Metagenomic Approach to Explore Microbial Diversity and Genetic Potential of Uncultured Microorganisms from Different Environment Niches

Pushpender Kumar Sharma\textsuperscript{a,b,*}, Monika Sharma\textsuperscript{b}, Jagdeep Kaur\textsuperscript{a,*}

\textsuperscript{a}Indian Institute of Science Education and Research (IISER), Mohali, Knowledge City, Sector 81, S.A.S. Nagar, Mohali 140306, India
\textsuperscript{b}Department of Biotechnology, Panjab University, Chandigarh- University, Chandigarh-160014, India

\textbf{ABSTRACT}

The term metagenomics refers to genomic analysis of those microorganisms, which are difficult to cultivate in standard cultivation medium. The 16S rRNA study from various environments has provided a strong evidence for the existence of uncultured microorganism. The novel genes and gene products discovered by metagenomic approach include many hydrolytic enzymes, novel molecules and antimicrobial compounds. Furthermore, metagenomic studies from various extreme environments shed light into genomic diversity and existence of various co-operations among the microorganisms. This review discusses the concepts, basic tools and applications of metagenomic gene cloning in discovering novel molecules and microbial diversities. In addition to this, this review also highlights the impact of uncultured microorganisms on human health and environment.

\textbf{Keywords}: Biocatalysts, Antimicrobial, biogeochemical, metagenomics, microflora, phylogenetic, uncultivable.

*Corresponding author:
Email:pushpg_78@rediffmail.com
INTRODUCTION

Microorganisms (Bacteria) contribute significantly to the earth’s biological diversity. They are present everywhere in the environment that includes thermal ducts, great depths of the oceans, international space station, and in general are heterogeneously distributed throughout the aggregates of soil particles [1-3]. Current estimates indicate that less than 1% bacteria present in most habitats are culturable [4-6]. The 16S rRNA study from diverse sources of environment has provided strong evidence for the existence of new lineages of microbes [7,8]. Assigning function to uncultured microorganisms in various environments (in absence of pure culture) present immense challenges for microbial ecologists [9, 10]. A molecular technique allows evaluation of the structure, dynamics and metabolic potential of environment samples. The word ‘metagenomics’ was coined to capture the notion of analysis of uncultured microorganism [5]. Metagenomic library construction and screening constitute a valuable tool for making industrial biotechnology, economically a sustainable success [11, 12]. In addition to this, understanding genomic diversity in various environment niches, especially extreme environment and human GIT (gastrointestinal tract) will shed light into their functional role in such environment. Furthermore, microorganisms play an important role in maintaining ecosystem; therefore metagenomic studies will decipher the role of unculturable microorganism in such system [13, 14]. This review article discusses the basic tools involved in metagenomic studies and further highlights the role of uncultured microorganisms in extreme environment, human and in ecosystem.

1. Tools and techniques used in metagenomic study

1.1. 16S rRNA and microbial diversity

Identification and characterization of rRNA genes help in the analysis of phylogeny and quantification of microbial diversity [15-17]. The 16S rRNA gene in bacteria consists of conserved sequences along with interspersed variable and hypervariable regions. The length of hyper variable region ranges from ~50 -100 bases long and differ with respect to variation in their corresponding utility for universal microbial identification. Various communities have been studied using metagenomic approach, here we are discussing few of them i.e. microbial diversity of a thermal environment from Yellowstone’s was described by Pace et al. [18], Fierer et al. [19] have used small-subunit RNA sequence for highlighting the richness of new bacteria, archaea, fungi and viruses from soil sample. Additionally, Hallam et al. [20] have cloned a 40 Kb DNA fragment from metagenome sample of sea water, harbouring a 16S rRNA gene belonging to archaea. Moreover, a culture-independent study has revealed ~ 40 bacterial divisions, suggesting that almost 30 major bacterial divisions have no cultured representative’s [21]. Meanwhile, various software’s have been developed for gaining access into microbial diversity. In this context, Meyer et al. [22] have developed software RAST for studying phylogeny and function, Huang et al. [23] have developed software for the identification of rRNA genes from metagenomic fragments, based on hidden Markov models (HMMs). This software provides rRNA gene predictions with high sensitivity and specificity on artificially fragmented genomic DNAs. In addition to this, techniques like fluorescent in situ hybridization (FISH) targets 16S
rRNA of metabolically active bacteria, which have high cellular rRNA content, meanwhile, specific probes can be used to monitor microorganisms at different levels of taxonomic specificity [24, 25]. More recently, Hess et al. [26] have reported diversity of a cellulose and hemicellulose degrading bacteria from rumen of cow. It was demonstrated that rumen environment contains vast majority of cellulolytic mesophilic microbes, described from any habitat till date. The data obtained from such studies will substantially expand the catalogue of genes and genomes participating in the deconstruction of cellulosic biomass.

1.2. Metagenomic library construction and screening

Metagenomic library construction and screening constitutes a valuable tool for exploring novel biocatalysts and other molecules, and is constructed by extraction and purification of heavy molecular weight DNA [27–29]. The DNA obtained is digested with suitable restriction enzyme, followed by cloning of gene of interest in suitable vector molecules (vector choice depend upon the size of insert to be cloned) (Table1). After successful construction of a metagenomic library, it is screened either based on the function of gene or based on sequence method. The latter is dependent on the hybridization of the conserved DNA sequences of the target genes [30, 31]. Furthermore, screening based upon function always resulted in novel sequences (readers are referred to Table 2 for more detail). In addition to this, new screening methods like micro-array further help in identification of active clones among millions of clones [32–34]. Genome enrichment is another method which is used to target the active components of microbial population [35]. Another useful technique SIGEX (substrate-induced gene expression screening) is further used to select catabolic active genes, induced by various substrates using green fluorescent protein (GFP), in concert with fluorescence activated cell sorting [36]. Implementation of such screening method will provide comprehensive functional information about the unknown molecules in shorter timescale.

<table>
<thead>
<tr>
<th>Different vectors used in library construction</th>
<th>Fragment size to be cloned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid vectors (pUC, pBR322)</td>
<td>1–9kb</td>
</tr>
<tr>
<td>Fosmid and Cosmid vector</td>
<td>35–40kb</td>
</tr>
<tr>
<td>Phagemid vector</td>
<td>7–20kb</td>
</tr>
<tr>
<td>BAC vector</td>
<td>80–120kb</td>
</tr>
<tr>
<td>YAC vector</td>
<td>200–800kb</td>
</tr>
</tbody>
</table>

1.3. PCR based cloning

Polymerase chain reaction (PCR) is another useful technique that is utilized for exploring the genomic diversity of uncultured microbes. PCR allows highly selective amplification of target DNA under optimized conditions. Here, we would like to discuss some of the important genes cloned by PCR e.g. two β-ketoacyl genes were cloned from environmental DNA by PCR amplification [37], by designing oligonucleotides against conserved domain of known keto synthase and acyl carrier protein encoding genes. In another study from East China Sea, PCR
was utilized successfully for isolation of new polyketide synthase gene, by designing primer against two most conserved motifs, DPQQR and HGTGT [38]. These polyketide synthases (PKS) are multifunctional enzymes and catalyzes the formation of a polyketide assembled from sequential condensation of short chain acyl coenzymes. Some of them possess antitumor or antibiotic activity. Another modified PCR technique followed by genome walking, resulted in the isolation of two complete genes of 2, 5-diketo-acid reductase. The gene product obtained showed valuable properties like lower \( k_m \) and high thermostability, as compared to earlier reported genes [39]. Recently, an inverse PCR technique (I-PCR) technique followed by pre-amplified I-PCR (PAI-PCR) was performed, to obtain two novel full length xylanases genes from digestive tract of horse, which displayed ~49-64% amino acid sequence similarities to the known xylanases [40]. A novel cytochrome b5 gene was amplified and cloned from a metagenome sample, by designing degenerate primers from conserved motifs of cyt b5 [41]. Additionally, taking the advantage of most conserved regions of the lipase genes i.e. catalytic triad and oxyanion hole, a novel lipase gene was amplified and cloned, that showed ~20% homology to the existing gene [42]. PCR technique has also been utilized to explore the functional diversity of chitinase genes in unculturable marine bacteria [43].

2. Application of metagenomic gene cloning

2.1. Discovery of novel biocatalysts/molecules from different environment niches

The range of the organic reactions performed by the biocatalysts is enormous; therefore identification of suitable enzyme candidate is very much important. Among different class of enzymes, the most exploited enzyme class belongs to hydrolases. Using metagenomic approach, many novel genes encoding enzymes and antibiotic molecules had been reported. Here, we would like to discuss some of the novel lipases with potential application in industries. Recently, a careful selection of metagenome sample resulted in detection of 2661 lipolytic clones, which shared very less homology (~40-60%) with the earlier reported lipases/esterases. These lipases showed high specific activity against long chain triacylglycerols [44]. Recently, a metagenomic library constructed from Antarctic soil resulted in number of hypothetical, putative and novel cold adaptive functional enzymes, and many other mobile genetic elements [45]. In another study, a metagenomic library screened, resulted in ~350 novel lipases and esterases from environmental DNA samples, which showed high affinity for the synthesis of 1, 2-Oleoyl-3-palmitoyl-sn-glycerol (OOP) and 1, 3-Oleoyl-2-palmitoyl-sn-glycerol (OPO) [46]. Lipase plays an important role in biosynthetic reaction, a metagenomic derived lipase that showed its promising application in hydrolyzing stereo-selectively ibuprofen-pNP ester, with a high preference for the \((R)\) enantiomer of >91% ee, was discovered from forest soil [47]. The discovered lipase showed ~ 90% identity with the enzymes from \(P.\) \textit{fluorescens}LS107d2, B52 or \textit{Pseudomonas} sp. KB700A. Table 2 will further summarize diverse class of novel molecules and biocatalysts reported from various metagenome sources. From Table 2 it is evident that these enzymes are novel and shared very less homology with the existing enzymes.
<table>
<thead>
<tr>
<th>Source of Metagenome</th>
<th>Gene</th>
<th>Name of Enzyme</th>
<th>% Homology with the previous</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep sea sediments</td>
<td></td>
<td>Lipase/esterases</td>
<td>33-58%</td>
<td>[48]</td>
</tr>
<tr>
<td>Mangrove sediment</td>
<td></td>
<td>Lipase</td>
<td>25-52%</td>
<td>[6]</td>
</tr>
<tr>
<td>Deep golden mine</td>
<td></td>
<td>Esterases</td>
<td>55%</td>
<td>[49]</td>
</tr>
<tr>
<td>Active sludge</td>
<td></td>
<td>Lipase</td>
<td>52-71%</td>
<td>[50]</td>
</tr>
<tr>
<td>Oil contaminated soil</td>
<td></td>
<td>Lipase</td>
<td>90-98%</td>
<td>[51]</td>
</tr>
<tr>
<td>Bovine rumen microflora</td>
<td></td>
<td>Polyphenol oxidases</td>
<td>42-69</td>
<td>[52]</td>
</tr>
<tr>
<td>Deep sea environment</td>
<td></td>
<td>Aminotransferase</td>
<td>≥45</td>
<td>[53]</td>
</tr>
<tr>
<td>Forest soil</td>
<td></td>
<td>Antifungal</td>
<td>24-90</td>
<td>[54]</td>
</tr>
<tr>
<td>Hot spring</td>
<td></td>
<td>Cyclomaltodextrinase 45-85</td>
<td>56-72</td>
<td>[55]</td>
</tr>
<tr>
<td>Human gut</td>
<td></td>
<td>Xylanases</td>
<td>53</td>
<td>[56]</td>
</tr>
<tr>
<td>Lake</td>
<td></td>
<td>Cellulase and esterase</td>
<td>68-95</td>
<td>[57]</td>
</tr>
<tr>
<td>Pacific deep-sea sediment</td>
<td></td>
<td>Alkane hydroxylases</td>
<td>56-72</td>
<td>[58]</td>
</tr>
<tr>
<td>Acid mine drainage</td>
<td></td>
<td>Ni resistance gene</td>
<td>36-74</td>
<td>[59]</td>
</tr>
<tr>
<td>Sludge</td>
<td></td>
<td>Aromatic degrading enzyme</td>
<td>51-99</td>
<td>[60]</td>
</tr>
<tr>
<td>Soil</td>
<td></td>
<td>Amylase</td>
<td>47-78</td>
<td>[61]</td>
</tr>
<tr>
<td>Soil</td>
<td></td>
<td>2-Deoxy-scyllo-inosose (DOI) synthase</td>
<td>97</td>
<td>[62]</td>
</tr>
<tr>
<td>hot spring</td>
<td></td>
<td>Neopullulanase</td>
<td>45-48</td>
<td>[63]</td>
</tr>
<tr>
<td>Tidal flat</td>
<td></td>
<td>Lipase</td>
<td>34</td>
<td>[64]</td>
</tr>
<tr>
<td>Compost Soil</td>
<td></td>
<td>RNase H</td>
<td>40-72%</td>
<td>[65]</td>
</tr>
<tr>
<td>Mangrove soil</td>
<td></td>
<td>Laccase</td>
<td>42-52%</td>
<td>[66]</td>
</tr>
<tr>
<td>Gobi and Death Valley Deserts</td>
<td></td>
<td>Serine protease</td>
<td>53%</td>
<td>[67]</td>
</tr>
<tr>
<td>Antarctic coastal sediment</td>
<td></td>
<td>Protease</td>
<td>41-51</td>
<td>[68]</td>
</tr>
<tr>
<td>Antarctic soil</td>
<td></td>
<td>Carboxymethylcellulase</td>
<td>47%</td>
<td>[69]</td>
</tr>
<tr>
<td>Marine sponge goat skin surface</td>
<td></td>
<td>DNA polymerase</td>
<td>53</td>
<td>[70]</td>
</tr>
<tr>
<td>metagenome</td>
<td></td>
<td>Serine protease</td>
<td>98</td>
<td>[71]</td>
</tr>
<tr>
<td>Glacier Soil</td>
<td></td>
<td>Lipase</td>
<td>51-82%</td>
<td>[72]</td>
</tr>
<tr>
<td>pitcher fluid of the carnivorous plant</td>
<td></td>
<td>Lipase</td>
<td>32-41</td>
<td>[73]</td>
</tr>
<tr>
<td>Nepenthes hybrid Metagenome</td>
<td></td>
<td>Mettaloproteases</td>
<td>49-92</td>
<td>[74]</td>
</tr>
<tr>
<td>Rumenof Chinese Holstein</td>
<td></td>
<td>Glycoside Hydrolases</td>
<td>35</td>
<td>[75]</td>
</tr>
<tr>
<td>Soil</td>
<td></td>
<td>Xylanase</td>
<td>56-71%</td>
<td>[76]</td>
</tr>
<tr>
<td>yak rumen</td>
<td></td>
<td>Xylanase/endoglucanase</td>
<td>37-72</td>
<td>[77]</td>
</tr>
<tr>
<td>cow manure</td>
<td></td>
<td>Xylanases</td>
<td>45</td>
<td>[78]</td>
</tr>
<tr>
<td>soil</td>
<td></td>
<td>Turbomycin A and B</td>
<td>47-97</td>
<td>[79]</td>
</tr>
<tr>
<td>Soil</td>
<td></td>
<td>Aminotransferase</td>
<td>68-96</td>
<td>[80]</td>
</tr>
</tbody>
</table>
2.2. Exploring genomic diversity from various extreme environment niches

Genes in extreme environment are mostly related to catabolism, transport and degradation of complex organic molecules, and therefore provide valuable information about the abundance of cooperation, source of energy and aggregation among the microorganisms. Recent study from deep Mediterranean region has revealed presence of diverse class of metabolic genes that included lux, dehydrogenases and cox genes. Presence of these quorum sensing genes in deep Mediterranean environment, at least suggest existence of abundance of co-operation between different organisms, for shelter or food [81]. Furthermore, most abundant marine proteobacteria SAR 11 reported from Sargasso Sea revealed ~71% identity (at amino acid level) to earlier reported genome, and further observed a significant evolutionary divergence between the coastal isolates and Sargasso Sea populations. Additionally, this study has documented that genome rearrangements in SAR11 are not random but are concentrated at a particular site, which are often operon boundaries [82]. Interestingly, largest metagenomic study from one of such extreme environment has acquired 17 million new ORF, which has significantly changed the landscape of current protein space [83]. Furthermore, there are many Archaeal groups which are able to survive in most extreme environments, i.e., deep in sea, at temperatures, over 100ºC, hot springs, and in extremely alkaline or acid waters. However, little is known about their physiology and biochemistry and therefore it provides an opportunity for the researchers, to study such aspects. In this context, several metagenomic studies from different extreme environment (springs) has revealed the presence of many bacterial groups involved in anoxygenic type of photosynthesis, sulfur reduction, anaerobic fermentation, ammonia oxidation, viruses and many novel delta-proteobacterial groups [84-87]. Additionally, a unique kind of DNA polymerase has been characterized from uncultivated psychrophillicarcheon Canarchaeumsymbiosum, which lives in specific association with marine sponge. This DNA polymerase gene of Canarchaeumsymbiosum exhibits greatest specific activity towards gapped duplex DNA, and towards single stranded DNA, as a substrate with 3’-5’ exonuclease and minor 5’-3’ exonuclease activity [70] (Schleper et al. 1997).

2.3. Deciphering role of uncultivable microorganism present in human

Metagenome studies from humans is generally focused on structure and dynamic co-operations of microbial communities with human cells, and provide valuable insights into improvement of nutrition, drug discovery and preventative medicine. Recently, culture independent comparisons for mice and humans have revealed that despite sharing common bacterial phyla, most genera and species found in mice were not seen in humans. This study has further suggested the possible role of these microbial communities in producing many digestive disorders, skin diseases, gum diseases and even obesity [88]. Yet, in another study, a metagenomic library prepared from fecal microbiota demonstrated that BSH (bile salt hydrolyzing) activity is a conserved microbial adaptation to human gut environment, with a high level of redundancy in this ecosystem. Phylogenetic analysis illustrated that there must be some selective pressure in form of conjugated bile acid, which has resulted in evolution of abundance of BSH related microorganisms in human gut flora [89]. Metagenomic study will further allow the evaluation of surface marker involved in the interactions of microbes with the
eukaryotic cells [90]. Additionally, viral community i.e. phages present in human gut has huge impact on bacterial microflora. In one such study, a viral community analyzed from DNA of human feces resulted in ~ 1200 viral genotypes with great variation in the bacteriophages present in human feces and another environment [91]. Metagenomic study from human has further established role of microorganism in inflammatory bowel disease [92].

2.4. Depicting role of uncultured microorganisms in recycling of nutrients

Microorganisms play an important role in recycling of nutrients to the environment. Advancement made in molecular biology have demonstrated significant role of uncultured microorganisms in maintaining ecology. Recent study has demonstrated role of microbial associations in anaerobic cycling of carbon [93]. Therefore, there is a pressing need to explore these associations, to better understand the microbial processes. Newly introduced techniques, like shot gun sequencing and clone free pyrosequencing provides valuable insight into the metabolic potential of such microorganisms [94, 95]. Additionally, microorganisms play an important role by acting as a sink for various chemical reactions along with biogeochemical cycling of sulphur, phosphorous and iron. Previously, such processes were reported to occur abiotically [96]. Microbial activities play an important role in maintaining the biogeochemistry of the planet, in this regard, a metagenomic profiling of 45 microbiomes and 42 viromes suggest their repository role in storing and sharing genes among their microbial hosts, and influences the global evolutionary and metabolic processes [97]. Using metagenomic approach certain microorganisms have been studied which hold the capability to use selenite as a terminal electron acceptor, by a process, known as dissimilatory reduction of selenate (DSeR) [98]. Yet another study reported from three water bodies showed anaerobic bioconversion of selenium in such environment [99].

CONCLUSION

Conclusions and future directions

Metagenomics can provide the tools to balance the abundance of knowledge attained from culturing, with an understanding of the uncultured majority of microbial life. Metagenomics may further increase our understanding of many of the exotic and familiar habitats that are attracting the attention of microbial ecologists. It includes deep sea thermal vents; acidic hot springs; permafrost; temperate; desert; cold soils; Antarctic frozen lakes and eukaryotic host organs. However, a number of barriers have limited the discovery of new genes that provide insight into microbial community structure and function, and can be used to solve medical, agricultural, or industrial problems. Furthermore, most DNA extraction methods have been tested on a limited number of soil types, so their general applicability is unknown for comparative ecological studies. Additionally, given the profound utility and importance of microorganisms to all biological systems, methods are needed to access the wealth of information within the metagenome. Finally, it will expand and continue to enrich our knowledge about unexplored microorganisms and their wide applications.
ACKNOWLEDGEMENT

Authors wish to thank Council of Scientific and Industrial Research (CSIR, India) for granting senior research fellowship to PKS to pursue research on metagenomic gene cloning.

REFERENCES


Hinrichs KU, Hayes JM, Bach W et al. Proc Natl Acad Sci USA 2006; 103: 14684-14689.

