Studii şi Cercetări Ştiinţifice 2018, **19** (1), pp. 091 – 096
Chimie şi Inginerie Chimică, Biotehnologii, Industrie Alimentară

Scientific Study & Research
Chemistry & Chemical Engineering, Biotechnology, Food Industry *ISSN 1582-540X*

SHORT COMMUNICATION

PYROSEQUENCING AND FUNCTIONAL ANNOTATION ANALYSIS OF MICROBIAL COMMUNITY IN FRESH WATER LAKE IN SAUDI ARABIA

Samy Selim[1,2*], Nashwa Hagagy[2]

1 Jouf University, College of Applied Medical Sciences, Department of Clinical Laboratory Sciences, Sakaka, P.O. 2014, Saudi Arabia
2 Suez Canal University, Faculty of Science, Botany Department, Ismailia, P.O. 41522, Egypt

*Corresponding author: sadomm2003@yahoo.com

Received: October, 04, 2017
Accepted: January, 31, 2018

Abstract: Metagenomics is providing conspicuous advantage to explore world of unculturable microorganisms in the natural samples to enhance our information about bacterial diversity. Here, we have performed metagenomic analysis of fresh water lake with focus on bacterial community using 454 pyrosequencing techniques. Roche GS FLX software indicated total of 156.253 reads; 15.226 contigs having > 100 bp sequence length whereas 10.481 contigs with > 500 bp sequence length. We have analyzed the bacterial community composition using BLASTN/BLASTX against NT/NR databases with E-value cutoff of 10^{-5}. We have observed a wide array of bacteria from phylum *proteobacteria* and family *Enterobacteriaceae* as well as very few viruses from *Podoviridae, Siphoviridae* and unclassified phages. Moreover, a functional analysis of all 5974 contigs using Rapid Annotation by Subsystem Technology (RAST) was performed and detected 15.319 coding sequences and 197 RNAs in 619 subsystems. The great diversity of microflora present in the lake may reflect the human activity in the area.

Keywords: *Freshwater Lake, functional genes, Metagenomics, microbial diversity, 454 pyrosequencing*

INTRODUCTION

Dumat AL-Jandal Lake is large, shimmering lake covers an area of about 500.000 square meters and has a depth of about 15-meters at its center. The water is uncontaminated enough to support plant and animal life; there is no trace of the usual smells associated with waste, and it is clean enough to attract families to picnic on its shores at weekends. The water floods down into the water and, after finding its way through filtering sand and rock underground, supplies the wells that irrigate many farms of Aljouf area. To our knowledge, no studies have been conducted to document microbial community in this geographically distinctive region with unique environmental conditions. Since the bacterial populations of the freshwater lakes play a vital role in many hydrological and biogeochemical cycles, thereby providing critical ecosystem facilities to humankind [1]. Hence, the objective of this study was to assess the microbial diversity and enzymatic characteristics of water samples collected from Lake Dumat Al-Jundal, Saudi Arabia, with main focus on bacterial community; using high throughput next generation sequencing.

MATERIALS AND METHODS

Five water samples were collected in sterile containers from Lake Dumat Al-Jundal in Aljouf region, Saudi Arabia (29°48'48.5"N 39°54'37.8"E) in June 2016. All samples were kept at 4^{o}C until reaching the laboratory for DNA extraction. The metacommunity DNA from selected water sources was trapped by filtering 5 liters of water (of each source) through 0.22 µm microfiber filter paper. The DNA from the whole membrane was extracted using the Power Soil kit (Mo Bio Laboratories, Carlsbad, USA) according the manufacturer's instructions with e final DNA elution 2 times 30 µL of TE buffer. The bacterial community was investigated through amplification of 16S rRNA gene using the primers 27f / 685r [2]. The PCR reaction of 50 µl was composed of 3 µl extracted DNA, 50 pmol of each primer, 200 µmol l−1 dNTP, 1.5 U HotStar Taq plus DNA polymerase (Qiagen, Hilden, Germany), 1× PCR buffer, and 2.1 mM Mg2+. Amplified DNA was sequenced using 454/Roche sequencing technique according to the manufacturer's instructions. Sequenced reads were then assembled and analyzed using several tools. To assemble the raw reads we have used Roche GS FLX Software (v 2.8) and all the contigs were further annotated using BLASTX/BLASTN. We performed functional analysis of all 5974 contigs using Rapid Annotation using Subsystems Technology (RAST) [3].

RESULTS AND DISCUSSION

Lakes are representing a critical natural resource for human societies [4]. Despite the recognition that bacteria occupy a prominent role in Lake Ecosystem processes and significantly impact lake water quality, the bacterial taxa contributing in these activities remain largely undescribed. Molecular biology tools now provide us with unique access to the diversity and composition of freshwater lake bacterial communities and have for

the first time allowed to recognize the numerically dominant organisms in these ecosystems and learn much about their distributions in time and space [5].

In this study, water samples were subjected to whole-community genome shotgun 454 pyrosequencing. Roche GS FLX Software gave us total 156,253 reads. There were 15,226 contigs having >100 bp sequence length whereas 10.481 contigs with > 500 bp sequence length. For further analysis we have taken contigs with >500 bp only. Further, we have analyzed the microbial community composition using BLASTN/BLASTX against NT/NR databases with E-value cutoff of 10^{-5} > =70 % of total contigs were mapped to the reference with > =60 % contig match coverage. Out of total 10,481 contigs, we have found hit for 5974 (57 %) contigs with NT/NR database. In which, we have got 5967 hits for bacteria, 1 for eukaryote, 5 for viruses and 1 for other sequence. More specifically, *proteobacteria* was the dominant phyla (5964) and *Enterobacteriaceae* was dominant family.

As shown in Figure 1, the results of meta-community analysis revealed that domain bacteria is predominantly present (99.8 %) in the water sample, followed by Eukaryota (0.02 %), viruses (0.08 %) and other sequences (0.02 %). Most abundant phyla was *Proteobacteria* (99.8 %) and the most dominant family was *Enterobacteriaceae* (89 %) followed by *Xanthomonadaceae* (10 %), *Vibrionaceae* (0.4 %), *Pasteurellaceae* (0.1 %), *Shewanellaceae* (0.07 %). According to a custom curated freshwater database by a previous study [6], 21 phyla have been recovered from freshwater lakes, with 5 phyla being recovered commonly (*Proteobacteria*, especially *Betaproteobacteria, Actinobacteria, Bacteroidetes, Cyanobacteria* and *Verrucomicrobia*). These five phyla were also the most numerous in the database compiled previously by Zwart and colleagues [7] and are in general agreement with the majority of studies, which reported the *Actinobacteria* [8 – 11] and the *Betaproteobacteria* [12 – 15] as being the most abundant bacterial phylum in Lake Epilimnia. The remaining 16 recovered phyla make up only ≈ 2.6 % of the total sequences collected in previous database[7] and include *Acidobacteria*, BRC1, *Chlorobi, Chloroflexi, Fibrobacteres, Firmicutes, Fusobacteria, Gemmatimonadetes, Lentisphaerae, Nitrospira,* OD1, OP10, *Planctomycetes, Spirochaetes,* SR1, and TM7. The results reported here may be nearby to the statement of "each lake has a different bacterial community" which stated by Newton *et al.* [6], however, we reported that *Proteobacteria* as the most abundant phyla (99.8 %).

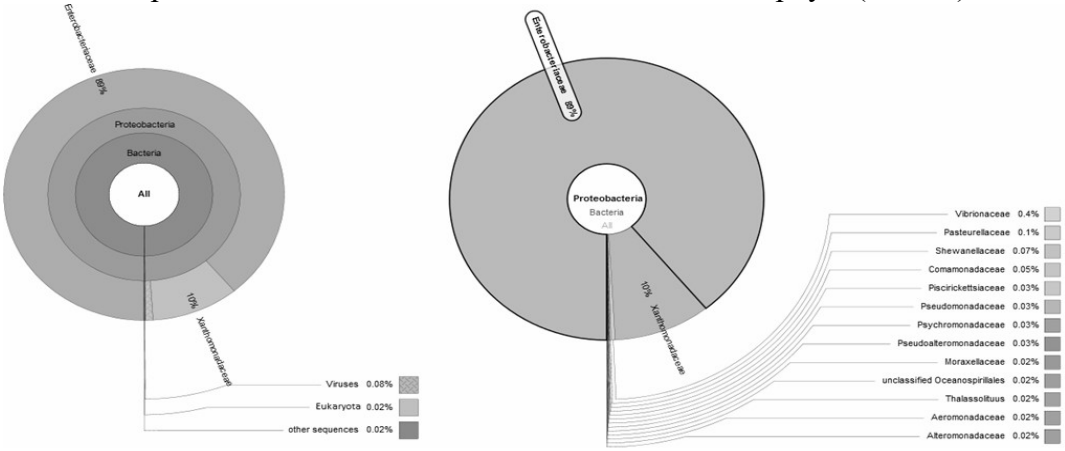

Figure 1. *Microbial community structure in Water Lake metagenome*

The functional analysis of all 5974 contigs using Rapid Annotation using Subsystems Technology (RAST) detected 15.319 coding sequences and 197 RNAs in 619 subsystems. Among the classified CDS from RAST showed major CDS hits for enzymes involved in the subsystems amino acids and derivatives and the carbohydrate metabolism. We could observe CDS for enzymes utilizing various monosaccharide and oligosaccharide like xylose.

Further, the subsystem information displays several antibiotic resistance genes encoding methicillin, efflux pumps and quinolone. There also exist multiple genes encoding hydrolytic enzymes. Overall, functional information demonstrates with potential for resistance to a wide range of antimicrobial and heavy metals. All the predicted coding sequence were further analyzed with KEGG pathway database and we found metabolic pathways were highly enriched (Figure 2).

This is dependable with a previous finding that the freshwater microbial genomes seem to harbor a higher proportion of certain putative genes involved in transport of sugars such as xylose and various polysaccharides [16, 17]. Further we constructed highly enriched metabolic pathways network using KEGG mapper where all highlighted in red represents metabolic pathways and pathway highlighted in bold red represents carbohydrate metabolism. Therefore, this metagenomic study provides a comprehensive description and analysis of microbial diversity and functional potential present in the metagenome of Dumat AL-Jandal Lake, KSA.

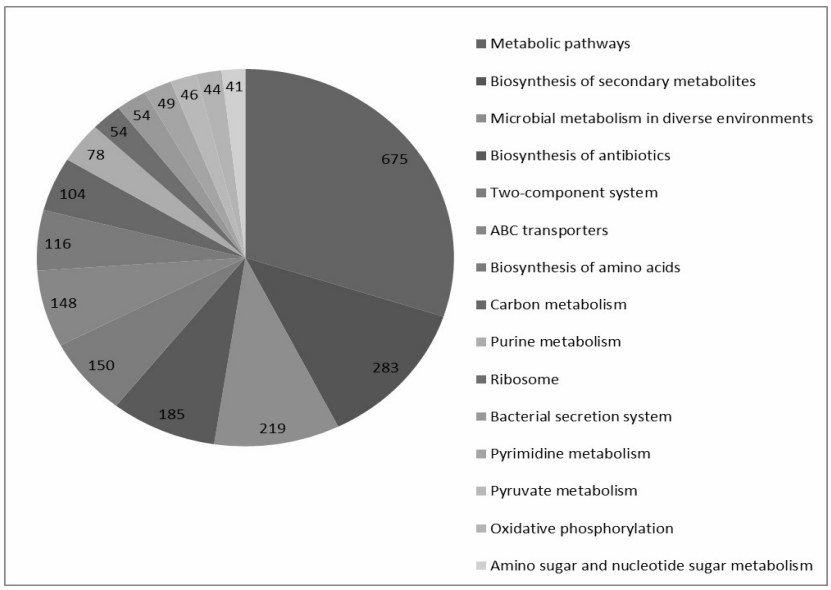

Figure 2. *Pathway enrichment analysis*

CONCLUSION

Freshwater is one of the most important resources for human and it endures very diverse ecosystems, nevertheless our knowledge of freshwater microbial ecology is still poor compared with marine ecosystems. The results suggested that metagenomic analysis of the microbial communities have the potential to characterize such ecosystem. Also, information about microbial interaction within the community can be revealed by metagenomic analysis. Therefore, this report was augment understanding of the

significance of freshwater microbial communities for ecosystem and also human health. However, further analysis with more study sites is needed to form a complete picture about the community and their reflection for human activity in this distinctive ecosystem.

ACKNOWLEDGEMENT

This work was funded by the Deanship of Scientific Research (DSR), Aljouf University, Aljouf, KSA, under grant no. (36/155, 2016).The authors, therefore, acknowledge with thanks DSR technical and financial support. We also would like to acknowledge ArrayGen Technologies, Pune, India for contributing to data analysis and bioinformatics support.

REFERENCES

1. Millennium Ecosystems Assessment, *Ecosystems and human well-being: current state and trends assessment*, Island Press, Washington, DC, **2005**;
2. Lane, D.J.: 16S/23S rRNA sequencing in: *Nucleic acid techniques in bacterial systematics* (Editors: Stackebrandt, E., Goodfellow, M.), Wiley and Sons, New York, **1991**,115-148;
3. 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.: The metagenomic RAST server-a public resource for the automatic phylogenetic and functional analysis of metagenomes, *BMC Bioinformatics*, **2008**, **9**, 386;
4. Downing, J.A., Prairie, Y.T., Cole, J.J., Duarte, C.M., Tranvik, L.J., Striegl, R.G.: The global abundance and size distribution of lakes, ponds, and impoundments, *Limnol Oceanogr*, **2006**, **51**, 2388-2397;
5. Eiler, A., Zaremba-Niedzwiedzka, K., Martínez-García, M., McMahon, K.D., Stepanauskas, R., Andersson, S.G.E., Bertilsson, S.: Productivity and salinity structuring of the microplankton revealed by comparative freshwater metagenomics, *Environmental Microbiology*, **2014**, **16** (9), 2682-2698;
6. Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D., Bertilsson, S.: A Guide to the natural history of freshwater lake bacteria, *Microbiology and molecular biology reviews*, **2011**, **75** (1), 14-49;
7. Zwart, G., Crump, C., Agterveld, M.P.K.V., Hagen, F., Han, S.K.: Typical fresh water bacteria: ananalysis of available 16S rRNA gene sequences from plankton of lakes and rivers, *Aquatic Microbial Ecology*, **2002**, **28**, 141-155;
8. Glo¨ckner, F.O.: Comparative 16S rRNA analysis of Lake Bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of actinobacteria, *Appl Environ Microbiol*, **2000**, **66**, 5053–5065;
9. Warnecke, F., Sommaruga, R., Sekar, R., Hofer, J.S., Pernthaler, J.: Abundances, identity, and growth state of actinobacteria in mountain lakes of different UV transparency, *Appl Environ Microbiol*, **2005**, **71**, 5551-5559;
10. Allgaier, M., Grossart, H.P.: Seasonal dynamics and phylogenetic diversity of free-living and particle-associated bacterial communities in four lakes in northeastern Germany, *Aquatic Microbial Ecology*, **2006**, **45**, 115-128;
11. Buck, U., Grossart, H.P., Amann, R., Pernthaler, J.: Substrate incorporation patterns of bacterioplankton populations in stratified and mixed waters of a humic lake, *Appl Environ Microbiol*, **2009**, **11**, 1854-1865;
12. Glo¨ckner, F.O., Fuchs, B.M., Amann, R.: Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization, *Appl Environ Microbiol*, **1999**, **65**, 3721-3726;

13. Perez, M.T., Sommaruga, R.: Differential effect of algal- and soil-derived dissolved organic matter on alpine lake bacterial community composition and activity, *Limnol Oceanogr*, **2006**, **51**, 2527-2537;
14. Wu, Q.L., Zwart, M., Schauer, M.P., Kamst-van, A., Hahn, M.W.: Bacterioplankton community composition along a salinity gradient of sixteen high mountain lakes located on the Tibet Plateau, China, *Appl Environ Microbiol*, **2006**, **72**, 5478-5485;
15. Hervas, A., Casamayor, E.O.: High similarity between bacterioneuston and airborne bacterial community compositions in a high mountain lake area, *FEMS Microbiol Ecol*, **2009**, **67**, 219-228;
16. Garcia, S.L., McMahon, K.D., Martinez-Garcia, M., Srivastava, A., Sczyrba, A., Stepanauskas, R., Grossart, H.P., Woyke,T., Warnecke, F.: Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton, *ISME Journal*, **2012**, **7**, 137-147;
17. Oh, S., Caro-Quintero, A., Tsementzi, D., DeLeon-Rodriguez, N., Luo, C., Poretsky, R., Konstantinidis, T.: Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of lake lanier, a temperate freshwater ecosystem, *Applied and Environmental Microbiology*, **2011**, **77** (17), 6000-6011.