METAGENOMICS AND 'OMICS' TECHNOLOGIES FOR ENVIRONMENTAL BIOREMEDIATION: A REVIEW

Vivehananthan K, Lilani EDC, Abeygunawardena I

Department of Basic Sciences, Faculty of Health Sciences

The Open University of Sri Lanka, Nawala, Nugegoda

Metagenomics is the study of a community's collective genome. It is a culture-independent study which combines various molecular tools established over the previous century, allowing researchers to better investigate the variety of microorganisms, their interdependence and unleash the possibilities of biotechnology. This review is focused on how metagenomic study and some of its "omics" approaches help to identify environmental microorganisms and their applications in the bioremediation of environmental pollutants in different studies. Metagenomics is a rapidly expanding and diversified discipline of environmental biology that aims to learn more about the genomes of environmental microbes and entire microbial communities. In metagenomic studies, DNA is directly extracted from the community, cloned into a host bacterium, created into a library, and then sequenced or screened for expression of activities of interest. Mainly, two methods are used to obtain the genetic data after the creation of a metagenomic library. In functional metagenomics, the host bacteria express the recombinant DNA in either growth-suppressive or growth-promoting ways. In sequence-based metagenomics, complementary oligonucleotides (oligos) are used to seek out a specific gene, or cloned DNA is randomly sequenced using vector-based primers. Innovative "omics" technologies, like transcriptomics, metaproteomic, metabolomics and in-silico research, have made it possible to expand the scope of metagenomics studies. Environmental microbes play an important role in the degradation and detoxification of various organic and inorganic pollutants and the biogeochemical cycling of minerals in the ecosystem. Therefore, it's necessary to understand the mechanism of bioremediation of a specific pollutant and identify the key enzyme of catabolic gene of microorganisms involved. The findings of several "omic" approaches can be combined and utilized to examine the metabolic activities of the bacteria involved in bioremediation processes. Such techniques have made it possible to examine metagenomic contaminated samples comprising diverse bacteria in a realistic and cost-effective manner. Characterization of environmental microbes will provide a fresh window into the search for undiscovered bacteria with unique catabolic genes and enzymes for the breakdown and detoxification of toxins in contaminated environments. In conclusion all environmental microbes are genetically described via metagenomics together with "omics" technologies to apply in different studies.

Keywords: Metagenomics, Omic approaches, Bioremediation

**Corresponding Author: chathurikalilani4@gmail.com*

METAGENOMICS AND 'OMICS' TECHNOLOGIES FOR ENVIRONMENTAL BIOREMEDIATION: A REVIEW

Vivehananthan K, Lilani EDC, Abeygunawardena I

Department of Basic Sciences, Faculty of Health Sciences

The Open University of Sri Lanka, Nawala, Nugegoda

INTRODUCTION

Microbes are invisible to the naked eye but are important inhabitants of the planet. They live in any environment, including water systems, soil, air and host organisms such as all animals, humans, plants and lower eukaryotes (Cui *et al*., 2016). Understanding their population dynamics is essential for understanding their genomic information. Isolating and sequencing the genome of a single organism may not be sufficient because no single isolate can reflect the entire genetic and metabolic capacity of its related members. Metagenomics is a non-culture-based approach used to examining large groups of genomes from a diverse population of bacteria (Neelakanta & Sultana., 2013). Metagenomics is a new field in genomics that uses a variety of genomic techniques to characterize the microbial communities in environmental samples and to sequence the genomes of uncultured microbes, revealing a variety of catabolic genes, taxonomically and phylogenetically significant genes, and entire operons. The development of sequencing techniques and other cost-effective methods for large-scale microbial community analysis has resulted in novel applications such as comparative community metagenomics, metatranscriptomics, metabolomic and metaproteomics. Metatranscriptomics study mRNA within a cell or organism while metaproteomics study protein patterns in microbial communities. Metabolomics focuses on metabolites released by the organism into its immediate surroundings to provide helpful information about the microbiome's characteristics and significant dependence on external influences (Kamble *et al*., 2020; Simon & Daniel., 2011).

The rapid industrialization and urbanization of the world result in the discharge of industrial wastewater from various industries which is thought to be the main cause of environmental contamination (soil and water). Numerous organic and inorganic contaminants found in industrial wastewater seriously harm the environment and cause health risks to humans and other organisms. Inorganic pollutants include a variety of toxic heavy metals like chromium (Cr), cadmium (Cd), arsenic (As), mercury (Hg), etc. Organic pollutants include phenols, polyaromatic hydrocarbons, chlorinated phenols, azo dyes, pesticides, polychlorinated biphenyls, etc. Removing organic and inorganic pollutants from industrial wastes using bioremediation is an environmentally acceptable waste management technique. The detoxification and breakdown of organic and inorganic contaminants from industrial wastewater is mostly dependent on microorganisms. Since microbes are the primary agents in bioremediation, changes in their composition and activity may have an impact on the fate of pollutants in the environment. Therefore, it is essential to understand the environmental microbial communities involved in the bioremediation of environmental pollutants. Metagenomics is currently being used to thoroughly analyze the makeup and activity of the microbial population during bioremediation in a contaminated environment. The primary enzymes and genes involved in the breakdown and detoxification of environmental contaminants can also be reliably identified using metagenomics methods, including advanced sequencing technologies [next-generation sequencing (NGS) and third-generation sequencing (TGS)] and its omics approaches (Zhang *et al*., 2021). This paper discusses how meatagenomics, as well as its 'omic' methodologies aid in identifying environmental microorganisms and their applications in bioremediation of environmental pollutants.

METAGENOMICS ANALYSIS

The concept of metagenomics was first proposed by Handelsman *et al.* (1998). It includes the extraction of the genomic DNA from every living organism inhabiting the sample collected from the environment, which is then used in cloning, transformation, and subsequent screening of the constructed metagenomic library. Metagenomic libraries can be explored in two ways: sequence-driven and function-driven screening. The sequenced-based analysis uses the DNA sequences of genomic fragements in order to identify the clone of origin using high-throughput sequencing technologies [such as 454 pyrosequencing, Illumina (Solexa) sequencing and SOLiD (Sequencing by Oligonucleotide Ligation and Detection) sequencing and third generation sequencing]. In function-based screening, the function of a gene's products is used to identify the clone of origin. Currently, substrate-induced gene expression screening (SIGEX), Fluorescence in situ hybridization and DNA stable-isotope probing (DNA-SIP) are further methods used to screen environmental metagenomic libraries. After sequencing, the tens of thousands to millions of short reads produced must be computationally turned into useful information that reflects the existence and abundance of the target microbes. Several cutting-edge "omics" technologies have made it possible to introduce new study areas into the metagenomics. (Chistoserdova., 2009; Neelakanta & Sultana., 2013; Ngara & Zhang., 2018; Thomas *et al*., 2012).

'META-OMICS' APPROACHES

The suffix "-ome," which comes from a Greek word that signifies "all," or "complete," is where the term "omics" gets its name. The term "-omics" is widely used to denote a field of research in the biological sciences that highlights the utilization of large-scale, high-throughput data and information in order to understand the "omes" of life. In the past two decades, a variety of omics tools have been created to gather and analyse high throughput data on proteins (proteomics), mRNA transcripts (transcriptomics), microbial diversity (metagenomics), metabolic profile (metabolomics) etc., of a particular cell, tissue, organ or whole organism at a specific time point. Using network modeling of omics data, it is possible to conduct unbiased studies of the molecular mechanisms, interactions and functions of individual cells, tissues, organs and the entire organism.

Unlike the genome, the transcriptome is dynamic and is made up of a variety of players. Metatranscriptomics is the study of the extraction and analysis of metagenomic mRNA (metatranscriptome). It provides details on the expression and regulation profiles of the complex microbial communities that were discovered in environmental samples. By capturing the total mRNA, it offers a snapshot of the gene expression in a given sample at a given time and under a specified set of conditions. Metatranscriptomics can be used to both identify the microorganisms that are a part of an ecosystem and to learn more about how they function. This is similar to metagenomics methodologies. This technique uses high-throughput methods, such as microarray and RNA-seq. Due to challenges with processing environmental RNA samples such as recovering high-quality mRNA from environmental samples, short half-lives of mRNA species and separating mRNA from other RNA species, in situ metatranscriptomic studies of microbial communities are currently uncommon. Applying direct cDNA sequencing using NGS technologies can get beyond these restrictions. This enables whole-genome expression profiling of a microbial population and offers affordable access to the metatranscriptome. Additionally, it is also possible to directly quantify the transcripts (Simon and Daniel, 2011). Pyrosequencing was originally used by Leininger *et al.* (2006) to identify the active genes in soil microbial communities which demonstrated the involvement of ammonia-oxidizing archaea in soil ecosystems. According to Shi *et al.* (2009), small RNAs (sRNAs) have been shown to play a role in a variety of environmental processes including nutrition intake and carbon metabolism. DNA microarray-based environmental metatranscriptome analysis has received a lot of attention lately. A microarray device created for whole-community genome analysis combined by Gao *et al*. (2007) with whole-community RNA amplification. Parro *et al.* (2007) used a similar methodology for the

analysis of the metatranscriptome of the Tinto River, and the results were limited to the transcriptome analysis of *Leptospirillum ferrooxidans*, the dominant species found in that environment (Aguiar-Pulido *et al*., 2016; Simon and Daniel, 2011; Lee *et al*., 2017).

Metaproteomics is often known as environmental proteomics or community proteomics. It is defined as the study of all the protein samples isolated from environmental sources. Ram *et al.* (2005) performed one of the most thorough metaproteomics investigations, examining the gene expression, essential activities and metabolic processes of a naturally occurring microbial biofilm from an acid mine drainage. This led to the identification of more than 2000 proteins from the five most prevalent bacteria. Detecting and identifying all proteins generated by a complex environmental microbial community remains a difficult undertaking. Uneven species distribution, the wide range of protein expression levels found in microorganisms and the significant genetic heterogeneity found in microbial communities are all difficulties for metaproteomics research. Despite these difficulties, the use of metaproteomics has the potential to significantly improve the understanding of how microbial communities' genetic diversity and activities affect ecosystem function (Simon and Daniel, 2011). Proteomics, like transcriptomics, is able to undertake qualitative and quantitative analysis of all the proteins produced by bio-mining bacteria under various stresses. To date, acid mine drainage (AMD) has been the primary source of the bio-mining microbial communities examined by metaproteomics technology. Metaproteomics and isotope labeling methods can also be utilized to examine the patterns of nutrient transport among the members of microbial communities and the metabolic responses in mixed microbial systems. For instance, because of their species uniqueness, the differential fractionation of stable hydrogen isotopes in proteins can disclose the nutritional levels of the constituents of microbial communities. A proteomic experiment using ^{2}H and ^{15}N isotopes shown that the archaea in AMD biofilms get nitrogen through recycling nitrogen-containing proteins (De Maayer *et al*., 2014; Li & Wen., 2021).

The analysis of the metabolome is a newly- emerging application field for omics techniques. Metabolomics is the comprehensive study that identifies and quantifies all a sample's metabolites (small chemicals that organisms release into its immediate surroundings). It is regarded as a helpful indicator of the health of an ecosystem. Metabolomics also seeks to advance knowledge of how the microbiome affects the host environment's homeostasis by transforming nutrients, toxins and other abiotic elements. Additionally, the metabolome can demonstrate signaling mechanisms involved in bacterial communication, such as quorum sensing, which links changes in gene expression to changes in cell population density. Metagenomics and metatranscriptomics, which largely rely on sequencing, are very different from metabolomics in terms of how data are generated. The identification and quantification of metabolites are typically accomplished by combining chromatography methods (such as liquid and gas chromatography) and detection techniques, such as mass spectrometry (MS) and nuclear magnetic resonance (NMR), which generate spectra consisting of patterns of peaks that enable both the identification and quantification of metabolites. (Aguiar-Pulido *et al*., 2016).

BIOINFORMATIC TOOLS FOR METAGENOMIC

Bioinformatics plays a variety of roles in metagenomic bioremediation, most notably during metagenomic data analysis. The Meta Genome Analyzer (MEGAN) is a popular software tool for evaluating huge amounts of metagenomic sequencing data. This programme is best suited for interactively analyzing and comparing taxonomic and functional metagenomic and metatranscriptomic data. Simple Metagenomics Analysis SHell for microbial communities is a standalone metagenomic annotation and analysis pipeline that shares design concepts and routines with SmashCell. It is appropriate for data supplied by Sanger and 454 sequencing technologies. Rapid Annotation Using Subsystems Technology for Metagenomes (MG-RAST) is an automated analytic tool for metagenomes that provides quantitative insights into

microbial populations based on sequencing data. The Integrated Microbial Genomes and Metagenomes (IMG/M) system allows for the annotation, analysis, and sharing of microbial genome and metagenome data (Zhang *et al.,* 2021).

APPLICATIONS OF METAGENOMICS AND OMICS IN BIOREMEDIATION

Bioremediation is a new bioengineering method that makes advantage of the natural cleaning capabilities of environmental microorganisms. The goal of bioremediation is to use the metabolic processes of the microorganisms in the treatment system to either reduce or make harmless contaminants. There are numerous conditions for bioremediation: (1) Microorganisms have the ability to break down organic pollutants. (2) The microorganisms have specific metabolic processes. (3) The environment is conducive to deterioration while still allowing the continuation of microbial activity. A novel microbial strain with high degrading efficiency, widespread applicability and stable expression may be created and screened out with the aid of metagenomics, which may also assist researchers in discovering new functional microorganisms and genes. In the sewage treatment project, it is necessary to eliminate the nitrogen and phosphorus components of the sewage in addition to the organic matter that is present. The wastewater treatment process interacts with a variety of microbial communities, including bacteria that oxidize ammonium under anaerobic conditions, bacteria that accumulate phosphorus and bacteria that are electrochemically active. The biological nitrogen removal mechanisms and improved biological phosphorus removal are both influenced by metagenomic studies of water microbes. Microorganisms capable of metabolizing pesticides, plastics, polycyclic aromatic hydrocarbons, petroleum hydrocarbons, and other organic contaminants have functional genes that can be crucial in environmental bioremediation. The use of these "omics" tools in bioremediation research has substantially benefited in identifying new biodegradation routes and describing or monitoring pollutant-biodegrading microbial populations. There are numerous researches that describe building and screening metagenomic libraries to find the genes involved in bioremediation. A combined proteomic technique based on 2DE/MS and cleavable isotope-coded affinity tag analysis was used by Kim *et al.* (2006) to examine the aromatic hydrocarbon catabolism pathways in *Pseudomonas putida* KT 2440. Keum *et al.* (2009) has discussed research on the comparative metabolome analysis of *Sinorhizobium sp*. C4 during the breakdown of phenanthrene. A fluxomics investigation was also carried out by Tang *et al.* (2009) on *Shewanella sp*., which is known to contain catabolic pathways for the bioremediation of radionuclides, hazardous metals and halogenated organic chemicals. The potential of bioremediation has yet to be effectively explored. One explanation for this is that bioremediation techniques that are effective in one location might not be effective in another. Furthermore, microbial systems that successfully remediate contaminants in the lab may not perform as well as in the field. The principles that govern microorganism development in polluted environments are poorly understood, which limits the applicability of bioremediation. Native microorganisms are used in the microbial-bioremediation technique to remove pollution. The cleaning of pollutants is regulated by a variety of parameters, including the structure of native microbial species and environmental factors. As a result, it is critical to consider a wide range of parameters while designing the bioremediation process in order to comprehend and foresee the effects of toxins in the environment. Environmental microbiologists have been able to address these difficulties because of the advancement of sequencing technologies (next-generation sequencing (NGS), and third-generation sequencing such as Oxford nanopore and SMRT sequencing) and in-silico research, which have benefited them in identifying active microorganisms in polluted environments. As the cost of next-generation sequencing (NGS) falls, transcriptomics/proteomics/bolomics become more viable choices for rapidly scanning contaminated locations for specific degradative processes and identifying microbes for pollutant breakdown. These omics technologies aid in the evaluation of bioremediation performance and the development of bioremediation strategies. Implementation of many bioremediation techniques would aid in improving remediation

efficiency (Amer & Baidoo., 2021; Breitwieser *et al*., 2017; Li & Wen., 2021; Neelakanta & Sultana., 2013; Sharma *et al*., 2022).

CONCLUSION

Metagenomics is a rapidly emerging study topic. The improvement of DNA isolation technologies, cloning methodologies and screening techniques enabled the evaluation and use of microbial populations from severe and unfriendly habitats. Metagenomic investigations can be supplemented by whole-genome analysis and related omics research to further understand microbial populations in a global setting. The omics techniques include meta-genomics, meta-transcriptomics, meta-proteomics and metabolomics. Microbial bioremediation is a low-cost and environmentally friendly method of eliminating contaminants from ecosystems. Knowledge of bacteria in contaminated ecological matrix is required to better understand the mechanism of bioremediation of a specific contaminant and identify the key enzyme of the catabolic gene involved. The introduction of new sequencing technologies [from shotgun sequencing to high-throughput, next-generation sequencing (NGS), and third-generation sequencing (TGS)] and advanced bioinformatics tools has provided crucial insights into microbial communities and underlying mechanisms in environmental contaminant bioremediation. Academics are employing omics tools in the bioremediation process to better understand the underlying mechanism and devise innovative ways. Bioremediation efficacy will surely improve if accurate molecular approaches are applied and scientifically researched. Eventually, metatranscriptomics, proteomics and metabolomics will be coupled with metagenomics enable to investigate everything from genes to proteins and from structure to function.

REFERENCES

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., & Narasimhan, G. (2016). Metatranscriptomics, and Metabolomics Approaches for Microbiome Analysis. *Evolutionary Bioinformatics*, 12(S1), 5–16.

Amer, B., & Baidoo, E.E.K., (2021).Omics-Driven Biotechnology for Industrial Applications.*Frontiers in Bioengineering and Biotechnology.* 9:613307.

Breitwieser, F. P., Lu, J., & Salzberg, S. L. (2017). A review of methods and databases for metagenomic classification and assembly. *Briefings in Bioinformatics*, 1–15.

Chistoserdova, L. (2009). Functional metagenomics: Recent advances and future challenges. *Biotechnology and Genetic Engineering Reviews*, 26(1), 335–352.

Cui, H., Li, Y., & Zhang, X. (2016). An overview of major metagenomic studies on human microbiomes in health and disease. *Quantitative Biology*, 4(3), 192–206.

De Maayer, P., Valverde, A., & Cowan, D. A. (2014). The Current State of Metagenomic Analysis. *Genome Analysis: Current Procedures and Applications*, 183–220.

Lee, P. Y., Chin, S.F., Neoh, H.M., & Jamal, R. (2017). Metaproteomic analysis of human gut microbiota : where are we heading ?. *Journal of Biomedical Science*, 24(36), 1-8.

Li, M., & Wen, J. (2021). Recent progress in the application of omics technologies in the study of bio-mining microorganisms from extreme environments. *Microbial cell factories*, 20(1), 178.

Neelakanta, G., & Sultana, H. (2013). The Use of Metagenomic Approaches to Analyze Changes in Microbial Communities. *Microbiology Insights*, 6, 37-48.

Sharma, P., Singh, S. P., Iqbal, H. M. N., & Tong, Y. W. (2022). Omics approaches in bioremediation of environmental contaminants: An integrated approach for environmental safety and sustainability. *Environmental research*, *211*, 113102.

Simon, C. R., Daniel, R. (2011). Metagenomic analyses: Past and future trends. *Applied and Environmental Microbiology*, 77(4), 1153-1161.

Ngara, T. R., & Zhang, H. (2018). Recent Advances in Function-based Metagenomic Screening. *Genomics, proteomics & bioinformatics*, 16(6), 405–415.

Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microbial informatics and experimentation*, 2(1), 3.
Zhang, L., Chen, F., Zeng, Z., Xu, M., Sun, F., Yang, L., Bi, X., Lin, Y., Gao, Y., Hao, H., Yi, W., Li, M., & Xie, Y. (2021). Advances in Metagenomics and Its Application in Environmental Microorganisms. *Frontiers in microbiology*, 12, 766364.