Metabarcoding Applications in Microbiota

Prashanth Kumar A
Scientist (Plant Pathology), JVR Horticultural Research Station, Malyal, SKLTSHU, Telangana
*Corresponding Author: adluriprashanth01@gmail.com

Metabarcoding refers to a molecular technique used to identify and quantify many species within a given environmental sample. Metabarcoding has demonstrated its efficacy as a cost-efficient approach for characterizing microbial communities. This method enables the assessment of biodiversity within samples, offering a high level of taxonomic precision. Additionally, it facilitates the comparison of sample communities that have been treated with various treatments. From a bioinformatic perspective, shotgun metagenomics is somewhat more challenging to handle because of its higher storage requirements and processing demands. Currently, this method is widely employed as the predominant molecular technique for characterizing microbiota in environmental samples (Francioli *et al*., 2021). This work provides novel insights into the examination of plant illnesses characterized by intricate causes occurring in both the aboveground and belowground components of plants.The efficacy of DNA metabarcoding is mostly contingent upon the careful selection of a suitable DNA marker gene. The selection of primers should possess suitable coverage of the target group, effective exclusion of outgroups, and the capability to differentiate taxa based on the nucleotide variability of the amplified marker (Tedersoo *et al*., 2022). The utilization of customer-designed primer pairs, as exemplified in the genome-enhanced detection and identification (GEDI) approach outlined below, is a viable option. However, it is equally crucial to incorporate universal primers on barcodes to facilitate cross-study comparisons. In order to analyze fungal communities, researchers commonly employ DNA barcoding and metabarcoding techniques. Among the various markers available, the internal transcribed sequence (ITS) region of the ribosomal RNA (rRNA) is widely utilized due to its numerous copies and ability to provide accurate species-level identification in most fungal groups (Tedersoo *et al*., 2022). However, it is important to note that certain groups, such as Trichoderma, *Fusarium*, or Oomycetes, may not exhibit complete resolution at the species level using the ITS region.

The fundamental aspects of analyzing metabarcoding sequencing data encompass (i) the process of grouping sequences and (ii) the attribution of taxonomic or functional information through comparison with existing databases. Sequence readings are grouped together based on their similarity. In most metabarcoding studies, operational taxonomic units (OTUs) are made by grouping readings that are similar to each other by a certain amount, usually between 95% and 99%. Typically, a threshold of 97% homology is employed. In order to enhance the accuracy and consistency of taxonomic identification, researchers have devised amplicon sequence variant (ASV) methodologies. For future assessments of the community, these methods only look at unique, identical sequences. They focus on OTUs (operational taxonomic units) that are 100% alike. After the clustering process, the reads need to be allocated to a taxon or a function, depending on the reference databases, in order to make taxonomic or functional assignments. The quality of the information obtained improves as the databases become more carefully selected and comprehensive. The identification of ITS sequences from fungal and other eukaryotic organisms is commonly conducted using the UNITE reference data set, which is accessed through the website https://unite.ut.ee (accessed on March 16, 2022). This database is widely utilized due to its extensive curation and inclusion of diverse non-

fungal sequences, which aid in the differentiation of fungi from other eukaryotes (Anslan *et al*., 2018; Nilsson *et al*., 2019). Functional assignments can be made using either the FUNGuild database (Nguyen et al., 2016) or the FungalTraits database (Polme *et al*., 2020). It is noteworthy to emphasize that next-generation sequencing (NGS) microbiome-based diagnostics provide a substantial volume of data, necessitating the utilization of machine learning or other resources for the evaluation of human, plant, and soil health (Oh *et al*., 2020; Krause *et al*., 2021; Wilhelm *et al*., 2022). The efficacy of bioinformatic methodologies in the retrieval of fungal strains and the corresponding proportions of the retrieved strains exhibited significant heterogeneity. The sequence analysis tools, namely USEARCH and VSEARCH, were successful in detecting nearly all strains present in the mock community. However, both methods tended to exaggerate the richness of the community. On the other hand, the DADA2 tool demonstrated more accuracy in retrieving both the true richness and composition of the mock community. The first two methods are better suited for identifying specific species, whereas the third method is better suited for doing studies on community ecology (Pauvert *et al*., 2019).

Fusarium Head Blight (FHB) is a major wheat disease caused by many *Fusarium* species, many of which can make mycotoxins (Karlsson *et al*., 2021). FHB can affect both above-ground and below-ground tissues. The wheat ear fungus community in a topographically varied environment was characterized using Illumina MiSeq with V3 Chemistry (Schiro *et al*., 2019). In other study, employed PacBio CCS long read sequencing to investigate the alterations occurring in *Fusarium* spp. by targeting a combination of the highly variable internal transcribed spacer (ITS) and the D1-D2-D3 portions of the large subunit (LSU) region. The investigation examines the impact of various cover crops on crop leftovers. In a similar manner, the amplification of bacterial 16S rRNA, fungal ITS, and *Fusarium* spp. is performed using the Illumina MiSeq platform in maize. The utilization of TEF1 areas provides valuable insights into the intricate epidemiology of *Fusarium* head blight (FHB) through the identification and concurrent appearance of various phytopathogenic and beneficial bacteria in maize stalks cultivated in conjunction with wheat. The investigation involved the analysis of fungal and bacterial community profiles in wheat straws that were intentionally inoculated with *Zymoseptoria tritici*. This allowed for comprehensive knowledge of the interactions and dynamics between the pathogen and the entire microbial community for a specified period of time (Kerdraon *et al*., 2019). The utilization of next-generation sequencing (NGS) techniques in grapevine cultivation presents numerous advantageous applications in determining the microbial species composition that is pertinent to the process of winemaking (Singh *et al*., 2019; Griggs *et al*., 2021).

The utilization of next-generation sequencing (NGS) techniques has been employed to characterize clusters of grapevine trunk diseases, specifically *Eutypa, Esca, Botryosphaeria*, *Phomopsis* dieback, and black foot. The accurate characterization of these diseases is essential in order to determine and implement the most suitable control strategies. The utilization of Illumina short read technology, together with optimized and universal primers designed to target both the ITS1 and ITS2 rDNA regions, has been employed to validate the existence of the most prominent species associated with each condition. Furthermore, this approach has facilitated the identification of species that have not yet been classified within this particular complex (Morales *et*

al., 2018). In contrast to the relatively high level of attention given to wood diseases, the detection of Vitis phylloplane diseases using next-generation sequencing (NGS) has garnered comparatively less interest. However, it is anticipated that the utilization of NGS for this purpose will increase in the near future (Cureau *et al*., 2021). Next-generation sequencing (NGS) was employed to investigate the impact of elicitors or biocontrol agents on the populations of microorganisms residing on the surface of leaves (Gobbi *et al*., 2020; Nerva *et al*., 2019). Apple Replant Disease (ARD) is a significant ailment characterized by a multifaceted origin that mostly impacts fruit trees, specifically apples and other members of the Rosaceae family that are replanted in a location previously used for agriculture (Mazzola *et al*., 1998). The primary species involved in many apple locales globally are oomycetes, including *Pythium* spp. and *Phytophthora* spp., as well as fungi, particularly *Cylindrocarpon* spp. and *Rhizoctonia solani*.

The management of acid rock drainage (ARD) poses challenges attributed to the limited availability of approved chemical treatments, which are further compounded by the intricate nature of the disease's causative factors. The application of next-generation sequencing (NGS) methods has revealed substantial disparities in microbial composition between newly established sites and replanted sites, particularly in populations of beneficial bacteria such as *Burkholderia* spp., Microcoleus, Nocardioides, sulfur-oxidizing bacteria, and those involved in nitrogen cycling. Additionally, these differences have been observed following the use of green manure with Brassica spp. The user's text does not contain any information. Next-Generation Sequencing (NGS) has been employed in other studies to investigate and describe the pathobiome of many crops, including oaks, ginseng, tomato, strawberry, potato, banana and ramie (*Boehmeria nivea*). The metabarcoding technique is commonly employed for the investigation of oomycetes, with a particular focus on *Phytophthora* spp. The second box The utilization of high-throughput sequencing (HTS) has significantly enhanced our capacity to evaluate biodiversity in fungal communities across various ecosystems, including soil, phylloplane, air, and water. This technological advancement has particularly facilitated the monitoring of a specific pathogen's dynamics within its dynamic environment. For instance, it enables the examination of the pathogen's behavior following chemical or biological interventions, as well as the investigation of the impact of climate change or agricultural practices on disease development. Understanding the structure and function of microbiota linked to various settings, such as roots, leaves, suppressive soils, and degraded soils, requires comprehensive knowledge of soil microbial communities and their compositions and diversity.

Several fungi have the ability to adopt several lifestyles, including harmful, saprophytic, or symbiotic. The boundaries of idioms are frequently ambiguous, resulting in a lack of clarity. Individuals have the ability to modify their way of existence, for instance, through the process of endophytes transitioning into parasites and vice versa. This can be achieved through the utilization of omics, which refers to a set of instruments and novel methodologies employed to investigate plant-microbe interactions and gain a better understanding of the various behaviors involved (Bahram *et al*., 2022). Microbiome studies have gained significant traction in the field of environmental research. These studies have been employed to assess biodiversity levels and facilitate conservation efforts in protected regions. Additionally, they have been utilized to investigate the influence of various factors such as host taxon,

tissues, and seasonality on the composition of fungi and bacteria in tropical forests (Li *et al*., 2022). Furthermore, microbiome studies have been instrumental in comparing the microbiomes of trees and associated herbaceous plants used for phytoremediation purposes (Yung *et al*., 2021).

Reference

Anslan, S., Nilsson, R.H., Wurzbacher, C., Baldrian, P., Tedersoo, L. and Bahram, M. (2018). Great differences in performance and outcome of high-throughput sequencing data analysis platforms for fungal metabarcoding. *MycoKeys*, *39*, 29–40.

Bahram, M. and Netherway, T. (2022). Fungi as mediators linking organisms and ecosystems. *FEMS Microbiol. Rev.*, *46*, fuab058.

Cureau, N., Threlfall, R., Savin, M., Marasini, D., Lavefve, L. and Carbonero, F. (2021). Year, location, and variety impact on grape-, soil-, and leaf-associated fungal microbiota of arkansas-grown table grapes. *Microb. Ecol.*, *82*, 73–86.

Francioli, D., Lentendu, G., Lewin, S. and Kolb, S. (2021). DNA Metabarcoding for the Characterization of Terrestrial Microbiota-Pitfalls and Solutions. *Microorganisms*, *9*, 361.

Gobbi, A., Kyrkou, I., Filippi, E., Ellegaard-Jensen, L. and Hansen, L.H. (2020). Seasonal epiphytic microbial dynamics on grapevine leaves under biocontrol and copper fungicide treatments. *Sci. Rep.*, *10*, 681.

Griggs, R.G., Steenwerth, K.L., Mills, D.A., Cantu, D. and Bokulich, N.A. (2021). Sources and assembly of microbial communities in vineyards as a functional component of winegrowing. *Front. Microbiol.*, *12*, 673810.

Krause, T., Wassan, J.T., McKevitt, P., Wang, H., Zheng, H. and Hemmje, M. (2021). Analyzing large microbiome datasets using machine learning and big data. *BioMedInformatics*, *1*, 138–165.

Mazzola, M. (1998). Elucidation of the microbial complex having a causal role in the development of apple replant disease in Washington. *Phytopathology*, *88*, 930–938.

Morales-Cruz, A., Figueroa-Balderas, R., García, J.F., Tran, E., Rolshausen, P.E., Baumgartner, K. and Cantu, D. (2018). Profiling grapevine trunk pathogens in planta: A case for community-targeted DNA metabarcoding. *BMC Microbiol.*, *18*, 214.

Nerva, L., Pagliarani, C., Pugliese, M., Monchiero, M., Gonthier, S., Gullino, M.L., Gambino, G., Chitarra, W. (2019). Grapevine phyllosphere community analysis in response to elicitor application against powdery mildew. *Microorganisms*, *7*, 662.

Nguyen, N.H., Song, Z., Bates, S.T., Branco, S., Tedersoo, L., Menke, J., Schilling, J.S. and Kennedy, P.G. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.*, *20*, 241–248.

Nilsson, R.H., Larsson, K.H., Taylor, A.F., Bengtsson-Palme, J., Jeppesen, T.S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F.O., Tedersoo, L., *et al.* (2019). The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.*, *47*, D259–D264.

Oh, M. and Zhang, L. (2020). DeepMicro: Deep representation learning for disease prediction based on microbiome data. *Sci. Rep.*,*10*, 6026.

Pauvert, C., Buee, M., Laval, V., Edel-Hermann, V., Fauchery, L., Gautier, A., Lesura, I., Vallance, J. and Vacher, C. (2019). Bioinformatics matters:

The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. *Fungal Ecol.*, *41*, 23–33.

Polme, S., Abarenkov, K., Nilsson, R.H., Lindahl, B.D., Clemmensen, K.E., Kauserud, H., Nguyen, N., Kjøller, R., Bates, S.T., Baldrian, P., *et al.* (2020). FungalTraits: A user-friendly traits database of fungi and fungus-like stramenopiles. *Fungal Divers.*, *105*, 1–16.

Schiro, G., Colangeli, P. and Müller, M.E.H. (2019). A metabarcoding analysis of the mycobiome of wheat ears across a topographically heterogeneous field. *Front. Microbiol.*, *10*, 2095.

Singh, P., Santoni, S., Weber, A., This, P. and Péros, J.P. (2019). Understanding the phyllosphere microbiome assemblage in grape species (Vitaceae) with amplicon sequence data structures. *Sci. Rep.*, *9*, 14294.

Tedersoo, L., Bahram, M., Zinger, L., Nilsson, R.H., Kennedy, P.G., Yang, T., Anslan, S. and Mikryukov, V. (2022). Best practices in metabarcoding of fungi: From experimental design to results. *Mol. Ecol.*, *31*, 2769–2795.

Wilhelm, R.C., van Es, H.M. and Buckley, D.H. (2022). Predicting measures of soil health using the microbiome and supervised machine learning. *Soil Biol. Biochem.*, *164*, 108472.

Yung, L., Bertheau, C., Tafforeau, F., Zappelini, C., Valot, B., Maillard, F., Selosse, M.A., Viotti, C., Binet, P., Chiapusio, G., *et al.* (2021). Partial overlap of fungal communities associated with nettle and poplar roots when co-occurring at a trace metal contaminated site. *Sci. Total Environ.*, *782*, 146692.

* * * * * * * *