
Metagenomics – A Technological Drift in Bioremediation

Pratap Devarapalli and Ranjith N. Kumavath

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/60749>

Abstract

Nature has its ways of resolving imbalances in its environment and microorganisms are one of the best tools of nature to eliminate toxic pollutants. The process of eliminating pollutants using microbes is termed Bioremediation. Metagenomics is a strategic approach for analysing microbial communities at a genomic level. It is one of the best technological upgradation to bioremediation. Identification and screening of metagenomes from the polluted environments are crucial in a metagenomic study. This chapter emphasizes recent multiple case studies explaining the approaches of metagenomics in bioremediation in different contaminated environments such as soil, water etc. The second section explains different sequences and function-based metagenomic strategies and tools starting from providing a detailed view of metagenomic screening, FACS, and multiple advanced metagenomic sequencing strategies dealing with the prevalent metagenomes in bioremediation and giving a list of different widespread metagenomic organisms and their respective projects. Eventually, we have provided a detailed view of different major bioinformatic tools and datasets most prevalently used in metagenomic data analysis and processing during metagenomic bioremediation.

Keywords: Metagenomics, Bioremediation, Microbial Metagenomes, Bioinformatics, Pollution

1. Introduction

From the day humans started invading this planet, Earth has been crammed with numerous toxic pollutants from multiple sources. Advance scientific technology has given rise to multiple

tools to reduce pollutants in different ways, and bioremediation is considered to be the best way to neutralise polluted environments on Earth [1, 2]. In this genomic era, metagenomic approaches have been developed and are known as effective methods of removing various kinds of pollutants [3, 4]. Metagenomics is a strategic approach of analyzing microbial communities at a genomic level. This provides a glimpse of the microbial community view of "Uncultured Microbiota". Recent studies suggest that microbial communities are the potential alternatives to eliminate toxic contaminants from our environment [5-8]. The term metagenomics was coined by Jo Handelsman et al. in 1998. They have accessed the collective genomes and the biosynthetic machinery of soil microflora during a study of cloning the metagenome [9]. Bioremediation has always been adapting new advances in science and technology for establishing better environments. Compared with the previous years, there has been a gradual increase of interest in metagenomics-based bioremediation studies [10-12]. These studies can prove that metagenomics is one of the best adaptations of bioremediation leading to the establishment of a pure nontoxic environment.

In this chapter, we discussed recent approaches of metagenomics in bioremediation with the help of recent multiple case studies. Preliminarily, we explained the methodology behind metagenomic analysis, starting from the sample screening and ending up with metagenomic analysis with respect to bioremediation. Metagenomic bioremediation reviews and extracts microbial communities applying their extensive biochemical pathways in degrading toxic pollutants. A part of our study aims to emphasize multiple case studies of metagenomic applications on air, water, and soil contaminations. Our analysis provided a topic-specific landscape with respect to metagenomic bioremediation of water contaminations, soil contaminations, and followed by air contaminations. The following part of our study focuses on recently developed sequence and function-based metagenomic strategies to analyze metagenomes from contaminated environments. In addition to this, our study explains the highly prevalent metagenomes derived from metagenomic communities which are also highly capable of degrading contaminations and toxins in the environment. Finally, we provided a landscape view of multiple bioinformatic tools used in the processing and analysis of metagenomic bioremediation data.

2. Applications of metagenomics in bioremediation

Environmental scientists consider metagenomic bioremediation as one of the potential tools to remove contaminants from the environment [13-15]. As cited earlier, recent multiple studies have reported metagenomic approaches in bioremediation. When this was compared with the other approaches of bioremediation, metagenomic bioremediation provided best outcomes with better degrading ratios. The results of a recent study emphasized the potential of metagenomic bacteria derived from petroleum reservoirs [16]. In this study, microbial strains and metagenomic clones have been isolated from petroleum reservoirs, and petroleum degradation abilities were evaluated either individually or in pools using seawater artificial ecosystems. The results showed that metagenomic clones were able to biodegrade up to 94% of phenanthrene and methyl phenanthrenes with rates ranging from 55% to 70% after 21 days

[16]. The authors concluded that bacterial strains and metagenomic clones showed high petroleum-degrading potential.

Metagenomic approaches in bioremediation aid in comprehending the characteristics of bacterial communities in different kinds of contaminated environments. A metaproteogenomic study was carried out on long-term adaptation of bacterial communities in metal-contaminated sediments [17]. The aim of this study was to understand the effect of a long-term metal exposure (110 years) on sediment microbial communities. In this study, the authors selected two freshwater sites differing by one order of magnitude in metal levels. The samples extracted from the two sites were compared by shotgun metaproteogenomics which resulted in a total of 69–118 Mpb of DNA and 943–1241 proteins. The two communities were found to be functionally very similar. However, significant genetic differences were observed for three categories: synthesis of exopolymeric substances, virulence and defense mechanisms, and elements involved in horizontal gene transfer. This study can be considered as a best example of advanced metagenomic approaches applied in bioremediation of different contaminated environments.

3. Metagenomic bioremediation of different contaminations

The environment where human activity abounds is being more polluted and contaminated by different kinds of toxic contaminants [18-20]. The contaminations are diverse and cover almost all sources of life including water, soil, and air which are considered the most important sources of life [21-23]. Metagenomic analysis is applied to multiple kinds of polluted environments primarily soil- and water-contaminated environments [24, 25].

3.1. Metagenomic bioremediation of soil contaminations

Soil contamination is a serious contamination [26, 27] as soil is considered as one of the major sources of life [28]. Compared with other approaches of bioremediation, microbial and environmental researches are more inclined in applying metagenomic approaches to bioremediation [10, 29, 30]. A recent case study discusses the metagenomic analysis of arctic soils contaminated by high concentration of diesel in Canada [31]. As this study was on arctic soils, the objective framed was to trace out microorganisms and their functional genes which are abundant and active during hydrocarbon degradation at cold temperature. In this study, scientists have sequenced the soil metagenome and performed reverse-transcriptase real-time PCR (RT-qPCR) to quantify the expression of several hydrocarbon-degrading genes. *Pseudomonas* species were detected as the most abundant organisms in diesel-contaminated soils at cold environments. RT-qPCR assays confirmed that *Pseudomonas* and *Rhodococcus* species actively expressed hydrocarbon degradation genes in arctic biopile soils. The results of this study indicated that biopile treatment leads to major shifts in soil microbial communities which favors aerobic bacteria to degrade hydrocarbons [31].

3.2. Metagenomic bioremediation of water contaminations

Water pollution has dramatically increased in comparison with the conditions of the 20th century [32, 33]. Metagenomic application in the bioremediation of water contamination is one of the best ways to reduce water contaminations [34-37]. Recent multiple case studies suggest that metagenomic applications have been widely used for the identification and treatment of pollutants and contaminations in the sea, ground water, and drinking water [34-37]. A recent research performed at the Gulf of Mexico beaches precisely talks about the longitudinal metagenomic analysis of water and soil affected by deepwater horizon oil spill [34]. Approximately 7×10^5 cubic meters of crude oil were released into the Gulf of Mexico as a consequence of deepwater horizon drilling rig explosion, where thousands of square miles of the earth's surface were covered in crude oil. During this study, researchers performed high throughput DNA sequencing of close-to-shore water and beach soil samples before and during the appearance of oil in Louisiana and Mississippi. The sequencing results have identified an unusual increase in the human pathogen *Vibrio cholera*, a sharp increase in *Rickettsiales sp.*, and decrease of *Synechococcus sp.* in water samples [34]. In addition, a metagenomic analysis was also performed for the bioremediation of hexavalent chromium-contaminated water that existed in fixed-film bioreactor [38]. This study talks about hexavalent chromium (Cr^{6+}) contamination from a dolomite stone mine in Limpopo Province, South Africa, causing extensive groundwater contamination. To restrict any further negative environmental impact at the site, an effective and sustainable treatment strategy for the removal of up to 6.49 mg/l Cr^{6+} from the groundwater was developed. The microbial community shifted in relative dominance during operation to establish an optimal metal-reducing community, including *Enterobactercloacae*, *Flavobacterium sp.* and *Ralstonia sp.*, which achieved 100% reduction. This study provides a glimpse of effective demonstration of a biological chromium (VI) bioremediation system [38].

4. Metagenomic strategies and tools for bioremediation

Advanced scientific technology has given rise to the advancements in research tools applied in different fields of scientific research [39]. These technologically advanced inventions have driven scientific researchers towards finding out some unrevealed things of nature [40]. Multiple technologies have started getting embedded to metagenomics for a better understanding of biological and life sciences [41]. Thus, in this section, we have discussed recent major metagenomic strategies and tools applied in the process of metagenomic bioremediation.

4.1. Screening of metagenomes from polluted environments

Identification and screening of metagenomes from polluted environments are crucial in a metagenomic study. The microbial community interaction can be detected precisely when metagenomes are finely screened from a contaminated environment. A methodology proposed from a recent study [42] suggested an updated technology of high throughput genetic

screening of a soil metagenomic library. The study was initiated by adding a typical composition of oligonucleotide probes to soil metagenomic DNA for hybridization. The pooled radio-labeled probes were designed to target genes encoding specific enzymes. The soil metagenomic DNA of fosmid clone library were spotted on high-density membranes before the addition of oligonucleotide probes. This next step was followed by affiliation of positive hybridizing spots to the corresponding clones in the library and sequencing of metagenomic inserts.

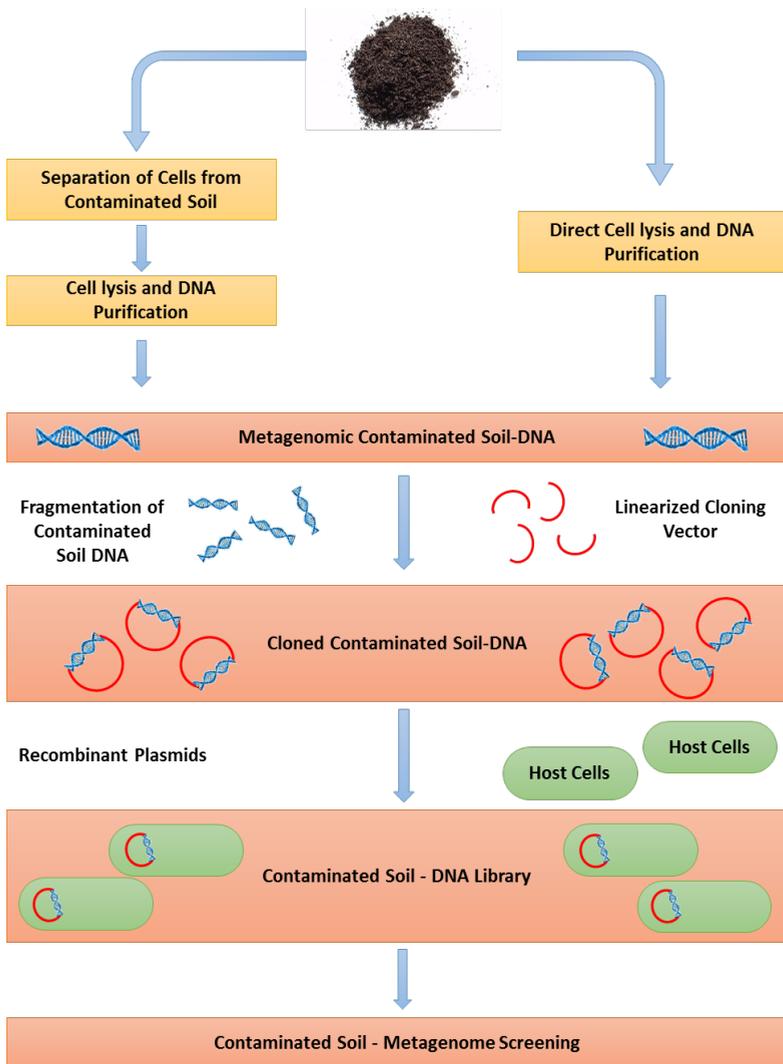


Figure 1. An ideal systematic workflow of steps involved in contaminated-soil metagenomics.

When assembly and annotation were completed, new coding DNA sequences related to genes of interest were identified with low protein similarity against the closest hits in the databases. This work basically highlights the sensitivity of DNA/RNA hybridization techniques as an effective and complementary way to recover novel genes from large metagenomic clone libraries with respect to soil microbiota. Nevertheless, multiple molecular biological-based techniques [43] may also be applied during the process of metagenome extraction and screening. The basic workflow of extracting metagenomes out of contaminated soil has been explained in Fig. 1. The steps were initiated by collecting contaminated soil from the environment. The collected contaminated soil sample can be processed in two ways; one is by direct cell lysis and DNA purification and second, by separation of cells from contaminated soil and then followed by cell lysis and DNA purification. The isolated DNA is then cloned using specific cloning vectors. The cloned contaminated soil DNA is then delivered into host cells using different gene delivery systems. The multiplied host cells containing contaminated soil DNA forms a Metagenome library and these contaminated soil metagenomes were then screened. A recent study conducted screening of biosurfactant producers from petroleum hydrocarbon contaminated sources in cold marine environments. In this study, the researchers have isolated and characterized 55 biosulphant microbiota of 8 different genera including 1 *Alcanivorax*, 1 *exiguobacterium*, and 2 *halomonas* strains [44].

4.2. Florescence-Activated Cell sorting (FACS)

Florescence-activated cell sorting is one of the most widely used cell sorting techniques which is applied to sort microbial cells based on florescence during the process of metagenomic screening [45], with an accuracy rate of 5,000 cells per second [46]. Figure 2 shows the schematic flow of SIGEX and intercellular biosensors methods. High-throughput screening does not require a selectable phenotype. This phenomenon has led to the focus on phenotypes such as pigments that are readily visible providing the use of fluorescence-activated cell sorting. Moreover, FACS can be used to detect expression of certain types of genes by regulation of a fluorescent biosensor present in the same cell as the metagenomic DNA [47, 48]. Hence, these screen methods will be a critical tool for rapid selection of cells from metagenomic libraries.

4.3. Metagenomic sequencing strategies

Genome sequencing technologies have been frequently upgraded [49] since the completion of the human genome at the beginning of the 21st century [50]. Multiple next-generation genomic sequencing strategies are applied to sequence the metagenomes of different microbial communities [51, 52]. Sequencing technologies were initiated by the Sangers sequencing method which was widely used during the process of human genome sequencing [53, 54]. Technological drift has gifted next-generation sequencing techniques like pyrosequencing [55, 56], ligation sequencing [57, 58], reverse terminator [60, 61], and single-molecule sequence by synthesis [62, 63], providing a high throughput that reads comparatively in less time [64-66]. A comparative overview of recent sequencing technologies applied in metagenome sequencing is provided in Table 1 for a more detailed understanding. However, most metagenomics researchers prefer the pyrosequencing method for sequencing the metagenomes of microbial communities [67-70].

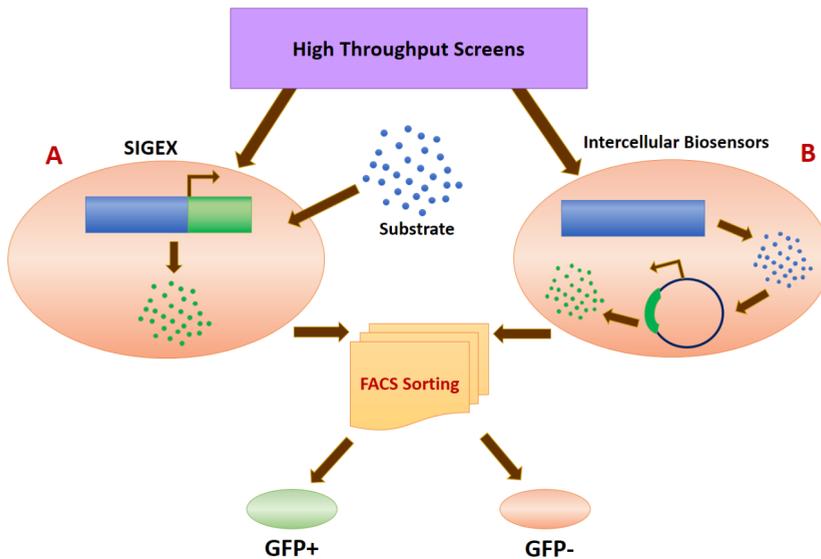


Figure 2. Systematic workflow representing the examples of high-throughput screens of (A) SIGEX and an (B) intracellular biosensor. SIGEX exploits the principle that catabolic genes are often substrate-induced by fusing a promoterless GFP to the metagenomic DNA and identifying clones in which GFP production is induced by the substrate of interest. An intracellular biosensor detects biologically active small molecules. GFP expression is dependent on the presence of a small molecule that activates a regulator. Finally, FACS is used to sort the GFP+ and GFP- cells separately.

Sequence Reaction Method	Read Length	Amplification	Data Production	Templates per run	Commercially Available as
Sanger's Method	~900 to 1,100	PCR	1 Mb per day	96	ABI 3730xl
Pyrosequencing	~400	Emulsion PCR	400 Mb per run in 7.5 to 8 hours	1,000,000	454 FLX Roche
Reverse Terminator	36 to 175	Bridge PCR	>17 Gb per run in 3 to 6 days	40,000,000	Illumina SOLEXA Genome Analyser
Ligation Sequencing	~50	Emulsion PCR	10 to 15 Gb per run in 6 days	85,000,000	ABI SOLiD
Single Molecule Sequence by Synthesis	30 to 35	None	21 to 28 Gb per run in 8 days	800,000,000	Helicos Heliscope

Table 1. A Comparative overview of next-generation sequencing technologies applied in metagenome sequencing

5. Prevalent metagenomes for bioremediation

Metagenomes extracted from uncultured microbial communities from multiple contaminant sites are screened and further identified for degrading properties [71]. Microbial communities

vary according to the characteristics of source and site of contamination [72]. A metagenomic analysis conducted on the heavy metal-contaminated groundwater revealed metagenomes of γ - and β -Proteobacteria dominated by *Rhodanobacter*-like γ -proteobacterial and *Burkholderia*-like β -proteobacterial species from the habitat of extremely high levels of uranium, nitrate, technetium and various organic contaminants [73]. Moreover, multiple metagenome projects have been taking place around the world; we have sorted out a list of multiple environmental metagenome projects with top microbe having the highest percentage of presence in the metagenomic community (Table 2). Studies on microbial adaptation of toxic environments may give rise to trace new metagenomic communities useful for efficient bioremediation. Specific functions and interactions of microbial communities with respect to contamination-degrading capabilities can be a result of environmental-based gene switching in the metagenomes.

Top Phylum	Percentage of Presence in Community	Domain	Metagenome Projects	Source
Actinobacteria	38.04	Bacteria	BASE - Biomes of Australian Soil Environments	Soil
Chlorobi	56.04	Bacteria	Antarctica Aquatic Microbial Metagenome	Environmental
Actinobacteria	38.21	Bacteria	American Lake Mendota metagenome	Water
Proteobacteria	31.62	Bacteria	Swedish Lake Vattern metagenome	Water
Proteobacteria	29.68	Bacteria	Detoxification of arsenic mediated by microbial sulphate reduction in Mediterranean marine sediments	Environmental
Proteobacteria	48.12	Bacteria	Illumina and 454-based metatranscriptomic analyses of a diatom-induced bacterioplankton bloom in the North Sea	Environmental
Unassigned Bacteria	34.8	Bacteria	Functional metagenomic profiling of Tibetan Plateau soils affected by permafrost or seasonal freezing	Soil
Euryarchaeota	22.71	Archaea	Lonar Lake Sediment prokaryotic metagenome	Water
Unassigned Bacteria	53.84	Bacteria	Metagenome of a microbial consortium obtained from the tuna oil field in the Gippsland Basin, Australia	Environmental
Actinobacteria	27.1	Bacteria	Meta soil	Soil

Table 2. List of multiple environmental metagenome projects with top microbe having the highest percentage of presence in the metagenomic community

6. Bioinformatic tools for metagenomic bioremediation

In the last two decades, bioinformatics has been advanced and simultaneously adapted to multiple fields of science such as basic sciences and advanced applied sciences [74]. Our previous study has given a glance of basic applications of bioinformatics in bioremediation [75]. Bioinformatics holds multiple tasks in the field of metagenomic bioremediation, majorly during metagenomic data analysis [76, 77]. A special issue on bioinformatics approaches and tools for metagenomic analysis has provided an advanced view towards comprehensive bioinformatic tools and methodologies used in metagenomics [78].

Multiple metagenomic projects are generating a large chunk of metagenomic sequence data challenging bioinformatics to develop more robust and better tools to analyze metagenomic sequence data. A recent study reveals the metagenomic characterization of soil microbial community using metagenomic approaches [79]. In this study, researchers have used 33 publicly available metagenomes obtained from diverse soil sites and integrated some state-of-the-art computational tools to explore the phylogenetic and functional characteristics of the microbial communities in soil. Recently, multiple advancements have taken place in the field of bioinformatics with respect to metagenomic bioremediation. In this section, most of our study focuses on recent bioinformatic tools and datasets majorly used in the analysis of metagenomic data in bioremediation. A comparative overview of functions and suitability of mostly used tools for metagenomic analysis is given in Table 3.

6.1. MEGAN

Meta Genome Analyzer (MEGAN) is one of the most widely used software tools for efficiently analyzing large chunks of metagenomic sequence data [80, 81]. This tool is most preferably used to interactively analyze and compare metagenomic and metatranscriptomic data, taxonomically and functionally. To perform taxonomic analysis, the program places reads onto the NCBI taxonomy and functional analysis is performed by mapping reads to the SEED, COG, and KEGG classifications. In addition, samples can be compared taxonomically and functionally, using a wide range of charting and visualization techniques like co-occurrence plots. This software also performs PCoA (Principle Coordinate Analysis) and clustering methods allowing high-level comparison of large numbers of samples [82]. Different attributes of the samples can be captured and used during analysis. Moreover, MEGAN supports different input formats of data and is capable of exporting the results of analysis in different text-based and graphical formats. Multiple methods of analysis, acceptance and comparison of high throughput data, robustness and being easy-to-handle are some of the features that made MEGAN as one of the most used metagenome analyzers.

6.2. SmashCommunity

Simple Metagenomics Analysis SHell for microbial communities (SmashCommunity) is a stand-alone metagenomic annotation and analysis pipeline that shares design principles and routines with SmashCell [83]. It is suitable for data delivered from Sanger and 454 sequencing technologies. It supports state-of-the-art software for essential metagenomic tasks such as

assembly and gene prediction. It also provides tools to estimate the quantitative phylogenetic and functional compositions of metagenomes, to compare compositions of multiple metagenomes, and to produce intuitive visual representations of such analyses [84]. It provides optimized parameter sets for Arachne and Celera for metagenome assembly, and GeneMark and MetaGene for predicting protein coding genes on metagenomes. SmashCommunity also includes scripts for downstream analysis of datasets. They can generate intuitive tree-based visualizations of results using the batch access API of the interactive Tree of Life (iTOL) web tool. SmashCommunity can also compare multiple metagenomes using these profiles, cluster them based on a relative entropy-based distance measure suitable for comparing such quantitative profiles, perform bootstrap analysis of the clustering, and generate visual representation of the clustering results.

6.3. CAMERA

Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA) is a database and associated computational infrastructure that provides a single system for depositing, locating, analyzing, visualizing, and sharing data about microbial biology through an advanced web-based analysis portal [85]. CAMERA holds a huge chunk of data including environmental metagenomic and genomic sequence data, associated environmental parameters, pre-computed search results, and software tools to support powerful cross-analysis of environmental samples. CAMERA works on a pattern of collecting and linking metadata relevant to environmental metagenome datasets with annotation in a semantically aware environment that allows users to write expressive semantic queries to the database. It also provides data submission tools to allow researchers to share and forward data to other metagenomic sites and community data archives. CAMERA can be best considered as a complete genome-analysis tool allowing users to query, analyze, annotate, and compare metagenome and genome data [86].

6.4. MG-RAST

Rapid Annotation using Subsystems Technology for Metagenomes (MG-RAST) is an automated analysis platform for metagenomes, providing quantitative insights into microbial populations based on sequence data [87]. This pipeline performs quality control, protein prediction, clustering, and similarity-based annotation on nucleic acid sequence datasets using a number of bioinformatic tools. Users can upload raw sequence data in FASTA format; the sequences will be normalized and processed, and summaries will be automatically generated. The MG-RAST server provides several methods of access to different data types, including phylogenetic and metabolic reconstructions, and has the ability to compare metabolism and annotations of one or more metagenomes and genomes. In addition, the server also offers a comprehensive search capability. The pipeline is implemented in Perl by using a number of open-source components, including the SEED framework, NCBI BLAST, SQLite, and Sun Grid Engine.

Suitability	SmashCommunity	MG-RAST	IMG/M	RAMMCP	MEGAN	MOTHUR
Single Genomes	-	-	+	+	-	-
Single Cell Genomes	-	-	-	-	-	-
Shotgun Metagenomes	+	+	+	+	+	-
16S rDNA Metagenomes	+	+	-	+	+	+
Functions						
Assembly	+	-	+	-	-	-
Gene Prediction	+	-	+	+	-	-
Functional Analysis	+	+	+	+	+	-
Taxonomy Assignment	+	+	+	+	+	+
Comparative Analyses	+	+	+	+	+	+
Data Management	+	-	+	-	-	-

+Yes, -No

Table 3. A comparative overview of functions and suitability of mostly used tools for metagenomic analysis

6.5. IMG/M

Integrated Microbial Genomes and Metagenomes (IMG/M) system supports annotation, analysis, and distribution of microbial genome and metagenome datasets. IMG/M provides comparative data using analytical tools extended to handle metagenome data, together with metagenome-specific analysis [88, 89]. IMG/M consists of samples of microbial community aggregate genomes integrated with IMG’s comprehensive set of genomes from all three domains of life: plasmids, viruses, and genome fragments. Function-based comparison of metagenome samples and genomes is provided by analytical tools that allow examination of the relative abundance of protein families, functional families or functional categories across metagenome samples and genomes. It seems like registered users can gain more advantage out of IMG/M as the tools focus on handling substantially larger metagenome datasets, are available only to registered users as part of the ‘My IMG’ toolkit, and support specifying, managing, and analyzing persistent sets of genes, functions, genomes or metagenome samples and scaffolds.

7. Summary

Metagenomics is a strategic approach for analyzing microbial communities at a genomic level. This gives a glimpse towards the microbial community view of “Uncultured Microbiota”. Bioremediation has always been adapting new advances in science and technology for

establishing better environments, and metagenomics can be considered as one of the best adaptations ever. Identification and screening of metagenomes from the polluted environments are crucial in a metagenomic study. The second section emphasizes recent multiple case studies explaining the approaches of metagenomics in bioremediation. Accordingly, the third section speaks about metagenomic bioremediation in different contaminated environments such as soil and water. The fourth section explains different sequences and function-based metagenomic strategies and tools starting from providing a detailed view of metagenomic screening, FACS, and multiple advanced metagenomic sequencing strategies. The fifth section deals with the prevalent metagenomes in bioremediation giving a list of different prevalent metagenomic organisms and their respective projects. The last section gives a detailed view of different major bioinformatic tools and datasets most prevalently used in metagenomic data analysis and processing during metagenomic bioremediation.

Author details

Pratap Devarapalli and Ranjith N. Kumavath*

*Address all correspondence to: rnkumavath@gmail.com

Department of Genomic Science, Central University of Kerala, Riverside Transit Campus, Padanakkad, Nileshwer, Kasaragod, Kerala, India

References

- [1] Baker, K. H., & Herson, D. S. (1994). Bioremediation. McGraw-Hill, Inc.
- [2] Cookson Jr, J. T. (1995). Bioremediation engineering: design and application. McGraw-Hill, Inc.
- [3] Bashir, Y., Pradeep Singh, S., & Kumar Konwar, B. (2014). Metagenomics: An Application Based Perspective. Chinese Journal of Biology..
- [4] Joshi, M. N., Dhebar, S. V., Bhargava, P., Pandit, A. S., Patel, R. P., Saxena, A. K., & Bagatharia, S. B. (2014). Metagenomic approach for understanding microbial population from petroleum muck. Genome Announcements, 2(3), e00533-14.
- [5] Duarte, M. "Functional soil metagenomics: elucidation of polycyclic aromatic hydrocarbon degradation potential after 10 years of *in situ* bioremediation." Annual Meeting and Exhibition (2014). Society for Industrial Microbiology & Biotechnology.
- [6] Prasad, Rajendra. "New approaches and insights into bioremediation of hazardous waste." Reviews on environmental health. 29.1-2 (2014): 33-35.

- [7] Miyazaki, Kentaro, and Hikaru Suenaga. "Extradiol Dioxygenases Retrieved from the Metagenome." *Encyclopedia of Metagenomics*. New York: Springer, 2014. 1-5.
- [8] Joshi, Madhvi N., et al. "Metagenomics of petroleum muck: revealing microbial diversity and depicting microbial syntrophy." *Archives of microbiology* (2014): 1-14.
- [9] Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & biology*, 5(10), R245-R249.
- [10] Shah, V., Jain, K., Desai, C., & Madamwar, D. (2011). Metagenomics and integrative 'omics' technologies in microbial bioremediation: current trends and potential applications. *Metagenomics: current innovations and future trends*. Caister Academic Press, Norfolk, 211-240.
- [11] Duarte, M. (2014, July). Functional soil metagenomics: elucidation of polycyclic aromatic hydrocarbon degradation potential after 10 years of *in situ* bioremediation. In Annual Meeting and Exhibition 2014 (July 20-24, 2014). Society for Industrial Microbiology & Biotechnology.
- [12] Uhlík, O., Strejček, M., Hroudová, M., Demnerová, K., & Macek, T. (2013). Identification and Characterization of Bacteria with Bioremediation Potential: from Cultivation to Metagenomics. *CHEMICKE LISTY*, 107(8), 614-622.
- [13] Röling, W. F. (2014). Maths on microbes: adding microbial ecophysiology to metagenomics. *Microbial biotechnology*.
- [14] Bell, T. H., Joly, S., Pitre, F. E., & Yergeau, E. (2014). Increasing phytoremediation efficiency and reliability using novel omics approaches. *Trends in biotechnology*, 32(5), 271-280.
- [15] Chemerys, A., Pelletier, E., Cruaud, C., Martin, F., Violet, F., & Jouanneau, Y. (2014). Characterization of Novel Polycyclic Aromatic Hydrocarbon Dioxygenases from the Bacterial Metagenomic DNA of a Contaminated Soil. *Applied and environmental microbiology*, 80(21), 6591-6600.
- [16] Dellagnezze, B. M., de Sousa, G. V., Martins, L. L., Domingos, D. F., Limache, E. E., de Vasconcellos, S. P., & de Oliveira, V. M. (2014). Bioremediation potential of microorganisms derived from petroleum reservoirs. *Marine pollution bulletin*, 89(1), 191-200.
- [17] Gillan, D. C., Roosa, S., Kunath, B., Billon, G., & Wattiez, R. (2014). The long-term adaptation of bacterial communities in metal-contaminated sediments: a metaproteogenomic study. *Environmental microbiology*.
- [18] Khan, M. A. Q. (Ed.). (2012). *Survival in toxic environments*. Elsevier.
- [19] Bind, M. A., Baccarelli, A., Zanobetti, A., Tarantini, L., Suh, H., Vokonas, P., & Schwartz, J. (2012). Air pollution and markers of coagulation, inflammation and en-

- dothelial function: Associations and epigene-environment interactions in an elderly cohort. *Epidemiology (Cambridge, Mass.)*, 23(2), 332.
- [20] Brimblecombe, P. (2011). *The big smoke: a history of air pollution in London since medieval times*. Routledge.
- [21] Lave, L. B., & Seskin, E. P. (2013). *Air pollution and human health (Vol. 6)*. Routledge.
- [22] Törnqvist, R., Jarsjö, J., & Karimov, B. (2011). Health risks from large-scale water pollution: trends in Central Asia. *Environment international*, 37(2), 435-442.
- [23] Chen, R., de Sherbinin, A., Ye, C., & Shi, G. (2014). China's soil pollution: farms on the frontline. *Science (New York, NY)*, 344(6185), 691-691.
- [24] Grace Liu, P. W., Chang, T. C., Whang, L. M., Kao, C. H., Pan, P. T., & Cheng, S. S. (2011). Bioremediation of petroleum hydrocarbon contaminated soil: effects of strategies and microbial community shift. *International Biodeterioration & Biodegradation*, 65(8), 1119-1127.
- [25] Gargouri, B., Mnif, S., Aloui, F., Karray, F., Mhiri, N., Chamkha, M., & Sayadi, S. (2012). Bioremediation of petroleum contaminated water and soils in tunisia. In *Clean Soil and Safe Water* (pp. 153-165). Springer Netherlands.
- [26] Wölz, J., Schulze, T., Lübcke-von Varel, U., Fleig, M., Reifferscheid, G., Brack, W., & Hollert, H. (2011). Investigation on soil contamination at recently inundated and non-inundated sites. *Journal of Soils and Sediments*, 11(1), 82-92.
- [27] Pennock, S. R., Abed, T. M., Curioni, G., Chapman, D. N., John, U. E., & Jenks, C. H. J. (2014). Investigation of soil contamination by iron pipe corrosion and its influence on GPR detection. In *Ground Penetrating Radar (GPR), 2014 15th International Conference on* (pp. 381-386). IEEE.
- [28] Robinson, D. A., Jackson, B. M., Clothier, B. E., Dominati, E. J., Marchant, S. C., Cooper, D. M., & Bristow, K. L. (2013). Advances in soil ecosystem services: Concepts, models, and applications for earth system life support. *Vadose Zone Journal*, 12(4).
- [29] Uhlik, O., Leewis, M. C., Strejcek, M., Musilova, L., Mackova, M., Leigh, M. B., & Macek, T. (2013). Stable isotope probing in the metagenomics era: a bridge towards improved bioremediation. *Biotechnology advances*, 31(2), 154-165.
- [30] Adetutu, E. M., Smith, R. J., Weber, J., Aleer, S., Mitchell, J. G., Ball, A. S., & Juhasz, A. L. (2013). A polyphasic approach for assessing the suitability of bioremediation for the treatment of hydrocarbon-impacted soil. *Science of the Total Environment*, 450, 51-58.
- [31] Yergeau, E., Sanschagrin, S., Beaumier, D., & Greer, C. W. (2012). Metagenomic analysis of the bioremediation of diesel-contaminated Canadian high arctic soils. *PloS One*, 7(1), e30058.

- [32] Sirés, I., & Brillas, E. (2012). Remediation of water pollution caused by pharmaceutical residues based on electrochemical separation and degradation technologies: a review. *Environment international*, 40, 212-229.
- [33] Rozell, D. J., & Reaven, S. J. (2012). Water pollution risk associated with natural gas extraction from the Marcellus Shale. *Risk Analysis*, 32(8), 1382-1393.
- [34] Widger, W. R., Golovko, G., Martinez, A. F., Ballesteros, E., Howard, J., Xu, Z., & Fofanov, Y. (2011). Longitudinal metagenomic analysis of the water and soil from Gulf of Mexico beaches affected by the Deep Water Horizon oil spill. *Nature Proceedings*.
- [35] Layton, A. C., Chauhan, A., Williams, D. E., Mailloux, B., Knappett, P. S., Ferguson, A. S., & Sayler, G. S. (2014). Metagenomes of microbial communities in arsenic-and pathogen-contaminated well and surface water from Bangladesh. *Genome announcements*, 2(6), e01170-14.
- [36] Shi, P., Jia, S., Zhang, X. X., Zhang, T., Cheng, S., & Li, A. (2013). Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. *Water research*, 47(1), 111-120.
- [37] Hansen, T., Skånseng, B., Hoorfar, J., & Löfström, C. (2013). Evaluation of direct 16S rDNA sequencing as a metagenomics-based approach to screening bacteria in bottled water. *Biosecurity and bioterrorism: biodefense strategy, practice, and science*, 11(S1), S158-S165.
- [38] Williams, P. J., Botes, E., Maleke, M. M., Ojo, A., DeFlaun, M. F., Howell, J., & van Heerden, E. (2014). Effective bioreduction of hexavalent chromium-contaminated water in fixed-film bioreactors. *Water SA*, 40(3), 549-554.
- [39] Garfield, E., & Merton, R. K. (1979). *Citation indexing: Its theory and application in science, technology, and humanities* (Vol. 8). New York: Wiley.
- [40] Tijssen, R. J. (2002). Science dependence of technologies: evidence from inventions and their inventors. *Research policy*, 31(4), 509-526.
- [41] Simon, C., & Daniel, R. (2009). Achievements and new knowledge unraveled by metagenomic approaches. *Applied microbiology and biotechnology*, 85(2), 265-276.
- [42] Jacquioid, Samuel, et al. "Characterization of new bacterial catabolic genes and mobile genetic elements by high throughput genetic screening of a soil metagenomic library." *Journal of biotechnology* (2014).
- [43] Godheja, Jai, Sudhir K. Shekhar, and D. R. Modi. "Advances in Molecular Biology Approaches to Gauge Microbial Communities and Bioremediation at Contaminated Sites." *International Journal of Environmental Bioremediation & Biodegradation* 2.4 (2014): 167-177.
- [44] Cai, Qinrong, Baiyu Zhang, Bing Chen, Zhiwen Zhu, Weiyun Lin, and Tong Cao (2014). "Screening of biosurfactant producers from petroleum hydrocarbon contami-

- nated sources in cold marine environments." *Marine pollution bulletin*, 86 (1), 402-410.
- [45] Handelsman, J. (2005). Sorting out metagenomes. *Nature biotechnology*, 23(1), 38-39.
- [46] Herzenberg, L. A., Sweet, R. G., & Herzenberg, L. A. (1976). Fluorescence-activated cell sorting. *Sci Am*, 234(3), 108-117.
- [47] Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and molecular biology reviews*, 68(4), 669-685.
- [48] Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., & Woyke, T. (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nature protocols*, 9(5), 1038-1048.
- [49] Metzker, M. L. (2009). Sequencing technologies—the next generation. *Nature Reviews Genetics*, 11(1), 31-46.
- [50] Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., & Moreno, R. F. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252(5013), 1651-1656.
- [51] Rodrigue, S., Materna, A. C., Timberlake, S. C., Blackburn, M. C., Malmstrom, R. R., Alm, E. J., & Chisholm, S. W. (2010). Unlocking short read sequencing for metagenomics. *PLoS One*, 5(7), e11840.
- [52] Tringe, S. G., & Rubin, E. M. (2005). Metagenomics: DNA sequencing of environmental samples. *Nature reviews genetics*, 6(11), 805-814.
- [53] Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463-5467.
- [54] Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., & Weissenbach, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285), 59-65.
- [55] Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., & Anastasi, C. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), 53-59.
- [56] Loman, N. J., Constantinidou, C., Chan, J. Z., Halachev, M., Sergeant, M., Penn, C. W., & Pallen, M. J. (2012). High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nature Reviews Microbiology*, 10(9), 599-606.
- [57] Roesch, L. F., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K., Kent, A. D., & Triplett, E. W. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME journal*, 1(4), 283-290.

- [58] Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A., & Versalovic, J. (2009). Metagenomic pyrosequencing and microbial identification. *Clinical chemistry*, 55(5), 856-866.
- [59] Pfeifer, G. P., Steigerwald, S. D., Mueller, P. R., Wold, B. A. R. B. A. R. A., & Riggs, A. D. (1989). Genomic sequencing and methylation analysis by ligation mediated PCR. *Science*, 246(4931), 810-813.
- [60] Ludwig, M., & Bryant, D. A. (2011). Transcription profiling of the model cyanobacterium *Synechococcus* sp. strain PCC 7002 by Next-Gen (SOLiD™) sequencing of cDNA. *Frontiers in microbiology*, 2.
- [61] Song, C. X., Clark, T. A., Lu, X. Y., Kislyuk, A., Dai, Q., Turner, S. W., & Korlach, J. (2012). Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nature methods*, 9(1), 75-77.
- [62] Harris, T. D., Buzby, P. R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., & Xie, Z. (2008). Single-molecule DNA sequencing of a viral genome. *Science*, 320(5872), 106-109.
- [63] Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., & Law, M. (2012). Comparison of next-generation sequencing systems. *BioMed Research International*.
- [64] Wall, P. K., Leebens-Mack, J., Chanderbali, A. S., Barakat, A., Wolcott, E., Liang, H., & Altman, N. (2009). Comparison of next generation sequencing technologies for transcriptome characterization. *BMC genomics*, 10(1), 347.
- [65] Claesson, M. J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J. R., Ross, R. P., & O'Toole, P. W. (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic acids research*, gkq873.
- [66] Niu, B., Fu, L., Sun, S., & Li, W. (2010). Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC bioinformatics*, 11(1), 187.
- [67] Morgan, J. L., Darling, A. E., & Eisen, J. A. (2010). Metagenomic sequencing of an in vitro-simulated microbial community. *PLoS One*, 5(4), e12029.
- [68] Schlüter, A., Bekel, T., Diaz, N. N., Dondrup, M., Eichenlaub, R., Gartemann, K. H., & Goesmann, A. (2008). The metagenome of a biogas-producing microbial community of a production-scale biogas plant fermenter analysed by the 454-pyrosequencing technology. *Journal of Biotechnology*, 136(1), 77-90.
- [69] Ronaghi, M. (2001). Pyrosequencing sheds light on DNA sequencing. *Genome research*, 11(1), 3-11.
- [70] Brennerova, M. V., Josefiova, J., Brenner, V., Pieper, D. H., & Junca, H. (2009). Metagenomics reveals diversity and abundance of meta-cleavage pathways in microbial

- communities from soil highly contaminated with jet fuel under air-sparging bioremediation. *Environmental microbiology*, 11(9), 2216-2227.
- [71] MacNaughton, S. J., Stephen, J. R., Venosa, A. D., Davis, G. A., Chang, Y. J., & White, D. C. (1999). Microbial population changes during bioremediation of an experimental oil spill. *Applied and environmental microbiology*, 65(8), 3566-3574.
- [72] Hemme, C. L., Deng, Y., Gentry, T. J., Fields, M. W., Wu, L., Barua, S., & Zhou, J. (2010). Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. *The ISME journal*, 4(5), 660-672.
- [73] Moody, G. (2004). *Digital code of life: how bioinformatics is revolutionizing science, medicine, and business*. John Wiley & Sons.
- [74] Kumavath, R. N., & Deverapalli, P. (2013). *Scientific Swift in Bioremediation: An Overview*. Intech Publishers.
- [75] Markowitz, V. M., Ivanova, N., Palaniappan, K., Szeto, E., Korzeniewski, F., Lykidis, A., & Kyrpides, N. C. (2006). An experimental metagenome data management and analysis system. *Bioinformatics*, 22(14), e359-e367.
- [76] Mitra, S., Klar, B., & Huson, D. H. (2009). Visual and statistical comparison of metagenomes. *Bioinformatics*, 25(15), 1849-1855.
- [77] Valiente, G., & Pesole, G. (2012). Bioinformatics approaches and tools for metagenomic analysis. Editorial. *Briefing in Bioinformatics*, 13(6):645.
- [78] Xu, Z., Hansen, M. A., Hansen, L. H., Jacquioid, S., & Sørensen, S. J. (2014). Bioinformatic Approaches Reveal Metagenomic Characterization of Soil Microbial Community. *PloS One*, 9(4), e93445.
- [79] Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome research*, 17(3), 377-386.
- [80] El Hadidi, M., Ruscheweyh, H. J., & Huson, D. (2013). Improved metagenome analysis using MEGAN5. In *Joint 21st Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) and 12th European Conference on Computational Biology (ECCB) 2013*.
- [81] Gower, J. C. (2005). Principal coordinates analysis. *Encyclopedia of Biostatistics*.
- [82] Harrington, E. D., Arumugam, M., Raes, J., Bork, P., & Relman, D. A. (2010). SmashCell: a software framework for the analysis of single-cell amplified genome sequences. *Bioinformatics*, 26(23), 2979-2980.
- [83] Arumugam, M., Harrington, E. D., Foerstner, K. U., Raes, J., & Bork, P. (2010). SmashCommunity: a metagenomic annotation and analysis tool. *Bioinformatics*, 26(23), 2977-2978.

- [84] Seshadri, R., Kravitz, S. A., Smarr, L., Gilna, P., & Frazier, M. (2007). CAMERA: a community resource for metagenomics. *PLoS biology*, 5(3), e75.
- [85] Sun, S., Chen, J., Li, W., Altintas, I., Lin, A., Peltier, S., & Wooley, J. (2010). Community cyberinfrastructure for advanced microbial ecology research and analysis: the CAMERA resource. *Nucleic acids research*, gkq1102.
- [86] Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., & Edwards, R. A. (2008). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics*, 9(1), 386.
- [87] Markowitz, V. M., Ivanova, N. N., Szeto, E., Palaniappan, K., Chu, K., Dalevi, D., & Kyrpides, N. C. (2008). IMG/M: a data management and analysis system for metagenomes. *Nucleic acids research*, 36(suppl 1), D534-D538.
- [88] Markowitz, V. M., Chen, I. M. A., Chu, K., Szeto, E., Palaniappan, K., Grechkin, Y., & Kyrpides, N. C. (2012). IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic acids research*, 40(D1), D123-D129.

