Unveiling of a novel bifunctional photoreceptor, Dualchrome1, isolated from a cosmopolitan green alga

Yuko Makita

RIKEN Center for Sustainable Resource Science

Shigekatsu Suzuki

Center for Environmental Biology and Ecosystem Studies, National Institute of Environmental Studies

Keiji Fushimi

Shizuoka University

Setsuko Shimada

RIKEN Center for Sustainable Resource Science

Aya Suehisa

RIKEN Center for Sustainable Resource Science

Manami Hirata

RIKEN Center for Sustainable Resource Science

Tomoko Kuriyama

RIKEN Center for Sustainable Resource Science

Yukio Kurihara

RIKEN

Hidehumi Hamasaki

RIKEN Center for Sustainable Resource Science

Kazutoshi Yoshitake

The University of Tokyo

Tsuyoshi Watanabe

Japan Fisheries Research and Education Agency

Takashi Gojobori

King Abdullah University of Science and Technology

Tomoko Sakami

Japan Fisheries Research and Education Agency

Rei Narikawa

Shizuoka University <https://orcid.org/0000-0001-7891-3510>

Haruyo Yamaguchi

National Institute for Environmental Studies <https://orcid.org/0000-0001-5269-1614>

Masanobu Kawachi

National Institute for Environmental Studies
Minami Matsui (✉ minami@riken.jp)
RIKEN Center for Sustainable Resource Science

Article

Keywords: Blue-light Sensing, Chimeric Gene, Prasinophyte, Light adaptation Mechanisms, Light-harvesting Complex

Posted Date: October 14th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-88004/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on June 16th, 2021. See the published version at <https://doi.org/10.1038/s41467-021-23741-5>.

**Title: Unveiling of a novel bifunctional photoreceptor, Dualchrome1, isolated
from a cosmopolitan green alga**

Authors: Yuko Makita^{1†}, Shigekatu Suzuki^{2†}, Keiji Fushimi^{3†}, Setsuko Shimada^{1†}, Aya Suehisa¹, Manami Hirata¹, Tomoko Kuriyama¹, Yukio Kurihara¹, Hidefumi Hamasaki^{1,4}, Kazutoshi Yoshitake⁵, Tsuyoshi Watanabe⁶, Takashi Gojobori⁷, Tomoko Sakami⁸, Rei Narikawa³, Haruyo Yamaguchi², Masanobu Kawachi², Minami Matsui^{1*}.

Affiliations:

¹Synthetic Genomics Research Group, RIKEN Center for Sustainable Resource Science, Yokohama 230-0045, Japan.

²Center for Environmental Biology and Ecosystem Studies, National Institute for Environmental Studies, Tsukuba 305-8506, Japan

³Graduate School of Integrated Science and Technology, Shizuoka University, 422-8529, Shizuoka, Japan; Research Institute of Green Science and Technology, Shizuoka University, Ohya, Suruga-ku, Shizuoka 422-8529, Japan; Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Saitama 332-0012, Japan.

⁴Yokohama City University Kihara Institute for Biological Research, Totsuka, Yokohama, Kanagawa, Japan.

⁵Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi,
Bunkyo, Tokyo, 113-8657, Japan

⁶Tohoku National Fisheries Research Institute, Japan Fisheries Research and Education Agency,
Shiogama, Miyagi 985-0001, Japan.

⁷Computational Bioscience Research Center, King Abdullah University of Science and
Technology, Thuwal, Kingdom of Saudi Arabia.

⁸Research Center for Aquaculture Systems, National Research Institute of Aquaculture, Japan
Fisheries Research and Education Agency, Minami-ise, Mie 516-0193, Japan.

*Corresponding author. Email: minami@riken.jp

[†]These authors contributed equally to this study.

Abstract

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. A search for blue-light photoreceptors in the marine metagenome uncovered a novel 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. Its complete genome revealed that *P. provasolii* facilitates light adaptation mechanisms via pheophorbide *a* oxygenase (Pao) and prasinoxanthin. 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.

Main text:

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)¹. 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 light². They have the ability to photoconvert between red-absorbing Pr and far-red-absorbing Pfr, and this conformational change enables interactions with signaling partners³. 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.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 novel 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, Pycnococcaceae (or prasinophytes clade V⁴) and was originally discovered in the pycnocline⁵. *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 intensities⁶. Thus, prasinophytes are key to understanding the diversity and evolutionary history of the light response system in the Viridiplantae.

Environmental DNA researches show *P. provasolii* lives at depths in the range of 0-100 m and in varying regions of the marine^{6,7}. It also has a unique pigment composition (prasinoxanthin and Mg-2,4-divinylphaeoporphyrin α_5) 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^{5,8}.

We unveil herewith DUC’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 sequenced the complete genome of *P. provasolii* and examined its general 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 novel photoreceptor *DUC1* from marine metagenome data

Since *CRY*s 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 (Fig. 1B). 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, *Pycnococcus provasolii*. Fortunately, a culture strain of it was stocked in the NIES collection as *P. provasolii* NIES-2893 (Fig. 1C). Metagenome data also supported the fact that there was enrichment of this species in the subsurface chlorophyll maximum layer of stations C12 and A21 (Fig. S1). After amplification of this gene from NIES-2893 by RT-PCR, we finally concluded that it does have both the *PHY* and *CRY* domains (5,073 bp, 183kDa protein) and designated it as *Dualchrome1* or *DUC1* (Fig. S2A).

Phylogenetic analysis using the *P. provasolii* *PHY* (PpPHY) and *CRY* (PpCRY) domains showed they branched with those of other chlorophytes (Fig. S3, S4), 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) (Fig. S2B, C).

PpDUC1 senses blue, orange and far-red light

To verify whether DUC1 possesses photoconverter 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 (Fig. S5A). 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 (Fig. S5A)⁹. The GAF Cys ligates to C3¹ of phycocyanobilin (PCB) or phytochromobilin (PΦB), whereas the N-terminal Cys ligates to C3² 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 (Fig. S6AB). 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)¹⁰. The purified PpPHY covalently bound PCB (Fig. S5B, Fig. S6C) 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, Table S1). Dark reversion from the photoproduct state to the dark state was not seen (Fig. S7A). 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 (Fig. S6C-E). 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) (Fig. S5C). Some redox state forms of FAD, such as oxidized FAD (FAD_{OX}), anion radical FAD (FAD^{•-}), neutral radical FAD (FADH[•]) and reduced FAD (FAD_{red}H⁻), are observed in the photocycle of CRYs (Fig. S5C)¹¹. As the PpCRY accumulated as insoluble inclusion bodies in the normal expression system, we expressed PpCRY in *E. coli* harboring the chaperon co-expression system (C41_pG-KJE8). The purified PpCRY non-covalently bound FAD (Fig. S5D, Fig. S6C). The protein showed a blue-absorbing (Pb) form (λ_{max} , 447 nm) before light irradiation (Fig. 2C, Table S1). The spectral form in the dark state was similar to that of the FAD_{OX}-bound CRYs¹¹. Blue-light irradiation resulted in conversion to a UV-absorbing (Puv) form (λ_{max} , 366 and 399 nm) (Fig. 2C, Table S1). The spectral form in the photoproduct state was similar to that of the FAD^{•-}-bound CRYs¹¹. Although we further irradiated the protein with UV-to-blue light, no spectral change was observed. Instead, Puv-to-Pb dark reversion occurred (Fig. S7B). To conclude, PpCRY showed photoconversion from the FAD_{OX}-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*¹², rather than CraCRY from the green algae *Chlamydomonas reinhardtii*¹³ and AtCRY1 and 2 from the land plant *Arabidopsis thaliana*¹⁴.

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 partly complements photoreceptor mutation in *Arabidopsis*

To address the *in vivo* function of PpDUC1, we examined whether PpDUC1 complements the defective phenotypes of *Arabidopsis* photoreceptor mutants. As PpDUC1 shows the highest similarity with AtPHYB (28% identity) and AtCRY2 (38% identity), PpDUC1 was expressed in *Arabidopsis phyB* and *cry2* mutants (Fig. S8A). Since AtPHYB and AtCRY2 mediate light inhibition of hypocotyl elongation in seedlings^{15,16}, *phyB* and *cry2* show a long-hypocotyl phenotype compared with wild type. We observed that *PpDUC1* partly suppressed the long-hypocotyl phenotype of *cry2* under blue light (Fig. 3AB), although no phenotype was seen in darkness (Fig. 3C). The *phyB* mutants expressing *PpDUC1* did not show any apparent effects on hypocotyl elongation under orange and red light (Fig. S9). Since PpPHY can incorporate PΦB *in vitro* with orange shift (Fig.S5B), it is possible that PpDUC1 cannot transduce downstream signals in *Arabidopsis*.

PpDUC1 localizes at nucleus in tobacco leaves

It is reported that light conditions change the intracellular localization of photoreceptors. 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. 3D). Expression of the GFP-fused proteins was confirmed by protein-blot analysis (Fig. 3E). *Arabidopsis phyB* is reported to localize in the nucleus after light irradiation¹⁷ but, although PpPHY::GFP contains all the PHY sequences homologous to AtPHYB, it localized in the cytoplasm (Fig. 3F). *Arabidopsis cry2* localizes in the nucleus¹⁸ and PpCRY::GFP was found mainly here with a weak signal in the cytoplasm. PpDUC1::GFP localized in the nucleus.

***P. provasolii* responds to orange 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 472 genes and 102 genes changed after orange and far-red light irradiation, respectively. Since there is no photoreceptor that absorbs orange and/or far-red light in *P. provasolii*, these changes of expression may be through photoperception by PpDUC1. Orange and blue light commonly regulate 388 genes (Fig. 4A). All of these genes fluctuate similarly in both orange and blue light, indicating the involvement of PpDUC1. GO enrichment analysis showed that many of their categories are related to photosynthesis or light responses (Fig. 4B). In *P. provasolii*, expression of the light-harvesting complex genes (LHC), including 16 duplicated Lhcp, are induced by both orange and blue light, not by far-red light (Fig. 4C).

***P. provasolii* possesses a unique gene repertoire for adaptation to various light conditions**

We sequenced the complete genome of *P. provasolii* NIES-2893 (Fig. 1CD). The reads were assembled into 43 scaffolds without gaps. The assembly showed high completeness (Table S2, S3; Supporting text), and the scaffolds likely correspond to chromosomes. The genome possesses some unique features (Fig. S10–S24; Supporting Text). The genome of *P. provasolii* is 22.7 Mbp, which is similar to other prasinophytes (12.5–22.0 Mbp) (Table 1). We predicted and annotated 11,337 genes in the genome (Table 1) and compared orthogroups¹⁹ among some prasinophytes: *P. provasolii*, *Chloropicon primus*, and five species in Mamiellophyceae (Fig.

S22). There were 3,996 conserved orthogroups (47.8% of *P. provasolii* genes) and a large number of unique ones (3,636 orthogroups, 43.5% of the total).

We annotated five *CRY*s and one *DUC1* but there is no *PHY* in the genome. Of the *CRY* genes, two and the *PpCRY* region of *PpDUC1* belong to the plant *CRY*s (*pCRY*¹³). These three genes form a monophyly and branch at the basal position of those of other chlorophytes (Fig. S3). Chlorophyte *CRY*s are a sister group of the streptophyte *pCRY*s, including *AtCRY1* and *AtCRY2*. The *PpPHY* region is a monophyly with other prasinophytes and streptophytes (Fig. S4). In the prasinophytes, the genomes of some mamiellophyceans lack genes for *pCRY* and *PHY* (Fig. 5).

The predicted genes suggest that several important light signal transduction genes in land plants are conserved. We detected *COP1*, *CK2alpha* and *HY5* homologs but not those for *PIF* and *SPA1* for red light, nor *CIB1* or *BIC* for blue-light signal transduction. 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 S5).

In addition to photoreceptor and signaling genes, other genes encoded related to adaptation to different light conditions. We predict that *P. provasolii* possesses 165 putative flagellar proteins (Table S6). These genes compose a full set for flagellate structure and maintenance, which indicates the existence of a flagellate stage of *P. provasolii*, as described in a previous observation⁵. These results indicate that *P. provasolii* can move to favourable environments, including those influenced by light conditions.

We found an expansion of some gene families (Table S7–S9; Supporting text). In particular, the *P. provasolii* genome encoded a large gene family for a unique prasinophyte-specific light-harvesting complex (Lhcp). *P. provasolii* possesses 16 genes for Lhcp, and 10 of these are clustered in specific regions of chromosome 29 (Fig. S23). These 10 genes had almost the same

sequences and were monophyletic (Fig. 6A). This suggests that the *lhcb* gene on chromosome 10 was duplicated on chromosome 29 in *P. provasolii*. Similar duplications of *lhcb* genes are observed in Mamiellophyceae, however, these have occurred independently to *P. provasolii*.

Interestingly, the *P. provasolii* genome encodes a gene for pheophorbide *a* oxygenase (*Pao*), which is related to chlorophyll degradation. This gene is monophyletic with cyanobacteria and streptophytes, while other prasinophytes do not possess it (Fig. 6). *P. provasolii* also possesses a gene for magnesium dechelataase (*Sgr*), but it lacks one for red chlorophyll catabolite reductase, indicating that *P. provasolii* can degenerate chlorophyll *a* to red chlorophyll catabolites.

Discussion

In this research, we have found a novel dual photoreceptor, PpDUC1, comprised 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 has been used successfully to speculate about direct physical interaction or indirect functional association²⁰. It is reported that phyB and cry2 physically interact to transduce light signals for controlling flowering time in *Arabidopsis*²¹. 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 ferns²², 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.

Possible function of PpDUC1

Expression analysis of *P. provasolii* enabled observation of gene expression under blue and orange light. It was found that 388 genes are expressed under these wavelengths indicating the involvement of PpDUC1 (Fig. 4). However, 76 and 49 genes are induced only by orange or far-red light, respectively, and only by the PHY domain of PpDUC1. PpDUC1 localizes exclusively in the nucleus under white light in tobacco cells (Fig. 3F) while the PpPHY domain mainly localizes in the cytoplasm. PpPHY may have a different function from the AtPHYs. This is consistent with the complementation assay that showed PpDUC1 partly complements *cry2* but it does not complement *phyB* (Fig. 3, Fig. S8). In the *P. provasolii* genome, there is HY5 but no PIF homologs (Table S5) so PpDUC1 may have a unique signal transduction for light-dependent gene expression.

Possible mechanism for blue-shifted spectral property of PpPHY

Photoreceptors of the Chlorophyta perceive various wavelengths of light to utilize sunlight to direct their development, and land plants also have some of these receptors. Whilst most PHYs of land plants perceive 660 nm and 730 nm for their photoconversion, PHYs of the Chlorophyta (among them those of prasinophytes) perceive various wavelengths². The PHY homologous domain of PpDUC1 absorbs 612 nm and 702 nm for its photoconversion *in vitro* (Fig. 2b). This shift to a shorter wavelength may be because of the ocean's light environment. Despite limited knowledge on algal PHYs, Rockwell *et al.* have reported that they show wide variation in their spectral properties². Some PHY molecules derived from other prasinophyte species show a Po/Pfr photocycle similar to PpPHY, indicative of the same origin². 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 absorption²³. 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 (Fig. S24)²⁴, 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.

Evolutionary diversity in chlorophyte photoreceptors

Many green algae share genes for *PHY*⁹ and *pCRY*²⁵. Of the prasinophytes, *Tetraselmis*, *Nephroselmis*, *Micromonas*, *Dolichomactix*, and *Prasinoderma* possess genes for functional *PHY*^{1,2}. However, the genomes of *Chloropicon*, *Ostreococcus*, and *Bathycoccus* lack any *PHY* genes (Fig. 5). These findings suggest that *PHY* may have disappeared multiple times, independently of the prasinophytes, from the ancestors of *Chlamydomonas*, *Chloropicon*, *Pycnococcus*, *Ostreococcus*, and *Bathycoccus*. However, *pCRY* is highly conserved in prasinophytes¹³, and only the genomes of *Ostreococcus* and *Bathycoccus* lack *pCRY*. In evolutionary terms, *PpDUC1* may have been generated via the fusion of genes for *PHY* and *pCRY* in an ancestor of *Pycnococcus* (and *Pseudoscurfieldia*), because it is not found in the genomes and transcriptomes (MMETSP²⁶) of the other prasinophytes. Whole genome duplication and large-scale genome rearrangements in *P. provasolii* (Supplementary Text) could induce such a gene fusion. The *P. provasolii* genome lacks a gene for *PHY*, so it is probable that it was deleted during the process of the fusion because its function was replaced by *PpDUC1*.

Dynamic adaptation mechanisms to various light spectra and intensities

Sensitivity to weak light is essential for marine algae. *P. provasolii* also has a unique pigment composition (prasinoxanthin and Mg 2,4-D) and an ability to adapt to spectral quality (blue and blue-violet) and low-light intensity^{5,8}. Interestingly, it is not only specialized for weak light; its genome possesses *Pao* and *Sgr* that are involved in chlorophyll (Chl) degradation in the chlorophyll-protein complexes²⁷. The *Pao* genes are mainly conserved in freshwater algae (cyanobacteria and chlorophytes) and land plants, and are absent from the available genomes of marine chlorophytes²⁸. Under high-light radiation, algae and land plants degenerate antenna complexes (LHCs and pigments) to decrease the absorbance of excess energy and prevent photodamage²⁹. The degradation of Chl triggers degradation of LHCs³⁰. 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^{5,31}. *P. provasolii* possesses LHCA, LHCB and a number of prasinophyte-specific LHCs (Lhcp) (Fig. 4 and S22). The Lhcp makes complexes with Chl *a/b* and carotenoids³². Our transcriptome data showed that all *LHC* genes were upregulated and some *Lhcps* were strongly upregulated under orange and blue-light irradiation (Fig. 4). This result is consistent with *LHC* regulation in land plants.

Our studies show the dynamic adaptation mechanisms to light spectra and intensities in *P. provasolii*. These mechanisms explain the strategies used to receive and utilize light, and to expand the niches of *Pycnococcus* in the marine ecosystem.

Methods

Prediction of novel CRYs 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 v5.40-77.0³³. Proteins were defined as candidate CRYs if they had the following three domains: "DNA photolyase N-terminal (IPR006050)", "Rossmann-like alpha/beta/alpha sandwich fold (IPR014729)" and "Cryptochrome/DNA photolyase, FAD-binding domain (IPR005101)", following the dbCRY method³⁴. We also excluded known CRYs that matched with the NCBI nr database³⁵ by a blastp search³⁶. To find the candidate host organism, we applied a blast search against MMETSP (The Marine Microbial Eukaryote Transcriptome Sequence Project) data³⁷. The culture strain of *P. provasolii* NIES-2893 was obtained from the National Institute for Environmental Studies, Japan.

Construction of phylogenetic trees for CRYs, photolyases and PHYs

The domain composition of PpDUC1 (1690 aa) from *P. provasolii* was assigned using SMART(49), Pfam(50) and NCBI CDD(51). We used the PAS-GAF-PHY-PAS-PAS domain for PHY, the PHR domain for CRYs and photolyases to construct the phylogenetic trees. Multiple sequence alignment was performed by MAFFT v7.450 with the "--auto" parameter(52). Neighbor-joining phylogenetic trees were constructed using MEGA7 or MEGA X software(53) with 10,000 bootstrap replicates.

Methods for spectral characterization of PHY and CRY

Plasmid construction for expression of PpPHY and PpCRY

The *Escherichia coli* strains Mach1 T1^R (Thermo Fisher Scientific) and DH5 α (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 μ g mL⁻¹ kanamycin or 100 μ g mL⁻¹ ampicillin. The *PpPHY* region (1–1986 bp in *PpDUCI*) 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 OneTM PCR Master Mix (Toyobo Life Science) with a synthetic gene that was codon optimized for expression in *E. coli* (Genscript) and an appropriate nucleotide primer set (forward primer 5'-CGCGGCAGCCCATATGCCGGTTCCGGTACCAGCC-3' and reverse primer 5'-CTCGAATTCGGATCCTTACTGCGCCACCAGCGAGCT-3'). The pET28a vector was amplified by PCR using DNA polymerase with template DNA and an appropriate nucleotide primer set (forward primer 5'-GGATCCGAATTCGAGCTC-3' and reverse primer 5'-CATATGGCTGCCGCGCGG-3').

The *PpCRY* region (3115–5073 bp in *PpDUCI*) 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 (forward primer 5'-ATGCGAATTCCCACCAGACTCCGATGTGAGCGGCG-3' and reverse primer 5'-GATCTCTAGACTACTGCGATCGATGGCGCTTCGCG-3').

Restriction sites on each primer are indicated by underlining. 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 α (BV), phycocyanobilin (PCB) and phytochromobilin (P Φ B) synthetic systems, respectively^{10,38}. Bacterial cells were grown in LB medium containing antibiotics (20 μ g mL⁻¹ kanamycin and 20 μ g mL⁻¹ chloramphenicol) at 37°C. For protein expression, after the cells reached an optical density of 0.4–0.8 at 600 nm, isopropyl β -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 5,000 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 (Avestin). Homogenates were centrifuged at 165,000 g for 30 min and then the supernatants were filtered through a 0.8 μ m cellulose acetate membrane before loading on to a nickel-affinity HisTrap HP column (GE Healthcare) using the ÄKTA pure 25 (GE Healthcare) system. His-tagged PpPHY incorporated with each chromophore was purified as described in previous studies^{39,40}. 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). Bacterial cells were grown in 8 L LB medium containing 500 $\mu\text{g mL}^{-1}$ L-arabinose and antibiotics (5 ng mL^{-1} tetracycline, 100 $\mu\text{g mL}^{-1}$ ampicillin and 20 $\mu\text{g mL}^{-1}$ 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 5,000 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 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.

SDS-PAGE 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. 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.

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 five-fold 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^{41–43}.

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 hour with shaking at 200 rpm. The precipitate was removed by centrifugation at 20,000 g for 5 min, and then 50 μ 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) with a solvent (MeOH / 20 mM KH_2PO_4 = 20 / 80) and separated using a reverse-phase HPLC column (InertSustainSwift C18, 4.6 i.d. $\times$ 250 mm, 5 μ 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.

Methods for *in vivo* functional analysis of PpDUC1 using plants

Generating transgenic *Arabidopsis* for complementation assay

PpDUC1 cDNA with an HA sequence and a stop codon was amplified from *P. provasolii* genomic DNA using the *DUC1-F* primer (5'-CACCATGCCAGTGCCAGTGCCTG-3') and *DUC1-HA-R* primer (5'-CTAAGCATAATCTGGTACATCATATGG ATACTGCGATCGATGGCGCTTCGC-3'). It was firstly cloned into pENTR/D-TOPO (Invitrogen) and subsequently into the binary vector pMDC32⁴⁴ containing a *CaMV 35S* promoter using the Gateway strategy (Thermo Fisher Scientific). The resulting *CaMV 35S-PpDUC1* constructs were transformed into *phyB* (Salk_022635) and *cry2-1*¹⁸ mutants via the floral dipping method⁴⁵. The transgenic plants were screened on 1/2 MS medium containing 25 mg/L hygromycin.

Measurement of hypocotyl length of transgenic *Arabidopsis*

For measurement of hypocotyls, seeds were sown on GM medium, cold treated for three days in the dark, and then exposed to white light for 6 hours to enhance germination. Following the white light treatment, plants were grown at 22°C under continuous blue light (peak 447 nm, half

bandwidth of 20 nm), orange light (peak 596 nm, half bandwidth of 20 nm), red light (peak 650 nm, half bandwidth of 25 nm) and in the dark. Seven-day-old seedlings were photographed, and the hypocotyl lengths measured using ImageJ software (<http://rsb.info.nih.gov/ij/>).

Western blot analysis of PpDUC1-GFP transiently expressed in plants

Total protein was extracted from 80 mg of *A. thaliana* whole plants and *N. benthamiana* leaf discs using the following buffer: 50 mM Tris-HCl (pH8.0), 150 mM NaCl, 0.1% 2-mercaptoethanol, 5% glycerol, 0.5% Triton X-100, and cOmpleteTM protease inhibitor mini cocktail (Sigma). The slurry was centrifuged twice to remove precipitates, and 10 µl of the SDS-denatured supernatant were loaded onto a 7.5% SDS-polyacrylamide gel for electrophoresis. After electrophoresis, proteins were electroblotted to a polyvinylidene difluoride (PVDF) membrane (Millipore) in the following buffer: 25 mM Tris, 192 mM glycine and 20% methanol. Subsequently, the membrane was incubated for 1.5 h in skimmed milk, rinsed two times with 1×TBST, and immunoblotted using peroxidase anti-HA (Roshe 1:1000 dilution) for 1.5 h to detect PpDUC1-HA or rabbit anti-GFP antibody (Torrey Pines Biolabs Inc. 1:2000 dilution) for 20 h to detect PpDUC1-GFP. After incubation, the membrane was rinsed three times and incubated for 0.5 h with AmershamTM protein A horseradish peroxidase linked antibody (GE Healthcare Corp. 1:2000 dilution) to detect PpDUC1-GFP. The membrane was then rinsed three times with 1×TBST as before and processed with a luminescent image analyzer (Chem Doc XRS plus, BIO-RAD) after incubation with enhanced chemiluminescence reagents (AmershamTM ECL SelectTM Western Blotting Detection Reagents) (GE Healthcare Corp.).

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* and *PpCRY* (without their stop codons) were amplified from *P. provasolii* genomic DNA by PCR with PpDUC1-F 5'-ggggacaagtttgataaaaaagcaggcttcATGCCAGTGCCAGTGCCTGCTGC -3' and PpDUC1-R 5'-ggggaccactttgtacaagaaagctgggtcCTGCGATCGATGGCGCTTCGC -3' for *PpDUC1*, PpPHY F 5'-ggggacaagtttgataaaaaagcaggcttcATGCCAGTGCCAGTGCCTGCTGC-3' and PpPHY-R 5'-ggggaccactttgtacaagaaagctgggtcAGCGTATGGCATGCACTGAAC-3' for *PpPHY*, PpCRY-F 5'-ggggacaagtttgataaaaaagcaggcttcATGCCACCAGACTCCGATGTGAGCGGCG-3' and PpCRY-R 5'-ggggaccactttgtacaagaaagctgggtcCTGCGATCGATGGCGCTTCGC-3' for *PpCRY* and cloned into the pDONR207 ENTRY vector by BP recombination, according to the manufacturer's instructions (Invitrogen Corp.). In the primer sequences, uppercase letters indicate sequences from genomic DNA, and lowercase letters indicate adaptor sequences for the Gateway system. These open reading frames were transformed into the pEarleyGate 103 destination vector⁴⁶ to fuse in frame with GFP and 6×his by LR reaction.

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* according to Lewis *et al.*, 2008⁴⁷. A FluoView FV1000 (Olympus Corp.) microscope with x20 and x40 objectives was used to detect GFP fluorescence. The excitation wavelength was 488 nm, and a band-pass filter of 510 to 525 nm was used for emission.

Methods for genome and transcriptome analysis

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 $\mu\text{mol photons/m}^2/\text{s}$) with a 14 h:10 h light:dark cycle. Cells were collected by gentle centrifugation. DNA was extracted as described in Suzuki *et al.*⁴⁸. 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 3.3.2⁴⁹. The nanopore reads were corrected using CONSENT v1.1.2⁵⁰. The paired-end reads were corrected using Trimmomatic version 0.36⁵¹.

The assembly was re-scaffolded by the corrected nanopore reads using Fast-SG⁵² with the k-mer size = 70. Gaps were filled using LR_Gapcloser v1.1⁵³ with the corrected nanopore reads.

Finally, we mapped the corrected paired-reads using BWA version 0.7.17⁵⁴, and polished the assembly using Pilon 1.22⁵⁵. 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 Trimmomatic⁵¹. Functional annotation was performed using the eggNOG web server⁵⁶. 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 2.17⁵⁷. We defined CDSs with RPK (read per kilobase) > 10 as “expressed genes”.

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 $\mu\text{mol m}^{-2} \text{s}^{-1}$ under a LD12:12 light/dark cycle. To monitor gene expression during irradiation of monochromatic light, cells were plated and 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 $\mu\text{mol m}^{-2} \text{s}^{-1}$. The following settings were used: for blue light, the peak was at 447 nm; for orange light, at 590 nm, and for far-red light, at 700 nm. After light treatment for 2 hours, 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, INK, 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, K. K, 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 v2.5.0c⁵⁸ after trimming the low quality reads (Phred quality of ≤ 20) with the FASTX-Toolkit-0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Read counts were normalized with DESeq

in an R package⁵⁹. Differentially expressed genes (DEGs) were defined with a q-value < 0.05 and fold-change (FC) < 2/3 or FC > 1.5.

Data and materials availability

The genome and transcriptome data were deposited in the DDBJ/EMBL/GenBank under the accession number of XXX. The genome browser and transcriptome data of *P. provasolii* are available at: <http://matsui-lab.riken.jp/JBrowse/index.html?data=data%2Fpyncococcus>.

References

1. Li, F. W. *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.* Eukaryotic algal phytochromes span the visible spectrum. *Proc. Natl. Acad. Sci. U. S. A.* **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., 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).
5. Guillard, R. R. L., Keller, M. D., O’Kelly, C. J. & Floyd, G. L. PYCNOCOCCUS PROVASOLII GEN. ET SP. NOV., A COCCOID PRASINOXANTHIN-CONTAINING

PHYTOPLANKTER FROM THE WESTERN NORTH ATLANTIC AND GULF OF
MEXICO1. *Journal of Phycology* vol. 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. & Vaultot, D. Diversity and ecology of green
microalgae in marine systems: an overview based on 18S rRNA gene sequences.
Perspectives in Phycology vol. 3 141–154 (2016).

8. Iriarte, A. & Purdie, D. A. Photosynthesis and growth response of the oceanic picoplankter
Pycnococcus provasolii Guillard (clone Ω 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. *New Phytol.* **225**, 2283–2300 (2020).

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

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

12. Berndt, A. *et al.* A novel photoreaction mechanism for the circadian blue light
photoreceptor *Drosophila* cryptochrome. *J. Biol. Chem.* **282**, 13011–13021 (2007).

13. 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).

14. Lin, C. *et al.* Association of flavin adenine dinucleotide with the Arabidopsis blue light receptor CRY1. *Science* **269**, 968–970 (1995).
15. Reed, J. W., Nagatani, A., Elich, T. D., Fagan, M. & Chory, J. Phytochrome A and Phytochrome B Have Overlapping but Distinct Functions in Arabidopsis Development. *Plant Physiol.* **104**, 1139–1149 (1994).
16. Lin, C. *et al.* Enhancement of blue-light sensitivity of Arabidopsis seedlings by a blue light receptor cryptochrome 2. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 2686–2690 (1998).
17. 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).
18. Guo, H., Yang, H., Mockler, T. C. & Lin, C. Regulation of flowering time by Arabidopsis photoreceptors. *Science* **279**, 1360–1363 (1998).
19. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
20. Marcotte, E. M. *et al.* Detecting protein function and protein-protein interactions from genome sequences. *Science* **285**, 751–753 (1999).
21. Más, P., Devlin, P. F., Panda, S. & Kay, S. A. Functional interaction of phytochrome B and cryptochrome 2. *Nature* **408**, 207–211 (2000).
22. Kawai, H. *et al.* Responses of ferns to red light are mediated by an unconventional photoreceptor. *Nature* **421**, 287–290 (2003).
23. Fushimi, K. & Narikawa, R. Cyanobacteriochromes: photoreceptors covering the entire UV-to-visible spectrum. *Curr. Opin. Struct. Biol.* **57**, 39–46 (2019).
24. Essen, L. O., Mailliet, J. & Hughes, J. The structure of a complete phytochrome sensory

- module in the Pr ground state. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14709–14714 (2008).
25. 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).
26. Johnson, L. K., Alexander, H. & Brown, C. T. Re-assembly, quality evaluation, and annotation of 678 microbial eukaryotic reference transcriptomes. *Gigascience* **8**, (2019).
27. Pružinská, A., Tanner, G., Anders, I., Roca, M. & Hörtensteiner, S. Chlorophyll breakdown: Pheophorbide a oxygenase is a Rieske-type iron–sulfur protein, encoded by the accelerated cell death 1 gene. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 15259–15264 (2003).
28. Hauenstein, M., Christ, B., Das, A., Aubry, S. & Hörtensteiner, S. A Role for TIC55 as a Hydroxylase of Phyllobilins, the Products of Chlorophyll Breakdown during Plant Senescence. *Plant Cell* **28**, 2510–2527 (2016).
29. 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).
30. Wu, S. *et al.* NON-YELLOWING2 (NYE2), a Close Paralog of NYE1, Plays a Positive Role in Chlorophyll Degradation in Arabidopsis. *Mol. Plant* **9**, 624–627 (2016).
31. Kunugi, M. *et al.* Evolution of Green Plants Accompanied Changes in Light-Harvesting Systems. *Plant Cell Physiol.* **57**, 1231–1243 (2016).
32. Swingle, W. D. *et al.* Characterization of photosystem I antenna proteins in the prasinophyte *Ostreococcus tauri*. *Biochim. Biophys. Acta* **1797**, 1458–1464 (2010).
33. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

34. Kim, Y. M. *et al.* dbCRY: a Web-based comparative and evolutionary genomics platform for blue-light receptors. *Database* **2014**, bau037 (2014).
35. O’Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–45 (2016).
36. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
37. 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).
38. 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).
39. Fushimi, K. *et al.* Rational conversion of chromophore selectivity of cyanobacteriochromes to accept mammalian intrinsic biliverdin. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 8301–8309 (2019).
40. Hasegawa, M. *et al.* Molecular characterization of D. *J. Biol. Chem.* **293**, 1713–1727 (2018).
41. 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).
42. Narikawa, R. *et al.* A biliverdin-binding cyanobacteriochrome from the chlorophyll d-bearing cyanobacterium *Acaryochloris marina*. *Sci. Rep.* **5**, 7950 (2015).
43. 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).
44. Curtis, M. D. & Grossniklaus, U. A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* **133**, 462–469 (2003).
45. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal* vol. 16 735–743 (1998).
46. Earley, K. W. *et al.* Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616–629 (2006).
47. 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).
48. 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).
49. 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 mega-reads algorithm. *Genome Research* vol. 27 787–792 (2017).
50. Morisse, P., Marchet, C., Limasset, A., Lecroq, T. & Lefebvre, A. CONSENT: Scalable long read self-correction and assembly polishing with multiple sequence alignment. doi:10.1101/546630.
51. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
52. Di Genova, A., Di Genova, A., Ruz, G. A., Sagot, M.-F. & Maass, A. Fast-SG: an alignment-free algorithm for hybrid assembly. *GigaScience* vol. 7 (2018).

53. Xu, G.-C. *et al.* LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *GigaScience* vol. 8 (2019).
54. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
55. Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* vol. 9 e112963 (2014).
56. Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
57. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
58. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
59. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).

Acknowledgements

We thank Dr. S. Ota (NIES, Japan) for the FCM analyses. This research was partially supported by the National BioResource Project (NBRP) from Japan Agency for Medical Research and Development (AMED).

Author contributions

Y.M. and A.S. identified the *DUC1* gene and analyzed its expression. S.Su., H.Y., M.K. determined the *P. pravasolii* genome. K.F., R.N. confirmed the absorbance spectra of PpPHY and PpCRY. S.Shi., M.H., T.K. contributed the RNA analysis and complementation assay in *Arabidopsis*. H.H. and S. Shi 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.Su., K.F., S.Shi., Y.K., R.N., H.Y., M.K., M.M. wrote the manuscript.

Competing interests

The authors declare no competing interests.

ORCID for corresponding author

Minami Matsui

<https://orcid.org/0000-0001-5162-2668>

Figure legends

Figure 1. Metagenome sampling locations and chimeric photoreceptor, *PpDUC1*. (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 FhlA, PHY: phytochrome, PHR: photolyase-homologous region, and FAD: flavin adenine dinucleotide. (B) Metagenome sampling locations (red dots). (C) Light

microscopic image of coccoid cells of *P. provasolii* NIES-2983. Scale bar = 5 μ m. **(D)** Phylogeny of *P. provasolii* and comparison of gene content in prasinophytes. Maximum likelihood (ML) tree of 126 orthologous proteins. The dataset was composed of 46,771 amino acids from 16 species that have genomes available, including *P. provasolii*. Bootstrap percentages (BPs) are shown on the nodes. Bold lines show BP = 100.

Figure 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_{ox}, Pb form) and the photoproduct state (FAD⁻, 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 Table S1. Assignment of chromophores incorporated in PpPHY and PpCRY was performed by comparison with each standard, which are reported in Figure S5.

Figure 3. Functional complementation of *Arabidopsis cry2* and subcellular localization of PpDUC1 in tobacco leaves. **(A)** Phenotypes of seven-day-old seedlings of wild type (WT, col-4), *cry2* (*cry2-1*), and transgenic lines with *PpDUC1-HA* overexpressed in the *cry2* background, grown in continuous blue light (5.2 μ mol m⁻² s⁻¹). Scale bar = 1 mm. **(B)** Mean hypocotyl

lengths of 7-day-old seedlings $\pm$ s.d. (** $P < 0.01$, $n = 10$ to 19). (C) Mean hypocotyl lengths of 7-day-old seedlings grown under dark conditions $\pm$ s.d. ($n = 14$ to 21). (D) Schematic diagrams of domains in the PpPHY, PpCRY and PpDUC1 constructs used in the *N. benthamiana* leaf injection assay. (E) 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. (F) Leaf injection assay in *N. benthamiana* pavement cells transiently expressing PpPHY-GFP, PpCRY-GFP and PpDUC1-GFP. DIC, differential interference contrast images (left side); GFP, fluorescence images (middle); and Merge, merged images (right side). Scale bar = $50 \mu\text{m}$.

Figure 4. 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. The threshold is $q\text{-value} < 0.05$ and fold-change (FC) $< 2/3$ or $\text{FC} > 1.5$. (B) GO enrichment analysis for 388 co-regulated DEGs under blue and orange light. (C) Expression values for the light-harvesting chlorophyll protein complex I (LHCI) and prasinophyte-specific LHC (LHCP).

Figure 5. Evolutionary scenario of PHY and pCRY in chlorophytes. The phylogeny of representative chlorophytes and the domain structures of PHY and pCRY are shown. Tree topology is based on previous studies. The NCBI-nr database was searched for PHYs and pCRYs. pCRYs of *Dolichomastix*, *Nephroselmis*, and *Prasinoderma* were found in MMETSP assemblies (MMETSP0033, 0034, and 0806, respectively). In *Micromonas*, *M. pusilla* possesses *PHY*, but the gene is absent in *M. commoda*.

Figure 6. Phylogenetic trees of prasinophyte-specific light-harvesting chlorophyll proteins. and pheophorbide a oxygenase. (A) Unlabeled ML tree of prasinophyte-specific light-harvesting chlorophyll proteins (Lhcp). The dataset was composed of 183 amino acids of 56 proteins, which were encoded in the prasinophyte genomes that were available in the PhycoCosm database (<https://phycocosm.jgi.doe.gov/phycocosm/home>). (B) Unlabeled ML tree of pheophorbide a oxygenase (Pao). The dataset was composed of 211 amino acids of 34 proteins, which were found using an NCBI-nr database search. Bootstrap support values are shown on each node. BP < 50 are not shown. Green color indicates *P. provasolii* proteins.

777 **Table 1.** Characteristics of genomes of 16 Viridiplantae species.

Species	Strain	No. Chr/ Assemblies	Genome/ Assembly size (Mb)	Number of genes
Prasinophytes				
<i>Pycnococcus provasolii</i>	NIES-2893	43	22.7	11,337
<i>Micromonas commoda</i>	RCC299	17	21	10,163
<i>Micromonas pusilla</i>	CCMP1545	21	22	10,707
<i>Ostreococcus lucimarinus</i>	CCE9901	21	13.2	7,839
<i>Ostreococcus tauri</i>	RCC4221	20	12.5	7,942
<i>Bathycoccus prasinus</i>	RCC1105	19	15	7,847
<i>Chloropicon primus</i>	CCMP1205	20	17.4	8,693
UTC clade				
<i>Chlamydomonas reinhardtii</i>	CC-503	17	111.1	18,115
<i>Volvox carteri</i>	Eve	19	131.2	15,194
<i>Chlorella variabilis</i>	NC64A	12	46.2	9,833
<i>Coccomyxa subelipsoidea</i>	C-169	20	48.8	9,947
<i>Ulva mutabilis</i>		n/a	98.5	12,924
Streptophyta				
<i>Mesostigma viride</i>	NIES-296	5	441.8	24,431
<i>Chlorokybus atmophyticus</i>	CCAC0220	n/a	85	9,300
<i>Klebsormidium nitens</i>	NIES-2285	n/a	103.9	16,244
<i>Arabidopsis thaliana</i>		5	135	27,416

778 n/a: not applicable

779

789 Figure 4

790
791

792 Figure 5

793
794

Figures

Figure 1

Metagenome sampling locations and chimeric photoreceptor, PpDUC1. (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 FhlA, PHY: phytochrome, PHR: photolyase-homologous region, and FAD: flavin adenine dinucleotide. (B) Metagenome sampling

locations (red dots). (C) Light microscopic image of coccoid cells of *P. provasolii* 724 NIES-2983. Scale bar = 5 μm . (D) Phylogeny of *P. provasolii* and comparison of gene content in prasinophytes. Maximum likelihood (ML) tree of 126 orthologous proteins. The dataset was composed of 46,771 amino acids from 16 species that have genomes available, including *P. provasolii*. Bootstrap percentages (BPs) are shown on the nodes. Bold lines show BP = 100.

Figure 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 (FADOX735 , Pb form) and the photoproduct state (FAD•–, 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 Table S1. Assignment of chromophores incorporated in PpPHY and PpCRY was performed by comparison with each standard, which are reported in Figure S5.

Figure 3

Functional complementation of *Arabidopsis cry2* and subcellular localization of PpDUC1 in tobacco leaves. (A) Phenotypes of seven-day-old seedlings of wild type (WT, col-4), *cry2* (*cry2-1*), and transgenic

lines with PpDUC1-HA overexpressed in the cry2 background, grown in continuous blue light (5.2 $\mu\text{mol m}^{-2} \text{s}^{-1}$). Scale bar = 1 mm. (B) Mean hypocotyl lengths of 7-day-old seedlings $\pm$ s.d. (** $P < 0.01$, $n = 10$ to 19). (C) Mean hypocotyl lengths of 7-day-old seedlings grown under dark conditions $\pm$ s.d. ($n = 14$ to 21). (D) Schematic diagrams of domains in the PpPHY, PpCRY and PpDUC1 constructs used in the *N. benthamiana* leaf injection assay. (E) 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. (F) Leaf injection assay in *N. benthamiana* pavement cells transiently expressing PpPHY-GFP, PpCRY-GFP and PpDUC1-GFP. DIC, differential interference contrast images (left side); GFP, fluorescence images (middle); and Merge, merged images (right side). Scale bar = 50 μm .

Figure 4

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. The threshold is q-value < 0.05 and fold-change (FC) < 2/3 or FC > 1.5. (B) GO enrichment analysis for 388 co-regulated DEGs under blue and orange light. (C) Expression values for the light-harvesting chlorophyll protein complex I (LHCI) and prasinophyte-specific LHC (LHCP).

Figure 5

Evolutionary scenario of PHY and pCRY in chlorophytes. The phylogeny of representative chlorophytes and the domain structures of PHY and pCRY are shown. Tree topology is based on previous studies. The NCBI-nr database was searched for PHYs and pCRYs. pCRYs of *Dolichomastix*, *Nephroselmis*, and *Prasinoderma* were found in MMETSP assemblies (MMETSP0033, 0034, and 0806, respectively). In *Micromonas*, *M. pusilla* possesses PHY, but the gene is absent in *M. commoda*.

Figure 6

Phylogenetic trees of prasinophyte-specific light-harvesting c chlorophyll proteins. and pheophorbide a oxygenase. (A) Unrooted ML tree of prasinophyte-specific light harvesting chlorophyll proteins (Lhcp). The dataset was composed of 183 amino acids of 56 proteins, which were encoded in the prasinophyte genomes that were available in the PhycoCosm database (<https://phycocosm.jgi.doe.gov/phycocosm/home>). (B) Unrooted ML tree of pheophorbide a oxygenase (Pao). The dataset was composed of 211 amino acids of 34 proteins, which were found using an NCBI-nr

database search. Bootstrap support values are shown on each node. BP < 50 are not shown. Green color indicates *P. provasolii* proteins.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformation.pdf](#)
- [SupplementaryData.xlsx](#)