DATA NOTE

The genome sequence of the spiny starfish, *Marthasterias glacialis* (Linnaeus, 1758) [version 1; peer review: 2 approved, 1 approved with reservations]

Mara K.N. Lawniczak[1], 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

[1]Tree of Life, Wellcome Sanger Institute, Cambridge, UK

v1 **First published:** 05 Nov 2021, **6**:295
https://doi.org/10.12688/wellcomeopenres.17344.1
Latest published: 05 Nov 2021, **6**:295
https://doi.org/10.12688/wellcomeopenres.17344.1

Abstract

We present a genome assembly from an individual *Marthasterias glacialis* (the spiny starfish; Echinodermata; Asteroidea; Forcipulatida; Asteriidae). The genome sequence is 521 megabases in span. The majority of the assembly, 99.44%, is scaffolded into 22 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, and is 16 kb in span.

Keywords

Marthasterias glacialis, spiny starfish, genome sequence, chromosomal

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ? ✓ ✓

	1	2	3
version 1 05 Nov 2021	? view	✓ view	✓ view

1. **Maurice R. Elphick**, Queen Mary University of London, London, UK
2. **Charles A Ettensohn**, Carnegie Mellon University, Pittsburgh, USA
3. **Shunsuke Yaguchi**, Tsukuba Daigaku, Tsukuba, Japan

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

Corresponding authors: Mara K.N. Lawniczak (mara@sanger.ac.uk), Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Lawniczak MKN: Conceptualization, Investigation, Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing;

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.

How to cite this article: Lawniczak MKN, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme *et al.* **The genome sequence of the spiny starfish, *Marthasterias glacialis* (Linnaeus, 1758) [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2021, **6**:295 https://doi.org/10.12688/wellcomeopenres.17344.1

First published: 05 Nov 2021, **6**:295 https://doi.org/10.12688/wellcomeopenres.17344.1

Species taxonomy

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Asterozoa; Asteroidea; Forcipulatacea; Forcipulatida; Asteriidae; Marthasterias; *Marthasterias glacialis* (Linnaeus, 1758) (NCBI:txid7609).

Background

The spiny starfish, *Marthasterias glacialis*, is an opportunistic and generalist feeder distributed widely throughout Europe (https://www.sealifebase.ca/summary/Marthasterias). In England and Scotland, it is only found on the west coast, and the specimen sequenced here was captured on the Isle of Cumbrae in Scotland. It was found at a depth of about 4 metres on the northeast of the island, and processed in the lab of FSC Millport in August 2020. An image of the sequenced specimen just prior to processing is provided in Figure 1. Cytochrome Oxidase I studies of the species throughout Europe have shown that there are two divergent lineages in the Mediterranean (Pérez-Portela *et al.*, 2010). This new reference genome will assist in better understanding population structure within the species across its full range.

Genome sequence report

The genome was sequenced from a single *M. glacialis* of unknown sex collected from Farland Point, Great Cumbrae, North Ayrshire, Scotland (latitude 55.746815, longitude -4.914907). A total of 49-fold coverage in Pacific Biosciences single-molecule long reads and 69-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 363 missing/misjoins and removed 42 haplotypic duplications, reducing the assembly size by 2.74% and scaffold number by 61.31%, and increasing the scaffold N50 by 37.41%.

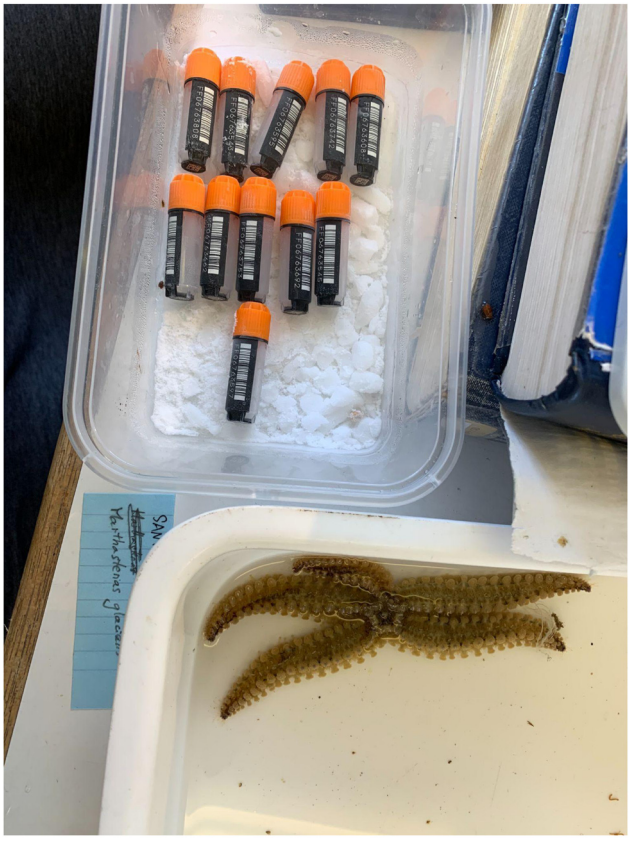

Figure 1. An image of the sequenced specimen, eaMarGlac1, taken immediately prior to processing and preservation.

The final assembly has a total length of 521 Mb in 106 sequence scaffolds with a scaffold N50 of 25 Mb (Table 1). Of the assembly sequence, 99.44% was assigned to 22 chromosomal-level scaffolds, representing 22 autosomes (numbered by sequence length) (Figure 2–Figure 5; Table 2). The assembly has a BUSCO (Simão *et al.*, 2015) v5.1.2 completeness of 98.4% using the

Table 1. Genome data for *Marthasterias glacialis*, eaMarGlac1.1.

Project accession data	
Assembly identifier	eaMarGlac1.1
Species	*Marthasterias glacialis*
Specimen	eaMarGlac1
NCBI taxonomy ID	7609
BioProject	PRJEB45116
BioSample ID	SAMEA7522991
Isolate information	Unknown sex; legs
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6436374
10X Genomics Illumina	ERR6054755-ERR6054758
Hi-C Illumina	ERR6054759
Illumina polyA RNA-Seq	ERR6054760
Genome assembly	
Assembly accession	GCA_911728455.1
Accession of alternate haplotype	GCA_911728445.1
Span (Mb)	521
Number of contigs	574
Contig N50 length (Mb)	1.9
Number of scaffolds	106
Scaffold N50 length (Mb)	25.2
Longest scaffold (Mb)	38.5
BUSCO* genome score	C:98.4%[S:98.1%,D:0.3%], F:0.9%,M:0.6%,n:954

*BUSCO scores based on the metazoa_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/eaMarGlac1.1/dataset/CAJVRT01/busco.

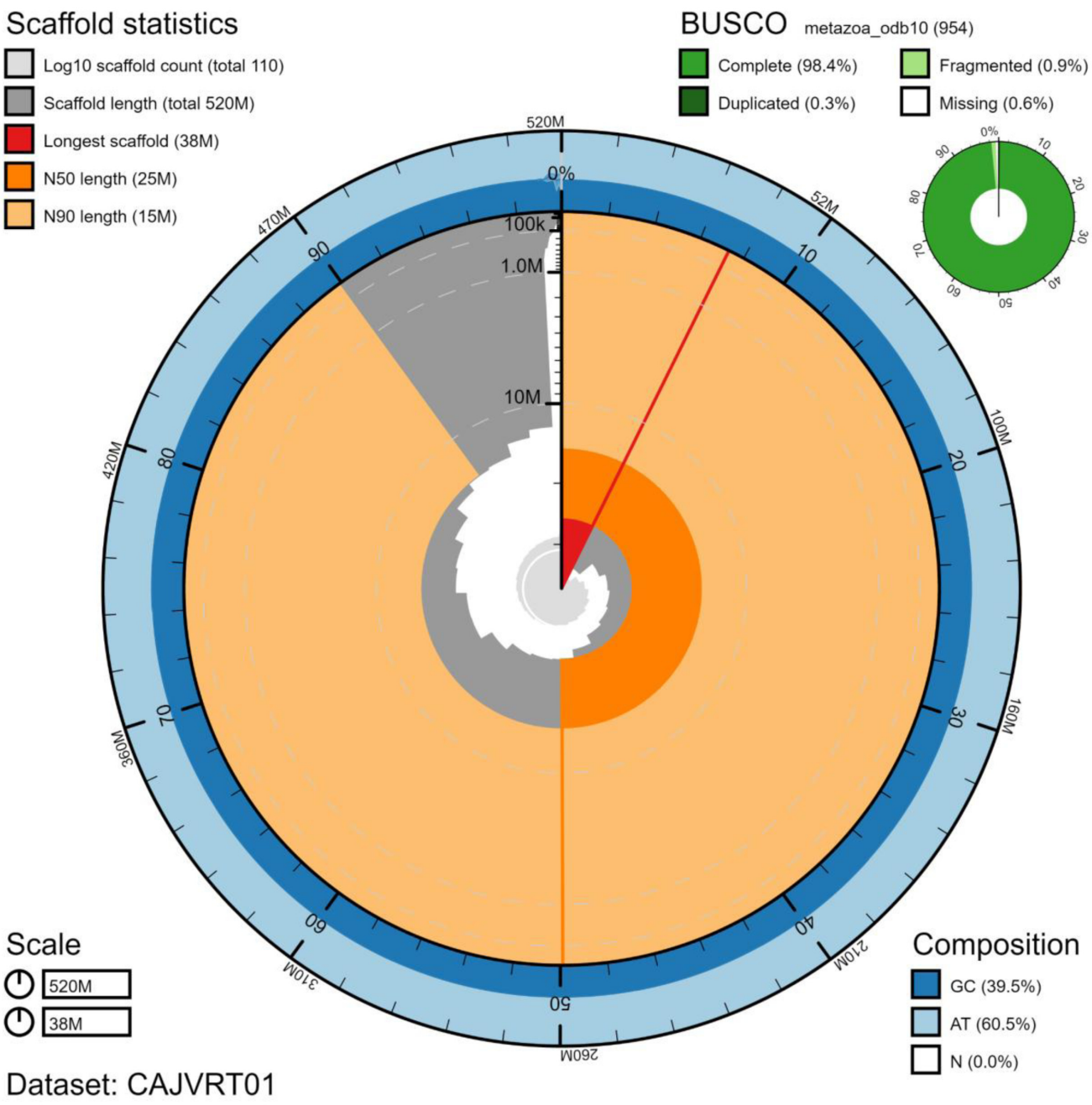

Figure 2. Genome assembly of *Marthasterias glacialis*, eaMarGlac1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 520,959,849 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (38,475,774 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (25,218,880 and 15,203,031 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the metazoa_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/eaMarGlac1.1/dataset/CAJVRT01/snail.

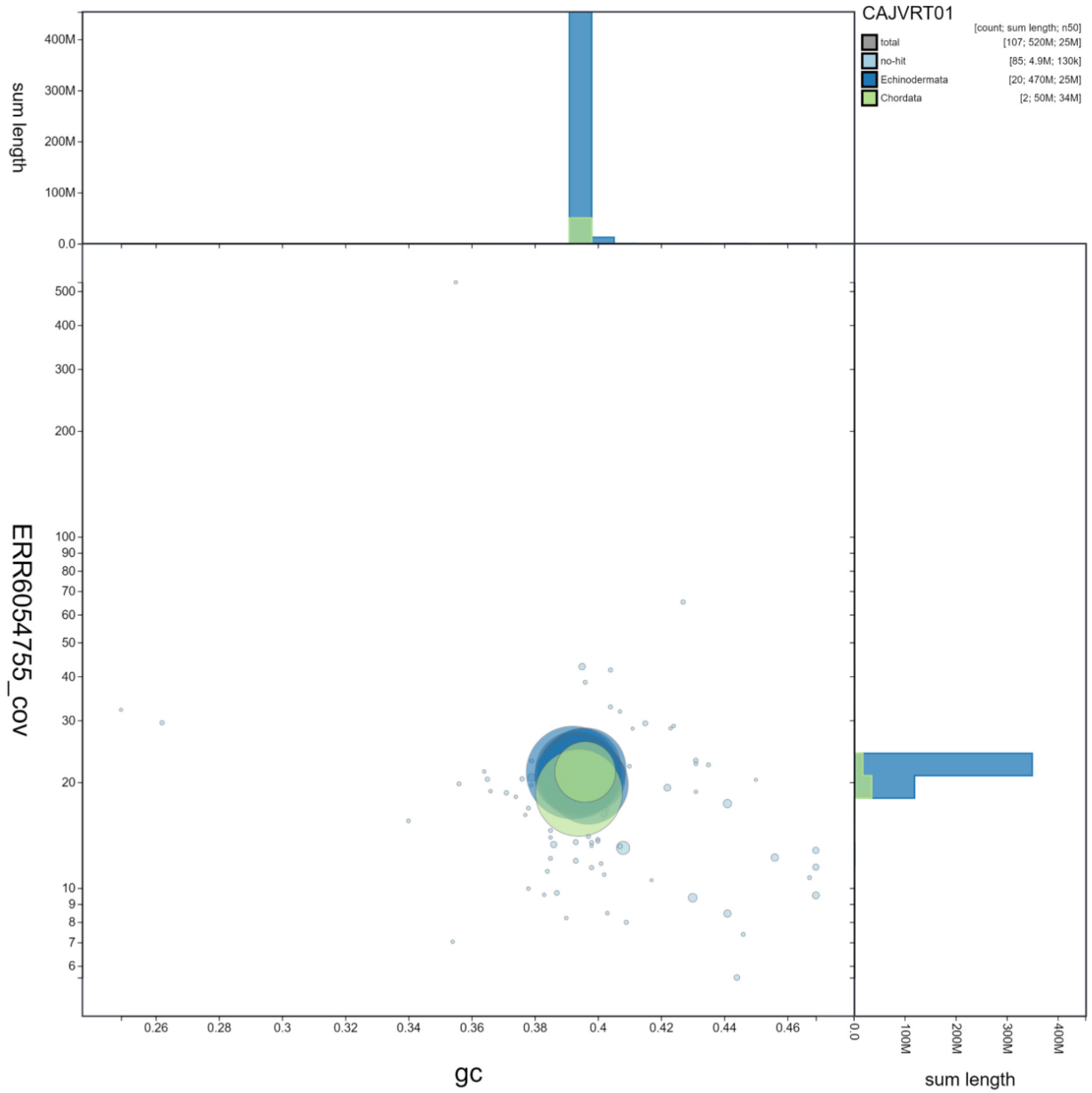

Figure 3. Genome assembly of *Marthasterias glacialis*, eaMarGlac1.1: GC-coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. Scaffolds labelled Chordata are spurious and assumed to reflect some confusion in the sequence databases from which data is pulled. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/eaMarGlac1.1/dataset/CAJVRT01/blob.

metazoa_odb10 reference set. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Methods

Sample acquisition and nucleic acid extraction

A single *M. glacialis* of unknown sex was collected from Farland Point, Great Cumbrae, North Ayrshire, Scotland (latitude 55.746815, longitude -4.914907) by Mara Lawniczak (Wellcome Sanger Institute (hereafter *Sanger*)). The specimen was collected by hand while snorkelling, and identified by Richard Durbin (University of Cambridge/Sanger) and Mark Blaxter (Sanger) and preserved and processed on dry ice by Mara Lawniczak.

DNA was extracted at the Tree of Life laboratory, WSI. The *M. glacialis* sample was weighed and dissected on dry ice with tissue set aside for RNA extraction and Hi-C sequencing.

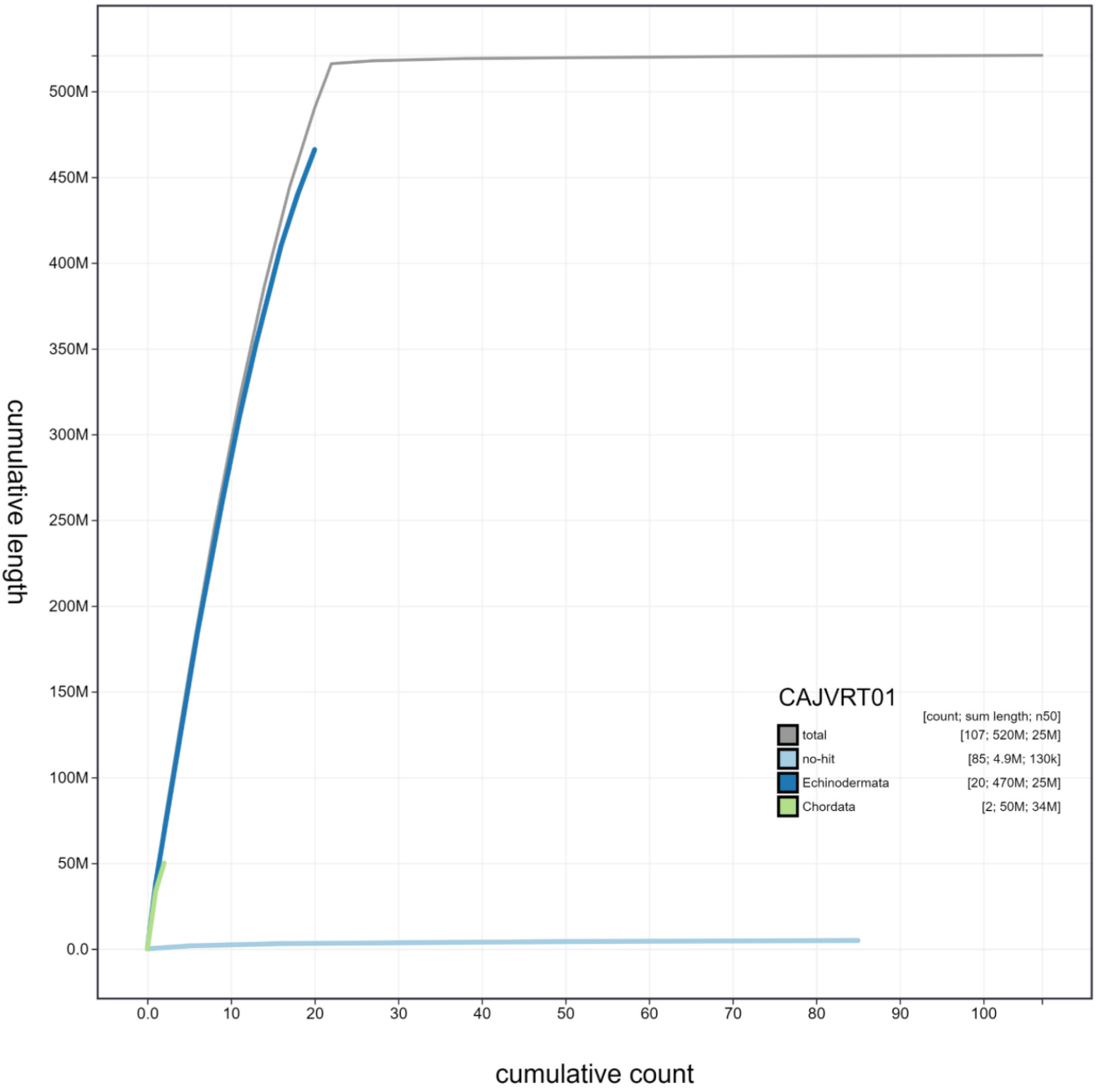

Figure 4. Genome assembly of *Marthasterias glacialis*, eaMarGlac1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/eaMarGlac1.1/dataset/CAJVRT01/cumulative.

Leg 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

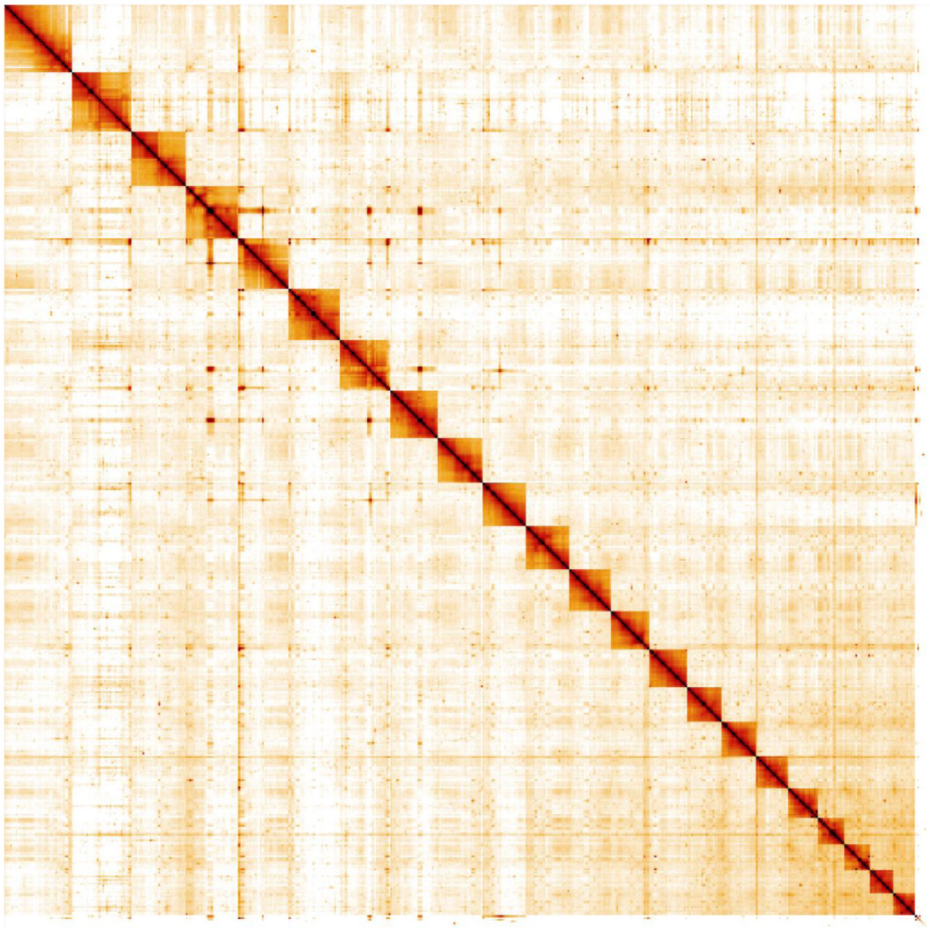

Figure 5. Genome assembly of *Marthasterias glacialis*, eaMarGlac1.1: Hi-C contact map. Hi-C contact map of the eaMarGlac1.1 assembly, visualised in HiGlass. Chromosomes are shown in order of size from left to right and top to bottom.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Marthasterias glacialis*, eaMarGlac1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
OU452219.1	1	38.48	39.2
OU452220.1	2	33.79	39.4
OU452221.1	3	30.68	39.4
OU452222.1	4	29.52	39.6
OU452223.1	5	29.13	39.3
OU452224.1	6	28.78	39.4
OU452225.1	7	28.21	39.7
OU452226.1	8	26.68	39.5
OU452227.1	9	25.22	39.3
OU452228.1	10	25.09	39.3
OU452229.1	11	24.75	39.5
OU452230.1	12	23.37	39.5
OU452231.1	13	21.44	39.4
OU452232.1	14	21.39	39.1
OU452233.1	15	19.79	39.8
OU452234.1	16	19.49	39.6
OU452235.1	17	18.12	39.6
OU452236.1	18	16.22	39.6
OU452237.1	19	15.20	39.7
OU452238.1	20	14.95	39.8
OU452239.1	21	13.34	39.8
OU452240.1	22	12.43	39.9
OU452241.1	MT	0.02	35.6
-	Unplaced	4.86	41

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 further leg tissue in the Tree of Life Laboratory at the WSI using TRIzol (Invitrogen), according to the manufacturer's instructions. RNA was then eluted in 50 μl RNAse-free water and its concentration 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 Chromium read cloud sequencing libraries were constructed according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. Sequencing was performed by the Scientific Operations core at Sanger on Pacific Biosciences SEQUEL II (HiFi), Illumina HiSeq X (10X) and Illumina HiSeq 4000 (RNA-Seq) instruments. Hi-C data were generated using the Arima v2 Hi-C kit and sequenced on an Illumina NovaSeq 6000 instrument.

Genome assembly

Assembly was carried out with HiCanu (Nurk *et al.*, 2020). Haplotypic duplication was identified and removed with purge_dups (with purging in the middle of contigs) (Guan *et al.*, 2020). Scaffolding with Hi-C data (Rao *et al.*, 2014) was carried out with SALSA2 (Ghurye *et al.*, 2019). The Hi-C scaffolded assembly was polished with the 10X Genomics Illumina data by aligning to the assembly with longranger align, calling variants with freebayes (Garrison & Marth, 2012). One round of the Illumina polishing was applied. 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 was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext. Regions of concern were identified and resolved using 10X longranger and genetic mapping data. 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.

Ethical/compliance issues

The materials that have contributed to this genome note were supplied by a Tree of Life collaborator. The Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible.

The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material;
- Legality of collection, transfer and use (national and international).

Each transfer of samples is undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Tree of Life collaborator, Genome Research

Table 3. Software tools used.

Software tool	Version	Source
HiCanu	2.1	Nurk *et al.*, 2020
purge_dups	1.2.3	Guan *et al.*, 2020
longranger	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
freebayes	1.3.1-17-gaa2ace8	Garrison & Marth, 2012
SALSA2	2.2	Ghurye *et al.*, 2019
MitoHiFi	1.0	Uliano-Silva *et al.*, 2021
gEVAL	N/A	Chow *et al.*, 2016
HiGlass	1.11.6	Kerpedjiev *et al.*, 2018
PretextView	0.1.x	https://github.com/wtsi-hpag/PretextView
BlobToolKit	2.6.2	Challis *et al.*, 2020

Limited (operating as the Wellcome Sanger Institute) and in some circumstances other Tree of Life collaborators.

Data availability

European Nucleotide Archive: Marthasterias glacialis (spiny starfish). Accession number PRJEB45116; https://identifiers.org/ena.embl/PRJEB45116.

The genome sequence is released openly for reuse. The *M. glacialis* 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 using 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 Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.4893704.

Members of the Wellcome Sanger Institute Tree of Life programme collective are listed here: https://doi.org/10.5281/zenodo.5377053.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.4790456.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013542.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783559.

References

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

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

Garrison E, Marth G: **Haplotype-Based Variant Detection from Short-Read Sequencing.** arXiv: 1207.3907. 2012.
Reference Source

Ghurye J, Rhie A, Walenz BP, *et al.*: **Integrating Hi-C Links with Assembly Graphs for Chromosome-Scale Assembly.** *PLoS Comput Biol.* 2019; **15**(8): e1007273.
PubMed Abstract | Publisher Full Text | Free Full Text

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and Removing Haplotypic Duplication in Primary Genome Assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–98.
PubMed Abstract | Publisher Full Text | Free Full Text

Howe K, Chow W, Collins J, *et al.*: **Significantly Improving the Quality of Genome Assemblies through Curation.** *GigaScience.* 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: Web-Based Visual Exploration and Analysis of Genome Interaction Maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | Publisher Full Text | Free Full Text

Nurk S, Walenz BP, Rhie A, *et al.*: **HiCanu: Accurate Assembly of Segmental Duplications, Satellites, and Allelic Variants from High-Fidelity Long Reads.** *Genome Res.* 2020; **30**(9): 1291–1305.
PubMed Abstract | Publisher Full Text | Free Full Text

Pérez-Portela R, Villamor A, Almada V: **Phylogeography of the Sea Star *Marthasterias Glacialis* (Asteroidea, Echinodermata): Deep Genetic Divergence between Mitochondrial Lineages in the North-Western Mediterranean.** *Mar Biol.* 2010; **157**(9): 2015–28.
Publisher Full Text

Rao SS, Huntley MH, Durand NC, *et al.*: **A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping.** *Cell.* 2014; **159**(7): 1665–80.
PubMed Abstract | Publisher Full Text | Free Full Text

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–12.
PubMed Abstract | Publisher Full Text

Uliano-Silva M, Nunes JGF, Krasheninnikova K, *et al.*: **marcelauliano/MitoHiFi: mitohifi_v2.0**. 2021.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ? ✓ ✓

Version 1

Reviewer Report 14 March 2023

https://doi.org/10.21956/wellcomeopenres.19174.r55125

✓ **Shunsuke Yaguchi**
Shimoda Marine Research Center, Tsukuba Daigaku, Tsukuba, Ibaraki Prefecture, Japan

The manuscript describes the whole genome assembly of *M. glacialis*. The authors detail the sampling, reading, and assembly process of the genome. Thanks to recent advancements in sequencing and bioinformatics technologies, it has become much easier to read genomes of many non-model organisms compared to the previous decade. However, not many journals accept simple "read and assembled" genome papers like this manuscript. Therefore, I'm personally pleased to see this Data Note on genome reading in this journal.

Minor points;

1. Why didn't the authors determine the specimen's sex?

2. Which part of the leg (generally referred to as the arm by other reviewers) did the authors use for sampling? As the leg or arm is an organ and not a tissue, the authors need to indicate whether they used the whole leg (arm) or just a part of it.

3. The scale of Figure 2 (520M/38M) may be a mistake.

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?
Yes

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Developmental biology and physiology of echinoderm embryos and larvae.

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 03 March 2023

https://doi.org/10.21956/wellcomeopenres.19174.r55122

Charles A Ettensohn
Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

This data note describes a genome assembly of an unsexed individual of the spiny starfish (*M. glacialis*). This work is part of the Darwin Tree of Life Project. The report is quite brief but the level of detail seems adequate and the assembly is publicly available. It has not been annotated and so will be of limited use to the community until that is completed. Certainly the annotated genome will be of interest to biologists of many stripes.

Minor questions:

1. Why was the individual not sexed? It would be trivial to sample a gonad and determine the sex, assuming the gonad had at at least some mature gametes.

2. I have never heard reference to the "legs" of a seastar. Do you mean the arms?

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?
Yes

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Developmental biology and genomics, using echinoderms as model organisms

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 15 December 2021

https://doi.org/10.21956/wellcomeopenres.19174.r47069

?

Maurice R. Elphick
School of Biological & Behavioural Sciences, Queen Mary University of London, London, UK

In this article the authors report the sequencing of the genome of the spiny starfish *Marthasterias glacialis*. This is the second starfish species to have its genome sequenced by the Wellcome Sanger Institute – a chromosomal assembly of the genome sequence of the common starfish *Asterias rubens* was submitted to NCBI on 12 February 2020 (PRJEB33974). This is relevant from a taxonomic perspective because both *Marthasterias glacialis* and *Asterias rubens* are species that belong to the same order (Forcipulatida) and family (Asteriidae). In contrast, the first starfish species to have its genome sequence reported, *Acanthaster planci* (Hall *et al.*, 2017[1]), belongs to a different order (Valvatida). It is to be expected, therefore, that the *A. rubens* and *M. glacialis* genomes will share more similarities than either genome shares with the genome of *A. planci*.

Assembly of the *M. glacialis* genome sequence is consistent with a karyotype comprising 22 chromosomes, which is consistent with the karyotype of *A. rubens* based on the assembled genome sequence (PRJEB33974) and the karyotype of several starfish species, including *Asterias amurensis*, which were determined by Saotome and Komatsu in 2002 (Saotome and Komatsu, 2002[2]).

The determination and public availability of the genome sequence of *Marthasterias glacialis* provides a superb resource for research on this species and starfish in general. Although *M. glacialis* is less widely distributed in UK waters than *A. rubens*, it has been used for experimental research. For example, the first detailed description of the anatomy of the starfish nervous system was largely based on analysis of specimens of *M. glacialis*, as reported by J.E. Smith in 1937 (Smith, 1937[3]). More recently, Yun *et al.* (2007[4]) reported the identification of four SALMFamide-type neuropeptides isolated from the radial nerve cords of *M. glacialis*.. With the availability of the genome sequence of *M. glacialis* there now exist exciting opportunities for molecular level investigations of many aspects of the biology of *M. glacialis* and comparison with *A. rubens* and other starfish species.

This article provides a detailed description of the methods employed for sequencing and assembly of the *M. glacialis* genome and links to sequence data. I recommend the following corrections/amendments to the article:

1. Throughout the article the authors refer to the “legs” of the starfish as the source of the DNA used for genome sequencing. However, the term “leg” not is used customarily. The correct term is “arm(s)” or “ray(s)”.

2. In the legend of Figure 3 the authors state “Scaffolds labelled Chordata are spurious and assumed to reflect some confusion in the sequence databases from which data is pulled.” I suggest that this issue should be resolved.

3. The name *Marthasterias glacialis* should be in italics throughout the article and in the reference list “*Marthasterias Glacialis*” should be changed to “*Marthasterias glacialis*” in the cited article by Pérez-Portela *et al.* (2010).

4. The article states that the specimen analysed was of unknown sex. Were some of the arms preserved in fixative? It may be possible determine the sex of the specimen by histological analysis of the gonads, which are located at the base of the arms.

5. I don’t understand the scale on the bottom left of Figure 2. The same symbol represents different lengths of DNA (520 M and 38 M).

References

1. Hall M, Kocot K, Baughman K, Fernandez-Valverde S, et al.: The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature*. 2017; **544** (7649): 231-234 Publisher Full Text
2. Saotome K, Komatsu M: Chromosomes of Japanese starfishes.*Zoolog Sci*. 2002; **19** (10): 1095-103 PubMed Abstract | Publisher Full Text
3. Smith JE: II - On the nervous system of the starfish Mathasterias glacialis (L.). *Phil. Trans. R. Soc. Lond.*1937; **227** (542): 111-173 Publisher Full Text | Reference Source
4. Yun SS, Thorndyke MC, Elphick MR: Identification of novel SALMFamide neuropeptides in the starfish Marthasterias glacialis.*Comp Biochem Physiol A Mol Integr Physiol*. 2007; **147** (2): 536-42 PubMed Abstract | Publisher Full Text

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

Partly

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: I collaborated with the Wellcome Sanger Institute in providing the material that was used for sequencing of the Asterias rubens genome, which I refer to in my report. I confirm that this potential conflict of interest did not affect my ability to write an objective and unbiased review of the article.

Reviewer Expertise: I have expertise on starfish biology and analysis of genome sequence data. I

do not have expertise on the technical details of genome sequencing and assembly.

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.