Identification of a dual orange/far-red and blue light photoreceptor from an oceanic green picoplankton

Yuko Makita1,13, Shigekatsu Suzuki2,13, Keiji Fushimi3,4,5,13, Setsuko Shimada1,13, Aya Suehisa1, Manami Hirata1, Tomoko Kuriyama1, Yukio Kurihara1, Hidefumi Hamasaki1,6, Emiko Okubo-Kurihara1, Kazutoshi Yoshitake7, Tsuyoshi Watanabe8, Masaaki Sakuta9, Takashi Gojobori10, Tomoko Sakami11, Rei Narikawa3,4,5,12, Haruyo Yamaguchi2, Masanobu Kawachi2 & Minami Matsui1,6

Photoreceptors are conserved in green algae to land plants and regulate various developmental stages. In the ocean, blue light penetrates deeper than red light, and blue-light sensing is key to adapting to marine environments. Here, a search for blue-light photoreceptors in the marine metagenome uncover a chimeric gene composed of a phytochrome and a cryptochrome (Dualchrome1, DUC1) in a prasinophyte, Pycnococcus provasolii. DUC1 detects light within the orange/far-red and blue spectra, and acts as a dual photoreceptor. Analyses of its genome reveal the possible mechanisms of light adaptation. Genes for the light-harvesting complex (LHC) are duplicated and transcriptionally regulated under monochromatic orange/blue light, suggesting P. provasolii has acquired environmental adaptability to a wide range of light spectra and intensities.

1 Synthetic Genomics Research Group, RIKEN Center for Sustainable Resource Science, Yokohama, Japan. 2 Biodiversity Division, National Institute for Environmental Studies, Tsukuba, Japan. 3 Graduate School of Integrated Science and Technology, Shizuoka University, Shizuoka, Japan. 4 Research Institute of Green Science and Technology, Shizuoka University, Shizuoka, Japan. 5 Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Saitama, Japan. 6 Yokohama City University, Kihara Institute for Biological Research, Yokohama, Japan. 7 Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan. 8 Fisheries Resources Institute, Japan Fisheries Research and Education Agency, Kushiro, Hokkaido, Japan. 9 Department of Biological Sciences, Ochanomizu University, Tokyo, Japan. 10 Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia. 11 Fisheries Resources Institute, Japan Fisheries Research and Education Agency, Minami-ise, Mie, Japan. 12 Department of Biological Sciences, Graduate School of Science, Tokyo Metropolitan University, Tokyo, Japan. 13 These authors contributed equally: Yuko Makita, Shigekatsu Suzuki, Keiji Fushimi, Setsuko Shimada. ✉email: minami@riken.jp
Photosynthetic organisms utilize various wavelengths of light, not only as sources of energy but also as clues to assess their environmental conditions. Blue light penetrates deeper into the ocean, whereas red light is absorbed and immediately decreases at the surface. Oceanic red algae possess blue-light receptor crytochromes (CRYs) but not red-light receptor phytochromes (PHYs). Similarly, most chlorophytes have CRYs but fewer have PHYs. PHYs are bilin-containing photoreceptors for the red/far-red-light response. Interestingly, algal PHYs are not limited to red and far-red responses. Instead, different algal PHYs can sense orange, green, and even blue light. They have the ability to photosense between red-absorbing Pr and far-red-absorbing Pfr, and this conformational change enables interactions with signaling partners. CRY is a photolyase-like flavoprotein and widely distributed in bacteria, fungi, animals and plants.

In 2012–2014, large-scale metagenome analyses were performed in Sendai Bay, Japan, and the western subarctic Pacific Ocean after the Great East Japan Earthquake to monitor its effects on the ocean (http://marine-meta.healthscience.waseda.ac.jp/crest/metacrest/graphs/). These metagenome analyses targeted the pycnocline 5. In 2012–2014, large-scale metagenome analyses were performed in Sendai Bay, Japan, and the western subarctic Pacific Ocean after the Great East Japan Earthquake to monitor its effects on the ocean (http://marine-meta.healthscience.waseda.ac.jp/crest/metacrest/graphs/). These metagenome analyses targeted the candidate genes identified in the pycnocline 5. We searched the marine metagenome data with cryptochrome (CRY) regions. We sequence the genome of P. provasolii (described later), which belongs to the PcyA cluster, and explain the environmental adaptation and success of P. provasolii under various light qualities and intensities. Thus, prasinophytes mainly inhabit marine environments and are dominant algae under various light qualities and intensities. Prasinophytes have a Cys residue either within the GAF domain or in the N-terminal loop region to covalently bind to the bilin chromophore (Supplementary Fig. 5b and Supplementary Fig. 6c) and undergo dual photoconversions at both the PHY and CRY domains (Supplementary Fig. 5b and Supplementary Table 1). Dark reversion from the photoproduct state (or excited state) of the PHYs (Supplementary Fig. 5a) shows light-induced isomerization that triggers a reversible photocycle between the dark state (or ground state) and the photoproduction state (or excited state) of the PHYs (Supplementary Fig. 5a). The GAF Cys ligates to C3 of phycocyanobilin (PCB) or phytochromobilin (PBΦ), whereas the N-terminal Cys ligates to C3 of biliverdin IXα (BV). As the PP PHY possesses the GAF Cys residue but not the N-terminal Cys, the binding chromophore is likely to be PCB or PBΦ. Furthermore, we detected a bilin reductase homolog from the P. provasolii NIES-2893 genome (described later), which belongs to the PcsA cluster, producing PCB from BV, but not to the HY2 cluster, producing PBΦ from BV (Supplementary Table 1). Taken together, we presumed that the native chromophore for PP PHY is PCB and not PBΦ. Therefore, we expressed PP PHY in E. coli harboring the PCB-synthetic system (C41 pKT271). The purified PP PHY covalently bound PCB (Supplementary Figs. 5b and 6c) and showed reversible photoconversion between an orange-absorbing (Pr) form (λmax, 612 nm) in the dark state and a far-red-absorbing (Pfr) form (λmax, 702 nm) in the photoproduction state (Fig. 2b and Supplementary Table 1). Dark reversion from the photoproduction state to the dark state was not seen (Supplementary Fig. 7a). Furthermore, we observed that PBΦ but not BV could be efficiently incorporated into the apo-PP PHY to show reversible photoconversion by using the PBΦ- and BV-synthetic systems, which is consistent with previous studies (Supplementary

Results

Identification of photoreceptor DUC1 from marine metagenome data. Since CRYs have been reported in several chlorophytes, we searched the marine metagenome data with cryptochrome PHR and FAD as baits to target CRY genes. We called 542 assembled metagenome sequences with the PHR, FAD and Rossmann-like_α/b/α_fold domains, and used a combination of these three domains as a hallmark for CRY candidates. Among the candidate genes identified, we found one with PHY-like sequence at its N-terminal region similar to Arabidopsis phytochrome B (PHYB) (Fig. 1). This gene has no introns and can encode angiosperm PHY domains at its N-terminus and CRY domains at its C-terminus (Fig. 1a). In metagenome data from four collection points, fragments of this gene were detected in open sea areas (C12 and A21 points) at both the Sendai Bay and A-line sampling stations (Supplementary Fig. 1a). To find the host plankton or its relative, we searched for a similar sequence in MMETSP (The Marine Microbial Eukaryote Transcriptome Sequencing Project). We found transcriptome fragments that matched with 100% identity to this gene, although we did not isolate the boundary sequence between the PHY and CRY regions. We identified the host plankton as a marine prasinophyte alga, P. provasolii. Fortunately, a culture strain of it was stocked in the NIES collection as P. provasolii NIES-2893 (Fig. 1b). Metagenome data also supported the fact that there was enrichment of this species in the subsurface chlorophyll maximum layer of stations C12 and A21 (Supplementary Fig. 1, Supplementary Note 1, Supplementary Method 1). After amplification of this gene from NIES-2893 by RT-PCR, we finally concluded that it does have both the PHY and CRY domains (5073 bp, 183 kDa protein) and designated it as Dualochrome1 or DUC1 (Supplementary Fig. 2a).

Phylogenetic analysis using the P. provasolii PHY (PpPHY) and CRY (PpCRY) domains showed they branched with those of other chlorophytes (Supplementary Figs. 3 and 4), which is consistent with the phylogenetic position of this species based on the 18S rRNA gene. Additionally, we found that three other strains of P. provasolii and Pseudoscurfella marina NIES-1419, which is a close relative of P. provasolii, also possess the DUC1 gene. P. marina DUC1 showed 98.2% amino acid identity with the P. provasolii DUC1 (PpDUC1) (Supplementary Fig. 2b, c).

PpDUC1 senses blue, orange, and far-red light. To verify whether DUC1 possesses photosensing activity, we expressed the PHY (PP PHY, 70.2 kDa, amino acid positions 1–662 in PpDUC1) and CRY (PP CRY, 72.2 kDa, amino acid positions 1039–1690 in PpDUC1) regions with His-tag and GST-tag, separately (Fig. 2a). PHY's have a Cys residue either within the GAF domain or in the N-terminal loop region to covalently bind to the bilin chromophore (Supplementary Fig. 5a). The bilin chromophore shows light-induced Z/E isomerization that triggers a reversible photocycle between the dark state (or ground state) and the photoproduction state (or excited state) of the PHYs (Supplementary Fig. 5a). The GAF Cys ligates to C3 of phycocyanobilin (PCB) or phytochromobilin (PBΦ), whereas the N-terminal Cys ligates to C3 of biliverdin IXα (BV). As the PP PHY possesses the GAF Cys residue but not the N-terminal Cys, the binding chromophore is likely to be PCB or PBΦ. Furthermore, we detected a bilin reductase homolog from the P. provasolii NIES-2893 genome (described later), which belongs to the PcsA cluster, producing PCB from BV, but not to the HY2 cluster, producing PBΦ from BV (Supplementary Table 1). Taken together, we presumed that the native chromophore for PP PHY is PCB and not PBΦ. Therefore, we expressed PP PHY in E. coli harboring the PBΦ-synthetic system (C41 pKT271). The purified PP PHY covalently bound PCB (Supplementary Figs. 5b and 6c) and showed reversible photoconversion between an orange-absorbing (Po) form (λmax, 612 nm) in the dark state and a far-red-absorbing (Pfr) form (λmax, 702 nm) in the photoproduction state (Fig. 2b and Supplementary Table 1). Dark reversion from the photoproduction state to the dark state was not seen (Supplementary Fig. 7a). Furthermore, we observed that PBΦ but not BV could be efficiently incorporated into the apo-PP PHY to show reversible photoconversion by using the PBΦ- and BV-synthetic systems, which is consistent with previous studies (Supplementary
Taking the presence of the PcyA homolog into consideration, it is likely that PpPHY reversibly senses orange and far-red light via PCB incorporation in vivo.

CRYs non-covalently bind to a flavin chromophore, in many cases, flavin adenine dinucleotide (FAD) (Supplementary Fig. 5c). Some redox state forms of FAD, such as oxidized FAD (FADOX), anion radical FAD (FAD•–), neutral radical FAD (FADH•), and reduced FAD (FADredH•), are observed in the photocycle of CRYs (Supplementary Fig. 5c)12. As the PpCRY accumulates as insoluble inclusion bodies in the normal expression system, we expressed PpCRY in E. coli harboring the chaperon co-expression system (C41_pG-KJE8)13. The purified PpCRY non-covalently bound FAD (Supplementary Figs. 5d and 6c). The protein showed a blue-absorbing (Pb) form (λmax, 447 nm) before light irradiation (Fig. 2c and Supplementary Table 1). The spectral form in the dark state was similar to that of the FADOX-bound CRYs12. Blue-light irradiation resulted in conversion to a UV-absorbing (Puv) form (λmax, 366 and 399 nm) (Fig. 2c and Supplementary Table 1). The spectral form in the photoproduct state was similar to that of the FAD•–-bound CRYs12. Although we further irradiated the protein with UV-to-blue light, no spectral change was observed. Instead, Puv-to-Pb dark reversion occurred (Supplementary Fig. 7b).

To conclude, PpCRY showed photoconversion from the FADOX-bound form to the FAD•–-bound one, and the reverse reaction occurred in the dark. This is similar to the photocycle of dCRY, a photoreceptor for circadian clock regulation in Drosophila melanogaster, rather than CraCRY from the green algae Chlamydomonas reinhardtii and AtCRY1 and 2 from the land plant Arabidopsis thaliana16.

These results suggest that PpDUC1 is a broadband light sensor that can detect long-wavelength light (i.e., orange to far-red) in the PpPHY region and short-wavelength light (i.e., UV to blue) in the PpCRY region.

PpDUC1 mainly localizes at nucleus in tobacco leaves. It is reported that light conditions change the intracellular localization of photoreceptors in higher plants and Micromonas pusilla.
We examined PpDUC1 localization using tobacco (Nicotiana benthamiana) cells. PpDUC1::GFP, as well as the PHY region of PpDUC1 (PpPHY::GFP) and the CRY region of PpDUC1 (PpCRY::GFP), were introduced into tobacco cells (Fig. 3a). Expression of the GFP-fused proteins was confirmed by protein-blot analysis (Fig. 3b).

Arabidopsis phyB is reported to localize in the nucleus after light irradiation but, although PpPHY::GFP contains all the PHY sequences homologous to AtPHYB, it localized in the cytoplasm (Fig. 3c). Arabidopsis cry2 localizes in the nucleus and PpCRY::GFP was found mainly here with a weak signal in the cytoplasm. PpDUC1::GFP exhibited mostly nuclear localization. We examined subcellular localization of PpDUC1::GFP in the dark and then on transfer to the light. In the darkness, PpDUC1::GFP also mostly localized in the nucleus, and, after white light irradiation, it did not show a clear change of localization (Supplementary Fig. 8).

P. provasolii possesses a unique gene repertoire for adaptation to various light conditions. We sequenced the complete genome of P. provasolii NIES-2893. The reads were assembled into 43 scaffolds without gaps and the scaffolds likely correspond to chromosomes (Supplementary Fig. 9, Supplementary Note 2, Supplementary Method 2). The genome size of P. provasolii is 22.7 Mbp, which is similar to other prasinophytes (12.5–22.0 Mbp) (Fig. 1c and Supplementary Method 3–5). We predicted and annotated 11,297 genes and evaluated gene annotation completeness with BUSCO. The results showed 89.1% (270/303) complete, 4.8% (16/303) fragmented and only 5.6% (17/303) missing BUSCOs. Among the seven prasinophytes that have their genomes determined, we identified 3996 conserved orthogroups (47.8% of P. provasolii genes) and a large number of unique genes (3636 orthogroups, 43.5% of the total) (Supplementary Fig. 9c).

With this genome sequence, we predicted five CRYs and one DUC1 but there is no PHY in the genome (Table 1). Of the CRY genes, two and the PpCRY region of PpDUC1 belong to the plant CRYs (pCRY). These three genes form a monophyly and branch at the basal position of those of other chlorophytes (Supplementary Fig. 3). Chlorophyte-specific CRYs are a sister group of the streptophyte pCRYs, including AtCRY1 and AtCRY2. The PpPHY region is a monophyly with other prasinophytes and streptophytes (Supplementary Fig. 4). In the prasinophytes, two genomes of mamiellophyceans lack genes for pCRY and PHY (Fig. 4).
The predicted genes also suggest that several important photoreceptors and light signal transduction genes in land plants are conserved. We detected PHOT, COP1, CK2alpha, some subunits of signalosome and HY5 homologs but not those for PIF, SPA1, SPA3, FHY1, FHY3, CIB1, or BIC for light signal transduction. The deduced amino acids sequence of LAF1 showed 44% identity with Arabidopsis LAF1 although this protein is suggested as an angiosperm-specific protein. We also confirmed the conservation of light signal transduction genes in 13 chlorophyte genomes, but PIF, CIB1 and BIC were not found in any (Table 1 and Supplementary Table 2).

For photosynthesis-related genes, its genome encoded a large gene family for a unique prasinophyte-specific light-harvesting complex (Lhcf). *P. provasolii* possesses a relatively large number of 16 Lhcf genes, and 10 of these are clustered in specific regions of chromosome 29 (Supplementary Fig. 10a). These 10 genes had almost the same sequences and were monophyletic (Supplementary Fig. 10a and Supplementary Method 6). Similar duplications of *lhc* genes are observed in Mamiellophyceae, however, these have occurred independently to *P. provasolii*.

Fig. 3 Subcellular localization of *PpDUC1* in tobacco leaves. a Schematic diagrams of the domains in the *PpPHY*, *PpCRY* and *PpDUC1* constructs used in the *N. benthamiana* leaf injection assay. b Immunoblot images of protein extract from *N. benthamiana* leaf tissue. Each protein was detected using an α-GFP antibody. Mock, non-injected leaf. Black arrows denote non-specific bands. Dot indicates the dye front. Experiments were repeated two times, and the results of one representative experiment are shown. c Leaf injection assay in *N. benthamiana* pavement cells transiently expressing *PpPHY-GFP*, *PpCRY-GFP* and *PpDUC1-GFP* with HY5-mCherry to indicate the nucleus. Observations were performed three times with similar results. DIC, differential interference contrast images; GFP, GFP fluorescence images; Merged, merged images of GFP and mCherry fluorescence images. Scale bar = 20 μm. The source data underlying Fig. 3b are provided as a Source Data file.
blue-light conditions (Fig. 5d). The HY5 homolog was also significantly expressed in blue and orange light with DCMU. Homologs of other light-signaling genes are listed in Fig. 5d.

Using real-time PCR, we found that the expression of two lhcb genes is induced mostly by orange and blue light, and DCMU treatment reduced their expression (Supplementary Fig. 12 and Supplementary Method 7). DUC1 expression was also induced by all light conditions and expression caused by orange and blue light was enhanced by DCMU treatment. These real-time PCR results were represented by RNA-Seq analysis.

Discussion
In this research, we have found a bifunctional photoreceptor, PpDUC1, composed of a fusion of PHY and CRY. In terms of evolution, it is often speculated that different domains of one organism’s protein are encoded by separate genes in another and this has been used successfully to speculate about direct physical interaction or indirect functional association. It is reported that phyB and cry2 physically interact to transduce light signals for controlling flowering time in Arabidopsis. Our discovery of PpDUC1 indicates that PHY and CRY interact actively to enable proper perception of light signals. Another example is Neurospora, which is found in ferns, that is composed of a PHY domain and a PHOT domain. This chimeric protein also supports the idea that there is dynamic interaction of photoreceptors.

From RNA-Seq analysis we found P. provasolii responds to orange, blue and far-red light and that 1,964 genes are expressed under orange and blue light. DCMU treatment reduced the number of DEGs to 1,094, orange DEGs being reduced from 1,503 to 471. Most of the genes whose expression was canceled by DCMU are genes involved in photosynthesis (Supplementary Fig. 11). Since DUC1 can sense orange, blue and far-red lights, some of the 119 DEGs of ORF light may be controlled by DUC1 (Fig. 5b). Interestingly 45 out of these 119 genes showed higher expression in blue compared to orange light. On the other hand, among 155 DEGs of orange and blue lights only 6 genes showed higher expression in blue compared to orange light (Fig. 5b). These 45 genes may be controlled by other cryptochromes enabling them to achieve their higher expression. DUC1, Plant-like CRY, CRY-dash and HY5 were all induced by blue light with DCMU treatment. Inhibition of photosynthesis may control expression of these genes (Fig. 5d).

In tobacco cells, PpDUC1 mostly localizes in the nucleus under white light (Fig. 3f) while the PpPHY domain is mainly localized in the cytoplasm. Although this investigation was done in a heterologous system, PpPHY’s intracellular localization did not change under different light conditions.

This may explain the results of the complementation assay that shows PpDUC1 does not complement phyB in respect of hypocotyl length (Supplementary Figs. 13 and 14, Supplementary Method 8 and 9). Plant and M. pusilla phytochromes translocate from the cytoplasm to the nucleus under light irradiation. Plant phytochrome is known to transduce its photoactivated signals through interaction with PIF protein in the nucleus and finally HY5 controls light-inducible gene expression. We did not observe light-dependent intracellular PpDUC1 translocation nor its enrichment in the nucleus by light in tobacco cells (Supplementary Fig. 8). Additionally, there is no PIF homolog in the P. provasolii genome (Table 1). Further analysis is needed to understand how DUC1 transduces light signals to control gene expression in P. provasolii.

In this study, we have revealed that PpPHY shows Po/Pf photocycle, which is similar to the PHY molecules derived from other prasinophyte species, indicative of the same origin. It is
well known in cyanobacteriochrome photoreceptors, distant relatives of the PHYs, that the trapped geometry of the rotating ring D is crucial for blue-shifted absorption. Residues unique to prasinophyte PHY molecules would be crucial for such a twist of ring D. Notably, two Tyr residues conserved among the plant and cyanobacterial PHYs that hold ring D are replaced with Phe, Met or Trp residues in the prasinophyte PHYs (Supplementary Fig. 15 and Supplementary Method 10), which may contribute to holding ring D in the trapped geometry, resulting in the absorption of blue-shifted orange light in the dark state Po form.

Many green algae share genes for PHY and pCRY. Of the prasinophytes, Tetraselmis, Nephroselmis, Micromonas, Dolichomastix, and Prasinoderma possess genes for functional PHY and pCRY. However, the genomes of Chloropicon, Ostreococcus, Bathycoccus, and M. commoda lack any PHY genes (Table 1 and Fig. 4). These PHYs are monophyletic and the tree topology coincides with the species tree (Fig. 1c and Supplementary Fig. 4), suggesting that PHY may have disappeared multiple times, independently. Our phylogenetic analysis inclusive of DUC1 supports the idea that the last common ancestor of the Archaeplastida had a phytochrome. In contrast, pCRY is widely shared in chlorophytes and streptophytes, and only the genomes of Mamiellales and Chlorella variabilis lack pCRY (Supplementary Fig. 3). In evolutionary terms, PpDUC1 was not found in other green algae except for P. provasolii and Pseudoscorephoria, and P. provasolii is sister to N. pyriformis, which possesses PHY and pCRY, which suggest that PpDUC1 may have been acquired in an ancestor of Pycnococcus (and Pseudoscorephoria). As the domain structure of PpDUC1 is similar to the Phy and Cry of N. pyriformis (Fig. 4), PpDUC1 may have been generated via the fusion of these genes. Recent fusion is suggested by the remaining original GC%, i.e., the PHY region has higher GC% than the pCRY region in PpDUC1 (PpPHY: 62.2% and PpCRY: 59.5%).

Sensitivity to weak light is essential for marine algae. P. provasolii also has a unique pigment composition (prasinoxanthin and Magnesium 2,4-divinylphoporphyrin a5 monomethyl ester) and an ability to adapt to spectral quality (blue and blue-violet) and low-light intensity. Under high-light radiation, algae and land plants degenerate antenna complexes (LHCs and pigments) to decrease the absorbance of excess energy and prevent photodamage. The degradation of Chl triggers degradation of LHCs. Interestingly, LHC induction by orange and blue light is strongly reduced by DCMU treatment but ELIP expression is induced by DCMU treatment. ELIPs are members of LHCs that...
Fig. 5 Transcriptome analysis under monochromatic blue, orange, and far-red light.  

a Number of DEGs under each monochromatic light (blue, orange, far-red) against dark conditions. Right: samples were treated with DCMU; left: no DCMU treatment. The DEG is defined as >1.5-fold for gene expression with a q-value < 0.05.

b Venn-diagram of number of DEGs in Fig. 5a.

c GO enrichment analysis for 686 co-regulated DEGs under blue and orange light.

d Expression values for DUC1 and light-signaling genes. The source data underlying Fig. 5b, d are provided as a Source Data file.
protection photosynthesis from high-light irradiation. Therefore, *P. provosolii* may adjust the amount of antenna complex by Chl degradation under strong light at the surface of the sea.

For photosynthesis in marine environments, Chl *b* is suitable because its absorption peak is shifted to blue-green compared with Chl *a*. This use of Chl *b* is different to land plants. Marine prasinophytes possess Chl *b*, not only in LHChs but also in PSI core antennae.

Phylogenetic analyses of Phy and Cry regions of DUC1

Analyses of DUC1, we divided it into Phy (1–3087 nucleotides) and Cry (3115–5073 nucleotides) regions. The datasets of Phy and Cry regions were composed of 54 operational taxonomic units (OTUs) and 96 OTUs, respectively, including available homologs from official databases. These sequences were aligned using maft v7.4539 with a linsi option, and ambiguous regions were trimmed using v1.4.rev1540 with the option automated.1 The trimmed dataset contained 457–932 amino acids for Phy and Cry regions, respectively. A model test was performed using ModelTest-NG v0.1.541. Maximum likelihood analysis was performed using RAxML-NG v0.9.0 with 200 bootstrap replicates.

Phylogenetic analyses of Phy and Cry regions of DUC1

For phylogenetic analyses of DUC1, we divided it into Phy (1–3087 nucleotides) and Cry (3115–5073 nucleotides) regions. The regions of Phy and Cry were composed of 54 OTUs. The DNA fragment of the Phy region was amplified by PCR using KOD OneTM PCR Master Mix (Toyobo Life Science) with a codon-optimized synthetic gene for expression in *E. coli* (GeneScript) and an appropriate nucleotide primer set (Supplementary Table 2). The Phy region was amplified by PCR using DNA polymerase with template DNA and an appropriate nucleotide primer set (Supplementary Table 2). The Cry region (3115–5073 bp in *PpDUC1*) was cloned into the EcoRI and XbaI sites of a pCold GST DNA vector with an N-terminal GST-tag (TaKaRa) using restriction enzymes and ligase. A DNA fragment of the Cry region was amplified by PCR using PrimeSTAR Max DNA Polymerase (TaKaRa) with genomic DNA from *P. provosolii* and an appropriate nucleotide primer set (Supplementary Table 2). All of the plasmid constructs were verified by nucleotide sequencing (FASMAC).

Expression and purification of PhyCh and His-tag

The *E. coli* strain C41 was used for GST-tagged PhyCh (amino acid positions 1–662 in *PpDUC1*) expression through the pGKJ8 construct as a chaperone expression system (TaKaRa).2 Bacterial cells were grown in 8 L LB medium containing 500 mg L−1 t-arabino and antibiotics (5 mg L−1 tetracycline, 100 µg mL−1 ampicillin, and 20 µg mL−1 chloramphenicol) at 37 °C. Protein expression was monitored since the optical density at 600 nm of the cells reached 0.4–0.8, IPTG was added (final concentration, 0.1 mM), and the cells were cultured at 18 °C overnight.

After incubation, the culture broth was centrifuged at 5000 × g for 15 min to collect the cells. The cells were resuspended in lysin buffer A (20 mM HEPES–NaOH pH 7.5, 100 mM NaCl and 10% (w/v) glycerol) with 0.5 mM tris(2-carboxyethyl)phosphine, and then disrupted using an Emulsiflex C5 high-pressure homogenizer at 12,000 psi (Avestin). Homogenates were centrifuged at 165,000 × g for 30 min and then the supernatants were filtered through a 0.8 µm cellulose acetate membrane before loading on to a nickel-affinity HisTrap HP column (GE Healthcare) using the AKTA pure 25 (GE Healthcare) system. The column was washed using the buffer containing 100 mM imidazole and, then, His-tagged PhyCh was eluted with a linear gradient of the imidazole buffer containing 100 to 400 mM imidazole (1 mL min−1, total 15 min). After incubation with 1 mM EDTA for 1 h, His-tagged PhyCh was dialyzed against the buffer with 1 mM dihydrothirol (DTT) to remove EDTA and imidazole. The purified proteins were dialyzed against the buffer containing 1 mM dihydrothirol (DTT). Protein concentration was determined by the Bradford method.

Expression and purification of PhyCh and His-tag

The *E. coli* strain C41 was used for GST-tagged PhyCh (amino acid positions 1039–1699 in *PpDUC1*) expression through the PGKJ8 construct as a chaperone expression system (TaKaRa).13 Bacterial cells were grown in 8 L LB medium containing 500 mg L−1 t-arabino and antibiotics (5 mg L−1 tetracycline, 100 µg mL−1 ampicillin, and 20 µg mL−1 chloramphenicol) at 37 °C. Protein expression was monitored since the optical density at 600 nm of the cells reached 0.4–0.8, IPTG was added (final concentration, 1 mM), and the cells were cultured at 15 °C overnight.

After incubation, the culture broth was centrifuged at 5000 × g for 15 min to collect the cells. They were resuspended in lysin buffer B (20 mM HEPES–NaOH pH 7.5, 100 mM NaCl, 5 mM MgCl2 and 10% (w/v) glycerol), and then disrupted using a homogenizer at 12,000 psi. The homogenate was centrifuged at 165,000 × g for 30 min and the supernatant filtered through a membrane before loading onto a glutathione-affinity GSTrap HP column (GE Healthcare). The column was washed with lysin buffer B to remove unbound proteins, and GST-tagged PhyCh was subsequently eluted with buffer containing 10 mM reduced glutathione. Protein concentration was determined by the Bradford method. The extracted protein was handled in the dark.

Sodium dodecyl sulphate-polycrylamide gel electrophoresis analysis for purified PhyCh and PhyCh

Purified proteins were diluted to 60 mM DTT, 2% (w/v) SDS and 60 mM Tris-HCl, pH 8.0, and then denatured at 95 °C for 3 min. These samples were electrophoresed at room temperature using 10% (w/v) polyacrylamide gels with SDS. The gels were soaked in distilled water for 30 min followed by monitoring of fluorescence bands for detection of chromophores covalently bound to the proteins.24 These bands were visualized through a 600-nm long-path filter upon excitation with blue (λmax = 470 nm) and green (λmax = 527 nm) light through a 562 nm short-path filter using WSE-6100 LuminoGraph (Lambda Sci, Tokyo) and WSE-5500 Varilux (ATTO) machines. After the monitoring, the gels were stained with Coomassie Brilliant Blue R-250.

UV-Vis spectroscopic analysis to monitor photocycles of PhyCh and PhyCh

Ultraviolet and visible absorption spectra of the proteins were recorded with a UV-2600 spectrophotometer (SHIMADZU) at room temperature (r.t., approximately 20–25 °C). An Opto-Spectrum Generator (Hamamatsu Photonics, Inc.) was used to produce monochromatic light of various wavelengths to induce photoconversion.

Assignment of the chromophores incorporated into PhyCh

Sample solutions containing the native PhyCh in the dark state and the photoproduct state were diluted fivefold in 8 M acidic urea (pH < 2.0). The absorption spectra were recorded at r.t. before and after 3 min of illumination with white light. Assignment of the chromophores was conducted by comparing the spectra between the denatured PhyCh and standard proteins.13,24

Assignment of the chromophores incorporated into PhyCh

Trichloroacetic acid was added to the sample solution containing the native PhyCh in the dark state to a final concentration of 440 mM. The solution was treated on ice for 1 h with shaking at 200 rpm. The precipitate was removed by centrifugation at 20,000 × g for 5 min, and then 50 µL of the supernatant was injected into a high-performance liquid chromatograph (Prominance HPLC system, Shimadzu). Released chromophores were monitored in the supernatant (1% ethanol) and visualized using a 600-nm long-path filter upon excitation with blue (λmax = 470 nm) and green (λmax = 527 nm) light through a 562 nm short-path filter using WSE-6100 LuminoGraph (Lambda Sci, Tokyo) and WSE-5500 Varilux (ATTO) machines. After the monitoring, the proteins were dialyzed against 10 mM reduced glutathione. Protein concentration was determined by the Bradford method. The extracted protein was handled in the dark.
The synthetic gene as a template and cloned into the pSK1 plasmid vector. DNA fragment was amplified using PpDUC1, the leaf epidermal cells of Arabidopsis thaliana, for transient transformation in Nicotiana benthamiana. Constructs for transient transformation in N. benthamiana leaves. Each construct was transformed into Agrobacterium GV3101. These agrobacteria were infiltrated into the leaf epidermal cells of N. benthamiana by cluting with 10 mM MES (pH 5.6) with 200 µM acetosyringone into the undersides of leaves of N. benthamiana. A Zeiss LSM880 Airyscan Fast-mode microscope (Zeiss) with AxisObserver Z1 20X and W40X objectives was used to detect GFP fluorescence. The excitation wavelength was 488 nm, and a band-pass filter of 493 to 536 nm was used for emission for GFP. The excitation wavelength was 561 nm, and a band-pass filter of 578 to 640 nm was used for emission for mCherry. Transient expression and observation of PpDUC1-GFP in N. benthamiana leaves. P. provasolii genomic DNA by PCR with primer set (Supplementary Table 2) and cloned into the pDONR207 ENTRY vector by BP recombination using with BP Clonase II enzyme (Invitrogen Corp.). These open reading frames were transformed into the pFlareGate103 destination vector fused to GFP and 6×His by LR reaction. A synthetic gene of Arabidopsis thaliana HY5 (Asg11266) fused to mCherry (Sequence ID: MF976504.1) was synthesized by Eurofins Genomics. The DNA fragment was amplified by PCR with primers (Supplementary Table 2) using the synthetic gene as a template and cloned into the pSk1 plasmid vector.

DNA extraction, genome sequencing, assembly, and annotation. P. provasolii NIES-2893 cells were cultivated for 4–6 days in −20 mL IMK medium (Nihon Pharmaceutical, Tokyo, Japan) under white LED light (~30 µmol m−2 s−1) with a 14 h:10 h light:dark cycle. The cells of P. provasolii were harvested by centrifugation and ground in a pre-cooled mortar with liquid nitrogen and 50 mg of 0.1 mm glass beads (Bertin, Rockville, MD, USA). The cells were incubated with 600 µL of CTAB extraction buffer and 0.5% 2-mercaptoethanol, 5% glycerol, 0.5% Triton X-100, and CompleteTM protease inhibitor mini cocktail (Sigma). After grinding in the protein extraction buffer total slurry was centrifuged twice at 12,000 rpm for 5 min to remove particulates. Ten microliters of supernatant was added with 5 µL of protein loading buffer and heat denatured. The denatured samples were loaded onto a 7.5% SDS-polyacrylamide gel for electrophoresis. After electrophoresis, proteins were electrobolted onto a polyvinylidene difluoride (PVDF) membrane (Millipore) in the blotting buffer consisting 25 mM Tris, 192 mM glycine, and 20% methanol. Subsequently, the membrane was incubated for 1.5 h in 1% of skimmed milk (Nakarai tesque), rinsed for 10 min twice with 1xTBST consisting 137 mM NaCl, 2.68 mM KCl, 25 mM Tris-HCl, pH 7.4, 0.1% w/v reagents (Millipore) in the blotting buffer composing 25 mM Tris, 192 mM glycine and 20% methanol. Both three times with 10 min intervals and incubated with protein A, horseradish peroxidase linked antibody (NA9120, GE Healthcare Corp., 1:2000) for 30 min. Both the membranes were rinsed three times for 10 min each with 1xTBST and after incubation with enhanced chemiluminescence reagents (ECL SelectTM Western Blotting Detection Reagents, GE Healthcare Corp.) They were processed with a chemi-luminescent image analyzer (Chem Doc XRS plus, BIO-RAD).

Blotting Detection Reagents, GE Healthcare Corp.) They were processed with a chemi-luminescent image analyzer (Chem Doc XRS plus, BIO-RAD).

The genome browser and transcriptome data of N. benthamiana are available at: http://marine-meta.healthscience.waseda.ac.jp/crnet/metacrest/graphs/1, and MMETST (The Marine Microbial Eukaryote Transcriptome Sequencing Project)[https://www.imicrobe.us/#/projects/104]. Source data are provided with this paper.

Received: 5 October 2020; Accepted: 11 May 2021; Published online: 16 June 2021

References

1. Liu, Y. F. et al. Phytochrome diversity in green plants and the origin of canonical plant phytochromes. Nat. Commun. 6, 7852 (2015).

2. Rockwell, N. C. et al. Eutrophication, the algal phytochromes span the visible spectrum. Proc. Natl Acad. Sci. USA 111, 3871–3876 (2014).

3. Rockwell, N. C., Su, Y. S. & Lagarias, J. C. Phytochrome structure and signaling mechanisms. Annu. Rev. Plant Biol. 57, 837–858 (2006).

4. Lemieux, C., Ottis, C. & Turmel, M. Six newly sequenced chloroplast genomes from prasinophyte algae provide insights into the relationships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. BMC Genomics 15, 857 (2014).

5. Guillard, R. R. L., Keller, M. D., O’Kely, C. J. & Floyd, G. L. Pycnococcus provasolii gen. et sp. nov., a coccoid prasinoxanthin-containing phytoplankter from the Western North Atlantic and Gulf of Mexico. J. Phycol. 27, 39–47 (1991).

The publicly available datasets in this study include: PhycoCosm database [https://phyocosm.igj.doe.gov/phyocosm/home/], the algal strains are available at the NIES collection: P. provasolii (NIES-2093) and P. marina (NIES-1419). The data is available online: Phytochromes in plants: Phytochrome A (http://www.broadinstitute.org/mcp/phytochrome/index.html). Read counts were normalized with DESeq 1.42.0 in an R package. Differentially expressed genes (DEGs) were defined as ≥1.5-fold for an gene expression with a q-value < 0.05.

Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this Article is available as a Supplementary Information file. The datasets and plant materials generated and analyzed during the current study are available from the corresponding author upon request. The genome and transcriptome data are deposited in the DDRI/EMBL/GenBank under the accession number of GCA_015473125.1, PRJDB10693, and PRJNA726377, respectively. The genome browser and transcriptome data of P. provasolii are available at: http://matu-ib.lab.riken.jp/Pbrown/index.html/data=databasePrynococcus. The algal strains are available in the NIES collection: P. provasolii (NIES-2093) and P. marina (NIES-1419). The data is available online: Phytochromes in plants: Phytochrome A (http://www.broadinstitute.org/mcp/phytochrome/index.html). Read counts were normalized with DESeq 1.42.0 in an R package. Differentially expressed genes (DEGs) were defined as ≥1.5-fold for an gene expression with a q-value < 0.05.

Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this Article is available as a Supplementary Information file. The datasets and plant materials generated and analyzed during the current study are available from the corresponding author upon request. The genome and transcriptome data are deposited in the DDRI/EMBL/GenBank under the accession number of GCA_015473125.1, PRJDB10693, and PRJNA726377, respectively. The genome browser and transcriptome data of P. provasolii are available at: http://matu-ib.lab.riken.jp/Pbrown/index.html/data=databasePrynococcus. The algal strains are available in the NIES collection: P. provasolii (NIES-2093) and P. marina (NIES-1419). The data is available online: Phytochromes in plants: Phytochrome A (http://www.broadinstitute.org/mcp/phytochrome/index.html). Read counts were normalized with DESeq 1.42.0 in an R package. Differentially expressed genes (DEGs) were defined as ≥1.5-fold for an gene expression with a q-value < 0.05.

Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this Article is available as a Supplementary Information file. The datasets and plant materials generated and analyzed during the current study are available from the corresponding author upon request. The genome and transcriptome data are deposited in the DDRI/EMBL/GenBank under the accession number of GCA_015473125.1, PRJDB10693, and PRJNA726377, respectively. The genome browser and transcriptome data of P. provasolii are available at: http://matu-ib.lab.riken.jp/Pbrown/index.html/data=databasePrynococcus. The algal strains are available in the NIES collection: P. provasolii (NIES-2093) and P. marina (NIES-1419). The data is available online: Phytochromes in plants: Phytochrome A (http://www.broadinstitute.org/mcp/phytochrome/index.html). Read counts were normalized with DESeq 1.42.0 in an R package. Differentially expressed genes (DEGs) were defined as ≥1.5-fold for an gene expression with a q-value < 0.05.

Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this Article is available as a Supplementary Information file. The datasets and plant materials generated and analyzed during the current study are available from the corresponding author upon request. The genome and transcriptome data are deposited in the DDRI/EMBL/GenBank under the accession number of GCA_015473125.1, PRJDB10693, and PRJNA726377, respectively. The genome browser and transcriptome data of P. provasolii are available at: http://matu-ib.lab.riken.jp/Pbrown/index.html/data=databasePrynococcus. The algal strains are available in the NIES collection: P. provasolii (NIES-2093) and P. marina (NIES-1419). The data is available online: Phytochromes in plants: Phytochrome A (http://www.broadinstitute.org/mcp/phytochrome/index.html). Read counts were normalized with DESeq 1.42.0 in an R package. Differentially expressed genes (DEGs) were defined as ≥1.5-fold for an gene expression with a q-value < 0.05.
6. Shiosaki, K. et al. Eukaryotic Phytoplankton contributing to a seasonal bloom and carbon export revealed by tracking sequence variants in the Western North Pacific. *PLoS Biol.* 10, e1002722 (2012).

7. Tragin, M., dos Santos, A. L., Christen, R. & Vaulot, D. Diversity and ecology of green microalga in marine systems: an overview based on 18s rRNA gene sequences. *Perspect. Phycol.* 3, 141–154 (2016).

8. Iriarte, A. & Purdie, D. A. Photosynthesis and growth response of the oceanic picophytoplankter *Prymnesium parvus* Greville (clone D48–23) (*Chlorophyta*) to variations in irradiance, photoperiod and temperature. *J. Exp. Mar. Biol. Ecol.* 168, 239–257 (1993).

9. Rockwell, N. C. & Lagarias, J. C. Phytochrome evolution in 3D: deletion, duplication, and diversification. *N. Phytol.* 225, 2283–2300 (2020).

10. Frankenberg, N., Mukougawa, K., Kohchi, T. & Lagarias, J. C. Functional genomic analysis of the HY2 family of ferredoxin-dependent bilin reductases from oxygenic photosynthetic organisms. *Plant Cell* 13, 965–978 (2001).

11. Miyake, K. et al. Functional diversification of two bilin reductases for light perception and harvesting in unique cyanobacterium *Acaryochloris marina* MBIC 11017. *FEBS J.* https://doi.org/10.1111/febs.15230 (2020).

12. Liu, B., Liu, H., Zhong, D. & Lin, C. Searching for a photocycle of the phytochrome photoreceptors. *Curr. Opin. Plant Biol.* 13, 578–586 (2010).

13. Nishihara, K., Kanemori, M., Yanagi, H. & Yura, T. Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. *Appl. Environ. Microb.* 66, 884–889 (2000).

14. Berndt, A. et al. A novel photo receptors for the circadian blue light receptor CRY1. *Science* 297, 1360–1363 (1998).

15. Han, X. et al. Origin and evolution of core components responsible for blue light receptor CRY1. *Proc. Natl Acad. Sci. USA* 111, 15827–15832 (2014).

16. Shiozaki, T. et al. Eukaryotic Phytoplankton contributing to a seasonal bloom in higher plants: environmental acclimation and photosynthetic harvesting systems. *Nature* 446, 171–172 (2007).

17. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a tool for automated trimming alignment in large-scale phylogenetic analyses. *Bioinformatics* 25, 1737–1741 (2009).

18. Darriga, D. et al. ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* 37, 291–294 (2020).

19. Köpke, K. & Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinformatics* 9, 286–298 (2008).

20. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973 (2009).

21. Hasegawa, M. et al. Molecular characterization of D. *Biol. Chem.* 239, 408–415 (1964).

22. Suzuki, S., Yamaguchi, H., Nakajima, N. & Kawachi, M. Cryptochrome family in photosynthetic organisms. *Biochim. Biophys. Acta* 1275, 387–395 (1996).

23. México, E. M. et al. Detecting protein function and protein-protein interactions from genome sequences. *Science* 285, 751–753 (1999).

24. Madsen, P., Devlin, P. A., Sandoz, A. L. & Rockwell, N. C. Functional interaction of cyanobacteriochrome photoreceptors. *Proc. Natl Acad. Sci. USA* 111, 15827–15832 (2014).

25. Kim, Y. M. et al. dbCRY: a Web-based comparative and evolutionary genomics platform for blue-light receptors. *Database 2014*, bau037 (2014).

26. O’Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D713–D745 (2016).
Acknowledgements
This research was partially supported by the National BioResource Project (NBRP) from Japan Agency for Medical Research and Development (AMED). This research was supported by JSPS KAKENHI Grant Number 20K21438.

Author contributions
Y.M. and A.S. identified the DUC1 gene and analyzed its expression. S. Suzuki, H.Y., M. K. determined the P. pravasolii genome. K.F., R.N. confirmed the absorbance spectra of PpPHY and PpCRY. S. Shimada, M.H., T.K., and M.S. contributed the RNA analysis and complementation assay in Arabidopsis. H.H., E.O.K., and S. Shimada examined the intracellular localization of DUC1. K.Y., T.W., T.G., T.S. provided the metagenome data and the gDNA samples. Y.M., S. Suzuki, K.F., S. Shimada, Y.K., R.N., H.Y., M.K., M.M. wrote the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-23741-5.

Correspondence and requests for materials should be addressed to M.M.

Peer review information Nature Communications thanks J. Clark Lagarias, Thomas Mock, and other, anonymous, reviewers for their contributions to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021, corrected publication 2021