Genomics and Marine Prokaryotes

Chandrasekaran. M and Sarita G Bhat

Department of Biotechnology
Cochin University of Science and Technology
Cochin-682022
mchandra@cusat.ac.in

Introduction

The dawn of 21st century is witnessing rapid advancements in the field of genomics, which has become an integral and essential element of biotechnology and molecular biology. On a broader scale, genomics is forging new perceptions of life and our concept of world. Marine microorganisms have received greater attention in biotechnology ever since scientists brought to light the potential of marine microbes for production of rare bioactive compounds and secondary metabolites with unique structures and of antimicrobial, antiviral, and antitumor activities in the field of biopharmaceuticals (Okami, 1986; Fenical and Jensen, 1993). Since then there is a surging interest around the world towards efficient utilization of marine microbes and biotechnology. In general, the recent advancements in the field of marine microbes and biotechnology, which is flowing into the scientific literature, could be attributed to the advancements in molecular biology and in bioprocess technology, in the light of an increased awareness on the potentials of biotechnology and their probable applications.

Marine microorganisms remain unexplored in terms of their biology and potential utilization in biotechnology compared to their terrestrial counterparts. Nevertheless, current research activities on marine microbes, around the world, include screening of new bioactive compounds; development of indigenous microbes for enhanced bioremediation; fermentation production of exoenzymes of industrial and pharmaceutical importance; development of strains for biopolymers, industrial chemicals, biomaterials, biofuel, etc.; molecular genetics of rare and useful genes; protoplast fusion; isolation of genes and expression in known hosts for better understanding of gene expression; and molecular cloning and expression of rare genes in *in vivo* conditions and in natural environments; microbial diversity in deep seas. In this context the emerging field of genomics holds immense scope for exploration and exploitation of marine microbes particularly prokaryotes for various applications besides understanding their biology.

Marine Environment and Marine Microbes

Marine environment, encompasses about 71 percent of the earth's surface, is rich not only with biodiversity but also a resource for potential microorganisms of useful applications, compared to the terrestrial environments. Microbes inhabit various habitats of marine environment which include neuston, plankton, nekton, seston, epibiotic, endobiotic, pelagic and benthic environments. In general oceans, seas, bays, fjords, estuaries and backwaters are considered as marine environments. These habitats harbor a diverse range of microbes including archaebacteria, cyanobacteia, eubacteria, actinomycetes, yeasts, filamentous fungi, microalgae, algae, and protozoa. In spite of the fact that there are abundant microorganisms in the marine environment, they are yet to draw the full attention, at global level, in terms of their application in biotechnology, and consequently they remain unexplored in terms of their biology and their potential utilization.

Marine Prokaryotes

In general, marine microorganisms are the group understudied. Although marine microbiology is about 70 years old, earlier attempts on marine microbes were primarily on heterotrophic bacteria associated with planktons, fish spoilage, sea food poisoning, bioremediation of coastal waters and cycling of elements and consequent fertility of seas. Of course all the studies dealt with the prokaryotes in quantitative terms and not in qualitative terms. The only group that attracted major attention with respect to marine prokaryotes is the *Vibrio* sp in terms their role in food poisoning and fish disease compared to other roles. Very few species of prokaryotes have been recognized for their potential and has been studied. There is a resurgence of interest in developing countries to study deep sea prokaryotes and Antarctic microbes in the last decade which may get strengthened in the coming years. Other groups of prokaryotes such as marine cyanobacteria and marine archaebacteria or other eubacteria never could draw the attention of scientists probably due to lack of appropriate cultivation strategies, culture medium and inadequate taxonomic tools besides know-how on maintenance of the same in laboratory conditions. Whereas, all the three groups namely archaebacteria, eubacteria and cyanobacteria of various marine environment holds immense potential. The emergence of modern molecular techniques holds promise to explore them for better understanding of their biology, ecology, and biotechnological potentials.

Genomic Approach

With the advent of molecular techniques such as, PCR amplification of genomic DNA from environmental samples, DNA shot gun sequencing and computation of metagenomes has led to the rapid development of genomic approaches that enable the understanding of biodiversity and ecosystems structures and dynamics. Further, techniques such as Throughput sequencing, Transcriptome analysis and Proteomics, enable scientists to investigate questions related to the protection of biodiversity and genetic resources and to the functioning of ecosystems, including the underpinning of ecological and behavioral research, the understanding of population structure, the identification of species and the understanding of evolutionary relationships. These approaches will be useful for developing new products and services in the area of environmental protection. In the twenty-first century, genomic innovations through genomic approaches will invariably bring about radical changes in medicine, agriculture, and the study of our evolutionary heritage.

Concept of Genomics

Genomics is the study of an organism's entire genome. The field of Genomics can be said to have appeared in the 1980s and took off in the 1990s with the initiation of genome projects for several species. A major branch of genomics is still concerned with sequencing the genomes of various organisms, although the knowledge of full genomes have created the possibility for the field of functional genomics, mainly concerned with patterns of gene expression during various conditions. The most important tools here are microarrays and bioinformatics.

Different Flavors of genomics

Today the field of genomics have developed such that there are several flavors attributed to genomics such as structural genomics, functional genomiccs, Phylogenomics, Metagenomics

Structural genomics consists in the determination of the three dimensional structure of all proteins of a given organism, by experimental methods such as X-ray crystallography, NMR spectroscopy or computational approaches such as homology modelling. While most structural biologists pursue structures of individual proteins or protein groups, specialists in structural genomics pursue structures of proteins on a genome wide scale. This implies large scale cloning, expression and purification. One main advantage of this approach is economy of scale.

Functional genomics is a field of molecular biology that attempts to make use of the vast wealth of data produced by genome sequencing to describe gene (and protein!) functions and interactions. Unlike genomics and proteomics, functional genomics focus on the dynamic aspects such as gene transcription, translation, and protein-protein interactions, as opposed to the static aspects of the genomic information such as DNA sequence or structures. Functional genomics includes function-related aspects of the genome itself such as mutation and polymorphism (such as SNP) analysis, as well as measurement of molecular activities. The latter comprise a number of "-omics" such as transcriptomics (gene expression), proteomics (protein expression), phosphoproteomics and metabolomics. Together these measurement modalities quantifies the various biological processes and powers the understanding of gene and protein functions and interactions.

Phylogenomics is a method of assigning a function to a gene based on its evolutionary history in a Phylogenetic tree. Phylogenomics uses knowledge on the revolution of a gene to improve function and prediction.

Metagenomics (also Environmental Genomics, Ecogenomics or Community Genomics) is the study of genomes recovered from environmental samples as opposed to from clonal cultures. Metagenomics is based on the genomic analysis of microbial DNA that is extracted directly from communities in environmental samples. This technology - genomics on a huge scale - enables a survey of the different microorganisms present in a specific environment, such as water or soil or sediment, to be carried out. Metagenomics represents a powerful tool to access the abounding biodiversity of native environmental samples and their tremendous but unutilized potential in application in many aspects of industry, therapeutics, and environmental sustainability.

Scope of Environmental genomics /Metagenomics

The valuable property of metagenomics is that it provides the capacity to effectively characterize the genetic diversity present in samples regardless of the availability of laboratory culturing techniques. Information from metagenomic libraries has the ability to enrich the knowledge and applications of many aspects of industry, therapeutics, and environmental sustainability. This information can then be applied to society in an effort to create a healthy human population that lives in balance with the environment. With the global political drive to promote white (industrial) biotechnology as a central feature of the sustainable economic future of modern industrialized societies, there is a rapid requirement in the development of novel enzymes, processes, products and applications.

Conventional sequencing begins with a culture of identical cells as a source of DNA. However, early metagenomic studies revealed that there are probably large groups of microorganisms in many environments that cannot be cultured and thus cannot be sequenced. These early studies focused on 16S ribosomal RNA sequences which are relatively short, often conserved within a species, and generally different between species. Many 16S rRNA sequences have been found which do not belong to any known cultured species, indicating that there are numerous unisolated organisms out there. Recovery of DNA sequences longer than a few thousand base pairs from environmental samples was very difficult until recent advances in molecular biological techniques, particularly related to constructing libraries in bacterial artificial chromosomes (BACs), provided better vectors for molecular cloning. In addition, advances in bioinformatics, refinements of DNA amplification, and proliferation of computational power have greatly aided the analysis of DNA sequences recovered from environmental samples by metagenomics. A 2004 metagenomic study of the Sargasso Sea found DNA from nearly 2000 different species including 148 types of bacteria never seen before. Another study, also from 2004, revealed the genomes of bacteria and archaea from an acid mine drainage system that had resisted attempts to culture them. Because the collection of DNA from an environment is largely uncontrolled, large samples, often sometimes prohibitively so, are needed to fully resolve the genomes of underrepresented members of a microbial community. On the other hand, many such underrepresented organisms might never be noticed without metagenomic analysis if they are difficult to isolate using traditional culturing techniques.

Environmental genome sequencing and analysis are rapidly increasing our knowledge of the genetic and functional diversity of bacteria and archaea. Diverse environments have been examined, including soil, marine sediments, marine waters, Antarctic systems, and acid mine drainage systems. These studies have produced physiological insights, such as the widespread nature of rhodopsins in marine bacteria, as well as insights into biogeochemistry, such as the genetic inference of the mechanism of methane oxidation in anoxic environments. Antarctic metagenomic studies have included an examination of phylogenetic diversity in deep polar front waters, comparative genomics of Antarctic and temperate archaeal genome fragments, and genome fragment analysis of a marine euryarchaeote.

Exploring unknown Microbial Diversity by environmental genomics

Molecular fingerprints indicate the presence of large numbers of so far uncultured microorganisms - potentially millions of species - in common marine habitats such as temperate and cold marine waters or sediments. More than 100 years of scientific microbiology have merely yielded an initial insight into this immense microbial diversity. Today, we often know little more of these microbes than the sequence of a single gene, usually the one coding for the 16S rRNA. From comparative sequence analysis affiliation - or lack of affiliation - with cultured bacteria can be deduced. Fluorescence in situ hybridization with rRNA-targeted oligonucleotide probes yields quantitative data on the abundance and spatio- temporal distribution of microorganisms. The use of horseradish peroxidase-labeled oligonucleotides is currently increasing the sensitivity of this method considerable. By means of automated microscopy and flow cytometry multiple samples can be analyzed. Thereby abundant marine "key species" have been identified. By targeted isolation more and more representative pure cultures are retrieved. Although isolation of pure cultures of many marine prokaryotes remains difficult, there is need to continue and strengthen the efforts in this respect. Full genome sequencing of many isolates is currently ongoing and this undoubtedly will result in a better understanding of the environmental adaptations and the biotechnological potential.

The shot gun sequencing of "metagenomes", a new term defined as the sum of the genomes that can be found in a particular habitat, has started. The huge effort Craig Venter is currently only one of several attempts to analyze the microbial diversity of the world oceans as detailed as possible. Along similar lines, large genomic fragments from marine planktonic and benthic communities, e.g., from recently discovered assemblages catalyzing the anaerobic oxidation of methane are retrieved and being analyzed. No doubt that the next decade of marine genomics will teach us much about the role of marine prokaryotes in the global cycling of elements, and, as a spin-off of basic research also result in the discovery of new natural products.

Accessing Metagenomes for industrial applications

Historically biotechnology has missed to screen 99% of existing microbial resources. As a consequence strategies of directly cloning and screening of environmental DNA are becoming increasingly popular to circumvent this restriction. Comprising the genetic blueprints of entire microbial consortia the metagenomes provide functionally meaningful molecular sequence space in terms of novel enzymes and biocatalysts, the construction of optimised „designer-bugs" and the identification of new natural compounds for industrial applications.

Different industries are interested in exploiting the resource of uncultivated micro-organisms: ***Ideal biocatalyst*** - Instead of designing a process to fit a mediocre enzyme, it is conceivable that the uncultivated microbial diversity, together with in vitro evolution technologies, might be used to find a suitable enzyme that that optimally fits process requirements. ***Elusive metabolites***- Many pharmacologically active secondary metabolites are produced by bacteria that live in complex consortia or by bacteria that inhabit niches that are difficult to reconstitute in isolated production hosts. So, the cloning and heterologous expression of biosynthetic gene clusters that encode secondary metabolites is the most straightforward method of accessing their biosynthetic potential. ***Novelty***- For industries a single enzyme backbone with superior functionality that has an entirely new sequence or a single pharmacologically active natural compound with a novel structure and new mode of action would be useful to avoid infringing competitors' intellectual property rights.

Studies are being conducted on evaluation and expression of high molecular weight DNA from metagenome cloned as Large-Insert Libraries (LIL®s) as well as the development of flexible LIL expression systems for heterologous expression of metagenome DNA in biosynthetically competent expression hosts. The biosynthetic potential is examined by screening for the production of novel enzymes as well as novel anti-microbial activities and the presence of secondary metabolite producing biosynthetic gene clusters.

New proteins from (meta)genomes of marine hydrocarbon-degrading organisms and their consortia

Recently discovered hydrocarbonoclastic marine bacteria are ubiquitous, possess a narrow substrate profile and have preference for aliphatic hydrocarbons and their derivatives. The genome of one of them, *Alcanivorax borkumensis*, has recently been sequenced and annotated. Its functional genome analysis revealed metabolic features of the "hydrocarbonoclastic" lifestyle: a large repertoir of monooxygenases and systems for scavenging oligominerals. From its genome a number of novel, not predicted in silico, carboxylesterases with a great enantioselectivity in the kinetic resolution of variety of chiral synthons, were retrieved and characterised. *Oleispira antarctica* is another new gamma-proteobacterium that efficiently degrades oil hydrocarbons at low temperatures. Activity- based screening of its genome library revealed a new carboxylesterase that was poorly expressed in *E. coli*. Trying to co-express this enzyme with Cpn60 chaperonin from Oleispira in *E. coli* to facilitate better enzyme solubility, it was discovered that E. coli became capable of growth at temperatures as low as 0°C. The consequent study of this phenomenon revealed a small group of essential proteins (Dps, ClpB, RpsB and DnaK) whose cold denaturation causes the systems failure in *E. coli* at low temperatures. An activity-based study of a metagenome library derived from the oil-degrading marine community from a brine-seawater chemocline of Urania deep Mediterranean hypersaline anoxic basin revealed that among few distinct groups of carboxylesterases retrieved from the librariy some enzymes exhibited unusual, habitat-specific characteristics (preference for high hydrostatic pressure, anoxia and high salinity). One exhibited an unusual structural signature incorporating three catalytic active centers mediating distinct hydrolytic activities and an adaptive tertiary-quaternary structure that altered between three molecular states, depending on the prevailing physicochemical conditions. Some of the esterases had high activities, specificities, enantioselectivities, and exceptional stability in polar solvents, and exhibited a good potential for industrial biotransformations.

Heterologous expression of natural product biosynthetic pathways

Natural products from symbiotic associations between marine invertebrate and prokaryotes show exceptional promise as pharmaceuticals in many therapeutic areas. The predicted lack of an economic and sustainable global market supply for production of marine-derived drugs, due to difficulty of synthesis, is often cited as the main obstacle to invest in and exploit these otherwise exciting bioactive compounds. Different strategies have been evoked to overcome this impediment as long-term harvesting of wild stocks from the