DATA NOTE



Prokaryotic communities profiling of Indonesian hot springs using long-read Oxford Nanopore sequencing



I Nengah Wirajana¹, Ni Putu Ariantari², Douglas J. H. Shyu³, Nilam Vaghamshi⁴, Komal Antaliya⁴ and Pravin Dudhagara^{1,4*}

Abstract

Objectives Indonesia's location at the convergence of multiple tectonic plates results in a unique geomorphological feature with abundant hot springs. This study pioneers the metagenomic exploration of Indonesian hot springs, harbouring unique life forms despite high temperatures. The microbial community of hot springs is taxonomically versatile and biotechnologically valuable. 16s rRNA amplicon sequencing of the metagenome is a viable option for the microbiome investigation. This study utilized Oxford Nanopore's long-read 16 S rRNA sequencing for enhanced species identification, improved detection of rare members, and a more detailed community composition profile.

Data description Water samples were taken from three hot springs of the Bali, Indonesia (i) Angseri, 8.362503 S, 115.133452 E; (ii) Banjar, 8.210270 S, 114.967063 E; and (iii) Batur, 8.228806 S, 115.404829 E. BioLit Genomic DNA Extraction Kit (SRL, Mumbai, India) was used to isolate DNA from water samples. The quantity and quality of the DNA were determined using a NanoDrop[™] spectrophotometer and a Qubit fluorometer (Thermo Fisher Scientific, USA). The library was created using Oxford Nanopore Technology kits, and the sequencing was done using Oxford Nanopore's GridION platform. All sequencing data was obtained in FASTQ files and filtered using NanoFilt software. This dataset is valuable for searching novel bacteria diversity and their existence.

Keywords 16s rRNA amplicon, Metagenome, Hot springs, Nanopore sequencing

*Correspondence:

Surat 395007, India

pravindudhagara@vnsgu.ac.in

¹Department of Chemistry, Faculty of Mathematics and Natural Science, Udayana University, Jimbaran-Badung, Bali 18036, Indonesia ²Department of Pharmacy, Faculty of Mathematics and Natural Science, Udayana University, Jimbaran-Badung, Bali 80361, Indonesia ³Department of Biological Science and Technology, National Pingtung University of Science and Technology, Neipu, Pingtung 912301, Taiwan ⁴Department of Biosciences, Veer Narmad South Gujarat University,

Objective

Indonesia lies at the intersection of the Ring of Fire and the Alpide belt, which fuels volcanic activity and geothermal heat, resulting in its abundance of hot springs [1]. Bali- an island in Indonesia, is well-known for its unique and diversified flora-fauna and hot springs. Hot springs in Indonesia are rich reservoirs of microbial life. Various researchers have investigated the hot springs for the discovery of novel thermophilic bacteria. However, the data about the exploration of thermophilic bacteria utilizing culturing and molecular techniques from the Indonesian hot springs is very low [2, 3]. Also, the metagenomic data aiming at microbial diversity identification is unavailable.



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http:// creativecommons.org/licenses/by-nc-nd/4.0/.

Pravin Dudhagara

Microbial profiling by the 16 S rRNA amplicon sequencing and shotgun metagenomic sequencing provides a comprehensive picture of the hot spring microbial community [4, 5] and leads to discovering the many novel and rare species, their metabolites, and biocatalysts [6].

Due to less cultiviability of the thermophiles, 16 S rRNA amplicon-based metagenomic analysis is the best way to determine the diversity of thermophilic bacteria living in hot springs [7]. It is suitable for the isolation of microbes having potential for the production of novel metabolites. Long-read 16 S rRNA gene amplicon sequencing using Oxford Nanopore Technologies (ONT) is better than other NGS platforms [8]. The long-reads sequencing using NGS has transformed microbiome taxonomic classification and profiling to understand microbial life and its potential for groundbreaking discoveries [9]. So, the 16 S rRNA amplicon sequencing data provides a window into the unseen world of microbes and offers invaluable insights into the composition, distribution, and potential of microbial communities. The present study explored the microbial diversity associated with three hot springs located in Bali, Indonesia. Data obtained from the present study may act as the benchmark for researchers aiming at mapping of microbial diversity associated with hot springs. It will also provide comprehensive view of the microbial community. The data is crucial for hot springs' health.

Data description

Sample collection

The sterile thermal bottles were used for the collection of water samples from three hot springs. Metadata like temperatures, pH, color and turbidity of water were recorded during sampling. Water samples were collected multiple times during the day in July and September 2023. Samples were brought to the laboratory on the same day. The samples were then pooled, and100 mL of each sample

Table 1	Overview of the	16s rRNA sequencing	data files/data
sets			

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or ac- cession number)
Data file 1	Angseri Hot Spring Metagenome	.fastq.gz	https://identi- fiers.org/ena.embl: ERX11741303. [14]. EMBL-EBI ENA database
Data file 2	Banjar Hot Spring Metagenome	.fastq.gz	https://identi- fiers.org/ena.embl: ERX11746727. [15]. EMBL-EBI ENA database
Data set 3	Batur Hot Spring Metagenome	.fastq.gz	https://identi- fiers.org/ena.embl: ERX12082498. [16]. EMBL-EBI ENA database

were filtered using a membrane filter, and retarded biomass on the filters was subjected to DNA isolation.

Metagenomic DNA extraction

The BioLit Genomic DNA Extraction Mini Kit (SRL, Mumbai, India) was used to isolate DNA from water samples from three hot springs. The quantity and quality of the DNA were determined using 0.8% agarose gel followed by a NanoDrop spectrophotometer and a Qubit fluorometer. After QC of isolated DNA, 50 μ l of each DNA sample was used for the sequencing.

16 S amplicon sequencing

The 16 S rRNA gene sequence libraries were created using the 16 S Rapid Amplicon Barcoding Kit (ONT, Oxford, UK) by following the manufacturer's instructions. LongAmp[®] Taq 2X master mix (New England Biolabs, Ipswich, USA) and the barcoded nanopore sequence primers 27 F 5'-AGA GTT TGA TCM TGG CTC AG-3' and 1492R: 5'-GGT TAC CTTGTT ACG ACT T-3' were used to amplify the full-length (1600 bp) 16 S rRNA gene. Following the quantification of 16 S rRNA gene amplicons, equal amounts of amplicons per sample were pooled, and the library was processed according to the manufacturer's instructions. After incubating the library with Library Loading Beads (ONT, Oxford, UK), the mixture was loaded into the GridION flow cell (version R.9.4, ONT, Oxford, UK). The GridION nanopore sequencer was used for 14 h of sequencing at PT. Genetika Science Indonesia (https://ptgenetika.com). Nanopore sequencing was operated by MinKNOW software version 23.04.5. Basecalling was performed using Guppy version 6.5.7 with a high-accuracy model [10].

Data processing

The output data (FASTQ files) generated more than 93,000 amplified sequences in each sample, subjected to QC using NanoPlot 1.40.0. Quality filtering was done using NanoFit 2.8.0. to obtain 0.34GB data in each sample with 1600 bp average sequence length in all three samples. The average sequence quality was 30 (Phred Score). Filtered reads were classified using the Centrifuge classifier [11]. The Bacteria and Archaea index was built using the NCBI 16 S RefSeq database [12]. Data is publicly available at EMBL-EBI ENA under the study ID PRJEB70710 (Table 1) [13]. The project is ongoing and no other data and analysis were published earlier.

Limitations

While 16 S rRNA sequencing successfully assigned taxonomy to the hot spring microbiome, it could not provide functional analysis of the microbes. 16 S rRNA sequencing doesn't detect fungi, viruses, or other non-bacterial/ archaeal organisms in the sample. Secondly, our sampling strategy involved collecting water samples multiple times throughout a single day in July and September 2023. The DNA was then isolated from pooled samples to capture a broader range of species. However, this approach only provides a snapshot and may not represent the seasonal variations within the hot spring's microbial community. To achieve a more comprehensive understanding of microbial dynamics, it would be ideal to collect water samples from all hot springs throughout the year.

Abbreviations

16S rRNA	16 S ribosomal ribonucleic acid
DNA	Deoxyribonucleic acid
NGS	Next generation sequencing
ONT	Oxford Nanopore technologies
QC	Quality control

Acknowledgements

We are thankful to Ni Made Yustikarini from the Department of Chemistry, Faculty of Mathematics and Natural Sciences at Udayana University for her support throughout the entire project.

Author contributions

INW, PD: Conceptualization, Methodology, funding acquisition and Investigation. INW, NPA: supervised the experiments and validation, NV, KA: performed the data submission and formal analysis. DJHS, PD: wrote and edited the data note.

Funding

This work was financially supported by the Udayana University International Senior Fellowship (UNISERF) grant for year 2023 (Grant Number: B/530-4/UN14.4 A/PT.01.03/2023), Udayana University, Bali, Indonesia.

Data availability

16s rRNA sequencing data files and data sets that support the findings of this study have been deposited in the European Nucleotide Archive with the primary accession code Bio Project: PRJEB70710. data can be access from given link https://www.ebi.ac.uk/ena/browser/view/PRJEB70710.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

Received: 20 June 2024 / Accepted: 12 September 2024 Published online: 02 October 2024

References

- Pambudi NA. Geothermal power generation in Indonesia, a country within the ring of fire: current status, future development and policy. Renew Sustain Energy Rev. 2018;81:2893–901. https://doi.org/10.1016/j.rser.2017.06.096.
- Lischer K, Putra ABRD, Guslianto BW, Avila F, Sitorus SG, Nugraha Y. The emergence and rise of indigenous thermophilic bacteria exploration from Hot Springs in Indonesia. Biodiversitas J Biol Divers. 2020;21(11):5474–81. https:// doi.org/10.13057/biodiv/d211156.
- Miyabayashi H, Tsuboi K, Sakai HD, Nur N, Suwanto A, Kurosawa N. Exploring of Culturable Novel Thermophiles from Indonesian Hot Springs by Enrichment Culture and 16S rRNA gene clone analysis. J Hot Spring Sciences/ Onsen Kagaku. 2021;71(1):38.
- Chan CS, Chan KG, Tay YL, Chua YH, Goh KM. Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. Front Microbiol. 2015;6:177. https://doi.org/10.3389/ fmicb.2015.00177.
- Mangrola AV, Dudhagara PR, Koringa PG, Josh CG, Patel RK. Metagenomic microbial community profiling of Unnai hot spring by Ion-Torrent based shotgun sequencing. Microbiology. 2018;87:143–46. https://doi.org/10.1134/ S0026261718010113.
- López-López O, Cerdán ME, González-Siso MI. Hot spring metagenomics. Life (Basel, Switzerland). 2013; 3(2): 308–20. https://doi.org/10.3390/life3020308
- Ghelani A, Patel R, Mangrola A, Dudhagara P. Cultivation-independent comprehensive survey of bacterial diversity in Tulsi Shyam Hot Springs, India. Genomics Data. 2015;4:54–6. https://doi.org/10.1016/j.gdata.2015.03.003.
- Oberle A, Urban L, Falch-Leis S, Ennemoser C, Nagai Y, Ashikawa K, Ulm PA, Hengstschläger M, Feichtinger M. Reprod Biomed Online. 2021;42(6):1097– 107. https://doi.org/10.1016/j.rbmo.2021.03.016. 165 rRNA long-read nanopore sequencing is feasible and reliable for endometrial microbiome analysis.
- Agustinho DP, Fu Y, Menon VK, Metcalf GA, Treangen TJ, Sedlazeck FJ. Unveiling microbial diversity: harnessing long-read sequencing technology. Nat Methods. 2024;21:954–66. https://doi.org/10.1038/s41592-024-02262-1.
- Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. 2019;20(1):129. https://doi. org/10.1186/s13059-019-1727-y.
- Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. 2016;26(12):1721–29. https://doi.org/10.1101/gr.210641.116.
- O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44(D1). https:// doi.org/10.1093/nar/gkv1189. ;D733-D745.
- EMBL-EBI ENA Bioproject. https://www.ebi.ac.uk/ena/browser/view/ PRJEB70710. (2023).
- 14. EMBL-EBI ENA database. https://identifiers.org/ena.embl:ERX11746727. (2024).
- EMBL-EBI ENA database. https://identifiers.org/ena.embl:ERX11746727. (2023).
- EMBL-EBI ENA database. https://identifiers.org/ena.embl:ERX12082498. (2024).

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.