Supplementary information for :

"Diatom phytochromes integrate the entire visible light spectra for photosensing in marine environments", Duchêne et al.,

1. Supplementary Method
2. Extended Data Figures
3. Extended Data Tables

Supplementary Data 1: (separate excel file)

Sheet 1: list of diatom genetic resources used in this study, Sheet2: Sequences of diatom phytochromes used to build hmm models, Sheet 3: Sequences of diatom aureochromes used to build hmm models. Sheet 4 and 5: Sequences of the *Tara* Oceans gene catalog (MATOU2-v2) identified as diatom aureochromes and diatom phytochromes. Sheet 6: List of *Tara* Oceans diatom MAGs with detected diatom aureochromes and diatom phytochromes presence.

Supplementary Method: Detailed description of DPH activity model construction based on equations described in (Mancinelli, 1994) *(13)*

Phytochrome model

Monomer:
In a model where phytochrome is acting as a monomer, phytochrome equilibrium can be theorized as the equilibrium between the Pfr→Pr transition (rate constant k1) and the Pr→Pfr reverse reaction (rate constant k2) with the following scheme:

$$Pfr \underset{k1}{\overset{k2}{\leftrightharpoons}} Pr$$

The rate constant of phytochrome photoconversion depends on phytochrome extinction coefficient, photoconversion yield and on light intensity, so k1 and k2 can be expressed as:
$k_1 = N * \sigma_{Pfr} = N * 2.3 * \varepsilon_{Pfr} * \phi_{Pfr}$ and $k_2 = N * 2.3 * \varepsilon_{Pr} * \phi_{Pr}$

where N is the fluence rate, or light intensity (µmol photon.m^{-2}.s^{-1}), and σ_{Pfr} is the photoconversion cross-section of the Pfr→Pr reaction, ε_{Pfr} is the extinction coefficient of Pfr and ϕ_{Pfr} the quantum yield of the Pfr→Pr reaction; σ_{Pr} is the photoconversion cross-section of the Pr→Pfr reaction, ε_{Pr} is the extinction coefficient of Pr and ϕ_{Pr} the quantum yield of the Pr→Pfr reaction.

At equilibrium, we have $\frac{dPrPr}{dt} = k_1 * Pfr - k_2 * Pr = 0$
Considering that the total amount of phytochrome P stays constant, i.e. synthesis and degradation are negligible, then $P = Pr + Pfr$, and we have:

$$\frac{Pr_{eq}}{P} = \frac{k1}{k1+k2} = \frac{1}{1+\frac{k2}{k1}} = \frac{1}{1+\frac{\varepsilon_{Pr}}{\varepsilon_{Pfr}}*\frac{\phi_{Pr}}{\phi_{Pfr}}} = \frac{1}{1+\frac{\varepsilon_{Pr}}{\varepsilon_{Pfr}}*\eta} \quad (1)$$

With $\eta = \frac{\phi_{Pr}}{\phi_{Pfr}}$

Dimer:
In a model where phytochrome is acting as a dimer, the scheme becomes:

$$PfrPfr \underset{2k1}{\overset{k2}{\leftrightharpoons}} PrPfr \underset{k1}{\overset{2k2}{\leftrightharpoons}} PrPr$$

We have $\frac{dPrPr}{dt} = k_1 * PrPfr - 2k_2 * PrPr$
At equilibrium, $\frac{dPrPr}{dt} = 0$
Once again, we consider that P is constant, $P = PrPr + PrPfr + PfrPfr$, hence

$$\frac{PrPr_{eq}}{P} = \left(\frac{k_1}{k_1+k_2}\right)^2 = \frac{1}{\left(1+\frac{\varepsilon_{Pr}}{\varepsilon_{Pfr}}*\frac{\phi_{Pr}}{\phi_{Pfr}}\right)^2} = \frac{1}{\left(1+\frac{\varepsilon_{Pr}}{\varepsilon_{Pfr}}*\eta\right)^2} \quad (2)$$

$$\text{And } \frac{PfrPfr_{eq}}{P} = \left(\frac{k_2}{k_1+k_2}\right)^2 = \left(\frac{\frac{\varepsilon_{Pr}}{\varepsilon_{Pfr}}*\frac{\phi_{Pr}}{\phi_{Pfr}}}{1+\frac{\varepsilon_{Pr}}{\varepsilon_{Pfr}}*\frac{\phi_{Pr}}{\phi_{Pfr}}}\right)^2 \quad (3)$$

Integration of Equation 3 gives the following expression for PrPr as a function of time:

$$\frac{\mathrm{PrPr}(t)}{P} = \left(\frac{k_1}{k_1+k_2} * \frac{PfrPfr(t0)}{P} + \frac{k_2}{k_1+k_2} * \frac{PrPr(t0)}{P} - \frac{k_1*k_2}{(k_1+k_2)^2}\right) e^{-2*(k_1+k_2)t} + \frac{k_1}{k_1+k_2}\left(\frac{PrPr(t0)}{P} - \frac{PfrPfr(t0)}{P} + \frac{k_2-k_1}{k_1+k_2}\right)e^{-(k_1+k_2)t} + \left(\frac{k_1}{k_1+k_2}\right)^2 \quad (4)$$

Use of YFP signal as indicative of DPH activity: estimation of the ratio of quantum yield

In equation 1, 2 or 3, $\varepsilon_{Pr}/\varepsilon_{Pfr}$ can be known from the recombinant PtDPH absorption spectra. The $\varepsilon_{Pr}/\varepsilon_{Pfr}$ ratio, in both monochromatic and bichromatic illumination, as been determined with the LED spectra with:

$$\frac{\varepsilon_{Pr,LED}}{\varepsilon_{Pfr,LED}} = \frac{\sum_{\lambda=300}^{\lambda=900} A_{Pr,\lambda} * N_\lambda}{\sum_{\lambda=300}^{\lambda=900} A_{Pfr,\lambda} * N_\lambda}$$

where $A_{Pr,\lambda}$ is the absorbance of Pr at wavelength λ measured with the recombinant protein, and N_λ the intensity of the LED measured at this wavelength.

We hypothesized that the YFP signal at saturating light is directly linked to phytochrome equilibrium. Our output measure is the YFP signal for a LED at wavelength λ at saturating intensities (the 4 points at highest light intensity were used), normalized between the signal from cells that went from constant green light (520 nm, growth condition) to darkness, and the signal of cells exposed 10 min to saturating far-red light (765 nm). This is expressed as following:

$$RelativeYFP(LED\lambda, sat) = \frac{YFP(LED\lambda, sat) - YFP(520nm)}{YFP(765nm, sat) - YFP(520nm)}$$

In a monomer model, Pr is the gene expression-inducting form[6]. If we consider that the YFP is linearly linked to Pr/P, we have:

$$RelativeYFP(LED\lambda, sat) = \frac{Pr_{eq,LED} - Pr_{eq,520nm}}{Pr_{eq,765nm} - Pr_{eq,520nm}} = \frac{\frac{1}{1+\frac{\varepsilon_{Pr,LED}}{\varepsilon_{Pfr,LED}}*\eta} - \frac{1}{1+\frac{\varepsilon_{Pr,520}}{\varepsilon_{Pfr,520}}*\eta}}{\frac{1}{1+\frac{\varepsilon_{Pr,765}}{\varepsilon_{Pfr,765}}*\eta} - \frac{1}{1+\frac{\varepsilon_{Pr,520}}{\varepsilon_{Pfr,520}}*\eta}} \quad (5)$$

In this equation, η is the only unknown.
To determine it, the equation was fitted with the nls function in R with η= 1 as starting value, on the experimental values obtained from monochromatic illuminations (excluding the red illumination), see Extended Data Table 3 for the results.

In this monomer model, there is no mathematical difference whether Pr is actively inducting gene expression or if Pfr is actively repressing it; in both cases, Pr is the only form that allows genes to be expressed.

In a dimer model however, if we considering PrPr as the active form (i.e. only PrPr will allow gene expression), we expect the YFP to reflect PrPr/P. On the other hand, if we consider that PfrPfr is the active form, i.e. repressing gene expression, the gene induction will be a function of $1 - \frac{PfrPfr}{P}$.

These two cases can be put to equation as follow:
If we consider that the YFP is linearly linked to PrPr/P, then :

$$RelativeYFP(LED\lambda, sat) = \frac{PrPr_{eq,LED} - PrPr_{eq,520nm}}{PrPr_{eq,765nm} - PrPr_{eq,520nm}} \frac{\frac{1}{(1+\frac{\varepsilon_{Pr,LED}}{\varepsilon_{Pfr,LED}}*\eta)^2} - \frac{1}{(1+\frac{\varepsilon_{Pr,520}}{\varepsilon_{Pfr,520}}*\eta)^2}}{\frac{1}{(1+\frac{\varepsilon_{Pr,765}}{\varepsilon_{Pfr,765}}*\eta)^2} - \frac{1}{(1+\frac{\varepsilon_{Pr,520}}{\varepsilon_{Pfr,520}}*\eta)^2}} \quad (6)$$

(6)

If we consider that the YFP is linearly linked to 1-PfrPfr/P, then :

$$RelativeYFP(LED\lambda, sat) = \frac{PfrPfr_{eq,520nm} - PfrPfr_{eq,LED}}{PfrPfr_{eq,520nm} - PrPr_{eq,765nm}} \frac{(\frac{\frac{\varepsilon_{Pr,520}}{\varepsilon_{Pfr,520}}*\eta}{1+\frac{\varepsilon_{Pr,520}}{\varepsilon_{Pfr,520}}*\eta})^2 - (\frac{\frac{\varepsilon_{Pr,LED}}{\varepsilon_{Pfr,LED}}*\eta}{1+\frac{\varepsilon_{Pr,LED}}{\varepsilon_{Pfr,LED}}*\eta})^2}{(\frac{\frac{\varepsilon_{Pr,520}}{\varepsilon_{Pfr,520}}*\eta}{1+\frac{\varepsilon_{Pr,520}}{\varepsilon_{Pfr,520}}*\eta})^2 - (\frac{\frac{\varepsilon_{Pr,765}}{\varepsilon_{Pfr,765}}*\eta}{1+\frac{\varepsilon_{Pr,765}}{\varepsilon_{Pfr,765}}*\eta})^2} \quad (7)$$

Once again in these equations, η is the only unknown.

To determine it, the equation was fitted with the nls function in R with η= 1 as starting value, on the experimental values obtained from monochromatic illuminations, (excluding the red illumination) see Extended Data Table 3 for the results.

The PrPr model was also tested on data including bichromatic illuminations, see Extended Data Table 3 for the results.

Use of YFP signal as indicative of DPH activity: sensitivity

From this point, we used the PrPr model. Sensitivity is here defined as the phytochrome photoconversion cross-section σ. In theory, we have the exponential constant in equation 4

$(k_1 + k_2) = N * (\sigma_{Pfr} + \sigma_{Pr}) = N * \sigma$ and

$$\sigma = 2.3 * \varepsilon_{Pr} * \phi_{Pr} + 2.3 * \varepsilon_{Pfr} * \phi_{Pfr} = A_{Pr} * \phi'_{Pr} + A_{Pfr} * \phi'_{Pfr}$$

With ϕ'_{Pr} and ϕ'_{Pfr} corresponding to the apparent photoconversion quantum yields.

The following equation was fitted to the YFP levels for each LED and each line (we used the ratio of photoconversion η estimated above to calculate PrPr in green and far-red):

$$RelativeYFP(LED\lambda, t) = \frac{PrPr(LED\lambda,t) - PrPr_{eq,520nm}}{PrPr_{eq,765nm} - PrPr_{eq,520nm}} = \left[\left(\frac{A_{Pfr}\phi'_{Pfr}}{A_{Pfr}\phi'_{Pfr} + A_{Pr}\phi'_{Pr}} * \frac{PfrPfr(t0)}{P} + \frac{A_{Pr}\phi'_{Pr}}{A_{Pfr}\phi'_{Pfr} + A_{Pr}\phi'_{Pr}} * \frac{PrPr(t0)}{P} - \frac{A_{Pfr}\phi'_{Pfr} * A_{Pr}\phi'_{Pr}}{(A_{Pfr}\phi'_{Pfr} + A_{Pr}\phi'_{Pr})^2} \right) e^{-2*(A_{Pfr}\phi'_{Pfr} + A_{Pr}\phi'_{Pr})*N*t} + \frac{A_{Pfr}\phi'_{Pfr}}{A_{Pfr}\phi'_{Pfr} + A_{Pr}\phi'_{Pr}} \left(\frac{PrPr(t0)}{P} - \frac{PfrPfr(t0)}{P} + \frac{A_{Pr}\phi'_{Pr} - A_{Pfr}\phi'_{Pfr}}{A_{Pfr}\phi'_{Pfr} + A_{Pr}\phi'_{Pr}} \right) e^{-(A_{Pfr}\phi'_{Pfr} + A_{Pr}\phi'_{Pr})*N*t} + \left(\frac{A_{Pfr}\phi'_{Pfr}}{A_{Pfr}\phi'_{Pfr} + A_{Pr}\phi'_{Pr}} \right)^2 - \frac{1}{\left(1 + \frac{A_{Pr,520}}{A_{Pfr,520}}*\eta\right)^2} \right] / \left[\frac{1}{(1 + \frac{A_{Pr,765}}{A_{Pfr,765}}*\eta)^2} - \frac{1}{(1 + \frac{A_{Pr,520}}{A_{Pfr,520}}*\eta)^2} \right]$$

(8)

where %PrPr at time 0 is either the equilibrium in green (for the induction spectra) or in far-red (for the reversion spectra), *t*, the time of illumination, i.e. 10 min, N the intensity $N = \sum_{\lambda=300}^{\lambda=900} N_\lambda$

Φ'Pr and Φ'Pfr are fitted for each LED and each line. For further analysis, we used the photoconversion cross section σ calculated from Φ'Pr and Φ'Pfr fitted values.

$$\sigma = A_{Pfr}\phi'_{Pfr} + A_{Pr}\phi'_{Pr}$$

In theory, σ should be a linear combination of (A_{Pr}* η +A_{Pfr}) (dotted lined on Fig.2C).

$$\sigma = \alpha * \left(A_{Pr,LED} * \eta + A_{Pfr,LED}\right)$$

where $A_{Pr,LED} = \frac{\sum_{\lambda=300}^{\lambda=900} A_{Pr,\lambda} * N_\lambda}{\sum_{\lambda=300}^{\lambda=900} N_\lambda}$, and α is a constant

Fitting the linear relationship between σ and (A_{Pr}* η +A_{Pfr}) with the lm function in R gave α=0.015154±0.002613 (adjusted R^2= 0.3855) when considering all the data points (Figure 3C), α=0.027574 ±0.004133 (adjusted R^2= 0.6085) when removing the data points above 700 nm and α=0. 0.0044465±0. 0.0003856 (adjusted R^2= 0.8461) when considering only the data points above 700 nm. Given this difference, we estimated the DPH cross section σ in the natural environment as the sum of σ below and σ above 700 nm see also below).

<u>Projection of DPH activity</u>

From the previous equations, we can calculate $\%PrPr(t) = \frac{\mathrm{PrPr}(t)}{P} * 100$ in a given environment. The light environment is described with $I_{tot} = l * \sum_{\lambda=300}^{\lambda=900} N_\lambda$ with l the bandwidth and N_λ the intensity at wavelength λ.

The ratio of absorption is calculated with $\frac{\varepsilon_{Pr}}{\varepsilon_{Pfr}} = \frac{\sum_{\lambda=300}^{\lambda=900} A_{Pr,\lambda} * N_\lambda}{\sum_{\lambda=300}^{\lambda=900} A_{Pfr,\lambda} * N_\lambda}$, from which we can calculate the DPH equilibrium:

$$\%PrPr_{eq} = \frac{1}{(1 + \frac{\varepsilon_{Pr}}{\varepsilon_{Pfr}} * \eta)^2} * 100$$

(η is either the value determined *in vivo* or *in vitro*)

We can calculate Pr and Pfr absorption in this environment with $A_{Pr} = \frac{\sum_{300}^{\lambda} A_{Pr,\lambda} * N_\lambda}{\sum_{300}^{\lambda} N_\lambda}$, from which we calculate the photoconversion cross-section

$$\sigma = 027574 * \left(A_{Pr,\lambda<700} * \eta + A_{Pfr,\lambda<700}\right) + 0.0044465 * \left(A_{Pr,\lambda>700} * \eta + A_{Pfr,\lambda>700}\right)$$

(this distinction is only relevant for real environmental data with λ from 350 to 800 nm, but not for modelled light fields, λ from 350 to 700 nm).

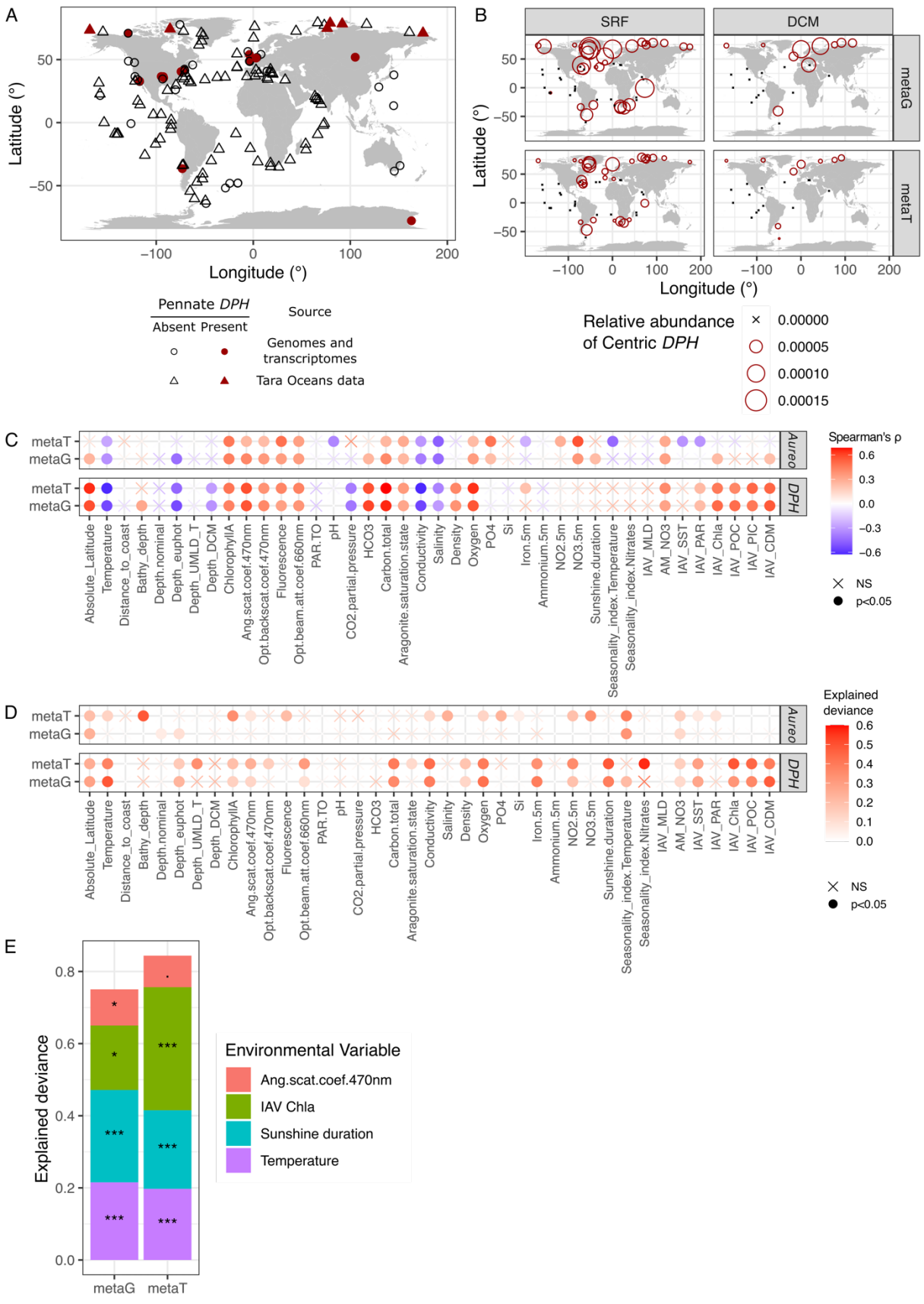
The rate constant displayed in Fig3.B is $k = \sigma * I_{tot}$

We also calculated PrPr(t) with PrPr(t0)=0, PfrPfr(t0)=1 and

$$\frac{\mathrm{PrPr}(t)}{P} = \left(\frac{k_1}{k_1+k_2} * \frac{PfrPfr(t0)}{P} + \frac{k_2}{k_1+k_2} * \frac{PrPr(t0)}{P} - \frac{k_1 * k_2}{(k_1+k_2)^2}\right) e^{-2*\sigma*Itot*t} + \frac{k_1}{k_1+k_2}\left(\frac{PrPr(t0)}{P} - \frac{PfrPfr(t0)}{P} + \frac{k_2-k_1}{k_1+k_2}\right)e^{-\sigma*Itot*t} + \left(\frac{k_1}{k_1+k_2}\right)^2$$

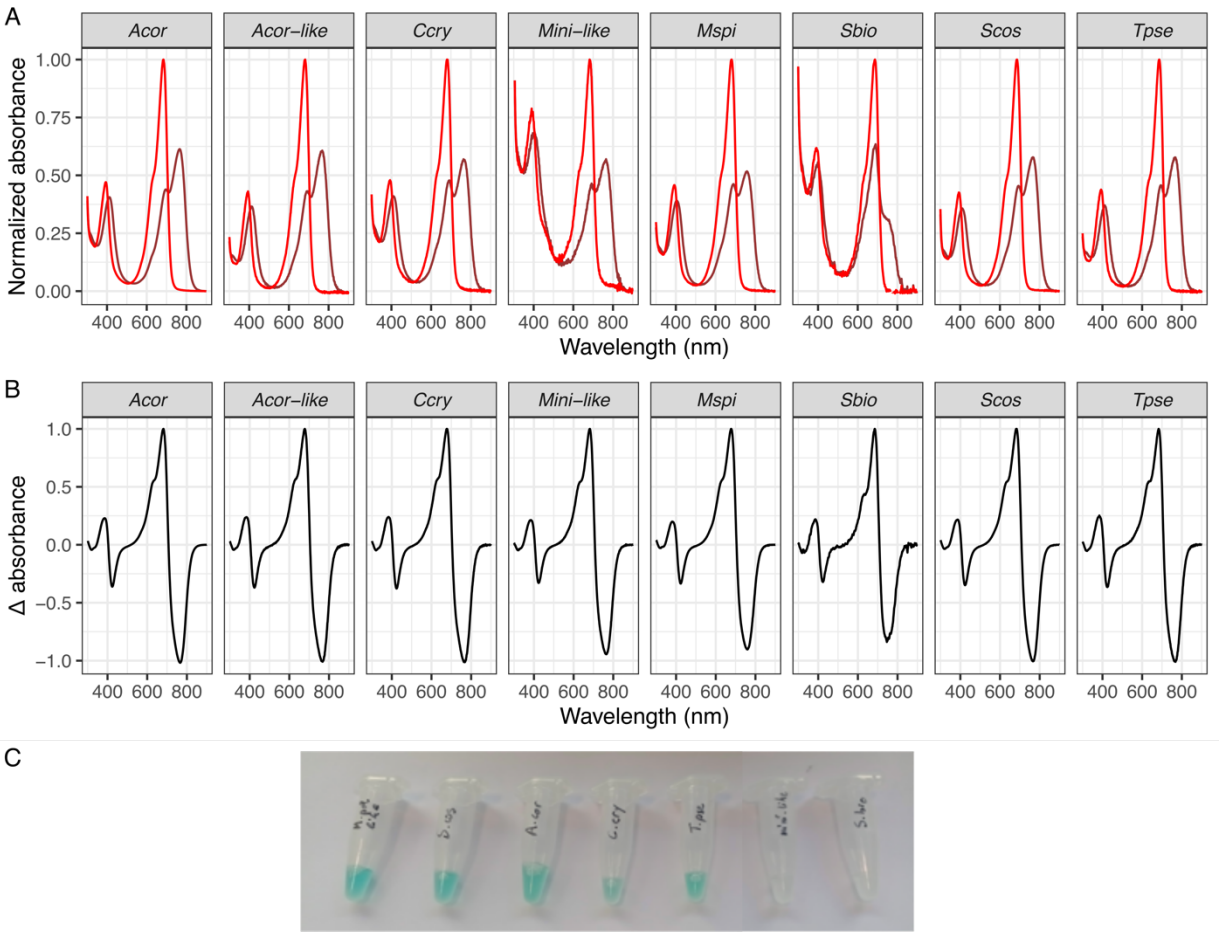
From this, we calculated $\sigma * Itot$ for which PrPr(t) ≥ 0.9 $\mathrm{PrPr_{eq}}$ as the limit for light detection by DPH, with t=10 min, 1h or 12h.

Extended Data Figures


A
Latitude (°)
Longitude (°)
Pennate DPH
Absent Present
Source
Genomes and transcriptomes
Tara Oceans data
B
SRF
DCM
metaG
metaT
Relative abundance of Centric DPH
0.00000
0.00005
0.00010
0.00015
C
metaT
metaG
Aureo
DPH
Spearman's ρ
NS
p<0.05
Absolute_Latitude
Temperature
Distance_to_coast
Bathy_depth
Depth.nominal
Depth_euphot
Depth_UMLD_T
Depth_DCM
ChlorophyllA
Ang.scat.coef.470nm
Opt.backscat.coef.470nm
Fluorescence
Opt.beam.att.coef.660nm
PAR.TO
pH
CO2.partial.pressure
HCO3
Carbon.total
Aragonite.saturation.state
Conductivity
Salinity
Density
Oxygen
PO4
Si
Iron.5m
Ammonium.5m
NO2.5m
NO3.5m
Sunshine.duration
Seasonality_index.Temperature
Seasonality_index.Nitrates
IAV_MLD
AM_NO3
IAV_SST
IAV_PAR
IAV_Chla
IAV_POC
IAV_PIC
IAV_CDM
D
Explained deviance
E
Explained deviance
Environmental Variable
Ang.scat.coef.470nm
IAV Chla
Sunshine duration
Temperature

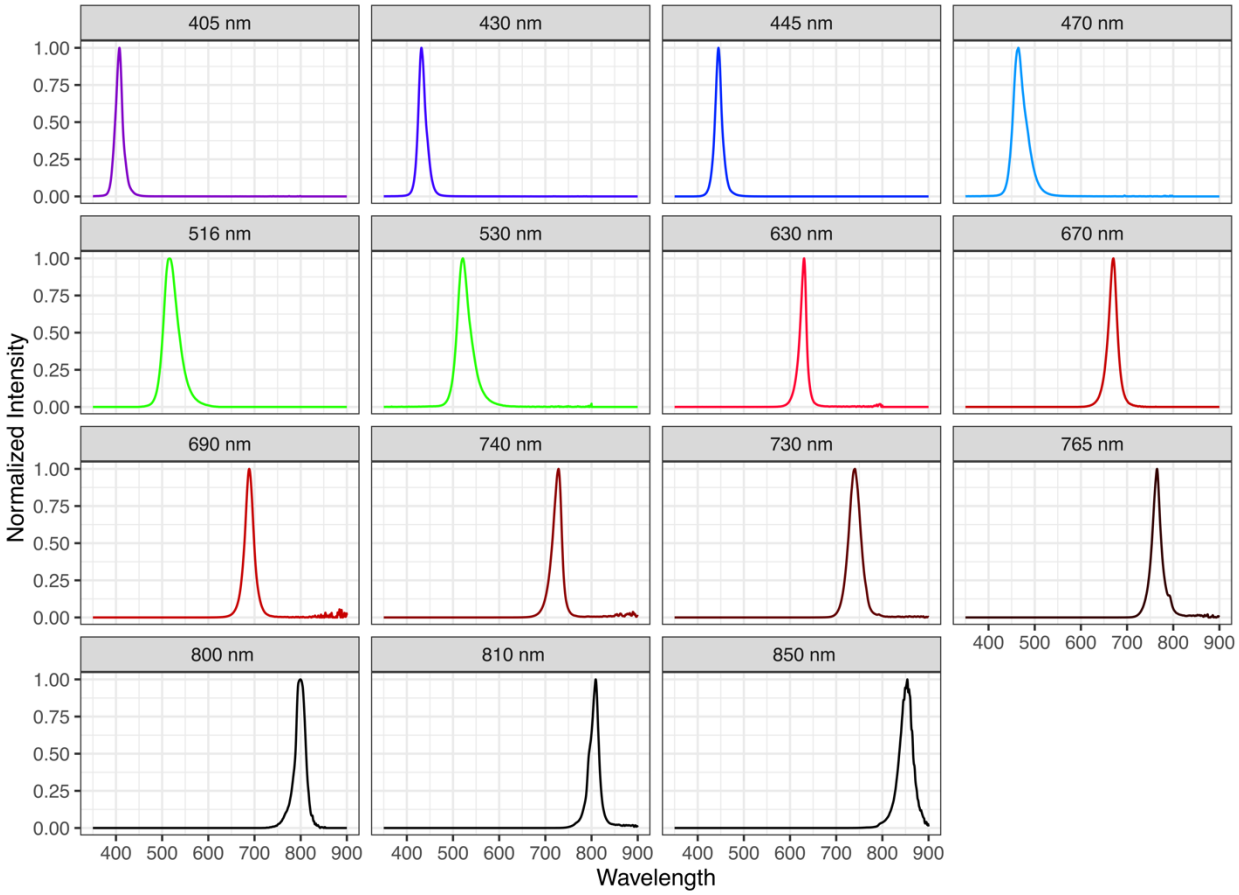
Extended Data Fig.1

DPH distribution is linked to latitude, temperature and optical parameters. (A) Map of the presence and absence of pennate *DPH* genes based on *Tara* Oceans sequence data and sampling location of pennate diatom strains with and without *DPH*. (B to E) Analysis of centric *DPH* biogeography from *Tara* Oceans data: (B) Map of the abundance of centric *DPH* genes and transcripts in *Tara* Oceans sampling stations (SRF: surface samples, DCM: Deep Chlorophyll Maximum; metaT: meta-transcriptomic reads, metaG: meta-genomic reads). Abundance is relative to total centric diatom genes or transcripts. (C) Spearman's correlation of *AUREO* and *DPH* relative abundance (in meta-transcriptomic reads, metaT, and meta-genomic reads, metaG) with different environmental parameters. DPH MetaT and MetaG reads also significantly correlate with each other (Pearson's r=0.758, p-value<<<0.01) (D) Generalized additive models (GAM) of *AUREO* and *DPH* relative abundances with different environmental parameters. (E) Complex GAMs explaining *DPH* relative abundance with a combination of environmental parameters.



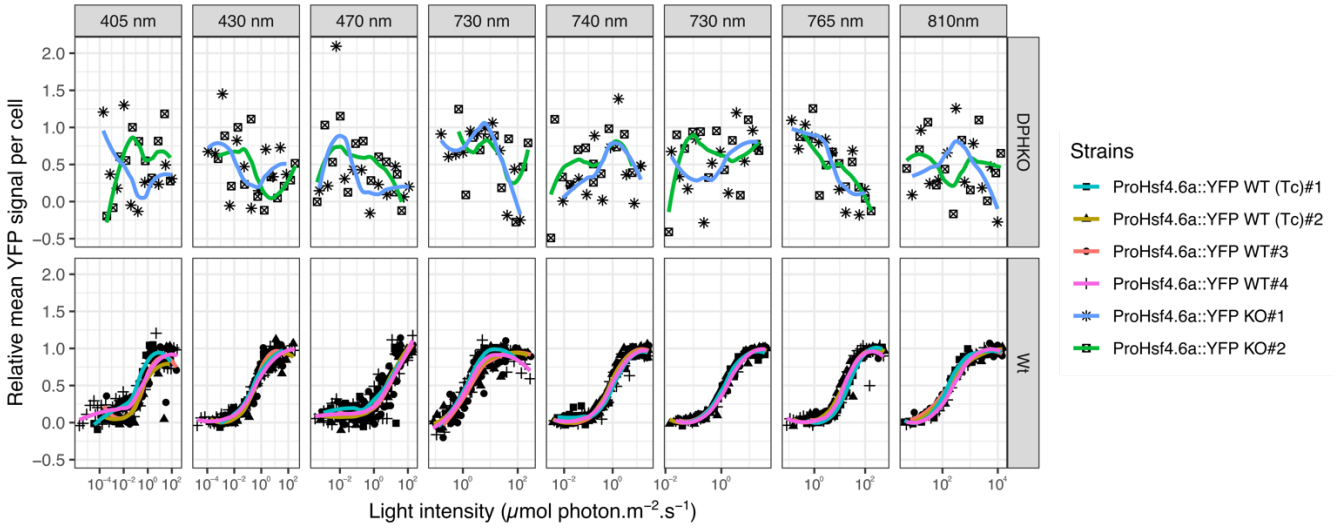
Extended Data Fig.2

Spectral properties of DPH from various species are conserved. (A) Normalized absorption spectra of far-red light illuminated (red line) and red light illuminated (darker red line) recombinant photosensory domains of DPH expressed with biliverdin as the conjugate chromophore. (B) Normalized differential absorption spectra between red- and far-red- illuminated DPH. (C) Purified recombinant DPH. *Acor Arcocellulus cornucervis, Ccry Cyclotella cryptica, Mspi Minidiscus spinulatus, Sbio Shionodiscus bioculatus, Scos Skeletonema costatum, Tpse Thalassiosira pseudonana*; Mini-like: synthetic environmental sequence close to *M. spinulatus DPH*, Acor-like synthetic environmental sequence close to *A. cornucervis DPH*.



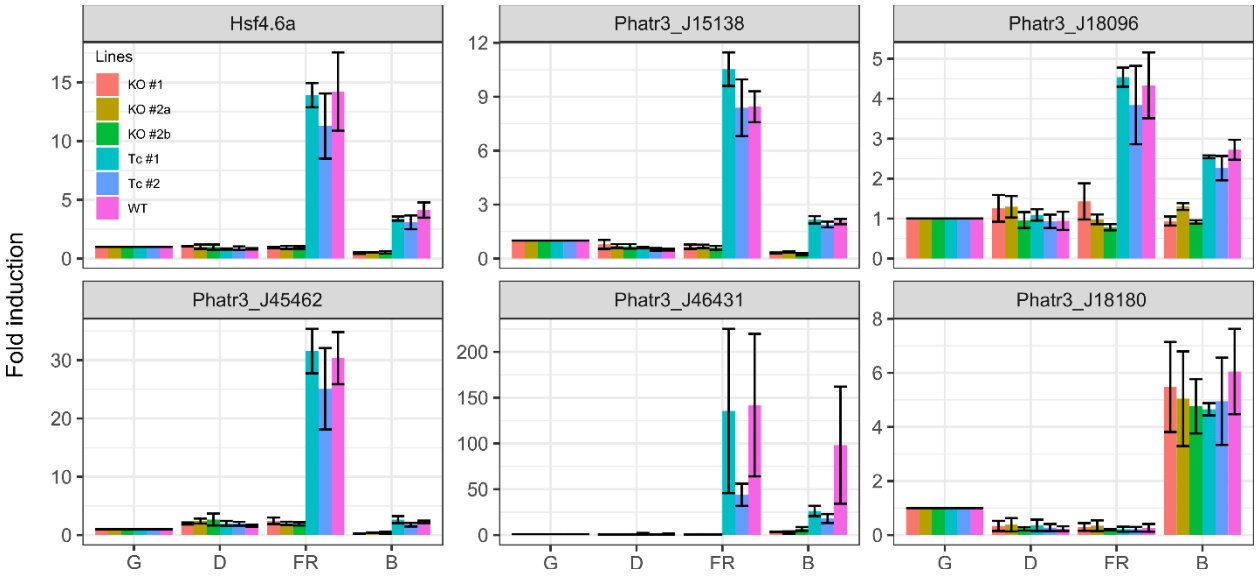
Extended Data Fig.3

Emission spectra of the LED used in this study. Intensity (photon.m^{-2}.s^{-1}) were normalized to the maximum for each LED.



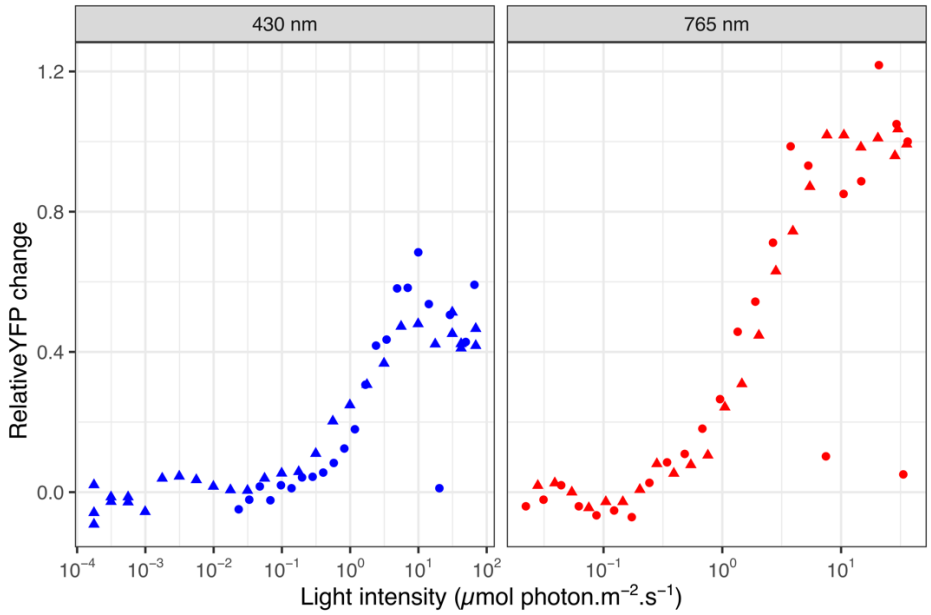
Extended Data Fig.4

The PtDPH response reporter system. Fluence rate-response curves induction of YFP in response to different lights in reporter lines with different genetic backgrounds: *P. tricornutum* wild-type (WT), knockout PtDPH mutants (KO), or non-mutated cells originated from the same transformed colony than the KO (WT (Tc), Transformation control). Cells were for 10 min exposed to an intensity gradient of monochromatic lights. YFP signal was measured after 6 h in the dark by flow cytometry and rescaled between minimum and maximum for each light treatment and each line (mean of the 3 minimum and the 3 maximum values). Curves represent the smooth trend (obtained with geom_smooth in R). WT panel are the same data as in Fig.2A with a different scaling for visualization.



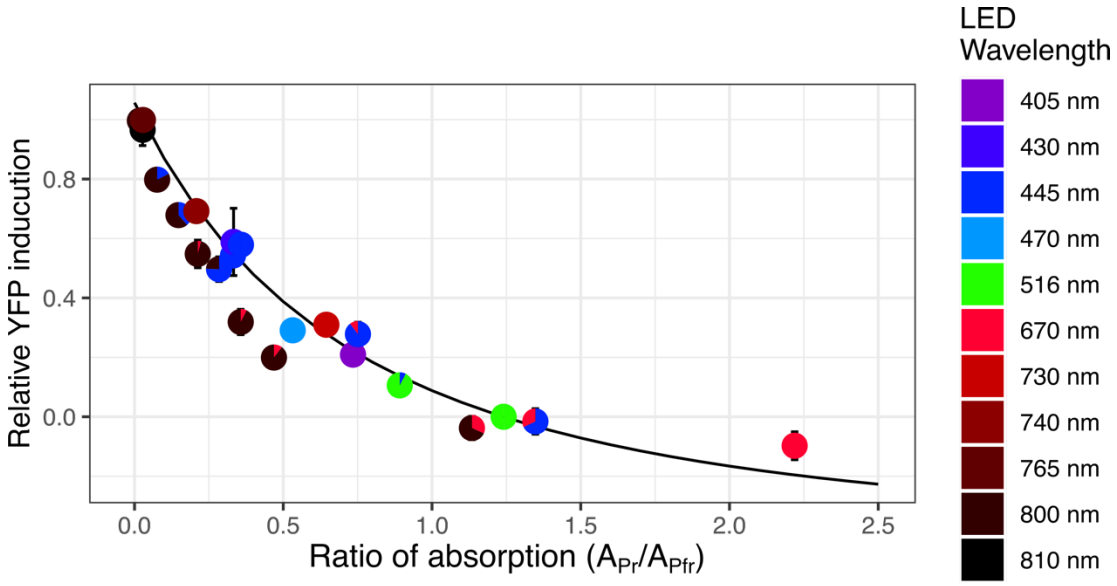
Extended Data Fig.5

Expression analysis of PtDPH-regulated (*HSF4.6a*, Phatr3_J15138, Phatr3_J18096, Phatr3_J45662, Phatr3_J46431) and PtDPH-non regulated (Phatr3_J18180) genes upon far-red and blue irradiation. *P. tricornutum* WT, DPH KO mutant and their corresponding non-mutated transformant (Tc) cells were collected in their continuous green light growth condition (G) and following a 30 min of irradiation with 445 nm LED at 30 µmol photon.m^{-2}.s^{-1} (B) or 800 nm LED at 200 µmol photon.m^{-2}.s^{-1}(FR) or kept in the dark for the same time (D). Gene expression quantifications were performed by RT-qPCR, with *H4* used as normalization and relativized to the green light condition. Values are the mean ±se on 3 independent biological replicates.



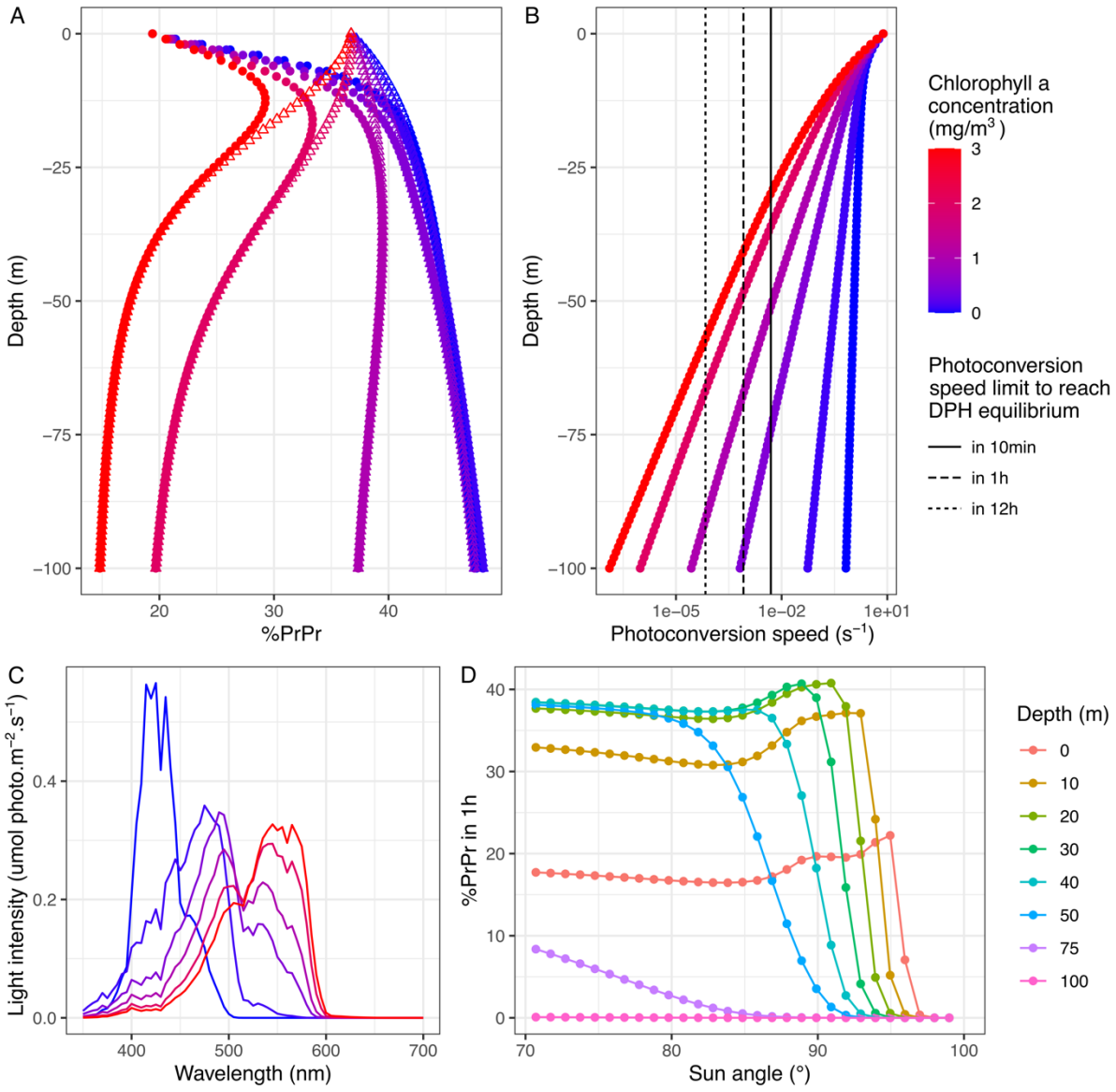
Extended Data Fig.6

TpDPH restores blue and far-red-dependent YFP induction in PtDPH KO line. PtDPH KO reporter line KO #1 was transformed with Tp-*DPH* gene under the control of Pt-*DPH* promoter and terminator, and subjected to a gradient of blue (430 nm) or far-red (765 nm) lights, according to the same experimental setup used for the induction spectra. The YFP increase compared to value in the dark were normalized to the maximal YFP increase, i.e. after exposure to saturating 765 nm light. The different symbols represent data for 2 independent lines.



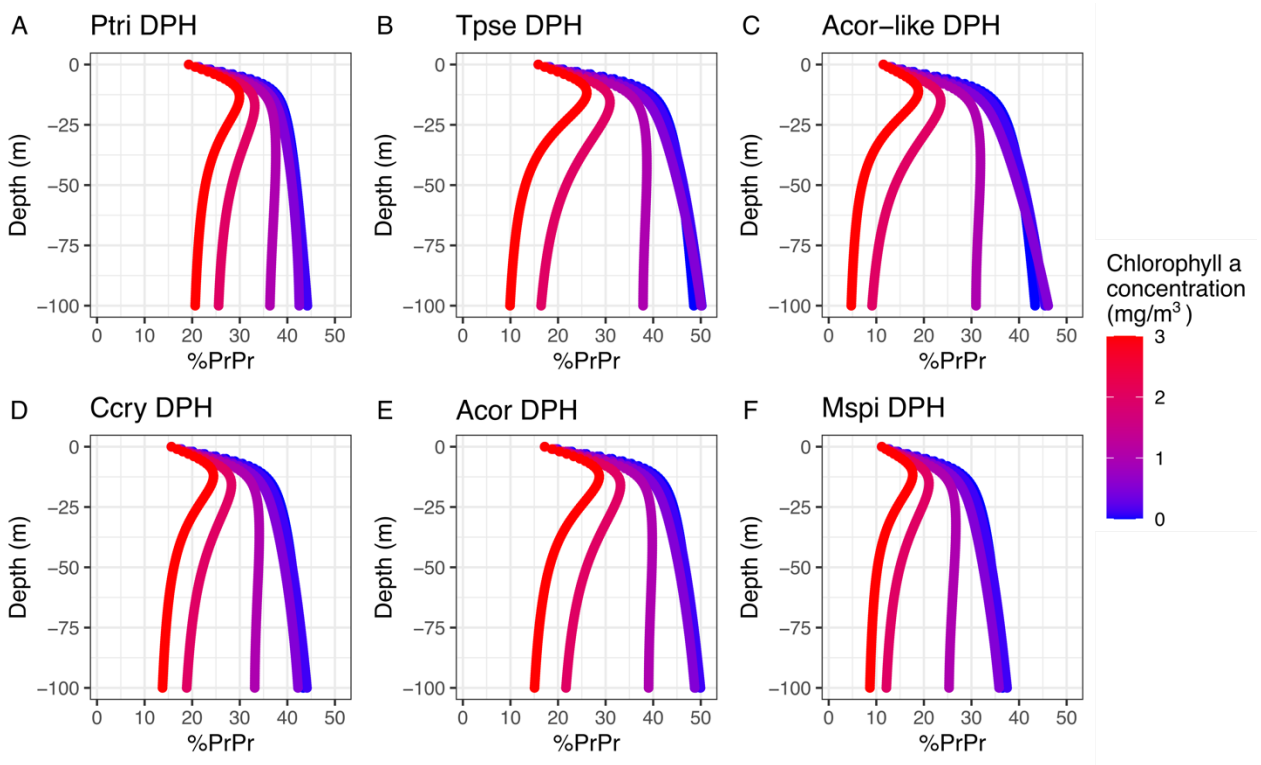
Extended Data Fig.7

PtDPH activity in response to bichromatic illuminations. Plot showing relative YFP signal at saturating monochromatic lights (same data as in Fig.3.C) and response to mixed wavelengths. The black curve indicates the η. Pie charts indicate the relative intensity of each LED color. The mixed illumination demonstrate that PtDPH can detect variation of the R/FR ratio, but also the FR/B, B/R and B/G ratios. Values are means ±se from the four independent WT lines



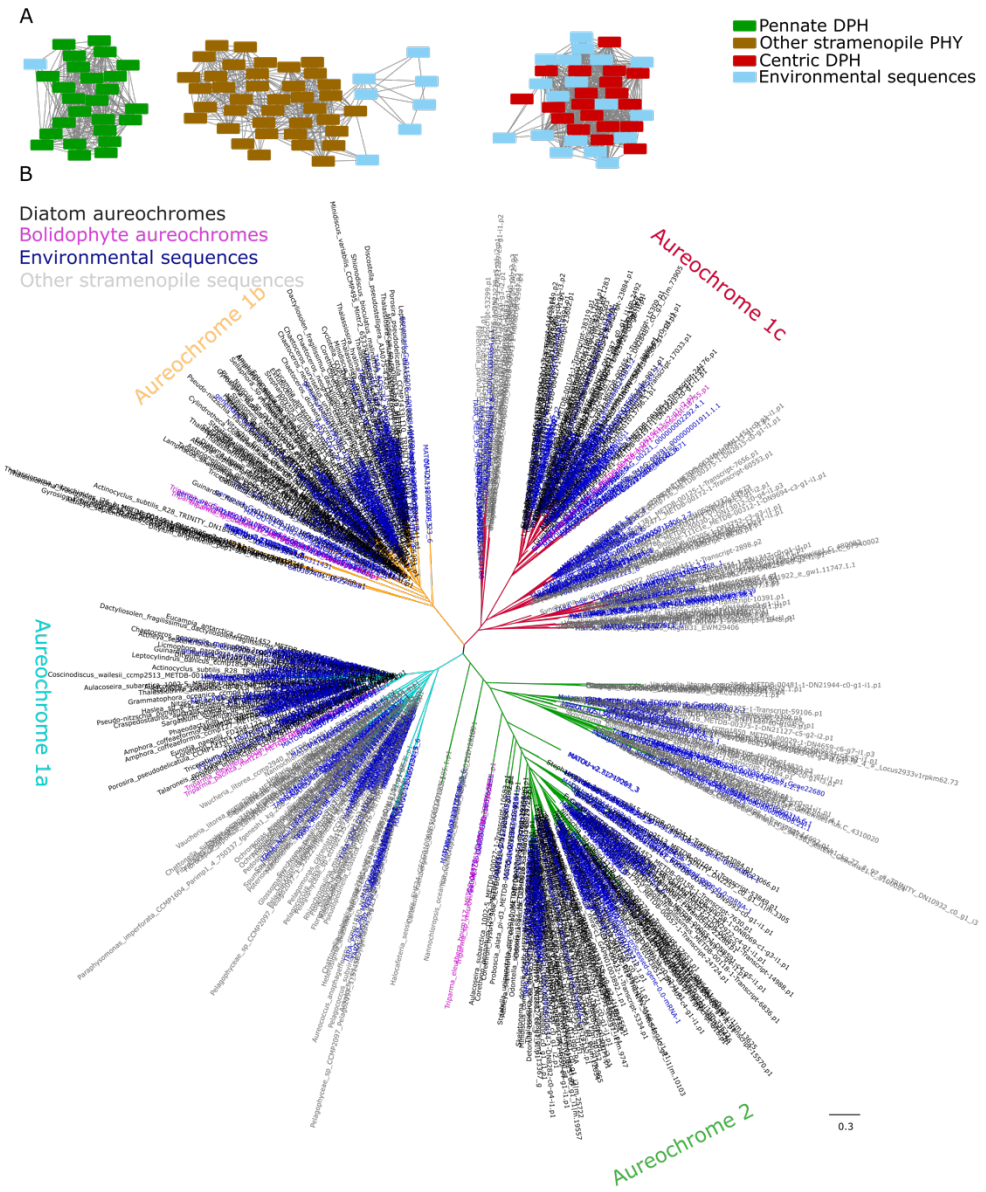
Extended Data Fig.7

PtDPH activity in modeled light fields (350 to 700 nm) mimicking different underwater scenarios of phytoplankton concentration: (A) Proportion of %PrPr at depth and contribution of the red band (removing the red band, i.e. 350 to 600nm spectra, triangles), and (B) the corresponding photoconversion rate constant. Vertical lines indicate the limit for DPH to reach equilibrium in 12 h, 1 h or 10 min (from left to right). (C) Modelled light fields at the bottom of the photic zone (1% of surface irradiance) for sun at zenith. Chlorophyll a scale is the same for A, B and C. (D) DPH photoequilibrium upon 1 h of exposure of light spectrum at different depths and different solar zenith angle (Chla 1 mg/m^3). Note the slight increase in %PrPr around 90° at 20 to 40 m deep.



Extended Data Fig.8

Projection of proportion of PrPr formed (%PrPr), in modeled light fields, with variations of chlorophyll a (Chl a) using properties of Ptri DPH (A), Tpse DPH (B), Acor-like synthetic DPH (C) Ccry DPH (D), Acor DPH (E) and Mspi DPH (F) determined *in vitro* from the recombinant protein absorption spectra (Extended Data Table2). The color legend is common to all panels. Note the influence of the absorption spectra (Pt DPH compared to Tp DPH) and the effect of the ratio of quantum yield (Acor DPH compared to Mspi DPH).



Extended Data Fig.9

Method for DPH and diatom AUREO protein search. (A**)** Example of Sequence Similarity Network of the DPH REC domain, searched in *Tara* Oceans data showing separation of the different phytochromes (centric diatom, pennate diatom, other stramenopiles) for an alignment score threshold of 20. Environmental sequences grouping with the centric or pennate DPH will be annotated as such. (B) Phylogenetic tree of diatom AUREO search. Diatom clades are identifiable (black label), delimited by the Bolidophyceae aureochromes (sister picoplanktonic group of diatoms, purple labels). Branches are colored by the different aureochrome types: aureochrome 1a (cyan), 1b (orange), 1c (red) and 2 (green).

Extended Data tables

Extended Data Table 1.

Species from which DPH have been characterized in this study and previous one* from *(6)* and information about their isolation site

Species	Strain	Sampling location, see also Data S1
*Phaeodactylum tricornutum**	CCMP2561	North Atlantic, off Blackpool, England
*Thalassiosira pseudonana**	CCMP1335	North Atlantic, Moriches Bay, Forge River, Long Island, New York USA
Cyclotella cryptica	CCMP332	North Atlantic, Martha's Vineyard, Massachusetts USA
Minidiscus spinulatus	RCC4659	Atlantic Ocean, English Channel, Brittany coast
Skeletonema costatum	RCC1617	Atlantic Ocean, English Channel, Brittany coast
Arcocellulus cornucervis	RCC2270	Arctic Ocean, Beaufort Sea, at 0m depth
Shionodiscus bioculatus	RCC1991	Arctic Ocean, Beaufort Sea, at 65m depth

Extended Data Table 2.

Spectral and photochemical properties of different recombinant DPH

Phytochrome	max ΔA in B (nm)	min ΔA in B (nm)	max ΔA in R (nm)	min ΔA in FR (nm)	η
A.cor	385	424	682	767	0.649
C.cry	384	424	677	768	0.766
Minidiscus-like	385	423	682	764	ND
Acor-like	384	424	678	768	0.855
S.cos	386	424	685	767	ND
M.spi	380	424	678	760	0.968
S.bio	386	424	685	746	ND
*T.pse**	384	425	683	768	0.678
*P.tri**	376	419	697	749	0.753

*absorption spectra from Fortunato et al, 2016 *(6)*. Max and min ΔA : maximum and minimum of the differential absorption spectra in the blue (B), red (R), and far-red (FR) bands. ND: not determined. η were calculated with method from *(27)*.

Extended Data Table 3.

Estimation of the ratio of quantum yield from *in vivo* data

Model	Active form	Saturating illuminations used in model	Average of models fitted for each strain		
			Mean estimated ratio of quantum yield	Std. Error	Average residual standard error
monomer	NA *	only monochromatic	1.256	0.271	0.07755
dimer	PrPr actif	only monochromatic	0.704	0.132	0.07732
dimer	PfrPfr actif	only monochromatic	3.728	0.542	0.07680
dimer	PrPr actif	both mono- and bi-chrmonatic	0.959	0.082	0.10387
dimer	PrPr actif	Monochromonatic and visible bichromonatic illumination (ie excluding FR mixes)	0.692	0.117	0.07875

*in a monomer model, the formula is exactly the same whether Pr or Pfr is the active from

Extended Data Table 4.

SSN parameters for the DPH and diatom aureochrome search in *Tara* Oceans data

Domain model	SSN alignment score threshold
Centric GAF	30
Centric PHY	10
Centric HisKA	20
Centric HATPase	20
Centric REC	20
Pennate GAF	30
Pennate PHY	10
Pennate HisKA	20
Pennate HATPase	20
Pennate REC	20
Diatom Aureochrome	70

Extended Data Table 5.

Sequences of primers used in this study

qPCR-Phatr3_45662FW	CGAGGGAGCTCGGTTTATGG
qPCR-Phatr3_45662RV	TGATGGGAACTGTTCTGCCC
qPCR-Phatr3_46431FW	GGTTTGCGAGTGCATTTGGT
qPCR-Phatr3_45662RV	TGTCAGCAACCTCATTCCCC
qPCR-Phatr3_18180FW	CCGGGAACGTAGGTTTGAT
qPCR-Phatr3_18180RV	CCGCGGCCAACATAGCAAG
qPCR-H4FW	AGGTCCTTCGCGACAATATC
qPCR-H4RV	ACGGAATCACGAATGACGTT
*Hsf4.6a*p_Fw0	TCTAGAGAGCTCGGATTTCGAATCTGTTTTGGGCA

*Hsf4.6a*p-*YFP*_RV	GATATCGGATCCTGTCAAAGGTTTAAGAGAATCGGC
*YFP-Hsf4.6a*p_Fw	AAACCTTTGACAGGATCCGATATCATGGTGAGCAAGGGCGAG
*FcpA*T_Rv0	TCTAGATGAAGACGAGCTAGTGTTATTCC
HO1xpET.HindIII.Fw	GCAAGCTTAGAAGGAGATATACATATGAG
HO1xpET.NotI.Rv	GCGCGGCCGCCTAGCCTTCGGAGGTGGCGAG
TpPHY.D.Fw	AGGCTGTCTCGTCTCGTCTCAGGTCTCAAGGTATGAGTGTCAAAAAGAGCAC
TpPHY.Sap1.Rv	CAATGTGTTTCGCATGAACAGCTCTCTCGCTTG
TpPHY.Sap1.Fw	CAAGCGAGAGAGCTGTTCATGCGAAACACATTG
TpPHY.BsaI.Rv	CAAGCTGTTTATCAGGGTCACCACCCCACGTGACAC
TpPHY.BsaI.Fw	GTGTCACGTGGGGTGGTGACCCTGATAAACAGCTTG
TpPHY.Sap2.Rv	GAGATCGAACAAGTGCTCCTCAATCTGAAGGTCGTATG
TpPHY.Sap2.Fw	CATACGACCTTCAGATTGAGGAGCACTTGTTCGATCTC
TpPHY.E.Rv	TGGTAATCTATGTATCCTGTTGGTCTCTAAGCTCATCGTTCATTTTTGTGAT
PrPtPHY.A.Fw	GGCTGTCTCGTCTCGTCTCAGGTCTCAGGAGCCCGGGGATATCGAAGATCC
PrPtPHY.C.Rv	TGGTAATCTATGTATCCTGGTGGTCTCGCATTTTTAAAGGCGTGGTTCCTTG
TrPtPHY.E.Fw	TCGTCTCGTCTCAGGTCTCAGCTTCATGGTCGTTCATTCATAGAAG
TrPtPHY.F.Rv	TGGTAATCTATGTATCCTGGTGGTCTCAAGCGCGCTCTTTCCACCTCATCTC
UNS1FL.FW	CATTACTCGCATCCATTCTCAGGCTGTCTCGTCTCGTCTC
UNSXFL.RV	GGTGGAAGGGCTCGGAGTTGTGGTAATCTATGTATCCTGG
SbioPHY.NheI.Fw	GCGGCTAGCATGTCTGCCAGTTCCACCAC
SbioPHY_PCD.SalI.Rv	GCGGTCGACCTAAAGATTTTCCTTTTGATCTTTG
AcorPHY.NheI.Fw	GCGGCTAGCATGTCGGCACCTGCGGCAGC
AcorPHY_PCD.SacI.Rv	GCGGAGCTCCTAGTAGCTTGTTGTGTTCTCGTC
MspiPHY.NheI.Fw	GCGGCTAGCATGACCTCCTCCTCAACCAAC
MspiPHY_PCD.SacI.Rv	GCGGAGCTCCTACTTCTGATCGGCAATCAATTC
ScosPHY.SpeI.Fw	GCGACTAGTATGTCGTCCACCAATAGCAC
ScosPHY_PCD.SacI.Rv	GCGGAGCTCCTAGAGGTTTTCCTTTTGATCTTG
CcryPHY.NheI.Fw	GCGGCTAGCATGGCAGCACCCCACAAAAC
CcryPHY_PCD.SacI.Rv	GCGGAGCTCCTACTTCTGATCTTTAATCAAATC