Supporting Online Material for

Candidatus Chloracidobacterium thermophilum: An Aerobic Phototrophic Acidobacterium

Donald A. Bryant,* Amaya M. Garcia Costas, Julia A. Maresca, Aline Gomez Maqueo Chew, Christian G. Klatt, Mary M. Bateson, Luke J. Tallon, Jessica Hostetler, William C. Nelson, John F. Heidelberg, David M. Ward

*To whom correspondence should be addressed. E-mail: dab14@psu.edu

Published 27 July 2007, *Science* **317**, 523 (2007)
DOI: 10.1126/science.1143236

This PDF file includes:

Supporting Online Material for

A phototrophic acidobacterium discovered by metagenomics

D. A. Bryant*, A. M. Garcia Costas, J. A. Maresca, A. Gomez Maqueo Chew, C. Klatt, M. M. Bateson, L. J. Tallon, J. Hostetler, W. C. Nelson, J. F. Heidelberg, and D. M. Ward

*To whom correspondence should be addressed. E-mail: dab14@psu.edu

This supplement contains:

Materials and Methods

SOM text

Figs. S1 to S13

Table S1

References

Supporting Online Material

MATERIALS AND METHODS

Library construction, DNA sequencing, and sequence analysis

The sequence of a representative *Roseiflexus* sp. isolate, strain RS1 (*1-3*) was determined as part of a Joint Genome Institute-Department of Energy project to determine genome sequences of filamentous anoxygenic phototrophs (http://genome.jgi-psf.org/; GenBank accession AAQU00000000). The *Synechococcus* sp. OS-A and OS-B' genome sequences (*4*) (GenBank accessions CP000239 and CP000240) and metagenomic clone libraries and databases (*4-6*) used in this study were developed for an NSF Frontiers in Integrative Biology Research project (*7*). Amplicon Express (Pullman, WA) constructed the bacterial artificial chromosome (BAC) clone library. Sequence data derived from BAC clone M60-018 J19 (see Fig. 6S) have been deposited in GenBank under accession number EF531339.

Sequencing and assembly of the metagenome were performed as described for genomes sequenced by TIGR and the J. Craig Venter Institute (*8*). BAC closure was done by mechanical shearing of the entire BAC followed by random shotgun sequencing on small insert (2-3 kbp) clones. The plasmid sequences were jointly assembled using the TIGR Assembler (*9*). The coverage criteria were that every position required at least double-clone coverage (or sequence from a PCR product amplified from genomic DNA), and either sequence from both strands or with two different sequencing chemistries. The sequence was edited manually, and additional PCR (*10*) and sequencing reactions were performed to close gaps, improve coverage, and resolve sequence ambiguities.

An initial set of open reading frames likely to encode proteins (coding sequence, CDS) were predicted for the BAC insert by Glimmer (*11*). All predicted proteins larger than 30 amino acids were searched against a non-redundant protein database as previously described (*12-14*). TopPred was used to predict the membrane-spanning domains of proteins (*15*). The 5' regions of each CDS were inspected to define initiation codons using homologies, and position of ribosomal binding sites and transcriptional terminators. Two sets of hidden Markov models were used to determine CDS membership in families and superfamilies: pfam v11.0 (*16*) and TIGRFAMs 3.0 (*17*). Pfam v11.0 hidden Markov models were also used with a constraint of a minimum of two hits to find and mask repeated domains within proteins.

Database searches were performed using the various blast algorithms to search the GenBank non-redundant databases (*18, 19*). For construction of phylogenetic trees, protein sequences were aligned using the ClustalW tool in MacVector versions 7.2.3 or 9.0.2 (Accelrys, San Diego, CA). Neighbor-joining trees for type-1 reaction centers and BchU were generated from the alignments using PAUP 4.0 (Sinauer Associates, Inc., Sunderland, MA). The neighbor-joining tree for RecA was generated using MacVector version 9.0.2 (Accelrys, San Diego, CA).

Enrichment culture methods and isolation of Anoxybacillus sp.

Enrichment cultures containing *Synechococcus* sp. strain JA-2-3B'a (2-13), Chloracidobacterium (Cab.) thermophilum, and *Anoxybacillus* sp. (*20*) were grown in B-HEPES medium at pH 8.0 (*21*) at 45-53°C. In cultures to maintain and enrich for Cab. thermophilum, the B-HEPES medium was supplemented with 1 mM ammonium chloride and a mixture of carbon sources: lactate, succinate, glycolate, butyrate, propionate, and acetate (1 mM each, added in the form of sodium salts). Nitrate was omitted from the growth medium of cultures that did not contain cyanobacteria. The medium was also supplemented with filter-sterilized vitamins at the

following final concentrations: (10 μg L^{-1}) riboflavin, vitamin B_{12} (100 μg L^{-1}), thiamine-HCl (100 μg L^{-1}), L-ascorbic acid (100 μg L^{-1}), D-Ca-panthothenate (100 μg L^{-1}), folic acid (100 μg L^{-1}), nicotinic acid (100 μg L^{-1}), 4-aminobenzoic acid (100 μg L^{-1}), pyridoxine-HCl (100 μg L^{-1}), and lipoic acid (100 μg L^{-1}). To inhibit the growth of the cyanobacterial cells in the original enrichment, atrazine was added to a final concentration of 10 μM.

Unless otherwise specified, all cultures were incubated under oxic conditions with or without shaking in constant light provided by a 60 W tungsten and a 60 W fluorescent bulb (total visible light intensity of 67 μmol photons m^{-2} s^{-1}). *Synechococcus* sp. strain JA-2-3B'a (2-13) was eliminated from an enrichment that also contained *Anoxybacillus* sp. and Cab. thermophilum by serial subculturing three times in the light in the presence of 10 μM atrazine. *Anoxybacillus* sp. was isolated in axenic culture by plating cells from the enrichment cultures on Luria-Bertani medium solidified with 1.5% (w/v) Difco Bacto-Agar (Becton-Dickinson, Franklin Lakes, NJ). *Anoxybacillus* sp. was identified by amplifying and DNA sequencing a region of its 16S rRNA by using universal bacterial primers (Table 1S).

Light microscopy

Cab. thermophilum cells were incubated in 1-hexanol-saturated phosphate-buffered saline to disrupt bacteriochlorophyll (BChl) *c* aggregates (*22*) and visualized in an Olympus BX-60 epifluorescence upright microscope equipped with an UplandFL 100× oil objective, a mercury vapor lamp and Olympus filter cubes 41001 (Olympus, Center Valley, PA).

Absorption and fluorescence emission spectra

Absorption spectra were either recorded with a Cary 14 UV-Vis-NIR spectrophotometer modified for computerized data acquistion by On-Line Systems, Inc. (Bogart, GA) or with a GENESYS 10 spectrophotometer (Thermo Electron Corp., Rochester, NY). Data were analyzed using Igor-Pro software (Wavemetrics, Inc. Lake Oswego, OR).

Chlorosome isolation and electron microscopy

Chlorosomes were isolated from fresh cells (2 L) as described by Vassilieva *et al.* (*23*). Isolated chlorosomes were negatively stained with 1% (w/v) uranyl acetate and visualized in a JEOL 1200 EXII Transmission electron microscope (Peabody , MA).

Pigment extraction and analysis

Pigments were extracted from 1-ml aliquots of cell suspension from late-exponential phase enrichment cultures by sonication in acetone-methanol (7:2, v/v). Cell debris was removed by centrifugation. The absorption spectrum of the pigment mixture was recorded after diluting the sample at least 1:10 in methanol, and the absorption spectrum of the extracts was recorded from 350 to 900 nm. Samples for HPLC analysis were filtered and 0.1 volume of 1 M ammonium acetate was added to the filtrate immediately before injection into the HPLC column. The pigments were separated on a 25 cm by 4.6 mm Discovery 5 μm C_{18} column (Supelco, Bellefonte, PA) attached to a 1,024-element diode array detector (model G1315B, 1100 series; Agilent Technologies, Palo Alto, CA) controlled with Agilent ChemStation software for HPLC. The solvent system was previously described (*24*). BChl *c* and BChl *a* extracted from *Chl. tepidum* and from *Cfx. aurantiacus* Y-400-fl served as standards (see Fig. 4). Elution of the pigments was monitored at 667 and 770 nm, the absorbance maxima for BChl *c* and BChl *a*

respectively; the absorption spectrum of the column eluate was recorded from 200 to 900 nm at two-second intervals.

Polymerase chain reaction and agarose gel electrophoresis

In general, the polymerase chain reaction conditions consisted of 5 minutes at 95^{o}C, 35 cycles of 45 seconds at 95^{o}C, 45 seconds at 57^{o}C and 1 minute at 72^{o}C, followed by a final elongation step of 10 minutes at 72^{o}C. About 10-100 ng of DNA were included in each reaction and primers (see Table 1S) were supplied at a final concentration of 1 μM. PCR products were electrophoresed in a 0.8% (w/v) agarose gel and stained with ethidium bromide prior to visualization.

SUPPORTING ON-LINE MATERIAL

History of discovery of photosynthetic bacteria and their differentiating characteristics

Five bacterial phyla (*Cyanobacteria, Chlorobi, Proteobacteria, Chloroflexi,* and *Firmicutes*) contain members, chlorophototrophs, capable of phototrophy based upon BChl or chlorophyll (Chl) (*25*). (Throughout this article, we have used the terms phylum and phyla as employed in the Second Edition of *Bergey's Manual of Systematic Bacteriology* (*26*), although there is debate as to whether the major sublineages of Domains should be considered kingdoms, as in the case of plants, animals, and fungi.). Studies of *Cyanobacteria* and photosynthetic *Proteobacteria* (*i.e.* the purple bacteria) date to the first half of the 19th century, and green sulfur bacteria (*Chlorobi*) were first described in 1906 (*26, 27*). Cyanobacteria perform oxygenic photosynthesis by virtue of their Chl *a*-containing, type-1 (photosystem I) and type-2 (photosystem II) reaction centers, and they fix carbon by the reductive pentose-phosphate

(Calvin-Benson-Bassham) cycle. Endosymbiosis of a cyanobacterium led to the acquisition of oxygenic photosynthesis by eukaryotic algae and plants. The purple chlorophototrophic bacteria, all members of the alpha, beta, and gamma subdivisions of the *Proteobacteria*, have relatively simple, type-2 reaction centers with either BChl *a* or BChl *b*, do not evolve oxygen, and also fix carbon via the reductive pentose phosphate pathway, although most can grow mixotrophically when presented with light and a usable organic carbon source. Many purple bacteria are aerobic chemoheterotrophs under oxic conditions, but are only facultatively phototrophic under anoxic conditions. *Chlorobi* are strictly anaerobic, photoautotrophic organisms that have type-1 reaction centers with chlorosomes as antennae; any given strain can produce three types of Chl (Chl *a,* BChl *a*, and either BChl *c, d,* or *e*) and fix CO_2 by the reverse TCA cycle.

In the 100 years since the initial description of the *Chlorobi* (*27*), only two additional bacterial phyla have been shown to contain members that can produce BChls and reaction centers. Pierson and Castenholz (*28*) isolated the filamentous anoxygenic phototroph, *Chloroflexus aurantiacus*, from the same microbial mats of Octopus Spring and other thermal features around the world. This moderate thermophile synthesizes BChls *a* and *c*, produces chlorosomes as light-harvesting antennae, has type-2 reaction centers similar to those of purple bacteria, and fixes CO_2 by the 3-hydroxypropionate cycle. *Cfx. aurantiacus* is the type species of the *Chloroflexi*, an early-diverging phylum of the *Bacteria*. Strains of both major subdivisions of the *Chloroflexi* contain the genes required for carbon fixation by the 3-hydroxypropionate cycle (*2*), although it should be noted that some chlorophototrophic *Chloroflexi* possess Rubisco and fix carbon by the reductive pentose-phosphate pathway (*29*). The most recently discovered reaction center-containing bacteria are the heliobacteria, which were originally described by Gest and Favinger in 1983 (*30*). Heliobacteria are Gram-positive bacteria, i.e., *Firmicutes*, which form

heat-resistant endospores and that are closely related to clostridia (see Fig. 8S). These strict anaerobes produce type-1 reaction centers containing BChl *g*, but no known isolate can grow photoautotrophically.

Sequence analysis of a* pscAB-fmoA *operon recovered from the metagenome of the phototrophic mat community of alkaline siliceous hot springs

Inserts encoding the highly divergent *pscA* gene of an organism from the phototrophic layer of the microbial mats of Octopus and Mushroom Springs in Yellowstone NP were completely sequenced from four independent plasmids and one BAC insert (see below; GenBank accession EF531339). The overlapping data revealed a probable operon, 5'*pscAB-fmoA*, encoding the type-1 reaction center (PscAB) and the FmoA apoprotein of the BChl *a*-binding Fenna-Matthews-Olson protein (Fig. 1B, Fig. 6S). Preliminary phylogenetic analyses of the flanking genes suggested that these DNA sequences were derived from an organism related to *Acidobacterium* sp. Ellin345 or *Solibacter usitatus* Ellin6076, two soil bacteria belonging to the poorly characterized phylum *Acidobacteria* (*26*), whose genomes have recently been sequenced (*3*).

The three sequences deduced from the *pscAB-fmoA* gene cluster were divergent from known homologs in other anoxygenic chlorophototrophs. PscA was predicted to be a polypeptide of 865 amino acids (aa), which was ~135 aa larger than the PscA subunits of green sulfur bacterial type-1 reaction centers, to which it was most similar (*31*) (Fig. 1A, 1S). This reaction center apoprotein was more distantly related to PshA (~610 aa) of heliobacteria and the PsaA and PsaB subunits (730-750 aa) of cyanobacterial photosystem I (Figs. 1, 1S-3S). Like the apoproteins of all other type-1 reaction centers, the mat PscA was predicted to contain eleven,

transmembrane α-helices. An insertion of ~165 aa between predicted transmembrane α-helices VII and VIII accounted for most of the size difference (Fig. 1B, 1S). This insertion corresponded to a large extrinsic loop, which might form a docking site for a periplasmic electron donor to the oxidized special pair. The C-terminal electron transfer domain of PscA, comprising the last four transmembrane α-helices of the protein, was most similar to the same domain of green sulfur bacterial PscA sequences (Figs. 1, 3S), but this domain was nevertheless quite divergent (48% identical, 69% similar). However, the highly conserved motif, FP**C**IGPVKGGT**C**GVS, containing two cysteine residues in the cytoplasmic loop between transmembrane helices VIII and IX (Fig. 3S), implied that this reaction center contains an intersubunit [4Fe-4S] cluster similar to F_X of PS I (*31*). The N-terminal, antenna domain of this PscA-like protein is not well conserved and was difficult to align to other reaction center sequences (Fig. 2S). It exhibits only ~29% identity and ~43% similarity to the N-terminal antenna domain of PscA from green sulfur bacteria.

Similar to PscB of green sulfur bacterial Type-1 reaction centers (*31*) and the PsaC protein of photosystem I (*32*, *33*), the mat-derived PscB was predicted to be a small protein with eight conserved cysteines, which should ligate two terminal, electron-accepting, [4Fe-4S] clusters (see Fig. 4S). The region encompassing these cysteines was well conserved, but the remainder of the protein was much less similar to PscB sequences of green sulfur bacteria. The *fmoA* gene encodes the Fenna-Matthews-Olson protein, a BChl *a*-binding protein that occurs in all green sulfur bacteria (*34*). The *fmoA* gene is only known to occur in green sulfur bacteria, but in 12 sequenced green sulfur bacterial strains this gene is unlinked to *pscAB* operons. Although the FmoA proteins of green sulfur bacteria are highly similar in sequence (~70% identity, ~85% similarity), the *pscAB*-linked *fmoA* gene encoded a protein that is only ~40% identical and 58%

similar to FmoA sequences of green sulfur bacteria (Fig. 5S). Based upon these data, we hypothesized that a thermophilic acidobacterium with a novel type-1 reaction center and FmoA occurs in the mats.

Identification and sequencing of a bacterial artificial chromosome insert encoding pscAB-fmoA, recA, *and the* rDNA *operon of an acidobacterium*

To compile more information about the putative acidobacterium with type-1 reaction centers, metagenomic reads from ~60°C and ~65°C sites of the Octopus and Mushroom Spring microbial mats were assembled to produce consensus contigs. Most of the biomass in these mats belongs to organisms closely related to *Synechococcus* sp. OS-A and OS-B' and *Roseiflexus* sp. RS1 (*35*, *36*), whose genomes have been sequenced (*2-4, 37*). Sequences showing >92% identity over 80% of the read to these three organisms were first removed. Contigs with G+C contents of ~56 mol% and with multiple gene products showing best hits from blastx analyses to acidobacteria and/or δ-proteobacteria were then binned as possibly being derived from a single type of mat bacterium. This bin contained two scaffolds that encoded well-established phylogenetic markers: a complete *recA* gene and a partial 16S rRNA gene. Bacterial artificial chromosome (BAC) clones, constructed from DNA isolated from the photic-zone mat layer of the ~60°C Mushroom Spring site, were end-sequenced and mapped onto the binned scaffolds (Fig. 6S). These analyses predicted that the binned *pscAB-fmoA*, *recA*, and the rRNA genes were encoded on a single BAC insert. Sequencing of this 271,846 bp BAC insert (Fig. 6S) confirmed that the *pscAB-fmoA* and rRNA operons were separated by ~183 kbp with the *recA* gene between them. The complete 16S rRNA sequence derived from the BAC clone is identical to the partial sequence assembled from the metagenome and is nearly identical (differences at 2 of 1348

shared positions) to that from an acidobacterium, denoted GFP1, the 16S rRNA sequence of which was recovered from Green Finger Pool in the Lower Geyser Basin of Yellowstone NP (a site ~5 km from Octopus Spring) (see Fig. 1 of *38*) (Fig. 7S). These data unequivocally establish that the highly divergent *pscAB-fmoA* genes are derived from a GFP1-like organism belonging to the phylum *Acidobacteria*. Phylogenetic analyses of RecA are consistent with this conclusion (Fig. 8S, 9S).

Analyses of sequences tentatively assigned to the GFP1-like acidobacterium.

The binned sequence assemblies putatively derived from the GFP1-like acidobacterium were annotated by computer and analyzed. The acidobacterial sequence bin contained several genes for BChl and carotenoid biosynthesis as well as genes associated with chlorosomes. For example, *bchU* (Fig. 10S, 11S), whose product is a C-20 methyltransferase required for the synthesis of BChl *c* of green bacteria (*39*) and *bchK*, whose product is the BChl *c* synthase (*40*), were found. A radical-SAM-type methyltransferase similar to BchR, which methylates C-12[1] of BChl *c* (*41*), was also identified. These data predicted that the acidobacterium should produce methylated Bchl *c* homologs and assemble chlorosomes like green sulfur bacteria. Consistent with this inference, a *csmA* gene, whose product is a BChl *a*-binding protein of the chlorosome baseplate (*42*), was also found in a putative operon with *bchG*, encoding a BChl *a* synthase (see Fig. 6S). Overlapping metagenome scaffolds and BAC clones physically linked these genes and *bchY*, a subunit of chlorophyllide *a* reductase, to the genes on the sequenced BAC insert (Fig. 6S). No oxygen-independent magnesium-protoporphyrin IX monomethyl ester oxidative cyclase *(bchE)* was identified. However, the presence of *acsF*, encoding an oxygen-dependent magnesium-protoporphyrin IX monomethyl ester oxidative cyclase (*41*, *43, 44*), implied that the

mat acidobacterium should synthesize BChl under *oxic* conditions (Fig. 12S). Genes with strong sequence similarity to *crtB*, *crtP*, and *crtH* further imply that the GFP1-like organism synthesizes carotenoids by a pathway similar to that in *Chlorobi* and *Cyanobacteria* (*45*). Although an *mcr* gene encoding malonyl-CoA reductase, an enzyme required for CO_2 fixation by the 3-hydroxypropionate pathway (*2*), was identified, no other genes specifically attributable to CO_2 fixation were identified. No genes encoding hydrogenases or enzymes for sulfide/sulfur oxidation were detected.

Supplemental Figure Legends

Figure 1S. Diagram comparing the relative positions of transmembrane α-helices and loops of various Type-1 reaction center polypeptides. Bars connecting the tops of the rectangles represent loops in the cytoplasm and bars connecting the bottoms of the rectangles represent loops in the periplasm or thylakoid lumen. Note the very large periplasmic loop between helices VII and VIII of the Cab. thermophilum. The positions of the F_X Cys ligands and the conserved histidine ligand to the reaction center special pair (P700, P840, P780) are indicated. The antenna (helices I to VI) and electron-transfer domains (helices VII to XI) of the polypeptides are indicated. Abbreviations and sequence alignments of selected Type-1 reaction center proteins can be found in Figures 2S and 3S.

Figure 2S. Sequence alignment of the N-terminal region of PscA of Cab. thermophilum and the homologous PscA, PshA, PsaA, and PsaB proteins of the Type-1 reaction centers and other photosynthetic bacteria. The region shown corresponds to that including transmembrane α–helices I through VII, including all of the antenna domain and the first transmembrane helix of the electron transfer domain; the positions of these helices for *Thermosynechococcus elongatus,* as derived from the X-ray structure for photosystem I (*47*), are indicated. Note that the long inserted loop sequence that follows helix VII of Cab. thermophilum has been truncated, since it is not related in sequence to any portion of any other reaction center protein. Hyphens indicate insertions/deletions included to optimize the alignment. Abbreviations: Cab., Chloracidobacterium; *Chl*., *Chlorobium; Hbc*., *Heliobacillus*; *Hbt*., *Heliobacterium*; *Pld*., *Pelodictyon; Ptc*., *Prosthecochloris*; *T.*, *Thermosynechococcus*.

Figure 3S. Sequence alignment of the C-terminal region of PscA and the homologous PshA, PsaA, and PsaB proteins of the reaction centers of Cab. thermophilum and other photosynthetic bacteria. The region shown corresponds to transmembrane α–helices VIII through XI, including most of the electron-transfer cofactor binding domain; the positions of these helices for *Thermosynechococcus elongatus* as derived from the X-ray structure for photosystem I (*47*) are shown. Hyphens indicate insertions/deletions included to optimize the alignment. The position of the conserved, cysteine-containing loop (red boxes enclose the conserved cysteine residues) that is predicted to provide the ligands to an intersubunit F_X-like, [4Fe-4S] cluster is indicated by the bar labeled F_X. The conserved His in Helix X that is the putative ligand to one BChl/Chl of the reaction center special pair is also indicated by a red box. Abbreviations are the same as in Figure 3S.

Figure 4S. Sequence alignment of PscB sequences from various organisms. Hyphens indicate insertions/deletions included to optimize the alignment. Abbreviations: *Cab*., Chloracidobacterium; *Chl*., *Chlorobium*; *Ptc*., *Prosthecochloris*. The eight conserved cysteines that are likely to be the ligands to two [4Fe-4S] clusters are indicated in red boxes and in the consensus sequence.

Figure 5S. Sequence alignment of selected FmoA proteins, the apoprotein of the BChl *a*-binding Fenna-Matthews-Olson protein. Hyphens indicate insertions/deletions included to optimize the alignment. Abbreviations: Cab., Chloracidobacterium; *Chl*., *Chlorobium*; and *Ptc*., *Prosthecochloris*.

Figure 6S. Physical map of a 950-kbp region of the consensus Cab. thermophilum genome, including BAC clone M60-018 J19 and showing the relative positions of the rRNA operon, *recA*, *pscA-pscB-fmoA, bchG-orf-csmA,* and *bchY* genes. A. Physical map of assembled scaffolds and contigs from the Mushroom Spring metagenome (dark blue bars) mapped onto the sequence of BAC clone M60-018 J19 (blue-green bar). This BAC clone is completely sequenced and comprises 271,846 bp (GenBank accession EF531339). The colored boxes indicate the positions of the 16S rRNA (purple), *recA* (red), and *pscAB-fmoA* (green) genes. B. Extension of the map in Panel A that shows the physical map of sequenced BAC clone M60-018 J19 (blue-green bar) and other BAC clones (gray bars) whose end-sequences (pink boxes) were mapped onto the scaffolds and contigs (dark blue bars) assembled from the Mushroom and Octopus Springs metagenome. The positions of the *bchG-orf-csmA* (yellow-green) and *bchY* (orange) genes are indicated, establishing that these genes are also physically linked to the 16S rRNA gene in BAC clone M60-018-J19.

Figure 7S. Comparison of the complete sequence of the 16S rRNA of Cab. thermophilum derived from BAC clone M60-018 J19 with the nearly complete sequence from *Acidobacterium* GFP1 (*37*). The red ellipses indicate the differences between the two sequences.

Figure 8S. Rooted neighbor-joining phylogram based upon the sequence alignment shown in Figure 9S, which shows the relationships of RecA sequences of Cab. thermophilum and various other organisms. The RadA-like protein (DR1105) of *Deinococcus radiodurans* was used as the outgroup sequence. Some of the major eubacterial phyla are listed at the right. Chlorophototrophs and the phyla that contain them are shown in colors, and all other phyla and

organisms are indicated in black. Bootstrap values were determined on the basis of 2000 resamplings.

Figure 9S. Alignment of the RecA sequences from the indicated organisms that was used to generate the tree shown in Figure 8S. Hypens indicate insertions/delections included to optimize the alignment. Abbreviations: Cab., Chloracidobacterium; *Car., Carboxydothermus*; *Cfx., Chloroflexus*; *Hb., Heliobacterium*; Pel., *Pelotomaculum*; *Pld., Pelodictyon*; *The., Thermoanaerobacter*.

Figure 10S. Rooted neighbor-joining phylogram derived from the sequence alignment shown in Figure 11S showing the relationship of the BchU sequence of Cab. thermophilum to the BchU and CrtF sequences of various organisms. The ComT methyltransferase of *Ara. thaliana* was used as the outgroup. Abbreviations: *A.*, *Arabidopsis;* Cab*.,* Chloracidobacterium*; Cfx.*, *Chloroflexus*; *Chl.*, *Chlorobium*; *Pld.*, *Pelodictyon*; *Ptc.*, *Prosthecochloris*; *Rhb.*, *Rhodobacter*; *Rps.*, *Rhodopseudomonas*; *Rvx.*, *Rubrivivax*. The bootstrap values shown are based on 1000 resamplings.

Figure 11S. Alignment of BchU, CrtF, and ComT sequences from the indicated organisms that was used to generate the tree shown in Figure 10S. Hyphens indicate insertions/deletions included to optimize the alignment. Abbreviations: *A., Arabidopsis*; Cab., Chloroacidobacterium; *Cfx., Chloroflexus*; *Chl., Chlorobium*; *Pld., Pelodictyon*; *Ptc., Prosthecochloris*; *Rhb., Rhodobacter*; *Rps.*, *Rhodopseudomonas*; *Rvx.*, *Rubrivivax*.

Figure 12S. Alignment of sequences for AcsF, the oxygen-dependent oxidative ring cyclase for Mg-protoporphyrin IX monomethyl ester in BChl and Chl biosynthesis (*41, 43, 44*). Hyphens indicate insertions/deletions included to optimize the alignment. Abbreviations: *A.*, *Arabidopsis;* Cab.*,* Chloracidobacterium*; Cfx.*, *Chloroflexus*; *G., Gloeobacter; Por., Porphyra; Rb.*, *Rhodobacter*; *Rfx., Roseiflexus*; *Rps.*, *Rhodopseudomonas*.

Figure 13S. Complete HPLC elution profile monitored at 667 nm for pigments extracted from an enrichment culture containing *Synechococcus* sp. strain JA-2-3B'a (2-13), Cab. thermophilum, and *Anoxybacillus* sp. The blue-colored block indicates the elution positions of BChl *c* homologs esterified with farnesyl (indicated by the subscript "F"). The yellow colored block indicates the elution position of BChl *c* homologs esterified with geranylgeraniol (indicated by the subscript "GG"). Most of the BChl *c* is esterified with alcohols more hydrophobic than geranylgeraniol (aqua colored block), and at least two distinct alcohol groups appear to be present because of the complexity of the pattern observed. The peaks numbered 1 to 4 indicate the typical pattern of [ethyl, methyl], [ethyl, ethyl], [propyl, ethyl] and [*i*-butyl, ethyl] BChl *c* homologs, and the letters "a" and "b" indicate peaks probably derived from two different esterifying alcohol groups. Peaks from Chl *a* of *Synechococcus* sp. strain JA-2-3B'a (2-13) are indicated by the two arrows. The inset shows the spectrum of peak 3b, which is characteristic of BChl *c* under these solvent conditions.

Figure 1S

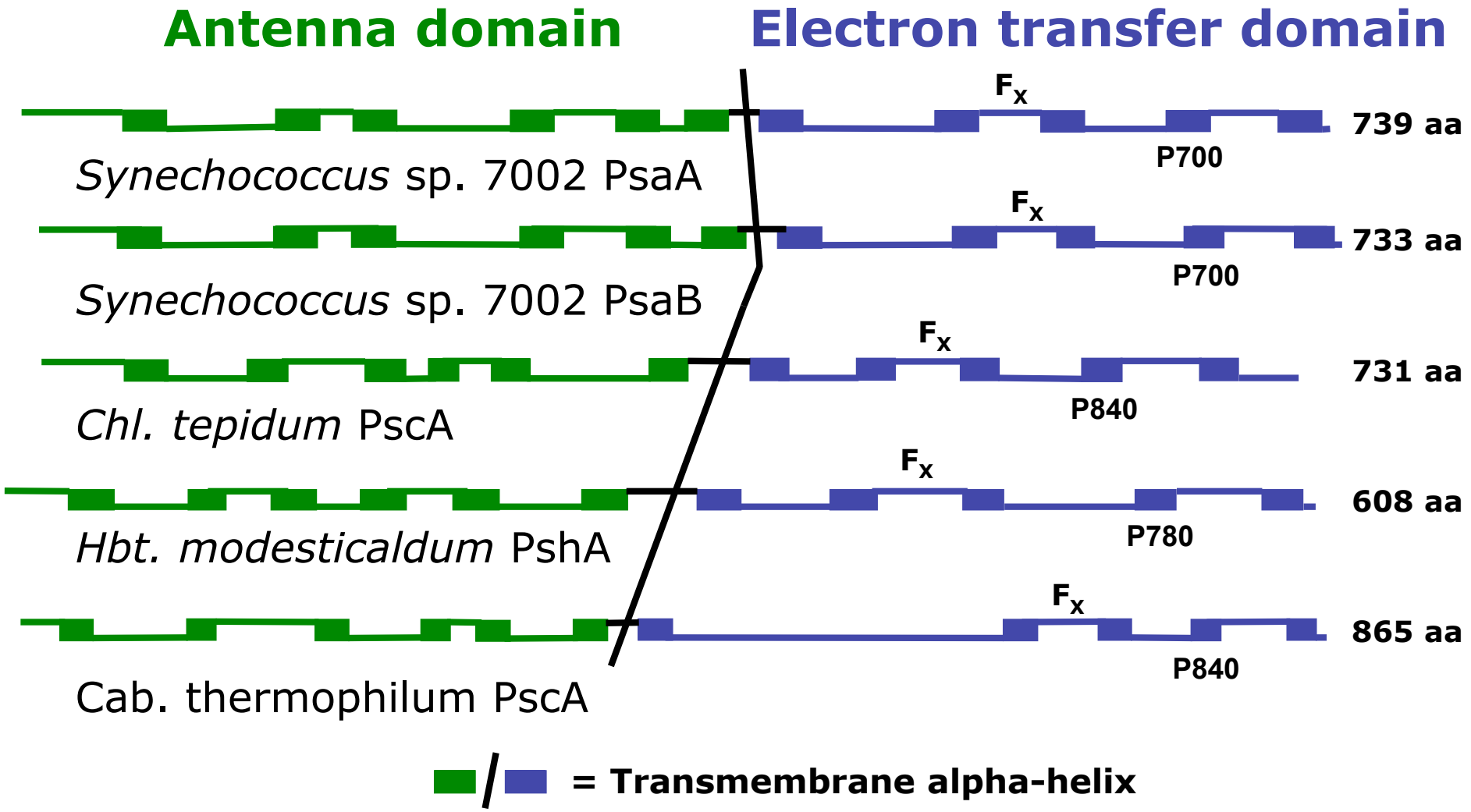
Antenna domain
Electron transfer domain
Fx
739 aa
P700
Synechococcus sp. 7002 PsaA
Fx
733 aa
P700
Synechococcus sp. 7002 PsaB
Fx
731 aa
P840
Chl. tepidum PscA
Fx
608 aa
P780
Hbt. modesticaldum PshA
Fx
865 aa
P840
Cab. thermophilum PscA
= Transmembrane alpha-helix

Figure 2S

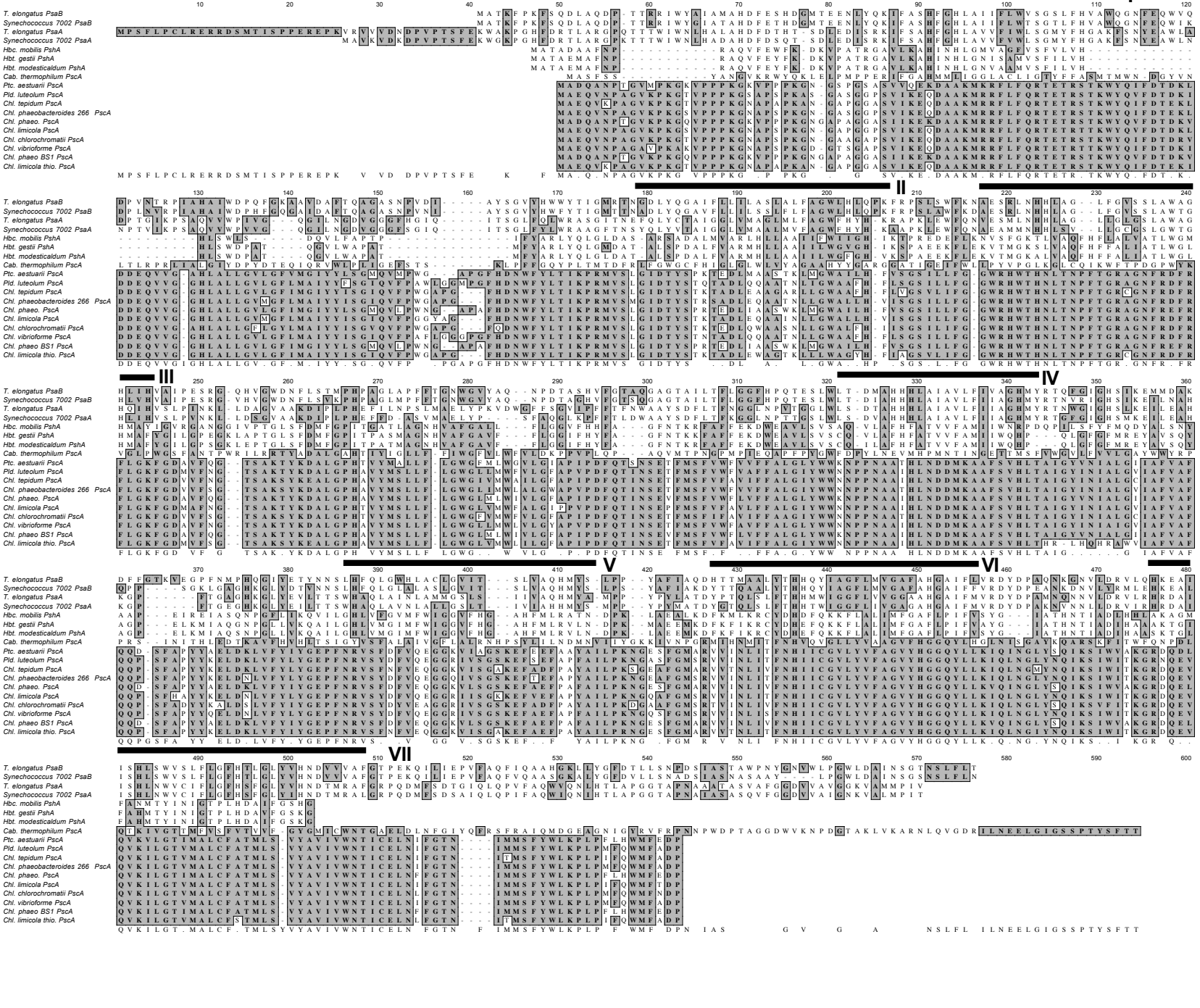
I
II
III
IV
V
VI
VII
T. elongatus PsaB
Synechococcus 7002 PsaB
T. elongatus PsaA
Synechococcus 7002 PsaA
Hbc. mobilis PshA
Hbt. gestii PshA
Hbt. modesticaldum PshA
Cab. thermophilum PscA
Ptc. aestuarii PscA
Pld. luteolum PscA
Chl. tepidum PscA
Chl. phaeobacteroides 266 PscA
Chl. phaeo. PscA
Chl. limicola PscA
Chl. chlorochromatii PscA
Chl. vibrioforme PscA
Chl. phaeo BS1 PscA
Chl. limicola thio. PscA

Figure 3S

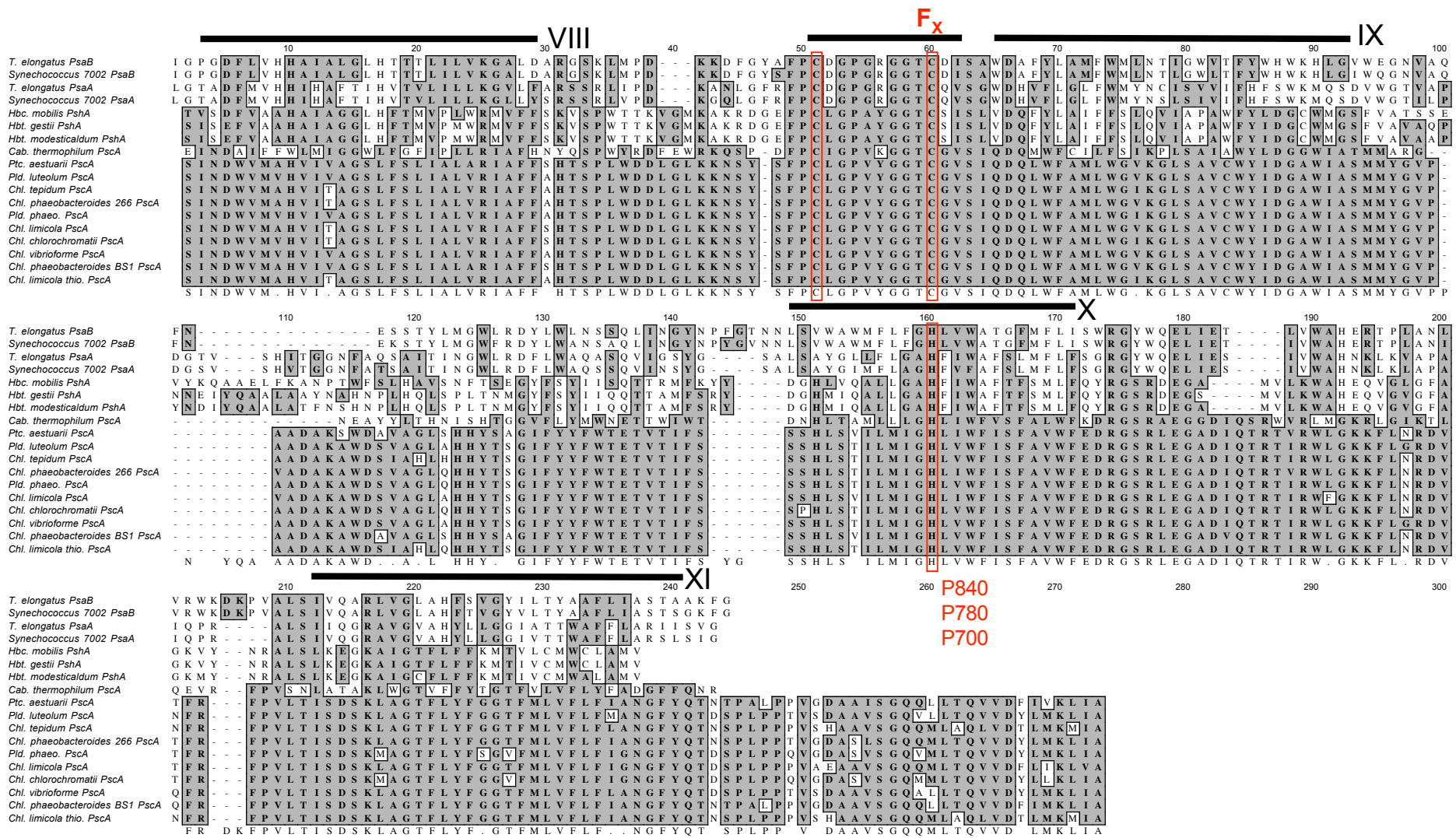
VIII
F_X
IX
X
XI
P840
P780
P700
T. elongatus PsaB
Synechococcus 7002 PsaB
T. elongatus PsaA
Synechococcus 7002 PsaA
Hbc. mobilis PshA
Hbt. gestii PshA
Hbt. modesticaldum PshA
Cab. thermophilum PscA
Ptc. aestuarii PscA
Pld. luteolum PscA
Chl. tepidum PscA
Chl. phaeobacteroides 266 PscA
Pld. phaeo. PscA
Chl. limicola PscA
Chl. chlorochromatii PscA
Chl. vibrioforme PscA
Chl. phaeobacteroides BS1 PscA
Chl. limicola thio. PscA

Figure 4S

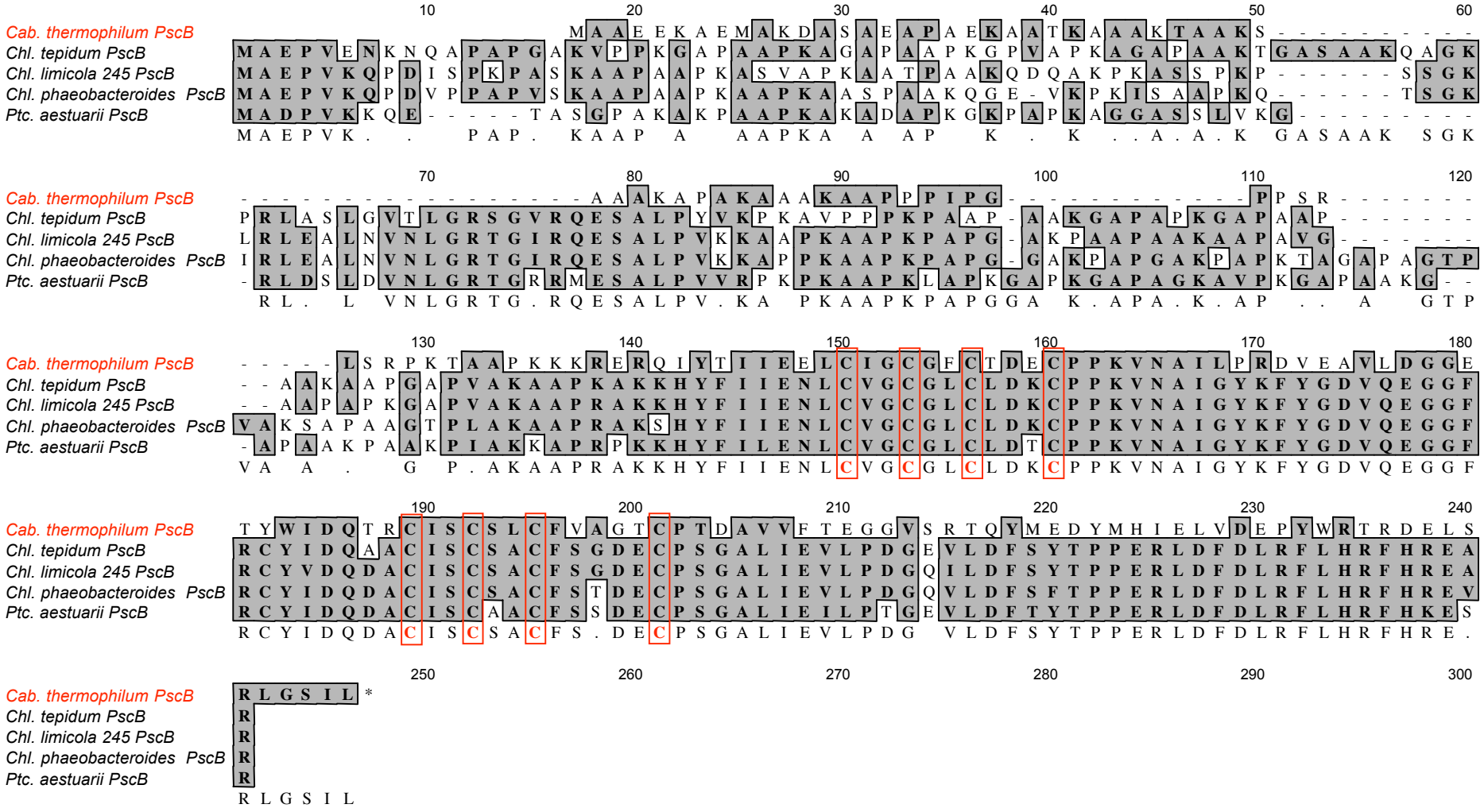
Cab. thermophilum PscB
Chl. tepidum PscB
Chl. limicola 245 PscB
Chl. phaeobacteroides PscB
Ptc. aestuarii PscB

Figure 5S

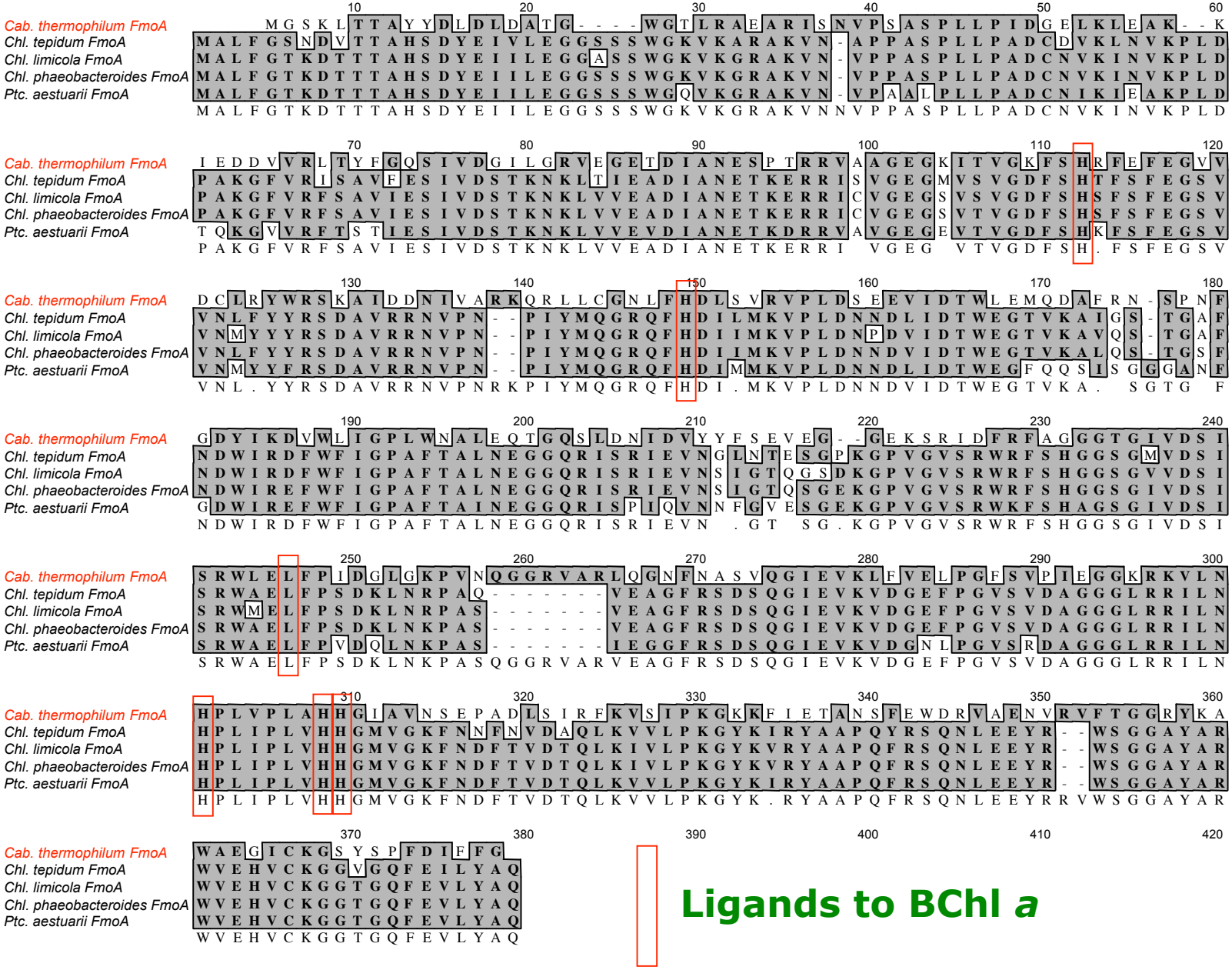
Cab. thermophilum FmoA
Chl. tepidum FmoA
Chl. limicola FmoA
Chl. phaeobacteroides FmoA
Ptc. aestuarii FmoA
Ligands to BChl a

Figure 6S

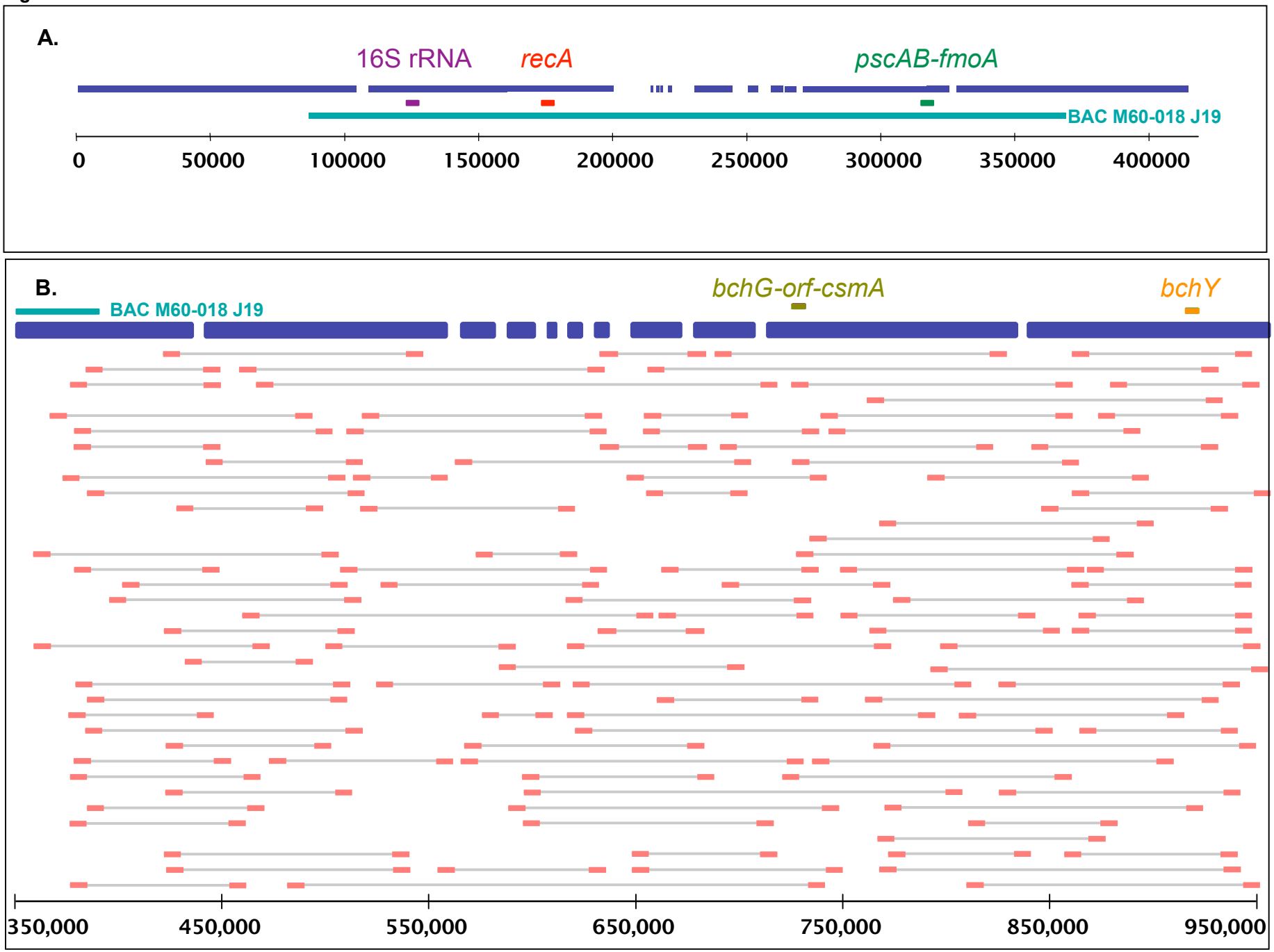
A.
16S rRNA
recA
pscAB-fmoA
BAC M60-018 J19
0
50000
100000
150000
200000
250000
300000
350000
400000
B.
BAC M60-018 J19
bchG-orf-csmA
bchY
350,000
450,000
550,000
650,000
750,000
850,000
950,000

Figure 7S

Figure 8S

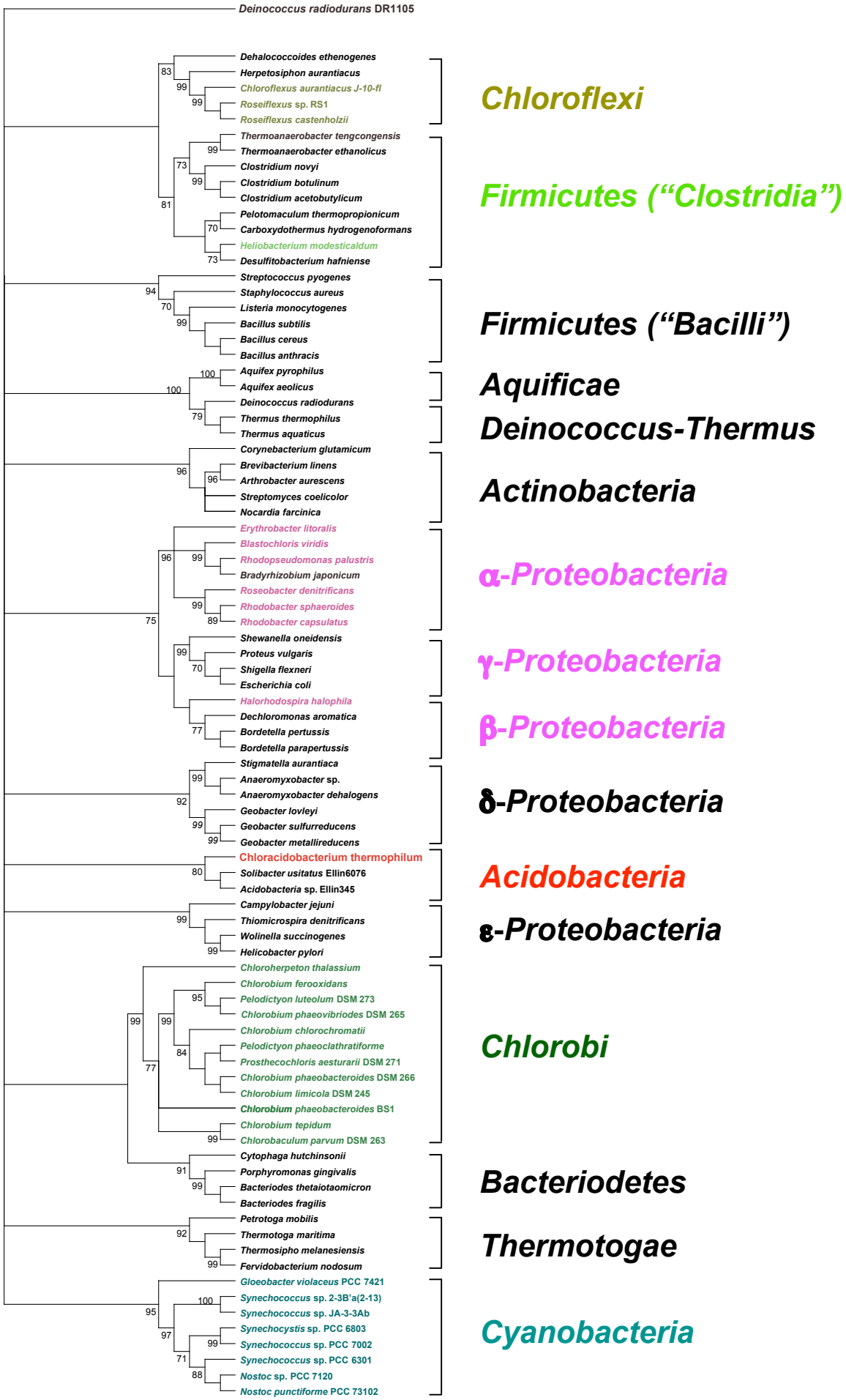
Deinococcus radiodurans DR1105
Dehalococcoides ethenogenes
Herpetosiphon aurantiacus
Chloroflexus aurantiacus J-10-fl
Roseiflexus sp. RS1
Roseiflexus castenholzii
Thermoanaerobacter tengcongensis
Thermoanaerobacter ethanolicus
Clostridium novyi
Clostridium botulinum
Clostridium acetobutylicum
Pelotomaculum thermopropionicum
Carboxydothermus hydrogenoformans
Heliobacterium modesticaldum
Desulfitobacterium hafniense
Streptococcus pyogenes
Staphylococcus aureus
Listeria monocytogenes
Bacillus subtilis
Bacillus cereus
Bacillus anthracis
Aquifex pyrophilus
Aquifex aeolicus
Deinococcus radiodurans
Thermus thermophilus
Thermus aquaticus
Corynebacterium glutamicum
Brevibacterium linens
Arthrobacter aurescens
Streptomyces coelicolor
Nocardia farcinica
Erythrobacter litoralis
Blastochloris viridis
Rhodopseudomonas palustris
Bradyrhizobium japonicum
Roseobacter denitrificans
Rhodobacter sphaeroides
Rhodobacter capsulatus
Shewanella oneidensis
Proteus vulgaris
Shigella flexneri
Escherichia coli
Halorhodospira halophila
Dechloromonas aromatica
Bordetella pertussis
Bordetella parapertussis
Stigmatella aurantiaca
Anaeromyxobacter sp.
Anaeromyxobacter dehalogens
Geobacter lovleyi
Geobacter sulfurreducens
Geobacter metallireducens
Chloracidobacterium thermophilum
Solibacter usitatus Ellin6076
Acidobacteria sp. Ellin345
Campylobacter jejuni
Thiomicrospira denitrificans
Wolinella succinogenes
Helicobacter pylori
Chloroherpeton thalassium
Chlorobium ferooxidans
Pelodictyon luteolum DSM 273
Chlorobium phaeovibrioides DSM 265
Chlorobium chlorochromatii
Pelodictyon phaeoclathratiforme
Prosthecochloris aesturarii DSM 271
Chlorobium phaeobacteroides DSM 266
Chlorobium limicola DSM 245
Chlorobium phaeobacteroides BS1
Chlorobium tepidum
Chlorobaculum parvum DSM 263
Cytophaga hutchinsonii
Porphyromonas gingivalis
Bacteriodes thetaiotaomicron
Bacteriodes fragilis
Petrotoga mobilis
Thermotoga maritima
Thermosipho melanesiensis
Fervidobacterium nodosum
Gloeobacter violaceus PCC 7421
Synechococcus sp. 2-3B'a(2-13)
Synechococcus sp. JA-3-3Ab
Synechocystis sp. PCC 6803
Synechococcus sp. PCC 7002
Synechococcus sp. PCC 6301
Nostoc sp. PCC 7120
Nostoc punctiforme PCC 73102
Chloroflexi
Firmicutes ("Clostridia")
Firmicutes ("Bacilli")
Aquificae
Deinococcus-Thermus
Actinobacteria
α-Proteobacteria
γ-Proteobacteria
β-Proteobacteria
δ-Proteobacteria
Acidobacteria
ε-Proteobacteria
Chlorobi
Bacteriodetes
Thermotogae
Cyanobacteria

Figure 9S

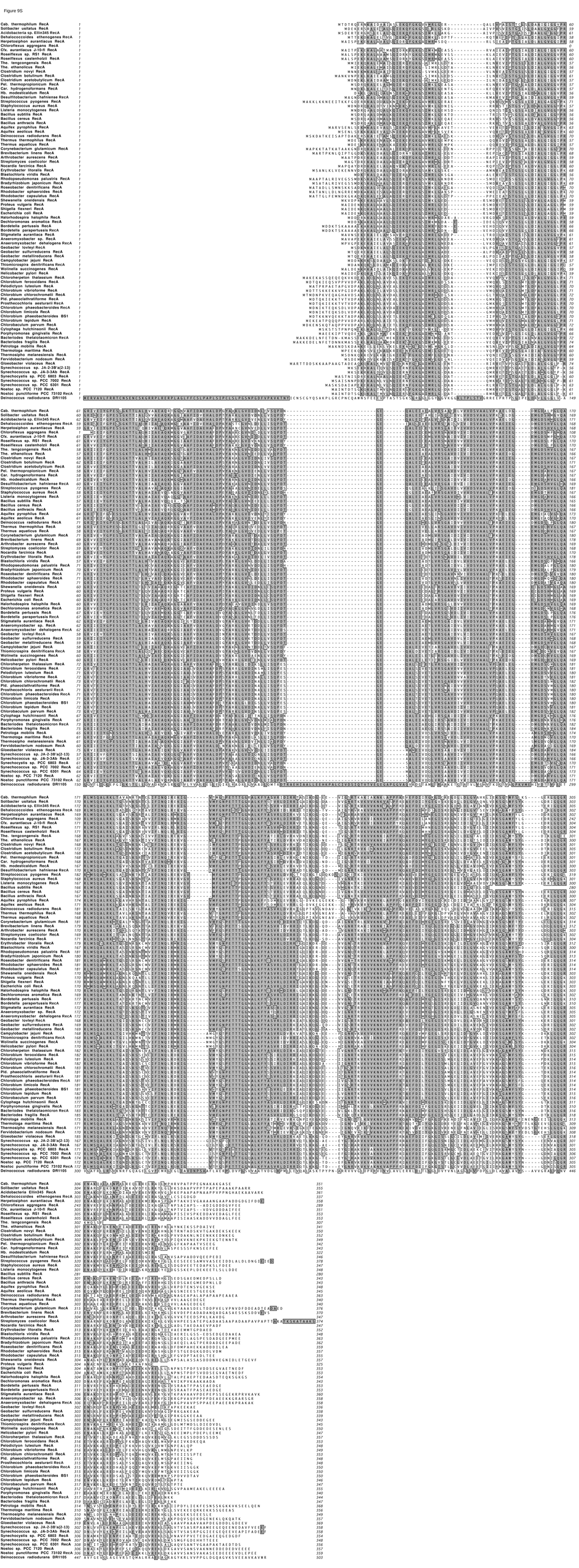

Figure 10S

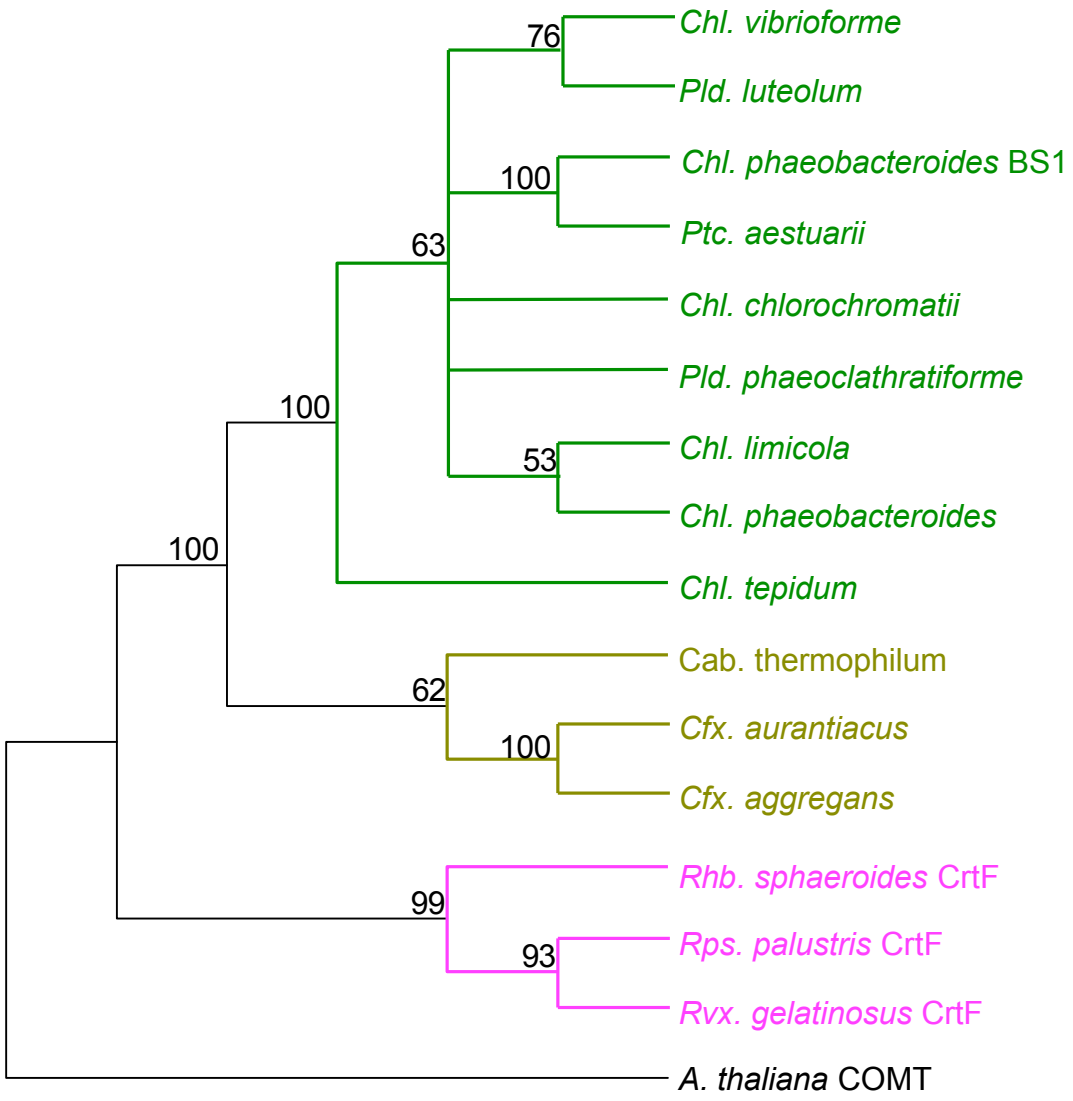
Chl. vibrioforme
Pld. luteolum
Chl. phaeobacteroides BS1
Ptc. aestuarii
Chl. chlorochromatii
Pld. phaeoclathratiforme
Chl. limicola
Chl. phaeobacteroides
Chl. tepidum
Cab. thermophilum
Cfx. aurantiacus
Cfx. aggregans
Rhb. sphaeroides CrtF
Rps. palustris CrtF
Rvx. gelatinosus CrtF
A. thaliana COMT
76
100
63
100
53
100
62
100
99
93

Figure 11S

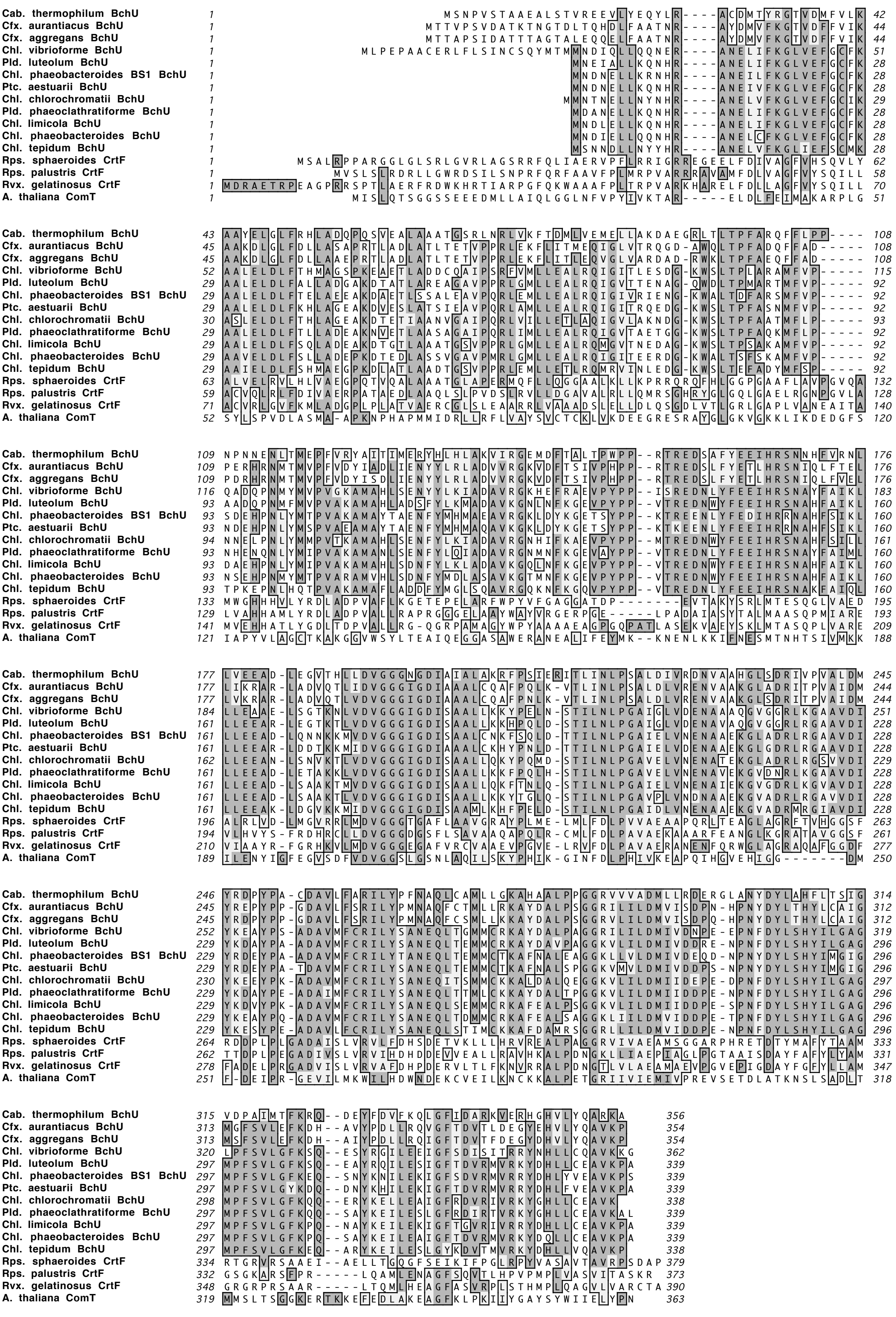
Cab. thermophilum BchU
Cfx. aurantiacus BchU
Cfx. aggregans BchU
Chl. vibrioforme BchU
Pld. luteolum BchU
Chl. phaeobacteroides BS1 BchU
Ptc. aestuarii BchU
Chl. chlorochromatii BchU
Pld. phaeoclathratiforme BchU
Chl. limicola BchU
Chl. phaeobacteroides BchU
Chl. tepidum BchU
Rps. sphaeroides CrtF
Rps. palustris CrtF
Rvx. gelatinosus CrtF
A. thaliana ComT

Figure 12S

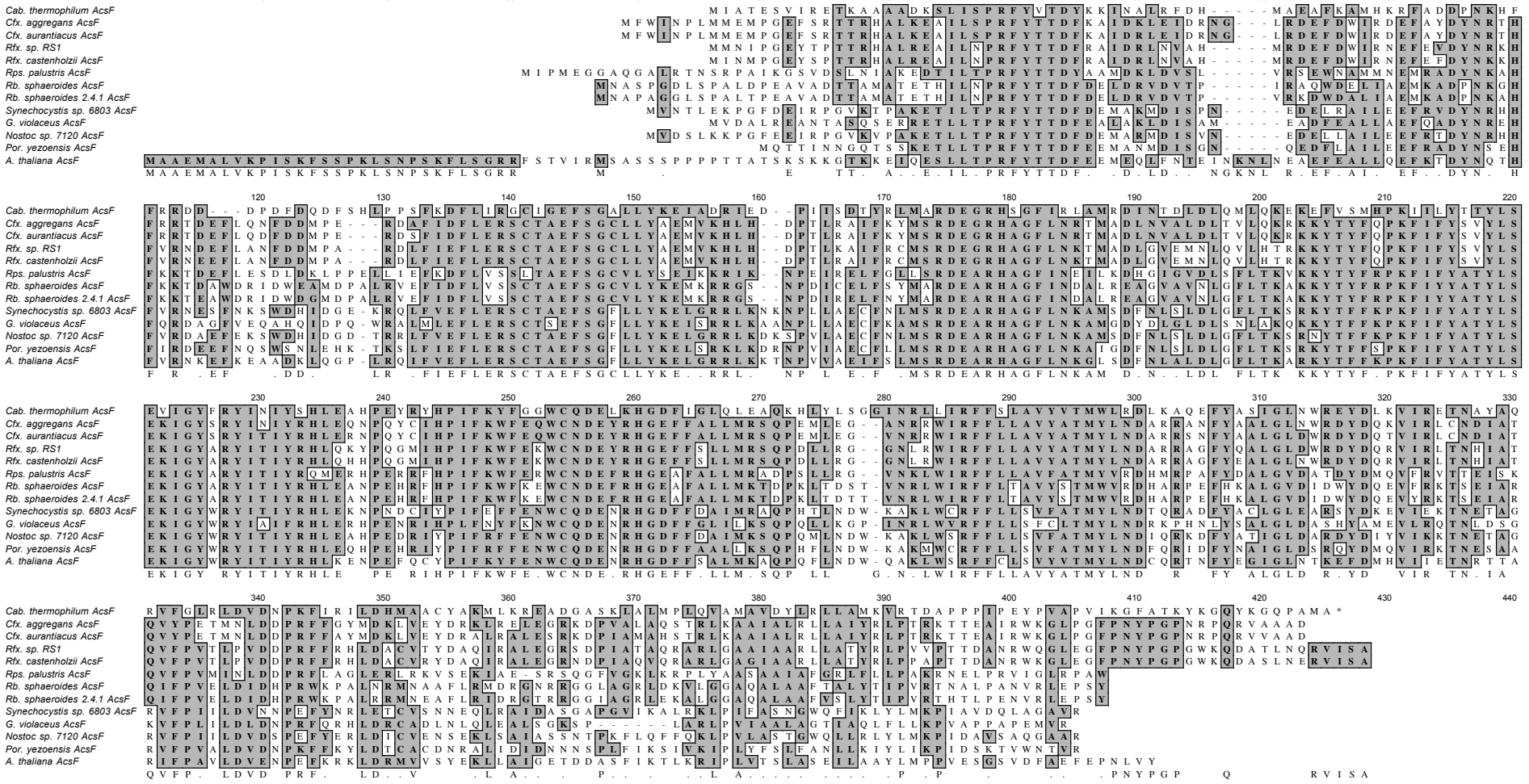
Cab. thermophilum AcsF
Cfx. aggregans AcsF
Cfx. aurantiacus AcsF
Rfx. sp. RS1
Rfx. castenholzii AcsF
Rps. palustris AcsF
Rb. sphaeroides AcsF
Rb. sphaeroides 2.4.1 AcsF
Synechocystis sp. 6803 AcsF
G. violaceus AcsF
Nostoc sp. 7120 AcsF
Por. yezoensis AcsF
A. thaliana AcsF

Figure 13S

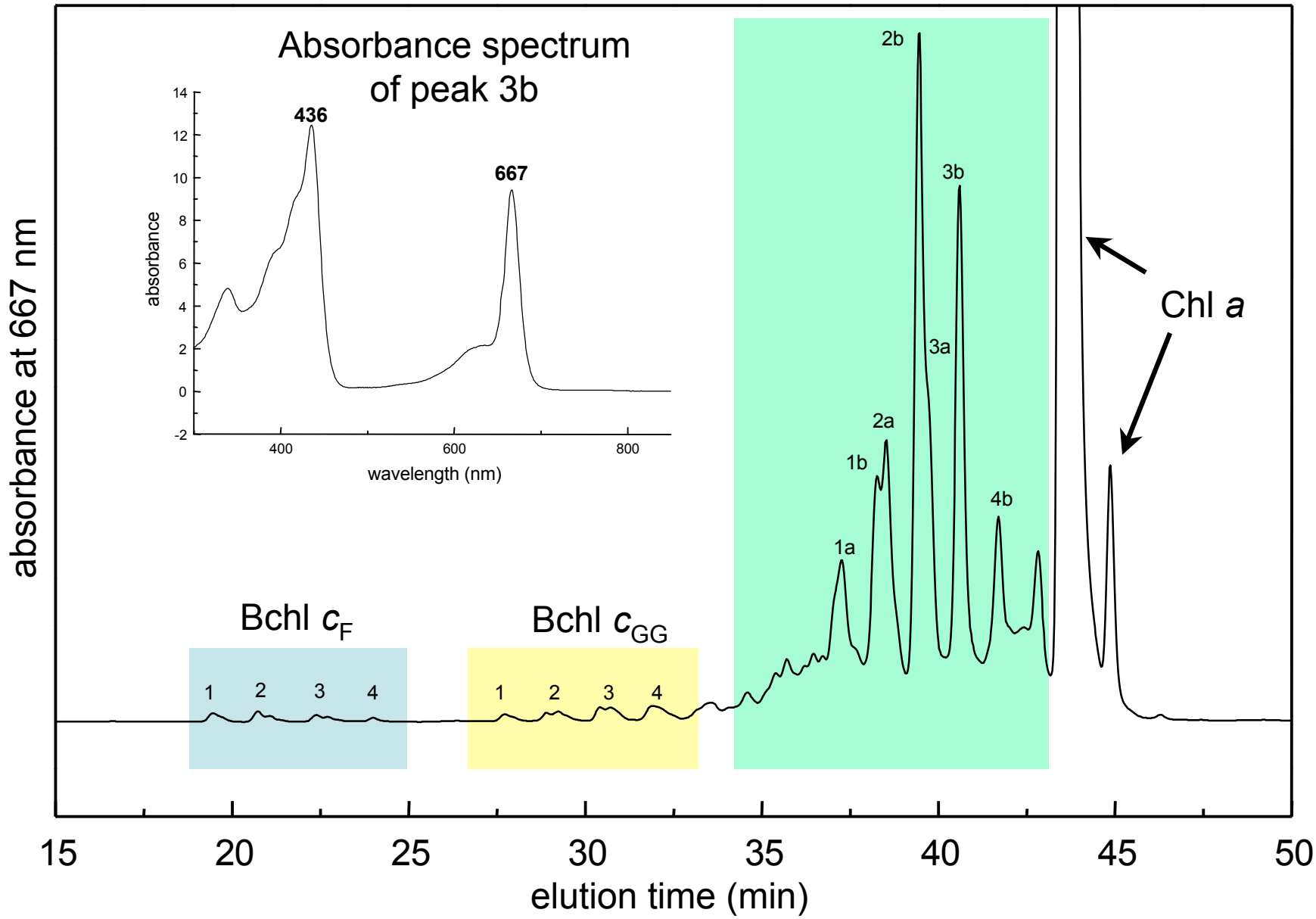
Absorbance spectrum
of peak 3b
436
667
absorbance
wavelength (nm)
absorbance at 667 nm
elution time (min)
Bchl c_F
Bchl c_{GG}
1
2
3
4
1a
1b
2a
2b
3a
3b
4b
Chl *a*

Table 1S. Oligonucleotide primers used in this study.

Gene	Description	Forward (5' to 3')	Reverse (5' to 3')
Cab. thermophilum 16S rRNA	Ribosomal RNA used as phylogenetic marker	GAT CCT GGC TCA GAA TC (*38*)	GGC TTA CAC AGG ATA CC
Anoxybacillus sp. 16S rRNA	Ribosomal RNA used as phylogenetic marker	CGA AAG TCT GAC GGA GCA AC	TCA TCG TTT ACG GCG TGG AC
Bacterial 16S rRNA	Ribosomal RNA used as phylogenetic marker	AGA GTT TGA TCA TGG CTC AG (*46*)	GGT TAC CTT GTT ACG ACT T (*46*)
Cab. thermophilum *recA*	DNA repair protein used as phylogenetic marker	GCA TGG AAG GCG CTG TAG CC	CAG ACA GCC TGC TTC CTA CC
Cab. thermophilum *pscA*	Type 1 reaction center protein	GCT TTT CGT CCT ATG CCA ACG	AGG GGT CAA ACC AAC CGT AAG G
Cab. thermophilum *fmoA*	BChl *a*-binding protein	GGC AAG ATT ACG GTC GGA AAG	CCC TCA ATA GGA ACA CTA AAC CCC
Cab. thermophilum *csmA*	BChl *a*-binding protein located in the baseplate of chlorosomes	GTC GTG GAC TCC ATC GTC	TCC TTC GGC TGT GAA ACG
Cab. thermophilum *bchU*	C-20 methyltransferase for BChl *c* biosynthesis	TTC TCT ACG AGC AGT ATT TGC GTG	CGG TGG ATT TCC TCA TAA AAC G
Cab. thermophilum *acsF*	Oxidative cyclase for BChl *a,c* biosynthesis under oxic conditions	CAG GAT TTC TCC CAT CTC CCA C	CCA ATA AAG TCG CCG TGT TTG AT

REFERENCES

1. M. T. Madigan *et al.*, in *Geothermal Biology and Geochemistry in YNP***,** W.I. Inskeep and T. R. McDermott, eds., pp. 203-219. Thermal Biol. Inst., Montana St. Univ., Bozeman, MT (2005).
2. C. G. Klatt *et al.*, *Environ. Microbiol.*, in press (doi: 10.1111/j.1462-2920.2007.01323.x) (2007).
3. http://genome.jgi-psf.org/mikc_home.html
4. D. Bhaya *et al.*, *ISME J.* **1**, in press (2007).
5. D. M. Ward *et al.*, *Heredity*, in press (2007).
6. http://www.tigr.org/tdb/ENVMGX/YNPHS/
7. http://landresources.montana.edu/FIBR/
8. J. C. Venter *et al.*, *Science* **304**, 66-74 (2004).
9. G. G. Sutton, O. White, M. D. Adams, A. R. Kerlavage*, Genome Seq. Technol.* **1**, 9-19 (1995).
10. H. Tettelin *et al.*, *Genomics* **62**, 500-507 (1999).
11. S. L. Salzberg, A. L. Delcher, S. Kasif, O. White, *Nucl. Acids Res.* **26**, 544-548 (1998).
12. J. F. Heidelberg *et al., Nat. Biotech.* **20**, 1118-1123 (2002).
13. J. A. Eisen *et al.*, *Proc. Natl. Acad. Sci. USA* **99**, 9509-9514 (2002).
14. J. F. Heidelberg *et al.*, *Nature* **406**, 477-483 (2000).
15. M. G. Claros, G. von Heijne, *Comput. Appl. Biosci.* **10**, 685-686 (1994).
16. A. Bateman *et al.*, *Nucl. Acids Res.* **28**, 263-266 (2000).
17. D. H. Haft *et al.*, *Nucl. Acids Res.* **29**, 41-43 (2001).
18. S. F. Altschul *et al., J. Mol. Biol.* **215**, 403-410 (1990)

19. S. F. Altschul *et al., Nucl. Acids Res.* **25**, 3389-3402 (1997).

20. J. P. Allewalt *et al.*, *Appl. Environ. Microbiol.* **72**, 544-550 (2006).

21. J. M. Dubbs, D. A. Bryant, *Mol. Microbiol.* **15**, 3073-3085 (1991).

22. D. C. Brune, T. Nozawa, R. E. Blankenship, *Biochemistry* **26**, 8644-8652 (1987).

23. E. V. Vassilieva *et al., Biochemistry* **41**, 4358-4370 (2002).

24. N.-U. Frigaard, G. D. Voigt, D. A. Bryant. *J. Bacteriol.* **184**, 3368-3372 (2002).

25. D. A. Bryant, N.-U. Frigaard, *Trends Microbiol.* **14**, 488-496 (2006).

26. D. R. Boone, R. W. Castenholz, eds., *Bergey's Manual of Systematic Bacteriology*, Volume 1, 2nd ed., Springer, New York, NY (2001).

27. G. A. Nadson, *Bull. Jard. Bot. St. Petersb.* **6**, 190 (1906)

28. B. K. Pierson, R. W. Castenholz, *Arch. Microbiol.* **100**, 5-24 (1974).

29. T. P. Turova *et al., Microbiol*. **75**, 235-244 (2006).

30. H. Gest, J. L. Favinger, *Arch. Microbiol.* **136**, 11-16 (1983).

31. G. Hauska, T. Schoedl, H. Remigy, G. Tsiotis, *Biochim. Biophys. Acta* **1507**, 260-277 (2001).

32. J. H. Golbeck, *Proc. Natl. Acad. Sci. USA* **90**: 1132-1136 (1993).

33. W.-D. Schubert *et al.*, *J. Mol. Biol.* **280**, 297-314 (1998).

34. A. Camara-Artigas, R. E. Blankenship, J. P. Allen, *Photosynth. Res.* **75**, 49-55 (2003).

35. U. Nübel *et al.*, *Appl. Environ. Microbiol.* **68**, 4593-4603 (2002).

36. R. T. Papke *et al.*, *Environ. Microbiol.* **5**, 650-659 (2003).

37. GenBank accession nos. CP000239, CP000240 and AAQU00000000

38. S. M. Barns, S. L. Takala, C. R. Kuske, *Appl. Environ. Microbiol.* **65**, 1731-1737 (1999).

39. J. A. Maresca *et al., J. Bacteriol.* **186**, 2558-2566 (2004).

40. N.-U. Frigaard, G. D. Voigt, D. A. Bryant, *J. Bacteriol.* **184**, 3368-3376 (2002).

41. A. Gomez Maqueo Chew, D. A. Bryant, *Annu. Rev. Microbiol.* **41**, in press (2007).

42. N.-U. Frigaard, D. A. Bryant, in *Complex Intracellular Structures in Prokaryotes,* J. M. Shively, Ed., pp. 79-114, Springer, Berlin, Gemany (2006).

43. V. Pinta, M. Picaud, F. Reiss-Husson, C. Astier, *J. Bacteriol.* **184**, 746-753 (2002).

44. S. Ouchane *et al.*, *J. Biol. Chem.* **279**, 6385-6394 (2004).

45. N.-U. Frigaard *et al.*, J. Baceriol. **186**, 5210-5220 (2004).

46. P. E. Eden *et al., Int. J. Syst. Bacteriol.* **41**, 324-325 (1991).

47. P. Jordan *et al.*, *Nature* **411**, 909-917 (2001).