Metagenomic analysis of Microbial Diversity of Tropical Sea Water of Georgetown Coast, Malaysia

Ramitha Arumugam[1] , Xin-Yue Chan[2], Wai Fong Yin[2], Siew Woh Choo[1*] and Kok Gan Chan [2*]

1. Dental Research and Training Unit, Faculty of Dentistry, University of Malaya, 50603 Malaysia. Tel: +60379676463. Email: lchoo@um.edu.my
2. Institute of Biological Sciences (Division of Genetics and Molecular Biology), Faculty of Science, University of Malaya, 50603 Malaysia. Tel: +603-79675162, Fax: +603-79674509. Email: kokgan@um.edu.my

Abstract: We studied using metagenomic approach, the sea surface of using Illumina MiSeq with the aim to uncover the microbial diversity. A total of 6,701,060 reads were generated from the shotgun sequencing. These raw sequences were trimmed and BLASTed against NCBI NR environmental database using the built-in BLASTN algorithm of the Cyber infrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA) portal. A total of 24 eubacteria phyla were identified, and of these, *Proteobacteria* made up the largest division followed by *Bacteroidetes* and *Actinobacteria*. Meanwhile, there were 5 phyla of Archaea identified namely *Eurkarchaeota*, *Thaumarchaeota*, *Crenarchaeota, Karoacheato* and *Nanoarchaeota* together with some unclassified Archaea 16S genes. Although a high diversity of microorganisms is observed, but more sequencing work needs to be done for better understanding on the microbes occurs as the DNA are yet to be sequence to saturation.
[Ramitha Arumugam, Xin-Yue Chan, Wai Fong Yin, Siew Woh Choo and Kok Gan Chan. **Metagenomic analysis of Microbial Diversity of Tropical Sea Water of Georgetown Coast, Malaysia.** *Life Sci J* 2013;10(3):2392-2396] (ISSN: 1097-8135). http://www.lifesciencesite.com. 348

Keywords: *Bacteroidetes* and *Actinobacteria*; microbial diversity and shotgun sequencing

Introduction

Prokaryote that includes bacteria and archaea is the most abundant organisms on Earth where soil and ocean were reported to have the highest numbers of these organisms. Based on previous studies, the oceanic subsurface has approximately 355×10^{28} prokaryotes cells whereas there are $>10^5$ cfu/ml of water on sea surface [1,2]. The metagenomic project on the Sargasso Sea surface water have found more than 1000 distinct rRNA genes in this warm sea by using different assemble protocols. Among them, Proteobacteria and Chlorobi make up the largest group of bacteria whereas Euryarchaeota and Thaumarchaeota contributes a large piece to the archaeal diversity found in the Sargasso Sea surface [3]. These microorganisms play very important roles in biogeochemical cycles and performing bioremediation work for the ocean [4-6].

Our previous studies (unpublished data) on the marine microbial diversity of Straits of Malacca via 16S rRNA amplicon-based sequencing revealed the occurrence of vast diversity of microbe in the water surface of tropical coast. Even though 16S rRNA sequencing enable profiling of bacterial community in a greater detail, in this study we opted for whole genome shotgun sequencing approach to omit PCR bias [3,7].

Our sampling site is situated at the tropical coast located at Georgetown, Malaysia. Georgetown is the capital of Penang state, currently the city center and the tourism area of the second metropolitan of Malaysia [8,9]. We believe that the diversity of prokaryotes in the tropical sea will be as interesting as the diversity of eukaryotes harbor in this warm protected strait. The curiosity towards the marine microbiome of the warm, calm, nutrient rich and protected straits of Malacca, has leaded us to this project.

The invention of high-throughput sequencing platform had enhanced the power for metagenomic studies [10]. In addition to this, the launching of MiSeq® by Illumina (Illumina, Inc., CA) had shorten the time for generating good quality reads whereby 2Gb of paired end reads (2×150bp) can be produced within 27 hours [11]. To date, there is no report on shotgun metagenomic analysis by using MiSeq as sequencing platform, thus, we applied this new sequencing technology platform on extensive marine metagenomic analysis.

Result and Discussion

MiSeq platform has been commonly used as a tool for bacteria whole genome sequencing. Its function has expanded when users start doing amplicons-based diversity studies using this platform. To our knowledge, none of the shotgun-based metagenomic studies was performed on this platform. Therefore, we decided to challenge the ability of the sequencer in sequencing a complex diversity. This is the first report for shotgun based metagenomic by using MiSeq.

A total of 6,701,060 raw reads was generated from this sequencing platform. After trimming of these reads at the limit of 0.05, the

number of processed reads is 5,170,106, which were used for downstream analyses. The processed reads were searched against NCBI NR environmental database using the built-in BLASTN algorithm of the Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA) portal [26]. BLAST outputs were imported into MEGAN4 software (version 4.69.4) (2) for taxonomical and functional classifications. This software will automatically calculate the taxonomic classification of the reads by mapping reads onto different taxa in the NCBI taxonomy.

Sea water harbor vast number and varieties of living microorganism and the concentration increase when it is closer to the coastal [12]. Among them, bacteria, Archaea, unicellular fungi and Protista made up the largest oceanic biomass. These microscopic organism are contributing to most of the primary production and biogeochemical cycle in the ocean which is crucial for the marine life [1,13]. Based on the taxonomic classification by MEGAN, we observed 24 phyla of Eubacteria (**Figure 1**) and 5 phyla of Archaea **(Figure 2)** together with a group of unclassified bacteria and Archaea. The 6 dominant phyla of Eubacteria on the sea surface of Georgetown are the *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, *Verrucomicrobia*, *Firmicutes* and *Cyanobacteria*, whereas the 5 phyla for Archaea are the *Eurkarchaeota*, *Thaumarchaeota*, *Crenarchaeota, Karoarchaeota* and *Nanoarchaeota*. .

Proteobacteria and *Bacteroidetes* are large and widespread group of bacteria that are commonly found in sea water [14] They has been reported to be the large division of bacteria found in marine in either the biodiversity studies of the sea ice [15], anoxic zone at Cariaco Basin [14], at the coastal, open ocean and estuary habitat in the studies of Sorcerer II Global Ocean Sampling project[16]. In our study, we observed the largest division of bacteria in our marine sample also *Proteobacteria* again supporting the finding in the above previous studies of sea water. A combination of α, β, δ, γ, ε, and the rare species of ζ-*Proteobacteria* made up the dominant group in this study. Second largest phylum of the bacteria presence on this sample is *Bacteroidetes*. Sometime the *Bacteroidetes* was group with its closely related phyla, *Chlorobi* forming a superphylum (Figure 1).

The abundance of sunlight on the sea surface of tropical country could be one of the reasons that made the autotrophic *Cyanobacteria* as one of the dominant species in this study. Scientist has made an estimation that the upper 200m of ocean contains an average cellular density of 4×10^4 cells/ml of autotrophic microbes such as *Cyanobacteria* [1]. Based on the reported findings, *Cyanobacteria* is an important organism which replenishes a significant number of oxygen in the Earth via biophotolysis by using the sunlight as sole energy source [17,18]. Its carbon fixation impacts the global carbon cycle [19].

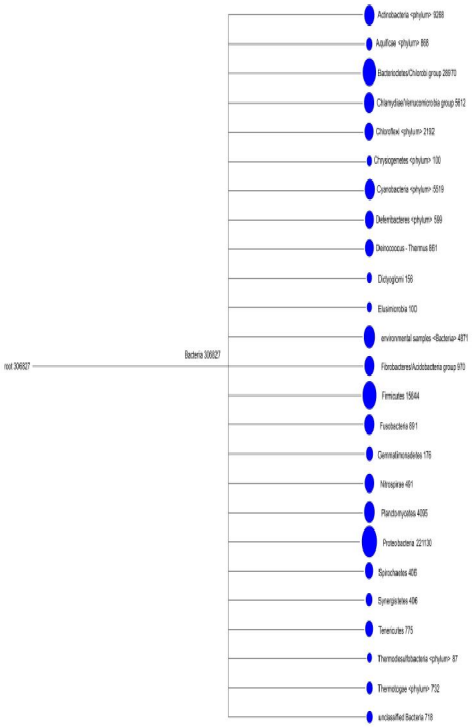

Figure 1: The diversity of Bacterial at phylum level

The third domain of life, *i.e.* Archaea was generally being perceive as a group of microbe that only leaves in extreme environment [20,21]. In addition to its common habitat such as hot spring, hydrothermal vent and salt lake, it can be isolated from sea and they play an important role in providing oceanic nitrogen source [14,21,22]. In this study, *Euryarchaeota* is the most abundant phyla followed by *Thaumarhaeota*, *Crenarchaeota*, and *Nanoarchaeota*. This finding show the distribution of the phylum in Archaea are defined where Euryarchaeota and *Thaumarhaeota* are known extreme halophiles survive in extreme salty environment such as seawater in this work. However, more studies are needed in order to further understand the role of archaea at sea surface.

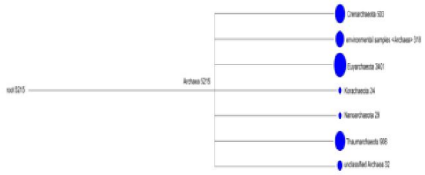

Figure 2: The diversity of Archaea at phylum level

Figure 3 illustrates the abundance of bacteria at phylum level. From the total number of 306,827 reads that mapped to bacteria, 221,130 (72%) reads are associated with *Proteobacteria* which represents the highest percentage among the bacterial phylum. The second most abundant phylum is *Bacteroidetes* (9.4%), followed by *Actinobacteria* (3%), *Verrucomicrobia* and *Cynanobacteria* (1.8%) and *Firmicutes* (5.1%). Numbers in the pie chart represents percentage.

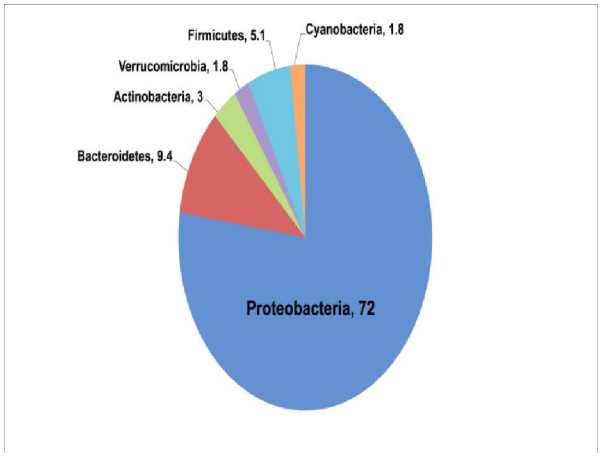

Figure 3: Distribution of bacterial percentage of at phylum level

Figure 4 illustrated the percentage of Archaea at phylum level present in our marine metagenome sample. Out of the 5 phyla, the most abundant phylum is *Euryarchaeota* (65.2%) with 3,401 from total reads of 5,215. *Thaumarhaeota* is the second most abundant phylum (17.4%; 857 reads), followed by *Crenarchaeota* (9.6%; 554 reads), *Karoacheato* (0.5%; 24 reads), and *Nanoarchaeota* (0.6%; 29 reads). This finding show the distributions of the phylum in Archaea are defined where Euryarchaeota and *Thaumarhaeota* are known extreme halophiles survive in extreme salty environment such as seawater in this work. Numbers in the pie chart represents percentage.

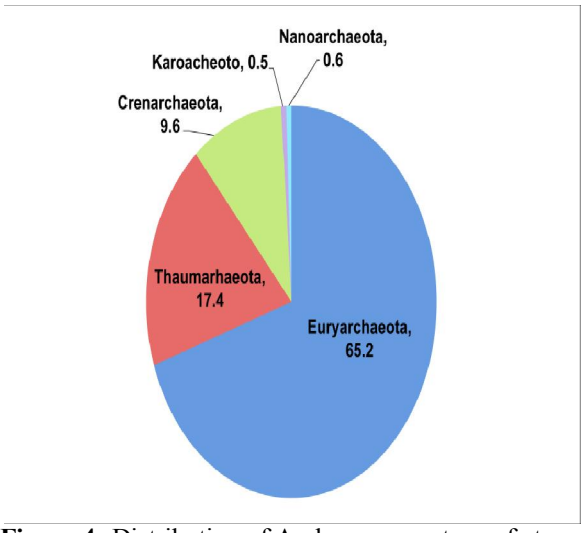

Figure 4: Distribution of Archaea percentage of at phylum level

Methods and Materials

Sample Collection

Sea water was collected from the sea surface of South Channel of Straits of Malacca, Georgetown coast, (N 05 °25.587, E 100° 19.591) Penang, Malaysia in 2011 during low tide and the water temperature was 30°C. The sample was transported back to our laboratory within 48 hours of collection in a sterile ample bottle.

DNA Extraction

Sea water (1L) was filtered through a filter membrane (pore size diameter of 0.22 µm) (Sartorius, Germany). The membrane was removed and submerged into 10 ml of modified Cetyltrimethylammonium bromide (CTAB) lysis buffer (100 mM Tris-HCl, 100mM EDTA, 100 mM K_2HPO_4, 1.5 M NaCI and 1 % CTAB) [23,24]. Subsequently, mechanical and enzymatically lysis was carried out by beating the membrane with sterile glass beads for 5 min followed by incubation at 37 °C for 4 hours in the presence of 100 µg/ml of lysozyme (Sigma, USA) with gentle swirled at 30 min interval. The solution was then boiled at 90°C for 1 hour [24].

Extraction of the DNA was started with the addition of SDS to the final concentration of 1% (w/v), followed by the supplementation of 100 µg/ml of proteinase K. The mixture was then incubated at 56°C for 2 hours with gentle shaking at 15 min interval. Successively, RNA treatment was performed

with addition of 2 µl of RNase A (Qiagen, USA) at ambient temperature for 30 min. Most of the proteins were removed by washing the DNA pellet twice with phenol/chloroform/isoamyl alcohol (25:24:1). The resulting DNA pellet was precipitated with 0.6 volume of isopropanol followed by 70% (v/v) ethanol and was rehydrated with 100 µl of TE buffer [25]

DNA Sequencing

Prior to sequencing, the purified DNA was quantified with Qubit™ (Invitrogen), and qualified with Nanodrop (Thermo Fisher Scientific) and electrophoresed in 1% (w/v) agarose gel. The high quality DNA was selected for library preparative with TruSeq DNA Sample Preparation Kit according to manufacturer's instruction, followed by shotgun sequencing (2 × 151 bp paired end run) on MiSeq platform (Illumina).

Microbial Diversity Metagenomic Analysis

Raw sequencing reads generated were trimmed using CLC Genomics Workbench (version 4.0). The trimmed sequences were BLASTed against NCBI non-redundant (nr) environmental database which is online tool using the built in BLASTN algorithm of the Cyber infrastructure of Advanced Microbial Ecology Research and Analysis (CAMERA) portal [26]. The BLAST output was channeled into MEGAN4 version 4.69.4 for microbial taxonomical and functional analysis. This software calculates the taxonomic classification by mapping reads onto different taxa with the reference from NCBI taxonomy.

The sequences of this metagenomic project can be obtained through SRA resipotary (ID=SRA 057075) on the National Center for Biotechnology Information (NCBI) website.

Acknowledgement

This project was funded by HIR grants (HIR/MOHE A-000001-50001; UM.C/HIR/MOHE/08) from the University of Malaya and Ministry of Higher Education, Malaysia. We would also like to thank Hazhir Hajian from Genome Informatics Research Group (GIRG), University of Malaya for his excellent IT support in this project.

Corresponding Author

Dr.Choo Siew Woh

Dental Research and Training Unit Faculty of Dentistry University of Malaya50603 Malaysia.

Email: lchoo@um.edu.my

References

1. Whitman WB, Coleman DC, Wiebe WJ (1998) Prokaryotes: the unseen majority. Proceedings of the National Academy of Sciences 95: 6578-6583.
2. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". Proceedings of the National Academy of Sciences 103: 12115-12120.
3. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. science 304: 66-74.
4. Sigman D, Hain M (2012) The Biological Productivity of the Ocean. Nature Education Knowledge 3: 4.
5. Naik MM, Pandey A, Dubey SK (2012) Bioremediation of Metals Mediated by Marine Bacteria. Microorganisms in Environmental Management: 665-682.
6. Gutierrez T (2011) Identifying polycyclic aromatic hydrocarbon-degrading bacteria in oil-contaminated surface waters at Deepwater Horizon by cultivation, stable isotope probing and pyrosequencing. Reviews in Environmental Science and Biotechnology: 1-5.
7. Shah N, Tang H, Doak TG, Ye Y. Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics; 2011. pp. 165-176.
8. Ismail S, Mohd-Ali N (2011) The Imaging of Heritage Conservation in Historic City of George Town for City Marketing. Procedia Engineering 20: 339-345.
9. Chai LT (2011) Culture Heritage Tourism Engineering at Penang: Complete The Puzzle Of "The Pearl Of Orient". Systems Engineering Procedia 1: 358-364.
10. Hugenholtz P, Tyson GW (2008) Metagenomics. Nature 455: 481-483.
11. Solovyev VV, Tatarinova TV (2011) Towards the integration of genomics, epidemiological and clinical data. Genome Medicine 3: 1-3.
12. Azam F, Fenchel T, Field J, Gray J, Meyer-Reil L, et al. (1983) The ecological role of water-column microbes in the sea. Marine ecology progress series Oldendorf 10: 257-263.
13. Atlas RM, Bartha R (1993) Microbial ecology: fundamentals and applications.
14. Madrid VM, Taylor GT, Scranton MI, Chistoserdov AY (2001) Phylogenetic diversity of bacterial and archaeal communities in the anoxic zone of the Cariaco Basin. Applied and environmental microbiology 67: 1663-1674.

15. Staley JT, Gosink JJ (1999) Poles apart: biodiversity and biogeography of sea ice bacteria. Annual reviews in Microbiology 53: 189-215.
16. Biers EJ, Sun S, Howard EC (2009) Prokaryotic genomes and diversity in surface ocean waters: interrogating the global ocean sampling metagenome. Applied and environmental microbiology 75: 2221-2229.
17. Yu J, Takahashi P (2007) Biophotolysis-based hydrogen production by cyanobacteria and green microalgae. Communicating current research and educational topics and trends in applied microbiology 1: 79-89.
18. Schütz K, Happe T, Troshina O, Lindblad P, Leitao E, et al. (2004) Cyanobacterial H2 production—a comparative analysis. Planta 218: 350-359.
19. Savage DF, Afonso B, Chen AH, Silver PA (2010) Spatially ordered dynamics of the bacterial carbon fixation machinery. science 327: 1258-1261.
20. Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences 87: 4576.
21. DeLong EF (1992) Archaea in coastal marine environments. Proceedings of the National Academy of Sciences 89: 5685.
22. Santoro AE, Buchwald C, McIlvin MR, Casciotti KL (2011) Isotopic signature of N2O produced by marine ammonia-oxidizing archaea. science 333: 1282-1285.
23. Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. Applied and environmental microbiology 62: 316-322.
24. M.G.Murray., W.F.Thompson. (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research 8: 4321-4325.
25. Manjula. A, Sathyavathi. S, Gunasekaran. P, Rajendhran. J (2011) Comparison of seven methods of DNA extraction from termitarium for functional metegenomic DNA library construction. Journal of Scientific and Industrial Research 70: 945-951.
26. Community cyberinfrastrucutre for Advance Micrial Ecology Research and Analysis: the CAMERA resource (2011) Shulei Sun; Jing Chen; Weizhong Li;Ilkay Altinatas; Abel Lin; Steve Peltier; Karen Stocks; Eric E. Allen; Mark Ellisman; Jeffrey Grethe; John Wooley Nucleic Acids Research 2010; doi: 10.1093/nar/gkq1102.

7/17/2013