Standard Review

Metagenomics - An advanced approach for non-cultivable micro-organisms

Md. Zeyaullah¹*, Majid R. Kamli², Badrul Islam¹, Mohammed Atif³, Faheem A Benkhayal¹, M. Nehal³, M.A. Rizvi⁴ and Arif Ali²

¹Department of Microbiology, Faculty of Medicine, Omar Al-Mukhtar University, Al-Baida, Libya.
²Department of Biotechnology, Jamia Millia Islamia, New Delhi- 110025, India.
³Ageing Research Laboratory, University Department of Zoology, L. N. Mithila University, Kameshwara Nagar, Darbhanga- 846 008, Bihar, India.
⁴Department of laboratory Medicine, Higher Institute of Medical Technology, Derna, Libya

Accepted 14 April, 2009

It is known that only 0.1 - 10% of all microorganisms observed in nature can be cultured under conventional laboratory conditions. This leaves researchers unable to study more than 99% of microorganisms in some environments - microorganisms that sometimes have unique and potentially very useful abilities such as waste degradation or synthesis of compounds that could find use as drugs or antibiotics. Metagenomics, the genomic reconstruction of unculturable microorganisms, is a powerful new tool for accessing the untapped resources of biodiversity in environmental samples. The ability to extract and purify high MW DNA from difficult samples such as waste effluents and soil, attempts to provide a breakthrough in representative metagenomic library construction for metagenomics that may enable the discovery of many future drugs and antibiotics.

Key words: Metagenomics, microorganisms, waste effluents and soil, drugs and antibiotics.

INTRODUCTION

Metagenomics is the culture-independent analysis of a mixture of microbial genomes (metagenome) using an approach based either on expression or on sequencing (Riesenfeld et al., 2004; Schloss et al., 2003; Susannah et al., 2005; Patrick et al., 2005). The term is derived and coined (Covacci et al., 1997) from the statistical concept of meta-analysis (the process of statistically combining separate analyses) and genomics (the comprehensive analysis of an organism’s genetic material) (Rondon et al., 2000) to capture the notion of analysis of a collection of similar but not identical items as in a meta-analysis which is an analysis of analyses (Glass, 1976). Metagenomic methodology has been developed as an effective tool for the discovery of new natural products and microbial functions (He et al., 2007) and as such it as the application of genomics suit of technologies to uncultivated microorganisms is coming of age (Gabor et al., 2007).

Metagenomics is employed as a means of systematically investigating, classifying and manipulating the entire genetic material isolated from environmental samples. This multi-step process relies on the efficiency of four main steps (Figure 1). Those steps are:

(i) The isolation of genetic material.
(ii) Manipulation of the genetic material.
(iii) Library construction.
(iv) The analysis of genetic material in the metagenomic library.

Biotechnological applications from metagenomics may be fostered by the pursuit of fundamental ecological studies and focused screens for bioprospecting, just as both basic and applied approaches have contributed to the discovery of antibiotics and industrial enzymes from cultured microorganisms.

The discovery of streptomycin and other bacterial antibiotics sprang from very basic studies of the taxonomy and ecology of actinomycetes in soil conducted by soil microbial ecologist, Selman Waksman (Waksman et al., 1937; Waksman et al., 1916) to develop the maximum number of biotechnological applications.

Metagenomics being still a young and exciting technique
has a broad scope of application in the fields of biology and biotechnology. Although many advances in heterologous gene expression, library construction, vector design and screening may still improve it more, the current technology has proved to be sufficiently powerful to yield products for solving real world problems, including the discovery of new antibiotics and enzymes. Recent studies in the Sargasso Sea (Venter et al., 2004), acid mine drainage (Tyson et al., 2004), soil (Tringe et al., 2005), and sunken whale skeletons, have used the shotgun-sequencing approach to sample the genomic content of these varied environments. In each study, environmental samples were obtained and the microbial DNA was extracted directly from the sample, sheared, cloned into *Escherichia coli* and random clones were sequenced. But genomics has failed to elucidate the functions of microbial communities where most microorganisms on Earth spend most of their time and provide the platform to shape plant, animal, environment and human health.

Even the metagenomics, coupled with gene arrays, proteomics, expression-based analyses, and microscopy, will provide insights into the studies of problems such as genome evolution and the members of particular niches that are currently hindered by our inability to culture most microorganisms in pure culture (Allen et al., 2005).

Making the metagenomic studies ecologically meaningful will require sampling strategies that account for spatial and temporal variability, thereby enabling comparisons between communities. These comparisons will also require standardized and aggressive methods for extracting DNA. It is unfortunate that all of the large metagenomic sequencing projects used chemical extraction methods to obtain DNA, whereas the technique of ‘bead beating’, which applies high shear forces to cells, is more effective than chemical lysis methods for breaking tough cells (Miller et al., 1999). Among the methods designed to gain access to the physiology and genetics of uncultured organisms, metagenomics being the genomic analysis of a population of microorganisms, has emerged as a powerful centerpiece (Whitaker et al., 2006). Direct isolation of genomic DNA from an environment circumvents culturing the organisms under study and cloning of it into a cultured organism captures it for study and preservation.

Advances have derived from sequence-based and functional analysis in samples from water and soil and associated with eukaryotic hosts. The idea of cloning DNA directly from environmental samples was first proposed by Pace (Pace et al., 1985) and in 1991, the first such cloning in a phage vector was reported (Schmidt et al., 1991). The next advance was the construction of a metagenomic library (Figure 2) with DNA derived from a mixture of organisms enriched on dried grasses in the laboratory (Healy et al., 1995). The work of DeLong’s group defined the field when they reported libraries constructed from prokaryotes from seawater (Stein et al., 1996). They identified a 40 kb clone that contained a 16S
**Figure 2.** Construction and screening of metagenomic libraries. Schematic representation of construction of libraries from environmental samples. The images at the top from left to right show bacterial mats at Yellowstone, soil from a boreal forest in Alaska, cabbage white butterfly larvae, and a tube worm (Patrick et al., 2005).

rRNA gene indicating that the clone was derived from an Archae which had never been cultured. A powerful yet challenging approach to metagenomic analysis is to identify clones that express a function. Success requires faithful transcription and translation of the gene or genes of interest and secretion of the gene product, if the screen or assay requires it to be extracellular. Functional analysis has identified novel antibiotics (Courtois et al., 2003; Gillespie et al., 2002), antibiotic resistance genes (Diaz-Torres et al., 2003; Riesenfeld et al., 1999), Na$_+/$(Li$_-$/H$_-$/) transporters (Majernik et al., 2001) and degradative enzymes (Healy et al., 1995; Henne et al., 1999; Henne et al., 2000).

**APPROACHES TO METAGENOMIC ANALYSIS**

Metagenomic analysis involves isolating DNA from an environmental sample, cloning the DNA into a suitable vector, transforming the clones into a host bacterium and screening the resulting transformants. The clones can be screened for phylogenetic markers or “anchors,” such as 16S rRNA and recA or for other conserved genes by hybridization or multiplex PCR (Stein et al., 1996) or for expression of specific traits, such as enzyme activity or antibiotic production (Courtois et al., 2003; Diaz-Torres et al., 2003; Gillespie et al., 2002; Knietsch et al., 2003; Lorenz et al., 2002; Lorenz et al., 2002; Mac Neil et al., 2000).
Sequence-based analysis

Sequence-based analysis can involve complete sequencing of clones containing phylogenetic anchors that indicate the taxonomic group that is the probable source of the DNA fragment. Alternatively, random sequencing can be conducted, and once a gene of interest is identified, phylogenetic anchors can be sought in the flanking DNA to provide a link of phylogeny with the functional gene. Sequence analysis guided by the identification of phylogenetic markers is a powerful approach first proposed by the DeLong group, which produced the first genomic sequence linked to a 16S rRNA gene of an uncultured Archaeon (Stein et al., 1996). Subsequently, they identified an insert from seawater bacteria containing a 16S rRNA gene that affiliated with the γ-Proteobacteria.

The sequence of flanking DNA revealed a bacteriorhodopsin-like gene. Its gene product was shown to be an authentic photoreceptor, leading to the insight that bacteriorhodopsin genes are not limited to Archaea but is in fact abundant among the Proteobacteria of the ocean (Beja et al., 2000; Beja et al., 2001). A promising application of phylogenetic anchor-guided sequencing is to collect and sequence many genomic fragments from one taxon. In more complex environments and taxa, reassembly of a genome may not be feasible but inference about the physiology and ecology of the members of the groups can be gleaned from sequence data. This approach has been initiated with clones from diverse soils carrying 16S rRNA genes that affiliate with the Acidobacteria phylum, which is abundant in soil and highly diverse (Barns et al., 1999; Buckley et al., 2003) and about which little is known (Liles et al., 2003; Quaiser et al., 2003). Complete sequencing of the estimated ~500 kb of Acidobacterium DNA in metagenomic libraries may provide insight into the subgroups of bacteria in this phylum that have never been cultured.

The alternative to a phylogenetic marker-driven approach is to sequence random clones, which has produced dramatic insights, especially when conducted on a massive scale. The distribution and redundancy of functions in a community, linkage of traits, genomic organization, and horizontal gene transfer can all be inferred from sequence-based analysis. The recent monumental sequencing efforts, which include reconstruction of the genomes of uncultured organisms in a community in acid mine drainage (Tyson et al., 2004) and the Sargasso Sea (Venter et al., 2004); illustrate the power of large-scale sequencing efforts to enrich our understanding of uncultured communities. These studies have made new linkages between phylogeny and function, indicated the surprising abundance of certain types of genes, and reconstructed the genomes of non-culturable organisms. The use of phylogenetic markers either as the initial identifiers of DNA fragments to study or as indicators of taxonomic affiliation for DNA fragments carrying genes of interest because of their function is limited by the small number of available markers that provide reliable placement in the Tree of Life. If a fragment of DNA that is of interest for other reasons does not carry a dependable marker, its organism of origin remains unknown. The collection of phylogenetic markers is growing, and as the diversity of markers increases, the power of this approach will also increase, making it possible to assign more fragments of anonymous DNA to the organisms from which they were isolated. Moreover, as more genomes are reconstructed, more genes will be linked to phylogenetic markers even though they were not cloned initially on the same fragment (Tyson et al., 2004; Venter et al., 2004).

Identifying active clones

Selection for antibiotic resistance led to the isolation of a tetracycline resistance determinant from samples of the oral microbiota of human (Díaz-Torres et al., 2003) and aminoglycoside resistance determinants from soil (Riesenfeld...
et al., 1998). The selection for aminoglyco-side resistance identified nine clones, six of which encoded 6'-acetyl transferases that formed a new cluster based on sequence analysis. These genes were discovered in libraries containing a total of 4 Gb of DNA, or approximately 1 million genes, and thus their infrequent representation would have made it prohibitively laborious to discover them by a screen without a selection. This review illustrates the power of functional metagenomics—genes that are expressed in an ordinary host such as E. coli may be extraordinary and novel. High-throughput screens can substitute when the functions of interest do not provide the basis for selection. For example, on certain indicator media, active clones display a characteristic and easily distinguishable appearance even when plated at high density. With the indicator dye tetrazolium chloride (Henne et al., 1999) detected clones that utilize 4-hydroxybutyrate in libraries of DNA from agricultural or river valley soil. Very rare lipolytic clones in the same libraries were detected by production of clear halos on media containing rhodamine and either triolein or tributyrin (Henne et al., 2000).

Conclusion

Metagenomics approach can be used to address the challenges of studying prokaryotes in the environment that are, as yet, unculturable and represent even more than 99% of the organisms in some environments (Amann et al., 1995). This approach has already opened new avenues of research by enabling unprecedented analyses of genome heterogeneity and evolution in environmental contexts and providing access to far more microbial diversity than that has been viewed in the petri dish.

Approaches that are enriched for a portion of the microbial community or for a collection of metagenomic clones will enhance the power of metagenomic analysis to address targeted questions in microbial ecology and to discover new biotechnological applications. To realize the full potential of metagenomics, however, a number of obstacles need to be overcome. Perhaps the most significant of these obstacles is the microbial complexity in most communities. Another factor for improvement in metagenomics is the use of robust sampling and DNA-extraction processes.

ACKNOWLEDGEMENTS

We are very thankful to the Dean of Faculty of Medicine, Vice-President and President of this university for providing us all the facilities for writing this article and also for their vibrant. We acknowledge Dr. Syed Jamal Mohammed highly for giving his valuable suggestions in editing and correcting the text of this manuscript.

REFERENCES


