Supporting Information

Bailleul et al. 10.1073/pnas.1007703107

SI Materials and Methods

Identification of the *P. tricornutum* LHCXs. LHCXs were identified in the *Phaeodactylum tricornutum* genome (protein ID 27278 for LHCX1, protein ID 25104 for LHCX2, protein ID 44733 for LHCX3, and protein ID 38720 for LHCX4; http://genome.jgi-psf.org/Phatr2/Phatr2.home.html) and annotated by B. Green as an ORF of 630 bp in length for the *LHCX1*; 714 bp in length for the *LHCX2*; 633 bp in length for the *LHCX3*; and 639 bp in length for the *LHCX4*, respectively. The corresponding full-length predicted proteins are around 220 aa long, with a calculated mass of 21.9 kDa for LHCX1; 24.7 kDa for LHCX2; 22.2 kDa for LHCX3; and 22.8 kDa for LHCX4. In all of the predicted proteins diatom-specific N-terminal bipartite presequences, consisting of a signal and a transit peptide-like and a conserved cleavage site of the signal peptides, have been identified.

Construction of the Vectors for Gene Silencing and Overexpression. Different vectors containing an antisense fragment from the different LHCXs were generated in a vector bearing a phleomycin resistance cassette, as reported in ref. 1. To avoid a specific silencing of the gene family, the 3′ end and part of the 3′-UTR for each *LHCX* gene were used as targeted regions. The four fragments were amplified by PCR from the the Pt1 cDNA using the following primers:

ASL1fw (containing an EcoRI site) 5′-ccgGAATTCGACCAAGGAACTCCAGAACG-3′
ASL1rv (containing a XbaI site) 5′-ctagTCTAGAGGTCGGTGAACAAGAGCTTC-3′,
(product size: 242);

ASL2 fw (containing an EcoRI site) 5′-ccgGAATTCTTGAACACTAGGCGCGATG-3′
ASL2 rv (containing a XbaI site) 5′-ctagTCTAGAGGGCAATAACTCGGAATACG-3′
(product size: 210);

ASL3fw (containing an EcoRI site) 5′-ccgGAATTCCTTCTGACCCGGAAGAACTG-3′
ASL3rv (containing a XbaI site) 5′-ctagTCTAGAGTGCATGTGACCAGCCATAC-3′
(product size: 182);

ASL4fw (containing an EcoRI site) 5′-ccgGAATTCACTACAGAATGGGCGTTTGG-3′
ASL4rv (containing a XbaI site) 5′-ctagTCTAGACCATTGCAACTCTTGGTTGA-3′
(product size: 182).

The resulting amplification products were digested with EcoRI and XbaI and subsequently introduced in the antisense orientation, between the *Sh ble* gene and the FcpA terminator, into the EcoRI–XbaI linearized pKS–hsa PtGUS vectors, replacing the *GUS* gene fragments vector (1).

Vectors for the *LHCX1* overexpression were generated by cloning the full-length *LHCX1* gene, amplified with the following primers: L1OEfw (containing a NotI site) 5′-acgtGCGGCCGCATGAAGTTCGCTGCCACCATC-3′ and L1OErv (containing a EcoRI site) 5′-acgtGAATTCTTAACCCTGAAGATTCTCAAGGATT-3′, in the NotI–EcoRI sites of FcpBp–A3 plasmid.

Gene Expression Analysis. The different *LHCX* transcripts were amplified with the following primers: *LHCX1* with the primers L1fw 5′-CCTTGCTCTTATCGGCTCTG-3′, L1rv 5′-ACGGTATCGCTTCAAAGTGG -3′; *LHCX2* with the primers L2fw 5′-CAGCACTAATGCCGCTTTCG-3′, L2rv 5′-CGTGAGTAACTTCCGCTTCC-3′; *LHCX3* with the primers L3fw 5′-TCCCGTTGGTATCTTTGATCC-3′, L3rv 5′-GAAGATCCTTCCACGGCTTC-3′; and *LHCX4* with the primers L4fw 5′-TCTTTGATCCACTCCGCTTC-3′, L4rv 5′-GGCGTTCCATAGAAAGTTCG-3′. The gene encoding the fucoxanthin chlorophyll a/c-binding protein (protein ID 9799), used as control for the silencing specificity (Fig. S2*B*), and was amplified with the primers Fcpfw 5′-GGAAATCGCCACCTTCAATA-3′ and Fcprv 5′-ACCCAGGAAGGCAATCATAG-3′. The *RPS* housekeeping gene (ribosomal protein small subunit 30S; protein ID 10847), was used as reference gene for data normalization and was amplified with the primers Rpsfw 5′-CGAAGTCAACCAGGAAACCAA-3′ and Rpsrv 5′-GTGCAAGAGACCGGACATACC-3′. Changes in gene expression in the wild type and *LHCX1* transgenic lines and in the time course experiments were determined using *RPS* as a reference gene and values were normalized to gene expression levels of the wild type (Pt1), when indicated.

Phylogenetic Analysis. For phylogenetic analysis, amino acid sequences were aligned using ClustalW in Bioedit (2). 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 (3) and MEGA 4.0.2 (4), respectively. ML tree was computed using the LG model for amino acid substitution, as selected by PROTTEST (5). NJ tree was computed using the JTT amino acid substitution model as present in MEGA. ML bootstraps were calculated from 300 replicates and NJ bootstraps from 1,000 replicates. Amino acid sequences were retrieved from JGI genome portal and NCBI GenBank. JGI protein IDs are: *Phaeodactylum tricornutum*, PtLhx1 (Phatr2:27278), PtLhcx2 (Phatr2:54065), PtLhcx3 (Phatr2:44733), and PtLhcx4 (Phatr2:38720); *Thalassiosira pseudonana*, TpLhcX1 (Thaps3:264921), TpLhcX2 (Thaps3:38879), TpLhcX4 (Thaps3:270228), TpLhcX5 (Thaps3:31128), and TpLhcX6 (Thaps3:12097); *Fragilariopsis cylindricus*, FcLhcX1 (Fracy1:272042), FcLhcX2 (Fracy1:190506), FcLhcX3 (Fracy1:218498), FcLhcX4 (Fracy1:244998), FcLhcX5 (Fracy1:192081), FcLhcX6 (Fracy1:271538), and FcLhcX7 (Fracy1:209887); *Ostreococcus sp*. RC809: OspLI818a (OstRCC809_1:45963),OspLI818b (OstRCC809_1:39546), and OspLI818c (OstRCC809_1:87531); and *Volvox carteri*, VcLhcX (Volca1:80312). NCBI accession numbers are: *Chlamydomonas reinhardtii*, CrLhcSR1 (XP_001696125), CrLhcSR2 (XP_001696064), and CrLhcSR3 (XP_001696138); *Chlamydomonas eugametos*, CeLI818 (Q03965); *Cyclotella cryptica*, CcFcpAC1 (CAA04404) and CcFcpAC2 (CAA04403); *Ostreococcus tauri*, OtFcpAC (CAL53934); *Ostreococcus lucimarinus*, OlFcpAC (XP_001417976); *Micromonas sp*. CCMP490, MspLI818_1 (DAA05943) and MspLI818_2 (DAA05937); *Mesostigma viride*, MvLI818_1 (ABD37894) and MvLI818_2 (DAA05932); *Bigelowiella natans*, BnLI818_1 (AAP79202) and BnLI818_3 (DAA05890); *Isochrysis galbana*, IgLhc10 (ABA55526); *Karlodinium micrum*, KmLhc2 (ABA55555) and KmLhc4 (ABV22205); *Physcomitrella patens*, Ppx (XP_001768071); and *Scenedesmus obliquus*, SoLI818 (ABD58893).

1. De Riso V, et al. (2009) Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res* 37:e96.
2. 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.
3. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704.
4. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
5. Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.

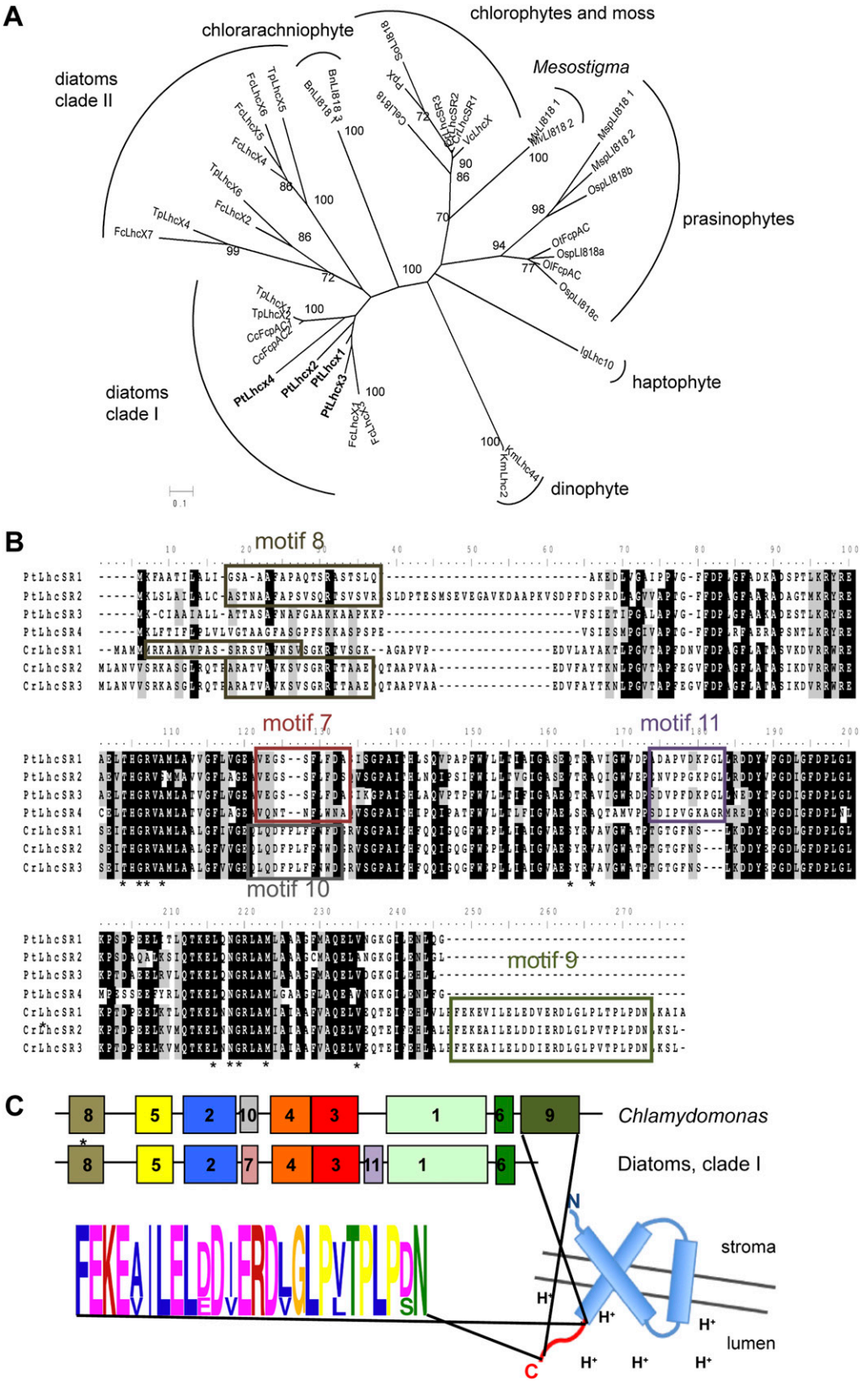

Fig. S1. LHCX1 phylogeny and protein motif characteristics. (*A*) Maximum-likelihood phylogenetic tree as inferred from LHCXs and related protein amino acid sequences (110 amino acid positions). Numbers at branches indicate ML bootstrap support. Only values ≥70% are indicated. ML tree topology was verified using the neighbor-joining method. Species abbreviations are: Pt, *Phaeodactylum tricornutum*; Tp, *Thalassiosira pseudonana*; Fc, *Fragilariopsis cylindrus*; Cc, *Cyclotella cryptica*; Bn, *Bigelowiella natans*; Ig, *Isochrysis galbana*; Km, *Karlodinium micrum*; Osp, *Ostreococcus* sp. RC809; Ot, *Ostreococcus tauri*; Ol, *Ostreococcus lucimarinus*; Msp, *Micromonas* sp. CCMP490; Mv, *Mesostigma viride*; Vc, *Volvox carteri*; Cr, *Chlamydomonas reinhardtii*; Ce, *Chlamydomonas eugametos*; *So, Scenedesmus obliquus*; Ppx, *Physcomitrella patens.* (*B*) *C. reinhardtii* and *P. tricornutum* ClustalW protein sequences alignment. Black and gray

Legend continued on following page

shading represent identical and similar residues, respectively. Specific motifs, identified by the MEME motif discovery tool v4.3.0, are indicated with boxes. (*C*) Schematic presentation of shared and specific *Chlamydomonas* and diatom protein motifs. The asterisk indicates absence in PtLHCX 3 and 4. The proposed LHCX protein model shows the localization of the additional C-terminal domain in *Chlamydomonas*. Assuming a similar topology for LHCSR and LHCII proteins, the "extra" domain of *Chlamydomonas* LHCSR should be localized in the thylakoid lumen, where it could confer the observed pH sensitivity to NPQ.

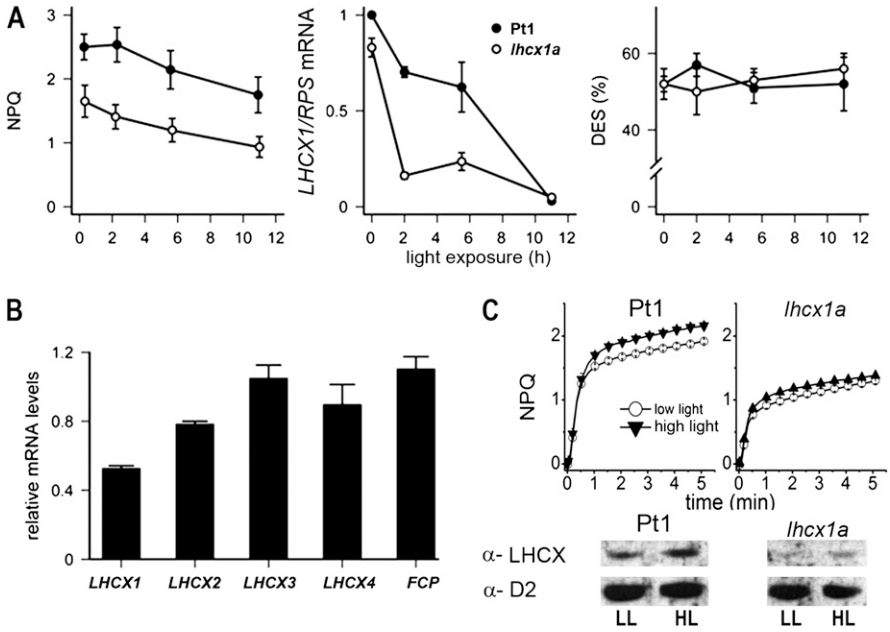

Fig. S2. Analysis of gene expression, NPQ and DES in Pt1 and in the *Ihcx1a* transgenic line. (*A*) Diurnal changes in NPQ (*Left*), *LHCX1* expression (*Center*) and DES (*Right*). Experiments have been performed as described in Fig. 2. (*B*) Relative transcript levels of the four *P. tricornutum LHCX* and *FCP* transcripts in the *Ihcx1a* strain, in cells grown at 70 μmol photons $m^{-2}{\cdot}s^{-1}$, as determined by qRT-PCR. *RPS* was used as reference gene and values normalized to gene expression levels of the wild type (Pt1). (*C*) NPQ level and LHCX protein content in cells grown in low light (70 μmol photons $m^{-2}{\cdot}s^{-1}$) and 1 h after a low light to high light shift (600 μmol photons $m^{-2}{\cdot}s^{-1}$).

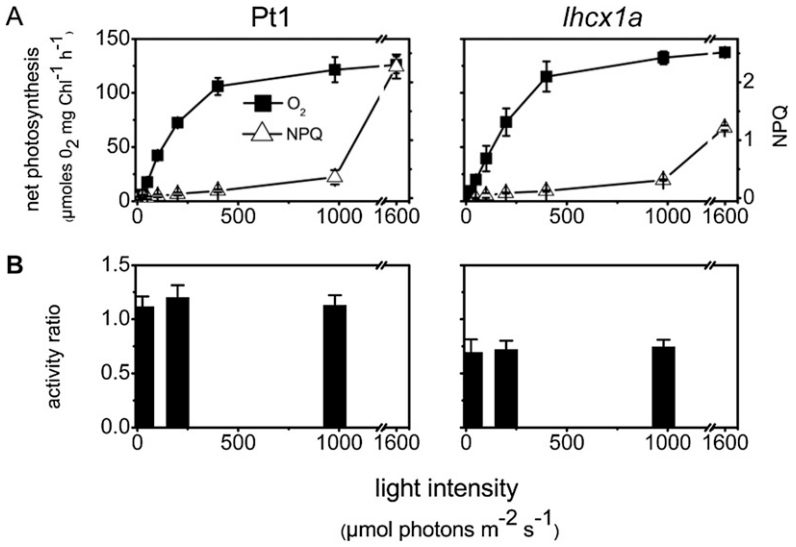

Fig. S3. Light saturation curves of oxygen evolution (squares) and NPQ (triangles) and photoinhibition in Pt1 and *Ihcx1a*. (A) Oxygen evolution was monitored using a Clark electrode under increasing light intensity, adjusted every 5 min. Photosynthetic activity was calculated as "net photosynthesis" (i.e., photosynthesis after correction for respiration) at any given light intensity. NPQ was calculated as (Fm-Fm′)/Fm′. (B) Inhibition of oxygen evolution after high light illumination. Cells were first exposed for ~20 min to 2,500 μmol $m^{-2}{\cdot}s^{-1}$ of white light. Then, photosynthesis activity was measured at different light intensities and normalized to its value before high light treatment. Error bars: SEM of three independent measurements. Note that because green light is strongly absorbed by *Phaeodactylum*, a lower NPQ was seen when cells were illuminated with 700 μmol photons $m^{-2}{\cdot}s^{-1}$ of white light than with the same "amount" of green light.

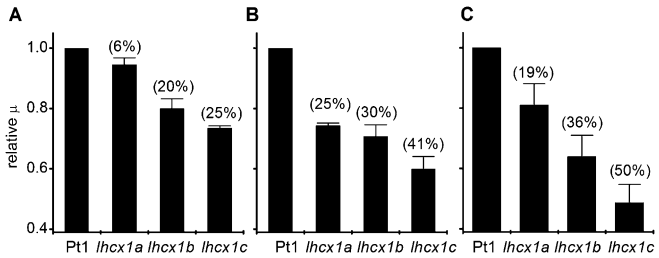

Fig. S4. Growth rates of Pt1 and *lhcx1* RNAi lines in different light conditions. Cells were grown under a 12 h:12 h light:dark regime at light intensity of 30 μmol photons $m^{-2}{\cdot}sec^{-1}$ (*A*) and 500 μmol photons $m^{-2}{\cdot}sec^{-1}$ (*B*) or under 1 h:1h light:dark intermittent regime at 70 μmol photons $m^{-2}{\cdot}sec^{-1}$ (*C*). The specific growth rates were calculated in the exponential phase. Relative growth rates (μ) were then obtained upon normalization to the value calculated for Pt1 (*A*) $0.49 \pm 0.01\ d^{-1}$; (*B*) $1.01 \pm 0.02\ d^{-1}$; (*C*) $1.11 \pm 0.07\ d^{-1}$). Error bars are relative to duplicate measurements in three independent biological samples. The percentage of the relative growth rate decrease in the mutants is shown in parentheses.

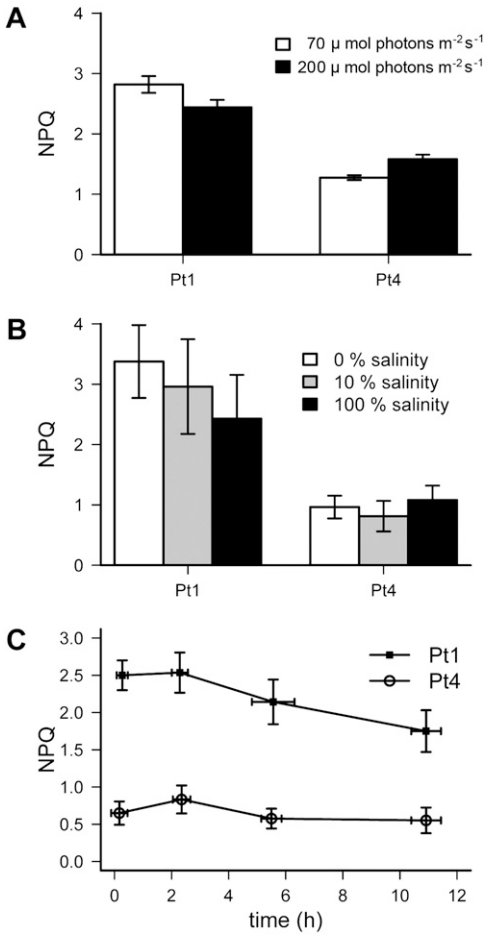

Fig. S5. NPQ capacity in Pt1 and Pt4 cells. NPQ levels are shown as a function of the growth light intensity (*A*) and salinity (*B*). Cultures were exposed to conditions indicated in the figure panels for at least 3 wk before measuring NPQ capacity. (*C*) NPQ profiles during the 12 h low light exposure (70 μmol photons $m^{-2}{\cdot}s^{-1}$). Error bars indicate SEM of three independent measurements.

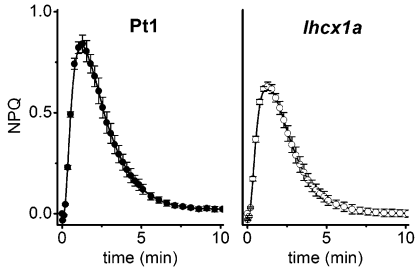

Fig. S6. Transient NPQ measured in Pt1 and *lhcx1a* cells during a dark to low light transition. Cells, collected during exponential growth (~10^6 cells·mL^{-1}), were dark adapted for 1.5 h. NPQ was measured during a 10-min illumination with 50 μmol photons m^{-2}·s^{-1} using the fluorescence CCD camera recorder, as described in *Materials And Methods*. Error bars are relative to duplicate measurements in two independent experiments.

Table S1. Pigment composition of dark adapted *P. tricornutum* strains

	Pt1	*lhcx 1a*	*lhcx 1b*	*lhcx 1c*	*Pt4*	*Pt4 OE1*	*Pt4 OE2*
chlc/chla	0.32 ± 0.08	0.31 ± 0.02	0.32 ± 0.03	0.33 ± 0.03	0.33 ± 0.04	0.35 ± 0.05	0.31 ± 0.04
Fx/chla	0.90 ± 0.20	0.90 ± 0.06	0.88 ± 0.07	0.91 ± 0.08	0.96 ± 0.08	1.01 ± 0.08	0.98 ± 0.06
DD/chla	0.17 ± 0.05	0.18 ± 0.02	0.15 ± 0.02	0.21 ± 0.03	0.18 ± 0.02	0.22 ± 0.02	0.19 ± 0.01
βcar/chla	0.06 ± 0.02	0.06 ± 0.02	0.07 ± 0.01	0.06 ± 0.02	0.04 ± 0.02	0.05 ± 0.03	0.04 ± 0.02

Cells were grown under a 12:12 h dark:light regime at a light intensity of 70 μmol photons m^{-2}·s^{-1}. They were dark adapted for 30 min before being frozen, and their pigment composition was measured after thawing as described in *Materials and Methods*. No diatoxanthin (DT) was detected in dark adapted samples. chlc, chlorophyll c; chla, chlorophyll a; Fx, fucoxanthin; βcar, β- carotene.