

The photosynthetic apparatus of *Prochlorococcus*: Insights through comparative genomics

Wolfgang R. Hess¹
Gabrielle Rocap²
Claire S. Ting³
Sallie W. Chisholm^{2,3}

¹Humboldt-University Berlin, Germany; Institute of Biology, Chausseestr. 117, D-10115
Berlin, Germany;

²Massachusetts Institute of Technology, Department of Civil and Environmental Engineering
and ³Department of Biology, Massachusetts Institute of Technology, 15 Vassar Street, 48-425
MIT, Cambridge MA 02139, USA

Submitted to: *Photosynthesis Research* (a special issue on genomics) May 2001

Key words: Chlorophyll *b* synthetase, carotenoids, reaction center, marine picophytoplankton, *Prochlorococcus*, comparative genomics, phycobiliprotein, Prochlorophyte Chlorophyll-Binding protein, high light-inducible proteins, cyanobacteria

Abstract

Within the vast oceanic gyres a significant fraction of the total chlorophyll belongs to the light-harvesting antenna systems of a single genus, *Prochlorococcus*. This organism, discovered only about ten years ago, is an extremely small, Chl *b*-containing cyanobacterium that sometimes constitutes up to 50 % of the photosynthetic biomass in the oceans. Various *Prochlorococcus* strains are known to have significantly different conditions for optimal growth and survival. Strains which dominate the surface waters, for example, have an irradiance optimum for photosynthesis of 200 $\mu\text{moles photons m}^{-2} \text{ s}^{-1}$, whereas those that dominate the deeper waters photosynthesize optimally at 30-50 $\mu\text{moles photons m}^{-2} \text{ s}^{-1}$. These high and low light adapted “ecotypes” are very closely related – less than 3% divergent in their 16S rRNA sequences – inviting speculation as to what features of their photosynthetic mechanisms might account for the differences in photosynthetic performance. Here we compare information obtained from the complete genome sequences of two *Prochlorococcus* strains, with special emphasis on genes for the photosynthetic apparatus. These two strains, *Prochlorococcus* MED4 and MIT9313, are representatives of high and low-light adapted ecotypes, characterized by their low or high Chl *b/a* ratio, respectively. Both genomes appear to be significantly smaller (1,700 and 2,300 kbp) than those of other cyanobacteria, and the low light adapted strain has significantly more genes than its high light counterpart. In keeping with their comparative light-dependent physiologies, MED4 has many more genes encoding putative high light inducible proteins (HLIP) and photolyases to repair UV induced DNA damage, whereas MIT9313 possesses more genes associated with the photosynthetic apparatus. These include two *pcb* genes encoding Chl-binding proteins and a second copy of the gene *psbA*, encoding the Photosystem II reaction center protein D1. In addition MIT9313 contains a gene cluster to produce chromophorylated phycoerythrin. The latter represents an intermediate form between the phycobiliproteins of non-Chl *b* containing cyanobacteria and an extremely modified phycoerythrin as the sole derivative of phycobiliproteins still present in MED4. Intriguing features found in both *Prochlorococcus* strains include a gene cluster for Rubisco and carboxysomal proteins that is likely of non-cyanobacterial origin and two genes for a putative and lycopene cyclase, respectively, explaining how *Prochlorococcus* may synthesize the β -branch of carotenoids that are common in green organisms but not in other cyanobacteria.

Abbreviations: bp – base pair, nt - nucleotide CAO - chlorophyllide *a* monooxygenase, Chl – chlorophyll, Chl *b*₂ – divinyl chlorophyll *b*, HLIP – high light-inducible protein, PCB - Prochlorophyte Chlorophyll-Binding protein, PE – phycoerythrin, Rubisco - Ribulose-1,5-bisphosphate carboxylase/oxygenase

Introduction

The genus *Prochlorococcus* (Chisholm et al., 1992; Chisholm et al., 1988) belongs to the cyanobacterial radiation (Palenik and Haselkorn, 1992; Urbach et al., 1992), and forms a branch distinct from the other two chlorophyll (Chl) *b* containing cyanobacteria (“Prochlorophytes”), *Prochloron* (Lewin, 1976; Lewin and Withers, 1975) and *Prochlorothrix* (Burger-Wiersma et al., 1986), and from the chloroplast lineage. The photosynthetic apparatus of *Prochlorococcus* possesses several intriguing features including the presence of the divinyl forms of Chl, the accumulation of Chl *b*₂ as the major light harvesting pigment, the occurrence of high concentrations of β -carotene and the lack of

recognizable phycobilisomes and phycobilins except phycoerythrin. Thus this photosynthetic prokaryote has evolved a variety of ways to harvest light. One of its key properties is the ability to perform photosynthesis successfully under irradiances from less than 1 up to $\sim 2,000 \mu\text{moles photons m}^{-2} \text{ s}^{-1}$ (Moore and Chisholm, 1999; Moore et al., 1995; Partensky et al., 1999b).

The complete genomes of three strains of *Prochlorococcus marinus* are being sequenced, two at the Joint Genome Institute in Walnut Creek, CA, and one at the Genoscope in Paris. Although not yet completely analyzed and annotated, the information obtained so far for two strains MED4 and MIT9313 (details see http://spider.jgi-psf.org/JGI_microbial/html/prochlorococcus_homepage.html), has begun to reveal some interesting features of the genetic inventory of this unusual cyanobacterium. These data provide a window into the evolutionary processes that formed the photosynthetic apparatus of this Chl *b* possessing oxyphototroph as it differentiated from its phycobilisome-containing ancestors. The genome of *Prochlorococcus* MED4 is extremely minimal, with a size of only 1,657,995 bp and 1,686 putative protein-coding genes. Of these about 10 % are directly or indirectly involved in photosynthesis.

The ecology, physiology and genetic diversity of *Prochlorococcus*

In open ocean ecosystems, carbon fixation is dominated by the closely related marine cyanobacteria *Prochlorococcus* and *Synechococcus*. Together they have been shown to contribute between 32 - 80% of the primary production in the oligotrophic oceans (Goericke and Welschmeyer, 1993; Li, 1995; Liu et al., 1997; Veldhuis et al., 1997). *Prochlorococcus* is numerically the most abundant, and it can grow deeper in the euphotic zone (Partensky et al., 1999a) presumably because Chl *b*₂ allows it to efficiently harvest the low light intensities and blue wavelengths characteristic of deep water (Moore et al., 1995; Morel et al., 1993).

Prochlorococcus isolates can be divided into two physiologically and genetically distinct groups, referred to as ecotypes because their differing physiologies have implications for their ecological distributions. The primary differentiator of the ecotypes is their Chl *b*₂/*a*₂ ratio, (Figure 1A) although they also differ in their light-dependent physiological responses (growth rate, pigment content, and photosynthetic rate as a function of light intensity). Isolates with a high Chl *b*₂/*a*₂ ratio (high B/A ecotype) are much more efficient at utilizing low light than those with a low Chl *b*₂/*a*₂ ratio (low B/A ecotype), but are incapable of growth at higher irradiances (Moore and Chisholm, 1999; Figure 1B). In addition, high B/A isolates are much more sensitive to copper than isolates of the low B/A ecotype (Mann, 2000). The physiological diversity among *Prochlorococcus* isolates is correlated with genetic diversity (Moore et al., 1998; Figure 2). Analysis of 16S rDNA and 16S-23S internal transcribed spacer (ITS) sequences demonstrates that low B/A isolates are closely related and form a well supported clade, which can be further divided into two clusters (Urbach et al., 1998; Rocap et al., 1999; Fig. 2). In contrast, the high B/A isolates, while distinctly different from the low B/A isolates, do not form a monophyletic clade of their own but instead are divided among four clusters (Rocap, 2000; Rocap et al., 1999). Thus the differences in physiology correlate well with the position of the respective genotype in a phylogenetic tree such as that shown in Figure 2.

The *Prochlorococcus* strains selected for genome analysis represent both physiological types and the genetically most different genotypes available so far. Since MIT9313 belongs to the most deeply branching lineage of *Prochlorococcus* while MED4 is more recently evolved

their comparative analysis has additional far-reaching implications. The low B/A strains are evolutionarily considerably younger than the high B/A strains and the former must have evolved from the latter (Urbach et al., 1998). Therefore their comparative analysis allows us to follow the dynamics of genome evolution in an unprecedented way.

General genome properties

The genomes of *Prochlorococcus* are very small for photosynthetic prokaryotes: The more recently evolved strain, MED4, has the smallest genome of any known oxygenic phototroph (1,657,995 bp). Even the somewhat less derived strains have relatively small genomes with about 2.4 Mbp in case of MIT9313 and 1.8 Mbp for SS120 (Strehl et al., 1999). The first cyanobacterial genome sequenced, that of the well studied model strain *Synechocystis* PCC 6803, is 3.6 Mbp (Kaneko et al., 1996), whereas the Chl *b* possessing prokaryote *Prochlorothrix hollandica* has an even bigger genome 5.5 Mbp (Schyns et al., 1997) and the soil cyanobacterium *Nostoc punctiforme* PCC 73102 is over 9 Mbp (J. Meeks, pers. comm.). When these data are superimposed on the phylogenetic relationships shown in Figure 2, it is very obvious that the known genomes form a gradient of decreasing size within the *Prochlorococcus* clade with the biggest genome belonging to MIT9313, the genotype closest to *Synechococcus* and the smallest to MED4, one of the strains most distal in the tree. Assuming a similar coding capacity of about 1 gene per 1,000 bp in both strains, about 40 % additional genes can be expected in MIT9313 than in MED4 and this is indeed what has been found - the former genome contains about 2,200 genes, whereas the latter has only 1,686 putative protein-coding genes (accessible at http://spider.jgi-psf.org/JGI_microbial/html/prochlorococcus_homepage.html).

A similar gradient can be seen for the total G+C content which decreases from about 50.7 % in MIT9313 to 30.79 % in MED4. In this respect the MED4 genome is surprisingly close to the bacterium *Ureaplasma urealyticum*, which has the lowest value (25.5 % G+C) of all bacterial genomes analyzed so far (Glass et al., 2000). AT-rich genomes can be expected to be more susceptible to mutations due to the formation of UV light-induced thymidine dimers, something that mucosal human pathogens such as *Ureaplasma* should not be readily exposed to. But as a photosynthetic organism, *Prochlorococcus* is regularly exposed to deleterious UV radiation and this is true in particular for the high light-adapted MED4, which lives preferentially in surface waters. Thus one might expect selection pressures to result in the opposite trend i.e. for MED4 to have the higher G+C content of the two ecotypes. Thus the driving force for the low G+C content in MED4 is unclear, and the question of how it copes the expected high mutation rate, for example by efficient UV-inducible repair systems, requires experimental analysis.

The total number of genes coding for the “inner circle” of photosynthesis, i.e. light-harvesting proteins, proteins of photosystems I and II, photosynthetic electron transfer chain and CO₂ fixation, is approximately 85 and 90 in *Prochlorococcus* MED4 and MIT9313, respectively, compared to about 100 in the cyanobacterium *Synechocystis* PCC 6803 (Kaneko et al., 1996). The majority of the genes from this set that are missing in *Prochlorococcus* are those encoding phycobiliproteins, and linkers and enzymes involved in the formation and attachment of chromophores to this group of proteins.

Light-harvesting systems

In most cyanobacteria, photosynthetic light harvesting occurs by complex supramolecular structures, the phycobilisomes. This is not the case in *Prochlorococcus*. As in the two other Chl *b*-possessing prokaryotes, *Prochloron* sp. and *Prochlorothrix hollandica*, the core protein of the light harvesting system is a membrane-bound protein with six putative membrane-spanning regions (La Roche et al., 1996). Past work has demonstrated that in *Prochlorococcus* these Prochlorophyte Chlorophyll-Binding proteins (PCBs) are encoded by a single gene in all the low B/A strains investigated, whereas multigene families have been found in several high B/A strains (Garczarek et al., 2000). The genome data support this to some extent, as *pcb* is indeed a single gene in the high light-adapted MED4. However, in MIT9313 only two such genes are present. This is somewhat surprising in view of the five or more copies found in the other high B/A strains investigated so far (Garczarek et al., 2000). The two Pcb in MIT9313 are not identical and their deduced amino acid sequences exhibit only 59% sequence identity. In comparison, the single Pcb in MED4 exhibits 69% and 54% amino acid sequence identity with Pcb1 and Pcb2 of MIT9313. In order to establish what the functional importance is of single and multiple *pcb* gene copies, further work is clearly required on the patterns of *pcb* gene expression under different conditions, and on the pigment binding capacity and structure of the different Pcb proteins. It is possible that possession of several *pcb* genes may assure the production of sufficient Chl *b*₂-binding protein to sustain photosynthesis at the extreme low light conditions that can sustain growth in the high B/A strains (Garczarek et al., 2000). Furthermore, the lower number of *pcb* genes in MIT9313 relative to SS120 (which has 7 *pcb* genes) may explain why MIT9313 is not quite as well adapted to grow at extremely low irradiances (below 5 $\mu\text{moles photons m}^{-2} \text{s}^{-1}$) and has a lower average Chl *b/a* ratio than SS120 (Garczarek et al., 2000; Moore and Chisholm, 1999).

Phycobiliprotein genes

Although *Prochlorococcus* does not possess phycobilisomes as its major light harvesting apparatus, earlier work has shown that several strains contain genes encoding phycoerythrin, which in most cyanobacteria make up the most peripheral part of the phycobilisome. These include *Prochlorococcus* SS120 (Hess et al., 1996), PAC1 and PAC2 (Penno et al., 2000), and MIT9313 and MIT9303 (Ting et al., 1999; Ting et al., submitted), which are all high B/A strains. In contrast, low B/A strains such as PCC 9511 (Rippka et al., 2000) or AS9601, appear to lack phycoerythrin and the corresponding genes as detected by immunological methods or PCR (Penno et al., 2000).

The whole genome data now illustrate in considerable detail the fate of the phycobilisome genes. In marine *Synechococcus* WH8020 or WH8102, which are closely related to *Prochlorococcus*, the proteins required for PBS are encoded by ~40 genes located in several operons and gene clusters. The largest such gene cluster has previously been described in *Synechococcus* WH8020 (Wilbanks and Glazer, 1993). Over a length of about 15 kb, eighteen closely linked genes encode phycocyanin, several polypeptides involved in the synthesis or attachment of chromophoric groups, and two different forms of phycoerythrin, PE(I) and PE(II), (Figure 3). The presence of a second set of phycoerythrin genes, *mpeB* and *mpeA*, in addition to *cpeB* and *cpeA* is a speciality of marine *Synechococcus* (Ong and Glazer, 1991). Surprisingly, in *Prochlorococcus* SS120 (Hess et al., 1999) and MIT9313 (Ting et al., submitted), although all genes encoding phycocyanin are lacking and there is no second set of genes for the phycoerythrin and subunits, the basic structure of the PE gene cluster has been retained. The presence of a set of genes encoding the phycoerythrin and subunits in MIT9313 supports the suggestion that this gene cluster is the minimal set of genes required to produce photophysiologicaly active phycoerythrin (Hess et al., 1999). The additional

presence of an homologue to *ho1* (Willows et al., 2000) in this genome region is intriguing: chromophores of phycobiliproteins are biosynthesized from heme in a pathway that begins with the opening of the tetrapyrrole macrocycle of protoheme to form biliverdin IX_b, in a reaction catalyzed by heme oxygenase; the gene encoding this activity is *ho1* (Cornejo et al., 1998; Richaud and Zabulon, 1997). Neither genome possesses genes encoding allophycocyanin, phycocyanin, core-membrane or rod-core linker proteins and phycobilisome degradation proteins such as NblA (Collier and Grossman, 1994).

It is appealing to suppose that the presence of phycoerythrin as an additional light-harvesting system in the high B/A ecotypes may support photosynthesis in these genotypes under very low irradiances, i.e. those receiving only about 0.21 % surface PAR (Johnson et al., 1999). The presence of phycourobilin as the dominating chromophore (Hess et al., 1996) is in agreement with such a scenario as its absorption peak at 495 nm corresponds exactly to the wavelength that penetrates best the water column down to depths relevant in this context, that is > 100 m (cf. Figure 5 in Johnson et al., 1999). However, an obvious difference between the phycoerythrin gene clusters is the lack of a *ppeC* homologue in MIT9313 (Figure 3), the sole gene that encodes a putative linker protein in SS120 (Hess and Partensky, 1999; Hess et al., 1999). The absence of just this protein should interfere with the anticipated role of *Prochlorococcus* phycoerythrins in light harvesting. Both their association to thylakoid membranes (Hess et al., 1999) as well as their properties in energy absorption and transfer (Lokstein et al., 1999) would depend on it strongly. Therefore only through experimental modification of such a gene cluster can the relevance of the light-harvesting properties of PE in *Prochlorococcus* be assessed.

Even more surprising is this genome region in the low B/A strain MED4. Two genes, *metK* and *uvrD*, that are located just outside the PE gene cluster in *Prochlorococcus* SS120, can serve as a marker. They are still in close proximity in MED4 (Figure 3), thus there was no translocation that would have split this genome segment in different pieces. However, the distance between the two genes has shrunk from about 11,500 bp in SS120 to 4,493 bp in MED4. Of the nine genes involved in the formation of PE only two have been retained, *cpeB* and *orf181*, along with a 175 nt region in front of *orf181* that might contain a promoter. Detailed comparisons and alignments of deduced amino acid sequences of -PE from MED4, MIT9303, MIT9313, and other cyanobacteria indicate that in MED4, the deduced -PE sequence is mutated at multiple sites, has lost two out of four widely conserved cysteines for chromophore binding, and represents an extremely degenerate form of phycoerythrin (Ting et al., submitted). These striking changes, in addition to the absence of the gene encoding the -phycoerythrin subunit in MED4, suggest that the phycoerythrin genes are in the process of being lost from the *Prochlorococcus* lineage (Ting et al., submitted), as appears to have occurred in several cyanobacterial lines (Apt et al., 1995). However, *cpeB* does not appear to be a pseudogene in MED4 because the reading frame has remained intact and the gene is expressed (Penno and Hess, unpublished). It is unlikely, however, that its gene product has a role in photosynthetic light harvesting given the absence of virtually all other genes involved in the biosynthesis of phycobiliproteins. A more realistic function might be that of a light sensor pigment. This hypothesis is supported indirectly by the presence of a tricistronic gene cluster *ho1-pebA-pebB* (Figure 3) for the biosynthesis of the putative chromophore, phycoerythrobin, in this strain (Frankenberg et al., 2001). Furthermore, neither of the *Prochlorococcus* genomes contain known photoreceptor genes, such as those encoding phytochromes, which have very important functions in cyanobacteria: CikA in *Synechococcus elongatus* serves to reset the clock in response to light (Schmitz et al., 2000), RcaE in *Fremyella diplosiphon* is critical for complementary chromatic adaptation (Kehoe and

Grossman, 1996), and *Synechocystis* Cph1 appears to be a taxis receptor (Vierstra and Davis, 2000; Yoshihara et al., 2000).

Chromosomal organization of photosystem genes

The chromosomal organization of several photosynthetic apparatus genes is similar between *Prochlorococcus* strains MED4 and MIT9313 and not fundamentally different from other cyanobacteria. In both strains, genes encoding the major proteins of the photosynthetic apparatus are distributed throughout the chromosome, with several forming distinct clusters that are most likely operons. For Photosystem II (PSII), these genes include *psbD* and *psbC*, which encode the D2 and CP43 polypeptides, respectively. While D2 is one of the major PSII reaction center subunits involved in binding the chlorophylls and cofactors involved in primary photochemistry, CP43 is one of the Chl-binding core antenna proteins of PSII. In both the MED4 and MIT9313 chromosomes, *psbD* and *psbC* are not only located adjacent to each other, but overlap, depending on which start codon is actually used, by at least 16 and not more than 91 base pairs. Interestingly, a similar overlap occurs in all cyanobacteria that have been investigated so far as well as in almost all chloroplast genomes. Genes encoding the α - (*psbE*) and β - (*psbF*) subunits of cytochrome *b₅₅₉* are found adjacent to each other and are separated by three base pairs in MED4 and four base pairs in MIT9313. Two genes encoding low molecular weight polypeptides (*psbJ*, *psbL*) associated with PSII are found next to these cytochrome *b₅₅₉* subunit genes in both chromosomes. For Photosystem I (PSI), genes encoding the major reaction center subunits, PsaA and PsaB, are located adjacent to each other and are separated by less than 30 base pairs in both MED4 and MIT9313. Furthermore, the *psaL* gene, which encodes a ~ 16 kDa protein that functions as a connecting protein in PSI trimers, can be found near *psaB* (separated by 1700 base pairs in MED4, 796 base pairs in MIT9313). The PSI proteins PsaF and PsaJ are believed to function in cyclic electron transport and plastocyanin/cytochrome *c₆* docking, respectively. In the MED4 chromosome, the genes encoding these proteins are separated by only 31 base pairs. Genes encoding the major components of the cytochrome *b_{6f}* complex exhibit similar organization in MED4 and MIT9313. The *petC* (Rieske-FeS) and *petA* (cytochrome *f*) genes are adjacent to each other. While these genes overlap by 22 base pairs in the MED4 chromosome, they are separated by 210 base pairs in the MIT9313 chromosome. Furthermore, the *petD* (subunit IV) and *petB* (cytochrome *b₆*) genes are found together and are separated by only 43 base pairs in MED4 and by 91 base pairs in MIT9313.

The gene *psbA* encodes the PSII reaction center protein D1 and is present in multiple copies in almost all cyanobacteria investigated so far. While *Anabaena* PCC7120 has four *psbA* copies (Vrba and Curtis, 1989), *Synechococcus* strains PCC7002, and PCC 7942 and *Synechocystis* strains PCC 6714 and PCC 6803 each have three *psbA* genes (Bouyoub et al., 1993; Gingrich et al., 1988; Golden et al., 1986). All the cyanobacteria which have been more closely investigated express only two different *iso*-forms of the D1 protein, which are differentially regulated by light. In *Synechococcus* PCC7942 both forms have been shown to participate greatly in photoacclimation (Öquist et al., 1995; Soitamo et al., 1996). In contrast, MED4 has only a single *psbA* gene, consistent with previous studies on SS120 (Hess et al., 1995), while MIT9313 possesses two genes encoding identical PSII D1 polypeptides. One copy is located in the same genomic context as in MED4 and in SS120 (Hess, 1997; Hess et al., 1995) whereas the second copy is at a different genomic location. The two *psbA* genes that are present in the sequenced MIT9313 genome differ in only one base pair. The deduced amino acid sequences of MIT9313 *psbA1* and *psbA2* share 87% identity with *psbA* of MED4. The presence of a single D1 protein in both *Prochlorococcus* strains suggests that the

mechanisms by which *Prochlorococcus* responds to high light-induced stress at the level of the PSII reaction center may differ from other cyanobacteria that possess multiple PSII reaction center genes, whose expression is transcriptionally regulated in response to photon flux density.

The *psbD* gene encodes the second PSII reaction center protein, D2, and is also often present in multiple copies. *Synechococcus* PCC 7942, *Synechococcus* PCC 7002 and *Synechococcus* PCC 6803 all possess two *psbD* genes that encode identical proteins in each strain (Gingrich et al., 1990; Golden and Stearns, 1988; Williams and Chisholm, 1987). The alteration of their expression in response to changes in irradiance has been well documented for several cyanobacteria. The second copy of the *psbD* gene in *Synechococcus* PCC7942, *psbDII*, has been shown to be important for light adaptation (Bustos and Golden, 1992). It provides an additional source for the mRNA (Colon-Lopez and Sherman, 1998) to produce more protein under conditions when an overexpression of D2 with regard to CP43 is required, that is to compensate for its more rapid turnover under light stress or illumination by UV-B (Bustos and Golden, 1992). However, in both MED4 and MIT9313 only a single *psbD* gene is present, located together with *psbC* in an operon. There is no additional free-standing *psbD* gene as in other cyanobacteria (Bustos and Golden, 1992). Consequently, there must be other mechanisms to adapt to conditions such as described above, or the capacity to cope with excess light and UV-B must be reduced or was lost in *Prochlorococcus* (Garczarek et al., 2001).

The presence or absence of seven amino acids at the carboxy-terminus of D1 proteins has previously been used as a phylogenetic marker (Lockhardt et al., 1993; Maid et al., 1990; Morden and Golden, 1989). While all land plant chloroplasts and the Chl *b* possessing cyanobacterium *Prochlorothrix hollandica* lack these 7 amino acids, they are present in *Cyanophora*, non-green algal plastids, and all other cyanobacteria examined to date (Morden and Golden, 1989).. The association of the D1 proteins of both MED4 and MIT9313 as well as that of SS120 (Hess et al., 1995) with an approximately seven amino acid carboxy terminus domain, is consistent with phylogenetic data indicating that *Prochlorococcus* evolved from a phycobilisome-containing, cyanobacterial-like ancestor and is not specifically closely related to land plant chloroplasts or other Chl *b* possessing cyanobacteria (Palenik and Haselkorn, 1992; Urbach et al., 1992).

Genes encoding High Light Inducible Proteins

The high light inducible proteins (HLIP), are proteins with a single transmembrane helix that may play a role in the acclimation of photosynthetic organisms to conditions associated with photooxidative stress (Dolganov et al., 1995; He et al., 2001). HLIPs are related to the early light-inducible proteins (ELIPs) that have been studied in several higher plants (Adamska et al., 1992; Grimm and Kloppstech, 1987; Lindahl et al., 1997; Meyer and Kloppstech, 1984; Potter and Kloppstech, 1993), and the stress enhanced proteins (Seps) characterized in *Arabidopsis* (Heddad and Adamska, 2000). These proteins are members of an extended family which also includes the Chlorophyll *a/b* (CAB)-binding light-harvesting antenna proteins (Dolganov et al., 1995; Green and Kuhlbrandt, 1995; He et al., 2001). Hydrophathy plots predict that ELIPs possess three transmembrane helices, the first and third of which are highly conserved between these proteins and light-harvesting complex polypeptides (Green and Kuhlbrandt, 1995; Green and Pichersky, 1994), whereas Seps have two membrane spanning helices (Heddad and Adamska, 2000). It has been hypothesized that the one-helix

stress-response proteins may have been the evolutionary precursors of light harvesting antennas (Chl *a/b* proteins, fucoxanthin Chl *a/c* proteins; Green and Kuhlbrandt, 1995). Through a series of duplication, fusion, and deletion events, genes encoding one helix HLIPs may have given rise to the genes encoding three-helix members of this extended family (Green and Kuhlbrandt, 1995).

Although HLIPs were first identified in cyanobacteria (Dolganov et al., 1995), genes encoding HLIPs have since been described in other organisms, including red algae (Reith and Munholland, 1995), glaucophytes (Stirewalt et al., 1995) and higher plants (*Arabidopsis*) (Jansson et al., 2000). The single transmembrane helix of these proteins generally shares sequence homology to transmembrane helix I (B) or III (A) of light-harvesting complex polypeptides (Dolganov et al., 1995; Green and Kuhlbrandt, 1995). *Prochlorococcus* strains MED4 and MIT9313 both possess several putative HLIP genes. Many of these were initially missed by commonly used gene modelling programs because of their short length and the low levels of overall identity often shared between sequences. Additional BLAST searches against the two *Prochlorococcus* genomes and manual analyses of conserved sequence motifs revealed the presence of several additional putative HLIP genes (B. Green and C. Ting, unpubl.). As many as 20 putative HLIP genes may be present in the MED4 genomes, nearly twice as many as found in MIT9313 (B. Green and C. Ting, unpublished data). The deduced amino acid sequences range in size from approximately 45 to 100 amino acids, with an average size of about 64 amino acids. Sequence alignments indicate that a highly conserved pigment binding motif present in helix III (A) of the light-harvesting complex (LHCII) polypeptide (He et al., 2001), is also found in the MED4 and MIT9313 HLIP sequences (B. Green and C. Ting, unpubl.). Hydropathy plots predict that the majority of the *Prochlorococcus* putative HLIPs possess one membrane spanning helix. Future studies may reveal that this difference in the number of HLIP genes may contribute to the ability of MED4 to grow successfully at higher photon flux densities than MIT9313.

Carbon assimilation

The rate limiting catalyst for carbon fixation is Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), encoded in cyanobacteria by *rbcL* and *rbcS* (or *cbbL* and *cbbS* in chemoautotrophs). Autotrophic organisms vary in their ability to carry out the CO₂ fixation reaction in the presence of O₂, measured by the substrate specificity factor (defined as $V_{CO_2}K_{O_2}/V_{O_2}K_{CO_2}$). Atmospheric CO₂ levels have decreased, and O₂ has increased over evolutionary time, and it has been suggested that Γ has increased in more recently derived organisms in response to the selective pressures (Raven, 1997a; Raven, 1997b; Tortell, 2000). Thus cyanobacteria such as *Synechococcus* PCC6301 and *Anabaena* whose direct ancestors evolved in a high CO₂ /low O₂ environment have values of Γ two to five fold lower than those of eukaryotic phototrophs such as diatoms, coccolithophores, green and red algae (Tabita, 1999). As a consequence, organisms with low Γ have evolved carbon concentrating mechanisms (CCMs) to make up for the low CO₂ affinity of their Rubisco (Raven, 1997a; Raven, 1997b; Tortell, 2000). These CCMs require energy investment, and thus may not be advantageous at very low light levels. Where does *Prochlorococcus*, a relatively recent lineage in the ancient cyanobacterial group, fit into this picture? The data from the two genomes suggest that carbon fixation in *Prochlorococcus* may be quite different from that of well studied cyanobacteria such as *Synechococcus* PCC 7942 and *Synechococcus* PCC 6301.

Carbon uptake

An efficient CCM requires both the active uptake of inorganic carbon in the form of CO₂ and/or HCO₃⁻ and the creation of an elevated local CO₂ concentration within the carboxysome, in close proximity to Rubisco (Kaplan and Reinhold, 1999). It is quite intriguing that there are no genes with homology to known transporters for inorganic carbon in either of the *Prochlorococcus* genomes. The genomes are lacking genes for an ABC-type bicarbonate transporter identified in *Synechococcus elongatus* PCC 7942 (Omata et al., 1999) and they have no homologue to *orf427* which was identified as being implicated in CO₂ uptake in *Synechococcus* PCC 7002 (Klughammer et al., 1999). Thus it is not at all clear how inorganic carbon is transported by *Prochlorococcus*.

Carbon concentration

Carboxysomes are polyhedral inclusion bodies found in all cyanobacteria and a number of chemoautotrophic bacteria (Shively et al., 1998b). A majority of the cellular Rubisco is located in the carboxysomes and they are an important part of the CCM. In both the chemoautotroph *Thiobacillus neapolitanus* and the freshwater cyanobacterium *Synechococcus* PCC 7942, insertional mutations in carboxysome genes result in an elevated CO₂ requirement for growth (Shively et al., 1998b). In *Prochlorococcus* carboxysomes are often visible by EM (Chisholm et al., 1992; Chisholm et al., 1988).

In both *Prochlorococcus* genomes several open reading frames located downstream of *rbcS* have high homology to genes from the *cso* operon of *Halothiobacillus neapolitanus* (Shively et al., 1998a; Shively et al., 1998b) which encodes carboxysomal shell proteins (Figure 4A). The identity of the polypeptide sequences deduced from MED4 open reading frames or1175, or1176 and or1177 to the *Thiobacillus* genes *csoS2*, *csoS3* and *orfA* (all acc. no. AF129925), is 33%, 43%, and 53% respectively. At the opposite end, upstream of *rbcL*, one more putative carboxysomal gene can be found (or1172). The encoded product is highly similar (91% identity) to CsoS1A from *Thiobacillus denitrificans* (AF0129292) and is even more similar to the atypical *ccmK* gene from *Synechococcus* WH7803 (P96485; Watson and Tabita, 1996).

Carbonic anhydrase has been found associated to the carboxysomes in several cyanobacteria (Price et al., 1992; So and Espie, 1998). Because it generates CO₂ from accumulated HCO₃⁻, it may play a central role in the enrichment of CO₂ in the carboxysome. Carbonic anhydrase is widespread in metabolically diverse species within the Archaea and Bacteria. It exists in three distinct types (designated α , β , and γ class) without significant sequence identity, suggesting that the different forms evolved independently from each other (Smith and Ferry, 2000). It is very surprising that both *Prochlorococcus* genomes do not contain any genes with homology to any of the known carbonic anhydrases. It should be noted however, that to some extent the H⁺ ions generated within the Rubisco-catalyzed reaction could stimulate the formation of CO₂ from HCO₃⁻ without the necessity of a carbonic anhydrase activity. Alternatively, the carboxysome's main function may be in creating an improved K_m value for Rubisco by inserting the enzyme in a particular microenvironment (Shively et al., 1998b). The very tight association of the *Thiobacillus*-type Rubisco small subunit to proteins of the carboxysome shell is consistent with this hypothesis (Holthuijzen et al., 1986a; Holthuijzen et al., 1986b).

Calvin cycle and Rubisco

The first determined *Prochlorococcus rbcL* sequence, from strain GP2, shares a higher percent similarity with form IA Rubiscos of purple bacteria (Shimada et al., 1995) than with typical cyanobacterial form IB Rubiscos. Other *Prochlorococcus* strains PAC1 and MED initially did not seem to follow this scheme (Pichard et al., 1997), and the anomalous branching pattern (with form IA sequences) was also noted for the *rbcL* of marine *Synechococcus* WH7803 (Watson and Tabita, 1996). Based on the total genome sequences both MED4 and MIT9313 have a *rbcL* sequence which is very closely related (>90% aa identity) to the marine *Synechococcus* and *Prochlorococcus* GP2 sequences and thus would appear to be of the Form IA type (Figure 4B). Between *Prochlorococcus* MED4 and GP2 for example, the difference is only one residue and this is a conservative amino acid exchange. Thus in phylogenetic analyses, the MIT9313 and MED4 Rubisco genes cluster closely with form IA *rbcL* sequences from species such as *Thiobacillus denitrificans*, *Hydrogenovibrio marinus*, *Chromatium vinosum*, and, in particular, *Prochlorococcus* GP2 and *Synechococcus* WH7803 and *rbcL* sequences amplified directly from the marine environment (Fig. 4B, J. Paul, pers. communication). Horizontal gene transfer is one possibility which may explain such an atypical appearance of different forms of genes in closely related organisms. Multiple cases of horizontal transfer of Rubisco genes have previously been suggested (Delwiche and Palmer, 1996). However, phylogenetic analyses based on currently available *rbcL* sequences cannot rule out completely convergent evolution as a mechanism for this branching pattern and sequence similarities rather than horizontal gene transfer (C. Cavanaugh personal communication).

Interestingly, the genes encoding two enzymes preceding Rubisco in the Calvin cycle, phosphoribulokinase and pentose-5-phosphate epimerase also bear little if any sequence similarity to their counterparts in cyanobacteria. The MED4 pentose-5-phosphate epimerase is most similar to the *Neisseria meningitidis cbbE* gene product (42/62% identity/similarity, AE002472), the phosphoribulokinase amino acid sequence has the closest match to the *Alcaligenes eutrophus* plasmid pHG1–encoded enzyme (60/73% identity/similarity, M33562). These bacteria belong to the branch of proteobacteria whereas the *Prochlorococcus* Rubisco might have been acquired from a proteobacterium.

Thus, the availability of whole genome sequences now allows a unique perspective on the phylogenetic and physiological implications of the presence of Rubisco form IA genes in *Prochlorococcus*. The genomic region that was putatively obtained by horizontal gene transfer from *Thiobacilli* or related bacteria is not restricted to *rbcL* and *rbcS*, but consists of one contiguous stretch of genes involved in carbon assimilation (seven and eight genes in MED4 and MIT9313, respectively). These include *rbcL*, *rbcS* and the carboxysomal shell proteins discussed above, which also bear closer homology to their counterpart genes in chemoautotrophs than to those other cyanobacteria except *Synechococcus* WH7803. In this marine cyanobacterium a similar situation was described for the *ccmK-rbcL-rbcS* gene cluster (Watson and Tabita, 1996). The gene orders in *Prochlorococcus* MED4 *csoS1A(ccmK)-rbcLS-csoS2-csoS3-orfA-orfB* and MIT9313 *csoS1A(ccmK)-rbcLS-csoS2-csoS3-orfA-orfB-csoS1A(ccmK)* (Figure 4A) are highly similar to those found in chemoautotrophs such as *Halothiobacillus neapolitanus* (*cbbL-cbbS-csoS2-csoS3-orfA-orfB-csoS1C-csoS1A-csoS1B*; Shively et al., 1998b). The important role of native Rubisco associated to carboxysome assembly (Kaplan and Reinhold, 1999) makes it very likely that the whole complex, consisting of carboxysome (CCM) and Rubisco was acquired by the common ancestor of marine *Synechococcus* and *Prochlorococcus* via horizontal gene transfer.

The implications that the putative horizontal gene transfer may have for altered substrate specificity of *Prochlorococcus* and marine *Synechococcus* Rubiscos are unclear, as available

values reported for the purified form IA enzyme from chemoautotrophs are not appreciably larger than those for cyanobacterial IB form Rubisco (Tabita, 1999). Further the lack of identifiable genes for inorganic carbon transport or carbonic anhydrase suggests that the CCM of *Prochlorococcus* is distinct from that of its cyanobacterial relatives and from the chemoautotrophs from whom it may have acquired its Rubisco and carboxysome. Ultimately biochemical characterization of the *Prochlorococcus* Rubisco and CCM will be necessary to determine its CO₂ selectivity and what ecological and physiological advantage it confers.

Pigments

Carotenoid biosynthesis

Carotenoids are versatile photosynthetic pigments in that they both trap light energy in the light-harvesting apparatus and dissipate excess radiant energy thereby protecting the structural integrity of pigment-protein complexes. They can also quench the triplet excited states of Chl and the reactive singlet oxygen (non-photochemical quenching) thus preventing damaging effects of free radicals in the cell. In *Synechococcus* PCC 7942, protection from UV-B radiation by zeaxanthin has been demonstrated (Gotz et al., 1999). In chlorophytes, non-photochemical quenching is achieved by the reversible deepoxidation of the xanthophyll violaxanthin to zeaxanthin (xanthophyll or violaxanthin cycle; Niyogi et al., 1997; Niyogi et al., 1998).

There is no xanthophyll cycle in *Prochlorococcus*. The major carotenoids are carotene and zeaxanthin, and some isolates also have cryptoxanthin and parasiloxanthin (Goericke et al., 2000; Goericke and Repeta, 1992). Thus, in contrast to other cyanobacteria *Prochlorococcus* must have the two different enzyme activities to form the carotenoid and rings, as is the case for chlorophytes. Genes for carotenoid biosynthesis enzymes are frequently clustered into large operons (Krubasik and Sandmann, 2000; Viveiros et al., 2000) but this does not appear to be the case in *Prochlorococcus*, where five single genes and one dicistronic operon are scattered throughout the genome whose encoded products may participate in the formation of these carotenoids from geranylgeranyl pyrophosphate (Figure 5). There are two genes, or182 and or0642 in MED4, that potentially encode desaturases participating in the formation of lycopene from phytoene. Two related desaturases, a phytoene desaturase for the first two reactions and a desaturase for the third and fourth desaturation step are also present in plants, but bacteria and fungi achieve the same result with just one enzyme (Armstrong, 1994; Sandmann, 1994). However, cyanobacteria occasionally possess a second enzyme of the desaturase type as well (Linden et al., 1994). Therefore the presence of two desaturases in *Prochlorococcus* fits with the situation in at least some cyanobacteria.

Another interesting question is whether there are actually two enzymes at the key branch point in the pathway of carotenoid biosynthesis that are to form the carotenoid and rings, and if so where the genes might have come from. Indeed, the two *Prochlorococcus* genomes each contain two genes coding for putative cyclases. Their amino acid sequence identity is 43 % to each other in MED4 and 48 % in MIT9313. The predicted amino acid sequences of these four putative lycopene cyclases bear little resemblance to the known lycopene cyclase enzymes from non-photosynthetic bacteria, but rather are more homologous to the plant and algal enzymes. But they are also not significantly more similar to the corresponding polypeptides from plants than to the single cyclase from *Synechococcus elongatus* PCC 7942 (Cunningham et al., 1994). With 48 – 52 % sequence identity and 65 – 67 % sequence similarity, all four putative enzymes are about equally related to the *Synechococcus elongatus*

PCC 7942 cyclase. Therefore these genes are more likely the result of an ancient duplication of this or a similar gene as *Prochlorococcus* evolved, rather than acquisition by horizontal gene transfer. It is interesting to note that the absence of carotenoids with two rings but the presence of carotenoids with two rings in *Prochlorococcus* also correlates with the inability of the but not the cyclase of *Arabidopsis* to catalyze two rings at the lycopene molecule (Cunningham et al., 1996). This is considered the key mechanism by which *Arabidopsis* adjusts the production and proportions of β -carotene and β -xanthophyll, carotenoids (Cunningham et al., 1996). There is a slightly higher similarity of one of the lycopene cyclase genes in MED4, or0970, to the cyclase of *Arabidopsis*, whereas the gene product of or1509 is slightly more similar to the corresponding cyclase (Figure 5).

Is there a chlorophyll b synthetase gene?

The defining feature of a “Prochlorophyte” is the presence of Chl *b*. Only very recently has Chl *b* synthetase or Chl *a* monooxygenase (CAO) been identified as the enzymatic activity catalyzing the formation of Chl *b* in higher plants (Oster et al., 2000; Tanaka et al., 1998) as well as in *Prochloron didemni* and *Prochlorothrix hollandica* (Tomitani et al., 1999). In *Prochlorococcus* MED4 and MIT9313 no gene with convincing homology to a *cao* gene of other Chl *b*-containing organisms can be found. Thus an independent phylogenetic origin for the *Prochlorococcus* Chl *b* synthetases cannot be rejected. This would be in direct conflict with the theory that all Chl *b* synthetases derived from one ancestral gene and entered eukaryotes via the cyanobacterial-like endosymbiotic progenitor to plastids (Tomitani et al., 1999).

Since the enzymatic activity of Chl *b* synthetase is an unusual two-step oxygenase reaction (Oster et al., 2000), we examined all putative oxygenase genes in the *Prochlorococcus* genomes and found one candidate gene in each strain, potentially encoding for monooxygenases of 440 and 436 residues in MED4 and in MIT9313, respectively. The deduced polypeptides have two particular protein regions, a putative binding domain for a [2Fe-2S] Rieske center and a binding pocket for a mononuclear iron (Figure 6A). Both of these two domains are essential for Chl *b* synthetase. In analyses using BLASTP there was homology to Chl *b* synthetases but with low scores (E values of 10^{-10} and a score of about 80) and less than 28 % sequence identity. Only if the most highly variable regions were taken out of the phylogenetic analysis, a stable position for both the anticipated CAO sequences could be found within the clade of all other CAOs (Figure 6B). In this case, the two *Prochlorococcus* sequences branch at the base of all Chl *b* synthetases but are part of the same sequence cluster. Such a level of similarity could well have been driven by the constraints of this biochemical reaction alone, starting with a gene coding for some kind of an oxygenase. That such a hypothetical convergent evolution did not result in an enzyme more similar to the green lineage CAO's could be a consequence of the slightly different substrate and product, respectively. In contrast to other chlorophyll-containing organisms, in *Prochlorococcus* the divinyl derivative of Chl is made. Experimental evidence is clearly required to solve this problem and to establish the true enzymatic conditions.

Conclusions and Overview

The comparative genomics of the two closely related *Prochlorococcus* strains analyzed herein provide us with clues as to how species formation and adaptation to their particular ecological niches took place in this important group of marine cyanobacteria. One obvious conclusion from this data is the small genome size of *Prochlorococcus*. In both ecotypes the genome is

not only more compactly organized than that of freshwater cyanobacteria, it also lacks many otherwise wide-spread genes or copies of genes. There are no detectable genes for the majority of phycobiliproteins, carbonic anhydrase, carbon uptake, or phytochrome-type photoreceptors. Also of interest are the genes that differ between the two ecotypes, either in their presence/absence or copy number. The high light adapted ecotype MED4 possesses nearly twice as many genes encoding HLIPS as does MIT9313. In contrast MIT9313 has a second *pcb* gene, a second *psbA* gene, and a *cpeA* gene, none of which are present in MED4. These genes may play a role in the ability of MIT9313 to thrive at low light levels where MED4 cannot. Thus we are beginning to get a glimpse of the genetic underpinnings of the differential distribution of the two ecotypes in the surface and deeper waters of the ocean euphotic zone. A concerted effort using the tools of functional genomics will be required to explore the function of these genes as well as to solve some of the questions raised in this review.

Acknowledgements

We wish to thank Lisa Moore for providing unpublished data, Frank Larimer, Jane Lamerdin and Stephanie Stilwagon for genome sequence data of *Prochlorococcus* MED4 and MIT9313 (accessible via the DOE Joint Genome Institute at http://spider.jgi-psf.org/JGI_microbial/html), and Murray R. Badger, Colleen Cavanaugh, Beverley Green, Jonathan King, Heiko Lokstein, Frédéric Partensky, John Paul, Anton Post, Gerhard Sandmann, Claudia Steglich and Nicole Tandeau de Marsac for helpful discussions and insights on the subject. Financial support was provided in part by the US Department of Energy (SWC; DE-FG02-99ER62814), the US National Science Foundation (SWC; OCE-9820035, OCE-0107472) and the Seaver foundation (SWC), the Deutsche Forschungsgemeinschaft, Bonn (WRH; SFB429-TP A4), and the European Union (WRH; PROMOLEC, MAS3-CT97-0128).

Figure Legends

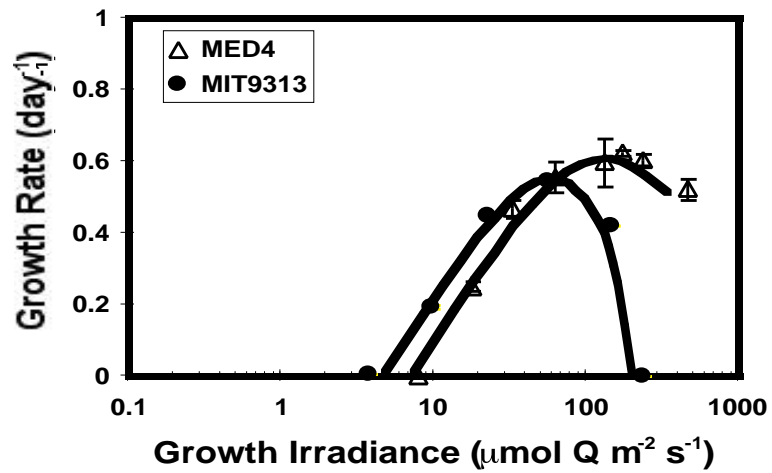


Figure 1 Light-dependent physiology of two *Prochlorococcus* strains, MED4 and MIT9313. (A) Growth rate as a function of light intensity. (B) Chlorophyll b/a_2 as a function of light intensity. Data replotted from Moore and Chisholm (1999).

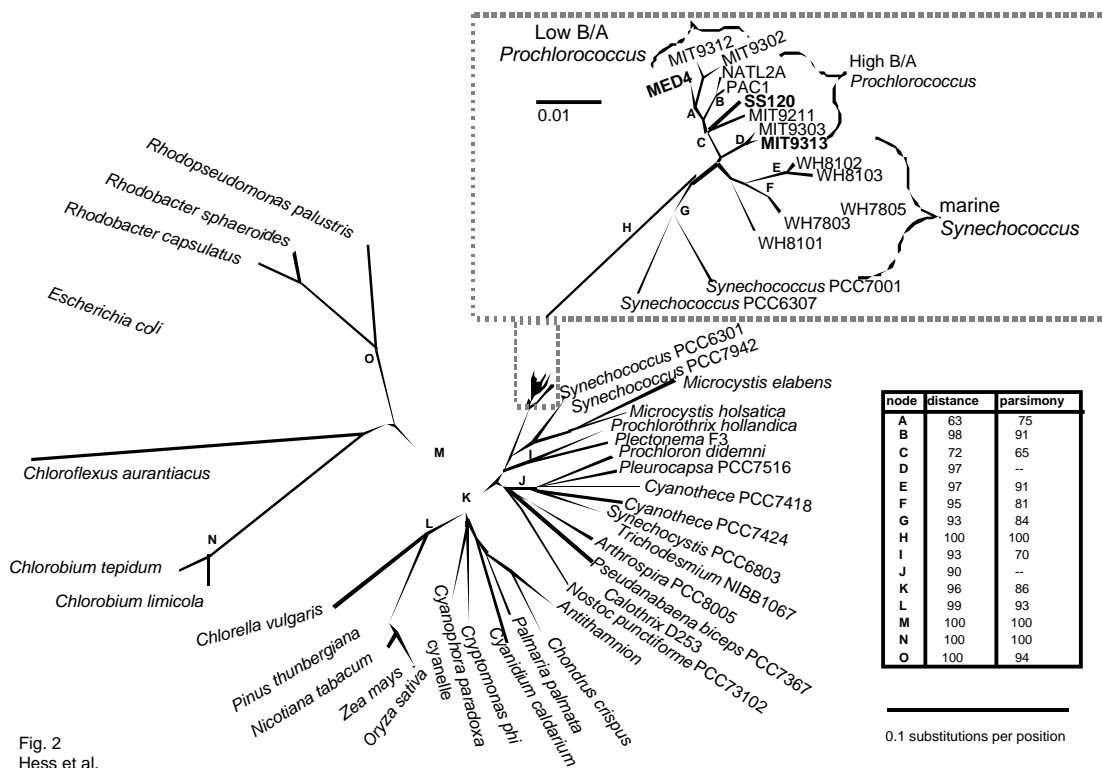


Fig. 2
Hess et al.

Figure 2 Phylogenetic tree based on 16S rDNA sequences showing position of *Prochlorococcus* relative to other member of the cyanobacterial lineage. High and low B/A ecotypes are indicated. The tree was constructed using paralinear distances and minimum evolution as the objective criterion. Bootstrap values represent 1000 resamplings. All phylogenetic analyses are based on 1042 positions common to all sequences and employed PAUP* (Swofford, 2000). Aligned sequences were downloaded from the Ribosomal Database Project (Maidak et al., 2000) except for *Synechococcus* WH8102 which was extracted and downloaded directly from the genome project (http://www.jgi.doe.gov/tempweb/JGI_microbial/html/synechococcus/synech_homepage.html). Genbank accession numbers are as follows: *Prochlorococcus* MED4, AF001466; *Prochlorococcus* MIT9302, AF053396; *Prochlorococcus* MIT9312, AF053398; *Prochlorococcus* NATL2A, AF001467; *Prochlorococcus* PAC1, AF001471; *Prochlorococcus* SS120, X63140; *Prochlorococcus* MIT9211, AF115270; *Prochlorococcus* MIT9303, AF053397; *Prochlorococcus* MIT9313, AF053399; *Synechococcus* WH8103, AF001479; *Synechococcus* WH7805, AF001478; *Synechococcus* WH7803, AF081834; *Synechococcus* WH8101, AF001480; *Synechococcus* PCC6307, AF001477; *Synechococcus* PCC7001, AB015058; *Microcystis holsatica*, U40336; *Microcystis elabens*, U40335; *Synechococcus* PCC6301, X03538; *Synechococcus* PCC7942, D88288; *Plectonema* F3, AF091110; *Cyanothece* PCC7418, AJ000708; *Pleurocapsa* PCC7516, X78681; *Prochloron didemni*, X63141; *Cyanothece* sp. PCC7424, AJ000715; *Synechocystis* PCC6803, D64000; *Calothrix* D253, X99213; *Nostoc punctiforme* PCC73102, AF027655; *Trichodesmium* NIBB1067, X70767; *Arthrospira* PCC8005, X70769; *Pseudanabaena biceps* PCC7367, AF091108; *Antithamnion*, X54299; *Chondrus crispus*, Z29521; *Palmaria palmata*, Z18289; *Cyanidium caldarium*, X52985; *Cryptomonas phi*, S73904; *Cyanophora paradoxa* cyanelle, U30821; *Oryza sativa*, X15901; *Zea mays*, X86563; *Nicotiana tabacum*, U12813; *Pinus*

thunbergiana, D17510; *Chlorella vulgaris*, D11347; *Rhodobacter sphaeroides*, X53853; *Rhodospseudomonas palustris*, D84187; *Rhodobacter capsulatus*, M34129; *Chloroflexus aurantiacus* M34116;

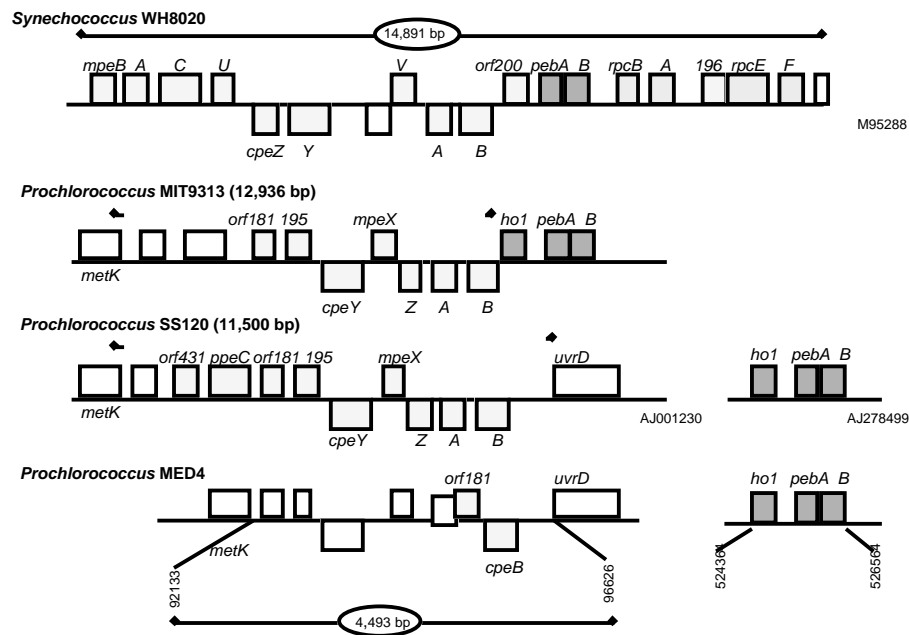


Fig. 3
Hess et al.

Figure 3. Comparison of the phycoerythrin-encoding genome regions in *Prochlorococcus* MIT9313, SS120 and MED4 compared to *Synechococcus* WH8020. This region is framed by two genes not related to phycobiliproteins, *uvrD*, coding for a DNA helicase and *metK*, encoding an S-adenosylmethionine transferase. Other gene names are shown where relevant. For each region the distance between the two black squares is given in bp. For the SS120 and *Synechococcus* sequences the GenBank accession numbers are shown. PE genes and associated reading frames are labelled by dotted boxes, phycocyanin genes by horizontal lines within the boxes, and genes involved in the biosynthesis of phycoerythrobilin are light gray. The heme oxygenase gene *ho1* together with *pebA* and *pebB*, which are part of the phycobiliprotein gene cluster in *Synechococcus*, are translocated in MED4 and SS120 about 200,000 bp away.

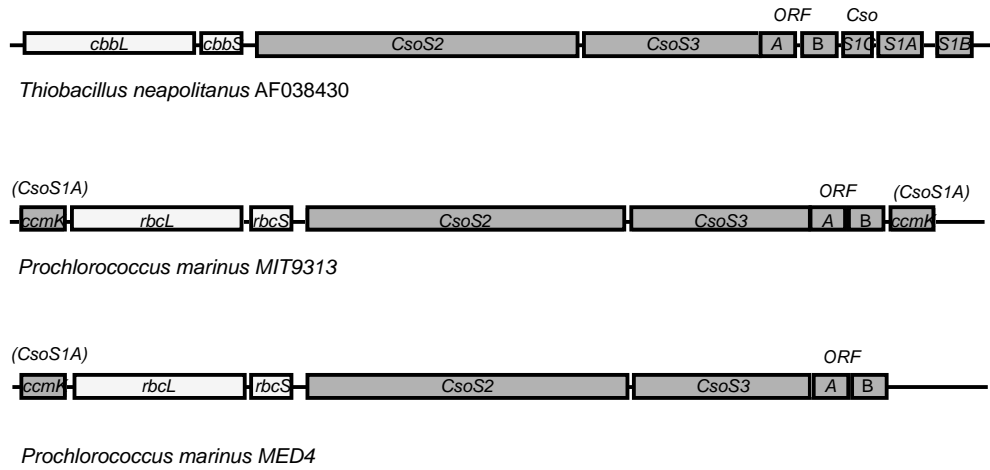


Fig. 4A
Hess et al.

Figure 4. Phylogeny of Rubisco sequences and similarity between the CCM operons of *Prochlorococcus* and *Thiobacillus*. (A) Comparison of operon structure of the Rubisco genes (*rbcL* and *rbcS* in *Prochlorococcus*, *cbbL* and *cbbS* in *Thiobacillus*) and the genes encoding proteins of the carbon concentrating mechanism.

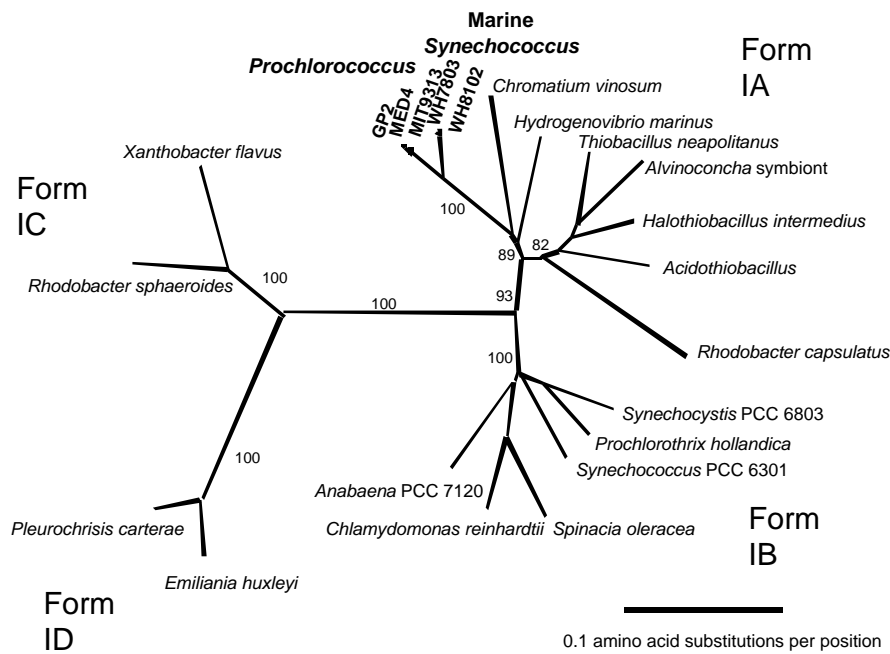


Fig. 4B

(B) Phylogenetic tree of Rubisco large subunit sequences, corresponding to amino acid positions 58-453 of the MED4 *rbcL* gene product. The tree was constructed using PAUP* v4.0b8 using mean distances and minimum evolution as the objective criterion. Bootstrap values represent 100 resamplings. Sequences for *Prochlorococcus* MED4 and MIT9313 and

Synechococcus WH8102 were downloaded from the JGI genome pages (http://www.jgi.doe.gov/tempweb/JGI_microbial/html). Other GenBank accession numbers are as follows: *Xanthobacter flavus*, CAA35115; *Rhodobacter sphaeroides*, P27997; *Pleurochrysis carterae*, BAA39175; *Emiliania huxleyi*, BAA08279; *Anabaena* PCC 7120, P00879; *Chlamydomonas reinhardtii*, AAA84449; *Spinacia oleracea*, CAA23473; *Synechococcus* PCC 6301, P00880; *Prochlorothrix hollandica*, P27568; *Synechocystis* PCC 6803, P54205; *Rhodobacter capsulatus*, O32740; *Acidithiobacillus ferrooxidans*, AAD30508; *Halothiobacillus intermedius*, AAD02445; *Alviniconcha hessleri* symbiont, P24672; *Thiobacillus neopolitanus*, AAC32549; *Hydrogenovibrio marinus*, BAA07731; *Chromatium vinosum*, P22859; *Prochlorococcus* GP2, BAA04861; *Synechococcus* WH7803, P96486.

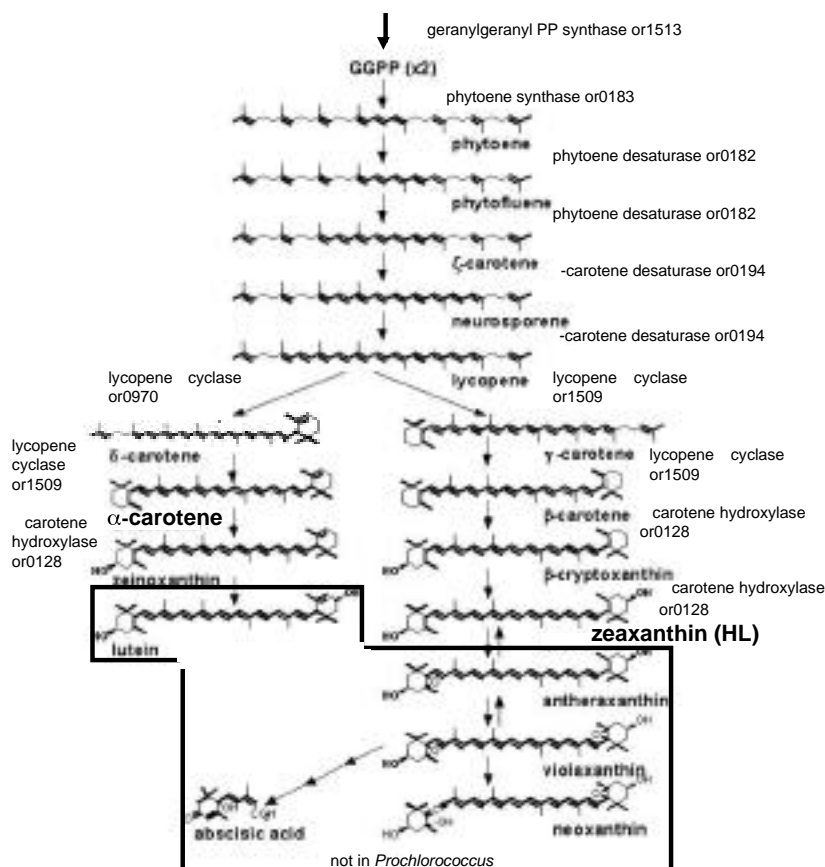


Fig. 5
Hess et al.

Figure 5. Comparison of carotenoid biosynthesis in plants and in *Prochlorococcus* MED4. The terminal products of this pathway in *Prochlorococcus* are shown in boldface letters, the higher plant carotenoids missing in *Prochlorococcus* are boxed. The designation of putative

enzymes is according to (Cunningham and Gantt, 1998) followed by the identifier of the respective *Prochlorococcus* MED4 gene. There are candidate genes for all required enzymatic activities except for the synthesis of parasiloxanthin, which indeed has not been found in cultures to date. Two different putative lycopene cyclases give rise to the - and - carotene branch, respectively.

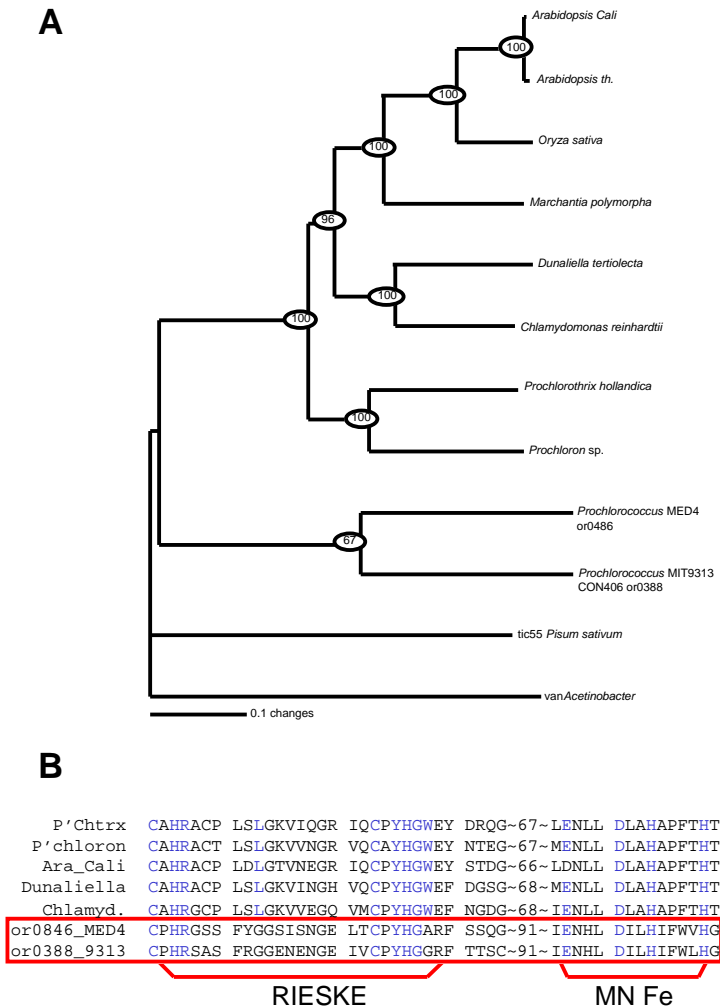


Fig. 6
Hess et al.

Figure 6. The candidate gene for a chlorophyll *b* synthetase in *Prochlorococcus*. (A) Phylogenetic tree of identified chlorophyll *b* synthetases from several higher plants, algae and two prochlorophytes compared to the candidate enzymes from *Prochlorococcus* MED4 and MIT9313. The optimality criterion was set to minimum evolution (PAUP 5.0; (Swofford, 2000)). The two proteins deduced from one gene in each of the two strains are located

together at a basal branch of a cluster containing all known chlorophyll *b* synthetases. The support values for each node are shown. The less conserved parts of the compared proteins (amino acid positions 1-230, 305-316, 347-361, 416-418, 446-455, 491-500, 535-704; alignment available upon request) were excluded from this analysis. Two proteins were taken as outgroups: tic55 from *Pisum sativum*, T06499; vanillin oxygenase from *Acetivobacter*, AF009672. ;Other GenBank accession numbers: *Oryza sativa*, BAA82479; *Arabidopsis thaliana* 1, BAA82484; *Chlamydomonas reinhardtii*, BAA33964; *Arabidopsis thaliana* 2, AF177200; *Dunaliella tertiolecta*, BAA82481; *Prochlorothrix hollandica*, BAA82482; *Prochloron* sp., BAA82483; *Marchantia polymorpha*, BAA82480. (B) Alignment of the *Prochlorococcus* candidate enzymes with the most conserved part of identified chlorophyll *b* synthetases (Tanaka et al., 1998). Critical residues are highlighted that participate in mononuclear iron binding (MN Fe) or binding a [2Fe-2S] Rieske center (RIESKE). The numbers within the sequences refer to the distance in amino acids between the two motifs.

Table 1. Summary of data about various genes discussed in the text

	protein/function	<i>Prochlorococcus</i>		% aa identity	other info
		MED4	MIT9313		
Antenna complex					
<i>pcb1</i>	antenna protein	or1216	or0186	69	pcb1 and pcb2 of MIT9313 are 59% identical
<i>pcb2</i>	antenna protein	----	or0393	54	
Phycobiliproteins					
<i>cpeA</i>	phycoerythrin	----	or0323	----	
<i>cpeB</i>	phycoerythrin	or2053	or0324	35	
Photosystem II					
<i>psbA1</i>	PSII D1	or2010	or1026	87	psbA1 and psbA1 of MIT9313 are 100% identical
<i>psbA2</i>	PSII D1	---	or0188	87	
<i>psbD</i>	PSII D2	or1562	or0217		
Pigment Synthesis					
	lycopene cyclase	or1509		55/43	putative; enzyme specificities to be determined experimentally
	lycopene cyclase	or970		42/58	
Carbon fixation					
<i>cbbP</i>	phosphoribulokinase	or0867		88	horizontal gene transfer from proteobacteria?
<i>rbcL</i>	Rubisco, large subunit	or1173		98	horizontal gene transfer from proteobacteria?
<i>rbcS</i>	Rubisco, small subunit	or1174		93	horizontal gene transfer from proteobacteria?

References

- Adamska I, Ohad I and Kloppstech K (1992) Synthesis of the early light-inducible protein is controlled by blue light and related to light stress. *Proc Natl Acad Sci USA* 89: 2610-2613
- Apt KE, Collier JL and Grossman AR (1995) Evolution of the phycobiliproteins. *J Mol Biol* 248: 79-96
- Armstrong GA (1994) Eubacteria show their true colors: genetics of carotenoid pigment biosynthesis from microbes to plants. *J Bacteriol* 176: 4795-802
- Bouyoub A, Vernotte C and Astier C (1993) Functional analysis of the two homologous *psbA* gene copies in *Synechocystis* PCC 6714 and PCC 6803. *Plant Mol Biol* 21: 149-258
- Burger-Wiersma T, Veenhuis M, Korthals HJ, Van de Wiel CCM and Mur LR (1986) A new prokaryote containing chlorophylls a and b. *Nature* 320: 262-264
- Bustos SA and Golden SS (1992) Light-regulated expression of the *psbD* gene family in *Synechococcus* sp. strain PCC 7942: evidence for the role of duplicated *psbD* genes in cyanobacteria. *Mol Gen Genet* 232: 221-230
- Chisholm SW, Frankel SL, Goericke R, Olson RJ, Palenik B, Waterbury JB, West-Johnsrud L and Zettler ER (1992) *Prochlorococcus marinus* nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll a and b. *Arch Microbiol* 157: 297-300
- Chisholm SW, Olson RJ, Zettler ER, Waterbury J, Goericke R and Welschmeyer N (1988) A novel free-living prochlorophyte occurs at high cell concentrations in the oceanic euphotic zone. *Nature* 334: 340-343
- Collier JL and Grossman AR (1994) A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria. *EMBO J.* 13: 1039-1047

- Colon-Lopez MS and Sherman LA (1998) Transcriptional and translational regulation of photosystem I and II genes in light-dark- and continuous-light-grown cultures of the unicellular cyanobacterium *Cyanothece* sp. strain ATCC 51142. *J Bacteriol* 180: 519-526
- Cornejo J, Willows RD and Beale SI (1998) Phytobilin biosynthesis: cloning and expression of a gene encoding soluble ferredoxin-dependent heme oxygenase from *Synechocystis* sp. PCC 6803. *Plant J* 15: 99-107
- Cunningham FX and Gantt E (1998) Genes and enzymes of carotenoid biosynthesis in plants. *Annu Rev Plant Physiol Plant Mol Biol* 49: 557-583
- Cunningham FX, Pogson B, Sun Z, McDonald KA, DellaPenna D and Gantt E (1996) Functional analysis of the beta and epsilon lycopene cyclase enzymes of *Arabidopsis* reveals a mechanism for control of cyclic carotenoid formation. *Plant Cell* 8: 1613-1626
- Cunningham FX, Sun Z, Chamovitz D, Hirschberg J and Gantt E (1994) Molecular structure and enzymatic function of lycopene cyclase from the cyanobacterium *Synechococcus* sp strain PCC7942. *Plant Cell* 6: 1107-1121
- Delwiche CF and Palmer JD (1996) Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. *Mol Biol Evol* 13: 873-882.
- Dolganov NA, Bhaya D and Grossman AR (1995) Cyanobacterial protein with similarity to the chlorophyll a/b binding proteins of higher plants: evolution and regulation. *Proc Natl Acad Sci USA* 92: 636-640
- Frankenberg N, Mukougawa K, Kohchi T and Lagarias JC (2001) Functional Genomic Analysis of the HY2 Family of Ferredoxin-Dependent Bilin Reductases from Oxygenic Photosynthetic Organisms. *Plant Cell*, in press.
- Garczarek L, Hess WR, Holtzendorff J, van der Staay GWM and Partensky F (2000) Multiplication of antenna genes as a major adaptation to low light in a marine prokaryote. *Proc Nat Acad Sci USA* 97: 4098-4101
- Garczarek L, Partensky F, Holtzendorff J, Babin M, Mary I, Thomas JC and Hess WR (2001) Differential expression of antenna and core genes in the marine oxychlorobacterium *Prochlorococcus* PCC 9511 grown under light-dark cycles. *Environ Microbiol* 3: 168-175
- Gingrich JC, Buzby JS, Stirewalt VL and Bryant DA (1988) Genetic analysis of two new mutations resulting in herbicide resistance in the cyanobacterium *Synechococcus* sp. PCC 7002. *Photosynth Res* 16: 83-99
- Gingrich JC, Gasparich GE, Saver K and Bryant DA (1990) Nucleotide sequence and expression of the two genes encoding D2 protein and the single gene encoding the CP43 protein of Photosystem II in the cyanobacterium *Synechococcus* sp. PCC 7002. *Photosynthesis Research* 24: 137-150
- Glass JI, Lefkowitz EJ, Glass JS, Heiner CR, Chen EY and Cassell GH (2000) The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. *Nature* 407: 757-762
- Goericke R, Olson RJ and Shalapyonok A (2000) A novel niche for *Prochlorococcus* sp. in low-light suboxic environments in the Arabian Sea and the eastern tropical North Pacific. *Deep-Sea Res I* 47: 1183-1205
- Goericke R and Repeta DJ (1992) The pigments of *Prochlorococcus marinus*: the presence of divinyl chlorophyll *a* and *b* in a marine prochlorophyte. *Limnol Oceanogr* 37: 425-433
- Goericke R and Welschmeyer NA (1993) The Marine Prochlorophyte *Prochlorococcus* contributes significantly to phytoplankton biomass and primary production in the Sargasso Sea. *Deep Sea Research* 40: 2283-2294
- Golden SS, Brusslan J and Haselkorn R (1986) Expression of a family of *psbA* genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. *EMBO J* 5: 2789-2798

- Golden SS and Stearns GW (1988) Nucleotide sequence and transcript analysis of three photosystem II genes from the cyanobacterium *Synechococcus* sp. PCC7942. *Gene* 67: 85-96
- Gotz T, Windhovel U, Boger P and Sandmann G (1999) Protection of photosynthesis against ultraviolet-B radiation by carotenoids in transformants of the cyanobacterium *Synechococcus* PCC7942. *Plant Physiol* 120: 599-604
- Green BR and Kuhlbrandt W (1995) Sequence conservation of light-harvesting and stress-response proteins in relation to the three-dimensional molecular structure of LHCII. *Photosyn Res* 44: 139-148
- Green BR and Pichersky E (1994) Hypothesis for the evolution of three-helix Chl a/b and Chl a/c light-harvesting antenna proteins from two-helix and four-helix ancestors. *Photosyn Res* 39: 149-162
- Grimm B and Kloppstech K (1987) The early light-inducible proteins of barley. Characterization of two families of 2-h-specific nuclear-coded chloroplast proteins. *Eur J Biochem* 167: 493-499
- He Q, Dolganov N, Bjorkman O and Grossman AR (2001) The high light-inducible polypeptides in *Synechocystis* PCC6803. Expression and function in high light. *J Biol Chem* 276: 306-314
- Heddad M and Adamska I (2000) Light stress-regulated two-helix proteins in *Arabidopsis thaliana* related to the chlorophyll a/b-binding gene family. *Proc Natl Acad Sci USA* 97: 3741-3746
- Hess WR (1997) Localization of an open reading frame with homology to human aspartoacylase upstream from *psbA* in the prokaryote *Prochlorococcus marinus* CCMP 1375. *DNA Seq* 7: 301-306
- Hess WR and Partensky F (1999) Analysis of a phycobiliprotein gene cluster in *Prochlorococcus marinus* CCMP 1375: Identification of a putative linker polypeptide and the phylogeny of *Prochlorococcus* phycoerythrins (Peschek, G. A., Löffelhardt, W. & G., S., eds) *The Photosynthetic Prokaryotes*, pp. 751-761. Plenum Press, New York
- Hess WR, Partensky F, van der Staay GW, Garcia-Fernandez JM, Börner T and Vaultot D (1996) Coexistence of phycoerythrin and a chlorophyll a/b antenna in a marine prokaryote. *Proc Natl Acad Sci USA* 93: 11126-11130
- Hess WR, Steglich C, Lichtlé C and Partensky F (1999) Phycoerythrins of the oxyphotobacterium *Prochlorococcus marinus* are associated to the thylakoid membranes and are encoded by a single large gene cluster. *Plant Mol Biol* 40: 507-521
- Hess WR, Weihe A, Loiseaux-de Goer S, Partensky F and Vaultot D (1995) Characterization of the single *psbA* gene of *Prochlorococcus marinus* CCMP 1375 (Prochlorophyta). *Plant Mol Biol* 27: 1189-1196
- Holthuijzen YA, van Breeman JFL, Konings WN and van Bruggen EFJ (1986a) Electron microscopic studies of carboxysomes in *Thiobacillus neapolitanus*. *Arch Microbiol* 144: 258-262
- Holthuijzen YA, van Breeman JFL, Kuenen JG and Konings WN (1986b) Protein composition of the of carboxysomes of *Thiobacillus neapolitanus*. *Arch Microbiol* 144: 398-404
- Jansson S, Andersson J, Kim SJ and Jackowski G (2000) An *Arabidopsis thaliana* protein homologous to cyanobacterial high-light- inducible proteins. *Plant Mol Biol* 42: 345-351
- Johnson Z, Landry ML, Bidigare RR, Brown SL, Campbell L, Gunderson J, Marra J and Trees C (1999) Energetics and growth kinetics of a deep *Prochlorococcus* spp. population in the Arabian Sea. *Deep-Sea Res II* 46: 1719-1743

- Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirose M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M and Tabata S (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 (supplement). *DNA Res* 3: 185-209
- Kaplan A and Reinhold L (1999) CO₂ concentrating mechanisms in photosynthetic microorganisms. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50: 539-570
- Kehoe DM and Grossman AR (1996) Similarity of a chromatic adaptation sensor to phytochrome and ethylene receptors. *Science* 273: 1409-1412
- Klughammer B, Sultemeyer D, Badger MR and Price GD (1999) The involvement of NAD(P)H dehydrogenase subunits, NdhD3 and NdhF3, in high-affinity CO₂ uptake in *Synechococcus* sp. PCC7002 gives evidence for multiple NDH-1 complexes with specific roles in cyanobacteria. *Mol Microbiol* 32: 1305-1315
- Krubasik P and Sandmann G (2000) A carotenogenic gene cluster from *Brevibacterium linens* with novel lycopene cyclase genes involved in the synthesis of aromatic carotenoids. *Mol Gen Genet* 263: 423-432
- La Roche J, van der Staay GW, Partensky F, Ducret A, Aebersold R, Li R, Golden SS, Hiller RG, Wrench PM, Larkum AW and 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
- Lewin RA (1976) Prochlorophyta as a proposed new division of algae. *Nature* 261: 697-698
- Lewin RA and Withers NW (1975) Extraordinary pigment composition of a prokaryotic alga. *Nature* 256: 735-737
- Li WKW (1995) Composition of ultraphytoplankton in the central North Atlantic. *Marine Ecology Progress Series* 122: 1-8
- Lindahl M, Funk C, Webster J, Bingsmark S, Adamska I and Andersson B (1997) Expression of ELIPs and PSII-S protein in spinach during acclimative reduction of the Photosystem II antenna in response to increased light intensities. *Photosyn Res* 54: 227-236
- Linden H, Misawa N, Saito T and Sandmann G (1994) A novel carotenoid biosynthesis gene coding for zeta-carotene desaturase: functional expression, sequence and phylogenetic origin. *Plant Mol Biol* 24: 369-379
- Liu H, Nolla HA and Campbell L (1997) *Prochlorococcus* growth rate and contribution to primary production in the Equatorial and Subtropical North Pacific Ocean. *Aquat Microb Ecol* 12: 39-47
- Lockhardt PJ, Penny D, Hendy MD and Larkum ADW (1993) Is *Prochlorothrix hollandica* the best choice as a prokaryotic model for higher plant Chl a/b photosynthesis? *Photosynth Res* 37: 61-68
- Lokstein H, Steglich C and Hess WR (1999) Light-harvesting antenna function of phycoerythrin in *Prochlorococcus marinus*. *Biochim Biophys Acta* 1410: 97-98
- Maid U, Valentin K and Zetsche K (1990) The *psbA*-gene from a red alga resembles those from Cyanobacteria and Cyanelles. *Curr Genet* 17: 255-259
- Maidak BL, Cole JR, Lilburn TG, Parker CT, Saxman PR, Stredwick JM, Garrity GM, Li BO, G. J., Pramanik S, Schmidt TM and Tiedje JM (2000) The RDP (Ribosomal Database Project) continues. *Nucleic Acids Research* 28: 173-174
- Mann EL (2000) Trace Metals and the Ecology of Marine Cyanobacteria. Ph.D. Thesis, MIT and WHOI
- Meyer G and Kloppstech K (1984) A rapidly light-induced chloroplast protein with a high turnover coded for by pea nuclear DNA. *Eur J Biochem* 138: 201-207

- Moore LR and Chisholm SW (1999) Photophysiology of the Marine Cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnology and Oceanography* 44: 628-638
- Moore LR, Goericke R and Chisholm SW (1995) Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. *Mar Ecol Progr Ser* 116: 259-275
- Moore LR, Rocap G and Chisholm SW (1998) Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* 393: 464-467
- Morden CW and Golden SS (1989) *psbA* genes indicate common ancestry of prochlorophytes and chloroplasts. *Nature* 337: 382-385
- Morel A, Ahn YW, Partensky F, Vaultot D and Claustre H (1993) *Prochlorococcus* and *Synechococcus*: a comparative study of their size, pigmentation and related optical properties. *J Mar Res* 51: 617-649
- Niyogi KK, Bjorkman O and Grossman AR (1997) The roles of specific xanthophylls in photoprotection. *Proc Natl Acad Sci USA* 94: 14162-14167
- Niyogi KK, Grossman AR and Bjorkman O (1998) *Arabidopsis* mutants define a central role for the xanthophyll cycle in the regulation of photosynthetic energy conversion. *Plant Cell* 10: 1121-1134
- Omata T, Price GD, Badger MR, Okamura M, Gohta S and Ogawa T (1999) Identification of an ATP-binding cassette transporter involved in bicarbonate uptake in the cyanobacterium *Synechococcus* sp. strain PCC 7942. *Proc Natl Acad Sci USA* 96: 13571-13576
- Ong LJ and Glazer AN (1991) Phycoerythrins of marine unicellular cyanobacteria. I. Bilin types and locations and energy transfer pathways in *Synechococcus* spp. phycoerythrins. *J Biol Chem* 266: 9515-27
- Öquist G, Campbell D, Clarke AK and Gustafsson P (1995) The cyanobacterium *Synechococcus* modulates Photosystem II function in response to excitation stress through D1 exchange. *Photosynth Res* 46: 151-158
- Oster U, Tanaka R, Tanaka A and Rudiger W (2000) Cloning and functional expression of the gene encoding the key enzyme for chlorophyll b biosynthesis (CAO) from *Arabidopsis thaliana*. *Plant J* 21: 305-310
- Palenik B and Haselkorn R (1992) Multiple evolutionary origins of prochlorophytes, the chlorophyll b- containing prokaryotes. *Nature* 355: 265-267
- Partensky F, Blanchot J and Vaultot D (1999a) Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review (Charpy, L. & Larkum, A. W. D., eds) *Marine Cyanobacteria*. Vol. 19, pp. 457-475, Monaco
- Partensky F, Hess WR and Vaultot D (1999b) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* 63: 106-127
- Penno S, Campbell L and Hess WR (2000) Presence of phycoerythrin in two strains of *Prochlorococcus* isolated from the sub-tropical North Pacific Ocean. *J Phycol* 36: 723-729
- Pichard SL, Campbell L and Paul JH (1997) Diversity of the ribulose biphosphate carboxylase/oxygenase form I gene (*rbcL*) in natural phytoplankton communities. *Appl Environ Microbiol* 63: 3600-3606
- Potter E and Klopstech K (1993) Effects of light stress on the expression of early light-inducible proteins in barley. *Eur J Biochem* 214: 779-786
- Price GD, Coleman JR and Badger MR (1992) Association of carbonic anhydrase with carboxysomes isolated from the cyanobacterium *Synechococcus* PCC7942. *Plant Physiol* 100: 784-793
- Raven JA (1997a) Putting the C in Phycology. *European Journal of Phycology* 32: 319-333

- Raven JA (1997b) The role of marine biota in the evolution of terrestrial biota: Gases and genes. *Biogeochemistry* 39: 139-164
- Reith ME and Munholland J (1995) Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Mol Biol Rep* 13: 333-335
- Richaud C and Zabulon G (1997) The heme oxygenase gene (*pbsA*) in the red alga *Rhodella violacea* is discontinuous and transcriptionally activated during iron limitation. *Proc Natl Acad Sci USA* 94: 11736-11741
- Rippka R, Coursin T, Hess WR, Lichtlé C, Scanlan DJ, Palinska K, Iteman I, Partensky F, Houmard J and Herdman M (2000) *Prochlorococcus marinus* Chisholm et al. 1992, subsp. nov. *pastoris*, strain PCC 9511, the first axenic chlorophyll a2/b2-containing cyanobacterium (Oxyphotobacteria). *Internat J Syst Env Microbiol* 50: 1833-1847
- Rocap G (2000) Genetic diversity and ecotypic differentiation in the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. Ph. D. Thesis, MIT and WHOI
- Rocap G, Moore LR and Chishom SW (1999) Molecular Phylogeny of *Prochlorococcus* Ecotypes (Charpy, L. & Larkum, A. W. D., eds) *Marine Cyanobacteria*, pp. 107-116. Bulletin de l'Institut Océanographique, Monaco, special issue N° 19 (632 pages)
- Sandmann G (1994) Carotenoid biosynthesis in microorganisms and plants. *Eur J Biochem* 223: 7-24
- Schmitz O, Katayama M, Williams SB, Kondo T and Golden SS (2000) CikA, a bacteriophytochrome that resets the cyanobacterial circadian clock. *Science* 289: 765-768.
- Schyns G, Rippka R, Namane A, Campbell D, Herdman M and Houmard J (1997) *Prochlorothrix hollandica* PCC 9006: Genomic properties of an axenic representative of the chlorophyll a/b-containing oxyphotobacteria. *Res Microbiol* 148: 345-354
- Shimada A, Kanai S and Maruyama T (1995) Partial sequence of ribulose-1,5-bisphosphate carboxylase/oxygenase and the phylogeny of *Prochloron* and *Prochlorococcus* (*Prochlorales*). *J Mol Evol* 40: 671-677
- Shively JM, Bradburne CE, Aldrich HC, Bobik AC and Mehlman JL (1998a) Sequence homologs of the carboxysomal polypeptide CsoS1 of the Thiobacilli are present in cyanobacteria and enteric bacteria that form carboxysomes/polyhedral bodies. *Can. J. Bot.* 76: 906-916
- Shively JM, van Keulen G and Mejer WM (1998b) Something from almost nothing: carbon dioxide fixation in chemoautotrophs. *Annu. Rev. Microbiol.* 52: 191-230
- Smith KS and Ferry JG (2000) Prokaryotic carbonic anhydrases. *FEMS Microbiol Rev* 24: 335-366
- So AKC and Espie GS (1998) Cloning, characterization and expression of carbonic anhydrase from the cyanobacterium *Synechocystis* PCC 6803. *Plant Mol Biol* 37: 205-215
- Soitamo AJ, Zhou G, Clarke AK, Öquist G, Gustafsson P and Aro EM (1996) Overproduction of the D1:2 protein makes *Synechococcus* cells more tolerant to photoinhibition of Photosystem II. *Plant Mol Biol* 30: 467-478
- Stirewalt VL, Michalowski CB, Löffelhardt W, Bohnert HJ and Bryant DA (1995) Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. *Plant Mol Biol Rep* 13: 327-332
- Strehl B, Holtzendorff J, Partensky F and Hess WR (1999) A small and compact genome in the marine cyanobacterium *Prochlorococcus marinus* CCMP 1375: lack of an intron in the gene for tRNA(Leu)UAA and a single copy of the rRNA operon. *FEMS Microbiol. Lett.* 181: 261-266
- Swofford DL. (2000) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts
- Tabita FR (1999) Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective. *Photosynthesis Research* 60: 1-28

- Tanaka A, Ito H, Tanaka R, Tanaka NK, Yoshida K and Okada K (1998) Chlorophyll a oxygenase (CAO) is involved in chlorophyll b formation from chlorophyll a. *Proc Natl Acad Sci USA* 95: 12719-12723
- Ting C, Rocap G, King J and Chisholm SW (1999) Characterization of phycoerythrin genes in the chlorophyll a_2/a_2 -containing prokaryote, *Prochlorococcus* sp. MIT9303 (Garab, G., ed) *Photosynthesis: Mechanisms and Effects*. Vol. I, pp. 225-228. Kluwer Academic Publishers, Dordrecht, The Netherlands
- Ting CS, Rocap G, Chisholm SW and King J (submitted) Phycobiliprotein genes of the marine photosynthetic prokaryote *Prochlorococcus*: evidence for rapid evolution of genetic heterogeneity.
- Tomitani A, Okada K, Miyashita H, Matthijs HC, Ohno T and Tanaka A (1999) Chlorophyll b and phycobilins in the common ancestor of cyanobacteria and chloroplasts. *Nature* 400: 159-162
- Tortell PD (2000) Evolutionary and ecological perspectives on carbon acquisition in phytoplankton. *Limnology and Oceanography* 45: 744-750
- Urbach E, Robertson DL and Chisholm SW (1992) Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. *Nature* 355: 267-270
- Urbach E, Scanlan DJ, Distel DL, Waterbury JB and Chisholm SW (1998) Rapid diversification of marine picophytoplankton with dissimilar light-harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (Cyanobacteria). *J Mol Evol* 46: 188-201
- Veldhuis MJW, Kraay GW, Van Bleijswijk JDL and Baars MA (1997) Seasonal and spatial variation in phytoplankton biomass, productivity and growth in the northwestern Indian Ocean: the southwest and northeast monsoon, 1992-1993. *Deep Sea Research* 44: 425-449
- Vierstra RD and Davis SJ (2000) Bacteriophytochromes: new tools for understanding phytochrome signal transduction. *Semin Cell Dev Biol* 11: 511-521.
- Viveiros M, Krubasik P, Sandmann G and Houssaini-Iraqi M (2000) Structural and functional analysis of the gene cluster encoding carotenoid biosynthesis in *Mycobacterium aurum* A+. *FEMS Microbiol Lett* 187: 95-101
- Vrba JM and Curtis SE (1989) Characterization of a four-member psbA gene family from the cyanobacterium *Anabaena* PCC 7120. *Plant Mol Biol* 14: 81-92
- Watson GM and Tabita FR (1996) Regulation, unique gene organization, and unusual primary structure of carbon fixation genes from a marine phycoerythrin-containing cyanobacterium. *Plant Mol Biol* 32: 1103-1115
- Wilbanks SM and Glazer AN (1993) Rod structure of a phycoerythrin II-containing phycobilisome. I. Organization and sequence of the gene cluster encoding the major phycobiliprotein rod components in the genome of marine *Synechococcus* sp. WH8020. *J Biol Chem* 268: 1226-1235
- Williams JGK and Chisholm DA (1987) Nucleotide sequences of both *psbD* genes from the cyanobacterium *Synechocystis* PCC 6803 (Biggins, J., ed) *Progress in Photosynthesis Research*, pp. 809-812. Martinus Nijhoff, Dordrecht
- Willows RD, Mayer SM, Foulk MS, DeLong A, Hanson K, Chory J and Beale SI (2000) Phytobilin biosynthesis: the *Synechocystis* sp. PCC 6803 heme oxygenase-encoding *ho1* gene complements a phytochrome-deficient *Arabidopsis thaliana* *hyl* mutant. *Plant Mol Biol* 113-120
- Yoshihara S, Suzuki F, Fujita H, Geng XX and Ikeuchi M (2000) Novel putative photoreceptor and regulatory genes required for the positive phototactic movement of the unicellular motile cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol* 41: 1299-1304