PREFACE

This book discusses metagenomics' methods, applications and perspectives.

Chapter 1 – Metagenomic analysis has extraordinary potential to improve our understanding of microbial populations in their natural environment and identify novel genes of interest. The key feature of such analyses is that they are performed using metagenomic libraries constructed from total DNA isolated from a particular niche rather than a laboratory culture. Thus, metagenomic analyses potentially allow access to all the genetic resources present in an environment, regardless of whether or not they belong to microorganisms that can be cultured in the laboratory. Sequence-based metagenomic analyses rely on comparisons with databases of known genomic sequences whilst functional analyses rely on screening libraries on the basis of the phenotypes cloned DNA can confer to host bacteria. Therefore, functional analysis allows the identification of novel genes with functions that could not have been predicted from their DNA sequence. However, a number of factors currently limit access to the full potential offered by functional metagenomic analyses. One major restriction is that despite the development of many procedures, indicators and genetic tools, we still lack effective screening methods for many activities. Another major limitation is the inefficient expression of some metagenomic genes in the host bacteria used for screening. Many metagenomic genes are derived from bacteria with highly divergent physiologies and gene expression machineries that are absent from the surrogate host. This review focuses on the main problems that limit the potential of functional analyses and on approaches that can be used to, at least partially, circumvent these problems and have allowed the identification of a large number of different activities from metagenomic libraries.

Chapter 2 – In the last decade, metagenomics has created a great revolution in microbial ecology. Metagenomics allows the study of uncultured microorganisms from the total DNA of environmental samples. This is an important approach to elucidate the structure of microbial communities and understand the functions that occur in complex environmental samples. At first, metagenomics covered only the construction of metagenomic libraries by cloning large fragments of DNA in appropriated vectors. Currently, with high-throughput DNA sequencing technologies, the metagenomic approach involves the massive sequencing of total DNA or total RNA, without the cloning step, of environmental samples. This chapter is a review of different strategies that can improve the library construction or massive sequencing of environmental samples and consequently increase the chances of finding a positive hit. It will also describe screening methods currently based on sequence and function to look for positive clones and interesting genes in the metagenomic library.

Chapter 3 – Characterizing the microbial community is the first step towards the understanding of the ecological aspects of the hydroelectric reservoirs, since those environments are highly complex and heterogeneous. One reason for this complexity is that foundations of Hydroelectric Power Stations (HPS) flood a wide area in order to create the reservoir, trapping a variety of organic matters from the vegetation around and sheltering a rich fauna of fish and other animals. It's estimated that all area inundated by hydroelectric reservoirs around the globe is equivalent to the area of the German territory. Those engineered environments arouse the curiosity of the scientific community, since they are considered a source of greenhouse effect gases (like methane and carbon dioxide), interfering on the life cycle from (micro)organism. This chapter consists of an original research that aimed to assess the microbial diversity of five HPS's reservoirs at the Brazilian territory. The authors used semiconductor sequencing by Ion Torrent Personal Genome Machine (PGM) to determine the bacterial diversity of five reservoirs from plants already in operation – Xingó, Itaipú, Três Marias, Balbina and Funil. They collected one liter of water from a series of points in the reservoirs to extract the total DNA, joining samples in a pool according to the HPS. The authors performed a polymerase chain reaction to amplify the 16S rRNA bacterial gene using universal primers with barcodes and sequencing adaptors to each pool of sample. The authors sequenced the amplicons in a single run at Ion Torrent PGM, chip 318. The sequencing resulted in 2,900,226 reads with quality value greater than 20 and length greater than 100 base pairs, originating around 581 Mbp of highly reliable genetic information. They submitted the sequences to alpha diversity and taxonomic analyses using RDP pipeline and Mothur, which we consider robust and highly reproducible. The most abundant bacterial phyla were Actinobacteria, Proteobacteria, Bacteroidetes, Verrucomicrobia and Cyanobacteria. Other less representative phylum found were OD1, Planctomycetes, TM7, Acidobacteria, Firmicutes, Nitrospira, Gemmatimonadetes, Chlamydiae, Chlorobi, Fusobacteria, Deinococcus-Thermus, WS3, Armatimonadetes, Chloroflexi, OP11, Lentisphaerae, SR1 and Thermotogae. The differences of microbial communities found in the reservoirs illustrate the vast variability between the HPS's reservoir, in spite of that the most abundant bacteria were assigned to the orders Burkholderiales, Rhizobiales and Actinomycetales, but photosynthetic and methanotrophic bacteria were also significant in all reservoirs.

Chapter 4 – Metagenomics provides new opportunities in environmental science and technology. Next-generation sequencing (NGS) technologies, coupled with advanced bioinformatics tools, have enabled rapid progress in microbial ecology and discovery of novel genes. In this chapter, an overview of the current state of analysis of metagenomic data and their possible applications for bioremediation of pesticides, strategies to combat global climate change and development of novel biomarkers for assessing water quality in light of currently available resources and tools are presented.

Chapter 5 – Analyses of microbial genome sequences have revealed numerous number of gene sequences, with the potential to code the enzymes for using an industrial, agricultural and pharmaceutical application. Metagenomics have identified a number of novel genes and enzymes in the environment. However, eukaryotic microorganisms are not being used to be a target of functional metagenomics. Since functional genes of eukaryotic microorganisms are separated by intron, it is very difficult to screen functional genes of eukaryotic microorganisms by DNA-based method. Metatranscriptomic is mRNA-based functional community analyses method based on expressed genes is a more suitable means to identify

eukaryotic genes and enzymes in the environment, because metagenomic analysis based on DNAs cannot determine the structural genes whose introns are excluded, let alone detect ecologically relevant active functions. An RNA-based metatranscriptomic approach can circumvent the recurrent problems in the conventional metagenomic approach, and 3' poly-A tails-specific purification and subsequent reverse transcription lead to construction of a cDNA library, allowing comprehensive analyses of the eukaryotic genes specifically expressed. This chapter summarizes the methods that have been developed for exploring the genetic and functional diversity of eukaryotes by applying a metatranscriptomic approach to target genes encoding an industrial application.

Chapter 6 – The search for ideal biocatalysts for specific applications is in progress all over the world. Although spectacular advancements have been made in improving the properties (stability to high pH and temperatures, affinity and activity) of biocatalysts using protein engineering and directed evolution, there is still a huge gap between what are available and those needed to be functional in extreme industrial process conditions. The majority of enzymes that are in use today have been sourced from mere 0.1% of the culturable microbes. There is a possibility of obtaining novel biocatalysts from the major portion of non-culturable microbial diversity. The recently emerged culture–independent approach, metagenomics, allows accessing genes encoding novel biocatalysts from the major portion of non-culturable microbial diversity. The emerging field of metagenomics has truly boosted the chances of discovering novel biocatalysts which will revolutionize biocatalysis. Metagenomics helps in understanding culturable and non-culturable microbial diversity in the environment besides discovering novel biocatalysts and other metabolites.

The recent developments in the retrieval of biocatalysts from a great variety of environmental samples employing metagenomics approaches will be reviewed in this chapter. First step in metagenomics is the extraction of humus-free DNA from the environmental samples. This has been achieved by treatment of the extracted DNA with activated charcoal and polyvinylpolypyrridine as well as using commercial kits. The second obstacle is the screening libraries obtained from metagenomes, which is labor intensive. Robot-assisted systems are now being used for this purpose. Sequence and activity driven analyses are the two approaches for screening clones obtained through metagenomic libraries. Sequence driven analysis is independent of expression of the cloned gene, while activity based screening relies on the expression of the gene. As activity based screening has nothing to do with already reported sequences in the database, the probability to access novel genes is more as compared to the sequence based analysis that fully depends on existing sequences for the respective genes. The function based approach always gives the full length gene of the expressed clone, while the sequence based approach primarily retrieve the partial sequences. Functional based approach is, therefore, more promising to reveal the hidden Pandora's box of the inaccessible microbiota.

Several starch, cellulose, xylan, protein, lipid, chitin, phytate and pectin hydrolyzing enzymes, and nitrilase, nitrile hydratase, and amidases have been discovered from metagenomes of a great variety of environmental (normal and extreme) samples.

The discovery of enzymes by culture-independent metagenomic approaches is no more a concept, but a reality. The Verenium Corporation, USA has commercialized several enzymes including phytase developed through metagenomics. It is surprising that several highly important enzymes like phytase, amylopullulanase, urease, superoxide dismutase, asperginase and carbonic anhydrase represent the class of biocatalysts that are either not touched or

properly exploited using this unconventional approach. Only one xylanase has been retrieved by metagenomic approach that can withstand extreme conditions prevailing in paper and pulp industries. Metagenomics is being peeled but extensive efforts are needed to understand the mechanisms involved in finding the association of microorganisms with their habitat and their unculturability on plates. Managing the metadata generated through metagenomics is another challenge. This selfish technology must continue exploration of novel biocatalysts and other biological products. The chapter focuses on recent developments made in the retrieval of biocatalysts from a variety of environmental genomes employing metagenomic approaches.

Chapter 7 – Metagenomics, enabling the study of majority of unculturable microbes, has opened new vistas of understanding the microbial diversity. Among various natural environments being studied by the microbiologists, soil is probably the most challenging reservoir with respect to microbial community size and diversity of species. Soils harbour a large number of microorganisms, of which only 0.1-1.0% have been reported to be culturable under standard conditions. Therefore, culture independent approach would play a major role in unravelling the hidden microbial flora not only from soil but from other environmental samples as well. The success of any soil metagenomic investigation is crucially dependent on method(s) used for isolation of metagenomic DNA from soil samples. An ideal protocol should enable an efficient in extraction of high molecular weight DNA free from inhibitors and amenable to biotechnological manipulations. High yield of DNA is also an important criterion for accessing the effectiveness of a protocol for isolation of the environmental DNA. Also, the extracted DNA should be an unbiased representation of the microbial community of the soil sample, i.e. the protocol should be able to extract DNA from hard to lyse cells and rare species efficiently. Several protocols for isolation of soil metagenomic DNA have been developed and these protocols have been grouped into two categories- strategies that consist of direct extraction of nucleic acids from soil through *in situ* lysis and the second approach is based on the separation of bacteria from the soil particles followed by the extraction of DNA. This chapter focuses on an overview of the challenges encountered during extraction of soil DNA and the strategies that can be adopted to achieve this challenging objective.

Chapter 8 – Over the past few decades microbial life in extreme environments has attracted broad scientific interests. Extreme environments harbor microorganism that represent the oldest inhabitants on Earth, and whose high adaptability has continued to challenge our understanding of biochemistry, biology and evolution. The majority of microorganisms cannot be cultivated using established laboratory methods thus requiring alternative approaches for their characterization. Metagenomics is a culture-independent genomic analysis, divided into sequence-based and function-driven analyses. These two branches of metagenomics address the challenge of studying microbial communities and functions in environments that are as yet unculturable and that represent more than 99% of the organisms in extreme environments. This new approach expands our understanding of the ecology and evolution of organisms, and aids in the discovery of diverse members of previously undefined classes of microorganisms. Metagenomics, does not need any selection (e.g. cultivation/enrichment) and greatly reduces technical biases often encounter in pure culture selection. This review will compare methods, highlight progress, discusses opportunities, challenges and perspectives in metagenomics with special reference to genome discoveries in extreme hot environments.