Supporting Information

Grébert et al. 10.1073/pnas.1717069115

SI Materials and Methods

For the modeling of the blue-to-green irradiance ratio ($Irr_{495:545}$) at *Tara* Oceans stations, we used the clear sky surface irradiance model of Frouin and McPherson in Fortran and translated to Matlab by Werdell [see Frouin et al. (1) and Tanre et al. (2) for the analytical formula used] using the date and latitude and longitude of each station, assuming sunny sky and at noon.

The spectral light distribution averaged over the mixed layer was computed from the following:

$$< Ir(\lambda) > = \frac{\int_0^{MLD} E(\lambda, 0^-)\, e^{-k(\lambda, chl)z} dz}{\text{MLD}} = \frac{I(\lambda, 0^-)}{\text{MLD}\; k(\lambda, chl)} \left\{1 - e^{-k(chl)\text{MLD}}\right\},$$

where *chl* denotes the average chlorophyll value in the mixed layer, MLD is the mixed layer depth that was computed based on a temperature threshold criterion, $k(\lambda, chl)$ is the diffuse attenuation coefficient at wavelength λ (495 or 545 nm using a 10-nm bandwidth), and [*chl*] was based on a fluorometer that was calibrated against HPLC data and corrected for nonphotochemical quenching. This parameter was computed using the equation of Morel and Maritorena (3):

$$k(\lambda, chl) = k_w(\lambda) + \chi(\lambda)[chl]^{e(\lambda)}.$$

k_w, χ, and e are provided in table 2 of Morel and Maritorena (3) and have the following values for the wavelengths of interest:

Wavelength, nm	$k_w(\lambda)$, m^{-1}	$\chi(\lambda)$	$e(\lambda)$
495	0.01885	0.06907	0.68947
545	0.05212	0.04253	0.65591

If the sampling depth was below the MLD, the irradiance was computed as follows:

$$Ir(\lambda, \text{sampling depth}) = (\lambda, 0^-)\, e^{-k(\lambda, chl)\text{sampling depth}}.$$

The ratio was then computed as $Irr_{495:545}$.

1. Frouin R, Ligner DW, Gautier C (1989) A simple analytical formula to compute clear sky total and photosynthetically available solar irradiance CC at the ocean surface. *J Geophys Res* 94:9731–9742.
2. Tanre D, Herman M, Deschamps P-Y, De Leffe A (1979) Atmospheric modeling for space measurements of ground reflectances, including bi-directional properties. *Appl Optics* 18:3587–3594.
3. Morel A, Maritorena S (2001) Bio-optical properties of oceanic waters: A reappraisal. *J Geophys Res* 106:7163–7180.
4. Six C, et al. (2007) Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: A comparative genomics study. *Genome Biol* 8:R259.
5. Humily F, et al. (2013) A gene island with two possible configurations is involved in chromatic acclimation in marine *Synechococcus*. *PLoS One* 8:e84459.

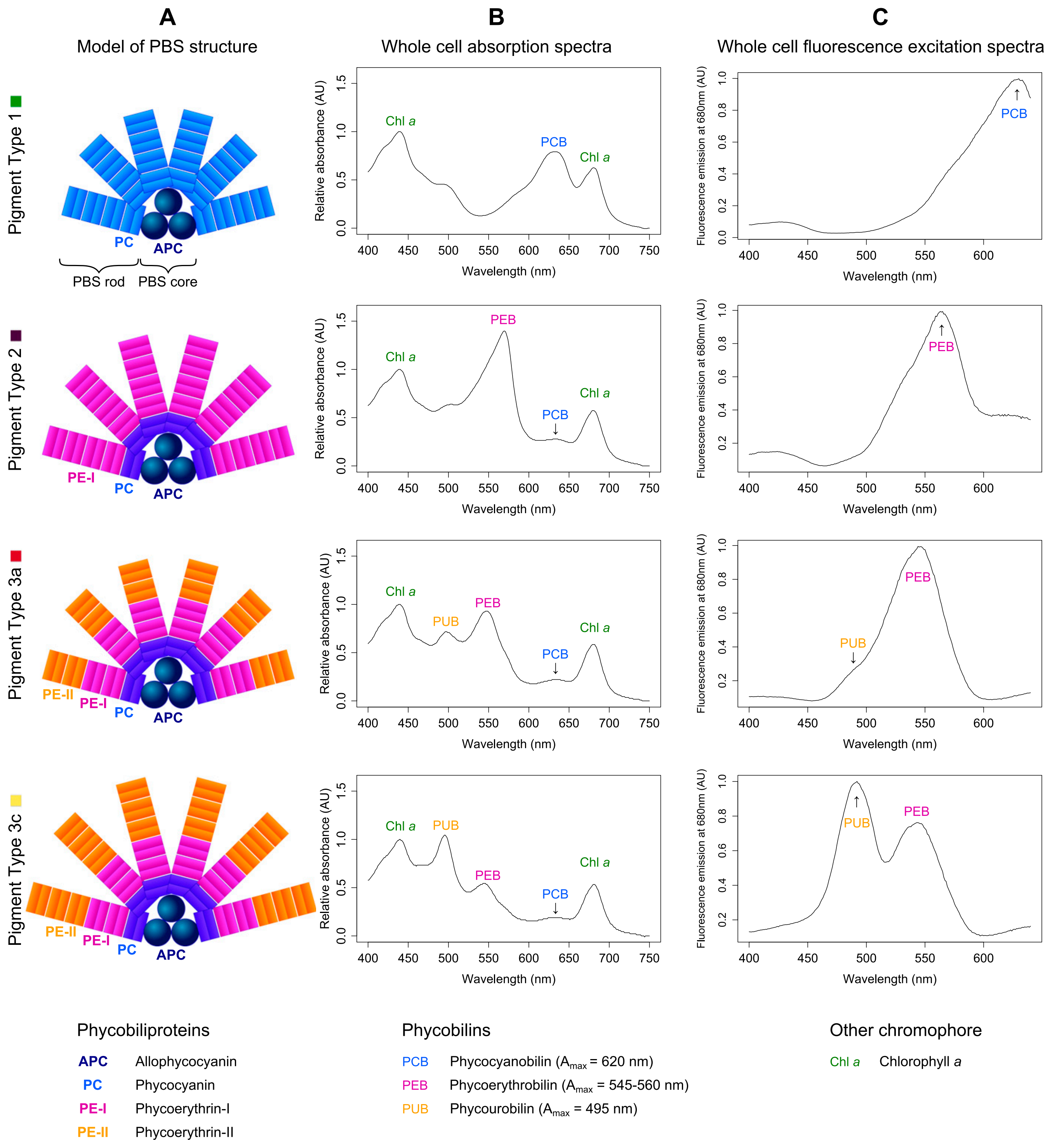

Fig. S1. Biochemical composition and biooptical properties of phycobilisomes (PBSs) of the main *Synechococcus* pigment types (PTs). (*A*) Models of PBS structure, highlighting the conserved core and variable rods of increasing complexity from PT1 to PT3 [redrawn after Six et al. (4)]. (*B*) Whole-cell absorption spectra of the different PTs [reproduced after Six et al. (4)]. Chromophores responsible of each absorption peaks are indicated. (*C*) Whole-cell fluorescence excitation spectra with emission at 680 nm. Note that, for chromatic acclimaters (PT 3d), the PBS structure is similar to other PT 3 but that the excitation ratio at 495 and 545 nm ($Ex_{495:545}$) varies from 0.6 in green light to 1.6 in blue light (5).

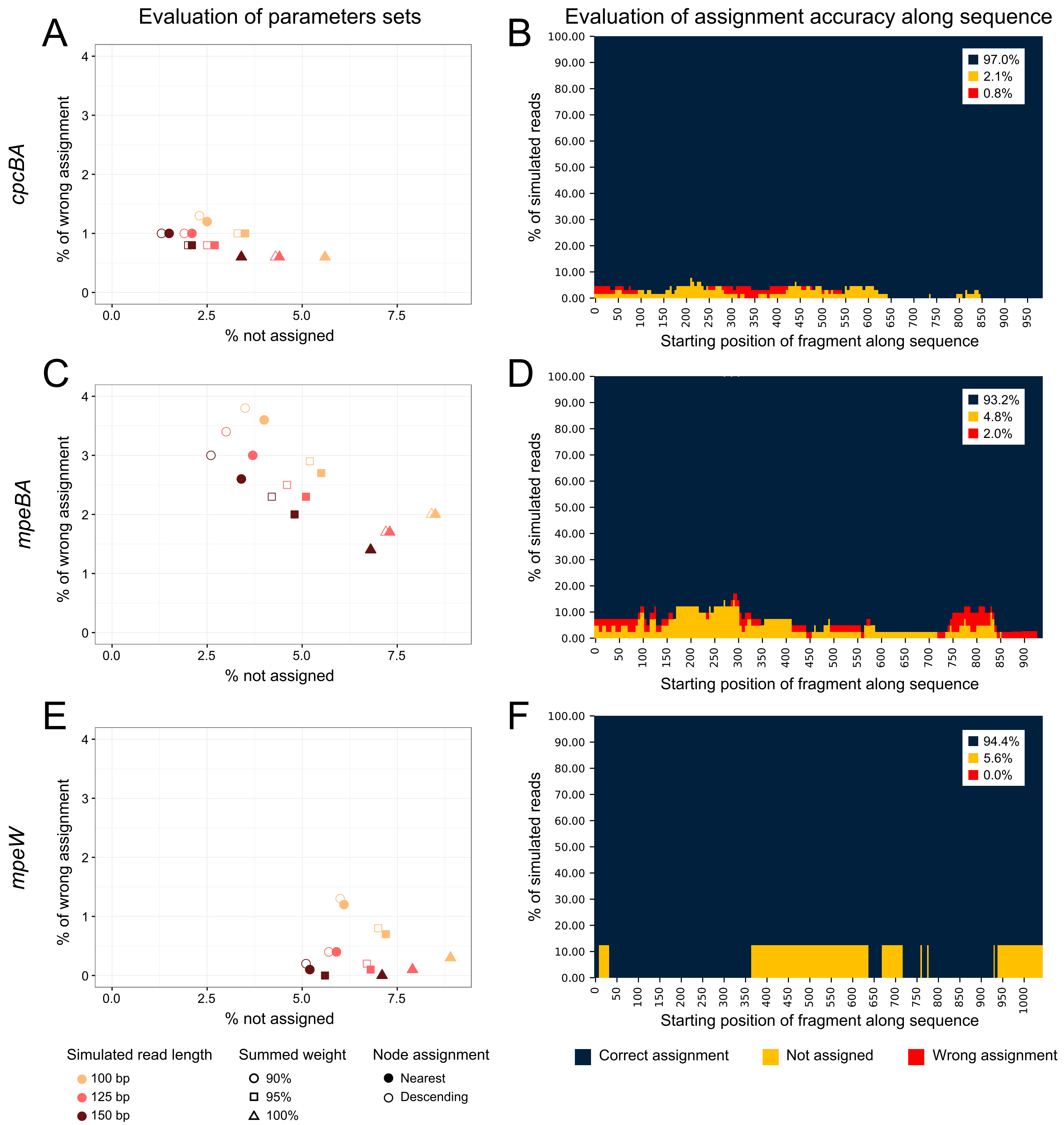

Fig. S2. Evaluation of the assignment pipeline and the resolution power of the different markers used in this study. Simulated reads were generated from the reference dataset and assigned using a custom-designed pipeline (*Materials and Methods*). (*A*, *C*, and *E*) Evaluation of different sets of parameters tested for read assignment for the different markers: *cpcBA* (*A*), *mpeBA* (*C*), and *mpeW* (*E*). The 100- (yellow), 125- (pink), and 150-bp-long (dark red) reads were simulated. For each read, pplacer returns a list of possible positions in the tree, each associated with a likelihood weight. From these placements, we considered only those that reached a summed likelihood weight of either 90% (circle), 95% (square), or 100% (triangle). The assignment was then performed based on the phenotype of either the nearest node (solid symbol) in the tree or the descending (child) node (empty symbol). (*B*, *D*, and *F*) Evaluation of the resolution power along *cpcBA* (*B*), *mpeBA* (*D*), or *mpeW* (*F*) for 150-bp simulated reads assigned using the parameters selected for *Tara* Oceans metagenomic read assignment (i.e., nearest-node assignment and summed weight of 95%). Note that *Tara* Oceans reads had a mean length of 164 bp.

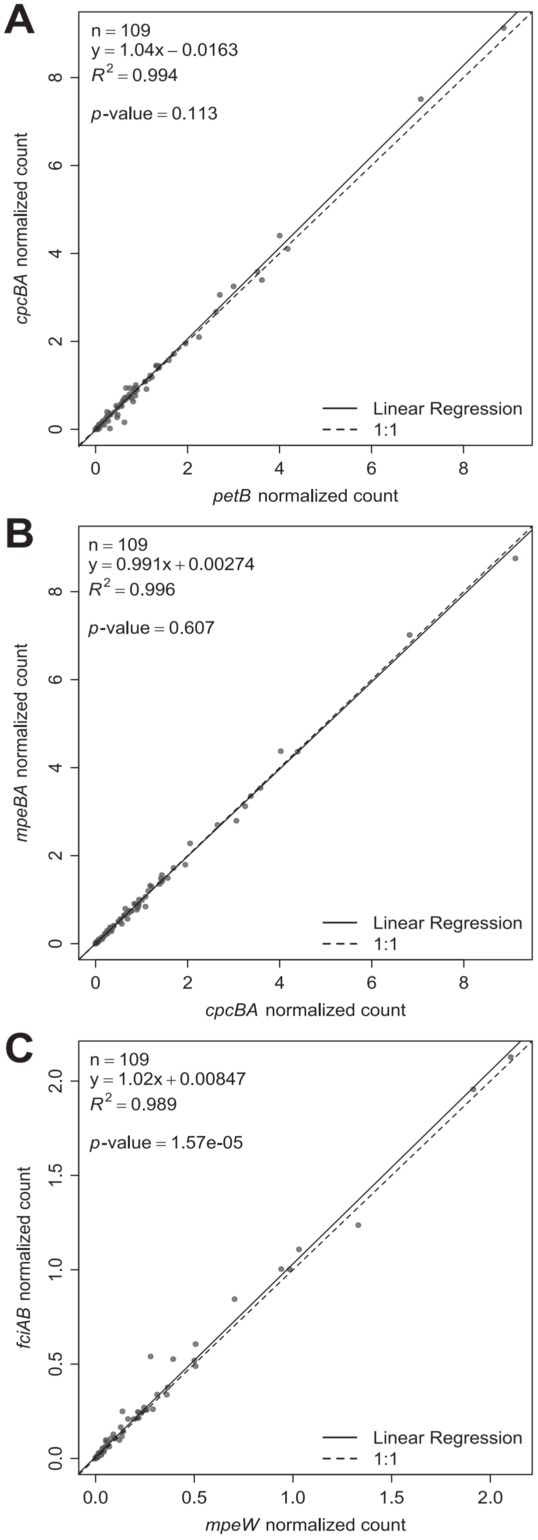

Fig. S3. Correlations between the number of reads recruited using the main markers used in this study. (*A*) Correlation between *petB* (vertical phylogeny) and *cpcBA* counts used to discriminate pigment types (PTs) 1, 2, and 3. (*B*) Correlation between PT 3 counts using *cpcBA* and total *mpeBA* counts. Note that *mpeBA* is a PT3-specific marker and is used to discriminate PTs 3a, 3dA, 3f, and 3c + 3dB. (*C*) Correlation between PT 3dB counts using *fciAB*, a PT 3dB- and 3dA-specific marker and total *mpeW* counts, a PT 3dB-specific marker.

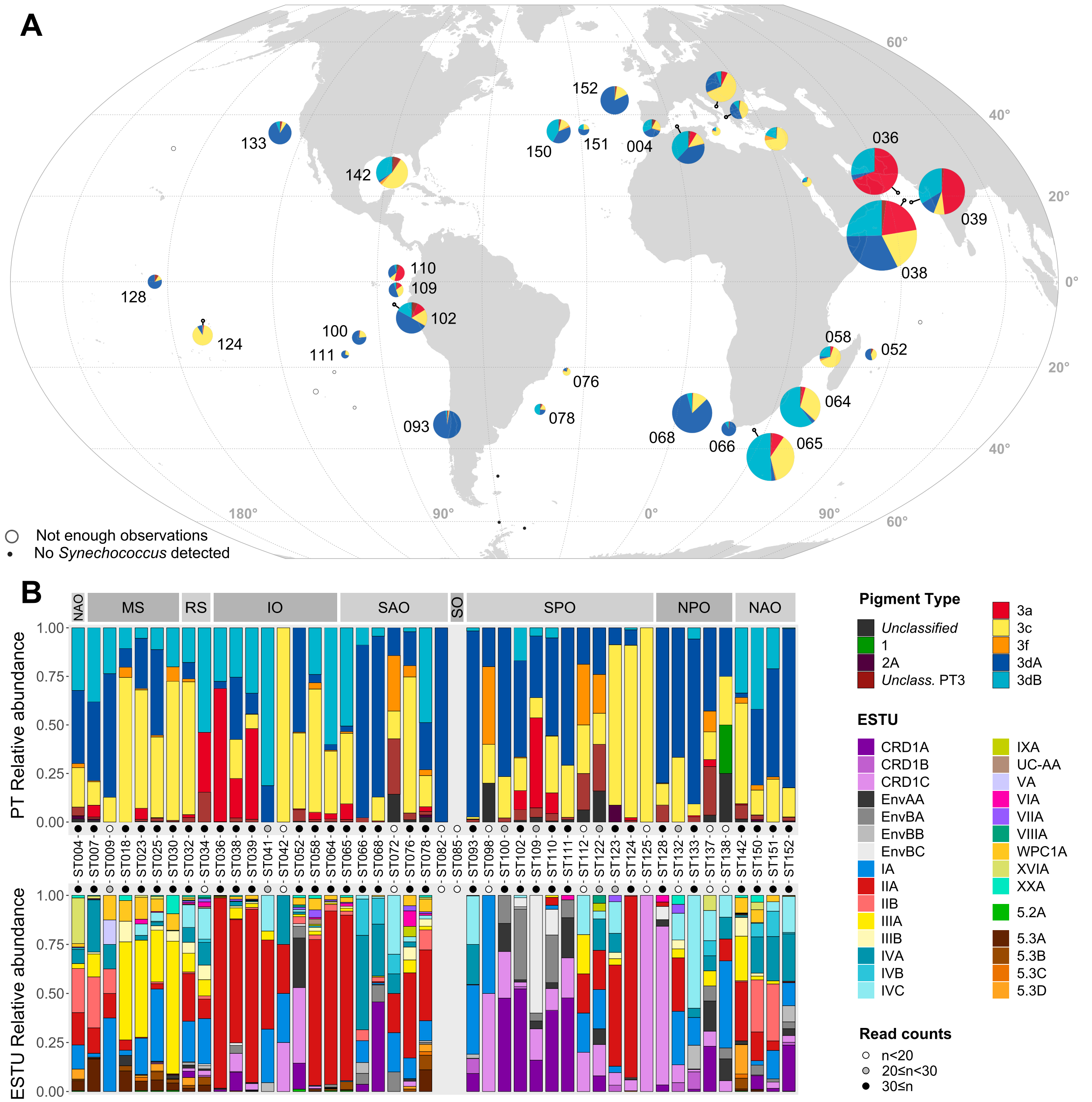

Fig. S4. Distribution of *Synechococcus* pigment types (PTs) at depth (deep chlorophyll maximum). (*A*) Map showing the global distribution of all *Synechococcus* PTs at depth along the *Tara* Oceans transect. Diameters of pies are proportional to the number of *cpcBA* reads normalized by the sequencing effort. Stations with less than 30 *cpcBA* or *mpeBA* reads are indicated by open circles, and those with no *cpcBA* reads by black dots. Numbers next to pies correspond to *Tara* Oceans stations. (*B*) PTs and ESTU relative abundance at depth for sampling station along the *Tara* Oceans transect. Oceanic provinces are indicated in the *Top* gray panels. IO, Indian Ocean; MS, Mediterranean Sea; NAO, North Atlantic Ocean; NPO, North Pacific Ocean; RS, Red Sea; SAO, South Atlantic Ocean; SO, Southern Ocean; SPO, South Pacific Ocean.

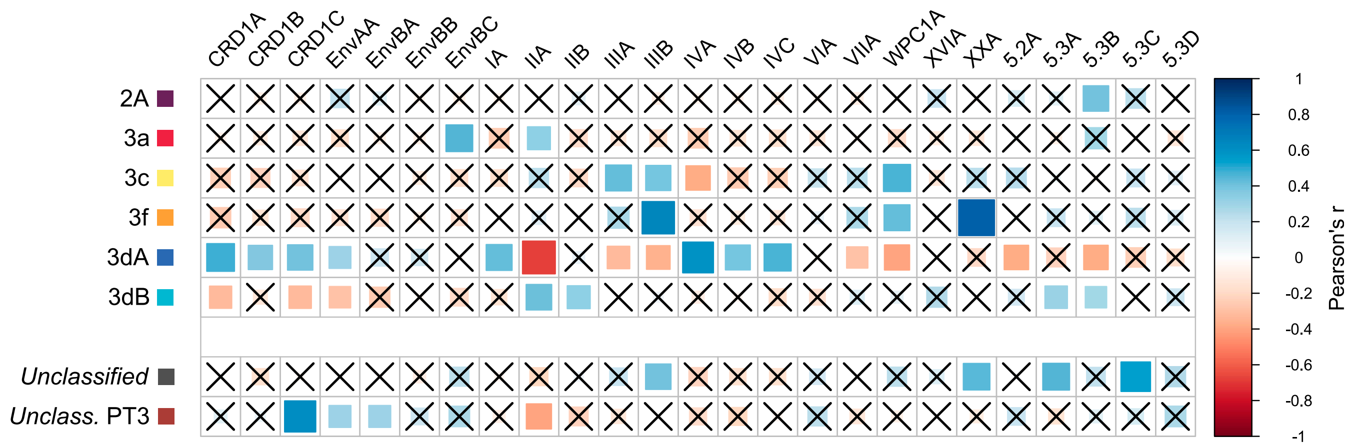

Fig. S5. Same as Fig. 4*A*, but for all ESTUs. Unclass., unclassified.

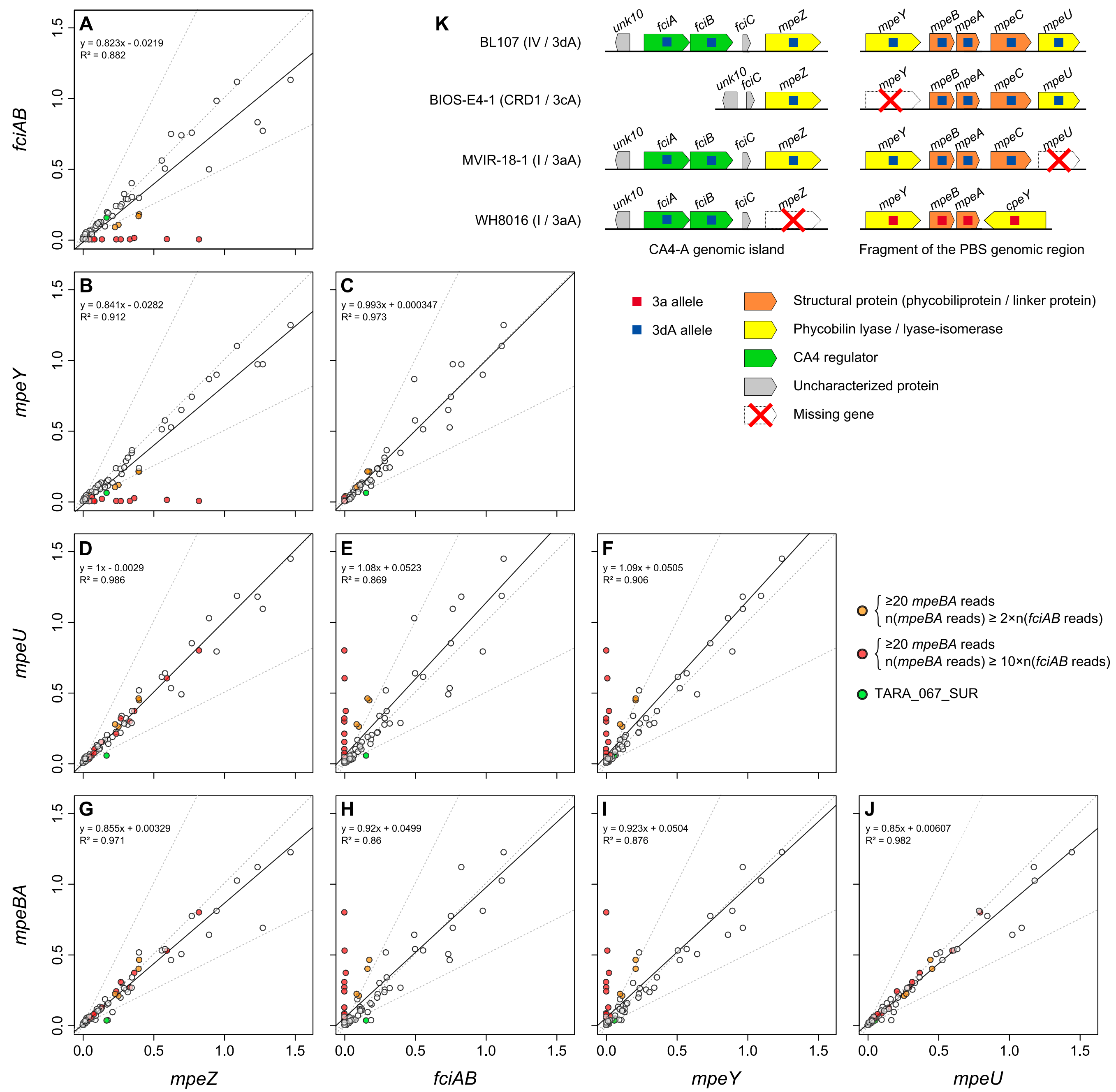

Fig. S6. Focus on pigment type (PT) 3dA natural mutants, exhibiting an altered gene content with regard to typical PT 3dA. (*A*–*J*) Correlation between the number of reads assigned as PT 3dA using different markers (all present in single gene copy in typical 3dA). Each circle corresponds to a *Tara* Oceans station and depth. Orange circles: stations with at least 20 *mpeBA* reads assigned to PT 3dA and at least twice more 3dA counted with *mpeBA* than with *fciAB*, corresponding to the surface sample of stations TARA_070, TARA_110, and TARA_137, and the DCM of stations TARA_038, TARA_058, and TARA_110. Red circles: same but with more than 10-fold 3dA counted with *mpeBA* than *fciAB*, corresponding to the surface sample of stations TARA_052, TARA_094, TARA_111, and TARA_122 to TARA_128, and DCM of stations TARA_052, TARA_100, TARA_111, and TARA_128. Green circle: surface of station TARA_067. (*K*) CA4-A genomic island and fragment of the phycobilisome (PBS) genomic region for a typical, CA4-able 3dA strain (strain BL107), and three CA4-deficient strains, which are stuck either in blue-light phenotype (similar to strain BIOS-E4-1) or green-light phenotype (as strains MVIR-18-1 and WH8016). Note that KORDI-49 and WH8016 strains have identical PBS gene complement and genomic arrangement. The complete PBS genomic region of the BL107 strain can be found in Six et al. (4). Note that, for readability, surface of station TARA_093 has been omitted since it has the highest normalized counts (2.7–3.2) for all markers and exhibited a good agreement between markers (ratio close to 1:1).

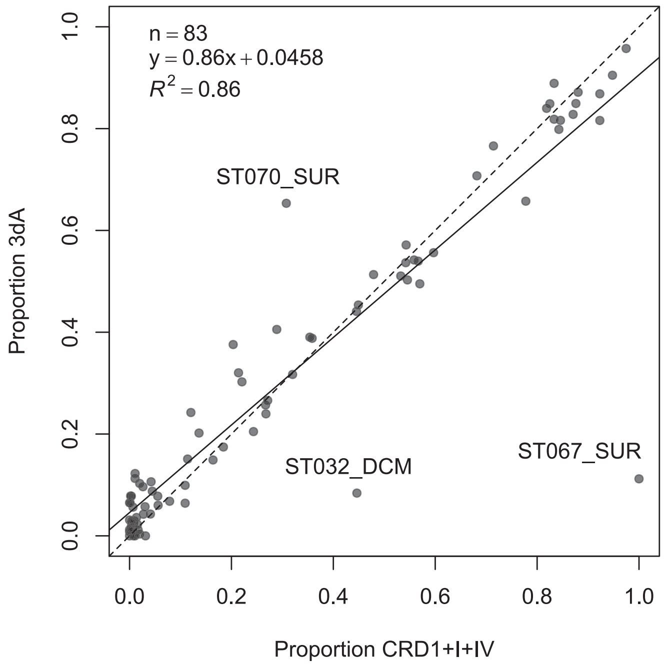

Fig. S7. Correlation between the proportion of clades I, IV, and CRD1, as assessed with *petB*, and the proportion of pigment type 3dA, as assessed with *mpeBA*, at each station.

Other Supporting Information Files

Dataset S1 (XLSX)
Dataset S2 (XLSX)
Dataset S3 (XLSX)
Dataset S4 (XLSX)