Biotechnology and Molecular Biology Reviews Vol. 4 (3), pp. 049-054, June, 2009
Available online at http://www.academicjournals.org/BMBR
ISSN 1538-2273 © 2009 Academic Journals

Standard Review

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

Md. Zeyaullah[1]*, Majid R. Kamli[2], Badrul Islam[1], Mohammed Atif[1], Faheem A Benkhayal[1], M. Nehal[3], M.A. Rizvi[4] and Arif Ali[2]

[1]Department of Microbiology, Faculty of Medicine, Omar Al-Mukhtar University, Al-Baida, Libya.
[2]Department of Biotechnology, Jamia Millia Islamia, New Delhi- 110025, India.
[3]Ageing 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

*Corresponding author. E. mail: zeya786@gmail.com.

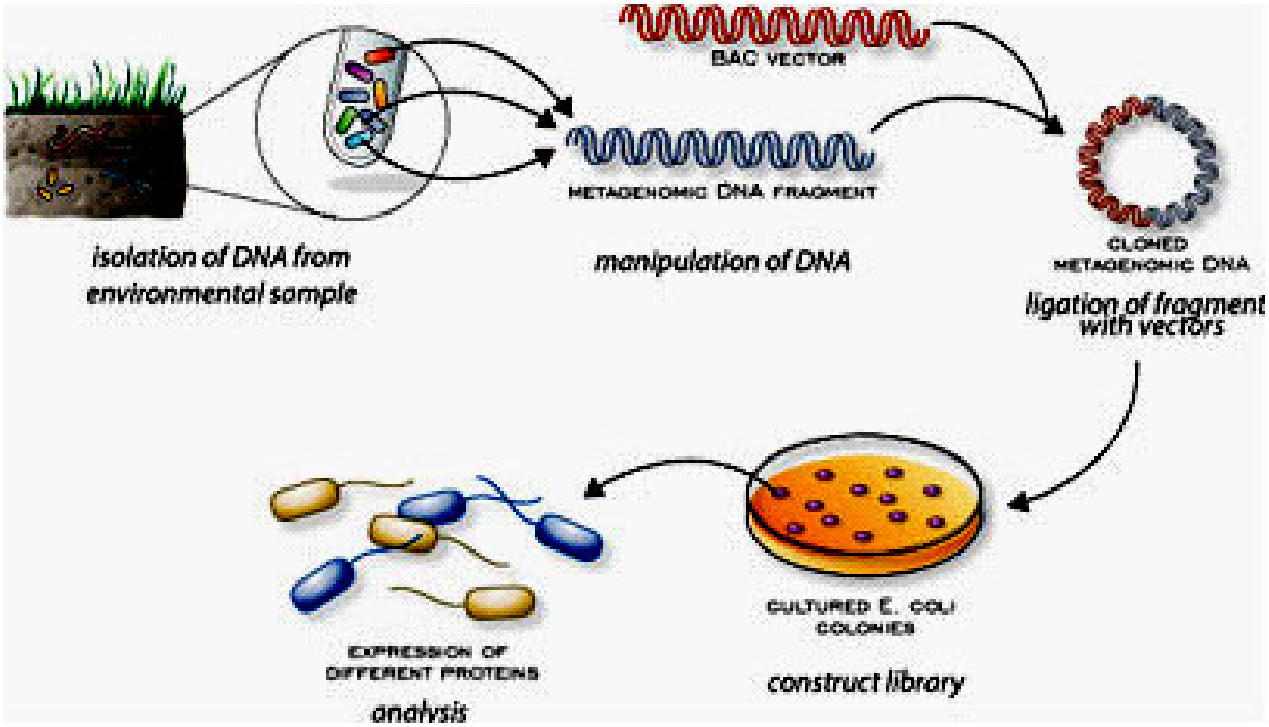

Figure 1. Steps involved in a metagenomics experiment.

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

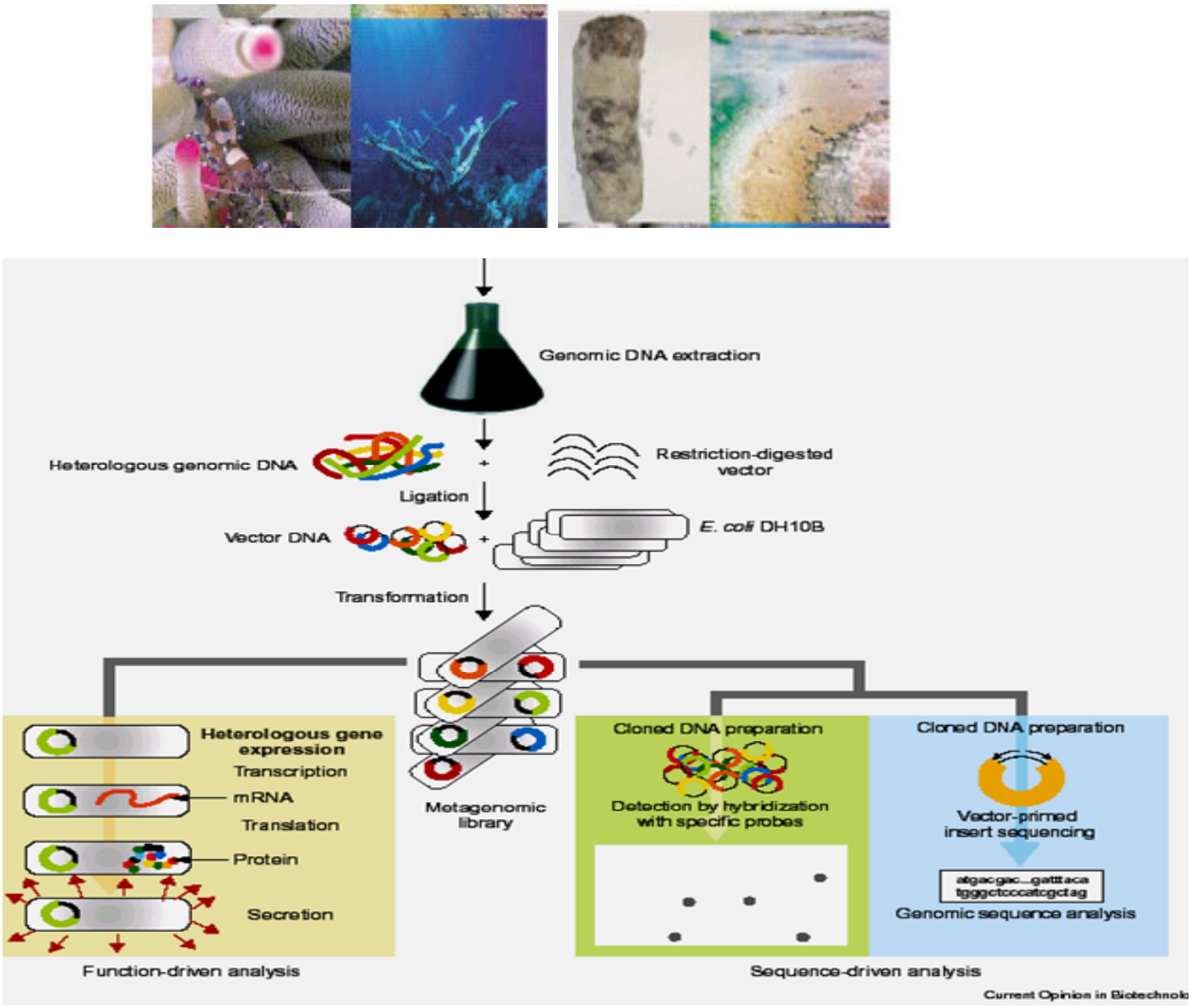

Figure 2. Construction and screening of metagenomic libraries. Schematic representation of construction of libraries from environmental samples. The images at the top from left to right show bacterial mats at Yellowstone, soil from a boreal forest in Alaska, cabbage white butterfly larvae, and a tube 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.,

2001; Majernik et al., 2001; Rondon et al., 2000; Schloss et al., 2003) or they can be sequenced randomly (Tyson et al., 2004; Venter et al., 2004). Each approach has strengths and limitations, together these approaches have enriched our understanding of the uncultured world, providing insight into groups of prokaryotes that are otherwise entirely unknown.

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-screens, selections, and functional anchors

The frequency of metagenomic clones that express any given activity is low. For example, in a search for lipolytic clones derived from German soil, only 1 in 730,000 clones showed activity (Henne et al., 2000). In a library of DNA from North American soil, 29 of a total of 25,000 clones expressed hemolytic activity (Rondon et al., 2000). The scarcity of active clones therefore necessitates development of efficient screens and selections for discovery of new activities or molecules. Just as bacterial genetics relies on selections to detect low-frequency events, metagenomics will be advanced by seeking selectable phenotypes to increase the collection of active clones that can be compared, analyzed, and used to build a conceptual framework for functional analysis. Several selections have proved to be fruitful. For example, the Daniel group designed a CVclever selection for Na+(Li+)/H+ antiporters that requires complementation of an *E. coli* mutant deficient in the three Na+/H+ antiporters (nhaA, nhaB and chaA) enabling growth on medium containing 7.5 mM LiCl (Majernik et al., 2001). This powerful selection facilitated the discovery of two novel antiporter proteins in a library of 1,480,000 clones containing DNA isolated from soil. Another selection strategy involved complementation of an *E. coli* mutant deficient in biotin production, which led to the isolation of seven new operons for biotin synthesis from enrichment cultures derived from samples of soil or horse excrement (Entcheva et al., 2001).

Selection for antibiotic resistance led to the isolation of a tetracycline resistance determinant from samples of the oral microbiota of human (Diaz-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'-acetyltransferases that formed a new cluster based on sequence analysis. These genes were disco-vered 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 tributryin (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 focus 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

Allen EE, Banfield JF (2005). Community genomics in microbial ecology and evolution. Nat. Rev. Microbiol. 3: 489-498.

Amann RI, Ludwig W, Schleifer KH (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. 59: 143-169.

Barns SM, Takala SL, Kuske CR (1999). Wide distribution and diversity of members of the bacterial kingdom Acidobacterium in the environment. Appl. Environ. Microbiol. 65: 1731-1737.

Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB, Gates CM, Feldman RA, Spudich JL, Spudich EN, DeLong EF (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Sci. 289: 1902-1906.

Beja O, Spudich EN, Spudich JL, Leclerc M, DeLong EF (2001). Proteorhodopsin phototrophy in the ocean. Nature 411: 786-789.

Buckley DH, Schmidt TM (2003). Diversity and dynamics of microbial communities in soils from agro-ecosystems. Environ. Microbiol. 5: 441-452.

Courtois S (2003). Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Appl. Environ. Microbiol. 69: 49-55.

Covacci A, Kennedy GC, Cormack B, Rappuoli R, Falkow S (1997). From microbial genomics to meta-genomics. Drug Dev. Res. 41: 180-192.

Diaz-Torres ML, McNab RD, Spratt A, Villedieu A, Hunt NM, Wilson and Mullany P (2003). Novel tetracycline resistance determinant from the oral metagenome. Antimicrob. Agents Chemother. 47: 1430-1432.

Entcheva P, Liebl W, Johann A, Hartsch T, Streit WR (2001). Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. Appl. Environ. Microbiol. 67: 89-99.

Gabor E, Liebeton K, Niehaus F, Eck J, Lorenz P (2007). Updating the metagenomics toolbox. Biotechnol. J. 2(2): 201-206.

Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy JR, Goodman M, Handelsman J (2002). Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. Appl. Environ. Microbiol. 68: 4301-4306.

Glass GV (1976). Primary, secondary and meta-analysis of research. Educ. Res. 5: 3-8.

He JZ, Zhu YG, Zheng YM, Zhang LM, Shen JP (2007). Methodology and application of soil metagenomics. Chinese Academy of Science (Article in Chinese) 18(1): 212-8.

Healy FG, Ray RM, Aldrich HC, Wilkie AC, ILO, Shanmugam KT (1995). Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. Appl. Microbiol. Biotechnol. 43: 667-674.

Henne A, Daniel R, Schmitz RA, Gottschalk G (1999). Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4- hydroxybutyrate. Appl. Environ. Microbiol. pp. 3901-3907.

Henne A, Schmitz RA, Bomeke M, Gottschalk G, Daniel R (2000). Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. Appl. Environ. Microbiol. 66: 3113-3116.

Knietsch A, Bowien S Whited G, Gottschalk G, Daniel R (2003). 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. 69: 3048-3060.

Liles MR, Manske BF, Bintrim SB, Handelsman J, Goodman RM (2003). A census of rRNA genes and linked genomic sequences within a soil metagenomic library. Appl. Environ. Microbiol. 69: 2684-2691.

Lorenz P, Liebeton K, Niehaus F, Eck J (2002). Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. Curr. Opin. Biotechnol. 13: 572-577.

Lorenz P, Schleper C (2002). Metagenome - a challenging source of enzyme discovery. J. Mol. Catal. B Enzym. 19: 13-19.

MacNeil IA, Tiong CL, Minor C, August PR, Grossman TH, Loiacono KA, Lynch BA, Phillips T, Narula S, Sundaramoorthi R, Tyler A,

Aldredge T, Long H, Gilman M, Holt D, Osburne MS (2001). Expression and isolation of antimicrobial small molecules from soil DNA libraries. J. Mol. Microbiol. Biotechnol. 3: 301-308.
Majernik A, Gottschalk G, Daniel R (2001). Screening of environmental DNA libraries for the presence of genes conferring Na_(Li_)/H_antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. J. Bacteriol. 183: 6645-6653.
Miller DN, Bryant JE, Madsen EL, Ghiorse WC (1999). Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. Appl. Environ. Microbiol. 65: 4715-4724.
Pace NR, Stahl DA, Lane DJ, Olsen GJ (1985). Analyzing natural microbial populations by rRNA sequences. ASM News. 51: 4-12.
Patrick D, Schloss, Handelsman Jo (2005). Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. Genome Biol. 6: 229.
Quaiser A, Ochsenreiter T, Lanz C, Schuster SC, Treusch AH, Eck J, Schleper C (2003). Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. Mol. Microbiol. 50: 563-575.
Riesenfeld CS, Goodman RM, Handelsman J (1998). Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. Environ. Microbiol. 6: 981-989.
Riesenfeld CS, Goodman RM, Handelsman J (1999). Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. Environ. Microbiol. 6: 981-989.
Riesenfeld CS, Schloss PD, Handelsman J (2004). Metagenomics: genomic analysis of microbial communities. Annu. Rev. Genet. 38: 525-552.
Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J, Goodman RM (2000). Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. 66: 2541-2547.
Schloss PD, Handelsman J (2003). Biotechnological prospects from metagenomics. Curr. Opin. Biotechnol. 14: 303-310.
Schmidt TM, DeLong EF Pace NR (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J. Bacteriol. 173: 4371-4378.
Stein JL, Marsh TL, Wu KY, Shizuya H, DeLong EF (1996). Characterization of uncultivated prokaryotes: isolation and analysis of a 40- kilobase-pair genome fragment front a planktonic marine archaeon. J. Bacteriol. 178: 591-599.
Susannah Green Tringe, Edward M Rubin (2005). Metagenomics: DNA sequencing of environmental samples. Nature Rev. Gen. 6: 805-814.
Tringe SG, Von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC (2005). Comparative metagenomics of microbial communities. Science 308: 554-557.
Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428: 37-43.
Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science 304: 66-74.
Waksman SA, Curtis RE (1916). The actinomyces of the soil. Soil Sci. 1: 99-134.
Waksman SA, Foster JW (1937). Associative and antagonistic effects of microorganisms: Antagonistic effects of microorganisms grown on artificial substrates. Soil Sci. 43: 69-76.
Whitaker RJ, Banfield JF (2006). Population genomics in natural microbial communities. Trends Ecol. Evol. 21(9): 508-16.