ORIGINAL RESEARCH PAPER

Volume-8 | Issue-6 | June-2019 | PRINT ISSN No. 2277 - 8179

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

METAGENOMICS: APPROACHES, RECENT ADVANCES AND APPLICATIONS

Zoology

Jyoti Research Scholar (UGC-JRF) at University of Delhi, Department of Zoology, Delhi-110007

ABSTRACT

Advancement in next generation sequencing (NGS) and development of high-performance bioinformatics tools have improved understanding of microbial communities. The massive data production and substantial cost reduction in NGS have led to the rapid growth of metagenomic research tools. The basic goal of any metagenomic project is the full characterization of a community i.e. “who's there? “what are they doing”? The science of metagenomics made it possible to investigate resource for the development of novel genes, enzymes, pathways and chemical compounds for use in biotechnology. Present review highlights the major breakthrough in the field of metagenomics, providing insights into current boundaries of the field and forthcoming challenges.

KEYWORDS

Metagenomics, Next-Generation sequencing, Amplicon sequencing, Shotgun sequencing.

1. INTRODUCTION

Microorganisms are present almost everywhere on the earth. Many challenges have been experienced during identification of unexplored entities and in their ways of interaction with their environment. Over a decade, metagenomics has been proven an advance field of research by explaining non-cultured microbes which represent the majority of life forms in most habitats of this planet. Metagenomics has risen as a strong weapon that can be used to study microbial communities despite inability of member organisms to be cultured in the laboratory using conventional isolation. Meta-genomics involves the extraction of the genetic material from the community so that genetic- material of all organisms in the community are pooled. Metagenomics is also described as environmental and community genomics that involves genome isolation from environmental sample, fragmentation and then cloned into plasmid which has capacity to replicate. Organisms are then cultured to create libraries and then the data is analysed. The two main and principally distinct outcomes of the metagenomic approach are the emerging new outlook at the complexity of the microbial communities, in forms of both species diversity and community dynamics, and identification of genetic determinants for production of biologically active molecules and processes that carry a potential for medical and biotechnological applications. This review intends to comprehensively explores the work on functional metagenomics, approaches and its applications.

2. BRIEF HISTORY OF METAGENOMICS

The history of metagenomics should probably be traced back to the work of Stanley and Konopka (1985), first reporting on 'great plate count anomaly', and the works of the Woase group, identifying the 16S rRNA gene as a marker gene for evaluating microbial diversity [1]. Later on, this idea of 16S rRNA marker gene analysis by Sanger sequencing method transformed the way microorganisms are classified taxonomically. The practical use of 16S rDNA analysis as a tool for phylogenetic profiling of microbial communities has been established by the Pace group in early nineties, constituting the onset of metagenomics as a sub-field of microbial biology [2]. Later the term 'metagenomics' was given by the Handelsman group [3]. The two influential works, first described the simple microbial community of an acid mine drainage [4] and the second one described a much more complex community of the Sargasso Sea [5]. Together they defined the most widely accepted meaning of metagenomics. Whole-genome shotgun (WGS) sequencing based analysis of microbial diversity in environmental samples, were published in 2004. Over the same time, advancement of alternative sequencing technologies like NGS, promising a significantly higher throughput and significantly reduced the cost of sequencing, known as next generation sequencing technologies have been tested in metagenomic applications [6]. These new sequencing technologies have also facilitated new fields of metagenomics, titled meta-transcriptomics (i.e. shotgun characterization of environmental transcripts) and meta-proteomics (i.e. analysis of community protein content) and marked the arrival of a whole new era of metagenomics.

3. Next Generation Sequencing (ngs) Technologies

The Sanger sequencing method is considered as 'first generation' technology while the advanced methods are known as 'next-generation sequencing' (NGS) technologies. It has been more than a decade to arrival of the next generation sequencing technologies. These technologies rely on template preparation, attachment to the solid surface, sequencing and imaging, base calling and quality control. Roche 454 is the world's first second generation sequencer. 454 technologies are based on sequencing-by-synthesis, with the release of pyrophosphate (PPi) from the DNA polymerization reaction as a consequence of nucleotide incorporation and used to generate light signal [7]. The read length of Roche 454 was initially 100-150 bps and 200000+ reads per run which is upgraded up to read length of 700 bps and output of 0.7GB data per run in 454 GS FLX system [8]. However, insertion-deletion errors in the homopolymeric regions are the main limitation of this technique. In 2006, Solexa/Illumina technology became one of the most active technologies because its low cost and high yields. This sequencer is based on reversible- termination sequencing by synthesis with fluorescently labelled nucleotides. The DNA fragments with fixed adaptors are denatured to single strand templates and then grafted to the flowcell, where the sequencing reaction takes place by incorporating a labelled nucleotide. The fluorophore of incorporated nucleotide is excited by laser and signal is captured by charged – coupled device (CCD).

Table1 | Comparison among sequencing technologies suitable for metagenomics

	Roche 454	SOLiD	Illumina MiSeq	PacBio RS2
Average read length (bp)	700	75	150	>1000
Output per run (GB)	0.7	180	1.5	1
Amplification for library construction	Yes	Yes	Yes	No
Error rate (%)	1	0.01	0.1	13
Run time	24h	14days	27h	2h
Pros	Large reads, short run time	Error correction	Short run time	Long reads, short run time
Cons	Homopolymeric errors	Short reads, long run time	Expensive	High error rate

Thereafter, fluorophore molecule is cleaved and followed by incorporation of next nucleotide [9]. The drawback of Illumina is substitution errors and occurs more frequently in the distal part of reads. In 2007, Applied Biosystems (ABI) launched the Sequencing by Oligo Ligation Detection (SOLiD) based on the nucleotide probe ligation and two-base encoding system [9]. In SOLiD, eight-nucleotide probe has five specific nucleotides complementary to template and rest three are non-specific bases. The probes are fluorescently labelled using a scheme for two-base encoding with four fluorophores. During sequencing reaction, probes containing all

possible combinations of first five bases are added. The probe with perfect match is hybridised and ligated and after imaging the fluorescent label and last three bases are cleaved off, and a new set of probes is added. Sequencing error rate is very low in case of SOLiD as each base is sequenced twice [10]. These techniques can sequence millions of molecules in parallel and eliminate the need of fragment-cloning methods and detect sequencing output directly without electrophoresis. The major pitfall of these second-generation sequencing technologies is short read length which create problem in assembly and metagenomics analysis [11]. Also, standard PCR is used to randomly amplify genomic fragments; hence there are chances of relative deviation in amplification of fragments, resulting in accuracies in gene expression analysis. To overcome these limitations of second-generation technologies, third generation sequencing technologies was developed. Third generation sequencing technologies allow single molecule based analysis as well as enable real time sequencing. At present, Nanopore [12], and SMRT [13] are single molecules, real-time sequencing technologies that lower amplification bias generated by PCR and short read length problem.

4. ANALYSIS METHODS OF METAGENOMICS

4.1 AMPLICON SEQUENCING ANALYSIS:

Amplicon sequencing method is based on the amplification of marker genes found in different organisms in the environmental sample. As 16SrRNA gene is highly conserved among prokaryotes (bacteria and archea), it is widely used as genetic marker to analyze the bacterial phylogeny and taxonomy. However, in case of fungi and eukaryote internal transcribed spacer (ITS) and 18SrRNA genes respectively are used as marker genes.

Marker gene profiling relies on PCR to amplify a range of different microorganisms as a wide as possible. Introduction of biases by 1) PCR as selection against particular group of organisms, 2) presence of different copy number of 16SrRNA genes in bacterial genomes and, 3) horizontal transfer among the organisms influence the relative abundance of microorganism in the sample. Mostly bioinformatic pipelines used to analyze amplicons of marker genes are designed for Roche 454 sequences. QIIME [14] Mothur [15] MEGAN [16] are some open source packages for amplicon analysis. Any typical pipeline involves three basic steps. First, filtration of sequence data through several quality filters as read length, phred score, low complexity sequences and removal of adaptor sequences. Second, clustering of sequences into operational taxonomic units based on their sequence similarity at particular level of taxonomy and then searched in reference databases to get closest match to an OTU to assign taxonomy classification. Some widely used databases are Greengenes for 16S [17], RDP [18], Silva for 16S and 18S [19] and Unite for ITS [20]. Third and final step is quantifying microbial diversity of the sample from resulting data.

4.2 SHOTGUN METAGENOMICS

a) Assembly

Assembly is a process of merging overlapped short reads generated from NGS technologies into larger genomic contigs and orientation of these into scaffolds. In metagenomic studies, for accurate functional annotation, it is advantageous to use large contigs as assembled data increases the probability of complete genes (or operons) reconstructions and also simplify the analysis by mapping long contigs instead of short reads [21]. Basically, there are two approaches for assembling short reads into contigs: 1) reference-based assembly, 2) de novo assembly. Earlier in metagenomic studies, reference-based approach was applied to metagenome assembly process. The reference-based assembly is based on the use of reference metagenomes as a "map" for generating contigs, which can represent genomes or part of genomes that belong to a particular species or genus. Newbler, MIRA 4 and AMOS are commonly used for referenced-based assemblies.

While De novo assembly concern to the generation of assembled contigs without using prior reference of known genome. De novo assemblers are preferred with low coverage and complex metagenome data. MetaVelvet [22], Ray-Meta [23] and Meta-IDBA [24] are some de novo assemblers. MetaVelvet and Met-IDBA decompose de-Bruijn graph based on k-mer frequencies and then assemble each subgraph. While Ray Meta is based on heuristics-guided graph approach to find optimal assembly [25].

b) Binning

Taxonomic classification of reads or grouping reads or contigs into individual genomes and assigning the groups to specific genus or species is called binning. Binning can be carried out using either reads or contigs. Two different binning strategies to get taxonomical classification are: 1) composition-based binning and, 2) similarity or homology-based binning. The former is based on k-mer frequencies i.e. individual genomes involve a unique allocation of k-mer sequences considered as "genomic signatures". Tetra [26] and MetaclusterTA [27] are the tools that work based on composition-based binning. Some other tools like Amphora and Maxbin rely on the k-mer signatures as well as GC content and coverage information. While homology-based methods use alignment algorithms such as hidden Markov Models and BLAST to map reads directly to individual reference genomes or pen-genomes. CARMA [28] and Megan [29] are currently used reference-based binning tools.

c) Gene Prediction and Functional Assignment

The fundamental step for functional annotation is gene prediction or gene calling i.e. identification of genes within the reads/contigs. MetaGeneMark [30] and Glimmer-MG [31] are some of the coding gene predictor that utilize different models for gene prediction such as Hidden Markov Models, di-codon usage. MetaGeneMark uses codon usage-incorporated HMM; MetaGeneMark is based on di-codon usage and FragGeneScan [32] uses sequencing error model and codon usage-incorporated HMM. Noncoding genes such as tRNAs and rRNAs genes are predicted using pipelines such as tRNAscan [33] and RDP [18] respectively.

Next step of functional annotation is the functional assignment to the predicted protein coding genes. This can be achieved by using homology-based searches (HMM) and similarity-based searches (BLAST). IMG-MER, MG-RAST, CAMERA [34]. IMG/MER [35] utilizes HMM search to associate genes to PFAM. It can utilize its own sources, using them as reference non-redundant databases from which it retrieves additional functional annotation. While MG-RAST [36] utilizes other databases for annotation as well as taxonomy classification.

d) Metabolic Pathways Reconstruction and Comparative Metagenomics

Metabolic pathways reconstruction is generally performed using the KEGG database [37]. Computational tools such as IMG-M, CAMERA AND MG-RAST use KEGG database and KEGG graphs, which allow analyses beyond annotation like taxonomy classification and pathways reconstructions. METATREP is a tool to compare annotated metagenomes by graphical information for taxonomical and functional classification [25]. Minpath [38] and MetaPath [39] are currently used programs to construct metabolic pathways utilize information stored in KEGG and MataCyc repositories.

Finally, relationship between different environments and the microbial population and their environments can be established by comparing the taxonomy and functional profiling. Parallel-meta [40] and MEGAN [29] are the useful computational tools to perform comparative metagenomics.

Table 2 | Metagenomic tools according to their functionality.

a) Marker gene analysis- Standalone softwares	QIIME Mothur Unifrac Megan
Databases	Greengenes RDP SILVA UNITE
b) Shotgun metagenomics Assembly	Velvet Metavelvet MIRA Meta-IDBA RayMeta ABySS Newbler
Binning	TETRA MetaClusterTA Amphora MaxBin CARMA Megan MetaPhyler

Annotation	MetaGeneMark Glimmer-MG RDP tRNAscan FragGeneScan metaGene
Analysis pipelines	IMG/MER MG-RAST Megan MetaTrep

5. APPLICATIONS

In microbiology, metagenomics has emerged as dynamic tool and has changed the way by which a microbiologist solves many problems. It has been estimated that less than 1% of the microorganisms in the natural environment can be cultured in laboratory. It is increasingly recognised that a huge number of natural products exists in non-cultural microbes with chemical, biological activities with potential uses in various industrial and biomedical applications [41]. Metagenomics provides an unlimited resource for the development of novel genes, enzymes, natural products, bioactive compounds, and bioprocesses.

5.1 METAGENOMICS AND NOVEL ENZYMES

Metagenomics has proven powerful approach for the ample demand of novel enzymes and biocatalysts [42]. Some industrially important novel enzymes such as cellulases, proteases, xylanases have been identified by using metagenomic approaches. Some of the main enzymes that have been unlocked from genetically unexploited resources have been described below.

A) Cellulases

Cellulase, a multi-component enzyme refers to class glycosyl hydrolases (GHs) that perform three major types of enzymatic activities: (1) endoglucanases (EC 3.2.1.4), (2) exoglucanases (3) β-glucosidases (EC 3.2.1.21) [43]. Because of wide diversity of cellulases applications, they have gained much interest. They are extensively used for processing of food and fruit drinks, textile area, improving the nutritional quality and digestibility of animal feeds, in processing of fruit juices. Besides all these, now a day's use of cellulases in paper industry for de-inking of papers' is yet another emerging application [44]. Cellulases have been isolated from numerous natural sites as soil, rumen, compost soil and many more using metagenomic techniques by constructing the metagenomic libraries followed by screening of the biologically functional clones. Many researchers reported isolation of cellulose enzymes from niche environments which include anaerobic digester, alkaline and saline lakes [45]. In 2013, Yeh and his coworkers [46] reported GH12 cellulase gene, RSC-EGI, encodes for 464 amino acids protein along with two other ORFs was extracted from metagenomic library derived from rice straw compost. This protein has more than 70% similarity at the amino acid level with cellulose from *Micromonospora aurantiaca* and *Thermobispora sp.* Kyong and his colleagues [47], in 2013, identified and characterized a novel CelEx-BR12 gene had an open reading frame (ORF) of 1140bps that encoded for a 380-amino acid-protein from ruminal bacteria using a robotic, high-throughput screening system. The observed enzyme is multifunctional as endocellulase/exocellulase/xylanases activities were observed against fluorogenic and natural glycosides, such as 4-methylumbelliferyl-β-d-cellobioside (0.3Umg (-1)), birch wood xylan (132.3U mg (-1), oat split xylan (67.9 U mg (-1)), and 2-hydroxyethyl-cellulose (26.3 U mg (-1)) that may useful for biotechnological applications in industrial area. Another novel enzyme halotolerant cellulase Cel5A, soil metagenome-derived enzyme (endoglucanase), is highly stable and also active at high pH and salt concentration revealing the importance of metagenomic cellulases in industrial applications [48].

B) Lipases

Lipases are triacylglycerol acylhydrolases that catalyzed the hydrolysis of triacylglycerol to glycerol and fatty acids. Due to their hostility on extreme conditions such as temperature, pH, organic solvents, they have been obtained special industrial attention. Due to their potential significance in industry largely in oil and fats, detergents, dairy and pharmaceutical industries the lipases from microbial sources like bacteria, yeast and fungi are presently earning particular attention [49]. Lipases have been reported using unique metagenomes e.g. from thermal environments, from saline lakes, and from marine sediments. Florence and his collegues, reported five solid-attached and four liquid-associated rumen bacteria clones exhibited lipolytic activity. Isolated lipases/esterases were shown activity against mostly short to medium-chain substrates over a spectrum of temperatures and pH indicating their potential industrial field [50]. Recently, three novel genes conferring lipolytic activity were identified from Kamchatka thermal spring volcanic metagenome [51]. Few lipolytic enzymes also have been extracted from marine sea sediments; one lipase h1Lip1 from metagenomic libraries of the Baltic Sea [52] and two esterases (estAT1 and estAT11) from the Arctic seashore [53]. In both cases metagenomic libraries were constructed using fosmid vectors. Out of these lipolytic enzymes, estearses, EstAT11, preferentially hydrolyzed (S)-racemic olfozacin butyl ester showing 70% enantiomeric excess value, displaying great potential for the chiral resolution of heat-liable substrates. Unique lipolytic enzymenamed Lipo l, was isolated from a metagenomic library of activated sludge source. Lipo l showed highest activity at temperature of 10°C and pH 7.5, decreases at temperature above 50°C, and resistance in presence of detergents, which might be useful as biocatalyst in organic chemistry and laundry industry [54]. EstATII, unusual esterase was isolated from Red Sea Atlantis 2 brine pool using metagenomic functional approach. EstATII was found thermophilic, halotolerant active up to 4.5M NaCl and maintains at least 60% of its activity in the presence of toxic (heavy) metals. These unique biochemical characteristics of the Red Sea Atlantis 2 brine pool isolated extremophilic esterases, i.e., halotolerance, thermophilicity, and resistance to heavy metals, shows its potential as biocatalyst [55].

C) Proteases

Proteases act as an invincible player in industrial biotechnology, especially in detergent, food and pharmaceutical field. Though, large numbers of proteases present in a wide range of living organisms; proteases from microbial origins have often been reported to have distinct prevalence because they have almost all characteristics desired for biotechnological applications. Several proteases have been discovered using metagenomic approaches in the recent times. Nowadays, alkaline-based detergents are used because of their better cleaning properties due to high stability and activity of proteases under harsh conditions such as elevated temperature and pH [56]. Protease from metagenomic DNA isolated from goat skin surface and metagenomic libraries were constructed using pSP1 plasmid. Upon screening, clones carrying recombinant plasmid pSP1 exhibited protease activity. The cloned insert DNA has an ORF of 1890bps encodes for 630 amino acids showed significant homology with peptidase S8 of *Shewalenna sp*. Although alkaline Serine protease (AS-protease) was over expressed and found inactive resulting in inclusion bodies formation [57]. Neveu *et al*. 2011, also isolated two serine proteases DV1 and M30 from metagenomic libraries derived from surface sand of the Gobi and Death Valley deserts. These proteases seem to belong to subtilisin family and exhibited unique biochemical properties. Protease DV1exhibit optimal activity at pH 8 and temperature 55°C while M30 works good at pH >11 and temperature 40°C [58].

d) Amylases

Amylases are starch degrading enzymes. These enzymes play potential roles in a number of industrial processes like food, fermentation and pharmaceutical industries for the hydrolysis of starch. An exceptionally cold-adopted alpha amylase Amy13C6, from a metagenomic library of cold and alkaline environment was purified and was shown to have significant homology to α-amylases from the class Clostridia. The enzyme was tested against two commercial detergents and was found that enzymes displayed activity in both of them, suggesting that the Amy13C6α-amylase may be effective as a detergent enzyme in environment friendly, low temperature laundry processes. Sharma *et al*., 2010, discovered a novel amylase from a soil metagenome having maximum activity at 40°C under neutral pH whereas retained 90% of activity even at low temperature suggesting its potential candidature for possible industrial applications [59]. A thermostable and calcium-dependent amylase was isolated from a soil metagenome and suggested its utilization in baking and de-starching [60].

5.2 METAGENOMICS AND BIODEGRADATION

Different kinds of waste such as petroleum discharge and incomplete explosion of fossils fuels, produced by industries are responsible for accretion of petroleum hydrocarbons in the environment. The generation of these anthropogenic compounds, through oil-related production, introduces ample amount of aromatic hydrocarbons into the surrounding, resulting in the contamination of environment [61]. Microorganisms are directly involved in the biogeochemical cycles

and responsible for the degradation of many carbon sources, can breakdown aromatic rings, like those of benzene, toluene, and mineralize their carbon skeleton [62]. Therefore, metagenomics is a tool that eliminates cultivation steps, as it consists of direct extraction of environmental DNA and its cloning into suitable vector. Isabel and coworkers in 2014 reported the potential of the metagenomic functional approach for the identification and characterization of new genes and the pathways in rarely studied environments and provide a broader aspect on the hydrocarbon degradation processes in oil reservoir [63]. Recently, researchers studied metagenomic libraries constructed from sludge DNA sample and obtained more than 400 positive clones out of 13,200 clones for phenol degradation. These clones were selected to evaluate potential of microorganisms present in wastewater treatment plant for degradation, focusing mainly on genes and their metabolic pathways related to degradation of phenol and other aromatic compounds in sludge samples [64].

5.3 Metagenomics And Fecal Microbiota Transplants

The existence of symbiotic relationship between gut microbiota and human health is well established and human intestine is known to harbor around 10^14 microbes with 3500 different species [65]. Human gut microbiota plays significant role in postnatal structural and functional maturation of gut, development of immune system and nervous system. Imbalance of micro-biota in body can lead to diseases such as antibiotic-associated diarrhea. Fecal micro-biota transplantation has proven to be valuable in restoring the disturbed micro-biota. At the first time, Ge Hong used fecal transplantation in treating food poisoning [66]. There are many clinical reports on using FMT for disease conditions such as autism, depression, obesity, and multiple sclerosis [67]. In this scenario metagenomics play an important role by determining the microbial diversity in both healthy and diseased gut before and after the microbial transplantation.

7. CONCLUSION

New emerging field of next generation sequencing (NGS) based metagenomics provide an access to the complete genetic material of microbial community of an environment. NGS based methods rule out the limitations of conventional genomic methods which are based on culture dependent isolation of genetic material. Metagenomics allows the identification of new genes, proteins and metabolic pathways with better accuracy than conventional molecular methods. NGS based methods have provided thorough understanding of genomics of microbial communities involved in degradation of xenobiotics. Also, mining of metagenomes has resulted in exploration of various novel enzymes like cellulases, proteases, lipases and amylases. Metagenomics of human gut resulted in identification of microbiome which has eventually led to the development of treatment technology called as Fecal Microbiota Transplant. This has been successfully employed for the treatment of diseases like colitis, obesity, and multiple sclerosis. In near future, metagenomics will be essential and elemental part of genome-based studies for comprehensive understanding of our environment in coordination with other omics-based methods like meta-transcriptomics and meta-proteomics.

REFERENCES

[1] Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proceedings of the National Academy of Sciences, 74(11), 5088-5090

[2] Schmidt, T. M., DeLong, E. F., & Pace, N. R. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. Journal of bacteriology, 173(14), 4371-4378.

[3] Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chemistry & biology, 5(10), R245-R249.

[4] Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., & Banfield, J. F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature, 428(6978), 37-43.

[5] Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., & Fouts, D. E. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. science, 304(5667), 66-74.

[6] Escobar-Zepeda, A., de León, A. V. P., & Sanchez-Flores, A. (2015). The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. Frontiers in genetics, 6.

[7] Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., & Dewell, S. B. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature, 437(7057), 376-380.

[8] Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L., & Welch, D. M. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. Genome biology, 8(7), R143.

[9] Mardis, E. R. (2008). Next-generation DNA sequencing methods. Annu. Rev. Genomics Hum. Genet., 9, 387-402.

[10] Berglund, E. C., Kiialainen, A., & Syvänen, A. C. (2011). Next-generation sequencing technologies and applications for human genetic history and forensics. Investigative genetics, 2(1), 23.

[11] van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. Trends in genetics, 30(9), 418-426.

[12] Derrington, I. M., Butler, T. Z., Collins, M. D., Manrao, E., Pavlenok, M., Niederweis, M., & Gundlach, J. H. (2010). Nanopore DNA sequencing with MspA. Proceedings of the National Academy of Sciences, 107(37), 16060-16065.

[13] Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., & Bibillo, A. (2009). Real-time DNA sequencing from single polymerase molecules. Science, 323(5910), 133-138.

[14] Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., & Huttley, G. A. (2010). QIIME allows analysis of high-throughput community sequencing data. Nature methods, 7(5), 335-336.

[15] Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., & Sahl, J. W. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and environmental microbiology, 75(23), 7537-7541.

[16] Huson, D. H., & Weber, N. (2013). Microbial community analysis using MEGAN. In Methods in enzymology (Vol. 531, pp. 465-485). Academic Press.

[17] DeSantis, T. Z., Hugenholtz, P., Keller, K., Brodie, E. L., Larsen, N., Piceno, Y. M., & Andersen, G. L. (2006). NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic acids research, 34(suppl_2), W394-W399.

[18] Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., & Tiedje, J. M. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic acids research, 42(D1), D633-D642.

[19] Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., & Glockner, F. O. (2013). The 658 SILVA ribosomal RNA gene database project: improved data processing and web-based tools. 659 Nucleic Acids Res 41. D590-596, 660.

[20] Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F., Bahram, M., & Douglas, B. (2013). Towards a unified paradigm for sequence–based identification of fungi. Molecular ecology, 22(21), 5271-5277.

[21] Segata, N., Boernigen, D., Tickle, T. L., Morgan, X. C., Garrett, W. S., & Huttenhower, C. (2013). Computational meta'omics for microbial community studies. Molecular systems biology, 9(1), 666.

[22] Namiki, T., Hachiya, T., Tanaka, H., & Sakakibara, Y. (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic acids research, 40(20), e155-e155.

[23] Boisvert, S., Raymond, F., Godzaridis, É., Laviolette, F., & Corbeil, J. (2012). Ray Meta: scalable de novo metagenome assembly and profiling. Genome biology, 13(12), R122.

[24] Peng, Y., Leung, H. C., Yiu, S. M., & Chin, F. Y. (2011). Meta-IDBA: a de Novo assembler for metagenomic data. Bioinformatics, 27(13), i94-i101.

[25] Kim, M., Lee, K. H., Yoon, S. W., Kim, B. S., Chun, J., & Yi, H. (2013). Analytical tools and databases for metagenomics in the next-generation sequencing era. Genomics & informatics, 11(3), 102-113.

[26] Teeling, H., Waldmann, J., Lombardot, T., Bauer, M., & Glöckner, F. O. (2004). TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. BMC bioinformatics, 5(1), 163.

[27] Wang, Y., Leung, H. C. M., Yiu, S. M., & Chin, F. Y. L. (2014). MetaCluster-TA: taxonomic annotation for metagenomic data based on assembly-assisted binni Krause, L., Diaz, N. N., Goesmann, A., Kelley, S., Nattkemper, T. W., Rohwer, F., & Stoye, J. (2008). Phylogenetic classification of short environmental DNA fragments. Nucleic acids research, 36(7), 2230-2239.ng. BMC genomics, 15(1), S12.

[28] Krause, L., Diaz, N. N., Goesmann, A., Kelley, S., Nattkemper, T. W., Rohwer, F., & Stoye, J. (2008). Phylogenetic classification of short environmental DNA fragments. Nucleic acids research, 36(7), 2230-2239.

[29] Huson, D. H., Mitra, S., Ruscheweyh, H. J., Weber, N., & Schuster, S. C. (2011). Integrative analysis of environmental sequences using MEGAN4. Genome research, 21(9), 1552-1560.

[30] Zhu, W., Lomsadze, A., & Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. Nucleic acids research, 38(12), e132-e132.

[31] Kelley, D. R., Liu, B., Delcher, A. L., Pop, M., & Salzberg, S. L. (2012). Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. Nucleic acids research, 40(1), e9-e9.

[32] Rho, M., Tang, H., & Ye, Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. Nucleic acids research, 38(20), e191-e191.

[33] Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic acids re Cole, J. R., Chai, B., Farris, R. J., Wang, Q., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., & Tiedje, J. M. (2007). The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. Nucleic acids research, 35(suppl 1), D169-D172.search, 25(5), 955-964.

[34] Seshadri, R., Kravitz, S. A., Smarr, L., Gilna, P., & Frazier, M. (2007). CAMERA: a community resource for metagenomics. PLoS Biol, 5(3), e75.

[35] Markowitz, V. M., Chen, I. M. A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M., & Huntemann, M. (2014). IMG/M 4 version of the integrated metagenome comparative analysis system. Nucleic Acids Research, 42(D1), D568-D573.

[36] Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., & Wilkening, J. (2008). The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC bioinformatics, 9(1), 386.

[37] Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research, 28(1), 27-30.

[38] Ye, Y., & Doak, T. G. (2009). A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. PLoS Comput Biol, 5(8), e1000465.

[39] Liu, B., & Pop, M. (2010, May). Identifying differentially abundant metabolic pathways in metagenomic datasets. In International Symposium on Bioinformatics Research and Applications (pp. 101-112). Springer Berlin Heidelberg.

[40] Su, X., Pan, W., Song, B., Xu, J., & Ning, K. (2014). Parallel-META 2.0: enhanced metagenomic data analysis with functional annotation, high performance computing and advanced visualization. PLoS One, 9(3), e89323.

[41] Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. Microbiol. Mol. Biol. Rev., 68(4), 669-685.

[42] Lorenz, P., Liebeton, K., Niehaus, F., & Eck, J. (2002). Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. Current opinion in biotechnology, 13(6), 572-577.

[43] Li, Y. H., Ding, M., Wang, J., Xu, G. J., & Zhao, F. (2006). A novel thermoacidophilic endoglucanase, Ba-EGA, from a new cellulose-degrading bacterium, Bacillus sp. AC-1. Applied Microbiology and Biotechnology, 70(4), 430-436.

[44] Soni, R., Chadha, B., & Saini, H. S. (2008). Novel sources of fungal cellulases of thermophilic/thermotolerant for efficient deinking of composite paper waste. Bioresources, 3(1), 234-246.

[45] Rees, H. C., Grant, S., Jones, B., Grant, W. D., & Heaphy, S. (2003). Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. Extremophiles, 7(5), 415-421.

[46] Yeh, Y. F., Chang, S. C. Y., Kuo, H. W., Tong, C. G., Yu, S. M., & Ho, T. H. D. (2013). A metagenomic approach for the identification and cloning of an endoglucanase from rice straw compost. Gene, 519(2), 360-366.

[47] Ko, K. C., Lee, J. H., Han, Y., Choi, J. H., & Song, J. J. (2013). A novel multifunctional cellulolytic enzyme screened from metagenomic resources representing ruminal

bacteria. Biochemical and biophysical research communications, 441(3), 567-572.
[48] Voget, S., Steele, H. L., & Streit, W. R. (2006). Characterization of a metagenome-derived halotolerant cellulase. Journal of biotechnology, 126(1), 26-36.
[49] Cardenas, F., Alvarez, E., de Castro-Alvarez, M. S., Sanchez-Montero, J. M., Valmaseda, M., Elson, S. W., & Sinisterra, J. V. (2001). Screening and catalytic activity in organic synthesis of novel fungal and yeast lipases. Journal of Molecular Catalysis B: Enzymatic, 14(4), 111-123.
[50] Privé, F., Newbold, C. J., Kaderbhai, N. N., Girdwood, S. G., Golyshina, O. V., Golyshin, P. N., ... & Huws, S. A. (2015). Isolation and characterization of novel lipases/esterases from a bovine rumen metagenome. Applied microbiology and biotechnology, 99(13), 5475-5485.
[51] Daniel, R. (2013). Microbial diversity and biochemical potential encoded by thermal spring metagenomes derived from the Kamchatka Peninsula. Archaea, 2013.
[52] Hårdeman, F., & Sjöling, S. (2007). Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. FEMS microbiology ecology, 59(2), 524-534.
[53] Jeon, J. H., Kim, J. T., Kang, S. G., Lee, J. H., & Kim, S. J. (2009). Characterization and its potential application of two esterases derived from the arctic sediment metagenome. Marine biotechnology, 11(3), 307-316.
[54] Roh, C., & Villatte, F. (2008). Isolation of a low−temperature adapted lipolytic enzyme from uncultivated micro−organism. Journal of applied microbiology, 105(1), 116-123.
[55] Mohamed, Y. M., Ghazy, M. A., Sayed, A., Ouf, A., El-Dorry, H., & Siam, R. (2013). Isolation and characterization of a heavy metal-resistant, thermophilic esterase from a Red Sea Brine Pool. Scientific reports, 3, 3358.
[56] Jaouadi, B., Abdelmalek, B., Jaouadi, N. Z., & Bejar, S. (2011). The Bioengineering and Industrial Applications of Bacterial Alkaline Proteases: The Case of SAPB and KERAB. INTECH Open Access Publisher.
[57] Pushpam, P. L., Rajesh, T., & Gunasekaran, P. (2011). Identification and characterization of alkaline serine protease from goat skin surface metagenome. AMB express, 1(1), 3.
[58] Neveu, J., Regeard, C., & DuBow, M. S. (2011). Isolation and characterization of two serine proteases from metagenomic libraries of the Gobi and Death Valley deserts. Applied microbiology and biotechnology, 91(3), 635-644.
[59] Sharma, S., Khan, F. G., & Qazi, G. N. (2010). Molecular cloning and characterization of amylase from soil metagenomic library derived from Northwestern Himalayas. Applied microbiology and biotechnology, 86(6), 1821-1828.
[60] Gillespie, D. E., Brady, S. F., Bettermann, A. D., Cianciotto, N. P., Liles, M. R., Rondon, M. R., & Handelsman, J. (2002). Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. Applied and environmental microbiology, 68(9), 4301-4306.
[61] Nazir, A. (2016). Review on Metagenomics and its Applications. Imperial Journal of Interdisciplinary Research, 2(3), 277-286.
[62] Alexander, M. (1999). Biodegradation and bioremediation. Gulf Professional Publishing.
[63] Sierra-García, I. N., Alvarez, J. C., de Vasconcellos, S. P., de Souza, A. P., dos Santos Neto, E. V., & de Oliveira, V. M. (2014). New hydrocarbon degradation pathways in the microbial metagenome from Brazilian petroleum reservoirs. PloS one, 9(2), e90087.
[64] Silva, C. C., Hayden, H., Sawbridge, T., Mele, P., De Paula, S. O., Silva, L. C., & Santiago, V. M. (2013). Identification of genes and pathways related to phenol degradation in metagenomic libraries from petroleum refinery wastewater. PloS one, 8(4), e61811.
[65] Frank, D. N., Amand, A. L. S., Feldman, R. A., Boedeker, E. C., Harpaz, N., & Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proceedings of the National Academy of Sciences, 104(34), 13780-13785.
[66] Zhang, F., Luo, W., Shi, Y., Fan, Z., & Ji, G. (2012). Should we standardize the 1,700-year-old fecal microbiota transplantation?. The American journal of gastroenterology, 107(11), 1755-author.
[67] Lee, W. J., Lattimer, L. D., Stephen, S., Borum, M. L., & Doman, D. B. (2015). Fecal microbiota transplantation: a review of emerging indications beyond relapsing clostridium difficile toxin colitis. Gastroenterology & hepatology, 11(1), 24.