

# Shotgun metagenomics of soil invertebrate communities reflects taxonomy, biomass and reference genome properties

Alexandra Schmidt<sup>1</sup>, Clément Schneider<sup>2</sup>, Peter Decker<sup>3</sup>, Karin Hohberg<sup>4</sup>, Jörg Römbke<sup>5</sup>, Ricarda Lehmitz<sup>6</sup>, and Miklos Balint<sup>7</sup>

<sup>1</sup>Universität Konstanz

<sup>2</sup>Senckenberg Gesellschaft für Naturforschung

<sup>3</sup>Senckenberg Museum of Natural History Görlitz Library

<sup>4</sup>Senckenberg Museum of Natural History

<sup>5</sup>ECT Oekotoxikologie, Flörsheim, Germany

<sup>6</sup>Senckenberg Museum für Naturkunde Görlitz

<sup>7</sup>Senckenberg Biodiversity Climate Research Center

November 29, 2021

## Abstract

Metagenomics - shotgun sequencing of all DNA fragments from a community DNA extract - is routinely used to describe the composition, structure and function of microorganism communities. Advances in DNA sequencing and the availability of genome databases increasingly allow the use of shotgun metagenomics on eukaryotic communities. Metagenomics offers major advances in the recovery of biomass relationships, in comparison to taxonomic marker gene based approaches (metabarcoding). However, little is known about the factors that influence metagenomics data from eukaryotic communities, such as differences among organism groups, properties of reference genomes and genome assemblies. We evaluated how shotgun metagenomics records composition and biomass in artificial soil invertebrate communities. We generated mock communities of controlled biomass ratios from 28 species from all major soil mesofauna groups: mites, springtails, nematodes, tardigrades and potworms. We shotgun-sequenced these communities and taxonomically assigned them with a database of over 270 soil invertebrate genomes. We recovered 90% of the species, and observed relatively high false positive detection rates. We found strong differences in reads assigned to different taxa, with some groups consistently attracting more hits than others. Biomass could be predicted from read counts after considering taxon-specific differences. Larger genomes more complete assemblies consistently attracted more reads than genomes. The GC content of the genome assemblies had no effect on the biomass-read relationships. The results show considerable differences in taxon recovery and taxon specificity of biomass recovery from metagenomic sequence data. Properties of reference genomes and genome assemblies also influence biomass recovery, and they should be considered in metagenomic studies of eukaryotes. We provide a roadmap for investigating factors which influence metagenomics-based eukaryotic community reconstructions. Understanding these factors is timely as accessibility of DNA sequencing, and momentum for reference genomes projects show a future where the taxonomic assignment of DNA from any community sample becomes a reality.

## Hosted file

Manuscript.pdf available at <https://authorea.com/users/448480/articles/547285-shotgun-metagenomics-of-soil-invertebrate-communities-reflects-taxonomy-biomass-and-reference-genome-properties>