

# Evolutionary Origins and Functions of the Carotenoid Biosynthetic Pathway in Marine Diatoms

Sacha Coesel<sup>1‡</sup>, Miroslav Oborník<sup>2</sup>, Joao Varela<sup>3</sup>, Angela Falciatore<sup>1</sup>, Chris Bowler<sup>1,4\*</sup>

**1** Cell Signalling Laboratory, Stazione Zoologica 'Anton Dohrn', Villa Comunale, Naples, Italy, **2** Institute of Parasitology, Biology Centre of the Academy of Sciences of the Czech Republic and Faculty of Science, University of South Bohemia, České Budejovice, Czech Republic, **3** Centre of Marine Sciences, University of Algarve, Faro, Portugal, **4** CNRS UMR8186, Department of Biology, Ecole Normale Supérieure, Paris, France

## Abstract

Carotenoids are produced by all photosynthetic organisms, where they play essential roles in light harvesting and photoprotection. The carotenoid biosynthetic pathway of diatoms is largely unstudied, but is of particular interest because these organisms have a very different evolutionary history with respect to the Plantae and are thought to be derived from an ancient secondary endosymbiosis between heterotrophic and autotrophic eukaryotes. Furthermore, diatoms have an additional xanthophyll-based cycle for dissipating excess light energy with respect to green algae and higher plants. To explore the origins and functions of the carotenoid pathway in diatoms we searched for genes encoding pathway components in the recently completed genome sequences of two marine diatoms. Consistent with the supplemental xanthophyll cycle in diatoms, we found more copies of the genes encoding violaxanthin de-epoxidase (VDE) and zeaxanthin epoxidase (ZEP) enzymes compared with other photosynthetic eukaryotes. However, the similarity of these enzymes with those of higher plants indicates that they had very probably diversified before the secondary endosymbiosis had occurred, implying that VDE and ZEP represent early eukaryotic innovations in the Plantae. Consequently, the diatom chromist lineage likely obtained all paralogues of *ZEP* and *VDE* genes during the process of secondary endosymbiosis by gene transfer from the nucleus of the algal endosymbiont to the host nucleus. Furthermore, the presence of a *ZEP* gene in *Tetrahymena thermophila* provides the first evidence for a secondary plastid gene encoded in a heterotrophic ciliate, providing support for the chromalveolate hypothesis. Protein domain structures and expression analyses in the pennate diatom *Phaeodactylum tricornutum* indicate diverse roles for the different *ZEP* and *VDE* isoforms and demonstrate that they are differentially regulated by light. These studies therefore reveal the ancient origins of several components of the carotenoid biosynthesis pathway in photosynthetic eukaryotes and provide information about how they have diversified and acquired new functions in the diatoms.

**Citation:** Coesel S, Oborník M, Varela J, Falciatore A, Bowler C (2008) Evolutionary Origins and Functions of the Carotenoid Biosynthetic Pathway in Marine Diatoms. PLoS ONE 3(8): e2896. doi:10.1371/journal.pone.0002896

**Editor:** Lennart Randau, Yale University, United States of America

**Received:** May 29, 2008; **Accepted:** July 7, 2008; **Published:** August 6, 2008

**Copyright:** © 2008 Coesel et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** Funding for this work was obtained from the EU-funded FP6 Diatomics project (LSHG-CT-2004-512035), the EU-FP6 Marine Genomics Network of Excellence (GOCE-CT-2004-505403), an ATIP "Blanche" grant from CNRS, and the Agence Nationale de la Recherche (France) to CB, the Fundação para a Ciência e a Tecnologia (Portugal) to SC (SFRH/BD/4839/2001), the Grant Agency of the Academy of Sciences of the Czech Republic to MO (IAA500220502), and the HFSP-CDA (0014/2006) to AF. Diatom genome sequencing at the Joint Genome Institute (Walnut Creek, CA, USA) was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396. P. tricornutum ESTs were generated at Genoscopy (Evry, Paris).

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: cbowler@biologie.ens.fr

‡ Current address: Institute for Systems Biology, Seattle, Washington, United States of America

## Introduction

The brown unicellular diatoms (Bacillariophyceae) constitute a successful group in the phytoplankton community estimated to be responsible for 30 to 40 % of marine primary production [1]. Their role in global carbon cycling is predicted to be comparable to that of all terrestrial rain forests combined [2,3], and they therefore represent an ecologically important group of bloom-forming phytoplankton. Diatoms can be classified into two major groups based on the symmetry of the frustule [4]. Centric diatoms are radially symmetrical and mostly planktonic, whereas pennate diatoms have a bilateral symmetry. Fossil records indicate that centric diatoms appeared at least 180 million years ago (Ma), whereas raphid pennate diatoms are thought to have evolved from centric diatoms prior to 90 Ma [5].

The importance of diatoms in marine ecosystems led to the sequencing of the nuclear, plastid and mitochondrial genomes of the centric diatom *Thalassiosira pseudonana*. The 34.5 Mb nuclear genome of this diatom was predicted to contain over 11,000 genes distributed on 24 chromosomes [6]. About half of the genes in this genome could not be assigned functions on the basis of similarity to genes in other organisms, and these genes may encode proteins involved in diatom-specific processes. The other half of the predicted genes have similar alignment scores to their closest homologs in either plant, red algal or animal genomes, which underscores the novel evolutionary history of diatoms (see below) [6–8]. More recently, the pennate diatom *Phaeodactylum tricornutum* was chosen as the second diatom for whole genome sequencing, partly because the physiology of this species has been studied for decades and because many molecular tools have been developed

for this alga [9,10]. Comparison between the two diatoms may also shed light on the evolution of diatoms. The nuclear *P. tricornutum* genome (27.4 Mb) is slightly smaller than the *T. pseudonana* genome but contains approximately the same number of genes and chromosomes (unpublished data). A recently assembled, and still expanding, diatom expressed sequence tag (EST) database [11,12] has been linked to the *T. pseudonana* and *P. tricornutum* genome databases, which aids the prediction of genes and also provides information on gene expression profiles.

Diatoms are believed to have obtained their plastid from a secondary endosymbiosis between a heterotrophic eukaryote and an ancient red alga. This event is postulated to have occurred at least 800 Ma [13–15] and may subsequently have given rise to the chromalveolates, which includes the group of algae collectively called Chromista. The chromist algae, Haptophyta (e.g., coccolithophorids), Cryptophyta (e.g., *Guillardia theta*), Heterokonts (diatoms and brown algae), and tertiary red-symbiotic dinoflagellates, all use chlorophyll *a*, chlorophyll *c*, and fucoxanthin as light-harvesting pigments [16]. Fucoxanthin, a carotenoid presumably derived from  $\beta$ -carotene (Fig. 1), absorbs light in the blue range of the light spectrum and is largely responsible for the characteristic brown color of chromist algae. The success of chromist algae may be explained by their ability to maintain photosynthetic activity in the blue-light-dominated oceanic environment [17–19]. In particular, diatoms have a huge capacity to dissipate excess absorbed light energy and their non-photochemical quenching (NPQ) levels can be as much as five times the levels registered for higher plants [20]. Moreover, diatoms are able to apply this photoprotective mechanism without significantly altering their light harvesting capacity [21], which allows them to maintain high growth rates over a wide range of light intensities [22].

The working model for energy-dependent NPQ ( $q_E$ ) in higher plants is based on the action of two oxygenated carotenoids, zeaxanthin and violaxanthin (Fig. 1). Essentially, the xanthophyll zeaxanthin binds protonated light-harvesting complex (LHC) proteins which may subsequently lead to a conformational change of the LHC and thereby allow the excess energy to be dissipated as heat [23–25]. The xanthophyll cycle allows plants to rapidly and tightly control their zeaxanthin content depending on the actual light conditions [26,27]. The cycle is performed by two enzymes: zeaxanthin epoxidase (ZEP) and violaxanthin de-epoxidase (VDE) [28,29]. Under dark/low light conditions, ZEP converts zeaxanthin, via antheraxanthin, to violaxanthin in two subsequent epoxidation reactions and when the photosynthetic apparatus receives too much light, VDE converts violaxanthin back to zeaxanthin in a reverse reaction (Fig. 1). In diatoms, the role of zeaxanthin has been supplemented by another xanthophyll molecule called diatoxanthin [Dtx; 30,31], and they appear to lack the PsbS protein, which has been implicated in the mechanism of NPQ in higher plants [23–25]. The diadinoxanthin cycle (Fig. 1) comprises the reversible conversion of diadinoxanthin (Ddx) and diatoxanthin (Dtx) [32,33]. Curiously, the ancestors of primary plastids (cyanobacteria) and secondary brown plastids, which are thought to originate from an engulfed rhodophyte, do not perform xanthophyll cycling, so the evolution of violaxanthin de-epoxidase and zeaxanthin epoxidase in the green and brown photosynthetic lineages is of particular interest.

In the current study the fully sequenced genomes of *T. pseudonana* and *P. tricornutum* were used to identify genes putatively involved in carotenoid biosynthesis and special attention was given to genes of the two diatom xanthophyll cycles. Candidate genes were subsequently compared with known sequences from other organisms in order to study the evolutionary history of carotenoid biosynthesis from the ancestral cyanobacterium to the modern

diatoms, and also from centric diatoms to raphid pennate diatoms. A number of environmental conditions, which may control the expression of diatom carotenoid biosynthesis-related genes, were assessed experimentally and *in silico* by means of quantitative real-time RT-PCR and by analysis of the *Phaeodactylum* Digital Gene Expression Database (<http://www.biologie.ens.fr/diatomics/EST3>).

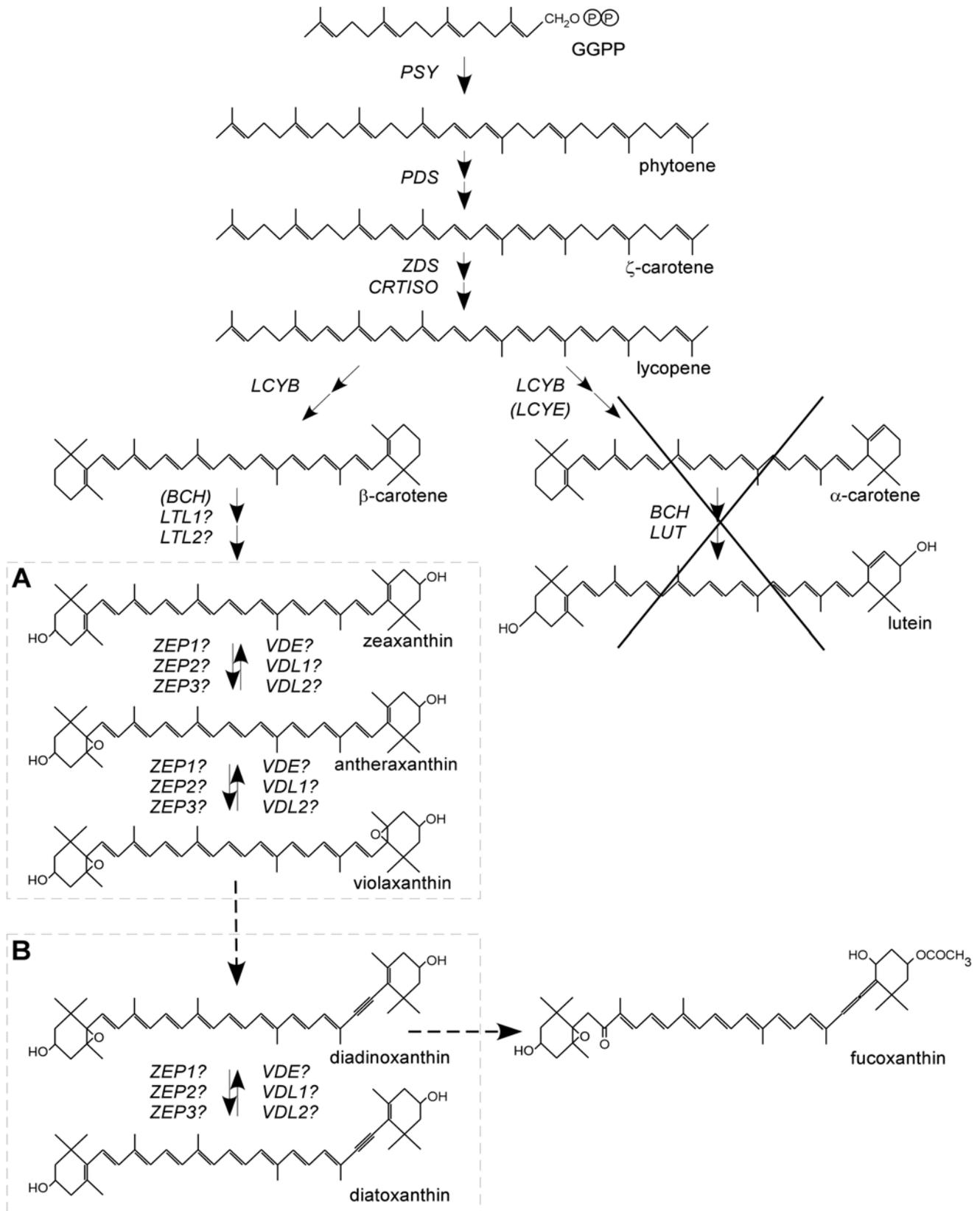
## Results and Discussion

The *P. tricornutum* and *T. pseudonana* nuclear genomes were examined for putative genes of the carotenoid biosynthetic pathway. The diatom EST database (<http://www.biologie.ens.fr/diatomics/EST3>), containing over 130,000 ESTs from *P. tricornutum*, was of great importance for accurate gene annotation. All selected gene models contain start and stop codons, unless stated otherwise, as well as putative upstream targeting sequences as determined by SignalP v3.0 and ChloroP v1.1 [34]. A summary of diatom genes putatively encoding carotenoid pathway components is shown in Table 1, and details concerning the early carotenoid biosynthetic pathway are given in the Supplementary Information. In short, our findings indicate that diatoms most likely have inherited the genes involved in the early reactions of carotenoid biosynthesis (up to  $\beta$ -carotene) from the algal endosymbiont. However, the gene encoding  $\beta$ -carotene hydroxylase (BCH), responsible for the hydroxylation of  $\beta$ -carotene into zeaxanthin (Fig. 1), is absent in the *P. tricornutum* genome and only a partial sequence is present in the *T. pseudonana* genome. This enzymatic reaction may therefore be catalyzed by other unrelated enzymes in diatoms, such as LUT-like P450 proteins (see Supplementary Information S1 and Figure S1). Due to their special interest in diatoms the enzymes involved in the xanthophyll cycle were studied in more detail.

### Domain structures of diatom violaxanthin de-epoxidases

Violaxanthin de-epoxidase (VDE)-encoding genes have been identified in a broad range of plants and they form a highly conserved family. VDE is characterized as a soluble protein located in the thylakoid lumen and the optimal VDE activity is found between pH 5.0 and 5.2 *in vivo* [26,35]. We have previously reported the presence of two VDE-encoding genes in the centric diatom *T. pseudonana* [8], one of which is similar to the VDE of higher plants, while the other (designated as violaxanthin de-epoxidase-like; *TpVDE*) is more distantly related. We had proposed that the first might be involved in the conventional xanthophyll cycle found also in green photosynthetic eukaryotes, while the latter may be more specialized in the chromist-specific diadinoxanthin cycle. Analysis performed in this work indicate that the *P. tricornutum* genome contains one VDE gene and two VDE-like genes, designated as *PtVDE* and *PtVDE1* and 2, respectively (Table 1).

A comparison between the domain structures of the plant and diatom VDE proteins shows that the diatom proteins are relatively similar to the plant counterparts and consist of a cysteine-rich N-terminal domain, a lipocalin domain, and a C-terminal glutamic acid-rich domain (Fig. 2). The cysteine residues in the first domain can form one or more disulfide bridges [28] and this domain seems essential for VDE function because deletion of any region in this domain leads to a total loss of VDE activity [36,37]. The second domain, which is thought to bind the xanthophyll molecule in the all-*trans* configuration, shows similarity to the eight-stranded  $\beta$ -barrel structure of the lipocalin protein family [38]. The small proteins ( $\pm 200$  aa) of this family are highly divergent in amino acid composition but are conserved in their tertiary structure, which allows them to bind small hydrophobic molecules [39].



**Figure 1. Hypothesized carotenoid biosynthetic pathway in diatoms.** The genes identified in this study, phytoene synthase (*PSY*), phytoene desaturase (*PDS*),  $\zeta$ -carotene desaturase (*ZDS*), lycopene  $\beta$ -cyclase (*LCYB*),  $\beta$ -carotene hydroxylase (*BCH*), lutein deficient-like (*LTL*), zeaxanthin epoxidase (*ZEP*), violaxanthin de-epoxidase (*VDE*) and violaxanthin de-epoxidase-like (*VDL*), are indicated. The *BCH*-encoding gene is absent in the *P. tricornutum* genome, *ZEP3* and *VDL2* are absent in the *T. pseudonana* genome. The two xanthophyll cycles are boxed, A) the violaxanthin cycle and B) the diadinoxanthin cycle.  $\alpha$ -carotene and lutein are not produced by diatoms. Dashed arrows indicate hypothetical conversion steps, according to Lohr and Wilhelm (1999, 2001).

doi:10.1371/journal.pone.0002896.g001

**Table 1.** Genes of the *P. tricornutum* and *T. pseudonana* carotenoid biosynthetic pathway.

protein	ID Phatr2	length (aa)	introns	SignalP (ChloroP***)	ID Thaps3	id/sim (%)	best hit NCBI	Accession	id/sim (%)
PSY1	56481	505	1	0.881 (0.526)	268908	61/71	<i>Pavlova lutheri</i>	ABA55571	45/61
PSY2	np	-	-	-	258309*	-	<i>Zea mays</i>	AAR31885	32/45
PDS1	45735	624	0	0.985 (0.531)	6524	81/89	<i>Ostreococcus lucimarinus</i>	ABO98307	74/83
PDS2	55102	589	1	0.999 (0.479)	1383	52/68	<i>Gentiana lutea</i>	BAB8246	57/74
ZDS	53974	591	0	1.000 (0.564)	37288	83/91	<i>Lyngbya sp. PCC 8106</i>	EAW35008	59/73
CRTISO	<i>up to 6 related genes+1 gene with closer homology to crtI</i>								
LCYB	56484	659	1	0.994 (0.561)	261407	57/70	<i>Capsicum annuum</i>	CAA60119	32/48
BCH	np	-	-	-	263437*	-	<i>Glycine max</i>	AAS88426	39/54
LUT-like1	50101	644	1	0.987 (0.564)	9541	75/85	<i>Ostreococcus tauri</i>	CAL50459	50/66
LUT-like2	26422	769	0	0.560 (0.474)	36235	69/81	<i>Skeletonema costatum</i>	AAL73435	70/82
ZEP1	45845	565	0	0.997 (0.574)	269147	66/75	<i>Chlamydomonas sp. W80</i>	AAO48941	33/47
ZEP2	56488	604	1	1.000 (0.566)	261390	68/75	<i>Vigna unguiculata</i>	BAB11934	31/45
ZEP3	56492	557	2	0.988 (0.560)	np	-	<i>Arabidopsis thaliana</i>	AAG38877	34/49
VDE	44635	437	0	0.902 (0.518)	7677	59/73	<i>Chrysanthemum × morifolium</i>	BAE79554	42/61
VDL1	46155	453	0	0.999 (0.464)	22076	74/82	<i>Spinacia oleracea</i>	CAB59211	20/34
VDL2	45846	555	1	0.997 (0.531)	np	-	<i>Medicago truncatula</i>	ABO83847	17/25
VDR	56450	587	0	0.912 (0.512)	270211	65/80**	<i>Oryza sativa</i>	EAZ13333	42/58**

The Phatr2 genome browser was used to identify *P. tricornutum* gene models encoding putative carotenoid biosynthetic enzymes (see text for abbreviations), and the following parameters are given: the protein identification number (ID) of the best gene model, the length of the immature (including putative signal peptide) protein, the number of introns, the probability of the presence of a signal peptide as determined by SignalP v3.0 and, in parentheses, the chloroplast targeting prediction score as determined with ChloroP v1.1. The Thaps3 genome browser was used to identify the *T. pseudonana* homologs and the protein IDs are shown together with the identity/similarity to the respective *P. tricornutum* gene. The identity/similarity was determined after aligning the mature protein sequences with ClustalW using the BLOSUM62 similarity matrix. BLASTP searches on the NCBI sequence browser were performed on 24-08-2007 and the NCBI accession number and identity/similarity with the respective *P. tricornutum* genes are given. When the Phatr2 gene model is absent, the Thaps3 model was used instead. When a gene model was not full length, the alignment was trimmed to the shortest sequence.

np) not present, \*) gene model not complete, \*\*) cleavage site of VDE was used, \*\*\*) Cut-off value of 0.500 used for ChloroP1.1.

doi:10.1371/journal.pone.0002896.t001

Both these domains are generally well conserved between diatoms and plants, indicating that these proteins show similar folding and can bind the same molecules. However, the C-terminal Glu-rich domain is considerably less conserved between diatoms and plants: whereas the C-terminal domain of plant VDEs contains an average of 47% charged residues of which about 25% are glutamic acid residues, the percentage of charged amino acids in the PtVDE and TpVDE domains is 29 and 37% (13 and 18% Glu), respectively. Partial protonation of the glutamic acid-rich domain is thought to increase the binding of VDE to the thylakoid membrane [28,40], and the divergence of this C-terminal domain may likely affect the pH-dependent binding of the diatom VDE to the thylakoid membrane [41,42].

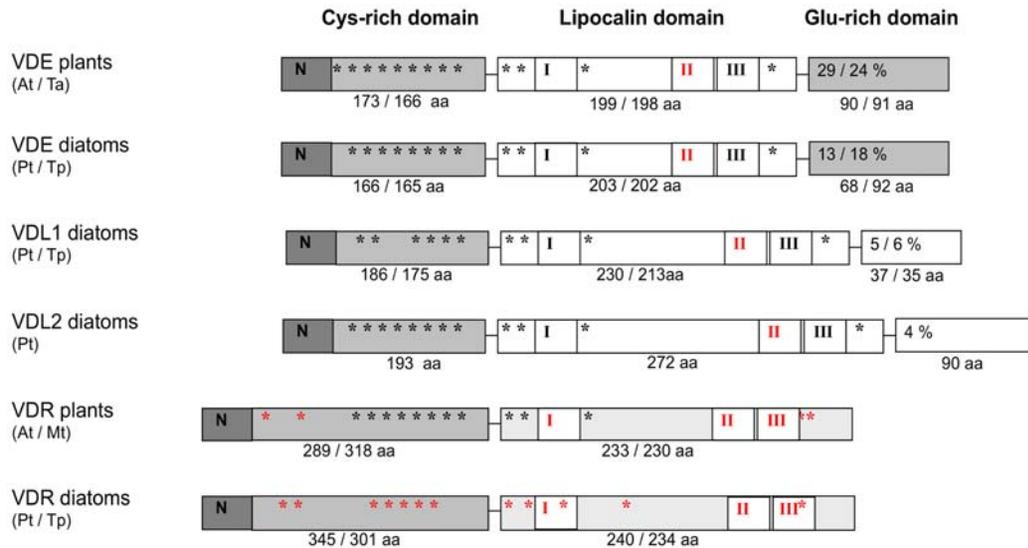
In the diatom VDL1 and VDL2 sequences, the Glu-rich domain has been completely replaced by an uncharged C-terminal domain. Therefore it is possible that the activation and/or localization of VDL proteins may be different than the VDE proteins. The lipocalin domain of the VDL proteins is considerably larger (Fig. 2), but they nevertheless contain the two typical lipocalin motifs (I and III), which are important for the correct folding of the  $\beta$ -barrel structure. Extensive searches in the available online genome and EST databases showed that genes containing this larger lipocalin domain are found only in chlorophyll *c*-containing chromist algae (see below). We therefore consider it possible that the VDL enzymes may be able to more efficiently bind, and eventually de-epoxidize, brown algal-specific molecules such as diadinoxanthin. However, it has to be noted that an *in vitro* enzyme assay with spinach VDE demonstrated that

the conventional VDE enzyme is also able to make the conversion of diadinoxanthin, albeit with a 30% reduced efficiency as compared to violaxanthin [43], so without experimental evidence we cannot rule out the possibility that the VDL proteins may have other unrelated or additional functions.

The genome of *Chlamydomonas reinhardtii* lacks a copy of the *VDE* gene (possibly due to incomplete sequencing), although careful scanning of the *C. reinhardtii* genome for VDE led to the discovery of a novel gene designated *VDE-related (VDR)*. VDR does not contain the typical VDE protein family signature (pfam 07137) which spans the lipocalin domain (Fig. 2) and does not contain a C-terminal domain, but nevertheless it still shares sequence similarity with VDE (*A. thaliana* VDE and VDR: 12% id, 27% sim). Following this first identification, *VDR* homologues were subsequently found in higher plants as well, and it appears that this gene of unknown function is ubiquitously present in all green photosynthetic eukaryotes. We found that the two diatom genomes also contain a copy of the *VDR* gene, and that the encoded amino acid sequences are relatively similar to the plant counterpart (Table 1), except for the position of the conserved cysteine residues (Fig. 2), which probably has implications for the folding of the protein.

Even though VDE, together with ZEP, were the first lipocalin-like proteins to be identified in plants [38,44], there is still considerable debate as to whether VDE can be truly ascribed to the lipocalin protein family because of the great divergence in size, function, and genomic organization between lipocalins and VDEs [45-47]. The most accepted evolutionary scenario for the

A



B

VDE N. tabacum CTLLKCRLEELAKIANPAANVAIQTNRRPDETEQIKGDLFENSVDV  
 VDE A. thaliana CAELLKCRLEELAKIANPAANVAIQTNRRPDETEQIKGDLFENSVDV  
 VDE P. tricornutum VSLFQKPLPLAKIANPKLANVINTTGRPDEIEQIEGNLFEVGVG  
 VDE T. pseudonana GLLVKRVPLAKITNPKLANVINSNGKEDETGQINGNVFEVGVG

VDL1 P. tricornutum VGSFINKGOOTKALFSNFRGIKGVSLGRKGG--ECSATRFAFEGSESN  
 VDL1 T. pseudonana VGSFINKGOOTKALFSNFRGIKGVSLGRKGG--ECSATRFAFEGSESDLD  
 VDL2 P. tricornutum QNLVSAISLOTKALODDESARKGLTAKLGG--DNAITGMARYGNAND

VDR A. thaliana LSMVKNGPQILNLLDNRKALQDNLQSPV--DQVSYRILASYEGPYFE  
 VDR C. reinhardtii IIMIKNGSKVVGVD--PTKALDNLNGTFN--DQVQYRILVSYBSPLE  
 VDR P. tricornutum VVSMVSKGCVVKGVD--DNRECELVITALDTR--DQVASYRTIVSYBSDLK  
 VDR T. pseudonana IFSMITKGCPIKTLND--ENSQCNALDKIDTR--DQVTSYRTVSYBSELLR

Contig A. castellanii AKLTIKPAQVMAIYED--EVLSAIAVQAIGPS--DVAATGTAQAGDPTFG

C

motif	I GxW (Y/F/W/H)	aa	II T (N/D) (Y/N) xxY	aa	III R/K
Lip Vc	VSDDELNNYL	60	SYVVFELDR	4	TEYLNLLSFP-TVE---RGIID--KF
Lip Dd	KKPYPEKYV	60	GYNVIAID	4	FDYMLILSRP-TIE---EKIKN--SY
Lip Gv	VAEVDNRYD	62	DYVLDLVEDPNSAG-DVAFAAI	4	RDFIPIISRP-AET--YQDVLAYQGI
Lip Py	VPALVAAAT	68	PYVWINLAG	10	SQFLPILLSFP-KER--YGVFEAVVYV
Lip Pt	AASCETAKFY	66	PYLPLEVD	4	RNYCHIMSRP-VMAEATYAKLTK--QL
TILL Ta	VLGLVARYM	68	DYRWLYVD	4	RKSLNHLCKT-HIE---EEVYN--QL
CHL Ta	AKNEDPVRYM	83	PYDVLATDY	3	TSFIQIYSFP-NPG---PEFIE--KY
PRBR Hs	KENEDKARFS	65	DHWVDVTDY	11	DSYSEVFSDFNGLP---PEAQK--I
ApoD Hs	QENEDVKNKYL	59	PYMLATDY	9	VDFAWILANP-NLP---PETVD--SL
CC-A2 Hg	DNEDLRRYA	60	PYEVLETDY	8	SEFAVFSFP-QTS---GPAVE--KT
VDE Nt	VQRDMKDFE	62	DWVLLSKVE-NSP--EDMIFVYV	9	YGCSSVLYTSA-VLP---ESIIP--EL
VDE At	VQNSNISDFN	63	DWVLLSKIE-NKP--EDMIFVYV	9	YGCSSVLYTSS-VLP---NSIIP--EL
VDE Pt	VPKDTPKFFD	63	DWVVDIYEDGNKDGVPPEAFVYV	9	YGCSSVLYTAA-QLP---ESLLP--RL
VDE Tp	VQSDDFKLWN	63	DWVVDVYAADNKGVPPEAFVYV	9	YGCSSVLYTDS-KLP---ESLLP--RL
VDL1 Pt	LRSDFPQSLV	94	NWVILG-ESD-GKGSVPPFKLVAN	8	YECAEFYAKEA-TVE---EAAKP--AI
VDL1 Lp	VKRDPSLTV	76	NWVILG-ESN-GDNDIPPFKLVAN	8	YEEAEFYAKES-VLP---KEAVG--AV
VDL2 Pt	VTAEDPKSLQ	133	NWVILG-ENDPGQ--PEFKFVYV	8	YECAEFYVSK-EVA---PESMA--KV
VDR At	CHDTAEDLFVGLG	64	PWTVVDVSD--LSWGLFHY	10	YTCFVLTVPDG-SYFA---EKDKE--RL
VDR Cr	THEAEDLFVGLG	64	YRPTVDCCDEN--LDWCLFHY	10	YSCAVLGTDPDG-GMFG---PQHTQ--RL
VDR Pt	TEDIARSLVGHNLN-DEAEPSS	64	PWTVGVADD--LSWVIFHY	10	YLCGLLCTADG-SLPD---ESQRP--EI
VDR Tp	TTDVARGIMIGHLEGAGGSLEGN	64	PWTVGVADD--LSWVVFHY	10	YLCGLLCTPTG-ELFP---EEDLG--HI

**Figure 2. Domain structure of violaxanthin de-epoxidases and related proteins.** A) Schematic representation of VDE, VDL and VDR proteins (not to scale). Three different domains are shown; the cysteine-rich domains include the N-terminal targeting sequence. Black and red asterisks indicate the positions of conserved and alternative cysteine residues, respectively. The central lipocalin domain contains the lipocalin binding fold. Conserved and divergent lipocalin motifs (roman numbers) are given in black and red, respectively. The size of the lipocalin motif was determined by sequence alignment of VDE sequences and a representative group of lipocalin proteins. The C-terminal glutamic acid-rich domain indicates the percentage of Glu residues in this domain. B) Alignment of the N-terminal cysteine-rich domains of several plant and diatom VDEs. Also included is a sequence derived from the amoeba *Acanthamoeba castellanii*. C) Alignment of the lipocalin motifs I, II and III of several different lipocalin VDE, VDL and VDR proteins. The distance (in amino acids) between the three lipocalin motifs is also indicated. The lipocalin motif consensus sequences, as derived from kernel lipocalins (Flower, 1996), are indicated above the alignment and conserved motifs within the alignment are indicated in red. The abbreviations used are: Lip, lipocalin; TIL, temperature induced lipocalin; CHL, chloroplast lipocalin; PRBR, plasma retinol-binding protein precursor; CC, crustacyanin; At, *Arabidopsis thaliana*; Cr, *Chlamydomonas reinhardtii*; Dd, *Dictyostelium discoideum*; Gv, *Gloeobacter violaceus*; Hg, *Homarus gammarus*; Hs, *Homo sapiens*; Mt, *Medicago truncatula*; Nt, *Nicotiana tabacum*; Pt, *Phaeodactylum tricornutum*; Py, *Porphyra yezoensis*; Ta, *Triticum aestivum*; Tp, *Thalassiosira pseudonana*; Vc, *Vibrio cholerae*. doi:10.1371/journal.pone.0002896.g002

appearance of plant VDEs involves a gene fusion event between a lipocalin and another gene, together giving rise to the larger VDE protein with a newly acquired function [45,46]. To find the ancestral lipocalin sequence, which may have given rise to the modern VDEs, Charron *et al.* [46] searched 14 complete and 2 partial cyanobacterial genomes, and found only 1 cyanobacterial species that contained a lipocalin gene. This particular species, *Gloeobacter violaceus* PCC7421, belongs to the most ancient members of the cyanobacterial lineage, which do not possess thylakoid membranes but instead use the plasma membrane to attach the phycobilisomes [48]. Charron *et al.* [46] speculated that the *G. violaceus* lipocalin might be the ancestor of recently identified chloroplastic lipocalins (CHL), VDEs and ZEPs of higher plants. However, there is currently no solid phylogenetic evidence to support the role of the cyanobacterial lipocalin in the occurrence of the xanthophyll cycle genes.

In order to find evidence for the proposed gene-fusion event between a lipocalin and a second gene, we used the well-conserved N-terminal cysteine-rich domain of VDE to search the microbial and protozoan genomes available at NCBI using BLAST. While the 915 microbial genomes did not yield any positive hits, a relatively conserved Cys-rich domain (>40% similarity) was present in the protozoan genomes of *Phytophthora sojae* and *P. ramorum*, *Entamoeba histolytica*, *Toxoplasma gondii*, and *Trypanosoma cruzi*. A similar search against the Taxonomically Broad EST Database (TBestDB) was performed, and an especially well conserved Cys-rich domain in an EST contig encoding a multi-domain protein of the protozoa *Acanthamoeba castellanii* was found. To illustrate the homology, this sequence was placed within the Cys-rich domain alignment of Figure 2B. Because protists, but not cyanobacteria, contain well-conserved expressed Cys-rich-encoding proteins in their genomes, the N-terminal domain of the VDE-encoding gene may have originated from the protozoan host. Furthermore, because only chlorophyll *b* and *c*-containing photosynthetic eukaryotes possess this Cys-rich domain in conjunction with a lipocalin domain, it can be hypothesized that VDE is the result of a gene-fusion event which probably occurred relatively early after the primary endosymbiosis of the cyanobacterium and the host cell.

### Phylogeny of violaxanthin de-epoxidases

The identification of plant-like VDEs in diatoms and other unicellular brown algae is most easily explained by the presence of VDE in the common ancestor of the green and red lineages. However, the main light-harvesting complexes of rhodophytes, the phycobilisomes, do not use a zeaxanthin-dependent NPQ mechanism to dissipate excess-absorbed light [49,50], and physiological experiments on a wide range of different red algal species indicate that red algae do not convert violaxanthin to zeaxanthin upon illumination [50], although some red algal species are able to produce antheraxanthin and violaxanthin [50,51]. Furthermore, no VDE sequence evidence from a red alga has been forthcoming from the *C. merolae* genome, nor from the EST sequences of the red macroalga *Porphyra yezoensis* (data not shown).

A phylogenetic tree of the diatom and plant VDEs and related genes was generated and rooted by related lipocalin family proteins from both prokaryotes and eukaryotes. As can be seen in the tree (Fig. 3), there is a considerable difference between the ancient lipocalins and VDEs, such that the ancestor of VDE proteins within these lipocalin proteins cannot be specified. The VDE proteins constituted three distinct clusters: VDE, VDL (VDE-like) and VDR (VDE-related) proteins. Two of these clusters (VDE and VDR) are composed of plant and diatom sequences with diatom sequences appearing on the root of each cluster

(Fig. 3). Plant sequences are absent from the VDL cluster, which contains, in addition to diatom proteins, sequences from other chromalveolates, such as heterokonts and dinoflagellates (Fig. 3). A single metazoan VDL sequence from the eastern oyster *Crassostrea virginica*, which has been obtained from ESTs at NCBI, appeared within the VDL proteins; however, the position of this oyster sequence is not supported by bootstraps. Its position is also questioned by the fact that the NJ tree showed a different topology than the ML tree, placing *C. virginica* at the root of all VDE and related proteins (data not shown). The sequence from *C. virginica* is the only VDE sequence known from non-photosynthetic eukaryotes and, moreover, when Blast searched in NCBI it yields hits only from photosynthetic eukaryotes. We therefore have serious reservations about the identity of this sequence and speculate that it may be derived from contamination of *C. virginica* cDNA with material from eukaryotic algae.

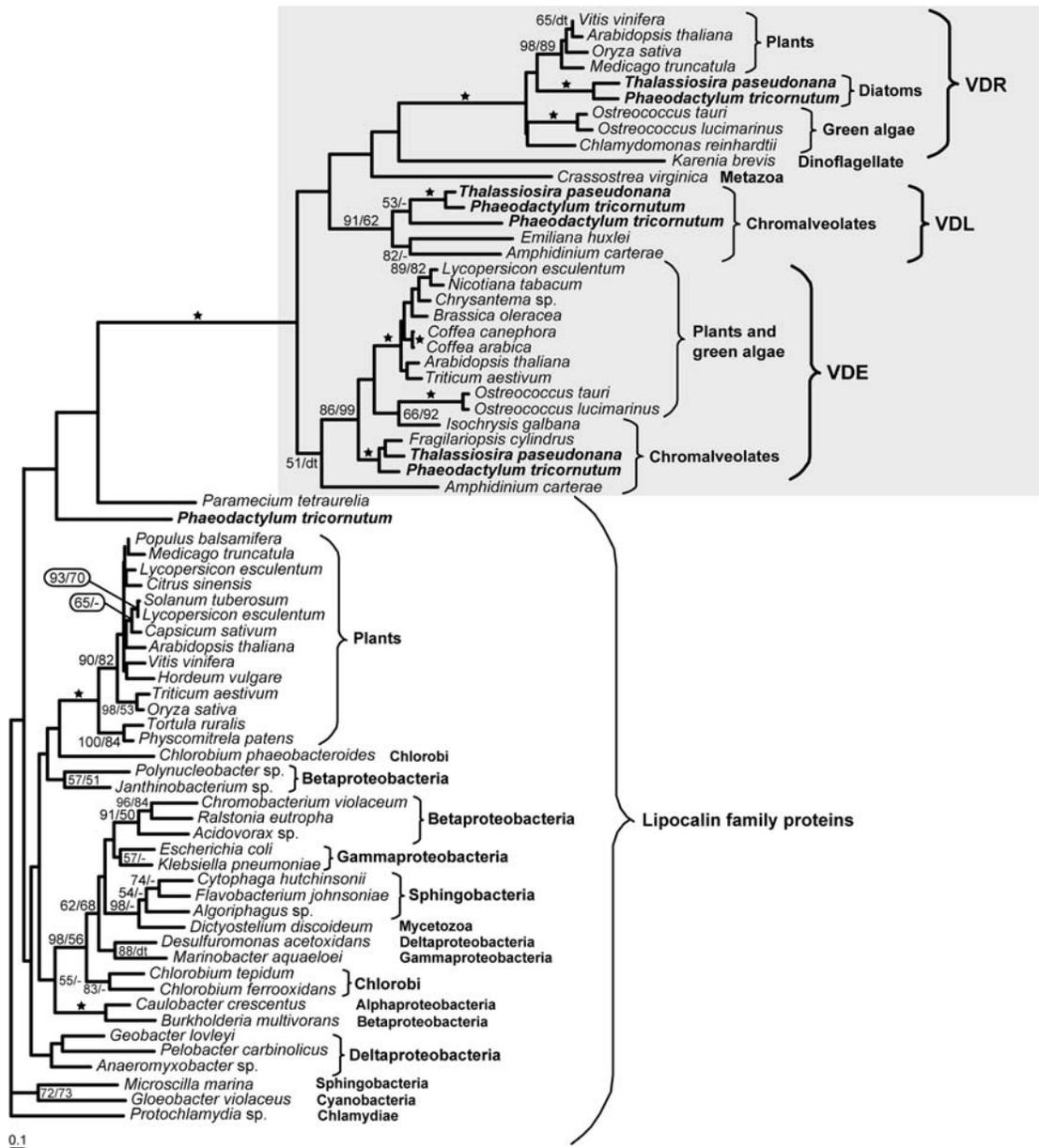
Within the VDE+VDL+VDR cluster, VDE sequences seem to constitute an ancestral form, which is basal to the advanced sisters VDL and VDR (Fig. 3). It is obvious that VDL and VDR proteins have arisen by two duplication events. The second duplication could be associated with the secondary endosymbiosis, since it has happened only in chromalveolates. Alternatively, this particular paralogue was previously present only in red algae or has been lost from green plants.

Based on the tree topology we suggest that VDE, VDL and VDR genes are out-paralogues of the original VDE gene, with the duplication event preceding the speciation as well as the secondary endosymbiosis that led to the appearance of complex plastids. The fact that no homolog of VDE proteins has been found in prokaryotes suggests that they represent an ancient eukaryotic innovation. This is also supported by the fact that in cyanobacteria, the putative prokaryotic donor of genes encoding proteins of plastid functions, no VDE proteins have ever been identified and the lipocalin homolog is found only in a single cyanobacterial species *Gloeobacter violaceus*. We propose that VDE proteins have arisen in diatoms by the endosymbiotic gene transfer from the nucleus of the algal endosymbiont. In such a scenario, VDE proteins have been lost from the current rhodophytes. Plants lack VDL proteins, possibly because they lost it during evolution, similarly to rhodophytes, or because VDLs are for some reason only found in photoautotrophs derived from the secondary endosymbiosis. Based on the phylogenetic position of plant, diatom and other chromalveolate VDEs, we suggest that VDE proteins of green and brown algae share a common origin. Alternatively, the VDE gene may have been introduced in diatoms by a lateral gene transfer from a green alga. However, this scenario seems less likely to explain the widespread appearance of VDE throughout the brown clade, as seen by the presence of a VDE sequence in haptophytes.

### Domain structures of diatom zeaxanthin epoxidases

Zeaxanthin epoxidase is localized in the stromal side of the thylakoids, where it catalyzes the conversion of zeaxanthin to antheraxanthin and violaxanthin [52]. The gene encoding ZEP was first identified in a transposon-tagged *Nicotiana plumbaginifolia* abscisic acid (ABA)-deficient mutant, called *aba2* [29], and was found to encode a chloroplast-imported flavin-containing monooxygenase (FMO), containing an ADP-binding fold and an FAD-binding domain similar to prokaryotic aromatic-substrate monooxygenases. As predicted by Bugos *et al.* [38], VDE and ZEP share a similar basic tertiary structure based on a lipocalin domain.

We have previously reported the presence of two ZEP encoding genes in the centric diatom *T. pseudonana* [8]. As with VDE, this is one copy more than is generally found in higher plants and may



**Figure 3. Maximum likelihood phylogenetic tree of violaxanthin de-epoxidases and related proteins.** A maximum likelihood phylogenetic tree ( $\log_{10}k = -13981.66253$ ) as inferred from amino acid sequences (141 amino acid characters) of violaxanthin de-epoxidases and related proteins was computed using WAG model for amino acid substitution (selected by PROTTEST) with discrete gamma distribution in four categories. All parameters ( $\gamma$  shape = 2.158; proportion of invariants = 0.000) were estimated from the dataset. Numbers above branches indicate ML/NJ bootstrap supports. ML bootstraps were computed using the above mentioned model in 300 replicates. An NJ tree was inferred using AsaturA program with cutoff value 0.906 and 1000 replicates. Black stars indicate both bootstraps over 90%. Nodes that display different NJ topology than the one obtained by ML are indicated by "dt".  
doi:10.1371/journal.pone.0002896.g003

again be a reflection of the two separate xanthophyll cycles found in heterokonts. As was the case for VDE, we found that the *P. tricorutum* genome contained a third copy of ZEP that is absent in the *T. pseudonana* genome (Table 1). As compared to the plant ZEPs, the amino acid region covering lipocalin motif I is considerably larger in the diatom ZEP1 and 2 proteins, but not in PtZEP3 (Fig. 4) and the motif I consensus sequence is not conserved in any of the diatom ZEPs. All diatom ZEPs lack the C-terminal forkhead-associated (FHA) domain that is normally found in plant ZEPs. In the diatom ZEP1 proteins this domain has been replaced with a conserved 35 amino acid sequence, whereas

considerably larger sequences are present in the Pt and TpZEP2 and PtZEP3 proteins. Interestingly, a transmembrane region is predicted in the C-terminal domain of PtZEP3, which may have an effect on the localization and/or regulation of the diatom ZEP proteins. In this context it is interesting to note that the epoxidation kinetics of diatoxanthin in diatoms under low light conditions is generally faster than the epoxidation of zeaxanthin in higher plants and green algae [53], but the presence of a proton gradient almost completely inhibits this reaction in diatoms [53,54]. This is in stark contrast with the epoxidation reaction of higher plants, which occurs in the dark as well as in the light, and

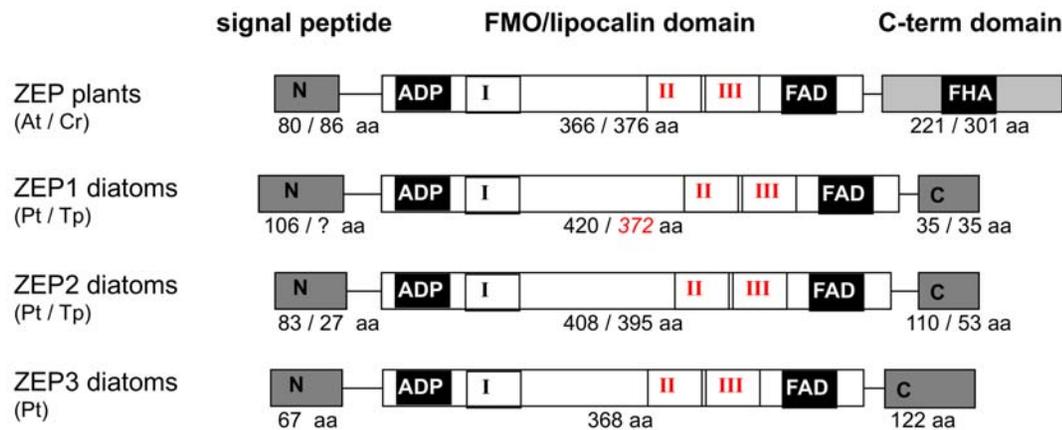
the extent of zeaxanthin accumulation in plants depends largely on the activity of VDE. The mechanism behind the peculiar light-dependent ZEP activation/deactivation in diatoms is not yet understood, but it is possible that the divergent C-terminal domains of ZEP1, ZEP2 or ZEP3 may play a role.

Whereas zeaxanthin epoxidation in higher plants is not only important for xanthophyll cycling, but also for ABA synthesis, diatoms do not appear to synthesize this phytohormone. Diatoms however are thought to produce their major light harvesting carotenoid, fucoxanthin, from violaxanthin [Fig. 1; 55,56], so the epoxidation reaction of zeaxanthin to synthesize fucoxanthin may require a different regulation than xanthophyll cycle-related epoxidation. The presence of two ZEP-encoding genes in *T. pseudonana* could also be a reflection of these two different processes, as was suggested by Wilhelm *et al.* [57]. The finding

that *P. tricornutum* possesses a third copy of both VDE and ZEP is also very striking. Results presented by Goss *et al.* [53] indicated that the NPQ capacity of *P. tricornutum* is higher than *T. pseudonana*, whereas the diatoxanthin epoxidation rate under low light is considerably slower in *P. tricornutum*.

As seen in Figure 4, zeaxanthin epoxidases are FAD-dependent mono-oxygenases (FMO) containing a putative lipocalin fold. Even though it is possible to recognize at least two of the three lipocalin motifs within the FMO domain, the overall similarity between lipocalin proteins and the putative lipocalin fold of ZEP is extremely weak (<10% similarity). Therefore, a possible gene fusion event between a lipocalin and another gene, as was proposed above for the *VDE* genes, seems less likely to explain the evolution of *ZEP* genes. More probably, the lipocalin fold of ZEPs may have evolved in a process driven by steric constraints for

**A**



**B**

motif	I GxW (Y/F/W/H)	aa	II T (N/D) (Y/N) xxY	aa	III R/K
Lip Vc	PVSDPELNNV	60	SNVVFELRR	4	TEYLGLLSDFTV
Lip Dd	AVKFDYPERV	60	GNVVIADL	4	FDYMLLSSEPTI
Lip Gv	TVAEVDNRYD	62	DMWIDLVDPNNSAGDFAFAAI	4	RFIFIFIISRELET
Lip Py	TVPALDVAAT	68	PYVWINLAGSKE	10	FSQGLELLSDFKL
Lip Pt	TAASCETAKFV	66	PVLLILEVD	4	RKSLGLLCKTHI
TiL1 Ta	VVLGLDVARVY	68	DMVVLVVD	4	RKSLGLLCKTHI
CHL Ta	TAKNDFPVRY	63	PMDVLATY	3	TSFIQIYSDFN
PRBR Hs	VKENFDKARFS	65	DHWIVDTY	11	DSYSFVFSDFNC
ApoD Hs	VQENFDVNRLL	59	PWVILATY	9	VDFATLAINFN
CC-A2 Hg	NQDNEDLRYA	60	PVEVIETV	8	SEFAFVFSDFDT
ZEP Np	NAGCITGDRINGLVDGVS	91	DFVPADLET	12	GKMOWYAFHHEPAC
ZEP Vu	RVGCITGDRINGLVDGVS	91	DFVPADLET	12	GKMOWYAFHHEPAC
ZEP At	EAGCITGDRINGLVDGVS	91	DFIPADIES	12	GKMOWYAFHHEPAC
ZEP Cr	REGCITGDRINGLVDGVT	101	DFTPADHDI	12	GKMOWYGFHHEPAC
ZEP CW80	DEGCITGDRINGLVDGVT	95	DFTPADHDI	12	GKMOWYGFHHEPAC
ZEP1 Pt	KAGTVADRVSGLKIYKKGVFLGLGK	129	EHRASNIEE	12	ERQOWEALIEPAC
ZEP1 Tp	TAGCCADRVSGLKIYKKGKLAG	94	PHRASNIEE	12	DRQOWEALIEPAC
ZEP2 Pt	DKFTFCTRTCCIKRDRADGSRFRMTEDRLDYLWNPADAPADNEVKEPLRQCAD	109	ILKTPDYA	12	GRVQWYAFALPEG
ZEP2 Tp	DKFTFCTRTCCIKRDRADGSRFRMTEDRLDYLWNPADAPADNEVKEPLRQCAD	108	VLKTPDYA	12	GRVQWYAFALPEG
ZEP3 Pt	ERFTFCTGKENGKRCIRT	98	ANDSPENQ	12	GKMOWYAFALPEG

**Figure 4. Domain structure of zeaxanthin epoxidases.** A) Schematic representation of ZEP proteins (not to scale). Three different domains are discriminated; the N-terminal targeting sequence (note that the *TpZEP1* gene model is not complete), the central flavin-containing monooxygenase (FMO) domain containing ADP and FAD binding sites, and the C-terminal domain containing the FHA motif. Conserved and divergent lipocalin motifs (roman numbers) are given in black and red, respectively. The size of the lipocalin motif was determined by sequence alignment of ZEPs and a representative group of FMO proteins. B) Alignment of the lipocalin motifs I, II and III of several different lipocalins and ZEP protein sequences. The lipocalin motif consensus sequence, as derived from kernel lipocalins, is indicated above the alignment and conserved motifs within the alignment are indicated in red. The asterisk indicates the Gly304 identified by Baroli *et al.* (2003). The distance (in amino acids) between the lipocalin motifs are also indicated. The abbreviations used are: Lip, lipocalin; TiL, temperature induced lipocalin; CHL, chloroplast lipocalin; PRBR, plasma retinol-binding protein precursor; CC, crustacyanin; At, *Arabidopsis thaliana*; Cr, *Chlamydomonas reinhardtii*; CW80, *Chlamydomonas sp. W80*; Dd, *Dictyostelium discoideum*; Gv, *Gloeobacter violaceus*; Hg, *Homarus gammarus*; Hs, *Homo sapiens*; Np, *Nicotiana plumbaginifolia*; Pt, *Phaeodactylum tricornutum*; Py, *Porphyra yezoensis*; Ta, *Triticum aestivum*; Tp, *Thalassiosira pseudonana*; Vc, *Vibrio cholerae*; Vu, *Vigna unguiculata*. doi:10.1371/journal.pone.0002896.g004



different conditions (<http://www.biologie.ens.fr/diatomics/EST3>). We have mined this extensive database for EST-derived contigs encoding genes of the carotenoid biosynthetic pathway, and have determined the frequency and distribution of each contig throughout the different EST libraries (Table 2). Although this type of expression analysis is not quantitative, it nevertheless provides useful information on general gene expression trends. The genes most represented in the EST database, *ZEP2*, *ZEP3* and *VDL1*, are all putatively involved in xanthophyll biosynthesis. Of these genes, *VDL1* appears to be the most constitutively transcribed.

The blue light (BL) library was found to be the most enriched in carotenogenesis-related ESTs. This library was generated from dark-adapted cells that were treated with 1 h of blue light with an intensity of  $25 \mu\text{mol m}^{-2} \text{s}^{-1}$  prior to RNA extraction, and it is enriched in ESTs related to photosynthesis and carbon fixation. The libraries from ammonium adapted (AA), urea adapted (UA) and iron limited (FL) cells are also rich in carotenogenesis-related ESTs compared to the other libraries. The cells of these libraries probably suffered a degree of photosynthetic inhibition due to nutrient shortage [58–60]. The ‘carotenogenesis-rich’ libraries all have a relatively high level of LTL1 and LTL2-encoding ESTs. These ESTs may encode Lut-like proteins involved in the hydroxylation of  $\beta$ -carotene (Supplementary Information S1). There are also some substantial differences in the EST distribution within the ‘carotenogenesis-rich’ libraries. For instance, the Fe-limited library does not contain many xanthophyll cycle-related ESTs but instead contains a high number of *PSY*-encoding sequences. The UA and AA libraries are particularly rich in *ZEP2*-encoding ESTs but do not contain any transcripts of the other two *ZEPs*, and the opposite is true for the BL library, which contains a high number of *ZEP1* and *ZEP3* but relatively few *ZEP2* transcripts (Table 2). The BL library is also relatively rich in *VDE* ESTs. Interestingly, cells grown in the AA and UA conditions contain significantly more diadinoxanthin than cells grown under standard nitrogen conditions (Andrew E. Allen, personal communication), whereas blue light treated cells are relatively enriched in violaxanthin (unpublished data).

The finding that the *P. tricornutum* blue light library was enriched in carotenoid biosynthesis-related ESTs as compared to all other libraries suggested that light may play an important role in controlling diatom carotenoid biosynthesis. Whereas the regulatory role of light in the nuclear gene expression of plastid-targeted proteins has been studied for decades in green algae and higher plants, study of these processes in diatoms is still in its infancy. To further explore the regulatory role of light in diatoms at the transcriptional level, we studied the gene expression profiles of *PSY*, *PDS1* and the different xanthophyll cycle genes in response to different light signals. To compare, we also studied the transcript levels of two diatom chlorophyll/carotenoid-binding LHC proteins, fucoxanthin/chlorophyll *a/c*-binding protein B (FCPB) and a protein denoted as ELIP-like, which shares structural similarity to early light-inducible proteins (ELIPs). Like the carotenoid biosynthesis-related genes, the transcripts of these nuclear genes are targeted to the plastid, and in higher plants and green algae the transcription of these genes are known to be regulated at the transcriptional level by light [61,62]. Likewise, the transcription of FCPs are also known to be light-regulated in diatoms [9,63].

The steady state transcript levels of 48 hour dark-adapted *P. tricornutum* cells treated with either continuous white ( $175 \mu\text{mol m}^{-2} \text{s}^{-1}$ ), blue ( $25 \mu\text{mol m}^{-2} \text{s}^{-1}$ ) or red ( $25 \mu\text{mol m}^{-2} \text{s}^{-1}$ ) light were determined for 12 subsequent hours by quantitative real-time PCR (qRT-PCR) using *histone H4* (*H4*) as a reference gene [9]. The transcript levels of both *PSY* and *PDS1* increased immediately

upon light exposure (Fig. 6), and the highest steady state transcript levels were measured after 3 to 5 h light. Transcript levels decreased again after longer exposure. The transcript levels of *PDS1* and, to a lesser degree, of *PSY* are similar at  $25 \mu\text{mol m}^{-2} \text{s}^{-1}$  blue light and  $175 \mu\text{mol m}^{-2} \text{s}^{-1}$  white light, indicating that the spectral quality of light plays a major role in the regulation of expression of these genes. This observation is further supported by the fact that red light, with an equal fluence rate as blue light, triggered a much weaker response. The induction of the two genes encoding LHC proteins, *FCPB* and *ELIP-like*, began slightly later than the carotenogenesis-related genes (Fig. 6). However, the amplitude of induction was more than 100-fold higher than for *PSY* and *PDS1*. The spectral quality of light was also of major influence for these genes: the transcript levels of both genes in blue and white light were similar, but the kinetics of increase were faster in blue light than in white light. By contrast, the amplitude and kinetics of transcription in response to red light were much lower.

We also determined the steady state transcript levels of the *P. tricornutum* *ZEPs* and *VDEs* in 48 h dark-adapted cells subsequently shifted to either white, blue and red light as described above (Fig. 7). We found a steady and strong increase of *ZEP1* transcript levels after 5 hours of white and blue light, and after this period the levels decreased again. The effect of white light was stronger than blue light, and red light did not result in a significant induction (Fig. 7A). The increase in *ZEP2* levels is approximately 10-fold lower than *ZEP1*, but the kinetics of *ZEP2* accumulation is faster and maximal levels were reached within 5 h of illumination (Fig. 7B). Blue light appears to have a stronger effect on *ZEP2* transcription than white light, even though the white light fluence rate was 7 times higher. Contrary to *ZEP1*, red light also has a slight effect on *ZEP2* transcript levels. *ZEP3* mRNA accumulation showed different kinetics, in which a minor induction after 1 h white or blue light was followed by a 2 h lag-phase (Fig. 7C). Maximal *ZEP3* transcript levels were found after 5 h of blue light and after 8 h of white light. *VDE* transcript levels very rapidly increased after stimulation with blue and white light and close to maximal levels were reached within 1h (Fig. 7D). The overall kinetics of steady-state *VDE* mRNA levels was much like *ZEP3* (Fig. 7C,D). These two genes are located next to each other on chromosome 8 (Fig. 7G) and are likely to form a co-regulated gene-cluster. Another such gene cluster is found on chromosome 4 composed of *ZEP1* and *VDL2*, and also in this case the transcript levels of the two clustered genes were comparable (Fig. 7A,F,G).

The data presented shows that mRNA levels for the first two genes committed to carotenoid biosynthesis, as well as the xanthophyll cycle-related genes, rapidly increase upon light exposure and that light of different spectral quality has differential effects. In higher plants and *C. reinhardtii*, *PSY* and *PDS* transcription has been demonstrated to be regulated through photoreceptors [64–66], and the genomes of *T. pseudonana* and *P. tricornutum* contain several putative photoreceptor-encoding genes, including a phytochrome, an animal-type cryptochrome/photolyase and several cryptochrome-DASH homologs [6,8,67], which may be important for diatom light-dependent gene-regulation. This is supported, at least for the *PSY* gene, by the finding that a short pulse (5 min) of blue light was sufficient to effectively induce its expression (Fig. S2).

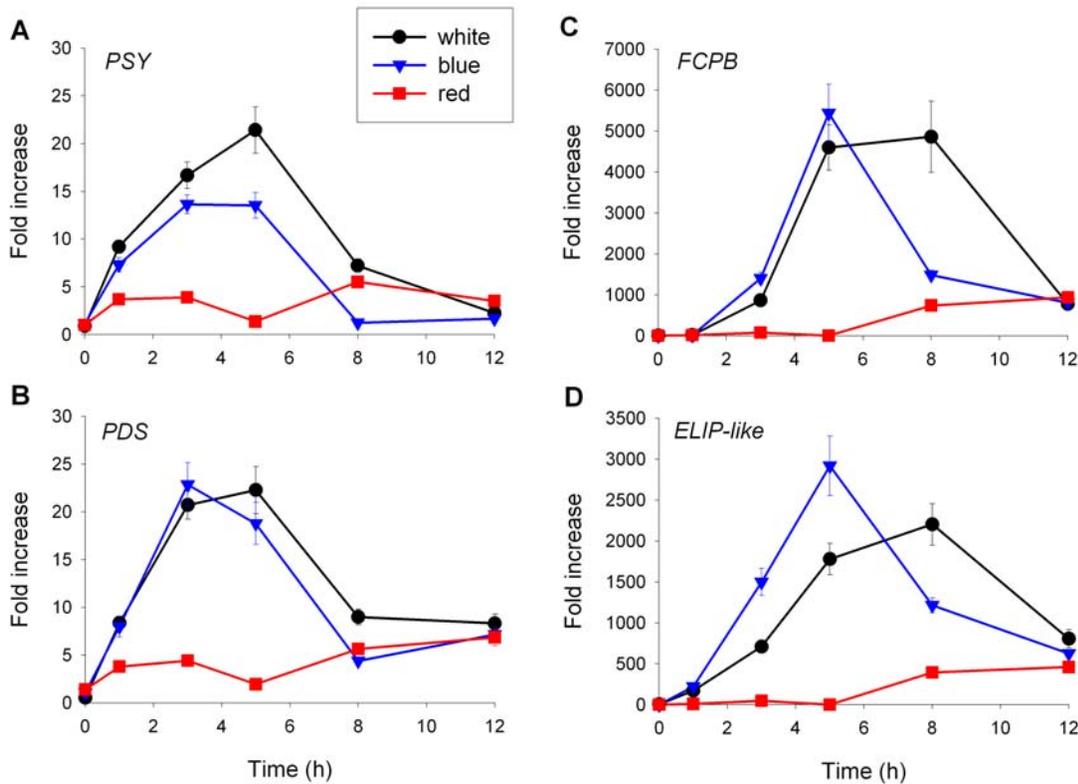
## Conclusions

In conclusion, we propose that both genes encoding enzymes of the xanthophyll cycle (*VDE* and *ZEP*) are eukaryotic innovations even though they are putatively located in the primary plastids of plants and the secondary plastids of diatoms. Genes for *VDE* and *VDR* have arisen through gene duplication events preceding

**Table 2.** EST distributions of carotenogenesis-related genes in the *Phaeodactylum* Digital Gene Expression Database.

gene	contig	OS	OM	TM	TA	SM	SP	NR	NS	AA	UA	FS	BL	LD	HD	C1	C4	total (%)	
PSY	G41878	-	-	-	-	-	-	-	-	-	-	0.73	0.08	0.11	-	0.10	0.10	1.11	
	G34382																		
PDS1	G35509	0.58	0.22	-	0.21	-	-	-	-	0.11	0.58	-	0.08	-	-	-	-	1.78	
PDS2	G55102	-	0.22	-	-	0.13	-	-	0.11	-	-	-	0.25	-	0.28	-	-	0.99	
ZDS	G9040	0.08	-	0.32	0.21	-	0.13	-	-	0.11	0.23	0.24	0.50	-	-	-	0.10	1.92	
LCYB	G8835	-	0.22	-	0.21	-	0.13	0.28	-	0.11	-	-	-	0.22	-	-	0.19	1.36	
	C193																		
LTL1	G16586	0.08	-	-	-	0.29	-	-	0.22	0.22	0.35	0.36	0.42	-	-	0.10	0.10	2.13	
LTL2	G26422	-	-	-	-	-	-	-	0.11	0.55	0.23	0.36	0.33	0.11	-	0.29	0.19	2.19	
ZEP1	G45845	0.08	-	-	-	-	-	-	0.22	-	-	0.24	0.58	-	0.28	0.10	-	1.50	
ZEP2	C962	0.25	0.44	-	-	0.14	-	-	0.11	1.11	0.94	0.12	0.33	0.22	-	0.10	0.10	<b>3.85</b>	
	S1206																		
	G5928																		
ZEP3	G10970	0.25	1.32	0.64	0.62	-	0.53	-	0.22	-	-	0.24	0.83	0.54	-	-	-	<b>5.19</b>	
	S1048																		
VDE	G51703	-	-	-	-	-	0.27	-	0.11	0.11	-	0.12	0.33	0.11	-	-	-	1.05	
VDL1	G36048	0.16	0.22	0.24	0.21	0.14	0.13	0.28	-	0.11	0.12	-	0.17	0.33	0.56	-	0.10	<b>2.76</b>	
VDL2	G45846	-	-	-	-	-	-	-	-	0.22	-	0.12	0.08	-	-	0.39	0.10	0.91	
VDR	G43240	-	-	-	-	-	-	-	-	0.22	0.12	-	-	0.11	-	-	-	0.45	
<b>total (%)</b>	12136	<b>2.64</b>	4544	12566	4821	6968	7508	3632	9122	9031	8552	<b>2.54</b>	<b>3.99</b>	9227	1.13	1.07	0.97		
<b>size library</b>																			10283

The database was searched for ESTs encoding genes of the diatom carotenoid biosynthetic pathway (see text for abbreviations of the genes) and their distribution in the different libraries was determined. The numbers given in the table reflect the amount of ESTs per one thousand ESTs (%) for each respective library. Also given are the EST contig ID numbers from the Gene Expression Database v3.0 and the total number of ESTs in each library. Abbreviations used for the libraries: OS) Original 12000 Standard, OM) Oval Morphotype in 10% seawater, TM) Triradiate Morphotype, TA) Tropical Accession at 15°C, SM) Silica Minus in artificial seawater, SP) Silica Plus in artificial seawater with 350  $\mu$ M metasilicate, NR) Nitrate Replete with 1.12 mM nitrate in chemostat, NS) Nitrate Starved for 3 days with 50  $\mu$ M nitrate in chemostat, AA) Ammonium Adapted at 75  $\mu$ M ammonium, UA) Urea Adapted at 50  $\mu$ M urea, FS) Iron Starved at 5 nM iron, BL) Blue Light for 1 h on 48 h dark adapted cells, LD) Low Decadialenol Treated for 6 h with 5  $\mu$ g/ml 2E,4E-decadialenol, HD) High Decadialenol Treated for 6 h with 5  $\mu$ g/ml 2E,4E-decadialenol, C1) 230  $\mu$ M/Kg of CO<sub>2</sub> for 1 day in chemostat, C4) 230  $\mu$ M/Kg of CO<sub>2</sub> for 4 days in chemostat. Note: this table gives *LTL1* and *LTL2* as alternative genes involved in  $\beta$ -carotene hydroxylation, because *BCH* is absent in *P. tricornutum*. doi:10.1371/journal.pone.0002896.t002



**Figure 6. mRNA levels of carotenoid biosynthesis and LHC-related genes upon white, blue or red light stimulation.** 48-hour-dark-adapted *P. tricornutum* cells were exposed to  $175 \mu\text{mol m}^{-2} \text{s}^{-1}$  continuous white light, or  $25 \mu\text{mol m}^{-2} \text{s}^{-1}$  continuous blue or red light and the relative transcript levels of *PSY* (A), *PDS1* (B), *FCPB* (C) and *ELIP-like* (D) were determined after 1, 3, 5, 8 and 12h by qRT-PCR using *H4* as a reference gene. The values were normalized to the transcript levels in the dark. Data are averages of triplicate measurements. The error bars represent standard deviation.

doi:10.1371/journal.pone.0002896.g006

secondary endosymbiosis. VDL proteins appeared due to a second duplication of VDR, which probably happened exclusively in organisms derived from secondary endosymbiosis. The ZEP genes are also duplicated and again this duplication likely occurred before the secondary endosymbiotic event. A single gene for ZEP from the ciliate *T. thermophila* has also been found, providing the first evidence for a secondary plastid gene being encoded in a heterotrophic ciliate.

The transcription of all *P. tricornutum* xanthophyll cycle-related genes was up-regulated by blue and white light. However, the amplitude and kinetics of mRNA accumulation varied between the different genes and it is possible that their transcription is regulated at multiple levels including photoreceptor- and plastid retrograde signaling mechanisms. The differential expression of  $\zeta\text{EP}$  and *VDE/VDL* homologs may be indicative of distinct roles for each gene product. Both  $\zeta\text{EP1/VDL2}$  and  $\zeta\text{EP3/VDE}$  form gene clusters and appear to be co-regulated upon light stimulation. These genes are not clustered in the *T. pseudonana* genome, providing an illustration of the continued divergence of this gene family even within the diatoms.

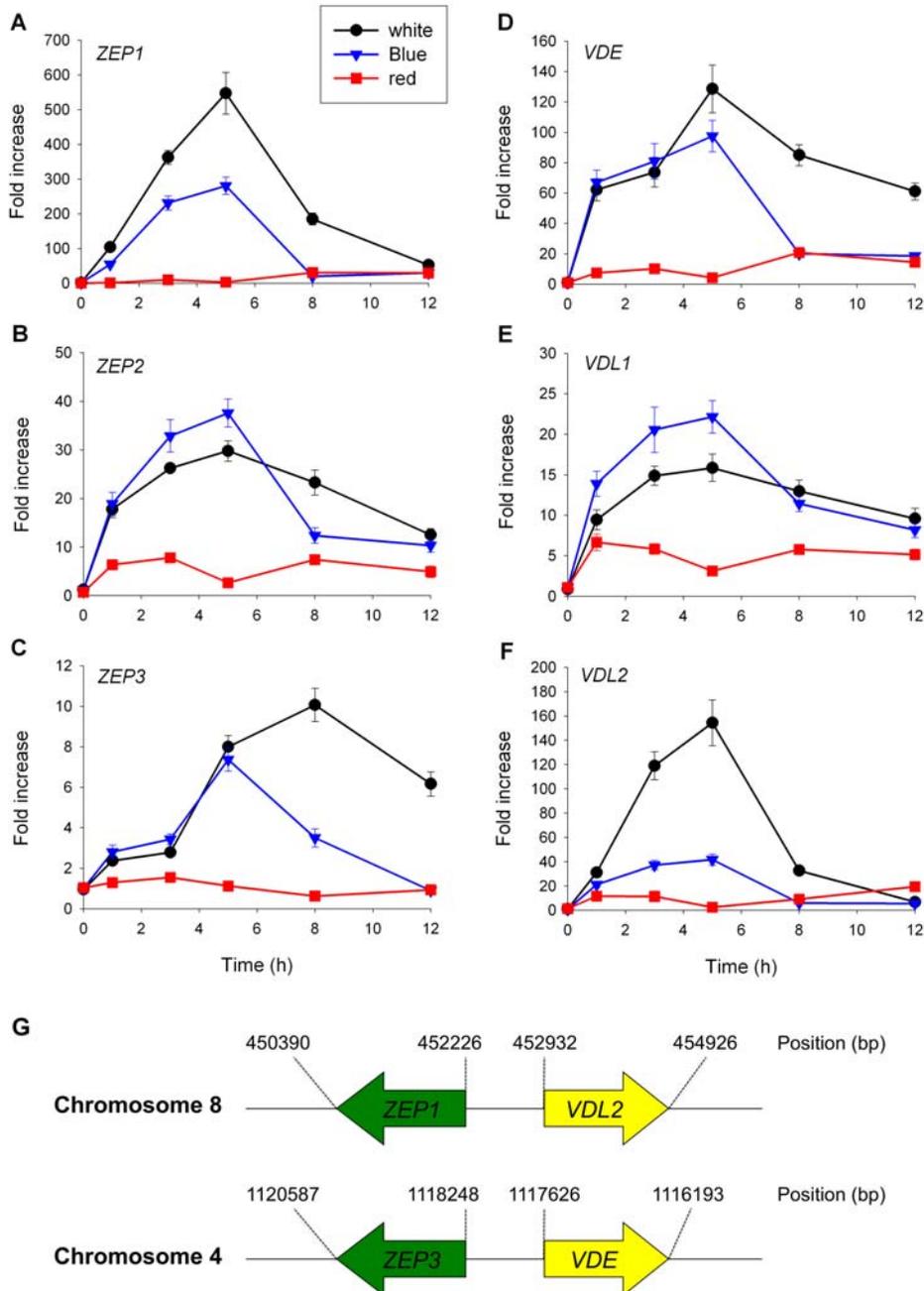
## Materials and Methods

### Identification, phylogeny and EST-distribution of carotenogenesis-related genes

The *P. tricornutum* genome sequence, available at the DOE Joint Genome Institute website (<http://genome.jgi-psf.org/Phatr2>), was analyzed for the presence of genes involved in the carotenoid

biosynthetic pathway using BLAST- and keyword searches. Using known sequences from other organisms and *P. tricornutum* EST sequence data available in the diatom EST database (<http://www.biologie.ens.fr/diatomics/EST3>), the best gene models were selected and annotated. The encoded genes were scanned for the presence of a signal peptide by SignalP v3.0 (<http://www.cbs.dtu.dk/services/SignalP>) and a plastid transit peptide (cTP) with ChloroP v1.1 (<http://www.cbs.dtu.dk/services/ChloroP>) [34]. The *P. tricornutum* sequences were used to identify homologs in the *T. pseudonana* genome (<http://genome.jgi-psf.org//Thaps3>), and keyword- and BLAST searches with sequences from other organisms were also performed. The identity and similarity of the *P. tricornutum* and *T. pseudonana* sequences were determined after ClustalW alignment of the mature protein sequences using Bioedit v7.0.5.3 and the BLOSUM62 similarity matrix [68]. The identity/similarity of the diatom sequences with the best BLASTP hits on the non-redundant protein sequences (nr) deposited at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>) were also determined.

For phylogenetic analysis, amino acid sequences were aligned using Kalign [69] at the EBI Tools Website (<http://www.ebi.ac.uk/kalign/>). Gaps and ambiguously aligned regions were excluded from further analysis. Phylogenetic trees were constructed using Maximum likelihood (ML) and Neighbor-joining (NJ) methods as implemented in PhyML [70] and AsaturA [40], respectively. ML trees were computed with particular models of amino acid substitutions chosen according to PROTTEST AIC results [70,71]. ML bootstraps were inferred from 300 replicates using



**Figure 7. mRNA levels of xanthophyll cycle-related genes upon white, blue or red light stimulation.** 48-hour-dark-adapted *P. tricornutum* cells were exposed to  $175 \mu\text{mol m}^{-2} \text{s}^{-1}$  continuous white light, or  $25 \mu\text{mol m}^{-2} \text{s}^{-1}$  continuous blue or red light and the relative transcript levels of ZEP1 (A), ZEP2 (B), ZEP3 (C), VDE (D), VDL1 (E) and VDL2 (F) were determined after 1, 3, 5, 8 and 12h by qPCR using *H4* as a reference gene. The values were normalized to the transcript levels in the dark. Data are averages of triplicate measurements. The error bars represent standard deviation.

doi:10.1371/journal.pone.0002896.g007

the same parameters. NJ trees were constructed using the same substitutional model as for ML computations, with NJ bootstraps calculated from 1000 replicates. The majority of the sequences used for phylogenetic analysis were derived from the NCBI database, but the genomes of *Ostreococcus tauri* and *O. lucimarinus* are available at JGI, the genome browser of *Cyanidioschyzon merolae* (<http://merolae.biol.s.u-tokyo.ac.jp>), the Taxonomically Broad EST Database (<http://tbestdb.bcm.umontreal.ca>), and the se-

quence database of the J. Craig Venter Institute (<http://www.jcvi.org/gme>) were also searched for homologous genes.

The *Phaeodactylum* Digital Gene Expression Database (v3.0) (<http://www.biologie.ens.fr/diatomics/EST3>) was used to gain insight into the differential expression of carotenogenesis-related genes. The BLASTN search algorithm linked to the database was used to identify the consensus sequence (contig) of multiple ESTs encoding the gene of interest, and the frequency and distribution

of the ESTs within each contig was determined. When genes were represented by more than one contig, the accuracy of the contigs was verified using the genome sequence and all gene-specific ESTs were taken into account.

### Strains and growth conditions

*P. tricornutum* Bohlin clone Pt1 8.6 (CCMP 2561) was obtained from the culture collection of the Provasoli-Guillard National Center for Culture of Marine Phytoplankton, Bigelow Laboratory for Ocean Sciences, USA. Cells were grown at 18°C under white fluorescent lights (TLD 58W/840, Philips) at approximately  $175 \mu\text{mol m}^{-2} \text{s}^{-1}$  in a 12 hr photoperiod, using f/2 medium made with 0.2- $\mu\text{m}$ -filtered and autoclaved local seawater supplemented with inorganic nutrients and vitamins according to Guillard [72]. Sterility was monitored microscopically and by occasional inoculation into peptone-enriched media to check for bacterial growth in darkness. To study the effect of light quality on steady state gene transcription, 3 L cultures were grown up to a cell density of  $2.10^6$  cells/mL as determined by microscopy using a Malassez counting chamber (Brand). The cultures were dark-adapted for 48 hours and subsequently submitted to either blue light ( $25 \mu\text{mol m}^{-2} \text{s}^{-1}$ ) with an emission between 380–450 nm (F 40 BB, Philips), red light ( $25 \mu\text{mol m}^{-2} \text{s}^{-1}$ ) provided by an LED lamp with a peak emission at 670 nm (QB1310CS-670-735, Quantum devices Inc.), or white fluorescent light ( $175 \mu\text{mol m}^{-2} \text{s}^{-1}$ ; TLD 58W/840, Philips). Samples for RNA extraction were collected at the end of the dark period and after the onset of light by centrifugation and the cell pellets were washed with PBS (1x) prior to storage in liquid nitrogen.

### Quantitative real-time RT-PCR

Total RNA was isolated from  $1.10^8$  cells using 1.5 mL TriPure isolation reagent (Roche) according to the instructions of the manufacturer. The RNA concentrations were determined photometrically at 260 nm and the integrity was confirmed by agarose gel electrophoresis. Residual genomic DNA was removed by incubating 1  $\mu\text{g}$  of RNA with 1 U/ $\mu\text{L}$  amplification grade DNaseI (Invitrogen) at 25°C for 10 min. First-strand cDNA synthesis for quantitative real-time RT-PCR (qRT-PCR) was obtained by incubating 430 ng of DNaseI-treated total RNA with 50 ng of random hexamers, 500  $\mu\text{M}$  dNTPs, and 50 U of SuperScript II reverse transcriptase (Invitrogen) in a 30  $\mu\text{L}$  reaction volume (1x RT buffer; Invitrogen) at 25°C for 10 min, followed by a 50 min incubation at 42°C. The reaction was stopped by heat inactivation at 70°C for 15 min, and RNA was removed with an RNase H treatment at 37°C for 20 min.

Gene-specific primers were designed with Primer3 (<http://frodo.wi.mit.edu>) and are listed in Supplementary Table S1. All primer pairs were initially tested by standard RT-PCR and the amplification of single products with the correct size was verified on 2% (w/v) agarose gels. Real Time-RT PCR amplification mixtures (25  $\mu\text{L}$ ) contained 1  $\mu\text{L}$  of cDNA obtained after the reverse transcription (estimated to represent 10 ng), 200 nM forward and reverse primers, and 2x FastStart SYBR Green I PCR Master Mix (Roche). Triplicate reactions were run in an Opticon Chromo4 MJ Research Thermal Cycler (Bio-Rad) in Low-Profile 0.2 mL PCR 8-Tube white Strips (Bio-Rad). The

cycling conditions comprised 10 min polymerase activation at 95°C and 40 cycles at 95°C for 15 sec and 60°C for 60 sec. The reaction was ended with a 5 min final elongation at 72°C. Amplicon dissociation curves, i.e., melting curves, were recorded after cycle 40 by heating from 60 to 95°C with a ramp speed of 0.5°C every second which served to confirm primer specificity. The results obtained in the Chromo4 Sequence Detector were exported as tab delimited text files and imported into Microsoft Excel for further analysis. Primer efficiencies were verified by real time RT-PCR on serial dilutions of cDNA. The relative steady state mRNA transcript levels were calculated by comparing the cycle threshold ( $C_T$ ) values of the target and the *histone H4*-control transcripts throughout the time-course (time x) with the  $C_T$  values of these transcripts at 48 hours dark (time 0) using the  $2^{-\Delta\Delta C_T}$  method [73].  $\Delta\Delta C_T$  is represented by the following formula:  $\Delta\Delta C_T = (C_{T,target} - C_{T,control\ gene})_{time\ x} - (C_{T,target} - C_{T,control\ gene})_{time\ 0}$

### Supporting Information

**Supplementary Information S1** Formation of  $\beta^2$ -carotene and zeaxanthin

Found at: doi:10.1371/journal.pone.0002896.s001 (0.07 MB DOC)

**Table S1** Primers used for quantitative real-time RT-PCR

Found at: doi:10.1371/journal.pone.0002896.s002 (0.06 MB DOC)

**Figure S1** ClustalW amino acid sequence alignment of diatom LTLs with LUT proteins from *A. thaliana* Amino acids shaded in black are identical, and those shaded in grey are similar (BLOSUM62). Sequence accession numbers (NCBI) are as follows: *A. thaliana* (At) Lut1, NP\_190881; Lut5, NP\_564384 and *Skeletonema costatum* (Sc) P450, AAL73435. Protein ID numbers of *P. tricornutum* (Pt; Phatr v2.0) are: LTL1, 50101 and LTL2, 26422 and protein ID numbers of *T. pseudonana* (Tp; Thaps v3.0) are: LTL1, 9541 and LTL2, 36235.

Found at: doi:10.1371/journal.pone.0002896.s003 (2.60 MB TIF)

**Figure S2** PSY and PDS1 gene transcription following a 5 minute blue light pulse 60-hour-dark-adapted *P. tricornutum* cells were exposed for 5 min to  $25 \mu\text{mol m}^{-2} \text{s}^{-1}$  blue light and subsequently transferred back to darkness. The relative transcript levels of PSY and PDS1 were determined after 30 min and 1, 3, 5 and 7h by qRT-PCR using H4 as a reference gene. The values were normalized to the transcript levels in the dark. Data are averages of triplicate measurements. The error bars represent standard deviation.

Found at: doi:10.1371/journal.pone.0002896.s004 (0.09 MB TIF)

### Acknowledgments

The authors thank Dr. Martin Lohr for critical reading of the manuscript.

### Author Contributions

Conceived and designed the experiments: SC AF CB. Performed the experiments: SC MO. Analyzed the data: SC MO. Contributed reagents/materials/analysis tools: JV AF CB. Wrote the paper: SC CB.

### References

- Nelson DM, Tréguer P, Brzezinski MA, Leynaert B, Quéguiner B (1995) Production and dissolution of biogenic silica in the ocean-Revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochem Cycles* 9: 359–372.
- Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281: 237–240.
- Mann DG (1999) The species concept in diatoms. *Phycologia* 38: 437–495.

4. Kooistra WH, De Stefano M, Mann DG, Medlin LK (2003) The phylogeny of the diatoms. *Prog Mol Subcell Biol* 33: 59–97.
5. Kooistra WH, Gersonde R, Medlin LK, Mann DG (2007) The origin and evolution of the diatoms: their adaptation to a planktonic existence; Falkowski PG, Knoll AH, eds. Amsterdam: Academic Press-Elsevier. 480 p.
6. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, et al. (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306: 79–86.
7. Allen AE, Vardi A, Bowler C (2006) An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. *Curr Opin Plant Biol* 9: 264–273.
8. Montsant A, Allen AE, Coesel S, De Martino A, Falcitore A, et al. (2007) Identification and comparative genomic analysis of signaling and regulatory components in the diatom *Thalassiosira pseudonana*. *J Phycol* 43: 585–604.
9. Siaut M, Hejide M, Mangogna M, Montsant A, Coesel S, et al. (2007) Molecular toolbox for studying diatom biology in *Phaeodactylum tricorutum*. *Gene* 406: 23–35.
10. Falcitore A, Bowler C (2002) Revealing the molecular secrets of marine diatoms. *Annu Rev Plant Biol* 53: 109–130.
11. Maheswari U, Montsant A, Goll J, Krishnasamy S, Rajyashri KR, et al. (2005) The Diatom EST Database. *Nucleic Acids Res* 33: D344–347.
12. Montsant A, Jabbari K, Maheswari U, Bowler C (2005) Comparative genomics of the pennate diatom *Phaeodactylum tricorutum*. *Plant Physiol* 137: 500–513.
13. Li S, Nosenko T, Hackett JD, Bhattacharya D (2006) Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. *Mol Biol Evol* 23: 663–674.
14. Patron NJ, Rogers MB, Keeling PJ (2004) Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. *Eukaryot Cell* 3: 1169–1175.
15. Yoon HS, Hackett JD, Cimiglia C, Pinto G, Bhattacharya D (2004) A molecular timeline for the origin of photosynthetic eukaryotes. *Mol Biol Evol* 21: 809–818.
16. Cavalier-Smith T (1981) Eukaryote kingdoms: seven or nine? *Biosystems* 14: 461–481.
17. MacIntyre HL, Kana TM, Geider RJ (2000) The effect of water motion on short-term rates of photosynthesis by marine phytoplankton. *Trends Plant Sci* 5: 12–17.
18. Strzepek RF, Harrison PJ (2004) Photosynthetic architecture differs in coastal and oceanic diatoms. *Nature* 431: 689–692.
19. Wagner H, Jakob T, Wilhelm C (2006) Balancing the energy flow from captured light to biomass under fluctuating light conditions. *New Phytol* 169: 95–108.
20. Ruban A, Lavaud J, Rousseau B, Guglielmi G, Horton P, et al. (2004) The super-excess energy dissipation in diatom algae: comparative analysis with higher plants. *Photosynth Res* 82: 165–175.
21. Lavaud J, Rousseau B, Etienne AL (2003) Enrichment of the light-harvesting complex in diadinoxanthin and implications for the nonphotochemical fluorescence quenching in diatoms. *Biochemistry* 42: 5802–5808.
22. Falkowski PG, Laroche J (1991) Acclimation to spectral irradiance in algae. *J Phycol* 27: 8–14.
23. Muller P, Li XP, Niyogi KK (2001) Non-photochemical quenching. A response to excess light energy. *Plant Physiol* 125: 1558–1566.
24. Holt NE, Fleming GR, Niyogi KK (2004) Toward an understanding of the mechanism of nonphotochemical quenching in green plants. *Biochemistry* 43: 8281–8289.
25. Niyogi KK, Li XP, Rosenberg V, Jung HS (2005) Is PsbS the site of non-photochemical quenching in photosynthesis? *J Exp Bot* 56: 375–382.
26. Hager A (1969) Lichtbedingte pH-erniedrigung in einem chloroplastenkompatibel als ursache der enzymatische violaxanthin- zu zeaxanthin- umwandlung; Beziehungen zur photophosphorylierung. *Planta* 89: 224–243.
27. Yamamoto HY, Kamite L, Wang YY (1972) An ascorbate-induced absorbance change in chloroplasts from violaxanthin de-epoxidation. *Plant Physiol* 49: 224–228.
28. Bugos RC, Yamamoto HY (1996) Molecular cloning of violaxanthin de-epoxidase from romaine lettuce and expression in *Escherichia coli*. *Proc Natl Acad Sci U S A* 93: 6320–6325.
29. Marin E, Nussaume L, Quesada A, Gonneau M, Sotta B, et al. (1996) Molecular identification of zeaxanthin epoxidase of *Nicotiana plumbaginifolia*, a gene involved in abscisic acid biosynthesis and corresponding to the ABA locus of *Arabidopsis thaliana*. *Embo J* 15: 2331–2342.
30. Lavaud J, Rousseau B, Etienne AL (2002) In diatoms, a transthylakoid proton gradient alone is not sufficient to induce a non-photochemical fluorescence quenching. *FEBS Lett* 523: 163–166.
31. Lavaud J, Rousseau B, van Gorkom HJ, Etienne AL (2002) Influence of the diadinoxanthin pool size on photoprotection in the marine planktonic diatom *Phaeodactylum tricorutum*. *Plant Physiol* 129: 1398–1406.
32. Olaizola M, La Roche J, Kolber Z, Falkowski P (1994) Non-photochemical fluorescence quenching and the diadinoxanthin cycle in a marine diatom. *Photosynth Res* 41: 357–370.
33. Stransky H, Hager A (1970) Das Carotinoidmuster und die Verbreitung des lichtinduzierten Xanthophyllzyklus in verschiedenen Algenklassen. *Arch Microbiol* 73: 315–323.
34. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2: 953–971.
35. Pfundel EE, Renganathan M, Gilmore AM, Yamamoto HY, Dilley RA (1994) Intrathylakoid pH in isolated pea chloroplasts as probed by violaxanthin de-epoxidation. *Plant Physiol* 106: 1647–1658.
36. Hieber AD, Bugos RC, Verhoeven AS, Yamamoto HY (2002) Overexpression of violaxanthin de-epoxidase: properties of C-terminal deletions on activity and pH-dependent lipid binding. *Planta* 214: 476–483.
37. Niyogi KK, Grossman AR, Bjorkman O (1998) *Arabidopsis* mutants define a central role for the xanthophyll cycle in the regulation of photosynthetic energy conversion. *Plant Cell* 10: 1121–1134.
38. Bugos RC, Hieber AD, Yamamoto HY (1998) Xanthophyll cycle enzymes are members of the lipocalin family, the first identified from plants. *J Biol Chem* 273: 15321–15324.
39. Flower DR (1996) The lipocalin protein family: structure and function. *Biochem J* 318 (Pt 1): 1–14.
40. Hager A, Holocher K (1994) Localization of the xanthophyll-cycle enzyme violaxanthin de-epoxidase within the thylakoid lumen and abolition of its mobility by a (light-dependent) pH decrease. *Planta* 192: 581–589.
41. Grouneva I, Jakob T, Wilhelm C, Goss R (2006) Influence of ascorbate and pH on the activity of the diatom xanthophyll cycle-enzyme diadinoxanthin de-epoxidase. *Physiol Plant* 126: 205–211.
42. Jakob T, Goss R, Wilhelm C (2001) Unusual pH-dependence of diadinoxanthin de-epoxidase activation causes chlororespiratory induced accumulation of diatoxanthin in the diatom *Phaeodactylum tricorutum*. *J Plant Physiol* 158: 383–390.
43. Goss R (2003) Substrate specificity of the violaxanthin de-epoxidase of the primitive green alga *Mantoniella squamata* (Prasinophyceae). *Planta* 217: 801–812.
44. Hieber AD, Bugos RC, Yamamoto HY (2000) Plant lipocalins: violaxanthin de-epoxidase and zeaxanthin epoxidase. *Biochim Biophys Acta* 1482: 84–91.
45. Charron JB, Ouellet F, Pelletier M, Danyluk J, Chauve C, et al. (2005) Identification, expression, and evolutionary analyses of plant lipocalins. *Plant Physiol* 139: 2017–2028.
46. Ganfornina MD, Gutierrez G, Bastiani M, Sanchez D (2000) A phylogenetic analysis of the lipocalin protein family. *Mol Biol Evol* 17: 114–126.
47. Salier JP (2000) Chromosomal location, exon/intron organization and evolution of lipocalin genes. *Biochim Biophys Acta* 1482: 25–34.
48. Nakamura Y, Kaneko T, Sato S, Mimuro M, Miyashita H, et al. (2003) Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. *DNA Res* 10: 181–201.
49. Joshua S, Bailey S, Mann NH, Mullineaux CW (2005) Involvement of phycoobilisome diffusion in energy quenching in cyanobacteria. *Plant Physiol* 138: 1577–1585.
50. Marquardt J, Hanelt D (2004) Carotenoid composition of *Delesseria lanicifolia* and other marine red algae from polar and temperate habitats. *Eur J Phycol* 39: 285–292.
51. Schubert N, Garcia-Mendoza E, Pacheco-Ruiz I (2006) Carotenoid composition of marine red algae. *J Phycol* 42: 1208–1216.
52. Bouvier F, d'Harlingue A, Huguency P, Marin E, Marion-Poll A, et al. (1996) Xanthophyll biosynthesis. Cloning, expression, functional reconstitution, and regulation of beta-cyclohexenyl carotenoid epoxidase from pepper (*Capsicum annuum*). *J Biol Chem* 271: 28861–28867.
53. Goss R, Ann Pinto E, Wilhelm C, Richter M (2006) The importance of a highly active and DeltapH-regulated diatoxanthin epoxidase for the regulation of the PS II antenna function in diadinoxanthin cycle containing algae. *J Plant Physiol* 163: 1008–1021.
54. Mewes H, Richter M (2002) Supplementary ultraviolet-B radiation induces a rapid reversal of the diadinoxanthin cycle in the strong light-exposed diatom *Phaeodactylum tricorutum*. *Plant Physiol* 130: 1527–1535.
55. Lohr M, Wilhelm C (1999) Algae displaying the diadinoxanthin cycle also possess the violaxanthin cycle. *Proc Natl Acad Sci U S A* 96: 8784–8789.
56. Lohr M, Wilhelm C (2001) Xanthophyll synthesis in diatoms: quantification of putative intermediates and comparison of pigment conversion kinetics with rate constants derived from a model. *Planta* 212: 382–391.
57. Wilhelm C, Buchel C, Fisahn J, Goss R, Jakob T, et al. (2006) The regulation of carbon and nutrient assimilation in diatoms is significantly different from green algae. *Protist* 157: 91–124.
58. Berges JA, Charlebois DO, Mauzerall DC, Falkowski PG (1996) Differential effects of nitrogen limitation on photosynthetic efficiency of photosystems I and II in microalgae. *Plant Physiol* 110: 689–696.
59. Green RM, Geider RJ, Falkowski PG (1991) Effect of iron limitation on photosynthesis in a marine diatom. *Limnol Oceanogr* 36: 1772–1782.
60. Paul MJ, Pellny TK (2003) Carbon metabolite feedback regulation of leaf photosynthesis and development. *J Exp Bot* 54: 539–547.
61. Nott A, Jung HS, Koussevitzky S, Chory J (2006) Plastid-to-nucleus retrograde signaling. *Annu Rev Plant Biol* 57: 739–759.
62. Kleine T, Kindgren P, Benedict C, Hendrickson L, Strand A (2007) Genome-wide gene expression analysis reveals a critical role for cryptochrome1 in the response of *Arabidopsis* to high irradiance. *Plant Physiol* 144: 1391–1406.
63. Leblanc C, Falcitore A, Watanabe M, Bowler C (1999) Semi-quantitative RT-PCR analysis of photoregulated gene expression in marine diatoms. *Plant Mol Biol* 40: 1031–1044.
64. Bohne F, Linden H (2002) Regulation of carotenoid biosynthesis genes in response to light in *Chlamydomonas reinhardtii*. *Biochim Biophys Acta* 1579: 26–34.
65. Im CS, Eberhard S, Huang K, Beck CF, Grossman AR (2006) Phototropin involvement in the expression of genes encoding chlorophyll and carotenoid

- biosynthesis enzymes and LHC apoproteins in *Chlamydomonas reinhardtii*. *Plant J* 48: 1–16.
66. von Lintig J, Welsch R, Bonk M, Giuliano G, Batschauer A, et al. (1997) Light-dependent regulation of carotenoid biosynthesis occurs at the level of phytoene synthase expression and is mediated by phytochrome in *Sinapis alba* and *Arabidopsis thaliana* seedlings. *Plant J* 12: 625–634.
67. Falcatore A, Bowler C (2005) The evolution and function of blue and red light photoreceptors. *Curr Top Dev Biol* 68: 317–350.
68. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41: 95–98.
69. Lassmann T, Sonnhammer EL (2005) Kalign—an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics* 6: 298.
70. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696–704.
71. Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21: 2104–2105.
72. Guillard RRL (1975) Culture of phytoplankton for feeding marine invertebrates; Smith WL, Chanley MH, eds. New York: Plenum Press.
73. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C(T)}$  method. *Methods* 25: 402–408.