Chromosomal-level genome assembly of the long-spined sea urchin *Diadema setosum* (Leske, 1778)

Hong Kong Biodiversity Genomics Consortium

Project Coordinator and Co-Principal Investigators: Jerome H.L. Hui[1], Ting Fung Chan[2], Leo L. Chan[3], Siu Gin Cheung[4], Chi Chiu Cheang[5,6], James K.H. Fang[7], Juan D. Gaitan-Espitia[8], Stanley C.K. Lau[9], Yik Hei Sung[10,11], Chris K.C. Wong[12], Kevin Y.L. Yip[13,14], Yingying Wei[15]

DNA extraction, library preparation and sequencing: Wai Lok So[1]

Genome assembly and gene model prediction: Wenyan Nong[1]

Sample collector, animal culture and logistics: Apple P.Y. Chui[16], Thomas H.W. Fong[16], Ho Yin Yip[1]

1. School of Life Sciences, Simon F.S. Li Marine Science Laboratory, State Key Laboratory of Agrobiotechnology, Institute of Environment, Energy and Sustainability, The Chinese University of Hong Kong, Hong Kong, China
2. School of Life Sciences, State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Hong Kong SAR, China
3. State Key Laboratory of Marine Pollution and Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR, China
4. State Key Laboratory of Marine Pollution and Department of Chemistry, City University of Hong Kong, Hong Kong SAR, China
5. Department of Science and Environmental Studies, The Education University of Hong Kong, Hong Kong SAR, China
6. EcoEdu PEI, Charlottetown, PE, C1A 4B7, Canada
7. Department of Food Science and Nutrition, Research Institute for Future Food, and State Key Laboratory of Marine Pollution, The Hong Kong Polytechnic University, Hong Kong SAR, China
8. The Swire Institute of Marine Science and School of Biological Sciences, The University of Hong Kong, Hong Kong SAR, China
9. Department of Ocean Science, The Hong Kong University of Science and Technology, Hong Kong SAR, China
10. Science Unit, Lingnan University, Hong Kong SAR, China
11. School of Allied Health Sciences, University of Suffolk, Ipswich, IP4 1QJ, UK
12. Croucher Institute for Environmental Sciences, and Department of Biology, Hong Kong Baptist University, Hong Kong SAR, China
13. Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China
14. Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA.
15. Department of Statistics, The Chinese University of Hong Kong, Hong Kong SAR, China
16. School of Life Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China

Correspondence on behalf of the consortium: jeromehui@cuhk.edu.hk

Abstract

The long-spined sea urchin *Diadema setosum* is an algal and coral feeder widely distributed in the Indo-Pacific and can cause severe bioerosion on the reef community. Nevertheless, the lack of genomic information has hindered the study its ecology and evolution. Here, we report the chromosomal-level genome (885.8 Mb) of the long-spined sea urchin *D. setosum* using a combination of PacBio long-read sequencing and Omni-C scaffolding technology. The assembled genome contained scaffold N50 length of 38.3 Mb, 98.1 % of BUSCO (Geno, metazoa_odb10) genes, and with 98.6% of the sequences anchored to 22 pseudo-molecules/chromosomes. A total of 27,478 genes including 23,030 protein-coding genes were annotated. The high-quality genome of *D. setosum* presented here provides a significant resource for further understanding on the ecological and evolutionary studies of this coral reef associated sea urchin.

Introduction

Similar to other echinoderms, sea urchins lack the vertebral column and can metamorphose from juvenile bilateral swimming larvae into radial symmetrical adults (Li et al., 2020; Lowe et al., 2015). Owing to its critical phylogenetic position, sea urchin offers the understanding of how deuterostomes evolved (e.g. Odekunle et al., 2019; Yañez-Guerra et al., 2020; Sakai et al., 2020; Logeswaran et al. 2021; Zheng et al., 2022). To date, 16 sea urchin genomes are available according to the data presented on NCBI (25 October 2023) and only five of them (all in the order Camarodonta) are assembled in chromosomal level (*Lytechinus variegatus* (Davidson et al., 2020), *Lytechinus pictus* (Warner et al., 2021), *Heliocidaris erythrogramma* (Davidson et al., 2022), *Heliocidaris tuberculata* (Davidson et al., 2022) and *Echinometra lucunter* (Davidson et al., 2023).

Diadema setosum (Leske, 1778) in the order Diadematoida, commonly known as the porcupine or long-spined sea urchin, is considered as one of the oldest known extant species in the genus *Diadema* (Coppard & Campbell, 2006). *D. setosum* displays features of typical sea urchin, including a dorso-ventrally compressed body and equipped with particularly long, brittle and hollow spines that are mildly venomous (Bilecenoğlu et al., 2019; Voulgaris et al., 2021). This species can be easily differentiated from other *Diadema* species by the presence of five distinctive white dots at the aboral side around the anal pore between the ambulacral grooves (Figure 1A). Sexually matured individuals have been documented to have an average weight from 35 to 80 g and an average test size ranges from 7 to 8 cm in diameter and approximately 4 cm in height (Alsaffar & Lone, 2000; Coppard & Campbell, 2006). Due to its high invasiveness to localities beyond its natural range, *D. setosum* is now widely distributed in the tropical regions throughout the Indo-Pacific basin and can now be found latitudinally from Japan to Africa and longitudinally from the Red Sea to Australia (Öndes et al., 2022).

D. setosum can live up as deep as 70 m below the sea level and is usually reef-associated (Lane et al., 2001). It is a prolific grazer that feeds on the macroalgae that can be found on the surface of various substrata, as well as the algae that are associated with the coral skeleton (Bak, 1994; Carreiro-Silva and McClanahan, 2001). Whilst a normal level of grazing eliminates competitive algae and can potentially offer a more suitable environments for coral settlement and development, overgrazing results in bioerosion, which thus in turn deteriorates the reef ecosystem and a reduction in the complexity of coral community (Carpenter and Edmunds, 2006; Dumont et al., 2013). Furthermore, overpopulated sea

urchins can reduce coral recruitment and the growth of juvenile coral can be hindered (Bak and van Eys, 1975; Sammarco, 1980; Davies and Vize, 2008).

Context

Here, we report a high-quality genome assembly of *D. setosum* in the order Diadematoida and family Diadematidae.

Methods

Collection and husbandry of samples

The long-spined sea urchin, *Diadema setosum*, was collected at the coastal area of the Tolo Habour in Hong Kong in November 2022. The animals were maintained at 35 ppt artificial sea water alive at 23°C until DNA and RNA isolation. Animals were fed with frozen clams or shrimps once a week.

Isolation of high molecular weight genomic DNA, quantification, and qualification

High molecular weight genomic DNA was isolated from a single individual. The urchin was first removed from the culture and the test was opened with a pair of scissors. The internal tissue, except the gut, was snap-frozen in liquid nitrogen and ground to fine powder. DNA extraction was performed with the Qiagen MagAttract HMW kit (Qiagen Cat. No. 67563) following the manufacturer's protocol. In brief, 1 g of powdered sample was put in a microcentrifuge tube together with 200 µl 1X PBS and RNase A, Proteinase K, and Buffer AL were added to the tube subsequently. The mixture was incubated at room temperature (~22 °C) for 3 hours. The sample was then eventually eluted with 120 µl of elution buffer (PacBio Ref. No. 101-633-500). Throughout the extraction progress, wide-bore tips were used whenever DNA was transferred. The eluted sample was quantified by the Qubit® Fluorometer and Qubit™ dsDNA HS and BR Assay Kits (Invitrogen™ Cat. No. Q32851). In total, 10 µg of DNA was collected for subsequent steps. The purity of the sample was examined by the NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer, with the standard A260/A280: ~1.8 and A260/A230: >2.0. The quality and the fragment distribution of the isolated genomic DNA was examined by the overnight pulse-field gel electrophoresis, together with three DNA markers (λ-Hind III digest, Takara Cat. No. 3403; DL15,000 DNA Marker, Takara Cat. No. 3582A and CHEF DNA Size Standard-8-48 kb Ladder, Cat. No. 170-3707). The DNA was then diluted in elution buffer to prepare a 300 ng solution for gel electrophoresis and the electrophoresis profile was set as follow: 5k as the lower end and 100k as the higher end for the molecular weight; Gradient=6.0V/cm; Run time = 15 h:16 min; Included angle = 120 °; Int. Sw. Tm = 22 s; Fin. Sw. Tm = 0.53 s; Ramping factor: a = Linear. The gel was run in 1.0% PFC agarose in 0.5X TBE buffer at 14 °C.

DNA shearing, library preparation and sequencing

10 µg of *D. setosum* DNA in 120 µl elution buffer was transferred to a g-tube (Covaris Part No. 520079). The sample was then proceeded to 6 passes of centrifugation with 2,000 xg for 2 min. The sheared DNA was collected with a 2 ml DNA LoBind® Tube (Eppendorf Cat. No. 022431048) at 4 °C until library preparation. Overnight pulse-field gel electrophoresis was performed to examine the fragment distribution of the sheared DNA, with the same electrophoresis profile as described in the last section.

A SMRTbell library was then constructed with the SMRTbell® prep kit 3.0 (PacBio Ref. No. 102-141-700), following the manufacturer's protocol. In brief, the sheared DNA was

first subjected to DNA repair, and both ends of each DNA strand was polished and tailed with an A-overhang. Ligation of T-overhang SMRTbell adapters was then performed and the SMRTbell library was purified with SMRTbell® cleanup beads (PacBio Ref. No. 102158-300). The concentration and size of the library were examined with the Qubit® Fluorometer and Qubit™ dsDNA HS and BR Assay Kits (Invitrogen™ Cat. No. Q32851), and an overnight pulse-field gel electrophoresis, respectively. A nuclease treatment step was performed afterwards to remove any non-SMRTbell structures in the library and a final size-selection step was performed to remove the short fragments in the library with 35% AMPure PB beads.

The Sequel® II binding kit 3.2 (PacBio Ref. No. 102-194-100) was used for final preparation of sequencing. In brief, Sequel II primer 3.2 and Sequel II DNA polymerase 2.2 were annealed and bound to the SMRTbell structures in the library. The library then was loaded at an on-plate concentration of 90 pM using the diffusion loading mode. The sequencing was performed on the Sequel IIe System with an internal control provided in the binding kit. The sequencing was prepared and run in 30-hour movies, with 120 min pre-extension. The movie was captured by the software SMRT Link v11.0 (PacBio) and HiFi reads are generated and collected for further analysis. In total, one SMRT cell was used in the sequencing. Details of the sequencing data are listed in Supplementary Information 1.

Omni-C library preparation and sequencing

An Omni-C library was constructed using the Dovetail® Omni-C® Library Preparation Kit (Dovetail Cat. No. 21005) according to the manufacturer's instructions. In brief, 60 mg of frozen powered tissue sample was added into 1 mL 1X PBS, where the genomic DNA was crosslinked with formaldehyde and the DNA was then digested with endonuclease DNase I. Subsequently, the concentration and fragment size of the digested sample was validated by the Qubit® Fluorometer and Qubit™ dsDNA HS and BR Assay Kits (Invitrogen™ Cat. No. Q32851), and the TapeStation D5000 HS ScreenTape, respectively. Afterwards, both ends of the DNA were polished and a biotinylated bridge adaptor was ligated at 22 °C for 30 min. Proximity ligation between crosslinked DNA fragments was then performed at 22 °C for 1 hour, followed by the reverse crosslinked of DNA, and it is then purified with SPRIselect™ Beads (Beckman Coulter Product No. B23317).

End repair and adapter ligation were performed with the Dovetail™ Library Module for Illumina (Dovetail Cat. No. 21004). In brief, DNA was tailed with an A-overhang, and ligate with Illumina-compatible adapters at 20 °C for 15 min. The Omni-C library was then sheared into small fragments with USER Enzyme Mix and purified with SPRIselect™ Beads. Subsequently, DNA fragments were isolated with Streptavidin Beads. Universal and Index PCR Primers from the Dovetail™ Primer Set for Illumina (Dovetail Cat. No. 25005) were used to amplify the DNA library. A final size selection step was completed with SPRIselect™ Beads with DNA fragments ranging between 350 bp and 1000 bp only. The concentration and fragment size of the sequencing library was assessed by the Qubit® Fluorometer and Qubit™ dsDNA HS and BR Assay Kits, and the TapeStation D5000 HS ScreenTape, respectively. Qualified library was sequenced on an Illumina HiSeq-PE150 platform. Details of the sequencing data are listed in Supplementary Information 1.

RNA extraction and transcriptome sequencing

Total RNA from the internal tissues were isolated from one individual using the TRIzol reagent (Invitrogen) following the manufacturer's protocol. The quality of the extracted RNA was validated with the NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer (Thermo Scientific™ Cat. No. ND-ONE-W) and 1% agarose gel electrophoresis. The qualified samples were sent to Novogene Co. Ltd (Hong Kong, China) for polyA selected RNA sequencing library construction using the TruSeq RNA Sample Prep Kit v2 (Illumina Cat. No. RS-122-2001), and 150 bp paired-end (PE) sequencing. Agilent 2100 Bioanalyser (Agilent DNA 1000 Reagents) was used to measure the insert size and concentration of the final library. Details of the sequencing data are shown in Supplementary Information 1.

Genome assembly and gene model prediction

De novo genome assembly was completed using Hifiasm (Cheng et al., 2021), and the Hifiasm output assembly was BLAST to the NT database and the BLAST output was used as input for BlobTools (v1.1.1) (Laetsch & Blaxter, 2017) to validate and remove any possible contaminations (Additional Information 1). Haplotypic duplications were detected and removed using purge_dups according to the depth of HiFi reads (Guan et al., 2020). Proximity ligation data from the Omni-C library were used to scaffold the PacBio genome by YaHS (Zhou et al., 2022). A kmer-based statistical method was performed to estimate the heterozygosity, while repeat content and size were analyzed by Jellyfish (Marçais & Kingsford, 2011) and GenomeScope (http://qb.cshl.edu/genomescope/genomescope2.0/) (Ranallo-Benavidez et al., 2020). Transposable elements (TEs) were annotated, using the automated Earl Grey TE annotation pipeline (version 1.2, https://github.com/TobyBaril/EarlGrey) (Baril et al., 2022). Mitochondrial genome was assembled using MitoHiFi (v2.2, https://github.com/marcelauliano/MitoHiFi) (Allio et al., 2020).

For gene model prediction, RNA sequencing data was first processed with Trimmomatic (Bolger, Lohse & Usadel, 2014) and transformed to transcripts using genome-guided Trinity (Grabherr et al., 2011). AUGUSTUS (Stanke et al., 2006) was trained using BUSCO (Manni et al., 2021), while GeneMark-ES (Ter-Hovhannisyan et al 2008) was used for *ab initio* gene prediction. Gene models were then predicted by funannotate using the parameters "--repeats2evm --protein_evidence uniprot_sprot.fasta --genemark_mode ET --optimize_augustus --organism other --max_intronlen 350000". The gene models from several prediction sources including GeneMark, high-quality Augustus predictions (HiQ), pasa, Augustus, GlimmerHM and snap were passed to Evidence Modeler to generate the annotation files. PASA was employed to update the EVM consensus predictions. In addition, UTR annotations were added and models for alternatively spliced isoforms were created.

Results

A total of 18.5 Gb of HiFi bases were generated with an average HiFi read length of 8,449 bp with 21x data coverage (Supplementary Information 1). The assembled genome size was 885.8 Mb, with 101 scaffolds and a scaffold N50 of 38.3 Mb in 11 scaffolds and the complete BUSCO estimation to be 98.1% (metazoa_odb10) (Figure 1B; Table 1). By incorporating 67.5 Gb Omni-C data, the assembly anchored 98.6% of the scaffolds into 22 pseudochromosomes, which matches the karyotype of *D. setosum* (2n = 44) (Shingaki & Uehara, 1984) (Figure 1C; Table 2). The assembled *D. setosum* genome size is comparable to

other published sea urchin genomes (Davidson et al., 2020; Warner et al., 2021; Davidson et al., 2022; Davidson et al., 2023) and to the estimated size of 804 Mb by GenomeScope with a 2.11% heterozygosity rate (Figure 1D; Supplementary Information 2). Moreover, telomeric repeats were identified in 16 out of 22 pseudochromosomes (Table 3).

Total RNA sequencing data was also obtained from a single *D. setosum* individual. The final assembled transcriptome contained 135,063 transcripts, with 113,391 Trinity annotated genes (with an average length of 838 bp and a N50 length of 1,456 bp), and was used to perform gene model prediction. A total of 27,478 gene models were generated with 23,030 predicted protein-coding genes, with a mean of 483 bp coding sequence length (Figure 1B; Table 1).

For repeat elements, a total repetitive content of 36.98% was identified in the assembled genome including 25.87% unclassified elements (Figure 1E; Table 4). Among the known repeats, DNA is the most abundant (4.18%), followed by LINE elements (3.64%) and LTR (1.92%), whereas Rolling Circle, SINE, Penelope and other are only present in low proportions (Rolling Circle: 0.92%, SINE: 0.23%, Penelope: 0.17%, other: 0.04%).

Conclusion and future perspectives

In sum, the high quality chromosomal-level genome assembly of *D. setosum* presented here provides further understanding of the ecology and evolution of sea urchins.

Data Records

The final genome assembly were deposited to NCBI under accession numbers GCA_033980235.1 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_033980235.1/). The raw reads generated in this study were submitted to the NCBI database under the BioProject accessions PRJNA973839 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA973839). The genome and genome annotation files, as well as an Additional Table and an Additional Information were also deposited in figshare (https://doi.org/10.6084/m9.figshare.24408319).

Data validation and quality control

Quality check of samples in DNA extraction and PacBio library preparation were processed with NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer, Qubit® Fluorometer, and overnight pulse-field gel electrophoresis. The Omni-C library was subjected to quality check by Qubit® Fluorometer and TapeStation D5000 HS ScreenTape.

For genome assembly, the validation of contamination scaffolds from the Hifiasm output was done by searching at the NT database through BLAST, from which the resultant output was analysed by BlobTools (v1.1.1) (Laetsch & Blaxter, 2017) (Supplementary Information 3). Furthermore, a kmer-based statistical approach was performed to estimate the genome heterozygosity. The repeat content and their size were estimated by Jellyfish (Marçais & Kingsford, 2011) and GenomeScope (Figure 1E and Additional Table 5) (http://qb.cshl.edu/genomescope/genomescope2.0/) (Ranallo-Benavidez et al., 2020). Benchmarking Universal Single-Copy Orthologs (BUSCO, v5.5.0) (Manni et al 2021) was run to

evaluate the completeness of the genome assembly and gene annotation with the metazoan dataset (metazoa_odb10).

Author' contribution

JHLH, TFC, LLC, SGC, CCC, JKHF, JDG, SCKL, YHS, CKCW, KYLY and YW conceived and supervised the study; APYC and THWF collected the sea urchin samples; HYY maintained the animal culture; WLS performed DNA extraction, library preparation and genome sequencing; WN carried out genome assembly and gene model prediction.

Competing interest

The authors declare that they do not have competing interests.

Funding

This work was funded and supported by the Hong Kong Research Grant Council Collaborative Research Fund (C4015-20EF), CUHK Strategic Seed Funding for Collaborative Research Scheme (3133356) and CUHK Group Research Scheme (3110154).

References

1. Coppard, S. E., & Campbell, A. C. Taxonomic significance of test morphology in the echinoid genera *Diadema* Gray, 1825 and *Echinothrix* Peters, 1853 (Echinodermata). *Zoosystema*, **28**, 93-112 (2006).
2. Bilecenoğlu, M., Baki Yokeş, M., & Draman, M. The invasive sea urchin *Diadema setosum* provides shelter for coastal fish–first observations from the Mediterranean Sea. *Zool Middle East*, **65**, 183-185 (2019).
3. Voulgaris, K., Varkoulis, A., Zaoutsos, S., Stratakis, A., & Vafidis, D. Mechanical defensive adaptations of three Mediterranean sea urchin species. *Ecol. Evol.*, **11**, 17734-17743 (2021).
4. Alsaffar, A. H., & Lone, K. P. Reproductive cycles of *Diadema setosum* and *Echinometra mathaei* (Echinoidea: echinodermata) from Kuwait (northern Arabian Gulf). *Bull. Mar. Sci.*, **67**, 845-856 (2000).
5. Öndes, F., Alan, V., Kaiser, M. J., & Güçlüsoy, H. Spatial distribution and density of the invasive sea urchin *Diadema setosum* in Turkey (eastern Mediterranean). *Mar. Ecol.*, **43**, e12724 (2022).
6. Lane, D. J., Marsh, L. M., VandenSpiegel, D., & Rowe, F. W. Echinoderm fauna of the South China Sea: an inventory and analysis of distribution patterns. *Raffles Bull. Zool.*, **48**, 459-494 (2001).
7. Bak, R. P. M. Sea urchin bioerosion on coral reefs: place in the carbonate budget and relevant variables. *Coral reefs*, **13**, 99-103 (1994).
8. Carreiro-Silva, M., & McClanahan, T. R. Echinoid bioerosion and herbivory on Kenyan coral reefs: the role of protection from fishing. *J. Exp. Mar. Biol. Ecol.*, **262**, 133-153 (2001).
9. Carpenter, R. C., & Edmunds, P. J. Local and regional scale recovery of *Diadema* promotes recruitment of scleractinian corals. *Ecol. Lett.*, **9**, 271-280 (2006).

10. Dumont, C. P., Lau, D. C., Astudillo, J. C., Fong, K. F., Chak, S. T., & Qiu, J. W. Coral bioerosion by the sea urchin *Diadema setosum* in Hong Kong: susceptibility of different coral species. *J. Exp. Mar. Biol. Ecol.*, **441**, 71-79 (2013).
11. Bak, R. P., & van Eys, G. Predation of the sea urchin *Diadema antillarum* Philippi on living coral. *Oecologia*, **20**, 111-115 (1975).
12. Sammarco, P. W. Diadema and its relationship to coral spat mortality: grazing, competition, and biological disturbance. *J. Exp. Mar. Biol. Ecol.*, **45**, 245-272 (1980).
13. Davies, S. W., & Vize, P. D. Effects of herbivore grazing on juvenile coral growth in the Gulf of Mexico. In The 11th International Coral Reef Symposium, Florida, USA (Vol. 1, pp. 1214-1218). (2008).
14. Lowe, C. J., Clarke, D. N., Medeiros, D. M., Rokhsar, D. S., & Gerhart, J. The deuterostome context of chordate origins. *Nature*, **520**, 456-465 (2015).
15. Odekunle, E. A., Semmens, D. C., Martynyuk, N., Tinoco, A. B., Garewal, A. K., Patel, R. R. et al. Ancient role of vasopressin/oxytocin-type neuropeptides as regulators of feeding revealed in an echinoderm. *BMC Biol.*, **17**, 1-23 (2019).
16. Yañez-Guerra, L. A., Zhong, X., Moghul, I., Butts, T., Zampronio, C. G., Jones, A. M. et al. Echinoderms provide missing link in the evolution of PrRP/sNPF-type neuropeptide signalling. *Elife*, **9**, e57640. (2020).
17. Sakai, T., Yamamoto, T., Matsubara, S., Kawada, T., & Satake, H. Invertebrate gonadotropin-releasing hormone receptor signaling and its relevant biological actions. *Int. J. Mol. Sci.*, **21**, 8544 (2020).
18. Logeswaran, D., Li, Y., Podlevsky, J. D., & Chen, J. J. L. Monophyletic origin and divergent evolution of animal telomerase RNA. *Mol. Biol. Evol.*, **38**, 215-228 (2021).
19. Zheng, Y., Cong, X., Liu, H., Wang, Y., Storey, K. B., & Chen, M. Nervous System Development and Neuropeptides Characterization in Embryo and Larva: Insights from a Non-Chordate Deuterostome, the Sea Cucumber *Apostichopus japonicus*. *Biol.*, **11**, 1538. (2022).
20. Davidson, P. L., Guo, H., Wang, L., Berrio, A., Zhang, H., Chang, Y. et al. Chromosomal-level genome assembly of the sea urchin Lytechinus variegatus substantially improves functional genomic analyses. *Genome Biol. Evol.*, **12**, 1080-1086 (2020).
21. Warner, J. F., Lord, J. W., Schreiter, S. A., Nesbit, K. T., Hamdoun, A., & Lyons, D. C. Chromosomal-level genome assembly of the painted sea urchin *Lytechinus pictus*: A genetically enabled model system for cell biology and embryonic development. *Genome Biol. Evol.*, **13**, evab061 (2021).
22. Davidson, P. L., Guo, H., Swart, J. S., Massri, A. J., Edgar, A., Wang, L. et al. Recent reconfiguration of an ancient developmental gene regulatory network in *Heliocidaris* sea urchins. *Nat. Ecol. Evol.*, **6**, 1907-1920 (2022).
23. Davidson, P. L., Lessios, H. A., Wray, G. A., McMillan, W. O., & Prada, C. Near-Chromosomal-Level Genome Assembly of the Sea Urchin *Echinometra lucunter*, a Model for Speciation in the Sea. *Genome Biol. Evol.*, **15**, evad093 (2023).
24. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat. Methods*, **18**, 170-175 (2021).
25. Laetsch, D. R., & Blaxter, M. L. BlobTools: Interrogation of genome assemblies. *F1000Research*, **6**, 1287 (2017).
26. Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., & Durbin, R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, **36**, 2896-2898 (2020).

27. Zhou, C., McCarthy, S. A., & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics*, **39**, 1- (2023).
28. Marçais, G., & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, **27**, 764-770 (2011).
29. Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.*, **11**, 1432. (2020).
30. Baril, T., Imrie, R. M., & Hayward, A. Earl Grey: a fully automated user-friendly transposable element annotation and analysis pipeline. *bioRxiv*. https://doi.org/10.1101/2022.06.30.498289 (2022).
31. Allio, R., Schomaker□Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., & Delsuc, F. MitoFinder: Efficient automated large□scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol. Ecol. Resour.*, **20**, 892-90 (2020).
32. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120. (2004).
33. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I. et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.*, **29**, 644 (2011).
34. Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res.*, 34(suppl_2), 435-439 (2006).
35. Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. BUSCO: assessing genomic data quality and beyond. *Curr. Protoc.*, **1**, e323 (2021).
36. Shingaki, M., & Uehara, T. (1984). Chromosome studies in the several species of sea urchin from Okinawa. *Zool. Sci.*, *1*, 1008.
37. Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O., & Borodovsky, M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.*, **18**, 1979-1990 (2008).

Table 1. Genome assembly statistic and sequencing information.

Table 2. Scaffold information of 22 pseudochromosomes.

Table 3. List of telomeric repeats identified in the genome.

Table 4. Statistics of annotated repetitive elements.

Figure 1. Genomic information of *Diadema setosum*. A) Photo of *D. setosum*; B) Statistics of the assembled genome; C) Omni-C contact map of the assembly visualised using Juicebox v1.11.08 (details can be found in Additional Table 6); D) Genomescope report with k-mer = 21; E) Repetitive elements distribution in the assembled genome.

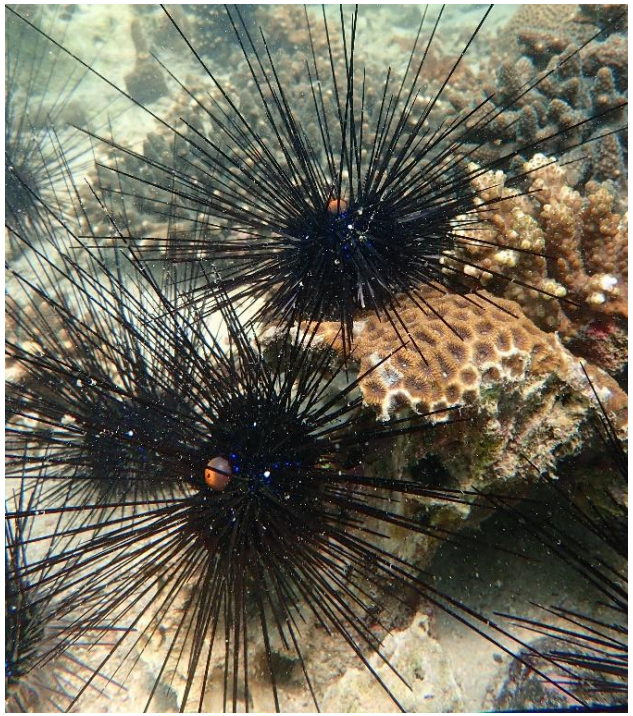
Supplementary Information 1. Genome and transcriptome sequencing information.

Supplementary Information 2. GenomeScope statistics with K-mer length 21.

Supplementary Information 3. Genome assembly QC and contaminant/cobiont detection.

Figure 1

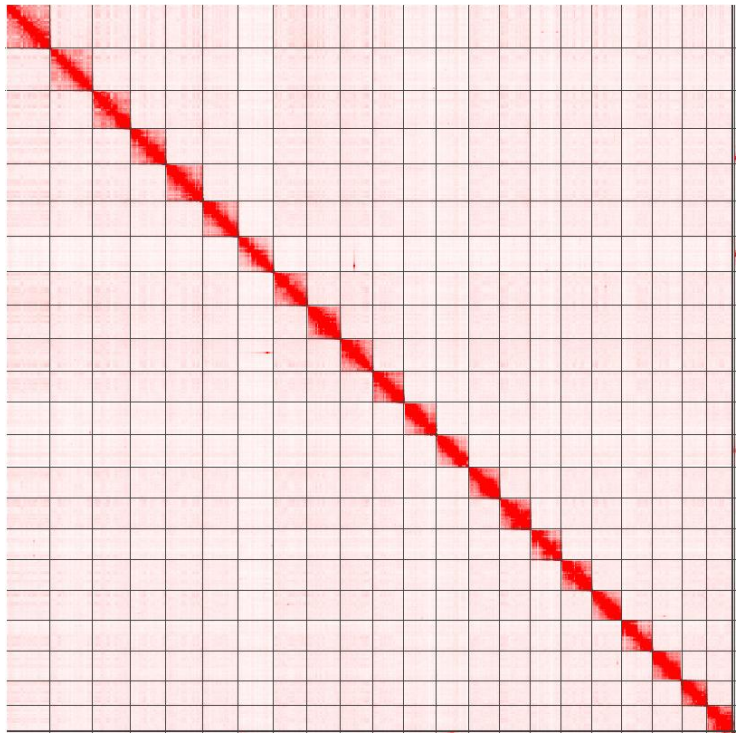
A)



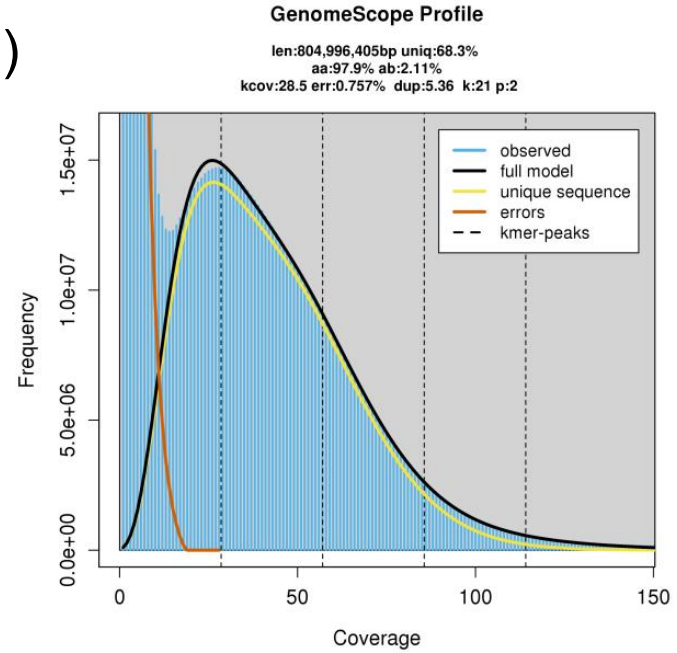
B)

	Diadema setosum
Accession number	JAWLSL000000000
Total length (bp	885,842,048
Number	101
Mean length (bp)	8,770,713
N_count	0.0093%
Gaps	412
N50	38,268,380
N50n	11
BUSCOs	98.10%
HiFi (X)	21
HiFi Reads	2,193,509
HiFi Bases	18,531,991,835
HiFi average length	8,449
Gene models	27,478

C)



E)



D)