An accessible metagenomic strategy allows for better characterization of invertebrate bulk samples

Callens Martijn, Le Berre Guillaume, Van den Bulcke Laure, Lolivier Marianne and Derycke Sofie

Marine Genomics lab, Flanders Research Institute for Agriculture, Fisheries and Food, Jacobsenstraat 1, Ostend, Belgium
E-mail: martijn.callens@ilvo.vlaanderen.be

Introduction
DNA metabarcoding has proven to be a cost- and time-effective alternative to morphological identifications for environmental monitoring. However, when a community contains a high phylogenetic diversity, it can be difficult to tailor PCR primers that effectively amplify a marker gene from all species. This results in PCR amplification bias, which affects monitoring data quality derived from metabarcoding. Several studies have proposed to use PCR-free approaches (i.e. shotgun metagenomics) to circumvent this issue, but generally indicate two hurdles preventing the wide applicability of this approach: 1) a lack of reference genomes for most species present in a given environment, and 2) computational intensive pipelines for processing shotgun metagenomic data. Here, we propose a strategy that tackles these two hurdles and apply it to classify shotgun metagenomic reads from macrobenthos samples.

Methods
We selected 25 macrobenthos species from various phyla for low-coverage Illumina whole genome sequencing. We build a k-mer index database directly from the sequencing reads, thus circumventing tedious genome assembly. This database was then used to classify shotgun metagenomic reads from macrobenthos samples using a very fast exact k-mer matching algorithm. The same samples were simultaneously characterized by morphological identification and metabarcoding to compare results obtained by different methods.

Results
We show that low-coverage genome sequencing allows us to build a database that equals the classification potential of a database build with fully assembled reference genomes. We are able to classify a large fraction of metagenomic reads from our samples (up to 96%). Results from shotgun metagenomics align better with biomass than those from metabarcoding due to the absence of PCR amplification bias.

Conclusions
Our strategy provides an easy, fast, and accessible way to assess community composition in metazoan bulk samples by shotgun metagenomics

Keywords
Metagenomics; Macrobenthos; Biodiversity Monitoring