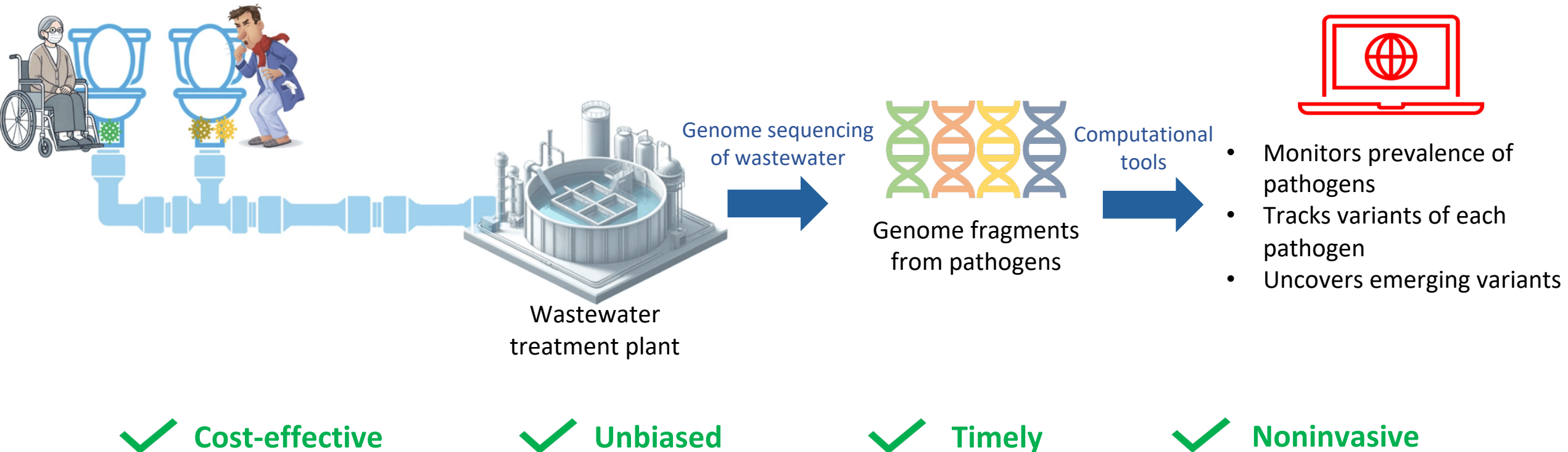# WEPP: Phylogenetic Placement Achieves Near-Haplotype Resolution in Wastewater-Based Epidemiology

Speakers: Pranav Gangwar and Pratik Katte

Wastewater-Based Epidemiology (WBE)

Genome sequencing of wastewater

Computational tools

Wastewater treatment plant

Genome fragments from pathogens

- Monitors prevalence of pathogens
- Tracks variants of each pathogen
- Uncovers emerging variants

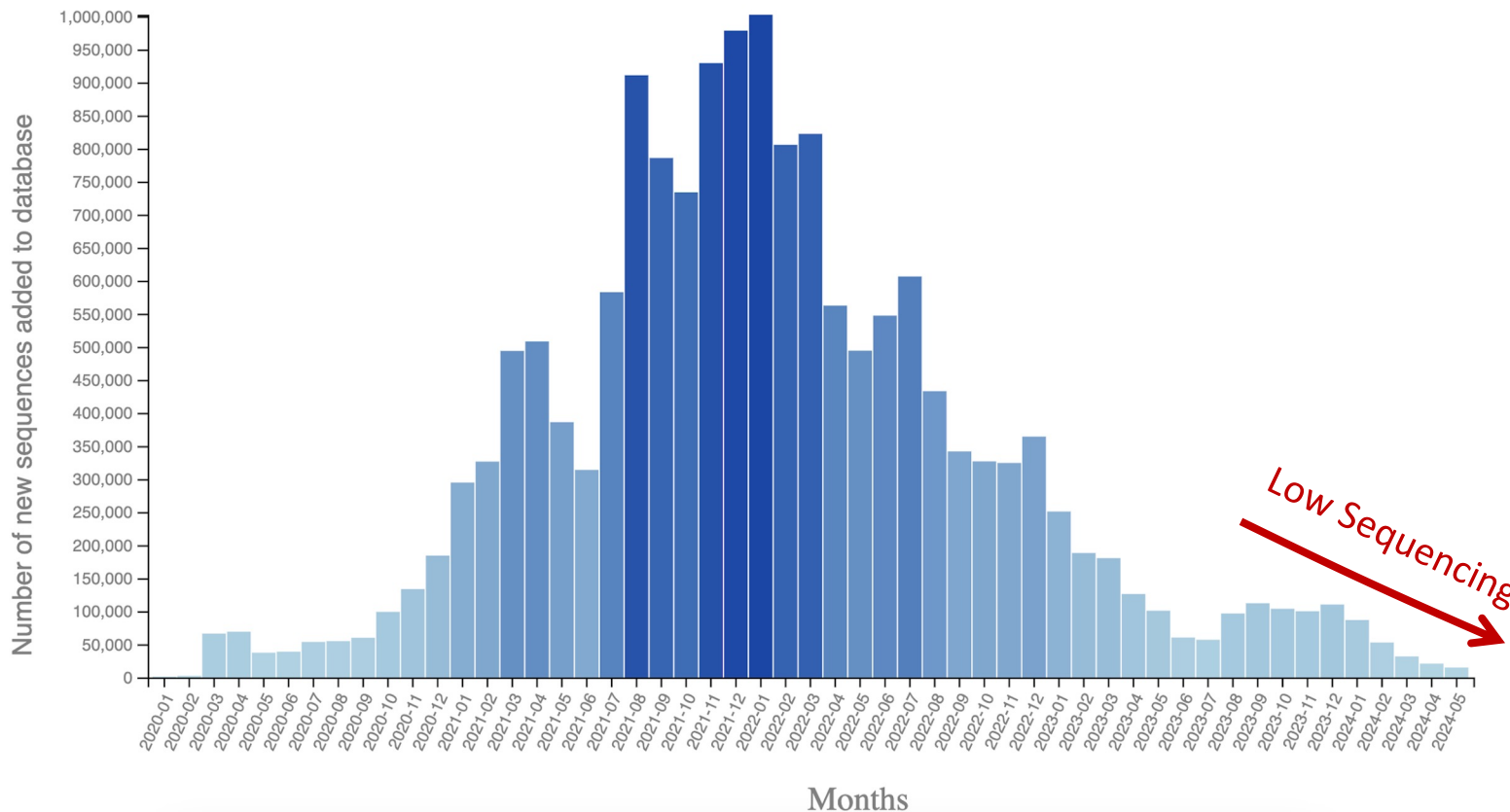✓ **Cost-effective**　　✓ **Unbiased**　　✓ **Timely**　　✓ **Noninvasive**

WBE Challenges

- Sequencing reads from multiple genomes with low genetic diversity

- Short reads with uneven coverage and sequencing errors

- <u>Existing tools focus on estimating lineage abundances</u>

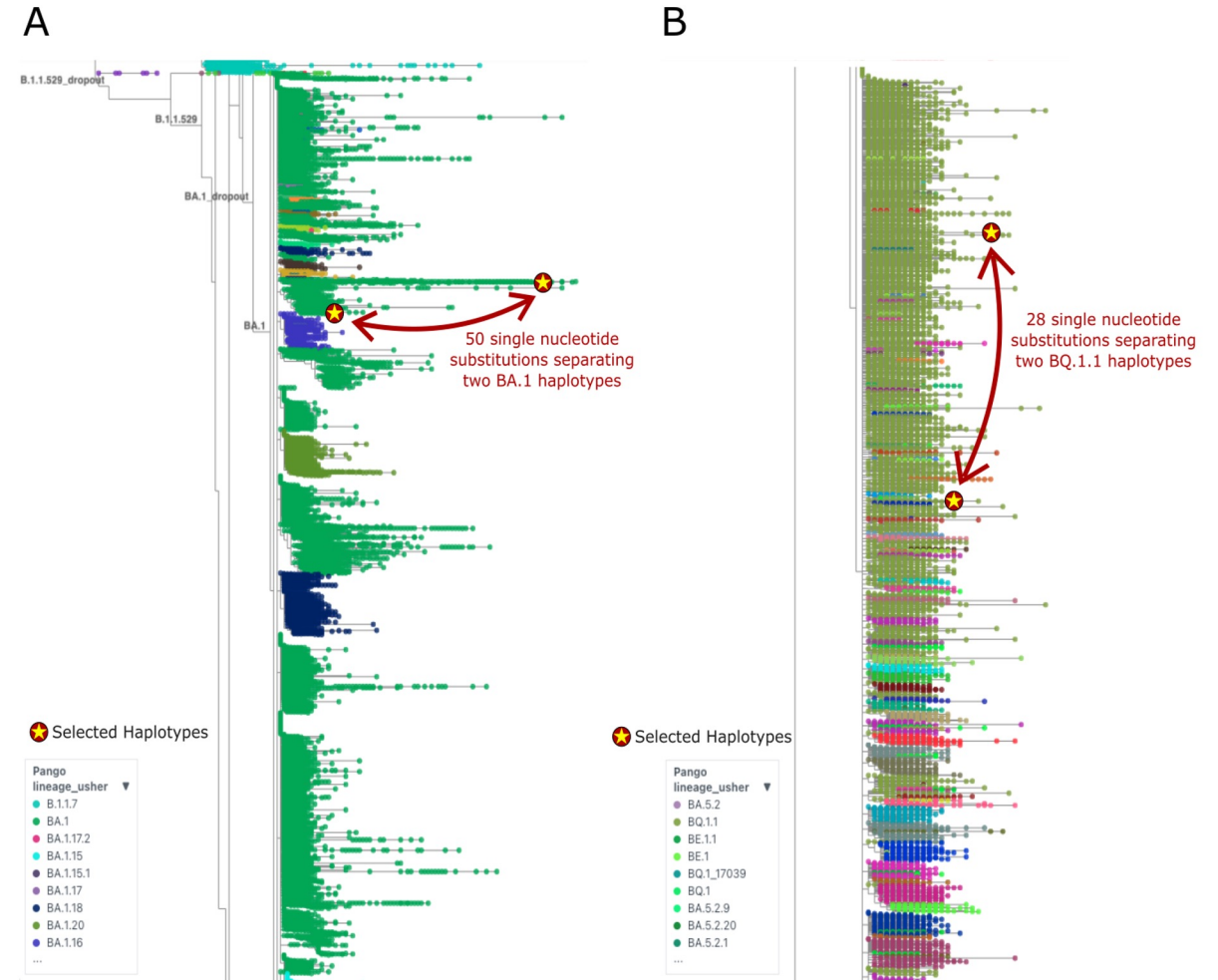# Reduced WBE effectiveness with Declining Clinical Sequencing

SARS-CoV-2 Clinical Sequencing

- Hard to detect new variants with reduced sequencing

- Longer delays in designating new lineages

- Losing valuable time in tracking new variants through wastewater with exiting tools

Limitations of Existing WBE Tools

1. **Resolution:** Cannot disambiguate intra-lineage diversity
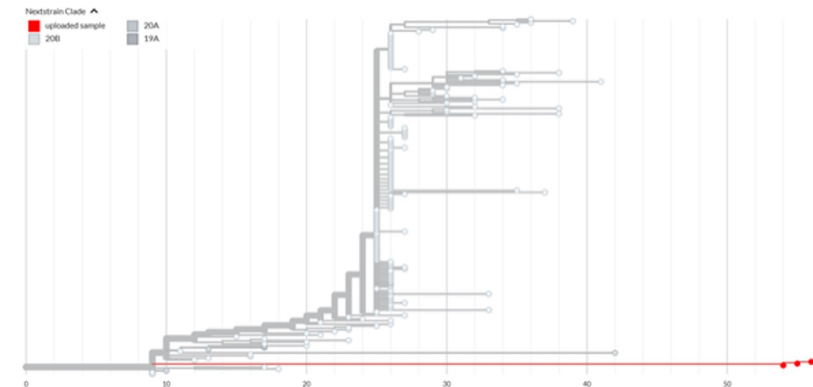
Limitations of Existing WBE Tools

1. **Resolution:** Cannot disambiguate intra-lineage diversity

2. **Accuracy:** Lineage abundance estimates get inaccurate, especially for large lineages

Lineage Abundance on SWAMPY's Simulated Dataset

Pipes et al. RMSE: 0.0939
Freyja RMSE: 0.0595
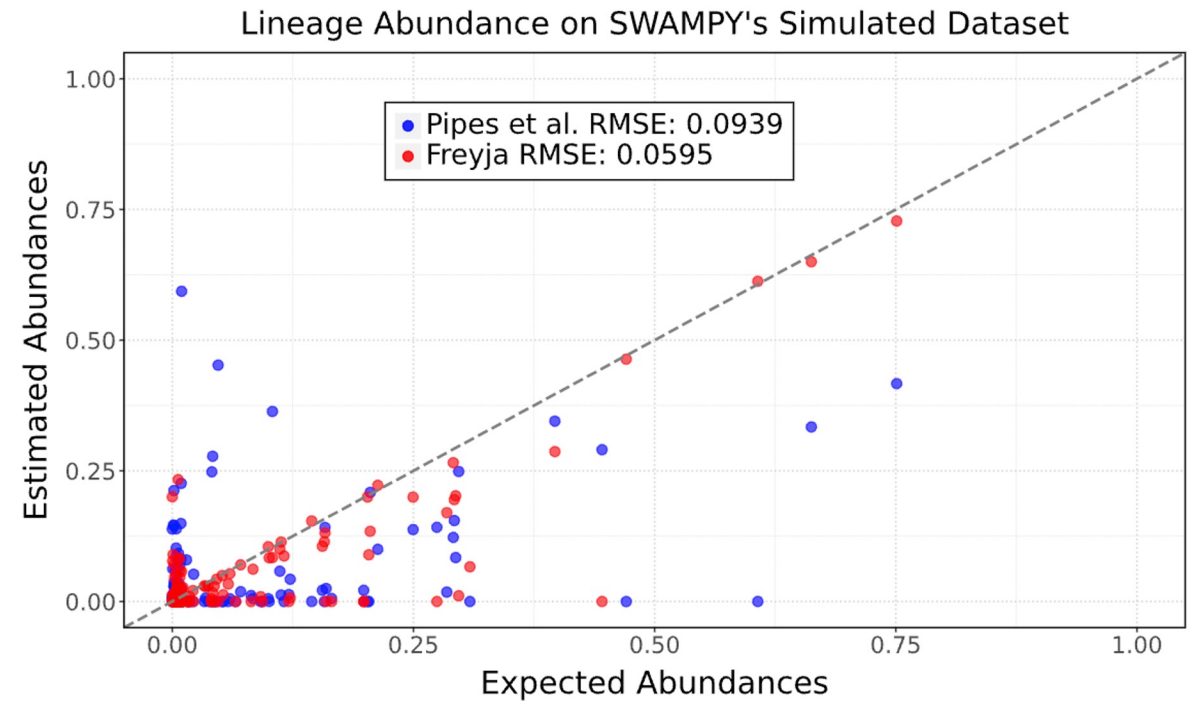
Limitations of Existing WBE Tools

1. **Resolution:** Cannot disambiguate intra-lineage diversity

2. **Accuracy:** Lineage abundance estimates get inaccurate, especially for large lineages

3. **Timeliness:** Unable to discover novel, undesignated lineages

Omicron introduction

https://github.com/cov-lineages/pango-designation/issues/343

WEPP Overview

Wastewater-Based **E**pidemiology using **P**hylogenetic **P**lacements

WEPP's Resolution Advantage

Previous WBE analysis
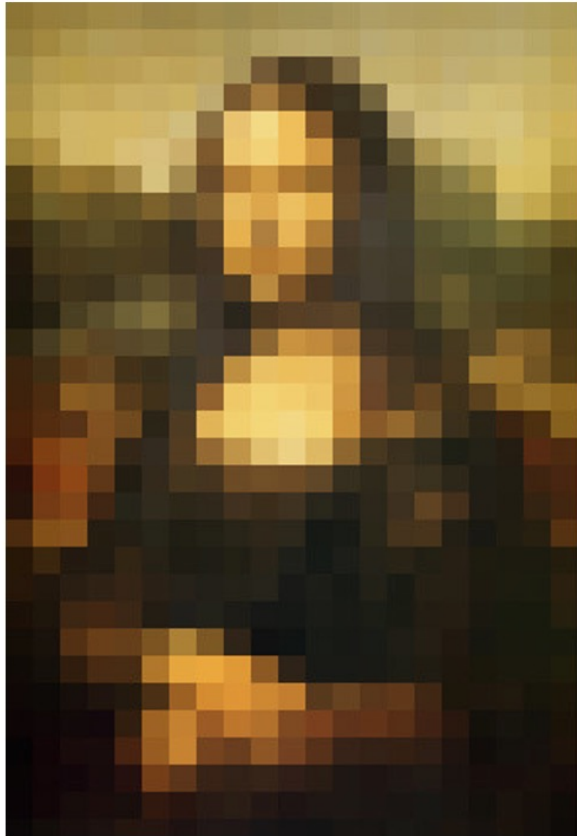
WBE analysis with WEPP

WEPP: Algorithm Overview
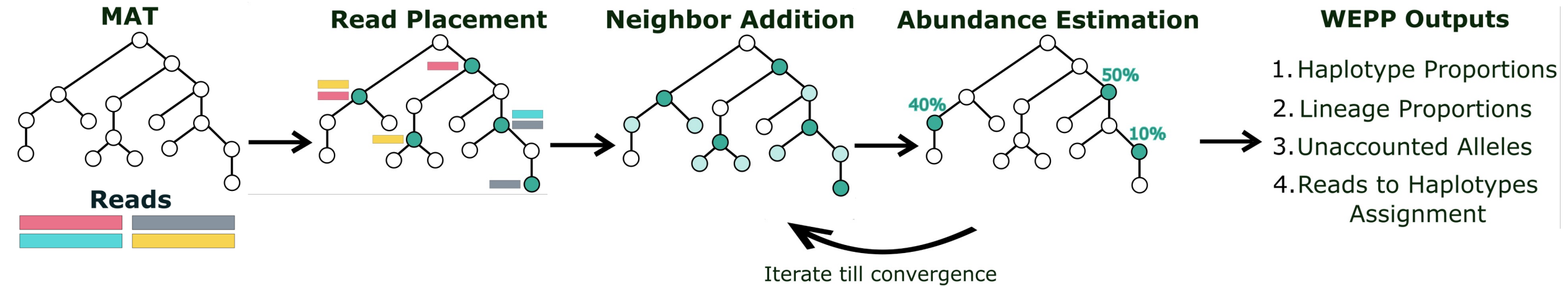
MAT

Reads

Read Placement

Neighbor Addition

Abundance Estimation

40% 50% 10%

Iterate till convergence

WEPP Outputs

1. Haplotype Proportions
2. Lineage Proportions
3. Unaccounted Alleles
4. Reads to Haplotypes Assignment

WEPP: Key Results

1. More accurate lineage abundances

Lineage Abundance on SWAMPY's Simulated Dataset

Pipes et al. RMSE: 0.0939
Freyja RMSE: 0.0595
WEPP RMSE: 0.0119
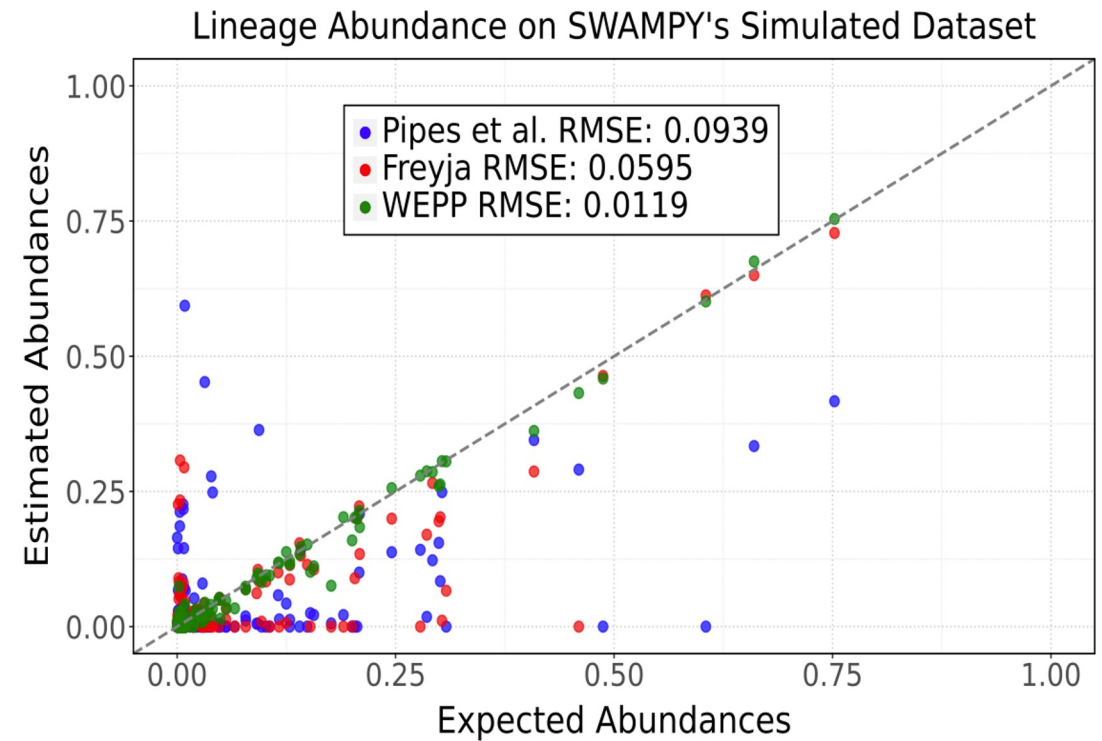
WEPP: Key Results

1. More accurate lineage abundances

2. Identifies intra-lineage clusters

EG.5.1.4 (21%)
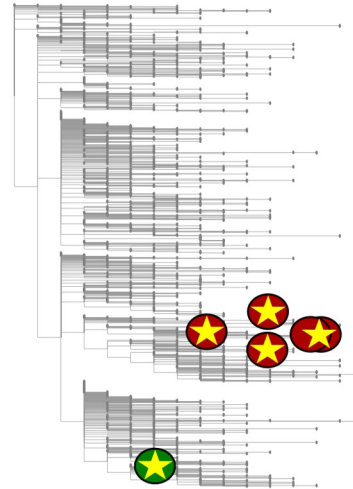
XBB.1.16.6 (14%)
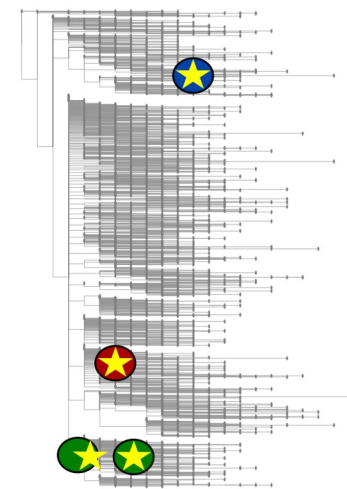
XBB.1.5.10 (10%)
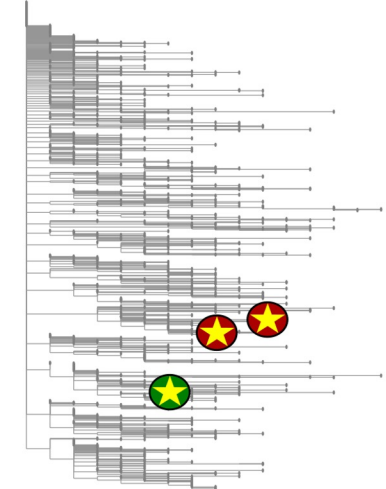
● Cluster - 1
● Cluster - 2
● Cluster - 3
★ WEPP Haplotypes

WEPP: Key Results

1. More accurate lineage abundances

2. Identifies intra-lineage clusters

3. Detects introduction of new clusters

CA -> Delaware introduction
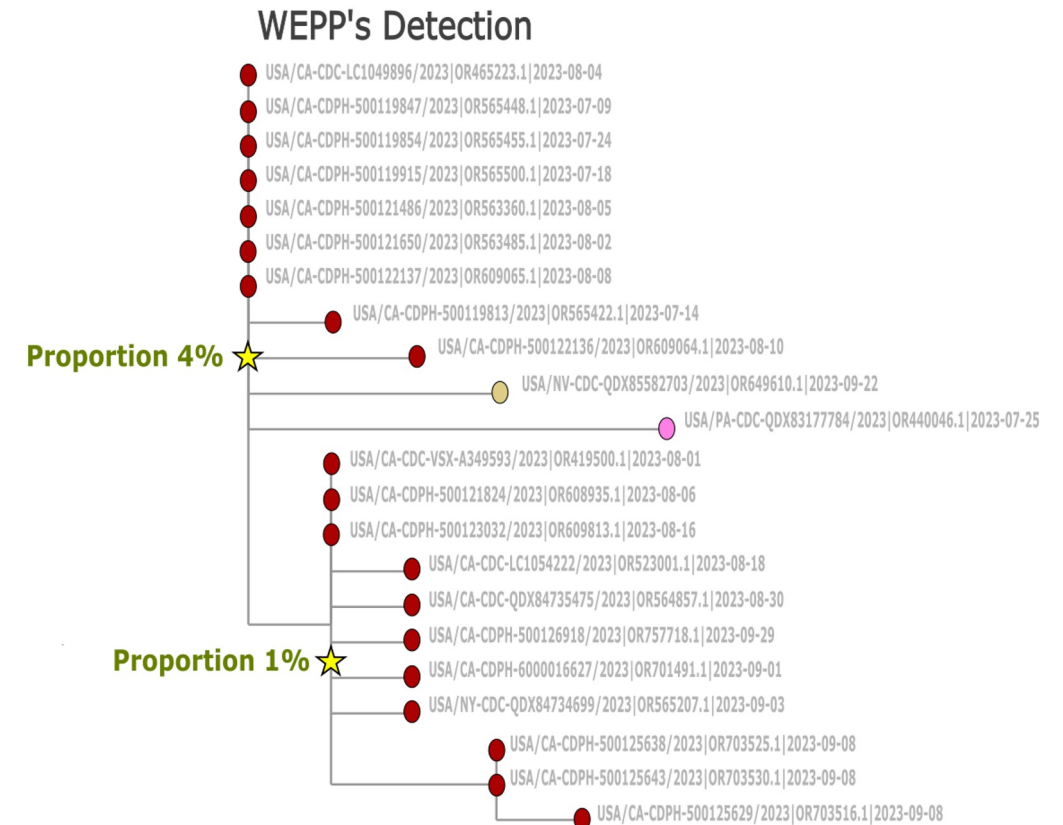
WEPP: Key Results

1. More accurate lineage abundances

2. Identifies intra-lineage clusters

3. Detects introduction of new clusters

4. Better correlated with clinical data

Lineage Proportion Correlation: Clinical Sequences vs WBE Tools

Clinical vs WEPP (Mean: 11.053)
Clinical vs Freyja (Mean: 12.820)
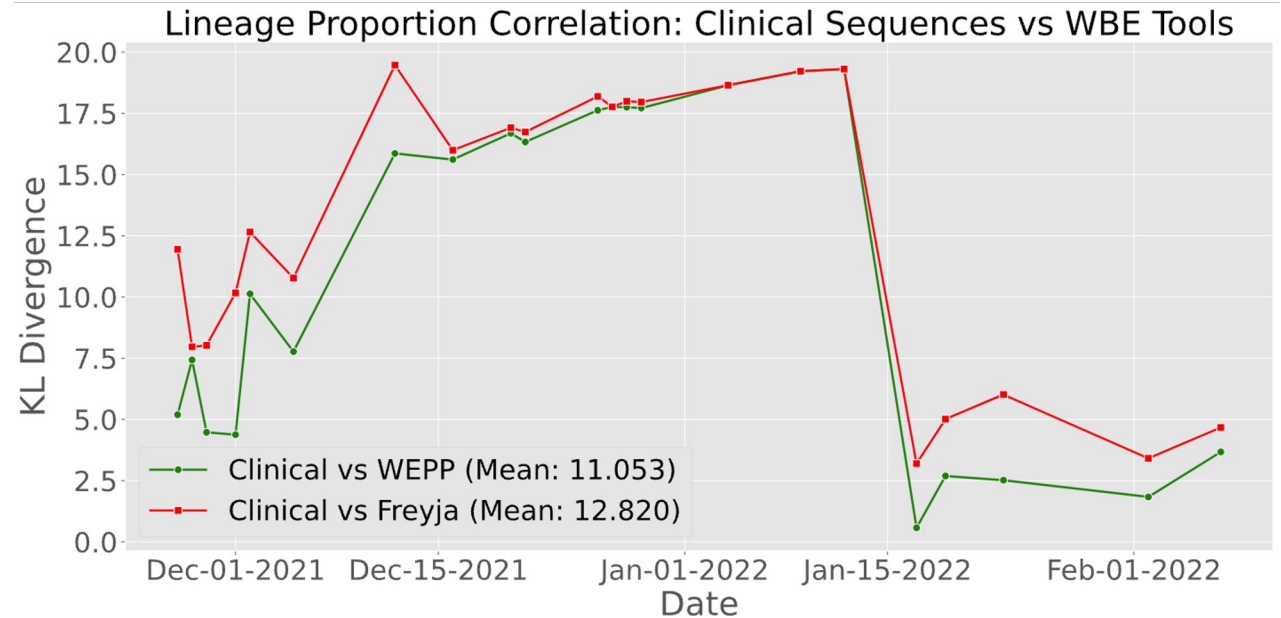
WEPP: Key Results

1. **More accurate** lineage abundances

2. Identifies **intra-lineage clusters**

3. Detects **introduction of new clusters**

4. **Better correlated** with clinical data

5. **Generalizable** to other pathogens

Lineage Abundance on ART's Simulated Dataset for RSV-A

WEPP RMSE: 0.0012

Lineage Abundance on ART's Simulated Dataset for Mpox

WEPP RMSE: 0.0324
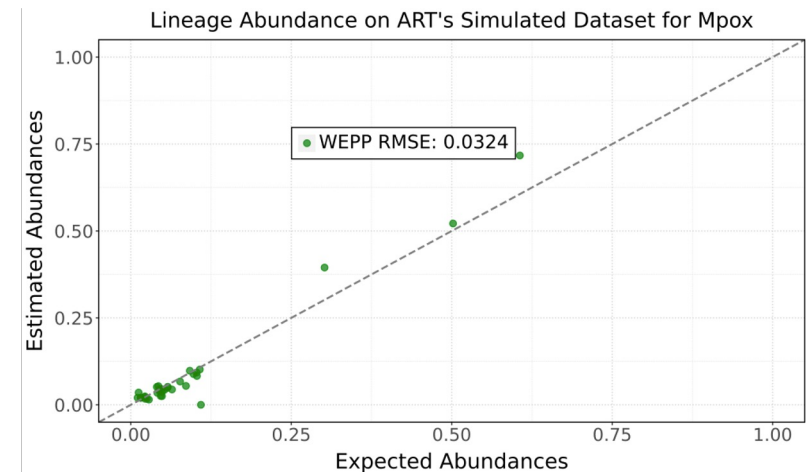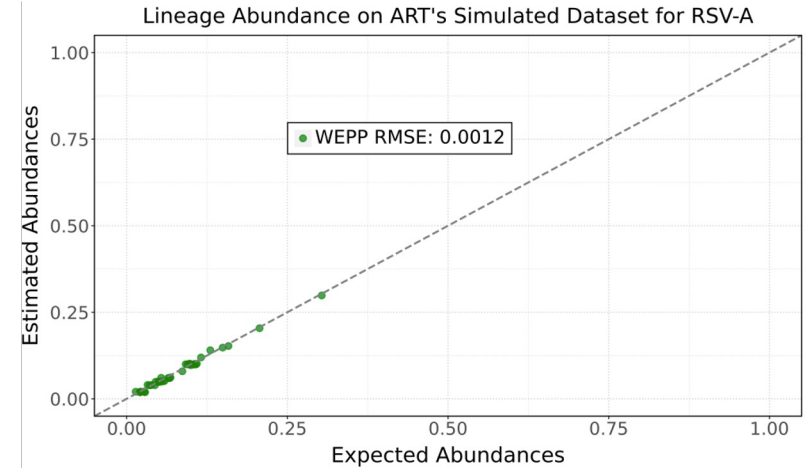
WEPP Dashboard: Interactive Analysis of Results

1. Phylogenetic View to explore the selected haplotypes.

2. Shows relative abundances of haplotypes and their corresponding lineages.

3. Inspect aligned reads mapped to the selected haplotypes.

4. Analyze unaccounted alleles in the context of reads and the selected haplotypes.
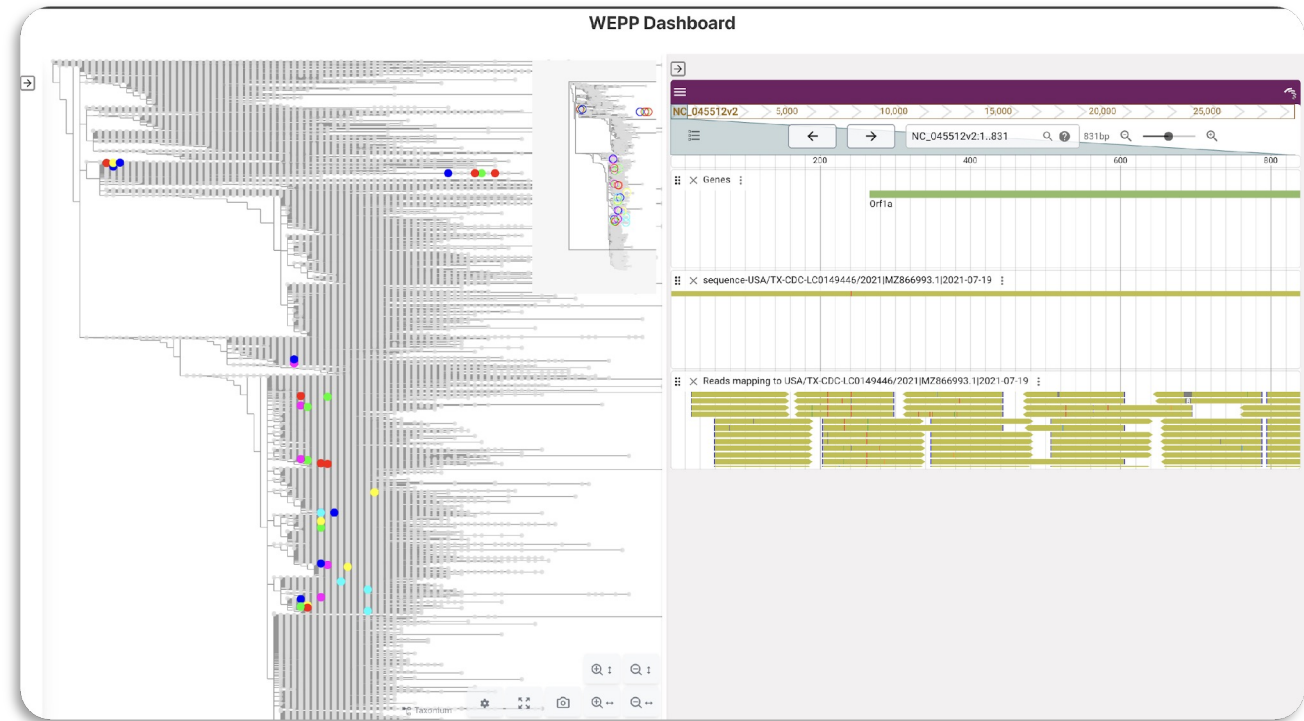
Upload Interface and Project Management

Phylogenetic View

Detected Haplotypes and Unaccounted Alleles

Haplotype Uncertainty

WEPP Dashboard

Zoom to haplotype

Display Uncertain Haplotypes

Zoomed region showing **uncertain haplotypes** associated with a **detected haplotype**

Name ⌄   node_3411

1 result 🔍 ●
Haplotype Proportion: 21.584% Haplotype Lineage: A.D.3

Uncertain Haplotypes   ↑
node_3412

Haplotype information

Haplotypes and Mapped Reads

Analyzing Unaccounted Alleles
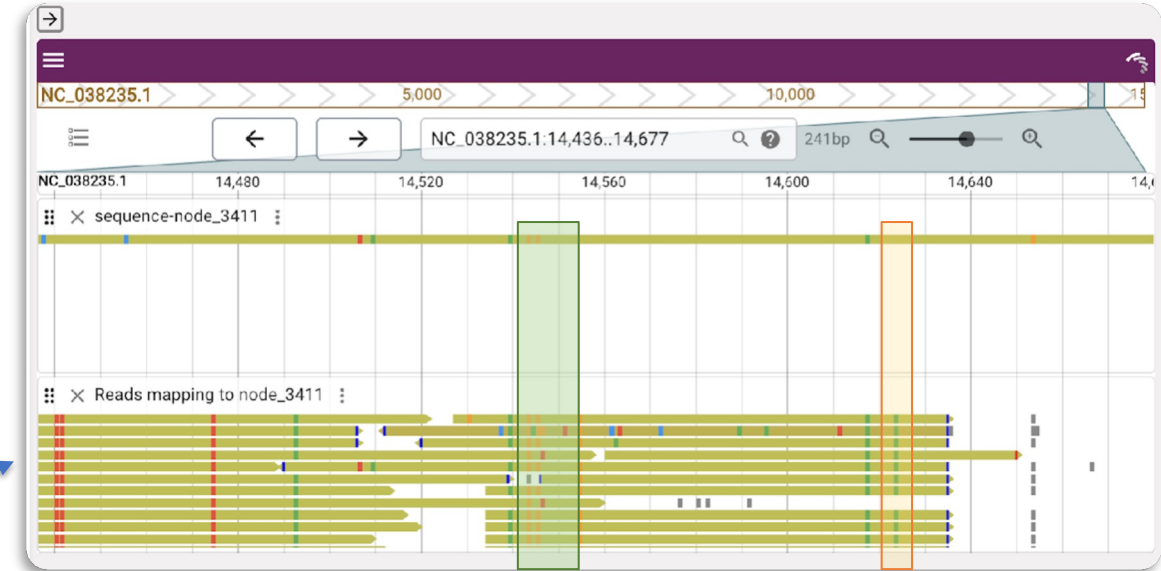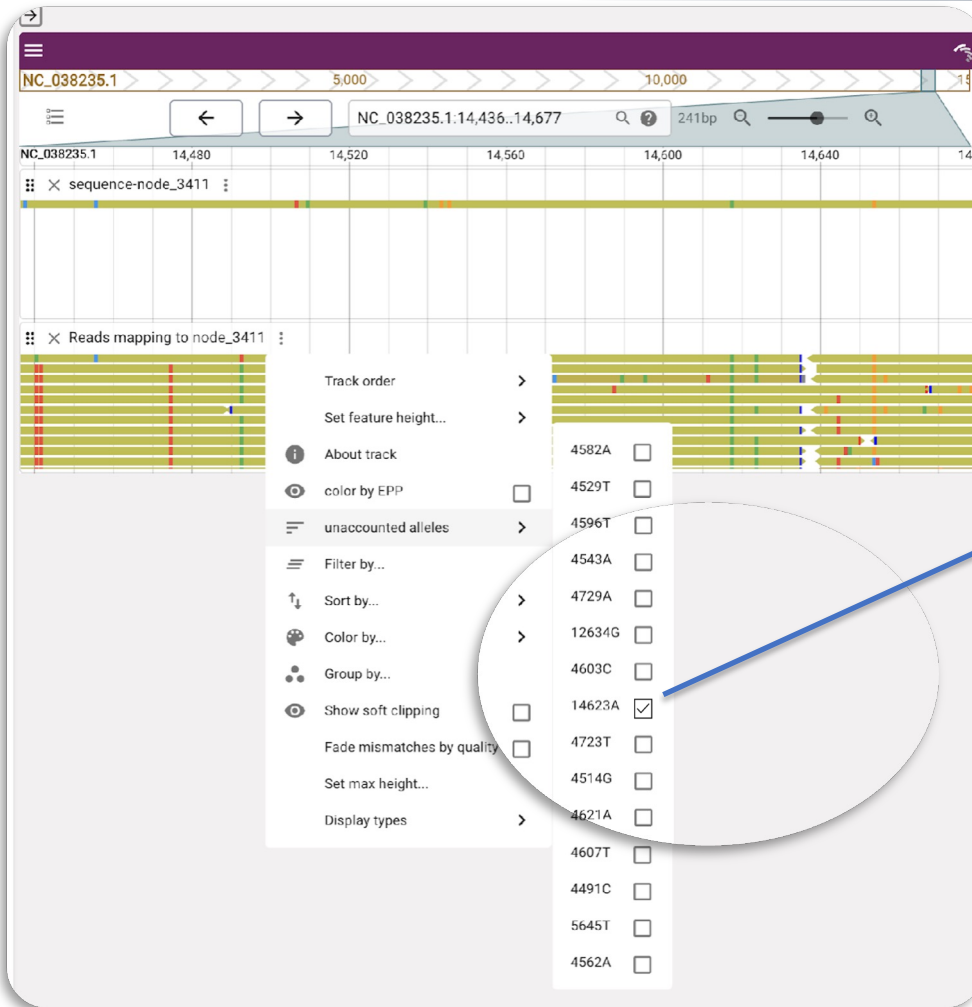
Accounted Alleles

Unaccounted Allele

Haplotype Details

Acknowledgements

Advisor

Yatish Turakhia

Collaborators

Manu Bhat

Sumit Walia

Angie Hinrichs

Joshua Levy

David Schaeper

Kyle Smith

Karthik Gangavarapu

Russ Corbett