Data in Brief 54 (2024) 110421

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

# Sea cucumber (*Holothuria glaberrima*) intestinal microbiome dataset from Puerto Rico, generated by shotgun sequencing

Edwin Omar Rivera-Lopez[a,b], Rene Nieves-Morales[a], Gabriela Melendez-Martinez[a], Jessica Alejandra Paez-Diaz[a], Sofia Marie Rodriguez-Carrio[a], Josue Rodriguez-Ramos[c], Luis Morales-Valle[a], Carlos Rios-Velazquez[a,*]

[a] *Microbial Biotechnology and Bioprospecting Laboratory, Biology Department, University of Puerto Rico, Mayagüez, P.R. 00681-9000, United States*
[b] *Food Science and Technology Program, University of Puerto Rico at Mayagüez, P.R. 00681-9000, United States*
[c] *Pacific Northwest National Laboratory, Biological Sciences Division, WA, United States*

ARTICLE INFO

*Article history:*
Received 9 January 2024
Revised 3 April 2024
Accepted 9 April 2024
Available online 15 April 2024

Dataset link: Holothuria glaberrima Eviscerated Gut Metagenome Raw sequence reads (Original data)

*Keywords:*
Holothuria glaberrima
Metagenome
Sea cucumber
Intestinal microbiome
Metagenomics

ABSTRACT

The sea cucumber (*H. glaberrima*) is a species found in the shallow waters near coral reefs and seagrass beds in Puerto Rico. To characterize the microbial taxonomic composition and functional profiles present in the sea cucumber, total DNA was obtained from their intestinal system, fosmid libraries constructed, and subsequent sequencing was performed. The diversity profile displayed that the most predominant domain was Bacteria (76.56 %), followed by Viruses (23.24 %) and Archaea (0.04 %). Within the 11 phyla identified, the most abundant was Proteobacteria (73.16 %), followed by Terrabacteria group (3.20 %) and Fibrobacterota, Chlorobiota, Bacteroidota (FCB) superphylum (1.02 %). The most abundant species were *Porvidencia rettgeri* (21.77 %), *Pseudomonas stutzeri* (14.78 %), and *Alcaligenes faecalis* (5.00 %). The functional profile revealed that the most abundant functions are related to transporters, MISC (miscellaneous information systems), organic nitrogen, energy, and carbon utilization. The data collected in this project on the diversity and functional profiles of the intestinal system of the *H.*

* Corresponding author.
*E-mail address:* carlos.rios5@upr.edu (C. Rios-Velazquez).

https://doi.org/10.1016/j.dib.2024.110421

*glaberrima* provided a detailed view of its microbial ecology. These findings may motivate comparative studies aimed at understanding the role of the microbiome in intestinal regeneration.



Specifications Table

| | |
|---|---|
| Subject | Microbiology |
| Specific subject area | Metagenomics |
| Data format | Raw data, Processed |
| Type of data | Fasta Qfile, Figures |
| Data collection | Sea cucumbers were collected from Piñones Beach in San Juan, Puerto Rico. Metagenomic DNA was extracted from three samples of the intestinal system of *Holothuria glaberrima*: the complete digestive system (DS), the washed intestine (WI), and the contents from the washed intestine (CW). The presented data resulted from the extraction of the fosmid pCC1FOS, shotgun sequencing and analysis in the National Microbiome Data Collaborative – Empowering the Development of Genomics Expertise (NMDC – EDGE) platform. |
| Data source location | Piñones Beach in San Juan, Puerto Rico (18.451141, -65.905634) |
| Data accessibility | Raw data and annotations of this metagenome are available in NCBI under BioProject PRJNA1061805 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1061805) and NMDC – EDGE under the project title "Holothuria glaberrima - Eviscerated Gut Metagenome EDGE Bioinformatics" under link: https://nmdc-edge.org/public/project?code=uTDw1Jv0NGWhTOQV |

## 1. Value of the Data

- This project represents the diversity and functional profiles of the intestinal system of the sea cucumber (H. glaberrima) in the Caribbean.
- The profiles generated can be used in comparative studies with the microbial flora present in other sea cucumbers.
- Diversity and functional profiles may prompt comparative bioprospecting studies to understand the potential role of the microbiome in processes such as intestinal regeneration.

## 2. Data Description

The sea cucumber (*H. glaberrima*) (Fig. 1) is a species found in the shallow waters near coral reefs and seagrass beds in Puerto Rico [1]. Members of the class Holothuroidea, including *H. glaberrima*, have the capability to regenerate their internal organs after the process of evisceration [2]. Here, we present descriptions of the diversity (Fig. 2) and functional profiles (Fig. 3) of the microbial communities within the intestinal system of the sea cucumber. These data were derived from a pool of three metagenomic libraries, each sourced from a specific anatomical region. From the evisceration process, the entire intestinal system, along with its internal contents were selected; this sample was identified as the complete digestive system (DS). Subsequently, only the intestine was selected and washed with a saline solution (0.85 % NaCl); this procedure generated two samples, the washed intestine (WI) and the contents from the wash (CW). To access, most of the microbial populations present in the anatomical regions of the sea cucumber


**Fig. 1.** *Holothuria glaberrima* before and after evisceration. **(A)** *H. glaberrima* exhibits an average length ranging from 10 to 15 cm (4 to 6 inches). Its predominant coloration is characterized by shades of black and dark brown [3]. **(B)** Evisceration process of *H. glaberrima* after induction with 2 mL of KCl [0.35 M].
Credit: Edwin Omar Rivera-Lopez and Carlos Ríos-Velázquez

at the genetic level, especially those in low abundance, independent DNA extractions from each region were performed and then combined to further be sequenced. Furthermore, the overarching goal is to understand the microbiome in each section of the sea cucumber's intestine. A total of 266,235 large insert (40kb) metagenomic clones were generated. After sequencing, a total of 9,259 sequences were obtained, with a median length of 653 bp, containing a cumulative length of 9,039,785 nucleotides. Furthermore, Table 1 provides a summary of sequence assembly statistics, including the total number of outputs read after filtering and trimming, the number of scaffolds, gaps percentage, the contig N50 value, and the number of genes per 1Mbp. Table 2 offers a summary of the metagenome the gene calls for the assembled scaffolds. Additionally, the Supplementary Data contains annotated genes associated with Fig. 3.

**Table 1**
Summary of sequence assembly statistics.

| | |
|---|---|
| Total number of outputs read after filters and trimming | 1,018,378 |
| Scaffolds | 9,259 |
| Gaps PCT | 0 |
| Ctg_N50 | 2,035 |
| Genes per 1Mbp | 1,852.04 |

**Table 2**
Summary of metagenome gene calls.

| Feature type | Prediction method | Number of seqs | Number of bps | Median length | Average length | Length shortest seq | Length longest seq | Standard deviation | Number of predicted features |
|---|---|---|---|---|---|---|---|---|---|
| CDS | Prodigal v2.6.3_patched | 8,911 | 7,224,504 | 438 | 496.666 | 75 | 6,081 | 355.181 | 14,546 |
| CDS | GeneMark.hmm-2 v1.25_lic | 1,614 | 602,838 | 207 | 322.546 | 90 | 3,447 | 307.198 | 1,869 |
| misc_feature | INFERNAL 1.1.3 (Nov 2019) | 3 | 177 | 59 | 59 | 59 | 59 | 0 | 3 |
| regulatory | INFERNAL 1.1.3 (Nov 2019) | 3 | 177 | 59 | 59 | 59 | 59 | 0 | 3 |
| rRNA | INFERNAL 1.1.3 (Nov 2019) | 3 | 177 | 59 | 59 | 59 | 59 | 0 | 3 |
| tmRNA | INFERNAL 1.1.3 (Nov 2019) | 3 | 177 | 59 | 59 | 59 | 59 | 0 | 3 |
| ncRNA | INFERNAL 1.1.3 (Nov 2019) | 3 | 177 | 59 | 59 | 59 | 59 | 0 | 3 |
| misc_binding | INFERNAL 1.1.3 (Nov 2019) | 3 | 177 | 59 | 59 | 59 | 59 | 0 | 3 |

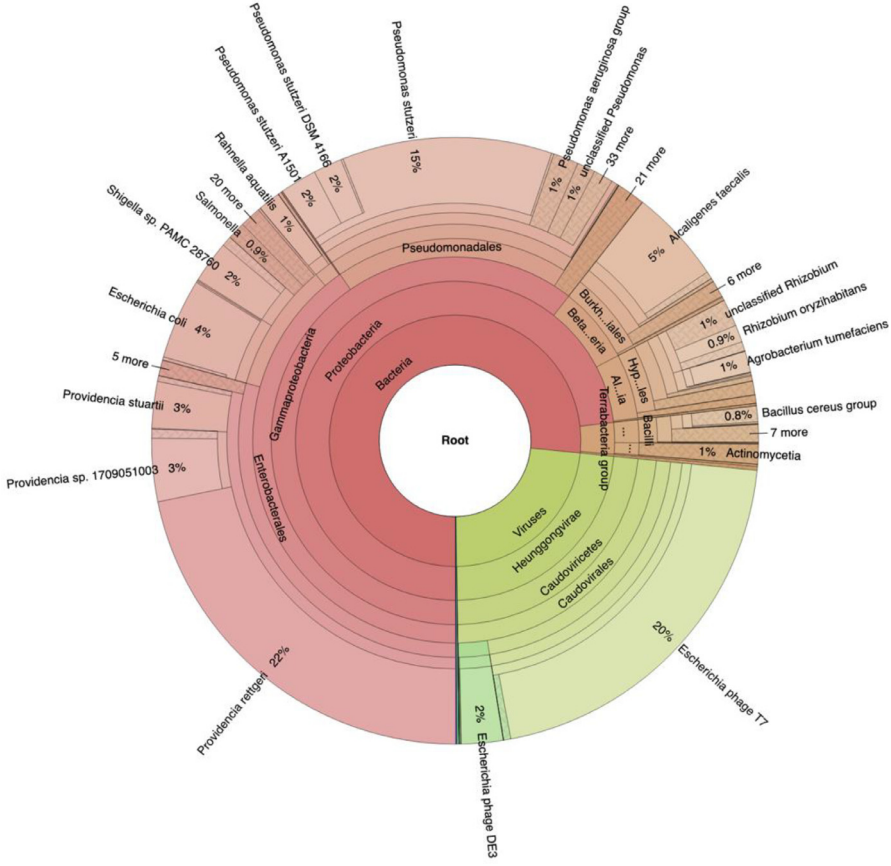


**Fig. 2.** Taxonomic diversity of the sea cucumber *(Holothuria glaberrima)* intestinal metagenome. Utilizing the Centrifuge Metagenomic Classification Tool through NMDC-EDGE showed that there were 5,557,757 classified reads and 314,616 species reads. The most abundant domain was Bacteria (76.56 %), followed by Virus (23.24 %), and Archaea (0.04 %). The most common Phyla out of the 11 found in Bacteria was Proteobacteria (73.16 %), followed by those included in the Terrabacteria group (3.20%) and Fibrobacterota, Chlorobiota, Bacteroidota (FCB) superphylum (1.02 %). The most abundant species were *Porvidencia rettgeri* (21.77 %), *Pseudomonas stutzeri* (14.78 %), and *Alcaligenes faecalis* (5.00 %).

## 3. Experimental Design, Materials and Methods

### 3.1. Sample Collection

The specimens were collected by Dr. Jose Garcia-Arraras' Laboratory from Piñones Beach in San Juan, Puerto Rico. A total of 13 specimens were collected and transported to the laboratory, where they were kept in a marine water aquarium (extracted from the natural environment) for 24 hours prior to the induction of evisceration (Fig. 1B) [4–6]. Subsequently, the samples DS, WI, and CW were collected in Falcon tubes and transported on dry ice to the Laboratory of Microbial Biotechnology and Bioprospecting, at the University of Puerto Rico at Mayagüez for genetic material extraction, processing and metagenomic libraries generation.

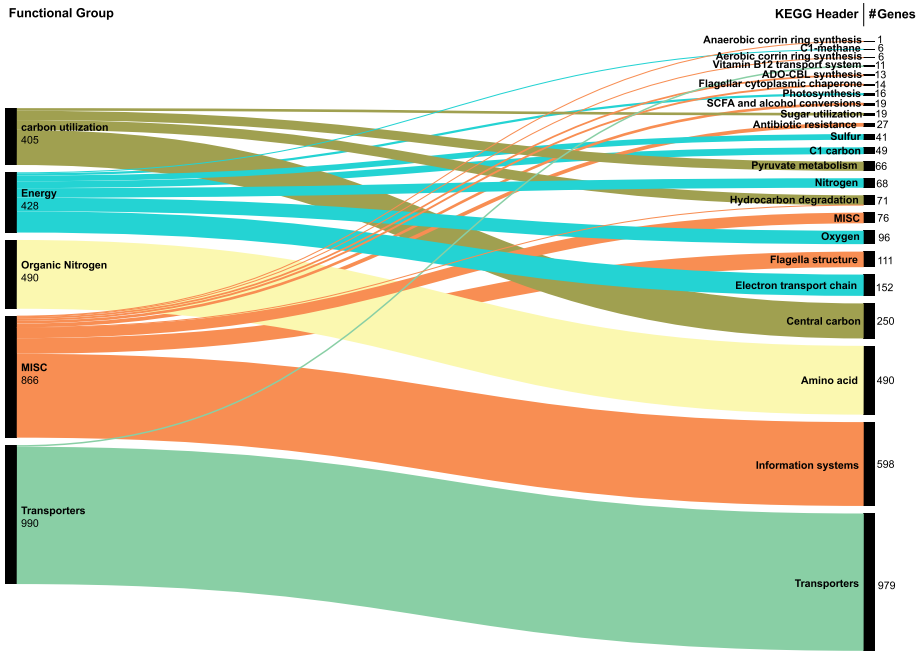


**Fig. 3.** Functional annotation of the *Holothuria glaberrima* anatomic regions metagenomic genes. Alluvial plot exhibits the total genes that were annotated and assigned to a KEGG category by DRAM metabolism sheet. The colors represent different types of functional groups. The numbers represent the total number of genes identified per category. Full functional annotation of metagenomes is publicly available via the NMDC EDGE platform and NCBI. From the assembly scaffolds, gene calls resulted in 16,615 total genes. Of those, 9,092 contained a K0 number, used to categorize them metabolically, obtaining 2,480 genes.

### *3.2. DNA Extraction and Library Preparation*

For DNA extraction, a direct method was employed, combining mechanical (freezing and thawing), enzymatic (lysozyme), and chemical approaches (SDS and GITC as a chaotropic agent) [7]. Fragments of 40 kb were chosen from the agarose gel. Subsequently, the purified DNA was ligated into the fosmid pCC1FOS and packed into Lambda phages (MaxPlax™ Lambda), followed by the transduction of the packed DNA into *Escherichia coli* EPI300-T1R. The clones were combined into a masterpool, which was subsequently stored at -80°C. The procedure was carried out with DS, WI, and CW parts of the sea cucumber's intestinal system.

### *3.3. Metagenome Sequencing*

Fosmid extraction (QIAGEN Plasmid Midi) was carried out from a masterpool culture of each sample (DS, WI, and CW), incubated for 5 hours at 37°C. The resulting metagenomic DNA was sent to Mr. DNA's laboratory (http://www.mrdnalab.com) for short-read sequencing (Illumina). In this process, the sample was fragmented, and the adapter sequences were incorporated. Subsequently, the library concentration was reduced to 4.0 nM and sequencing was performed with 600 cycles using the Illumina MiSeq system.

### *3.4. Metagenomic Data Processing*

Metagenomic data based on the pooled fosmids was processed through the National Microbiome Data Collaborative (NMDC) open informatics platform, EDGE [8] using their

standardized bioinformatic workflows. Docker images and full commands for each of the processes can be found on their GitHub (https://github.com/microbiomedata). Briefly, raw reads files were uploaded to the NMDC EDGE platform via their graphic user interface (GUI) (https://nmdc-edge.org/home) and the full metagenomics workflow was employed. Read quality control was performed by rqcfilter2 from BBTools [9] as described on the NMDC GitHub (https://github.com/microbiomedata/ReadsQC). Read-based taxonomy was assigned using three different tools: Kraken2 [10], GOTTCHA [11], and Centrifuge [12] (https://github.com/microbiomedata/ReadbasedAnalysis). With the resulting KRONA plot from Centrifuge shown in Figure 2. Metagenome assembly was performed with MetaSpades [13] (https://github.com/microbiomedata/metaAssembly). Assembled scaffolds were then annotated with tRNAscan_SE [14], RFAM [15], CRT [16], Prodigal [17], and GeneMarkS [18] as described in the NMDC workflow (https://github.com/microbiomedata/mg_annotation). The resulting K0s were then assigned to KEGG functional groups using the DRAM metabolism hierarchy [19] and plotted in Figure 3. Additionally, metagenome assembled genomes (MAGs) were as described in the NMDC GitHub (https://github.com/microbiomedata/metaMAGs).

## Limitations

Not applicable.

## Ethics Statement

This article is an original work of the authors. *Holothuria glaberrima* is found in large numbers along the coast of Puerto Rico. This species is not threatened with extinction nor subject to protective measures. Since they are invertebrates, there is no need for specific permits when collecting them.

## Data Availability

Holothuria glaberrima Eviscerated Gut Metagenome Raw sequence reads (Original data) (NCBI).

## CRediT Author Statement

**Edwin Omar Rivera-Lopez:** Investigation, Writing – original draft; **Rene Nieves-Morales:** Investigation, Writing – original draft; **Gabriela Melendez-Martinez:** Investigation, Writing – original draft, Formal analysis; **Jessica Alejandra Paez-Diaz:** Investigation, Writing – original draft; **Sofia Marie Rodriguez-Carrio:** Investigation, Writing – original draft; **Josue Rodriguez-Ramos:** Formal analysis; **Luis Morales-Valle:** Formal analysis; **Carlos Rios-Velazquez:** Supervision, Writing – review & editing.

## Acknowledgements

This work was supported by the National Science Foundation (NSF), [grant number 2100494]. Special thanks to the Laboratory of Dr. Jose Garcia-Arraras at the University of Puerto Rico Rio Piedras for providing us with the sea cucumber samples.

## Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2024.110421.

## References

[1] J.J. Alvarado, Echinoderm diversity in the Caribbean Sea, Marine Biodiver. 41 (2011) 261–285.
[2] J.E. García-Arrarás, L. Estrada-Rodgers, R. Santiago, I.I. Torres, L. Díaz-Miranda, I. Torres-Avillán, Cellular mechanisms of intestine regeneration in the sea cucumber, *Holothuria glaberrima* Selenka (Holothuroidea: Echinodermata), J. Experim. Zool. 281 (4) (1998) 288–304.
[3] E.H. Kaplan, A field guide to coral reefs: Caribbean and Florida, 27, Houghton Mifflin Harcourt, 1999.
[4] M. Pagán-Jiménez, J.F. Ruiz-Calderón, M.G. Dominguez-Bello, J.E. García-Arrarás, Characterization of the intestinal microbiota of the sea cucumber *Holothuria glaberrima*, PLoS One 14 (1) (2019) e0208011.
[5] L.M. Díaz-Díaz, N. Rosario-Meléndez, A. Rodríguez-Villafañe, Y.Y. Figueroa-Vega, O.A. Pérez-Villafañe, A.M. Colón-Cruz, P.I. Rodríguez-Sánchez, J.M. Cuevas-Cruz, S.J. Malavez-Cajigas, S.M. Maldonado-Chaar, J.E. García-Arrarás, Antibiotics modulate intestinal regeneration, Biology 10 (3) (2021) 236.
[6] J.L. Quiñones, R. Rosa, D.L. Ruiz, J.E. García-Arrarás, Extracellular matrix remodeling and metalloproteinase involvement during intestine regeneration in the sea cucumber *Holothuria glaberrima*, Dev. Biol. 250 (1) (2002) 181–197.
[7] J.M. Cruz, M.A. Ortega, J.C. Cruz, P. Ondina, R. Santiago, C. Rios-Velazquez, Unraveling activities by functional-based approaches using metagenomic libraries from dry and rain forest soils in Puerto Rico. Current Research Technology and Education Topics, Appl. Microbiol. Microbial. Biotechnol. 2 (2) (2010) 1471–1478.
[8] E.A. Eloe-Fadrosh, F. Ahmed, A. Anubhav, M. Babinski, J. Baumes, M. Borkum, L. Bramer, S. Canon, D. s. Christianson, Y. e. Corilo, K. w. Davenport, B. Davis, M. Drake, W. d. Duncan, M. c. Flynn, D. Hays, B. Hu, M. Huntemann, J. Kelliher, S. Lebedeva, P.-E. Li, M. Lipton, C.-C. Lo, S. Martin, D. Millard, K. Miller, M. a. Miller, P. Piehowski, E.P. Jackson, S. Purvine, T.B.K. Reddy, R. Richardson, M. Rudolph, S. Sarrafan, M. Shakya, M. Smith, K. Stratton, J.C. Sundaramurthi, P. Vangay, D. Winston, E. m. Wood-Charlson, Y. Xu, P.S.G. Chain, L.A. McCue, D. Mans, C. j. Mungall, N. j. Mouncey, K. Fagnan, The National microbiome data collaborative data portal: an integrated multi-omics microbiome data resource, Nucl. Acids Res. 50 (2022) D828–D836.
[9] B. Bushnell, BBMap: A Fast, Accurate, Splice-Aware Aligner, Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States), 2014 https://www.osti.gov/servlets/purl/1241166 (Accessed March 23, 2022).
[10] D.E. Wood, J. Lu, B. Langmead, Improved metagenomic analysis with Kraken 2, Genome Biol. 20 (2019) 1–13.
[11] T.A.K. Freitas, P.E. Li, M.B. Scholz, P.S. Chain, Accurate read-based metagenome characterization using a hierarchical suite of unique signatures, Nucl. Acids Res. 43 (10) (2015) e69 -e69.
[12] D. Kim, L. Song, F.P. Breitwieser, S.L. Salzberg, Centrifuge: rapid and sensitive classification of metagenomic sequences, Genome Res. 26 (12) (2016) 1721–1729.
[13] S. Nurk, D. Meleshko, A. Korobeynikov, P.A. Pevzner, metaSPAdes: a new versatile metagenomic assembler, Genome Res., 27 (5) (2017) 824–834.
[14] P.P. Chan, T.M. Lowe, tRNAscan-SE: searching for tRNA genes in genomic sequences, Gene Pred.: Methods Prot. (2019) 1–14.
[15] S. Griffiths-Jones, S. Moxon, M. Marshall, A. Khanna, S.R. Eddy, A. Bateman, Rfam: annotating non-coding RNAs in complete genomes, Nucl. Acids Res. 33 (suppl_1) (2005) D121–D124.
[16] C. Bland, T.L. Ramsey, F. Sabree, M. Lowe, K. Brown, N.C. Kyrpides, P. Hugenholtz, CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats, BMC Bioinfor. 8 (1) (2007) 1–8.
[17] D. Hyatt, G.L. Chen, P.F. LoCascio, M.L. Land, F.W. Larimer, L.J. Hauser, Prodigal: prokaryotic gene recognition and translation initiation site identification, BMC Bioinfor. 11 (2010) 1–11.
[18] J. Besemer, A. Lomsadze, M. Borodovsky, GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions, Nucl. Acids Res. 29 (12) (2001) 2607–2618.
[19] M. Shaffer, M.A. Borton, B.B. McGivern, A.A. Zayed, S.L. La Rosa, L.M. Solden, P. Liu, A.B. Narrowe, J. Rodriguez-Ramos, B. Bolduc, M.C. Gazitua, R.A. Daly, G.J. Smith, D.R. Vik, P.B. Pope, M.B. Sullivan, S. Roux, K.C. Wrighton, DRAM for distilling microbial metabolism to automate the curation of microbiome function, Nucl. Acids Res. 48 (16) (2020) 8883–8900.