

Decoding Complex Soil Microbial Communities through New Age “Omics”

Girish R Nair and Suresh SS Raja*

Department of Microbiology, Bharathidasan University Constituent College, Kurumbalur, Perambalur-621107, Tamil Nadu, India

Abstract

Omics employs a suite of high-throughput techniques coupled with robust computational analysis to extract every aspect of information from microbial communities, thus giving access to their genome, metagenome, transcriptome, proteome and metabolome. Over the past 5-10 years omics has revolutionized the field of microbial ecology and diversity by overcoming several challenges associated with isolation and characterization of unknown microorganisms from different environments. With increasing technological advancements omics is growing rapidly by incorporating newer areas of study like metabolomics and culturomics and could be referred as new age omics. New age omics package involves unparalleled techniques of detection and analysis that could be employed to study soil microbiota, which was once considered as a challenging task. Present review summarizes chronological scientific discoveries that contributed towards study of soil microorganisms and familiarizes with omics methodologies and its importance in soil microbial ecology.

Keywords: Soil microbiota; Omics; Metagenomics; NGS; MG-RAST; Culturomics

Introduction

Soil microbiota and its significance

In the terrestrial ecosystem, soil is a major component that sustains variety of life forms among which the microbes are the most abundant ones. One gram of chernozem contains about 10 billion or more living microorganisms (10 tons per hectare) that strictly depend upon the soil microenvironment for their survival. The microbiota of the soil has a symbiotic-mutualistic behavior and can develop by direct as well as indirect ways in the soil. Microbes which originate directly in the soil or by decomposition of plant matter are known as the native microbiota. Some microbes like the pathogens and aquatic microbes can enter accidentally from the gastrointestinal tract of animals and man and through agricultural runoff to become part of the soil microbial community. Soil microbiota comprises both prokaryotes (bacteria, actinomycetes, blue-green algae) and eukaryotes (fungi, microscopic algae, protozoans) but bacteria are the most abundant group of soil microbiota in terms of species and are the first colonizers [1-4]. Soil microbiota contribute in soil structure formation, involve in decomposing organic matter and recalcitrant xenobiotics, modulate the global biogeochemical cycle and recycle nutrients as well as important elements such as carbon, nitrogen, phosphorous and sulphur [5-8]. Hence it is imperative to understand the dynamics of soil microbial communities which is under the influence of abiotic factors like soil fertility [9], substrate availability, pH [10], climate [11], soil temperature [12] and moisture, as well as shifts in seasonality [9,13] as well as biotic factors like plant communities [14], microbe food web interactions [15,16] and farming practices [17]. Enormous studies have been carried out in order to study the microbial communities thriving in soil which maintains the balance of ecosystem, these studies involve both culturing the bacteria in a nutrient rich media and identify them through direct DNA isolation and sequencing. Over the past years researchers are trying to fill the gaps generated by older techniques of isolation and identification, which are the primary requirements for any downstream application. The gaps in techniques of isolation may involve inability to grow or culture targeted microorganism and novel microorganisms present in the soil [18]. Drawbacks of identification strategies include inaccurate identification that lead to the reclassification of many bacterial genera [19]. The new era of microbial ecology was originated with the advent of high-throughput sequencing technologies and smarter culturing techniques which are

far more superior to the earlier methods of detection. These newer techniques of isolation and identification form the basis of the suffix “omics” that define today’s soil microbial studies. Omics amalgamates advanced instrumentation capability with sophisticated computational analysis fulfills the important fields of study that involve analyzing the genome, proteome, transcriptome and metabolome of a single bacteria or microbial community thriving in soil. This review primarily discusses the apparent technological advancements in the microbial diversity studies targeted to unfold the community structure and function of the soil microbiota.

Trends in microbial diversity studies

The study of microbes predates to 1600 AD, when Leeuwenhoek described his oral organisms in 1676 [20], which was followed by Robert Koch who designed nutrient media using potato slices or gelatin to isolate bacteria and count them [21]. Later, with the advancement in microscopy and staining techniques like Gram, Ziehl-Neelsen and Schaeffer and Fulton, the identification and characterization of microbes was significantly improved [21,22]. It was the work of Russian botanist Sergei winogradsky in early 1930’s on lithotrophy that threw light on the function of soil microorganisms and revolutionized the field of microbiology with the origin of the concept of “microbial ecology”. His work involved methods for culturing soil bacteria, studying iron bacteria, nitrifying bacteria, nitrogen fixation by azotobacter and cellulose degrading bacteria [23] (Figure 1). Thereafter, a plenty of reports published which involved the study of soil microbial communities in different terrestrial ecosystems exhibiting an independent microenvironment, like the desert ecosystem where researchers studied the xerophytic microbiota present in the soils of hot

*Corresponding author: Suresh SS Raja, Department of Microbiology, Bharathidasan university constituent college, Kurumbalur, Perambalur-621107, Tamil Nadu, India, Tel: 09443075015; E-mail: sudalaimuthuraja@yahoo.co.in

Received November 05, 2017; Accepted December 18, 2017; Published December 26, 2017

Citation: Nair GR, Raja SSS (2017) Decoding Complex Soil Microbial Communities through New Age “Omics”. J Microb Biochem Technol 9: 301-309. doi: [10.4172/1948-5948.1000382](https://doi.org/10.4172/1948-5948.1000382)

Copyright: © 2017 Nair GR, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

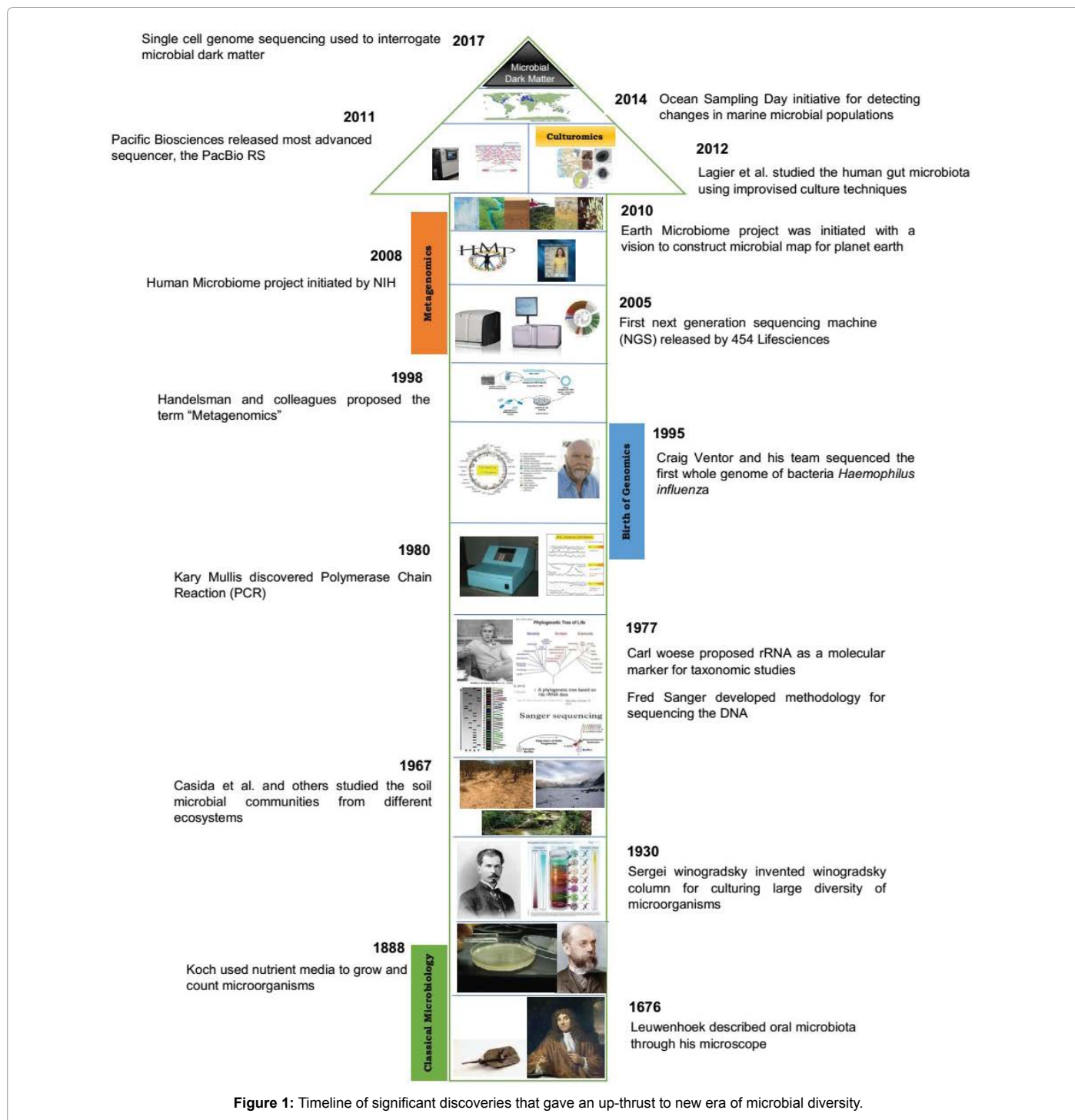


Figure 1: Timeline of significant discoveries that gave an up-thrust to new era of microbial diversity.

and cold deserts [24,25]. But all these studies provided limited insight into microbial world of the terrestrial ecosystem. A thrust to unravel and understand the complex soil microbial communities was given by the molecular characterization approaches with the introduction of 16S rRNA gene as a molecular marker for identification of eubacteria in the late 1970's by Carl Woese [26] together with the automated sequencing method developed by Sanger [27]. Further, these technologies became a platform for newer techniques like PCR based amplification of 16S rRNA gene to generate clonal library and identification using sequencing

[28-31], Amplified ribosomal DNA restriction analysis (ARDRA) denaturing gradient gel electrophoresis (DGGE & TGGE), restriction-fragment length polymorphism (RFLP), terminal restriction-fragment length polymorphism (T-RFLP) and Fluorescent in situ hybridization (FISH) etc. But, these techniques were limited by the fact that they could only detect the microbial community, they cannot be applied to understand the function of genes present in bacteria. Therefore, cloning approach was started to test the function of genes present in microorganisms and to express certain genes in bacteria that could

produce bioactive agents such as enzymes, antibiotics and metabolites and soil DNA libraries were extensively generated for such purposes [32]. With technological improvements in sequencing techniques and development of computational databases, studying the genome of bacteria was possible through whole genome sequencing. The first whole genome of bacteria sequenced was *Haemophilus influenzae*, a rod-shaped bacterium which causes meningitis in 1995 by Craig Venter and his team at The Institute for Genomic Research in Rockville, Maryland, USA. He used a technique called “shotgun” sequencing which involves fragmenting the DNA into smaller pieces and then sequencing to reconstruct the genome. This was further assembled using assembler software with information accuracy to obtain the genomic information of the bacteria [33]. Following which, many reports were published that involved genome sequencing of bacteria isolated from different environments. In 1998 Handelsman et al., contributed towards research in microbial diversity by studying the microbial community from specific environment that involved multiple genomes and coined the term “Metagenomics”. The offshoot of genomics and metagenomics saw promising strategies that could very well unravel the unseen microbial diversity and function of soil microbial communities and led to development of more advanced sequencing technologies which is known as the Next Generation Sequencing (NGS). The NGS technology came into existence when 454 Lifesciences launched the first NGS machine based on “pyrosequencing”, which is now marketed by Roche. Later on, Illumina/Solexa also developed NGS machines. In June 2009, Virginia bioinformatics institute and Virginia Tech together sequenced the complete genome of nitrogen-fixing, soil-living bacterium *Azotobacter vinelandii* which led to understanding of its features and metabolic potential [34]. Following years, the culture-independent techniques became indispensable to understand the genetic diversity, population structure, and ecological roles of the majority of microorganisms. Metagenome sequencing involves the functional and sequencing based analysis of collective genomes isolated from a particular niche using NGS and advanced bioinformatics tools. Microorganisms being ubiquitous, a new trend of studying the microbiota, which was present in soil, ocean and human body, was started by several research groups. These studies were facilitated by Metagenomics, following which several Metagenome databases and sequencing projects came into existence, one of which was the Human Microbiome project initiated by National Institute of Health (NIH), USA in 2008 aimed at characterizing the human microbiota and analyzing their role in human health and disease. Another project is the EMP or the Earth Microbiome project started in August 2010, targeted to generate the microbial map of earth by sequencing and analyzing over 200,000 samples from different biomes of earth. Although Metagenomics has the potential to decipher the uncultured microbial population, it has limitations as it ignores the minority bacterial populations during bacterial community studies from an ecosystem. Some other problems associated with Metagenomics are cloning biases [35], sampling biases, misidentification of “decorating enzymes” and incorrect promoter sites in genomes and dispersion of genes involved in secondary metabolite production [36]. These biases could only be resolved using statistical approaches in order to detect the difference between the expected and observed bacterial diversity and to determine the actual species richness [37]. Cultivation based studies, once considered outdated is now been re-considered for microbial diversity studies. Lagier et al. used improved culture techniques to isolate 31 new bacterial species, a large human virus, the largest bacteria and largest archaea from human in 2012. Apparently, the researchers saw the need of new methods and strategies for cultivation of bacteria in diversity related studies which cannot be fulfilled by Metagenomics alone. This led to development of

enhanced culture techniques which gained superiority over traditional cultivable approach is now known as “Culturomics”. The field of single-cell genomics is rapidly growing as it aids in exploring “microbial dark matter”, which constitutes the unknown population of microbes that were considered uncultivable at laboratory [38]. Over several years these population of microbes stayed hidden from scientists, but recent studies have shown promising approaches to cultivate these unknown bacteria in laboratory and exploiting their potential to generate new antibiotics that can target multidrug resistant pathogens [39]. Single-cell genomics involves sequencing the individual cell DNA with optimized NGS technologies. Single-cell genome sequencing provides a precise genetic map of a single bacterium living in a microbial community. This genetic map provides a better understanding of the function of an individual bacterium in the context of its microenvironment, thereby complementing the cultivation based approaches through satisfying the specific growth requirements of uncultivable bacterium from metabolic profile of an individual cell [40] (Figure 1).

Technological advancements in the new era of “omics”

The approaches to study and understand the microbial life thriving in different micro-environments has undergone a remarkable change since Koch used to culture bacteria using synthetic media. Modern era of microbial ecology witnessed the technological revolution, which when combined together with the classical microbiology gave birth to “omics” that aims at the collective high-throughput characterization and quantification of pools of biological molecules that translate into structure, function and dynamics of an organism and organisms. “Omics” is an English-language neologism which provides a set of advanced tools to study the genome, metagenome, proteome, metabolome and transcriptome of a micro-organism. It is used as a suffix to generate these important fields of study like genomics, metagenomics, proteomics, metabolomics, lipidomics, transcriptomics and culturomics. It all starts with isolating and sequencing the DNA and the success majorly depends upon the quality of DNA isolated and depth of sequencing. Nowadays, sequencing the DNA has undergone a dramatic shift with the arrival of new sequencing chemistries, instruments, and bioinformatics which we call now NGS. NGS is much faster and carries a greater sequencing depth compared to traditional Sanger’s dideoxy sequencing and in the past years, the rapid and substantial cost reduction in NGS technologies has accelerated the use of it in Genomic and Metagenomic studies. With NGS researchers could very well explore the rare bacterial groups present in the community of an environmental sample that could provide a potentially inexhaustive genetic reservoir. Currently, the NGS platforms could be classified based on the parameters like maximum read length, cost, run-time and error rate (Table 1) [41]. The first generation sequencing technologies involved the use of cloning DNA into vectors or bacterial host for library preparation, which had a greater chance of DNA contamination and error in sequencing results. This problem was resolved by the second generation sequencing platforms that do not require cloning process for library preparation. Yet amplification biases and short read length were persistent, which was sorted out by single molecule real time technologies (SMRT) like the PacBio RS from Pacific Bioscience [42] which are considered as third generation sequencing platforms. Now, which sequencing platform is better for use is still debatable as each one has its own shortcoming. But one could ascertain it based on the characteristics of these sequencing platforms to use a customized approach of sequencing based on the type of environmental sample, requirement of data and depth of analysis. As of now, Illumina’s sequencing platform has become the most widely used platform for Genomic and Metagenomic sequencing because of its low sequencing

	Roche 454	IonTorrent PGM	Illumina	PacBio RSIIa
Sequencing chemistry	Pyrophosphate (PPI) release transformed to luminous signal	Measured hydrogen potential converted to signal during proton release	Laser excitation of the incorporated fluorescently labeled nucleotide	Zero-mode wave guide system detects the fluorescently labeled single nucleotide insertion
Maximum read length (bp)	1200	400	300b	50,000
Output per run (Gb)	1	2	1000c	1
Amplification for library construction	Yes	Yes	Yes	No
Cost/Gb (USA Dollar)	\$9538.46	\$460.00	\$29.30	\$600
Error kind	Indel	Indel	Substitution	Indel
Error rate (%)	1	~ 1	~ 0.1	~ 13
Run time	20 h	7.3 h	6 days	2 h

Table 1: Shows the comparison of present sequencing modalities employed in genomics and metagenomics.

cost and higher yield. Another important aspect in the NGS is the processing of the crude data from the sample DNA runs that has to be translated into important information. One would require higher computational resources, more complex bioinformatic analysis and large data storage for processing NGS datasets, which means we not only require high end servers but also LINUX operative system skills [43]. Although, programming and scripting knowledge are desirable to run and install the available metagenomics software for processing the raw data and interpreting the results. But it is expected that the researchers working with NGS data for genomic and metagenomic analysis are trained in basic computational skills. The first step in data processing involves analyzing the quality of the raw reads generated by the sequencer. All sequencing platforms have their own quality check (QC) module in their software suite which assists in initial processing of raw reads that involves filtering of low quality reads, trimming and adapter removal. The results of QC analysis often contain information regarding sequencing output that involves number of reads, read length, GC content, overrepresented sequences, etc. Nevertheless, this can also be achieved by using easily operational software FastQC [44] and several others which are capable of performing initial quality check of high-throughput sequencing data, thus not entirely relying on the software suite based on the sequencing platform. Second step in the sequence processing after initial QC is called assembly and annotation that requires a whole lot of computing power along with programming skills. Therefore a computer with latest processor and large memory depending upon the dataset is advisable for a hassle free and less time consuming analysis. This job of assembly and annotation is made now easier by several centralized server pipelines like RAST, MG-RAST, IMG/M and CAMERA which analyze genomic and metagenomic datasets and later allow data storage and sharing of the computational results. But for a customized analysis, where one can modify parameters according to the needs is not supported by these server pipelines. This can be achieved using some legacy software packages like Mothur [45] QIIME [46] MEGAN [47], CARMA [48] and requires a sound knowledge in Linux operating system.

Materials and Methods

Bacterial genomics

Genomics was one of the pioneering fields which involved the use of omic technologies. Genomics is the study of entire genome of an organism, i.e., the set of genes that translate into proteins which ultimately determine all the cellular functions. Bacterial genomics has led to identification of putative gene products by a comparative approach called comparative genomics. In comparative genomics, genome of a pathogenic bacterium is compared with a non-pathogenic bacterium to identify molecular targets that can be manipulated for

designing new drugs. This approach could unravel the pathogenicity factors present in bacterium or the genes that code for bacterial survival [49]. Bacterial Genomics involves the use of NGS technologies to sequence the whole genome of bacteria and identify genetic functions using advanced computational tools [50]. Further, the data from the whole sequencing projects is deposited in searchable databases like Microbial Genomes Resources at the National Center for Biotechnology Information (NCBI), which contains more than 1000 prokaryotic genomes. Similar to this is the Genomes Online Database (GOLD) that contains comprehensive information of ongoing and completed genome sequencing projects (<http://www.genomesonline.org>). Bacterial genome annotation can be achieved using online web servers like Rapid Annotation Using Subsystem Technology (RAST) and Integrated Microbial Genomics (IMG).

Metagenomics

Metagenomics is community genomics and gives access to the genetic content of entire communities of organisms and is useful for studying ecological role of a particular microbial community present in different ecosystems. Phylogenetic and functional diversity of uncultured microorganisms present in soil, phyllosphere, ocean and acid mine drainage could be very well investigated using metagenomics [51]. Before NGS, metagenomic study involved isolating the DNA from environmental sample and cloning it into a suitable vector, transforming the clones into host bacterium and screening the resultant transformants. Later the clones were screened for phylogenetic markers like 16S rDNA and *recA* etc., this approach is often called a sequence based metagenomic approach. Another type of metagenomic study is a function based approach in which the expression of specific genes like the genes for antibiotic production or enzyme activity are screened using expression vectors [52-54]. Today most of the research laboratories use NGS for metagenomic studies which not only accelerated the process but also removed the biases and artifacts that were present in clone based approach. The NGS based metagenomic process is similar to previous approach, but the need for cloning the DNA fragments into a vector is replaced by labeling them with fluorescent adapters and measuring the signal during the single nucleotide addition with a great accuracy using a NGS machine. Further the massive data generated from the NGS machine is processed using bioinformatic tools. Accessing the microbial diversity using NGS can be achieved by using two different approaches 1) amplicon sequencing 2) shotgun metagenomics. In first approach the DNA fragments are amplified using a specific primer targeting a single gene like the 16S rDNA for eubacteria [55,56]. In second approach, the large DNA fragments or even complete genomes from organisms in a community can be characterized using shotgun libraries and amplifying them using multiple primer sets. This method is often called whole metagenome approach and requires more

expenditure and sophisticated bioinformatic analysis. An initial study which involves determining the bacterial community present in soil or marine ecosystem will require only amplicon sequencing approach targeting the 16S rDNA. Further whole metagenome sequencing can be done in order to identify the functional genes present in the bacterial community with prospects of novel metabolites that have antimicrobial or anticancer properties. In simple terms metagenomics determines the phylogeny and function of group/community of organisms thriving in different ecosystems. As mentioned earlier, the analysis of metagenomic datasets and data interpretation is a crucial step in metagenomics. Therefore data storage, sharing and statistical analysis has become an important aspect in metagenomics. Large-scale databases that process and deposit metagenomic datasets include centralized servers like MG-RAST, IMG/M and CAMERA [57-59]. Reference databases like KEGG [60], eggNOG [61], COG/KOG [62], PFAM [63] and TIGRFAM [64] can be used to give functional context to the metagenome data after assembly. In addition one can also perform the analysis of metagenomic data without depending on the centralized servers if we have strong bioinformatic skills using different software based on UNIX/Linux platforms (Table 2). The major steps in metagenomic data analysis involve assembly, binning, annotation, statistical analyses and storage. In assembly the raw sequence data after QC is processed to get full length CDS in the form of contigs. The assembly algorithms are usually based on de Bruijn graphs and can be of two types 1) reference based 2) *denovo*. The reference based assembly is fast and involves a reference genome or metagenome dataset, whereas the *denovo* assembly requires more computational power, large memory and processing time. The second step is binning which involves sorting the DNA sequences in the metagenome to represent an individual genome or genomes from closely related organisms. Through binning we can create the genome or partial genome of uncultured organisms. The process of annotation involves gene calling and has two aspects namely feature prediction and functional annotation. In feature prediction the specific features of gene of interest is identified and protein coding and non-coding sequences can be annotated. Functional annotation is a challenging task as assigning function to datasets requires accessing the protein families that have sequence homology with the metagenomic data, which relies on the protein structure prediction methods like NMR, X-ray crystallography and other biochemical methods. Now, to make meaningful results from larger metagenomic datasets requires statistical analysis which can be achieved through various statistical software packages that can perform statistical analysis based on the experimental design and objectives (Table 2). For data storage and sharing several data repositories, most of which are the centralized servers like the US National Center for Biotechnology Information (NCBI). For storing the metadata a set of guidelines is provided by Minimum Information about any (x) Sequence checklists (MIxS) [65]. MIxS is composed MIGS (the Minimum Information about a Genome Sequence), MIMS (the Minimum Information about a Metagenome Sequence) and MIMARKS (Minimum Information about a MARKer Sequence) that outlay standard formats for recording environmental and experimental data [66].

Meta-transcriptomics

Meta-transcriptomics involves random sequencing of microbial community mRNA that can be employed to identify the RNA based regulation and expression of biological signatures of a microbiome under different conditions. Expression profiling studies such as Microarray and RT-qPCR depend upon factors like primer design, array conditions and hybridization conditions. Therefore presently these are replaced by transcriptome sequencing as the sequence of a gene transcript is always the same and the meta-transcriptome data can be stored and retrieved through centralized server databases such as MG-RAST, CAMERA and IMG/M [50]. The meta-transcriptome sequencing begins with isolation of total RNA from microbial community followed by enrichment of selective RNA to be sequenced (i.e., mRNA, lincRNA and microRNA). Later, RNA is fragmented into smaller pieces (the fragment sizes in bp depend of the selected sequencing platform) followed by cDNA synthesis using reverse transcriptase and random hexamers or oligo (dT) primers which are subjected to high throughput sequencing [67]. Other methods which do not require synthesizing cDNA and involve direct sequencing of RNA were also developed to avoid the sequencing biases introduced in quantification of transcripts that arise during conversion of RNA to cDNA [68-71] (Figure 2). The processing of the meta-transcriptomic data is similar to the metagenomic data and two strategies are employed for this (1) mapping sequence reads to reference genomes and genes and (2) *de novo* assembly of new transcriptomes. Metagenomic assembly programs like SOAPdenovo [72], ABySS [73] and Velvet-Oases [74] could be very well applied for meta-transcriptome data assembly [75-78], further the function of the expressed genes can be identified using databases like KEGG [79]. Trinity is now most widely used program for *de novo* assembly of short read RNA sequence-data because of its efficiency in recovering full length transcripts [80-82]. This direct sequencing of RNA from environmental sample is emerging as a powerful technique for elucidating the in-situ activities of soil microbial communities. But extraction of RNA and its processing from soil microbial communities has its own limitations as well difficulties. One major limitation is the short half-life of RNA [83] and its variation in different species [84]. Likewise, there are difficulties which arise during isolation of RNA from soils due to inaccessibility of cells located on and within soil particles, inefficient cell lysis, the adsorption of RNA to soil particles and the presence of RNases [85]. Recently some studies have shown improved meta-transcriptomics approach to study soil microbiota by overcoming the above mentioned limitations [86,87].

Meta-proteomics

Microbial diversity studies not only involve studying the evolutionary relationship of microbial communities, but also finding the link between the phylogeny and function which requires accessing the mRNA and proteins [88-90]. The meta-proteomic approach basically employs studying the total protein isolated from the meta-community of an environmental sample in order to get the functional insight of the microbial environments. Unlike mRNA expression, expression of protein is a reflection of specific microbial activities in a

Assembly		Binning			Annotation	Statistical analyses
Reference based	<i>denovo</i>	Compositional-based	Similarity- based	Both	FragGeneScan	Primer-E package
Newbler	SOAP	Phylopythia	SOrt-ITEMS	PhymmBL	MetaGeneMark	Metastats
AMOS	Velvet	S-GSOM	MetaPhyler	MetaCluster	MetaGeneAnnotator (MGA)	R statistical package
MIRA	MetaVelvet	PCAHIER	MEGAN	-	Metagene	-
-	Meta-IDBA	TACAO	CARMA	-	Orphelia	-

Table 2: Shows the computational resources employed in different steps of metagenomic data analysis.

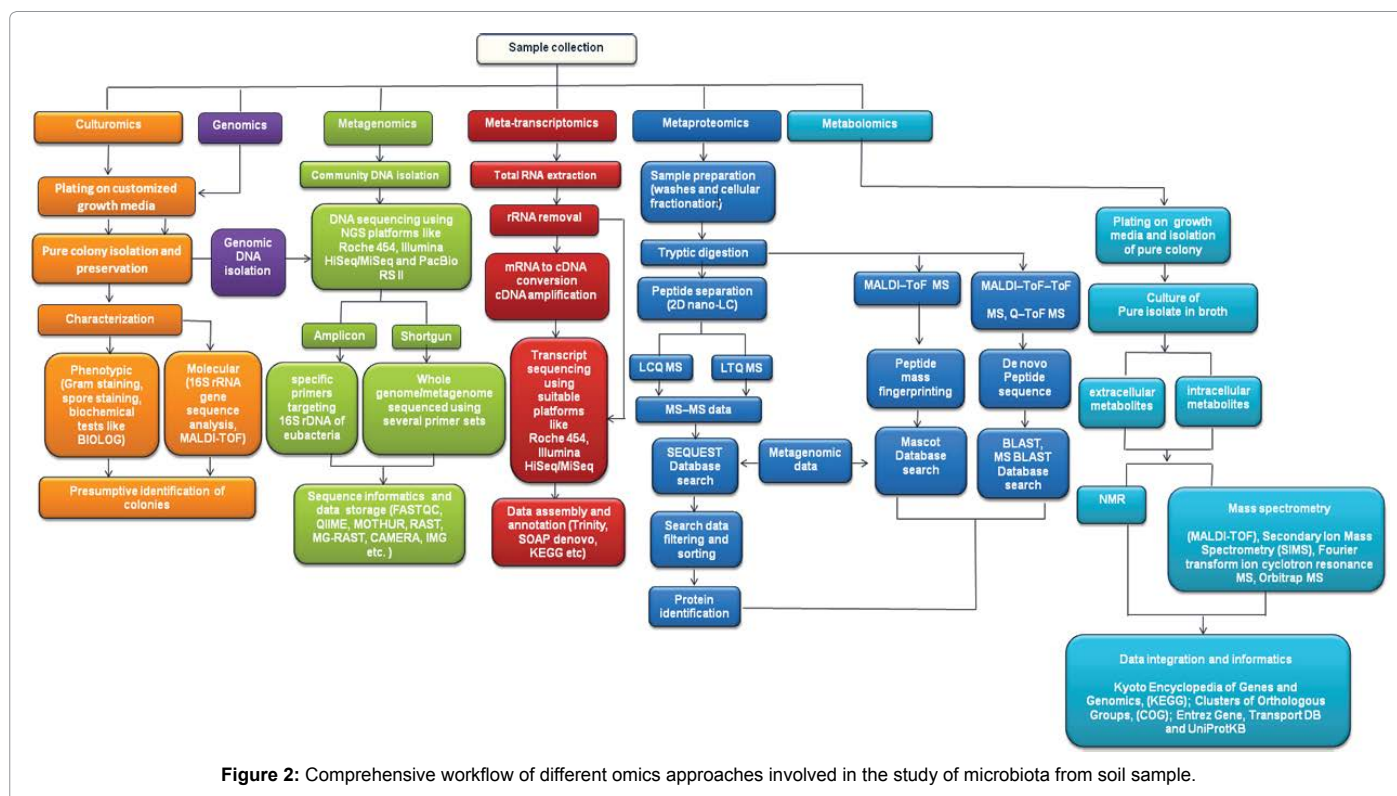


Figure 2: Comprehensive workflow of different omics approaches involved in the study of microbiota from soil sample.

given ecosystem, therefore it has great potential for functional analysis of microbial communities [91,92]. The procedure involves sample preparation, protein extraction, separation of protein or peptides using techniques like 2D gel electrophoresis and finally identification using mass spectrometry (MS) (Figure 2). As the technologies advance the challenges in protein separation and identification are met with improved extraction techniques like combination of citrate extraction method, SDS lysis method and NaOH extraction, followed by direct identification using LC-MS/MS [93]. Omics part in metabolomics involves the generation of peptide databases based on the specific peptide mass fingerprint (PMF) generated during MS analysis which can be matched to identify the proteins. In another approach the *de novo* peptide sequence can be generated from the isolated proteins by peptide sequencing. Proteome Discoverer software suite (v1.4, Thermo Fisher Scientific) and the Mascot search engine (v2.5, Matrix Science 49) are usually used for peptide identification and quantification [94].

Metabolomics

This omic technique enables us to map the entire metabolic profile of an organism. In context to microbiology, metabolomics can be regarded as “microbial metabolomics” which involves identifying the complete set of metabolites produced within a microbe that in turn reflects the enzymatic pathways and other network processes involved in the functioning of the cell [95-97]. Metabolomics assesses the interaction between cell’s genome and its environment thereby understanding the novel biosynthetic and degradative pathways exploited by the microbe in utilizing the organic content present in a particular soil ecosystem [98,99]. Metabolomics can also be employed in studying the entire microbial community present in the soil in order to understand the community responses to changes in soil through “community metabolomics” [100]. Soil metabolites include amino acids, organic acids and other natural and synthetic compounds, which can

be identified using various techniques such as GC-MS for amino acids, N-containing compounds and di-peptides and ¹H NMR, HPLC for identification of phenolics, flavonoids, carbohydrates and organic acids, etc. Once all metabolites are identified a metabolic model is generated and stored in databases (Figure 2). A handful number of computational tools facilitate this reconstruction and model process, one such is the BiGG (biochemical, genomic and genetic) knowledge server that provides the reconstructions of genome scale metabolic networks from six organisms spanning three major branches of the Tree of Life. Among them are *Escherichia coli* (a model organism), *Helicobacter pylori* (Gram negative bacterium), *Staphylococcus aureus* (Gram positive bacterium) and *Methanosarcina barkeri* (archaea) [101-105]. Another resource that integrates metabolic data is the MetaCyc database containing highly curated small molecule metabolites [106,107]. Metabolomics has its own restrictions and many gaps are needed to be filled using the upcoming technological advancements. Nevertheless, metabolomics serves as a crucial link to other omic technologies like metagenomics, meta-transcriptomics meta-proteomics, etc. and circumvents the limitations posed by them.

Culturomics

The term “culturomics” is often misleading as it resonates with the other fields of study such as the study of human society and cultural behavior. But in context to microbiology “culturomics” is a new terminology which comes with the omics package. Culturomics employs improvised culture techniques to isolate microbes in laboratory. Culturomics came in to existence from studies carried out by Lagier et al. 2011 on human gut microbiome [107]. In this study 3 people (2 lean African and one obese European), were analyzed by 212 different culture conditions of culture that recovered 32500 colonies, which was traced to 340 species of bacteria from seven phyla and 117 genera, five fungi and a giant virus (Senegalvirus). Advantages of culturing microbes

instead of detecting them from their DNA signatures include obtaining information about their viability. Further these newer strains could be exploited for several microbial byproducts that have therapeutic value. Like the NGS, next generation of culture technique is fuelled by next generation technologies like smart incubators, automated colony-picking systems, miniaturization, automated detection of microbial growth, innovative culture conditions, customized media supplements and high throughput identification using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS). Now regarding soil microbiota, high throughput culture studies are yet to come and many researchers are now trying to isolate soil microbes using soil supplements in their media. No doubt culturomics has abolished the term uncultivable or non-culturable microbes and will be a promising omics strategy to isolate novel microorganisms.

Results and Discussion

This review packs comprehensive information on present omic technologies employed in studying soil microorganisms and details its path of evolution on a timeline scale. Research on the diversity of soil microbial community has undergone a paradigm shift with the arrival of omic technologies like metagenomics, metaproteomics and metabolomics, etc. which was previously studied using conventional culture techniques, DGGE, TGGE and 16S rRNA clone library method. The review also highlights the importance of informatics along with the technology and its need in data analysis and storage. Rebirth of microbial culture techniques using improvised strategies (culturomics) and its choice over other culture independent methods is increasingly adopted by many research groups shows the importance of culture based techniques as a fundamental methodology for isolating and studying the novel microorganisms.

Conclusion

As the technological advancements take place in the field of microbial diversity studies, we gain more and more insight into the unexplored microbial community. At present both Metagenomics and Culturomics approaches must go hand in hand to unravel the unseen diversity of microbial world that will act as a platform for discovery of new generation drugs and antibiotics.

References

- Garbeva P, Van Veen JA, Van Elsas JD (2004) Microbial diversity in soil: Selection microbial populations by plant and soil type and implications for disease suppressiveness. *Annu Rev Phytopathol* 42: 243-270.
- Singh BK, Campbell CD, Sorenson SJ, Zhou J (2009) Soil genomics. *Nat Rev Microbiol* 7: 756.
- Classen AT, Sundqvist MK, Henning JA, Newman GS, Moore JAM, et al. (2015) Direct and indirect effects of climate change on soil microbial and soil microbial-plant interactions: What lies ahead? *Ecosphere* 6: 130.
- Panikov NS (1999) Understanding and prediction of soil microbial community dynamics under global change. *Appl Soil Ecol* 11: 161-176.
- Schimel J (1995) Ecosystem consequences of microbial diversity and community structure. In: Chapin FS, Körner C (eds) arctic and alpine biodiversity: Patterns, causes and ecosystem consequences. *Ecological studies (Analysis and Synthesis)*, Springer, Berlin, Heidelberg.
- Balser T, Kinzig A, Firestone M (2002) In the functional consequences of biodiversity, eds. Kinzig, A., Pacala, S. and Tilman, D. Princeton Univ. Press, Princeton.
- Roesch LF, et al. (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* 1: 283-290.
- Schmidt S, Costello Ek, Nemergut DR, Cleveland CC, Reed SC, et al. (2007) Biogeochemical consequences of rapid microbial turnover and seasonal succession in soil. *Ecology* 88: 1379-1385.
- Bååth E, Frostegård Å, Pennanen T, Fritze H (1995) Microbial community structure and pH response in relation to soil organic matter quality in wood-ash fertilized, clear-cut or burned coniferous forest soils. *Soil Biol Biochem* 27: 229-240.
- Bardgett RD, Lovell RD, Hobbs PJ, Jarvis SC (1999) Seasonal changes in soil microbial communities along a fertility gradient of temperate grasslands. *Soil Biol Biochem* 31: 1021-1030.
- Batten KM, Scow KM, Davies KF, Harrison SP (2006) Two invasive plants alter soil microbial community composition in serpentine grasslands. *Biol Invasions* 8: 217-230.
- De Deyn G, Raaijmakers C, Zoomer H, Berg M, De Ruyter P, et al. (2003) Soil invertebrate fauna enhances grassland succession and diversity. *Nature* 422: 711-713.
- Petersen SO, Debosz K, Schjonning P, Christensen BT, Elmholt S (1997) Phospholipid fatty acid profiles and C availability in wet-stable macro-aggregates from conventionally and organically farmed soils. *Geoderma* 78: 181-196.
- Steenwerth KL, Jackson LE, Carlisle EA, Scow KM (2006) Microbial communities of a native perennial bunchgrass do not respond consistently across a gradient of land-use intensification. *Soil Biol Biochem* 38: 1797-1811.
- Steinberger Y, Zelles L, Bai QY, von Lutzow M, Munch JC (1999) Phospholipid fatty acid profiles as indicators for the microbial community structure in soils along a climatic transect in the Judean Desert. *Biol Fert Soil* 28: 292-300.
- Wardle DA (2006) The influence of biotic interactions on soil biodiversity. *Ecol Lett* 9: 870-886.
- Zogg GP, Zak DR, Ringelberg DB, Macdonald NW, Pregitzer KS, et al. (1997) Compositional and functional shifts in microbial communities because of soil warming. *Soil Sci Soc Am J* 61: 475-481.
- Schierbeek A (1959) *Measuring the Invisible World: The Life and Works of Antoni van Leeuwenhoek*. London: Abelard-Schuman.
- Blevins SM, Bronze MS (2010) Robert Koch and the "golden age" of bacteriology. *Inter J Infect Dis* 14: e744-e751.
- Beveridge TJ (2001) Use of Gram stain in microbiology. *Biotech Histochem* 76: 111-118.
- Ackert L (2012) *Sergei vinogradskii and the cycle of Life: From the thermodynamics of life to ecological microbiology 1850-1950*. Netherlands: Springer Science and Business Media.
- Schmidt TM, Reiman DA (1994) Phylogenetic identification of uncultured pathogens using ribosomal RNA sequences. *Methods Enzymol* 235: 205-222.
- Kuske CR, Barns SM, Busch JD (1997) Diverse uncultivated bacterial groups from soils of the arid southwestern United States that are present in many geographic regions. *Appl Environ Microbiol* 63: 3614-3621.
- Friedmann EI (1980) Endolithic microbial life in hot and cold deserts. *Orig Life* 10: 223-235.
- Casida LE Jr, (1965) Abundant microorganisms in soil. *Appl Microbiol* 13: 327-334.
- Opfell JB, Zebal GP (1967) Ecological patterns of micro-organisms in desert soils. *Life Sci Space Res* 5: 187-203.
- Baublis JA, Wharton RA Jr, Volz PA (1991) Diversity of micro-fungi in an antarctic dry valley. *J Basic Microbiol* 31: 3-12.
- Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc Natl Acad Sci U S A* 74: 5088-5090.
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463-5467.
- Dunbar J, Takala S, Barns SM, Davis JA, Kuske CR (1999) Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. *Appl Environ Microbiol* 65: 1662-1669.
- Kielak A, Pijl AS, van Veen JA, Kowalchuk GA (2009) Phylogenetic diversity of Acidobacteria in a former agricultural soil. *ISME J* 3: 378-382.
- Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, et al. (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl Environ Microbiol* 68: 4301-4306.

33. Fleishmann RD (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269: 496-512.
34. Setubal JC, dos Santos P, Goldman BS, et al. (2009) Genome sequence of *Azotobacter vinelandii*, an obligate aerobe specialized to support diverse anaerobic metabolic processes. *J Bacteriol* 191: 4534-4545.
35. Morgan JL, Darling AE, Eisen JA (2010) Metagenomic sequencing of an *in vitro*-simulated microbial community. *PLoS ONE* 5: e10209.
36. Keller M, Zengler K (2004) Tapping into Microbiol diversity. *Nat Rev Microbiol* 2: 141-150.
37. Escobar-Zeped A, Vera-Poncede León A, Sanchez-Flores A (2015) The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Front Genet* 6: 348.
38. Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, et al. (2010) Pathway Tools version 13.0: Integrated software for pathway/genome informatics and systems biology. *Brief Bioinform* 11: 40-79.
39. Lagier JC, Armougom F, Million M, et al. (2012) Microbial culturomics: Paradigm shift in the human gut microbiome study. *Clin Microbiol Infect* 18: 1185-1193.
40. Gawad C, Koh W, Quake SR (2016) Single-cell genome sequencing: Current state of the science. *Nat Rev Genet* 17: 175-188.
41. Glenn T (2014) NGS field guide: Overview. *Mol Ecol*.
42. Fichot EB, Norman RS (2013) Microbial phylogenetic profiling with the Pacific Biosciences sequencing platform. *Microbiome* 1: 10.
43. Logares R, Haverkamp THA, Kumar S, Lanzén A, Nederbragt A, et al. (2012) Environmental microbiology through the lens of high-throughput DNA sequencing: Synopsis of current platforms and bioinformatics approaches. *J Microbiol Methods* 91.
44. Andrews S (2015) Babraham bioinformatics-fastQC a control tool for high throughput sequence data.
45. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. (2009) Introducing mothur: Open-source, platform independent, community supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75: 7537-7541.
46. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. (2010) QIIME allows analysis of highthroughput community sequencing data. *Nat Methods* 7: 335-336.
47. Huson DH, Weber N (2013) Microbial community analysis using MEGAN. *Meth Enzymol* 531: 465-485.
48. Krause L, Diaz NN, Goesmann A, Kelley S, Nattkemper TW, et al. (2008) Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Res* 36: 2230-2239.
49. Brian Fritz, Gregory A, Racznik (2002) Bacterial genomics. Potential for antimicrobial drug discovery. *BioDrugs* 16: 331-337.
50. Metzker ML (2010) Sequencing technologies - The next generation. *Nat Rev Genet* 11: 31-46.
51. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chem Biol* 5: R245-249.
52. Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, et al. (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl Environ Microbiol* 68: 4301-4306.
53. Diaz-Torres ML, McNab R, Spratt DA, Villedieu A, Hunt N, et al. (2003) Novel tetracycline resistance determinant from the oral metagenome. *Antimicrob Agents Chemother* 47: 1430-1432.
54. Galperin MY (2004) Metagenomics: From acid mine to shining sea. *Environ Microbiol* 6: 543-545.
55. Sharpton TJ (2014) An introduction to the analysis of shotgun metagenomic data. *Front Plant Sci* 5: 209.
56. Tonge DP, Pashley CH, Gant TW (2014) Amplicon-based metagenomic analysis of mixed fungal samples using proton release amplicon sequencing. *PLoS ONE* 9: e93849.
57. Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F (2010) Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb Protoc* 1: 5368.
58. Markowitz VM, Ivanova NN, Szeto E, Palaniappan K, Chu K, et al. (2008) IMG/M: A data management and analysis system for metagenomes. *Nucleic Acids Res* 36: D534-538.
59. Sun S, Chen J, Li W, Altintas I, Lin A, et al. (2011) Community cyber infrastructure for advanced microbial ecology research and analysis: The CAMERA resource. *Nucleic Acids Res* 39: D546-551.
60. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32: D277-280.
61. Muller J, Szklarczyk D, Julien P, Letunic I, Roth A, et al. (2010) eggNOG v2.0: Extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res* 38: D190-195.
62. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, et al. (2003) The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* 4: 41.
63. Finn RD, Mistry J, Tate J, Coggill P, Heger A, et al. (2010) The Pfam protein families database. *Nucleic Acids Res* 38: D211-222.
64. Selengut JD, Haft DH, Davidsen T, Ganapathy A, Gwinn-Giglio M, et al. (2007) TIGRFAMs and Genome Properties: Tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Res* 35: D260-264.
65. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, et al. (2011) Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nat Biotechnol* 29: 415-420.
66. Field D, Amaral-Zettler L, Cochrane G, Cole JR, Dawyndt P, et al. (2011) The Genomic Standards Consortium: Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *PLoS Biol* 9: e1001088.
67. Liu D, Graber JH (2006) Quantitative comparison of EST libraries requires compensation for systematic biases in cDNA generation. *BMC Bioinformatics* 7: 77.
68. Jorth P, Ciulla DM, Huang K, Haas BJ, Izard J, et al. (2014) Metatranscriptomics of the human oral microbiome during health and disease. *MBio* 5: e01012-1014.
69. Ozsolak F, Milos PM (2011) Single-molecule direct RNA sequencing without cDNA synthesis. *Wiley Interdiscip Rev RNA* 2: 565-567.
70. Ozsolak F, Platt AR, Jones DR, Reifengerger JG, Sass LE, et al. (2009) Direct RNA sequencing. *Nature* 461: 814-818.
71. Hickman SE, Kingery ND, Ohsumi TK, Borowsky ML, Wang LC, et al. (2013) The microglial sensome revealed by direct RNA sequencing. *Nat Neurosci* 16: 1896-905.
72. Li R, Yu C, Li Y, Lam TW, Yiu SM, et al. (2009) SOAP2: An improved ultrafast tool for short read alignment. *Bioinformatics* 25: 1966-1967.
73. Birol I, Jackman SD, Nielsen CB, Qian JQ, Varhol R, et al. (2009) De novo transcriptome assembly with ABySS. *Bioinformatics* 25: 2872-2877.
74. Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28: 1086-1092.
75. Shi CY, Yang H, Wei CL, Yu O, Zhang ZZ, et al. (2011) Deep sequencing of the camellia sinensis transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC Genomics* 12: 131.
76. Robertson G, Schein J, Chiu R, Corbett R, Field M, et al. (2010) De novo assembly and analysis of RNA-seq data. *Nat Methods* 7: 909-912.
77. Garg R, Patel RK, Tyagi AK, Jain M (2010) De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Res* 18: 53-63.
78. Ness RW, Siol M, Barrett SC (2011) De novo sequence assembly and characterization of the floral transcriptome in cross-and self-fertilizing plants. *BMC Genomics* 12: 298.
79. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28: 27-30.

80. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29: 644-652.
81. Ghaffari N, Sanchez-Flores A, Doan R, Garcia-Orozco KD, Chen PL, et al. (2014) Novel transcriptome assembly and improved annotation of the white leg shrimp (*Litopenaeus vannamei*), a dominant crustacean in global seafood mariculture. *Sci Rep* 4: 7081.
82. Luria N, Sela N, Yaari M, Feygenberg O, Kobiler I, et al. (2014) *De novo* assembly of mango fruit peel transcriptome reveals mechanisms of mango response to hot water treatment. *BMC Genomics* 15: 957.
83. Deutscher MP (2006) Degradation of RNA in bacteria: Comparison of mRNA and stable RNA. *Nucleic Acids Res* 34: 659-666.
84. Carvalhais LC, Dennis PG, Tyson GW, Schenk PM (2012) Application of metatranscriptomics to soil environments. *J Microbiol Methods* 91: 246-251.
85. Tveit AT, Urich T, Svenning MM (2014) Meta-transcriptomic analysis of arctic peat soil microbiota. *Appl Environ Microbiol* 80: 5761-5772.
86. Geisen S, Tveit AT, Clark IM, Richter A, Svenning MM, et al. (2015) Metatranscriptomic census of active protists in soils. *ISME J* 9: 2178-90.
87. Wilmes P, Bond PL (2004) Application of two dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ Microbiol* 6: 911-920.
88. Bastida F, Hernández T, García C (2014) Metaproteomics of soils from semiarid environment: Functional and phylogenetic information obtained with different protein extraction methods. *J Proteomics* 101: 31-42.
89. Wang HB et al. (2011) Characterization of metaproteomics in crop rhizospheric soil. *J Proteome Res* 10: 932-940.
90. Schulze WX et al. (2005) A proteomics fingerprint of dissolved organic carbon and of soil particles. *Oecologia* 142: 335-343.
91. Benndorf D, Balcke GU, Harms H, Von Bergen M (2007) Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. *ISME J* 1: 224-234.
92. Zampieri E (2016) Soil metaproteomics reveals an inter-kingdom stress response to the presence of black trufes. *Sci Rep* 6.
93. Lacerda CMR, Choe LH, Reardon KF (2007) Metaproteomic analysis of a bacterial community response to cadmium exposure. *J Proteome Res* 6: 1145-1152.
94. Bundy JG, Willey TL, Castell RS, Ellar DJ, Brindle KM (2005) Discrimination of pathogenic clinical isolates and laboratory strains of *Bacillus cereus* by NMR-based metabolomics profiling. *FEMS Microbiol Lett* 242: 127-136.
95. Garcia DE, Baidoo EE, Benke PI, Pingitore F, Tang YJ, et al. (2008) Separation and mass spectrometry in microbial metabolomics. *Curr Opin Microbiol* 11: 233-239.
96. Nicholson JK, Lindon JC (2008) Metabolomics. *Nature* 455: 1054-1056.
97. Kessler JD, Valentine DL, Redmond MC, Du M, Chan EW, et al. (2011) A persistent oxygen anomaly reveals the fate of spilled methane in the deep Gulf of Mexico. *Science* 331: 312-315.
98. Hazen TC, Dubinsky EA, De Santis TZ, Andersen GL, Piceno YM, et al. (2010) Deep sea oil plume enriches indigenous oil-degrading bacteria. *Science* 330: 204-208.
99. Jones OA, Sdepanian S, Lofts S, Svendsen C, Spurgeon DJ, et al. (2014) Metabolomic analysis of soil communities can be used for pollution assessment. *Environ Toxicol Chem* 33: 61-64.
100. Durot M, Bourguignon PY (2009) Schachter models of bacterial metabolism: Reconstruction and applications. *FEMS Microbiol Rev* 33: 164-190.
101. Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO (2009) Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* 7: 129-143.
102. Breitling R, Vitkup D, Barrett MP (2008) New surveyor tools for charting microbial metabolic maps. *Nat Rev Microbiol* 6: 156-161.
103. Thomas S, Bonchev D (2010) A survey of current software for network analysis in molecular biology. *Hum Genomics* 4: 353-360.
104. Schellenberger J, Park JO, Conrad TM, Palsson BO (2010) BiGG: A Biochemical Genetic and Genomic knowledge base of large scale metabolic reconstructions. *BMC Bioinformatics* 11: 213.
105. Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, et al. (2010) The MetaCyc database of metabolic pathways and enzymes and the Bio Cyc collection of pathway/genome databases. *Nucleic Acids Res* 38: D473-479.
106. Greub G (2012) Culturomics: A new approach to study the human microbiome. *Clinical Microbiology and Infection* 18: 1157-1159.
107. Hayden EC (2013) Researchers glimpse microbial 'dark matter' Single-cell DNA sequencing unlocks trove of microbial diversity. *Nature News and Comment*.