

Shotgun metagenomic dataset of seawater bacterial communities from Pari Islands, Indonesia

Buntora Pasaribu,^{1,2} Christopher Vincent Mishael Dilens,^{2,3} Muhammad Wahyudin Lewaru,¹ Diah Ayuningrum,⁴ Mufti Petala Patria,⁵ Donny Juliandri Prihadi,¹ Noir P. Purba,¹ Mochamad Untung Kurnia Agung,¹ Ismail Maqbul,¹ Sulistiowati Sulistiowati⁶

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT Pari Island is located in Seribu Islands, Indonesia, and is well-known for its marine biodiversity. Shotgun metagenomic sequencing was performed using the DNaseq-G400 platform, and bioinformatics approaches were applied to analyze the sequence data.

KEYWORDS metagenomics, microbial communities

Pari Island is a region of significant ecological value with rich marine biodiversity, including *Ascidia* spp., coral reefs, mangroves, and seagrass (1–3). Furthermore, Pari Island currently faces a severe environmental crisis that is influenced by various tourist activities, climate change, and pollution (3–5). However, the data on microbial communities in the water column are still unknown.

Seawater samples were collected in August 2024 from Pari Islands (5.85205°S, 106.60571°E), with measured conditions of temperature (29.55°C), salinity (32 ppt), pH (7.25), and dissolved oxygen (6.06 mg/L). Two-liter water samples were aseptically filtered using a vacuum filter with a 0.22-μm nitrocellulose membrane (Sartorius, Germany). The filter membranes used for filtration were then employed in the subsequent process. DNA extraction was performed using the ZymoBIOMICS DNA Kit (Zymo Research, USA). DNA quality was assessed using agarose gel electrophoresis and the Qubit BR-DNA Assay (Thermo Fisher Scientific, USA), following the manufacturer's instructions. DNA concentration was quantified using Qubit fluorometer dsDNA BR Assay (Thermo Fisher Scientific), and 13–15 ng/μL of each sample was used for sequencing, following the manufacturer's instructions. The DNA library was prepared through end-repaired adapter and PCR amplification using the MGIEasy FS DNA Library Prep Kit (MGI Tech, China). Sequencing was carried out on MGI DNBSeg-G400 (MGI Tech) in the 2 × 150 bp paired-end mode, using the DNBSeg-G400RS High Throughput Sequencing Kit (FCL PE150) (MGI Tech). The raw data and quality of reads were obtained using the DNBSeg-G400RS, as detailed in Table 1 (6). Raw sequencing reads were evaluated for quality using fastp v1.0, and adapter sequences, along with low-quality bases, were trimmed from both ends to improve data reliability for downstream analyses (7, 8). The metagenomic assembly was performed using Metaspades v4.2.0 with the k-mer size set to 31 (9) (<https://doi.org/10.5281/zenodo.18476699>). The results of the metagenomic assembly were evaluated and compared with the closest reference using metaQUAST (10). Functional and taxonomic analyses were performed using Diamond v0.9.25 (11) + MEGAN v6 (12). Diamond v0.9.25 was used to align all DNA fragments to a protein database. The alignment results were then analyzed for taxonomic and functional information using MEGAN v6.

Taxonomic analysis revealed that Proteobacteria was the most abundant phylum (75.5%), followed by Cyanobacteria (14.5%), with minor contributions from viruses, eukaryotes, and other organisms (10%) (<https://doi.org/10.5281/zenodo.18476606>). The

Editor Vanja Klepac-Ceraj, Wellesley College, Wellesley, Massachusetts, USA

Address correspondence to Buntora Pasaribu, buntora.pasaribu@unpad.ac.id.

Buntora Pasaribu and Christopher Vincent Mishael Dilens contributed equally to this article. Author order was determined both alphabetically and based on increasing seniority.

The authors declare no conflict of interest.

Received 18 December 2025

Accepted 24 February 2026

Published 23 March 2026

Copyright © 2026 Pasaribu et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

TABLE 1 Accession and data for metagenome information for the analyzed seawater samples

Sample	NCBI BioProject	SRA accession no.	No. of raw sequence reads	No. of reads that passed QC	N50 (bp)	Max contig
P1	PRJNA1218935	SRX27269221	40,073,216	40,073,120	360,742	121,199
P2	PRJNA1218910	SRX27269222	40,286,300	40,286,186	290,644	90,545

distribution of the most abundant bacterial species identified through metagenomic analysis demonstrated *Prochlorococcus marinus*, *Synechococcus* sp. *M16.1*, *Synechococcus* sp. *TAK9802*, *Methylobacterium oryzae*, *Candidatus Pelagibacter* sp. *FZCC0015*, *Candidatus Pelagibacter* sp. *RS39*, *Cognatishimia activa*, and *Thalassovita* sp. Metagenomic functional profiling revealed that metabolism (225,000 reads) was the predominant functional category in the microbial community of the water column sample, followed by energy production (75,000 reads) and protein processing (73,000 reads). Functional analysis of metabolism identified amino acid transport and metabolism (60,000 reads) as the most abundant metabolic function, followed by energy production and conversion (40,000 reads), inorganic ion transport and metabolism (38,000 reads), carbohydrate transport and metabolism (32,000 reads), lipid transport and metabolism (21,000 reads), and coenzyme transport and metabolism (20,000 reads).

ACKNOWLEDGMENTS

This research was funded by Riset Kolaborasi Indonesia under grant number 2213/UN6.3.1/T.U00/2024 to B.P.

AUTHOR AFFILIATIONS

¹Department of Marine Sciences, Faculty of Fisheries and Marine Sciences, Universitas Padjadjaran, Bandung, Indonesia

²Shallow Coastal and Aquatic Research Forensic (SCARF) Laboratory, Faculty of Fishery and Marine Sciences, Universitas Padjadjaran, Bandung, Indonesia

³Marine Science Study Programme, Faculty of Fisheries and Marine Sciences, Universitas Padjadjaran, Bandung, Indonesia

⁴Department of Aquatic Resource Management, Diponegoro University, Semarang, Indonesia

⁵Department of Biology, University of Indonesia, Depok, Indonesia

⁶Department of Fisheries, Faculty of Agriculture, Gadjah mada University, Yogyakarta, Indonesia

AUTHOR ORCIDs

Buntora Pasaribu  <http://orcid.org/0000-0002-9639-1506>

AUTHOR CONTRIBUTIONS

Buntora Pasaribu, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Writing – original draft, Writing – review and editing | Christopher Vincent Mishael Dilens, Investigation, Writing – original draft | Muhammad Wahyudin Lewaru, Data curation, Formal analysis | Diah Ayuningrum, Visualization | Mufti Petala Patria, Validation | Donny Juliandri Prihadi, Software | Noir P. Purba, Data curation, Formal analysis, Methodology | Mochamad Untung Kurnia Agung, Formal analysis | Ismail Maqbul, Writing – review and editing | Sulistiowati Sulistiowati, Methodology

DATA AVAILABILITY

The sequences were deposited and are available in an NCBI SRA data set under BioProject numbers [PRJNA1218935](#) and [PRJNA1218910](#). The accession numbers are listed in Table 1.

REFERENCES

1. Utami PB, Barid VB, Lukman KM, Gunawan B, Ihsan YN, Jogaswara H, Sjafrie NDM. 2024. The paradox between economic gain and environmental sustainability: a case of small island tourism in Pari Island, Indonesia. *Tourism in Marine Environments* 19:1–21. <https://doi.org/10.3727/154427323X17026460566015>
2. Hirose E, Iskandar BH, Wardiatno Y. 2014. Photosymbiotic ascidians from Pari Island (Thousand Islands, Indonesia). *Zookeys* 422:1–10. <https://doi.org/10.3897/zookeys.422.7431>
3. Kurniawan TA, Pasaribu B, Kusworo TD, Wibisono Y, Goh HH, Zhang D, Ali I, Aziz F, Anouzla A, Sarangi PK, Mahmoud M, Othman MHD, Al-Sultan GA, Casila JCC. 2024. Building disaster resilience in thousand islands (Indonesia): unlocking climate adaptation strategies to navigate sea level rise in coastal regions while safeguarding crop productivity and local biodiversity. *ACS EST Water* 4:3213–3224. <https://doi.org/10.1021/acsestwater.4c00099>
4. Faizal I, Anna Z, Utami ST, Mulyani PG, Purba NP. 2022. Baseline data of marine debris in the Indonesia beaches. *Data Brief* 41:107871. <https://doi.org/10.1016/j.dib.2022.107871>
5. Chen CS, Lu CY, Jiang PL. 2014. Nutrient limitation-induced qualitative changes of fatty acids and caleosin expression in *Auxenochlorella protothecoides*. *Biotechnol Lett* 36:175–180. <https://doi.org/10.1007/s10529-013-1365-9>
6. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27:863–864. <https://doi.org/10.1093/bioinformatics/btr026>
7. Chen S. 2023. Ultrafast one - pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* 2. <https://doi.org/10.1002/imt2.107>
8. Zhao S, Tadesse BT, Gu L, Solem C. 2025. Draft genome sequence of *Brevibacterium linens* MA5, an isolate from Danbo cheese that carries the bacteriocin gene cluster. *Microbiol Resour Announc* 14:e0096325. <https://doi.org/10.1128/mra.00963-25>
9. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27:824–834. <https://doi.org/10.1101/gr.213959.116>
10. Mikheenko A, Saveliev V, Gurevich A. 2016. MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics* 32:1088–1090. <https://doi.org/10.1093/bioinformatics/btv697>
11. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60. <https://doi.org/10.1038/nmeth.3176>
12. Huson DH, Beier S, Flade I, Górski A, El-Hadidi M, Mitra S, Ruscheweyh H-J, Tappu R. 2016. MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput Biol* 12:e1004957. <https://doi.org/10.1371/journal.pcbi.1004957>