

Pranav Gangwar<sup>1</sup>, Pratik Katte<sup>2</sup>, Manu Bhat<sup>1</sup>, Yatish Turakhia<sup>1</sup>

<sup>1</sup>University of California San Diego, San Diego, CA 92093, United States

<sup>2</sup>University of California Santa Cruz, Santa Cruz, CA 95064, USA

## Wastewater-Based Epidemiology (WBE): Introduction and Key Challenges

- WBE enables cost-effective and unbiased monitoring of community health, detecting pathogens and their variants days to weeks before clinical reporting.
- Current WBE tools focus on estimating lineage proportions, **limiting critical epidemiological applications** that require *haplotype-level* resolution, such as:
  - Tracking intra-lineage haplotype clusters
  - Identifying regional introductions of haplotype clusters
  - Detecting emerging variants before their lineage designation

## Wastewater-Based Epidemiology using Phylogenetic Placements (WEPP)

- WEPP is a pathogen-agnostic pipeline that places wastewater reads on a global phylogeny to report (Figure 1):
  - Haplotype and Lineage Proportions
  - 'Unaccounted Alleles' – alleles observed in the sample but unexplained by selected haplotypes
  - Parsimonious read to haplotype assignment
- Provides an interactive dashboard to explore haplotypes and 'Unaccounted Alleles' in the phylogenetic context (Figure 2).

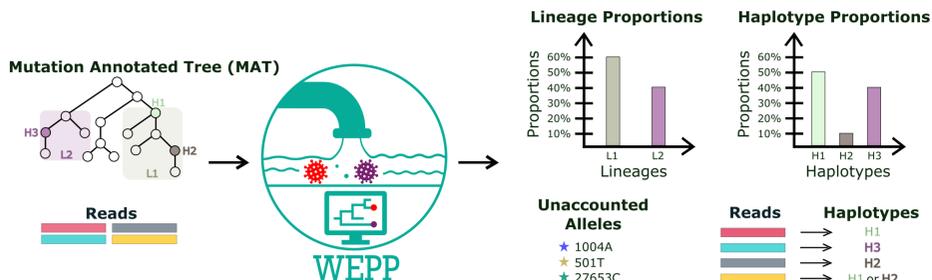


Figure 1: WEPP input and output

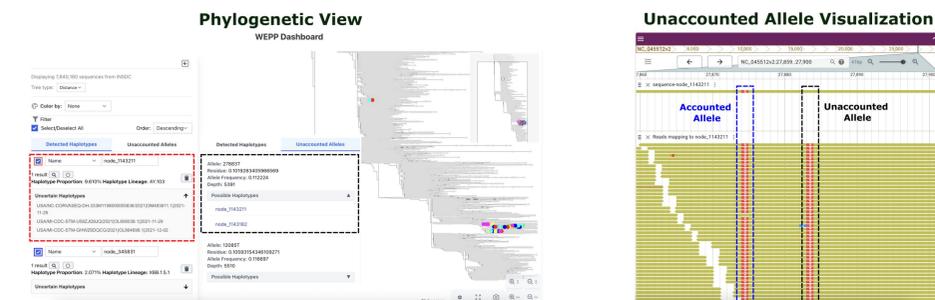


Figure 2: WEPP Dashboard: Phylogenetic View (left) and Read Analysis (right).

## Key Results and Contributions

### Detects Intra-Lineage Clusters

WEPP accurately identified multiple intra-lineage haplotype clusters circulating in San Diego in September 2023 (Figure 3).

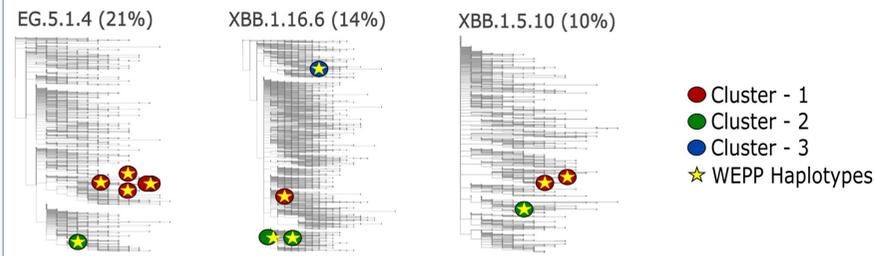


Figure 3: Intra-lineage cluster detection by WEPP.

### New Insights from Real Wastewater

- WEPP outperformed Freyja<sup>1</sup> in correlating lineage abundance with one-week-delayed "Oracle" clinical sequences from San Diego (Figure 5).
- Two independent Omicron introductions were identified in San Diego; 32 detected haplotypes matched clinical sequences from the city.
- 35.8% 'Unaccounted Alleles' were absent from clinical sequences, and another 43% were rare (<1%), indicating hidden transmissions.

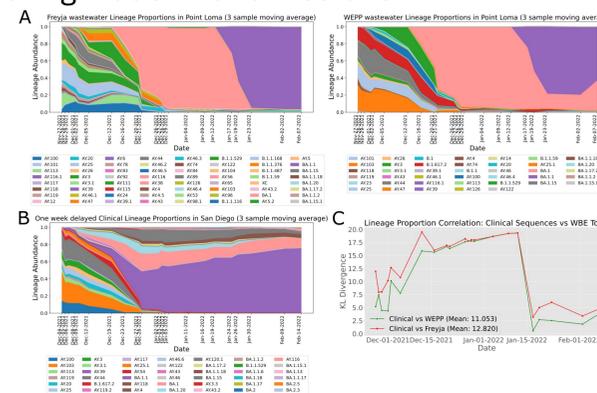


Figure 5: (A) SARS-CoV-2 Lineage proportions from Freyja (left) and WEPP (right). (B) Proportions from "Oracle" clinical sequences. (C) KL divergence: Clinical vs Freyja and WEPP.

### Early Detection of Cluster Introductions

WEPP detected a California-originating haplotype cluster in Delaware using a MAT that excluded local haplotypes, enabling early detection of new introductions (Figure 4).

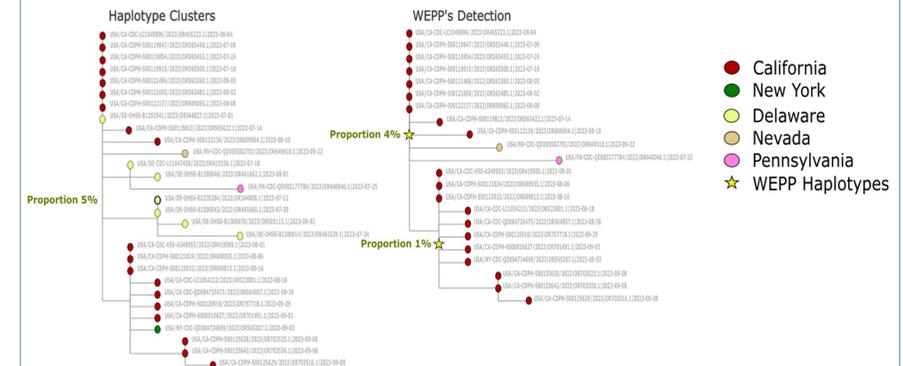


Figure 4: Likely variant introduction from California to Delaware (left). WEPP detection with Delaware haplotypes excluded from MAT (right).

### Reveals Clinically Overlooked RSV-A Lineages

WEPP consistently detected lineages A.D and A.D.5 in wastewater from Geneva, undetected by clinical sequencing and V-Pipe<sup>2</sup> (Figure 6).

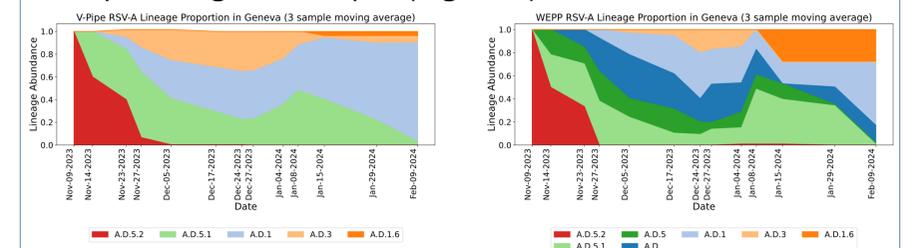


Figure 6: RSV-A lineage proportions: V-Pipe (left) vs WEPP (right).

## References

- [1] Karthikeyan, S. et al. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature* 609, 101–108 (2022).
- [2] De Korne-Elenbaas, J. et al. Wastewater-based sequencing of Respiratory Syncytial Virus enables tracking of lineages and identifying mutations at antigenic sites. Preprint at <https://doi.org/10.1101/2025.02.28.25321637> (2025).

This project is a part of **Turakhia Lab**: <https://turakhia.ucsd.edu>  
 WEPP source code is available under the MIT License on GitHub: <https://github.com/TurakhiaLab/WEPP>