AgriCos e-Newsletter
ISSN: 2582-7049
Volume : 01 Issue : 06 October 2020 Article No: 08

Metagenomics: New Approach in Identifying Novel Gene

Bahatkar B. P.
Ph.D. Scholar, Biotechnology Centre, Department of Agricultural Botany, Dr. PDKV Akola, (M.S)

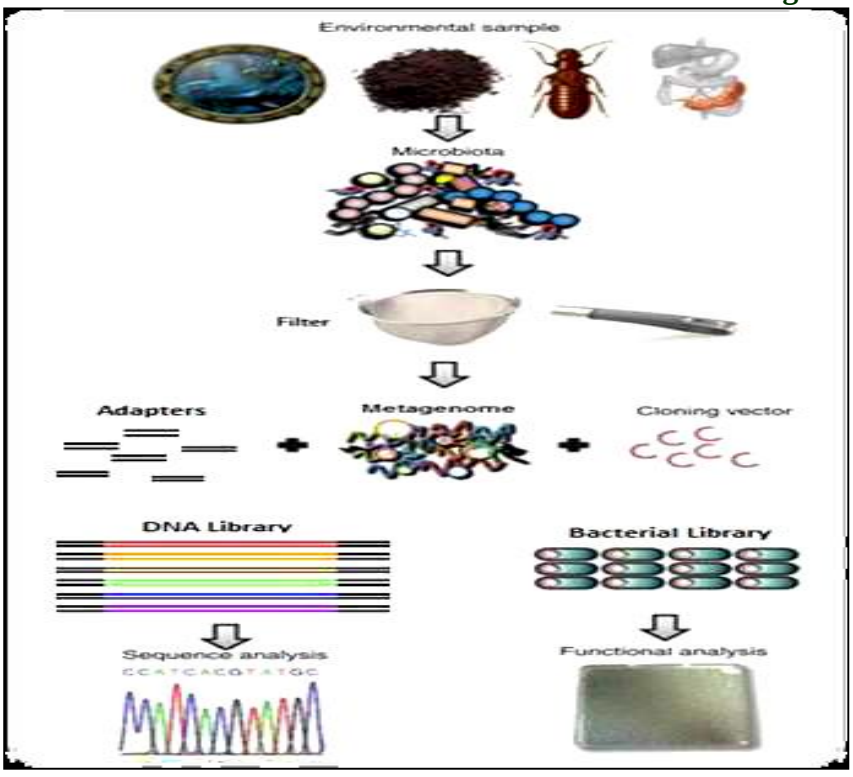
SUMMARY

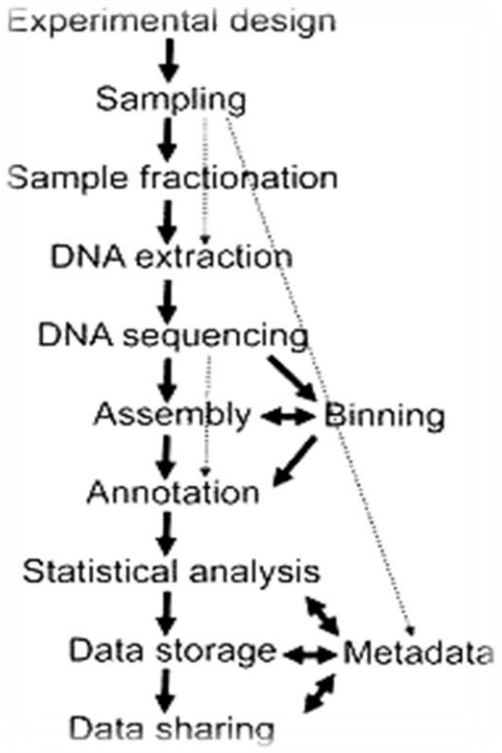
Microorganisms constitute 2/3rd of the Earth's biological diversity out of which almost 99% of the microorganism cannot be cultured by standard techniques. So there is requirement of culture- independent methods to know the genetic diversity, population structure and ecological roles of the majority of organisms. Metagenomics deals with the isolation of genetic material straightly recovered from environmental samples. Metagenomics as an approach has come up over the past two decades to elucidate a host of microbial communities inhabiting a specific niche with the aim of understanding their genetic diversity, population structure, and ecological role played by them. A number of new and novel molecules with significant functionalities and applications have been identified through this approach. Metagenomics can also be applied to solve practical challenges in the field of medicine, agriculture, sustainability, and ecology. Here we discuss overview of metagenomics, the various methodologies and tools developed to understand the biology of uncultured microbes including metagenomic analysis.

INTRODUCTION

Metagenomics is the study of metagenomes (collective genome of microorganism), genetic material recovered directly from environmental samples, also be referred to as environmental genomics, ecogenomics or community genomics. The term "metagenomics" was introduced by Jo Handelsman, Jon Clardy, Robert M. Goodman, and first appeared in publication in 1998. In Greek, Meta means "transcendent" (combination of separate analysis) and Genomics refers to the study of the genome. Metagenomics can unlatch the massive uncultured microbial diversity present in the environment for new molecule for therapeutic and biotechnological application. It also studies identification of many novel microbial genes coding for metabolic pathways such as energy acquisition, carbon and nitrogen metabolism in natural environments that were previously considered to lack such metabolism.

Overview of Metagenomics





Why there is a Boom of Metagenomics?

- Microorganisms constitute two third of the Earth's biological diversity.
- As many as 99% of the microorganisms present in certain environments cannot be cultured by standard techniques.
- Culture-independent methods were required to understand the genetic diversity, population structure and ecological roles of the majority of organisms.
- Science of metagenomics makes it possible to investigate resource for the development of novel genes, enzymes and chemical compounds for use in biotechnology.
- Microbes, as communities, are key players in maintaining environmental stability.
- Investigate microbes in their natural environment, the complex communities in which they normally live in.
- High-throughput gene-level studies of communities.

Overview of Metagenomics

1.Site Selection

Selection of ecosystem from where you have to isolate the sample for example; ocean, sea, volcanoes, soil etc.
Two important notes-

a. Collect more information about the habitat (physical, chemical, and ecological) so more insight can be derived from the metagenomic data.
b. The discovery of the ***keystone species replies on*** knowledge of the site.

2.Sampling

- First and most crucial step in metagenomics.
- Sampling Strategy
- Type, size, scale, number
- The samples must be representative to the habitats

3.Filtering

Size based separation
This step is to ensure that sufficient enrichment of the target is achieved and that minimal contamination of non-target material occurs.
Goals:
1. Get as much as of what we need, and
2. Leave out as much as what we don't need.

4.DNA Extraction

Macromolecule recovery

- The quality and completeness of data obtained from metagenomic analysis of any of the community will be only as good as the procedures used for the extraction of DNA from a sample.
- DNA from dead cells (may be important in drawing conclusions about the overall metabolic capabilities of a microbial community)
- **DNA extracted should be representative of all cells present in the sample and sufficient amounts of high quality nucleic acids must be obtained for subsequent library production and sequencing.**
- DNA extraction based on sample type
- Invertebrate or plant — Fractionation or selective lysis
- Soil projects — Physical separation & isolation of cells
- Biopsies or ground water — Multiple displacement amplification

5.Sequencing Technology

There are two basic types of Metagenomics studies

1. Sequence-based Metagenomics
Involves sequencing and analysis of DNA from environmental samples.
2. Function-based Metagenomics
It involves screening for a specific function or activity.
1.Sequenced based metagenomics-

- Ribotyping
- Whole genome sequencing
- Pyrosequencing
- 16S rRNA Gene Sequencing
- Shortgun sequencing

6. Assembly

It is the comparison of each sequence read to every other and put them in proper order based on their overlap resulting in collection of correctly ordered big genome stretch. The reads are assembled into contigs, and finally to the whole genome.

Software's used

1.Newbler 4.Celera assembler
2.AMOS 5.Phrap
3.MIRA 6.Atlas
Works well if the metagenomic dataset contains sequences where closely related reference genomes are available

7. Binning

- Sorting of DNA sequences into groups that represent an individual genome or genome from closely related organisms.
- Binning is the process of grouping reads or contigs into individual genomes and assigning the group to specific species, subspecies or genus.
- Binning methods can be characterized in two different ways depending on information contained within a given DNA sequence.

1. Composition based binning 2. Similarity based binning (homology)

8. Annotation

- Annotation is the process of assigning functional, positional, and species of- origin information to the genes in a database. Involves identification of protein-coding regions, rRNA and tRNA genes.
- Annotation of metagenome is specifically designed to work with mixtures of genomes and contig of varying length.
- Identification of genes within the reads/ assembly contig, a process often denoted as "gene calling"
- Annotation pipeline involves functional assignment to the predicted protein coding genes. This is currently achieved by homology based searches of query sequences against databases containing known functional and/or taxonomic information.
- Both IMG/MER and MGRAST are widely used data management repositories and comparative genomics environments. They are fully automated pipelines that provide quality control, gene prediction, and functional annotation.

Application of Metagenomics

1. Food applications

Papaya proteases (papain) are widely used in food applications as meat tenderizers, in brewing, baking, and production of protein hydrolysates. Chymosin (rennin), an acidic protease obtained from stomachs of calves, is required for the manufacture of cheese (Wong, 1995). Proteases are also used by the food industry, particularly in the preparation of protein hydrolysates from casein, soy, fish, and whey proteins to produce formulated or texturized food products.

2. Novel Proteases from Metagenomics

- A *Bacillus* species was isolated from screening soil metagenomes using a medium containing leather powder (from grinding chrome-tanned cow leather) to produce a protease that hydrolysed not only casein, gelatin but leather powder at alkaline pH (Ogino *et al., 2008).*
- A novel metalloprotease gene was identified from a metagenomic library constructed using DNA extracted from deep-sea sediments (Lee *et al., 2007). The purified enzyme* hydrolysed azocasein as well as fibrin. Fibrinolytic enzymes have been used as thrombolytic drugs. These are mostly serine proteases that work by converting plasminogen to the natural fibrinolytic agent plasmin, which functions to dissolve clot by breaking down the fibrinogen and fibrin contained in a clot.
- Sana *et al.* (2006) reported the isolation, purification and characterization of a novel protease with versatile properties from deep-sea sediment. The enzyme was stable in the presence of high salt, oxidizing/reducing agents, commercial detergents and bleaches, and capable of complete removal of blood and egg stains, and acting on a wide range of substrates.

3. Antibiotic Resistance Genes

- The emergence of antibiotic-resistant bacteria in recent years underscores the urgent need for better understanding of microbial ecology of resistance genes in the environment, because they may migrate to clinical settings and become a threat to human health.
- A functional metagenomic analysis of a remote Alaska soil sample has identified 14 clones conferring resistance on the *E. coli host.* The clones harboured genes encoding all four classes of -lactamases (active site serine -1actamases classes A, C, D, and metallo--lactamases class B) (Allen *et al., 2008).* The discovery indicated that the soil is abundant in -lactarnases that can hydrolyse -lactam type antibiotics, such as penicillin and cephalosporin.
- Riesenfeld *et al. (2004) have also identified tetracycline* and aminoglycoside antibiotic resistance genes from cloned soil DNA.

4. Antibiotics and Secondary Metabolites

- Plant and microbial natural products have been the main source of antibiotics, therapeutic drugs and other pharmaceutical compounds. Some of the potent anticancer drugs, such as bleomycin, doxorubicin, and actinomycin, are isolated from *Streptomyces and Actinomyces species (Petit,* 2004).
- Two bioactive metabolites, indirubin and indigo were found produced by *E. coli clones from soil metagenomic libraries* (MacNeil *etal., 2001; Lim et al., 2005). Indirubin* a compound with antimicrobial activity and indigo is used as a textile dye.
- The antibiotics turbomycin A and B have been isolated from a soil metagenomic library by screening active clones conferred with dark brown colour (Gillespie *et al., 2002).*

5. Industrial Enzymes

There is ample demand for novel enzymes and biocatalysts, and metagenomics nowadays thought to be one of the most likely technologies to provide the required candidate molecules. Cellulases, lipases, xylanases, amylases, proteases, and various other industrially important enzymes have been mass-producing through metagenomics.

CONCLUSION

Metagenomics allows us to discover new genes and proteins or even the complete genomes of non-cultivable organisms in less time and with better accuracy than classical microbiology or molecular methods. It also helps to solve practical challenges in the field of medicine, agriculture, sustainability, and ecology.

REFERENCES

Breitbart, M et. al. (2002). "Genomic analysis of uncultured marine viral communities". *Proceedings of the National cademy of Sciences of the United States of America.* 99 (22): 14250–14255.

Gillespie et. al. (2002). "Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA". Appl Environ Microbiol 68:1

Handelsman J et. al. (1998). "Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products". *Chemistry & Biology.* 5 (10): R245–R249.

Lee et. al. (2007). "Screening and characterization of a novel fibrinolytic metalloprotease from a metagenomic library". Biotechnol Lett.29:465–472.

MacNeil et. al. (2001). "Expression and isolation of antimicrobial small molecules from soil DNA libraries". J Mol Microbiol Biotechnol 3:301

Pace. NR et. al. (1991). "Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing". *Journal of Bacteriology.* 173 (14): 4371– 4378.

Venter, JC et. al. (2004) "Environmental Genome Shotgun Sequencing of the Sargasso Sea". *Science.* 304 (5667): 66–74.