

# Metagenomics: A New Approach to Explore Microbiome

Purnima Sharma

Address: Department of Bioinformatics, Hans Raj Mahila Maha Vidyalaya, Jalandhar-144008, Punjab, India

## Abstract

Microbes influence almost every process of life in one or the other way. We have been employing various techniques to study microbial communities and their interactions with the environment. With advances in molecular biology, we are now able to study microbes which were unculturable a few decades ago. The study of microbial communities by sequencing DNA samples from various ecological niches is called metagenomics. Parallel advancements in Next Generation Sequencing (NGS) technologies have made it possible to explore complex biological samples cost-effectively. Generation of billions of reads in a single NGS run poses a significant challenge to store, manage and analyze this vast data. Development of novel bioinformatics applications, specifically for metagenomics data, is therefore vital to assign biological significance to metagenomics data. Science of meta-genomics makes it possible to investigate resource for the development of novel genes, enzymes and chemical compounds for use in biotechnology. Microbes, as communities, are key players in maintaining environmental stability.

## Introduction

Microorganisms constitute two third of the Earth's biological diversity. Importance of microbes in the ecosystem can be estimated from that fact that these unseen microbes influence every part of almost all life on this mother Earth in one way or the other. The number of ways in which microbes influence daily life is countless. Almost every process this living world is touched by the seemingly endless capacity of microbes. All plants and animals have microbial communities as their inherent partners which help them with the availability of essential nutrients, metals, and vitamins. The microbial communities hosted by plants or living around them play a key role in maintaining their health and productivity. Soil is a house to most complexes of these communities most of which have developed synergistic relationship with it. A typical example is nitrogen fixing bacteria which converts the nitrogen from unusable form to the form readily usable by plants. Many specialized soil microbes operate on decaying plants and animals to recycle nutrients while others convert elements, such as iron and manganese, to forms suitable for plant nutrition. Microbes are known to create foods or add values to them through a number of natural processes including fermentation making them useful for human consumption.

Additionally, microbes have been useful to remediate toxins in the environment, both natural as well as produced by human activities including oil and chemical spills. They have been evolved to be a part of diverse strata of this biosphere including soil, water and air, apart from being co- evolving within their living hosts including gut of higher organisms.

The microbes present in the human intestine and mouth helps to extract energy from food that we could not digest apart for providing us protection against a variety of disease-causing pathogens (National Research Council (US), 2007). Recent surge in antibiotic resistance of infectious pathogens makes it more important for us to understand role of microbial communities in disease protection. The detailed understanding of mechanisms by which microbial communities can aid our protection from infectious agents may enable us to exploit specific genes or mechanisms to develop future medicines. Additionally, microbes have been the source of numerous innovations and technologies that have playing key role in improving the human life (Pharmacopeia).

Marine microbes play a very significant role in maintaining the global biogeochemical cycles through a number of chemical transformations. The combine activities of marine microbial communities have global effects on fluxes of energy and matter in the sea, on the composition of Earth's atmosphere, and on global climate. Therefore, the role played by microbial communities in maintaining the chemistry of the oceans if critical in maintaining the habitability of the entire planet (Falkowski *et al.*, 1998).

It is surprising to observe that on the one hand, microbial organisms important players in preserving environmental stability and the health of individuals, while on the other hand, they can cope with extremities of temperature, pressure and pH levels to survive, where no other organism can thrive. This ability of microbes to survive under extreme environment has developed due to their smart strategies for survival. Their genomes have gone through countless biochemical transformations, to make their countless generations of their communities to adapt to enormous environmental changes for the billions of years (Sleator *et al.*, 2008). This diversity presents a bountiful of genetic and biological pool that may be exploited to discover novel products including novel genes, metabolic pathways and their products (Cowan *et al.*, 2005). However, majority of this genetic and functional diversity is still unexploited. Around  $4 \times 10^{30}$  -  $6 \times 10^{30}$  estimated prokaryotic cells are yet to be characterize which represent the largest proportion of individual organisms, comprising  $10^6$  to  $10^8$  separate genospecies (Turnbaugh and Gordon. 2008; Sleator *et al.*, 2008). The study of microorganisms that pervade every nook and corner of this planet has faced many challenges over time, such as discovering unknown organisms and recognizing how they interact with their environment (Escobar-Zepeda *et al.*, 2015). Moreover, in the present scenario marked by unprecedented and dramatic global change, understanding the dynamic role of microbial communities has become more challenging. In the past, microbes were hard to study in their own environments; microbiologists studied individual species in the laboratory one by one. Historically, study of microbes in the laboratories has been limited to single species. This approach has limited our knowledge about different aspects microbial communities as compared to our understanding about microbes as individuals. However, with the advancement in science and technology, we have developed strategies to study microbes in the complex communities in their natural environments thus making us more capable understand more about their capabilities and specific roles. No doubt, traditional approaches to study microbes have established their importance in the ecosystem, the novel approach of metagenomics will further extend our ability to discover much more and novel about microbial

capabilities. The knowledge thus gained, will definitely be applied to better human life, and life as a whole in the current challenges and dynamics of the biosphere (National Research Council (US), 2007).

### **Microbial community study: A historical perspective**

Microbial communities can be defined as the set of organisms (in this case, microorganisms) coexisting in the same space and time (Begon *et al.*, 1986). The roots of microbiology are firmly associated with the microscope with the first record of a human being's seeing a bacterial cell in 1663. Using a homemade microscope, Antonie van Leeuwenhoek watched bacteria recovered from his own teeth. Since the first report of microbes made by Leeuwenhoek and their oral organisms in 1676 (Schierbeek, 1959), our understanding about microbes and their communities has gone far ahead using the current molecular techniques. However, this journey was not straight forward but, in fact, has been marked by some great contributions from some pioneers.

Isolation and cultivation of these “invisible” organisms by Robert Koch later on helped to understand microorganisms' physiologies (Blevins and Bronze, 2010). Resolution of microscopy was significantly enhanced by development of practical staining techniques such as Gram, Ziehl– Neelsen, and Schaeffer and Fulton (Beveridge, 2001; Blevins and Bronze, 2010). Winogradsky achieved a break through by establishing formulations of the culture media which were similar to the natural growth conditions of microbes (McFall-Ngai, 2008). This helped a lot to bridge the gap between the number of observed microorganisms in a microscope and the number of microorganism obtained in culture plates (Staley and Konopka, 1985). Winogradsky's ideas and his contribution to ecology revolutionized microbiology and gave birth to a new concept named “microbial ecology,” which refers to the study of microorganisms and their environmental roles (Ackert, 2012). The study of microorganisms has been based on morphology, characteristics, growth, and selection of some biochemical profiles for almost 300 years (Roszak *et al.*, 1984; Oliver *et al.*, 1991; Colwell *et al.*, 1996). These techniques provided insight into the microbial world but in the present scenario, they extract limited information, to be used for other applications. Our understanding of life at molecular level gave us new opportunities to explore the microbial world using advanced molecular biological techniques. In the late 1970s, Carl Woese gave an idea of using ribosomal RNA genes as molecular markers to classify life (Woese and Fox, 1977). Fortunately, the introduction of Sanger automated sequencing (Sanger *et al.*, 1977), at the same time, made it possible to sequence ribosomal RNA genes from several species. Both these developments revolutionized the study and classification of microorganisms. Due to established evidence that many microbes resist being cultured; the use of molecular markers provided a successful alternative identifying and enumerating microbes without culturing them in laboratories (Pace, N. R. 1997). Later on, further advances in molecular biological techniques such as polymerase chain reaction (PCR), rRNA genes cloning and sequencing, fluorescent *in situ* hybridization (FISH), denaturing gradient gel electrophoresis (DGGE and TGGE), restriction-fragment length polymorphism, and terminal restriction- fragment length polymorphism (T-RFLP) accelerated the exploration and understanding of microbial diversity, making access to a “new uncultured world” of microbial communities (Escobar-Zepeda *et al.*, 2015).

In spite of such advancements in studying microbial world, many aspects of microbial life still remain unanswered especially pertaining to their metabolic and ecological functions. The journey continued with motivation from efforts to discover new genes, explore new functions, and to extract novel metabolic products. These increasing technological applications gave birth to biotechnology. Some successful discoveries including “terragines” from *Streptomyces lividians* (Wang *et al.*, 2000) or genes related to broad-spectrum antibiotics from soil-DNA libraries (Gillespie *et al.*, 2002) set the stage for a new specialization called “metagenomics analysis,” which was later defined as the theoretical collection of all genomes from members in a microbial community from a specific environment (Handelsman *et al.*, 1998).

### **Metagenomics: new and unparalleled insights into microbial communities**

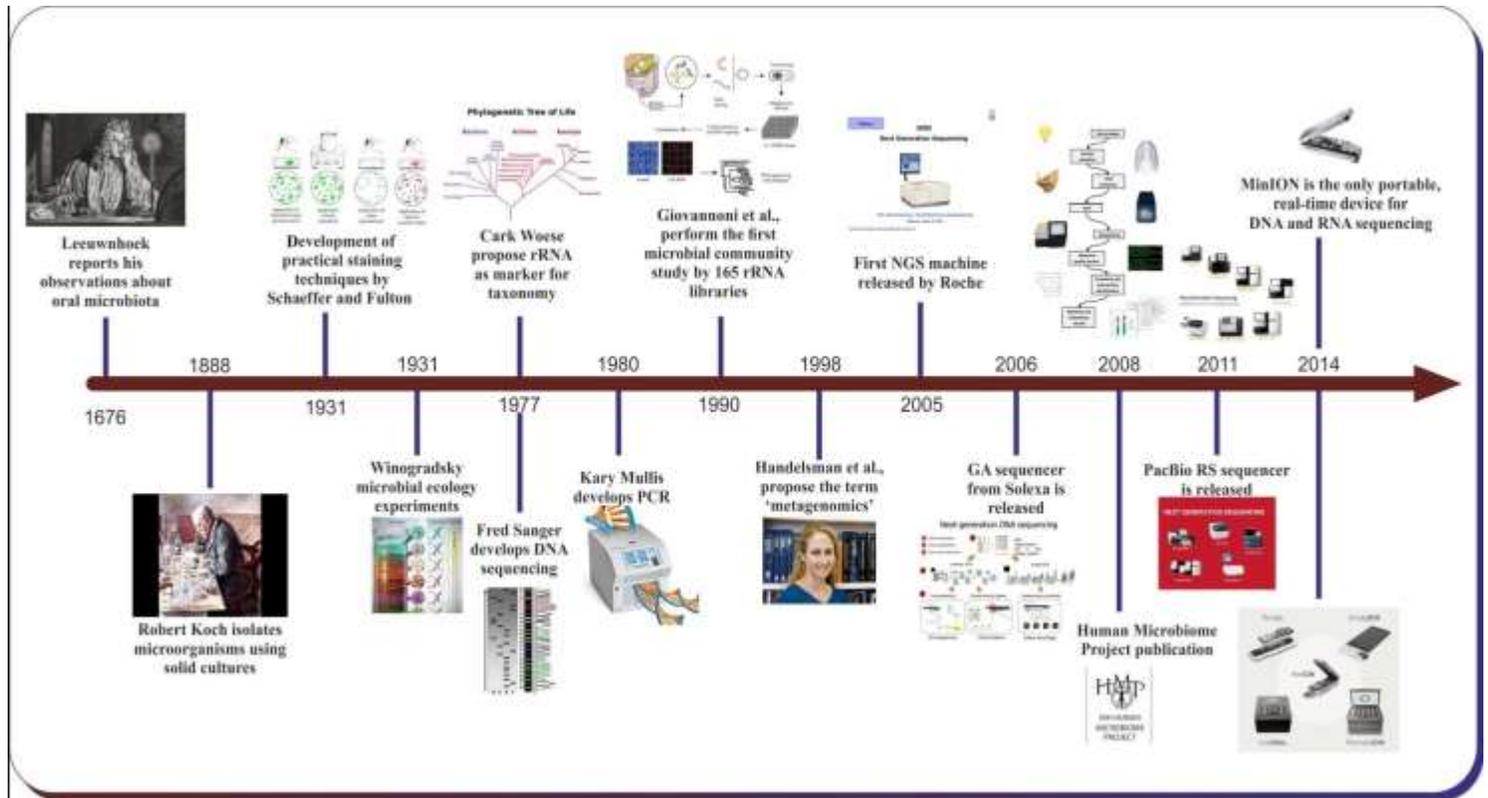
Metagenomics is a powerful combination of genomics, bioinformatics, and systems biology which can be used to study of the genomes of many organisms simultaneously (National Research Council (US), 2007). In fact, ‘meta’, in Greek means “beyond,” and ‘genomics’ mean the study of entire DNA content of an organisms (Handelsman *et al.*, 1998). The metagenomics, therefore simple means study of many genomes at a time. Metagenomics provide new access to the microbial world by making it possible to study uncultured microbial communities which were otherwise impossible to explore. It provides access to the untapped reservoir of novel enzymes, metabolites and other chemicals. Metagenomic approaches can be used as a powerful tool to directly isolate nucleic acids from environmental samples to compare and explore the ecology (Biddle *et al.*, 2008), metabolic profiling (DeLong *et al.*, 2006, Tringe *et al.*, 2005) and identifying novel biomolecules (Daniel, 2005, Ferrer *et al.*, 2009, Handelsman, 2004, Simon and Daniel, 2010; Steele *et al.*, 2009).

The construction of metagenomic libraries starts with extraction of total DNA from environmental samples followed by shotgun cloning of the randomly generated fragments into appropriate cloning vectors. These vectors are then transformed into a host bacterium and followed by screening either by sequencing or testing based on some physiological functions (National Research Council (US), 2007). Metagenomic libraries are constructed in such a way that each part of DNA is at least represented 2–3 times. Small DNA fragments (2–3 kb) provide better coverage of the as compared to the larger fragments. Small fragments are useful in phenotypic studies involving single genes and while reconstructing metagenomes for genotypic analysis. However, large fragments are very much desirable when exploring multigene metabolic pathways. Roughly, at least 1011 genomic clones are required to genomes from rare members of microbial communities (Rastogi and Sani, 2011).

Parallel advancements in Next Generation Sequencing (NGS) technologies have made it possible to explore complex samples from biomedical as well as environmental experiments at molecular level. Major milestones of metagenomics have been summarized in figure 1. It is now possible to sequence billions of reads in a single NGS run. Also generation of novel and robust bioinformatics software programs and pipelines has further revolutionized the analysis of single genes and proteins to a collection of entire set of molecules from whole genomes. Furthermore, in the recent past, innovations and improvements in sequencing instruments lead to

exponential decreases in sequencing costs accelerated the adoption of NGS technology by various stakeholders (Charles and Steven, 2019). A comparative account of various NGS platforms, their running costs, length and quality has been provided in Table 1.

We can summarize that the science of metagenomics has a great potential to explore resources for the development of novel genes, enzymes and chemical compounds for use in biotechnology, healthcare, agriculture, ecology and environment.



**Figure 1:** Timeline of major milestones in metagenomics. Figure adapted from Escobar-Zepeda *et al.*, 2015

## An overview of experimental techniques for metagenomics

### Conventional Methods

This section gives a quick overview of various methods used that have been used to examine and study microbial communities.

#### Denaturing gel electrophoresis (DGGE)

DGGE was first introduced by Muyzer *et al.* in 1993 which has been used to separate fragments of DNA sequence having same size but differ in their composition. This difference in composition gives DNA fragments unique melting points and makes them behave differently under denaturing conditions. Principally, GC pair forms 3 hydrogen bonds while AT pair forms 2 hydrogen bonds. Therefore, DNA fragments possessing higher GC content tends to melt at higher temperature as compared to those possessing higher AT content. Sample DNA fragments are loaded into a polyacrylamide gel with gradient of denaturing strength to separate them as they move along the gel. This method can also be used to detect transversions and transitions in the DNA sequence

(Fischer and Lerman, 1982). Another technique, temperature gel gradient electrophoresis (TGGE) is also used to separate DNA fragments of similar size. However, TGGE uses temperature gradient instead of chemical gradient.

### **Single Strand Conformation Polymorphism (SSCP)**

SSCP is similar to DGGE in terms of using electrophoresis and PCR but it differs in using non-denaturing polyacrylamide gel for separating DNA fragments (Fakruddin *et al.*, 2013; Schwieger and Tebbe, 1998). It is based on the principle that single stranded DNA (ssDNA) folds uniquely, leading to differences in mobility. DNA from environmental sample is extracted, amplified by PCR, and then denatured to produce ssDNA fragments that are subjected to gel electrophoresis for separation. This technique is very suitable for DNA fragments of 150 to 400 bp in length.

### **2. Ribosomal Intergenic Spacer Analysis (RISA)**

RISA is a molecular technique introduced by Carlse Woese in 1987 that uses the polymerase chain reaction (PCR) based amplification of a specific portion of the intergenic spacer region (ISR) between the genes coding for 16S and 23S ribosomal subunits. These ISRs are very informative in the sense that they contain a unique combination of well conserved as well as significantly variable regions between species. The conserved regions provide an advantage that common primers can be used across species to carry our PCR reaction to amplify ISRs (Fisher *et al.*, 1999), while variable regions are used to differentiate within the species and aid in further downstream analysis. A more sophisticated variant of RISA called automated ribosomal intergenic spacer analysis (ARISA) uses a fluorescently labeled forward primer and a detector (Rastogi and Sani, 2011). This automated version not only allows more than one sample but also generates highly reproducible bacteria community profiles.

### **Restriction Fragment Length Polymorphism (RFLP) or Terminal Restriction Fragment Length Polymorphism (T-RFLP)**

This is another PCR based molecular technique and uses specifically, 5' fluorescently labeled primer during the polymerase chain reaction. It is used to study microbial diversity and it is based on single nucleotides polymorphisms in DNA sequences. The 16S rDNA fragments amplified using PCR are digested with restriction enzymes and electrophoresed in agarose or acrylamide gels. RFLP technique has been utilized in various metagenomic studies including estimating the prokaryotic diversity in hyper saline ponds. (Martinez-Murcia *et al.*, 1995).

### **Fluorescent In-situ Hybridization (FISH)**

FISH is one most versatile technique that allows at source (in situ) phylogenetic identification and enumeration of individual microbial cells using whole cell hybridization to oligonucleotides (18-30 nucleotides long) probes (Amann *et al.*, 1995). The 5' end of the probe is fluorescently labeled with a dye which facilitates its detection epifluorescent microscope. The signal strength is correlated to the growth rates and cellular rRNA content and therefore can be applied to study metabolism dynamics of the communities. Use of FISH is sometimes limited

low intensity of signal, background noises, and poor accessibility of targets. However, its new variants including catalyzed reporter deposition (CARD) FISH, flow cytometry coupled FISH, and ion mass spectrophotometry coupled FISH are used to get better resolutions (Rastogi and Sani, 2011).

### Next Generation Sequencing (NGS) Technologies

Sanger sequencing (Sanger and Coulson 1975) which was developed on dideoxyribonucleotide (ddNTP), chain termination technology has revolutionized many initial studies on microbial communities. However, the field of sequencing has evolved a lot in capacity, capability and applications (Levy and Myers, 2016). The present day sequencing experiments are much more demanding in terms of sequence length as well as sequence yield while looking for low cost alternatives. Advancements in sequencing chemistry has lead to emergence of high throughput sequencing technologies called Next Generation Sequencing technologies (NGS) and third generation sequencing which allows running millions of sequencing experiments in parallel. In addition their underline chemistry, the NGS technologies (Table 1) also differ in the template preparation, sequence yields, read length, quality of reads and sequencing cost (Loman and Pallen, 2015; Escobar-Zepeda *et al.*, 2015).

**Table 1:** Comparative view of sequencing technologies for metagenomics. Adapted from Glenn, 2014 and Escobar-Zepeda *et al.*, 2015. a: P6-C4 chemistry, b:MiSeq read length, Illumina HiSeq 2500 dual flow cell yield.

SEQUENCING PLATFORMS	Roche 454	IonTorrent PGM	Illumina	PacBio RSII <sup>a</sup>
Maximum read length (bp)	1200	400	300b	50,000
Output per run (Gb)	1	2	1000 <sup>c</sup>	1
Library Amplification	Yes	Yes	Yes	No
Cost per Gb (US Dollar)	\$9538.46	\$460.00	\$29.30	\$600
Error kind	Indel	Indel	Substitution	Indel
Error rate (%)	1	~1	~0.1	~13
Run time	20h	7.3h	6 days	2h

NGS offers following major advantages over Sanger sequencing (Cao *et al.*, 2017): -

1. In vitro construction of the sequencing libraries
2. In vitro clonal amplification of DNA fragments
3. Array based sequencing allowing DNA fragments to be multiplexed
4. Solid phase immobilization of DNA.

The following three NGS technologies have been developed based on the way DNA is immobilized on solid surface (Cao *et al.*, 2017): -

- (a) High throughput pyrosequencing on beads
- (b) Sequencing by ligation on beads
- (c) Sequencing by synthesis on a glass substrate

**High throughput pyrosequencing on beads:** Pyrosequencing, developed at the Royal Institute of Technology (KIT) was NGS technology to be used as an alternative to the conventional Sanger method implemented using the Roche 454 GS20 pyrosequencing platform (Margulies, 2005). In this technology, a DNA template is sheared in to fragments (using sonication or enzyme based digestion) which are ligated with oligonucleotide adapters. These primed fragments are then attached to the microbeads (28- $\mu\text{m}$ ) and then amplified using emulsion PCR. The amplicon bearing beads are then captured in the wells of Pico TiterPlate followed by addition of DNA polymerase, ATP sulfurylase, luciferase, apyrase, luciferin and adenosine 5'-phosphosulfate (APS) (Margulies, 2005). In every sequencing cycle, pyrophosphate is released with the addition of appropriate dNTP, which gets converted into ATP in the presence of APS. The ATP released converts luciferin to oxyluciferin catalyzed by luciferase, producing visible light signal which is detected by a charge coupled detector (CCD) camera and then converted into nucleotide sequence. The amount of this signal is proportional to the amount of ATPs. Pyrosequencing, also known as sequencing by synthesis, is prone to high error rates. Whenever the polymerase runs through homopolymeric sequences, it does not accurately reports the number of positions and often  $n$  nucleotides are read as  $n-1$  nucleotides (Cao *et al.*, 2017). Also, pyrosequencing produces artificial replicate sequences making difficult to estimate actual gene abundance (Thomas *et al.*, 2012). Pyrosequencing has been successfully used in identification of several microbial populations including discovery of *Myxococcus xanthus*, from soil (Vos and Velicer, 2006) and discovery of a novel cellulase from a thermophilic cellulose- degrading microbial consortium (Zhao *et al.*, 2017). Additionally many microbial communities were surveyed from different environments *viz.* underground mine water, marine, fresh water, fish, corals terrestrial animals and mosquitoes using pyrosequencing (Dinsdale *et al.*, 2008).

### **Sequencing by ligation over beads:**

SOLiD (Sequencing by Oligonucleotide Ligation and Detection) is a NGS technology which based on the Multiplex Polony Sequencing technology (Shendure *et al.*, 2005). This NGS technology was developed by Life Technologies and has been used commercially since 2006. A fragment library is prepared by shearing the sample DNA. The fragments obtained are attached to paramagnetic beads (1- $\mu\text{m}$ ) using universal p1 adapters and are amplified using PCR to get clonally amplified populations in such a way that only one species of fragment is present on the surface of each magnetic bead. These beads populated with PCR amplicons are then covalently immobilized to a glass slide and are hybridized using universal PCR primers which are complementary to the adapter used. During each sequencing cycle, a fluorescently labeled DNA octamer is ligated to the universal

primer revealing the positional identity of the nucleotide (Cao *et al.*, 2017). This ligation process is followed by chemical cleavage leaving a pentamer on the DNA template. This process is repeated through a number of cycles to get information about the DNA sequence. This technology generates 108- 109 small sequence reads at one time. This platform utilizes 2 base encoding to decode the raw data and therefore it easily identify miscalls and gives about 99.94% accuracy.

### Sequencing by synthesis on a glass solid phase surface:

Illumina Genome Analyser (SOLEXA) (Fedurco *et al.*, 2006; Turcatti *et al.*, 2008) immobilizes random DNA fragments on a surface and then performs solid surface PCR amplification, resulting in clusters of identical DNA fragments. The widely adopted Illumina sequencing utilizes massive parallel sequencing and detects every base as it gets incorporated in the growing

DNA strand. This process starts by ligating both 3' and 5' ends of the random DNA fragments with universal adapters. The forward and reverse primers, complementary to the adapters are first immobilized on the glass surface using flexible linkers. The adapter ligated DNA fragments are then hybridized to the forward and reverse primers. Bridge PCR (Cao *et al.*, 2017) is then

amplifies the DNA fragment generating a "cluster" of clonal amplicons. After cluster generation, the sequencing primer hybridizes to the universal sequence flanking the region of interest. Sequencing process completes in a cyclic manner using a strategy known as sequencing by synthesis (SBS). Nucleotides are labeled with a chemically cleavable fluorescent reporter group at the 3'-OH end such that only a single base can be added in each cycle. Each cycle extends a single base followed by the chemical cleavage of the fluorescent reporter that will identify the incorporated nucleotide. With continuous improvements in Illumina sequencing, we can achieve read length of about 150 bp. Moreover, clustered fragments can be sequenced from both increasing the sequencing yields (Thomas *et al.*, 2012). The basic challenges of Illumina technology are signal decay and dephasing caused by incomplete fluorescent label cleavage or terminating moieties leading to about 1-1.5% average raw error rates (Cao *et al.*, 2017). **Ion Torrent:** Ion Torrent was introduced by Life Technologies in 2012 as light independent sequencing technology (Rothberg *et al.*, 2011) Ion Torrent semiconductor sequencing is also sequencing by synthesis approach in which hydrogen ions are emitted with the incorporation of nucleotide bases in the growing DNA chain. The release of the hydrogen ions causes a change in the pH of the surrounding solution which is proportional to the number of nucleotides added in the DNA chain. A pH sensor on a silicon substrate is used as a detector that can detect minuscule (0.02 pH unit) and transient (with a half-life <1 s) pH changes. Millions of pH sensors are arranged on the semiconductor chips for fast and cost efficient sequencing. Till now, a number of ion torrent sequencers have been released including Ion S5 Systems, Ion AmpliSeq Technology,

Ion Proton System, Ion PGM System, Ion Chef System, Ion Reporter Software and Server, Ion AmpliSeq Exome Certified Service Provider Program

### Third Generation Sequencing (Single Molecule Long Read Sequencing-SMRT)

All NGS technologies discussed above are based on PCR amplification to generate clonally amplified DNA fragments which becomes a source of bias in read distribution, ultimately affecting coverage. Third generation sequencing technologies have been developed which does not PCR based amplification and can sequence single DNA molecules thereby eliminating amplification associated bias, intensity averaging, phasing or synchronization problems, reducing error rates (Cao *et al.*, 2017)

a. **Single-molecule-real-time (SMRT) Technology:** SMRT was the first long read based sequencing technology commercially introduced by Pacific Biosciences (Eid *et al.*, 2009). In

SMRT, adapter molecules are ligated to both ends of the DNA fragment to be sequenced, which then forms a closed circular DNA molecule. The circular DNA molecule is then loaded

into a SMRT cell containing 150,000 zeptolitre wells in such a way that each well contains a single immobilized DNA polymerase at its base. DNA polymerase then binds to the hairpin adaptors present on the circular DNA molecule to initiate the process of replication. Four fluorescently labeled nucleotides are then introduced into the reaction wells. As each base is enzymatically incorporated, a light pulse is produced that identifies the base and analyzed iteratively to generate the DNA sequence (Rhoads and Au, 2015). The main advantage of the

SMRT sequencing is the read length obtained which varies from 1500 bp in original C1 generation sequencers to 10-kbp in present day C5 generation sequencers. In an average, a PacBio RS II system gives sequence yield of about 0.5–1 billion bases per SMRT cell.

However, the error rate of this technology is significantly higher (approximately 11- to 15%) (Cao *et al.*, 2017).

b. **HeliScope:** HeliScope (Braslavsky *et al.*, 2003) is also based on single DNA molecule sequencing technology. In HeliScope, DNA molecules are sequenced by synthesis using a highly sensitive fluorescence detection system. A DNA library is prepared by generating random DNA fragments followed by poly A tailing. The poly-A tail is then hybridized to a surface containing tethered poly T oligomers. This creates an array of primer annealed single molecule DNA templates. In the sequencing step, DNA polymerase adds a single nucleotide at a time to these primed, immobilized fragments to achieve a template-dependent extension. Each fluorescently labeled nucleotide is added one at a time and therefore detected by the analysis of its corresponding signal image. The cycle is then repeated with a new species of nucleotide (Cao *et al.*, 2017).

c. **MinION:** MinION (Oxford Nanopore Technology; Kasianowicz *et al.*, 1996) was released in 2014 through the MinION Access Programme (MAP). MinION uses electrophoresis move the DNA/RNA molecule through a nanopore. A MinION flow cell consists of an array of 512 sensors, each of which is connected to four nanopores which allows it to measure the current through the pore thousands of times a second. As the nucleic

acid passes through the nanopore, the change in the current pattern and magnitude is measured (Meller *et al.*, 2000). During the library preparation step, double stranded DNA is sheared using a Covaris g-TUBE and fragmented DNA is repaired using a PreCR step. Blunt ended DNA molecules are then created using an end repair step before a poly A tail is added to the 3-OH end. Two adaptors are then added to the DNA, a Y adapter (so called, due to its shape) and a hair pin adaptor. A motor protein unzips the double stranded DNA at the Y adapter and feeds the DNA as a single strand through the nanopore. The sequence of this single stranded DNA is read through Base calling using software called MinKNOW. In this way, read length of a few hundred thousand base pairs is achieved with an accuracy ranging from 65 to 88% (Cao *et al.*, 2017). If information from only one strand is used, base calling is 1-dimensional (1D), otherwise it is 2-dimensional (2D) system (Lu *et al.*, 2016). Due to the small size of the instrument, low cost and the real-time nature, the MinION platform is becoming a method of choice among genomics community particularly for pathogen surveillance and diagnostics (Judge *et al.*, 2015; Quick *et al.*, 2015). For example MinION technology has been used for in-field genomic epidemiology studies of the West Africa Ebola virus outbreak (Quick *et al.*, 2016). Also, this technology was used in Environmental metagenomics to identify ecologically and epidemiologically important components of sediments, soils, waters, and surfaces, employing whole genome sequence analysis to (Brown *et al.*, 2017).

Apart from being fast and able to generate high throughput data, the major advantage of second (NGS) and third generation sequencing technologies is that the library preparation does not require DNA cloning vectors or bacterial hosts, simplifying the library preparation and reducing DNA contamination from other organisms that are not part of the metagenome (Escobar-Zepeda *et al.*, 2015). The powerful new generation technologies have been applied to study novel microbial communities and even exploring entirely new environments, they do have some limitations (Table 1), giving room for further improvement and innovation. It is important to note that the mammoth of data generated through new generation sequencing technologies poses challenges to store, manage and interpret this data, requiring state-of-the-art computational resources and complex bioinformatics strategies.

### **Meta-genomics Applications**

In the field of microbial world, meta-genomics have demonstrated as quickly developing weapon and has changed the way, which microbiologist confronted numerous issues. Among the techniques intended to access the physiology and hereditary qualities of uncultured life forms, metagenomics, the genomic examination of a populace of microorganisms, has risen as a ground-breaking highlight. It has been assessed that under 1% of the microorganisms in the common habitat can be refined in the research facility. It is progressively perceived that countless normal items exists in unculturable organisms with concoction, natural, and utilitarian exercises for potential uses in different mechanical and biomedical applications (Handelsman *et al.*, 2004). Metagenomics gives a boundless asset to the improvement of novel qualities, chemicals, normal items, bioactive mixes, and bioprocesses that may significantly affect modern and biotechnological applications.

## 1 Enzymes and Metagenomics:-

Metagenomics has proven powerful approach for the ample demand of novel enzymes and biocatalysts (Lorenz *et al.*, 2002). Cellulases, lipases, xylanases, amylases, proteases, and various other industrially important enzymes have been produced through metagenomics. The following are some of the main enzymes that have been unlocked from genetically untapped resources.

### Cellulases

Cellulases are a type of glycosyl hydrolases (GHs) that contain three major types of enzymatic activities: (i) endoglucanases (EC 3.2.1.4), (ii) exoglucanase (iii)  $\beta$ -glucosidases (EC 3.2.1.21) (Li *et al.*, 2008). Cellulases have been of special interest due to the diversity of applications they present including, their use as animal feeds as an agent to improve the nutritional quality as well as digestibility, fruit juice processing, baking. Various natural environments contain cellulases including soil, rumen, compost etc. from which they have been isolated by using different metagenomic techniques. Cellulases have been isolated from niche environment including anaerobic digester, alkaline and saline lakes (Rees *et al.*, 2003). These metagenomic strategies helped to isolate many novel enzymes from a variety of environments including GH12 cellulase from rice straw compost (Yeh *et al.*, 2013), endo- $\beta$ -1,4-glucanases and  $\beta$ -glucosidases from microbial consortium collected from forest soil, dung of elephant, cow rumen and rotted tree .

### Xylanases

Xylanases (endo-1,4-xylanases, EC 3.2.1.8) acts on xylan and cleave its backbone into smaller oligosaccharides. Like cellulases, Xylanases are another important enzymes having wide range of industrial applications including in extraction and preparation of beverages (Wong *et al.*, 1988), purification of juices (Pawan *et al.*, 2012), detergents (Kumar *et al.*, 2004), generation of plant cell protoplast (Kulkarni *et al.*, 1999), production of antimicrobial agents (Christakopoulos *et al.*, 2001), antioxidants (Katapodis *et al.*, 2003), surfactants (Kashyap *et al.*, 2014) etc. Xylanases have been produced by a variety of microbes living on a variety of natural environments thereby making them one of the favorite candidates for metagenomic studies.

### Lipases

Lipases are triacylglycerol acylhydrolases (EC 3.1.1.3) that catalyze the hydrolysis of triacylglycerol to glycerol and fatty acids. They have gained special attention from industry due to their robust nature making them resistance to extremities of temperature, pH, Organic solvents etc. Many plants, animals and microbes contain lipases with lipases of microbial origin gaining special owing to their vast industrial applications including making of oils, fats, detergents, dairy products and pharmaceuticals (Cardenas *et al.*, 2001). Metagenomics have been used to isolate lipases from various environmental samples including thermal (Rhee *et al.*, 2005), saline lake (Rees *et al.*, 2003), field soil (Henne *et al.*, 2000), marine sediments (Sivasubramani *et al.*, 2012) and drinking water.

## Proteases

Proteases represent one of the most important classes of enzymes having immense applications in research as well as industry. They have been present in almost all living organisms including plants, animals and microbes. Proteases of microbial origins have gained special interest due to their unique applications in biotechnological and pharmaceutical industry. Recently, several novel proteases have been isolated using metagenomic approach in the recent times.

## 2 Metagenomics and Medicine:-

For a long time, microbes have been the source of many antibiotics and other medical agents and played a great role in improving human health. However, we have almost reached a saturation point of discovering novel useful products from microorganisms using traditional culturing methods results. We need new strategies for discovering new medicinal products from microorganisms. The metagenomic approach has been considered promising to discover new avenues of novel antibiotics. Recently, many novel antibiotics have been discovered using this strategy including isolation of Turbomycin A and B from Soil, didemnin B (Aplidine™) and thiocoraline for cancer treatment (Liu et al., 2010), red indirubin pigment possessing antibacterial activity (Lim *et al.*, 2005).

## 3. Metagenomics and Biosurfactants:-

As chemical surfactants pose toxicity and environmental concerns, focus is being shifted to developing biosurfactants. With the advances in biotechnology, biosurfactants are emerging as potential replacement of chemical surfactants in industrial applications including lubrication, wetting, foaming, de-foaming, emulsification, softening, fixing etc. Also biosurfactants have been sought out in biomedical, food and pharmaceutical industry (Henkel *et al.*, 2012). Metagenomics has been used to make DNA libraries of the petroleum-contaminated samples (soil, water, etc.) followed by the screening of biosurfactant producing clones. Morikawa and his co-workers in 1992 reported two bacteria (A-1 and B-1) which exhibited large emulsified halos around their colonies on oil-L-agar plates were isolated (Morikawa *et al.*, 1992).

## 4. Metagenomic and Biodegradation:-

There are heaps of various types of waste, for example, oil spills and the deficient ignition of non-renewable energy sources, created by enterprises have caused a gathering of oil hydrocarbons in the earth. The seepage of these anthropogenic mixes, through industry involved in production oil and related products, leads to deposition of lot of sweet-smelling hydrocarbons, thus polluting natural systems (Jacques *et al.*, 2007). Microorganisms play significant role in biogeochemical cycles, specifically in degrading a variety of carbon based natural products. These can be therefore utilized to degrade sweet-smelling rings such as benzene, toluene, and xylene rings, and mineralize their carbon skeleton (Alexander M *et al.*, 1994). Recently, researchers have discovered genes and their respective metabolic pathways, involved in degradation of a variety of aromatic compounds in sludge waste samples from waste water treatment unit of a petroleum refinery. They have used a metagenomic approach to capture wide range of the extant functional diversity (Cynthia *et al.*, 2013). Also, bacterial populations, capable

of degrading polycyclic aromatic hydrocarbon (PAH), have been identified from cold marine ecosystems (Marcos *et al.*, 2009).

## Conclusion

Metagenomics is very dynamic and robust strategy to study the microbial world and has provided significant insights in from its inception. We have even been able to study microbial populations which were considered unculturable in the recent past. Advancements in molecular techniques, computational resources and data analytics have made it possible to scale up metagenomics to deal with mega projects which were never thinkable few decades ago. As many other technologies, metagenomics is developing and therefore face a lot of challenges. However, one cannot rule out the opportunities this technology offers in terms of making new discoveries into the microbial world in particular and environment as a whole. Metagenomics, thus provides a window into a world of unseen microbial diversity which can be exploited using biotechnological tools thereby paving way to novel scientific, environmental, pharmaceutical and industrial applications.

## References

- Escobar-Zepeda, A., Vera-Ponce de Leon, A., & Sanchez-Flores, A. (2015). The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in genetics*, 6, 348.
- Falkowski, P. G., Barber, R. T., & Smetacek, V. (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science*, 281(5374), 200-206.
- Staley, J. T., & Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual review of microbiology*, 39(1), 321-346.
- Schierbeek, A. (1959). *Measuring the invisible world: the life and works of Antoni van Leeuwenhoek* (No. 37). Abelard-Schuman.
- Cowan, D., Meyer, Q., Stafford, W., Muyanga, S., Cameron, R., & Wittwer, P. (2005). Metagenomic gene discovery: past, present and future. *TRENDS in Biotechnology*, 23(6), 321-329.
- Begon, M., Harper, J. L., & Townsend, C. R. (1986). *Ecology. Individuals, populations and communities*. Blackwell scientific publications.
- Blevins, S. M., & Bronze, M. S. (2010). Robert Koch and the 'golden age' of bacteriology. *International Journal of Infectious Diseases*, 14(9), e744-e751.
- Beveridge, T. J. (2001). Use of the Gram stain in microbiology. *Biotechnic & Histochemistry*, 76(3), 111-118.
- McFall-Ngai, M. (2008). Are biologists in 'future shock'? Symbiosis integrates biology across domains. *Nature Reviews Microbiology*, 6(10), 789-792.
- Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science*, 276(5313), 734-740.
- Ackert, L. (2012). *Sergei Vinogradskii and the Cycle of Life: From the Thermodynamics of Life to Ecological Microbiology, 1850-1950* (Vol. 34). Springer Science & Business Media.
- Roszak, D. B., Grimes, D. J., & Colwell, R. R. (1984). Viable but nonrecoverable stage of *Salmonella enteritidis* in aquatic systems. *Canadian journal of microbiology*, 30(3), 334-338.
- Oliver, J. D., Nilsson, L., & Kjelleberg, S. (1991). Formation of nonculturable *Vibrio vulnificus* cells and its relationship to the starvation state. *Appl. Environ. Microbiol.*, 57(9), 2640-2644.
- Colwell, R. R., Brayton, P., Herrington, D., Tall, B., Huq, A., & Levine, M. M. (1996). Viable but non-culturable *Vibrio cholerae* O1 revert to a cultivable state in the human intestine. *World Journal of Microbiology and biotechnology*, 12(1), 28-31.

- Fox, G. E., & Woese, C. R. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA*, 74, 5088-5090.
- National Research Council. (2007). *The new science of metagenomics: revealing the secrets of our microbial planet*. National Academies Press.
- Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.*, 68(4), 669-685.
- Sanger, F. N. S. and Coulson, AR (1977) Proc. In *Natl. Acad. Sci. USA* (Vol. 74, pp. 5463-5467).
- Wang, G. Y. S., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J., ... & Davies, J. (2000). Novel natural products from soil DNA libraries in a streptomycete host. *Organic letters*, 2(16), 2401-2404.
- 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. *Appl. Environ. Microbiol.*, 68(9), 4301-4306.
- Sleator, R. D., Shortall, C., & Hill, C. (2008). Metagenomics. *Letters in applied microbiology*, 47(5), 361-366.
- Turnbaugh, P. J., & Gordon, J. I. (2008). An invitation to the marriage of metagenomics and metabolomics. *Cell*, 134(5), 708-713.
- Tringe, S. G., Von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., ... & Bork, P. (2005). Comparative metagenomics of microbial communities. *Science*, 308(5721), 554-557.
- Simon, C., & Daniel, R. (2010). Construction of small-insert and large-insert metagenomic libraries. In *Metagenomics* (pp. 39-50). Humana Press, Totowa, NJ.
- Steele, H. L., Jaeger, K. E., Daniel, R., & Streit, W. R. (2009). Advances in recovery of novel biocatalysts from metagenomes. *Journal of molecular microbiology and biotechnology*, 16(1-2), 25-37.
- 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.
- DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N. U., ... & Chisholm, S. W. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science*, 311(5760), 496-503.
- Daniel, R. (2005). The metagenomics of soil. *Nature Reviews Microbiology*, 3(6), 470-478.
- Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.*, 68(4), 669-685.
- 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), R245R249.
- Rastogi, G., & Sani, R. K. (2011). Molecular techniques to assess microbial community structure, function, and dynamics in the environment. In *Microbes and microbial technology* (pp. 29-57). Springer, New York, NY.
- Muyzer, G., De Waal, E. C., & Uitterlinden, A. G. (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(3), 695-700.
- Fischer, S. G., & Lerman, L. S. (1983). DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. *Proceedings of the National Academy of Sciences*, 80(6), 1579-1583.
- Schwieger, F., & Tebbe, C. C. (1998). A new approach to utilize PCR–single-strand-conformation polymorphism for 16S rRNA gene-based microbial community analysis. *Appl. Environ. Microbiol.*, 64(12), 4870-4876.
- Fisher, M. M., & Triplett, E. W. (1999). Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Appl. Environ. Microbiol.*, 65(10), 4630-4636.
- Fakruddin, M., & Mannan, K. (2013). Methods for analyzing diversity of microbial communities in natural environments. *Ceylon Journal of Science (Biological Sciences)*, 42(1).

- Martinez-Murcia, A. J., Acinas, S. G., & Rodriguez-Valera, F. (1995). Evaluation of prokaryotic diversity by restrictase digestion of 16S rDNA directly amplified from hypersaline environments. *FEMS Microbiology Ecology*, 17(4), 247-255.
- Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Mol. Biol. Rev.*, 59(1), 143-169.
- Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology*, 94(3), 441-448.
- Levy, S. E., & Myers, R. M. (2016). Advancements in next-generation sequencing. *Annual review of genomics and human genetics*, 17, 95-115.
- Loman, N. J., & Pallen, M. J. (2015). Twenty years of bacterial genome sequencing. *Nature Reviews Microbiology*, 13(12), 787-794.
- 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.
- Vos, M., & Velicer, G. J. (2006). Genetic population structure of the soil bacterium *Myxococcus xanthus* at the centimeter scale. *Appl. Environ. Microbiol.*, 72(5), 3615-3625.
- Dinsdale, E. A., Edwards, R. A., Hall, D., Angly, F., Breitbart, M., Brulc, J. M., ... & Desnues, C. 508 Haynes M, Li L, McDaniel L, Moran MA, Nelson KE, Nilsson C, Olson R, Paul J, Brito 509 BR, Ruan Y, Swan BK, Stevens R, Valentine DL, Thurber RV, Wegley L, White BA, 510 Rohwer F. 2008. Functional metagenomic profiling of nine biomes. *Nature*, 452, 629-32.
- Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M., ... & Church, G. M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, 309(5741), 1728-1732.
- Fedurco, M., Romieu, A., Williams, S., Lawrence, I., & Turcatti, G. (2006). BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic acids research*, 34(3), e22-e22.
- Turcatti, G., Romieu, A., Fedurco, M., & Tairi, A. P. (2008). A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic acids research*, 36(4), e25-e25.
- Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics-a guide from sampling to data analysis. *Microbial informatics and experimentation*, 2(1), 3.
- Cao, Y., Fanning, S., Proos, S., Jordan, K., & Srikumar, S. (2017). A review on the applications of next generation sequencing technologies as applied to food-related microbiome studies. *Frontiers in microbiology*, 8, 1829.
- 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.
- Rhoads, A., & Au, K. F. (2015). PacBio sequencing and its applications. *Genomics, proteomics & bioinformatics*, 13(5), 278-289.
- Zhao, C., Chu, Y., Li, Y., Yang, C., Chen, Y., Wang, X., & Liu, B. (2017). High-throughput pyrosequencing used for the discovery of a novel cellulase from a thermophilic cellulose-degrading microbial consortium. *Biotechnology letters*, 39(1), 123-131.
- Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., ... & Hoon, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475(7356), 348-352.
- Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION sequencing and genome assembly. *Genomics, proteomics & bioinformatics*, 14(5), 265-279.
- Judge, K., Harris, S. R., Reuter, S., Parkhill, J., & Peacock, S. J. (2015). Early insights into the potential of the Oxford Nanopore MinION for the detection of antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy*, 70(10), 2775-2778.
- Quick, J., Ashton, P., Calus, S., Chatt, C., Gossain, S., Hawker, J., ... & De Pinna, E. (2015). Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome biology*, 16(1), 114.
- Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., ... & Ouédraogo, N. (2016). Realtime, portable genome sequencing for Ebola surveillance. *Nature*, 530(7589), 228-232.
- Kasianowicz, J. J., Brandin, E., Branton, D., & Deamer, D. W. (1996). Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences*, 93(24), 13770-13773.

- Meller, A., Nivon, L., Brandin, E., Golovchenko, J., & Branton, D. (2000). Rapid nanopore discrimination between single polynucleotide molecules. *Proceedings of the National Academy of Sciences*, 97(3), 1079-1084.
- Brown, B. L., Watson, M., Minot, S. S., Rivera, M. C., & Franklin, R. B. (2017). MinION™ nanopore sequencing of environmental metagenomes: a synthetic approach. *Gigascience*, 6(3), gix007.
- 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.
- 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), 430436.
- 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.
- 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.
- Wang, F., Li, F., Chen, G., & Liu, W. (2009). Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches. *Microbiological research*, 164(6), 650-657.
- Wong, K. K., Tan, L. U., & Saddler, J. N. (1988). Multiplicity of beta-1, 4-xylanase in microorganisms: functions and applications. *Microbiological reviews*, 52(3), 305.
- Sharma, P. K., & Chand, D. (2012). *Pseudomonas* sp. xylanase for clarification of Mausambi and Orange fruit juice. *International Journal of Advancements in Research and Technology*, 1(2), 1-3.
- Kumar, B. K., Balakrishnan, H., & Rele, M. V. (2004). Compatibility of alkaline xylanases from an alkaliphilic *Bacillus* NCL (87-6-10) with commercial detergents and proteases. *Journal of Industrial Microbiology and Biotechnology*, 31(2), 83-87.
- Kulkarni, N., Shendye, A., & Rao, M. (1999). Molecular and biotechnological aspects of xylanases. *FEMS microbiology reviews*, 23(4), 411-456.
- Christakopoulos, P., Katapodis, P., Kalogeris, E., Kekos, D., Macris, B. J., Stamatis, H., & Skaltsa, H. (2003). Antimicrobial activity of acidic xylo-oligosaccharides produced by family 10 and 11 endoxylanases. *International Journal of Biological Macromolecules*, 31(4-5), 171-175.
- Katapodis, P., Vardakou, M., Kalogeris, E., Kekos, D., Macris, B. J., & Christakopoulos, P. (2003). Enzymic production of a feruloylated oligosaccharide with antioxidant activity from wheat flour arabinoxylan. *European Journal of Nutrition*, 42(1), 55-60.
- Kashyap, R., & Subudhi, E. (2014). A novel thermoalkaliphilic xylanase from *Gordonia* sp. is salt, solvent and surfactant tolerant. *Journal of basic microbiology*, 54(12), 1342-1349.
- 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(4-6), 111-123.
- Rhee, J. K., Ahn, D. G., Kim, Y. G., & Oh, J. W. (2005). New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. *Appl. Environ. Microbiol.*, 71(2), 817-825.
- Henne, A., Schmitz, R. A., Bömeke, M., Gottschalk, G., & Daniel, R. (2000). Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl. Environ. Microbiol.*, 66(7), 3113-3116.
- Sivasubramani, K., SINGH, J. R., & JAYALAKSHMI, D. (2012). ISOLATION OF METAGENOMIC DNA FROM MANGROOVE SOIL AND CONSTRUCTION OF COSMID LIBRARY. *International Journal of Pharma and Bio Sciences*, 3(2).
- Liu, X., Ashforth, E., Ren, B., Song, F., Dai, H., Liu, M., ... & Zhang, L. (2010). Bioprospecting microbial natural product libraries from the marine environment for drug discovery. *The Journal of Antibiotics*, 63(8), 415-422.
- Lim, H. K., Chung, E. J., Kim, J. C., Choi, G. J., Jang, K. S., Chung, Y. R., ... & Lee, S. W. (2005). Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. *Appl. Environ. Microbiol.*, 71(12), 7768-7777.

- Henkel, M., Müller, M. M., Kügler, J. H., Lovaglio, R. B., Contiero, J., Sydatk, C., & Hausmann, R. (2012). Rhamnolipids as biosurfactants from renewable resources: concepts for next-generation rhamnolipid production. *Process Biochemistry*, 47(8), 1207-1219.
- Morikawa, M., Ito, M., & Imanaka, T. (1992). Isolation of a new surfactin producer *Bacillus pumilus* A-1, and cloning and nucleotide sequence of the regulator gene, *psf-1*. *Journal of fermentation and bioengineering*, 74(5), 255-261.
- Jacques, R. J. S., Bento, F. M., Antonioli, Z. I., & Camargo, F. A. D. O. (2007). Biorremediação de solos contaminados com hidrocarbonetos aromáticos policíclicos. *Ciência Rural*, 37(4), 1192-1201.
- Alexander, M. (1999). *Biodegradation and bioremediation*. Gulf Professional Publishing.
- 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).
- Marcos, M. S., Lozada, M., & Dionisi, H. M. (2009). Aromatic hydrocarbon degradation genes from chronically polluted Subantarctic marine sediments. *Letters in applied microbiology*, 49(5), 602-608.

