Accepted Manuscript

Metagenome-based diversity analyses suggest a strong locality signal for bacterial communities associated with oyster aquaculture farms in Ofunato Bay

Atsushi Kobiyama, Kazuho Ikeo, Md. Shaheed Reza, Jonaira Rashid, Yuichiro Yamada, Yuri Ikeda, Daisuke Ikeda, Nanami Mizusawa, Shigeru Sato, Takehiko Ogata, Mitsuru Jimbo, Toshiaki Kudo, Shinnosuke Kaga, Shiho Watanabe, Kimiaki Naiki, Yoshimasa Kaga, Katsuhiko Mineta, Vladimir Bajic, Takashi Gojobori, Shugo Watabe

PII: S0378-1119(18)30455-4
DOI: doi:[10.1016/j.gene.2018.04.073](https://doi.org/10.1016/j.gene.2018.04.073)
Reference: GENE 42797

To appear in: *Gene*
Received date: 25 April 2018
Accepted date: 25 April 2018

Please cite this article as: Atsushi Kobiyama, Kazuho Ikeo, Md. Shaheed Reza, Jonaira Rashid, Yuichiro Yamada, Yuri Ikeda, Daisuke Ikeda, Nanami Mizusawa, Shigeru Sato, Takehiko Ogata, Mitsuru Jimbo, Toshiaki Kudo, Shinnosuke Kaga, Shiho Watanabe, Kimiaki Naiki, Yoshimasa Kaga, Katsuhiko Mineta, Vladimir Bajic, Takashi Gojobori, Shugo Watabe, Metagenome-based diversity analyses suggest a strong locality signal for bacterial communities associated with oyster aquaculture farms in Ofunato Bay. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. *Gene*(2017), doi:[10.1016/j.gene.2018.04.073](https://doi.org/10.1016/j.gene.2018.04.073)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Metagenome-based diversity analyses suggest a strong locality signal for bacterial communities associated with oyster aquaculture farms in Ofunato Bay

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

^a*Kitasato University School of Marine Biosciences, Minami-ku, Sagami-hara, Kanagawa 252-0373, Japan;*

^b*Iwate Fisheries Technology Center, Kamaishi, Iwate 026-0001, Japan;*

^c*King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Center (CBRC), Thuwal 23955-6900, Kingdom of Saudi Arabia*

*Corresponding authors:

Professor Shugo Watabe, Department of Marine Biochemistry, Kitasato University School of Marine Biosciences, Minami-ku, Sagami-hara, Kanagawa 252-0373, Japan.

E-mail: swatabe@kitasato-u.ac.jp

Professor Takashi Gojobori, Computational Bioscience Research Center and Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia

Email: takashi.gojobori@kaust.edu.sa

Present addresses: ¹Department of Fisheries Technology, Bangladesh Agricultural University, Mymensingh-2202, Bangladesh; ²Ofunato Fisheries Promotion Center, Iwate Prefectural Government, Ofunato, Iwate 022-8502, Japan.

Running title: Metagenomic bacterial communities in oyster farms of Ofunato Bay

Abstract

Ofunato Bay, in Japan, is the home of buoy-and-rope-type oyster aquaculture activities. Since the oysters filter suspended materials and excrete organic matter into the seawater, bacterial communities residing in its vicinity may show dynamic changes depending on the oyster culture activities. We employed a shotgun metagenomic technique to study bacterial communities near oyster aquaculture facilities at the center of the bay (KSt. 2) and compared the results with those of two other localities far from the station, one to the northeast (innermost bay, KSt. 1) and the other to the southwest (bay entrance, KSt. 3). Seawater samples were collected every month from January to December 2015 from the surface (1 m) and deeper (8 or 10 m) layers of the three locations, and the sequentially filtered fraction on 0.2- μ m membranes was sequenced on an Illumina MiSeq system. The acquired reads were uploaded to MG-RAST for KEGG functional abundance analysis, while taxonomic analyses at the phylum and genus levels were performed using MEGAN after parsing the BLAST output. Discrimination analyses were then performed using the ROC-AUC value of the cross validation, targeting the depth (shallow or deep), locality [(KSt. 1+KSt. 2) vs. KSt. 3; (KSt. 1+KSt. 3) vs. KSt. 2 or the (KSt. 2+KSt. 3) vs. KSt. 1] and seasonality (12 months). The matrix discrimination analysis on the adjacent 2 continuous seasons by ROC-AUC, which was based on the datasets that originated from different depths, localities and months, showed the strongest discrimination signal on the taxonomy matrix at the phylum level for the datasets from July to August compared with those from September to June, while the KEGG matrix showed the strongest signal for the datasets from March to June compared with those from July to February. Then, the locality combination was subjected to the same ROC-AUC discrimination analysis, resulting in significant differences between KSt. 2 and KSt. 1+KSt. 3 on the KEGG matrix. These results suggest that aquaculture activities markedly affect bacterial functions.

1. Introduction

Ofunato Bay is an enclosed, deep water bay located in the northeast of Japan, Iwate Prefecture, at the Sanriku Rias Coast. With an average depth of 10.3 m at the inner part and 38.5 m at the entrance, the bay is famous for its intricate coastline (Okaichi, 2004; Yamada et al., 2017). Chlorophyll *a* (chl-*a*) concentrations have been reported to be high in the bay during spring (JAXA, 2004; Reza et al., 2018b). With its enclosed nature and a breakwater that narrows down at the entrance to a width of 200 m (Yamada et al., 2017), the bay has a relatively calm environment and is considered suitable for buoy-and-rope-type aquaculture facilities (Komatsu et al., 2012). Most aquaculture activities are concentrated in the central part of the bay, where the giant Pacific oyster (*Crassostrea gigas*) is reared for two years. They filter suspended materials as food at an estimated rate of 1.24 tons C/day and excrete organic matter into the seawater at a rate of 1.05 tons C/day (Terasawa et al., 1995), indicating their substantial roles in the biogeochemical cycle of Ofunato Bay. Although the annual bivalve production of the bay has been halved since the 1970s to below 400 mt, primarily due to the construction of the breakwater and the massive tsunami generated by the 2011 Tohoku Earthquake off the Pacific coast (Miyazawa and Hayakawa, 1994; Yamada et al., 2017), the roles of the bivalves in influencing the planktonic and microbial communities are still considered to be vital to the ecology of the bay. Considering the importance of bivalve aquaculture and other economic activities in the bay, several studies have been conducted on its planktonic populations, mainly focusing on the surroundings of the oyster rafts and nearby areas (Ogata et al. 1982; Yamada et al., 2017). However, detailed analyses on the microbial communities in Ofunato Bay and their functional roles have remained unclear. Accordingly, we performed metagenomic analysis on the taxonomic profiles of the free-living microbial community in the entire length of Ofunato Bay, revealing 29 bacterial phyla (~80% of assigned reads), 3 archaeal phyla (~4%) and 59 eukaryotic phyla (~15%) (Reza et al., 2018a) in the bay. Furthermore, we investigated the year-round variation in free-

living bacterioplankton and showed phylogenetic differences among the bacterioplankton in the spring, summer, autumn and winter, suggesting that members of Proteobacteria tended to decrease at high water temperatures and increase at low temperatures (Reza et al., 2018b).

In this study, we analyzed the metagenomics datasets, mainly focusing on the bacterial communities near the oyster culture facilities in the center of Ofunato Bay and compared them with those of two other localities far from the station, one in the northeast (innermost bay) and the other in the southwest (bay entrance). The obtained results suggest that aquaculture activities markedly affect bacterial functions in the community.

2. Materials and methods

2.1. Site information

Three sampling stations in Ofunato Bay were selected along the length of the bay: KSt. 1 (141.73457E, 39.063370N; avg. depth 10.3 m), KSt. 2 (141.73245E, 39.044612N; avg. depth 25.3 m) and KSt. 3 (141.72820E, 39.019030N; avg. depth 38.5 m), where KSt. 1 and KSt. 3 are located in the northeast (innermost bay) and southwest (bay entrance) parts of the bay, respectively, while KSt. 2 is at the center of the bay and around the highly dense oyster aquaculture facilities (Reza et al., 2018a).

2.2. Sample collection

Sample collection and processing were performed according to Reza et al. (2017a). Briefly, seawaters were collected monthly from January to December 2015 using a vertical water sampler from 1 m (KSt. 1, KSt. 2 and KSt. 3) and 8 (KSt. 1) or 10 m (KSt. 2 and KSt. 3) below the surface. The seawater samples were then filtered sequentially through 20- μ m-pore sized Nylon Net Filters NY20 with 47 mm diameter (Merck Millipore Ltd., Tullagreen, Ireland) and each of 5, 0.8 and 0.2- μ m Isopore Membrane Filters with 142 mm diameter (Merck Millipore Ltd.).

2.3. Analytical procedures

Genomic DNA was then extracted from the 0.2- μ m filters using a PowerWater® DNA Isolation Kit (MoBio Laboratories Inc., Carlsbad, CA, USA), following the manufacturer's recommendations. The extracted DNA was used for metagenomic shotgun sequencing with an Illumina MiSeq (Illumina Inc., San Diego, CA, USA). The whole-genome sequence (WGS) datasets have been registered in the DDBJ Sequence Read Archive under the accession number DRA005744. Acquired Illumina paired reads from each library were first joined by overlapping forward and reverse reads of the same DNA fragment (paired-end

sequences), using the software FLASH (Magoc and Salzberg, 2011) with default parameters (overlap minimum, 10 nt; maximum allowed ratio between the number of mismatched base pairs and the overlap length, 0.25).

These WGS reads were subjected to quality filtering by removing <50 bp and quality trimmed to Phred 20 using the Genomics Workbench (CLCbio, Cambridge, MA, USA). BLASTn was performed locally in a server using the quality- and size-filtered sequences against the NCBI-nt reference database. Taxonomic analysis at the phylum and genus levels was then performed using MEGAN ver 5.10.3 (Huson et al., 2007), after parsing the BLAST output. Comparative analysis in MEGAN was also performed after normalizing the counts according to the recommendations of the authors. The least common ancestor (LCA) assignment algorithm used the following parameters: min support = 1; min score = 50; top percent = 10 and maximum expected $1.0E - 10$.

Functional analyses of the WGS datasets targeting the mining of the rRNA gene sequences were conducted in the Meta Genome Rapid Annotation Subsystem Technology (MG-RAST) server (<http://metagenomics.anl.gov/>). Data were uploaded to MG-RAST, processed after joining and overlapping the paired-ends reads, and filtered using QC pipelines (Meyer et al., 2008) to remove replicated reads. Only alignments longer than 100 bp were considered to have a minimum percent identity >95%. Hierarchical functional predictions were performed using Kyoto Encyclopedia of Genes and Genomes (KEGG) Ontology (KO) (<http://www.genome.jp/kegg/>) (Kanehisa et al., 2004).

Discrimination analyses were finally performed by the Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) value of the cross validation using the stratified k-fold method (Kohavi, 1995), targeting the two depths (1 m and 8 or 10 m), three stations (KSt. 1, KSt. 2 and KSt. 3) and 12 months (from January to December 2015) for seasonality (12 months to adjacent continuous 2 seasons) and local combination [(KSt. 1+KSt. 2) vs. KSt. 3; (KSt. 1+KSt. 3) vs. KSt. 2; or (KSt. 2+KSt. 3) vs. KSt. 1].

Statistical analyses were also performed by employing the non-parametric Mann-Whitney U-test.

ACCEPTED MANUSCRIPT

3. Results and discussion

3.1. Seasonal changes revealed by the taxonomical analyses

Seasonal changes in the abundances of heterotrophic bacteria and cyanobacteria, as well as in the free-living bacterial community in Ofunato Bay, have already been reported in our previous report (Reza et al., 2018b). This study focused on KSt. 2, where oyster aquaculture activities are concentrated, using the same datasets from January to December 2015. In Ofunato Bay, winter is usually delayed due to the late arrival of the cold *Oyashio Current* from the north, and in our study, we found the seawater temperature in March 2015 to be lowest, ranging from 4.6 to 6.4°C at the three sampling stations (Reza et al., 2018b; Rashid et al., 2018). Thus, May, August, November and March are spring, summer, autumn and winter, respectively, in the bay.

Seasonal changes in the relative abundances of the major bacterial phyla at the 1 m and 10 m depths at KSt. 2, based on the metagenomic shotgun sequencing datasets, are shown in Fig. 1, whereas those assigned to specific genera are shown in Supplementary Fig. S1.

Diverse communities of free-living bacteria were apparent at KSt. 2 in Fig. 1 at the phylum level, exhibiting inter-seasonal differences from January to December 2015 as reported previously, where the bacterial communities were reported at the genus level (Reza., 2018b). The most frequently available bacterial sequences belonged to Proteobacteria (Alpha, Beta and Gamma), followed by Bacteroidetes, collectively accounting for ~85% of all the WGS bacterial reads (Fig. 1). Other phyla included Firmicutes, Actinobacteria, Cyanobacteria and Tenericutes, consisting of ~1% by read counts.

Alphaproteobacteria was the most dominant bacterial phylum at the 1 m and 10 m depth through the year and remained high between May and August at the 1 m depth and between May and Jun at the 10 m depth (Fig. 1). At the genus level, *Planktomarina* and *Candidatus Pelagibacter*, belonging to Alphaproteobacteria, were found to be the most dominant (Supplementary Fig. S1). Similarly, Gammaproteobacteria was abundant at both water depths

throughout the year and included *Shewanella* and *Alteromonas*. Betaproteobacteria, which are mostly freshwater bacteria (Zwart et al., 2002; Gucht et al., 2005), were marginal at KSt. 2 and were less than 1% of the WGS reads. Bacteroidetes were the second largest bacterial phylum found in Ofunato Bay and were dominated by a diverse clade, including *Polaribacter*, *Flavobacterium*, *Sphingobacterium* and *Cellulophaga*. Many genera of this phylum, including *Polaribacter*, were found to be high during September. Actinobacteria accounted for 1.5 – 2.1% of the WGS reads, with *Candidatus Actinomarina* as the most dominant genus, while Cyanobacteria constituted approximately 1 to 10% of the WGS reads. The abundance of Cyanobacteria increased in the 10 m depth during the higher temperature periods of the year, where *Synechococcus* was found to be the most dominant. Other minor phyla found in the KSt. 2 station included Aquificae, Planctomycetes, Chlamydiae, and Chlorobi.

3.2. Functional analyses in the bacterial communities by KEGG

Then, we analyzed changes in the core functional traits using KO for KSt. 2 (Fig. 2). It was revealed that >30% of the WGS reads encoded proteins involved in "metabolism," followed by >10% for "environmental information processing," which were among the first-tier KO categories at both the surface and deep waters. Significant differences were observed among the first-tier functional categories between the 1 m and 10 m depth datasets for "metabolism" and "environmental information processing" (data not shown). When the abundances of the second- and third-tier KO functional gene annotations were considered, "translation", "amino acid metabolism", and "metabolism of cofactors and vitamins" were significantly higher ($p < 0.05$) in the surface than the deep waters (Table 1). In contrast, the functional abundance of "energy metabolism" was significantly higher ($p < 0.05$) in the deep waters than the surface waters. It was noted that the KO category of organismal systems was apparent only in February and between April and September at the 1 m depth, while it appeared between March and June at the 10 m depth (Fig. 2).

3.3. Discrimination analyses by ROC-AUC

To identify possible locality-specific differences among the stations KSt. 1, KSt. 2 and KSt. 3 along the length of Ofunato Bay, we employed discrimination analyses using both taxonomic and functional datasets obtained from the three sampling stations. Discrimination analyses were performed by ROC-AUC value of the cross validation using the stratified k-fold method (Kohavi, 1995), targeting the two depth (1 m and 8 or 10 m), three stations (KSt. 1, KSt. 2 and KSt. 3) and 12 months (from January to December 2015) for seasonality (12 months to adjacent continuous 2 seasons) and local combination [(KSt. 1+KSt. 2) vs. KSt. 3; (KSt. 1+KSt. 3) vs. KSt. 2; or (KSt. 2+KSt. 3) vs. KSt. 1].

First, we performed the matrix discrimination analysis for seasonality at the phylum and genus levels and at the KO level. In this analysis, the 12 months were divided into two provisional seasons with any combinations of two adjacent months. At the taxonomic levels, the highest ROC-AUC value (0.933) was obtained when comparing July – August and other months. Significant values were also obtained in the comparisons including January, February, May, July, October and December. Non-parametric Mann-Whitney U-tests for the discrimination that provided the highest value between June – August and other months revealed that Verrucomicrobia was the most important factor for the matrix discrimination, followed by Acidobacteria, Spirochaetes, Firmicutes and Deinococcus-Thermus at the phylum level (Supplementary Table S1). Verrucomicrobia have been found in Arctic marine bacterial communities (Galand et al., 2009; Cardman et al., 2014) and are also well known to inhabit marine, freshwater and soil environments (Zwart et al., 1998; Liles et al., 2003; Freitas et al., 2012). In the case of the genus level, bacteria belonging to unclassified Verrucomicrobia were again the most discriminating factor, followed by those belonging to unclassified Proteobacteria, *Coralimargarita*, which also belong to Verrucomicrobia (Yoon et al., 2007), and *Akkermansia* (Supplementary Table S2).

In the KEGG matrix, the highest significance was obtained for March-June against the other months (0.977). Significant values were also obtained for other combinations in the two provisional seasons, including those implicated from March to July. Non-parametric Mann-Whitney U-tests demonstrated that glycerophospholipid metabolism was the most discriminating factor and had the highest value from March-June compared with other months, followed by photosynthesis, valine, leucine and isoleucine degradation, RNA transport, and photosynthesis - antenna proteins (Supplementary Table S3). It is noted that these factors belong to either amino acid metabolism, energy metabolism, or translation in the KO first tier (level 1) by the KEGG analyses, which all showed significant differences between the two different depths of 1 m and 8 m at KSt. 2 (Table 1).

ROC-AUC analyses were then performed for local combinations in the taxonomy and KEGG matrices. No significant locality combination was observed in the taxonomy matrix (Table 2). However, KSt. 2 showed highly significant differences from KSt. 1 plus KSt. 2 in the KEGG matrix. A non-parametric Mann-Whitney U-test revealed that the gap junction was the most responsible for this discrimination, followed by the cell cycle, endocytosis, lysosome and peroxisome (Supplementary Table S4). It is curious that the lysosome and peroxisome are characteristics of eukaryotes, suggesting a limitation of our analysis. Despite such uncertainty, the obtained data contain apparent bacteria-specific factors, including flagellar assembly and bacterial chemotaxis, as discriminating factors, along with those associated with metabolism. These results indicate that functional genes have important roles in discriminating KSt. 2 from KSt. 1 and KSt. 3, possibly in association with bacterial activities.

In this study, we analyzed metagenomic datasets, mainly focusing on the bacterial communities near the highly dense oyster culture facilities in the center of Ofunato Bay and compared them with those of two other localities far from the station, one in the northeast (the innermost bay) and the other in southwest (the bay entrance). ROC-AUC discrimination

analysis demonstrated the significant differences between KSt. 2 and KSt. 1 plus KSt. 3 on the KEGG matrix, suggesting that aquaculture activities markedly affect bacterial functions. Further analyses are needed to better understand the relationship between bacterial functions and the availability of nutrients and other environmental factors around aquaculture facilities.

Acknowledgments

We thank Toshitsugu Okayama, of the National Institute of Genetics, Japan, for assisting us with the ROC-AUC analysis. 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.

References

- Cardman, Z., Arnosti, C., Durbin, A., Ziervogel, K., Cox, C., Steen, A. D., Teske, A., 2014. *Verrucomicrobia* are candidates for polysaccharide-degrading bacterioplankton in an Arctic Fjord of Svalbard. *Appl. Environ. Microbiol.* 80, 3749-3756.
- Freitas, S., Hatosy, S., Fuhrman, J. A., Huse, S. M., Welch, D. B. M., Sogin, M. L., Martiny, A. C., 2012, Global distribution and diversity of marine *Verrucomicrobia*. *ISME J.*, 6, 1499–1505.
- Galand, P. E., Casamayor, E. O., Kirchman, D. L., Love-joy, C., 2009. Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc. Natl. Acad. Sci.*, 106, 22427–22432.
- Huson, D. H., Auch, A. F., Qi, J., Schuster, S. C., 2007. MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386.
- JAXA, 2004. Images and Data Seen from Space by Japan Aerospace Exploration Agency, January 26, 2004 (<http://www.eorc.jaxa.jp/en/imgdata/topics/2004/tp040126.html>)
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., Hattori, M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280.
- Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence* (pp. 1137–1143). San Francisco, CA: Morgan Kaufmann.
- Komatsu, T., Sagawa, T., Sawayama, S., Tanoue, H., Mohri, A., Sakanishi, Y., 2012. Mapping is a key for sustainable development of coastal waters: examples of seagrass beds and aquaculture facilities in Japan with use of ALOS images, sustainable development - education, business and management - In: *Architecture and Building Construction - Agriculture and Food Security*, Chaouki Ghenai (Ed.), ISBN: 978-953-51-0116-1.

- Liles, M. R., Manske, B. F., Bintrim, S. B., Handelsman, J., Goodman, R. M., 2003. A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Appl. Environ. Microbiol.* 69, 2684- 2691.
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A., 2008. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9, 386–393.
- Magoc, T., Salzberg, S.L., 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963.
- Miyazawa, K., Hayakawa, Y. 1994. A large-scale structure and the environmental capacity for aquaculture – effects of an artificial structure (breakwater at the entrance of the Ofunato estuary) on aquaculture and water qualities in the estuary. *Bul. Coast. Oceanogr.* 32, 29–38 (in Japanese, with English abstract).
- Okaichi, T., 2004. Red-tide species and the environmental conditions. In: *Red Tides* (Okaichi, T., ed.). Terra Scientific Publishing Company, Tokyo, Japan, and Kluwer Academic Publishers, London, Boston, p. 367.
- Ogata, T., Kodama, M., Fukuyo, Y., Inoue, T., Kamiya, H., Matsuura, F., Sekiguchi, K., Watanabe, S., 1982. The occurrence of *Protogonyaulax* spp. in Ofunato Bay, in association with the toxification of the scallop (*Patinopecten yessoensis*). *Bull. Jpn. Soc. Sci. Fish.* 48, 563–566.
- Rashid, J., Kobiyama, K., Reza, M. S., Yamada, Y., Ikeda, Y., Ikeda, D., Mizusawa, N., Ikeo, K., Sato, S., Ogata, T., Kudo, T., Kaga, S., Watanabe, S., Naiki, K., Kaga, Y., Mineta, K., Bajic, V., Gojobori, T., Watabe, S., 2018. Seasonal changes in the communities of photosynthetic picoeukaryotes in Ofunato Bay as revealed by shotgun metagenomic sequencing. *Gene* (under review).

- Reza, M. S., Kobiyama, A., Yamada, Y., Ikeda, Y., Ikeda, D., Mizusawa, N., Ikeo, K., Sato, S., Ogata, T., Jimbo, M., Kudo, T. Kaga, S., Watanabe, S., Naiki, K., Kaga, Y., Mineta, K., Bajic, V., Gojobori, T., Watabe, S., 2018a. Taxonomic profiles in metagenomic analyses of free-living microbial communities in the Ofunato Bay (under review).
- Reza, M. S., Kobiyama, A., Yamada, Y., Ikeda, Y., Ikeda, D., Mizusawa, N., Ikeo, K., Sato, S., Ogata, T., Jimbo, M., Kudo, T., Kaga, S., Watanabe, S., Naiki, K., Kaga, Y., Mineta, K., Bajic, V., Gojobori, T., Watabe, S., 2018b. Basin-scale seasonal changes in marine free-living bacterioplankton community in the Ofunato Bay (under review).
- Terasawa, T., Nakata, K., Taguchi, K., 1995. Numerical modeling on the Ofunato Bay ecosystem including the oyster farming. In: Proceedings of 4th International Conference on Estuarine and Coastal Modeling, ASCE, pp. 530–541.
- Yamada, Y., Kaga, S., Kaga, Y., Naiki, K., Watanabe, S., 2017. Changes of seawater quality in Ofunato Bay, Iwate, after the 2011 off the Pacific coast of Tohoku Earthquake. *J. Oceanogr.* 73, 11–24.
- Yoon, J., Yasumoto-Hirose, M., Katsuta, A., Sekiguchi, H., Matsuda, S., Kasai, H., Yokota, A., 2007. *Coralimargarita akajimensis* gen. nov., sp. nov., a novel member of the phylum 'Verrucomicrobia' isolated from seawater in Japan. *Int J Syst Evol Microbiol* 57, 959–63.
- Zwart, G., Crump, B. C., Kamst-Van Agterveld, M. P., Hagen, F., Han, S. K., 2002. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat. Microbiol. Ecol.* 28, 141–155.
- Zwart, G., Huismans, R., van Agterveld, M. P., van de Peer, Y., De Rijk, P., Eenhoorn, H., Muyzer, G., van Hanne, E. J., Gons, H. J., Laanbroek, H. J., 1998. Divergent members of the bacterial division *Verrucomicrobiales* in a temperate freshwater lake. *FEMS Microbiol. Ecol.* 25, 159–169.

Figure Legends

Fig. 1. The relative abundance of major bacterial phyla at the 1 m and 10 m depths at KSt. 2 around the oyster aquaculture facilities in Ofunato Bay based on metagenomic shotgun sequencing. Phylogenetic groups that account for <1% of the WGS reads were summarized in the artificial group “Less abundant phyla”.

Fig. 2. Percentages of first-tier KO functions among all the shotgun metagenomic functional predictions obtained from KSt. 2.

Fig. 3 Measured values for the discrimination in the taxonomy (phylum plus genus) and KEGG matrices in seasonality (two groups for 12 months to adjacent continuous 2 seasons) evaluated by ROC-AUC of the cross validation with the stratified k-fold method. Targeted datasets were from the two depths (1 m and 8 or 10 m), three stations (KSt. 1, KSt. 2 and KSt. 3), and 12 months from January to December 2015 in Ofunato Bay.

Title: Metagenome-based diversity analyses suggest a strong locality signal for bacterial communities associated with oyster aquaculture farms in Ofunato Bay
Atsushi Kobiyama et al.

Abbreviations:

KEGG: Kyoto Encyclopedia of Genes and Genomes

MG-RAST: Meta Genome Rapid Annotation Subsystem Technology

ROC-AUC: Receiver Operating Characteristic-Area Under the Curve

WGS: whole-genome sequence

ACCEPTED MANUSCRIPT

Table 1

Correlations of predicted second-tier functional abundances between the 1 m and 10 m metagenomic datasets from KSt. 2 in Ofunato Bay.

First- tier (Level 1)	Second- tier (Level 2)	KSt2-1m	KSt2-10m	<i>p</i> -value
Metabolism	Amino acid metabolism	8.34±5.54	0.00±0.00	0.000
	Biosynthesis of other secondary metabolites	0.00±0.01	0.00±0.00	0.339
	Carbohydrate metabolism	5.92±4.98	6.96±5.01	0.613
	Energy metabolism	4.72±11.02	34.48±8.19	0.000
	Glycan biosynthesis and metabolism	0.53±0.54	0.85±0.71	0.230
	Lipid metabolism	1.84±1.60	1.80±1.63	0.958
	Metabolism of cofactors and vitamins	4.39±3.08	1.90±1.28	0.021
	Metabolism of other amino acids	4.92±6.78	9.04±7.62	0.177
	Metabolism of terpenoids and polyketides	0.45±0.47	0.93±0.94	0.133
	Nucleotide metabolism	0.24±0.28	0.44±0.50	0.254
Genetic information processing	Xenobiotics biodegradation and metabolism	0.32±0.81	0.07±0.12	0.303
	Folding, sorting and degradation	5.28±4.26	3.70±2.78	0.295
	Replication and repair	1.89±1.91	1.81±1.54	0.908
	Transcription	4.40±3.58	3.30±2.48	0.393
Environmental information processing	Translation	31.01±12.80	14.21±4.70	0.001
	Membrane transport	4.47±3.06	5.22±4.22	0.625
	Signal transduction	0.94±0.66	1.38±1.05	0.241
	Signaling Molecules and Interaction	0.00±0.00	0.09±0.15	0.059
Cellular processes	Cell communication	1.22±1.33	1.28±1.02	0.890
	Cell growth and death	2.40±2.18	2.52±1.83	0.891
	Cell motility	0.06±0.09	0.08±0.07	0.604
	Transport and catabolism	1.00±0.99	1.57±1.63	0.314

*Bold faced categories are highlighted to indicate significantly different second-tier functional abundances

Table 2

Measured values for the discrimination for seasonality between the two groups (12 months to adjacent continuous 2 seasons) evaluated by ROC-AUC of the cross validation for the taxonomy (phylum and genus) and KEGG matrices with the datasets from the two depth (1 m and 8 or 10 m), three stations (KSt. 1, KSt. 2 and KSt. 3) and 12 months (from January to December 2015).

Locality combination	Taxonomy matrix	KEGG matrix
KSt. 1 vs. [KSt. 2 + KSt. 3]	0.348	0.505
KSt. 2 vs. [KSt. 1 + KSt. 3]	0.245	0.999
KSt. 3 vs. [KSt. 1 + KSt. 2]	0.607	0.627

Figure 1

Figure 2

Figure 3