





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

Yuko Makita<sup>1,13</sup>, Shigekatsu Suzuki<sup>2,13</sup>, Keiji Fushimi<sup>3,4,5,13</sup>, Setsuko Shimada<sup>1,13</sup>, Aya Suehisa<sup>1</sup>, Manami Hirata<sup>1</sup>, Tomoko Kuriyama<sup>1</sup>, Yukio Kurihara<sup>1</sup>, Hidefumi Hamasaki<sup>1,6</sup>, Emiko Okubo-Kurihara<sup>1</sup>, Kazutoshi Yoshitake<sup>7</sup>, Tsuyoshi Watanabe<sup>8</sup>, Masaaki Sakuta <sup>9</sup>, Takashi Gojobori<sup>10</sup>, Tomoko Sakami<sup>11</sup>, Rei Narikawa <sup>3,4,5,12</sup>, Haruyo Yamaguchi <sup>2</sup>, Masanobu Kawachi<sup>2</sup> & Minami Matsui <sup>1,6</sup>✉

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.

<sup>1</sup> Synthetic Genomics Research Group, RIKEN Center for Sustainable Resource Science, Yokohama, Japan. <sup>2</sup> Biodiversity Division, National Institute for Environmental Studies, Tsukuba, Japan. <sup>3</sup> Graduate School of Integrated Science and Technology, Shizuoka University, Shizuoka, Japan. <sup>4</sup> Research Institute of Green Science and Technology, Shizuoka University, Shizuoka, Japan. <sup>5</sup> Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Saitama, Japan. <sup>6</sup> Yokohama City University, Kihara Institute for Biological Research, Yokohama, Japan. <sup>7</sup> Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan. <sup>8</sup> Fisheries Resources Institute, Japan Fisheries Research and Education Agency, Kushiro, Hokkaido, Japan. <sup>9</sup> Department of Biological Sciences, Ochanomizu University, Tokyo, Japan. <sup>10</sup> Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia. <sup>11</sup> Fisheries Resources Institute, Japan Fisheries Research and Education Agency, Minami-ise, Mie, Japan. <sup>12</sup> Department of Biological Sciences, Graduate School of Science, Tokyo Metropolitan University, Tokyo, Japan. <sup>13</sup> These authors contributed equally: Yuko Makita, Shigekatsu Suzuki, Keiji Fushimi, Setsuko Shimada. ✉email: [minami@riken.jp](mailto: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 cryptochromes (CRYs) but not red-light receptor phytochromes (PHYs)<sup>1</sup>. 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<sup>2</sup>. They have the ability to photosense between red-absorbing Pr and far-red-absorbing Pfr, and this conformational change enables interactions with signaling partners<sup>3</sup>. CRY is a photolyase-like flavo-protein 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.sci.waseda.ac.jp/crest/metacrest/graphs/>). These metagenome analyses targeted eukaryotic marine microorganisms as well as bacteria. Blue-light sensing is key to adapting to marine environments. From a search for CRYs in the marine metagenomic data, we found a chimeric photoreceptor, designated as Dualchrome1 (DUC1), consisting of a two-domain fusion of PHY and CRY. We found that DUC1 originated from a prasinophyte alga, *Pycnococcus provasolii*.

*P. provasolii* is a marine coccoid alga in Pseudoscourfieldiales, Pycnococaceae (or prasinophyte clade V<sup>4</sup>) and was originally discovered in the pycnocline<sup>5</sup>. *P. provasolii* is classified in Chlorophyta, which is a sister group of the Streptophyta in Viridiplantae. Chlorophyta contain three major algal groups, Ulvophyceae, Trebouxiophyceae and Chlorophyceae (UTC clade), and “prasinophytes”, which have several characteristics considered to represent the last common ancestor of Viridiplantae. Prasinophytes mainly inhabit marine environments and are dominant algae under various light qualities and intensities<sup>6</sup>. Thus, prasinophytes are key to understanding the diversity and evolutionary history of the light response system in the Viridiplantae.

Environmental DNA research shows *P. provasolii* lives at depths in the range of 0–100 m and in varying regions of the marine environment<sup>6,7</sup>. It also has a unique pigment composition (prasinolanthin and Magnesium 2,4-divinylpheoporphyryl *a*<sub>5</sub> monomethyl ester) and an ability to adapt to the spectral quality (blue and blue-violet) and low fluxes of light found in the deep euphotic zone of the open sea<sup>5,8</sup>.

We unveil here with DUC1's ability to detect a wide range of the light spectrum (orange to far-red for PHY and UV to blue for CRY) and undergo dual photoconversions at both the PHY and CRY regions. We sequence the genome of *P. provasolii* and examine its light-associated features. These findings will help to understand the evolutionary diversity of photoreceptors in algae and explain the environmental adaptation and success of *P. provasolii*.

## 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 a/b/a\_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*

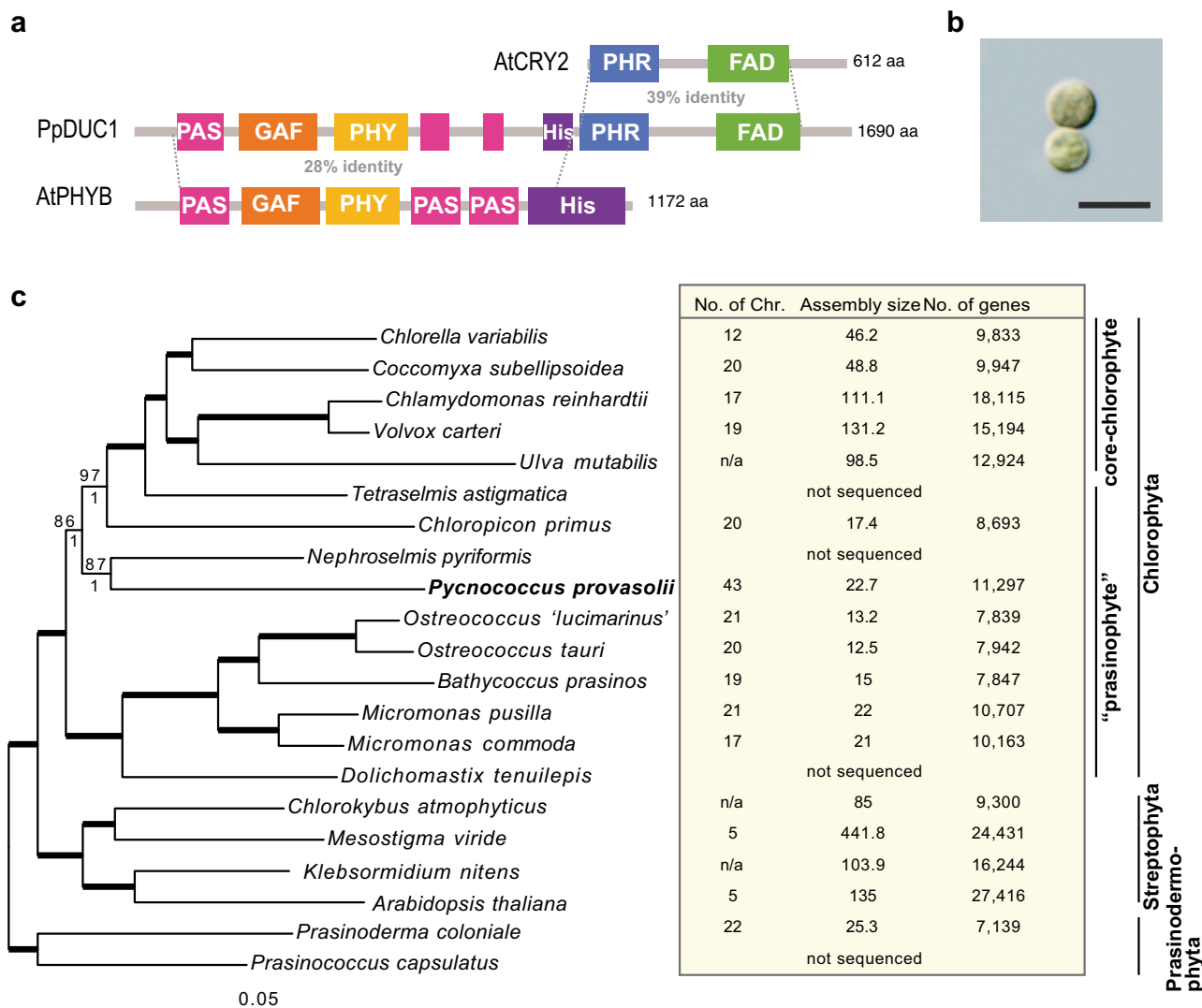
phytochrome B (PHYB) (Fig. 1). This gene has no introns and can encode angiosperm PHY domains at its N-terminus and CRY domains at its C-terminus (Fig. 1a).

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 *Dualchrome1* 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 *Pseudoscourfieldia 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 (PpPHY, 70.2 kDa, amino acid positions 1–662 in PpDUC1) and CRY (PpCRY, 72.2 kDa, amino acid positions 1039–1690 in PpDUC1) regions with His-tag and GST-tag, separately (Fig. 2a).

PHYs 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 photoproduct state (or excited state) of the PHYs (Supplementary Fig. 5a)<sup>9</sup>. The GAF Cys ligates to C3<sup>1</sup> of phycocyanobilin (PCB) or phytochromobilin (PΦB), whereas the N-terminal Cys ligates to C3<sup>2</sup> of biliverdin IXα (BV). As the PpPHY possesses the GAF Cys residue but not the N-terminal Cys, the binding chromophore is likely to be PCB or PΦB. Furthermore, we detected a bilin reductase homolog from the *P. provasolii* NIES-2893 genome (described later), which belongs to the PcyA cluster, producing PCB from BV, but not to the HY2 cluster, producing PΦB from BV (Supplementary Fig. 6a, b)<sup>10</sup>. Taken together, we presumed that the native chromophore for PpPHY is PCB and not PΦB. Therefore, we expressed PpPHY in *E. coli* harboring the PCB-synthetic system (C41\_pKT271)<sup>11</sup>. The purified PpPHY covalently bound PCB (Supplementary Figs. 5b and 6c) and showed reversible photoconversion between an orange-absorbing (Po) form ( $\lambda_{\max}$ , 612 nm) in the dark state and a far-red-absorbing (Pfr) form ( $\lambda_{\max}$ , 702 nm) in the photoproduct state (Fig. 2b and Supplementary Table 1). Dark reversion from the photoproduct state to the dark state was not seen (Supplementary Fig. 7a). Furthermore, we observed that PΦB but not BV could be efficiently incorporated into the apo-PpPHY to show reversible photoconversion by using the PΦB- and BV-synthetic systems, which is consistent with previous studies (Supplementary



**Fig. 1 Chimeric photoreceptor PpDUC1 and phylogeny of its host *P. provasolii*.** **a** Domain conservation between PpDUC1 and *Arabidopsis* CRY2 and PHYB. Domain structure of PpDUC1 constructed with PAS: Per-Arnt-Sim, GAF: cGMP-specific phosphodiesterases, adenylyl cyclases and FhIA. PHY: phytochrome, His: histidine kinase domain, PHR: photolyase-homologous region, and FAD: flavin adenine dinucleotide. **b** Light microscopic image of coccoid cells of *P. provasolii* NIES-2893. Scale bar = 5 μm. Observations were repeated more than three times and representative cells are shown. **c** Phylogeny of *P. provasolii* and genome statistics of the main lineages of Viridiplantae. Maximum likelihood (ML) tree of 105 orthologous proteins. The dataset was composed of 41,023 amino acids from 21 species that have genomes available, including *P. provasolii*. Bootstrap percentages (BPs) and Bayesian posterior probabilities (BPP) are shown above and below the lines, respectively. Bold lines show BP = 100 and BPP = 1.00.

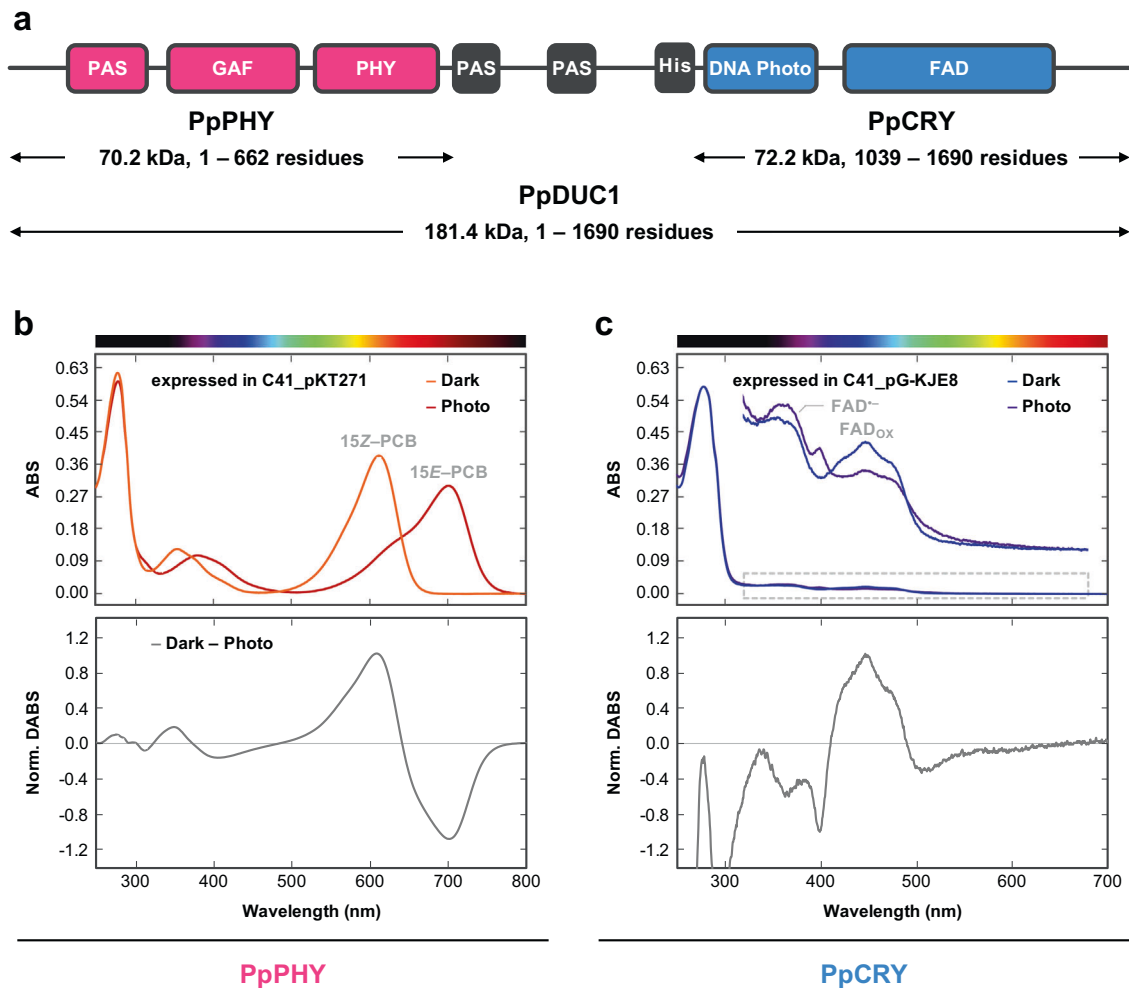
Fig. 6c–f). Taking the presence of the PcyA homolog into consideration, it is likely that PpPHY reversibly senses orange and far-red light via PCB incorporation in vivo.

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 (FAD<sub>OX</sub>), anion radical FAD (FAD<sup>•-</sup>), neutral radical FAD (FADH<sup>•</sup>) and reduced FAD (FAD<sub>red</sub>H<sup>-</sup>), are observed in the photocycle of CRYs (Supplementary Fig. 5c)<sup>12</sup>. 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)<sup>13</sup>. The purified PpCRY non-covalently bound FAD (Supplementary Figs. 5d and 6c). The protein showed a blue-absorbing (Pb) form (λ<sub>max</sub>, 447 nm) before light irradiation (Fig. 2c and Supplementary Table 1). The spectral form in the dark state was similar to that of the FAD<sub>OX</sub>-bound CRYs<sup>12</sup>. Blue-light irradiation resulted in conversion to a UV-absorbing (Puv) form (λ<sub>max</sub>, 366 and 399 nm) (Fig. 2c and Supplementary Table 1). The spectral form in

the photoproduct state was similar to that of the FAD<sup>•-</sup>-bound CRYs<sup>12</sup>. 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 FAD<sub>OX</sub>-bound form to the FAD<sup>•-</sup>-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*<sup>14</sup>, rather than CraCRY from the green algae *Chlamydomonas reinhardtii*<sup>15</sup> and AtCRY1 and 2 from the land plant *Arabidopsis thaliana*<sup>16</sup>.

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*



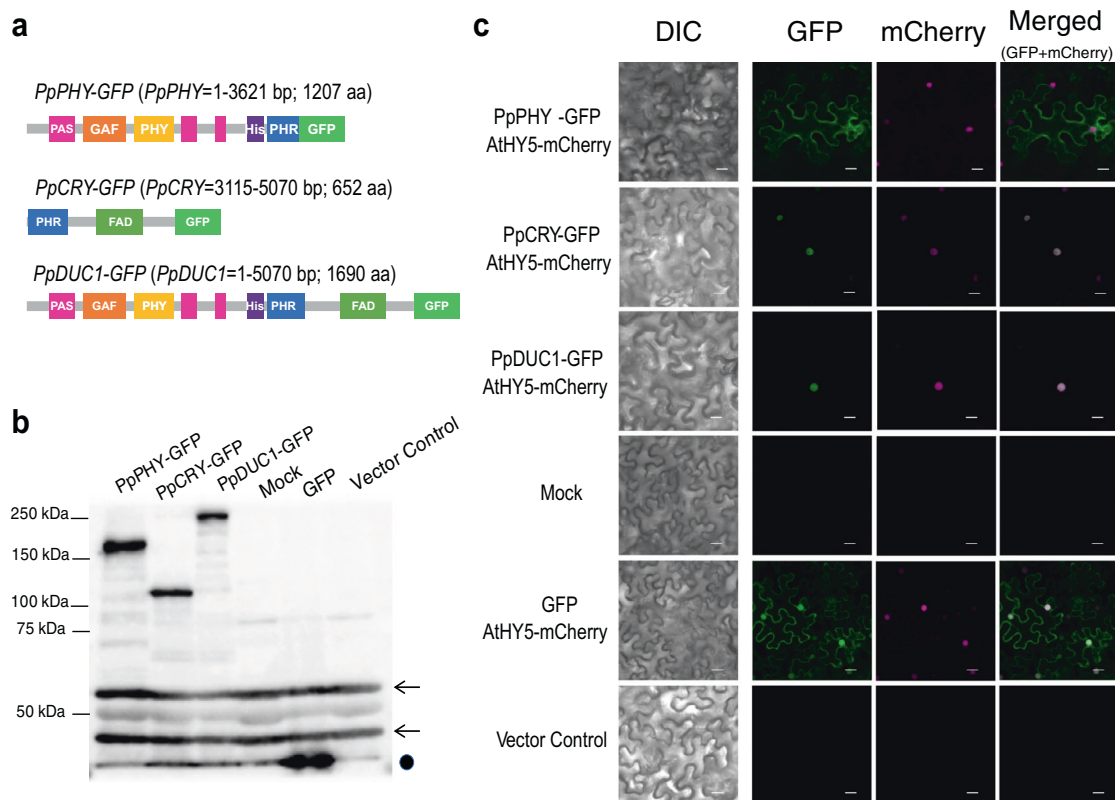
**Fig. 2** Photoconversion of PpPHY and PpCRY. **a** Domain structure of PpDUC. PpPHY (magenta) and PpCRY (cyan) regions are shown with their molecular sizes and amino acid positions. **b** Absorption spectra of the dark state (15Z-PCB, Po form) and the photoproduct state (15E-PCB, Pfr form) of PpPHY expressed in C41\_pKT271 (harboring the PCB-synthetic system). The normalized difference spectrum (dark state—photoproduct state) was calculated from these absorption spectra. **c** Absorption spectra of the dark state (FAD<sub>ox</sub>, Pb form) and the photoproduct state (FAD<sup>-</sup>, Puv form) of PpCRY expressed in C41\_pG-KJE8 (harboring the chaperone expression system). The normalized difference spectrum (dark state— photoproduct state) was calculated from these absorption spectra. The absorption maxima of PpPHY and PpCRY are reported in Supplementary Table 1. Assignment of chromophores incorporated in PpPHY and PpCRY was performed by comparison with each standard, which are reported in Supplementary Fig. 5.

CCMP1545, a marine microalgae<sup>17</sup>. 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<sup>18</sup> 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<sup>19</sup> 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*<sup>15</sup>). 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 *pCRY*s, 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).



**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  $\alpha$ -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  $\mu$ m. The source data underlying Fig. 3b are provided as a Source Data file.

The predicted genes also suggest that several important photoreceptors and light signal transduction genes in land plants are conserved. We detected *PHOT*, *COPI*, *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<sup>20</sup>. 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 (*Lhcp*). *P. provasolii* possesses a relatively large number of 16 *Lhcp* 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 *lhcb* genes are observed in Mamiellophyceae<sup>21</sup>, however, these have occurred independently to *P. provasolii*.

***P. provasolii* responds to orange, blue, and far-red light for gene expression.** We performed RNA-seq analysis to know how *P. provasolii* responds to monochromatic light after dark acclimatization. The expression of 1,503 genes and 1,147 genes changed after orange- and blue-light irradiation, respectively (Fig. 5a, b left). Through gene ontology (GO) enrichment analysis of the common differentially expressed genes (DEGs) in orange-

and blue-light conditions, photosynthesis and light-harvesting proteins were identified (Fig. 5c). To distinguish the effect of photosynthesis, we applied the photosynthesis inhibitor DCMU (3-(3,4-dichlorophenyl)-1,1-dimethylurea), which inhibits the electron flow from photosystem II to plastoquinone. As a result, the expression of many photosynthesis-related genes, including *LHCs* were not induced under orange light with DCMU (Supplementary Figs. 10b and 11). The number of DEGs with DCMU are shown in Fig. 5a, b right. About 69% of orange light-controlled genes are affected by DCMU (Fig. 5a). Blue light also gave DEGs although the affected number was smaller compared to orange light. There were almost no differences in dark conditions with/without DCMU (Supplementary Fig. 11a).

We summarized the DEGs induced by DCMU shown in Fig. 5b. In total 1,108 genes are indicated with their normalized expression values and ratios between two light conditions. The 119 DEGs seen in OBF (orange, Blue, and Far-red) light contain genes for elongation factor 2 kinase, Mcm2-7 hexameric complex component, DNA replication licensing factor, mcm4 component, and DNA replication protein psf2. Among these DEGs, we observed 45 genes that showed higher expression (>1.5x) in blue light compared to orange light and two genes with lower expression in blue light compared to orange light. Twenty-one genes showed higher expression in blue light compared to far-red light and 31 genes showed lower expression in orange light compared to far-red light.

Among the DEGs under monochromatic light conditions with DCMU was *DUC1*, which was significantly expressed in

**Table 1 Homologs of Arabidopsis photoreceptors and signal-transducing factors in chlorophytes with known genomes.**

Organism	PHY	Plant CRY	Plant-like CRY (Chlorophyte-specific CRY)	CRY-DASH	PIF	COP1	SPA	CK2 alpha	HY5	BIC	CIB1	FT	PHOT	UVR8	CSN1-8	EIN3	EBF1	FHY1,3	LAF1	DET1	DDB1a	CUL4
Angiospermae	4	2	-	1	8	1	4	4	1	2	1	1	2	1	10	1	1	2	1	1	1	1
Prasinophytes	DUC1	2 + DUC1	2	1	-	1	-	1	1	-	-	-	1	-	7	-	-	-	1	-	-	1
<i>Arabidopsis thaliana</i>																						
<i>Pycnococcus provasolii</i>																						
<i>Bathycoccus prasinos</i>																						
<i>Ostreococcus lucimarinus</i>																						
<i>Ostreococcus sp. RCC809</i>																						
<i>Ostreococcus tauri</i>																						
<i>Micromonas commoda</i>																						
<i>Micromonas pusilla</i>																						
<i>Chloropicon primus</i>																						
<i>Chlorella variabilis NC64A</i>																						
<i>Coccomyxa subellipsoidea</i>																						
<i>Volvox carterii</i>																						
<i>Chlamydomonas reinhardtii</i>																						
<i>Ulva mutabilis</i>																						

“-” indicates that there is no homologous gene in the genome annotation. Detailed gene names and criteria of homologous genes are provided as a Source Data file. COP1 constitutive photomorphogenic 1, SPA suppressor of phytochrome A, CK2 alpha casein kinase 2 alpha, HY5 elongated hypocotyl 5, BIC B light inhibitors of cryptochromes, CIB1 cryptochrome-interacting basic-helix-loop-helix 1, FT flowering locus 1, PHOT phototropin, UVR8 UV resistance locus 8, CSN cop9 signalosome subunit, EIN3 ethylene-insensitive 3, EBF1 EIN3-binding F box protein 1, FHY far-red elongated hypocotyl, LAF1 long after far-red light 1, DDB1a damaged DNA binding protein 1A, CUL4 cullin 4.

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 *lhcp* 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**

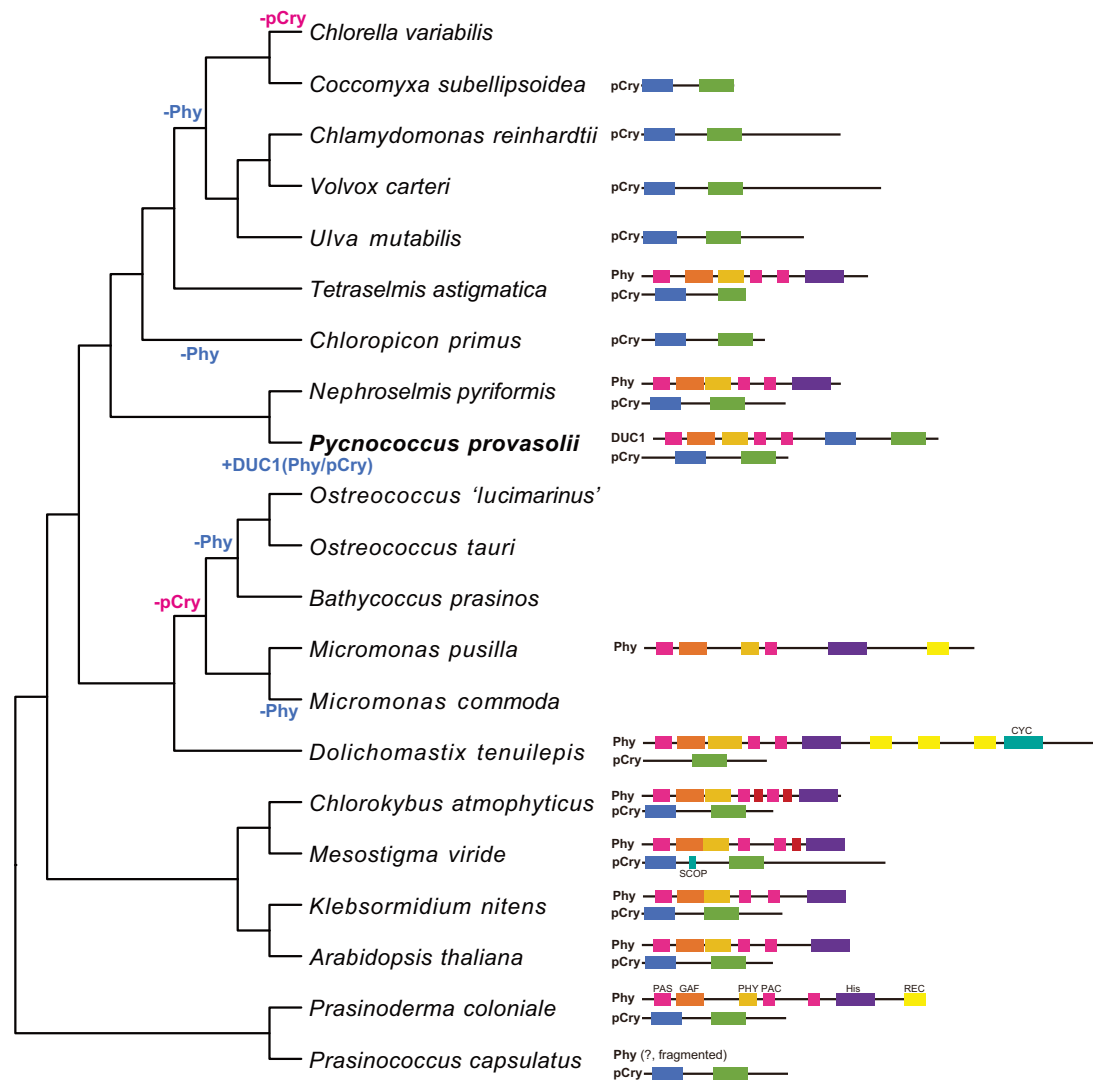
In this research, we have found a bifunctional photoreceptor, PpDUC1, composed of a fusion of PHY and CRY. In terms of evolution, it is often speculated that different domains of one organism’s protein are encoded by separate genes in another and this has been used successfully to speculate about direct physical interaction or indirect functional association<sup>22</sup>. It is reported that phyB and cry2 physically interact to transduce light signals for controlling flowering time in *Arabidopsis*<sup>23</sup>. Our discovery of PpDUC1 indicates that PHY and CRY interact actively to enable proper perception of light signals. Another example is Neochrome, which is found in ferns<sup>24</sup>, 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<sup>25</sup>.

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 OBF 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<sup>17</sup>. 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<sup>26</sup>. We did not observe light-dependent intracellular PpDUC1 translocation nor its enrichment in the nucleus by light in tobacco cells (Supplementary Fig. 8). Also, 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/Pfr photocycle, which is similar to the PHY molecules derived from other prasinophyte species, indicative of the same origin<sup>2</sup>. It is



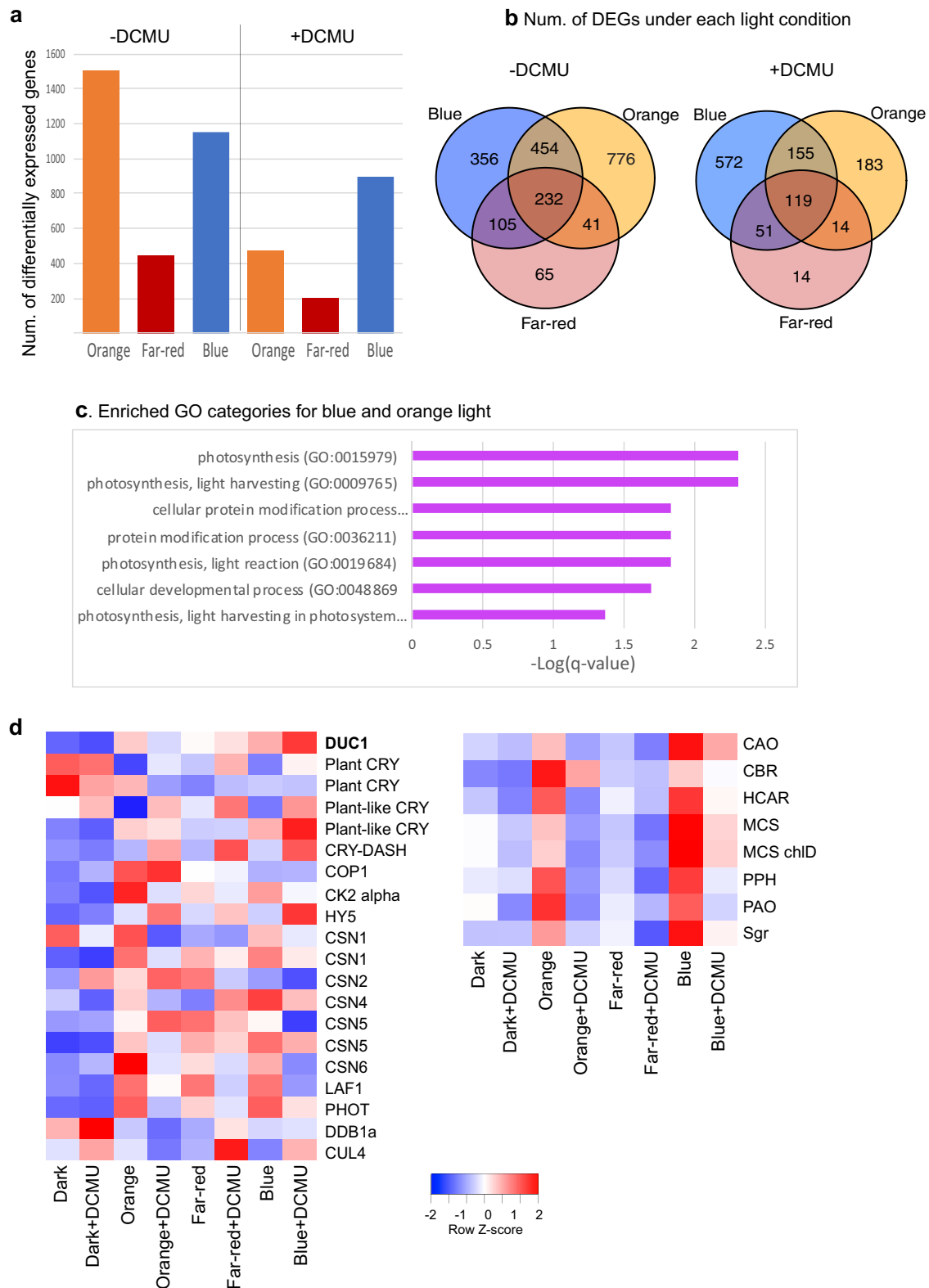
**Fig. 4 Evolutionary scenario of PHY and pCRY in chlorophytes.** The phylogenetic relationships of representative chlorophytes and the domain structures of their PHYs and pCRYs are shown. The tree topology is based on this study. The NCBI-nr database or MMETSP assemblies were searched for PHYs and pCRYs. Each domain is depicted by a different color.

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<sup>27</sup>. 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)<sup>28</sup>, 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<sup>9</sup> and pCRY<sup>29</sup>. Of the prasinophytes, *Tetraselmis*, *Nephroselmis*, *Micromonas*, *Dolichomastix*, and *Prasinoderma* possess genes for functional PHY<sup>1,2</sup>. 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<sup>9</sup>. Our phylogenetic analysis inclusive of DUC1 supports the idea that the last common ancestor of the Archaeplastida had a phytochrome<sup>17</sup>. In contrast, pCRY is widely shared in chlorophytes and streptophytes<sup>15</sup>, 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 *Pseudoscourfieldia*, 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 *Pseudoscourfieldia*). 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-divinylpheoporphyryn *a*<sub>5</sub> monomethyl ester) and an ability to adapt to spectral quality (blue and blue-violet) and low-light intensity<sup>5,8</sup>. Under high-light radiation, algae and land plants degenerate antenna complexes (LHCs and pigments) to decrease the absorbance of excess energy and prevent photodamage<sup>30</sup>. The degradation of Chl triggers degradation of LHCs<sup>31</sup>. 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.



protect photosynthesis from high-light irradiation. Therefore, *P. provasolii* 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 LHCs but also in PSI core antennae<sup>5,32</sup>. *P. provasolii* possesses LHCA, LHCB and a number of prasinophyte-specific LHCs (Lhcp) (Fig. 5). The Lhcps make complexes with Chl *a/b* and carotenoids<sup>33</sup>. Our transcriptome data showed that all LHC genes were upregulated and some Lhcps were strongly upregulated under orange- and blue-light irradiation (Fig. 5). This result is consistent with LHC regulation in land plants.

Our studies show the possible dynamic adaptation mechanisms to light spectra and intensities in *P. provasolii*, through the blue-, orange- and far-red-sensing DUC1. Acquisition of the dual functional photoreceptor, DUC1, has opened up a way for *P. provasolii* to widen its spectral utilization.

## Methods

**Prediction of CRY genes and identification of the host organism.** We used the assembled marine metagenome data sampled at Sendai Bay and the western subarctic Pacific Ocean 2012–2014 (<http://marine-meta.healthscience.sci.waseda.ac.jp/crest/metacrest/graphs/>) and checked the protein domains using InterProScan v5.40-77.0<sup>34</sup>. Proteins were defined as candidate CRYs if they had the following three domains: “DNA photolyase N-terminal (IPR006050)”, “Rossman-like alpha/beta/alpha sandwich fold (IPR014729)”, and “Cryptochrome/DNA photolyase, FAD-binding domain (IPR005101)”, following the dbCRY method<sup>35</sup>. We also excluded known CRYs that matched with the NCBI nr database<sup>36</sup> by a blast search<sup>37</sup>. To find the candidate host organism, we applied a blast search against MMETSP (The Marine Microbial Eukaryote Transcriptome Sequence Project) data<sup>38</sup>. The culture strain of *P. provasolii* NIES-2893 was obtained from the National Institute for Environmental Studies, Japan.

**Phylogenetic analyses of Phy and Cry regions of DUC1.** For phylogenetic analyses of DUC1, we divided it into Phy (1–3087 nucleotides) and Cry (3143–5073 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 mafft v7.453<sup>39</sup> with a linsi option, and ambiguous regions were trimmed using v1.4.rev15<sup>40</sup> with the option automated1. The trimmed dataset contained 457 and 353 amino acids for Phy and Cry regions, respectively. A model test was performed using ModelTest-NG v0.1.5<sup>41</sup>. Maximum likelihood analysis was performed using RAXML-NG v0.9.0<sup>42</sup> with 200 bootstrap replicates.

**Plasmid construction for spectral characterization of PpPHY and PpCRY.** The *Escherichia coli* strains Mach1 T1<sup>R</sup> (Thermo Fisher Scientific) and DH5 $\alpha$  (TaKaRa) were used for DNA cloning. These cells were grown on Lysogeny Broth (LB) agar medium at 37 °C. The transformed cells were selected by 20  $\mu$ g mL<sup>-1</sup> kanamycin or 100  $\mu$ g mL<sup>-1</sup> ampicillin.

The PpPHY region (1–1986 bp in PpDUC1) was cloned into *Nde* I and *Bam*H I sites of the pET28a vector with N-terminal His-tag (Novagen) using the Gibson Assembly System (New England Biolabs, Japan). The DNA fragment of the PpPHY region was amplified by PCR using KOD One<sup>TM</sup> PCR Master Mix (Toyobo Life Science) with a codon-optimized synthetic gene for expression in *E. coli* (Genscript) and an appropriate nucleotide primer set (Supplementary Table 2). The pET28a vector was amplified by PCR using DNA polymerase with template DNA and an appropriate nucleotide primer set (Supplementary Table 2).

The PpCRY region (3115–5073 bp in PpDUC1) was cloned into the *Eco*R I and *Xba* I sites of a pCold GST DNA vector with an N-terminal GST-tag (TaKaRa) using restriction enzymes and ligase. A DNA fragment of the PpCRY region was amplified by PCR using PrimeSTAR Max DNA Polymerase (TaKaRa) with genomic DNA from *P. provasolii* and an appropriate nucleotide primer set (Supplementary Table 2). All of the plasmid constructs were verified by nucleotide sequencing (FASMAC).

**Expression and purification of PpPHY fused with His-tag.** The *E. coli* strain C41 (Cosmo Bio) was used for His-tagged PpPHY (amino acid positions 1–662 in PpDUC1) expression through the pKT270, pKT271 and pKT272 constructs as biliverdin IX $\alpha$  (BV), phycocyanobilin (PCB) and phytychromobilin (P $\Phi$ B) synthetic systems, respectively<sup>11,43</sup>. Bacterial cells were grown in LB medium containing antibiotics (20  $\mu$ g mL<sup>-1</sup> kanamycin and 20  $\mu$ g mL<sup>-1</sup> chloramphenicol) at 37 °C. For protein expression, after the cells reached an optical density of 0.4–0.8 at

600 nm, isopropyl  $\beta$ -D-1-thiogalactopyranoside (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  $\times$  g for 15 min to collect the cells. The cells were resuspended in lysis 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. Homogenates were centrifuged at 165,000  $\times$  g for 30 min and then the supernatants were filtered through a 0.8  $\mu$ m cellulose acetate membrane before loading on to a nickel-affinity HisTrap HP column (GE Healthcare) using the ÄKTA pure 25 (GE Healthcare) system. The column was washed using the lysis buffer containing 100 mM imidazole and, then, His-tagged PpPHY incorporated with each chromophore was eluted with a linear gradient of the lysis buffer containing 100 to 400 mM imidazole (1 mL min<sup>-1</sup>, total 15 min). After incubation with 1 mM EDTA for 1 h, His-tagged PpPHY was dialyzed against the lysis buffer with 1 mM dithiothreitol (DTT) to remove EDTA and imidazole<sup>44,45</sup>. The purified proteins were dialyzed against lysis buffer A containing 1 mM dithiothreitol (DTT). Protein concentration was determined by the Bradford method.

**Expression and purification of PpCRY fused with GST-tag.** The *E. coli* strain C41 was used for GST-tagged PpCRY (amino acid positions 1039–1690 in PpDUC1) expression through the pG-KJE8 construct as a chaperone expression system (TaKaRa)<sup>13</sup>. Bacterial cells were grown in 8 L LB medium containing 500  $\mu$ g mL<sup>-1</sup> L-arabinose and antibiotics (5 ng mL<sup>-1</sup> tetracycline, 100  $\mu$ g mL<sup>-1</sup> ampicillin, and 20  $\mu$ g mL<sup>-1</sup> chloramphenicol) at 37 °C. For protein expression, once 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  $\times$  g for 15 min to collect the cells. They were resuspended in lysis buffer B (20 mM HEPES–NaOH pH 7.5, 500 mM NaCl, 5 mM DTT and 10% (w/v) glycerol), and then disrupted using a homogenizer at 12,000 psi. The homogenate was centrifuged at 165,000  $\times$  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 lysis buffer B to remove unbound proteins, and GST-tagged PpCRY 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–polyacrylamide gel electrophoresis analysis for purified PpPHY and PpCRY.** Purified proteins were diluted into 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<sup>46</sup>. These bands were visualized through a 600-nm long-path filter upon excitation with blue ( $\lambda_{\text{max}} = 470$  nm) and green ( $\lambda_{\text{max}} = 527$  nm) light through a 562 nm short-path filter using WSE-6100 LuminoGraph (ATTO) and WSE-5500 VariRays (ATTO) machines. After the monitoring, the gels were stained with Coomassie Brilliant Blue R-250.

## UV-Vis spectroscopic analysis to monitor photocycles of PpPHY and PpCRY.

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 PpPHY.** Sample solutions containing the native PpPHY 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 PpPHY and standard proteins<sup>47–49</sup>.

**Assignment of the chromophore incorporated into PpCRY.** Trichloroacetic acid was added to the sample solution containing the native PpCRY 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  $\times$  g for 5 min, and then 50  $\mu$ L of the supernatant were injected into a high-performance liquid chromatograph (Prominence HPLC system, Shimadzu). Released chromophore included in the supernatant was eluted isocratically (flow rate, 1 mL min<sup>-1</sup>) with a solvent (MeOH/20 mM KH<sub>2</sub>PO<sub>4</sub> = 20/80) and separated using a reverse-phase HPLC column (InertSustainSwift C18, 4.6 i.d.  $\times$  250 mm, 5  $\mu$ m; GL Sciences) at 35 °C. The chromophore was detected by its absorption at 360 nm. Assignment of the chromophores was performed based on their retention times ( $t_R$ ) compared with standard compounds; 5,10-methenyltetrahydrofolate chloride (MTHF chloride, Schircks Laboratories), flavin adenine dinucleotide (FAD, Tokyo Chemical Industry), flavin mononucleotide (FMN, Tokyo Chemical Industry) and riboflavin

(RF, Tokyo Chemical Industry). The sample and standard compounds were handled under dark conditions.

**Protein blot of PpDUC1-GFP and PpDUC1-HA expressed in plants.** Total protein was extracted from 80 mg of *A. thaliana* whole plants and *N. benthamiana* leaf discs using 240  $\mu$ L of protein extraction buffer composing 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 0.1% 2-mercaptoethanol, 5% glycerol, 0.5% Triton X-100, and cOmplete™ 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 precipitates. Ten microliters of supernatant was added with 5  $\mu$ L 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 electroblotted onto a polyvinylidene difluoride (PVDF) membrane (Millipore) in the blotting buffer composing 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 1 $\times$ TBST composing 137 mM NaCl, 2.68 mM KCl, 25 mM Tris-HCl, pH 7.4, 0.1 w/v% Tween-20. For detection of PpDUC1-HA, skim-milk blocked membrane was incubated with anti-HA peroxidase conjugate (1201381900, Roche 1:1000 dilution) for 1.5 h. For detection of PpDUC1-GFP, blocked membrane was incubated with rabbit anti-GFP antibody (TP401, Torrey Pines Biolabs Inc., 1:2000) for 20 h. The membrane was rinsed 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 membranes were rinsed three times for 10 min each with 1 $\times$ TBST and after incubation with enhanced chemiluminescence reagents (ECL Select™ Western Blotting Detection Reagents, GE Healthcare Corp.) They were processed with a chemi-luminescent image analyzer (Chem Doc XRS plus, BIO-RAD).

**Constructs for transient transformation in *N. benthamiana*.** For infiltration into the leaf epidermal cells of *N. benthamiana*, the full-length open reading frames of PpDUC1, and parts of PpPHY (1–3621 bp in DUC1) and PpCRY (3115–5070 bp in DUC1 without stop codon) were amplified from *P. provasolii* genomic DNA by PCR with each 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 pEarleyGate 103 destination vector<sup>50</sup> to fuse in frame with GFP and 6 $\times$  His by LR reaction. A synthetic gene of *Arabidopsis thaliana* HY5 (At5g11260) fused to mCherry (Sequence ID: MH976504.1) 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<sup>51</sup>.

**Transient expression and observation of PpDUC1-GFP in *N. benthamiana* leaves.** Each construct was transformed into Agrobacterium GV3101. These agrobacteria were infiltrated into the leaf epidermal cells of *N. benthamiana* by eluting with 10 mM MES (pH 5.6) with 200  $\mu$ M acetosyringone into the undersides of leaves of *N. benthamiana* plants<sup>52</sup>. A Zeiss LSM880 Airyscan Fast-mode microscope (Zeiss) with AxioObserver 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 556 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.

**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 (~50  $\mu$ mol photons  $m^{-2} s^{-1}$ ) with a 14 h:10 h light:dark cycle. DNA was extracted as described in Suzuki et al.<sup>53</sup> Cells were collected by gentle 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  $\mu$ L of CTAB extraction buffer<sup>54</sup> at 65 °C for 1 h. DNA was separated by mixing with 500  $\mu$ L of chloroform and centrifuging at 20,000  $\times$ g for 1 min; it was concentrated by standard EtOH precipitation. The paired-end library was constructed using a TruSeq DNA PCR-free Kit (Illumina Inc., San Diego, CA, USA), according to the manufacturer's protocol. The mate-pair library was constructed using a Nextera Mate Pair Library Prep Kit (Illumina), according to the manufacturer's "gel-free" protocol. The paired-end and mate-pair libraries were sequenced using the MiSeq System (Illumina) with MiSeq Reagent Kits v3 (300 bp  $\times$  2) (Illumina). For nanopore sequencing, the library was constructed using a Rapid Sequencing Kit (SQK-RAD003, Oxford Nanopore Technologies, Oxford, UK). Sequencing was performed using a MinION with a SpotON Flow Cell (FLO-MIN106).

We sequenced 40,351,502 paired-end (9.7 Gbp), 10,652,248 mate-pair (1.7 Gbp), and 90,971 nanopore reads (0.50 Gbp, N50 = 18.8 Kbp). The raw reads were assembled into 47 scaffolds (22.9 Mbp) with one gap using MaSuRCA 3.3.2<sup>55</sup>. The nanopore reads were corrected using CONSENT v1.1.2<sup>56</sup>. The paired-end reads were corrected using Trimmomatic version 0.36<sup>57</sup>. The assembly was re-scaffolded by the corrected nanopore reads using Fast-SG<sup>58</sup> with the k-mer size = 70. Gaps were filled using LR\_Gapcloser v1.1<sup>59</sup> with the corrected nanopore reads. Finally, we mapped the corrected paired-reads using BWA version 0.7.17<sup>60</sup>, and polished the assembly using Pilon 1.22<sup>61</sup>. The polishing was performed twice.

We removed the two sequences of the complete chloroplast and mitochondrial genomes. Genome annotation was performed using the funannotate pipeline v1.5.3 (<https://github.com/nextgenusfs/funannotate>). For construction of the gene models, we used our RNA-seq reads (19 Gbp) under various light conditions (described in the transcriptome section) after read correction with Trimmomatic<sup>57</sup>. Functional annotation was performed using the eggNOG web server<sup>62</sup>. To check expression of the gene models, we mapped the RNA-seq reads used for the gene-model construction to the coding sequences (CDSs) using minimap2 2.17<sup>63</sup>, and 99.5% and 98.4% of the predicted CDSs were mapped by at least one and five RNA-seq reads, respectively. The completeness of the gene prediction was assessed using BUSCO v2/v31 via gVolante<sup>61</sup>.

**Transcriptome analysis under monochromatic light.** The cells of *P. provasolii* NIES-2893 were grown at 19.5 °C with shaking at 120 rpm in a light intensity of 13  $\mu$ mol  $m^{-2} s^{-1}$  under a LD12:12 light/dark cycle. To monitor gene expression during irradiation of monochromatic light, cells were synchronized by a 12 h dark treatment before being exposed to monochromatic constant light. Light treatments were performed using LEDs (Nippon Medical & Chemical Instruments Co., Ltd., Japan) of specific wavelengths at an intensity of 7  $\mu$ mol  $m^{-2} s^{-1}$ . The following settings were used: for blue light, the peak was at 447 nm; for orange light, at 596 nm, and for far-red light, at 697 nm. The electron transport inhibitor 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) was added to the samples before dark treatment. The cells were treated with 40  $\mu$ M DCMU in 0.1% DMSO or with only 0.1% DMSO (control). After light treatment for 2 h with or without the addition of DCMU, cells were harvested and frozen with liquid nitrogen. After cell disruption using glass beads, total RNA was extracted from *P. provasolii* using the TRIzol reagent (Thermo Fisher Scientific Inc, USA). The extracted total RNA was purified using RNeasy Plant Mini Kits (QIAGEN, The Netherlands). Libraries for RNA-seq analysis were synthesized using Illumina® TruSeq® Stranded mRNA Sample Preparation Kits (Illumina Inc, USA) and sequenced on an Illumina HiSeq 2000 using directed paired-end technology. The sequenced reads were mapped to the *P. provasolii* genome with STAR v2.5.0c<sup>62</sup> after trimming the low-quality reads (Phred quality of  $\leq$ 20) with the FASTX-Toolkit-0.0.14 ([http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)). Read counts were normalized with DESeq 1.42.0 in an R package<sup>63</sup>. Differentially expressed genes (DEGs) were defined as >1.5-fold for gene expression with a *q*-value < 0.05.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## 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 DDBJ/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://matsui-lab.riken.jp/Browse/index.html?data=data%2Fpyncococcus>. The algal strains are available in the NIES collection: *P. provasolii* (NIES-2893) and *P. marina* (NIES-1419). The publicly available datasets in this study include: PhycoCosm database [<https://phyocosm.jgi.doe.gov/phyocosm/home>], Ocean Monitoring Database of Sendai Bay and the western subarctic Pacific Ocean [<http://marine-meta.healthscience.sci.waseda.ac.jp/crest/metacrest/graphs/>], and MMETSP (The Marine Microbial Eukaryote Transcriptome Sequence 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

- Li, F. W. et al. Phytochrome diversity in green plants and the origin of canonical plant phytochromes. *Nat. Commun.* **6**, 7852 (2015).
- Rockwell, N. C. et al. Eukaryotic algal phytochromes span the visible spectrum. *Proc. Natl Acad. Sci. USA* **111**, 3871–3876 (2014).
- Rockwell, N. C., Su, Y. S. & Lagarias, J. C. Phytochrome structure and signaling mechanisms. *Annu. Rev. Plant Biol.* **57**, 837–858 (2006).
- Lemieux, C., Otis, C. & Turmel, M. Six newly sequenced chloroplast genomes from prasinophyte green algae provide insights into the relationships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. *BMC Genomics* **15**, 857 (2014).
- Guillard, R. R. L., Keller, M. D., O'Kelly, C. J. & Floyd, G. L. *Pyncococcus provasolii* gen. et sp. nov., a coccoid prasinanthin-containing phytoplankton from the Western North Atlantic and Gulf of Mexico. *J. Phycol.* **27**, 39–47 (1991).

6. Shiozaki, T. et al. Eukaryotic Phytoplankton contributing to a seasonal bloom and carbon export revealed by tracking sequence variants in the Western North Pacific. *Front. Microbiol.* **10**, 2722 (2019).
7. Tragin, M., dos Santos, A. L., Christen, R. & Vaulot, D. Diversity and ecology of green microalgae 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 picoplankter *Pycnococcus provasolii* Guillard (clone  $\Omega 48-23$ ) (Chlorophyta) to variations in irradiance, photoperiod and temperature. *J. Exp. Mar. Bio. 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 cryptochrome 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. Microbiol.* **66**, 884–889 (2000).
14. Berndt, A. et al. A novel photoreaction mechanism for the circadian blue light photoreceptor *Drosophila* cryptochrome. *J. Biol. Chem.* **282**, 13011–13021 (2007).
15. Kottke, T., Oldemeyer, S., Wenzel, S., Zou, Y. & Mittag, M. Cryptochrome photoreceptors in green algae: Unexpected versatility of mechanisms and functions. *J. Plant Physiol.* **217**, 4–14 (2017).
16. Lin, C. et al. Association of flavin adenine dinucleotide with the Arabidopsis blue light receptor CRY1. *Science* **269**, 968–970 (1995).
17. Duanmu, D. et al. Marine algae and land plants share conserved phytochrome signaling systems. *Proc. Natl Acad. Sci. USA* **111**, 15827–15832 (2014).
18. Yamaguchi, R., Nakamura, M., Mochizuki, N., Kay, S. A. & Nagatani, A. Light-dependent translocation of a phytochrome B-GFP fusion protein to the nucleus in transgenic *Arabidopsis*. *J. Cell Biol.* **145**, 437–445 (1999).
19. Guo, H., Yang, H., Mockler, T. C. & Lin, C. Regulation of flowering time by *Arabidopsis* photoreceptors. *Science* **279**, 1360–1363 (1998).
20. Han, X. et al. Origin and evolution of core components responsible for monitoring light environment changes during plant terrestrialization. *Mol. Plant* **12**, 847–862 (2019).
21. Li, L. et al. The genome of *Prasinoderma coloniale* unveils the existence of a third phylum within green plants. *Nat. Ecol. Evol.* **4**, 1220–1231 (2020).
22. Marcotte, E. M. et al. Detecting protein function and protein-protein interactions from genome sequences. *Science* **285**, 751–753 (1999).
23. Más, P., Devlin, P. F., Panda, S. & Kay, S. A. Functional interaction of phytochrome B and cryptochrome 2. *Nature* **408**, 207–211 (2000).
24. Kawai, H. et al. Responses of ferns to red light are mediated by an unconventional photoreceptor. *Nature* **421**, 287–290 (2003).
25. Jaedicke, K., Lichtenthaler, A. L., Meyberg, R., Zeidler, M. & Hughes, J. A phytochrome-phototropin light signaling complex at the plasma membrane. *Proc. Natl Acad. Sci. USA* **109**, 12231–12236 (2012).
26. Xu, X., Paik, I., Zhu, L. & Huq, E. Illuminating progress in phytochrome-mediated light signaling pathways. *Trends Plant Sci.* **20**, 641–650 (2015).
27. Fushimi, K. & Narikawa, R. Cyanobacteriochromes: photoreceptors covering the entire UV-to-visible spectrum. *Curr. Opin. Struct. Biol.* **57**, 39–46 (2019).
28. Essen, L. O., Mailliet, J. & Hughes, J. The structure of a complete phytochrome sensory module in the Pr ground state. *Proc. Natl Acad. Sci. USA* **105**, 14709–14714 (2008).
29. Fortunato, A. E., Annunziata, R., Jaubert, M., Bouly, J.-P. & Falciatore, A. Dealing with light: the widespread and multitasking cryptochrome/photolyase family in photosynthetic organisms. *J. Plant Physiol.* **172**, 42–54 (2015).
30. Schöttler, M. A. & Tóth, S. Z. Photosynthetic complex stoichiometry dynamics in higher plants: environmental acclimation and photosynthetic flux control. *Front. Plant Sci.* **5**, 188 (2014).
31. Wu, S. et al. NON-YELLOWING2 (NYE2), a close paralog of NYE1, plays a positive role in chlorophyll c degradation in *Arabidopsis*. *Mol. Plant* **9**, 624–627 (2016).
32. Kunugi, M. et al. Evolution of green plants accompanied changes in light-harvesting systems. *Plant Cell Physiol.* **57**, 1231–1243 (2016).
33. Swingle, W. D. et al. Characterization of photosystem I antenna proteins in the prasinophyte *Ostreococcus tauri*. *Biochim. Acta* **1797**, 1458–1464 (2010).
34. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
35. Kim, Y. M. et al. dbCRY: a Web-based comparative and evolutionary genomics platform for blue-light receptors. *Database* **2014**, bau037 (2014).
36. O’Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
37. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinforma.* **10**, 421 (2009).
38. Keeling, P. J. et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **12**, e1001889 (2014).
39. Katoh, K. & Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **9**, 286–298 (2008).
40. 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).
41. Darriba, 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).
42. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
43. Mukougawa, K., Kanamoto, H., Kobayashi, T., Yokota, A. & Kohchi, T. Metabolic engineering to produce phytochromes with phytochromobilin, phycocyanobilin, or phycoerythrobilin chromophore in *Escherichia coli*. *FEBS Lett.* **580**, 1333–1338 (2006).
44. Fushimi, K. et al. Rational conversion of chromophore selectivity of cyanobacteriochromes to accept mammalian intrinsic biliverdin. *Proc. Natl Acad. Sci. USA* **116**, 8301–8309 (2019).
45. Hasegawa, M. et al. Molecular characterization of D. *J. Biol. Chem.* **293**, 1713–1727 (2018).
46. Berkelman, T. R. & Lagarias, J. C. Visualization of bilin-linked peptides and proteins in polyacrylamide gels. *Anal. Biochem.* **156**, 194–201 (1986).
47. Fushimi, K. et al. Photoconversion and fluorescence properties of a red/green-type Cyanobacteriochrome AM1\_C0023g2 that binds not only phycocyanobilin but also biliverdin. *Front. Microbiol.* **7**, 588 (2016).
48. Narikawa, R. et al. A biliverdin-binding cyanobacteriochrome from the chlorophyll d-bearing cyanobacterium *Acaryochloris marina*. *Sci. Rep.* **5**, 7950 (2015).
49. Eichenberg, K. et al. *Arabidopsis* phytochromes C and E have different spectral characteristics from those of phytochromes A and B. *FEBS Lett.* **470**, 107–112 (2000).
50. Earley, K. W. et al. Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616–629 (2006).
51. Kojima, S. et al. A binary vector plasmid for gene expression in plant cells that is stably maintained in *Agrobacterium* cells. *DNA Res.* **6**, 407–410 (1999).
52. Lewis, J. D., Abada, W., Ma, W., Guttman, D. S. & Desveaux, D. The HopZ family of *Pseudomonas syringae* type III effectors require myristoylation for virulence and avirulence functions in *Arabidopsis thaliana*. *J. Bacteriol.* **190**, 2880–2891 (2008).
53. Suzuki, S., Yamaguchi, H., Nakajima, N. & Kawachi, M. *Raphidocelis subcapitata* (= *Pseudokirchneriella subcapitata*) provides an insight into genome evolution and environmental adaptations in the Sphaeropleales. *Sci. Rep.* **8**, 8058 (2018).
54. Lang, B. F. & Burger, G. Purification of mitochondrial and plastid DNA. *Nat. Protoc.* **2**, 652–660 (2007).
55. Zimin, A. V. et al. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA megareads algorithm. *Genome Res.* **27**, 787–792 (2017).
56. Morisse, P., Marchet, C., Limasset, A., Lecroq, T. & Lefebvre, A. Scalable long read self-correction and assembly polishing with multiple sequence alignment. *Sci. Rep.* **11**, 761 (2021).
57. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
58. Di Genova, A., Di Genova, A., Ruz, G. A., Sagot, M.-F. & Maass, A. Fast-SG: an alignment-free algorithm for hybrid assembly. *GigaScience* **7**, giy048 (2018).
59. Xu, G.-C. et al. LR\_GapCloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *GigaScience* **8**, giy157 (2019).
60. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
61. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
62. Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
63. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

## 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