











RESEARCH

Open Access



# Chromosome-level genome assembly of the sacoglossan sea slug *Elysia timida* (Risso, 1818)

Lisa Männer<sup>1,2\*</sup>, Tilman Schell<sup>1,2†</sup>, Julia Spies<sup>1,3</sup>, Carles Galià-Camps<sup>4,5</sup>, Damian Baranski<sup>1,2</sup>, Alexander Ben Hamadou<sup>1,2</sup>, Charlotte Gerheim<sup>1,2</sup>, Kornelia Neveling<sup>6</sup>, Eric J. N. Helfrich<sup>1,2,3</sup> and Carola Greve<sup>1,2\*</sup>

## Abstract

**Background** Sequencing and annotating genomes of non-model organisms helps to understand genome architecture, the genetic processes underlying species traits, and how these genes have evolved in closely-related taxa, among many other biological processes. However, many metazoan groups, such as the extremely diverse molluscs, are still underrepresented in the number of sequenced and annotated genomes. Although sequencing techniques have recently improved in quality and quantity, molluscs are still neglected due to difficulties in applying standardized protocols for obtaining genomic data.

**Results** In this study, we present the chromosome-level genome assembly and annotation of the sacoglossan sea slug species *Elysia timida*, known for its ability to store the chloroplasts of its food algae. In particular, by optimizing the long-read and chromosome conformation capture library preparations, the genome assembly was performed using PacBio HiFi and Arima HiC data. The scaffold and contig N50s, at 41.8 Mb and 1.92 Mb, respectively, are approximately 30-fold and fourfold higher compared to other published sacoglossan genome assemblies. Structural annotation resulted in 19,904 protein-coding genes, which are more contiguous and complete compared to publicly available annotations of Sacoglossa with respect to metazoan BUSCOs. We found no evidence for horizontal gene transfer (HGT), i.e. no photosynthetic genes encoded in the sacoglossan nucleus genome. However, we detected genes encoding polyketide synthases in *E. timida*, indicating that polypropionates are produced. HPLC–MS/MS analysis confirmed the presence of a large number of polypropionates, including known and yet uncharacterised compounds.

**Conclusions** We can show that our methodological approach helps to obtain a high-quality genome assembly even for a "difficult-to-sequence" organism, which may facilitate genome sequencing in molluscs. This will enable a better understanding of complex biological processes in molluscs, such as functional kleptoplasty in Sacoglossa, by significantly improving the quality of genome assemblies and annotations.

**Keywords** High-quality reference genome, Arima HiC, PacBio HiFi, *Elysia timida*, Sacoglossa, Mollusca, Kleptoplasty, Polyketides, Biosynthesis

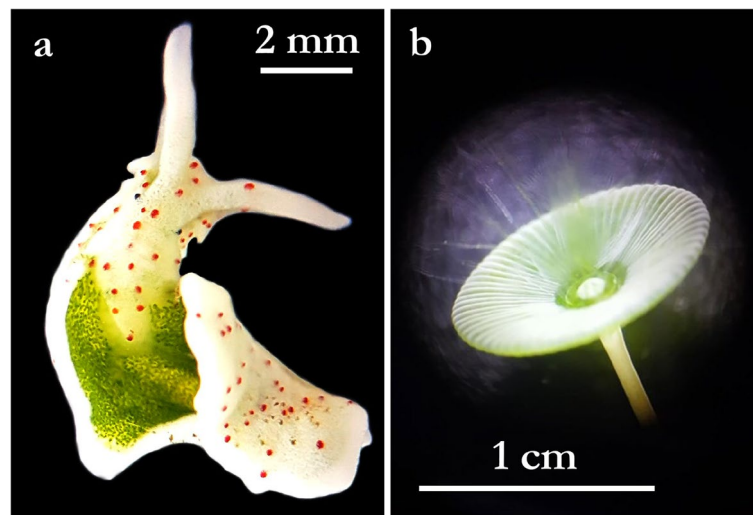
†Lisa Männer and Tilman Schell shared first authors.

\*Correspondence:

Lisa Männer  
lisa.maenner@senckenberg.de  
Carola Greve  
carola.greve@senckenberg.de

Full list of author information is available at the end of the article





**Fig. 1** The Sacoglossa *Elysia timida* [24] (a) and its unicellular food algae *Acetabularia acetabulum* (b). Photos were taken by C. Greve

## Introduction

Studying genomes of species is essential to comprehend the biology of organisms [1]. Third generation sequencing technologies, such as PacBio HiFi or Oxford Nanopore sequencing, have opened up the possibility of rapidly sequencing high-quality reference genomes of different organism groups at a reasonable price. However, sequencing methods and protocols are mainly developed and optimized for model organisms, especially human samples. In addition, DNA isolation for many non-model organisms is challenging. Sometimes even well-established sequencing methods sometimes do not work as expected, requiring special and adjusted handling [2]. Because of the existing bias towards developing methods for model species, certain taxonomic groups are still severely underrepresented in terms of genomic data and high-quality genome assemblies [3, 4].

Molluscs represent the second-largest animal phylum consisting of approximately 200,000 species – many of them still undescribed [5–7]. The diversity in molluscs is not only reflected in their manifold appearances, but also in divergent life cycles and habitats. Furthermore, they are of great ecological, economic, and medical significance [8, 9]. Considering their ecological and economical importance as well as species richness of the phylum, genomic resources of molluscs are still disproportionately low. It is worth noting that the number of molluscan reference genome assemblies on NCBI has more than doubled in recent years [10]. However, more high-quality and less fragmented genomes need to be sequenced and annotated to gain a deeper insight into the genomic diversity and evolutionary characteristics of molluscs.

An extraordinary group within molluscs are the sacoglossans, also known as “solar-powered” sea slugs. Some sacoglossan species have evolved an exceptional photosynthetic association with chloroplasts and are able to functionally store them from their food algae in the cells of their digestive gland, a process which is also referred to as functional kleptoplasty [11–13]. The role of the incorporated functional chloroplasts in the nutrition and metabolism of sacoglossan sea slugs is still highly controversial among scientists [14–18]. Most sacoglossans digest the kleptoplasts immediately (non-retention types) or after a few weeks (short-term retention types). However, six sacoglossan species are known to be capable of long-term chloroplast retention, in which the incorporated chloroplasts remain photosynthetically active for a period of two to ten months [19–23], such as the shallow-water Mediterranean species *E. timida* [24–26] (Fig. 1a; <https://youtu.be/MZRep08-81Y>).

It still remains unknown how these sacoglossan species keep the chloroplasts active in their digestive gland cells without support from the algal nucleus, since chloroplasts need to import many proteins encoded by nuclear genes for their activity [27]. Previous research suggested horizontal gene transfer (HGT) of photosynthetic genes from the algal nuclear genome to the nuclear DNA of the sacoglossan sea slug [28–30], but this was not supported by subsequent studies [31–33]. To determine whether the nuclear genome of *E. timida* contains traces of HGT from algae, the assembled contigs were screened for algal-like sequences. Polypropionate pyrones, produced in sacoglossans, have been suggested to be involved in the establishment and the maintenance of the association between sacoglossa and the incorporated chloroplasts by serving

antioxidant and photoprotective roles [34]. Therefore, to gain a better insight into the evolution and functionality of certain gene families, such as those encoding the polyketide synthases (PKS) involved in polypropionate pyrone biosynthesis, high-quality genome assemblies of *Sacoglossa* species are required in terms of completeness and contiguity.

In this work, we present the first annotated chromosome-level genome assembly of *E. timida*, and compare it with publicly available sacoglossan genomes. In addition, we provide a laboratory protocol that can be used to improve the sequencing of DNA from organisms whose sequencing is severely hampered by, for example, precipitated contaminants and DNA-bound metabolites that inhibit the sequencing polymerase. However, sequencing of such organisms can be improved by amplification-based protocols using currently available technologies and library kits [35]. This newly generated high-quality genome assembly may serve as a reference genome for future genetic investigations on kleptoplasty in *E. timida* and as a high-quality resource for studies on sacoglossans and molluscs in general.

## Material & methods

### Sample collection and sequencing

Specimens of *E. timida* (Fig. 1a) were collected in Cadaqués, Girona, Spain, in June 2021 (coordinates: 42.285173, 3.296461). Genomic DNA was extracted from a single individual using a CTAB-based method [36]. First, we prepared two PacBio ultra-low input libraries (including a long-range PCR amplification step using the PacBio polymerases A/B) using the SMRTbell® gDNA Sample Amplification Kit and the SMRTbell® Express Template Preparation Kit 2.0. Two SMRT cell sequencing runs were performed on the Sequel System IIe in CCS mode. In addition, to reduce potential PCR biases of the amplification polymerases A/B supplied by PacBio, we prepared two further libraries using the KOD Xtreme™ Hot Start DNA Polymerase (Merck), optimized for amplification of long and GC-rich DNA templates. For this, we combined the buffer, dNTPs and KOD polymerase from the KOD Xtreme Hot Start DNA Polymerase Kit with the ultra-low input primers from the PacBio SMRTbell gDNA Sample Amplification Kit. Otherwise, we followed the PacBio ultra-low input protocol (see also protocol adaptation and cycling conditions in Supplemental Tables S5). To obtain sufficient DNA to generate the SMRTbell libraries with the customized PacBio ultra-low input protocol, we performed two independent amplifications with the KOD polymerase and then pooled the DNA. These two customized PacBio ultra-low input libraries were then each sequenced on a single SMRTcell using the PacBio Sequel IIe and Revio

instruments, respectively. An initial attempt to sequence a PacBio standard/low-input library of these animals resulted in very poor sequencing results (Supplemental Table S6). The same was true for the attempt to sequence *E. timida* with Oxford Nanopore technology.

Chromatin conformation capture libraries were prepared using the Arima HiC Kit v01 (Arima Genomics) according to the manufacturer's low-input protocol with a slight modification in the initial sample preparation steps. Another whole specimen of *E. timida* from the same locality was first washed in seawater, then in deionised water, and finally ground with a pestle in a 1.5 mL tube. After preparing the specimen, we followed the manufacturer's instructions for proximity ligation. The proximally-ligated DNA was then converted into an Arima High Coverage HiC library according to the protocol of the Swift Biosciences® Accel-NGS® 2S Plus DNA Library Kit. The fragment size distribution and concentration of the Arima High Coverage HiC library was assessed using the TapeStation 2200 (Agilent Technologies) and the Qubit Fluorometer and Qubit dsDNA HS reagents Assay kit (Thermo Fisher Scientific, Waltham, MA), respectively. The library was sequenced on the NovaSeq 6000 platform at Novogene (UK) using a 150 paired-end sequencing strategy, with an expected output of 30 Gb.

RNA was extracted from a third individual from the same locality using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. Quality and concentration were assessed using the TapeStation 2200 (Agilent Technologies) and the Qubit Fluorometer with the RNA BR Reagents Assay Kit (Thermo Fisher Scientific, Waltham, MA). The RNA extraction was then sent to Novogene (UK) for Illumina paired-end 150 bp RNA-seq of a cDNA library (insert size: 350 bp) sequenced on a NovaSeq 6000, with an expected output of 12 Gb.

### Genome size estimation

The genome size was estimated following a flow cytometry (FCM) protocol with propidium iodide-stained nuclei described by Hare and Johnston (2012) [37]. Two fresh individuals of *E. timida* from Cadaqués were homogenized in a 1.5 mL tube with a pestle, and, as an internal reference standard, neural tissue from *Acheta domesticus* (female, 1C=2 Gb) was chopped with a razor blade in a petri dish. Ice-cold Galbraith buffer (2 mL) was used as the suspension medium. The suspensions were filtered each through a 42-µm nylon mesh, then stained with the intercalating fluorochrome propidium iodide (PI, Thermo Fisher Scientific) (final concentration 25 µg/mL), and treated with RNase A (Sigma-Aldrich) (final concentration 250 µg/mL). The mean red PI fluorescence signal of stained nuclei was quantified using a Beckman-Coulter

CytoFLEX flow cytometer with a solid-state laser emitting at 488 nm. Fluorescence intensities of at least 10,000 nuclei per measurement were recorded. We used the software CytExpert 2.3 for histogram analyses. After measuring the suspensions of *E. timida* and the internal reference standard separately, they were mixed. With this suspension mix, the total quantity of DNA per nuclei of *E. timida* was calculated as the ratio of the mean red fluorescence signal of the 2C peak of the stained nuclei of *E. timida* divided by the mean fluorescence signal of the 2C peak of the stained nuclei of the reference standard times the 1C amount of DNA in the reference standard. In total, two suspensions from two *E. timida* individuals were measured, each with four replicates that were measured on four different days to minimize possible random instrumental errors. The average of these eight measurements was calculated to estimate the genome size (1C) of *E. timida*. The value of the robust coefficient of variance (rCV), which should be about 5% or less, provides an estimate of the confidence level of the measurements.

Genome size and heterozygosity were estimated from a k-mer profile of the HiFi reads. First, count from Jellyfish 2.3.0 [38] was run with the additional parameters “-F 4 -C -m 21 -s 1,000,000,000 -t 96” and all HiFi reads as input. Second, a histogram was created from the resulting database with “jellyfish histo -t 96”. Third, GenomeScope 2.0 [39] in combination with R 4.3.1 was executed using the histogram as input. Additionally, *E. timida*'s genome size was also estimated from coverage distribution of mapped PacBio reads using ModEst [40], as implemented in backmap 0.5 [3] (<https://github.com/schellt/backmap>).

### Assembly strategy

Bioinformatic analyses were conducted with default parameters if not stated otherwise.

HiFi reads were called using a pipeline, which is running PacBio's tools ccs 6.4.0 (<https://github.com/PacificBiosciences/ccs>), actc 0.3.1 (<https://github.com/PacificBiosciences/actc>), samtools 1.15 [41], and DeepConsensus 1.2.0 [42]. All commands were executed as recommended in the respective guide for DeepConsensus ([https://github.com/google/deepconsensus/blob/v1.2.0/docs/quick\\_start.md](https://github.com/google/deepconsensus/blob/v1.2.0/docs/quick_start.md)) except --all was applied instead of --min-rq=0.88 for ccs.

To remove PCR adapters and PCR duplicates, which might originate from the PCR amplification during the ultra-low library preparation, PacBio's tools lima 2.6.0 (<https://github.com/PacificBiosciences/barcoding>) with options “--num-threads 67 --split-bam-named --same --ccs” and pbmarkdup 1.0.2-0 with options “--num-threads 84 --log-level INFO --log-file pbmarkdup.log --cross-library --rmdup” (<https://github.com/PacificBiosciences/pbmarkdup>) were applied, respectively.

We assembled the genome of *E. timida* from filtered PacBio HiFi reads of four SMRT cells using hifiasm 0.19.8 [43]. Subsequently, the primary contigs were processed. Contamination was filtered out by first running “screen genome” of FCS-GX 0.5.0 [44] with the corresponding database (downloaded on Dec 5th, 2023) and the NCBI taxonomy ID (154,625). Second, “clean genome” was executed with the action report created by “screen genome” and a minimum sequence length of 1 (--min-seq-len 1). Subsequently, the FCS filtered assembly was polished using a workflow which includes DeepVariant. First, the HiFi reads used for assembly were mapped against the contigs with minimap2 2.26 [45, 46] and the options “-a -x map-hifi”. The bam file was sorted by coordinate with samtools 1.19.1 and duplicated HiFi reads were removed with Picard 3.1.0 [47] MarkDuplicates and the option --REMOVE\_DUPLICATES. The assembly fasta and filtered bam files were indexed with samtools faidx and index commands, respectively. To call SNPs, DeepVariant 1.5.0 [48] was applied. To keep only homozygous variants, SNPs were subsequently filtered using bcftools view 1.13 [41] with the options -f 'PASS' -i 'GT="1/1"'. Then, the vcf file containing the homozygous SNPs was indexed with tabix from htlib 1.17 [49] to finally apply the variants in the filtered assembly with bcftools consensus, which is from here on referred to as polished assembly. Haplotigs were purged from the polished assembly with purge\_dups 1.2.6 ([https://github.com/dfgvan/purge\\_dups](https://github.com/dfgvan/purge_dups)) together with minimap 2.24 for mapping HiFi reads and self-alignment of the assembly according to the guidelines ([https://github.com/dfgvan/purge\\_dups/tree/v1.2.6?tab=readme-ov-file#--pipeline-guide](https://github.com/dfgvan/purge_dups/tree/v1.2.6?tab=readme-ov-file#--pipeline-guide)), except “-x map-hifi” was applied during HiFi mapping and high coverage contigs were kept when running get\_seqs (-c). Prior to HiC scaffolding, blobtoolkit 4.1.4 [50] was used to evaluate if contamination was still present. Taxonomic assignment for blobtools was conducted with blastn 2.15.0+ [51] and the options “-task megablast -outfmt '6 qseqid staxids bitscore std' -num\_threads 96 -evaluate 1e-25”. Information on coverage per contig was obtained from mapping HiFi reads used for assembly back to the assembly itself via backmap 0.5 [3, 40] in combination with minimap 2.26 [45], samtools 1.17, Qualimap 2.3 (bamqc; [52]), bedtools 2.30.0 [53] and R 4.0.3 [54]. Contigs were excluded if taxonomic assignment was not one of no-hit, Mollusca, Chordata or Arthropoda, GC was lower than 0.287, or average coverage was lower than 15.

The polished contigs that were filtered via blobtools, were scaffolded with Arima HiC reads in yahs 1.1 [55]. To do so, HiC reads were first mapped with the Arima mapping pipeline ([https://github.com/VGP/vgp-assembly/blob/master/pipeline/salsa/arima\\_mapping\\_pipeline.sh](https://github.com/VGP/vgp-assembly/blob/master/pipeline/salsa/arima_mapping_pipeline.sh)), in combination with bwa mem 0.7.17 [56], samtools

1.15.1, picard 2.27.1, and java 1.8.0 [57]. Afterwards, the bam file was processed along with the contigs in yahs. A hic file was created with the tool “juicer pre” from yahs, which was loaded together with the respective assembly file into Juicebox 1.11.08 [58, 59] for manual curation.

To test for support on HGT between photosynthetic genes from the algal nuclear genome to the nuclear DNA of *E. timida*, the blast search used for taxonomic assignment in blobtools was screened for hits of *Acetabularia acetabulum* and *Ulva compressa* outside the contigs assigned as Chlorophyta. Furthermore, Chlorophyta blast hits were screened if respective database sequences are chloroplast or nuclear sequences. Finally, a HiC scaffolding was conducted as explained above, except that all contigs, which were filtered out due to taxonomic assignment to Chlorophyta, were added again to the polished and blobtools filtered contig set.

Contigs of mitochondrial origin were filtered out after manual curation. To do so, MitoHiFi 3.2.1 [60] was applied together with the available mitochondrial genome sequence of *E. timida* (KU174946.1; [61]) and the scaffolds after manual curation. Subsequently, blastn 2.15.0+ was applied to find curated scaffolds similar to our *E. timida* mitochondrial genome sequence. All scaffolds with an alignment length equal to the mitogenome length were filtered out. Additionally, all contigs flagged as circular by hifiasm were filtered out, which only included sequences that remained un-scaffolded.

During different stages of the assembly, quality controls were conducted. Basic contiguity statistics were calculated with Quast 5.2.0 [62]. Single copy orthologs of the provided metazoan set were searched with BUSCO 5.5.0 [63]. Completeness regarding k-mers and QV values were obtained with Meryl 1.3 and Merqury 1.3 [64]. Mapping coverage distribution of HiFi reads was checked using backmap 0.5.

### Annotation

Masking of repetitive regions was conducted by identifying repeat families with RepeatModeler 2.0.5 [65, 66]

and the dependencies rmbast 2.14.1+ as search engine, and TRF 4.09 [67], RECON 1.08 [68], RepeatScout 1.0.6 [69], as well as RepeatMasker 4.1.6 [70]. LTR structural discovery pipeline was enabled (-LTRStruct) with the dependencies GenomeTools 1.6.2 [71], LTR\_Retriever 2.9.9 [72], Ninja 0.97 [73], MAFFT 7.520 [74], and CD-HIT 4.8.1 [75]. The resulting repeat families were used as repeat library in RepeatMasker 4.1.6 [70] together with the options “-xsmall -no\_is -e ncbi -pa 10 -s” and the dependencies rmbast 2.14.1+ as search engine, HMMer 3.4 (hmmer.org) and TRF 4.09 [67].

Structural annotation of the *E. timida* genome assembly, was conducted with BRAKER 3.0.8 [76–91]. RNAseq data from *E. timida* was mapped against the genome assembly using HISAT 2.2.1 [92], and the bam file sorted by coordinate with samtools 1.19.1 was provided to BRAKER with --bam. In addition, we downloaded the protein sets from the high-quality annotations of six molluscan species, which were then used as evidence during structural annotation: *Aplysia californica* (GCF\_000002075.1; [93]), *Gigantopelta aegis* (GCF\_016097555.1; [94]), *Mizuhopecten yessoensis* (RefSeq: GCF\_002113885.1; [95, 96]), *Octopus sinensis* (GCF\_006345805.1; [97]), *Pecten maximus* (GCF\_902652985.1; [98]), and *Pomacea canaliculata* (RefSeq: GCF\_003073045.1; [99, 100]) (Table 1). Before the protein sequences were concatenated and provided with --prot\_seq to BRAKER, we evaluated BUSCO completeness, number of genes, and corresponding contiguity of the annotation to verify the high quality of the mollusc protein set. BRAKER was executed with the additional parameters --gff3 --threads=96 --busco\_lineage=metazoa\_odb10.

Quality controls of the *E. timida* annotation were conducted by searching for single copy orthologs with BUSCO and by calculating basic contiguity statistics of the annotated features (e.g. genes, mRNAs, etc.). The functional annotation of the predicted *E. timida* proteins was conducted using InterProScan 5.64–96.0 [101]. Databases and tools which were used during the

**Table 1** The protein sets used as evidence for the annotation of the genome assembly

Species	Metazoa BUSCO (n: 954)	Number of protein sequences
<i>Aplysia californica</i>	C:97.8%[S:75.8%,D:22.0%],F:0.8%,M:1.4%	26,656
<i>Gigantopelta aegis</i>	C:98.5%[S:70.5%,D:28.0%],F:0.8%,M:0.7%	33,249
<i>Mizuhopecten yessoensis</i>	C:98.6%[S:75.2%,D:23.4%],F:0.4%,M:1.0%	41,567
<i>Octopus sinensis</i>	C:98.4%[S:69.2%,D:29.2%],F:0.8%,M:0.8%	36,112
<i>Pecten maximus</i>	C:98.5%[S:74.7%,D:23.8%],F:0.4%,M:1.1%	39,918
<i>Pomacea canaliculata</i>	C:98.2%[S:74.9%,D:23.3%],F:0.4%,M:1.4%	40,391

operation with InterProScan [102], are shown in detail in Supplemental Table S1. Furthermore, InterProScan was executed in the same way using the four protein sets from annotations of *Elysia chlorotica*, *Elysia crispata*, *Elysia marginata*, and *Plakobranthus ocellatus* as input (Table 4).

#### Comparison of PKS encoding genes from sacoglossans

To detect the presence and copy number of PKS genes in the annotation of *E. timida*, the PKS coding sequences from the genomes of *E. chlorotica*, *Elysia diomedea* and *P. ocellatus* from Torres et al. (2020) [34] were used as a reference (Supplemental Table S2). We also evaluated the presence and copy number of PKS genes in the annotations of *E. chlorotica*, *E. crispata*, *E. marginata* and *P. ocellatus*.

Nucleotide sequences were translated into their corresponding six protein reading frames using Geneious Prime 2023.0.4. We proceeded with the reading frames which did not contain any stop codon. These sequences were then blasted against the annotations of *E. timida*, *E. chlorotica*, *E. crispata*, *E. marginata*, and *P. ocellatus* using blastp 2.14.0+ [51] and the options “-task blastp -outfmt '6 qseqid staxids bitscore std qlen slen' -evalue 1e-25”. We furthermore filtered out all blast hits with an identity lower than 80%.

#### Phylogenetic analysis of FASs and PKSs

Transcript sequences of the FASs and PKSs from *E. timida*, *E. chlorotica*, *E. diomedea* and *P. ocellatus* were retrieved, MUSCLE aligned, manually refined and realigned with MUSCLE. A maximum likelihood phylogenetic tree was created in MEGA11 using 1000 bootstrap replications [103]. The branch lengths are scaled according to the number of substitutions per site.

#### Breeding conditions of *E. timida* for polypropionate extraction

Individuals of *E. timida* from Cadaqués were reared in a climate chamber at 20 °C, where they spawned. Hatching specimens (F1) were reared to adulthood under the same conditions as their parental line. The sea slugs were housed in small and transparent plastic containers and kept in a 12:12 day-night rhythm (153 lx;  $\lambda$ p: 545 nm) (Supplemental Figure S1). Once a week water was changed and food algae (*A. acetabulum*; Fig. 1b) were provided.

#### Extraction and HR-HPLC–MS measurement

Three F1 specimens were independently homogenized by blending in 0,2 Mol Tris–HCl at pH 7 in a volume of 1 mL at room temperature. The extraction was performed using ethyl acetate in tenfold excess. The crude extract

was dried under reduced pressure and re-dissolved in methanol. The extracts were measured by high-performance liquid chromatography-electrospray ionization high-resolution mass spectrometry (HPLC-ESI-HR-MS) using an Ultimate 3000 LC system coupled to an Impact II QTOF (Bruker) high-resolution mass spectrometer. The extract was separated on an Acquity UPLC BEH C18 column (130 Å, 1.7  $\mu$ m particle size, 2.1 mm  $\times$  100 mm) with a gradient flow of 0,4 mL/min from 5 to 95% solvent B (acetonitrile + 0,1% formic acid) over a time span of 14 min. The data was acquired in positive mode at a scan range between  $m/z$  100 to  $m/z$  1200 and analyzed using the Bruker software DataAnalysis 4.3 and MetaboliteDetect 2.1. HPLC was used to separate the constituents of the crude extract based on their physico chemical properties, high resolution mass spectrometry was used to determine the exact mass and sum formula of the compounds of interest and tandem mass spectrometry results in characteristic fragmentation patterns (fingerprints) of each molecule. The MS/MS fingerprints of characterized polypropionates were used to identify related compounds based on their similar MS/MS fragmentation patterns to characterized polypropionates as reported by Torres et al. (2020) [34]. Similar MS/MS fragmentation patterns are indicative of structural relatedness between two compounds.

#### Molecular networking and visualization

The HPLC–MS/MS datasets obtained from analysis of the crude extracts from *E. timida* specimens were uploaded to Global Natural Products Molecular Networking (GNPS) to generate a molecular network, setting the minimum matched peaks to 7 and the cosine score to 0.6 [104]. The software Cytoscape 3.9.1 was used to visualize the molecular network [105]. The excerpt of the network that visualizes the polypropionates was identified by the presence of nodes representing characterized compounds [34]. Putative polypropionates were identified based on their exact mass (sum formula) and similar fragmentation patterns to characterized polypropionates.

#### Construction of proposed sequence for EtPKS1 mRNA

The putative sequence of the mRNA for *E. timida* PKS1 (EtPKS1) was proposed based on sequence similarity with the transcript for the EcPKS1 (Accession number: MT348433). The conserved “GHSMGE” motif in the acyltransferase domain of PKS1 was identified on nucleotide level and used as a bait for the identification of the genomic area encoding the EtPKS1 [34]. An excerpt of the genomic sequence around this sequence motif was translated in the three forward translation frames to map the translations with the EtPKS1 transcript and annotate the exons.

## Results

### Genome size estimation

Flow cytometry results were represented as histograms displaying the relative propidium iodide fluorescence intensity which we received after a simultaneous analysis of *E. timida* 2C and the house cricket *A. domesticus* 2C as an internal reference standard (Supplemental Figure S2). The obtained average haploid genome sizes for the two individuals of *E. timida* were 898.00 and 891.78 Mb, respectively. From all eight measurements, including both individuals, a genome size of 894.89 Mb was estimated (Supplemental Table S3).

Mapping based genome size estimation with ModEst based on the final genome assembly resulted in 632 Mb. Based on k-mers, the genome size was estimated to be 548.2 Mb and heterozygosity 0.794% (Supplemental Figure S3). The heterozygosity value of *E. timida* is higher compared to other sacoglossan species (Supplemental Table S4).

### Sequencing

The four PacBio sequencing runs using the standard PacBio ultra-low and the customized PacBio ultra-low DNA input library preparations yielded a total polymerase read length of 510 Gb, 431 Gb, 612 Gb, and 1,360 Gb, respectively (Supplemental Table S6 and Supplemental Figure S4). Illumina sequencing of Arima HiC and RNAseq libraries resulted in 95.7 and 40.6 million read pairs, corresponding to 28.7 and 12.2 Gb, respectively.

Raw data of PacBio HiFi reads (subreads) and raw Arima HiC reads which were used for genome assembly, raw RNAseq reads as well as the final assembly and annotation can be publicly accessed via BioProject PRJNA1119176 and this link: <https://genome.senckenberg.de/download/etim/>.

### Assembly

HiFi calling and subsequent PCR duplicate removal resulted in more than 19 million HiFi reads with a total length of 108 Gb and an N50 of 6,143 bp (more details in Supplemental Figure S4). Given the genome size estimates based on FCM (895 Mb) and ModEst (632 Mb), the theoretical coverages were calculated to be 120x and 170x, respectively. FCS-GX identified 873 sequences (53.3 Mb) containing contamination (Supplemental Table S7), which were excluded or trimmed (Supplemental Table S8). Nevertheless, after polishing and purging, the blobplot still showed the presence of contamination (Supplemental Figure S5). Blobtools returned five sequences taxonomically assigned to Chlorophyta, of which three were assigned to *A. acetabulum* and two to *U. compressa* (Supplemental Table S9). In total, 3084

additional contigs (100.8 Mb), including the five assigned to Chlorophyta, were removed before proceeding with HiC scaffolding (blobplot after removal of contamination in Supplemental Figure S11). After manual curation, scaffolding resulted in 15 chromosome-scale sequences (Fig. 2), showing no contamination (Supplemental Figure S6). These 15 scaffolds made up 89.1% of the assembly's total length. The total length of the final *E. timida* genome assembly was 754 Mb with a scaffold and contig N50 of 41.8 Mb and 1.92 Mb, respectively. Furthermore, the assembly had a QV of 58.3. Additional quality metrics of the final *E. timida* assembly and a comparison to publicly available sacoglossan genome assemblies are shown in Fig. 3 and Table 2.

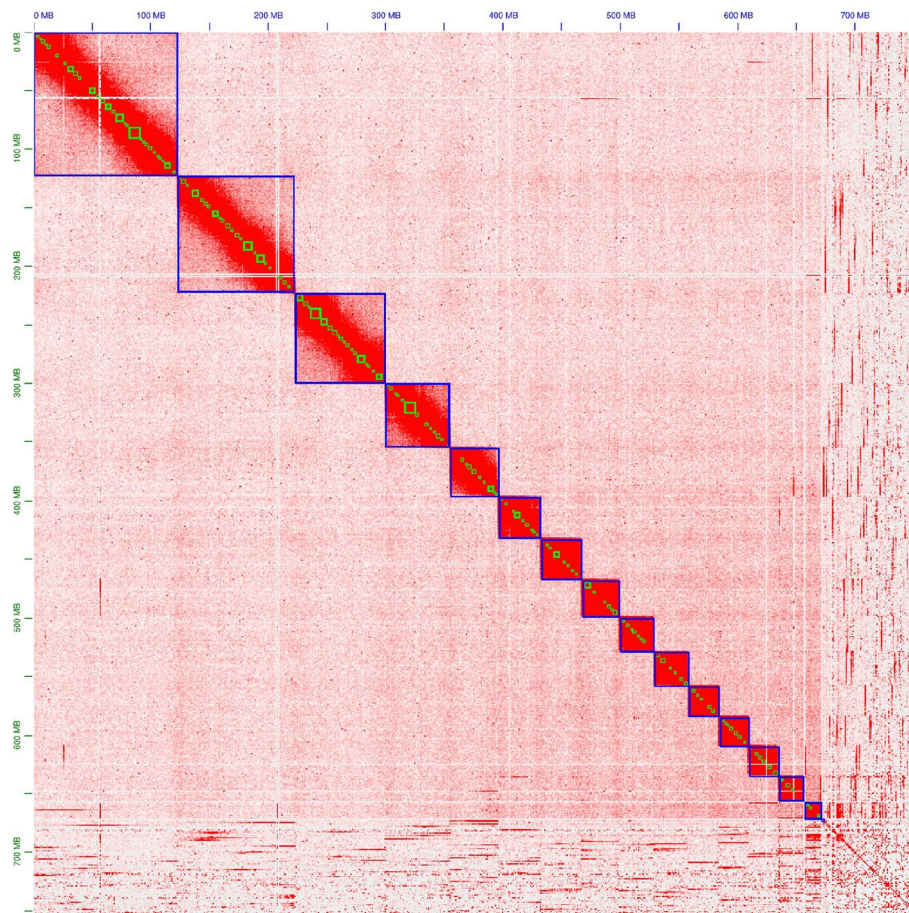
The contig N50 of the *E. timida* genome assembly is fourfold higher than the highest contig N50 so far achieved for a sacoglossan species' genome assembly (*E. crispata*: 0.45 Mb), while the BUSCO values of *E. timida* were similar to those of the other sacoglossans (Table 2). Only duplicated BUSCOs were higher for *E. timida* compared to the other assemblies. As HiC data have not yet been generated for other Sacoglossa, the *E. timida* assembly has a manyfold higher scaffold N50 in comparison (Table 2).

### Checking for support regarding horizontal gene transfer

The sequence similarity search for taxonomic assignment in blobtools returned five contigs, which were classified as *A. acetabulum* or *U. compressa* (see assembly results). Further inspection of the blast output showed that the five Chlorophyta contigs only generated hits leading to taxids of *A. acetabulum* (NCBI:txid35845) or *U. compressa* (NCBI:txid63659) and not to any *E. timida* contigs (Supplemental Table S10). All target sequences of blast hits with taxids of *A. acetabulum* or *U. compressa* are sequences from a chloroplast of these two species (Supplemental Table S11). HiC scaffolding was performed to test whether the five Chlorophyta contigs could be linked to other nuclear sequences of *E. timida*. This resulted in splitting one contig but none of the Chlorophyta contigs was linked to other *E. timida* sequences (Supplemental Table S12).

### Annotation

RepeatModeler identified sequences of 1,886 repeat families with a total length of 1,705,165 bp. Subsequently, RepeatMasker annotated 44.3% of the assembly as repetitive, of which the majority of repetitive families were labelled as unclassified (32.5% of the assembly). Contiguity statistics and BUSCO results of the *E. timida* annotation compared to other Sacoglossa annotations are shown in Table 3. The values of the complete BUSCOs range between 86.1% and 92.9%



**Fig. 2** Contact map after yahs scaffolding using Arima HiC data and manual curation. Blue and green squares mark scaffolds and contigs, respectively. Higher number of contacts is represented by higher intensity of the colour

which confirm the high quality of the protein sets of the published sacoglossan annotations.

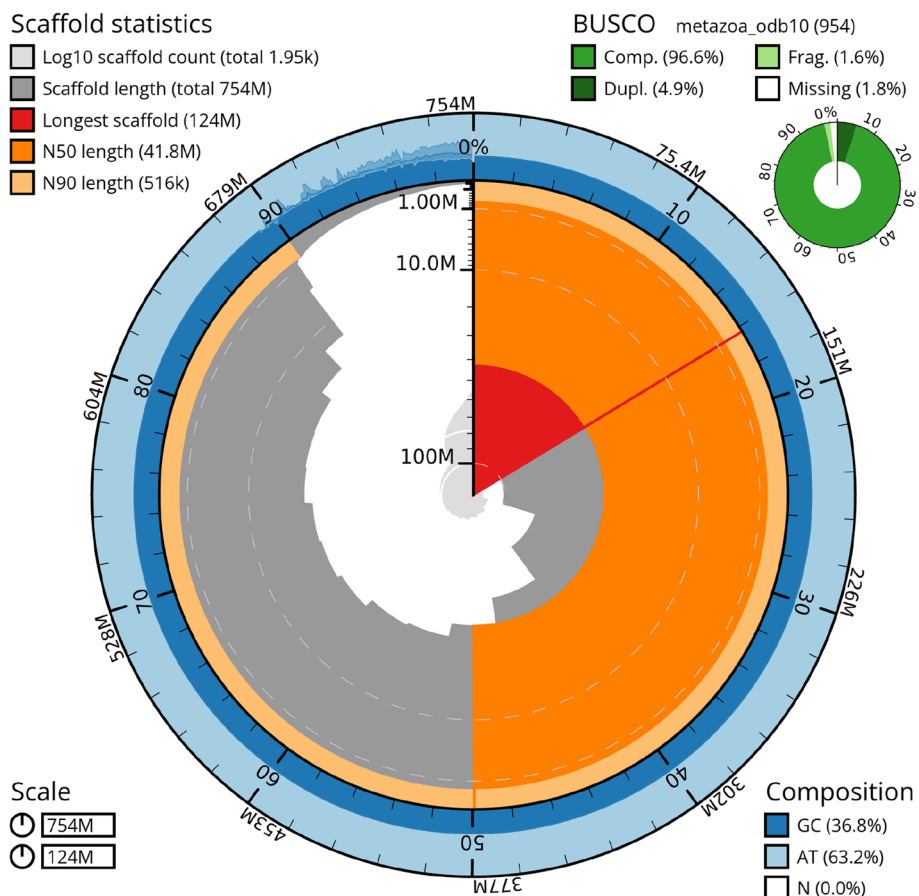
In comparison to annotations of other sacoglossan genome annotations, *E. timida* was the most complete and least fragmented in terms of BUSCOs. Additionally, mean CDS per mRNA and median gene length among others were also highest for *E. timida*, while the number of single CDS mRNAs was lowest. Total gene and CDS space were similar or lower compared to other annotated sacoglossan species.

InterProScan assigned at least one functional annotation to 19,489 (97.9%) different *E. timida* protein sequences with one of the applied analyses. Furthermore, 12,368 (62.1%) different *E. timida* protein sequences were annotated with at least one Gene Ontology (GO) term, whereas this percentage was lower in other Sacoglossa with a wider range (from 22 to 46%). A comparison of functionally annotated protein sequences with other Sacoglossa can be found in Table 4.

#### PKS presence in *E. timida* and comparison of PKSs in sacoglossans

All genes encoding FAS and PKS from *E. chlorotica*, *E. diomedea* and *P. ocellatus* [34] were found in the annotation of *E. timida* using BLAST search. The FAS and PKS protein sets were also blasted against the genome annotations of *E. chlorotica*, *E. crispata*, *E. marginata* and *P. ocellatus* and found in all of them (Supplemental Figure S7 and Supplemental Table S13). The transcripts of the genes associated with FASs form a distinct clade from the PKS transcripts. The PKS1 transcripts and PKS2 transcripts also form two separate clades. The transcripts of *P. ocellatus* were the most phylogenetically distinct sequences for FAS, PKS1 and PKS2 compared to the transcripts of the *Elysia* species. Before filtering the blast hits, the nine PKS or FAS encoding genes from three sacoglossan species were detected in all of the other sacoglossans. To correct for false positives, we filtered out all the blast hits with an e-value above  $1e^{-25}$  and a percentage of identical positions below 80%. After filtering,





**Fig. 3** Snail plot of the final genome assembly. The plot created with blobtoolkit visualizes amongst others scaffold count, lengths, length distribution, nucleotide composition, and recovered BUSCOs

**Table 2** Quality metrics of the *E. timida* genome assembly in comparison to available sacoglossan species

	<i>Elysia timida</i>	<i>Elysia chlorotica</i>	<i>Elysia marginata</i>	<i>Plakobranthus ocellatus</i>	<i>Elysia crispata</i>
Citation	This study	[106]	[33]	[33]	[107]
Accession number		GCA_003991915.1	GCA_019649035.1	GCA_019648995.1	GCA_033675545.1
<b>Contigs</b>					
Number of sequences	2539	41,624	211,550	280,271	8089
Total length (Mb)	754.4	540.5	707.4	849.1	786.4
N50 (Mb)	1.92	0.03	0.006	0.005	0.45
<b>Scaffolds</b>					
Number of sequences	1946	9987	14,149	8647	-
Total length (Mb)	754.5	557.5	790.3	927.9	-
N50 (Mb)	41.8	0.44	0.23	1.45	-
% of Ns	0.016	3.04	10.43	8.48	-
<b>BUSCO (%; N = 954)</b>					
C	96.6	94.8	89.2	95.3	96.9
S	91.7	94.1	88.5	94.7	96.3
D	4.9	0.7	0.7	0.6	0.6
F	1.6	3.1	7.2	2.7	1.6
M	1.8	2.1	3.6	2.0	1.5

**Table 3** Quality metrics of the *E. timida* genome annotation in comparison to available sacoglossan species. \*For annotations of *E. marginata* and *P. ocellatus* only “gene” and “CDS” features are annotated, so that no metrics regarding mRNAs and mean CDSs/mRNA as well as single CDS genes are shown

	<i>Elysia timida</i>	<i>Elysia chlorotica</i>	<i>Elysia marginata</i>	<i>Plakobranthus ocellatus</i>	<i>Elysia crispata</i>
Number					
Gene	15,809	23,871	70,752	77,230	67,429
mRNA	19,904	23,871	-	-	70,120
CDS	211,587	155,011	249,028	271,856	329,620
Mean ratio					
mRNAs/ gene	1.26	1.00	-	-	1.04
CDSs/ mRNA	10.63	6.49	3.52*	3.52*	4.70
Median length (bp)					
Gene	9,325	5,452	1,708	1,561	2,419
mRNA	10,579	5,452	-	-	2,514
CDS	126	132	147	148	134
Total space (bp)					
Gene	236,595,571	215,824,722	314,620,341	363,188,163	375,920,318
mRNA	236,595,571	215,824,722	-	-	375,682,798
CDS	28,112,659	30,877,231	55,967,689	61,040,586	64,065,507
Number of single					
Exon mRNAs	1,441	3,514	-	-	16,366
CDS mRNAs	1,441	3,514	23,332*	28,380*	16,366
BUSCO (%; N=954)					
C	98.7	91.5	86.1	91.1	92.9
S	75.4	91.2	85.3	89.9	87.2
D	23.3	0.3	0.8	1.2	5.7
F	0.4	3.8	9.1	6.2	2.7
M	0.9	4.7	4.8	2.7	4.4

**Table 4** Number of protein sequences annotated structurally (Total) and functionally with at least one analysis in InterProScan or with at least one GO term

	<i>Elysia timida</i>	<i>Elysia chlorotica</i>	<i>Elysia marginata</i>	<i>Plakobranthus ocellatus</i>	<i>Elysia crispata</i>
Total	19,904	23,871	70,752	77,230	70,120
Functional annotated	19,489 (97.9%)	21,889 (91.7%)	61,319 (86.7%)	59,564 (77.1%)	46,665 (66.6%)
GO term annotation	12,368 (62.1%)	10,979 (46.0%)	18,150 (25.7%)	17,772 (23.0%)	15,626 (22.3%)

the number of hits reduced drastically in all sacoglossans (Supplemental Table S13). Before filtering, between two and 38 gene hits were found in all sacoglossan species. However, after filtering, the maximum number of gene copies was found in the annotation of *E. marginata*, where we had four blast hits of the EcPKS1 gene. Some sacoglossan annotations had no FAS or PKS gene hits after filtering. The genes encoding EtFAS, EtPKS1 and EtPKS2 were annotated in the genome of *E. timida* (Table 5). The genes encoding the EtFAS and EtPKS2

**Table 5** Sequence identifiers for the nucleotide and amino acid sequences of EtFAS, EtPKS1 and EtPKS2

Coding region in genome assembly	Transcript level	Protein level
Scaffold_4:11,989,323–12,013,722	g10811.t1	EtFAS
Scaffold_6:1,080,882–1152789	Assembled manually	EtPKS1
Scaffold_12:1,113,801–1144318	g4188.t1	EtPKS2

were automatically annotated by BRAKER 3.0.8. In contrast, the gene encoding the EtPKS1 was not automatically recognized, and was manually annotated based on the alignment with the amino acid sequence of EcPKS1 (Supplemental Figure S8). The presence of PKS encoding genes in the genome of *E. timida* was investigated to obtain an indication of which polypropionates can be produced by the sea slug.

### Identification of putative polypropionates

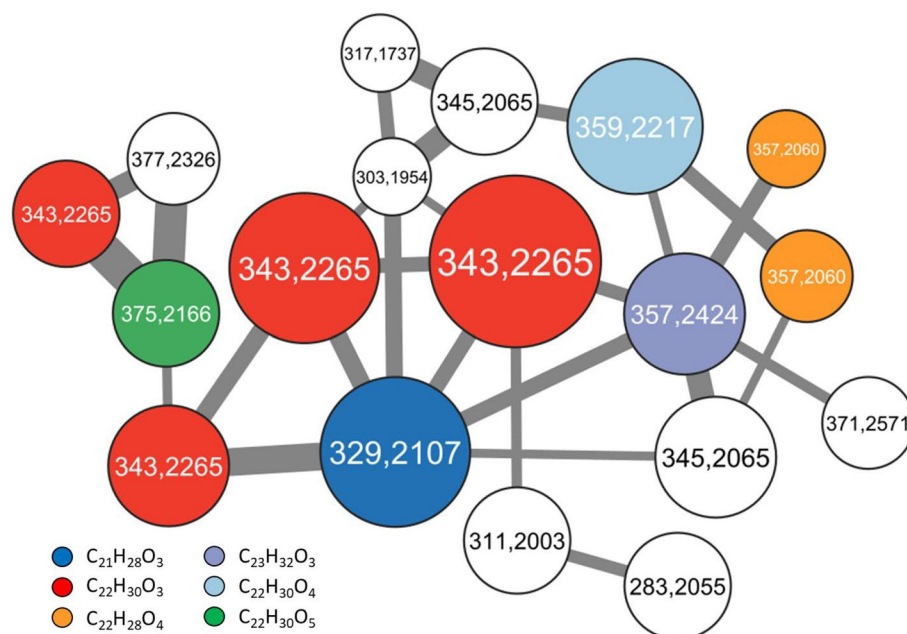
To investigate the spectrum of polypropionates produced by the sea slug, three adult specimens were extracted. Analysis of high-performance liquid chromatography coupled to high resolution tandem mass spectrometry (HPLC–MS/MS) data of the crude extract resulted in the identification of putative polypropionates (Supplemental Figure S9). The obtained HPLC–MS/MS data were visualized to represent the relatedness between the putative compounds (Fig. 4). Each node in the molecular network represents a polypropionate and is labelled with the detected mass. The edge line width visualizes the degree of relatedness between two compounds. More than half of the 18 putative polypropionates detected matched characterized compounds that were reported from different *Elysia* and *Plakobran-chus* species (Table 6) [108]. Additionally, eight putative

polypropionates were detected that have not been structurally characterized.

The stereochemistry of the identified polypropionates cannot be determined by mass spectrometry. As a result, planar structures are shown. In the excerpt of the molecular network isomers of polypropionates appear as separate nodes which are labelled with the same mass-to-charge ratio. The complex polyketide scaffolds are hypothesized to derive from spontaneous light-induced cyclizations and tailoring reactions [109]. This results in the wide spectrum of polypropionates produced by sacoglossans (Supplemental Figures S9 and S10) [108].

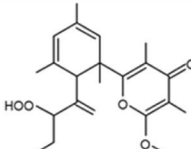
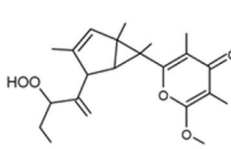
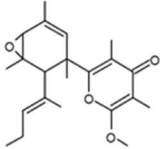
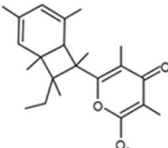
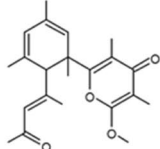
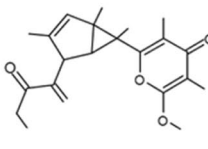
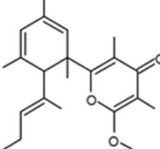
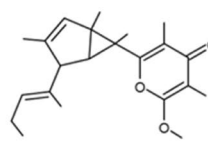
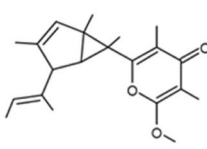
### Discussion

The adaptive potential and remarkable survival mechanisms of Sacoglossa have been the subject of many studies. The inclusion of fully annotated Sacoglossa genomes in these studies is essential to properly investigate the genetic processes underlying functional kleptoplasty and to understand its functional role. Our *E. timida* genome assembly achieves the highest values of contiguity, of BUSCOs completeness and accuracy compared to available sacoglossan genome assemblies and therefore makes a valuable contribution to this understanding. In addition, we identified genes encoding PKS1 and PKS2 in the genome annotation of *E. timida*, suggesting that *E. timida* is a putative producer of polypropionates.



**Fig. 4** Excerpt of a molecular network showing detected polypropionates from crude extracts of *E. timida*. Putative polypropionates were clustered based on their similar MS/MS fragmentation patterns. Each node represents a polypropionate and is labelled with the detected mass. Masses that correspond to characterized polypropionates are color-coded. White nodes correspond to putative polypropionates that were not characterized yet. The node size corresponds to the production level and the edge width represents the relatedness between two compounds. The thresholds for the cluster were set to 7 minimum matched peaks and a cosine score of 0.6

**Table 6** Overview of polypropionates isolated from kleptoplastic sacoglossans. Compounds were detected by LC-HR-MS/MS analysis of crude extracts. Only masses corresponding to characterized polypropionates are shown

Detected mass-to-charge ratio	Predicted sum formula	Polypropionates
$m/z$ 375,2166 [M+H] <sup>+</sup>	C <sub>22</sub> H <sub>30</sub> O <sub>5</sub> Δ 0 ppm	 Tridachiapyrone J  Phototridachiapyrone J
$m/z$ 359,2213 [M+H] <sup>+</sup>	C <sub>22</sub> H <sub>30</sub> O <sub>4</sub> Δ -1,07 ppm	 Tridachione
$m/z$ 357,2424 [M+H] <sup>+</sup>	C <sub>23</sub> H <sub>32</sub> O <sub>3</sub> Δ -0,05 ppm	 Ocellapyrone A
$m/z$ 357,2060 [M+H] <sup>+</sup>	C <sub>22</sub> H <sub>28</sub> O <sub>4</sub> Δ -0,10 ppm	 Tridachiapyrone I  Tridachiapyrone E
$m/z$ 343,2267 [M+H] <sup>+</sup>	C <sub>22</sub> H <sub>30</sub> O <sub>3</sub> Δ -0,21 ppm	 (iso)-9,10-deoxytridachione  Photodeoxytridachione
$m/z$ 329,2106 [M+H] <sup>+</sup>	C <sub>21</sub> H <sub>28</sub> O <sub>3</sub> Δ -1,52 ppm	 15-norphotodeoxytridachione

It is hypothesized that polypropionates are involved in the establishment and maintenance of the association between Sacoglossa and the incorporated chloroplasts [34]. Ireland and Scheuer (1979) [110] found that in Sacoglossa, fixed carbon acquired from de novo chloroplast photosynthesis was incorporated into polypropionates. The polypropionates act via oxidative and photocyclization pathways which are suggested to behave like sunscreen and prevent the sea slugs from photosynthetic

damage [110–115]. Polypropionates would therefore be needed in species with a ‘photosynthetic’ lifestyle, such as *E. timida*.

The hypothesis of HGT of algal nucleic photosynthetic genes to the nuclear DNA of the sea slug is not supported by the data presented in this study. First, despite the high coverage, all contigs assigned to Chlorophyta were assembled separately and could not be linked to other *E. timida* sequences with HiC data. Second, all

blast hits returning a taxid of the two algae detected (*A. acetabulum* and *U. compressa*) are on the five contigs assigned as Chlorophyta, which implies that there are no sequences similar to these algae in the *E. timida* nuclear genome. Thirdly, only sequences similar to chloroplasts were found in the contamination screening, showing no support for the similarity to the nuclear sequences of the two detected algae sequenced together with *E. timida*. These last two statements may be less reliable, as the database probably does not reflect the true diversity of *A. acetabulum* and *U. compressa*, which would be required for a comprehensive assignment. Overall, neither the HiFi nor the HiC data support a linkage of the detected algal sequences to *E. timida* sequences, and no algal-like sequences were found in the assembly, apart from the expected chloroplast sequences (which were filtered out during contamination screening). These results are consistent with those of Maeda et al. (2021) [33], Chan et al. (2018) [32] and Wägele et al. (2011) [31], who also found no support for HGT. Nevertheless, HGT between algae and *E. timida* needs to be further investigated in future studies.

It is surprising that *U. compressa* sequences were found in addition to *A. acetabulum*. As a wild caught animal was used for sequencing, we cannot exclude the possibility of contamination from the body surface. Nevertheless, the specimen was starved in a lab culture, where *U. compressa* was not present, and previous experiments have shown that *E. timida* can feed on algae other than *A. acetabulum* [116, 117]. *Ulva compressa* is a common alga of the Catalan coast and is reported in the Global Biodiversity Information Facility (GBIF; <https://www.gbif.org/es/species/52733229>) for the region of Cadqués where the *E. timida* individual was sampled. Since only *U. compressa* chloroplast sequences were detected, we suggest that *U. compressa* may be a food alga for *E. timida* in the wild. This would contradict previous studies claiming that *A. acetabulum* is the only food source for *E. timida* [118, 119].

Despite their high diversity, molluscs are still very poorly studied in terms of publicly accessible high-quality reference genomes, partly due to the aforementioned difficulties in DNA extraction, library preparation and sequencing [120–122]. Currently available molluscan genome assemblies in the National Center for Biotechnology Information (NCBI) cover only ~0.1% of the described species in the entire phylum. One reason for the often rather fragmented genome assemblies in molluscs may be that the sequencing polymerases are hindered by contaminants, such as the polysaccharide-containing mucus of molluscs, or metabolites bound to the DNA [123, 124]. The sacoglossan genome assemblies published so far are quite fragmented, with contig N50s

of 0.005 to 0.45 Mb [33, 106, 107]. The reason is that in most cases long-read sequencing, but also the chromatin conformation capture library preparation did not work, as so often in molluscs. This resulted in very low sequencing yield. To enable the sequencing of these animals, we have therefore switched to the PacBio ultra-low input protocol, which includes a long-range PCR amplification step to increase the amount of DNA relative to possible contaminants and to obtain ‘artificial but clean’ DNA that can then be easily sequenced. In general, PCR amplification used in the preparation of ultra-low input libraries can lead to bias towards some genomic regions. However, using different PCR polymerases for amplification can counteract this bias and complementary amplify different genomic regions. Combining these data thus leads to improved contiguity of genome assemblies [35]. Bein et al. (2024) [35] included various species from different taxa when investigating the effect of different polymerases on long-range PCR amplification and subsequent assembly results. Other parameters, such as different sequencing machines, were not analysed by Bein et al. (2024) [35]. However, they were able to observe a different performance of KOD polymerase in mammals compared to molluscs and collembolans. Nevertheless, even in molluscs, the KOD polymerase contributes to a better contiguity of the resulting assembly when combined with data generated by PacBio’s polymerase A/B amplification. Although error rates of polymerases used for long range amplification are low (PacBio polymerases A/B: below 1 in  $10^5$ ; KOD polymerase:  $13.1 \times 10^{-6}$ ), PCR errors might be present in the reads or even in the assembly. We cannot investigate the effect of these errors in this study, as no other reference for *E. timida* is known. By using the Arima HiC low-input library preparation protocol, higher cross-linking yields were achieved, ultimately resulting in increased coverage and improved distances between cross-linked genomic loci (compared to the Arima HiC standard library and Dovetail Genomic’s Omni-C library (data not shown)). With a scaffold and contig N50 of 41.8 Mb and 1.92 Mb, respectively, the *E. timida* genome assembly of this study has approximately 30-fold and fourfold better scaffold and contig N50 values than the other sacoglossan genome assemblies.

However, there is a discrepancy of total assembly length (754 Mb) and genome size estimates (FCM: 895 Mb; ModEst: 632 Mb; GenomeScope: 548 Mb). The smaller total length compared to the FCM estimate may be due to collapsed repeats that have not yet been resolved. Although PacBio sequencing was successful, the overall N50 of the HiFi reads (6 kb) may, in some cases, still be too short to resolve some long repetitive regions of the genome. The mapping-based genome size estimate of ModEst is considerably smaller than the

FCM estimate and the total assembly length. ModEst assumes that differences in coverage are due to technical problems in assembly. This may not be entirely correct in our case, as differences in coverage may be caused by bias in PCR amplification. Due to its assumption, the ModEst estimate appears to be less reliable than the FCM estimate with rCV values < 5%. Similarly, genome sizes appear to be consistently underestimated by k-mer-based methods, partly due to repeats [40]. Therefore, a comparison between the total length of the high-complexity regions in the assembly and a k-mer-based estimate of genome size is useful. For example, adding the number of masked bases in the assembly (335 Mb) and the k-mer-based genome size estimate (548 Mb) yields 883 Mb, which is very close to the FCM-based genome size estimate (895 Mb). However, the fact that we performed PCR amplification prior to PacBio sequencing may have resulted in a shorter genome assembly length, and thus the mapping based genome size estimate may underestimate the true genome size, as PCR amplification can introduce errors, mainly homopolymer length changes and dinucleotide repeat compression, which may have led to this underestimation, or parts of the genome may not have been amplified at all. The number of chromosome-level sequences of 15, obtained from the HiC data for *E. timida*, is consistent with the karyotype of the sacoglossan species *Oxynoe olivacea* [125]. In earlier publications, 17 chromosomes were found in other more closely related sacoglossan species [126, 127]. However, the karyotype of *E. timida* was not included and needs to be verified in future studies.

The BUSCO values were similar to other sequenced sacoglossan species. However, the duplicated BUSCOs in the *E. timida* genome assembly were higher (4.9%) than in the other genome assemblies (< 1%). These duplications can be caused by true biological events, replicating loci which contain genes thought to be single copy orthologs. In addition, duplicated BUSCOs can result from high heterozygosity, resulting in the same genomic locus (in a diploid organism) being assembled twice, creating so-called haplotigs. Haplotypic duplications are searched for and collapsed with both hifiasm and purge\_dups (here only at the ends of contigs). The heterozygosity of the *E. timida* genome (0.794%) is estimated to be higher than in other Sacoglossa (0.18%-0.42%; Supplementary Table S6; [128]). However, with only 4.9% duplicated BUSCOs, the proportion is still quite low.

Contamination can lead to major problems when dealing with genome assemblies from public databases [44, 129, 130]. Therefore, contamination screening is a fundamental part of the genome assembly. Regarding contamination, the presented assembly of *E. timida* was screened with two different tools. The advantage of blobtools over

the sequence similarity-based method FCS-GX was that coverage and GC content were also taken into account. In cases where taxa are underrepresented in a database for sequence similarity searches (e.g. Mollusca), false positive and false negative hits occur more frequently. In addition, shorter sequences are less likely to be identified in general. However, taking into account the read coverage and GC content, short sequences can still be identified as contaminants with a high probability (see cluster at the bottom left of Supplemental Figure S5). False-positive hits are still possible due to the taxonomic assignment by blobtools. Nevertheless, we did not filter out sequences that were assigned to Chordata or Arthropoda because the nt database does not contain the necessary diversity of Mollusca sequences to reliably identify this phylum in a de novo assembly. The contigs of the *E. timida* assembly therefore generate hits for the closest related species in the database, likely due to conserved elements of the genome across different phyla (e.g. protein domains).

The structural annotation of the *E. timida* genome assembly shows excellent quality metrics and is the most complete in terms of BUSCOs compared to available annotated genome assemblies of Sacoglossa. In particular, the higher number of CDSs/mRNA and the longer median gene length, while total gene space is comparable, indicate a higher contiguity of annotated genes. This seems to be strongly influenced by the contiguity and accuracy of the underlying genome assembly. Furthermore, over 60% of the protein sequences resulting from the *E. timida* annotation were annotated with at least one GO term, which is 1.4- to 2.8-fold higher compared to the annotation of other Sacoglossa (Table 4). The absolute numbers of protein sequences annotated with a GO term are probably higher for other Sacoglossa due to gene fragmentation. When a gene is split between two contigs or scaffolds, it is annotated as two different genes, but if both parts are large enough to make a reliable match, a GO term (maybe even the same one) is assigned to both gene fragments. The overall functional annotation rate is lower for the other Sacoglossa, probably due to general fragmentation and sequences becoming too short to reliably match against protein sequences of known functions. However, the high number of genes annotated in other Sacoglossa may not be due to fragmentation of the assembly alone. While most of the annotation's statistics are satisfactory, we can only speculate why EtPKS1 was not annotated by BRAKER. A gene was predicted by Augustus at the same locus, which was not taken to the final gene set of BRAKER. There are less RNAseq reads mapping to EtPKS2 (263 reads) compared to EtPKS1 (680), while both genes have a similar size (total CDS length; EtPKS1: 6785 bp; EtPKS2: 6834 bp). Therefore, the amount of RNAseq reads might not be the

reason. To show that many genes could be true or false positive annotations, orthologous clustering could be performed with all available *Sacoglossa* and other high-quality annotations from other molluscs. As many downstream analyses (e.g. comparative, evolutionary) depend on high-quality data as input, the presented annotation will enhance or even enable future studies.

Although it has been claimed that polypropionates are not produced by the animals themselves, but by symbiotic bacteria or dietary organisms [131, 132], there is increasing evidence that animals are capable of producing various compounds themselves [133–137]. In *Sacoglossa*, it is assumed that the produced polypropionate pyrones contribute, among other things, to the establishment and maintenance of the association of *Sacoglossa* and incorporated chloroplasts [34]. The identification of genes encoding PKS1 and PKS2 in the genome annotation of *E. timida* indicates that *E. timida* is a likely producer of polypropionates. Depending on the genome annotation and the origin of the proteins, FAS, PKS1 and PKS2 genes were found in the sacoglossan assemblies. The quality of the genome assemblies and the protein sequences seemed to have a strong influence on the number of gene copies found.

Polypropionates are abundant in molluscs worldwide and have been found in the sacoglossan species *E. chlorotica*, *E. diomedea* and *P. ocellatus*. We have now expanded their presence to *E. timida*. In sacoglossans, polypropionates appear not only to bind fixed carbon from the chloroplast via de novo photosynthesis [110], but also to act via oxidative and photocyclization pathways [110–113]. These reactions may protect sacoglossans from damage caused by photosynthetic reactive oxygen products and may therefore play an important role in life with functional kleptoplasty [111–115]. Interestingly, Torres et al. (2020) [34] found the mRNA of PKSs in the transcriptome of *E. timida* [61, 114]. However, transcriptomes only show the genes that are expressed in a given tissue at a given time, resulting in a subset of all genes present in the genome. We are also aware that a blast search is not the adequate tool to do a gene orthology prediction. However, the aim in this study was getting an overview of the presence and abundance of PKS genes in the assembled and annotated sacoglossan genomes. Therefore, we chose to use the BLAST search for the PKS gene analyses. Future research might provide a deeper insight into and knowledge about the gene orthology of sacoglossan PKS genes. After the polypropionate scaffold is produced, it is decorated by tailoring enzymes. A C-methyltransferase and cytochrome P450 are probably required to produce the large number of polypropionates in *Sacoglossa* [108]. As in other eukaryotes, the genes for the enzymes involved in the production of natural products are not

adjacent to each other. This makes it difficult to identify the corresponding genes for the decorating tailoring enzymes [34]. The genomic environments of the genes encoding PKS1 and PKS2 were searched using BLASTp, mainly yielding uncharacterised and hypothetical proteins with no indication of their catalytic activity. Only highly complete and continuous genome assemblies, such as that of *E. timida*, can provide a comprehensive picture of the genes present.

In the future, the *E. timida* genome assembly may help to shed light not only on polypropionates and their role in the functional kleptoplasty, but also on immune genes. Immune genes are well studied in cnidarians, and have recently been discussed in *Sacoglossa*, as the innate immune system probably plays an important role in the establishment of the process of photosymbiosis (e.g. [138–144]). Although we know more and more about the process of functional kleptoplasty, it is still unknown how especially short-term and long-term sacoglossans correctly identify the chloroplast of their food algae as a symbiont rather than a pathogen nor how chloroplasts are absorbed. Melo Clavijo et al. (2020) [144] found that sacoglossans—including *E. timida*—have a divergent collection of specific scavenger receptors and the thrombospondin-type-1 repeat protein superfamily, comparable to photosymbiotic cnidarians (e.g. [145–151]). Furthermore, they detected species-specific candidate genes that may be important for the symbiont identification in sacoglossans. We investigated the presence of polypropionate encoding genes in *E. timida* and hope that our genome assembly can also serve as a reference genome for immune gene studies in *Sacoglossa*. We expect the genome assembly to contribute to future genetic studies on kleptoplasty and to serve as a high-quality resource for studies on sacoglossans and molluscs in general.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10829-7>.

Supplementary Material 1: Supplemental Figure S1. Climate chamber in which the *E. timida* slugs were kept in artificial sea water in plastic cups as aquariums. The green tubes provided the air supply. Supplemental Table S1. Databases and tools which were used while operating InterProScan version 5.64-96.0 [101]. Supplemental Table S2. Table of PKS and fatty acid synthase (FAS) sequences from Torres et al. (2020) [34] including the animal species they were received from and the accession number. Supplemental Figure S2. Genome size estimation of *E. timida* using flow cytometry. The histogram shows the relative propidium iodide fluorescence intensity obtained after simultaneous analysis of *E. timida* 2C (in green) and the house cricket *A. domesticus* 2C as an internal standard reference (in red). The PI fluorescent dyes were excited with a solid-state laser emitting at 488 nm. The y-axis gives the counts of propidium iodide (PI) stained nuclei. The x-axis displays the relative red PI fluorescence signal. To obtain the mean relative red PI fluorescence signals, the peaks were enclosed by line segments. The percentages in brackets are the portions of all events in the histogram enclosed by the respective line segments.

Supplemental Table S3. Genome size estimates from two individuals of *E. timida*. The measured individual is given in brackets. Chopping buffer was prepared as described by Galbraith et al. (1983) [152]. Propidium iodide was used as a fluorescent dye. We used the house cricket *A. domesticus* as standard reference (genome size: 2000 Mb). Supplemental Figure S3. K-mer profile and estimates based on HiFi reads. Supplemental Table S4. Sacoglossan heterozygosity values. The heterozygosity values from all species except for *E. timida*, were inferred by Theisen & Jensen (1991) [128]. Supplemental Tables S5. PacBio ultra-low library preparation based on PCR amplification with KOD Xtreme™ Hot Start DNA Polymerase (Merck). Supplemental Table S6. Sequencing output and subread mean length of the PacBio low- and ultra-input libraries. Supplemental Figure S4. HiFi read length distribution and statistics. Standard PacBio ultra-low input libraries are listed as SMRT1 and SMRT2. PacBio ultra-low libraries amplified with KOD polymerase are shown as SMRT3 (Sequel IIe) and SMRT4 (Revio). N50 values are presented in bp. Supplemental Table S7. FCS-GX contamination summary. Supplemental Table S8. FCS-GX action summary. Supplemental Figure S5. Blobplot of the assembly after polishing and purging. At this stage of the assembly process, contamination filtering with FCS was already conducted. Supplemental Table S9. Blobtools taxonomic assignment. The table shows all contigs classified as Chlorophyta by “bestsumorder”, which were filtered out among others. Sequences marked with asterisk were categorized as “HICOV” by `purge_dups`. Supplemental Figure S6. Blobplot of the final genome assembly. Supplemental Figure S7: Maximum likelihood phylogenetic tree of FAS, PKS1 and PKS2 transcripts from *E. timida*, *E. chlorotica*, *E. diomedea* and *P. ocellatus*. For the alignment the transcriptomic data from the sequences listed in Table 5 and Supplemental Table S2 were used. The branches are labelled with their length and scaled according to the number of substitutions per site. The percentage of trees in which the associated data clustered together is shown next to the branches. The transcript of EtPKS1 was manually constructed based on sequence homology to EcPKS1, EdPKS1 and PoPKS1 as described previously. The transcripts from *E. timida* are labelled with an asterisk. Supplemental Table S10. Number of blast hits with taxid of *Acetabularia acetabulum* or *Ulva compressa* against contigs of the polished *E. timida* genome assembly. Supplemental Table S11. Number of blast hits for targets with a taxid of *Acetabularia acetabulum* or *Ulva compressa*. All target sequences originate from a chloroplast. Supplemental Table S12. Section of the app file from scaffolding including the 5 Chlorophyta sequences showing all resulting scaffolds containing these Chlorophyta sequences. Except for splitting one of the sequences, none was linked to other nuclear sequences of the *E. timida* genome assembly. Supplemental Figure S8. The genes encoding EtFAS, EtPKS1 and EtPKS2 are annotated in the genome of *E. timida*. The exons are labelled in black on the excerpt of the genomic sequence. The arrows present the transcript of the a) EtFAS, b) EtPKS1 and c) EtPKS2. The domains of the enzymes are presented in bubbles below the arrow. The gene encoding EtPKS1 was annotated manually based on sequence homology with the EcPKS1. The label with an  $x^0$  indicates an inactive domain. Supplemental Figure S9. Isotopic patterns of the putative polypropionates produced by *E. timida* and identified by HPLC-ESI-HRMS analysis. Supplemental Figure S10. HPLC-MS data of *E. timida* extracts. Base peak chromatogram (BPC) and extracted ion chromatograms (EICs) of polypropionates shown in fig. 4. The BPC shows all detected ions present in the crude extract and the EICs show the peaks corresponding to putative polypropionates. Supplemental Table S13. Result of the polypropionate blast search in the annotations of *E. timida*, *E. chlorotica*, *E. diomedea* and *P. ocellatus*. Both unfiltered (F-) as well as filtered (F+) blast hits are shown. Supplemental Figure S11. Blobplot of the assembly after removing sequences identified as contamination and before HiC scaffolding.

### Acknowledgements

The present study is a result of the LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG) and was supported through the program ‘LOEWE-Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz’ of Hesse’s Ministry of Higher Education, Research, and the Arts (HMWK). We thank the Genome Technology Center (RGTC) at Radboudumc for the use of the Sequencing Core Facility (Nijmegen, The Netherlands), which provided the PacBio SMRT sequencing service on the Sequel IIe platform. We would

also like to thank the Bioscientia Institut für med. Diagnostik GmbH, especially Prof. Dr. Hanno Jörn Bolz and Dr. Christian Betz, for providing the PacBio SMRT sequencing service on the PacBio Revio platform. Dr. Bruno Hüttel provided valuable advice and support in establishing the PacBio ultra-low input library protocol in our laboratory at the LOEWE-TBG Centre. We would especially like to thank Dr. Vesa Havurinne and Dr. Sónia Cruz for their support in establishing an *Acetabularia acetabulum* culture in our laboratory in Frankfurt. We also want to thank Dr. Julian Taffner (@terra\_aliens, Instagram @terra\_aliens) for providing the video of *Elysia timida* feeding on its food alga *Acetabularia acetabulum*: <https://youtu.be/MZRep08-81Y>. The funding code for LOEWE-TBG is: LOEWE/1/10/519/03/03.001(0014)/52.

### Authors’ contributions

CGC collected the *Elysia timida* samples. ABH, DB, ChG, and CG performed the DNA extraction and the library preparations. KN assisted with the PacBio SMRT sequencing. LM and TS assembled, annotated and analyzed the genome, and conducted the downstream analyses. JS and EJNH performed the HPLC-MS/MS analysis. LM, JS and EJNH conducted the PKS search/comparative analysis. CG supervised the project. LM, CG and TS wrote the manuscript with input from ABH, CGC, ChG, EJNH and JS. All authors read and approved the final manuscript before submission.

### Funding

Open Access funding enabled and organized by Projekt DEAL. The present study is a result of the LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG) and was supported through the program ‘LOEWE-Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz’ of Hesse’s Ministry of Higher Education, Research, and the Arts (HMWK). The funding code for LOEWE-TBG is: LOEWE/1/10/519/03/03.001(0014)/52.

### Availability of data and materials

Raw data of PacBio HiFi reads (subreads) and raw Arima HiC reads which were used for genome assembly, raw RNAseq reads as well as the final assembly and annotation can be publicly accessed via BioProject PRJNA1119176 and this link: <https://genome.senckenberg.de/download/etim/>.

### Data availability

PacBio HiFi reads and Arima HiC reads which were used for genome assembly, RNAseq reads as well as the final assembly and annotation can be publicly accessed via BioProject PRJNA1119176 and this link: <https://genome.senckenberg.de/download/etim/>.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not Applicable.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>LOEWE Centre for Translational Biodiversity Genomics, Senckenberganlage 25, Frankfurt 60325, Germany. <sup>2</sup>Senckenberg Research Institute, Senckenberganlage 25, Frankfurt 60325, Germany. <sup>3</sup>Institute for Molecular Bio Science, Goethe University Frankfurt, Max-Von-Laue Straße 9, Frankfurt am Main 60438, Germany. <sup>4</sup>Centre d’Estudis Avançats de Blanes (CEAB, CSIC), Accés Cala St. Francesc 14, Blanes, Girona 17300, Spain. <sup>5</sup>Institut de Recerca de La Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain. <sup>6</sup>Department of Human Genetics, Radboud University Medical Centre (Radboudumc), Nijmegen, Netherlands.

Received: 4 June 2024 Accepted: 23 September 2024

Published online: 07 October 2024



## References

- Salzberg SL. Next-generation genome annotation: We still struggle to get it right. *Genome Biol.* 2019;20(1):92. <https://doi.org/10.1186/s13059-019-1715-2>.
- da Fonseca RR, Albrechtsen A, Themudo GE, Ramos-Madrigr J, Sibbesen JA, Marett L, Zepeda-Mendoza ML, Campos PF, Heller R, Pereira RJ. Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Mar Genomics.* 2016;30:3–13. <https://doi.org/10.1016/j.margen.2016.04.012>.
- Schell T, Feldmeyer B, Schmidt H, Greshake B, Tills O, Truebano M, Rundle SD, Paule J, Ebersberger I, Pfenninger M. An Annotated Draft Genome for *Radix auricularia* (Gastropoda, Mollusca). *Genome Biol Evol.* 2017;9(3):585–92. <https://doi.org/10.1093/gbe/evx032>.
- Sigwart JD, Lindberg DR, Chen C, Sun J. Molluscan phylogenomics requires strategically selected genomes. *Philos Trans R Soc B.* 2021;376(1825):20200161. <https://doi.org/10.1098/rstb.2020.0161>.
- Wells SM. Molluscs and the conservation of biodiversity. In van Bruggen AC, Wells SM, Kemperman ThCM. (eds), Biodiversity and Conservation of the Mollusca, Proceedings of the Alan Solem Memorial Symposium on the Biodiversity and Conservation of the Mollusca, Eleventh International Malacological Congress. Siena, Italy, 1992, 21–36. 1995. [https://doi.org/10.14825/kaseki.76.0\\_100](https://doi.org/10.14825/kaseki.76.0_100).
- Groombridge B, Jenkins MD, Jenkins M. World atlas of biodiversity: Earth's living resources in the 21st century. Berkeley: Univ of California Press; 2002.
- Chapman AD. Numbers of Living Species in Australia and the World. 2nd ed. Canberra, Australia: Australian Biological Resources Study; 2011.
- Haszprunar G, Wanninger A. Molluscs. *Curr Biol.* 2012;22(13):R510–4.
- Wanninger A, Wollesen T. The evolution of molluscs. *Biol Rev.* 2019;94(1):102–15. <https://doi.org/10.1111/brv.12439>.
- Gomes-dos-Santos A, Lopes-Lima M, Castro LFC, Froufe E. Molluscan genomics: The road so far and the way forward. *Hydrobiologia.* 2020;847(7):1705–26. <https://doi.org/10.1007/s10750-019-0411-1>.
- Kawaguti S. Electron microscopy on the symbiosis between an elysiid gastropod and chloroplasts of a green alga. *Biology J Okuyama University.* 1965;11:57–65.
- Trench RK, Trench ME, Muscatine L. Symbiotic chloroplasts; their photosynthetic products and contribution to mucus synthesis in two marine slugs. *Biol Bull.* 1972;142(2):335–49. <https://doi.org/10.2307/1540236>.
- Trench RK, Boyle EJ, Smith DC. The association between chloroplasts of *Codium fragile* and the mollusc *Elysia viridis* II. Chloroplast ultrastructure and photosynthetic carbon fixation in *E. viridis*. *Pro Royal Soc London Series Biol Sci.* 1973;184(1074):63–81. <https://doi.org/10.1098/rspb.1973.0031>.
- Hinde R, Smith DC. Persistence of Functional Chloroplasts in *Elysia viridis* (Opisthobranchia, Sacoglossa). *Nat New Biol.* 1972;239:30–1. <https://doi.org/10.1038/newbio239030a0>.
- Hinde R, Smith DC. The role of photosynthesis in the nutrition of the mollusc *Elysia viridis*. *Biol J Lin Soc.* 1975;7(2):161–71. <https://doi.org/10.1111/j.1095-8312.1975.tb00738.x>.
- Christa G, Zimorski V, Woehle C, Tielens AGM, Wägele H, Martin WF, Gould SB. Plastid-bearing sea slugs fix CO<sub>2</sub> in the light but do not require photosynthesis to survive. *Proc Royal Soc Biol.* 2014;281:20132493. <https://doi.org/10.1098/rspb.2013.2493>.
- Cartaxana P, Trampe E, Kühl M, Cruz S. Kleptoplast photosynthesis is nutritionally relevant in the sea slug *Elysia viridis*. *Sci Rep.* 2017;7:7714. <https://doi.org/10.1038/s41598-017-08002-0>.
- Cruz S, LeKieffre C, Cartaxana P, Hubas C, Thiney N, Jakobsen S, Escrig S, Jesus B, Kühl M, Calado R, Meibom A. Functional kleptoplasts intermediate incorporation of carbon and nitrogen in cells of the Sacoglossa sea slug *Elysia viridis*. *Sci Rep.* 2020;10(1):10548. <https://doi.org/10.1038/s41598-020-66909-7>.
- Evertsen J, Burghardt I, Johnsen G, Wägele H. Retention of functional chloroplasts in some sacoglossans from the Indo-Pacific and Mediterranean. *Mar Biol.* 2007;151:2159–66. <https://doi.org/10.1007/s00227-007-0648-6>.
- Händeler K, Grzybowski YP, Krug PJ, Wägele H. Functional chloroplasts in metazoan cells—a unique evolutionary strategy in animal life. *Front Zool.* 2009;6:1–18. <https://doi.org/10.1186/1742-9994-6-28>.
- Wägele H, Raupach MJ, Burghardt I, Grzybowski Y, Händeler K. Solar powered seaslugs (Opisthobranchia, Gastropoda, Mollusca): Incorporation of photosynthetic units: a key character enhancing radiation? *Evolution in Action: Case Studies in Adaptive Radiation, Speciation and the Origin of Biodiversity*, 263–282. 2010. [https://doi.org/10.1007/978-3-642-12425-9\\_13](https://doi.org/10.1007/978-3-642-12425-9_13).
- Christa G, Gould SB, Franken J, Vleugels M, Karmeinski D, Händeler K, Martin WF, Wägele H. Functional kleptoplasty in a limapontioidean genus: Phylogeny, food preferences and photosynthesis in *Costasiella*, with a focus on *C. ocellifera* (Gastropoda: Sacoglossa). *J Molluscan Studies.* 2014;80(5):499–507. <https://doi.org/10.1093/mollus/eyu026>.
- Wägele H, Martin WF. Endosymbioses in sacoglossan seaslugs: Plastid-bearing animals that keep photosynthetic organelles without borrowing genes. In: Löffelhardt, W. (eds) *Endosymbiosis* (pp. 291–324). Springer, Vienna. 2014. [https://doi.org/10.1007/978-3-7091-1303-5\\_14](https://doi.org/10.1007/978-3-7091-1303-5_14).
- Risso A. Memoire sur quelques Gasteropodes nouveaux, Nudibranches et Tectibranches observes dans la mer de Nice (1). *J Phys Chim Hist Nat Arts (Paris).* 1818;87:368–77.
- Bouchet P. Les Elysiidae de Méditerranée (Gastropoda, Opisthobranchiata). *Ann Inst Océanogr Paris (NS).* 1984;60:19–28.
- Thompson TE, Jaklin A. Eastern Mediterranean Opisthobranchia: Elysiidae (Sacoglossa=Ascoglossa). *J Molluscan Studies.* 1988;54:59–69. <https://doi.org/10.1093/mollus/54.1.59>.
- Green BR. Chloroplast genomes of photosynthetic eukaryotes. *Plant J.* 2011;66(1):34–44. <https://doi.org/10.1111/j.1365-313X.2011.04541.x>.
- Pierce SK, Curtis NE, Hanten JJ, Boerner SL, Schwartz JA. Transfer, integration and expression of functional nuclear genes between intercellular species. *Symbiosis.* 2007;43:57–64.
- Rumpho M, Worful JM, Lee J, Kannan K, Tyler MS, Bhattacharya D, Moustafa A, Manhart JR. Horizontal gene transfer of the algal nuclear gene *psbO* to the photosynthetic sea slug *Elysia chlorotica*. *Proc Natl Acad Sci USA.* 2008;105:17867–71. <https://doi.org/10.1073/pnas.0804968105>.
- Schwartz JA, Curtis NE, Pierce SK. Using Algal Transcriptome Sequences to Identify Transferred Genes in the Sea Slug, *Elysia chlorotica*. *Evol Biol.* 2010;37:29–37. <https://doi.org/10.1073/pnas.0804968105>.
- Wägele H, Deusch O, Händeler K, Martin R, Schmitt V, Christa G, Pinzger B, Gould SB, Dagan T, Klusmann-Kolb A, Martin W. Transcriptomic Evidence That Longevity of Acquired Plastids in the Photosynthetic Slugs *Elysia timida* and *Plakobranchea ocellatus* Does Not Entail Lateral Transfer of Algal Nuclear Genes. *Mol Biol Evol.* 2011;28(1):699–706. <https://doi.org/10.1093/molbev/msq239>.
- Chan CX, Vaysberg P, Price DC, Pelletreau KN, Rumpho ME, Bhattacharya D. Active Host Response to Algal Symbionts in the Sea Slug *Elysia chlorotica*. *Mol Biol Evol.* 2018;35(7):1706–11. <https://doi.org/10.1093/molbev/msy061>.
- Maeda T, Takahashi S, Yoshida T, Shimamura S, Takaki Y, Nagai Y, Toyoda A, Suzuki Y, Arimoto A, Ishii H, Satoh N, Nishiyama T, Hasebe M, Maruyama T, Minagawa J, Obokata J, Shigenobu S. Chloroplast acquisition without the gene transfer in kleptoplastic sea slugs, *Plakobranchea ocellatus*. *ELife.* 2021;10:e60176. <https://doi.org/10.7554/eLife.60176>.
- Torres JP, Lin Z, Winter JM, Krug PJ, Schmidt EW. Animal biosynthesis of complex polyketides in a photosynthetic partnership. *Nat Commun.* 2020;11:2882. <https://doi.org/10.1038/s41467-020-16376-5>.
- Bein B, Chrysostomakis I, Arantes L, Brown T, Gerheim C, Schell T, Schneider C, Leushkin E, Chen Z, Sigwart J, Gonzalez V, Wong NLWS, Santos FR, Blom MPK, Mayer F, Mazzoni CJ, Böhne A, Winkler S, Greve C, Hiller M. "Long-read sequencing and genome assembly of natural history collection samples and challenging specimens." *bioRxiv.* 2024:2024–03.
- Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 1980;8(19):4321–6. <https://doi.org/10.1093/nar/8.19.4321>.
- Hare EE, Johnston JS. Chapter 1 of propidium iodide-stained nuclei. *Methods.* 2012;772: 3–12. [https://doi.org/10.1007/978-1-61779-228-1\\_1](https://doi.org/10.1007/978-1-61779-228-1_1)
- Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics.* 2011;27(6):764–70. <https://doi.org/10.1093/bioinformatics/btr011>.
- Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Commun.* 2020;11(1):1432. <https://doi.org/10.1038/s41467-020-14998-3>.

40. Pfenninger M, Schönnenbeck P, Schell T. ModEst: accurate estimation of genome size from next generation sequencing data. *Mol Ecol Res.* 2022;22:1454–64. <https://doi.org/10.1111/1755-0998.13570>.
41. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. *GigaSci.* 2021;10(2):giab008. <https://doi.org/10.1093/gigascience/giab008>.
42. Baid G, Cook DE, Shafin K, Yun T, Llinares-López F, Berthet Q, Belyaeva A, Töpfer A, Wenger AM, Rowell WJ, Yang H, Kolesnikov A, Ammar W, Vert J-P, Vaswani A, McLean CY, Nattestad M, Chang P-C, Carroll A. DeepConsensus improves the accuracy of sequences with a gap-aware sequence transformer. *Nat Biotechnol.* 2023;41:232–8. <https://doi.org/10.1038/s41587-022-01435-7>.
43. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods.* 2021;18:170–5. <https://doi.org/10.1038/s41592-020-01056-5>.
44. Astashyn A, Tvedte ES, Sweeney D, Sapojnikov V, Bouk N, Joukov V, Mozes E, Strophe PK, Sylla PM, Wagner L, Bidwell SL, Brown LC, Clark K, Davis EW, Smith-White B, Hlavina W, Pruitt KD, Schneider VA, Murphy TD. Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biol.* 2024;25:60. <https://doi.org/10.1186/s13059-024-03198-7>.
45. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34:3094–100. <https://doi.org/10.1093/bioinformatics/bty191>.
46. Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics.* 2021;37:4572–4. <https://doi.org/10.1093/bioinformatics/btab705>.
47. “Picard Toolkit” 2019. Broad Institute, GitHub Repository. <https://broadinstitute.github.io/picard/>; Broad Institute.
48. Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, Dijamco J, Nguyen N, Afshar PT, Gross SS, Dorfman L, McLean CY, DePristo MA. A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol.* 2018;36:983–7. <https://doi.org/10.1038/nbt.4235>.
49. Bonfield JK, Marshall J, Danecek P, Li H, Ohan V, Whitwham A, Keane T, Davies RM. HTSLIB: C library for reading/writing high-throughput sequencing data. *GigaSci.* 2021;10(2):giab007. <https://doi.org/10.1093/gigascience/giab007>.
50. Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies. *F1000Res.* 2017;6:1287. <https://doi.org/10.12688/f1000research.12232.1>.
51. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421. <https://doi.org/10.1186/1471-2105-10-421>.
52. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics.* 2016;32(2):292–4. <https://doi.org/10.1093/bioinformatics/btv566>.
53. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033>.
54. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. URL <https://www.R-project.org/>.
55. Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics.* 2023;39(1):btac808. <https://doi.org/10.1093/bioinformatics/btac808>.
56. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at *arXiv*. 2013. <https://doi.org/10.48550/arXiv.1303.3997>.
57. Arnold K, Gosling J, Holmes D. The Java Programming Language. The Java Series. 4th ed. Reading, MA: Addison-Wesley; 2005.
58. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* 2016;3(1):99–101. <https://doi.org/10.1016/j.cels.2015.07.012>.
59. Dudchenko O, Shamim MS, Batra SS, Durand NC, Musial NT, Mostofa R, Pham M, St Hilaire BG, Yao W, Stamenova E, Hoeger M, Nyquist SK, Korchina V, Pletch K, Flanagan JP, Tomaszewicz A, McAloose D, Pérez Estrada C, Novak BJ, Omer AD, Aiden EL. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under \$1000. *BioRxiv*. 2018:254797.
60. Uliano-Silva M, Ferreira JGRN, Krashenninnikova K, Darwin Tree of Life Consortium, Formenti G, Abueg L, Torrance J, Myers EW, Durbin R, Blaxter M, McCarthy SA. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics.* 2023;24:288. <https://doi.org/10.1186/s12859-023-05385-y>.
61. Rauch C, Christa G, de Vries J, Woehle C, Gould SB. Mitochondrial genome assemblies of *Elysia timida* and *Elysia cornigera* and the response of mitochondrion-associated metabolism during starvation. *Genome Biol Evol.* 2017;9(7):1873–9. <https://doi.org/10.1093/gbe/evx129>.
62. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics.* 2018;34(13):i142–50. <https://doi.org/10.1093/bioinformatics/bty266>.
63. Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. BUSCO Update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol.* 2021;38:4647–54. <https://doi.org/10.1093/molbev/msab199>.
64. Rhie A, McCarthy SA, Fedrigo O, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature.* 2021;592:737–46. <https://doi.org/10.1038/s41586-021-03451-0>.
65. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, Mager DL, Feschotte C. Ten things you should know about transposable elements. *Genome Biol.* 2018;19(199). <https://doi.org/10.1186/s13059-018-1577-z>.
66. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci.* 2020;117(17):9451–7. <https://doi.org/10.1073/pnas.1921046117>.
67. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999;27(2):573–80. <https://doi.org/10.1093/nar/27.2.573>.
68. Bao Z, Eddy SR. Automated *de novo* Identification of Repeat Sequence Families in Sequenced Genomes. *Genome Res.* 2002;12:1269–76. <https://doi.org/10.1101/gr.88502>.
69. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics.* 2005;21:i351–8. <https://doi.org/10.1093/bioinformatics/bti1018>.
70. Smit AFA, Hubley R, Green P. (2013–2015). RepeatMasker Open-4.0. <http://www.repeatmasker.org>.
71. Gremme G, Steinbiss S, Kurtz S. GenomeTools: A Comprehensive Software Library for Efficient Processing of Structured Genome Annotations. *IEEE/ACM Trans Comput Biol Bioinf.* 2013;10(3):645–56. <https://doi.org/10.1109/TCBB.2013.68>.
72. Ou S, Jiang N. LTR\_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. *Plant Physiol.* 2018;76(2):1410–22. <https://doi.org/10.1104/pp.17.01310>.
73. Al-Ghalith G, Montassier E, Ward H, Knights D. NINJA-OPS: Fast Accurate Marker Gene Alignment Using Concatenated Ribosomes. *PLoS Comput Biol.* 2016;12(1): e1004658. <https://doi.org/10.1371/journal.pcbi.1004658>.
74. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol.* 2013;30(4):772–80. <https://doi.org/10.1093/molbev/mst010>.
75. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics.* 2010;26(5):680–2. <https://doi.org/10.1093/bioinformatics/btq003>.
76. Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics.* 2006;7(1):62. <https://doi.org/10.1186/1471-2105-7-62>.
77. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics.* 2008;24(5):637–44. <https://doi.org/10.1093/bioinformatics/btn013>.
78. Gotoh O. A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. *Nucleic Acids Res.* 2008;36(8):2630–8. <https://doi.org/10.1093/nar/gkn105>.

79. Iwata H, Gotoh O. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Res.* 2012;40(20):e161–e161. <https://doi.org/10.1093/nar/gks708>.
80. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using Diamond. *Nat Methods.* 2015;12(1):59. <https://doi.org/10.1038/nmeth.3176>.
81. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. Braker1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics.* 2016;32(5):767–9. <https://doi.org/10.1093/bioinformatics/btv661>.
82. Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. In *Gene Prediction* (pp. 65–95). Humana, New York, NY. 2019. [https://doi.org/10.1007/978-1-4939-9173-0\\_5](https://doi.org/10.1007/978-1-4939-9173-0_5).
83. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 2019;20(1):1–13. <https://doi.org/10.1186/s13059-019-1910-1>.
84. Pertea G, Pertea M. GFF utilities: GffRead and GffCompare. *F1000Research.* 9, 304. 2020. <https://doi.org/10.12688/f1000research.23297.2>.
85. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. Braker2: Automatic Eukaryotic Genome Annotation with GeneMark-EP+ and AUGUSTUS Supported by a Protein Database. *NAR Genom Bioinformatics.* 2021;3(1):lqaa108. <https://doi.org/10.1093/nargab/lqaa108>.
86. Brůna T, Li H, Guhlin J, Honsel D, Herbold S, Stanke M, Nenasheva N, Ebel M, Gabriel L, Hoff KJ. Galba: genome annotation with miniprot and AUGUSTUS. *BMC Bioinformatics.* 2023;24:327. <https://doi.org/10.1186/s12859-023-05449-z>.
87. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210–2. <https://doi.org/10.1093/bioinformatics/btv351>.
88. Li H. Protein-to-genome alignment with miniprot. *Bioinformatics.* 2023;39(1):btad014. <https://doi.org/10.1093/bioinformatics/btad014>.
89. Huang N, Li H. compleasm: a faster and more accurate reimplementation of BUSCO. *Bioinformatics.* 2023;39(10):btad595. <https://doi.org/10.1093/bioinformatics/btad595>.
90. Gabriel L, Hoff KJ, Brůna T, Borodovsky M, Stanke M. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics.* 2021;22:1–12. <https://doi.org/10.1186/s12859-021-04482-0>.
91. Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. Braker3: Fully automated genome annotation using RNA-Seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. *Biorxiv.* 2023. <https://doi.org/10.1101/2023.06.10.544449>.
92. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37:907–15. <https://doi.org/10.1038/s41587-019-0201-4>.
93. Knudsen B, Kohn AB, Nahir B, McFadden CS, Moroz LL. Complete DNA sequence of the mitochondrial genome of the sea-slug, *Aplysia californica*: conservation of the gene order in Euthyneura. *Mol Phylogenet Evol.* 2006;38:459–69. <https://doi.org/10.1016/j.ympev.2005.08.017>.
94. Lan Y, Sun J, Chen C, Sun Y, Zhou Y, Yang Y, Zhang W, Li R, Zhou K, Wong WC, Kwan YH, Cheng A, Bougouffa S, Van Dover CL, Qiu J-W, Qian P-Y. Hologenome analysis reveals dual symbiosis in the deep-sea hydrothermal vent snail *Gigantopelta aegis*. *Nature Communication.* 2021;12:1165. <https://doi.org/10.1038/s41467-021-21450-7>.
95. Sato M, Nagashima K. Molecular Characterization of a Mitochondrial DNA Segment from the Japanese Scallop (*Patinopecten yessoensis*): Demonstration of a Region Showing Sequence Polymorphism in the Population. *Mar Biotechnol.* 2001;3:370–9. <https://doi.org/10.1007/s10126001-0015-4>.
96. Wang S, Zhang J, Jiao W, et al. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nature Ecology & Evolution.* 2017;1:0120. <https://doi.org/10.1038/s41559-017-0120-7>.
97. Yokobori S, Fukuda N, Nakamura M, Aoyama T, Oshima T. Long-Term Conservation of Six Duplicated Structural Genes in Cephalopod Mitochondrial Genomes. *Mol Biol Evol.* 2004;21(11):2034–46. <https://doi.org/10.1093/molbev/msh227>.
98. Kenny NJ, McCarthy SA, Dudchenko O, James K, Betteridge E, Corton C, Dolucan J, Mead D, Oliver K, Omer AD, Pelan S, Ryan Y, Sims Y, Skelton J, Smith M, Torrance J, Weisz D, Wipat A, Aiden EL, Howe K, Williams ST. The gene-rich genome of the scallop *Pecten maximus*. *GigaScience.* 2020;9(5):giaa037. <https://doi.org/10.1093/gigascience/giaa037>.
99. Zhou X, Chen Y, Zhu S, Xu H, Liu Y, Chen L. The complete mitochondrial genome of *Pomacea canaliculata* (Gastropoda: Ampullariidae). *Mitochondrial DNA Part A, DNA Mapp Seq Anal.* 2016;27:884–5. <https://doi.org/10.3109/19401736.2014.919488>.
100. Liu C, Zhang Y, Ren Y, Wang H, Li S, Jiang F, Yin L, Qiao X, Zhang G, Qian W, Liu B, Fan W. The genome of the golden apple snail *Pomacea canaliculata* provides insight into stress tolerance and invasive adaptation. *GigaScience.* 2018;7, 9. <http://orcid.org/0000-0001-5036-8733>.
101. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014;30(9):1236–40. <https://doi.org/10.1093/bioinformatics/btu031>.
102. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A. InterPro in 2022. *Nucleic Acids Res.* 2023;51(D1):D418–27. <https://doi.org/10.1093/nar/gkac993>.
103. Tamura K, Stecher G, Kumar S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol Biol Evol.* 2021;38(7):3022–7. <https://doi.org/10.1093/molbev/msab120>.
104. Wang M, Carver JJ, Phelan VV, et al. Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat Biotechnol.* 2016;34:828–37. <https://doi.org/10.1038/nbt.3597>.
105. Shannon P, Markiel A, Ozier O, Baliga NS, Jonathan TW, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–504. <https://doi.org/10.1101/gr.1239303>.
106. Cai H, Li Q, Fang X, Li J, Curtis NE, Altenburger A, Shibata T, Feng M, Maeda T, Schwartz JA, Shigenobu S, Lundholm N, Nishiyama T, Yang H, Hasebe M, Li S, Pierce SK, Wang J. A draft genome assembly of the solar-powered sea slug *Elysia chlorotica*. *Scientific Data.* 2019;6: 190022. <https://doi.org/10.1038/sdata.2019.22>.
107. Eastman KE, Pendleton AL, Shaikh MA, Suttiyut T, Ogas R, Tomko P, Gavelis G, Widhalm JR, Wisecaver JH. A reference genome for the long-term kleptoplast-retaining sea slug *Elysia crispata* morphotype clarki. *G3 Genes|Genomes|Genetics.* 2023;13(12):jkad234. <https://doi.org/10.1093/g3journal/jkad234>.
108. Li F, Lin Z, Krug PJ, Catrow JL, Cox JE, Schmidt EW. Animal FAS-like polyketide synthases produce diverse polypropionates. *Proc Natl Acad Sci U S A.* 2023;120:e2305575120. <https://doi.org/10.1073/pnas.2305575120>.
109. Bouthillette LM, Aniebo V, Colosimo DA, Brumley D, MacMillan JB. Nonenzymatic reactions in natural product formation. *Chem Rev.* 2022;122(18):14815–41.
110. Ireland C, Scheuer PJ. Photosynthetic Marine Mollusks: In vivo <sup>14</sup>C Incorporation into Metabolites of the Sacoglossan *Placobranchus ocellatus*. *Science.* 1979;205:922–3. <https://doi.org/10.1126/science.205.4409.922>.
111. Zuidema DR, Miller AK, Trauner D, Jones PB. Photosensitized Conversion of 9,10-Deoxytridachione to Photodeoxytridachione. *Org Lett.* 2005;7(22):4959–62. <https://doi.org/10.1021/ol051887c>.
112. Zuidema DR, Jones PB. Photochemical Relationships in Sacoglossan Polypropionates. *Org Lett.* 2005;68(4):481–6. <https://doi.org/10.1021/np049607+>.
113. Zuidema DR, Jones PB. Triplet photosensitization in cyercene A and related pyrones. *J Photochem Photobiol B.* 2006;83(2):137–45. <https://doi.org/10.1016/j.jphotobiol.2005.12.016>.
114. de Vries J, Woehle C, Gregor C, Wägele H, Tielsen AGM, Jahns P, Gould SB. Comparison of sister species identifies factors underpinning plastid compatibility in green sea slugs. *Proc R Soc B.* 2015;282:20142519. <https://doi.org/10.1098/rspb.2014.2519>.
115. Powell KJ, Richens JL, Bramble JP, Han L-C, Sharma P, O'Shea P, Moses JE. Photochemical activity of membrane-localised polyketide derived marine natural products. *Tetrahedron.* 2018;74(12):1191–8. <https://doi.org/10.1016/j.tet.2017.10.056>.
116. Marín A, Ros JD. Ultrastructural and ecological aspects of the development of chloroplast retention in the Sacoglossan gastropod *Elysia*

- Timida*. *J Molluscan Studies*. 1993;59(1):95–104. <https://doi.org/10.1093/mollus/59.1.95>.
117. Schmitt V, Händeler K, Gunkel S, Escande M-L, Menzel D, Gould SB, Martin WF, Wägele H. Chloroplast incorporation and long-term photosynthetic performance through the life cycle in laboratory cultures of *Elysia timida* (Sacoglossa, Heterobranchia). *Front Zool*. 2014;11:5. <https://doi.org/10.1186/1742-9994-11-5>.
  118. Marín A, Ros JD. Dynamics of a peculiar plant-herbivore relationship: the photosynthetic ascoglossan *Elysia timida* and the chlorophycean *Acetabularia acetabulum*. *Mar Biol*. 1992;112:677–82. <https://doi.org/10.1007/BF00346186>.
  119. Jesus B, Ventura P, Calado G. Behaviour and a functional xanthophyll cycle enhance photo-regulation mechanisms in the solar-powered sea slug *Elysia timida* (Risso, 1818). *J Exp Mar Biol Ecol*. 2010;395(1–2):98–105. <https://doi.org/10.1016/j.jembe.2010.08.021>.
  120. van Bruggen AC, Wells SM, Kemperman TCM. Biodiversity of the Mollusca: time for a new approach. In: van Bruggen AC, editor. *Biodiversity and Conservation of the Mollusca*. Oegstgeest-Leiden: Buckhuys Publishers; 1995. p. 1–19.
  121. Stork NE. Estimating the number of species on Earth. In: Ponder WF, Lunney D, editors. *The Other 99%: The Conservation and Biodiversity of Invertebrates*. Sydney: Royal Zoological Society of New South Wales; 1999. p. 1–7.
  122. Ponder WF, Lindberg DR. *Phylogeny and evolution of the Mollusca*. Oakland: University of California Press; 2008.
  123. Aoki Y, Koshihara H. Inhibitory effects of acid polysaccharides from sea urchin embryos on RNA polymerase activity. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis*, 1972;272(1):33–43. [https://doi.org/10.1016/0005-2787\(72\)90030-5](https://doi.org/10.1016/0005-2787(72)90030-5).
  124. Sokolov EP. An improved method for DNA isolation from mucopolysaccharide-rich molluscan tissues. *J Molluscan Studies*. 2000;66(4):573–5.
  125. Vitturi R, Gianguzzo P, Colomba M, Jensen KR, Riggio S. Cytogenetics in the sacoglossan *Oxyno olivacea* (Mollusca: Opisthobranchia): karyotype, chromosome banding and fluorescent in situ hybridization. *Mar Biol*. 2020;137:577–82. <https://doi.org/10.1007/s00227000037>.
  126. Mancino G, Sordi M. Conferma del numero cromosomico di *Bosellia mimetica* (Gastropodi, Opisthobranchi). *Atti Soc Toscana Sci Nat Residente Pisa Memorie*. 1965;71:1–12.
  127. Burch JB, Natarajan R. Chromosomes of some opisthobranchiate mollusks from Eniwetok Atoll, western Pacific. *Pacific Sci*. 1967;21(2):252–259. <https://hdl.handle.net/10125/7390>.
  128. Theisen BF, Jensen KR. Genetic variation in six species of Sacoglossan Opisthobranchs. *J Molluscan Studies*. 1991;57(2):267–75. <https://doi.org/10.1093/mollus/57.2.267>.
  129. Steinegger M, Salzberg SL. Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biol*. 2020;21:1–12. <https://doi.org/10.1186/s13059-020-02023-1>.
  130. Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit—interactive quality assessment of genome assemblies. *G3: Genes, Genomes, Genetics*, 2020;10(4):1361–1374. <https://doi.org/10.1534/g3.119.40090>.
  131. Morita M, Schmidt EW. Parallel lives of symbionts and hosts: chemical mutualism in marine animals. *Nat Prod Rep*. 2018;35:357–78. <https://doi.org/10.1039/C7NP00053G>.
  132. Zan J, Li Z, Tianero MD, Davis J, Hill RT, Donia MS. A microbial factory for defensive kahalalides in a tripartite marine symbiosis. *Science*. 2019;364:eaaw6732. <https://doi.org/10.1126/science.aaw6732>.
  133. Castoe TA, Stephens T, Noonan BP, Calestani C. A novel group of type I polyketide synthases (PKS) in animals and the complex phylogenomics of PKSs. *Gene*. 2007;392:47–58. <https://doi.org/10.1016/j.gene.2006.11.005>.
  134. Shou Q, Feng L, Long Y, Han J, Nunnery JK, Powell DH, Butcher RA. A hybrid polyketide–nonribosomal peptide in nematodes that promotes larval survival. *Nat Chem Biol*. 2016;12:770–2. <https://doi.org/10.1038/nchembio.2144>.
  135. Cooke TF, Fischer CR, Wu P, Jiang T-X, Xie KT, Kuo J, Doctorov E, Zehnder A, Khosla C, Chuong C-M, Bustamante CD. Genetic mapping and biochemical basis of yellow feather pigmentation in budgerigars. *Cell*. 2017;171:427–39. <https://doi.org/10.1016/j.cell.2017.08.016>.
  136. Sabatini M, Comba S, Altabe S, Recio-Balsells AI, Labadie GR, Takano E, Gramajo H, Arabolaza A. Biochemical characterization of the minimal domains of an iterative eukaryotic polyketide synthase. *FEBS J*. 2018;285:4494–511. <https://doi.org/10.1111/febs.14675>.
  137. Torres JP, Schmidt EW. The biosynthetic diversity of the animal world. *J Biol Chem*. 2019;294(46):17684–92. <https://doi.org/10.1074/jbc.REV119.006130>.
  138. Koike K, Jimbo M, Sakai R, Kaeriyama M, Muramoto K, Ogata T, Maruyama T, Kamiya H. Octocoral chemical signaling selects and controls dinoflagellate symbionts. *Biol Bull*. 2004;207(2):80–6. <https://doi.org/10.2307/1543582>.
  139. Davy SK, Allemand D, Weis VM. Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiol Mol Biol Rev*. 2012;76(2):229–61. <https://doi.org/10.1128/mmr.05014-11>.
  140. Fransolet D, Roberty S, Plumier JC. Establishment of endosymbiosis: The case of cnidarians and *Symbiodinium*. *J Exp Mar Biol Ecol*. 2012;420:1–7. <https://doi.org/10.1016/j.jembe.2012.03.015>.
  141. Lehnert EM, Mouchka ME, Burriesci MS, Gallo ND, Schwarz JA, Pringle JR. Extensive differences in gene expression between symbiotic and aposymbiotic cnidarians. *G3: Genes, Genomes, Genetics*, 2014;4(2):277–295. <https://doi.org/10.1534/g3.113.009084>.
  142. van der Burg CA, Prentis PJ, Surm JM, Pavasovic A. Insights into the innate immunome of actinarians using a comparative genomic approach. *BMC Genomics*. 2016;17(1):850. <https://doi.org/10.1186/s12864-016-3204-2>.
  143. Melo Clavijo J, Donath A, Seródio J, Christa G. Polymorphic adaptations in metazoans to establish and maintain photosymbioses. *Biol Rev*. 2018;93(4):1006–2020. <https://doi.org/10.1111/brv.12430>.
  144. Melo Clavijo J, Frankenbach S, Fidalgo C, Seródio J, Donath A, Preisfeld A, Christa G. Identification of scavenger receptors and thrombospondin-type-1 repeat proteins potentially relevant for plastid recognition in Sacoglossa. *Ecol Evol*. 2020;10:12348–63. <https://doi.org/10.1002/ece3.6865>.
  145. Miller DJ, Hemmrich G, Ball EE, Hayward DC, Khalturin K, Funayama N, Agata K, Bosch TCG. The innate immune repertoire in Cnidaria - ancestral complexity and stochastic gene loss. *Genome Biol*. 2007;8:R59. <https://doi.org/10.1186/gb-2007-8-4-r59>.
  146. Wood-Charlson EM, Weis VM. The diversity of C-type lectins in the genome of a basal metazoan. *Nematostella vectensis* *Developmental & Comparative Immunology*. 2009;33(8):881–9. <https://doi.org/10.1016/j.dci.2009.01.008>.
  147. Zou J, Chang M, Nie P, Secombes CJ. Origin and evolution of the RIG-I like RNA helicase gene family. 2009. Available from: <https://link.springer.com/article/10.1186/1471-2148-9-85>. Cited 3rd June 2024.
  148. Neubauer EF, Poole AZ, Detournay O, Weis VM, Davy SK. The scavenger receptor repertoire in six cnidarian species and its putative role in cnidarian-dinoflagellate symbiosis. *PeerJ*. 2016;4:e2692. <https://doi.org/10.7717/peerj.2692>.
  149. Brennan JJ, Messerschmidt JL, Williams LM, Matthews BJ, Reynoso M, Gilmore TD. Sea anemone model has a single Toll-like receptor that can function in pathogen detection, NF- $\kappa$ B signal transduction, and development. *PNAS*. 2017;114(47):E10122–31. <https://doi.org/10.1073/pnas.1711530114>.
  150. Brennan JJ, Gilmore TD. Evolutionary Origins of Toll-like Receptor Signaling. *Mol Biol Evol*. 2018;35(7):1576–87. <https://doi.org/10.1093/molbev/msy050>.
  151. Dimos BA, Butler CC, Ricci CA, MacKnight NJ, Mydlarz LD. Responding to Threats Both Foreign and Domestic: NOD-Like Receptors in Corals. *Integr Comp Biol*. 2019;59(4):819–29. <https://doi.org/10.1093/icb/icz111>.
  152. Galbraith DW, Harkins KR, Maddox JM, Ayres NM, Sharma DP, Firoozabady E. Rapid flow cytometric analysis of the cell-cycle in intact plant-tissues. *Science*. 1983;220:1049–51. <https://doi.org/10.1126/science.220.4601.1049>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.