Metagenomics in animal gastrointestinal ecosystem: a microbiological and biotechnological perspective

B. Singh · T. K. Bhat · N. P. Kurade · O. P. Sharma

Received: 15 May 2007 / Accepted: 15 December 2007 / Published online: 17 June 2008

Abstract Metagenomics- the application of the genomics technologies to nonculturable microbial communities, is coming of age. These approaches can be used for the screening and selection of nonculturable rumen microbiota for assessing their role in gastrointestinal (GI) nutrition, plant material fermentation and the health of the host. The technologies designed to access this wealth of genetic information through environmental nucleic acid extraction have provided a means of overcoming the limitations of culture-dependent microbial genetic exploitation. The molecular procedures and techniques will result in reliable insights into the GI microbial structure and activity of the livestock gut microbes in relation to functional interactions, temporal and spatial relationships among different microbial consortia and dietary ingredients. Future developments and applications of these methods promise to provide the first opportunity to link distribution and identity of rumen microbes in their natural habitats with their genetic potential and in situ activities.

Keywords Metagenomics · Rumen ecosystem · Gastrointestinal tract microbes · Livestock production

Introduction

Microorganisms play essential roles in specialized niches including inside of the body as well as external surfaces, deep seas, saline marshes, glaciers and boiling streams. Microbes have many of the properties similar to complex organisms like humans. They also exhibit unique properties such as ability to degrade waste products. As a result, the genetic and biological diversity of microorganisms is an important area of scientific research [1]. The global microbial diversity presents an enormous, largely untapped genetic and biological pool that could be exploited for the recovery of novel genes, metabolic pathways and valuable products [2]. Unfortunately, scientists are able to grow in vitro only less than 1% of all microorganisms present in nature. This leaves more than 99% of the microbial diversity unexploited. The apparent underestimation of true microbial diversity is largely due to a reliance on culture-based microbiological studies. Recent years have seen a remarkable evolution in the development and application of molecular tools that are allowing the microbiologists to characterize and understand microbial communities in unprecedented ways. By creatively leveraging these new data sources, microbial ecology has potential to have a transition from a purely descriptive to a predictive framework, in which ecological principles are integrated and exploited to engineer the systems that are biologically optimized for the desired goal. Many of these biotechniques are based on 16S ribosomal DNA (rDNA) sequences and exploit either hybridization or PCR techniques.

Metagenomics approaches enable the microbiologists to have a look at a more complete scenario of microbial communities and thus, to better understand the microbial systems. The advancements in metagenomic technologies

B. Singh (✉) · T. K. Bhat · N. P. Kurade · O. P. Sharma
Animal Biotechnology Lab. Regional Station,
Indian Veterinary Research Institute,
Palampur - 176 061, India

e-mail: bsbpalampur@yahoo.co.in

over the past few years have provided an inexhaustible access to much of the prokaryotic genetic resources available in nature. Though DNA-DNA hybridization, which has been used for many years, is still considered to be the standard protocol for bacterial species identification, PCR-based methods investigating 16S rRNA gene sequences, and other approaches, such as bioinformatics tools are being increasingly applied to study microbial diversity for obtaining novel genes with potential pharmaceutical and biotechnological applications. In view of these advancements, there is a growing belief that the term ‘unculturable’ is inappropriate [3, 4], and that in reality we have yet to discover the correct microbial culture conditions. We, here, present an overview of the process of metagenomics, recent status of the metagenomic approaches aimed at exploring and tapping the potential of GI tract, specifically the ruminants’ gut ecosystem. Examples of the metagenomic process as applied for some other microbial habitats are also incorporated.

What is metagenomics?

The term ‘metagenomics’ (encompassing the various biological fields, also known as ecogenomics, microbial community genomics or environmental genomics) was first coined by Jo Handelsman and co-workers [5] to study genomes from all the microbes in an environment as opposed to the genome from one organism isolated from the environment and maintained in vitro. Metagenomics refers to the study of the collective genomes in an environmental community, and makes it feasible to obtain information about entire microbial communities like deep seas, soil or gut ecosystem etc. Aimed at elucidating the genomes of nonculturable microbes, and to better understand the global microbial ecology on one side and, on the other side driven by increasing industrial biotechnological demands for novel enzymes and biomolecules, metagenomics has emerged as a promising tool for exploiting different microbial ecosystems. Furthermore, the microbial ecologists have ubiquitously realized the need for culture-independent investigations in view of the observations that culture-dependent microbiological approaches for studying microbial ecology were left behind by the astounding inferences derived from sequencing small subunit rRNA genes of microbial populations.

Initially, non-cultured microflora and ancient DNA analyses had been the main targets for metagenomic studies, now the technology has been applied to study an array of microbial ecosystems like deep-sea aquatic microflora, soil microbes and GI ecosystem in ruminants and monogastrics including humans [6, 7]. Only 0.001–0.1% of the

total microbes in sea water, 0.25% in freshwater, 0.25% in sediments and only 0.3% soil microorganisms were found to be cultivable [8]. As reviewed recently [9, 10], the metagenomic studies have largely progressed due to the construction of efficient cloning vectors (e.g. bacterial artificial chromosomes- BACs, and yeast artificial chromosomes, YACs etc.), which allow cloning and expression of larger DNA segments or genes. Study sites have expanded tremendously and gene sequence information is available online; and finally, the number of sequences generated from a particular microbe has been increasing.

Methodology in microbial metagenomics

Sample enrichment

In metagenomic analysis process (Fig. 1) the target gene(s) may represent a small proportion of the total nucleic acid fraction. Pre-enrichment of the sample, thus, provides an attractive means of enhancing the screening hit rate [11]. The rare or less represented microbial target genes need to be enriched by applying suitable enrichment methods [11], ranging from whole-cell enrichment, to the selection of a particular microbial group. Among the different methods, differential centrifugation could be a promising sample enrichment protocol, as it has been used to enrich *Buchnera aphidicola* and *Cenarchaeum symbiosum* symbionts by isolating them from their natural environment in preparation for whole genome sequencing [12]. Culture enrichment on a selective medium favors the growth of target microorganisms. The inherent selection pressure can be based on nutritional, physical or chemical criteria, although substrate utilization is most commonly employed. For example, a four-fold enrichment of cellulase genes in a small insert expression library was obtained by microbial culture enrichment on carboxymethylcellulose [13]. Although culture enrichment will result in the loss of a large proportion of the microbial diversity, the problem can be partially minimized by reducing the selection pressure to a mild level after a short period of stringent treatment. Although quantitative PCR is a sensitive technique to detect small changes in community composition, it does not determine microbial growth yields. Furthermore, studying slow-growing microbes require specialized culture conditions. For instance, membrane coupled bioreactors (MBRs) were reported to be the useful tools to study ecology of slow-growing bacteria [75]. Another method to enrich mixed bacterial population comprising of slow-growing microorganism, using the gel microdrop (GMD) growth assay, combined with fluorochrome staining and flow cytometry has been developed

by Akselband and co-workers [76]. They demonstrated the feasibility of using this experimental approach to detect clonogenic growth of individual bacteria within GMD in less than 3h and to separate various microbial subpopulations based on differential growth rates. The microorganisms in this study were found to remain viable after sorting and capable of being propagated for further analysis [76].

Growing methane-oxidizing anaerobic microbial communities on ^{15}N -labelled ammonium as N-source has also been shown to be a sensitive and fast way to detect growth of extremely slow-growing microorganisms in different sediment systems [77]. However, information is not currently available on applications of these techniques for enrichment of slow-growing rumen or GI bacteria.

Fig. 1 Metagenomic analysis of the animal GI ecosystem. The herbivore's GI ecosystem is supposed to harbor a wealth of enzymes, microbes and novel products for applications to livestock nutrition, health and for industrial applications.

Nucleic acids extraction

Successful application of molecular techniques relies on effective recovery of high purity nucleic acids from the environmental samples. In metagenomic analysis, DNA extracted from environmental sample is to be cloned for constructing a metagenomic library, therefore, purity and recovery of open reading frames from metagenomic nucleic acids is necessary. An ideal procedure for recovering nucleic acids from environmental samples should meet several criteria like, high and unbiased recovery of nucleic acids representative of the entire microbial community, large enough size of pure nucleic acids (RNA/DNA) sequences, and simplicity of the extraction protocols [14]. However, owing to the physiochemical diversity in matrices serving as microbial habitats, there is not any universally applicable protocol for isolating the microbial nucleic acids and therefore, different extraction protocols have been in use [15, 16]. The principal strategies for the recovery of metagenomic DNA include the microbial recovery and their cell lysis [17]. Extraction and purification of metagenomic DNA is difficult, and the process as whole is a compromise between vigorous extraction required for the microbial genomes, minimized mechanical or chemical DNA shearing and the co-extraction of contaminating macromolecules. Different DNA isolation and purification methods from complex microbial ecosystem have been described [18]. However, chemical or enzymatic lyses are more gentle methods for recovering higher molecular weight DNA. The technologies for recovering RNA from environmental samples are largely similar to those used for DNA isolation, modified to optimize the yield of intact mRNA by minimizing single-stranded polynucleotide degradation [19–21].

Total nucleic acids extracted directly from samples do not typically represent the genomes of all the microbes within the sample. Some microbial species are likely to be overshadowed by a few dominant microbial populations and therefore, rare organisms contribute a relatively low proportion of the total nuclear materials [22]. This leads to a selective bias in downstream analyses such as PCR amplification of the nucleic acids, sequencing of the genes and interpretation of the data. The problem could be partially resolved by means of experimental normalization [23]. Separation of genotypes is achieved by cesium chloride gradient ultracentrifugation in the presence of intercalating agents, such as bisbenzimidazole or ethidium bromide. Normalization of the genomic materials can also be achieved by denaturing the genomic DNA fragments, and re-annealing single stranded DNA (ssDNA) under stringent conditions (e.g. 68.8°C for 12–36 h). Abundant ssDNAs anneal more rapidly to generate double-stranded nucleic acids than rare

DNA species. The remaining single-stranded sequences are then separated from the double-stranded nucleic acids, resulting in an enrichment of rarer sequences within an environmental sample [23].

Microbial genome and gene enrichment

Genome enrichment strategies are aimed at targeting the active components of a specific microbial population. With the advent of genome enrichment and amplification techniques in metagenomics, new options for overcoming limitations in DNA purity and yield have become easier.

Stable-isotope probing (SIP) is a method used in microbial ecological studies wherein specific functional groups of organisms that incorporate particular substrates are identified without prerequisite of their *in vitro* cultivation. Nucleic acids-SIP techniques involve labeling uncultivated microorganisms in environmental samples using a substrate enriched with certain stable isotopes (^{13}C and/ or ^{15}N etc.), assimilated by microbes and subsequently incorporated into the microbial genomes. Labeled biomarkers, such as phospholipids-derived fatty acid (PLFA), ribosomal RNA, and DNA can then be probed using a range of molecular analytical techniques, and used to identify the organisms that have incorporated the labeled substrates. For instance, $^{13}\text{CH}_3\text{OH}$ -labeling of forest soil metagenomic DNA resulted in the identification of both known α -proteobacterial methylotrophs and novel methanol dehydrogenase (*mxhF*) gene variants belonging to Acidobacterial taxa [24]. In another study, analysis of ^{13}C -phenol enriched anaerobic populations by RNA-SIP demonstrated that phenol degradation was dominated by microbes belonging to the genus *Thaurea*, earlier unknown for phenol degradation. Actively growing microorganisms can also be labeled with 5-bromo-2-deoxyuridine (BrdU) and the labeled DNA or RNA be separated by immunocapture or density gradient ultracentrifugation [25]. Limitations of these methods include cross-feeding and recycling of the label within the community, resulting in loss of specific enrichment.

Another method, termed suppression subtractive hybridization (SSH) is an effective approach to identify the genes that vary in expression levels during the biological processes. SSH identifies genetic differences between microorganisms and is, therefore, a powerful tool for specific gene enrichment and detecting the specific functional genes or DNA markers in microorganisms. Key features of this method are- the simultaneous normalization and subtraction steps. The normalization step of the process equalizes the abundance of DNA fragments within the target population, and the subtraction step excludes sequences that are common to the two populations being compared. SSH has typi-

cally been applied to analyze genetic differences between two closely related bacteria (e.g. identification of genetic elements contributing to pathogenesis) [26], but has also recently been used to identify differences between complex DNA samples from the rumens of two hay-fed steers [27]. SSH is likely to be a promising tool for comparison of metagenomic samples from diverse environments including gut ecosystems

To selectively enrich for a specific target gene within a metagenome, a more practical approach would be to use differential gene expression technologies that rely on the isolation of transcriptome to target transcriptional differences. The gene expression profile of a culture grown from a metagenomic sample can be compared following pre- and post- exposure to a specific substrate, thus offering the detection of expression of the particular genes. These approaches appear to be promising to identify herbivore gut flora from the animals exhibiting a natural adaptation to the diets enriched with an array of detrimental phytometabolites. A review by Green and co-workers [28] describes some of the recent innovative molecular biological methodologies on the subject.

Gene targeting

Gene targeting is the process of studying the identified genes without removing them from natural position in genome in living cells. The process relies much on suitability of a gene cloning vector. Gene-specific PCR is used extensively to probe the microorganisms with specific metabolic or biodegradative traits. For example, the targeting of genes such as methane monooxygenase, methanol dehydrogenase and ammonia monooxygenase was used to identify methanotrophic [29], and chemolithotrophic ammonium-oxidizing bacteria [30, 31]. Gene targeting approaches have been used in understanding the key community-regulators in intestinal bacteria in diseases like Crohn's disease [32], and developing the new antibiotics targeting pathogenic bacterial-genes whose expression is essential for their in vivo viability [33].

As a tool for biocatalysts investigations, gene-specific PCR has some limitations. First, the design of primers is dependent on existing gene sequence information and skews the search in favor of known DNA sequence types. Functionally similar genes resulting from convergent evolution are not likely to be detected by a single gene-family-specific set of PCR primers. Second, only a fragment of a structural gene will typically be amplified by gene-specific PCR, thus requiring additional steps to access the full-length genes in new microbial groups. Amplicons could be labeled as probes to identify the putative full-length gene(s) in con-

ventional metagenomic libraries. Alternatively, PCR-based strategies for the recovery of either the up- or down-stream flanking regions can be used to access the full-length gene. For example, universal 'fast walking' [34, 35], panhandle PCR (to amplify known sequence flanked by unknown sequence) [36], inverse PCR and adaptor-ligation PCR [37] are the important tools in recent microbial genomic analyses. The methods are technically difficult to apply at a metagenomic level owing to the increased complexity of a metagenomic DNA sample, but have been used successfully for the recovery of novel gene variants of 2, 5 diketo-D-gluconic acid reductase from environmental DNA [38]. However, these techniques are likely to revolutionize the current approaches to study microbial ecology in the rumen/ GI tract and to provide, not simply a refinement or increased understanding, but a complete description of the rumen ecosystem for the first time. At the moment some of these approaches remain laborious and time-consuming, hence, innovative alternatives are being developed.

Analysis of metagenomic libraries

Among the various approaches for metagenome data analysis, the function-driven analysis (screening the metagenomic libraries for an expressed and detectable trait) and sequence-driven analysis (metagenomic libraries screened for particular DNA sequences) are used to draw information from microbial metagenomic libraries.

The function-driven analysis

This approach is based on the identification of the clones that express a desired trait, followed by characterization of the active clones by biochemical and molecular (gene sequence) features. This is a quick approach and identifies the clones (expressing enzymes or other proteins) that have medicinal, agricultural or industrial applications. Though a highly preferred and promising approach, this method suffers the limitation that it requires expression of functional product in host cell which further needs clustering of all the genes encoding a particular product. Furthermore, it also depends on availability of an assay for the function of interest that can be performed efficiently on vast libraries, as the frequency of active clones is quite low. Approaches are being developed to overcome these limitations. Improved systems for heterologous gene expression are being developed with shuttle vectors that facilitate screening of metagenomic DNA in diverse host species and with modifications of *Escherichia coli* as host system to expand the range of gene expression. Although the genes encoding most of

the enzymes required for synthesis of secondary metabolites are usually clustered on a single contiguous piece of DNA, it is still necessary to have large intact DNA fragments encoding the complex molecules (e. g. antibiotics). Methods to improve the preparation of large fragments of DNA that are pure enough to be cloned are being pursued. Functional screening of metagenomic libraries has identified several novel and previously described antibiotics [8, 39–41], an antibiotic-resistance gene [42], and some commercially important hydrolytic enzymes [43, 44].

Phenotypic characterization and 16S rRNA gene analysis has been found to be promising in identifying novel bacterial species in avian GI tract [45]. Furthermore, from mammalian GI tract some novel hydrolases [46, 47] and polyphenol oxidases [48] have been identified.

The sequence-driven analysis

Although activity based approaches have clear conceptual advantages for identifying economically relevant biocatalysts, identification of potential enzyme candidates in metagenomes based on sequence similarity is viable and rewarding approach [49]. Sequence-driven analysis relies on the conserved DNA sequences to design hybridization probes or PCR primers for screening the metagenomic libraries for clones that are expected to contain sequences of interest. Recent advances in bioinformatics tools have greatly facilitated the process of drawing information from sequence-driven metagenome analysis.

Sequencing of clones carrying phylogenetic anchors, such as the 16S rRNA gene and Archaeal DNA repair gene *radA* [50–53] has led to functional information about uncultivated marine organisms from which these clones were derived.

There has been disagreement about the utility of random sequencing of metagenomic clones [12]. Although some members view this approach as too undirected to yield biological understanding, others stress that there is so little known about some divisions of bacteria that any genomic sequence is helpful in guiding the design of experiments to reveal their biology. Sequencing many clones derived from one division may provide sufficient data to begin to discern patterns for developing hypotheses [12].

The *Acidobacterium* (acid-tolerant, encapsulated Gram-negative member, of ubiquitous but poorly understood phylum), offers a prime candidate for metagenomic sequencing. A collection of metagenomic clones carrying 16S rRNA genes that align with *Acidobacterium* division, have been assembled and full length sequencing is underway [54].

The sequence conservation of regions of phylogenetic anchors facilitates their isolation without prior knowledge of full gene sequence. By contrast, the sequences of most genes of interest are far too divergent to make it possible to identify new homologues by PCR or hybridization. However, a few classes of genes contain sufficiently conserved regions to facilitate their identification by sequence instead of function. Thus, application of this approach is likely to be broadened as patterns emerge that define more gene families containing highly conserved features. The genes distantly related to the known genes are of particular interest for biotechnological applications, but are the least likely to be detected by probes or primers whose designs are based on already available database entries.

MEGAN (MetaGenome ANalyzers), a new computer program that generates specific profiles from sequencing data by assigning the reads to NCBI taxonomy using a straight-forward assigned algorithm, is a recently developed method for analysis of metagenomic data [55]. The MEGAN approach has been found applicable to several data sets including subset of Sargasso Sea data set (obtained by Sanger sequencing), data obtained from mammoth (*Mammuthus primigenius*) bone (obtained by “sequencing-by-synthesis” approach) and identifying the microbial species based on already available microbial (*Escherichia coli* and *Bdellovibrio bacteriovorus*) genome sequence information [55].

Plasmids confer the bacteria additional advantages for survival, and have also been exploited as cloning vectors in genetic engineering of the microorganisms. The transposon aided capture (TRACA), a PCR based detection of bacterial genes, has been developed for detection of mobile genetic elements (MGEs) including plasmids. In TRACA, plasmids are identified from the metagenomic DNA extracts, and the identification is independent of any of the known plasmid-encoded traits (such as selectable markers or ability to mobilize and replicate in surrogate host species) [56]. The plasmids in the extracted microbial metagenomes are provided with an origin of replication, and a selectable marker, by tagging plasmids with a transposon encoding these functions, in an in vitro reaction. The method has been reported to be very promising in identification of plasmids from a range of GI bacteria, some of which were found lacking any selectable markers [56].

The application of TRACA to further study plasmids resident in the gut microbiota, and those found in other bacterial ecosystems, is likely to identify new plasmids encoding diverse functions important for adaptation, survival, interaction between bacteria within an ecosystem, and interactions between bacteria and their host species or environment [56]. The authors suggest the scope to further modify

and improve the original TRACA method for applications to other microbial ecosystem.

Gut microbial and biotechnological applications of metagenomics

The mammalian gastrointestinal (GI) ecosystem evolved over millions of years with a diverse population of symbiotic microbes, comprises of more than 400 different species [57]. The gut microbes have a profound effect on immunological, nutritional and physiological processes of the host [57, 58]. Also called as “normal flora” the gut microorganisms have adapted in such a manner that they have no adverse effects on their host’s overall health, and often they are beneficial or even obligatory to the host, especially in the herbivores. In view of the abundance and significance of the gut flora to the host, Goodacre [59] emphasizes human as a human-microbe hybrid, and the health of this super organism will be affected by intrinsic properties such as human genetics, diurnal cycles, and

age and by extrinsic factors, such as, lifestyle choices and acquisition of a stable “healthy” gut microbiota (so called microbiome).

Gut flora have also been identified as inexhaustible sources of biotherapeutics, enzymes, genes and novel products for applications in livestock health, nutrition and industrial development. Biotechnological applications (Table 1) from metagenomics will be fostered by the pursuit of fundamental ecological studies and focused screen for bioprospecting, just as both basic and applied approaches have contributed to the discovery of antibiotics and enzymes from cultured microbes. As the concept of microbial diversity has dramatically expanded within the past decade, the need and search for new microbial products has emerged. In a recent article on industrial applications of microbial metagenomics, Lorenz and Eck [60] emphasize that industries have different motivations to probe the enormous resource of uncultivated microbial diversity; metagenomics has the potential to substantially impact industrial revolution, and the process can lead to development of novel enzymes, processes, products and applications.

Table 1 Metagenomic tools in microbiological and biotechnological interventions in animals’ GI ecosystem

Animals	Targets	Future prospects
Ruminants	Methanogens	Identification of novel methanogens, and study of methanogenesis in rumen Manipulation of rumen ecosystem for minimizing green-house gases emissions Developing vaccines, and inducing host immune system mediated antibodies against rumen protozoa/ methanogens
	Hydrolytic-enzymes yielding microbes	Identifying the genes/ enzymes with industrial (food and animal feed industry, pulp and paper, and textile) applications Enhanced plant biomass utilization for growth and production of animals Combining classical in vitro microbial culture-, and molecular techniques for metabolic engineering of the rumen ecosystem for livestock and industrial development Quantitative estimation of different GI tract microbes
	Novel microbial species	Establishing cultures conditions for newly identified microbes Defining the host-microbial interactions, and exploring the potential of novel microbial species for health and nutrition of the host Studying the rumen ecosystem of the animals adapted to diets enriched with anti-nutritional plant secondary metabolites Exploring the potential of novel microbial species as direct-fed microbials, and developing strategies for bio-monitoring of the trans-inoculated microbes in new hosts
	Novel genes and enzymes	Using the novel gut bacterial plasmids and genes and enzymes as tools for genetic engineering of the resident GI flora Studying the interactions among the different microbial consortia, and between microbes and the feed particles
Monogastrics (Porcine and poultry etc.)	Novel microbial species	Establishing cultures conditions for newly identified GI microbes Studying the host gut microbial interactions Exploiting the potential of gut microbes for livestock feed industrial applications Exploring the potential of novel microbial species as direct-fed microbials, and developing strategies for bio-monitoring of the transferred gut microbes in new host GI ecosystems

Nucleic acids-based studies indicating that a majority of the bacteria in gut ecosystems are different from those described in *in vitro* culture, has led to an extensive development and application of culture-independent approaches to study the GI microbial ecosystem. Gut microbes are supposed to possess a wealth of genes encoding valuable proteins and other macromolecules. To better understand the roles that specific organisms or consortia of organisms play, investigators need to identify gut microbial species associated with these attributes.

Applications of metagenomics have been extended to the normal flora of animals species as well. Investigating the potential of metagenomics in gut ecosystem, a study [55] involved construction of a bacterial artificial chromosome (BAC) library from the DNA directly extracted from BALB/c mice large bowel microbiota. A functional screen of the library for β -glucanase activity was carried out, and from the annotation of the sequences obtained, glucanolytic enzymes (lichenases and laminarianases) activities were detected. The metagenomic evidence was also for presence of other putative genes, some of which might have a role in environmental sensing, nutrient acquisition or GI microbial coaggregation [61].

The increased ease with which molecular analysis can be carried out has led to an explosion in sequencing of rDNA sequences from different bacterial species and strains from extreme environments. Gut metagenomic studies are underway: a recent comprehensive rRNA survey [62] provides a base line for such metagenomic studies, by generating 13355 rRNA gene sequences from multiple sampling by biopsies performed on caecum and rectum. Notable findings from the study included substantial differences between three individuals sampled, indicating that fecal microbes are not indicative of entire GI tract microbiota, that only 20% of the sequences represent currently cultivated microorganisms, and that gut lumen and mucosal surface bacterial populations are significantly different [63].

Metagenomics and rumen ecosystem

The rumen, due the presence of unique diverse microbial consortia and continuous microbial products, has been identified as fountain head of valuable fibrolytic enzymes and novel products that could be exploited in feed (plant biomass saccharification, deriving and supplying critical nutrients to the animals from low quality forages), pulp and paper processing [64–66]. The progress in exploiting the rumen ecosystem is slow, primarily due to the fact that rumen microbial ecosystem cannot be defined using the traditional culture-dependent techniques. Hence, different

molecular techniques based on 16S rRNA or rDNA are now being used to study the animals' gut microbial populations [67–69].

A metagenomic expression library of bulk DNA extracted from the rumen content of dairy cattle was established in a phage vector, and activity-based screening was employed to explore the functional activity of the rumen microbes. Sequence analysis of the retrieved enzymes revealed that 36% (8/22) sequences were entirely new and formed deep-branched phylogenetic lineages with no close relatives among known esters and glycosyl hydrolases [70]. The studies demonstrated the usefulness of the metagenomic approach to identify novel hydrolytic enzymes from bovine rumen, thus exploring the potential applications of rumen microbial enzymes in industrial biocatalysis [46, 70]. In another study, RL5, a gene coding for a novel polyphenol oxidase, was identified through activity screening of a metagenome expression library from the bovine rumen microflora [48].

An important application of metagenomic analysis to ruminant nutrition is- the quantitative determination of total rumen microbial proteins and differentiating between bacterial and protozoal biomass. This is because the ciliate protozoa are present in rumens of most ruminants (10^5 – 10^6 cells/ml of rumen fluid) and can represent up to half of the total microbial nitrogen. Despite the importance of protozoal ecology, there is no widely applicable marker to measure and differentiate protozoal protein fractions from bacterial proteins. In an attempt to solve this problem, Skillman et al. [71] developed a real-time PCR assay to quantify *Entodinium* and *Dasytricha* populations in total rumen contents. The method was found to be more accurate and repeatable compared to the conventional microscopic counts of the rumen protozoa.

Although early initiatives in microbial gene sequencing were focused on microbes with either small genomes (mycoplasma and other archaeobacteria) and/ or economically important pathogens, similar approaches are now being applied to agriculturally and environmentally relevant microflora. The rumen bacteria with relevance to fiber degradation, for which genome sequences are now available, include *Fibrobacter succinogenes*, *Ruminococcus albus* and *Prevotella ruminicola* strain 23. The *F. succinogenes* genome has yielded significant insights into an unexplored gut microbial niche [72], and from the gene-list at least 24 genes encoding endoglucanases and cellodextrinases have been identified, far exceeding the six genes previously characterized by recombinant DNA strategies employing *E. coli* as cloning host.

Other molecular tools, like real time PCR and bioinformatics are also in use for studying GI or rumen ecosystem.

A real time PCR CYBER Green assay, using PCR primers to target total bovine rumen fungi, and the major cellulolytic bacteria *R. flavefaciens*, and *F. succinogenes*, revealed the time dependent changes in rumen bacterial populations after feeding the animals. The assay has been reported to be more rapid and accurate for monitoring total rumen fungal population in the rumen [78]. In another application, using real time PCR (targeting 16S rDNA), the effects on rumen cellulolytic populations in response to dietary ingredients manipulations were measured, which could have not been feasible using any classable culture methods [79]. Real time PCR coupled with cytological analysis tools like fluorescent in situ hybridization (FISH) could be more useful for species- and group-specific detection of rumen bacteria, and studying the dietary fiber degradation dynamics, the role played by fibrolytic bacterial genera, and even the phylogenetic groups within a single bacterial species [69].

Furthermore, animal nutritionists are interested in integrating the current knowledge of rumen functioning with a future perspective regarding how metagenomics could be used to integrate rumen microbiology and animal nutrition. A better understanding of mechanistic process altering the production and uptake of amino nitrogen will help the livestock nutritionists to improve the overall conversion of dietary nitrogen into microbial protein. It will provide key information needed to further improve mechanistic models describing rumen function and evaluating dietary conditions that influence the efficiency of conversion of dietary nitrogen into milk protein [73]. Another important application of metagenomic analysis in livestock production system would be the identification of rumen methanogens which have not been cultured till date. The work has been initiated and metagenomic analyses have proved to be a promising tool in identification of the methanogenic rumen archaea. A temporal temperature gradient gel electrophoresis method developed to determine the diversity of methanogens in cattle and sheep rumen showed that uncultured methanogens had significant population densities in each of the rumen samples examined [74]. The metagenomic analysis and understanding of the association patterns between archaea and rumen protozoa would be helpful in developing strategies to reduce methane emissions by dietary or genetic manipulation of the rumen ecosystem.

Additionally, metagenomics offers a huge potential for discovering new microbial proteins and enzymes with known functions, new proteins with novel functions, known proteins with unique functions, and other novel molecules. Metagenomics along with transcriptomics and proteomic analysis of the GI ecosystems would make it feasible to develop the functional genomics of important microbial communities, which finally aims to characterize the molecular

signs and various types of interactions, not only between the microorganisms, but with the host as well.

Bottlenecks of the technology

The technology has certain practical limitations that slow down its implementation to explore many microbial ecosystems. Only a very limited number of metagenomically-revealed genes and gene-products (enzymes) are currently in use in any of biotechnology processes. Furthermore, within the many novel DNA sequences identified, though new enzymatic functions are identified, but none of them has been isolated [80]. Furthermore, despite the hype of metagenomic approaches, emerging technologies and a revival in culturing efforts may make the metagenomic approaches unnecessary in many cases [81].

Conclusion and future prospects

Gut microbes produce a wealth of novel substances, and metagenomics has become a powerful tool appraising and exploiting them for biotechnological and pharmaceutical applications. The techniques are in the initial phase of their applications for exploring herbivores' GI microbial ecosystems. Applications of these techniques to the rumen ecosystem have primarily been focused on providing nutrient supply and utilization by the animals through three main components- supply of critical nutrients from the dietary forages, enhanced fiber digestion by the gut flora, and nutrients-host tissue interactions. The information derived from the molecular analyses of the gut flora would greatly facilitate the new isolation procedures for novel gut microflora. The long term goal of metagenomics is to reconstruct the genomes of unculturable gut microorganisms by identifying overlapping fragments in metagenomic libraries and 'walking', clone to clone, to assemble each chromosome. To exploit the maximum potential of biotechnological applications of the gut flora, it is essential that both basic biology and utility streams be pursued as a part of the new field of metagenomics.

References

1. Shelswell KJ (2004) <http://www.scq.ubc.ca/?p=509>
2. Cowan DA (2000) Microbial genomes- the untapped resource. Trends Biotechnol 18:14–16
3. Rappe MS and Giovannoni SJ (2003) The uncultured microbial majority. Annu Rev Microbiol 57:369–394
4. Stevenson BS, Eichorst SA, Wertz JT, Schmidt TM and Breznak JA (2004) New Strategies for cultivation and de-

- tection of previously uncultured microbes. *Appl Environ Microbiol* 70:4748–4755
5. Handelsman J, Rondon MR, Brady SF, Clardy J and Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5:R245–R249
 6. Lu J, Santo Domingo JW and Shanks OC (2007) Identification of chicken-specific fecal microbial sequences using a metagenomic approach. *Water Res* 41:3561–3574
 7. Shanks OC, Santo Domingo JW, Lamendella R, Kelty CA and Graham JE (2006) Competitive metagenomic DNA hybridization identifies host-specific microbial genetic markers in cow fecal samples. *Appl Environ Microbiol* 72: 4054–4060
 8. Amann RL, Ludwig W and Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59:143–169
 9. Ward N (2006) New directions and interactions in metagenomic research. *FEMS Microbiol Ecol Rev* 55:331–338
 10. Xu J (2006) Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol Ecol* 15:1713–1731
 11. Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R and Wittwer P (2005) Metagenomic gene discovery: past, present and future. *Trends Biotechnol* 23:321–329
 12. Schloss PD and Handelsman J (2003) Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* 14: 303–310
 13. Rees HC, Grant WD, Jones BE and Heaphy S (2004) Diversity of Kenyan soda lake alkaliphiles assessed by molecular methods. *Extremophiles* 8:63–71
 14. Hurt RA, Qiu X, Wu L, Roh Y, Palumbo AV, Tiedje JM and Zhou J (2001) Simultaneous recovery of RNA and DNA from soils and sediments. *Appl Environ Microbiol* 67: 4495–4503
 15. Meyer QC, Burton SG and Cowan DA (2007) Subtractive hybridization magnetic bead capture: a new technique for the recovery of full-length ORFs from the metagenome. *Biotechnol J* 2:36–40
 16. Zhou J, Bruns MA and Tiedje JM (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* 62:316–322
 17. Roose-Amsaleg CL, Garnier-Sillam E and Harry M (2001) Extraction and purification of microbial DNA from soil and sediment samples. *Appl Soil Ecol* 18:47–60
 18. Niemi RM, Heiskanen I, Wallenius K and Lindstrom K (2001) Extraction and purification of DNA in rhizosphere soil samples for PCR-DGGE analysis of bacterial consortia. *J Microbiol Methods* 45:155–165
 19. Alm EW, Zheng D and Raskin L (2000) The presence of humic substances and DNA in RNA extracts affects hybridization results. *Appl Environ Microbiol* 66:4547–4554
 20. Frischer ME, Danforth JM, Newton Healy MA and Saunders FM (2000) Whole-cell versus total RNA extraction for analysis of microbial community structure with 16S rRNA-targeted oligonucleotide probes in salt marsh sediments. *Appl Environ Microbiol* 66:3037–3043
 21. Griffiths RI, Whiteley AS, O'Donnel AG and Bailey MJ (2000) Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl Environ Microbiol* 66:5488–5491
 22. Bohannan BJ and Hughes J (2003) New approaches to analyzing microbial biodiversity data. Review. *Curr Opin Microbiol* 6:282–287
 23. Short JM and Mathur EJ (1999) Production and use of normalized DNA libraries. US Patent 6001574
 24. Radajewski S, Webster G, Reay DS, Morris SA, Ineson P, Nedwell DB, Prosser JI and Murrell JC (2002) Identification of active methylotroph populations in an acidic forest soil by stable-isotope probing. *Microbiology* 148:2331–2342
 25. Urbach ER, Vergin KL and Giovannoni SJ (1999) Immunochemical detection and isolation of DNA from metabolically active bacteria. *Appl Environ Microbiol* 65:1207–1213
 26. Bart A, Dankert J and van der Ende A (2000) Representational difference analysis of *Neisseria meningitidis* identifies sequences that are specific for the hyper-virulent lineage III clone. *FEMS Microbiol Lett* 188:111–114
 27. Galbraith EA, Antonopoulos DA and White BA (2004) Suppressive subtractive hybridization as a tool for identifying genetic diversity in an environmental metagenome: the rumen as a model. *Environ Microbiol* 6:928–937
 28. Green CD, Simons JF, Taillon BE and Lewin DA (2001) Open systems: panoramic views of gene expression. Review. *J Immunol Methods* 250:67–79
 29. 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
 30. 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
 31. McDonald IR, Kenna EM and Murrell JC (1995) Detection of methanotrophic bacteria in environmental samples with the PCR. *Appl Environ Microbiol* 61:116–121
 32. Kobayashi KS, Chamaillard M, Ogura Y, Henegariu O, Inohara N, Nunez G and Flavell RA (2005) Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* 307:731–734
 33. Clatworthy AE, Pierson E and Hung DT (2007) Targeting virulence: a new paradigm for antimicrobial therapy. Review. *Nat Chem Biol* 9:541–548
 34. Misra RN, Singla-Pareek SL, Nair S, Sopory SK and Reddy MK (2002) Directional genome walking using PCR. *Biotechniques* 33:830–834
 35. Myrick KV and Gelbart WM (2002) Universal fast walking for direct and versatile determination of flanking sequence. *Gene* 284:125–131
 36. Megonigal MD, Rappaport EF, Wilson RB, Jones DH, Whitlock JA, Ortega JA, Slater DJ, Nowell PC and Felix CA (2000) Panhandle PCR for cDNA: a rapid method for isolation of MLL fusion transcripts involving unknown partner genes. *Proc Natl Acad Sci USA* 97:9597–9602
 37. Ochman H, Ayala FJ and Hartl DL (1993) Use of polymerase chain reaction to amplify segments outside boundaries of known sequences. *Methods Enzymol* 218:309–321
 38. Eschenfeldt WH, Stols L, Rosenbaum H, Khambatta ZS, Quate-Randall E, Wu S, Kilgore DC, Trent JD and Donnelly

- MI (2001) DNA from uncultured organisms as a source of 2,5-diketo-D-gluconic acid reductases. *Appl Environ Microbiol* 67:4206–4214
39. Rondon MR, August PR, Bettrmann 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 and 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
 40. 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 metagenome library of soil microbial DNA. *Appl Environ Microbiol* 68:4301–4306
 41. Piel J (2002) A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetle. *Proc Natl Acad Sci USA* 99:14002–14007
 42. Diaz-Torres ML, McNab R, Spratt DA, Villedieu A, Hunt N, Wilson M, Mullany P (2003) Novel tetracycline resistance determinant from the oral metagenome. *Antimicrob Agents Chemother* 47:1430–1432
 43. Henne A, Schimitz RA, Bomeke M, Gottschalk G and 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
 44. Lammle K, Zipper H, Breuer M, Hauer B, Buta C, Brunner H and Rupp S (2007) Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *J Biotechnol* 127:575–592
 45. Lan PT, Sakamoto M, Sakata S and Benno Y (2006) *Bacteroides barnesiae* sp. nov., *Bacteroides salanitronis* sp. nov., and *Bacteroides gallinarum* sp. nov., isolated from chicken caecum. *Int J Syst Evol Microbiol* 56:2853–2859
 46. Ferrer M, Belouqui A, Golyshina OV, Plou FJ, Neef A, Chernikova TN, Fernandez-Arrojo L, Ghazi I, Ballesteros A, Elborough K, Timmis KN and Golyshin PN (2007) Biochemical and structural features of a novel cyclodextrinase from cow rumen metagenome. *Biotechnol J* 2:207–213
 47. Feng Y, Duan CJ, Pang H, Mo XC, Wu CF, Yu Y, Hu YL, Wei J, Tang JL and Feng JX (2007). Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. *Appl Microbiol Biotechnol* 75:319–328
 48. Belouqui A, Pita M, Polaina J, Martinez-Arias A, Golyshina OV, Zumarraga M, Yakimov MM, Garcia-Arellano H, Alcalde M, Fernandez VM, Elborough K, Andreu, JM, Ballesteros A, Plou FJ, Timmis KN, Ferrer M and Golyshin PN (2006) Novel polyphenol oxidase mined from a metagenome expression library of bovine rumen: biochemical properties, structural analysis, and phylogenetic relationships. *J Biol Chem* 281:22933–22942
 49. Gabor E, Liebeton K, Niehaus F, Eck J and Lorenz P (2007) Updating the metagenomics toolbox. *Biotechnol J* 2:201–206
 50. 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 (2000a) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* 2:516–529
 51. Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB, Gates CM, Feldman RA, Spudich JL, Spudich AN and DeLong EF (2000b) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* 289:1902–1906
 52. Beja O, Koonin EV, Aravind L, Taylor LT, Seitz H, Stein JL, Bensen DC, Feldman RA, Swanson RV and DeLong EF (2002) Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl Environ Microbiol* 68:335–345
 53. Quaiser A, Ochsenreiter T, Klenk HP, Kletzin A, Treusch AH, Meurer G, Eck J, Sensen, CW and Schelper C (2002) First insight into the genome of an uncultivated crenarchaeote from soil. *Environ Microbiol* 4:603–611
 54. Liles MR, Manske BF, Bintrim SB, Handelsman J and Goodman RM (2003) A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Appl Environ Microbiol* 69:2684–2691
 55. Huson DH, Auch AF, Qi J and Schuster SC (2007) MEGAN analysis of metagenomic data. *Genome Res* 17:377–386
 56. Jones BV and Marchesi JR (2007) Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome. *Nature Methods* 4:55–61
 57. Savage DC (1977) Microbial ecology of gastrointestinal tract. *Annu Rev Microbiol* 31:107–133
 58. O'Hara AM and Shanahan F (2007) Gut microbiota: mining for therapeutic potential. *Clin Gastroenterol Hepatol* 5: 274–284
 59. Goodacre R (2007) Metabolomics of a superorganism. *J Nutr* 137 (Suppl 1):259S–266S
 60. Lorenz P and Eck J (2005) Metagenomics and industrial applications. *Nat Rev Microbiol* 3:510–516
 61. Walter J, Mangold M and Tannock GW (2005) Construction, analysis, and β -glucanase screening of a bacterial artificial chromosome library from the large-bowel microbiota of mice. *Appl Environ Microbiol* 71: 2347–2354
 62. 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
 63. Beriberi M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P and Rohwer F (2003) Metagenomics of uncultured viral community from human feces. *J Bacteriol* 185: 6220–6223
 64. Selinger LB, Forsberg CW and Cheng KJ (1996) The rumen: a unique source of enzymes for enhancing livestock production. *Anaerobe* 2:263–284
 65. Singh B, Bhat TK and Singh B (2001) Exploiting gastrointestinal microbes for livestock and industrial development. *Asian-Aust J Anim Sci* 14:567–586
 66. Palackal N, Lyon CS, Zaidi S, Luginbuhl P, Dupree P, Goubet F, Macomber JL, Short JM, Hazlwood GP, Robertson DE and Steer BA (2007) A multifunctional hybrid glycosyl hydrolase discovered in an uncultured microbial consortium from ruminant gut. *Appl Microbiol Biotechnol* 74:113–124
 67. Leser TD, Amenuvor JZ, Jensen TK, Lindencrona RH, Boye M and Møller K (2002) Culture-independent analysis of gut bacteria: the pig gastrointestinal tract microbiota revisited. *Appl Environ Microbiol* 68:673–690

68. Stevenson DM and Weimer PJ (2007) Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. *Appl Microbiol Biotechnol* 75:165–174
69. Shinkai T and Kobayashi Y (2007) Localization of ruminal cellulolytic bacteria on plant fibrous materials as determined by fluorescence in situ hybridization and real-time PCR. *Appl Environ Microbiol* 73:1646–1652
70. Ferrer M, Golyshina OV, Chernikova T, Khachane AN, Reyes-Durate D, Santos VA, Strompl C, Elborough K, Jarvis G, Neef AN, Yakimov MM, Timmis KN and Golyshin PN (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ Microbiol* 7:1996–2010
71. Skillman LC, Toovey AF, Williams AJ and Wright AG (2006) Development and validation of a real-time PCR method to quantify rumen protozoa and examination and validity between *Entodinium* populations in sheep offered a hay-based diet. *Appl Environ Microbiol* 72:200–206
72. Nelson KE, Zinder SH, Hance I, Burr P, Odongo D, Wasawo D, Odenyo A and Bishop R (2003) Phylogenetic analysis of the microbial populations in the wild herbivore gastrointestinal tract: insight into an unexplored niche. *Environ Microbiol* 5:1212–1220
73. Firkins JL, Yu Z and Morrison M (2007) Ruminal nitrogen metabolism: perspectives for integration of microbiology and nutrition for dairy. *J Dairy Sci* 90 (Suppl 1):E1–E16
74. Nicholson MJ, Evans PN and Joblin KN (2007) Analysis of metahnogens diversity in the rumen using temporal temperature gradient gel electrophoresis: identification of uncultured methanogens. *Microb Ecol* 54:141–150
75. Chen R and LaPara TM (2006) Aerobic biological treatment of low-strength synthetic wastewater in membrane-coupled bioreactors: the structure and function of bacterial enrichment cultures as the net growth rate approaches zero. *Microb Ecol* 51:99–108
76. Akselband Y, Cabral C, Castor TP, Chikarmane HM and McGrath P (2006) Enrichment of slow-growing marine microorganisms from mixed cultures using gel microdrop (GMD) growth assay and fluorescent-activated cell sorting. *J Exp Marine Biol Ecol* 329:196–205
77. Kruger M, Wolters H, Gehre M, Joye SB and Richnow HH (2008) Tracking the slow growth of methane-oxidizing communities by (15)N-labeling techniques. *FEMS Microbiol Ecol* 63:401–411
78. Denman SE and McSweeney CS (2006) Development of a real-time PCR assay for monitoring anaerobic fungal and cellulolytic bacterial populations within the rumen. *FEMS Microbiol Ecol* 58:572–582
79. Mosoni P, Chaucheyras-Durand F, Bera-Maillet C and Forono E (2007) Quantification by real-time PCR of cellulolytic bacteria in the rumen of sheep after supplementation of a forage diet with readily fermentable carbohydrates: effect of a yeast additive. *J Appl Microbiol* 103:2676–2685
80. Schmeisser C, Steele H and Streit WR (2007) Metagenomics, biotechnology with non-culturable microbes. *Appl Microbiol Biotechnol* 75:955–962
81. Kowalchuk GA, Speksnijder AGC, Zhang K, Goodman RM and van Veen JA (2007) Finding the needles in metagenome haystack. *Microbial Ecol* 53:475–485