International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

Metagenomics: A Helping Hand

Mishika Ahuja[1], Sonali Thapa[2]

University Institute of Biotechnology, Chandigarh University, Mohali, Punjab, India

Abstract: ***Increasing population, industrialization and removal of natural resources leads to increase in more pollutants in the atmosphere, creating problems for living organisms. Bioremediation is the most effective way to cure the problems related to the environment. Using the genomic level of microorganisms, the dataset is collected, and then annotation is done using next-generation sequencing methodologies. Hence, it is referred to as a peerless technology for bioremediation up-gradation. Without any need for the previous culture, the evaluation of DNA from microbial communities is usually termed as metagenomics. For the analysis, many computational tools have been developed and used for the exploitation of great invasion of data. This review mainly focuses on the high-tech computational tools used in sequencing technologies along with its further aspects. Moreover, we also provide applications for future related issues. We also discussed the evaluation of comparative metagenomics of diverse datasets using already developed datasets and annotation specifications.***

Keywords: Bioremediation, computational tools, sequencing technologies, metagenomics, datasets, next-generation sequencing

1. Introduction

Over the last few decades, the witness of the harmful effects on the environment of pollution has been administered. Pollutants contaminate the natural ambiance and also agitate the pure environment causing global cause. A natural pollutant, as well as anthropogenic activities, lead to the disturbance of ambiance which results in pollution or global cause in the environment state. The only solution to this major disaster is that an individual must attain to manage the ecosystem, to achieve sustainable development. For this, current technologies and innovative methods are used to refine their actions.

The major pollutants involved in contamination are nitrates, hydrocarbons, aromatic substances, radioactive substances, heavy metals, arsenic, lead, organic pollutants, etc. By contaminating the environment, these substances lead to an increase in gases like carbon dioxide, nitrogen oxide, Sulphur oxide, and some traces of particulate matter with the increase in the graph of global warming, ozone layer depletion, and many health hazards.

Microorganisms such as bacteria, fungi, algae, etc. are used toimprove these vulnerable environmental conditions. Novel methodologies have been derived for bioremediation using nucleic acids of these microorganisms. Nowadays, the field of Metagenomics has proven a good helping hand for scientists, in the Bioremediation of various environmental pollutants. “Metagenomics refers to the idea of extraction of DNA, directly from the environment samples by analyzing them with DNA sequencing techniques”. This study comprises genome level identification of communities with methods imitated from genomics. As observed in the “Metagenomics” section, metagenomics also has multiple properties. Despite all, its foremost purpose is to figure out the taxonomic sketch of microorganisms. Though whole-metagenome sequencing (WMS) provides a fractional glimpse of functional data for the microbial community, it’s better to imply metatranscriptomics, which includes sequencing of complete (Meta) transcriptome of the microbial community. Thus, this form of recovery of waste uses no toxic chemicals, although the microorganisms used may be destructive under certain conditions. The process is slow and often requires enhancement by a nutrient supply (Bio stimulation). Uprising bioremediation, using Metagenomics showed great interest in studies done recently. Metagenomics evolved to be an effective technique of bioremediation in the formation of a non-toxic (pure) environment.

This study mainly targets on the freshly modified sequence and function-based metagenomic techniques to evaluate metagenomes from polluted sites. Moreover, the study also explains the highly new metagenomes derived from metagenomic communities, and also immensely capable of degrading contaminations and toxins in the environment (Martínez-Porchas, M., & Vargas-Albores, F. 2017). The study mainly signifies several cases of applications of Metagenomics in water, air, and soil contaminants and this review mainly focuses on the major developments in the field of environmental microbiology and biotechnology which can be credited to Metagenomics research.

2. Sequencing Technologies

Nowadays, two types of next-generation sequencing technologies are used: the 454 Life Sciences and the Illumina systems, so it is essential, to sum up, their advantages and disadvantages of the sequencing of metagenomic samples (Aguiar et al. 2016).

The first next-generation sequencing technology was 454 pyrosequencing and introduced in 2004. It works on the principle of immobilization of DNA fragments on DNA - secured bead in a water-oil emulsion and PCR is used to magnify the fixed fragments. A Picotiter Plate is used to place beads on it along with the DNA polymerase and then pyrosequencing is performed. The difference between the classic Sanger sequencing and the pyrosequencing is that pyrosequencing depends on the detection of pyrophosphate release on nucleotide incorporation in place of chain termination with dideoxynucleosides. The walkout of pyrophosphate is transmitted to light using enzyme reactions, which is then converted into actual sequence information (Oula et al. 2015).

For the production of local colonies through the attachment of DNA molecules to primers on a slide, followed by amplification of that DNA is done using Illumina dye sequencing. This type of "DNA clusters" is cultivated by the augmentation of fluorescently labeled, reversible terminator bases (adenine, cytosine, guanine, and thymine) attached with a blocking group (Tripathi et al. 2016). The bound four bases on the template DNA will be then sequenced and the non-bonded once are washed off. After every cycle of synthesis, a laser is applied to elicit the dyes and a high-consistency scan of the fused base is made. At last, a chemical unclog step confirms the removal of the 3' terminal blocking group and the dye in a single step. Until the full DNA molecule is not sequenced the process will be performed repeatedly. It consists of a variety of instruments, for example, 1) MiSeq: it has an output of 15 GB and 25 million sequencing reads of 300 bp in length; clustered fragments can also be sequenced from both ends (paired-end sequencing), which can be merged so that 600 bp reads can be obtained. 2) For greater output HiSeq2500 is used, which offers 125bp reads (Chandra, R. (Ed.). 2015).

3. Shotgun Metagenomics

Collection of metagenomic based data

These studies are commonly used to evaluate the specific genome which is present in the environmental community. While performing shotgun metagenomics the whole sequences of protein-coding genes as well as the operon sequences provides specific knowledge about the community. Due to these judgments, a collection of shorter segments into genomic contigs and accumulation of these into scaffolds is mostly done to obtain a short and concise form of the sequenced community under evaluation.

A recently developed IDBA-UD algorithm is used to identify the major issues of metagenomic sequencing technologies at its major depth. It uses multiple depth-relative k-mer thresholds for the removal of false k-mers in both low-depth and high-depth regions. Comparison between both the tools concerning their capabilities is done using N50 length score (Megharaj et al. 2011), usually referred to as a collection of all the contigs in the assembly. A recent study shows that IDBA-UD can regenerate longer contigs with higher accuracy.

The mechanism of collection of shorter reads into contigs can be stated in two different ways: 1) reference-based collection and 2) de novo collection, the dataset is first examined and then the specific route is followed. For the representation of each genome belonging to specie, one or more reference genome is used for the contigs formation. Commonly used tools for the reference-based collection are Newbler (Roche), MIRA 4, MetaAMOS (Megharaj et al. 2011). When no previous reference is used o already known genome for their contigs formation is referred to de novo collection. Tools related to this technique work based on the graphical algorithm for example de-Bruijn graphs, EULER, Velvet, SOAP, and Abyss (Tikhonovich et al. 2017).

Tools used in Binning

Binning is the technique used to bind contigs into the specific genome, and impute the groups to individual species. There are two ways by which binning can be performed 1). Binning based on combination states that single genome have different partitioning of K-mer sequences, using these preserved species, a grouping of sequences can be made into their corresponding genomes. 2). Harmony based binning- refers to the the technique of using algorithms based on alignment such as BLAST and or profile hidden Markov Models (pHMMs), to provide identifying information about individual sequences from provided databases (Tikonovich et al. 2017).

Accessible combination-based binning algorithms are numbered as TETRA, S-GSOM, Phylopythia and PhylopythiaS, TACAO, PCAHIER, ESOM, and ClaMS, although, the examples of similarly-based binning tools consist of CARMA, MetaPhyler, and SOrt-ITEMS. Based on the type of algorithm binning of tools are classified into: *ab initio* unsupervised classifiers and 2) training-based classifiers. The method of using pre-existing bins copied from the genome's sequence to distinguish a given dataset without the user's supervision is referred to as unsupervised binning. While, in supervised binning, the user's supervision and interference are allowed in the whole process. (Tikonovich et al. 2017).

Annotation of metagenomic sequences

While working with the mixtures of genomes and contigs of various lengths, annotation of metagenomes is strictly required. The first step requires the arrangement of reads for annotation, these include: 1) *Trimming of low-quality reads* using FASTX-Toolkit; 2) *Masking of low-complexity* using DUST; 3) removal of 95% of similar sequences is done using *de-replication* step 4) *screening* of genomes is performed using Bowtie2; 4) *Identification of genes* also called *gene calling*, genes are specified as coding DNA sequences (CDSs) and non-coding RNA genes and some annotation pipelines also conclude regulatory elements, for example, clustered regularly interspaced short palindromic repeats (CRISPRs). Tools used for the identification of CDSs are MetaGeneMark, Metagene, Prodigal, Orphelia, and FragGeneScan. For recognition of CRISPER elements following tools will be used: CRT and PILER-CR, IMG/MER and examination of non-coding RNA genes will be done using programs like tRNAscan for tRNA, ribosomal RNA (rRNA) genes (5s, 16s, and 23s) will be detected using IMG/MER, and MG-RAST (Tikonovich et al. 2017).

Functional assessment of protein-coding genes

This step can be performed using a homology-based evaluation of sequences across datasets, mainly using BLAST or another sequence-based similarity algorithm. A latterly developed EBI metagenomic tool uses metadata network that fulfills the Genomic Standards Consortium (GSC) guidelines. It works on the principle of extraction of rRNA data from shotgun metagenomics using tools such as rRNASelector for metagenomic evaluation. CAMERA is another tool used in the reduction of sequencing cost by sample mixing (Tikonovich et al. 2017).

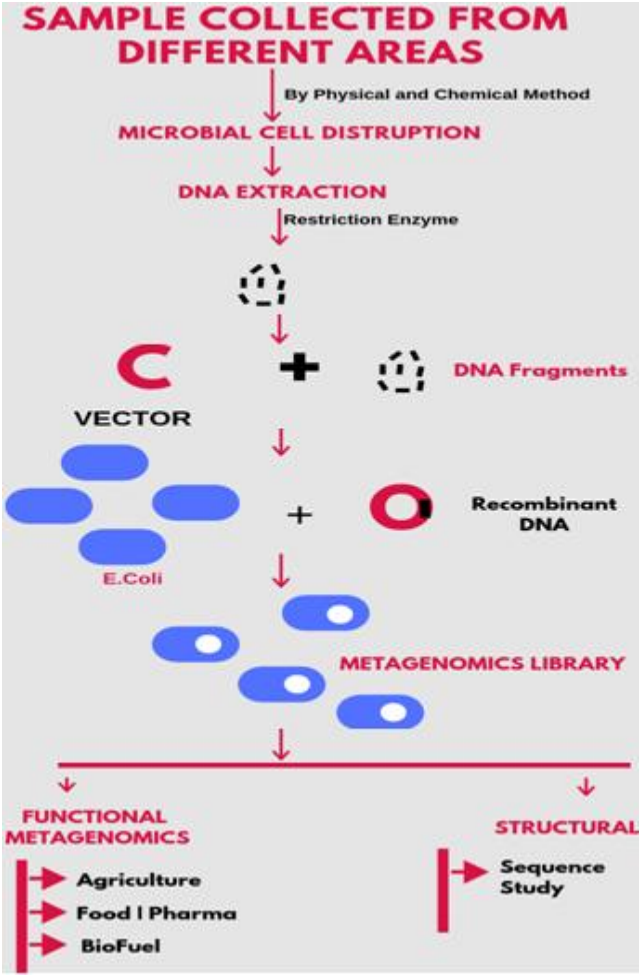

Figure: Steps include in the metagenomic process

4. Application of metagenomics

Bioremediation

Environmental pollution is the major and utmost problem worldwide. The decline in the health of the environment is directly linked to human health. Urbanization and industrialization are the main reason for the raising of pollution by directly discharging their waste into the natural bodies, that are loaded with chemical (organic, inorganic, and heavy metals) and other toxic compounds. Due to the poor biodegradability and non-biodegradability of organic compounds, it becomes a major challenge for environmental stability and human health safety (Bharagava et al.2017). Bioremediation is the removal or conversion of toxic waste into the less toxic compound; we use predominantly bacteria (fungi and algae sometimes) for the removal of toxic waste from natural bodies and the environment (Bharagava et al.2017). Metagenomics and the sequencing methods are used for the identification of key enzymes involved in the biotransformation and biodegradation of toxic waste. Numerous conventional or culture depending and non-culture depending techniques are used for the identification of microbial community that helps in the bioremediation (Xing et al. 2017).

Biofuel Production

Due to the shortage in the mineral oil and continuously rising to pollution the bioenergy and biofuel sector is gaining more attention to researchers. These are an environmentally friendly product, derived by the manipulation of waste such as plant waste known as biomass. Biomass contains lignocellulose and cellulose sperms; the process include constitutes the conversion of cellulose present in biomass into cellulosic ethanol (Xing et al. 2017). Metagenomics is employed for the identification of novel enzyme through high throughput sequencing method (Zimmer and Carl 2010). Enzymes such as ligases, endoglucanase, xylanase, etc are for bioethanol and lipolytic enzyme for biodiesel (Zimmer and Carl 2010). Cellulose and xylanase is the most preferred hydrolysis enzyme across the world, vast method are used for the identification of this enzyme from various microorganisms (including extremophiles), one them is HTS (high throughputs screening) which is time consuming and precise which identifies the microbe producing the enzyme of interest by sequence screening method (Ayala et al. 2016).

Gut Microbe Characterization

Microbial biota plays a significant role in maintaining equanimity in human or living body but their mechanism and composition remain unknown (Nelson et al. 2010). Any disturbance in the microbial community leads to the declination of health and increase the susceptibility to a pathogen (This part of metagenomics is related to the human microbiome initiative and the the primary goal is to identify if there is core human microbiota and their relation with human health through the techniques of metagenomics (Qin et al.2010).

METHIT (metagenomics of human intestinal tract) is the project which constitutes 124 individuals from Spain and Denmark are healthy, overweight and patient of irritable bowel disease. To characterize the phylogenetic diversity of gut microbe, through metagenomics technologies it was revealed that patient with bowel disease have less microbial diversity than the healthy one

Food and Pharmaceutical Industries

Microbes play significant role in food and pharma industries to increase the aroma, taste and shelf life of food and the microbial enzymes helps in the manufacture of medicines through the the biocatalytic activity of enzymes (Koonin and E.V 2018). In food and pharmaceutical industries through metagenomics techniques the clones of actual enzymes are prepared that are phylogenetically screened to identify homology with the actual enzyme (Boddu et al. 2018). Lipase is the hydrolytic an enzyme that plays an important role in food and pharmaceuticals industry, in 2014 Peng et al. discovered highly alkaline-stable lipase which has high specificity to buttermilk fat Easter, trough the screening of metagenomic virus library of cloned E. coli host (Boddu et al. 2018).

Disease Diagnosis

Identification of infection and pathogen responsible for the cause of infection, biodefence is the necessity of biosecurity towards the infectious disease-causing war pathogens for the protection against bioterrorism. To cure the disease outbreak new strategies and approaches are applied and the metagenomics analysis is a time-saving tool for the identification and detection of pathogens and disease caused by them. Next-generation sequencing is the method of the

choice for diagnosis of diseases because the identification of the causative a pathogen that may be bacteria, fungi or virus can be revealed by this method by obtaining sequence data (Zhang et al.2018).

5. Conclusion

In this era pollution is becoming a serious problem, Metagenomics acting as an actual helping hand to recover the ecosystem. The research in this field rising every single second, this seems metagenomics has endless benefits for humans and the environment. Researchers aim to share, estimate, and evaluate the outcome. The downside of this field that, this is still expensive for massive projects of sequencing, and during shotgun sequencing method only genes that are in dominating form are represented. But the metagenomics is boon to Bioremediation for the restoration of the environment. Functional metagenomics has a great advantage for the making of clones and the formation of novel enzymes that has massive advantages in industries and the agriculture field.

References

[1] Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., & Narasimhan, G. (2016). Metagenomics, Metatranscriptomics, and Metabolomics Approaches for Microbiome Analysis: Supplementary Issue: Bioinformatics Methods and Applications for Big Metagenomics Data.Evolutionary Bioinformatics, 12, EBO-S36436.

[2] Ayala-Mendivil, N., de Los Angeles Calixto-Romo, M., Amaya-Delgado, L., Casas-Godoy, L., & Sandoval, G. (2016). High Throughput Screening: Developed Techniques for Cellulolytic and Xylanolytic Activities Assay.

[3] Bharagava, R. N., Saxena, G., Mulla, S. I., & Patel, D. K. (2017). Characterization and identification of Hitler recalcitrant organic pollutants (ROPs) in tannery wastewater and its phytotoxicity evaluation for environmental safety.

[4] Boddu, R. S., & Divakar, K. (2018). Metagenomic sights into Environmental Microbiome and Their Application in Food/Pharmaceutical Industry. Coughlan, L. M., Cotter, P. D., Hill, C., & Alvarez-Ordóñez, A. (2015).

[5] Chandra, R. Koonin, E. V. (2018). Environmental microbiology and metagenomics: the Brave New World is here, what's next?.Environmental microbiology.

[6] Coughlan, L. M., Cotter, P. D., Hill, C., & Alvarez-Ordóñez, A. (2015). Biotechnological applications of functional metagenomics in the food and pharmaceutical industries.Frontiers in microbiology, 6, 672.

[7] (Ed.). (2015). Advances in biodegradation and bioremediation of industrial waste. CRC Press.

[8] Megharaj, M., Nelson, K. E., White, B. A., & Marco, D. (2010). Metagenomics and its applications to the study of the human microbiome. Metagenomics: Theory, Methods and Applications, 171-182.

[9] Microbial metagenomics in aquaculture: a potential tool for a deeper insight into the activity. Reviews in Aquaculture, 9(1), 42-56.

[10] Oulas, A., Pavloudi, C., Polymenakou, P., Pavlopoulos, G. A., Papanikolaou, N., Kotoulas, G., ... & Iliopoulos, L. (2015). Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies.Bioinformatics and biology insights, 9, BBI-S12462.

[11] Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., ... & Mende, D. R. (2010). A human gut microbial gene catalog established by metagenomic sequencing. nature, 464(7285).

[12] Tikhonovich, I. A., Ivanova, E. A., Pershina, E. V., &Andronov, E. E. (2017). Metagenomic technologies of detecting genetic resources of microorganisms. Herald of the Russian Academy of Sciences, 87(2), 115-119

[13] Tripathi, M., Singh, D. N., Vikram, S., Singh, V. S., & Kumar, S. (2018). A metagenomic approach towards bioprospection of novel biomolecule (s) and environmental bioremediation

[14] Xing, M. N., Zhang, X. Z., & Huang, H. (2012). Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. Biotechnology advances,30(4), 920-929.

[15] Zhang, X., Zhang, M., Ho, C. T., Guo, X., Wu, Z., Weng, P., ... & Cao, J. (2018). Metagenomics analysis of gut microbiota modulatory effect of green tea polyphenols by high fat diet-induced obesity mice model. Journal of Functional Foods, 46, 268-27

[16] Zimmer, Carl (13 July 2010). "How Microbes Defend and Define Us". New York Times. Retrieved 29 December 2011.