Optimizing targeted interventions: Identifying Imported SARS-CoV-2 Lineages and Reconstructing Local Dispersal in Large Phylogenetic Datasets

Leke Lyu[a], *Guppy Stott*[a], *Gabriella Veytsel*[a], *Armand Brown*[b], *and Justin Bahl*[a]
a. Institute of Bioinformatics, Department of Infectious Diseases, Department of Epidemiology and Biostatistics, Center for the Ecology of Infectious Diseases, University of Georgia, Athens, GA
b. Houston Health Department, Houston, TX

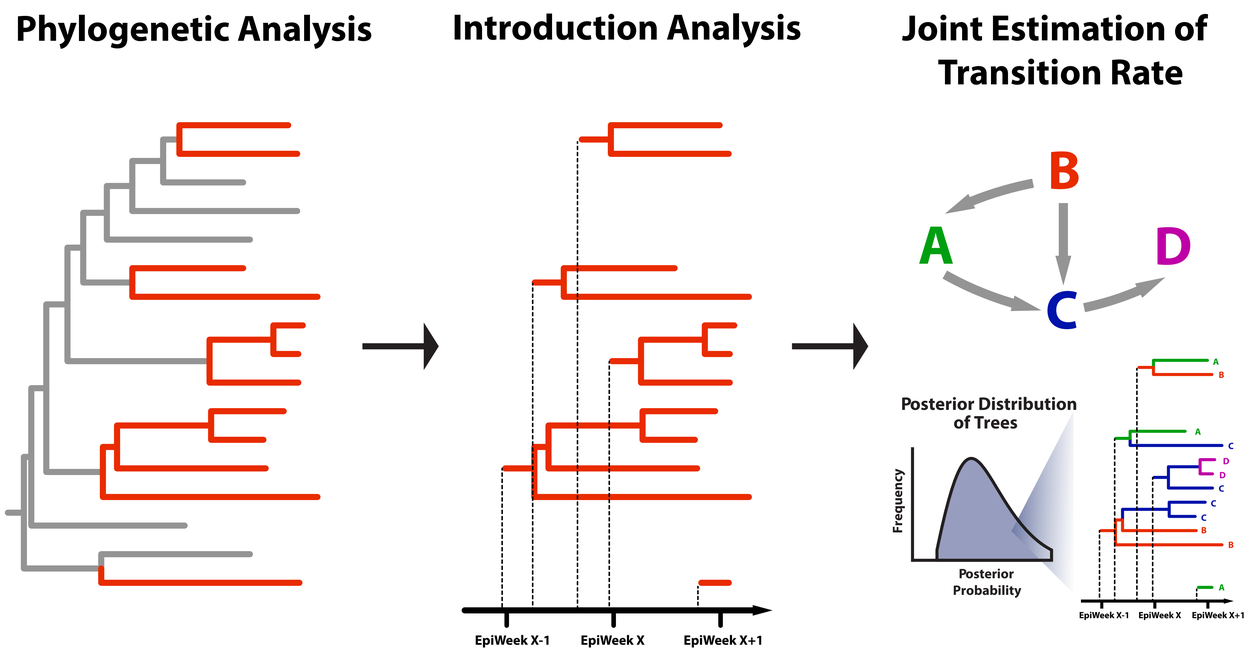
UNIVERSITY OF GEORGIA
College of Public Health
College of Veterinary Medicine
Institute of Bioinformatics

Abstract

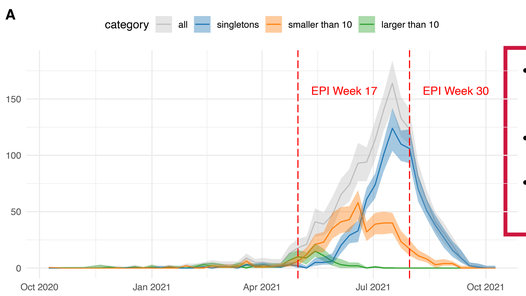
BACKGROUND:
Viral genomes contain records of geographic movements and cross-scale transmission dynamics. The growing recognition that genome sequencing is critical for outbreak response has led to an explosion in the availability of sequence data. Currently, over 16 million SARS-CoV-2 viral samples accessible in GISAID offer unprecedented opportunities for high-resolution investigations of spatial transmission history but also challenge the computational capacity of current methodologies. How can we effectively translate such vast, time-calibrated phylogenetic data, with its inferred transitions of traits, into readable epidemiological inferences?

METHOD & RESULTS:
Here, we propose an optimized workflow that utilizes a graph database to integrate posterior tree sets, clinical data, and epidemiological data, facilitating rapid importation analysis. The City of Houston, the fourth most populated city in the US, experienced significant impact from the delta outbreak. Through a collaboration with the Houston Health Department, we accessed an extensive dataset enabling the linkage of epidemiological data with sequencing data. To quantify introduction events and reconstruct regional dissemination, we applied our workflow to analyze 9,186 SARS-CoV-2 genomes from the focal area and 16,952 genomes from the rest of the world for context.

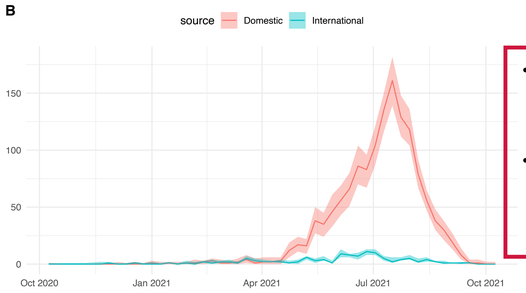
From Context-Specific Introduction Analysis to Joint Estimation of Transition Rate



Introduction of SARS-CoV-2 in Houston

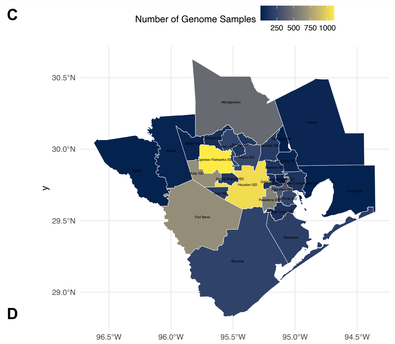


- Most introduction from outside Houston led to singletons.
- The distribution of cluster sizes was highly uneven.
- Early introduction events led to larger cluster sizes.



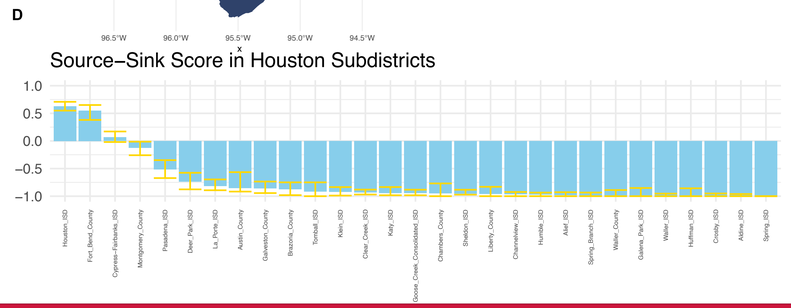
- Domestic introduction significantly outnumbering the international introduction.
- International introduction tended to occur earlier and were associated with larger cluster sizes on average.

Local Phylogeographic reconstructions

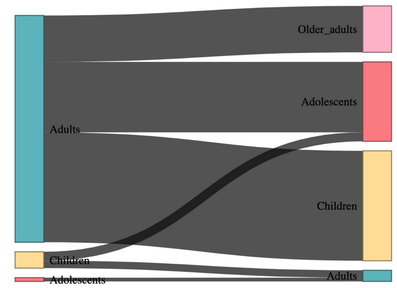


The Source Sink Score allows us to determine whether a localized outbreak may be a source or sink in other regions.

$$Source\ Sink\ Score = \frac{C_t(Export) - C_t(Import)}{C_t(Export) + C_t(Import)}$$



Demographic Determinants of Transmission



We performed an ancestral state reconstruction of sex and transmission risk characteristics using Bayesian Stochastic Search Variable Selection (BSSVS). All the transmission linkages shown here have a Bayes Factor (BF) > 3.

Fitted to our dataset, the following transitions have the highest rates: from adult to children and from female to male.

This work has been funded in part from the National Institute of Allergy and Infectious Diseases, a component of the NIH, Department of Health and Human Services, under contract no. 75N93021C00018 (NIAID Centers of Excellence for Influenza Research and Response, CEIRR) and Centers for Disease Control and Prevention, Department of Health and Human Services, under contracts 75D30121C10133 and NU50CK000626. We acknowledge the GISAID contributors (acknowledgment table of genomes used is provided on our GitHub repository) for sharing genomic data.

CDC Centers for Disease Control and Prevention · CIDER · CEIRR Centers of Excellence for Influenza Research and Response · CAIPE Center for Applied Pathogen Epidemiology and Outbreak Control