

RESEARCH ARTICLE

Open Access

Chlorophyll-binding proteins revisited - a multigenic family of light-harvesting and stress proteins from a brown algal perspective

Simon M Dittami^{1,2*}, Gurvan Michel^{1,2}, Jonas Collén^{1,2}, Catherine Boyen^{1,2}, Thierry Tonon^{1,2*}

Abstract

Background: Chlorophyll-binding proteins (CBPs) constitute a large family of proteins with diverse functions in both light-harvesting and photoprotection. The evolution of CBPs has been debated, especially with respect to the origin of the L1818 subfamily, members of which function in non-photochemical quenching and have been found in chlorophyll a/c-containing algae and several organisms of the green lineage, but not in red algae so far. The recent publication of the *Ectocarpus siliculosus* genome represents an opportunity to expand on previous work carried out on the origin and function of CBPs.

Results: The *Ectocarpus* genome codes for 53 CBPs falling into all major families except the exclusively green family of chlorophyll a/b binding proteins. Most stress-induced CBPs belong to the L1818 family. However, we highlight a few stress-induced CBPs from *Phaeodactylum tricornutum* and *Chondrus crispus* that belong to different sub-families and are promising targets for future functional studies. Three-dimensional modeling of two L1818 proteins revealed features common to all L1818 proteins that are likely to interfere with their capacity to bind chlorophyll b and lutein, but may enable binding of chlorophyll c and fucoxanthin. In the light of this finding, we examined the possibility that L1818 proteins may have originated in a chlorophyll c/fucoxanthin containing organism and compared this scenario to three alternatives: an independent evolution of L1818 proteins in different lineages, an ancient origin together with the first CBPs, before the separation of the red and the green lineage, or an origin in the green lineage and a transfer to an ancestor of haptophytes and heterokonts during a cryptic endosymbiosis event.

Conclusions: Our findings reinforce the idea that the L1818 family of CBPs has a role in stress response. In addition, statistical analyses of phylogenetic trees show an independent origin in different eukaryotic lineages or a green algal origin of L1818 proteins to be highly unlikely. Instead, our data favor an origin in an ancestral chlorophyll a/c-containing organism and a subsequent lateral transfer to some green algae, although an origin of L1818 proteins in a common ancestor of red and green algae cannot be ruled out.

Background

Photosynthesis is a central process in plant physiology, which involves the collection of solar energy *via* two types of light-harvesting complexes (LHC-I and LHC-II). LHC-II is the most abundant of these complexes in thylakoid membranes, and consists of pigments (chlorophylls and carotenoids) which are bound to chlorophyll-binding proteins (CBPs). LHC proteins constitute a

large family of proteins [1,2], which includes chlorophyll a/b-binding proteins (CABs), fucoxanthin chlorophyll a/c-binding proteins (FCPs) [3,4], high light-induced proteins (HLIPs), early light-induced proteins (ELIPs), the psbS subunit of photosystem II (psbS), and stress-enhanced proteins (SEPs). CABs, as well as FCPs (together referred to as CBPs in this manuscript), have been suggested to have emerged from cyanobacterial HLIPs as a result of two duplication events and the subsequent loss of one transmembrane helix [1], but this evolutionary scenario has been recently challenged by Engelken et al. [5]. CABs and FCPs were frequently

* Correspondence: simon.dittami@bio.uio.no; tonon@sb-roscoff.fr

¹UPMC Univ Paris 06, UMR 7139 Marine Plants and Biomolecules, Station Biologique, 29680, Roscoff, France

Full list of author information is available at the end of the article

reported to be transcriptionally repressed in response to light stress [6-8].

Although CBPs are classically considered as light-harvesting proteins, increasing amounts of data point to possible additional functions within this protein family. For example, a number of genes encoding FCPs were observed to be down-regulated in a developmental mutant of the brown alga *Ectocarpus siliculosus* [9]. Moreover, several recent transcriptomic studies of stress response [10-14] highlighted FCPs that were up-regulated in response to heat-, salt-, oxidative-, or light stress in both brown algae and diatoms. Within the green lineage, similar observations were made concerning CBPs in *Chlamydomonas reinhardtii* after high light treatment [15,16]. The up-regulated CBPs were referred to as stress-induced CBPs, LI818 proteins, or LHCSR [15,17,18]. They were shown to constitute one of several distinct families of LHC proteins [14,19-21], and can be found in a range of eukaryotic lineages.

A role of these proteins in light-harvesting seems improbable considering their expression profiles. Functional analyses of both stress-induced CBPs and closely related non stress-induced CBPs carried out in a few model eukaryotes suggest that at least some of these proteins function in non-photochemical quenching (NPQ), *i.e.* the dissipation of excess light energy. In *C. reinhardtii*, the LI818 polypeptide, unlike other CBPs, was shown to be only loosely embedded in the thylakoid membranes and to be localized in stroma-exposed regions [22], and mutants of this protein were affected in their capacity to adapt to high light [18]. Paralogs of LI818 proteins in the moss *Physcomitrella patens* (named LHCSRs), even though not stress-induced on a transcriptomic level, were also shown to be active in promoting NPQ and to contribute to photoprotection under high light conditions [23]. Moreover, in the diatom *Cyclotella meneghiniana*, proteins of this family have been suggested to bind diadinoxanthin and diatoxanthin [24], and are important to quench fluorescence [25]. In another diatom, *P. tricornutum*, a LI818 protein has been shown to be required for efficient light response and to influence natural variability in photoresponse [26]. Similar suggestions have also been made for LI818 proteins from the haptophyte *Emiliana huxleyi* [21].

In addition to uncertainties about the function of these LI818 and LI818-like proteins, their evolutionary origin has not yet been finally resolved. Neilson and Durnford [20] argued that, since LI818 and LI818-like proteins have been identified in diverse groups of photosynthetic organisms, they are likely to have been amongst the first eukaryotic light-harvesting proteins. On the other hand, several reports have suggested that these proteins have evolved within the green lineage and

were transferred to the heterokonts and other chlorophyll *a/c*-containing algae during a cryptic primary endosymbiosis event, which involved the uptake of a green algal endosymbiont by an ancestral chromalveolate [18,27,28]. However, this latter theory, as well as the chromalveolate hypothesis in general, is highly debated [29-32], particularly because the only genomic data on red algae come from an atypical alga with a reduced genome and without cell wall [33].

The recent publication of the genome of the brown alga *Ectocarpus siliculosus* [34] represents an interesting opportunity to expand on previous work carried out on the origin and function of CBPs. The *Ectocarpus* genome codes for a total of 53 CBPs belonging to different families [34], several of which have already been shown to be induced in response to stress [12]. Taking into consideration the new sequence data from *E. siliculosus*, we aimed to explore both the diversity and evolution of this large family of proteins, focusing our attention on stress-induced CBPs. First, we examined the phylogenetic position of known stress-induced CBPs from different organisms. We then sought to assess structural and evolutionary differences between CABs and the LI818 family, the latter comprising most of the stress-induced CBPs. To this means, homology modeling and structural superimposition of *E. siliculosus* and *C. reinhardtii* proteins belonging to the LI818 clade with a previously crystallized spinach CAB were performed. Finally, several hypotheses on the origin of the LI818 family were tested.

Results and discussion

Genomes of photosynthetic haptophytes and heterokonts code for high numbers of CBPs and LI818 family proteins

An extensive search for CBPs in fully sequenced genomes revealed that, on average, heterokonts and haptophytes contain the highest total numbers of CBPs (Table 1, Additional file 1), although, within the green lineage, *Volvox cateri* and *Physcomitrella patens* also possess numerous copies of these proteins. LI818 family proteins are only scarcely represented within the green lineage. They were found in one to three copies in several green algae, in the genome of the moss *P. patens*, and in EST libraries produced for gymnosperms, notably *Picea glauca* (two copies, gb|DR591434.1 and gb|CO250289.1) and *Picea sitchensis* (three copies, Figure 1). Incomplete copies were also found in the angiosperms *Medicago truncatula* (one copy, gb|BG452558.1) and *Quercus robur* (one copy, gb|FR633552.1). In the genome of the red alga *Cyanidioschyzon merolae* and EST libraries for *Chondrus crispus* (4,114 ESTs), *Porphyra yezoensis* (20,069 ESTs), and *Gracilaria changii* (8,147 ESTs), no gene encoding any LI818 protein was identified; neither was any homolog found among the

Table 1 Total number of proteins containing a chlorophyll a/b-binding domain (PFAM00504/IPR022796) in a selection of eukaryotic genomes, as well as the number of CBPs belonging to the LI818 family

Organism	Total number	LI818 family proteins	Genome size (Mbp)
<i>Arabidopsis thaliana</i>	22	0	125
<i>Populus trichocarpa</i>	23	0	480
<i>Oryza sativa</i>	16	0	206
<i>Selaginella moellendorffii</i>	12	0	213
<i>Physcomitrella patens</i>	47	2	480
<i>Ostreococcus lucimarinus</i>	16	1	13
<i>Ostreococcus tauri</i>	15	1	13
<i>Chlamydomonas reinhardtii</i>	25	3	112
<i>Micromonas</i> sp. RCC299	21	2	21
<i>Coccomyxa</i> sp. C-169	24	1	49
<i>Chlorella</i> sp. NC64A	20	1	46
<i>Volvox cateri</i>	37	1	140
<i>Phaeodactylum tricornutum</i>	40	4	26
<i>Thalassiosira pseudonana</i>	40	5	32
<i>Ectocarpus siliculosus</i>	53	13	200
<i>Emiliana huxleyi</i>	87	15	168
<i>Cyanidioschyzon merolae</i>	3	0	16

Accession numbers for each of the proteins are given in Additional file 1.

available sequences for the cryptophytes *Guillardia theta* (15,173 ESTs) and *Rhodomonas* sp. [35]. In contrast, numerous proteins of this family were present in haptophytes and heterokonts, where 4 to 15 copies were found in the examined genomes. These data indicate the importance of CBPs, and in particular of LI818 proteins, in the marine environment and among haptophytes and heterokonts.

CABs and FCPs contain several subfamilies

In order to establish relationships between the *E. siliculosus* sequences and other LHC proteins identified in chlorophyll a/b-, chlorophyll a-, and chlorophyll a/c-containing organisms, phylogenetic analyses with CBPs from a wide range of taxa were performed. The topology of the tree presented in Figure 1 is similar to results previously published [14,19-21]. Two groups are clearly distinguishable. A first one comprised sequences only from the green lineage and contained most of the CABs from terrestrial plants (LHCa + LHCb). It will not be further discussed here, because it was already considered in detail in previous articles [19,36]. The remaining part of the tree constitutes the second group, which can be split into several major divisions, each comprising sequences from photosynthetic heterokonts and haptophytes, and containing a certain number of subfamilies and taxon-specific subgroups.

The first division to consider was originally highlighted by Koziol et al. [19] and was composed of members from the cryptomonads, haptophytes, and chlorarachniophytes. It was named LHCZ in absence of any indication on

function or localization of this class of proteins. In our phylogeny, LHCZ also contained one brown algal as well as several diatom sequences.

The second division corresponds to the LI818 family (also named LHCX in diatoms, [10,11,13]), and contains most of the stress-induced CBPs. It features three subgroups with moderate statistical support, which correspond to different taxonomic groups. One is formed almost exclusively by stress-induced genes from several species of diatoms, and contains the recently functionally characterized PtLHCX1 [26] (Ptr_25 in our phylogeny). Another one comprises several *E. siliculosus* sequences, many of which were induced in response to abiotic stress conditions and most of them are likely to have been subject to concerted evolution and/or to result from recent duplications. Effectively, proteins Esi_12 to Esi_22 are all highly similar and located in close proximity on supercontig 0085 in linkage group 16 of the *Ectocarpus* genome [37]. This group also contains one *Fucus* sequence. The third subgroup is the most heterogeneous and comprises sequences belonging mainly to haptophytes and to organisms of the green lineage. This subgroup also contains the LI818 proteins that have been recently functionally characterized in *Chlamydomonas* and in *Physcomitrella*, and which function in NPQ.

The remaining sequences were previously divided into two groups [19,20]: one named cryptomonad/red algal LHC clade [14,19], LhcaR [20], and LHCR in diatoms [13,38], and a second one named FCP or LHCF clade [19,20]. However, in this study as well as in the study of

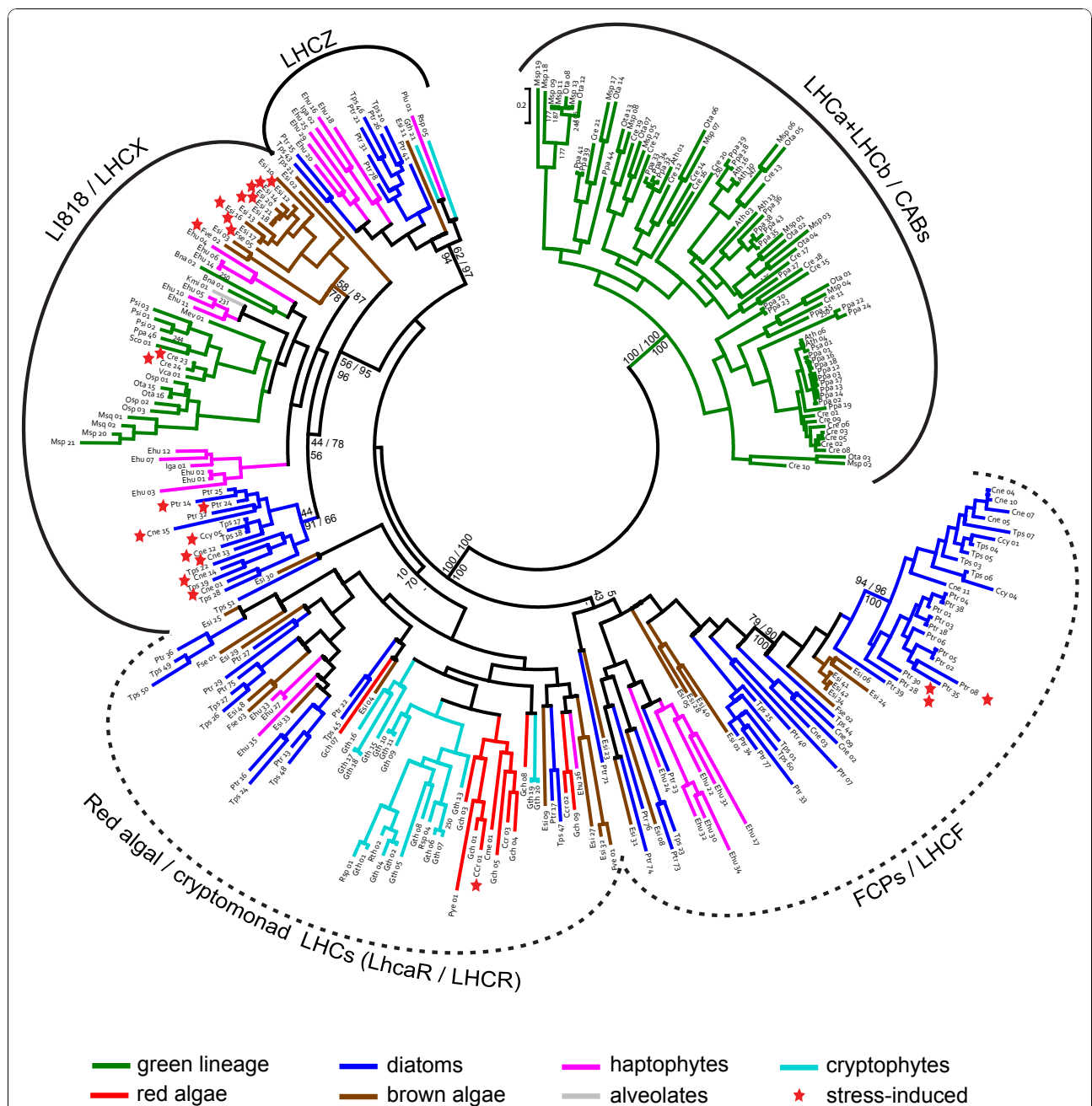


Figure 1 Phylogenetic tree of CBPs. Red stars next to the sequence name indicate genes that have been shown to be induced in response to stress. Only selected confidence values were plotted (PhyML bootstrap, PhyML Approximate Likelihood test, and MrBayes posterior probabilities respectively, with dash indicating no support by MrBayes). Dotted lines were used to indicate groups according to current naming conventions, but which are poorly resolved in our phylogeny (see text). *Abbreviations:* Ath = *Arabidopsis thaliana*, Bna = *Bigeloviella natans*, Ccr = *Chondrus crispus*, Cme = *Cyanidioschyzon merolae*, Cne = *Chaetoceros neograticus*, Cre = *Chlamydomonas reinhardtii*, Ccy = *Cyclotella meneghiniana*, Ehu = *Emiliania huxleyi*, Esi = *Ectocarpus siliculosus*, Fse = *Fucus serratus*, Fve = *Fucus vesiculosus*, Gch = *Gracilaria changii*, Gth = *Guillardia theta*, Iga = *Isochrysis galbana*, Kmi = *Karlodinium micrum*, Mev = *Mesostigma viride*, Msp = *Micromonas sp. RCC299*, Msq = *Micromonas sp. CCMP490*, Osp = *Ostreococcus sp. RCC809*, Ota = *Ostreococcus tauri*, Plu = *Pavlova lutheri*, Ppa = *Physcomitrella patens*, Psa = *Pisum sativum*, Psi = *Picea sitchensis*, Ptr = *Phaeodactylum tricornutum*, Pye = *Porphyra yezoensis*, Rsp = *Rhodomonas sp. CS24*, Sco = *Scenedesmus obliquus*, Tps = *Thalassiosira pseudonana*, Vca = *Volvox carteri*.

Neilson and Durnford [20], statistical support for the separation of these two subgroups is insufficient (10 and 5 for PhyML/bootstrap, 70 and 43 for PhyML/SH test, and no support by MrBayes). Moreover, the position of these two sub-groups with respect to the LI818 and LHCZ divisions is poorly resolved in both studies [20]. We mention these divisions in our study because they correspond to current naming conventions, but the biological relevance of such a distinction is questionable. In contrast, among the LHCF sequences, we identified a well-supported subgroup of proteins containing sequences from several species of diatoms, three of which have been shown to be induced in response to light stress [13], as well as a group of closely related *Ectocarpus* sequences.

Stress-induced chlorophyll-binding proteins

One of the questions we attempted to answer in this study was whether all stress-induced CBPs, both in *Ectocarpus* and in other photosynthetic organisms, fall into the same clade. Figure 1 clearly demonstrates that most of the stress-induced CBPs from most examined organisms (all for *Ectocarpus*) belong to the LI818 clade. This is in agreement with the supposed function of these proteins in NPQ [18,23,25]. Nevertheless, not all LI818 proteins have been shown to be transcriptionally induced in response to stress. This can be seen for example in the case of the moss *P. patens*, where neither of the two known LI818 proteins (represented by Ppa_46 in Figure 1) were up-regulated in response to osmotic-, salt-, drought, and UV-B stress in two independent studies [39,40]. Another recent study in *P. tricornutum* [26] has shown the protein LHCX1 to be induced at a protein level in response to high light, although no changes were observed on the mRNA level.

In addition to non stress-induced LI818 proteins, we also found two cases of stress-induced CBPs from other clades. The first of these cases are high light-induced CBPs from *P. tricornutum*. While four of them clearly fell into the LI818 clade, forming a subgroup with other diatom sequences, three others fell into the LHCF group and are part of a strongly supported group of diatom proteins with a sister group of *E. siliculosus* proteins. The second case is the red algal protein Ccr_01, which fell into the LhcaR clade and which was strongly induced in response to different abiotic stresses: 87-fold induced in high temperature, 21-fold in high light, and 3.6-fold in low salinity ($p < 0.05$ for all treatments) [41]. This finding is particularly interesting considering that in public red algal EST databases or in the (strongly reduced) genome of *Cyanidioschyzon*, we were not able to find either LI818 or psbS proteins. Should the absence of these proteins be confirmed by the ongoing *Chondrus* and *Porphyra* genome projects, Ccr_01 could

be an excellent candidate for an NPQ-regulating protein in red algae. In any case, the discovery of stress-induced CBPs in different families supports the hypothesis that functions other than light-harvesting may have evolved independently in different CBP clades and in different organisms.

Modeling of LI818 family proteins from *Chlamydomonas reinhardtii* and *Ectocarpus siliculosus*

After having observed that most of the stress-induced CBPs belong to the LI818 family, we attempted to assess the structural and evolutionary changes that have occurred in LI818 proteins compared to CABs by homology modeling of two sequences of this subfamily, one representative for the “brown” lineage (Esi_02 from *E. siliculosus*) and another for the “green” lineage (Cre_23 from *C. reinhardtii*). The N- and C-terminal ends of Cre_23 and Esi_02 were not included in the models because they are too divergent, and only the residues corresponding to residues Pro 19 to Asp 215 of the spinach CAB were considered. The modeled CBPs from *C. reinhardtii* and *E. siliculosus* display 58% and 60% sequence similarity, respectively, with the spinach CAB, for which the three-dimensional structure has been solved [42]. This similarity level was sufficient to generate a reliable 3D model for Cre_23 and Esi_02.

The α -helices α_1 , α_3 and α_4 and most of the β -turns are well conserved for these proteins (Figure 2), which was confirmed by the multiple sequence alignment of the LI818-like proteins (Figure 3). In these regions, LI818-like proteins feature key residues strictly conserved with the spinach CAB: Asp47 (Asp28, Esi_02 numbering), Glu65 (Glu48), His68 (His51), Glu180 (Glu158), Arg185 (Arg163) and Gln197 (Gln175). The lysine 179 is also well conserved or replaced by polar glutamine. Almost all of these residues are involved in direct binding of chlorophyll a molecules (Chla 602, 603, 610, 612, and 613). The arginine 185 stabilizes Glu65 through an ionic interaction, which coordinates the magnesium ion of Chla 602 [42].

The main difference in the secondary structure is observed in the loop connecting the helices α_1 and α_2 and in the helix α_2 itself. This region is shorter in the LI818 proteins (between 6 and 12 residues shorter) and the sequence composition is completely different (Figure 3), resulting in a significant displacement of the helix α_2 in the models of Cre_23 and Esi_02 (Figure 4). All LI818 proteins feature a conserved signature G-P-A-X-X-[H/Q] in this region. The N-terminal end of this loop is more divergent and the sequences can be grouped according to three types of sequence patterns (Figure 3). These three groups roughly correspond to several subfamilies that could be distinguished within the LI818 clades (Figure 1). The group including all LI818 proteins

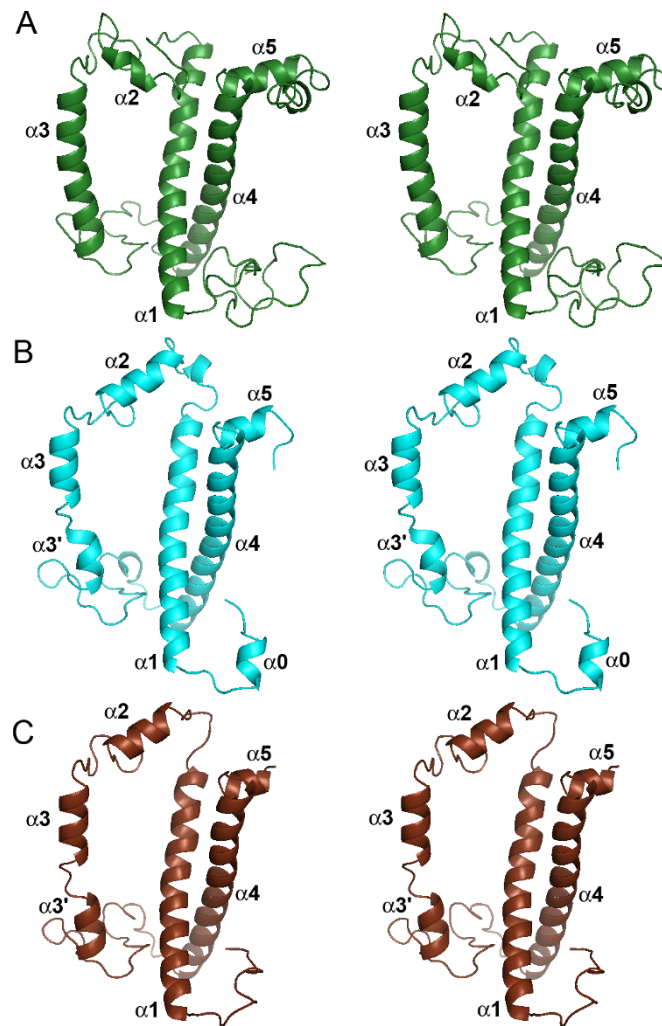
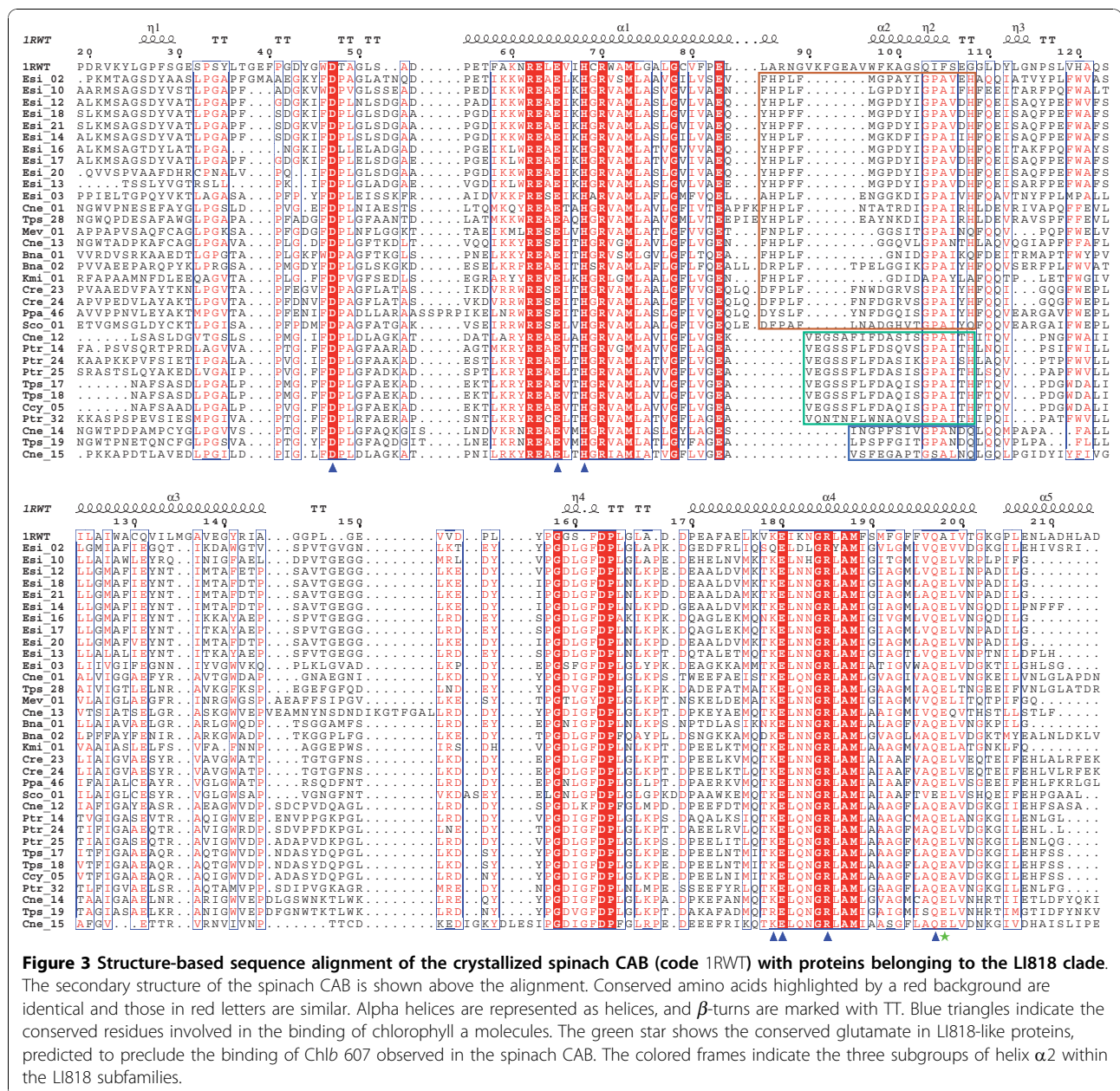


Figure 2 Structural comparison of the crystallized CAB from spinach with the 3D models of the LI818 protein Cre_23 from *Chlamydomonas reinhardtii* and of the stress-induced LHC protein Esi_02 from *Ectocarpus siliculosus*. Stereo ribbon representation of the crystal structure of the spinach CAB (PDB code: 1RWT) (A), and of the modeled proteins Cre_23 (B) and Esi_02 (C).

from *E. siliculosus* also comprises some CBPs from green algae (Cre_23, Cre_24, and Sco_01) and from moss (Ppa_46). In the spinach CAB, the helix $\alpha 2$ is involved in the binding of two chlorophyll b molecules (Chlb 607 and Chlb 606) and one lutein molecule (Lut 621) [42]. As shown in our 3D models and in the alignment, the location of helix $\alpha 2$ in LI818-like proteins is completely different and cannot provide similar binding interactions for these chlorophyll b and lutein molecules. Moreover, LI818 proteins possess a conserved glutamate residue (Glu176, Esi_02 numbering) instead of an alanine at this position in the spinach CAB (Ala198). Such a substitution would result in a steric clash with the Chlb 607, suggesting that this molecule is likely absent in LI818 proteins. These structural predictions are consistent with the fact that CBPs from heterokonts

and haptophytes do not bind chlorophyll b but bind chlorophyll c. Similarly, brown algal LHCs contain fucoxanthin, instead of lutein. In the spinach CAB, one end of the lutein is bound to Asp47, while the opposite end interacts with Trp97, which is localized in helix $\alpha 2$. In the LI818-like proteins, Asp47 is strictly conserved, but Trp97 is missing, since the equivalent region is completely altered (Figure 3). In addition, fucoxanthin is a shorter carotenoid than lutein, suggesting that one end of a fucoxanthin molecule could interact with Asp47 (Asp28, Esi_02 numbering), but that the recognition of the other end differs from the spinach protein structure. In the Cre_23 model, the helix $\alpha 2$ adopts a conformation similar to the Esi_02 helix $\alpha 2$ (Figure 4), as predicted by the conservation of the signature G-P-A-X-X-[H/Q] (Figure 3). This signature is also



conserved in other green algal proteins (Cre_24 and Sco_01) and in the protein Ppa_46 from *P. patens*. Therefore, the “green” LI818 proteins cannot recognize chlorophyll b and lutein in the same way that the spinach CAB binds these molecules [42].
 Altogether, these analyses support the conservation of the chlorophyll a binding sites in LI818 proteins. However, these proteins can probably not recognize chlorophyll b and lutein molecules as observed in the spinach CAB structure [42], due to the large displacement of helix $\alpha 2$ (Figure 4). These structural features raise questions about the nature of the pigments bound in the conserved helix $\alpha 2$ region of the LI818

proteins. In heterokont CBPs, these changes in topology could be related to the potential binding of chlorophyll c and fucoxanthin molecules, easily explaining the differences with the spinach CAB; in contrast, this hypothesis is not relevant for “green” LI818 CBPs, which are expected to bind chlorophyll b and lutein molecules. In addition, this alteration of helix $\alpha 2$ is similar in “brown” and “green” LI818 proteins, supporting that these proteins have diverged from a close, common ancestor. This suggests that green LI818 proteins may have evolved in an ancestral chlorophyll c/fucoxanthin-containing organism and were possibly acquired by green algae later.

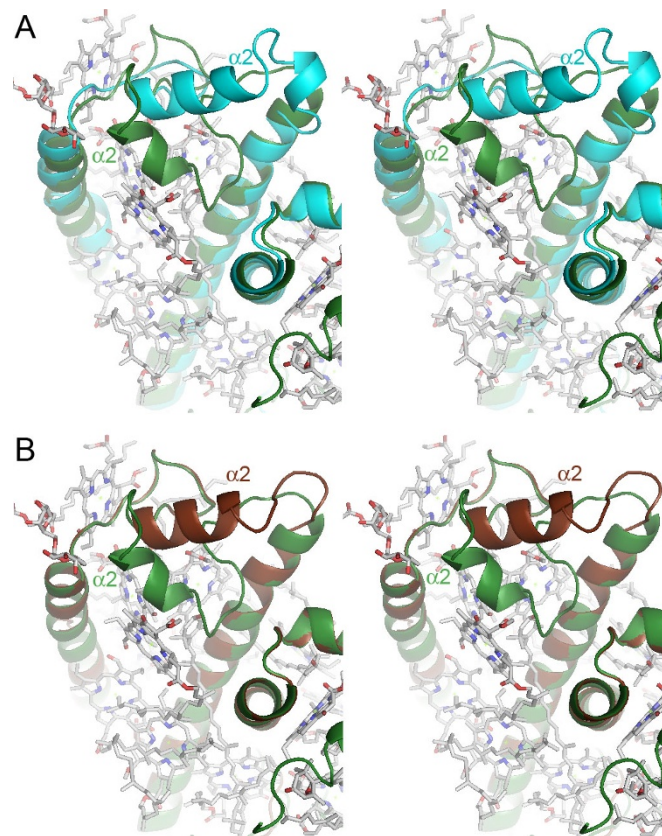


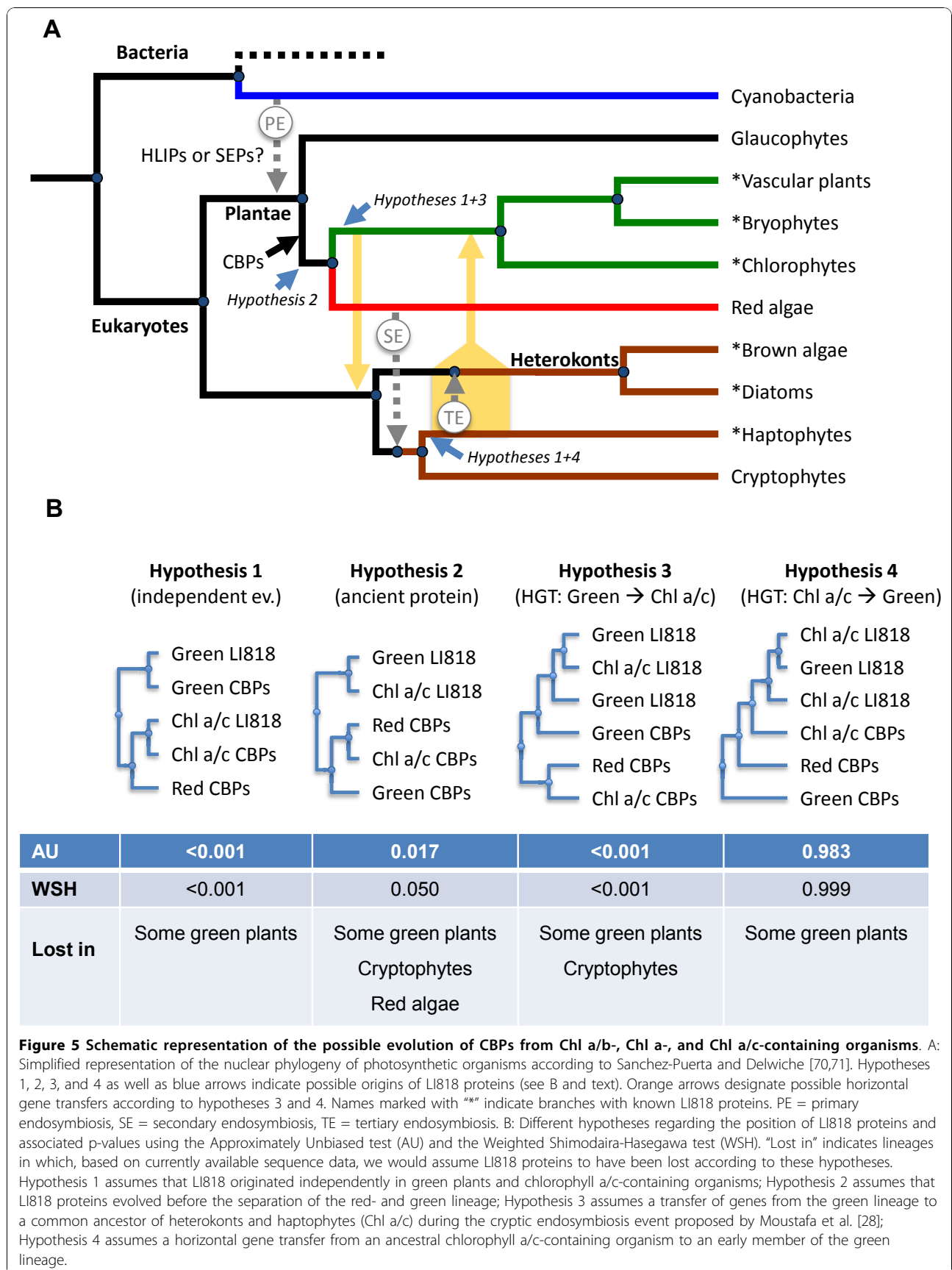
Figure 4 Comparison of the binding site of chlorophyll b and lutein molecules. (A) Stereo representation of the superimposition of the crystallized spinach CAB (green) and the modeled LI818 protein Cre_23 from *Chlamydomonas reinhardtii* (cyan). (B) Stereo representation of the superimposition of the crystallized spinach CAB (green) and the modeled LHC Esi_02 (brown). The view is a zoom on the region comprising the helix $\alpha 2$. The chlorophyll and lutein molecules bound to the spinach CBP are represented in balls and sticks, with the CPK color code.

The origin of LI818 proteins in different lineages

This latter observation prompted us to reexamine the origin of LI818 proteins in greater detail. The absence of CBP proteins in cyanobacteria and in an extensive EST library of the glaucophyte *Cyanophora paradoxa* [43] indicates that ancestral CBPs evolved after the separation of glaucophytes from red algae and green plants (Figure 5A). The presence of red-, haptophyte-, and heterokont CBPs within the same group is in agreement with the hypothesis that these proteins originate from a photosynthetic organism from the red lineage, and were transferred during secondary endosymbiosis. This event, however, cannot explain the presence of green LI818 proteins in a branch of the tree (Figure 1) otherwise comprising purely red- and chlorophyll *a/c*-containing algae. Interestingly, very similar observations were previously made for another group of proteins: Frommelt et al. [27] found that 5 of 16 protein sequences from heterokonts, haptophytes, and cryptophytes involved in carotenoid synthesis were closest to prasinophytes and not to red algal sequences. Although genes involved in carotenoid biosynthesis and

chlorophyll-binding proteins are not related, they are both involved in photosynthesis. Notably, the enzymes violaxanthin de-epoxidase and zeaxanthin epoxidase, both of which were closest to the green lineage in the study of Frommelt et al. [27], are part of the xanthophyll cycle, and thus an important actor in the regulation of NPQ, just as it has been suggested for LI818 proteins [18,23,25].

To go deeper in the analysis, we tested four different hypotheses regarding the phylogenetic position of the LI818 proteins by manually rearranging trees and performing statistical analyses. The first hypothesis assumes independent but convergent evolution of LI818 proteins in the “green” and the “brown” lineages. Our results show this scenario to be highly unlikely (Hypothesis 1, p -values < 0.00001; Figure 5B). While our model cannot account for environmental factors that might favor such a scenario, the aforementioned structural data provide additional strong arguments against this hypothesis. The second possibility was that LI818 proteins may be ancestral CBPs that have evolved in a common ancestor of the red and green lineage [20] and have subsequently



been lost in many embryophytes and probably also in red algae (Figure 5B, Hypothesis 2). In this case the position of the green LI818 proteins in our phylogenetic tree should be in reality at the root of the green branch. The fact that LI818 proteins were not found in red algae or cryptophytes could be explained either by insufficient sequence data for these organisms or by a secondary loss of these proteins. Based on the statistical analysis of a tree manually rearranged to fit this hypothesis, this option cannot be clearly ruled out ($p = 0.017$ and 0.05 with the AU and WSH tests, respectively).

A third hypothesis was previously proposed by Peers et al. [18] and consists in a horizontal transfer of green LI818 genes to a common ancestor of heterokonts and haptophytes during a cryptic primary endosymbiosis event. Such an event is believed to have taken place before secondary endosymbiosis [28], and could explain the absence of LI818 proteins in red algae but not in cryptophytes. Under the assumption that LI818 proteins were transferred during this event, we would expect LI818 proteins to branch early from the green lineage. A tree manually rearranged to fit this hypothesis received p -values < 0.001 in both AU and WSH tests (Figure 5B, Hypothesis 3). Hence, based on our analysis, this option seems highly unlikely. In addition, several studies have recently provided independent evidence against the "Chromalveolate hypothesis" [44], favoring distinct rhodobiont endosymbioses in chlorophyll *a/c*-containing algae [30,45,46]. Under this evolutionary scenario, it seems unlikely that a cryptic green algal endosymbiosis could have preceded the plastidic secondary endosymbiosis in each lineage of chlorophyll *a/c*-containing algae.

The last hypothesis, which was prompted by our examination of the structure of LI818 proteins, was that the first members of this family could have evolved in a chlorophyll *a/c*-containing organism after secondary endosymbiosis, and that a horizontal gene transfer could have taken place to an early green organism (Figure 5B, Hypothesis 4). Based on the close proximity of green and haptophyte LI818 proteins in Figure 1, such a transfer could have involved an ancestral haptophyte. Molecular clock studies date the split of the red- and green lineage to the late mesoproterozoic, over 900 million years ago (Mya) [47,48], and secondary plastid endosymbiosis is assumed to have occurred shortly thereafter [47]. The same studies date diversification of the green lineage to approximately 730 to 700 Mya, respectively, leaving a time-frame of about 200 million years during the early neoproterozoic for such an event to have taken place. This hypothesis does not assume any losses of LI818 proteins except in parts of the green lineage, and best fits our phylogenetic tree because all LI818 proteins constitute a subfamily within the group of chlorophyll

a/c-containing and red algal proteins (p -values of 0.983 to 0.999 with AU and WSH tests respectively). In accordance with this, the phylogenetic trees obtained for enzymes involved in carotenoid biosynthesis [27] could also be interpreted in the same sense, since, just as for CBPs, several proteins from green algae (e.g. phytoene desaturase, isopentenyl diphosphate:dimethylallyl diphosphate isomerase, zeaxanthin epoxidase, violaxanthin de-epoxidase) branched within a group of haptophyte and/or heterokont sequences, and not as a sister group. Gene acquisition by green algae is not unprecedented. The genomic analysis of *Micromonas* unraveled the transfer of bacterial genes involved in biosynthesis of peptidoglycan [49], and this alga has also acquired several brown algal genes responsible for mannitol synthesis [31].

Although the evidence presented in this study is not strong enough to definitely reject Hypothesis 2 (Figure 5B), our findings provide a strong indication against the putative transfer of LI818 proteins during a cryptic green algal endosymbiosis event preceding red algal secondary endosymbiosis (Hypothesis 3), and favor a scenario with a transfer in the other direction (Hypothesis 4).

Conclusion

This study shows that *E. siliculosus* contains a wide variety of CBPs from different divisions and with potentially different functions. In all of the examined species, most of the CBPs known to be stress-induced belong to the LI818 family, which is particularly well represented among haptophytes and heterokonts, but also present in several organisms of the green lineage. According to our analyses, the possibility that LI818 proteins may be ancestral CBPs, which have evolved before the separation of the red and the green lineage, remains viable. However, structural alignments and three-dimensional modeling illustrated several elements common to LI818 proteins, which are likely to interfere with their capacity to bind certain chlorophyll *b* molecules, but may enable the binding of chlorophyll *c* and/or fucoxanthin. This finding, together with the statistical analysis of our phylogenetic tree, point to a new possibility not dealt with in previous studies, *i.e.* that LI818 proteins may have originated in an ancient chlorophyll *a/c*-containing organism and could have been later transferred to the green lineage. Understanding the evolutionary history of CBPs may also increase our understanding of the evolution of the different eukaryotic lineages as a whole. If our hypothesis is correct, LI818 proteins will not be found in red algae. We thus anticipate that ongoing genome projects such as those of the red algae *Chondrus* and *Porphyra* will confirm this prediction.

Methods

Sequence retrieval

Four approaches were taken to select sequences for this study. First, CBPs were searched for in the sequenced genomes of photosynthetic eukaryotes representing the major photosynthetic lineages (Table 1). These searches were carried out using “blastp” and reference sequences representing the four major clades in our phylogeny. All hits (e-value < 10) were submitted to InterProScan [50] using default parameters, and considered CBPs if a “Chlorophyll A-B binding protein” domain corresponding to the PFAM00504/IPR022796 motif was detected. All sequence identifiers, database references, and InterProScan domains are given in Additional file 1. Many of these sequences were included in our phylogenetic analysis; however, in cases where the results obtained for related species were highly similar, only a representative was selected.

In a second step we attempted to fill gaps in our phylogenetic tree, where no or only limited genome sequences were available, by using data from EST libraries. For these searches “tblastn” was used instead of “blastp”, and in case of duplicate sequences, the longest EST was chosen. The identified CBP sequences were translated using NCBI OrfFinder (<http://www.ncbi.nlm.nih.gov/projects/gorf/>). In the case of red algae, only the reduced genome of *C. merolae* was available for public use, and we added EST sequences from *C. crispus* (4,114 ESTs, 3 coding for CBPs; [51]), *P. yezoensis* (20,069 ESTs, 4 coding for CBPs; [52]), and *G. changii* (8,147 ESTs, 7 coding for CBPs; Teo et al., unpublished), all of which were obtained from the NCBI EST database (dbEST). Furthermore, as no genome sequences are available for cryptophytes, the same strategy was applied for an EST library of *Guillardia theta* (15,173 ESTs, 21 coding for CBPs; [53]), and we also took advantage of the availability of five CBP sequences obtained in a previous study by Broughton et al. on *Rhodomonas sp.* CS24 [35].

As we were particularly interested in stress-induced CBPs and the possible role of these CBPs in NPQ, we considered additional sequences relevant for these topics in a third step: seventeen ESTs coding for CBPs from the Antarctic diatom *Chaetoceros neogracile* (total number of ESTs in the library 1,744; [10]), and six previously identified CBPs from *Fucus serratus* and *F. vesiculosus* were selected because transcriptomic data were available for these sequences [10,11,14]; also, three sequences from the diatom *Cyclotella meneghiniana* were included, because these proteins had been shown to be related to NPQ [25].

Finally, several individual sequences were added to our phylogenetic tree in order to facilitate comparisons with previous phylogenetic studies [19,20,36,54], and to include a wider selection of LI818 proteins from the green lineage. For the latter purpose, blast searches were performed in the NCBI nr and EST databases, and the retrieved sequences were treated as described above. We added two sequences from the chlorarachniophyte (Rhizaria) *Bigelowiella natans*, which is thought to have acquired its plastids through an independent secondary endosymbiosis event with a green alga [55]; one sequence from the dinophyte *Karlodinium micrum*; two sequences from the unicellular green algae *Mesostigma viride* (Charophyta) and *Scenedesmus obliquus* (Chlorophyta); one sequence from the haptophyte *Pavlova lutheri*; two sequences from a second strain of *Micromonas sp.* (CCMP490); three sequences from *Ostreococcus sp.* RCC809; one sequence each from *O. tauri* and *V. carteri* (these had also been identified in our analysis of the corresponding genomes); and three sequences from the gymnosperm *Picea sitchensis*. The sequence corresponding to the first available CAB structure from *Pisum sativum* [56] was included to enable homology-based three-dimensional modeling, but was later replaced by the spinach protein mentioned below. A complete list of sequences and accession numbers is available in Additional file 2, and further LI818 proteins are reported in the first paragraph of the results and discussion section, even though they were not included in the phylogenetic analyses due to the high degree of sequence similarity with other represented sequences.

Phylogenetic analyses

A phylogenetic tree was constructed for all sequences except those of poplar and maize, which were highly similar to those of *A. thaliana*, and those of *O. lucimarinus*, as they were highly similar to *O. tauri*. To this aim, sequences were aligned using MAFFT [57] and the “E-INS-i” strategy with default settings. Automatic alignments were then manually refined using Bioedit 7.0.9.0 [58], taking into consideration only the conserved regions of the proteins for further analyses. A total of 131 residues were manually selected (see Additional file 3). In some cases, this resulted in two sequences with 100% sequence identity, and one of the identical sequences was excluded for further analyses. In parallel, an automatic selection of conserved residues was carried out using the Gblocks algorithm [59]. Even with the least stringent settings, this resulted in the selection of only 53 residues. Phylogenetic analyses based on these 53 residues yielded similar tree topologies as the manual

selection (data not shown). CBPs and FCPs were considered stress-induced if they were identified as such in the original publications, or in the case of *E. siliculosus*, if the mean ratio stress/control for hyposaline-, hypersaline-, and oxidative stress reported in our previous study was greater than 2, and the p-value (ANOVA) associated to this change < 0.05 [60].

Maximum likelihood trees were generated using PhyML (bootstrapping with 250 iterations as well as Approximate Likelihood-Ratio testing [61]). We chose the "Blosum62" substitution model allowing gamma-distributed rate heterogeneity over sites (4 categories) as well as invariable sites, as this was shown to best represent our data using ProtTest [62] and the AIC criterion. The results were complemented by bayesian inference analysis using MrBayes [63]. Except for the use of "Blosum62" as prior for the substitution model and gamma-distributed rate heterogeneity in the likelihood model (see above), default parameters were applied. The analysis was run for 8,000,000 generations (samples were taken every 100 generations). At this point, the average standard deviation of split frequencies was stable at approximately 0.03 and decreased no further. The first 25% of samples were discarded as burn in. The PhyML tree was used as a basis to manually introduce targeted modifications in the tree and to run the tests evaluating different evolutionary scenarios on the origin of LI818 proteins. Different trees were tested using treepuzzle [64] to calculate site-wise log-likelihood (same substitution model as above), and CONSEL [65] to perform the Approximately Unbiased (AU, [66]) and the Weighted Shimodaira-Hasegawa tests (WSH).

Sequence analysis and three-dimensional modeling

A multiple sequence alignment of the LI818-like sequences was generated as described above, and manually refined using BioEdit, on the basis of the crystal structure of the CAB from spinach (PDB code 1RWT; [42]). The chain A of this structure was also used to generate a three dimensional (3D) model of the proteins Cre_23 from *C. reinhardtii* and Esi_02 from *E. siliculosus* using Modeller [67] with default parameters. The multiple sequence alignment and the 3D models were displayed using ESPript [68] and PyMOL [69], respectively.

Additional material

Additional file 1: Databases and sequence identifiers of CBPs found in the genomes of the organisms listed in Table 1. Each organism is represented by a separate sheet. "SignalP" indicates the presence of a signal peptide.

Additional file 2: Accession numbers of all sequences considered for the phylogenetic tree in Figure 1 (including those removed due to high sequence identity). For stress-induced proteins, the PubMed Id (PMID) of the corresponding publication is given.

Additional file 3: Alignment used for the construction of the phylogenetic tree in Figure 1.

List of abbreviations

CAB: chlorophyll a/b-binding protein; CBP: chlorophyll-binding protein; ELIP: early light-induced protein; EST: expressed sequence tag; FCP: fucoxanthin chlorophyll a/c-binding protein; HLIP: high light-induced protein; LHC: light-harvesting complex; NPQ: non-photochemical quenching; SEP: stress-enhanced proteins;

Acknowledgements

We would like to thank Odile Richard, Catherine Leblanc, Gildas Le Corguillé, Shuhei Ota, Martin Lohr, and the reviewers of this paper for helpful discussions and/or comments. SD received funding from the European community's Sixth Framework Programme (contract n° MESTCT 2005-020737).

Author details

¹UPMC Univ Paris 06, UMR 7139 Marine Plants and Biomolecules, Station Biologique, 29680, Roscoff, France. ²CNRS, UMR 7139 Marine Plants and Biomolecules, Station Biologique, 29680, Roscoff, France.

Authors' contributions

SD, GM and TT conceived the study, together with JC and CB, SD, GM, and TT performed and interpreted the analyses. SD wrote the manuscript together with GM, CB, JC, and TT. All authors approved the final manuscript.

Received: 28 August 2010 Accepted: 26 November 2010

Published: 26 November 2010

References

1. Green BR, Kühlbrandt W: Sequence conservation of light-harvesting and stress-response proteins in relation to the three-dimensional molecular structure of LHCII. *Photosynth Res* 1995, **44**:139-148.
2. Green BR, Pichersky E, Kloppstech K: Chlorophyll a/b-binding proteins: an extended family. *Trends Biochem Sci* 1991, **16**:181-186.
3. Grossman A, Manodori A, Snyder D: Light-harvesting proteins of diatoms: their relationship to the chlorophyll a/b binding proteins of higher plants and their mode of transport into plastids. *Mol Gen Genet* 1990, **224**:91-100.
4. Caron L, Douady D, de Martino A, Quinet M: Light harvesting in brown algae. *Cah Biol Mar* 2001, **42**:109-124.
5. Engelken J, Brinkmann H, Adamska I: Taxonomic distribution and origins of the extended LHC (light-harvesting complex) antenna protein superfamily. *BMC Evol Biol* 2010, **10**:233.
6. Tonkyn JC, Deng X-W, Gruijssem W: Regulation of plastid gene expression during photooxidative stress. *Plant Physiol* 1992, **99**:1406-1415.
7. Maxwell DP, Laudenbach DE, Huner N: Redox regulation of Light-Harvesting Complex II and CAB mRNA abundance in *Dunaliella salina*. *Plant Physiol* 1995, **109**:787-795.
8. Teramoto H, Nakamori A, Minagawa J, T-aki Ono: Light-intensity-dependent expression of Lhc gene family encoding light-harvesting chlorophyll-a/b proteins of photosystem II in *Chlamydomonas reinhardtii*. *Plant Physiol* 2002, **130**:325-333.
9. Peters AF, Scornet D, Ratin M, Charrier B, Monnier A, Merrien Y, Corre E, Coelho SM, Cock JM: Life-cycle-generation-specific developmental processes are modified in the immediate upright mutant of the brown alga *Ectocarpus siliculosus*. *Development* 2008, **135**:1503-1512.
10. Hwang YS, Jung G, Jin E: Transcriptome analysis of acclimatory responses to thermal stress in Antarctic algae. *Biochem Biophys Res Comm* 2008, **367**:635-641.
11. Park S, Jung G, Y-sic Hwang, Jin E: Dynamic response of the transcriptome of a psychrophilic diatom, *Chaetoceros neogracile*, to high irradiance. *Planta* 2010, **231**:349-360.
12. Dittami SM, Scornet D, Petit J-L, Ségurens B, Da Silva C, Corre E, Dondrup M, Glatting K-H, König R, Sterck L, et al: Global expression analysis of the brown alga *Ectocarpus siliculosus* (Phaeophyceae) reveals

- large-scale reprogramming of the transcriptome in response to abiotic stress. *Genome Biol* 2009, **10**:R66.
13. Nymark M, Valle KC, Brembu T, Hancke K, Winge P, Andresen K, Johnsen G, Bones AM: An integrated analysis of molecular acclimation to high light in the marine diatom *Phaeodactylum tricornutum*. *PLoS One* 2009, **4**:e7743.
 14. Pearson GA, Hoarau G, Lago-Leston A, Coyer James A, Kube M, Reinhardt R, Henckel K, Serrão ETA, Corre E, Olsen JL: An expressed sequence tag analysis of the intertidal brown seaweeds *Fucus serratus* (L.) and *F. vesiculosus* (L.) (Heterokontophyta, Phaeophyceae) in response to abiotic stressors. *Mar Biotechnol* 2010, **12**:195-213.
 15. Savard F, Richard C, Guertin M: The *Chlamydomonas reinhardtii* LI818 gene represents a distant relative of the cabl/II genes that is regulated during the cell cycle and in response to illumination. *Plant Mol Biol* 1996, **32**:461-473.
 16. Miura K, Yamano T, Yoshioka S, Kohinata T, Inoue Y, Taniguchi F, Asamizu E, Nakamura Y, Tabata SY, Katsuyuki T, et al: Expression profiling-based identification of CO₂-responsive genes regulated by CCM1 controlling a carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. *Plant Physiol* 2004, **135**:1595-1607.
 17. Elrad D, Grossman AR: A genome's-eye view of the light-harvesting polypeptides of *Chlamydomonas reinhardtii*. *Curr Genet* 2004, **45**:61-75.
 18. Peers G, Truong TB, Ostendorf E, Busch A, Elrad D, Grossman AR, Hippler M, Niyogi KK: An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. *Nature* 2009, **462**:518-521.
 19. Koziol AG, Borza T, Ishida K-I, Keeling P, Lee RW, Durnford DG: Tracing the evolution of the light-harvesting antennae in chlorophyll a/b-containing organisms. *Plant Physiol* 2007, **143**:1802-1816.
 20. Neilson JAD, Durnford DG: Structural and functional diversification of the light-harvesting complexes in photosynthetic eukaryotes. *Photosynth Res* 2010, **106**:57-71.
 21. Lefebvre SC, Harris G, Webster R, Leonardos N, Geider RJ, Raines CA, Read BA, Garrido JL: Characterization and expression analysis of the LHCF gene family in *Emiliania huxleyi* (Haptophyta) reveals differential responses to light and CO₂. *J Phycol* 2010, **46**:123-134.
 22. Richard C, Ouellet H, Guertin M: Characterization of the LI818 polypeptide from the green unicellular alga *Chlamydomonas reinhardtii*. *Plant Mol Biol* 2000, **42**:303-316.
 23. Alboresi A, Gerotto C, Giacometti GM, Bassi R, Morosinotto T: *Physcomitrella patens* mutants affected on heat dissipation clarify the evolution of photoprotection mechanisms upon land colonization. *PNAS* 2010, **107**:11128-11133.
 24. Beer A, Gundermann K, Beckmann J, Büchel C: Subunit composition and pigmentation of fucoxanthin-chlorophyll proteins in diatoms: evidence for a subunit involved in diadinoxanthin and diatoxanthin binding. *Biochem* 2006, **45**:13046-13053.
 25. Gundermann K, Büchel C: The fluorescence yield of the trimeric fucoxanthin-chlorophyll-protein FCPa in the diatom *Cyclotella meneghiniana* is dependent on the amount of bound diatoxanthin. *Photosynth Res* 2008, **95**:229-235.
 26. Bailleul B, Rogato A, de Martino A, Coesel S, Cardol P, Bowler C, Falciatore A, Finazzi G: An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light. *PNAS* 2010, **1-6**.
 27. Frommolt R, Werner S, Paulsen H, Goss R, Wilhelm C, Zauner S, Maier UG, Grossman AR, Bhattacharya D, Lohr M: Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. *Mol Biol Evol* 2008, **25**:2653-2667.
 28. Moustafa A, Beszteri B, Maier UG, et al: Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* 2009, **324**:1724-1726.
 29. Burki F, Shalchian-Tabrizi K, Minge M, Skjæveland A, Nikolaev SI, Jakobsen KS, Pawlowski J: Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One* 2007, **2**:e790.
 30. Michel G, Tonon T, Scornet D, Cock JM, Kloareg B: The cell wall polysaccharide metabolism of the brown alga *Ectocarpus siliculosus*. Insights into the evolution of extracellular matrix polysaccharides in eukaryotes. *New Phytol* 2010, **188**:82-97.
 31. Michel G, Tonon T, Scornet D, Cock JM, Kloareg B: Central and storage carbon metabolism of the brown alga *Ectocarpus siliculosus*: insights into the origin and evolution of storage carbohydrates in eukaryotes. *New Phytol* 2010, **188**:67-81.
 32. Dagan T, Martin W: Microbiology. Seeing green and red in diatom genomes. *Science* 2009, **324**:1651-1652.
 33. Matsuzaki M, Misumi O, Shin-I T, Maruyama S, Takahara M, Miyagishima S-Y, Mori T, Nishida K, Yagisawa F, Nishida K, et al: Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 2004, **428**:653-657.
 34. Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury J-M, Badger JH, et al: The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* 2010, **465**:617-621.
 35. Broughton MJ, Howe CJ, Hiller RG: Distinctive organization of genes for light-harvesting proteins in the cryptophyte alga *Rhodomonas*. *Gene* 2006, **369**:72-79.
 36. Alboresi A, Caffarri S, Nogue F, Bassi R, Morosinotto T: In silico and biochemical analysis of *Physcomitrella patens* photosynthetic antenna: identification of subunits which evolved upon land adaptation. *PLoS One* 2008, **3**:e2033.
 37. Heesch S, Cho GY, Peters AF, Le Corguillé G, Falentin C, Boutet G, Coëdel S, Jubin C, Samson G, Corre E, et al: A sequence-tagged genetic map for the brown alga *Ectocarpus siliculosus* provides large-scale assembly of the genome sequence. *New Phytol* 2010, **188**:42-51.
 38. Zhu S-H, Green BR: Light-Harvesting and photoprotection in diatoms: identification and expression of L818-like proteins. In *Photosynthesis. Energy from the Sun*. Edited by: Allen JF, Gantt E, Golbeck JH, Osmond B. Dordrecht: Springer; 2008:261-264.
 39. Cuming AC, Cho SH, Kamisugi Y, Graham H, Quatrano RS: Microarray analysis of transcriptional responses to abscisic acid and osmotic, salt, and drought stress in the moss, *Physcomitrella patens*. *New Phytol* 2007, **176**:275-287.
 40. Wolf L, Rizzini L, Stracke R, Ulm R, Rensing SA: The molecular and physiological response of *Physcomitrella patens* to UV-B radiation. *Plant Physiol* 2010, **153**:1123-1134.
 41. Collén J, Guisles-Marsollier I, Leger JJ, Boyen C: Response of the transcriptome of the intertidal red seaweed *Chondrus crispus* to controlled and natural stresses. *New Phytol* 2007, **176**:45-55.
 42. Liu Z, Yan H, Wang K, Kuang T, Zhang J, Gui L, An X, Chang W, et al: Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature* 2004, **428**:287-292.
 43. Reyes-Prieto A, Hackett JD, Soares MB, Bonaldo MF, Bhattacharya D: Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. *Curr Biol* 2006, **16**:2320-2325.
 44. Cavalier-Smith T: Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Euk Microbiol* 1999, **46**:347-366.
 45. Stiller JW, Huang J, Ding Q, Tian J, Goodwillie C: Are algal genes in nonphotosynthetic protists evidence of historical plastid endosymbioses? *BMC Genomics* 2009, **10**:484.
 46. Baurain D, Brinkmann H, Petersen J, Rodríguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger G, Lang BF, Philippe H: Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol Biol Evol* 2010, **27**:1698-1709.
 47. Douzery EJP, Snell EA, Baptiste E, Delsuc F, Philippe H: The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? *PNAS* 2004, **101**:6-1538.
 48. Berney C, Pawlowski J: A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc Biol Sci* 2006, **273**:1867-1872.
 49. Worden AZ, Lee J-H, Mock T, Rouzé P, Simmons MP, Aerts AL, Allen AE, Cuvelier ML, Derelle E, Everett MV, et al: Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* 2009, **324**:268-272.
 50. Zdobnov EM, Apweiler R: InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 2001, **17**:847-848.
 51. Collén J, Roeder V, Rousvoal S, Collin O, Kloareg B, Boyen C: An expressed sequence tag analysis of thallus and regenerating protoplasts of *Chondrus crispus* (Gigartinales, Rhodophyceae). *J Phycol* 2006, **42**:104-112.
 52. Nikaido I, Asamizu E, Nakajima M, Nakamura Y, Saga N, Tabata S: Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, *Porphyra yezoensis*. *DNA Res* 2000, **7**:223-227.

53. Gould SB, Fan E, Hempel F, Maier U-G, Klösgen RB: **Translocation of a phycoerythrin alpha subunit across five biological membranes.** *J Biol Chem* 2007, **282**:30295-30302.
54. La Roche J, Henry D, Wyman K, Sukenik A, Falkowski P: **Cloning and nucleotide sequence of a cDNA encoding a major fucoxanthin-, chlorophyll a/c-containing protein from the chrysophyte *Isochrysis galbana*: implications for evolution of the CAB gene family.** *Plant Mol Biol* 1994, **25**:355-368.
55. Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ: **The complete chloroplast genome of the chlorarachniophyte *Bigeloviella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts.** *Mol Biol Evol* 2007, **24**:54-62.
56. Kühlbrandt W, Wang DN, Fujiyoshi Y: **Atomic model of plant light-harvesting complex by electron crystallography.** *Nature* 1994, **367**:614-621.
57. La Roche J, van der Staay GW, Partensky F, Ducret A, Aebersold R, Li R, Golden SS, Hiller RG, Wrench PM, Larkum AW, Green BR: **Independent evolution of the prochlorophyte and green plant chlorophyll a/b light-harvesting proteins.** *PNAS* 1996, **93**:15244-15248.
58. Katoh K, Misawa K, Kuma K-ichi, Miyata T: **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.** *Nucleic Acids Res* 2002, **30**:3059-3066.
59. Hall T: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symp Ser* 1999, **41**:95-98.
60. Talavera G, Castresana J: **Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments.** *Syst Biol* 2007, **56**:564-577.
61. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52**:696-704.
62. Abascal F, Zardoya R, Posada D: **ProtTest: selection of best-fit models of protein evolution.** *Bioinformatics* 2005, **21**:2104-2105.
63. Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP: **Bayesian inference of phylogeny and its impact on evolutionary biology.** *Science* 2001, **294**:2310-2314.
64. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18**:502-504.
65. Shimodaira H, Hasegawa M: **CONSEL: for assessing the confidence of phylogenetic tree selection.** *Bioinformatics* 2001, **17**:1246-1247.
66. Shimodaira H: **An approximately unbiased test of phylogenetic tree selection.** *Syst Biol* 2002, **51**:492-508.
67. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M-Y, Pieper U, Sali A: **Comparative protein structure modeling using Modeller.** *Curr Protoc Bioinformatics* 2006, **Chapter 5**(Unit 5.6).
68. Gouet P, Robert X, Courcelle E: **ESPrpt/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins.** *Nucleic Acids Res* 2003, **31**:3320-3.
69. DeLano WL: **The case for open-source software in drug discovery.** *Drug Discov Today* 2005, **10**:213-7.
70. Sanchez-Puerta MV, Delwiche CF: **A hypothesis for plastid evolution in chromalveolates.** *J Phycol* 2008, **44**:1097-1107.
71. Archibald JM: **The puzzle of plastid evolution.** *Curr Biol* 2009, **19**:R81-88.

doi:10.1186/1471-2148-10-365

Cite this article as: Dittami et al.: Chlorophyll-binding proteins revisited - a multigenic family of light-harvesting and stress proteins from a brown algal perspective. *BMC Evolutionary Biology* 2010 **10**:365.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

