The Evolution of the Cytochrome c_6 Family of Photosynthetic Electron Transfer Proteins

Barnaby Slater [†], Darius Kosmützky[†], R. Ellen R. Nisbet [‡,*], and Christopher J. Howe [*]

Department of Biochemistry, University of Cambridge, United Kingdom

*Corresponding authors: E-mails: ellen.nisbet@nottingham.ac.uk; ch26@cam.ac.uk.

Accepted: 10 June 2021

[†] These authors contributed equally to this work.

[‡]Present address: School of Biosciences, University of Nottingham, Sutton Bonington Campus, Sutton Bonington, Leicestershire, United Kingdom

Abstract

During photosynthesis, electrons are transferred between the cytochrome b_6f complex and photosystem I. This is carried out by the protein plastocyanin in plant chloroplasts, or by either plastocyanin or cytochrome c_6 in many cyanobacteria and eukaryotic algal species. There are three further cytochrome c_6 homologs: cytochrome c_{6A} in plants and green algae, and cytochromes c_{6B} and c_{6C} in cyanobacteria. The function of these proteins is unknown. Here, we present a comprehensive analysis of the evolutionary relationship between the members of the cytochrome c_6 family in photosynthetic organisms. Our phylogenetic analyses show that cytochromes c_{6B} and c_{6C} are likely to be orthologs that arose from a duplication of cytochrome c_6, but that there is no evidence for separate origins for cytochromes c_{6B} and c_{6C}. We therefore propose renaming cytochrome c_{6C} as cytochrome c_{6B}. We show that cytochrome c_{6A} is likely to have arisen from cytochrome c_{6B} rather than by an independent duplication of cytochrome c_6, and present evidence for an independent origin of a protein with some of the features of cytochrome c_{6A} in peridinin dinoflagellates. We conclude with a new comprehensive model of the evolution of the cytochrome c_6 family which is an integral part of understanding the function of the enigmatic cytochrome c_6 homologs.

Key words: photosynthesis, evolutionary model, phylogeny.

Significance

The cytochrome c_6 family of proteins plays an essential role in photosynthetic electron transfer, but the evolutionary relationships among the members of the family remain unclear. We show that a previously drawn distinction between cytochromes c_{6B} and c_{6C} probably reflects taxon sampling, that cytochromes c_{6BC} arose from cytochrome c_6, and that cytochrome c_{6A} subsequently arose from cytochrome c_{6B} after the divergence of the green photosynthetic lineage. These conclusions, together with a survey of the distribution of the family among eukaryotes, give us a much better understanding of the evolution of this important protein family.

Introduction

The Cytochrome c_6 Family of Proteins

Photosynthesis is one of the most important processes in the natural world and has played a vital role in shaping the planet and its atmosphere. One essential feature of oxygenic photosynthesis is the photosynthetic electron transfer chain (PETC), where the oxidation of water to generate reducing equivalents and chemical energy as ATP is driven through light energy absorption. In the plant PETC, electrons can be transferred between the cytochrome b_6f complex and photosystem I by the copper-containing protein plastocyanin (Gross 1993). Many cyanobacteria and eukaryotic algae have an alternative electron transfer protein as a substitute for plastocyanin, the hemoprotein cytochrome c_6, which is used when copper is not readily available (Wood 1978). It is believed that cytochrome c_6 is a more ancient protein than

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

plastocyanin, with the latter evolving after increasing atmospheric oxygen concentrations led to a decrease in the ready availability of iron in the environment (De la Rosa et al. 2002). Green plants were believed to have lost cytochrome c_6, retaining only plastocyanin (Kerfeld and Krogmann 1998).

However, in 2002 a homolog of cytochrome c_6 was found in green plants (Gupta et al. 2002; Wastl et al. 2002). This protein was subsequently named cytochrome c_{6A} (Wastl et al. 2004; fig. 1*A*). The sequence of cytochrome c_{6A} was found to be highly similar to that of c_6, with a major difference that cytochrome c_{6A} contains a 12-amino acid insertion in a loop region of the protein (fig. 1*B*). This insertion has been named the loop insertion peptide (LIP) (Howe et al. 2006), and contains two cysteines that form a disulfide bridge (Marcaida et al. 2006). Further homologs of cytochrome c_{6A} (in addition to the conventional cytochrome c_6) were then discovered in cyanobacteria, and named cytochromes c_{6B} and c_{6C} (Nomura 2001; Bialek et al. 2008; fig. 1*A* and *B*). These cytochromes were split into B and C homologs based on a phylogenetic analysis, which showed that cytochrome c_{6B} shared a more recent common ancestry with cytochrome c_{6A}, and cytochrome c_{6C} shared a more recent common ancestor with cytochrome c_6 (Bialek et al. 2008).

Cytochromes c_{6A}, c_{6B}, and c_{6C} have a redox midpoint potential around 200 mV lower than cytochrome c_6, suggesting that cytochromes c_{6A}, c_{6B}, and c_{6C} are unable to oxidize cytochrome *f* and have a different function from cytochrome c_6 (Molina-Heredia et al. 2003; Bialek et al. 2008, 2014). This suggestion of a difference in function was supported by studies on the reaction between cytochrome c_{6A} and photosystem I in vitro and the demonstration that plastocyanin is essential in plants (Molina-Heredia et al. 2003; Weigel et al. 2003). The difference in redox midpoint potential between cytochromes c_{6A} and c_6 is proposed to be largely due to a single amino acid residue, found at position 52 in *Arabidopsis thaliana* cytochrome c_{6A} (Marcaida et al. 2006; fig. 1*B*). In the low redox midpoint potential cytochrome c_6-like proteins, this residue is hydrophobic (leucine, isoleucine, or valine), with cytochrome c_6 having a conserved glutamine in the same position. Substituting the *A. thaliana* cytochrome c_{6A} valine 52 with a glutamine has been shown to increase the redox midpoint potential of the protein by around 100 mV (Worrall et al. 2007). The function of these low redox midpoint cytochrome c_6-like proteins is currently unclear, though a role in alternative pathways in electron transfer has been proposed (Howe et al. 2016).

The Current Model of Cytochrome c_6 Family Ancestry

The current hypothesis for the evolution of the cytochrome c_6 family in photosynthetic organisms has been outlined by Howe et al. (2016) (fig. 1*C*). The model suggested that duplication(s) of cytochrome c_6 in an ancestral cyanobacterium led to the genesis of the low redox midpoint potential cytochromes grouped under the umbrella-term cytochrome c_{6ABC}. Phylogenetic analysis by Bialek et al. (2008) suggested that at least two duplications had occurred in cyanobacteria, resulting in cytochromes c_{6B} and c_{6C}. Following primary endosymbiosis, cytochrome c_6 was lost in the green plant lineage leaving only a low redox midpoint potential sequence, cytochrome c_{6A}. Secondary endosymbiosis involving the green lineage (e.g., as seen for *Euglena*), was believed to have failed to transfer the low redox midpoint potential cytochrome c_6. In contrast, cytochrome c_{6ABC} was believed to have been lost in the red algal and glaucophyte lineages (which contain primary plastids) sometime after the origin of the haptophytes (containing a secondary plastid; Yoon et al. 2002), which retain both cytochrome c_6 and c_{6ABC}. A recent study, however, has identified a cytochrome c_{6BC} homolog in the glaucophyte *Cyanophora paradoxa* (Kleiner et al. 2021).

Aims of the Study

With the availability of more sequence data, this study expanded the search for cytochrome c_6 family sequences in a wider range of photosynthetic taxa, both prokaryotic and eukaryotic. We particularly wished to identify whether cyanobacterial cytochrome c_{6BC} proteins were derived from cytochrome c_6 or vice versa, what the evolutionary relationship is between cytochrome c_{6B} and cytochrome c_{6C}, and how widely distributed the cytochrome c_{6ABC} family is among eukaryotes.

Results

Mapping Cytochromes c_6, c_{6B}, and c_{6C} on an Established Cyanobacterial Species Tree Shows That c_{6B} and c_{6C} Are Orthologs That Arose from a Single c_6 Gene Duplication Event

To examine the distribution of cytochromes c_6, c_{6B}, and c_{6C} across cyanobacteria, the presence or absence of c_6 family cytochrome sequences was mapped onto a phylogenetic tree of cyanobacterial species inferred from a concatemer of conserved sequences (Walter et al. 2017; fig. 2). This phylogenetic framework represents a comprehensive phylogenomic analysis of cyanobacteria based on 31 conserved gene markers, using Maximum Likelihood (ML). Putative cytochrome $c_{6B/C}$ sequences were found by database searching and defined as cytochrome $c_{6B/C}$ if they had both an appropriately located haem-binding motif (CXXCH; Barker and Ferguson 1999) and a valine, leucine, or isoleucine rather than glutamine at the equivalent of position 52 in *A. thaliana*. (The presence of valine at this position was linked to a lower redox midpoint potential relative to cytochrome c_6 [Worrall et al. 2007; Bialek et al. 2008].)

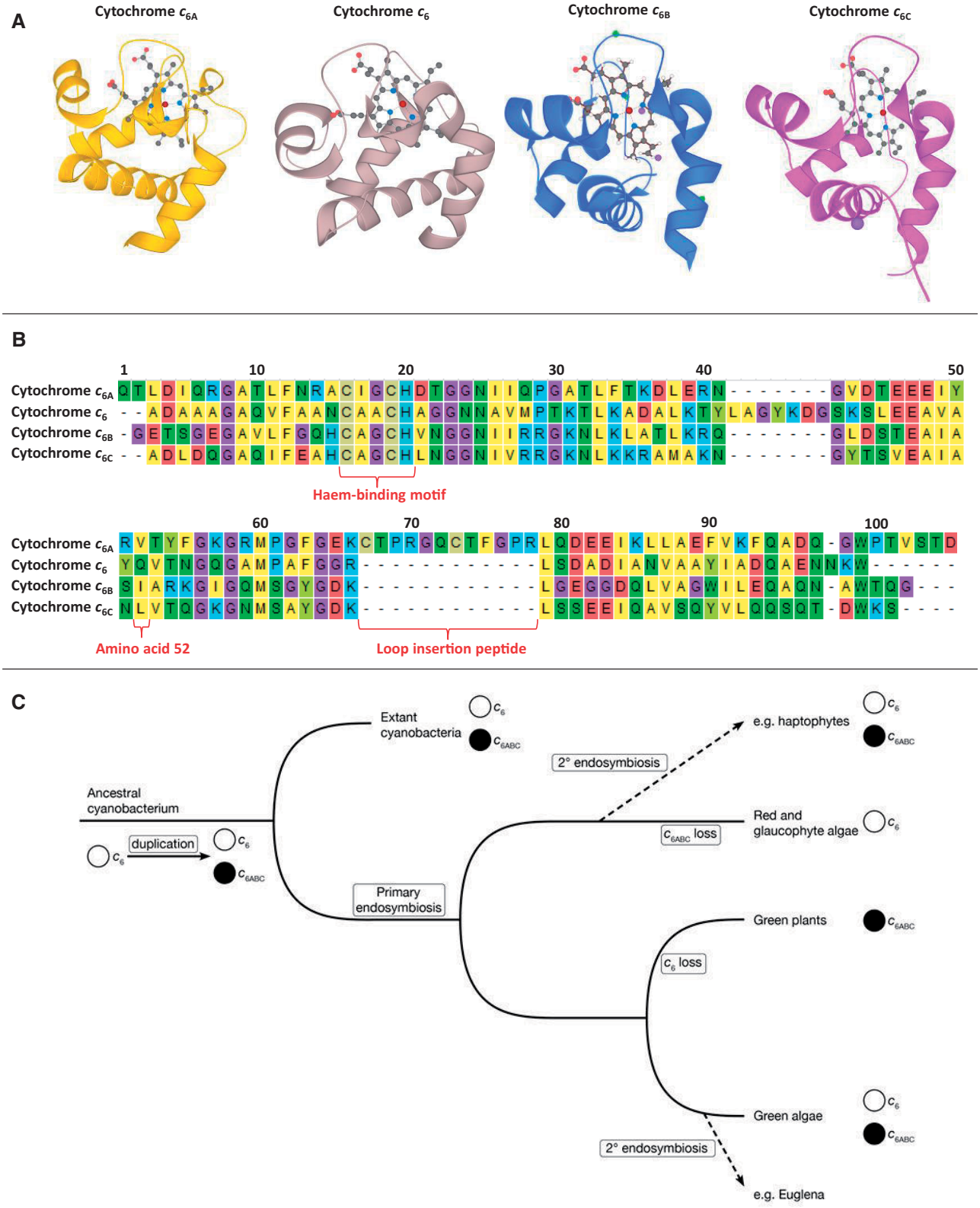

Fig. 1.—(*A*) X-ray crystal structures of cytochromes c_{6A} (yellow), c_6 (beige), c_{6B} (blue), and c_{6C} (magenta). Secondary structure for each protein is shown in ribbon form, and the haem prosthetic groups shown in ball and stick (carbon—black, oxygen—red, nitrogen—blue, iron—deep red, and sodium—purple). (*B*) Protein sequence alignment of cytochromes c_{6A}, c_6, c_{6B}, and c_{6C} from *A. thaliana* (accession Q93VA3.1), *Synechococcus* sp. PCC 7002 (accession

With this expanded sampling, the sequences identified as cytochrome c_{6B} by Bialek et al. (2008) were found exclusively in one clade, which contained cyanobacteria of the genera *Prochlorococcus* or *Parasynechococcus*, as shown in figure 2. In contrast, the sequences previously identified as cytochrome c_{6C} (Bialek et al. 2008) were widespread across the cyanobacterial species tree, but not found in the genera *Prochlorococcus* or *Parasynechococcus*. In addition, no species in the tree contained more than one cytochrome $c_{6B/C}$-like sequence. These observations suggest that the prior separation observed between cytochrome c_{6B} and c_{6C} could be accounted for by taxon sampling, without needing to propose them as two separate families.

The cyanobacterial tree shows a split (labeled with a red arrow) separating taxa which have only cytochrome c_6 from those which also have a cytochrome c_{6B} or c_{6C} sequence. This branch point separates the basally diverging *Gloeobacter violaceus* PCC7421 (Criscuolo and Gribaldo 2011; Mareš et al. 2013) and a few other taxa from the rest. This distribution suggests that cytochrome c_6 appeared first, and that cytochromes c_{6B} and c_{6C} may have arisen through duplication and neofunctionalization of cytochrome c_6.

A Phylogenetic Tree Using a Wider Taxon Selection Suggests a Single Origin for Cytochromes c_{6B} and c_{6C}

To investigate further the hypothesis of cytochromes c_{6B} and c_{6C} being duplicates of cytochrome c_6 with a new function, a phylogenetic tree was inferred from an alignment of cytochromes c_6, c_{6B}, and c_{6C} sequences covering all the organisms used in two independent phylogenetic analyses of the cyanobacterial lineage (Schirrmeister et al. 2015; Walter et al. 2017). A condensed tree is shown in figure 3*A* (the full tree can be found in supplementary fig. 1, Supplementary Material online). As the sequences are short, a Neighbor-Net analysis was also performed (fig. 3*B*) and showed that the data were treelike, and that tree-based phylogenetic analysis was appropriate (Huson and Bryant 2006).

The cytochromes predicted to have a low redox midpoint potential, including those assigned as cytochromes c_{6B} and c_{6C} previously, all grouped to the exclusion of the predicted cytochrome c_6 sequences in both the phylogenetic tree (bootstrap value of 84%) and the Neighbor-Net analysis, and maintained a similar general topology to that of the cyanobacterial tree of figure 2. This distribution showed cytochrome c_{6B} as a clade derived from within the cytochrome c_{6C} clade, as with figure 2. Once again, there was no evidence of both cytochromes c_{6B} and c_{6C} within the same organism. (*Crocosphaera watsonii* has been shown to have two low redox midpoint potential cytochrome c_6 sequences, but both were assigned as cytochrome c_{6C} in prior studies [Bialek et al. 2008].) These observations suggest a single origin for the cytochrome c_{6BC} family, and that they are orthologs rather than paralogs. It is worth noting that the bootstrap values in this tree were considerably lower than those in the tree inferred by Walter et al. (2017), which is to be expected with a larger number of taxa for sequences of short sequence length (Rokas and Carroll 2005) such as with the cytochrome c_6 family peptides. However, the tree-like appearance of the Neighbor-Net analysis inferred from the same sequence alignment suggests that phylogenetic inference is appropriate.

The neighboring open-reading frames of cytochromes c_{6B} and c_{6C} in cyanobacterial genomes were compared (data not shown). Most of the species that possess a cytochrome c_{6B} were observed to have neighboring genes coding for Nif1 domain-containing, YciI family, or DUF3136 domain-containing proteins, except for *Prochlorococcus marinus* str. MIT 9312 and MIT 9301. The most closely related species possessing a cytochrome c_{6C} had different neighboring genes from those in the cytochrome c_{6B} species. However, it is difficult to determine whether this difference in genetic neighborhoods is due to cytochromes c_{6B} and c_{6C} being paralogs, or due to the overall similarity between the genomes of species containing cytochrome c_{6B}, which comprise only two genera of cyanobacteria. Additionally, there was a high diversity of genetic neighborhoods amongst cytochrome c_{6C} possessing species, which represent a wider range of cyanobacteria. There was therefore no evidence from synteny to indicate that cytochromes c_{6B} and c_{6C} are paralogs.

Taken together, there is no evidence that would support a functional differentiation between cytochromes c_{6B} and c_{6C}. Cytochrome c_{6B} is found only in a clade of organisms known for a high protein substitution rate (Dufresne et al. 2005), cytochromes c_{6B} and c_{6C} both have a low redox midpoint potential, share common ancestry to the exclusion of cytochrome c_6, and are not found together in one organism. The distinction between cytochromes c_{6B} and c_{6C} does not seem to represent functional divergence, and we propose to refer to all as cytochromes c_{6B} in future.

Distribution of Cytochrome c_6 Family Members across Photosynthetic Eukaryotes

The recent sequencing of genomes and transcriptomes of a wider range of eukaryotic photosynthetic organisms allowed

O30881.1), *Synechococcus* WH8103 (CRY92441.1), and *Synechococcus* sp. PCC 7002 (accession AAN03578.1) respectively. The sequences have their putative signal peptides excluded. Amino acids are colored with yellow—hydrophobic residues, green—polar residues, beige—cysteines, blue—positively charged residues, and red—negatively charged residues, and the haem-binding motif (CXXCH), the LIP, and amino acid 52 are indicated below the alignment. Figure uses crystallography and sequence data from Marcaida et al. (2006), Bialek et al. (2009), and Zatwarnicki et al. (2014) (PDB cytochrome c6: 3DR0, c6A: 2CE0, c6B: 4KMG, c6C: 4EIE). (*C*) Current model of cytochrome c_6 family evolution in photosynthetic organisms. Adapted from Howe et al. (2016).

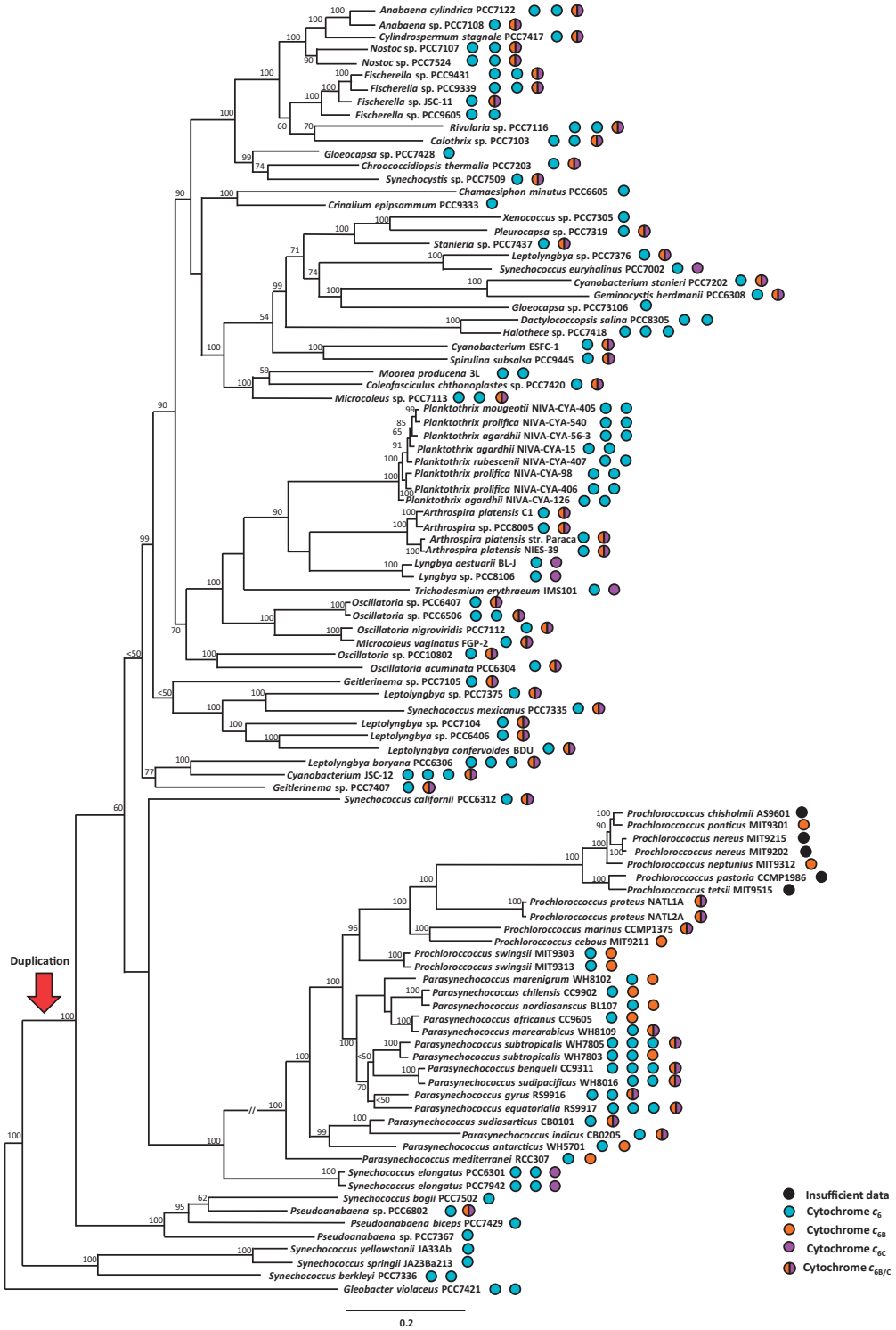

Fig. 2.—Cyanobacterial species tree with presence of cytochromes c_6, c_{6B}, and c_{6C} mapped onto it (colored blue, orange, and magenta, respectively). Circles with both orange and magenta semicircles contain a putative low redox midpoint potential cytochrome c_6, which was not included in Bialek et al. (2008). Proteins represented by orange or purple full circles were described as c_{6B} or c_{6C}, respectively by Bialek et al. (2008). Black circles represent species whose peptide or nucleotide data are not readily available to probe. Tree is reproduced from Walter et al. (2017). The potential branch where neofunctionalisation led to the evolution of cytochrome c_{6B} is indicated with a red arrow. Scale bar represents branch length. Bootstrap values were calculated as a percentage using 1,000 iterations Walter et al. (2017). Accessions of the sequences used for this figure can be found in supplementary table 1, Supplementary Material online.

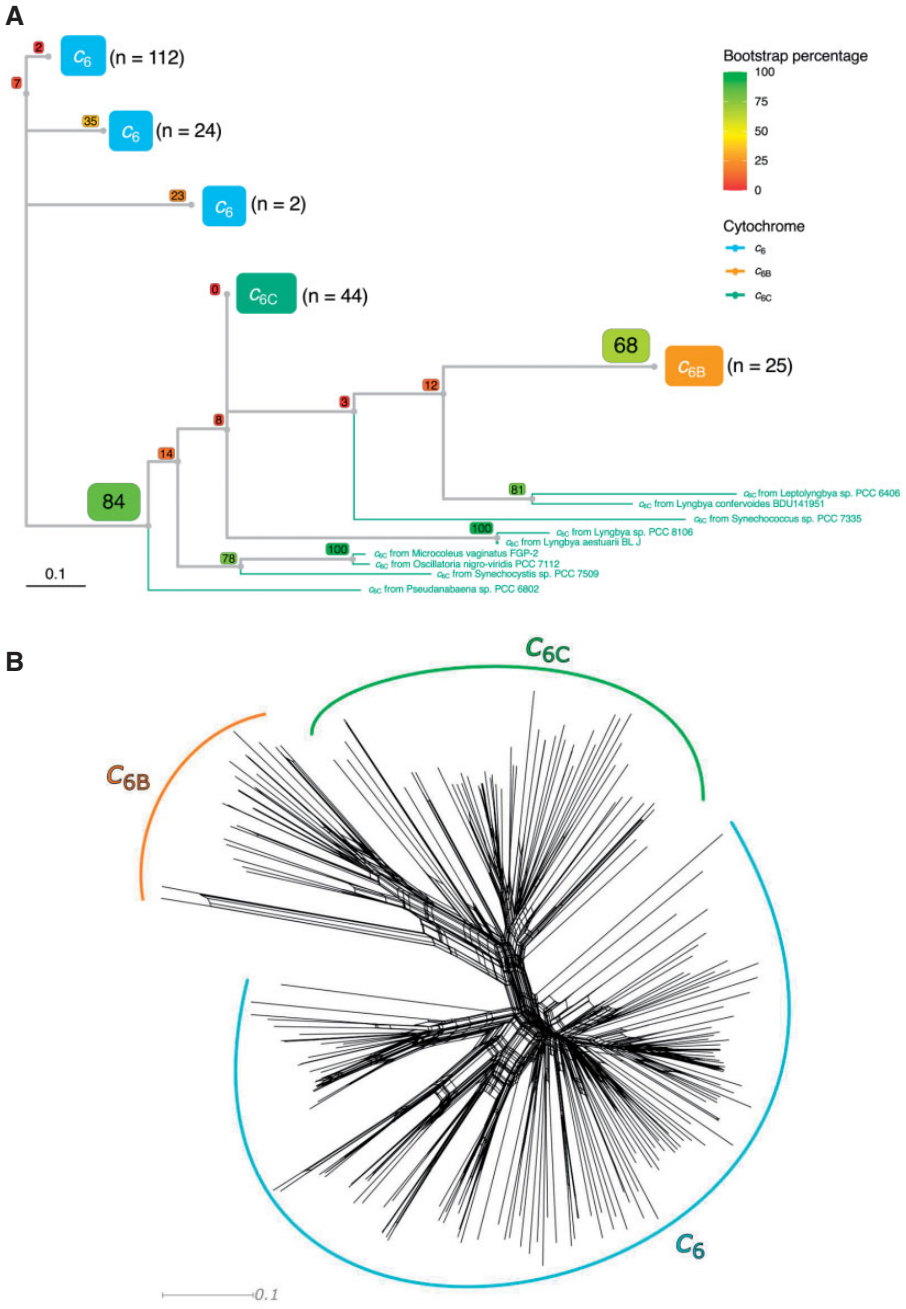

Fig. 3.—Condensed phylogenetic tree (*A*) and Neighbor-Net splits graph (*B*) inferred from an alignment of cytochrome c_6, c_{6B}, and c_{6C} peptide sequences from cyanobacterial species (colored blue, orange, and green, respectively). Alignments were performed using MUSCLE algorithm and can be found in the supplementary information along with accessions for each sequence used. The phylogenetic tree was built using ML inference using a WAG model with Gamma distribution and invariable sites (WAG + G + I). Bootstrap values for each branch point, using 100 iterations, are shown in colored boxes. The *n* value next to each group represents the number of sequences found within each clade. The full tree is shown in supplementary figure 1, Supplementary Material online and the alignment from which the tree was inferred can be found in supplementary table 2, Supplementary Material online. SplitsTree4 was used to obtain the Neighbor-Net splits graph.

for a more thorough search for c_6-like cytochromes, including cytochrome c_{6A}. Protein and nucleotide databases across eukaryotes were searched using BLAST for cytochrome c_6 family sequences. Sequences recovered were defined as cytochrome c_{6A} if they contained a hydrophobic residue (valine, leucine, or isoleucine) at the equivalent of position 52, indicating a low redox midpoint potential, and an insertion containing a disulfide bridge (the LIP) in the loop region

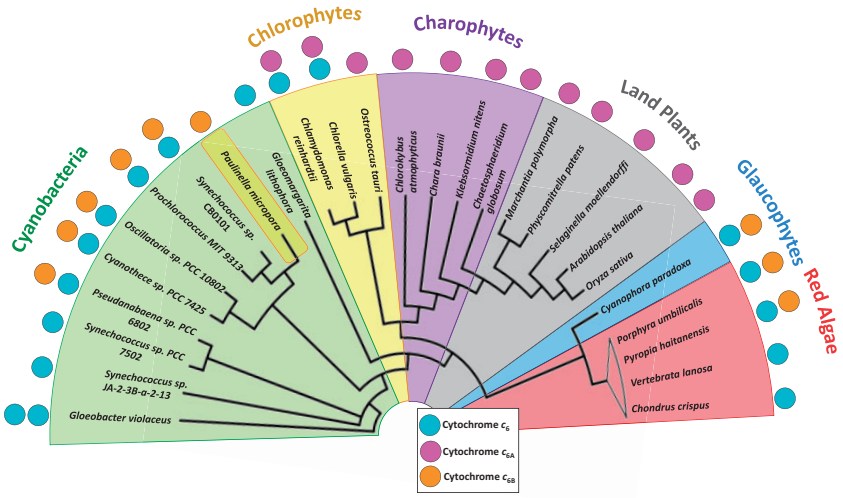

Fig. 4.—Distribution of c_6-type cytochromes across photosynthetic lineages. Presence of a colored circle adjacent to a species name indicates that a sequence of the relevant cytochrome was found in sequence database searches, with multiple copies of the same colored circle indicating a potential paralog. *Paulinella chromatophora* is highlighted to segregate it from the cyanobacteria. Phylogenetic tree branch lengths are not to scale. Diagram based on phylogenetic tree from Ponce-Toledo et al. (2017). Accession numbers for individual gene sequences can be found in supplementary table 3, Supplementary Material online.

compared with cytochrome c_6. Sequences recovered were defined as cytochrome c_{6B} if they contained the hydrophobic residue implying a low redox midpoint but not the LIP.

Distribution of Cytochrome c6 after Primary Endosymbiosis

The distribution of cytochromes c_6, c_{6B}, and c_{6A} across cyanobacteria and in eukaryotes after primary endosymbiosis was mapped onto a phylogenetic tree based on an alignment of concatemers of plastid genes and cyanobacterial homologs (fig. 4). The presence of cytochromes c_6 and c_{6B} in the glaucophyte and red algal lineages suggests that the cyanobacterium involved in the primary endosymbiosis event contained both a cytochrome c_6 and a c_{6B}.

The results indicate that after primary endosymbiosis, cytochrome c_{6B} was replaced by cytochrome c_{6A} in the green plant and algal lineage. This suggests that cytochrome c_{6A} was derived from cytochrome c_{6B}, possibly through an insertion of the LIP early in the green chloroplast lineage. The insertion of the LIP into an existing sequence rather than duplication and divergence is supported by the observation that no species have been found to contain both a cytochrome c_{6A} and a c_{6B} sequence. Although cytochrome c_6 was identified in some chlorophyte species, it was not identified in any charophytes or land plants, suggesting that the loss of cytochrome c_6 occurred in the ancestor to the charophyte lineage. This in turn has resulted in land plants exclusively containing cytochrome c_{6A}. In contrast, the glaucophytes and many red algal species have retained the original cytochrome c_6. However, some red algal species (though not all) appeared to have lost cytochrome c_{6B}, for example *Chondrus crispus*. Finally, the eukaryotic protist *Paulinella* contains a cytochrome c_{6B}-like sequence. This is likely to reflect the recent, independent primary endosymbiosis of a cyanobacterium that gave rise to the *Paulinella* chloroplast (Marin et al. 2005; Yoon et al. 2006), although the *Paulinella* line also appears to have lost cytochrome c_6.

Distribution of Cytochrome c6 Family Members after Secondary Endosymbioses

Many photosynthetic eukaryotes contain chloroplasts of secondary origin. We therefore searched for the presence of cytochromes c_6, c_{6B}, and c_{6A} in these organisms. *Euglena gracilis*, which contains a chloroplast of secondary green origin (Turmel et al. 2009), was predicted to contain a cytochrome c_{6A} sequence in addition to cytochrome c_6 (Novák Vanclová et al. 2020). The chlorarachniophytes, a class of Rhizaria with a secondary chloroplast of green origin (Rogers et al. 2007), on the other hand only had a cytochrome c_6 sequence and no evidence of cytochrome c_{6A}.This suggests that either the green algal endosymbiont of the chlorarachniophytes did not have cytochrome c_{6A} or that the gene was lost after secondary endosymbiosis.

Different organisms with a secondary red chloroplast also varied in cytochrome c_6 family gene distribution. Haptophytes, cryptomonads, and some ochrophytes, which contain a chloroplast of a red algal origin (Yoon et al. 2002), contained cytochrome c_6 and c_{6B} sequences, as expected. However, many ochrophytes and some haptophytes and cryptomonads had no evidence of c_{6B} sequences. This suggests that the red algal endosymbiont to haptophytes, cryptomonads, and ochrophytes had retained cytochrome c_{6B}, and that the gene was lost afterwards downstream in certain

lineages, although widespread lateral transfer cannot be excluded.

The situation in dinoflagellate algae is complex. The peridinin dinoflagellates contain a chloroplast of secondary red origin (Dorrell and Howe 2015). Two peridinin dinoflagellates (*Amphidinium carterae* and *Symbiodinium microadriaticum*) contain a cytochrome c_6 sequence and what appeared to be a cytochrome c_{6A} sequence. In contrast, *Karlodinium veneficum*, a fucoxanthin dinoflagellate (which obtained its chloroplast via loss of the red algal chloroplast and serial endosymbiosis of a haptophyte [Dorrell and Howe 2015; Klinger et al. 2018]), contains a cytochrome c_6 and cytochrome c_{6B}. The existence of a cytochrome c_{6B} in the *Karlodinium* lineage, whose chloroplast is of red algal origin, is not surprising. However, the peridinin dinoflagellate chloroplast is also of red algal origin, so the apparent existence of cytochrome c_{6A} in this lineage is unexpected.

The sequences resembling cytochrome c_{6A} in peridinin dinoflagellate algae were compared with those of other cytochrome c_{6A} proteins (fig. 5). This revealed that the dinoflagellate sequences show some sequence similarity with the cytochromes c_{6A} from the green chloroplast lineage. The LIP insertion, however, shows very little sequence similarity between dinoflagellates and green plants, except for the two characteristic cysteine residues. It should also be noted that the dinoflagellate sequences are longer than the cytochrome c_{6A} sequences from green plants, with the first position of each of the dinoflagellate sequences in figure 5 being residues 123 and 50 for *A. carterae* and *S. microadriaticum*, respectively. These observations suggest that the putative dinoflagellate cytochrome c_{6A} sequences have a functional similarity to cytochrome c_{6A} from green plants, but are likely to have an independent origin.

A summary of the photosynthetic eukaryotes and the presence of each cytochrome c_6 family sequence discovered can be found in table 1. Taxon IDs of clades searched are in supplementary table 4, Supplementary Material online, and accession numbers of the sequences found are in supplementary table 1, Supplementary Material online.

Cytochrome c_{6A} Arose from Cytochrome c_{6B} Rather Than Directly from Cytochrome c_6

To test if cytochrome c_{6A} (non-dinoflagellate) arose from cytochrome c_{6B}, rather than by independent modification of a cytochrome c_6, a phylogenetic tree was inferred using cytochrome c_{6A} sequences from eukaryotic algae and green plants, together with cytochromes c_6 and c_{6B} from a wide range of cyanobacteria (fig. 6 shows a condensed version of this tree, and the full tree can be found in supplementary fig. 2, Supplementary Material online). Cytochromes c_{6A} and c_{6B} grouped together to the exclusion of cytochrome c_6 (bootstrap value of 75), suggesting that cytochrome c_{6A} shares most recent common ancestry with cytochrome c_{6B}. This supports the conclusion above (fig. 4) that cytochrome c_{6A} was derived from cytochrome c_{6B} through an insertion event in the loop region, rather than independently of cytochrome c_{6B}. Once again, the bootstrap values in the tree were considerably lower than those of the species tree established by Walter et al. (2017), but this is to be expected as the c_6 family cytochrome sequences are short. (A Neighbor-Net splits graph for this alignment was also constructed (not shown) but was less clearly resolved and did not give any additional information.)

Discussion

Cytochromes c_{6B} and c_{6C} Are Orthologs

The original differentiation of cytochromes c_{6B} and c_{6C} was based on the sequence data available at the time (Bialek et al. 2008). However, now that more genomic sequence data are available, cytochrome c_6 family sequences from a larger range of taxa can be analyzed. Our analysis indicates that the distinction between cytochromes c_{6B} and c_{6C} can be accounted for by taxon sampling rather than differences in function. (Although cytochromes c_{6B} and c_{6C} lie on opposite sides of the root of the cytochrome c_6 family in the tree of Bialek *et al.* [2008], the placing of the root should be viewed with caution given that it depends on other c-type cytochromes of very different function from the cytochrome c_6 family.) In addition, as the crystal structures, surface charge distribution, and redox midpoint potentials of cytochromes c_{6B} and c_{6C} are notably similar (Zatwarnicki et al. 2014; fig. 1*A* and *B*), it seems likely that cytochromes c_{6B} and c_{6C} perform a similar function and are thus orthologs.

Two Independent Origins of c_{6A}

Although the presence of cytochrome c_{6A} in plants and green algae has been known for time, the presence of cytochrome c_{6A} in peridinin dinoflagellates was unexpected. Dinoflagellates contain chloroplasts of secondary or tertiary origin, depending on species. The chloroplast found in *S. microadriaticum* and *A. carterae* contains peridinin, and is believed to represent the ancestral dinoflagellate chloroplast. This chloroplast was most likely obtained through secondary endosymbiosis of red algae (Dorrell and Howe 2015). Therefore, these species might be expected to contain a cytochrome c_{6B}. Instead, the peridinin dinoflagellates contain a cytochrome c_{6A}-like sequence. Two hypotheses for this are 1) the result of lateral gene transfer from an organism with cytochrome c_{6A} and the loss of the cytochrome c_{6B} or 2) the insertion of a LIP-like sequence into an existing cytochrome c_{6B} sequence. Although lateral gene transfer to dinoflagellates from other organisms has been well documented (Takishita et al. 2003; Hackett et al. 2005; Chan et al. 2012; Wisecaver et al. 2013), and it is difficult to exclude conclusively lateral transfer of cytochrome c_{6A} into the dinoflagellates, the low sequence similarity between the dinoflagellate c_{6A} and those from the green

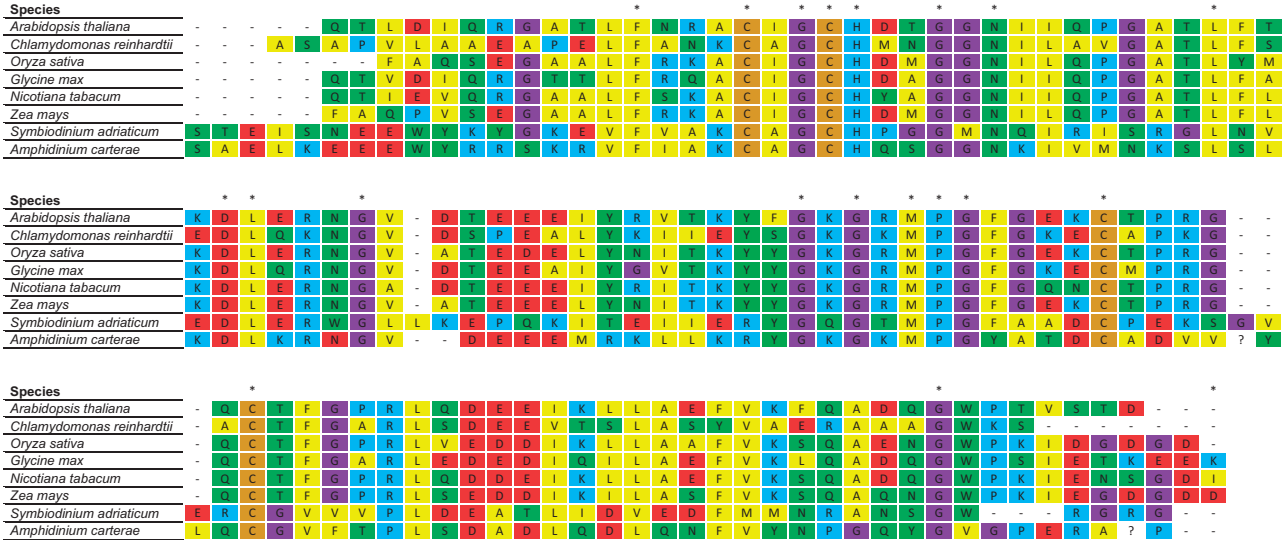

FIG. 5.—MUSCLE alignment of cytochrome c_{6A} from green eukaryotic lineages and the putative cytochrome c_{6A} sequences from the dinoflagellates *S. microadriaticum* and *A. carterae*. Amino acids are colored such that yellow—hydrophobic residues, green—polar residues, blue—positively charged residues, beige—cysteines, purple—glycines, light green—tyrosine, and red—negatively charged residues. Dashes indicate inserted gaps. Asterisks represent conserved residues. The sequences of the mature dinoflagellate proteins have been cut for a clearer depiction of the alignment, with the *S. microadriaticum* sequence beginning at amino acid 50 and the *A. carterae* sequence beginning at amino acid 123. Accessions used: *A. thaliana* (AT5G45040.1), *C. reinhardtii* (XP_001692119.1), *O. sativa* (EAZ04378.1), *G. max* (KRH50430.1), *N. tabacum* (XP_016489567.1), *Z. mays* (ACN28933.1), *S. microadriaticum* (OLP91854.1), and *A. carterae* (CF065358.1).

Table 1

The Presence or Absence of Cytochrome c_6 Family Members across Photosynthetic Eukaryotes

Supergroup	Clade[a]	Plastid	Cytochrome		
			c_6[b]	c_{6A}[b]	c_{6B}[b]
Glaucophyta	Glaucophyta ($n = 1$)	Primary	✓ (1)	O	✓ (1)
Rhodophyta	Rhodophyta ($n = 190$)	Primary	✓ (190)	O	✓ (4)
Viridiplantae	Chlorophyta ($n = 42$)	Primary	✓ (32)	✓ (34)	O
Viridiplantae	Streptophyta ($n = 217$)	Primary	O	✓ (215)	O
SAR—Rhizaria	*Paulinella* ($n = 3$)	Primary cyanobacteria	O	O	✓ (3)
	Chlorarachniophyta ($n = 2$)	Secondary green	✓ (2)	O	O
Cryptomonads	Cryptomonads ($n = 7$)	Secondary (rhodophyta)	✓ (2)	O	✓ (1)
Haptista	Haptophyta ($n = 6$)	Secondary (rhodophyta)	✓ (6)	O	✓ (3)
SAR—Heterokonta	Ochrophyta ($n = 80$)	Secondary (rhodophyta)	✓ (78)	O	✓ (2)
SAR—Alveolata	Dinoflagellata ($n = 7$)	Secondary (rhodophyta)/ Tertiary (haptophyta)	✓ (6)	✓ (2)	✓ (1)
Discoba—Euglenozoa	Euglenids ($n = 2$)	Secondary (chlorophyta)	✓ (2)	✓ (1)	O

[a]Where *n* represents the number of organisms searched.
[b]The number of sequences found is given in brackets.

plant lineage would suggest an independent LIP insertion into cytochrome c_{6B} in dinoflagellates is more likely.

The Current Model of Cytochrome c_6 Family Ancestry

The analysis of the cytochromes c_6, c_{6A}, and c_{6B} in this study provides a revised evolutionary model for cytochrome *c* homologs consistent with more extensive taxon sampling (fig. 7). As more anciently diverged cyanobacterial species such as *Gloeobacter* appear to contain cytochrome c_6 exclusively, this suggests that the low redox midpoint potential cytochromes are more recent than cytochrome c_6. A duplication of cytochrome c_6, followed by point mutations that lowered the redox midpoint potential, led to the evolution of cytochrome c_{6B}. This is supported by the presence of both cytochromes c_6 and c_{6B} in most extant cyanobacteria today.

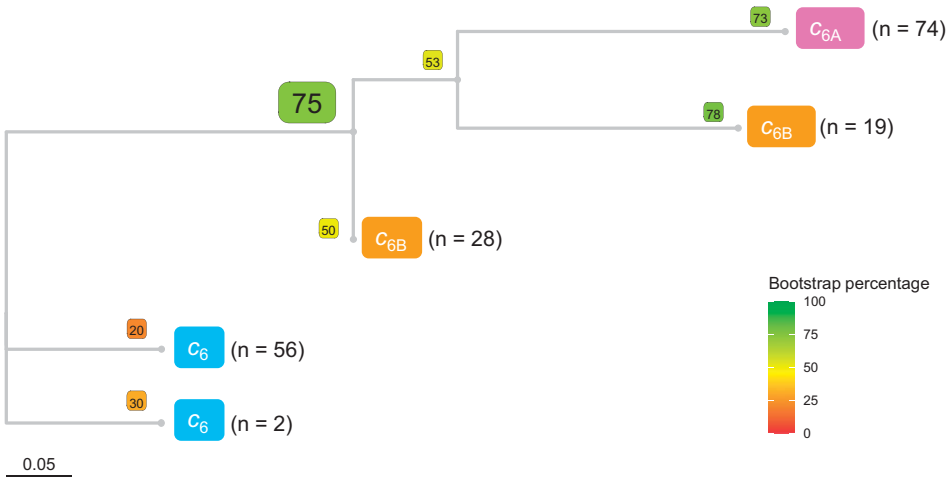

Fig. 6.—Condensed phylogenetic tree inferred from an alignment of cytochrome c_6, c_{6A}, and c_{6B} peptide sequences (colored blue, pink, and orange, respectively) from eukaryotic algae, green plants, and cyanobacteria. Alignments were performed using MUSCLE algorithm and can be found in the supplementary information along with accessions for each sequence used. The tree was built using ML inference using a WAG model with Gamma distribution and invariable sites (WAG + G + I). Bootstrap values for each branch point, using 100 iterations, are shown in colored boxes. The full tree is shown in supplementary figure 2, Supplementary Material online and the alignment from which the tree was inferred can be found in supplementary table 5, Supplementary Material online.

At primary endosymbiosis, giving rise to the red, green, and glaucophyte chloroplasts, the genes were transferred to photosynthetic eukaryotes.

Red algal lineages and the glaucophytes contain cytochromes c_6 and c_{6B}, although some red algal species have lost cytochrome c_{6B}. In the green chloroplast lineages, cytochrome c_{6B} was replaced by cytochrome c_{6A}. This was probably due to an insertion of the LIP into cytochrome c_{6B}, as cytochrome c_{6A} is monophyletic within cytochrome c_{6B} (fig. 6). In many chlorophyte species, both cytochromes c_6 and c_{6A} are present. In the charophytes, ancestors to the green land plants, cytochrome c_6 was lost. In consequence, land plants contain only a cytochrome c_{6A}.

Organisms containing chloroplasts of secondary origin appear have inherited their cytochrome c_6 family genes from the relevant endosymbiont. The haptophytes obtained both cytochromes c_6 and c_{6B} from the red algal chloroplast, and these genes were transferred to the fucoxanthin dinoflagellates following serial endosymbiosis. In contrast, the peridinin dinoflagellates, containing chloroplasts of secondary red origin, probably converted the cytochrome c_{6B} into a cytochrome c_{6A}-like protein through the insertion of a novel LIP. With green plastid secondary endosymbiosis, genes for both cytochromes c_6 and c_{6A} were passed to the euglenids (Novák Vanclová et al. 2020).

Overall, it is clear that the low potential cytochrome c_{6AB} family is widely, but not universally, present among oxygenic photosynthetic organisms. It is unlikely to be essential under all conditions, but there is no obvious environmental feature common to those organisms that retain a member of the family. The function of the cytochrome c_{6AB} family remains to be determined.

Materials and Methods

Construction of Phylogenetic Trees

A cytochrome c_{6B} sequence (BAD79758.1) was used for searching the "non-redundant protein sequences (*nr*)" database with the NCBI BLASTp algorithm, limiting the results to the organisms used in two independent phylogenetic analyses of the cyanobacterial lineage (Schirrmeister et al. 2015; Walter et al. 2017). A cytochrome c_{6A} protein sequence (AED95193.1) was used to search the *nr* protein database and the "nucleotide collection (*nr/nt*)" nucleic acid database with the BLASTp and tBLASTn algorithms respectively, limiting the results to the orders used in a phylogenetic analysis of green plants (Ruhfel et al. 2014). The resulting sequences of both searches were downloaded from the NCBI BLAST result page.

The retrieved peptide sequences were imported into MEGA7 (Kumar et al. 2016). Sequences were deleted from the selection if they were too short or too long to be a valid cytochrome $c_{6(A/B)}$ sequence (less than 80 and more than 200 amino acids before N-terminal targeting peptide trimming) or did not have a CxxCH haem binding motif. The sequences were aligned using the MUSCLE algorithm with UPGMA clustering method and a gap-opening penalty of −2.9 and no gap extension penalty. Subsequently the putative signal peptides were trimmed from the sequences.

The WAG model with gamma distribution and invariable sites (WAG+G + I) (Whelan and Goldman 2001) was determined to be optimal for tree inference with maximum likelihood (ML) by the "Find Best DNA/Protein Models (ML)" tool in MEGA7 and was thus used in the algorithm parameters for tree inference. Statistical testing was performed using the

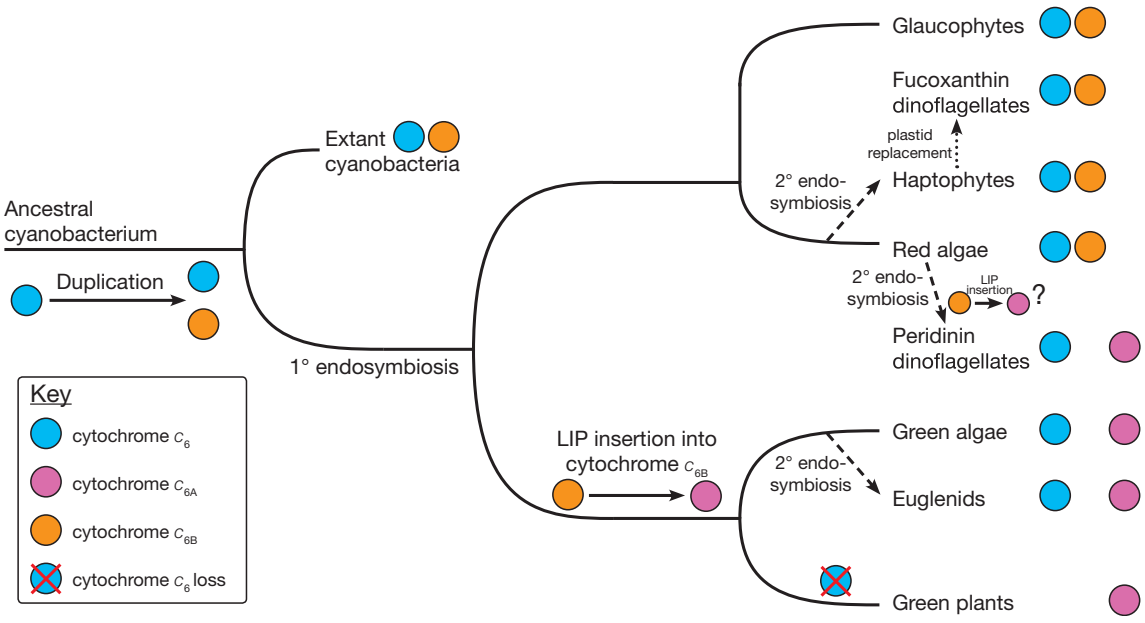

Fig. 7.—Updated ancestry model for the cytochrome c_6 family in photosynthetic organisms.

bootstrap method with 100 iterations. The final trees were visualized using Treeio in RStudio (Wang et al. 2020). The accessions of the sequences used for inference of phylogenetic trees can be found in supplementary tables 2 and 5, Supplementary Material online.

The same aligned sequences were imported into SplitsTree4 (Huson and Bryant 2006), and the software was used to build a Neighbor-Net splits graph.

Database Queries for Peptide Sequences

Searches for protein sequences homologous to the cytochrome c_6 family, or nucleotide sequences encoding them, were performed using NCBI BLAST both in BLASTp and tBLASTn (https://www.ncbi.nlm.nih.gov/). The cytochrome c_6 c_{6A}, c_{6B}, and c_{6C} peptide sequences (without targeting) used in both BLASTp and tBLASTn searches were from accessions ALJ67080.1, AED95193.1, AAP99622.1, and ACB00369.1, respectively. For BLASTp searches, the database searched was "Non-redundant protein sequences (nr)" with default parameters. For tBLASTn searches, the databases searched were "nucleotide collection (nr/nt)," "Whole-genome shotgun contigs (wgs)," and "Expressed sequence tags (est)" with default parameters. Having identified a putative cytochrome c_6 sequence, each organism that provided a query sequence was searched again to confirm the query as the best hit.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was supported by a Biotechnology and Biological Sciences Research Council DTP PhD studentship (BB/M011194/1 to B.S.); the German Academic Scholarship, the Gates Cambridge Trust and the Benn W Levy Trust to D. K.; and the Gordon and Betty Moore Foundation (GBMF4976, doi: https://doi.org/10.37807/GBMF4976 and GBMF9358, doi: https://doi.org/10.37807/GBMF9358 to R.E.R.N. and C.J.H.).

Data Availability

The data underlying this article are available at the NCBI. References for individual genes are given in the article and/or in the Supplementary Material online. All alignments are provided in the supplementary Material online.

Literature Cited

Barker PD, Ferguson SJ. 1999. Still a puzzle: why is haem covalently attached in c-type cytochromes? Structure 7(12):R281–R290.

Bialek W, Krzyda S, Jaskolski M, Szczepaniak A. 2009. Atomic-resolution structure of reduced cyanobacterial cytochrome c_6 with an unusual sequence insertion. Febs J. 276(16):4426–4436.

Bialek W, Nelson M, Tamiola K, Kallas T, Szczepaniak A. 2008. Deeply branching c_6-like cytochromes of cyanobacteria. Biochemistry 47(20):5515–5522.

Bialek W, et al. 2014. Insights into the relationship between the haem-binding pocket and the redox potential of c_6 cytochromes: four atomic resolution structures of c_6 and c_6-like proteins from *Synechococcus* sp. PCC 7002. Acta Crystallogr D Biol Crystallogr. 70(Pt 11):2823–2832.

Chan CX, et al. 2012. Analysis of *Alexandrium tamarense* (dinophyceae) genes reveals the complex evolutionary history of a microbial eukaryote. J Phycol. 48(5):1130–1142.

Criscuolo A, Gribaldo S. 2011. Large-Scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria. Mol Biol Evol. 28(11):3019–3032.
De la Rosa MA, et al. 2002. An evolutionary analysis of the reaction mechanisms of photosystem I reduction by cytochrome c_6 and plastocyanin. Bioelectrochemistry 55(1–2):41–45.
Dorrell RG, Howe CJ. 2015. Integration of plastids with their hosts: lessons learned from dinoflagellates. Proc Natl Acad Sci U S A. 112(33):10247–10254.
Dufresne A, Garczarek L, Partensky F. 2005. Accelerated evolution associated with genome reduction in a free-living prokaryote. Genome Biol. 6(2):R14.
Gross EL. 1993. Plastocyanin: structure and function. Photosynth Res. 37(2):103–116.
Gupta R, He Z, Luan S. 2002. Functional relationship of cytochrome c_6 and plastocyanin in *Arabidopsis*. Nature 417(6888):567–571.
Hackett JD, et al. 2005. Insights into a dinoflagellate genome through expressed sequence tag analysis. BMC Genomics. 6(1):80.
Howe CJ, Nimmo RH, Barbrook AC, Bendall DS. 2016. Cytochrome c_{6A} of chloroplasts. In: Cramer WA, and Kallas T, editors. Cytochrome complexes: evolution, structures, energy transduction, and signaling. Dordrecht: Springer. p. 701–712.
Howe CJ, Schlarb-Ridley BG, Wastl J, Purton S, Bendall DS. 2006. The novel cytochrome c_6 of chloroplasts: a case of evolutionary bricolage? J Exp Bot. 57(1):13–22.
Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 23(2):254–267.
Kerfeld CA, Krogmann DW. 1998. Photosynthetic cytochromes *c* in Cyanobacteria, algae, and plants. Annu Rev Plant Physiol Plant Mol Biol. 49:397–425.
Kleiner FH, Vesteg M, Steiner JM. 2021. An ancient glaucophyte c_6-like cytochrome related to higher plant cytochrome c_{6A} is imported into muroplasts. J Cell Sci. 134(9):jcs255901.
Klinger CM, et al. 2018. Plastid transcript editing across dinoflagellate lineages shows lineage-specific application but conserved trends. Genome Biol Evol. 10(4):1019–1038.
Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 33(7):1870–1874.
Marcaida MJ, et al. 2006. Structure of cytochrome c_{6A}, a novel dithiocytochrome of *Arabidopsis thaliana*, and its reactivity with plastocyanin: implications for function. J Mol Biol. 360(5):968–977.
Mareš J, et al. 2013. The primitive thylakoid-less cyanobacterium *Gloeobacter* is a common rock-dwelling organism. PLoS One. 8(6):e66323.
Marin B, Nowack ECM, Melkonian M. 2005. A plastid in the making: evidence for a second primary endosymbiosis. Protist 156(4):425–432.
Molina-Heredia FP, et al. 2003. A new function for an old cytochrome? Nature 424(6944):33–34.
Nomura CT. 2001. Electron transport of Synechococcus sp. PCC 7002 [dissertation]. Pennsylvania: Pennsylvania State University.
Novák Vanclová AMG, et al. 2020. Metabolic quirks and the colourful history of the *Euglena gracilis* secondary plastid. New Phytol. 225(4):1578–1592.
Ponce-Toledo RI, et al. 2017. An early-branching freshwater cyanobacterium at the origin of plastids. Curr Biol. 27(3):386–391.
Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ. 2007. The complete chloroplast genome of the chlorarachniophyte *Bigelowiella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. Mol Biol Evol. 24(1):54–62.
Rokas A, Carroll SB. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol Biol Evol. 22(5):1337–1344.
Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. 2014. From algae to angiosperms–inferring the phylogeny of green plants (*Viridiplantae*) from 360 plastid genomes. BMC Evol Biol. 14:23.
Schirrmeister BE, Gugger M, Donoghue PCJ. 2015. Cyanobacteria and the great oxidation event: evidence from genes and fossils. Palaeontology 58(5):769–785.
Takishita K, Ishida KI, Maruyama T. 2003. An enigmatic GAPDH gene in the symbiotic dinoflagellate genus *Symbiodinium* and its related species (the order Suessiales): possible lateral gene transfer between two eukaryotic algae, dinoflagellate and euglenophyte. Protist 154(3–4):443–454.
Turmel M, Gagnon MC, O'Kelly CJ, Otis C, Lemieux C. 2009. The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids'. Mol Biol Evol. 26(3):631–648.
Walter JM, et al. 2017. Ecogenomics and taxonomy of *Cyanobacteria phylum*. Front Microbiol. 8:2132.
Wang LG, et al. 2020. treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. Mol Biol Evol. 37(2):599–603.
Wastl J, Bendall DS, Howe CJ. 2002. Higher plants contain a modified cytochrome *c*(6). Trends Plant Sci. 7(6):244–245.
Wastl J, Purton S, Bendall DS, Howe CJ. 2004. Two forms of cytochrome c_6 in a single eukaryote. Trends Plant Sci. 9(10):474–476.
Weigel M, et al. 2003. Plastocyanin is indispensable for photosynthetic electron flow in *Arabidopsis thaliana*. J Biol Chem. 278(33):31286–31289.
Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol. 18(5):691–699.
Wisecaver JH, Brosnahan ML, Hackett JD. 2013. Horizontal gene transfer is a significant driver of gene innovation in dinoflagellates. Genome Biol Evol. 5(12):2368–2381.
Wood PM. 1978. Interchangeable copper and iron proteins in algal photosynthesis: studies on plastocyanin and cytochrome *c*-552 in *Chlamydomonas*. Eur J Biochem. 87(1):9–19.
Worrall JAR, et al. 2007. Modulation of heme redox potential in the cytochrome c_6 family. J Am Chem Soc. 129(30):9468–9475.
Yoon HS, Hackett JD, Pinto G, Bhattacharya D. 2002. The single, ancient origin of chromist plastids. J Phycol. 38:40–40.
Yoon HS, Reyes-Prieto A, Melkonian M, Bhattacharya D. 2006. Minimal plastid genome evolution in the Paulinella endosymbiont. Curr Biol. 16(17):R670–672.
Zatwarnicki P, et al. 2014. Cytochrome c_{6B} of *Synechococcus* sp. WH 8102 - crystal structure and basic properties of novel c_6-like family representative. Biochem Biophys Res Commun. 443(4):1131–1135.

Associate editor: Rebecca Zufall