

Metagenomic Approach towards Bioprospection of Novel Biomolecule(s) and Environmental Bioremediation

**Manikant Tripathi^{1†}, Durgesh Narain Singh^{2†}, Surendra Vikram³,
Vijay Shankar Singh² and Shailendra Kumar^{1*}**

¹*Department of Microbiology, Dr. Ram Manohar Lohia Avadh University, Faizabad, (U.P.),
Pin: 224001, India.*

²*Laboratory of Synthetic Biology, CSIR-Central Institute of Medicinal and Aromatic Plants, Lucknow,
(U.P.), Pin: 226015, India.*

³*Centre for Microbial Ecology and Genomics, Department of Genetics, University of Pretoria, Pretoria,
Pin: 0028, South Africa.*

Authors' contributions

This work was carried out in collaboration between all authors. Authors MT and DNS developed the concept and wrote the first draft of manuscript. Author VSS created figures. Author SV explored literature and edited manuscript. Author SK proof read manuscript and approved it.

Article Information

DOI: 10.9734/ARRB/2018/38385

Editor(s):

(1) George Perry, Dean and Professor of Biology, University of Texas at San Antonio, USA.

Reviewers:

(1) Hanan E.-S. Ali, Egyptian Petroleum Research Institute, Egypt.

(2) Oscar M. Mosquera, Universidad Tecnologica de Pereira, Colombia.

(3) Osazuwa Omoregbee, University of Benin, Nigeria.

Complete Peer review History: <http://www.sciencedomain.org/review-history/22652>

Review Article

Received 24th November 2017
Accepted 28th December 2017
Published 9th January 2018

ABSTRACT

Microorganisms have developed several physiological adaptations to survive within extreme ecological niches including environments contaminated with heavy metals, pesticides, polycyclic aromatic hydrocarbons, and nuclear wastes. Microorganisms in extreme habitat are potential source of "novel biomolecule(s)" such as whole microbial cells, extremozymes and extremolytes, significantly required for environmental, industrial, and red medical/pharmaceutical biotechnology. These novel biomolecule(s) are valuable resources and may help improve economic development. The scanty information about the factors governing the microbial growth within stressed

*Corresponding author: E-mail: shailendra.microbio@gmail.com;

E-mail: manikant.microbio@gmail.com;

† Contributed Equally

environments is the major constraint in the recovery of novel biomolecule(s) from extreme habitats. Understanding the structure, metabolic capabilities, microbial physiology, and factors governing the composition and role of indigenous microorganism is the key to success of any study. In recent past the problems associated with classical cultivation techniques have been resolved by an emerging approach referred to as “metagenomics”. Metagenomic studies give an insight into details of the structure, metabolic and physiological capabilities of indigenous microbial communities. High-throughput sequencing technologies in conjunction with metagenomics has aided in the identification and characterization of novel culturable and uncultured microorganisms with unique capabilities. Metagenomic studies have been used for isolation and characterization of novel biomolecule(s) relevant for white, grey, and red biotechnologies. The major objective of this review is to discuss the applications of metagenomic approach for bioprospection of novel biomolecule(s) and environmental bioremediation.

Keywords: Environment; extremozymes; metagenomics; microorganism; novel biomolecule(s).

1. INTRODUCTION

The best indicator for the existence of uncultured microorganisms is “great plate count anomaly”, which is defined as the variance between the estimated populations of microbial communities determined using serial dilution method and microscopic observations [1]. Since more than 99% of microorganisms in the environment are unculturable, the microbiologists focused mainly on culturable microorganisms for the discovery of novel biomolecule(s). This reflects that a substantial portion of microbial community has remained unexplored due to our inability to grow them in artificial culture media. Lane et al. [2] described a new method based on 16S rRNA gene sequence analysis for phylogenetic study of microbial community. A new method based on direct analysis of universal markers of taxonomy including 5S and 16S rRNA gene has been described for identification of microbial communities within an environment without culturing the microorganism [3]. Moreover, it's not always easy to isolate and culture some of the microorganisms (unculturable) due to their unknown physiological requirements. In addition to evolutionary studies, sequencing of 16S rRNA gene provides information regarding physiological requirements for culturing the unculturable microorganisms. Recent advancement in metagenomics has revolutionized the studies in microbiology. Metagenomic approach, whereby the whole DNA is directly extracted from sample and sequenced using the next generation sequencing platforms (Illumina, SOLiD, PacBio, Oxford NanoPore) can be used to circumvent the problem associated with the isolation of unculturable microorganisms [4]. The next generation sequencing technologies

produces large amount of data with high accuracy, seamlessly suitable for the shotgun metagenomic sequencing projects. The longest reads produced by Illumina platform are up to 250 x 2 (paired end) with the accuracy more than 99.9% [5]. Conversely, relatively new and developing sequencing technologies produce much longer reads such as PacBio (read length 10-15Kb) and Oxford Nanopore (read length up to 900 Kb) with high error rates up to 15% [5,6].

The next generation sequencing and application of bioinformatics enabled researchers to reveal untapped microbial diversity (uncultivable) using environmental DNA (Fig. 1). The insight of metabolic pathways of indigenous microbial community derived from analysis of metagenome can further be applied to design the special media and conditions for enrichment and isolation of novel uncultured microorganisms possessing unique capabilities [7,8]. In addition to novel biocatalysts, novel microorganisms can be potential sources of protective organic biomolecule(s) (extremolytes) and other bioproducts (exopolysaccharide, biosurfactant, peptide and biopolymers) that enable them to survive under harsh environmental conditions. Metagenomics can reveal information about the identity and physiological requirements by applying the sequence based strategy i.e. “gene centric” or “genome centric” analysis of metagenomics DNA [9]. The gene centric approach is to study the metagenomics data for the presence of different taxa and metabolic and functional genes in the environmental DNA, whereas genome centric approach is to reconstruct the complete or near complete genomes from the environmental sequence data [9].

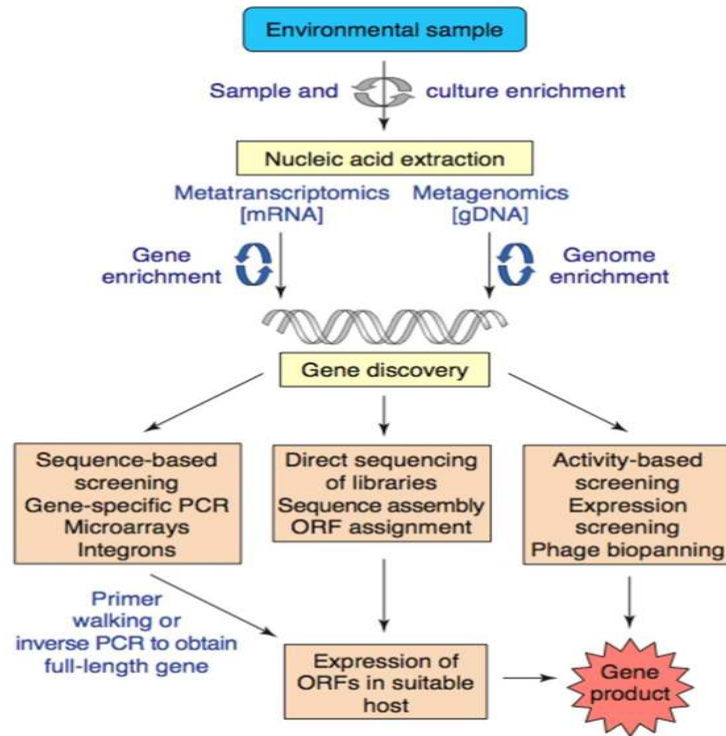


Fig. 1. Schematics presentation of essential steps to explore and exploit the genomic diversity of soil microbial communities by metagenomics approach; nucleic acid is extracted from the environmental sample (metagenome) or meta transcriptomes (environmental mRNA). Downstream screening approaches can be activity-based through the screening of expression libraries, sequence-dependent or sequence-independent. The final expression requires a full-length open reading frame (ORF) expressed in a suitable host to generate a functional gene product (Adapted from Cowan et al. [10])

Moreover, metagenomics has been used for acceleration of the process of *in situ* bioremediation. In the process of bioremediation, the decontamination rate is accelerated either by addition of whole cells (bioaugmentation) or rate-limiting nutrients such as electron acceptors (biostimulation) [11,12]. Novel uncultured microorganisms (biomolecule) isolated from contaminated sites can be used for bioaugmentation. On the other hand, physiological as well as nutritional requirements of indigenous microorganisms derived from metagenomic analysis of contaminated sites can be used for biostimulation of indigenous microbial community.

The development of novel biomolecule(s) is the primary requirement for the economic growth of white (industrial) biotechnological industries in a sustainable manner. In addition metagenomics can be successfully applied to solve the practical challenges in the field of medicinal and

agricultural sector. Besides the application of metagenomics for tapping hidden microbial diversity (discovery of novel biocatalysts from uncultivable microorganisms), it also has wide applications in the field of environmental microbiology such as bioremediation and metabolism of xenobiotics.

2. METAGENOMICS- BIOPROSPECTION OF NOVEL BIOMOLECULE(S)

The growth of white (industrial) biotechnology requires the development of novel enzymes and different other products which are relevant to mankind. Application of metagenomics promises the production of new molecules with diverse functions. With the advent of metagenomics we can target microorganisms which are recalcitrant to isolation on common laboratory medium due to the lack of nutrients or inappropriate incubation temperature and/or pressure and composition of atmospheric gases [13].

Phylogenetic studies based on molecular tools have confirmed that representation of cultured bacteria is only a miniature (< 1%) of bacterial diversity, as culturing conditions do not imitate the natural habitat conditions [14,15]. The great majority of microorganisms that cannot be grown in laboratory are known as ‘unculturables’ but are now often referred to as ‘yet to be cultured’ or ‘yet uncultured’ since it is likely that when provided right conditions we would be able to culture these organisms. Consequently, the conventional methods of cultivation are insufficient for deciphering the substantial reservoir of unseen natural diversity of ‘yet to be cultured’ microorganisms. To overcome the natural loss of diversity due to unknown culture conditions, the collective genomes of all the microorganisms (metagenome) present in the particular habitat, could be used in the metagenomic approach (cultivation independent approach). Metagenomic analyses of universal marker of taxonomy such as 16S rRNA gene would help to identify the diversity, metabolic capabilities and physiological requirement of indigenous microbial communities of a given habitat [13,16]. Bacterial 16S rRNA genes are comprised of nine hyper-variable regions (V1-V9) which reflects significant sequence variation among different bacteria. Initial phylogenetic

studies included Sanger-sequencing strategies where complete 1.5 kb fragments of 16S rRNA gene were targeted for sequencing and identification. Nowadays, the hypervariable regions (V1-V9) that are interspersed within conserved regions (C1-C9) are targeted for sequencing and identification (Fig. 2).

Sequencing of hypervariable regions enabled researchers to identify many individual genera that could not have been identified with traditional biochemical methods. Moreover, structural and physiological information of indigenous microbial communities derived by metagenomic approach is used for designing or selection of special media and conditions for enrichment and isolation of novel microorganisms which are “yet to be cultured”. These un-cultivable microorganisms are untapped source of different novel biomolecule(s) including biocatalysts, antibiotics, extremolytes, exopolymeric substances (EPS), and other bio-products that enables them to survive under harsh environments (Fig. 3). Microbial extremolytes are primarily used in cosmetic and medical sectors, and some examples include UV radiation-protective compounds such as melanin, ectoines and scytonemin [17,18].

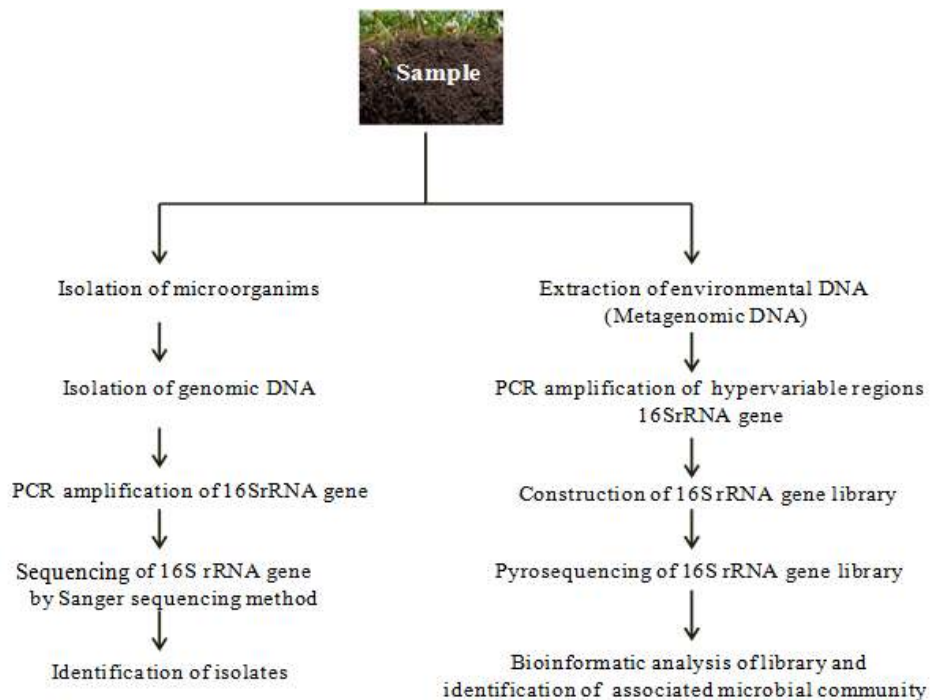


Fig. 2. Flow chart showing 16S rRNA gene based method for identification of microorganism associated with sample

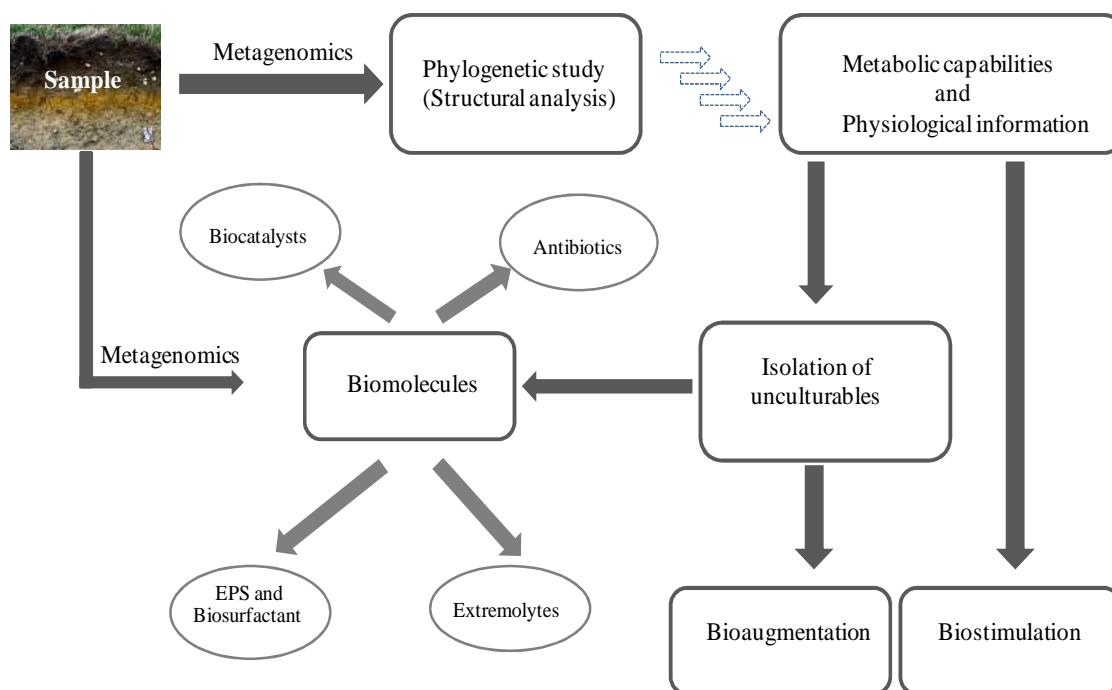


Fig. 3. Schematic illustration representing applications of metagenomic approach for bioprospection of novel biomolecule(s) and acceleration of environmental bioremediation

Metagenomic analysis has emerged as a powerful technique for exploiting different biomolecules from the hidden diversity present in our environment [13]. Metagenomic approach involves extraction of DNA from the environmental sample, cloning DNA in a suitable vector (cosmid, fosmid or BAC vectors), transformation of recombinant vector into a suitable host bacterium, and eventually screening of the clones for a particular function. The metagenomic clones are screened following “function based” or “sequence based” approaches. In “function based” approach recombinant clones are screened for the desired activity on different media containing a suitable substrate [13,16,19]. However function based metagenomics has its limitations such as incompatibility of gene expression between host and cloned DNA. On the other hand, sequence based screening has been widely accepted which involve the generation of metagenomics DNA library, sequencing on the next generation sequencing platform, assembly of the raw data into longer contigs and identification of the protein sequences by aligning database against reference databases such as NCBI-NR or NCBI-Refseq. The taxonomic and functional annotation of the metagenomics data can be further taken into consideration for the next step of gene

synthesis, codon optimization according to the destination expression in host, and biochemical characterization of identified enzymes. However, it is potentially possible that limited number of diverse sequences in the database and large number of novel sequences in the environmental sample is problematic for the annotations of metagenomics sequences. Moreover, sequence based screening methods often end with partial genes, which is a major drawback because it does not allow inferring the functionality of encoded enzymes [20].

With the onset of next generation sequencing (NGS) and bioinformatic tools, sequence based screening has facilitated the identification, isolation and characterization of several novel enzymes of industrial importance, and phylogenetic studies [21]. The gene encoding polyketide synthase involved in polyketide synthesis was cloned from soil metagenome using sequence based approach [13]. Moreover, shotgun sequencing is not cost effective with the target to discover genes encoding specific enzymes. In contrast, function based screening methods involve the selection of clones expressing desired traits, and further characterization of genes and encoded products [21]. The screening of metagenomic libraries on

the basis of function has enabled identification and isolation of several genes encoding antibiotics, genes conferring resistance to antibiotics, Na⁺(Li⁺)/H⁺ transporters, and various degradative enzymes [22]. Genes encoding novel lipases [23], proteases [24], amylases [25], melanin, antibiotics [13,26], and several other biomolecule(s) have been identified and characterized. Using functional screening method a novel phosphodiesterase enzyme has been identified and characterized from an Indian coal bed [27]. However, major limitations for the functional screening method are the requirement for the presence of a cluster of all genes required for particular function and their expression within a heterologous host. An appropriate assay is

also required for the detection of the function [28]. To resolve the problems related to screening, different alternate strategies have been explored. In addition to *E. coli*, heterologous expression of secondary metabolites has been optimized in *Streptomyces lividans* or *Pseudomonas putida* [29,30]. Using standard inhibition assay, terragine, an antibiotic against *Mycobacterium* has been discovered from a metagenomic clone expressed in *Streptomyces lividans* [31]. Databases and bioinformatic tools used in metagenomic studies are listed below (Table 1). Some of the enzymes discovered through metagenomic approach are listed in Table 2.

Table 1. List of important bioinformatic tools and databases used in next generation sequencing based metagenomics analysis

Database/Software	Application	References (Link websites)
16S rRNA and marker genes based tools		
Qiime	Phylogenetic and diversity analysis based on Greengenes and SILVA database	http://qiime.org/
Mothur	From reads quality analysis to taxonomic classification, calculus of diversity estimators and comparison of ribosomal gene metaprofiling	http://www.mothur.org/
ARB	Interrelated tools for sequence databases(RNA/DNA/amino acids)	http://www.arb-home.de
SILVA	i) Ribosomal RNA database of bacteria, archaea, eukaryotes. ii) 'Parc' datasets of rRNA gene sequences	http://www.arb-silva.de
Greengenes	16S rRNA database of bacteria and archaea	http://greengenes.secondgenome.com/
Phyloseq	R Package for the analysis of OTUs data obtained from the QIIME, Mothur, etc. Statistical analysis of diversity and graphics production.	https://joey711.github.io/phyloseq/
FunGene	Functional gene pipeline/ repository analysis pipeline	http://fungene.cme.msu.edu/
PICRUSt	Prediction of metabolic capabilities from taxonomic information obtained from 16S rRNA metaprofiling projects	http://picrust.github.io/picrust/
MG-RAST	Taxonomic and functional annotation of metagenomic short reads and assembled contigs comparative metagenomics	http://metagenomics.anl.gov/
GenBank	i) Redundant archival database ii) Provides alternate views of proteins, names and extra information	http://www.ncbi.nlm.nih.gov/Genbank
RefSeq	Non-redundant collection of well-annotated sequences from genomic data, transcripts and proteins	http://www.ncbi.nlm.nih.gov/RefSeq

Metagenome and genome analysis tools		
IMG/M	Functional annotation, phylogenetic distribution of genes and comparative metagenomics	https://img.jgi.doe.gov/cgi-bin/m/main.cgi
GOLD	Metadata information for worldwide projects related to genomes and metagenomes	http://www.genomesonline.org
MEGAN	Geographical and statistical analysis of several data sets	http://www-ab.informatik.uni-tuebingen.de/software/megan
Megx.net	Database tools used for the analysis of marine metagenome	http://www.megx.net
RAST	Phylogenetic and functional study of genome	https://rast.nmpdr.org/
Database/Software	Application	References (Link websites)
NR-protein	Non-redundant protein sequences from GenPept, Swissprot, PIR, PDF, PDB, and NCBI RefSeq Non redundant collection of protein sequences	ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz
XplorSeq	Provides large compilation phylogenetic analysis of DNA sequences	http://vent.colorado.edu/phloware
Parallel-meta	Taxonomic annotation of ribosomal gene markers sequences obtained by metaprofiling or metagenomic data. Functional annotation based on BLAST. Comparative metagenomics	http://www.computationalbioenergy.org/parallel-meta.html
Genometa	Taxonomic and functional annotation of short-reads metagenomic data. Graphical interface	http://genomics1.mh-hannover.de/genometa/
GlimmerMG	Predictor of gene coding sequences from metagenomic sequences by unsupervised clustering	http://www.cbcb.umd.edu/software/glimmer-mg/
MetaPhlan	Tool to analyze the microbial composition based on the unique clade specific marker genes	https://bitbucket.org/nsegata/metaphlan/wiki/Home
PhyloPythiaS+	Support vector machine based taxonomic annotation of metagenomic contigs (>1kb)	https://github.com/algbioi/ppsp
KRAKEN	Taxonomic classification metagenomic sequences based on custom databases (bacteria, archaea, fungi, viruses etc.)	http://ccb.jhu.edu/software/kraken
PhymBL	Taxonomic classification of metagenomic reads	https://www.cbcb.umd.edu/software/phymm/
Naïve Bayesian Classification (NBC) tool	Taxonomic Classification of metagenomics reads.	http://nbc.ece.drexel.edu/
Statistical tools for analysis of sequence data		
Statistical analysis of taxonomic and functional profiles	Tool to analyze the metagenomic data obtained from the MEGAN, MG-RAST, etc.	https://github.com/dparks1134/S-TAMP
UniFrac	Comparative study of microbial communities using phylogenetic information	bmf2.colorado.edu/unifrac/index.psp

Table 2. List of enzymes discovered through metagenomic approach

Enzymes	Substrate	References
Esterase/Lipase	Tributyryn	Henne et al. [32]
Esterase/Lipase	Triolein	Henne et al. [32]
Esterase/Lipase	Bacto lipid	Rondon et al. [33]
Alcohol oxidoreductase	Glycerol/1,2-propanediol	Knietsch et al. [34]
Amidase	D-phenylglycine-L-leucine	Gabor et al. [35]
Amylase	Starch	Gabor [36]
Amylase	Starch	Rondon et al. [33]
Cellulase	Carboxy methyl cellulose	Rees et al. [37]
Chitinase	Methyl umbelliferyl diacetyl chitobioside	Cottrell et al. [38]
Dehydratase	Glycerol	Knietsch et al. [39]
β -Lactamase	Ampicillin	Gabor [36]

3. METAGENOMICS- ECOLOGICAL INFERENCE

Certain groups of microorganisms make symbiotic relationship with eukaryotes, and compete for nutrients to produce energy [13]. Several medicinally important natural products have been obtained from invertebrate animals, e.g., sponges. Some sponges host large numbers of bacteria (40–60% of total biomass) within their tissues. Many bacterial symbionts do not readily grow on culture medium because of highly specialized and ancient relationship with their hosts [40]. Using metagenomic analysis of 16S rRNA gene sequences, different microorganisms have been identified from various environments such as sponges, acid mine drainage, extreme desert environments, etc. [22]. Such studies have been conducted with *Cenarchaeum symbiosum*: marine sponge symbiosis [41], *Pseudomonas*-like bacterium: *Paederus* beetles symbiosis [42], *Buchnera aphidicola*: aphids [43], and *Proteobacterium: Riftia pachyptila* symbiosis [44]. *Riftia pachyptila* is a tube worm, dwelling 2600 m under the sea surface, close to the thermal vents rich in sulfide and having temperature about 400 °C. Tube worms lack mouth and digestive tract and thus, completely rely on symbiotic bacterial partner for food and energy. The worms transport hydrogen sulfide and CO₂ from environment to the bacterial symbiont. The bacterial partner residing inside the trophosome of worm oxidize hydrogen sulfide for CO₂ fixation, and provide food and energy to worm. The bacterium has not been isolated in pure culture on synthetic medium, while, following metagenomic analysis the bacterium was identified as γ -Proteobacteria. With the objective to understand the physiology of bacterial symbiont a metagenomic DNA library was constructed in fosmid vector [45]. Robinson et al. [46], identified a gene encoding RubisCO

enzyme following screening of fosmid library, which was similar to the RubisCO of *Rhodospirillum rubrum*. Moreover, metagenomic analysis revealed the presence of gene which encode for flagellin, and this gene was heterologously over-expressed within *E. coli* and was involved in synthesis of bacterial flagella. This suggested that endosymbiont may be free-living and infects every generation of the worm instead of being moved through maternal transfer [47]. Certain groups of microorganisms exert competition for resources and survival. Sequence based screening for genes conferring resistance is difficult, however functional genomics can be helpful in identification of genes encoding biologically active compounds that confer resistance to their producer organisms [13]. Intriguingly, metagenomics also provides insight into paleogenomic, and metagenomic approach coupled to high-throughput sequencing is an efficient tool to study the nuclear genome of extinct microorganisms. Phylogenetic study has been performed by comparative analysis of 27 kb metagenomic DNA sequence of *Ursus spelaeus* (cave bear) and PCR amplified orthologous sequence from black brown and polar bears [22].

4. METAGENOMICS- ENVIRONMENTAL BIOREMEDIATION

Environmental pollution is a serious global concern. All biological processes involved in removal or transformation of pollutants from the environment are referred to as bioremediation. In recent years, microorganisms mainly bacteria are used for bioremediation of pollutants at different sites due to their ability to resist, transform or degrade the pollutants. Microorganisms adapt to contaminated environment with the help of catabolic enzymes and other biomolecule(s) produced by them. Exopolymeric substances (EPS) and biosurfactants are involved in

bioremediation of aromatic hydrocarbon and toxic heavy metals; EPS is a combination of polysaccharides, muco-polysaccharides and proteins. The composition of EPS mainly depends on bacterial strains and the culture conditions. Ionizable functional group of EPS enables it to bind heavy metal ions, and protects bacterial cells from heavy metal toxicity [48]. Biosurfactants have their role in petroleum industry for transport of crude oil, and enhanced oil recovery by effectively reducing interfacial tensions of oil and water [49]. Using metagenomic approach, a strain of *Pseudomonas stutzeri* was isolated from Indian coal bed which solubilized coal and produced copious amount of biosurfactant when coal was added to medium [50]. The process of bioremediation can be conducted either in-situ or ex-situ. In-situ bioremediation utilizes microbial processes that degrade or transform contaminants on site. It is further subdivided in to intrinsic bioremediation and in situ engineered bioremediation. The former utilizes the indigenous microorganisms present at site, whereas engineered bioremediation utilizes different process such as availability of water, aeration, nutrients amendments etc.

In *ex-situ* bioremediation processes contaminated sample is moved to another location before treatment. In the process of bioremediation, the decontamination rate is accelerated either by addition of whole cells (bioaugmentation) or rate-limiting nutrients/electron acceptors (biostimulation) [11, 12]. The major constrain in bioremediation is the requirement of information about parameters which govern the growth and metabolism of indigenous microorganisms residing within contaminated environments [51]. A better understanding of microbial adaptations to environment is key to success of any bioremediation study [52]. With conventional cultivation techniques it was difficult to ascertain the role of isolated microorganism for in situ bioremediation. The 16S rRNA gene sequence based cultivation-independent analyses help to understand the structure, metabolic capabilities, physiology, and other factors governing the indigenous microbial community (Fig. 1). However, only 16S rRNA gene based study cannot reveal the biochemical and genetic basis of bioremediation.

Microarray technique viz., GeoChip [53] allows the fast and economical profiling of sample. Using GeoChip the presence of genes required

for anaerobic degradation of alkane, toluene and ethyl benzene has been detected in the formation water of San Juan Basin [54]. Unlike, 16S rRNA gene sequence based studies; the disadvantage of microarray approach is that it can be used only to analyze the targets which are identical to already known gene sequences. Sampling based studies are, thus, troubled by regular re-isolation of abundant sequences at the cost of identifying atypical species. Therefore, metagenomic approaches are typically employed with a view to identify functional genes of metabolic pathways [55]. Metagenomic studies provide all the essential information for *in situ* bioremediation, and are centered on reconstruction of the genome for “yet uncultured” microorganisms. The newly reconstructed genomes of “yet uncultured” microorganisms are screened for novel biomolecule(s) involved in the process of bioremediation. A microcosm study employing the enrichment technique (using mercury, ethanol and diesel) followed by next generation sequencing and genome reconstruction produces the genome bins of selectively enriched bacteria. Another genome resolved metagenomic study analyzing the microbial communities from the thiocyanate degrading bioreactors (solid particulate tailing and solid free reactor) revealed the differences in the bacterial strains and their metabolic capacities. Metagenomics based phylogenetic studies provide metabolic and physiological information required for the isolation of unculturable microorganisms. This information is used for the isolation of unculturable or acceleration of bioremediation by bioaugmentation and biostimulation processes. Nutritional and physiological information derived from cultivation independent analysis of microbial community associated with Indian coal bed has been utilized for accelerating the rate of coal biotransformation into methane gas [56]. The advantage of bioaugmentation and biostimulation approach is due to its economical and eco-friendly nature. However, some factors which can limit the process of bioaugmentation and biostimulation are strain selection, strain adaptation and survival, differential environmental constrains, and process of introducing microbial culture and nutrient supplements to contaminated sites [55].

5. CONCLUSIONS

Metagenomics is a promising field that has enabled microbiologists to access hidden microbial resources relevant for different

biotechnological sectors. Comprehensive study of composition and function of indigenous microbial community enables to find novel biomolecule(s). Earlier the microbiological processes for the production of biomolecules were labor and cost intensive while following the start of metagenomics it has become time saving and cost effective development of technology for the same. With the start of high-throughput sequencing technology and functional screening of metagenomics libraries facilitate the detection of novel biocatalysts and other bioactive molecules which are indispensable for mankind. The large data set derived from metagenomic studies are used to construct database which can be exploited to discover potentially undiscovered biomolecule(s). Future alteration of methods that enrich the genes encoding any meticulous function would help out in discovery of novel biomolecule(s). In addition, metagenomics has potential application in the field of environmental bioremediation. Metagenomics based metabolic and physiological assessment of indigenous microbial community support to design technologies for efficient *in situ* bioremediation.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Staley JT, Konopka A. Measurement of *in situ* activities of non photosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol.* 1985;39: 321–46.
2. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci USA.* 1985;82:6955–59.
3. Pace NR, Stahl DA, Lane DJ, Olsen GJ. The analysis of natural microbial populations by ribosomal RNA sequences. *Adv Microb Ecol.* 1986;9:1–5.
4. Maphosa F, Van Passel MWJ, De Vos WM, Smidt H. Metagenome analysis reveals yet unexplored reductive dechlorinating potential of *Dehalobacter* sp. E1 growing in coculture with *Sedimentibacter* sp. *Environ Microbiol Rep*; 2012. DOI: 10.1111/j.1758- 2229.2012.00376.x
5. Glenn TC. Field guide to next-generation DNA sequencers. *Mol Ecol Resour.* 2011; 11:759–69.
6. Jain M, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *bioRxiv.* DOI: <https://doi.org/10.1101/128835>
7. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chem Biol.* 1998;5:245–49.
8. Lorenz P, Liebeton K, Niehaus F, Eck J. Screening for novel enzymes for biocatalytic processes: Accessing the metagenome as a resource of novel functional sequence space. *Curr Opin Biotechnol.* 2002;13:572–77.
9. Prosser JI. Dispersing misconceptions and identifying opportunities for the use of ‘omics’ in soil microbial ecology. *Nature Rev Microbiol.* 2015;13:439-46.
10. Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R, Wittwer P. Metagenomic gene discovery: Past, present and future. *Trends Biotechnol.* 2005;23:321-29.
11. Atlas RM. Microbial-degradation of petroleum-hydrocarbons an environmental perspective. *Microbiol Rev.* 1981;45:180–09.
12. Nikolopoulou M, Kalogerakis N. Biostimulation strategies for fresh and chronically polluted marine environments with petroleum hydrocarbons. *J Chem Technol Biotechnol.* 2009;84:802–07.
13. Handelsman J. Metagenomics. Application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev.* 2004;68:669–85.
14. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev.* 1995;59: 143–69.
15. Zengler K, Toledo G, Rappe M, Elkins J, Mathur EJ, Short JM, Keller M. Cultivating the uncultured. *Proc Natl Acad Sci.* 2002; 99:15681–86.
16. Streit WR, Schmitz RA. Metagenomics: The key to the uncultured microbes. *Curr Opin Microbiol.* 2004;7:492–98.
17. Singh SP, Klisch M, Sinha RP, Hader DP. Genome mining of mycosporine-like amino acid (MAA) synthesizing and non synthesizing cyanobacteria, a bioinformatics study. *Genomics.* 2010;95: 120–28.

18. Singh OV, Gabani P. Extremophiles: Radiation resistance microbial reserves and therapeutic implications. *J Appl Microbiol.* 2011;110:851–61.
19. Jeon JH, Kim JT, Kim YJ, Kim HK, Lee HS, Kang SG, Kim SJ, Lee JH. Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome. *Appl Microbiol Biotechnol.* 2009;81:865-74.
20. Telling H, Glockner FO. Current opportunities and challenges in microbial metagenome analysis-a bioinformatic perspective. *Brief Bioinform.* 2012;13(6): 728-42.
21. Yun J, Ryu S. Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. *Microbial Cell.* 2005;4:8.
DOI: 10.1186/1475-2859-4-8
22. Singh J, BehalA, Singla N, Joshi A, BirbianN, Singh S, Bali V, Batra N. Metagenomics: Concept, methodology, ecological inference and recent advances. *Biotechnol J.* 2009;4:480–94.
23. Lee SW, Won K, Lim HK, Kim JC, Choi GJ, Cho KY. Screening for novel lipolytic enzymes from uncultured soil microorganisms. *Appl Microbiol Biotechnol.* 2004;65:720–726.
24. Gupta R, Beg QK, Lorenz P. Bacterial alkaline proteases: Molecular approaches and industrial applications. *Appl Microbiol Biotechnol.* 2002;59:15–2.
25. Yun J, Kang S, Park S, Yoon H, Kim MJ, Heu S, Ryu S. Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Appl Environ Microbiol.* 2004;70:7229–35.
26. Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM, Handelsman J. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl Environ Microbiol.* 2002;68:4301–06.
27. Singh DN, Gupta A, Singh VS, Mishra R, Kateriya S, Tripathi AK. Identification and characterization of a novel phosphodiesterase from the metagenome of an Indian coal bed. *PLOS ONE* 2015;10(2):e0118075.
DOI: 10.1371/journal.pone.0118075.
28. Schloss PD, Handelsman J. Biotechnological prospects from metagenomics. *Curr Opin Biotechnol.* 2003;14: 303–10.
29. Henne A, Schmitz RA, Bomeke M, Gottschalk G, Daniel R. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl Environ Microbiol.* 2000;66:3113-16.
30. Courtois S, Cappellano CM, Ball M, Francou FX, Normand P, Helynck G, Martinez A, Kolvek SJ, Hopke J, Osburne MS, August PR, Nalin R, Guerineau M, Jeannin P, Simonet P, Pernodet JL. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol.* 2003;69:49-5.
31. Wang GY, Graziani E, Waters B, Pan W, Li X, McDermott J, Meurer G, Saxena G, Andersen RJ, Davies J. Novel natural products from soil DNA libraries in a streptomycete host. *Org Lett.* 2000;2: 2401–04.
32. Henne A, Schmitz RA, Bömeke M, Gottschalk G, and Daniel R. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl. Environ. Microbiol.* 2000;66:3113–3116.
33. Rondon MR, et al. Cloning the soil metagenome: A strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol.* 2000;66:2541–2547.
34. Knietsch A, Waschowitz T, Bowien S, Henne A, Daniel R. Construction and screening of metagenomic libraries derived from enrichment cultures: Generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Appl Environ Microbiol.* 2003;69:1408–1416.
35. Gabor EM, de Vries EJ, Janssen DB. Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environ Microbiol.* 2004;6:948–958.
36. Gabor EM. Harvesting novel biocatalysts from the metagenome. Thesis, Univ. Groningen; 2004.
37. Rees HC, Grant S, Jones B, Grant WD, Heaphy S. Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles.* 2003;7:415–421.
38. Cottrell MT, Moore JA, Kirchman DL. Chitinases from uncultured marine

- microorganisms. Appl. Environ. Microbiol. 1999;65:2553–2557.
39. Knietsch A, Bowien S, Whited G, Gottschalk G, Daniel R. Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. Appl Environ Microbiol. 2003;69:3048–3060.
 40. Hentschel U, Usher KM, Taylor MW. Marine sponges as microbial fermenters. FEMS Microbiol Ecol. 2006;55:167–77.
 41. Preston CM, Wu KY, Molinski TF, DeLong EF. A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. Proc Natl Acad Sci USA. 1996;93:6241–46.
 42. Piel J. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. Proc Natl Acad Sci USA. 2002;99:14002–07.
 43. Abbot P, Withgott JH, Moran NA. Genetic conflict and conditional altruism in social aphid colonies. Proc Natl Acad Sci USA. 2001;98:12068–71.
 44. Hughes DS, Felbeck H, Stein JL. A histidine protein kinase homolog from the endosymbiont of the hydrothermal vent tubeworm *Riftia pachyptila*. Appl Environ Microbiol. 1997;63:3494–98.
 45. Robinson JJ, Stein JL, Cavanaugh CM. Cloning and sequencing of a form II ribulose-1,5-biphosphate carboxylase / oxygenase from the bacterial symbiont of the hydrothermal vent tubeworm *Riftia pachyptila*. J Bacteriol. 1998;180:1596–99.
 46. Millikan DS, Felbeck H, Stein JL. Identification and characterization of a flagellin gene from the endosymbiont of the hydro thermal vent tubeworm *Riftia pachyptila*. Appl Environ Microbiol. 1999; 65:3129–33.
 47. Zhang D, Wang J, Pan X. Cadmium sorption by EPSs produced by anaerobic sludge under sulfate-reducing conditions. J Hazard Mater B. 2006;138:589–93.
 48. Singh A, Van Hamme JD, Ward OP. Surfactants in microbiology and biotechnology: Part 2 Application aspects. Biotechnol Adv. 2006;25:99–21.
 49. Singh DN, Tripathi AK. Coal induced production of a rhamnolipid biosurfactant by *Pseudomonas stutzeri*, isolated from the formation water of Jharia coal bed. Biores Technol. 2013;128:215–21.
 50. Rogers SL, McClure N. In Bioremediation: A critical review. (eds Head, I. M., Singleton, I. & Milner, M. G.), Horizon Scientific Press, Wymondham, UK. 2003; 27–9.
 51. Head IM, Larter SR, Gray ND, Sherry A, Adams JJ, et al. Hydrocarbon degradation in petroleum reservoirs. Handbook of Hydrocarbon and Lipid Microbiology. 2010; 3097–09.
 52. He Zi, Deng Y, Nostrand JDV, Tu Q, Xu M, Hemme CL, Li X, Wu L, Gentry TJ, Yin Y, Liebich J, Hazen TC, Zhou J. GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. ISME J. 2010;4:1167-79.
 53. Wawrik B, Callaghan AV, Davidova I, Liang Y, Martin K, et al. Field and laboratory studies on the bioconversion of coal to methane. Int Soc Microb Ecol (ISME), 13th, Seattle; 2010.
 54. Strapoc D, Mastalerz M, Dawson K, Maclady J, Callaghan AV, Wawrik B, Turich C, Ashby M. Biogeochemistry of microbial coal bed methane. Annu Rev Earth Planet Sci. 2011;39:617-56.
 55. Singh DN, Kumar A, Sarbhai MP, Tripathi AK. Cultivation independent analysis of archaeal and bacterial communities of the formation water in an Indian coal bed to enhance biotransformation of coal into methane. Appl Microbiol Biotechnol. 2012; 93:1337–50.
 56. Tyagi M, da Fonseca MM, de Carvalho CC. Bioaugmentation and biostimulation strategies to improve the effectiveness of bioremediation processes. Biodegrad. 2011;22:231-41.

© 2018 Tripathi et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
<http://sciencedomain.org/review-history/22652>