JUST AGRICULTURE
THE FUTURE OF AGRI INNOVATION

Genome Analysis: from uncertainty to Precision of Soil Microbiome

Barkha Sharma, Harshita Singh, Shalini Tiwari

Ph.D Research Scholar, G.B. Pant University of Agriculture & Technology, Pantnagar

Article ID: 040

"Soil is a region on the earth crust where geology and biology unite."

Three factors i.e., physical fertility, chemical fertility and biological fertility comprise soil fertility. Biological fertility includes the organisms that live and interact with the other components present in soil. The soil organisms are highly dynamic and vary greatly in number but are least explored fertility factor. Soil constitutes diverse organisms like bacteria, archaea, fungi, algae, insects, annelids, protozoa and other invertebrates which exhibit close relationship with each other and plants. Plants secret out their photosynthetically fixed carbon into direct surroundings, i.e., spermosphere, phyllosphere, rhizosphere and mycorrhizosphere, thereby nourishing the microbial community and controlling their activities. It has been estimated that one gram of soil contains approximately 10^9 bacteria, 10^6 fungi, 10^4 nematodes, 10^8 actinomycetes, 10^5 microalgae, 10^5 protozoa and 10^2 other invertebrates. If agricultural production is to meet the needs of a growing world population, a better insight of soil microbiology is required. Soil microbial communities influence the agricultural productivity and ecological processes such as the nutrient cycling, biodegradation, enhancement of crop yield and suppression of soil-borne pathogens.

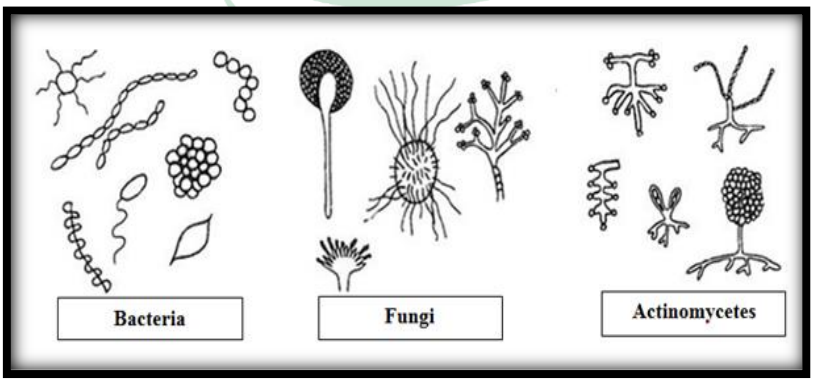

Figure 1: Morphological features of different microorganisms

Molecular Characterization of Soil Microflora

The purpose of phylogenetic analysis is to understand the past evolutionary path of organisms. Even though we will never know for certain the true phylogeny of any organism, phylogenetic analysis provides best assumptions, thereby providing a framework for various disciplines in microbiology. Due to the technological advancement of molecular biology and in computational science, accurate inference of the phylogeny of a gene or organism seems possible in the near future.

A. **Genetic fingerprinting:** Developments in DNA-based molecular techniques have revolutionized the field of bacterial taxonomy and bacterial identification system. The techniques include Restriction Fragment Length Polymorphism (RFLP), Plasmid profiling, Ribotyping, Amplified Ribosomal DNA Restriction Analysis (ARDRA), Pulse Field Gel Electrophoresis (PFGE) and Randomly Amplified Polymorphic DNA (RAPD).

- **PFGE** uses in situ lysis of bacterial whole-cells in agarose plugs to release DNA which protect the chromosomal DNA from mechanical damages. The DNA fragments are visualized on the gel following staining.
- **RFLP** is a genetic analysis that allows individual to be identified on the basis of their unique pattern of restriction enzyme cutting in specific regions of DNA.
- **Inribotyping** rRNA, rDNA or gene specific oligonucleotides are used as probes against enzyme restricted DNA. Due to multiple copy number of 16S rRNA gene (1-14 depending upon the group of microbe), a complex profile is obtained in riboprinters.
- **ARDRA** employs digestion of amplified ribosomal DNA with different restriction enzymes and a profile is obtained using the combination of these patterns.
- **AFLP** (Amplified Fragment Length Polymorphism) is another offshoot of RFLP in which specific adaptors are ligated to enzyme restricted DNA which is subsequently amplified using primers from the adaptor and restriction site-specific sequences.
- **RAPD** also referred to as arbitrary Primed PCR (AP-PCR) is another technique in which short primer sequences (octa- to decamer) randomly anneal to genomic DNA and initiate amplification.

16S rRNA Sequencing: 16S Ribosomal RNA sequencing is widely used in microbiology studies to identify the diversities in prokaryotic organisms as well as other organisms and thereby studying the phylogenetic relationships between them. The 16S rRNA gene provides

a species specific signature sequence which is useful for bacterial identification process. For eukaryotic species identification 18s rRNA gene sequencing is performed. The 16S ribosomal RNA gene codes for the RNA component of the 30S subunit of the bacterial/archaeal ribosome. Different microbial species have one to multiple copies of the 16S rRNA gene. 16S rRNA gene sequencing is by far one of the most common methods targeting housekeeping genes to study bacterial phylogeny and genus/species classification. According to proposed guidelines for bacterial classification, strains with less than 97% similarity in 16S rRNA gene sequence represent different bacterial species; the ones that show more than 97% similarity should be classified using an alternative approach. 16S sequence analysis includes analytical processing (trimming, screening and aligning sequences), followed by microbial profiling by comparisons to 16S rRNA sequences in public databases, or from operational taxonomic units (OTU) using the frequency distribution of sequences found in bins (using an accepted threshold of a 3% dissimilarity level for species and 5% for genus). The taxonomic depth chosen is frequently situation-dependent.

B. G+C ratio :

Analysis of DNA G+C ratio or mole percent of guanosine and cytosine is one of the classical genotyping methods in the bacterial systematics. The variation in the percent GC content is not more than 3% within a well-defined species and not more than 10% within a well-defined genus and it varies from 24 to 76% in the bacterial world.

C. Nucleic acid hybridization:

Nucleic acid hybridization technique is based on a comparison between whole genome of two bacterial species. According to this technique, soil bacterial species generally would include the strain with 70% or greater DNA-DNA hybridization values with 5°C or less ΔTm values and both the values must be considered.

D. FISH (Fluorescence In Situ Hybridization):

FISH detects nucleic acid sequencesby a fluorescently labeled probe that hybridizes specifically to its complementary target sequence within the intact cell. The procedure involves the following steps:

(i) Specimen fixation,

(ii) Preparation of the sample,

(iii) Hybridization with the specific probes for the detection of the respective target sequences,

(iv) Washing to remove the unbound probes, and

(v) Mounting, visual representation and then, documentation of results.

Dyes commonly used for FISH in the microbiology are fluorescein-derivates (Fluorescein Isothiocyanate (FITC), 5-(-6-)carboxyfluorescein–*N*- hydroxysuccimide-ester (FluoX)) and rhodamine-derivates (Tetramethyl–Rhodamine–Isothiocyanate (TRITC), Texas Red.

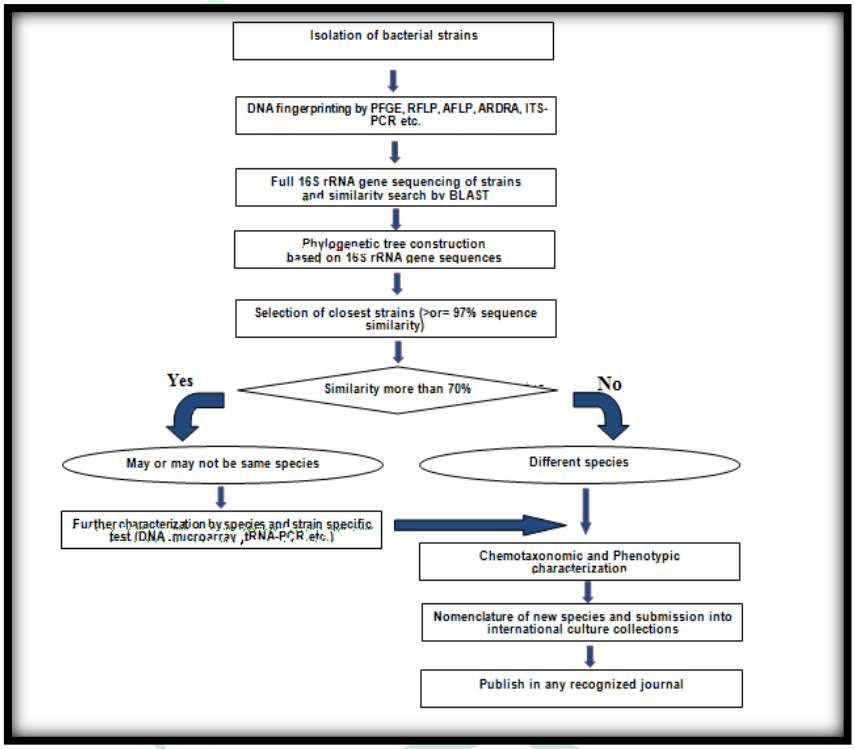

Figure 2: Flow diagram depicting the step-by step procedure for taxonomical characterization of newly isolated strains

Omics techniques:

The terms 'Ome' and 'Omics' are derivations of the suffix -ome, which has been appended to a variety of previously existing biological terms to create names for fields of endeavor like genome, proteome, transcriptome and metabolome that have some actual meaning in specific case.

Genomics: Genomics is the study of an organism's genome, or genetic material. Genomics is an emerging area of molecular biology where novel techniques were employed in order to get an insight of the complex, biological function of the genome.

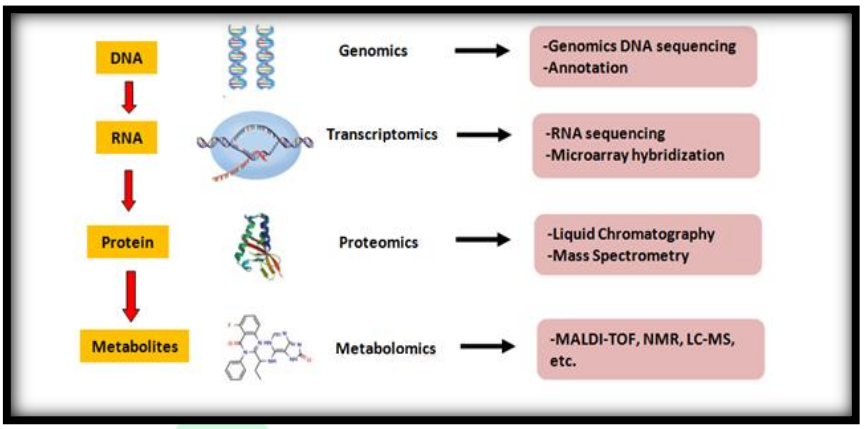

Figure 3: Schematic representation of omics dogma describing the network the processing of DNA, RNA, proteinsand metabolites (Yellow) and method used (Pink)

Metagenomics:

Metagenomics is the culture-independent technique to analyze complete genome of microbial communities present in a particular environment sample. The term is obtained from the statistical view of meta-analysis (the process of statistically combining separate analyses) and genomics (whole genome analysis of organism). Metagenomics can be used to address the challenge of studying 99% unculturable prokaryotes in the environment, as yet, only 1% of the total microbial population is cultureable.

Transcriptomics:

Transcriptomics is the overall scanning of the approx. 50,000 known genes that are transcribed into mRNA from the three-billion-letter human genome. Each cell expresses different genes at different times during its development and under various physiological conditions. The advantage of this new technique is that instead of examining one gene at a time, it analyses the complete transcriptome in one single go, which is the full set of all mRNA molecules present at a particular time in a defined cell population.

Proteomics:

Proteomics is the implementation of technologies to examine gene products, *proteins*, on a large scale which includes protein expression profiles,protein networks and protein modifications in relation to function and biological processes of cell e.g. development, health and disease.

Metabolomics:

It is the integrated study of the metabolome, the entire set of small metabolites (e.g. hormones, metabolic intermediates, signalling molecules, and secondary metabolites) of

biological sample. The metabolic state reflects the encoded product of genome which is being modified by environmental factors. NMR and other spectroscopic methods such as MS, GC/MS can be used to generate metabolic profiles.

Integrated omics technology:

Data integration with multiple omics: Integrative study is exercised for roughly two purposes: first, prediction of gene functions and second, characterization of the systematic interaction of biological processes. In this section, we focus on the current opportunities and challenges of integrative studies with multiple omics data in plant biology.