

Protein Interactions Limit the Rate of Evolution of Photosynthetic Genes in Cyanobacteria

Tuo Shi,* Thomas S. Bibby,* Lin Jiang,* Andrew J. Irwin,*¹ and Paul G. Falkowski*†

*Environmental Biophysics and Molecular Ecology Program, Institute of Marine and Coastal Sciences, Rutgers University; and †Department of Geological Sciences, Rutgers University

Using a bioinformatic approach, we analyzed the correspondence in genetic distance matrices between all possible pairwise combinations of 82 photosynthetic genes in 10 species of cyanobacteria. Our analysis reveals significant correlations between proteins linked in a conserved gene order and between structurally identified interacting protein scaffolds that coordinate the binding of cofactors involved in photosynthetic electron transport. Analyses of amino acid substitution rates suggest that the tempo of evolution of genes encoding core metabolic processes in the photosynthetic apparatus is highly constrained by protein-protein, protein-lipid, and protein-cofactor interactions (collectively called “protein interactions”). These interactions are critical for energy transduction, primary charge separation, and electron transport and effectively act as an internal selection pressure governing the conservation of clusters of photosynthetic genes in oxygenic prokaryotic photoautotrophs. Consequently, although several proteins within the photosynthetic apparatus are biophysically and physiologically inefficient, selection has not significantly altered the genes encoding these essential proteins over billions of years of evolution. In effect, these core proteins have become “frozen metabolic accidents.”

Introduction

Oxygenic photosynthesis is the most complex energy-transducing process in biology. To function effectively, the structural components of the photosynthetic machinery require coordinated synthesis and assembly of a large number of proteins, which can be broadly clustered into five complexes: light-harvesting antennae, photosystem II (PSII), cytochrome *b₆f*, photosystem I (PSI), and the proton-adenosine triphosphate (ATP) synthase (CF₁-CF₀). These protein complexes form scaffolds that coordinately bind a variety of cofactors including pigments, hemes, iron-sulfur clusters, and metal ions. In addition to structural proteins, the functional photosynthetic apparatus requires enzymes for chlorophyll and carotenoid biosynthesis, CO₂ fixation, and electron transport. Altogether, over 100 genes are devoted to the synthesis and regulation of the photosynthetic apparatus (see Table S1 in the Supplementary Material online).

Although many of the genes required for photosynthetic energy transduction are conserved, the pattern of gene organization (i.e., the genome “landscape”) fundamentally differs among the photosynthetic taxa. In anoxygenic purple bacteria (Naylor et al. 1999) and the gram-positive bacterium, *Heliobacillus mobilis* (Xiong, Inoue, and Bauer 1998), most of the photosynthetic genes are contained in a large, continuous photosynthesis gene cluster (PGC), consisting of tightly linked operons in which expression is cotranscriptionally controlled (Bauer and Bird 1996). In contrast, small clusters of two to three genes that are conserved in linkage appear frequently in green sulfur (Naterstad, Kolsto, and Sirevag 1995) and green nonsulfur bacteria (Xiong et al. 2000). Cyanobacteria, the only prokaryotes that carry out oxygenic photosynthesis, provide an intermediate situation (Wollman, Minai, and

Nechushtai 1999), where genes coding for the light-harvesting phycobilisomes, the carbon-concentrating mechanism, and ATP synthase complex are tightly clustered, while most of the remaining genes of the photosynthetic apparatus are scattered either singly or in small clusters along the genome.

To explore the evolutionary constraints on the organization of the photosynthetic genes in cyanobacteria, we employed a bioinformatic approach (Goh et al. 2000; Pazos and Valencia 2001) in which we measured the correlation between the genetic distance matrices used to build the phylogenetic trees. This approach was used to objectively assess the similarity between the phylogenetic trees of potentially interacting proteins as a metric of coordinated evolution; that is, we seek to understand whether amino acid substitutions between two potentially interacting partners are more correlated than between two noninteracting proteins. Our goals are twofold: (1) to examine the patterns of coevolution between individual proteins that comprise the working units of the photosynthetic apparatus in cyanobacteria and (2) to understand the factors that may have shaped the genome “landscape” in these photosynthetic prokaryotes.

In this paper we report on the analysis of genomes from 10 species of cyanobacteria for conservation in the linkage of photosynthetic genes. We discovered significant correlations in the genetic distance matrices between proteins linked in a conserved gene order, as well as interacting partners that are conformationally important for the stabilization of the cofactors involved in electron transport. These biophysical constraints appear to reduce the tempo of evolution of core photosynthetic genes (e.g., *psbA* encoding D1 protein of PSII), consequently inefficiencies in photosynthetic energy transduction are conserved. In effect, protein interactions in the core photosynthetic machinery have led to “frozen metabolic accidents”; that is, the metabolic functions the proteins mediate are energetically inefficient, but apparently cannot be significantly altered via selection.

¹ Present address: Department of Mathematics and Computer Science, Mount Allison University, New Brunswick, Canada.

Key words: cyanobacteria, photosynthesis, coevolution, bioinformatics, gene order, protein-protein interactions.

E-mail: falko@imcs.rutgers.edu.

Mol. Biol. Evol. 22(11):2179–2189. 2005

doi:10.1093/molbev/msi216

Advance Access publication July 13, 2005

Materials and Methods

Genome Organization of Photosynthetic Bacteria

The organization of the photosynthetic genes was visualized and compared using *Rhodobacter capsulatus*,

Table 1
Photosynthetic Prokaryotic Genomes Used in This Study

Strain	Size (kb)	ORFs	Databases	Reference
<i>Rhodobacter capsulatus</i>	3,700	3,709	IG	
<i>Chlorobium tepidum</i>	2,154	2,253	TIGR	Eisen et al. (2002)
<i>Chloroflexus aurantiacus</i>	~3,000	3,372	JGI/DOE	
<i>Synechocystis</i> sp. PCC 6803	3,573	3,264	CyanoBase	Kaneko et al. (1996)
<i>Anabaena</i> sp. PCC 7120	6,414	6,133	CyanoBase	Kaneko et al. (2001)
<i>Gloeobacter violaceus</i> PCC 7421	4,659	4,430	CyanoBase	Nakamura et al. (2003)
<i>Thermosynechococcus elongatus</i> BP-1	2,594	2,475	CyanoBase	Nakamura et al. (2002)
<i>Synechococcus</i> sp. WH8102	2,434	2,526	JGI/DOE	Palenik et al. (2003)
<i>Prochlorococcus marinus</i> MED4	1,657	1,713	JGI/DOE	Rocap et al. (2003)
<i>P. marinus</i> SS120	1,751	1,884	Roscoff	Dufresne et al. (2003)
<i>P. marinus</i> MIT9313	2,410	2,267	JGI/DOE	Rocap et al. (2003)
<i>Nostoc punctiforme</i>	~9,500	7,281	JGI/DOE	Meeks et al. (2001)
<i>Trichodesmium erythraeum</i> IMS101	~6,500	5,266	JGI/DOE	

NOTE.—The data are from original articles and from the latest variants of the genome sequences at their respective Web sites. The databases are Integrated Genomics Inc. (IG; <http://www.integratedgenomics.com>), The Institute for Genomic Research (TIGR; <http://www.tigr.org>), Department of Energy Joint Genome Institute (JGI/DOE; <http://www.jgi.doe.gov>), CyanoBase (<http://www.kazusa.or.jp/cyanobase/>), and Roscoff (<http://www.sb-roscoff.fr/>).

Chlorobium tepidum, *Chloroflexus aurantiacus*, and *Synechocystis* sp. PCC 6803 as models. Annotated genome sequences for the three anoxygenic photoautotrophs were downloaded from the Integrated Genomics Inc. Web site (<http://www.integratedgenomics.com/>), The Institute for Genomic Research Web site (<http://www.tigr.org/>), and the Department of Energy Joint Genome Institute Web site (http://www.jgi.doe.gov/JGI_microbial/html/index.html), respectively, with *C. aurantiacus* in draft form. Sequences of *Synechocystis* sp. PCC 6803 were downloaded from CyanoBase Web site (<http://www.kazusa.or.jp/cyanobase/>) (see table 1 for a list of all the photosynthetic prokaryotes analyzed in this study).

Sequence Comparisons Within Cyanobacteria

Orthologs of protein sequences were retrieved with Blast (Altschul et al. 1997) using an e value of 10^{-4} as a lower limit cutoff. Genomes of *Nostoc punctiforme* and *Trichodesmium erythraeum* IMS101 were in draft form when this work was undertaken, and preliminary protein-coding sequences were downloaded based on the most recently released contig assemblies. Ortholog candidates between two species are defined primarily by reciprocal top scores of gapped Blast. This relationship is extended to multiple genomes so that every open reading frame (ORF) from an assembled set is each other's best Blast match when comparing these genomes. Protein families related to photosynthesis (see Table S1 in the Supplementary Material online) were then selected from this pool of orthologs. In cases where one species contained more than one homologous sequence of a given protein (paralogous sequences), only one of them was selected. We used a simple criterion to reduce the redundancy, choosing only the sequence more similar to proteins of *Prochlorococcus marinus* MED4, the sequenced cyanobacterium with extremely reduced genome where almost all of the photosynthetic genes are in single copies.

Sequence Alignment and Similarity Matrix

Protein sequences were aligned with ClustalW (Thompson, Higgins, and Gibson 1994) with manual

corrections, followed by selecting unambiguous parts of the alignments and concatenating sequences excluding all gap sites with Gblocks program (Castresana 2000). To prepare a data set that is small but still contains sufficient information to provide enough cases for the analysis, we refined the initial alignments by selecting only sequences that are common to all 10 cyanobacterial species examined. The alignment was then used to generate a distance matrix by percentage sequence divergence with the ClustalW phylogeny program. Eighty-two orthologous protein families (see Table S1 in the Supplementary Material online) related to photosynthesis were analyzed, leading to a final number of 3,321 pairs of matrices for the coevolutionary analysis.

Coevolutionary Analysis

To quantify the coevolution of interacting partners, we calculated the correlation between pairwise matrices each containing the distances among all proteins in a multiple sequence alignment (Goh et al. 2000). Such distance matrices implicitly contain the structure of the phylogenetic trees. The correlation coefficient r (Pearson's correlation coefficient) between two matrices was calculated according to

$$r = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N (X_{ij} - \bar{X})(Y_{ij} - \bar{Y})}{\sqrt{\sum_{i=1}^{N-1} \sum_{j=i+1}^N (X_{ij} - \bar{X})^2} \sqrt{\sum_{i=1}^{N-1} \sum_{j=i+1}^N (Y_{ij} - \bar{Y})^2}}$$

with $-1 \leq r \leq +1$, where N is the number of sequences in the multiple sequence alignments, X_{ij} and Y_{ij} represent equivalent elements of the upper triangle distance matrices X and Y (excluding zeros on the diagonal), and \bar{X} and \bar{Y} are the means of X_{ij} and Y_{ij} , respectively.

Rate of Molecular Evolution

We compared the rate of evolution of different genes within cyanobacteria by taking into account their branch lengths (amino acid substitution rate per site) for two strains of *P. marinus*, MED4 and MIT9313, relative to the out-group *Gloeobacter violaceus* PCC7421 or *Anabaena* sp. PCC7120 in cases where genes were absent

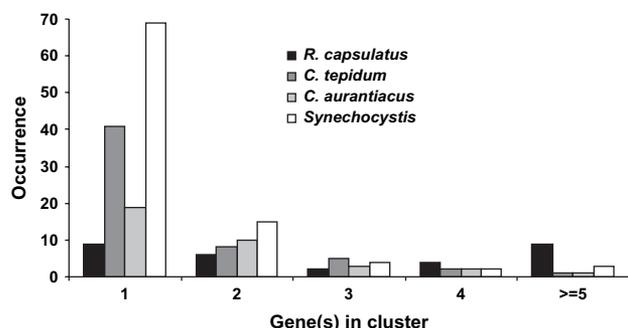


FIG. 1.—Organization of photosynthetic genes in prokaryotic photoautotrophs. Gene clusters are identified based on the synteny and orientation of transcription of genes in an operon. The frequency of occurrence for genes distributed both as singletons and clusters is manually counted.

from the former. In this three-species analysis, we compared 49 proteins (see Table S4 in the Supplementary Material online). The branch lengths of the tree were inferred by the least squares method (Nei and Kumar 2000).

We also performed a relative rate test to compare the divergence of photosynthetic genes between cyanobacterial (K_1) and plastid- or nuclear-encoded (K_2) homologues using RRTree (Robinson-Rechavi and Huchon 2000). K_1/K_2 is the mean divergence among all pairs of sequences within each clade relative to the out-group, *G. violaceus* PCC7421. We estimated the divergence for nine proteins making up the core photosynthetic apparatus (CP43, CP47, D1, D2, PetA, PetB, PetD, PsaA, and PsaB) and three nuclear-encoded proteins PetC, PsbO, and PsbP. The null hypothesis for the test is that the rate of substitution of the tested clade is the same as that of the reference group ($K_1 - K_2 = 0$). The GenBank accession numbers for the investigated sequences are included in Tables S2 and S3 of the Supplementary Material online. To aid the analysis, we introduced a neighbor-joining tree based on concatenated sequences (see Fig. S1 in the Supplementary Material online) for topological weights in each comparison. This phylogenetic weighting scheme improves the accuracy of the test in the case of unbalanced taxonomic sampling (Robinson et al. 1998).

Results

Organization of Photosynthetic Genes in Photosynthetic Prokaryotes

We compared the distribution of clusters of photosynthetic genes in genomes of *R. capsulatus*, *C. tepidum*, *C. aurantiacus*, and *Synechocystis* PCC 6803 and considered a cluster within each species as a group of continuous genes encoding for proteins that form a multimeric protein complex. For example, in *Synechocystis* PCC 6803, *coxB* and *coxAC* are located approximately 906 and 746 kb downstream of the *coxBAC* gene cluster, respectively. In this case, we defined *coxBAC* as a gene cluster, while *coxB* and *coxAC* were omitted from the analysis. The organization of photosynthetic genes is fundamentally different in four groups of photosynthetic bacteria (fig. 1). The purple bacterium *R. capsulatus* has, by far, the most compact PGC, where up to 60% of the photosynthetic genes are

located in clusters of at least five ORFs. In the green sulfur bacterium, *C. tepidum*, and green nonsulfur bacterium, *C. aurantiacus*, most of the photosynthetic genes are organized in small clusters of two to three genes, where synteny of some pigment biosynthesis genes is fully conserved. *Synechocystis* PCC 6803 has three large clusters, which code for the phycobilisome components (*cpcBACCD*), the carbon-concentrating mechanism (*ccmKKLMN*), and ATP synthase complex (*atpIHGFDAC*), respectively. However, more than half (69 out of 136) of the genes are dispersed as singletons. The remaining 37% of the genes are scattered over the entire genome as small clusters of two to four genes.

The divergent clustering pattern of photosynthetic genes appears to have emerged as a result of progressive operon splitting, which may have led to different gene recombinations (table 2). For instance, the cyanobacterial genes for the various subunits of the cytochrome b_6f complex are split between two operons, namely the *petCA* and *petBD*, whereas in *R. capsulatus* these genes are contained in a single operon, *petPRCBA*, in which *petB* is a fused product with *petD* (Widger et al. 1984). Nicotinamide adenine dinucleotide (NADH) dehydrogenase displays a similar pattern of progressive operon splitting. *Rhodobacter capsulatus* contains the longest cluster with 14 genes (18,343 bp) coding for the subunits of bacterial NADH ubiquinone oxidoreductase (*nuo*) (Dupuis et al. 1998), an equivalent of type I NADH dehydrogenase (NDH1), which is encoded by *ndhCKJHAIGEFDB* in *C. tepidum*. In *Synechocystis*, the *ndh* genes are distributed among three separate operons, *ndhCKJ*, *ndhAIGE*, and *ndhFD*.

Despite progressive operon splitting, small clusters of two to four photosynthesis genes appear to be tightly linked in all cyanobacteria (fig. 2). Such conservation of gene order includes, for example, *psaAB*, *psbDC*, and *petCA* and *petBD*. Both the direction and the order in which the genes are transcribed in each pair/cluster are conserved for all clusters. Altogether 37 proteins in the photosynthetic apparatus in cyanobacteria have retained identical pattern of gene organization.

Prediction Power of r for Protein-Protein Interaction

To examine how the pattern of gene organization is related to the evolution, we carried out a correlation analysis to identify protein-protein interactions in the linked photosystems. To distinguish real interactions from potentially “false” positives, we calculated correlation coefficients between subunits within a multimeric energy-transducing or enzyme complex and compared the r values to those between components in the complex and proteins not in that complex. We assume that the distribution of r values derived in this fashion could, on the basis of either experimental data or biological context, distinguish between true interacting proteins in the former case (H_1) and for presumably noninteracting proteins in the latter (H_0). We analyzed each of the nine complexes including the core components of PSI (PsaA and PsaB), PSII (PsbA/D1, PsbB/CP47, PsbC/CP43, and PsbD/D2), cytochrome b_6f (PetA, PetB, PetC, and PetD), and subunits (RbcL and RbcS) of ribulose 1,5 biphosphate carboxylase/oxygenase (Rubisco), ATP synthase

Table 2
Photosynthetic Gene Clusters Subject to Operon Splitting

	<i>Rhodobacter capsulatus</i>	<i>Chloroflexus aurantiacus</i>	<i>Chlorobium tepidum</i>	<i>Synechocystis</i>
ATP synthase	<i>atpDACBE</i>	<i>atpDAC</i>	<i>atpAC; atpBE</i>	<i>atpDAC; atpBE</i>
NADH dehydrogenase	<i>nuoABCDE?F?G?H?I?JKLMN</i>	<i>nuoEFG??: nuoIHJKL;</i> <i>nuoMN</i>	<i>ndhCKJHAIGEFDB</i>	<i>ndhCKJ; ndhAIGE;</i> <i>ndhFD</i>
Cytochrome $b_6/fbcI$ complex	<i>petPRCBA</i>	—	<i>petCB</i>	<i>petCA; petBD</i>
Reaction center	<i>pufQBALMX</i>	<i>pufBAC; pufLM</i>	<i>pscAB</i>	<i>psaAB; psbDC</i>
Chlorophyll biosynthesis	<i>bchCXYZ; bchIDO; bchEJG?P?;</i> <i>bchFNBHLM</i>	<i>bchID; bchJE; bchNB</i>	<i>bchFCX; bchIDH;</i> <i>bchHME; bchNBL</i>	<i>chlLN</i>
Carotenoid biosynthesis	<i>CrtIBK; crtDC; crtEF</i>	—	—	<i>crtDB</i>
Calvin cycle	<i>cbbLSQO; cbbRFPTGAME??Y</i>	<i>cbbIE</i>	—	<i>cbbLXS</i>

NOTE.—The data are from original articles and from the latest variants of the genome annotation at their respective Web sites (table 1). In the case of *C. aurantiacus*, gene cluster information is based on draft genome annotation, which requires confirmation from the final assemblage. Hypothetical proteins with unknown functions and genes not related to photosynthesis are designated by “?”.

(proteins encoded by *atpIHGFDAC* and *atpBE* operons), NADH dehydrogenase (*ndhAIGE*, *ndhFD*, and *ndhCKJ*), cytochrome oxidase (*coxBAC*), Mg-chelatase (*chlDHI*), and protochlorophyllide reductase (*chlNBL*). The results of this comparison reveal clear differences in the average correlation coefficients between subunits within a structurally known multimeric complex (i.e., real interactions, $r > 0.8$) and between proteins that are physically isolated (i.e., false or weak interactions, $r < 0.8$) (fig. 3), suggesting that an r value $> ca. 0.8$ appears to accurately predict real protein-protein interactions based on known protein structures. The representation in figure 4 shows how this approach separates potential interactions.

Conservation of Gene Order As an Indication of Protein Interactions

Inspection of genes linked in a conserved gene order reveals that proteins encoded by these conserved gene pairs/clusters are highly correlated. A detailed map of the correlations is illustrated in figure 5, where extremely high correlation indices were found between proteins comprising the core photosynthetic apparatus, including PsaA-PsaB of PSI, PsbD-PsbC of PSII, PetA-PetC and PetB-PetD of cytochrome b_6/f complex and similarly for enzyme subunits of protochlorophyllide reductase, cytochrome oxidase, and Rubisco. This result implies that conserved gene order tends to predict convergence or coevolution of genes; that is, genes arranged within a cluster are more likely to evolve in concert than nonclustered genes.

Protein-Protein Interactions in Linked Photosystems

Based on the correlation matrices, we predicted possible interactions within each photosynthetic complex (fig. 6). The core of the PSII reaction center and its associated antennae, namely D1, D2, CP47, and CP43, were predicted to be highly interactive with each other. Each of these proteins was predicted to interact with PsbO, one of the required components responsible for oxygen production in the oxygen-evolving complex. In PSI, the reaction center proteins PsaA and PsaB exhibit strongest interaction ($r = 0.94$). Both PsaA and PsaB correlate with the soluble electron carrier ferredoxin (Fd, PetF). In the cytochrome b_6/f complex, the four major polypeptides cytochrome f (PetA), cytochrome b_6 (PetB), subunit IV

(PetD), and Rieske Iron Sulfur Protein (ISP, PetC) interact with each other, however, the correlation between cytochrome b_6 and ISP appears to be relatively weak ($r = 0.67$). All of the subunits comprising the ATP synthase exhibit pairwise correlation value of $r > 0.8$ except the subunit c chain (AtpH).

The analyses also predict interactions at a distance between proteins among complexes, i.e., between different proteins that form part of the linked photosystem even though they may not interact directly. In particular, significant interactions were indicated between the PSI reaction center core proteins, PsaA and PsaB, and PSII antenna proteins, CP47 and CP43, whose pairwise correlation values are between 0.86 and 0.95 (fig. 6). Additionally, the cytochrome b_6/f complex exhibits a high degree of interaction with two complexes that are involved in respiratory electron transport, cytochrome oxidase, and NADH dehydrogenase. The interactions of cytochrome f , b_6 , and subunit IV, with at least one subunit of cytochrome oxidase are predicted. Other predicted interactions include cytochrome f with the major components of NADH dehydrogenase, NdhA, NdhB, NdhC, and NdhD. Both NdhA and NdhB were predicted to interact with cytochrome b_6 , subunit IV, as well as the Rieske ISP.

Mutation Rate As an Evolutionary Constraint

The evolutionary rates of different photosynthetic genes were computed by comparing the branch lengths for *P. marinus* MED4 and MIT9313, when *G. violaceus* PCC7421 was used as the out-group (Fig. 7; Table S4 in the Supplementary Material online). The root-to-tip distances are generally very similar for the two *Prochlorococcus* strains (i.e., the slope of the MED4 vs. MIT9313 trendline is ≈ 1), but vary for different photosynthetic genes. We compared the amino acid substitution rates (1) for proteins constituting the photosynthetic apparatus core, which contain more than one transmembrane helix, (2) for small subunits surrounding the core but still within the thylakoid membrane, and (3) for subunits that do not have any transmembrane helices, i.e., proteins either dissolved in the cytoplasm or located at the stromal or luminal side of the thylakoid membrane. The substitution rates are < 0.35 per site for (1), whereas most (80%) of the proteins in (2) correspond to substitution rates > 0.35 . Most of the



FIG. 2.—Organization of photosynthetic gene clusters in cyanobacteria with complete genomes. Conserved small clusters of two to four photosynthetic genes shown here tend to encode proteins that physically interact with each other. The ATP synthase operons are not shown. Orientations of transcription for each gene are represented by arrowed boxes. Predicted protein-coding regions are colored according to biological role and white boxes indicate hypothetical proteins. Intergenic regions of unspecified lengths are indicated by “//”.

proteins in (3), except PsaC and ATP synthase α and β subunits, exhibit rates >0.4 substitutions per site.

A similar pattern of mutation was also discovered in our extended analysis of the relative divergence between cyanobacterial and plastid- or nuclear-encoded genes, i.e., highly conservative evolution of the functional core (e.g., D1 and D2 proteins), but elevated mutation of the flanking or soluble proteins (e.g., PsbO and PsbP; table 3). Part of the core of cytochrome b_6f complex, the *petCA* operon, exhibits a relatively high mutation frequency. We attribute this largely to the structural domains of cytochrome *f* and ISP at the luminal side (Kurisu et al. 2003), which are subject to fewer constraints from membrane protein-protein interactions. Another feature of the relative rate test is that there is no significant variation between cyanobacterial and plastid- or nuclear-encoded proteins. The hypothesis that K_1 (cyanobacterial located homologues) and K_2 (plastid- or nuclear-encoded homologues) evolve at the same rate cannot be statistically rejected (table 3).

Discussion

In extant photosynthetic prokaryotes, the genomic organization of photosynthetic genes appears to follow a pattern in which large and continuous clusters of genes

are retained in purple photosynthetic bacteria, but the clusters become increasingly fragmented in green photosynthetic bacteria and cyanobacteria (fig. 1). This pattern is arguably consistent with the acquisition of the whole

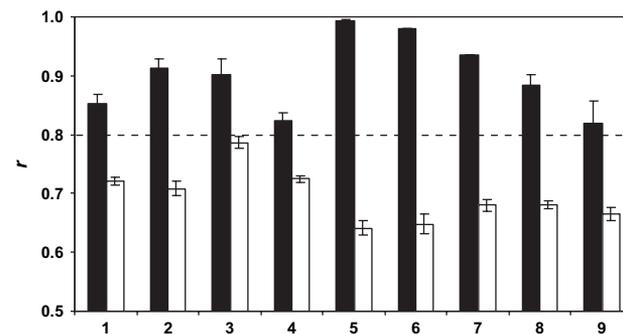


FIG. 3.—Averaged correlation coefficients of real interacting protein pairs (black) and the presumably noninteracting pairs (white) for (1) ATPase, (2) cytochrome oxidase, (3) Mg-chelatase, (4) NADH dehydrogenase, (5) protochlorophyllide reductase, (6) Rubisco, (7) PSI, (8) PSII, and (9) cytochrome b_6f . The correlations between subunits within each multimeric enzymatic complex were interpreted as real protein interactions, whereas potentially noninteracting correlation was calculated between an enzymatic subunit and proteins not in that complex. Similar analysis was also extended to the photosystems where only the core components were considered as real interactions. Error bars denote the standard error of the *r* values and the dotted line indicates the >0.8 cutoff threshold.

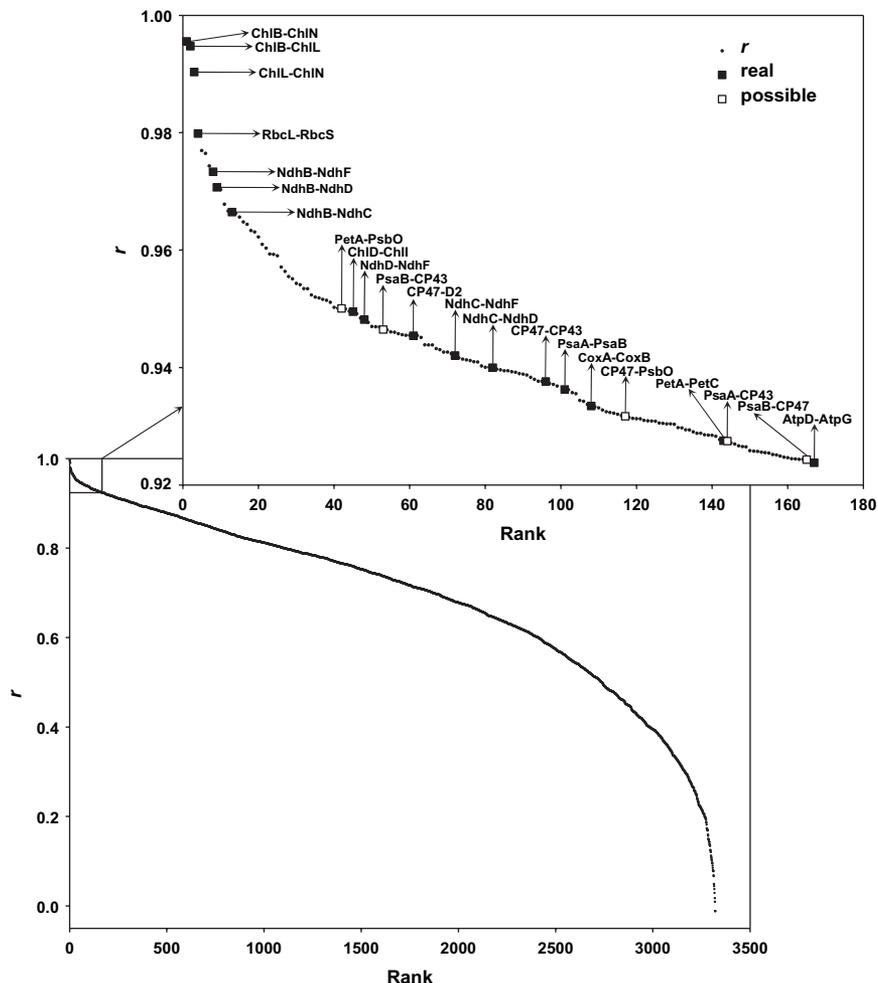


FIG. 4.—Distribution of the correlation values of r corresponding to 3,321 pairs of protein distance matrices. The top 5% (167 pairs) of the correlations are highlighted. Pairs representing “real” interactions—the two structural domains/subunits of the same multimeric protein complex—are marked with closed squares and possible ones with open squares. Representative pairs are labeled with the name of the corresponding proteins.

PGC by cyanobacteria from ancestral analogues of purple photosynthetic bacteria, followed by operon splitting and gene rearrangement. Environmental selection potentially allows genes contained within small operons more flexibility in expression (i.e., greater physiological plasticity) resulting from independent regulation of transcription. However, because operon splitting has not progressed to the fullest extent possible (i.e., clusters of photosynthetic genes are retained in all extant cyanobacteria), there must be strong selective pressure to maintain a conserved genomic organization in which structurally and/or functionally interacting photosynthetic genes are grouped and thus reduce the chance of genetic recombination perturbing coadapted pairs of genes.

Although not all biophysical interactions between proteins can be revealed by the conservation of gene order, the idea that biophysical interactions between encoded proteins is one of the reasons for evolutionary conservation of gene order is well established (Casjens and Hendrix 1974; Botstein 1980; Campbell 1994) and has subsequently been extended by analyses of nonphotosynthetic bacterial and archaeal species (Dandekar et al. 1998). The coupling of gene order with interactions between some protein pairs

in chlorophyll biosynthesis has also been experimentally confirmed (Trumpower 1990; Papenbrock et al. 1997; Grafe et al. 1999; Fujita and Bauer 2000; Xiong et al. 2000). Our systematic analysis of 10 cyanobacterial genomes supports the hypothesis that the conservation of gene order (fig. 2) is generally highly correlated with the evolution and coadaptation of interacting proteins (fig. 6). Such a correlation is observed in proteins that comprise the core photosynthetic energy-transducing machinery as well as those involved in chlorophyll biosynthesis, carbon fixation, and ATP synthesis. Paradoxically, while the α (PsbE) and β (PsbF) subunits of Cyt b_{559} are encoded by genes involved in the conserved operon *psbEFLJ*, their correlation based on the statistical method employed is relatively low ($r = 0.61$). This could result from the bias in composition among sites due to short sequences (none of the PsbF, L, and J contains >50 amino acids, and thus insufficient information to infer meaningful correlation). On the other hand, it also implies that Cyt b_{559} might be less critical to the photochemical function of PSII. Indeed, while PSII most probably derived from an ancestral type II reaction center of purple bacteria (Schubert et al. 1998), homologues of Cyt b_{559} are absent from purple bacteria.

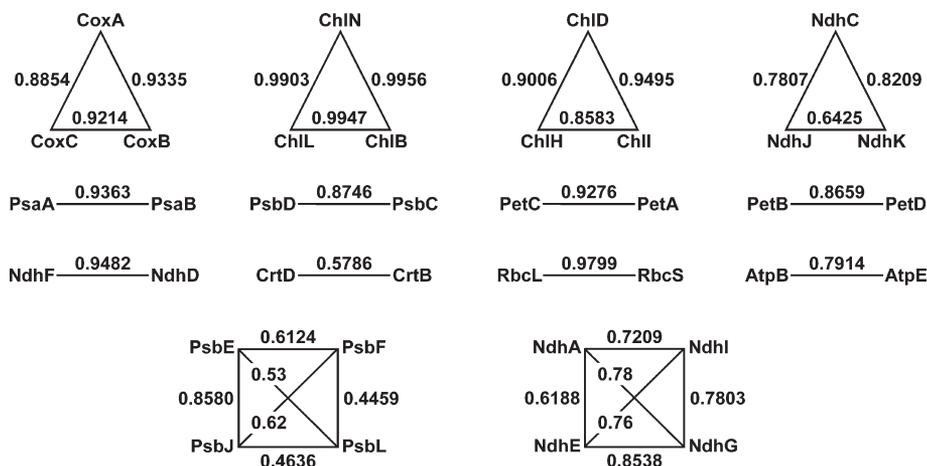


FIG. 5.—Numerical correlation of proteins linked in conserved gene order. Values stand for the correlation coefficient between the corresponding pair of proteins that are linked in a gene cluster. Pairs of proteins corresponding to high correlation values, which can be used to infer protein-protein interactions, make up either subunits of a single enzyme or components of the structural complex of the photosynthetic apparatus. Subunits of magnesium protoporphyrin IX chelatase, encoded by *chID*, *chlH*, and *chlI*, which do not cluster in cyanobacteria as they do in anoxyphotobacteria (Xiong et al. 2000) also exhibit high degree of correlation.

Hence, Cyt *b*₅₅₉, with poorly understood functions (Barber and Rivas 1993), appears to have evolved only in oxygenic photoautotrophs and flanks the core photosynthetic reaction center as supplemental “add ons” (Zouni et al. 2001; Ferreira et al. 2004) and as such may be less evolutionarily constrained. Furthermore, the cistronic nature of the *psbEFLJ* operon may possibly drive its conservation for coordinated transcription of the genes and subunit stoichiometry, as only a single copy of each subunit is present per PSII reaction center (Zouni et al. 2001; Ferreira et al. 2004).

Most of our understanding of protein interactions in photosynthesis comes from structural features indicative

of protein interaction partners, yet our bioinformatic approach provides a powerful tool in predicting interactions beyond physical contact (Jordan et al. 2001; Kurisu et al. 2003; Ferreira et al. 2004). Our analysis predicts strong interactions between the PSI apoproteins PsaA and PsaB with PSII antenna proteins CP47 and CP43 (figs. 4 and 6). This observation suggests coevolution and a functional link between the two photosystems ever since their derivation from a common ancestor (Vermaas 1994; Schubert et al. 1998; Xiong, Inoue, and Bauer 1998). We propose, therefore, that although the two photosystems are physically separated within the thylakoid membrane, they must

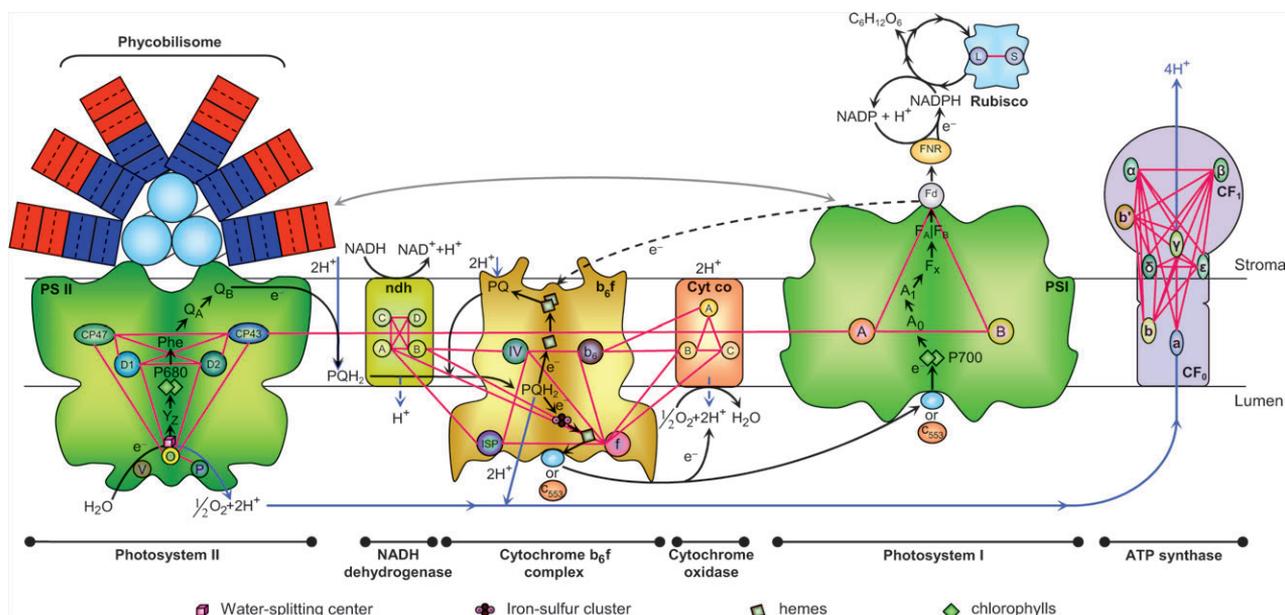


FIG. 6.—Protein-protein interactions in linked photosystems revealed by the coevolutionary analysis. Red lines represent predicted interactions with coefficient values better than 0.8. Also shown is a network of protein-protein interactions in the ATP synthase complex. The pattern of protein-protein interactions suggests coevolution of photosynthetic genes driven by electron transport and redox state of the primary photochemistry. The figure was modified from Bryant (1994) and Kühlbrandt (2003). Black arrows, electron transfer; blue arrows, proton transfer; and gray arrows, proposed lateral diffusion of phycobilisomes between PSII and PSI.

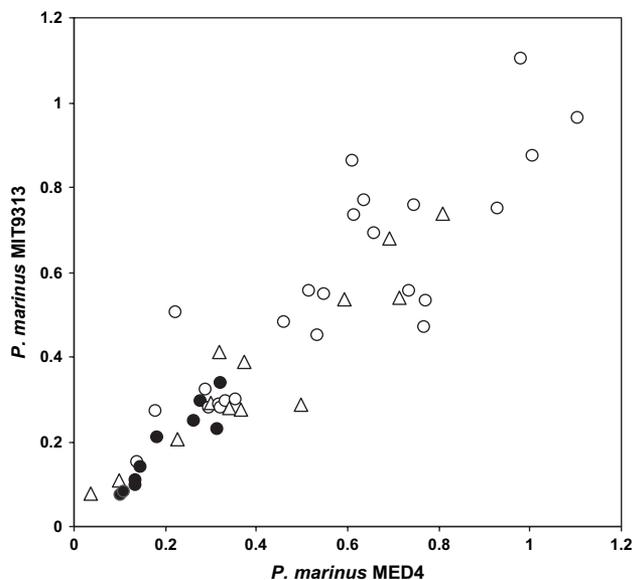


FIG. 7.—Root-to-tip branch lengths of orthologous proteins for two close strains of *Prochlorococcus marinus* MED4 and MIT9313 when *Gloeobacter violaceus* PCC7421/*Anabaena* sp. PCC 7120 was used as the out-group. Closed circles indicate proteins with more than one transmembrane helix that form the functional physical core of the photosynthetic apparatus, whereas open ones are flanking or peripheral proteins surrounding the core but still within the membrane. Triangles represent proteins that are located on the stromal or luminal sides of the membranes or soluble proteins in the cytoplasm. The abscissa and the ordinate represent the number of amino acid replacement per site. Note that proteins constituting the core photosynthetic apparatus are highly conservative in their substitution rate.

be functionally linked in order to prevent their divergent evolution. What is the physical link? One possibility is a mobile pool of light-harvesting complexes. The two photosystems share light-harvesting antenna systems (i.e., the water-soluble phycobilisomes of cyanobacteria or Pcb proteins of prochlorophytes) (Rijgersberg and Ames 1980; Mullineaux 1994; Bibby et al. 2003), a fraction of which physically migrate between the photosystems during state transitions (Mullineaux, Tobin, and Jones 1997). Both photosystems therefore must maintain a conserved docking site for the light-harvesting antenna such that excitation energy can be efficiently transferred to the reaction center.

These docking sites are probably as ancient as the reaction centers themselves, and we propose that the physically separated protein pairs in the two photosystems (PsaA, PsaB of PSI and CP43, CP47 of PSII) have high r values because they possess similar, conserved docking sites for the mobile antenna system.

A putative cytochrome b origin of type II reaction centers has recently been proposed to be an evolutionary link between photosynthesis and respiration (Xiong and Bauer 2002). We predicted protein-protein interactions of cytochrome b_{6f} with respiratory components (fig. 6), including NADH dehydrogenase and cytochrome oxidase, providing additional lines of evidence to support this hypothesis. In fact, the cytochrome b_{6f} and the respiratory cytochrome bc_1 complexes share the basic elements involved in the electron transfer and proton translocation from NADH to the quinone pool (Rich 1984; Trumppower 1990; Kramer and Crofts 1993), which can be subsequently oxidized by cytochrome oxidase. For example, cytochrome b_6 and subunit IV are homologous to the N- and C-terminal halves of cytochrome b of the bc_1 complex (Widger et al. 1984), as is the ISP between the two complexes (Carrell et al. 1997).

It is apparent that highly correlated proteins are more likely to be physically adjacent to each other or structurally suited for optimal conformation in coordinating the cofactors involved in photosynthetic electron transport (fig. 6). These interactions involve the functions of genes expected to be encoded in all oxygenic photoautotrophs. Indeed, the plastids in all photosynthetic eukaryotes are derived from a single cyanobacterial ancestor through endosymbiotic events (Wolfe et al. 1994; Bhattacharya and Medlin 1998; Delwiche 1999). Despite the considerable variability in the number of genes retained in various plastid genomes (Martin et al. 1998; Grzebyk et al. 2003), a “core” set of genes coding for photosynthetic reaction center proteins, electron transport components, Rubisco, as well as the ribosomal machinery regulating their rapid synthesis is retained in all plastids (Race, Herrmann, and Martin 1999; Grzebyk et al. 2003). Such conservation greatly improves the ability to extrapolate specific biophysical, structural and biochemical processes that are encoded in the chloroplast genome (Falkowski and Raven 1997).

Table 3
Relative Divergence of Cyanobacterial, Plastid-, and Nuclear-Encoded Genes

Protein	Group 1	Group 2	K_1	K_2	P
D1	Cyanobacteria	Plastids	0.2034	0.2198	0.274
D2	Cyanobacteria	Plastids	0.1986	0.2136	0.232
CP43	Cyanobacteria	Plastids	0.3118	0.3284	0.229
CP47	Cyanobacteria	Plastids	0.4249	0.4429	0.312
PetA	Cyanobacteria	Plastids	0.6641	0.7107	0.203
PetB	Cyanobacteria	Plastids	0.1891	0.2073	0.337
PetD	Cyanobacteria	Plastids	0.3519	0.3918	0.183
PsaA	Cyanobacteria	Plastids	0.2962	0.2760	0.025
PsaB	Cyanobacteria	Plastids	0.2621	0.2664	0.672
PetC	Cyanobacteria	Nucleus	0.6968	0.7594	0.291
PsbO	Cyanobacteria	Nucleus	1.2314	1.2369	0.956
PsbP	Cyanobacteria	Nucleus	1.4559	1.3160	0.609

NOTE.—Sequences from cyanobacteria are designated in Table S1 of the Supplementary Material online. The GenBank accession numbers for plastid- and nuclear-encoded proteins are included in Tables S2 and S3 of the Supplementary Material online. K_1/K_2 indicates mean amino acid substitution rate for each internal group relative to the out-group *Gloeobacter violaceus* PCC7421.

Moreover, proteins whose genes are most resistant to transfer to the nucleus constitute the functional physical core of the photosynthetic apparatus (Race, Herrmann, and Martin 1999). These core proteins, including D1 and D2 of PSII, Cytochrome b_6 and subunit IV, and PsaA and PsaB of PSI, form scaffolds of the photosynthetic apparatus via their membrane intrinsic structural domains. Even a tiny variation in the conformation of their transmembrane α helices would have fatally affected the core proteins in binding the prosthetic groups and pigments involved in electron transfer. Hence mutations of the α helices of the core photosynthetic proteins are strongly selected against. This hypothesis is supported by our evolutionary rate analysis, which reveals that the core photosynthetic genes are the most conserved (fig. 7). Furthermore, the rates of amino acid substitution are basically very similar for prokaryotic, organellar, and nuclear-encoded photosynthetic genes (table 3). This finding contrasts with the enhanced mutation rate of functional genes that have been reported previously in endosymbiotic bacteria or organelles (Moran 1996; Lynch 1997; Lambert and Moran 1998), which is often explained by Muller's ratchet effect, whereby mildly deleterious mutations tend to accumulate at random in small populations. However, considering a metabolic pathway that still maintains its original function after hundreds of millions of years of evolution, the ratchet hypothesis cannot explain the extremely slow mutation rates of core photosynthetic genes. If deleterious mutations accumulate continuously in a gene, the gene will eventually deteriorate and lose its original function (Pamilo, Nei, and Li 1987; Kondrashov 1995; Rispé and Moran 2000). Given the fact that no acceleration of substitution rate was observed for photosynthetic genes, we propose that strong purifying selection may have offset the effect of deleterious substitution in these genes. In fact, consistent with our analysis of photosynthetic genes, the substitution rate in genes coding for key proteins of respiration and ATP synthesis in a protist mitochondrion is also highly attenuated (Itoh, Martin, and Nei 2002). The retention of a core set of proteins that are extremely conservative in mutation of their transmembrane structural domains strongly suggests purifying selective constraints imposed by the protein-lipid, protein-cofactor, as well as protein-protein interactions, which are critical to energy transduction.

The core processes in the photosynthetic apparatus evolved under anaerobic conditions, yet were appropriated for oxygenic photosynthesis with relatively minor modifications (Blankenship and Hartman 1998; Rutherford and Faller 2003). The structural similarities between the two reaction centers in cyanobacteria and their evolutionary homologues in anoxygenic bacteria reveal that the electron transfer scheme within the complex remains fundamentally unchanged. However, the Mn-binding protein, encoded by *psbO*, and other peripheral or flanking proteins (e.g., PsbE and F) are critical for oxygenic photosynthesis. While these supplemental "add ons" to a bacterial type II reaction may have led to a functional water splitting machine, it came at a cost. In all oxygenic photoautotrophs, D1 turns over approximately once every 30 min in the light (Prasil, Adir, and Ohad 1992). A large fraction of this turnover is a consequence of attack on the core protein by reactive oxygen species (ROS), which is produced by the reaction center

itself. The process leads to light-dependent inactivation of a fraction of reaction centers (i.e., a type of photoinhibition), which is not observed in anaerobic photoautotrophs or in oxygenic photoautotrophs in which electron transport is blocked with inhibitors. Alterations to the fundamental structure of the D1 protein do not appear to be tolerated (Diner and Rappaport 2002), hence selection pressure has been directed on the gene's promoter. The result is that D1 turnover is compensated by an elaborate "repair" cycle in which the damaged protein is removed from the reaction center and replaced with a protein synthesized de novo (Melis 1999). This repair cycle requires a large investment in energy and a commitment of genes specifically designed for this purpose. We propose that the conservation of D1 in the face of such a huge selection pressure is a consequence of protein-protein, protein-cofactor, and protein-substrate interactions that are critical for energy transduction, primary charge separation, and electron transport. This interpretation of the retention of a seemingly inefficient core metabolic process in such a fundamentally critical energy-transducing pathway is supported by our analysis of the preserved gene clustering and reduced evolutionary tempo of the core photosynthetic gene products. In effect, the reaction center of PSII has become a "frozen metabolic accident."

Another example of a "frozen metabolic accident" is Rubisco. The enzyme, derived from a methionine-scavenging pathway (Ashida et al. 2003), is relatively efficient in fixing CO_2 at high partial pressures of the gas and when O_2 concentrations are low. However, over the past billion years, as CO_2 levels declined and O_2 levels rose, Rubisco became increasingly inefficient, using O_2 as an alternative substrate while maintaining a relatively low affinity for CO_2 . Again, to compensate for these inefficiencies, two novel solutions emerged. One was the evolution of carbon-concentrating mechanisms, which in cyanobacteria, strongly increases the intracellular partial pressure of CO_2 , thereby providing Rubisco with a near-saturating concentration of substrate regardless of the external concentration (Kaplan et al. 2001). The second solution was to increase the production of Rubisco; indeed, between 20% and 40% of the total soluble protein in a cyanobacterium is dedicated to this single enzyme. The constraints on evolution, imposed again by protein-protein and protein-substrate interactions of this multimeric enzyme, have imposed such inefficiencies in the photosynthetic machinery that in higher C3 plants, 40% of the photosynthetic energy is lost to photorespiratory O_2 consumption in Rubisco.

Not all "frozen metabolic accidents" are inefficient. Under normal physiological conditions, PSI and the cytochrome b_6f complex operate with little loss of energy or need for replacement of damaged machinery due to attack from ROS or other radicals. However, the damage endured by PSII and the lack of catalytic selectivity and affinity by Rubisco reduce the overall energy transduction efficiency of the photosynthetic apparatus by at least 50% over what might be achievable were these two components selected without their present "flaws." These inefficiencies not only constrain the rate of growth of cyanobacteria, but also appear to have retarded the evolutionary tempo of key proteins since the origins of oxygenic photosynthesis over approximately 2.8 billion years ago.

Supplementary Material

Supplementary Tables S1–S4 and Supplementary Figures S1 and S2 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank Quansong Tong for programing assistance; Jason Raymond, Daniel Grzebyk, and Colombar de Vargas for stimulating discussions; and Yael Helman and Diana Nemergut for critical reading. We also thank Peter Lockhart and two anonymous reviewers for constructive comments. This work is supported by the National Science Foundation through the Biocomplexity Program under grant OCE-0084032.

Literature Cited

- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- Ashida, H., Y. Saito, C. Kojima, K. Kobayashi, N. Ogasawara, and A. Yokota. 2003. A functional link between RuBisCO-like protein of *Bacillus* and photosynthetic RuBisCO. *Science* **302**:286–290.
- Barber, J., and J. Rivas. 1993. A functional model for the role of cytochrome b559 in the protection against donor and acceptor side photoinhibition. *Proc. Natl. Acad. Sci. USA* **90**:10942–10946.
- Bauer, C. E., and T. H. Bird. 1996. Regulatory Circuits Controlling Photosynthesis Gene Expression. *Cell* **85**:5–8.
- Bhattacharya, D., and L. Medlin. 1998. Algal phylogeny and the origin of land plants. *Plant Physiol.* **116**:9–15.
- Bibby, T. S., I. Mary, J. Nield, F. Partensky, and J. Barber. 2003. Low-light-adapted *Prochlorococcus* species possess specific antennae for each photosystem. *Nature* **424**:1051–1054.
- Blankenship, R. E., and H. Hartman. 1998. The origin and evolution of oxygenic photosynthesis. *Trends Biochem. Sci.* **23**:94–97.
- Botstein, D. 1980. A theory of modular evolution for bacteriophages. *Ann. NY Acad. Sci.* **354**:484–491.
- Bryant, D. A. 1994. *The molecular biology of cyanobacteria*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Campbell, A. 1994. Comparative molecular biology of lambdaoid phages. *Annu. Rev. Microbiol.* **48**:193–222.
- Carrell, C. J., H. Zhang, W. A. Cramer, and J. L. Smith. 1997. Biological identity and diversity in photosynthesis and respiration: structure of the lumen-side domain of the chloroplast Rieske protein. *Structure* **5**:1613–1625.
- Casjens, S. R., and R. Hendrix. 1974. Comments on the arrangement of the morphogenetic genes of bacteriophage lambda. *J. Mol. Biol.* **90**:20–25.
- Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**:540–552.
- Dandekar, T., B. Snel, M. Huynen, and P. Bork. 1998. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* **23**:324–328.
- Delwiche, C. F. 1999. Tracing the thread of plastid diversity through the tapestry of life. *Am. Nat.* **154**:S164–S177.
- Diner, B. A., and F. Rappaport. 2002. Structure, dynamics, and energetics of the primary photochemistry of photosystem II of oxygenic photosynthesis. *Annu. Rev. Plant Biol.* **53**:551–580.
- Dufresne, A., M. Salanoubat, F. Partensky et al. (17 co-authors). 2003. Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc. Natl. Acad. Sci. USA* **100**:10020–10025.
- Dupuis, A., M. Chevallet, E. Darrouzet, H. Duborjal, J. Lunardi, and J. P. Issartel. 1998. The Complex I from *Rhodobacter capsulatus*. *Biochim. Biophys. Acta* **1364**:147–165.
- Eisen, J. A., K. E. Nelson, I. T. Paulsen et al. (32 co-authors). 2002. The complete genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic, green-sulfur bacterium. *Proc. Natl. Acad. Sci. USA* **99**:9509–9514.
- Falkowski, P. G., and J. A. Raven. 1997. *Aquatic Photosynthesis*. Blackwell Science, Oxford.
- Ferreira, K. N., T. M. Iverson, K. Maghlaoui, J. Barber, and S. Iwata. 2004. Architecture of the photosynthetic oxygen-evolving center. *Science* **303**:1831–1838.
- Fujita, Y., and C. E. Bauer. 2000. Reconstitution of light-independent protochlorophyllide reductase from purified Bchl and BchN-BchB subunits. In vitro confirmation of nitrogenase-like features of a bacteriochlorophyll biosynthesis enzyme. *J. Biol. Chem.* **275**:23583–23588.
- Goh, C.-S., A. A. Bogan, M. Joachimiak, D. Walther, and F. E. Cohen. 2000. Co-evolution of proteins with their interaction partners. *J. Mol. Biol.* **299**:283–293.
- Grafe, S., H.-P. Saluz, B. Grimm, and F. Hanel. 1999. Mg-chelatase of tobacco: the role of the subunit CHL D in the chelation step of protoporphyrin IX. *Proc. Natl. Acad. Sci. USA* **96**:1941–1946.
- Grzebyk, D., O. Schofield, C. Vetriani, and P. G. Falkowski. 2003. The mesozoic radiation of eukaryotic algae: the portable plastid hypothesis. *J. Phycol.* **39**:259–267.
- Itoh, T., W. Martin, and M. Nei. 2002. Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc. Natl. Acad. Sci. USA* **99**:12944–12948.
- Jordan, P., P. Fromme, H. T. Witt, O. Klukas, W. Saenger, and N. Krauß. 2001. Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* **411**:909–917.
- Kaneko, T., Y. Nakamura, C. P. Wolk et al. (19 co-authors). 2001. Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res.* **8**:205–213.
- Kaneko, T., S. Sato, H. Kotani et al. (21 co-authors). 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**:109–136.
- Kaplan, A., Y. Helman, D. Tchernov, and L. Reinhold. 2001. Acclimation of photosynthetic microorganisms to changing ambient CO₂ concentration. *Proc. Natl. Acad. Sci. USA* **98**:4817–4818.
- Kondrashov, A. S. 1995. Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? *J. Theor. Biol.* **175**:583–594.
- Kramer, D. M., and A. R. Crofts. 1993. The concerted reduction of the high and low potential chains of the bf complex by plastoquinol. *Biochim. Biophys. Acta* **1183**:72–84.
- Kühlbrandt, W. 2003. Dual approach to a light problem. *Nature* **426**:399–400.
- Kurusu, G., H. Zhang, J. L. Smith, and W. A. Cramer. 2003. Structure of the cytochrome b₆f complex of oxygenic photosynthesis: tuning the cavity. *Science* **302**:1009–1014.
- Lambert, J. D., and N. A. Moran. 1998. Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. USA* **95**:4458–4462.
- Lynch, M. 1997. Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes. *Mol. Biol. Evol.* **14**:914–925.
- Martin, W., B. Stoebé, V. Goremykin, S. Hansmann, M. Hasegawa, and K. V. Kowallik. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**:162–165.

- Meeks, J. C., J. Elhai, T. Thiel, M. Potts, F. Larimer, J. Lamerdin, P. Predki, and R. Atlas. 2001. An overview of the genome of *Nostoc punctiforme*, a multicellular, symbiotic cyanobacterium. *Photosynth. Res.* **70**:85–106.
- Melis, A. 1999. Photosystem-II damage and repair cycle in chloroplasts: what modulates the rate of photodamage in vivo? *Trends Plant Sci.* **4**:130–135.
- Moran, N. A. 1996. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. USA* **93**:2873–2878.
- Mullineaux, C. W. 1994. Excitation energy transfer from phycobilisomes to photosystem I in a cyanobacterial mutant lacking photosystem II. *Biochim. Biophys. Acta* **1184**:71–77.
- Mullineaux, C. W., M. J. Tobin, and G. R. Jones. 1997. Mobility of photosynthetic complexes in thylakoid membranes. *Nature* **390**:421–424.
- Nakamura, Y., T. Kaneko, S. Sato et al. (18 co-authors). 2002. Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. *DNA Res.* **9**:123–130.
- Nakamura, Y., T. Kaneko, S. Sato et al. (16 co-authors). 2003. Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. *DNA Res.* **10**:137–145.
- Naterstad, K., A. Kolsto, and R. Sirevag. 1995. Physical map of the genome of the green phototrophic bacterium *Chlorobium tepidum*. *J. Bacteriol.* **177**:5480–5484.
- Naylor, G. W., H. A. Adlense, L. C. D. Gibson, and C. N. Hunter. 1999. The photosynthesis gene cluster of *Rhodospira rubra*. *Photosynth. Res.* **62**:121–139.
- Nei, M., and S. Kumar. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, New York.
- Palenik, B., B. Brahamsha, F. W. Larimer et al. (15 co-authors). 2003. The genome of a motile marine *Synechococcus*. *Nature* **424**:1037–1042.
- Pamilo, P., M. Nei, and W.-H. Li. 1987. Accumulation of mutations in sexual and asexual populations. *Genet. Res.* **49**:135–146.
- Papenbrock, J., S. Grafe, E. Kruse, F. Hanel, and B. Grimm. 1997. Mg-chelatase of tobacco: identification of a Chl D cDNA sequence encoding a third subunit, analysis of the interaction of the three subunits with the yeast two-hybrid system, and reconstitution of the enzyme activity by co-expression of recombinant C. *Plant J.* **12**:981–990.
- Pazos, F., and A. Valencia. 2001. Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng.* **14**:609–614.
- Prasil, O., N. Adir, and I. Ohad. 1992. Dynamics of photosystem II: mechanism of photoinhibition and recovery processes. Pp. 295–348 in J. Barber, ed. *The photosystems, topics in photosynthesis*. Elsevier, Amsterdam, The Netherlands.
- Race, H. L., R. G. Herrmann, and W. Martin. 1999. Why have organelles retained genomes? *Trends Genet.* **15**:364–370.
- Rich, P. R. 1984. Electron and proton transfers through quinones and cytochrome bc complexes. *Biochim. Biophys. Acta* **768**:53–79.
- Rijgersberg, C. P., and J. Amesz. 1980. Fluorescence and energy transfer in phycobiliprotein-containing algae at low temperature. *Biochim. Biophys. Acta* **593**:261–271.
- Rispe, C., and N. A. Moran. 2000. Accumulation of deleterious mutations in endosymbionts: Muller's ratchet with two levels of selection. *Am. Nat.* **156**:425–441.
- Robinson, M., M. Gouy, C. Gautier, and D. Mouchiroud. 1998. Sensitivity of the relative-rate test to taxonomic sampling. *Mol. Biol. Evol.* **15**:1091–1098.
- Robinson-Rechavi, M., and D. Huchon. 2000. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics* **16**:296–297.
- Rocap, G., F. W. Larimer, J. Lamerdin et al. (21 co-authors). 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**:1042–1047.
- Rutherford, A. W., and P. Faller. 2003. Photosystem II: evolutionary perspectives. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358**:245–253.
- Schubert, W. D., O. Klukas, W. Saenger, H. T. Witt, P. Fromme, and K. Krausz. 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.
- Thompson, J., D. Higgins, and T. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
- Trumppower, B. L. 1990. Cytochrome bc1 complexes of microorganisms. *Microbiol. Rev.* **54**:101–129.
- Vermaas, W. F. J. 1994. Evolution of heliobacteria: implications for photosynthetic reaction center complexes. *Photosynth. Res.* **41**:285–294.
- Widger, W. R., W. A. Cramer, R. G. Herrmann, and A. Trebst. 1984. Sequence homology and structural similarity between cytochrome b of mitochondrial complex III and the chloroplast b₆-f complex: position of the cytochrome b hemes in the membrane. *Proc. Natl. Acad. Sci. USA* **81**:674–678.
- Wolfe, G. R., F. X. Cunningham, D. Durnford, B. R. Green, and E. Gantt. 1994. Evidence for a common origin of chloroplasts with light-harvesting complexes of different pigmentation. *Nature* **367**:566–568.
- Wollman, F.-A., L. Minai, and R. Nechushtai. 1999. The biogenesis and assembly of photosynthetic proteins in thylakoid membranes. *Biochim. Biophys. Acta* **1411**:21–85.
- Xiong, J., and C. E. Bauer. 2002. A cytochrome b origin of photosynthetic reaction centers: an evolutionary link between respiration and photosynthesis. *J. Mol. Biol.* **322**:1025–1037.
- Xiong, J., W. M. Fischer, K. Inoue, M. Nakahara, and C. E. Bauer. 2000. Molecular evidence for the early evolution of photosynthesis. *Science* **289**:1724–1730.
- Xiong, J., K. Inoue, and C. E. Bauer. 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.
- Zouni, A., H.-T. Witt, J. Kern, P. Fromme, N. Krauss, W. Saenger, and P. Orth. 2001. Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. *Nature* **409**:739–743.

Peter Lockhart, Associate Editor

Accepted July 4, 2005