Data in Brief 48 (2023) 109171

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Genomic and 16S metabarcoding data of *Holothuria tubulosa* Gmelin, 1791

Maria Kyritsi[a], George Tsiolas[a], Sofia Michailidou[a], Konstantinos Koukaras[a], Anagnostis Argiriou[a,b,*]

[a] *Institute of Applied Biosciences / CERTH, P.O. Box 60361, Thermi, Thessaloniki 57001, Greece*
[b] *Department of Food Science and Nutrition, University of the Aegean, Lemnos 81400, Greece*

ARTICLE INFO

Article history:
Received 31 December 2022
Revised 29 March 2023
Accepted 17 April 2023
Available online 22 April 2023

Dataset link: Holothuria tubulosa Whole Genome Sequencing (Original data)
Dataset link: Holothuria tubulosa digestive tract 16S rRNA metabarcoding (Original data)

Keywords:
Sea cucumber
Echinoderms
Holothuria tubulosa
Whole genome sequencing
Gastrointestinal tract microbiome
Metagenomics
16S rRNA
Illumina

ABSTRACT

Holothuria tubulosa Gmelin, 1791 is an edible sea cucumber species widespread in the Mediterranean Sea with ecological and increasing economic importance. Genome data of holothurian species is limited and the availability of genomic data resources is crucial in understanding their biology and adaptability mechanisms. This dataset presents the raw genome sequence data of *H. tubulosa* sequenced on an Illumina NextSeq 2000 platform. Genome size estimation was performed based on k-mer frequency approach. Additionally, the bacterial microbiome in the stomach and intestine of *H. tubulosa* collected from the Strymonian Gulf (North Aegean Sea, Greece) through 16S rRNA amplicon metabarcoding sequencing is reported. Sequencing was performed on an Illumina MiSeq platform. Analysis was conducted using the QIIME2 software package, the DADA2 algorithm and a trained classifier for taxonomy assignment. The datasets presented in this work serve as valuable resources for a comprehensive investigation of *H. tubulosa* at the genome level and for comparative genomics and echinoderms gut microbial studies.

* Corresponding author.
E-mail address: argiriou@certh.gr (A. Argiriou).

https://doi.org/10.1016/j.dib.2023.109171

Specifications Table

Subject	Biological sciences: Omics: Genomics, Microbiology: Microbiome
Specific subject area	Short-read whole genome sequencing and high throughput 16S rRNA metabarcoding analysis in the gastrointestinal tract of *Holothuria tubulosa* Gmelin, 1791.
Type of data	Illumina genomic short-read sequences (FASTQ files) Partial 16S rRNA Illumina sequences (FASTQ files) Table Image Figure
How the data were acquired	Whole genome sequencing (WGS) was conducted on an Illumina NextSeq 2000 system using Illumina's NextSeq 1000/2000 P3 Reagents (300 cycles). Software tools used for sequence analysis included Trim Galore!, FastQC and GenomeScope2.0. 16S rRNA amplicon sequencing was conducted on a MiSeq system using the MiSeq Reagent Kit v3 (600 cycles). Microbiome analysis was performed into the QIIME2 platform with the DADA2 algorithm and sequences were classified to taxonomic ranks using a naive Bayes classifier in the q2-feature-classifier plugin.
Data format	Raw and Analyzed data
Description of data collection	Adult sea cucumbers were collected from the coastal area of Strymonian Gulf (North Aegean Sea, Greece). Genomic DNA was extracted from the body wall of a single male and the WGS library was constructed using the Illumina DNA Prep Kit. Stomach and intestine samples were isolated from four sea cucumbers. Total DNA was extracted and the V3-V4 hypervariable regions of the 16S rRNA gene were amplified.
Data source location	Institution: Institute of Applied Biosciences – Centre for Research and Technology Hellas City/Town/Region: Thessaloniki Country: Greece Latitude and longitude for collected samples: 40.555943, 23.879743
Data accessibility	Repository name: NCBI BioProject Data identification number: PRJNA916651, PRJNA917268 Direct URL to data: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA916651 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA917268

Value of the Data

This dataset collection provides the first whole genome raw sequence data for the edible and ecologically important holothurian *Holothuria tubulosa* Gmelin, 1791, as well as the first investigation of the bacterial microbiome along its gastrointestinal tract.

- The metabarcoding data analysis provides the taxonomic profiles of the bacterial communities in the stomach and intestine of *H. tubulosa* and the characterization of the structural variations of the bacterial microbiome along the holothurian digestive tract.
- The metabarcoding data provide useful information on the bacterial diversity and abundance in the digestive system of *H. tubulosa* that can be useful to researchers conducting holothurian and environmental microbiome interaction studies.
- The deposited genomic short-read sequences can serve as a basis in the scientific community for further assembly and annotation of the *H. tubulosa* genome together with other long-read sequence data and transcriptome sequences.
- This genomic data collection can benefit future research on the phylogenomic analysis of holothurians and other echinoderms. Also, the data contributes to the investigation of the coevolution and spatial distribution of bacterial symbionts along the holothurian digestive tract.

1. Objective

Holothurians (Phylum: Echinodermata, Class: Holothuroidea) or sea cucumbers are ecologically and economically important marine invertebrates. Their deposit-feeding activity contributes to the sediment's nutrients, pH, and bacterial community equilibrium [1,2]. Sea cucumbers are exploited as a culinary delicacy mostly in Asian countries and are attracting interest in the aquaculture and pharmaceutical sectors due to their high nutrient content and bio-active substances [3,4]. *Holothuria tubulosa* Gmelin, 1791 is one of the most common holothurian species found in the Mediterranean Sea, living in dense populations mostly in *Posidonia* meadows [5]. High exportation demands and overharvesting resulted in the drastic depletion of natural population levels [6]. Nevertheless, genomic information on holothurian species inhabiting the Mediterranean Sea is limited [7]. Therefore, we generated the first whole genome sequencing data for future use in the characterization of the *H. tubulosa* genome to better understand *H. tubulosa* biology and support conservation efforts. Furthermore, in light of climate change's impact on marine ecosystems, we provide a 16S rRNA metabarcoding dataset of bacterial communities in two parts of the digestive tract of *H. tubulosa* to investigate the potential involvement in holothurians' role in ecological conservation.

2. Data Description

Here we report the first short-read raw genomic sequences and bacterial community profiling of different digestive compartments of *Holothuria tubulosa* Gmelin, 1791. The genome of an adult male *H. tubulosa* was sequenced on an Illumina NextSeq 2000 platform, generating 610,737,456 high-quality 150 bp paired-end reads. After quality filtering and adapter trimming 610,679,490 reads remained, with 93.64% corresponding to Phred quality score Q30, which are available

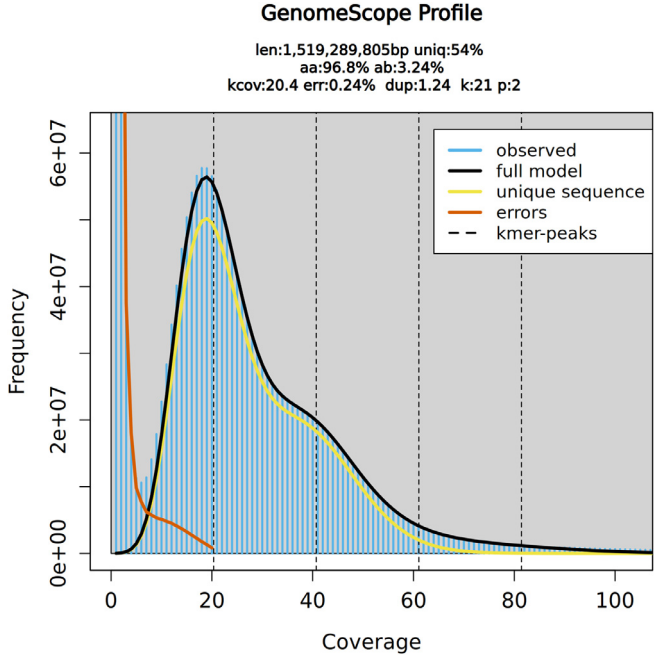

Fig. 1. Plot of k-mer spectra and model fitting of the *Holothuria tubulosa* Gmelin, 1791 genome.

through the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI); SRA accession SRR22992055, BioProject PRJNA917268, BioSample SAMN32533584. The genome of *H. tubulosa* is diploid and is estimated to be 1.52 Gb according to k-mer-based analysis. The k-mers peak was at a coverage depth of 19X, heterozygosity levels at 3.24%, and GC and repeat content were 39.5% and 46.0%, respectively (Fig. 1).

The V3-V4 hypervariable regions of the 16S rRNA gene from stomach and intestine samples were sequenced resulting in 359,825 raw reads. Raw reads are available through the National Center for Biotechnology Information (NCBI) BioProject PRJNA916651 (Table 1). After quality and taxonomy filtering 212,296 reads remained, corresponding to 2,351 unique Amplicon Sequence Variants (ASVs) (Supplementary table 1). The final number of sequences and ASVs for each sample resulted from a series of filtering steps and the results are presented in detail in Table 2. The number of ASVs was increased in the stomach compared to the intestine in three out of four individuals under study (HT1S, HT2S, and HT3S). Alpha diversity analysis indicated that samples derived from the stomach had higher bacterial species richness compared to intestine samples (Fig. 2).

Table 1
List of holothurian samples included in the 16S rRNA metabarcoding analysis; Individuals' sex, tissue part, BioSample and SRA accession numbers.

Sample ID	Sea cucumber sex	Tissue part	BioSample accession	SRA accession
HT1S	Male	Stomach	SAMN32489615	SRR22991838
HT1I	Male	Intestine	SAMN32489616	SRR22991837
HT2S	Male	Stomach	SAMN32489617	SRR22991836
HT2I	Male	Intestine	SAMN32489618	SRR22991835
HT3S	Female	Stomach	SAMN32489619	SRR22991834
HT3I	Female	Intestine	SAMN32489620	SRR22991833
HT5S	Female	Stomach	SAMN32489621	SRR22991832
HT5I	Female	Intestine	SAMN32489622	SRR22991831

Table 2
Number of total raw reads, filtered reads after DADA2 analysis, final number of reads after taxonomy-based filtering, and final observed ASVs for each sample.

Sample ID	Raw reads	DADA2 filtered reads	Taxonomy filtered – final reads	Inferred ASVs
HT1S	42,391	25,378	22,209	797
HT1I	54,747	35,701	35,034	262
HT2S	54,044	35,209	30,191	1158
HT2I	8856	4753	4287	155
HT3S	53,097	35,064	31,900	1001
HT3I	48,259	29,582	29,266	203
HT5S	61,247	42,091	40,220	157
HT5I	37,184	25,134	19,189	444
Total	359,825	232,912	212,296	2351

Principal Coordinates Analysis (PCoA) based on Bray-Curtis dissimilarities of the classified ASVs indicated the clustering of samples according to tissue type (Fig. 3). Stomach samples were clustered together, except HT5S which was found closer to the intestine group, while HT5I was differentiated from all the rest. Proteobacteria was the most abundant Phylum in both tissue types indicating 35.4% total relative abundance and was represented by Gammaproteobacteria and Alphaproteobacteria. The second most abundant Phylum was Bacteroidota (29.5%) (Fig. 4). Verrucomicrobiota (6.2%) and Planctomycetota (4.7%) were mostly observed in stomach samples compared to intestine samples. Acidobacteriota and Myxococcota were observed in low abundances in stomach samples, with 2.6% and 1.8% relative abundance, respectively, while they were almost absent in intestine samples (0.1% and 0.3%, respectively). At the Family taxonomic level, *Vibrionaceae* and *Flavobacteriaceae* were the most abundant families representing 11.8% and 10.6%

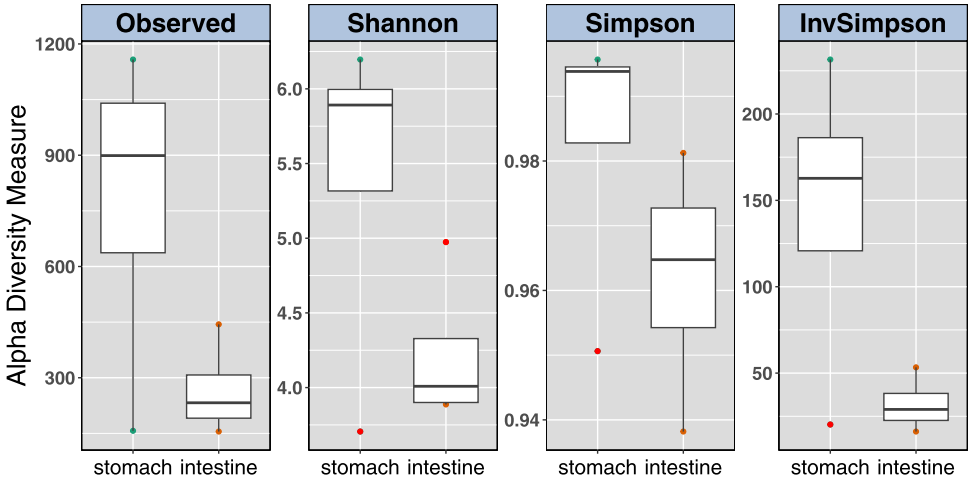

Fig. 2. Alpha diversity indices (Observed ASVs, Shannon, Simpson, and Inverse Simpson) calculated for each digestive tract compartment.

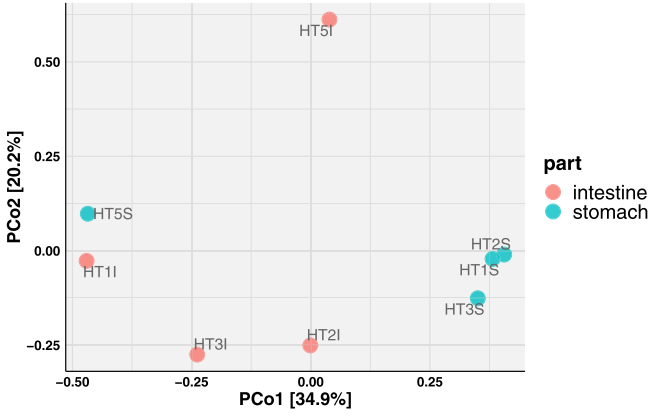

Fig. 3. Principal Coordinates Analysis (PCoA) plot of gastrointestinal tract samples based on bacterial microbiome composition. The first two principal coordinates captured 55.1% of the total variation.

of total relative abundance, respectively (Fig. 5). Members of the *Flavobacteriaceae* were more abundant compared to *Vibrionaceae* Family in the stomach. *Rubritaleaceae* and *Cryomorphaceae* were more abundant in stomach samples, while *Marinilabiliaceae* were mostly observed in intestine samples.

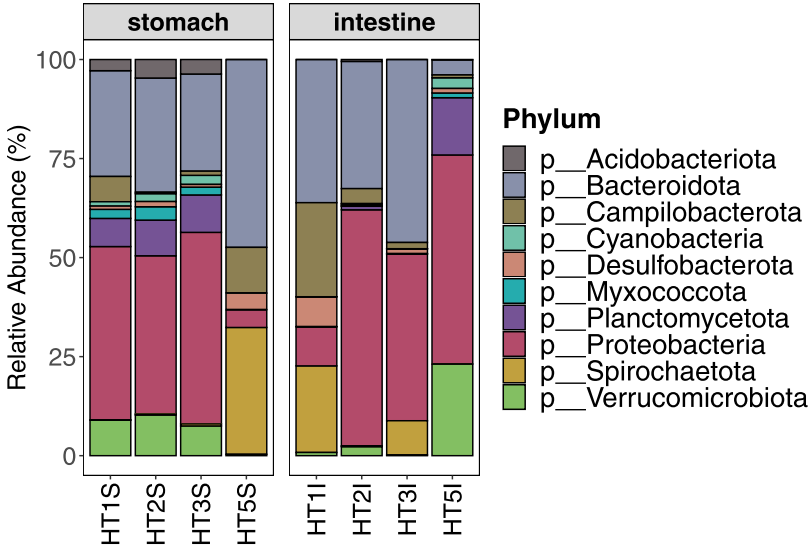

Fig. 4. Barplots depicting the percentage of relative abundances of the bacterial microbiome composition for the ten most abundant Phyla.

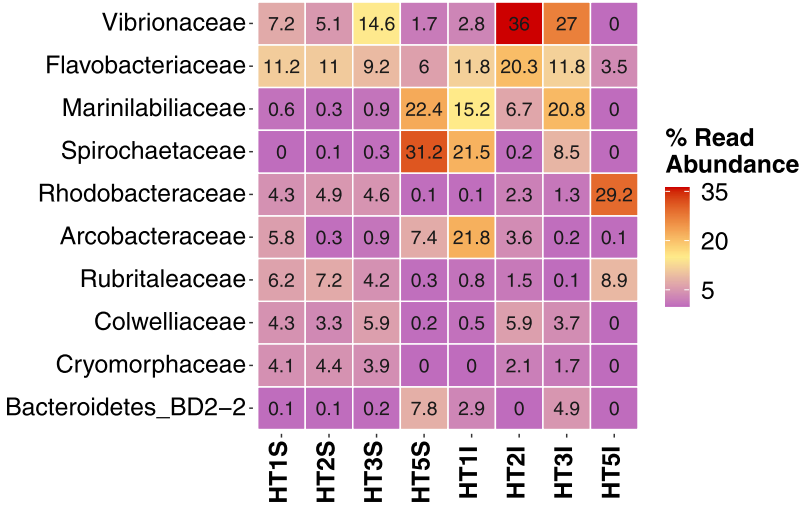

Fig. 5. Family level heatmaps of the percent relative abundances in each sample. The number in each box represents the percentage of aggregated ASVs of each Family.

3. Experimental Design, Materials and Methods

3.1. Sample Collection

Adult *H. tubulosa* individuals were collected from the coastal waters of the Strymonian Gulf, North Aegean Sea, Greece. Species identification was based on external morphological characteristics (brown dorsal and lighter ventral side, presence of small dark elongated papillae on the dorsal surface, and lack of Cuvierian tubules). Animals were kept in seawater during transportation to the laboratory. Dissection was performed on ice using sterile equipment and aseptic conditions.

3.2. Whole Genome Sequencing and Data Analysis

Genomic DNA was isolated from ~30 mg of body wall tissue of a single adult male individual using the Quick-DNA™ HMW MagBead Kit (ZYMO RESEARCH; Irvine, CA, USA) according to the manufacturer's instructions. gDNA integrity was evaluated by agarose gel electrophoresis and DNA concentration was measured on a Qubit 4.0 Fluorometer using the Qubit dsDNA BR assay kit (Invitrogen, Carlsbad, CA, USA). Whole genome library construction was conducted according to the Illumina DNA prep kit instructions (Illumina, San Diego, CA, USA). The indexed libraries' concentration was assessed by fluorometric method with the Qubit dsDNA BR kit. Quality assessment and average size estimation were analyzed by capillary electrophoresis on a Fragment Analyzer system (Agilent Technologies, Santa Clara, USA), using the DNF-474 High Sensitivity NGS Fragment Analysis Kit (DNF-474-0500; Agilent Technologies, Santa Clara, USA). Final quantification was performed by quantitative PCR (qPCR) on a Rotor-Gene Q thermocycler using the KAPA Library Quantification kit for Illumina sequencing platforms (Kapa Biosystems, Woburn, MA, USA). Molarity was calculated according to the average fragment size and the calculated concentration from the qPCR. WGS was performed on an Illumina NextSeq 2000 platform using Illumina's NextSeq 1000/2000 P3 Reagents (300 Cycles) (Illumina, San Diego, CA, USA).

Filtering for sequencing adapters, unidentified nucleotides (Ns), and mean quality score (Q>20) was conducted using the *trim_galore* script for paired-end reads of the Trim Galore! v0.6.7 wrapper, with default parameters [8]. Read quality was evaluated using FastQC v0.11.7 [9]. Genome size, repeat and GC content, and heterozygosity were estimated based on k-mer frequencies with the GenomeScope2.0 tool [10]. In particular, 21-mers were mapped with the *jellyfish count* tool with the options –C –m 21 –s 100000000000. Results were visualized using the *jellyfish histo* command. The resulting histogram was used in the R script *genomescope.R* with k-mer length 21, for genome size estimation and generation of the model and plots.

4. 16S rRNA Amplicon Sequencing and Bioinformatics Analysis

For the 16S rRNA amplicon sequencing four sea cucumbers (two males and two females) were used. The internal cavity was exposed after an abdominal incision, starting from the mouth to the opening of the cloaca; gonads and the respiratory tree were removed to reveal the alimentary tract which was collected and dissected. Stomach and intestine tissues were immediately frozen in liquid nitrogen and ground using a mortar and pestle. Microbial DNA was extracted from ~200 mg homogenized tissue using the ZymoBIOMICS™ DNA Miniprep Kit (ZYMO RESEARCH; Irvine, CA, USA) according to the manufacturer's instructions. DNA concentration was measured on a Qubit 4.0 Fluorometer. 16S rRNA amplicon libraries were constructed as previously described by Michailidou et al., 2021 [11]. Particularly, the primer set S-D-Bact-0341-b-S-17 (5′ - CCTACGGGNGGCWGCAG - 3′) and S-D-Bact-0785-a-A-21 (5' - GACTACHVGGGTATCTAATCC – 3') was used for the amplification of a ~460 bp fragment of the V3-V4 hypervariable regions [12]. All PCR reactions were performed on a Rotor-Gene Q thermocycler (Qiagen, Hilden, Ger-

many). Each PCR reaction was performed with 1X KAPA HiFi HotStart Ready enzyme mix (KAPA BIOSYSTEMS, Woburn, MA, USA), 0.4 μM primer mix, 0.625 μM Syto9 (Thermo Fisher Scientific, USA) and 12.5 ng/μl DNA. PCR conditions were as follows: initial denaturation at 95°C for 3 minutes, 30 cycles of denaturation at 95°C, annealing at 55°C and extension at 72°C for 30 seconds each, followed by a final extension step at 72°C for 5 minutes. Free primers and primer–dimer by-products were removed using AMPure XP beads (Agencourt, Beckmann-Coulter, USA). Dual index barcodes were added to all amplicons through an additional 8-cycle PCR, followed by a second bead clean-up step. Library validation and quantification were performed as described above for the WGS procedure. Sequencing was performed on a MiSeq platform, using the MiSeq reagent kit v3 (2 $\times$ 300 cycles) (Illumina, San Diego, CA, USA).

Raw 16S rRNA amplicon sequences (.fastq files) were imported into the QIIME2 v2021.11 (Quantitative Insights into Microbial Ecology 2) platform [13]. Sequences were processed with the denoising strategy and the final ASV table was generated using the *denoised-paired* method of the DADA2 (Divisive Amplicon Denoising Algorithm 2) plugin inside QIIME2 [14]. ASVs represent true biological sequences within each sample. For taxonomy classification, the scikit-learn naive-Bayes classifier methodology implemented in QIIME2 was used. Firstly, a V3-V4 region-specific classifier was trained on the SILVA 138 database (SILVA 138 99% OTUS) using the RESCRIPt plugin and the *feature-classifier fit-classifier-naive-bayes* command [15]. ASVs were then classified based on the trained classifier, using the *classify-sklearn* method of the q2-feature-classifier plugin with default parameters [16]. A final filtering step was performed to exclude non-16S rRNA sequences prior to diversity analysis. Specifically, the ASVs which were classified as Eukaryota or Archaea at the Kingdom level, mitochondria, and chloroplasts or were not assigned to any classification were removed from the dataset. For further processing and visualization of the results the ASV table (.biom file) was imported into the R environment v4.1.1 [17] and R studio software v2022.2.3.49 [18] using the *import_biom* function of the phyloseq package [19]. For barplot generation data were agglomerated at the desired taxonomic level with the *tax_glom* function, sorted with the *sort* function and the abundances were normalized as the percentage with the *transform_sample_counts* tool. Alpha diversity analysis was conducted with the *plot_richness* command. The ampvis2 package was used for Principal Coordinates Analysis (PCoA) with the Bray-Curtis method using *amp_ordinate*, and heatmap generation with the *amp_heatmap* command [20].

Ethics Statements

All experiments were carried out in accordance with the EU Directive 2010/63/EU for animal experiments.

CRediT Author Statement

Maria Kyritsi: Methodology, Investigation, Formal analysis, Visualization, Writing – Original Draft, Writing – Review & Editing; **George Tsiolas:** Methodology, Formal analysis, Writing – Review & Editing; **Sofia Michailidou:** Supervision, Methodology, Investigation, Writing – Review & Editing; **Konstantinos Koukaras:** Resources, Methodology, Writing – Review & Editing; **Anagnostis Argiriou:** Conceptualization, Writing – review & editing, Funding acquisition, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

Data Availability

Holothuria tubulosa Whole Genome Sequencing (Original data) (NCBI BioProject).
Holothuria tubulosa digestive tract 16S rRNA metabarcoding (Original data) (NCBI BioProject).

Acknowledgments

This research received funding from the BLUE_BOOST project of the Interreg ADRION Program Contract n. PAR.045013.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2023.109171.

References

[1] S. Purcell, C. Conand, S. Uthicke, M. Byrne, Ecological Roles of Exploited Sea Cucumbers, 2016, pp. 367–386, doi:10.1201/9781315368597-8.
[2] V. Costa, A. Mazzola, S. Vizzini, Holothuria tubulosa Gmelin 1791 (Holothuroidea, Echinodermata) enhances organic matter recycling in Posidonia oceanica meadows, J. Exp. Mar. Bio. Ecol. 461 (2014) 226–232, doi:10.1016/j.jembe.2014.08.008.
[3] C.C. Roggatz, M. González-Wangüemert, H. Pereira, M.J. Rodrigues, M.M. da Silva, L. Barreira, J. Varela, L. Custódio, First report of the nutritional profile and antioxidant potential of Holothuria arguinensis, a new resource for aquaculture in Europe, Nat. Prod. Res. 30 (2016) 2034–2040, doi:10.1080/14786419.2015.1107555.
[4] C. Luparello, D. Ragona, D.M.L. Asaro, V. Lazzara, F. Affranchi, M. Celi, V. Arizza, M. Vazzana, Cytotoxic potential of the coelomic fluid extracted from the sea cucumber holothuria tubulosa against triple-negative MDA-MB231 breast cancer cells, Biology (Basel) 8 (2019), doi:10.3390/biology8040076.
[5] M. Mavidis, M. Aplikioti, I. Kirmitzoglou, A. Koukouras, The euphausiacean fauna (Malacostraca) of the Aegean Sea, and comparison with those of the neighbouring seas, Crustaceana 78 (2005) 19–27, doi:10.1163/1568540054024538.
[6] D. Meloni, G. Esposito, Hygienic and commercial issues related to the illegal fishing and processing of sea cucumbers in the Mediterranean: A case study on over-exploitation in Italy between 2015 and 2017, Reg. Stud. Mar. Sci. 19 (2018) 43–46, doi:10.1016/j.rsma.2018.03.009.
[7] K.A. Long, C.W. Nossa, M.A. Sewell, N.H. Putnam, J.F. Ryan, Low coverage sequencing of three echinoderm genomes: The brittle star Ophionereis fasciata, the sea star Patiriella regularis, and the sea cucumber Australostichopus mollis, Gigascience 5 (2016) 1–4, doi:10.1186/s13742-016-0125-6.
[8] F. Krueger, Trim galore: a wrapper tool around cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, (2015). doi:10.5281/zenodo.5127899.
[9] S. Andrews, FastQC: A Quality Control Tool for High Throughput Sequence Data, 2010 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
[10] T.R. Ranallo-Benavidez, K.S. Jaron, M.C. Schatz, GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes, Nat. Commun. 11 (2020), doi:10.1038/s41467-020-14998-3.
[11] S. Michailidou, E. Pavlou, K. Pasentsis, J. Rhoades, E. Likotrafiti, A. Argiriou, Microbial profiles of Greek PDO cheeses assessed with amplicon metabarcoding, Food Microbiol. 99 (2021) 103836, doi:10.1016/j.fm.2021.103836.
[12] A. Klindworth, E. Pruesse, T. Schweer, J. Peplies, C. Quast, M. Horn, F.O. Glöckner, Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies, Nucleic Acids Res. 41 (2013) 1–11, doi:10.1093/nar/gks808.
[13] E. Bolyen, J.R. Rideout, M.R. Dillon, N.A. Bokulich, C.C. Abnet, G.A. Al-Ghalith, H. Alexander, E.J. Alm, M. Arumugam, F. Asnicar, Y. Bai, J.E. Bisanz, K. Bittinger, A. Brejnrod, C.J. Brislawn, C.T. Brown, B.J. Callahan, A.M. Caraballo-Rodríguez, J. Chase, E.K. Cope, R. Da Silva, C. Diener, P.C. Dorrestein, G.M. Douglas, D.M. Durall, C. Duvallet, C.F. Edwardson, M. Ernst, M. Estaki, J. Fouquier, J.M. Gauglitz, S.M. Gibbons, D.L. Gibson, A. Gonzalez, K. Gorlick, J. Guo, B. Hillmann, S. Holmes, H. Holste, C. Huttenhower, G.A. Huttley, S. Janssen, A.K. Jarmusch, L. Jiang, B.D. Kaehler, K. Bin Kang, C.R. Keefe, P. Keim, S.T. Kelley, D. Knights, I. Koester, T. Kosciolek, J. Kreps, M.G.I. Langille, J. Lee, R. Ley, Y.X. Liu, E. Loftfield, C. Lozupone, M. Maher, C. Marotz, B.D. Martin, D. McDonald, L.J. McIver, A.V. Melnik, J.L. Metcalf, S.C. Morgan, J.T. Morton, A.T. Naimey, J.A. Navas-Molina, L.F. Nothias, S.B. Orchanian, T. Pearson, S.L. Peoples, D. Petras, M.L. Preuss, E. Pruesse, L.B. Rasmussen, A. Rivers, M.S. Robeson, P. Rosenthal, N. Segata, M. Shaffer, A. Shiffer, R. Sinha, S.J. Song, J.R. Spear, A.D. Swafford, L.R. Thompson, P.J. Torres, P. Trinh, A. Tripathi, P.J. Turnbaugh, S. Ul-Hasan, J.J.J. van der Hooft, F. Vargas, Y. Vázquez-Baeza, E. Vogtmann, M. von Hippel, W. Walters, Y. Wan, M. Wang, J. Warren, K.C. Weber, C.H.D. Williamson, A.D. Willis, Z.Z. Xu, J.R. Zaneveld, Y. Zhang, Q. Zhu, R. Knight, J.G. Caporaso, Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2, Nat. Biotechnol. 37 (2019) 852–857, doi:10.1038/s41587-019-0209-9.

[14] B.J. Callahan, P.J. McMurdie, M.J. Rosen, A.W. Han, A.J.A. Johnson, S.P. Holmes, DADA2: High-resolution sample inference from Illumina amplicon data, Nat. Methods. 13 (2020) 581–583, doi:10.1007/978-1-4614-9209-2_118-1.
[15] M.S. Robeson, D.R. O'Rourke, B.D. Kaehler, M. Ziemski, M.R. Dillon, J.T. Foster, N.A. Bokulich, RESCRIPt: Reproducible sequence taxonomy reference database management, 2021. doi:10.1371/journal.pcbi.1009581.
[16] N.A. Bokulich, B.D. Kaehler, J.R. Rideout, M. Dillon, E. Bolyen, R. Knight, G.A. Huttley, J.Gregory Caporaso, Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin, Microbiome 6 (2018) 1–17, doi:10.1186/s40168-018-0470-z.
[17] R Development Core TeamR: A Language and Environment for Statistical Computing, R, 2011 https://www.r-project.org/.
[18] RStudio Rstudio-team, Integrated Development for R. RStudio, PBC, 2020 http://www.rstudio.com/.
[19] P.J. McMurdie, S. Holmes, Phyloseq: an R Package for reproducible interactive analysis and graphics of microbiome census data, PLoS One 8 (2013), doi:10.1371/journal.pone.0061217.
[20] K.S. Andersen, R.H. Kirkegaard, S.M. Karst, M. Albertsen, ampvis2: an R package to analyse and visualise 16S rRNA amplicon data, BioRxiv (2018) 10–11, doi:10.1101/299537.