Title

Chloroplast acquisition without the gene transfer in kleptoplastic sea slugs, *Plakobranchus ocellatus*

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: ******@nibb.ac.jp

ST: ******@nibb.ac.jp

TY: ******@jamstec.go.jp

S. Shimamura: ******@jamstec.go.jp

YT: ******@jamstec.go.jp

YN: ******@jamstec.go.jp

AT: ******@nig.ac.jp

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

AA: ******@hiroshima-u.ac.jp

KI: ******@gmail.com

NS: ******@oist.jp

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

MH: ******@nibb.ac.jp

T. Maruyama: ******@nifty.ne.jp

JM: ******@nibb.ac.jp

JO: ******@setsunan.ac.jp S.

Shigenobu: ******@nibb.ac.jp

Abstract

Some sea slugs sequester chloroplasts from algal food in their intestinal cells and photosynthesize for months. This phenomenon, kleptoplasty, poses a question of how the chloroplast retains its activity without the algal nucleus. There have been debates on the horizontal transfer of algal genes to the animal nucleus. To settle the arguments, this study reported the genome of a kleptoplastic sea slug *Plakobranchus ocellatus* and found no evidence of photosynthetic genes encoded on the nucleus. Nevertheless, it was confirmed that light illumination prolongs the life of mollusk under starvation. These data presented 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 analysis showed that genes for proteolysis and immunity undergo gene expansion and are up-regulated in chloroplast-enriched tissue, suggesting that these molluskan genes are involved in the DNA-independent phenotype acquisition.

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 has rekindled the 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 a secondary plastid acquisition scenario in dinoflagellates, (1) a non-phototrophic eukaryote sequesters a unicellular archaeplastid; (2) the endogenous gene transfer to the non-phototrophic eukaryote leads to the 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 algae-derived phenotype acquisition. Some species of sacoglossan sea slugs (Mollusca: Gastropoda: Heterobranchia) can photosynthesize using the chloroplasts of their algal food (Figure 1A and B; 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 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 photosynthesis differs among sacoglossan species (1 to >300 days; Figure 1B; Christa et al., 2015, 2014a, 2014b; Evertsen et al., 2007; Laetz and Wägele, 2017) and 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 maintains photosynthetic activity without algal nuclei (Hirose, 2005; Kawaguti, 1965; Laetz and Wägele, 2018; Martin et al., 2015; Pierce and Curtis, 2012). Because the algal nucleus, rather than the plastids, encodes most photosynthetic proteins, the mechanism to maintain photosynthetic proteins is especially

81 intriguing, given that photosynthetic proteins have a high turnover rate (de Vries and Archibald, 2018; Pelletreau
82 et al., 2011). Previous polymerase chain reaction (PCR)-based studies have suggested the HGT of algal nucleic
83 photosynthetic genes (e.g., *psbO*) to the nucDNA of the sea slug *Elysia chlorotica* (Pierce et al., 1996, 2010,
84 2007, 2003; Rumpho et al., 2008; Schwartz et al., 2014). A genomic study of *E. chlorotica* (N50 = 824 bases)
85 provided no reliable evidence of HGT but predicted that fragmented algal DNA and mRNAs contribute to its
86 kleptoplasty (Bhattacharya et al., 2013). Schwartz et al. (2014) reported *in situ* hybridization-based evidence for
87 HGT and argued that the previous *E. chlorotica* genome might overlook the algae-derived gene. Although an
88 improved genome of *E. chlorotica* (N50 = 442 kb) was published recently, this study made no mention of the
89 presence or absence of algae-derived genes (Cai et al., 2019). The genomic studies of sea slug HGT have been
90 limited to *E. chlorotica*, and the studies have used multiple samples with different genetic backgrounds for
91 genome assembling (Bhattacharya et al., 2013; Cai et al., 2019). The genetic diversity of sequencing data may
92 have inhibited genome assembling. Although transcriptomic analyses of other sea slug species failed to detect
93 HGT (Chan et al., 2018; Wägele et al., 2011), transcriptomic data were insufficient to ascertain genomic gene
94 composition (de Vries et al., 2015; Rauch et al., 2015).

95 Here, the genome sequences of another sacoglossan species, *Plakobranchus ocellatus* (Figure 1C–E), are
96 presented to clarify whether HGT is the primary system underlying kleptoplasty. For more than 70 years,
97 multiple research groups have studied *P. ocellatus* for its long-term (>3 months) ability to retain kleptoplasts
98 (Christa et al., 2013; Evertsen et al., 2007; Greve et al., 2017; Kawaguti, 1941; Trench et al., 1970; Wade and
99 Sherwood, 2017; Wägele et al., 2011). However, a recent phylogenetic analysis showed that *P. ocellatus* is a
100 species complex (a set of closely related species; Figure 1F; Christa et al., 2014c; Krug et al., 2013; Maeda et al.,
101 2012; Meyers-Muñoz et al., 2016; Yamamoto et al., 2013). Therefore, it is useful to revisit previous studies on *P.*
102 *ocellatus*. This study first confirmed the photosynthetic activity and adaptive relevance of kleptoplasty to *P.*
103 *ocellatus* type black [a species confirmed by Krug et al. (2013) via molecular phylogenetics; hereafter "PoB"].
104 The genome sequences of PoB (N50 = 1.45 Mb) and a related species, *Elysia marginata* (N50 = 225 kb), were
105 then constructed. By improving the DNA extraction method, the genome sequences from a single sea slug
106 individual in each species were successfully assembled. The comparative genomic and transcriptomic analyses
107 of these species demonstrate the complete lack of photosynthetic genes in these sea slug genomes and supported
108 an alternative hypothetical kleptoplasty mechanism.

Results

Does the kleptoplast photosynthesis prolong the life of PoB?

To explore the photosynthetic activity of PoB, three photosynthetic indices were measured: photochemical efficiency of kleptoplast photosystem II (PSII), oxygen production rate after starvation for 1 to 3 months, and effect of illumination on PoB longevity. The Fv/Fm value, which reflects the maximum quantum yield of PSII, was 0.68 to 0.69 in the "d38" PoB group (starved for 38 days) and 0.57 to 0.64 in the "d109" group (starved for 109–110 days; Figure 2A). These values were only slightly lower than those of healthy *Halimeda borneensis*, a kleptoplast donor of PoB (Maeda et al. 2012), which showed Fv/Fm values of 0.73 to 0.76. The donor algae of PoB consisted of at least eight green algal species, and they are closely related to *H. borneensis* (Maeda et al. 2012). Although it is not clear whether the Fv/Fm is the same for all donor algae, Fv/Fm is almost identical (~0.83) in healthy terrestrial plants regardless of species (Maxwell and Johnson. 2000). Moreover the values of *H. borneensis* were similar to those of other green algae (e.g., *Chlamydomonas reinhardtii*, Fv/Fm = 0.66–0.75; Bonente et al. 2012). Hence, we assumed that Fv/Fm values are similar among the donor species. The Fv/Fm value suggested that PoB kleptoplasts retain a similar photochemical efficiency of PSII to that of the food algae for more than 3 months.

Based on the measurement of oxygen concentrations in seawater, starved PoB individuals ("d38" and "d109") displayed gross photosynthetic oxygen production (Figure 2B and C). The mock examination without the PoB sample indicated no light-dependent increase in oxygen concentrations, i.e., no detectable microalgal photosynthesis in seawater (Figure 2–Figure Supplement 1). The results demonstrated that PoB kleptoplasts retain photosynthetic activity for more than 3 months, consistent with previous studies on *P.* cf. *ocellatus* (Christa et al., 2014c; Evertsen et al., 2007).

The longevity of starved PoB specimens was then measured under different light conditions. The mean longevity was 156 days under continuous darkness and 195 days under a 12 h light/12 h dark cycle ($p = 0.022$; Figure 2D), indicating that light exposure significantly prolongs PoB longevity. This observation was consistent with that 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 the survival rate (Christa et al., 2014c), our results indicated that this finding does not apply to PoB.

136 Given the three photosynthetic indices' data, we concluded that the increase in PoB survival days was due to
137 photosynthesis. A previous study using the short-term (retention period of 4–8 days) kleptoplastic sea slug,
138 *Elysia atroviridis*, indicated no positive effect of light illumination on their survival rate (Akimoto et al. 2014).
139 These results supported that light exposure does not affect sacoglossan longevity in the absence of kleptoplasts.
140 Although plastid-free PoB may directly indicate that kleptoplasty extends longevity, there is no way to remove
141 kleptoplasts from PoB, except during long-term starvation. PoB feeds on nothing but algae and retains the
142 kleptoplast for 3 to 5 months. Although future research methods may allow for more experimental analyses of
143 kleptoplast functions, the currently most straightforward idea is that kleptoplast photosynthesis increases the
144 starvation resistance of PoB.

145 **Do kleptoplasts encode more photosynthetic genes than general plastid?**

146 To reveal the proteins synthesized from kleptoplast DNA (kpDNA), we sequenced the whole kpDNA from
147 PoB and compared the sequences with algal plastid and nuclear genes. Illumina sequencing provided two types
148 of circular kpDNA and one whole mitochondrial DNA (mtDNA) (Figure 3A; Supplementary File 1; Figure 3–
149 Figure Supplements 1–3). The mtDNA sequence was almost identical to the previously sequenced *P.* cf.
150 *ocellatus* mtDNA (Figure 3–Figure Supplement 3B; Greve et al., 2017). The sequenced kpDNAs corresponded
151 to those of the predominant kleptoplast donors of PoB (Maeda et al., 2012), i.e., *Rhipidosiphon lewmanomontiae*
152 (AP014542; hereafter "kRhip") and *Poropsis* spp. (AP014543; hereafter "kPoro"; Figure 3B; Figure 3–Figure
153 Supplement 4).

154 To determine if kpDNA gene repertoires were similar to green algal chloroplast DNAs (cpDNAs), *H.*
155 *borneensis* cpDNA was sequenced, and 17 whole cpDNA sequences were obtained from public databases
156 (Figure 3C; Figure 3–Figure Supplements 5 and 6). PoB kpDNAs contained all the 59 conserved chloroplastic
157 genes in Bryopsidales algae (e.g., *psbA* and *rpoA*), although they lacked four to five of the dispensable genes
158 (i.e., *petL*, *psb30*, *rpl32*, *rpl12*, and *ccs1*; Figure 3C).

159 To test whether kpDNAs contained no additional photosynthetic genes, a dataset of 614 photosynthetic genes
160 was then used (hereafter the "A614" dataset), which were selected from the algal transcriptomic data and public
161 algal genomic data (Supplementary Files 2 and 3). A tblastn homology search using A614 obtained no reliable
162 hits (*E*-value < 0.0001) against the kpDNA sequences, except for the *chlD* gene, which resembled the kpDNA-

encoded *chlL* gene (Figure 3D and E). A positive control search against an algal nucDNA database (*Caulerpa lentillifera*; https://marinegenomics.oist.jp/umibudo/viewer/info?project_id=55; Arimoto et al., 2019) found reliable matches for 93% (575/614) of the queries (Figure 3D and E), suggesting that the method has high sensitivity. Hence, it was concluded that kpDNAs lack multiple photosynthetic genes (e.g., *psbO*) as general green algal cpDNAs.

Are there horizontally transferred algal genes to the PoB 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. A search for algae-like sequences in gene models, genomic sequences, and pre-assembled short reads were conducted. The genome assembly contained 927.9 Mbp (99.1% of the estimated genome size; 8647 scaffolds; N50 = 1.45 Mbp; 77,230 gene models; Supplementary File 4; Figure 4–Figure Supplements 1–3). Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis using the eukaryota_odb9 dataset showed high coverage (93%) of the eukaryote conserved gene set, indicating that the gene modeling was sufficiently complete to enable HGT searches (Supplementary File 4). The constructed gene models were deposited in INSD (PRJDB3267) and FigShare (DOI: 10.6084/m9.figshare.13042016) with annotation data.

Searches of PoB gene models found no evidence of algae-derived HGT. Simple homology searches (blastp) against the RefSeq database found 127 PoB gene models with top hits to Cyanobacteria or eukaryotic algae. Eighty-two of the 127 top hits encoded functionally unknown protein (annotated as “hypothetical protein” or “uncharacterized protein”; Supplementary File 5). Viridiplantal *Volvox* was the most abundant source of algae of the top-hit sequences (17 genes), and cryptophytan *Guillardia* was the second most abundant source (10 genes). However, 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 File 5). 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 HGT evidence (Supplementary File 5).

Using LCA analysis for all PoB gene models, some genes were predicted to originate from species other than Lophotrochozoa. These genes may be due to HGT, but no photosynthesis-related genes were found from them. MEGAN predicted that 20,189 of the PoB genes originated from Lophotrochozoa and its subtaxa. The prediction also assigned 5550 genes to higher adequate taxa as Lophotrochozoa (e.g., Bilateria). MEGAN

191 assigned the 2964 genes to Opisthokonta other than Lophotrochozoa (e.g., Ecdysozoa) and 312 genes derived
192 from proteobacteria as promising horizontally transferred genes. However, these 3276 genes contained no
193 photosynthetic gene. Many of the PoB genes assigned as promising horizontally transferred genes encoded
194 reverse transcriptase. For the remaining 48,215 genes, MEGAN assigned no species. A homology search against
195 the public database annotated 41,203 of the no-taxon-assigned genes as functionally unknown genes (FigShare;
196 DOI: 10.6084/m9.figshare.13042016). Other no-taxon-assigned 7012 genes were not associated with
197 photosynthesis, except one gene (p288c60.92) annotated to "photosystem I assembly protein ycf3." However,
198 this annotation to p288c60.92 was not reliable because it seems to be derived from incorrect annotations on
199 public databases. Our reannotation via blastp search against the RefSeq database indicated the similarity of the
200 PoB gene (p288c60.92) to "XP_013785360, death-associated protein kinase related-like" of a horseshoe crab.
201 The details of the analyses are summarized in Supplementary File 6. MEGAN-based analysis, hence, indicated
202 that several non-photosynthetic genes might have originated from proteobacteria or other organisms but
203 provided no evidence of algae-derived HGT.

204 In gene function assignment with Gene Ontology (GO) terms, no PoB gene model was annotated as a gene
205 relating "Photosynthesis (GO:0015979)," although the same method found 72 to 253 photosynthesis-related
206 genes in the five referential algal species (Figure 4A). Six PoB genes assigned to the child terms of "Plastid
207 (GO:0009536)" were found (Figure 4A; Supplementary File 7). However, an ortholog search with animal and
208 algal genes did not support these six genes' algal origin (Figure 4–Figure Supplement 5). It was considered that
209 the sequence conservation beyond the kingdom caused these pseudo-positives in the GO assignment. The
210 function of these six genes had no relationships to photosynthesis (Supplementary File 7); six genes relate
211 proteasome (p2972c65.3), arginine kinase (p197c68.18), DNA-binding response regulator (p234c64.89),
212 chromatin structure (p466c59.83), mitochondrial inner membrane translocase (p503c65.126), and functionally
213 unknown protein (p45387c41.1).

214 To confirm that the gene modeling did not overlook a photosynthetic gene, the A614 dataset against the PoB
215 and *C. lentillifera* genome sequences was directly searched using tblastn and Exonerate software (Slater and
216 Birney, 2005). Against the *C. lentillifera* genome (positive control), 455 (tblastn) and 450 (Exonerate) hits were
217 found; however, using the same parameters, only 1 (tblastn) and 2 (Exonerate) hits against the PoB genome
218 were detected (Figure 4B; Figure 4–Figure Supplement 5). The three PoB loci contain the genes encoding

219 serine/threonine-protein kinase LATS1 (p258757c71.5, Figure 4–Figure Supplement 5), deoxyribodipyrimidine
220 photolyase (p855c67.9, Figure 4B), and phosphoglycerate kinase (p105c62.89, Figure 4B). A phylogenetic
221 analysis with homologous genes showed the monophyletic relationships of these three genes with molluskan
222 homologs (Figure 4–Figure Supplement 6–8). This result indicated that the three PoB loci contain molluskan
223 genes rather than algae-derived genes. It was thus concluded that the PoB genome assembly contains no algae-
224 derived photosynthetic genes.

225 To examine if the genome assembly failed to construct algae-derived regions in the PoB genome, a search
226 for reads resembling photosynthetic genes among the pre-assembled Illumina data was performed. From 1065
227 million pre-assembled reads, 1698 reads showed similarity against 261 of the A614 dataset queries based on
228 MMseq2 searches. After normalization by query length, the number of matching reads against algal queries was
229 about 100 times lower than that against PoB single-copy genes: the normalized read count (mean ± SD) was 25
230 ± 105 for the A614 dataset and 2601 ± 2173 for the 905 PoB single-copy genes ($p < 0.0001$, Welch's two-
231 sample *t*-test; Figure 4C). This massive difference indicated that most algae-like reads come from contaminating
232 microalgae rather than PoB nucDNA. For six algal queries (e.g., peptidyl-prolyl cis-trans isomerase gene), the
233 number of matching reads was comparable between PoB and the A614 dataset (Figure 4C; the normalized read
234 count was higher than 100). However, the presence of homologous genes on the PoB genome (e.g., p310c70.15)
235 suggests that the reads were not derived from an algae-derived region but rather resemble molluskan genes. For
236 example, a simple alignment showed that the *C. lentillifera*-derived g566.t1 gene (encoding peptidyl-prolyl cis-
237 trans isomerase) was partially similar to the PoB p310c70.15 gene, and 76% (733/970) of reads that hit against
238 g566.t1 were also hitting against p310c70.15 under the same MMseq2 parameter (Figure 4–Figure Supplement
239 9). Hence, it was considered that no loss of algal-derived HGT regions occurred in the assembly process.

240 Changing the focus to HGT of non-photosynthetic algal genes, the indices for prokaryote-derived HGT (h-
241 index; Boschetti et al., 2012) and algae-derived HGT (hA-index; see Materials and Methods) were calculated for
242 PoB and two non-kleptoplastic gastropods (negative controls; Supplementary File 8). Three PoB gene models as
243 potential algae-derived genes were detected (Figure 4D); however, two of these encoded a transposon-related
244 protein, and the other encoded an ankyrin repeat protein with a conserved sequence with an animal ortholog
245 (Figure 4–Source Data 1). Furthermore, non-kleptoplastic gastropods (e.g., *Aplysia californica*) had similar

246 numbers of probable HGT genes (Figure 4D). Taking these results together, it was concluded that there is no
247 evidence of algae-derived HGT in the PoB genomic data.

248 The presence of algae-like RNA in PoB was then examined, as a previous study of another sea slug species *E.*
249 *chlorotica* hypothesized that algae-derived RNA contributes to kleptoplasty (Bhattacharya et al., 2013). This
250 previous study used short-read-based blast searches and reverse transcription PCR analyses to detected algal
251 mRNA (e.g., *psbO* mRNA) in multiple adult *E. chlorotica* specimens (no tissue information was provided;
252 Bhattacharya et al., 2013). To analyze the algal RNA distribution in PoB, 15 RNA-Seq libraries were
253 constructed from six tissue types [digestive gland (DG), parapodium, DG-exenterated parapodium (DeP), egg,
254 head, and pericardium; Figure 4–Figure Supplement 10], and MMseq2 searches were conducted (Figure 4C).
255 Although almost all (594/614) of the A614 dataset queries matched no reads, 19 queries matched 1 to 10 reads
256 and the *C. lentillifera*-derived g566.t1 query matched more than 10 reads (Figure 4E). This high hit rate for
257 g566.t1 (peptidyl-prolyl cis-trans isomerase in Figure 4E), however, is due to its high sequence similarity with
258 PoB ortholog, p310c70.15, as mentioned above. A previous anatomical study showed the kleptoplast density in
259 various tissues to be DG > parapodium > DeP = head = pericardium >>> egg (Hirose, 2005). Therefore, the
260 amount of algae-like RNA reads did not correlate with the kleptoplast richness among the tissues. Enrichment of
261 algae-like RNA was only found in the egg (Figure 4E). The PoB egg is considered a kleptoplast-free stage and is
262 covered by a mucous jelly, which potentially contains environmental microorganisms (Figure 4–Figure
263 Supplement 10). Hence, it was presumed that these RNA fragments are not derived from kleptoplasts but from
264 contaminating microalgae. The searches for algae-like RNA in PoB found no credible evidence of algae-derived
265 RNA transfer and no correlation between algal RNA and kleptoplasty.

266 **What kind PoB genes are reasonable candidates for kleptoplasty-related genes?**

267 Because the PoB genome was found to be free of algae-derived genes, it was considered that a
268 neofunctionalized molluskan gene might contribute to kleptoplasty. To find candidate kleptoplasty-related
269 molluskan (KRM) genes, we assessed genes that were up-regulated in DG (the primary kleptoplasty location)
270 versus DeP in the RNA-Seq data described above. As a result, 162 DG-up-regulated genes were found [false
271 discovery rate (FDR) < 0.01, triplicate samples; Figure 5A; Figure 5–Figure Supplements 1 and 2;
272 Supplementary File 9]. By conducting GO analysis, the functions of 93 DG-up-regulated genes were identified.
273 These up-regulated genes are enriched for genes involved in proteolysis (GO terms: "Proteolysis," "Aspartic-

274 type endopeptidase activity," "Cysteine-type endopeptidase inhibitor activity," and "Anatomical structure
275 development"), carbohydrate metabolism ("Carbohydrate metabolic process," "One-carbon metabolic process,"
276 "Cation binding," and "Regulation of pH"), and immunity ("Defense response"; Supplementary File 10).
277 Manual annotation identified the function of 42 of the remaining DG-up-regulated genes. Many of these are also
278 related to proteolysis and immunity: three genes relate to proteolysis (i.e., genes encoding interferon-γ-inducible
279 lysosomal thiol reductase, replicase polyprotein 1a, and phosphatidylethanolamine-binding protein), and 21
280 genes contribute to natural immunity (i.e., genes encoding lectin, blood cell aggregation factor, and
281 MAC/perforin domain-containing protein; Figure 5–Figure Supplement 1 and 2). Manual annotation also found
282 four genes encoding apolipoprotein D, which promotes resistance to oxidative stress (Charron et al., 2008), and
283 three genes involved in nutrition metabolism (i.e., genes encoding betaine-homocysteine S-methyltransferase 1-
284 like protein, intestinal fatty acid-binding protein, and cell surface hyaluronidase). Because the analyzed slugs
285 were starved for 1 month, it was considered that DG-up-regulated genes contribute to kleptoplasty rather than
286 digestion.

287 A comparative genomic analysis was then conducted to find orthogroups that expand or contract in size along
288 the metazoan lineage to PoB. The phylogenomic analysis showed that 6 of the 193 orthogroups that underwent
289 gene expansion in this lineage contained DG-up-regulated genes (Figure 5B; Figure 5–Figure Supplement 3–5;
290 Supplementary File 11). This result supported the notion that these DG-up-regulated genes play a role in
291 kleptoplasty. The most distinctive orthogroup in the six groups was OG0000132, which contained 203 cathepsin
292 D-like genes of PoB. Forty-five of the 203 genes were DG-up-regulated, and Fisher's exact test supported the
293 significant enrichment of DG-up-regulated genes ($p < 0.0001$; Supplementary File 11). Other heterobranchian
294 mollusks only had four to five genes belonging to OG0000132 (Figure 5B). These gene duplications in PoB
295 might reduce selection pressure to maintain function via redundancy and promote new function acquirement of
296 the paralogs, as occurs in the well-known neofunctionalization scenario (Conrad and Antonarakis, 2007).
297 Significant enrichment of DG-up-regulated genes was also detected in OG0000005 (18 genes) and OG0000446
298 (4 genes; both $p < 0.0001$; Figure 5B; Supplementary File 11), which contain lectin-like and apolipoprotein D-
299 like genes, respectively. DG-up-regulated genes were also detected in OG0000002, OG0009179, and
300 OG0000194. However, no significant enrichment of these genes was found in these orthogroups ($p > 0.05$).
301 OG0000002 contains DG-up-regulated gene for type 2 cystatin and various genes with reverse transcriptase

302 domains. This result suggested that the reverse transcriptase domain clustered the various genes as one
303 orthogroup, and the gene number expansion was due to the high self-duplication activity of the retrotransposon
304 in PoB. OG0009179 and OG0000194 contain DG-up-regulated genes of unknown function. From the above
305 results, 67 genes were finally selected as promising targets of study for PoB kleptoplasty: 45 genes for cathepsin
306 D-like proteins, 18 genes for lectin-like proteins, and 4 genes for apolipoprotein D-like protein (Supplementary
307 File 9).

308 **Evolution of candidate KRM genes**

309 For a more detailed analysis of the evolution of KRM genes in sacoglossan lineages, a new draft genome
310 sequence of another sacoglossan sea slug *E. marginata* (previously included in *E. ornata*; Krug et al., 2013) was
311 constructed. PoB and *E. marginata* belong to the same family (Plakobranchidae; Figure 1B). Both species
312 sequester plastids from Bryopsidales algae; however, the kleptoplast retention time is limited to a few days in *E.*
313 *marginata* (Yamamoto et al., 2009). These features suggested that their common ancestor obtained a mechanism
314 to sequester algal plastids, but *E. marginata* did not develop a system for their long-term retention. Hence, it was
315 considered that comparing gene expansion in these species would clarify the genetic basis of plastid
316 sequestration and long-term retention.

317 Using the same methods as described for PoB, one complete circular kpDNA, one complete mtDNA, and
318 790.3 Mbp nucDNA (87.8% of the estimated genome size; 14,149 scaffolds; N50 = 0.23 Mbp; 70,752 genes;
319 Supplementary Files 1, 4, 13, and 14) were sequenced for *E. marginata*. The constructed gene models covered
320 89.5% of the BUSCO eukaryota_odb9 gene set (Supplementary File 4). No credible photosynthetic gene was
321 detected from *E. marginata* nucDNA (annotation data; DOI: 10.6084/m9.figshare.13042016).

322 The evolution of representative candidate PoB KRM genes (i.e., cathepsin D-like, apolipoprotein D-like, and
323 lectin-like genes) was then phylogenetically analyzed. In cathepsin D-like genes, sacoglossan (PoB and *E.*
324 *marginata*) genes formed a specific subgroup in the OG0000132 phylogenetic tree (Figure 6A), and gene
325 duplication in OG0000132 seemed to be accelerated along the PoB lineage (203 genes in PoB versus 5 genes in
326 *E. marginata*; Figure 5B). All sacoglossan cathepsin D-like genes belonged to a clade with several other
327 heterobranchian homologs; this clade contained three subclades (α, β, and γ in Figure 6B). The basal α-clade
328 contained three *A. californica* genes, one *Biomphalaria glabrata* gene, and three sacoglossan genes. The β- and
329 γ-clades contained sacoglossan genes only and an *A. californica* gene located at the basal position of the β- and

330 γ-clades. Almost all duplicated PoB genes (201/203) belonged to the γ-clade, including one *E. marginata* gene
331 (e8012c40.2). These phylogenetic relationships suggested that the γ-clade has undergone dozens of gene
332 duplication events in the PoB lineage. Interestingly, all DG-up-regulated differentially expressed genes (DEGs)
333 belonged to the γ-clade, and the PoB paralogs that belong to the α- and β-clades showed different expression
334 patterns from the γ-clade paralogs; the gene p609c69.52 (α-clade) was ubiquitously expressed in the examined
335 tissues, and p374c67.53 (β-clade) was expressed only in the egg (Figure 6B). The mammalian genes encoding
336 cathepsin D and its analog (cathepsin E) were ubiquitously expressed on various tissue types (Benes et al., 2008).
337 Therefore, we considered that (1) the ubiquitously expressed p609c69.52 gene in α-clade is a functional ortholog
338 of the mammalian cathepsin D gene, (2) the p374c67.53 gene in β-clade relates sea slug embryo development,
339 and (3) the γ-clade genes have been acquired with the development of plastid sequestration. To test the positive
340 or relaxed selection on the duplication event, the ratio of substitution rates at non-synonymous (dN) and
341 synonymous (dS) sites (ω = dN/dS) was used. Longly alignable heterobranchian genes from OG0000132 were
342 used and statistically tested using CodeML and RELAX software (Supplementary File 12). The CodeML
343 analysis (branch-site model) determined no significant positive selection on the basal node of the γ-clade ($p >$
344 0.05, χ^2 test), but RELAX suggested that the γ-clade evolved under relaxed selection relative to the rest
345 branches ($K = 0.44$, $p < 0.001$). The cathepsin D-like genes formed multiple tandem repeat structures in the PoB
346 genome, although other Heterobranchia had no tandem repeat (Figure 6B; Figure 6–Figure Supplement 1 and 2).
347 In *E. marginata*, the gene e8012c40.2 located at the basal position of the γ-clade had no repeat structure,
348 although two genes from the α-clade (e4981c37.5 and e4981c37.4) and two from the β-clade (e2244c39.16 and
349 e2244c39.20) made two tandem repeats (Figure 6–Figure Supplement 1 and 2). The tree and repeat structure
350 suggested that the γ-clade was separated from the β-clade as a single-copy gene in the common ancestor of PoB
351 and *E. marginata* and duplicated in the PoB lineage (Figure 6B). The revealed genomic structure indicated that
352 the duplicates were not due to whole-genome duplication but rather a combination of several subgenomic
353 duplication events: the dispersed duplicates are likely due to replicative transposition by a transposable element,
354 and the tandem repeats are likely due to a local event (e.g., unequal crossover).

355 The gene duplication in OG0000446 seems to have happened in the PoB lineage and at the node between
356 PoB and *E. marginata* (Figure 5B; Figure 6–Figure Supplement 3 and 4). The sacoglossan genes were
357 duplicated in a monophyletic clade (clade I) only, and all DG-up-regulated DEGs were contained in clade I.

358 CodeML found no significant positive selection on the basal node of clade Ia ($p > 0.05$), but RELAX suggested
359 evolution under relaxed selection relative to the rest branches ($K = 0.69$; $p = 0.003$; Supplementary File 12). On
360 the other subclades, no significant positive selection nor relaxed selection was found (Supplementary File 12).
361 All DG-up-regulated DEGs were contained in clade I. It was hypothesized that duplication on the common
362 lineage relates to plastid sequestration, and the PoB-specific duplication events contribute to long-term
363 kleptoplasty.

364 In OG0000005 (lectin-like gene group), the gene counts were comparable between PoB (367 genes) and *E.*
365 *marginata* (213 genes; Figure 5B); however, the phylogenetic tree suggests that this similarity is due to different
366 gene duplication events in each species. One-by-one orthologous gene pairs were rare between PoB and *E.*
367 *marginata* (only 14 were detected), and many of the other homologs formed species-specific subclades (Figure
368 6–Figure Supplement 5). Lectins are carbohydrate-binding molecules that mediate the attachment of bacteria
369 and viruses to their intended targets (Lis and Sharon, 1998; Yamasaki et al., 2008). The observed gene
370 expansions of lectin-like genes in each of these sea slugs may widen their targets and complicate the natural
371 immune system to distinguish the kleptoplast from other antigens. Four clades having expanded homologs on
372 the sacoglossan lineage were detected (clades A–D in Figure 6–Figure Supplement 5). These clades contained
373 all 34 DEGs between DG and DeP tissues. The DG-up-regulated and DG-down-regulated genes are located on
374 different clades, except for one DG-up-regulated gene (p3334c67.98) in clade A. These results indicated that the
375 determined DEGs have different expression patterns depending on the clade.

376

377 # Discussion

378 Here, it was demonstrated that (1) kleptoplast photosynthesis extends the lifetime of PoB under starvation,
379 (2) the PoB genome encodes no algal nucleus-derived genes, and (3) PoB individuals up-regulate genes for
380 carbohydrate metabolism, proteolysis, and immune response in their DG. Candidate KRM genes, i.e.,
381 apolipoprotein D-like, cathepsin D-like, and lectin-like genes, were also identified using a combination of
382 genomic and transcriptomic analysis. Data present a paradigm of kleptoplasty in which PoB obtains the adaptive
383 photosynthesis trait by a DNA-independent phenotype acquisition.

384 The PoB genomic sequence clarified the gene repertory of its kpDNA, mtDNA, and nucDNA. The whole
385 kpDNA sequence indicated that PoB kleptoplasts could produce some proteins involved in photosynthesis (e.g.,
386 PsbA, a core protein in PSII; Figure 3) if gene expression machinery is sufficiently active, as reported in *E.*
387 *chlorotica* (Green et al., 2000; Pierce et al., 2007). The absence of core photosynthetic genes in the PoB genome
388 was then demonstrated. For instance, no genes encoding PsaF of photosystem I, PsbO of PSII, or RbcS of
389 Rubisco in kpDNAs, mtDNA, or nucDNA (Figures 3 and 4) were detected despite the queries (e.g., A612
390 dataset; Supplementary File 3) containing multiple algal orthologs of these genes; these gene products are
391 essential for photosynthesis in various plants and algae and located on their nuclear genome, not plastid genome
392 (Farah et al., 1995; Izumi et al., 2012; Pigolev et al., 2009). This result means that PoB can perform
393 photosynthesis (Figure 2) without *de novo* synthesis of these gene products. The absence of algae-derived HGT
394 is consistent with previous transcriptomic analyses of *P.* cf. *ocellatus* (Wägele et al., 2011) and other
395 sacoglossan species (de Vries et al., 2015; Han et al., 2015; Wägele et al., 2011). A previous genome study of *E.*
396 *chlorotica* predicted that fragmented algal DNA and mRNAs contribute to its kleptoplasty (Bhattacharya et al.,
397 2013), and fluorescence *in situ* hybridization study detected algal gene signals on *E. chlorotica* chromosomes
398 (Schwartz et al., 2014). Although our manual check of gene annotation found no photosynthesis-related gene
399 from a more complete version of the *E. chlorotica* nuclear genome sequence, Cai et al. (2019) provided no
400 discussion about HGT. After all, the PoB results proposed that algal DNA and/or RNA is not an absolute
401 requirement for kleptoplast photosynthesis.

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

404 the photosynthetic proteins in algal cells and/or *in vitro* (Roberts et al., 2013), previous studies have discussed
405 the elongation of algal protein lifespans via protective sea slug proteins as an alternative hypothesis to HGT (de
406 Vries and Archibald, 2018; Serôdio et al., 2014). This study shows three types of potential KRM genes:
407 apolipoprotein D-like, cathepsin D-like, and lectin-like genes. Some of these KRM genes may encode proteins
408 that protect algae-derived proteins. Although further studies are needed to elucidate their functional roles in
409 detail, the obtained results are consistent with HGT's alternative hypothesis, algal protein retention.

410 Previous RNA-Seq studies of *Elysia timida* and *E. chlorotica* found up-regulation of superoxide dismutase
411 (SOD) genes in response to photostress and postulated that SOD protects algal proteins in the kleptoplasts from
412 oxidative damage (Chan et al., 2018; de Vries et al., 2015). No significant up-regulation of the SOD gene in PoB
413 DG was found (Figure 5–Figure Supplement 1). However, there was up-regulation of apolipoprotein D-like
414 genes and the expansion of these genes in PoB lineages (Figure 5). Apolipoprotein D, a lipid antioxidant,
415 confers resistance to oxidative stress in higher plants and animal brains (Bishop et al., 2010; Charron et al.,
416 2008). Ortholog analysis of the new *E. chlorotica* data found no gene number expansion of these homologs in *E.*
417 *chlorotica*, i.e., only three apolipoprotein D-like homologs (Supplementary File 15). Although the retention
418 process details may differ among species and abiotic conditions, it is attractive to speculate that oxidative stress
419 resilience is of major importance for kleptoplasty in multiple sacoglossan species.

420 Previous kleptoplasty studies have not focused on cathepsin D-like and lectin-like genes. The involvement of
421 the mechanisms of these proteins in kleptoplasty remains to be solved. Cathepsin D degrades intracellular
422 proteins and contributes to the degradation of damaged mitochondria (Benes et al., 2008). In general, damaged
423 photosynthetic proteins generate abundant ROS, promoting further protein damage. In PoB, cathepsin D-like
424 protease may degrade damaged proteins and prevent further protein inactivation. Lectin is involved in self-
425 recognition in natural immunity (Geijtenbeek and Gringhuis, 2009; Worthley et al., 2005). The diverse lectins
426 expressed in PoB DG tissue may bind the antigens of algae-derived molecules, mediate the detection of non-
427 self-proteins and/or saccharides, and lead to the selective degradation and retention of algae-derived proteins and
428 organelles.

429 These genomic data indicate that a proteomic analysis of kleptoplasts is warranted. A previous isotopic study
430 indicated that *E. chlorotica* kleptoplasts contain function-unknown proteins transported from the sea slug
431 cytoplasm (Pierce et al., 1996). Although several algal photosynthetic proteins have been immunoassayed in

432 kleptoplasts, animal nuclear-encoded proteins have been out of the target (Green et al., 2000; Pierce et al., 1996).
433 The *in silico* study found no typical chloroplast localization signal in PoB KRM genes (Figure 6–Figure
434 Supplement 1 and 3). However, the genomic data will help identify kleptoplast-localized sea slug proteins by
435 peptide mass fingerprinting.

436 This study provides the first genomic evidence of photosynthesis acquisition without horizontal DNA or
437 RNA transfer. Previous studies have demonstrated that DNA is the core material for heredity (Hershey and
438 Chase, 1952; Watson and Crick, 1953) and assumed that horizontal DNA transfer causes cross-species
439 phenotype acquisition (Acuña et al., 2012; Anderson, 1970; Boto Luis, 2014; Dehal et al., 2002). However, our
440 studies indicate that PoB gains adaptive photosynthetic activity without acquiring any of the many algal nucleic
441 genes involved in photosynthesis. This is evidence of a phenotype (and organelle) acquisition, although
442 temporally, from the donor without DNA or RNA transfer. Recent studies of shipworms (wood-feeding
443 mollusks in the family Teredinidae) showed that they utilize several symbiont-derived proteins for their food
444 digestion (O'Connor et al., 2014). Golden sweeper fish (*Parapriacanthus ransonneti*) gain the luciferase for
445 their bioluminescence from ostracod prey, suggesting phenotype acquisition via the sequestration of a non-self-
446 protein (kleptoprotein; Bessho-Uehara et al., 2020). However, these two examples of DNA/RNA-independent
447 phenotype acquisition are limited to the transfer of simple phenotypes that depend on just a few enzymes. In
448 contrast, the well-known complexity of photosynthesis suggests that sea slug kleptoplasty depends on
449 DNA/RNA-independent phenotype acquisition of complex pathways requiring multiple enzymes (e.g., entire
450 photosystems and the Calvin cycle). It is attractive to speculate that other symbiont-derived organelles (e.g.,
451 mitochondria), and obligate endosymbiotic bacteria and protozoan kleptoplasts (e.g., in *Dinophysis acuminata*;
452 Hackett et al., 2003) can obtain traits derived from another species via a DNA-independent system. Although
453 several organisms have multiple HGT-derived functional genes, it is still unclear how the organism
454 evolutionarily obtained the appropriate expression control system of the non-self-gene (Sasakura et al., 2016).
455 These PoB data suggested that the transfer of complex adaptive phenotypes sometimes precede the gene transfer
456 from the donor species, having the potential to explain the process of cross-species development of complex
457 phenotypes. Some organisms might evolutionarily obtain the HGT-derived genes and appropriate control system
458 of the gene expression after the transfer of phenotype. Our finding of DNA-independent complex phenotype
459 acquisition may open new viewpoints on cross-species evolutionary interaction.

460

Materials and Methods

Samplings of sea slugs and algae

Samples were collected from southwestern Japan. Specifically, PoB and *H. borneensis* were collected on Okinawa Island shores, and *E. marginata* was collected from Kinkowan Bay. Regarding *B. hypnoides*, a cultivated thallus was initially collected from Kinkowan Bay and used in the authors' laboratory for several years. The collected samples of PoB and *E. marginata* in seawater were transported to the laboratories at the National Institute for Basic Biology (NIBB) and Kyoto Prefectural University, respectively, 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 light/12 h 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, Lincoln, NE, USA). The seawater was not changed during the incubation period, except to adjust the salinity using distilled water. Photosynthetic activity indices (oxygen generation rate and PAM fluorometry) were measured using oxygen sensor spots (Witrox 4; Loligo Systems, Tjele, Denmark) and PAM-2500 (Walz, Effeltrich, Germany), respectively (Figure 2–Figure Supplement 1B–F). *H. borneensis*, which was easily collected and kept in the laboratory, was used as a reference for comparison in the PAM analysis. This species is a donor alga of PoB kleptoplasts and closely related species with the major donors of PoB, *R. lewmanomontiae* and *Poropsis* spp. These two major donors firmly adhered to the rock, were tangled up in other algal species, and were difficult to be collected.

The oxygen sensor spots were affixed to the inside of a glass respirometry chamber. Before performing the 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 (23°C–24°C) using a water

487 jacket attached to a temperature-controlled water flow. The Witrox temperature probe for calibration was
488 immersed in the water jacket (Figure 2–Figure Supplement 1B–F).

489 The oxygen concentration was measured sequentially under changing light conditions. The percent O_2
490 saturation was monitored continuously and recorded using AutoResp software (Loligo Systems) for 10 min after
491 the respirometry chamber acclimation period (10 min). The oxygen consumption rate by respiration was
492 measured under dark conditions. Next, the respirometry chambers were exposed to the red LED light (800 μmol
493 photons m^{-2} s^{-1}) to measure the change in the oxygen concentration under light conditions due to the balance
494 between the photosynthesis and respiration rates. The chambers were illuminated from the sides because a
495 mounted LED light increased the oxygen meter's noise. The light direction to the sea slug was inconsistent
496 because the slug kept moving around the chamber during the measurement period. In sea slugs, the rate of
497 photosynthesis was unaffected by the direction of illumination because the rate of O_2 generation rate under a
498 particular constant illumination did not change regardless of the sea slug's position in the chamber. The percent
499 O_2 saturation was measured for 10 min. One blank (i.e., without sea slug) condition was run as a negative
500 control to account for background biological activity in seawater. AutoResp software was used to convert the
501 percent saturation to an oxygen concentration ([O_2], mg O_2 l^{-1}) based on the rate of change in the percent O_2
502 saturation, water temperature, and barometric pressure (fixed at 1013 Pa). A regression analysis was performed
503 using the "lm" function in R version 3.5.2 (tidyverse 1.2.1 package) to calculate the changing oxygen
504 concentration rate under dark and light conditions. A gross oxygen generation rate was then obtained by
505 photosynthesis (oL + oD = oG; oL, oxygen production rate under light conditions; oD, oxygen consumption rate
506 under dark conditions; oG, gross oxygen generation by photosynthesis).

507 During PAM fluorometry analysis, each sea slug was caged in a single well of a 12-well cell culture plate
508 (Corning, Corning, NY, USA) after adaptation to the dark for 15 min. To ensure reproducibility, the sea slug
509 was caged upside down (i.e., the ventral surface was brought to the upside), the animal was softly squeezed with
510 a plastic sponge, and the PAM light probe was connected to the plate from underneath the well. Consequently,
511 the samples could not move during the measurement, and the PAM light probe always measured the
512 fluorescence of the dorsal surface. The maximal quantum yield, Fv/Fm, was determined by a saturation pulse of
513 >8000 μmol photons m^{-2} s^{-1} and a measurement light of 0.2 μmol photons m^{-2} s^{-1}.

514 **Effect of light conditions on *P. ocellatus* longevity**

515 The longevity of PoB was measured using a modified medaka (Japanese rice fish) housing rack system
516 (Iwaki, Japan). For the longevity measurement, different individuals from the photosynthetic activity
517 measurement were used. The longevity was investigated from the samples used in the above-mentioned
518 photosynthetic activity measurement. Using centrally filtered systems, the water tank rack maintained consistent
519 water conditions (e.g., temperature and mineral concentrations) among the incubation chambers (subtanks) and
520 enabled a focus on the effect of light conditions. After acclimating the collected sea slugs under the same
521 conditions for 1 week in an aquarium, the organisms were incubated separately for 8 months under different
522 light conditions (continuous dark and 12 h light/12 h dark cycle). The sea slugs' conditions were measure daily,
523 and death was defined when a sea slug remained motionless for 30 s after stimulation (i.e., touching with plastic
524 bar).

525 **Sequencing of *H. borneensis* cpDNA**

526 A combination of pyrosequencing and Sanger sequencing was used to evaluate *H. borneensis* cpDNA.
527 Collected *H. borneensis* thalli were washed with tap water to remove the attached organisms. The cleaned thalli
528 (27 g) were frozen in liquid nitrogen, ground with a T-10 basic homogenizer (IKA, Germany) to a fine powder,
529 suspended in 15 mL AP1 buffer from a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany), and centrifuged
530 (500 × *g*, 1 min) to remove the calcareous parts. Total DNA was purified from the supernatant according to the
531 protocol supplied with the DNeasy Plant Mini Kit. The resulting DNA yield (76.8 μg) was measured using a
532 dsDNA HS Assay Qubit Starter Kit (Thermo, Waltham, MA, USA) and used to prepare a single-fragment
533 library for pyrosequencing. The Pyrosequencer GS-FLX Titanium Platform (Roche, Germany) was used to
534 generate 23.04 Mb of total singleton reads (68032 reads; average read length, 368 bp). After filtering low-
535 quality reads, the remaining 21865 reads (~8 Mb) were submitted for assembly by Newbler (Roche). Of the
536 6309 obtained contigs (N50 = 663 bp), the cpDNA sequences were identified by blastx searches (version 2.2.28)
537 against the protein coding sequences of cpDNA from the chlorophyte alga *B. hypnoides* (NC_013359). The gaps
538 between the four identified contigs (55551, 21264, 8076, and 3435 bp) and ambiguous sites in the contigs were
539 amplified by PCR using inverse primers. The PCR products were sequenced by primer walking and Sanger
540 sequencing with Takara LA Taq (Takara, Japan), a dGTP BigDye Terminator Cycle Sequencing FS Ready
541 Reaction Kit (Thermo), and an ABI PRISM 3130xl DNA Sequencer (Thermo). Regions that could not be read

542 by direct sequencing were amplified using specific primers, cloned with a TOPO TA cloning kit (Thermo), and
543 sequenced with plasmid-specific primers. Complete cpDNA sequences were obtained by assembling the GS-
544 FLX contigs and reads generated by Sanger sequencing using Sequencher version 4.10 (Gene Codes
545 Corporation, Ann Arbor, MI, USA). All open reading frames >100 bp were annotated using a blastx search
546 version 2.2.31+ against the non-redundant protein sequences (nr) database in GenBank and a tblastx search of
547 chloroplast genes from other algae (*B. hypnoides*, *C. reinhardtii*, *Vaucheria litorea*, and *H. borneensis*). Introns
548 were detected using RNAweasel (Gautheret and Lambert, 2001), Rfam (Kalvari et al., 2018), and Mfold (Zuker,
549 2003). The splicing sites were confirmed via alignment with orthologous genes from other green algal cpDNAs.
550 MAFFT version 7.127b (Katoh and Standley, 2013) was used to perform the alignment. The origins of bacteria-
551 like proteins were explored using a blastx search against the nr database and a phylogenetic analysis with blast-
552 hit sequences. MAFFT (Katoh and Standley, 2013) was used for alignment, Trimal version 1.4 (Capella-
553 Gutiérrez et al., 2009) was used for trimming, and RaxML version 8.2.4 (Stamatakis, 2014) was used for
554 phylogenetic tree construction. The resulting gene maps were visualized using Circos version 0.69-2.

555 **Sequencing of kpDNAs from *P. ocellatus***

556 Kleptoplast sequences were generated using an Illumina system. Total DNA was extracted from DG
557 (kleptoplast-rich tissue) and parapodium (including DG, kleptoplast-less muscle, and reproductive systems) of a
558 single PoB individual using a CTAB-based method (Murray and Thompson, 1980). Two Illumina libraries with
559 180 and 500 bp insertions were constructed from each DNA pool (DRR063261, DRR063262, DRR063263, and
560 DRR063264). An S220-focused ultrasonicator (Covaris, Woburn, MA, USA), Pippin Prep (Sage Science,
561 Beverly, MA, USA), and a TruSeq DNA Sample Prep Kit (Illumina, San Diego, CA, USA) were used for DNA
562 fragmentation, size selection, and library construction, respectively. Libraries were sequenced (101 bp from each
563 end) on a HiSeq 2500 platform (Illumina). A total of 42,206,037 raw reads (8.53 Gb) were obtained. After
564 filtering the low-quality and adapter sequences, the remaining 5.21 Gb of sequences were used for assembly.
565 Paired sequences from 180 bp libraries were combined into overlapping extended contigs using FLASH version
566 1.2.9 (Magoc and Salzberg, 2011) with the default settings. An input of 14,867,401 paired-end sequences and
567 FLASH were used to construct 13,627,554 contigs (101–192 bp). The joined fragments and filtered paired
568 sequences from 500 bp libraries were assembled using Velvet assembler version 1.2.07 (Zerbino and Birney,
569 2008) with parameters that were optimized based on the nucDNA and kpDNA sequence coverage depths; the

570 estimated nucDNA depth was approximately 30× based on the k-mer analysis, and the predicted kpDNA depth
571 was 272× based on the read mapping to previously obtained kleptoplast *rbcL* sequences (AB619313; 1195 bp)
572 using Bowtie2 version 2.0.0 (Langmead and Salzberg, 2012). After several tests to tune the Velvet parameters,
573 the best assembly was achieved with a k-mer of 83 and exp_cov of 50. The resulting assembly comprised 1537
574 scaffolds (>2000 bp) containing 4,743,113 bp (N50 = 2830 bp). Two kpDNAs (AP014542 and AP014543) from
575 this assembly were then identified based on the sequence similarity with *H. borneensis* cpDNA and a mapping
576 back analysis. blastx version 2.2.31 assigned bit scores >1000 to the two scaffolds (AP014542 = 1382,
577 AP014543 = 1373, database = coding sequences in the obtained *H. borneensis* cpDNA, and query = all
578 constructed scaffolds). Mapping back, which was performed using BWA version 0.7.15-r1140, showed that the
579 coverage depths of AP014542 and AP014543 correlated with the relative abundance of kleptoplasts; the read's
580 average coverage depth was increased by twofold to fourfold in a library from kleptoplast-rich tissue
581 (DRR063263) relative to a library from kleptoplast-poor tissue (DRR063261). The same degree of change was
582 never observed in other scaffolds (Figure 3–Figure Supplement 7).

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

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

589 The *de novo* transcript profiles of *H. borneensis* and *B. hypnoides* were obtained from Illumina RNA-Seq
590 data. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen). An Illumina library was constructed
591 for each species using the TruSeq RNA Sample Prep Kit. The libraries were sequenced on a HiSeq 2500
592 platform (101 bp from each end). A total of 290,523,622 reads (29 Gb) and 182,455,350 raw reads (18 Gb) were
593 obtained for *H. borneensis* and *B. hypnoides*, respectively (Supplementary File 1). After filtering the low-quality
594 and adapter sequences, the obtained reads were assembled using Trinity version 2.4.0 (Grabherr et al., 2011) and
595 clustered using CD-Hit version 4.6 (Fu et al., 2012) with the -c 0.95 option. The TransDecoder version 2.0.1 was
596 used to identify 26,652 and 24,127 candidate coding regions from *H. borneensis* and *B. hypnoides*, respectively
597 (Supplementary File 2). The gene completeness of the transcripts was estimated using BUSCO version 2.0

598 (Waterhouse et al., 2018). The obtained *H. borneensis* transcripts covered 86.5% (262/303) of the total BUSCO
599 groups, and the *B. hypnoides* transcripts covered 92.7% (281/303) of the conserved genes in Eukaryota
600 (database, eukaryota_odb9). The transcripts were annotated using AHRD version 3.3.3
601 (https://github.com/groupschoof/AHRD) based on the results of a blastp search against nr, RefSeq, and
602 *Chlamydomonas* proteome dataset on UniProt. The composed functional domains on the transcripts were
603 annotated using InterProScan version 5.23-62 (Jones et al., 2014). To distinguish the reliable target species
604 transcripts, the original transcript species were predicted using MEGAN version 5 (Huson et al., 2007) and
605 11,629 and 8630 transcripts were selected as viridiplantal genes. The annotation procedure details are visually
606 presented in Figure 3–Figure Supplement 8 and 9.

607 A query dataset from the algal transcripts was manually selected to search algae-derived genes on the sea
608 slug DNAs. Then, 176 and 129 transcripts were selected from *H. borneensis* and *B. hypnoides*, respectively. To
609 perform more comprehensive searches, queries from three public genomic datasets derived from *C. lentillifera*,
610 *C. reinhardtii*, and *Cyanidioschyzon merolae* were also obtained; these queries were termed as the A614 dataset
611 (Supplementary File 3; Figure 3–Figure Supplement 10).

612 **Sequencing of the PoB genome**

613 The mean nucDNA size in three PoB individuals was estimated using flow cytometry (Figure 4–Figure
614 Supplement 1). Dissected parapodial tissue (5 mm^2) was homogenized in 1 ml phosphate-buffered saline (PBS)
615 containing 0.1% Triton X-100 (Thermo) and 0.1% RNase A (Qiagen) using a BioMasher (Nippi, Tokyo, Japan).
616 The homogenate was then filtered through a 30 μm CellTrics filter (Sysmex, Hyogo, Japan), and the filtrate was
617 diluted with PBS at a density of $<5 \times 10^6$ cells/ml. The resulting solution was mixed with genome size standard
618 samples and stained with a 2% propidium iodide solution (SONY, Tokyo, Japan). *Acyrthosiphon pisum* (genome
619 size = 517 Mb) and *Drosophila melanogaster* (165 Mbp) samples were used as genome size standards. The
620 references were processed using the same method described for PoB. The mixture was analyzed on Cell Sorter
621 SH800 (SONY) according to the manufacturer's instructions. The above procedure was repeated for three PoB
622 individuals, and an estimated genome size of 936 Mb was determined.

623 Genomic DNA was extracted from a single PoB individual using the CTAB-based method (Murray and
624 Thompson, 1980). The adapted buffer compositions are summarized in Supplementary File 16. A fresh PoB
625 sample (collected on October 17, 2013 and starved for 21 days) was cut into pieces and homogenized in 2×

626 CTAB using a BioMasher. To digest the tissues, 2% volume of proteinase K solution (Qiagen) was added, and
627 the sample was incubated overnight at 55°C. The lysate was emulsified by gentle inversion with an equal
628 volume of chloroform; after centrifugation (12,000 × *g*, 2 min), the aqueous phase was collected using a pipette.
629 This phase was combined with a one-tenth volume of 10% CTAB, mixed well at 60°C for 1 min, and again
630 emulsified with chloroform. These 10% CTAB and chloroform treatment steps were repeated until a clear
631 aqueous phase was achieved. The aqueous phase was then transferred to a new vessel, an equal volume of
632 CTAB precipitation buffer was overlaid onto the mixture, and the liquids were gently mixed by tapping. The
633 resulting filamentous precipitations (DNA) were removed using a pipette chip and incubated at room
634 temperature for 10 min in high salt TE buffer. DNA was purified according to the protocol supplied with the
635 DNeasy Blood and Tissue Kit (Qiagen); briefly, the supernatant was transferred after vortex mixing, equal
636 volumes of buffer AL (supplied with the kit) and EtOH were added to the supernatant, and the sample was
637 processed on a Qiagen spin column according to the protocol. We finaly obtained a DNA quantity of 15 μg from
638 a PoB individual (Figure 4–Figure Supplement 11).

639 The PoB genomic sequence was obtained via Illumina paired-end and mate-pair DNA sequencing
640 (Supplementary File 1; Figure 4–Figure Supplement 2). Two paired-end Illumina libraries containing 250 and
641 600 bp insertions (DRR029525 and DRR029526) were constructed using a TruSeq DNA Sample Prep Kit
642 (Illumina). Three mate-pair libraries with 3000, 5000, and 10,000 bp insertions (DRR029528, DRR029529, and
643 DRR029530) were constructed using a Nextera Mate Pair Library Prep Kit (Illumina). The libraries were
644 sequenced (150 bp from each end) on a HiSeq 2500 platform (Illumina). A total of 1,130,791,572 and
645 787,040,878 raw reads were obtained for the paired-end (170 Gb) and mate-pair (118 Gb) libraries. After
646 filtering the low-quality and adapter sequences, the remaining 161 Gb of sequences were assembled using
647 Platanus assembler version 1.2.1 (Kajitani et al., 2014) with the default settings. The assembly comprised 8716
648 scaffolds containing 928,345,517 bp. The repetitive regions were masked using a combination of
649 RepeatModeler version open-1.0.8 and RepeatMasker version open-4.0.5 (http://www.repeatmasker.org). The
650 default parameters were used for identification and masking. A total of 268,300,626 bp (29%) of the assemblies
651 were masked with RepeatMasker.

652 Strand-specific RNA-Seq sequencing was used for gene modeling (Supplementary File 1; Figure 4–
653 Figure Supplement 2). PoB RNA was extracted from an individual after starvation for 20 days (collected on

654 October 17, 2013). TRIzol™ Plus RNA Purification Kit (Thermo) was used to extract RNA according to the
655 manufacturer's protocol. A paired-end Illumina library (DRR029460) was constructed using the TruSeq
656 Stranded mRNA LT Sample Prep Kit (Illumina). Libraries were sequenced (150 bp from each end) on a HiSeq
657 2500 platform (Illumina). The library produced a total of 286,819,502 raw reads (28 Gb).

658 The PoB assemblies were processed using a single transcript-based gene model construction pipeline
659 (AUGUSTUS version 3.2; Stanke and Morgenstern, 2005), two transcriptomic data mapping tools (Trinity
660 version 2.4.0 and Exonerate version 2.2.0; Grabherr et al., 2011; Slater and Birney, 2005), and two non-
661 transcript-based model construction pipelines (GeneMark-ES version 4.33 and glimmerHMM version 3.0.4;
662 Majoros et al., 2004; Ter-Hovhannisyan et al., 2008). The four obtained gene sets were merged with the
663 EVidenceModeler version 1.1.1 pipeline (Haas et al., 2008) to yield a final gene model set. For AUGUSTUS,
664 Braker pipeline version 1.9 (Hoff et al., 2019) was used to construct PoB-specific probabilistic models of the
665 gene structure based on strand-specific RNA-Seq data. After filtering the low-quality and adapter sequences, the
666 remaining 181,873,770 RNA-Seq reads (16 Gb) were mapped to the PoB genome assembly using TopHat
667 version 2.1.1 (Kim et al., 2013) with the default setting and the Braker pipeline-constructed PoB-specific
668 probabilistic models from mapped read data. TopHat mapped 132,786,439 of the reads (73%) to the PoB model.
669 AUGUSTUS then predicted 78,894 gene models from the TopHat mapping data (as "splicing junction" data)
670 and the Braker probabilistic model. Using Trinity and Exonerate, *de novo* transcriptomic data from RNA-Seq
671 data were then constructed, and these were aligned to the genome. Trinity constructed 254,336 transcripts,
672 which were clustered to 194,000 sequences using CD-Hit version 4.6 (-c 0.95); subsequently, Transdecoder
673 identified 44,596 protein coding regions from these sequences. Exonerate (--bestn 1 --percent 90 options) then
674 aligned 13,141 of the transcripts to the genome. GeneMark-ES with the default settings predicted 107,735 gene
675 models, and glimmerHMM predicted 115,633 models after the model training, with 320 manually constructed
676 gene models from long scaffolds. EVidenceModeler was then used to merge the model with the following
677 weight settings: AUGUSTUS = 9, Exonerate = 10, GeneMark-ES = 1, and glimmerHMM = 2. Finally,
678 EVidenceModeler predicted 77,444 gene models.

679 The removal of contaminant sequences was minimized to avoid missing horizontally transferred genes.
680 Bacterial scaffolds were defined as those encoding >1 bacterial gene with no lophotrochozoan gene, and the
681 potential bacterial scaffolds were removed from the PoB assemblies. The bacterial genes were predicted using

682 MEGAN software according to a blastp search against the RefSeq database. Of the 40,330 gene hits identified
683 from the RefSeq data, MEGAN assigned the origins for 39,113 genes. Specifically, 719 and 23,559 genes were
684 assigned as bacterial and lophotrochozoan genes, respectively. MEGAN results before removing bacterial
685 scaffolds are summarized in Supplementary File 17 with the detailed data of removed scaffolds. Fifty-five of the
686 8716 scaffolds contained two or more bacterial genes and no lophotrochozoan gene. The mean depth of read
687 coverage of the bacterial scaffolds differed from the mean coverage of sea slug scaffolds.

688 Kleptoplast- or mitochondria-derived scaffolds were also removed from the assemblies (Supplementary
689 File 18). The source of scaffolds was determined based on a blast bit score against the three referential
690 organelle DNAs (kRhip AP014542, kPoro AP014543, and PoB mtDNA AP014544) and the difference of
691 the read depth value from the other (nuclear-derived) scaffolds. The blastn search detected 13 and 1
692 scaffolds as sequences of kleptoplast or mitochondrial origin, respectively (bit score >1000; Supplementary
693 File 18). Mapping back of the Illumina read (DRR029525) by Bowtie version 2.4.1 indicated that the depth
694 values of the 14 scaffolds were 537 to 5143, and the averaged depth of the other scaffolds was 31
695 (Supplementary Files 17 and 18), indicating that the 14 scaffolds were derived from organelle DNAs or
696 repetitive regions on the nucDNA. Mapping of the reads derived from DG (kleptoplast enriched tissue;
697 DRR063263) and parapodium (including a muscle, gonad, and DG; DRR063261) indicated that the relative
698 read depth of the 13 kpDNA-like scaffolds was higher in the DG sample than in the parapodium sample,
699 supporting that the 13 scaffolds are derived from the kleptoplasts (Figure 4–Figure Supplement 12). It was
700 then confirmed that the 14 scaffolds contain no algal nuclear-derived photosynthetic gene using two
701 methods: dot plots with referential organelle DNAs and homology search using A612 query set (Figure 4–
702 Figure Supplement 13–15). Therefore, even if these removed 14 scaffolds are part of the nucDNA, they do
703 not provide HGT evidence. The removed scaffold sequence data were deposited in FigShare (DOI:
704 10.6084/m9.figshare.12587954).

705 The final PoB assembly comprised 8647 scaffolds containing 927,888,823 bp (N50 = 1,453,842 bp) and
706 77,230 genes. Gene completeness was estimated using BUSCO version 2.0 (Waterhouse et al., 2018). The
707 predicted gene models were annotated using AHRD version 3.3.3. The results of a blastp search against the
708 SwissProt, Trembl, and *A. californica* proteome datasets on UniProt were used as reference data for AHRD

709 under the following weight parameter settings: SwissProt = 653, Trembl = 904, and *A. californica* = 854. The
710 functional domains were annotated using InterProScan version 5.23-62 (Figure 4–Figure Supplement 3).

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

716 The GO annotation was assigned using Blast2GO version 5.2.5 according to the blastp searches against
717 the RefSeq database and InterProScan results. SonicParanoid version 1.0.11 was then used for orthogroup
718 detection. All parameters of SonicParanoid were left at the default values. The species analyzed in the
719 orthogroup detections are summarized in Supplementary File 19. The phylogenetic tree was constructed using
720 IQ-tree. The resulting trees were visualized using iTol version 4.

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

725 MMseq2 version 2.6 (--orf-start-mode 1) was used to search for algae-like reads among the trimmed
726 Illumina reads. The matching threshold was set using a default *E*-value < 0.001. As a positive control, PoB
727 genes were selected from the BUSCO analysis of the genomic model data, and 911 gene models detected by
728 BUSCO version 2 were selected as single-copy orthologs of the metazoa_odb9 gene set and named the dataset
729 P911 (DOI: 10.6084/m9.figshare.13042016).

730 The HGT index "h" and the modified index "hA" were calculated using the R script, HGT_index_cal.R
731 (https://github.com/maedat/HGT_index_cal; R version 3.6.1). The h-index was calculated as the difference in bit
732 scores between the best prokaryote and best eukaryote matches in the blast alignments, and the hA-index was
733 calculated as the difference in bit scores between the best lophotrochozoan and best algae matches. The blast
734 databases of adapted species are summarized in Figure 4–Source Data 1.

735 **RNA-Seq analysis of PoB tissues**

736 Total RNA samples from five PoB individuals and one egg mass were obtained for a gene expression
737 analysis. An overview of sample preparation is illustrated in Figure 4–Figure Supplement 10. Collected adult
738 PoB individuals were dissected manually after incubation for 21 to 94 days. An egg mass was obtained via
739 spontaneous egg lying in an aquarium. A TRIzol™ Plus RNA Purification Kit (Thermo) was used to extract
740 RNA according to the manufacturer's protocol. Six paired-end and nine single-end Illumina libraries were
741 constructed using a combination of the RiboMinus Eukaryote Kit (Thermo), RiboMinus concentration module
742 (Thermo), and TruSeq RNA Sample Preparation Kit version 2 (Illumina) according to the manufacturers'
743 protocols. Libraries were sequenced (101 bp) on a HiSeq 2500 platform (Illumina). A total of 280,445,422 raw
744 reads (28 Gb) were obtained from the libraries (Supplementary File 1).

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

749 The Hisat-stringtie-DESeq2 pipeline was used to conduct a differential gene expression analysis of DGs
750 and DeP (epidermis, muscle, and reproductive systems, DeP). According to the Stringtie protocol manual
751 (http://ccb.jhu.edu/software/stringtie/index.shtml?t=manual), trimmed RNA-Seq reads were mapped to the PoB
752 genome assembly using Hisat version 2.1.0 with the default setting. The obtained BAM files were processed
753 using Stringtie version 1.3.4d (-e option) with PoB gene model data (gff3 format) acquired through the above-
754 mentioned EVidenceModeler analysis. Following the default settings of Stringtie software, if more than 95% of
755 the reads aligned in a gene locus are multi-mapped, processing of that locus is skipped (Pertea et al. 2015). It
756 was assumed that this reduces the type I error on the gene enrichment analysis on the duplicated orthogroup. The
757 pairwise sequence similarity of three KRM gene-including orthogroups is summarized on Supplementary File
758 21. The averaged pairwise sequence similarities were 70.3%, 71.0%, and 66.9% in OG0000005, OG0000132,
759 and OG0000446, respectively. The resulting count data were analyzed using R and the DESeq2 package, and
760 1490 DEGs ($p < 0.01$ and $p_{adj} < 0.05$) were identified between the tissues. The GOseq (Young et al., 2010) and
761 topGO packages in R were used to apply a GO enrichment analysis to the up-regulated genes in DG tissue
762 (threshold: $p < 0.01$).

763 **Sequencing of the *E. marginata* genome**

764 The *E. marginata* genome sequencing process was nearly identical to the methodology applied to PoB.
765 Flow cytometry yielded an estimated genome size of 900 Mb. Genomic DNA was extracted from an individual
766 using a CTAB-based method. Four types of Illumina libraries were constructed: two paired-end libraries with
767 250 and 500 bp insertions and two mate-pair libraries with 3000 and 5000 bp insertions (Supplementary File 1).
768 Using the HiSeq 2500 platform (Illumina), 562,732,268 and 608,977,154 raw reads were obtained for the
769 paired-end (84 Gb) and mate-pair (91 Gb) libraries, respectively. After filtering the low-quality and adapter
770 sequences, the remaining 40 Gb of sequences were assembled using the Platanus assembler. The assembly
771 comprised 14,285 scaffolds containing 791,005,940 bp.

772 For gene modeling, strand-specific RNA-Seq sequencing of *E. marginata* was performed. A paired-end
773 Illumina library (DRR029460; see also Supplementary File 1) was constructed using a TruSeq Stranded mRNA
774 LT Sample Prep Kit (Illumina) and an RNA sample extracted from an *E. marginata* individual via a TRIzol™
775 Plus RNA Purification Kit (Thermo). Libraries were sequenced (150 bp from each end) on a HiSeq 2500
776 platform (Illumina). A total of 286,819,502 raw reads (28 Gb) were obtained from the library. Using the gene
777 modeling procedure described for PoB, EVidenceModeler constructed 71,137 gene models of the genomic
778 assemblies based on the RNA-Seq data.

779 Next, contaminant-derived bacterial scaffolds were removed from the genomic assemblies. Using the
780 same gene annotation as applied to PoB, it was determined that the 110 of the 14,285 scaffolds contained >1
781 bacterial gene and no lophotrochozoan gene, and these scaffolds were removed. The organelle-derived scaffolds
782 (kpDNA and mtDNA) were identified using blastn searches and removed from the final assemblies. A blastn
783 search (query = all scaffolds, database = cpDNA of *B. hypnoides* NC_013359.1 or mtDNA of PoB) identified 25
784 kleptoplast-matching scaffolds and 1 mitochondria-matching scaffold (bit score > 1000). The complete kpDNA
785 and mtDNA was then reassembled using the same method as described for PoB organellar DNA assembling.

786 **Ortholog analysis of sacoglossan genes**

787 Orthologous relationships were classified using OrthoFinder version 2.2.3 (Emms and Kelly, 2015) and
788 rapidly expanded/contracted families were identified using CAFE version 4.2 (Han et al., 2013) based on the
789 OrthoFinder results. All parameters of OrthoFinder were left at the default values, and no "user-specified rooted

species tree" was set. OrthoFinder and CAFE analyses used 16 metazoan species as reference species (Figure 5–Figure Supplement 3). Phylogenetic trees for CAFE were constructed based on 30 single-copy genes in 15 major species according to the OrthoFinder results (Figure 5–Source Data 2). The sequences were trimmed using PREQUAL version 1.02 (Whelan et al., 2018), aligned with MAFFT version 7.407, and analyzed with IQ-tree version 1.6.1 for the ML analysis. The obtained ML tree was converted to an ultrametric tree for CAFE analysis based on the divergence times of Amphiesmenoptera-Antliophora (290 Myr) and Euarchontia-Glires (65 Myr) using r8s version 1.81 (Sanderson, 2003).

The expanded/contracted families was also analyzed using the *Z*-scores of the assigned gene numbers for each orthogroup. The analysis included all 16 referential species and 2 sacoglossan species, and an expanded group of 38 was determined on the PoB lineage (Threshold: *Z*-score > 2; Figure 5–Figure Supplement 5). The phylogenetic relationships in the orthogroups were analyzed using a PREQUAL-MAFFT-IQ-tree as in the CAFE analysis. The domain structures and gene positions on the constructed genomic data were visualized using the GeneHere script (https://github.com/maedat/GeneHere) and Biopython packages.

Acknowledgments

We appreciate the incisive comments of Masayoshi Kawaguchi, Atsushi J. Nagano, and Kan Tanaka. The sample collection was partially performed 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), the NIBB Functional Genomics Facility, the NIBB Data Integration and Analysis Facility, and the Japan Advanced Plant Science Network. Computations were partially performed on the NIG supercomputer at the 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 under BioProjects PRJDB4939, PRJDB3267, PRJDB10060, and PRJDB5024. All data collected in this study are available on FigShare at DOI: 10.6084/m9.figshare.12587954 and 10.6084/m9.figshare.13042016. The codes used to analyze the HGT index and visualize gene distribution on scaffolds are available on https://github.com/maedat/HGT_index_cal and https://github.com/maedat/GeneHere.

822 # References

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

826 Akimoto A, Hirano YM, Sakai A, Yusa Y. 2014. Relative importance and interactive effects of photosynthesis and
827 food in two solar-powered sea slugs. *Mar Biol* 161:1095–1102.

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

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

830 Arimoto A, Nishitsuji K, Higa Y, Arakaki N, Hisata K, Shinzato C, Satoh N, Shoguchi E. 2019. A siphonous
831 macroalgal genome suggests convergent functions of homeobox genes in algae and land plants. *DNA Res*
832 **26**:183–192.

833 Benes P, Vetvicka V, Fusek M. 2008. Cathepsin D--many functions of one aspartic protease. *Crit Rev Oncol Hematol*
834 **68**:12–28.

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

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

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

841 Bonente G, Pippa S, Castellano S, Bassi R, Ballottari M. 2012. Acclimation of *Chlamydomonas reinhardtii* to
842 different growth irradiances. *J Biol Chem* 287:5833–5847.

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

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

847 Cai H, Li Q, Fang X, Li J, Curtis NE, Altenburger A, Shibata T, Feng M, Maeda T, Schwartz JA, Shigenobu S,

848 Lundholm N, Nishiyama T, Yang H, Hasebe M, Li S, Pierce SK, Wang J. 2019. A draft genome assembly of the
849 solar-powered sea slug *Elysia chlorotica*. *Sci Data* **6**:190022.

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

852 Cartaxana P, Trampe E, Kühl M, Cruz S. 2017. Kleptoplast photosynthesis is nutritionally relevant in the sea slug
853 *Elysia viridis*. *Sci Rep* **7**:7714.

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

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

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

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

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

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

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

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

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

874 Cruz S, Cartaxana P, Newcomer R, Dionísio G, Calado R, Serôdio J, Pelletreau KN, Rumpho ME. 2015.

875 Photoprotection in sequestered plastids of sea slugs and respective algal sources. *Sci Rep* **5**:7904.

876 Curtis NE, Pierce SK, Massey SE, Schwartz J, Maugel TK. 2007. Newly metamorphosed *Elysia clarki* juveniles feed
877 on and sequester chloroplasts from algal species different from those utilized by adult slugs. *Marine Biology*
878 **150**:797–806.

879 Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M,
880 Goodstein DM, Harafuji N, Hastings KEM, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D,
881 Meinertzhagen IA, Necula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang H-
882 G, Awazu S, Azumi K, Boore J, Branno M, Chin-Bow S, DeSantis R, Doyle S, Francino P, Keys DN, Haga S,
883 Hayashi H, Hino K, Imai KS, Inaba K, Kano S, Kobayashi K, Kobayashi M, Lee B-I, Makabe KW, Manohar C,
884 Matassi G, Medina M, Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H,
885 Ohta F, Oishi K, Rigoutsos I, Sano M, Sasaki A, Sasakura Y, Shoguchi E, Shin-i T, Spagnuolo A, Stainier D,
886 Suzuki MM, Tassy O, Takatori N, Tokuoka M, Yagi K, Yoshizaki F, Wada S, Zhang C, Hyatt PD, Larimer F,
887 Detter C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS.
888 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298**:2157–
889 2167.

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

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

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

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

899 Fan X, Qiao H, Xu D, Cao S, Ye N. 2014. Short-term retention of kleptoplasty from a green alga (*Bryopsis*) in the sea
900 slug *Placida* sp. YS001. *Biologia* **69**:635–643.

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

903 subunit. *EMBO J* **14**:4976–4984.

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

906 Gautheret D, Lambert A. 2001. Direct RNA motif definition and identification from multiple sequence alignments
907 using secondary structure profiles. *Journal of Molecular Biology* **313**:1003–1011.

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

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

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

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

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

922 Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated
923 eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments.
924 *Genome Biol* **9**:R7.

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

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

929 Han JH, Klochkova TA, Han JW, Shim J, Kim GH. 2015. Transcriptome analysis of the short-term photosynthetic sea

930 slug *Placida dendritica*. *Algae* **30**:303–312.

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

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

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

937 Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. *Methods Mol Biol*
938 **1962**:65–95.

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

940 Izumi M, Tsunoda H, Suzuki Y, Makino A, Ishida H. 2012. RBCS1A and RBCS3B, two major members within the
941 *Arabidopsis* RBCS multigene family, function to yield sufficient Rubisco content for leaf photosynthetic
942 capacity. *J Exp Bot* **63**:2159–2170.

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

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

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

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

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

956 Kawaguti S. 1941. Study on the invertebrates associating unicellular algae I. *Placobranchus ocellatus* von Hasselt, a

957 nudibranch. *Palau Tropical Biological Station Studies* **2**.

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

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

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

965 Laetz EMJ, Wägele H. 2017. Chloroplast digestion and the development of functional kleptoplasty in juvenile *Elysia*
966 *timida* (Risso, 1818) as compared to short-term and non-chloroplast-retaining sacoglossan slugs. *PLoS One*
967 **12**:e0182910.

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

969 Lis H, Sharon N. 1998. Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem Rev* **98**:637–
970 674.

971 Maeda T, Hirose E, Chikaraishi Y, Kawato M, Takishita K, Yoshida T, Verbruggen H, Tanaka J, Shimamura S,
972 Takaki Y, Tsuchiya M, Iwai K, Maruyama T. 2012. Algivore or phototroph? *Plakobranchus ocellatus*
973 (Gastropoda) continuously acquires kleptoplasts and nutrition from multiple algal species in nature. *PLoS One*
974 **7**:e42024.

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

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

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

981 Maxwell K, Johnson GN. 2000. Chlorophyll fluorescence—a practical guide. *J Exp Bot* 51:659–668.

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

984 Opisthobranchia, Sacoglossa). *Zookeys* 73–98.

985 Murray MG, Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* **8**:4321–
986 4325.

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

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

993 Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved
994 reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33:290–295.

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

997 Pierce SK, Curtis NE. 2012. Cell biology of the chloroplast symbiosis in sacoglossan sea slugs. *Int Rev Cell Mol Biol*
998 **293**:123–148.

999 Pierce SK, Curtis NE, Hanten JJ, Boerner SL, Schwartz J. 2007. Transfer, integration and expression of functional
1000 nuclear genes between multicellular species. *Symbiosis* **43**:57–64.

1001 Pierce SK, Curtis NE, Schwartz J. 2010. Chlorophyll a synthesis by an animal using transferred algal nuclear genes.
1002 *Symbiosis* **49**:121–131.

1003 Pierce SK, Massey SE, Hanten JJ, Curtis NE. 2003. Horizontal transfer of functional nuclear genes between
1004 multicellular organisms. *Biol Bull* **204**:237–240.

1005 Pigolev AV, Zharmukhamedov SK, Klimov VV. 2009. The *psbO* mutant of *Chlamydomonas reinhardtii* is capable of
1006 assembling stable, photochemically active reaction center of photosystem II. *Biochemistry (Moscow) Supplement*
1007 *Series A: Membrane and Cell Biology* **3**:33–41.

1008 Rauch C, Vries J de, Rommel S, Rose LE, Woehle C, Christa G, Laetz EM, Wägele H, Tielens AGM, Nickelsen J,
1009 Schumann T, Jahns P, Gould SB. 2015. Why it Is time to look beyond algal genes in photosynthetic slugs.
1010 *Genome Biol Evol* **7**:2602–2607.

1011 Reyes-Prieto A, Weber APM, Bhattacharya D. 2007. The origin and establishment of the plastid in algae and plants.
1012 *Annu Rev Genet* **41**:147–168.

1013 Roberts IN, Miranda H, Tâm LX, Kieselbach T, Funk C. 2013. PsbO degradation by Deg proteases under reducing
1014 conditions In: Kuang T, Lu C, Zhang L, editors. Photosynthesis Research for Food, Fuel and the Future. Springer
1015 Berlin Heidelberg. pp. 599–602.

1016 Rumpho ME, Pelletreau KN, Moustafa A, Bhattacharya D. 2011. The making of a photosynthetic animal. *J Exp Biol*
1017 **214**:303–311.

1018 Rumpho ME, Worful JM, Lee J, Kannan K, Tyler MS, Bhattacharya D, Moustafa A, Manhart JR. 2008. Horizontal
1019 gene transfer of the algal nuclear gene *psbO* to the photosynthetic sea slug *Elysia chlorotica*. *Proc Natl Acad Sci*
1020 *U S A* **105**:17867–17871.

1021 Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a
1022 molecular clock. *Bioinformatics* **19**:301–302.

1023 Sasakura Y, Ogura Y, Treen N, Yokomori R, Park S-J, Nakai K, Saiga H, Sakuma T, Yamamoto T, Fujiwara S,
1024 Yoshida K. 2016. Transcriptional regulation of a horizontally transferred gene from bacterium to chordate. *Proc*
1025 *Biol Sci* **283**. doi:10.1098/rspb.2016.1712

1026 Schwartz JA, Curtis NE, Pierce SK. 2014. FISH labeling reveals a horizontally transferred algal (*Vaucheria litorea*)
1027 nuclear gene on a sea slug (*Elysia chlorotica*) chromosome. *Biol Bull* **227**:300–312.

1028 Serôdio J, Cruz S, Cartaxana P, Calado R. 2014. Photophysiology of kleptoplasts: photosynthetic use of light by
1029 chloroplasts living in animal cells. *Philos Trans R Soc Lond B Biol Sci* **369**:20130242.

1030 Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC*
1031 *Bioinformatics* **6**. doi:Artn 31 10.1186/1471-2105-6-31

1032 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.
1033 *Bioinformatics* **30**:1312–1313.

1034 Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-
1035 defined constraints. *Nucleic Acids Res* **33**:W465–7.

1036 Taylor DL. 1968. Chloroplasts as symbiotic organelles in the digestive gland of *Elysia viridis* (Gastropoda:
1037 opisthobranchia). *J Mar Biol Assoc U K* **48**:1–15.

1038 Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes
1039 using an *ab initio* algorithm with unsupervised training. *Genome Res* **18**:1979–1990.

1040 Teugels B, Bouillon S, Veuger B, Middelburg JJ, Koedam N. 2008. Kleptoplasts mediate nitrogen acquisition in the
1041 sea slug *Elysia viridis*. *Aquatic Biology* **4**:15–21.

1042 Trench ME, Trench RK, Muscatine L. 1970. Utilization of photosynthetic products of symbiotic chloroplasts in
1043 mucus synthesis by *Placobranchus ianthobapsus* (Gould), Opisthobranchia, Sacoglossa. *Comparative*
1044 *Biochemistry and Physiology* **37**:113–117.

1045 Trench RK. 1969. Chloroplasts as functional endosymbionts in the mollusc *Tridachia crispata* (Bërgh),
1046 (Opisthobranchia, Sacoglossa). *Nature* **222**:1071.

1047 Wade RM, Sherwood AR. 2017. Molecular determination of kleptoplast origins from the sea slug *Plakobranchus*
1048 *ocellatus* (Sacoglossa, Gastropoda) reveals cryptic bryopsidalean (Chlorophyta) diversity in the Hawaiian
1049 Islands. *J Phycol* **53**:467–475.

1050 Wägele H, Deusch O, Händeler K, Martin R, Schmitt V, Christa G, Pinzger B, Gould SB, Dagan T, Klussmann-Kolb
1051 A, Martin W. 2011. Transcriptomic evidence that longevity of acquired plastids in the photosynthetic slugs
1052 *Elysia timida* and *Plakobranchus ocellatus* does not entail lateral transfer of algal nuclear genes. *Mol Biol Evol*
1053 **28**:699–706.

1054 Wägele M, Johnsen G. 2001. Observations on the histology and photosynthetic performance of "solar-powered"
1055 opisthobranchs (Mollusca, Gastropoda, Opisthobranchia) containing symbiotic chloroplasts or zooxanthellae.
1056 *Org Divers Evol* **1**:193–210.

1057 Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM.
1058 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*
1059 **35**:543–548.

1060 Watson JD, Crick FH. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*
1061 **171**:737–738.

1062 Whelan S, Irisarri I, Burki F. 2018. PREQUAL: detecting non-homologous characters in sets of unaligned
1063 homologous sequences. *Bioinformatics* **34**:3929–3930.

1064 Wickner RB, Shewmaker FP, Bateman DA, Edskes HK, Gorkovskiy A, Dayani Y, Bezsonov EE. 2015. Yeast prions:

1065 structure, biology, and prion-handling systems. *Microbiol Mol Biol Rev* **79**:1–17.

1066 Worthley DL, Bardy PG, Mullighan CG. 2005. Mannose-binding lectin: biology and clinical implications. *Intern Med*
1067 *J* **35**:548–555.

1068 Yamamoto S, Hirano YM, Hirano YJ, Trowbridge CD, Akimoto A, Sakai A, Yusa Y. 2013. Effects of photosynthesis
1069 on the survival and weight retention of two kleptoplastic sacoglossan opisthobranchs. *J Mar Biol Assoc U K*
1070 **93**:209–215.

1071 Yamamoto YY, Yusa Y, Yamamoto S, Hirano Y, Hirano Y, Motomura T, Tanemura T, Obokata J. 2009.
1072 Identification of photosynthetic sacoglossans from Japan. *Encocytobiosis Cell Res* **19**:112.

1073 Yamasaki S, Ishikawa E, Sakuma M, Hara H, Ogata K, Saito T. 2008. Mincle is an ITAM-coupled activating receptor
1074 that senses damaged cells. *Nat Immunol* **9**:1179–1188.

1075 Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for
1076 selection bias. *Genome Biol* **11**:R14.

1077 Zapata F, Wilson NG, Howison M, Andrade SCS, Jörger KM, Schrödl M, Goetz FE, Giribet G, Dunn CW. 2014.
1078 Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proc Biol Sci* **281**:20141739.

1079 Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using *de Bruijn* graphs. *Genome Res*
1080 **18**:821–829.

1081 Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**:3406–
1082 3415.

1083

1084 Figure Legends

1085 **Figure 1. Kleptoplasty in sea slugs.**

1086 **(A)** Process of algal chloroplast retention by a sacoglossan sea slug (Pierce and Curtis, 2012). Sacoglossan sea
1087 slugs puncture the cell wall of food algae to suck out the protoplasm. The chloroplasts in the protoplasm are
1088 transported to the sea slug's intestinal tract, and the intestinal epithelial cells sequester chloroplasts by
1089 phagocytosis. The sequestered chloroplasts (kleptoplasts) maintain the photosynthetic activity in the cell for
1090 days to months. The sacoglossan cell contains no algal nuclei. Kleptoplast has never been found in germ cells of

1091 sea slug. **(B)** Phylogenetic distribution of kleptoplasty in the order Sacoglossa. Phylogenetic analysis showed
1092 that a common ancestor of Sacoglossa acquired non-functional chloroplast retention phenomena (without the
1093 maintenance of photosynthetic function), and multiple sacogolossan groups subsequently acquired the ability to
1094 maintain photosynthetic activity. Phylogenetic tree and kleptoplasty states are simplified from Christa et al.
1095 (2015). Christa et al. (2014b) defined functional chloroplast retention for less than two weeks as "short term
1096 retention", and for more than 20 days as "long term retention". Relationships within Heterobranchia are
1097 described according to Zapata et al. (2014). The red-colored taxa include the species used in the present study (*P.*
1098 *ocellatus* and *E. marginata*). **(C–E)** Photo images of PoB starved for 21 days. **(C)** Dorsal view. H, head; R,
1099 rhinophores; P, parapodia (lateral fleshy flat protrusions). Almost always, PoB folded parapodia to the back in
1100 nature. **(D)** The same individual of which parapodium was turned inside out (without dissection). The back of
1101 the sea slug and inside of the parapodia are green. This coloration is caused by the kleptoplasts in DG, which are
1102 visible through the epidermis. **(E)** Magnified view of the inner surface of parapodium. The diagonal green
1103 streaks are ridged projections on the inner surface of the parapodium. The cells containing kleptoplasts are
1104 visible as green spots. **(F)** Phylogeny of the *P.* cf. *ocellatus* species complex based on mitochondrial *cox1* genes
1105 (ML tree from 568 nucleotide positions) from INSD and the whole mtDNA sequence. The sequence data for the
1106 phylogenetic analysis are listed in Figure 1–source data 1. Clade names in square brackets are based on Krug et
1107 al. (2013). Asterisks mark the genotypes from Krug et al. (2013). Study topics analyzed by previous researchers
1108 were described within the colored boxes for each cluster. Small black circles indicate nodes supported by a high
1109 bootstrap value (i.e., 80–100%). *Thuridilla gracilis* is an outgroup. *Plakobranchus papua* is a recently described
1110 species and previously identified as *P. ocellatus* (Meyers-Muñoz et al., 2016).

1111

1112 **Figure 2. Photosynthetic activity of PoB.**

1113 **(A)** Jitter plot of Fv/Fm values indicating the photochemical efficiency of PSII. Habo, *H. borneensis*; d38,
1114 starved PoB for 38 days; d109, starved for 109 to 110 days (12 h light/12 h dark cycle; the light phase
1115 illumination was 10 μmol photons m^{-2} s^{-1}). The magenta line indicates the mean, and the black dot indicates
1116 each individual's raw value (n = 3 per group). **(B)** Dynamics of seawater oxygen concentration. To better
1117 demonstrate the PoB photosynthesis, a change of oxygen concentration every second was visualized. A PoB
1118 individual, one of the d38 samples (ID1), was put into the measurement chamber, and the light conditions were

1119 changed in tens of minutes. The measurements in other individuals are visualized in Figure 2–Figure
1120 Supplement 1, along with the diagrams of the measuring equipment. Each point represents the oxygen
1121 concentration value per second. Gray signifies a dark period, and yellow means an illuminated period (50 μmol
1122 photons m^{-2} s^{-1}). Temp, water temperature. Although the values fluctuated by the noise due to ambient light and
1123 other factors, the oxygen concentration decreased in the dark conditions and increased in the light conditions.
1124 When the illumination condition was changed, the changing pattern kept the previous pattern for a few minutes.
1125 This discrepancy may reflect the distance between the organism and oxygen sensor and the time for adaptation
1126 to brightness by the kleptoplast. **(C)** Jitter plots of PoB oxygen consumption and generation. D, dark conditions;
1127 L, light conditions; G, gross rate of light-dependent oxygen generation (L minus D). **(D)** Jitter plots of PoB
1128 longevity (n = 5 per group). D, Continuous dark; L/D, 12 h light/12 h dark cycle. The p from Welch's two-
1129 sample t-test was used. A source file of Fv/Fm jitter plot, time course of oxygen concentration, and longevity
1130 analysis are available in Figure 2–source data 1–3, respectively.

1131

1132 **Figure 3. Gene composition of PoB kpDNAs.**

1133 **(A)** Gene map of two kpDNAs from PoB. Gene positions are described in circles colored according to the
1134 gene's functional category (see keys in the box). Genes on the outside and inside of each circle are transcribed in
1135 the clockwise and anticlockwise directions, respectively (for detailed maps, see Figure 3–Figure Supplements 1
1136 and 2). **(B)** Phylogenetic positions of sequenced kleptoplasts among green algal plastids. The original ML tree
1137 (Figure 3–Figure Supplement 4) was created based on *rbcL* genes (457 positions) and converted to the tree. Red
1138 indicates sequenced kpDNA or cpDNA in the present study. Underlines indicate algal species used in RNA-Seq
1139 sequencing. **(C)** An UpSet plot of plastid gene composition. Species abbreviations are defined in Figure 3–
1140 source data 1. The horizontal bar chart indicates the gene numbers in each species. The vertical bar chart
1141 indicates the number of genes conserved among the species. Intersect connectors indicate the species
1142 composition in a given number of genes (vertical bar chart). Connections corresponding to no gene were omitted.
1143 Connectors were colored according to the gene's conservation level in Bryopsidales: Core gene, conserved
1144 among all analyzed Bryopsidales species; Dispensable gene, retained more than two Bryopsidales species; Rare
1145 gene, determined from a single or no Bryopsidales species. Gray shading indicates non-Viridiplantae algae, and
1146 magenta shading indicates PoB kleptoplasts. Cyme (*C. merolae*) and Vali (*V. litorea*) had more than 100 genes

1147 that Bryopsidales does not have (e.g., left two vertical bars). **(D)** Box plots of tblastn results. The vertical axis
1148 shows the database searched (kPoro and kRhip, PoB kpDNAs; nCale, nucDNA of *C. lentillifera*). Each dot
1149 represents the tblastn result (query is the A614 dataset). Red dots show the result using the *chlD* gene (encoding
1150 magnesium-chelatase subunit ChlD) as the query sequence; this sequence is similar to the kleptoplast-encoded
1151 *chlL* gene. The right pie chart shows the proportion of queries with hits (*E*-value < 0.0001). **(E)** Heat map of
1152 tblastn results of representative photosynthetic nuclear genes (a subset of data in Figure 3D). The source species
1153 of the query sequences are described on the top. The source files of plastid gene composition and tblastn
1154 analysis are available in Figure 3–source data 1 and 2.

1155

1156 **Figure 4. Search for horizontally transferred algal genes in the PoB genome.**

1157 **(A)** Heatmap of the number of genes assigned to photosynthesis- or plastid-related GO terms. The abundance of
1158 photosynthesis- or plastid-related genes was compared among PoB, two non-kleptoplastic mollusk species (Aca,
1159 *A. californica*; Bgl, *B. glabrata*), and five algal species (see abbreviations in Figure 3 and Figure 3–source data
1160 1). A color scale was visualized in the box on the left. **(B)** Scatter plot of the Exonerate results, the alignment of
1161 the A614 gene set (query sequences) to the PoB genome. The enlarged view with the tblastn results; Figure 4–
1162 Figure Supplement 5. The dot color shows the source algae of each query sequence (see keys in the box). The
1163 horizontal axis shows the percentage of the query length aligned to the hit sequences (PoB genome). The
1164 vertical axis shows the similarity of the aligned sequences between the query and PoB genome; alignment score
1165 (the sum of the substitution matrix scores and the gap penalties) divided by aligned length (bp). Dashed lines are
1166 thresholds for a credible query hit (i.e., a hit covering >60% of the query sequence and a normalized Exonerate
1167 alignment score of >2). **(C)** Scatter plot of MMseq2 results for the A614 dataset (algal photosynthetic genes,
1168 red) and P911 reference dataset (PoB single-copy genes, blue) used as query sequences against the database of
1169 pre-assembled read sets from paired-end DNA libraries of PoB. The top panel shows the probability density
1170 distribution of the number of hit-reads (normalized with TPM: transcripts per kilobase million; *x*-axis) versus the
1171 averaged "pident" value (percentage of identical matches) from the hit reads (*y*-axis). **(D)** Scatter plot of HGT
1172 indices (hA-index versus h-index) for genes in PoB, the two non-kleptoplastic mollusk species (Aca and Bgl),
1173 and one algae species (*C. lentillifera* [Cale]). Each dot represents a gene. A high hA-index or h-index value
1174 means the possibility of algal or prokaryote origin, respectively. Dashed red lines represent the conventional

1175 threshold for HGT (−100 for h-index and 100 for hA-index). Three red arrowheads indicate PoB genes
1176 exceeding the thresholds. **(E)** Heatmap of the results of searches for algae-like RNA fragments in the PoB RNA-
1177 Seq data. Pa, parapodium; EG, egg; He, head; Pe, pericardium. The blue gradient indicates the number of RNA-
1178 Seq reads assigned as algae-like fragments (see key). The *y*-axis labels show the RNA-Seq library name and
1179 analyzed tissue types. The *x*-axis labels indicate the query protein; the queries having no corresponding reads
1180 were omitted from the figure. For queries using the P911 reference dataset, the mean of the hit-read counts from
1181 each library was described. The total number of reads for each library is given on the far right. The source files
1182 of homology-based algal gene searching analysis using Exonerate, MMseq2, and blastp are available in raw data
1183 Figure 4–source data 1.

1184

1185 **Figure 5. Probable KRM genes in PoB.**

1186 **(A)** Volcano plot comparing gene expressions in DG and DeP tissues of PoB. Red, orthogroup OG0000132
1187 (cathepsin D-like protease genes); blue, orthogroup OG0000005 (lectin-like genes); orange, other orthogroups
1188 (for details, see Figure 5–Figure Supplement 2). Up and down arrows signify up- and down-regulation,
1189 respectively, in DG. **(B)** Orthogroups expanded on the *P. ocellatus* lineage and containing DG-up-regulated
1190 genes. Ema, *E. marginata*; Bgl, *B. glabrata*; Lgi, *Lottia gigantea*; Hdi, *Haliotis discus*; Obi, *Octopus*
1191 *bimaculoides*; Pfu, *Pinctada fucata*; Cgi, *Crassostrea gigas*; Cte, *Capitella teleta*; Dme, *D. melanogaster*; Bmo,
1192 *Bombyx mori*; Api, *A. pisum*; Hsa, *Homo sapiens*; Mmu. *Mus musculus.* The phylogenetic tree was scaled to
1193 divergence time based on 30 conserved single-copy genes. Mya, million years. The numbers of rapidly
1194 expanded (blue) and contracted (magenta) orthogroups on the lineages to PoB are provided at the nodes (for
1195 details, see Figure 5–Figure Supplement 4 and 5). Below the tree is information for the six expanded
1196 orthogroups that contained DG-up-regulated genes. The left-side heatmap shows the gene numbers (number in
1197 boxes) and *Z*-score of gene numbers (color gradient) for each orthogroup. The table shows the expanded/not
1198 expanded status of each orthogroup (P–E, *Plakobranchus-Elysia* node; PoB, *Plakobranchus* node). The right-
1199 side heat map indicates the number of DEGs between DG and DeP tissue in each orthogroup. Representative
1200 gene products are given on the far right. The source files of RNA-Seq analysis and comparative genomic
1201 analysis are available in raw data Figure 5–source data 1 and 2, respectively.

1202 **Figure 6. Evolutional process of KRM gene candidates of OG0000132**

1203 **(A)** ML phylogenetic tree for OG0000132 genes (for details, see Figure 6–Figure Supplement 1). Red circle,
1204 PoB gene; blue square, *E. marginata* (Ema) gene. Genes mentioned in the text are indicated with their IDs
1205 beside the circles. Pictograms represent non-sacoglossan species defined in Figure 5–source data 2. **(B)**
1206 Expression of each PoB gene and gene position on the genomic assemblies. The left-side heat plot indicates
1207 normalized expression degree (FPKM) of the PoB genes in various tissues; tissue abbreviations are defined in
1208 Figure 4 and Figure 4–Figure Supplement 10. The vertical gene order corresponds with the position in Figure
1209 6A. The white box with a diagonal line means a non-PoB gene (no expression data is available). Tissue samples
1210 for the FDR calculation (DG versus DeP) were enclosed with square brackets. Color scales for FPKM and FDR
1211 are visualized on the right-side boxes. The FDR's black color indicates that the value could not be calculated
1212 (NA) due to its undetectable gene expression. The right panel shows the genomic positions of the genes. The
1213 bands' color indicates the correlation between the gene and position, purple text and arrows indicate the scaffold
1214 ID and direction, and colored boxes on the scaffolds mean the genes' positions (magenta, OG0000132 gene;
1215 gray box, other protein-encoding genes). Scaffolds having less than five OG0000132 genes were omitted from
1216 the figure.

1217

1218 List of source data

1219 **Figure 1–source data 1**

1220 **Sequence ID list used for phylogenetic analysis on Figure 1B**

1221 **Figure 2–source data 1**

1222 **Summarized data of Fv/Fm and oxygen generation activity analysis of *P. ocellatus* used for Figure 2**

1223 **Figure 2–source data 2**

1224 **Raw data of oxygen concentration dynamics**

1225 This zip archive contains raw data of oxygen concentration dynamics. The measurement data for each
1226 time series are stored in separate files, with the experiment ID (ID1–6) and types of the experiment (mock
1227 or *P. ocellatus* sample) indicated in the name.

1228 **Figure 2–source data 3**

1229 **Summary of the longevity of analyzed *P. ocellatus* individuals used for Figure 2D**

1230 **Figure 3–source data 1**

1231 **Plastid gene composition**

1232 This zip archive contains source files of plastid gene composition and of orthologous analysis.

1233 **Figure 3–source data 2**

1234 **Tblastn analysis of kleptoplast/chloroplast DNA**

1235 This zip archive contains source and result files of tblastn analysis of kleptoplast/chloroplast DNA.

1236 **Figure 4–source data 1**

1237 **Algal gene detection from PoB genome**

1238 This zip archive contains source files of algal gene searching from PoB genome and pre-assembled reads.

1239 **Figure 5–source data 1**

1240 **KRM gene searching using RNA-Seq**

1241 This zip archive contains source files of kleptoplasty relating molluscan (KRM) genes searching based on

1242 RNA-Seq analysis.

1243 **Figure 5–source data 2**

1244 **KRM gene searching using orthologues analysis**

1245 This zip archive contains source files of KRM gene searching based on the comparative genomic

1246 (orthologous) analysis.

1247

1248

1249 Legends for supplementary figures

1250 **Figure 2–figure supplement 1**

1251 **Light-dependent oxygen generation by *P. ocellatus***

1252 **(A)** Time-course changes of dissolved oxygen in the sea-water in the vial containing the sea-slug. Each

1253 plot indicates the data of a single PoB individual. Raw data are available in Figure 2–source data 2. d38

1254 group, three PoBs starved for 38 days; d109 group, three PoBs starved for 109 days. **(B–E)** Oxygen

1255 measurement system. **(B)** Overview of the dissolved oxygen measurement system. C, iron chilling block;

1256 Mw, mixer for lagging water; Mv, mixer for sea slug rearing seawater; V, vial caging a sea slug; W,

1257 plastic tray filled with lagging water for the vial. **(C)** Horizontal view of the glass vial. O, oxygen-
1258 sensitive optode; P, rubber-made partition plate to separate sea slug from optode and rotor; R, rotor for
1259 seawater agitation. **(D)** Top view of the vial. H, holes for water ventilation. **(E)** Top vial view with a sea
1260 slug. G, glass plate sealing the top of the vial. **(F)** Diagram of PAM measurement method. Horizontal
1261 view of the sea slug in a well of the cell culture plate and positions of a sponge and a light probe. The sea-
1262 slug was upside-down facing the dorsal side to the PAM light probe.
1263

1264 **Figure 3–figure supplement 1**

1265 **Gene map of a *P. ocellatus* kleptoplast DNA sequestered from *Rhipidosiphon lewmanomontiae***

1266 The outer circle shows the distribution of all genes on kpDNA. As shown in the box, colors indicate
1267 functional categories of the genes, respectively. Circles on different colored backgrounds show the
1268 names and positions of genes related to photosynthesis (2nd circle), translation (3rd), transcription (4th),
1269 other conserved chloroplast genes (5th), genes relating to intron-splicing and transposition (6th), and
1270 bacteria-like genes (7th), respectively. Genes on the outer-side of each circle are transcribed in the
1271 clockwise direction; genes on the inner-side are transcribed in the anticlockwise direction. Hatched boxes
1272 of genes on the circles are introns.
1273

1274 **Figure 3–figure supplement 2**

1275 **Gene map of a *P. ocellatus* kleptoplast DNA sequestered from *Poropsis* spp.**

1276 Circles and colors are the same as those in Figure 3–figure supplement 1.
1277

1278 **Figure 3–figure supplement 3**

1279 **Mitochondrial gene map of the *P. ocellatus* type black**

1280 **(A)** The outermost circle gives the scale in base pairs. Inner circles show the names and positions of
1281 protein-encoding genes (2nd circle) and structure RNA encoding genes (3rd). **(B)** Sequence similarity
1282 between our *P. ocellatus* type black mitochondrial DNA and previously sequenced ones from *P.* cf.
1283 *ocellatus* HW-2015 (KX853083). Genes on outside of each circle are transcribed clockwise and those on
1284 inside, anticlockwise. We used mVISTA software to calculate and visualize the similarity.

1285

1286 **Figure 3–figure supplement 4**

1287 **Bryopsidales ML tree made by RAxML (a -x 12345 -p 12345 -# 300 -m GTRGAMMAI) based on**

1288 **457 positions of *rbcL* genes**

1289 Bootstrap values are described at each node. We omitted the values less than 50. Scale bar represents the

1290 branch length. The color and underlines at the OTUs have the same meanings as in Figure 3B.

1291

1292 **Figure 3–figure supplement 5**

1293 **Gene map of the *H. borneensis* chloroplast DNA**

1294 Circles and colors are the same as those in Figure 3–figure supplement 1.

1295

1296 **Figure 3–figure supplement 6**

1297 **Gene compositions of 18 algal cpDNAs and *P. ocellatus* kpDNAs**

1298 The presence (cyan) and absence (magenta) of the genes were visualized with the color tiles. The abscissa

1299 -axis shows the species, and the vertical-axis shows the gene. kPoro, PoB kleptoplast derived from

1300 *Poropsis* sp.; kRhip, PoB kleptoplast derived from *R. lewmanomontiae*. Other abbreviations were

1301 explained in Figure 3–source data 1. Top tree shows the ML tree made by IQ-tree based on 51 conserved

1302 plastid genes (*ycf4, ycf3, tufa, rps9, rps8, rps7, rps4, rps3, rps2, rps19, rps18, rps14, rps12, rps11, rpoa,*

1303 *rpl5, rpl36, rpl23, rpl20, rpl2, rpl16, rpl14, rbcL, psbZ, psbT, psbN, psbL, psbK, psbJ, psbI, psbH, psbF,*

1304 *psbE, psbD, psbC, psbB, psbA, psaJ, psaC, psaB, psaA, petG, petD, petB, petA, atpI, atpH, atpF, atpE,*

1305 *atpB, atpA*, 9882 amino acid positions). Left side genes were clustered base on the GO annotation of each

1306 gene. GO annotations were described on the right-side boxes. Source files of gene compositions and

1307 annotations are available in Figure 3–source data 1.

1308

1309 **Figure 3–figure supplement 7**

1310 **Read coverage depth of the constructed *P. ocellatus* genomic assemblies**

1311 We mapped the two libraries to the constructed assemblies to determine the origin of each assembly. The

1312 "Poc_DNA_Pa_180" is a kleptoplast-poor sample (Parapodium). The "Poc_DNA_Di_180" is a

1313 kleptoplast-enriched sample (Digestive gland). Considerable change of depth on AP014543 and

1314 AP014542 reflects the different origin of the assemblies from the rest (mitochondria- or nuclear-derived
1315 assemblies).
1316
1317 **Figure 3–figure supplement 8**
1318 **Annotation procedure of *H. borneensis* transcripts**
1319 **(A)** *De novo* transcripts assembling process. **(B)** Annotation procedure of the constructed transcripts.
1320 Dark blue, submitted *H. borneensis* transcripts for DDBJ; light red, reference data; aquamarine,
1321 annotation method (for detail, see material and methods); yellow, annotated gene number for each step.
1322 **(C–D)** Upset plots of annotated gene number for each annotation step. The source file of the orthologous
1323 relationship of the kleptoplast/chloroplast-encoded genes is available in Figure 3–source data 1. We
1324 deposited detailed annotation data of *H. borneensis* transcripts in FigShare (DOI
1325 10.6084/m9.figshare.13042016).
1326
1327 **Figure 3–figure supplement 9**
1328 **Annotation procedure of *B. hypnoides* transcripts**
1329 Colors and boxes are the same as those in Figure 3–figure supplement 8. Detailed annotation data were
1330 deposited in FigShare (DOI 10.6084/m9.figshare.13042016).
1331
1332 **Figure 3–figure supplement 10**
1333 **Annotation procedure of *C. lentillifera***
1334 Colors and boxes are the same as those in Figure 3–figure supplement 8. Detailed annotation data were
1335 deposited in FigShare (DOI 10.6084/m9.figshare.13042016).
1336
1337 **Figure 4–figure supplement 1**
1338 **Genome size estimation of *P. ocellatus* type black by flow cytometry**
1339 **(A)** Histograms showing the fluorescence intensities of the analyzed nuclei (*x*-axis) and count of the
1340 nuclei (*y*-axis). Studied species were represented with pictograms and abbreviations as in the left bottom
1341 box. Each nuclear suspension was analyzed for each species to confirm that the peak (particle) was free
1342 from contamination (Upper panels). On lower panels, we mixed the suspensions of three species to

estimate the PoB genome size. The mixed ratio of the suspensions is shown on the upper-side of each histogram. We changed the mixing ratio to identify the respective peaks derived from the species. **(B)** box-plots of genomic size of *P. ocellatus* estimated from the mixed flowcytometry with other genome size standards. The black dots are the values of PoB genome size estimated from triplicate experiments (runs). Red diamond, an averaged genome size estimated of the experimental triplicates (estimated genome size for each biological triplicate). A blue star means an averaged genome size of the biological triplicates. We made biological triplicates from different individuals of each species.

Figure 4–figure supplement 2

Genome assembling and gene modeling approach for the *P. ocellatus* genome analysis

Light orange, raw read data; light blue, intermediate assembly data; dark blue, a final assembly for DDBJ submission; black, eliminated assemblies. Information on the raw read data are available in Supplementary file 1.

Figure 4–figure supplement 3

Annotation procedure for the genomic gene models of *P. ocellatus* type black

Dark blue, *P. ocellatus* genome assembly for DDBJ submission; light red, reference data; aquamarine, annotation method (for detail, see material and methods); yellow, annotated *P. ocellatus* gene number. Other colors and boxes are the same as those in Figure 3–figure supplement 8.

Figure 4–figure supplement 4

ML tree for the candidate HGT genes with algal and molluscan homologous genes

The abbreviations represent the species from which the genes are derived (see Supplementary file 19, Figure 5–source data 1). The full gene ID (INSDC ID, OIST Marine Genomics Unit database ID, and our annotation ID) were noted in the boxes. Green, algal gene; magenta, *P. ocellatus* gene; black, other molluscan genes. Our GO-analysis found six PoB genes assigned with the child terms of "Plastid" (GO:0009536). In these PoB genes, p2972c65.3 and p503c65.126 (annotated as "Chloroplast envelope" GO:0009941) made orthogroups "4058" and "2017" with several algal genes by the SonicParanoid analysis, respectively. Rest four genes (gene ID: p197c68.18, p234c64.89, p45387c41.1, and

1372 p466c59.83) showed no orthologous relationships with algal gene and were omitted from the
1373 phylogenetic analysis. **(A)** ML tree for the orthologous group "4058". **(B)** ML tree for the orthologous
1374 group "2017". The p2972c65.31 and p503c65.126 formed a monophyletic clade with molluscan orthologs
1375 rather than algal gene, indicating this gene is of molluscan origin (no HGT). Source files are available in
1376 Figure 4–source data 1.
1377
1378 **Figure 4–figure supplement 5**
1379 **Summary of Exonerate and tblastn results**
1380 Scatter plots of the analyzed results by Exonerate **(A)** and tblasn **(B)**. Adopted databases were *P. ocellatus*
1381 genome (top plate), *C. lentillifera* genome (middle plate), or kpDNA-derived scaffolds (bottom plate).
1382 The query sequence was the algal A614 gene set. The top box shows the source algae of each query.
1383 Query hit statistics were described on the left shoulder of each plate. **(A)** Dashed lines are thresholds for
1384 the credible query hit (Covering over 60% of the query sequence and over two normalized aligned scores
1385 for aligned length). All seven hit queries on kpDNAs were algal *chlD* gene (encoding Magnesium-
1386 chelatase subunit ChlD) resembling the kleptoplast-encoded *chlL* gene. **(B)** Dashed lines are thresholds
1387 (bit score = 100) for the credible query hit. All seven significant hit queries on the kpDNA-derived
1388 scaffolds were algal *chlD* gene (encoding Magnesium-chelatase subunit chlD) resembling the kleptoplast-
1389 encoded *chlL* gene (see Supplementary file 20 and Figure 4–source data 1).
1390
1391 **Figure 4–figure supplement 6**
1392 **ML tree for the orthologous group "65" by SonicParanoid**
1393 Colors are the same as those in Figure 4–figure supplement 4. Source files are available in Figure 4–
1394 source data 1.
1395
1396 **Figure 4–figure supplement 7**
1397 **ML tree for the orthologous group "501" by SonicParanoid**
1398 Colors are the same as those in Figure 4–figure supplement 4. Source files are available in Figure 4–
1399 source data 1.
1400

1401 **Figure 4–figure supplement 8**

1402 **ML tree for the orthologous group "456" by SonicParanoid**

1403 Colors are the same as those in Figure 4–figure supplement 4. Source files are available in Figure 4–
1404 source data 1.

1405

1406 **Figure 4–figure supplement 9**

1407 **Sequence similarity between g566.t1 (*C. lentillifera*) and p310c70.15 (*P. ocellatus*)**

1408 (A) Sequence alignment and similarity. (B) Venn's plot of number of hit reads by MMseq2 analysis. The
1409 adopted database is pre-assembling Illumina DNA sequencing read (DRR029525 and DRR029526).

1410

1411 **Figure 4–figure supplement 10**

1412 **Overview of the sample preparation for *P. ocellatus* type black transcriptomic analysis**

1413 **(A–C)** Photographs of the PoB dissection and the positions of sample tissues. **(A)** Dorsal views of a fresh
1414 PoB individual with spread (or inside-out) parapodia. **(B)** The upper plate shows an enlarged view of the
1415 tissue dissected from the parapodium (-PA sample). The plate on the lower left shows the parapodium,
1416 excluding the digestive gland part, collected in a plastic tube (-DeP sample). The lower-right plate shows
1417 the digestive gland cut out from the parapodium using a razor (-DG sample). **(C)** Egg mass of PoB. **(D)**
1418 Incubation states of the dissected sea slugs and the relationships among the samples, libraries, and RNA-
1419 seq analysis. "Mean coverageData" means the averaged depth of coverage of each library to the PoB
1420 genome. The libraries from three individuals (T1–3) were used for DEG calling using the Wald test.

1421

1422 **Figure 4–figure supplement 11**

1423 **Agarose gel electrophoresis image displaying extracted *P. ocellatus* DNA for the nuclear genome**
1424 **sequencing with two fragment size markers**

1425 The extracted total DNA from *P. ocellatus* was mixed with gel loading buffer, electrophoresed at 50V in
1426 0.7 % (w/v) agarose gel in TAE buffer, and stained by SYBR safe (Thermo).

1427

1428 **Figure 4–figure supplement 12**

1429 **Read depth of the *P. ocellatus* genome assemblies**

1430 To determine the scaffolds from kleptoplasts and mitochondria, we mapped the Illumina libraries of *P.*
1431 *ocellatus* (DRR063261 and DRR063263, Supplementary file 1) to the genomic reference data and
1432 compared the coverage depth. We used two types of *P. ocellatus* DNA (DG = digestive gland, PA =
1433 parapodium) to construct the Illumina libraries. The genomic reference data contained 13 kleptoplastic
1434 (kp), one mitochondrial (mt), and 8647 nuclear (nuc) scaffolds. **(A)** Violin and scatter plots of the mapped
1435 read depths. Each dot indicates a scaffold. A horizontal bar indicates the averaged depth values of
1436 coverage. The right boxes show the legend of coloration. **(B)** Line graphs of the relative values of the read
1437 depths of the kleptoplastic scaffolds against the averaged depth of all nuclear scaffolds in libraries.
1438 Kleptoplastic scaffold IDs were described on the tops of respective graphs.
1439
1440 **Figure 4–figure supplement 13**
1441 **Dot plot of the referential kleptoplast DNAs and PoB genomic scaffolds determined as the**
1442 **kleptoplast DNA (kpDNA)**
1443 Abscissa-axis, referential kleptoplast DNAs (kPoro and kRhip); horizontal-axis, 13 scaffolds selected as
1444 kpDNA sequences. The diagonals indicate contiguously aligned regions. The coloration means the
1445 identity of the aligned regions. The top left box indicates the legend of similarity coloration. Arrows
1446 indicate the aligned sequences and direction. We used D-GENIES software to make the dot-plots. Both
1447 referential kpDNAs indicated high similarity with the 11 scaffolds, supporting kpDNA-origin of the 11 of
1448 13 scaffolds. For the rest two scaffolds, "scaffold81938_cov17" and "scaffold268313_cov495" indicated
1449 a similarity to kPoro and kRhip, respectively. This discrepancy may be due to the sequence diversity of
1450 plastidial DNA among the species (donor algae). Overall, these results support that the sequence of 13
1451 scaffolds are not derived from anything other than plastidial (kleptoplast) DNA.
1452
1453 **Figure 4–figure supplement 14**
1454 **Dot plot of the two referential kleptoplast DNAs**
1455 Arrows and colors are the same as those in Figure 4–figure supplement 13.
1456
1457 **Figure 4–figure supplement 15**
1458 **Dot plot of the referential mtDNAs and one scaffold determined as mtDNA sequence**

1459 Arrows and colors are the same as those in Figure 4–figure supplement 13.

1460

1461 **Figure 5–figure supplement 1**

1462 **DG-upregulated genes by the RNA-seq comparison with -DeP samples**

1463 Differentially expressed gene calling among the libraries from DG (digestive gland) and DeG (DG-
1464 exenterated parapodium) found 162 significantly DG-upregulated genes. The expression levels and
1465 annotation information were summarized in Figure 5–source data 1. Fold changes (FC) and false
1466 discovery rate (FDR) were calculated between the DG and DeP samples and visualized after logarithmic
1467 transformation.

1468

1469 **Figure 5–figure supplement 2**

1470 **Volcano plot of the cross-tissue comparison between DG and DeP samples**

1471 Each dot represents a gene. Colors and dot shapes mean the belonging orthogroups of genes. Source files
1472 are available in Figure 5–source data 1

1473

1474 **Figure 5–figure supplement 3**

1475 **The procedure of gene family history analyses**

1476 The square box and the circle mean the adopted analysis and the obtained orthogroup (OG) datasets,
1477 respectively.

1478

1479 **Figure 5–figure supplement 4**

1480 **CAFE-based all rapidly** expanding/contracting **genes on the sacoglossan linage**

1481 The gene count of all rapidly expanding/contracting orthogroups on *P. ocellatus* (left side) and the
1482 common node between *P. ocellatus* and *E. marginata* (right side) were plotted with a heat map. The left
1483 top tree is an ultrametric tree made by r8s based on 30 genes. The red color means the number of genes in
1484 each species. The blue color indicates the number of differentially expressed genes between DG and DeP
1485 tissue. The up and down arrows mean up- and down-regulation on the DG sample, respectively. Source
1486 files are available in Figure 5–source data 1 and 2

1487

1488 **Figure 5–figure supplement 5**

1489 ***Z*-score-based rapidly expanding/contracting genes on the *P. ocellatus* linage (Threshold: Z-score**

1490 **>2)**

1491 Color legends were described in the top right box.

1492

1493 **Figure 6–figure supplement 1**

1494 **Enlarged view of Gene tree and annotation data for all member of OG0000132 (Cathepsin D like)**

1495 From left, gene tree, domain structure, predicted intracellular location with DeepLoc-1.0, expression

1496 profiles in cross-tissue RNA-Seq, and the gene position on *P. ocellatus* genome. Color legends are shown

1497 in the left box.

1498

1499 **Figure 6–figure supplement 2**

1500 **Positions of genes belonging to OG0000132 on the *E. marginata* genome**

1501 Genes on the upperside are transcribed from left to right; genes on the bottomside are transcribed from

1502 right to left. Magenta, OG0000132 genes; gray, other genes. †Clade-γ gene.

1503

1504 **Figure 6–figure supplement 3**

1505 **Gene tree and annotation data for all genes of OG0000446 (Apolipoprotein D-like)**

1506 From left, gene tree, domain structure, predicted intracellular location with DeepLoc-1.0, expression

1507 profiles in cross-tissue RNA-Seq, and the gene position on *P. ocellatus* genome. Color legends are shown

1508 in the bottom box.

1509

1510 **Figure 6–figure supplement 4**

1511 **Gene positions of OG0000446 on *E. marginata* genome**

1512 Genes on the rightside are transcribed from top to bottom; genes on the leftside are transcribed from

1513 bottom to top. Magenta, OG0000446 genes; gray, other genes.

1514

1515 **Figure 6–figure supplement 5**

1516 **Gene tree for the member of OG0000005 (Lectin-like)**

The black dot on the node indicates the Bootstrap support over 80. One-by-one ortholog pairs are masked with gray color.

List of supplementary files

Supplementary file 1

Summary of the raw sequencing data

Supplementary file 2

Assembly statistics of de novo RNA-seq of donor algae

Supplementary file 3

Algal genes used for referential query (A614 dataset)

Supplementary file 4

Assembly statistics of *Plakobranchus ocellatus* and *Elysia marginata* genome

Supplementary file 5

PoB genes having algal top-hit reslts of blastp

This zip archive contains a file for statistics of PoB genes having algal top-hit reslts of blastp analysis and a file for detailed annotation of the algal top-hit PoB genes.

Supplementary file 6

Detailed result of MEGAN analysis

Supplementary file 7

Annotation data of the six PoB genes assigned with the child terms of "Plastid(GO:0009536)"

Supplementary file 8

Applied species for HGT index analysis

Supplementary file 9

Gene annotation of the DG-upregulated PoB genes

Supplementary file 10

Enriched GOs of the significantly upregulated genes on the digestive gland of PoB

Supplementary file 11

1544 **Cross-tabulation table for the significant enrichment analysis of DG-upregulated gene number on**
1545 **the PoB-expanded orthogroup orthologous group**
1546 **Supplementary file 12**
1547 **Summary of the results of dN/dS-based selection test**
1548 **Supplementary file 13**
1549 ***E. marginata* scaffolds determined as bacterial contaminants**
1550 **Supplementary file 14**
1551 ***E. marginata* scaffolds selected as kleptoplast/mitochondrial DNA**
1552 **Supplementary file 15**
1553 **OrthoFinder result with additional data from *E. chlorotica* gene set (gene counts)**
1554 **Supplementary file 16**
1555 **Composition information of DNA extraction solutions for *P. ocellatus* genomic DNA**
1556 **Supplementary file 17**
1557 **Scaffolds determined as bacterial contaminants during the PoB genome assmbling and the result**
1558 **of MEGAN analysis before the removing of the bacterial scaffolds.**
1559 **Supplementary file 18**
1560 **Scaffolds determined as kleptoplast/mitochondrion sequences**
1561 **Supplementary file 19**
1562 **Analyzed protein sequences by SonicParanoid analysis**
1563 **Supplementary file 20**
1564 **Result of the Exonerate search**
1565 Database, the 13 PoB-derived scaffolds determined as kpDNA; query, algal photosynthetic genes
1566 [A614])
1567 **Supplementary file 21**
1568 **Sequence similarity among the genes belonging OG0000005, OG0000132, and OG0000446**
1569
1570
1571

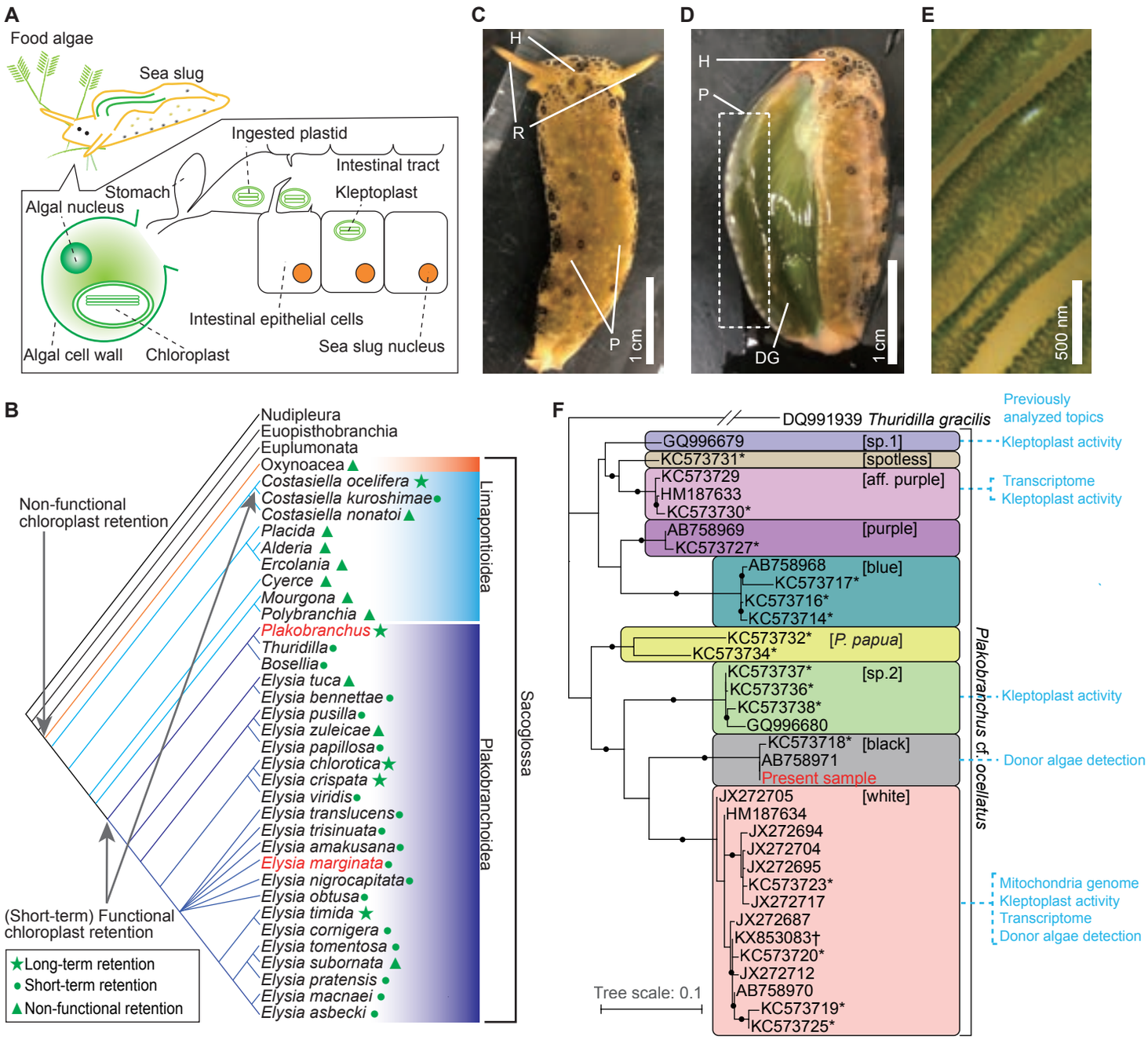
A
Food algae
Sea slug
Ingested plastid
Intestinal tract
Stomach
Kleptoplast
Algal nucleus
Intestinal epithelial cells
Algal cell wall
Chloroplast
Sea slug nucleus
C
H
R
P
1 cm
D
H
P
DG
1 cm
E
500 nm
B
Nudipleura
Euopisthobranchia
Euplumonata
Oxynoacea
Costasiella ocelifera
Costasiella kuroshimae
Costasiella 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
Previously analyzed topics
GQ996679 [sp.1]
Kleptoplast activity
KC573731* [spotless]
KC573729
HM187633
KC573730*
[aff. purple]
Transcriptome
Kleptoplast activity
AB758969
KC573727*
[purple]
AB758968
KC573717*
KC573716*
KC573714*
[blue]
KC573732*
KC573734*
[P. papua]
KC573737*
KC573736*
KC573738*
GQ996680
[sp.2]
Kleptoplast activity
KC573718*
AB758971
Present sample
[black]
Donor algae detection
JX272705
HM187634
JX272694
JX272704
JX272695
KC573723*
JX272717
JX272687
KX853083†
KC573720*
JX272712
AB758970
KC573719*
KC573725*
[white]
Mitochondria genome
Kleptoplast activity
Transcriptome
Donor algae detection
Plakobranchus cf. ocellatus
Tree scale: 0.1

Figure 1

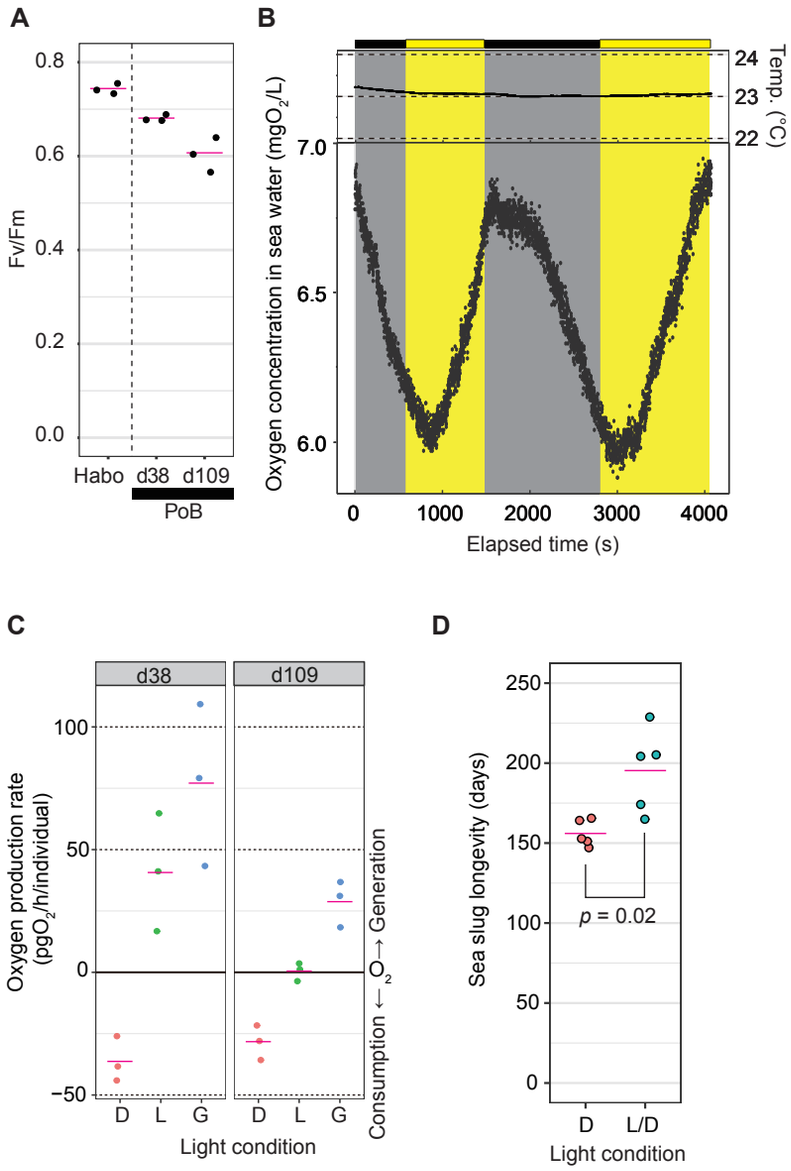
A
Fv/Fm
0.8
0.6
0.4
0.2
0.0
Habo
d38
d109
PoB
B
Oxygen concentration in sea water (mgO2/L)
7.0
6.5
6.0
Temp. (°C)
24
23
22
0
1000
2000
3000
4000
Elapsed time (s)
C
d38
d109
Oxygen production rate (pgO2/h/individual)
100
50
0
−50
D
L
G
Light condition
Consumption ← O2 → Generation
D
Sea slug longevity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

Figure 2

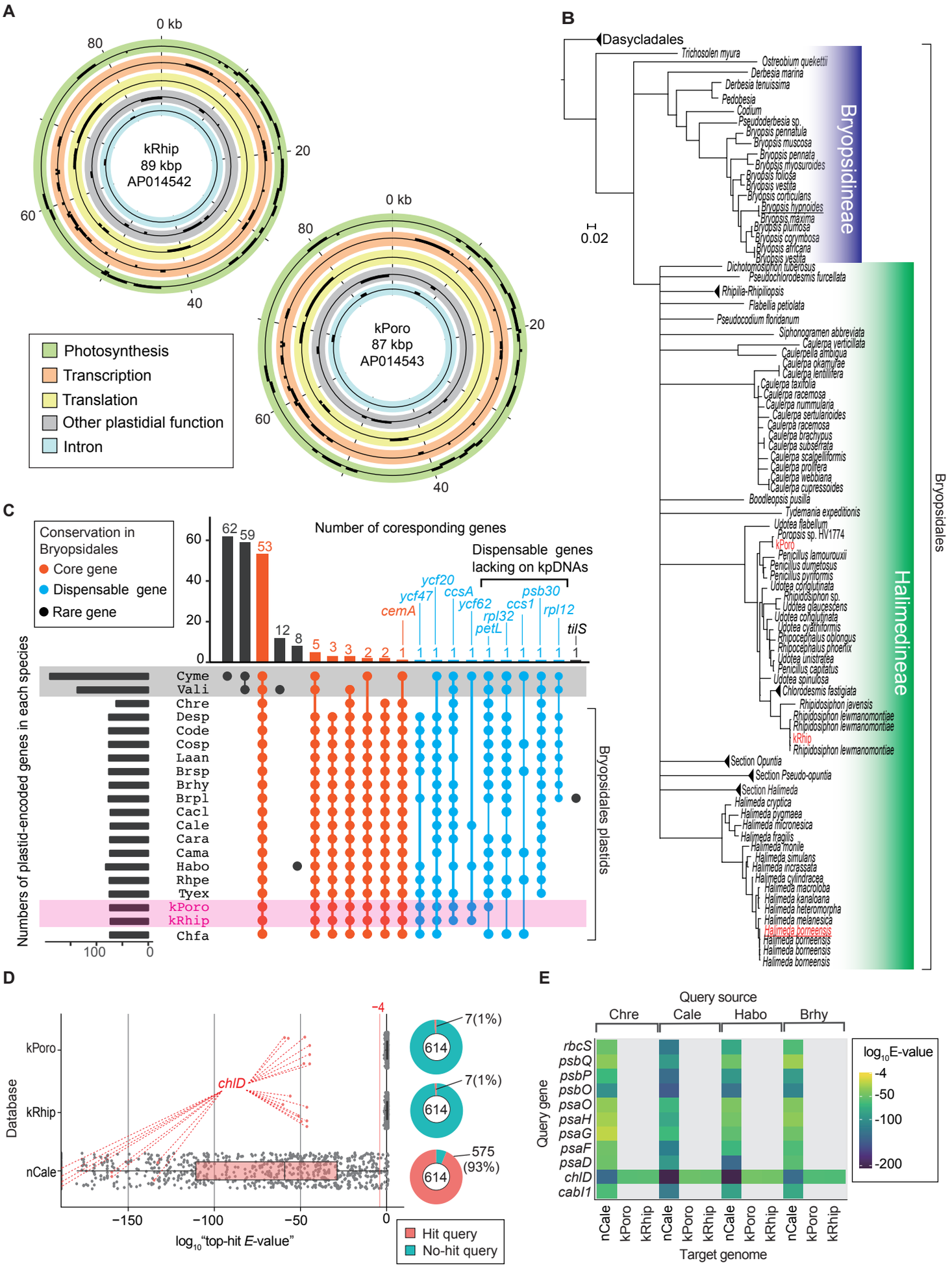
A
0 kb
80
20
60
40
kRhip
89 kbp
AP014542
kPoro
87 kbp
AP014543
Photosynthesis
Transcription
Translation
Other plastidial function
Intron
B
Dasycladales
Trichosolen myura
Bryopsidineae
Halimedineae
Bryopsidales
0.02
C
Conservation in Bryopsidales
Core gene
Dispensable gene
Rare gene
Number of coresponding genes
Dispensable genes lacking on kpDNAs
cemA
ycf47
ycf20
ccsA
ycf62
petL
rpl32
ccs1
psb30
rpl12
tilS
Numbers of plastid-encoded genes 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
cabI1
log10E-value
-4
-50
-100
-200
nCale
kPoro
kRhip
Target genome

Figure 3

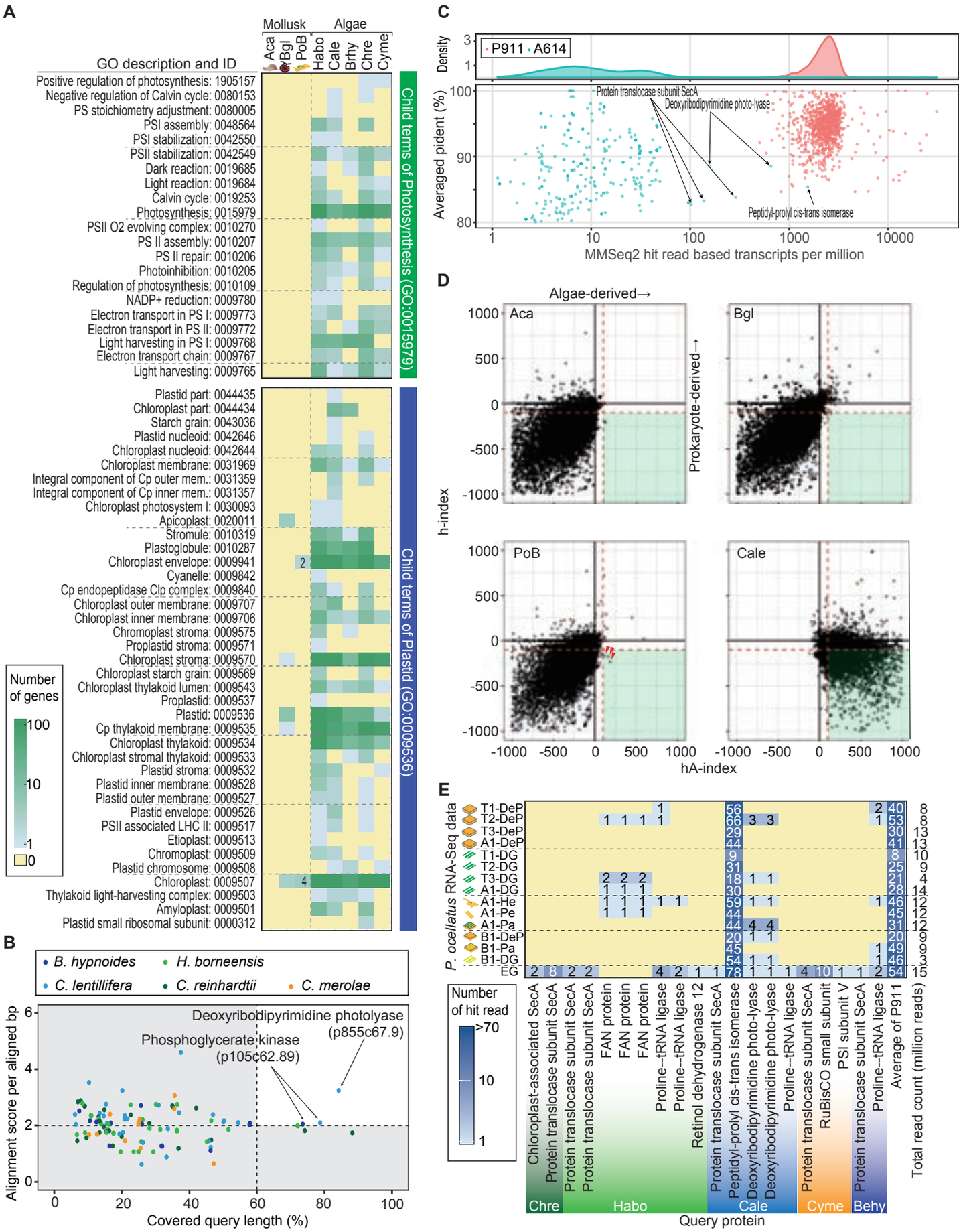
A
Mollusk
Algae
Aca
Bgl
PoB
Habo
Cale
Brhy
Chre
Cyme
GO description and ID
Positive regulation of photosynthesis: 1905157
Negative regulation of Calvin cycle: 0080153
PS stoichiometry adjustment: 0080005
PSI assembly: 0048564
PSI stabilization: 0042550
PSII stabilization: 0042549
Dark reaction: 0019685
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
Averaged pident (%)
Protein translocase subunit SecA
Deoxyribodipyrimidine photo-lyase
Peptidyl-prolyl cis-trans isomerase
MMSeq2 hit read based transcripts per million
D
Algae-derived→
Prokaryote-derived→
Aca
Bgl
PoB
Cale
h-index
hA-index
E
P. ocellatus RNA-Seq 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

Figure 4

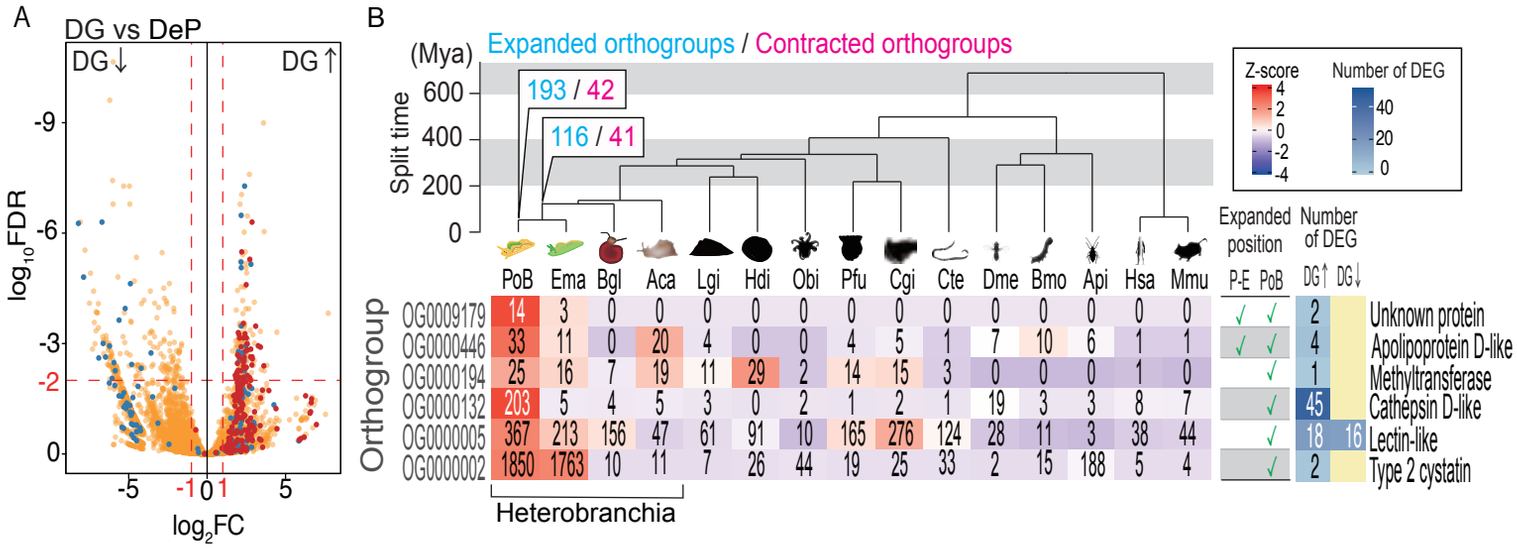
A
DG vs DeP
DG↓
DG↑
-9
-6
-3
-2
0
-5
-1
0
1
5
log10FDR
log2FC
B
(Mya)
Expanded orthogroups / Contracted orthogroups
Split time
600
400
200
0
193 / 42
116 / 41
Z-score
4
2
0
-2
-4
Number of DEG
40
20
0
PoB Ema Bgl Aca Lgi Hdi Obi Pfu Cgi Cte Dme Bmo Api Hsa Mmu
Expanded position
Number of DEG
P-E PoB
DG↑ DG↓
Orthogroup
OG0009179 14 3 0 0 0 0 0 0 0 0 0 0 0 0 0 ✓ ✓ 2 Unknown protein
OG0000446 33 11 0 20 4 0 0 4 5 1 7 10 6 1 1 ✓ ✓ 4 Apolipoprotein D-like
OG0000194 25 16 7 19 11 29 2 14 15 3 0 0 0 1 0 ✓ 1 Methyltransferase
OG0000132 203 5 4 5 3 0 2 1 2 1 19 3 3 8 7 ✓ 45 Cathepsin D-like
OG0000005 367 213 156 47 61 91 10 165 276 124 28 11 3 38 44 ✓ 18 16 Lectin-like
OG0000002 1850 1763 10 11 7 26 44 19 25 33 2 15 188 5 4 ✓ 2 Type 2 cystatin
Heterobranchia

Figure 5

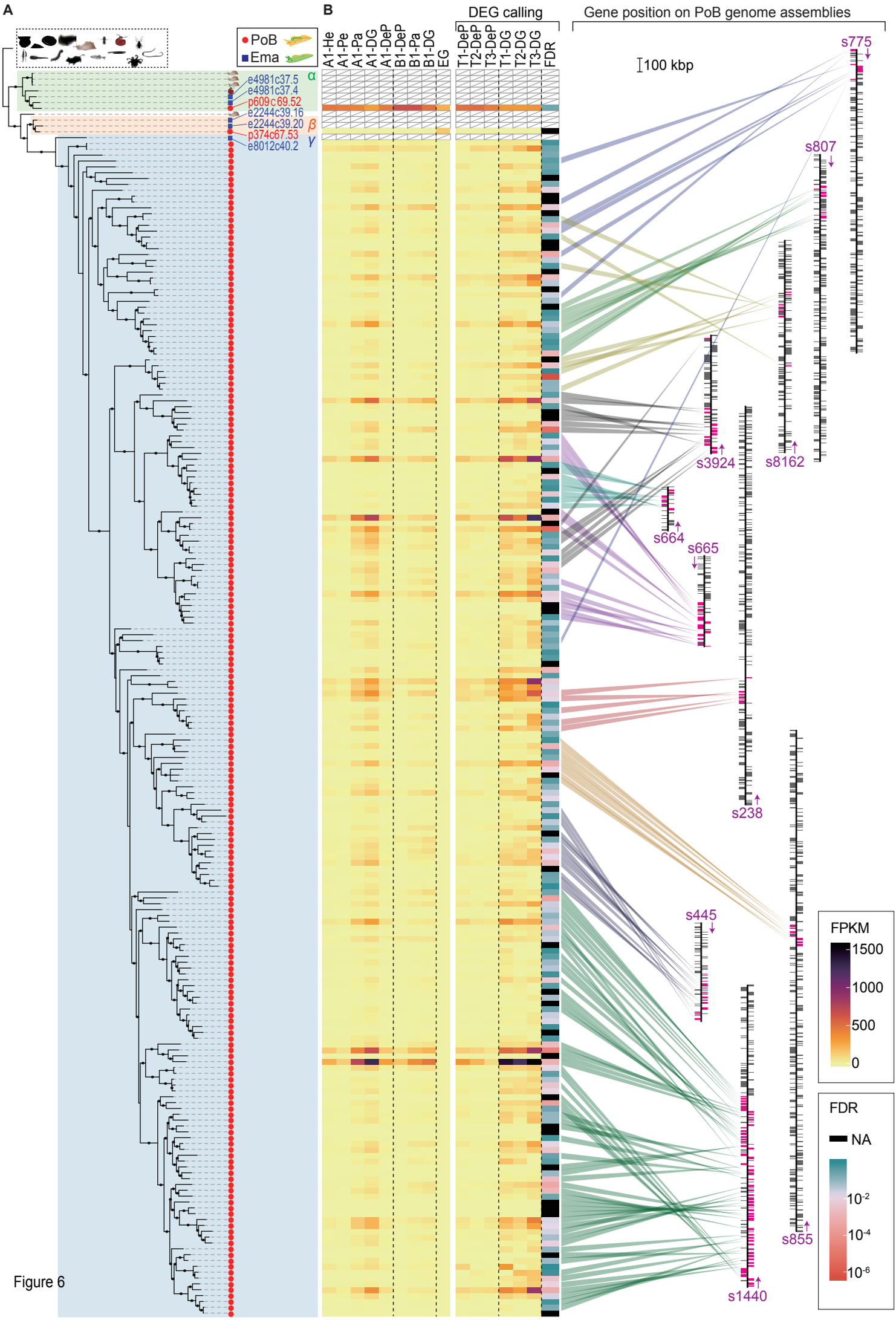
A
B
PoB
Ema
e4981c37.5
e4981c37.4
p609c69.52
e2244c39.16
e2244c39.20
p374c67.53
e8012c40.2
α
β
γ
A1-He
A1-Pe
A1-Pa
A1-DG
A1-DeP
B1-DeP
B1-Pa
B1-DG
EG
DEG calling
T1-DeP
T2-DeP
T3-DeP
T1-DG
T2-DG
T3-DG
FDR
Gene position on PoB genome assemblies
100 kbp
s775
s807
s3924
s8162
s664
s665
s238
s445
s855
s1440
FPKM
1500
1000
500
0
FDR
NA
10-2
10-4
10-6

Figure 6