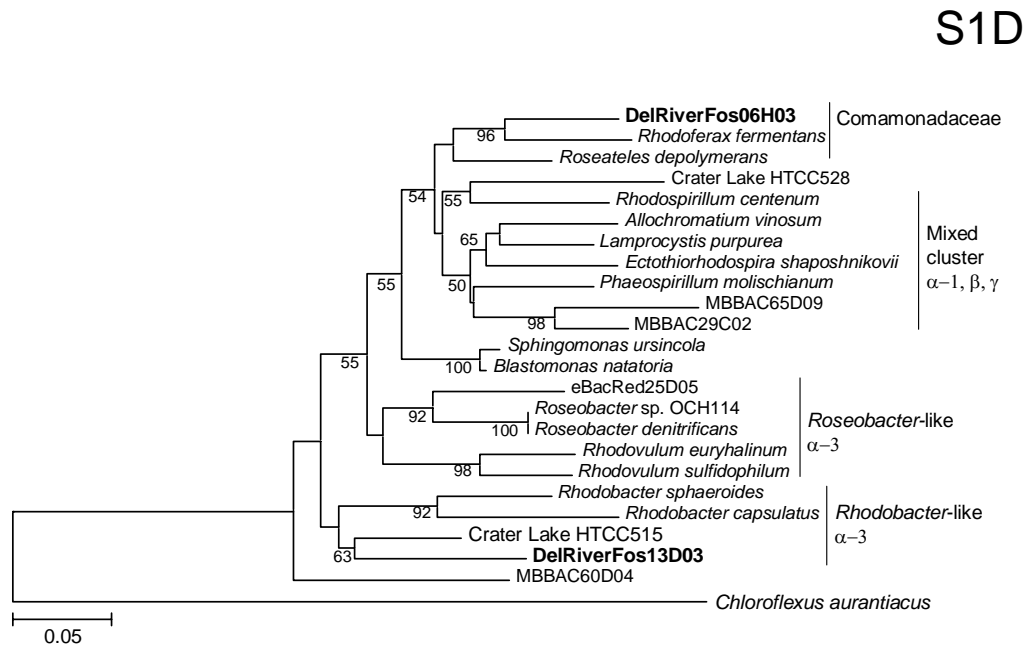
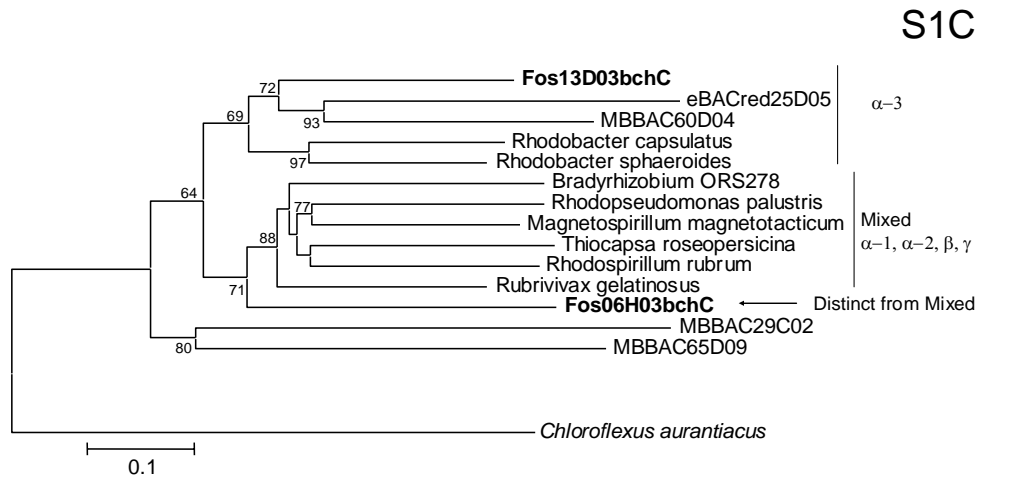


Figure S1 continued

S1E

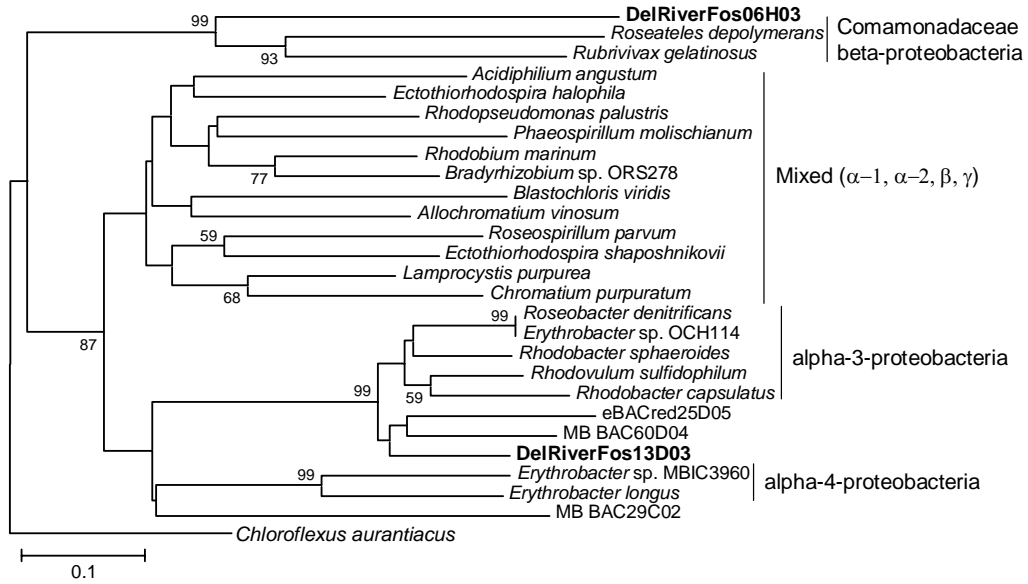


Figure S2. Examples of genes placing the beta-proteobacterial fosmid clone in the mixed cluster of proteobacteria. (A) *bchX* protein analysis, based on 410 amino acids. (B) *pufC* protein, based on 503 amino acids. (C) *crtD* protein, based on 539 amino acids. (D) *acsF* nucleotide analysis, 798 nucleotide positions.

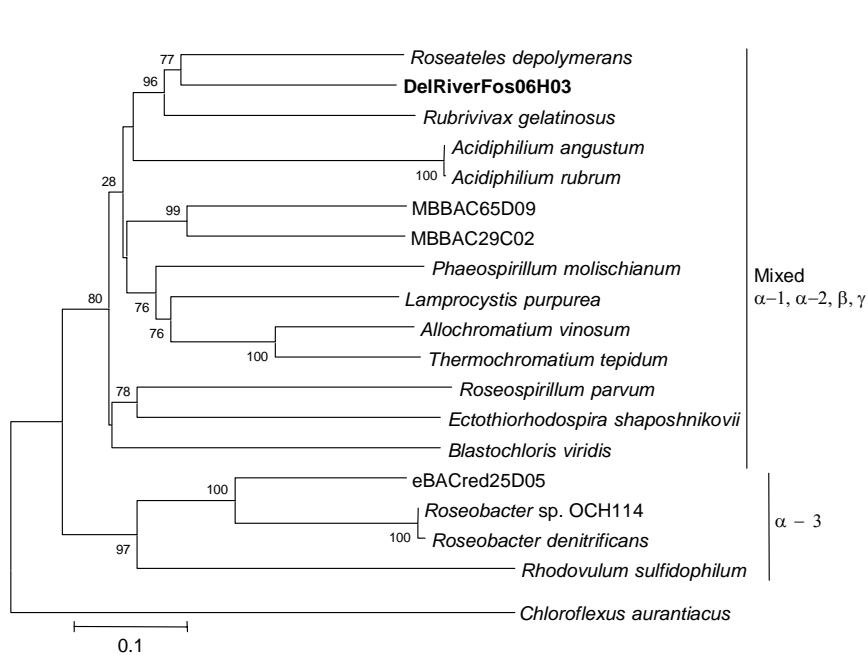
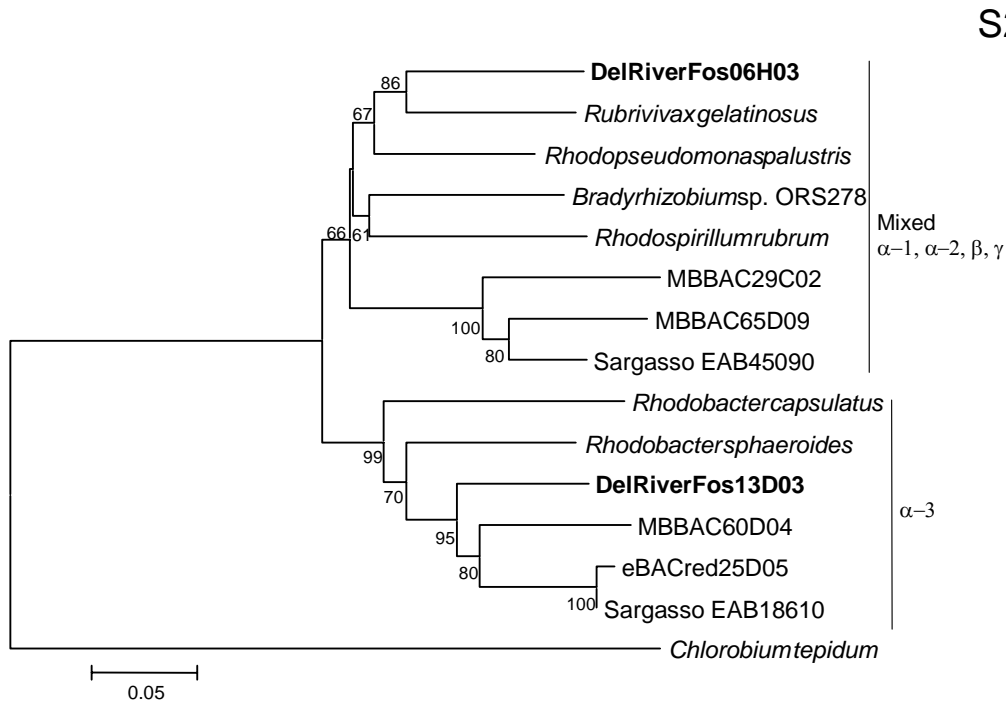
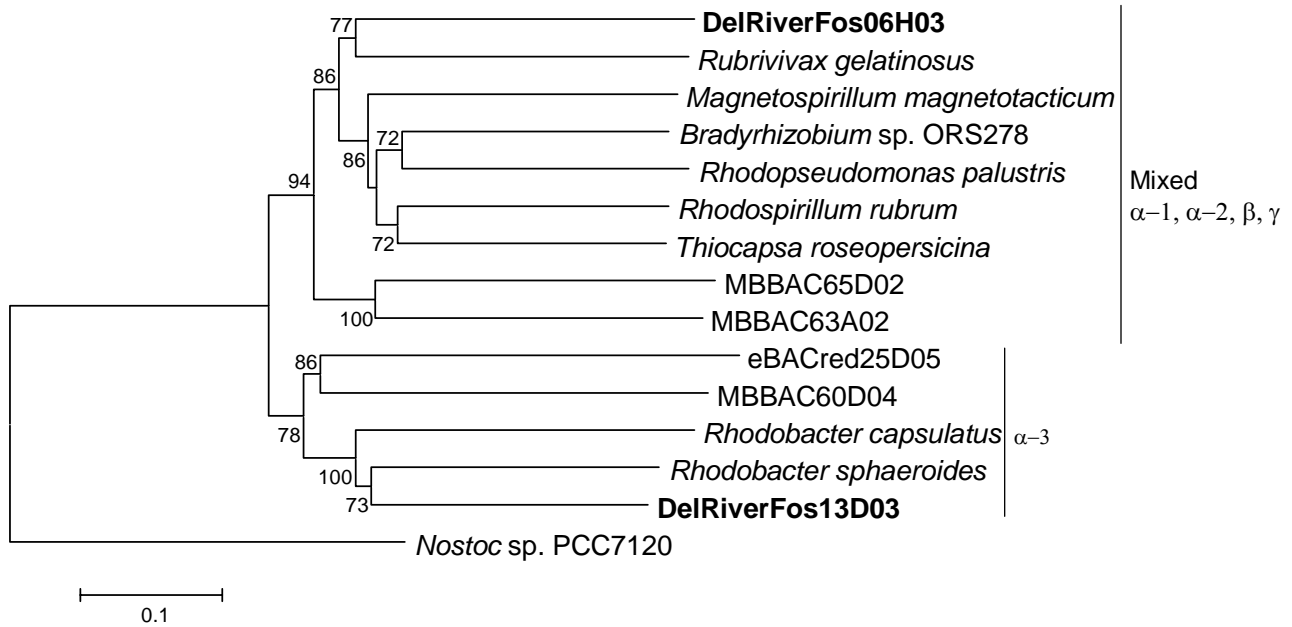


Figure S2 continued.

S2C



S2D

