EasyChair Preprint

№ 14610

Metagenomics a Multifurcated Approach to Retrieve the Untapped Resources

Jeyaraj Pandiarajan, Muthukalingan Krishnan,
Thangaiyan Suganya and Karuppasamy Muthuvel Prasath

August 30, 2024

Metagenomics a multifurcated approach to retrieve the untapped resources

Jeyaraj Pandiarajan[1&2*], Muthukalingan Krishnan[2&3], Thangaiyan Suganya[2], Karuppasamy Muthuvel Prasath[1].

[1] Post Graduate Department of Biotechnology,
Ayya Nadar Janaki Ammal College,
Sivakasi 626124,
Tamil Nadu,
India

[2]Department of Environmental Biotechnology,
School of Environmental Sciences,
Bharathidasan University,
Tiruchirappalli 620024,
Tamil Nadu,
India

[3]Central University of Tamil Nadu, Neelakudi,
Thiruvarur - 610 005, India.

***Author for Correspondence**

Dr. Jeyaraj Pandiarajan
Assistant Professor
Post Graduate Department of Biotechnology,
Ayya Nadar Janaki Ammal College,
Sivakasi 626124,
Tamil Nadu,
India
E-mail: pandiarajan_sf424@anjaconline.org;
jeyapandiyah@gmail.com

Abstract

Metagenomic studies facilitate the detection of microbial communities without culturing of bacteria; in addition, direct molecular methods preserve the *in situ* metabolic status and microbial community composition-because samples are frozen immediately after acquisition. Also direct extraction of nucleic acids from experimental samples accounts for the very large proportions of microorganisms (90-99.9%) that are not readily cultured in the laboratory, but that may be responsible for the majority of the biodegradable activity of interest. The development of metagenomics and the application of high throughput technologies for handling and screening of very large numbers of clones represent important factors to increase speed and efficacy of enzyme discovery, thus accelerating and broadening the establishment of biocatalytic applications. In addition to application-oriented benefits, phylogenetic analysis of novel genes could also provide insight into evolutionary affinity, natural distribution and development of distinct enzymatic activities and could serve to elucidate evolutionary structure-function relationships. When combined with classic microbiological methods, these molecular biological methods and metagenomics approach will pave a new way to explore novel catabolic genes responsible for degradation of toxic pollutants.

Keywords

Metagenomics; Next Generation Sequencing; Ion torrent platform; Meta data analysis.

Introduction

Most of the microorganisms on earth are shrouded from our curious eyes because of their reluctance to grow in pure culture.Conventionally, characterization of various microbial communities in contaminated soil and water has been limited to the ability to culture microorganisms from environmental samples. Unfortunately, only a few numbers of the microorganisms can be cultured in the laboratory which involved in the biodegradation of contaminants in soil and water (Malik *et al.*, 2008). A recent study has predicted that1g of soil may contain >1 million distinct bacterial genomes (Gans *et al.,* 2005), drastically greater than the estimates of 4,000 different species (Torsvik *et al.,* 2002), but less than 1% of these may be culturable. More than 99% of the microorganisms present in many natural environments are not easily culturableand so not accessible for biotechnology or basic research (Kimura, 2006).The fast growing organisms or strains adapted to particular culture conditions grow favorablywhich do not accurately represent the actual microbial community composition (Gilbride *et al.,*2006). Hence Cultivation based methods elucidating only smallnumber of microorganisms at contaminated sites may limit the scope of microbial biodiversity and the ecological importance of unculturable organisms at contaminated sites may go undetected (Widada *et al.,* 2002; Chakraborty *et al.,* 2022).

To overcome the necessity for culturing microorganisms,modern molecular techniques provide an opportunityin further understanding of microbial diversity and its functionality.It is based on the characterization of cellular constituents such as nucleic acids, proteins, fatty acids and other taxa-specific compounds (Rossello-Mora and Amann, 2001; Godheja *et al.*, 2014)which can be extracted directly from environmental samples without the need for culturing and also preserve its*in situ* metabolic function (Wilson *et al.,* 1999).

16S rRNA hypervariable region

Here, we reviewed the species abundance and diversity of the bacterial community using 16s rRNA hypervariable region through ion torrent sequencingat each stage of the CETP. The small subunit16S ribosomal RNA molecule has been found to be the most versatile and reliable marker for the detection of bacterial populations. It includes nine hypervariable regions (V1–V9) used to distinguish sequences flanking with nine highly conserved regions (C1–C9) used as primer for PCR amplification (Baker *et al.,* 2003). The four hypervarible regions are commonly used for the analysis namely V3, V4, V5 and V6wherein V6 region with 100 bp is used in ion torrent platform. Even thoughV6 is a shorter region, it explicated the phylogeny and bacterial community was profiled clearly.

PCR-based pyrosequencing of 16S rRNA genes (commonly defined as '16S pyrotags') allows researchers to profile highly complex bacterial community compositions in depth, especially for exploring currently uncultivated and unknown bacteria. All commonly used 16S rRNA gene universal primers are designed in the highly conserved regions that are believed to exhibit different coverage for different bacterial lineages. This has been revealed by computational simulation (*in silico* prediction) using the collected 16S rRNA gene databases (Nossa *et al.,* 2010).

Novel microbes from various niches

Microorganisms are ubiquitous contributors in global biogeochemical cycles, and their ability to degrade or transform environmental pollutants that are otherwise persistent in the environment has sparked an interest in the mechanisms underlying these natural processes. Harnessing those biochemical reactions in the practice of bioremediation is an attractive option

for eliminating harmful compounds from contaminated sites (Liu and Suflita, 1993), but understanding the microbial physiology underlying biodegradation processes that take place *in situ* can also be helpful for improving the process of biostimulation or co-metabolism (Nzila, 2013). However, our current understanding of bacterial metabolism is limited by our apparent inability to grow the vast majority of environmental bacteria in a pure culture (Pham and Kim, 2012). In the last decade molecular cultivation-independent techniques have allowed the access to yet uncultivated microorganisms in several environmental niches, providing significant insights into bacterial communities in wastewater treatment processes (Silva *et al.,* 2010). Many studies related to an activity-based strategy have demonstrated that the metagenomic approach allows for finding various biocatalysts such as lipases, esterases, amylases, cellulases, xylanases, proteases and nitrilases (Jeon *et al.,*2009) and novel antibiotics, such as turbomycin A and B (Gillespie *et al.,* 2002). Metagenomic screening of extreme environments, soda-lakes in Africa and Egypt, detected more than a dozen cellulases, some of which displayed habitat related halotolerant characteristics (Rees *et al.,* 2005; Ilmberger and Streit, 2017). One of the earliest articles presenting metagenome-derived biocatalysts reported the detection of cellulases from a thermophilic, anaerobic digester fueled by lignocelluloses (Healy *et al.,* 1995). While most metagenomic surveys for novel cellulases concentrate on extreme environments, there is sufficient evidence that non-extreme, and therefore highly genetically diverse, environments also contain a range of cellulases which are highly stable and suitable for industrial applications (Voget *et al.,* 2006).

Metagenomic DNA extraction and purification

The physical and chemical structure of each microbial community affects the quality, size and amount of microbial DNA that can be extracted. Accessing planktonic communities requires

equipment that is capable of handling large volumes of water to concentrate sufficient microbial biomass to obtain enough DNA to build libraries (Griffiths *et al.,* 2000; Felczykowska *et al*., 2015). Contaminating chemicals and enzymes often remain in the water, making it relatively easy to isolate DNA without abundant contaminants. In contrast, inorganic soil components, such as negatively and positively charged clay particles and biochemical contaminants, such as humic acids and DNases, make DNA extraction from soils and subsequent manipulation, challenging. The process for removing contaminants determines both the clonability and the size of the DNA because many of the processes that effectively remove contaminants that inhibit cloning also shear the DNA (Gupta *et al*., 2018). Physical disassociation of microbes from the semisolid matrix, typically termed 'cell separation', can yield a cell pellet from which DNA (Roose-Amsaleg *et al.,* 2001), especially high molecular weight (>20 kb) DNA, can be obtained. Immobilization of cells in an agarose matrix further reduces DNA shear forces and, following electrophoresis, facilitates separation of high molecular weight DNA from humic acids and DNases. Various commercial kits can be used to extract DNA from soil and other semisolid matrices. Applying multiple extraction methods to a DNA sample can yield minimally contaminated DNA. For example, a Fast DNA Spin (Qbiogene) preparation followed by a cetyltrimethylammonium bromide (CTAB) extraction yields high quality DNA. Due to shear forces, low molecular weight DNA is typically isolated, but the physically vigorous nature of some of these methods can facilitate lysis of encapsulated bacteria, spores, and other microbial structures that are resistant to more 'gentle' lysis methods, thereby providing access to a greater, more diverse proportion of the microbial community for cloning (Handelsman *et al.,* 1998; Handelsman, 2004).

Library construction

The choice of cloning vector and strategy largely reflects the desired library structure (i.e., insert size and number of clones) and target activities sought. To obtain a function encoded by a single gene, small DNA fragments (20 kb) must be cloned into fosmids, cosmids or Bacterial Artificial Chromosomes (BACs), all of which can stably maintain large DNA fragments (Mokoena, 2019). Two vectors, namely pCC1FOS and pWE15, have been used for cloning large DNA fragments from various microbial communities (Viszwapriya and Karuthapandian, 2017). The pCC1FOS vector has the advantage that, its copy number can be controlled by addition of arabinose in the medium to increase DNA yield in the appropriate host (e.g. *E. coli* Epi300).

BAC, Yeast Artificial Chromosome (YAC), Cosmid and Fosmid are commonly used to navigate high elevate clone, linear level of non- clustered subcloning fragment library can be achieved through pUC vector series (Rajendren and Gunasekaran, 2008). Nevertheless, actinomycetes possess interesting genes coding for enzymes and antibiotic synthesis. Actinomycetes devote a large part of their genomes to the synthesis of secondary metabolites(Hug *et al.,* 2018). An average actinomycetes strain has the genetic potential to produce 10–20 secondary metabolites (Omura *et al.*, 2001). Thus, it is important to develop metagenomic libraries for the identification of functional genes from environmental samples that are rich in actinomycetes. For this, a submetagenomic library may be generated using *E. coli*–Streptomyces Artificial Chromosome (ESAC) vectors to create expression libraries in Streptomyces organism (Rajendren and Gunasekaran, 2008; Sosio *et al.,* 2000).

Community complexity and metagenomic library structure

The determination of target insert size, cloning vector, and minimum number of library clones is governed by the type of genes that are sought and the complexity of the microbial community (Ferrer *et al.,* 2009). Shotgun sequencing is usually conducted on small-insert clones, whereas successful functional studies can be performed on small- and large-insert clones. Small-insert metagenomic libraries constructed in plasmids that stably maintain up to 10 kb of DNA require 3–20 times more clones compared to libraries constructed in fosmids (30–40 kb inserts) or BACs (up to 200 kb inserts) to obtain comparable coverage of the same microbial community (Benbelgacem *et al.,* 2018). Community coverage is possible only in relatively simple microbial communities like the acid mine drainage, which contains only about five members. This community was sampled and sequenced deeply with a high-density, large-insert metagenomic library, making it possible to reconstruct the genome of one member. In contrast, the metagenome of a very complex community such as that in soil can only be sampled, not exhaustively sequenced. With today's technology, a community can be sequenced to closure more quickly and cheaply with small than large insert clones, but future technology development may change this. In activity based analysis, large inserts are preferable because the probability that the target activity is encoded by any one clone is positively correlated with the size of the insert, and if the activity is encoded by a cluster of genes, they are more likely to be captured in a large insert (Handelsman *et al.,*1998).

Screening of enzymes from metagenomic library

Metagenomics has increased the probability to discover several novel enzymes with high and specific catalytic properties sought for various industrial applications like synthetic organic chemistry, chiral resolutions, food and flavour, biofuels and synthetic biology (Kennedy *et*

*al.,*2008) obtained from microbial communities that either confounded cultivation or failed to yield new culture isolates upon repeated attempts through metagenomics (Madhuri *et al.,* 2019). Assays that have historically been used to identify enzymes (e.g., amylases, cellulases, chitinases, and lipases) in cultured isolates have been applied successfully to functional metagenomics (Henne *et al.,* 1999). Function-based metagenomic analysis of Wisconsin agricultural soil yielded 41 clones having either antibiotic, lipase, DNase, amylase, or hemolytic activities among BAC libraries containing 28,000 clones with an average insert size of 43 kb. The frequency of finding active clones in these libraries ranged from 1:456 to 1:3648, which is similar to the results from other metagenomic surveys for biocatalysts, thereby highlighting the need for robust assays for functional analysis (Rondon *et al.,* 2000).

Functional metagenomics

Plenty of novel genes encoding industrially important enzymes have been isolated from the metagenome. Typical reviews summarize the discovery of biocatalysts using metagenomics (Ferrer *et al.,* 2005 a & b). A cost effective production of second-generation biofuels has been a hot spot of research and industries since a decade. Biofuels have attracted great interest as an alternative, renewable source of energy. Energy dependence on ongoing fossil fuels and growing environmental awareness of the critical consequences of burning such fuels has forced the necessity of production of cost effective, eco-friendly and renewable resources. Plant biomass is the most abundant biopolymer containing the largest reservoirs of organic carbon on Earth is largely inaccessible to most organisms (Gupta and Sharma,2011). These biopolymers, mainly present in the form of celluloses, hemicelluloses, and lignin, have been well-recognized as a potential sustainable source for biofuel production. Examples of microbial degradation of major plant biomass i.e. lignocelluloses for biofuel production are diverse. Still our current

understanding of the enzymes involved in these processes is limited to a handful of model organisms such as the fungus *Trichoderma reesei* and the bacterium *Clostridium thermocellum* (Rubin, 2008).There are several biofuels which have already been produced commercially e.g. bioethanol, biobutanol, biodiesel and biogases; however cost-effective production methods are still elusive. In order to identify novel microbes and enzymes to accomplish the goal of cost-effective production, metagenomic libraries from several resources are being targeted (Singh *et al.,* 2009).

Advancement in screening of metagenomes

Number of bottlenecks was faced in metagenomic screening that need to be resolved in order to improve the discovery rate of target enzymes. These limitations can be grouped in three main categories. First, there is a need for improvement in host organism capabilities with the aim of improving the expression of the target enzymes. Second, the development of new genetic tools is necessary in order to improve the construction of metagenomic libraries suitable for screening in different hosts. Finally, continuation of ongoing research to elaborate novel screening strategies that enhance the discovery rate of the enzymes of interest is needed (Guazzaroni*et al.,* 2015). The design aspect focuses on the planning and construction of new gene circuits for the desired application (Canton *et al.,* 2008). Modelling involves computational simulation of the proposed gene circuits in order to both evaluate performance and capabilities and to guide the selection of the suitable molecular components necessary for its construction (Koide *et al.,* 2009). The implementation step encompasses the physical assembly of the DNA elements encoding the appropriate components (such as promoters, regulators, terminators, enzymes, transporters, etc.) and follows a specific assembly standard (Alves *et al.,* 2017; Arkin, 2008).

M13 based insert amplification from metagenome library

The limitation of small fragment cloning or subcloning vector and *E. coli* as a host for comprehensive mining of metagenomic samples is highlighted by the low number of positive clones obtained during a single round of screening (typically less than 0.01%). A recent study indicates that it is virtually impossible to recover translational fusion products owing to the high number of clones (O107) that would need to be screened through SIGEX and M13 (Gabor *et al.,* 2004). Intuitively, expression from native promoters and read-through transcription from the vector-based M13 promoter offer the best chance for recovery of heterologously expressing genes. Statistically, for a small insert (>10 Kb) library, between 10^5 and 10^6 clones need to be screened for a single hit (Henne *et al.,* 1999). This suggests that without sample enrichment the discovery of specific genes in a complex metagenome is technically challenging. The assumption that expression in an *E. coli* host will not impose a further bias is largely untested. Although the vector and *E. coli* transcriptional machinery is known to be relatively promiscuous in recognizing foreign expression signals, a bias in favour of Firmicutes genes has been established (Srivastava*et al.,* 2013; Gabor *et al.,* 2004). The further development of host screening systems is therefore a fruitful approach for the more effective future exploitation of metagenomes.

A laboratory-scale sequencing batch reactor was successfully operated for different levels of Enhanced Biological Phospahate Removal (EBPR), removing around 25, 40 and 55 mg/l. The microbial communities were dominated by the uncultured polyphosphate-accumulating organism ‘‘*CandidatusAccumulibacterphosphatis*’’ (CAP). When EBPR failed, the sludge was dominated by tetrad-forming a-Proteobacteria. Representative and reproducible 2D gel protein separations were obtained for all sludge samples. 638 protein spots were matched across gels generated from the phosphate removing sludges. A 111 of these were excised and 46 proteins were identified

using recently available sludge metagenomic sequences. Many of these closely match proteins from ‘‘*CandidatusAccumulibacterphosphatis*’’ and could be directly linked to the EBPR process. They included enzymes involved in energy generation, olyhydroxyalkanoate synthesis, glycolysis, gluconeogenesis, glycogen synthesis, glyoxylate/TCA cycle, fatty acid b oxidation, fatty acid synthesis and phosphate transport. Several proteins involved in cellular stress response were detected (Wilmes*et al.,* 2008).

Novel molecules recovery

The list of published biocatalysts and bioactive compounds discovered by screening of complex environment DNA libraries is still rather small, but grows rapidly. The list includes biocatalysts of industrial interest such as amylases, lipases, proteases, dehydratases, oxidoreductases and agarases (Daniel, 2004). Richardson *et al.,* (2002) described a high-performance amylase for starch liquefaction that was generated by combining environmental discovery and laboratory evolution for the identification and optimization of the enzyme.

To date at total of about 80 esterase or lipase-positive clones derived from metagenomes have been reported (Pindi *et al.,* 2009). The level of characterisation of these novel lipolytic genes ranges from DNA restriction and sequencing analysis to determine clone diversity (Rondon *et al.,* 2000; Voget *et al.,* 2006) to detailed biochemical analysis of the purified enzyme. The esterases from the deep sea hypersaline basin display habitat related properties in that they were most active at alkaline pH and displayed higher activities under high pressure conditions. They displayed unique substrate spectra with EstA3 being able to hydrolyse substrates such as 7-(3-octylcarboxy-(3-hydroxy- 3-methyl-butyloxy))-coumarin, a normally unreactive secondary ester (Elend *et al.,* 2006; Schmeisser *et al.,* 2007).

Amylases have been in the focus of a number of many metagenome studies. Many articles reporting the detection and characterisation of novel amylolytic enzymes from metagenomic DNA libraries have been published (Richardson *et al.*, 2002; Voget *et al.,* 2003; Yun *et al.*, 2004; Vidya *et al.,* 2011; Ferrer *et al.*, 2005b; Wang *et al.*, 2019; Ahmad *et al.*, 2019; Yadav*et al.,* 2019). Aninteresting characteristics of amylases from a soil metagenome is exhibitedas stable and active under alkaline conditions, with a pH optimum of 9.0, a characteristic required of amylases in detergents (Yun *et al.,* 2004; Yadav *et al.,* 2019). Functional screening of a soil metagenomic library for cellulases revealed a total of eight cellulolytic clones, one of which was purified and characterised (Voget *et al.*, 2006). Despite the fact that this library had been generated from a soil sample collected from a non-extreme environment, the cellulase displayed a high level of stability over a broad pH range, up to pH 9.0, it was stable at 40°C for up to 11 h and it was highly halotolerant being active and stable in 3M NaCl.

Chitinases have several biotechnological applications including increasing plant resistance to fungal disease (Howard *et al.,* 2003; Stoykov *et al.,* 2015; Singh *et al.*, 2021). There are limitedmetagenomic studies that focus on chitinases from marine environments to date (Cottrell *et al.,* 1999) and one study using an enrichment approach in combination with metagenomic technologies (Hoster *et al.,* 2005, Jacquiod *et al.,* 2013). The metagenomic survey of the insect gut discovered four xylanases, which were phylogenetically distant from all other known xylanases suggesting that they had evolved independently (Brennan *et al.,* 2004; Wu, 2018). These xylanases are novel and produce unique hydrolysis products. Screening of the lagoon resulted in the detection and characterisation of one xylanase which displayed habitat related properties in that it was most active at lower temperatures (Lee *et al.,* 2006; Pindi *et al.,*

2009). Thus, combining discovery of novel robust activities from the screening of metagenomic libraries with modification by directed evolution or gene shuffling offers a great potential for encountering and evolving commercially valuable biomolecules (Daniel, 2004).

Ion torrent PGM sequencing is light independent where sequence composition is determined by measuring pH changes due to hydrogen ion liberation as nucleotides are incorporated during strand synthesis in picolitre wells (Rothberg *et al.,* 2011; Whiteley*et al.,* 2012). Integrated circuits are used to measure pH changes to identify the base incorporation which removes need for expensive light detection systems, substantially reduces costs and theoretically, it is infinitely scalable; since the number of sequences obtained simply equates to the physical dimensions of the integrated sensor (Whiteley *et al.,* 2012). Economical and rapid in sequencing are the two important positive aspects of ion torrent technology (Glenn, 2011, Krishnan *et al*., 2016), which has been used in diverse ecological systems, such as the Athabasca river (Yergeau *et al.,* 2012), human sub-gingival plaque (Zhou *et al.,* 2013), sewage sludge (Bibby *et al.,* 2011), arctic soil (Bell *et al.,* 2013) and oil-contaminated soil (Roy *et al.,* 2013).

Apart from microbial community studies, ion torrent sequencing has been used to draft the genomic sequences of bacteria, including *Pseudomonas aeruginosa* from crude oil-contaminated soil (Roy *et al.,* 2013), *Sulfobacillusthermosulfidooxidans* from mines (Travisany *et al.,* 2012), halotolerant *Pontibacter* sp. (Joshi *et al.*, 2012) and *Alkalibacillus haloalkaliphilis* from saline sediment. Similarly, CETP comprisediverse microbial populationincludes bacteria, fungi, Archae, algae, protozoa and viruses. In aeration systems, bacteria are the dominant phyla that are responsible for the digestion and removal of pollutants during the treatment (Sanapareddy*et al.,* 2009) where bacterial habitant in activated sludge are nitrifiers (Wang *et al.,* 2014), denitrifiers (Gentile *et al.,* 2007), phosphorus-accumulating organisms and methanogens

(Fernandez *et al.*, 2000). The presence of β-proteobacteria, Bacteroidetes and γ-proteobacteria groups is profusion in aerobic and anaerobic chamber of wastewater bioreactor identified by 16S rRNA gene clone library (Niu *et al.*, 2006; Kapley *et al.*, 2007). The bacterial diversity has been evaluated in aeration lagoon I and II of tannery CETP through PCR-RFLP (Chandra *et al.*, 2010).

The characterisation of bacterial diversity of tannery CETP at the different treatment stages is done by ion torrent sequencing (Krishnan *et al*., 2016). Usually, microbiota is the first community of organisms to be influenced by the release of pollutants into the environment which affects the microorganisms functionally that lead to change over in their physiology and diversity (Silver, 1996). Thereby Bacterial diversity indices reflect the degree of environmental contaminationpresumably by occurrence of resistant strains (Aiking *et al.*, 1984). Microbial communities in the water system can be used as a potential indicator of pollution because they intakeof pollutants, even at minor levels (Yergeau *et al.*, 2012).

For the development of bioremediation strategies, it is important to understand indigenous microbial taxa of contaminated site (Pollmann et al., 2006). In the four treatment stages of tannery Common Effluent Treatment Plant (CETP), the bacterial composition was assessed through culture reliable approach. Firmicutes was found to be the dominant group among all treatment stages of about 90%. Notably, putative Bacillus sp., B. firmus and B. licheniformis were copious in all the treatment stages (Suganya *et al*., 2013; Krishnan *et al*., 2016).

Among the diverse ecosystems, CETP is characterised by a dynamic microbial consortium. Its function can be understood through studying the co-existing populations

(Manefield *et al*., 2005). In addition, the microbial population is altered with respect to the reactor's operational conditions and physico-chemical parameters (Kapley *et al*., 2007). The next generation sequencing (NGS) (Sogin *et al*., 2006) is implemented for the better exploration and understanding of microbial diversity and its function in complex environment such as municipal effluent (Ma *et al*., 2013), marine water (Brown *et al*., 2009), soil (Jones *et al*., 2009), river sediments (Suriya *et al*., 2017) and wastewater (Sanapareddy *et al*., 2009).

Emergence of Next Generation Sequencing

Over the past decades, both traditional culture dependent and independent approaches including isolation, cloning, functional identification, DGGE (Denaturing Gradient Gel Electrophoresis) (Muyzer *et al.,* 1993), T-RFLP (Terminal Restriction Fragment Length Polymorphism) (Liu *et al.*, 1997), FISH (Fluorescence In Situ Hybridization) (Wagner *et al.*, 1998) and Gene chips (He *et al.*,2012) etc., always cause bottlenecks in exploring global bacterial diversities and their potential functions (Cai *et al.*,2013). To overcome these limitations, high-throughput sequencing technologies have been well developed to promote the relevant research fields. The Roche 454, SOLID, Ion Torrent and Illumina platforms have dominated the next-generation sequencing market and made notable contributions to the genomic and metagenomic studies in the past five years (Glenn, 2011; Dubey *et al.,* 2022).

Since previous decade, next-generation sequencing (NGS) technology has greatly advanced our knowledge of microbial diversity and function in complex ecosystems (Dubey *et al.,* 2022). Both PCR based analysis of 16S rRNA and shotgun metagenomic studies have been used recently to characterise soils (Bowers *et al.,* 2011), oceans (Caporaso *et al.*,2011), the atmosphere (Bowers *et al.*,2011) as well as the human microbiome (Kuczynski *et al.*,2011). The other technological advances include the high throughput pyrosequencing (Margulies *et*

al.,2005), developments ofcomplexbarcoding analyses (Hamady *et al.*,2008), ultra-high throughput systems (Caporaso *et al.*,2012) and extensive increases in the capability to analyse and handle data outputs (Meyer *et al.*,2008).

In the past, most commonly used approach was cloning and sequencing of the 16S ribosomal RNA gene (rDNA) using conserved broad-range PCR primers (Nossa *et al.,* 2010; Klindworth *et al.,* 2013). With the advent of massive parallel sequencing technologies, direct sequencing of PCR amplicons became feasible (Armougom and Raoult, 2009). In 2006, Roche's 454 GS 20 pyrosequencing (Margulies *et al.*,2005) became the first high-through-put sequencing technology to be successfully applied for large scale biodiversity analysis and was crucial to reveal the 'rare biosphere' (Sogin *et al.,* 2006; Nathani *et al.,* 2020). The continuous advancement of the technology, offering read lengths of up to 1000 bp; further improved throughput and resolution of 16S rDNA sequencing also developed (Liu *et al.*,2007; Klindworth *et al.*,2013). Since then, additional high-throughput sequencing technologies have become commercially available. The attractiveness of Illumina (Bennett, 2004) lies in the reduced per base costs and comparatively high sequencing depth (Caporaso *et al.,* 2012), despite having short read lengths (Klindworth *et al.,* 2013). While the major advantages of Ion Torrent sequencing (Rothberg *et al.,* 2011) are low cost and rapid sequencing speed.

4.1 Ion torrent platform

Recently, the Ion Torrent Personal Genome Machine (Life Technologies) introduced a distinctiveapproach of 'ion' sequencing, which is light independent where sequence composition is determined by measuring pH changes due to hydrogen ion liberation as nucleotides are incorporated during strand synthesis in picolitre wells (Rothberg *et al.*,2011; Whiteley *et al.,* 2012). The integrated circuits is used to measure pH changes, which substantially reduces costs

and, since the number of sequences obtained simply equates to the physical dimensions of the integrated sensor (Glenn, 2011).

In 2010, Life Technologies introduced the Ion PGM as a post light sequencing technology. This system relies on the real-time detection of hydrogen ion concentration, released as a by-product when a nucleotide is incorporated into a strand of DNA by the polymerase action (Shokralla*et al.,* 2012). Ion Torrent uses a high-density array of micro-machined wells to perform this bio-chemical process in a massively parallel way. Each well holds a single DNA template from the library. Beneath the wells are an ion-sensitive layer and a proprietary ion sensor to detect the change in hydrogen ion concentration because of nucleotides incorporation (Rothberg *et al.,* 2011; Shokralla*et al.,* 2012). The Ion Torrent plat-form can utilize one of the three available ion chips: 314, 316 or 318, which can generate up to 10 Mb, 100 Mb or 1 Gb respectively, according to the required sequencing coverage. In 2012, Life Technologies introduced a new generation of Ion semi-conductor sequencers; the Ion Proton bench top sequencer. Ion Proton chips will deliver the human genome or human exome in just a few hours. Ion Proton chips will be available in two versions: Ion Proton I chip with 165 million wells (about 100-fold more than the Ion 314 chip) and Ion Proton II chip with 660 million wells (about 1000-fold more than the Ion 314 chip) (Shokralla *et al.,* 2012). Both of these chips use CMOS semiconductor chip technology to capture chemistry changes instead of light and translate these changes into digital data.

The implementation of the Hi-Q enzyme has significantly improved the accuracy of PGM reads by reducing indel errors and enhancing genome assembly quality (Veras*et al.,* 2014). Regarding the error rates, the Ion Torrent platform report a pattern of premature sequence truncation, specific to semiconductor sequencing. Read truncation was dependent on both the

directionality of sequencing and the target species, resulting in organism-specific biases in community profiles. These sequencing artifacts could be minimized by using bidirectional amplicon sequencing and an optimized flow order on the Ion Torrent platform, which have enhanced the accuracy of microbial community profiling(Salipante *et al.,* 2014).

Tools for meta data analysis

A number of different systems and resources for metagenome or similar analysis, which are offered in the form of databases, web portals, web services and basic stand alone programs (Overbeek *et al.*,2005; Mitra *et al.*,2011; Dudhagara *et al.*,2015). The currently available comprehensive platforms, integrative and annotation tools are reviewed by Tas*et al.*,2021 which includes Metagenomics Rapid Annotation usingSubsystem Technology (MG-RAST) (Meyer *et al.*,2008),Integrated Microbial Genomesand Metagenomes (IMG/M) (Markowitz *et al.*,2014),EBI Metagenomics (Hunter *et al.*,2014), and pipelinessuch as MEGAN (Huson and Weber, 2013),MetaMetaDB(Yang and Iwasaki, 2014),Anvi'o platform, etc. Formerly, many of those resources are suitable only for taxonomic analysis and mainly focused on the analysis of individual metagenomes,at present, they areable to performrapid and highly interactive comparison of multiple datasets.

MG-RAST is a user-friendly opensource metagenomics analysis server(Meyer *et al.,* 2008, Keegan *et al.,* 2016) current version 4.0.3 provides 512,950 metagenomes containing 2,287 billion sequences (as on 03 June 2024),a readily useable service for analyzing a new metagenomic dataset by using the pipelines of SEED subsystem, COG, etc.Besides producing initial assignments of gene function and a metabolic reconstruction, the RAST server provides an environment for browsing the annotated genome and comparing it to the hundreds of genomes maintained within the integration. The genome viewer included in MG-RAST supports detailed

comparison against existing genomes, determination of genes that the genome has in common with specific sets of genomes (genes that distinguish the genome from those in a set of existing genomes), the ability to display genomic context around specific genes, and the ability to download relevant information and annotations as desired (Aziz *et al.*,2008).

The IMG/Mis an online metagenome data storage, management,and analysis system at the Department of Energy (DOE) Joint Genome Institute (JGI) continues to provide support for users to perform comparative analysis of isolate and single cell genomes, metagenomes, and metatranscriptomes and integrates metagenome datasets with isolatedmicrobial genomes from the IMG system(Markowitz et al., 2007). The multiple reference datasets are used for the analysis of IMG/M data in threelevels such as phylogenetic composition, functional or metabolicpotential within individual microbiomes and comparisonsacross microbiomes.In addition to datasets produced by the JGI, IMG v.7 also includes datasets imported from public sources such as NCBI Genbank, SRA, and the DOE National Microbiome Data Collaborative (NMDC), or submitted by external users (Chen *et al.,* 2023). It facilitates environmental and organismal metagenomicsanalysis (Dudhagara *et al.,* 2015).

EBI Metagenomics was developed by EMBL-EBI in 2011 which contains QIIME (Caporaso *et al.,* 2010) embeddedtools for data management, analysis, storage, andsharing of metagenomes (Hunter *et al.,*2014). It is continuallymodernizingthe analysis pipeline with the development of additional analysisand visualization tools. The Community Cyberinfrastructure for Advanced MarineMicrobial Ecology Research and Analysis (CAMERA) wasdeveloped in 2007 with the aim of monitoring microbial communitiesof the ocean and their response to environmentalchanges (Seshadri *et al.,* 2007; Sun *et al.,* 2010).

MEGAN is easy to install and use, and requires only a BLAST output file as input to operate. The program is designed to allow both high-level analysis that summarizes data at different ranks of the NCBI taxonomy, and detailed analysis that drills down to individual reads and their BLAST matches (Mitra *et al.,* 2011).Using MEGAN program,the unusual abundance of ecologically importantbacteria and archaea are revealed by the analysis of metagenome from tropical mangrove swamp in Malaysia (Ismail *et al.,* 2017).

MetaMetaDB (Yang and Iwasaki, 2014) was developed for a meta-analysis ofdifferent environments inhabited by a microbe and the investigation of specific functional genes or metabolic flexibility that facilitate adaptation. This database may be used topredict all possible habitats of microbes by searching for thepresence of microbes in metagenomic and 16S rRNA ampliconsequencing datasets derived from diverse environments such as heavy metal resistance genes and salt-stress responsive genes (Hiraoka *et al.,* 2016).

Anvi'o is an advanced integrated analysis and visualization platform that offers automated and human-guided characterization of microbial genomes for genomics, metagenomics, metatranscriptomics, pangenomics, metapangenomics, phylogenomics (Eren *et al.,* 2015). By using Anvi'o analysis, one thousand non-redundant microbial population genomes were identified from the TARA Oceans metagenomes (Delmont *et al.,* 2018).Combined use of gene centric metagenomics, metatranscriptomics and metaproteomics allows elucidation of microbial interaction and regulatory functions. The analysis of Metagenome Assembled Genomes (MAGs) (Yang *et al.,* 2021) has also shown novel insights intomicrobial metabolic diversity and niche differentiation (Tas *et al.,* 2021).

Reference

Ahmad, T., Singh, R. S., Gupta, G., Sharma, A., & Kaur, B. (2019). Metagenomics in the search for industrial enzymes. In Advances in Enzyme Technology (pp. 419-451). Elsevier.

Aiking H, Stinamn A, Van-Ganderen C, Van-Heerikhuizen H &Vant-Riet J (1984). In organic phosphate accumulation and cadmium detoxification in *Klebsiella aerogenesa* NCTC 418 growing in continuous culture. Appl. Environ. Microbiol. 47: 374-377.

Alves, L. D. F., Silva-Rocha, R., &Guazzaroni, M. E. (2017). Enhancing metagenomic approaches through synthetic biology. Functional metagenomics: Tools and applications, 75-94.

Arkin AP (2008).Setting the standard in synthetic biology. Nature Biotechnol. 26(8): 771-774.

Armougom F &Raoult D (2009). Exploring microbial diversity using 16S rRNA high-throughput methods. J. Comput. Sci. Syst. Biol. 2: 74–92.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., ...&Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. BMC genomics, 9, 1-15.

Baker GC, Smith JJ & Cowan DA (2003). Review and re-analysis of domain-specific 16S primers. J. Microbiol. Methods. 55: 541-555.

Bell TH, Yergeau, E, Maynard C, Juck D, Whyte LG & Greer CW (2013). Predictable bacterial composition and hydrocarbon degradation in Arctic soils following diesel and nutrient disturbance. The ISME J.7 (6): 1200-1210.

Benbelgacem, F. F., Salleh, H. M., &Noorbatcha, I. A. (2018). Construction of Metagenomic DNA Libraries and Enrichment Strategies. Multifaceted Protocol in Biotechnology, 23-42.

Bibby K, Viau E &Peccia J (2011). Viral metagenome analysis to guide human pathogen monitoring in environmental samples. Lett. Appl. Microbiol. 52: 386-392.

Bowers RM, Sullivan AP, Costello EK, Collett JL, Knight R &Fierer N (2011). Sources of bacteria in outdoor air across cities in the midwestern United States. Appl. Environ. Microbiol. 77: 6350-6356.

Brennan Y (2004). Unusual microbial xylanases from insect guts. Appl. Environ. Microbiol. 70: 3609-3617.

Brown MV, Philip GK, Bunge JA, Smith MC &Bissett A (2009). Microbial community structure in the North Pacific Ocean. ISME J. 3: 1374-1386.

Cai, L., Ye, L., Tong, A. H. Y., Lok, S., & Zhang, T. (2013). Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. PloS one, 8(1), e53649.

Canton B, Labno A&Endy DN (2008). Refinement and standardization of synthetic biological parts and devices. Biotechnol. 26(7): 787-93.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G & Knight R (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6(8): 1621-1624.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Gonzalez A, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J & Knight R (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods. 7: 335-336

Chakraborty, J., Palit, K., & Das, S. (2022). Metagenomic approaches to study the culture-independent bacterial diversity of a polluted environment—a case study on north-eastern coast of Bay of Bengal, India. In Microbial Biodegradation and Bioremediation (pp. 81-107). Elsevier.

Chandra R, Bharagava RN, Kapley A&Purohit HJ (2010). Bacterial diversity, organic pollutants and their metabolites in two aeration lagoons of common effluent treatment plant (CETP) during the degradation and detoxification of tannery wastewater. Biores. Technol 102(3): 2333-2341.

Chen, I. M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., ...&Ivanova, N. N. (2023). The IMG/M data management and analysis system v. 7: content updates and new features. Nucleic acids research, 51(D1), D723-D732.

Cottrell MT, Moore JA &Kirchman DL (1999). Chitinases from uncultured marine microorganisms. Appl. Environ. Microbiol. 65:2553–2557

Daniel, R. (2004). The soil metagenome–a rich resource for the discovery of novel natural products. Current opinion in biotechnology, 15(3), 199-204.

Daniel, R. (2004). The soil metagenome–a rich resource for the discovery of novel natural products. Current opinion in biotechnology, 15(3), 199-204.

Delmont, T. O., Quince, C., Shaiber, A., Esen, Ö. C., Lee, S. T., Rappé, M. S., ...&Eren, A. M. (2018). Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nature microbiology, 3(7), 804-813.

Dubey, A., Malla, M. A., & Kumar, A. (2022). Role of next-generation sequencing (NGS) in understanding the microbial diversity. In Molecular genetics and genomics tools in biodiversity conservation (pp. 307-328). Singapore: Springer Nature Singapore.

Dudhagara, P., Bhavsar, S., Bhagat, C., Ghelani, A., Bhatt, S., & Patel, R. (2015). Web resources for metagenomics studies. Genomics, Proteomics and Bioinformatics, 13(5), 296-303.

Elend C (2006). Isolation and biochemical characterization of two novel metagenome-derived esterases. Appl. Environ. Microbiol. 72:3637–3645.

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ, 3, e1319.

Felczykowska, A., Krajewska, A., Zielińska, S., &Łoś, J. (2015). Sampling, metadata and DNA extraction-important steps in metagenomic studies. ActaBiochimicaPolonica, 62(1), 151-160.

Fernandez AS, Hashsham, SA, Dollhopf SL, Raskin L, Glagoleva O, Dazzo FB, Hickey RF., Criddle, CS &Tiedje JM (2000). Flexible community structure correlates with stable community function in methanogenic bioreactor communities perturbed by glucose. Appl. Environ. Microbiol. 66: 4058–4067

Ferrer M (2005a). Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. Chem. Biol. 12:895–904.

Ferrer M (2005b). Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. Environ. Microbiol. 7: 1996–2010.

Ferrer, M., Beloqui, A., Timmis, K. N., &Golyshin, P. N. (2009). Metagenomics for mining new genetic resources of microbial communities. Journal of molecular microbiology and biotechnology, 16(1-2), 109-123.

Gabor EM & Janssen DB (2004). Increasing the synthetic performance of penicillin acylase PAS2 by structure-inspired semi-randommutagenesis. Protein. Eng. Des. Sel. 17: 571–579.

Gans, J., Wolinsky, M., & Dunbar, J. (2005). Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science, 309(5739), 1387-1390.

Gentile ME, Jessup CM, Nyman JL &Criddle CS (2007). Correlation of functional instability and community dynamics in denitrifying dispersed-growth reactors. Appl. Environ. Microbiol. 73(3): 680-690.

Gilbride KA, Lee DY &Beaudette LA (2006). Molecular techniques in wastewater: understanding microbial communities, detecting pathogens, and real-time process control. J. Microbiol. Meth. 66: 1-20

Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon M R &Handelsman J (2002). Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. Appl. Environ. Microbio. 68(9): 4301-4306.

Glenn TC (2011). Field guide to next-generation DNA sequencers. Mol. Ecol. Resour. 11: 759-769.

Godheja, J., Shekhar, S. K., & Modi, D. R. (2014). Advances in molecular biology approaches to guage microbial communities and bioremediation at contaminated sites. International Journal of Environmental Bioremediation & Biodegradation, 2(4), 167-177.

Griffiths BS, Ritz K, Bardgett RD, Cook R, Christensen S &Ekelund F (2000). Ecosystem response of pasture soil communities to fumigation-induced microbial diversity reductions: an examination of the biodiversity–ecosystem function relationship. Oikos. 90: 279–294.

Guazzaroni, M. E., Silva-Rocha, R., & Ward, R. J. (2015). Synthetic biology approaches to improve biocatalyst identification in metagenomic library screening. Microbial Biotechnology, 8(1), 52-64.

Gupta, N., Vats, S., & Bhargava, P. (2018). Sustainable agriculture: role of metagenomics and metabolomics in exploring the soil microbiota. In silico approach for sustainable agriculture, 183-199.

Gupta, R. D., & Sharma, R. (2011). Metagenomics for environmental and industrial microbiology. Sci Cult, 77(1-2), 27-31.

Hamady M, Walker JJ, Harris JK, Gold NJ & Knight R (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat. Methods. 5: 235-237.

Handelsman J (2004). Metagenomics: Application of genomics to uncultured microorganisms. Microbiol. Mol. Biol. Rev. 68(4): 669-685.

Handelsman J, Rondon MR, Brady SF, Clardy J & Goodman RM (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem. Biol. 5: 245-249.

He ZL, Van Nostrand JD & Zhou JZ (2012). Applications of functional gene microarrays for profiling microbial communities. Curr. Opin. Biotech. 23: 460-466.

Healy PM & Palepu KG (1995). The challenges of investor communications: the case of CUC International. Inc. J. Financ. Econ. 38: 111-141.

HenneA, Daniel R, Schmitz RA & Gottschalk G (1999). Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. Appl. Environ. Microbiol. 65: 3901- 3907.

Hiraoka, S., Yang, C. C., & Iwasaki, W. (2016). Metagenomics and bioinformatics in microbial ecology: current status and beyond. Microbes and environments, 31(3), 204-212.

Hoster F, Schmitz JE & Daniel R (2005). Enrichment of chitinolytic microorganisms: isolation and characterization of a chitinase exhibiting antifungal activity against phytopathogenic fungi from a novel *Streptomyces strain*. Appl. Microbiol. Biotechnol. 66:434– 442.

Howard RL, Abotsi E, van RensburgELj, & Howard S (2003). Lignocellulose biotechnology: issues of bio conversion and enzyme production.Arf. J. Biotechnol. 2(12): 602-619.

Hug, J. J., Bader, C. D., Remškar, M., Cirnski, K., & Müller, R. (2018). Concepts and methods to access novel antibiotics from actinomycetes. Antibiotics, 7(2), 44.

Hunter, S., Corbett, M., Denise, H., Fraser, M., Gonzalez-Beltran, A., Hunter, C., ...&Sansone, S. A. (2014). EBI metagenomics—a new resource for the analysis and archiving of metagenomic data. Nucleic acids research, 42(D1), D600-D606.

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

Ilmberger, N., &Streit, W. R. (2017). Screening for cellulase encoding clones in metagenomic libraries. Metagenomics: Methods and Protocols, 205-217.
Ion Torrent Sequencing from Different Treatment Stages of CETP for Tannery. *Expert Opin Environ Biol 5:3. doi:10.4172/2325-9655.1000136*

Ismail, Z., Sam, C. K., Yin, W. F., & Chan, K. G. (2017). Tropical mangrove swamp metagenome reveals unusual abundance of ecologically important microbes. Current Science, 1698-1703.

Jacquiod, S., Franqueville, L., Cecillon, S., M. Vogel, T., &Simonet, P. (2013). Soil bacterial community shifts after chitin enrichment: an integrative metagenomic approach. PLoS One, 8(11), e79699.

Jeon JH, Kim JT, Kang SG, Lee JH & Kim SJ (2009). Characterization and its potential application of two esterases derived from the arctic sediment metagenome. Mar. Biotechnol. (NY). 11(3): 307-316.

Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R &Fierer N (2009). A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. ISME J. 3: 442-453.0

Joshi MN, Sharma AC, Pandya RV, Patel RP, Saiyed ZM, Saxena AK &Bagatharia SB (2012). Draft genome sequence of *Pontibacter sp*. nov. BAB1700, a halotolerant, industrially important bacterium. J. Bacteriol.194(22): 6329-6330.

Kapley A, De Baere T &Purohit HJ (2007). Eubacterial diversity of activated biomass from a common wastewater treatment plant. Res. Microbiol. 158: 494-500.

Keegan, K. P., Glass, E. M., & Meyer, F. (2016). MG-RAST, a metagenomics service for analysis of microbial community structure and function. Microbial environmental genomics (MEG), 207-233.

Kennedy M, Hu J, Gao P, Li L, Ali-Reynolds A, Chal B, Gupta V, Ma C, Mahajan N, Akrami A &Surapaneni S (2008). Enhanced bioavailability of a poorly soluble VR1 antagonist using an amorphous solid dispersion approach: a case study. Mol Pharm. 5(6): 981-993.

Kimura, N. (2006). Metagenomics: access to unculturable microbes in the environment. Microbes and Environments, 21(4), 201-215.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., &Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic acids research, 41(1), e1-e1.

Koide A, Wojcik J, Gilbreth RN, Reichel A, Piehler J & Koide S (2009). Accelerating phage-display library selection by reversible and site-specific biotinylation. Protein Eng. Des. Sel. 22: 685-690.

Krishnan, M., Suganya, T., &Pandiarajan, J. (2016). Bacterial Community Exploration through

Kuczynski J, Lauber CL, Walters WA, Parfrey LW, Clemente JC, Gevers D & Knight R (2011). Experimental and analytical tools for studying the human microbiome. Nat. Rev. Genet. 13: 47-58.

Lee CC, Kibblewhite-Accinelli RE, Wagschal K, Robertson GH &Wong DW (2006a). Cloning and characterization of a cold active xylanase enzyme from an environmental DNA library. Extremophiles. 10: 295-300.

Liu S &Suflita JM (1993). Ecology and evolution of microbial populations for bioremediation. Trends Biotechnol. 11(8): 344-352.

Liu Z, Lozupone C, Hamady M, Bushman FD & Knight R (2007). Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids. Res. 35: 120-125.

Ma J, Wang Z, Zhu C, Liu S & Wang Q (2013). Analysis of Nitrification Efficiency and Microbial Community in a Membrane Bioreactor Fed with Low COD/N-Ratio Wastewater. PLoS ONE. 8(5): 630-659.

Madhuri, R. J., Saraswathi, M., Gowthami, K., Bhargavi, M., Divya, Y., &Deepika, V. (2019). Recent approaches in the production of novel enzymes from environmental samples by enrichment culture and metagenomic approach. Recent developments in applied microbiology and biochemistry, 251-262.

Malik, S., Beer, M., Megharaj, M., & Naidu, R. (2008). The use of molecular techniques to characterize the microbial communities in contaminated soil and water. Environment international, 34(2), 265-276.

Manefield M, Griffiths RI, Leigh MB, Fisher R & Andrew SW (2005). Functional and compositional comparison of two activated sludge communities remediating coking effluent. Environ. Microbiol. 7: 715-722.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF & Rothberg JM (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature. 437: 376-380.

Markowitz, V. M., Chen, I. M. A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M., ...&Kyrpides, N. C. (2014). IMG/M 4 version of the integrated metagenome comparative analysis system. Nucleic acids research, 42(D1), D568-D573.

Markowitz, V. M., Ivanova, N. N., Szeto, E., Palaniappan, K., Chu, K., Dalevi, D., ...&Kyrpides, N. C. (2007). IMG/M: a data management and analysis system for metagenomes. Nucleic acids research, 36(suppl_1), D534-D538.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J & Edwards RA (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinforma. 9: 386.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., ...& Edwards, R. (2008). The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC bioinformatics, 9, 1-8.

Mitra, S., Rupek, P., Richter, D. C., Urich, T., Gilbert, J. A., Meyer, F., ...&Huson, D. H. (2011). Functional analysis of metagenomes and metatranscriptomes using SEED and KEGG. BMC bioinformatics, 12, 1-8.

Mokoena, M. I. (2019). Mining of genes encoding for DNA-manipulating enzymes from hot springs using metagenomic techniques (Doctoral dissertation, Vaal University of Technology).

Muyzer G, de Waal EC &Uitterlinden AG (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. 59: 695-700.

Nathani, N. M., Mitra, A., Mootapally, C., & Kothari, R. K. (2020). Advancements in Bioinformatics and Role in Marine Genomics and Metagenomics. Encyclopedia of Marine Biotechnology, 3, 1791-1807.

Niu SQ, Fukushima J, Jiang Y, Ishikawa Y, Ueda T & Matsumoto S (2006). Analysis of bacterial community structure in the natural circulation system wastewater bioreactor by using a 16S rRNA gene clone library. Microbiol. Immunol. 50(12): 937-950.

Nossa CW, Oberdorf WE, Yang L, Aas JA, Paster BJ, Desantis TZ, Brodie EL, Malamud D, Poles MA & Pei Z (2010). Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. WJG. 16: 4135-4144.

Nossa, C. W., Oberdorf, W. E., Yang, L., Aas, J. A., Paster, B. J., DeSantis, T. Z., ... & Pei, Z. (2010). Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. World journal of gastroenterology: WJG, 16(33), 4135.

Nzila A(2013). Update on the cometabolism of organic pollutants by bacteria. Environ. Pollut. 178: 474-482.

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., ...&Vonstein, V. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic acids research, 33(17), 5691-5702.

Pham VHT & Kim J (2012). Cultivation of unculturable soil bacteria. Trends. Biotechnol. 30: 475-484.

Pindi, P. K., Kishore, N., & Reddy, S. (2009). Metagenomics: Principles, Methodology and Biotechnological Potentials. A Textbook of Molecular Biotechnology, 343.

Pollmann K, Raff J, Merroun M, Fahmy K &Selenska-Pobell S (2006). Metal binding by bacteria from uranium mining waste piles and its technological applications. Biotechnol. Advan. 24(1): 58-68.

Rajendhran J & Gunasekaran P (2008). Strategies for accessing soil metagenome for desired applications. Biotechnol. Advances. 26: 576-590.

Rees HC, Grant S, Jones B, Grant WD &Heaphy S (2003). Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. Extremophiles. 7: 415-421.

Richardson TH (2002). A novel, high performance enzyme for starch liquefaction. Discovery and optimization of a low pH, thermostable alpha-amylase. J. Biol. Chem. 277: 26501-2650.

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J&Goodman RM (2000). Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. 66(6): 2541-2547.

Roose-Amsaleg CL, Garnier-Sillam E & Harry, M., (2001). Extraction and purification of microbial DNA from soil and sediment samples. Appl. Soil Ecol. 18: 47-60.

Rosselló-Mora, R., &Amann, R. (2001). The species concept for prokaryotes. FEMS microbiology reviews, 25(1), 39-67.

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ & Edwards M (2011). An integrated semiconductor device enabling non-optical genome sequencing. Nature. 475: 348-352.

Roy AS, Baruah R, Gogoi D, Borah M, Singh AK & Boruah HPD (2013). Draft genome sequence of *Pseudomonas aeruginosa* strain N002, isolated from crude oil-contaminated soil from Geleky, Assam, India. Gen. Announce. 1(1): 104-112.

Rubin, E. M. (2008). Genomics of cellulosic biofuels. Nature, 454(7206), 841-845.

Salipante, S. J., Kawashima, T., Rosenthal, C., Hoogestraat, D. R., Cummings, L. A., Sengupta, D. J., & Hoffman, N. G. (2014). Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. Applied and environmental microbiology, 80(24), 7583-7591.

Sanapareddy N, Hamp TJ, Gonzalez LC, Hilger HA, Fodor AA & Clinton SM (2009). Molecular diversity of a North Carolina wastewater treatment plant as revealed by Pyrosequencing. Appl. Environ. Microbiol. 75: 1688–1696.

Schmeisser, C., Steele, H., &Streit, W. R. (2007). Metagenomics, biotechnology with non-culturable microbes. Applied microbiology and biotechnology, 75, 955-962.

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

Shokralla, S., Spall, J. L., Gibson, J. F., &Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. Molecular ecology, 21(8), 1794-1805.

Silva CC, Viero AF, Dias ACF, Andreote FD, Jesus EC, De Paula SO, Torres APR,Santiago VMJ & Oliveira VM (2010). Monitoring the bacterial community dynamics in a petroleum refinery wastewater membrane bioreactor fed with a high phenolic load. J. Microbiol. Biotechnol. 20: 17-25.

Silver S (1996). Bacterial resistances to toxic metal ions-a review. Gene. 179(1): 9-19.

Singh B, Bhat TK, Sharma OP, Kanwar SS, Rahi P & Gulati A (2012). Isolation of tannase-producing *Enterobacter ludwigii* GRT-1 from the rumen of migratory goats. Small Ruminant. Res. 102: 172-176.

Singh, R. V., Sambyal, K., Negi, A., Sonwani, S., & Mahajan, R. (2021). Chitinases production: A robust enzyme and its industrial applications. Biocatalysis and Biotransformation, 39(3), 161-189.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM &Herndl GJ (2006). Microbial diversity in the deep sea and the underexplored 'rare biosphere' Proc. Natl Acad. Sci. 103: 12115-12120.

Sosio M, Bianchi A, Bossi E &Donadio S (2000). Teicoplanin biosynthesis genes in Actinoplanesteichomyceticus. Antonie Van Leeuwenhoek. 78: 379-384.

Srivastava, S., Ghosh, N., & Pal, G. (2013). Metagenomics: mining environmental genomes. Biotechnology for Environmental Management and Resource Recovery, 161-189.

Stoykov, Y. M., Pavlov, A. I., &Krastanov, A. I. (2015). Chitinase biotechnology: production, purification, and application. Engineering in Life Sciences, 15(1), 30-38.

Suganya T, Pandiarajan J, Arunprasanna V, Shanmugam P & Krishnan M (2013). Census of cultivable bacterial community in common effluent treatment plant (CETP) of tannery discharge and computational scrutiny on their leading residents. Bioinformation. 9(2): 101-105.

Sun, S., Chen, J., Li, W., Altintas, I., Lin, A., Peltier, S., ...& Wooley, J. (2010). Community cyberinfrastructure for advanced microbial ecology research and analysis: the CAMERA resource. Nucleic acids research, 39(suppl_1), D546-D551.

Suriya, J., Chandra Shekar, M., Nathani, N. M., Suganya, T., Bharathiraja, S., & Krishnan, M. (2017). Assessment of bacterial community composition in response to uranium levels in

sediment samples of sacred Cauvery River. Applied microbiology and biotechnology, 101, 831-841.

Taş, N., de Jong, A. E., Li, Y., Trubl, G., Xue, Y., & Dove, N. C. (2021). Metagenomic tools in microbial ecology research. Current Opinion in Biotechnology, 67, 184-191.

Torsvik, V., Øvreås, L., &Thingstad, T. F. (2002). Prokaryotic diversity--magnitude, dynamics, and controlling factors. Science, 296(5570), 1064-1066.

Travisany D, Genova AD, Sepúlveda A, Bobadilla-Fazzini RA, Parada P &Maass A (2012). Draft genome sequence of the SulfobacillusthermosulfidooxidansCutipay strain, an indigenous bacterium isolated from a naturally extreme mining environment in Northern Chile. J. Bacteriol. 194(22): 6327-6328.

Veras, A. A. O., de Sá, P. H. C. G., Pinheiro, K. C., Assis das Graças, D., AzevedoBaraúna, R., Cruz Schneider, M. P., ... & Silva, A. (2014). Efficiency of Corynebacterium pseudotuberculosis Cp31 genome assembly with the Hi-Q enzyme on an Ion Torrent PGM sequencing platform. J Proteomics Bioinform, 7, 374-378.

Vidya, J., Swaroop, S., Singh, S., Alex, D., Sukumaran, R., & Pandey, A. (2011). Isolation and characterization of a novel α-amylase from a metagenomic library of Western Ghats of Kerala, India. Biologia, 66(6), 939-944.

Viszwapriya, D., &Karutha Pandian, S. (2017). Metagenomic Approaches for Novel Active Metabolites. Bioresources and Bioprocess in Biotechnology: Volume 1: Status and Strategies for Exploration, 275-302.

Voget S, Steele HL &Streit WR (2006). Characterization of a Metagenome derived halotolerant cellulase. J. Biotechnol. 126: 26-36.

Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K. E., &Streit, W. R. (2003). Prospecting for novel biocatalysts in a soil metagenome. Applied and Environmental Microbiology, 69(10), 6235-6242.

Wagner M, Noguera DR, Juretschko S, Rath G &Koops HP (1998). Combining fluorescent in situ hybridization (FISH) with cultivation and mathematical modeling to study population structure and function of ammonia-oxidizing bacteria in activated sludge. Water Sci. Technol. 37: 441-449.

Wang Z, Zhang XX, Lu X, Liu B, Li Y, Long C & Li A (2014). Abundance and Diversity of Bacterial Nitrifiers and Denitrifiers and Their Functional Genes in Tannery Wastewater Treatment Plants Revealed by High-Throughput Sequencing. PloS one. 9(11): 113-603.

Wang, H., Hart, D. J., &An, Y. (2019). Functional metagenomic technologies for the discovery of novel enzymes for biomass degradation and biofuel production. BioEnergy research, 12, 457-470.

Whiteley, A. S., Jenkins, S., Waite, I., Kresoje, N., Payne, H., Mullan, B., ...& O'Donnell, A. (2012). Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. Journal of microbiological methods, 91(1), 80-88.

Widada J, Nojiri H &omori T (2002). Recent developments in molecular techniques for identification and monitoring of xenobiotic-degrading bacteria and their catabolic gene in bioremediation. Appl. Microbiol. Biotech. 60: 45-59.

Wilmes P, Wexler M & Bond PL (2008). Metaproteomics Provides Functional Insight into Activated Sludge Wastewater Treatment. PLoS ONE. 3(3):1778-1789.

Wilson WA, Mahrenholz AM&Roach PJ (1999). Substrate targeting of the yeast cyclin-dependent kinase Pho85p by the cyclin Pcl10p. Mo.l Cell. Biol. 19(10): 7020-7030.

Wu, H. (2018). Characterizing xylan-degrading enzymes from a putative xylan utilization system derived from termite gut metagenome (Doctoral dissertation, INSA de Toulouse).

Yadav, D., Tanveer, A., & Yadav, S. (2019). Metagenomics for novel enzymes: a current perspective. Environmental Contaminants: Ecological Implications and Management, 137-162.

Yang, C. C., & Iwasaki, W. (2014). MetaMetaDB: A database and analytic system for investigating microbial habitability. PLoS One, 9(1), e87126.

Yang, C., Chowdhury, D., Zhang, Z., Cheung, W. K., Lu, A., Bian, Z., & Zhang, L. (2021). A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. Computational and Structural Biotechnology Journal, 19, 6301-6314.

Yergeau E, Lawrence JR, Sanschagrin S, Waiser MJ, Korber DR, Greer CW (2012). Next-generation sequencing of microbial communities in the Athabasca River and its tributaries in relation to oil sands mining activities. *Appl. Environ. Microbiol.*78(21): 7626-7637.

Yun J, Kang S, Park S, Yoon H, Kim MJ, Heu S&Ryu S (2004). Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. Appl. Environ. Microbiol. 70: 7229-7235.

Zhou M, Rong R, Munro D, Zhu C & Gao X (2013) Investigation of the Effect of Type 2 Diabetes Mellitus on Subgingival Plaque Microbiota by High-Throughput 16S rDNA Pyrosequencing. PLoS ONE. 8(4): 615-616.