Review Article

Current Research and Applications of Meta-omics Stratagems in Bioremediation: A Bird's-Eye View

Arockiyajainmary Michealsamy[*1] iD **, Lokesh Thangamani**[1] iD **, Gowdham Manivel**[1]**, Praveen Kumar**[1]**, Shobana Sundar**[1] iD **,**
Shanmughavel Piramanayagam[1]**, Jeyakumar Natarajan**[2]

[1]DBT Computational Biology Lab, Bharathiar University, Coimbatore, India
[2]Department of Bioinformatics, Bharathiar University, Coimbatore, Tamilnadu, India

Corresponding Author: Arockiyajainmary Michealsamy, PhD, Research Scholar, DBT Computational Biology Lab, Bharathiar University, Coimbatore, India. Tel: +91-8870384834, Email: arockiyajainmary13@gmail.com

Abstract
Microorganisms are ubiquitous in nature. They are found across diverse biosphere and greatly vary among them. Several beneficial microbes engaging in degradation and their services to the ecosystem have not been fully explored. This tangled module could be resolved by Meta 'omics' approach and analysing their molecular interaction with the environment. In our day-to-day life, human beings are exposed to various xenobiotics in the form of drugs/pharmaceuticals, pesticides, artificially flavoured food and beverages. Newer diseases are also emerging due to environmental pollutants. Bioremediation offers an effective way to resilient our fragile planet. Next-generation sequencing (NGS) became an ultimate technique to unravel the significance of the microbiome in remediating polluted lands and sludges. Integrating the meta omics data would open new perspectives in the clean-up of toxic contaminants from our environment. Through this review, we attempted to explore the potential of meta-omics approaches in deciphering the eavesdropping of complex microbes. This review provides novel insights into the Meta-omics techniques that aid in the field of bioremediation.
Keywords: Microbiota, Meta-omics, NGS, Pollutants, Biodegradation, Bioremediation
Citation: Michealsamy A, Thangamani L, Manivel G, Kumar P, Sundar S, Piramanayagam S, et al. current research and applications of meta-omics stratagems in bioremediation: a bird's-eye view. J Appl Biotechnol Rep. 2021;8(2):109-115. doi:10.30491/JABR.2020.237662.1248.

Introduction

Microorganisms are omnipresent. Micro biota comprises a diverse range of archaea, bacteria, fungi, protists, and viruses. They are found in nook and cranny of environmental niches such as soil, air, oceans, hot springs etc. and live as symbionts with the living host. In humans, a tremendous abundance of microcosm is found in the gut, skin, urogenital and oral cavity.[1] They have a strong dependence on factors such as diet, exposure to xenobiotics, anthropogenic and ecological stressors. Through anthropogenic activities, various noxious chemicals are released into the environment and they persist for several decades.[2] Extremophilic bacteria are of prime significance in eradicating those organo-nitro compounds.[3-5] Soil microorganisms have the potential to degrade environmental contaminants and thereby clean up the ecosystem. They share functional genes and their products, produce metabolites and signalling molecules thereby conferring many beneficial services to the host and biological system.[6] It is essential to decipher those signal-based communications that take place between microbial communities in degrading anthropogenic substances. Next-generation sequencing (NGS) turned into a decisive method to unwind the noteworthiness of microbes in the remediation of contaminated grounds and mucks.

The performance and efficacy of NGS technologies were progressed. Fostering advances in meta-omics techniques provide valuable information not just about the characteristics of the microbiome but also reveal the interactions of the microbial community with the host, their environment, and the neighbouring organisms.[7] This review provides deeper insights into the meta-omics application in the field of bioremediation. It has also dealt with new data analytics and computational resources for metagenomics research.

In recent years, metagenomics is a fast-growing and diverse field in biological sciences directed towards acquiring information on genomes of entire microbial networks present in environmental samples, discarding the cultivation bias.[8] It also computes the functional profile by focusing on gene content and utilizing the available functional annotations of the corresponding proteins from the databases. Meta-taxonomics encourages us to allocate the taxonomic profile of a microbial community. Metatranscriptomics involves the investigation of transcripts (mRNA) isolated directly from the environment or microcosms. This method would reach beyond the microbial community's genomic potential, and associate more directly the taxonomic makeup of the community to its *in situ* activity (function), via profiling most abundant transcripts and

correlating them with explicit environmental conditions.[9] For large-scale metatranscriptomics experiments, the NGS technologies are exclusively alluring for transcript analysis. It represents the next logical step in the meta-omic approach. Metabolomics offers ample analysis of all metabolites in an environmental sample.[10] The small molecules directly released by the organism into the environment are identified and evaluated. Several tools and standalone pipelines have been created to analyse this information. By integrating these omics datasets, we would be furnished with a complete picture from genotype to phenotype (Figure 1).

Computational Resources for Data Analysis

Bioinformatics is a boon to the research society from all perspectives. In this genomic era, progress in metagenomic sequencing and data analysis contribute a technical drift in bioremediation developments. Bioremediation is an effective strategy in environmental clean-up by microbial action.[11,12] The microorganisms utilize those accessible organic compounds as their energy supply. The crucial function of the microbes is to convert them into mineral constituents which could serve as energy sources for them. In earlier days, it was difficult for us to get a vivid picture of global microbial communities occurring in a specific habitat. The practical application of debasement of environmental toxins by beneficial microbial flora is still in the cradle stage because of the challenges in their cultivation bias. With the arrival of high throughput NGS technologies, it got conceivable to discover those microbes even which are not amenable to cultivation. They offer an effective way to discover and analyse the microorganisms by the implementation of newer tools and software and go beyond routine analyses. The more profound comprehension of xenobiotic degradation pathways, enzymes, and knowledge of structural properties of pollutants supports the swot-up bioremediation process.[13,14]

All living organisms are subjected to numerous pollutants, drugs, microplastics,[15,16] and organic polymers. Humans are probably to be more exposed. They distort our health and lead to various neurodegenerative diseases and disorders.[17] Lu et al[18] conducted an experimental investigation in mice and proved that polystyrene induces hepatic lipid metabolism disorder and gut microbiota dysbiosis. Studies have explored microbial networks through the metagenome sequencing of the gut, and the faecal samples of human beings. The metabonomic and metabolomic contemplates investigating the interaction of toxins in the human gut cells. The gut microbiota varies with races, demography, age, and among the sound and ailing individuals.[19] Environmental pollutants disrupt the normal metabolic function of the human body.[20] In addition to hepatic enzymes, the secondary metabolites of the human gut microbiome play a pivotal role in metabolizing these xenobiotics.[21] Notably, the plethora of gut microbiota in xenobiotic degradation is still not fully explored.[22] The integrative meta-omic research would serve as an ultimate way to unravel the microbial communities found in the specific environment and their interactive mechanism.

Multi-omic studies involve meta-genomics, metataxonomics, meta-transcriptomics, meta-proteomics, metabolomics, and metabolomics studies. The innovative advancements in the NGS era prompted the invention of many conventional sequencing platforms.[23] Prior metagenomic studies include the clone-based library preparation in blend with sanger sequencing. Direct sequencing of the environmental genome by shotgun amplicon sequencing was developed. Sequencing could be done by NGS frameworks such as SOLiD/Ion Torrent, Genome Analyzer, GS FLX Titanium, Nanopore. A microbial community occupying a well-defined habitat could be explored through whole-metagenome sequencing.

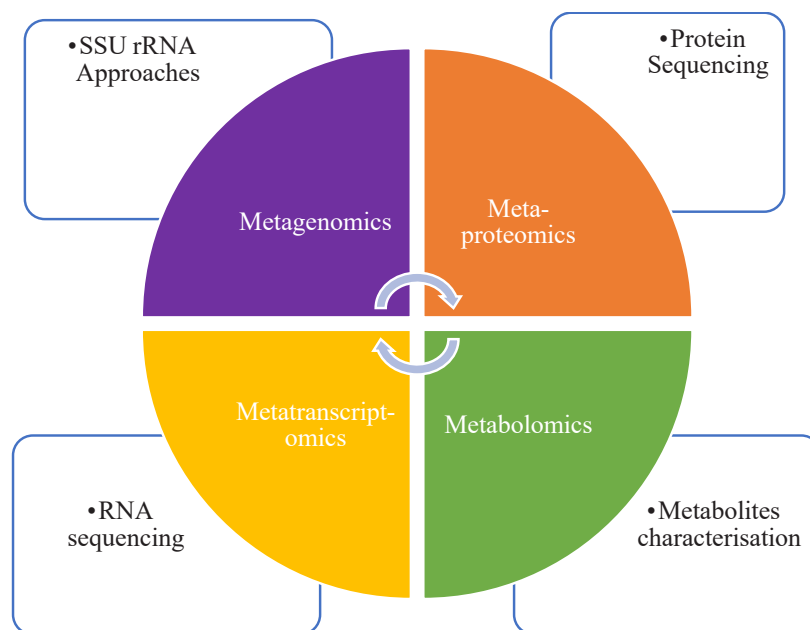Nowadays, various strategies have been applied to gather

Figure 1. Integration of Meta-'omics' Approach.

complete information on microbiome. For instance, amplicon sequencing, whole-genome shotgun (metagenome) analysis, and whole-transcriptome shotgun analysis (metatranscriptome) analysis. The whole-transcriptomic profile depicts the gene expression patterns of the microbiota. The glimpse of gene expression is inferred through metatranscriptome and metaproteomic studies. Gene sequences of 16S rRNA, 18S rRNA, ITS, and COI are amplified to distinguish microcosms in the environmental samples. This environmental tag sequencing determines their taxonomic profile and phylogenetic structure. The more defined experiment aimed to characterize the hypervariable regions flanked by the conserved regions. Polymerase chain reaction (PCR) enhancement of one or more hypervariable regions is done by using universal primer, resulting in amplicons of targeted conserved regions. The NGS data analysis involves read pre-processing, data analysis and annotation, and binning. The read quality control, base calling, adapter trimming, k-mer counting were done. The raw reads were then assembled and processed. From the metagenomic contigs and singletons sequences, the complete or partial gene sequences can be predicted. Binning was performed to identify the microbial strain. Taxonomic binning is the most crucial step in the reconstruction of metagenomic assemblies. An efficient binning relies on sufficient read coverage, sequence similarity, and sequence composition. Bins represent the clade of the population community. Binning requires longer contiguous sequences or scaffolds from the assembled genomes. They were assigned based on two methods, namely, taxonomy dependent and taxonomy independent. The ribosomal gene-based databases viz., RDP-II,[24] GreenGenes,[25] OTT, SILVA,[26] PR2 database,[27] and ExTaxon[28] were built on a phylogenetic order. Evolutionary placement methods, such as pplacer,[29] EPA/RaxML,[30] APPLES,[31] and SEPP[32] aligned to a reference tree. The other strategy is a composition-based approach that utilizes a fragment-wide GC composition, codon usage, and k-mers. For recognizing the fungal communities, the internal transcribed spacers 1 and 2 (ITS1 and ITS2) with flanking rDNA operons were employed. Generally, the UNITE and ITSoneDB database is preferred as a fungal reference database in metagenomic studies of environmental fungal communities. The functional profile of the microbial flora could be assigned through homology-based, context-based, and/or motif and domain-based similarity search. The homology-based function prediction is performed through BLAST, BLAT, DIAMOND, HMMER search tools, and publicly available resources including Pfam, TIGRfam, KEGG, SEED, and STRINGS, etc. Many databases and web servers are available to investigate the metagenomes such as metagenomics RAST server,[33] IMG/M,[34] QIIME 2,[35] Phymm,[36] METAREP,[37] CoMet,[38] METAGENassist,[39] PhyloPytiaS,[40] WebMGA,[41] MOCAT2,[42] MGnify[43] etc. The InterProScan database search results in motifs and fingerprints in the metagenomic assembled sequences. Dedicated gene prediction tools particularly developed for metagenomics, namely, MetaGene, MetaGeneAnnotator, and FragGeneScan. Mande et al.[44] have stated that, alignment-

based binning manner exhibit higher binning accuracy and specificity. MetaVelvet, an extension of Velvet assembler to de novo metagenome assembly from short sequence reads.[45] Meta-IDBA de novo assembler performs local assembly to fill gaps and iteratively removes short inaccurate contigs. DNA metabarcoding and metagenomic approaches were used to assign the taxonomy of the microbial communities. Variants of BLAST such as BlastX, BlastP, PSI-Blast, and RPS-Blast were preferred. The gene-centric metagenomic analysis was employed to explore the functional profile of microbial communities. For instance, the nitrogen-fixing bacteria could be explored by targeting the Nitrogen-fixing genes i.e., nif genes along with amplifying the 16S rRNA gene marker. Quite a lot of bioinformatic tools and pipelines were being developed to characterize the metagenomic datasets (Table 1). MEGAN, an open-source software, assigns the phylogeny by blasting 16S sequences against the SILVA database. Greengenes, is a 16S reference database available for downloading and browsing. The Ribosomal Database Project (RDP) Classifier, a naive Bayesian placement of 8-mers can accurately classify bacterial 16S rRNA sequences into higher-order taxonomy. MOTHUR relies on the nearest neighbour searching for k-mers. TANGO is a homology-based taxonomic classifier which provides better accuracy than the RDP classifier and MOTHUR. Numerous specialized reference databases are available and it is necessary to choose the updated one and chimera checked database. Mapping-based classification is the most precise method for taxonomic classification which relies on mapping the reads to the reference genome. The pipeline and web servers like SAMSA2,[46] TaxMapper,[47] MetaTrans[48] are available. Taxonomic assignment dependent is alignment-based, composition-based and hybrid. The web resources for the alignment-based method are MG-RAST/CAMERA, MEGAN, Sort-ITEMS, DiScRIBinATE,[49] MARTA,[50] MetaPhyler,[51] CARMA, AMPHORA,[52] MLTreeMap,[53] TreePhyler,[54] pplacer, Papara. The compositional-based taxonomic assignment resources incorporate Phylophytha, NBC Classifier, TACOA,[55] Phymm, RAIPhy,[56] INDUS,[57] ClaMS.[58] SPHINX[59] and PhymmBL[60] uses both composition

Table 1. Web Servers and Software for Metagenomics

Web Servers and Software for Metagenomics	References
MG-RAST	[33]
IMG/M	[34]
QIIME	[35]
PhymmBL	[36]
METAREP	[37]
CoMet	[38]
METAGENassist	[39]
PhyloPythiaS	[40]
WebMGA	[41]
MOCAT	[42]
EBI-MGnify	[43]
Mothur	[73]
TANGO	[74]

and alignment-based algorithms. The hybrid binning exhibits higher specificity. The alignment independent method is performed by TETRA,[61] CompostBin, AbundanceBin,[62] and MetaCluster[63] algorithms. Most of the algorithms use hidden Markov model or BLAST/BLAT and the BWA read mapping approach. The wide range of '-omics has attempted to biologically characterize the microbiota from the environment of interest. The metagenome phylogenetic placement approach uses mathematical approaches such as Markov models, vector machines, non-negative least squares, or mixture modelling. Calculating the species abundance or richness in a particular logical community provides wide insights into the impact of the microbial flora on that ecosystem.

The functional profiles of microbial biota could be explored by metatranscriptomics (cDNA sequencing). The crucial step in metatranscriptomics relies on the extraction of environmental mRNA. Metatranscriptomic data analysis and annotation convey gene expression profiles, regulatory mechanisms, and complex metabolic interactive pathways. From the differential expression analysis, we can interpret the adaptive mechanism of microbes to a certain environment. This review put forth a comprehensive review of high throughput technologies, bioinformatics resources, and algorithms developed in the past five years.

The Metatranscriptomics shares ample information on the expressed gene of the microbial community which is responsible for biodegrading pollutants in the contaminated sites. A comparative metagenomics approach is employed to discover the association between the microbial communities of two different niches and also among the healthy and diseased individuals. The functional annotation of the assembled contigs grants paramount information on the phenotype of the microbiota. Although a huge number of beneficial microbial flora subsists for the degradation of xenobiotics, its availability and practical application are mere. The ultimate aim of the function-based metagenomic study is to unlock their functional potential through annotation and metabolic pathway inference. A biological pathway represents a series of molecular interactions that lead to certain cellular functions and productivity. The metabolic pathway is designed as a result of manual curation made by researchers in various fields aimed at building a network of genes with relationships such as physical or chemical association, substrate-product link, etc., which have been experimentally proven.

The metabolic pathway investigation has a deep impact on the functional assessment of genes or genome of the organism. The publicly available centralized repositories for metabolic pathways are KEGG and MetaCyc. Describing the biological processes and taxonomic composition of the sample could deliver sufficient knowledge on the unknown target species. The whole metabolite study provides the quantitative description of metabolites and metabolic response to the external stimuli. Metabonomics aims to characterize and quantify all the complex small molecules of living systems. The peptide mass fingerprints produced from MALDI-ToF-MS or LC-ESI-MS were compared against known peptide fragments in databases. The deeper analysis of the metaproteomic study reveals the activity of the microbiome through functional annotation of proteins within the reference databases including eggnog, MetaCys, Pfam, UniProtKB, etc. The iMetaLab, Galaxy-P, and MetaGOmics, MetaProteomeAnalyser were the resources for metaproteomic analysis.

Employing NGS Technology in Bioremediation Studies

Compared to the last decade, there has been a steady increase in metagenomics-based bioremediation studies. Nowadays, microplastics have become a global health concern. It is noxious to human, aquatic and terrestrial animals. Microplastics could be an opportunistic intake in our routine life through the plastic bead in cosmetics, toothpaste, and personal care products. The heavy metals could be accumulated through the food chain. These plastic particles have been found in the gastrointestinal tract of birds, animals, and marine organisms. Microplastics consumed by humans and animals produce lethal effects and disrupt the normal functioning of the cell. However, the consequence of the microplastics on the gut microbiota is still not obvious. Understanding the microbial response to the pollutants can also help improve their metabolic potential to degrade xenobiotics. Microorganisms drive biogeochemical cycles to degrade plastics procured in marine environments.[64] Kachienga et al,[65] assessed the microbial diversity in petroleum-contaminated aquifers through whole genome sequencing and found that most of them were protozoan (62.04%) followed by fungi (24.49%). *Pseudomonas alk*B gene appeared to be abundantly expressed in diesel contaminated soils.[66] *Geobacter* species predominately occupying uranium-contaminated sites were studied and gene expressions were quantitatively analysed by qRT-PCR.[67] Some putative roles of the metagenomic sequences could be inferred by identifying the target sites. The evolutionary process of carbon metabolism is studied from the metabolic flux analysis of *Shewanella* spp.[68] Orellana et al,[69] conducted an *in silico* metagenomic study on the abundance of NosZ genes involved in the nitrogen cycle in the soil ecosystem other than agricultural soil metagenomes. The Human Microbiome Project (HMP) or integrative HMP enabled by the National Institute of Health generates all the available resources for the study of microbial communities inhabiting the human body and their impact on health and disease. There is a strong correlation between the gut microbiota and the disease.

Future Aspects and Challenges

The evolutions of HTS from 2G to 4G sequencers overcome many challenges. The second-generation sequencing suffers from short read length and enormous PCR error rate. The 2G NGS platforms are 454, illumina, SOLiD, and Ion Torrent. The 3G NGS platforms Oxford Nanopore and Pacific Biosciences use Single-molecule Real-time Sequencing Technology. Nanopore sequencer, PacBio RS sequencer generates up to 1 Mb long reads while the illumina sequencers produce a maximum of 2 × 300 bp length. The common issues in metagenome sequencing could be overwhelmed by increasing the genome coverage or switching to long-read technology.

Analysis of those reads obtained from sequencers represent the compositional properties of their source genome. Long read sequences and accurate binning would anticipate successful metagenomic sequencing. Illumina sequencer is more favoured than the 454 pyro sequencers due to its coverage efficiency. However, it is cost-effective. Metagenomic assembly is complicated due to the presence of intergenomic and intragenomic repeats. Also, finding true SNPs in draft genomes has drastically increased. A high frequency of polymorphism leads to complexity in metagenome assembly. For instance, Kurokawa et al,[70] investigated the human gut microbiomes of 13 healthy individuals in Japan and found a high variation in individual taxonomic and gene composition. The disadvantage of employing Pfam hits is that they occur less frequently than BLAST hits. Teeling and Glocknwer[71] address the challenges and opportunities in metagenome analysis. Recently, a signature-based method has been developed for the fast-taxonomic profiling of metagenomes. Many unexplored microbial diversities could be revealed through metagenomic study. This approach is very sensitive which needs attention on sample preparation and sequencing method. They potentially affect the functional analysis of metagenomic datasets. The length of the reads determines the quality of the sequence analysis. The short-read metagenomic probes possess an average length of 200-300 bp. A well-developed strategy that involves long-reads of metagenomic data provides more insights into the functional analysis. The rapid arising of Stable Isotope Probing serves its application in identifying active members of the microbial community and genes associated with the biodegradative process.[72] Improvements in siRNA research might evolve in the future. The predicament in metagenome sequencing is incomplete coverage, numerous raw reads generated by the sequencers, producing short-read length, massive species diversity etc. They actually produce noise. Incomplete coverage results in shorter contigs which leads to difficulty in taxonomic binning or genome assembly. The metagenome datasets are much complicated and greatly vary with the whole genome studies due to the presence of unambiguous data. The most crucial step is the assembly of raw metagenomic reads. Some assemblers yield chimeric contigs. Dedicated metagenomic assemblers address this issue by improving the quality of the assembling methods. De novo assembler such as Genovo, Meta-IDBA, MetaVelvet, and MAP lowers the risk of chimeric assemblies. MAP handles the longer reads. More importantly, an effective *de novo* metagenomic assembly requires long paired-end sequences. The major challenge in multi-omics studies is the dearth of integrating approaches.

Conclusions

Microbes play an important role in their habitats and are known to be one of the influential factors in the ecosystem. Bioremediating microbes in contaminated environments such as soils contaminated with heavy metals, industrial effluents and sludge, through small molecular signals or the transfer of functional genes effective in bioremediation through horizontal transmission, have a significant ecological role in environmental decontamination. Evaluating the omics dataset provides in-depth insights into the world of microbiomes responsible for bioremediation. These data represent a significant source of genetic information of microbial communities in environmental samples. This review provides insights into current aspects of meta-omic techniques that could emerge as an atlas for researchers.

Authors' Contributions
AM wrote the review. LT, GM, PK, SS supported in writing and SP and JN edited the review.

Conflict of Interest Disclosures
The authors report no conflicts of interest.

References

1. Henke MT, Clardy J. Molecular messages in human microbiota. Science. 2019;366(6471):1309-1310. doi:10.1126/science.aaz4164.

2. Ali H, Khan E, Ilahi I. Environmental chemistry and ecotoxicology of hazardous heavy metals: environmental persistence, toxicity, and bioaccumulation. J Chem. 2019;2019:6730305. doi:10.1155/2019/6730305.

3. Bang C, Dagan T, Deines P, et al. Metaorganisms in extreme environments: do microbes play a role in organismal adaptation? Zoology (Jena). 2018;127:1-19. doi:10.1016/j.zool.2018.02.004.

4. Marvin-Sikkema FD, de Bont JA. Degradation of nitroaromatic compounds by microorganisms. Appl Microbiol Biotechnol. 1994;42(4):499-507. doi:10.1007/bf00173912.

5. Ju KS, Parales RE. Nitroaromatic compounds, from synthesis to biodegradation. Microbiol Mol Biol Rev. 2010;74(2):250-272. doi:10.1128/mmbr.00006-10.

6. Braga RM, Dourado MN, Araújo WL. Microbial interactions: ecology in a molecular perspective. Braz J Microbiol. 2016;47(Suppl 1):86-98. doi:10.1016/j.bjm.2016.10.005.

7. Aguiar-Pulido V, Huang W, Suarez-Ulloa V, Cickovski T, Mathee K, Narasimhan G. Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis. Evol Bioinform Online. 2016;12(Suppl 1):5-16. doi:10.4137/ebo.s36436.

8. Cocolin L, Mataragas M, Bourdichon F, et al. Next generation microbiological risk assessment meta-omics: the next need for integration. Int J Food Microbiol. 2018;287:10-17. doi:10.1016/j.ijfoodmicro.2017.11.008.

9. Carvalhais LC, Dennis PG, Tyson GW, Schenk PM. Application of metatranscriptomics to soil environments. J Microbiol Methods. 2012;91(2):246-251. doi:10.1016/j.mimet.2012.08.011.

10. Fiehn O. Metabolomics--the link between genotypes and phenotypes. Plant Mol Biol. 2002;48(1-2):155-171.

11. Dvořák P, Nikel PI, Damborský J, de Lorenzo V. Bioremediation 3.0: engineering pollutant-removing bacteria in the times of systemic biology. Biotechnol Adv. 2017;35(7):845-866. doi:10.1016/j.biotechadv.2017.08.001.

12. Pande V, Pandey SC, Sati D, Pande V, Samant M. Bioremediation: an emerging effective approach towards environment restoration. Environ Sustain. 2020;3(1):91-103. doi:10.1007/s42398-020-00099-w.

13. Ghosal D, Ghosh S, Dutta TK, Ahn Y. Current state of knowledge in microbial degradation of polycyclic aromatic hydrocarbons (PAHs): a review. Front Microbiol. 2016;7:1369. doi:10.3389/fmicb.2016.01369.

14. Malla MA, Dubey A, Yadav S, Kumar A, Hashem A, Abd Allah EF. Understanding and designing the strategies for the microbe-mediated remediation of environmental contaminants using omics approaches. Front Microbiol. 2018;9:1132. doi:10.3389/fmicb.2018.01132.

15. Galloway TS. Micro- and nano-plastics and human health. In: Bergmann M, Gutow L, Klages M, eds. Marine Anthropogenic Litter. Cham: Springer; 2015. p. 343-366. doi:10.1007/978-3-319-16510-3_13.

16. Thompson RC, Moore CJ, vom Saal FS, Swan SH. Plastics, the environment and human health: current consensus and future trends. Philos Trans R Soc Lond B Biol Sci. 2009;364(1526):2153-2166. doi:10.1098/rstb.2009.0053.

17. Smith M, Love DC, Rochman CM, Neff RA. Microplastics in seafood and the implications for human health. Curr Environ Health Rep. 2018;5(3):375-386. doi:10.1007/s40572-018-0206-z.

18. Lu L, Wan Z, Luo T, Fu Z, Jin Y. Polystyrene microplastics induce gut microbiota dysbiosis and hepatic lipid metabolism disorder in mice. Sci Total Environ. 2018;631-632:449-458. doi:10.1016/j.scitotenv.2018.03.051.

19. Valles-Colomer M, Falony G, Darzi Y, et al. The neuroactive potential of the human gut microbiota in quality of life and depression. Nat Microbiol. 2019;4(4):623-632. doi:10.1038/s41564-018-0337-x.

20. Le Magueresse-Battistoni B, Vidal H, Naville D. Environmental pollutants and metabolic disorders: the multi-exposure scenario of life. Front Endocrinol (Lausanne). 2018;9:582. doi:10.3389/fendo.2018.00582.

21. Rajakovich LJ, Balskus EP. Metabolic functions of the human gut microbiota: the role of metalloenzymes. Nat Prod Rep. 2019;36(4):593-625. doi:10.1039/c8np00074c.

22. Thomas AM, Segata N. Multiple levels of the unknown in microbiome research. BMC Biol. 2019;17(1):48. doi:10.1186/s12915-019-0667-z.

23. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. J Appl Genet. 2011;52(4):413-435. doi:10.1007/s13353-011-0057-x.

24. Cole JR, Wang Q, Fish JA, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014;42(Database issue):D633-642. doi:10.1093/nar/gkt1244.

25. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72(7):5069-5072. doi:10.1128/aem.03006-05.

26. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-596. doi:10.1093/nar/gks1219.

27. Guillou L, Bachar D, Audic S, et al. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 2013;41(Database issue):D597-604. doi:10.1093/nar/gks1160.

28. Yoon SH, Ha SM, Kwon S, et al. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol. 2017;67(5):1613-1617. doi:10.1099/ijsem.0.001755.

29. Matsen FA, Kodner RB, Armbrust EV. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics. 2010;11:538. doi:10.1186/1471-2105-11-538.

30. Barbera P, Kozlov AM, Czech L, et al. EPA-ng: massively parallel evolutionary placement of genetic sequences. Syst Biol. 2019;68(2):365-369. doi:10.1093/sysbio/syy054.

31. Balaban M, Sarmashghi S, Mirarab S. APPLES: scalable distance-based phylogenetic placement with or without alignments. Syst Biol. 2020;69(3):566-578. doi:10.1093/sysbio/syz063.

32. Mirarab S, Nguyen N, Warnow T. SEPP: SATé-enabled phylogenetic placement. Pac Symp Biocomput. 2012:247-258. doi:10.1142/9789814366496_0024.

33. Meyer F, Paarmann D, D'Souza M, et al. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 2008;9:386. doi:10.1186/1471-2105-9-386.

34. Chen IA, Chu K, Palaniappan K, et al. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. Nucleic Acids Res. 2019;47(D1):D666-D677. doi:10.1093/nar/gky901.

35. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852-857. doi:10.1038/s41587-019-0209-9.

36. Brady A, Salzberg SL. Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. Nat Methods. 2009;6(9):673-676. doi:10.1038/nmeth.1358 .

37. Goll J, Rusch DB, Tanenbaum DM, et al. METAREP: JCVI metagenomics reports--an open source tool for high-performance comparative metagenomics. Bioinformatics. 2010;26(20):2631-2632. doi:10.1093/bioinformatics/btq455.

38. Lingner T, Asshauer KP, Schreiber F, Meinicke P. CoMet--a web server for comparative functional profiling of metagenomes. Nucleic Acids Res. 2011;39(Web Server issue):W518-523. doi:10.1093/nar/gkr388.

39. Arndt D, Xia J, Liu Y, et al. METAGENassist: a comprehensive web server for comparative metagenomics. Nucleic Acids Res. 2012;40(Web Server issue):W88-95. doi:10.1093/nar/gks497.

40. Patil KR, Roune L, McHardy AC. The PhyloPythiaS web server for taxonomic assignment of metagenome sequences. PLoS One. 2012;7(6):e38581. doi:10.1371/journal.pone.0038581.

41. Wu S, Zhu Z, Fu L, Niu B, Li W. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics. 2011;12:444. doi:10.1186/1471-2164-12-444.

42. Kultima JR, Coelho LP, Forslund K, et al. MOCAT2: a metagenomic assembly, annotation and profiling framework. Bioinformatics. 2016;32(16):2520-2523. doi:10.1093/bioinformatics/btw183.

43. Mitchell AL, Almeida A, Beracochea M, et al. MGnify: the microbiome analysis resource in 2020. Nucleic Acids Res. 2020;48(D1):D570-D578. doi:10.1093/nar/gkz1035.

44. Mande SS, Mohammed MH, Ghosh TS. Classification of metagenomic sequences: methods and challenges. Brief Bioinform. 2012;13(6):669-681. doi:10.1093/bib/bbs054.

45. Sakakibara Y. MetaVelvet: An Extension of Velvet Assembler to de novo Metagenome Assembly from Short Sequence Reads (Metagenomics Informatics Challenges Workshop: 10K Genomes at a Time). Berkeley, CA, United States: DOE Joint Genome Institute (JGI), Lawrence Berkeley National Laboratory (LBNL); 2011. doi:10.4016/37705.01.

46. Westreich ST, Treiber ML, Mills DA, Korf I, Lemay DG. SAMSA2: a standalone metatranscriptome analysis pipeline. BMC Bioinformatics. 2018;19(1):175. doi:10.1186/s12859-018-2189-z.

47. Beisser D, Graupner N, Grossmann L, Timm H, Boenigk J, Rahmann S. TaxMapper: an analysis tool, reference database and workflow for metatranscriptome analysis of eukaryotic microorganisms. BMC Genomics. 2017;18(1):787. doi:10.1186/s12864-017-4168-6.

48. Martinez X, Pozuelo M, Pascal V, et al. MetaTrans: an open-source pipeline for metatranscriptomics. Sci Rep. 2016;6:26447. doi:10.1038/srep26447.

49. Ghosh TS, Monzoorul Haque M, Mande SS. DiScRIBinATE: a rapid method for accurate taxonomic classification of metagenomic sequences. BMC Bioinformatics. 2010;11(Suppl 7):S14. doi:10.1186/1471-2105-11-s7-s14.

50. Horton M, Bodenhausen N, Bergelson J. MARTA: a suite of Java-based tools for assigning taxonomic status to DNA sequences. Bioinformatics. 2010;26(4):568-569. doi:10.1093/bioinformatics/btp682.

51. Liu B, Gibbons T, Ghodsi M, Treangen T, Pop M. Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. BMC Genomics. 2011;12(Suppl 2):S4. doi:10.1186/1471-2164-12-s2-s4.

52. Kerepesi C, Bánky D, Grolmusz V. AmphoraNet: the webserver implementation of the AMPHORA2 metagenomic workflow suite. Gene. 2014;533(2):538-540. doi:10.1016/j.gene.2013.10.015.

53. Stark M, Berger SA, Stamatakis A, von Mering C. MLTreeMap-accurate Maximum Likelihood placement of environmental DNA sequences into taxonomic and functional reference phylogenies. BMC Genomics. 2010;11:461. doi:10.1186/1471-2164-11-461.

54. Schreiber F, Gumrich P, Daniel R, Meinicke P. Treephyler: fast taxonomic profiling of metagenomes. Bioinformatics. 2010;26(7):960-961. doi:10.1093/bioinformatics/btq070.

55. Diaz NN, Krause L, Goesmann A, Niehaus K, Nattkemper TW. TACOA: taxonomic classification of environmental genomic fragments using a kernelized nearest neighbor approach. BMC Bioinformatics. 2009;10:56. doi:10.1186/1471-2105-10-56.

56. Nalbantoglu OU, Way SF, Hinrichs SH, Sayood K. RAIphy: phylogenetic classification of metagenomics samples using iterative refinement of relative abundance index profiles. BMC Bioinformatics. 2011;12:41. doi:10.1186/1471-2105-12-41.

57. Mohammed MH, Ghosh TS, Reddy RM, Reddy CV, Singh NK, Mande SS. INDUS - a composition-based approach for rapid and accurate taxonomic classification of metagenomic sequences. BMC Genomics. 2011;12(Suppl 3):S4. doi:10.1186/1471-2164-12-s3-s4.

58. Pati A, Heath LS, Kyrpides NC, Ivanova N. ClaMS: a classifier for metagenomic sequences. Stand Genomic Sci. 2011;5(2):248-253. doi:10.4056/sigs.2075298.

59. Mohammed MH, Ghosh TS, Singh NK, Mande SS. SPHINX--an algorithm for taxonomic binning of metagenomic sequences. Bioinformatics. 2011;27(1):22-30. doi:10.1093/bioinformatics/btq608.

60. Kelley DR, Salzberg SL. Clustering metagenomic sequences with interpolated Markov models. BMC bioinformatics. 2010;11(1):1-2. doi: 10.1186/1471-2105-11-544

61. Teeling H, Waldmann J, Lombardot T, Bauer M, Glöckner FO. TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. BMC Bioinformatics. 2004;5:163. doi:10.1186/1471-2105-5-163.

62. Wu YW, Ye Y. A novel abundance-based algorithm for binning metagenomic sequences using l-tuples. J Comput Biol. 2011;18(3):523-534. doi:10.1089/cmb.2010.0245.

63. Wang Y, Leung HC, Yiu SM, Chin FY. MetaCluster 4.0: a novel binning algorithm for NGS reads and huge number of species. J Comput Biol. 2012;19(2):241-249. doi:10.1089/cmb.2011.0276.

64. Carreres-Calabuig JA, Rogers KL, Gorokhova E, Posth NR. Micro-by-micro interactions: how microorganisms influence the fate of marine microplastics. Preprints. 2019. doi:10.20944/preprints201910.0125.v1.

65. Kachienga L, Jitendra K, Momba M. Metagenomic profiling for assessing microbial diversity and microbial adaptation to degradation of hydrocarbons in two South African petroleum-contaminated water aquifers. Sci Rep. 2018;8(1):7564. doi:10.1038/s41598-018-25961-0.

66. Yergeau E, Sanschagrin S, Beaumier D, Greer CW. Metagenomic analysis of the bioremediation of diesel-contaminated Canadian high arctic soils. PLoS One. 2012;7(1):e30058. doi:10.1371/journal.pone.0030058.

67. Holmes DE, O'Neil RA, Chavan MA, et al. Transcriptome of *Geobacter uraniireducens* growing in uranium-contaminated subsurface sediments. Isme J. 2009;3(2):216-230. doi:10.1038/ismej.2008.89.

68. Tang YJ, Martin HG, Dehal PS, et al. Metabolic flux analysis of *Shewanella* spp. reveals evolutionary robustness in central carbon metabolism. Biotechnol Bioeng. 2009;102(4):1161-1169. doi:10.1002/bit.22129.

69. Orellana LH, Rodriguez RL, Higgins S, et al. Detecting nitrous oxide reductase (NosZ) genes in soil metagenomes: method development and implications for the nitrogen cycle. mBio. 2014;5(3):e01193-01114. doi:10.1128/mBio.01193-14.

70. Kurokawa K, Itoh T, Kuwahara T, et al. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. DNA Res. 2007;14(4):169-181. doi:10.1093/dnares/dsm018.

71. Teeling H, Glöckner FO. Current opportunities and challenges in microbial metagenome analysis--a bioinformatic perspective. Brief Bioinform. 2012;13(6):728-742. doi:10.1093/bib/bbs039.

72. Uhlik O, Leewis MC, Strejcek M, et al. Stable isotope probing in the metagenomics era: a bridge towards improved bioremediation. Biotechnol Adv. 2013;31(2):154-165. doi:10.1016/j.biotechadv.2012.09.003.