Title

Chloroplast acquisition without the gene transfer in kleptoplastic sea slugs

Authors

Taro Maeda*[1], Shunichi Takahashi[1,2] , Takao Yoshida[3], Shigeru Shimamura[3], Yoshihiro Takaki[3], Yukiko Nagai[3], Atsushi Toyoda[5], Yutaka Suzuki[6], Asuka Arimoto[7], Hisaki Ishii[8], Noriyuki Satoh[9], Tomoaki Nishiyama[10], Mitsuyasu Hasebe[1,2], Tadashi Maruyama[11], Jun Minagawa[1], Junichi Obokata[8], Shuji Shigenobu*[1,2]

Affiliation

1 National Institute for Basic Biology, Okazaki, Aichi, Japan

2 SOKENDAI, the Graduate University for Advanced Studies, Mishima, Shizuoka, Japan

3 Japan Agency for Marine-Earth Science and Technology, Yokosuka, Kanagawa, Japan

4 Tokai University, Hiratsuka, Kanagawa, Japan

5 National Institute of Genetics, Mishima, Shizuoka, Japan

6 The University of Tokyo, Bunkyo-ku, Tokyo, Japan

7 Marine Biological Laboratory, Hiroshima University, Onomichi, Hiroshima, Japan

8 Kyoto Prefectural University, Kyoto, Kyoto, Japan

9 Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan

10 Advanced Science Research Center, Kanazawa University, Kanazawa, Japan

11 Kitasato University, Sagmihara, Kanagawa, Japan

***Corresponding author**

Email addresses:

T. Maeda: maedat@nibb.ac.jp

ST: shun@nibb.ac.jp

TY: tyoshida@jamstec.go.jp

S. Shimamura: shimas@jamstec.go.jp

YT: takakiy@jamstec.go.jp

YN: nagai.y@jamstec.go.jp

AT: atoyoda@nig.ac.jp

YS: ysuzuki@k.u-tokyo.ac.jp

AA: aarimoto@hiroshima-u.ac.jp

KI: hsk.ishii@gmail.com

NS: norisky@oist.jp

TN: tomoakin@staff.kanazawa-u.ac.jp

MH: mhasebe@nibb.ac.jp

T. Maruyama: pee02660@nifty.ne.jp

JM: minagawa@nibb.ac.jp

JO: junichi.obokata@setsunan.ac.jp

S. Shigenobu: shige@nibb.ac.jp

Abstract

Some sea slug species sequester chloroplasts from algal food in their digestive gland cells and photosynthesize for up to several months. This phenomenon, kleptoplasty, poses a question of how the chloroplast retains its activity within the animal cell without the algal nucleus. As a mechanism of the retention, there have been debates on horizontal gene transfer from the algal nucleus to animal nucleus. To settle the arguments, we, here, report the genome of a kleptoplastic sea slug *Plakobranchus ocellatus* and found no evidence that photosynthetic genes are encoded on its genome. Nevertheless, our long-term incubation experiment confirmed that kleptoplast photosynthesis prolongs the life of sea slug under starvation. The data present a paradigm that a complex adaptive trait, as typified by photosynthesis, can be transferred between eukaryotic kingdoms by a unique organelle transmission without nuclear gene transfer. Our phylogenomic and transcriptomic analysis showed that animal genes for proteolysis and immune response underwent gene expansion in the lineage to the sea slug and are upregulated in the chloroplast-enriched tissue, suggesting that these molluscan genes are involved in this DNA-independent transformation.

Introduction

Since the Hershey–Chase experiment (Hershey and Chase, 1952) which proved that DNA is the material transferred to bacteria in phage infections, horizontal gene transfer (HGT) has been considered essential for cross-species transformation (Arber, 2014). Although the prion hypothesis rekindled interest in proteins as an element of phenotype propagation (Crick, 1970; Wickner et al., 2015), HGT is still assumed to be the cause of transformation. For example, in secondary plastid acquisition scenario, 1) a non-phototrophic eukaryote sequesters an unicellular archaeplastid; 2) endogenous gene transfer to the non-phototrophic eukaryote leads to shrinkage of the archaeplastidan nuclear DNA (nucDNA); and 3) the archaeplastidan nucleus disappears and its plastid becomes a secondary plastid in the host (Reyes-Prieto et al., 2007).

Chloroplast sequestration in sea slugs has attracted much attention due to the uniqueness of the phenotype acquisition from algae. Some species of sacoglossan sea slugs (Mollusca: Gastropoda: Heterobranchia) can photosynthesize using the chloroplasts of their algal food (Fig. 1) (de Vries and Archibald, 2018; Kawaguti, 1965; Pierce and Curtis, 2012; Rumpho et al., 2011; Serôdio et al., 2014). These sacoglossans ingest species-specific algae and sequester the chloroplasts into their intestinal cells. This phenomenon is called kleptoplasty (Gilyarov, 1983; Pelletreau et al., 2011). The sequestered chloroplasts (named kleptoplasts) retain their electron-microscopic structure (Fan et al., 2014; Kawaguti, 1965; Martin et al., 2015; Pelletreau et al., 2011; Trench, 1969) and their photosynthetic activity (Cartaxana et al., 2017; Christa et al., 2014a; Cruz et al., 2015; Händeler et al., 2009; Taylor, 1968; Teugels et al., 2008; Wägele and Johnsen, 2001; Yamamoto et al., 2009). The retention period of the photosynthesis differs among sacoglossan species (one day to >300 days) (Christa et al., 2015, 2014a, 2014b; Evertsen et al., 2007a; Laetz and Wägele, 2017), development stages and depends on the plastid “donor” species (Curtis et al., 2007; Laetz and Wägele, 2017).

The absence of algal nuclei in sacoglossan cells makes kleptoplasty distinct from other symbioses and plastid acquisitions (de Vries and Archibald, 2018; Rauch et al., 2015). Electron microscopic studies have indicated that the sea slug maintain photosynthetic activity without algal nuclei (Hirose, 2005; Kawaguti, 1965; Laetz and Wägele, 2018; Martin et al., 2015; Pierce and Curtis, 2012). Because most photosynthetic proteins are encoded on the algal nucleus rather than plastids, the mechanism by which photosynthetic proteins are maintained in kleptoplasty is especially intriguing, given that the photosynthetic proteins have high turnover rate

(de Vries and Archibald, 2018; Pelletreau et al., 2011). Previous PCR-based studies have suggested the HGT of algal-nucleic photosynthetic genes (e.g., *psbO*) to the nucDNA of the sea slug *Elysia chlorotic*a (Pierce et al., 1996, 2010, 2007, 2003; Rumpho et al., 2008; Schwartz et al., 2014). However, genomic study of sea slug HGT has been limited to *E. chlorotica*, and provided no reliable evidence of HGT (Bhattacharya et al., 2013). Schwartz *et al*. (2014) reported *in situ* hybridization–based evidence for HGT and argued that the genome study might overlook algae-derived genes. Although transcriptome analyses of other sea slug species failed to detect HGT (Chan et al., 2018; Wägele et al., 2011), the transcriptomic data was insufficient to ascertain genomic gene composition (de Vries et al., 2015; Rauch et al., 2015).

Here, we present genome sequences of another sacoglossan species, *Plakobranchus ocellatus*, to clarify whether HGT is the primary system underlying kleptoplasty. For over 70 years, *P. ocellatus* has been studied for its ability to retain kleptoplasts for long-term (>3 months) (Christa et al., 2013; Evertsen et al., 2007a; Greve et al., 2017; Kawaguti, 1941; Trench et al., 1970; Wade and Sherwood, 2017; Wägele et al., 2011). However, recent phylogenetic analysis showed *P. ocellatus* to be a species complex (set of closely related species) (Fig. 1) and, therefore, useful to revisit previous studies about *P. ocellatus* (Christa et al., 2014c; Krug et al., 2013; Maeda et al., 2012; Meyers-Muñoz et al., 2016; Yamamoto et al., 2013). Hence, we first confirmed the photosynthetic activity and adaptive relevance of kleptoplasty to *P. ocellatus* type black (a species confirmed by Krug *et al*. (2013) via molecular phylogenetics, hereafter "PoB"). We then constructed genome sequences of PoB and of a related species, *Elysia marginata*. Our comparative genomic and transcriptomic analyses of these species demonstrate the complete lack of photosynthetic genes in these sea slug genomes and provide evidence for an alternative hypothetical mechanism of kleptoplasty.

Results

Kleptoplast photosynthesis prolongs the life of *P. ocellatus* type black

To explore the photosynthetic activity of PoB, we measured three photosynthetic indexes: the photochemical efficiency of kleptoplast photosystem II (PSII), the oxygen production rate after starvation for 1-3 months, and the effect of illumination on PoB longevity. The value of Fv/Fm, which reflects the maximum quantum yield of PSII, was 0.68–0.69 in the "d38" PoB group (starved for 38 days), and was 0.57–0.64 in the "d109" group (starved for 109–110 days). These values were only slightly lower than those of healthy *Halimeda borneensis*, a kleptoplast donor of PoB, which showed Fv/Fm values of 0.73–0.76 (Fig. 2a, Supplementary Table 1), indicating that PoB kleptoplasts retain a similar photochemical efficiency of PSII to that of the food algae for over three months. On the measurement of oxygen concentrations in seawater, starved PoB ("d38" and "d109") displayed gross photosynthetic oxygen production (Fig. 2c). Without sea slugs, there was no light-dependent increase in oxygen concentration: i.e., no detectable microalgal photosynthesis in the seawater (Supplementary Fig. 1). The results demonstrate that PoB kleptoplasts retain photosynthetic activity for over three months. consistent with previous *P.* cf. *ocellatus* studies (Christa et al., 2014c; Evertsen et al., 2007b). We then measured the longevity of starved PoB specimens under different light conditions. Mean longevity was 156 days under continuous darkness and 195 days under a 12 h:12 h light-dark cycle ($p = 0.022$) (Fig. 2d, Supplementary Table 2), indicating that longevity was significantly extended when the animals were exposed to light. Our observation is consistent with the observation of Yamamoto et al. (2013) that the survival rate of PoB after 21 days under starvation is light-dependent. Although a study using *P.* cf. *ocellatus* reported that photosynthesis had no positive effect on survival rate (Christa et al., 2014c), our results indicate that this finding is not applicable to PoB. Taken together, the data for the three photosynthetic indexes indicate that kleptoplast photosynthesis increases resistance to starvation in PoB.

Photosynthetic genes in kleptoplasts

To reveal the genetic autonomy of kleptoplasts, we sequenced whole kleptoplast DNAs (kpDNA) from PoB and compared the sequences with algal plastid and nuclear genes. Illumina sequencing provided two types of circular kpDNA and one whole mitochondrial DNA (mtDNA) (Fig. 3, Supplementary Fig. 2-7). The mtDNA

sequence was almost identical to the previously sequenced *P.* cf. *ocellatus* mtDNA (Greve et al., 2017) (Supplementary Fig. 4). The sequenced kpDNAs corresponded with those of the predominant kleptoplast donors of PoB (Maeda et al., 2012): i.e., *Rhipidosiphon lewmanomontiae* (AP014542, hereafter "kRhip"), and *Poropsis* spp. (AP014543, hereafter "kPoro") (Fig. 3b and Supplementary Fig. 5).

To determine if the kpDNA gene repertoires were similar to those of green algal chloroplast DNAs (cpDNAs), we sequenced *H. borneensis* cpDNA and obtained 17 whole cpDNA sequences from public database (Supplementary Fig. 6, and Supplementary Tables 3 and 4). The PoB kpDNAs contained all of the 59 conserved chloroplastic genes in Bryopsidales algae (*e.g.*, *psbA, rpoA*), although they lacked 4–5 of the dispensable genes (i.e., *petL*, *psb30*, *rpl32*, *rpl12*, and *ccs1*) (Fig. 3c and Supplementary Fig. 7).

To test whether the kpDNAs contained no additional photosynthetic genes, we then used a dataset of 614 photosynthetic genes (hereafter, the "A614" dataset), which were selected from our algal transcriptome data and public algal genome data (Supplementary Table 5 and 6). A tblastn homology search using the A614 obtained no reliable hits (E-value < 0.0001) against our kpDNA sequences, except for the gene *chlD*, which resembled kpDNA-encoded *chlL* (Fig. 3de, Supplementary Figs. 8–10). A positive control search against an algal nucDNA (*C. lentillifera*) database (Arimoto et al., 2019) found reliable matches for 93% (575/614) of the queries (Fig. 3d), suggesting that the method has high sensitivity. Thus, the comparison with algal plastid and nucleic genes clarified that the kpDNAs lack multiple photosynthetic genes (e.g., *psbO*).

Absence of horizontally transferred algal genes in the *P. ocellatus* type black nucleic genome

To determine if the PoB nucleic genome contains algae-derived genes (i.e., evidence of HGT), we sequenced the nuclear genome of PoB and searched for algae-like sequences in the gene models, genomic sequences, and pre-assembled short reads. Our genome assembly contained 927.9 Mbp (99.1% of the estimated genome size, 8,647 scaffolds, N50 = 1.45 Mbp, 77,230 gene models) (INSD; PRJDB3267, Supplementary Figs. 11–14, Supplementary Tables 7–10). BUSCO (Benchmarking Universal Single-Copy Orthologs) analysis using the eukaryota_odb9 dataset showed high coverage (93%) of the eukaryote conserved gene set (Supplementary Table 7), indicating that our gene modeling was sufficiently complete to enable HGT searches.

Searches of the PoB gene models found no evidence of HGT from algae. Although simple homology searches (blastp) against the RefSeq database found 127 PoB gene models with top-hits against Cyanobacteria

or eukaryotic algae, the prediction of taxonomical origin using multiple blast hit results (LCA analysis with MEGAN software) (Huson et al., 2007) denied the algal-origin of the genes (Supplementary Table 11). A blastp search using the A614 dataset which contains sequences of the potential gene donor (e.g. transcriptomic data of *H. borneensis*) also determined no positive evidence of HGT (Supplementary Table 11).

In gene function assignment with gene ontology (GO) terms, no PoB gene model was annotated as a "Photosynthesis (GO:0015979)"-related gene, although the same method found 72–253 photosynthesis-related genes in the five algal species used as references (Fig. 4a). Our GO-analysis found six PoB genes assigned with the child terms of "Plastid" (GO:0009536). However, ortholog search with animal and algal genes did not support the algal origin of these genes (Supplementary Fig. 15 and 16; Supplementary Table 10). We consider that these pseudo-positives in the similarity search and GO assignment are caused by sequence conservation of the genes beyond the kingdom.

To confirm that our gene modeling did not overlook a photosynthetic gene, we directly searched the A614 dataset against the PoB and *C. lentillifera* (algal, positive control) genome sequences with tblastn and Exonerate software (Slater and Birney, 2005). Against the *C. lentillifera* genome, we found 455 (tblastn) and 450 (Exonerate) hits; however, using the same parameters we only detected 1 (tblastn) and 2 (Exonerate) hits against the PoB genome (Fig. 4b, Supplementary Fig. 17). The three loci detected in PoB contain the genes encoding serine/threonine-protein kinase LATS1 (p258757c71.5), Deoxyribodipyrimidine photolyase (p855c67.9), and phosphoglycerate kinase (p105c62.89). Phylogenetic analysis with homologous genes showed the monophyletic relationships of these three genes with molluscan homologs (Supplementary Figs. 18–20). This indicates that the three PoB loci contain molluscan genes rather than algae-derived ones. We thus conclude that our PoB genome assembly contains no algae-derived photosynthetic genes.

To examine if our genome assembly failed to construct algae-derived regions in the PoB genome, we searched for reads resembling photosynthetic genes among the pre-assembled Illumina data. From the 1,065 million pre-assembled reads, 1,698 reads showed similarity against 261 of the A614 dataset queries based on MMseq2 searches. After normalization by query length, the number of matching reads against algal queries was about 100 times lower than that against PoB single-copy genes: the normalized read count was 25 ± 105 (mean ± SD) for the A614 dataset and 2601 ± 2173 for the 905 PoB single-copy genes ($p < 0.0001$, Welch's two-sample *t*-test) (Fig. 4c). This large difference indicates that many of the algae-like reads were derived from

contaminating microalgae rather than PoB nucDNA. Although, for five algal queries (e.g., peptidyl-prolyl cis-trans isomerase gene), the number of matching reads was comparable between PoB and the A614 dataset (Fig. 4c), the presence of homologous genes on the PoB genome suggests that the reads were not derived from an algae-derived region, but rather resemble molluscan genes. For example, simple alignment showed that the *C. lentillifera*–derived g566.t1 gene, which encodes peptidyl-prolyl cis-trans isomerase, was partially similar to the PoB p310c70.15 gene (Supplementary Fig. 21), and 76% (733/ 970) of reads that hit against g566.t1 were also hitting against p310c70.15 under the same MMseq2 parameter (Supplementary Fig. 21). We hence consider that no loss of algal-derived regions occurred in our assembly process.

Changing the focus to HGT of non-photosynthetic algal genes, we calculated the indexes for prokaryote-derived HGT (h index) (Boschetti et al., 2012) and eukaryotic-algae-derived HGT (hA index, see Methods, Supplementary Table 12) for PoB and two non-kleptoplastic gastropods (negative controls). We detected three PoB gene models as potential algae-derived genes (Fig. 4d); however, two of these encoded a transposon-related protein and the other encoded an ankyrin repeat protein that has a conserved sequence with an animal ortholog. Furthermore, the non-kleptoplastic gastropods (e.g., *Aplysia californica*) had similar numbers of probable HGT genes (Fig. 4d). Taking these results together, we conclude that there is no evidence of algae-derived HGT in our PoB genomic data.

We then examined if algae-like RNA is present in PoB, because a previous study of another sea slug species *E. chlorotica* hypothesized that algae-derived RNA contributes to kleptoplasty (Bhattacharya et al., 2013). This previous study used short-read–based BLAST searches and RT-PCR analyses to detected algal mRNA (e.g., *psbO* mRNA) in multiple adult *E. chlorotica* specimens (no tissue information was provided) (Bhattacharya et al., 2013). To analyze the algal RNA distribution in PoB, we constructed 15 RNA-Seq libraries from six tissue types (digestive gland [DG], parapodia, DG-exenterated parapodia [DeP], egg, head, and pericardium; Supplementary Fig. 22) and conducted MMseq2 searches (Fig. 4c). Although almost all (594/614) of the A614 dataset queries matched no reads, 19 queries matched 1–10 reads, and the *C. lentillifera*-derived g566.t1 query matched over 10 reads (Fig. 4e). This high hit rate for g566.t1 (Peptidyl-prolyl cis-trans isomerase in Fig. 4e), however, is due to its high sequence similarity with PoB ortholog, p310c70.15 as mentioned above. A previous anatomical study of PoB showed the kleptoplast density in various tissues to be DG > parapodia > DeP = head = pericardium >>> egg (Hirose, 2005). Therefore, the amount of algae-like RNA reads did not correlate with the

kleptoplast richness among the tissues. Enrichment of algae-like RNA was only found in the egg. The PoB egg is considered to be a kleptoplast-free stage and is covered by a mucous jelly, which potentially contains environmental microorganisms (Fig. 4e and Supplementary Fig. 22c). We hence presume that these RNA fragments are not derived from kleptoplasts but from contaminating microalgae. We conclude that our searches for algae-like RNA in PoB found no credible evidence of algae-derived RNA transfer and no correlation between algal RNA and kleptoplasty.

Kleptoplasty-related *P. ocellatus* type black genes

Because we found the PoB genome to be free of algae-derived genes, we considered that a neofunctionalized molluscan gene might contribute to kleptoplasty. To find candidate kleptoplasty-related molluscan genes (KRMs), we focused on genes that were upregulated in DG (the primary kleptoplasty location) versus DeP in the RNA-Seq data described above. We found 162 DG-upregulated genes (FDR < 0.01, triplicate samples) (Fig. 5a and Supplementary Figs. 23–25). By conducting GO analysis, we identified the functions of 93 of the DG-upregulated genes and showed that they are enriched for genes involved in proteolysis (GO terms: “Proteolysis”, “Aspartic-type endopeptidase activity”, “Cysteine-type endopeptidase inhibitor activity”, and “Anatomical structure development”), carbohydrate metabolism (“Carbohydrate metabolic process”, “One-carbon metabolic process”, “Cation binding”, and “Regulation of pH”), and immunity (“Defense response”) (Supplementary Table 13). Manual annotation identified the function of 42 of the remaining DG-upregulated genes. Many of these are also related to proteolysis and immunity: three genes relate to proteolysis (i.e., genes encoding interferon-gamma-inducible lysosomal thiol reductase, replicase polyprotein 1a, and phosphatidylethanolamine-binding protein), and 21 genes contribute to natural immunity (i.e., genes encoding lectin, blood cell aggregation factor, and MAC/Perforin domain-containing protein) (Supplementary Fig. 24 and 25). Our manual annotation also found four genes encoding apolipoprotein D, which promotes resistance to oxidative stress (Charron et al., 2008), and three genes involved in nutrition metabolism (i.e., genes encoding betaine–homocysteine S-methyltransferase 1-like protein, intestinal fatty acid-binding protein, and cell surface hyaluronidase). Because the analyzed slugs were starved for a month, we consider that the DG-upregulated genes contribute to kleptoplasty rather than digestion.

We then conducted a comparative genomic analysis to find orthogroups that expand or contract in size along the metazoan lineage to PoB. Our phylogenomic analysis showed that 6 of the 193 orthogroups that underwent gene expansion in this lineage contained DG-upregulated genes, supporting the notion that these genes play a role in kleptoplasty (Fig. 5b, Supplementary Figs 26 and 27, Supplementary Table 14 and 15). The most distinctive orthogroup was OG0000132, which contained 203 cathepsin D-like genes, 45 of which were DG-upregulated, in PoB; Fisher's exact test supported the significant enrichment of the DG-upregulated genes ($p < 0.0001$, Supplementary Table 15). Other heterobranchian mollusks only had 4–5 genes in OG0000132 (Fig. 5b and Supplementary Fig. 28). These gene duplications in PoB might reduce selection pressure to maintain function via redundancy, and promote new function acquirement of the paralogs, as occurs in the well-known neofunctionalization scenario (Conrad and Antonarakis, 2007). DG-upregulated genes were also detected as significantly enriched in OG0000005 (18 genes) and OG0000446 (4 genes) (both $p < 0.0001$) (Fig. 5b), which contain lectin-like and apolipoprotein D-like genes, respectively. DG-upregulated genes were detected in OG0000002, OG0009179, and OG0000194, but were not significantly enriched in these orthogroups ($p > 0.05$). OG0000002 contains the DG-upregulated gene for type 2 cystatin, but also contains various genes with reverse transcriptase domains. This suggests that the reverse transcriptase domain clustered the various genes as one orthogroup and the gene number expansion was due to the high self-duplication activity of the retrotransposon in PoB. OG0009179 and OG0000194 contain DG-upregulated genes of unknown function. From the above results, we finally selected 67 genes as promising targets of study for PoB kleptoplasty: 45 genes for cathepsin D-like proteins, 18 genes for lectin-like proteins, and four genes for apolipoprotein D-like protein (detained gene annotation data has been deposited in 10.6084/m9.figshare.12318977).

Evolution of the candidate KRMs

For a more detailed analysis of the evolution of the kleptoplasty-related in sacoglossan lineages, we constructed a new draft genome sequence of another sacoglossan sea slug *E. marginata* (previously identified as *E. ornata* (Krug et al., 2013)). PoB and *E. marginata* belong to the same family (Plakobranchidae, Fig. 1e). Both species sequester plastids from Bryopsidales algae; however, the kleptoplast retention time is limited to a few days in *E. marginata (Yamamoto et al., 2009)*. This suggests that their common ancestor obtained a mechanism to sequester algal plastids, but *E. marginata* did not develop a system for their long-term retention. Hence, we

considered that comparing gene expansion in these species would clarify the genetic basis of plastid sequestration and long-term retention.

Using the same methods as described for PoB, we sequenced one complete circular kpDNA, one complete mtDNA, and 790.3 Mbp of nucDNA (87.8% of estimated genome size; 14,149 scaffolds; N50 = 0.23 Mbp; 70,752 genes, Supplementary Tables 7, 16, and 17) for *E. marginata*. The constructed gene models covered 89.5 % of the BUSCO eukaryota_odb9 gene set (Supplementary Table 7). No credible photosynthetic gene was detected (annotation data, 10.6084/m9.figshare.12318977).

We then phylogenetically analyzed the evolution of representative candidate PoB KRMs (i.e., cathepsin D-like, apolipoprotein D-like, and lectin-like genes). In the case of the cathepsin D-like genes, the sacoglossan (PoB + *E. marginata*) genes formed a specific subgroup in the OG0000132-based phylogenetic tree (Fig. 5c, far left), and the gene duplication in OG0000132 seemed to be accelerated along the PoB lineage (203 genes in PoB versus 5 in *E. marginata*; Fig. 5b). All sacoglossan cathepsin D-like genes belonged to a clade with several other heterobranchian homologs; this clade contained three sub-clades (α, β, γ in Fig. 5c). The basal α-clade contained three *Aplysia californica* genes, one *Biomphalaria glabrata* gene, and three sacoglossan genes. The β- and γ-clades contained sacoglossan genes only, and an *A. californica* gene was located at the basal position of the β- and γ-clades. Almost all duplicated PoB genes (201/203) belonged to the γ-clade, which also included one *E. marginata* gene (e8012c40.2). These phylogenic relationships suggest that the γ-clade has undergone dozens of gene duplication events in the PoB lineage. Interestingly, all DG-upregulated DEGs were contained in the γ-clade, and the PoB paralogs belonging to the α- and β-clades showed different expression patterns from the γ-clade paralogs; the gene p609c69.52 (α-clade) was ubiquitously expressed in the examined tissues, and p374c67.53 (β-clade) was only expressed in the egg (Fig. 5c, center right). The mammalian genes encoding cathepsin D and its analog (cathepsin E) are ubiquitously expressed on various tissues types (Benes et al., 2008). We, therefore, consider that 1) the ubiquitously expressed p609c69.52 gene in α-clade is a functional ortholog of the mammalian cathepsin D gene; 2) the p374c67.53 gene in β-clade relates sea slug embryo development; 3) the γ-clade genes have been acquired with the development of plastid sequestration. The cathepsin D-like genes formed multiple tandem repeat structures in the PoB genome, although other Heterobranchia had no tandem repeat (Fig. 5c, far right; Supplementary Figs. 28 and 29). In *E. marginata*, the gene e8012c40.2 located at the basal position of the γ-clade had no repeat structure, although two genes from the α-clade (e4981c37.5 and

e4981c37.4) and two from the β-clade (e2244c39.16 and e2244c39.20) made two tandem repeats (Supplementary Figs. 28 and 29). The tree and repeat structure suggest that the γ-clade separated from the β-clade as a single copy gene in the common ancestor of PoB and *E. marginata*, and duplicated in the PoB lineage (Fig. 5c). The revealed genomic structure indicates that the duplicates were not due to whole-genome duplication but rather a combination of several subgenomic duplication events: the dispersed duplicates are likely due to replicative transposition by a transposable element, and the tandem repeats are likely due to a local event (e.g., unequal crossover).

The gene duplication in OG0000446 seems to have happened in the PoB lineage and at the node between PoB and *E. marginata* (Fig. 5b, Supplementary Figs. 30 and 31). The sacoglossan genes were duplicated in a monophyletic clade (Clade I) only, and all DG-upregulated DEGs were contained in the Clade I. We hypothesize that duplication on the common lineage relates to plastid sequestration, and the PoB-specific duplication events contribute to long-term kleptoplasty. All DG-upregulated DEGs were contained in the Clade I. We hypothesize that duplication on the common lineage relates to plastid sequestration, and the PoB-specific duplication events contribute to long-term kleptoplasty.

In OG0000005 (lectin-like gene group), the gene counts were comparable between PoB (367) and *E. marginata* (213) (Fig. 5b); however, the phylogenetic tree suggests that this similarity is due to different gene duplication events in each species. One-by-one orthologous gene pairs were rare between PoB and *E. marginata* (only 14 were detected), and many of the other homologs formed species-specific sub-clades (Supplementary Fig. 32). Lectins are carbohydrate-binding molecules that mediate attachment of bacteria and viruses to their intended targets(Lis and Sharon, 1998; Yamasaki et al., 2008). The observed gene expansions of lectin-like genes in each of these sea slugs may widen their targets and may complicate the natural immune system to distinguish the kleptoplast from other antigens. We detected four clades having expanded homologs on the sacoglossan lineage (clades A–D in Supplementary Fig. 32). These clades contained all of the 34 DEGs between DG and DeP tissues. The DG-upregulated and -downregulated genes were placed on different clades, except for one DG-upregulated gene (p3334c67.98) in Clade A. These results indicate that the determined DEGs duplicated after the specification of PoB, and the homologs have different expression patterns depending on the clade.

Discussion

Here we demonstrate that 1) kleptoplast photosynthesis extends the lifetime of PoB under starvation, 2) the PoB genome encodes no algal-nucleus-derived genes, and 3) PoB individuals upregulate genes for carbohydrate metabolism, proteolysis and immune response in their digestive gland. Combination of the genomic and transcriptomic analysis identified candidate KRM genes (apolipoprotein D-like, cathepsin D-like, and lectin-like genes), which relates proteolysis and immune response. Together, the data present a paradigm of kleptoplasty in which PoB obtains the adaptive photosynthesis trait by DNA-independent transformation.

Our genomic sequence of PoB clarified the gene repertory of its kpDNA, mtDNA, and nucDNA. The whole kpDNA sequence indicated that PoB kleptoplasts can produce some proteins involved in photosynthesis (e.g., PsbA, a core protein in PSII) (Fig. 3) if gene expression machinery is sufficiently active, as reported in *E. chlorotica*. We then demonstrated the absence of core photosynthetic genes in PoB genome. For instance, we did not detect the genes encoding psaF of photosystem I, psbO of photosystem II, or rbcS of Rubisco in kpDNAs, mtDNA, or nucDNA (Figs. 3 and 4), despite our queries (e.g. A612 dataset, Supplementary Table 6) containing multiple algal orthologs of these genes; these gene products are essential for photosynthesis in various plants and algae, and are encoded on their nuclear genome, not plastid genome(Farah et al., 1995; Izumi et al., 2012; Pigolev et al., 2009). This means that PoB can perform adaptive photosynthesis (Fig. 2) without *de novo* synthesis of these gene products. The absence of algae-derived HGT is consistent with previous transcriptomic analyses of *P.* cf. *ocellatus* (Wägele et al., 2011) and other sacoglossan species (de Vries et al., 2015; Han et al., 2015; Wägele et al., 2011). A previous genome study of *E. chlorotica* predicted that fragmented algal DNA and/or mRNAs contribute to its kleptoplasty (Bhattacharya et al., 2013), and fluorescence *in situ* hybridization study detected algal gene signals on *E. chlorotica* chromosomes (Schwartz et al., 2014). Although no photosynthesis-related algal genes were found in a more comprehensive draft version of the *E. chlorotica* nuclear genome sequence, Cai *et al* provided no discussion about the HGT (Cai et al., 2019). From our results for PoB, we propose that algal DNA and/or RNA are not an absolute requirement for kleptoplast photosynthesis.

Our combination of genomics and transcriptomics suggests that the maintenance of algae-derived protein activity is the most probable mechanism for retaining PoB photosynthesis. Because of the limited longevity of

the photosynthetic proteins in algal cells and/or *in vitro* (Roberts et al., 2013), previous studies have discussed elongation of algal protein lifespans via protective sea slug proteins as an alternative hypothesis to HGT (de Vries and Archibald, 2018; Serôdio et al., 2014). Our study shows three types of molluscan genes as candidate protective sea slug proteins: apolipoprotein D-like, cathepsin D-like, and lectin-like genes.

Previous RNA-Seq studies of *Elysia timida* and *E. chlorotica*, found upregulation of superoxide dismutase (SOD) genes in response to photostress (Chan et al., 2018; de Vries et al., 2015), and postulated that SOD protests algal proteins in the kleptoplasts from oxidative damage. We found no significant upregulation of the SOD gene in PoB DG (Supplementary Fig 24). However, we did find the upregulation of apolipoprotein D-like genes and cathepsin D-like genes and the expansion of these genes in the PoB lineages. Both proteins are candidates for protective proteins against oxidative stress. Apolipoprotein D, a lipid antioxidant, confers resistance to oxidative stress in higher plants and animal brain (Bishop et al., 2010; Charron et al., 2008). Cathepsin D degrades intracellular proteins and contributes to the degradation of damaged mitochondria (Benes et al., 2008). In general, damaged photosynthetic proteins generate abundant ROS, which promote further protein damage. In PoB, the chain of protein inactivation may be broken through 1) inhibition of ROS accumulation by apolipoprotein D-like proteins and 2) active degradation of damaged proteins by cathepsin D-like proteases. Ortholog analysis of our new *E. chlorotica* data found no gene number expansion of these homologs in *E. chlorotica*: i.e., only three apolipoprotein D-like homologs and four cathepsin D-like homologs (Supplementary Table 18). Although the details of the retention process may differ among species and abiotic conditions, it is attractive to speculate that oxidative stress resilience is of major importance for kleptoplasty in multiple sacoglossan species.

The observed increase in expression of lectin genes in PoB DG tissue and their expansion in PoB lineages (Fig. 5b, Supplementary Figs. 24 and 32) are also consistent with the kleptoplast photosynthetic protein retention hypothesis. Almost all metazoans display natural immunity, and lectins are involved in self-recognition in immunity (Geijtenbeek and Gringhuis, 2009; Worthley et al., 2005). The diverse lectins expressed in PoB DG tissue may bind the antigens of algae-derived molecules, mediate detection of non-self-proteins and/or saccharides, and lead to the selective degradation and retention of algae-derived proteins and organelles.

Our genomic data indicate that proteomic analysis of kleptoplasts is warranted. A previous isotopic study indicated that function-unknown sea slug proteins are transported to *E. chlorotica* kleptoplasts (Pierce et al.,

1996). Although several algal photosynthetic proteins have been immunoassayed in kleptoplasts, animal nuclear-encoded proteins have not been examined (Green et al., 2000; Pierce et al., 1996). Our *in-silico* study found no typical chloroplast localization signal in PoB KRMs (Supplementary Figs. 28 and 30), however, our genomic data will help the future identification of kleptoplast-localized sea-slug proteins by peptide mass fingerprinting.

Here, we provide the first genomic evidence of photosynthesis acquisition without horizontal DNA or RNA transfer. Previous studies have demonstrated that DNA is the core material for heredity (Hershey and Chase, 1952; Watson and Crick, 1953) and have assumed that horizontal DNA transfer causes cross-species phenotype acquisition (Acuña et al., 2012; Anderson, 1970; Boto Luis, 2014; Dehal et al., 2002). Our studies, however, indicate that PoB gains adaptive photosynthetic activity without acquiring any of the many algal-nucleic-genes involved in photosynthesis. This is evidence that a phenotype (and organelle) can move beyond a species without DNA or RNA transfer from the donor. Recent studies of shipworms (wood-feeding mollusks in the family Teredinidae) showed that they utilize several symbiont-derived proteins for their food digestion (O'Connor et al., 2014). Golden sweeper fish (*Parapriacanthus ransonneti*) gain the luciferase for their bioluminescence from ostracod prey, suggesting phenotype acquisition via sequestration of a non-self-protein (kleptoprotein) (Bessho-Uehara et al., 2020). These two examples of DNA/RNA-independent transformation, however, are limited to the transfer of simple phenotypes that depend on just a few enzymes. In contrast, the well-known complexity of photosynthesis suggests that sea slug kleptoplasty depends on DNA/RNA-independent transformation of complex pathways requiring multiple enzymes (e.g., entire photosystems and the Calvin-cycle). It is attractive to speculate that other symbiont-derived organelles (e.g., mitochondria), and obligate endosymbiotic bacteria and protozoan kleptoplasts (e.g., in *Dinophysis acuminata*) *(Hackett et al., 2003)* can move beyond the species via a DNA-independent system. Although several organisms have multiple HGT-derived functional genes, it is still unclear how the organism evolutionary obtained the appropriate expression control system of the non-self gene (Sasakura et al., 2016). Our PoB data suggest that the transfer of adaptive complex phenotypes sometimes precede gene transfer from the donor species, having the potential to explain the process of cross-species development of complex phenotypes. Some organisms may evolutionary obtained the HGT-derived genes and appropriate control system of the gene expression after the transfer of

phenotype. Our finding of DNA-independent complex phenotype acquisition may open new viewpoints on cross-species evolutionary interaction.

#Materials and Methods

Sampling of sea slugs and algae

Samples were collected from southwestern Japan; specifically, PoB and *H. borneensis* were collected at shares of the island of Okinawa, and *E. marginata* was collected from Kinkowan bay. Regarding *B. hypnoides*, a cultivated thallus was initially collected from Kinkowan bay and used in our laboratory for several years. Collected samples of PoB and *E. marginata* in seawater were transported respectively to laboratories at NIBB and Kyoto Prefectural University under dark conditions within 2 days. The samples were then acclimated in an aquarium filled with artificial seawater (REI-SEA Marine II; Iwaki, Japan) at 24 °C.

Photosynthetic activity of sea slugs and algae

The photosynthetic activity of PoB was measured after 38, 109, or 110 days of incubation under a 12 h:12 h light-dark cycle without food. During the light phase, the photosynthetic photon flux density was 10 µmol photons m^{-2} s^{-1} (LI-250A Light Meter with LI-193 Underwater Spherical Quantum Sensor, LI-COR). We did not change the seawater during the incubation period except to adjust salinity using distilled water. Photosynthetic activity indexes (oxygen generation rate and PAM Fluorometry) were measured using oxygen-sensor spots (Witrox 4; Loligo Systems, Tjele, Denmark) and PAM-2500 (WALZ, Effeltrich, Germany, Supplementary Fig. 1), respectively.

The oxygen-sensor spots were affixed to the inside of a glass respirometry chamber. Before performing measurements, the system was calibrated using sodium sulfite (0% O_2 saturation) and fully O_2-saturated seawater (100% O_2 saturation). A sea slug was placed into a respirometry chamber filled with fully O_2-saturated filtered artificial seawater (7 ml). The top of the chamber was closed with a glass slide. All visible bubbles were removed from the chamber. The chamber was maintained at a constant temperature (24 ºC) using a water jacket attached to temperature-controlled water flow. The Witrox temperature probe for calibration was immersed in the water jacket.

The oxygen concentration was measured sequentially under changing light conditions. The percent O_2 saturation was monitored continuously and recorded using AutoResp software (Loligo Systems) for 10 min after the respirometry chamber acclimation period (10 min). The oxygen consumption rate by respiration was

measured under the dark condition. Next, we exposed the respirometry chambers to the red LED light (800 µmol photons m^{-2} s^{-1}) to measure the change in the oxygen concentration under the light condition due to the balance between the photosynthesis and respiration rates. The chambers were illuminated from the sides because a top-mounted LED light increased the noise measured by the oxygen meter. The light direction with respect to the sea slug was inconsistent because the slug continued to move around the chamber during the measurement period. In sea slugs, the rate of photosynthesis appeared to be unaffected by the direction of illumination because the rate of O_2 generation rate under a certain constant illumination did not change, regardless of the sea slug position in the chamber. The percent O_2 saturation was measured for 10 min. One blank (i.e., without sea slug) condition was run as a negative control to account for background biological activity in the seawater. AutoResp software was used to convert the percent saturation to an oxygen concentration ([O_2], mg O_2 l^{-1}) based on the rate of change in the percent O_2 saturation, the water temperature, and barometric pressure (fixed at 1013 Pa). We performed a regression analysis by using the "lm" function in R (version 3.5.2, tidyverse 1.2.1 package) to calculate the changing oxygen concentration rate under dark and light conditions, and obtained a gross oxygen generation rate by photosynthesis (oL + oD = oG, oL; Oxygen production rate under light, oD; Oxygen consumption rate under dark, oG; gross oxygen generation by photosynthesis).

During the PAM Fluorometry analysis, each sea slug was caged in a single well of a 12-well cell culture plate (Corning, Corning, NY, USA) after adaptation to the dark for 15 min. To ensure reproducibility, we caged the sea slug upside down (i.e., the ventral surface was brought to the upside), softly squeezed the animal with a plastic sponge, and connected the PAM light probe to the plate from underneath the well. Consequently, the samples could not move during the measurement, and the PAM light probe always measured the fluorescence of the dorsal surface. The maximal quantum yield, Fv/Fm, was determined by a saturation pulse of >8000 µmol photons m^{-2} s^{-1} and a measurement light of 0.2 µmol photons m^{-2} s^{-1}.

Effect of the light condition on *P. ocellatus* longevity

We measured the longevity of PoB using a modified medaka (Japanese rice fish) housing rack system (Iwaki, Japan). For the longevity measurement, we used different individuals from the photosynthetic activity measurement. The longevity was investigated from the samples used in the above-mentioned photosynthetic

activity measurement. Using centrally filtered systems, our water tank rack maintained consistent water conditions (e.g., temperature and mineral concentration) among the incubation chambers (sub-tanks) and enabled a focus on the effect of the light condition. After acclimating the collected sea slugs under the same conditions for 1 week in an aquarium, the organisms were incubated separately for 8 months under different light conditions (continuous dark and 12 h:12 h light-dark cycle). We evaluated the conditions of the sea slugs daily and defined death as a sea slug that remained motionless for 30 seconds after stimulation (i.e., touching with plastic bar).

Sequencing of *H. borneensis* chloroplast DNA

We used a combination of pyrosequencing and Sanger sequencing to evaluate *H. borneensis* cpDNA. Collected *H. borneensis* thalli were washed with tap water to remove the attached organisms. The cleaned thalli (27 g) were frozen in liquid nitrogen, ground with a T-10 Basic Homogenizer (IKA, Germany) to a fine powder, suspended in 15 mL of AP1 buffer from a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany), and centrifuged (500 × g, 1 min) to remove the calcareous parts. Total DNA was purified from the supernatant according to the protocol supplied with the DNeasy Plant Mini Kit. The resulting DNA yield (76.8 µg) was measured using a dsDNA HS Assay Qubit Starter Kit (Thermo, Waltham, MA, USA), and was used to prepare a single-fragment library for pyrosequencing. The Pyrosequencer GS-FLX Titanium platform (Roche, Germany) was used to generate 23.04 Mb of total singleton reads (68,032 reads, average read length: 368 bp). After filtering low-quality reads, the remaining 21,865 reads (~8 Mb) were submitted for assembly by Newbler (Roche). Of the 6,309 obtained contigs (N50 = 663 bp), the cpDNA sequences were identified by BLASTX searches (version 2.2.28) against the protein-coding sequences of cpDNA from the chlorophyte alga *B. hypnoides* (NC_013359). The gaps between the four identified contigs (55,551, 21,264, 8,076, and 3,435 bp) and ambiguous sites in the contigs were amplified by PCR using inverse primers. PCR products were sequenced by primer walking and Sanger sequencing with Takara LA Taq (Takara, Japan), a dGTP BigDye Terminator Cycle Sequencing FS Ready Reaction Kit (Thermo), and an ABI PRISM 3130xl DNA Sequencer (Thermo). Regions that could not be read by direct sequencing were amplified using specific primers, cloned with a TOPO TA cloning kit (Thermo), and sequenced with plasmid-specific primers. Complete cpDNA sequences were obtained by assembling the GS-FLX contigs and reads generated by Sanger sequencing using Sequencher ver. 4.10 (Gene Codes Corporation, MI, USA). All open reading frames >100 bp were annotated using a BLASTX search (version

2.2.31+) against the non-redundant protein sequences (nr) database in GenBank and a TBLASTX search of chloroplast genes from other algae (*B. hypnoides*, *Chlamydomonas reinhardtii*, *V. litorea*, and *H. borneensis*). Introns were detected using RNAweasel (Gautheret and Lambert, 2001), Rfam (Kalvari et al., 2018), and Mfold (Zuker, 2003). We confirmed the splicing sites via alignment with orthologous genes from other green algal cpDNAs. We used MAFFT v7.127b (Katoh and Standley, 2013) to perform the alignment. The origins of bacteria-like proteins were explored using a BLASTX search against the nr database and a phylogenetic analysis with BLAST-hit sequences. We used MAFFT (Katoh and Standley, 2013) for alignment, Trimal v1.4 (Capella-Gutiérrez et al., 2009) for trimming, and RaxML v8.2.4 (Stamatakis, 2014) for phylogenetic tree construction. The resulting gene maps were visualized using Circos version 0.69-2.

Sequencing of kpDNAs from *P. ocellatus*

Kleptoplast sequences were generated using an Illumina system. Total DNA was extracted from the digestive gland (kleptoplast-rich tissue) and parapodia (including the digestive gland, kleptoplast-less muscle, and reproductive systems) of a single PoB individual using a CTAB-based method (Murray and Thompson, 1980). Two Illumina libraries with 180- and 500-bp insertions were constructed from each DNA pool (DRR063261, DRR063262, DRR063263, DRR063264). An S220 Focused-Ultrasonicator (Covaris, MA, USA), Pippin Prep (Sage Science, MA, USA), and a TruSeq DNA Sample Prep Kit (Illumina, CA, USA) were used for DNA fragmentation, size selection, and library construction, respectively. Libraries were sequenced (101 bp from each end) on a HiSeq 2500 platform (Illumina). A total of 42,206,037 raw reads (8.53 Gb) were obtained. After filtering the low-quality and adapter sequences, the remaining 5.21 Gb of sequences were used for assembly. Paired sequences from 180-bp libraries were combined into overlapping extended contigs using FLASH ver. 1.2.9 (Magoc and Salzberg, 2011) with the default settings. An input of 14,867,401 paired-end sequences and FLASH were used to construct 13,627,554 contigs (101–192 bp). The joined fragments and filtered paired sequences from 500-bp libraries were assembled using Velvet assembler ver. 1.2.07 (Zerbino and Birney, 2008) with parameters that were optimized based on the nucDNA and kpDNA sequence coverage depths; the estimated nucDNA depth was approximately 30× based on the k-mer analysis, and the predicted kpDNA depth was 272× based on the read mapping to previously obtained kleptoplast *rbcL* sequences (AB619313; 1195 bp) using Bowtie2 ver. 2.0.0 (Langmead and Salzberg, 2012). After several tests to tune the Velvet parameters, the best assembly was achieved with a k-mer of 83 and exp_cov of 50. The resulting

assembly comprised 1,537 scaffolds (>2000 bp) containing 4,743,113 bp (N50 = 2830 bp). We then identified two kpDNAs (AP014542 and AP014543) from this assembly based on the sequence similarity with *H. borneensis* cpDNA and a mapping back analysis. BLASTX ver. 2.2.31 assigned bit scores >1000 to the two scaffolds (AP014542 = 1382, AP014543 = 1373, database = coding sequences in the obtained *H. borneensis* cpDNA, query = all constructed scaffolds). Mapping back, which was performed using BWA ver. 0.7.15-r1140, showed that the coverage depths of AP014542 and AP014543 correlated with the relative abundance of kleptoplasts; the average coverage depth of the read was increased by 2–4 fold in a library from kleptoplast-rich tissue (DRR063263) relative to a library from kleptoplast-poor tissue (DRR063261). The same degree of change was never observed in other scaffolds (Supplementary Fig. 33).

The two kleptoplast sequences were annotated and visualized using the same method described for *H. borneensis* cpDNA. The phylogenetic positions of PoB kleptoplasts and algal chloroplasts were analyzed using the *rbcL* gene sequences from 114 ulvophycean green algae according to the maximum likelihood method (Fig. 3b, Supplementary Fig. 5). A phylogenetic tree was constructed according to the same method used to search for the origins of bacteria-like genes in *H. borneensis* cpDNA.

Analysis of the *H. borneensis* and *B. hypnoides* transcriptomes

De novo transcript profiles of *H. borneensis* and *B. hypnoides* were obtained from Illumina RNA-Seq data. We extracted total RNA using the RNeasy Plant Mini Kit (Qiagen), constructed single Illumina libraries using the TruSeq RNA Sample Prep Kit, and sequenced the library of each species (101 bp from each end) on a HiSeq 2500 platform. A total of 290,523,622 reads (29 Gb) and 182,455,350 raw reads (18 Gb) were obtained for *H. borneensis* and *B. hypnoides*, respectively. After filtering the low-quality and adapter sequences, the obtained reads were assembled using Trinity version 2.4.0 (Grabherr et al., 2011) and clustered using CD-Hit version 4.6 (Fu et al., 2012) with the -c 0.95 option. The TransDecoder (Haas et al., 2016) version 2.0.1 was used to identify 26,652 and 24,127 candidate coding regions from *H. borneensis* and *B. hypnoides*, respectively.. The gene completeness of the transcripts was estimated using BUSCO version 2.0 (Waterhouse et al., 2018). The obtained *H. borneensis* transcripts covered 86.5% (262/303) of the total BUSCO groups), while the *B. hypnoides* transcripts covered 92.7% (281/303) of the conserved genes in Eukaryota (eukaryota_odb9 database). The transcripts were annotated using AHRD version 3.3.3 (https://github.com/groupschoof/AHRD) based on the

results of a BLASTP search against nr, RefSeq, and *Chlamydomonas* proteome dataset on UniProt. The composed functional domains on the transcripts were annotated using InterProScan version 5.23-62 (Jones et al., 2014). To distinguish the reliable target species transcripts, we predicted the original transcript species using MEGAN version 5 (Huson et al., 2007) and selected 11,629 and 8,630 transcripts as viridiplantal genes. We have presented the details of the annotation procedure visually in Supplementary Figs 8 and 9.

We manually selected a query dataset from the algal transcripts to search algae-derived genes on the sea slug DNAs. We selected 176 and 129 transcripts from *H. borneensis* and *B. hypnoides*, respectively. To perform more comprehensive searches, we also obtained queries from three public genomic datasets derived from *Caulerpa lentillifera*, *Chlamydomonas reinhardtii*, and *Cyanidioschyzon merolae*; these queries were termed the A614 dataset (10.6084/m9.figshare.12318947).

Sequencing of the *P. ocellatus* type black genome

The mean nucDNA size in three PoB individuals was estimated using flow cytometry. Dissected parapodial tissue (5 mm^2) was homogenized in 1 ml of PBS buffer containing 0.1 % triton X-100 (Thermo) and 0.1 % RNase A (Qiagen) using a BioMasher (Nippi, Tokyo, Japan). We then filtered the homogenate through a 30-μm CellTrics filter (Sysmex, Hyogo, Japan) and diluted the filtrate with PBS buffer to a density of $<5\times10^6$ cell/ml. The resulting solution was mixed with genome size standard samples and stained with a 2% propidium iodide solution (SONY, Tokyo, Japan). We used *Acyrthosiphon pisum* (genome size = 517 Mb) and *Drosophila melanogaster* (165 Mbp) samples processed using the same method described for PoB as genome size standards. The mixture was analyzed on a Cell Sorter SH800 (SONY) according to the manufacturer's instructions. We repeated the above procedure for three PoB individuals and determined an estimated genome size of 936 Mb.

Genomic DNA was extracted from a single PoB individual using the CTAB-based method (Murray and Thompson, 1980). The adapted buffer compositions are summarized in Supplementary Table 18. A fresh PoB sample (collected on October 17, 2013, and starved for 21 days) was cut into pieces and homogenized in 2×CTAB buffer using a BioMasher. To digest the tissues, we added a 2% volume of Proteinase K solution (Qiagen) and incubated the sample overnight at 55º C. The lysate was emulsified by gentle inversion with an equal volume of chloroform; after centrifugation (12,000× g, 2 min), the aqueous phase was collected using a pipette. This phase was combined with a one-tenth volume of 10% CTAB buffer, mixed well at 60 ºC for 1 min,

and again emulsified with chloroform. These 10% CTAB buffer and chloroform treatment steps were repeated until a clear aqueous phase was achieved. We then transferred the aqueous phase to a new vessel, overlaid an equal volume of CTAB precipitation buffer, and mixed the liquids gently by tapping. The resulting filamentous precipitations (DNA) were removed using a pipette chip and incubated at room temperature for 10 min in High Salt TE buffer. We then purified the DNA according to the protocol supplied with the DNeasy Blood and Tissue Kit (Qiagen); briefly, we transferred the supernatant after vortex mixing, added equal volumes of buffer AL (supplied with the kit) and EtOH to the supernatant, and processed the sample on a Qiagen spin column according to the protocol. We obtained a final DNA quantity of 15 μg from a PoB individual.

The genomic sequence of PoB was obtained via Illumina paired-end and mate-pair DNA sequencing. Two paired-end Illumina libraries containing 250- and 600-bp insertions (DRR029525, DRR029526) were constructed using a TruSeq DNA Sample Prep Kit (Illumina). Three mate-pair libraries with 3k-, 5k-, and 10k-bp insertions (DRR029528, DRR029529, DRR029530) were constructed using a Nextera Mate Pair Library Prep Kit (Illumina). The libraries were sequenced (150 bps from each end) on a HiSeq 2500 platform (Illumina). A total of 1,130,791,572 and 787,040,878 raw reads were obtained for the paired-end (170 Gb) and mate-pair (118 Gb) libraries, respectively. After filtering the low-quality and adapter sequences, the remaining 161 Gb of sequences were assembled using Platanus assembler version 1.2.1 (Kajitani et al., 2014) with the default setting. The assembly comprised 8,716 scaffolds containing 928,345,517 bp. Repetitive regions were masked using a combination of RepeatModeler version open-1.0.8 and RepeatMasker version open-4.0.5 (http://www.repeatmasker.org). We used the default parameters for identification and masking. A total of 268,300,626 bp (29%) of the assemblies were masked with RepeatMasker.

We used strand-specific RNA-Seq sequencing for gene modeling. PoB RNA was extracted from an individual after starvation for 20 days (collected on October 17, 2013). We used TRIzol™ Plus RNA Purification Kit (Thermo) to extract RNA according to the manufacturer's protocol. A paired-end Illumina library (DRR029460) was constructed using the TruSeq Stranded mRNA LT Sample Prep Kit (Illumina). Libraries were sequenced (150 bp from each end) on a HiSeq 2500 platform (Illumina). The library produced a total of 286,819,502 raw reads (28 Gb).

The PoB assemblies were processed using single transcript-based gene model construction pipeline (AUGUSTUS version 3.2) (Stanke and Morgenstern, 2005), two transcriptomic data mapping tools (Trinity

version 2.4.0 and Exonerate version 2.2.0) (Grabherr et al., 2011; Slater and Birney, 2005), and two non-transcript-based model construction pipelines (GeneMark-ES version 4.33 and glimmerHMM version 3.0.4) (Majoros et al., 2004; Ter-Hovhannisyan et al., 2008). The four obtained gene sets were merged with the EVidenceModeler version 1.1.1 pipeline (Haas et al., 2008) to yield a final gene model set. For AUGUSTUS, we used Braker pipeline version 1.9 (Hoff et al., 2019) to construct PoB-specific probabilistic models of the gene structure based on strand-specific RNA-Seq data. After filtering the low-quality and adapter sequences, the remaining 181,873,770 RNA-Seq reads (16 Gb) were mapped to the PoB genome assembly using TopHat version v2.1.1 (Kim et al., 2013) with the default setting, as well as to the Braker pipeline-constructed PoB-specific probabilistic models from mapped read data. TopHat mapped 132,786,439 of the reads (73%) to the PoB model. AUGUSTUS then predicted 78,894 gene models from the TopHat mapping data (as splicing junction data) and the Braker probabilistic model. Using Trinity and Exonerate, we then constructed de novo transcriptomic data from the RNA-Seq data and aligned these to the genome. Trinity constructed 254,336 transcripts, which were clustered to 194,000 sequences using CD-Hit version 4.6 (-c 0.95); subsequently, Transdecoder identified 44,596 protein-coding regions from these sequences. Exonerate (--bestn 1 --percent 90 options) then aligned the 13,141 of the transcripts to the genome. GeneMark-ES with the default setting predicted 107,735 gene models, and glimmerHMM predicted 115,633 models after the model training, with 320 manually constructed gene models from long scaffolds. EVidenceModeler was then used to merge the model with the following weight settings: AUGUSTUS = 9, Exonerate = 10, GeneMark-ES = 1, glimmerHMM = 2. Finally, EVidenceModeler predicted 77,444 gene models.

We then removed the contaminant-derived bacterial scaffolds from the PoB assemblies. We defined bacterial scaffolds as those encoding >1 bacterial gene with no lophotrochozoan gene. The bacterial genes were predicted using MEGAN software according to a BLASTP search against the RefSeq database. Of the 40,330 gene hits identified from the RefSeq data, MEGAN assigned the origins for 39,113 genes. Specifically, 719 and 23,559 genes were assigned as bacterial and lophotrochozoan genes, respectively. Fifty-five of the 8,716 scaffolds contained two or more bacterial genes and no lophotrochozoan gene and were removed (Supplementary Table 8).

We also removed kleptoplast- or mitochondria-derived scaffolds based on the blast search results. A Blastn search (query; all scaffolds, database; previously obtained PoB kleptoplast- and mitochondrial DNA)

identified 12 kleptoplast-hitting scaffolds and one mitochondria-hitting scaffold (bit score >1000, Supplementary Table 9). Hence, we removed 13 scaffolds from the final PoB DNA assemblies.

The final PoB assembly comprised 8,647 scaffolds containing 927,888,823 bp (N50 = 1,453,842 bp) and 77,230 genes. Gene completeness was estimated using BUSCO version 2.0 (Waterhouse et al., 2018). The predicted gene models were annotated using AHRD version 3.3.3. The results of a Blastp search against the SwissProt, Trembl, and *Aplysia californica* proteome datasets on UniProt were used as reference data for AHRD under the following weight parameter settings: SwissProt = 653, Trembl = 904, and *A. californica* = 854. The functional domains were annotated using InterProScan version 5.23-62.

We performed a Blastp analysis against the RefSeq database to identify algae-derived genes from the constructed genes models. After translating the protein-encoding region to amino acid sequence data, we adapted the BLASTP search to include the "-e-value 0.0001" option. The output was analyzed using MEGAN software with the following LCA and analysis parameters: Min Score = 50, Max Expected = 1.0E-4, Top Percent = 20, Min Support Percent = 0.1, Min Support = 1, LCA percent = 90, and Min Complexity = 0.3.

The GO annotation was assigned using Blast2GO version 5.2.5 according to the BLASTP searches against the RefSeq database and InterProScan results. We then used SonicParanoid version 1.0.11 for orthogroup detection. The species analyzed in the orthogroup detections are summarized in Supplementary Table 10. The phylogenetic tree was constructed using IQ-tree. The resulting trees were visualized using iTol version 4.

We used Exonerate version 2.2.0 (with the --bestn 1 --model protein2genom options) to identify algal genes in the PoB genome. We used the A614 dataset as a query after translating the sequence to amino acids. *Caulerpa lentillifera* (green algae) genomic data were used as a control to estimate the sensitivity of our method. The results were handled and visualized using R (tidyverse packages version 1.2.1).

We used MMseq2 version 2.6 (--orf-start-mode 1) to search for algae-like reads among the trimmed Illumina reads. The matching threshold was set using a default E-value <0.001. As a positive control, we selected PoB genes from the BUSCO analysis of our genomic model data. We selected 911 gene models detected by BUSCO version 2 as single-copy orthologs of the metazoa_odb9 gene set and named the dataset P911 (10.6084/m9.figshare.12318977).

The horizontal gene transfer index “h” and the modified index “hA” were calculated using our R script, HGT_index_cal.R (https://github.com/maedat/HGT_index_cal; R version 3.6.1). The “h” index was calculated as the difference in bit scores between the best prokaryote and best eukaryote matches in the BLAST alignments, and “hA” was calculated as the difference in bit scores between the best lophotrochozoan and best algae matches. The BLAST databases of adapted species are summarized in Supplementary Table 12.

RNA-Seq analysis of *P. ocellatus* type black tissues

Total RNA samples from five PoB individuals and one egg mass were obtained for a gene expression analysis. An overview of sample preparation is illustrated in Supplementary Fig. 22. Collected adult PoB individuals were dissected manually after an incubation of 21–94 days. An egg mass was obtained via spontaneous egg lying in an aquarium. We used a TRIzol™ Plus RNA Purification Kit (Thermo) to extract RNA according to the manufacturer’s protocol. We constructed six paired-end and nine single-end Illumina libraries ("Poc_RNA-" libraries) using a combination of the RiboMinus Eukaryote Kit (Thermo), RiboMinus concentration module (Thermo), and TruSeq RNA Sample Preparation Kit v2 (Illumina) according to the manufacturers’ protocols. Libraries were sequenced (101 bp) on a HiSeq 2500 platform (Illumina). A total of 280,445,422 raw reads (28 Gb) were obtained from the libraries (DOI 10.6084/m9.figshare.12301277).

After filtering the low-quality and adapter sequences, 150,701,605 RNA-Seq reads (13 Gb) were obtained. We used only the R1 reads for the six paired-end datasets. We used MMseq2 to identify algae-derived reads from the trimmed reads, using the A614 dataset as a query. We applied the same parameters as the above-described DNA read search. The PoB gene dataset P911 was also used as a positive control.

We used the Hisat-stringtie-DESeq2 pipeline to conduct a differential gene expression analysis of the digestive gland (DG) and the DG-exenterated parapodia (epidermis, muscle, and reproductive systems, Dep). According to the Stringtie protocol manual (http://ccb.jhu.edu/software/stringtie/index.shtml?t=manual), trimmed RNA-Seq reads were mapped to the PoB genome assembly using Hisat2 version 2.1.0 with the default setting. The obtained BAM files were processed using Stringtie version 1.3.4d (-e option) with PoB gene model data (gff3 format) acquired through the above-mentioned EVidenceModeler analysis. The resulting count data were analyzed using R and the DESeq2 package, and 1,490 differentially expressed genes (p-value <0.01 and padj <0.05) were identified between the tissues. We used the Goseq (Young et al., 2010) and topGO (Alexa and

Rahnenfuhrer, 2010) packages in R to apply a GO enrichment analysis to the upregulated genes in DG tissue (threshold: p-value <0.01).

Sequencing of the *E. marginata* genome

The *E. marginata* genome sequencing process was nearly identical to the methodology applied to PoB. Flow cytometry yielded an estimated genome size of 900 Mb. We extracted genomic DNA from an individual using a CTAB-based method. We constructed four types of Illumina libraries: two paired-end libraries with 250- and 500-bp insertions, and two mate-pair libraries with 3k- and 5k-bp insertions (DOI 10.6084/m9.figshare.12301277). Using the HiSeq 2500 platform (Illumina), 562,732,268 and 608,977,154 raw reads were obtained for the paired-end (84 Gb) and mate-pair (91 Gb) libraries, respectively. After filtering the low-quality and adapter sequences, the remaining 40 Gb of sequences were assembled using the Platanus assembler. The assembly comprised 14,285 scaffolds containing 791,005,940 bp.

For gene modeling, strand-specific RNA-Seq sequencing of *E. marginata* was performed. A paired-end Illumina library (DRR029460, also see DOI 10.6084/m9.figshare.12301277) was constructed using a TruSeq Stranded mRNA LT Sample Prep Kit (Illumina) and an RNA sample extracted from an *E. marginata* individual via a TRIzol™ Plus RNA Purification Kit (Thermo). Libraries were sequenced (150 bp from each end) on a HiSeq 2500 platform (Illumina). A total of 286,819,502 raw reads (28 Gb) were obtained from the library. Using the gene modeling procedure described for PoB, EVidenceModeler constructed 71,137 gene models of the genomic assemblies based on the RNA-Seq data.

We next removed the contaminant-derived bacterial scaffolds from the genomic assemblies. Using the same gene annotation as applied to PoB, we determined that the 110 of the 14,285 scaffolds contained >1 bacterial genes and no lophotrochozoan gene and removed these scaffolds (Supplementary Table 8).

The organelle-derived scaffolds (kleptoplast DNA and mitochondrial DNA) were identified using BLASTN searches and removed from the final assemblies. A BLASTN search (query = all scaffolds, database = chloroplast DNA of *B. hypnoides* NC_013359.1 or mitochondrial DNA of PoB) identified 25 kleptoplast-matching and one mitochondria-matching scaffold (bit score >1000). We then reassembled the complete kleptoplast DNA and mitochondrial DNA using the same method as described for PoB organellar DNA assembling. We obtained two circular sequences using the "Eor_DNA_WL_250" library, FLASH, and the

Velvet assembler. We then aligned the *B. hypnoides* cpDNA and PoB mitochondrial DNA to the assemblies to confirm the respective origins of these assemblies (kpDNA and mtDNA).

Ortholog analysis of sacoglossan genes

Orthologous relationships were classified using Orthofinder (ver. 2.2.3) (Emms and Kelly, 2015), and rapidly expanded/contracted families identified from the Orthofinder results were analyzed using CAFE (ver. 4.2) (Han et al., 2013). CAFE analysis used 16 metazoan species as reference species (Supplementary Fig. 23). Phylogenetic trees for CAFE were constructed using PREQUAL (ver. 1.02) (Whelan et al., 2018) for sequence trimming, MAFFT (ver. 7.407) for sequence alignment, IQ-tree (ver. 1.6.1) for the maximum likelihood (ML) analysis, and r8s (v1.81) (Sanderson, 2003) for conversion to an ultrametric tree. An ML tree was constructed from 30 single-copy genes in 15 major species according to the Orthofinder results (DOI 10.6084/m9.figshare.12319862), and was converted to an ultrametric tree based on the divergence times of Amphiesmenoptera-Antliophora (290 Myr) and Euarchontia-Glires (65 Myr).

We also analyzed the expanded/contracted families using the Z-scores of the assigned gene numbers for each orthogroup. The analysis included all 16 reference species and two sacoglossan species, and an expanded group of 38 was determined on the PoB lineage (Threshold: Z-score > 2) (Supplementary Fig. 27). The phylogenetic relationships in the orthogroups were also analyzed using a PREQUAL-MAFFT-IQ-tree. The domain structures and gene positions on the constructed genome data were visualized using the GeneHere script (https://github.com/maedat/GeneHere) and Biopython packages.

Author contributions:

T.Maeda, S.T., and S.Shigenobu. conceived of and designed the experiments; T.Maeda, S.T., J.M., performed the photochemical and physiological experiments and analyses; T.Maeda, A.A., T.Y., S.Shimamura., Y.T., Y.N., K.T., T.T., Y.S., M.K., N.S., T.N., M.H., T. Maruyama, J.O., and S.Shigenobu performed the genomic and transcriptomic experiments and analyses. T.Maeda and S.Shigenobu wrote the paper.

Acknowledgment:

We appreciate the incisive comments offered by Masayoshi Kawaguchi, Atsushi J. Nagano, and Kan Tanaka. The sample collection was supported by Katsuhiko Tanaka, Tohru Iseto, Rie Nakano, and Hirose Euichi. This work was supported by the MEXT/JSPS KAKENHI (grant numbers 16H06279, 25128713 and 22128001), National Institute for Basic Biology (NIBB) Functional Genomics Facility, NIBB Data Integration and Analysis Facility, and the Japan Advanced Plant Science Network. Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics.

Competing Interests:

The authors declare no competing financial interests.

Data availability:

All of the raw sequence data obtained in this research have been deposited in the DDBJ Sequence Read Archive (DRA) under BioProject PRJDB4939, PRJDB3267 and PRJDB5024. The data that support the findings of this study are available from the corresponding author upon request. A part of the supporting data has been deposited in the FigShare (https://figshare.com) under DOI

10.6084/m9.figshare.12300869 (Sequence data set for *Plakobranchus* phylogeny *i*n Fig. 1)

10.6084/m9.figshare.12301865 (Raw dataset for oxygen concentration measurement in Fig. 2)

10.6084/m9.figshare.12301277 (Summary of the sequenced raw data)

10.6084/m9.figshare.12311990 (Encoded genes on kleptoplast chloroplast DNA)

10.6084/m9.figshare.12316163 (Orthologous relationship of the kleptoplast/chloroplast-encoded genes)

10.6084/m9.figshare.12316283 (Annotation list of *H. borneensis* de novo transcripts)

10.6084/m9.figshare.12316868 (Annotation list of *Bryopsis hypnoides* de novo transcripts)

10.6084/m9.figshare.12316895（Annotation list of *Caulerpa lentillifera* gene models)

10.6084/m9.figshare.12318947（Algal genes used for the referential query, A614 data set)

10.6084/m9.figshare.12318962 (TBLASTN hit result)

10.6084/m9.figshare.12318974（TBLASTN E-values for Fig. 3)

10.6084/m9.figshare.12318977 (*P. ocellatus* gene annotation)

10.6084/m9.figshare.12318989 (*P. ocellatus* gene models showing top-hit with alga sequence)

10.6084/m9.figshare.12318992 (*Aplysia californica* annotation)

10.6084/m9.figshare.12318998 (*Biomphalaria glabrata* annotation)

10.6084/m9.figshare.12319001 (Genes assigned for the child gene ontology "Plastid/Photosynthesis")

10.6084/m9.figshare.12319424 (Plastid-relating *P. ocellatus* genes)

10.6084/m9.figshare.12319532 (Orthogroups constructed by SonicParanoid)

10.6084/m9.figshare.12318920 (Exonerate searches against *P. ocellatus* and *C. lentillifera* genome)

10.6084/m9.figshare.12319739 (Result of MMseq2 search against raw DNA/RNA-Sequence reads)

10.6084/m9.figshare.12319736 (HGT indexes obtained from BLASTP results)

10.6084/m9.figshare.12319802 (DEG results between the digestive gland and the parapodia)

10.6084/m9.figshare.12319826 (Differentially expressed genes between DG and DeP tissue)

10.6084/m9.figshare.12319832 (Enriched GOs of the upregulated genes on the digestive gland)

10.6084/m9.figshare.12319844 (Clustered gene set by OrthoFinder analysis)

10.6084/m9.figshare.12318908 (Applied species for the comparative genomic analysis)

10.6084/m9.figshare.12319853 (Gene number of each orthologous groups)

10.6084/m9.figshare.12319859 (Rapidity expanded/contracted orthogroups by CAFE analysis)

10.6084/m9.figshare.12319862 (Genes for the phylogenetic tree construction for CAFE analysis)

10.6084/m9.figshare.12319889 (Annotation of the predicted gene models on the *E. marginata* genome)

Figure legends

Fig. 1. Kleptoplasty in sea slugs.

a The process of algal chloroplast retention by a sacoglossan sea slug(Pierce and Curtis, 2012). **b–d** Images of *P. ocellatus* type black starved for 21 days. **b** Dorsal view. **c** Spread parapodia. H, head; P, parapodium; DG, digestive gland. **d** Enlarged view of the surface of the parapodium and digestive gland. Kleptoplasts are visible as a green color. **e** Phylogenetic distribution of kleptoplasty in the order Sacoglossa. Phylogenetic tree and kleptoplasty states are simplified from Christa *et al.* 2015(Christa et al., 2015). Relationships within Heterobranchia are described based on Zapata *et al.* 2014(Zapata et al., 2014). **f** Phylogeny of the *P.* cf. *ocellatus* species complex based on mitochondrial *cox1* genes (Maximum Likelihood tree from 568 nucleotide positions) from INSDC and our whole mitochondrial DNA sequence. Clade names in square brackets are based on Krug *et al.* (2013)(Krug et al., 2013). Asterisks mark genotypes from Krug *et al.* (2013)(Krug et al., 2013). Previously analyzed topics for each cluster are described within the colored boxes. Small black circles indicate nodes supported by a high bootstrap value (i.e., 80%–100%). *Thuridilla gracilis* is outgroup. *Plakobranchus papua* is recently described species and previously identified as *P. ocellatus(Meyers-Muñoz et al., 2016)*. The detailed data for the samples were registered in FigShare (DOI, 10.6084/m9.figshare.12300869)

Fig. 2. Photosynthetic activity of *P. ocellatus* type black

a Jitter plot of Fv/Fm values indicating the photochemical efficiency of photosystem II. Habo, *H. borneensis*; PoB, *P. ocellatus* type black; d38, starved for 38 days; d109, starved for 109–110 day (12 h:12 h light-dark cycle, the light phase illumination was 10 µmol photons m^{-2} s^{-1}). Magenta line indicates the mean value, black dot indicates the raw value of each individual (n = 3, each group). **b** Time-course of oxygen concentration in the water in which PoB was reared. Gray color signifies a dark period; yellow color signifies an illuminated period (50 µmol photons m^{-2} s^{-1}). Temp, temperature. **c** Jitter plots of PoB oxygen consumption and generation. D, dark condition; L, light condition; G, gross rate of light-

dependent oxygen generation (L minus D). **d** Jitter plots of PoB longevity (n = 5, each group; raw data: Supplementary Table 1). D, Continuous dark; L/D, 12 h:12 h light-dark cycle. The *p-value* from Welch's two-sample *t*-test is shown. Raw data, Supplementary Table 1.

Fig. 3. Gene composition of *P. ocellatus* type black kleptoplast DNAs

a Gene map of two kleptoplast DNAs from PoB. Gene positions are described in circles colored according to the functional category of the gene (see key). Genes on the outside and inside of each circle are transcribed in the clockwise and anticlockwise direction, respectively (for detailed maps, see Supplementary Figs. 2 and 3). **b** Phylogenetic positions of sequenced kleptoplasts among green algal plastids. The tree was simplified from an ML tree based on *rbcL* genes (457 positions). When multiple sequences derived from the same species were registered in the INSDC due to intra-species variation, these multiple sequences were included in the analysis (original tree: Supplementary Fig. 5). Red color indicates sequenced kpDNA or cpDNA in the present study. Underlining indicates algal species used for our RNA-Seq sequencing. **c** An UpSet plot of plastid gene composition. Species abbreviations are defined in Supplementary Table 3. The horizontal bar chart indicates the gene numbers in each species. The vertical bar chart indicates the number of genes conserved among the species. Intersect connectors indicate species composition in which a given number of genes (vertical bar chart). We omitted connections corresponding to no gene. Connectors are colored according to the conservation level of the gene (see key): Core gene, conserved among all of the analyzed Bryopsidales species; Dispensable gene, retained over 2 Bryopsidales species; Rare gene, determined from a single or no Bryopsidales species. Gray shading indicates non-Viridiplantae algae, and magenta shading indicates PoB kleptoplasts. **d** Box plots of tblastn results. The y-axis shows the database searched (kPoro and kRhip, PoB kleptoplast DNAs; nCale, nucDNA of *Caulerpa lentillifera*). Each dot represents the tblastn result (query is A614 dataset). Red dots show the result using the *chlD* gene (encoding Magnesium-chelatase subunit ChlD) as the query sequence; this sequence is similar to the kleptoplast-encoded *chlL* gene. The right pie chart shows the proportion of queries with hits (E-value < 0.0001). **e** Heat map of tblastn results of representative photosynthetic nuclear genes (a subset of data in Fig. 3d). The source species of the query sequences are described on the top: Abbreviations are defined in Supplementary Table 3. Raw data: DOI

10.6084/m9.figshare.12311990, 10.6084/m9.figshare.12316163, 10.6084/m9.figshare.12318962, and 10.6084/m9.figshare.12318974.

Fig. 4 Search for horizontally transferred algal-genes in the *P. ocellatus* type black genome

a Heatmap of gene ontology (GO) comparison analysis among PoB, two non-kleptoplastic mollusk species (Aca, *Aplysia californica*; Bgl, *Biomphalaria glabrata*), and 5 algae species (abbreviations as defined in Supplementary Table 3). For each species, the number of genes assigned to various photosynthesis- or plastid-related GO terms are visualized on a color scale (see key). **b** Scatter plot of the results of the alignment of the A614 gene set (query sequences) of photosynthetic genes to the PoB genome using Exonerate software. The enlarged view is shown on Supplementary Fig. 17 with tblastn result). The dot color shows the source algae of each query sequence (see key). The horizontal axis shows what percentage of the query sequences were aligned to the hit sequences (PoB genome). The vertical axis shows the similarity of the aligned sequences between the query and PoB genome; alignment score (the sum of the substitution matrix scores and the gap penalties) divided by aligned length (bp). Dashed lines are thresholds for a credible query hit (i.e., a hit covering >60% of the query sequence, and a normalized Exonerate alignment score of >2). **c** Scatter plot of MMseq2 results for the A614 dataset (algal photosynthetic genes, red) and P911 reference dataset (PoB single-copy genes, blue) used as query sequences against our database of preassembled read sets from paired-end DNA libraries of PoB. The upper panel shows the probability density distribution of the number of hit-reads (normalized with TPM: transcripts per kilobase million) (x-axis) versus averaged “pident” value (percentage of identical matches) from the hit reads (y-axis). **d** Scatter plot of HGT indexes (hA- versus h-index) for genes in PoB, the two non-kleptoplastic mollusk species (Aca and Bgl), and one algae species (*Caulerpa lentillifera* [Cale]). Each dot represents a gene. A high hA or h-index value means the possibility of algal or prokaryote origin, respectively. Dashed red lines represent the conventional threshold for HGT (-100 for h index(Boschetti et al., 2012) and 100 for hA index).. **e** Heatmap of the results of searches for algae-like RNA fragments in the PoB RNA-Seq data. DG, digestive gland; Pa, parapodia; DeP, DG-exenterated parapodia; EG, egg; He, head; Pe, pericardium. The blue color gradient indicates the number of RNA-Seq reads assigned as algae-like fragments (see key). The y-axis labels show the RNA-seq library name and analyzed tissue

types (see Supplementary Fig. 21). The x-axis labels indicate the query protein; those with no corresponding RNA reads were omitted from the figure. For queries using the P911 reference dataset, we describe the mean value of the hit-read counts from each library. The total number of reads for each library is given on the far right. Detailed and raw data: Supplementary Figs 14–21, 10.6084/m9.figshare.12318920, 10.6084/m9.figshare.12319739, and 10.6084/m9.figshare.12319736.

Fig. 5 Probable kleptoplasty-related molluscan genes in *P. ocellatus* type black

a Volcano plot comparing gene expression in the digestive gland (DG) and DG-exenterated parapodia (DeP) tissue of PoB. Red indicates orthogroup OG0000132 (cathepsin D-like protease genes); blue indicates orthogroup OG0000005 (lectin-like genes); and orange indicates other orthogroups (for details, see Supplementary Fig. 25). Up and down arrows signify up- and down-regulated, respectively, in DG. **b** Orthogroups that were expanded on the *P. ocellatus* lineage and contained DG-upregulated genes. PoB, *P. ocellatus* type black; Ema, *E. marginata*; Bgl, *Biomphalaria glabrata*; Lgi, *Lottia gigantea*; Hdi, *Haliotis discus*; Obi, *Octopus bimaculoides*; Pfu, *Pinctada fucata*; Cgi, *Crassostrea gigas*; Cte, *Capitella teleta*; Dme, *Drosophila melanogaster*; Bmo, *Bombyx mori*; Api, *Acyrthosiphon pisum*; Hsa, *Homo sapiens*; Mmu. *Mus musculus.* The phylogenetic tree is scaled to divergence time based on 30 conserved single-copy genes. Mya = million years. The numbers of rapidly expanded (blue) and contracted (magenta) orthogroups on the lineages to PoB are provided at the nodes (Detail data: Supplementary Fig. 24). Below the tree is information for the six expanded orthogroups that contained DG-upregulated genes. The left-side heatmap shows the gene numbers (number in boxes) and z-score of gene numbers (color gradient) for each orthogroup. The table shows the expanded/not expanded status of each orthogroup (P–E, Plakobranchus–Elysia node; Poc, Plakobranchus node) (for details, see Supplementary Figs. 26 and 27). The right-side heat map indicates the number of differentially expressed genes (DEGs) between DG and DeP tissue in each orthogroup. Representative gene products are given on the far right.. **c** Details for OG0000132 (phylogeny, domain structure, gene expression, and gene positions on the assemblies). The phylogenetic tree on the far left is a part of the ML tree for OG0000132 genes (Supplementary Fig. 28). Red circles mark PoB genes; blue squares mark *E. marginata* (Ema) genes. Other species are represented with pictograms defined in 10.6084/m9.figshare.12319844. The domain structure of the proteins encoded

by OG0000132 genes is shown in the centre left (see “Domain structure” key on far right). The expression of each gene in OG0000132 in various tissues is shown in the centre right; tissue abbreviations are as defined in the Fig. 4 legend (see “Gene expression” key on far right). Genes derived from organisms other than PoB are shown in gray. The false discovery rate (FDR) was calculated by the comparison between DG and DeP samples. Genes for which no expression was observed on DG nor DeP (FDR could not be calculated) are shown in black. In the right panel, the genomic positions of OG0000132 genes on scaffolds are shown as red boxes, and other protein-encoding genes are shown as blue boxes. Purple text and arrows indicate the scaffold ID and direction, respectively. The color of the bands indicates the correlation between the gene and position. Scaffolds having less than 5 genes belonging to OG0000132 were omitted from the figure. Raw data: DOI 10.6084/m9.figshare.12319802, 10.6084/m9.figshare.12319826, 10.6084/m9.figshare.12319832, 10.6084/m9.figshare.12319844, 10.6084/m9.figshare.12318908, 10.6084/m9.figshare.12319853 10.6084/m9.figshare.12319859, and 10.6084/m9.figshare.12319862

References

Acuña R, Padilla BE, Flórez-Ramos CP, Rubio JD, Herrera JC, Benavides P, Lee S-J, Yeats TH, Egan AN, Doyle JJ, Rose JKC. 2012. Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. *Proc Natl Acad Sci U S A* **109**:4197–4202.

Alexa A, Rahnenfuhrer J. 2010. topGO: enrichment analysis for gene ontology. *R package version* **2**:2010.

Anderson NG. 1970. Evolutionary significance of virus infection. *Nature* **227**:1346–1347.

Arber W. 2014. Horizontal gene transfer among bacteria and Its role in biological evolution. *Life* **4**:217–224.

Arimoto A, Nishitsuji K, Higa Y, Arakaki N, Hisata K, Shinzato C, Satoh N, Shoguchi E. 2019. A siphonous macroalgal genome suggests convergent functions of homeobox genes in algae and land plants. *DNA Res* **26**:183–192.

Benes P, Vetvicka V, Fusek M. 2008. Cathepsin D--many functions of one aspartic protease. *Crit Rev Oncol Hematol* **68**:12–28.

Bessho-Uehara M, Yamamoto N, Shigenobu S, Mori H, Kuwata K, Oba Y. 2020. Kleptoprotein bioluminescence: Parapriacanthus fish obtain luciferase from ostracod prey. *Sci Adv* **6**:eaax4942.

Bhattacharya D, Pelletreau KN, Price DC, Sarver KE, Rumpho ME. 2013. Genome analysis of *Elysia chlorotica* egg DNA provides no evidence for horizontal gene transfer into the germ line of this kleptoplastic mollusc. *Mol Biol Evol* **30**:1843–1852.

Bishop NA, Lu T, Yankner BA. 2010. Neural mechanisms of ageing and cognitive decline. *Nature* **464**:529–535.

Boschetti C, Carr A, Crisp A, Eyres I, Wang-Koh Y, Lubzens E, Barraclough TG, Micklem G, Tunnacliffe A. 2012. Biochemical diversification through foreign gene expression in bdelloid rotifers. *PLoS Genet* **8**:e1003035.

Boto Luis. 2014. Horizontal gene transfer in the acquisition of novel traits by metazoans. *Proceedings of the Royal Society B: Biological Sciences* **281**:20132450.

Cai H, Li Q, Fang X, Li J, Curtis NE, Altenburger A, Shibata T, Feng M, Maeda T, Schwartz JA, Shigenobu S, Lundholm N, Nishiyama T, Yang H, Hasebe M, Li S, Pierce SK, Wang J. 2019. A draft genome assembly of the solar-powered sea slug *Elysia chlorotica*. *Sci Data* **6**:190022.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:1972–1973.

Cartaxana P, Trampe E, Kühl M, Cruz S. 2017. Kleptoplast photosynthesis is nutritionally relevant in the sea slug

Elysia viridis. *Sci Rep* **7**:7714.

Chan CX, Vaysberg P, Price DC, Pelletreau KN, Rumpho ME, Bhattacharya D. 2018. Active host response to algal symbionts in the sea slug *Elysia chlorotica*. *Mol Biol Evol* **35**:1706–1711.

Charron J-BF, Ouellet F, Houde M, Sarhan F. 2008. The plant Apolipoprotein D ortholog protects *Arabidopsis* against oxidative stress. *BMC Plant Biol* **8**:86.

Christa G, Gould SB, Franken J, Vleugels M, Karmeinski D, Händeler K, Martin WF, Wägele H. 2014a. Functional kleptoplasty in a limapontioidean genus: phylogeny, food preferences and photosynthesis in *Costasiella*, with a focus on *C. ocellifera* (Gastropoda: Sacoglossa). *J Molluscan Stud* **80**:499–507.

Christa G, Händeler K, Kück P, Vleugels M, Franken J, Karmeinski D, Wägele H. 2015. Phylogenetic evidence for multiple independent origins of functional kleptoplasty in Sacoglossa (Heterobranchia, Gastropoda). *Org Divers Evol* **15**:23–36.

Christa G, Händeler K, Schäberle TF, König GM, Wägele H. 2014b. Identification of sequestered chloroplasts in photosynthetic and non-photosynthetic sacoglossan sea slugs (Mollusca, Gastropoda). *Front Zool* **11**:15.

Christa G, Wescott L, Schäberle TF, König GM, Wägele H. 2013. What remains after 2 months of starvation? Analysis of sequestered algae in a photosynthetic slug, *Plakobranchus ocellatus* (Sacoglossa, Opisthobranchia), by barcoding. *Planta* **237**:559–572.

Christa G, Zimorski V, Woehle C, Tielens AGM, Wägele H, Martin WF, Gould SB. 2014c. Plastid-bearing sea slugs fix CO2 in the light but do not require photosynthesis to survive. *Proc Biol Sci* **281**:20132493.

Conrad B, Antonarakis SE. 2007. Gene duplication: a drive for phenotypic diversity and cause of human disease. *Annu Rev Genomics Hum Genet* **8**:17–35.

Crick F. 1970. Central dogma of molecular biology. *Nature* **227**:561–563.

Cruz S, Cartaxana P, Newcomer R, Dionísio G, Calado R, Serôdio J, Pelletreau KN, Rumpho ME. 2015. Photoprotection in sequestered plastids of sea slugs and respective algal sources. *Sci Rep* **5**:7904.

Curtis NE, Pierce SK, Massey SE, Schwartz J, Maugel TK. 2007. Newly metamorphosed *Elysia clarki* juveniles feed on and sequester chloroplasts from algal species different from those utilized by adult slugs **150**:797–806.

Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, Harafuji N, Hastings KEM, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen IA, Necula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang H-G, Awazu S, Azumi K, Boore J, Branno M, Chin-Bow S, DeSantis R, Doyle S, Francino P, Keys DN, Haga S, Hayashi H, Hino K, Imai KS, Inaba K, Kano S, Kobayashi K, Kobayashi M, Lee B-I, Makabe KW, Manohar C,

Matassi G, Medina M, Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H, Ohta F, Oishi K, Rigoutsos I, Sano M, Sasaki A, Sasakura Y, Shoguchi E, Shin-i T, Spagnuolo A, Stainier D, Suzuki MM, Tassy O, Takatori N, Tokuoka M, Yagi K, Yoshizaki F, Wada S, Zhang C, Hyatt PD, Larimer F, Detter C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS. 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298**:2157–2167.

de Vries J, Archibald JM. 2018. Plastid autonomy vs nuclear control over plastid function In: Chaw S-M, Jansen RK, editors. Advances in Botanical Research. Academic Press. pp. 1–28.

de Vries J, Woehle C, Christa G, Wägele H, Tielens AGM, Jahns P, Gould SB. 2015. Comparison of sister species identifies factors underpinning plastid compatibility in green sea slugs. *Proc Biol Sci* **282**. doi:10.1098/rspb.2014.2519

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**:157.

Evertsen J, Burghardt I, Johnsen G, Wägele H. 2007a. Retention of functional chloroplasts in some sacoglossans from the Indo-Pacific and Mediterranean. *Mar Biol* **151**:2159–2166.

Evertsen J, Burghardt I, Johnsen G, Wägele H. 2007b. Retention of functional chloroplasts in some sacoglossans from the Indo-Pacific and Mediterranean. *Mar Biol* **151**:2159–2166.

Fan X, Qiao H, Xu D, Cao S, Ye N. 2014. Short-term retention of kleptoplasty from a green alga (*Bryopsis*) in the sea slug *Placida* sp. YS001 **69**. doi:10.2478/s11756-014-0355-y

Farah J, Rappaport F, Choquet Y, Joliot P, Rochaix JD. 1995. Isolation of a psaF-deficient mutant of *Chlamydomonas reinhardtii*: efficient interaction of plastocyanin with the photosystem I reaction center is mediated by the PsaF subunit. *EMBO J* **14**:4976–4984.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**:3150–3152.

Gautheret D, Lambert A. 2001. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *Journal of Molecular Biology*. doi:10.1006/jmbi.2001.5102

Geijtenbeek TBH, Gringhuis SI. 2009. Signalling through C-type lectin receptors: shaping immune responses. *Nat Rev Immunol* **9**:465–479.

Gilyarov MS. 1983. Appropriation of functioning organelles of food organisms by phytophagous and predatory opisthobranch mollusks as a specific category of food utilization. *Zh Obshch Biol* **44**:614–620.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**:644–652.

Green BJ, Li WY, Manhart JR, Fox TC, Summer EJ, Kennedy RA, Pierce SK, Rumpho ME. 2000. Mollusc-algal chloroplast endosymbiosis. Photosynthesis, thylakoid protein maintenance, and chloroplast gene expression continue for many months in the absence of the algal nucleus. *Plant Physiol* **124**:331–342.

Greve C, Ruiz-Tagle Lui M, Sivalingam S, Ludwig KU, Wägele H, Donath A. 2017. The complete mitochondrial genome of the "solar-powered" sea slug *Plakobranchus* cf. *ocellatus* (Heterobranchia: Panpulmonata: Sacoglossa). *Mitochondrial DNA Part B* **2**:130–131.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**:R7.

Haas B, Papanicolaou A, Others. 2016. TransDecoder (find coding regions within transcripts). *Google Scholar*.

Hackett JD, Maranda L, Yoon HS, Bhattacharya D. 2003. Phylogenetic evidence for the *Cryptophyte* origin pf the plastid of *Dinophysis* (Dinophysiales, Dinophyceae). *J Phycol* **39**:440–448.

Händeler K, Grzymbowski YP, Krug PJ, Wägele H. 2009. Functional chloroplasts in metazoan cells - a unique evolutionary strategy in animal life. *Front Zool* **6**:28.

Han JH, Klochkova TA, Han JW, Shim J, Kim GH. 2015. Transcriptome analysis of the short-term photosynthetic sea slug *Placida dendritica*. *Algae* **30**:303–312.

Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol* **30**:1987–1997.

Hershey AD, Chase M. 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol* **36**:39–56.

Hirose E. 2005. Digestive system of the sacoglossan *Plakobranchus ocellatus* (Gastropoda: Opisthobranchia): light- and electron-microscopic observations with remarks on chloroplast retention. *Zoolog Sci* **22**:905–916.

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-Genome Annotation with BRAKER. *Methods Mol Biol* **1962**:65–95.

Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Res* **17**:377–386.

Izumi M, Tsunoda H, Suzuki Y, Makino A, Ishida H. 2012. RBCS1A and RBCS3B, two major members within the

Arabidopsis RBCS multigene family, function to yield sufficient Rubisco content for leaf photosynthetic capacity. *J Exp Bot* **63**:2159–2170.

Jones P, Binns D, Chang HY, Fraser M, Li WZ, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**:1236–1240.

Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H, Kohara Y, Fujiyama A, Hayashi T, Itoh T. 2014. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res* **24**:1384–1395.

Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, Petrov AI. 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res* **46**:D335–D342.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**:772–780.

Kawaguti. 1965. Electron microscopy on the symbiosis between an elysioid gastropod and chloroplasts of a green alga. *Biol J Okuyama Univ* **11**:57–65.

Kawaguti S. 1941. Study on the invertebrates associating unicellular algae I. *Placobranchus ocellatus* von Hasselt, a nudibranch. *Palau Tropical Biological Station Studies* **2**.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**:R36.

Krug PJ, Vendetti JE, Rodriguez AK, Retana JN, Hirano YM, Trowbridge CD. 2013. Integrative species delimitation in photosynthetic sea slugs reveals twenty candidate species in three nominal taxa studied for drug discovery, plastid symbiosis or biological control. *Mol Phylogenet Evol* **69**:1101–1119.

Laetz EMJ, Wägele H. 2018. Comparing amylose production in two solar-powered sea slugs: the sister taxa *Elysia timida* and *E. cornigera* (Heterobranchia: Sacoglossa). *J Molluscan Stud*. doi:10.1093/mollus/eyy047

Laetz EMJ, Wägele H. 2017. Chloroplast digestion and the development of functional kleptoplasty in juvenile *Elysia timida* (Risso, 1818) as compared to short-term and non-chloroplast-retaining sacoglossan slugs. *PLoS One* **12**:e0182910.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:357–359.

Lis H, Sharon N. 1998. Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem Rev* **98**:637–674.

Maeda T, Hirose E, Chikaraishi Y, Kawato M, Takishita K, Yoshida T, Verbruggen H, Tanaka J, Shimamura S, Takaki Y, Tsuchiya M, Iwai K, Maruyama T. 2012. Algivore or phototroph? *Plakobranchus ocellatus* (Gastropoda) continuously acquires kleptoplasts and nutrition from multiple algal species in nature. *PLoS One* **7**:e42024.

Magoc T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**:2957–2963.

Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**:2878–2879.

Martin R, Walther P, Tomaschko K-H. 2015. Variable retention of kleptoplast membranes in cells of sacoglossan sea slugs: plastids with extended, shortened and non-retained durations. *Zoomorphology* **134**:523–529.

Meyers-Muñoz MA, van der Velde G, van der Meij SET, Stoffels BEMW, van Alen T, Tuti Y, Hoeksema BW. 2016. The phylogenetic position of a new species of *Plakobranchus* from west Papua, Indonesia (Mollusca, Opisthobranchia, Sacoglossa). *Zookeys* 73–98.

Murray MG, Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* **8**:4321–4325.

O'Connor RM, Fung JM, Sharp KH, Benner JS, McClung C, Cushing S, Lamkin ER, Fomenkov AI, Henrissat B, Londer YY, Scholz MB, Posfai J, Malfatti S, Tringe SG, Woyke T, Malmstrom RR, Coleman-Derr D, Altamia MA, Dedrick S, Kaluziak ST, Haygood MG, Distel DL. 2014. Gill bacteria enable a novel digestive strategy in a wood-feeding mollusk. *Proc Natl Acad Sci U S A* **111**:E5096–104.

Pelletreau KN, Bhattacharya D, Price DC, Worful JM, Moustafa A, Rumpho ME. 2011. Sea slug kleptoplasty and plastid maintenance in a metazoan. *Plant Physiol* **155**:1561–1565.

Pierce S, Biron R, Rumpho M. 1996. Endosymbiotic chloroplasts in molluscan cells contain proteins synthesized after plastid capture. *J Exp Biol* **199**:2323–2330.

Pierce SK, Curtis NE. 2012. Cell biology of the chloroplast symbiosis in sacoglossan sea slugs. *Int Rev Cell Mol Biol* **293**:123–148.

Pierce SK, Curtis NE, Hanten JJ, Boerner SL, Schwartz J. 2007. Transfer, integration and expression of functional nuclear genes between multicellular species **43**:57–64.

Pierce SK, Curtis NE, Schwartz J. 2010. Chlorophyll a synthesis by an animal using transferred algal nuclear genes **49**:121–131.

Pierce SK, Massey SE, Hanten JJ, Curtis NE. 2003. Horizontal transfer of functional nuclear genes between

multicellular organisms. *Biol Bull* **204**:237–240.

Pigolev AV, Zharmukhamedov SK, Klimov VV. 2009. The psbO mutant of *Chlamydomonas reinhardtii* is capable of assembling stable, photochemically active reaction center of photosystem II. *Biochemistry (Moscow) Supplement Series A: Membrane and Cell Biology* **3**:33–41.

Rauch C, Vries J de, Rommel S, Rose LE, Woehle C, Christa G, Laetz EM, Wägele H, Tielens AGM, Nickelsen J, Schumann T, Jahns P, Gould SB. 2015. Why It Is time to look beyond algal genes in photosynthetic slugs. *Genome Biol Evol* **7**:2602–2607.

Reyes-Prieto A, Weber APM, Bhattacharya D. 2007. The origin and establishment of the plastid in algae and plants. *Annu Rev Genet* **41**:147–168.

Roberts IN, Miranda H, Tâm LX, Kieselbach T, Funk C. 2013. PsbO degradation by Deg proteases under reducing conditions In: Kuang T, Lu C, Zhang L, editors. Photosynthesis Research for Food, Fuel and the Future. Springer Berlin Heidelberg. pp. 599–602.

Rumpho ME, Pelletreau KN, Moustafa A, Bhattacharya D. 2011. The making of a photosynthetic animal. *J Exp Biol* **214**:303–311.

Rumpho ME, Worful JM, Lee J, Kannan K, Tyler MS, Bhattacharya D, Moustafa A, Manhart JR. 2008. Horizontal gene transfer of the algal nuclear gene psbO to the photosynthetic sea slug *Elysia chlorotica*. *Proc Natl Acad Sci U S A* **105**:17867–17871.

Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**:301–302.

Sasakura Y, Ogura Y, Treen N, Yokomori R, Park S-J, Nakai K, Saiga H, Sakuma T, Yamamoto T, Fujiwara S, Yoshida K. 2016. Transcriptional regulation of a horizontally transferred gene from bacterium to chordate. *Proc Biol Sci* **283**. doi:10.1098/rspb.2016.1712

Schwartz JA, Curtis NE, Pierce SK. 2014. FISH labeling reveals a horizontally transferred algal (*Vaucheria litorea*) nuclear gene on a sea slug (*Elysia chlorotica*) chromosome. *Biol Bull* **227**:300–312.

Serôdio J, Cruz S, Cartaxana P, Calado R. 2014. Photophysiology of kleptoplasts: photosynthetic use of light by chloroplasts living in animal cells. *Philos Trans R Soc Lond B Biol Sci* **369**:20130242.

Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**. doi:Artn 31 10.1186/1471-2105-6-31

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313.

Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* **33**:W465–7.

Taylor DL. 1968. Chloroplasts as symbiotic organelles in the digestive gland of *Elysia viridis* [Gastropoda: opisthobranchia]. *J Mar Biol Assoc U K* **48**:1–15.

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res* **18**:1979–1990.

Teugels B, Bouillon S, Veuger B, Middelburg JJ, Koedam N. 2008. Kleptoplasts mediate nitrogen acquisition in the sea slug *Elysia viridis* **4**:15–21.

Trench ME, Trench RK, Muscatine L. 1970. Utilization of photosynthetic products of symbiotic chloroplasts in mucus synthesis by *Placobranchus ianthobapsus* (Gould), Opisthobranchia, Sacoglossa **37**:113–117.

Trench RK. 1969. Chloroplasts as functional endosymbionts in the mollusc *Tridachia crispata* (Bërgh), (Opisthobranchia, Sacoglossa). *Nature* **222**:1071.

Wade RM, Sherwood AR. 2017. Molecular determination of kleptoplast origins from the sea slug *Plakobranchus ocellatus* (Sacoglossa, Gastropoda) reveals cryptic bryopsidalean (Chlorophyta) diversity in the Hawaiian Islands. *J Phycol* **53**:467–475.

Wägele H, Deusch O, Händeler K, Martin R, Schmitt V, Christa G, Pinzger B, Gould SB, Dagan T, Klussmann-Kolb A, Martin W. 2011. Transcriptomic evidence that longevity of acquired plastids in the photosynthetic slugs *Elysia timida* and *Plakobranchus ocellatus* does not entail lateral transfer of algal nuclear genes. *Mol Biol Evol* **28**:699–706.

Wägele M, Johnsen G. 2001. Observations on the histology and photosynthetic performance of “solar-powered” opisthobranchs (Mollusca, Gastropoda, Opisthobranchia) containing symbiotic chloroplasts or zooxanthellae. *Org Divers Evol* **1**:193–210.

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol Biol Evol* **35**:543–548.

Watson JD, Crick FH. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**:737–738.

Whelan S, Irisarri I, Burki F. 2018. PREQUAL: detecting non-homologous characters in sets of unaligned homologous sequences. *Bioinformatics* **34**:3929–3930.

Wickner RB, Shewmaker FP, Bateman DA, Edskes HK, Gorkovskiy A, Dayani Y, Bezsonov EE. 2015. Yeast prions:

structure, biology, and prion-handling systems. *Microbiol Mol Biol Rev* **79**:1–17.

Worthley DL, Bardy PG, Mullighan CG. 2005. Mannose-binding lectin: biology and clinical implications. *Intern Med J* **35**:548–555.

Yamamoto S, Hirano YM, Hirano YJ, Trowbridge CD, Akimoto A, Sakai A, Yusa Y. 2013. Effects of photosynthesis on the survival and weight retention of two kleptoplastic sacoglossan opisthobranchs. *J Mar Biol Assoc U K* **93**:209–215.

Yamamoto YY, Yusa Y, Yamamoto S, Hirano Y, Hirano Y, Motomura T, Tanemura T, Obokata J. 2009. Identification of photosynthetic sacoglossans from Japan. *Enco cyto b io sis Cell Re s* **19**:112.

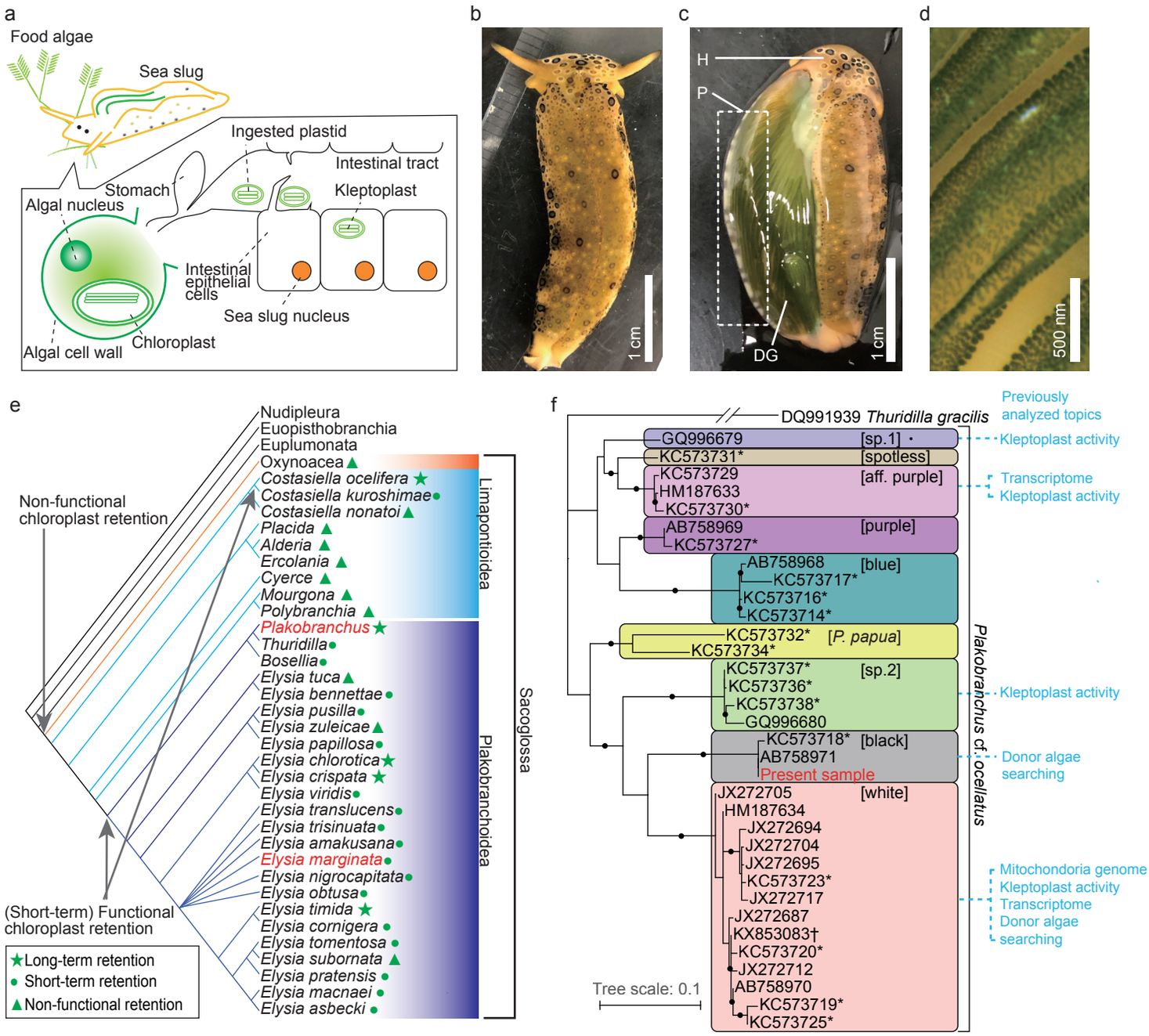
Yamasaki S, Ishikawa E, Sakuma M, Hara H, Ogata K, Saito T. 2008. Mincle is an ITAM-coupled activating receptor that senses damaged cells. *Nat Immunol* **9**:1179–1188.

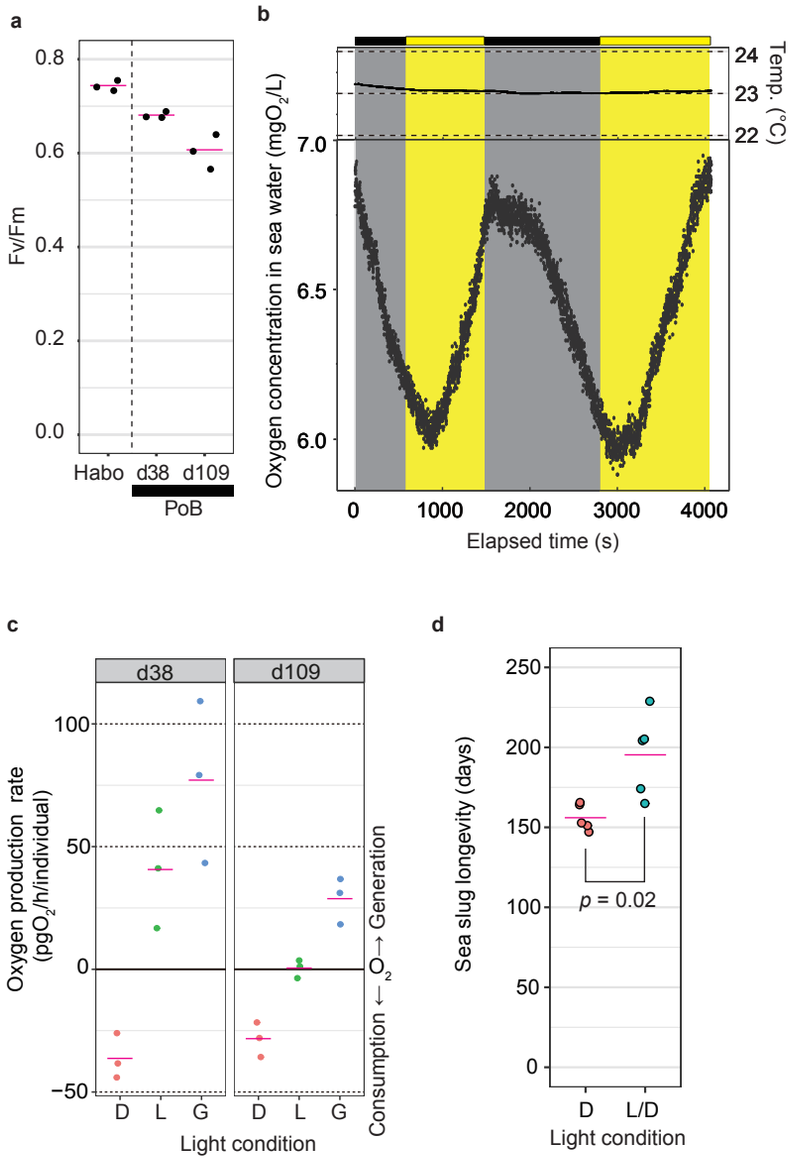
Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* **11**:R14.

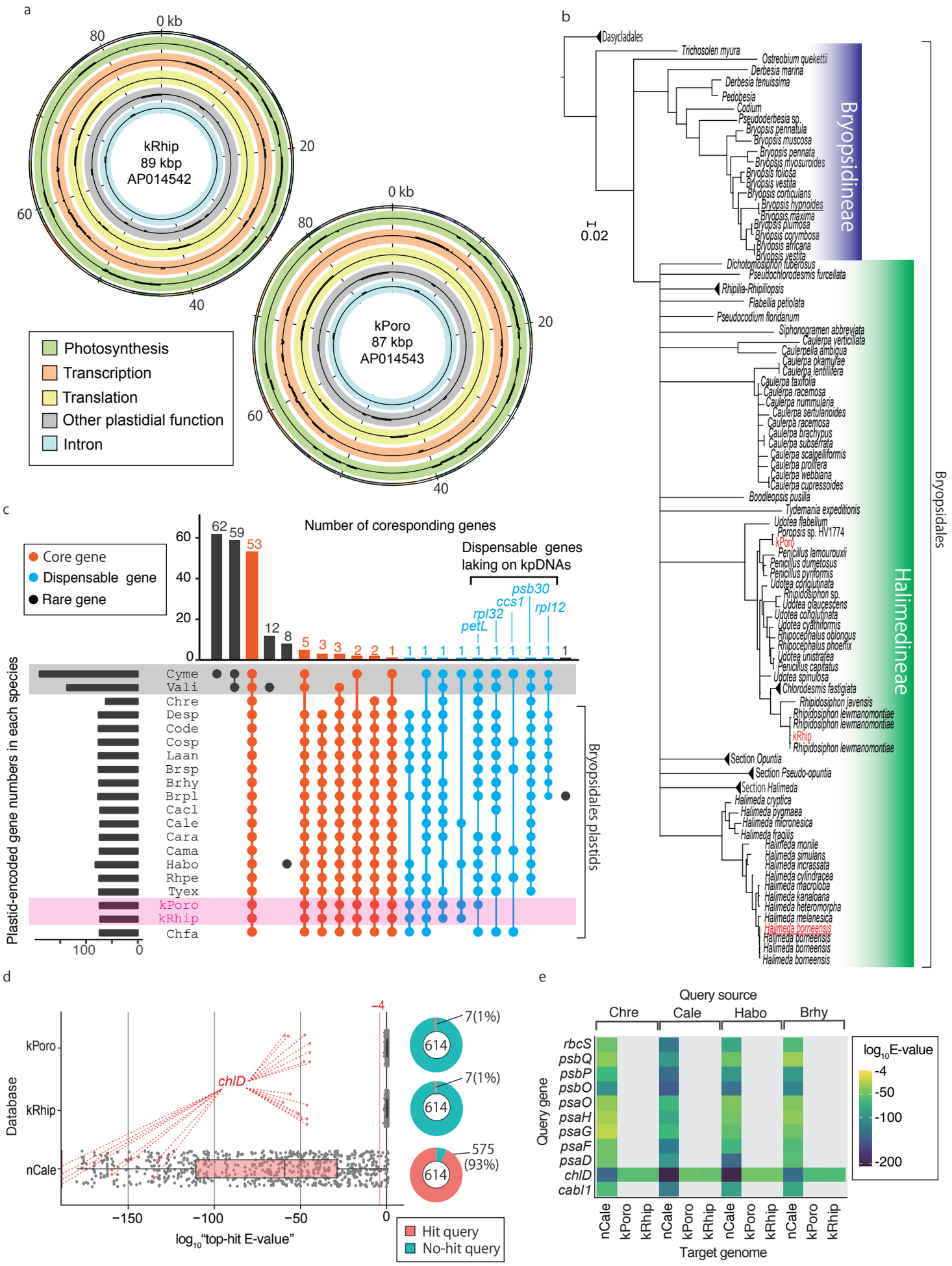
Zapata F, Wilson NG, Howison M, Andrade SCS, Jörger KM, Schrödl M, Goetz FE, Giribet G, Dunn CW. 2014. Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proc Biol Sci* **281**:20141739.

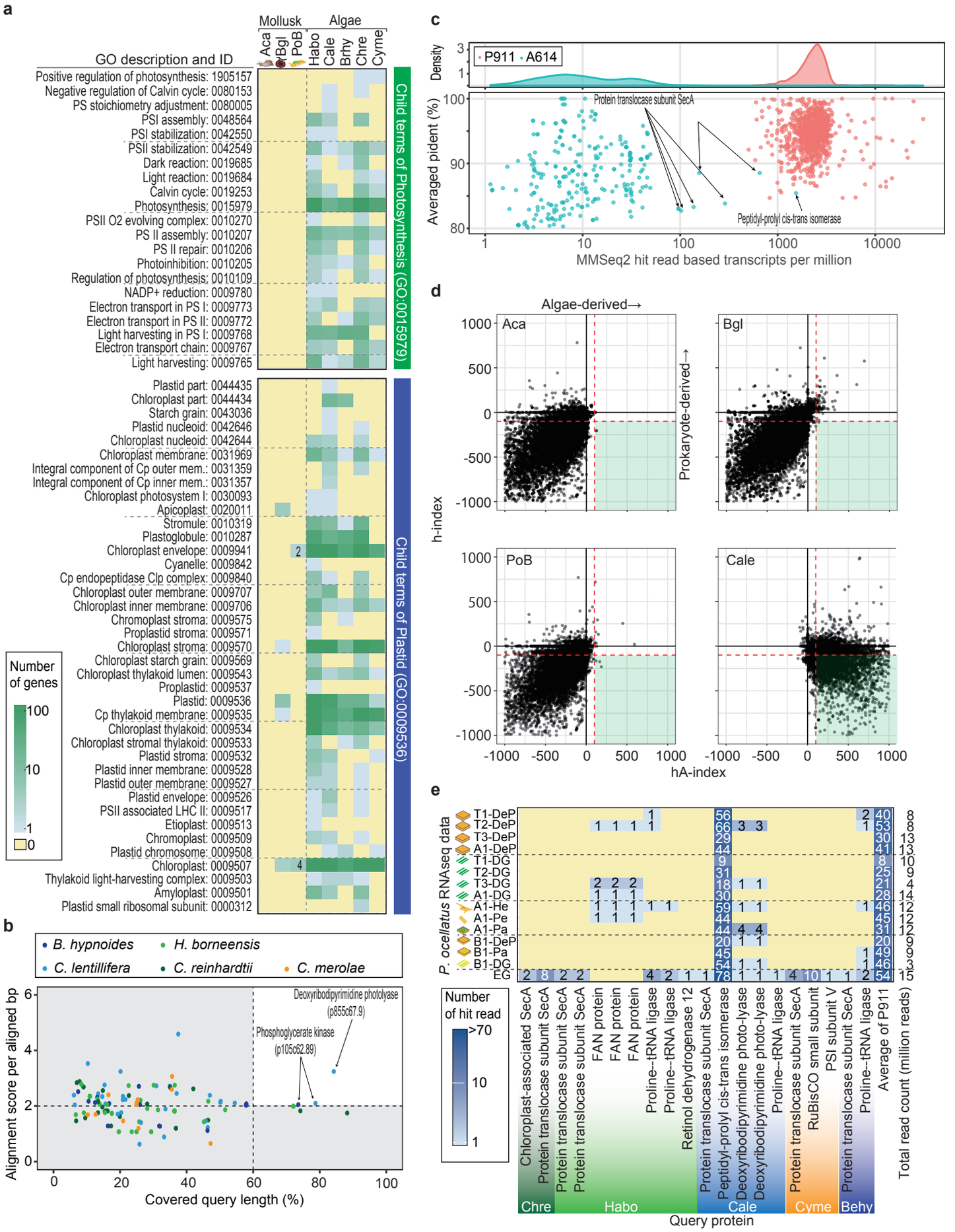
Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using *de Bruijn* graphs. *Genome Res* **18**:821–829.

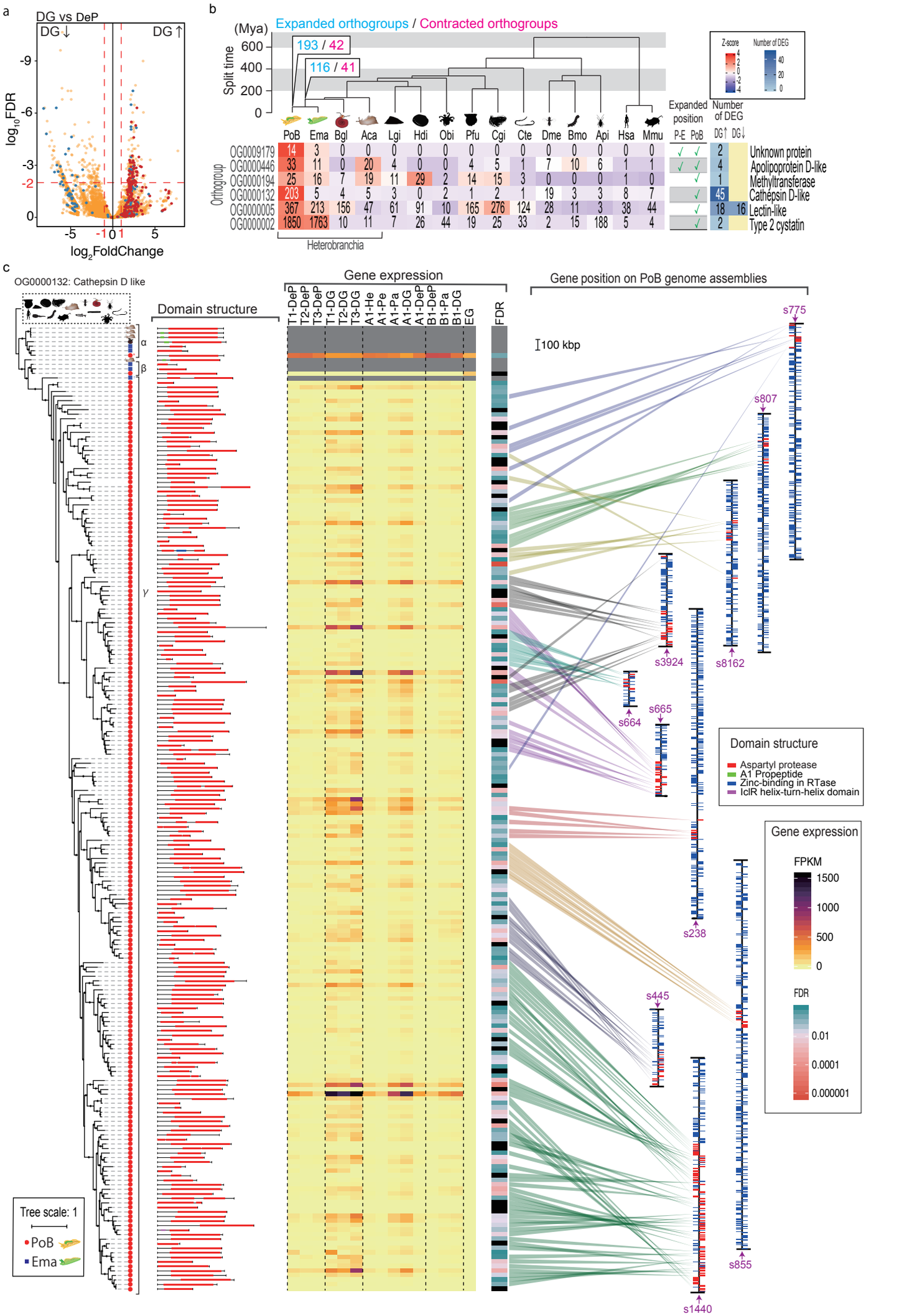
Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**:3406–3415.


a
Food algae
Sea slug
Ingested plastid
Intestinal tract
Stomach
Kleptoplast
Algal nucleus
Intestinal epithelial cells
Sea slug nucleus
Chloroplast
Algal cell wall
b
1 cm
c
H
P
DG
1 cm
d
500 nm
e
Nudipleura
Euopisthobranchia
Euplumonata
Oxynoacea
Costasiella ocelifera
Costasiella kuroshimae
Costasiella nonatoi
Placida
Alderia
Ercolania
Cyerce
Mourgona
Polybranchia
Plakobranchus
Thuridilla
Bosellia
Elysia tuca
Elysia bennettae
Elysia pusilla
Elysia zuleicae
Elysia papillosa
Elysia chlorotica
Elysia crispata
Elysia viridis
Elysia translucens
Elysia trisinuata
Elysia amakusana
Elysia marginata
Elysia nigrocapitata
Elysia obtusa
Elysia timida
Elysia cornigera
Elysia tomentosa
Elysia subornata
Elysia pratensis
Elysia macnaei
Elysia asbecki
Limapontioidea
Plakobranchoidea
Sacoglossa
Non-functional chloroplast retention
(Short-term) Functional chloroplast retention
Long-term retention
Short-term retention
Non-functional retention
f
DQ991939 Thuridilla gracilis
GQ996679 [sp.1]
KC573731* [spotless]
KC573729
HM187633
KC573730*
[aff. purple]
AB758969
KC573727*
[purple]
AB758968
KC573717*
KC573716*
KC573714*
[blue]
KC573732*
KC573734*
[P. papua]
KC573737*
KC573736*
KC573738*
GQ996680
[sp.2]
KC573718*
AB758971
Present sample
[black]
JX272705
HM187634
JX272694
JX272704
JX272695
KC573723*
JX272717
JX272687
KX853083†
KC573720*
JX272712
AB758970
KC573719*
KC573725*
[white]
Plakobranchus cf. ocellatus
Tree scale: 0.1
Previously analyzed topics
Kleptoplast activity
Transcriptome
Kleptoplast activity
Kleptoplast activity
Donor algae searching
Mitochondoria genome
Kleptoplast activity
Transcriptome
Donor algae searching


a
Fv/Fm
0.8
0.6
0.4
0.2
0.0
Habo
d38
d109
PoB
b
Oxygen concentration in sea water (mgO2/L)
7.0
6.5
6.0
Temp. (°C)
24
23
22
0
1000
2000
3000
4000
Elapsed time (s)
c
d38
d109
Oxygen production rate (pgO2/h/individual)
100
50
0
−50
D
L
G
Light condition
Consumption ← O2 → Generation
d
Sea slug longevity (days)
250
200
150
100
50
0
p = 0.02
D
L/D
Light condition


a
0 kb
80
20
60
40
kRhip
89 kbp
AP014542
kPoro
87 kbp
AP014543
Photosynthesis
Transcription
Translation
Other plastidial function
Intron
b
Dasycladales
Trichosolen myura
Ostreobium quekettii
Derbesia marina
Derbesia tenuissima
Pedobesia
Codium
Pseudoderbesia sp.
Bryopsis pennatula
Bryopsis muscosa
Bryopsis pennata
Bryopsis myosuroides
Bryopsis foliosa
Bryopsis vestita
Bryopsis corticulans
Bryopsis hypnoides
Bryopsis maxima
Bryopsis plumosa
Bryopsis corymbosa
Bryopsis africana
Bryopsis vestita
Bryopsidineae
0.02
Dichotomosiphon tuberosus
Pseudochlorodesmis furcellata
Rhipilia-Rhipiliopsis
Flabellia petiolata
Pseudocodium floridanum
Siphonogramen abbreviata
Caulerpa verticillata
Caulerpella ambigua
Caulerpa okamurae
Caulerpa lentillifera
Caulerpa taxifolia
Caulerpa racemosa
Caulerpa nummularia
Caulerpa sertularioides
Caulerpa racemosa
Caulerpa brachypus
Caulerpa subserrata
Caulerpa scalpelliformis
Caulerpa prolifera
Caulerpa webbiana
Caulerpa cupressoides
Boodleopsis pusilla
Tydemania expeditionis
Udotea flabellum
Poroposis sp. HV1774
kPoro
Penicillus lamourouxii
Penicillus dumetosus
Penicillus pyriformis
Udotea corticulinata
Rhipidosiphon sp.
Udotea glaucescens
Udotea conglutinata
Udotea cyathiformis
Rhipocephalus oblongus
Rhipocephalus phoenix
Udotea unistratea
Penicillus capitatus
Udotea spinulosa
Chlorodesmis fastigiata
Rhipidosiphon javensis
Rhipidosiphon lewmanomontiae
Rhipidosiphon lewmanomontiae
kRhip
Rhipidosiphon lewmanomontiae
Section Opuntia
Section Pseudo-opuntia
Section Halimeda
Halimeda cryptica
Halimeda pygmaea
Halimeda micronesica
Halimeda fragilis
Halimeda monile
Halimeda simulans
Halimeda incrassata
Halimeda cylindracea
Halimeda macroloba
Halimeda kanaloana
Halimeda heteromorpha
Halimeda melanesica
Halimeda borneensis
Halimeda borneensis
Halimeda borneensis
Halimeda borneensis
Halimedineae
Bryopsidales
c
Core gene
Dispensable gene
Rare gene
Number of coresponding genes
Dispensable genes laking on kpDNAs
psb30
ccs1
rpl32
rpl12
petL
Plastid-encoded gene numbers in each species
Cyme
Vali
Chre
Desp
Code
Cosp
Laan
Brsp
Brhy
Brpl
Cacl
Cale
Cara
Cama
Habo
Rhpe
Tyex
kPoro
kRhip
Chfa
Bryopsidales plastids
d
Database
kPoro
kRhip
nCale
chlD
−4
−150
−100
−50
0
log10"top-hit E-value"
7(1%)
614
575
(93%)
Hit query
No-hit query
e
Query source
Chre
Cale
Habo
Brhy
Query gene
rbcS
psbQ
psbP
psbO
psaO
psaH
psaG
psaF
psaD
chlD
cabl1
log10E-value
-4
-50
-100
-200
nCale
kPoro
kRhip
Target genome


a
Mollusk
Algae
Aca
Bgl
PoB
Habo
Cale
Brhy
Chre
Cyme
GO description and ID
Positive regulation of photosynthesis: 1905157
Negative regulation of Calvin cycle: 0080153
PS stoichiometry adjustment: 0080005
PSI assembly: 0048564
PSI stabilization: 0042550
PSII stabilization: 0042549
Dark reaction: 0019685
Light reaction: 0019684
Calvin cycle: 0019253
Photosynthesis: 0015979
PSII O2 evolving complex: 0010270
PS II assembly: 0010207
PS II repair: 0010206
Photoinhibition: 0010205
Regulation of photosynthesis: 0010109
NADP+ reduction: 0009780
Electron transport in PS I: 0009773
Electron transport in PS II: 0009772
Light harvesting in PS I: 0009768
Electron transport chain: 0009767
Light harvesting: 0009765
Child terms of Photosynthesis (GO:0015979)
Plastid part: 0044435
Chloroplast part: 0044434
Starch grain: 0043036
Plastid nucleoid: 0042646
Chloroplast nucleoid: 0042644
Chloroplast membrane: 0031969
Integral component of Cp outer mem.: 0031359
Integral component of Cp inner mem.: 0031357
Chloroplast photosystem I: 0030093
Apicoplast: 0020011
Stromule: 0010319
Plastoglobule: 0010287
Chloroplast envelope: 0009941
Cyanelle: 0009842
Cp endopeptidase Clp complex: 0009840
Chloroplast outer membrane: 0009707
Chloroplast inner membrane: 0009706
Chromoplast stroma: 0009575
Proplastid stroma: 0009571
Chloroplast stroma: 0009570
Chloroplast starch grain: 0009569
Chloroplast thylakoid lumen: 0009543
Proplastid: 0009537
Plastid: 0009536
Cp thylakoid membrane: 0009535
Chloroplast thylakoid: 0009534
Chloroplast stromal thylakoid: 0009533
Plastid stroma: 0009532
Plastid inner membrane: 0009528
Plastid outer membrane: 0009527
Plastid envelope: 0009526
PSII associated LHC II: 0009517
Etioplast: 0009513
Chromoplast: 0009509
Plastid chromosome: 0009508
Chloroplast: 0009507
Thylakoid light-harvesting complex: 0009503
Amyloplast: 0009501
Plastid small ribosomal subunit: 0000312
Child terms of Plastid (GO:0009536)
Number of genes
100
10
1
0
b
B. hypnoides
H. borneensis
C. lentillifera
C. reinhardtii
C. merolae
Deoxyribodipyrimidine photolyase (p855c67.9)
Phosphoglycerate kinase (p105c62.89)
Alignment score per aligned bp
Covered query length (%)
c
Density
P911
A614
Protein translocase subunit SecA
Peptidyl-prolyl cis-trans isomerase
Averaged pident (%)
MMSeq2 hit read based transcripts per million
d
Algae-derived→
Prokaryote-derived→
Aca
Bgl
PoB
Cale
h-index
hA-index
e
P. ocellatus RNAseq data
T1-DeP
T2-DeP
T3-DeP
A1-DeP
T1-DG
T2-DG
T3-DG
A1-DG
A1-He
A1-Pe
A1-Pa
B1-DeP
B1-Pa
B1-DG
EG
Number of hit read
>70
10
1
Chloroplast-associated SecA
Protein translocase subunit SecA
Protein translocase subunit SecA
Protein translocase subunit SecA
FAN protein
FAN protein
FAN protein
Proline--tRNA ligase
Proline--tRNA ligase
Retinol dehydrogenase 12
Protein translocase subunit SecA
Peptidyl-prolyl cis-trans isomerase
Deoxyribodipyrimidine photo-lyase
Deoxyribodipyrimidine photo-lyase
Proline--tRNA ligase
Protein translocase subunit SecA
RuBisCO small subunit
PSI subunit V
Protein translocase subunit SecA
Proline--tRNA ligase
Average of P911
Total read count (million reads)
Chre
Habo
Cale
Cyme
Behy
Query protein


a
DG vs DeP
DG↓
DG↑
log10FDR
log2FoldChange
b
(Mya)
Expanded orthogroups / Contracted orthogroups
Split time
193 / 42
116 / 41
PoB Ema Bgl Aca Lgi Hdi Obi Pfu Cgi Cte Dme Bmo Api Hsa Mmu
Orthogroup
OG0009179 14 3 0 0 0 0 0 0 0 0 0 0 0 0 0
OG0000446 33 11 0 20 4 0 0 4 5 1 7 10 6 1 1
OG0000194 25 16 7 19 11 29 2 14 15 3 0 0 0 1 0
OG0000132 203 5 4 5 3 0 2 1 2 1 19 3 3 8 7
OG0000005 367 213 156 47 61 91 10 165 276 124 28 11 3 38 44
OG0000002 1850 1763 10 11 7 26 44 19 25 33 2 15 188 5 4
Heterobranchia
Z-score
Number of DEG
Expanded position
Number of DEG
P-E PoB
DG↑ DG↓
Unknown protein
Apolipoprotein D-like
Methyltransferase
Cathepsin D-like
Lectin-like
Type 2 cystatin
c
OG0000132: Cathepsin D like
Domain structure
Gene expression
T1-DeP T2-DeP T3-DeP T1-DG T2-DG T3-DG A1-He A1-Pe A1-Pa A1-DG A1-DeP B1-DeP B1-Pa B1-DG EG
FDR
Gene position on PoB genome assemblies
100 kbp
s775
s807
s3924
s8162
s664
s665
s238
s445
s1440
s855
Domain structure
Aspartyl protease
A1 Propeptide
Zinc-binding in RTase
IclR helix-turn-helix domain
Gene expression
FPKM
1500
1000
500
0
FDR
0.01
0.0001
0.000001
Tree scale: 1
PoB
Ema