ViOTUcluster: A High-Speed, All-in-One Pipeline for Viromic Analysis of Metagenomic Data

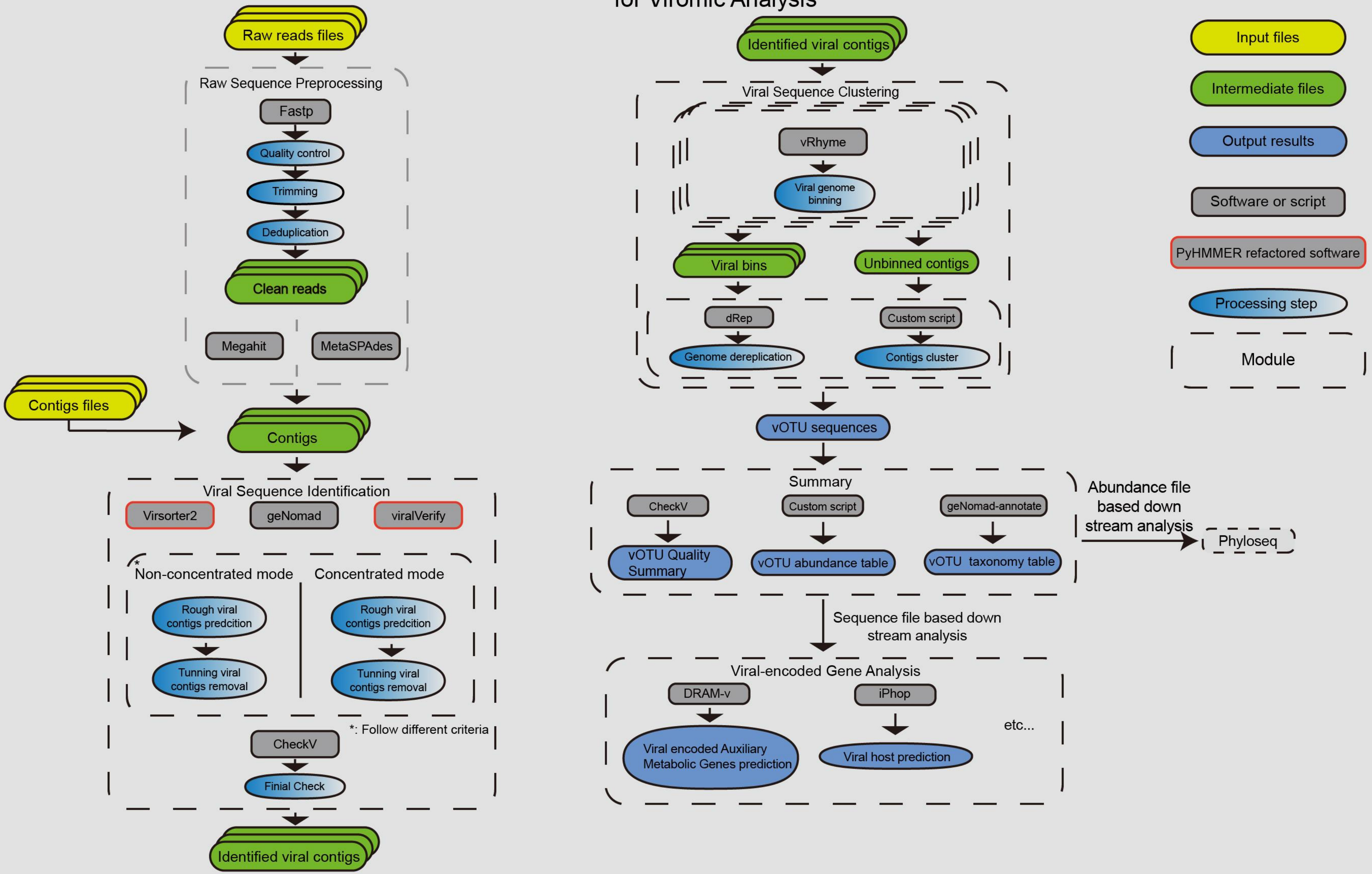
Sihang Liu[1,2], Yinyin Ye[3], Bing Guo[4], Yuxing Hu[1,2], Kaiyang Jiang[1,2],
Chengyu Liang[1,2], Siqing Xia[1,2*], Hong Wang[1,2,5*]

[1]State Key Laboratory of Water Pollution Control and Green Resource Recycling,
College of Environmental Science and Engineering, Tongji University,China
[2]Shanghai Institute of Pollution Control and Ecological Security, China
[3]Department of Civil, Structural and Environmental Engineering,
University at Buffalo, United States
[4]School of Sustainability, Civil and Environmental Engineering, University of Surrey, UK
[5]Jiaxing-Tongji Environmental Research Institute,, China

Sihang Liu, Yinyin Ye, Bing Guo, Yuxing Hu, Kaiyang Jiang, Chengyu Liang, Siqing Xia, et al. 2025.
ViOTUcluster: A High-Speed, All-in-One Pipeline for Viromic Analysis of Metagenomic Data.
iMetaOmics 3: e70023. https://doi.org/10.1002/imo2.70023

Introduction

ViOTUcluster: A High-Speed, All-in-One Pipeline for Viromic Analysis



Raw reads files

Raw Sequence Preprocessing

Fastp

Quality control

Trimming

Deduplication

Clean reads

Megahit

MetaSPAdes

Contigs files

Contigs

Viral Sequence Identification

Virsorter2

geNomad

viralVerify

*Non-concentrated mode

Rough viral contigs predcition

Tunning viral contigs removal

Concentrated mode

Rough viral contigs predcition

Tunning viral contigs removal

*: Follow different criteria

CheckV

Finial Check

Identified viral contigs

Identified viral contigs

Viral Sequence Clustering

vRhyme

Viral genome binning

Viral bins

Unbinned contigs

dRep

Genome dereplication

Custom script

Contigs cluster

vOTU sequences

Summary

CheckV

vOTU Quality Summary

Custom script

vOTU abundance table

geNomad-annotate

vOTU taxonomy table

Abundance file based down stream analysis

Phyloseq

Sequence file based down stream analysis

Viral-encoded Gene Analysis

DRAM-v

Viral encoded Auxiliary Metabolic Genes prediction

iPhop

Viral host prediction

etc...

Input files

Intermediate files

Output results

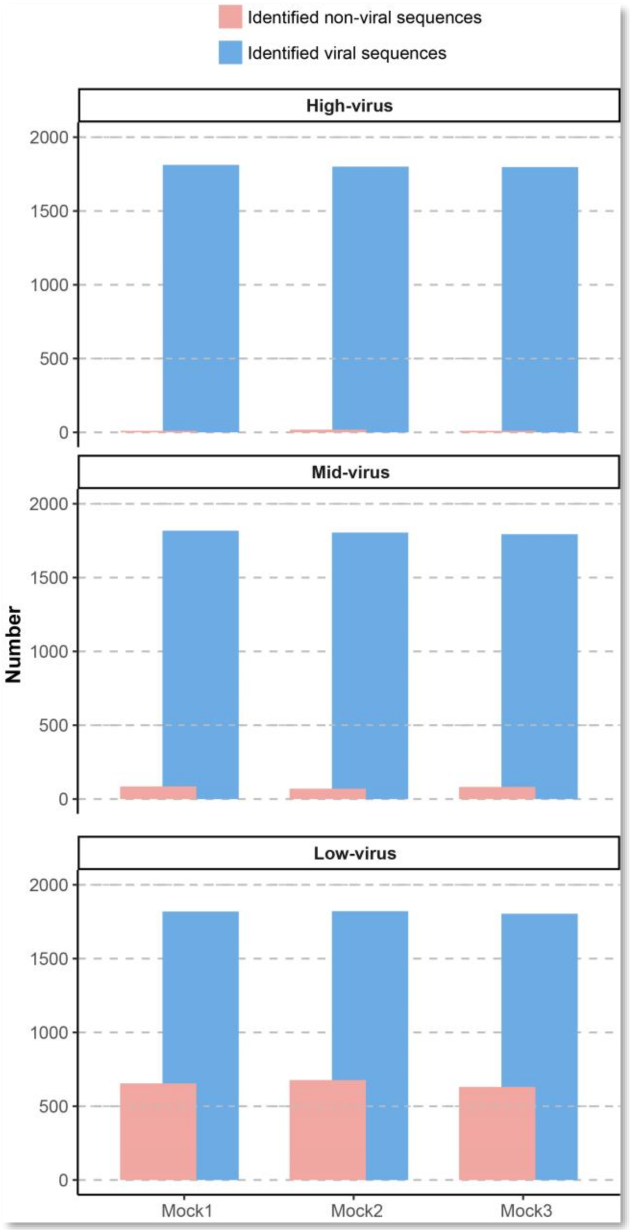
Software or script

PyHMMER refactored software

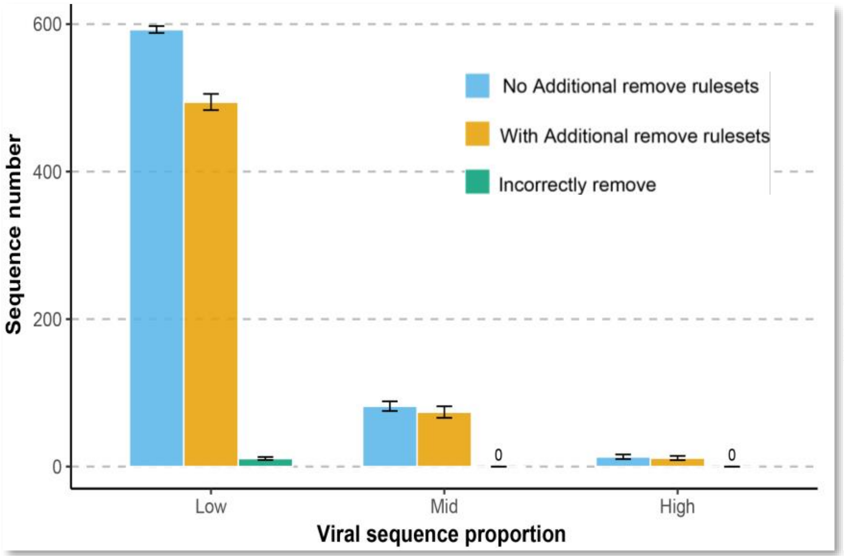
Processing step

Module

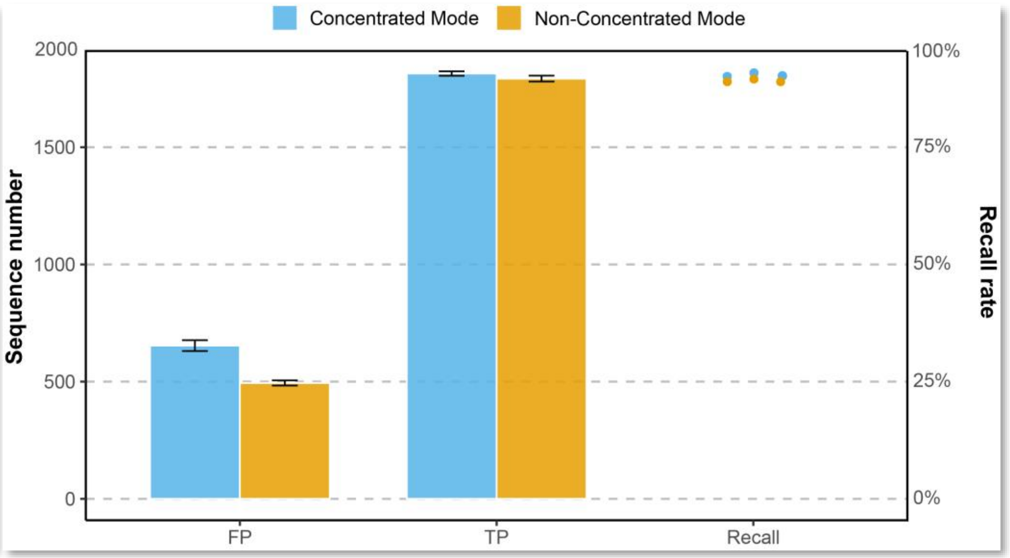
Performance metrics of Viral-Sequence-Identification module



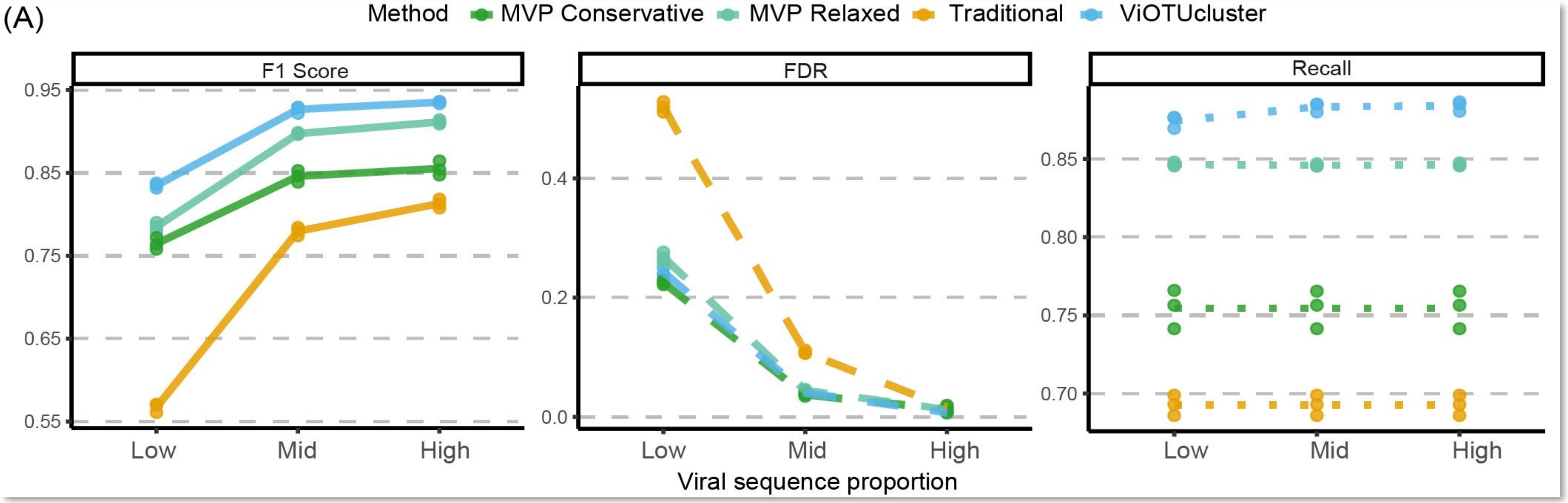
Contigs classified as viruses in the first step of ViOTUcluster



Effect of additional removal rulesets on the removal of non-viral sequences

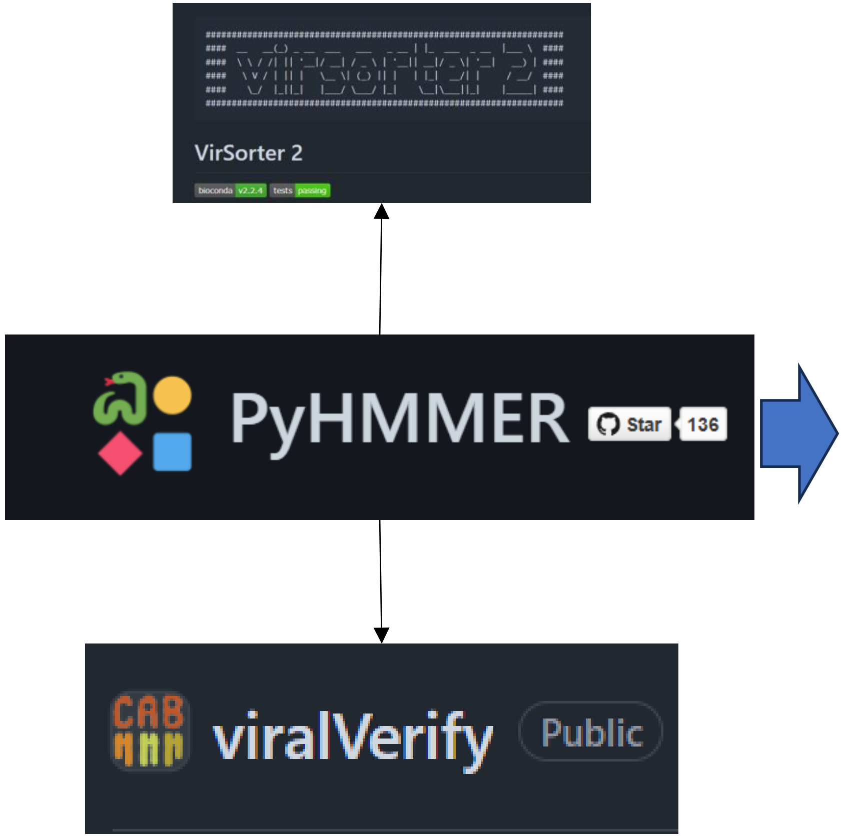


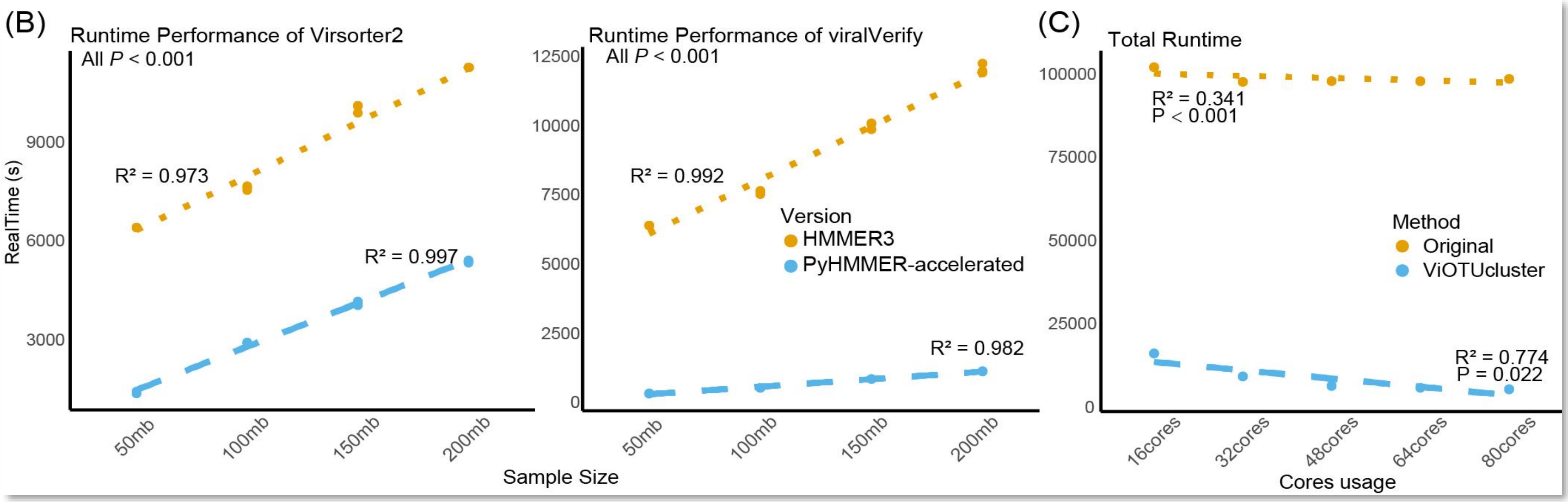
Performance of ViOTUcluster “Concentrate” mode and “Non-concentrated” mode



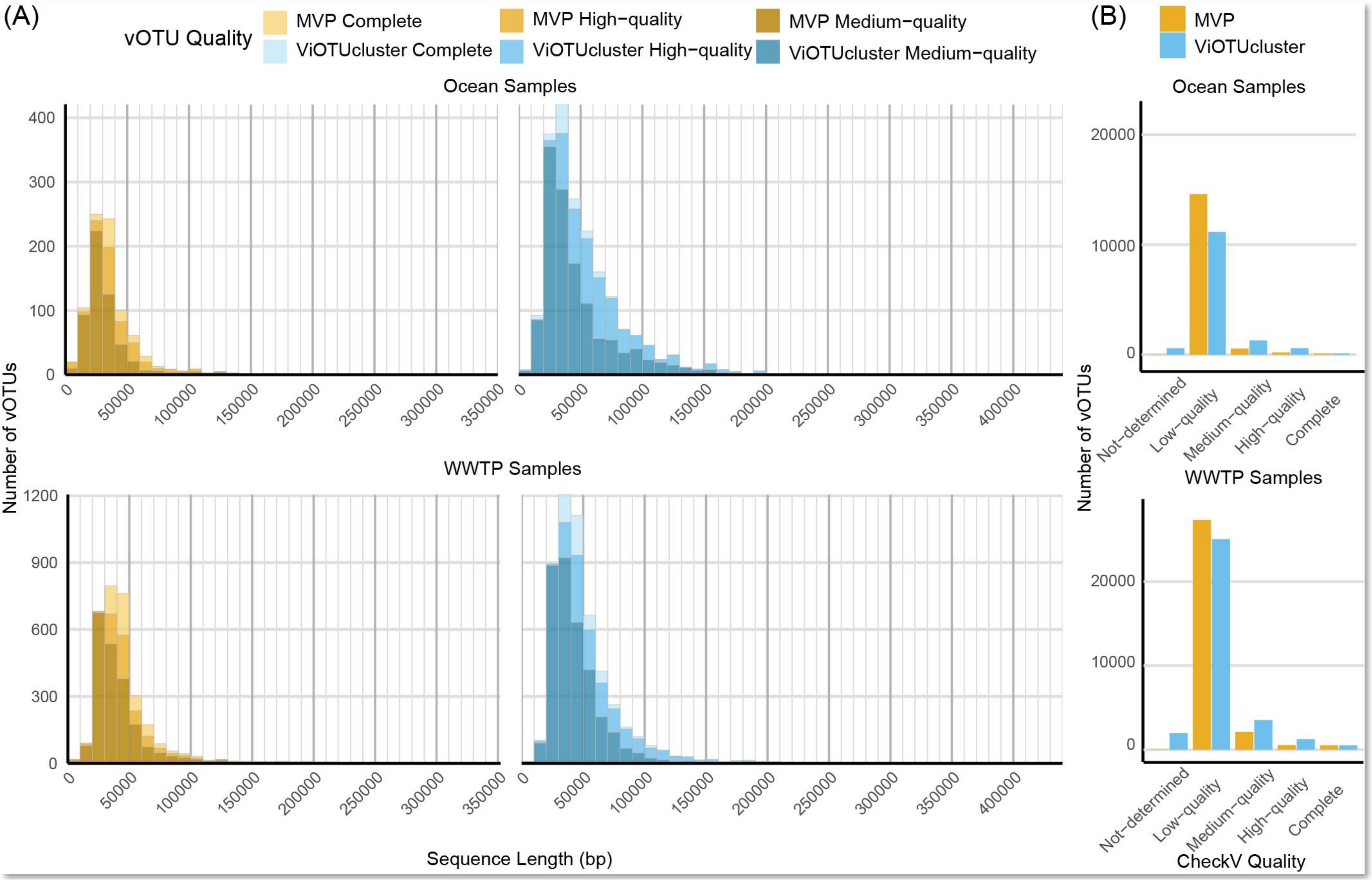
Overall performance metrics comparison

ViOTUcluster with refactored tools reduce computational time for large-scale analyses

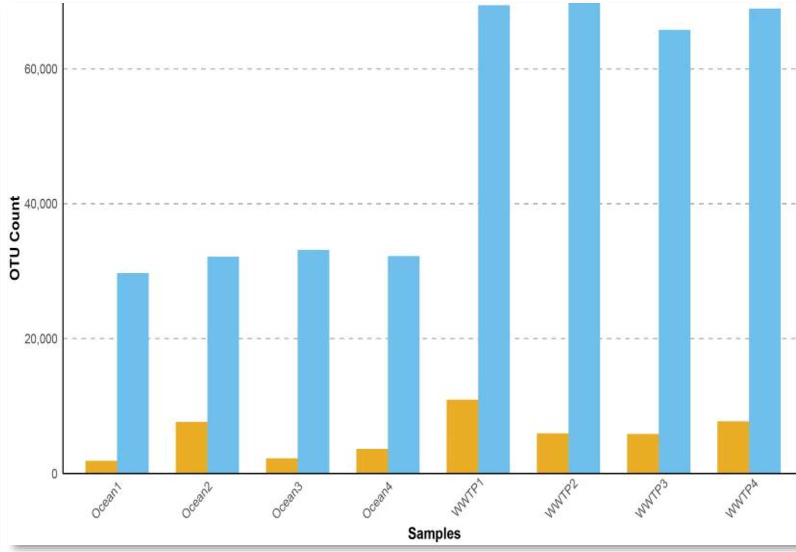




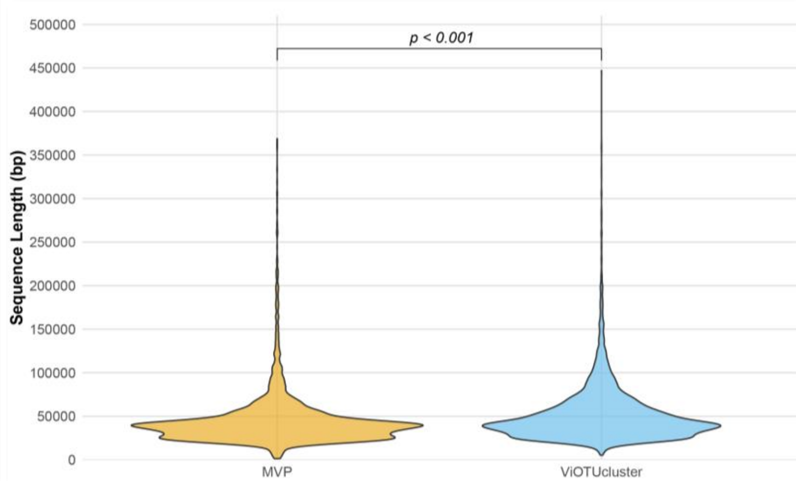
Benchmarking with Real-World Metagenomic Data



Reconstructed vOTUs quality in ocean and WWTP samples



The number of identified vOTUs in each sample between MVP and ViOTUcluster



Medium-, high-, and complete vOTU sequence lengths (in base pairs) obtained using MVP and ViOTUcluster

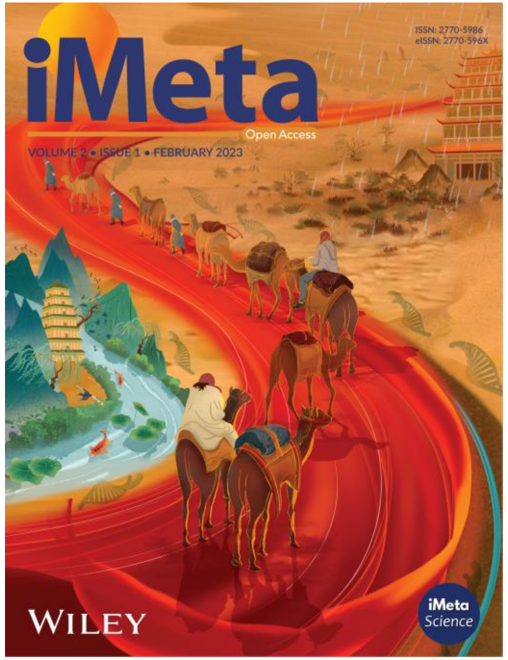
Summary

- Comprehensive & Streamlined Solution:
  - Integrates multiple steps (preprocessing to gene analysis) into one user-friendly workflow.
- High Accuracy & Adaptability:
  - Achieves high F1 scores (>0.83 low, >0.94 high viral abundance samples) via optimized modes.
  - Outperforms traditional methods and MVP, especially in recall for low-abundance samples.
- Significant Computational Efficiency Gains:
  - Drastically reduces runtime (up to ~16x faster) through tool refactoring (PyHMMER) and efficient parallelization.
- Superior Real-World Performance:
  - Recovers more high-quality, longer viral genomes from complex metagenomes compared to MVP.
  - Processes real data substantially faster.
- Website: https://github.com/liusihang/ViOTUcluster

Sihang Liu, Yinyin Ye, Bing Guo, Yuxing Hu, Kaiyang Jiang, Chengyu Liang, Siqing Xia, et al. 2025.
ViOTUcluster: A High-Speed, All-in-One Pipeline for Viromic Analysis of Metagenomic Data.
iMetaOmics 3: e70023. https://doi.org/10.1002/imo2.70023

***iMeta*:** Integrated metaomics to understand the biology, med and environment

WILEY



"***iMeta***" launched by iMeta Science Society in 2022, impact factor (IF) **23.8**, ranking top 107/21973 in world and 2/161 in the microbiology. It aims to publish innovative and high-quality papers with broad and diverse audiences. **Its scope is similar to *Cell*, *Nature*, *Science***, *Nature Biotechnology/Methods/Microbiology/Medicine/Food*. Its unique features include video abstract, bilingual publication, and social media with 600,000 followers. Indexed by **SCIE/ESI**, **PubMed**, **Google Scholar** etc.

"***iMetaOmics***" launched in 2024, with a **target IF>10, and its scope is similar to *Nature Communications***, ***Cell Reports***, *Microbiome, ISME J, Nucleic Acids Research, Briefings in Bioinformatics,* etc.

"***iMetaMed***" launched in 2025, with a **target IF>15, similar to *Med, Cell Reports Medicine***, *eBioMedicine, eClinicalMedicine* etc.

Society: http://www.imeta.science
Publisher: https://wileyonlinelibrary.com/journal/imeta
iMeta: https://wiley.atyponrex.com/journal/IMT2
Submission: iMetaOmics: https://wiley.atyponrex.com/journal/IMO2
iMetaMed: https://wiley.atyponrex.com/journal/IMM3

iMetaScience
iMetaScience

office@imeta.science
imetaomics@imeta.science

Promotion Video

Update
2025/5/21