What's in my Pot? (WIMP), a quantitative analysis tool for real-time species identification

WIMP is a quantitative analysis tool for real-time species identification for bacteria, fungi, archaea and viruses. We apply it here to DNA extracted from soil around the roots of a Basmati rice plant

Contact: **publications@nanoporetech.com** More information at: **www.nanoporetech.com** and **publications.nanoporetech.com**

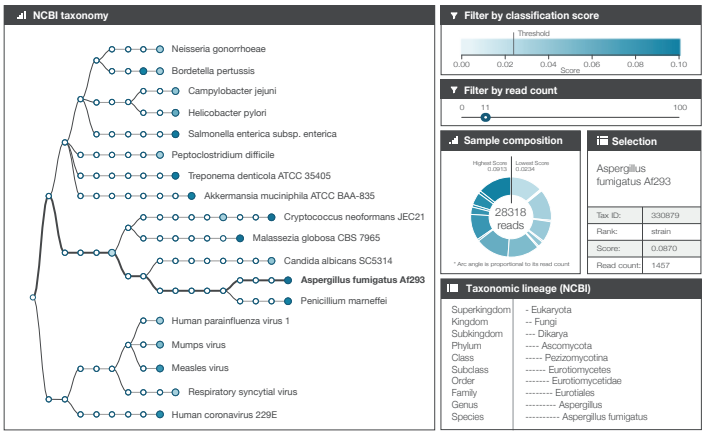

Fig. 1 WIMP report, shown for a sample containing bacteria, viruses and fungi

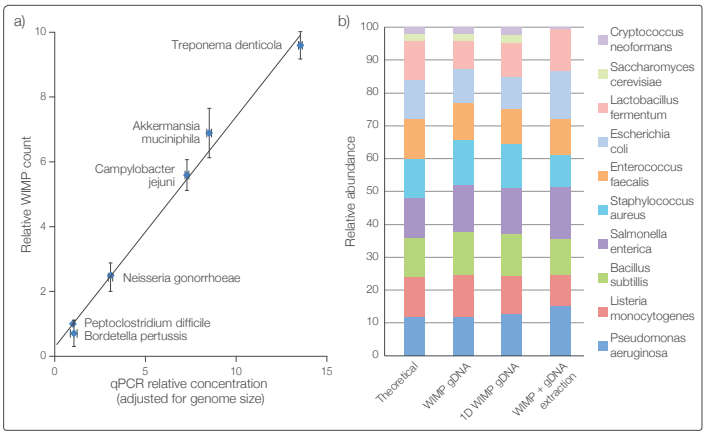

Fig. 2 WIMP species quantification a) vs. qPCR b) on the Zymo microbial community standard

The WIMP application for bacterial, viral, fungal and archaeal species identification

The WIMP workflow classifies and identifies species in real time: as soon as a strand of DNA passes through the pore it can be basecalled and analysed. WIMP makes use of Centrifuge, which is capable of accurately identifying reads when using databases containing multiple highly similar reference genomes, such as different strains of a bacterial species. Centrifuge works by identifying unique segments of those genomes and building an FM-index that can be used for efficient searches of sequenced reads. WIMP processes Centrifuge results to determine the most reliable placement in the taxonomy tree, assigning a score to each taxonomic placement. WIMP currently supports bacteria, archaea, viruses and fungi (Fig. 1).

WIMP species counts correlate with qPCR and microbial standard quantification

To demonstrate that the WIMP workflow (including sequencing and library preparation) is quantitative, we took a selection of bacterial genomes, prepared sequencing libraries and pooled them in arbitrary ratios. We then compared the WIMP counts to relative abundance as measured by qPCR. The qPCR counts agreed well with the WIMP data (Fig. 2a). Next, we compared the WIMP counts on a microbial community standard to the manufacturer's theoretical values and again saw good concordance (Fig. 2b). As a consequence, the limit of sensitivity of the workflow is essentially a question of how many reads are obtained from the sequencing run, which is governed by how long the device is left to generate data.

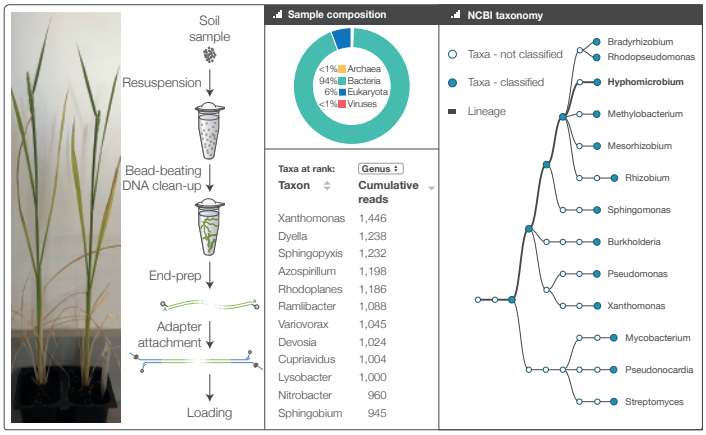

Fig. 3 Analysis of microbiome from soil around the roots of Basmati rice, shown at genus level

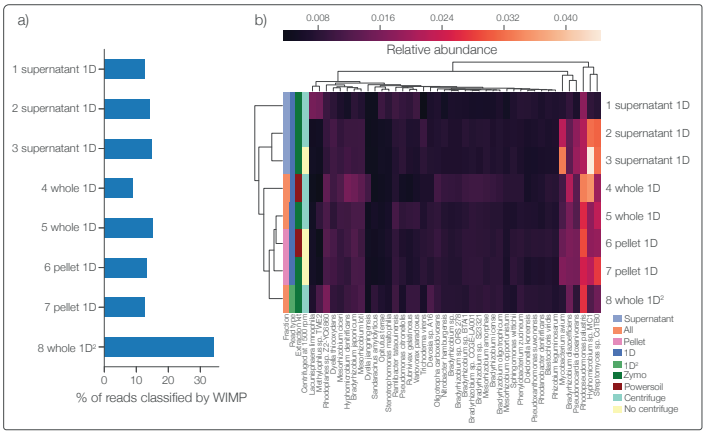

Fig. 4 Analysis of rice soil microbes with 1D and 1D² reads a) classification b) identified species

WIMP analysis of complex soil microbial community from Basmati rice roots

To demonstrate the effectiveness of WIMP for the analysis of a highly complex microbial community, we took a sample of soil from around the roots of a Basmati rice plant and extracted total genomic DNA by bead-beating. We prepared and sequenced an LSK-108 library and identified the organisms present with WIMP. The microbial community consisted largely of bacteria, though some fungal species were also identified (Fig. 3). WIMP identified many species of nitrogen-fixing bacteria, such as those belonging to the Bradyrhizobium and Rhizobium genera. Notably, WIMP also revealed the presence of several phytopathogens, including Xanthomonas translucens, which causes leaf streak, and can reduce crop yields substantially.

Reproducible species identification from 1D and 1D² reads

We prepared 1D and 1D² libraries from the rice soil metagenomic sample, to investigate the effect of using higher-accuracy reads on our ability to classify to species level. The results showed that approximately twice as many 1D² reads could be classified by WIMP, compared to 1D reads (Fig. 4a). This is presumably due to a higher proportion of 1D² reads passing WIMP's inbuilt quality-score filtering. We compared the species identified from several 1D libraries with the 1D² results, by exporting CSV files from WIMP, and by generating heat-maps from the data (Fig. 4b). The results from all libraries were in very close agreement, indicating that it is not essential to use the higher-accuracy 1D² reads for robust species identification with WIMP.