Transcriptomic Evidence for the Expression of Horizontally Transferred Algal Nuclear Genes in the Photosynthetic Sea Slug, *Elysia chlorotica*

Sidney K. Pierce,†,*[1,2] Xiaodong Fang,†[3] Julie A. Schwartz,[1] Xuanting Jiang,[3] Wei Zhao,[3] Nicholas E. Curtis,[1] Kevin M. Kocot,[4] Bicheng Yang,[3] and Jian Wang*[3,4]

[1]Department of Integrative Biology, University of South Florida

[2]Department of Biology, University of Maryland

[3]BGI-Shenzhen, Shenzhen, China

[4]Department of Biological Sciences, Auburn University

†These authors contributed equally to this work.

***Corresponding author:** E-mail: Wangjian@genomics.org.cn; pierce@usf.edu.

Associate editor: Charles Delwiche

Abstract

Analysis of the transcriptome of the kleptoplastic sea slug, *Elysia chlorotica*, has revealed the presence of at least 101 chloroplast-encoded gene sequences and 111 transcripts matching 52 nuclear-encoded genes from the chloroplast donor, *Vaucheria litorea*. These data clearly show that the symbiotic chloroplasts are translationally active and, of even more interest, that a variety of functional algal genes have been transferred into the slug genome, as has been suggested by earlier indirect experiments. Both the chloroplast- and nuclear-encoded sequences were rare within the *E. chlorotica* transcriptome, suggesting that their copy numbers and synthesis rates are low, and required both a large amount of sequence data and native algal sequences to find. These results show that the symbiotic chloroplasts residing inside the host molluscan cell are maintained by an interaction of both organellar and host biochemistry directed by the presence of transferred genes.

Key words: transcriptome, HGT, horizontal gene transfer, chloroplast symbiosis, kleptoplasty, *Elysia chlorotica*, *Vaucheria litorea*.

Introduction

Certain cells lining the digestive tubules of several species of sacoglossan, opisthobranch, sea slugs are able to sequester chloroplasts from their algal food. The plastids remain intact inside the digestive cells for some length of time, depending on the species involved (Pierce and Curtis 2012). Also, in several of these slug species, the captured plastids, often called kleptoplasts, are capable of photosynthesis, and in a few of the species, photosynthesis continues almost unabated for as long as a year in the complete absence of any contact with the algal source of the chloroplasts (Pierce and Curtis 2012).

Most detailed information about various aspects of the mechanism of long-term maintenance of the chloroplast symbiosis has come from work on *Elysia chlorotica* (Gould), where the sequestered chloroplasts come from the chromophytic alga, *Vaucheria litorea* (C. Agardh). Once this slug sequesters the chloroplasts, it can continue to photosynthesize for 10–12 months in the absence of any algal food. Several chloroplast proteins and chlorophyll a are synthesized during that starvation period, and polymerase chain reaction (PCR) experiments have demonstrated the presence of at least 11 algal nuclear genes, all involved in photosynthesis, in *E. chlorotica* adult and veliger larval genomic DNA as well as in adult slug RNA (Pierce et al. 2007, 2009; Rumpho et al. 2008, 2009; Schwartz et al. 2010). All of these results show that translationally competent algal nuclear genes are present in the slug and that plastid protein and pigment turnover, necessary for sustained photosynthesis, is taking place in the slug cell, supported by horizontal gene transfer (HGT) between the two multicellular species. In addition, a variety of chloroplast-encoded chloroplast proteins are also synthesized while the plastid resides inside the *E. chlorotica* digestive cell (Mujer et al. 1996; Green et al. 2000). However, recent partial analyses of the transcriptomes of two other slug species, *E. timida* and *Plakobranchus ocellatus*, failed to find any transcriptome sequence reads corresponding to algal nuclear genes, which lead to the provocative conclusion that, in spite of the entire foregoing, HGT has not occurred between slug and algae and that "sacoglossan are not, in genetic terms, what they eat" (Wägele et al. 2011). Perhaps not, at least in the case of the two species investigated by Wägele et al. (2011), but several aspects of this study seem to make such a broad conclusion premature. For instance, the data set represented only a small fraction of the transcriptome. Also, the RNA came from just a few specimens and was extracted from whole animals, although only a small fraction of cells contain chloroplasts. In addition, Wägele et al. (2011) assumed that expression levels

of algal genes in the slug cells would be equivalent to transcription rates in the algae even though expression levels of at least two nuclear-encoded genes for plastid-targeted proteins are much lower in *E. chlorotica* than in its food alga (Soule 2009). Thus, generalizing negative results obtained from problematical data to the rest of the species of kleptoplastic slugs, especially in the presence of the large amount of biochemical data, including PCR, and showing the presence of transferred genes, seem inappropriate.

However, another recent report failed to find any evidence for transferred algal nuclear genes in the transcriptome of *E. chlorotica*, although 19 chloroplast-encoded gene sequences emerged from the analysis (Pelletreau et al. 2011). Unfortunately, while this study investigated *E. chlorotica*, the transcriptome sequencing yielded <20% (14,000) of the contigs of the *P. ocellatus* transcriptome and was not compared with *V. litorea* native sequences. Thus, the lack of both sequencing depth and database voids suggests that rare transcripts could have been easily missed. Indeed, a conclusion of this analysis was that "more exhaustive sequencing may be required" to adequately test for the presence of transferred genes in *E. chlorotica* (Pelletreau et al. 2011).

In spite of the issues with these negative studies (Pelletreau et al. 2011; Wägele et al. 2011), both nonetheless point to the likelihood that if transferred algal nuclear genes are present in sacoglossan slugs, they will be of very low copy number, their expression will be low, and knowledge of the native algal sequence will facilitate, perhaps even be required for, the annotation. Therefore, instead of assuming that a large number of transcripts for nuclear-encoded and plastid-encoded proteins would be present in the slug cell, we have hypothesized that such would be exceedingly rare in the transcriptome and have done our own analysis of *E. chlorotica* creating large amounts of Illumina-generated sequence data. In addition, to facilitate an accurate annotation, we have sequenced the genome of *V. litorea* as well as the algal transcriptome, to provide a database of native algal transcript sequences. Our *E. chlorotica* transcriptome data contain a variety, albeit rare, of chloroplast-encoded transcripts and, in addition, rarer still, at least 111 reads that match 52 algal nuclear-encoded sequences, including one that matches exactly an algal nuclear sequence found by PCR in *E. chlorotica* genomic DNA and cDNA in the earlier studies.

Materials and Methods

Animals and Algae

Specimens of *E. chlorotica* were collected from a salt marsh near Menemsha on Martha's Vineyard, MA, and shipped to Tampa, FL. The sea slugs were placed into aerated aquaria containing artificial seawater made from Instant Ocean salts dissolved in sterilized deionized water at 1,000 mOsm. The aquaria were kept at 10 °C in a cold box equipped with fluorescent lights set on a 14:10 light–dark cycle. The slugs were starved for at least 2 months under the foregoing conditions before use in experiments.

Filaments of *V. litorea* used in the experiments came from a culture that has been maintained in our lab for more than 10 years. The initial filament used to establish the culture came from the same marsh that provided the slugs. The *V. litorea* filaments are grown in a modified F/2 enriched seawater at 250 mOsm in an incubator at 20 °C, illuminated by fluorescent lights on a 14:10 light–dark cycle (Pierce et al. 1996). Since the original culture was established from a single filament and has only grown vegetatively, the filaments used here are clonal.

Extraction of Genomic DNA from *V. litorea*

Genomic DNA was extracted from 10 g of algal filaments using a protocol modified from Al-Samarrai and Schmid (2000). The filaments were rinsed with fresh culture media, blotted to remove excess liquid, frozen in liquid nitrogen, and ground to a powder with a precooled mortar and pestle. Lysis buffer (40 mM Tris–acetate, 20 mM sodium acetate, 1 mM ethylenediaminetetraacetic acid, and 1% sodium dodecyl sulfate, pH 7.0) was added to the frozen algal powder and extracted for 10 min at room temperature (RT) while rotating. To facilitate the precipitation of proteins and polysaccharides, 5 M NaCl was added to each tube (2:5 v/v), the tubes were vortexed and then centrifuged at 12,000 × g at 4 °C. The supernatant was extracted with phenol/chloroform (1:1 v/v) and spun at 10,000 × g 4 °C. The aqueous phase was extracted with chloroform (1:1 v/v) and centrifuged again at 10,000 × g 4 °C. DNA was precipitated from the aqueous phase by adding isopropanol (1:1 v/v) and spun at 10,000 × g 4°C to pellet the nucleic acids. The precipitated DNA was resuspended in lysis buffer containing 100 µg/ml RNase A (Qiagen, Valencia, CA) and rotated for 30 min at RT. Following that incubation, the DNA was run through the purification process a second time. The final precipitated DNA was washed twice with 75% ethanol, air dried, resuspended in nuclease-free water, quantified spectrophotometrically at 260 nm, and then express shipped to BGI in Hong Kong, China.

Algal Transcriptome

Two *V. litorea* transcriptome data sets were used in our analysis. One set was produced earlier by 454 sequencing (Schwartz et al. 2010) and the second set using the Illumina platform as described below. The 454 data are referred to as EST and the Illumina data as RNA-seq, hereafter.

Extraction of RNA from *V. litorea*

Total RNA was isolated from about 1 gm of algal filaments after 5 h of exposure to light. The filaments were ground into a frozen powder as described above, RNA was extracted using the RNeasy Plant mini kit (Qiagen, Valencia, CA) following the manufacturer's instructions, quantified spectrophotometrically at 260 nm, placed on dry ice, and express shipped immediately to BGI in Hong Kong, CN.

Extraction of RNA from *E. chlorotica*

Total RNA was isolated from >2-month starved slugs after 8 h of light exposure by homogenization in Trizol Reagent (Invitrogen, Carlsbad, CA). The homogenate was

centrifuged at 12,000 × g at 4 °C to remove cellular debris. The supernatant was extracted with chloroform (1:6 v/v) and centrifuged at 10,000 × g at 4 °C. RNA was precipitated from the aqueous phase by adding isopropanol (1:4 v/v) followed by 0.8 M $Na_3C_6H_5O_7$/1.2 M NaCl solution (1:4 v/v) and was spun at 12,000 × g at 4 °C. The RNA pellet was washed twice with 75% ethanol, air dried, and resuspended in nuclease-free water. The RNA in this final solution was quantified spectrophotometrically and express shipped to BGI in Hong Kong, CN on dry ice.

Algal Genome Sequencing and Assembly

The genome of *V. litorea* was sequenced (Illumina HiSeq 2,000 platform) using a whole genome shotgun strategy. To reduce bias, eight paired-end sequencing libraries were constructed, with various insert sizes (350 bp, 400 bp, 800 bp, 2 kbp, 5 kbp, 10 kbp, and 20 kbp), and sequenced.

Genome Assembly

In total, 8 Gb, or 86-fold coverage, of high-quality reads were used in the assembly. The algal genome was assembled using SOAPdenovo (Li, Zhu, et al. 2010) software (http://soap.genomics.org.cn), using the same procedures described for assembly of the giant Panda genome (Li, Fan, et al. 2010).

A total of 7.9 Gb (or 85.9X) data were retained for assembly. All high-quality paired-end reads were aligned into contigs for scaffold building. This paired-end information was subsequently used to link contigs into scaffolds, step-by-step, from short insert sizes to long insert sizes. About 3.9 Gb (or 42.4X) data were used to build contigs for the algal genome, and all the high-quality data were used to build scaffolds. Some intrascaffold gaps were filled by local assembly using the reads in a read pair, where one end uniquely aligned to a contig while the other end was located within a gap. The final total contig size and N50 were 83.6 Mb and 59.6 Kb, respectively. The total scaffold size and N50 were 93.2 Mb and 333.3 Kb, respectively. More than 97% of long ESTs (>500 bp) mapped to the algal genome, which indicated the high quality of the genome in the transcribed regions.

Gene Annotation Pipeline and Evaluation of Gene Quality

Since there is only scant *V. litorea* sequence information in the public databases, we used both homology based and de novo methods to localize gene sequences in the algal genomic data incorporating reads from both the RNA-seq and EST data. To identify homologous genes, protein sequences from algal species that were available in NCBI (http://www.ncbi.nlm.nih.gov/), including *Ectocarpus siliculosus*, *Phaeodactylum tricornutum*, *Micromonas sp.*, *M. pusilla*, *Ostreococcus lucimarinus*, *O. tauri*, and *Volvox carteri*, were mapped to the *V. litorea* genome using TBlastN (Kent 2002). Then, homologous genome sequences were aligned against the matching proteins, using GeneWise (Birney et al. 2004) to define gene models. For de novo discovery of coding genes, AUGUSTUS (Stanke and Waack 2003), GlimmerHMM, and SNAP were used with appropriate parameters. ESTs were mapped to the genome with BLAST and assembled to genes with PASA. RNA-seq were mapped to the genome using TopHat (http://tophat.cbcb.umd.edu/), and transcriptome-based gene structures were obtained with Cufflinks (http://cufflinks.cbcb.umd.edu/). Finally, the homology-based de novo derived EST prediction and transcript gene sets were merged to form a comprehensive and nonredundant reference gene set using GLEAN (http://sourceforge.net/projects/glean-gene/), removing all the genes which had only de novo method support. This procedure produced a reference set of 17,988 *V. litorea* genomic coding genes.

Transcriptome Sequencing and Assembly

To prepare the slug and algal transcriptomes for sequencing, poly $(A)^+$ RNA was enriched from the total RNA of each species, sheered into fragments, and cDNA was synthesized by reverse transcription. The cDNA from each species was then sequenced using standard high throughput techniques (Illumina HiSeq2000). All high-quality reads were assembled into contigs longer than 100 bp using SOAPdenovo software (Li et al. 2010). Contigs were linked into scaffolds by mapping reads back to contigs and combining paired-end information.

Sequence Analysis

Alignment and annotation of the *E. chlorotica* transcriptome sequences were done using the *V. litorea* coding sequence database as a reference source (fig. 1) utilizing MPI BLAST 1.5.0 software (Darling et al. 2003; Lin et al. 2005), by means of the University of South Florida's 120 node computer cluster platform consisting of dual Intel Xeon X5460 Quad Core processors each with 16 GB of memory. Briefly, the slug contigs, scaffolds, and raw reads from the transcriptome data were formatted as databases and then compared with all 17,988 *V. litorea* coding sequences using the BlastN algorithm (Altschul et al. 1990) with a cutoff of 1×10^{-10}. Slug transcripts with significant hits were aligned to the corresponding *V. litorea* reference sequence using the ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) sequence alignment program to determine the location of the matching slug transcripts in the coding sequence. The corresponding *V. litorea* matching sequences were then analyzed using the BlastX algorithm (Altschul et al. 1990) to determine if they were nuclear- or chloroplast-encoded in origin by comparison with the *V. litorea* chloroplast genome database in NCBI. Last, slug transcripts matching to algal nuclear-encoded sequences were compared with the assembled slug and algal transcriptomes using BlastN to determine if the sequences were located in the contig or raw-read data sets.

Results

Chloroplast-Encoded Transcripts in *E. chlorotica*

Our sequencing runs of the cDNA made from the *E. chlorotica* transcriptome produced 98,238,204 reads, which assembled into 459,299 contigs and 378,851 scaffolds (table 1).

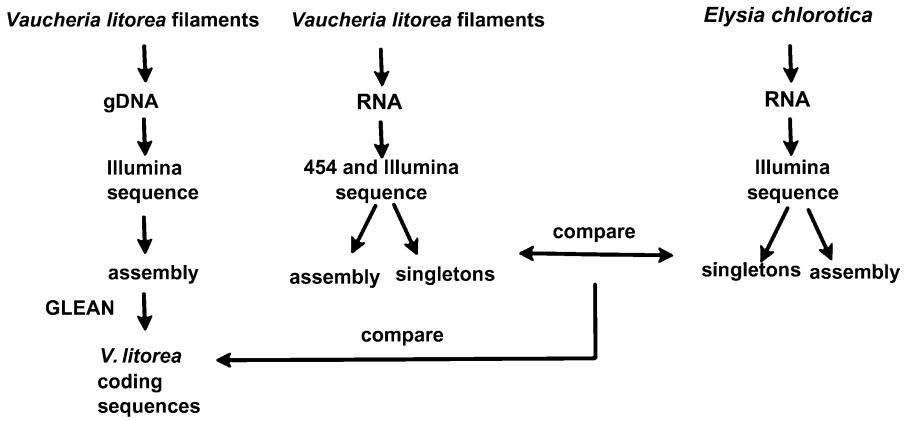

Fig. 1. Flow chart of *Elysia chlorotica* and *Vaucheria litorea* transcriptome analyses.

Although relatively rare in number among the contigs, a functionally diverse set of chloroplast-encoded sequences were present in the *E. chlorotica* transcriptome (table 2). One hundred one chloroplast-encoded protein coding genes were identified in the slug transcriptome sequence data that matched sequences in both the *V. litorea* EST or RNA-seq data, of which 60 were located among the contigs and the remaining 41 were present in the raw reads, and also matched sequences in the *V. litorea* chloroplast genome (table 2). Transcript sequences coding for proteins in Photosystem I and II reaction centers, both subunits of RuBisCO, cytochrome, chlorophyll and ATP synthesis, and RNA polymerase were all discovered (table 2). In addition, matching transcripts associated with protein synthesis and trafficking were present in the slug and algal transcriptomes as well as the *V. litorea* chloroplast genome data (table 2). An additional 36 sequences, which mapped to a variety of ribosomal protein subunits, matched between the slug and algal transcriptomes as well as with the chloroplast genome sequences (table 2). Most of the matching sequences in the chloroplast-encoded data set were 100% identical, both between slug and algal transcripts as well as compared with the chloroplast genome, with the exceptions containing either nonassigned or miscalled bases.

Nuclear-Encoded Algal Sequences in the Slug Transcriptome

Although much less common in the transcriptome data than the chloroplast-encoded transcripts, and not present in any of the contigs, 111 sequences from the *E. chlorotica* transcriptome database matched *V. litorea* reference gene sequences, representing 52 putative nuclear-encoded *V. litorea* genes (table 3). Among the 111 slug transcripts, 88 were 100% identical, 15 contained 1 bp difference, 4 contained 2 bp difference, 2 contained 3 bp differences, 1 contained 4 bp differences, and 1 contained 6 bp differences to aligned portions of the *V. litorea* genome coding sequences. All of the 52 *V. litorea* genomic coding sequences contained at least two and as many as four nonoverlapping slug transcriptome sequences (fig. 2 and table 3; supplementary material online). So even though the raw reads were only 90 bp in length, the match within each *V. litorea* sequence was at least 180 bp to as much as 360 bp. Furthermore, in addition to matching the algal coding sequences, 106 of the *E. chlorotica* transcript sequences were also present in the *V. litorea* transcriptome (table 3). Twenty-seven of the transcript sets matching *V. litorea* genomic coding sequences were homologous to genes involved in photosynthesis, carbon fixation, carbohydrate metabolism, thylakoid structure, chaperone activity, and other unique chloroplast processes. Nineteen *V. litorea* gene sequences were annotated as hypothetical or unknown proteins, mostly within the *E. siliculosus* genome data or other algal species, and the final six sequences returned no Blast hit, but all were present in the slug transcriptome, the *V. litorea* genomic sequences and, all but five, in the *V. litorea* transcriptome (table 3).

As usual, the possibility of contamination of *E. chlorotica* mRNA with algal material in this kind of study is of concern, especially considering the rarity of the matching transcripts identified in these data. However, the RNA was extracted from animals that were starved for at least 2 months, to ensure that the gut was clear of any algal food. Our previous PCR-based studies with DNA or cDNA extracted from similarly starved slugs have produced negative results for nuclear-encoded, non–chloroplast-targeted sequences for *V. litorea* "ITS" regions as well as "SPDS" (Pierce et al. 2007, 2009, Schwartz et al. 2010). In addition, the algal nuclear sequences that were found here in the slug

Table 1. Summary of the *Elysia chlorotica* Transcriptome Sequence Data.

	Sequences (*n*)	Base Pairs (Mbp)	Mean Length (bp)	N50 (bp)
Raw reads	**98,238,204**	**8,841.4**	**90**	
Contigs (≥100 bp)	**459,299**	**80.7**	**175**	**180**
Scaffold sequences (≥100 bp)	**378,851**	**86.5**	**228**	**254**

Table 2. List of the Chloroplast-Encoded Transcripts in the *E. chlorotica* Transcriptome and the Corresponding Sequences in the *V. litorea* Transcriptome and Chloroplast Genome.

Protein Category	Protein Name	Number of Slug Transcript Read Matches[a]	Assembled in Slug Contigs (C) or Raw Reads (R)	*E. chlorotica* Transcriptome Sequences	*V. litorea* Transcriptome Matching Sequence Accession Number	*V. litorea* Chloroplast Genome Match Accession Number (YP0002327xxx.1)
Photosystem I reaction center	Subunit II	5	R	EC Illumina reads[b]	JP709230	575
	Subunit IV	8	R	EC Illumina reads	JP709231	498
	Subunit VII	22	C	JP709299[c]	JP709232	573
	Subunit XI	24	C	JP709300	JP709233	559
	PSI assembly related protein	4	R	EC Illumina reads	JP709234	542
	Plastocyanin-binding subunit III	12	C	JP709301	JP709235	477
	Photosystem assembly protein Ycf3	2	R	EC Illumina reads	E52OB4B01EZIRC[d]	461
	Apoprotein A1	568	C	JP709302	JP709236	590
	P700 apoprotein A2	911	C	JP709303	JP709236	591
Photosystem II reaction center	D1	4234	C	JP709304	E52OB4B02FP4GF	545
	D2	97	C	JP709305	JP709237	479
	CP43 apoprotein	131	C	JP709306	JP709237	480
	CP47 chlorophyll apoprotein	97	C	JP709307	JP709238	500
	Protein T	1	R	EC Illumina reads	JP709238	501
	Protein N	1	R	EC Illumina reads	JP709238	502
	PSII phosphoprotein	3	R	EC Illumina reads	No match	503
	Protein E	6	R	EC Illumina reads	JP709239	562
	Protein J	4	R	EC Illumina reads	No match	565
RuBisCO	Small subunit	20	C	JP709308	JP709240	543
	Large subunit	28	R	EC Illumina reads	JP709241	544
Cytochrome synthesis	Apocytochrome f	57	C	JP709309	JP709242	462
	Cytochrome b6	36	C	JP709310	JP709243	486
	Cytochrome b6/f complex subunit 4	17	C	JP709310	JP709243	487
	Cytochrome b6/f complex petM subunit	3	R	EC Illumina reads	No match	578
	Cytochrome c553	492	C	JP709311	E52OB4B02GOVIB	555
	Cytochrome c550 oxygen evolving complex component	178	C	JP709312	JP709244	556
	Ferredoxin	23	R	EC Illumina reads	JP709245	504
	Ferredoxin-thioredoxin reductase catalytic chain	57	C	JP709313	JP709246	567
	Protein involved in cytochrome c biogenesis	17	R	EC Illumina reads	JP709247	548
	Thiamin biosynthesis protein	16	R	EC Illumina reads	JP709248	572
Chlorophyll synthesis	Protochlorophyllide reductase ChlB subunit	16	R	EC Illumina reads	JP709249	464
	Protochlorophyllide reductase ATP binding subunit (chlI)	181	C	JP709314	JP709250	455
	Protochlorophyllide reductase ATP binding subunit (chlL)	10	R	EC Illumina reads	JP709251	547
	Protochlorophyllide reductase ChlN chain	8	R	EC Illumina reads	JP709252	546
ATP synthesis or binding	CfxQ-like protein	31	C	JP709315	JP709253	549
	ATP synthase CF1 subunit b	272	C	JP709316	JP709254	459
	ATP synthase CF1 α subunit	26	C	JP709317	JP709255	465
	ATP synthase CF0 subunit III	12	C	JP709318	JP709256	469
	ATP synthase CF0 subunit IV	16	C	JP709319	JP709257	470
	ABC ATPase	6	R	EC Illumina reads	No match	541
RNA polymerase	α chain	6	R	EC Illumina reads	JP709258	512
	β″ subunit	10	R	EC Illumina reads	No match	473
	β′ subunit	116	C	JP709320	JP709259	474
	β subunit	567	C	JP709321	JP709260	475

Table 2
Continued

Protein Category	Protein Name	Number of Slug Transcript Read Matches[a]	Assembled in Slug Contigs (C) or Raw Reads (R)	*E. chlorotica* Transcriptome Sequences	*V. litorea* Transcriptome Matching Sequence Accession Number	*V. litorea* Chloroplast Genome Match Accession Number (YP0002327xxx.1)
Ribosomal	L1	1	R	EC Illumina reads	No match	496
	L2	179	C	JP709322	JP709261	530
	L3	7	R	EC Illumina reads	JP709262	533
	L4	43	C	JP709323	JP709263	532
	L5	15	R	EC Illumina reads	JP709264	521
	L6	294	C	JP709324	JP709265	519
	L9	4	R	EC Illumina reads	No match	493
	L11	8	C	JP709325	JP709266	495
	L12	2	R	EC Illumina reads	JP709231	497
	L13	18	C	JP709326	JP709267	511
	L14	88	C	JP709327	JP709268	523
	L16	117	C	JP709328	E52OB4B01BJN0V	526
	L18	106	C	JP709324	E52OB4B01AMWPF	518
	L19	2	R	EC Illumina reads	No match	551
	L20	3	R	EC Illumina reads	No match	536
	L21	1	C	JP709329	JP709269	588
	L22	47	C	JP709330	JP709270	528
	L24	7	C	JP709327	E52OB4B01DUAI4	522
	L27	18	C	JP709329	JP709269	587
	L31	14	C	JP709331	No match	509
	L32	1	R	EC Illumina reads	No match	585
	L33	1	R	EC Illumina reads	JP709271	456
	S2	8	R	EC Illumina reads	No match	472
	S3	56	C	JP709328	JP709272	527
	S4	10	C	JP709332	JP709273	485
	S5	556	C	JP709333	JP709274	517
	S7	71	C	JP709334	JP709275	507
	S8	27	C	JP709324	JP709264	520
	S9	18	C	JP709335	JP709267	510
	S11	32	C	JP709336	JP709276	513
	S12	224	C	JP709331	JP709277	508
	S13	341	C	JP709337	JP709276	514
	S14	35	C	JP709338	JP709278	592
	S16	6	R	EC Illumina reads	JP709279	483
	S18	2	R	EC Illumina reads	JP709271	457
	S19	2	R	EC Illumina reads	No match	529
Hypothetical	YCF3	13	C	JP709339	JP709280	458
	YCF24	24	C	JP709340	JP709281	540
	YCF66	2	R	EC Illumina reads	JP709282	557
	Conserved chloroplast protein	21	C	JP709341	JP709283	553
	Putative cell division protein FtsH	76	C	JP709342	JP709284	499
Others	Hsp70-type chaperone	922	C	JP709343	JP709285	534
	60 kDa chaperonin	292	C	JP709344	JP709286	574
	Elongation factor Ts	2	R	EC Illumina reads	JP709287	471
	CAB/ELIP/HLIP-like protein	2	R	EC Illumina reads	No match	481
	Acetohydroxy acid synthase large subunit	37	C	JP709345	JP709288	489
	DNA-replication helicase	15	C	JP709346	JP709289	494
	Translation elongation factor Tu	404	C	JP709347	JP709275	506
	Preprotein-translocase subunit Y	1	R	EC Illumina reads	E52OB4B02G87QP	516
	Preprotein-translocase subunit	9	C	JP709348	JP709290	583
	Acetolactate synthase small subunit	65	C	JP709349	JP709291	535

Table 2
Continued

Protein Category	Protein Name	Number of Slug Transcript Read Matches[a]	Assembled in Slug Contigs (C) or Raw Reads (R)	*E. chlorotica* Transcriptome Sequences	*V. litorea* Transcriptome Matching Sequence Accession Number	*V. litorea* Chloroplast Genome Match Accession Number (YP0002327xxx.1)
	Thiol-specific antioxidant protein	13	R	EC Illumina reads	No match	538
	Isoflavone reductase	9	R	EC Illumina reads	JP709292	539
	Caseinolytic-like Clp protease	220	C	JP709350	JP709293	552
	Acyl carrier protein precursor	2	R	EC Illumina reads	JP709294	589
	Orf 32	2	R	EC Illumina reads	No match	586
	Unidentified reading frame 1	4	R	EC Illumina reads	JP709295	463

[a] Cutoff = e^{-30}.
[b] "EC Illumina reads" can be located using NCBI Study accession number SRP009263.2 and are all listed in supplementary table 2, Supplementary Material online.
[c] Sequence identifiers "JPxxxxxx" are NCBI database accession numbers.
[d] All singleton reads in this column labeled "E52OB4B0xxxxxx" are included in Sequence Read Archive Study number SRP009267.2.

transcripts, corresponded mostly to chloroplast-targeted proteins involved in photosynthesis, or genes involved in protein processing and chaperone activity. If algal RNA contamination was present in the slug material, we would have likely identified highly expressed, non–photosynthesis-related genes in the data set. But, none were found.

Discussion

The transcriptome of *E. chlorotica* contains a variety of diverse, although relatively rare, transcripts not only of *V. litorea*, chloroplast-encoded, origin, but also of algal nuclear-encoded origin. Clearly, the symbiotic plastid genome is transcriptionally active within the host cell and is producing transcripts for a variety of chloroplast proteins. More importantly, we have detected the presence of several transcripts for nuclear-encoded algal proteins in the slug RNA. While some of the nuclear-encoded reads are for products of presently unknown function, most of the annotatable reads code for proteins involved in the sustenance of photosynthesis. Two reads, in particular, correspond to *prk*, which we (Schwartz et al. 2010) and others (Rumpho et al. 2009) found in slug genomic DNA and cDNA with PCR experiments. More than 60 algal nuclear-encoded genes have now been demonstrated in *E. chlorotica* DNA and RNA, either by PCR (Pierce et al. 2007, 2009; Rumpho et al. 2009; Schwartz et al. 2010) or by our transcriptome sequencing reported here.

Among the chloroplast-encoded transcripts in the *E. chlorotica* transcriptome are several that code for components of both PSI and PSII reaction centers, including D1, D2, and CP43, whose months-long synthesis by the slug had been demonstrated by immunoblot experiments (Mujer et al. 1996; Green et al. 2000). Similarly, the presence of chloroplast-encoded transcripts for cyt F and both the large and small subunits of RuBisCO verify the long-term synthesis of those proteins within the *E. chlorotica* cell (Green et al. 2000). Also similarly, the presence of transcripts for a plastid-encoded subunit of ChlI in the chlorophyll synthesis pathway confirms

```
ATGCGTCTCACCTCCCTGATTTCGATCTCAAACCTCGTGGTTCACCAATC    50
CCACGCCTTCCTAAATGGCGGGGCCTTCGTGGGCAACCAAATTGTCGGCA
GAGCTTCGACTTCGTTGACCAGAGGAGGAAGACAAGTCCTCAAAGCCGTG
GCCACCGCGCCAGTCGTCACGGAATCCCTCACGAAGGCTGCGGTTGAGGC   200
TCGCGGTTTAGCGATGGACTCCATCGGTTACTGTCACTCAGGTCACTTAG
GCTTACCCCTCGGCTGCGCGGAAATCGGGGCTGTTCTTTTCGGTCACGAT
TCGGGTCTGACTTACAACCCGGAAGACCCTCAATGGATCAACAGAGACAG
GTTTGTCCTTTCGGCAGGTCACGGATCAATGTTCTTGTACTCGTGGCTGC   400
ACCTCGCAGGATACGACCTCCCCTTAGAAGAGGTCAAGAACTTCCGTCGC
TTTCACTCGATGACCCCTGGTCACCCTGAATTTCCCAATTCTGAGCATAA
CACTCCTGGAGTTGAAGCTACAACTGGTCCGTTGGGTCAAGGAGTGGCCA
ACGCTGCTGGTATGGCTGCCGCTGCTAAAATGGCTGCTGCTAGGTTCAAC   600
ACACCATCGCATGAAATATTCAGCAATACGATTGTTGCGTTGTGTGGCGA
TGGATGCATTCAAGAAGGAGTTTCGTCTGAAGCGGCATGTTTCTCGGCTC
AAGAAGGCCTCGACAATCTTATCGTCGTGTATGACGCCAATGATGTGACC
CTTGATAAAATGGCTGAGTTCACTCAGAAGGAAGATGTTGCCAAAAGATA   800
CGAAGCGTATGGATGGGATGTTGTCACCATCGATGGTCATGATTTGAACG
CTGTTAGCAAAGCTTTGGCCGATGCCAAATCGAACAAAAATGGAAACCCA
AAGCTGATCATTGCAAAGACAGTGATCGGGAAAGGTATAAGTGCGATCGA
AGGCACGAATGCAGCACATGGTGAAGCTGGGGTTAAGTACGTGGATGAGT  1000
CAAGAAAAGAGTTAGGCTTACCTGAAGAGAAGTGGTACGTTTCACCCGAA
ACCCGCTCGTTCTTCAAAGAAAAAGCTTCCTCTCTAAAAGACACTTACAA
CAAATGGCAAGACACTTATAAAGCGTGGCAATCGGCGAATCCCGACCTTG
CCAAGCTTCTCCAAAGCGGAATCGACCAATCTACACCTTCCGCACAAGAA  1200
CTCCTCAAAGTCATCCCTGAGTTCGATCAAACTAAAAATATCGCTACTCG
CGAAGCTGGATCTGTTGTCCTGCAACCCCTGGCAGATGCCCTTCCTCTTT
ATGTAAGCGGAAGCGCTGACCTCCATGGCTCAAACAAAAACTACATTAAA
AACGGTGGGGACTTCGGTGTTAATTTCGGTAAATCGTACACGGGCCGCAA  1400
TTTCTACTACGGCATCAGAGAACATGCGATGGGCTCCATCCTCAACGGCA
TCAACTACTTCGGACTGTTTCGTGCTTCTGGTGCAACCTTCCTTGTGTTC
GCCGATTACCTCAGAGCAACCATCCGTGTGGCTTCCCTTGCTGAGCTCCC
AGTCGGCTACATCTTCACACACGACTCAATCGGCGTGGGAGAAGATGGGC  1600
CAACTCATCAACCTGTTGAGACAGTTTCGGGACTCAGAGTCATTCCAAAC
TTAGACGTGATACGCCCGGCTGACCCCGAAGAGACCGCTGCTGCTTACGC
TGCCTCCATCGAAAGAAAGGAGGGGCCTACTGCATTGATTTTAACCCGTC
AAAATGTGAGAACACTTTCAGAGATTCCTGTGTCAGTGAGGAGAGAAGGA  1800
ACACTGAAAGGCGCATACGTCGCCAAGAAAGAGACATCGGACTTGCAGCT
CATCATATTGGCTTCAGGCTCTGAAGTACAATGGGCCATGGACGTCGCTA
AAGACATCCCTGGCGCACGAGTTGTTTCCATGCCCTCAATGTATAGGTTT
GACATGCAACCTGAATCGTACAAGGAGTCGGTGTTGCCGGCGGCGTGCAC  2000
CAAGCGCATCTCAATCGAGGCTGGCGTCACTCCCCTTTGGTGGAAATATG
TAGGGACTGCTGGGAAAGTTTTGGGGACTGACAAATTCGGGTTGTCTGCG
CCCGGAGACAAAGTGATGGAAGTCTTTGGAATGGATTATAAAGGACTTAA
AAGGGAGGTCGAGGCCTACATGTAA 2175
```

Fig. 2. The location of four raw reads (highlighted in gray) in the *Elysia chlorotica* transcriptome within the *Vaucheria litorea* genomic transketolase coding region sequence (NCBI accession number JQ062398). All the reads have a 90 bp match with their respective regions except for the third one (bp 1857–1946), which has a single bp difference marked in bold.

Table 3. List of Algal Nuclear-Encoded Transcripts Identified in the *E. chlorotica* Transcriptome and Corresponding Sequence Matches in the Algal Transcriptome and Genome Data.

Protein Identity (top BLAST species hit)[a] NCBI Accession Number	*Elysia chlorotica* Transcript Identification Number[b]	*Vaucheria litorea* Transcriptome NCBI Accession Number[c]	*V. litorea* Genomic Coding Sequence NCBI Accession Number	bp/bp Match (*E. chlorotica*/ *V. litorea* sequence)
Phosphoribulokinase (*V. litorea*) AAK21910.1	44555290_d1	JP709200	JQ062392	90/90
	44555290_d2	JP709200		90/90
Light harvesting complex protein (*V. litorea*) ADD60136.1	33743048_d1	E52OB4B01DBNCL	JQ062393	90/90
	33743048_d2	E52OB4B02I8050		90/90
Carbonic anhydrase (*Ectocarpus siliculosus*) CBN77745.1	5998831_d1	E52OB4B01BKQIY	JQ062394	90/90
	5998831_d2	JP709201		90/90
Carbonic anhydrase α type (*E. siliculosus*) CBN76519.1	23892389_d1	JP709202	JQ062395	90/90
	23892389_d2	E52OB4B02JG2P9		90/90
Carbonic anhydrase (*Saccharina japonica*) AEF33616.1	4706782_d1	JP709203	JQ062396	47/47[d]
	4706782_d1	JP709203		90/90
Pyrophosphate-dependent phosphofructokinase (*E. siliculosus*) CBJ30811.1	22229246_d1	JP709204	JQ062397	88/90
	22229246_d2	JP709204		90/90
	22946887_d2	JP709204		90/90
Transketolase (*E. siliculosus*) CBJ48487.1	15424893_d1	JP709205	JQ062398	90/90
	15424893_d2	JP709205		90/90
	16607113_d1	JP709206		90/90
	16607113_d2	JP709206		89/90
Farnesyltransferase (*Phaeodactylum tricornutum*) XP_002178555.1	26003592_d1	JP709207	JQ062399	84/90
	26003592_d2	JP709207		90/90
Hsp-GrpE (*E. siliculosus*) CBN76646.1	20676998_d1	JP709208	JQ062400	89/90
	20676998_d2	JP709208		90/90
Chaperonin cpn60 (*E. siliculosus*) CBJ29911.1	17979526_d1	E52OB4B02GZ9VY	JQ062401	90/90
	17979526_d2	E52OB4B02GZ9VY		90/90
Molecular chaperones Hsp70/Hsc70 Hsp superfamily (*E. siliculosus*) CBJ32839.1	19589706_d1	JP709209	JQ062402	90/90
	19589706_d2	JP709209		90/90
Soluble pyridine nucleotide transhydrogenase (*E. siliculosus*) CBN79739.1	26709754_d1	JP709210	JQ062403	84/84[d]
	26709754_d2	JP709210		90/90
Methyltransferase type 11 (*Arthrospira maxima*) ZP_03275870.1	42534601_d1	JP709211	JQ062404	90/90
	42534601_d2	JP709211		90/90
Threonyl-tRNA synthetase (*Phytophthora infestans*) XP_002895146.1	43733749_d1	JP709296	JQ062405	90/90
	43733749_d2	JP709296		90/90
Rab8E.RAB family GTPase (*E. siliculosus*) CBN76228.1	32083728_d1	JP709212	JQ062406	89/90
	32083728_d2	JP709212		34/34[d]
Rab11A.RAB family GTPase (*E. siliculosus*) CBN78685.1	44639873_d1	E52OB4B01EGA6H	JQ062407	90/90
	44639873_d2	E52OB4B01D56HY		90/90
ABC transporter-like protein (*E. siliculosus*) CBN74513.1	39105431_d1	JP709213	JQ062408	90/90
	39105431_d2	JP709213		90/90
ATPase (*E. siliculosus*) CBJ25903.1	5321251_d1	E52OB4B01BQB96	JQ062409	90/90
	5321251_d2	None		87/90
Inorganic pyrophosphatase (*E. siliculosus*) CBJ29369.1	41485915_d1	E52OB4B01ES29L	JQ062410	90/90
	41485915_d2	E52OB4B01ES29L		90/90
Trigger factor (TF) (*E. siliculosus*) CBJ32857.1	35301097_d1	JP709214	JQ062411	89/90
	35301097_d2	JP709214		89/90
Cellulase 2 (*Pristionchus pacificus*) ADV58320.1	26209836_d1	E52OB4B02IOHXW	JQ062412	90/90
	26209836_d2	JP709297		85/85[e]
Tryptophan synthase (α/β chains) (*E. siliculosus*) CBN77109.1	38497509_d1	E52OB4B01AUQXK	JQ062413	90/90
	38497509_d2	E52OB4B01AUQXK		88/90
Tocopherol polyprenyltransferase-like protein (*Thalassiosira pseudonana*) XP_002293015.1	45450736_d1	JP709215	JQ062414	89/90
	45450736_d2	JP709215		90/90
Ribosomal protein L3 (*E. siliculosus*) CBJ28620.1	14972294_d1	E52OB4B02GOMSI	JQ062415	88/90
	14972294_d2	E52OB4B02GOMSI		90/90
Conserved unknown protein (*E. siliculosus*) CBN75765.1	600662_d1	JP709216	JQ062416	90/90
	600662_d2	JP709216		90/90
Conserved unknown protein (*E. siliculosus*) CBJ30937.1	20695943_d1	E52OB4B01BEOXI	JQ062417	90/90
	20695943_d2	E52OB4B01BEOXI		89/90
Conserved unknown protein (*E. siliculosus*) CBJ33840.1	926198_d1	JP709217	JQ062418	90/90
	926198_d2	JP709217		90/90
Conserved unknown protein (*E. siliculosus*) CBJ29091.1	48792667_d1	JP709218	JQ062419	90/90
	48792667_d2	JP709218		90/90
Conserved unknown protein (*E. siliculosus*) CBJ28899.1	9456896_d1	E52OB4B01C4II9	JQ062420	90/90
	9456896_d2	E52OB4B01C4II9		90/90

Table 3
Continued

Protein Identity (top BLAST species hit)[a] NCBI Accession Number	*Elysia chlorotica* Transcript Identification Number[b]	*Vaucheria litorea* Transcriptome NCBI Accession Number[c]	*V. litorea* Genomic Coding Sequence NCBI Accession Number	bp/bp Match (*E. chlorotica*/*V. litorea* sequence)
Conserved unknown protein (*E. siliculosus*)	12058179_d1	JP709219	JQ062421	89/90
CBJ26871.1	12058179_d2	JP709219		89/90
Conserved unknown protein (*E. siliculosus*)	13706224_d1	None	JQ062422	90/90
CBN79086.1	13706224_d2	E52OB4B02FYN72		90/90
Conserved unknown protein (*E. siliculosus*)	24090625_d1	JP709220	JQ062423	90/90
CBJ26383.1	24090625_d2	JP709220		90/90
Conserved unknown protein (*E. siliculosus*)	30948636_d1	JP709221	JQ062424	90/90
CBN76079.1	30948636_d2	JP709221		90/90
	41324700_d1	JP709221		89/90
	41324700_d2	JP709221		90/90
Conserved unknown protein (*E. siliculosus*)	19593353_d1	JP709222	JQ062425	89/90
CBN76436.1	19593353_d2	JP709222		90/90
Conserved unknown protein (*E. siliculosus*)	3907773_d1	E52OB4B01CTM1Z	JQ062426	90/90
CBN74931.1	3907773_d2	E52OB4B02IV83X		90/90
Conserved unknown protein (*E. siliculosus*)	12842227_d1	JP709223	JQ062427	90/90
CBN78635.1	12842227_d2	JP709223		90/90
Conserved unknown protein (*E. siliculosus*)	26886148_d1	JP709224	JQ062428	90/90
CBN74099.1	26886148_d2	JP709224		90/90
Conserved unknown protein (*E. siliculosus*)	48772180_d1	JP709225	JQ062429	90/90
CBJ30349.1	48772180_d2	JP709225		90/90
Expressed unknown protein (*E. siliculosus*)	20219847_d1	JP709226	JQ062430	90/90
CBJ32627.1	20219847_d2	JP709226		89/90
Putative sulfate permease family protein (*E. siliculosus*)	6037510_d1	JP709227	JQ062431	90/90
CBN74518.1	6037510_d2	JP709227		90/90
Hypothetical protein Esi 0531 0003 (*E. siliculosus*)	8333052_d1	E52OB4B01BAHK0	JQ062432	90/90
CBJ33624.1	8333052_d2	E52OB4B01BAHK0		87/90
	30302247_d1	E52OB4B01BAHK0		90/90
	30302247_d2	E52OB4B01BAHK0		86/90
Hypothetical protein Esi 0531 0003 (*E. siliculosus*)	42187880_d1	None	JQ062433	90/90
CBJ33624.1	42187880_d2	None		90/90
Hypothetical protein THAPSDRAFT_22565	19828103_d1	None	JQ062434	90/90
(*T. pseudonana*) XP_002290531.1	19828103_d2	None		90/90
Similar to CG7697-PA (*E. siliculosus*) CBN77571.1	15303709_d1	JP709228	JQ062435	82/83[d]
	15303709_d2	JP709228		90/90
Putative NIN-like transcription factor	22750097_d1	JP709298	JQ062436	90/90
(*E. siliculosus*) CBN75385.1	22750097_d2	E52OB4B01AL89S		89/90
Predicted protein (*P. tricornutum*)	21613523_d1	None	JQ062437	90/90
XP_002179782.1	21613523_d2	None		90/90
No BLAST hit	6290510_d1	JP709229	JQ062438	88/90
	6290510_d2	JP709229		90/90
No BLAST hit	36768416_d1	None	JQ062439	90/90
	36768416_d2	None		44/44[d]
No BLAST hit	34498593_d1	E52OB4B02IHZHE	JQ062440	90/90
	34498593_d2	E52OB4B02IHZHE		90/90
No BLAST hit	39685080_d1	None	JQ062441	90/90
	39685080_d2	E52OB4B01CY5TH		90/90
No BLAST hit	16533926_d1	None	JQ062442	26/26[d]
	16533926_d2	None		89/90
No BLAST hit	35063098_d1	E52OB4B02G8Q7G	JQ062443	90/90
	35063098_d2	E52OB4B02G8Q7G		90/90

[a] Cutoff = 10^{-10}.
[b] All EC Illumina reads in this column are included in Sequence Read Archive Study number SRP009263.2. The labels in this column are identifiers used in supplementary table 1, Supplementary Material online.
[c] Singleton reads labeled E52OB4B0xxxxxx are included in Sequence Read Archive Study number SRP009267.2.
[d] Fragment extended beyond the stop codon.
[e] Fragment extended beyond the start codon.

the synthesis of chlorophyll a by *E. chlorotica* (Pierce et al. 2009). Of particular interest, chloroplast-encoded transcripts for a few chaperone and other protein-processing proteins were present in the slug RNA. This is the first time that any evidence for the synthesis of these kinds of chloroplast proteins has been found in a host cell, and they may provide clues to the underlying mechanisms of long-term plastid function in the foreign cytoplasm.

Rarer than the chloroplast-encoded sequences in the *E. chlorotica* transcriptome were transcripts that matched *V. litorea* nuclear-encoded genes for chloroplast proteins. All of these transcripts matched coding sequences in the algal genome and most were present in one or both of our algal transcriptome data sets. Annotation revealed genes coding for Calvin cycle enzymes essential for the continuation of photosynthesis, like algal specific sequences for *prk*, which had been found by PCR previously, as well as for carbonic anhydrase, *pfk* and *tkt*. One of the *V. litorea* thylakoid membrane light harvesting components, *lhc*, as well as *thf1* (thylakoid formation protein) were also present in the slug sequence data. Both of these, along with the other *lhc*'s found previously by PCR (Pierce et al. 2007; Schwartz et al. 2010) are important for the longevity and maintenance of the sequestered plastid. As was the case with the chloroplast-encoded transcripts, several algal nuclear-encoded sequences for molecular chaperones or other protein-processing enzymes, including a trigger factor homologue, which is involved in protein folding, appeared in the data. One of the more interesting aspects of this intracellular symbiosis is the biochemical mechanisms that integrate the chloroplast and host cell. While mostly unknown at present, the mechanisms that prolong chloroplast survival in the slug cells may rely on proteins such as those.

Interestingly, 23 *E. chlorotica* transcripts that matched nuclear-encoded *V. litorea* sequences varied slightly from the native algal reference sequences by 1–6 bp. Nucleotide substitutions have also been observed in *uroD* (uroporphyrinogen decarboxylase) located in slug cDNA and gDNA in earlier studies (Pierce et al. 2009). These differences suggest that at least some of the transfer occurred sufficiently long ago to permit some of divergence to occur between the native algal sequences and the corresponding sequences in the sea slugs.

Several other non–chloroplast-encoded sequences in the *E. chlorotica* transcripts were annotated in the "conserved-", "unknown-," or "hypothetical protein" categories. It is not presently possible to determine their functional significance without more sequence information. However, almost all of those listed in those categories of our data set had the *E. siliculosus* genome as the top BLAST species hit, indicating that they are most likely of algal origin.

Thus, a variety of algal nuclear-encoded transcripts are present in the *E. chlorotica* transcriptome in spite of the two earlier reports to the contrary (Pelletreau et al. 2011; Wägele et al. 2011). Several important issues need to be considered to resolve these dramatic differences. First, different slug species were used by Wägele et al. (2011) and, except that it is indeed long standing, beyond the demonstration of photosynthesis (Wägele and Johnsen 2001), almost nothing is known about the chloroplast symbiosis in *P. ocellatus*, including the algal species that donates the chloroplasts. Second, the *P. ocellatus* RNA was extracted from only a single specimen under undescribed conditions that may or may not have maximized the presence of transcripts for chloroplast proteins.

Third, a bit more is known about the chloroplast symbiosis in the other species, *E. timida*, used by Wägele et al. (2011). Although it can last for several months, the chloroplast maintenance strategy of this species is behavioral shading of the chloroplast-residing body regions by means of the parapodia, rather than any biochemical accommodations (Rahat 1976; Rahat and Monselise 1979; Casalduero and Muniain 2008; Schmitt and Wägele 2011). Also, while these species are fairly distant from each other phylogenetically, in both, the anatomical location of the symbiotic plastids is similar. Namely, they are mostly located in a dorsal patch, where the digestive tubules rise to just below the epithelial surface between the two thick parapodial flaps, which can be extended over the chloroplast regions, effectively shading them from light (Wägele et al. 2011). Without some sort of biochemical replacement, continual exposure to light will cause both the thylakoid components and photopigments to degrade. Indeed, degraded chloroplasts and rapid chlorophyll loss are evident in starved *E. timida* and parapodial shading behavior occurs in response to light (Marín and Ros 1989, 1992). About 10% of the symbiotic plastids are replaced over a 4-day period if food is available, so *E. timida* seems to rely on feeding and behavioral responses to sustain a fraction of its symbiotic plastids long term and may not be an appropriate species to test for horizontally transferred algal genes. Less is known with respect to *P. ocellatus*, but the anatomy and similar shading behavior suggest chloroplast protection mechanisms similar to *E. timida*. In *E. chlorotica*, on the other hand, the chloroplast containing ends of the digestive tubules underlie the epithelium over the entire surface of the slug, including the parapodia, so many of the plastids are in anatomical locations where shading is not possible (http://www.seaslugforum.net/find/13756). So among the sacoglossans, more than one mechanism of plastid maintenance occurs (see also Pierce and Curtis 2012) and not all may require algal genes, as in *E. chlorotica*.

Fourth, Wägele et al. (2011) assumed that, as in plant cells, there would be a very large chloroplast-targeted transcript presence in the slug cells and thus, easily found. However, there is no basis for this assumption and, at least in *E. chlorotica*, photosynthetic gene expression is much slower than in the alga (Soule 2009).

Fifth, the symbiotic chloroplasts reside in only a few cells within the slugs. However, Wägele et al. (2011) extracted the RNA from the entire animal in both slug species. Thus, almost all the RNA in their analyses came from cells that do

not have chloroplasts and may not make transcripts for plastid proteins.

Sixth, in the case of both species, Wägele et al. (2011) examined only a small fraction of the transcriptomes and attempted to annotate them by comparison to the RefSeq database, which contains few algal sequences and none from Udoteacea, the algal taxon that seems to be the source of the symbiotic plastids. Not only did this process fail to find any algal nuclear gene-like sequences, it returned only about 6,000 BLAST hits from more than 77,000 assembled contigs from the *P. ocellatus* data, for example, and of those, only 79 had a top homologue hit among plants. Of some importance, 29 of these latter matched chloroplast-encoded genes, indicating some plastid genome translational activity. The remaining 50 sequences matched genes "widely distributed among eukaryotes." The *E. timida* transcriptome analysis returned only two chloroplast-encoded gene matches (Wägele et al. 2011). Perhaps more information would be found by comparison to data from species phylogenetically closer to the actual chloroplast sources, especially since sequence conservation between species among several of the transferred algal nuclear genes found in *E. chlorotica* by PCR is low and do not hit anything in the genomic data from *Arabidopsis*, *Chlamydomonas*, or *Volvox* (Pierce et al. 2007; Rumpho et al. 2008; Schwartz et al. 2010). So while the conclusion that neither *P. ocellatus* nor *E. timida* contain transferred algal genes to support their symbiotic chloroplasts may ultimately be correct, more evidence in addition to that presented by Wägele et al. (2011) seems necessary to establish that.

Finally, it is clear from our results that finding rare transcripts requires a relatively large sequencing effort and identifying foreign sequences from phylogenetically distant species with little representation in the sequence databases is facilitated by knowledge of native sequence. Sweeping conclusions based on small partial transcriptome analyses can be misguided. Like the other two sacoglossan studies, our RNA came from whole animal extracts, so most of it was from cells that do not contain chloroplasts. In addition, the rarity of the transcripts in our data and the low rate of expression in the slug relative to the alga (Soule 2009; Pelletreau et al. 2011) suggest a low expression rate. Taken together, a low expression rate by a relatively few cells would conspire against finding the sequences in limited amounts of data obtained from small amounts of RNA extracted mostly from nonexpressing cells. Even our large amount of data (8.8 Gbp) only returned a weak coverage of a few, mostly not overlapping, 90 bp fragments that did not occur in sufficient number or variety to be formed into contigs by the software. Indeed, the much smaller (148 Mbp) *E. chlorotica* transcriptome data set was unsuccessful at locating transferred genes (Pelletreau et al. 2011). No doubt, there are more transferred algal genes to be found in the *E. chlorotica* DNA, which will require even more sequencing.

We are presently working on *E. chlorotica* genome sequencing which may be the only way to get an estimate of the entire transferome. Although the initial discovery of transferred genes in the slug genome was done with PCR techniques and, as a result of the specificity of PCR, returned only a few genes (Pierce et al. 2007; Rumpho et al. 2008), the transcriptome analysis here indicates that many algal nuclear genes have somehow arrived into the *E. chlorotica* genome, and it is likely that others will be found. The size of the transferome suggests the possibility that pieces of DNA, rather than single genes, on multiple occasions may be involved in the transfer and integration process, but comparative genomics may provide the important answer to that.

Supplementary Material

Supplementary tables 1 and 2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The portion of the work done at the University of South Florida was funded by a private donor, who wishes to remain anonymous. The portion of the work done at BGI-Shenzhen was done with funds provided by the BGI Science and Technology Foundation.

References

Al-Samarrai TH, Schmid J. 2000. A simple method for extraction of fungal genomic DNA. *Lett Appl Microbiol.* 30:53–56.

Altschul SF, Gish W, Miller W, Myer EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.

Birney E, Clamp M, Durbin R. 2004. GeneWise and Genomewise. *Genome Res.* 14:988–995.

Casalduero FG, Muniain C. 2008. The role of kleptoplasts in the survival rates of *Elysia timida* (Risso, 1818) (Sacoglossa: opisthobranchia) during periods of food shortage. *J Mar Biol Ecol.* 357:181–187.

Darling A, Carey L, Feng W. 2003. The design, implementation, and evaluation of mpiBLAST. In: Proceedings of the ClusterWorld Conference Expo and the 4th International Conference on Linux Cluster: The HPC Revolution; June, 2003; San Jose, CA.

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

Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Res.* 12:656–664.

Li R, Fan W, Tian G, et al. (116 co-authors). 2010. The sequence and *de novo* assembly of the giant Panda genome. *Nature* 463(7279):311–317.

Li R, Zhu H, Ruan J, et al. (14 co-authors). 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20:265–272.

Lin H, Ma X, Chandramohan P, Geist A, Samatova N. 2005. Efficient data access for parallel BLAST. In: Proceedings of 19th IEEE International Parallel and Distributed Processing Symposium; 2005 Apr 3–8; Denver, CO. Los Alamitos (CA): IEEE Computer Society Conference Publishing Services.

Marín A, Ros J. 1989. The chloroplast-animal association of four Iberian Sacoglossan opisthobranchs: *elysia timida, Elysia translucens, Thuridilla hopei* and *Bosellia mimetica*. In: Ros J, editor.

Topics in marine biology. Proceedings of the 22nd European Marine Biology Symposium. *Sci Mar*. 53:429–440.

Marín A, Ros J. 1992. Dynamics of a peculiar plant–herbivore relationship: the photosynthetic ascoglossan *Elysia timida* and the chlorophycean *Acetabularia acetabulum*. *Mar Biol*. 112: 677–682.

Mujer CV, Andrews DL, Manhart JR, Pierce SK, Rumpho ME. 1996. Chloroplast genes are expressed during intracellular symbiotic association of *Vaucheria litorea* plastids with the sea slug *Elysia chlorotica*. *Proc Nat Acad Sci*. 93:12333–12338.

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

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

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

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

Pierce SK, Curtis NE, Schwartz JA. 2009. Chlorophyll a synthesis by an animal using transferred algal nuclear genes. *Symbiosis* 49: 121–131.

Rahat M. 1976. Direct development and symbiotic chloroplasts in *Elysia timida* (Mollusca: Opisthobranchia). *Israel J Zool*. 25: 186–193.

Rahat M, Monselise EBI. 1979. Photobiology of the chloroplast-hosting mollusc, *Elysia timida* (Opisthobranchia). *J Exp Biol*. 79: 225–233.

Rumpho ME, Pochareddy S, Worful JM, Summer EJ, Bhattacharya D, Pelletreau KN, Tyler MS, Lee J, Marhart JR, Soule KM. 2009. Molecular characterization of the Calvin cycle enzyme phosphoribulokinase in the stramenopile alga *Vaucheria litorea* and the plastid hosting mollusc *Elysia chlorotica*. *Mol Plant*. 2:1384–1396.

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

Schmitt V, Wägele H. 2011. Behavioral adaptations in relation to long-term retention of endosymbiotic chloroplasts in the sea slug *Elysia timida* (Opisthobranchia, Sacoglossa). *Thalassas* 27:225–238.

Schwartz JA, Curtis NE, Pierce SK. 2010. Using algal transcriptome sequences to identify transferred genes in the sea slug, *Elysia chlorotica*. *Evol Biol*. 37:29–37.

Soule KM. 2009. Light-regulated photosynthetic gene expression and enzyme activity in the heterokont alga *Vaucheria litorea* and its symbiotic partner the sacoglossan mollusc, *Elysia chlorotica* [MS thesis]. [Orono (ME)]: University of Maine. p. 120.

Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19(Suppl 2): ii215–ii225.

Wägele H, Deusch O, Händeler K, et al. (11 co authors). 2011. Transcriptomic evidence that longevity of acquired plastids in the photosynthetic slugs *Elysia timida* and *Plakobranchus ocellatus* does not entail lateral transfer of algal nuclear genes. *Mol Biol Evol*. 28:699–706.

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