A Nucleomorph-Encoded CbbX and the Phylogeny of RuBisCo Regulators

Uwe-G. Maier,* Martin Fraunholz,* Stefan Zauner,* Susanne Penny,† and Susan Douglas†

*Cell Biology and Applied Botany, Philipps-University Marburg, Marburg, Germany; and †National Research Council of Canada and Program in Evolutionary Biology, Canadian Institute of Advanced Research, Institute for Marine Biosciences, Halifax, Nova Scotia, Canada

Chloroplasts contain proteins that are encoded by different genetic systems, the plastid genome and the nuclear chromosomes. By comparing the gene content of plastid genomes of different taxa, some predictions about nuclear-encoded genes for plastid proteins are possible. However, early in evolution, many genes were transferred from the plastid to the cell nucleus and are therefore missing from all known plastid genomes and escape such predictions. By sequencing the miniaturized chromosomes of the nucleomorph of the cryptophyte Guillardia theta, as well as the plastid genome, we uncovered two genes encoding CbbX which are predicted to be involved in plastid function. Our findings suggest that (1) red-type plastid rbcLS genes evolved together with cbbX, which is related to cbbX genes of purple bacteria; (2) early in rhodoplast evolution, the cbbX gene was duplicated and transferred into the nucleus; (3) the plastid-encoded LysR transcriptional activator gene, rbcR, is homologous to rbcR and cbbR transcriptional activator genes of purple bacteria and cyanobacteria; and (4) the ancestral plastid probably harbored both types of form I RuBisCo.

Introduction

In the course of plastid evolution, the cyanobacterium-like endosymbiont that gave rise to this semi-autonomous organelle lost most of its genes. Some were deleted, but most were transferred into the host nucleus. Interestingly, the numbers of transferred genes differ significantly in different taxa. Comparisons of plastid gene content in green and nongreen plastids (glauco-cystophytes, red algae, heterokonts, and cryptophytes) show that more than 100 genes are differentially distributed among distinct evolutionary lineages (Martin et al. 1998). Furthermore, independent parallel gene losses from plastid DNA vastly outnumber phylogenetically unique gene losses. Nevertheless, a handful of genes can easily be identified that are typical of the plastid genomes of glauco-cystophytes (Stirewalt et al. 1995) or the red lineage that includes red algae, heterokont algae, and cryptophytes (Kowallik et al. 1995; Reith and Mun-holland 1995; Douglas and Penny 1999) but are missing from the green plastid lineage (Ohyama et al. 1988; Hiratsuka et al. 1989; Sugiuara 1992; Hallick et al. 1993; Wakasugi et al. 1994, 1997; Maier et al. 1995). If the products of those genes restricted to nongreen plastids are involved in general plastid functions, homologs will be identified in the cell nucleus of the green line. On the other hand, former endosymbiotic genes that cannot be identified in at least one plastid genome escape such predictions.

One of the first reported examples of differential gene flow from the plastid to the cell nucleus in phylogenetically different lines was the identification of rbcS, the gene encoding the small subunit of RuBisCo, in the plastid genomes of nongreen plastids (Douglas and Durnford 1989; Martin et al. 1998). This differs from the situation for green algae and land plants, for which rbcS is nuclear-encoded (Coruzzi et al. 1983; Martin and Schnarrenberger 1997). In red algae, heterokonts, and cryptophytes, rbcS and rbcL are arranged as an operon and seem to coevolve with cbbX, whose gene product is shown to be necessary for photoautotrophic growth (Gibson and Tabita 1997). The cbbX gene (cfxQ in some database entries, renamed cbbX by Stöbe, Martin, and Kowallik [1998]) is located downstream of the rbc genes in red algae and cryptophytes (Reith and Munholland 1995; Ohta 1997; Douglas and Penny 1999), whereas in heterokonts, it is plastid-encoded but not located adjacent to the Rubisco genes (Kowallik et al. 1995).

Recently, a curious RuBisCo gene was identified in the Methanococcus genome encoding an O2-sensitive enzyme (Watson, Yu, and Tabita 1999), but data about regulation of its expression do not exist. In α-proteobacteria, the rbcL and rbcS genes are part of a large operon encoding genes involved in carbon dioxide fixation. The rbcL and rbcS genes of Xanthobacter flavus, Rhodobacter sphaeroides, and Alcaligenes eutrophus are flanked by two open reading frames (ORFs), with the upstream one exhibiting similarities to a member of the LysR-type transcriptional activator family CbbR, and the downstream one exhibiting similarities to CbbX (Meijer et al. 1991; Gibson, Falcone, and Tabita 1991; Kusian et al. 1992; Gibson and Tabita 1996; Kusian and Bowien 1997). A transmissible plasmid present in A. eutrophus encodes a duplicate copy of the carbon dioxide fixation genes with the exception of the upstream LysR-type transcriptional regulator gene (Kusian et al. 1995).

The RuBisCo found in eukaryotes can be classified into three types according to its structure and biochemistry, but its evolution is a mystery. Form I RuBisCo exists in two subtypes: the green type I RuBisCo of glauco-cystophytes, green algae, land plants and organisms harboring secondarily evolved green plastids (euglenoids and chlorarachniophytes), and the red type I RuBisCo of red algae and organisms harboring secondarily evolved red plastids (e.g., heterokonts and cryptophytes). The green type I RuBisCo shows phyloge-
with cyanobacterial homologs, whereas the red type I RuBisCo is more closely related to homologs from α-proteobacteria (Douglas, Durnford, and Morden 1990). Furthermore, the red type I RuBisCo genes are usually found adjacent to cbbX as in α-proteobacteria. Among eukaryotes, form II RuBisCo is to date exclusively found in peridinin-containing dinoflagellates (Morse et al. 1995; Whitney, Shaw, and Yellowless 1995).

Several hypotheses have been put forward to explain the presence of an α-proteobacterial (rather than cyanobacterial) form I RuBisCo in red algae, heterokont algae, and cryptophytes. Cavalier-Smith (1989) argued that the Rubisco operon of chromophytes could have been acquired from the α-proteobacterial symbiont that gave rise to mitochondria. Another possibility is the lateral gene transfer of an α-bacterial form I RuBisCo into the ancestors of plastids either before or after endosymbiosis (Douglas, Durnford, and Morden 1990). This seems an attractive hypothesis given the presence of the RuBisCo operon on a transmissible plasmid in certain α-proteobacteria (Kusian et al. 1995). If both types of form I RuBisCo genes were present in the progenitor of plastids, as is found, for example, in the α-proteobacterium *Rhodobacter capsulatus* (Paoli et al. 1998), alternate copies could have been differentially lost in the red and green lineages during the course of plastid evolution, resulting in the distribution seen today (Delwiche and Palmer 1997; Martin and Schnarrenberger 1997).

Cryptophytes are links in the evolution of complex plastids (Douglas 1994; Frauenholz et al. 1997). In addition to a nucleus, a mitochondrion, and a plastid, they harbor a fourth DNA-containing organelle, the nucleomorph (Gibbs 1981; McFadden et al. 1997). This dwarf nucleus is the remnant of the nucleus of a red alga-like symbiont (Douglas et al. 1991; Maier et al. 1991; Van de Peer et al. 1996; Van der Auwera et al. 1998) and contains a subset of the genes once present in the nucleus. These genes are located on three miniaturized chromosomes that are presently being sequenced in an international collaborative effort (Gilson, Maier, and McFadden 1997).

Here, we present an example from our *Guillardia theta* sequencing project: two genes for CbbX, one encoded on the nucleomorph genome and one on the plastid genome (both gene products located in the plastid), and a distinct LysR-type transcriptional activator (RbcR) encoded in the plastid genome. The detection of these factors, associated with Rubisco expression but encoded in two different chromosomal backgrounds, provides insights into the regulation of plastid processes and the dependence of the plastid on the nucleus. Furthermore, phylogenetic studies demonstrate that the *cbbX* genes, including those of *G. theta*, are similar to the stage-specific *Bacillus* sporulation factors and a group of mycobacterial hypothetical proteins, and suggest that the plastid-encoded RbcR, a homolog of the prokaryotic transcriptional activators CbbR and RbcR, is of cyanobacterial origin.

### Materials and Methods

**Cloning and Sequencing**

DNA of the nucleomorph chromosomes was restricted with Xba1 (Amersham-Pharmacia) and cloned into pBluescript plasmid vector (Stratagene). Plasmids were sequenced on an ALFExpress (Amersham-Pharmacia) automated sequencer using the Thermo Sequenase fluorescent labeled primer cycle sequencing kit (Amersham-Pharmacia) with 7-deaza-dGTP. The chloroplast genome was subcloned into pUC19 (Amersham-Pharmacia) using a variety of restriction enzymes and sequenced using a Perkin Elmer ABI 373 automated sequencer and the AmpliTaq FS dye terminator cycle sequencing ready reaction kit (Douglas and Penny 1999).

**Phylogenetic Analysis**

LysR-type homologous sequences were obtained from GenBank. Amino acid alignments (166 amino acids) were used to construct the phylogenetic tree. Phylogenetic analyses were performed using the PHYLIP, version 3.5, package (Felsenstein 1989).

### Results and Discussion

By shotgun sequencing of the nucleomorph chromosomes of the cryptomonad alga *G. theta*, we identified an ORF, nmCbbX (GenBank accession number AJ251479), whose amino acid sequence is 56% identical to the plastid-encoded CbbX (ptCbbX), a protein thought to be involved in RuBisCo expression (fig. 1) (Douglas and Penny 1999). The nmCbbX gene is transcribed as shown by the detection of a polyadenylated cDNA in our EST project (data not shown). An additional N-terminal extension on the nmCbbX characteristic of plastid targeting signals suggests that it resides in the plastid. Therefore, at least two different *cbbX* genes are maintained in cryptophytes, one in the plastid and one in the nucleomorph genome. In addition, the gene for RbcR, an ORF related to the CbbR/RbcR group of LysR-type transcriptional activators, is maintained on the plastid genome (GenBank accession number AF041468) (Douglas and Penny 1999).

The plastid *rbcL/rbcS/cbbX* operon arrangement of cryptophytes (Douglas and Penny 1999) and that of the red algae *Porphyra purpurea, Cyanidioschyzon merolae,* and *Cyanidium caldarium* (Reith and Munholland 1995; Ohta 1997) are identical. An ORF with high similarity to the amino terminus of CbbX is also evident downstream of *rbcS* of *Antithamnion* sp. if a frameshift is introduced in the sequence reported by Kostrzewa et al. (1990). Therefore, it is unlikely that the possession of two *cbbX* genes in different chromosomal backgrounds in cryptomonads represents redundant genetic material and that the plastid-located gene is destined for elimination. More likely, both genes are needed for plastid functions, by acting either as heterodimers or with different accessory proteins.
Fig. 1.—Alignment and chromosomal localization of CbbX polypeptides. 

a. Alignment of deduced amino acid sequences of CbbX polypeptides encoded by the nucleomorph (Nm) and chloroplast genomes of Guillardia theta. Amino acid residues shared by both CbbX molecules are boxed and shaded. Gaps introduced to improve the alignment are indicated by dashes.

b. Chromosomal localization of cbbX on nucleomorph chromosome II and the chloroplast genome. 26S Proteasome S10B subunit S10B of the 26S proteasome; SnRNP core protein D3 of a small nuclear ribonucleoprotein particle; ORF open reading frame; cbbX Calvin-Benson-Bassham-cycle protein X; rpS27 ribosomal protein S27; ilvH gene for the small subunit of the acetohydroxyacid synthetase.

Phylogeny of CbbX and Members of the CbbX Family

In order to reconstruct the evolution of CbbX and homologous proteins, we constructed phylogenetic trees. Database searches showed a protein family dominated by several classes, including genes for sporulation factors and protein genes found in Mycobacteria (fig. 2). As shown in figure 2, the nucleomorph-encoded CbbX branches together with some α-proteobacterial CbbX sequences, but not within the plastid-encoded CbbX group. This phylogenetic separation of the two crypto-
monad CbbX proteins may be a result of the higher AT content of the nucleomorph genome relative to the plastid genome, which could cause differences in the distribution of sites free to vary (Lockhart et al. 1998). Each of the organisms in this CbbX group, whether prokaryotes or eukaryotes, contains a red-type RuBisCo, and with the exception of Odontella sinensis and the G. theta nucleomorph, the RuBisCo genes are physically linked to the cbbX gene. This suggests that the genes for the CbbX group originated in the α-proteobacterial RuBisCo gene cluster and were integrated as a unit into the genome of the red-type plastid as portrayed in figure 3. The sister group of the CbbX proteins are the proteins for sporulation factors and CbbX-related ones found in Mycobacteria (fig. 2), indicating that this group of proteins acquired alternative functions. At least in the case of the sporulation factors, the CbbX-related proteins are expressed in a stage-specific manner (Piggot 1996).
None of the sequences including the sporulation factors and the mycobacterial homologs are DNA-binding proteins, due to the lack of domains known for DNA-binding. However, a chaperone-like function of CbbXs and related proteins is possibly indicated by an ATP/GTP-binding site.

Phylogeny of RbcR and Members of the LysR Family

A phylogenetic tree of RbcR and related LysR transcriptional activators is presented in figure 4. (Due to inconsistencies in nomenclature, a mixture of gene names exist in the database, especially with respect to CbbR, RbcR, and CfxR. In our tree, we named the groups, with respect to phylogeny, RbcR or CbbR subgroup I and II, independent of database entries.) Members of this group (which includes a mycobacterial and a streptomycete LysR, various RbcR’s and CbbR’s, and plastid-encoded Ycf30’s/RbcR’s) lack an ATP/GTP-binding site. It is known that at least some CbbR and RbcR polypeptides are DNA-binding proteins, acting in close proximity to the rbcLS genes of green- and red-type RuBisCo (e.g., Kusian and Bowien 1997; Paoli, Vichivanives, and Tabita 1998), and that in Xanthobacter flavus CbbR is a NADPH sensor (van Keulen et al. 1998). Plastid-encoded RbcR (like bacterial CbbR and RbcR) contains a helix-turn-helix domain, suggesting that it, too, acts as a DNA-binding protein. Interestingly, a RbcR homolog was identified not only in Synechocystis and in the plastid genome of glaucocystophytes (which possess green-type RuBisCo), but also in those of organisms with red-type RuBisCo (red algae, heterokont algae, cryptomonads). All plastid-encoded RbcR’s, including the Synechocystis sequence sl0998 form a group with 100% bootstrap support, indicating that they share a common cyanobacterial ancestry.

Red-Type RuBisCo Expression: A Puzzle of Different Origins

As proposed in figure 3, the α-proteobacterial cbbX, together with the red-type RuBisCo genes, were transferred and integrated into the ancestral plastid, but the transcriptional activator, CbbR, either was not inte-
Fig. 4.—Phylogenetic tree derived from members of the LysR-type family of transcriptional regulators using the neighbor-joining method with a protein distance matrix. The consensus tree of 100 bootstrap replicates is shown, and bootstrap values higher than 50% are indicated at the nodes. The genes from *Alcaligenes eutrophus*, *Xanthobacter flavus*, and *Rhodobacter sphaeroides* are those found upstream of *rbcL*/*rbcS* in the gene cluster from these organisms. Abbreviations are as given in figure 2, with the following exceptions: *C. paradoxa*, *Cyanophora paradoxa*; *C. vinosum*, *Chromatium vinosum*; *S. coelicolor*, *Sci35.38c*; *R. rubrum*, *A. eutrophus*, *X. flavus*, *R. capsulatus*, *R. sphaeroides*.

Impact of the Nucleomorph Sequencing

As recently shown in the case of *groEL/cpn60* (Wastl et al. 1999), and now for the CbbX family, our
nucleomorph sequencing project provides new information about the complement of plastid proteins and their gene distribution and can allow predictions to be made about a nucleus-encoded second copy of \( cbbX \) in red algae and chromophytes. Our data provide insights into host control of functions of red alga–like plastids. Obviously, these plastids are more under host control than previously thought. The existence of plastid-located genes encoding general regulatory functions, such as gene expression and GroEL chaperonin-mediated protein folding, does not guarantee its independence. Here, at least one further factor is under direct control of the host genetic system, thereby allowing the host to regulate the whole RuBisCo complex. Therefore, the increased gene content of red/glaucocystophyte plastid genomes relative to those of green plastids does not lead to a more autonomous status in the former type of plastids; as in green plastids, regulatory functions tend to be fixed in the nuclei of organisms containing red alga–like plastids (Herrmann 1997; Martin and Herrmann 1998).

Acknowledgments

We thank Drs. Mike Reith and Bettina Stöbe for critical reading and comments on the manuscript. The authors are supported by the Deutsche Forschungsgemeinschaft (Germany) and the Natural Sciences and Engineering Research Council (Canada). This is NRC publication number 42304.

LITERATURE CITED


MAIER, R. M., K. NECKERMANN, G. L. IGLOI, and H. J. KÖSSL. 1995. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tun-


William Martin, reviewing editor

Accepted December 13, 1999