

Tracking molecular evolution of photosynthesis by characterization of a major photosynthesis gene cluster from *Heliobacillus mobilis*

(bacteriochlorophyll biosynthesis/cytochrome *bc* complex/phylogenetic analysis/photosystem origin)

JIN XIONG*, KAZUHITO INOUE†, AND CARL E. BAUER*‡

*Department of Biology, Indiana University, Bloomington, IN 47405; and †Department of Biological Sciences, Kanagawa University, Hiratsuka, Kanagawa 259-1293, Japan

Communicated by Norman R. Pace, University of California, Berkeley, CA, October 13, 1998 (received for review July 21, 1998)

ABSTRACT A DNA sequence has been obtained for a 35.6-kb genomic segment from *Heliobacillus mobilis* that contains a major cluster of photosynthesis genes. A total of 30 ORFs were identified, 20 of which encode enzymes for bacteriochlorophyll and carotenoid biosynthesis, reaction-center (RC) apoprotein, and cytochromes for cyclic electron transport. Donor side electron-transfer components to the RC include a putative RC-associated cytochrome *c*₅₅₃ and a unique four-large-subunit cytochrome *bc* complex consisting of Rieske Fe-S protein (encoded by *petC*), cytochrome *b*₆ (*petB*), subunit IV (*petD*), and a diheme cytochrome *c* (*petX*). Phylogenetic analysis of various photosynthesis gene products indicates a consistent grouping of oxygenic lineages that are distinct and descendent from anoxygenic lineages. In addition, *H. mobilis* was placed as the closest relative to cyanobacteria, which form a monophyletic origin to chloroplast-based photosynthetic lineages. The consensus of the photosynthesis gene trees also indicates that purple bacteria are the earliest emerging photosynthetic lineage. Our analysis also indicates that an ancient gene-duplication event giving rise to the paralogous *bchI* and *bchD* genes predates the divergence of all photosynthetic groups. In addition, our analysis of gene duplication of the photosystem I and photosystem II core polypeptides supports a “heterologous fusion model” for the origin and evolution of oxygenic photosynthesis.

Bacterial photosynthesis is found in five eubacterial phyla: cyanobacteria, purple bacteria, heliobacteria, green sulfur bacteria, and green nonsulfur bacteria. Analysis of photosystems from purple and green nonsulfur bacteria shows that these organisms synthesize primitive photosystems that are ancestral to the more complex oxygen-evolving photosystem II (PSII) from cyanobacteria and chloroplasts (1). Additional studies of heliobacteria and green sulfur bacteria indicate that they synthesize photosystems that are ancestral to the photosystem I (PSI) complex from cyanobacteria and chloroplasts (2).

With the exception of the sequence analysis of PSI and PSII structural polypeptides, there is little information on the origin and evolution of photosynthesis. Complete sequence information on photosynthesis genes is only available for the purple nonsulfur bacterium, *Rhodobacter capsulatus*, which has a major clustering of photosynthesis genes (3), and for the cyanobacterium *Synechocystis* sp. PCC 6803, for which sequence analysis has been completed for the entire chromosome. Information on photosynthesis genes from the three other eubacterial photosynthetic phyla is limited mainly to the reaction-center (RC) core polypeptides and a few cytochromes.

With the aim of providing more substantive information on the evolution of photosynthesis, we have undertaken an extensive characterization of photosynthesis genes from *Heliobacillus mobilis*. A gene cluster was found to encode enzymes for bacteriochlorophyll and carotenoid biosynthesis, as well as the RC core polypeptide and cytochromes involved in photosynthetic electron transfer. Phylogenetic studies based on this new sequence information provide important clues for the unique evolutionary position of *H. mobilis* in the course of development of photosynthesis.

MATERIALS AND METHODS

DNA Sequencing. A genomic library was constructed by partial digestion of *H. mobilis* (ATCC 43427) genomic DNA with *Sau3A* I, followed by ligating the DNA fragments into the *Bam*HI site of the cosmid vector pJRD215 (4). Cosmid containing transformants were pooled and used for complementation of bacteriochlorophyll-biosynthesis mutants of *R. capsulatus* through bipartite matings as described (5). One cosmid, pHM6, which complemented a *bchB* mutant (6), was chosen for sequencing analysis. Additional DNA sequences flanking the insert in pHM6 were obtained by the inverse PCR method as described (7). DNA sequencing was performed with a fluorescent dye-labeled dideoxynucleotide sequencing reaction kit (Amersham) with an Applied Biosystems DNA sequencer (model 377, Perkin—Elmer). Oligonucleotide primers were synthesized by Operon Technologies (Alameda, CA) and GIBCO/BRL. All of the DNA sequences reported in this study were determined completely on both strands.

Phylogenetic Analysis. The nucleotide sequence was analyzed with SEQUENCHER software (version 3.0, Gene Codes, Ann Arbor, MI). Database searches were performed by using the Basic Local Alignment Search Tool (BLAST; ref. 8). Sequence compilation and analysis of translated ORFs were performed with the aid of the Genetics Computer Group's sequence-analysis package (version 9.0, Madison, WI). Multiple sequence alignments of translated gene sequences were carried out with the computer program CLUSTAL W (version 1.7; ref. 9). Phylogenetic trees were generated by the neighbor-joining method with the CLUSTAL W program (with 100 bootstrap replicates), by the maximum-parsimony method (with 100 branch-bound replicates) with the PAUP program (version 3.1.1; ref. 10), and by the maximum-likelihood method (1,000 replicates) with the PUZZLE programs (version 3.1).

RESULTS AND DISCUSSION

Photosynthesis Gene Cluster. *H. mobilis* photosynthesis genes were cloned initially by complementing *R. capsulatus*

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/9514851-6\$2.00/0 PNAS is available online at www.pnas.org.

Abbreviations: PSI, photosystem I; PSII, photosystem II; RC, reaction center.

Data deposition: The nucleotide sequence reported in this paper has been deposited in the GenBank database (accession no. AF080002).

‡To whom reprint requests should be addressed. e-mail: cbauer@sunflower.bio.indiana.edu.

bacteriochlorophyll-biosynthesis mutants with a *H. mobilis* genomic-cosmid library. Complete sequence analysis of cosmid pHM6, which complemented mutations in the protochlorophyllide reductase genes *bchL*, *bchN*, and *bchB*, indicated that it contained an 8,355-bp region coding for several genes involved in bacteriochlorophyll biosynthesis (*bchM*, *bchE*, *bchL*, *bchN*, *bchB*, *bchI*, and *bchD*), as well as part of the RC structural polypeptide coded by the previously sequenced *pshA* gene (Fig. 1). The DNA sequence flanking the region contained in pHM6 was determined by chromosome walking (7) until regions of nonphotosynthesis genes were obtained. This analysis generated a 35,618-bp contiguous sequence that contains 30 closely linked ORFs. Functions for all but three of the gene products were assigned with a relatively high degree of confidence based on sequence homology analysis (Table 1). We identified 20 clustered genes encoding enzymes and proteins involved in photosynthesis. Most photosynthesis genes are closely spaced, with several gene pairs (*bchE*–*bchL*, *bchL*–*bchN*, *bchN*–*bchB*, and *bchI*–*bchD*) overlapping. With the exception of two genes, *cobQ* and *murC*, all of the genes are transcribed in the same orientation.

Pigment-Biosynthesis Genes. Heliobacteria synthesize a unique photopigment, bacteriochlorophyll *g*, which has an interesting structural similarity to chlorophyll *a* from oxygenic phototrophs (11). Bacteriochlorophyll *g* contains an ethylidene side group at C8 of ring II that, in the presence of oxygen and light, isomerizes to produce a vinyl group, giving rise to a ring structure that is the same as in chlorophyll *a*. Inspection of ORFs in the photosynthesis gene cluster indicates that all of the genes involved in steps of the bacteriochlorophyll/chlorophyll biosynthesis pathway between Mg-chelation and formation of chlorophyllide (*bchI*, *bchD*, *bchH*, *bchJ*, *bchM*, *bchE*, *bchL*, *bchN*, and *bchB*) are present in the cluster. The cluster also contains *bchG*, which is involved in esterification of the tetrapyrrole with farnesol. No unique genes are present in the photosynthesis gene cluster that have an obvious role in synthesizing/stabilizing the ethylidene side group at C8. This indicates that either an unidentified gene(s) may be involved in the formation of the ethylidene side group or that it may be generated as a product by one of the homologs of the chlorophyll *a* biosynthesis genes. *bchI*, *bchD*, and *bchH*, which code for subunits of Mg-chelatase, are linked tightly, as are *bchL*, *bchN*, and *bchB*, which code for three subunits of light-independent protochlorophyllide reductase. The gene cluster also contains both *hemA*, which codes for glutamyl tRNA reductase, the enzyme involved in the first step of tetrapyrrole biosynthesis, and *tspO* (previously known as *crtK*), which codes for a homolog to a peripheral-type benzodiazepine receptor that has been implicated in the transport of porphyrins (12).

Heliobacteria are unique among photosynthetic organisms in that they synthesize a C₃₀ carotenoid (4,4'-diaponeurosporene) rather than a C₄₀ carotenoid (neurosporene; ref. 13). There are only a few other eubacterial species, such as

Staphylococcus aureus, known to synthesize C₃₀ carotenoids. Two genes in *S. aureus* carotenoid biosynthesis have been identified. One (*CrtM*) codes for dehydroqualene synthase, and the other (*CrtN*) codes for diapophytoene desaturase (14). In the *H. mobilis* photosynthesis gene cluster, a homolog to *crtN* was found downstream of *bchH*.

Reaction Center and Cytochromes. The gene cluster contains at least six genes involved in primary photochemistry (RC) and electron transport (cytochromes). The RC gene *pshA* is located among the major group of *bch* genes. Although this gene was previously sequenced from the same strain used in our study (15), our analysis indicated several variations from the published sequence, including four frame shifts that resulted in an alteration of 34 amino acid codons (5.5% of the entire sequence) near the center and the C terminus of the protein. There is also a putative rho-independent termination site located in the intervening sequence between *pshA* and *bchM*, indicating that *pshA* may not be cotranscribed with the large cluster of downstream *bch* genes.

The cytochrome gene *petJ* is found downstream of *hemA*. This gene was found to encode cytochrome *c*₅₅₃, and a partial sequence of *petJ* was recently reported (16). It also shows a high degree of similarity to cytochrome *c*₅₅₃ from *Heliobacterium gestii* (17). Cytochrome *c*₅₅₃ is involved in electron transfer between the cytochrome *bc* complex and the RC in *H. mobilis* (18). *PetJ* has a single heme *c* binding motif, CNSCH (starting at amino acid residue 79), with Cys-79 and Cys-82 as covalent heme binding ligands and His-83 and Met-125 as axial heme ligands. The *petJ* coding sequence is followed immediately by a putative rho-independent transcription-termination site.

Located downstream of *petJ* is a cluster of *pet* genes that form an apparent operon that codes for a unique cytochrome *bc* complex. A comparison of the *H. mobilis* *pet* gene organization with the *pet* operons from several other species and organelles is shown in Fig. 2. Purple bacteria synthesize a three-subunit cytochrome *bc*₁ complex consisting of cytochrome *b*, cytochrome *c*₁, and the Rieske Fe-S protein (19). Cyanobacteria, algae, and plants synthesize a functionally similar cytochrome *b₆f* complex that is composed of four major core subunits, the Rieske Fe-S protein, cytochrome *f* (equivalent to cytochrome *c*₁), cytochrome *b₆*, and subunit IV (20). Cytochrome *b₆* and subunit IV are considered equivalent to cytochrome *b* of purple bacteria, in which these two regions are fused into a single polypeptide.

Inspection of *pet* genes of *H. mobilis* indicates that this species synthesizes an unconventional *bc* complex resembling both cytochrome *bc*₁ of purple bacteria and cytochrome *b₆f* of cyanobacteria. Specifically, this putative operon contains *petC*, which codes for a Rieske Fe-S protein that has a conserved 2Fe-2S site (Cys-87, His-89, Cys-108, and His-111) as observed in other species. The Rieske Fe-S protein of *H. mobilis* also shares high sequence similarity to a homolog from the green sulfur bacterium *Chlorobium limicola*, which coincides with a

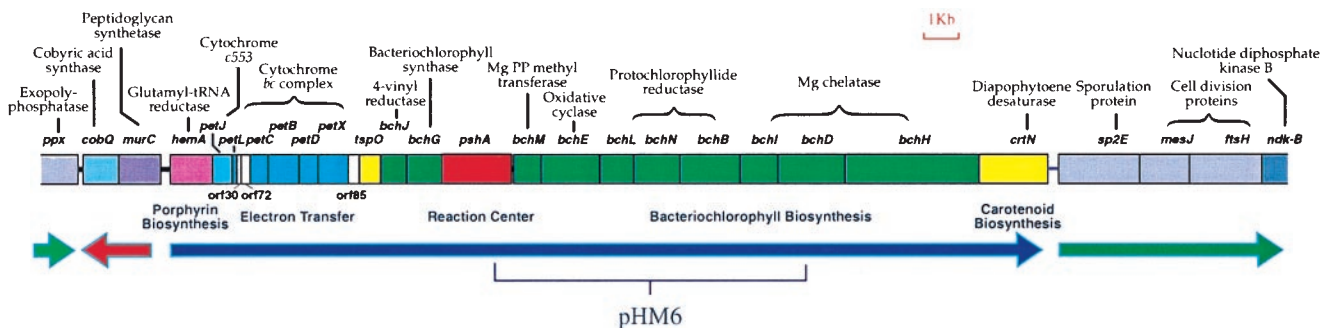


FIG. 1. The *H. mobilis* photosynthesis gene cluster. Directions of transcription are indicated by thick solid arrows with the plasmid pHM6 insert sequence delineated with a bracket.

Table 1. The putative functions and similarity matches of the ORFs of *Heliobacillus mobilis* identified in this study

Gene	Best database match, organism, function	Alignment score, E-value
<i>hemA</i>	Glutamyl-tRNA reductase, <i>Bacillus subtilis</i> , heme biosynthesis	319, 2 × 10 ⁻⁸⁶
<i>petJ</i>	Cytochrome <i>c</i> ₅₅₃ , <i>Heliobacterium gestii</i> , electron donor to the photosynthetic reaction center	163, 4 × 10 ⁻⁴⁰
<i>petC</i>	Rieske iron sulfur protein, <i>Thermus thermophilus</i> , subunit of cytochrome <i>bc</i> complex	63, 9 × 10 ⁻¹⁰
<i>petB</i>	Cytochrome <i>b</i> ₆ , liverwort, subunit of cytochrome <i>b</i> ₆ <i>f</i> complex	220, 4 × 10 ⁻⁵⁷
<i>petD</i>	Cytochrome <i>b</i> /subunit IV, <i>Cairina moschata</i> , subunit of cytochrome <i>bc</i> ₁ complex	79, 1 × 10 ⁻¹⁴
<i>petX</i>	Cytochrome <i>c</i> , <i>Bradyrhizobium japonicum</i> , subunit of cytochrome <i>bc</i> complex	56, 2 × 10 ⁻⁷
<i>tspO</i>	Peripheral-type benzodiazepine receptor, <i>Archaeoglobus fulgidus</i> , involved in transport of porphyrin	105, 2 × 10 ⁻²²
<i>bchJ</i>	Bacteriochlorophyll synthase, <i>Rhodobacter capsulatus</i> , 4-vinyl reduction	94, 4 × 10 ⁻¹⁹
<i>bchG</i>	Chlorophyll synthase, <i>Arabidopsis thaliana</i> , esterification of chlorophyllide with geranylgeraniol-PP	203, 1 × 10 ⁻⁵¹
<i>pshA</i>	Reaction-center core polypeptide, <i>Heliobacillus mobilis</i> , primary photochemical reaction	1182, 0
<i>bchM</i>	Mg protoporphyrin IX methyl transferase, <i>Synechocystis</i> sp. 6803	142, 1 × 10 ⁻³³
<i>bchE</i>	Mg-protoporphyrin IX monomethyl ester oxidative cyclase, <i>Rhodobacter capsulatus</i>	624, 1 × 10 ⁻¹⁷⁸
<i>bchL</i>	Subunit of light-independent protochlorophyllide reductase, <i>Cyanophora paradoxa</i>	260, 6 × 10 ⁻⁶⁹
<i>bchN</i>	Subunit of light-independent protochlorophyllide reductase, <i>Plectonema boryanum</i>	345, 2 × 10 ⁻⁹⁴
<i>bchB</i>	Subunit of light-independent protochlorophyllide reductase, <i>Plectonema boryanum</i>	375, 1 × 10 ⁻¹⁰³
<i>bchI</i>	Subunit of Mg-chelatase, <i>Synechocystis</i> sp. 6803	393, 1 × 10 ⁻¹⁰⁹
<i>bchD</i>	Subunit of Mg-chelatase, <i>Pisum sativum</i>	509, 1 × 10 ⁻¹⁴³
<i>bchH</i>	Subunit of Mg-chelatase, <i>Synechocystis</i> sp. 6803	1165, 0
<i>crtN</i>	Diapophytoene dehydrogenase, <i>Myxococcus xanthus</i> , carotenoid biosynthesis	351, 7 × 10 ⁻⁹⁶
<i>ppx</i>	Exopolyphosphatase, <i>Synechocystis</i> sp., response to signals of amino acid starvation	129, 2 × 10 ⁻²⁹
<i>cobQ</i>	Cobyrinic acid synthase, <i>Methanobacterium thermoautotrophicum</i> , coenzyme B ₁₂ biosynthesis	65, 4 × 10 ⁻¹⁰
<i>murC</i>	UDP-N-acetylmuramyl tripeptide synthetase, <i>Methanobacterium thermoautotrophicum</i> , cell-wall synthesis	130, 2 × 10 ⁻²⁹
<i>sp2E</i>	Stage II sporulation protein E, <i>Bacillus megaterium</i> , sporulation	236, 4 × 10 ⁻⁶¹
<i>mesJ</i>	Cell-cycle protein, <i>Bacillus subtilis</i> , cell division	115, 7 × 10 ⁻²⁵
<i>ftsH</i>	ATP-dependent zinc metalloproteinase, <i>Bacillus subtilis</i> , cell division	677, 0
<i>ndk-B</i>	Nucleoside diphosphate kinase B, <i>Flaveria bidentis</i> , biosynthesis of nucleoside triphosphates	159, 6 × 10 ⁻³⁹

finding from the EPR spectral analysis that showed a close relationship between the Rieske Fe-S proteins of these two species (21).

The next gene is *petB*, the product of which exhibits a significant sequence similarity to cytochrome *b*₆ from plants and cyanobacteria. This cytochrome contains two conserved histidine pairs (His-84 and His-185 pair and His-98 and His-200 pair), which noncovalently bind two heme groups, located in the transmembrane helices II and IV. The presence of cytochrome *b*₆ in *H. mobilis* is supported by recent electron-transfer studies that indicated that flash-induced kinetics of cytochrome *b* oxidation of the *H. mobilis* *bc* complex were more similar to those of algal and plant *b*₆*f* than those of purple bacterial *bc*₁ (22). Interestingly, the *C. limicola* *bc* complex contains a cytochrome *b* composed of fused *b*₆ and subunit IV domains (Fig. 2). This *bc* complex also has biochemical characteristics between those of the *bc*₁ and *b*₆*f* complexes with key features resembling the *b*₆*f* type (23).

Downstream of *petB* is *petD*, the product of which exhibits significant homology (30% identity and 61% similarity) to subunit IV from plants and cyanobacteria. Like its counterparts, the *H. mobilis* subunit IV contains three transmembrane domains (determined by hydropathy analysis; data not shown).

Study with subunit IV of the *b*₆*f* complex has indicated a role for the chloroplast *b*₆*f* complex in quinone binding, which is essential for the catalytic function (24).

Immediately downstream of *petD* is an ORF that codes for a protein exhibiting similarity to a membrane-bound diheme cytochrome *c* from *Bradyrhizobium japonicum* (25). This cytochrome contains two putative heme *c* binding sites, CASCH and CAACH, starting at amino acid residues 35 and 133, respectively, with His-39, Met-81, His-137, and Met-176 forming potential axial ligands. In the absence of biochemical data, we tentatively designated this putative cytochrome *c* subunit of the *bc* complex as PetX. This ORF also shows homology to a monoheme cytochrome (CytM) of cyanobacteria (with a slightly lower score), which apparently is involved in electron transfer between plastocyanin/cytochrome *c*₅₅₃ and cytochrome oxidase (26). As heliobacteria are obligate anaerobes, it is possible that this protein performs a dual function, first in its association with other subunits of the *bc* complex and second as an alternative electron donor to the RC, as was suggested for cytochrome *c*₅₅₁ in *C. limicola* (23).

In addition to genes coding for the major cytochrome *bc* subunits, there are three additional small ORFs located upstream of *petC* (orf29, orf30, and orf72) and one downstream of *petX* (orf85). Biochemical analysis indicates that cytochrome *b*₆*f* has three small subunits (subunits V, VI, and VII, or PetG, PetL, and PetM, respectively) in addition to the four major subunits discussed above. These additional subunits have a common feature of being of low molecular mass (≈4 kDa) and having one single transmembrane helix (27). Direct sequence comparison of the *H. mobilis* small ORFs with the *b*₆*f* small subunits indicated considerable sequence similarity between PetL of *Cyanophora paradoxa* and orf29 (28% identity and 38% similarity). This ORF was thus designated as *petL*. No obvious similarity is, however, observed between orf30, which does not seem to contain transmembrane segments, and the small *b*₆*f* subunits or any known sequences in GenBank. In addition, there is limited sequence similarity of orf72 and orf85 with subunits 10 and 11, respectively, of the bovine mitochon-

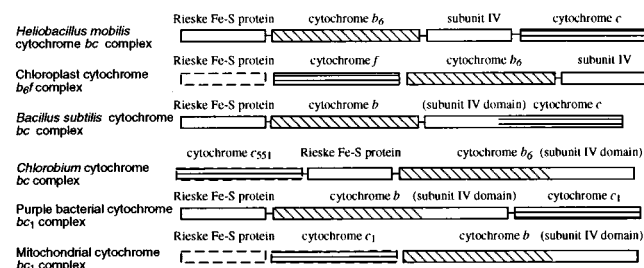


Fig. 2. Comparison of organizations of *pet* genes coding for cytochrome *bc*/*bc*₁/*b*₆*f* complexes from various organisms. The nuclear-encoded genes are enclosed in dashed lines. Separated boxes indicate nonlinked genes on chromosome.

drial cytochrome *bc*₁ complex (19.4% identity and 51.6% similarity relative to subunit 10; 17.9% identity and 58.9% similarity for subunit 11). These additional *H. mobilis* ORFs may thus code for components of this unique cytochrome *bc* complex.

Evolutionary Analyses. *H. mobilis* photosynthesis genes were compared with homologous sequences in the GenBank database to study evolutionary relationships. Phylogenetic trees were generated from multiple aligned sequences by using the maximum-parsimony, neighbor-joining, and maximum-likelihood methods. As the results from the three methods are highly congruent, we shall mainly discuss results obtained from the neighbor-joining tree analysis. A BchL/ChlL phylogenetic tree was constructed by using the nitrogenase iron protein NifH from an archaeon species *Methanococcus jannaschii* as an outgroup (Fig. 3A). Fujita *et al.* (28) and Burke *et al.* (29) found remarkable similarity between NifH and BchL. NifH was also used previously as an outgroup for studying phylogenetic relationships among BchL homologs from bacteria and plants (30). Its choice as an outgroup is based on the premise that nitrogen fixation occurs in both archaea and bacteria, whereas tetrapyrrole-based photosynthesis occurs only in bacteria and bacterially derived chloroplasts. Thus, unlike nitrogen fixation, photosynthesis may not predate the divergence of archaea and bacteria. A notable feature of the highly supported BchL tree is that oxygenic phototrophs together form a monophyletic group distinct from the anoxygenic phototrophs *R. capsulatus* and *H. mobilis* (Fig. 3A). The paraphyletic *R. capsulatus* and *H. mobilis* lineages are rooted closer to the base of the tree and form an outgroup to the oxygenic clade, suggesting that anoxygenic phototrophs are ancestral to the oxygenic phototrophs. This result is in agreement with the previous phylo-

genetic analysis by Burke *et al.* (30), who concluded that bacteriochlorophyll biosynthesis evolved earlier than chlorophyllous biosynthesis. However, this conclusion was later challenged by Lockhart *et al.* (31) who argued for an alternative hypothesis. Another notable feature of the tree is that *R. capsulatus* is a deeper branching lineage than *H. mobilis*, suggesting its more primitive evolutionary status.

A similar ancestral relationship between anoxygenic and oxygenic phototrophs is observed with the rooted BchH/ChlH tree with CobN from *M. jannaschii* as an outgroup (Fig. 3B). CobN is a subunit of Co-chelatase responsible for insertion of Co into tetrapyrrole during biosynthesis of coenzyme B₁₂. Rooting with CobN is based on the extensive sequence similarity of CobN to the Mg-chelatase subunit BChH/ChlH (32) and on the argument that the biosynthesis of coenzyme B₁₂, which is involved in a wide variety of cellular biochemical processes including the biosynthesis of nucleotides, may have preceded photosynthesis. A significant feature of this rooted tree is that it shows a better resolution of phylogenetic relationships among the phototrophic lineages with high bootstrap support. In addition to separation of anoxygenic and oxygenic lineages, with the former being an exclusive common ancestor to the latter, this tree places *H. mobilis* nearest to *Synechocystis* sp. PCC 6803, which is consistent with the previous 16S rRNA analysis (33). *C. vibrioforme* and *R. capsulatus* are deeper branching lineages relative to *H. mobilis*, although the placement of *C. vibrioforme* in a more basal position than *R. capsulatus* is considered tentative, because this placement is supported neither by the phylogenetic trees discussed below nor by the 16S rRNA tree (33). A monophyletic cyanobacterial origin of green plants is also evident in this tree with the support of an extremely high bootstrap value

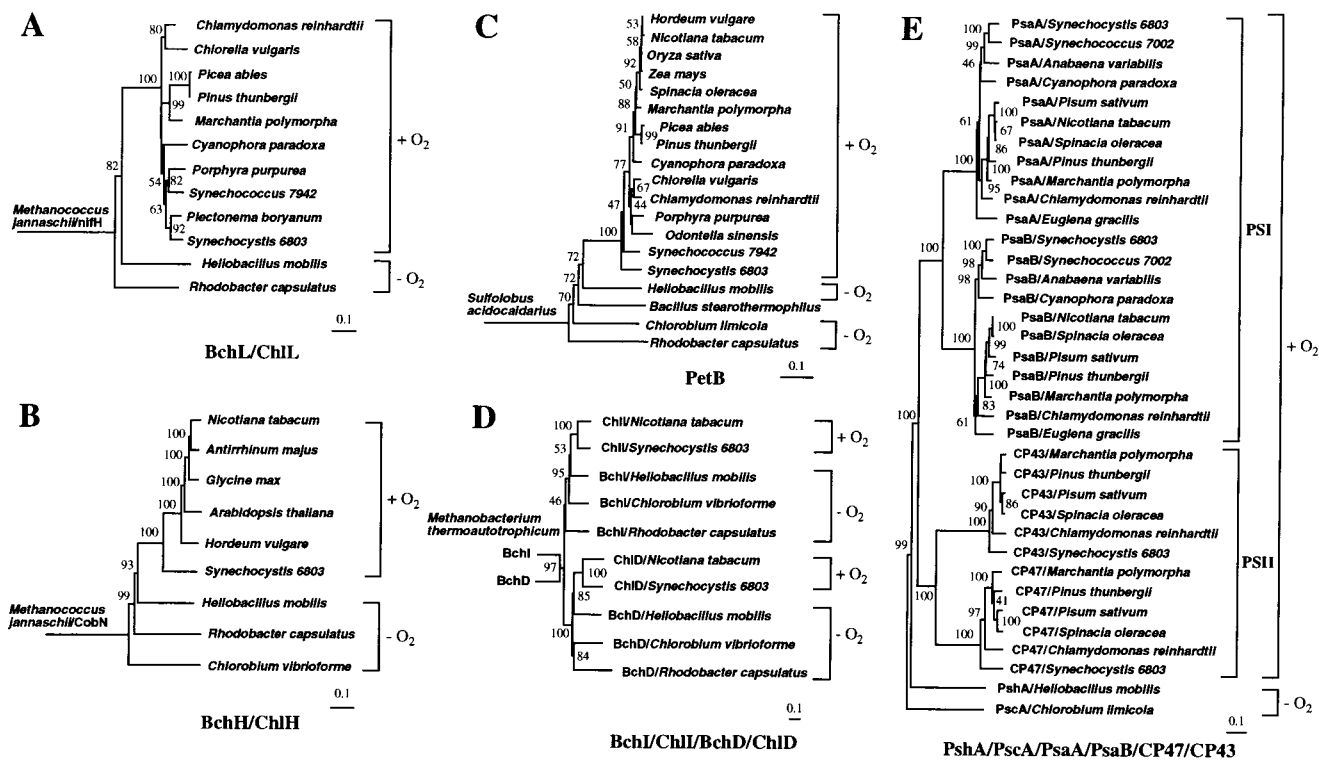


FIG. 3. (A) This phylogenetic tree of BchL/ChlL was created with the neighbor-joining method. NifH from *M. jannaschii* is used as an outgroup. Photosynthetic taxa are delineated by brackets and identified as oxygenic (+O₂) or anoxygenic (-O₂). Horizontal branch lengths represent relative evolutionary distances with the scale bar corresponding to 0.1 amino acid substitutions per site. Bootstrap values above 40% are indicated at each node, except basal nodes. (B) This phylogenetic tree of BchH/ChlH used CobN from *M. jannaschii* as an outgroup. (C) This phylogenetic tree of PetB used PetB from *Sulfolobus acidocaldarius* as an outgroup. (D) This composite phylogenetic tree of paralogous gene products Mg-chelatase subunits BchI/ChlI/BchD/ChlD used BchI and BchD homolog from *Methanobacterium thermoautotrophicum* as an outgroup. (E) This composite phylogenetic tree of PSI-like RC polypeptides (PshA/PscA/PsaA/PsaB) together with two PSII core antenna polypeptides (CP47 and CP43) used PscA of *C. limicola* as an outgroup.

(100%), which confirms many lines of evidence that cyanobacteria are the endosymbiotic origin of plastids in photosynthetic eukaryotes (e.g., ref. 34).

A cytochrome *b/b₆* (PetB) parsimony tree was constructed that was rooted with PetB of an archaeon *Sulfolobus acidocaldarius* (Fig. 3C). This rooted tree corroborates the grouping of oxygenic lineages distinct from that of anoxygenic ones as was observed with the Bch/Chl trees. In this tree, *H. mobilis* is shown again as the closest relative to cyanobacteria. The neighbor-joining tree also places *R. capsulatus* as the deepest lineage and *Bacillus stearothermophilus* intermediately in between *C. limicola* and *H. mobilis*, which may indicate the possibility of lateral gene transfer for the cytochrome *b₆* gene from a green sulfur bacteria-like common ancestor to the low-G+C Gram-positive bacteria. The more basal placement of *R. capsulatus* relative to *C. limicola* in this tree is in congruence with the 16S rRNA tree without the inclusion of green nonsulfur bacteria (33). Another notable feature of this tree is significantly less sequence diversity among oxygenic photosynthetic lineages as shown by relatively short branch lengths, which reflects a highly conserved nature of cytochrome *b₆* in these organisms, indicating that PetB may also be a good phylogenetic marker for oxygenic photosynthesis.

Phylogenetic analyses of other photosynthesis gene products (BchI, BchD, BchB, BchN, and BchM) are in agreement with the three representative trees shown in Fig. 3 A–C (data not shown). Thus, one can conclude that photosynthesis genes from anoxygenic photosynthetic organisms are ancestral to those found in oxygenic photosynthetic organisms. Comparing phylogenetic trees of photosynthesis genes vs. nonphotosynthesis genes, however, has been complicated by the various versions of trees available in the literature (e.g., 16S rRNA; refs. 33, 35), which further justifies the use of photosynthesis genes for tracking the molecular evolution of photosynthesis.

In terms of phylogenetic position, it is the current belief, based on the common RC type, that heliobacteria share the closest relationship with green sulfur bacteria. However, our quantitative analyses of evolutionary distances for various other photosynthesis genes from cyanobacteria, heliobacteria, and green sulfur bacteria do not support this idea. The analysis of pairwise adjusted mean distances among the three bacterial groups calculated for 15 genes (data not shown) indicates that heliobacteria are closer to cyanobacteria for 12 genes and that all three groups are of equal distance for 2 genes (including the RC polypeptides), whereas only 1 gene supports the idea that heliobacteria are closer to green sulfur bacteria.

Gene-Duplication Events. Analysis of several photosynthesis genes also indicates that some may have undergone a different developmental pathway, such as gene duplication. For example, *bchI* (or *chlI*) and *bchD* (or *chlD*), which encode two Mg-chelatase subunits, share significant similarity at the amino acid level (for *H. mobilis*, BchI and BchD polypeptides exhibit 31.0% identity and 67.9% similarity). These genes, which are nearly always linked in eubacterial genomes, are considered paralogous genes with a common evolutionary ancestry (36). To determine whether a gene-duplication event occurred before or after the divergence of current photosynthetic organisms, we conducted phylogenetic analysis of all available BchI and BchD pairs in the GenBank database. A joint phylogenetic tree with high bootstrap support was generated with a BchI and BchD homolog pair from *Methanobacterium thermoautotrophicum* as root. (The function of BchI and BchD homologs in the nonphotosynthetic archaeon is not yet known. However, we speculate that they may be subunits of a Ni-chelatase for the biosynthesis of the Ni-containing coenzyme F430, which is essential for the production of methane in methanogens.) As indicated in Fig. 3D, all of the BchI/ChlI sequences are clustered in a clade distinct from the BchD/ChlD clade, suggesting that an ancient gene duplication may have occurred well before photosynthetic speciation events. In

addition, the topology in each subtree corroborates the result from the PetB tree (Fig. 3C), indicating that (i) *H. mobilis* is closest to the last common ancestor of all oxygenic lineages, (ii) *C. vibrioforme* and *R. capsulatus* are ancestral to *H. mobilis*, and (iii) in the absence of green nonsulfur bacteria, *R. capsulatus* is the earliest emerging photosynthetic lineage.

Another striking example of gene duplication is found with the gene for the PSI-like RC core polypeptide. *H. mobilis* and *C. limicola* use a single RC core polypeptide (PshA and PscA, respectively) to form a homodimeric PSI-type RC complex. In contrast, oxygenic species synthesize a heterodimeric PSI RC with two highly related polypeptides, PsaA and PsaB. A composite phylogenetic tree (Fig. 3E) of the PSI RC core polypeptides from a broad representation of oxygenic lineages with PscA of *C. limicola* as root indicates that PsaA and PsaB are divided into two distinct clades and are derived from an ancient gene duplication from a PscA/PshA-like common ancestor. Thus, the RC gene duplication seems to have occurred well before diversification of all oxygenic lineages. Interestingly, PSI RC polypeptides also share significant similarity to CP47 (PsbB) and CP43 (PsbC; ref. 2), both of which are core antenna polypeptides of PSII. Our phylogenetic analysis (Fig. 3E) indicates that these genes have undergone gene duplication from a *pshA*-like common ancestor that preceded the divergence of all oxygenic lineages.

Origin of Oxygenic Photosynthesis. Several models have been put forward to account for the origin and evolution of PSI and PSII in oxygenic organisms. Many of these models (e.g., ref. 37) are based on subunit composition and functionality of photosystems rather than on rigorous phylogenetic reconstruction based on multiple sequence comparisons. Currently, there are three models that have obtained the most attention: the selective-loss model (38), a model proposed by Vermaas (2), and the fusion model (1, 39).

Olson and Pierson (38) propose a selective-loss model in which two linked photosystems analogous to PSI and PSII in cyanobacteria developed very early in evolution. Subsequent loss of one of the photosystems would give rise to the existing anoxygenic photosynthetic bacteria that have only a single PSI- or PSII-type photosystem. However, this model is not supported by our phylogenetic analysis (Fig. 3), which clearly indicates that photosynthesis genes from anoxygenic photosynthetic organisms are ancestral to those from oxygenic photosynthetic organisms.

A model described by Vermaas (2) involves the assumption that an ancestral RC resembling that of heliobacteria or *Chlorobium* evolved into both PSI- and PSII-type RCs. Part of the model suggests that PSI in oxygenic organisms was evolved from gene duplication and divergence from an anoxygenic homodimeric RC, a suggestion supported by our phylogenetic data (Fig. 3E). The model also proposes that the PSII-type RC in both anoxygenic and oxygenic organisms is derived from an anoxygenic PSI-type photosystem through gene splitting, operon duplication, and loss of the Fe-S center. The implicit assumptions are that heliobacteria and green sulfur bacteria preceded both oxygenic and purple bacteria and that purple bacteria were the latest emerging photosynthetic lineage. This part of the model is clearly not supported by our phylogenetic analysis, which indicates that purple bacteria are ancestral to heliobacteria and cyanobacteria (Fig. 3). Recent x-ray crystal-structure determination of PSI (37) has provided additional information and added complexity to the complete picture regarding the origin and evolution of oxygenic photosynthesis. The structure shows significant similarity between the purple bacterial RC and PSI core in cofactor arrangements and transmembrane helix topology; this similarity was taken to support the hypothesis that the two types of photosystems share a common ancestry and that the progenitor of all photosystems may be a PSI-like RC (37). However, the conclusion was drawn in the absence of identifiable sequence

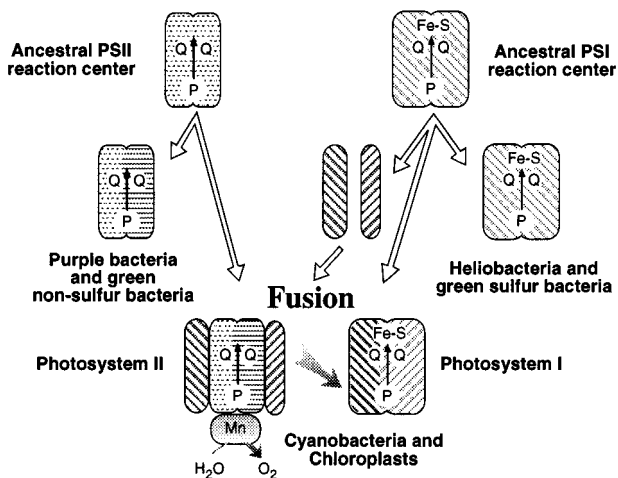


FIG. 4. A heterologous fusion model for the evolution of oxygenic photosynthesis based on phylogenetic analysis. Details of the model are provided in the text.

similarity between the two types of RC polypeptides. The structural similarity may thus indicate that the two types of RCs are either highly remote homologs or results from convergent evolution from two distinct origins. Our sequence comparisons support the convergent-evolution scenario.

Mathis (39) and Blankenship (1) suggest a fusion model in which a primordial photosynthetic organism containing a single homodimeric photosystem gave rise to the existing lineages of anoxygenic bacteria that contain only a single PSI- or PSII-type photosystem. Subsequent fusion of these two types of anoxygenic photosystems and development of water-oxidizing capability gave rise to oxygenic photosynthesis. This model is supported by our rooted-parsimony analysis of various photosynthesis genes (Fig. 3), which clearly indicates that purple bacterial and heliobacterial photosynthesis genes are ancestral to cyanobacterial, algal, and plant photosynthesis genes. However, this model must be modified in light of the observation of Vermaas (2), who showed that the CP47 and CP43 PSII core antenna polypeptides exhibit sequence similarity with the light-harvesting domains of the anoxygenic PSI-type RC polypeptide, which is corroborated by our parsimony-tree analysis (Fig. 3E). This evidence indicates that the oxygenic PSII RC contains antenna components derived from an ancient anoxygenic PSI RC. Consequently, we propose a "heterologous fusion model" (Fig. 4) that indicates that PSII from oxygenic organisms are derived from a fusion of the ancestral form of anoxygenic PSI- and PSII-type RCs. In this model, the PSII D1 and D2 core polypeptides are obtained from an anoxygenic PSII-type ancestor, presumably a homodimer form that is most similar to the purple bacterial L/M polypeptides (1). In addition, the PSII CP47 and CP43 core polypeptides are derived from gene fragmentation and subsequent duplication from a common ancestor to the anoxygenic PSI-type RC polypeptide.

We thank Drs. Yuichi Fujita and Chris Parkinson for valuable suggestions. This work was supported by National Institutes of Health Grants GM53940 and GM00618 (to C.E.B.) and by a grant from the US-Japan Cooperative Program on Photoconversion and Photosynthesis Research (to K.I.).

- Blankenship, R. E. (1992) *Photosynth. Res.* **33**, 91–111.
- Vermaas, W. F. J. (1994) *Photosynth. Res.* **41**, 285–294.
- Alberti, M., Burke, D. H. & Hearst, J. E. (1995) in *Anoxygenic Photosynthetic Bacteria*, eds. Blankenship, R. E., Madigan, M. T.

- & Bauer, C. E. (Kluwer, Dordrecht, the Netherlands), pp. 1083–1106.
- Davison, J., Heusterspreute, M., Chevalier, N., Ha-Thi, V. & Brunel, F. (1987) *Gene* **51**, 275–280.
- Young, D. A., Bauer, C. E., Williams, J. C. & Marrs, B. L. (1989) *Mol. Gen. Genet.* **218**, 1–12.
- Bollivar, D. W., Suzuki, J. Y., Beatty, J. T., Dobrowski, J. M. & Bauer, C. E. (1994) *J. Mol. Biol.* **237**, 622–640.
- Ochman, H., Gerber, A. S. & Hartl, D. L. (1988) *Genetics* **120**, 621–623.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
- Swofford, D. L. (1993) PAUP, Phylogenetic Analysis Using Parsimony (Illinois Natural History Survey, Champaign, IL), Version 3.1.1.
- Brockmann, H., Jr., & Lipinski, A. (1983) *Arch. Microbiol.* **136**, 17–19.
- Yeliseev, A. A. & Kaplan, S. (1995) *J. Biol. Chem.* **270**, 21167–21175.
- Takaichi, S., Inoue, K., Akaike, M., Kobayashi, M., Oh-oka, H. & Madigan, M. T. (1997) *Arch. Microbiol.* **168**, 277–281.
- Marshall, J. & Wilmoth, G. (1981) *J. Bacteriol.* **147**, 900–913.
- Liebl, U., Mockensturm-Wilson, M., Trost, J. T., Brune, D. C., Blankenship, R. E. & Vermaas, W. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7124–7128.
- Lee, W.-Y., Blankenship, R. E. & Kim, S. H. (1997) *J. Microbiol.* **35**, 206–212.
- Albert, I., Rutherford, A. W., Grav, H., Kellermann, J. & Michel, H. (1998) *Biochemistry* **37**, 9001–9008.
- Nitschke, W., Liebl, U., Matsuura, K. & Kramer, D. M. (1995) *Biochemistry* **34**, 11831–11839.
- Trumpower, B. L. (1990) *Microbiol. Rev.* **54**, 101–129.
- Gray, K. A. & Daldal, F. (1995) in *Anoxygenic Photosynthetic Bacteria*, eds. Blankenship, R. E., Madigan, M. T. & Bauer, C. E. (Kluwer, Dordrecht, the Netherlands), pp. 747–774.
- Liebl, U., Rutherford, A. W. & Nitschke, W. (1990) *FEBS Lett.* **261**, 427–430.
- Kramer, D. M., Schoepp, B., Liebl, U. & Nitschke, W. (1997) *Biochemistry* **36**, 4203–4211.
- Schütz, M., Zirngibl, S., le Coutre, J., Büttner, M., Xie, D.-L., Nelson, N., Deutzmann, R. & Hauska, G. (1994) *Photosynth. Res.* **39**, 163–174.
- Li, L. B., Zou, Y. P., Yu, L. & Yu, C. A. (1991) *Biochim. Biophys. Acta* **1057**, 215–222.
- Preisig, O., Anthamatten, D. & Hennecke, H. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 3309–3313.
- Manna, P. & Vermaas, W. (1997) *Plant Mol. Biol.* **35**, 407–416.
- de Vitry, C., Breyton, C., Pierre, Y. & Popot, J.-L. (1996) *J. Biol. Chem.* **271**, 10667–10671.
- Fujita, Y., Takahashi, Y., Chuganji, M. & Matsubara, H. (1992) *Plant Cell Physiol.* **33**, 81–92.
- Burke, D. H., Alberti, M. & Hearst, J. E. (1993) *J. Bacteriol.* **175**, 2407–2413.
- Burke, D. H., Hearst, J. E. & Sidow, A. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7134–7138.
- Lockhart, P. J., Larkum, A. W. D., Steel, M. A., Waddell, P. J. & Penny, D. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 1930–1934.
- Hudson, A., Carpenter, R., Doyle, S. & Coen, E. S. (1993) *EMBO J.* **12**, 3711–3719.
- Woese, C. R. (1987) *Microbiol. Rev.* **51**, 221–271.
- Gray, M. W. (1989) *Trends Genet.* **5**, 294–299.
- Neefs, J.-M., Van der Peer, Y., De Rijk, P., Chapelle, S. & De Wachter, R. (1993) *Nucleic Acids Res.* **21**, 3025–3049.
- Jensen, P. E., Gibson, L. C. D., Henningsen, K. W. & Hunter, C. N. (1996) *J. Biol. Chem.* **271**, 16662–16667.
- Schubert, W.-D., Klukas, O., Saenger, W., Witt, H. T., Fromme, P. & Krauss, N. (1998) *J. Mol. Biol.* **280**, 297–314.
- Olson, J. M. & Pierson, B. K. (1987) *Int. Rev. Cytol.* **108**, 209–248.
- Mathis, P. (1990) *Biochim. Biophys. Acta* **1018**, 163–167.