Perspective

# Various Methods and Resources for Metagenomics Agriculture Data Analysis

Francesca Pasolli*

*Department of Agricultural Sciences, University of Naples Federico II, Naples, Italy*

## DESCRIPTON

Agriculture's linked microbial ecosystem has a significant impact on productivity. Metagenomics is now known to play a crucial part in microbial world research and thanks to advances in omics technology by revealing the uncultured microbial populations that are prevalent in the environment. As a diagnostic tool, metagenomics can be used to identify specific signature loci of animal and plant diseases as well as advantageous microbes in samples. Methods that drive microbial community population research on heterogeneity and complexity through time are used to reveal metagenomics (community and environmental genomics), which was originally described as a key demand in microbial analysis. To comprehensively characterize a microbial community is the aim of any metagenomic study. This aids in understanding the composition of the microbial community, the functional roles played by each microbial constituent, and the information regarding intra-species heterogeneity.

With the development of computer technologies and sequencing, metagenomics analysis underwent a revolution. The Next-Generation Sequencing (NGS) cost decline and rise in analytical applications have had a significant impact on the discipline of microbiology. The characterization of the microbial community advances in utilizing several sequencing techniques. The ability of next-generation sequencing to target a variety of distinctive characteristic loci of plant and animal diseases as well as advantageous bacteria from samples makes it a potent diagnostic tool in agriculture metagenomics. Metagenomic research employs a variety of techniques, including:

## Amplicon-based methods

These are read-based reconstructions of the metagenome's functional and taxonomic elements. It is a targeted technique that amplifies a taxonomic marker rRNA gene prior to sequencing in order to target it specifically. Amplicon sequences are PCR products that are unique to each individual organism and define a region inside a gene. This approach uses the 16s rRNA, RNA polymerase, and heat shock protein 90 as well as protozoan and fungus identification with the 18s rRNA gene to characterize bacteria. Viral RNA polymerase is employed in viral population investigations. A sample can be deep sequenced using next-generation technologies and a variety of amplicons. To find the conserved gene of a particular organism, these sequences are combined into a reference database. By using this technique, the organism is identified together with its relative abundance. Overall, this method becomes more sensitive for strain identification when it uses amplicons to target a specific species. Even so, there is a chance for bias in this method, such as in the abundance ratio, which could result from PCR amplification that artificially inflates the counts of some taxa or from the use of universal primers that could be misleading or fail to identify other organisms.

## Shotgun metagenomics

An environmental sample's whole genome is sequenced as part of a broader strategy that enables taxonomic profiling of the bacteria and a better understanding of their functions. In contrast to amplicon-based techniques, the sequenced reads are compared to the reference database, which is distinct and substantial. This method reflects the better microbial community structure in a sample and is computationally intensive and less biassed. Additionally, this technique has the ability to distinguish between closely related species or strains. This technique can also be utilized to find new ammonia oxidizers for fertilizer as well as stable microbial community structures in agricultural soil.

Following sample collection and DNA/RNA extraction, the processing of biological samples, sequencing, and bioinformatics analyses utilizing various tools are all part of the overall workflow of metagenomics. Each phase, which involves a variety of methods and equipment, is thoroughly detailed. Metagenomic analysis entails extracting all of a sample's nucleic acids, depending on the study's goals. Depending on the technique, the microbial population to be examined, such as bacteria or viruses, is either targeted with DNA or RNA. When studying viruses using metagenomics, virions are retrieved by first eliminating biological components, then using nucleases to remove non-viral

nucleic acids and extract viral nucleic acids intact within the nucleocapsid. Metagenomics also involves the collecting of geographical data, such as GPS coordinates, depth, height, date of sample collection, extraction procedure, and clone library details, as well as environmental sample associated data, or metadata, which includes temperature, pH, and salinity.

## Sequencing

The sequence of the samples, the size of the sample, and the cost of all affect sequencing, which is an important stage. There are numerous platforms for sequencing. Illumina and 454 pyrosequencing are the two most often utilized sequencing systems in metagenomics study.

## Bioinformatics analysis

The computational analysis is dependent on sequencing techniques, which employ two primary strategies for reconstructing the genome: Read assembly, then taxonomic and functional classification ad hoc sequencing, method of read based

based reconstruction sequencing of amplicons. Bioinformatics techniques and tools have been specifically created and developed as a result of extensive research on the sequencing of bacterial 16s rRNA and fungal ITS amplicons as well as whole sample sequencing by shotgun metagenomics. This is crucial to comprehend many microorganisms that live in various environments.

## Tools

Depending on how the data is sequenced, such as amplicon-based or shotgun, different computational tools and techniques are used. When analyzing shotgun data, many pieces of software are employed at various stages.

The following algorithms, starting with the conventional method that incorporates overlap-based assembly and the Kmer technique, support genome assembly. Apriori algorithm, association rule mining algorithm, binning method like MaxBin 2, SPHINX, MetaCluster 5, sequence classification algorithm, and others are also available.