Evolution and diversity of CO₂ concentrating mechanisms in cyanobacteria

Murray R. Badger, David Hanson and G. Dean Price

Molecular Plant Physiology Group, Research School of Biological Sciences, Australian National University, PO Box 475, Canberra, ACT 2601, Australia.

Abstract. Cyanobacteria have developed an effective photosynthetic CO₂ concentrating mechanism (CCM) for improving the efficiency of carboxylation by a relatively inefficient Rubisco. The development of this CCM was presumably in response to the decline in atmospheric CO₂ levels and rising O₂, both of which were triggered by the development of oxygenic photosynthesis by cyanobacteria themselves. In the past few years there has been a rapid expansion in our understanding of the mechanism and genes responsible for the CCM. In addition, there has been a recent expansion in the availability of complete cyanobacterial genomes, thus increasing our potential to examine questions regarding both the evolution and diversity of components of the CCM across cyanobacteria. This paper considers various CCM and photosynthesis gene components across eight cyanobacteria where significant genomic information is available. Significant conclusions from our analysis of the distribution of various genes indicated the following. Firstly, cyanobacteria have developed with two types of carboxysomes, and this is correlated with the form of Rubisco present. We have coined the terms α-cyanobacteria to refer to cyanobacteria containing Form 1A Rubisco and α-carboxysomes, and β-cyanobacteria having Form 1B Rubisco and β-carboxysomes. Secondly, there are two NDH-1 CO₂ uptake systems distributed variably, with Prochlorococcus marinus species appearing to lack this CO₂ uptake system. There are at least two HCO₃⁻ transport systems distributed variably, with some α-cyanobacteria having an absence of systems identified in β-cyanobacteria. Finally, there are multiple forms of carbonic anhydrases (CAs), but with only β-carboxysomal CA having a clearly shown role at present. The α-cyanobacteria appear to lack a clearly identifiable carboxysomal CA. A pathway for the evolution of cyanobacterial CCMs is proposed. The acquisition of carboxysomes triggered by the rapid decline of atmospheric CO₂ in the Phanerozoic is argued to be the initial step. This would then be followed by the development of NDH-1 CO₂-uptake systems, followed by the development of low- and high-affinity HCO₃⁻ transporters. An intriguing question is, were carboxysomes developed first in cyanobacteria, or did they originate by the lateral transfer of pre-existing proteobacterial bacterial microcompartment genes? The potentially late evolution of the CCM genes in cyanobacteria argues for a polyphyletic and separate evolution of CCMs in cyanobacteria, algae, and higher plants.

Introduction

Cyanobacteria have developed an effective photosynthetic CCM for improving carboxylation by their relatively inefficient Rubiscos (Badger and Price 1992; Price et al. 1998; Kaplan and Reinhold 1999). The development of this CCM was presumably in response to the decline in atmospheric CO₂ levels and rising O₂, both of which were triggered by the development of oxygenic photosynthesis by cyanobacteria themselves. In the past few years there has been a rapid expansion in our understanding of the mechanism and genes involved in cyanobacterial CCMs. In addition, there has been a recent expansion in the availability of complete cyanobacterial genomes, thus increasing our potential to examine questions regarding both the evolution and diversity of components of the CCM across cyanobacterial species. This paper uses the existing cyanobacterial genome information to ask questions about CCM evolution and diversity.

A current view of the CCM

Figure 1 shows a schematic diagram of a cyanobacterial cell, with components of the CCM (see reviews by Price et al. 1998; Kaplan and Reinhold 1999). Central to the CCM is the carboxysome, which contains the Rubisco of the cell and a carboxysomal CA, which converts cytosolically accumulated HCO₃⁻ into CO₂. CO₂ is subsequently elevated within
the localised domain of the carboxysome by some property of the carboxysomes, perhaps its protein shell, restricting CO$_2$ efflux. To enable HCO$_3^-$ to be accumulated within the cytosol, a number of active CO$_2$ and HCO$_3^-$ transporters are involved, and many of these have been recently discovered (see later sections). These transporters are located on both the plasma membrane and the thylakoid membrane, and exist in both high-affinity and low-affinity transporter forms. The ability of cyanobacterial cells to improve their affinity for inorganic carbon (Ci) when grown at limiting Ci levels is based primarily on the changes that occur in the presence of various affinity Ci transporters associated with the cells.

To assess the diversity and evolution of CCM genes in cyanobacteria, we have focused on the following functional elements:

1. Elements associated with CO$_2$ fixation, which include Rubisco, carboxysome structural elements, CAs, glycolate metabolism, and key photosynthetic carbon reduction (PCR) cycle components.

2. Ci transporters, including HCO$_3^-$ transporters and CO$_2$ uptake components, including NDH-1 subunit components.

**The phylogeny of cyanobacteria**

Previous studies of cyanobacterial phylogeny have divided them into groups based on their photosynthetic pigment composition. Conventional cyanobacteria possess phycobilisomes and lack chlorophyll $b$, while prochlorophytes lack phycobilisomes and contain chlorophyll $b$. However, it is obvious that there have been multiple evolutionary origins of so-called prochlorophytes within widely different cyanobacterial taxonomic groups (Urbach et al. 1992). However, for the purposes of examining the evolution and diversity of cyanobacterial CCMs, recent more extensive genomic studies strongly suggest that a division of cyanobacterial species based on their type of Rubisco may be much more appropriate (see Fig. 2).

A phylogenetic tree for Rubisco from various photosynthetic bacteria is shown in Fig. 2. The four groups of photosynthetic bacteria shown are grouped according to their Form 1 ($L_8S_8$) Rubisco types, originally described by Delwiche and colleagues (Delwiche 1999) with Form 1A, B, C and D groups shown. What is obvious is that cyanobacterial species are contained within both the Form 1A and 1B domains, as initially noted by Tabita and colleagues (Tabita 1999). This divergence of cyanobacteria is associated with the division between $\beta$-proteobacterial-like cyanobacteria such as *P. marinus* and *Synechococcus* WH8102, and blue-green cyanobacteria such as *Synechocystis* PCC6803. For the purposes of this review, we will refer to the two primary cyanobacterial groups as being either $\alpha$-cyanobacteria with Form 1A Rubisco or $\beta$-cyanobacteria with Form 1B Rubisco.

Also indicated on Fig. 1 is the presence of carboxysomes in various photosynthetic and chemautotrophic bacterial species. Carboxysomes are protein bodies, which are surrounded by a protein shell and contain the Rubisco. Hence, there is potential significance between the nature of the carboxysome structure and the type of Rubisco that it contains. The carboxysome story is interesting and is elaborated upon in a later section. However, since the sequencing of the $\alpha$-cyanobacterial genomes, it has become apparent that $\alpha$-cyanobacteria possess carboxysomes that are significantly different from the carboxysomes found in $\beta$-cyanobacteria. For the purposes of this review and future nomenclature, we devise here that $\alpha$-carboxysomes are found in Form 1A photosynthetic bacteria, including $\alpha$-cyanobacteria and other proteobacteria such as *Thiobacillus* species (Shively et al. 1998a, b) while $\beta$-carboxysomes are associated with Form 1B Rubisco in $\beta$-cyanobacteria (Price et al. 1998). Form 1A Rubiscos are clearly found in chemautotrophic bacteria both with and without carboxysomes, so that carboxysomes are apparently an optional adaptation to enabling more
efficient carbon fixation in chemosynthetic bacteria. In this regard, little is known about the differences in the kinetic properties of Rubisco between these organisms, or the potential ecological advantages that the possession of carboxysomes might confer. However, all cyanobacteria characterised to date have carboxysomes.

Cyanobacterial genome sequencing

Over the past two years, there has been a rapid increase of complete bacterial genome sequences, and analysis is providing insights into many issues concerning microbial evolution and function. In this paper, we have made use of genetic information available for the species indicated in Table 1, with links and information sources acknowledged in the legend. In addition to this we have also utilised sequences available for other bacteria indicated in Fig. 2, as a means of sequence comparison for phylogenetic relationships for various gene components.

Carboxysome components

Table 1 shows genes that have been identified as being associated with carboxysomes. Listed in the table are components for both α- and β-carboxysomes, with β-carboxysomes characterised in β-cyanobacteria such as *Synechococcus* PCC7942 (Price *et al.* 1993) and *Synechococcus* PCC7002 (Ludwig *et al.* 2000), while α-carboxysomes were originally identified in *Thiobacillus*-proteobacteria species (Shively *et al.* 1973a, b). Of the genes listed, *ccmK*, *ccmL*, *ccmO*, *csoS1*, peptide A and peptide B, are all related to each other by coding for proteins containing one or more regions of homology to bacterial microcompartment domains. CcmK, CcmO and CsoS1 proteins form one homologous grouping within this family, while CcmL, peptide A and peptide B form another. These conserved structural domains have been identified by comparison of CcmK, CcmL or CsoS1-like proteins involved in carboxysome formation in α- and β-cyanobacteria and β-proteobacteria (Price *et al.* 1998; Shively *et al.* 1999b), as well as more recently discovered genes associated with enteric proteobacteria containing carboxysome-like microcompartments specialised in both propanediol and ethanolamine metabolism and detoxification (Bobik *et al.* 1999; Kofoi *et al.* 1999).

The *ccmM* and *ccmN* genes are specific to β-cyanobacterial carboxysomes, while *csoS2* and *csoS3* are specific

![Carboxysomes](image-url)
for α-carboxysomes. In general, these genes and their proteins have no functional homologs in other bacterial systems, except to say that CcmM protein has regions within it that are homologous to both γ-CA and to the small subunit of the Form 1B Rubisco protein (Ludwig et al. 2000). The functional significance of these homologies is unknown. The Rubisco activase protein is included, as it has stronger homology to the CcmM protein than to Rubisco activase from both higher plants and *Chlamydomonas* (data not shown), but is only found in *Nostoc* and *Anabaena* species. Rubisco and carboxysomal CA are included as the other functional components of the carboxysome.

The most obvious aspect of this carboxysome gene comparison is that for the first time, it is now apparent that

<table>
<thead>
<tr>
<th>Table 1. CCM related genes in cyanobacterial genomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some data for <em>Synechococcus</em> PCC7002 were kindly provided by D. Bryant (Penn State University, PA, USA), who is currently sequencing the genome. The following resources were used to make gene comparisons: for general information, Genbank protein and nucleotide sequence databases (<a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>); for specific information on finished and unfinished cyanobacterial genomes, PEDANT (<a href="http://pedant.gsf.de/index.html">http://pedant.gsf.de/index.html</a>), DOE Microbial Genomics (<a href="http://www.jgi.doe.gov/tempweb/JGI_microbial/html/index.html">http://www.jgi.doe.gov/tempweb/JGI_microbial/html/index.html</a>); for <em>Synechocystis</em> and <em>Anabaena</em>, Cyanobase (<a href="http://www.kazusa.or.jp/cyano/cyano.html">http://www.kazusa.or.jp/cyano/cyano.html</a>). References are not given to each gene, however, the presence or absence of each gene was assessed by its protein homology to reference sequences for each gene. The protein references used are given. SLL and SLR numbers are from Cyanobase references for <em>Synechocystis</em> PCC6803. —, genes not currently found, but genome not completely sequenced; n.f., genes not found in mostly completed genomes; *, species for which mostly completed genomes are currently available; ?, possible presence</td>
</tr>
<tr>
<td><strong>Gene</strong></td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td><strong>Synecochytris</strong> PCC6803*</td>
</tr>
<tr>
<td><strong>β-Carboxysomes</strong></td>
</tr>
<tr>
<td>ccmK</td>
</tr>
<tr>
<td>ccmL</td>
</tr>
<tr>
<td>ccmM</td>
</tr>
<tr>
<td>rbcL activase</td>
</tr>
<tr>
<td>ccmN</td>
</tr>
<tr>
<td><strong>α-Carboxysomes</strong></td>
</tr>
<tr>
<td>csoSI</td>
</tr>
<tr>
<td>peptide A</td>
</tr>
<tr>
<td>peptide B</td>
</tr>
<tr>
<td>csoS2</td>
</tr>
<tr>
<td>csoS3</td>
</tr>
<tr>
<td><strong>Rubisco</strong></td>
</tr>
<tr>
<td><strong>Carboxysome CA</strong></td>
</tr>
<tr>
<td><strong>NDH-1 genes</strong></td>
</tr>
<tr>
<td>ndhD1/D2</td>
</tr>
<tr>
<td>ndhD5/6</td>
</tr>
<tr>
<td>ndhF1</td>
</tr>
<tr>
<td><strong>CO2 uptake</strong></td>
</tr>
<tr>
<td>ndhD3</td>
</tr>
<tr>
<td>ndhD4</td>
</tr>
<tr>
<td>ndhF3</td>
</tr>
<tr>
<td>ndhF4</td>
</tr>
<tr>
<td>chpY</td>
</tr>
<tr>
<td>chpX</td>
</tr>
<tr>
<td><strong>Bicarbonate transporters</strong></td>
</tr>
<tr>
<td>cmpA</td>
</tr>
<tr>
<td>cmpB</td>
</tr>
<tr>
<td>cmpC</td>
</tr>
<tr>
<td>cmpD</td>
</tr>
<tr>
<td>slr1515</td>
</tr>
<tr>
<td>slr1512</td>
</tr>
<tr>
<td><strong>Glycolate metabolism</strong></td>
</tr>
<tr>
<td>PGK</td>
</tr>
<tr>
<td>glcD</td>
</tr>
<tr>
<td>glcE</td>
</tr>
<tr>
<td>glcF</td>
</tr>
<tr>
<td>Glycolate oxidase</td>
</tr>
<tr>
<td><strong>Carbonic anhydrases</strong></td>
</tr>
<tr>
<td>α-CAs</td>
</tr>
<tr>
<td>β-CAs</td>
</tr>
<tr>
<td>γ-CAs</td>
</tr>
</tbody>
</table>
α-cyanobacteria have α-carboxysomes rather than the β-carboxysomes found in β-cyanobacteria. This is evidenced by the presence of single copies of csoS1, peptide A, peptide B, csoS2 and csoS3. This would suggest that these two carboxysomes have either been inherited or developed in parallel in association with the Form 1A or Form 1B Rubisco found in each cyanobacterial group. It is possible that both Rubisco and carboxysome genes could be inherited by a single lateral gene transfer event as in many cyanobacteria they are often found organised on contiguous regions of the chromosome. This is shown in Fig. 3 for carboxysome genes from *Synechococcus PCC7942, Synechococcus PCC7002, Synechococcus WH8102* and *P marinus* MED4. In other cyanobacteria, such as *Synechocystis PCC6803, Nostoc punctiforme* and *Anabaena PCC7120*, although the carboxysome ccm genes are clustered together, they are found in a region of the chromosome separated from Rubisco genes (data not shown and Fig. 7).

Carboxysomal CA is interesting in that such a CA gene has only been identified in Type B carboxysomes from β-cyanobacteria (Fukuzawa et al. 1992; Price et al. 1992; Yu et al. 1992). None of the α-cyanobacteria has a recognisable carboxysomal β-CA homolog, although one β-CA is present in the *Synechococcus WH8102* genome. The absence of any clearly identifiable CA gene in the two *Prochlorococcus* species is intriguing, and clearly points to the potential for a different mode of carboxysome function in these cyanobacteria. No carboxysomal CA gene has been identified to date for any α-carboxysome containing bacteria.

### NDH-1 genes and CO₂ uptake

Previous studies have shown that the NDH-1 dehydrogenase complex is involved in enabling CO₂ uptake by cyanobacteria (Ogawa 1992; Price et al. 1998; Klughammer et al. 1999; Ohkawa et al. 2000a, b). However, within β-cyanobacterial species there may be a number of NDH-1 complexes with different roles within the cell. Using *Synechocystis PCC6803* as a model, it can be shown that there are up to six ndhD genes and three ndhF genes, while there are single copies of the other nine polypeptides and these may combine in a manner to produce at least three different classes of NDH-1 complex (Price et al. 1998). The NdhD1/D2 polypeptides, together with NdhF1, probably are involved in forming a conventional respiratory NDH-1 complex, oxidising NADPH and NADH, and reducing plastoquinone (Ohkawa et al. 2000a). The NdhD3/D4, together with NdhF3 and F4 components, are implicated in forming two types of a specialised NDH-1 complex involved in driving active CO₂ uptake and converting CO₂ to HCO₃⁻ within the cell (Ohkawa et al. 2000a). The exact role of NdhD5/D6 polypeptides is unclear. In addition to this, two other genes/proteins are involved in enabling the CO₂ uptake activity of the NDH-1 complex, and these are referred to in the table as chpX and chpY (note that Shibata, Ogawa and colleagues have named these genes as cupA and cupB, while Price and co-workers have used chpX and chpY nomenclature; Maeda et al. 2002; Shibata et al. 2001, 2002; Price et al. 2002). Among all the cyanobacteria in Table 1, there are examples of the organisation of ndhF3, ndhD3 and chpY into an identifiable operon, and ndhF4, ndhD4 and chpX into another. Recent evidence produced from both Ogawa’s and Price’s laboratories (Maeda et al. 2002; Shibata et al. 2001) has clearly indicated that the F3/D3/chpY genes code for high-affinity CO₂ uptake, while the F4/D4/chpX genes code for low-affinity CO₂ uptake. Price et al. (2002) have speculated that the ChpX and Y polypeptides may in fact be a part of the NDH-1 CO₂-uptake complex, and are involved directly in the conversion of CO₂ to HCO₃⁻, linked to electron transport and proton translocation associated with the complex.

All the cyanobacterial species listed in Table 1 have two or more copies of ndhD1/D2 homologs, except *Synechococcus WH8102*. However, there was variability in the content of ndhD5/6 homologs, with α-cyanobacteria appearing to lack representatives of this class. There appears to be a single ndhF1 homolog in all species, except *Prochlorococcus MIT9313*.

For the CO₂-uptake related NDH-1 genes, all five β-cyanobacteria had single copies of ndhf, ndhD and chp genes, coding for both high- and low-affinity CO₂ uptake (note that in *Nostoc*, the possible ndhD4 gene is at the end of a contig, and its presence is somewhat problematic). The α-cyanobacterial species are different. *Synechococcus WH8102* has single copies of the ndhf4, ndhD4 and chpX genes (coding for low-affinity CO₂ uptake) arranged adjacent to each other and the carboxysome genes (see Fig. 3). However, the two *Prochlorococcus* species have no homologs of either the low-affinity or high-affinity CO₂ uptake NDH-1 genes. From what we know about
Fig. 4. Unrooted phylogenetic trees for a number of proteins associated with the CCM and photosynthetic carbon metabolism in photosynthetic bacteria. Proteins shown are (A) NdhF, (B) Chp, (C) GlcD, (D) PGA kinase, (E) phosphoribulokinase (PRK). Protein names are derived from the gene names shown in Table 1. Phylogenies are based on protein sequences that were aligned using the program ClustalW version 1.7 (Thompson et al. 1994). All phylogenetic trees are unrooted and were generated using the program PHYLIP version 3.6 alpha (J. Felsenstein, University of Washington, Seattle, USA). All trees were created using the neighbour joining method but they did not differ significantly from maximum likelihood parsimony analyses with respect to the conclusions drawn in the text. The species shown in the figure include the higher plant relatives, spinach (Spinacea oleracea), Marchantia polymorpha, wheat (Triticum aestivum), and Arabidopsis (Arabidopsis thaliana); the Proteobacteria, Nitrosomonas europaea, Rhodopseudomonas palustris, Rhodobacter sphaeroides, Rhodobacter capsulatus, Ralstonia metallidurans, Thiobacillus ferrooxidans, Nitrobacter vulgaris, Ralstonia eutropha, Nitrobacter vulgaris, Sinorhizobium meliloti, Hygenophilus thermoluteus, Vibrio cholerae, Xanthobacter flavus, and Escherichia coli; the α-cyanobacteria Prochlorococcus marinus MED4, P. marinus MIT9313, and Synechococcus WH8102; the β-cyanobacteria Synechococcus PCC7942, Synechococcus PCC7002, Synechocystis PCC6803, Anabaena PCC7120, and Nostoc pustuliforme; the green alga Chlamydomonas reinhardtii, and the heterokont alga Odontella sinensis. Subdivisions indicated by grouping and shading include the α- and β-cyanobacteria, high- and low-affinity CO₂ uptake protein homologs of NdhF and Chp, and the NdhF1 subunits associated with respiratory NDH-1 complexes.
β-cyanobacteria, this absence should mean that they lack the capacity for active CO$_2$ uptake, unless they possess another active CO$_2$ uptake system that is presently uncharacterised.

**Bicarbonate transporters**

Our current understanding of the genes and proteins involved in HCO$_3^-$ transport is less well developed than is the case for CO$_2$ uptake. The only well characterised HCO$_3^-$ transporter is the ABC-HCO$_3^-$ transporter (Bct1), described by Omata and colleagues (Okamura et al. 1997; Omata et al. 1999; Maeda et al. 2000). This is coded for by a four gene operon consisting of cmpA,B,C and D genes. These genes code for a high-affinity HCO$_3^-$ transport complex located on the plasma membrane, and have some homology to ABC nitrate transporters in β-cyanobacteria. The cmpABCD operon is found in the freshwater β-cyanobacteria, but is absent from the marine *Synechococcus* PCC7002 (D. Bryant, pers. comm.). This operon is also entirely absent from all three marine α-cyanobacterial species.

Two other genes may also code for proteins associated with HCO$_3^-$ transport. These are *slr*1515 or *ict*B (Bonfil et al. 1998), and *slr*1512 (Shibata et al. 2002). The *slr*1512 gene is probably present in all β-cyanobacteria, however it appears to be absent from all three α-cyanobacterial species. The *slr*1515 gene is more widely distributed, with only *Prochlorococcus* MED4 lacking a clearly identifiable homolog.

**Glycolate metabolism**

In examining the adaptation of cyanobacteria to limiting Ci, it is of interest to explore the presence of enzymatic steps for metabolising phosphoglycolate and glycolate, which are the products of the Rubisco oxygenase reaction. The *PGP* gene shown in Table 1 is the bacterial phosphoglycolate
phosphatase (PGP). Unfortunately, the gene for the higher plant PGP equivalent has not yet been identified. Two glycolate processing options are shown. One is the three-polypeptide glcDEF operon, identified in E. coli for glycolate metabolism (glycolate dehydrogenase; Pellicer et al. 1996). The other is the single-polypeptide glycolate oxidase gene identified in higher plants.

The PGP gene is found in all β-cyanobacteria. However, a clearly identifiable homolog appears to be absent from the three α-cyanophytes. The glcDEF gene assemblage appears also to be present in β-cyanobacteria, and is also present in Synechococcus WH8102 and Prochlorococcus MIT9313. At this stage, no clear homologs can be found in Prochlorococcus MED4. A homolog of the higher plant glycolate oxidase is present only in the closely related Nostoc and Anabaena species.

Carbonic anhydrases

CAs have been intimately implicated in the operation of the CCM, for their ability to interconvert CO₂ and HCO₃⁻. However, as noted above, only the carboxysomal β-CA has been shown to have a direct involvement in the CCM of β-cyanobacteria. An analysis of possible α-, β-, and γ-CAs (see Smith and Ferry 2000) in the cyanobacterial genomes shows that there is a wide diversity in CA gene content. A β-carboxysomal CA gene is probably present in all β-cyanobacteria, although it is not clearly obvious in Anabaena at present. In addition to this, one or more other β-CAs may also be present, including eccB (So and Espie 1998). The β-cyanobacteria may also possess an α-CA (Soltes-Rak et al. 1997). There is a paucity of any identifiable CA genes in the α-cyanophytes, with one β-CA (not a carboxysome homolog) present in only Synechococcus WH8102. There are no clearly proven γ-CAs in any of the cyanobacteria, although the CcmM protein in β-cyanobacteria has an amino terminal domain that could potentially contain a γ-CA active site, and ferripyochelin has some homology to γ-CA enzymes (Smith and Ferry 2000).

CCM gene phylogenies

With the apparent divergence of the carboxysomes and Rubisco between α- and β-cyanobacteria it is pertinent to examine the phylogenetic relatedness of other proteins associated with both the CCM and photosynthetic carbon metabolism.

Figure 4a shows an unrooted tree for various NdhF homologs distributed among various proteobacteria and cyanobacteria. Homologs of the NdhF1 subunit, found in the respiratory NDH-1 complex are found in all species, but it is apparent that the α- and β-cyanobacteria are more closely related to each other and higher plants than to proteobacterial species. However, there are clear separations between both the α- and β-cyanobacteria and higher plants. Homologs of the NdhF3 and NdhF4 subunits associated with the CO₂-uptake NDH-1 complex are found only in α- and β-cyanobacteria and are closely related to each other. A similar unrooted tree is also produced considering the NdhD homologs in all cyanobacteria are again more closely related to each other and higher plants than to other proteobacteria, and the NdhD3/D4 homologs are only found in cyanobacteria.

The ChpX and ChpY proteins in cyanobacteria are closely related to each other, and again there are no homologs found outside either α- or β-cyanobacteria (Fig. 4b). HCO₃⁻ transporters are not considered here except to say that significant homologs of the CmpABCD proteins, slr1515 and slr1512, are not found outside the cyanobacteria.

Considering other genes associated with the PCR and photosynthetic carbon oxidation (PCO) cycles, GlcD proteins are representative of the phylogeny of the glycolate GlcD,E,F complex (Fig. 4c). The GlcD and PGA kinase proteins (Fig. 4d) are more closely related to other cyanobacterial homologs, than to other proteobacteria. However, with both trees there are clear α- and β-cyanobacterial grouping and relatedness.

The above phylogenies are consistent with α- and β-cyanobacteria having a common ancestral origin, as has been previously shown by examining the phylogenies based on 16S ribosomal sequences (Urbach et al. 1998; Honda et al. 1999). However, one other PCR cycle enzyme shows a phylogenetic divergence similar to that found for Rubisco and carboxysomes. Figure 4e indicates that Phosphoribulokinase (PRK) phylogeny shows two distinct groupings within cyanobacteria. The β-cyanobacteria possess a PRK that is higher plant-like in type, while α-cyanobacterial PRK is clearly proteobacterial in nature, being more closely related to proteins found β-proteobacteria such as Nitrosomonas europaea and E. coli rather than to other β-cyanobacteria. The reason for this divergence is unclear. PRK is the enzyme that supplies Rubisco with its substrate, ribulose bisphosphate, but our understanding of carboxysomes and the PCO cycle is that carboxysomes only contain Rubisco and CA, and that PRK is soluble in the cytosol. Rubisco and CA are the only two enzymes to have been found to be associated with either α- or β-carboxysomes (Shively and English 1991; Price et al. 1998). Is it possible that PRK is associated with Rubisco and carboxysomes in a manner that would necessitate a PRK protein that would match the Rubisco and carboxysomes present in the cell?

Evolution of the cyanobacterial CCM

With our current knowledge of cyanobacterial CCM genes (Table 1) and view of past evolutionary and climatological processes, it is possible to speculate on the timing and evolutionary pathway for the development of CCMs in cyanobacteria. Figure 5 shows an historical view of the development of photosynthetic cyanobacteria and algae over the past 3.5 billion years. This is plotted together with
Evolution and diversity of CCMs in cyanobacteria deduced changes in CO₂ and O₂ over the past 600 million years (the Phanerozoic era).

The atmospheric CO₂ level when cyanobacteria first arose was probably over 100-fold higher than present day conditions. This, combined with low O₂ conditions, would have meant that the original cyanobacteria would not have needed a CCM to achieve effective photosynthesis. The initial development of a CCM in cyanobacteria would have been triggered by changes in CO₂ and O₂ that caused CO₂ to be a limiting resource for photosynthesis and the Rubisco oxygenase reaction to be a problem. The CO₂ and O₂ levels at which this occurred would have depended on the kinetic properties of the original cyanobacterial Rubiscos (both Form 1A and Form 1B), and the extent to which a cyanobacterial community may have been limited by diffusion factors, such as living in cyanobacterial mats. Clear records for changes in O₂ and CO₂ before about 600 million years ago are lacking, but it has been inferred that O₂ was near present levels by the beginning of the Phanerozoic, and CO₂ may have been around 15–20 times current atmospheric levels. Given the properties of current cyanobacterial Rubiscos (see Badger et al. 1998), these enzymes should have been able to achieve efficient photosynthesis under these conditions. This would be even more certain if we entertain the likely possibility that prior to the advent of CCMs, cyanobacterial Rubiscos were perhaps more efficient than they are today. About 400 million years ago, there was a large decline in CO₂ levels and an almost doubling of the O₂ concentration. These changes would have placed significant pressures on both cyanobacterial and algal photosynthesis. It can be argued that this was the first time that major pressure was applied to photosynthetic organisms to develop CCMs.

The steps involved in developing a cyanobacterial CCM may have been quite simple at first, and speculation is outlined in Fig. 6. In the first stages of CO₂ decline, the first step towards developing a CCM would have been the evolution of a carboxysome structure for Rubisco. The cyanobacterial CCM is totally dependent on this structure, and all other additions would have revolved around its presence. If atmospheric CO₂ was 10-times the present levels, then cytosolic HCO₃⁻ would have been around 10–20 mM (depending on pH), and would have been sufficient substrate to elevate CO₂ within the carboxysome appreciably, allowing effective CO₂ fixation (Badger et al. 1991). A carboxysome CA would probably have been required at this stage, as the rate of chemical conversion of HCO₃⁻ to CO₂ would have been too slow. The presence of a cytosolic CA in this early phase would be possible, as Ci

---

**Fig. 5.** The evolution of photosynthetic cyanobacteria, algae, higher plants, and chemoautotrophic proteobacteria over the past 3.5 billion years, with particular reference to the changes in CO₂ and O₂ during the Phanerozoic period (drawn with reference to Raven 1997b). Also shown on the graph are the proposed changes in CO₂ and O₂ during the Phanerozoic period (Berner 2001; Berner and Kothavala 2001). CO₂ is shown as the ratio of past to present levels (RCO₂), while O₂ is shown as percent content. The shaded section indicates the period of CO₂-limitation combined with O₂ increase that may have initiated the development of CCMs in aquatic photosynthetic organisms.
would be in passive equilibrium between the external medium and the cytosol. However, when active Ci accumulation was developed, cytosolic CA would have been lost in order to avoid wasteful CO₂ leakage. As CO₂ limitation became more severe, the development of the NDH-1 based low- and high-affinity CO₂ hydration process would have maintained adequate internal HCO₃⁻ pools and CO₂ levels around Rubisco in the carboxysome. This process would have been based around the modification of an existing respiratory NDH-1 complex, and would have resulted in the efficient recycling of leaked CO₂, as well as net acquisition of CO₂ from outside the cell. Finally, as more extreme CO₂ limitation was imposed, the evolution of low- and high-affinity HCO₃⁻ transport systems and high-affinity CO₂ uptake would have been necessary.

Of some interest is the close linkage between the NDH-1 CO₂-hydration genes and carboxysome genes which is apparent in some cyanobacteria, such as Anabaena PCC7120, N. punctiforme and Synechococcus WH8102 (Fig. 7). Here, the ndhF4, D4 and chpX low-affinity CO₂ hydration genes are located adjacent to the carboxysome genes. In the case of Synechococcus WH8102, Rubisco is also located in this region. The presence of NDH-1 CO₂ uptake systems in some α-cyanobacteria could be due to gene transfer from β-cyanobacteria.

The pathway for the initial acquisition of carboxysomes is interesting to consider, particularly as it is now evident that the α- and β-cyanobacteria have carboxysomes with different protein components (see Table 1). Examining the genes involved in cyanobacterial carboxysomes and proteobacterial microcompartments, it is obvious that the common components are the small ccmK and ccmL-bacterial micro-

![Fig. 6. Possible pathways for the evolution of the CCM and its components in cyanobacteria.](image)

![Fig. 7. Contiguous organisation of carboxysome genes and ndhF4, D4, and chpX low-affinity CO₂ uptake genes in Synechococcus WH8102, Anabaena PCC7120, and Nostoc punctiforme. Note that in Nostoc the ndhF4 genes occur at the end of a DNA contig, and the position of the other ndhD4 and chpX genes are not certain at present. DNA sequences were obtained from sources listed in Table 1.](image)
compartment genes. The csoS2 and csoS3 genes are specific for Form 1A Rubisco α-carboxysomes, while the ccmM and ccmN genes are specific for Form 1B Rubisco β-carboxysomes. Thus, the question becomes, were the small bacterial microcompartment genes developed first in cyanobacteria or did they originate in β-proteobacterial species? Some potential scenarios are outlined in Fig. 6, and these incorporate the increasingly evident view that lateral gene transfer has been a major part of bacterial evolution and the acquisition of important functional genetic units (Eisen 2000; Brown et al. 2001; Sicheritz-Ponten and Andersson 2001).

If microcompartment genes originated with β-cyanobacteria, then you can postulate that proteobacteria may have acquired these genes by lateral gene transfer. These genes would then have supported the development of ethanolamine and propanediol metabolism compartments, as well as carboxysomes in autotrophic proteobacteria.

The appearance of α-carboxysomes in α-cyanobacteria is largely dependent on where the initial origin of Form 1A Rubisco is assumed to have occurred. As indicated in Fig. 6, Form 1A Rubisco in cyanobacteria may have preceded the appearance in proteobacteria, and this would have occurred with the evolution of the α-cyanobacterial lineage. If this is the case then, as indicated in Fig. 6, the origin of α-carboxysomes in autotrophic proteobacteria may be the result of gene transfer from the α-cyanobacteria, and would have been triggered by the low CO2 conditions in the Phanerozoic (see Fig. 5). With the contiguous arrangement of carboxysome genes and Rubisco genes on the chromosome, the transfer of a complete α-carboxysome and Form 1A Rubisco may have been possible.

The evolutionary divergence of α- and β-cyanobacteria is probably a fairly ancient event, and it has been argued that the types of cyanobacteria split before the advent of the primary endosymbiosis some 2 billion years ago (Tomitani et al. 1999). If this is the case, then it could be argued that both cyanobacterial groups developed CCM mechanisms independently of each other, rather than from a common ancestor having CCM components.

Are CCMs in cyanobacteria related to those in algae and higher plants?

If CO2 limitation were not imposed on cyanobacteria until the Phanerozoic, then this would strongly imply that the cyanobacteria that were the basis for the original primary endosymbiotic event(s) did not have CCMs. Thus, the original Chlorophyte and Rhodophyte algae would have lacked any pre-existing, common CCM genetic elements with cyanobacteria that may have aided their adaptation to falling CO2 levels in the Phanerozoic. Chlorophytes and Rhodophytes, as well as the secondary and tertiary endosymbiont algae that arose during the CO2 limitation of the Phanerozoic, would have all needed to develop independent strategies for adapting to low CO2. Indeed, it has been suggested that the development of secondary endosymbiont algae may have been driven by this decline in CO2, as engulfment into an acidic vacuolar structure may have made CO2 more available by the conversion of HCO3− to CO2 (Lee and Kugrens 2000).

A polyphyletic origin of CCM mechanisms in algae and cyanobacteria has been previously argued (Raven 1997a, b), and this is in agreement with the above interpretation. If the original cyanobacterial endosymbiont had a CCM, then one might expect to find homologs of CCM genes in either the algae or perhaps higher plants. A search of the existing higher plant and algal databases (see phylogenetic trees in Fig. 4) indicates there are no homologs of cyanobacterial CCM genes to be found. Thus, carboxysome genes are restricted to cyanobacteria and some proteobacteria, NDH-1 CO2-uptake genes are restricted to cyanobacteria, as are the HCO3− transport genes. However, the algal genome databases are limited at present and are dominated by algal chloroplast genomes and Chlamydomonas reinhardtii expressed sequence tags, and it is possible that further sequencing of other algal species may uncover some CCM homologs in the nuclear genome. However, it would seem reasonable at this stage to assume that it was probable that there were multiple origins of aquatic CCMs in algae and the ancestors of higher plants. Although it is likely that there are similar functional elements in all CCMs, based on Ci acquisition, HCO3− to CO2 conversion, and diffusion restrictions, the nature of the genetic elements of each system may be quite different.

Genetic diversity in cyanobacteria and ecological adaptation

The analysis shown in Table 1 indicates a number of areas in which CCM diversity exists among cyanobacteria, with the major differences being most apparent between α-cyano-bacteria and β-cyanobacteria.

The α- and β-carboxysomes are perhaps the most striking variation. However, as we have no idea about the relative effectiveness of each type of carboxysome, it is difficult to speculate on the photosynthetic carboxylase advantages that each structure and its Rubisco may confer.

All β-cyanophyte genomes examined to date have genes correlated with both low- and high-affinity NDH-1 CO2-uptake systems. This presumably implies that all these cyanobacteria are able to induce high-affinity CO2 uptake systems when Ci becomes most limiting. Common high- and low-affinity HCO3− transport systems may also be present, if slr1512 and slr1515 are in fact HCO3− transport-related genes. The Ci transport properties of α-cyanobacteria are again quite different. *Synechococcus* WH8102 possesses genes that could code only for a low-affinity NDH-1 CO2-uptake system, however, these genes are absent from both *Prochlorococcus* species. Observations with
Synechococcus WH7803 (a close relative of WH8102) that they were able to evolve CO₂ during photosynthesis using HCO₃⁻ (Tchernov et al. 1997), may be consistent with either the absence of CO₂ uptake genes in this species, or the presence of only the low-affinity uptake system. Extensive physiological studies of α-cyanobacterial species are lacking, but for Prochlorococcus, it would be expected that if these cyanobacteria do possess active HCO₃⁻ transport then they should evolve CO₂ during active photosynthesis. However, the nature of HCO₃⁻ transport systems in α-cyanobacteria may be quite different from β-cyanobacteria. A homolog of slr1515 appears to be present, but slr1512 appears to be absent (Table 1). If α- and β-cyanobacteria diverged in their evolution prior to the development of HCO₃⁻ transport systems, as carboxysome differences may suggest, then different types of HCO₃⁻ transport may have evolved independently.

The ecological significance of the differences in CCMs between α- and β-cyanobacteria remains to be determined. However, it is clear that the α-cyanobacteria in Table 1 occupy quite different habitats compared with most β-cyanobacteria. The Prochlorococcus species and Synechococcus WH8102 occur as dominant primary producers throughout oligotrophic oceanic waters roughly spanning the latitudes 40° N and 40° S (Moore et al. 1998; Partensky et al. 1999). In these environments, it may be expected that Ci is never severely depleted, and light and other nutrients may be major limiting factors. The β-cyanobacteria in Table 1 occupy environments such as mats, films, estuarine situations, and alkaline lakes, where higher population densities of organisms may occur, other nutrients may be more abundant and overall, situations where Ci is a limiting resource may be much more common. Hence, the oceanic α-cyanobacteria may have developed a physiology where they may not have the ability to acquire or induce high-affinity Ci transport systems, and in some species no active CO₂ uptake system may be present. On the other hand, many β-cyanobacteria have the ability to induce various CO₂ and HCO₃⁻ transport systems as their environmental conditions change.

A comment on cyanobacterial classification

In undertaking this analysis of CCM gene diversity and evolution, it has become apparent that the classification system used in this paper based on the presence of Form 1A or Form 1B Rubisco and corresponding α- or β-carboxysomes in α- and β-cyanobacteria may be a useful basis for further understanding of basic genetic variation and relatedness within cyanobacteria. Previous variation based on pigment composition as encompassed in the cyanobacteria and prochlorophyte descriptions is relatively unhelpful, as there is a polyphyletic appearance of pigment variation within widely different cyanobacterial groups. The α- and β-cyanobacteria, however, would appear to be a much more robust and basic subdivision, and worthy of further consideration as a revised classification tool for cyanobacterial subdivision.

References


Manuscript received 19 September 2001, accepted 9 November 2001.

http://www.publish.csiro.au/journals/fpb