Int J Biochem Mol Biol 2023;14(4):62-75
www.ijbmb.org /ISSN:2152-4114/IJBMB0152493

Review Article

Assessing the impact of meta-genomic tools on current cutting-edge genome engineering and technology

Tuward J Dweh*, Subhashree Pattnaik, Jyoti Prakash Sahoo*

*Department of Agriculture and Allied Sciences, C.V. Raman Global University, Bhubaneswar 752054, Odisha, India. *Equal contributors.*

Received July 29, 2023; Accepted August 15, 2023; Epub August 15, 2023; Published August 30, 2023

Abstract: Metagenomics is defined as the study of the genome of the total microbiota found in nature and is often referred to as microbial environmental genomics because it entails the examination of a group of genetic components (genomes) from a diverse community of organisms in a particular setting. It is a sub-branch of omics technology that encompasses Deoxyribonucleic Acid (DNA), Ribonucleic acid (DNA), proteins, and various components associated with comprehensive analysis of all aspects of biological molecules in a system-wide manner. Clustered regularly interspaced palindromic repeats and its endonuclease, CRISPR-associated protein which forms a complex called CRISPR-cas9 technology, though it is a different technique used to make precise changes to the genome of an organism, it can be used in conjunction with metagenomic approaches to give a better, rapid, and more accurate description of genomes and sequence reads. There have been ongoing improvements in sequencing that have deepened our understanding of microbial genomes forever. From the time when only a small amount of gene could be sequenced using traditional methods (e.g., "the plus and minus" method developed by Allan and Sanger and the "chemical cleavage" method that is known for its use in the sequencing the phiX174 bacteriophage genome via radio-labeled DNA polymerase-primer in a polymerization reaction aided by polyacrylamide gel) to the era of total genomes sequencing which includes "sequencing-by-ligation" and the "sequencing-by-synthesis" that detects hydrogen ions when new DNA is synthesized (Second Generation) and then Next Generation Sequencing technologies (NGS). With these technologies, the Human Genome Project (HGP) was made possible. The study looks at recent advancements in metagenomics in plants and animals by examining findings from randomly selected research papers. All selected case studies examined the functional and taxonomical analysis of different microbial communities using high-throughput sequencing to generate different sequence reads. In animals, five studies indicated how Zebrafish, Livestock, Poultry, cattle, niches, and the human microbiome were exploited using environmental samples, such as soil and water, to identify microbial communities and their functions. It has also been used to study the microbiome of humans and other organisms, including gut microbiomes. Recent studies demonstrated how these technologies have allowed for faster and more accurate identification of pathogens, leading to improved disease diagnostics. They have also enabled the development of personalized medicine by allowing for the identification of genetic variations that can impact drug efficacy and toxicity. Continued advancements in sequencing techniques and the refinement of CRISPR-Cas9 tools offer even greater potential for transformative breakthroughs in scientific research and applications. On the other hand, metagenomic data are always large and uneasy to handle. The complexity of taxonomical profiling, functional annotation, and mechanisms of complex interaction still needs better bioinformatics tools. Current review focuses on better (e.g., AI-driven algorithms) tools that can predict metabolic pathways and interactions, and manipulate complex data to address potential bias for accurate interpretation.

Keywords: Metagenomics, CRISPR-cas9, genome engineering, sequencing, next generation sequencing

Introduction

Metagenomics is defined as the study of the genome of the total microbiota found in nature and is often referred to as microbial environmental genomics because It entails the examination of a group of genetic components (genomes) from a diverse community of organisms in a particular setting. The exploration of microbial diversity, population structure, genetic and evolutionary links, functional activity, and cooperative linkages with the environment is made possible through metagenomics [1]. It is a part of omics technology, a sub-field of sys-

tem biology that accounts for the bigger picture of molecules at cellular, organismal, and tissue levels. Therefore, to understand the concepts of metagenomics, it is important to get a little overview of the components of omics. Omics encompasses DNA, RNA, proteins, metabolites, and various components associated with the comprehensive analysis of all aspects of biological molecules in a system-wide procedure. Beyond these molecular interactions, omics as a field extends as interdisciplinary approaches and integrative analysis are often employed to gain insights regarding how it serves as a driving force in biomedical research, drug discovery, agriculture, and the environment. Genomics is concerned with the study and analysis of an organism's genes in their entirety.

DNA, as a chemical compound, contains coded instructions through which all organisms develop. It is made of two strands that look much like a twisted ladder loaded with four nucleotides (i.e., Adenine (A), Guanine (G), Cytosine (C), and Thymine (T)) that pair up as a result of complementarity in structures to form sequences that make up the double-helical structure. In these arrangements, A-T and G-C form amino acids, which are used in building proteins that code for genes. As stipulated by the National Human Genome Research Institute, it is estimated that the human genome is composed of 20,000 to 25,000 genes, and an average of three proteins can be coded for each gene. Proteomics is a complete set of proteins encoded by the genome followed by their analysis, characterized by protein synthesis, gene expression, and posttranscriptional modifications. According to some scientists [2], the primary structural proteins make up the extracellular matrix with an average of 1045 amino acids.

Similarly, a group of researchers [3] reported that Mass spectrometry-based proteomics for schizophrenia biomarker discovery and pathway analysis in human peripheral fluids were assessed using a thorough systematic review and meta-analysis. In comparison to healthy controls, the study found 217 changed proteins in schizophrenia patients, including ficolin-3, which consistently had elevated levels. The findings show the promise of mass spectrometry-based proteomics for therapeutic uses in finding biomarkers and identifying proteome patterns linked to schizophrenia. Transcriptomics studies all RNAs (e.g., messenger RNAs (mRNAs), microRNAs (miRNAs), transfer RNA (tRNA), ribosomal RNA (rRNA), small interference RNA (siRNA), and small nucleolar RNA (snoRNA)) expressed at the cellular, tissue, or organismal level as well as their expression, trafficking, and degradation levels (**Figure 1**). Unlike eukaryotes, due to the lack of the 3'-end poly(A) tail, which is thought to be a characteristic of mature mRNA in eukaryotes, prokaryotes' transcriptome research is limited. It was also noted that understanding transcriptomes gives insight into how intricate development processes work or helps people grasp disorders. Sequencing the genome of larger organisms requires extensive efforts in analyzing all transcriptomes.

Additionally, the National Institute of Health (NIH) constructed a library containing sequences of humans, mice, and rats and also instituted other programs such as the Genotype-Tissue Expression Project (GTex) and the Encyclopedia of DNA Element (ENCODE). Metagenomics is the entire set of genetic material, which can be obtained straight from specimens from the environment or from a complex association set of genetic material such as a waste-treatment system or bioreactor [4]. This genetic material is typically total DNA, but it can also include complementary DNA (cDNA) created by reverse transcription of extracted community RNA. Microbiological studies cover cutting-edge, futuristic, state-of-the-art, and high-throughput sequencing analysis [5]. Such technologies are required because traditional microbiology methods have some shortcomings and restrictions. With the right tools, bacteria that are present in our gut, bowel, and soil can be investigated using metagenomics. For instance, although our intestines contain 500 different types of bacteria, a single gram of soil has 4000-5000 different species of microbes. The study of microbes in any system is made possible by metagenomics. In essence, metagenomics uses multiple lab-based and in-silico technologies for the detection of harmful viruses without the need for special aseptic conditions. Metagenomics research has been challenging up until this point due to viruses' lack of distinctive conserved genes. Moreover, cellular DNA contamination, free ambient DNA contamination, and the ongoing evolution of

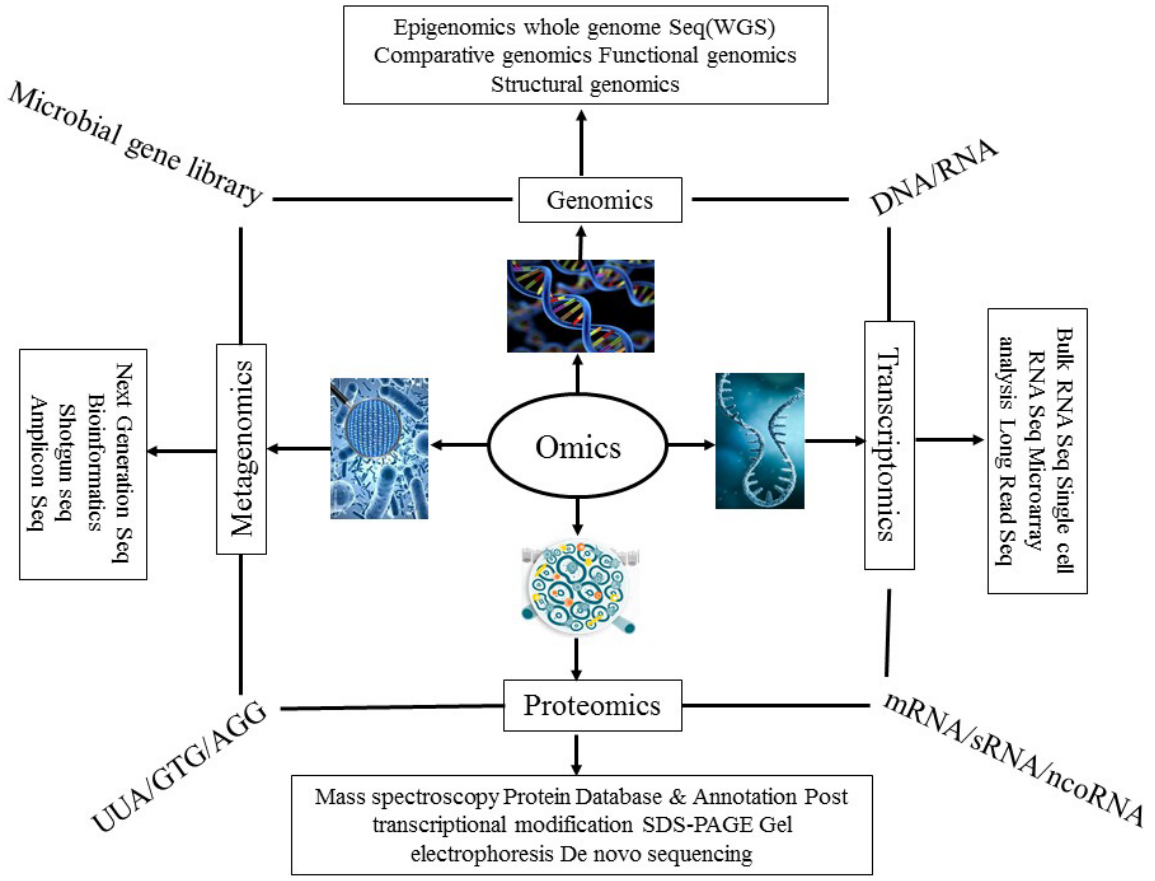

Figure 1. An illustration of omics technologies in gene formation.

viruses are additional difficulties in the field of metagenomics. The significance of metagenomics is not only the discovery of genes and categorization of various microbiota, but it can also be useful in the discovery of pathogenic microbes in cases of disease outbreaks to identify drug-resistance genes and their mechanisms to improve therapies. Scientists are still having issues finding the best approaches to handle large data omics technology variation, how missing values are handled, how to understand complex systems models, and issues with the annotation of data, storage, and computing resources [6, 7]. As CRISPR-cas9 technology is used in conjunction with metagenomics approaches, its challenges also have to be emphasized. Some researchers also noted the current difficulties in creating long arrays of gRNAs and in predicting how these gRNAs will behave in living cells [8].

The discovery of novel biocatalysts from environmental samples has increased most recently as a result of the efficient application of metagenomics. Also, because functional metagenomics is still in its early stages and there are many technical difficulties with a direct screening of potential enzyme-coding genes and the availability of suitable host-vector combinations for the successful expression of metagenome-derived enzymes, the technique's great potential is constrained. This review examines current developments in metagenomic sequencing, discussing its implications for plants and animals and the interplay between gene engineering (CRISPR-cas9) and metagenomics. It analyses the emergence, growth, use, and effect of cutting-edge genome engineering and technology and its clinical applications.

A synopsis of the evolution and current developments of CRISPR and Cas proteins

The acronym "CRISPR" stands for Clustered regularly inter-spaced palindromic repeat, which refers to the unique pattern of DNA sequences found in the genomes of bacteria

and archaea. Cas9 refers to the CRISPR-associated protein 9, which is an endonuclease that cuts DNA molecules at a precise location, specifically in the "Protospacer Adjacent Motif" region. Both CRISPR and cas9 proteins together form a complex called the CRISPR-cas9 complex, which can be used to introduce specific genetic changes in organisms, including humans, animals, and plants. The CRISPR-cas9 complex relies on repeated sequences derived from the genome of a bacteriophage that cleaves DNA molecules at a precise location. Upon its discovery in the 1980s, the technology that was used naturally by bacteria as adaptive immunization has now become a revolutionary tool and a game changer in modern molecular biology. To finally reach this level of CRISPR-cas9 technology, it took scientists more than two decades until the first engineered cas proteins (cas9) were made in 2012 to cut any part of the genome using modified sequences made in the laboratory. A breakthrough that has since been implemented in different genetic experiments. Originally, CRISPR was foreign to the genes of its host. Mechanically, it enters the host cell through infection by a bacteriophage, is picked up by cas9, gets cleaved, and is incorporated into the host cell's genome. By using this mechanism, CRISPR-Cas9 technology allows researchers to make precise insertions and deletions in the DNA. Some scientists [9] found that by directly reversing disease-causing mutations, CRISPR-Cas9 also possesses tremendous therapeutic promise for the treatment of genetic conditions. The acronym "CAS" stands for CRISPR-associated proteins whose function is to search the genome of a host cell and make a cut into a region known as the protospacer adjacent motif (PAM). According to some scientists [10], the CAS family is categorized into two classes, six types, and 33 subtypes.

Class one uses multiple proteins for cleavage, while class two usually requires one protein with guide RNA and cas genes as basic components. Furthermore, class two are the most commonly used proteins because they can easily be constructed and are simple. The most commonly used class two cas proteins include cas1, cas2, cas9, and 13. The classification is based on the mechanism of action of the CRISPR-Cas system, which is viewed in three steps: adaptation, expression, maturation, and interference (collectively, these three make up the effector module). Similarly, CRISPR-associated endonuclease from Prevotella and Francisella (Cpf1), also called Cas12a, emphasizes how guide RNAs (gRNAs) may be used to program Cas12a and drive it to specific DNA targets [11]. Cas12a and Cas9 are often employed for transcriptional control and genetic editing. To enable transcriptional control, including CRISPR-mediated inhibition (CRISPRi) and activation (CRISPRa), Cas12a can be nuclease-null and coupled with effector domains. These endonucleases have played key functions in recent metagenomic approaches, for instance, two CRISPR-cas systems, notably CRISPR-casX and Y in bacteria, were found when sequenced DNA that was directly taken from microbial communities to examine the existence and effectiveness of CRISPR-cas systems in uncultivated species [12]. Their study also uncovered the presence of CRISPR-cas genes in nano archaea.

CRISPR-cas9 recent advancements and strategies in metagenomics

CRISPR-cas9 technology is one of the most frequent methods used in recent developments involving the investigation of samples from microbial communities and other molecular biology research. Many studies highlight the challenges in CRISPR-cas9 gene editing in cancer research and therapy while also referencing a study that utilizes Sinefungin Derivative 7 (Scr7) as a non-homologous ends joints (NHEJs) inhibitor to avoid targets as NHEJs are known to be prone to errors. During the same time, some scientists [13] used CRISPR-cas9 to implement a highly rapid and effective technique to address the plasmodium parasite issue by transfecting the parasite with cas9 expression and linear donor templates, permitting fast cleavage of the targeted locus.

This was accomplished by employing a cloned DNA fragment that encoded the cas9 and dihydrofolate reductase-yeast cytosine deaminase and uracil phosphoribosyl transferase (hdhfr-yfcu) genes into a plasmid, respectively. Here, the linear donor was required to prevent unwanted recombination. Because the method is new and easy to implement, every step became possible and led to the discovery and dissemination of additional information about the technology. Reports also had it that a computation-

al tool referred to as FuNcLib was employed in creating a cas9, and after its activities in yeast were evaluated, it was discovered that the engineered regions cohesively bonded to produce overstimulated cas9 variants, which led to the conclusion that cas9 variants function not only in yeast but also in mammalian cells. Additionally, a novel V CRISPR-cas system called CRISPR-casπ (cas12l) was discovered in environmental metagenomics, and it has the power to alter mammalian cells.

In contrast to cas12 and cas9, the findings showed that the casπ enzyme has the distinctive property of surrounding the target DNA [14]. Currently, CRISPR-cas9 is used to manipulate DNA sequences in organisms of interest. Researchers have demonstrated its therapeutic potential in treating genetic diseases, and it is also used to create Genetically Modified Organisms (GMOs) for agricultural purposes. This method also paved the way for newly created proteins like cas12 and cas13. However, reducing the off-targets has been the driving force behind the expansion of the method, as scientists are reviewing various strategies for better results while legal and ethical concerns surrounding the use of this technology in certain experiments (such as editing the embryonic germline, for example) are still being discussed.

Traditional sequencing methods used in metagenomics

Early sequencing exploited an analytical chemistry technique that only detected two nucleotides. It was using this technique that Robert Holley and his team created the first entire genome nucleic acid sequence (Alanine) in 1965. Following this, a method utilizing radiolabel to partly digest fragments after they were separated into two parts allowing for a relative amount of rRNA and tRNA sequences. A technique employed by Walter Fier's laboratory created the coat of Bacteriophage MS2, which is the first sequence of a protein-coding gene, in 1972. Later, the "plus and minus" method developed by Allan and Sanger and the "chemical cleavage" method developed by Maxam and Gilbert were used to sequence the phi-X174 bacteriophage genome via radiolabeled DNA polymerase-primer in a polymerization reaction aided by polyacrylamide gel (Heather, 2016). Finally, the Chain termination method introduced by Sanger and his team in 1977, which would later be named Sanger's sequencing method, was utilized to achieve 16,569 base pairs from sequencing the human mitochondria gene and 48,502 base pairs for the bacterium genome.

The basic elements of this approach are a primer, DNA polymerase enzyme, single-stranded DNA template (ssDNA), designed di-deoxynucleotide triphosphate (ddNTPs), and regular dinucleotide triphosphate (dNTPs) to solve problem of time and energy associated with this method. Leroy Hood and partner Michael Hunkapiller made some modifications to the techniques in 1987 by using fluorescence dye in place of the radioactive label and generating data through computer analysis. This was recognized among methods of first-generation sequencing as the first significant development and was adopted for sequencing for over three decades [15].

Second generation sequencing

In 1996, second-generation sequencing kicked off with the creation of a novel DNA sequencing approach, pyrosequencing, by Mostafa Ronaghi, Mathias Uhlen, and Pl Nyen. Pyrosequencing is a machine-readable technique that looks at the luminescence emitted during the sequencing process. It is otherwise known as pyrophosphate sequencing and has been categorized as high-throughput sequencing. The "sequencing-by-ligation" platform created via the SOLiD system in 2007 and the "sequencing-by-synthesis" (Ion Torrent) that detects hydrogen ions when new DNA is synthesized are additional platforms at this level. These platforms differ from first-generation technologies because they are rapid, high throughput sequencing, relatively more samples, and a whole genome could be sequenced compared to earlier methods with only small fragments (low scalability). However, the second-generation platforms offer more expensive approaches.

Next-generation sequencing and the human genome project

Due to its capacity to process numerous DNA sequences at once, NGS is the most sophisticated of all the earlier techniques. According to

Mobley [16], the Massive Parallel Sequencing (MPSS) technique, created in 2000, was the first of these. Later, 454 Life Science devised the Roche GS20 from pyrosequencing technology with a high rate of generating up to 20 million base pairs, which became the first commercially available platform in 2004. Similarly, a sequencer called High-Throughput Sequencing X (HiSeq X) Ten was made by Illumina to claim that it had created the first $1,000 genome. This was seen as far less expensive than other methods that could require millions to achieve. Beyond the limits of conventional Sanger procedures, it can provide a very efficient, quick, inexpensive, and accurate way of sequencing DNA. Sadly, rapid technologies had not yet been created at the time of the Human Genome Project (HGP). The 1990-started project came to an end in 2003; in addition to the length of time it lasted, it cost 3 billion dollars.

The Human Genome Project, which cost over $300 million and took 13 years to complete, was able to establish the DNA sequence of the whole human genome in 2003. A complete human genome may now be sequenced in a single day for less than $1,000. The 100,000 Genomes initiative, launched by UK Prime Minister David Cameron in December 2012, surpassed its milestone more quickly than the initial initiative (2003) and led to the establishment of Genomics England. The project was split into two stages. Human chromosomes were separated into DNA segments of adequate sizes for subsequent sub-division into smaller, overlapping DNA pieces during stage one, which was referred to as the shotgun stage. As the shotgun's capacity was insufficient for the project's goal of using the physical map to obtain a vast amount of DNA sequence, it quickly transitioned to the next stage. Stage 2 was characterized by Gap filling and resolving DNA sequences in unclear regions not captured during shotgun. Today, improved technologies can be used to complete HGP in a day and cost far less than before, which is a clear indication that tremendous progress has been made in the last two decades (**Figure 2**).

Identification of microbial genes and pathways

Researchers have created versions of the CRISPR-Cas system in addition to Cas proteins to improve the system's functionality and accuracy. For instance, "base editors" and "prime editors" have been developed to allow for more accurate editing of individual DNA bases or the insertion of particular alterations without the requirement for DNA cleavage. A phylogenetic analysis was carried out to identify and define cas13 proteins by constructing a computational pipeline in which sequenced data was used for detecting CRISPR arrays (regions with short repeated interspaced sequences) and 20 kb regions of DNA analysis for an Open Reading Frame (flanking arrays), selecting the proteins with more than 400 residues for analysis (longer proteins with functional significance), and then building a library with Hidden markers models (HMMs) of all known cas13 genes (cas13a, b, c, and d).

Some scientists [17] noted that It is estimated that higher than 10^{30} microbial cells are thought to exist on Earth while More than 99 percent of the microorganism from habitats are thought to be uncultured with a majority of cultivated prokaryotic species coming from just four phyla: Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria. In metagenomics, researchers study the diversity of microbes leading to the discovery of new genes, metabolic pathways, biological and pharmaceutical, and therapeutic significance. The sequence-based and functional approaches are the two main methods usually adapted for metagenomics analysis. According to some scientists [18], gene extraction (microbial samples), vector cloning of genes of interest, Library construction using clones, and screening and analysis are basic steps of metagenomic sequencing.

Microbial DNA extraction and construction of the library

To extract metagenomic DNA, extraction of all genes in the sample and avoiding fragment contamination (maintaining its purity and integrity) are key points to consider. To preserve larger DNA and more from environmental microbes as feasible, the sample collection must adhere to a precise protocol. Microbes are lysed using physical by physically crushing or sonication, chemical - the use of chemical agents such as Sodium dodecyl Sulfate (SDS), and enzymatic (such as breaking of N-acetyl Glucosamine-N-acetyl muramic acid bond in

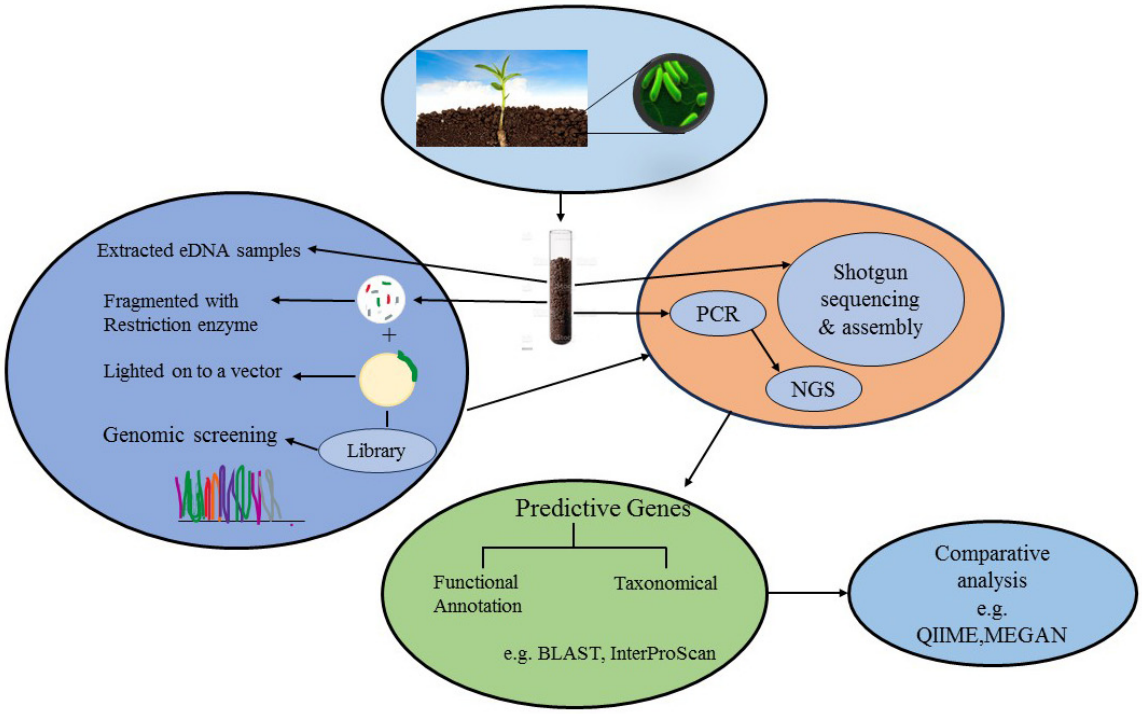

Figure 2. Short protocol of gene extraction and metagenomic Sequencing (This figure summarizes the steps involved in gene extraction and sequencing. 1. Indicate the different microbe and soil. 2. Shows the extraction of DNA from soil sample into the test tube. The extracted DNA is restriction digested with enzymes and ligated to a vector, cloned and used to setup a genomic library. 3. Screening of libraries involves harnessing of various sequencing techniques and PCR. 4. Predictive genes are derived based on functional genes and taxon using various software such as BLAST. 5. Analysis of genomes using different measures requires techniques such as QIIME, MEGAN etc.).

the case of bacteria) methods. And depending on whether the cells were isolated, they could be broken down into direct extraction and indirect extraction [18]. Generally, DNA extraction is carried out after the collection of samples from the environment (soil, water host organisms). The use of physical or chemical methods ensures the fragmentation of the DNA into smaller pieces and then the fragments are ligated onto suitable vectors (e.g., plasmids, bacteriophages), etc. to create a recombinant DNA which is then transformed into hosts such as bacteria or yeast cells. The host organisms are used as replication machinery for the inserted DNA. Through culturing, the transformed host organism begins to replicate thereby making multiple copies of the gene of interest.

A vector can take a specific amount of DNA depending on the type used. For instance, phage (λ) based vectors have a capacity of 20 kb as a desirable fragment compare to other vectors such as cosmids that can take up to 40 kb. When employing an enzymatic approach, a thorough breakdown generally produces fragments that are excessively minuscule for cloning in case the targeted gene to be cloned encompasses various sites for a specific restriction enzyme. This scenario can result in certain segments being absent from the library. To address this concern, partial digestion offers a solution by employing varying amounts of restriction enzymes to acquire fragments of optimal dimensions, factored by both sticky and blunt ends. Initially, the construction of a genomic Library was a problem especially when cloning large, as well as fragments from multiple sources. If care is not taken, some clones may not be represented in the library and this can lead to an incomplete analysis of an organism's genome. To solve this, Carke and Carbon proposed this equation in 1978; $[N=\frac{Ln(1-p)}{Ln(1-\frac{1}{n})}]$ an equation used to generate the required number of clones, where: P = the probability of including all DNA sequences in a random Library, and N = the number of individual recombinants to be screened. For instance, to obtain 95% probability (e.g., Human genome)

using 20 kb as insert size, the number of clones required would be $N = \frac{Ln(1-95)}{Ln(1-\frac{1}{1}.4\times10^{5})} = 4.2\times10^{5}$.

Identifying, classifying, and analysing, clones in genomic Libraries requires the use of methods such as hybridization methods, assays, or NGS techniques. Lambda or cosmids are widely used since it is possible to create a copy of a complete gene that includes both the coding sequence and the regulatory elements within a single clone. In the case of hybridization, labeled probes flanked with sequences of the desired gene sequence(s) in a solution are used. The work of some scientists [19], showed that the PowerSoil DNA extraction kit (MOBIO laboratories, Calshal, "Max" version) for DNA analysis using the Standard Operation Procedure (SOP) was used in the DNA extraction protocol.

Sometimes the DNA is subjected to Propidium Monoxide (PMA) to distinguish between viable and dead cells using Shotgun sequencing, whereas PCR is used in conjunction with Amplicon Sequencing and Illumina Sequencing to generate paired-end Reads. Data analysis and interpretation follow after DNA extraction. Several platforms are available, such as QIIME to obtain sequences, PICTRUST for functional prediction and analysis, ANOVA and LSD as multiple comparison methods (ANOVA is used for comparing 2 different groups, while LSD is a Post hoc test for specific differences between 2 pairs of a group), Quant-IT PiCOGreen dsDNA assay kit for fluorescence quantification, Illumina MiSeq for paired-end sequences (Library preparation), Shotgun sequencing for library preparation, Next Generation sequencing (NGS) for metagenomic analysis, etc. [19, 20]. Functional annotation uses OUT clustering for taxonomical classification, Quantitative Insights into Microbial Ecology (QIIME), Mothur Version 1.2.1-used for calculating the richness of species in samples, etc.

A description of how functional analysis of complex microbial community is carried out

Researchers can unravel the ecological functions, interactions, and contributions of individual microbes within the community by using functional analysis. Most metagenomics methodologies use a variety of sequencing techniques to examine Meta-transcriptomic or metaproteomic. The selection of a particular community or the collection of samples using a few techniques, such as Polymerase Chain Reaction (PCR) and other sequencing methods, are fundamental to the process (**Figure 3**). Also, it is significant to the underlying bedrock such as functional categories and diversity, microbial abundance sequencing, and community structure. For instance, a comparative analysis study [21] revealed for functional categories and diversity (metabolism-high, nutrient cycling-moderate, energy production-moderate, signaling-low & stress response-moderate), Illumina MiSeq PE250 was adopted as a sequencing technique resulting in proteobacteria (19.86%) as the abundant group in a study involving the use of 30 Chinese Cordyceps samples from 3 plots randomly. The same was observed in a study conducted in the open pit mining area of Ke'e's village in Shanxi province (China) [20] where function diversity & categories were high for metabolism and absent (or not investigated) in the other factors using 16s rRNA gene sequencing techniques and presenting proteobacteria as the highest in percentage across all samples. Additionally, some scientists [19] found 34 poly extremophiles in high abundance, as metabolism was also high in functional categories and diversity while investigating microbial communities in extreme terrestrial environments using Next generation sequencing techniques. Functional analysis as emphasized in the examples of past studies above is well defined as (i) metabolism (functional diversity: high/moderate/low; correlation with other functions: +/-/no correlation with nutrient cycling); nutrient cycling (functional diversity: high/moderate/low; correlation with other functions: +/-/no correlation with stress response); energy production (functional diversity: high/moderate/low; correlation with other functions: +/-/no correlation with other functions etc.).

Current implications of metagenomics in plants

Metagenomics has emerged as a powerful tool in plant research with diverse implications. Researchers have indicated that metagenomics enables the study of microbial communities associated with plants, providing insights into functional diversity, metabolism, and interactions. It involves DNA extraction, high-throu-

Figure 3. An illustration of the availability of which revolutionize our understanding of biological systems, enabling targeted interventions, improved diagnostics, and innovative strategies for sustainability. With these, there is more we can do with genes.

ghput sequencing, and bioinformatic analysis. Metatranscriptomics and metaproteomics complement the analysis by assessing gene expression and protein activity. The impacts of metagenomics on plants include enhanced crop productivity, disease management, and ecosystem understanding. Beneficial microbial communities can be identified, leading to targeted microbial inoculants for improved plant growth and disease resistance. Metagenomics aids in disease detection, allowing timely control measures. It also provides insights into microbial interactions, nutrient cycling, and ecosystem functioning. Practical applications span various areas. In plant breeding, metagenomics aids in selecting beneficial microbes associated with desired traits. See **Figure 4**. Disease management benefits from early pathogen detection. Ecosystem studies benefit from assessing soil health and the impacts of land management practices. Recent studies on the impact of metagenomics on plants can be found in **Table 1**.

Implication of metagenomics in animals

Recent studies encompass a range of topics related to metagenomics in animals, including the contribution of the rumen microbiome to global methane emissions, the role of the ruminal microbial community in ruminants, the impact of metagenomics on animal health, and the potential for genetic control of the rumen microbiome. sequencing methods, such as metagenomic Next-Generation Sequencing (mNGS), to analyze the microbial communities present in animals. The provision of insights into the composition and functions of microbial communities in animals, particularly in the rumen of ruminants, plays a crucial role in the digestion of complex plant carbohydrates and the production of nutrients that are beneficial for the host animal. Metagenomic studies have been instrumental in disease surveillance, early detection of pathogens, and understanding zoonotic risks.

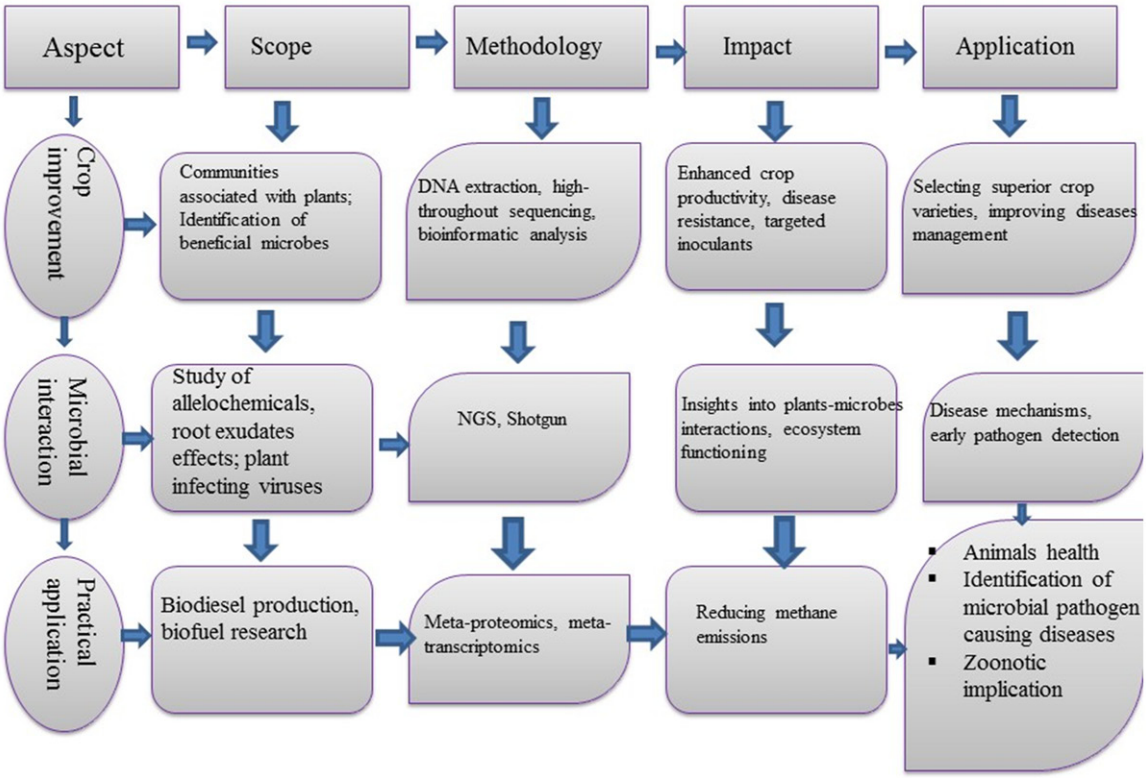

Figure 4. A Flow Chart showing the integration of Diverse Sequencing Techniques in crops improvement and disease control. It illustrates how microbes can be harnessed through sequencing and bioinformatic analysis and engineered to increase crops output (Agriculture), used to identify microbials pathogens in plants, animal and humans and as well as to ascertain the function of microbes.

Table 1. Implication of metagenomics in plants using five recent studies

Research title	Sample size	Gene Extraction method	Seq. method	No. of Reads	References
Analysis of the bacterial populations living on Chinese Cordyceps: structure and estimated functional analysis	5 different samples (QF, XF, ZF, NaF, NyF)	High-throughput sequencing method	Illumina MiSeq PE250	591,672	[21]
Microbial Community Composition and Predicted Functional Assessment in Reclaimed Soil with Different Vegetation Types (e.g., Xiaoyi Mine Waste Dump in Shanxi)	Not mentioned	E.Z.N.A. DNA/ RNA Isolation kit & E.Z.N.A Soil DNA Kit	16sRNA gene Sequencing	-	[20]
Intact communities of microbes living in harsh, essential for astrobiology anaerobic environments: Taxonomic and functional analysis	(6) IM, PPF2, PF2, RR, SS1, Hypersaline environment	Arch344F and Arch915R primers	Illumina MiSeq paired-end sequencing	37,516 reads	[19]
The genetic and enzymatic capability of the microbial community participating in the decomposition of a multifaceted microbial compound is shown by cultivation-independent and cultivation-dependent metagenomes	Not specified	ORF prediction	Shotgun metagenomics sequencing	18,762,958 reads (average of 1,563,580 reads/sample)	[22]

NB: This Table shows summary of 4 studies from different microbial communities through the Metagenomic tools are utilized to analyse diverse microbial communities in various environments and investigate their genetic and functional characteristics. Study 1: "Analysis of the bacterial populations living on Chinese Cordyceps" had five different samples (QF, XF, ZF, NaF, NyF) analysed using high-throughput sequencing (Illumina MiSeq PE250), resulting in 591,672 reads. Study 2: "Microbial Community Composition and Predicted Functional Assessment in Reclaimed Soil with Different Vegetation Types" used gene extraction kits and 16sRNA gene sequencing, but sample size and read count are not specified. Study 3: "Intact communities of microbes living in harsh anaerobic environments" examined six samples using Illumina MiSeq paired-end sequencing, generating 37,516 reads with Arch344F and Arch915R primers. Study 4: "Genetic and enzymatic capability of microbial community in decomposing multifaceted microbial compound" employed ORF prediction and shotgun metagenomics sequencing, resulting in an average of 18,762,958 reads (1,563,580 reads per sample) with an unspecified sample size.

By analyzing metagenomic data, specific microbial pathogens are identified and effective measures are implemented for disease prevention and control. For instance, recent studies high-

Table 2. Implications of metagenomics in animals enhanced by CRISPR-Cas9 technique from recent studies

Targeted species	Purpose	Sequencing method	CRISPR-protein	Impact on the studied organism	Reference
Livestock	Identify genetic markers, perform genome-wide association studies, and genome selection	SNP chips, Whole-genome	Cas9	Identify genetically superior animals based on specific traits, (efficiency), reproduction, disease resistance, and animal welfare)	[23]
Zebrafish	Study the physiological consequences of PARN loss-of-function	Not specified	Cas9	Role of PARN in oogenesis and gonadal maturation in zebrafish; demonstrate the impact of PARN deficiency on sex determination and gonadal development	[24]
Poultry	Study the diversity and functions of the poultry gut microbiome	Amplicon sequencing, metagenomics	Not mentioned	Enable comprehensive analysis of the entire poultry gut microbiome, improved health and productivity of poultry birds	[25]
Cattle	Genome engineering to enhance food production, animal health, and welfare	Not specified	Cas9	Improves the efficiency of food production, enhances animal health and welfare, and introduces valuable traits	[26]
Niches, arthropod disease vectors, and the human microbiome	Microbiome analysis in diseased and healthy states	NGS	Cas9	Analysis of human host genes and transcriptome in response to infection, aiding in disease diagnosis and evaluation of disease risk	[27]

NB: The tables provide a visual representation of key findings, comparisons, trends, or relationships between different variables from recent studies including: 1. Targeted species in each study, 2. Sequencing method applied, 3. CRISPR endonuclease, and 4. Impact on the study organisms. It provides a comprehensive overview of the application of metagenomics and CRISPR-Cas9 in various animal species, highlighting the potential benefits in terms of disease resistance, allergen eradication, product generation, gender-specific birth, the introduction of valuable traits, and overall animal health and welfare. The table also emphasizes the importance of accurate guide RNA design and the use of computational tools to ensure efficiency and specificity in the CRISPR editing process. insights into the current state of metagenomics and gene editing in animals, identify successful applications.

light the importance of specific microbial taxa in the rumen, such as Firmicutes, Bacteroidetes, Methanobrevibacter, and Fibrobacter, in the digestion and nutrient utilization of ruminants. The viromes of common livestock, including cattle, small ruminants, poultry, and pigs, have been characterized, revealing known and novel viruses with potential zoonotic implications. The identification of pathogens causing disease, such as avian and swine influenza, not only offer opportunities to reduce methane emissions but also improve animal health and enhance feed efficiency. Refer to **Table 2** for metagenomic impart in animals.

Clinical applications

Metagenomics sequencing techniques are highly recommended to be used in case of a disease outbreak, in the medical diagnosis of pathogenic microorganisms, and in the identification of infectious diseases in hospitals and communities. By sequencing the genomes present in clinical samples and analyzing the final data, unexpected microbes can be uncovered, including fastidious ones. A large amount of sequenced data generated can be analyzed using bioinformatics tools such as Basic Local Alignment Tools (BLAST) to compare sequence reads to a reference database such as the National Center for Biotechnology Information (NCBI) GenBank. The matching sequence of known genes or genetic markers aids in providing a detailed understanding of the identity and abundance of microorganisms in the samples. Here are some ways in which metagenomic techniques can be harnessed:

i. Assessing the sensitivity of microorganisms to drugs: Studying genetic sequences in clinical samples, metagenomics detects resistance genes or mutations, revealing microorganisms' drug responses. To identify the presence of resistance factors, data can be compared with databases of known genes for antibiotic resistance, such as the Comprehensive Antibiotic Resistance Database (CARD) or the ResFinder database. This information may then be used to determine whether the microbes are susceptible to different antimicrobial agents or resistant to them. This helps to assist in choosing suitable antimicrobial treatments and deriving more targeted and efficient therapeutic approaches [28, 29].

ii. Detection of communicable illnesses: It is used as a procedure for diagnosing infectious illnesses that involves reading the genetic information in patient samples and analysing the results. Compared to traditional approaches like cultivating cultures or certain PCR tests, this technology makes it possible to identify a

larger range of microorganisms and has been shown to work in a variety of settings that include infectious illnesses, including infections of the respiratory system, circulation, digestive system, brain, and spinal cord. delivering quicker outcomes in a matter of hours.

iii. Identification of hard-to-culture or dead microorganisms: Metagenomics can help identify these cases by spotting microorganisms that are tough to grow in labs or that might have died because of previous exposure to drugs Metagenomics can help identify these cases.

iv. Personalizing drug regimens for infections: Metagenomics can help customize treatment plans for individual patients by using information about the microorganisms identified and how they respond to drugs.

Future prospects

As indicated in various research studies lately, metagenomic sequencing has been promising and fast progressing since the 1970s. Metagenomics offers the potential for more efficient and less costly solutions in molecular biology, research, medicine, and agriculture and highlights the rise of new sequencing techniques that can provide more accurate and high-throughput analysis of complex microbial communities [30]. These advancements in metagenomics sequencing have implications for various fields, including the discovery of novel biocatalysts, understanding of microbial interactions, and the development of targeted microbial inoculants for improved crop productivity and disease management [31], which also play crucial roles in animal health, disease surveillance, and identifying zoonotic risks [32]. The integration of omics technologies, such as metagenomics, proteomics, and transcriptomics, further enhances our understanding of molecular biology and enables genetic engineering advancements. Metagenomics sequencing holds immense potential for transformative advancements in scientific research and applications. With the enhancement of new techniques such as NGS, and shotgun, Whole Exome Sequencing (WES), genome engineering has become easier with CRISP-cas technologies, reducing the level of off-target activities and yielding far better outcomes that provide a more extensive understanding of the complexity of genes compared to two decades ago. Soon, with the basis of these methods, scientists will enhance new technologies that will offer unbiased, efficient, and most importantly easy to handle and less costly.

Conclusion

Metagenomics and sequencing technologies have deepened our understanding of biological systems, opening doors to targeted interventions, improved diagnostics, and better strategies for sustainability. In the realm of plant metagenomics, researchers have explored the composition and functional analysis of bacterial communities in Chinese Cordyceps, reclaimed soil with different vegetation types, and extreme, astrobiology-relevant, anoxic sites, among other selected recent studies. These studies have utilized high-throughput sequencing methods, such as Illumina MiSeq, and have provided insights into the microbial diversity and functional potential of these ecosystems. In animals, the use of CRISPR-Cas9 technology has enhanced our ability to study and alter the microbiome and genomes of species. Here are a few instances: the identification of genetic markers in livestock, the physiological consequences of gene loss-of-function in zebrafish, the analysis of the poultry gut microbiome, genome engineering in cattle, and microbiome analysis in disease states, etc. These applications have the potential to improve animal health, disease resistance, productivity, and overall welfare. The availability of a wide variety of sequencing methods has enabled researchers to delve deep into the genetic and functional diversity of microbial communities associated with plants and animals. Using CRISPR-Cas9 in metagenomic approaches is a boast to the sequencing era. The ability to extract valuable insights from complex biological systems, coupled with the ability to manipulate genes and microbial communities, opens up new avenues for improved disease diagnostics, personalized medicine, enhanced crop productivity, and sustainable animal farming practices. However, challenges remain in fully realizing these transformative possibilities. One key limitation is the vastness and complexity of metagenomic data, which demands sophisticated computational tools for accurate interpretation. Ongoing research is focusing on developing AI-driven algorithms that can predict metabolic pathways, identify potential interactions,

and prioritize gene targets for manipulation. We can envision clinical diagnostics that incorporate metagenomic profiling to swiftly identify pathogens, track disease outbreaks, and personalize treatment regimens aided by bioinformatic techniques.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Jyoti Prakash Sahoo, Department of Agriculture and Allied Sciences, C.V. Raman Global University, Bhubaneswar 752054, Odisha, India. E-mail: jyotiprakashsahoo2010@gmail.com; jyotiprakash.sahoo@cguodisha.ac.in

References

[1] Zhang L, Chen F, Zeng Z, Xu M, Sun F, Yang L, Bi X, Lin Y, Gao Y, Hao H, Yi W, Li M and Xie Y. Advances in metagenomics and its application in environmental microorganisms. Front Microbiol 2021; 12: 766364.

[2] Naba A. Ten years of extracellular matrix proteomics: accomplishments, challenges, and future perspectives. Mol Cell Proteomics 2023; 22: 100528.

[3] Rodrigues JE, Martinho A, Santa C, Madeira N, Coroa M, Santos V, Martins MJ, Pato CN, Macedo A and Manadas B. Systematic review and meta-analysis of mass spectrometry proteomics applied to human peripheral fluids to assess potential biomarkers of schizophrenia. Int J Mol Sci 2022; 23: 4917.

[4] Ladino-Orjuela G, Gomes E, da Silva R, Salt C and Parsons JR. Metabolic pathways for degradation of aromatic hydrocarbons by bacteria. Rev Environ Contam Toxicol 2016; 237: 105-21.

[5] Humix. Exploring the microbial world: an introduction to metagenomic studies. 2023; Available at: https://geneticeducation.co.in/humix/video/1a0a015679455d370918e0b48e3c813a60dd8a76ebb6343f06bee8eb7996f141. Accessed on: 16th August 2023.

[6] Bhukya PL and Nawadkar R. Potential applications and challenges of metagenomics in human viral infections. InTech 2018; 75023.

[7] Tarazona S, Arzalluz-Luque A and Conesa A. Undisclosed, unmet and neglected challenges in multi-omics studies. Nat Comput Sci 2021; 1: 395-402.

[8] McCarty NS, Graham AE, Studená L and Ledesma-Amaro R. Multiplexed CRISPR technologies for gene editing and transcriptional regulation. Nat Commun 2020; 11: 1281.

[9] Tycko J, Myer VE and Hsu PD. Methods for optimizing CRISPR-Cas9 genome editing specificity. Mol Cell 2016; 63: 355-70.

[10] Zhang D, Zhang Z, Unver T and Zhang B. CRISPR/Cas: a powerful tool for gene function study and crop improvement. J Adv Res 2020; 29: 207-221.

[11] McCarty NS, Graham AE, Studená L and Ledesma-Amaro R. Multiplexed CRISPR technologies for gene editing and transcriptional regulation. Nat Commun 2020; 11: 1281.

[12] Burstein D, Harrington LB, Strutt SC, Probst AJ, Anantharaman K, Thomas BC, Doudna JA and Banfield JF. New CRISPR-Cas systems from uncultivated microbes. Nature 2017; 542: 237-241.

[13] Shinzawa N, Nishi T, Hiyoshi F, Motooka D, Yuda M and Iwanaga S. Improvement of CRISPR/Cas9 system by transfecting Cas9-expressing Plasmodium berghei with linear donor template. Commun Biol 2020; 3: 426.

[14] Sun A, Li CP, Chen Z, Zhang S, Li DY, Yang Y, Li LQ, Zhao Y, Wang K, Li Z, Liu J, Liu S, Wang J and Liu JG. The compact Casπ (Cas12l) 'bracelet' provides a unique structural platform for DNA manipulation. Cell Res 2023; 33: 229-244.

[15] Lavelle TA, Feng X, Keisler M, Cohen JT, Neumann PJ, Prichard D, Schroeder BE, Salyakina D, Espinal PS, Weidner SB and Maron JL. Cost-effectiveness of exome and genome sequencing for children with rare and undiagnosed conditions. Genet Med 2022; 24: 1349-1361.

[16] Mobley I. A brief history of Next Generation Sequencing (NGS) - front line genomics. 2021; Available at: https://frontlinegenomics.com/a-brief-history-of-next-generation-sequencing-ngs/. Accessed on: 16th August 2023.

[17] Kharaghani D, Gitigard P, Ohtani H, Kim KO, Ullah S, Saito Y, Khan MQ and Kim IS. Design and characterization of dual drug delivery based on in-situ assembled PVA/PAN core-shell nanofibers for wound dressing application. Sci Rep 2019; 9: 12640.

[18] Zhang L, Chen T, Wang Y, Zhang S, Lv Q, Kong D, Jiang H, Zheng Y, Ren Y, Huang W, Liu P and Jiang Y. Comparison analysis of different DNA extraction methods on suitability for long-read metagenomic nanopore sequencing. Front Cell Infect Microbiol 2022; 12: 919903.

[19] Bashir AK, Wink L, Duller S, Schwendner P, Cockell C, Rettberg P, Mahnert A, Beblo-Vranesevic K, Bohmeier M, Rabbow E, Gaboyer F, Westall F, Walter N, Cabezas P, Garcia-Descalzo L, Gomez F, Malki M, Amils R, Ehrenfreund P, Monaghan E, Vannier P, Marteinsson V, Erlacher A, Tanski G, Strauss J, Bashir M, Riedo A and Moissl-Eichinger C. Taxonomic and functional analyses of intact microbial communi-

ties thriving in extreme, astrobiology-relevant, anoxic sites. Microbiome 2021; 9: 50.
[20] Dong Z, Hou HP, Liu HY, Wang C, Ding ZY and Xiong JT. Microbial community structure and predictive functional analysis in reclaimed soil with different vegetation types: the example of the Xiaoyi mine waste dump in Shanxi. Land 2023; 12: 456.
[21] Xia F, Zhou X, Liu Y, Li Y, Bai X and Zhou X. Composition and predictive functional analysis of bacterial communities inhabiting Chinese Cordyceps insight into conserved core microbiome. BMC Microbiol 2019; 19: 105.
[22] Costa OYA, de Hollander M, Pijl A, Liu B and Kuramae EE. Cultivation-independent and cultivation-dependent metagenomes reveal genetic and enzymatic potential of microbial community involved in the degradation of a complex microbial polymer. Microbiome 2020; 8: 76.
[23] Chakraborty D, Sharma N, Kour S, Sodhi SS, Gupta MK, Lee SJ and Son YO. Applications of omics technology for livestock selection and improvement. Front Genet 2022; 13: 774113.
[24] Nanjappa DP, De Saffel H, Kalladka K, Arjuna S, Babu N, Prasad K, Sips P and Chakraborty A. Poly (A)-specific ribonuclease deficiency impacts oogenesis in zebrafish. Sci Rep 2023; 13: 10026.
[25] Aruwa CE, Pillay C, Nyaga MM and Sabiu S. Poultry gut health - microbiome functions, environmental impacts, microbiome engineering and advancements in characterization technologies. J Anim Sci Biotechnol 2021; 12: 119.
[26] Singh P and Ali SA. Impact of CRISPR-Cas9-based genome engineering in farm animals. Vet Sci 2021; 8: 122.
[27] Chiu CY and Miller SA. Clinical metagenomics. Nat Rev Genet 2019; 20: 341-355.
[28] Charalampous T, Kay GL, Richardson H, Aydin A, Baldan R, Jeanes C, Rae D, Grundy S, Turner DJ, Wain J, Leggett RM, Livermore DM and O'Grady J. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol 2019; 37: 783-792.
[29] Tarazona S, Arzalluz-Luque A and Conesa A. Undisclosed, unmet and neglected challenges in multi-omics studies. Nat Comput Sci 2021; 1: 395-402.
[30] Junier T, Huber M, Schmutz S, Kufner V, Zagordi O, Neuenschwander S, Ramette A, Kubacki J, Bachofen C, Qi W, Laubscher F, Cordey S, Kaiser L, Beuret C, Barbié V, Fellay J and Lebrand A. Viral metagenomics in the clinical realm: lessons learned from a Swiss-wide ring trial. Genes (Basel) 2019; 10: 655.
[31] Parveen R, Kumar M, Swapnil, Singh D, Shahani M, Imam Z and Sahoo JP. Understanding the genomic selection for crop improvement: current progress and future prospects. Mol Genet Genomics 2023; 298: 813-821.
[32] Hosokawa M, Endoh T, Kamata K, Arikawa K, Nishikawa Y, Kogawa M, Saeki T, Yoda T and Takeyama H. Strain-level profiling of viable microbial community by selective single-cell genome sequencing. Sci Rep 2022; 12: 4443.