

RESEARCH ARTICLE

Open Access

# Evolution of light-harvesting complex proteins from Chl *c*-containing algae

Gabriel E Hoffman<sup>1,3</sup>, M Virginia Sanchez Puerta<sup>1,4</sup> and Charles F Delwiche<sup>1,2\*</sup>

## Abstract

**Background:** Light harvesting complex (LHC) proteins function in photosynthesis by binding chlorophyll (Chl) and carotenoid molecules that absorb light and transfer the energy to the reaction center Chl of the photosystem. Most research has focused on LHCs of plants and chlorophytes that bind Chl *a* and *b* and extensive work on these proteins has uncovered a diversity of biochemical functions, expression patterns and amino acid sequences. We focus here on a less-studied family of LHCs that typically bind Chl *a* and *c*, and that are widely distributed in Chl *c*-containing and other algae. Previous phylogenetic analyses of these proteins suggested that individual algal lineages possess proteins from one or two subfamilies, and that most subfamilies are characteristic of a particular algal lineage, but genome-scale datasets had revealed that some species have multiple different forms of the gene. Such observations also suggested that there might have been an important influence of endosymbiosis in the evolution of LHCs.

**Results:** We reconstruct a phylogeny of LHCs from Chl *c*-containing algae and related lineages using data from recent sequencing projects to give ~10-fold larger taxon sampling than previous studies. The phylogeny indicates that individual taxa possess proteins from multiple LHC subfamilies and that several LHC subfamilies are found in distantly related algal lineages. This phylogenetic pattern implies functional differentiation of the gene families, a hypothesis that is consistent with data on gene expression, carotenoid binding and physical associations with other LHCs. In all probability LHCs have undergone a complex history of evolution of function, gene transfer, and lineage-specific diversification.

**Conclusion:** The analysis provides a strikingly different picture of LHC diversity than previous analyses of LHC evolution. Individual algal lineages possess proteins from multiple LHC subfamilies. Evolutionary relationships showed support for the hypothesized origin of Chl *c* plastids. This work also allows recent experimental findings about molecular function to be understood in a broader phylogenetic context.

## Background

Light harvesting complex (LHC) proteins are fundamental to oxygenic photosynthesis, and members of the LHC family are present in most photosynthetic eukaryotes, although variation in nomenclature sometimes obscures their widespread occurrence (Table 1). These transmembrane proteins bind chlorophyll (Chl) and carotenoid pigments which function to absorb light and transfer energy to the reaction center Chl of photosystems (PS) in the thylakoid membrane [1]. The biochemistry, physical interactions and molecular phylogeny of

multiple types of LHCs have been characterized in plants and chlorophytes (green algae) [2-4], but less is known about homologs in Chl *c*-containing algae. Consequently, to improve our understanding of the evolution of the LHC gene family in Chl *c*-containing lineages, we undertook a phylogenetic analysis of expressed sequence tags (ESTs) and genomic data. Chl *c*-containing algae, along with their non-photosynthetic relatives are also known as “chromalveolates” [5] under the hypothesis that these lineages descend from a photosynthetic common ancestor. However, the monophyly of the “chromalveolates” has been questioned and remains controversial [6,7]. This analysis is intended to relate the molecular phylogeny of the Chl *c*-containing algal LHCs to their function, and provides insight into

\* Correspondence: delwiche@umd.edu

<sup>1</sup>Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD, 20742, USA

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

**Table 1 Major clades of the LHC gene superfamily.**

Systematic name	Other names/subclades	Included in this analysis	Lineages
Lhca	LHC I	-	plants
Lhcb	LHC II, CP24,CP26, CP29	-	plants
Lhcc	Cac	+	cryptophytes
Lhcd	Lhcp	+	peridinin-containing dinoflagellates
Lhcf	FCP, cac	+	haptophytes, heterokonts
Lhcr	LhcaR	+	rhodophytes
-	Lhcz	+	cryptophytes, haptophytes, heterokonts
-	L1818, LHCSR	+	chlorarachniophytes, chlorophytes, fucoxanthin-containing dinoflagellates, haptophytes, heterokonts

The systematic nomenclature was established by Jansson et al. [71]. The clades termed Lhcz and L1818 do not have systematic names and were identified by Koziol et al. [2] and Gagne and Guertin [72], respectively. None of the other clades in the current analysis have been previously described. The genes in Lhca, Lhcb and related clades recently identified by Koziol et al. [2] form a monophyletic outgroup to the genes analyzed in the current analysis, but the root of the phylogeny cannot be confidently inferred.

gene duplication, expression, differential biochemical activity and evolution of the LHC family, and although we use the term here, the study is not dependent upon the validity of the chromalveolate hypothesis.

All eukaryotic lineages that are capable of oxygenic photosynthesis obtained this ability by (directly or indirectly) engulfing photosynthetic cyanobacteria and incorporating them as plastids [8], which are most familiar as the chloroplasts of green algae and plants. Chlorophytes, rhodophytes and glaucocystophytes appear to be the result of a single primary endosymbiotic event, whereby a cyanobacterium was incorporated as a plastid [9]. It should be noted that the 'event' in question could have involved many individual cells in a population, and might have resulted from processes that took place over a very long time, but all primary plastids seem to share a common origin, and hence it is appropriate to think of it as a single event [10,11]. Chl *c*-containing plastids are the result of at least one secondary endosymbiotic incorporation of a rhodophyte, and tertiary events involving a plastid of rhodophyte origin [5,7,8]. Chl *c*-containing algae include four major lineages: heterokonts, haptophytes, dinoflagellates and cryptophytes that are united (at least) by their use of Chl *c* as a photopigment. These lineages have often been treated separately, but recent evidence suggests that heterokonts and dinoflagellates (together with other alveolates) may form a single clade (possibly also including Rhizaria), and that cryptophytes and haptophytes are sister lineages [12,13]. The precise origin of the plastids (and number of endosymbiotic events involved) in these lineages is still unclear, but their ultimate rhodophyte ancestry is uncontroversial [14-16]. In addition, some dinoflagellates are thought to have replaced their plastids with new plastids from other lineages, including an environmentally important group (fucoxanthin-containing

dinoflagellates) that have incorporated a haptophyte as a plastid in a tertiary endosymbiotic event [17]. Most recently (and controversially), genome analyses led Moustafa et al. [18] to hypothesize that the common ancestor of heterokonts and haptophytes acquired a transient green algal plastid, subsequently replaced by a red algal plastid and now only represented by residual genes of green algal origin.

Members of the LHC gene family are nuclear-encoded, and in organisms with secondary or tertiary plastids the proteins are targeted from the cytosol to the ER and then to the plastid using a bipartite signal sequence at the amino terminus. Functional LHCs are located in the thylakoid membrane of the plastid and possess three  $\alpha$ -helical transmembrane regions (TMR) that are evolutionarily conserved [19]. The 3-dimensional structure and carotenoid binding sites of Chl *a/b* LHCs have been characterized [20,21], but the structural details and specific binding properties of the Chl *a/c* LHCs addressed here are likely different given the degree of sequence divergence and duration of independent evolution of these two gene subfamilies. Moreover, even closely related members of the diverse Chl *a/c* LHC family exhibit differential carotenoid and Chl binding, and differential associations in trimers or higher oligomers [22,23]. Thus, functional inferences made for plant LHCs can only tenuously be extended to the Chl *a/c* LHCs. However, a first step towards understanding these LHCs is to unravel evolutionary relationships within the gene family.

Major groups of algae can be characterized by the type of Chl associated with their LHCs. Chlorophyte and charophyte LHCs bind Chls *a* and *b*, rhodophyte LHCs bind Chl *a*, and Chl *c*-containing algae have LHCs that bind Chls *a* and *c*. These proteins also bind carotenoids that expand the absorption spectrum and

are especially important in aquatic photosynthesis [24]. A range of carotenoids is present in algal plastids, with the relative abundance of each carotenoid varying between species. There is diversity in the most abundant carotenoid among these groups, with chlorophytes having lutein, rhodophytes zeaxanthin, and lineages of Chl *c*-containing algae have several distinct primary carotenoids. More specifically, cryptophytes have alloxanthin, most heterokonts and haptophytes have fucoxanthin, dinoflagellates with a plastid of red algal origin have peridinin (hereafter “peridinin-containing dinoflagellates”), while dinoflagellates with a plastid of haptophyte origin (hereafter “fucoxanthin-containing dinoflagellates”) have a fucoxanthin derivative found in haptophytes [17,25]. It is generally thought that the primary carotenoid in each lineage constitutes the major component of LHC-bound pigments. Thus, LHCs naturally bind the Chl and carotenoid molecules present in their native lineage, but reconstruction experiments have shown that they are capable of binding non-native pigments [26]. Differences in primary pigment binding are at least as much a function of organismal biology as they are the result of fundamental differences in LHC biochemistry, although affinities for specific pigments do vary.

Most of the current biochemical, structural and expression research has focused on chlorophytes or plants and there are relatively few data on Chl *c*-containing algal LHCs. Previous biochemical studies of Chl *c*-containing algal LHCs have focused on *Cyclotella* spp., *Phaeodactylum tricornutum* and *Amphidinium carterae* [22,27,28]. Molecular sequences from a significant number of species have been published, but previous comparative or phylogenetic analyses have been limited to individual genomic sequences [e.g. [29]] or have not focused on Chl *a/c* sequences [e.g. [2]]. Presumably because of these limitations, some previous studies suggested that each Chl *c*-containing lineage possessed a single unique LHC subfamily [e.g. [29]]. Other studies identified a subfamily of LHCs that is highly conserved between chlorophytes and diatoms, and on that basis concluded that a single lineage possesses multiple subfamilies of LHCs and that the subfamily was not lineage specific [30,31]. However, no previous study has undertaken a comprehensive analysis of this complex gene family in Chl *c*-containing algae. Consequently, much current research on these LHCs is based on tenuously supported assumptions about relationships within the gene family. In addition, new biochemical [22,23,32] and expression data [32-36] on LHCs have recently become available that have not previously been related to the phylogeny of the protein family.

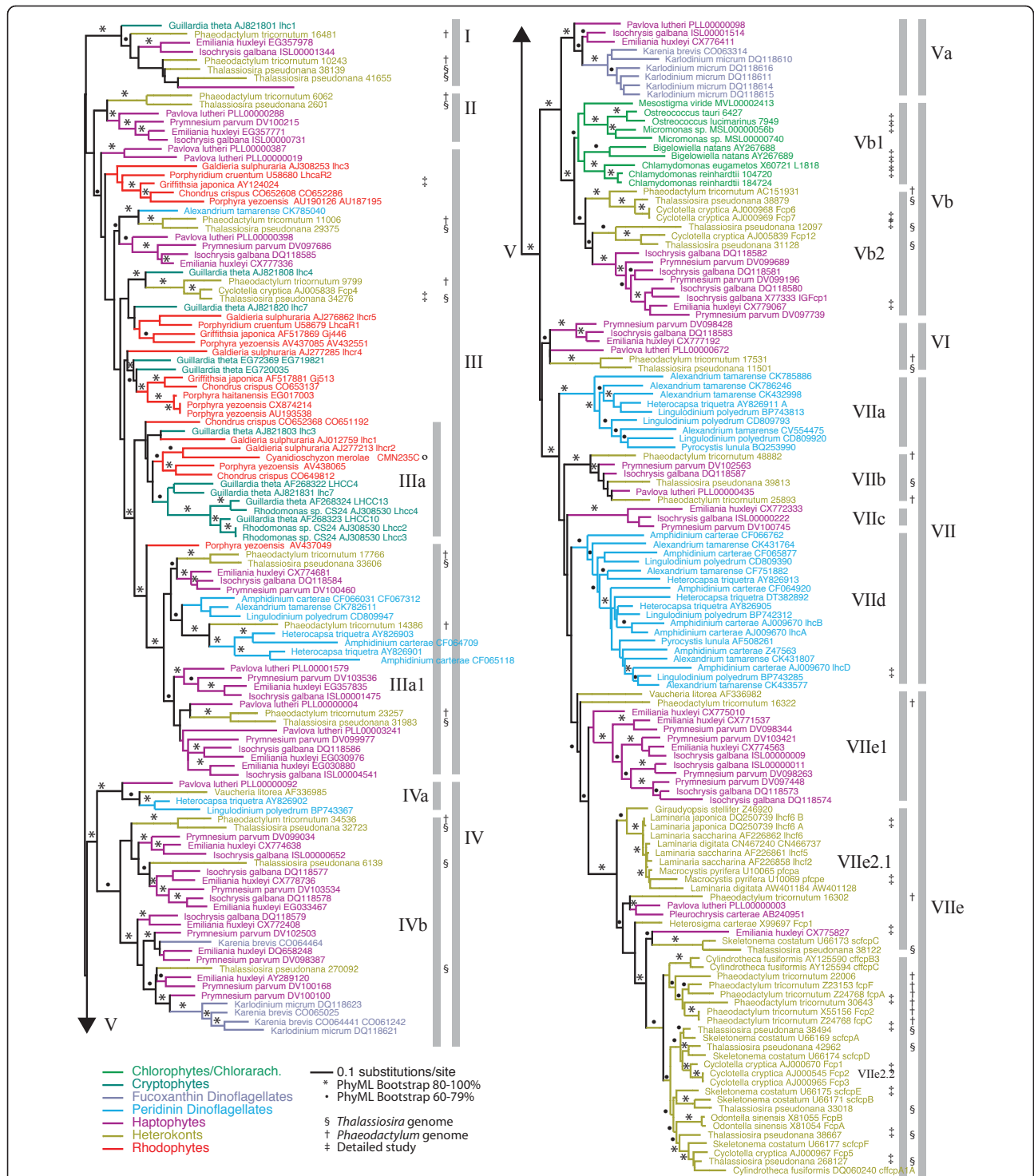
## Results and Discussion

### Overview of the Phylogeny

To help ensure that the gene phylogeny was comprehensive, but bounded by objective criteria, sequences were selected by BLAST analysis with a relatively low threshold, and then screened for the feasibility of end-to-end alignment. Analysis by motif detection programs indicated that the three TMRs are well conserved compared to the regions exposed to the plastid stroma and thylakoid lumen. These conserved regions generally included carotenoid and Chl binding sites. This protein architecture of three conserved TMRs was used to identify a set of LHCs that were practical for analysis.

Phylogenetic analyses were performed with a variety of methods and analytical conditions to ensure that the major features of the tree topology were not method-dependent. Analyses were performed using maximum likelihood methods implemented in PhyML [37] for an amino acid alignment and Garli [38] for the corresponding nucleotide alignment. Bayesian methods implemented in MrBayes [39] were used to analyze both alignments. In general, agreement was good among the analyses for features with strong bootstrap support (>80%) or posterior probability (>0.95), but substantial differences were observed in features with low support. Among the analyses performed (see *Materials and Methods* below), the tree found by PhyML with amino acid data (PhyML/AA) was a good representation of the overall analytical findings, and had relatively high support for features also observed in the other analyses (Figure 1). Particularly among deep branches, the PhyML/AA tree is well resolved compared to other analyses. The support values from the other analyses are given in Additional File1, Figure S1. To facilitate discussion of what is potentially a very confusing gene phylogeny, we have established a systematic nomenclature for well-supported clades (Figure 1). The tree is also labeled with corresponding protein names that are in most cases based on biochemical analyses in specific model systems, but because not all of the clades identified here have been characterized biochemically (or directly shown to be functionally distinct), we prefer a nomenclature that reflects the gene phylogeny because we feel it is likely to be relatively stable.

Our analysis is particularly reliable because we apply multiple robust statistical methods of tree reconstruction and include roughly 10-fold denser sampling of LHCs from Chl *c*-containing taxa and related species. Simulations have demonstrated that the probability-based methods employed here are superior in terms of reconstruction accuracy [40] compared to the methods used by Durnford et al. [29]. Moreover, Zwickl and Hillis [41] demonstrate that increased taxon sampling has a



**Figure 1** The maximum likelihood tree from PhyML, with 246 taxa based on a 266 aa alignment (PhyML/AA). The individual LHCs are colored by taxonomic lineage: chlorophytes and chlorarachniophytes (green), cryptophytes (turquoise), fucocanthin-containing dinoflagellates (light purple), peridinin-containing dinoflagellates (light blue), haptophytes (dark purple), heterokonts (brown), rhodophytes (red). The taxon name and sequence identifier are given for each sequence, and gene name is given when available. See Additional File 1, Additional Figure S1 for complete support values from other analyses.

marked effect on reconstruction accuracy so that the phylogeny presented here is expected to be substantially more reliable for the relevant LHC proteins than that of Koziol et al. [2]. Distance optimality, neighbor-joining and parsimony based methods had lower support for most clades, but were compatible with the maximum likelihood phylogeny (data not shown).

The inferred phylogeny is unrooted (see below, in *Comparison to LHC superfamily*) and resolves the LHCs into seven major clades (I-VII), several of which are defined by a long, well-supported branch (Figure 1). The clades vary substantially in number of members and representation of algal lineages.

It is clear that most organismal lineages contain LHCs from multiple subfamilies in addition to duplicates with very similar sequences. This refutes the still widely accepted hypothesis that each LHC subfamily is characteristic of a particular group of organisms. Furthermore, there are many cases where clusters of genes represent several different organismal lineages, and some such clusters appear multiple times within the tree. The most common of these is the association of sequences from three Chl *c*-containing lineages (haptophytes, heterokonts and dinoflagellates). There are also some distinctive absences, for example, cryptophyte LHCs are only present in clades I and III, and rhodophyte LHCs are present only in clade III, while sequences from chlorophytes are present only in clade V. Taken together, these observations strongly imply that LHCs subfamilies are functionally distinct and under differential selection in many organisms.

### The role of endosymbiosis in the phylogeny

There is reason to think that the transfer of LHCs from endosymbiont to host genome has occurred repeatedly in the evolution of the gene family. The ancestral rhodophyte and chlorophyte plastid genomes were inherited from cyanobacterial ancestors, with relatively early and rapid transfer from the plastid to the host cell nuclear genome [42]. This transfer presumably occurred subsequent to the permanent incorporation of the plastid in the host cell, although this is not necessarily the case. Particularly in the context of kleptoplastidic organisms that appear to have obtained some genes from their prey [e.g. [43]], it is reasonable to think that some transfer of genes may have occurred prior to the acquisition of a plastid *per se* [10]. This is further complicated by the transfer of plastids via secondary and tertiary endosymbioses [8], followed by gene transfer from the eukaryotic endosymbiont to the host nuclear genome.

The overall structure of the phylogeny is consistent with the hypothesis that Chl *c*-containing algal LHCs descended from the rhodophyte proteins to create the diversity of extant sequences, although the high level of

sequence divergence makes it difficult to infer the root. Related LHCs from chlorophytes (Lhca and Lhcb) constitute an outgroup, but have such low sequence similarity to those studied here that it was not possible to infer the root with reasonable confidence. Koziol et al. [2] and Durnford et al. [44] placed representatives of the set of proteins studied here (specifically clades V and VII) as sister to the group containing Lhca and Lhcb. Durnford et al. [29] placed the set as sister to Lhcb and some Lhca proteins with other Lhca proteins as an outgroup, although this arrangement did not have strong bootstrap support. The rooting shown here is compatible with those hypotheses, but several alternate rootings would be equally plausible.

The position of LHCs from the chlorarachniophyte *Bigeloviella natans* deeply nested among chlorophytes is expected because the *B. natans* plastid is of chlorophyte origin. Similarly, LHCs from fucoxanthin-containing dinoflagellates cluster with haptophytes, the source of their tertiary plastids [17]. If the diversification of the Chl *c*-containing algal LHCs entirely predated the acquisition of plastids from red algae then one would predict that each major clade in the gene phylogeny would have red algal sequences at its base. This is not observed, but it may be an artifact of incomplete data, as the only complete red algal genome analyzed here was from the highly reduced picoeukaryote *Cyanidioschyzon merolae*. We expect that other red algae will show a more diverse set of LHC genes.

### Diversity within individual lineages

Much of the diversity of LHC proteins - both within and among organisms - seen here (Figure 1) has not been previously characterized. Previous phylogenetic studies have included a relatively small number of LHC subfamilies, most of which were assumed to be lineage specific [29,44,45]. In addition, some algal lineages were thought to have multiple LHCs belonging to a single or a small number of LHC subfamilies. The current analysis greatly expands the number of putative subfamilies and indicates that some are lineage specific while others are not.

Previous studies [29,30,45-47] concluded that dinoflagellates possess only a single LHC subfamily (i.e. VIId), which was taken to be lineage specific. However, these analyses included only between 1 and 4 dinoflagellate sequences, all from *Amphidinium carterae*. Similarly, haptophytes were thought to possess LHCs from only a single subfamily that grouped with diatom and chlorophyte sequences in subfamily Vb [29,30,46]. A haptophyte LHC was later identified in clade VII [47]. A single large lineage specific gene subfamily was identified in heterokonts in the above analyses, corresponding to the clade VIIe2 of heterokonts in this study. Eppard

et al. [30], Green [47] and Koziol et al. [2] incorporated more sequence data and identified heterokont sequences that grouped with rhodophytes in clade III and chlorophytes in clade V.

The incorporation of newly available sequence data illustrates the diversity and complexity of LHCs. In clade VII, which constitutes the major Chl *c*-containing clade, heterokonts contain a lineage specific LHC subfamily with sequences from diatoms and brown algae (VIIe2). Similarly, peridinin dinoflagellates possess LHCs from the lineage specific subfamilies VIIa and VIId. Subfamilies specific to haptophytes are VIIc and VIIe1, but in other clades sequences from this lineage often have close homologs in diatoms (e.g. VI and VIIb). Clade V comprises subfamilies from haptophytes, diatoms, fucoxanthin-containing dinoflagellates, chlorophytes and chlorarachniophytes. Clade III includes sequences from many lineages, and has subfamilies that are unresolved or poorly supported. In this case, cryptophyte and rhodophyte LHCs group together, as do those from diatoms, dinoflagellates and haptophytes.

#### LHCs from complete genomes and possible biases in the data

At the time of this study, the only relevant complete genomes with substantial representation of the LHC family considered in this analysis were the diatoms (heterokont) *Thalassiosira pseudonana* [48] and *Phaeodactylum tricornerutum* [49] (Additional File 2, Additional Table S1). Our analysis of *T. pseudonana* and *P. tricornerutum* identified members of all seven LHC subfamilies, representing nearly the full diversity of sequences studied (Figure 1). The LHCs of these species are the only representatives from diatoms in most subfamilies outside of VIIe2, (with the exception of those from *Cy. cryptica*). This is an important indication that other techniques are likely to miss members of the family. Indeed, EST and other transcriptomic studies almost always represent a subset of all of the genes present within the genome. Furthermore, with the exception of those from complete genome analyses (Additional File 2, Additional Table S1), the individual genomic sequences included in the analysis were obtained via Western blots using an antibody from a similar LHC [e.g. [50]], Southern blots using a probe from a similar LHC [e.g. [51]] or PCR using primers based on an LHC alignment [e.g. [52]]. Because these methods only detect proteins that are similar to the LHC used to create the antibody, probe, or primers, the set of sequences identified by these methods is intrinsically limited, and consequently this creates another source of sampling bias.

Taken together, these observations indicated that the diversity in LHCs reported here is an underestimate. It is also noteworthy that only a small fraction of the

sequences included in this analysis have been characterized biochemically. Because sequence annotation often relies on the assumption that homologous sequences share biochemical function, a complex gene family of this type is at risk of erroneous inferences due to transitive annotation. Consequently, one important prediction of this analysis is that there are uncharacterized functional differences among the several subfamilies.

At the same time, it is clearly possible for an organism to function with a small complement of LHC genes. The chlorophytes *Ostreococcus tauri* and *O. lucimarinus* have highly reduced genomes, and encode clade V LHCs that are similar to other proteins from chlorophytes (including *Chlamydomonas reinhardtii*). The rhodophyte *Cyanidioschyzon merolae*, which also has a highly reduced genome [53], possesses only a single family IIIa LHC as well as two very similar sequences identified by Koziol et al. [2] and additional homologs that did not have the conserved domain architecture or degree of sequence conservation required for inclusion in this study. This is strikingly different from the diversity of genes already identified from EST analyses of red algae. *C. merolae* has been shown to be unusual in other ways, and this is not necessarily surprising [54,55].

#### A single subfamily can comprise LHCs that bind different Chl and carotenoids

Clade V comprises LHCs from diatoms, chlorophytes, haptophytes and fucoxanthin-containing dinoflagellates. Chlorophytes possess Chl *a/b* and use lutein as the primary LHC carotenoid, while diatoms, haptophytes and dinoflagellates use Chl *a/c* and fucoxanthin (or derivatives). The grouping of these sequences into a single clade implies that similar LHCs can bind a diverse set of Chls and carotenoids. This observation parallels *in vitro* reconstruction experiments performed by Grabowski et al. [26] where non-native Chl and primary carotenoids were functionally inserted into LhcaR1 from the rhodophyte *P. cruentum* in clade III. The reconstructed LHC was able to functionally bind Chl *a*, *b* and *c* as well as the carotenoids lutein, fucoxanthin and peridinin in addition to its native zeaxanthin, albeit with some variation in affinity. Thus, it appears that the binding of primary carotenoids and certain chlorophylls is due to the nature of the endogenous biosynthesis of Chlorophyll and carotenoids, and not necessarily because of a profound incongruence in the structure or binding properties of the protein; this inference remains to be tested *in vivo*.

#### Functional specialization

Despite the plasticity of pigment binding, the persistence of multiple subfamilies of LHC over hundreds of millions of years of evolutionary divergence (Figure 1)

suggest that these subfamilies are functionally distinct. Unfortunately, only a small fraction of the gene products corresponding to the sequences studied here have been characterized biochemically. Nonetheless, placing differences in expression patterns, pigment binding, and differential physical associations among LHCs in a phylogenetic context, it is apparent that individual subfamilies have undergone specialization (See Additional File 3, Supplementary Information).

Plants and chlorophytes have two major types of LHCs, namely LHC I and II, which have distinct biochemistries and associate primarily with PS I and II, respectively. However, the LHCs of rhodophytes and Chl *c*-containing algae constitute a sister clade to most chlorophyte sequences, meaning that they diverged independently of those in the green lineage. Consequently, the diversity of LHCs in Chl *c*-containing algae is generally distinct and independent of that observed in plants and chlorophytes [29], with the notable exception of subfamily Vb. Moreover, the rhodophyte LHCs associate only with PS I, so proteins that associate with PS II like those of chlorophytes are thought to have evolved independently in Chl *c*-containing algae and chlorophytes and generally do not show PS specificity [25]. In rhodophytes an unrelated set of phycobiliproteins are associated with PS II [56].

The sequence diversity and biochemical differences observed in the LHC gene phylogeny suggest that members of the protein family have undergone functional specialization. This complements the work of Koziol et al. [2] on chlorophyte LHCs and extends it to the LHCs of Chl *c*-containing algae. The conservation of LHCs from multiple and diverse subfamilies in individual organisms or closely related lineages suggests that the subfamilies are functionally distinct and could exhibit differences in pigment binding, absorption spectrum, strength of interaction with thylakoid membrane and differences in expression and associations with other LHCs. In addition to the advantages of increased gene dosage, having such a diversity of LHCs may allow an organism to carry out photosynthesis or photoprotection under a range of conditions.

The phylogeny of LHCs observed here cannot be easily reconciled with the currently accepted organismal phylogeny [13], and the presence of many subfamilies that contain proteins from diverse algal lineages is problematic. One explanation would be that the complex phylogeny is the result of extensive horizontal gene transfer between taxa in distant lineages. However, even the most parsimonious scenario would require a very large number of transfer events, and we favor the interpretation that some of the subfamilies are ancestral, and that there is substantial sequence diversity that reflects selection on functionally distinct proteins.

This presents a new perspective on LHC diversity, and illustrates the difficulty inferring functional similarity from sequence similarity. Although there is no doubt that the sequences studied here are all homologous (i.e., are derived from a single common ancestral sequence), the sequence diversity and evolutionary complexity of the gene family are far greater than was appreciated until recently. Most troubling is the biochemical difference between the *Cy. cryptica* Fcp1-3 and Fcp5 in subfamily VIIe2.2 because the former is thought to associate in heterotrimers with Fcp6 and Fcp7 but the latter in higher oligomers despite high sequence similarity. If the relationships shown here are accurate and the biochemical data reliable, then one would predict great potential for functional differences among the many uncharacterized LHCs.

There can be little doubt that a complex history of gene duplication, gene loss, and functional specialization shaped the phylogeny presented here. Closely related paralogs in the same lineage are likely the result of recent gene duplication events, while more distantly related proteins or LHCs that are shared by distant lineages are likely the result of ancestral duplications, or perhaps horizontal gene transfer. Following these ancestral events, gene loss led to the phylogeny observed in this study, with extreme (but not convergent) reduction having occurred in both *Ostreococcus* and *Cyanidioschyzon*. Although much of the apparent gene loss might be attributed to incomplete sampling, the absence of cryptophyte, rhodophyte and peridinin dinoflagellate LHCs from clade V is noteworthy. Proteins from this group are thought to be involved in photoprotection by binding xanthophyll cycle carotenoids and facilitating non-photochemical quenching [22,57]. The presence of a xanthophyll cycle has been demonstrated in all of the lineages with clade V proteins, but to our knowledge has not been demonstrated in cryptophytes or peridinin-containing dinoflagellates. One hypothesis would be that proteins from clade V are involved in the xanthophyll cycle. Brown algae have a xanthophyll cycle, but a brown algal homolog in clade V had not been characterized at the time that the analysis was performed. We speculated that this was a result of the sampling bias described above, and indeed, two very recent studies of brown algae (*Fucus spp.* and *Ectocarpus siliculosus*) found sequences that belong to the clade V LHCs [58,59].

This study characterized the large diversity of LHC sequences and demonstrated that individual species have LHCs from multiple subfamilies, and that subfamilies are not lineage specific. Relating the molecular phylogeny to the biochemistry of the protein has the potential to add substantial insight into the evolution and functional divergence of LHCs. It also tells a

cautionary tale for the inference of function solely on the basis of sequence similarity, and suggests the importance of independent lines of evidence in the analysis of genomic data.

## Conclusions

This study provides a comprehensive overview of an important subfamily of LHC genes. Phylogenetic analysis of gene family sequences can provide valuable insights into the diversification of gene families, and a lack of understanding of relationships among LHC sequences has very likely led to incorrect inference of function. Further work will be needed to validate the predictions made here, but it is very likely that several of the clades identified here represent functionally distinct entities, and correctly distinguishing among these sequence variants should help understand their biochemical function. It is also possible that horizontal gene transfer and gene duplication *without* functional differentiation of the duplicated genes have played a role in the evolution of LHCs. Taken together, these observations illustrate the power of comparative sequence analysis for functional inference, but also demonstrate the potential pitfalls in overly simplistic analyses.

## Methods

### Sequence data acquisition

Sequence data from EST surveys of 19 algal species, individual genomic sequences from 25 species, and complete genomic data from 6 species were screened for LHC homologs (Additional File 2, Additional Table S1). Putative LHCs were identified with BLAST [60] using Chl *c*-containing algae LHC sequences from Durnford et al. [29] as queries. The sequences identified by BLAST were clustered with Sequencher (Gene Codes, Ann Arbor, MI, USA) to remove duplicate and nearly identical sequences from the same species. Reads were clustered if they had > 92% identity for a 40 bp overlapping region.

MEME [61] and MAST [62] were used to further characterize the conserved domain structure of the gene family. The Chl *c*-containing algae sequences from Durnford et al. [29] were used as a training dataset to create a local motif profile of the gene family using MEME. MAST was then used to compare the set of BLAST hits against this profile. The characterization of local motifs in the dataset facilitated the identification of a subset of sequences that could be aligned. As expected from previous studies [19,29] the conserved regions corresponded to the three transmembrane regions plus surrounding sequence including Chl and carotenoid binding sites. BLAST hits above an *e*-value of  $1e-19$  or without the expected

domain structure could not be aligned well with other LHCs and were not included in subsequent analyses. In addition, individual LHCs from polycistronic [63] or polyprotein [64,65] sequences were treated as separate sequences. Sequence data were organized and manipulated using the BioPerl toolkit [66].

### Sequence alignment and phylogenetic analyses

An amino acid alignment of LHC sequences was constructed with ClustalW [67] and was edited manually using MacClade 4.0 (Sinauer Associates, Sunderland, MA, USA). Other publicly available alignment programs were used, but the alignment from ClustalW was deemed to be the most biologically relevant based on the location of gaps in relation to the previously reported conserved regions and the known tertiary structure of the LHC protein. The amino acid alignment was used to create the nucleotide alignment for further analysis [68]. Poorly conserved regions of the protein were omitted from the phylogenetic analysis, so that the alignment consisted mostly of TMRs and each sequence had an average of 149 amino acids out of 266 positions in the alignment. Since these regions span the membrane, the amino acid composition is biased in favor of hydrophobic residues. Because this bias is not adequately addressed by amino acid substitution models and due to the short length of the alignment, both the amino acid and corresponding nucleotide alignment were used for phylogenetic analysis. Preliminary datasets consisted of 470 LHC sequences that fit the criteria above, but individual operational taxonomic units (OTUs) were omitted to reduce the computation time and facilitate the interpretation of the phylogenetic tree. The OTUs corresponding to very similar orthologs within a single lineage were omitted so that the final dataset containing 246 OTUs maintained the same sequence diversity and taxon distribution as the original. Orthologs from complete genomes were not omitted. Comparisons of the topologies from the two datasets indicated that they were consistent. The alignment has been deposited in TreeBase under the accession number 10532.

Phylogenetic analyses based on nucleotides were conducted under the maximum likelihood optimality criterion using the GTR + I +  $\Gamma$  model with 4 rate categories implemented in MrBayes 3.1.2 [39] and Garli 0.942 [38]. The model was selected using MrModeltest 2.2 [69]. MrBayes was run for 10 000 000 generations, trees were sampled every 1000 generations and a burnin of 25% was used to create a consensus tree. Garli was run with the default parameters for the tree search and 200 bootstrap replicates. The same parameters were used to analyze the nucleotide alignment without third codon-position nucleotides.



Phylogenetic analyses of the amino acid alignment were conducted with MrBayes and PhyML 2.4.5 [37]. MrBayes was run with the mixed fixed-rate model and the parameters used above. The consensus tree was used as the starting topology for the PhyML analysis, which used WAG + I +  $\Gamma$  and 8 rate categories for the tree search and 100 bootstrap replicates. The model was selected using ProtTest 1.3 [70]. This tree is presented in Figure 1. See Additional File 1, Additional Figure S1 for equivalent tree with support values from other analyses.

## Additional material

**Additional file 1: List of organisms from which LHC sequence data was examined in the current analysis.** Data sources and citation for sequence data use in this analysis

**Additional File 2: Detailed LHC phylogeny.** The maximum likelihood tree from PhyML, with 246 taxa based on a 266 aa alignment (PhyML/AA). This tree is equivalent to that in Figure 1, but contains all the sequence names and support values from additional analyses. The support values for each branch were obtained as follows: PhyML with amino acids (top left), MrBayes with amino acids (bottom left), Garli with nucleotides (top right), MrBayes with nucleotides (bottom right). The individual LHCs are colored by taxonomic lineage: chlorophytes (green), cryptophytes (light blue), fucoxanthin-containing dinoflagellates (light purple), peridinin-containing dinoflagellates (dark blue), haptophytes (pink), heterokonts (orange), rhodophytes (red).

**Additional File 3: Supplementary Information.** This file contains added information detailing the analytical methods used and details on the interpretation of the tree that are not included in the main manuscript.

## Acknowledgements

The sequence data from *Chlamydomonas reinhardtii*, and *Ostreococcus tauri*, *Ostreococcus lucimarinus*, *Phaeodactylum tricorutum* and *Thalassiosira pseudonana* were produced by the US Department of Energy Joint Genome Institute, <http://www.jgi.doe.gov> and are provided for use in this publication only. This project was supported by NSF grant MCB-9984284 and the Howard Hughes Medical Institute Undergraduate Science Education Program at the University of Maryland. We would also like to thank John D. Hall for his help throughout the entire process.

## Author details

<sup>1</sup>Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD, 20742, USA. <sup>2</sup>Maryland Agricultural Experiment Station, University of Maryland, College Park, MD, 20742, USA. <sup>3</sup>Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY, 14853, USA. <sup>4</sup>IBAM, Facultad de Ciencias Agrarias, Universidad Nacional de Cuyo, Chacras de Coria, 5500 Mendoza, Argentina.

## Authors' contributions

GEH carried out the analysis and drafted the manuscript. MVSP and CFD conceived of the study, and participated in its design and helped to draft the manuscript. All authors read and approved the final manuscript.

Received: 20 May 2010 Accepted: 15 April 2011 Published: 15 April 2011

## References

1. Green BR, Durnford DG: The chlorophyll-carotenoid proteins of oxygenic photosynthesis. *Annu Rev Plant Physiol Plant Mol Biol* 1996, **47**:685-714.

2. Koziol A, Borza T, Ishida KI, 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.
3. van Amerongen H, Dekker JP: Light-harvesting in photosystem II. In *Light-Harvesting Antennas in Photosynthesis*. Edited by: Green BR, Parson BR. Kluwer Academic Publishers; 2003.
4. Fromme P, Schlodder E, Jansson S: Structure and function of antenna system in photosystem I. In *Light-Harvesting Antennas in Photosynthesis*. Edited by: Green BR, Parson BR. Kluwer Academic Publishers; 2003.
5. Cavalier-Smith T: Principles of protein and lipid targeting in secondary symbiogenesis: Euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Eukaryot Microbiol* 1999, **46**(4):347-366.
6. Baurain D, Brinkmann H, Petersen J, Rodriguez-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.
7. Sanchez-Puerta MV, Delwiche CF: A Hypothesis for plastid evolution in chromalveolates. *J Phycol* 2008, **44**(5):1097-1107.
8. Delwiche CF: Tracing the thread of plastid diversity through the tapestry of life. *Am Nat* 1999, **154**:S164-S177.
9. Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, Bohnert HJ, Philippe H, Lang BF: Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. *Curr Biol* 2005, **15**(14):1325-1330.
10. Delwiche CF: Algae in the warp and weave of life: bound by plastids. In *Unravelling the algae: the past, present, and future of algal systematics Systematics Association Special Volume # 75*. Edited by: Brodie J, Lewis J. London: Taylor and Francis; 2007.
11. Stiller JW, Reel DC, Johnson JC: A single origin of plastids revisited: Convergent evolution in organellar genome content. *J Phycol* 2003, **39**(1):95-105.
12. Patron NJ, Inagaki Y, Keeling PJ: Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Curr Biol* 2007, **17**(10):887-891.
13. Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rummele SE, Bhattacharya D: Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of Rhizaria with Chromalveolates. *Mol Biol Evol* 2007, **24**(8):1702-1713.
14. Sanchez-Puerta MV, Bachvaroff TR, Delwiche CF: Sorting wheat from chaff in multi-gene analyses of chlorophyll c-containing plastids. *Mol Phylogeny Evol* 2007, **44**(2):885-897.
15. Yoon HS, Hackett JD, Pinto G, Bhattacharya D: The single, ancient origin of chromist plastids. *Proc Natl Acad Sci USA* 2002, **99**(24):15507-15512.
16. Bachvaroff TR, Puerta MVS, Delwiche CF: Chlorophyll c-containing plastid relationships based on analyses of a multigene data set with all four chromalveolate lineages. *Mol Biol Evol* 2005, **22**(9):1772-1782.
17. Tengs T, Dahlberg OJ, Shalchian-Tabrizi K, Klaveness D, Rudi K, Delwiche CF, Jakobsen KS: Phylogenetic analyses indicate that the 19' hexanoxyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol Biol Evol* 2000, **17**(5):718-729.
18. Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D: Genomic Footprints of a Cryptic Plastid Endosymbiosis in Diatoms. *Science* 2009, **324**(5935):1724-1726.
19. Green BR, Pichersky E: Hypothesis for the evolution of 3-helix chl a/b and chl a/c light-harvesting antenna proteins from 2-helix and 4-helix ancestors. *Photosynthesis Res* 1994, **39**(2):149-162.
20. Liu ZF, Yan HC, Wang KB, Kuang TY, Zhang JP, Gui LL, An XM, Chang WR: Crystal structure of spinach major light-harvesting complex at 2.72 angstrom resolution. *Nature* 2004, **428**(6980):287-292.
21. Standfuss R, van Scheltinga ACT, Lamborghini M, Kuhlbrandt W: Mechanisms of photoprotection and nonphotochemical quenching in pea light-harvesting complex at 2.5A resolution. *EMBO J* 2005, **24**(5):919-928.
22. Beer A, Gundermann K, Beckmann J, Buchel C: Subunit composition and pigmentation of fucoxanthin-chlorophyll proteins in diatoms: Evidence for a subunit involved in diadinoxanthin and diatoxanthin binding. *Biochemistry* 2006, **45**(43):13046-13053.
23. Buchel C: Fucoxanthin-chlorophyll proteins in diatoms: 18 and 19 kDa subunits assemble into different oligomeric states. *Biochemistry* 2003, **42**(44):13027-13034.

24. Mimuro M, Akimoto S: **Carotenoids of light harvesting systems: Energy transfer processes from fucoxanthin and peridinin to chlorophyll.** In *Photosynthesis in Algae*. Edited by: Larkum AWD, Douglas SE, Raven JA: Kluwer Academic Publishers; 2003:335-349.
25. Macpherson AN, Hiller RG: **Light-harvesting systems in chlorophyll c-containing algae.** In *Light-Harvesting Antennas in Photosynthesis*. Edited by: Green BR, Parson WW. Kluwer Academic Publishers; 2003:323-352.
26. Grabowski B, Cunningham FX, Gantt E: **Chlorophyll and carotenoid binding in a simple red algal light-harvesting complex crosses phylogenetic lines.** *Proc Natl Acad Sci USA* 2001, **98**(5):2911-2916.
27. Polivka T, van Stokkum IHM, Zigmantas D, van Grondelle R, Sundstrom V, Hiller RG: **Energy transfer in the major intrinsic light-harvesting complex from Amphidinium carterae.** *Biochemistry* 2006, **45**(28):8516-8526.
28. Lepetit B, Volke D, Szabo M, Hoffmann R, Garab GZ, Wilhelm C, Goss R: **Spectroscopic and molecular characterization of the oligomeric antenna of the diatom Phaeodactylum tricorutum.** *Biochemistry* 2007, **46**(34):9813-9822.
29. Durnford DG, Deane JA, Tan S, McFadden GI, Gantt E, Green BR: **A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastid evolution.** *J Mol Evol* 1999, **48**(1):59-68.
30. Eppard M, Krumbain WE, von Haeseler A, Rhiel E: **Characterization of fcp4 and fcp12, two additional genes encoding light harvesting proteins of Cyclotella cryptica (Bacillariophyceae) and phylogenetic analysis of this complex gene family.** *Plant Biol* 2000, **2**(3):283-289.
31. Richard C, Ouellet H, Guertin M: **Characterization of the LI818 polypeptide from the green unicellular alga Chlamydomonas reinhardtii.** *Plant Mol Biol* 2000, **42**(2):303-316.
32. Becker F, Rhiel E: **Immuno-electron microscopic quantification of the fucoxanthin chlorophyll a/c binding polypeptides Fcp2, Fcp4, and Fcp6 of Cyclotella cryptica grown under low- and high-light intensities.** *Int Microbiol* 2006, **9**(1):29-36.
33. Janssen M, Bathke L, Marquardt J, Krumbain WE, Rhiel E: **Changes in the hot apparatus of diatoms in response to low and high light intensities.** *Int Microbiol* 2001, **4**:27-33.
34. Oeltjen A, Krumbain WE, Rhiel E: **Investigations on transcript sizes, steady state mRNA concentrations and diurnal expression of genes encoding fucoxanthin chlorophyll a/c light harvesting polypeptides in the centric diatom Cyclotella cryptica.** *Plant Biol* 2002, **4**(2):250-257.
35. Oeltjen A, Marquardt J, Rhiel E: **Differential circadian expression of genes fcp2 and fcp6 in Cyclotella cryptica.** *Int Microbiol* 2004, **7**(2):127-131.
36. Westermann M, Rhiel E: **Localisation of fucoxanthin chlorophyll a/c-binding polypeptides of the centric diatom Cyclotella cryptica by immuno-electron microscopy.** *Protoplasma* 2005, **225**(3-4):217-223.
37. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52**(5):696-704.
38. Zwickl DJ: **Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion.** The University of Texas at Austin; 2006, Ph.D. dissertation.
39. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**(12):1572-1574.
40. Huelsenbeck JP: **Performance of phylogenetic method in simulation.** *Syst Biol* 1995, **44**(1):17-48.
41. Zwickl DJ, Hillis DM: **Increased taxon sampling greatly reduces phylogenetic error.** *Syst Biol* 2002, **51**(4):588.
42. Palmer JD, Delwiche CF: **The origin and evolution of plastids and their genomes.** In *Molecular Systematics of Plants II*. Edited by: Soltis DE, Soltis PS, Doyle JJ. Boston: Kluwer Academic Publishers; 1998:375-409.
43. Johnson MD, Oldach D, Delwiche CF, Stoecker DK: **Retention of transcriptionally active cryptophyte nuclei by the ciliate Myrionecta rubra.** *Nature* 2007, **445**(7126):426-428.
44. Durnford DG, Aebersold R, Green BR: **The fucoxanthin-chlorophyll proteins from a chromophyte alga are part of a large multigene family: Structural and evolutionary relationships to other light harvesting antennae.** *Mol Gen Genet* 1996, **253**(3):377-386.
45. Caron L, Douady D, Quinetszely M, deGoer S, Berkaloff C: **Gene structure of a chlorophyll a/c-binding protein from a brown alga: Presence of an intron and phylogenetic implications.** *J Mol Evol* 1996, **43**(3):270-280.
46. Deane JA, Fraunholz M, Su V, Maier UG, Martin W, Durnford DG, McFadden GI: **Evidence for nucleomorph to host nucleus gene transfer: Light-harvesting complex proteins from cryptomonads and chlorarachniophytes.** *Protist* 2000, **151**(3):239-252.
47. Green BR: **The evolution of light-harvesting antennas.** In *Light-Harvesting Antennas in Photosynthesis*. Edited by: Green BR, Parson BR. Kluwer Academic Publishers; 2003:.
48. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou SG, Allen AE, Apt KE, Bechner M, et al: **The genome of the diatom Thalassiosira pseudonana: Ecology, evolution, and metabolism.** *Science* 2004, **306**(5693):79-86.
49. Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, O'tillar RP: **The Phaeodactylum genome reveals the evolutionary history of diatom genomes.** *Nature* 2008, **456**(7219):239-244.
50. Apt KE, Clendennen SK, Powers DA, Grossman AR: **The gene family encoding the fucoxanthin chlorophyll proteins from the brown alga Macrocyctis pyrifera.** *Mol Gen Genet* 1995, **246**(4):455-464.
51. Kroth-Pancic PG: **Nucleotide sequence of 2 cDNAs encoding fucoxanthin chlorophyll a/c proteins in the diatom Odontella sinensis.** *Plant Mol Biol* 1995, **27**(4):825-828.
52. Smith GJ, Gao Y, Alberte RS: **The fucoxanthin-chlorophyll a/c proteins comprise a large family of coexpressed genes in the marine diatom Skeletonema costatum (Greve).** *Plant Physiol* 1997, **114**:1136.
53. Matsuzaki M, Misumi O, Shin-I T, Maruyama S, Takahara M, Miyagishima SY, Mori T, Nishida K, Yagisawa F, Nishida K, et al: **Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D.** *Nature* 2004, **428**(6983):653-657.
54. Toplin JA, Norris TB, Lehr CR, McDermott TR, Castenholz RW: **Biogeographic and Phylogenetic Diversity of Thermoacidophilic Cyanidiales in Yellowstone National Park, Japan, and New Zealand.** *Appl Environ Microbiol* 2008, **74**(9):2822-2833.
55. Claudia C, Hwan Su Y, Antonino P, Gabriele P, Debashish B: **Hidden biodiversity of the extremophilic Cyanidiales red algae.** *Mol Ecol* 2004, **13**(7):1827-1838.
56. Grossman AR, Schaefer MR, Chiang GG, Collier JL: **The phycobilisome, a light-harvesting complex responsive to environmental conditions.** *Microbiol Rev* 1993, **57**(3):725-749.
57. Muller P, Li XP, Niyogi KK: **Non-photochemical quenching. A response to excess light energy.** *Plant Physiol* 2001, **125**(4):1558-1566.
58. Pearson GA, Hoarau G, Lago-Leston A, Coyer JA, Kube M, Reinhardt R, Henckel K, Serrao 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**(2):195-213.
59. Dittami SM, Michel G, Collen J, Boyen C, Tonon T: **Chlorophyll-binding proteins revisited- a multigenic family of light-harvesting and stress proteins from a brown algal perspective.** *BMC Evol Biol* 2010, **10**(1):365.
60. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**(17):3389-3402.
61. Bailey TL, Elkan C: **Fitting a mixture model by expectation maximization to discover motifs in biopolymers.** *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology: 1994 Menlo Park, California: AAAI Press; 1994.*
62. Bailey TL, Gribskov M: **Combining evidence using p-values: application to sequence homology searches.** *Bioinformatics* 1998, **14**(1):48-54.
63. de la Espina SMD, Alverca E, Cuadrado A, Franca S: **Organization of the genome and gene expression in a nuclear environment lacking histones and nucleosomes: the amazing dinoflagellates.** *Eur J Cell Biol* 2005, **84**(2-3):137-149.
64. Hiller RG, Wrench PM, Sharples FP: **The light-harvesting chlorophyll a/c-binding protein of dinoflagellates - a putative polyprotein.** *FEBS Lett* 1995, **363**(1-2):175-178.
65. Koziol AG, Durnford DG: **Euglena light-harvesting complexes are encoded by multifarious polyprotein mRNAs that evolve in concert.** *Mol Biol Evol* 2008, **25**(1):92-100.
66. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JGR, Korf I, Lapp H, et al: **The bioperl toolkit: Perl modules for the life sciences.** *Genome Res* 2002, **12**(10):1611-1618.
67. Thompson JD, Higgins DG, Gibson TJ: **Clustal-W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**(22):4673-4680.

68. Bininda-Emonds ORP: **transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences.** *BMC Bioinformatics* 2005, **22**(6):156.
69. Nylander JAA: **MrModeltest v2.2.** *Evolutionary Biology Centre, Uppsala University* 2004 [<http://www.abc.se/~nylander/>].
70. Abascal F, Zardoya R, Posada D: **ProtTest: selection of best-fit models of protein evolution.** *Bioinformatics* 2005, **21**(9):2104-2105.
71. Jansson S, Green B, Grossman AR, Hiller R: **A proposal for extending the nomenclature of light-harvesting proteins of the three transmembrane helix type.** *Plant Molecular Biology Reporter* 1999, **17**(3):221-224.
72. Gagne G, Guertin M: **The early genetic response to light in the green unicellular alga *Chlamydomonas eugametos* grown under light dark cycles involves genes that represent direct responses to light and photosynthesis.** *Plant Mol Biol* 1992, **18**(3):429-445.

doi:10.1186/1471-2148-11-101

**Cite this article as:** Hoffman et al.: Evolution of light-harvesting complex proteins from Chl *c*-containing algae. *BMC Evolutionary Biology* 2011 11:101.

**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](http://www.biomedcentral.com/submit)

