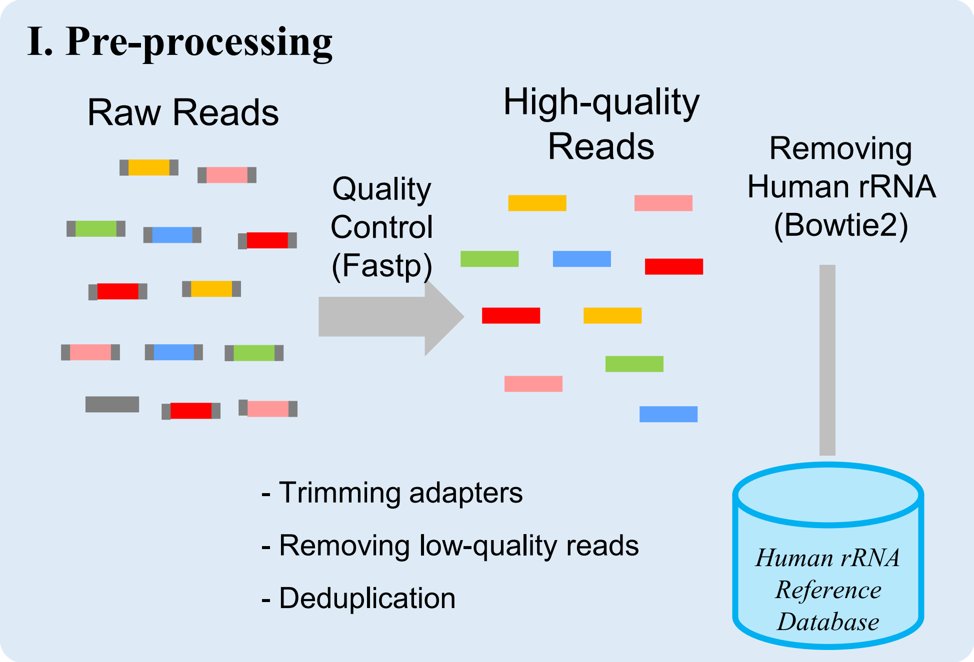
RTTAP: Empowering Metatranscriptomic Data Analysis with a Read-based Total-Infectome Taxonomic Solution

Wei Jiang[1], Herui Liao[1], Mang Shi[2], Liangjun Chen[3*], Yanni Sun[1*]

[1]Department of Electrical Engineering, City University of Hong Kong,
Hong Kong, China
[2]School of Medicine, Shenzhen campus of Sun Yat-sen University,
Shenzhen, China
[3]Department of Laboratory Medicine, Zhongnan Hospital, Wuhan University,
Wuhan, China

Wei Jiang, Herui Liao, Mang Shi, Liangjun Chen, Yanni Sun. 2025. RTTAP: Empowering metatranscriptomic data analysis with a read-based total-infectome taxonomic solution. ***iMetaOmics*** 2: e70044.
https://doi.org/10.1002/imo2.70044.

Introduction



I. Pre-processing

Raw Reads

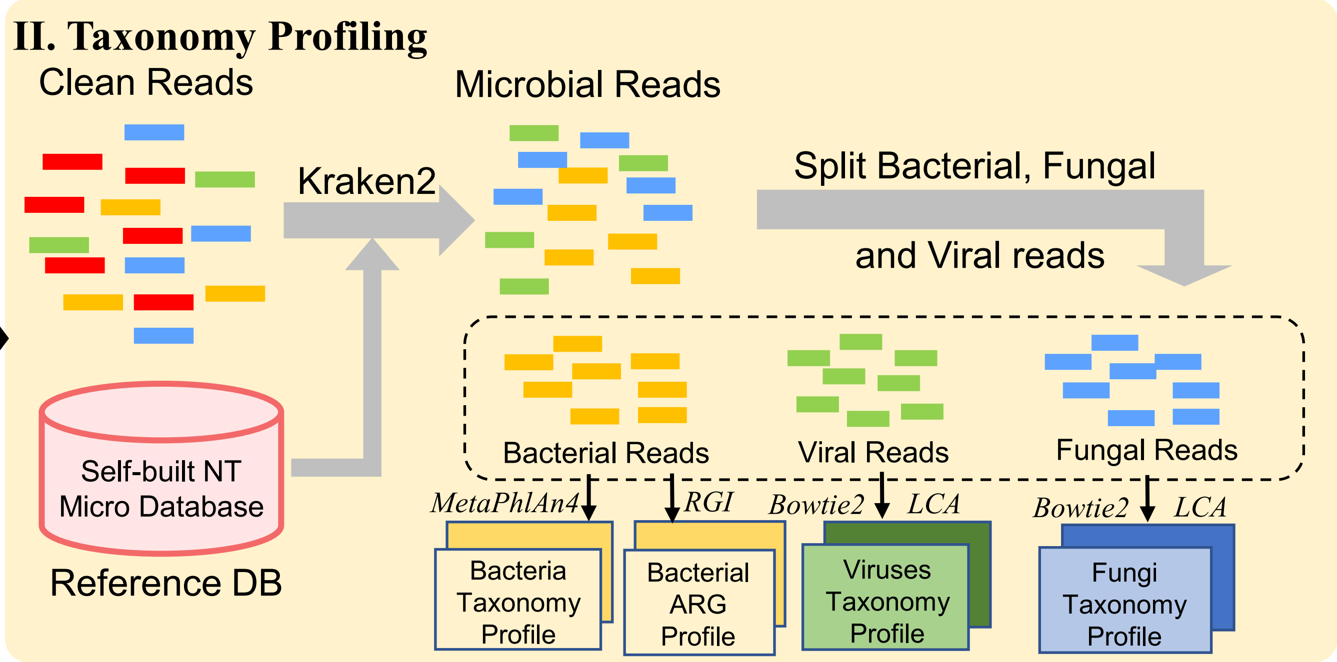
Quality Control (Fastp)

High-quality Reads

Removing Human rRNA (Bowtie2)

- Trimming adapters
- Removing low-quality reads
- Deduplication

Human rRNA Reference Database



II. Taxonomy Profiling

Clean Reads

Kraken2

Microbial Reads

Split Bacterial, Fungal and Viral reads

Self-built NT Micro Database

Reference DB

Bacterial Reads

Viral Reads

Fungal Reads

MetaPhlAn4

RGI

Bowtie2 *LCA*

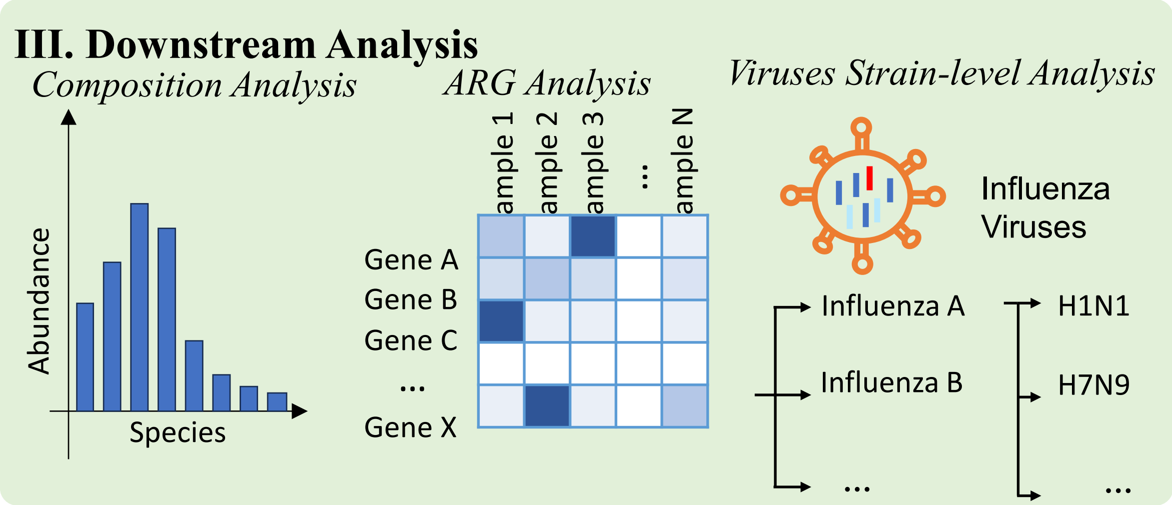
Bowtie2 *LCA*

Bacteria Taxonomy Profile

Bacterial ARG Profile

Viruses Taxonomy Profile

Fungi Taxonomy Profile



III. Downstream Analysis

Composition Analysis

ARG Analysis

Viruses Strain-level Analysis

Influenza Viruses

Influenza A

Influenza B

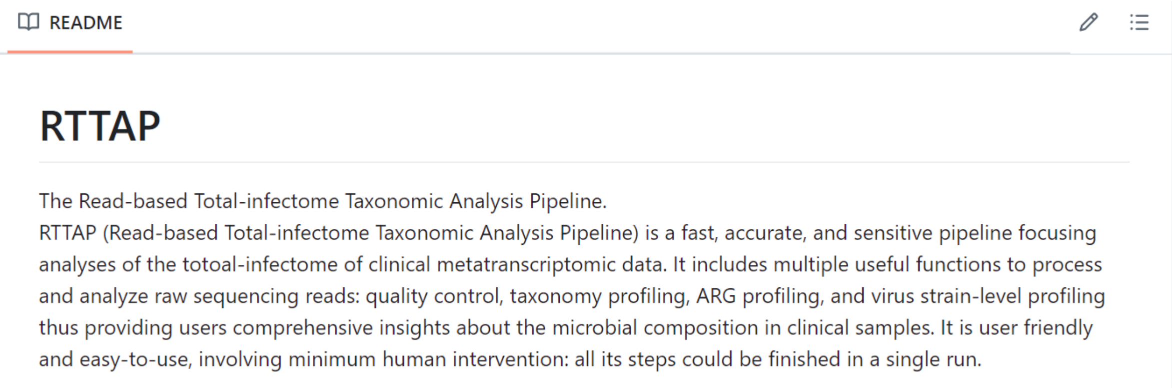
…

H1N1

H7N9

…

Documentation:



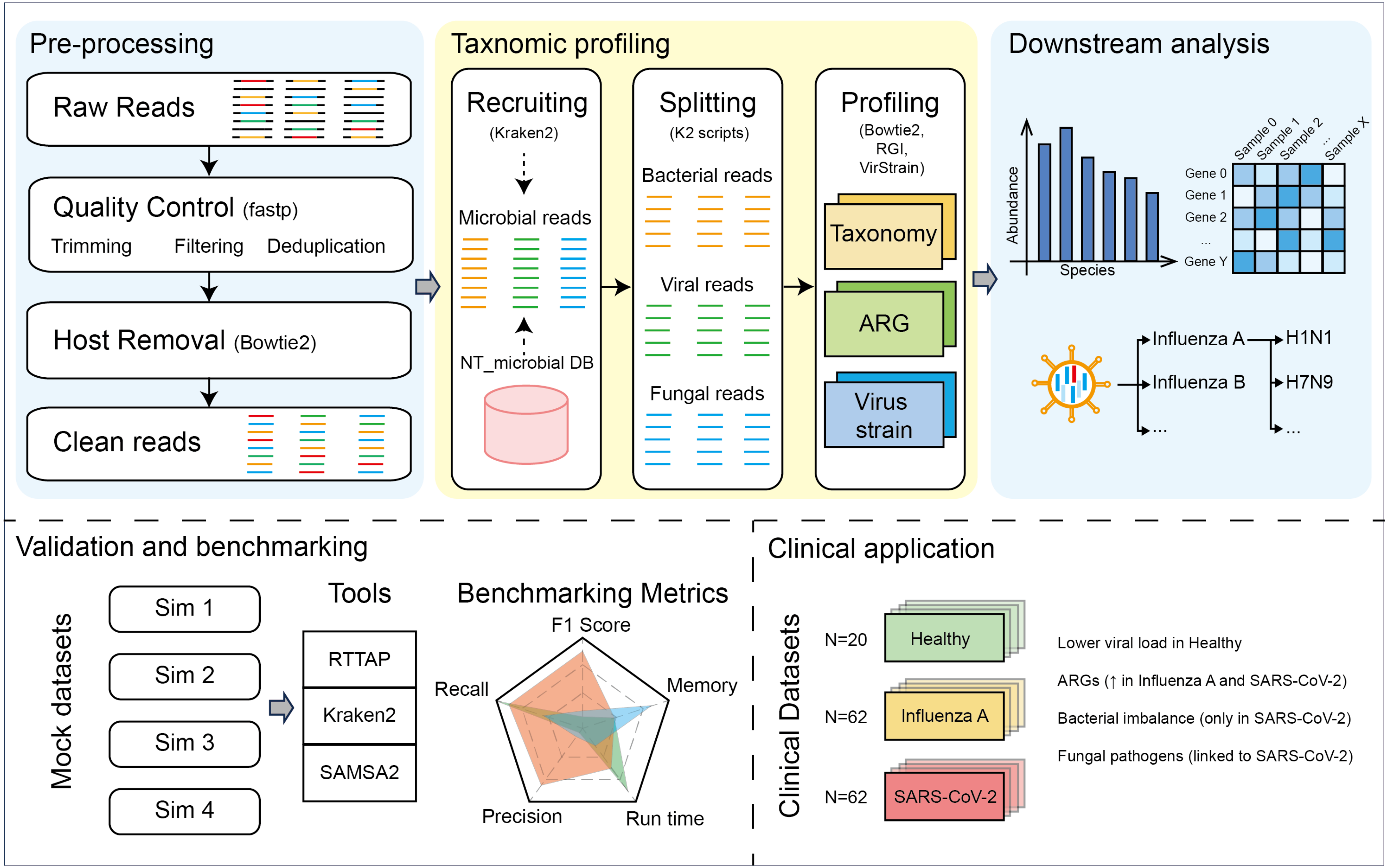
RTTAP

The Read-based Total-infectome Taxonomic Analysis Pipeline.

RTTAP (Read-based Total-infectome Taxonomic Analysis Pipeline) is a fast, accurate, and sensitive pipeline focusing analyses of the totoal-infectome of clinical metatranscriptomic data. It includes multiple useful functions to process and analyze raw sequencing reads: quality control, taxonomy profiling, ARG profiling, and virus strain-level profiling thus providing users comprehensive insights about the microbial composition in clinical samples. It is user friendly and easy-to-use, involving minimum human intervention: all its steps could be finished in a single run.

For more details, please visit: https://github.com/weijiang34/RTTAP

Highlights



Pre-processing

Raw Reads

Quality Control (fastp)

Trimming Filtering Deduplication

Host Removal (Bowtie2)

Clean reads

Taxnomic profiling

Recruiting (Kraken2)

Microbial reads

NT_microbial DB

Splitting (K2 scripts)

Bacterial reads

Viral reads

Fungal reads

Profiling (Bowtie2, RGI, VirStrain)

Taxonomy

ARG

Virus strain

Downstream analysis

Influenza A → H1N1

Influenza B → H7N9

... → ...

Validation and benchmarking

Mock datasets

Sim 1

Sim 2

Sim 3

Sim 4

Tools

RTTAP
Kraken2
SAMSA2

Benchmarking Metrics

Clinical application

Clinical Datasets

N=20	Healthy
N=62	Influenza A
N=62	SARS-CoV-2

Lower viral load in Healthy

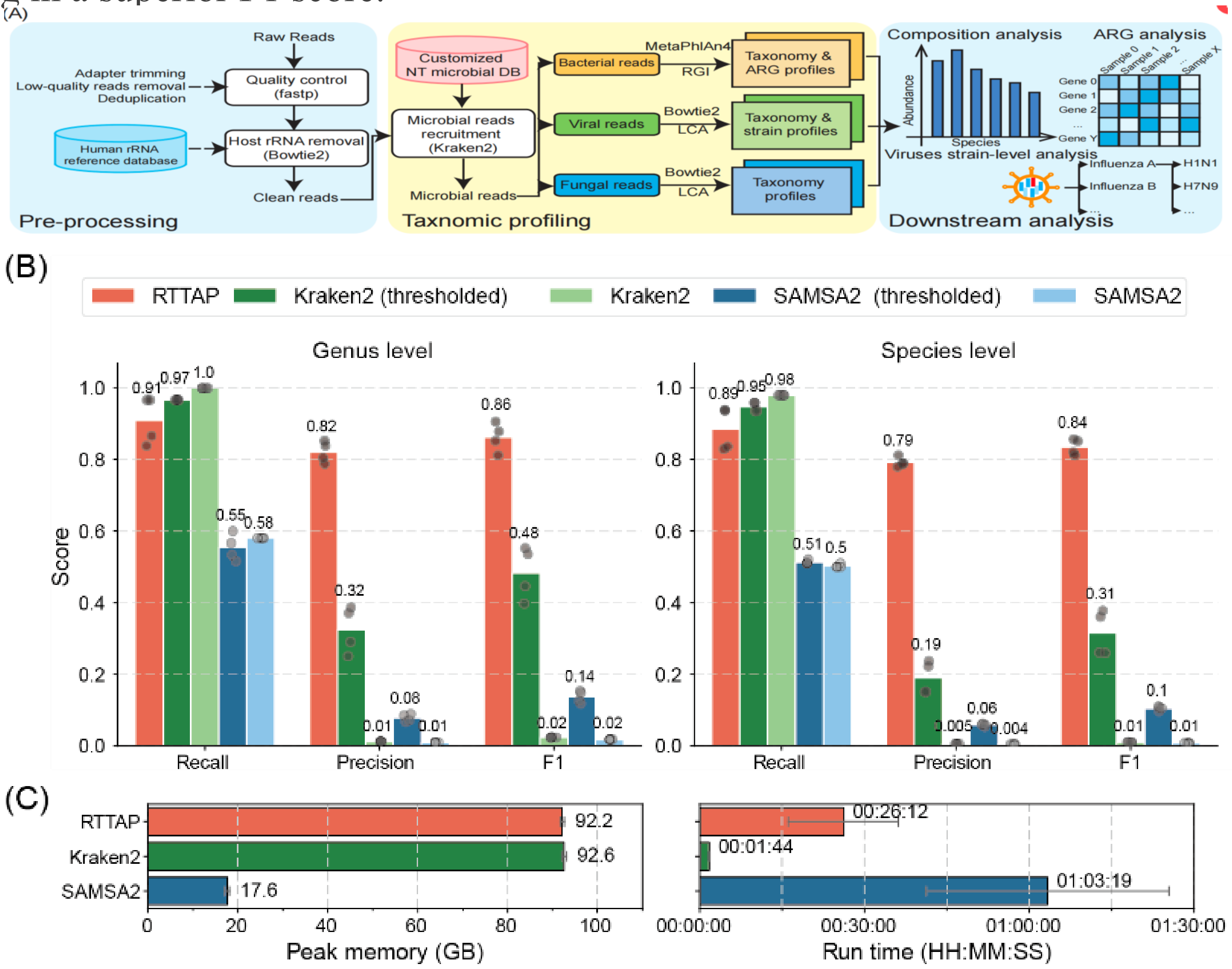
ARGs (↑ in Influenza A and SARS-CoV-2)

Bacterial imbalance (only in SARS-CoV-2)

Fungal pathogens (linked to SARS-CoV-2)

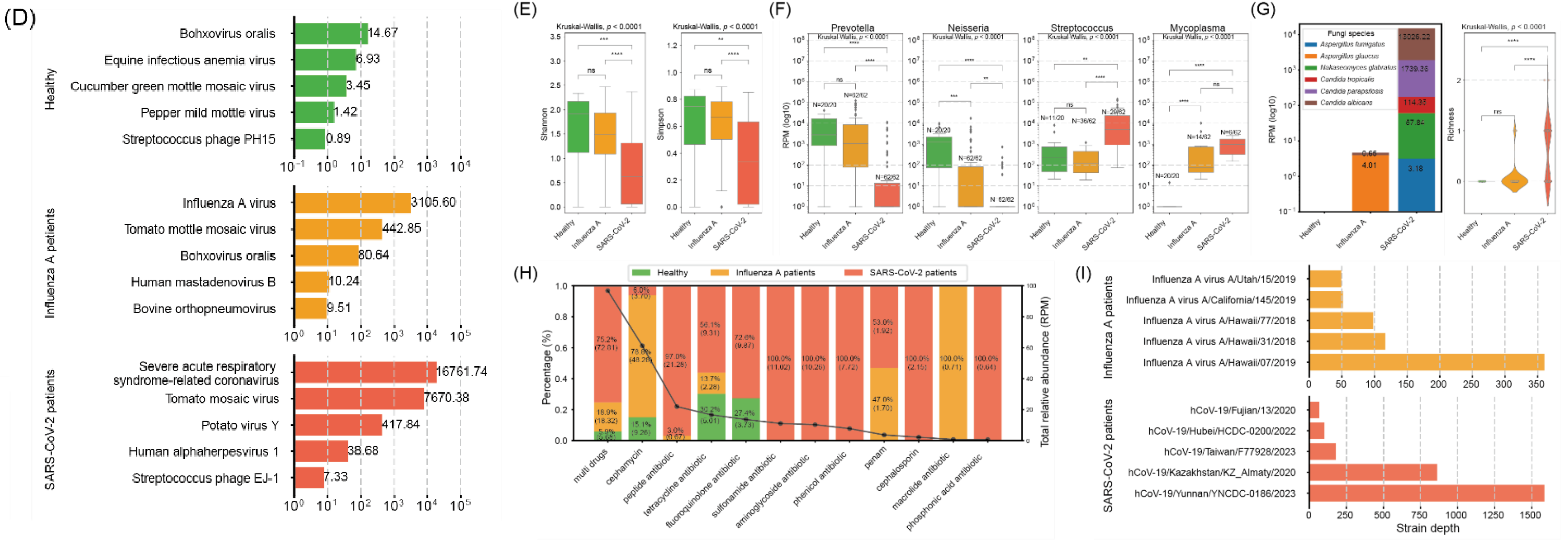
Validation and performance of RTTAP

Based on benchmarking tests using four mock samples, the RTTAP pipeline exhibits high accuracy along with high sensitivity, resulting in a superior F1 score.



Application to clinical samples

Through testing on 144 clinical samples, the RTTAP pipeline demonstrated accurate detection and classification capabilities while providing comprehensive information for downstream analysis.



Summary

- ❑ RTTAP represents an efficient and sensitive taxonomic analysis pipeline specifically optimized for metatranscriptomic infectome profiling, encompassing viruses, bacteria, and fungi.

- ❑ Rigorous validation using both simulated and clinical datasets ensures the RTTAP pipeline delivers reliable analytical performance.

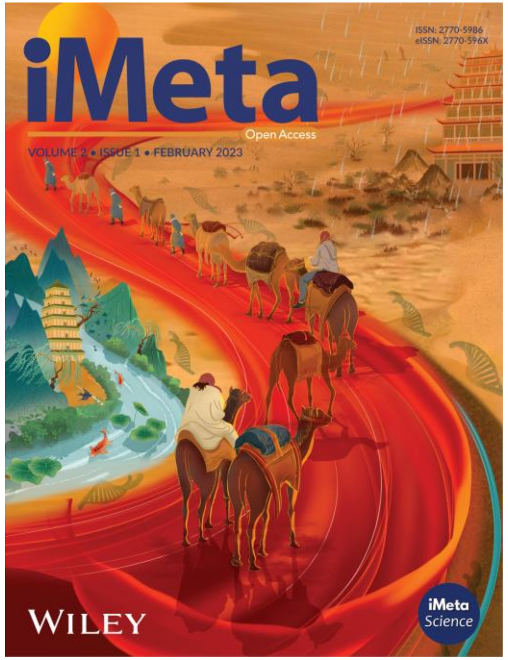
- ❑ The user-friendly design accommodates both routine analysis requirements and customizable needs for complex scenarios, ensuring accessibility for researchers with diverse backgrounds.

- ❑ Website: https://github.com/weijiang34/RTTAP

Wei Jiang, Herui Liao, Mang Shi, Liangjun Chen, Yanni Sun. 2025. RTTAP: Empowering metatranscriptomic data analysis with a read-based total-infectome taxonomic solution. ***iMetaOmics*** 2: e70044.
https://doi.org/10.1002/imo2.70044.

***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