ID: W2949342052

TITLE: Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold

AUTHOR: ['Martin Steinegger', 'Milot Mirdita', 'Johannes Söding']

ABSTRACT:

The open-source de novo protein-level assembler, Plass (https://plass.mmseqs.com), assembles six-frame-translated sequencing reads into protein sequences. It recovers 2?10 times more protein sequences from complex metagenomes and can assemble huge datasets. We assembled two redundancy-filtered reference protein catalogs, 2 billion sequences from 640 soil samples (soil reference protein catalog) and 292 million sequences from 775 marine eukaryotic metatranscriptomes (marine eukaryotic reference catalog), the largest free collections of protein sequences. The protein-level assembler can assemble protein catalogs from raw metagenomic sequencing data, enabling large-scale metagenomics studies.

PDF URL: None

CITED BY COUNT: 281

PUBLICATION YEAR: 2019

TYPE: article

CONCEPTS: ['Metagenomics', 'Computational biology', 'Biology', 'Protein sequencing', 'Sequence assembly', 'Redundancy (engineering)', 'Genetics', 'Gene', 'Peptide sequence', 'Computer science', 'Operating system', 'Gene expression', 'Transcriptome']