VIBRANT: Automated recovery, annotation and curation of microbial viruses, and evaluation of virome function from genomic sequences

CURRENT STATUS: POSTED

Kristopher Kieft
University of Wisconsin Madison
ORCiD: https://orcid.org/0000-0002-9120-1961

Zhichao Zhou
University of Wisconsin Madison

Karthik Anantharaman
karthik@bact.wisc.edu Corresponding Author
ORCiD: https://orcid.org/0000-0002-9584-2491

DOI:
10.21203/rs.2.18878/v1

SUBJECT AREAS
General Microbiology

KEYWORDS
Virome, virus, bacteriophage, metagenome, machine learning, auxiliary metabolism
Abstract
Background Viruses are central to microbial community structure in all environments. The ability to generate large metagenomic assemblies of mixed microbial and viral sequences provides the opportunity to tease apart complex microbiome dynamics, but these analyses are currently limited by the tools available for analyses of viral genomes and assessing their metabolic impacts on microbiomes.

Design Here we present VIBRANT, the first method to utilize a hybrid machine learning and protein similarity approach that is not reliant on sequence features for automated recovery and annotation of viruses, determination of genome quality and completeness, and characterization of virome function from metagenomic assemblies. VIBRANT uses neural networks of protein signatures and a novel v-score metric that circumvents traditional boundaries to maximize identification of lytic viral genomes and integrated proviruses, including highly diverse viruses. VIBRANT highlights viral auxiliary metabolic genes and metabolic pathways, thereby serving as a user-friendly platform for evaluating virome function. VIBRANT was trained and validated on reference virus datasets as well as microbiome and virome data.

Results VIBRANT showed superior performance in recovering higher quality viruses and concurrently reduced the false identification of non-viral genome fragments in comparison to other virus identification programs, specifically VirSorter and VirFinder. When applied to 120,834 metagenomically derived viral sequences representing several human and natural environments, VIBRANT recovered an average of 94.5% of the viruses, whereas VirFinder and VirSorter achieved less powerful performance, averaging 48.1% and 56.0%, respectively. Similarly, VIBRANT identified more total viral sequence and proteins when applied to real metagenomes. When compared to PHASTER and Prophage Hunter for the ability to extract integrated provirus regions from host scaffolds, VIBRANT performed comparably and even identified proviruses that the other programs did not. To demonstrate applications of VIBRANT, we studied viromes associated with Crohn’s Disease to show that specific viral groups, namely Enterobacteriales-like viruses, as well as putative dysbiosis associated viral proteins are more abundant compared to healthy individuals, providing a possible
viral link to maintenance of diseased states.

Conclusions The ability to accurately recover viruses and explore viral impacts on microbial community metabolism will greatly advance our understanding of microbiomes, host-microbe interactions and ecosystem dynamics.

Full-text
Due to technical limitations, full-text HTML conversion of this manuscript could not be completed.

However, the manuscript can be downloaded and accessed as a PDF.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

Additional_File_8.xlsx
Additional_File_4.xlsx
Additional_File_11.xlsx
Additional_File_6.xlsx
Additional_File_7.xlsx
Additional_File_5.xlsx
Additional_File_22.pdf
Additional_File_3.xlsx
Additional_File_2.xlsx
Additional_File_1.xlsx
Additional_File_21.pdf
Additional_File_20.xlsx
Additional_File_9.xlsx
Additional_File_12.xlsx
Additional_File_23.pdf
Additional_File_18.xlsx
Additional_File_15.xlsx
Additional_File_17.xlsx
Additional_File_14.xlsx
Additional_File_16.xlsx
Additional_File_13.xlsx
Additional_File_21.xlsx
Additional_File_19.xlsx