

METAGENOMICS THROUGH NEXT GENERATION SEQUENCING- A LEAP IN THE EXPLORATION OF THE HIDDEN WORLD OF ENDOPHYTIC MICROBIAL COMMUNITY

Abstract

Endophytes are microbes residing in plant tissue and they are clandestine masters of plant growth regulation. Quantifying taxon frequency and community membership of this microflora will allow us to assess their potential at unprecedented depth. The generation of highly accurate genomic data with advanced computational tools points to application-based perspectives of these phytomicrobes. Whole metagenome analysis and metagenome analysis of 16S rDNA sequences are indispensable technologies to explore this valuable resource of microorganisms. Applications of metagenomic analysis in the field of endophyte study are a promising aspect of research focusing on plant-microbe interaction. Different sequencing technologies and bioinformatic tools employed for the metagenome analysis are discussed in brief. Recognising the advantages and challenges of metagenomics and an awareness of emerging technologies for sequencing will make endophytic microbiome studies more appreciable. This is the era of tera-base-scale metagenomics, therefore the expansion of data in different public domain databases and bioinformatic tools for comparative analysis will lead to ice-breaking discoveries of novel microbes, genes, proteins or enzymes.

Author

Sithara K Urumbil

Department of Botany,
Little Flower College, Guruvayoor,
P O Puthanpalli-680103, Kerala, India
sithara@littleflowercollege.edu.in

METAGENOMICS THROUGH NEXT GENERATION SEQUENCING- A LEAP IN THE EXPLORATION OF
THE HIDDEN WORLD OF ENDOPHYTIC MICROBIAL COMMUNITY

I. INTRODUCTION

Plants are complex microecosystems in which different niches are present on both external surfaces and internal tissues (Azevedo *et al.*, 2000). Microorganisms are ubiquitous and they have been identified even in extreme environmental conditions. The magnitude of the microbial genome present in the biosphere is considered to be three-fold times higher than the plant and animal cells (Whitman *et al.*, 1998). According to Strobel *et al.* (2004), almost all the plant species existing on the planet harbours one or more microorganisms as endophytes. There is almost zero percentage chance of finding a plant without endophyte mentioned as a rare exception (Partida-Martinez and Heil, 2011) and if present, would be more susceptible to pathogens and difficult to withstand environmental stress (Timmusk *et al.*, 2011). The concept of ‘plant microbiome’ including the plant and symbiont, altered scientific approaches of the research community and they focused on the interactions taking place in this co-existence.

Environmental conditions play an influential role in the development and survival of plants. In this sustainable development process, plants show radiant interactions with different microbial communities which may or may not be beneficial. The established development of plants associated with microbes indicates the high range of plant-microbe interaction which enhances plant growth. Symbiotic or mutualistic microbial communities residing in plant tissues have a significant role in the production of plant growth hormones, fighting pathogens, nutrient uptake and stress management. Isolation and culturing of endophytic microbes in growth medium and their screening for functional characterisation have always been done to elucidate a clear idea about this plant-microbe interaction. One of the limitations of culture-dependent screening is the hindrance of endophytes to grow in invitro culture conditions (Hong *et al.*, 2019). Specific studies on culture-based isolation point to another struggle that in some cases after initial isolation bacterial colonies show difficulties to grow in subcultures. The researchers working in this area speculated that the same host-specific metabolites were essential for the proper growth and multiplication of these endophytes which may not be available in further subculture conditions (Eevers *et al.*, 2015)

The limitations of traditional culture-based isolation urged the need for advanced techniques including genome-based studies. Handelsman *et al.*, (1998) initiated studies on the total genome of the microbial samples from the environment and mentioned it as the study focusing on ‘the genome of the total microbiota found in nature’. Now the metagenome analysis attained a wider perspective which includes the entire genome of various microbial communities like bacteria and fungi in environmental samples. The environmental samples include soil samples, ecto- and endo-microbial colonies from plants and animals, water samples etc. The focus of metagenomic analysis mainly relies on microbial population diversity and structure, phylogenetic trends, multifaceted interactions, relationships of microbial communities and the applied aspects of identifying beneficial microbes.

Metagenomic analysis is more relevant in endophytic microbial population studies. The rhizobia population of fungi and bacteria can be easily isolated and cultured in invitro conditions (Bafana, 2012). The microbes which reside inside plant tissues show much more complicated host-microbe interactions and they are very much adapted and established in

**METAGENOMICS THROUGH NEXT GENERATION SEQUENCING- A LEAP IN THE EXPLORATION OF
THE HIDDEN WORLD OF ENDOPHYTIC MICROBIAL COMMUNITY**

host tissues. Therefore some of them are modified into an extent of obligate endophytes which are very difficult to cultivate under normal culture conditions (Croes et al., 2013, Kamnev et al., 2005).

II. METAGENOMICS: AN EMERGING STRATEGY FOR ENDOPHYTIC MICROFLORA STUDIES

Genomic analysis of individual microbial colonies is in use for screening of novel metabolites and its large-scale industrial applications. The whole metagenome-based analysis widened the identification and application of beneficial microbes residing inside plant tissues. These microbes were reported to have a specific role in plant growth and health promotion (Rat et al., 2021, Eid et al., 2021, Lacava et al., 2022). Advances in metagenomics attained a fast pace of progression through Next Generation Sequencing Technologies (NGST). Biodiversity cluster analysis of endophytic microbial populations was conducted in different plant species with beneficial aspects like in the field of health care and agriculture. Metagenomic analysis was employed to identify microbial endophytic populations from different plants (Table 1). The metagenome studies led to the evaluation of applied aspects of plant-microbial interactions. The endophytes possess a specific role in host growth and development. As part of that it divulges the information of some rare microbes and their metabolic potential. Identification of gene sequences associated with novel metabolites makes the industrial application of that microorganism more significant. Production of plant growth hormones on a large scale, diverse antibiotics for fighting pathogenic microbes for the host, biochemicals involved in phytoremediation and different stress management systems for the host are some of the functions associated with specific genes from endophytes identified in metagenome analysis

Table 1: Metagenomic studies of bacterial and fungal endophytes from different plants

Plant	Organism	Endophytes Identified Using Metagenome Analysis	References
<i>Aloe vera</i>	Bacteria	Proteobacteria, Firmicutes, Actinobacteria and Bacteroidetes	Akinsanya et al., 2015.
<i>Panax ginseng</i>	Bacteria	Proteobacteria and Actinobacteria	Hong et al., 2019
<i>Senna italica</i>	Bacteria	Cyanobacteria, Proteobacteria, Actinobacteria, Firmicutes	Alsaedi et al, 2022
<i>Eleusine coracana</i> (L.) Gaertn. subsp. coracana	Bacteria	Proteobacteria, Actinobacteria and Bacteroidetes	Prasannakumar et al., 2019
<i>Quercus mongolica</i> and <i>Quercus serrata</i>	Fungi	Peltaster, Cladosporium and Monochaetia	Nguyen, et al., 2021
<i>Oryza sativa</i> L.	Bacteria	Proteobacteria, Firmicutes, Cyanobacteria and Actinobacteria	Sengupta et al., 2017
<i>Vitis vinifera</i> L.	Bacteria	Alstonia, Burkholderia and Pseudomonas genera	Campisano et al., 2014

**METAGENOMICS THROUGH NEXT GENERATION SEQUENCING- A LEAP IN THE EXPLORATION OF
THE HIDDEN WORLD OF ENDOPHYTIC MICROBIAL COMMUNITY**

<i>Paeonia</i> Sect. Moutan	Bacteria	Proteobacteria, Firmicutes, Bacteroidetes, Acidobacteria, Actinobacteria	Yang et al., 2017
<i>Armoracia rusticana</i>	Fungi	Plectosphaerella, Thanatephorus, Podospora, Monosporascus, Exophiala, Setophoma	Plaszko et al., 2022
<i>Dendrobium officinale</i>	Fungi	Ascomycota	Zhuoyan et al., 2017
<i>Huperzia serrata</i>	Fungi	Ascomycota, Basidiomycota	Fan et al., 2020

III. METAGENOMICS AND TECHNOLOGY ADVANCEMENT

Molecular studies were initiated when the DNA-based identification and characterisation of environmental microbial samples came into existence. This was a great leap in microbial studies which involves large and diverse community studies. 16S rRNA gene sequences, which are signature sequences of particular microbes make the identification of microbes at the species level easy but only with microbes which are available in invitro culture conditions. 16S rRNA retrieved from shotgun sequencing methods provides better information in species characterisation. However, this was not highly efficient for the identification of rare species in communities (Shah et al., 2011).

The understanding of interactions of large microbial populations increased in recent times due to effective sequencing technology advancements like Next Generation Sequencing technology (NGS) (Lozupone, and Knight, 2007). The 16S rRNA sequencing technique and shotgun sequencing technology were effective for characterising individual colonies and small sample populations but with a relatively high cost. NGS with the new Illumina sequencing platform changed the pace of endophyte research with Illumina mass communities that can be sequenced and analysed which generated huge data sets (Caporaso et al., 2012). 16S rRNA amplicon of microbial communities (discussing bacterial endophytes) consists of conserved and hypervariable regions. These hypervariable regions show high specificity and it attracted the attention of the researchers. Therefore, high throughput sequencing technologies can be employed for sequencing these regions which will reveal the diversity as well as phylogenetic relationships of these communities.

The sequence read length while sequencing 16S rRNA is one of the key factors in molecular studies including 16SrRNA amplicon. High efficiency better throughput sequencing instruments and techniques are available like 454 GS pyrosequencing, MiSeq Illumina etc. Over time better sequencing techniques with high throughput per turn and reduced error rate were developed by combined efforts of researchers working in the field of molecular biology and bioinformatics (Loman et al., 2012).

IV. OVERVIEW OF WHOLE METAGENOMIC ANALYSIS METHODOLOGY

Sample Collection: Sample collection is a key process for the metagenome study of endophytic microbiota. Different plant parts like root, stem and leaf can be collected based on the objectives of the study (Calvert, 2021). The growth period, age of the host plant, geographical location and environmental conditions and the time taken for the transport of

METAGENOMICS THROUGH NEXT GENERATION SEQUENCING- A LEAP IN THE EXPLORATION OF THE HIDDEN WORLD OF ENDOPHYTIC MICROBIAL COMMUNITY

plant materials for further studies are some of the factors to be considered for sample collection (Dos Reis et al., 2022). Surface sterilization of the collected samples to be done to make sure that any microbes on the surface of plant parts were eliminated.

DNA Extraction: Plant parts like root, stem and leaf can be used for the DNA extraction which should be crushed and homogenised. Endosphere microbiological studies reached new dimensions through metagenomic analysis because it includes the study of hidden and diverse microbial populations. Therefore, direct DNA extraction from the sample is preferable even though the purity is low. Different advanced and efficient DNA isolation protocols and isolation kits are available according to the requirements and sample (Wang et al., 2011). The quality and purity of the isolated DNA can be analysed and verified by PCR.

Metagenomic Library Construction: Next Generation Sequencing technology workflow includes the library construction through DNA fragmentation, ligation of adaptor oligos and library quality analysis (Navgire et al., 2022). Quality of metagenomic library construction depends on the sample, complexity or heterogeneity of community and the procedure used for library construction. Endophyte metagenome may vary based on the host plant, Geographical locations and environmental conditions and based on that the amount of DNA extracted and library preparations procedure should be standardised. Different library preparation methods or procedures were analysed in detail by Gaulke et al., (2021) and used different communities with varying complexities.

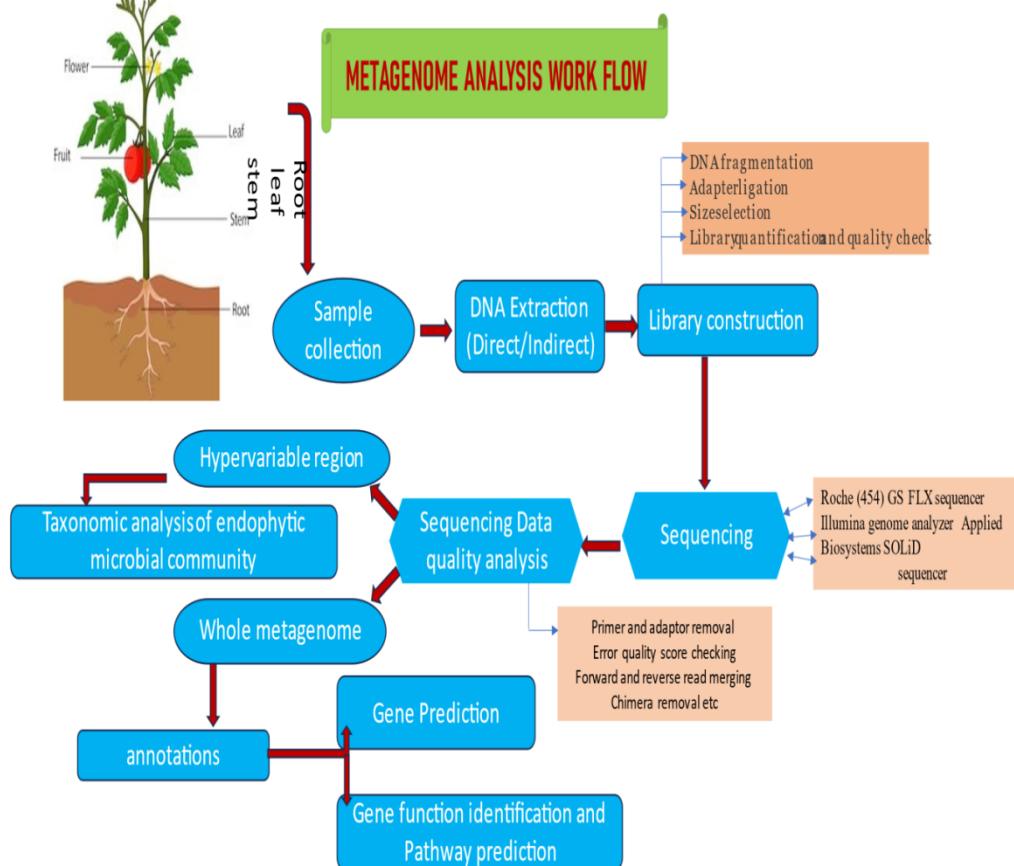
Sequencing: A gradual shift of sequencing technologies from shot-gun sequencing methods to next-generation sequencing occurred in the past two decades. 54/Roche and the Illumina/Solexa systems are the commonly used NGS technologies for large microbial community analyses. Millions of sequence reads can be processed parallel and don't require any vector-based cloning methods for library construction are some points which make NGS a researcher-friendly technology. Three major sequencers available are the Roche (454) GS FLX sequencer, Illumina genome analyser and Applied Biosystems SOLiD sequencer (Mardis et al., 2008). One of the widely used amplicon sequencing techniques is Illumina which will be performed in HiSeq2000 and MiSeq platforms. The HiSeq and MiSeq platforms differ in the scale of sequenced reads generated in time. The HiSeq2000 produces 450 Gb paired-end base pairs per day whereas MiSeq generates 1.5 Gb bp per day.

Sequence data quality analysis and optimization: Different bioinformatics tools are available to analyse the sequence quality. Forward and reverse reads can be trimmed to maintain the required base-pair length. The primer sequences should be removed before further analysis. Error checking and calculation of maximum expected error from the quality score ($EE = \sum (10^{-Q/10})$) is essential for identifying higher quality reads. High throughput sequences generated will be processed for further preliminary quality filtering. Forward and reverse read merging and chimera removal are the two major steps before progressing to the preparation of rarefaction curves. Rarefaction curves denote the saturation values when the curve reaches the plateau based on the number of reads that passed the preliminary filtering (Kuźniar et al., 2020).

Biodiversity and Taxonomic analysis: In metagenome analysis for taxonomic classification and biodiversity analysis only some focused areas in the 16S rDNA are to be analysed and

**METAGENOMICS THROUGH NEXT GENERATION SEQUENCING- A LEAP IN THE EXPLORATION OF
THE HIDDEN WORLD OF ENDOPHYTIC MICROBIAL COMMUNITY**

sequenced. In the 16S rRNA coding region of the genome nine highly conserved and nine hypervariable regions (V1-V9) are present which are specific to the particular microorganism. Sequencing of these regions using NGS technology allows taxonomic identification and classification of microbial flora. Comparison of these sequences is possible because a huge amount of data is already available in the public database domain (Chakravorty et al., 2007). Different bioinformatic tools with high accuracy and precision are available for 16S rDNA sequence analysis. The latest version of RDP data including Naïve Bayesian Classifier is one of the promising tools for microbial taxonomy analysis. Further taxonomic analysis and read counts can be done by bioinformatic tools like QIIME, UPARSE, MOTHUR, DADA2, (MED) etc. Identification of Operational Taxonomic Units (OTUs) enables the study of phylogeny progression (Niu et al., 2018). Identification and categorization of OTUs using the Ribosomal Database Project Classifier indicates the heterogeneity and richness of the microbial community. Different diversity index analyses can be employed for the study and representation of these communities like the Alpha diversity index, Shannon index, Chao 1 index, Simpson Index, Goods Coverage index etc.



High volume data generation and a tremendous increase of endophytic studies require more advanced bioinformatic tools for a better understanding of heterogeneity and structure of microbial endospheric population. Advanced phylogenetic analysis tools like PhyloPythiaS and Kraken are available to elucidate the evolutionary relationship which will unravel new areas of colonization potential of microbial communities in plant tissues. These phylogenetic

**METAGENOMICS THROUGH NEXT GENERATION SEQUENCING- A LEAP IN THE EXPLORATION OF
THE HIDDEN WORLD OF ENDOPHYTIC MICROBIAL COMMUNITY**

analyses use different algorithms for taxonomic profiling like K-mer (compares with organisms of a wide range of clades) and marker genes (aligned with preselected clade-specific marker genes). An algorithm based on the read map approach compares the metagenomic data with the available databases.

Gene Function Analysis: Metagenomic analysis especially in the case of the microbial population inside plant tissues exposed an unexplored area of research. the plant-microbe interactions. For a deeper understanding of the involvement of the microbial genome in plant growth and metabolism, NGS technologies can be employed. Metagenomics with metatranscriptomics and metabolomics clearly explains the role of endophytes in various plant growth regulatory activities like the production of plant growth hormones, biotic and abiotic stress management etc.

The sequence data generated through whole metagenome data can be processed with different bioinformatic tools for function-driven analysis. Binning is the term used in metagenome analysis denoting comparison and grouping of sequenced data from closely related genomes available in the databases to generate contigs. Different binning algorithms are available like Phylopythia, S-GSOM, PCAHIER, TACAO, IMG/M, MG-RAST, MEGAN, CARMA, SOrt-ITEMS and MetaPhyler (Thomas et al., 2012). Feature prediction and functional annotation are mandatory for identifying genes coding for different proteins and regulatory compounds. Advanced research in molecular bioinformatics enhanced studies based on gene structure prediction and developed programs focused on three different perspectives- ab initio methods, homology-based methods, and combined methods. EasyGene and GeneMark are the two commonly used gene prediction programs (Ejigu and Jung, 2020). Progressive sequencing technologies like NGS and TGS generated huge amounts of sequence data which are easily available for comparison in various public domain databases. Homology similarity searches connect the sequenced data under investigation with these already deposited data, which makes the comparison and functional characterisation of gene annotations from endophytes an effortless one. Some commonly used databases for gene structure and function studies are summarised in the table below.

Table 2: Different Databases Available for Gene Functional Annotation Studies

Data Base	Significance	Link to the Data Base	Reference
KEGG (Kyoto Encyclopedia of Genes and Genomes)	For systematic analysis of gene functions, linking genomic information with higher order functional information	http://www.genome.ad.jp/kegg/	Kanehisa and Goto, 2008
KEGG Orthology (KO) database	Molecular functions of genes and proteins are associated with ortholog groups and stored in the KEGG Orthology	http://www.kegg.jp/ http://www.genome.jp/kegg/	Kanehisa et al., 2016

*METAGENOMICS THROUGH NEXT GENERATION SEQUENCING- A LEAP IN THE EXPLORATION OF
THE HIDDEN WORLD OF ENDOPHYTIC MICROBIAL COMMUNITY*

GenBank	Contains publicly available nucleotide sequences for more than 420 000 formally described species	(www.ncbi.nlm.nih.gov/genbank/)	Sayers et al., 2019
European Nucleotide Archive (ENA)	Organized and presented the world's public domain output of sequence information	http://www.ebi.ac.uk/ena/	Cochrane et al., 2013
DNA Data Bank of Japan (DDBJ)	Member of the International Nucleotide Sequence Database Collaboration (INSDC), in collaboration with the US National Center for Biotechnology Information and the European Bioinformatics Institute	http://www.ddbj.nig.ac.jp	Kodama et al., 2018
UniProtKB	A protein sequence database that combines UniProtKB/Swiss-Prot (over 560,000 manually curated sequences) and UniProtKB/TrEMBL provides access to proteomes for over 84 thousand species with completely sequenced genomes	https://www.uniprot.org/proteomes/ https://www.ebi.ac.uk/proteins/api/doc/	Consortium, 2019
InterPro database	classifies protein sequences into families and predicts the presence of functionally important domains and sites	http://www.ebi.ac.uk/interpro/	Mitchell et al., 2019
Gene3D	database of globular domain annotations for millions of	http://gene3d.biochem.ucl.ac.uk	Lewis et al., 2018

**METAGENOMICS THROUGH NEXT GENERATION SEQUENCING- A LEAP IN THE EXPLORATION OF
THE HIDDEN WORLD OF ENDOPHYTIC MICROBIAL COMMUNITY**

	available protein sequences		
PROSITE	consists of documentation entries describing protein domains, families and functional sites	http://prosite.expasy.org/	Sigrist et al., 2012
NCBI's Conserved Domain Database (CDD)	Database for protein domains conserved during molecular evolution	https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml	Lu et al., 2020
Reactome Knowledgebase	Pathways of signal transduction, transport, DNA replication, metabolism and other cellular processes	www.reactome.org	Fabregat et al., 2016
MetaCyc database	metabolic pathways and enzymes	MetaCyc.org	Caspi et al., 2016,

V. METAGENOME ANALYSIS FOR PLANT ENDOPHYTE INTERACTION STUDIES

Metagenome analysis uncovered a vast array of functionally active and interactive endophytic microbial communities and the majority of them are either bacteria or fungi. The endophytes and their genes characterised through next-generation sequencing revealed the involvement of endophytes in plant growth. Functional metagenomics focuses on the identification of genes that code for bioactive components which have multifaceted applications. Functional metagenomics opens up a new arena of research in which the screening of a targeted gene and its function can be analysed from huge sequence information without any culture conditions. Designing and large-scale industrial production of essential bioactive metabolites using beneficial non-pathogenic endophytic microorganisms will always be a great advantage to biomedical drug development systems.

The majority of the endophytes isolated and identified are either rhizosphere or soil microorganisms. Genome analysis of endophytes reveals genes with specific functions like identification of plant root exudates, internalization, establishment and colonization inside the plant tissues without causing any apparent adverse effect on the host (Martinez-Garcia et al., 2015, Pinski et al., 2020). Gene annotations were identified in the phytohormone Indole Acetic Acid production pathway from tryptophan in endophytes which indicates its active role in the growth promotion of the host (Andrea et al., 2015). Gene functional annotations were identified and some genes which indicated its presence in endophytes are expected to be due to lateral gene transfer, which points to plant-microbe interactive co-existence and co-evolution.

METAGENOMICS THROUGH NEXT GENERATION SEQUENCING- A LEAP IN THE EXPLORATION OF THE HIDDEN WORLD OF ENDOPHYTIC MICROBIAL COMMUNITY

Biotic or abiotic stress management is always an important aspect of study in agriculture and crop breeding centres. Isolation and identification of endophytes which can harness diverse stress conditions will be beneficial for its host plant for managing the stress. Production of a wide range of metabolites in coexistence with the metabolic pathways of host metabolism marks the role of endophytes as a potential candidate for screening and isolation of bioactive metabolites. Genome analysis of isolated microbes and whole metagenome analysis of the endophytic community give the details of specific gene annotations responsible for the synthesis of diverse compounds with role in plant growth and health regulation.

Table 3: Gene annotations reported from endophytes with functional role various host metabolic pathways

Endophyte	Gene Identified	Functional Role of the Gene	References
Endophytic bacterium <i>Pseudomonas putida</i>	gene for mannitol dehydrogenase	an enzyme involved in defence against phytopathogenic fungi	Wu et al., 2011
Endophytic bacterium <i>Enterobacter</i> sp, which are important	genes for amino acid/iron transport, hemolysin, and hemagglutinin	Iron transport and interaction with host	Taghavi et al., 2010
Endophytes from rice plants	Glutathione synthases and Glutathione-S-transferases	Managing reactive oxygen species and stress tolerance	Sessitsch et al., 2012
endophytic fungus of tea <i>Pestalotiopsis fici</i>	polyketide synthases, non-ribosomal peptide synthases (NRPSs), dimethylallyl tryptophan synthases, putative PKS-like enzymes	Secondary metabolite production	Wang et al., 2015
<i>Bacillus subtilis</i> strain RS10	Phosphate transport (pstACS)	Phosphate transport	Wang et al., 2022
Endophytic bacteria of <i>Emilia sonchifolia</i>	haloacetate dehalogenase (dehH), haloalkane dehalogenase (dha A)	dehalogenate toxic halogenated pollutants	Urumbil and Anilkumar, 2023
Endophytic bacteria of <i>Lotus corniculatus</i> and <i>Oenothera biennis</i>	alkane monooxygenase, alkane hydroxylase	Hydrocarbon degradation	Pawlik et al., 2017

VI. NGS TO TGS FOR METAGENOMIC SEQUENCING

De novo sequencing technology development based on single molecule sequencing and large parallel sequencing representing the third-generation sequencing is a new milestone in sequencing based large sample analysis. Different methods can be adopted for TGS sequencing like single-molecule fluorescence sequencing and nanopore sequencing. In single molecule fluorescence sequencing, fluorescently labelled nucleotides are detected based on light wavelength and peak value where as in nanopore sequencing electrification and passage through the nanopore are used to detect the ATGC sequences. TGS does not require PCR amplification and large sequence data can be generated in less time are the advantages of TGS over NGS. But the increased rate of error and high cost are the disadvantages of TGS. Research works were going on these aspects to improve the efficiency of third generation sequencing and TGS is in its infancy (Zhang et al., 2021).

VII. METAGENOMICS: ADVANTAGES AND CHALLENGES FOR ENDOPHYTES SCREENING

DNA sequencing, identification of new microorganisms, and sequence databases are increasing at an exponential rate due to the decrease in the cost of NGS sequencing techniques like Illumina. Endophytic microorganisms reside in a very complex environment by interacting with host cell responses and co-existing with other microbial communities. The isolation and culturing of these microbes are very difficult in invitro culture conditions because of these complex associations. The metagenome-based analysis in that perspective is a new hole for the entry into the unexplored world of endophytic microorganisms that are hitherto unknown. Metagenome searches unravelled the complexities of phytomicrobiomes and detected some unexpected novel microbes. Whole metagenome screening opened a new gateway to the identification of specific gene annotations from endophytes with non-axenic bioprospecting potential. Metagenomic analysis has become a benchmark technology for endophyte study because of unbeatable data generation, analysis and output with low-volume samples like small pieces of root, stem or leaf.

Sequencing technologies used for metagenomic data generation are in a surge and new technologies are developing at a swift tempo. Up to the present time, some technical issues exist in large-scale endophytic microbiome genome sequencing due to its complexities and multifarious identity. In taxonomic profiling, low-abundance microbes may be excluded from the genome data and in gene expression profile analysis the chance of missing some key gene annotations is a bottleneck. The recent upswing in the metagenomic study of endophytic microflora adds information to the databases. Therefore, in some cases, the absence of information about very rare novel microbes or genes may be a struggle for the researcher.

CONCLUSION

The awareness about microbial flora existing inside the plant tissues is elevated by modern metagenome studies which analyse large quantities of environmental samples. Metagenomic analysis provides high taxonomic resolution of the microbial flora and generates high throughput sequencing data for future studies. Distinctive technologies are emerging in this area of research. Metagenome analysis progresses through a specific

**METAGENOMICS THROUGH NEXT GENERATION SEQUENCING- A LEAP IN THE EXPLORATION OF
THE HIDDEN WORLD OF ENDOPHYTIC MICROBIAL COMMUNITY**

pipeline where biotechnology and bioinformatic tools are merged for data generation and analysis. DNA extraction, library construction, amplification, sequencing, sequence data quality analysis, filtering, assembly, binning, contig creation, taxonomic profiling and annotation are some attention-worthy steps in metagenomics. An enormous number of protocols and bioinformatic tools are developed by researchers for screening environmental samples. A combination of different analysis platforms is used for different studies based on the nature of the sample.

Endophytes marked their involvement in almost all fields of plant growth promotion. As part of that, endophytes are involved in the production of plant growth regulators, different bioactive compounds with putative roles. Identification of endophytes with plant growth promotion activity is an added advantage for its application in the agriculture sector. Stress management and stress tolerance generated either by biotic factors or abiotic factors is a major issue faced by plants for their sustenance. Active participation of endophytes in these aspects makes the plant-microbe interaction an inevitable one. Diversified aspects of stress tolerance come under the topics of drought tolerance, biodegradation of pollutants and prevention of the establishment of other pathogenic microorganisms inside the plant tissue. The metagenome analysis focusing on gene function studies contributed new insights into the untapped area of plant-microbe interaction. NGS technology with advanced computer algorithms and programs provided genetic information on the functional expression of genes associated with endophytes harbouring inside plant tissues. Integration of more than one 'omics' tool like transcriptomic, proteomics and metabolomics with metagenomics can surpass some major struggles in endophyte studies.

REFERENCES

- [1] Akinsanya MA, Goh JK, Lim SP, Ting AS (2015) Metagenomics study of endophytic bacteria in *Aloe vera* using next-generation technology. *Genom Data*.10(6):159-63. DOI: 10.1016/j.gdata.2015.09.004.
- [2] Alsaedia ZS, Ashya RA, Shamib AY, Majeeda MA, Alswata A M, Bazc L, Baeshena MN, Jalala RS (2002) Metagenomic study of the communities of bacterial endophytes in the desert plant *Senna italica* and their role in abiotic stress resistance in the plant. *Brazilian Journal of Biology*. 82: e267584 | <https://doi.org/10.1590/1519-6984.267584>.
- [3] Andrea F, Robert O, Asaf S, Jeremy S, Zareen K, Regina R, Neil F, Erika L, Igor G, Sharon D (2015) Genome sequence of the plant growth promoting endophytic yeast *Rhodotorula graminis* WP1. *Frontiers in Microbiology*. 6: DOI=10.3389/fmicb.2015.00978
- [4] Azevedo JL, Maccheroni WJr, Pereira JO, Araujo WL (2000) Endophytic microorganism: a review on insect control and recent advances on tropical plants. *Electron. J. Biotechno. (online)*. 3(1).
- [5] Bafana A (2012) Diversity and metabolic potential of culturable root-associated bacteria from *Origanum vulgare* in sub-Himalayan region. *World J Microb Biot*. 29: 63–74.
- [6] Calvert J (2021) Sampling of plant material to study endophytes in small, large, and woody plants. *Methods Mol Biol*. 2232:37-42. doi: 10.1007/978-1-0716-1040-4_4.
- [7] Campisano A, Antonielli L, Pancher M, Yousaf S, Pindo M, Pertot I (2014) Bacterial endophytic communities in the grapevine depend on pest management. *PLoS One*. 9: e112763. doi: 10.1371/journal.pone.0112763.
- [8] Caporaso JG, Lauber CL, Walters WA, et al., (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*. 6:1621-1624
- [9] Caspi R, Billington R, Ferrer L, Foerster H, Fulcher C A, Keseler IM, et al., (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res*. 44:D471–D480. doi: 10.1093/nar/gkv1164
- [10] Chakravorty S, Helb D, Burday M, Connell N, Alland D (2007) A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods*.69:330–9 .

**METAGENOMICS THROUGH NEXT GENERATION SEQUENCING- A LEAP IN THE EXPLORATION OF
THE HIDDEN WORLD OF ENDOPHYTIC MICROBIAL COMMUNITY**

- [11] Cochrane G, Alako B, Amid C, Bower L, Cerdeño-Tárraga A, Cleland I, Gibson R, Goodgame N, Jang M, Kay S, Leinonen R, Lin X, et al., (2013) Facing growth in the European Nucleotide Archive. *Nucleic Acids Research*. 41(1):30–D35 <https://doi.org/10.1093/nar/gks1175>
- [12] Consortium U (2019) UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res*. 47: D506–D515. doi: 10.1093/nar/gky1049.
- [13] Croes S, Weyens N, Janssen J, Vercampt H, Colpaert JV, Carleer R, Vangronsveld J (2013) Bacterial communities associated with *Brassica napus* L. grown on trace element-contaminated and non-contaminated fields: a genotypic and phenotypic comparison. *Microb Biotechnol*. 6: 371–384.
- [14] Dos Reis JBA, Lorenzi AS, do Vale HMM (2022) Methods used for the study of endophytic fungi: A review on methodologies and challenges, and associated tips. *Arch Microbiol*. 204(11):675. doi: 10.1007/s00203-022-03283-0.
- [15] Eevers N, Gielen M, Sánchez-López A, Jaspers S, White JC, Vangronsveld J, Weyens N (2015) Optimization of isolation and cultivation of bacterial endophytes through addition of plant extract to nutrient media. *Microbial Biotechnol*. 8:707–715.
- [16] Eid AM, Fouda A, Abdel-Rahman MA, Salem SS, Elsaied A, Oelmüller R, Hijri M, Bhowmik A, Elkelish A, Hassan SE-D (2021) Harnessing bacterial endophytes for promotion of plant growth and biotechnological applications: An overview. *Plants*. 10:935. <https://doi.org/10.3390/plants10050935>
- [17] Ejigu GF, Jung J (2020) Review on the computational genome annotation of sequences obtained by Next-Generation Sequencing. *Biology (Basel)*. 18; 9(9):295. doi: 10.3390/biology9090295.
- [18] Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, et al., (2016) The reactome pathway knowledgebase. *Nucleic Acids Res*. 44: D481–D487. doi: 10.1093/nar/gkv1351
- [19] Fan S, Miao L, Li H, Lin A, Song F, Zhang P (2020) Illumina-based analysis yields new insights into the diversity and composition of endophytic fungi in cultivated *Huperzia serrata*. *PLoS ONE*. 15(11): e0242258. doi: 10.1371/journal.pone.0242258.
- [20] Gaulke CA, Schmeltzer ER, Dasenko M, Tyler BM, Vega Thurber R, Sharpton TJ 2021. Evaluation of the Effects of Library Preparation Procedure and Sample Characteristics on the Accuracy of Metagenomic Profiles. *mSystems*. 6(5): e0044021. doi: 10.1128/mSystems.00440-21.
- [21] 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–R249. doi: 10.1016/s1074-5521(98)90108-9.
- [22] Hong CE, Kim JU, Lee JW, Bang KH, Jo IH (2019) Metagenomic analysis of bacterial endophyte community structure and functions in *Panax ginseng* at different ages. *3 Biotech*. 9(8):300. doi: 10.1007/s13205-019-1838-x.
- [23] Kamnev AA, Tugarova AV, Antonyuk LP, Tarantilis PA, Polissiou MG, Gardiner PHE (2005) Effects of heavy metals on plant-associated rhizobacteria: comparison of endophytic and non-endophytic strains of *Azospirillum brasilense*. *J Trace Elem Med Biol*. 19: 91–95.
- [24] Kanehisa M, Goto S (2008) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 28:27–30. doi: 10.1093/nar/28.1.27
- [25] Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 44: D457–D462. doi: 10.1093/nar/gkv1070
- [26] Kodama Y, Mashima J, Kosuge T, Kaminuma E, Ogasawara O, Okubo K, Nakamura Y, Takagi T (2018) DNA data bank of Japan: 30th anniversary. *Nucleic Acids Res*. 46: D30–D35. doi: 10.1093/nar/gkx926.
- [27] Kuźniar A, Włodarczyk K, Grządziel J, Goraj W, Gałązka A, Wolińska A (2020) Culture-independent analysis of an endophytic core microbiome in two species of wheat: *Triticum aestivum* L. (cv. ‘Hondia’) and the first report of microbiota in *Triticum spelta* L. (cv. ‘Rokosz’). *Syst Appl Microbiol*. 43(1):126025
- [28] Lacava PT, Bogas AC, Cruz FdPN (2022) Plant growth promotion and biocontrol by endophytic and rhizospheric Microorganisms from the tropics: A review and perspectives. *Front Sustain Food Syst*. 6:796113. doi: 10.3389/fsufs.2022.796113
- [29] Lewis TE, Sillitoe I, Dawson N, Lam SD, Clarke T, Lee D, Orengo C, Lees J (2018) Gene 3D: Extensive prediction of globular domains in proteins. *Nucleic Acids Res*. 46: D435–D439. doi: 10.1093/nar/gkx1069.
- [30] Loman N, Misra R, Dallman T et al., (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol*. 30:434–439. <https://doi.org/10.1038/nbt.2198>
- [31] Lozupone CA, Knight R (2007) Global patterns in bacterial diversity. *Proc Natl Acad Sci*. 104:11436–11440

**METAGENOMICS THROUGH NEXT GENERATION SEQUENCING- A LEAP IN THE EXPLORATION OF
THE HIDDEN WORLD OF ENDOPHYTIC MICROBIAL COMMUNITY**

- [32] Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, et al., (2020) CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res.* 48:D265–D268. doi: 10.1093/nar/gkz991.
- [33] Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends Genet.* 24(3):133–141. doi: 10.1016/j.tig.2007.12.007
- [34] Martinez-Garcia PM, Ruano-Rosa D, Schiliro E, Prieto P, Ramos C, Rodríguez-Palenzuela P, et al. (2015) Complete genome sequence of *Pseudomonas fluorescens* strain PICF7, an indigenous root endophyte from olive (*Olea europaea* L.) and effective biocontrol agent against *Verticillium dahlia*. *Stand. Genomic Sci.* 10, 10. doi: 10.1186/1944-3277-10-10
- [35] Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI, et al. (2019) InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47:D351–D360. doi: 10.1093/nar/gky1100.
- [36] Navgire GS, Goel N, Sawhney G et al., (2022) Analysis and Interpretation of metagenomics data: an approach. *Biol Proced Online.* 24(18). <https://doi.org/10.1186/s12575-022-00179-7>
- [37] Nguyen MH, Shin KC, Lee JK (2021) Fungal community analyses of endophytic fungi from two Oak Species, *Quercus mongolica* and *Quercus serrata*, in Korea. *Mycobiology.* 9: 385–395.
- [38] Niu S-Y, Yang J, McDermaid A, Zhao J, Kang Y, Ma Q (2018) Bioinformatics tools for quantitative and functional metagenome and metatranscriptome data analysis in microbes. *Briefings in Bioinformatics.* 19(6):1415–1429
- [39] Partida-Martinez LP, Heil M (2011) The microbe-free plant: fact or artifact? *Front Plant Sci.* 2:100.
- [40] Pawlik M, Cania B, Thijs S, Vangronsveld J, Piotrowska-Seget Z. Hydrocarbon degradation potential and plant growth-promoting activity of culturable endophytic bacteria of *Lotus corniculatus* and *Oenothera biennis* from a long-term polluted site. *Environ Sci Pollut Res Int.* 2017 Aug;24(24):19640-19652. doi: 10.1007/s11356-017-9496-1. Epub 2017 Jul 6. PMID: 28681302; PMCID: PMC5570797.
- [41] Pinski A, Betekhtin A, Hupert-Kocurek K, Mur LAJ, Hasterok R (2019) Defining the genetic basis of plant-endophytic bacteria interactions. *Int J Mol Sci.* 20;20(8):1947. doi: 10.3390/ijms20081947.
- [42] Plaszkó T, Szűcs Z, Cziáky Z, Ács-Szabó L, Csoma H, Géczi L, Vasas G, Gonda S (2022) Correlations between the metabolome and the endophytic fungal metagenome suggests importance of various metabolite classes in community assembly in horse radish (*Armoracia rusticana*, Brassicaceae) Roots. *Front. Plant Sci.* 13:921008. doi: 10.3389/fpls.2022.921008
- [43] Prasannakumar MK, Mahesh HB, Desai RU, Kunduru B, Narayan KS, Teli K, Puneeth ME, Rajadurai RC, Parivallal B, Babu GV (2020) Metagenome sequencing of finger millet-associated microbial consortia provides insights into structural and functional diversity of endophytes. *3 Biotech.* 10(1):15. doi: 10.1007/s13205-019-2013-0.
- [44] Rat A, Naranjo HD, Krigas N, Grigoriadou K, Maloupa E, Alonso AV, Schneider C, Papageorgiou VP, Assimopoulou AN, Tsafantakis N, Fokialakis N, Willems A (2021) Endophytic bacteria from the roots of the medicinal plant *Alkanna tinctoria* Tausch (Boraginaceae): Exploration of plant growth promoting properties and potential role in the production of plant secondary metabolites. *Front Microbiol.* 12:633488. doi: 10.3389/fmicb.2021.633488
- [45] Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I (2019) GenBank. *Nucleic Acids Res.* 47: D94–D99. doi: 10.1093/nar/gky989.
- [46] Sengupta S, Ganguli S, Singh PK (2017) Metagenome analysis of the root endophytic microbial community of Indian rice (*O. sativa* L.) *Genom Data.* 12:41–43. doi: 10.1016/j.gdata.2017.02.010.
- [47] Sessitsch A, Hardoim P, Döring J, Weilharter A, Krause A, Woyke T, Mitter B, Hauberg-Lotte L, Friedrich F, Rahalkar M, Hurek T, Sarker A, Bodrossy L, van Overbeek L, Brar D, van Elsas JD, Reinhold-Hurek B (2012) Functional characteristics of an endophyte community colonizing rice roots as revealed by metagenomic analysis. *Mol Plant Microbe Interact.* 25(1):28-36. doi: 10.1094/MPMI-08-11-0204. PMID: 21970692.
- [48] Shah N, Tang H, Doak TG, Ye Y (2011) Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. *Pac Symp Biocomput.* 2011:165–176.
- [49] Sigríst CJA, De Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I (2012) New and continuing developments at PROSITE. *Nucleic Acids Res.* 41:D344–D347. doi: 10.1093/nar/gks1067.
- [50] Strobel G, Daisy B, Castillo U, Harper J (2004) Natural products from endophytic microorganisms. *J. Nat. Prod.* 67: 257-268.

**METAGENOMICS THROUGH NEXT GENERATION SEQUENCING- A LEAP IN THE EXPLORATION OF
THE HIDDEN WORLD OF ENDOPHYTIC MICROBIAL COMMUNITY**

- [51] Taghavi S, van der Lelie D, Hoffman A, Zhang YB, Walla M D, Vangronsveld J, et al., (2010) Genome sequence of the plant growth promoting endophytic bacterium *Enterobacter* sp 638. PLoS Genet. 6: e1000943. doi: 10.1371/journal.pgen.1000943
- [52] Thomas T, Gilbert J, Meyer F (2012) Metagenomics - a guide from sampling to data analysis. Microb Inform Exp. 2(1):3. doi: 10.1186/2042-5783-2-3.
- [53] Timmusk S, Paalme V, Pavlicek T, Bergquist J, Vangala A, Danilas T, Nevo E (2011) Bacterial distribution in the rhizosphere of wild barley under contrasting microclimates. PLoS ONE. 6:e17968.
- [54] Urumbil sk, Anilkumar M (2023). Unveiling multiple strategic bioremediation potential of bacterial endophytes from medicinal plant *Emilia sonchifolia*(linn.) Dc. Through metagenomic detailing Eur. Chem. Bull. 12(SI:6), 522-531doi: doi: 10.48047/ecb/2023.12.Si6.047
- [55] Wang Y, Zhao Q, Sun Z, Li Y, He H, Zhang Y, Yang X, Wang D, Dong B, Zhou H, Zhao M, Zheng H (2022) Whole-genome analysis revealed the growth-promoting mechanism of endophytic bacterial strain Q2H1 in potato plants. Front Microbiol. 13:1035901. doi: 10.3389/fmicb.2022.1035901
- [56] Wang TY, Wang L, Zhang JH, Dong WH (2011) A simplified universal genomic DNA extraction protocol suitable for PCR. Genet Mol Res. 10: 519–525. doi: 10.4238/vol10-1gmr1055
- [57] Wang X, Zhang X, Liu L, et al., (2015) Genomic and transcriptomic analysis of the endophytic fungus *Pestalotiopsis fici* reveals its lifestyle and high potential for synthesis of natural products. BMC Genomics.16:28 (2015). <https://doi.org/10.1186/s12864-014-1190-9>
- [58] Whitman WB, Coleman DC, Wiebe WJ (1998) Prokaryotes: The unseen majority. Proc. Natl. Acad. Sci. USA. 95:6578-6583.
- [59] Wu X, Monchy S, Taghavi S, Zhu W, Ramos J, Van der Lelie D (2011) Comparative genomics and functional analysis of niche specific adaptation in *Pseudomonas putida*. FEMS Microbiol Rev. 35:299–323. doi: 10.1111/j.1574-6976.2010.00249.x
- [60] Yang R, Liu P, Ye W (2017) Illumina-based analysis of endophytic bacterial diversity of tree peony (*Paeonia* Sect. *Moutan*) roots and leaves. Braz J Microbiol. 48:695–705. doi: 10.1016/j.bjm.2017.02.009.
- [61] Zhang L, Chen F, Zeng Z, Xu M, Sun F, Yang L, Bi X, Lin Y, Gao Y, Hao H, Yi W, Li M, Xie Y (2021) Advances in metagenomics and its Application in environmental microorganisms. Front Microbiol. 12:766364. doi: 10.3389/fmicb.2021.766364
- [62] Zhuoyan LWZ, Liu Y, Hu X, Guo Y, Junfeng L (2017) Application of high-throughput internal transcribed spacer rRNA metagenomics analysis in deciphering endophytic fungi diversity of *Dendrobium Officinale*. J Bio Mater Bioenerg 11:106–118