

A reference genome for the long-term kleptoplast-retaining sea slug *Elysia crispata* morphotype clarki

Katharine E. Eastman ,^{1,2} Amanda L. Pendleton ,^{1,2} Mearaj A. Shaikh ,^{2,3} Thiti Suttiyut ,^{2,3} Raeya Ogas,^{1,2} Paxton Tomko ,^{2,4} Gregory Gavelis ,^{1,2} Joshua R. Widhalm ,^{2,3,*} Jennifer H. Wisecaver ,^{1,2,*}

¹Department of Biochemistry, Purdue University, West Lafayette, IN 47907, USA

²Purdue Center for Plant Biology, Purdue University, West Lafayette, IN 47907, USA

³Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN 47907, USA

⁴Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA

*Corresponding author: Department of Biochemistry, Purdue University, 175 S University St., West Lafayette, IN 47907, USA. Email: jwisecav@purdue.edu;

*Corresponding author: Department of Horticulture and Landscape Architecture, Purdue University, 625 Agriculture Mall Dr, West Lafayette, IN 47907, USA. Email: jwidhalm@purdue.edu

Several species of sacoglossan sea slugs possess the incredible ability to sequester chloroplasts from the algae they consume. These "photosynthetic animals" incorporate stolen chloroplasts, called kleptoplasts, into the epithelial cells of tubules that extend from their digestive tracts throughout their bodies. The mechanism by which these slugs maintain functioning kleptoplasts in the absence of an algal nuclear genome is unknown. Here, we report a draft genome of the sacoglossan slug *Elysia crispata* morphotype clarki, a morphotype native to the Florida Keys that can retain photosynthetically active kleptoplasts for several months without feeding. We used a combination of Oxford Nanopore Technologies long reads and Illumina short reads to produce a 786-Mb assembly (N50 = 0.459 Mb) containing 68,514 predicted protein-coding genes. A phylogenetic analysis found no evidence of horizontal acquisition of genes from algae. We performed gene family and gene expression analyses to identify *E. crispata* genes unique to kleptoplast-containing slugs that were more highly expressed in fed versus unfed developmental life stages. Consistent with analyses in other kleptoplastic slugs, our investigation suggests that genes encoding lectin carbohydrate-binding proteins and those involved in regulation of reactive oxygen species and immunity may play a role in kleptoplast retention. Lastly, we identified four polyketide synthase genes that could potentially encode proteins producing UV- and oxidation-blocking compounds in slug cell membranes. The genome of *E. crispata* is a quality resource that provides potential targets for functional analyses and enables further investigation into the evolution and mechanisms of kleptoplasty in animals.

Keywords: *Elysia crispata*; Oxford Nanopore Technologies; MinION; long-read assembly; gene expression; comparative genomics; Mollusca; Gastropoda

Introduction

The ability to feed on algae and temporarily sequester functional chloroplasts (called kleptoplasts) has convergently evolved multiple times in invertebrates (Van Steenkiste et al. 2019). In no lineage is this strategy more common than in the gastropod superorder Sacoglossa, a group of sap-sucking sea slugs (Händeler et al. 2009; Maeda et al. 2021). In slug digestive cells, phagocytosed algal material is selectively degraded, leaving only kleptoplasts intact, which provide the slug with nitrogen and fixed carbon (Trench et al. 1974; Raven et al. 2001; Curtis et al. 2010; Cruz et al. 2020). The length of time kleptoplasts remain photosynthetically active differs between slug species, ranging from a few days in species such as *Elysia cornigera* to over eleven months in *Plakobranchus ocellatus* (Clark et al. 1990; Evertsen et al. 2007; Händeler et al. 2009). Species that can incorporate functional kleptoplasts for up to 2 weeks of starvation are referred to as short-term-retention forms,

and those that retain functional kleptoplasts for over 20 days are considered long-term forms (Christa, Händeler, et al. 2014). Phylogenetic analyses indicate that the ability to retain short-term and long-term kleptoplasts has evolved multiple times in sacoglossan slugs (Christa et al. 2015; Maeda et al. 2021). Moreover, kleptoplasts acquired from the same algal donor species can have wildly different retention times in different slug species. For example, *Elysia crispata* (synonym *Elysia clarki*) retains long-term kleptoplasts for several months from multiple algal donors, including *Penicillius capitatus*, whereas *Elysia patina* also harbors chloroplasts from *P. capitatus*, but the organelles are fully degraded within 2 weeks (Curtis et al. 2010). This suggests that slug-dependent factors contribute to kleptoplast stability and retention. Yet the mechanisms by which sacoglossan sea slugs maintain functional kleptoplasts for extended periods of time remain unclear.

In algae, most chloroplast genes are encoded in the nuclear genome, and key proteins for photosynthesis such as photosystem

Received on 07 July 2023; accepted on 22 September 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of The Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

subunits and light harvesting antennae are synthesized in the cytoplasm and subsequently imported into the chloroplast (Martin and Herrmann 1998; Nowack and Weber 2018). Kleptoplast longevity was, therefore, once hypothesized to stem from horizontal gene transfer (HGT) of algal nuclear-encoded genes to the nuclear genome of kleptoplastic slugs with subsequent targeting of encoded proteins to the kleptoplast (Rumpho et al. 2008; Pierce et al. 2012). However, transcriptome and genome level analyses of multiple species of kleptoplastic sacoglossans have found no evidence of HGT from algae (Wägele et al. 2011; Bhattacharya et al. 2013; Maeda et al. 2021). Similarly, it was hypothesized that kleptoplastic slugs may translate proteins from retained algal mRNAs to control turnover, stability, and/or repair damaged kleptoplasts (Pierce and Curtis 2012), but a recent transcriptomic analysis of *P. ocellatus* found no evidence to support this “transferome” hypothesis (Maeda et al. 2021).

Combined, the lack of evidence for horizontally acquired algal genes or transcripts suggests that evolved nuclear-encoded slug proteins play a role in prolonged kleptoplast retention. For example, kleptoplastic slugs produce propionate pyrones that may function as “screens” in cell membranes (Powell et al. 2018). These compounds are de novo synthesized from carbon dioxide by slugs in light (Ireland and Scheuer 1979) and are predicted to help limit the formation of damaging reactive oxygen species (ROS) produced by excessive light absorption. Recently, Torres et al. (2020) discovered an *Elysia chlorotica* polyketide synthase (EcPKS1) and showed in vitro that it can produce propionate pyrone precursors from methylmalonyl-CoA. The authors also found that less UV-reactive 5-propionate products occur in nonphotosynthetic and short-term kleptoplastic slugs while longer 7- and 8-propionate pyrones are restricted to slugs performing long-term kleptoplasty, which suggests that novel PKSs have evolved in slugs capable of long-term kleptoplasty.

Genome assemblies are available for 2 long-term-retention species, *E. chlorotica* and *P. ocellatus*, as well as 1 short-term species, *Elysia marginata* (Cai et al. 2019; Maeda et al. 2021). To gain a better understanding of the evolution and underlying mechanisms of kleptoplasty in sacoglossan sea slugs, genomic data across additional species are necessary. Here, we report a high-quality draft genome assembly of *E. crispata* morphotype clarki (Fig. 1a and b), a kleptoplastic sea slug native to mangrove swamps and waterways of the Florida Keys (Krug et al. 2016) that can retain long-term functional kleptoplasts for 3 months or more (Middlebrooks et al. 2012). A phylogenetic approach was used to determine whether the *E. crispata* genome contains any genes that appear horizontally acquired from algae. Gene expression data and comparative genome analysis identified gene candidates that may be involved in prolonging the life of kleptoplasts. Taken together, the *E. crispata* genome offers a quality resource for exploring the evolution and mechanisms of long-term kleptoplast retention in Sacoglossa.

Materials and Methods

Slug and algal material and laboratory culturing in aquaria

Wild-caught *E. crispata* were purchased from a local aquarium shop in Lafayette, Indiana (United States). The slugs were shipped overnight to the shop after being collected from the vertical walls of an undisclosed canal in the Florida Keys (United States).

Slugs were acclimated in 20 L aquaria containing artificial sea-water (ASW) prepared using distilled water and Instant Ocean Reef Crystals at a specific gravity of 1.023 at 25°C (77°F). Slugs were then transferred to connected 180–200 L aquaria containing

ASW recirculated through a 280 L sump tank and a Pentair SMART UV sterilizer to reduce microorganismal growth. The water temperature was maintained at 25°C using a standard aquarium heater. Aquaria were illuminated with white LED lights (Fluval Sea Marine 3.0 LED Aquarium Lights set at 75% of maximum intensity) on a 12:12 light:dark cycle. Every week, ~50% of the system volume was replaced with fresh ASW, and freshly cultured algae was introduced. Aquaria were monitored daily for the deposition of egg masses. Egg masses were collected within 1 day of deposition.

Macroalgae were cultured on 15 × 30 × 1 cm gridded plastic racks in a separate 220 L aquarium containing recirculated ASW filtered through a 20 L sump tank. To help promote algal growth, a constant turbulent water flow was maintained using 3 circulation pumps positioned across the top, middle, and bottom of the aquarium in alternating directions. Weekly maintenance consisted of treating the water with 50 µg mL⁻¹ rifampicin and changing out 100% of the ASW 24 h later. Aquarium ASW was supplemented 3 times per week with 1 h of gentle bubbling of compressed carbon dioxide and dosing with 100 mL each of 20 g L⁻¹ potassium nitrate (Thermo Fisher, Waltham, MA, United States), 1.5 g L⁻¹ potassium dihydrogen phosphate (Thermo Fisher, Waltham, MA, United States), and 1:20 diluted Guillard's F/2 trace elements (Algae Research Supply, San Diego, CA, United States).

Sampling, RNA extraction, and RNA sequencing and processing

Wild-caught adult slugs deposited egg masses ~2–3 weeks after being introduced in the laboratory aquaria. Egg samples (Fig. 1d) for RNA extraction were rinsed in fresh ASW, flash-frozen in liquid nitrogen, and stored at -80°C within 24 h of deposition. Each biological replicate was derived from using the entire egg mass deposited by an individual slug.

To obtain free-swimming veligers, crawling larvae, and juvenile slugs, each remaining collected and rinsed egg mass was transferred to an individual petri dish containing 25 mL of freshly prepared ASW and supplemented with 5 µg mL⁻¹ rifampicin and 500 ng mL⁻¹ ivermectin to eliminate the potential growth of cyanobacteria, bacterial pathogens, and predators. Individual egg masses were maintained at room temperature under white LED light (Fluval Sea Marine 3.0 LED Aquarium Lights set at 75% of maximum intensity) on a 12:12 light:dark cycle. Egg masses were carefully rinsed and filtered through 52-micron nylon mesh and transferred to new ASW with rifampicin and ivermectin every 2 days. Free-swimming veliger larvae (Fig. 1e) hatched ~21–28 d after egg deposition and were collected by centrifuging the ASW media at 1000 × g for 1 min. Pellets of free-swimming veliger samples were rinsed in ASW, flash-frozen in liquid nitrogen, and stored at -80°C until RNA extraction. Each biological replicate was derived from using all the free-swimming veligers that hatched from a single egg mass deposited by an individual slug. The hatchlings from the remaining unused egg masses were allowed to metamorphose into larvae. Crawling larvae (Fig. 1f) were maintained as pools derived from individual egg masses and kept in petri dishes as described for egg masses. Crawling larvae samples were prepared as described for veliger samples ~3–5 d after metamorphosis. Each biological replicate was derived from using all the larvae derived from a single egg mass. The larvae derived from the remaining unused egg masses were introduced into 16 × 20 cm Pyrex dishes containing freshly prepared ASW, 5 µg mL⁻¹ rifampicin, 500 ng mL⁻¹ ivermectin, and ~1 g of macroalgae. After reaching a length of ~1–2 cm, juvenile slugs (Fig. 1c and g) were collected, flash-frozen in liquid nitrogen, and

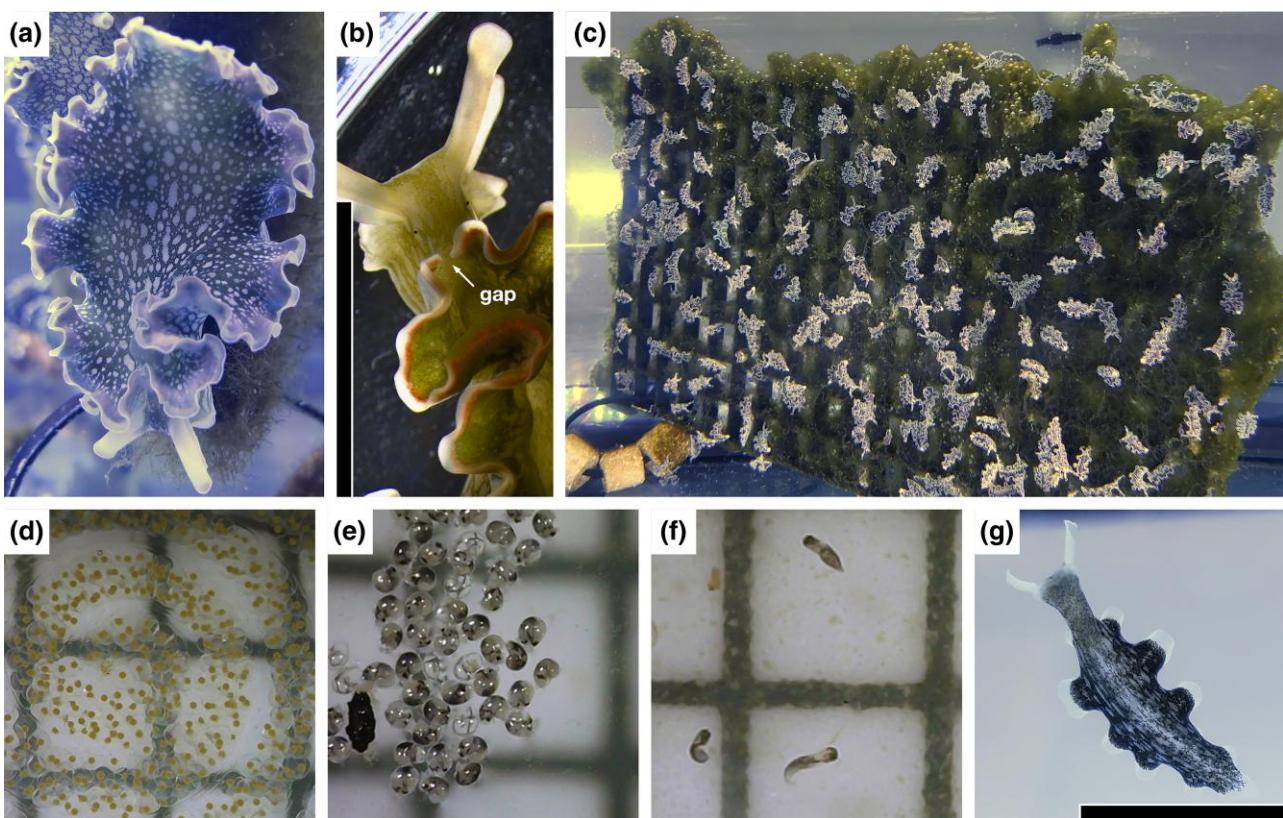


Fig. 1. *Elysia crispata* morphotype clarki. a) Wild-caught adult slug, dorsal view, showing near-uniform dark coloration with small round light spots indicative of clarki morphotype. b) Close up of wild-caught adult slug showing parapodial gap indicative of clarki morphotype. c) Lab-reared cornucopia of slugs. d) Egg clutch 24-h postdeposition. e) Free-swimming veliger. Larvae. f) Crawling larvae. g) Lab-reared juvenile slug, ventral view, showing transparent and tapered foot indicative of clarki morphotype; pigmentation indicates presence of kleptoplasts. Scale bar in b and g = 10 mm. Scale grid in d, e, and f = 2 x 2 mm.

stored at -80°C until RNA extraction. Each biological replicate consisted of 3 individual slugs.

Total RNA was isolated from 100 mg of pulverized flash-frozen *E. crispata* eggs, free-swimming veligers, crawling larvae, or juvenile slugs using TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, United States) according to the manufacturer's protocol. Extracted RNA samples were digested with DNaseI (New England Biolabs, Ipswich, MA, United States) and column purified using an RNA concentrator and cleanup system (Zymo Research, Irvine, CA, United States). The eluted RNA samples were quantified using a NanoDrop spectrophotometer. Purified samples were measured for RNA integrity (RIN) scores at the Purdue Genomics Core Facility. Samples with RIN scores ≥ 7 were selected for library prep. Three biological replicates of each *E. crispata* developmental stage of the slug were sent to Novogene Corporation Inc. (Sacramento, CA, United States) for library construction (NEBNext Ultra RNA Library Prep Kit, New England Biolabs) from 1 μg total RNA and Illumina sequencing. Paired-end (2×150 bp) reads were quality filtered and adapter trimmed using fastp v0.20.1 (Chen et al. 2018).

Sampling, DNA extraction, and genome sequencing and assembly

Genomic DNA (gDNA) for Illumina sequencing was extracted from *E. crispata* eggs reared in petri dishes (without algae). Samples were pulverized in liquid nitrogen using a TissueLyser II (Qiagen, Hilden, Germany). DNA was extracted using 1 of 2 methods, a CTAB-based protocol for samples with high mucopolysaccharide

content (dx.doi.org/10.17504/protocols.io.kxygxp81z18j/v1) modified from Winnepenningckx et al. (1993) or a DNAzol Reagent (Invitrogen, Waltham, MA, United States) protocol (dx.doi.org/10.17504/protocols.io.n92ldprrnl5b/v1). Quantity and quality of DNA for each sample were assessed using a Qubit 4 fluorometer (Invitrogen, Waltham, MA, United States) and a TapeStation 4150 (Agilent, Santa Clara, CA, United States). No difference in quality between the CTAB- and DNAzol-based protocols was observed, and the samples were pooled prior to Illumina sequencing. The sequencing library was constructed using an NEBNext DNA library prep kit (New England Biolabs), and 2×150 -bp paired-end reads were sequenced using an Illumina NovaSeq 6000 by Novogene Corporation Inc. (Sacramento, CA, United States). Illumina gDNA read quality was assessed by FastQC v0.10.0 (Babraham Bioinformatics 2011). Illumina TruSeq adapters and low-quality reads were removed with fastp v0.20.1 (Chen et al. 2018) using default parameters.

Illumina reads derived from suspected bacterial contamination were identified using BlobTools v1.1.1 (Laetsch and Blaxter 2017). A first-pass Illumina-only genome assembly was performed by Abyss v2.2.4 (Simpson et al. 2009) using a k-mer size of 96. Illumina gDNA reads were aligned to the Abyss assembly using BWA-MEM v0.7.15 (Li and Durbin 2009) to generate a coverage BAM file. Abyss contigs were queried against the NCBI nucleotide (nt) database (accessed 11 September 2021) using blastn v2.11.0 (Camacho et al. 2009). DIAMOND v2.0.8.146 (Buchfink et al. 2015) was used to query Abyss contigs against a custom protein databases that consisted of NCBI RefSeq release 207 (O'Leary et al.

2016) sequences supplemented with additional predicted protein sequences from MMETSP (Keeling et al. 2014) and the 1000 Plants transcriptome sequencing project (1KP) (Matasci et al. 2014). The custom protein database used in the BlobTools analysis is available from the authors as well as through the following link: <https://www.datadepot.rcac.purdue.edu/jwisejav/custom-refseq/2021-08-02>. The BlobTools taxrule “bestsumorder” determined the taxonomic assignment of each contig, prioritizing information from protein hits first. Contigs denoted as noneukaryotic in origin were flagged, and BBSplit v38.87 (Bushnell 2017) was used to exclude Illumina reads that mapped to these contigs. This step removed 3.4% of reads, as most contigs were either taxonomically unassigned or determined to be of eukaryotic origin (Fig. S1). Estimates of genome size and heterogeneity were calculated using GenomeScope2.0 (Ranallo-Benavidez et al. 2020).

gDNA for Oxford Nanopore Technologies (ONT) sequencing was extracted from 2 *E. crispata* sample types: brains dissected from wild adults and whole bodies of wild adults starved for a minimum of 2 weeks (Table S1). Both the CTAB and DNAzol protocols referenced above were used to generate high-molecular-weight DNA for ONT long-read sequencing (Table S1). Short DNA fragments less than 10 kb were depleted using an SRE XS kit (Circulomics, Baltimore, MD, United States) according to the manufacturer’s protocol. At least 1.9 µg of high-molecular-weight gDNA was used as input for ONT LSK-109 library ligation kits and sequenced on R9 MinION flow cells. Six flow cells were necessary to generate sufficient read depth for whole genome assembly (Table S1). Base calling of ONT reads was performed with Guppy v6.1.2 (Oxford Nanopore Technologies, Oxford, United Kingdom). ONT reads were quality filtered by Filtlong (Wick 2022) using the Illumina reads as an external reference and with the following parameter settings: –trim –split 500 –keep_percent 90 –min_length 3000.

Two programs for genome assembly were evaluated: CANU v2.2 (Koren et al. 2017) using an estimated genome size of 711 Mb and Flye v2.9 (Kolmogorov et al. 2019) using the –keep-haplotypes and –nano-raw parameter settings. Three programs for error correction and genome polishing were evaluated: Racon v1.5.0 (Vaser et al. 2017), Medaka v1.6.0 (nanoporetech 2022), and NextPolish v1.4.0 (Hu et al. 2020). Racon and Medaka polishing runs were performed using ONT reads only. NextPolish polishing runs were performed using ONT and Illumina reads as references. Racon polishing was iteratively performed up to 4 times. All polishing programs were run using default parameters. Polished assemblies were compared to each other and to the unpolished starting assemblies using QUAST-LG v5.2.0 (Mikheenko et al. 2018) and BUSCO v4.0.6 (Manni et al. 2021) run in genome mode with the eukaryote_odb10, metazoan_odb10, and mollusca_odb10 lineage data sets. The Flye assembly was selected as the final published assembly as it resulted in longer contigs (Table S2) and better BUSCO recovery compared to CANU (Table S3). Polishing results were comparable across the different iterations. No single iteration performed best for all BUSCO odb10 data sets: 3 rounds of Racon polishing yielded the best BUSCO recovery of the eukaryote_odb10 data set (96.08% complete and single copy), 1 round of Racon polishing followed by Medaka and NextPolish yielded the best BUSCO recovery of the metazoan_odb10 data set (93.82% complete and single copy), and a single round of Medaka polishing yielded the best BUSCO recovery of the mollusca_odb10 data set (87.76% complete and single copy). Therefore, the 3 best iterations were all carried forward to gene annotation and protein prediction (see Genome and gene annotation methods section below).

Ultimately, 1 round of Racon polishing followed by Medaka and NextPolish was selected as the final polishing strategy based on its superior BUSCO recovery of all three odb10 data sets when run in protein mode (Table S3). A flowchart of the genome assembly workflow is provided in Fig. S2.

Genome and gene annotation

De novo repeat identification was performed using RepeatModeler v2.0.1 (Flynn et al. 2020), and the resulting repeat library was used to inform repeat masking with RepeatMasker v4.0.7 (Smit et al. 2017). Gene model and protein prediction was conducted using BRAKER2 v2.1.5 (Hoff et al. 2019; Brúna et al. 2021). For the initial run, BRAKER2 was supplied the *E. crispata* assembly with (1) repeats soft-masked, (2) a custom protein database comprised of metazoa_odb10 and the predicted proteomes of 23 mollusk genome assemblies on NCBI (Table S4), and (3) the *E. crispata* Illumina RNA-Seq data aligned to the genome using STAR v2.7.10 (Dobin et al. 2013). We noticed that highly conserved, multicopy genes (e.g. tandem duplicates of histone protein H3) were missing from this initial run, and manual investigation revealed that these types of gene families had been incorrectly modeled as repetitive elements by RepeatModeler. Therefore, we performed a second BRAKER2 run using the unmasked assembly. Gene models resulting from BRAKER2 annotation with and without the use of the masked reference assembly were independently filtered to retain the highest-quality gene models. Unique gene models resulting from the unmasked run were identified through coordinate intersection with the masked gene set using BEDTools (Quinlan and Hall 2010) and were added to the final gene set if they satisfied the following criteria: (1) had a significant BLASTp hit (*e*-value $< 1e^{-3}$) to a database of model animal proteomes that included human (GRCh38), zebrafish (GRCz11), mouse (GRCm39), and 2 mollusks *Crassostrea virginica* (*C. virginica*-3.0) and *Pomacea canaliculata* (v1.0), (2) did not intersect with a classified repeat element (LINE, SINE, LTR, etc.), and (3) did not have exons that were $> 50\%$ simple sequence repeats. A flowchart of these curation steps is provided in Fig. S3.

To identify mitochondria- and plastid-derived contigs, predicted proteins from the *P. ocellatus* mitochondrial genome (NCBI accn: AP014544.1) and *Rhipidophoron lewmanomontiae* (order: Bryopsidales) chloroplast genome (NCBI accn: AP014542.1) were queried using tblastn v2.11.0 (Camacho et al. 2009) against the *E. crispata* genome assembly. Contigs with 2 or more significant hits (*e*-value $\leq 1e^{-5}$) to these organellar proteins were annotated using GeSeq (Tillich et al. 2017), and OGDRAW (Greiner et al. 2019) was used to generate a physical map of each contig. The physical maps were manually inspected, and 1 contig was flagged as mitochondrial in origin (Fig. S4). An additional 9 contigs were flagged as plastid or plastid fragments (Fig. S5). All organelle-derived contigs were excluded from the nuclear genome assembly and can be downloaded from FigShare (see Data availability).

During the Alien Index (AI) analysis (see Methods section on HGT below), we flagged 194 contigs as bacterial contamination, which were excluded from the nuclear genome assembly. Gene models were removed if they were located on any contig that was flagged from the AI pipeline. Additionally, we flagged the proximal end of contig_7836 as misassembly likely resulting from plastid contamination. The region consisted of 16,750 kb and contained 4 genes, including 2 genes annotated as Photosystem I P700 chlorophyll a apoprotein (PsaA/PsaB)-like. Typically, PsaA and PsaB are both encoded on the chloroplast genome of chlorophyte algae. To investigate the accuracy of the assembly at this locus, Illumina

gDNA and ONT reads were aligned to the genome using BWA MEM (Li and Durbin 2009) and minimap2 (Li 2016) using default parameters, respectively. Alignments relative to genes and repeat elements were visualized using Integrated Genomics Viewer (Robinson et al. 2011). No Illumina reads joined the locus in question to the rest of contig_7836, only a single ONT read mapped to the “gap” between the contig ends, but it mapped poorly (Fig. S6). Altogether, this indicated that this region was likely incorporated into the nuclear genome assembly in error rather than a true HGT event. Therefore, this region of contig_7836 was manually removed from the final genome assembly and the 4 gene models removed from the final gene set. The resulting *E. crispata* assembly and gene annotations following these steps were designated as final (v1).

Search for HGT

We searched the *E. crispata* genome for possible HGT using the AI score as previously described (Wisecaver et al. 2016). Briefly, each predicted protein sequence was queried against the same custom protein database used for BlobTools (see above) with DIAMOND v2.0.8.146 (Buchfink et al. 2015). A custom python script sorted the DIAMOND results based on the normalized bitscore (*nbs*), where *nbs* was calculated as the bitscore of the single best-scoring High-scoring Segment Pair to the subject sequence divided by the best bitscore possible for the query sequence (i.e. the bitscore of the query aligned to itself). The AI score is given by the formula: $AI = nbsO - nbsM$, where *nbsO* is the normalized bit score of the best hit to a species outside of the Metazoa lineage (NCBI: txid33208), *nbsM* is the normalized bit score of the best hit to a species within Metazoa skipping all hits to Placobranchioidea (NCBI: txid71491), a sublineage within Sacoglossa that contains all currently available genomes from kleptoplastic slugs including *E. crispata*. AI scores range from -1 to 1 , being greater than 0 if the predicted protein sequence had a better hit to a nonmetazoan sequence, suggestive of either HGT or contamination (Wisecaver et al. 2016). Contigs were flagged as contamination if the minimum AI score for all genes on a contig was > 0 or if at least half of genes on a contig had an AI score > 0.1 . All contamination-flagged contigs were bacterial in origin and were excluded from the nuclear genome assembly (see Methods section on genome and gene annotation above).

We filtered HGT candidates ($AI > 0$) to those that were most likely to be phylogenetically informative by requiring $AI > 0.1$ and total database hits ≥ 50 . Phylogenetic trees of protein sequences were constructed for all filtered AI-flagged HGT candidates. Full-length proteins corresponding to the top 200 hits (e -value $< 1e^{-10}$) to each HGT candidate were extracted from the local database using esl-sfetch (Eddy 2009). Protein sequences were aligned with MAFFT v7.471 using the E-INS-i strategy and the BLOSUM30 amino acid scoring matrix (Katoh and Standley 2013) and trimmed with trimAL v1.4.rev15 using its gappyout strategy (Capella-Gutierrez et al. 2009). Phylogenies were inferred using maximum likelihood as implemented in IQ-TREE v1.6.12 (Nguyen et al. 2015) using an empirically determined substitution model and 1000 rapid bootstrap replications. The phylogenies were midpoint-rooted, and branches with local support < 95 were collapsed using the ape and phangorn R packages (Paradis et al. 2004; Schliep 2011). Phylogenies were visualized using ETE v3 (Huerta-Cepas et al. 2016) and inspected manually to identify phylogenetically supported HGT candidate proteins.

Species phylogenies

Sacoglossan 28S ribosomal DNA and histone H3 sequences were downloaded from the NCBI nucleotide database (Fig. S7), and

corresponding 28S and H3 sequences in *E. crispata* were identified using BLAST. For the 28S phylogeny, sequences were aligned with MAFFT v7.471 using the G-INS-i iterative refinement method (Katoh and Standley 2013). H3 sequences were aligned with MAFFT using the RevTrans online web portal (Wernersson and Pedersen 2003). Maximum likelihood phylogenies were constructed using IQ-TREE v2.2.0 (Minh et al. 2020) using the built-in ModelFinder to determine the best-fit nucleic acid substitution model (Kalyaanamoorthy et al. 2017) and 1000 ultrafast bootstrap replicates. The concatenated ML tree was constructed based on the combined 28S and H3 data matrix (Chernomor et al. 2016) consisting of 2 partitions and 2006 sites. Phylogenies were visualized using ETE v3 (Huerta-Cepas et al. 2016).

Gene expression quantification and differential expression analysis

Quantification of gene expression was performed using Kallisto v0.46.2 (Bray et al. 2016). The Kallisto index was built using all BRAKER2 predicted transcripts with the default k-mer size of 31. Transcripts per million (TPM) gene abundance values from Kallisto were scaled using the average transcript length averaged over samples and to library size, using the lengthScaledTPM option in tximport v1.18.0 (Soneson et al. 2016). Quality control of the data was performed via principal component analysis, which showed clustering of all biological replicates and strong separation of crawling larva and juvenile slug sample types. Veliger and egg sample types also showed distinct clusters but with slight overlap of 1 egg replicate close to 1.5 standard deviations from the veliger cluster in PC1 and PC2 space (Fig. S8).

A differential expression analysis was performed to compare gene expression in the juvenile slug samples (postfeeding) to the 3 prefeeding developmental stages (egg, veliger, and crawling larvae). Raw gene counts from Kallisto were passed to EdgeR v3.32.1 (Robinson et al. 2009), and only genes with an average TPM > 1 in at least 1 of the 4 developmental stages were retained. Raw gene counts were normalized using the TMM (trimmed mean of M values) method (Robinson and Oshlack 2010). Exact tests were conducted using a trended dispersion value and a double tail reject region. The false discovery rate (FDR) was calculated using the Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg 1995).

Identification and analysis of orthologous gene families

Homology between the predicted proteomes of *E. crispata* and 4 other gastropods (Table S5) was determined using OrthoFinder v2.5.4 (Emms and Kelly 2019) with sequence similarity searches performed by DIAMOND (Buchfink et al. 2015). When multiple protein isoforms were present, the longest protein sequence per gene was selected for the OrthoFinder analysis. Functional annotations were performed on the predicted proteome of *E. crispata* using the eggNOG-mapper v2 online service using default settings (Cantalapiedra et al. 2021). Hypergeometric tests for enrichment of functional categories were performed in python using the SciPy library hypergeom (Virtanen et al. 2020), and P-values were adjusted for multiple comparisons using the StatsModels library multitest with the BH method (Benjamini and Hochberg 1995; Seabold and Perktold 2010). Gene ontology (GO) terms were collapsed to medium length lists of representative GO terms using the simRel somatic similarity score in Revigo (Supek et al. 2011).

Table 1. Summary statistics of *E. crispata* gene models.

Gene model statistics	Total gene set	High-confidence gene set
No. protein-coding genes	68,514	27,767
No. transcripts	71,367	30,043
Mean gene length	5,779.57 bp	10,610.16 bp
Median gene length	2,483 bp	6,891 bp
Mean no. exons per transcript	4.82	8.11
Median no. exons per transcript	3	5
Mean exon length	207.29 bp	201.20 bp
Median exon length	132 bp	128 bp
No. genes supported by RNA-Seq ^a	45,593 (66.55%)	26,778 (96.44%)
No. genes with functional annotation ^b	17,960 (26.21%)	17,529 (63.13%)
No. genes assigned to an orthogroup ^c	28,489 (41.58%)	26,488 (95.39%)
No. genes with PFAM annotation	16,601 (24.23%)	16,220 (58.41%)
No. genes with GO annotation	10,287 (15.01%)	10,191 (36.70%)
No. genes with KEGG annotation	10,624 (15.51%)	10,498 (37.81%)

^a Average length-scaled TPM ≥ 1 in at least 1 developmental stage: egg, veliger, larval slug, and juvenile slug.

^b Assigned 1 or more annotation via EggNOG-mapper.

^c Excluding orthogroups unique to *E. crispata*.

Ketoacyl synthase domain phylogeny

The *E. crispata* polyketide synthase gene EcPKS1 (Torres et al. 2020) was queried using blastp against the same custom database used for the AI analysis, and full-length sequences corresponding to the top 2000 significant hits (e -value $< 1e^{-10}$) were extracted. All significant hits were queried using hmmscan (Eddy 2009) against 3 ketoacyl synthase domain profiles present in fatty acid and polyketide synthases (including EcPKS1): PF00109, beta-ketoacyl synthase, N-terminal domain; PF02801, beta-ketoacyl synthase, C-terminal domain; and PF16197, ketoacyl synthetase C-terminal extension. For each homolog, the maximum sequence region that spanned all 3 domain profiles was extracted and retained for phylogenetic analysis; extracted domain regions < 350 or > 650 amino acids were excluded. To reduce redundancy in the final phylogeny, database sequences were collapsed with CD-HIT v4.8.1 using a sequence identity threshold of 0.8 (Fu et al. 2012); sequences from kleptoplastic slugs were not collapsed. The final sequence set was aligned with MAFFT v7.471 using the L-INS-i strategy (Katoh and Standley 2013). Phylogenies were inferred using maximum likelihood as implemented in IQ-TREE v1.6.12 (Nguyen et al. 2015) using an empirically determined substitution model and 1000 rapid bootstrap replications. The phylogeny was visualized using ETE v3 (Huerta-Cepas et al. 2016).

Results

E. crispata genome assembly and annotation

Sequencing data for the *E. crispata* genome assembly consisted of 18.98 Gbp of ONT long reads (2,504,236 reads with an N50 of 8,728 bp) and 75.25 Gbp of paired-end Illumina short reads. A k-mer frequency analysis of the Illumina data indicated an estimated heterozygosity of 0.88% and a projected genome size of 711.11 Mb (Fig. S9). The final genome assembly consisted of 8,089 contigs spanning 786.3 Mb, with an N50 of 458.73 kb. To evaluate completeness and coverage, we aligned the ONT and Illumina gDNA reads to the *E. crispata* genome assembly. Coverage histograms showed a single dominant peak at 20 \times and 86 \times coverage for ONT and Illumina reads, respectively (Fig. S9), indicating that the genome assembly was largely homozygous without a significant amount of redundant haplotigs. The proportion of the genome with ≥ 10 \times read support was high: 751.15 Mb (96.29%) using ONT reads and 757.17 Mb (96.29%) using Illumina reads.

Repetitive elements made up 29.85% of assembly bases (Table S6). Most repetitive elements were unclassified, which comprised 17.7% of the genome assembly. Simple repeats made up 7.19% of the genome. Classified retroelements and DNA transposons comprised 2.85% and 1.47% of the genome, respectively. The repeat content in *E. crispata* is comparable to the repeat content observed in the other sequenced sacoglossan genomes (29–33% repetitive) (Cai et al. 2019; Maeda et al. 2021).

A total of 68,514 genes encoding 71,367 transcripts were predicted using a combination of ab initio, homology-based, and transcriptome-based prediction methods (Table 1). The average protein-coding gene was 5,779.57-bp long and contained 4.82 exons. Functional annotations were ascribed to 24.23%, 15.51%, and 15.01% of genes using the PFAM, KEGG, and GO databases, respectively (Table 1). Within the *E. crispata* protein-coding gene set, 96.65% (922/954) of BUSCO conserved metazoan genes were identified as complete, and of those, 94.03% were present in single copy, and 5.97% were duplicated (Table S3). Furthermore, 99.22% (253/255) of BUSCO conserved eukaryota genes were identified as complete; of those, 92.50% were present in single copy, and 7.51% were duplicated (Table S3). This level of BUSCO recovery is the greatest of any currently available genome assembly from kleptoplastic sacoglossans (Fig. 2a; Table S7).

The *E. crispata* genome assembly is similar to *E. marginata* in genome assembly size and number of protein-coding genes (Fig. 2a). Gene count is noticeably elevated in *E. crispata*, *E. marginata*, and *P. ocellatus* ($n > 68,000$) compared to a fourth kleptoplastic *Sacoglossa*, *E. chlorotica* ($n = 24,980$), as well as the nonkleptoplastic sea hare, *Aplysia californica* ($n = 19,945$). We suspected the difference in gene count was due to an abundance of de novo-predicted gene calls in the former species set. To address this, we determined the number of *E. crispata*-predicted protein-coding genes with transcriptome- or homology-based support. Most of the de novo predicted genes in *E. crispata* (66.55%) showed evidence of expression, having an average length-scaled TPM > 1 in at least 1 of the 4 developmental stages surveyed (Fig. 2b; Table 1). Fewer genes had homology-based support; 41.58% were assigned to gene families containing 1 or more species in addition to *E. crispata*, and 26.21% of genes were assigned 1 or more functional annotations via EggNOG-mapper (Fig. 2b; Table 1). We created a high-confidence gene set of 27,767 genes, which we defined as those genes with 2 or more sources of support (i.e. RNA-seq, orthology, and EggNOG-mapper). Median gene length and exons

per transcript were increased in the high-confidence gene set compared to the total set (Table 1). Both gene sets can be downloaded from FigShare (see Data availability). When accessing the data, we encourage users to select the gene set most appropriate for their biological question.

Assessment of HGT

We calculated the AI for all *E. crispata* predicted proteins to identify possible contamination and cases of HGT. The AI screen flagged 194 contigs as bacterial contamination, which were subsequently excluded from the final assembly (see Materials and methods). Only 7 proteins were flagged as candidate HGTs with top hits to diverse organisms in Apicomplexa, Ochrophyta, Proteobacteria, and Riboviria (Table S8). None of the candidate HGTs had predicted functions associated with photosynthesis or plastids (Table S9). We used a custom phylogenetic pipeline to manually evaluate all candidate HGTs and found that all but 1 lacked phylogenetic support for HGT. The only HGT with

phylogenetic support was Ecla3748g428790 (Fig. 3), a single-exon gene with homology to a fragment of RNA-dependent RNA polymerase (RdRp) from negative-sense RNA viruses of invertebrates. The Ecla3748g428790 transcript had low but measurable expression support in our RNA-seq data set with length-scaled TPM > 1 in 3 of 4 developmental stages surveyed (Table S9). RdRp genes appear to be frequently transferred into the nuclear genomes of their eukaryotic hosts (Dolja and Koonin 2018), supporting the inference that this sequence was horizontally acquired from a virus in *E. crispata*.

Phylogenetic placement

Slugs were identified as the clarki morphotype of *E. crispata* based on macroscopic morphological features, including a near-uniform green coloration with small round white spots, a transparent and tapered foot, and the presence of a gap in the parapodia in the midline near the head (Fig. 1a, b, and g) (Pierce et al. 2006; Krug et al. 2016). To confirm the placement of our assembled genome

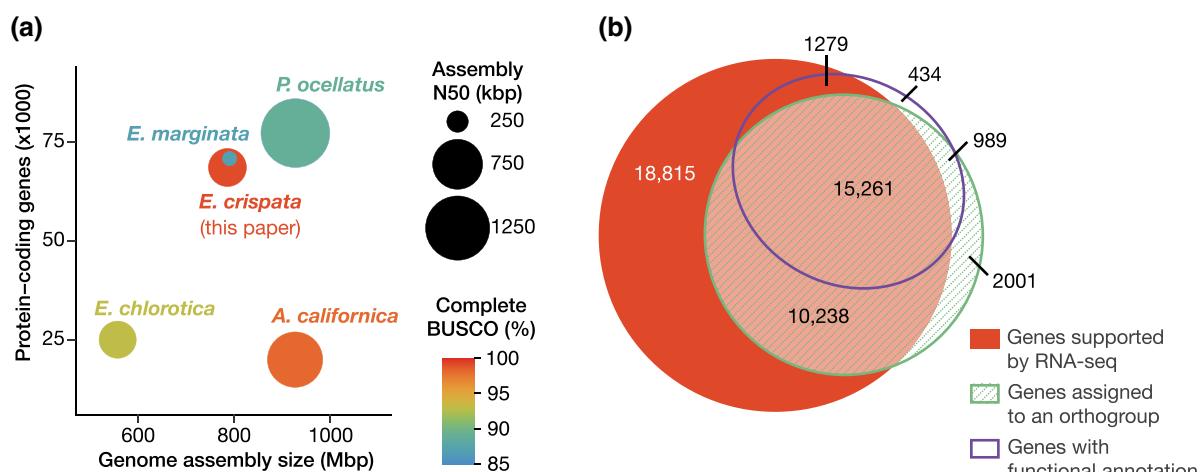


Fig. 2. Genome assembly and annotation statistics. a) Comparison of assembly completeness and contiguity of 4 kleptoplastic sea slugs (*E. crispata*, *E. chlorotica*, *P. ocellatus*, and *E. marginata*) and the sea hare *A. californica*. b) Area-proportional Venn diagram (Micallef and Rodgers 2014) depicting overlapping support for *E. crispata* gene models. Support categories are as described in Table 1. Genes with 2 or more sources of support were assigned to the high-confidence gene set.

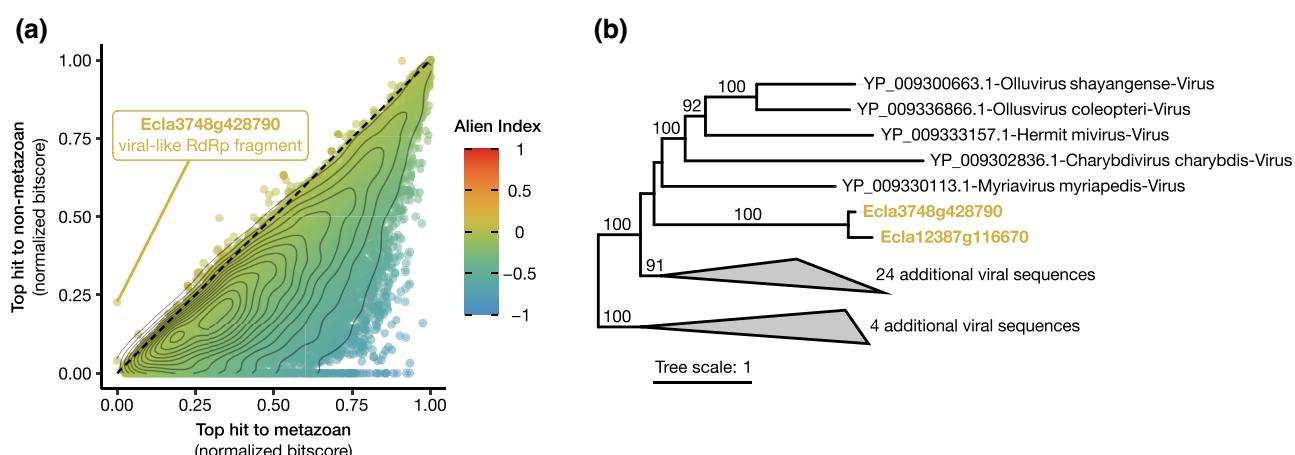


Fig. 3. Detection of HGT in *E. crispata*. a) AI analysis. Normalized bitscores for the top hit to the metazoan lineage (x-axis) are compared to the strength of the top hit outside of metazoans (y-axis). Genes above the dashed line have an AI > 0 and are candidates for HGT. Contour lines indicate a 2D kernel density estimation of the distribution of AI scores. Highlighted HGT candidate gene, Ecla3748g428790, has top hit to RdRp from *Huberia myriapoda* virus. No other HGT candidates were supported following phylogenetic analysis. b) RdRp phylogeny showing phylogenetic support for Ecla3748g428790 being horizontally acquired from viruses. Numbers along branches indicate ultrafast bootstrap support values (≥ 95) for descendant nodes.

in the Sacoglossa species tree, we constructed concatenated maximum likelihood phylogeny consisting of 2 nuclear loci coding for the 28S large ribosomal subunit and histone protein H3. Both 28S and H3 have served as DNA barcode genes in a previous phylogenetic analysis of the lineage (Händeler et al. 2009; Christa, Gould, et al. 2014, 2015; Krug et al. 2016; Martín-Hervás et al. 2021). The 28S sequence from the *E. crispata* genome was 99.58% identical (1406/1412 bp) to 28S amplified from another *E. crispata* collected from the Florida Keys (GenBank accn: KM230502, Krug et al. 2016), and the H3 genome sequence was 100% identical (327/327 bp) H3 amplified from the same sample (GenBank accn: KM040828, Krug et al. 2016), supporting its identification as *E. crispata* based on morphology. Genome assembly sequences grouped with *E. crispata* and sister species *Elysia ellena* (Ortea et al. 2013; Krug et al. 2016) in both the concatenated and single-gene phylogenies with strong support (ultrafast bootstrap ≥ 98 ; Fig. 4; Fig. S7). Other *Elysia* species that grouped closely to *E. crispata* included *Elysia canguzua*, *E. chlorotica*, *Elysia diomedea*, *Elysia evelinae*, *Elysia serca*, and *Elysia viridis* (Fig. 4), consistent with these species being members of a West Atlantic subclade complex (Krug et al. 2016).

Gene expression across developmental life stages

We tracked *E. crispata* gene expression across 4 developmental life stages. Three life stages (egg, veliger, and crawling larva) were grown in the absence of algae and contained no kleptoplasts. The fourth life stage, juvenile slug, was allowed to feed on macroalgae and acquired green pigmentation indicative of active kleptoplast sequestration (Fig. 1g). Most predicted genes were not expressed under any of the four life stages ($n = 22,925$; average length-scaled TPM < 1); most genes in this category (95.7%; $n = 21,932$) were not assigned to the high-confidence set (Fig. 5a; Table S9). In contrast, gene models that were expressed across all 4 developmental stages ($n = 19,772$; average length-scaled TPM > 1) were mostly comprised of high-confidence genes (Fig. 5a). A total of 5,863 gene models were uniquely expressed in the kleptoplast-containing juvenile slugs with no measurable expression in the unfed life stages; of those, 2,044 gene models (34.9%) were considered high-confidence (Fig. 5a).

RNA-seq analysis identified 3318 differentially expressed genes (DEGs) that were significantly more abundant in kleptoplast-containing juvenile slugs compared to the unfed developmental life stages ($-\log_2 FC > 2$; BH adjusted P-value < 0.01; Fig. 5b). A test for significant enrichment of GO functional categories in DEGs identified 561 enriched GO terms (hypergeometric test, BH-adjusted P-value < 0.05; Table S10). Among the most abundant were GO terms such as GO: 0051239, regulation of multicellular organismal process; GO: 0032879, regulation of localization; GO: 0009605, response to external stimulus; and GO: 0002376, immune system process (Table S10). We hypothesize that DEGs in this analysis include candidate genes involved in long-term kleptoplast retention.

Candidates for kleptoplasty-associated genes

To distinguish conserved gastropod developmental genes from genes potentially involved in kleptoplasty, we performed an OrthoFinder analysis (Emms and Kelly 2019) using the predicted proteome of *E. crispata* and 4 additional gastropod genomes (Fig. 6). Included in our analysis were genomes from 2 sacoglossan slugs capable of long-term kleptoplast retention, *E. chlorotica* (Cai et al. 2019) and *P. ocellatus* (Maeda et al. 2021). A third Sacoglossa was capable of short-term kleptoplast retention, *E. marginata* (Maeda et al. 2021). Lastly, 1 nonkleptoplastic gastropod relative (the sea hare, *A. californica*) was also included (see Table S6 for

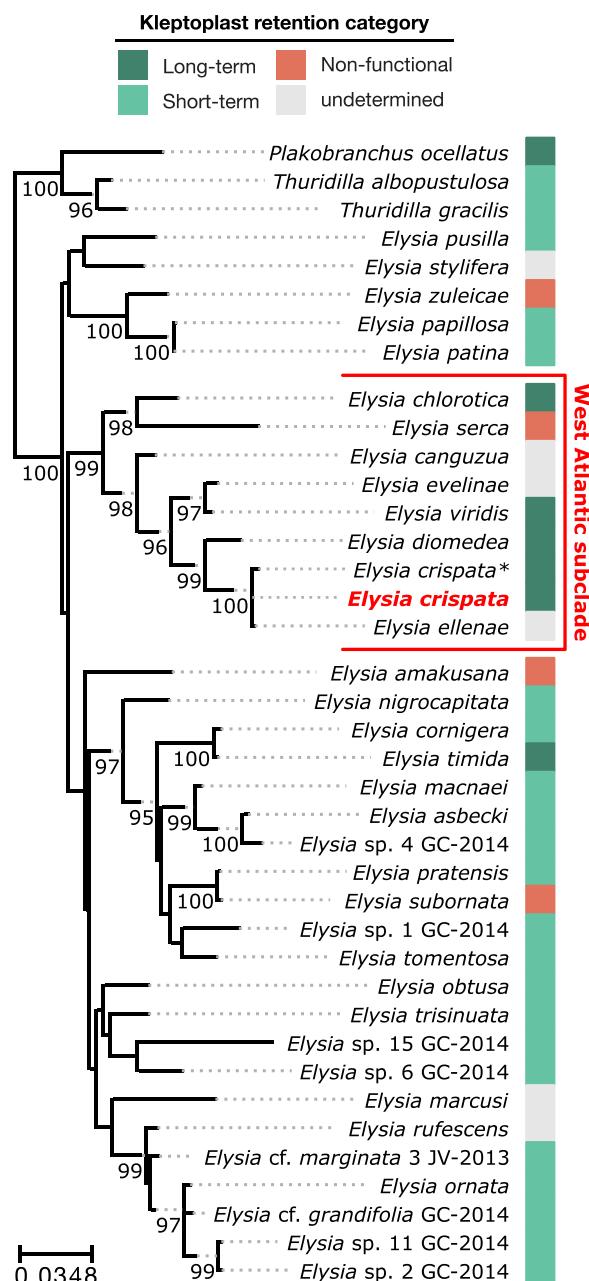


Fig. 4. Phylogenetic analysis of sacoglossan sea slugs. Maximum likelihood analysis of concatenated 28S and H3 loci. *E. crispata* genome sequence is in red and bolded. Asterisk (*) indicates additional *E. crispata* sequence from Krug et al. (2016). Tree was rooted on the *Thridilla*/*Plakobranchus* clade. Numbers along select branches indicate IQ-TREE ultrafast bootstrap support values (≥ 95) for the descendant nodes. The color bar indicates the kleptoplast retention state for each species: long-term retention (>20 days) and short-term retention (>1 and <20 days); see Table S12 for references. See Fig. S7 for single-gene phylogenies of 28S and H3.

genome accessions). The OrthoFinder analysis identified 119,336 orthogroups (predicted gene families), of which 21,821 were present in 2 or more species in the analysis. Excluding singletons, most orthogroups were present in all 5 gastropod species ($n = 11,249$; Fig. 6a). Within these conserved orthogroups, 1,054 *E. crispata* sequences had significantly higher expression in juvenile slugs compared to the earlier developmental stages (Fig. 6b). Functional enrichment of GO terms indicated that many of these genes in *E. crispata* have putative functions in the regulation of

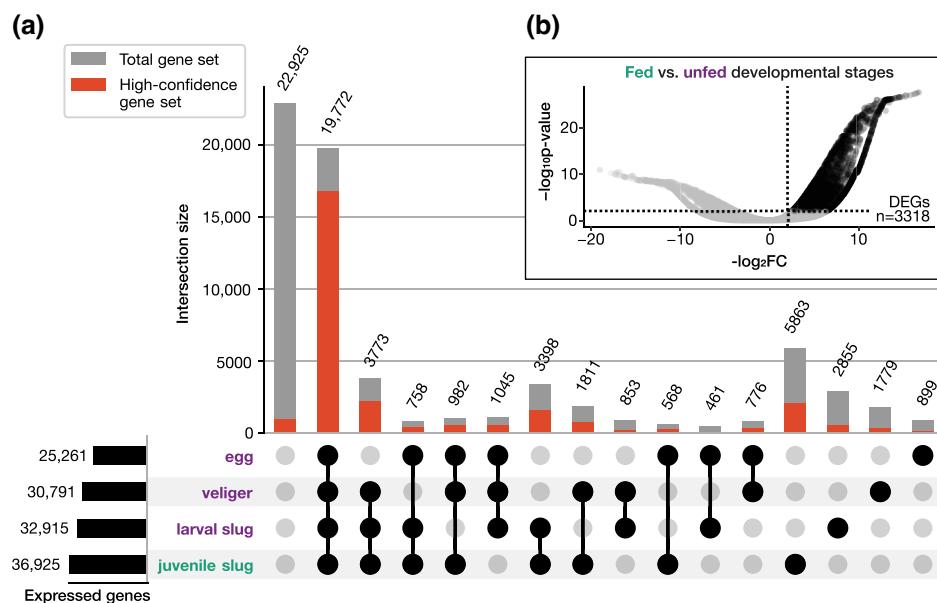


Fig. 5. Gene expression in different slug life stages. a) Upset plot displaying intersection of gene expression across 3 unfed life stages (purple: eggs, veligers, and crawling larvae) and in alga-fed juvenile slugs (green). Filled circles and connecting lines indicate expression support under those conditions (average length-scaled TPM > 1). Numbers above bars indicate the total number of genes present in each intersection. Red bars indicate the proportion of each intersection comprised of high-confidence genes. b) Distribution of *E. crispata* gene expression comparing alga-fed juvenile slugs to unfed developmental stages. Dashed lines indicate significance thresholds for differential expression ($-\log_2\text{FC} > 2$; BH adjusted P-value < 0.01). Number indicates count of DEGs with increased abundance in fed slugs.

multicellular organismal processes (GO: 0051239), developmental processes (GO: 0032502, and GO: 0050793), and anatomical structure morphogenesis (GO: 0009653) (Fig. 6c; Table S10).

A total of 2,165 orthogroups were present in all kleptoplastic Sacoglossa and absent in the nonkleptoplastic sea hare (Fig. 6a). Within these sacoglossan-specific orthogroups, 554 *E. crispata* sequences had significantly higher expression in juvenile slugs compared to the earlier developmental stages (Fig. 6b). These genes showed significant enrichment of GO categories involved in the regulation of ROS (GO: 2000377 and GO: 1903427), regulation of nitric oxide (GO: 0045019), and adaptive immune response (GO: 0002819) (Fig. 6d; Table S10).

Previous genomic and transcriptomic analyses of *P. ocellatus* identified 3 types of genes that were highly duplicated in the *P. ocellatus* genome and overexpressed in kleptoplast-containing tissues, suggesting they may be involved in kleptoplast retention (Maeda et al. 2021). We searched the *E. crispata* genome for sequences belonging to these 3 gene families: apolipoprotein D-like (KEGG: K03098), cathepsin D-like (KEGG: K01379), and lectin-like genes (KEGG: ko04091). The *E. crispata* genome contained 11 apolipoprotein D-like genes and 6 cathepsin D-like genes (Table S9), but neither gene family was enriched in genes with higher expression in kleptoplast-containing slugs compared to earlier developmental stages. In contrast, the *E. crispata* genome contained 307 lectin-like genes, 40.4% of which had higher expression in kleptoplast-containing slugs (Table S9). Moreover, lectins was the only KEGG functional category significantly enriched in the sacoglossan-specific gene set (hypergeometric test, BH-adjusted P-value = 1.3e⁻⁶; Table S11).

A polyketide synthase in *E. chlorotica* (EcPKS1) was recently shown in vitro to synthesize the precursor for 7- and 8-propionate pyrones (Torres et al. 2020), which are proposed to function as UV- and oxidation-blocking compounds in slug cell membranes (Powell et al. 2018). Type 1 iterative polyketide synthase genes in kleptoplastic slugs, including EcPKS1, appear to have evolved

from fatty acid synthase (FAS) genes present in all animals that are required for primary metabolism (Torres et al. 2020). We identified several PKS/FAS homologs encoded in the *E. crispata* genome (Fig. 7). A phylogeny of the ketoacyl synthase domain from these sequences showed that 1 *E. crispata* gene (Ecla12149g328170) grouped with FAS genes from *E. chlorotica* and other gastropods. Four genes in *E. crispata* were members of an expanded clade of sacoglossan-specific PKS-like genes (Fig. 7b). Ecla12132g95200 (EclaPKS1) grouped with EcPKS1 in a subclade specific to species capable of long-term kleptoplast retention. Ecla7432g605620 (EclaPKS2) and Ecla7432g605630 (EclaPKS2-like) grouped with EcPKS2. In a third PKS clade, Ecla2359g352500 (EclaPKS3) grouped with an uncharacterized gene in *E. chlorotica* (GenBank accn: RUS77019), which we called EcPKS3 following the naming convention used previously by others (Torres et al. 2020). EclaPKS1 and EclaPKS3 were mostly highly expressed in kleptoplast-containing juvenile slugs (Fig. 7c); however, neither was statistically significantly overexpressed in juvenile slugs compared to the unfed developmental stages (Table S9).

Discussion

The *E. crispata* genome assembly serves as a high-quality reference for future functional and evolutionary studies in kleptoplastic sea slugs. With calculated BUSCO scores of 93–99% (Table S7), the *E. crispata* gene space is the most complete of all sequenced sacoglossans (Fig. 2) and contends with the most complete of all mollusk genomes (Caurcel et al. 2021). As in previous genome level analyses of kleptoplastic sea slugs (Wägele et al. 2011; Bhattacharya et al. 2013; Maeda et al. 2021), we found no evidence of HGT from host algae. Nevertheless, our search strategy detected 1 case of probable HGT from viruses (Fig. 3). Alternative mechanisms for kleptoplast maintenance must be considered and investigated to better understand how these sea slugs can retain

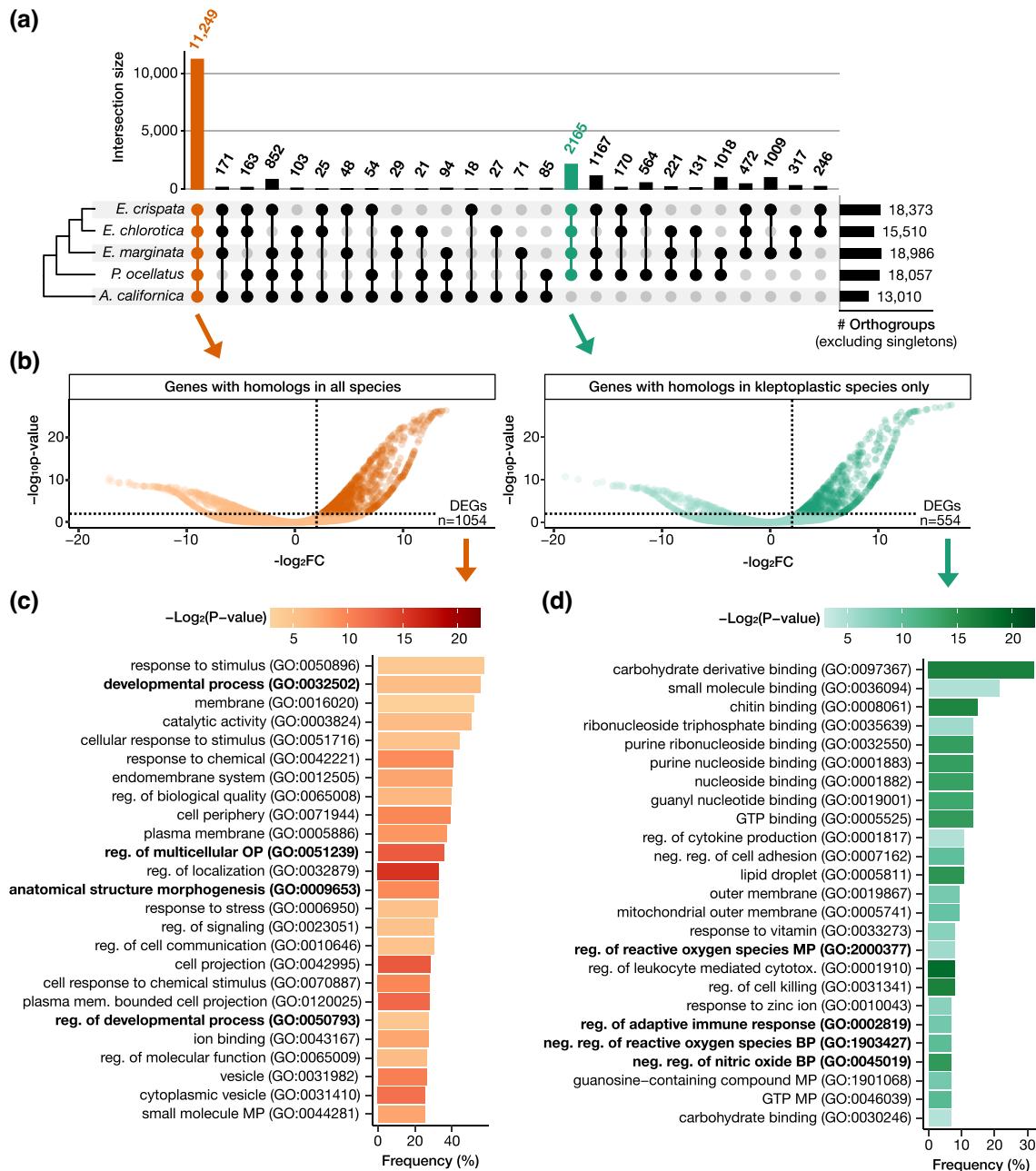


Fig. 6. Gene family analysis of DEGs. a) Upset plot depicting intersection of orthogroups among four kleptoplastic sea slugs and the sea hare *A. californica*. Orange intersection indicates orthogroups present in all species; green intersection indicates orthogroups present in all kleptoplastic species and absent in the outgroup. b) For the orange and green intersections, DEGs with increased abundance in fed slugs compared to unfed life stages (see Fig. 5b). Tests for enrichment of functional categories were performed on these 2 DEG categories: c) top 25 most abundant GO categories significantly enriched in DEGs with homologs in all other species and d) top 25 most abundant GO categories significantly enriched in DEGs with homologs in all other kleptoplastic species and absent in the outgroup (see Table S10). GO categories highlighted in the text are bolded.

functional kleptoplasts for so long in the absence of algal nuclear-encoded chloroplast genes.

The ability to retain long-term kleptoplasts has evolved multiple times independently in Sacoglossa, but the exact number and timing of transitions from short-term to long-term kleptoplasty remains unclear (Christa et al. 2015). *E. crispata* belongs to the *Elysia* West Atlantic subclade (Krug et al. 2016), a lineage that is unique in that it contains several species that retain long-term kleptoplasts including *E. crispata*, *E. chlorotica*, *Elysia diamedea*, and *E. viridis* (Fig. 4; Table S12). The prevalence of long-term kleptoplasts in this subclade suggests that long-term retention may be

an ancestral characteristic of this lineage. However, another species within the subclade, *E. serca*, does not retain functional kleptoplasts (Clark et al. 1990), and several members of this subclade (e.g. *E. canguzua*, *E. ellenae*, and *E. evelinae*) are yet to be studied for kleptoplast retention. Characterizing the retention time for these species and determining the timing of transition(s) to long-term kleptoplasty in the West Atlantic subclade would inform our understanding of how kleptoplasty evolved in sacoglossan slugs. A single ancestral origin for long-term retention in this subclade would suggest that 1 or more rare innovations facilitated this key transition. In contrast, if long-term retention arose repeatedly

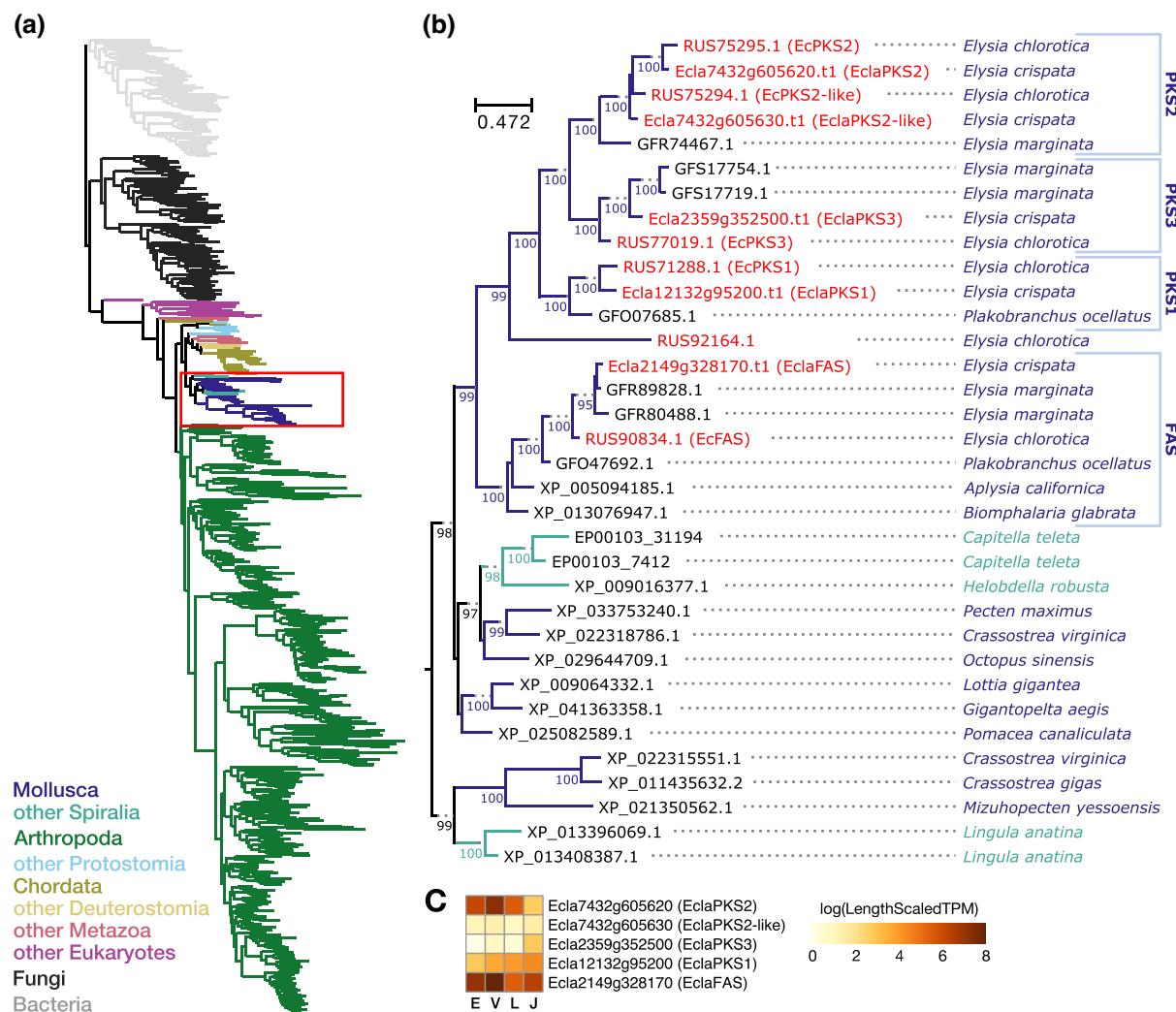


Fig. 7. Phylogenetic analysis of the ketoacyl synthase domain from FAS and PKS genes in kleptoplastic slugs. a) Maximum likelihood phylogeny of the ketoacyl synthase domains with significant sequence similarity to the *E. chlorotica* type 1 PKS (EcPKS1; Torres et al. 2020). Tree is rooted on bacterial PKS sequences, and branches are color-coded according to their taxonomic identity. Molluscan sequences are indicated by the red box. b) Detailed view of molluscan sequences. *E. crispata* and *E. chlorotica* sequences are in red. Numbers along branches indicate ultrafast bootstrap support values (≥ 95) for the descendant nodes. c) Average gene expression for *E. crispata* FAS/PKS homologs in 4 developmental stages (E, egg; V, veliger; L, larval slug; J, juvenile slug).

in the subclade (i.e. parallel or convergent evolution), the transition to long-term retention may be the result of relatively simple genetic mechanisms. Addressing this question requires additional taxonomic sampling and phenotyping in this group.

Combined, comparative genomic and gene expression analyses identified 554 genes in *E. crispata* that were both unique to kleptoplast-containing slugs and more highly expressed in juvenile slugs compared to unfed early developmental stages. Genes within this intersection included lectin carbohydrate-binding proteins (Table S11) as well as those involved in regulation of ROS, production of nitric oxide, and adaptive immunity (Fig. 6d). Similar gene families were identified in a genomic analysis of *P. ocellatus*, another long-term kleptoplastic Sacoglossa (Maeda et al. 2021), providing additional evidence that genes in these categories may play a role in prolonged kleptoplast retention in these slugs. Recently, Torres et al. (2020) identified 3 type 1 polyketide synthase genes in *E. chlorotica* that may play a role in protecting the kleptoplast from UV and oxidative damage. Our analysis of this gene family identified 4 PKS genes in *E. crispata* (Fig. 7), including 1 belonging to a new PKS3 subclade, suggesting that polyketide-based products in kleptoplastic slugs may be more

diverse than previously recognized. Future work is necessary to develop *E. crispata* into a genome-enabled model system for functional investigation of these genes and others to better understand the genetic and cellular mechanisms of kleptoplasty.

Data availability

All supplemental material as well as genome assembly, predicted CDS and protein sequences, multiple sequence alignments, tree files, and other related data files are available through FigShare (<https://doi.org/10.6084/m9.figshare.23635812.v1>). Scripts are available through GitHub (https://github.com/WisecaverLab/elysia_crispata_ECLA1_genome). The *E. crispata* Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAWDGP000000000. The version described in this paper is version JAWDGP010000000. *E. crispata* 28S ribosomal RNA gene sequence has been deposited at the NCBI GenBank database under accession OR177832. Raw sequencing reads used for de novo whole-genome assembly have been deposited in the Sequence Read Archive database under BioProject PRJNA987316.

Supplemental material available at G3 online.

Acknowledgments

We thank members of the Wisecaver and Widhalm labs for helpful discussions. We also thank Dr. Dave Sandstrom (University of Maryland) for guidance on sea slug rearing and culturing algae. This work was conducted in part using the resources of the Rosen Center for Advanced Computing at Purdue University.

Funding

This work was supported by Purdue University through research seed grants provided by the Center for Plant Biology in the College of Agriculture and the Agricultural Science and Extension for Economic Development to J.R.W. and J.H.W. This work was also supported by the National Science Foundation under grant DEB-1831493 to JHW, and by Research Corporation for Science Advancement award number 26212, by Showalter Research Trust award 41000747, and by Gordon and Betty Moore Foundation award 9331 to J.R.W. This work was also supported by the USDA National Institute of Food and Agriculture Hatch Project numbers 177845 to J.R.W and 1016057 to J.H.W.

Conflict of interest

The authors declare no conflict of interest.

References

Babraham Bioinformatics. 2011. FastQC A quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc. Series B (Methodol)*. 57(1):289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.

Bhattacharya D, Pelletreau KN, Price DC, Sarver KE, Rumpho ME. 2013. Genome analysis of *Elysia chlorotica* egg DNA provides no evidence for horizontal gene transfer into the germ line of this kleptoplastic mollusc. *Mol Biol Evol*. 30(8):1843–1852. doi:10.1093/molbev/mst084.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 34(5):525–527. doi:10.1038/nbt.3519.

Brüna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform*. 3(1):lqaa108. doi:10.1093/nargab/lqaa108.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 12(1):59–60. doi:10.1038/nmeth.3176.

Bushnell B. 2017. BBTools Software Package. <https://sourceforge.net/projects/bbmap/>.

Cai H, Li Q, Fang X, Li J, Curtis NE, Altenburger A, Shibata T, Feng M, Maeda T, Schwartz JA, et al. 2019. A draft genome assembly of the solar-powered sea slug *Elysia chlorotica*. *Sci Data*. 6(1):190022. doi:10.1038/sdata.2019.22.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL, et al. 2009. BLAST+: architecture and applications. *BMC Bioinform*. 10(1):421. doi:10.1186/1471-2105-10-421.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol*. 38(12):5825–5829. doi:10.1093/molbev/msab293.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 25(15):1972–1973. doi:10.1093/bioinformatics/btp348.

Caurcel C, Laetsch DR, Challis R, Kumar S, Gharbi K, Blaxter M, et al. 2021. MolluscDB: a genome and transcriptome database for molluscs. *Phil. Trans. R. Soc. B*. 376(1825):20200157. doi:10.1098/rstb.2020.0157.

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 34(17):i884–i890. doi:10.1093/bioinformatics/bty560.

Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol*. 65(6):997–1008. doi:10.1093/sysbio/syw037.

Christa G, Gould SB, Franken J, Vleugels M, Karmeinski D, Händeler K, Martin WF, Wägele H, et al. 2014. Functional kleptoplasty in a limapontioidean genus: phylogeny, food preferences and photosynthesis in *Costasiella*, with a focus on *C. ocellifera* (Gastropoda: Sacoglossa). *J Molluscan Stud.* 80(5):499–507. doi:10.1093/mollus/eyu026.

Christa G, Händeler K, Kück P, Vleugels M, Franken J, Karmeinski D, Wägele H. 2015. Phylogenetic evidence for multiple independent origins of functional kleptoplasty in Sacoglossa (Heterobranchia, Gastropoda). *Org Divers Evol*. 15(1):23–36. doi:10.1007/s13127-014-0189-z.

Christa G, Händeler K, Schäferle TF, König GM, Wägele H. 2014. Identification of sequestered chloroplasts in photosynthetic and non-photosynthetic sacoglossan sea slugs (Mollusca, Gastropoda). *Front Zool*. 11(1):15. doi:10.1186/1742-9994-11-15.

Clark K, Jensen K, Stirts HM. 1990. Survey for functional kleptoplasty among West Atlantic Ascoglossa (=Sacoglossa) (Mollusca, Opisthobranchia). *Veliger*. 33:339–345.

Cruz S, LeKieffre C, Cartaxana P, Hubas C, Thiney N, Jakobsen S, Escrig S, Jesus B, Kühl M, Calado R, et al. 2020. Functional kleptoplasts intermediate incorporation of carbon and nitrogen in cells of the Sacoglossa sea slug *Elysia viridis*. *Sci Rep*. 10(1):10548. doi:10.1038/s41598-020-66909-7.

Curtis NE, Schwartz JA, Pierce SK. 2010. Ultrastructure of sequestered chloroplasts in sacoglossan gastropods with differing abilities for plastid uptake and maintenance. *Invertebr Biol*. 129(4):297–308. doi:10.1111/j.1744-7410.2010.00206.x.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR, et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 29(1):15–21. doi:10.1093/bioinformatics/bts635.

Dolja VV, Koonin EV. 2018. Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Res*. 244:36–52. doi:10.1016/j.virusres.2017.10.020.

Eddy SR. 2009. A new generation of homology search tools based on probabilistic inference. *Genome Inform*. 23:205–211.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 20(1):238. doi:10.1186/s13059-019-1832-y.

Evertsen J, Burghardt I, Johnsen G, Wägele H. 2007. Retention of functional chloroplasts in some sacoglossans from the Indo-Pacific and Mediterranean. *Mar Biol*. 151(6):2159–2166. doi:10.1007/s00227-007-0648-6.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA*. 117(17):9451–9457. doi:10.1073/pnas.1921046117.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 28(23): 3150–3152. doi:[10.1093/bioinformatics/bts565](https://doi.org/10.1093/bioinformatics/bts565).

Greiner S, Lehwerk P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47(W1): W59–W64. doi:[10.1093/nar/gkz238](https://doi.org/10.1093/nar/gkz238).

Händeler K, Grzymbowski YP, Krug PJ, Wägele H. 2009. Functional chloroplasts in metazoan cells - a unique evolutionary strategy in animal life. *Front Zool.* 6(1):28. doi:[10.1186/1742-9994-6-28](https://doi.org/10.1186/1742-9994-6-28).

Hoff K, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. *Methods Mol Biol.* 1962:65–95. doi:[10.1007/978-1-4939-9173-0_5](https://doi.org/10.1007/978-1-4939-9173-0_5).

Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*. 36(7): 2253–2255. doi:[10.1093/bioinformatics/btz891](https://doi.org/10.1093/bioinformatics/btz891).

Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol.* 33(6): 1635–1638. doi:[10.1093/molbev/msw046](https://doi.org/10.1093/molbev/msw046).

Ireland C, Scheuer PJ. 1979. Photosynthetic marine mollusks: in vivo ¹⁴C incorporation into metabolites of the sacoglossan *Placobranchus ocellatus*. *Science*. 205(4409):922–923. doi:[10.1126/science.205.4409.922](https://doi.org/10.1126/science.205.4409.922).

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 14(6):587–589. doi:[10.1038/nmeth.4285](https://doi.org/10.1038/nmeth.4285).

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780. doi:[10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).

Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, et al. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* 12(6):e1001889. doi:[10.1371/journal.pbio.1001889](https://doi.org/10.1371/journal.pbio.1001889).

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 37(5): 540–546. doi:[10.1038/s41587-019-0072-8](https://doi.org/10.1038/s41587-019-0072-8).

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM, et al. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27(5):722–736. doi:[10.1101/gr.215087.116](https://doi.org/10.1101/gr.215087.116).

Krug PJ, Vendetti JE, Valdés Á. 2016. Molecular and morphological systematics of *Elysia* Risso, 1818 (Heterobranchia: Sacoglossa) from the Caribbean region. *Zootaxa*. 4148(1):1–137. doi:[10.11646/zootaxa.4148.1.1](https://doi.org/10.11646/zootaxa.4148.1.1).

Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. *F1000Res.* 6:1287. doi:[10.12688/f1000research.12232.1](https://doi.org/10.12688/f1000research.12232.1).

Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*. 32(14):2103–2110. doi:[10.1093/bioinformatics/btw152](https://doi.org/10.1093/bioinformatics/btw152).

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 25(14):1754–1760. doi:[10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324).

Maeda T, Takahashi S, Yoshida T, Shimamura S, Takaki Y, Nagai Y, Toyoda A, Suzuki Y, Arimoto A, Ishii H, et al. 2021. Chloroplast acquisition without the gene transfer in kleptoplastic sea slugs, *Placobranchus ocellatus*. *eLife*. 10:e60176. doi:[10.7554/eLife.60176](https://doi.org/10.7554/eLife.60176).

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol.* 38(10):4647–4654. doi:[10.1093/molbev/msab199](https://doi.org/10.1093/molbev/msab199).

Martin W, Herrmann RG. 1998. Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol.* 118(1):9–17. doi:[10.1104/pp.118.1.9](https://doi.org/10.1104/pp.118.1.9).

Martín-Hervás MR, Carmona L, Malaquias MAE, Krug PJ, Gosliner TM, Cervera JL. 2021. A molecular phylogeny of *Thuridilla* Berg, 1872 sea slugs (Gastropoda, Sacoglossa) reveals a case of flamboyant and cryptic radiation in the marine realm. *Cladistics*. 37(6):647–676. doi:[10.1111/cla.12465](https://doi.org/10.1111/cla.12465).

Matasci N, Hung L-H, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T, Ayyampalayam S, Barker M, et al. 2014. Data access for the 1,000 Plants (1KP) project. *GigaSci.* 3(1):17. doi:[10.1186/2047-217X-3-17](https://doi.org/10.1186/2047-217X-3-17).

Micallef L, Rodgers P. 2014. eulerAPE: drawing area-proportional 3-venn diagrams using ellipses. *PLoS One*. 9(7):e101717. doi:[10.1371/journal.pone.0101717](https://doi.org/10.1371/journal.pone.0101717).

Middlebrooks ML, Bell SS, Pierce SK. 2012. The kleptoplastic sea slug *Elysia clarki* prolongs photosynthesis by synthesizing chlorophyll a and b. *Symbiosis*. 57(3):127–132. doi:[10.1007/s13199-012-0187-x](https://doi.org/10.1007/s13199-012-0187-x).

Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*. 34(13):i142–i150. doi:[10.1093/bioinformatics/bty266](https://doi.org/10.1093/bioinformatics/bty266).

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 37(5):1530–1534. doi:[10.1093/molbev/msaa015](https://doi.org/10.1093/molbev/msaa015).

Nanoporetech. 2022. Medaka. GitHub repository. <https://github.com/nanoporetech/medaka>.

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274. doi:[10.1093/molbev/msu300](https://doi.org/10.1093/molbev/msu300).

Nowack ECM, Weber APM. 2018. Genomics-informed insights into endosymbiotic organelle evolution in photosynthetic eukaryotes. *Annu Rev Plant Biol.* 69(1):51–84. doi:[10.1146/annurev-arplant-042817-040209](https://doi.org/10.1146/annurev-arplant-042817-040209).

O’Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44(D1):D733–D745. doi:[10.1093/nar/gkv1189](https://doi.org/10.1093/nar/gkv1189).

Ortea J, Espinosa J, Buske Y, Caballer M. 2013. Additions to the inventory of the sea slugs (Opisthobranchia and Sacoglossa) from Guadeloupe (Lesser Antilles, Caribbean Sea). *Rev Acad Canar Cienc.* 25:163–194.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* (Oxford, England). 20(2):289–290. doi:[10.1093/bioinformatics/btg412](https://doi.org/10.1093/bioinformatics/btg412).

Pierce SK, Curtis NE, Jeon KW. 2012. In: Jeon KW, editors. *International Review of Cell and Molecular Biology*. Cambridge, MA: Academic Press. p. 123–148.

Pierce SK, Curtis NE, Massey SE, Bass AL, Karl SA, Finney CM, et al. 2006. A morphological and molecular comparison between *Elysia crispata* and a new species of kleptoplastic sacoglossan sea slug (Gastropoda: Opisthobranchia) from the Florida Keys, USA. *Molluscan Res.* 26(1):23–38.

Pierce SK, Fang X, Schwartz JA, Jiang X, Zhao W, Curtis NE, Kocot KM, Yang B, Wang J, et al. 2012. Transcriptomic evidence for the expression of horizontally transferred algal nuclear genes in the photosynthetic sea slug, *Elysia chlorotica*. *Mol Biol Evol.* 29(6): 1545–1556. doi:[10.1093/molbev/msr316](https://doi.org/10.1093/molbev/msr316).

Powell KJ, Richens JL, Bramble JP, Han L-C, Sharma P, O’Shea P, Moses JE, et al. 2018. Photochemical activity of membrane-localised

polyketide derived marine natural products. *Tetrahedron*. 74(12): 1191–1198. doi:10.1016/j.tet.2017.10.056.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 26(6):841–842. doi: 10.1093/bioinformatics/btq033.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 11(1):1432. doi:10.1038/s41467-020-14998-3.

Raven JA, Walker DI, Jensen KR, Handley LL, Scrimgeour CM, McInroy SG, et al. 2001. What fraction of the organic carbon in sacoglossans is obtained from photosynthesis by kleptoplastids? An investigation using the natural abundance of stable carbon isotopes. *Mar Biol*. 138(3):537–545. doi:10.1007/s002270000488.

Robinson MD, McCarthy DJ, Smyth GK. 2009. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 26(1):139–140. doi:10.1093/bioinformatics/btp616.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 11(3):R25. doi:10.1186/gb-2010-11-3-r25.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP, et al. 2011. Integrative genomics viewer. *Nat Biotechnol*. 29(1):24–26. doi:10.1038/nbt.1754.

Rumpho ME, Worful JM, Lee J, Kannan K, Tyler MS, Bhattacharya D, Moustafa A, Manhart JR, et al. 2008. Horizontal gene transfer of the algal nuclear gene *psbO* to the photosynthetic sea slug *Elysia chlorotica*. *Proc Natl Acad Sci USA*. 105(46):17867–17871. doi:10.1073/pnas.0804968105.

Schliep KP. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* (Oxford, England). 27:592–593.

Seabold S, Perktold J. 2010. Statsmodels: econometric and statistical modeling with Python. *Proc Python Sci Conf*. 92–96. doi:10.25080/Majora-92bf1922-011.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I, et al. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res*. 19(6):1117–1123. doi:10.1101/gr.089532.108.

Smit AF, Hubley R, Green P. 2017. RepeatMasker Open-4.0. <http://repeatmasker.org>.

Soneson C, Love MI, Robinson MD. 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res*. 4:1521. doi:10.12688/f1000research.7563.2.

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 6(7): e21800. doi:10.1371/journal.pone.0021800.

Tillich M, Lehark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S, et al. 2017. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 45(W1):W6–W11. doi:10.1093/nar/gkx391.

Torres JP, Lin Z, Winter JM, Krug PJ, Schmidt EW. 2020. Animal biosynthesis of complex polyketides in a photosynthetic partnership. *Nat Commun*. 11(1):2882. doi:10.1038/s41467-020-16376-5.

Trench RK, Boyle JE, Smith DC, Harley JL. 1974. The association between chloroplasts of *Codium fragile* and the mollusc *Elysia viridis* - III. Movement of photosynthetically fixed 14C in tissues of intact living *E. viridis* and in *Tridachia crispata*. *Proc. R. Soc. Lond. B*. 185(1081):453–464. doi:10.1098/rspb.1974.0029.

Van Steenkiste NWL, Stephenson I, Herranz M, Husnik F, Keeling PJ, Leander BS, et al. 2019. A new case of kleptoplasty in animals: marine flatworms steal functional plastids from diatoms. *Sci Adv*. 5(7):eaaw4337. doi:10.1126/sciadv.aaw4337.

Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res*. 27(5):737–746. doi:10.1101/gr.214270.116.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 17(3):261–272. doi:10.1038/s41592-019-0686-2.

Wägele H, Deusch O, Händeler K, Martin R, Schmitt V, Christa G, Pinzger B, Gould SB, Dagan T, Klussmann-Kolb A, et al. 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(1):699–706. doi:10.1093/molbev/msq239.

Wernersson R, Pedersen AG. 2003. RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res*. 31(13):3537–3539. doi:10.1093/nar/gkg609.

Wick R. 2022. Filtlong. GitHub repository. <https://github.com/rrwick/Filtlong>.

Winnepernincx B, Backeljau T, De Wachter R. 1993. Extraction of high molecular weight DNA from molluscs. *Trends Genet*. 9(12): 407. doi:10.1016/0168-9525(93)90102-N.

Wisecaver JH, Alexander WG, King SB, Hittinger CT, Rokas A. 2016. Dynamic evolution of nitric oxide detoxifying flavohemoglobins, a family of single-protein metabolic modules in bacteria and eukaryotes. *Mol Biol Evol*. 33(8):1979–1987. doi:10.1093/molbev/msw073.

Editor: R. Mallarino