The ongoing COVID-19 pandemic marks the first time that genomic sequencing in real time has been able to inform the public health response to a pandemic2. While this wealth of genomic data can reveal powerful information, such as viral lineage and evolutionary origins, there is need for a novel framework that integrates different sources of data (e.g., spatial, demographic, genomic, and wastewater) to investigate critical public health questions. Integrating genomic and epidemiological data can elucidate transmission dynamics within and between populations, identify clusters for public health response, and determine key drivers of transmission in real time. There is increasing evidence that environmental data may be a particularly valuable addition, as it is independent of health-seeking behavior and captures asymptomatic and pre-symptomatic cases. A 2021 systematic review of wastewater surveillance for infectious diseases identified two major needs for future studies: 1) integration of epidemiological measures of disease transmission, such as disease incidence and hospitalizations and 2) assessment of the utility of wastewater surveillance to guide policy and public health intervention1.

As part of a CDC grant, we receive viral RNA extracts from the Baylor College of Medicine and the Houston Health Department (HHD) collected during clinical diagnostic testing; the Georgia Genomics and Bioinformatics Core (GGBC) sequences these samples using Illumina and Pacbio platforms. They HHD will also begin sending wastewater sequence data as part of a separate CDC grant. These different sources of data all have various strengths and limitations; combined, they can complement each other’s shortcomings and help us more accurately understand SARS-CoV-2 transmission dynamics, such as spatial diffusion within and between communities and across the United States. By describing the epidemiological characteristics of clusters, such as zip code, race/ethnicity, and age (metadata collected by HHD), we can inform targeted public health response efforts (e.g., increasing clinical testing, deploying mobile testing units, and education/outreach).

Aim 1 will be to develop and validate a novel approach integrating genomic, epidemiological, and environmental data in a statistical phylodynamic framework. We would evaluate the utility of wastewater data in this framework to inform public health action.

Aim 2 will be to investigate the molecular epidemiology and transmission dynamics of SARS-CoV2 in Houston, TX. The framework established in Aim 1 will allow us to identify high-risk communities for targeted intervention; the rich dataset we generate will also be a valuable source to answer other critical questions, such as evaluation of community mitigation strategies and detection of variants in wastewater samples not yet observed in clinical samples.

Aim 3 will be to create and maintain a Houston Nextstrain instance as a tool for the Houston Health Department. This phylogenetic visualization tool builds off of Aims 1 and 2, allowing for rapid data sharing with our partners to inform real-time public health decision making.

This study would produce outputs that will assist the Houston Health Department with responding to and controlling the COVID-19 pandemic. Furthermore, our novel framework will overcome current limitations and could serve as a model to mitigate future infectious disease outbreaks in real time. Sharing SARS-CoV-2 raw sequence reads and consensus sequences through GISAID and NCBI also supports the broader scientific community.

1. <https://doi.org/10.1101/2021.07.26.21261155>
2. <https://www.who.int/publications/i/item/9789240018440>