

RIVET: A visual interactive browser for tracking and curating SARS-CoV-2 recombinants

Kyle Smith¹, Cheng Ye¹, Yatish Turakhia¹
¹University of California, San Diego

UC San Diego
JACOBS SCHOOL OF ENGINEERING
Electrical and Computer Engineering

TURAKHIA LAB

Motivation

SARS-CoV-2 Recombination

- Recombination is a known contributor to genetic novelty within SARS-CoV-2
- Recombination can lead to drastic advances in viral fitness
- The current dominant SARS-CoV-2 variant is a sublineage of the identified XBB recombinant
- XBB sublineages, such as XBB.1.5 and XBB.1.16 have been shown to be some of the most transmissible SARS-CoV-2 variants^{12,13}

Current Approach for Recombinant Lineage Designation

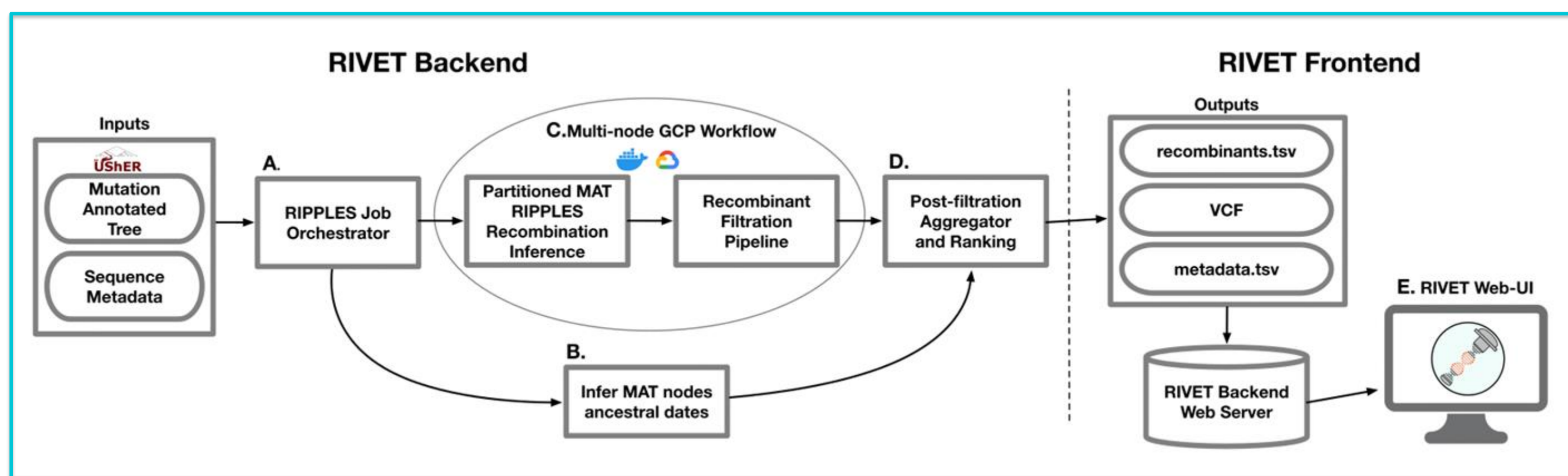
1. New sequences added to online databases are manually inspected for recombination events
2. Putative recombinants are proposed via Pangolin-GitHub Issues (LINK)
3. Evidence of recombination and growth accumulated over weeks to months
4. Decision to designate a unique lineage is reached

Problem: Confidently identifying recombinants from thousands of new sequences is challenging and often suffers from weeks to months of delay, becoming a bottleneck for online genomic epidemiology.

Our Approach

RIVET: SARS-CoV-2 Recombination Viewer and Tracker

- RIVET is a software pipeline and visual web application designed to greatly accelerate the process of identifying and tracking recombinants
- RIVET's architecture has two components:
 - Backend to perform recombination inference using a fast implementation of the RIPPLES¹⁰ algorithm
 - Frontend web-browser interface for interactive visualization and analysis of recombinants



RIVET Backend

- Multithreaded recombinant inference with sensitive parameters
- Can be massively parallelized across many servers on Google Cloud
- Quality Control/Filtration Pipeline
- Recombinant date of origin inference using Chronomutal
- Recombinant Ranking

RIVET Frontend

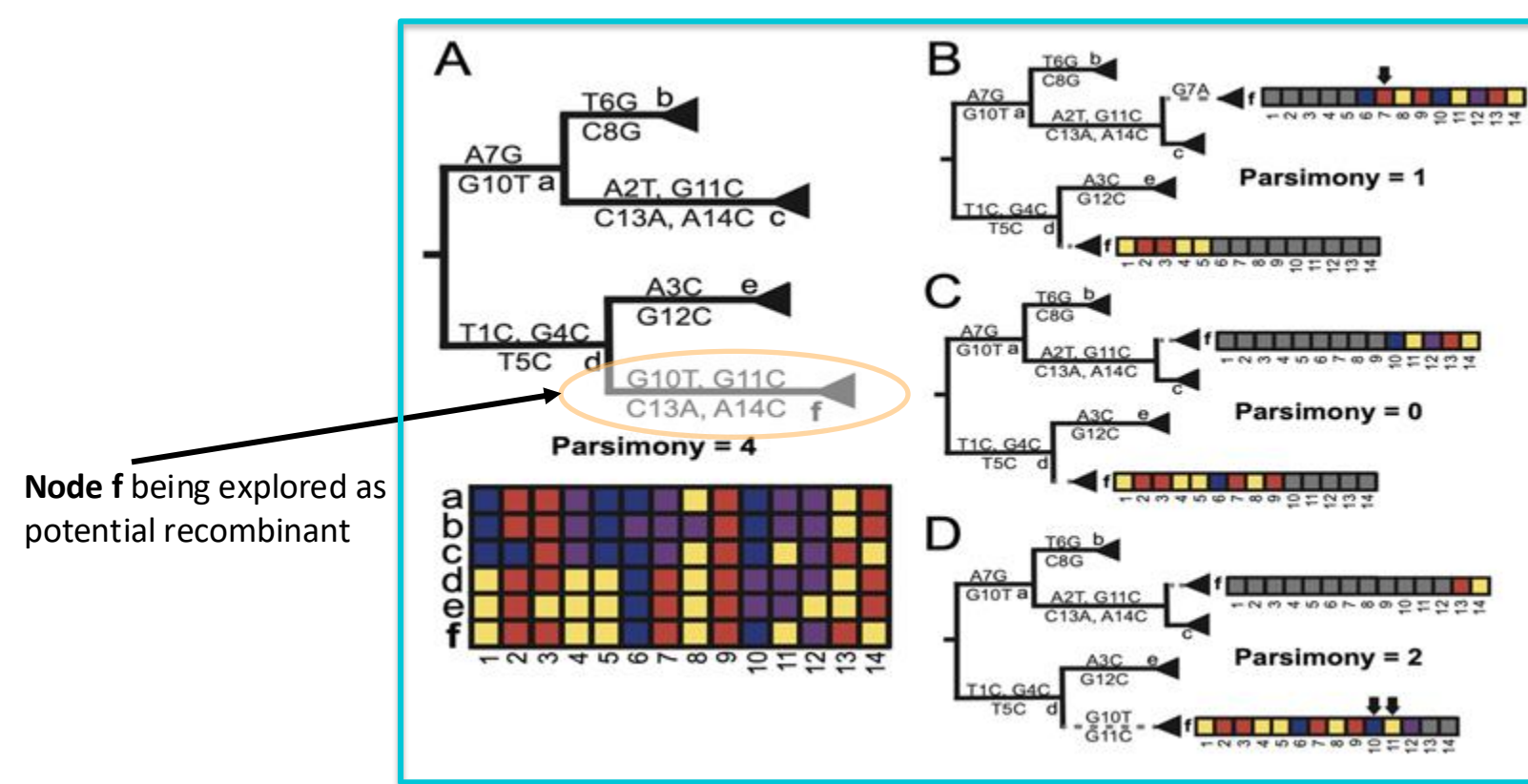
- Visualization of recombinant and parental sequences
- Integration with USHER, Taxonium, and Nextstrain/Auspice
- Integration with sequence metadata for built-in queries
- Aggregate all necessary evidence for recombinant Pangolin-lineage proposal

RIVET Backend

RIPPLES Algorithm

Key innovations:

- a. Restrict search space to long branches (~90,000) and mutated sites only
- a. Restrict to potential parents that improve partial-placement parsimony
- a. Exhaustively break the possible recombinant sequence into shorter segments, then place segments back onto phylogeny using maximum parsimony



Optimized implementation of RIPPLES Algorithm

- Achieves 1-2 orders of magnitude in speedup vs original implementation with identical results
- Key performance optimizations:
 - a. Amortizing parsimony improvement computations between breakpoint intervals
 - a. Improved memory locality
 - a. Fine-grained parallelism using available SSE-based vector instructions

RIVET backend runtime and cost estimates

Date of MAT	Number of sequences in the MAT	Number of unique recombinants discovered	Ripples-fast runtime	Total runtime (including QC, ranking etc.)	Estimated compute cost	
October 31, 2022	6,427,951	1,413	421	20m 22s	45m 25s	\$2.06
November 30, 2022	6,497,825	1,452	441	20m 19s	49m 17s	\$2.23
December 31, 2022	6,612,971	1,473	455	17m 01s	43m 10s	\$1.97
January 31, 2023	6,716,605	1,470	453	23m 35s	51m 10s	\$2.34
July 2, 2023	15,360,149	3,665	847	94m 43s	135m 47s	\$3.68

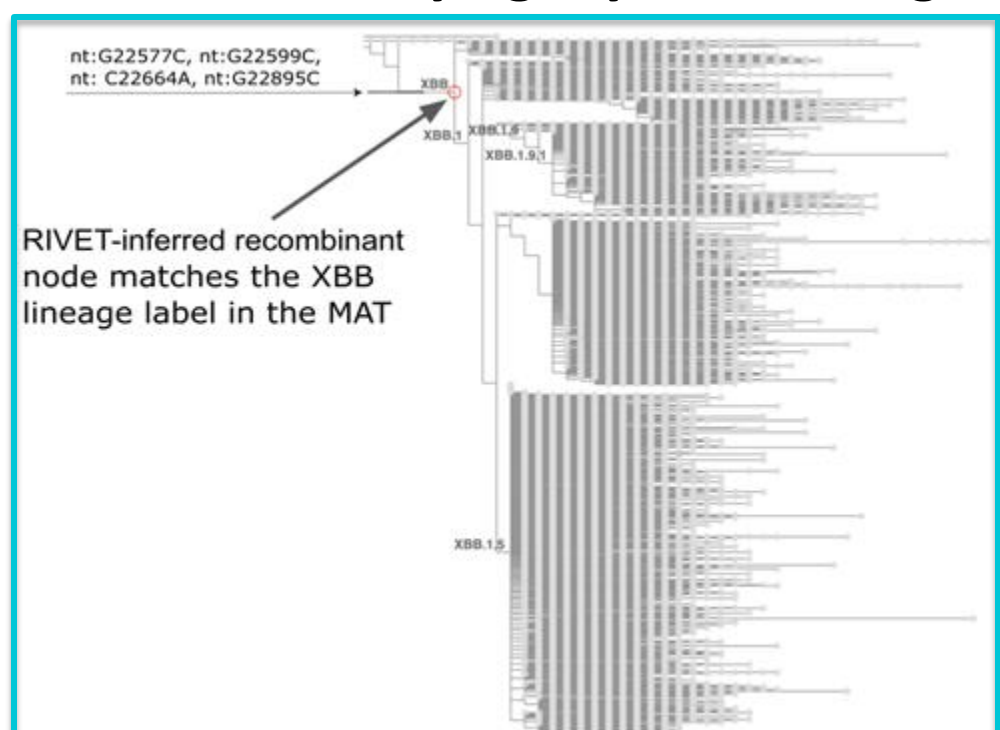
Ranking Putative Recombinants by Growth Metric

1. RIVET backend infers dates of origin for all RIPPLES-inferred recombinant nodes in the MAT using sequence metadata and Chronomutal⁸
1. To support tracking emerging recombinants of epidemiological interest, all detected recombinants are ranked based on recency of the recombinant ancestor and relative growth of its descendant sample set

$$G(R) = 2^{-m(R)} \cdot \sum_{s \in S} 2^{-m(s)}$$

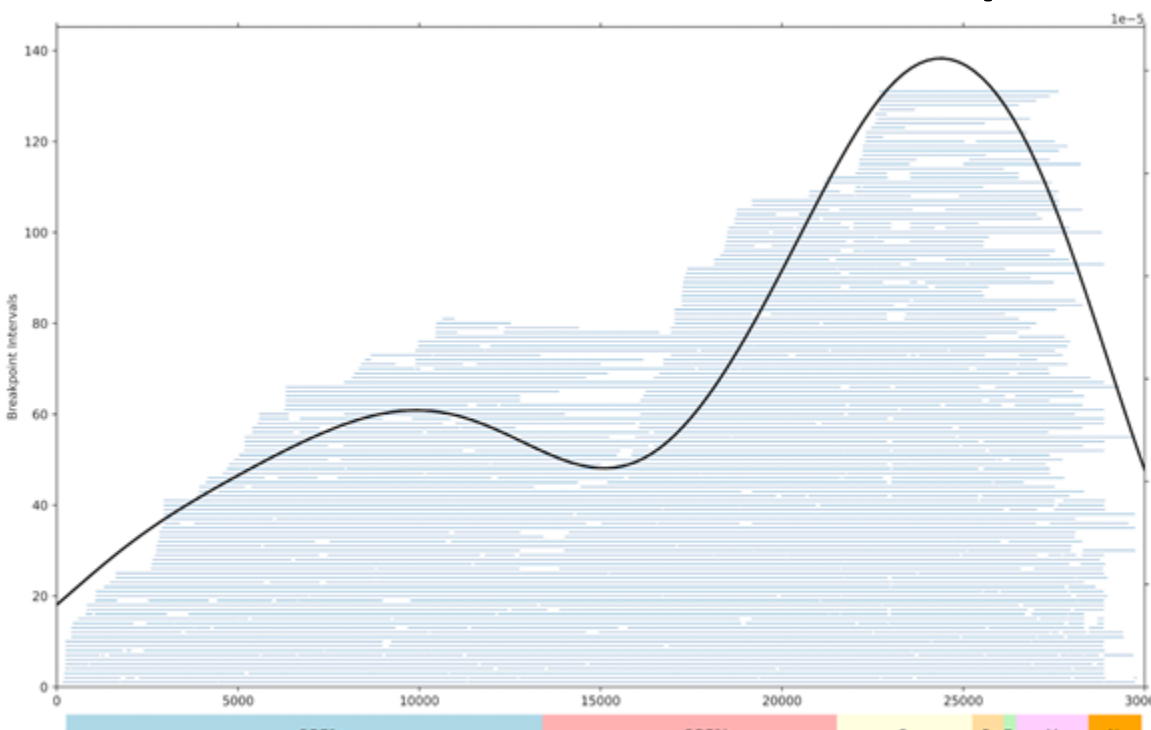
Results

SARS-CoV-2 Phylogeny XBB Lineage



- RIVET inferences are largely consistent with the manually curated recombinants

Distribution of RIVET-inferred breakpoints



- The breakpoint interval (blue bars) distribution of RIVET-inferred high-quality recombinants that pass our filtration criteria showed increased recombination rates in the 3' portion of SARS-CoV-2 genome

RIVET Web Interface

Visualization and Analysis of RIVET-Inferred SARS-CoV-2 Recombinants

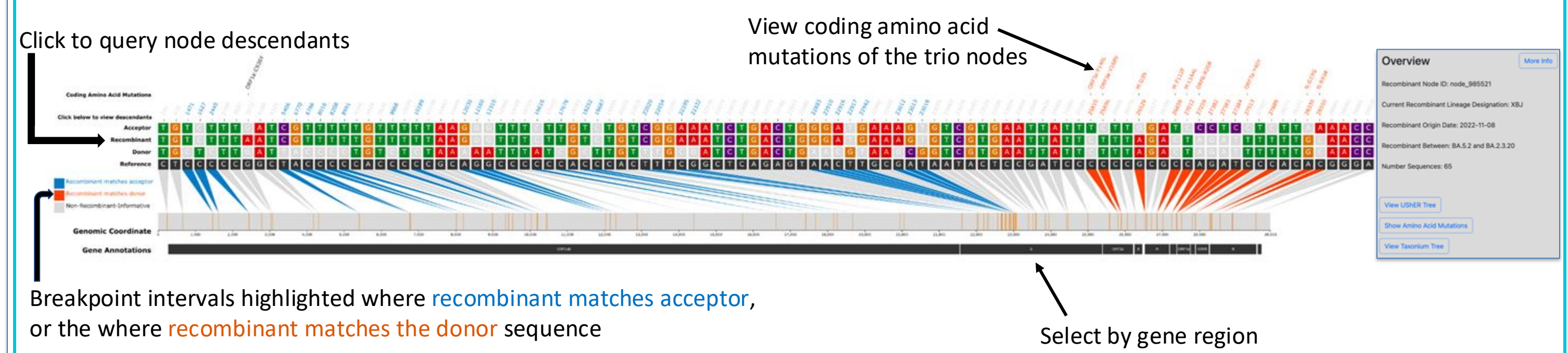
- Typical RIVET web-interface workflow analyzing previously Pango-identified XBJ recombinant

1. Select or search for recombinant ancestry in results table

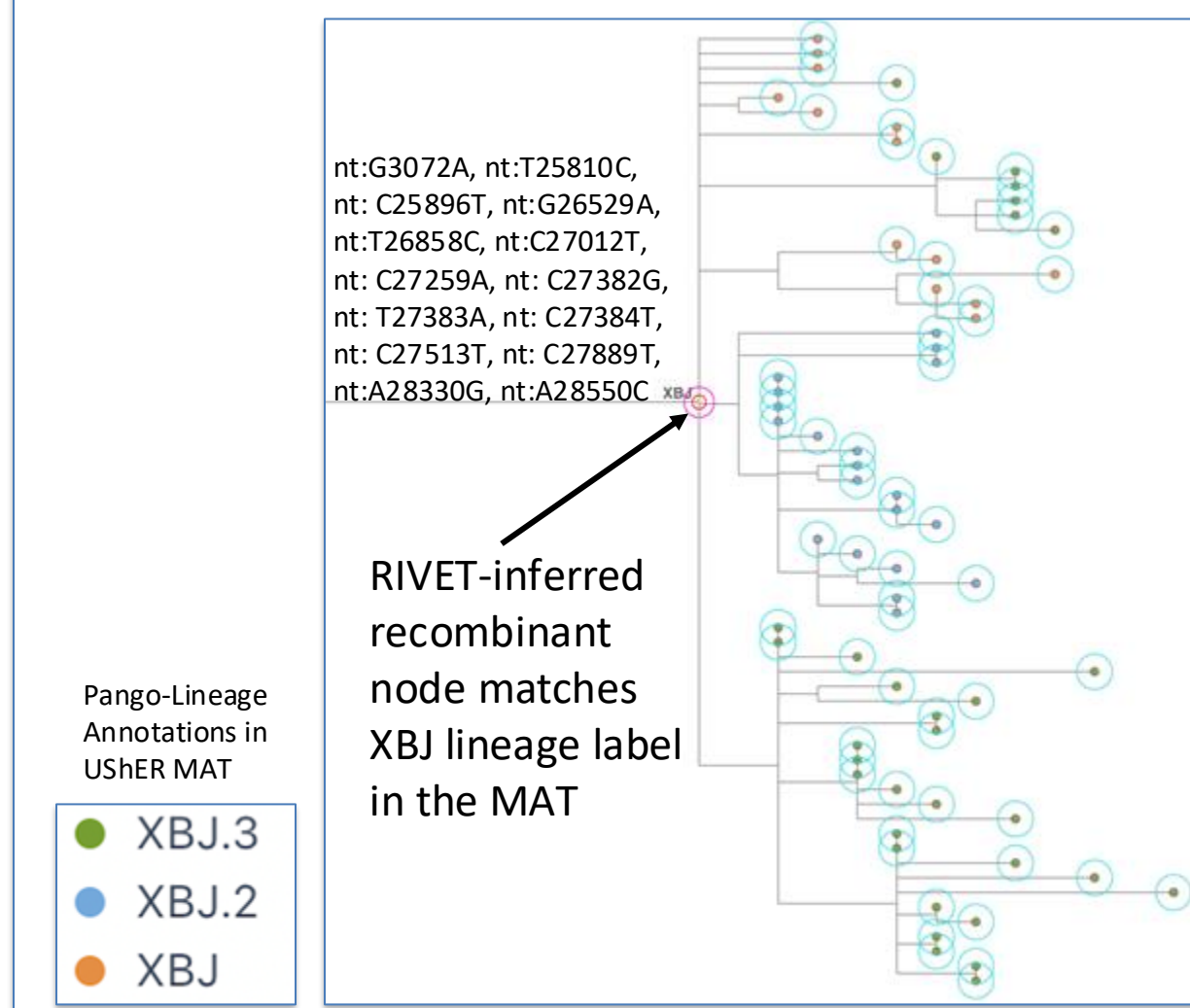
RIVET detects thousands of unique recombinants

Select recombinant row to view SNVs

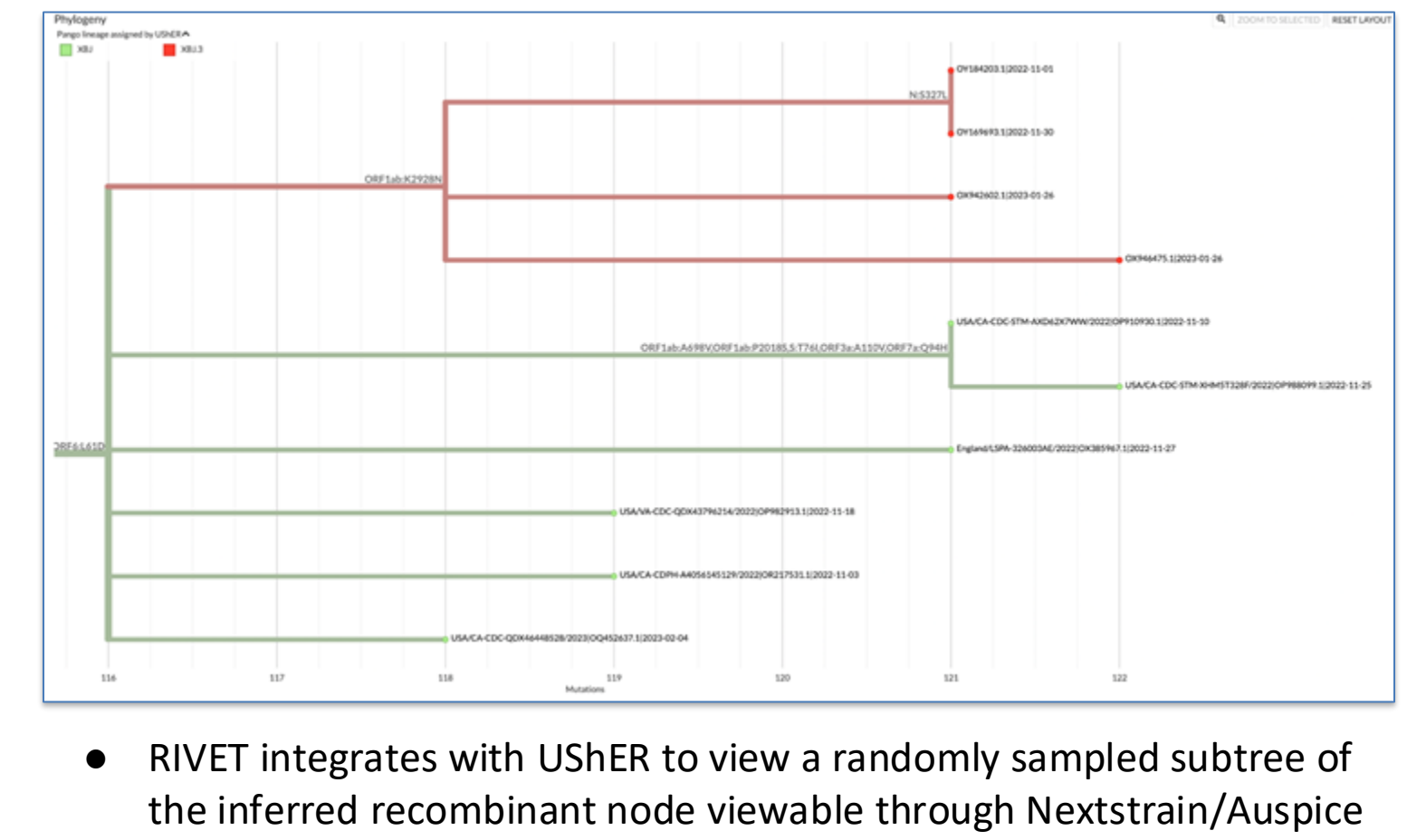
2. View single-nucleotide variants of recombinant and inferred parents (donor and acceptor)



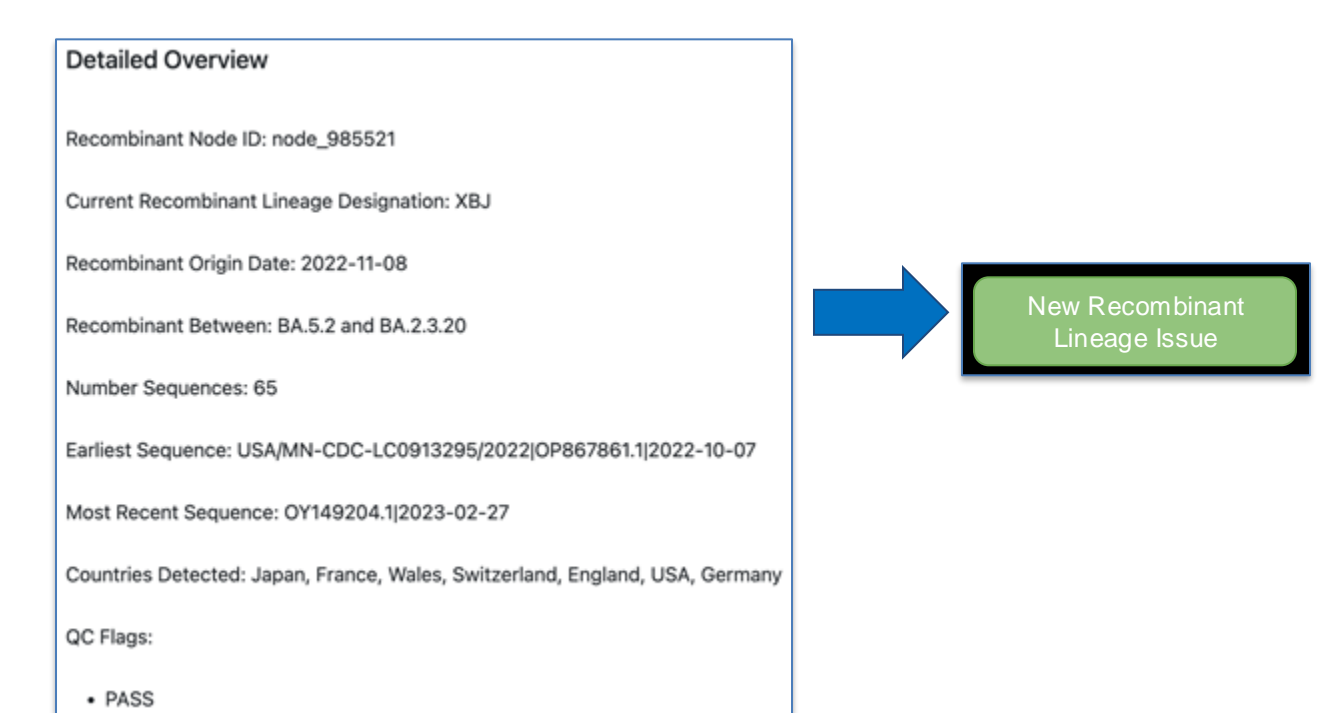
3. View recombinant in the context of global phylogeny using Taxonium



4. View randomly sampled USHER subtree in Nextstrain/Auspice



5. Rapidly build new recombinant lineage proposals



View latest SARS-CoV-2 Recombinants:

<https://rivet.ucsd.edu/>

Wiki/Documentation:

<https://turakhialab.github.io/rivet/>

GitHub:

<https://github.com/TurakhiaLab/rivet>

- RIVET aggregates all the necessary information to immediately propose a putative recombinant lineage designation for manual expert review

Future Work

1. Integration with Autoln⁶ to fully automate lineage proposals with special treatment of recombinant lineages
1. Improved representation of sequence evolution using phylogenetic networks
1. Extending RIVET to enable recombination inference and analysis of other pathogens beyond SARS-CoV-2

Acknowledgements

We are grateful for the feedback and support from our following collaborators:

- Professor Russell Corbett-Detig (UCSC)
- Angie Hinrichs (UCSC)
- Jakob McBroom (UCSC)
- Alexander Kramer (UCSC)
- Laura Hughes (Scripps Research)



- Funding:
- Genomic Sequence Databases:

References

1. Hadfield, J. et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34, 4121–4123 (2018).
2. Jackson, B. et al. (2021) Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic. *Cell* 184, 5179–5188.e8.
3. Kramer, A.M. et al. (2023) Treemane Browser: co-visualization of enormous phylogenies and millions of genomes. *Bioinformatics*, 39, btac772.
4. Martin, D.P. et al. (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*, 26, 2462–2463.
5. McBroom, J. et al. (2021) A Daily-Updated Database and Tools for Comprehensive SARS-CoV-2 Mutation-Annotated Trees. *Molecular Biology and Evolution*, 38, 5819–5824.
6. McBroom, J. et al. (2023) Automated Agnostic Designation of Pathogen Lineages Evolutionary Biology.
7. Rambaut, A. et al. (2020) A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*, 5, 1403–1407.
8. Sanderson, T. (2021) Chronomutal: time tree estimation from very large phylogenies.
9. Sanderson, T. (2022) Taxonium, a web-based tool for exploring large phylogenetic trees. *PLoS* 11, e82392.
10. Turakhia, Y. et al. (2022) Pandemic-scale phylogenetics reveals the SARS-CoV-2 recombination landscape. *Nature*, 609, 994–997.
11. Turakhia, Y. et al. (2021) Ultrafast Sample Placement on Existing tRees (USHER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. *Nat Genet*, 53, 809–816.
12. Yamasoba, D. et al. (2023) Virological characteristics of the SARS-CoV-2 omicron XBB.1.16 variant. *The Lancet Infectious Diseases*, 23, 855–856.
13. Yue, C. et al. (2023) Enhanced transmissibility of XBB.1.5 is contributed by both strong ACE2 binding and antibody evasion. 2023.01.03.52247.
14. Zhan, S. H. et al. (2023) Towards Pandemic-Scale Ancestral Recombination Graphs of SARS-CoV-2.