DATA NOTE

**REVISED** The genome sequence of the grey top shell, *Steromphala cineraria* (Linnaeus, 1758) [version 3; peer review: 2 approved]

Patrick Adkins¹, Robert Mrowicki¹, Robert Mrowicki², Joanna Harley¹, Nova Mieszkowska¹, João G. R. N. Ferreira⁴, Marine Biological Association Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

¹Marine Biological Association, Plymouth, Devon, UK
²Natural History Museum, London, UK
³Bio Bureau Biotechnology, Rio de Janeiro, Brazil
⁴Tree of Life, Wellcome Sanger Institute, Cambridge, UK

**Abstract**
We present a genome assembly from an individual *Steromphala cineraria* (the grey topshell; Mollusca; Gastropoda; Trochida; Trochidae). The genome sequence is 1,270 megabases in span. Most of the assembly (99.23%) is scaffolded into 18 chromosomal pseudomolecules.

**Keywords**
Steromphala cineraria, grey topshell, genome sequence, chromosomal, Mollusca

This article is included in the Tree of Life gateway.

---

**v3 First published**: 08 Feb 2022, 7:44
https://doi.org/10.12688/wellcomeopenres.17677.1

**Second version**: 24 Mar 2022, 7:44
https://doi.org/10.12688/wellcomeopenres.17677.2

**Latest published**: 21 Mar 2024, 7:44
https://doi.org/10.12688/wellcomeopenres.17677.3

**Open Peer Review**

| Approval Status |
|-----------------|
| ✓   ✓            |

**version 3**
(revision)
21 Mar 2024
view view

**version 2**
(revision)
24 Mar 2022
view view

**version 1**
08 Feb 2022

1. André Gomes-dos-Santos, University of Porto, Porto, Portugal
2. Taro Maeda, Keio University, Mizukami, Japan

Any reports and responses or comments on the article can be found at the end of the article.
Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Adkins P: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; Mrowicki R: Investigation, Resources; Harley J: Investigation, Resources; Mieszkowska N: Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; Ferreira JGRN: Investigation, Methodology, Software;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (206194) and the Darwin Tree of Life Discretionary Award (218328). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2024 Adkins P 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: Adkins P, Mrowicki R, Harley J et al. The genome sequence of the grey top shell, Steromphala cineraria (Linnaeus, 1758) [version 3; peer review: 2 approved] Wellcome Open Research 2024, 7:44 https://doi.org/10.12688/wellcomeopenres.17677.3

First published: 08 Feb 2022, 7:44 https://doi.org/10.12688/wellcomeopenres.17677.1
Species taxonomy
Eukaryota; Metazoa; Spiralia; Lophotrochozoa; Mollusca; Gastropoda; Vetigastropoda; Trochida; Trochoidea; Trochidae; Cantharidinae; Steromphala; Steromphala cineraria (Linnaeus, 1758) (NCBI:txid216125).

Background
Steromphala cineraria (Linnaeus, 1758), commonly called the grey topshell, is a gastropod common to rocky shores in the UK. It typically occurs among boulders and cobbles on the lowshore and sub-tidally, where it grazes among Fucus and Laminaria species. Intertidally, it is most common on the lower shore, but can also be found in pools higher on the shore. Sub-tidally it extends to depths of 130 m, although it is most common in the kelp forests between 30 m and low water spring tide (Fretter & Graham, 1976). Its geographical distribution ranges from northern Norway to southern Portugal, becoming rarer at its range edges due to thermal limits being approached (Høisaeter, 2009; Nekhaev, 2013).

An important grazing species, S. cineraria is distinguished from other species of trochids by its bluntly conical shell and grey/yellowish finely striped patterning on the shell. In smaller shells, the umbilicus is large, becoming smaller and elliptical with age and in large shells sometimes becoming overgrown by the columellar lip (Fretter & Graham, 1976).

As S. cineraria is found across a large range of latitudes, it is exposed to a wide range of thermal environments in temperature, both due to time of year and geographical distribution. It is important to understand how populations may change in response to climate change, especially in its southern and northern range limits, and the knock-on effects this may have on macroalgae due to changes in grazing populations (Mieszkowska et al., 2007). A high quality genome sequence for this species will allow future studies to understand more about the mechanisms driving the observed response of this species to a changing climate.

Genome sequence report
The genome was sequenced from a single S. cineraria (Figure 1) collected from Mount Batten, Devon, UK (latitude 50.36084, longitude -4.12833). A total of 42-fold coverage in Pacific Biosciences single-molecule long reads and 35-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 408 missing/misjoins and removed 70 haplotypic duplications, reducing the assembly size by 2.51% and the scaffold number by 53.835%, and increasing the scaffold N50 by 123.17%.

The final assembly has a total length of 1,270 Mb in 283 sequence scaffolds with a scaffold N50 of 70.7 Mb (Table 1). Of the assembly sequence, 99.23% was assigned to 18 chromosomal-level scaffolds (numbered by sequence length) (Figure 2–Figure 5; Table 2). Viewing the high-resolution Pretext map shows that there are large inversions between sister chromatids can be seen on chromosome 5 at 29.7–60.7 Mb and chromosome 11 at 17.7–39.7 Mb. Possible inversions are also seen on chromosome 11 at Mb 3.4–39.4 and 18–66 Mb. The assembly has a BUSCO v5.1.2 (Manni et al., 2021) completeness of 85.4% (single 84.6%, duplicated 0.8%) using the mollusca_odb10 reference set (n=5295). However, we believe that this relatively low BUSCO score is a result of limitations with the current mollusca_odb10 geneset. Using the metazoa_odb10 reference set (n=954), the assembly has a completeness of 97.6% (single 97.0%, duplicated 0.6%), which we believe is evidence of high completeness. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Metadata for specimens, barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/216125.

Methods
Sample acquisition and nucleic acid extraction
A single S. cineraria specimen (xgSteCine2) was collected from Mount Batten, Devon, UK (latitude 50.36084, longitude -4.12833) by Rob Mrowicki (Natural History Museum),
Table 1. Genome data for *Steromphala cineraria*, xgSteCine2.1.

| Project accession data          |        |
|--------------------------------|--------|
| Assembly identifier            | xgSteCine2.1 |
| Species                        | *Steromphala cineraria* |
| Specimen                       | xgSteCine2 |
| NCBI taxonomy ID               | NCBI:txid216125 |
| BioProject                     | PRJEB45667 |
| BioSample ID                   | SAMEA7536348 |
| Isolate information            | Muscle   |

| Raw data accessions            |        |
|--------------------------------|--------|
| PacificBiosciences SEQUEL II   | ERR6939216, ERR6939217 |
| 10X Genomics Illumina          | ERR6363284-ERR6363287 |
| Hi-C Illumina                  | ERR6363289 |
| PolyA RNA-Seq Illumina         | ERR6688409 |

| Genome assembly                |        |
|--------------------------------|--------|
| Assembly accession             | GCA_916613615.1 |
| Accession of alternate haplotype| GCA_916613985.1 |
| Span (Mb)                      | 1,270  |
| Number of contigs              | 842    |
| Contig N50 length (Mb)         | 6.2    |
| Number of scaffolds            | 283    |
| Scaffold N50 length (Mb)       | 70.7   |
| Longest scaffold (Mb)          | 98.8   |
| BUSCO* genome score            | C:85.4%, [S:84.6%, D:0.8%], F:4.9%, M:9.7%, n:5295 |

*BUSCO scores based on the mollusca_odb10 BUSCO set using v5.1.2. C=complete, S=single copy, D=duplicated, F=fragmented, M=missing, n=number of orthologues in comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/Steromphala%20cineraria/dataset/CAKAJN01/busco.*

Patrick Adkins and Joanna Harley (both Marine Biological Association), by hand. The samples were identified by the same individual and snap-frozen in liquid nitrogen.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute. The xgSteCine2 sample was weighed and dissected on dry ice with tissue set aside for Hi-C and RNA sequencing. Muscle tissue was cryogenically disrupted to a fine powder using a Covaris cryoPREP Automated Dry Pulveriser, receiving multiple impacts. Fragment size analysis of 0.01–0.5 ng of DNA was then performed using an Agilent FemtoPulse. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 200-ng aliquot of extracted DNA using 0.8X AMPure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size between 12–20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA was purified by solid-phase reversible immobilisation using AMPure PB beads with a 1.8X ratio of beads to sample to remove the shorter fragments and concentrate the DNA sample. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit.
Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from muscle tissue in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer’s instructions. RNA was then eluted in 50 μl RNAse-free water and its concentration RNA assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

**Sequencing**

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud sequencing libraries were constructed according
to the manufacturers’ instructions. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II and Illumina NovaSeq 6000 instruments. Hi-C data were generated from additional muscle tissue of xgSteCine2 using the Arima v2.0 kit and sequenced on an Illumina NovaSeq 6000 instrument.

**Genome assembly**

Assembly was carried out with Hifiasm (Cheng et al., 2021). Haplotypic duplication was identified and removed with purge_dups (Guan et al., 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with longranger align, calling variants with freebayes (Garrison & Marth, 2012). The assembly was then scaffolded

---

**Figure 3. Genome assembly of Steromphala cineraria, xgSteCine2.1. GC coverage.** BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at [https://blobtoolkit.genomehubs.org/view/Steromphala%20cineraria/dataset/CAKAJN01/blob](https://blobtoolkit.genomehubs.org/view/Steromphala%20cineraria/dataset/CAKAJN01/blob).
with Hi-C data (Rao et al., 2014) using SALSA (Ghurye et al., 2019). The mitochondrial genome was assembled with MitoHiFi (Uliano-Silva et al., 2021), which performed annotation using MitoFinder (Allio et al., 2020). The assembly was checked for contamination and corrected using the gEVAL system (Chow et al., 2016) as described previously (Howe et al., 2021). Manual curation (Howe et al., 2021) was performed using gEVAL, HiGlass (Kerpedjiev et al., 2018) and Pretex. The genome was analysed within the BlobToolKit environment (Challis et al., 2020). Table 3 contains a list of all software tool versions used, where appropriate.

**Ethics/compliance issues**

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject
Table 2. Chromosomal pseudomolecules in the genome assembly of *Steromphala cineraria*, xgSteCine2.1.

| INSDC accession | Chromosome | Size (Mb) | GC% |
|-----------------|------------|-----------|-----|
| OU744720.1      | 1          | 98.78     | 35.1|
| OU744721.1      | 2          | 90.22     | 35.1|
| OU744722.1      | 3          | 87.49     | 35.6|
| OU744723.1      | 4          | 76.76     | 35.9|
| OU744724.1      | 5          | 75.56     | 35.5|
| OU744725.1      | 6          | 74.15     | 35.3|
| OU744727.1      | 7          | 70.75     | 35.0|
| OU744726.1      | 8          | 71.35     | 35.3|
| OU744728.1      | 9          | 70.17     | 35.7|

Figure 5. Genome assembly of *Steromphala cineraria*, xgSteCine2.1: Hi-C contact map. Hi-C contact map of the xgSteCine2.1 assembly, visualised in HiGlass. Chromosomes are shown in size order from left to right and top to bottom.
Table 3. Software tools used.

| Software tool   | Version | Source                           |
|-----------------|---------|----------------------------------|
| Hifiasm         | 0.15    | Cheng et al., 2021              |
| purge_dups      | 1.2.5   | Guan et al., 2020               |
| SALSAs2         | 3.0     | Ghurye et al., 2019             |
| longranger align| 2.2.2   | https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines |
| freebayes       | v1.3.1  | Garrison & Marth, 2012          |
| MitoHiFi        | 2       | https://github.com/marcelauliano/MitoHiFi |
| gEVAL           | N/A     | Chow et al., 2016               |
| HiGlass         | 1.11.6  | Kerpedjiev et al., 2018         |
| PretextView      | 0.2.x   | https://github.com/wtsi-hpag/PretextView |
| BlobToolKit     | 2.6.4   | Challis et al., 2020            |

to the Darwin Tree of Life Project Sampling Code of Practice. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances other Darwin Tree of Life collaborators.

Data availability
European Nucleotide Archive: Steromphala cineraria (grey top shell). Accession number PRJEB45667; https://identifiers.org/ena.embl/PRJEB45667.

The genome sequence is released openly for reuse. The S. cineraria genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated with the RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1.

Author information
Members of the Marine Biological Association Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.591383.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.5744972.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: https://doi.org/10.5281/zenodo.5744840.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.5746904.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5743293.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.5638618.

References
Allio R, Schomaker-Bastos A, Romiguier J, et al.: MitoFinder: Efficient Automated Large-Scale Extraction of Mitogenomic Data in Target Enrichment Phylogenomics. Mol Ecol Resour. 2020; 20(4): 892–905. PubMed Abstract | Publisher Full Text | Free Full Text
Challis R, Richards E, Rajan J, et al.: BlobToolKit - Interactive Quality Assessment of Genome Assemblies. G3 (Bethesda). 2020; 10(4): 1361-74. PubMed Abstract | Publisher Full Text | Free Full Text
Cheng H, Concepcion GT, Feng X, et al.: Haplotype-Resolved de Novo Assembly Using Phased Assembly Graphs with Hifiasm. Nat Methods. 2021; 18(2): 170-75. PubMed Abstract | Publisher Full Text | Free Full Text
Chow W, Brugger K, Caccamo M, et al.: gEVAL - a Web-Based Browser for Evaluating Genome Assemblies. Bioinformatics. 2016; 32(16): 2508-10. PubMed Abstract | Publisher Full Text | Free Full Text
Open Peer Review

Current Peer Review Status:

✅✅

Version 3

Reviewer Report 06 May 2024

https://doi.org/10.21956/wellcomeopenres.23196.r77343

© 2024 Gomes-dos-Santos 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.

✅ André Gomes-dos-Santos

Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Porto, Portugal

The authors, Adkins et al., have addressed some of my previous concerns in their revised manuscript. However, not all of my concerns have been adequately addressed. I would appreciate the authors' opinions on the remaining issues I have raised. Since I did not receive a response to my previous review comments, I am uncertain as to why some of my concerns were not addressed in the revision.

Nevertheless, I believe that the manuscript has been significantly improved.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics, Phylogenomics, 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 26 April 2024

https://doi.org/10.21956/wellcomeopenres.23196.r77344

© 2024 Maeda T. 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.

✅ Taro Maeda

Institute for Advanced Biosciences, Keio University, Mizukami, Japan
I am grateful for the opportunity to review the revised manuscript once again. I am pleased that the authors have provided clear information about the source samples for the RNA data. However, I have observed that some of my previous comments still need to be addressed in their comment or revised manuscript. Although I regret that I did not receive several corresponding comments from the author, Considering that this is a brief communication note, the information in question is not essential for this type of article format. If the editorial team determines to accept without the additional information, I respect and agree with that decision.

Below, I list the questions for which I could not find answers in the revised text or comments, along with my new comments on this version of the manuscript.

“The authors need to clarify the covering degree of their genome data against the total genome size and the number of chromosomes. Were the total scaffold length and the number of chromosomal pseudomolecules consistent with the predicted genome size and chromosome? At least the authors should describe the genome size predicted by the k-mer frequency analysis. If the authors did not measure the number of chromosomes, it should be made clear that they inferred the number of chromosomes from the Hi-C results.”

"The authors opened PolyA RNA-seq results (ERR6688409), but it is unclear how this data contributed to the analysis. Although I expect it to be sequenced to predict gene regions, I found no description of the construction of the gene model. I also found no gene model data from the submitted genome data on the public database. The publication of raw RNA-Seq data is valuable for future analysis. However, the authors should clarify in the paper why this data was not used in the study."

"A total of 42-fold coverage in Pacific Biosciences". Please clarify the predicted genome size of this species and the prediction method. What kind "-m" options were used for the gene model prediction for the BUSCO analysis? Did you use RNA-seq data during the BUSCO analysis? Please describe the number of gap regions (N-base) on the chromosomal-level scaffolds."

The authors have responded as follows, but the corresponding link was not found in the text. "We have included a link to the TOLQC page, which supplies required information about the genome sequencing and assembly, including k-mer frequency spectra analyses."

When I investigated the TOLQC page, I believed the following page was relevant, but I could not find the corresponding link. https://tolqc.cog.sanger.ac.uk/darwin/molluscs/Steromphala_cineraria/

Additionally, I could not find a genome size description based on k-mer analysis on the relevant web page.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Sacoglossa, Genome, Symbiosis, Kleptoplasty

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.
I am honored to have the opportunity to review the paper by Adkins et al. This paper reports the genomic data of a sea snail, *Steromphala cineraria*. The genome of this species has never been revealed. There is no doubt that this data is worthy of publication. I think, however, some additional information needs to be provided to ensure the reliability of this data, as follows.

Unfortunately, their genome data included no gene model and annotation information, and I found no description of how the RNA-Seq data contributed to the analysis. This deficiency may confuse the reader. Please clarify that the genomic data contain no gene model and annotation data and the role of RNA-Seq.

**Abstract**

The authors need to clarify the covering degree of their genome data against the total genome size and the number of chromosomes. Were the total scaffold length and the number of chromosomal pseudomolecules consistent with the predicted genome size and chromosome? At least the authors should describe the genome size predicted by the k-mer frequency analysis. If the authors did not measure the number of chromosomes, it should be made clear that they inferred the number of chromosomes from the Hi-C results.

**Table 1**

The authors opened PolyA RNA-seq results (ERR6688409), but it is unclear how this data contributed to the analysis. Although I expect it to be sequenced to predict gene regions, I found no description of the construction of the gene model. I also found no gene model data from the submitted genome data on the public database.

The publication of raw RNA-Seq data is valuable for future analysis. However, the authors should clarify in the paper why this data was not used in the study.

Moreover, please describe whether this RNA-Seq data were obtained from the same individual used for genome sequencing. If the RNA is obtained from different samples, I recommend clarifying the sampling point and the preincubation state before the RNA extraction.
"A total of 42-fold coverage in Pacific Biosciences". Please clarify the predicted genome size of this species and the prediction method.

What kind "-m" options were used for the gene model prediction for the BUSCO analysis? Did you use RNA-seq data during the BUSCO analysis? Please describe the number of gap regions (N-base) on the chromosomal-level scaffolds.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Partly

Are the datasets clearly presented in a useable and accessible format?
Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Sacoglossa, Genome, Symbiosis, Kleptoplasty

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.

Reviewer Report 14 April 2022

https://doi.org/10.21956/wellcomeopenres.19632.r49928

© 2022 Gomes-dos-Santos 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.

**André Gomes-dos-Santos**
Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Porto, Portugal

Thank you for the opportunity to review this Data Note. In the Manuscript entitled " The genome sequence of the grey top shell, Steromphala cineraria (Linnaeus, 1758) ", the Authors Adkins et al, produced the high quality, chromosome level assembly of the grey top shell, using the de novo approach of the Darwin Tree of Life Project. The genome here presented has been made publicly available and is undoubtedly an important genomic resource. However, as seen in detail below, I have some concerns regarding possible technical issues. Thus, I am in favour of indexing, but only
after the raised issues are answered.

Abstract

Please also include the overall stats of the mtDNA assembly in the abstract.

Background

Regarding the species distribution, the authors report “Its geographical distribution ranges from southern Portugal and north to the White Sea in northern Russia, becoming rarer at its range edges as thermal limits are approached (Nekhaev, 2013).”

However, in the Nekhaev 2013, the following is stated (please note that Gibbula cineraria is a synonym of Steromphala cineraria):

- “Distribution: In Atlantic G. cineraria is distributed from Morocco to Northern Norway [Fretter, Graham, 1977; Poppe, Goto, 1991]. It is common along the Norway coast, but live animals were not found in East Finnmark [Høisøeter, 2009].”

- “Remarks: This species was previously reported from Ura Bay [Knipowitsch, 1900] but not mentioned in recent Russian faunistic and taxonomic literature [Galkin, 1955; Golikov, 1995; Golikov et al., 2001; Kantor, Sysoev, 2006].”

It is important that the authors revised this information accordingly to the citations they provided or any other recent citation that may support their claim.

Regarding the last sentence of the Background: “A high quality genome sequence for this species will allow future studies to understand more about the mechanisms driving the observed response of this species to a changing climate.”

Although I recognize the fundamental importance of having a high-quality genome assembly to study such adaptative responses, by itself it is not enough. Especially given that no genome annotation is provided (which I understand is not a requirement). Given that, I would recommend changing the statement, so it is understood that the genome is important as it serves as a basal tool for future studies on the subject.

Genome sequence report

Figure 1 - Images B-D have very poor resolution, the text on the labels is very hard to read. I recommend new photos.

Regarding sequencing coverage, i.e., 42-fold in PacBio and 35-fold in 10X Genomics. Were these estimations based on any previous expectation of genome size? Have the authors produced a k-mer frequency spectrum analysis with the 10X Genomic reads? How did the authors decide the amount of sequencing output to produce? Molluscan genome sizes are highly variable, so it is customary to have guided assumptions prior to sequencing.

Regarding the chromosome level scaffolding, given that SALSA does not require prior knowledge of the number of chromosomes and the high percentage of scaffolds assigned to chromosomes, the results seem highly reliable. However, I wonder if any prior estimation of the number of chromosomes for this group is available, that may further support the results.
Regarding the statement “Large inversions between sister chromatids can be seen on chromosome 5 at 29.7–60.7 Mb and chromosome 11 at 17.7–39.7 Mb. Possible inversions are also seen on chromosome 11 at Mb 3.4–39.4 and 18–66 Mb.” This is not evident from the Hi-C contact map. How did the authors determine this?

Methods

The authors have produced a mitogenome assembly as well. However, they did not provide a characterization of the mitogenome nor have included an accession number for the mitogenome. According to Nekhaev (2013), identification based on morphological characters may be misleading for Steromphala cineraria. Since the whole mtDNA has proven to be a great complementary tool in Systematics, I think that is important to highlight this new resource as well. Please provide a general characterization of the mitogenome and include the accession number in Table 1.

I noticed that the mtDNA available on NCBI (accession OU744738.1) is not annotated. I assume that it is not mandatory. As a note for the journal, I think that in the future, mtDNA annotation, which is very easily accomplished, should be a requirement. For instance, I noticed that this mtDNA sequence is almost 20.00bp long, largely due to a 2,853bp long tandem repeat of the motif TA. This repeat is not present in any of the other mtDNA available for the same genus (i.e., Steromphala (Gibbula) umbilicaris, accession NC_035682.1). Interestingly, the same pattern is present in another species of the same Family, i.e., Phorcus lineatus (accession OVI12099.1), which is also associated with another Wellcome Sanger Institute Tree of Life project, also not annotated. Have the authors confirmed that this repetitive pattern is not a result of a sequencing error? Have the authors validated this pattern with both sequencing strategies?

If it is not a sequencing error, then it is a novel character that may have phylogenetic implications and highlight the importance of using long read technologies to produce mtDNA assemblies. Please note that I limited my search to sequences on NCBI. Furthermore, I produced a quick annotation using Mitos2 web server, that resulted in the following warning “Translational exceptions: start=nad4, stop=nad4,nad4” which is reflected in the fact that the gene is apparently truncated. This was neither reported for the species of the same genus in NCBI (accession NC_035682.1) nor when I produced an annotation on Mitos2. I also annotated the Phorcus lineatus (accession OVI12099.1) with Mitos2, which did not report such a warning. This, highlights the importance of producing such resources, given that if confirmed, it reveals a novel uncharacterized feature. On the other hand, it may provide additional small-scale validation of the overall quality of the genome assembly (given that the whole genome annotation is not provided). The fact that such an error might be present in the mtDNA may be a warning for similar errors in the whole genome.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of methods and materials provided to allow replication by others?

Yes
Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics, Phylogenomics, 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, however I have significant reservations, as outlined above.