DATA NOTE

ERGA-BGE reference genome of *Diadema setosum:* the Black Longspine Urchin invading the Mediterranean sea

[version 1; peer review: 1 approved, 2 approved with reservations]

Previously titled "BGE-ERGA reference genome of *Diadema setosum:* the Black Longspine Urchin invading the Mediterranean sea"

Katerina Vasileiadou[1], Tereza Manousaki[2], Astrid Böhne[3], Rosa Fernández[4], Nuria Escudero[4], Genoscope Sequencing Team, Alice Moussy[5], Corinne Cruaud[5], Karine Labadie[5], Lola Demirdjian[6], Benjamin Istace[6], Arnaud Couloux[6], Patrick Wincker[6], Pedro H. Oliveira[6], Jean-Marc Aury[6], Rita Monteiro[3]

[1]Institute of Oceanography, Hellenic Centre for Marine Research, Heraklion, Creete, 70014, Greece
[2]Institute of Marine Biology Biotechnology and Aquaculture, Hellenic Centre for Marine Research, Heraklion, Crete, 70014, Greece
[3]Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig Bonn, Adenauerallee 127, Bonn, 53113, Germany
[4]Metazoa Phylogenomics Lab, Institute for Evolutionary Biology (CSIC-UPF). Passeig marítim de la Barceloneta 37-49. 08003, Barcelona, Spain
[5]Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, 91057, France
[6]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, 91057, France

v1 **First published:** 12 Mar 2026, **6**:75
https://doi.org/10.12688/openreseurope.20681.1
Latest published: 12 Mar 2026, **6**:75
https://doi.org/10.12688/openreseurope.20681.1

Abstract

The *Diadema setosum* reference genome is important for understanding the species adaptation to the Mediterranean marine environment, where it has been newly introduced. The species is a Lessepsian migrant, gradually expanding to the eastern Mediterranean basin, driven by rising water temperature. *Diadema setosum* often dominates over native sea urchin species or coexists with them in rocky habitats.

Sea urchins are environment-forming species, as they are intensive grazers responsible for habitat degradation and bottom erosion. A high-quality reference genome could provide valuable insights into the adaptive ability of *D. setosum* populations, supporting better monitoring and conservation efforts. Additionally, the genome will also contribute towards having a record of recently introduced populations to the Mediterranean, allowing researchers to track their evolution over time.

A total of 22 contiguous chromosomal pseudomolecules were

Open Peer Review

Approval Status ? ? ✓

	1	2	3
version 1 12 Mar 2026	? view	? view	✓ view

1. **Abdelmalek Lekired**, University of New Mexico, Albuquerque, USA
2. **Bingpeng Xing**, Third Institute of Oceanography Ministry of Natural Resources, Xiamen, China
3. **Qinggang Xue**, Zhejiang Wanli University, Ningbo, China

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

assembled from the genome sequence. This chromosome-level assembly encompasses 0.91 Gb, composed of 745 contigs and 101 scaffolds, with contig and scaffold N50 values of 2.2 Mb and 39.8 Mb, respectively.

Keywords
Diadema setosum, genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome Project, Diadematidae, Black Longspine Urchin

This article is included in the Horizon Europe gateway.

This article is included in the Genome Reports from the Biodiversity Genomics Europe Project collection.

Corresponding author: Rita Monteiro (R.Monteiro@leibniz-lib.de)

Author roles: Vasileiadou K: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Manousaki T**: Methodology, Supervision, Writing – Review & Editing; **Böhne A**: Funding Acquisition, Supervision, Writing – Review & Editing; **Fernández R**: Funding Acquisition, Methodology, Supervision, Writing – Review & Editing; **Escudero N**: Methodology, Supervision, Validation, Writing – Review & Editing; **Moussy A**: Investigation, Supervision, Writing – Review & Editing; **Cruaud C**: Investigation, Supervision, Writing – Review & Editing; **Labadie K**: Investigation, Supervision, Writing – Review & Editing; **Demirdjian L**: Data Curation, Formal Analysis, Writing – Review & Editing; **Istace B**: Data Curation, Formal Analysis, Writing – Review & Editing; **Couloux A**: Data Curation, Formal Analysis, Writing – Review & Editing; **Wincker P**: Investigation, Supervision, Writing – Review & Editing; **H. Oliveira P**: Investigation, Supervision, Writing – Review & Editing; **Aury JM**: Data Curation, Formal Analysis, Supervision, Writing – Review & Editing; **Monteiro R**: Methodology, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Biodiversity Genomics Europe (Grant no.101059492) is funded by Horizon Europe under the Biodiversity, Circular Economy and Environment call (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. This work was supported by the Genoscope, the Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA), France Génomique (ANR-10-INBS-09-08), and the exploratory research programme 'ATLASea: Atlas of marine genomes' and its targeted project SEQ-Sea (ANR-22-EXAT-0003-SEQ-Sea).
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: Vasileiadou K, Manousaki T, Böhne A *et al.* **ERGA-BGE reference genome of *Diadema setosum:* the Black Longspine Urchin invading the Mediterranean sea [version 1; peer review: 1 approved, 2 approved with reservations]** Open Research Europe 2026, **6**:75 https://doi.org/10.12688/openreseurope.20681.1

First published: 12 Mar 2026, **6**:75 https://doi.org/10.12688/openreseurope.20681.1

Introduction

Diadema setosum or Μακράκανθος Αχινός in Greek, also known as the Black Longspine Urchin, belongs to the Diadematidae family. The phylogeny of the *Diadema* species has shown that *D. setosum* is more divergent than the other *Diadema* species, forming two distinct clades: Clade A, which expands into the Indo-Pacific and Clade B, which is restricted to the Persian Gulf and Red Sea (Lessios, 2000). Thus far, *D. setosum* spreading in the Mediterranean has been found to belong to the second clade (Bronstein et al., 2017). The first record of the species was in 2006 from the coasts of Turkey. Since then, its distribution has expanded to the north-eastern Mediterranean coasts with the last references reporting occurrences up to the south Ionian Sea and the Sea of Marmara, with the abundances observed to increase dramatically since 2018 (Zirler et al., 2023). Thermophilia of the Longspine Urchin favors the organism against the indigenous species, as the water temperature is rising in the Mediterranean due to the shifting of climatic conditions.

Sea urchins are considered habitat-forming species as they are intense algal grazers, which can cause the complete collapse of entire systems. *Diadema setosum* can alter benthic communities by regulating algal growth, which can lead to shifts in larval settlement and development (Zirler et al., 2023). It has significant effects on fisheries, and in cases where it was removed from habitats, biodiversity and fish biomass massively increased (Huseyinoglu et al., 2024). These aspects render the species an important ecosystem regulator; therefore, controlling its populations is of ecological and economic importance. Competition from the indigenous urchin species is forcing *D. setosum* to live on deeper bottoms, while its great size drives species to occupy open spaces, leaving rocks and small crevices free for the indigenous species, which are much smaller in size. The gonad size of the invasive urchin is also greater than that of the Mediterranean urchins. This attribute could favour *D. setosum* to dominate in common habitats.

Invasive *D. setosum* is spreading in the Eastern Mediterranean basin, and although mass mortality events have already been noted (Skouradakis et al., 2024; Dinçtürk et al., 2024), the species attributes seem to support successful establishments. Moreover, overfishing of the indigenous urchin species benefits *D. setosum*, which is not exploited commercially. Therefore, describing a high-quality reference genome will expand our knowledge on traits, evolution, and adaptation of the Longspine Urchin and will contribute towards designing effective conservation strategies to control the species populations.

The generation of this reference resource was coordinated by the European Reference Genome Atlas (ERGA) initiative's Biodiversity Genomics Europe (BGE) project, supporting ERGA's aims of promoting transnational cooperation to promote advances in the application of genomics technologies to protect and restore biodiversity https://www.zotero.org/google-docs/?mS7euC (Mazzoni et al., 2023; https://www.zotero.org/google-docs/?mS7euC).

Materials & Methods

ERGA's sequencing strategy includes Oxford Nanopore Technology (ONT) and/or Pacific Biosciences (PacBio) for long-read sequencing, along with Hi-C sequencing for chromosomal architecture, Illumina Paired-End (PE) for polishing (i.e. recommended for ONT-only assemblies), and RNA sequencing for transcriptomic profiling, to facilitate genome assembly and annotation.

Sample and sampling information

On 31st May 2023, a female adult of *Diadema setosum* was sampled and identified by Katerina Vasileiadou. The species identification through COI barcoding was confirmed by Katerina Vasileiadou. The specimen was collected by hand picking in Vlychia, Heraklion, Crete (Greece), under permission ΥΠΕΝ\ΔΔΔ\34284\1131 from the Ministry for Environment and Energy Secretariat General for Natural Environment & Water Directorate General for Forests & Forest Environment Directorate for Forest Management. The specimen's gonads were snap-frozen immediately after harvesting and stored in liquid nitrogen until DNA extraction.

Vouchering information

Physical reference materials for the sequenced specimen were deposited in the Natural History Museum of Crete https://www.nhmc.uoc.gr/, under the accession number NHMC.65.24.

Frozen reference tissue material of the gonad is available from a proximal individual at the Biobank of the Natural History Museum of Crete https://www.nhmc.uoc.gr/, under the proxy voucher ID NHMC.65.24.

An electronic voucher image of the sequenced individual is available from ERGA's EBI BioImageArchive dataset www.ebi.ac.uk/biostudies/bioimages/studies/S-BIAD1012?query=ERGA under accession ID https://ftp.ebi.ac.uk/biostudies/fire/S-BIAD/012/S-BIAD1012/Files/ERGA/SAMEA114349570_1.jpg.

Data availability

Diadema setosum and the related genomic study were assigned to Tree of Life ID (ToLID) 'eeDiaSeto1', and all sample, sequence, and assembly information are available under the umbrella BioProject PRJEB77220 https://www.ebi.ac.uk/ena/browser/view/PRJEB77220. The sample information is available at the following BioSample accessions: SAMEA114349572, SAMEA114349578, SAMEA114349580. The genome assembly is accessible from ENA under accession number GCA_964275005.1.

Sequencing data produced as part of this project are available from ENA at the following accessions: ERX12733445, ERX12733463, ERX12737198, ERX12737199. Documentation related to the genome assembly and curation can be found in the ERGA Assembly Report (EAR) document available at https://github.com/ERGA-consortium/EARs/tree/main/Assembly_Reports/Diadema_setosum/eeDiaSeto1. Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/31175.

Genetic information

The estimated genome size, based on ancestral taxa, is 1.17 Gb. This is a diploid genome with a haploid number of 22 chromosomes (2n = 11), and unknown sex chromosomes. All information for this species was retrieved from Genomes on a Tree https://www.zotero.org/google-docs/?wmNEAd (Challis et al., 2023).

DNA/RNA processing

DNA was extracted from 200 mg of gonads using a conventional CTAB extraction followed by a commercial purification using Qiagen Genomic tips (QIAGEN, MD, USA). A detailed protocol is available on protocols.io (https://www.protocols.io/view/hmw-dna-extraction-for-long-read-sequencing-using-bp2l694yzlqe/v1). DNA fragment size selection was performed using Short Read Eliminator (PacBio, CA, USA). Quantification was performed using a Qubit dsDNA HS Assay kit (Thermo Fisher Scientific) and integrity was assessed in a FemtoPulse system (Agilent). DNA was stored at 4 ^{o}C until usage.

RNA was extracted from gonads (50 mg) using the RNeasy Plus Universal kit (Qiagen) following the manufacturer's instructions. Residual genomic DNA was removed with 6U of TURBO DNase (2 U/μL) (Thermo Fisher Scientific). Quantification was performed using a Qubit RNA HS Assay kit, and integrity was assessed in a Bioanalyzer system (Agilent). RNA was stored at -80 °C.

Library preparation and sequencing

Long-read DNA libraries were prepared with the SMRTbell prep kit 3.0 following manufacturers' instructions and sequenced on a Revio system (PacBio).

Hi-C libraries were generated from gonads (50 mg) of the same individual using the Arima High Coverage HiC kit (following the Animal Tissues low input protocol v01) and sequenced on a NovaSeq6000 instrument (Illumina) with 2x150 bp read length.

Poly(A) RNA-Seq libraries were constructed using the Illumina Stranded mRNA Prep, Ligation Prep kit (Illumina) and sequenced on the Illumina NovaSeq6000 instrument (Illumina) with 2x150 bp read length.

Genome assembly methods

The genome of *Diadema setosum* was assembled using the Genoscope GALOP pipeline (https://workflowhub.eu/workflows/1200). Briefly, raw PacBio HiFi reads were assembled using Hifiasm v0.19.5-r593. Retained haplotigs were removed using purge_dups v1.2.5 with default parameters and the proposed cutoffs. The purged assembly was scaffolded using YaHS v1.2, and assembled scaffolds were then curated through manual inspection using PretextView v0.2.5 to remove false joins and incorporate sequences not automatically scaffolded into their respective locations within the chromosomal pseudomolecules.

Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. The mitochondrial genome was assembled using Oatk v1.0 and included in the released assembly. Summary analysis of the released assembly was performed using the ERGA-BGE Genome Report ASM Galaxy workflow (https://doi.org/10.48546/workflowhub.workflow.1104.1).

Results

Genome assembly

The genome assembly has a total length of 912,728,646 bp in 101 scaffolds, including the mitogenome (Figures 1 and 2), with a GC content of 38.35%. It features a contig N50 of 2,178,457 bp (L50 = 133) and a scaffold N50 of 39,828,611 bp (L50 = 11). There are 644 gaps, totaling 66,200 kb in cumulative size. The single-copy gene content analysis using the Eukaryota database with BUSCO (Manni et al., 2021) resulted in 98.4% completeness (98.0% single and 0.4% duplicated). 76.9% of reads k-mers were present in the assembly, and the assembly has a base accuracy Quality Value (QV) of 55.6% as calculated by Merqury https://www.zotero.org/google-docs/?8tN3un (Rhie et al., 2020).

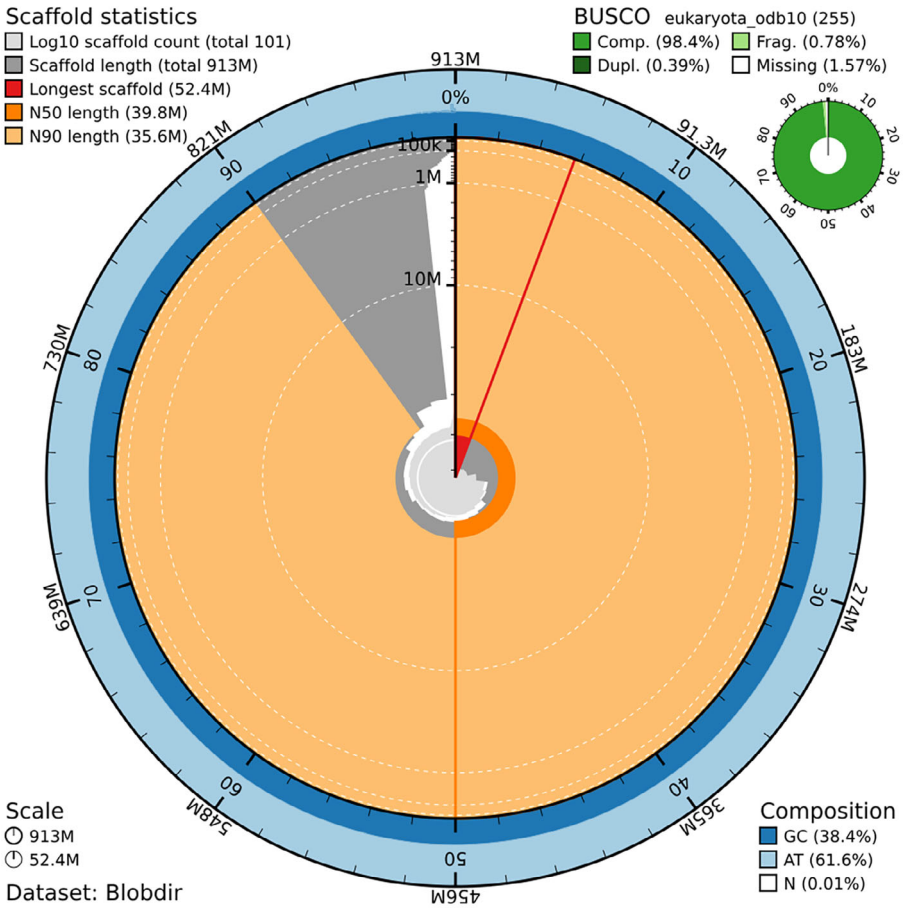

Figure 1. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 912,728,646 bp assembly, including the mitochondrial genome. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (52.4 Mb, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (39.8 Mb and 35.6 Mb), respectively. The pale grey spiral shows the cumulative sequence 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 found in the assembled genome from the Eukaryota database (odb10) is shown on the top right.

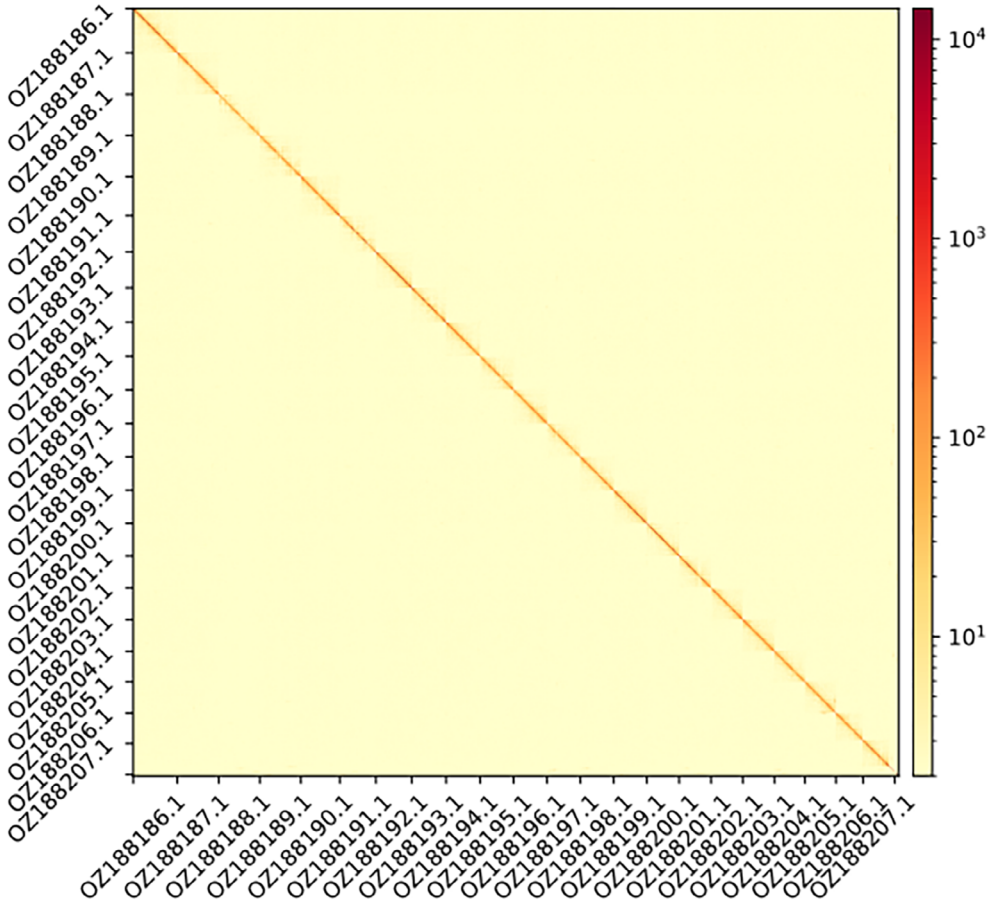

Figure 2. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 150 kb bin size for plotting.

Author contributions

KV and TM coordinated the project, KV collected the species, identified the species, sampled and preserved biological material and provided metadata, RM, AsB, RF and NE provided sampling and metadata support and management, GST extracted DNA, prepared libraries, and performed sequencing under the supervision of AM, CC, KL and PHO; LD, AC, BI and JMA performed genome assembly and curation under the supervision of JMA; RM generated the analysis and report. All authors contributed to the writing, review, and editing of this genome note and read and approved the final version.

Author information

Members of the Genoscope Sequencing Team are listed here: https://doi.org/10.5281/zenodo.14611490.

Acknowledgements

We acknowledge the support of the Freiburg Galaxy Team: Saim Momin and Björn Grüning, Bioinformatics, University of Freiburg (Germany), funded by the German Federal Ministry of Education and Research BMBF grant 031 A538A de. NBI-RBC and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de. NBI Freiburg.

References

Bronstein O, Georgopoulou E, Kroh A: **On the distribution of the invasive long-spined echinoid Diadema setosum and its expansion in the Mediterranean Sea.** *Mar. Ecol. Prog. Ser.* 2017; **583**: 163–178.
Publisher Full Text

Challis R, Kumar S, Sotero-Caio C, *et al.*: **Genomes on a Tree (GoaT): A versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life [version 1; peer review: 2 approved].** *Wellcome Open Research.* 2023; **8**: 24.
Publisher Full Text

Dinçtürk E, Öndes F, Alan V, *et al.*: **Mass Mortality of the Invasive Sea Urchin Diadema setosum in Türkiye, Eastern Mediterranean Possibly Reveals Vibrio Bacteria Infection.** *Mar. Ecol.* 2024; **45**(6): e12837.
Publisher Full Text

Huseyinoglu MF, Jimenez C, Resaikos V, *et al.*: **Assisted recovery of mediterranean native fish species by removal of two established invasive alien species, Pterois miles and Diadema setosum.** *Estuar. Coast. Shelf Sci.* 2024; **304**: 108826.
Publisher Full Text

Lessios HA: **Molecular phylogeny of Diadema: Systematic implications.** *Echinoderms.* 2000; **2000**.

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO: Assessing genomic data quality and beyond.** *Current Protocols.* 2021; **1**: e323.
Publisher Full Text

Mazzoni C, Ciofi C, Waterhouse R: **Biodiversity: An atlas of European reference genomes.** *Nature.* 2023; **619**: 252–252.
Publisher Full Text

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**: 245.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Skouradakis G, Vernadou E, Koulouri P, *et al.*: **Mass mortality of the invasive echinoid Diadema setosum (Leske, 1778) in Crete, East Mediterranean Sea. Mediterranean Marine.** *Science.* 2024; **25**(2): 480–483.
Publisher Full Text

Zirler R, Leck LA, Farkash TF, *et al.*: **Gaining a (tube) foothold–trends and status following two decades of the long-spined echinoid Diadema setosum (Leske, 1778) invasion to the Mediterranean Sea.** *Front. Mar. Sci.* 2023; **10**: 1152584.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ? ? ✓

Version 1

Reviewer Report 13 June 2026

https://doi.org/10.21956/openreseurope.22371.r75610

✓ **Qinggang Xue**
Zhejiang Wanli University, Ningbo, Zhejiang, China

This high-quality chromosome-level genome assembly of the invasive *Diadema setosum* fills a critical gap in understanding Mediterranean Lessepsian migrant adaptation. With strong BUSCO completeness and EBP-aligned metrics, it supports conservation efforts—minor clarifications on assembly details will enhance its utility. Recommended for Indexing

Reviewer Comments

1. The manuscript states a diploid chromosome number of 2n=11 but later reports 22 chromosomal pseudomolecules. Resolve this contradiction (likely a typo; should be 2n=22) and confirm with cytogenetic evidence or ancestral taxon comparisons.
2. The Materials and Methods do not specify key parameters for the GALOP pipeline (e.g., Hifiasm assembly settings, purge_dups cutoff values). Supplement with detailed parameters to ensure reproducibility of the assembly workflow.
3. The RNA-seq data are mentioned for transcriptomic profiling but not utilized for genome annotation in this note. Clarify when the annotated genome will be released and whether preliminary analyses of adaptive genes (e.g., temperature tolerance, grazing-related loci) were conducted.
4. The Results report 76.9% k-mer completeness, which is lower than typical standards for reference genomes. Discuss potential reasons (e.g., repeat content, heterozygosity) and how this impacts the assembly's utility for downstream analyses.
5. The study focuses on a single female specimen—note whether sex-specific genomic features (e.g., sex chromosomes) were investigated and if sampling a single individual limits insights into population-level adaptive variation.
6. The Hi-C contact map (Figure 2) lacks a clear legend for the logarithmic heatmap scale and chromosome labeling. Revise the figure to include these details and improve readability.
7. The Introduction highlights *D. setosum*'s ecological impacts but does not compare its genome to native Mediterranean urchin species. Add a brief comparative context to emphasize unique adaptive traits of the invader.

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: marine biology, genetics, evolution

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 08 April 2026

https://doi.org/10.21956/openreseurope.22371.r71383

?

Bingpeng Xing
Third Institute of Oceanography Ministry of Natural Resources, Xiamen, China

The manuscript reports a chromosome-level reference genome for Diadema setosum. As an important invasive sea urchin in the Mediterranean, this species has clear ecological relevance, so the genomic resource itself is valuable. The authors used PacBio HiFi and Hi-C for genome assembly, which is an appropriate strategy for generating a high-quality reference genome.

The main issue at present is that several key methodological details are missing or described too briefly, making the assembly workflow insufficiently clear and difficult to reproduce. For example, the manuscript mentions ONT, PacBio, Illumina, and RNA-seq in the overall sequencing strategy, but it does not clarify whether ONT data were actually used for this species, and if so, at which step. Although hifiasm is listed with its version, key parameters are not provided. The description of purge_dups is also too vague, as the actual cutoff values are not reported. I suggest that the authors describe the assembly workflow more clearly, step by step, and provide the relevant software versions and key parameter settings.

A second major issue is that RNA-seq and annotation are mentioned in the Methods, but no corresponding annotation results are presented in the Results. If genome annotation was indeed carried out, the manuscript should at least report the number of predicted genes and provide a brief summary of the annotation. If annotation is not part of the formal scope of this paper, then

the wording in the Methods should be revised accordingly. As written, the manuscript gives the impression that annotation was completed, but no results are actually shown.

In addition, there are inconsistencies in the genome quality assessment. Different BUSCO datasets and versions are mentioned in different parts of the manuscript, including Metazoa_odb12, Eukaryota, and odb10, and these should be made consistent throughout. Another concern is that the reported genome size of 1.17 Gb appears to come from previously available database information rather than from an independent estimate based on the present sample. Given that the final assembly size is only about 0.91 Gb, I recommend adding a k-mer–based estimate of genome size, heterozygosity, and repeat content, and briefly discussing the discrepancy.

Contamination assessment is also a relatively weak point in the current manuscript. It is not clearly stated whether contamination screening was performed, how potentially contaminant contigs were identified, or whether any contigs were removed.

In addition, the manuscript states that the mitochondrial genome was assembled and included in the final release, but no basic information is provided in the main text. At minimum, the authors should report the mitochondrial genome size, gene content, and annotation method.

There are also a few issues of accuracy and presentation that should be corrected. For example, the statement “a diploid genome with a haploid number of 22 chromosomes (2n = 11)” is biologically incorrect and should be revised. The Merqury result should also be written as QV 55.6 rather than 55.6%. More generally, the manuscript would benefit from a careful check of wording and formatting to ensure consistency in terminology, software/database versions, and reference formatting throughout.

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?

Partly

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

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: genome，marine biology，DNA，phylogenetic analysis

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 30 March 2026

https://doi.org/10.21956/openreseurope.22371.r71388

?

Abdelmalek Lekired
University of New Mexico, Albuquerque, New Mexico, USA

Reviewer Report
I would like to thank the authors for submitting their manuscript entitled *"Reference genome of Diadema setosum: the Black Longspine Urchin invading the Mediterranean Sea."* The generation of a chromosome-level genome for this ecologically important and invasive species represents a valuable contribution to the field. The use of HiFi long reads combined with Hi-C scaffolding is appropriate and has strong potential to produce a high-quality reference genome.
However, several methodological details are currently missing, and some analyses require clarification or improvement to meet current standards for genome assembly and annotation studies. I outline below my major comments and suggestions.

Abstract
The abstract would benefit from including key assembly statistics to better reflect the quality of the genome. I recommend adding:

- Genome size
- Contig/scaffold N50
- BUSCO completeness scores

Including these metrics will provide readers with a clearer overview of the assembly quality.

Materials and Methods
The authors state that ONT, PacBio, Illumina, and RNA-seq data were used for assembly and annotation. However, several important details are missing:

1. The role of ONT reads in the assembly is not described. It is unclear whether they were used for scaffolding, gap filling, polishing, or not used at all. This must be clarified.
2. The number of RNA-seq samples, their origin (tissues, developmental stages), and sequencing depth are not provided. These details are essential for evaluating the annotation quality.
3. Annotation methods are mentioned, but no corresponding results are presented. Either the annotation results should be included, or the annotation section should be removed or revised.
4. The authors mention using the BUSCO dataset *Metazoa_odb12*, but the results are not reported anywhere in the manuscript. This inconsistency must be addressed.

General Concerns

1. The genome size is reported as 1.17 Gb based on previously published data. The authors should estimate the genome size independently using k-mer–based approaches (e.g., GenomeScope2) and report heterozygosity and repeat content.
2. The statement:
   "This is a diploid genome with a haploid number of 22 chromosomes (2n = 11)"

is incorrect and biologically inconsistent. If 2n = 22, then n = 11. This should be corrected.

Genome Assembly Methods

This section lacks sufficient detail and is currently not reproducible.

- The authors mention the use of ONT reads but do not describe how they were incorporated into the assembly workflow.
- If *hifiasm* was used, relevant parameters should be reported (e.g., options related to duplicate purging such as -l 0 or -l 3, if applicable).
- The statement:
  "Retained haplotigs were removed using purge_dups v1.2.5 with default parameters and the proposed cutoffs"
  is unclear. The authors should clearly define what is meant by "retained haplotigs" and specify the cutoff values used.

Overall, the assembly pipeline should be described step-by-step with software versions and parameters to ensure reproducibility.

Results

Several inconsistencies and missing details need to be addressed:

1. Two BUSCO versions (*odb10* and *odb12*) are mentioned. The authors should clarify which dataset was used and ensure consistency throughout the manuscript.
2. The results should include BUSCO evaluation using an appropriate lineage dataset (preferably Metazoa). If a eukaryotic dataset was used, the rationale should be explained.
3. The manuscript does not address potential contamination:
   - Were contaminants assessed (e.g., using BlobToolKit or similar tools)?
   - If so, how were contaminant contigs identified and removed?
4. Regarding the blob plot:
   - The dataset labeling should reflect the genome name rather than generic directory names (e.g., "BlobDir").
   - The figure and legend should be revised for clarity.
5. The snail plot:
   - It should be clearly stated whether this plot corresponds to the final chromosome-level assembly.
6. The mitochondrial genome assembly is mentioned, but no details are provided. The authors should report (if available):
   - Genome size
   - Number of protein-coding genes
   - rRNA and tRNA content
   - Annotation approach used

Conclusion

This study presents a valuable genomic resource for *Diadema setosum*, but the manuscript requires **substantial revision** to improve clarity, reproducibility, and completeness. Addressing the issues outlined above will significantly strengthen the manuscript and bring it in line with current standards for genome assembly and annotation studies.

References

1. Lekired A, Mudge J, Laidemitt M, Mutuku M, et al.: Whole-genome sequencing and genome-wide transcriptome profiling of the freshwater planorbid snail Bulinus ugandae (Mollusca: Gastropoda), a Nilotic bulinine refractory to Schistosoma haematobium. *BMC Genomics*. 2025; **26**

(1). Publisher Full Text

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?
No

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Computational genomics and bioinformatics, with specialization in de novo genome assembly (long-read and illumina short reads), genome curation, gene prediction and functional annotation, transcriptome assembly, RNA-Seq analysis, ploidy inference, and genome quality assessment (e.g., BUSCO, k-mer analysis). My research primarily focuses on non-model organisms, including metazoan genomes.

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.