

## **ABSTRACT**

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the etiological agent of Corona Virus Disease 2019 (COVID-19), has rapidly spread across the global human population, with up to 101 million confirmed cases and approximately 2.19 million have resulted in death (as of January 29, 2021). Despite tremendous scientific efforts and breakthrough including 4 approved vaccines, in the last one year alone, the biology and evolutionary dynamics of SARS-CoV-2 remain to be fully described. In Africa, there are currently (as of April 6, 2021) 11,325 publicly available whole genome sequences from GISAID, against a backdrop of ~4.3 million confirmed cases (Africa CDC). Owing to this paucity of genome data, understanding the genomic epidemiology of SARS-CoV-2 at different spatial scales remains a challenge. Whole genome sequence data in addition to available epidemiological data will be utilized to understand the diversity and within county transmission of SARS-CoV-2 cases at ward level and identify potential clusters of transmission at household levels where data is available. To do this, an epidemiological and genomic audit of available diagnostic and contact tracing data will be carried out. Further, phylogenetic analysis of sequenced samples will be done, and these data will be utilized to explore sequence diversity and potential clusters of transmission at smaller spatial and temporal scales.