

# COVID-19 sampling and whole genome sequencing informs makeup of virus variants and affected populations in Bangladesh

**Catherine Martini** - Columbia University College of Dental Medicine, Class of 2024  
**Mentors: Sylvana Sinha, JD and Dr. Kim Hekimian, PhD** - Columbia University

**Research Question:** The focus of this study was to analyze the whole genome sequences of COVID-19 samples collected in Bangladesh and to identify the composition of those variants and the populations affected.

## BACKGROUND

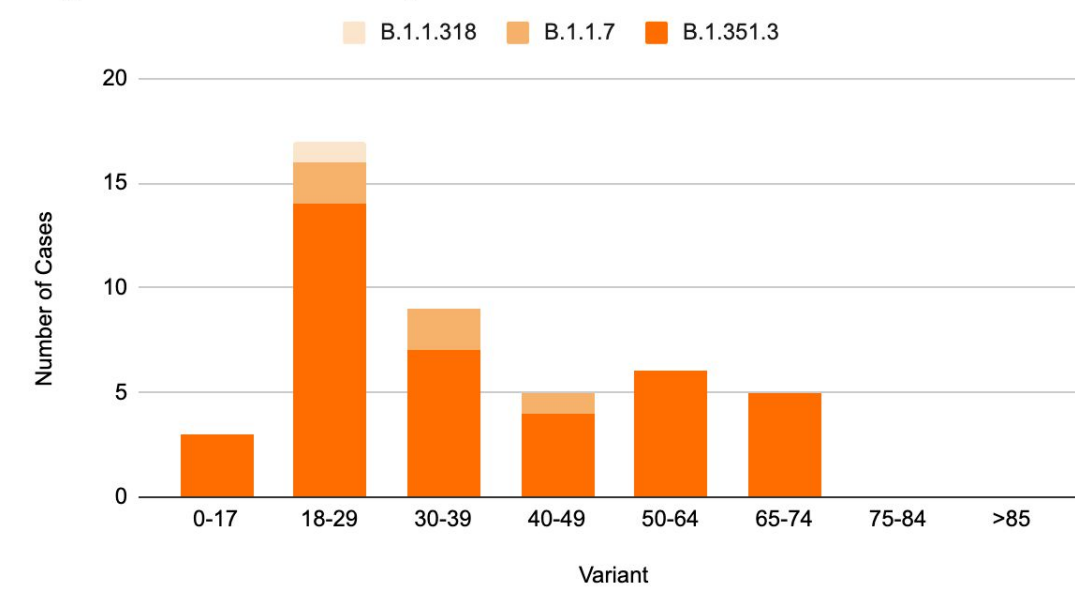
In March 2020, Bangladesh detected its first case of SARS-CoV-2. Although cases subsided in the summer and fall of 2020, a surge in March 2021 conveyed the mutability and changing infectivity of this virus. During the most recent surge in Bangladesh, the Child Health Research Foundation (CHRF) became the first to sequence SARS-CoV-2 in Bangladesh. In collaboration with Praava Health, CHRF has been able to collect and sequence COVID-19 samples from across the divisions of Bangladesh. The goal of this collaboration is to expand knowledge surrounding the strains of COVID-19 within Bangladesh in hopes to improve prevention processes.

## DESCRIPTION OF ORGANIZATION

Praava Health was founded in 2018, they have continuously strived to deliver convenient, affordable, and high-quality healthcare to the people of Bangladesh. Patients can access health services available under the Praava umbrella including pharmaceutical services, dental care, physiotherapy, and laboratory testing. Praava's laboratory participates in the world's largest quality assessment scheme, offering over 250 diagnostic tests and average accuracy score of 99.9%— making it one of the best in Bangladesh. Following the rise of the COVID-19 pandemic in March of 2020, Praava began bringing their services directly to their patients in the comfort and safety of their homes through their Video Consultation and Home Health Check services. In addition to their service-oriented changes, Praava Health has begun to collaborate with the Child Health Research Foundation to help contribute to the knowledge about the COVID-19 situation in Bangladesh.

## TABLES

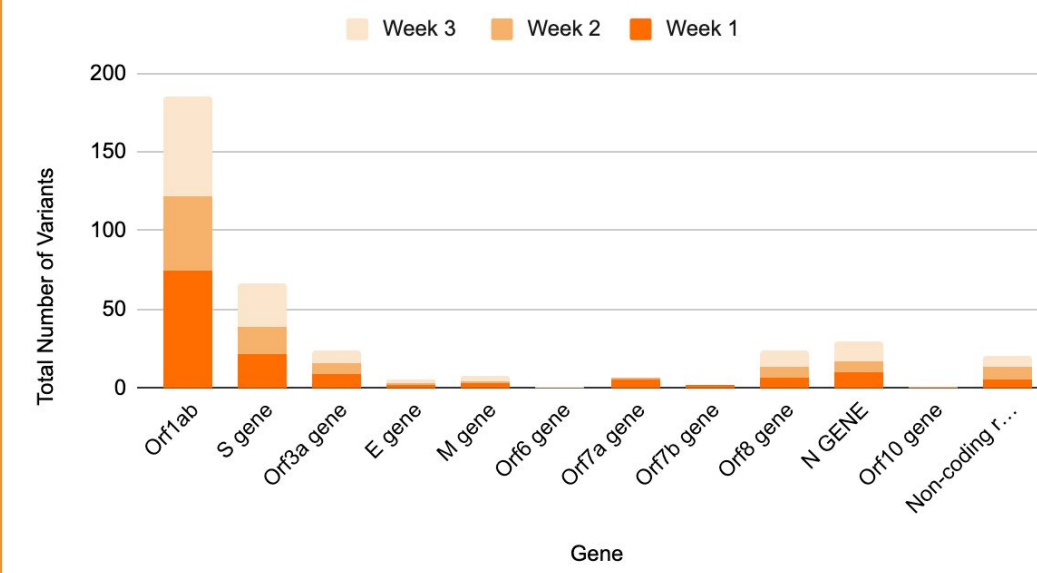
**Age Distribution of Sequenced Variants**



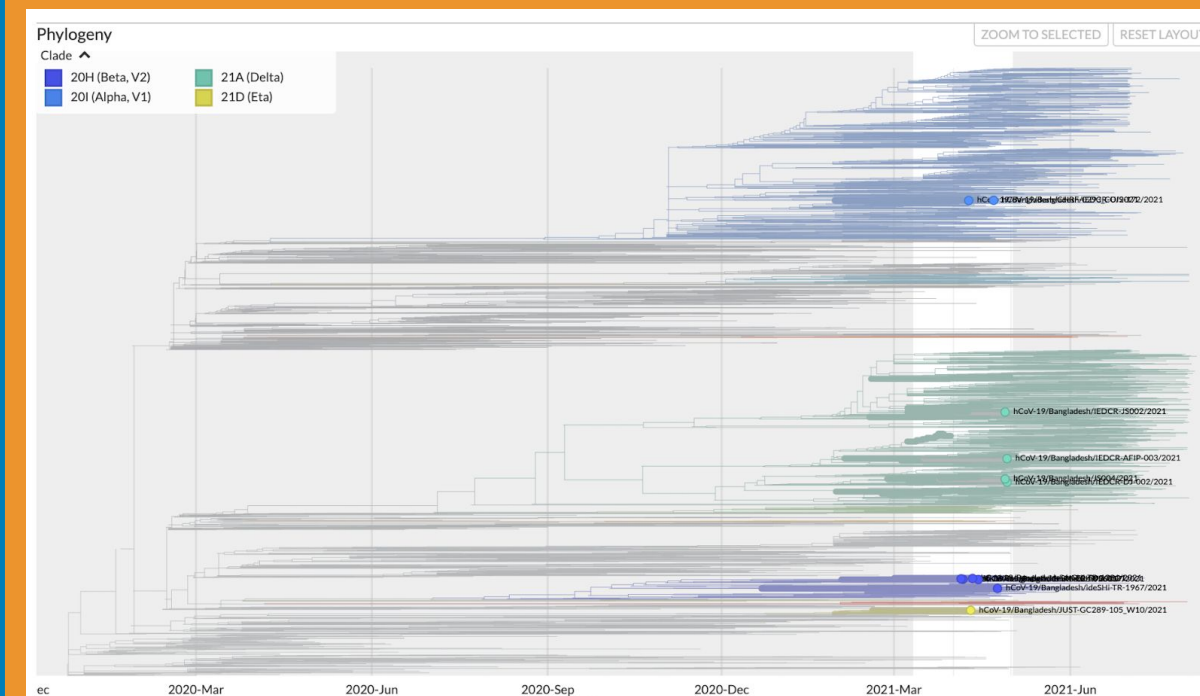
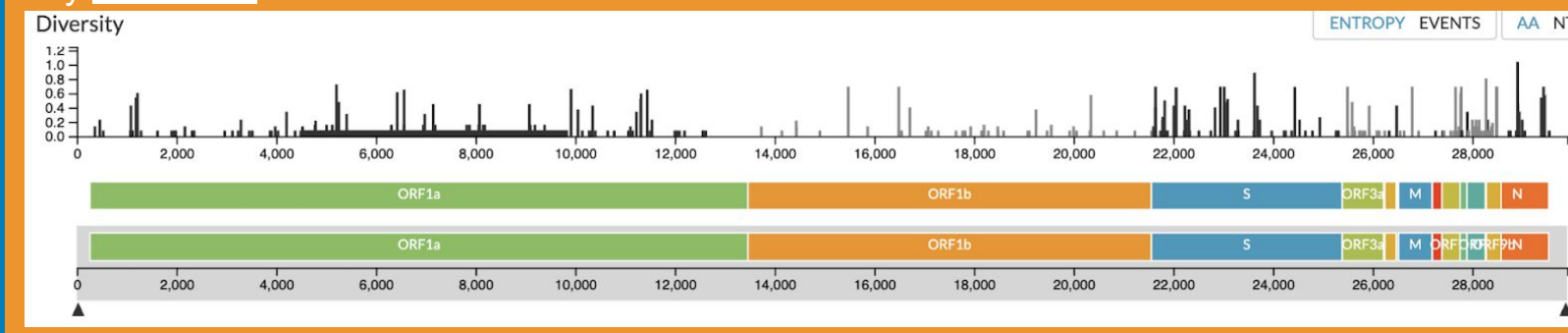
**Figure 1:** Each sample was characterized by its Pango lineage (B.1.1.318, B.1.1.7, or B.1.351.3) and was then categorized based on the patient's age. The age groups were decided based on the categorizations made by the United States Centers for Disease Control and Prevention's (CDC) [COVID-19 Death Data and Resources](#).

**Figure 2.1:** Each sample was characterized by its Pango lineage (B.1.1.318, B.1.1.7, or B.1.351.3). The variants were then categorized by the locations of their respective mutations in the genome.

**Sequencing Variants Characterized By Mutation Location**



**Figure 2.2:** The exact genomic location of each mutation and its occurrence was graphed by [Nextstrain](#).



**Figure 3:** [Nextstrain](#) has developed a SARS-CoV-2 cladogram beginning in December of 2019. The global cladogram was narrowed to showcase Bangladesh samples within the timeframe of March 2021 until the beginning of May 2021.

## METHODS

Samples were collected from Praava patients who received a positive COVID-19 test. In order to collect additional demographic data, chronic illness history, vaccination status, and informed consent, Praava Health's call center contacted each positive patient. All data was recorded and samples were sent to CHRF in weekly batches of 10 samples. The samples were amplified using PCR and then whole genome sequencing was performed. A weekly report was used to communicate the raw findings. The sequences and related demographic data were then analyzed and summarized into a more digestible format.

## DISCUSSION

The data gathered from Praava Health's collaboration has shed light on the various strains currently circulating in Bangladesh. The three variants of SARS-CoV-2 appearing in the sequencing include B.1.1.318, B.1.1.7, or B.1.351.3. Most of the mutations in these three variants arise in the ORF1ab gene, the S gene, and the N gene. Although the variants seem to disproportionately occur in ages 18-29, the original samples were skewed towards this age group due to a reluctance of older age groups to give consent for the study.

## ACKNOWLEDGEMENT

I deeply appreciate the support of my funders, both the Global and Population Health Research Fellowship and the Tamer Center for Social Enterprise. Additionally, I would like to thank Sylvana Sinha and Dr. Zaheed Hussain from Praava Health as well as the individuals within the Program for Education in Global and Population Health, Tania Genel, Dr. Kim Hekimian, and Sofia Tinio for making this work possible.

## REFERENCES

Centers for Disease Control and Prevention. (2021, August 18). *COVID-19 provisional counts - weekly updates by SELECT demographic and Geographic Characteristics*. Centers for Disease Control and Prevention. [https://www.cdc.gov/nchs/nvss/vsrr/covid\\_weekly/index.htm#SexAndAge](https://www.cdc.gov/nchs/nvss/vsrr/covid_weekly/index.htm#SexAndAge).

CHRF Bangladesh Genomics Team et al. (n.d.). *Genomic epidemiology of novel coronavirus - Bangladesh-focused subsampling*. Nextstrain. <https://nextstrain.org/community/CHRF-Genomics/ncovBangladesh@main>.