International Journal Of Public Mental Health And Neurosciences
ISSN: 2394-4668
(An Official publication of Sarvasumana Association)

NGS and Bioinformatics: A Comprehensive Review of Metagenomic Applications

Akshay Wabale*, Chaitali Dhande, Komal Devgirikar and Virendra Sontakke
Institute of Bioscience and Technology, MGM University, Chhatrapati Sambhajinagar, India
*Corresponding Author Email: akshuwabale2001@gmail.com

Abstract: Metagenomics is a field that investigates microbial communities' genetic content directly from environmental samples that has revolutionized our understanding of microbial diversity and function. Compared to traditional methods, advances in metagenomics analysis brought about by next-generation sequencing have expanded our understanding of microbial communities and offered a sophisticated method for identifying new and uncultivable microorganisms based on their genetic information derived from a specific environment. Examining the DNA of the entire community without the need for PCR amplification is known as shotgun metagenomics. It makes it possible to examine every gene in the sample. Conversely, amplicon sequencing focuses on taxonomically significant marker genes, whose study is limited to previously identified DNA sequences. Sequence-based metagenomics analyses DNA sequences straight from the environment without the need to build libraries and can identify a limited number of novel genes and products that can be complemented by functional genomics. In contrast, function-based metagenomics necessitates cloning and fragmenting the extracted metagenome DNA in a suitable host, followed by functional screening and sequencing clones to detect novel genes. Despite metagenomics' advancements, new problems still exist. This review sheds light on developments in metagenomic techniques in conjunction with next-generation sequencing, their current uses, the difficulties they present, and some of the more recent ones. The various forms of metagenomics are also covered in this study, along with developments in bioinformatics tools and their importance for metagenomic dataset processing.

Keywords: Metagenomics, Bioinformatics, Next generation sequencing, Computational tools, Environmental microbiology

I. INTRODUCTION

Metagenomics is a field at the nexus of genomics and ecology which represents a revolutionary approach to studying microbial communities. It is estimated that less than 1% of bacteria on common laboratory media may be identified by microbial community investigation using culture-based techniques [1]. It is impossible to detect the proximity of any microbial community complex in a pure culture setting. The efficacy of genetic and biotechnological research is still hampered by this [2]. These conventional approaches do not take into account the integration and interaction of different microorganisms and the overall climate; thus, they are insufficient to disclose the genetic evolution and natural capacity of the microbial communities in the relationship [3]. Traditional microbiological methods that depends on isolating and culturing individual microorganisms while metagenomics offers a holistic view by directly sequencing and analysing DNA extracted from environmental samples. However, this approach circumvents the limitations of culture-based techniques which often fail to capture the full diversity of microbial life due to the inability to culture certain species under laboratory conditions [4]. By capturing the collective genomic blueprint of entire microbial community's metagenomics provides unprecedented insights into their genetic diversity, functional potential and ecological dynamics. It allows researchers to explore the intricate web of interactions among microorganisms and their environment which are shedding light on fundamental questions about microbial ecology, evolution and biogeography [5]. Conventional microbiological methods are outperformed by Next-Generation Sequencing (NGS) and molecular microbiology technologies, which reveal the vast majority of the microbial world that is invisible. Important strides have been made to make previously uncultured microorganisms more cultivable. When it comes to identifying and extracting gene products from species found in the environment, the culture-independent method is the most promising and economical [6]. The ongoing development of NGS technology is crucial to ushering in a new era in metagenomics [8]. High throughput and rapidly declining DNA sequencing costs are made possible by this innovation's

sequencing effort [3, 9]. In the sequencing endeavour, genome-based disclosure has expanded to a new level for biotechnological applications that address unanswerable biological issues [7]. Another crucial development that contributes to metagenomics' potency is the growth of increasingly sophisticated bioinformatics and computational tools for sequence analysis across all phases of analysis, including pre-processing, assembly, binning, and annotation of sequencing data obtained from NGS technology.

The significance of metagenomics extends across diverse ecosystems, from terrestrial and aquatic environments to the human body's microbiome. In environmental microbiology, metagenomics enables the exploration of microbial communities' roles in nutrient cycling, bioremediation, and ecosystem stability. In human health, it offers insights into the complex interplay between the microbiome and host physiology, with implications for understanding diseases, such as obesity, inflammatory bowel disease, and infectious disorders [10]. However, the sheer volume and complexity of metagenomic data pose significant challenges for analysis and interpretation. This is where bioinformatics emerges as a critical component of metagenomics research. Bioinformatics encompasses a diverse set of computational methods and tools designed to extract meaningful biological insights from raw sequencing data. From pre-processing and quality control to taxonomic classification, functional annotation, and data visualization, bioinformatics provides the analytical framework necessary to decipher the wealth of information encoded in metagenomic sequences [11].

This review provides an overview of recent developments in metagenomic techniques, next-generation sequencing advancements, bioinformatics software and computational tool advancements, metagenomics applications, and a framework of the latest developments in this rapidly evolving field. The purpose of this study is to investigate the developments in NGS instruments and computational tools that enable metagenomics to advance to a mature state. This review's main goal is to provide a better understanding of metagenomics, its fundamental techniques, and its applications, especially in the most current and developing studies. Additionally, this review provides an examination of fundamental methodological and technical challenges as well as possible entanglements arising from metagenomic techniques, including data processing, interpretation, and research design.

II. METAGENOMICS

The study of the metagenome, or collective genome of microorganisms from an environmental sample, is known as metagenomics. Its purpose is to provide details about the ecology and microbial diversity of a particular habitat. There are two main types of metagenomics: amplicon sequencing, which focuses on a key portion of the gene for taxonomy studies, like the internal transcribed spacer (ITS) region, 16s rRNA, and 18s rRNA. Marker gene-based metagenomics is another name for amplicon sequencing. Creating a nonspecific primer from a preserved area of a microbial community is the main requirement for this method. ITS region for fungus, 18s rRNA gene for eukaryotic organisms, and 16s rRNA gene for bacteria and archaea are the most often used marker genes. A low-cost method for profiling the genus makeup of the microbiota is often the 16s rRNA marker gene [12]. The hypervariable region (V1-V9) located in the conserved area of the 16s rRNA gene is amplified using the generic primer designed from the conserved region to provide a species-specific signature [13]. The nucleotide order is ascertained by sequencing the amplified area. A consensus sequence is generated to identify the taxonomy of bacteria by matching in an appropriate reference database after the amplified region is sequenced to ascertain the nucleotide order. The sequence reads are then binned to a specific taxon based on sequence closeness [14]. Shotgun metagenomics, which involves sequencing all of the genes in the microbial community. In shotgun metagenomics, every gene found in a specific environment is analysed. With a moderately low cost, better time requirements, and more information, it provides a coordinated understanding of community structure, composition, genetic and population heterogeneity, functional capacity, and probable metabolic pathways. By directly sequencing extracted genomic DNA without amplifying for the particular gene, it circumvents the bias brought about by PCR [15]. Castañeda & Barbosa (2017) investigated the taxonomic, functional, and metabolic potential of the microbial population in Chilean vineyards and nearby natural woods using shotgun metagenomic sequencing [16]. Sequence-driven metagenomics and functional metagenomics are two more

subcategories of the whole-genome shotgun sequencing technique. The sequence-driven metagenomics approach, which includes gene prediction, assembly, binning, and annotation of recently discovered metagenomic sequences, specifically relies on sequence analysis for the retrieved metagenome. This method necessitates reference sequence data but may concentrate on sequencing a specific gene of interest (marker genes) amplified by PCR [7] or direct sequencing of isolated metagenomic DNA without amplification. Every protein encoded by the metagenome is expressed in functional metagenomics, which requires building libraries from extracted metagenomes and cloning libraries in the proper culturable microorganism as host.

III. NEXT GENERATION SEQUENCING IN METAGENOMICS

Early microbial community investigations were greatly impacted by Sanger sequencing technologies. Since Sanger sequencing, the sequencing yield and length have significantly changed in the modern day. With an average length of 650 bp, Sanger sequencing can currently recover up to 96 sequences per run, which may be sufficient for phylogenetic marker analysis. However, millions of DNA molecules with varying yields and sequence lengths can be sequenced in parallel using inexpensive platforms called Next Generation Sequencing Technologies (NGS), which have beneficial effects in various fields. The process of ascertaining the nucleotide base order of DNA fragments is known as DNA sequencing. It was crucial to the development of molecular biology. From the earliest manual sequencing procedure (Sanger sequencing) to the portable next-generation sequencing used today (Oxford Nanopore), this field has undergone significant progress [17]. Next-generation sequencing is a kind of DNA sequencing technique that determines the sequence by sequencing several tiny DNA fragments in parallel. With next-generation sequencing technology, a single run can generate 10^5–10^7 short sequence reads quickly and at a lower cost. Because NGS can generate a lot of datasets in a single run, it is referred to as "high-through-put technology." Compared to traditional Sanger sequencing, NGS is a potential method for metagenomic research because of its speed, volume of data produced, and lower cost. [18]. In 2005, 454 Roche pyrosequencing was introduced as the first next-generation sequencer. Since then, a number of next-generation sequencing techniques have emerged, including second and third-generation sequencing. This includes the sequencing by synthesis technology that has been used by Solexa/Illumina sequencers since 2006.

To determine which microbes are present and in what proportions, metagenomic NGS (mNGS) entails simply running all of the nucleic acids in a sample which may contain mixed populations of microorganisms and assigning them to respective reference genomes shown in Figure 1. This is an extremely potent new technology that can concurrently identify genetic material from completely distinct kingdoms of creatures by sequencing and identifying nucleic acids from several different taxa for metagenomic study. The potential clinical uses are enormous and include, among many other things, the diagnosis of infectious diseases, the tracking of outbreaks, infection control surveillance, and the identification of mutations and pathogens. mNGS, also known as shotgun sequencing, has been used for a variety of clinical samples, such as blood, respiratory, gastrointestinal, ophthalmic, and cerebrospinal fluid. To targeted polymerase chain reaction (PCR) techniques, which depend on primers to identify particular targets to be amplified and detected, mNGS's greatest strength is its impartiality and lack of hypothesis. Because they amplify unique nucleic acid sequences that can be bioinformatically classified into bacteria/archaea or fungi, respectively, using specific primers of conserved 16S ribosomal RNA (rRNA) gene and internal transcribed spacer (ITS) sequences, even universal or broad-range PCR methods are not broad enough to be classified as metagenomic. universal primers present a challenge for molecular testing used to diagnose polymicrobial illnesses. Using 16S sequencing in the presence of polymicrobial populations will result in several base-calls per nucleotide, creating an unintelligible mixed nucleotide chromatogram. Many laboratories rely on next-generation sequencing of the 16S gene for polymicrobial samples, as opposed to de-convolutional computational approaches, which are available for predicting organisms detected. Next-generation sequencing is extremely dynamic and quick due to ongoing instrument improvements and the subsequent emergence of contemporary breakthroughs eliminating the need for vector cloning by determining the sequence information

from a single DNA fragment of a library. The primary benefit provided by these stages is the development of an enabling platform to directly obtain DNA sequence data from environmental samples and the use of next-generation sequencing technologies to diagnostic virology [19, 20].

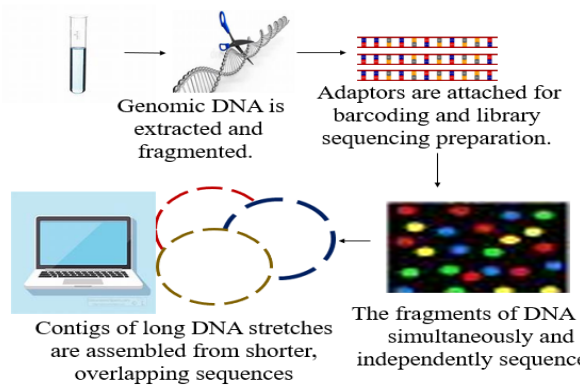

Fig 1. Metagenomic Next Generation Sequencing

The least prevalent microbe in the environment can be found thanks to NGS's great sequencing depth. However, NGS's efficacy is limited by its comparatively short read duration and massive data generation [6]. Oxford Nanopore Technology's (ONT) MinION is the first sequencer that uses nanopores that is sold commercially. By monitoring the shift in electrical conductivity as DNA molecules move through biological pores, ONT can distinguish between different nucleotides [21]. Given that plants have huge, complex genomes with numerous repetitive regions, MinION sequencing is helpful for sequencing their genomes [22]. In the clinical context, Yohe and Thyagarajan [23] describe the technique, applications, and limits of using NGS technology to test for tumor mutations and hereditary illnesses. Using pyrosequencing, Li and his associates investigated the organization of the microbial community in a mesophilic anaerobic digester in a different study [24]. The possibilities and progress of NGS for the examination of fish gut microbiota were further emphasized by Ghanbari and his associates. Reddington and associates used MinION Nanopore sequencing, a portable platform that would be helpful for future worldwide river monitoring, to more recently in 2020 uncover the microbial consortia taxonomic and functional potential of 11 rivers worldwide [25]. Even though next generation sequencing technologies are strong and have helped us explore new settings and uncover new microbial worlds, they also have unique drawbacks and biases that must be overcome. It is crucial to remember that the processing of data derived from second or third generation sequencing technologies requires certain computer resources. More sophisticated bioinformatics analysis and greater processing resources are required as the size of the resulting dataset increases. Large data storage is also required for data processing and archiving [26]. High-end servers are necessary for bioinformatic analysis, but so are knowledge of the UNIX operating system. The existing metagenomics software requires programming and scripting skills to run and install in order to parse and analyse the data. Therefore, in order to benefit from metagenomic data, biologists or biological scientists should learn the fundamentals of computation.

IV. BIOINFORMATICS TOOLS AND TECHNIQUES IN METAGENOMICS

Metagenomics has emerged as a powerful tool for studying microbial communities, regardless of whether the member organisms can be isolated and cultured in a lab. It has also provided a way to characterize the diversity of microorganisms in the environment, since many of them are still uncultivable. Extracting DNA from a community in order to combine the genomes of all the creatures inside it is known as metagenomics. Environmental and community genomics, another name for metagenomics, is the study of microorganisms by directly extracting and copying their DNA from a collection of microorganisms found in the majority of the world's habitats, such as soil or water. An organism's genome is typically extracted from its surroundings, fragmented, and then cloned using its replicating plasmid. After that, the organisms are cultivated to produce metagenomic libraries, which are analysed via DNA sequencing. It is the study of genetic material recovered directly from environmental samples, relies on sophisticated bioinformatics tools and techniques to unravel the complexity of microbial communities. Niu et al. [27] describe how bioinformatics is used in metagenomic analysis, for as when analysing 16S rRNA data. By analysing 16S rRNA data, one can estimate the metabolic pathways of the bacteria in the sample and ascertain the diversity of the samples. MOTHUR serves as an illustration of a diversity analysis tool. The use of 16S rRNA sequencing data to predict a community's metabolic pathway from a sample using the PICRUSt software has been documented by Mallick et al. [28]. Bioinformatics tools are

essential for pre-processing, assembling, binning, annotating, and analysing the vast amount of sequencing data generated from metagenomic studies. Pipelines for metagenomic sequence analysis differ according on the kind of sequence analysis steps.

To eliminate chimeras and low-quality sequences, the raw sequence read that was taken straight from the sequencer needs to be pre-processed. To make the raw sequence read standard for further bioinformatics analysis, this procedure is necessary. Pre-processing tools, FastQC is a popular tool for assessing the quality of raw sequencing reads. It provides detailed reports on various quality metrics, including per-base sequence quality, GC content, and sequence duplication levels [29]. Trimmomatic is a versatile tool for pre-processing sequencing data by removing adapter sequences, low-quality bases, and PCR artifacts. It employs various filtering and trimming algorithms to improve the overall quality of sequencing reads [30]. Seqtk is another preprocessing tool used for sequence quality control, format conversion, and sub-sampling of large sequencing datasets. It offers efficient algorithms for filtering low-quality reads and removing sequencing artefacts [31].

The reconstruction of short metagenome reads linked to create a lengthy sequence is called assembly. The process of joining short sequence reads into lengthy contigs or scaffolds and ultimately rebuilding entire genomes is known as genome assembly [32]. In order to accurately estimate the microbial population in a given environment, genome assembly makes it easier to predict full-length protein-coding genes or transcripts and identify strain-specific genomic islands. A full genome or a good draft may be produced using the assembly method [33]. Reference-based and de novo assembly are the two methods of genome assembly. In order to align short reads with the reference database, reference-based assembly depends on the availability of a reference genome. Due to reference databases' preference for previously identified sequences, this approach significantly understates microbial diversity. When assembling a sequence that has never been assembled before, a de novo assembly is the optimum choice. However, the cost of this approach is significant, and it requires more gigabytes of Random-Access Memory. Assembly makes use of one of two often employed techniques: the de Bruijn graph and overlap layout consensus (OLC) [12]. Other scholars have also created hybrid and iterative joining techniques for assembly [34, 35]. The de Bruijn graph, however, is the most widely used technique. The de Bruijn graph has the advantage of being less expensive than OLC due to its ability to be constructed without pairwise comparisons [12]. The bioinformatic tools BBAP, Genovo, MegaGT, and MEGAHIT can be used for assembly [36]. Assembly tools MEGAHIT is a state-of-the-art metagenomic assembler optimized for reconstructing high-quality contigs from short-read sequencing data. It employs an efficient de Bruijn graph algorithm to assemble complex metagenomic datasets [37]. MetaSPAdes is a metagenomic assembler designed to handle heterogeneous datasets containing multiple species with varying abundances. It integrates information from multiple sequencing libraries to improve assembly accuracy and contiguity [38].

The clustering of sequences created during the assembly process is known as binning. Binning represents a biological taxon by classifying sequences known as contigs [12]. MetaWatt [39] and CONCOCT [40] are two examples of software alternatives utilized for binning analysis. Higher accuracy compared to current approaches and ease of use are two benefits of MetaWatt [39]. CONCOCT is a highly accurate software that can classify complex microbial communities [40]. A technique for identifying pieces of the same biological sequence is sequence analysis. Simple alignment and multiple alignments are the two categories into which sequence analysis is separated. Aligning two sequences together is known as simple alignment, but aligning more than two sequences is known as multiple alignment [41]. The Basic Local Alignment Tool, or BLAST, is one of the alignment tools. A tool called BLAST is used to compare the sequences of different kinds of organisms. The expectation value (E value), a gauge of statistical significance, is assigned to each alignment's score [42].

Annotation is the process of adding and allocating biological information to the sequence of short reads and assembled contigs in order to make sense of them. Various bioinformatics tools designed for this purpose are used to accomplish this. The unassembled read's taxonomic and functional annotation primarily depends on the database. For short reads, some annotation tools are USEARCH,

BLAT, RapSearch, and DIAMOND; for assembled reads, they are IMG/ER, MG RAST Server, WebMGA, and ggKbase. Gene prediction tools for metagenomic sequence reads include Prodigal, MetaGeneAn-notator, FragGeneScan, Orphelia, and MetaGeneMark [43]. Taxonomic classification tools QIIME (Quantitative Insights Into Microbial Ecology) is a comprehensive bioinformatics pipeline for analysing microbial community composition and diversity from marker gene sequencing data, such as 16S rRNA. It provides modules for sequence processing, taxonomic assignment, and statistical analysis [44]. Kraken is a fast and accurate taxonomic classification system that assigns taxonomic labels to sequencing reads using k-mer-based alignment algorithms. It leverages reference databases to classify millions of reads within minutes [45]. Functional annotation tools MG-RAST (Metagenomics Rapid Annotation using Subsystem Technology) is a web-based platform for annotating metagenomic sequences with functional information. It utilizes reference databases such as SEED and KEGG to assign putative functions to genes and pathways [46]. HUMAnN (HMP Unified Metabolic Analysis Network) is a pipeline for functional profiling of microbial communities using metagenomic sequencing data. It quantifies the abundance of microbial metabolic pathways and gene families, providing insights into community-level metabolic activities [47]. Comparative analysis tools STAMP (Statistical Analysis of Metagenomic Profiles) is a software package for statistical analysis and visualization of metagenomic data. It offers a wide range of statistical tests and graphical tools for comparing microbial community compositions across different experimental conditions [48]. LEfSe (Linear discriminant analysis Effect Size) is a computational tool for identifying differentially abundant microbial taxa or functional features between multiple sample groups. It employs linear discriminant analysis (LDA) to detect biomarkers associated with specific biological conditions [49].

Protein families are listed in the Pfam database. In Pfam's analysis, the double alignment generated by the hidden Markov model is referred to. Investigating the relationships between protein sequences at the family level is the aim of Pfam's investigation [42]. Amino acid sequence is referred to as the fundamental structure. The genes that encode the fundamental structure are sequenced in a certain way. Protein structures are divided into three categories: secondary, tertiary, and quaternary. Understanding protein structure is essential to comprehending how proteins work. Bioinformatics prediction study of protein structure can aid in comprehending a protein's physical properties and functions [50]. Phylogenetic analysis of functional metagenomics describes methods for predicting specific molecular properties and reconstructing the evolutionary links between groups of protein molecules. The likelihood, parsimony, and distance approaches are the techniques used to create phylogenetic trees. Every approach has unique advantages and disadvantages, and no one approach is flawless. MOLPHY, PHYLIP, and MEGA (Molecular Evolutionary Genetics Analysis) are a few examples of phylogenetic analysis tools [51].

V. APPLICATION OF METAGENOMICS

Bioinformatics serves as the backbone of metagenomics, enabling researchers to decode the genetic information of complex microbial communities and extract valuable insights across various fields. Illustrating the profound impact of bioinformatics in metagenomic applications:

5.1. Environmental Microbiology

In a ground-breaking study by Xiong et al. (2015), metagenomic sequencing coupled with sophisticated bioinformatics analysis was employed to explore microbial communities thriving in oil-contaminated soil [52]. By leveraging tools like QIIME for taxonomic classification and MG-RAST for functional annotation, researchers identified key microbial taxa and metabolic pathways involved in hydrocarbon degradation. This not only provided insights into the microbial ecology of polluted environments but also offered valuable strategies for bioremediation efforts. Both natural and man-made elements contaminate the environment, but in recent years, pollutants brought on by the latter such as different waste products from industry and other household activities have gotten worse. Population increase and industrialization are the primary causes of environmental pollution. To preserve the natural ecosystem and create a safer environment for human health, pollutants must be removed from contaminated soil and wastewater. Recently, there has been an increase in interest in bioremediation-based techniques for the economical and sustainable removal, degradation, and adsorption of

contaminants [53]. Pollutants such as hydrocarbons, polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons (PAHs), and other aromatic chemicals can be broken down and changed by microbes [54]. An improved technique for identifying putative genes involved in the removal of metal pollutants and the breakdown of aromatic hydrocarbons is the use of metagenomics methods. A comprehensive understanding of microbial communities is made possible by recent developments in next-generation sequencing and bioinformatics, which also offer an invisible mechanism for the breakdown of contaminants. Finding new microbes that are essential to the breakdown of pollutants and identifying new genes that enable microbial remediation from contaminated environmental samples is made easier by metagenomics, all without the need to separate and cultivate the microorganisms. Furthermore, metagenomics identifies many methods and approaches to improve clean-up [55]. The fungus community in the soil sample affected by hydrocarbons was identified using a metagenomic technique in the recent study by Ogbonna and colleagues [56]. Chakraborty and associates use shotgun nanopore sequencing to reveal the functional analysis and spatiotemporal taxonomic diversity of hypersaline and hyperalkaline lonar lakes [57]. For benzoate degradation routes (KO 00362) and nitrotoluene degradation pathways (KO 00633) classified up to function level, they examined xenobiotic degrading genes and metabolic pathways. Another study looked at how the addition of nitrogen, which promotes the breakdown of total petroleum hydrocarbon in various treatments, altered the composition and organization of the microbial community [55]. Using the KEGG orthologous group (KOs), they attempted to determine the primary functional enzymes involved in the metabolism of carbon and nitrogen. Another study also reported the bioremediation capacity of the Metagenome-Assembled Genome (MAGs) of Oceancaulis Alexandarii NP7 isolated from the Mediterranean Sea-polluted marine sediments in terms of hydrocarbon breakdown and metal detoxification [58].

5.2. Human Health

Metagenomic analysis of the human gut microbiome has revolutionized our understanding of its role in health and disease. Qin et al. (2010) conducted a seminal study wherein metagenomic sequencing and bioinformatics tools like MetaPhlAn and HUMAnN were employed to compare gut microbiota in individuals with and without type 2 diabetes [59]. The analysis revealed significant differences in microbial composition and metabolic pathways associated with disease. Such findings hold promise for developing personalized therapeutic interventions targeting the gut microbiome. Novel viral compositions can be found using metagenomics techniques in nasopharyngeal swabs taken from patients infected with COVID-19 [60]. Understanding the makeup of the gastrointestinal microbiota in individuals infected with SARS-CoV-2 is the other significant application. It is established that the composition of the intestinal microbiome plays a major role in both human health and disease. There could be major health consequences if the microbiota in our intestine's changes or declines. The gut microbiome's discovery and characterisation offer crucial information about people's health. Gestural symptoms may be experienced by patients with a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection [61]. A potent method for identifying and detecting the gastrointestinal microbiome composition of individuals infected with SARS-CoV-2 is advanced high-throughput metagenomics sequencing [62]. In contrast, the metagenomic technique greatly aids and facilitates the analysis of the microbial population from patient samples to detect the presence of harmful bacteria and genes responsible for toxicity and pathogenicity [63]. Identifying microorganisms in the examined samples, including those from infectious diseases, is possible through the use of a metagenomic method, which is not specific to a single isolate [64, 65]. Without the need for isolation in pure culture, metagenomics-based high-throughput sequencing makes it possible to identify possible pathogens including bacteria, viruses, fungus, and parasites in patient samples. NGS is appropriate for analysing the complete genomic contents of clinical samples because it produces millions to billions of readings in a single run. In the diagnosis of infectious diseases, clinical metagenomics is used to define the taxonomy of the pathogenic bacterium, predict antibiotic resistance genes in patient samples, identify virulence factors, and discover antiviral resistance genes. Additionally, this data is utilized for oncology, transcriptomics to study the human host's response to infection, and microbiome studies. Potential applications of viral metagenomics include

environmental monitoring, disease outbreak response, clinical diagnostics, and the identification of new viral pathogens [66]. Shotgun metagenomics is superior to other conventional microbiology and molecular techniques for pathogen detection because it is faster, less expensive, can identify low-level pathogens, can identify novel or variant pathogens, can detect transmission events, and can detect multiple pathogens in a single test [67].

5.3. Microbial Ecology and Evolution

Metagenomic data coupled with advanced bioinformatics tools have provided unprecedented insights into microbial ecology and evolution. Locey & Lennon (2016) utilized metagenomic datasets to explore microbial diversity patterns across diverse ecosystems [68]. By applying ecological models and network analysis techniques, researchers deciphered fundamental principles governing community assembly, biodiversity dynamics, and ecosystem functioning. Such studies contribute to our understanding of microbial ecosystems and their response to environmental changes. The external environment is unforgiving and uncomfortable for microorganisms to thrive and reproduce. Extreme salinity and nutrients, high temperatures, changed gravity, and intense radiation are the characteristics of those environmental circumstances. Microbial communities persist in those settings despite the adverse effects those conditions have on microbial life. Extremophiles are microorganisms that can endure in those environments. Investigating extremophiles on Earth helps discover new biomarkers that could be applied in liveable areas beyond Earth and is crucial for comprehending their adaptation tactics [69]. In order to investigate if life exists on other extra-terrestrial planets, these liveable habitats are regarded as analog environments on Earth. 16S Amplicon Sequencing and shotgun metagenomics were used to assess the taxonomic and functional diversity of microbial communities in Costa Rica's Poás Volcano, which is perhaps one of the planet's most harsh Martian analog settings [70]. Utilizing culture-dependent approaches is not feasible for studying microbial communities from those similar habitats. Without culturing, we can thus directly access the genetic material of the extremophilic bacteria present in the sample through metagenomics, meta-transcriptomics, and metabolomics research. The metagenomics technique does not necessitate sustaining the harsh environmental conditions and nutrient needs of every microorganism in lab settings [71]. A living cell's diverse range of chemical reactions and small molecules cannot be sufficiently explained by methods based on DNA and RNA. Extremophiles and environmental samples include tiny compounds that can be accurately identified by metabolomics [72]. These tiny chemicals are the primary means by which microorganisms survive in hostile environments. In the meanwhile, those tiny molecules are crucial for studying extra-terrestrial life. To profile microbial communities in sulfidic springs, subterranean, hypersaline, acidic lakes, acidic rivers, and permafrost settings, amplicon sequencing and whole-genome shotgun sequencing were employed [73]. All of these include the developments in metagenomic approaches that have made it possible to get insight into the intricate and multifaceted nature of life in a more effective manner. A study that used metagenomic analyses to provide a comprehensive view of the surface microbiome of the International Space Station (ISS) has shown a microbial burden, the prevalence of bacterial and fungal species, changes in the microbiome and resistome over time and space, as well as the functional capabilities and microbial interactions of this specially constructed microbiome [74]. In order to guarantee an ISS surface microbiota that supports astronaut health and spacecraft integrity, these investigations help shape regulations for next space missions.

5.4. Biotechnological Applications

Metagenomics has emerged as a powerful tool for bioprospecting and discovering novel enzymes with biotechnological potential. Ferrer et al. (2005) employed metagenomic sequencing and bioinformatics analysis to identify thermostable enzymes from hot springs [75]. Protein molecules called enzymes reduce the activation energy of biological reactions, speeding them up. Because of their sensitivity to the substrate and specificity, enzymes were chosen by many businesses. Enzymes are now widely used in sectors like textiles, paper, detergents, food, and beverages as a result of the development of technology associated to the manufacturing of enzyme-linked goods. Higher activity, heat tolerance, and pH tolerance are some of these enzymes' unique characteristics that help them meet the demands of particular industries [63]. The requirement for new enzymes in various sectors has been satisfied by metagenomics, which has

yielded enzymes including lipases, β-lactamases, polysaccharide modifying enzymes, nitrilases, dehydrogenases, oxidoreductases, cellulose [76], amylase, xylanase, and proteases. The process of cloning and screening the intended clone according to the function expressed by the host is necessary to identify a new biocatalyst [7]. Mirete and his associates summed up the significance of employing functional metagenomics to identify extremophiles—organisms that are unable to thrive under typical laboratory conditions—and their extremozymes [2]. By leveraging bioinformatics tools for sequence annotation and functional prediction, researchers isolated enzymes with industrial applications in bio catalysis and bioremediation. This illustrates the transformative potential of metagenomics in biotechnology and industrial processes. Using metagenomics techniques, new pathways, enzymes, and microbes involved in the breakdown of cellulosic biomass were discovered [77]. Motahar and his associates discovered, isolated, and described new acidic and/or thermostable α-amylase (persiAmy3) from the microbiota of sheep rumen. By raising the overall lowering sugar content, PersiAmy3 is utilized in chicken feed enhancement to improve starch breakdown and boost feed nutrient availability [78]. A more recent study looked at the functional characteristics and microbial composition of the cecum of Chinse bamboo rats [79]. They discovered a number of metabolic pathways for the genes and genomes involved in the breakdown of cellulosic and lignocellulosic biomass. Another study identified more carbohydrate-active genes and novel thermostable cellulose-degrading enzymes from anaerobic digestion sludge [80].

Bioinformatics and next generation sequencing, coupled with metagenomics, has opened new avenues for exploring microbial diversity, understanding ecosystem dynamics, and harnessing the biotechnological potential of microorganisms. Through interdisciplinary collaborations and technological innovations, bioinformatics continues to drive ground-breaking discoveries in environmental microbiology, human health, agriculture, microbial ecology, and biotechnology, shaping our understanding of microbial life and its myriad applications in diverse fields.

VI. CHALLENGES IN METAGENOMICS

Metagenomic data analysis, while incredibly powerful, is not without its hurdles. The complexity of microbial communities and the inherent limitations of sequencing technologies pose significant challenges. Although metagenomics has been shown to have significant applications across several industries, there are still many issues that must be resolved for the field to profit effectively [3]. The contamination of extracted DNA with soil-derived extracts, such as humic acid, is one of the main and most important problems in metagenomics. While the absence of necessary purification disrupts downstream processes, additional purification to eliminate these contaminants results in sample loss [81,82]. To produce the library, adequate length and high-quality DNA are needed; meeting these requirements, as well as the sensitivity of packing concentrates and of cleaned, digested, and dephosphorylated vector DNA for ligation, continues to be a challenge [81]. Another issue with amplicon-based metagenomics in the future is the limited resolution for novel microorganisms and their propensity for unusual amplification [3]. For metagenomics research, assembling the numerous short reads generated by the current sequencing method into continuous fragments (contigs) presents significant hurdles as well. When numerous lengthy repeat patterns are displayed within the genes, the assembly process becomes more difficult [3]. By advancing high-throughput gene sequencing technologies, these difficulties can be overcome and longer reads can be sequenced with more precision and depth. Data storage and processing power are the two major issues brought on by the massive volume of data produced by NGS [83]. Metagenomics examination is challenging for beginners due to the lack of advanced bioinformatics understanding and the difficulty to view all NGS data in an interpretable format [53]. Another drawback in metagenomics is the bias of the reference database that is used to annotate sequence reads. Sequences from readily available, efficiently culturable organisms and species with industrial or pathogenic implications are preferred in certain databases [11]. The majority of binning techniques are tailored to binning reads that come from bacterial genomes [11]. Metagenomic investigation is made simpler by carrying out more metagenomic research to build larger genomic databases [3].

VII. CONCLUSION

This review has provided a comprehensive overview of the role of bioinformatics and next generation sequencing in advancing metagenomics research. Through a synthesis of literature and case studies, several key findings and insights have been highlighted. Bioinformatics and next generation sequencing serve as a cornerstone in metagenomic studies, facilitating data analysis, interpretation, and integration across diverse applications such as environmental microbiology, human health and biotechnology. Through the use of advanced computational methods, algorithms, and software tools, researchers can extract meaningful biological insights from complex metagenomic datasets, uncovering the taxonomic composition, functional potential, and ecological dynamics of microbial communities. Bioinformatics has revolutionized our understanding of microbial ecology, evolution, and biotechnological applications. By enabling the exploration of microbial diversity, interactions, and metabolic capabilities, bioinformatics has paved the way for transformative discoveries in fields ranging from environmental remediation and disease diagnostics to bioprospecting and synthetic biology. The importance of bioinformatics and next generation sequencing in metagenomics cannot be overstated. As technological advancements continue to generate vast amounts of sequencing data, bioinformatics plays a crucial role in processing, analysing, and interpreting this information, thereby bridging the gap between raw data and biological insights. Without robust bioinformatics approaches, the full potential of metagenomics to revolutionize our understanding of microbial life and its applications would remain unrealized.

The future prospects of bioinformatics and next generation sequencing in metagenomic studies are promising. Emerging trends such as multi-omics integration, single-cell analysis, machine learning, spatial profiling, and environmental monitoring are poised to further expand the frontiers of metagenomics research. By embracing these advancements and fostering interdisciplinary collaborations, researchers can unlock new dimensions of microbial complexity, diversity, and functionality, leading to innovative solutions for addressing global challenges in health, agriculture, environment, and biotechnology. Bioinformatics and next generation sequencing is indispensable in advancing metagenomics research, providing the analytical framework and computational tools necessary for unravelling the mysteries of microbial communities. As we embark on this exciting journey into the microbial world, bioinformatics will continue to be a guiding light, illuminating our path towards a deeper understanding of microbial life and its implications for humanity.

REFERENCES

1. Coughlan LM, Cotter PD, Hill C, Alvarez-Ordóñez A (2015) Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. Front Microbiol 6:1–22.
2. Mirete S, Morgante V, González-Pastor JE (2016) Functional metagenomics of extreme environments. Curr Opin Biotechnol 38:143–149. https:// doi. org/ 10. 1016/j. copbio. 2016. 01. 017
3. Behzad H, Gojobori T, Mineta K (2015) Challenges and opportu-nities of airborne metagenomics. Genome Biol Evol 7(5):1216–1226. https:// doi. org/ 10. 1093/ gbe/ evv064
4. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chemistry & Biology, 5(10), R245-R249.
5. Riesenfeld, C. S., Schloss, P. D., & Handelsman, J. (2004). Metagenomics: genomic analysis of microbial communities. Annual Review of Genetics, 38, 525-552.
6. Degois J, Clerc F, Simon X, Bontemps C, Leblond P, Duquenne P (2017) First metagenomic survey of the microbial diversity in bioaerosols emitted in waste sorting plants. Ann Work Expo Heal 61(9):1076–1086.
7. dos Santos DFK, Istvan P, Quirino BF, Kruger RH (2017) Functional metagenomics as a tool for identification of new antibi-resistance genes from natural environments. Microb Ecol 73(2):479–491.
8. Scholz MB, Lo CC, Chain PSG (2012) Next generation sequencing and bioinformatic bottlenecks: The current state of metagenomic data analysis. Curr Opin Biotechnol 23(1):9–15.
9. Ju F, Zhang T (2015) experimental design and bioinformatics analysis for

the application of metagenomics in environmental sciences and biotechnology. Environ Sci Technol 49(21):12628–12640.
10. Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project: exploring the microbial part of ourselves in a changing world. Nature, 449(7164), 804-810.
11. Mande, S. S., Mohammed, M. H., & Ghosh, T. S. (2012). Classification of metagenomic sequences: methods and challenges. Briefings in Bioinformatics, 13(6), 669-681.
12. Roumpeka DD, Wallace RJ, Escalettes F, Fotheringham I, Wat-son M (2017) A review of bioinformatics tools for bio-prospecting from metagenomic sequence data. Front Genet.
13. Bikel S et.al (2015) Combining metagenomics, metatranscrip-tomics and viromics to explore novel microbial interactions: Towards a systems-level understanding of human microbiome. Comput Struct Biotechnol J 13:390–401.
14. Chiang AD, Dekker JP (2020) From the pipeline to the bedside: Advances and challenges in clinical metagenomics. J Infect Dis 221(3):331–340.
15. Charles TC, Liles MR, Sessitsch A (2017) Functional metagenomics: Tools and applications. Springer link.
16. Castañeda LE, Barbosa O (2017) Metagenomic analysis exploring taxonomic and functional diversity of soil microbial com-munities in Chilean vineyards and surrounding native forests. PeerJ 3:1–19.
17. Zhou XG, Ren LF, Li YT, Zhang M, Yu YD, Yu J (2010) The next-generation sequencing technology: A technology review and future perspective. Sci China Life Sci 53(1):44–57.
18. Ghanbari M, Kneifel W, Domig KJ (2015) A new view of the fish gut microbiome: Advances from next-generation sequencing. Aquaculture 448:464–475.
19. Barzon L, Lavezzo E, Militello V, Toppo S, Palù G (2011) Applications of next-generation sequencing technologies to diagnostic virology. Int J Mol Sci 12(11):7861–7884.
20. Fadiji AE, Babalola OO (2020) Metagenomics methods for the study of plant-associated microbial communities: A review. J Microbiol Methods 170(12):105860.
21. Sevim V et.al (2019) Shotgun metagenome data of a defined mock community using Oxford Nanopore PacBio and Illumina technologies. Sci Data 6(1):1–9.
22. Dumschott K, Schmidt MHW, Chawla HS, Snowdon R, Usadel B (2020) Oxford Nanopore sequencing: new opportunities for plant genomics? J Exp Bot 71(18):5313–5322.
23. Yohe S, Thyagarajan B (2017) Review of clinical next-generation sequencing. Arch Pathol Lab Med 141(11):1544–1557.
24. Li L, He Q, Ma Y, Wang X, Peng X (2016) A mesophilic anaerobic digester for treating food waste: Process stability and micro-bial community analysis usingpyrosequencing. Microb Cell Fact 15(1):1–11.
25. Reddington K et.al (2020) Metagenomic analysis of planktonic riverine microbial consortia using nanopore sequencing reveals insight into river microbe taxonomy and function. Gigascience 9(6):1–12.
26. Logares, R., Haverkamp, T. H. A., Kumar, S., Lanzén, A., Nederbragt, A. J., Quince, C., et al. (2012). Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches. J. Microbiol. Methods 91, 106–113.
27. Niu S, Yang J, Mcdermaid A, Zhao J, Kang Y (2018) Bioinformatics tools for quantitative and functional metagenome and metatranscriptome data analysis in microbes. Brief Bioinform 19:1415–1429.
28. Mallick H, Franzosa EA, Mclver LJ, Banerjee S, Sirota-Madi A, Kostic AD, Clish CB, Vlamakis XRJ, Huttenhower C (2019) Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. Nat Commun 10:1–11.
29. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at:

http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
30. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114-2120.
31. Li, H. (2012). Seqtk: Toolkit for processing sequences in FASTA/Q formats. Available online at: https://github.com/lh3/seqtk.
32. Ghurye JS, Cepeda-Espinoza V, Pop M (2016) Metagenomic assembly: Overview, challenges and applications. Yale J Biol Med 89(3):353–362.
33. Teeling H, Glöckner FO (2012) Current opportunities and challenges in microbial metagenome analysis-A bioinformatic perspective. Brief Bioinform 13(6):728–742.
34. Gregor I, Scho A, Mchardy AC (2016) Snowball: strain aware gene assembly of metagenomes. Bioinformatics 32:i649–i657
35. Ruby JG, Bellare P, & DeRisi JL (2013) PRICE: Software for the targeted assembly of components of (Meta) genomic sequence data. G3 (Bethesda) 3: 865–880.
36. Ayling M, Clark MD, Leggett RM (2019) New approaches for metagenome assembly with short reads. Brief Bioinform 00:1–11.
37. Li, D., et al. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics, 31(10), 1674-1676.
38. Nurk, S., et al. (2017). MetaSPAdes: a new versatile metagenomic assembler. Genome Research, 27(5), 824-834.
39. Strous M, Kraft B, Bisdorf R, Tegetmeyer HE (2012) The binning of metagenomic contigs for microbial physiology of mixed cultures. Front Microbiol 3:1–11.
40. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C (2014) Binning metagenomic contigs by coverage and composition. Nat Methods 11:1144–1146.
41. Diniz WJS, Canduri F (2017) Bioinformatics: an overview and its applications. Genet Mol Res 16:1–21.
42. Khandelwal I, Sharma A, Agrawal PK, Shrivastava R (2017) Bioinformatics Database Resources. In: Library and information services for bioinformatics education and research. IGI Global, Hershey, pp 45–90.
43. Zhu W, Lomsadze A, Borodovsky M (2010) Abinitio gene identification in metagenomic sequences. Nucleic Acids Res 38(12):1–15.
44. Caporaso, J. G., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nature Methods, 7(5), 335-336.
45. Wood, D. E., et al. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biology, 15(3), R46.
46. Meyer, F., et al. (2008). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics, 9(1), 386.
47. Abubucker, S., et al. (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS Computational Biology, 8(6), e1002358.
48. Parks, D. H., et al. (2014). STAMP: Statistical analysis of taxonomic and functional profiles. Bioinformatics, 30(21), 3123-3124.
49. Segata, N., et al. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. Nature Methods, 9(8), 811-814.
50. Raza K (2012) Application of data mining in bioinformatics. Indian J Comp Sci Eng 1:114–118
51. Mehmood MA, Sehar U, Ahmad N (2014) Use of bioinformatics tools in different spheres of life sciences. Data Mining Genom Proteomics 5:1–13.
52. Xiong, J., et al. (2015). Soil fungi possess the potential for lignocellulosic biomass degradation and biofuel production. Critical Reviews in Biotechnology, 35(2), 164-174.
53. Sharma P, Kumar S, Pandey A (2021) Bioremediated techniques for remediation of metal pollutants using metagenomics approaches: A review. J Environ Chem Eng 9(4):105684.

54. Bekele GK et.al (2022) Isolation and characterization of diesel degrading bacteria from hydrocarbon-contaminated sites, flower farms, and soda lakes. Int J Microbiol.
55. Ogbonna DN, Udotong IR, Basi N, Nrior RR (2021) Culture-dependent and metagenomic profiling of Eukaryal diversity in petroleum hydrocarbon-polluted soil from B-Dere, Gokana, Rivers state Nigeria. J Adv Microbiol 21(12):64–73.
56. Gao Y et.al (2021) Metagenomics analysis identifies nitrogen metabolic pathway in bioremediation of diesel contaminated soil. Chemosphere 271:129566.
57. Chakraborty J, Rajput V, Sapkale V, Kamble S, Dharne M (2021) Spatio-temporal resolution of taxonomic and functional microbiome of Lonar soda lake of India reveals metabolic potential for bioremediation. Chemosphere 264:128574.
58. Dell'Anno F et.al (2021) Metagenome-assembled genome (MAG) of Oceancaulis alexandrii NP7 isolated from Mediterranean Sea polluted marine sediments and its bioremediation potential. Genes, Genomes, Genet 11(9):1–4.
59. Qin, J., et al. (2010). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature, 490(7418), 55-60.
60. Rodriguez C et.al (2021) Viral genomic, metagenomic and human transcriptomic characterization and prediction of the clinical forms of COVID-19. PLoS Pathog 17(3):1–21.
61. Liu Y et.al (2021) (2021) Distinct metagenomic signatures in the SARS-CoV-2 Infection". Front Cell Infect Microbiol 11(12):1–12.
62. Sehli S et.al (2022) Metagenomics approaches to investigate the gut microbiome of COVID-19 patients. Bioinform Biol Insights.
63. Datta S, Rajnish KN, Samuel MS, Pugazlendhi A, Selvarajan E (2020) Metagenomic applications in microbial diversity, bioremediation, pollution monitoring, enzyme and drug discovery: A review. Environ Chem Lett 18(4):1229–1241.
64. Berry MI et.al (2022) Next generation sequencing and bioinformatics methodologies for infectious disease research and public health: approaches, applications, and considerations for development of laboratory capacity. J Infect Dis 221:S292–S307.
65. Chiang AD, Dekker JP (2020) From the pipeline to the bedside: Advances and challenges in clinical metagenomics. J Infect Dis 221(3):S331–S340.
66. Bibby K (2013) Metagenomic identification of viral pathogens. Trends Biotechnol 31(5):275–279.
67. Miller RR, Montoya V, Gardy JL, Patrick DM, Tang P (2013) Metagenomics for pathogen detection in public health. Genome Med.
68. Locey, K. J., & Lennon, J. T. (2016). Scaling laws predict global microbial diversity. Proceedings of the National Academy of Sciences, 113(21), 5970-5975.
69. Ehrenfreund P et.al (2011). Astrobiology and habitability studies in preparation for future Mars missions: Trends from investigating minerals, organics and biota. Int J Astrobiol 10(3):239–253.
70. Wang JL, Dragone NB, Avard G, Hynek BM (2022). Microbial survival in an extreme martian analog ecosystem: Poás volcano Costa Rica. Front Astron Sp 9(1):1–16.
71. Thombre RS, Vaishampayan PA, Gomez F (2020). Applications of extremophiles in astrobiology. Inc.
72. Seyler L et.al (2020). Metabolomics as an emerging tool in the search for astrobiologically relevant biomarkers. Astrobiology 20(10):1251–1261.
73. Bashir AK et.al (2021). Taxonomic and functional analyses of intact microbial communities thriving in extreme, astrobiology relevant, anoxic sites. Microbiome 9(1):50.
74. Urbaniak C et.al (2022) Microbial Tracking-2, a metagenomics analysis of bacteria and fungi onboard the International Space Station. Microbiome 10(1):1.
75. Ferrer, M., et al. (2005). Metagenomics: bacterial diversity and enzyme properties of sediment and water from a shallow submarine hydrothermal vent in Milos

Island, Greece. Journal of Biotechnology, 127(3), 165-177.

76. Fan MZ, Wang W, Cheng L, Chen J, Fan W, Wang M (2021). Metagenomic discovery and characterization of multi-functional and monomodular processive endoglucanases as biocatalysts. Appl Sci 11(11):1–36.
77. Thompson CE et.al (2013) A potential source for cellulolytic enzyme discovery and environmental aspects revealed through metagenomics of Brazilian mangroves. AMB Express 3:1–35.
78. Motahar SFS et.al (2020) A novel metagenome-derived thermostable and poultry feed compatible α-amylase with enhanced biodegradation properties. Int J Biol Macromol 164:2124–2133.
79. Bai DP et al (2012) Metagenomics approach to identify lignocellulose-degrading enzymes in the gut microbiota of the Chinese bamboo rat cecum. Electron J Biotechnol 50:29–36.
80. Xia Y, Ju F, Fang HHP, Zhang T (2013) Mining of novel thermostable cellulolytic genes from a thermophilic cellulose-degrading consortium by metagenomics. PLoS ONE 8(1):e53779.
81. Lam KN, Cheng J, Engel K, Neufeld JD, Charles TC (2015) Current and future resources for functional metagenomics. Front Microbiol 6:1–8.
82. Singh BK, Macdonald CA (2010) Drug discovery from uncultivable microorganisms. Drug Discov Today 15(17–18):792–799.
83. Laudadio I, Fulci V, Stronati L, Carissimi C (2019) Next-gener-ation metagenomics: methodological challenges and opportunities. Omi A J Integr Biol 23(7):327–333