



GENOME NOTE

Two draft genomes of enigmatic Solenogastres (Mollusca, Aplacophora) *Epimenia babai* and *Neomenia megatrapezata*

[version 1; peer review: 1 approved, 1 approved with reservations]

Meghan K. Yap-Chiongco  ¹, Stacy Pirro², Rebecca M. Varney³, Hiroshi Saito⁴, Kenneth M. Halanych⁵, Kevin M. Kocot^{1,6}

¹Department of Biological Sciences, The University of Alabama, Tuscaloosa, Alabama, 35487, USA

²Iridian Genomics, Inc, Bethesda, MD, 20817, USA

³University of California Santa Barbara, Santa Barbara, California, 93106, USA

⁴Department of Zoology, National Museum of Nature and Science, 4-1-1 Amakubo, Tsukuba, Ibaraki, 305-005, Japan

⁵Center for Marine Science, University of North Carolina Wilmington, Wilmington, North Carolina, 28409, USA

⁶Alabama Museum of Natural History, The University of Alabama, Tuscaloosa, Alabama, 35487, USA

V1 First published: 23 Apr 2024, 13:318
<https://doi.org/10.12688/f1000research.148511.1>

Latest published: 23 Apr 2024, 13:318
<https://doi.org/10.12688/f1000research.148511.1>

Abstract

Abstract*

Many molluscan genomes have been published to date, however only three are from representatives of the subphylum Aculifera (Polyplacophora, Caudofoveata, and Solenogastres), the sister taxon to all other molluscs. Currently, genomic resources are completely lacking for Solenogastres. This gap in knowledge hinders comparative and evolutionary studies. Here, we sequenced the genomes of the solenogaster aplacophorans *Epimenia babai* Salvini-Plawen, 1997 and *Neomenia megatrapezata* Salvini-Plawen & Paar-Gausch, 2004 using a hybrid approach combining Oxford Nanopore and Illumina reads. For *E. babai*, we produced a 628 Mbp haploid assembly (N50 = 413 Kbp, L50 = 370) that is rather complete with a BUSCO completeness score of 90.1% (82.0% single, 8.1% duplicated, 6.0% fragmented, and 3.9% missing). For *N. megatrapezata*, we produced a 412 Mbp haploid assembly (N50 = 132 Kbp, L50 = 881) that is also rather complete with a BUSCO completeness score of 85.1% (81.7% single, 3.4% duplicated, 8.1% fragmented, and 6.8% missing). Our annotation pipeline predicted 25,393 gene models for *E. babai* with a BUSCO score of

Open Peer Review

Approval Status

	1	2
version 1	 view	 view
23 Apr 2024		
1. Gloria Arriagada		
Universidad Andres Bello, Santiago, Chile		
2. Alice Dennis		
University of Namur, Namur, Belgium		

Any reports and responses or comments on the article can be found at the end of the article.

92.4% (80.5% single, 11.9% duplicated, 4.9% fragmented, and 2.7% missing) and 22,463 gene models for *N. megatapeza* with a BUSCO score of 90.2% (81.0% single, 9.2% duplicated, 4.7% fragmented, and 5.1% missing). Phylogenomic analysis recovered Solenogastres as the sister taxon to Polyplacophora and Aculifera as the sister taxon to all other sampled molluscs with maximal support. These represent the first whole-genome resources for Solenogastres and will be valuable for future studies investigating this understudied group and molluscan evolution as a whole.

Keywords

Aculifera, Aplacophora, Solenogastres, genome

Corresponding author: Kevin M. Kocot (kmkocot@ua.edu)

Author roles: **Yap-Chiongco MK:** Data Curation, Formal Analysis, Investigation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Pirro S:** Investigation, Writing – Review & Editing; **Varney RM:** Investigation, Writing – Review & Editing; **Saito H:** Resources, Writing – Review & Editing; **Halanych KM:** Resources, Writing – Review & Editing; **Kocot KM:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Resources, Supervision, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by U.S. National Science Foundation funding to K.M.K. (Award DEB-1846174) and K.M.H. (ANT-1043745 and OPP-0338218). Sequencing of the *Neomenia megatapeza* transcriptomes was supported as part of the Best Practices for using Next Generation Sequencing (NGS) Datasets to Determine Robust Evidence of Positive Selection and Convergent Evolution of Polar Organisms workshop sponsored by the U.S. National Science Foundation (OPP-1744877 to Scott Santagata). Funding for Illumina sequencing of *Epimenia babai* was provided by Iridian Genomes, grant #IRGEN_RG_2021-1345 Genomic Studies of Eukaryotic Taxa.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2024 Yap-Chiongco MK *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Yap-Chiongco MK, Pirro S, Varney RM *et al.* **Two draft genomes of enigmatic Solenogastres (Mollusca, Aplacophora) *Epimenia babai* and *Neomenia megatapeza* [version 1; peer review: 1 approved, 1 approved with reservations]** F1000Research 2024, 13:318 <https://doi.org/10.12688/f1000research.148511.1>

First published: 23 Apr 2024, 13:318 <https://doi.org/10.12688/f1000research.148511.1>

Introduction

With their incredible diversity in morphology, life history and ecology, molluscs are of great interest to diverse fields of biology ranging from ecology to evolutionary developmental biology to biomechanics. To date there are high-quality genomes available for six of the eight molluscan classes publicly available, although sampling is heavily biased toward the economically and ecologically important members of the clade Conchifera (e.g. Gastropoda, Bivalvia and Cephalopoda; reviewed by [Gomes-dos-Santos et al. 2020](#) and [Sigwart et al. 2021](#)). As the sister taxon to all other molluscs, the clade Aculifera (Polyplacophora + Aplacophora) is important to our understanding of molluscan evolution and innovation. Despite their importance, only three aculiferan genomes have been published to date, two species of chitons (Polyplacophora; [Varney et al. 2021](#); [Varney and Yap-Chiongco et al. 2022](#)) and one species of Caudofoveata ([Wang et al. 2024](#)). Therefore, we sequenced the genomes of two species of Solenogastres – *Neomenia megatrapezata* [Salvini-Plawen & Paar-Gausch 2004](#) and *Epimenia babai* [Salvini-Plawen, 1997](#) – to facilitate further study of this evolutionarily important taxon.

The two clades of aplacophoran molluscs are united by their distinctive vermiform shape, loss or reduction of the foot, and a chitinous cuticle covered by calcareous integumental spines and scales (sclerites). Solenogastres are distinctive with the presence of a ciliated foot restricted to a ventral groove, an undifferentiated gut, and the lack of true ctenidia, while members of Caudofoveata lack a foot entirely and have the presence of a specialized oral shield used for burrowing in mud ([Todt 2013](#)). The current taxonomy of Solenogastres divides the group into 24 families and four orders (Pholidoskepia, Cavibelonia, Sterrofustia, and Neomeniamorpha) although recent phylogenomic analyses have indicated the need for significant taxonomic revision within this group ([Kocot et al. 2019](#); [Yap-Chiongco et al. 2024](#)).

Whereas most solenogasters are just a few millimeters in length ([Todt 2013](#)) both *N. megatrapezata* and *E. babai* are large enough to obtain high-quality genomic and transcriptomic data from a single individual. *E. babai* is a species of Cavibelonia that is relatively easily collected at SCUBA-accessible depths where it feeds on the soft coral *Scleronephthya gracillima* ([Okusu 2002](#)). Because of its relatively large-body size (up to 20 cm in length) and reliable spawning within the laboratory, *E. babai* remains one of very few species of Solenogastres in which development has been described ([Baba 1938, 1940, 1951](#); [Okusu, 2002](#); [Todt and Wanninger 2010](#)). Sequencing of its genome coupled with characterized development and reliability of spawning make *E. babai* a desirable species to serve as a developmental model within Solenogastres. *N. megatrapezata* (up to 18 cm in length) is a species of Neomeniamorpha found in the Southern Ocean off Antarctica ([Salvini-Plawen & Paar-Gausch 2004](#); [Kocot et al. 2019](#)). *Neomenia* is interesting to study trait evolution within Solenogastres as species within this genus have a complex accessory genital apparatus and lack both ventral foregut glands and a radula ([García-Álvarez & Salvini-Plawen 2007](#)). Together, the sequencing of these two species expands our ability to explore solenogaster evolution and, more broadly, provides an important resource in identifying large-scale patterns in molluscan evolution.

Methods

Specimen collection

Epimenia babai was collected by Akiko Okusu as described in [Okusu \(2002\)](#). The specimen of *Neomenia megatrapezata* (collector number Ap227) used for genome and transcriptome sequencing was collected by Ken Halanych and Kevin Kocot on 3 February 2013 near the Ross Ice Shelf in Antarctica (NBP12-10 Station 22; 76° 59.8965' S, 175° 05.5920' W, 541 m depth) via Blake trawl and is deposited in the Alabama Museum of Natural History (ALMNH) under catalog number ALMNH:Inv:25736.

Additional solenogasters were used for transcriptome sequencing to provide evidence for genome annotation. A second individual of *N. megatrapezata* (Ap259) was collected by Ken Halanych and Kevin Kocot on 28 November 2013 near Recovitza Island, Antarctica (LMG13-12 Station 2; 64° 24.672' S, 61° 57.790' W, 664 m depth) via Blake trawl. A second individual from the same collection event has been deposited in ALMNH under catalog number ALMNH: Inv:25732 as a voucher. A specimen of *Neomenia* aff. *herwigi* (Ap20; same individual used to generate a 454 transcriptome by [Kocot et al. 2011](#); NCBI SRA accession number SRR108985) was collected by Ken Halanych on 6 December 2004 off Argentina (63°23' 03.0" S, 60° 03' 24.0" W, 277 m depth) via Blake trawl. A specimen of *Neomenia permagna* (Ap26) was collected on 15 May 2006 by Ken Halanych off Argentina (LMG06-05; 53°47' S, 60° 42' W; 170 m depth) via Blake trawl and is deposited in ALMNH under catalog number ALMNH:Inv:25733. A specimen of an unidentified species of *Epimenia* (possibly also *E. babai*) was collected on 29 June 2015 by Hiroshi Saito off Osezaki, Japan (55 m depth) by SCUBA diving and a tissue sample is deposited in ALMNH under catalog number ALMNH:Inv:25734.

Extraction, library preparation, and sequencing

To produce short-read genomic data, DNA was extracted using the CTAB-phenol-chloroform method employed by [Varney et al. \(2021\)](#), which was based on [Doyle & Doyle \(1991\)](#). For *E. babai*, DNA was extracted from a piece of midbody tissue and sent to Iridium Genomics for Illumina TruSeq sequencing library preparation and 2 X 100 bp paired-end (PE) sequencing on an Illumina HiSeq X. For *N. megatraperzata*, DNA was extracted from hemolymph and sent to the NY Genome Center for Illumina TruSeq sequencing library preparation and sequencing on an Illumina HiSeq X.

To produce long-read data via Oxford Nanopore sequencing, a library for *N. megatraperzata* was produced from DNA extracted from haemolymph using a CTAB-phenol-chloroform method following [Varney et al. \(2021\)](#). For the three libraries for *E. babai* and the second library for *N. megatraperzata*, genomic DNA was extracted from cryo-preserved tissue using the EZNA Tissue DNA Kit (Omega Bio-tek) followed by cleaning and size selection for high-molecular-weight fragments with Ampure XP beads. Sequencing libraries were prepared with the LSK-109 ligation-based library preparation kit and sequenced in-house using R9.4.1RevD flow cells on a GridION. Reads were base called with [Guppy 4.0](#) and trimmed with [PoreChop \(Wick 2018\)](#) with the --discard_middle flag.

New transcriptome data were generated from the same individual used for WGS for each species and from other individuals of the same or closely related species for genome annotation. For *E. babai*, *Epimenia* sp., *Neomenia* aff. *herwigi*, and *Neomenia permagna*, RNA was extracted as described by [Kocot et al. \(2019\)](#). RNA concentration was measured using a Qubit 3.0 (Thermo Fisher) fluorometer with the RNA High Sensitivity kit, RNA purity was assessed by measuring the 260/280 nm absorbance ratio using a Nanodrop Lite (Thermo Fisher), and RNA integrity was evaluated using a 1% SB agarose gel. Complementary DNA (cDNA) synthesis was performed using 1 ng of total RNA using the Clontech SMART-Seq HT kit. An Illumina sequencing library was then prepared in house using the Nextera XT DNA Library Preparation Kit with 0.15 ng of cDNA. Final library size and concentration were assessed using an Agilent Fragment Analyzer with the NGS 1-6000 bp kit and sent to Psomagen (Cambridge, MA) for sequencing on an Illumina NovaSeq using a S4 flowcell with 2 X 150 bp PE reads. For *N. megatraperzata*, RNA was extracted from the individual used for genome sequencing (Ap227) and a second individual (Ap259) as described in [Kocot et al. \(2019\)](#) and sent to the High-Throughput Genomics Shared Resource of the Huntsman Cancer Institute at the University of Utah for Illumina TruSeq Stranded mRNA Sample Prep Kit with oligo (dT) selection. Sequencing was performed on the Illumina HiSeq 2500 system with 2 X 125 bp PE reads.

Genome assembly and annotation

Genome size and heterozygosity were estimated based on trimmed reads using GenomeScope2 ([Ranallo-Benavidez et al. 2020](#)) with a k-mer of 21. Hybrid genome assembly was performed with [MaSuRCA 3.3.5 \(Zimin et al. 2017\)](#). Recommended settings for eukaryotes with >20X Illumina coverage were used. Genome quality was assessed following assembly (and after each step involving filtering or polishing the genome assembly; see below) with [QUAST 5.0.2 \(Mikheenko et al. 2018\)](#) and completeness with [BUSCO 5.2.2 \(Manni et al. 2021\)](#) using the Metazoa odb_10 dataset and the “--long” flag. We then removed redundant haplotigs with Redundans ([Pryszcz & Gabaldón 2016](#)) using the flags --noscaffolding --norearrangements. Finally, the remaining scaffolds were polished with four rounds of Pilon 1.23 using the Illumina paired-end reads, which were first quality- and adapter-trimmed with [trimmomatic 1.8.0 \(Bolger et al. 2014\)](#) using “ILLUMINACLIP:adapters.fasta:2:30:10 LEADING 10 TRAILING 10 SLIDINGWINDOW:4:15 MINLEN:50” for *N. megatraperzata* and “ILLUMINACLIP:adapters.fasta:3:30:10 LEADING:10 TRAILING:10 SLIDINGWINDOW:4:20 MINLEN:50” for *E. babai*.

For genome annotation, we followed a bioinformatic pipeline that masked repetitive DNA and leveraged both RNA-seq data from that species and protein sequences from diverse aplacophorans as evidence for gene modeling. Repeats in the final assemblies were annotated and softmasked with [RepeatMasker](#) using a custom repeat database generated with [RepeatModeler \(Smit & Hubley 2015\)](#). For RepeatModeler, a maximum genome sample size of 1M and the -LTRStruct option were used. For RepeatMasker, the slow and gcalc options were used. The engine used for both programs was [rmblast](#). Available aplacophoran and select other mollusc proteomes (see data on Figshare for details) were then aligned to the final genome assemblies with [ProtHint 2.6 \(Brñna et al. 2020\)](#) with an e-value cutoff of 1e-25. We ran [TrimGalore \(Krueger et al. 2021\)](#) on the transcriptome reads with the following settings: “-q 30 --illumina --trim-n --length 50.” The trimmed and filtered transcriptome reads were then mapped to the genome using [STAR 2.4.0k \(Dobin et al. 2013\)](#) with “--genomeChrBinNbits 15 --chimSegmentMin 50.” Annotation of protein-coding genes was performed with [BRAKER 2.1.6 \(Brñna et al. 2021\)](#) using the output of ProtHint and STAR with the following settings: “--eptmode --softmasking --crf.”

Phylogenomic analysis

Homologous protein sequences in our newly sequenced solenogaster genomes and the proteomes of 21 other lophotrochozoans, including 16 other molluscs, two annelids, one brachiopod, one phoronid, and one nemertean were used as input for orthology inference using *OrthoFinder 2.4.0* (Emms & Kelly 2019). We then further refined the output of OrthoFinder to construct a matrix of one-to-one orthologs following the pipeline of Varney and Yap-Chiongco *et al.* (2022) except we retained only genes sampled for 18/23 taxa using *PhyloPyPruner* (<https://gitlab.com/fethalen/phyloypypruner>). Maximum likelihood phylogenetic analysis was employed on the concatenated supermatrix of amino acid sequences in IQ-Tree 2.1.3 (Minh *et al.* 2020) using the best-fitting model for each partition (-m MFP) with 1000 rapid bootstraps. The tree was arbitrarily rooted with all non-molluscan taxa. Notably, as the genome of *Chaetoderma* sp. (Wang *et al.* 2024) was released following the completion of this analysis, it was not included.

Results

For *N. megatrapezata*, Illumina genome sequencing yielded 443.02 M reads. Illumina transcriptome sequencing of the same individual of *N. megatrapezata* used for genome sequencing (Ap227) yielded 29.72 M reads and 32.93 M reads for a second individual of *N. megatrapezata* (Ap259). Transcriptomes of other *Neomenia* species resulted in 110.68 M reads for *N. aff. herwigi* and 24.40 M reads for *N. permagna*. GenomeScope analysis estimated the genome size of *N. megatrapezata* to be 294 Mbp and a heterozygosity of 0.142% based on the trimmed PE reads. Following adapter trimming, two flowcells of Oxford Nanopore sequencing yielded 101.31 M reads. Assembly with MaSuRCA resulted in an assembly of 9,870 contigs totaling 347 Mbp (N50 = 119 Kbp, L50 = 1,041) with a BUSCO completeness score of 87.1% (83.9% single, 3.2% duplicated, 8.0% fragmented, and 4.9% missing). After polishing and purging redundant haplotigs, the final *N. megatrapezata* assembly was reduced to 6,168 contigs totaling 412 Mbp (N50 = 132 Kbp, L50 = 881) with a BUSCO completeness score of 85.1% (81.7% single, 3.4% duplicated, 8.1% fragmented, and 6.8% missing). BRAKER predicted 65,328 gene models with 90.6% of BUSCOs detected (80.5% single, 10.1% duplicated, 4.5% fragmented, and 4.9% missing). Removal of gene models not supported by transcriptome or protein evidence resulted in a final gene set for *N. megatrapezata* of 22,463 with a BUSCO completeness score of 90.2% (81.0% single, 9.2% duplicated, 4.7% fragmented, and 5.1% missing).

For *E. babai*, Illumina PE sequencing yielded 693.24 M reads. Illumina transcriptome sequencing of *E. babai* and *Epimenia* sp. yielded 70.06 M and 131.47 M reads, respectively. GenomeScope analysis of PE reads failed to converge for *E. babai* with k-mer sizes of 17 and 21. Therefore, we used the trimmed nanopore data in the program **KMC 3** (Kokot *et al.* 2017) with the options -k21 -fm -t16 -m500 -ci1 -cs10000 to obtain k-mer statistics followed by kmc-tools with the transform option to create a histogram for input into GenomeScope2. Oxford Nanopore sequencing of three flowcells data yielded 14.79M reads post adapter trimming with PoreChop. The resulting analysis based on Nanopore data estimated a genome size for *E. babai* of 543 Mbp and heterozygosity of 1.16%. MaSuRCA yielded an assembly of 11,695 contigs totaling 572 Mbp (N50 = 271 Kbp, L50 = 618) with a BUSCO completeness score of 91.3% (77.7% single, 13.6% duplicated, 5.2% fragmented, and 3.5% missing). The final polished and purged *Epimenia* assembly was reduced to 5,965 contigs totaling 628 Mbp (N50 = 413 Kbp, L50 = 370) with a BUSCO completeness score of 90.1% (82.0% single, 8.1% duplicated, 6.0% fragmented, and 3.9% missing). BRAKER predicted 123,904 gene models with a BUSCO completeness score of 93.3% (80.9% single, 12.4% duplicated, 4.7% fragmented, and 2.0% missing). Removal of genes not supported by transcriptome or protein evidence resulted in a final gene set for *E. babai* of 25,393 with a BUSCO completeness score of 92.4% (80.5% single, 11.9% duplicated, 4.9% fragmented, and 2.7% missing).

Molluscan genomes sequenced to date range in size from 359 Mbp (Simakov *et al.* 2013) to over 6 Gbp (Song *et al.* 2023) and genome size estimates by Adachi *et al.* (2021) for 141 diverse molluscs (Polyplacophora, Gastropoda, Bivalvia, and Cephalopoda) based on flow cytometry ranged from 469.4 Mbp (measured as a C-value of 0.48) to 5.31 Gbp (C-value of 5.43) with a mean size of 2.61 Gbp (C-value of 2.67) among the sampled molluscan taxa. Kocot *et al.* (2016) utilized Feulgen image analysis densitometry (FAID) to estimate genome sizes in select species of aplacophorans, including the same individual of *Epimenia babai* sequenced here. Genome size of *E. babai* was estimated to be 750 Mbp and that of *Neomenia permagna* was estimated to be 293 Mbp (Kocot *et al.* 2016). Genome size for *E. babai* inferred by GenomeScope based on Oxford Nanopore data is smaller than expected based on FIAD at 543 Mbp. However, we were unable to obtain an estimate based on more accurate PE data as the analysis failed to converge. The estimated genome size for *N. megatrapezata* of 294 Mbp inferred by GenomeScope is comparable to the FIAD estimate of 293 Mbp for the closely related species *N. permagna*. With final assembly sizes of 628 Mbp and 412 Mbp for *E. babai* and *N. megatrapezata*, respectively, solenogaster genomes are fairly small with respect to other molluscs. Based on computational and FAID genome size estimates, genome size within Aculifera is quite variable ranging from as small as 293 Mbp in *N. permagna* to 2.84 Gbp in *Cryptochiton stelleri* (Kocot *et al.* 2016; Hinegardner, 1974). Comparing the aplacophoran classes, the estimated genome size of *Chaetoderma* sp. (Caudofoveata) is quite large (2.45 Gbp) in comparison to both *E. babai* and *N. megatrapezata* (Table 1). Although in some groups it has been shown that genome

Table 1. Comparison of genome assembly statistics for available aculiferan genomes.

Species	Class	Size (Gbp)	N50 (Mbp)	BUSCO complete	# Genes	Citation
<i>Epimenia babai</i>	Solenogastres	0.63	0.413	90.1	25,393	This study
<i>Neomenia megatapeza</i>	Solenogastres	0.41	0.132	85.10%	22,463	This study
<i>Chaetoderma</i> sp.	Caudofoveata	2.45	141.46	89.52%	23,675	Wang <i>et al.</i> 2024
<i>Acanthopleura granulata</i>	Polyplacophora	0.61	23.9	97.40%	81,691	Varney <i>et al.</i> 2021
<i>Hanleya hanleyi</i>	Polyplacophora	2.52	0.065	92.00%	69,284	Varney and Yap-Chiongco <i>et al.</i> 2022
<i>Liophura japonica</i>	Polyplacophora	0.61	37.34	96.10%	28,010	Hui <i>et al.</i> 2024

size may co-evolve with life-history traits, the observed high level of variation in genome size within Aculifera raises questions about the underlying mechanisms driving this heterogeneity in genome size (Ritchie *et al.* 2017; Beaudreau *et al.* 2021). Further comparisons between solenogaster, caudofoveate, and chiton genomes will aid in our understanding

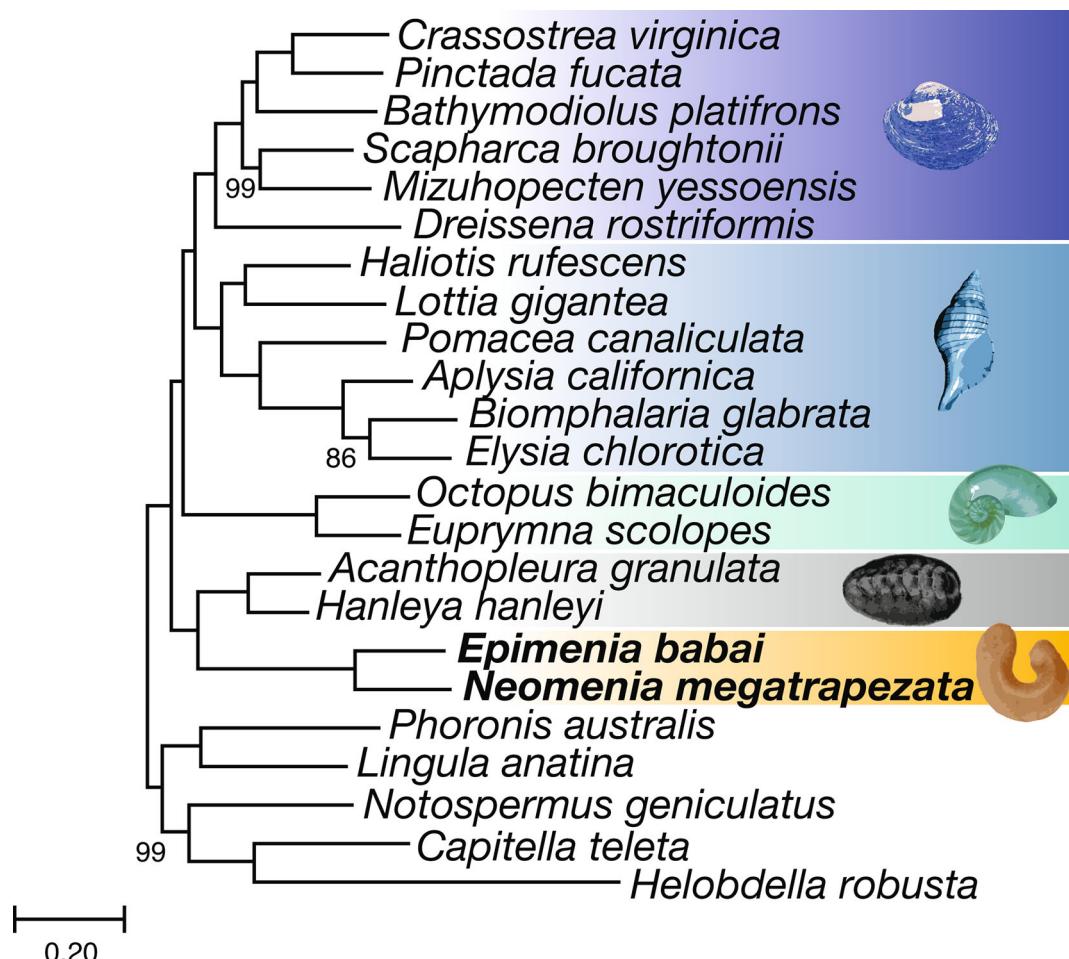


Figure 1. Phylogenetic analysis of Mollusca based on 3,577 nuclear protein-coding genes. Bootstrap support values below 100 are displayed at each node. Image of *Neomenia megatapeza* was altered from a photo taken by Christoph Held.

of the evolutionary history of Aplacophora and Aculifera and the potential interplay between genome size and, e.g., life history.

Comparison of the full set of *Epimenia* and *Neomenia* gene models to the gene models from 21 other lophotrochozoans in OrthoFinder resulted in 251,468 groups of homologous sequences. Our pipeline selected 7,561 single-copy genes sampled for at least 18 of the 23 taxa. The final PhyloPyPruner output resulted in a supermatrix of 3,577 genes with a total of 1,169,375 amino acids and 20.3% missing data. Of these, *E. babai* was sampled for 2,319 and *N. megatrapezata* was sampled for 2,232. The resulting tree (Figure 1) is strongly supported overall with maximal support for Solenogastres and for Aculifera as the sister taxon to all other sampled molluscs, consistent with recent studies (e.g., Kocot *et al.* 2020; Song *et al.* 2023).

In summary, the genomes of *N. megatrapezata* and *E. babai* are relatively complete and comparable to other published aculiferan genomes (Table 1) with BUSCO completeness scores of 85.1% and 90.1%, respectively. These represent the first sequenced genomes for Solenogastres and are a valuable resource for future studies of molluscan evolution.

Ethical considerations

Ethics permits were not required to undertake this research.

Data availability

Underlying data

Associated raw data is available under NCBI Bioproject PRJNA1071799: <https://identifiers.org/bioproject:PRJNA1071799>

NCBI Sequence Read Archive (SRA): RNA-Seq of *Epimenia babai*. Accession number [SRX23536357](https://identifiers.org/sra:SRX23536357)

NCBI SRA: RNA-Seq of *Epimenia* sp. Accession number [SRX23536362](https://identifiers.org/sra:SRX23536362)

NCBI SRA: RNA-Seq of *Neomenia megatrapezata*. Accession numbers [SRX23536358](https://identifiers.org/sra:SRX23536358) and [SRX23536359](https://identifiers.org/sra:SRX23536359)

NCBI SRA: RNA-Seq of *Neomenia* aff. *herwigi*. Accession number [SRX23536361](https://identifiers.org/sra:SRX23536361)

NCBI SRA: RNA-Seq of *Neomenia permagna*. Accession number [SRX23536360](https://identifiers.org/sra:SRX23536360)

NCBI SRA: Illumina Sequencing of *Epimenia babai* gDNA. Accession numbers [SRX5511159](https://identifiers.org/sra:SRX5511159) and [SRX6730477](https://identifiers.org/sra:SRX6730477).

NCBI SRA: Illumina Sequencing of *Neomenia megatrapezata* gDNA. Accession number [SRX23536356](https://identifiers.org/sra:SRX23536356)

NCBI SRA: GridION Sequencing of *Epimenia babai* gDNA. Accession number [SRX23536354](https://identifiers.org/sra:SRX23536354).

NCBI SRA: GridION Sequencing of *Neomenia megatrapezata* gDNA. Accession number [SRX23536355](https://identifiers.org/sra:SRX23536355)

Extended data

Figshare: Additional tables with the list of proteomes used for gene annotation and the full BUSCO reports of gene models are available under the file: YapChiongco_2024_SupplementaryTables.doc. *Neomenia_Nanopore_GenomeFiles.zip* and *Epimenia_Nanopore_GenomeFiles.zip* each contain the content below for the respective species, <https://doi.org/10.6084/m9.figshare.25111997.v1> (Yap-Chiongco, 2024).

- 01_GenomeScope
- 02_MaSuRCA
- 03_Redundans
- 04_Pilon
- 05_GenomeAssembly_BUSCOandQuast

- 06_RepeatModelerandMasker
- 07_BRAKER
- 08_AnnotationBUSCO

Data are available under the terms of the [Creative Commons Attribution 4.0 International license \(CC-BY 4.0\)](#).

Acknowledgements

We are extremely grateful to Akiko Okusu for kindly sharing the specimen of *Epimenia babai* used for genome and transcriptome sequencing. We thank Hiroshi Takashige who assisted in sampling of *Epimenia* sp. We thank the crew and scientists of the Icy Inverts cruises aboard the RV *Lawrence M. Gould* and RVIB *Nathaniel B. Palmer* for help in collecting Antarctic solenogasters. We also thank Scott Santagata for facilitating sequencing of the *Neomenia megatrapezata* transcriptomes.

The authors gratefully acknowledge use of the resources of the Alabama Water Institute at The University of Alabama.

References

Adachi K, Yoshizumi A, Kuramochi T, et al.: **Novel insights into the evolution of genome size and AT content in mollusks.** *Mar. Biol.* 2021; **168**: 1–7. [Publisher Full Text](#)

Baba K: **The later development of a solenogastre, *Epimenia verrucosa*.** *Journal of the Department of Agriculture, Kyusyu Imperial University.* 1938; **6**: 21–40. [Publisher Full Text](#)

Baba K: **The early development of a Solenogastre, *Epimenia verrucosa* (Nierstrasz).** *Annot. Zool. Jpn.* 1940; **19**: 107–113.

Baba K: **General sketch of the development in a Solenogastre, *Epimenia verrucosa* (Nierstrasz).** *Miscellaneous Reports of the Research Institute for Natural Resources.* 1951; **19–21**: 38–46.

Beaudreau N, Massamba-N'Siala G, Belzile C, et al.: **Life-history traits display strong associations to genome size in annelids.** *Hydrobiologia.* 2021; **848**: 799–810. [Publisher Full Text](#)

Bolger AM, Lohse M, Usadel B: **Trimmomatic: A flexible trimmer for Illumina sequence data.** *Bioinformatics.* 2014; **30**(15): 2114–2120. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Brúna T, Hoff KJ, Lomsadze A, et al.: **BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database.** *NAR Genom. Bioinform.* 2021; **3**(1): lqaa108. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Brúna T, Lomsadze A, Borodovsky M: **GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins.** *NAR Genom. Bioinform.* 2020; **2**(2): lqaa026. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Dobin A, Davis CA, Schlesinger F, et al.: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics.* 2013; **29**(1): 15–21. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Doyle J, Doyle JL: **CTAB total DNA isolation.** *Molecular Techniques in Taxonomy.* 1991; 283–293. [Publisher Full Text](#)

Emms DM, Kelly S: **OrthoFinder: phylogenetic orthology inference for comparative genomics.** *Genome Biol.* 2019; **20**(1): 214–238. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

García-Álvarez O, Salvini-Plawen LV: **Species and diagnosis of the families and genera of Solenogastres (Mollusca).** *Iberus.* 2007; **25**(2): 73–143.

Gomes-dos-Santos A, Lopes-Lima M, Castro LFC, et al.: **Molluscan genomics: the road so far and the way forward.** *Hydrobiologia.* 2020; **847**(7): 1705–1726. [Publisher Full Text](#)

Hinegardner R: **Cellular DNA content of the Mollusca.** *Comp. Biochem. Physiol. A Physiol.* 1974; **47**(2): 447–460. [Publisher Full Text](#)

Hui JH, Chan TF, Chan L, et al.: **Chromosome-level genome assembly of the common chiton, *Liophura japonica* (Lischke, 1873).** *bioRxiv.* 2024; 2024-01. [Publisher Full Text](#)

Kocot KM, Cannon JT, Todt C, et al.: **Phylogenomics reveals deep molluscan relationships.** *Nature.* 2011; **477**(7365): 452–456. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [Reference Source](#)

Kocot KM, Jeffery NW, Mulligan K, et al.: **Genome size estimates for Aplacophora, Polypolacophora and Scaphopoda: small solenogasters and sizeable scaphopods.** *J. Molluscan Stud.* 2016; **82**(1): 216–219. [Publisher Full Text](#)

Kocot KM, Todt C, Mikkelsen NT, et al.: **Phylogenomics of Aplacophora (Mollusca, Aculifera) and a solenogaster without a foot.** *Proc. R. Soc. B.* 2019; **286**(1902): 20190115. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Kocot KM, Poustka AJ, Stöger I, et al.: **New data from Monoplacophora and a carefully-curated dataset resolve molluscan relationships.** *Sci. Rep.* 2020; **10**(1): 101. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Kokot M, Dlugosz M, Deorowicz S: **KMC 3: counting and manipulating k-mer statistics.** *Bioinformatics.* 2017; **33**(17): 2759–2761. [PubMed Abstract](#) | [Publisher Full Text](#)

Krueger F, James F, Ewels P, et al.: **TrimGalore: v0. 6.7.** *Zenodo.* 2021. [Publisher Full Text](#) | [Reference Source](#)

Manni M, Berkeley MR, Seppey M, et al.: **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**(10): 4647–4654. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Mikheenko A, Prijibelski A, Saveliev V, et al.: **Versatile genome assembly evaluation with QUAST-LG.** *Bioinformatics.* 2018; **34**(13): i142–i150. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Minh BQ, Schmidt HA, Chernomor O, et al.: **IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era.** *Mol. Biol. Evol.* 2020; **37**(5): 1530–1534. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Okusu A: **Embryogenesis and development of *Epimenia babai* (Mollusca Neomeniomorpha).** *Biol. Bull.* 2002; **203**(1): 87–103. [PubMed Abstract](#) | [Publisher Full Text](#)

Pryszcz LP, Gabaldón T: **Redundans: An assembly pipeline for highly heterozygous genomes.** *Nucleic Acids Res.* 2016; **44**(12): e113–e113. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Ranallo-Beñavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat. Commun.* 2020; **11**(1): 1432. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Ritchie H, Jamieson AJ, Piernley SB: **Genome size variation in deep-sea amphipods.** *R. Soc. Open Sci.* 2017; 4(9): 170862.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Salvini-Plawen L: **Systematic revision of the Epimeniidae (Mollusca: Solenogastres).** *J. Molluscan Stud.* 1997; 63(2): 131-55.
[Reference Source](#)

Salvini-Plawen LV, Paar-Gausch I: **Three new species of *Neomenia* (Mollusca: Solenogastres) from the southern hemisphere.** *N. Z. J. Mar. Freshw. Res.* 2004; 38(1): 137-162.
[Publisher Full Text](#)

Sigwart JD, Lindberg DR, Chen C, et al.: **Molluscan phylogenomics requires strategically selected genomes.** *Philos. Trans. R. Soc. B.* 2021; 376(1825): 20200161.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Simakov O, Marletaz F, Cho SJ, et al.: **Insights into bilaterian evolution from three spiralian genomes.** *Nature.* 2013; 493(7433): 526-531.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Smit AFA, Hubley R: *RepeatModeler Open-1.0*. 2008-2015. Accessed December 29, 2020.
[Reference Source](#)

Song H, Wang Y, Shao H, et al.: **Scaphopoda is the sister taxon to Bivalvia: Evidence of ancient incomplete lineage sorting.** *Proc. Natl. Acad. Sci.* 2023; 120(40): e2302361120.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Todt C, Wanninger A: **Of tests, trochs, shells, and spicules: Development of the basal mollusk *Wirenia argentea* (Solenogastres) and its bearing on the evolution of trochozoan larval key features.** *Front. Zool.* 2010; 7(1): 6-17.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Todt C: **Aplacophoran mollusks—still obscure and difficult?** *Am. Malacol. Bull.* 2013; 31(1): 181-187.
[Publisher Full Text](#)

Varney RM, Speiser DI, McDougall C, et al.: **The iron-responsive genome of the chiton *Acanthopleura granulata*.** *Genome Biol. Evol.* 2021; 13(1).
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Varney RM, Yap-Chiongco MK, Mikkelsen NT, et al.: **Genome of the lepidopleurid chiton *Hanleya hanleyi* (Mollusca, Polyplacophora).** *F1000Res.* 2022; 11: 555.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Wang Y, Wang M, Li J, et al.: **A chromosome-level genome assembly of a deep-sea symbiotic Aplacophora mollusc *Chaetoderma* sp.** *Sci. Data.* 2024; 11(1): 133.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Wick RR: *Porechop* (Version 0.2.1) [Computer software]. 2018.
[Reference Source](#)

Yap-Chiongco MK, Bergmeier FS, Roberts NG, et al.: **Phylogenomic reconstruction of Solenogastres (Mollusca, Aplacophora) informs hypotheses on body size evolution.** *Mol. Phylogenet. Evol.* 2024; 194(108029): 108029.
[PubMed Abstract](#) | [Publisher Full Text](#)

Yap-Chiongco M: *Neomenia megatrapezata* and *Epimenia babai* Extended Genome Data. Dataset. *figshare.* 2024.
[Publisher Full Text](#)

Zimin AV, Puiu D, Luo M-C, et al.: **Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm.** *Genome Res.* 2017; 27(5): 787-792.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 30 May 2024

<https://doi.org/10.5256/f1000research.162830.r273592>

© 2024 Dennis A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Alice Dennis

University of Namur, Namur, Belgium

The manuscript is very clear and well written. It presents assemblies of two new molluscan genomes, from the highly neglected clade *Aculifera*. These genomes are the first to be sequenced in the class *Solenogastres*. These two genomes are of good quality and are well constructed, using hybrid sequencing methods and assembly. The genomes have been annotated using appropriate models and same-species evidence. I found it interesting to see the comparison of genome size estimates based on Feulgen and kmer methods, next to the final assemblies. After annotation, single copy genes were extracted from both assemblies and well as a set of representative lophotrochozoans species. These were used to construct a phylogeny. This demonstrates the value of this genomic resource for future work in taxonomy and beyond.

I think this work is suitable for indexing as-is. The only place I might ask for a bit more clarification is in the annotation curation. I am curious as to why there might be so many gene models for *Epimenia* (123,904). You have reduced this to a more reasonable 25,393 by only retaining genes with direct evidence. This is a good approach, but I worry that a few valid things have been excluded, and that this inflated number of genes perhaps reflects some important aspect of the genome (for example some sort of novel repetitive element).

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a usable and accessible format, and the assembly and

annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Evolutionary genetics, adaptive physiology, mollusc genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 22 May 2024

<https://doi.org/10.5256/f1000research.162830.r273593>

© 2024 Arriagada G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Gloria Arriagada

Universidad Andres Bello, Santiago, Santiago Metropolitan Region, Chile

The article entitled "Two draft genomes of enigmatic Solenogastres (Mollusca, Aplacophora) Epidemia babai and Neomenia megatapeza" described the genome sequencing of two mollusk form an underrepresented subphylum among the sequenced mollusks. They clearly explain the methodology used for the sequencing and assembling as well as its annotation. Although they provide access to all the information generated, the results are very succinct and only present the statistics and numbers, and do not explain or discuss anything about the annotated genes, as a reader and not been an expert in sequencing I missed that. It was not clear to me how many of the annotated genes are indeed transcribed, since they use transcriptomics also, how many are protein coding genes and where in the physiological or cellular processes are they most represented. Are there repetitive or mobile elements in these genomes? Which fraction of the total genome are they? Are they expressed? I believe a bit of more functional insight of this very relevant genomic information, will make this a more amicable paper for readers interested in both mollusk evolution and functional genomics.

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a usable and accessible format, and the assembly and

annotation available in an appropriate subject-specific repository?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: virology, host-pathogen interaction, mobile elements, mollusk LTR-retrotransposons

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research