



## RETRACTED ARTICLE: Metagenomics for taxonomy profiling: tools and approaches

Mukesh Kumar Awasthi, B. Ravindran, Surendra Sarsaiya, Hongyu Chen, Steven Wainaina, Ekta Singh, Tao Liu, Sunil Kumar, Ashok Pandey, Lal Singh & Zengqiang Zhang

To cite this article: Mukesh Kumar Awasthi, B. Ravindran, Surendra Sarsaiya, Hongyu Chen, Steven Wainaina, Ekta Singh, Tao Liu, Sunil Kumar, Ashok Pandey, Lal Singh & Zengqiang Zhang (2020) RETRACTED ARTICLE: Metagenomics for taxonomy profiling: tools and approaches, Bioengineered, 11:1, 356-374, DOI: [10.1080/21655979.2020.1736238](https://doi.org/10.1080/21655979.2020.1736238)

To link to this article: <https://doi.org/10.1080/21655979.2020.1736238>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 09 Mar 2020.



Submit your article to this journal [↗](#)



Article views: 11427



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 82 View citing articles [↗](#)

## RETRACTED ARTICLE: Metagenomics for taxonomy profiling: tools and approaches

Mukesh Kumar Awasthi<sup>1a,b</sup>, B. Ravindran<sup>c</sup>, Surendra Sarsaiya<sup>d</sup>, Hongyu Chen<sup>e</sup>, Steven Wainaina<sup>1b</sup>, Ekta Singh<sup>f</sup>, Tao Liu<sup>a</sup>, Sunil Kumar<sup>f</sup>, Ashok Pandey<sup>g</sup>, Lal Singh<sup>f</sup>, and Zengqiang Zhang<sup>a</sup>

<sup>a</sup>College of Natural Resources and Environment, Northwest A&F University, Yangling, Shaanxi Province, China; <sup>b</sup>Swedish Centre for Resource Recovery, University of Borås, Borås, Sweden; <sup>c</sup>Department of Environmental Energy and Engineering, Kyonggi University Youngtong-Gu, Suwon, South Korea; <sup>d</sup>Key Laboratory of Basic Pharmacology of Ministry of Education, Zunyi Medical University, Zunyi, Guizhou, China; <sup>e</sup>Institute of Biology, Freie Universität Berlin Altensteinstr, Berlin, Germany; <sup>f</sup>CSIR-National Environmental Engineering Research Institute, Nagpur, India; <sup>g</sup>Centre for Innovation and Translational Research CSIR-Indian Institute of Toxicology Research, Lucknow, India

### ABSTRACT

The study of metagenomics is an emerging field that identifies the total genetic materials in an organism along with the set of all genetic materials like deoxyribonucleic acid and ribose nucleic acid, which play a key role with the maintenance of cellular functions. The best part of this technology is that it gives more flexibility to environmental microbiologists to instantly pioneer the immense genetic variability of microbial communities. However, it is intensively complex to identify the suitable sequencing measures of any specific gene that can exclusively indicate the involvement of microbial metagenomes and be able to advance valuable results about these communities. This review provides an overview of the metagenomic advancement that has been advantageous for aggregation of more knowledge about specific genes, microbial communities and its metabolic pathways. More specific drawbacks of metagenomes technology mainly depend on sequence-based analysis. Therefore, this 'targeted based metagenomics' approach will give comprehensive knowledge about the ecological, evolutionary and functional sequence of significantly important genes that naturally exist in living beings either human, animal and microorganisms from distinctive ecosystems.

### ARTICLE HISTORY

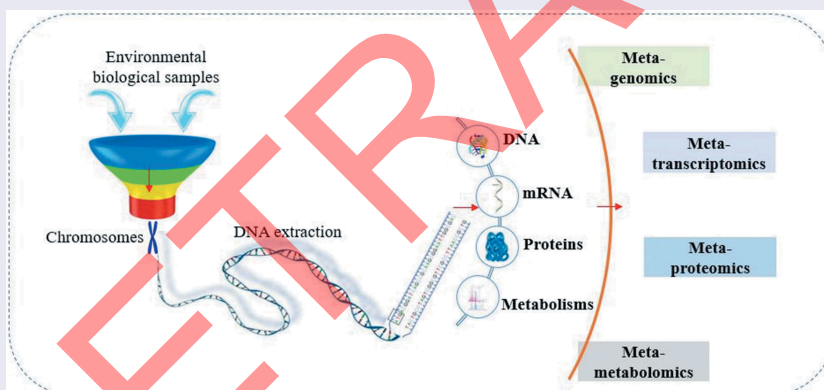
Received 24 December 2019

Revised 20 February 2020

Accepted 21 February 2020

### KEYWORDS

metagenomics; genomics; deoxyribonucleic acid; ribose nucleic acid; microbial metagenomes



## 1. Introduction

Since the invention of the microscope and other analogous observational instruments that could be exploited to explore the micro world, we humans are becoming conscious that microorganisms are of equal significance with plants and animals and have played a vital role in the dynamic derivation of earth biosphere. With the passage of time, scientists

recognized that a single study of one pure species in the laboratory scale contradicts laws of nature, since organisms cannot survive and grow as one single species. They could only exist as a member of interdependent communities in various natural environments. Thus, the amount of microbes existing in nature cannot be incubated alone in the laboratory. Hence, in order to establish gene inventories, a further

**CONTACT** Mukesh Kumar Awasthi  [mukesh\\_awasthi45@yahoo.com](mailto:mukesh_awasthi45@yahoo.com)  College of Natural Resources and Environment, Northwest A&F University, Yangling, Shaanxi Province 712100, PR China

This article has been retracted. Please see Retraction Statement (<http://dx.doi.org/10.1080/21655979.2022.2126093>)

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

better understanding to functions of populations and communities in an ecosystem, etc., is necessary and inevitable to study the hereditary information of all entangled species. In 1998, 'metagenomic' was first proposed as 'the collective genomes of soil microflora' and approached to investigate genetic information of all populations [1]. Later, 'metagenomic' was defined as 'the functional and sequence-based analysis of the collective microbial genomes that are contained in an environmental sample' by Allen and Banfield [2] and Cowan et al. [3]. Since the first application of metagenomics, now it has been comprehensively utilized to study microbial diversities in a certain environmental community by analyzing selected genes such as 16 S rRNA.

### 1.1. Classification of metagenomic

The purpose of the taxonomic classification of metagenomic sequences is to find the exact species of origin, data of which are also primarily performed with the objective of cataloging/classifying various microbial groups inhabiting a given environment. Therefore, taxonomic classification is a crucial procedure for different applications of metagenomic such as disease diagnostics, microbiome analyses, and outbreak tracing, etc. Current taxonomic classification algorithms mainly utilize handcrafted sequence composition features such as oligonucleotide frequency [4,5].

Given massive bio-sequencing data, it is extremely critical to efficiently analyze the categories of these

biological sequences within a limited time, so as to capture characteristics of the microbial community which researchers are interested in. Afterward, researchers can further investigate the species diversity of microbial communities by applying classification methods of metagenomics samples to classify sequencing data from environmental samples. The metagenomics sample classification methods study the relationship between changes of environmental or host state and microbial community by identifying or predicting categories of metagenomics samples. The classification methods of metagenomics sequencing data can be divided into two categories according to different sample data processed methods: one is based on the sequence of marker genes such as 16SrRNA, the other is based on whole-genome sequencing fragments.

### 1.2. Basic tools and techniques used in metagenomic

Metagenomic is based on genetic recombinant technology. The general procedure for metagenomic research has been depicted in Figure 1, which generally includes the enrichment of samples or genes (groups), extraction of genomic DNA in a specific environment, construction of metagenomic DNA library, screening of the target gene, and expression of active product of the target gene. Among them, the methods of direct cracking

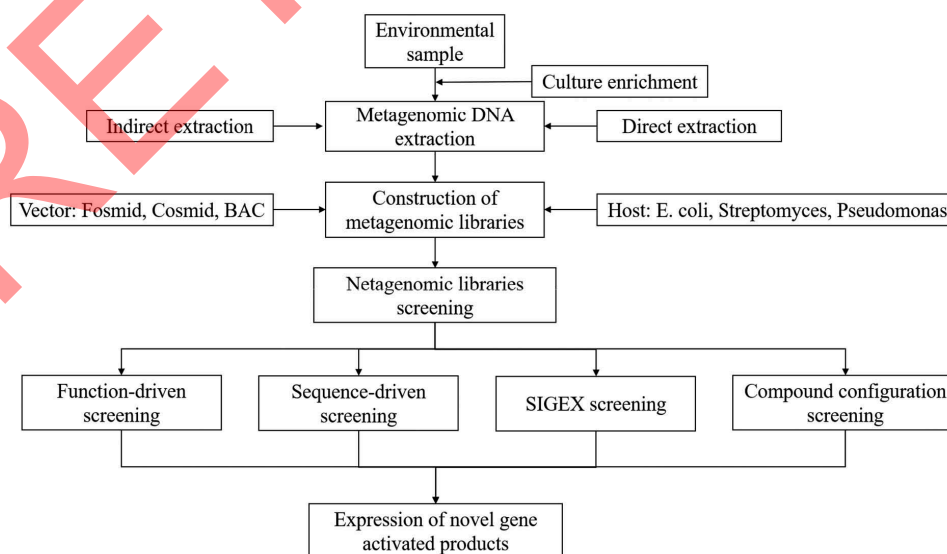
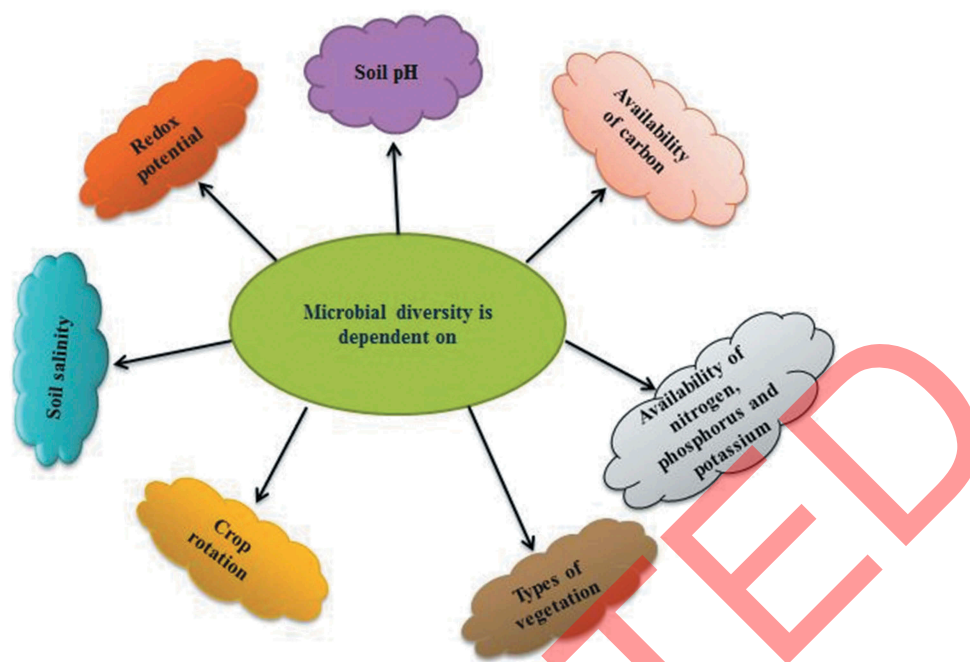
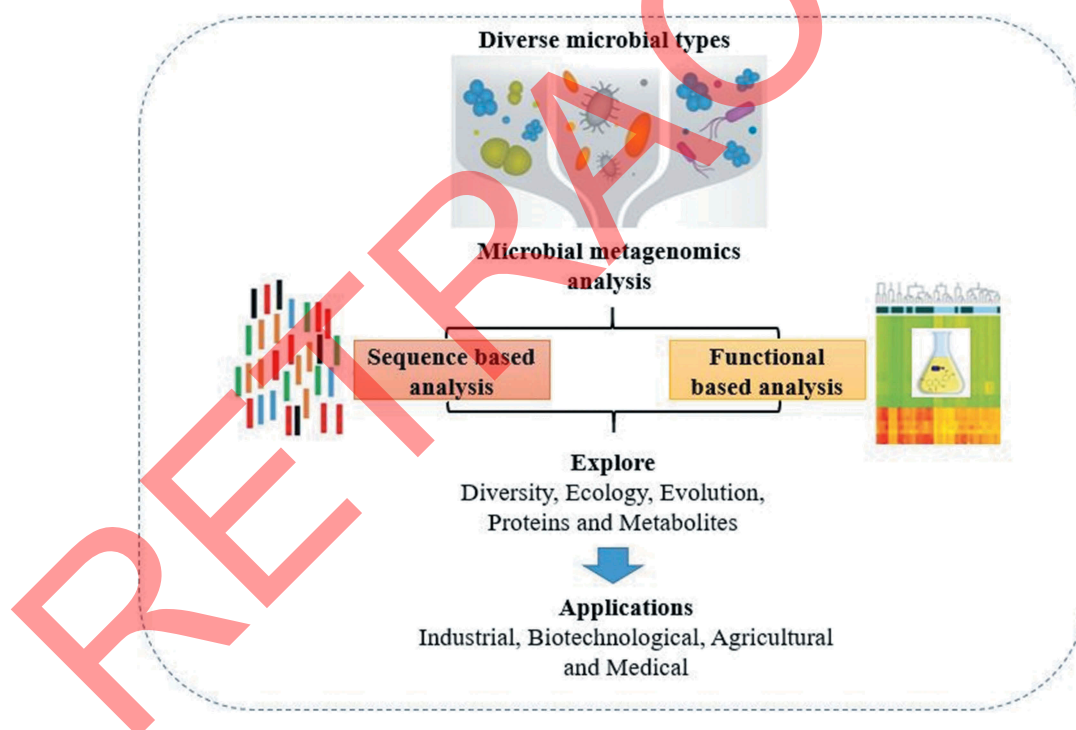


Figure 1. General process of metagenomic workflow.



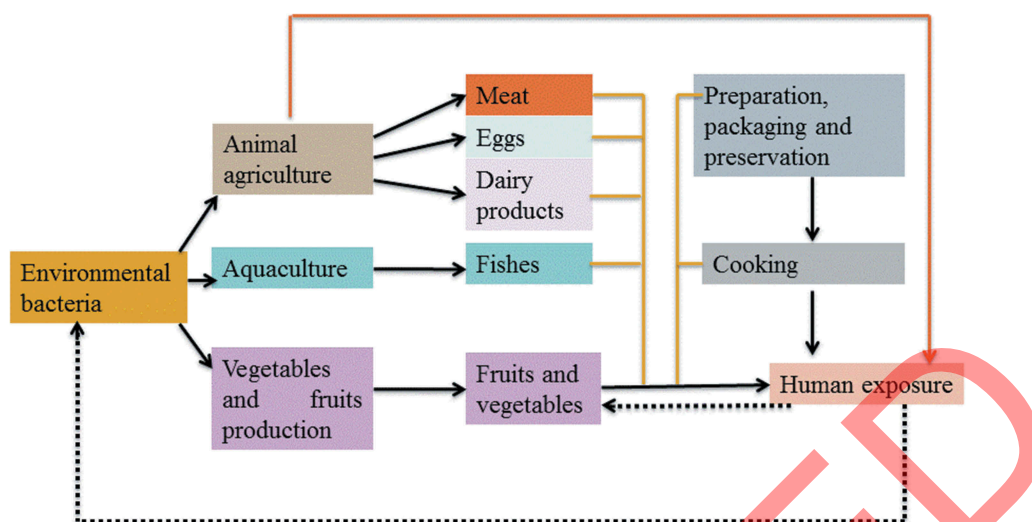
**Figure 2.** Systematic presentation of several environmental factors influencing soil microbial diversity.



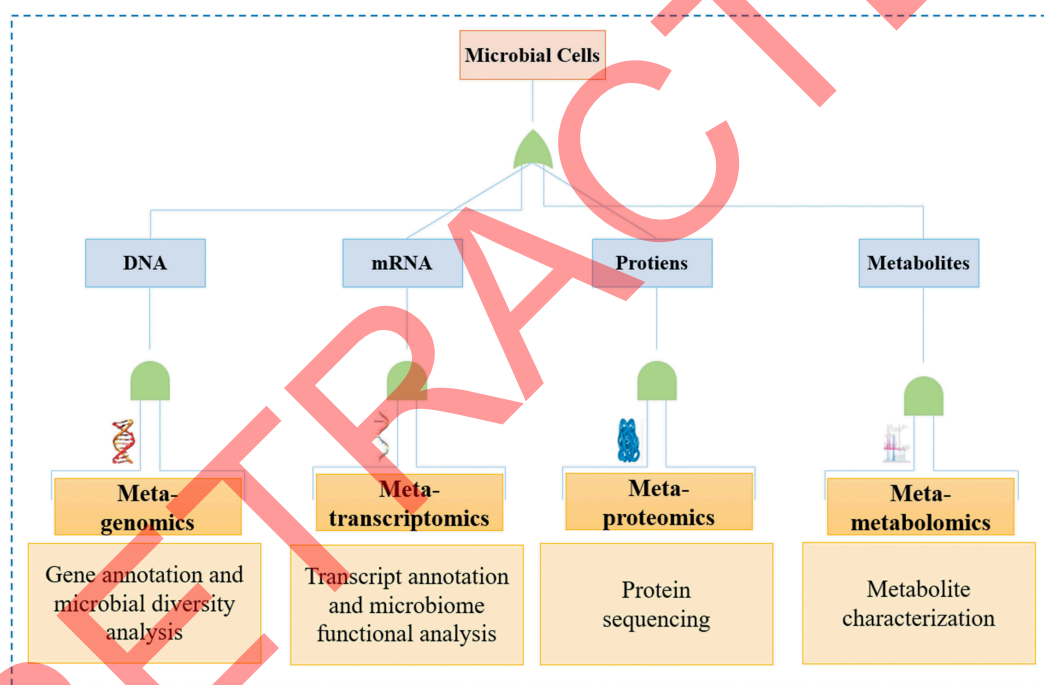
**Figure 3.** Systematic presentation of metagenomics.

and indirect extraction are two conventional technologies used for DNA enrichment and extraction from environmental samples in the metagenomic analysis. DNA from the samples should be extracted completely and maintain larger

fragments to obtain unbroken target genes or gene clusters as much as possible. The construction of the genomic library is a prerequisite for revealing new genes. During this procedure, the choice of vector and host system depends on the



**Figure 4.** Schematic overview of the flow of resistant microbes (bacteria) and resistance genes in the food supply chain.



**Figure 5.** An overview of biotechnological tools for microbial community analysis.

quality of extracted DNA and purpose of research, which the size of insert, number of vector copies required, host selected as well as screening methods all need to be taken into consideration. The next step is to effectively utilize the abundant resources of the library to discover new biomolecules. Due to the high complexity of the environmental genome, high-throughput and highly sensitive technologies are required to screen and identify useful genes in libraries. Currently, widely adopted screening technologies can be broadly

divided into four categories: function-driven screening, sequence-driven screening, SIGEX (substrate-induced gene-expression) screening and compound configuration screening.

With the expeditious development of the metagenomic research in recent years, the quantity of sequencing microbial genomes increased exponentially, which means that the comparison requirement for novel sequencing reads is also synchronously increasing. This requests the development of efficient tools which could analysis metagenomics data accurately.

**Table 1.** Characteristics of different classifiers.

Type	Classifier	Custom databases	Generates abundance profile	Memory required	Time required	Reference
DNA	CLARK	Yes	Yes	80 Gb	2 min	[119]
	Kraken	Yes	Yes	190 Gb	1 min	[120]
	MegaBLAST	Yes	No	61 Gb	4 h	[121]
	Prophyle	Yes	No	40 Gb	40 min	[122]
Protein	DIAMOND	Yes	No	110 Gb (varies)	10 min	[123]
Markers	MetaPhlan2	No	Yes	2 Gb	1 min	[124]
	mOTUs2	No	Yes	2 Gb	1 min	[125]

Numerous tools have been exploited to classify metagenomic data and estimate taxon abundance profiles. Table 1 compares the characteristics of various classifiers that are commonly chosen to apply in metagenomics technologies.

## 2. Novel biological resources screened from uncultured bacteria by metagenomic method

Nowadays, more than 99% of microorganisms are currently difficult or cannot be purely cultured in microbiology field, which means our knowledge of microbes is basically limited in less than 1% of microorganisms [1]. Those uncultured microorganisms possess great application potential, their metabolites may produce amounts of valuable compounds. The emergence and development of metagenomic have made up for this deficiency and serve us commence to gradually explore and understand the potential micro-world. It has been widely recognized that metagenomic can be utilized to discover novel genes, enzymes, as well as compounds from uncultured bacteria. Currently, researchers have employed metagenomic to obtain numerous novel functional genes and active substances, which have been extensively applied in various fields. Two primary categories are ordinarily utilized in metagenomic to discover and identify novel natural genes and compounds from environmental samples, which are sequence-based screening and function-based screening, respectively. For sequence-based screening, the PCR technology is employed to detect the existence of genes and compounds, while function-based screening in metagenomic is conducted by investigating the activity of enzymes [6–8]. These two technologies have deciphered amounts of novel genes encoding valuable biocatalysts from uncultured bacteria existing in different specific environments, which have already been broadly practiced in

many industries such as environmental restoration, new medical compounds, and alternative energy, etc.

In addition, notable companies include DNA Star, Inc (US), Illumina, Inc. (US), Oxford Nanopore Technologies Ltd. (UK), IntegraGen SA (France), Macrogen Inc. (South Korea), Thermo Fisher Scientific, Inc (US), Novogene Corporation (China), Allwegene Co. Ltd (China), CD Genomics (USA), BGI (USA), Zymo Research Corporation (US) and Microsynth AG (Switzerland) that are providing services regarding metagenomic analysis.

## 3. Metagenomic analysis and its applications in different sector

### 3.1. Agriculture

In the last few decades, the production of crops has been increased worldwide. The quality of soil can alter plant production in agricultural soil. In general, the soil system contains an enormous amount of microbial population (e.g., 10 billion microbes/gram) as well documented by Rosselló-Mora and Amann [9]. Microorganisms and their metabolic activities can enhance crop production [10,11]. In addition, agriculture is playing an important role in the world economy [11,12]. Therefore, there is an urgent requirement of metagenomic analyses to understand the microbial communities in the soil system [13]. The metagenomic analysis is normally used to examine various microbial communities from the environment (Figure 4). The microbes can stimulate the cycling of macro as well as micronutrients and also microbes release the essential enzymes which enhance crop production [14]. Metabolic network and modeling of microorganisms provide the information on various biochemical activities [13]. At present, an enormous number of metabolic re-constructions associated with microorganisms have been modeled [15,16]. Metabolic rebuild

of microorganisms, namely, *Saccharomyces cerevisiae* and *Escherichia coli* is analyzed using the dynamic flux balance analysis [17]. According to previous literature, metagenomic models were not well addressed. These approaches can detect the metabolic features from sequence data.

### 3.2. Biofuel

Due to the growing interest of minimizing the negative environmental impacts accompanied by the utilization of petroleum during energy production, alternative energy sources, such as biofuels, are being explored [18]. The availability of feedstocks needed for the production of biofuels (e.g. waste gas, biomass and organic residues) can be largely guaranteed [18–20]. The types of biofuels that can be produced include bioethanol, biogas, biohydrogen, and biodiesel [18,21]. The bioconversion of the waste materials and residues into biofuels involves biochemical reactions that essentially control the microbial activities [22]. During the bioconversion process, metagenomic analysis is critical in characterizing the functional microbial communities involved as well as the associated metabolic pathways [23]. The metagenomic analysis tools can also be applied during the biologically mediated substrate preparation steps prior to the actual fermentation. For instance, Kundu et al. [24] classified the metabolic interaction network of gut compartments of *Nasutitermes corniger* used for lignin disintegration. The pretreatment of lignin-rich materials is critical for biomass utilization in the bioethanol production process [25]. Biogas is produced when organic materials are broken down in oxygen-deficient conditions in the so-called anaerobic digestion process catalyzed by naturally-occurring microflora. The process involves four interdependent stages facilitated by hydrolytic, acidogenic, acetogenic, and methanogenic microbes (Table 2). The diverse groups of microorganisms that are active during the biogas production process have been studied by several authors [18,23,26].

A remarkable research that was conducted by Campanaro et al. [27] revealed the phylogenetic and functional structure of the microbial population in bioreactors digesting cattle manure. Their novel approach entailed the application of high-

throughput sequencing and binning procedures which resulted in a comprehensive report on the metabolic pathways involved. Biohydrogen can be produced in anaerobic digesters in the acidogenic and acetogenic steps, alongside other metabolites such as volatile fatty acids (VFAs) [28]. When the anaerobic digestion processes are aimed at these intermediate metabolites, the activity of methanogenic archaea must be deactivated while appropriate methods are applied to manipulate the microbial pathways toward the targeted product [29]. The VFAs are potential precursors in the biodiesel production process due to their use as carbon sources for cultivating oleaginous microbes [30]. The identification of bacteria responsible for the acidogenic fermentation has been widely performed, sometimes alongside the methanogenic archaea profiling [27,31,32]. Some of the analytical tools that can be mentioned include 454 pyrosequencing and fluorescent in situ hybridization (FISH) [33].

Microalgae are promising microorganisms that can also be used to synthesize biofuels coupled with CO<sub>2</sub> fixation [34]. Metagenomic analysis has also been performed on these microorganisms. For instance, Sambles et al. [35] used high-throughput Illumina sequencing and 16S rDNA analysis to study the microbial composition of *Botryococcus braunii* at diverse treatments. The microbial analysis is also crucial in understanding other important components during the bioprocessing of microalgae, such as biofouling of photo-bioreactors [36]. A previous research was conducted to identify the metagenome of a bacterial biofilm in a photo-bioreactor using 16S rRNA phylogeny after which a KEGG pathway analysis was performed. The bacterial species originated from *Chlorella vulgaris* and *Scenedesmus obliquus* [36].

### 3.3. Biotechnology

In the present century, various molecular techniques include terminal restriction fragment length polymorphism (T-RFLP), denaturing gradient gel electrophoresis (DGGE), polymerase chain reaction (PCR) cloning, fluorescence in situ hybridization (FISH) are applied to microorganisms, these modified microbes contain different applications [37–39]. These microbial techniques can enhance the knowledge in the field

**Table 2.** Correlation between soil physicochemical parameters and bacterial OTUs.

Taxa	Parameters	Positive correlation (+) and negative correlation (-)	References
Proteobacteria	pH	+	[126]
	Total organic carbon	+	[127]
	Total nitrogen	+	[128]
Actinobacteria	Phosphorus	+	[129]
	pH	+	[130]
	Total organic carbon	-	[126]
Acidobacteria	Total nitrogen	-	[126]
	Phosphorus	-	[126]
	pH	-	[128]
Firmicutes	Total organic carbon	+	[124]
	Total nitrogen	+	[125]
	Phosphorus	+	[127]
Bacteroidetes	pH	-	[129]
	Total organic carbon	+	[124]
	Total nitrogen	ns	[125]
Verrucomicrobia	Phosphorus	ns	[128]
	pH	+	[123]
	Total organic carbon	-	[124]
Planctomycetes	Total nitrogen	-	[125]
	Phosphorus	+	[125]
	pH	-	[121]
Cyanobacteria	Total organic carbon	+	[124]
	Total nitrogen	ns	[130]
	Phosphorus	+	[130]
Armatimonadetes	pH	-	[131]
	Total organic carbon	+	[121]
	Total nitrogen	+	[125]
Chloroflexi	Phosphorus	+	[125]
	pH	-	[132]
	Total organic carbon	+	[132]
Chloroflexi	Total nitrogen	+	[133]
	Phosphorus	ns	[125]
	pH	+	[125]
Chloroflexi	Total organic carbon	-	[125]
	Total nitrogen	-	[125]
	Phosphorus	ns	[125]

ns – no significance.

of microbiology, but the interaction of microbes and their mechanisms are not well addressed. Metagenomics is one of the suitable methods to study microbial genetics collected from various environmental samples [40]. Furthermore, metagenomics

has provided the whole-genome sequences of microorganisms [41–43]. Metagenomics is also used and examined the metatranscriptomes from recovered samples to detect the microbes and their gene expression [44,45]. Moreover, many scientists have been using metagenomics to detect the novel microbial genes for biodegradation and also monitoring the biohazards associated with microorganisms (*i.e.*, viruses, various pathogens and, ARGs) as well documented by Wexler et al. [41] and Xia et al. [46].

### 3.4. Ecology

The ecosystem contains many types of microorganisms which regulate the different biological processes and biogeochemical cycles. The large level of microbial diversity can enhance ecosystem stability and productivity [47–49]. Understanding of microbe's interaction between living organisms and their internal mechanisms from the molecular view. Metagenomic methods are promising tools for analyses of microbial ecosystems [50–52].

### 3.5. Environmental remediation

The major sources of environmental pollution associated with industrial based on wastewater/wastewater sludge. In general, industries are consumed different types of chemical substances for the production of various materials and release several hazardous substances at the end of the process. Various industries release different types of hazardous waste materials (e.g., polyaromatic hydrocarbons, phenols, pesticides and toxic heavy metals) which generate serious environmental pollution [53–56]. Bioremediation is environment-friendly and low-cost technology to the reclamation of polluted soils, particularly microbes play an important role in biodegradation of pollutants [53,55,57,58]. Active microbes in contaminated sites can be assessed by genome enhancement using metagenomic analysis (Figure 2). Particularly, stable isotope probing (SIP) techniques employing for enhancement of different macromolecules include RNA, DNA, phospholipid-derived fatty acids [59]. Metagenomic methods are one of the emerging techniques for sequences of various microbes and enhanced microbes through



metagenomics can remove the different pollutants from the contaminated environment.

### 3.6. Pharmaceutical and medical science

At present, many more peoples are affected by various diseases particularly type 1 diabetes, bowel disease (IBD), Crohn's disease and ulcerative colitis. These diseases severely affected the human immune system. In addition, weakness alleles have generated deadly diseases in human beings. Apart from this, some microorganisms can contribute to pathogenesis. The detection of proteins associated with pathogenesis is very complicated. Metagenomics-based technologies can identify specific genes [60,61].

## 4. Microbial metagenomics, metatranscriptomics, and metaproteomics

### 4.1. Microbial metagenomics

With the rapid advancement of high-throughput sequencing, a cultivation-independent methodology, i.e., metagenomics, can become a competent approach to the examination of the ability of certain microbes for refining. Metagenomic microbial sequencing is a competent alternative to rRNA sequencing to examine the complex microbial community structure, and allow screening of a microbial community's innate composition and utilitarian potential [62]. The development of sophisticated molecular approaches in conjunction with computational regulations has resulted in a decline in the sequencing of aggregate genome DNA and several opportunities related to marine microbial exploration (Figure 5). As a result, an entire era of metagenomic sequencing has contributed to more in-depth considerations of representing a wider range of species from varying tests [63–66]. Li et al. [67] showed that metagenomics-based population analyses can provide both community and valuable information without an upgrade slant. This approach has been utilized for comparing functional and requested profiles of the microbial communities related to specific types.

Commercial devices have risen which are arranged to address this issue by upgrading for microbial DNA. One methodology, Unused

Britain Biolab's NEB Next Microbiome DNA Improvement unit, takes advantage of human and other higher orchestrate eukaryotic DNA having tall CpG methylation rates. By utilizing the methylated CpG-specific official protein MBD2 entwined to a human IgG Fc portion, human DNA is particularly bound and confined utilizing Protein A-bound attractive dots [68]. Multiple omics methods are used to govern the microbial community alignment, co-occurrence network along with metabolic pathways. Deng et al. [69] hypothesis revealed that niche transformations caused a distinct microbial assembly and increased microbial ecology circulation, which has the possibility to improve system effectiveness and stability. The higher microbial diversity, efficient redundancy and observed higher system stability are related to the huge niche differences in reactors [69]. Deng et al. [70] research further revealed a similar difference trend in the phylum taxonomy level between floc bacteria and gut microbiota. The microbial community distribution for both water and gut samples was visualized by principal coordinate analysis (PCoA) based on the detected operational taxonomic unit.

### 4.2. Metatranscriptomics

Metagenomics and metatranscriptomics can collect the microorganism's entire genome and transcriptome range by sequencing add up to DNA/RNA from different natural tests, delivering useful data with high determination. The metatranscriptomics represents valuable devices for all-inclusive cataloging microbial quality and transcript profiles and reflecting generally metabolic capacities [71]. Falk et al. [72] uncovered that microbial genomic potential can be remade by metagenomic investigation of environmentally extracted DNA, metatranscriptomics can shed light on the deciphered items of DNA, giving riches of data on add up to flag-bearer RNA (mRNA) yield; an intermediary for quality expression. As microorganisms are coupled to numerous biogeochemical cycles within the environment, metatranscriptomic considers of anthropogenically affected dregs are profitable in numerous aspects; 1) understanding critical reactions of the biosphere to irritations, 2) evaluating in case and how microbial communities are

working to direct the impacts of bequest defilement in oceanic frameworks, and 3) investigating microbial differing qualities as a work of human-induced scene changes. As of late, metagenomic or metatranscriptomic examinations of Italian cheese, French cheese, Mexican cheese and Belgian cheese have been performed, creating modern information with respect to differences and metabolic highlights of microbial species in cheeses [73].

With the improvement of microarray and high-throughput sequencing, the metatranscriptomics has uncovered the startling high microbial differences in corrosive mine seepage, which given a profound knowledge that a much larger assortment of life forms may adjust to this extraordinary environment than already thought. Genome and transcriptome investigations had given an expansive sum of quality data about potential capacities of acidophiles, particularly within the metabolic pathways of carbon, nitrogen, hydrogen and sulfur Liyuan et al. [74]. Lim et al. [75] uncovered that metatranscriptomics of host-associated communities is especially troublesome. Enhancement of the microbial RNA by strategies that utilize manufactured polyadenylation is not pertinent for tests that contain huge amounts of eukaryotic mRNA. The added poly-A tails decrease the sum of valuable arrangement information, particularly when pyrosequencing advances such as Roche/454 are utilized. Due to the brief half-life and small amount of mRNA, test filtration and control with buffer ought to be dodged earlier to RNA extraction. This definitely causes an increment in host RNA defilement when managing with host-associated microbial tests. Metatranscriptomics uncovered critical eCO<sub>2</sub> impacts on the composition and movement of the meadow microbiomes. The decrease in soil parasitic movement was affirmed by RT-qPCR of 18S rRNA; in great agreement with a noteworthy diminish in contagious mRNA included in oxidative phosphorylation (rhizosphere soil), and in collapsing, sorting and debasement (root-associated). Among Organisms, the relative plenitude of most bunches (e.g., *Agaricomycetes* [soil] and *Leotiomycetes* [roots]) diminished, but that of the *Glomeromycetes* (both compartments) expanded. Utilitarian investigation of root mRNA recommends that the generation of plant auxiliary metabolites was expanded within

the summer [76]. Besides, the effect of worldwide warming on the endophytic and epiphytic living beings related with *C. quitensis* remains unclear. Here, a metatranscriptomic approach was utilized to decide the impact of an in situ reenacted worldwide warming situation on *C. quitensis* plants. We found a huge number of differentially communicated qualities effectively explained (2997), recommending that climate alter tweaks the metatranscriptome of *C. quitensis* plants and related endophytes and epiphytes [77–79].

### 4.3. Metaproteomics

Recently, metaproteomics and metabolomics have emerged as effective devices for the characterization of energetic host–microbiome interactions, especially in combination with metagenomics and metatranscriptomics approaches [80]. Metaproteomics can be utilized to consider protein expression from blended cultures and thus this approach can supply coordinate to prove the metabolic and physiological exercises occurring in a given framework (Figure 3). This approach has been utilized somewhere else to consider the relationship between the structure of microbial communities and the working of a biological system [81]. Soil metaproteomics has the potential to contribute to be distant for a better higher understanding of warming impacts on soil living beings as proteins especially talk to dynamic living beings and their physiological working. The contrasts in community work may be related to particular phyla utilizing metaproteomics, showing that microbial adjustment to long-term soil warming basically changed microbial capacities, which is related to upgraded soil breath [82]. Metaproteomics exploits the control of mass spectrometry to recognize wide protein profiles in complex tests, such as digestive tract microbiota. The foremost recent mechanical advances inside the field of mass spectrometry have opened the field of large-scale characterization of microbial proteins. In spite of these hardware progressions, bioinformatics examination remains a fundamental challenge [83].

Shotgun metaproteomics may uncover the functional geography of microbial population, but needs suitable strategies for complicated tests with obscure formulations. Beyter et al. [84] have

displayed Proteo-Storm, a competitive database looks like a framework for large-scale metaproteomics that identifies high-confidence peptide-spectrum matches while accomplishing a speedup of two-to-three orders-of-magnitude over reigning systems. Metaproteomics could be an effective tool for getting information on all proteins recuperated specifically from natural tests at a given time. The proteome information gives to coordinate prove on the biological processes in the soil environment and is the primary detailed reference information from the maize rhizosphere (Table 3). The LC-MS/MS proteomic information is accessible by ProteomeXchange with identifier PXD014519

[85]. The induction of proteins from the respective peptides and the confirmation of the source and function of these proteins are the major steps toward meta-proteomics information analysis. In this way, the instructional exercise introduces the Unipept command-line interface as a platform-independent method for such metaproteomics knowledge assessments. To continue with, the open Unipept commands are given point by point. Next is the regulation of the Unipept command-line interface utilizing two case studies examining a single tryptic peptide and a group of peptides retrieved independently of a shotgun metagenomics test [86].

**Table 3.** Summary of recent research literature on the microbial community.

Region	Year	Published research themes	Process, tools and techniques	Key research outcomes	References
Saudi Arabia	2020	Metagenomics-based evaluation of microbial profiles	Shotgun metagenomic tools; CosmosID metagenomic software; high performance k-mer algorithm databases	Metagenomics revealed a lower alpha diversity for both bacteria and virus	[85]
United States of America	2019	Whole metagenomic sequencing for microbial community	Whole metagenome sequencing; Descriptive analyses using SAS® 9.4	whole metagenomics approach to characterize sediment microbial biodiversity	[63]
China	2019	Metagenomic sequencing for microbial gene catalog	Metagenomic sequencing using an Illumina Hiseq 2000 sequencing system; MetaGene predicted an open reading frames (ORFs); Gene taxonomic classification using BLASTP search	Metagenomic sequencing is conducted to investigate the microbial gene catalog	[62]
Singapore	2019	Bioinformatics analysis of metagenomics data of biogas-producing microbial communities	Next-generation sequencing (NGS) technologies used for qualitative and quantitative analysis of the microbial communities	Described the procedure of processing metagenomics data of microbial communities for revealing metagenomics characterization using bioinformatics approaches	[86]
China	2019	Metatranscriptomics for rhizosphere microbiomes	Random-primed cDNA libraries using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina; The cDNA libraries using the Experion DNA 12 K Analysis Kit (Bio-Rad) and quantified using the Qubit dsDNA HS Assay kit	Metatranscriptomics revealed significant eCO <sub>2</sub> effects on the composition and activity of the grassland microbiomes	[87]
China	2020	Metagenomic insights for microbial community and antimicrobial resistance genes	Shotgun metagenomics sequences processed using Trimmomatic 0.36 and Bowtie2 via the KneadData wrapper software; MEGAHIT v1.1.3 used for de novo metagenomic assembly analyses; The draft genomes annotated using Prokka software and eggNOG-mapper (based on eggNOG 4.5 orthology	The results give first metagenomic insights into the salt-mediated changes in microbial community composition and a broad view of associated antibiotic resistance genes in the process of food fermentation.	[134]
Canada	2018	Mining metagenomic and metatranscriptomic data for microbial metabolic functions	Metagenomics and metatranscriptomics, including nucleic acid extraction, sequencing platform selection, library construction, quality control of sequencing outputs, compositional analysis, assembly-based functional analysis, assembly-free functional analysis, and comparative analysis.	Metagenomics and metatranscriptomics can arrest the whole genome and transcriptome range of microorganisms through sequencing entire DNA/RNA, providing both taxonomic and functional information with high resolution	[69]
Switzerland and Israel		Microbial Metagenomics Mock Scenario-based Sample Simulation (M3S3)	Microbial Metagenomics Mock Scenario-based Sample Simulation (M3S3) workflow generate virtual samples from raw reads or assemblies	The M3S3 tool is support the progress and authentication of identical metagenomics applications	[88]

## 5. Metagenome analysis and interpretation

Metagenomics uncovered a lower alpha difference for microbes from different sources. The current approaches to recognizing changes in the microbial community incorporate qualitative approaches like terminal restriction fragment length polymorphism (T-RFLP) or quantitative approaches like real-time PCR (qPCR). T-RFLP produces fingerprinting profiles for comparative investigation but does not give data on the ordered personalities of microbial contaminants. qPCR may be a focused-on approach that does not give a comprehensive viewpoint on the wide assortment of microscopic organisms, and quality contaminants display in biowaste, soil and wastewater streams [87]. Researchers consider both the focal points and impediments correlated with the 16S rRNA approach with that of the characteristic predisposition. There are outstanding distinctions between 16S rRNA screens and WMGS (whole metagenomic sequencing) strategies for organized and functional population profiling. For example, WMGS provides a thorough sample assessment by sequencing a large majority of random, small fragments in a shotgun approach individually, resulting in more sequencing coverage [63].

The quality ordered classification is rendered utilizing BLASTP look against the NCBI-NR database beneath a scale of e-value  $<10^{-5}$  for metagenome analysis. The repositories of KEGG (Kyoto Encyclopedia of Genes and Genomes) and COG (Clusters of Orthologous Groups) are utilized to bunch and classify the non-redundant quality set. Use of BLASTP look (Blast Version 2.2.28+) for COG and KEGG databases (e-value  $<10^{-5}$ ) is a valuable clarification of consistency. Based on a phylogenetic list of proteins expressed in various genomes of bacteria, archaea and eukaryotes, the COG database is commonly used for practical analysis on recent genomes and developmental study. By influencing the non-reductant content collection against the COG database, the percentage of usable categories of COG was established [62]. Examination of high-throughput sequencing information and a reasonable bioinformatics investigation approach in this manner played a really basic part within the examination of the microbial metagenome. These considered through bioinformatics tools such as MG-RAST, MetaVelvet,

Genovo, etc., most of which borrowed from the areas of data mining, manufactured insights and measurements, have incredibly enhanced our information on microbial compositions and network-based relationship examinations [88]. Qi et al. [89] have utilized the metagenomics by utilizing the Illumina MiSeq high-throughput sequencing of the 16S rRNA quality to uncover the composition of the microbial community. Advance investigation of the shotgun metagenomic sequencing comes about uncovered a *Scedosporium* beginning CYP450 encoding quality, assigned cyp450q1 (GenBank no. MK388855), which offers 99% amino corrosive arrangement likeness (six amino corrosive changes among 525 amino acids) with that of the already detailed short-chain vaporous alkanes and ethers degradation-associated CYP450 quality CYP52L1 (Gen Bank no. AY438638) from *Graphium* ATCC 58400 [90].

Employing a metagenomics approach the relationship between nitrogen expulsion execution and nitrogen change quality sort and abundance, as well as dispersion of important qualities in copious genomes are assessed. This allowed an appraisal of the ordered dispersion, plenitude and genotypes of the prevailing individuals of the community counting anaerobic alkali oxidation microscopic organisms and other individuals of the community [67]. Shotgun metagenomics arrangements are preprocessed for quality sifting of peruses and expulsion of a contaminant peruses (human and/or PhiX genome) utilizing Trimmomatic 0.36 and Bowtie2 (GRCh\_38PhiX database) through the Knead Data wrapper program with 'SLIDING WINDOW: 4:20, MINLEN 150' parameters [67]. To perform metagenomics and metatranscriptomics, the primary step is to disconnect high-quality DNA and RNA from assorted tests. To date, there are two commonly utilized sequencing stages for metagenomics and metatranscriptomics: the Roche/454 stage (out of trade currently) and the Illumina framework (<https://www.illumina.com>). Some considers have moreover attempted the Pacific Biosciences (PacBio) stage and the Ion Torrent stage (<https://www.thermofisher.com>). For metagenomic and metatranscriptomic library development, Illumina gives a wide assortment of commercial packs. A few programs have been created for this step, primarily counting Trimmomatic, PRINSEQ, FASTX-Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)), and NGS QC Toolkit. Among these

programs, Trimmomatic may be an adaptable and proficient preprocessing apparatus that is particularly planned for Illumina yields and can create results prevalent to, or at slightest competitive with, other programs such as PRINSEQ and Fastx-Toolkit [71].

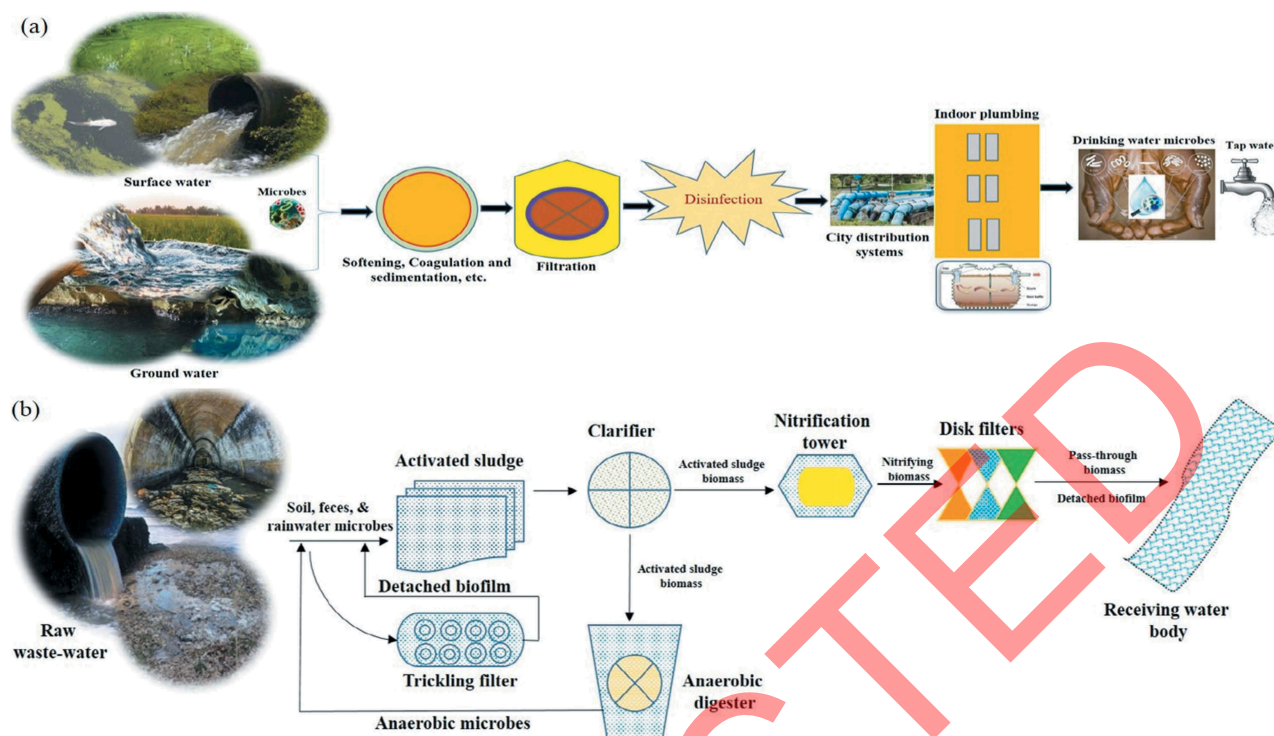
On the other hand, Motro et al. [88] developed the workflow of Microbial Metagenomics Deride Scenario-based Test Reenactment (M3S3) which permits clients to produce simulated tests from simple peruses. The M3S3 yield could be FASTQ or FASTA deride check. M3S3 is tested by creating virtual tests for 10 challenging irresistible illness scenarios, including a foundation framework with pathogens counting blends spiked in silicon. M3S3 is competent in creating on-demand simulated tests, reenacting a variety of complexity-changing symptomatic scenarios. The M3S3 instrument can subsequently bolster the improvement and approval of standardized metagenomics applications. Li et al. [91] have given a Bayesian system for genotype estimation for blends of numerous microscopic organisms, named as Genetic Polymorphisms Assignments (GPA). Recreation results appeared that GPA has diminished the wrong revelation rate and mean absolute error in CNV and single nucleotide variation (SNV) recognizable proof. This system was approved by whole-genome sequencing and Pool-seq information from *Klebsiella pneumoniae* with numerous microscopic organism's blend models, and appeared the high precision within the allele division discoveries of CNVs and SNVs in AMR qualities between two populaces. The quantitative consider on the changes of AMR quality division between two tests appeared a great consistency with the AMR design watched within the person strains. Moreover, the system at the side of the genome explanation and populace comparison apparatuses have been coordinated into an application, which might give a total arrangement for AMR quality recognizable proof and measurement in un-cultural clinical tests.

## 6. Quantifying the contribution of microbial immigration in ecosystem

Microbial populations in any environment be it natural or engineered play major roles in the biogeochemical cycles [92,93]. Different methods have been established, including the niche and neutral hypotheses, to investigate how different microorganisms

aggregate into a population and add up to the ecosystem's capacity [94]. Immigration is amongst the major stochastic processes in the neutral theory of biodiversity and biogeography, which together with death and birth alters population assemblage [95]. This is referred to as migration [96] and was originally used in macroecology to determine the rate of new species reaching a remote island from the closest landmass, i.e. the likelihood of colonization, which plays a critical role in maintaining the balance of faunal diversity of the island [97]. Since the definition of immigration can vary significantly [95], this review adopts the one previously used by Bell [98] and characterizes it as the procedure for introducing a microbial entity to a local community from the metacommunity species pool, which comprises of a lot of communities truly connected by immigration and which can exchange colonies of different species as well.

One such commonly used, the equivalent word is dispersal [99–101]. While there may be some difference between dispersal and immigration in a particular context and one may even integrate the other [95], there is still no consensus on its differentiation. Throughout natural microbial processes, the shift from the upstream to a downstream network can be easily shown by the immigration of microbes from African dust to European marine condition [98], from flowing waters to woodland lakes [99] and also from river water to offshore environment and estuarines [100]. Immigration occurs more frequently in engineered conditions than in natural environments. Microorganisms found in various sources, such as surface and groundwater supplies, act as inocula in drinking water systems (Figure 6(a)) to propagate populations along the treatment procedure [99]. Microorganisms occurring on certain treatment systems, such as filters or reactor wall surfaces, can subsequently be released during or after processing [102] to the delivery units that depend on whether or not disinfection has occurred [103,104]. Bidirectional immigration amidst the water process and the biofilms on the inner pipe surface is feasible in the distribution network [105], which can further be elevated via stagnating the water [106]. Another case of microbial immigration can be seen in the wastewater treatment plant (WWTP) (Figure 6(b)), on the grounds that different bioreactors are directly linked to each other, and the biomass transfer is greater than that in the natural environment. In WWTPs, microbes in raw sewage system [106] and trickling filters [105–107] were recorded to have an effect on the



**Figure 6.** Illustration of potential microbial immigration in (a) waterbodies and distribution system, and (b) wastewater treatment plants.

populations of downstream-activated sludge. The activated sludge biomass [108] also serves as an immigration source for the downstream anaerobic digester [109–111]. The effluent [112] from the WWTP will influence the water supply network [113], thus raising the number of microbes and antibiotic-resistant genes associated with the human gut [114]. While microbial immigration is reported very often in engineered water systems, it remains very challenging to answer to what degree immigration contributes to the assemblage and operation of the downstream community. In this study, focus has been made on the procedure assessing microbial migration using survey methods that are now being used and also in establishing the drawbacks related to them. Immigration is a process that can have a strong impact on microbial communities in natural and engineered environments as well. In any case, it remains a challenge to quantitatively assess the commitment of this procedure to the microbial diversity and its proper functioning in the receiving ecosystems [115,116].

Many approaches such as shared microbial species counting, microbial source monitoring, and commonly used neutral community modeling rely on an abundance profile to predict the degree of coverage between upstream and downstream

networks [117,118]. Therefore, they cannot propose the contribution of immigrants to the role of the downstream communities because individual immigrants are not taken into account after reaching the receiving environment. This impediment can be solved by using a technique that combines a couple of mass balance with the sequencing of DNA, i.e. ecogenomics-based mass balance. This determines the net growth rate of individual microbial immigrants and splits the whole group into several communities that add to a community's production and inactive ones that have almost zero functional benefit. Linking immigrant activities to their abundance also offer quantifying the contribution of an upstream community toward the downstream one. Considering that only active communities may improve the accuracy of categorizing the main ecological parameters which guide the execution of the procedure by using various techniques such as machine learning.

## 7. Conclusions and future prospects

Metagenomics is a dynamic approach to identify the microbial world and has provided significant

advancement between the research group and industrial application with regard to knowledge and importance of non-cultural microbes. Functional metagenomics is an important technique to identify the metabolites source from non-cultivable microorganisms. However, functional metagenomics is still facing many challenges that are yet to be solved because this technology is passing from development stages. Improvement of a metagenomic technology has already revealed to be beneficial. Even in its early years, it recommended that there can be important rewards if suitable advancement and further optimizations are conducted. The emergence of many metagenomic branches confirms the advancement in the techniques along with rapid and economic feasible sequencing assessment and will also acknowledge the amplification of a very encouraging research area in microbiology, genetic engineering, molecular science, pharmaceutical and food industries. This review article provides brief overview of the aptitude of developing metagenomics technology to produce novel industrial products obtained from uncultured microorganisms. Subsequently, the recovery of beneficial metabolites, upcoming challenges in another research area, is the advancement of a customer requirement and economic feasibility as well as the viable end product that can be further produced at industrial scale. In addition, it can also sustain its activity when scaled up and can be obtained purified and formulated appropriately into a final end product and keep its stability during preservation and transportation. A successful association of various metagenomics-based approaches is a requisite for identifying the complication of microbial communities, biological system and correlation with environmental factors in the last few years. In conclusion, the advancement of new 'metagenomics' technologies could be a competent tool to identify the microbial communities, contingent it is made more understandable, better informative, multipurpose, and economically feasible.

## Acknowledgements

We are also thankful to all our laboratory colleagues and research staff members for their constructive advice and help.

## Disclosure Statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by a research fund from the International Young Scientists from National Natural Science Foundation of China [Grant No. 31750110469], China; the Shaanxi Introduced Talent Research Funding [A279021901], China and The Introduction of Talent Research Start-up Fund [Grant No. Z101021904], College of Natural Resources and Environment, Northwest A&F University, Yangling, Shaanxi Province 712100, China. The Mobility for Regional Excellence-2020 [Grant No. RUN 2017-00771], European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No. 754412 at University of Borås, Borås, Sweden, and Guizhou Science and Technology Corporation Platform Talents Fund, China [Grant No.: [2017]5733-001 & CK-1130-002]. We are also thankful to all our laboratory colleagues and research staff members for their constructive advice and help.

## ORCID

Mukesh Kumar Awasthi  <http://orcid.org/0000-0003-0879-184X>

Steven Wainaina  <http://orcid.org/0000-0003-4709-5126>

## References

- [1] Handelsman J, Rondon MR, Brady SF, et al. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol.* 1998;5:245–249.
- [2] Allen EE, Banfield JF. Community genomics in microbial ecology and evolution. *Nat Rev Microbiol.* 2005;3:489–498.
- [3] Cowan DA, Arslanoglu A, Burton SG. Metagenomics, gene discovery, and the ideal biocatalyst. *Biochem Soc Trans.* 2004;32:298–302.
- [4] Rosen GL, Reichenberger ER, Rosenfeld AM. NBC: the naïve bayes classification tool webserver for taxonomic classification of metagenomic reads. *Bioinformatics.* 2011;27:127–129.
- [5] Vervier K, Mahé P, Tournoud M, et al. Large-scale machine learning for metagenomics sequence classification. *Bioinformatics.* 2016;32:1023–1032.
- [6] Kimura N, Sakai K, Nakamura K. Isolation and characterization of a 4-nitrotoluene-oxidizing enzyme from activated sludge by a metagenomic approach. *Microbes Environ.* 2010;25:133–139.
- [7] Fujita MJ, Kimura N, Sakai A, et al. Cloning and heterologous expression of the vibrioferrin biosynthetic gene cluster from a marine metagenomic library. *Biosci Biotechnol Biochem.* 2011;75:2283–2287.
- [8] Fujita MJ, Kimura N, Yokose H, et al. Heterologous production of bisucaberin using a biosynthetic gene cluster cloned from a deep-sea metagenome. *Mol Biosyst.* 2012;8:482–485.

- [9] Rosselló-Mora R, Amann R. The species concept for prokaryotes. *FEMS Microbiol Rev.* 2001;25:39–67.
- [10] Hughes L. Climate change and Australia: trends, projections and impacts. *Aust Ecol.* 2003;28:423–443.
- [11] Hollister EB, Engledow AS, Hammett AJM, et al. Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. *Isme J.* 2010;4(6):829–838.
- [12] Karlsson T. Carbon and nitrogen dynamics in agricultural soils model applications at different scales in time and space. [Doctoral Thesis]. Swedish University of Agricultural Science; 2012.
- [13] Durot M, Bourguignon PY, Schachter V. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol Rev.* 2009;33:164–190.
- [14] Sylvia DM, Hartel PG, Fuhrman JJ, et al. Principles and applications of soil microbiology. In: Principles and applications of soil microbiology. 2nd ed. Upper Saddle River, New Jersey: Pearson Prentice Hall; 2005.
- [15] Forster J, Famili I, Palsson BO, et al. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* 2003;13:244–253.
- [16] Duarte NC, Herrgott MJ. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* 2004;14:1–12.
- [17] Antoniewicz MR. Dynamic metabolic flux analysis—tools for probing transient states of metabolic networks. *Curr Opin Biotechnol.* 2013;24(6):973–978.
- [18] Tsapekos P, Kougias PG, Treu L, et al. Process performance and comparative metagenomic analysis during co-digestion of manure and lignocellulosic biomass for biogas production. *Appl Energy.* 2017;185:126–135.
- [19] Slezak R, Grzelak J, Krzystek L, et al. The effect of initial organic load of the kitchen waste on the production of vfa and h<sub>2</sub> in dark fermentation. *Waste Manage.* 2017;68:610–617.
- [20] Wikandari R, Taherzadeh MJ. Rapid anaerobic digestion of organic solid residuals for biogas production using flocculating bacteria and membrane bioreactors – a critical review. *Biofuel Bioprod Biorefin.* 2019;13(4):1119–1132.
- [21] Schnürer A. Biogas production: microbiology and technology. In: Hatti-Kaul R, Mamo G, Mattiasson B, editors. *Anaerobes in biotechnology*. Cham: Springer International Publishing; 2016. p. 195–234.
- [22] Marshall CW, LaBelle EV, May HD. Production of fuels and chemicals from waste by microbiomes. *Curr Opin Biotechnol.* 2013;24(3):391–397.
- [23] Luo G, Fotidis IA, Angelidaki I. Comparative analysis of taxonomic, functional, and metabolic patterns of microbiomes from 14 full-scale biogas reactors by metagenomic sequencing and radioisotopic analysis. *Biotechnol Biofuels.* 2016;9:51.
- [24] Kundu P, Manna B, Majumder S, et al. Species-wide metabolic interaction network for understanding natural lignocellulose digestion in termite gut microbiota. *Sci Rep.* 2019;9(1):16329.
- [25] Taherzadeh MJ, Karimi K. Pretreatment of lignocellulosic wastes to improve ethanol and biogas production: a review. *Int J Mol Sci.* 2008;9(9):1621–1651.
- [26] Schlüter A, Bekel T, Diaz NN, et al. The metagenome of a biogas-producing microbial community of a production-scale biogas plant fermenter analysed by the 454-pyrosequencing technology. *J Biotechnol.* 2008;136(1):77–90.
- [27] Campanaro S, Treu L, Kougias PG, et al. Metagenomic analysis and functional characterization of the biogas microbiome using high throughput shotgun sequencing and a novel binning strategy. *Biotechnol Biofuels.* 2016;9:26.
- [28] Wainaina S, Parchami M, Mahboubi A, et al. Food waste-derived volatile fatty acids platform using an immersed membrane bioreactor. *Bioresour Technol.* 2019b;274:329–334.
- [29] Wainaina S, Lukitawesa, Awasthi MK, et al. Bioengineering of anaerobic digestion for volatile fatty acids, hydrogen or methane production: a critical review. *Bioengineered.* 2019a;10(1):437–458.
- [30] Béligon V, Noblecourt A, Christophe G, et al. Proof of concept for biorefinery approach aiming at two bioenergy production compartments, hydrogen and biodiesel, coupled by an external membrane. *Biofuels.* 2016;9(2):163–174.
- [31] Sarkar O, Kumar AN, Dahiya S, et al. Regulation of acidogenic metabolism towards enhanced short chain fatty acid biosynthesis from waste: metagenomic profiling. *RSC Adv.* 2016;6(22):18641–18653.
- [32] Shanthi JS, Butti SK, Sarkar O, et al. Electrofermentation of food waste – regulating acidogenesis towards enhanced volatile fatty acids production. *Chem Eng J.* 2018;334:1709–1718.
- [33] Zheng X, Su Y, Li X, et al. Pyrosequencing reveals the key microorganisms involved in sludge alkaline fermentation for efficient short-chain fatty acids production. *Environ Sci Technol.* 2013;47(9):4262–4268.
- [34] Razzak SA, Ali SAM, Hossain MM, et al. Biological CO<sub>2</sub> fixation with production of microalgae in wastewater – a review. *Renew Sust Energ Rev.* 2017;76:379–390.
- [35] Sambles C, Moore K, Lux TM, et al. Metagenomic analysis of the complex microbial consortium associated with cultures of the oil-rich alga *Botryococcus braunii*. *MicrobiologyOpen.* 2017;6(4):e00482.
- [36] Krohn-Molt I, Wemheuer B, Alawi M, et al. Metagenome survey of a multispecies and alga-associated biofilm revealed key elements of bacterial-algal interactions in photobioreactors. *Appl Environ Microbiol.* 2013;79(20):6196–6206.
- [37] Grady CL Jr, Daigger GT, Love NG, et al. *Biological wastewater treatment*. London, U.K: IWA Publishing; 2011.
- [38] Sieber JR, McInerney MJ, Gunsalus RP. Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. *Annu Rev Microbiol.* 2012;66(1):429–452.



- [39] Fang HHP, Zhang T, Eds. *Anaerobic biotechnology: environmental protection and resource recovery*. London: Imperial College Press; 2015.
- [40] Van Dijk EL, Auger H, Jaszczyszyn Y, et al. Ten years of next-generation sequencing technology. *Trends Genet*. 2014;30(9):418–426.
- [41] Wexler M, Bond PL, Richardson DJ, et al. A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environ Microbiol*. 2005;7(12):1917–1926.
- [42] Martín HG, Ivanova N, Kunin V, et al. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat Biotechnol*. 2006;24(10):1263–1269.
- [43] Krause L, Diaz NN, Edwards RA, et al. Taxonomic composition and gene content of a methane producing microbial community isolated from a biogas reactor. *J Biotechnol*. 2008;136(1):91–101.
- [44] Meyerdierks A, Kube M, Kostadinov I, et al. Metagenome and mRNA expression analyses of anaerobic methanotrophic archaea of the ANME-1 group. *Environ Microbiol*. 2010;12(2):422–439.
- [45] Yu K, Zhang T. Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge. *PLoS One*. 2012;7(5):e38183.
- [46] Xia Y, Wang Y, Fang HH, et al. Thermophilic microbial cellulose decomposition and methanogenesis pathways recharacterized by metatranscriptomic and metagenomic analysis. *Sci Rep*. 2014;4:6708–6716.
- [47] Roesch LFW, Fulthorpe RR, Riva A, et al. Pyrosequencing enumerates and contrasts soil microbial diversity. *Isme J*. 2007;1:283–290.
- [48] Hingamp P, Grimsley N, Acinas SG, et al. Exploring nucleocytoplasmic large DNA viruses in tara oceans microbial metagenomes. *Isme J*. 2013;7:1678–1695.
- [49] de Vargas C, Audic S, Henry N, et al. Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science*. 2015;348:1261605.
- [50] DeLong EF. Microbial community genomics in the ocean. *Nat Rev Microbiol*. 2005;3:459–469.
- [51] Karsenti E, Acinas SG, Bork P, et al. A holistic approach to marine eco-systems biology. *PLoS Biol*. 2011;9:e1001177.
- [52] Ranjard L, Dequiedt S, Chemidlin P-BN, et al. Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. *Nat Commun*. 2013;4:1434.
- [53] Megharaj M, Ramakrishnan B, Venkateswarlu K, et al. Bioremediation approaches for organic pollutants: a critical perspective. *Environ Int*. 2011;37:1362–1375.
- [54] Saxena G, Bharagava RN. Ram Chandra: advances in biodegradation and bioremediation of industrial waste. *Clean Technol Environ Policy*. 2016;18:979–980.
- [55] Saxena G, Bharagava RN. Organic and inorganic pollutants in industrial wastes, their ecotoxicological effects, health hazards and bioremediation approaches. In: Bharagava RN, editor. *Environmental pollutants and their bioremediation approaches*. Boca Raton, Florida USA: CRC Press, Taylor & Francis Group; 2017. p. 23–56.
- [56] Goutam SP, Saxena G, Singh V, et al. Green synthesis of TiO<sub>2</sub> nanoparticles using leaf extract of *Jatropha curcas* L. for photocatalytic degradation of tannery wastewater. *Chem Eng J*. 2018;336:386–396.
- [57] Maszenan AM, Liu Y, Ng WJ. Bioremediation of wastewaters with recalcitrant organic compounds and metals by aerobic granules. *Biotechnol Adv*. 2011;29:111–123.
- [58] Saxena G, Bharagava RN. Persistent organic pollutants and bacterial communities present during the treatment of tannery wastewater. In: Chandra R, editor. *Environmental waste management*. Boca Raton, FL: CRC Press, Taylor and Francis Group; 2015. p. 217–247.
- [59] Neufeld JD, Dumont MG, Vohra J, et al. Methodological considerations for the use of stable isotope probing in microbial ecology. *Microb Ecol*. 2007;53(3):435–442.
- [60] Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet*. 2010;42:1118–1125.
- [61] Anderson CA, Boucher G, Lees CW, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet*. 2011;43:246–252.
- [62] Lou J, Liu M, Gu J, et al. Metagenomic sequencing reveals microbial gene catalogue of phosphinothricin-utilized soils in South China. *Gene*. 2019;711:143942.
- [63] Sarsaiya S, Jain A, Jia Q, et al. Molecular identification of endophytic fungi and their pathogenicity evaluation against *dendrobium nobile* and *dendrobium officinale*. *Int J Mol Sci*. 2020;21(1):316.
- [64] Sarsaiya S, Shi J, Chen J. A comprehensive review on fungal endophytes and its dynamics on orchidaceae plants: current research, challenges, and future possibilities. *Bioengineered*. 2019a;10(1):316–334.
- [65] Sarsaiya S, Jia Q, Fan X, et al. First report of leaf black circular spots on *dendrobium nobile* caused by *Trichoderma longibrachiatum* in Guizhou Province, China. *Plant Dis*. 2019b;103(12):3275.
- [66] Sarsaiya S, Shi J, Chen J. Bioengineering tools for the production of pharmaceuticals: current perspective and future outlook. *Bioengineered*. 2019c;10(1):469–492.
- [67] Wei L, Zhuang J-L, Zhou Y-Y, et al. Metagenomics reveals microbial community differences lead to differential nitrate production in anammox reactors with differing nitrogen loading rates. *Water Res*. 2020;169:115279.
- [68] Thoendel M, Jeraldo PR, Greenwood-Quaintance KE, et al. Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing. *J Microbiol Methods*. 2016;127:141–145.
- [69] Deng Y, Ruan Y, Ma B, et al. Multi-omics analysis reveals niche and fitness differences in typical denitrification microbial aggregations. *Environ Int*. 2019a;132:105085.

- [70] Deng Y, Xu X, Yin X, et al. Effect of stock density on the microbial community in biofloc water and pacific white shrimp (*Litopenaeus vannamei*) gut microbiota. *Appl Microbiol Biotechnol*. 2019b;103:4241–4252.
- [71] Li F, Neves AL, Ghoshal B. Symposium review: mining metagenomic and metatranscriptomic data for clues about microbial metabolic functions in ruminants1. *J Dairy Sci*. 2018;101(6):5605–5618.
- [72] Falk N, Reid T, Skoyles A, et al. Microbial metatranscriptomic investigations across contaminant gradients of the Detroit River. *SciTotal Environ*. 2019;690:121–131.
- [73] Duru IC, Laine P, Andreevskaya M, et al. Metagenomic and metatranscriptomic analysis of the microbial community in swiss-type maasdam cheese during ripening. *Int J Food Microbiol*. 2018;281:10–22.
- [74] Liyuan M, Hongmei W, Jiangjun W, et al. Metatranscriptomics reveals microbial adaptation and resistance to extreme environment coupling with bioleaching performance. *Bioresour Technol*. 2019;280:9–17.
- [75] Lim YW, Schmieder R, Haynes M, et al. Metagenomics and metatranscriptomics: windows on CF-associated viral and microbial communities. *J Cyst Fibros*. 2013;12(2):154–164.
- [76] Bei Q, Moser G, Wu X, et al. Metatranscriptomics reveals climate change effects on the rhizosphere microbiomes in European grassland. *Soil Biol Biochem*. 2019;138:107604.
- [77] Ballesteros GI, Torres-Díaz C, Bravo LA, et al. In silico analysis of metatranscriptomic data from the antarctic vascular plant *colobanthus quitensis*: responses to a global warming scenario through changes in fungal gene expression levels. *Fungal Ecol*. 2020;43:100873.
- [78] Jain A, Sarsaiya S, Wu Q, et al. New insights and rethinking of cinnabar for chemical and its pharmacological dynamics. *Bioengineered*. 2019a;10(1):353–364.
- [79] Jain A, Sarsaiya S, Wu Q, et al. A review of plant leaf fungal diseases and its environment speciation. *Bioengineered*. 2019b;10(1):409–424.
- [80] Hernández-Jarguín A, Díaz-Sánchez S, Villar M, et al. Integrated metatranscriptomics and metaproteomics for the characterization of bacterial microbiota in unfed ixodes ricinus. *Ticks Tick-Borne Dis*. 2018;9(5):1241–1251.
- [81] Lin YW, Tuan NN, Huang SL. Metaproteomic analysis of the microbial community present in a thermophilic swine manure digester to allow functional characterization: a case study. *Int Biodeter Biodegr*. 2016;115:64–73.
- [82] Liu D, Keiblinger KM, Schindlbacher A, et al. Microbial functionality as affected by experimental warming of a temperate mountain forest soil—A metaproteomics survey. *Appl Soil Ecol*. 2017;117–118:196–202.
- [83] Isaac NI, Philippe D, Nicholas A, et al. Metaproteomics of the human gut microbiota: challenges and contributions to other OMICS. *Clin Mass Spectrom*. 2019;14(A):18–30.
- [84] Beyter D, Lin MS, Yu Y, et al. ProteoStorm: an ultrafast metaproteomics database search framework. *Cell Syst*. 2018;7(4):463–467.
- [85] Renu, Gupta SK, Rai AK, et al. Metaproteomic data of maize rhizosphere for deciphering functional diversity. *Data Brief*. 2019;27:104574.
- [86] Mesuere B, Van der Jeugt F, Willems T, et al. High-throughput metaproteomics data analysis with unipept: a tutorial. *J Proteomics*. 2018;171:11–22.
- [87] Zaouri N, Jumat MR, Cheema T, et al. Metagenomics-based evaluation of groundwater microbial profiles in response to treated wastewater discharge. *Environ Res*. 2020;180:108835.
- [88] Zhang L, Loh KC, Lim JW, et al. Bioinformatics analysis of metagenomics data of biogas-producing microbial communities in anaerobic digesters: a review. *Renew Sust Energ Rev*. 2019;100:110–126.
- [89] Qi M, Huang H, Zhang Y, et al. Novel tetrahydrofuran (THF) degradation-associated genes and cooperation patterns of a THF-degrading microbial community as revealed by metagenomic. *Chemosphere*. 2019;231:173–183.
- [90] Motro Y, Moran-Gilad J. Microbial metagenomics mock scenario-based sample simulation (M3S3). *Clin Microbiol Infect*. 2018;24(3):308.1–308.4.
- [91] Li J, Du P, Ye AY, et al. GPA: a microbial genetic polymorphisms assignments tool in metagenomic analysis by Bayesian estimation. *Genomics Proteomics Bioinf*. 2019;17(1):106–117.
- [92] Mei R, Liu WT. Quantifying the contribution of microbial immigration in engineered water systems. *Microbiome*. 2019;7:144.
- [93] Cira NJ, Pearce MT, Quake SR. Neutral and selective dynamics in a synthetic microbial community. *Proc Natl Acad Sci*. 2018;115(42):E9842–8.
- [94] Fargione J, Brown CS, Tilman D. Community assembly and invasion: an experimental test of neutral versus niche processes. *Proc Natl Acad Sci*. 2003;100(15):8916–8920.
- [95] Hubbell SP. The unified neutral theory of biodiversity and biogeography (MPB-32) Princeton: Princeton University Press;2001, p. 4. Bell G. The distribution of abundance in neutral communities. *Am Nat*. 2000; 155 (5):606–17.
- [96] MacArthur RH, Wilson EO. An equilibrium theory of insular zoogeography. *Evolution*. 1963;17(4):373–387.
- [97] Schneider F. Dispersal and migration. *Annu Rev Entomol*. 1962;7(1):223–242.
- [98] Leibold MA, Holyoak M, Mouquet N, et al. The meta-community concept: a framework for multi-scale community ecology. *Ecol Lett*. 2004;7(7):601–613.
- [99] Vellend BM. Conceptual synthesis in community ecology. *Q Rev Biol*. 2010;85(2):183–206.
- [100] Konopka A, Lindemann S, Fredrickson J. Dynamics in microbial communities: unraveling mechanisms to identify principles. *Isme J*. 2015;9(7):1488–1495.
- [101] Stegen JC, Lin X, Fredrickson JK, et al. Quantifying community assembly processes and identifying features that impose them. *Isme J*. 2013;7(11):2069–2079.
- [102] Fodelianakis S, Lorz A, Valenzuela-Cuevas A, et al. Dispersal homogenizes communities via immigration even at low rates in a simplified synthetic bacterial metacommunity. *Nat Commun*. 2019;10(1):1314.

- [103] Frigon D, Wells G. Microbial immigration in wastewater treatment systems: analytical considerations and process implications. *Curr Opin Biotechnol.* 2019;57:151–159.
- [104] Gilpin ME, Diamond JM. Calculation of immigration and extinction curves from the species-area-distance relation. *Proc Natl Acad Sci.* 1976;73(11):4130–4134.
- [105] Hervàs A, Camarero L, Reche I, et al. Viability and potential for immigration of airborne bacteria from Africa that reach high mountain lakes in Europe. *Environ Microbiol.* 2009;11(6):1612–1623.
- [106] Lindström ES, Bergström A-K. Influence of inlet bacteria on bacterioplankton assemblage composition in lakes of different hydraulic retention time. *Limnol Oceanogr.* 2004;49(1):125–136.
- [107] Kisand V, Andersson N, Wikner J. Bacterial freshwater species successfully immigrate to the brackish water environment in the northern Baltic. *Limnol Oceanogr.* 2005;50(3):945–956.
- [108] Pinto AJ, Xi C, Raskin L. Bacterial community structure in the drinking water microbiome is governed by filtration processes. *Environ Sci Technol.* 2012;46(16):8851–8859.
- [109] Kotlarz N, Rockey N, Olson TM, et al. Biofilms in full-scale drinking water ozone contactors contribute viable bacteria to ozonated water. *Environ Sci Technol.* 2018;52(5):2618–2628.
- [110] Zhang Y, Oh S, Liu WT. Impact of drinking water treatment and distribution on the microbiome continuum: an ecological disturbance's perspective. *Environ Microbiol.* 2017;19(8):3163–3174.
- [111] Lautenschlager K, Hwang C, Ling F, et al. Abundance and composition of indigenous bacterial communities in a multi-step biofiltration-based drinking water treatment plant. *Water Res.* 2014;62:40–52.
- [112] Ling F, Hwang C, LeChevallier MW, et al. Core-satellite populations and seasonality of water meter biofilms in a metropolitan drinking water distribution system. *Isme J.* 2015;10(3):582–595.
- [113] Ling F, Whitaker R, LeChevallier MW, et al. Drinking water microbiome assembly induced by water stagnation. *Isme J.* 2018;12(6):1520–1531.
- [114] Lee S-H, Kang H-J, Park H-D. Influence of influent wastewater communities on temporal variation of activated sludge communities. *Water Res.* 2015;73:132–144.
- [115] Wells GF, Wu CH, Piceno YM, et al. Microbial biogeography across a full-scale wastewater treatment plant transect: evidence for immigration between coupled processes. *Appl Microbiol Biotechnol.* 2014;98(10):4723–4736.
- [116] Ju F, Lau F, Zhang T. Linking microbial community, environmental variables, and methanogenesis in anaerobic biogas digesters of chemically enhanced primary treatment sludge. *Environ Sci Technol.* 2017;51(7):3982–3992.
- [117] Price JR, Ledford SH, Ryan MO, et al. Wastewater treatment plant effluent introduces recoverable shifts in microbial community composition in receiving streams. *Sci Total Environ.* 2018;613:1104–1116.
- [118] Zhang S-Y, Tsementzi D, Hatt JK, et al. Intensive allochthonous inputs along the Ganges River and their effect on microbial community composition and dynamics. *Environ Microbiol.* 2018;21(1):182–196.
- [119] Ounit R, Wanamaker S, Close TJ, et al. CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. *BMC Genomics.* 2015;16:236.
- [120] Wood DE, Salzberg SL. Kraken: ultrafast metagenomics sequence classification using exact alignments. *Genome Biol.* 2014;15:R46.
- [121] Morgulis A, Coulouris G, Raytselis Y, et al. Database indexing for production MegaBLAST searches. *Bioinformatics.* 2008;24:1757–1764.
- [122] Brinda K, Salikhov K, Pignotti S, et al. karel-brinda/prophyle: ProPhyle 2017;0.3.1.0. (version 0.3.1.0). Zenodo.
- [123] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2015;12:59–60.
- [124] Truong DT, Franzosa EA, Tickle TL, et al. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat Methods.* 2015;12:902–903.
- [125] Milanese A, Mende DR, Paoli L, et al. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nature Commun.* 2019;10:1014.
- [126] Nacke H, Thürmer A, Wollherr A, et al. Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. *PLoS ONE.* 2011;6(2):e17000.
- [127] Trivedi P, Anderson IC, Singh BK. Microbial modulators of soil carbon storage: integrating genomic and metabolic knowledge for global prediction. *Trends Microbiol.* 2013;21:641–651.
- [128] Zeng Q, Dong Y, An S. Bacterial community response to soil along a latitudinal and vegetation gradient on loess plateau, China. *PLoS ONE.* 2016;11(4):e0152894.
- [129] Yun Y, Wang H, Man B, et al. The relationship between pH and bacterial communities in a single karst ecosystem and its implication for soil acidification. *Front Microbiol.* 2016;7:1955.
- [130] Liu Y, Zhang J, Zhao L, et al. Spatial distribution of bacterial communities in high-altitude freshwater wetland sediment. *Limnology.* 2014;15:249–256.
- [131] Sheng P, Yu Y, Zhang G, et al. Bacterial diversity and distribution in seven different estuarine sediments of Poyang Lake, China. *Environ Earth Sci.* 2016;75:479.
- [132] Canfora L, Bacci G, Pinzari F, et al. Salinity and bacterial diversity: to what extent does the concentration of salt affect the bacterial community in a saline soil? *PLoS ONE.* 2014;9(9):e106662.

- [133] Wolinska A, Kuzniar A, Zielenkiewicz U, et al. Metagenomic analysis of some potential nitrogen-fixing bacteria in arable soils at different formation processes. *Microb Ecol.* [2017](#);73:162–176.
- [134] Yashiro E, Pinto-Figueroa E, Buri A, et al. Local environmental factors drive divergent grassland soil bacterial communities in the Western Swiss Alps. *Appl Environ Microbiol.* [2016](#);82(21):6303–6316.
- [135] Zhihua L, Ling D, Chi Z, et al. Metagenomic insights into the changes in microbial community and antimicrobial resistance genes associated with different salt content of red pepper (*Capsicum annuum* L.) sauce. *Food Microbiol.* [2020](#);85::103295.

RETRACTED