Horizontal gene transfer of the algal nuclear gene psbO to the photosynthetic sea slug Elysia chlorotica

Mary E. Rumpho1, Jared M. Worful2, Jungho Lee2, Krishna Kannan1, Mary S. Tyler2, Debashish Bhattacharya2, Ahmed Moustafa2,3, and James R. Manharte1

1Department of Biochemistry, Microbiology, and Molecular Biology, University of Maine, Orono, ME 04469; 2Green Plant Institute, Seoul National University, Gwonneun, Suwon, Gyeonggi 441-853, Korea; 3School of Biology and Ecology, University of Maine, Orono, ME 04469; 4Department of Biological Sciences and the Roy J. Carver Center for Comparative Genomics, Interdisciplinary Program in Genetics, University of Iowa, Iowa City, IA 52242-1324; and 5Department of Biology, Texas A&M University, College Station, TX 77843

Edited by Lynn Margulis, University of Massachusetts, Amherst, MA, and approved September 17, 2008 (received for review June 9, 2008)

The sea slug Elysia chlorotica acquires plastids by ingestion of its algal food source Vaucheria litorea. Organelles are sequestered in the mollusc’s digestive epithelium, where they photosynthesize for months in the absence of algal nucleotropism. This is perplexing because plastid metabolism depends on the nuclear genome for >90% of the needed proteins. Two possible explanations for the persistence of photosynthesis in the sea slug are (i) the ability of V. litorea plastids to retain genetic autonomy and/or (ii) more likely, the mollusc provides the essential plastid proteins. Under the latter scenario, genes supporting photosynthesis have been acquired by the animal via horizontal gene transfer and the encoded proteins are retargeted to the plastid. We sequenced the plastid genome and confirmed that it lacks the full complement of genes required for photosynthesis. In support of the second scenario, we demonstrated that a nuclear gene of oxygenic photosynthesis, psbO, is expressed in the sea slug and has integrated into the germline. The source of psbO in the sea slug is V. litorea because this sequence is identical from the predator and prey genomes. Evidence that the transferred gene has integrated into sea slug nuclear DNA comes from the finding of a highly diverged psbO 3’ flanking sequence in the algal and mollusc nuclear homologues and gene absence from the mitochondrial genome of E. chlorotica. We demonstrate that foreign organelle retention generates metabolic novelty (“green animals”) and is explained by anastomosis of distinct branches of the tree of life driven by predation and horizontal gene transfer.

Symbiotic associations and their related gene transfer events are postulated to contribute significantly to evolutionary innovation and biodiversity. This comes from extensive analysis of organelles such as plastids (e.g., chloroplasts) that originated via primary endosymbiosis of a free-living cyanobacterium (1, 2). The cyanobacterial genome was greatly reduced by endosymbiotic gene transfer (EGT) to the host nucleus and wholesale gene loss, giving rise to the primary lineages of plants and green algae (streptophytes and chlorophytes), red algae (rhodophytes), and glaucophytes (3–6) [see the scheme in supporting information (SI) Fig. S1]. The diverse group of secondary or “complex” alga (e.g., chloromaloeates, euglenids), in turn, arose by secondary endosymbiosis—the uptake of a eukaryotic alga (green or red lineage) by a heterotrophic eukaryotic host. In this case, in addition to EGT, transfer of genes between the unrelated organisms by lateral or horizontal gene transfer (HGT) and loss of genes occurred as a result of the “merger” of the two nuclei (host and endosymbiont) (7). As a result of primary and secondary endosymbiosis, plastid genomes (ptDNAs) encode less than 10% of the predicted 1,000 to 5,000 proteins required to sustain the metabolic capacity of the plastid (8, 9).

Examples of HGT between unrelated or nomating species are abundant among prokaryotes (10, 11) but less so between prokaryotes and unicellular (12–14) or multicellular eukaryotes (15–20). Most of these latter examples are associated with parasitism or phagotrophy, including the elegant studies of HGT from the α-proteobacteria Wolbachia to insects and nematodes (16–18), and the finding of rhizobial-like genes in plant parasitic nematodes (19, 20). The exchange of genetic material between two eukaryotes is extremely rare, or at least not well documented to date. The best-studied cases include the transfer of mitochondrial DNA from achlorophyllous or epiphytic plants to the mitochondrial genome (mtDNA) of their closely related photosynthetic hosts (21), the exchange of transposons between two animal (22) or two plant (23) species, and the presence of plant genes in plant parasitic nematodes (in addition to the rhizobial genes discussed previously), which are hypothesized to be “defense” genes whose products protect the parasite from host detection (20).

The sacoglossan mollusc (sea slug) Elysia chlorotica represents a unique model system to study the potential for interdomain HGT between two multicellular eukaryotes—in this case, from a filamentous secondary (heterokont) alga (Vaucheria litorea) to a mollusc. This emerald green sea slug owes its coloring and photosynthetic ability to plastids acquired during herbivorous feeding (24–29). The plastids do not undergo division in the sea slug and are sequestered intracellularly in cells lining the finely divided digestive diverticula. The plastids continue to carry out photosynthesis, providing the sea slug with energy and carbon during its approximately 10-month life span (27, 28). Long-term plastid activity continues despite the absence of algal nuclei (27, 29), and hence a source of nuclear-encoded plastid-targeted proteins. We hypothesize that the algal nuclear genes encoding essential plastid proteins are present in the sea slug, presumably as a result of HGT. Here, we present evidence for such interdomain HGT of psbO, a nuclear gene encoding the plastid manganese-stabilizing protein (MSP = PsbO). MSP is a subunit of the photosystem II complex associated with photosynthetic oxygen evolution (30, 31), which is, unquestionably, the most important enzyme complex of oxygenic life.

Results and Discussion

Plastid Genetic Autonomy. The plastids in E. chlorotica are not transmitted vertically; rather, they must be acquired with each generation early in development to ensure maturation to the

Author contributions: M.E.R. and J.R.M. designed research; J.M.W., J.L., and K.K. performed research; J.L., M.S.T. contributed new reagents/analytic tools; K.K., D.B., A.M., and J.R.M. analyzed data; and M.E.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos. EU912438 (V. litorea complete ptDNA), EU599581 (E. chlorotica complete mtDNA), DQ514337 (V. litorea psbO ptDNA), EU621881 (V. litorea psbO ptDNA), and EU621882 (E. chlorotica psbO ptDNA)].

1To whom correspondence should be addressed. E-mail: mrumpho@umit.maine.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0804968105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA
adult sea slug (32). Laboratory coculturing studies were carried out to establish that the alga *V. litorea*, a derived heterokont alga that contains secondary plastids of red algal origin (33) (Fig. S1), was the sole source of plastids in the sea slugs (Fig. 1, Movie S1, Movie S2, and SI Methods). Subsequent sequencing and mapping of the *V. litorea* ptDNA (only the fifth heterokont ptDNA to be published to date) revealed a very compact 115,341-bp double-stranded circular genome encoding 169 genes, including 139 protein-encoding genes (14 are conserved ORFs [ORF-designated ycf5] and 2 are unknown ORFs), 27 tRNA genes, and 3 rRNA genes (Fig. 2). The genes are densely arranged with an average intergenic region of 74bp and 11.1% noncoding DNA. The overall G + C content is 28%, which is low compared with most plastids, including other related heterokonts (34, 35). The genome is separated into two large single-copy regions (62,002bp and 43,469bp) by two smaller inverted repeats (both 4935bp). All the plastid rRNA genes are found on both copies of the inverted repeat in the highly conserved operon *rrs-trnl-trnA-rrl-ref*. Unlike the four published heterokont ptDNAs that lack introns (34, 35), *V. litorea* contains one intron in the *trnL* UAA gene—an ancient intron that is also present in cyanobacteria (36). In addition, *V. litorea* has retained the genes for the light-independent chlorophyll biosynthesis pathway: *chlB*, *chlL*, and *chlN*. However, as expected, the *V. litorea* ptDNA shares more similarity with heterokonts and red algae than it does with green plastids (34, 35, 37) (see complete inventory of plastid genes by category in Table S1).

Examining the genetic autonomy of *V. litorea* ptDNA revealed the absence of the major core protein of the oxygen evolving complex of photosystem II, MSP (encoded by *psbO*). MSP has been reported to be critical to the stability of the water-splitting reaction of photosynthesis that generates atmospheric oxygen (30, 31). The evolutionary conservation of this reaction is demonstrated by the presence of MSP in all oxygenic photosynthetic organisms (30). Likewise, animal genomes have never been shown to contain *psbO*; hence, MSP cannot be made by the sea slug in the absence of HGT. We have previously demonstrated that oxygen evolution is linked to photosynthetic electron transport in the sea slug for at least 5 months after being removed from its algal prey (27), and photosystem II is generally highly susceptible to photo-oxidative damage requiring *de novo* synthesis and reassembly of its subunits (38, 39). For these reasons, we targeted *psbO* for HGT from *V. litorea* to *E. chlorotica*.

HGT and Expression of *psbO*. Heterologous degenerate primers (Table S2) were designed based on alignments of published *psbO* sequences to amplify an internal fragment using reverse transcriptase (RT)-PCR. A 452-bp fragment was amplified from both algal and sea slug cDNA (5 months after algal feeding) (Fig. 3A). The translated product was used to blast the GenBank database, which revealed a relatively high identity (48%–68%) to several secondary red algal–derived MSP amino acid sequences. The entire *V. litorea* *psbO* cDNA sequence was then obtained using 5′- and 3′-RACE, and this sequence was used to design homologous primers to amplify a 963-bp internal fragment of the *V. litorea* *psbO* cDNA (Fig. 3C). These same primers yielded a similar sized PCR product from sea slug cDNA and DNA templates (Fig. 3C), the sequences of which were identical to the *V. litorea* *psbO* cDNA sequence beginning with the start codon (there are no introns in the algal gene; Fig. S2).

Although it had been several months since the sea slugs had been in contact with any algal prey, the possibility of algal nuclei remaining in the gut of the sea slug and contaminating the total genomic DNA preparation was eliminated by carrying out the same PCR on sea slug egg DNA. Because plastids are
not inherited in *E. chlorotica*, eggs provide a source of animal DNA and RNA that is free of algal contamination (27). Amplification of the sea slug egg DNA with the same primers resulted in a 963-bp fragment (Fig. 3C) with a sequence identical to the algal and sea slug psbO product (28). As further PCR controls, primers complementary to the *V. litorea* internal transcribed spacer region (*ITS1*; GenBank EF441743) were used as a positive algal nuclear control (27); no product was observed from sea slug or sea slug egg DNA templates. Likewise, attempts to amplify psbO from negative controls (pufferfish and *Dicyostelium* DNA) using the same primers and reagents did not yield a positive PCR product. Finally, identical PCR results have been obtained from sea slugs collected from the same site on multiple occasions over a 3-year period (results not shown). Expression of the foreign gene in the sea slug was further supported by Northern blot analysis. The *V. litorea* psbO probe cross-reacted with a 1.2-kb transcript for both *V. litorea* and *E. chlorotica* RNA, along with a slightly larger transcript (1.6 kb) in the sea slug (Fig. 3B).

The identical translated MSP amino acid sequences for both *V. litorea* and *E. chlorotica* (Fig. 3D) were analyzed by phylogenetic methods (40, 41) incorporating 23 published MSP sequences. This revealed nesting of the sequences in the red algal lineage in a clade containing other heterokonts (Fig. S3). As expected for this secondary lineage, the *V. litorea* MSP preprotein contains a tripartite targeting sequence (Fig. 3D); as predicted by refs. 42, 43. This reflects the presence of chloroplast endoplasmic reticulum around the complex plastids, which must be transmitted by proteins targeted to the plastid (44). What is interesting is that the MSP preprotein in the sea slug has retained the entire tripartite targeting sequence (Fig. 3D), especially in light of the observation that the chloroplast endoplasmic reticulum does not appear to be retained around the engulfed chloroplasts (28).

Recently, it was reported that nuclear genes encoding plastid-localized light harvesting complex proteins (*fcp*, *lhcv1*, and *lhcv2*) have also been transferred from *V. litorea* to *E. chlorotica* (45). Using a similar PCR approach, identical nucleotide sequences were reported for sea slug and algal *fcp* and *lhcv1*, and only a single base substitution was found between *lhcv2* and adult sea slug or alga *lhcv2*. Although evidence from Southern blotting has not been achieved in the study reported here or for the light harvesting complex protein genes (45), we were able to obtain sequence information using genome walking for the 3’ untranslated flanking region of the psbO gene from both algal DNA and sea slug egg DNA. A nested gene-specific primer coupled with an adapter-specific primer (Table S2) yielded a 3’ flanking sequence from both organisms that was identical for the first 81 bp corresponding to the 3’ end of the *psbO* gene and ending with the stop codon (Fig. S4). This sequence was followed by a highly diverged sequence corresponding to the 3’ untranslated region in each genome. These results support the interdomain transfer of an algal gene to a mollusc, its expression in the foreign host, and also that the gene has been inserted into the germline, even though the plastids are not yet transmitted vertically in the sea slug.

**Mechanism and Site of Integration of Transferred Genes.** Similar to many other phagocytic or parasitic relations that lead to presumptive HGT events, the *E. chlorotica/V. litorea* plastid endosymbiosis involves intimate physical contact between predator and prey. During the sea slug’s phagocytic feeding, the algal nuclei come into direct physical contact with the sea slug digestive epithelium. Upon nuclear rupture in the gut, pieces of algal chromosomal DNA (and possibly transcripts) may have been randomly transferred by “bulk transfer” or viral transmission (46) to the sea slug. Two potential sites for insertion of foreign genes in the sea slug are the nuclear genome and the mtDNA. Mitochondrion-to-mitochondrion gene transfer is now recognized as a dominant mode of HGT in plants because of the larger and more plastic mtDNAs in these taxa (21). The smaller compact animal or metazoan mitochondrial genome is generally believed to be a poorer target for foreign gene insertion. However, some basal metazoans do exhibit greater variation in mtDNA size and gene content (47). This includes multiple examples of HGT of group I intron sequences (normally not found in animals) into the mtDNA of a sponge (48), a sea anemone (49), and a coral (50).

To determine if the mtDNA of *E. chlorotica* serves as a target for any foreign genes, including *psbO*, we used PCR and primer walking to obtain the complete sequence and map the 14,132-bp mtDNA from sea slug eggs (Fig. 4). The genome was found to escape the standard 37 genes found in other typical animal mitochondria (see ncbi.nlm.nih.gov-genomes/ORGANELLES/33208). No introns were identified, and only 0.0125% of the DNA was noncoding. By measuring the G + C content over adjacent windows of 500 nt with 200-nt overlaps, the values were found to be uniformly distributed across the windows, suggesting homogeneity in GC content of the mtDNA and not supporting
the existence of a chimeric region (Fig. S5). To assess further the possibility of HGT in *E. chlorotica* mitochondria, we did phylogenetic analyses with nucleotide data generated using a sliding window approach with the genome data (i.e., DNA sequences that are independent of gene structure) and using the complete translated ORFs. The maximum likelihood phylogenetic trees inferred with these alignments showed that the *E. chlorotica* sequences are monophyletic with molluscs, consistent with a vertical evolutionary history for *E. chlorotica* mtDNA (e.g., see the phylogenetic tree of cytochrome b in Fig. S6). These analyses point to an intact and "typical" animal mtDNA in *E. chlorotica*. We, however, do not argue absolutely against the possibility of a partial DNA insertion from an algal or other source in this genome; rather, that if such an insertion exists, it is not detectable using the approaches described here. In any case, it is undoubtedly more likely that large-scale gene insertion would be able using the approaches described here. In any case, it is undoubtedly more likely that large-scale gene insertion would be able using the approaches described here.

**Conclusions**

Molecular evidence is presented supporting eukaryotic multicellular interdomain HGT (including into the germline) using a mollusc model and expression of an essential algal nuclear gene. The implications for evolution and speciation through acquisition of foreign parts and selected genes to produce new lineages, as proposed by Margulis (2), are heightened by this unusual photosynthetic mollusc.

**Methods**

**Experimental Materials.** *E. chlorotica* was collected from Martha’s Vineyard Island in Massachusetts and maintained without algae in aquaria containing aerated artificial seawater (925 mosmol; Instant Ocean, Aquarium Systems) at 12 °C during a 14-h photoperiod (27). After 3 months, eggs produced by *E. chlorotica* were used to initiate culturing experiments as described in *SI Methods*. *V. litorea* CCMP2940 filaments were maintained in a modified f/2 medium (27).

**Nucleic Acid Preparation.** DNA and RNA were isolated from sea slugs (5 months after feeding or collection), sea slug eggs, and algal filaments using DNAzol or DNAzol extra strength (Molecular Research Center, Inc.) and the RNeasy mini kit (Qiagen), respectively, unless noted differently. RNase and DNase were added during the extraction process for DNA and RNA, respectively, and negative controls were run on each. First-strand DNA was synthesized using SuperScript II ribonuclease H- RT (Invitrogen) and oligo d(T) priming on DNase-treated RNA.

**PCR Amplification and Northern Blotting.** Degenerate primers (psbO R and psbO L2; see Table S2) were designed to amplify an internal fragment of *V. litorea* psbO based on the conserved regions of several heterokont and red alga psbO sequences (for list, see *SI Methods*). This 452-bp psbO fragment was then used as a probe for Northern blot analysis with the Northern Max kit (Ambion) and RediprimeII random prime labeling system (Amersham Biosciences).

**RACE and Phylogenetic Analysis of psbO.** The complete *V. litorea* psbO gene was obtained by rapid amplification of cDNA ends (RACE) using the Gene-Racer Kit (Invitrogen) and primers listed in Table S2. Homologous primers (psbO L5 and psbO R8) were then designed to amplify a larger (963-bp) internal fragment of the *V. litorea* psbO cDNA. Phylogenetic analysis of psbO was based on amino acid sequences of 25 mature proteins (for list, see *SI Methods*) and carried out using maximum parsimony in PAUP 4.0b10 (41).

**ptDNA Sequencing.** ptDNA was purified from *V. litorea* filaments as described (51). ptDNA was digested with the restriction endonucleases *PstI*, *HindIII*, and *EcoRI*. The *PstI* and *HindIII* fragments were ligated and a restriction site map using all three enzymes was produced as described by Lehman and Manhart (52). A total of 104 kb was obtained from cloned restriction fragments. The remainder of the genome (11 kb) was obtained by PCR amplification (53). Thirty-eight oligonucleotide primers were used to fill gaps between cloned restriction fragments and to check fragment connections, using all possible combinations of these primers. Fragments were sequenced by primer walking (53).

**Genome Walking.** Clontech’s Genome Walking Kit was used with gene-specific and adapter primers (sequences presented in Table S2) to amplify the 3’ end of the psbO gene and the flanking untranslated region using algal DNA and sea slug egg DNA (see *SI Methods* for additional details).

**mtDNA Sequencing.** Universal primers (ref. 54; Table S3) were used to amplify fragments of the mitochondrial *rrnL*, *cob*, and *cox1* regions from sea slug egg DNA and then in various combinations to amplify the entire mtDNA. The mitochondrial sequence was annotated using Dual Organellar GenoMe Annotator (DOGMA) (55), and the map was drawn using OrganellarGenomeDRAW (OGDRAW) (56). Additional information, including analysis for HGT, is provided in *SI Methods*.

**ACKNOWLEDGMENTS.** This research was supported by National Science Foundation grants IBN-9808904 (M.R. and J.M.) and IOS-0726178 (M.R. and M.T.); the American Society of Plant Biologists’ Education Foundation (M.R. and J.L.); the National Institutes of Health (grant R01ES013579 to D.B.); and the University of Maine (M.R.). This is manuscript no. 3024 of the Maine Agriculture and Forestry Experiment Station Hatch Project no. ME08361-08MRF (NC 1168).

![Fig. 4.](image-url) **Map of the mtDNA of the sacoglossan mollusc *E. chlorotica*.** Genes transcribed clockwise are shown on the outside of the circle, whereas those transcribed counterclockwise are shown to the inside of the circle. Names of tRNA genes are indicated by the three-letter amino acid code with the two leucine and two serine tRNAs differentiated by + and − signs, recognizing codons UAG and UAA for leucine and AGN and UCN for serine, respectively.