J Mol Evol (2005) 60:153–163
DOI: 10.1007/s00239-003-0181-2

JOURNAL OF MOLECULAR EVOLUTION

Phylogenetic Analyses of the Core Antenna Domain: Investigating the Origin of Photosystem I

Lucas J. Mix, David Haig, Colleen M. Cavanaugh

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

Received: 30 December 2003 / Accepted: 29 July 2004 [*Reviewing Editor:* Dr. W. Ford Doolittle]

Abstract. Phototrophy, the conversion of light to biochemical energy, occurs throughout the Bacteria and plants, however, debate continues over how different phototrophic mechanisms and the bacteria that contain them are related. There are two types of phototrophic mechanisms in the Bacteria: reaction center type 1 (RC1) has core and core antenna domains that are parts of a single polypeptide, whereas reaction center type 2 (RC2) is composed of short core proteins without antenna domains. In cyanobacteria, RC2 is associated with separate core antenna proteins that are homologous to the core antenna domains of RC1. We reconstructed evolutionary relationships among phototrophic mechanisms based on a phylogeny of core antenna domains/proteins. Core antenna domains of 46 polypeptides were aligned, including the RC1 core proteins of heliobacteria, green sulfur bacteria, and photosystem I (PSI) of cyanobacteria and plastids, plus core antenna proteins of photosystem II (PSII) from cyanobacteria and plastids. Maximum likelihood, parsimony, and neighbor joining methods all supported a single phylogeny in which PSII core antenna proteins (PsbC, PsbB) arose within the cyanobacteria from duplications of the RC1-associated core antenna domains and accessory antenna proteins (IsiA, PcbA, PcbC) arose from duplications of PsbB. The data indicate an evolutionary history of RC1 in which an initially homodimeric reaction center was vertically transmitted to green sulfur bacteria, heliobacteria, and an ancestor of cyanobacteria. A heterodimeric RC1 (=PSI) then arose within the cyanobacterial lineage. In this scenario, the current diversity of core antenna domains/proteins is explained without a need to invoke horizontal transfer.

Key words: Cyanobacteria — Green sulfur bacteria — Heliobacteria — Phylogeny — Photosystem I — Photosystem II — Reaction center 1 — Reaction center 2

Introduction

The conversion of solar radiation into biochemical reducing power by phototrophic bacteria is by far the greatest contributor to the energy budget of the biosphere. The search for the evolutionary origin of this important process, and the related oxygenic reaction, has spanned the fields of biology, chemistry, physics, and geology for most of the last century (e.g., van Niel 1944). This breadth and the large quantities of data have made it difficult to incorporate all the information available. Here we limit our attention to the evolution of accessory antenna domains. We show that the available data can only be used to support a limited number of hypotheses, and discuss the support for each.

The origin of phototrophy was perhaps the most dramatic event in the history of life on Earth. Enzyme

Correspondence to: Lucas J. Mix; 2450 LeConte Avenue, Room 305, Berkeley, CA 94709, USA; *email:* lmix@oeb.harvard.edu

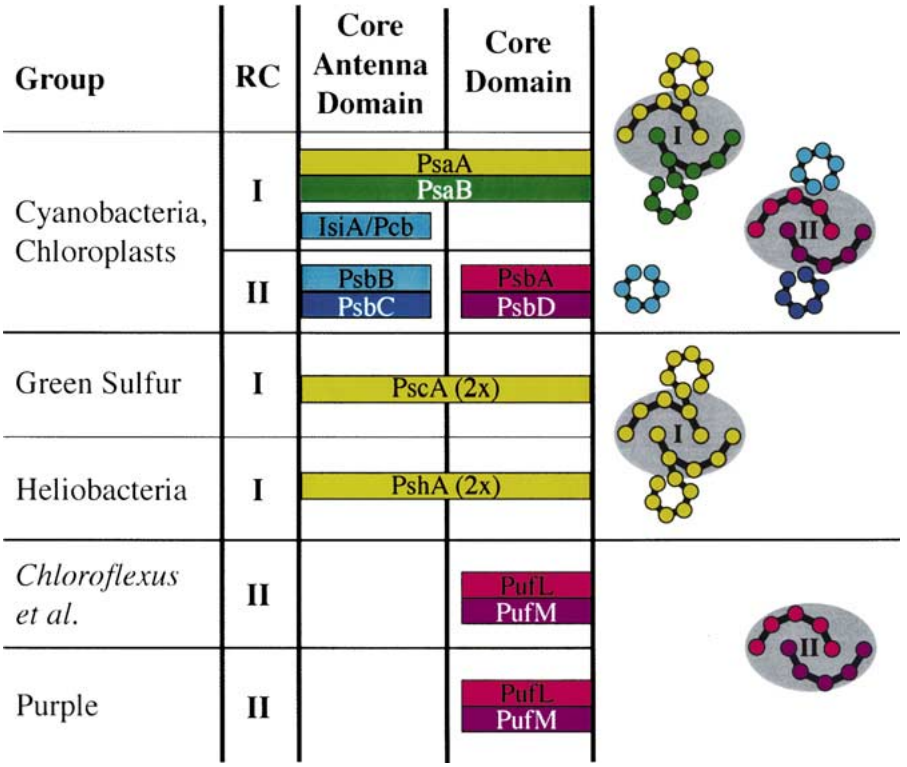

Fig. 1. Homology and structure of core antenna and core domains in phototrophic reaction centers. Each row represents a group of phototrophic bacteria. The second column indicates which domains correspond to RC1 and RC2. Proteins are labeled by gene name and arranged N terminal to C terminal so that the core antenna domain and core domain fall within the appropriate column. Note that all domains in a column are homologous. The final column diagrams the relative locations of α-helices in the photosynthetic membrane. Gray ovals correspond to the core domain and are marked with a I or II to reflect the type of reaction center in which they occur. Different circle colors correspond to the different proteins in the domain columns. The lines emphasize that several helices occur within the same polypeptide; they do not represent a loop. Heliobacteria and green sulfur bacteria have core heterodimers; (2×) after the protein name reflects that a single gene codes for two proteins for each reaction center.

complexes, referred to as reaction centers (RCs), arose within some bacteria, allowing them to convert light into chemical energy. At a RC, photons excite electrons, which are then passed down a redox chain to a final electron acceptor. In the process, reduced compounds are produced that store chemical energy for later use. The simplest forms of phototrophy have insufficient reducing power to strip hydrogen and electrons from water molecules, leaving O_2; hence they are termed anoxygenic. In oxygenic phototrophy, chlorophyll absorbs light at a shorter wavelength, generating enough reducing power to form oxygen (Blankenship 2002, 241–243; Blankenship and Hartman 1998). Oxygenic phototrophy arose in an ancestor of the modern cyanobacteria and made vast amounts of energy available to living processes wherever light and water were present (Des Marais 2000). Oxygenic bacteria caused the Precambrian rise in oxygen levels, producing the modern atmosphere, allowing modern respiration (Dismukes et al. 2001), and forming the ozone for UV protection. The evolutionary history of phototrophy, in particular, the origin of oxygenic photosynthesis, is central to understanding the history and diversity of life.

Phototrophy is present in all three domains of life—Archaea, Bacteria, and Eukarya—yet can be described as primarily a bacterial phenomenon. Bacteriorhodopsins, first observed in the Archaea, act as light-activated proton pumps but do not produce storable chemical energy in the form of reduced compounds (Oesterhelt and Stoeckenius 1971; DasSarma et al. 2001). Homologous proteins have also been observed in γ-proteobacteria (Beja et al. 2000) and fungi (Bieszke et al. 1999). All other forms of phototrophy employ a bacterial type 1 (RC1) or type 2 RC (RC2) (or both). Phototrophic eukaryotes have acquired their RCs by the endosymbiotic incorporation of a cyanobacterium within a eukaryotic cell. Chloroplasts are the descendants of these engulfed bacterial phototrophs (McFadden 2001).

Proteins forming phototrophic RCs in the Bacteria are presented in Fig. 1. RC1s have a long core protein that includes an antenna domain, while RC2s have a short core protein that does not possess an antenna domain. RC2s in cyanobacteria, however, are associated with antenna proteins that perform a function analogous to that of the antenna domain of RC1 core proteins (Schubert et al. 1998). Although structure and function clearly indicate homology of RC1 and RC2 core proteins (Schubert et al. 1998), sequence alignment and the exact nature of evolutionary relationships remain elusive.

Two hypotheses have been invoked to explain the length difference between RC1 and RC2 core proteins. In the first scenario, a short core protein, as occurs in RC2, is ancestral. RC1 would then have been derived by the fusion of this core protein with an antenna domain (Mathis 1990; Xiong et al. 1998). In the second scenario, a long core protein with an antenna domain, as occurs in RC1, is ancestral. RC2 would then be derived by splitting the shorter RC2 core protein and the RC2-associated core antenna proteins from this longer ancestral molecule (Baymann et al. 2001; Mulkidjanian and Junge 1997;

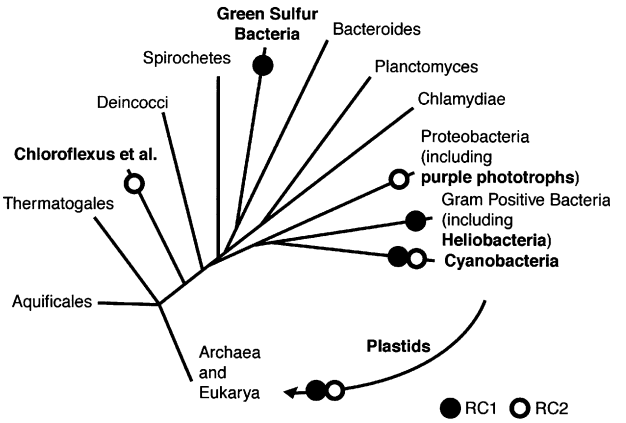

Fig. 2. Phylogenetic tree showing bacterial divisions, with phototrophic groups indicated in bold face. The tree is based on 16S rRNA data (adapted from Pace 1996).

Otsuka et al. 1992). Other authors are convinced that RC1 and RC2 have a common origin but remain agnostic as to which type of reaction center is ancestral (Blankenship 1992; Nitschke and Rutherford 1991). Here we present evidence for a third scenario combining elements of the first two: a short core protein was ancestral to the long core protein of RC1, but the RC2-associated core antenna proteins of cyanobacteria were derived by cleavage from an RC1 core protein.

Five divisions of Bacteria include phototrophic members (Fig. 2). No two phototrophic groups are closely related on 16S rRNA phylogenies, and RC1 and RC2 are interspersed on the tree without any clear pattern (Fig. 2). Oxygenic cyanobacteria possess both types of RC and form a cohesive monophyletic group. The other groups, all anoxygenic, possess either RC1 or RC2 but not both. The heliobacteria are a monophyletic group within the gram-positive bacteria and possess RC1. Green sulfur bacteria also possess RC1 but are placed in their own division. Purple phototrophs are interspersed among nonphototrophic taxa in the proteobacteria and possess RC2. *Chloroflexus* and related species are filamentous anoxygenic phototrophs that possess RC2. This broad, yet sporadic, distribution continues to be a challenge in attempts to locate the origin of phototrophy.

Evolutionary relationships among phototrophic bacteria have long been controversial (reviewed by Baymann et al. 2001). Phylogenies based on molecules that are not directly required for photosynthesis have produced conflicting results. In trees based on 16S rRNA, *Chloroflexus* is on an early branch, followed by green sulfur bacteria, proteobacteria, and then heliobacteria and cyanobacteria (Fig. 2**) (**Pace 1996**)**. By contrast, an analysis of conserved amino acid sequence insertions suggested that heliobacteria were basal, followed by *Chloroflexus*, cyanobacteria, then green sulfur bacteria and proteobacteria (Gupta et al. 1999). A combined data set of 14 proteins produced an entirely different tree, which groups cyanobacteria and proteobacteria, to the exclusion of gram-positive bacteria (Brown et al. 2001). In one recent analysis, orthologs of 188 genes were obtained for a single representative of each phototrophic group (Raymond et al. 2002). From that set, maximum likelihood phylogenies found genes that supported all 15 possible unrooted trees, with no tree supported by more than 15% of the genes. These disagreements may result from the use of different phylogenetic methods, convergent evolution, a lack of phylogenetic signal, and/or horizontal transfer. There seems little prospect that a definitive phylogenetic tree of the major divisions of the Bacteria will be available in the near-future (Doolittle 1999; Raymond et al. 2002). The recent discovery of photosynthetic core genes in a virus presents a means for horizontal transfer (Mann et al. 2003). While relationships between organisms present major problems, real progress can be made toward understanding the evolution of phototrophy by limiting our scope to a single protein or mechanism.

The patchy distribution of phototrophy, and of RC1 and RC2, has raised the possibility of horizontal transfer among bacterial lineages. If this is the case, phototrophic mechanisms may have a different phylogeny from the organisms with which they are currently associated. Unfortunately, phylogenies based on genes directly involved in photosynthetic function have also failed to give consistent results. In a phylogeny based on nine bacteriochlorophyll biosynthesis genes, *Chloroflexus* and green sulfur bacteria formed one clade, cyanobacteria and heliobacteria another, and proteobacteria were basal (Xiong et al. 2000). All nine genes were reported to independently support the same topology or an alternative with the *Chloroflexus* + green sulfur bacteria clade moved onto the branch leading to the proteobacteria. Our attempts to repeat this analysis have resulted in multiple conflicting topologies. Further, a clade that unites *Chloroflexus* and green sulfur bacteria is difficult to reconcile with their distant positions on the 16S rRNA tree and their different RC types. Raymond et al. (2002) found that a set of 36 phototrophy-related genes failed to agree on a favored tree. No clear evolutionary signal stands out and it may be impossible to reconstruct the history of RCs by analyzing molecules not directly involved. Core proteins, and their associated antenna proteins, present a unique opportunity to view the evolution of phototrophy directly. The history of a process must be consistent with the history of the molecules mediating that process.

In this study a molecular phylogeny of the core antenna domains of RC1 and the related core an-

tenna proteins of RC2 is used to shed light on the origin of phototrophy in the Bacteria, addressing three questions. First, does the current genetic nomenclature describe natural categories? In other words, do current gene names correspond to clades? Second, did photosynthesis arise in a common ancestor of all current photosynthetic groups, with its current patchy distribution explained by loss in many lineages, or was photosynthesis acquired by some groups as the result of horizontal transfer? Third, are the recognized homologies between RC1 and RC2 explained by a fragmentation of core and core antenna domains in an ancestor of RC2 or by their fusion in an ancestor of RC1?

Dramatis Personae

Reaction Centers. RC1 and RC2 share many structural features. At the center are 10 transmembrane α-helices forming a handshake motif. This acts as a scaffold for a series of six pigment molecules, either chlorins or bacteriochlorins (Fig. 1). A "special pair" of (bacterio)chlorins acts as a photon trap. Photons excite an electron in the special pair, which then reduces one of the adjacent (bacterio)chlorin molecules, starting a redox chain. The arrangement of pigments is always symmetric or pseudosymmetric, with three on each side. Homodimeric RC1 complexes from heliobacteria and green sulfur bacteria probably have true symmetry, but evidence remains inconclusive (Hauska et al. 2001). The remaining RCs are all heterodimeric, and not perfectly symmetrical. Structural studies on heterodimeric RC1 complexes show a pseudosymmetric arrangement with electron transfer occurring primarily along one branch (Schubert et al. 1998). The conservative nature of the (bacterio)chlorin and α-helix placements suggests a single origin of phototrophy for all Bacteria (Blankenship 1992; Schubert et al. 1998).

The terminology of RCs and associated proteins can be confusing, with different names being given to related proteins in different groups, and with the names of genes sometimes differing from the names of the polypeptides they encode. We use gene names —which are standardized and consistent across species—to refer to both a gene (in italics) and its encoded polypeptide (capitalized and in roman type).

Reaction Centers of Type 2 (RC2). RC2, the simpler type of reaction center, is organized around two core proteins, each with five transmembrane α-helices (Fig. 1). After photon excitation, the special pair initiates a redox chain that involves two additional (bacterio)chlorins and a quinone. The electron is then returned to the special pair (cyclic electron transport) or temporarily stored in a pool of reduced quinones (Blankenship 2002, chap. 7). The crystal structure of RC2 from the purple proteobacterium *Rhodopseudomonas viridis* is the most detailed picture currently available (Deisenhofer et al. 1995). This RC is a heterodimer of two core proteins, PufL and PufM. Homologous proteins have also been isolated from *Chloroflexus aurantiacus* and used to identify *pufL* and *pufM* genes in this species (Shiozawa et al. 1989), but crystallographic analysis has proved difficult (Feick et al. 1996).

The RC2 complex of oxygenic phototrophs is called Photosystem II (PSII). Structural analyses of this complex are available from a cyanobacterium (Zouni et al. 2001) and a chloroplast (Rhee et al. 1998). In addition to its core proteins (PsbA and PsbD), PSII contains core antenna proteins (PsbB and PsbC), each composed of six transmembrane α-helices with paired histidines acting as chlorophyll binding sites (Barber et al. 2000). These chlorophylls trap photons and channel their energy to the reaction center. PsbB and PsbC show homology to the antenna domains of RC1 core proteins (Barber et al. 2000) but appear unrelated to the light-harvesting complexes found in anoxygenic species. Thus RC2 core sequences are represented by PsbA and PsbD (from cyanobacteria and plastids) and PufL and PufM (from phototrophic proteobacteria), while core antenna proteins PsbB and PsbC (from cyanobacteria and plastids) are also present at the reaction center.

Reaction Centers of Type 1 (RC1). The core proteins of RC1 have 11 transmembrane α-helices. These proteins can be subdivided into a core domain that corresponds to the entire core protein of RC2 and a core antenna domain that is homologous to the PsbB/PsbC core antenna proteins of PSII (Fig. 1) (Schubert et al. 1998). The core domain consists of five α-helices from the C-terminal end of the core protein. The remaining six α-helices form the core antenna domain (Schubert et al. 1998). After photon excitation, the special pair initiates a redox chain, involving two additional (bacterio)chlorins and an iron–sufur cluster. The chain can be circular, recycling a single electron, or noncircular, requiring an electron donor and reducing NAD^+ to NADH (Blankenship 2002, chap. 7).

Sequence and structural data are readily available for Photosystem I (PSI), the RC1 in oxygenic phototrophs, but scarce for RC1 in anoxygenic species. The most detailed structure is for PSI of *Synechococcus elongatus* (2.5-Å resolution [Jordan et al. 2001]). In cyanobacteria and plastids, the core heterodimer is composed of two proteins (PsaA and PsaB). In contrast, heliobacteria and green sulfur bacteria have core homodimers (subunits known as PshA and PscA, respectively) (Schubert et al. 1998). Liebl (1993) isolated and sequenced the *pshA* gene from *Heliobacillus mobilis*, noting that the translated sequence had regions

Table 1. Species and gene sequences used in phylogenetic analyses

		RC2		RC1		Core antenna	
Group	Abbrev.	Gene	GenBank	Gene	GenBank	Gene	GenBank
Heliobacteria							
Heliobacillus mobilis	Hm			*pshA*	L19604		
Heliophilum fasciatum	Hf			*pshA*	AY525370		
Green sulfur bacteria							
Chlorobium limicola	Cl			*pscA*	M94675		
Chlorobium tepidum	Ct			*pscA*	AE012952		
Cyanobacteria							
Fischerella muscicola	Fm					*isiA*	AJ295840
						pcbC	AJ296146
Nostoc punctiforme	Np	*psbB*	NC003272	*psaA*	NC003272		
		psbC	”	*psaB*	”		
Prochlorothrix hollandica	Ph	*psbB*	P27200	*psaA*	AY026898	*pcbA*	X97043
		psbC	PHU40144	*psaB*	”		
Prochloron didemni	Pd					*	Z72475
Prochlorococcus marinus	Pm					*pcbA*	U57661
Synechocystis PCC 6803	Sy	*psbB*	M17109	*psaA*	X58825	*isiA*	M23639
		psbC	M21538	*psaB*	”		
Eukaryotes							
Cyanidium caldarium	Cc	*psbB*	AF022186	*psaA*	AF022186		
		psbC	”	*psaB*	”		
Euglena gracilis	Eg	*psbB*	X70810	*psaA*	X70810		
		psbC	”	*psaB*	”		
Lotus japonicus	Lj	*psbB*	AP002983	*psaA*	AP002983		
		psbC	”	*psaB*	”		
Mesostigma viride	Mv	*psbB*	AF166114	*psaA*	AF166114		
		psbC	”	*psaB*	”		
Odontella sinensis	Oi	*psbB*	Z67753	*psaA*	Z67753		
		psbC	”	*psaB*	”		
Oryza sativa	Os	*psbB*	X15901	*psaA*	X15901		
		psbC	”	*psaB*	”		

Note. (*) Light-harvesting protein.

of high amino acid identity with PsaA and PsaB (cyanobacterial RC1 cores) and PsbC (cyanobacterial RC2 core antenna). A core antenna domain sequence is also available from *Heliophilum fasciatum* (Mix et al. 2004). While redox kinetics have been studied, no crystal structures have been published for heliobacteria (Neerken and Amesz 2001). To date, only two *pscA* sequences are available from *Chlorobium limicola* and *C. tepidum*, and these sequences are nearly identical (96% aa identity). The structure of RCs and related proteins in green sulfur bacteria has been reviewed by Hauska et al. (2001). Thus, the known diversity of RC1 core sequences is represented by PsaA and PsaB (from cyanobacteria and plastids), PshA (from heliobacteria), and PscA (from green sulfur bacteria).

Accessory Antennas. Accessory antennas —polypeptides that contain a core antenna domain similar to PsbC—have been found associated with PSI in some cyanobacteria. One such protein, IsiA (iron stress induced A), forms 18-member antenna rings around PSI when access to iron is limited. Prochlorophytes (cyanobacteria containing chlorophyll *b*) living in low-light environments have been found to possess multiple genes for accessory antenna proteins known as *pcbA–pcbG* (Garczarek et al. 2001). Groups of eight pcbs (prochlorophyte chlorophyll binding proteins) have been found associated with PSII (Bibby et al. 2003). Electron microscopy suggests that the antenna arrays allow each reaction center to capture more photons (Bibby et al. 2001; Boekema et al. 2001). Sequence and functional similarities have been noted by several authors (Bibby et al. 2003; LaRoche et al. 1996).

Methods

Sequence Selection and Alignment

Forty-six genes were chosen from GenBank to reflect the broad diversity of core antenna domains (Table 1). These include all known sequences of RC1 core proteins from heliobacteria (PshA) and green sulfur bacteria (PscA). Cyanobacterial and plastid sequences were chosen considering both taxonomic diversity and sequence availability. For most taxa, our analyses included sequences from both the PSI core (*psaA, psaB*) and the PSII core antenna (*psbB, psbC*) genes. Cyanobacteria are represented by sequences from a prochlorophyte (*Prochlorothrix*) as well as from filament-forming (*Nostoc*) and single-celled (*Synechocystis*) genera. Chloroplasts are represented by sequences from a monocot (*Or-*

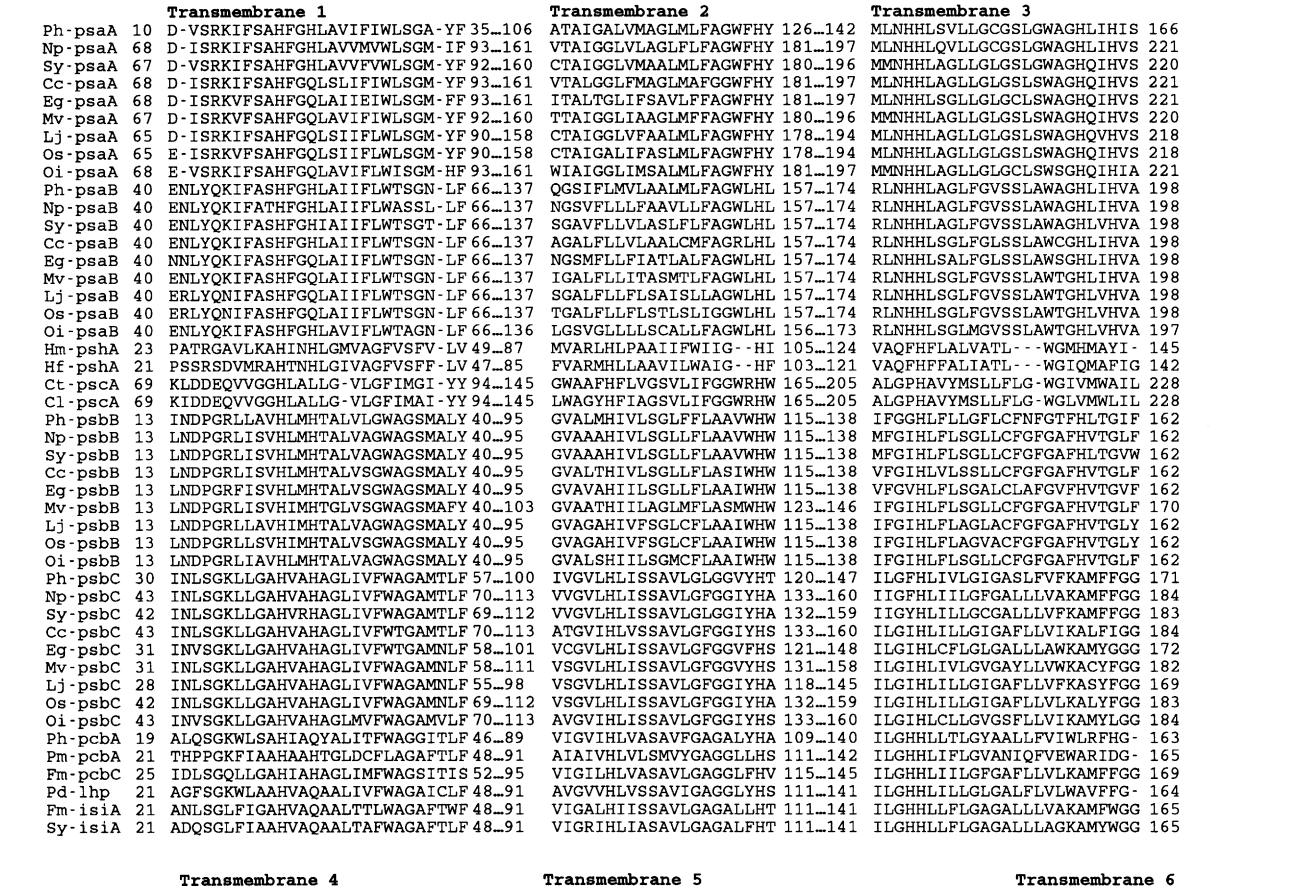

Fig. 3. Aligned regions of core antenna domains from photosynthetic bacteria. Names in the first column include two-letter species abbreviations from Table 1 and the name of the protein.

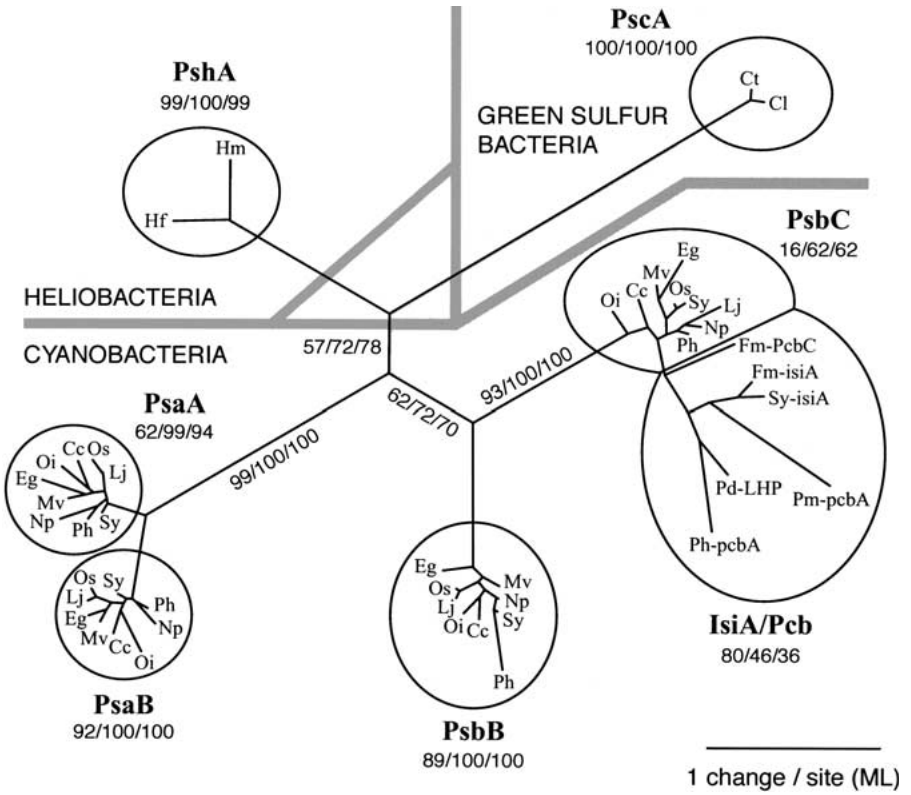

Fig. 4. Phylogenetic tree based on analyses of 159 amino acid residues from core antenna domains. Branch lengths are derived from the maximum likelihood analysis. Numbers indicate bootstrap percentages for 1000 replicates each of maximum likelihood, parsimony, and neighbor joining (ML/P/NJ). Ovals indicate clusters of proteins with the same name and function. Two-letter species abbreviations are listed in Table 1.

yza), a dicot (*Lotus*), a green alga (*Mesostigma*), a rhodophyte (*Cyanidium*), a diatom (*Odontella*), and a euglenid (*Euglena*).

Accessory antenna sequences used include IsiA from *Fischerella* and *Synechocystis*, PcbA from *Prochlorococcus* and *Prochlorothrix*, PcbC from *Fischerella*, and a gene from *Prochloron didemni* identified simply as a "light-harvesting-protein." While many more sequences are available, the sequences in our analysis appear to reflect the full range of diversity of known antenna proteins (Garczarek et al. 2001).

Inferred amino acid sequences were aligned using the BioEdit program (Hall 1999) with CLUSTALW, then adjusted manually. Length differences and divergent evolution have rendered full alignment extremely difficult, however, functional constraints insure that some residues and structures are conserved. Though structural data are only available for a few organisms, all the sequences should contain six transmembrane α-helices with binding sites for pigment molecules. Six hydrophobic regions were identified in each, and while binding sites have only been directly observed for some of the proteins (e.g., Redding et al. 1998), others can be postulated at the location of rare amino acids. Histidine is particularly effective at binding pheophytins, either by coordinating the central Mg^{2+} or by forming hydrogen bonds (Ivancich et al. 1998; Ivancich and Mattioli 1998) and seems to be conserved. The final data set included 159 amino acid sites from those regions of the alignment judged to correspond to the six transmembrane domains acting as a core antenna domain in each protein (Fig. 3).

Phylogenetic Analyses

The data set was analyzed using the PHYLIP package of phylogenetic software (Felsenstein 1989). When sequences are only distantly related, DNA data can contain high levels of noise, as a large number of sites mutate through only four states. Having stronger constraints and more states, amino acid sequences are expected to change more slowly and have a higher signal-to-noise ratio. Consequently, amino acid sequences were used in the analyses. SEQBOOT was used to create 1000 bootstrapped data sets, which were analyzed using PROML, PROTPARS, and PROTDIST/NEIGHBOR with default settings. PROML constructed maximum likelihood (ML) trees using the Jones–Taylor–Thornton (1994) data matrix. PROTPARS constructed minimum-length trees weighting all sites equally. PROTDIST constructed matrices representing the distance between each pair of taxa, using the Jones–Taylor–Thornton data matrix, then NEIGHBOR used a neighbor joining (NJ) algorithm to connect the nearest species to form a tree. CONSENSE was used to make a consensus of all 1000 trees for each method (Felsenstein 1989).

Results

Phylogenies based on ML, parsimony (P), and NJ all produced trees showing the same relationship between protein clusters. Figure 4 presents the ML tree for core antenna domains. The NJ tree had similar relative branch lengths. Bootstrap scores appear as a list of three numbers representing ML, P, and NJ analyses, respectively.

In the likelihood analysis, we assumed a constant rate of evolution across all sites, believing that the data would not support additional variables. To ensure the robustness of this assumption, we repeated the analysis with gamma-distributed rates. That analysis produced the same tree with higher support (data not shown). Nonetheless, we believe the bootstrap values from the constant rate analysis more accurately reflect our confidence.

Clusters of core antenna domains corresponding to the gene names PshA, PscA, PsaA, PsaB, and PsbB all receive strong support as monophyletic groups. Only

PsbB and PsaA failed to achieve 90% confidence from all three methods. The PsbB clade received only 89% ML bootstrap support. Most of the 110 dissenting trees, however, grouped PscA with subsets of PsbB and are most likely an artifact of the unstable position of the long branch leading to PscA. The PsaA clade received only 62% ML bootstrap support. The most common alternative arrangement placed plastidial PsaA sequences from angiosperms (*O. sativa* and *L. japonicus*) at the base of a PsaA/PsaB clade and the sequences from other photosynthetic eukaryotes with cyanobacteria. Since considerable evidence has been presented indicating that plastids have a single origin (Palmer 2003), the basal placement of the two angiosperm sequences is almost certainly an artifact. The next most common alternative arrangement placed *Synechocystis* PCC6803 at the base of the PsaA/PsaB clade (6.2% ML bootstrap support). No other alternative tree received ML bootstrap support above 1%. PsaA monophyly is clearly indicated by high P (99%) and NJ (94%) bootstrap support.

Internal branches, representing older evolutionary events showed mixed support. No method gave strong support for monophyly of either PsbC or IsiA/Pcb, but all methods provide strong support for a group that contains both (93/100/100). Either the accessory antennas arose from within PsbC or vice versa. Analyses found almost unequivocal support (99/100/100) for the branch separating PsaA and PsaB from the remaining core antenna domains.

The PsbB/PsbC/IsiA/Pcb clade receives only moderate bootstrap support (62/72/70). The best-supported alternative hypothesis (26/23/28) groups the PSII core antenna protein PsbB with the core antenna domains of PscA (green sulfur bacteria) and PshA (heliobacteria), but this requires separating the long RC1s on the tree. Such a hypothesis requires either loss and regain of the longer form or two origins, both of which are less parsimonious then the primary hypothesis shown in Fig. 4. Grouping PsbB/PsbC/IsiA/Pcb, all of the proteins containing only a core antenna domain, is both evolutionarily appealing and supported by the data.

Anoxygenic RCs PshA and PscA likewise cluster together in all three analyses (57/72/78). Some support is given for a group including PscA and PsbB (13/1/10), but this once again unparsimoniously separates the long and short proteins. Other alternative hypotheses group the long PSI cores (PsaA and PsaB) with PscA (20/19/9) or PshA (18/8/4), which maintains the long/short division but separates the cyanobacterial species on the tree.

Discussion

Phylogenetic analyses answer the first question presented in the introduction: protein names reflect

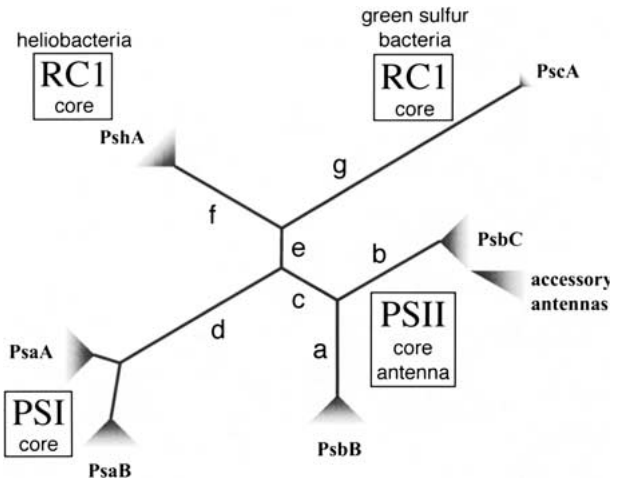

Fig. 5. Diagrammatic, unrooted tree showing the reaction center and group in which each protein cluster occurs. Internal branches are labeled for discussion of rooting possibilities.

natural categories in all cases except PSII core antenna protein PsbC and the accessory antenna proteins IsiA/Pcb. Horizontal transfer need not have occurred from within one of the modern protein clusters, though a *psbC* or *isiA/pcb* gene must have been duplicated at some point within the cyanobacterial lineage. Addressing the second and third questions—involving earlier horizontal transfer of core antenna domains and the relationship between RC1 and RC2—will require a closer look at the tree and where it should be rooted.

A connection between PsbC and IsiA/Pcb has been proposed previously by several authors (LaRoche et al. 1996; van der Staay et al. 1998), based solely on visual similarity. LaRoche et al. (1996) do phylogenetic analyses only within the cluster, while van der Staay et al. (1998) presuppose the relationship by using PsbB as an outgroup. In neither case are relationships with other proteins or the direction of evolution on the tree addressed, so no conclusions could be drawn about monophyly of the groups.

In our analysis, two lines of reasoning support a derivation of the accessory antenna proteins (IsiA/Pcb) from within the PsbC cluster. First, PsbC is an integral part of PSII, present in all cyanobacteria and all plastids, whereas the use of accessory antennas appears to be an optional feature of PSI, found in only a few cyanobacteria (Garczarek et al. 2000). Second, primary and tertiary structural features that are shared by PsbB and PsbC (Barber et al. 2000), but not by the accessory antennas, are easier to explain by divergence of the accessory antennas from PsbC than by convergence of PsbB and PsbC after the latter was derived from an accessory antenna.

Bearing these things in mind, the data can be diagrammed as a tree of core antenna domains spanning RC1, RC2, and accessory antennas (Fig. 5). This tree, however, is nondirectional, as none of the phyloge-

netic methods used provides a root. The analyses, indicating that PshA, PscA, PsaA, PsaB, PsbB, and PsbC (+IsiA/Pcb) are monophyletic groups, implicitly assume that the root lies not within these clusters but somewhere on the internal branches. This assumption seems reasonable; a placement of the root within any of the clusters would necessitate a complicated history of RCs with extremely unequal rates of change in different parts of the tree.

The homodimeric RC is ancestral to the heterodimeric equivalent (Blankenship 2002, p 235), suggesting that the root should not be placed between the two core proteins of PSI (PsaA and PsaB) or between the core antenna proteins of PSII (PsbB and PsbC). The RC1 cores of heliobacteria (PshA) and green sulfur bacteria (PscA) are homodimeric. Other things being equal, heterodimer formation from a homodimer is more reasonable than the reverse. Equally, a heterodimer shows greater specialization, is more likely to have complex interactions with surrounding proteins, and is therefore less likely to be horizontally transferred. An ancestral homodimer is the most elegant solution: all three groups inherited a homodimeric RC1 from a common ancestor or there has been horizontal transfer of a homodimeric RC1 to, or from, the cyanobacterial lineage. In each scenario the branches connecting psaA and psaB do not hold the root of the RC1 tree.

One can imagine a scenario wherein the core antenna domain originated on branch a or b in Fig. 5. This would entail an early free antenna much like the modern accessory antennas, which was co-opted by a RC2 and later joined the core protein of RC1 by gene fusion. We do not find this scenario compelling for two reasons. First, it would require that one branch of the heterodimer evolved much more rapidly than the other. With branches a and b of similar length, and representing genes with, ultimately, the same expression pattern and function, common descent from their intersection seems the simpler explanation. Additionally, the ancestral presence of both PSI and PSII heterodimers in the cyanobacteria makes an origin on branch e, f, or g aesthetically pleasing in that this would allow similar evolution time/rate within the cyanobacteria for psaA, psaB, psbB, and psbC. Second, gene fission is easier to imagine than gene fusion when the physical location of the expressed proteins is so important and consistent. This is not to say that fusion in this case is impossible. At some point in history, the long core protein of RC1 must have been assembled from smaller functional pieces. It seems improbable that such fusion events would have maintained the tertiary structure, while a fission event almost necessitates such a conservation. We cannot definitively rule out origin on branch a or b, however, such a scenario would require additional explanations regarding deviation in evolutionary rates and conservation of tertiary structure.

Placement of the root on branch c would necessitate one of two possible hypotheses: origin of a short RC core protein, followed by fusion, or origin of a long RC core protein with core antenna domain, followed by fission. The former hypothesis faces difficulties similar to those presented for rooting on branches a and b. The latter is very similar in nature to rooting the tree near the intersection of branches c, d, and e as discussed below. Thus rooting on branch c should be considered as a possibility, but not a uniquely compelling one.

Placing the root somewhere on internal branches d–g would imply that a homodimeric RC was ancestral and that its core protein included an antenna domain. Core antenna proteins psbB and psbC would then have originated on internal branch c by fission of the antenna domain from the longer core protein. Note that this in no way requires that the short RC2 protein is the result of such a fission, only the core antenna proteins, psbB and psbC, and the accessory antenna proteins. The relationship between RC1 and RC2 remains an open question.

Rooting on any of branches d–g would be compatible with the origin of RC1 in an ancestor of cyanobacteria and its subsequent transfer (either directly or indirectly) to heliobacteria and green sulfur bacteria. Rooting on any of branches e–g would also be compatible with an RC1 origin outside the cyanobacterial lineage. Possibilities include horizontal transfer(s) and RC1 origin in a common ancestor of green sulfur bacteria, heliobacteria, and cyanobacteria, with RC1 vertically transmitted to all three descendant groups. The origin of a heterodimeric RC1 (=PSI) and the fission of the core antenna domain to produce PsbB and PsbC would then be limited to the cyanobacterial lineage.

Placing the root on branch g and invoking purely vertical transmission of core antenna domains/proteins is most consistent with the data and our knowledge of the molecules. This is the branch that separates RC1 of green sulfur bacteria from RC1 of heliobacteria and cyanobacteria. A root close to the origin of branch g would be near the midpoint of the unrooted tree and would thus be compatible with a rough molecular clock. Although it cannot be determined whether the tree is clock-like, PsaA, PsaB, PsbB, and PsbC are represented in the trees by the same taxa, and these sequences are all roughly equidistant from their presumed common ancestor (if the root is somewhere on branches e–g). Furthermore, the accessory antennas appear on longer branches, consistent with their having fewer interactions with other molecules and being less constrained than the core molecules. A rooting on branch g would also be consistent with the 16S rRNA phylogeny that places green sulfur bacteria basal to heliobacteria and cyanobacteria (Fig. 2). A

root on one of the adjacent branches (e or f) would present similar advantages and affect the interpretation minimally. Branch g has been chosen here, simply because it is longest and consistent with the 16S tree presented in Fig. 2.

The data from core antenna domains are entirely consistent with a scenario wherein no horizontal transfer occurs in the history of RC1. The hypothesis, however, rests on an appeal to the tree presented above, in particular, on branch e separating RC1 core antenna domains from anoxygenic phototrophs and cyanobacteria. This hypothesis might be further defended by clear signature sequences on both sides of the branch. At this time we feel the limited number of representative genes from heliobacteria and green sulfur bacteria make such an endeavor highly speculative. Likewise, when more sequences are available, it is hoped that a clearer, more confident phylogeny will appear. At present, it is sufficient and significant to note that the data are more consistent with the topology in Fig. 5 than any other hypothesis. Alternatives cannot be ruled out. Nonetheless, it is interesting to note that the favored topology is entirely consistent with the 16S rRNA phylogeny, and horizontal transfer is not necessary to explain the current distribution of RC1.

It could be argued that a hypothesis without horizontal transfer in the case of RC1 cannot be truly parsimonious if it necessitates horizontal transfer of RC2. The history of RC2 cannot be addressed on the basis of the data presented here. It may be that the distribution of RC1 which we present makes it more difficult to imagine an RC2 evolution and distribution which would not require invoking horizontal transfer. Even if this were the case, however, we are inclined to believe that horizontal transfer is more common for RC2. One case has been documented (Igarashi et al. 2001) and several more are highly likely (Mix, unpublished). The evolution of RC1 is, here, our sole concern.

The most compelling hypothesis then posits that a homodimeric RC1 was already present in a common ancestor of green sulfur bacteria, heliobacteria, and cyanobacteria. The descendants of this common ancestor include many nonphototrophic members. Therefore, the elimination of horizontal transfer from the history of RC1 requires multiple losses of phototrophy. We suggest that RC1 was lost in most intervening taxa as surviving lineages found new sources of energy (heterotrophy) and as obligatory phototrophs were outcompeted by oxygenic cyanobacteria.

In summary, the following hypotheses arise from our analyses explaining evolutionary relationships among core antenna domains/proteins.

1. The most recent common ancestor of all RC1-containing bacteria had a homodimeric RC that has been retained in heliobacteria and green sulfur bacteria.
2. A segment of the RC1 core protein gene was duplicated early in the history of cyanobacteria, resulting in the production of a core antenna protein that may initially have acted as an additional antenna for PSI or have been directly recruited to its current function (if RC2 core proteins were already present).
3. This core antenna protein duplicated to produce the ancestors of PsbB and PsbC.
4. The RC1 core protein duplicated to produce the ancestors of PsaA and PsaB. This duplication resulted in the current heterodimeric form of the PSI RC.
5. The PsbB/C and PsaA/B duplications were already present in the common ancestor of modern cyanobacteria, as was a heterodimeric RC2 core (composed of psbA and psbD). Relative branch lengths suggest that the duplication that produced the PSII core antenna proteins, PsbB and PsbC, preceded the duplication that produced the PSI core proteins, PsaA and PsaB. This suggests that the pseudosymmetric core of PSII (psbA + psbD) predates the pseudosymmetric core of PSI (PsaA + PsaB).
5. Accessory antenna proteins, IsiA and the Pcb proteins, were derived from PsbC in different lineages of the cyanobacterial radiation.

Acknowledgments. Special thanks go to Michael Madigan, Robert Blankenship, Joseph Felsenstein, Charles Marshall, and John Wakeley for learned advice. This work was supported by NASA/Goddard Space Flight Center through the Graduate Student Researchers Program NGT5-101 (to L.J.M) and NASA Research Grant NAGS-10906 (to C.M.C.)

References

Barber J, Morris E, Buchel C (2000) Revealing the structure of the photosystem II chlorophyll binding proteins, CP43 and CP47. Biochim Biophys Acta Bioenerg 1459:239–247

Baymann D, Brugna M, Muhlenhoff U, Nitschke W (2001) Daddy, where did (PS)I come from? Biochim Biophys Acta Bioenerg 1507:291–310

Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich S, Gates CM, Feldman RA, Spudich JL, Spudich EN, DeLong EF (2000) Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. Science 289:1902–1906

Bibby TS, Nield J, Barber J (2001) Iron deficiency induces the formation of an antenna ring around trimeric photosystem I in cyanobacteria. Nature 412:743–745

Bibby TS, Mary I, Nield J, Barber J (2003) Low-light-adapted *Prochlorococcus* species possess specific antennae for each photosystem. Nature 424:1051–1054

Bieszke JA, Braun EL, Bean LE, Kang S, Natvig DO, Borkovich KA (1999) The *nop-1* gene of *Neurospora crassa* encodes a seven transmembrane helix retinal-binding protein homologous to archaeal rhodopsins. Proc Natl Acad Sci USA 96:8034–8039

Blankenship RE (1992) Origin and early evolution of photosynthesis. Photosynth Res 33:91–111

Blankenship RE (2002) Molecular Mechanisms of Photosynthesis Blackwell Science, Maiden, MA

Blankenship RE, Hartman H (1998) The origin and evolution of oxygenic photosynthesis. Trends Biochem Sci 23:94–97

Boekema EJ, Hifney A, Yakushevska AE, Piotrowski M, Keegstra W, Berry S, Michel KP, Pistorius EK, Kruip J (2001) A giant chlorophyll-protein complex induced by iron deficiency in cyanobacteria. Nature 412:745–748

Brown JR, Douady CJ, Italia MJ, Marshall WE, Stanhope MJ (2001) Universal trees based on large combined protein sequence data sets. Nature Genet 28:281–285

Deisenhofer J, Epp O, Sinning I, Michel H (1995) Crystallographic refinement at 2.3 angstrom resolution and refined model of the photosynthetic reaction center from *Rhodopseudomonas viridis*. J Mol Biol 246:429–457

Des Marais DJ (2000) Evolution—When did photosynthesis emerge on earth? Science 289:1703–1705

Dismukes GC, Klimov VV, Baranov SV, Kozlov YN, DasGupta J, Tyryshkin A (2001) The origin of atmospheric oxygen on Earth: The innovation of oxygenic photosynthesis. Proc Natl Acad Sci USA 98:2170–2175

Doolittle WE (1999) Phylogenetic classification and the universal tree. Science 284:2124–2128

Feick R, Ertlmaier A, Ermler U (1996) Crystallization and X-ray analysis of the reaction center from the thermophilic green bacterium *Chloroflexus aurantiacus*. FEBS Lett 396:161–164

Felsenstein J (1989) PHYLIP—Phylogeny inference package (version 3.2). Cladistics 5:164–166

Garczarek L, Hess WR, Holtzendorff J, van der Staay GWM, Partensky F (2000) Multiplication of antenna genes as a major adaptation to low light in a marine prokaryote. Proc Natl Acad Sci USA 97:4098–4101

Garczarek L, van der Staay GWM, Hess WR, Le Gall F, Partensky F (2001) Expression and phylogeny of the multiple antenna genes of the low-light-adapted strain *Prochlorococcus marinus* SS120 (Oxyphotobacteria). Plant Mol Biol 46:683–693

Gupta RS, Mukhtar T, Singh B (1999) Evolutionary relationships among photosynthetic prokaryotes (*Heliobacterium chlorum, Chloroflexus aurantiacus*. cyanobacteria, *Chlorobium tepidum* and proteobacteria): Implications regarding the origin of photosynthesis. Mol Microbiol 32:893–906

Hall TA (1999) BIOEDIT: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acid Symp Series 41:95–98

Hauska G, Schoedl T, Remigy H, Tsiotis G (2001) The reaction center of green sulfur bacteria. Biochim Biophys Acta Bioenerg 1507:260–277

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

Ivancich A, Mattioli TA (1998) A comparative study of conserved protein interactions of the primary electron donor in photosynthetic purple bacterial reaction centers. Photosynth Res 55:207–215

Ivancich A, Artz K, Williams JC, Allen JP, Mattioli TA (1998) Effects of hydrogen bonds on the redox potential and electronic structure of the bacterial primary electron donor. Biochemistry 37:11812–11820

Jones DT, Taylor WR, Thornton JM (1994) A mutation data matrix for transmembrane proteins. FEES Lett 339:269–275

Jordan P, Fromme P, Witt HT, Klukas O, Saenger W, Krauss N (2001) Three-dimensional structure of cyanobacterial photosystem I at 2.5 angstrom resolution. Nature 411:909–917

LaRoche J, vander Staay GWM, Partensky F, Ducret A, Aebersold R, Li R, Golden SS, Hiller RG, Wrench PM, Larkum AWD, Green BR (1996) Independent evolution of the prochlorophyte and green plant chlorophyll *a*/*b* light-harvesting proteins. Proc Natl Acad Sci USA 93:15244–15248

Liebl U, Mockensturm-Wilson M, Trost JT, Brune DC, Blankenship RE, Vermaas W (1993) Single core polypeptide in the reaction center of the photosynthetic bacterium *Heliobacillus mobilis*—structural implications and relations to other photosystems. Proc Natl Acad Sci USA 90:7124–7128

Mann NH, Cook A, Millard A, Bailey S, Clokie M (2003) Bacterial photosynthesis genes in a virus. Nature 424:741

Mathis P (1990) Compared structure of plant and bacterial photosynthetic reaction centers—evolutionary implications. Biochim Biophys Acta 1018:163–167

Mix LJ, Harmer TL, Cavanaugh CM (2004) Sequence of the core antenna domain from the anoxygenic phototroph *Meliophilum fasciatum*: Implications for diversity of reaction Center type I. Curr Microbiol 48:438–440

McFadden GI (2001) Primary and secondary endosymbiosis and the origin of plastids. J Phycol 37:951–959

Mulkidjanian AY, Junge W (1997) On the origin of photosynthesis as inferred from sequence analysis—A primordial UV-protector as common ancestor of reaction centers and antenna proteins. Photosynth Res 51:27–42

Neerken S, Amesz J (2001) The antenna reaction center complex of heliobacteria: Composition, energy conversion and electron transfer. Biochim Biophys Acta Bioenerg 1507:278–290

Nitschke W, Rutherford AW (1991) Photosynthetic reaction centers—Variations on a common structural theme. Trends Biochem Sci 16:241–245

Otsuka J, Miyachi H, Horimoto K (1992) Structure model of core proteins in photosystem I inferred from the comparison with those in photosystem II and bacteria: An application of principal component analysis to detect the similar regions between distantly related families of proteins. Biochim Biophys Acta 1118:194–210

Pace NR (1996) New perspective on the natural microbial world: Molecular microbial ecology. ASM News 62:463–470

Raymond J, Zhaxybayeva O, Gogarten JP, Gerdes SY, Blankenship RE (2002) Whole genome analysis of photosynthetic prokaryotes. Science 298:1616–1620

Redding K, MacMillan F, Leibl W, Brettel K, Hanley J, Rutherford AW, Breton J, Rochaix JD (1998) A systematic survey of conserved histidines in the core subunits of Photosystem I by site-directed mutagenesis reveals the likely axial ligands of P-700. EMBO J 17:50–60

Rhee KH, Morriss EP, Barber J, Kuhlbrandt W (1998) Three-dimensional structure of the plant photosystem II reaction centre at 8 angstrom resolution. Nature 396:283–286

Schubert WD, Klukas O, Saenger W, Witt HT, Fromme P, Krauss N (1998) A common ancestor for oxygenic and anoxygenic photosynthetic systems: A comparison based on the structural model of photosystem I. J Mol Biol 280:297–314

Shiozawa JA, Lottspeich F, Oesterhelt D, Feick R (1989) The primary structure of the *Chloroflexus aurantiacus* reaction-center polypeptides. Eur J Biochem 180:75–84

van der Staay GWM, Yurkova N, Green BR (1998) The 38 kDa chlorophyll a/b protein of the prokaryote Prochlorothrix hollandica is encoded by a divergent pcb gene. Plant Mol Biol 36:709–716

van Niel CB (1944) The culture, general physiology, morphology, and general classification of the non-sulfur purple and brown bacteria. Bacteriol Rev 8:1–118

Xiong J, Inoue K, Bauer CE (1998) Tracking molecular evolution of photosynthesis by characterization of a major photosynthesis gene cluster from Heliobacillus mobilis. Proc Natl Acad Sci USA 95:14851–14856

Xiong J, Fischer WM, Inoue K, Nakahara M, Bauer CE (2000) Molecular evidence for the early evolution of photosynthesis. Science 289:1724–1730

Zouni A, Witt HT, Kern J, Fromme P, Krauss N, Saenger W, Orth P (2001) Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 angstrom resolution. Nature 409:739–743