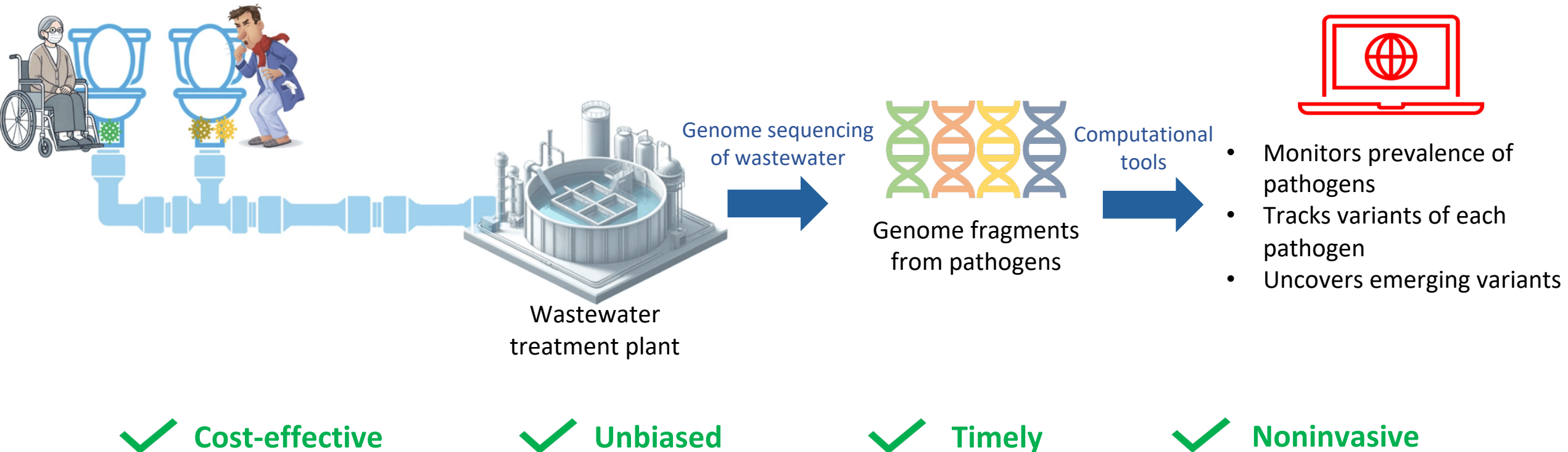


WEPP: Phylogenetic Placement Achieves Near-Haplotype Resolution in Wastewater-Based Epidemiology

Speakers: Pranav Gangwar and Pratik Katte

Wastewater-Based Epidemiology (WBE)

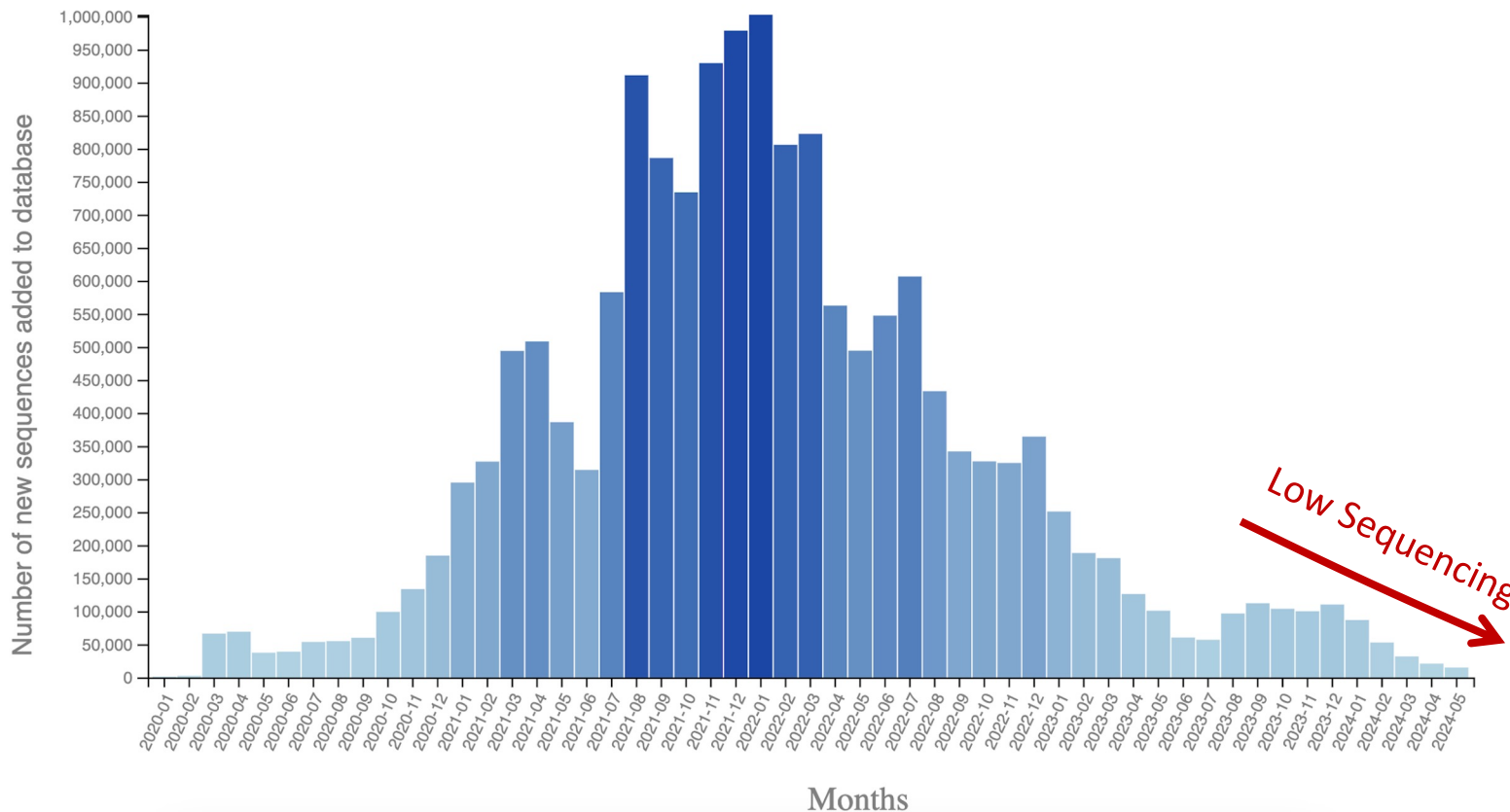


WBE Challenges

- Sequencing reads from multiple genomes with low genetic diversity
- Short reads with uneven coverage and sequencing errors
- Existing tools focus on estimating lineage abundances

Reduced WBE effectiveness with Declining Clinical Sequencing

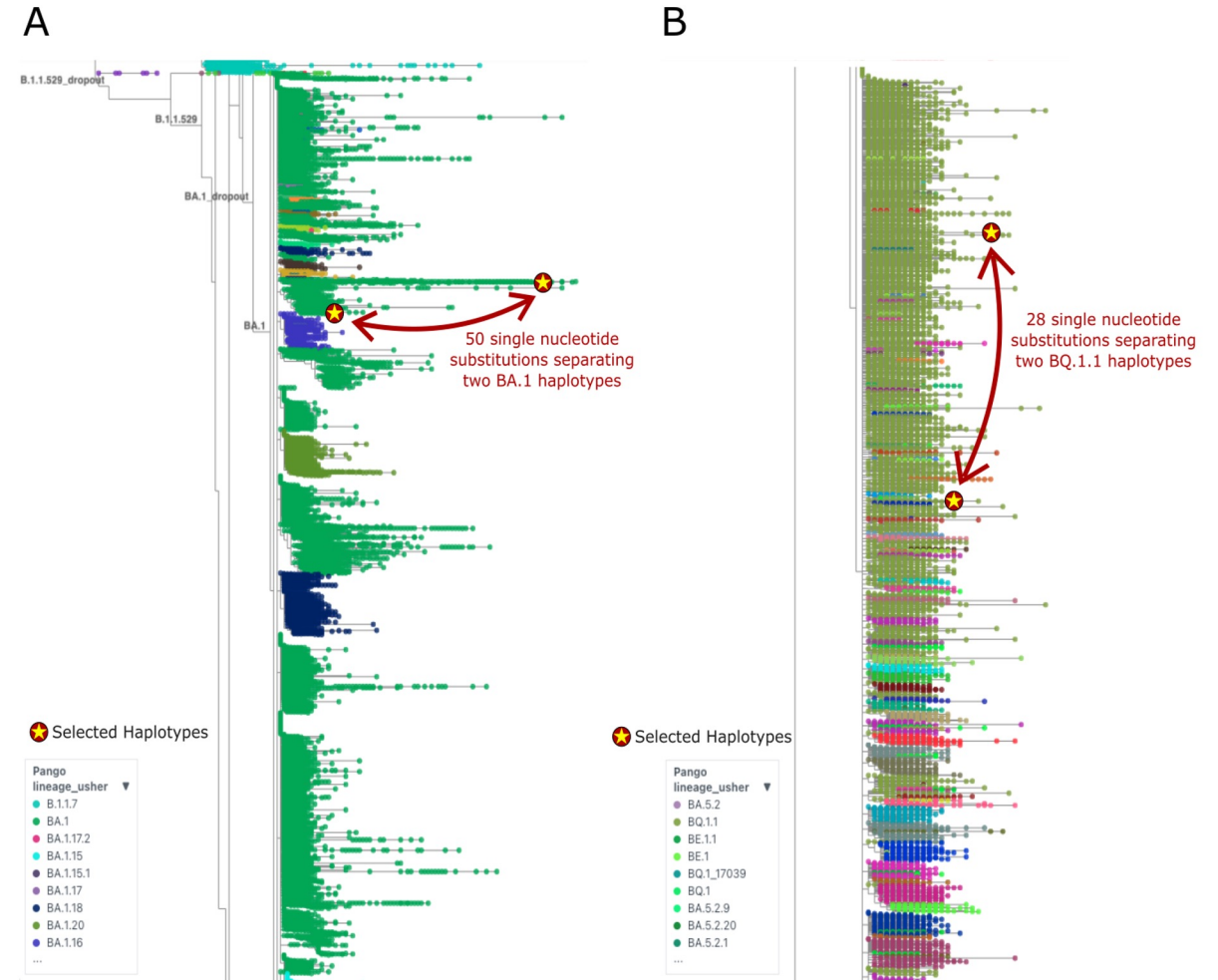
SARS-CoV-2 Clinical Sequencing



- 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

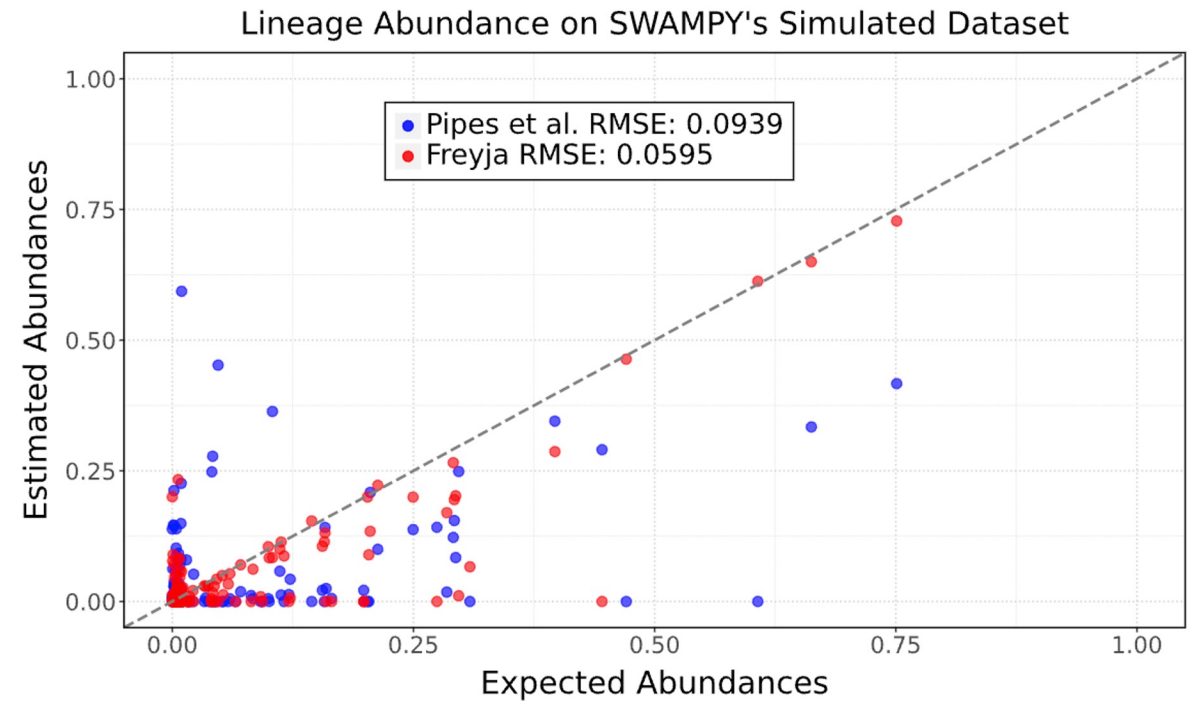
Limitations of Existing WBE Tools

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



Limitations of Existing WBE Tools

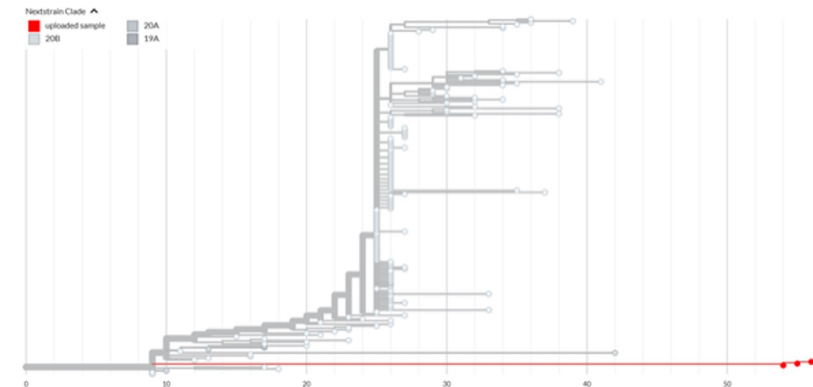
1. **Resolution:** Cannot disambiguate intra-lineage diversity
2. **Accuracy:** Lineage abundance estimates get inaccurate, especially for large lineages



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, undesigned lineages

Omicron introduction



thomaspeacock commented on Nov 23, 2021 • edited by chrisruis

New proposed lineage
By Tom Peacock

Description
Sub-lineage of: B.1.1
Earliest Sequence: 2021-11-11
Latest Sequence: 2021-11-13

Countries circulating: Botswana (3 genomes), Hong Kong ex S. Africa (1 genome, partial)

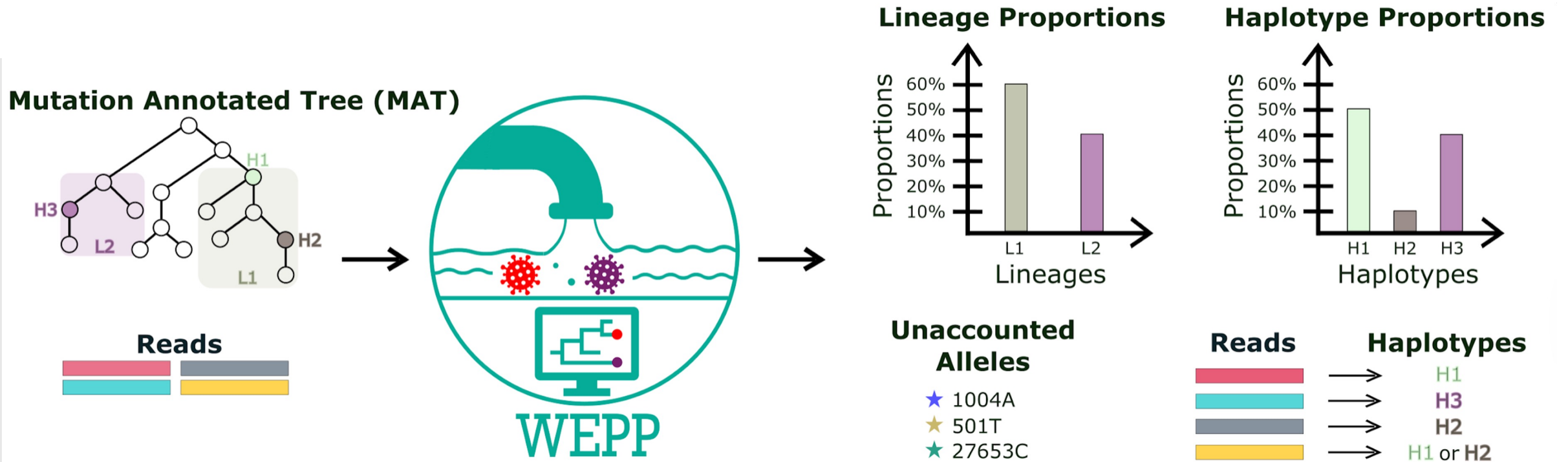
Description:
Conserved Spike mutations - A67V, Δ69-70, T95I, G142D/Δ143-145, Δ211/L212I, ins214EPE, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F

Conserved non-Spike mutations - NSP3 - K38R, V1069I, Δ1265/L1266I, A1892T; NSP4 - T492I; NSP5 - P132H; NSP6 - Δ105-107, A189V; NSP12 - P323L; NSP14 - I42V; E - T9I; M - D3G, Q19E, A63T; N - P13L, Δ31-33, R203K, G204R

Currently only 4 sequences so would recommend monitoring for now. Export to Asia implies this might be more widespread than sequences alone would imply. Also the extremely long branch length and incredibly high amount of spike mutations suggest this could be of real concern (predicted escape from most known monoclonal antibodies)

Genomes:
EPI_ISL_6590608 (partial RBD Sanger sequencing from Hong Kong)
EPI_ISL_6640916
EPI_ISL_6640919
EPI_ISL_6640917

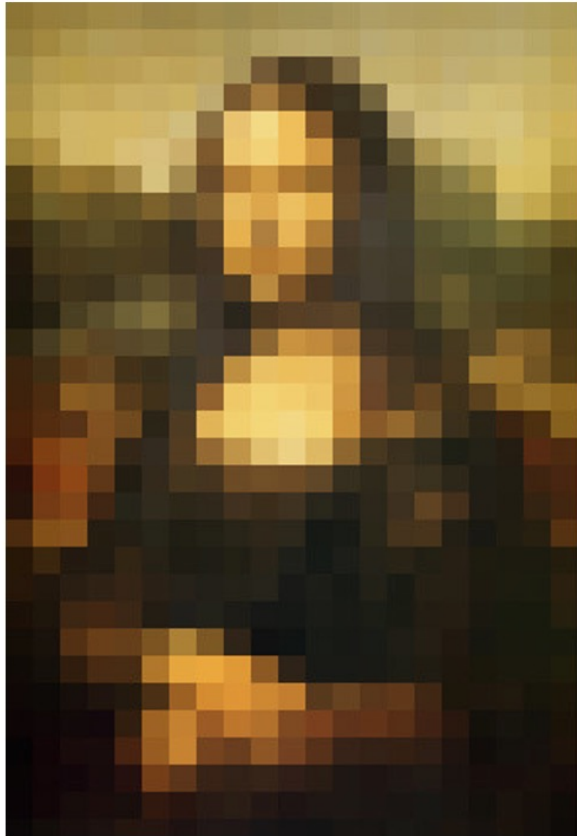
WEPP Overview



Wastewater-Based Epidemiology using Phylogenetic Placements

WEPP's Resolution Advantage

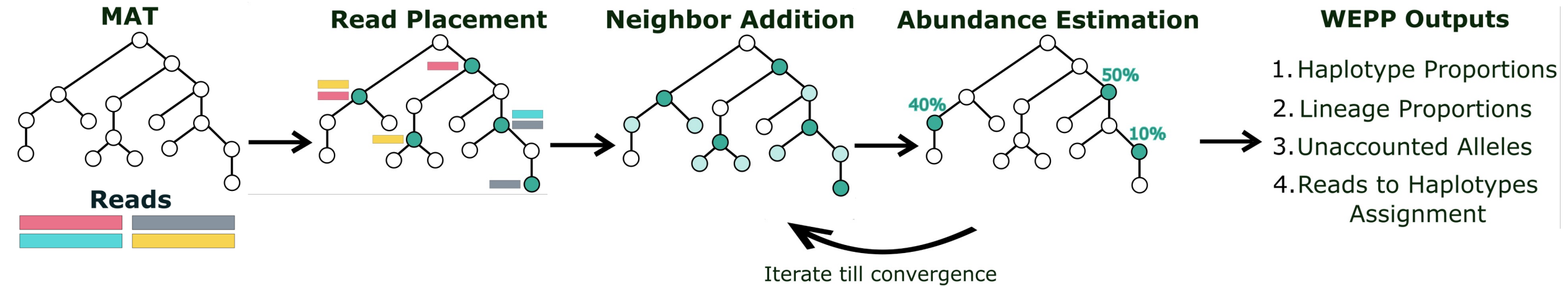
Previous WBE analysis



WBE analysis with WEPP

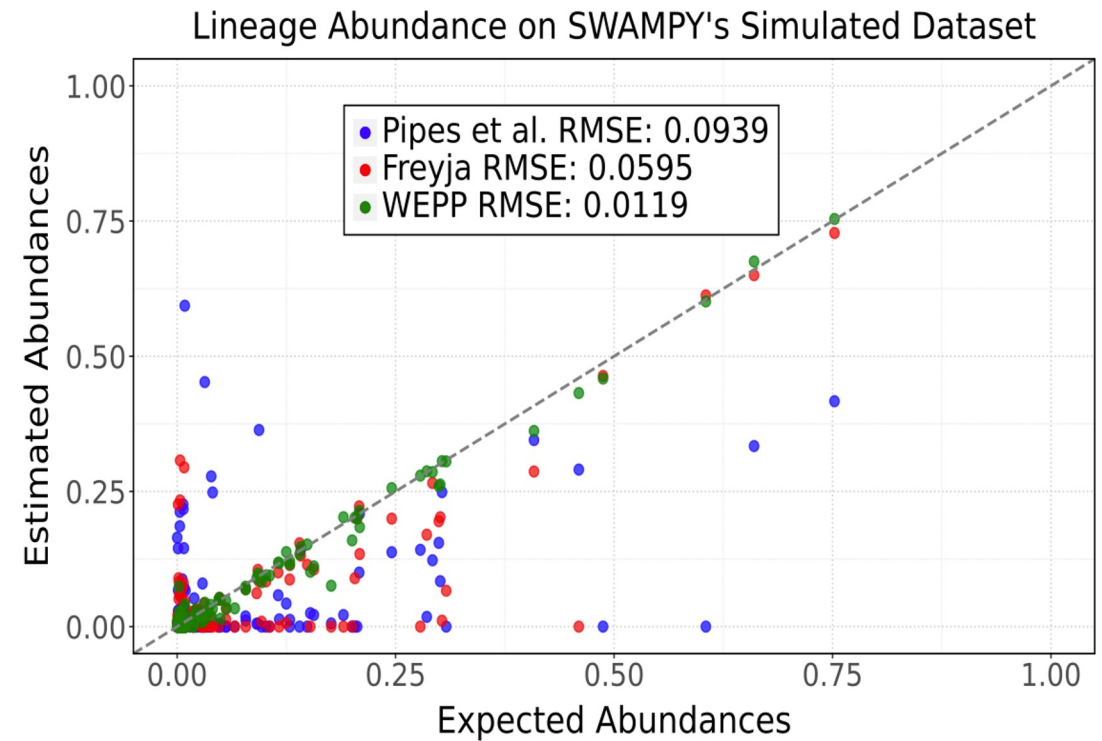


WEPP: Algorithm Overview



WEPP: Key Results

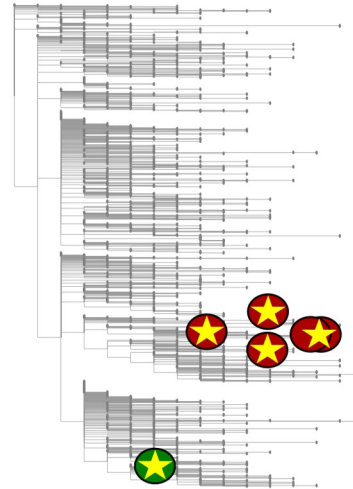
1. More accurate lineage abundances



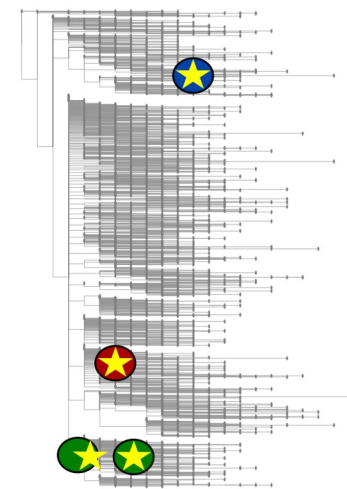
WEPP: Key Results

1. More accurate lineage abundances
2. Identifies intra-lineage clusters

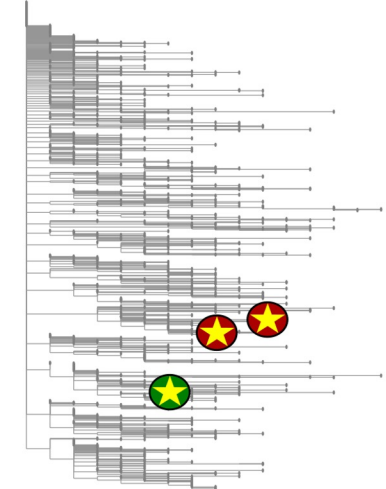
EG.5.1.4 (21%)



XBB.1.16.6 (14%)



XBB.1.5.10 (10%)

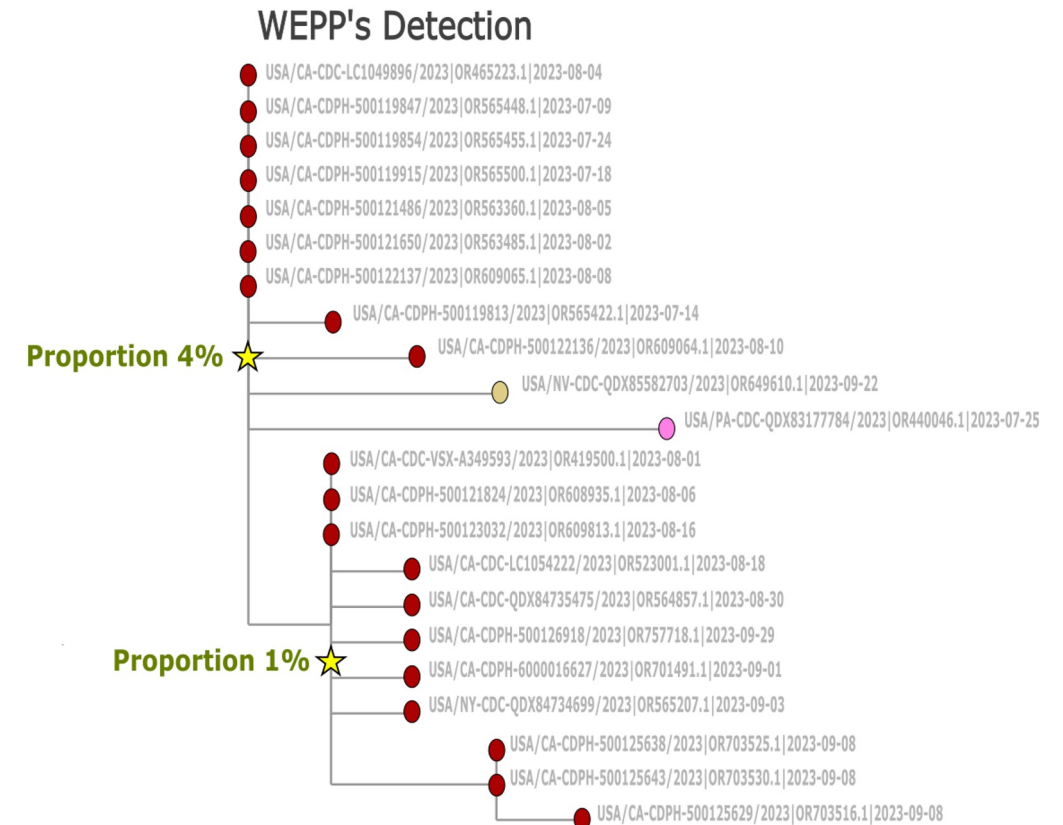


- Cluster - 1
- Cluster - 2
- Cluster - 3
- ★ WEPP Haplotypes

WEPP: Key Results

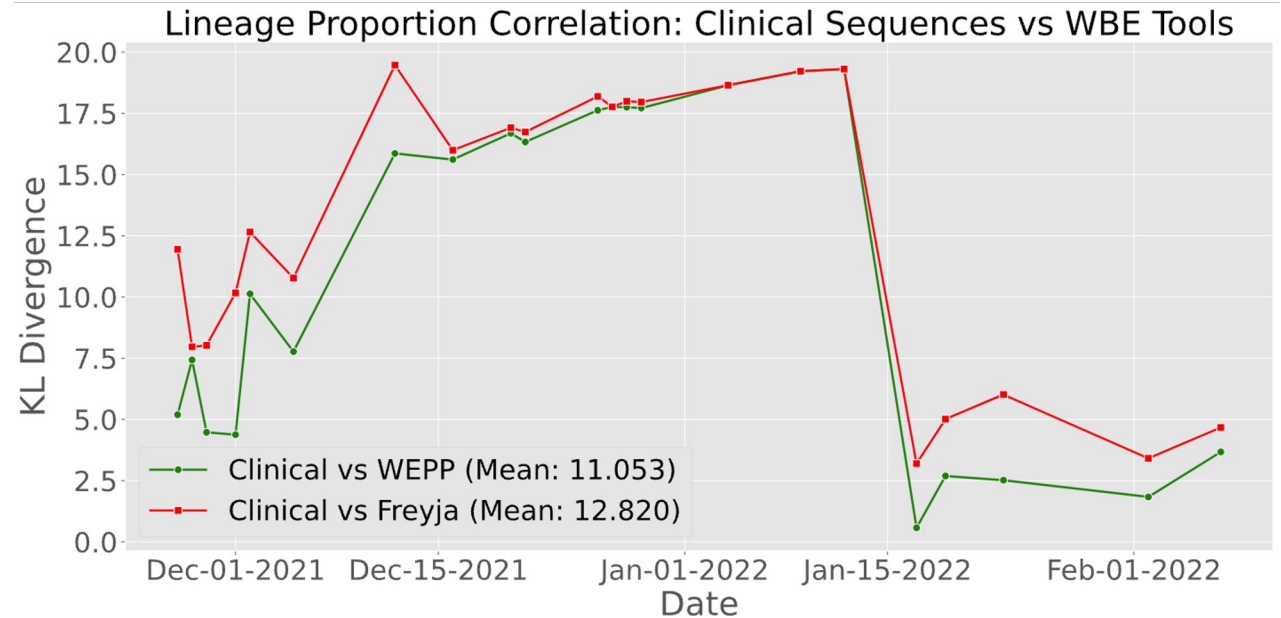
1. More accurate lineage abundances
2. Identifies intra-lineage clusters
3. Detects introduction of new clusters

CA -> Delaware introduction



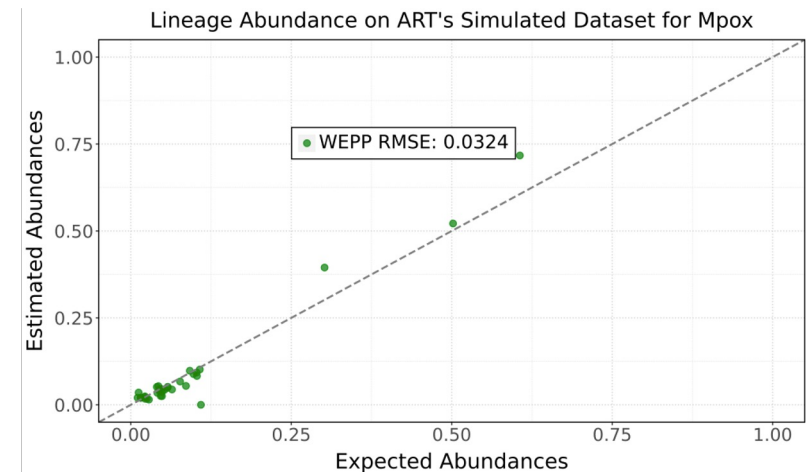
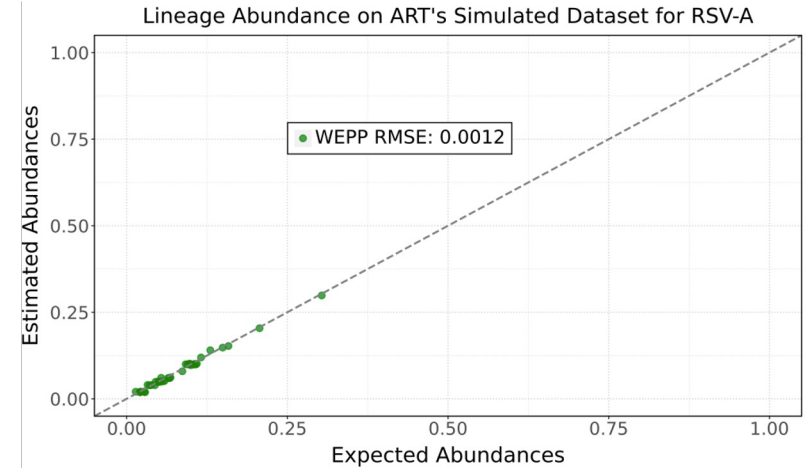
WEPP: Key Results

1. More accurate lineage abundances
2. Identifies intra-lineage clusters
3. Detects introduction of new clusters
4. Better correlated with clinical data



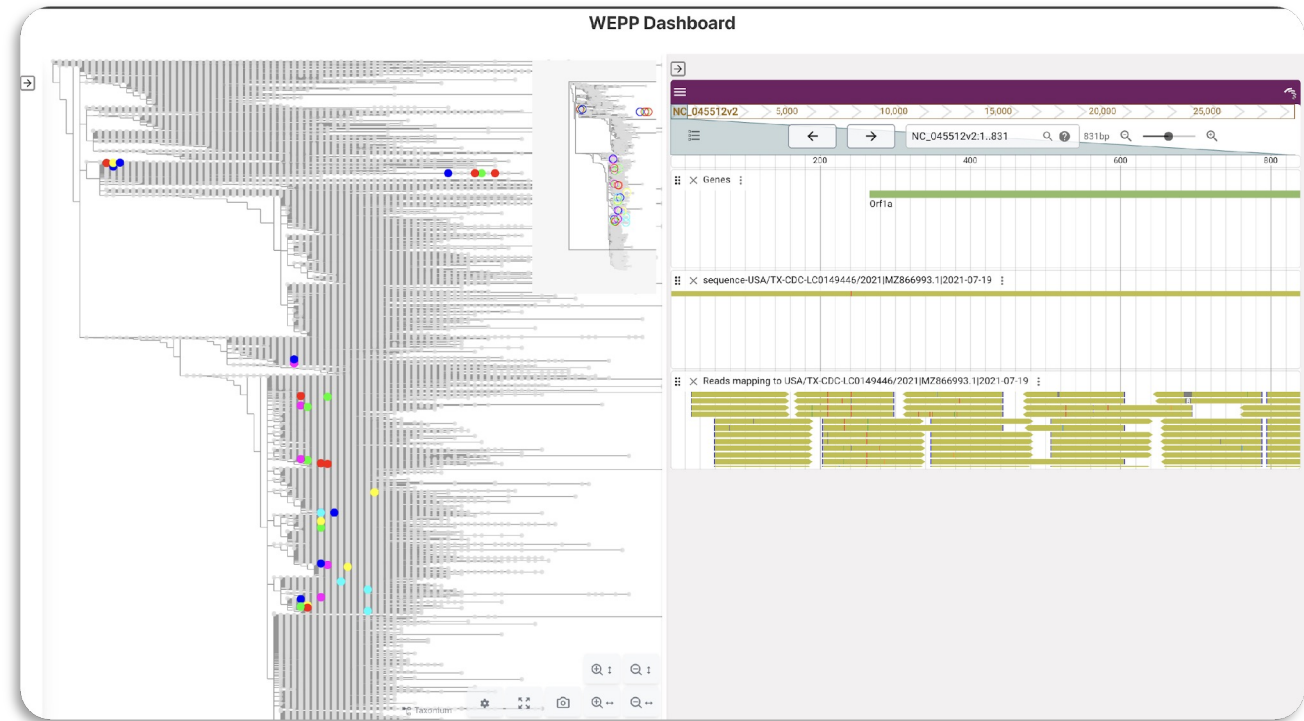
WEPP: Key Results

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



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

The screenshot displays the WEPP Dashboard interface. At the top, the title "WEPP Dashboard" is centered. Below it, a descriptive paragraph states: "This dashboard provides insights into wastewater sequencing data by enabling interactive estimation of haplotype and lineage proportions, detection of unaccounted (novel) alleles, and read-level analysis." A note follows: "Note: Please upload your BAM and reference files with their index files for complete analysis or click on one of the results to explore." The main upload section features a "Choose Files" button, a text field showing "No file chosen", and a green "Upload" button. A blue arrow points from a green callout box labeled "Upload Bam files with reference file" to the "Upload" button. Below the upload section, the text "Loaded Taxonium file: public-2021-12-05.all.masked.pb.jsonl.gz" is displayed. A blue arrow points from a green callout box labeled "Loaded Taxonium file" to this text. A horizontal line separates the upload section from the results section, which is labeled "results". Below the line, three project buttons are shown: "RSVA_real", "SARS_COV_2_real", and "SARS_COV_2_real_1". A blue arrow points from a green callout box labeled "List of Projects that were run" to the "SARS_COV_2_real" button.

WEPP Dashboard

This dashboard provides insights into wastewater sequencing data by enabling interactive estimation of haplotype and lineage proportions, detection of unaccounted (novel) alleles, and read-level analysis.

Note: Please upload your BAM and reference files with their index files for complete analysis or click on one of the results to explore.

Choose Files No file chosen Upload

Loaded Taxonium file: public-2021-12-05.all.masked.pb.jsonl.gz

results

RSVA_real SARS_COV_2_real SARS_COV_2_real_1

List of Projects that were run

Upload Bam files with reference file

Loaded Taxonium file

Phylogenetic View

WEPP Dashboard

Phylogenetic View Panel

Displaying 8,215 sequences from INSDC

Tree type: Distance

Colour by: None

Filter

☒ Select/Deselect All Order: Descending

Detected Haplotypes Unaccounted Alleles

☒ Name node_3377
1 result
Haplotype Proportion: 23.460% Haplotype Lineage: A.D.3

☒ Name node_3411
1 result
Haplotype Proportion: 21.584% Haplotype Lineage: A.D.3

Uncertain Haplotypes

☒ Name node_887
1 result
Haplotype Proportion: 21.510% Haplotype Lineage: A.D.1

☒ Name node_886
1 result
Haplotype Proportion: 20.237% Haplotype Lineage: A.D.1

☒ Name node_33
1 result
Haplotype Proportion: 7.243% Haplotype Lineage: A.D.

Detected Haplotype annotation

The image displays a WEPP Dashboard interface for a Phylogenetic View. On the left, a sidebar contains a list of detected haplotypes and uncertain haplotypes, each with a checkbox, a name field, and a search icon. The main area shows a large phylogenetic tree with a dark blue border. A green box labeled 'Phylogenetic View Panel' points to the tree. A green box labeled 'Detected Haplotype annotation' points to specific nodes in the tree, which are highlighted with colored circles (red, blue, green, purple, cyan). The tree is a distance-based tree showing the relationships between 8,215 sequences from INSDC. The bottom right corner of the dashboard includes a 'Taxonium' logo and several icons for zooming and navigation.

Detected Haplotypes and Unaccounted Alleles

The image displays a software interface for analyzing genetic data, specifically focusing on detected haplotypes and unaccounted alleles. The interface is divided into two main panels, each with a filter and a table of results. Annotations with arrows point to specific features in the interface.

Left Panel: Detected Haplotypes

- Filter:** Select/Deselect All (checked), Order: Descending.
- Table:** Contains a list of detected haplotypes. Each entry includes a checkbox, a name dropdown, a text input, and a result summary.
- Annotations:**
 - List of detected Haplotypes:** Points to the table header.
 - Sort based on haplotype proportion:** Points to the Order dropdown.

Right Panel: Unaccounted Alleles

- Filter:** Select/Deselect All (checked), Order: Descending.
- Table:** Contains a list of unaccounted alleles. Each entry includes a checkbox, a name dropdown, a text input, and a result summary.
- Annotations:**
 - List of Unaccounted Alleles:** Points to the table header.
 - Unaccounted Allele Information:** Points to the Allele, Residue, Allele Frequency, and Depth fields.
 - List of Possible Haplotypes:** Points to the Possible Haplotypes dropdown.

Map: A map on the right side of the interface shows the geographical distribution of the data, with colored dots indicating specific locations. A label **Unaccounted Allele Information** points to a specific location on the map.

Haplotypes and Mapped Reads

WEPP Dashboard

Displaying 8,215 sequences from INSDC

Tree type: Distance

Colour by: None

Filter

☒ Select/Deselect All Order: Descending

Detected Haplotypes **Unaccounted Alleles**

☒ Name

node_3377

1 result

Haplotype Proportion: 23.460%

Haplotype Lineage: A.D.3

☒ Name

node_3411

1 result

Haplotype Proportion: 21.584%

Haplotype Lineage: A.D.3

node_3411 close

Descendant sequences: 105

Mutations at this node: ☒ aa ☒ nt

nt:T5503C

On clicking

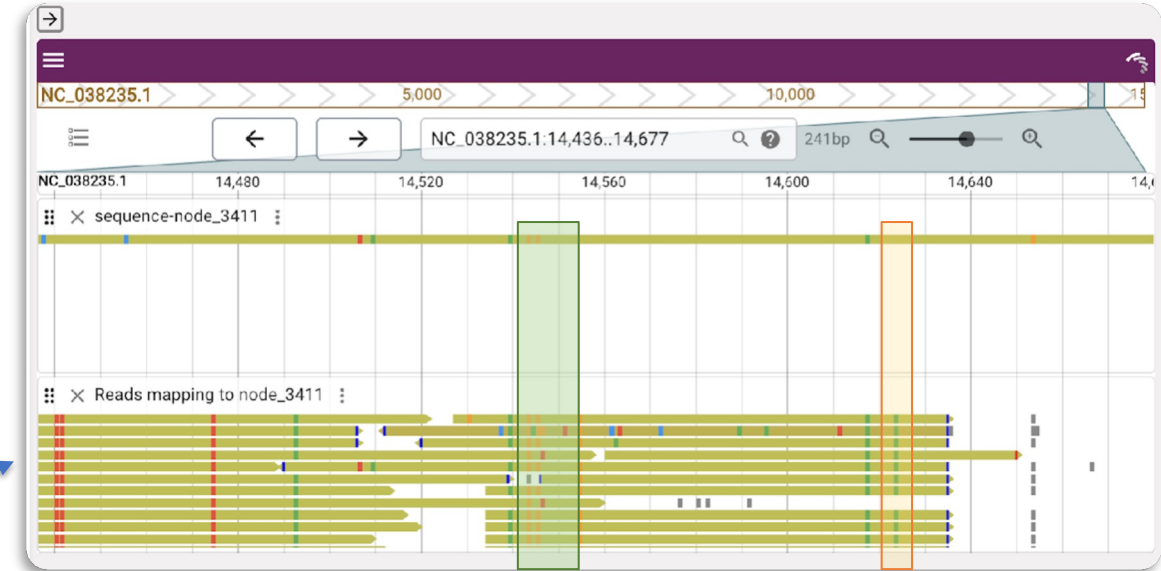
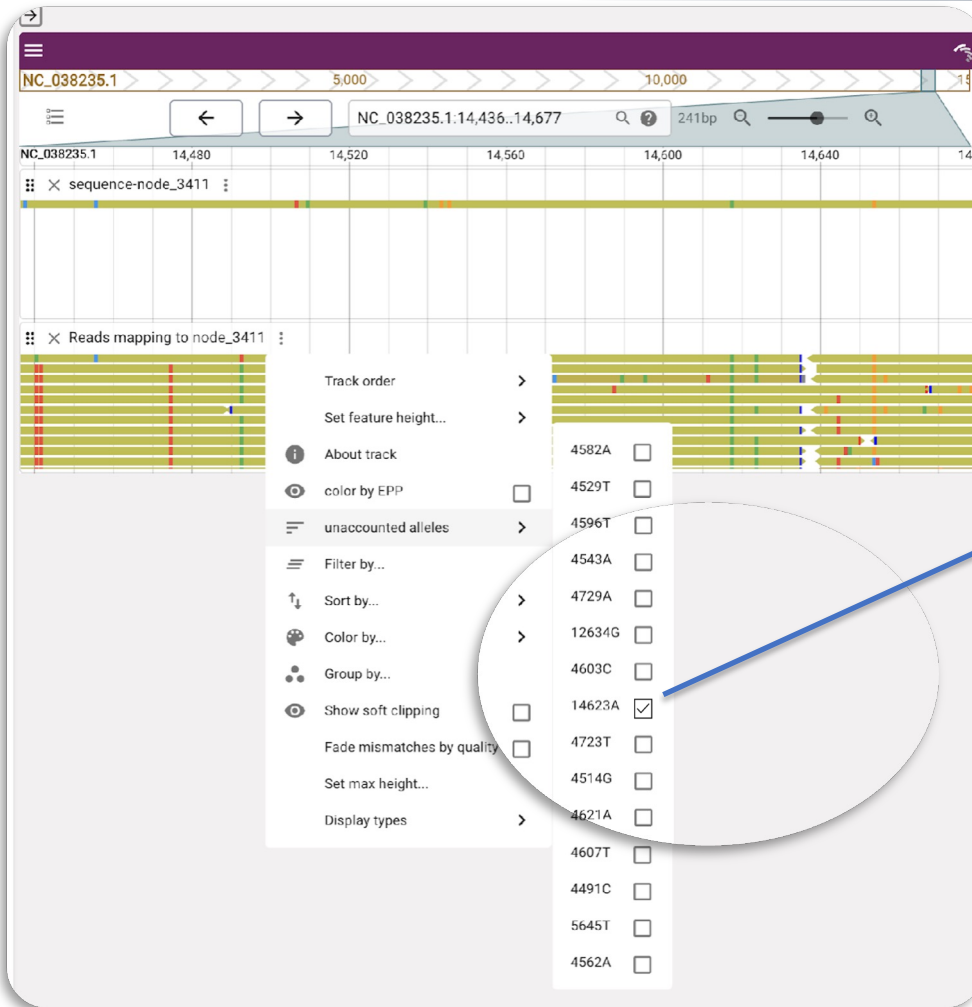
Genomic View Panel

Haplotype Aligned Sequence

Wastewater reads mapped to this specific haplotype

The screenshot displays the WEPP Dashboard interface. On the left, a sidebar contains filters and a list of detected haplotypes. The main area shows a phylogenetic tree with node_3411 highlighted. A circular callout labeled 'On clicking' points to node_3411. To the right, the 'Genomic View Panel' shows the sequence NC_038235.1. Below it, the 'Haplotype Aligned Sequence' panel shows wastewater reads mapped to node_3411. A callout labeled 'Wastewater reads mapped to this specific haplotype' points to the aligned reads.

Analyzing Unaccounted Alleles



Accounted Alleles

Unaccounted Allele

Haplotype Details

Feature details

NODE_3411 - MATCH

Core details

Position

NC_038235.1:1..15,222 (+)

Name

node_3411

Length

15,222

Type

match

Attributes

score

60

tags.MD

CopyShow more
29A79T26A6T49T11T8C17G29G53A27C43T11T5T3C94T2T11T5T52A8C26G2A19C10A0T35G2G2A8C11A11C11C52A18C8C17A8A...

seq

CopyShow more
ACGCGAAAAAATGCGTACAACAACTTGCCTAAACCAAAAAATGGGGCAAAATAAGAATTGTAAGTACCACCTAAATTAACCTCCCTTGTTAGAGAT...

SHOW FEATURE SEQUENCE ?

UNCERTAIN HAPLOTYPES

node_3412

POSSIBLE UNACCOUNTED ALLELES

Allele	Residue	Allele Frequency	Depth
4582A	0.632	0.999	29250
4529T	0.811	1.000	31062
4596T	0.445	0.999	30756
4543A	0.334	1.000	26850
4729A	0.340	1.000	32393
12634G	0.504	1.000	119310
4603C	0.445	1.000	30253
14623A	0.474	0.571	38963
4723T	0.443	0.999	29680
4514G	0.812	1.000	31166

WEPP Dashboard

NC_038235.1

5,000

10,000

NC_038235.1:1,191..2,021

831bp

sequence-node_3411

Reads mapping to node_3411

Acknowledgements

Advisor



Yatish Turakhia

Collaborators

Manu Bhat

Sumit Walia

Angie Hinrichs

Joshua Levy

David Schaeper

Kyle Smith

Karthik Gangavarapu

Russ Corbett

