GENOME NOTE

Genome of the lepidopleurid chiton *Hanleya hanleyi* (Mollusca, Polyplacophora) [version 1; peer review: 2 approved]

Rebecca M. Varney[1*], Meghan K. Yap-Chiongco[2*], Nina T. Mikkelsen[3], Kevin M. Kocot[2,4]

[1]Ecology, Evolution and Marine Biology, University of California, Santa Barbara, Santa Barbara, CA, 93106, USA
[2]Department of Biological Sciences, The University of Alabama, Tuscaloosa, Alabama, 35487, USA
[3]University Museum of Bergen, Univeristy of Bergen, Bergen, 5020, Norway
[4]Alabama Museum of Natural History, The University of Alabama, Tuscaloosa, AL, 35487, USA

* Equal contributors

v1 **First published:** 23 May 2022, **11**:555
https://doi.org/10.12688/f1000research.121706.1
Latest published: 23 May 2022, **11**:555
https://doi.org/10.12688/f1000research.121706.1

Abstract

Mollusca is the second most species-rich phylum and includes animals as disparate as octopuses, clams, and chitons. Dozens of molluscan genomes are available, but only one representative of the subphylum Aculifera, the sister taxon to all other molluscs, has been sequenced to date, hindering comparative and evolutionary studies. To facilitate evolutionary studies across Mollusca, we sequenced the genome of a second aculiferan mollusc, the lepidopleurid chiton *Hanleya hanleyi* (Bean 1844), using a hybrid approach combining Oxford Nanopore and Illumina reads. After purging redundant haplotigs and removing contamination from this 1.3% heterozygous genome, we produced a 2.5 Gbp haploid assembly (>4X the size of the other chiton genome sequenced to date) with an N50 of 65.0 Kbp. Despite a fragmented assembly, the genome is rather complete (92.0% of BUSCOs detected; 79.4% complete plus 12.6% fragmented). Remarkably, the genome has the highest repeat content of any molluscan genome reported to date (>66%). Our gene annotation pipeline predicted 69,284 gene models (92.9% of BUSCOs detected; 81.8% complete plus 11.1% fragmented) of which 35,362 were supported by transcriptome and/or protein evidence. Phylogenomic analysis recovered Polyplacophora sister to all other sampled molluscs with maximal support. The *Hanleya* genome will be a valuable resource for studies of molluscan biology with diverse potential applications ranging from evolutionary and comparative genomics to molecular ecology.

Keywords

Aculifera, Lepidopleurida, genome, repetitive DNA

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 23 May 2022	✓ view	✓ view

1. **Vanessa L. Gonzalez**, Smithsonian Institution, Washington, USA

2. **Samuel Abalde**, Swedish Museum of Natural History, Stockholm, Sweden

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

This article is included in the Genomics and Genetics gateway.

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

Author roles: Varney RM: Data Curation, Formal Analysis, Investigation, Methodology, Resources, Writing – Review & Editing; **Yap-Chiongco MK**: Data Curation, Formal Analysis, Investigation, Methodology, Resources, Writing – Review & Editing; **Mikkelsen NT**: Funding Acquisition, Investigation, Resources, Writing – Review & Editing; **Kocot KM**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the United States National Science Foundation to K.M.K. (NSF DEB 1846174) and The Norwegian Biodiversity Information Centre to N.T.M. (26-19).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Varney RM, Yap-Chiongco MK, Mikkelsen NT and Kocot KM. **Genome of the lepidopleurid chiton *Hanleya hanleyi* (Mollusca, Polyplacophora) [version 1; peer review: 2 approved]** F1000Research 2022, **11**:555
https://doi.org/10.12688/f1000research.121706.1

First published: 23 May 2022, **11**:555 https://doi.org/10.12688/f1000research.121706.1

Introduction

Mollusca is the second most diverse animal phylum and includes many economically and ecologically important species. Molluscs have been the focus of significant genomic research in recent years, which has enabled exciting comparative and evolutionary genomic investigations (reviewed by Gomes-dos-Santos *et al.* 2020). However, although dozens of molluscan genomes have been sequenced to date, all but one belong to the subphylum Conchifera, which includes the familiar gastropods, bivalves, and cephalopods. The subphylum Aculifera, which includes the eight-shelled chitons (Polyplacophora) and worm-like aplacophorans (Solenogastres and Caudofoveata), is the sister taxon to all other molluscs (Kocot *et al.* 2020). Surprisingly, just one species from this clade, the chiton *Acanthopleura granulata* (Gmelin, 1791), has been sequenced to date (Varney *et al.* 2021). Aculiferans are of great interest because as the sister taxon of all other molluscs they are important to understanding molluscan evolution. Further, species in this clade exhibit interesting traits such as iron-hardened teeth (Brooker and Shaw 2012), a complex armature of scales and spines (García-Álvarez & Salvini-Plawen 2007), and the only eyes in a living animal with a mineralized lens (Speiser *et al.* 2011).

Here, we expanded available genomic resources for Aculifera by sequencing and annotating the genome of the chiton *Hanleya hanleyi* (Bean 1844). Extant chitons can be divided into two major clades: Chitonida, the clade represented by the previously published *Acanthopleura* genome, and Lepidopleurida, which includes *Hanleya.* Lepidopleurida is interesting from an evolutionary standpoint as these chitons are thought to be plesiomorphic, with shell features like those of ancient fossil chitons, gills restricted to the posterior region of the body, and simple gamete structure (Sigwart *et al.* 2011). Because of this suite of putatively ancestral characteristics and its phylogenetic position as the sister taxon to all other chitons, Lepidopleurida is thought to be critical to understanding large-scale patterns in molluscan evolution (Sigwart 2008). *Hanleya hanleyi* is a widely distributed, sponge-feeding lepidopleurid that is relatively common off Bergen, Norway and it is the largest lepididopleurid chiton known (Sirenko *et al.* 2016), making it an excellent choice for genome sequencing.

Methods

The specimen of *Hanleya hanleyi* used for genome sequencing (Figure 1A) was collected by N.T.M. off Bergen, Norway in 2018 and is deposited in the University Museum of Bergen under catalog number ZMBN 146951. The genome was sequenced with a combination of short and long reads. To produce short-read data, genomic DNA was extracted from 96% ethanol-preserved samples of foot tissue using a CTAB-phenol-chloroform method following Varney *et al.* (2021). A sequencing library was prepared in-house using the Illumina TruSeq DNA PCR-Free kit with dual indexing according to the manufacturer's instructions. This library was sequenced by Macrogen USA on one lane of the Illumina HiSeq X instrument with 150 bp paired-end (PE) sequencing. To produce long-read data via Oxford Nanopore sequencing, genomic DNA was extracted with an EZNA Tissue DNA Kit (Omega Bio-tek) and cleaned and enriched for high-molecular-weight fragments with the Short-Read Eliminator kit (Circulomics) according to the manufacturer's instructions. Three sequencing libraries were prepared with the LSK-109 ligation-based library preparation kit and sequenced in-house on three 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.

A different specimen of *Hanleya hanleyi* collected by dredging near Bergen, Norway in summer 2008 was gifted to the authors by Dr. Hans Torre Rapp for transcriptome sequencing and is deposited in the Alabama Museum of Natural History under catalog number ALMNH:Inv:23399. Notably, tissue from this same individual was used to generate the 454 pyrosequencing-based foot tissue transcriptome for this species (SRR108987) published by Kocot *et al.* (2011). For Illumina transcriptome sequencing, RNA extraction was performed on mantle tissue preserved in RNAlater and stored at -80°C using the Omega Bio-tek EZNA Mollusc RNA Extraction Kit with an on-column DNAse digestion. 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. RNA was sent to Psomagen (Cambridge, MA, USA) for Illumina TruSeq RNA v2 library preparation (polyA enrichment) and sequencing on the Illumina HiSeq 2500 system with 100 bp PE sequencing.

Genome size and heterozygosity were estimated based on the PE Illumina reads using GenomeScope 2 (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), which consolidates PE data into super reads and then uses long-read data to scaffold and gap-fill. Recommended settings for eukaryotes with >20X Illumina coverage and "PE= pe 587 88" were used. At this point (and after each step involving filtering or polishing the genome assembly; see below), we assessed assembly quality with QUAST 5.0.2 (Mikheenko *et al.* 2018) and completeness with BUSCO 5.1.3 (Manni *et al.* 2021) using the Metazoa odb_10 dataset and the "--long" flag. We then removed redundant haplotigs with purge_dups. Finally, the remaining scaffolds were polished with POLCA (Zimin & Salzberg 2020) using the Illumina paired-end reads, which were first quality- and

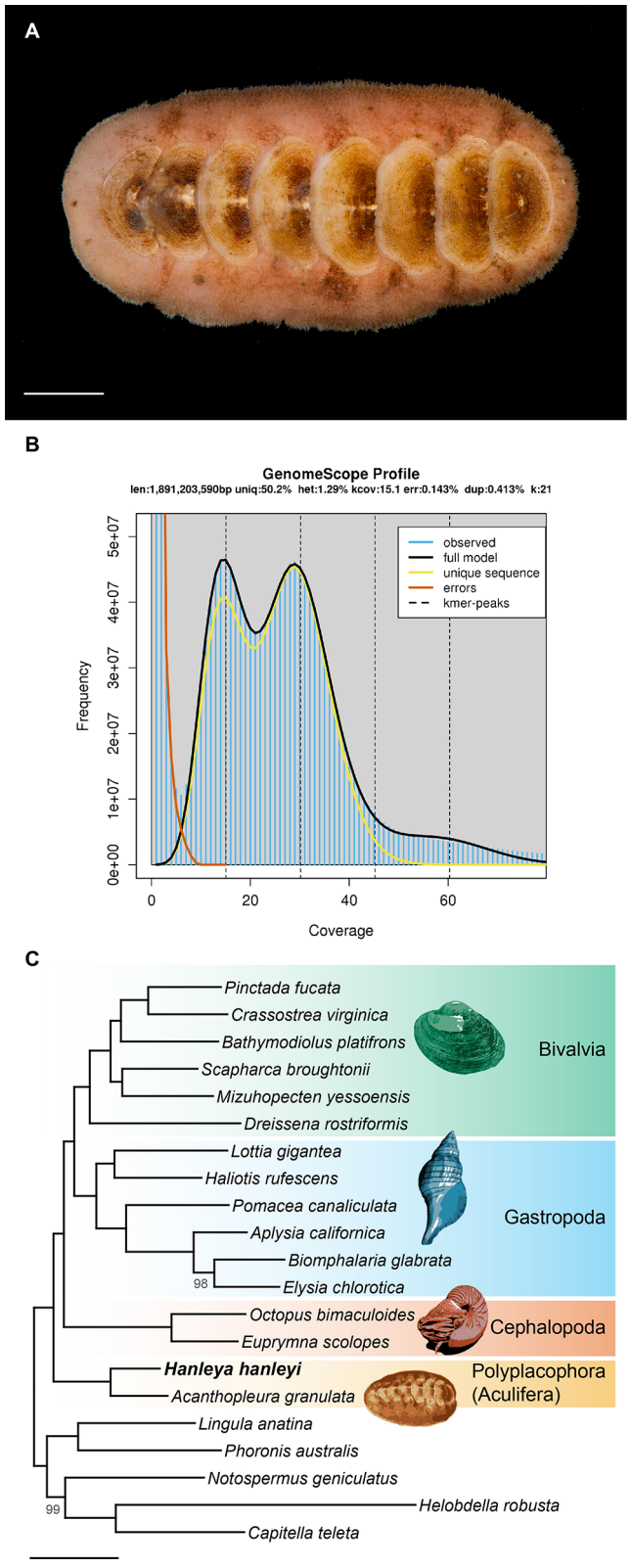

Figure 1. A. Specimen of *Hanleya hanleyi* used for genome sequencing (ZMBN 146951). Scale bar = 8 mm. B. GenomeScope analysis of the paired-end Illumina data. The presence of two peaks indicates that Hanleya has a diploid genome, as expected. Heterozygosity is measured via k-mer distribution (presented at top of graph as "het"). C. Phylogenetic analysis of 2,331 nuclear protein-coding genes. Bootstrap support values below 100 are displayed at each node. Scale bar = 0.2 substitutions per site.

adapter-trimmed with trimmomatic 1.8.0 (Bolger *et al.* 2014) using the following settings: "ILUMINACLIP:adapters.fasta:2:30:10 LEADING 10 TRAILING 10 SLIDINGWINDOW:4:15 MINLEN:50."

Contamination was then screened for and removed with BlobTools2 (Challis *et al.* 2020). The POLCA-polished assembly was searched against the Uniprot reference proteomes (02-Jun-2021 release) with Diamond 2.0.14 (Buchfink *et al.* 2015) using the following settings: "--sensitive --index-chunks 1 --block-size 10 --max-target-seqs 1 -evalue 1e-25 --outfmt 6." The quality- and adapter-trimmed genomic PE reads were then mapped to the genome with minimap 2.23 (Li 2018) with the following settings: "-ax sr." The output of these tools as well as full_table.tsv generated by BUSCO were then used as input files to run BlobTools2. We removed scaffolds with fewer than 10 mapping Illumina reads, scaffolds not annotated as Metazoa, and scaffolds with a GC content <0.30 or >0.55, which appeared as clear outliers when GC content was plotted against coverage.

For genome annotation, repeats in the final contamination-filtered assembly 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 gccalc options were used. The engine used for both programs was rmblast. Available chiton and select other mollusc proteomes (see data on Dryad for details) were then mapped to the final genome assembly with ProtHint 2.6 (Brůna *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." 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 (Bruna *et al.* 2021) using the output of ProtHint and STAR with the following settings: "--eptmode --softmasking --crf." Predicted transcripts with at least partial support from the *Hanleya* transcriptome and/or other chiton proteomes were identified with the selectSupportedSubsets.py bundled with BRAKER.

Building on the phylogenomic analysis of Varney *et al.* (2021), we identified homologous protein sequences in the full set of *Hanleya hanleyi* gene models (including those with no transcript or protein evidence) to the complete proteome of the only other available chiton genome, *Acanthopleura granulata*, and the proteomes of 19 other lophotrochozoans, including 14 other molluscs, 2 annelids, 1 brachiopod, 1 phoronid, and 1 nemertean using OrthoFinder 2.4.0 (Emms & Kelly 2019). We then identified orthologous genes from the homogroups produced by OrthoFinder using the pipeline of Varney *et al.* (2021) except we retained only genes sampled for 18/21 taxa using PhyloPyPruner. Phylogenetic analysis on the concatenated supermatrix in IQ-Tree 2.1.3 (Minh *et al.* 2020) using the best-fitting model for each partition (-m MFP). The tree was arbitrarily rooted with all non-molluscan taxa.

Results

Illumina transcriptome sequencing yielded 25.8M reads or 5.8 Gbp. For the genome, Oxford Nanopore sequencing of three flowcells yielded 13.30, 12.47, and 13.91 Gbp (4,401,106, 4,551,630, and 7,027,597 reads respectively) and Illumina sequencing yielded 129 Gbp (860,037,886 reads). GenomeScope analysis of the PE genomic data inferred a genome size of 1.89 Gbp and a heterozygosity of 1.3% (Figure 1B). Assembly with MaSuRCA yielded an initial assembly consisting of 81,742 scaffolds totaling 3.11 Gbp with an N50 of 59.9 Kbp. After polishing and purging redundant haplotigs, the assembly was reduced to 62,284 scaffolds totaling 2.77 Gbp with an N50 of 66.1 Kbp. Despite being somewhat fragmented, the resulting assembly is rather complete with 94.9% of BUSCOs detected (83.3% complete plus 11.6% fragmented). After removing putative contaminant scaffolds – those with fewer than 10 mapping Illumina reads, not annotated as Metazoa (Proteobacteria, Firmicutes, and "Bacteria-undef") or as "no-hit" in BlobTools, and/or with a GC content <0.30 or >0.55 – the final assembly consisted of 57,495 scaffolds totaling 2.52 Gbp with an N50 of 65.0 Kbp, an N90 of 19.97 Kbp, an L50 of 10.42 Kbp, an L90 of 38.44 Kbp, and a longest scaffold of 0.8 Mbp. After removal of putative contamination, 92.0% of BUSCOs could be detected (79.4% complete [74.4% single-copy and 5.0% duplicated], 12.6% fragmented, and 8.0% missing).

At 2.5 Gbp, the *Hanleya hanleyi* genome is over four times the size of that of the only other chiton with a genome sequenced to date, *Acanthopleura granulata.* RepeatModeler identified 327 families of repeats across five major classes (Table 1). The diversity of repetitive DNA motifs in the *Hanleya* genome is on par with that of other molluscan genomes with the exception of long terminal repeats (LTRs), which are much more diverse (100 different types) in *Hanleya* than any other molluscan genome we examined. A majority of repeats were annotated by RepeatClassifier as unclassified, likely because there are still few molluscan genomes incorporated in repetitive element databases. The genome of *Hanleya* has more than double the total repetitive content of that of *Acanthopleura*: 66.41% total interspersed repeats in *Hanleya* compared to 23.56% in *Acanthopleura* (Varney *et al.* 2020). Moreover, to our knowledge, the genome of *Hanleya* has an overall repetitive content higher than *any* mollusc sequenced to date (Gomes-dos-Santos *et al.* 2020).

Table 1. **The number of repetitive elements of various types across several molluscan genomes as indicated by RepeatModeler.**

Repetitive element	*Hanleya hanleyi*	*Acanthopleura granulata*	*Haliotis rufescens*	*Pinctada fucata*	*Crassostrea virginica*	*Bathymodiolus platifrons*	*Scapharca broughtonii*	*Lottia gigantea*
buffer	0	4	1	3	0	0	1	2
DNA	65	76	154	132	322	224	186	108
LINE	151	44	119	161	78	156	81	49
SINE	7	23	13	16	13	7	32	26
LTR	100	22	31	21	47	38	19	21
RC	4	10	13	0	85	52	60	3
Satellite	0	4	10	0	2	7	7	1

F1000Research 2022, 11:555 Last updated: 23 MAY 2023

BRAKER predicted 69,284 gene models with 92.9% of BUSCOs detected (81.8% complete [75.6% single-copy and 6.2% duplicated], 11.1% fragmented, and 7.1% missing). Of these, 35,362 were supported by transcriptome and/or protein evidence. Removal of gene models not supported by transcriptome or protein evidence had little effect on the estimated completeness of the gene models as 92.2% of BUSCOs were detected (81.3% complete [75.2% single-copy and 6.1% duplicated] 10.9% fragmented, and 7.8% missing).

Comparison of the full set of *Hanleya* gene models to the gene models from 20 other lophotrochozoans in OrthoFinder resulted in 185,272 groups of homologous sequences. Our pipeline selected 2,331 single-copy genes sampled for at least 18 of the 21 taxa. Of these, *Hanleya* was sampled for 2,168 genes (93%), further demonstrating the completeness of this genome. For comparison, *Lottia gigantea* (Gastropoda) was sampled for 2,243, *Crassostrea virginica* (Bivalvia) was sampled for 2,076, and *Acanthopleura granulata* (Polyplacophora) was sampled for 1,999. Concatenation resulted in a supermatrix 831,793 amino acids in length with 16.7% missing data. Phylogenetic analysis resulted in a strongly supported tree with maximal support for Polyplacophora and placement of Polyplacophora as sister to all other sampled molluscs (Figure 1C).

Sequencing data has been uploaded to NCBI SRA (see *Underlying data)* and all other results to Figshare (see *Extended data* (Kocot 2022)).

Conclusions

Despite challenges in assembling this relatively large (2.5 Gbp), heterozygous (1.3%), and repetitive (66.4%) genome, BUSCO analysis indicates that it is rather complete with 92.0% of BUSCOs detected in the final, decontaminated genome and 92.9% and 92.2% of BUSCOs detected in the full and evidence-supported predicted transcript sets, respectively. Our orthology inference pipeline recovered 93% of the genes sampled from at least 18/21 lophotrochozoan genomes in the *Hanleya*, further supporting the near completeness of this genome.

Data availability

Underlying data

NCBI Sequence Read Archive (SRA): RNA-Seq of *Hanleya hanleyi* mantle. Accession number SRX8235059. https://identifiers.org/ncbiprotein:SRX8235059.

NCBI SRA: Illumina Sequencing of *Hanleya hanleyi* gDNA. Accession number SRR18273088. https://identifiers.org/ncbiprotein:SRR18273088.

NCBI SRA: GridION Sequencing of *Hanleya hanleyi* gDNA. Accession numbers SRX14411365, https://identifiers.org/ncbiprotein:SRX14411365; SRX14411366, https://identifiers.org/ncbiprotein:SRX14411366; and SRX14411367, https://identifiers.org/ncbiprotein:SRX14411367.

Extended data

Figshare: *Hanleya hanleyi* genome extended data. https://doi.org/10.6084/m9.figshare.19672449.v2 (Kocot 2022).

This project contains the following extended data:

- 01_Jellyfish_and_GenomeScope.zip (Jellyfish and GenomeScope results)
- 02_MaSuRCA.zip (genome assembly produced by MaSuRCA)
- 03_purge_dups.zip (heterozygosity-purged genome assembly)
- 04_POLCA.zip (purge_dups output polished with Illumina reads in POLCA)
- 05_QUAST_and_BUSCO_on_final_genome_assembly.zip (QC of final assembly after POLCA)
- 06_RepeatMasker_and_RepeatModeler.zip (RepeatMasker & RepeatModeler output)
- 07_BlobTools.zip (BlobTools contamination screening results)
- 08_BRAKER.zip (Genome annotation with BRAKER)

- 09_BUSCO_on_gene_models.zip (QC on final gene models produced by BRAKER)

- final_genome_assembly_and_annotations.zip (final genome assembly and annotation)

Ethical approval

Ethics permits were not required to undertake this research because Institutional Animal Care and Use Committee (IACUC) review is not required for use of invertebrates in research activities at the University of Alabama.

Acknowledgments

We thank the late Dr. Hans Tore Rapp for gifting us the specimen of *Hanleya hanleyi* that inspired this project. We thank Dr. John Sutton for advice regarding Oxford Nanopore library preparation.

References

Blaxter M, Challis R: *BlobToolkit.* BlobToolKit; 2018.
Reference Source

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

Brooker LR, Shaw JA: **The chiton radula: a unique model for biomineralization studies.** *Advanced Topics in Biomineralization.* 2012; **1**: 65–84.

Brůna 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

Brůna 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

Buchfink B, Xie C, Huson DH: **Fast and sensitive protein alignment using DIAMOND.** *Nat. Methods.* 2015; **12**(1): 59–60.
PubMed Abstract | Publisher Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – Interactive Quality Assessment of Genome Assemblies.** *G3 Genes|Genomes|Genetics.* 2020; **10**(4): 1361–1374.
Publisher 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

Emms DM, Kelly S: **OrthoFinder: phylogenetic orthology inference for comparative genomics.** *Genome Biol.* 2019; **20**(1): 1–14.
Publisher 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

Kocot K: **Hanleya hanleyi genome extended data. figshare. [Dataset].** 2022.
Publisher 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): 101–108.
PubMed Abstract | Publisher Full Text

Krueger F: 2017. *TrimGalore.*
Reference Source | Reference Source

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
PubMed Abstract | Publisher Full Text

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

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

Ranallo-Benavidez 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

Sigwart JD: **Gross anatomy and positional homology of gills, gonopores, and nephridiopores in "basal" living chitons (Polyplacophora: Lepidopleurina).** *Am. Malacol. Bull.* 2008; **25**(1): 43–49.
Publisher Full Text

Sigwart JD, Schwabe E, Saito H, *et al.*: **Evolution in the deep sea: a combined analysis of the earliest diverging living chitons (Mollusca: Polyplacophora: Lepidopleurida).** *Invertebr. Syst.* 2011; **24**(6): 560–572.
Publisher Full Text

Sirenko BI, Sigwart J, Dell'Angelo B: ***Hanleya hanleyi* (Bean in Thorpe, 1844) (Mollusca, Polyplacophora) and the influence of the Gulf Stream System on its distribution.** *Ruthenica.* 2016; **26**(2).

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

Speiser DI, Eernisse DJ, Johnsen S: **A Chiton Uses Aragonite Lenses to Form Images.** *Curr. Biol.* 2011; **21**(8): 665–670.
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

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

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

Zimin AV, Salzberg SL: **The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies.** *PLoS Comput. Biol.* 2020; **16**(6): e1007981.
PubMed Abstract | Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 10 August 2022

https://doi.org/10.5256/f1000research.133601.r145650

✔ **Samuel Abalde**
Department of Zoology, Swedish Museum of Natural History, Stockholm, Sweden

Mollusca is an important animal phylum, and as such it has received a lot of attention from the scientific community. However, it is important to note that much of this attention has been drawn towards the three most diverse and economically important molluscan classes, while all other have been relatively neglected. From a genomic perspective the scenario is similar, with all but two mollusk genomes published to date sequenced from the same three classes. Despite the importance of the Aculifera, the clade containing chitons and aplacophorans, one of the two main clades of mollusks and hence fundamental to fully understanding mollusk evolution, only one genome has been generated to date, hampering comparative studies.

In this manuscript, Varney et al. report the complete genome of the chiton *Hanleya hanleyi*, the second aculiferan (and chiton) genome. The genome is relatively fragmented but seems to be very complete, and it will become an important addition to future studies of mollusk evolution.

I would like to congratulate the authors for their work. The manuscript is concise but it presents all the relevant information and the methods look sound. I have only three minor comments that, although will not change substantially the manuscript, I think the authors should consider:

- **"Extant chitons can be divided into two major clades: Chitonida, the clade represented by the previously published *Acanthopleura* genome, and Lepidopleurida, which includes *Hanleya*."** I am not an expert on chiton systematics, but to the best of my knowledge there are three main groups: Callochitonida, Chitonida, and Lepidopleurida. The same three groups were recovered in a recent phylogeny[1]. I am not aware of more recent updates on this matter, but if so then I think this should be referenced in the text to avoid misunderstandings.

- Pertaining to the previous comment: **"Because of this suite of putatively ancestral characteristics and its phylogenetic position as the sister taxon to all other chitons, Lepidopleurida is thought to be critical to understanding large-scale patterns in molluscan evolution."** If we accept there are only two main groups, then Lepidopleurida is as sister to all other chitons as Chitonida, so this sentence is technically correct but

misleading, because it makes you think of a ladderized tree and not in a sister relationship. If we consider the three groups mentioned above, then this sentence is correct.

- As for the repeat content, I wonder about their distribution in the genome. This number is not high enough to rise suspicions, there are other genomes above 50%, but since it will set the new upper limit for repeat content in molluscan genomes I would like to double check this figure is correct. Are the repeats scattered around the genome? Is it possible that they might be concentrated in a few contigs that should be quality-checked?

References

1. Irisarri I, Uribe J, Eernisse D, Zardoya R: A mitogenomic phylogeny of chitons (Mollusca: Polyplacophora). *BMC Evolutionary Biology*. 2020; **20** (1). Publisher Full Text

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: Phylogenomics, Bioinformatics, Genomics, Invertebrate Biology

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 24 June 2022

https://doi.org/10.5256/f1000research.133601.r138695

Vanessa L. Gonzalez
Global Genome Initiative, National Museum of Natural History, Smithsonian Institution,

Washington, DC, USA

The manuscript is clear, concise, well written and scientifically sound. This phylogenetically important taxon's genome assembly is a much-needed addition to the currently sampling of available molluscan genomes. All methods are explicitly outlined and are appropriate for the genome assembly (hybrid assembly, annotation & phylogenetic methods). The outcome of the annotation process is expected with the resulting contiguity of the genome (fragmented).

Not sure if I have missed it in the text, but is seems as though the database and database version that was used to calculate the BUSCO scores is not listed (maybe Metazoa?). If it was the Metazoan database, I think it would be helpful to also add the BUSCO scores for the Molluscan specific database as well.

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: Biodiversity Genomics, Invertebrate Biology, Bioinformatics

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.