Indian J. Microbiol. (June 2008) 48:202–215

REVIEW

Metagenomics: Future of microbial gene mining

J. Vakhlu · Avneet Kour Sudan · B. N. Johri

Received: 26 May 2007 / Accepted: 7 April 2008

Abstract Modern biotechnology has a steadily increasing demand for novel genes for application in various industrial processes and development of genetically modified organisms. Identification, isolation and cloning for novel genes at a reasonable pace is the main driving force behind the development of unprecedented experimental approaches. Metagenomics is one such novel approach for engendering novel genes. Metagenomics of complex microbial communities (both cultivable and uncultivable) is a rich source of novel genes for biotechnological purposes. The contributions made by metagenomics to the already existing repository of prokaryotic genes is quite impressive but nevertheless, this technique is still in its infancy. In the present review we have drawn comparison between routine cloning techniques and metagenomic approach for harvesting novel microbial genes and described various methods to reach down to the specific genes in the metagenome. Accomplishments made thus far, limitations and future prospects of this resourceful technique are discussed.

Keywords Metagenomics · Uncultivable · Microbes · Novel genes

J. Vakhlu[1] (✉) · A. K. Sudan[1] · B. N. Johri[2]
[1]Department of Biotechnology,
University of Jammu,
Jammu - 180 006,
India
[2]Department of Biotechnology and bioinformatics centre,
Barkatutullah University,
Bhopal - 462 026,
India
e-mail: jyotivakhlu@gmail.com

Introduction

Microbes underpin all life processes on earth. Being pioneer colonizers of this planet they have evolved and accumulated remarkable physiological and functional diversity. It would not be far-fetched to say that microbes are 'micro biotechnology' units designed by nature millions of years ago, much before we discovered their influence in every sphere of human life. Today, they represent the chief source of most biocatalysts, antibiotics and secondary metabolites. The traditional method for isolating novel microbes is by cultivating them randomly and subsequently screening for pure strains. This method is not only cumbersome but also suffers strong limitation as has been revealed by molecular ecological studies. It is now common knowledge that the microbial diversity retrieved by standard cultivation techniques represents less than 1% of diversity present in nature [1–9]. This huge gap between actual microbial abundance and the total recovery made so far using standard culturing techniques can be bridged by metagenomics. The significance of the field is evident by the number (12) of reviews published in last five years on varied aspects of metagenomics (Table 1) .

In the first half of last century the structure and function of the all powerful molecule on the earth *i.e.* DNA; was not elucidated. Therefore, the characterization of all living beings, both macro- and microscopic, was done on the basis of phenotypic characters. To characterize microbes, culturing was the only way. Bergey's Manual (1923) categorically states that no microorganisms can be characterized without being cultured [2]. But now we have come to realize that there are many bacteria which are not amenable to culturing however this does not disprove their existence. Some glaring impediments in cultivating microorganisms are lack

Table 1 Reviews published in last few years on metagenomics

S.No.	Name	Authors	Name of the Journal	Impact Factor of the journal	Year
1.	Metagenomics, Biotechnology with non-culturable microbes	Schmeisser, C, Steele H and Streit W [95]	Applied microbiology and Biotechnology	2.85	2007
2.	Metagenomics: Access to uncultuable microbes in the environment.	Kimura, N [8]	Microbes and Environment	2.411	2006
3.	New direction and interaction in metagenomic research.	Ward, N [9]	FEMS Microbiol Ecol	3.157	2006
4.	Metagenomics gene discovery : past, present and future.	Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R and Wittwer P [6]	Trends in Biotechnology	7.955	2005
5.	Diversity of the human intestinal microbial flora.	Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, and Relman DA [90]	Science	21.911	2005
6.	Metagenomics and industrial applications.	Lorenz, P and Eck, J. [91]	Nature	27.074	2005
7.	The soil metagenome- a rich source for the discovery of novel natural products.	Daniel R [4]	Current opinion in Biotechnology	6.89	2004
8.	Microbial population genomics and ecology.	Delong EF [92]	Environmental Microbiology	3.818	2004
9.	Metagenomics, gene discovery and ideal biocatalyst.	Cowan DA, Arslanoglu A, Burton SG, Baker GC, Cameron RA, Smitch JJ and Meyer Q [22]	Biochemical Society Transaction	2.962	2004
10.	To BAC or not BAC: marine ecogenomics.	Beja O [93]	Current opinion in Biotechnology	6.89	2004
11.	Metagenomics; application of genomics to uncultured microorganisms.	Handelsman J [60]	Microbiology Molecular Biology Review	15.500	2004
12.	Microbial genomes- the untrapped resource.	Cowan DA [96]	Trends in Biotechnology	7.955	2004
13.	Exploring prokaryotics diversity in the genomics era.	Hugenholtz P [94]	Genome Biology	7.71	2002

of necessary symboint, nutrient or surface, excess of inhibitory compounds, incorrect combination of temperature, pressure or atmospheric gas composition, accumulation of toxic waste products from their own metabolism in the culture medium and intrinsically slow growth rate or rapid dispersion from colonies [10–11].

Torsvik (1978) and Pace (1986) had individually put forth an idea that the genomes of microorganisms can be archived and characterized without necessarily cultivating them first [12–14]. Pace (1986) and his colleagues used direct analysis of 5S and 16S rRNA gene sequences directly from environment to describe the diversity of microorganism in a particular environment [15]. Looking back at events now, one wonders why this method of microbial genome analysis took so long to materialize. Once standardized, the use of genome technology to characterize naturally occurring microbes (both cultivable and uncultivable) has gained wide spread acceptance. The foundation of metagenomics rests on several different technologies and disciplines which include genomics, ecology, evolution, high through-put DNA sequencing, microarray technology and bioinformatics. In literature there are several synonyms to the metagenomic cloning approach such as, Soil DNA libraries, eDNA library, Microbial population genomics, Recombinant environmental libraries, Community genomics, Whole genome shotgun sequencing, Environmental genomics and Ecogenomics. The term metagenomics was coined by Handelsman in 1998**,** and has been defined as culture independent genomic analysis of microbial communities [16]. Few other definitions of metagenomics based on

the techniques that are used or the aim of this young budding field have also been suggested by other workers;

Definition I Leveau and his group suggest that "Metagenomic libraries are databases of bacterial clones, usually *E.coli* carrying DNA fragments that originate from collective genomes of all the organisms present in the particular environment, habitat and assemblage" [17].

Definition II(a) "Metagenomics describes the functional and sequence based analysis of collective microbial genomes contained in an environmental samples" [18].

Definition II(b) " The metagenome approach is the culture independent genomic analysis of microbial communities in the environment" [18].

As isolation of nucleic acids from environmental samples is not that straight forward, therefore many researchers have resorted to analyses of pooled genome of cultivable microbes from environmental sample/s as well [19–21]. The basic definition therefore, can be modified to "collective genome analysis of pooled genomes of cultivable and/or uncultivable microorganisms" in agreement with the definition IIa given by the Riesenfeld and his group.

The present review concerns with advantages of metagenomics over routine gene mining techniques, strategies for harvesting novel genes from pooled genomes, achievements made so far, limitations and future prospects of the youthful discipline of metagenomics.

Metagenomics versus routine microbial gene cloning

Metagenomics is not the only way to harvest unknown genes as there are other techniques to engender novel genes. In fact developments in protein engineering and evolution have provided methods not only for isolating but also for generating new genes. The classical technique of random and site specific mutagenesis can be used to replace a single or limited number of amino acids to produce a new gene-peptide sequence. More recently other methods like error–prone PCR, saturation mutagenesis and domain shuffling give access to more varied probable sequences of genes/peptides under study [22]. The protein sequences available to researchers are from direct sequencing of the protein, from the translation of directly cloned genes and from sequenced genomes. Our understanding of the occupancy of protein sequence space is driven from the sequences available in protein and nucleic acid data banks. The sequences are in turn dependent upon the range of the known and available genes/genomes. In order to move in "sequence space" and expand genes for invitro evolution, new technologies that circumvent the limitations of microbial culturability must be applied. These technologies which currently include metagenomic library screening, gene specific amplification methods and even full metagenomic sequencing provide access to the volume of "sequence space" that is not addressed by traditional screening [22, 6, 1].

The pronouncement that 99.9 % of the microbial species represented in any biotope are not culturable at the moment, highlights the limitation of any gene discovery protocol that is dependent on culturing. However metagenomics is direct route to novel gene cloning as it is not dependent on culturing. Finding a novel gene or functional sequence space within reasonable time depends as much on efficient/sensitive screening strategies as on input of diverse candidate genomes, which can be made available by following metagenomic approach. The major advantages of metagenomics are:

1. Much higher proportion of genomes present in any environmental samples can be collected in comparison to genomes obtained through classical microbial isolation,
2. the target gene isolation is matter of ingenuity and the design of the assay system and,
3. searching for target genes in a pool of genomes of cultivable and/ or uncultivable microbes will result in increased probability of finding target gene (than through analyzing individual microbes).

Although our present capacity to produce genomic information (in particular sequence related) surpasses our ability to analyze, interpret and utilize it. Development of effective analytical bioinformatic tools and ingenuous assay systems will propagate the road ahead, for this approach which is still in its infancy. Nevertheless studies on the genome characterization of uncultivated microbes, gene inventories of diverse microbial assemblages, functional analysis of novel proteins and process and quantitative gene surveys all are now increasing both in quality and quantity.

Strategies for reaching out to microbial metagenome

(a) Source metagenome and enrichment techniques. Target gene isolation can be started with selecting the right target environmental site as it will increase the probability of finding novel target gene. Fig. 1 gives the overview of various routes to identify, isolate and clone novel microbial genes. It is difficult to presume the abundance of target genes in the any metagenome, but usually target gene/s represent a small proportion of the total nucleic acid fraction. One of the straight forward routes is to pre- enrich the target environmental site for specific gene encoding microbes. Venter and his associates (2004) while undertaking Sargasso Sea genome sequencing project did size selective filteration to effectively remove eukaryotic population so

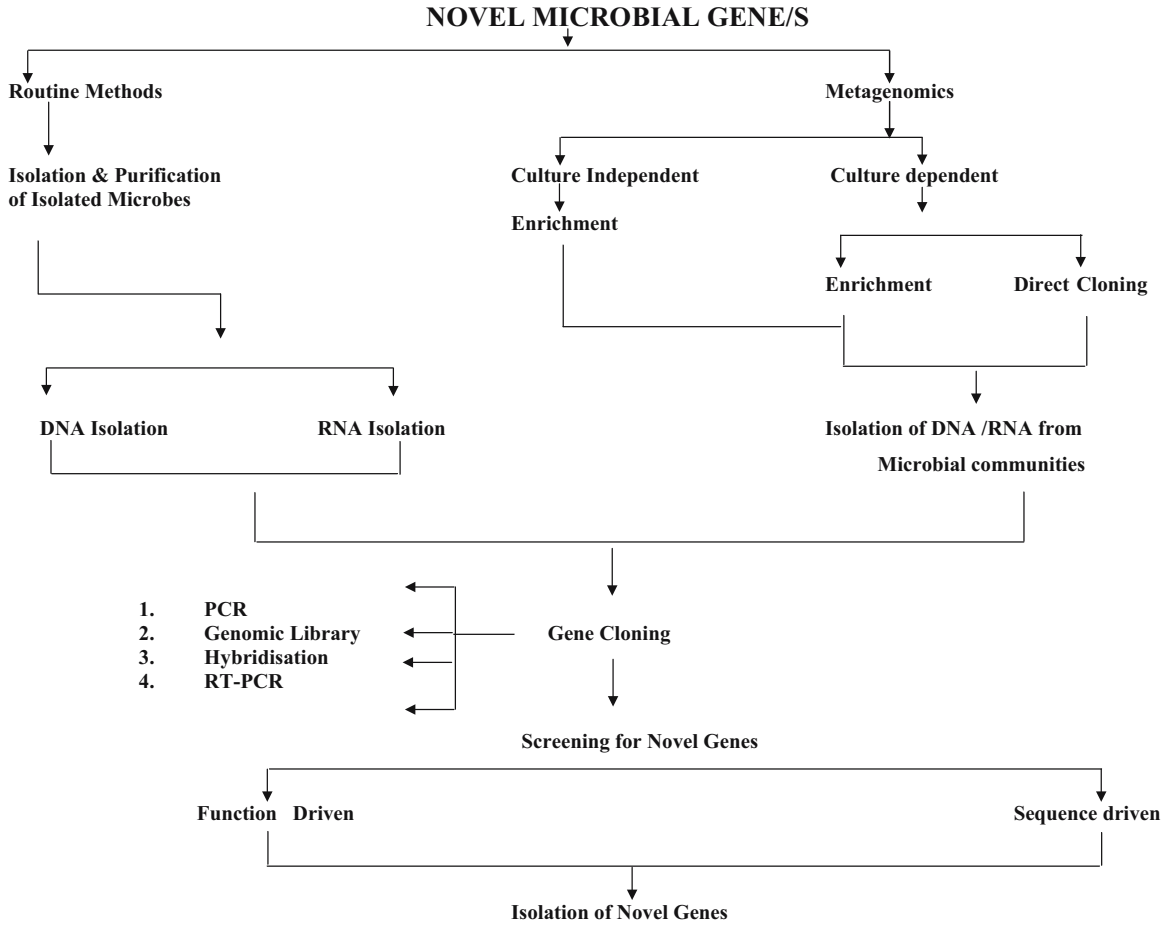

Fig. 1 Polyphasic approaches for isolation of novel microbial genes

as to enrich target prokaryotic, archea population [23]. The first complete genome sequence obtained from uncultivable bacterium *Buchnera aphidicola* was enriched for bacterial genome by differential centrifugation for removing insect tissue DNA[16]. Culture enrichment for a target gene can be done by exposing the cells to the selective substrate. The selection pressure for enrichment can be nutritional, physical and chemical, although substrate utilization is most commonly employed [19–20]. Culture enrichment on carboxymethylcellulose resulted in four fold enrichment of cellulose genes in a small insert expression library [24]. Stating the obvious, culture enrichment will inevitably lead to loss of some microbial diversity by selecting fast growing specific culturable species but it is straight forward if we are looking for a specific genes and not the overall diversity. One alternative is to introduce a mild selection pressure so that presence of non enriched cultures is not lost totally. Enrichment of microorganisms with special traits and construction of metagenomic library by direct cloning of environmental DNA has great potential for identifying genes and gene products for biotechnological purposes. Various types of enrichment techniques like stable isotope probing (SIP) and 5-Bromo–2-deoxyuridine labelling, suppressive subtractive hybridization (SSH), differential expression analysis (DEA), phage display and affinity capture have been developed to increase the possibility of hitting the target gene [6].

b) Community nucleic acid isolation. The direct isolation of microbial DNA from the environment for the sake of unbiased genomic representation was proposed and pioneered by Torsvik (1978) and Pace (1986) along with their group as mentioned earlier [12, 15, 25]. Pace and his group extracted total DNA from marine plankton and partially digested it to obtain ligation competent fragments of suitable sizes. Subsequently it was cloned in *E.coli* using bacteriophage λ vector and screening for genes encoding rRNAs was done thereby illustrating the potential of isolating any kind of gene from un cultivable organisms. Since then numerous studies have dealt with methods extracting 'metagenome' from a variety of different environments. Extraction methods fall into two categories depending on whether bacteria are lysed directly in the context of sub statum (*in situ*) or

are first separated from surrounding material. Balance has to be struck between high DNA recovery yields and ease of operation on one hand and superior DNA integrity and purity on the other hand [26–28].

As in any DNA isolation method the first step is breaking open the cell and isolating the DNA. The procedure to break open the cell is either to shear it mechanically or chemically. Both methods have there own advantages and limitations. Chemical treatment of the sample (cells) is a gentle method and results in recovery of higher molecular weight DNA but chemical lysis can select certain species only by exploiting their biochemical characteristics. Mechanical shearing on the other hand does not show any such bias and is known to recover nucleic acid from more diverse cells as compared to the chemical method, however quality of DNA is not so good. Therefore microbial (metagenomic) community DNA isolation at present is a compromise between vigorous extraction required for the representation of all microbial genomes and minimization of DNA shearing and the inhibition of co-extraction of organic acid contaminants [29–30, 6, 8]. Despite various types of metagenomic DNA isolation protocols available there is no sure shot way, by which, all microbial community nucleic acid representation in the extract can be ensured.

c) Normalization of the metagenome. Total nucleic acids directly extracted from the environmental samples will not represent all the genomes equally as it will directly depend upon the relative abundance of the each type of population in the environment [31]. Environmental metagenomes can be normalized artificially in the laboratory. Separation of genotypes (specific genomes) is achieved by Cesium Chloride density centrifugation in presence of intercalating agents such as bis-benzimide, for buoyant density separation of genomes based on their % G and C content. Equal amounts of the each band on the gradient are combined to represent a normalized metagenome. Normalization can also be done by denaturing fragmented genomic DNA and reannealing under stringent conditions. Abundant ssDNA will anneal more rapidly to generate double-stranded nucleic acids than rare DNA species. Single stranded sequences are then separated from the double–stranded nucleic acids, resulting in an enrichment of rarer sequences within the environmental sample [32].

Recently Yokouchi and his group (2006) have devised a strategy called multiple displacement amplification (MDA) for whole- genome amplification (WGA) to overcome these limitations. This method can generate large quantities of high quality DNA from a small amount of sample using φ29 DNA polymerase and random exonuclease - resistant primers. Authors have demonstrated that employment of this method can ensure representation of all the genomes isolated in the metagenome [33].

Stratagem for sifting genes from environmental metagenome

Once the source metagenome is identified and metagenomic DNA isolated the goal shifts to isolation of gene of interest. There can be two approaches, one is to clone all the metagenome isolated by making genomic libraries first and subsequently screening it for specific gene of interest. The alternate approach is to specifically isolate the gene of interest directly by performing a polymerase chain reaction. The first approach includes metagenomic /metatranscriptomic library construction.

A. i) Metagenomic library: The basic steps in the construction of a metagenomic library and genomic library are broadly similar and it seems now so surprising as to why this technique took so long to come alive. Metagenomic library is constructed by using routine cloning and expression vectors but the choice usually is to use cosmid or bacterial artificial chromosome (BAC) as they can accommodate large sized DNA fragment. Cloning in either cosmid or BAC allows entire functional operons to express together. This way operons for several multigenic pathways have been successfully cloned and isolated [6]. Phage display expression metagenomic libraries are made by affinity selection of the surface–displayed expression product. Phage display is yet another method to enrich even rare DNA present in metgenome but the only limitation is that the protein size upper limit is 50 kD [34]. The construction of genomic libraries or gene banks is one of the pioneering experiments of genetic engineering but metagenomic libraries were constructed only in 1990's [35–40, 19–20, 23]. In the last two and a half year since the birth of this field large number of metagenomic libraries have been constructed and the number of novel genes isolated from them is equally impressive (Table 2).

A.ii) Metatranscriptomic libraries: Construction of a metatranscriptomic library requires the isolation of RNA from environmental samples. Protocols are modified to take care to minimize single- stranded polynucleotide degradation, to limit physical degradation and RNAse activity as they are the major cause of yield loss [41–44]. Metagenomic libraries that contain eukaryotic genes will not be suitable for mining eukaryotic genes, due to presence of the introns. This problem can be solved but to a limited extent by large scale sequencing of clones. The other alternative is to make metatranscriptomic libraries. Though there are inherent difficulties in the isolation of mRNA and cDNA library construction but the potential of direct expression of uncultivable eukaryotic genes is tremendous. Metatranscriptomic library unlike metagenomic libraries will not represent non-expressed DNA. The additional hurdle are limits of

Table 2 Various genes cloned following technique of metagenomics

S.No	Name of Gene	Vector/ Host	E/N	Methodology	Site	References
1	Chitinase	*E.coli*/Bacteriophage/ phagemid λ ZAP III	N	Genomic library & functional assay	1. Coastal sea water from bay estuary	Cottrell et al 1999 [76]
2	PKS I	*E.coli/ Streptomyces lividans/* plasmid / cosmid shuttle vector	N	Cosmid library & PCR amplification	1. arable field la cote saint Andre(France)	Courtois et al 2003 [36]
3	DNase anti bacterial Comp. 2 lipase	*E.coli DHIOB* /BAC Bacillus	-	BAC library	West Madison Agriculture in Madison	Rondon et al 2000 [39]
4	Glycerol dehydratase Diol dehydratase	*E.coli* / Plasmid	E	PCR Based amplification (sequence driven) Screening of genomic library constructed.	a) Sediment if river grone (Germany) b) Sugar beet field gothein –gon (Germany) c) Sediment soil solar lake (Egypt)	Knietsch et al 2003 [20]
5	Type I polyketide synthase	-	-	Genomic library & functional Screening	Soil	Osburnis and his group cited from Handelsman 2004 [60]
6	Oxido- reductases	*E.coli DH5ά* / plasmid pblue script Sk^+	N	Genomic library & functional Screening	Sediment 1. River Grone 2. Sugar beet field (Germany) 3. Solar lake (Egypt) 4. Sediments from gulf Eilol (Israel)	Knietsch et al 2003 [20]
7	Amylase	*E.coli* /cosmid, BAC	-	Genomic library	Soil	Rondon et al 2000 [39]; Richardson et al 2002 [77];Voget et al 2003 [40], Ward 2006 [9]
8	Violacein, Deoxyviolacein	*E.coli* /cosmid	-	Genomic library	Soil	Brady et al 2002 [78]
9	Amidase	*E.coli* /cosmid		Genomic library	Soil	Knietsch et al 2003 [20]
10	Terragine A	Streptomycete/ *plasmid*	-	Genomic library functional screen	Soil metagenome	Wang et al 2000 [79]
11	Long-chain N-acyl amino acid antibiotic	*E.coli*/plasmid	-	Genomic library	soil	Brady et al 2002 [80]
12	Acyl.tyrosine	*E.col*i	-	Genomic library	Soil	Brady et al 2000 [80], 2001 [81]
13	Type II polyketide antibiotic			PCR amplification and cloning	Soil	Seow et al 1997 [83]
14	Turbomycin	*E.coli*/BAC	-	Genomic library	Soil	Gillespie et al 2002 [84]
15	Indirubin	*E.coli*/BAC	-	Genomic library	Soil	MacNeil et al 2001 [85]
16	Bacteriorhodpsin, Photeorhodpsin		N	Genomic library	Ocean	Beja *et al* 2000 [38], 2001 [69]
17	Biotin	*E. Coli* / plasmid	E.	Genomic library and functional screening	Forest Soil Agricultural site Horse excreme meadow Beach- Germany Naent hood in oregon	Entcheva et al 2001 [19]

Table 2 *(Continued)*

18	Agarase	*E.coli* /cosmid	-	Genomic library	soil	Voget et al 2003 [40]
19	4-hydroxy butyrate dehydrogenase	*E. coli*/ plasmid	-			Henne et al 1999 [35]
20	Lipolytic activity	*E. coli* DH5ά/ pblue script Sk^+	N	Environmental libraries and functional screening	1. Meadow near north Germany 2. Sugar beet field Germany. Mieme river Germany.	Henne et al 2000 [37] Rondon et al 2000 [39]; Bell et al 2002 [84]; Voget et al 2003 [40] Rhee et al 2005 [85], Jiang et al 2006 [86]· Hardeman & Sjoling 2007 [97]
21	Cellulose	*E. coli*/cosmid	E	Genomic library	Microbes enriched on dried grassess	Healey et al 1995 [24]; Voget et al 2003 [40]
22	Esterase	*E. coli*/plasmid	N	Genomic library	Soil	Henne et al 2000 [37], Elend et al 2006 [87]
26	Nitrilase	*E. coli*/λ-*ZAP*	-	Genomic library	Various samples	Robertson et al 2004 [88]
27.	Pectate lyase	*E. coli*/λ-*ZAP*	-	Genomic library	Soil	Voget et al 2003 [40]
28	Xylanase	*E. coli*/λ-*ZAP*	-	Genomic library	Insert gut	Brennam et al 2004 [90]

E stands enriched

N stands for non enriched

amplification and limits of the size of insert in cDNA while making metatranscriptomic libraries [45, 38].

B) Polymerase chain reaction based gene/s retrieval. The alternate PCR based approach for novel gene retrieval from metagenome isolates can be categorized into two major classes; PCR based on specific gene amplification technique and integron based gene capture technique.

B i) Gene- specific PCR based techniques: Gene- specific PCR has been used extensively to isolate specific nucleic acids from community DNA. Instead of cloning whole isolated metagenomes, primers are designed specifically against an identified target gene. Methane mono-oxygenase gene, 2, 3-dioxygenase, chlorocatechol dioxygenase and phenol hydroxylase are a few examples of the genes that have been isolated using this PCR based technique from the metagenomic DNA [46–50]. Other genes that also have been isolated by sequence based approach using PCR are Glycerol dehydatrase, diol dehydratase [20, 36] and type II Polyketide antibiotic [51]. RT-PCR has also been used to recover genes from environmental samples. Napthalene deoxygenase enzymes encoding genes from micro-organisms present in coal tar waste has been isolated by using RT-PCR by Wilson and his group (1999) [52]. The PCR based on gene specific primers has two major draw backs;

1. Primers designed for gene targeting are dependent on existing sequence information therefore, the bar dips in favour of known sequence types and ending up logically in not very diverse (novel) sequences, and,
2. only a fragment of main target gene will be amplified and we will have to resort to fishing out the fulllength gene from main metagenome / metatranscriptome.

To access fulllength gene alternative PCR based strategies like universal fast walking, panhandle PCR, random primed PCR, inverse PCR and adaptor ligation PCR are to be used.

B.ii) Integron gene capture technique: The genomic era has clearly indicated that a large proportion of bacterial genes have been acquired by horizontal gene transfer [53]. Horizontal gene transfer is facilitated by a number of genetic elements in bacteria including plasmids, transposons and integrons. Traditionally, attention was being focused on plasmids and transposons. Integrons have been recently demonstrated to be present on the chromosome of diverse bacteria [54]. Thus there is the opportunity to recover a significant proportion of undiscovered bacterial gene pool, not by targeting gene sequences themselves but rather by targeting conserved sequences associated with these mobile elements [55–58]. Integrons are naturally occurring gene capture, dissemination and expression systems that have, until recently been associated with antibiotic resistant and pathogenic bacteria. An integron includes a gene cassette integration site (att I), and int I gene that encodes an integrase. It also contains two promoters that drives the expression of the integrase gene and the incorporated gene cassettes. The gene cassettes are inserted into or excised from

integron by a site specific recombination reaction catalyzed by the integron integrase, Int I. Features of the integrons gene cassettes system suggest that it might provide a useful means by which intact novel genes could be recovered directly from the environment DNA by PCR without prior sequence data [54].

Metagenome analysis for gene isolation

The analysis of metagenome is done either by direct analysis of the isolated sequence i.e sequence driven approach or by analysis the isolated sequence for function i.e function driven approach [4]. Each approach has its own strengths and limitations; together these approaches have enriched our understanding of unexplorable microbial world.

a) Sequence driven approach. Sequence driven approach may involve complete sequencing of clones containing phylogenetic/conserved sequence ' anchors' (Nucleic acid probe or Primer based) that indicate the taxonomic group that is the probable source of the DNA fragment or a related gene. Random sequencing can also be conducted, and a new microbial taxa or gene of interest can be identified. 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 effort, which include reconstruction of the genomes of uncultivated communities in acid mine drainage [59] and the Sargaso sea [23] illustrate the power of large scale sequencing effort to enrich our understanding of uncultivable microflora [5].

The assessment of microbial diversity in any particular environment is done by various techniques of which the prominent ones are I) 16S rDNA gene analysis for identification of member community, ii) substrate utilization for measuring metabolic diversity, iii) DNA-DNA re-association kinetics to view genetics complexity of given sample iv) amplified- ribosomal DNA restriction gel electrophoresis (ARDRGE) v) and temperature gradient gel electrophoresis (TGGE) [16]. The scientists at the European Molecular Biology Laboratory (EMBL) have used these techniques and the result is a new appreciation for rich diversity of life that exists in most unlikely places.

b) Function- driven approach. In function driven approach, metagenomic libraries are screened for expressed traits and once identified the clones are characterize by biochemical and sequence analysis. Screening by heterologous gene expression is a powerful yet challenging approach to metagenome analysis. Faithful transcription and translation of the gene/s are required for a successful hit during screening in this approach. Despite bottlenecks many novel antibiotics, antibiotic resistance genes, ion (Na^+/H^+) transporters and degradative enzymes have isolated using function driven screen [4, 83]. As is evident, this approach is not dependent on any prior sequence data, hence is the only approach to isolate novel genes of known function and more importantly entirely new class of genes for totally new functions. Nevertheless, there is inherent contradiction in this approach - all novel genes of known and unknown function will not be expressed in any particular host but at the same time the diversity of genes isolated from metagenome by successfull expression in *E.coli* is surprising [4, 60]. The frequency of metagenomic clones that express any given activity, as expected is low and hence it is necessary that development of efficient screens and diverse host expression-vector systems are undertaken. The discovery of new biological motifs will depend in part on functional analysis of metagenomic clones. High through-put screens can substitute when the functions of interest do not provide the basis for selection. One such high through-put screen that is modification of function driven screening of metagenome is SIGEX- Substrate-Induced-Gene-Expression-Screening. This is designed to select the clones harboring catabolic genes induced by various substrates in concert with fluorescence activated cell sorting (FACs) [61]. Functional screens of metagenomic libraries have led to the assignment of functions of numerous 'hypothetical proteins' in databases. An emerging and powerful direction for metagenome analysis is use of functional anchors, which are the functional analogs of phylogenetic anchors.

Produce of metagenomics

a) Microbial diversity assessment. Two near – complete microbial genomes and partial recovery of three others from an environmental sample is a significant advancement in the study of natural microbial communities [59]. Genome reconstruction was facilitated by the fact that the community was dominated by few genomically distinct species. The effectiveness of this approach in other environments will be limited by high species richness, heterogeneities in the abundance of community members as well as by extensive genome rearrangements. However in more complex environments, it should be possible to extend random shotgun sequencing approach to recover genomes of uncultivated strains and species. These data can be used to explore the nature of the community metabolic network, to find conditions for cultivating previously uncultivable organism, to monitor community structure over time and to construct DNA microarray to monitor global community gene expression patterns [62–63]. Quaiser and co-worker (2002)

[64] studied and analyzed genome of an uncultivable *Crenarchaeote* using direct genome analysis. Authors have developed a procedure for fast and reliable purification of concentrated and clonable high molecular weight DNA that is primarily based on two phase electrophoresis technique. Analysis of genomic fragments indicated that archarea from soil differ from uncultivable marine relatives. Their work also resulted in isolation of 17 putative protein encoding genes whose analysis will give some insight into the physiological potential of these organisms.

Recent advances in soil community analysis using molecular methods agree with earlier data on total genetic diversity by indicating the enormous microbial diversity in soil [65–67]. Soil diversity exceeds that of aquatic environment and is a great resource for biotechnological exploration of novel organisms, products and processes. Studies of sequence information from organisms in soil microhabitats and their gene expression under different conditions will provide guidelines for designing new and improved culturing methods that resemble their natural niches. New tools in bioinformatics and statistical analysis enable us to handle the huge amount of data obtained through multidimentional studies that combine growth independent molecular analysis, and analysis of microbial growth activity and physiology and integrate measures of environmental parameters. Such polyphasic studies integrate different aspects of microbial diversity and provide a more complete picture of microbial diversity and a deeper understanding of the interactions in soil microbial ecosystems. Studies of microbial sequences, comparative genomics and microarray technology will improve our understanding of the structure function relationship and the effect of abiotic and biotic factors on soil microbial communities. It is conceivable that the research field dealing with the interaction of genomes with the environment will be an important topic in the future.

b) Microbial genes mined so far. Classical cultivation based methods and modern direct cloning strategies have their advantages and limitations. The natural molecular diversity found in metagenomic DNA of uncultivated microorganism is so enormous that the likelihood of retrieving known genes is very small. The access of novel natural 'sequence space' via direct cloning of metagenomic DNA could significantly contribute to the identification of valuable biocatalysts. The technology developed for human genome project vectors were combined with Oceanographic methods to produce the first environmental genomic library [68]. The environmental library constructed was a fosmid library prepared from coastal Oregon waters; and new screening methodologies like multiplex PCR was developed. This study gave science the first glimpse into genome organization and protein composition of still elusive and uncultivable marine archaea. The first metagenomics library of marine proplankton and soil were initiated in year 2000 [38–39]. In the summer of 2003 a meeting entitled 'Metagenomics of 2003' organized by Dr. Christa Schleper and colleagues at the Darmstadt Technical University, USA. It was clear from the meeting's presentations that the pace of genomic investigations in environmental microbiology and ecology is accelerating. The repertoire of tools for 'post – environmental genomics' is also expanding with microarray proteomic and metabolic applications following fresh on heels of environmental genome discoveries. Quaiser and his group (2002) identified a metagenomic DNA fragment that along with hallmark rDNA genes carries open reading frames for polyhydroxy alkanoate (PHA) synthase, glycosyltransferase and other potentially important industrial enzymes. Metagenomic library prepared by Healey and his group (1995) from a mixture of organisms enriched on dried grasses in laboratory referred to as zoo libraries yielded clones expressing cellulotytic activity. Yet another revelation made by metagenome is the presence of ocean proteobacteria when Bacteriorhodopsin like gene were cloned following metagenomic technique [36, 69]. Combined metagenomic and susceptibility testing of 8000 + clones by Song *et al* (2005) [70] revealed that the presence of antibiotic resistant genes similar in resistance profile to antibiotic resistant stains isolated during the period in which antibiotics have been widely used. This led authors to suggest that the diversity of a type of beta lactamases may reflect an ancient evolutionary origin rather than being a consequence of antibiotic overuse. The list of microbial gene mines is quite long and has been tabulated in table 2 for easy comprehension .

Since the inception of two pioneering commercial metagenomic ventures in late 1990's by recombinant Biocatalysis Ltd. of La Jolla and Terra Gen Discover Inc, these technologies have been taken up by several of the biotechnology giants and have been the focal area of several start-up companies. Recombinant biocatalysts Ltd., Now Diversa Corporation (www.diversa.com), is the acknowledged leader in the field with an impressive list of libraries derived from global biotopes and of cloned enzymes. Several other biotechnology companies appear to be competing in the same market sector. Small size of industrial enzyme market compared with the pharmaceuticals market suggests that a switch in product focus might not be unexpected [6].

c) Microbial community behavior. Metagenomics helps us to extract and sequence the complete DNA of all organisms in a single habitat and therefore so called "group specific gene patterns" can be identified. Metagenomics for assessment of the gene function in particular habitat is also known as "MetaFunctions' and the goal is to

correlate metagenome sequence data with ecological data to find group specific gene pattern in order to elucidate the function of "conserved hypothetic protein". Metagenomics so far has provided insight into the prokaryotic symbiosis with eukaryotes, how microbes compete and communicate with other organisms, acquire nutrition and provide energy [58, 71].

d) Fossil DNA analysis. Metagenomics can also be used to gather information stored in ancient bacteria which have been fossilized. There is no other way but to extract collective DNA and analyze it, though sequence analysis of ancient DNA is challenging because of the technological obstacles of degradation and contamination with modern DNA. These difficulties not withstanding *Micrococcus luteus* preserved in 120 – million-year- old amber have been identified using metagenomics [72]. The other glaring example is microbes detected in the stomach of Neolithic "Tyrolean Iceman". Along with typical members of human faecal flora genus *Vibrio* have also been identified in the stomach of neolithic man [73].

Future prospects

a) Development of bioinformatic tools. Our present ability to produce sequence data surpasses our ability to analyze and use it. In addition there has been exponential increase in the published metagenome sequences in the last 5 year [1, 7] and future explosion of Meta-sequences is anticipated. Until now metagenome sequence data was deposited in public domain i.e 'all purpose' databases but need for separate storage space, processing and data mining protocols have been realized. The"metagenomes Mapserver" is the first attempt to systematically integrate genomic and metagenomic data into consistent, curated databases including geographical and ecological databases context information resulting in Geographic Information System (GIS). The metagenome Metaserver is part of "Metafunctions" a three year project funded by the European union. The goal of this project is to elucidate the unknown functions of the "conserved hypothetical proteins" by correlating the occurrence of such proteins to habitat specific parameters. The number of new metagenomic sequences being isolated makes it necessary to develop an Information Extraction (IE) system that will be able to extract the environmental and geographical data from biological scientific literature. The result will be the fundamental backbone of the Metagenomics Mapserver and will also enable scientists to perform "comparative metagenomics" [18]. Rower and his group (2005) are actively engaged in the development of tools associated with analysis of viral metagenomic data, one such tool is PHACCS i.e. phage communities from contig spectrum [74]. Gabor et al (2004) [75] for increasing the possibility of positive hits have calculated a formulae that describe the chance of isolating a gene by random expression cloning, taking into account different modes of expression. Development of innovative bioinformatics tools is life line of this new born field of metagenomics.

b) High through-put screening assay. Most reliable tool for assigning the function to any DNA sequence today also is good old functional assay. The need of the hour is to direct evolution of these function based assays so that a specific novel characters is identified. This will reduce the time exhausted in the primary screening. The future of the specific or generalized screening lies in the miniaturization and automation. Therefore the key to increase the positive clone hit rate lies in automation of cloning, expression techniques and development of functional assays, akin to biochip hybridization, microarray and mass sequencing.

c) Heterologous gene expression. A challenging yet powerful approach identifying the positive clone is the one that expresses the function. Though the large number of the functional genes have been isolated using functional assays and the number and types of genes that get expressed in *E.coli* are surprising but innovations are required to overcome barriers of heterologous gene expression, assignment of function to the numerous 'Hypothetical proteins' in data bases and the development of functional anchors to isolate active clones.

Conclusions

Once the potential of metagenomics was realized by international scientific fraternity, it took no time to direct research to gather information for other macromolecules like RNA, protein and macromolecules engaged in metabolism. This has resulted in the emergence of new fields like metatranscriptomics, metaproteomics and metablomics. It is perhaps too early to state that metagenomic gene discovery is a technology that has 'come of age'. New approaches and technological innovations are reported on a regular basis and many of the technical difficulties have yet to be fully resolved. However their can be little doubt that the field of metagenomic gene discovery offers enormous scope and potential for both fundamental microbiology and biotechnological development.

Although many advances in heterologous gene expression, library construction, vector design and screening will improve it, the current technology is sufficiently powerful to yield products for solving real world problems, including discovery of new antibiotics and enzymes. Approaches that enrich collection of metagenomic clones will enhance the

power of metagenomic analysis to address targeted questions in microbial ecology and to discover new biotechnological applications.

A genome is a resource for providing access to further understanding of basic molecular biology processes and yielding novel genes, gene products or reactions to biochemists as well as biotechnologists. Yet the surface of this resource has been barely scratched as far as microbial genomes are concerned. The awareness of the real scope of microbial genome diversity and growing interest in biotechnological application of microbial products as pharmaceuticals, bioactives, biocatalysts, biomaterials and so forth must prompt the development of new research techniques for the direct or indirect acquisition of these genomes. Although there is unarguably, a great need for future leaps in techniques for isolating and culturing novel microorganisms the recent development of metagenomics, a field that effectively circumvent the current limitation of microbial isolation has been a major breakthrough.

Acknowledgement J vakhlu is thankful to Professor A. K. Koul, Department of Biotechnology, University of Jammu, India and Dr. R.K.Jain, IMtech Chandigarh for giving valuable suggestions in the preparation of the manuscript.

References

1. Streit WR and Schmitz R (2004) Metagenomics- the key to uncultured microbes. Curr Opin Biotechnol 7:492–498
2. Handelsman J (2005) Sorting metagenomics. Nature Biotech 23:38–39
3. Sharma R (2005) Unculturable' bacterial diversity. An untapped resource. Curr Sci 89:72–77
4. Daniel R (2005) The metagenomics of soil. Nat Microbiol Rev 3:470–478
5. Tringe SG, Von mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Deter JC, Bork P, Hugenholtz P, and Rubin EM (2005) Comparative Metagenomics of Microbial Communities. Science 22: 554–555
6. Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R and Wittwer P (2005) Metagenomic gene discovery: past: present and future. Trends Biotechnology 23:322–329
7. Gill, pop M., Deboy RT, Eckburg PB., Turnbaugh PJ., Samuel BS., Gordon, JI., Relman DA, Liggott CM, Fraser and Nelson KE (2006) Metagenomic analysis of the human distal gut microbiome. Science 312:1355–1359
8. Kimura N (2006) Metagenomics: Access to unculturable microbes in the environment. Microbes Environ 2:201–215
9. Ward, N (2005) New direction and interaction in metagenomic research. FEMS Microbiol Ecol 55:331–338
10. Kaeberlein T, Lewis E and Epstein SS (2002) Isolating uncultivable microorganism in pure culture by a simulated natural environment. Science 296:1127–1129
11. Zengler S, Teledo G, Rappe M, Elhens J, Mathur EJ, Short JM and Killer M *(*2002). Cultivating the uncultured. PNAS 99:15681–15686
12. Torsvik KV and Goksoyr J (1978). Determination of bacterial DNA in soil. Soil Biol Biochem 10:7–12
13. Lane DJ, Pace Bernadettl, Olsen GJ, Stahl DA, Sogin ML and Pace NR (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci USA 82:6955–6959
14. Stahl DA, Lane DJ, Olsen GJ and Pace NR (1985) Characterization of yellowstone hot spring microbial community by 5S rRNA sequences. Appl Environ Microbiol 49: 1379–1384
15. Pace NR, Stahl DA, Lane DJ. and Olsen GJ (1986) The analysis of natural microbial populations by ribosomal RNA. Adv Microb Ecol 9:1–55
16. Schloss P and Handelsman (2003) Biotechnoligical prospects from metagenomics. Curr Opin Biotechnol 14:303–310
17. Leveau JH, Gerards S, Voer W and VanVeen JA (2004) Phylogeny – function analysis of (Meta) genomic libraries: screening for expression of ribosomal RNA genes by large-insert library fluorescent in situ hybridization (lil-fish). Environ Microbiol 6:990–998
18. Riesenfeld GS, Schloss PD and Handelsman J (2004) METAGENOMICS: Genomic analysis of microbial communities. Annual Review of Genetics 38:525–552
19. Entcheva P, Liebl W Johann A Hartsch T and 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
20. Knietsch A, Waschkowitz T, Bowein S, Henne A and Daniel R (2003) Construction and screening of metagenomic libraries derived from enrichment cultures: Generation of a gene bank for genes conferring alcohol oxido reductase activity *E. coli*. Appl Environ Microbiol 69:1408–1416
21. Knietsch A Waschkowitz T Bowein S Henne A and Daniel R (2003) Identification and characterization of coenzymes B_{12} – dependent glycerol dehydratase – and diol dehydratase- encoding genes from metagenomic DNA libraries derived from enrichment cultures. Appl Environ Microbiol 69: 3048–3060
22. Cowan D (2004) Metagenomics, gene discovery and ideal biocatalyst. Biochem Society Transac 32:298–302
23. Venter JC, Reminton K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J and Smith HO (2004) Environmental genome shortgun sequencing of Sargasso Sea. Science 304: 66–74
24. Healy FG, Ray RM, Aldrich HC, Wilkie HC and Shanmugam KT (1995) Direct isolation of functional genes encoding celluloses from the microbial consortia in hemophilic anaerobic digester maintained on lignocelluloses. Appl Microbiol Biotechnol 43:667–674
25. Schmidt T, Delong E and Pace N (1991) Analysis of a marine Pico- Plankto community of 16S rRNA gene cloning and sequencing. J Bacteriol 173:4371–4378
26. Tsai YL and Olson BH (1992) Rapid methods for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. Appl Environ Microbiol 58:2292–2295

27. Zhou J, Brun MA and Tiedje JM (1996). DNA recovery from soils of diverse composition. Appl Environ Microbiol 62: 316–322
28. Santosa DA (2001) Rapid extraction and purification of environmental DNA for molecular cloning applications and molecular diversity studies. Mol Biotech 1759–1764
29. Bergman H (2001) A strategy for optimizing quality and quantity of DNA extracted from soil. J Microbiol method 45:7–20
30. Lorenz P, Liebeton K, Niehaus, F and 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
31. Bohannan, BJM and Hughes J (2003) New approaches to analyzing microbial biodiversity data. Curr opin Microbiol 6: 282–287
32. Short, JM and Mathur EJ (1999) Production and use of normalized DNA Libraries. US Patent 6001574
33. Yokouchi H, fukuoka Y, Mukoyama D, Calugay R, Takeyama H and Matsunaga T (2006) Whole metagenome amplification of a microbial community associated with scleractinian coral by multiple displacement amplification using φ29 polymerase. Environ Microbiol 8:1155–1163
34. Crameri R and Suter M (1993) Display of biologically active proteins on the surface of filamentous phages: cDNA cloning system for the selection of functional gene products linked to the genetic information responsible for their production. Gene 137:69–75
35. Henne A, Daniel R, Schmitz RA and Gottschalk G (1999) Construction of environmental DNA libraries in *E.coli* and screening for the presence of gene conferring utilization of 4-hydroxybutyate. Appl Environ Microbiol 65:3901–3907
36. Courtoiss S, Cappellano CM, Ball M, Francou FX, Normand P, Helynck G,Martinez A, Kolvek SJ, Hopke J, Osborne MS, Paul R, Guerineau M, Jeannin P, Simonel P and parnodet JL (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Appl Environ Microbiol 69:6235–6242
37. Henne A, Schmitz RA, Bomeke M, Gottschalk G and Daniel R (2000) Screening of environmental DNA libraries for presence of genes conferring lipolytic activity on *E.coli.* Appl Environ Microbiol 66:3113–3116
38. Beja O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP, Villacorta R, Amjadi M, Garrigues C, Jovanovich SB, Feldman RA and Delong EF (2000) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. Environ Microbiol 2:516–529
39. Rondon MR, August PR, Betterman AD, Brady SF, Grossman TH, Liles MR, Lynch BA, Macneil IA and Goodman RM (2000) Cloning the soil metagenome; a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Applied Environ Microbiol 66:2541–2547
40. Voget S, Leggewie C, Uesbeck C, Raasch C, Jaeger KE and Streit WR (2003) Prospecting for novel biocatalysts in a soil metagenome. App Environ Microbiol 69:6235–6242
41. Alm E (2000) The presence of humic substances and DNA in RNA extracts affects hybridization results. Appl Environ Microbiol 66:4547–4554
42. Frischer ME, Danfouth JM, Healy AN, Saunder FM (2000) Whole- cell versus total RNA extraction for analysis of microbial community structure with 16S- rRNA- targeted oligonucleotide probes in salt marsh sediment. Appl Environ Microbiol 66:3037–3043
43. Griffiths RT, Whiteley AS, Donnell AG and Bailey MJ (2000) Rapid method for co extraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and r-RNA- based microbial community composition. Appl Environ Microbio 66:5488–5491
44. Rebel F, Renard JP and Duranthon V (1995) PCR-generated cDNA libraries from reduced numbers of mouse oocytes. Zygote 3:241–250
45. Starkey MP, Umrania Y, Nundy CR and Bishop MJ (1998) Reference cDNA libraries available from Europian sources. Mol Biotech 9:35–57
46. Watanabe K, Teramoto M, Futamata H and Harayama S (1998) Molecular detection, isolation and physiological characterization of functionally dominant phenol- degrading bacteria in activated sludge. Appl Environ Microbiol 64:4396–4402
47. Henckel T, Friedrich M and Conrad R (1999) Molecular analyses of the methane oxidizing microbial community in rice field soil by targeting the genes of the 16S rRNA particulate methane monooxygenase, and methanol dehydrogenase Appl Environ Microbiol 65:1980–1990
48. Henckel T, Jackel U, Schnell S and Conrad R. (2000). Molecular analyses of novel methanotrophic communities in forest soil that oxidize atmospheric methane Appl Environ Microbiol 66:1801–1808
49. Futamata H, Harayama S and Kazuya W (2001) Group specific monitoring of phenol hydroxylase genes for a functional assessment of phenol stimulated trichloroethylene bioremediation. Appl Environ Microbiol 67:4671–4677
50. Mesarch MB, Nakabsu CH and Nies Loring (2000). Development of catechol 2,3 dioxygenase specific primers for monitoring bioremediation by competitive quantitative PCR. Appl Environ Microbiol 66:678–683
51. Seow KT, Meurer G, Gerlitz M, Wendt-Pienkowski E, Hutchinson CR and Davis J (1997) A study of interactive type II polyketide synthases, using bacterial genes cloned from soil DNA: means to access and use genes from uncloned microorganisms. J Bacteriol 179:7360–7368
52. Wilson MS, Bakerman C and Madsen EC (1999) In- situ real –time catabolic gene expression: extraction and characterization of Napthalene dioxygenase mRNA transcript from ground water. Appl Environ Microbiol 65:80–87
53. Ochman H (2000) Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304
54. Strokes HW, Holmes AJ, Nield BS, Holley MP, Nevalainen KM, Mabbutt and Gillings MR (2001) Gene cassette PCR: Sequence independent recovery of entire genes from environmental DNA. App Environ Microbiol 67:5240–5246
55. Bennet PM (1999) Integrons and gene cassette : a genetic construction kit for bacteria. J Antimicrob Chemo 43:1–4
56. Collis CM and Hall RM (1992) Gene cassettes from insert region of integrons are excised as covalent closed circles. Mol Microbio 6:2875–2885
57. Ricchia GD and Hall RM (1995) Gene cassettes: a new class of mobile elements. Microbiology 141:3015–3027
58. Rowe–Magnus DA and Mazel D (2001) Integrons: natural tool for bacterial genome evolution. Curr Opin Microbiol 4: 565–569

59. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS and Banfield JF (2004) Community structure and metabolism through reconstruction of microbial genome from environment. Nature 428:37–43
60. Handelsman J (2004) Metageneomics: application of genomics to uncultured microorganism. Microbiol Mole Biol Rev 68:669–685
61. Yun J and Ryu S (2005) Screening for novel enzymes from metagenome and sigex, as way to improve it. Microbial cell factories 4:8
62. Sebat JL, Colwell FS and Crawford RL (2003) Metagenomic profiling: Microarray analysis of an environmental genomic library. Appl Environ Microbiol 69:4927–4934
63. Dennis P, Edward EA, Liss SN and Fulthorpe R (2003) Monitoring gene expression in mixed microbial communities by using DNA microarrays. Appl Environ Microbiol 69: 769–778
64. Quaiser A, Ochsenreiter T, Klenk A, Kletzin A, Treusch AH, Meurer G, Eck J, Sensen CW and Schleper C (2002) First insight into genome of an uncultivated crenarchaeote from soil. Environ Microbiol 4:603–611
65. Bintrim SB, Donohue TJ, Handelsman J, Roberts GP and Goodman RM (1997) Molecular phylogeny of archea from soil. Proc Natl Acad Sci USA 94:277–282
66. Hugenholtz P, Goebel BM. and Pace NR (1998) Impact of culture independent studies on the emerging phylogentic view of bacterial diversity. J Bacteriol 180:4765–4774
67. Kellenberger E (2001) Exploring the unknown. EMBO rep 2:5–7
68. Stein JL, Marsh TL, Wu KY, Shizuya H and Delong EF (1996) Characterization of uncultivated prokaryotes: isolation and analysis of 40- kilobase- pair genomic fragment from a marine archaeon. J Bacteriol 178:591–599
69. Beja O, Spudich EN, Spudich M, Leclerc and Delong EF (2001) Proteorhodopsin phototrophy in ocean. Nature 411: 786–789
70. Song JS, Jeon JH, Lee JH, Jeong SH, Jeong BC, Kim SJ, Lee JH and Lee SH (2005). Molecular characterization of TEM-type beta-Lactamases identified in cold seep sediments of Edison Seamount. J Microbiol 43:172–178
71. Kottmann 2005. Data acquisition, Design and implementation of an integrative database for metagenomic (from net). *Marmic.mpg.de/marimic- cms/uploads-edits/Renzo-Kottmann.pdf*
72. Greenbelt CL, Baum J, klein BY, Nachshon S**,** Koltnov V and Cano RJ (2004) Micrococcus luteus- survival in amber. Microb Ecol 48:120–127
73. Cano RJ, Tiefenbrunner F, Ubaldi M, Del Cueto C, Luciani S, Cox T, Orkand P,Kunzel KH and Rollo F (2000) Sequence analysis of bacterial DNA in colon and stomach of the Tyrolean Iceman. Am J Phys Anthropol 112:297–309
74. Angly F, Rodriguez-Brito B, Bangor D, McNairnie P, Breitbart M, Salamon P, Felts B, Nulton J, Mahaffy J and Rohwer F (2005) PHACCS, online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. BMC Bioinformatics 6:41
75. Gabor EM, Alkema WB and Janssen DB (2004) Quantifying the accessibility of metagenome by random expression cloning techniques. Environ Microbiol 6:876–886
76. Cottrell MT, Jessica A, Moore and David L. Kirchman (1999) Chitinases from Uncultured marine microorganisms. Appl Environ Microbol 65:2553–2557
77. Richardson TH, Tan X, Frey G, Callen W, Cabell M, Lam D, Macomber J, Short JM, Robertson DE and Miller C (2002) A novel high performance enzyme for starch liquefaction. J Biol Chem 277:26501–26507
78. Brady SF, Chao CJ and Clardy J (2002) New natural product families from an environmental DNA (eDNA) gene cluster. J Am Chem Soc 124:9968–9969
79. Wang GY, Graziani E, Water B, Pan W, Li X, McDermott J, Meurer G, Saxena G, Anderson RJ and Davies J (2000) Novel natural products from soil DNA libraries in a *Streptomycete* host. Org Letter 2:2401–2404
80. Brady SF and Clardy J (2000) Long- chain N- acyl amino acid antibiotics isolated from heterogeneously expressed environmental DNA. J Am Chem Soc 122: 12903–12904
81. Brady SF, Chao CJ, Handelsman J and Clardy J (2001) Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. Organ Lett 3: 1981–1984
82. Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM and Handelsman J (2002) Isolation of antibiotics Turbomycin A and B from a metagenomic library of soil microbial DNA. Appl Environ Microbiol 68:4301–4306
83. Mac Neil IA, Tiong CA, Minor C, August PR, Grossman TH, Lioacono KA, Lynch BA, Phillips T, Narula S and Sundaramoorthi R (2001) Expression and isolation antimicrobial small molecules from soil DNA libraries. J Mol Microbiol Biotechnol 3:301–308
84. Bell JC, Sunna A, Gibbs MD, Curach NC, Nevaleerinen H and Bergquist P (2002) Prospecting for novel lipase gene using PCR. Microbiology 148:2283–2291
85. Rhee JK, Ahn DG, Gukin Y and Oh WJ (2005) New thermophilic and thermostable esterase with sequence similarity to hormone sensitive lipase family cloned from a metagenomic library. Appl Environ Microbiol 71: 817–825
86. Jiang Z, Wang H, Ma Y and Wei D (2006) Characterization of two novel lipase genes isolated directly from environmental sample. Appl Microbiol Biotechnol 70:327–332
87. Elend C, Schmeisser C, Leggewie C, Babiak P, Carballeira JD, Steele HL, Reymond JL, Jaeger KE and Streit WR (2006) Isolation and Biochemical characterization of two novel metagenomic derived esterases. Appl Environ Microbiol 72:3637–3645
88. Robertson DE, Chaplin JA, Desant G, Podar M, Maden M *et al* (2004) Exploring nitrilase sequence space for enantioselective catalysis. Environ Microbiol 70:2429–2433
89. Brennan Y, allen W, Christoffersen L, Dupree P, Healay S, Hernandez M, Keller M, lile, Palachal N, Tamayo G, Wells S, Mathur EJ, Short M, Robertson DE and Steer BA (2004) Unusual microbial xylanases from insect guts. Appl Environ Microbiol 70:3609–3617
90. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent, M, Gill SR, Nelson KE and Relman DA (2005). Diversity of the human intestinal microbial flora Science 308:1635–1638

91. Lorenz P and Eck J (2005) Metagenomics and industrial application. Nat Rev Microbiol 3:510–516
92. Delong EF (2004). Microbial population genomics and ecology: The road ahead. Environ Microbiol 6: 875–878
93. Beja O (2004) To BAC or not to BAC: marine ecogenomics. Curr Opin Biotechnol 15:187–190
94. Hugenholtz P (2002) Exploring prokaryotic diversity in the genomic era. Genome Biol 3:3
95. Schmeisser C, Steele H and Streit W (2007) Metagenomics, Biotechnology with non- culturable microbes. Appl Microbiol Biotechnol 75:955–962
96. Cowan DA (2000) Microbial genomes – the untrapped resource. Trends in Biotechnolo 18:14–16
97. Hardeman F and Sjoling S (2007) Metagenomic approach for the isolation of a novel low- temperature– active lipase from uncultured bacteria of marine sediment. FEMS Microbiol 59:524–534