Advancing Microbial Identification in Complex Matrices: Oxford Nanopore Sequencing in High-Throughput Analysis

Ida Romano[1*], Natacha Bodenhausen[2], Valeria Ventorino[1,3], Michael Schneider[2], Sonja Reinhard[2], Giuseppina Magaraci[1], Ralf Müller[2], Jean-Claude Walser[4], Olimpia Pepe[1,3]

[1] Department of Agricultural Sciences, Division of Microbiology, University of Naples Federico II, Naples, Piazza Carlo di Borbone 1, Portici (NA), Italy

[2] Department of Soil Sciences, Research Institute of Organic Agriculture (FiBL), Ackerstrasse 113, Frick, Switzerland

[3] Task Force on Microbiome Studies, University of Naples Federico II, Naples, Italy

[4] ETH Zurich Genetic Diversity Center, Universitätstrasse 16, Zurich, Switzerland.

* Corresponding author Ida Romano (e-mail: ida.romano@unina.it)

Accurate species-level identification of bacteria and fungi by high-throughput sequencing in complex matrices is important in many fields, including biotechnological processes, environmental monitoring, and clinical diagnostics. The techniques used today to identify species in complex habitats are frequently time- and labor-intensive. Illumina's Amplicon sequencing technology makes it possible to characterize microbial communities; nevertheless, its short fragment length makes it difficult to identify specific species. More precise but quick approaches are needed to allow species-level identification. In this context, the application of Oxford Nanopore sequencing technology (ONT), also referred to as third-generation sequencing, to sequence complex samples is extremely relevant, because the large sequence length enables species-level identification.

This study utilized lignocellulosic biomass samples mixed with soil, alongside bacterial bulk samples obtained by plating the biomass onto a selective CMC substrate for cellulolytic bacteria isolation. The latest generation of flow cells (R10.4.1) were utilized, and the PacBio M13 barcodes two-step PCR procedure was employed in conjunction with the Ligation Sequencing Kit V14 (SQK-LSK114). This strategy targeted the 16S rRNA gene to characterize the bacterial community. In order to validate the new technology, samples underwent Illumina (V3-V4) sequencing for comparison.

Sequencing of 16S rRNA with ONT resulted in 3'908'291 sequences of ~1500 bp. Bioinformatics analysis was performed using USEARCH. Comparable outcomes were observed between Illumina and ONT sequencing methods for bulk samples, which exhibit lower microbial diversity. However, the standard bioinformatic pipeline underperformed when analyzing data from lignocellulosic biomass mixed with soil.

These results demonstrate the potential of ONT as a promising tool to study bacterial community in complex matrices. Further optimization of protocols and bioinformatics pipelines could enhance its applicability. Long-read sequencing technologies offer unprecedented opportunities for comprehensive genomic analysis and warrant continued exploration in microbial ecology and biotechnology applications.