

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Introductory Chapter: Metagenomics and Metagenomic Approaches

Wael N. Hozzein

1. Introduction

Metagenomics can be defined as the techniques and procedures that are used for the culture-independent analysis of the total genomic content of microorganisms living in a certain environment [1]. It has many useful applications with very promising potential in both medical and environmental microbiology. The most common use of metagenomics in environmental microbiology is studying the diversity of microbial communities in particular environments through the analysis of rRNA genes and how these communities change in response to changes in physical and chemical properties of these environments [2].

Metagenomics also provides an opportunity to obtain and identify novel enzymes with industrial applications from extreme environments where unculturable extremophiles live. In such circumstances, functional metagenomics enables the isolation of genes coding for extremozymes, enzymes that are capable of being catalytically active in extreme conditions, or genes that will allow for better understanding of the mechanisms that make such organisms resistant to extreme environmental conditions [3].

Metagenomics has special importance when it comes to studying soil microbiology. It is estimated that the number of distinct microorganisms in 1 gram of soil exceeds the number of microbial species cultured so far [4]. Therefore, metagenomics seems to be the ideal culture-independent technique for unraveling the biodiversity of soil microorganisms and to study how this biodiversity is affected with continuously changing conditions.

2. Sequencing technologies and metagenomics

Recently, taxonomic profiling, characterization, and analysis of microbial communities are being mostly performed using different next-generation sequencing (NGS) platforms. Metagenomic samples are high-throughput, short-read sequences, and the cost is relatively decreasing. In addition, these platforms are advantageous, avoiding the need for cloning of DNA fragments [5].

Recent advances in NGS technologies were developed to suit various numbers of applications, cost, and capabilities [6]. The most commonly used platforms are the 454 Life Sciences (Roche) and Illumina systems (Solexa) [7]. The 454 sequencing technology, which was the first commercially available next-generation technology, is based on the pyrosequencing technique. It provides high throughput and relatively cheap analysis [8]. During the sequencing reaction in this technique,

nucleotide incorporation into the growing chain is detected by the capture of the released pyrophosphate, which is converted into a light through an enzymatic reaction. Different nucleotides are sequentially added into each nucleotide incorporation event; therefore the light signal can be attributed to a specific nucleotide. Finally, the light signals are converted into sequence information. In the 454 pyrosequencer, the DNA fragments are amplified after being fixed on beads in a water-oil emulsion [9]. Pyrosequencing has been employed widely in the analysis of microbial diversity in many environments including marine environments [10] and different soil environments [11, 12].

Illumina sequencing technology relies on the use of fluorescently labeled reversible terminator nucleotides. Instead of being chemically modified to prevent further DNA synthesis (dideoxynucleotides) which is the case with Sanger sequencing, the terminator nucleotides are attached with blocking group that can be removed from the nitrogen base in a single step. DNA synthesis takes place on a chip where primers are attached. After each cycle, the dyes attached to each nucleotide are excited by laser followed by scanning of the incorporated bases. In order for the next synthesis cycle to proceed, the blocking group and the dye are first removed by a chemical reaction. Illumina sequencing platform was successfully used to study microbial diversity in many environments [13–15].

In addition to the abovementioned technologies, recently developed sequencing technologies are available and being employed in metagenomic studies. These include SOLiD 5500 W Series developed by Applied Biosystems, single-molecule real-time (SMRT) DNA sequencing from Pacific Biosciences, and Ion Torrent semiconductor sequencing [7]. More innovative technologies are being developed that could be of great use for metagenomic studies in the near future. Strand sequencing technologies, currently being developed by Oxford Nanopore technologies, enable the sequencing of intact DNA strand that passes through a protein nanopore [16]. Irys Technology, developed by BioNano Genomics, represents one of the very promising new technologies in genomics era [7].

3. The metagenomic approaches

Metagenomics research strategy starts with selecting a proper ecological or biological environment of interest that hosts a wide variety of microbial communities which may have potential biotechnological applications. Environments that attract metagenomic researchers are mainly those characterized with extreme conditions or unique environmental conditions. These include environments with highly acidic or alkaline pH; high metal concentrations, pressures, or radiation; and high salinity or extreme temperatures [3].

Metagenomic analysis starts with isolating genomic DNA that represents the whole community in the soil sample, constructing a DNA library from the isolated DNA, and screening the available library for a target gene. It is important here to select the DNA extraction method that will provide enough yield and DNA that represents the diversity of the whole microbial community in the target environment. This is still one of the most challenging steps of metagenomic analysis. The chemical and physical characteristics of soils are very wide and complex, depending on the type of the soil examined, that will make it difficult to develop a reference method for DNA extraction from soils. Besides, soils contain many substances that are co-extracted with genomic DNA and harbor inhibitory effects on the downstream processing of the extracted DNA. Examples include humic and fulvic acids [17]. Therefore, optimization and comparison between different extraction methods are usually required for each type of soil [18–21].

A DNA library is then constructed from the genomic DNA isolated from the target environment. This is performed by fragmenting the isolated DNA into fragments with appropriate sizes that would allow for their cloning. This is performed by either using restriction enzyme digestion or mechanical shearing. DNA fragments obtained from such processes are cloned into the proper cloning vector. Plasmid vectors are used for small DNA fragments, and the libraries generated are called small-insert genomic libraries. Large inserts are cloning into cosmid or fosmid vectors which can hold inserts up to 40 kb in size or BAC vector which can carry inserts with sizes that exceed 40 Kb [22].

DNA libraries are usually constructed in a microorganism that is well-studied and is easy to manipulate inside the laboratory such as *Escherichia coli*. In case there is a need for expressing the genes within the DNA inserts in other microorganisms, shuttle vectors are used to transfer the libraries into a proper host [23].

Finally, a screening assay is applied to search for a gene of a particular function, and the gene product is functionally analyzed. There are two different metagenomic strategies that are commonly used in research. The first one is focused on the use of marker genes such as the ribosomal genes 16S rRNA [24] and 18S rRNA [25] to study the composition of the microbial community in a certain environment or specific protein-coding gene with medical or industrial importance [26–28]. Such a strategy is called targeted metagenomics. The second approach is the shotgun metagenomics, in which a wide coverage of genomic DNA sequences is achieved using high-throughput next-generation sequencing to assess the entire taxonomic structure or functional potential of microbial communities [29].

The most challenging aspect of the screening process in metagenomics is the analysis of a huge amount of sequence data that are generated from the constructed library. A wide range of bioinformatic tools has been developed over the years to help analyze the metagenomic data and compare it to available online databases.

Acknowledgements

This work was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University, through the Research Groups Program Grant no. (RGP-1438-0006).

Author details

Wael N. Hozzein
Botany and Microbiology Department, Faculty of Science, Beni-Suef University,
Beni-Suef, Egypt

*Address all correspondence to: hozzein29@yahoo.com

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Daniel R. The metagenomics of soil. *Nature Reviews Microbiology*. 2005; 3(6):470
- [2] Delmont TO, Robe P, Cecillon S, Clark IM, Constancias F, Simonet P, et al. Accessing the soil metagenome for studies of microbial diversity. *Applied and Environmental Microbiology*. 2011; 77(4):1315-1324
- [3] Mirete S, Morgante V, González-Pastor JE. Functional metagenomics of extreme environments. *Current Opinion in Biotechnology*. 2016;38: 143-149
- [4] Torsvik V, Øvreås L. Microbial diversity and function in soil: From genes to ecosystems. *Current Opinion in Microbiology*. 2002;5(3):240-245
- [5] Ansorge WJ. Next-generation DNA sequencing techniques. *New Biotechnology*. 2009;25(4):195-203
- [6] Scholz MB, Lo CC, Chain PS. Next generation sequencing and bioinformatic bottlenecks: The current state of metagenomic data analysis. *Current Opinion in Biotechnology*. 2012;23(1):9-15
- [7] Oulas A, Pavloundi C, Polymenakou P, Pavlopoulos GA, Papanikolaou N, Kotoulas G, et al. Metagenomics: Tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinformatics and Biology Insights*. 2015;9:BBI-S12462
- [8] Ronaghi M. Pyrosequencing sheds light on DNA sequencing. *Genome Research*. 2001;11(1):3-11
- [9] Goodwin S, McPherson JD, McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*. 2016;17(6):333
- [10] Egge E, Bittner L, Andersen T, Audic S, de Vargas C, Edvardsen B. 454 pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: A test for marine haptophytes. *PLoS One*. 2013;8(9):e74371
- [11] Wang X, Hu M, Xia Y, Wen X, Ding K. Pyrosequencing analysis of bacterial diversity in 14 wastewater treatment systems in China. *Applied and Environmental Microbiology*. 2012; 78(19):7042-7047
- [12] Alex A, Antunes A. Pyrosequencing characterization of the microbiota from Atlantic intertidal marine sponges reveals high microbial diversity and the lack of co-occurrence patterns. *PLoS One*. 2015;10(5):e0127455
- [13] Fan W, Huo G, Li X, Yang L, Duan C, Wang T, et al. Diversity of the intestinal microbiota in different patterns of feeding infants by Illumina high-throughput sequencing. *World Journal of Microbiology and Biotechnology*. 2013;29(12):2365-2372
- [14] Lentini V, Gugliandolo C, Bunk B, Overmann J, Maugeri TL. Diversity of prokaryotic community at a shallow marine hydrothermal site elucidated by Illumina sequencing technology. *Current Microbiology*. 2014;69(4): 457-466
- [15] Hong C, Si Y, Xing Y, Li Y. IlluminaMiSeq sequencing investigation on the contrasting soil bacterial community structures in different iron mining areas. *Environmental Science and Pollution Research*. 2015;22(14): 10788-10799
- [16] Jain M, Olsen HE, Paten B, Akeson M. The Oxford NanoporeMinION: Delivery of nanopore sequencing to the genomics community. *Genome Biology*. 2016;17(1):239

- [17] Young JM, Rawlence NJ, Weyrich LS, Cooper A. Limitations and recommendations for successful DNA extraction from forensic soil samples: A review. *Science & Justice*. 2014;**54**(3): 238-244
- [18] Finley SJ, Lorenzo N, Mulle J, Robertson BK, Javan GT. Assessment of microbial DNA extraction methods of cadaver soil samples for criminal investigations. *Australian Journal of Forensic Sciences*. 2016;**48**(3):265-272
- [19] Lim NY, Roco CA, Frostegård Å. Transparent DNA/RNA co-extraction workflow protocol suitable for inhibitor-rich environmental samples that focuses on complete DNA removal for transcriptomic analyses. *Frontiers in Microbiology*. 2016;**7**:1588
- [20] Gupta P, Manjula A, Rajendhran J, Gunasekaran P, Vakhlu J. Comparison of metagenomic DNA extraction methods for soil sediments of high elevation Puga hot spring in Ladakh, India to explore bacterial diversity. *Geomicrobiology Journal*. 2017;**34**(4): 289-299
- [21] Mazziotti M, Henry S, Laval-Gilly P, Bonnefoy A, Falla J. Comparison of two bacterial DNA extraction methods from non-polluted and polluted soils. *Folia Microbiologica*. 2018;**63**(1):85-92
- [22] Simon C, Daniel R. Construction of small-insert and large-insert metagenomic libraries. In: *Metagenomics*. New York, NY: Humana Press; 2017. pp. 1-12
- [23] Lam KN, Cheng J, Engel K, Neufeld JD, Charles TC. Current and future resources for functional metagenomics. *Frontiers in Microbiology*. 2015;**6**:1196
- [24] Païssé S, Valle C, Servant F, Courtney M, Burcelin R, Amar J, et al. Comprehensive description of blood microbiome from healthy donors assessed by 16 S targeted metagenomic sequencing. *Transfusion*. 2016;**56**(5): 1138-1147
- [25] West D. Use of an 18s rRNA metagenomics approach as a method of detection of multiple infections in field blood samples collected on FTA cards from cattle [Doctoral dissertation]. University of Salford; 2018
- [26] Nurdiani D, Ito M, Maruyama T, Terahara T, Mori T, Ugawa S, et al. Analysis of bacterial xylose isomerase gene diversity using gene-targeted metagenomics. *Journal of Bioscience and Bioengineering*. 2015;**120**(2): 174-180
- [27] Ufarte L, Laville E, Duquesne S, Potocki-Veronese G. Metagenomics for the discovery of pollutant degrading enzymes. *Biotechnology Advances*. 2015;**33**(8):1845-1854
- [28] Lanza VF, Baquero F, Martínez JL, Ramos-Ruiz R, Gonzalez-Zorn B, Andremon A, et al. In-depth resistome analysis by targeted metagenomics. *Microbiome*. 2018;**6**(1):11
- [29] Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*. 2017; **35**(9):833