

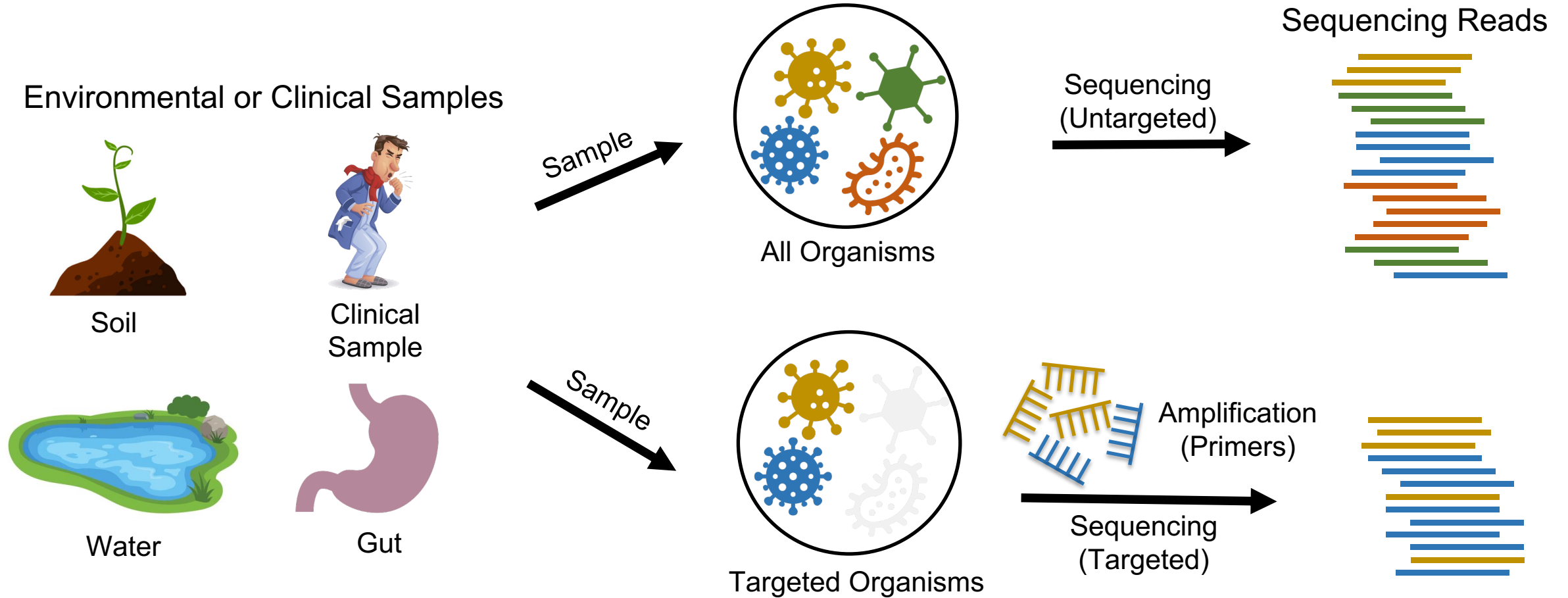
Phylogenetic Placement Enables Near-Haplotype Resolution of Metagenomic Samples using metaWEPP

Pranav Gangwar, Qiwen Xu, Jaden Seangmany, Pratik Katte, Yatish Turakhia

Outline

- Metagenomics: Introduction and Applications
- metaWEPP: Algorithm and Implementation Details
- Key Results
- Conclusion

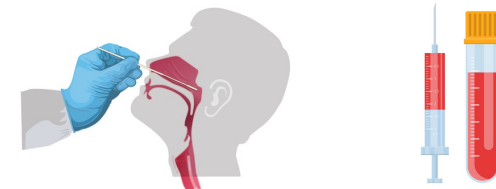
Metagenomics



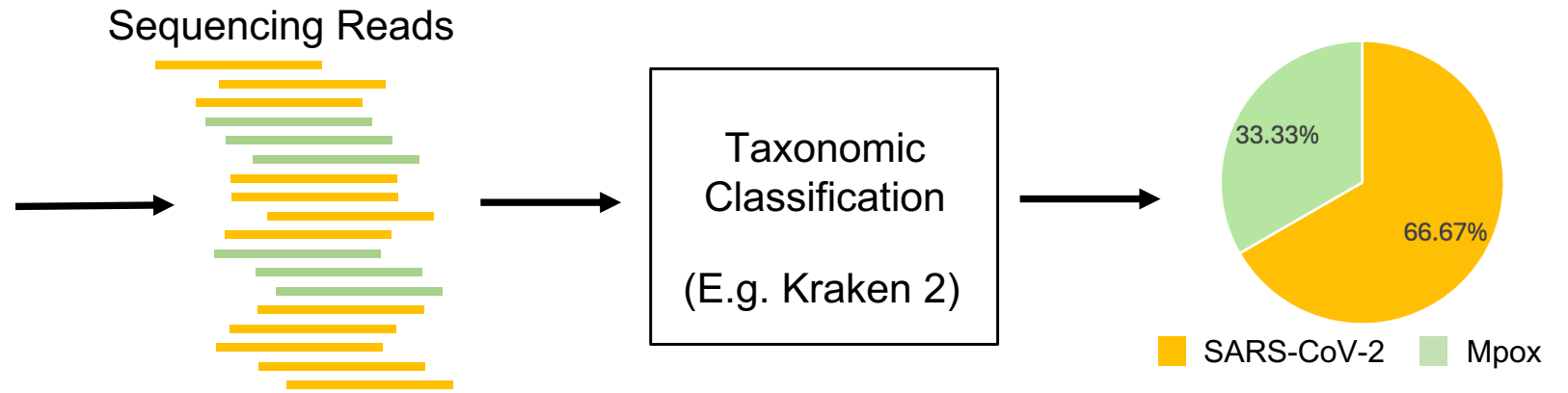
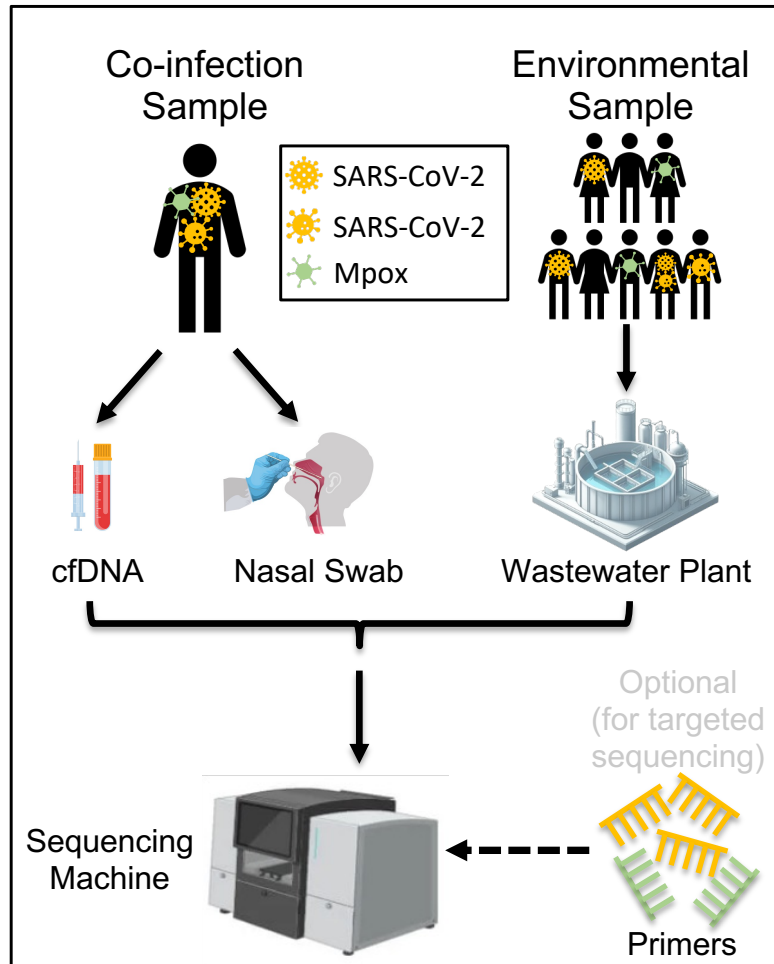
Metagenomics in Public Health and Point-of-Care Diagnostics



Applications of Interest



Existing Metagenomic Software - Species-Level Resolution



✗ No haplotype information



Pathogenicity

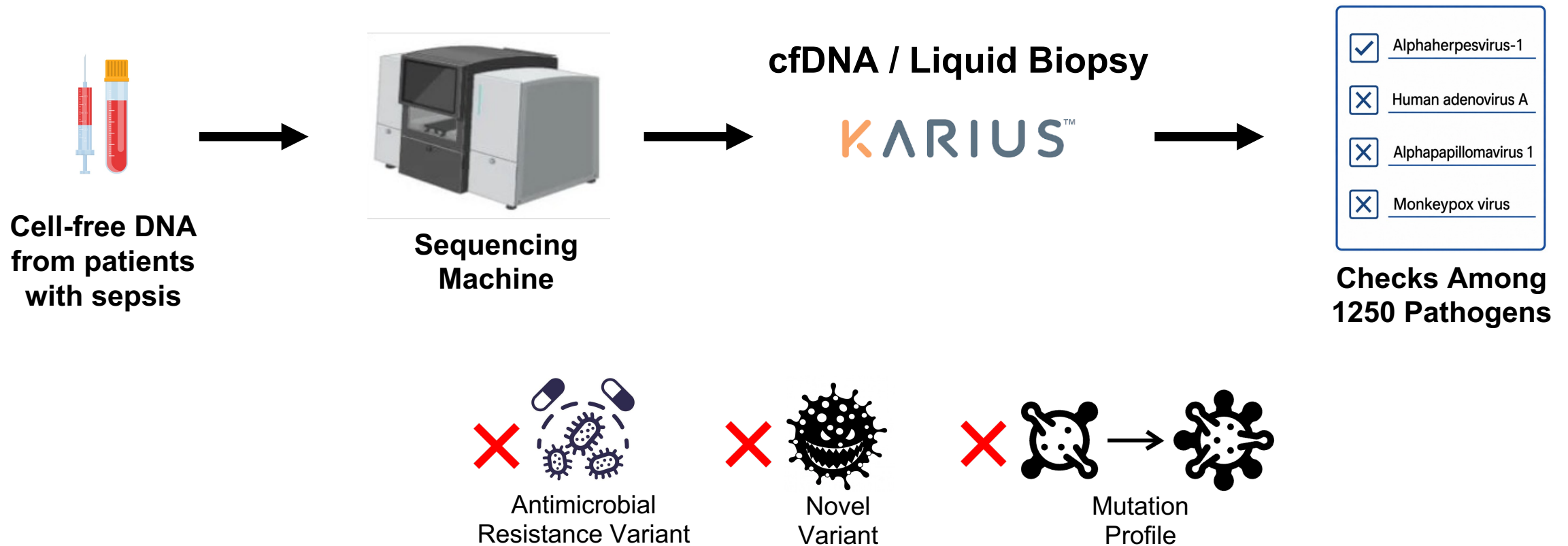


Antimicrobial
Resistance

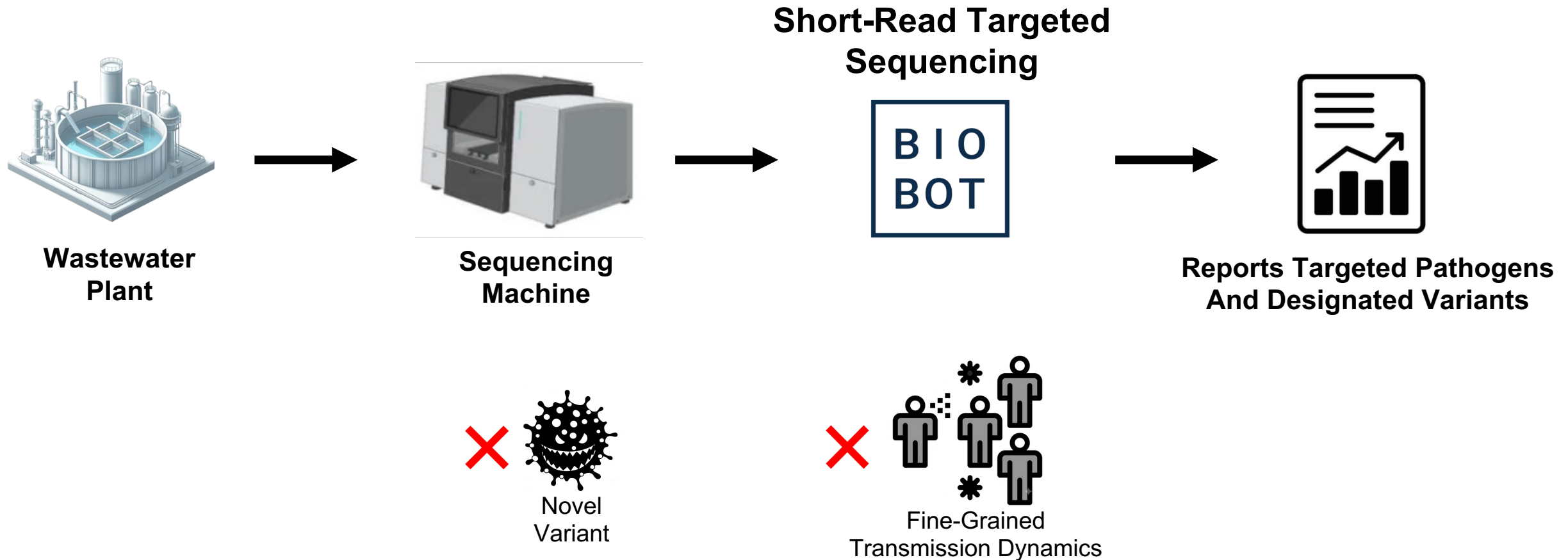


Immune
Evasion

Limitations of Point-of-Care Diagnostics Pipelines



Limitations of Public Health Surveillance Pipelines



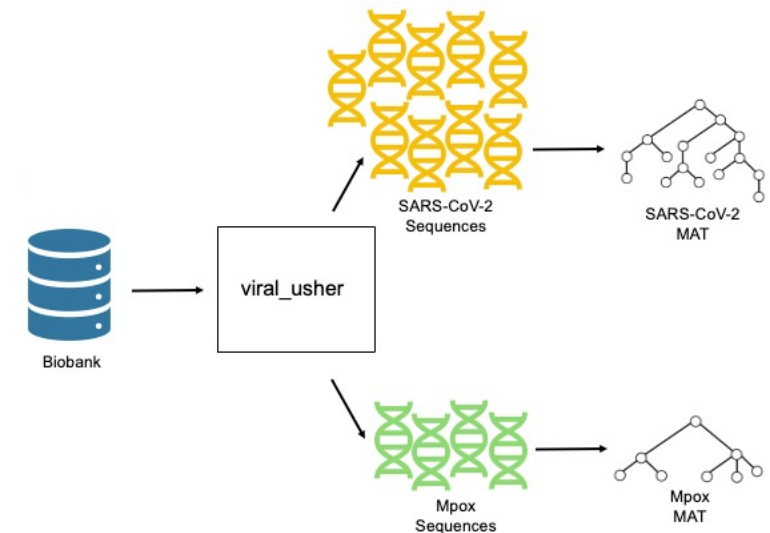
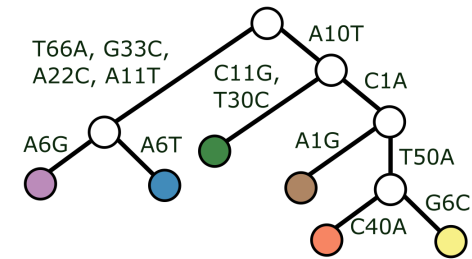
Outline

- Metagenomics: Introduction and Applications
- metaWEPP: Algorithm and Implementation Details
- Key Results
- Conclusion

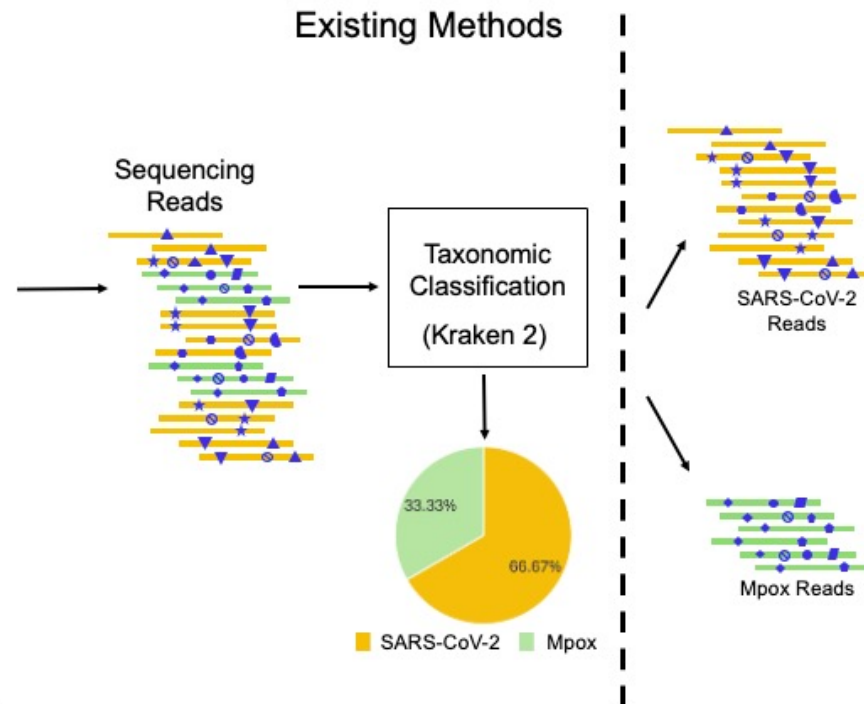
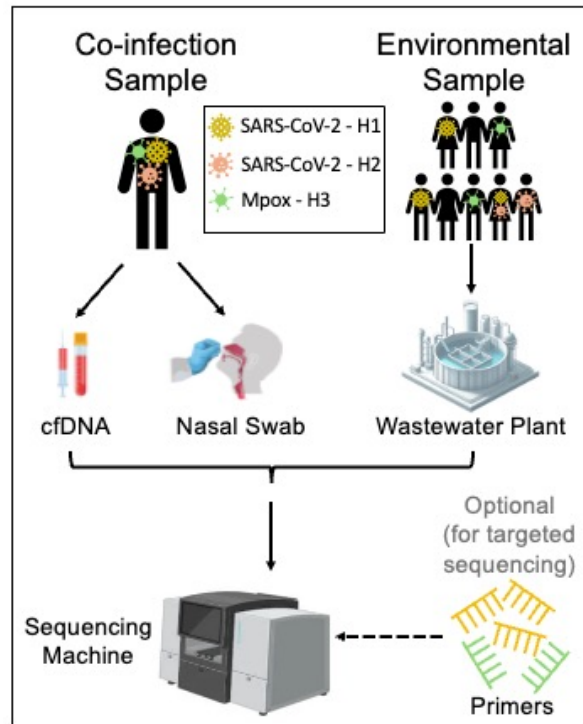
Mutation-Annotated Tree (MAT)

- Data structure storing evolutionary relationships between haplotypes, each branch annotated with mutations occurring along it.
- Prebuilt MATs: <https://dev.usher.bio>
- Viral_usher software for custom MATs: https://github.com/AngieHinrichs/viral_usher

Mutation-Annotated Tree

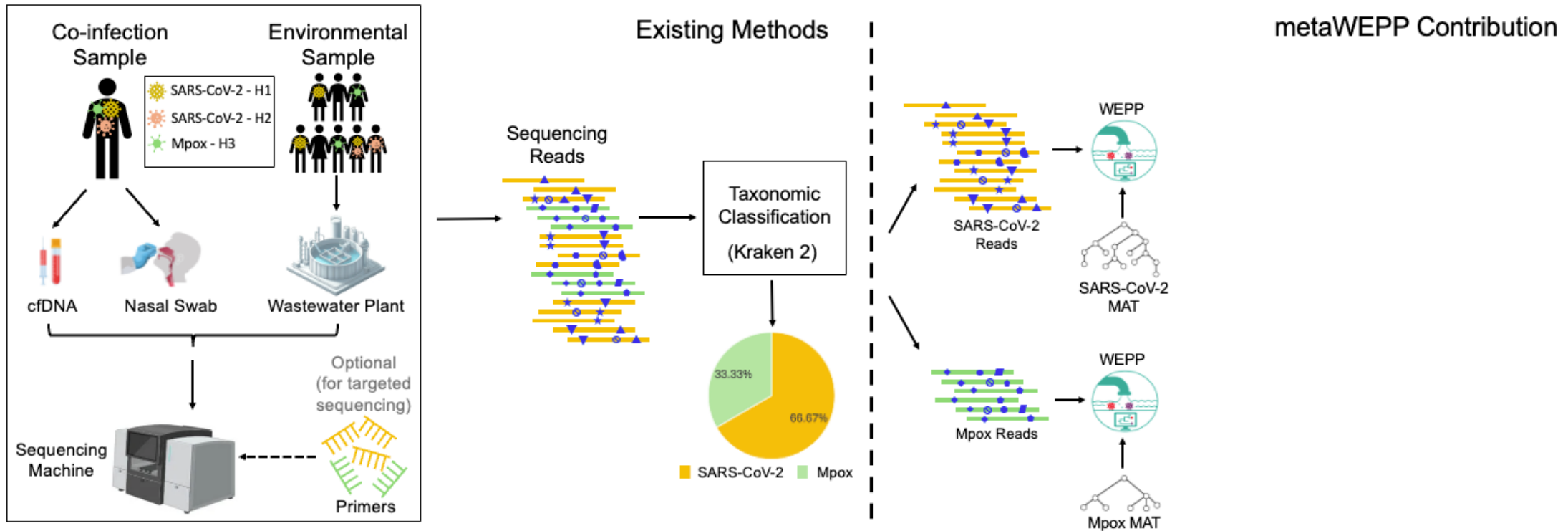


metaWEPP Overview

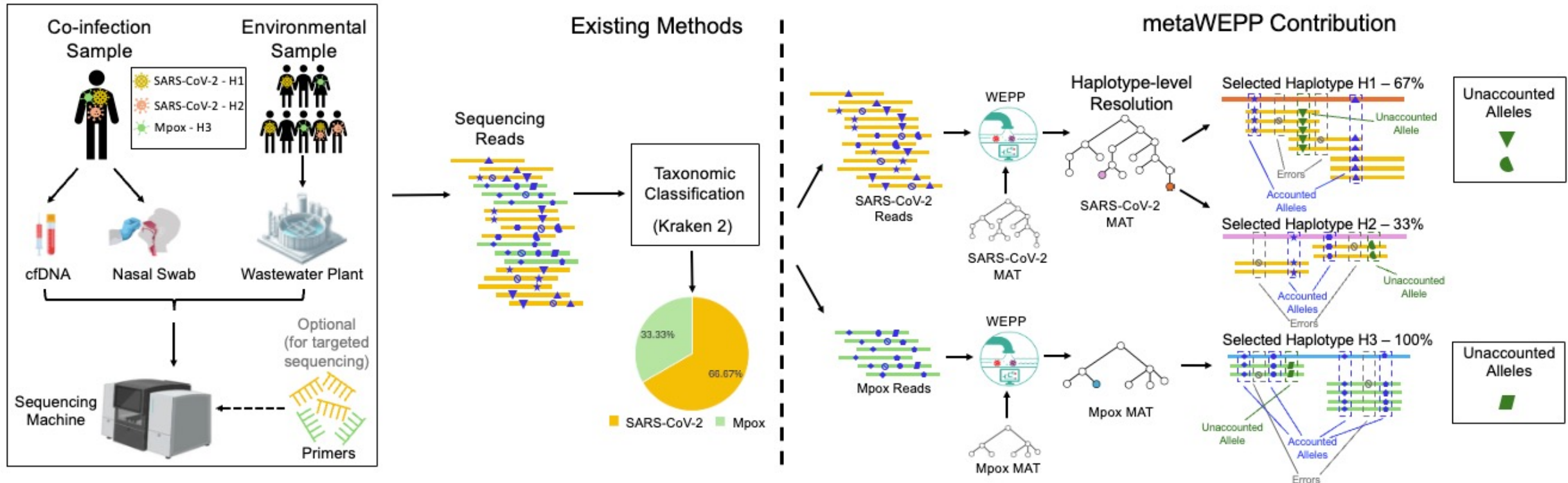


metaWEPP Contribution

metaWEPP Overview

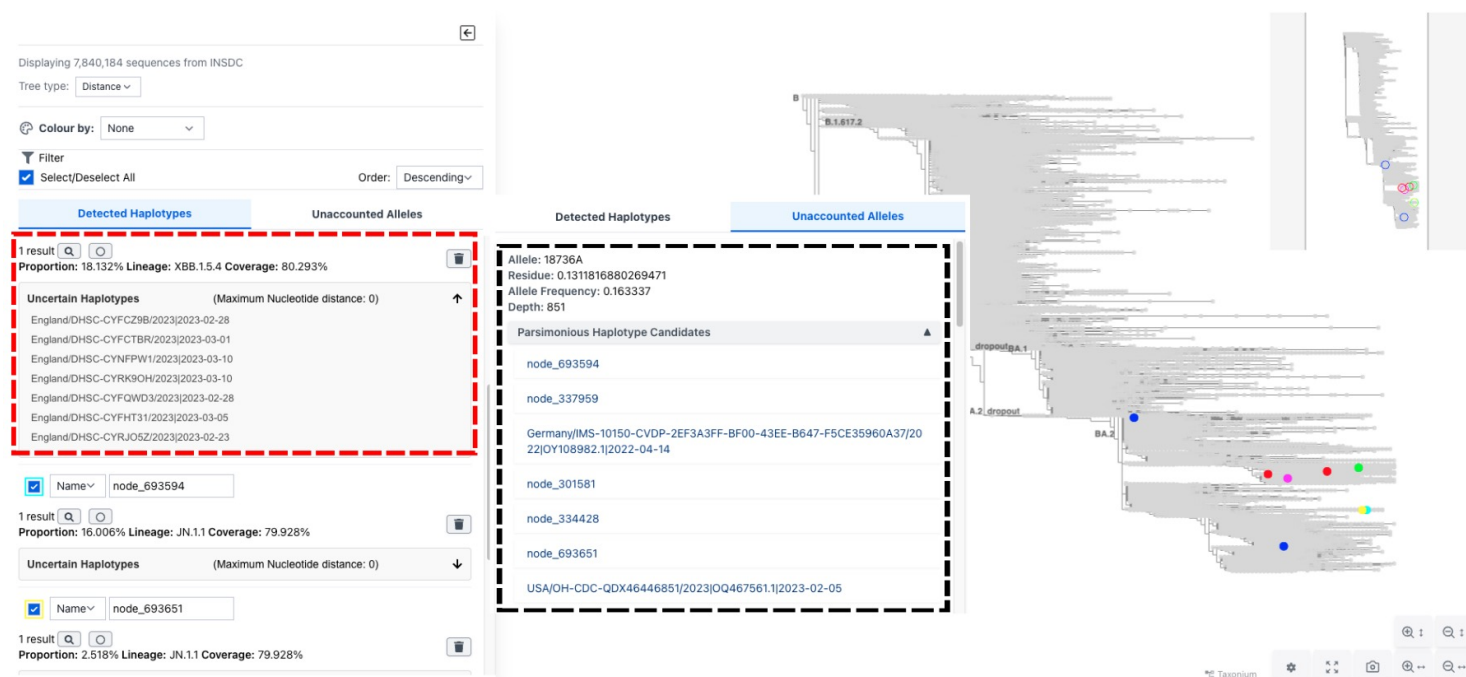


metaWEPP Overview

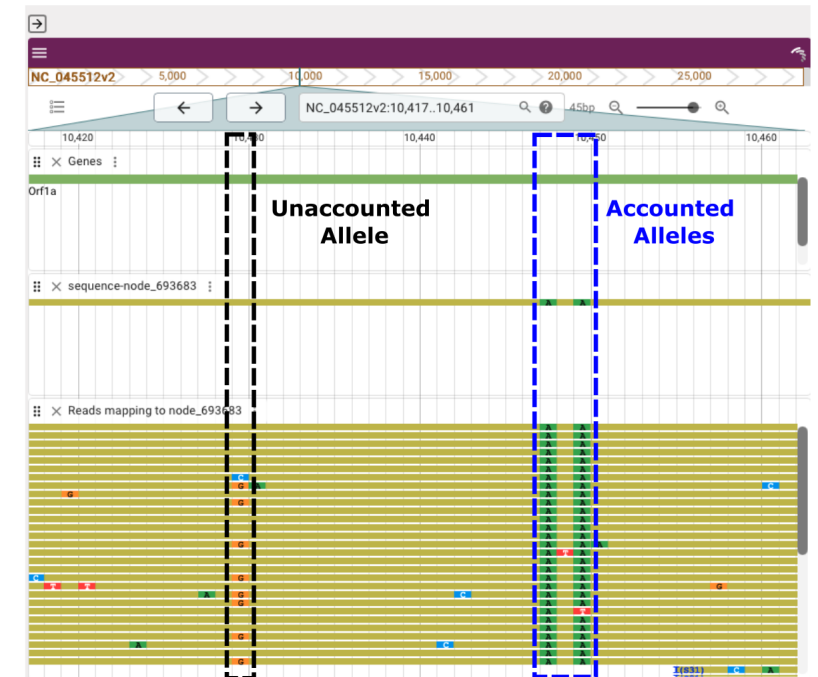


metaWEPP Dashboard

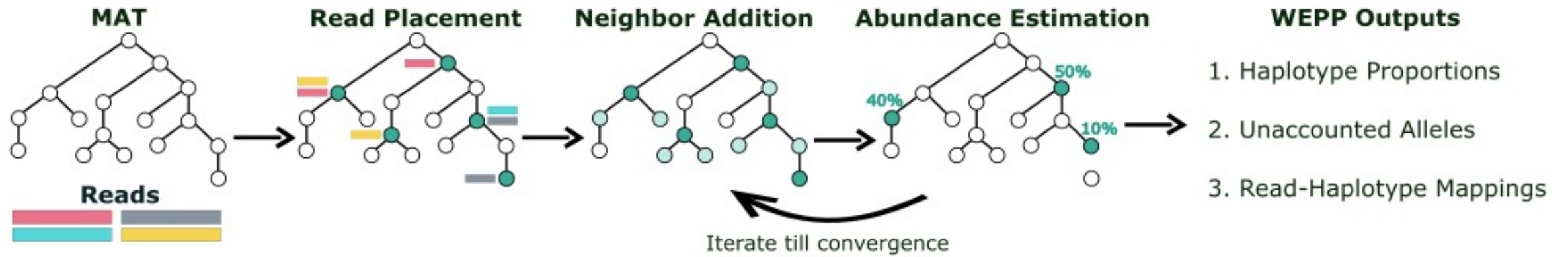
Phylogenetic View of Detected Haplotypes



Read-level Analysis



WEPP Algorithm



Outline

- Metagenomics: Introduction and Applications
- metaWEPP: Algorithm and Implementation Details
- Key Results
- Conclusion

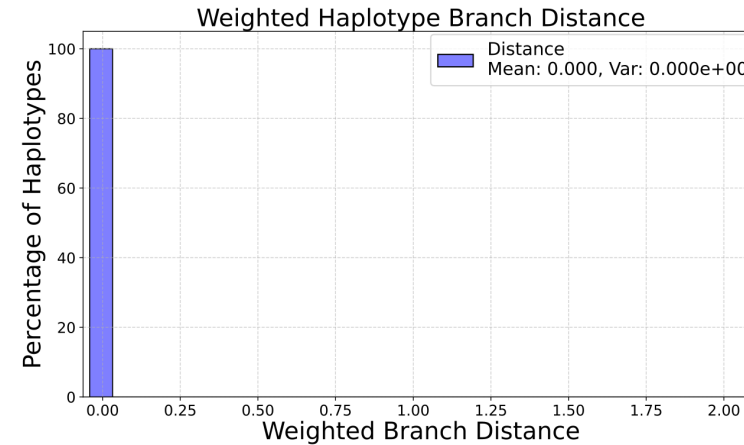
metaWEPP Results

- Simulated Metagenomic Datasets
- Targeted Sequencing of Clinical Samples
 - Multi-Pathogen
 - Single-Pathogen, Multi-Variant
- Untargeted Sequencing of Wastewater Samples
 - Multi-Pathogen

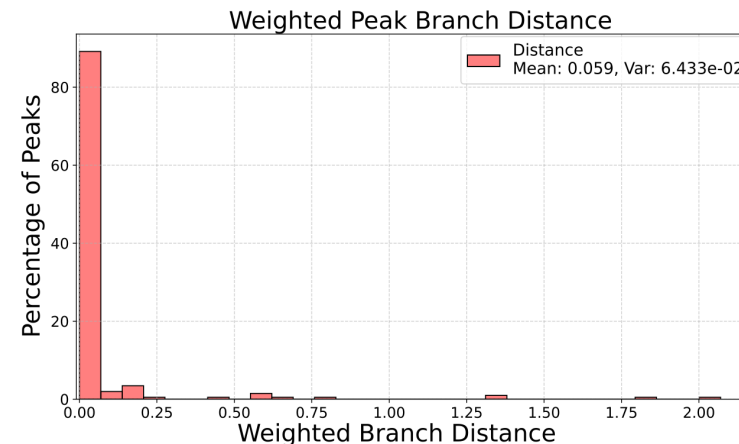
Detecting Haplotypes from Simulated Data

- Simulated 24 samples, similar in complexity as low-complexity CAMI dataset
- Weighted haplotype branch distance (Sensitivity)
Mean Value - 0
- Weighted peak branch distance (Precision)
Mean value - 0.059

High Sensitivity, High Precision



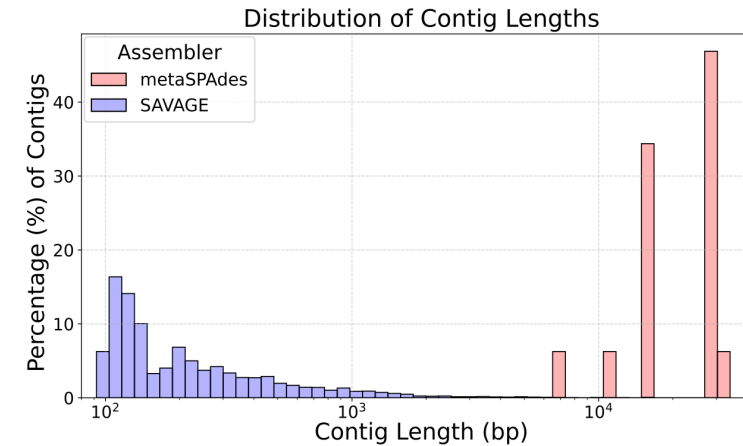
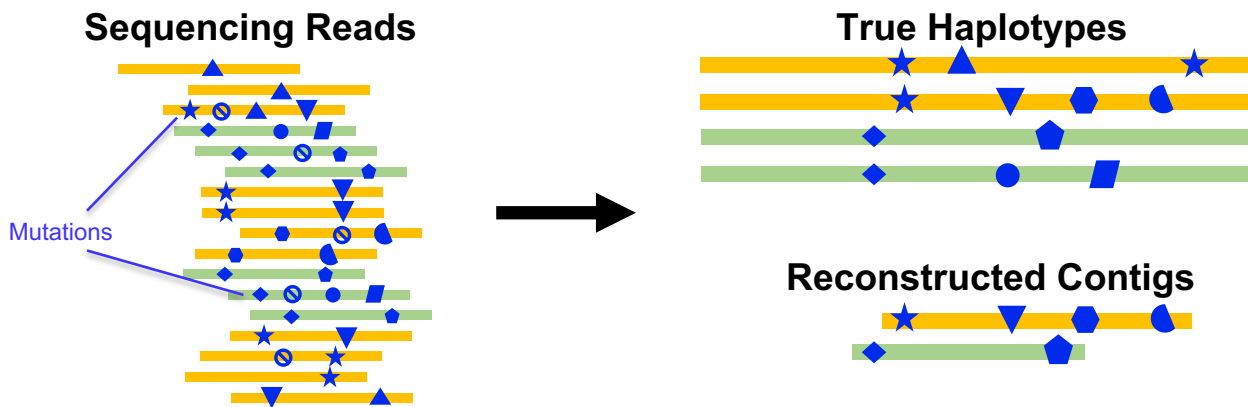
Sensitivity Proxy
(Lower is better)



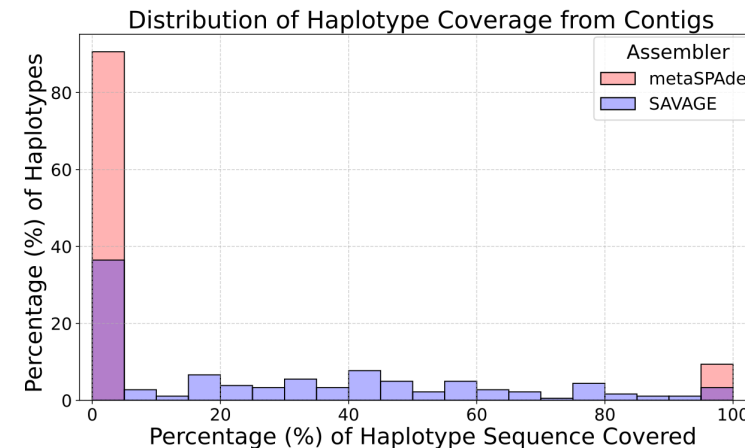
Precision Proxy
(Lower is better)

Detecting Haplotypes from Simulated Data

- De novo assemblers, SAVAGE and metaSPAdes, produced fragmented assemblies and missed several haplotypes
- Low-diversity variants collapse into single variant



Fragmented
Contigs

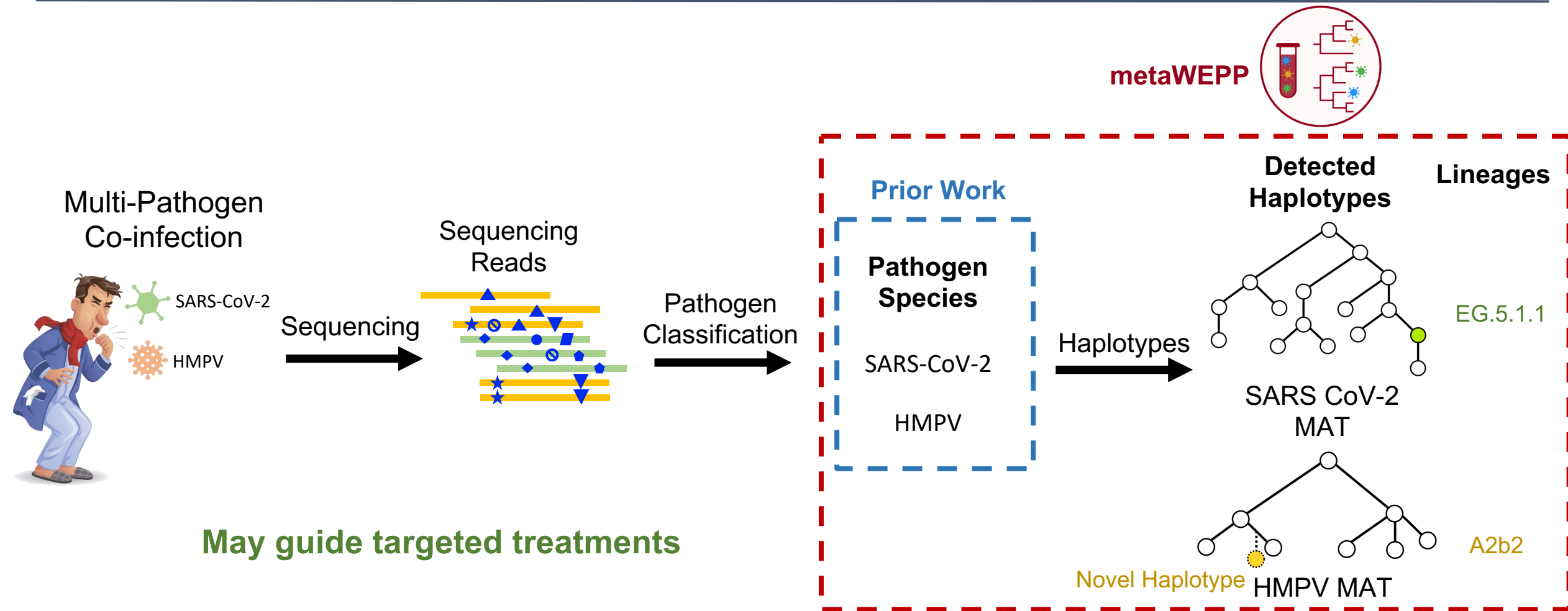


Low Haplotype
Coverage

metaWEPP Results

- Simulated Metagenomic Datasets
- Targeted Sequencing of Clinical Samples
 - Multi-Pathogen
 - Single-Pathogen, Multi-Variant
- Untargeted Sequencing of Wastewater Samples
 - Multi-Pathogen

Detecting Clinical Co-Infections with Multiple Pathogens

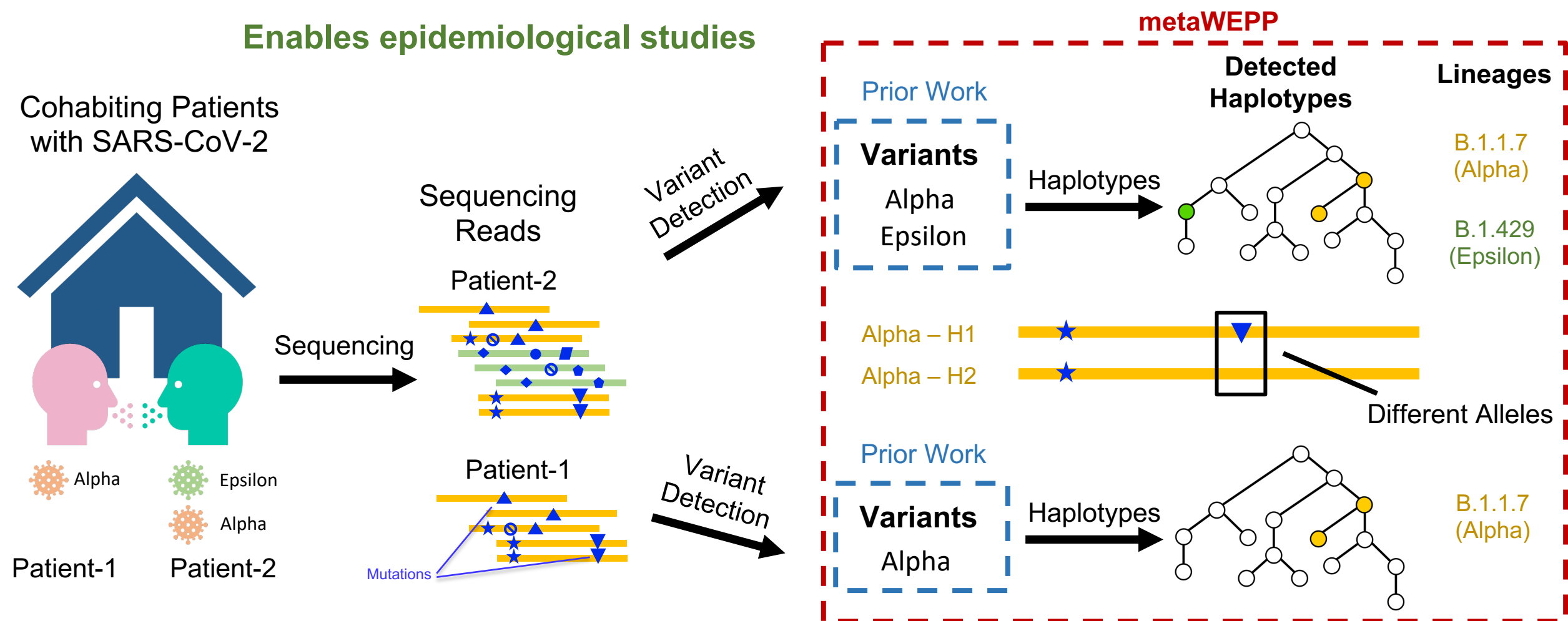


metaWEPP Results

- Simulated Metagenomic Datasets
- Targeted Sequencing of Clinical Samples
 - Multi-Pathogen
 - Single-Pathogen, Multi-Variant
- Untargeted Sequencing of Wastewater Samples
 - Multi-Pathogen

Detecting Clinical Co-Infections with Multiple Variants of SARS-CoV-2

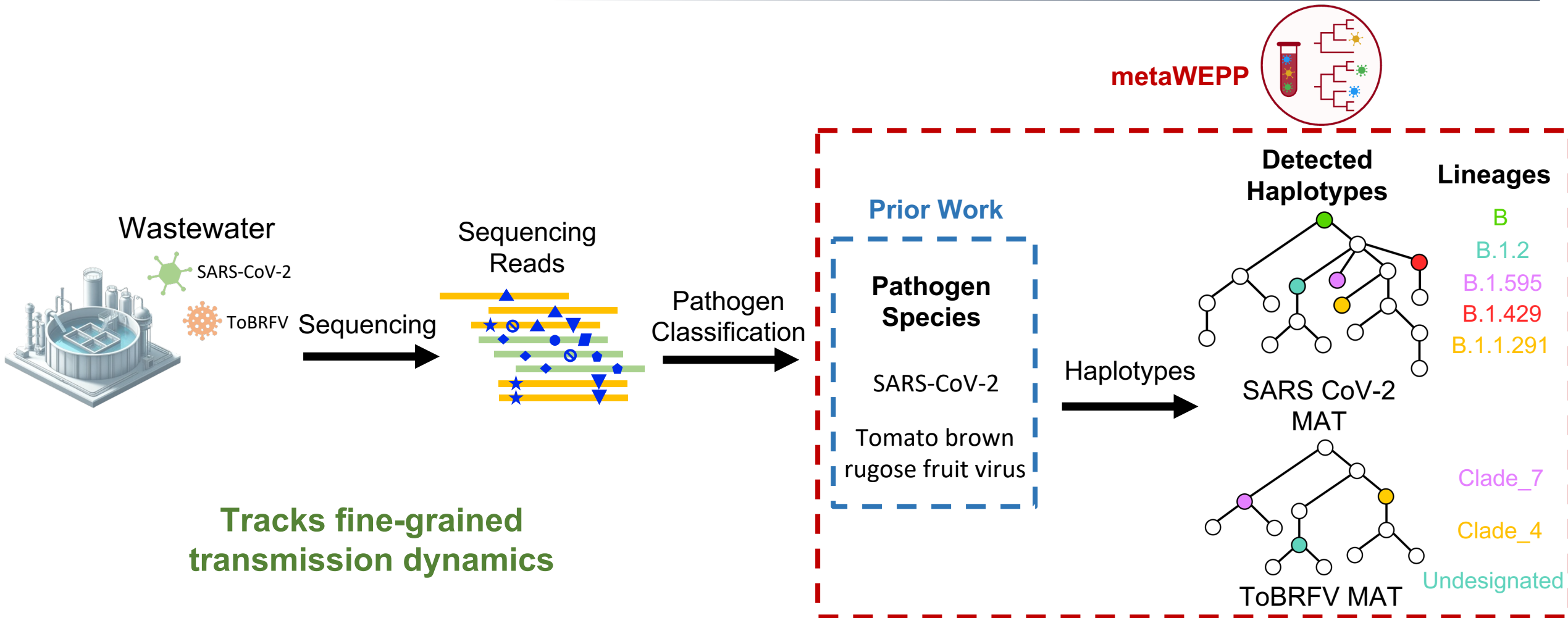
Enables epidemiological studies



metaWEPP Results

- Simulated Metagenomic Datasets
- Targeted Sequencing of Clinical Samples
 - Multi-Pathogen
 - Single-Pathogen, Multi-Variant
- Untargeted Sequencing of Wastewater Samples
 - Multi-Pathogen

Detecting Haplotype Clusters Across Pathogens from Wastewater Samples



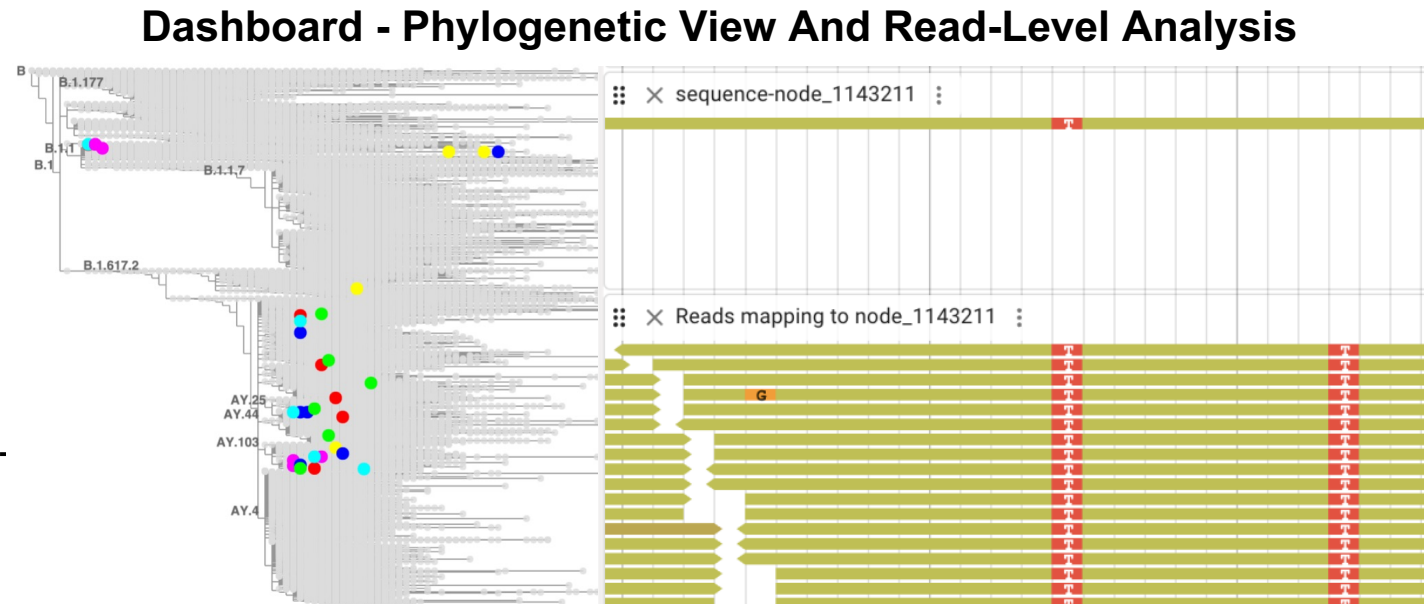
Outline

- Metagenomics: Introduction and Applications
- metaWEPP: Algorithm and Implementation Details
- Key Results
- Conclusion

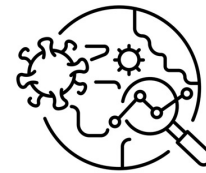
metaWEPP Summary

Contributions:

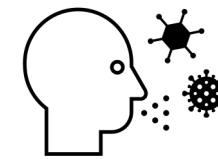
1. Near-haplotype resolution
 - i. Detects haplotype clusters in environmental surveillance
 - ii. Detects co-infecting variants in clinical samples
2. Novel variant detection
3. Dashboard's phylogenetic and read-level analysis



New Opportunities



Pathogen
Surveillance



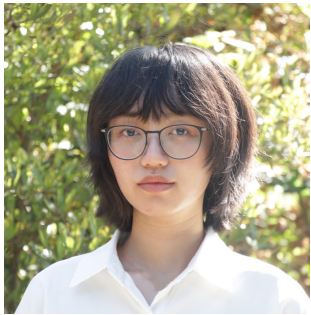
Co-infection
Analysis



Targeted Drug
Therapies

Acknowledgements

Co-authors



Qiwen Xu



Jaden Seangmany



Pratik Katte



Yatish Turakhia

Thanks for Feedback

Angie Hinrichs

Jason Caravas

Russell Corbett-Detig

Funding



Published in



metaWEPP
Documentation