

Review Article

Periodic survey by metagenomic analysis on the marine microbial communities in an enclosed bay locating at Sanriku coast off northern Japan in the Pacific Ocean

Shugo Watabe ^{1,*}, Md. Shaheed Reza ^{1,2}, Atsushi Kobiyama ¹, Kazuho Ikeo ^{1,3}, Jonaira Rashid ¹, Yuri Ikeda ¹, Yuichiro Yamada ¹, Daisuke Ikeda ¹, Nanami Mizusawa ¹, Shigeru Sato ¹, Takehiko Ogata ¹, Mitsuru Jimbo ¹, Toshiaki Kudo ¹, Shinnosuke Kaga ⁴, Shiho Watanabe ⁵, Kimiaki Naiki ⁵, Yoshimasa Kaga ⁵, Satoshi Segawa ⁵, Katsuhiko Mineta ⁶, Vladimir Bajic ⁶ and Takashi Gojorori ⁶

¹ School of Marine Biosciences, Kitasato University, Minami-ku, Sagamihara, Kanagawa 252-0373, Japan

² Department of Fisheries Technology, Bangladesh Agricultural University, Mymensingh-2202, Bangladesh

³ National Institute of Genetics, Yata, Mishima, Shizuoka 411-8540, Japan

⁴ Ofunato Fisheries Promotion Center, Iwate Prefectural Government, Ofunato, Iwate 022-8502, Japan

⁵ Iwate Fisheries Technology Center, Kamaishi, Iwate 026-0001, Japan

⁶ Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

*Correspondence: swatabe@kitasato-u.ac.jp; Tel.: +81-42-778-9094

Keywords: Bacterial diversity; Bioinformatics; Free-living bacteria; Shotgun metagenomics.

Received: 15 August 2017 / Accepted: 6 September 2017

© 2017 by the authors.

Introduction

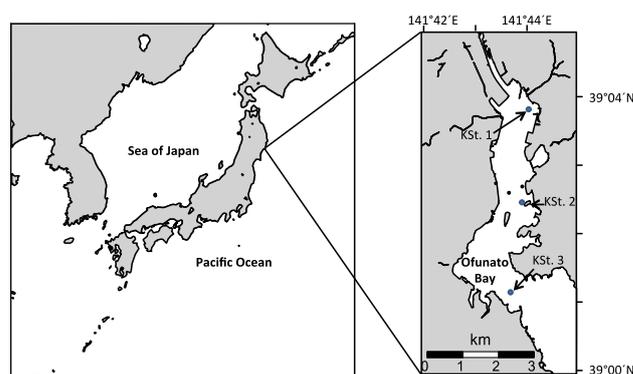
Japan's northeastern coastal and offshore areas in the Pacific Ocean with its temperate nature have the highest biodiversity of marine organisms in the world and the Ofunato Bay (an enclosed bay) at the Sanriku Rias Coast in the Iwate Prefecture is one of them. In contrast, the Red Sea is a model of the warmest ocean in the world and also holds one of the most diverse marine ecosystems, primarily due to its pristine coral reef ecosystem. Considering the economic and environmental importance of the Ofunato Bay, a joint research project was initiated in 2014 with King Abdullah University of Science and Technology (KAUST) to model, optimize and rationally design metabolic processes, proteins, and pathways in microbial cell factories by comparing metagenomes between the Ofunato Bay and the Red Sea which will later be used as an input for genome engineering.

Here we report our strategies for metagenomic analysis on the marine microbial communities in the Ofunato Bay together with datasets of metagenomic analyses on samples obtained from the bay.

Sampling location

Sampling stations are KSt. 1 at the innermost, KSt. 2 at the center and KSt. 3 at the entrance of the Ofunato Bay (Fig. 1). Seawaters were collected every month from September 2014 to December 2015 at 1 m and 8 (KSt. 1) or 10 m (KSt. 2 and KSt. 3) depths from the three

stations. Basic water quality parameters were recorded on the point using water quality profiler (RINKO AAQ176, JFE Advantech Co. Ltd., Japan) and other parameters including NO₂₊₃-N, NH₄-N, PO₄-P, SiO₂-Si and chlorophyll *a* were measured after bringing water samples into the laboratory of Iwate Fisheries Technology Center.



©Geospatial Information Authority of Japan

Fig. 1. Location of seawater sampling stations in the Ofunato Bay.

Seawater treatments

Collected seawaters were filtered sequentially through 5.0, 0.8 and 0.22- μ m filters (Fig. 2), while DNA was extracted only from the 0.22 μ m-filters in the present study, targeting free-living bacteria, and subsequently applied to a MiSeq sequencer (Illumina, San Diego, CA, USA) to obtain the whole genome sequencing reads (Fig. 3).

The JSFS 85th Anniversary-Commemorative International Symposium
“Fisheries Science for Future Generations”

Symposium Proceedings, No. 08003

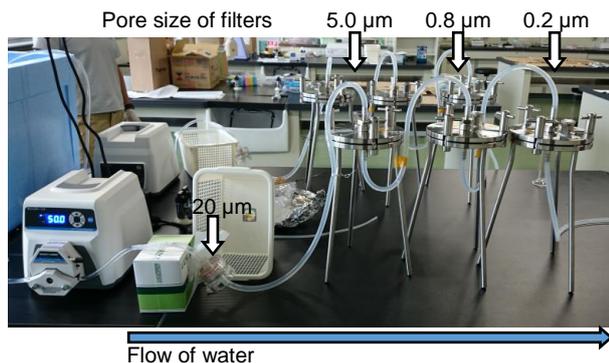


Fig. 2. Filtration of seawater.

Bioinformatic analysis and data processing

As shown in Fig. 3, acquired Illumina reads were uploaded to Meta Genome Rapid Annotation using Subsystem Technology (MG-RAST) server [1] and processed after joining overlapping paired-ends reads. These data were filtered by using QC pipelines [2] to remove replicated reads, since the platforms of high throughput sequencing (HTS) occasionally produce large numbers of reads that are nearly identical [3]. Only one representative read in the clusters of replicated reads, whose first 50 base pairs were identical, was preserved. The reads which contained 5 or more ambiguous bases were then removed. Typical datasets recently obtained between July 2016 and April 2017 from the three sampling stations and two water depths are shown in Table 1. The filtered reads were used for the following bioinformatic analyses.

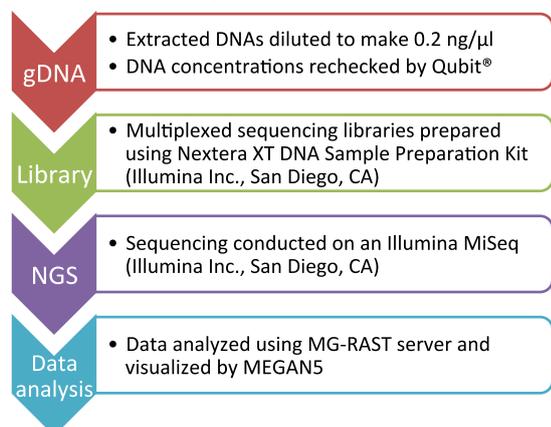


Fig. 3. Next generation sequencing and data processing.

Taxonomic analysis was then performed using MEGAN ver 5.10.3 after parsing the BLAST output, whereas comparative analysis in MEGAN was performed after normalizing the counts [4] (Fig. 3). The least common ancestor (LCA) assignment algorithm had the following parameters: min support = 1; min score = 50; top percent = 10 and maximum expected $1.0E^{-10}$. A taxonomic analysis of metagenomic datasets was also conducted in the MG-RAST server. Data were first uploaded to MG-RAST and processed after joining overlapping paired-ends reads and filtered using QC

pipelines to remove replicated reads. Taxonomic analysis was then conducted by comparison against the Ribosomal Database Project (RDP) [5]. Only alignments longer than 100 bp were considered for determining the minimum percent identity >95%. A hierarchical cluster dendrogram based on common genera and principal component analyses (PCA) based on the relative abundance of bacterial genera were carried out using R software [5]. The number of differentially abundant taxa among different months of the year was estimated, and these values were subjected to a Student's *t*-test ($p < 0.05$).

The data thus obtained will be shown in the following paper in this Proceedings.

Table 1. Datasets of metagenomic analyses on samples obtained from Ofunato Bay between July 2016 and April 2017 using MG-RAST statistics

| Sampling period | Station number | Water depth (m) | Uploaded bp count | Uploaded sequences count | Post QC sequence count | Post QC mean sequence length | Failed QC (%) | rRNA gene (%) | Predicted protein% (known) | Predicted protein% (unknown) |
|-----------------|----------------|-----------------|-------------------|--------------------------|------------------------|------------------------------|---------------|---------------|----------------------------|------------------------------|
| Jul-2016 | KSt. 1 | 1 | 345 M | 1.22 M | 1.14 M | 246+107 | 6.67 | 0.41 | 69.06 | 30.53 |
| | | 8 | 495 M | 1.84 M | 1.74 M | 244+108 | 5.54 | 0.43 | 68.36 | 31.21 |
| | | 10 | 425 M | 1.50 M | 1.39 M | 260+109 | 5.53 | 0.40 | 71.09 | 28.51 |
| KSt. 2 | 1 | 1 | 498 M | 1.78 M | 1.64 M | 240+104 | 8.25 | 0.40 | 54.31 | 45.28 |
| | | 8 | 465 M | 1.86 M | 1.75 M | 222+102 | 6.05 | 0.35 | 68.84 | 30.81 |
| | | 10 | 498 M | 2.10 M | 1.99 M | 215+104 | 5.48 | 0.43 | 64.84 | 34.73 |
| KSt. 3 | 1 | 1 | 652 M | 2.43 M | 2.27 M | 239+105 | 6.44 | 0.37 | 74.37 | 25.26 |
| | | 8 | 582 M | 2.21 M | 2.08 M | 235+101 | 5.90 | 0.43 | 67.77 | 31.79 |
| | | 10 | 482 M | 1.74 M | 1.64 M | 247+107 | 6.64 | 0.40 | 65.80 | 33.79 |
| Aug-2016 | KSt. 1 | 1 | 647 M | 2.17 M | 1.99 M | 259+108 | 8.24 | 0.36 | 75.77 | 23.87 |
| | | 8 | 473 M | 1.90 M | 1.78 M | 223+105 | 6.10 | 0.39 | 71.21 | 28.39 |
| | | 10 | 354 M | 1.34 M | 1.25 M | 232+105 | 6.52 | 0.39 | 64.16 | 35.46 |
| KSt. 2 | 1 | 1 | 642 M | 2.50 M | 2.40 M | 226+98 | 4.44 | 0.34 | 69.51 | 30.15 |
| | | 8 | 748 M | 2.81 M | 2.70 M | 247+101 | 3.74 | 0.37 | 67.71 | 31.92 |
| | | 10 | 280 M | 1.01 M | 0.96 M | 246+99 | 4.36 | 0.30 | 64.66 | 35.04 |
| KSt. 3 | 1 | 1 | 648 M | 1.63 M | 2.44 M | 238+101 | 3.66 | 0.37 | 65.78 | 33.85 |
| | | 8 | 469 M | 1.63 M | 1.56 M | 260+102 | 4.40 | 0.30 | 67.14 | 32.57 |
| | | 10 | 316 M | 1.16 M | 1.13 M | 254+105 | 2.70 | 0.31 | 64.66 | 35.03 |
| Oct-2016 | KSt. 1 | 1 | 205 M | 0.82 M | 0.79 M | 226+100 | 3.77 | 0.30 | 61.44 | 38.26 |
| | | 8 | 392 M | 1.76 M | 1.69 M | 208+100 | 3.86 | 0.33 | 61.33 | 38.34 |
| | | 10 | 720 M | 2.55 M | 2.45 M | 260+102 | 3.75 | 0.33 | 68.88 | 30.79 |
| KSt. 2 | 1 | 1 | 284 M | 0.98 M | 0.96 M | 267+104 | 2.49 | 0.31 | 62.44 | 37.25 |
| | | 8 | 621 M | 2.08 M | 1.97 M | 266+105 | 5.00 | 0.35 | 70.75 | 28.91 |
| | | 10 | 586 M | 1.96 M | 1.89 M | 273+105 | 3.72 | 0.32 | 66.05 | 33.63 |
| KSt. 3 | 1 | 1 | 310 M | 1.16 M | 1.14 M | 236+65 | 1.61 | 0.42 | 65.68 | 33.89 |
| | | 8 | 374 M | 1.43 M | 1.40 M | 236+66 | 1.74 | 0.42 | 62.69 | 36.89 |
| | | 10 | 280 M | 1.07 M | 1.05 M | 234+65 | 1.69 | 0.40 | 64.56 | 35.04 |
| Nov-2016 | KSt. 1 | 1 | 512 M | 2.07 M | 2.03 M | 226+67 | 2.08 | 0.43 | 64.51 | 35.06 |
| | | 8 | 392 M | 1.76 M | 1.69 M | 233+65 | 1.32 | 0.46 | 62.45 | 37.09 |
| | | 10 | 135 M | 0.51 M | 0.50 M | 222+67 | 2.24 | 0.38 | 65.06 | 34.56 |
| KSt. 2 | 1 | 1 | 418 M | 1.67 M | 1.63 M | 222+67 | 2.24 | 0.38 | 65.06 | 34.56 |
| | | 8 | 416 M | 1.86 M | 1.82 M | 206+69 | 2.12 | 0.41 | 64.41 | 35.56 |
| | | 10 | 297 M | 1.14 M | 1.12 M | 237+66 | 1.75 | 0.43 | 66.48 | 33.10 |
| KSt. 3 | 1 | 1 | 419 M | 1.57 M | 1.54 M | 236+65 | 1.90 | 0.39 | 64.58 | 35.03 |
| | | 8 | 315 M | 1.21 M | 1.19 M | 230+65 | 1.71 | 0.39 | 65.57 | 34.05 |
| | | 10 | 309 M | 1.18 M | 1.05 M | 267+105 | 3.28 | 0.42 | 71.22 | 28.36 |
| Jan-2017 | KSt. 1 | 1 | 187 M | 0.71 M | 0.69 M | 248+104 | 2.90 | 0.41 | 66.50 | 33.10 |
| | | 8 | 532 M | 2.10 M | 2.01 M | 238+102 | 4.13 | 0.42 | 70.94 | 28.64 |
| | | 10 | 833 M | 3.01 M | 2.94 M | 257+105 | 4.15 | 0.44 | 74.06 | 25.50 |
| KSt. 2 | 1 | 1 | 767 M | 3.06 M | 2.91 M | 239+104 | 4.84 | 0.41 | 73.29 | 26.30 |
| | | 8 | 194 M | 0.78 M | 0.76 M | 234+103 | 3.17 | 0.43 | 69.02 | 30.55 |
| | | 10 | 428 M | 1.54 M | 1.47 M | 257+102 | 4.48 | 0.43 | 73.24 | 26.33 |
| KSt. 3 | 1 | 1 | 825 M | 2.95 M | 2.82 M | 262+104 | 4.41 | 0.41 | 76.73 | 22.86 |
| | | 8 | 309 M | 1.08 M | 1.05 M | 267+105 | 3.28 | 0.42 | 71.22 | 28.36 |
| | | 10 | 650 M | 2.21 M | 2.11 M | 275+106 | 4.32 | 0.41 | 76.66 | 22.93 |
| Feb-2017 | KSt. 1 | 1 | 456 M | 1.54 M | 1.46 M | 270+105 | 5.43 | 0.39 | 73.84 | 25.77 |
| | | 8 | 361 M | 1.15 M | 1.11 M | 288+108 | 3.72 | 0.44 | 73.59 | 25.97 |
| | | 10 | 284 M | 0.89 M | 0.86 M | 290+107 | 3.68 | 0.43 | 72.29 | 27.28 |
| KSt. 2 | 1 | 1 | 254 M | 0.96 M | 0.96 M | 239+101 | 4.38 | 0.43 | 71.58 | 27.99 |
| | | 8 | 369 M | 1.23 M | 1.23 M | 272+106 | 4.83 | 0.41 | 74.93 | 24.66 |
| | | 10 | 742 M | 2.66 M | 2.66 M | 260+104 | 5.55 | 0.42 | 78.08 | 21.49 |
| KSt. 3 | 1 | 1 | 479 M | 1.49 M | 1.49 M | 290+110 | 4.83 | 0.43 | 75.01 | 24.56 |
| | | 8 | 309 M | 1.08 M | 1.05 M | 267+105 | 3.28 | 0.42 | 71.22 | 28.36 |
| | | 10 | 735 M | 2.59 M | 2.59 M | 256+103 | 5.91 | 0.41 | 75.75 | 23.84 |
| Mar-2017 | KSt. 1 | 1 | 332 M | 1.12 M | 1.12 M | 270+106 | 4.66 | 0.39 | 71.75 | 27.86 |
| | | 8 | 709 M | 2.57 M | 2.57 M | 256+104 | 7.72 | 0.37 | 72.04 | 27.59 |
| | | 10 | 386 M | 1.40 M | 1.40 M | 251+107 | 6.15 | 0.39 | 73.07 | 26.54 |
| KSt. 2 | 1 | 1 | 366 M | 1.11 M | 1.11 M | 295+111 | 4.51 | 0.42 | 73.03 | 26.55 |
| | | 8 | 762 M | 2.68 M | 2.68 M | 263+106 | 6.84 | 0.37 | 75.71 | 23.92 |
| | | 10 | 797 M | 2.77 M | 2.77 M | 267+105 | 4.23 | 0.45 | 72.51 | 27.04 |
| KSt. 3 | 1 | 1 | 762 M | 2.68 M | 2.68 M | 263+106 | 6.84 | 0.37 | 75.71 | 23.92 |
| | | 8 | 797 M | 2.77 M | 2.77 M | 267+105 | 4.23 | 0.45 | 72.51 | 27.04 |
| | | 10 | 797 M | 2.77 M | 2.77 M | 267+105 | 4.23 | 0.45 | 72.51 | 27.04 |

Acknowledgment

This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No URF/1/1976-03.

References

1. <http://metagenomics.nmpdr.org>
2. Meyer FD et al. (2008) BMC Bioinformatics 9: 386–393
3. Gomez-Alvarez VTK et al. (2009) ISME J 3: 1314–1317
4. Huson DH et al. (2007) Genome Res 17: 377–386
5. R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0