

SARS-CoV 2 Geographic and Phylodynamic Distribution from Environmental Sources

Elijah Kolawole Oladipo^{1,2,*}, Boluwatife Ayobami Irewolede², Emmanuel Oluwatobi Dairo^{2,3}, Ayodele Eugene Ayeni^{2,4}, Moyosoluwa Precious Oyewole^{2,5}, Charles Oluwaseun Adetunji⁶, Julius Kola Oloke⁷, Tolu David Oladunni^{2,8}, Oluseyi Rotimi Taiwo^{2,9}, Sodiq Opeyemi Hammed^{2,10}, Olatunji Matthew Kolawole¹¹

¹Department of Microbiology, Laboratory of Molecular Biology, Immunology and Bioinformatics, Adeleke University, Ede, Osun State, Nigeria.

²Genomics Unit, Helix Biogen Institute, Ogbomoso, Oyo State, Nigeria.

³Department of Virology, College of Medicine, University of Ibadan, Ibadan, Oyo State, Nigeria.

⁴Department of Medical Microbiology and Parasitology, University of Ibadan, Ibadan, Oyo State, Nigeria.

⁵Department of Biochemistry, Ladoké Akintola University of Technology, Ogbomoso, Oyo State, Nigeria.

⁶Department of Microbiology, Edo State University, Uzairue, PMB 04, Auchi, 312101, Edo State, Nigeria.

⁷Department of Natural Science, Precious Cornerstone University, Ibadan, Oyo State, Nigeria.

⁸Department of Microbiology, Bowen University, Iwo, Osun State, Nigeria

⁹Faculty of Veterinary Medicine, University of Ibadan, Oyo State, Nigeria

¹⁰Department of Physiology, Ladoké Akintola University Of Technology, Ogbomoso, Oyo State, Nigeria

¹¹Department of Microbiology, University Of Ilorin, Kwara state, Nigeria

Corresponding author: Elijah Kolawole Oladipo^{1,2,*}

Abstract:- The highly contagious coronavirus disease 2019 (COVID-2019) pandemic's etiology is the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The disease is highly infectious and spreads majorly from person to person as well as by contact with viral droplets in the environment. While the role of human-to-human transmission is well elucidated, the significance of environmental transmission is not so clear. This study aims to evaluate the importance of the environment in COVID-19 spread through phylodynamic analysis of environmental isolates. Two thousand and ten (2010) whole-genome sequences from samples such as wastewater, sewage, freshwater, and hospital beds from 14 countries were retrieved from GISAID and investigated to determine trends and patterns. R, MAFFT, and Figtree software were used to conduct the analysis. Results from this study showed that Austria recorded the highest isolate caseloads with 1287 samples (about 64%). Wastewater recorded the majority of isolates with 1918 samples (97%) followed by sewage with 22 samples (1%). Isolates from freshwater and hospital beds were the least with 1 sample apiece. The most occurring clade and lineage is the GK clade lineage B.1.617.2. The research shows similarities in the phylogeny of isolates from different geographical areas that were all found in the same cluster, further suggesting the possibility of a common ancestor. It also sheds light on the role of novel transmission routes like wastewater in the spread of COVID-19 as well as the importance of measures like hand washing as a control measure.

Keywords:- SARS-CoV-2, Sewage, Waste Water, COVID-19, Environment, Phylodynamic

I. INTRODUCTION

In addition to respiratory droplets from an infected individual (Bourouiba, 2020; Chan *et al.*, 2020), SARS-CoV-2 can also be spread by fomites (Martin *et al.*, 2020), fecal-oral (Gu *et al.*, 2020), and human-to-human contact pathways. It can also be transmitted through contact with mucous membranes of the mouth, nose, and eyes with contaminated surfaces through touch. Given that SARS-CoV-2 has been found in wastewater, cleaning swabs, air, sewage sludge, and river water around the globe, the ability of COVID-19 patients to disseminate the virus raises concerns about the virus invading many ecological systems. (Patel *et al.*, 2021).

The role of environmental sources like fomites, food, and air as well as the impact of factors such as climate change and water transfer on the transmission of SARS-CoV-2 has been emphasized in several recent studies (Eslami *et al.*, 2020). In the study conducted by (Kampf *et al.*, 2020), The importance of inanimate surface contact in the transmission of SARS-CoV-2 was emphasized. SARS-CoV-2 is known to last on surfaces for days at room temperature, increasing the likelihood that it will be transmitted by contact (Kampf *et al.*, 2020; Van *et al.*, 2020). Moreover, the presence of the SARS-CoV-2 virus in wastewater treatment facilities has been validated by other studies (Hindson *et al.*, 2020; Medema *et al.*, 2020; Naddeo *et al.*, 2020). According to the WHO Coronavirus Disease Situation Dashboard, as of August 8th, 2022, there were 579,092,623 confirmed cases of COVID-19 worldwide, including 6,407,556 fatalities. A significant number of these cases may have been caused by environmental sources.

Regardless of the potential modes of COVID-19 transmission, it is critical that there is clarity on the mode of transmission of this disease in order to aid in the prevention of its spread among health workers and the general public (Hadei *et al.*, 2020; Lewis, 2020; Noorimotlagh *et al.*, 2021).

Genomic sequencing and phylogenetic analysis are recent tools used in tracing the spread and evolution of coronavirus disease both locally and globally (Martins *et al.*, 2021; du Plessis *et al.*, 2021). These tools can help to unravel the cryptic transmission chains that are not detected by travel history or contact tracing (Bedford *et al.*, 2020). The combination of genomic studies, contact tracing, and evaluation of travel records is vital in enhancing current knowledge of how this virus is transmitted (Aggarwal *et al.*, 2020). The study of genomic epidemiology aids in reconstructing the geographic distribution of viruses and is crucial for determining key epidemiological parameters such as a virus's fundamental reproductive capacity, its growth rate, the incidence of diseases, and patterns of disease prevalence. (Miller *et al.*, 2020). This knowledge has been useful in making policy on the prevailing pandemic (Bedford *et al.*, 2020; Fauver *et al.*, 2020).

The focus of this study is to investigate the prevalence of COVID-19 in the environment globally and to determine, by phylodynamic analysis, major patterns and trends among different environmental isolates. It is believed that this would enable us to have a better understanding of the possible means by which the virus is transmitted within an

environment. In this study, we analyze the phylodynamics of SARS-CoV-2 in relation to its environmental distribution to better understand the degree of environmental contamination. This will help in expanding current knowledge about the virus' transmission route with the aim of proffering effective public health measures in order to curtail its spread.

II. RESULT

Of the 2010 SARS-CoV-2 isolates' metadata obtained from GISAID, only 1986 isolates were analyzed as 24 of them had incomplete datasets and could not be analyzed. The analysis was based on geography, specimen, clade, lineage, and collection date.

➤ Geographic Distribution

• Case distribution per country

The analysis of the dataset with respect to the global distribution of reported occurrences revealed that, with 1287 samples, Austria has the greatest dispersion. followed by the USA with 625 samples. These two nations constitute the bulk of the SARS-CoV-2 sequences from environmental isolates available on the GISAID database, collectively accounting for about 97% of all reported observations. The remaining 12 nations made up 3% of the cases, with China and Liechtenstein both taking 1% apiece. The distribution of cases per country is shown in Figure 1

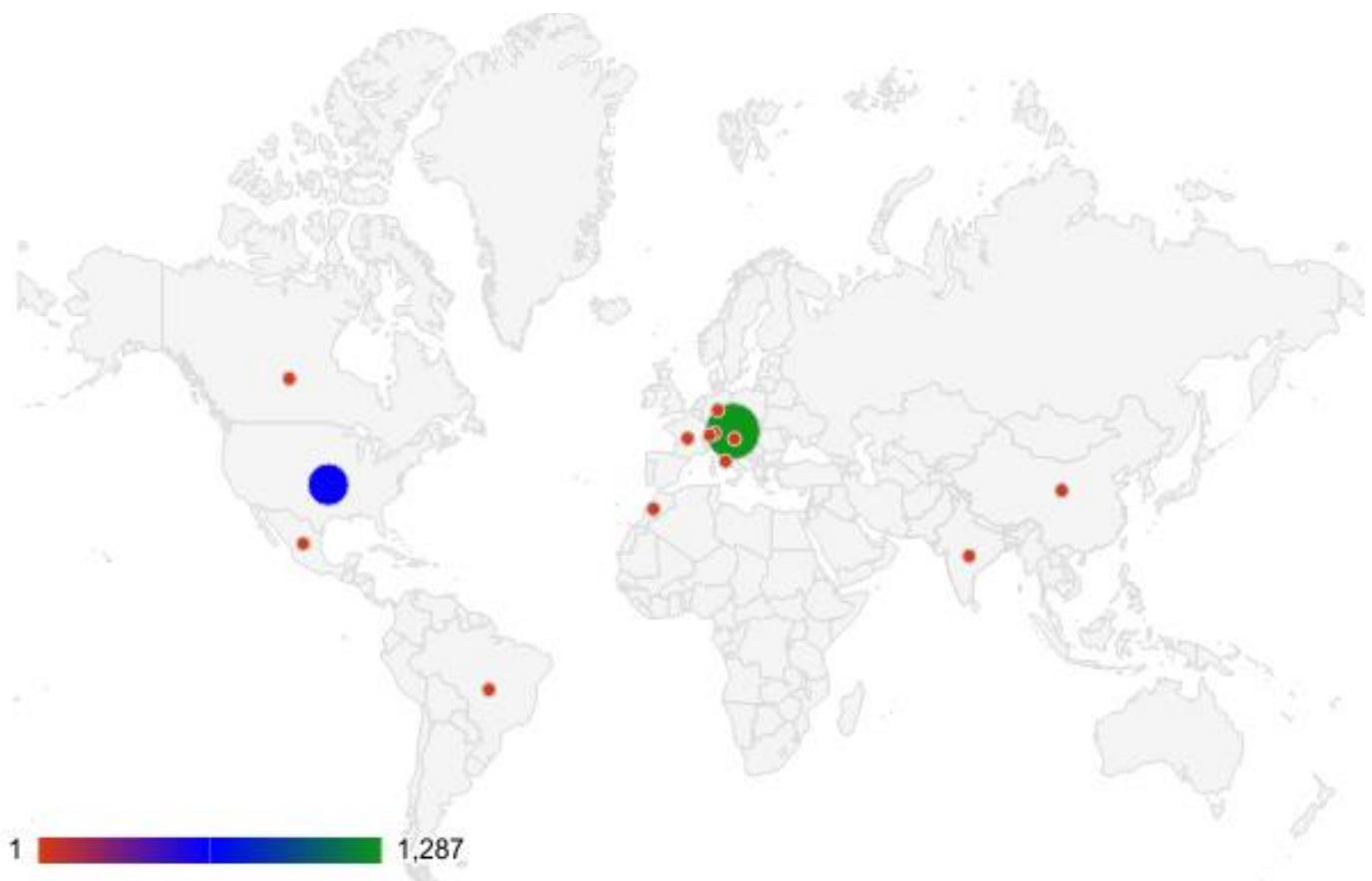


Fig 1 A Map showing the case distribution per country

• *Specimen distribution per country*

Further analysis also showed that Austria, Liechtenstein, and the United States had the bulk of their isolates obtained from wastewater, while China had the majority of positive samples obtained from swabs of outer packaging of cold-chain products. (Figure 2)

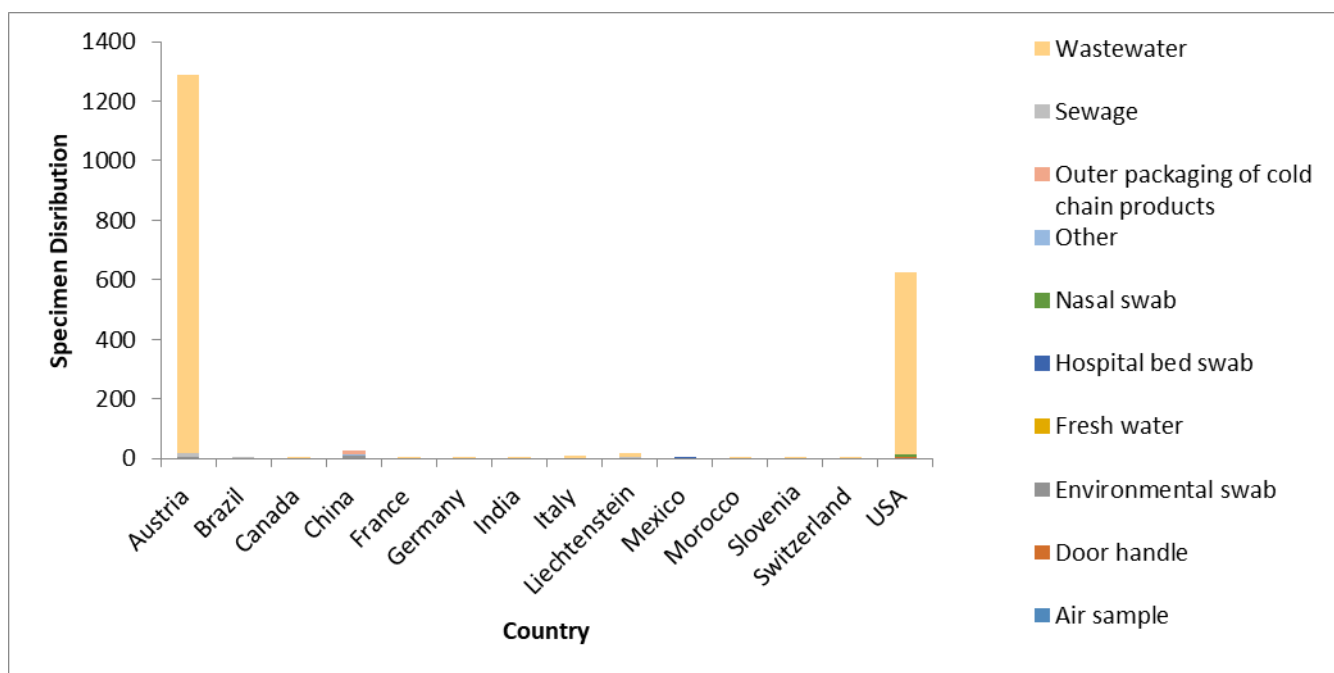


Fig 2 Chart Showing Specimen Distribution Per Country

• *Clade and lineage distribution per country*

The majority of the isolates obtained from the USA, Liechtenstein, and Austria were part of the GK clade, while China had more of the GR clade. Austria also had a significant proportion of GRA and GRY clades. Figure 3 shows the distribution of these clades per country.

Furthermore, while there was a fairly even distribution of the lineages across the countries, Austria had more of the AY.43, B.1.1.7, and B.1.617.2 lineages; China had more of the B.1.1 and B.1.1.317 lineages; Liechtenstein had a more even distribution across the lineages along with USA, but a sizable number of USA isolates were under the B.1.617.2 lineage. A detailed analysis of the lineage per country can be found in the supplementary file (S3)

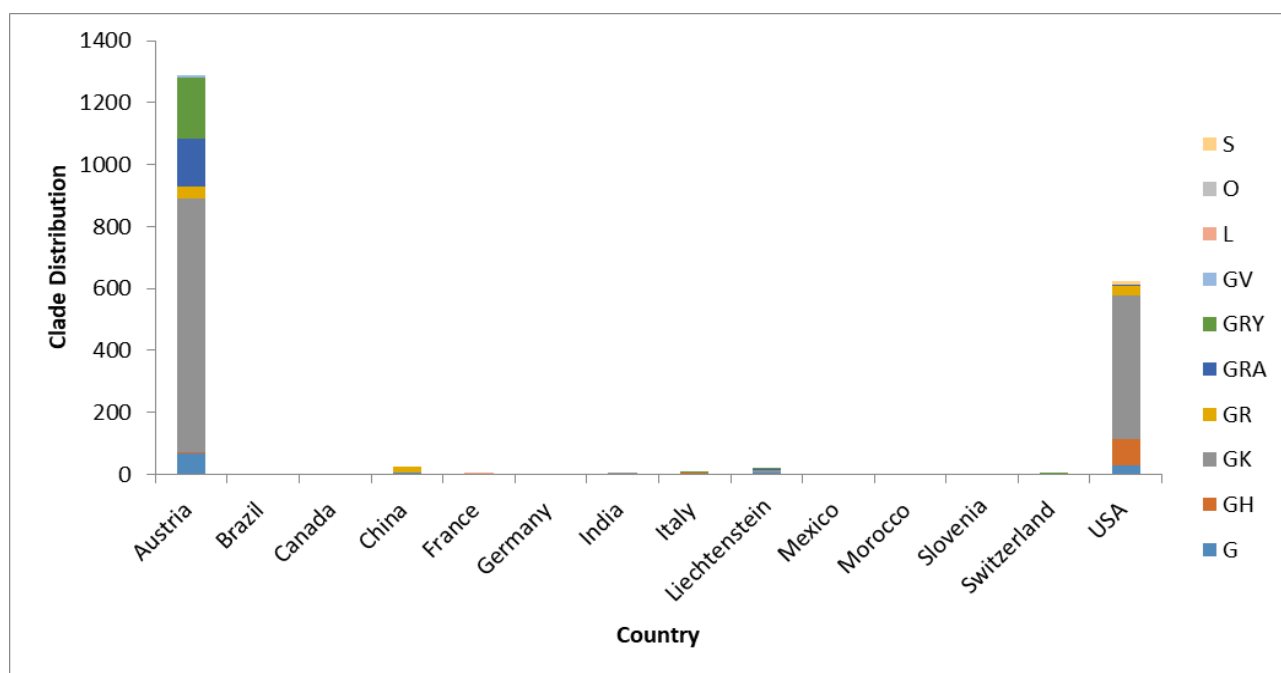


Fig 3 Chart Showing the Clade Distribution of Countries

• *Time Distribution per country*

It was also deduced from the study that the USA recorded the first environmental isolate in February 2020 (Figure 4). This was closely followed by isolates from China the following month. Austria recorded its first case in November 2020 and frequently had the most cases per month for the rest of the period under review, with the exception of three months in which more cases were recorded in the USA. It is of note that Austria recorded the highest number of monthly cases in the month of November 2021. The number of cases started to decline from January 2022, and by May 2022, only 5 cases were recorded across all countries.

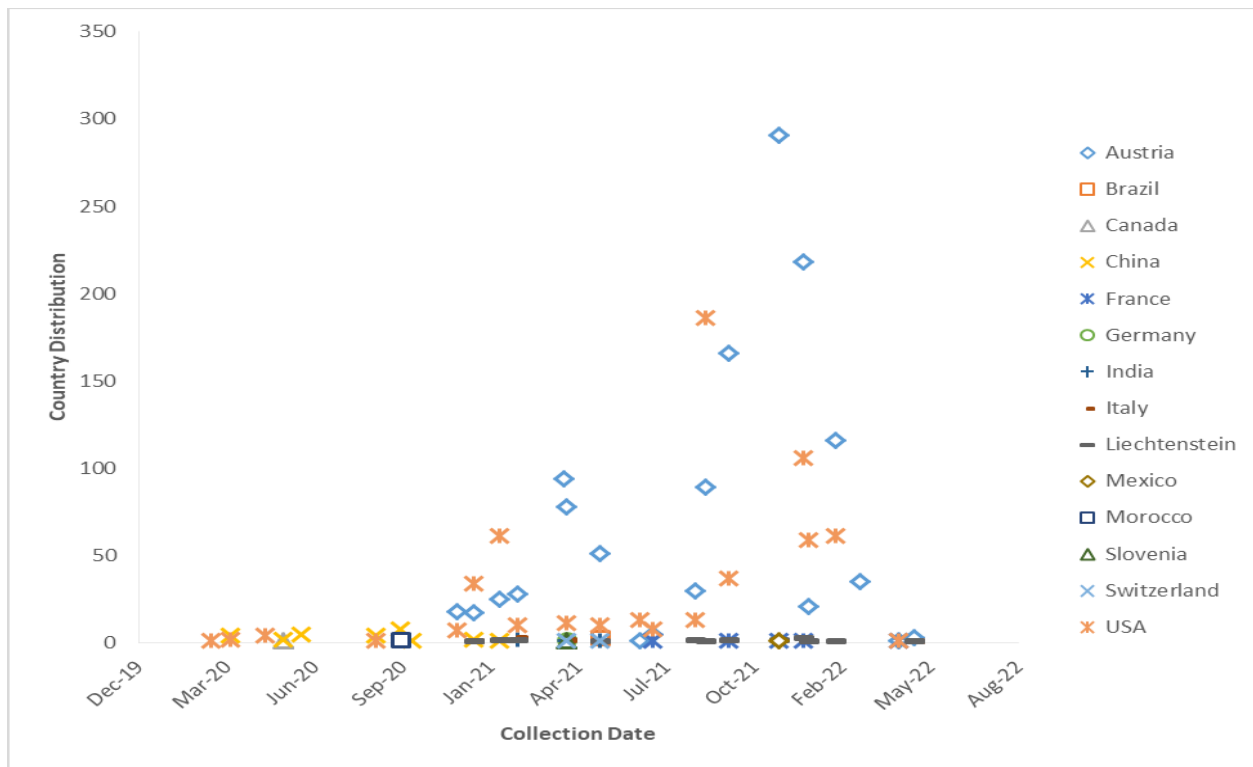


Fig 4 Chart Showing the Collection Date of Isolates by Country

• *Specimen distribution across continents*

On a continental level, Europe had the highest number of cases, followed by North America, while Africa had the lowest number of reported cases. The majority of isolates were obtained from wastewater across all continents, with the exception of Asia, which had swabs from the outer packaging of cold chain products accounting for approximately 80% of their cases. (Figure 5)

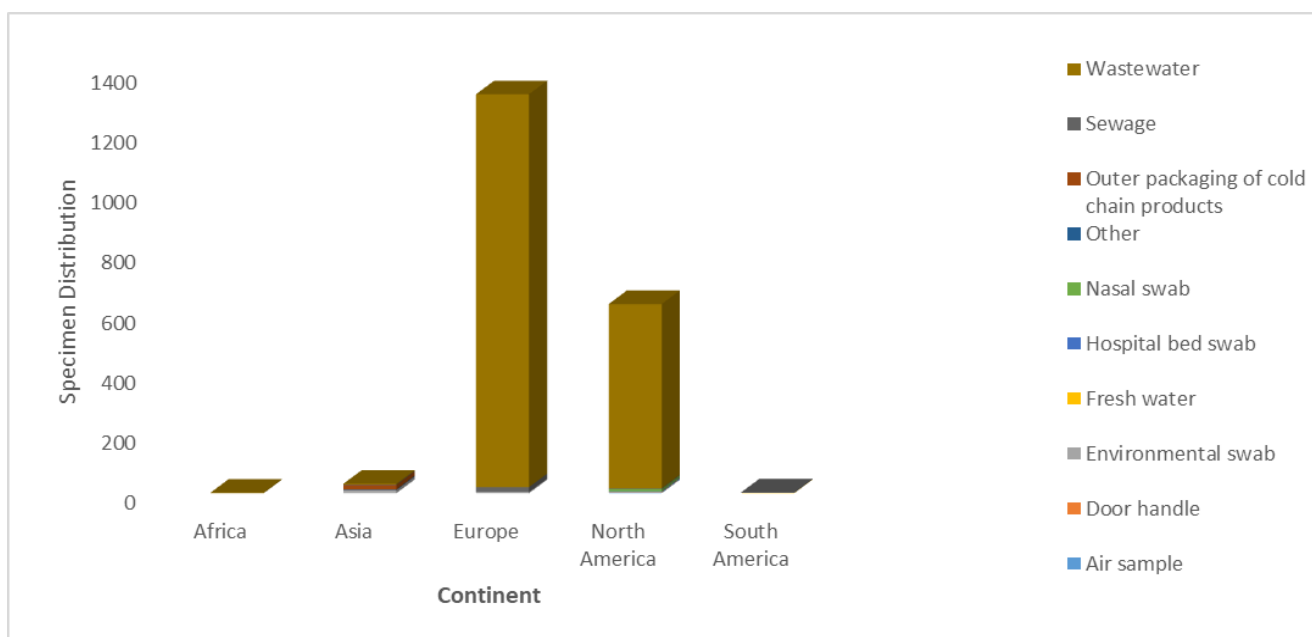


Fig 5 Chart Showing Specimen Distribution Per Continent

• *Clade and lineage distribution across the different continents*

The isolates from Europe and North America were majorly from the GK clade, while Asian and South American isolates were predominantly from the GRA clade. Figure 6 revealed that all of Africa's cases were within the G clade.

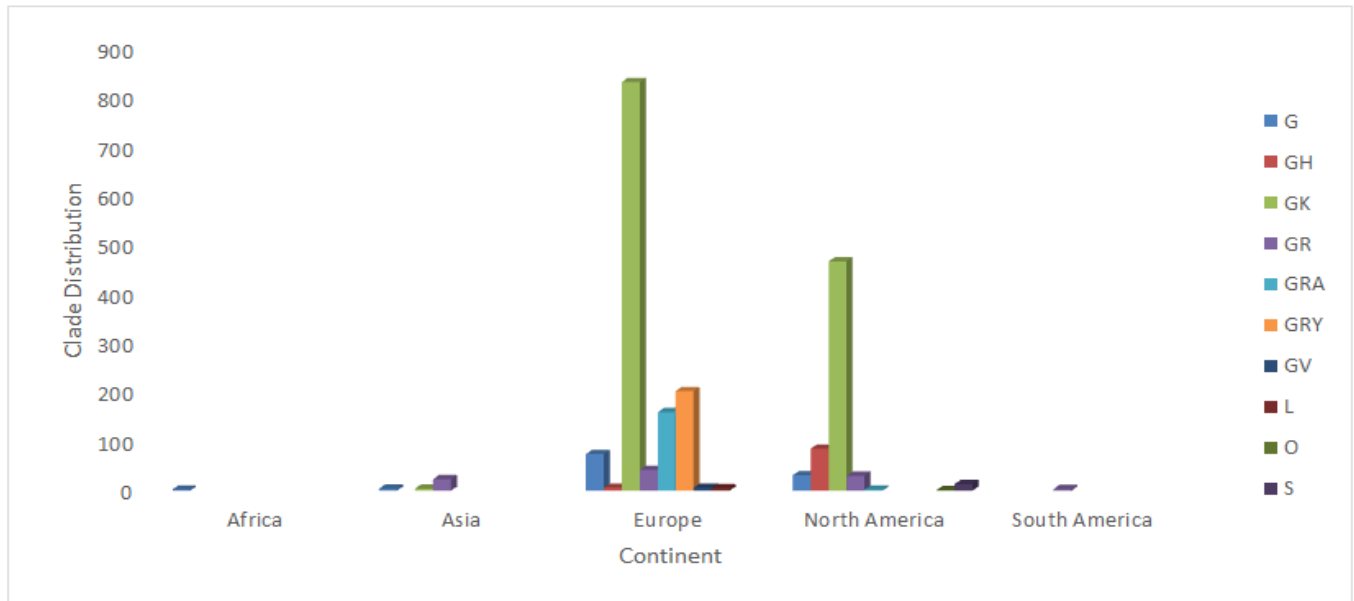


Fig 6 Chart Showing the Clade Distribution of the Continent.

Most of Europe's cases fell within the AY.43, B.1.1.7, and B.1.617.2 lineages, while Asia had more of the B.1.1 and B.1.1.317 lineages; North America had a more even distribution across the lineages but with a concentration of cases at the B.1.617.2 lineage. A detailed analysis of the lineage distribution by continent can be seen in Supplementary File (S4)

• *Time distribution per continent*

It was also deduced from the study that North America recorded the first environmental isolate in February 2020. This was closely followed by isolates from Asia the following month. Europe recorded its first case in November 2020 and consistently recorded the highest number of cases per month for the rest of the period under review, apart from a few months in which North American cases superseded. It is of note from Figure 7 that the highest number of cases per month was recorded in Europe in November 2021. The number of cases started to decline from January 2022, and by May 2022, only 5 cases were recorded across all continents.

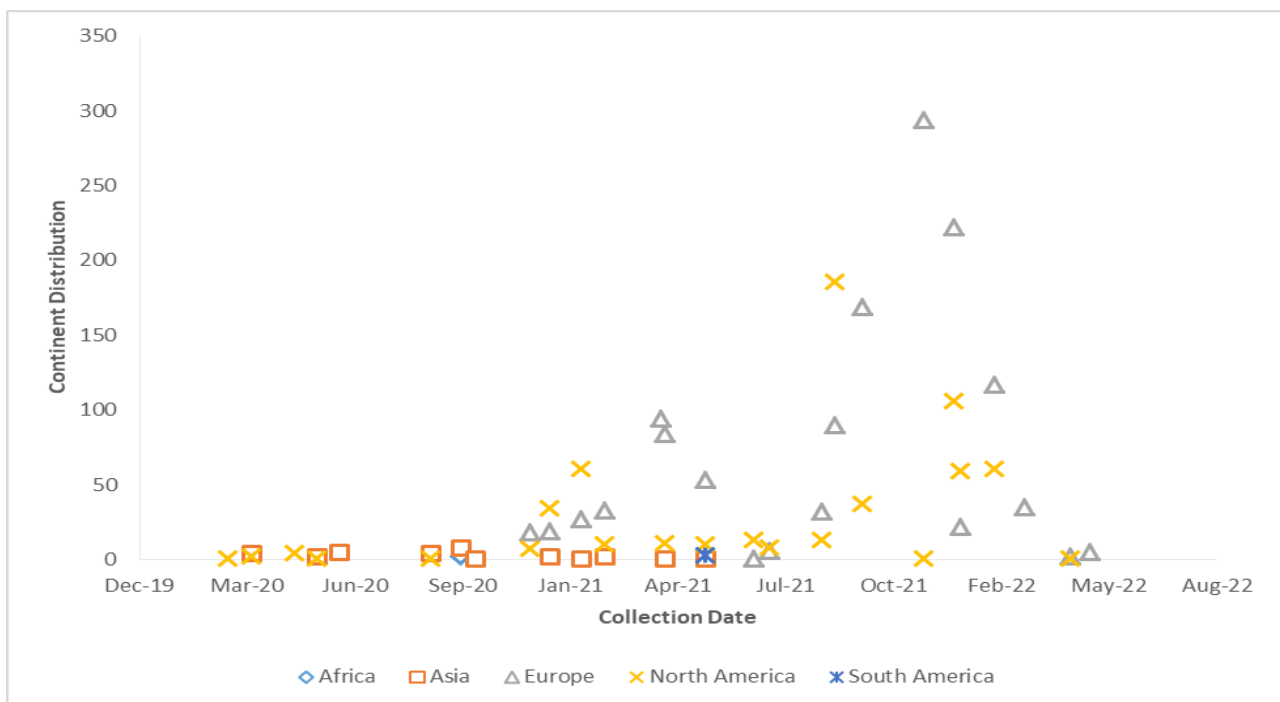


Fig 7 Chart Showing the Collection Date of Isolates Per Continent

➤ Specimen Distribution

• Clade distribution among different specimens

Evaluation of the dataset with respect to specimens showed that most isolates from wastewater, which had the bulk of samples, were from the GK clade (about 67%) (Figure 8). Sewage is a distant second and has an even distribution across the different clades. Other specimens like air samples and fresh water are solely from the GH and GR clades, respectively.

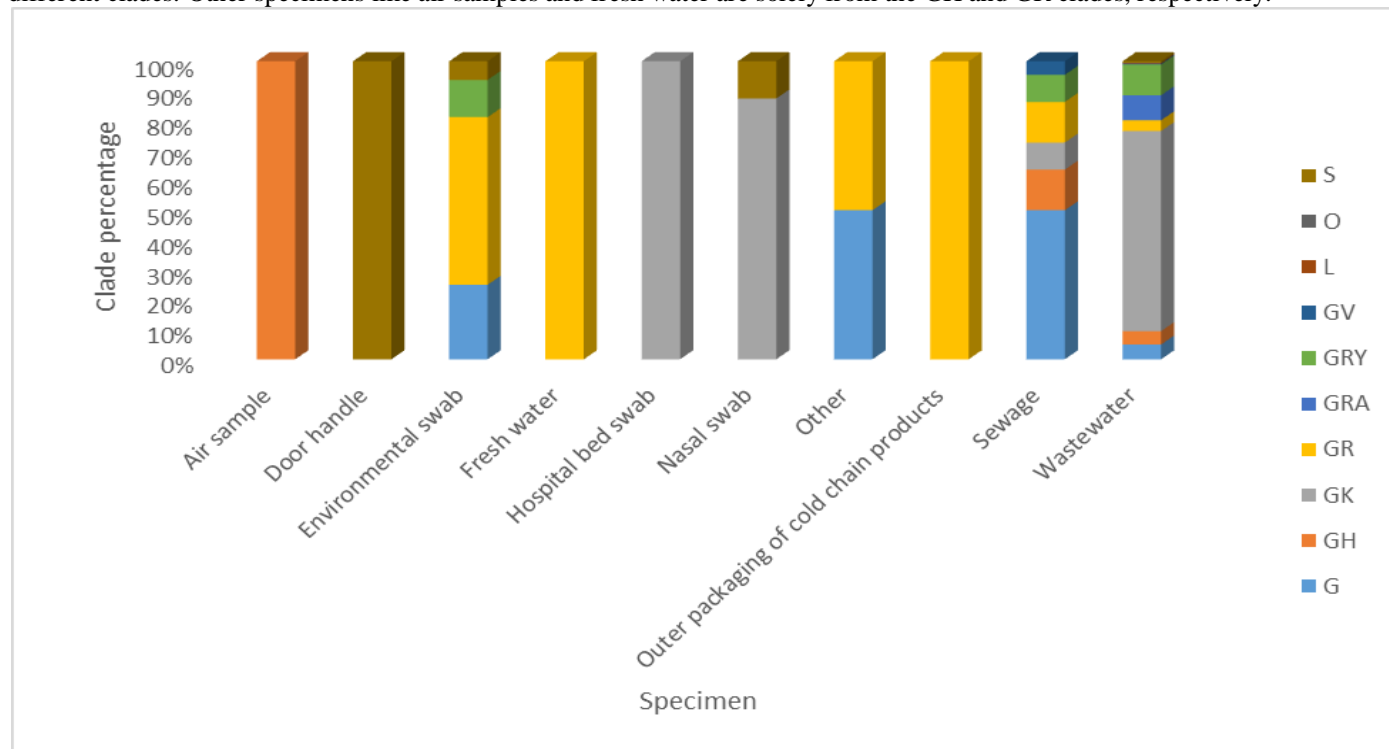


Fig 8 Chart Showing Clade Percentage Per Specimen

• Time distribution of specimen collection

Door handle swabs were the first samples collected in February 2020, followed closely by environmental and nasal swabs the following month. It can be seen from Figure 9 that wastewater had the highest collection rate from November 2020, peaking in November 2021 before declining monthly till May 2022 when just 5 cases were recorded.

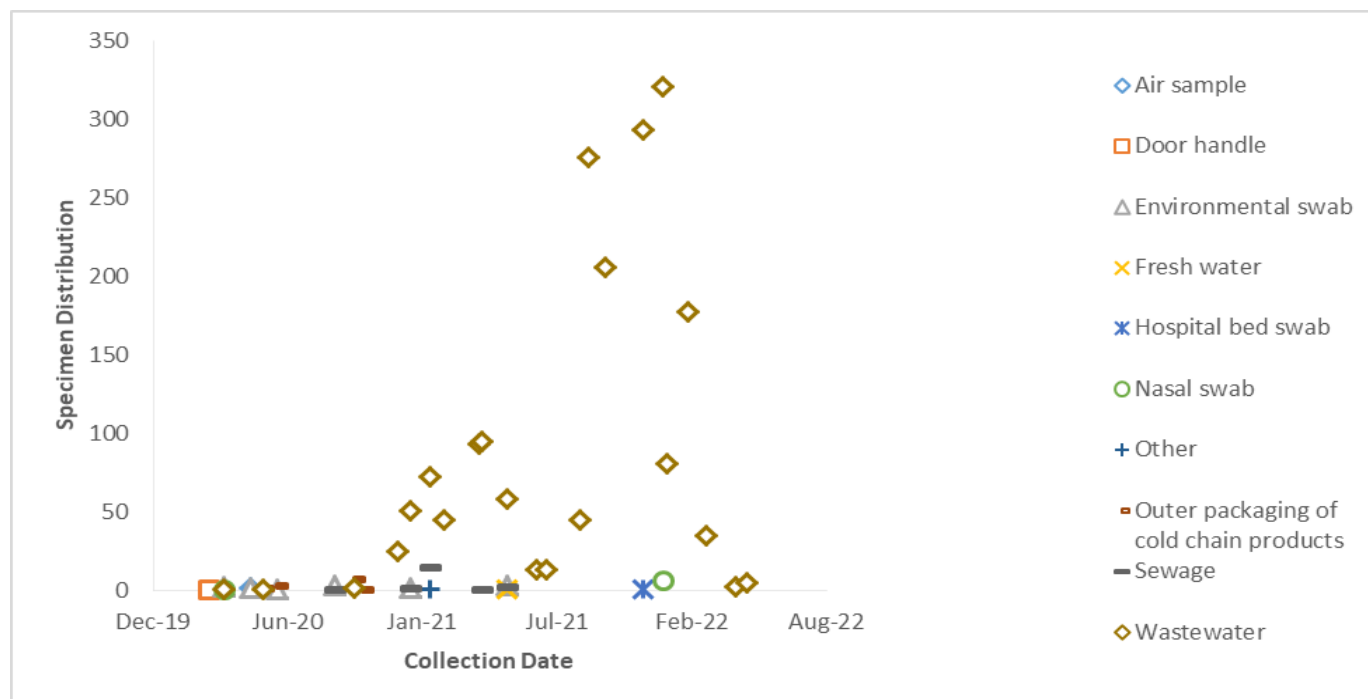


Fig 9 Chart Showing Specimen Distribution by Collection Dates

• *Lineage distribution across different specimen*

Wastewater had a fairly even distribution with most lineages represented. However, there was a dominance of the B.1.617.2 lineage with about 40% representation. Freshwater, hospital bed, and door handle samples had absolute representation within the P1, AY 100, and A1 lineages. Environmental swab was distributed across B.1, B.1.1, B.1.1.294, B.1.1.317, and B.1.1.7 lineages. A detailed analysis of the lineage distribution can be seen in the supplementary file (S5)

➤ *Clades Distribution*

• *Lineage Distribution per clade*

Of the 10 clades present, GK had the highest number of lineages and the highest number of isolates; it has B.1.617.2 as its highest occurring lineage.

GRY comes second in the number of isolates but has just two lineages: B.1.1 and B.1.1.7. A detailed analysis of the lineage distribution per clade can be seen in the supplementary file (S6)

• *Time distribution of Clade*

The S clade was the pioneer clade detected in February 2020, followed by the G and GH clades the following month. GK, the most predominant clade, was detected most often in December 2021. (Figure 10). G and GH also had a significant presence, occurring in most months under review. The GRA clade, reported in May 2022, was the last clade reported in the period under review.

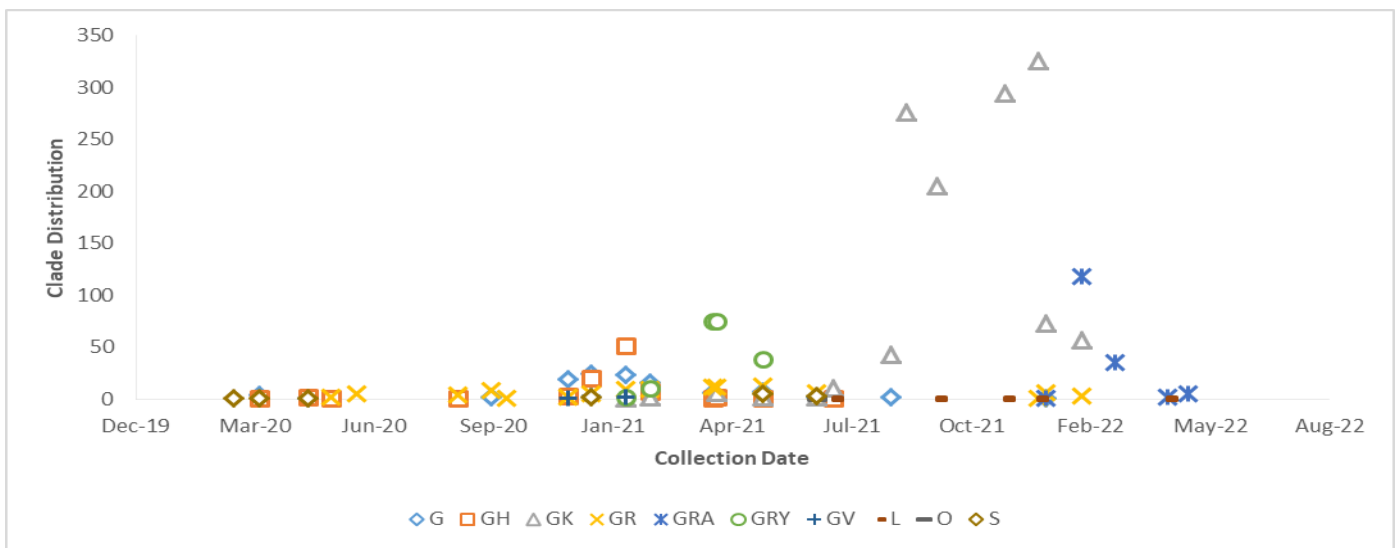


Fig 10 Chart Showing Clade Distribution Per Collection Date

➤ *Lineages*

There were 98 lineages in total appearing across the specimens collected over the period under review. B.1.617.2 was the most prevalent lineage, followed by B.1.1.7. A few lineages had just one isolate each, reflecting them. (Figure 11)

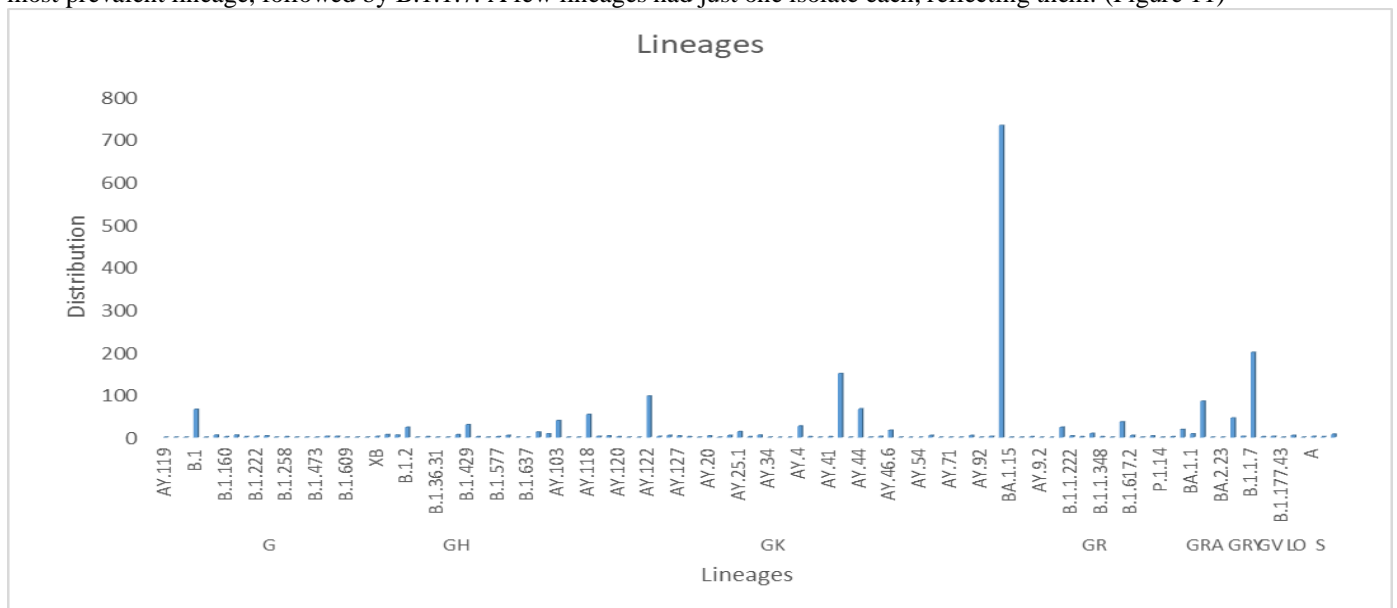


Fig 11 Chart Showing Lineage Distribution

➤ *Collection Date*

The collection of environmental samples started in February 2020 and a total of 1986 well-documented samples were sequenced as of 31st June 2022. December 2021 was the month with the highest sample collection, followed by November 2021. The number of cases started declining from January 2022, and just five cases were recorded for the month of May 2022. Figure 12 shows a detailed view of the number of cases recorded per month for the review period.

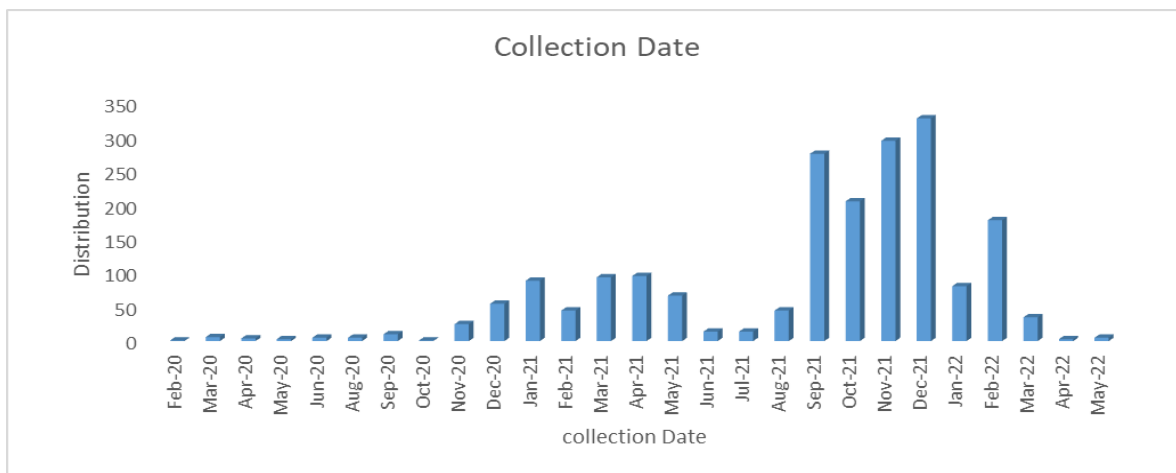


Fig 12 Chart Showing Collection Date Distribution

➤ *Phylogeny*

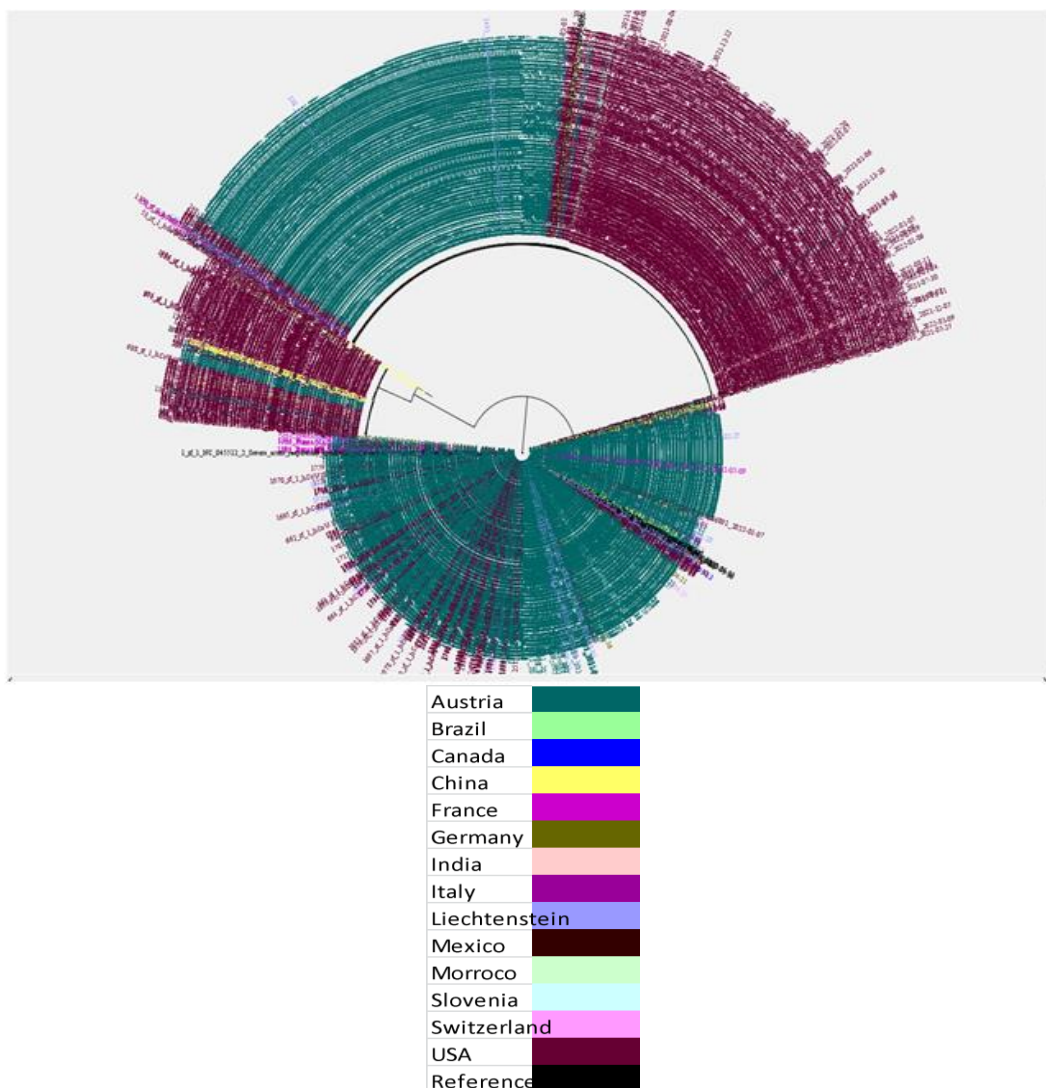


Fig 13 Phylogeny Tree of Isolates from SARS CoV 2 Isolates from the Environment

It can be observed from the tree that there were three distinct clades and 55% of the isolates, majorly isolates from Austria, China, France, and USA, share close similarity with the reference genome. The intermediate clade consists mainly of isolates from USA while the most recently evolved has both USA and Austria closely linked together

III. DISCUSSION

Upon filtering the retrieved dataset, only 24 isolates were removed due to an incomplete dataset, indicating that adequate record-keeping practices were used in documenting the cases. Analysis of the remaining data revealed a number of trends as regards the proliferation of the isolate in the environment.

According to the geographic distribution of the SARS-CoV-2 sequences from environmental isolates based on their locales, Brazil and Liechtenstein had the fewest samples, while Austria had the most cases that had been reported. Austria is a central European nation with 8.8 million inhabitants and a highly advanced national epidemiological monitoring network.. Contact tracing had been completed for all 21,821 reported SARS-CoV-2-positive patients as of August 7, 2020. 10,385 of these cases were connected to epidemiological clusters. A considerable overlap between these two lines of evidence was found when epidemiological data and phylogenetic analysis of Austrian SARS-CoV-2 sequences were combined. 199 of the 345 sequences (or 65 percent) were allocated to epidemiological clusters (Popa et al., 2020).. In contrast, SARS-CoV-2 epidemiological surveillance was insufficient in Brazil and Liechtenstein because of their middle-income status (World Bank, 2022). The Brazilian government's response to the outbreak has also generated some controversy. Their President's opposition to social isolationist measures and underestimation of the importance of the COVID-19 safety protocols serve as proof of this (Lancet, 2020). The USA was second in terms of caseload, though it reported less than half of the cases Austria reported. With a population of about 329 million and spanning about half the landmass of the North American continent, it shows a less-than-par surveillance program compared to that of Austria.

It was also discovered that the most common specimen isolates retrieved from Austria and USA were wastewater. This could be as a result of the policy adopted by these nations to test environmental samples to gauge viral load among the population, as it was becoming increasingly difficult to get people to undergo tests (*National Wastewater Surveillance System*, 2022). China instead had a larger percentage of their isolates from the outer packaging of cold chain products. This is in line with China's position as a manufacturing and exporting hub. According to a study by Ma *et al.* (2021), there have been occasional instances of the nucleic acid of the COVID-19 virus being found on the packaging of imported cold-chain chicken and cold-chain shrimp in China. The study strongly suggested that the Dalian COVID-19 outbreak was caused by COVID-19 virus-contaminated imported cold-chain seafood products, suggesting the possibility of COVID-19 virus transmission

through cold-chain products. This is also corroborated by epidemiological investigations, laboratory testing (nucleic acid and antibody detection), and full-length genome sequencing. The low number of cases, on the other hand, could be an indication of a lack of commitment to goods surveillance, which is required to prevent virus exportation or an indication of the success of China's aggressive zero-case policy.

USA, Liechtenstein, and Austria isolates were largely of the GK clade. This shows the consistency of the strain circulating in the environment of these countries. In contrast, China had more of the GR clade. This variation can be chalked up to the distance between the countries and the effectiveness of transboundary control measures. This is in contrast to a study by Mercatelli *et al.* (2020), which proposes that the G clade and its progeny, GH and GR, currently make up the majority of SARS-CoV-2 genomes that have been sequenced, making up 74 percent of all worldwide sequences. The SARS-CoV-2 population is currently represented most frequently globally by the GR clade, which carries the Spike D614G and Nucleocapsid RG203KR mutations. Clade L, which represents the initial viral strain, still makes up 7% of the sequenced genomes, while the derived clades S and V have comparable frequencies in the whole dataset. Our study also highlights the organism's inability to mutate effectively outside of the human host. This is in line with a study carried out by Mercatelli *et al.* (2020), in which about 48,635 samples were investigated. This study suggests that with an average of 7.23 mutations per sample in comparison to the reference, the virus has a low mutation rate of the genomic sequences for SARS-CoV-2

There were 98 lineages identified across all the isolates reported, and each country had its own unique set of lineages. For example, while Austria had more of the AY.43, B.1.1.7, and B.1.617.2 lineages, China had more of the B.1.1 and B.1.1.317 lineages, and a sizable number of the USA isolates were under the B.1.617.2 lineage. This is in line with the inference from the clade distribution that showed that there was a significant difference between the types of isolates circulating within different regions.

The first environmental sample was taken and sequenced in February 2020. This was three months after the virus was discovered in China. The USA was the first nation to record an environmental case. China followed closely the next month, but the USA showed more monthly consistency in the following months. Austria started its first environmental testing in November 2020 and quickly surpassed the USA in the number of cases per month. This is an attestation of Austria's sequencing capabilities. As earlier noted, China's lack of cases in the following months after its first case can be chalked up to its zero-case policy as China was able to bring the spread of the virus under control earlier and better than other nations. (Hazra, 2022)

On a continental level, Europe had the highest number of cases, most of which were reported by Austria. The ease of movement across European nations could have contributed to the high number of cases. This was highlighted by Al-Salem *et al.* (2021) in their study that stated that the spread of COVID-19 was made worse by free travel within Europe. North America came second, with the USA contributing the most. Africa had the least number of cases, and while the continent reported a relatively low number of human cases at the peak of the pandemic, the low environmental cases could be a result of the gross lack of sequencing capabilities across the continent rather than the low proliferation of the virus. It is also of note that the few cases reported were from Morocco and none from the sub-Saharan region of the continent.

Wastewater was the most collected sample and hence the most reported across Africa, Europe and North America. Asia had outer packaging of cold storage goods and environmental swabs forming the bulk of its samples, while South America had sewage. The continents with a large proportion of wastewater samples comprise countries with good sequencing technologies like Austria and are able to make use of environmental tests to gauge the viral load among the population. Asia, being dominated by China, as a major goods production hub, had environmental swabs and outer packaging of cold storage goods as its main sample types.

Europe and North America had the GK strain circulating predominantly in the environment. Other strains were also represented but in lesser quantities than the GK strain. The similarities in strains across the different continents are an indicator of the travel patterns between the two continents. Asia and South America had more of the GRA clade, reflecting the effect of the large distance between the eastern and western continents as well as the effectiveness of transboundary control measures.

Most European and North American cases fell within the AY.43, B.1.1.7, and B.1.617.2 lineages, reflecting the similarities in clade, while Asia had more of the B.1.1 and B.1.1.317 lineages, reflecting the differences in clades as mentioned earlier.

At a distance of around 5 m, droplets transmit SARS viruses predominantly via respiratory infections, but they can also be indirectly transmitted by fomite through contact with contaminated items, fecal-oral pathways, and airborne transmission (Wang *et al.*, 2005; Naddeo *et al.*, 2020). According to the results of the distribution of SARS-CoV-2 environmental sources, samples separated from wastewater had the highest frequency of sequences, whereas samples from hospital bed swabs, door handles, and freshwater had the lowest frequency. The prevalence of SARS-CoV 2 in wastewater could be a result of the high nutrient level in the source. This corresponds to Cheng *et al.* (2020) observations on air and ambient samples of SARS-CoV-2 near hospital patients with COVID-19. Their analysis showed that all samples taken from the air sources for patients who were individually segregated in isolation rooms for airborne

infections were negative for SARS-CoV-2 RNA. Door knobs and mobile phone surfaces were found to have lower contamination levels than other ambient samples. Following this, Gundy *et al.* (2020) and Ahmed *et al.* (2020) found the presence of coronavirus RNA in both symptomatic and asymptomatic SARS and COVID-19 patients' feces when they were exposed to raw sewage. They proposed a hypothesis for the virus' persistence in sewage for 14 days at 4 °C and 2 days at 20 °C, along with the likelihood of finding its RNA within eight days, as their study's conclusion.

Results from our study are at variance with those of other studies (Randazzo *et al.*, 2020). This suggests that because of the virus' vulnerability to disinfectants and weak environmental stability, SARS-CoV-2 may pose a lesser danger in wastewater than in sewage. According to this study, there were more findings in wastewater than in sewage. This could be due to the fact that selected countries relied more on wastewater as a means of surveillance (*National Wastewater Surveillance System*, 2022). SARS-CoV-2 RNA may be found in raw sewage, according to reports from two hospitals in Beijing, China, even if there may not always be any live SARS-CoV-2 left after chlorine treatment. SARS-CoV-2 was recovered from both feces and urine, and its RNA was discovered in both. (Wang *et al.*, 2019). However, SARS-CoV-2 was discovered for days on personal objects, bathroom, room, and floor surfaces, as well as plastic, stainless steel, copper, and cardboard. (Van *et al.*, 2020). Days-long persistence was also observed on inanimate surfaces, including wood, ceramics, metal, glass, garbage, containers, and bags. (Van *et al.*, 2020).

Aerosolization from toilet flushing, a defective plumbing system, and other sources might result in fecal-oral transmission of SARS-CoV-2 due to its presence in sewage and feces (Heller *et al.*, 2020). The results of Van *et al.* (2020), who observed a significant rate of SARS-CoV-2 viral fecal-aerosol transmission among affected people in Italian cities, possibly related to aerosol exposure, may lend credence to this. Aerosol transmission may be a more likely method of COVID-19 transfer than fecal-oral transmission in terms of its worldwide dissemination. According to data from earlier studies, several researchers concur that the primary mode of transmission for SARS-CoV-2 is through the air.

It was shown by Morawska *et al.* (2020) that SARS-CoV-2 may survive and spread through aerosols for up to hours and on surfaces for up to two days, underscoring the likelihood of both aerosol and fomite transmission. Also noted by Fear *et al.* (2020) is the fact that SARS-CoV-2 typically retains infectivity while airborne over short distances and remains for longer durations than when formed as respiratory particles. When transferring contaminated solid waste away from local populations in developing nations with subpar waste management procedures, garbage collectors and pickers risk becoming a source of infection for their neighborhood if proper handling practices are not used (Nghiem *et al.*, 2020).

Among the different samples used, varying clade prevalence was also observed. For instance, while wastewater showed the presence of all the different clades, GK was the most prevalent, consisting of about 67% of the cases. Similarly, sewage also showed the presence of a variety of clades but had a dominance of the GRA clade, whereas samples like air samples, door swabs, freshwater, and hospital bed swabs had predominantly one clade each: GH, S, GR, and GK, respectively. This is in contrast to Hamed *et al.* (2021), which showed that GR was the most prevalent clade followed by GV and GH, indicating a difference between clades prevalent in the environment and in humans. This could be an indication of the potential of different clades to thrive in different environments. GK and GRA show a propensity to exist in human waste products, while G, GH, and GR can better survive on non-biological materials.

The first sets of samples to be collected were door handles, environmental and nasal swabs. Wastewater samples were first taken in November, which coincides with the timeline at which scientists started using waste products to conduct viral load tests (*National Wastewater Surveillance System*, 2022). This was done consistently until December 2021. The number of tests done per month experienced a sharp decline, a reflection of waning efforts at viral surveillance.

B.1.617.2 was the most prevalent lineage in the wastewater samples, dwarfing all other lineages in terms of representation. Nevertheless, the study shows the ability of the different lineages to thrive in biological samples as opposed to the non-biological samples that showed a tendency to allow the existence of a limited number of lineages. For example, freshwater, hospital bed, and door handle samples had a singular lineage presence of P1, AY 100, and A1 lineages, respectively.

There were 10 clades represented from all the isolates sequenced, with GK having the highest representation followed by GRY. GK had the B.1.617.2 lineage mostly expressed, while GRY had two lineages: B.1.1 and B.1.1.7 expressed. The S clade was the first, followed by the G and GH clades. While the appearance of the G clade spans the entire period under review, clades like GR and ERA were evident only for a short period of time (January 2021 to May 2021 and January 2022 to May 2022, respectively). This signifies the ability of clades to either spread simultaneously, as is the case of G, GH, and GR, or the clades to dominate, as is the case of GRY and GRA. This is buttressed by Singh *et al.* (2021), who stated that clades V and G first emerged around the middle of January, then clades GH and GR toward the end of February, clade GV toward the end of June, and clade GRY toward the end of September 2020. Clades L and V are most likely extinct, whereas clades G, GH, GR, and GRY account for the vast majority of SARS-CoV-2 sequences found worldwide. Since Clade G first appeared, Clade S has likewise been in decline. It was also discovered during analysis that the GRA was the most prevalent clade at the time of this study. Contained in the 10 clades are 98 lineages appearing across

all countries during the period under review. The highest occurring lineage was B.1.617.2, followed by B.1.1.7.

The number of environmental tests done per month has declined over the past 6 months prior to this study. This could be due to the virus's decreased prevalence, a lack of public attention to the virus, or governments' lax response policies.

The majority of isolates being of the root clade confirms the earlier stated inference that the virus is unable to effectively mutate outside of a human host. The predominance of the Austrian and USA isolates in the most recently evolved clade can be said to be as a result of the cross-border transfer of the virus. This is probably due to the combination of relaxed COVID-19 travel protocols along with the high level of trade between the nations (*U.S. Relations With Austria*, 2021). Also of note is the fact that the highest traded goods between both nations are flavored water, serums, vaccines, and cultures (Guichard, 2022), which could be viable reservoirs for the virus.

IV. METHODOLOGY

➤ Data Retrieval

The GISAID database's SARS-CoV-2 isolates from environmental sources that had whole genome sequences available as of June 30, 2022, were retrieved. A filtering criterion that solely considers high coverage level, environmental samples, and whole genome was applied, which led to the retrieval of a total of 2010 SARS-CoV-2 complete sequences. Sequences were obtained from a total of 14 countries. To learn more about specific sequences, the metadata for the sequences was also taken from the GISAID database. A total of 24 sequences with incomplete metadata sets were then removed to avoid analytical errors, thereby reducing the number of valid sequences available for analysis to 1986. The Supplementary file (S1) includes a table with comprehensive details on the 1986 SARS-CoV-2 sequences utilized in this work. Changes were also made to sort samples used, as isolates from similar sources were grouped under a uniform heading. These changes can be seen in the Supplementary file (S2).

➤ Analysis of Data

This study examines and identifies trends and patterns in SARS-CoV-2 genomes derived from environmental sources using information that was retrieved along with nucleotide sequences from the GISAID database. The relationships among the different characteristics under examination were analyzed and determined, and the results were expressed graphically using Microsoft Excel (*MICROSOFT EXCEL*, 2010) and R (v1.2.1; Appelhans *et al.*, 2015). Graphical representations of the comparisons were created to compare SARS-CoV-2 isolates according to their location, source, clade, and lineage. The environmental source from where the isolates were collected was highlighted in order to provide light on potential channels via which this virus can spread within a certain geographic area. Knowing this will make it easier to propose ethical

laws, rules, and practices that will reduce COVID-19's exponential expansion and future environmental stress.

➤ *Phylogenetic Analysis*

The NC 045512.2 SARS-CoV-2 reference Wuhan genome was retrieved from the NCBI Gen Bank. The recovered 1986 SARS-CoV-2 isolates' whole genome sequences were aligned with reference NC 045512 using MAFFT version 7 (Katoh et al., 2019; Kuraku et al., 2013) set to the default settings. To determine the common ancestor shared by each strain, a neighbor-joining tree was built using the MAFFT phy ML engine based on the aligned sequences and bootstrapped at 500. (Tingting et al., 2020). The tree was then viewed using Figtree v1.1.4 (Rambaut, 2010)

V. ACKNOWLEDGMENT

The authors acknowledge the GISAID community and the authors from the original laboratories who contributed to the genetic sequence data provided through the GISAID Initiative, which served as the foundation for this study. Special appreciation for the technical assistance goes to Helix Biogen Institute in Ogbomoso, Nigeria.

VI. CONCLUSION

Because coronaviruses have a complex genetic makeup and a strong susceptibility to genetic mutations and gene recombination, the likelihood of interspecies transmission is increased (Helmy et al., 2020). Our findings indicate that Austria has significantly contributed the most sequences, clades, and lineages to the database. The least number of sequences, clades, and lineages are found in sequences from Brazil and Liechtenstein. The Wuhan-Hu-1 reference genome had a 90% similarity with 1108 nucleotide sequences from this study's multiple sequence alignment, indicating shared evolutionary origins. The patients' travel history to the infection hotspot may be connected to this (China). The results of the phylodynamic analysis showed the period of the most recent common ancestor of the viral isolates that were collected and the beginning of an epidemic in a particular population (Sang, The phylogenetic study of the SARS-CoV-2 sequences shows that while there are many areas of contact between the American and Austrian sequences, the bulk of the American sequences differ significantly from the Austrian sequences in terms of evolution.

On the other hand, the Austrian and Chinese sequences have homology in common. Regardless of the geographic location of the sequences, the phylogeny of SARS-CoV-2 whole-genome sequences from Austria, the United States, China, Brazil, and Liechtenstein showed that the SARS-CoV-2 viruses were all grouped together. This raises the idea that all SARS-CoV-2 strains have a recent common ancestor or that the virus strain was spread throughout the various nations.

In summary, our work demonstrates how SARS-CoV-2 isolates from distinct geographic regions are genetically connected and trapped in the same cluster regardless of their environmental circumstances. This demonstrates how their common progenitor may have existed. Additionally, wastewater may be a route for virus transmission, as evidenced by the high prevalence of SARS-CoV-2 from environmental sources in specimens isolated from wastewater samples. Wastewater is also a useful tool for determining the viral load in a population without the constraints of human cooperation. Before being released into the environment, wastewater produced from various facilities must undergo thorough treatment in order to further stop the spread of the dangerous virus. Investing in sequencing technology can help nations better understand the trends and patterns of viral propagation domestically.

REFERENCES

- [1]. Aggarwal, D., Page, A. J., Schaefer, U., Savva, G. M., Myers, R., Volz, E., ... & Gallagher, E. (2020). An integrated analysis of contact tracing and genomics to assess the efficacy of travel restrictions on SARS-CoV-2 introduction and transmission in England from June to September. *medRxiv*.
- [2]. Al-Salem, W., Moraga, P., Ghazi, H., Madad, S., & Hotez, P. J. (2021). The emergence and transmission of COVID-19 in European countries, 2019–2020: a comprehensive review of timelines, cases and containment. *International Health*, 13(5), 383–398. <https://doi.org/10.1093/inthealth/ihab037>
- [3]. Appelhans, T., Detsch, F., & Nauss, T. (2015). remote: Empirical Orthogonal Teleconnections in R. In *Journal of Statistical Software* (Vol. 65, Issue 10, pp.1–19). <https://www.jstatsoft.org/article/view/v065i10>
- [4]. Bárcena, M., Oostergetel, G. T., Bartelink, W., Faas, F. G., Verkleij, A., Rottier, P. J., ... & Bosch, B. J. (2009). Cryo-electron tomography of mouse hepatitis virus: Insights into the structure of the coronavirus. *Proceedings of the National Academy of Sciences*, 106(2), 582-587.
- [5]. Bedford, T., Greninger, A. L., Roychoudhury, P., Starita, L. M., Famulare, M., Huang, M. L., Nalla, A., Pepper, G., Reinhardt, A., Xie, H., Shrestha, L., Nguyen, T. N., Adler, A., Brandstetter, E., Cho, S., Giroux, D., Han, P. D., Fay, K., Frazar, C. D., . . . Wolf, C. R. (2020). Cryptic transmission of SARS-CoV-2 in Washington state. *Science*, 370(6516), 571–575. <https://doi.org/10.1126/science.abc0523>
- [6]. Bourouiba, L. (2020). Turbulent gas clouds and respiratory pathogen emissions: Potential implications for reducing transmission of COVID-19. *JAMA*, 323(18), 1837–1838. <https://doi.org/10.1001/jama.2020.4756>

- [7]. Chan, J.F.-W., Yuan, S. F., Kok, K.-H., To, K.K.-W., Chu, H., Yang, J., Xing, F. F., Liu, J. L., Yip, C.C.-Y., Poon, R.W.-S., Tsoi, H.-W., Lo, S.K.-F., Chan, K.-H., Poon, V.K.-M., Chan, W.-M., Ip, J. D., Cai, J.-P., Cheng, V.C.-C., Chen, H., ... Yuen, K.-Y. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *Lancet (London, England)*, 395(10223), 514–523. [https://doi.org/10.1016/S0140-6736\(20\)30154-9](https://doi.org/10.1016/S0140-6736(20)30154-9)
- [8]. Chen Y, Liu Q, Guo D. 2020. Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol* 92:418–423. <https://doi.org/10.1002/jmv.25681>.
- [9]. Collins AR, Knobler RL, Powell H, Buchmeier MJ. 1982. Monoclonal antibodies to murine hepatitis virus-4 (strain JHM) define the viral glycoprotein responsible for attachment and cell–cell fusion. *Virology* 119:358–371. [https://doi.org/10.1016/0042-6822\(82\)90095-2](https://doi.org/10.1016/0042-6822(82)90095-2).
- [10]. *Data for Brazil, Upper middle income | Data.* (2022). The World Bank Group. Retrieved July 30, 2022, from <https://data.worldbank.org/?locations=BR-XT>
- [11]. du Plessis L, McCrone JT, Zarebski AE, Hill V, Ruis C, Gutierrez B, et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science*.2021;371(6530):708–12. <https://doi.org/10.1126/science.abf2946>.
- [12]. Eslami, H., & Jalili, M. (2020). The role of environmental factors to transmission of SARS-CoV-2 (COVID-19). *AMB Express*, 10(1). <https://doi.org/10.1186/s13568-020-01028-0>
- [13]. Eurosurveillance Editorial T. (2020). Note from the editors: World Health Organization declares novel coronavirus (2019-nCoV) sixth public health emergency of international concern. *Euro Surveill*, 25(5), 2001231. <https://doi.org/10.2807/1560-7917.ES.2020.25.5.200131e>
- [14]. Fauver, J. R. *et al.* Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. *Cell* 181, 990–996.e5 (2020).
- [15]. Fears, A. C., Klimstra, W. B., Duprex, P., Hartman, A., Weaver, S. C., Plante, K. S., ... & Roy, C. J. (2020). Persistence of severe acute respiratory syndrome coronavirus 2 in aerosol suspensions. *Emerging infectious diseases*, 26(9), 2168.
- [16]. Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L. M., Samborskiy, D. V., Sidorov, I. A., Sola, I., & Ziebuhr, J. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, 5(4), 536–544. <https://doi.org/10.1038/s41564-020-0695-z>
- [17]. Gu, J., Han, B., & Wang, J. (2020). COVID-19: gastrointestinal manifestations and potential fecal-oral transmission. *Gastroenterology*, 158(6), 1518–1519. <https://doi.org/10.1053/j.gastro.2020.02.054>
- [18]. Guichard, A. (2022). *United States (USA) and Austria (AUT) Trade | OEC.* OEC - The Observatory of Economic Complexity. <https://oec.world/en/profile/bilateral/country/usa/partner/aut#:~:text=The%20main%20products%20that%20Austria,to%20%2410.6B%20in%202020.>
- [19]. Guo, M., Yan, J., Hu, Y., Xu, L., Song, J., Yuan, K., ... & Wang, D. (2022). Transmission of SARS-CoV-2 on Cold-Chain Food: Precautions Can Effectively Reduce the Risk. *Food and Environmental Virology*, 1-9.
- [20]. Hadei, M., Mohebbi, S. R., Hopke, P. K., Shahsavani, A., Bazzazpour, S., Alipour, M., & Yarahmadi, M. (2021). Presence of SARS-CoV-2 in the air of public places and transportation. *Atmospheric Pollution Research*, 12(3), 302–306.
- [21]. Hamed, S. M., Elkhatib, W. F., Khairalla, A. S., & Noreddin, A. M. (2021). Global dynamics of SARS-CoV-2 clades and their relation to COVID-19 epidemiology. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-87713-x>
- [22]. Hazra, D. (2022, July 30). *China abandons economic growth target but insists on zero-Covid: Why the politburo has made this a priority.* TimesNow. <https://www.timesnownews.com/exclusive/china-abandons-economic-growth-target-but-insists-on-zero-covid-why-the-politburo-has-made-this-a-priority-article-93228238>
- [23]. Hindson, J. (2020). COVID-19: faecal–oral transmission?. *Nature Reviews Gastroenterology & Hepatology*, 17(5), 259-259.
- [24]. Hu, B., et al. (2020). Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology*, 19, 11740–21534.
- [25]. Kampf, G., Todt, D., Pfaender, S. and Steinmann, E. 2020. Persistence of coronaviruses on inanimate surfaces and their inactivation with biocidal agents. *J. Hosp. Infect.* 2020, 104, 246–251.
- [26]. Lancet, T. (2020). COVID-19 in Brazil: “So what?”. *Lancet (London, England)*, 395(10235), 1461.
- [27]. Lewis, D. (2020). Is the coronavirus airborne? Experts can’t agree. *Nature*, 580(7802), 175.
- [28]. Ma, H., Zhang, J., Wang, J., Qin, Y., Chen, C., Song, Y., Wang, L., Meng, J., Mao, L., Li, F., Li, N., Cai, J., Zhang, Y., Wang, D., Xia, Y., Wang, H., Jiang, S., Zhao, X., Niu, P., . . . Xu, W. (2021). COVID-19 Outbreak Caused by Contaminated Packaging of Imported Cold-Chain Products — Liaoning Province, China, July 2020. *China CDC Weekly*, 3(21), 441–447. <https://doi.org/10.46234/ccdcw2021.114>
- [29]. Martin MA, VanInsberghe D, Koelle K. Insights from SARS-CoV-2 sequences. *Science*. 2021;371(6528):466–7. <https://doi.org/10.1126/science.abf3995>.
- [30]. McBride R, van Zyl M, Fielding BC. 2014. The coronavirus nucleocapsid is a multifunctional protein. *Viruses*6:2991–3018.<https://doi.org/10.3390/v6082991>
- [31]. Medema, G., Heijnen, L., Elsinga, G., Italiaander, R. and Brouwer, A., 2020. Presence of SARS-Coronavirus-2 in sewage. *MedRxiv*.

- [32]. Mercatelli, D., & Giorgi, F. M. (2020). Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Frontiers in Microbiology*, *11*. <https://doi.org/10.3389/fmicb.2020.01800>
- [33]. Microsoft Corporation. (2010). *Microsoft Excel*. Retrieved from <https://office.microsoft.com/excel>
- [34]. Miller, D., Martin, M. A., Harel, N., Tirosh, O., Kustin, T., Meir, M., Sorek, N., Gefen-Halevi, S., Amit, S., Vorontsov, O., Shaag, A., Wolf, D., Peretz, A., Shemer-Avni, Y., Roif-Kaminsky, D., Kopelman, N. M., Huppert, A., Koelle, K., & Stern, A. (2020). Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. *Nature Communications*, *11*(1). <https://doi.org/10.1038/s41467-020-19248-0>
- [35]. Morawska, L., & Cao, J. (2020). Airborne transmission of SARS-CoV-2: The world should face the reality. *Environment international*, *139*, 105730.
- [36]. Naddeo, V., Liu, H., 2020. Editorial perspectives: 2019 novel coronavirus (SARS-CoV 2): what is its fate in urban water cycle and how can the water research community respond? *Environmental Science Water Research & Technology* *6*, 1213–1216. <https://doi.org/10.1039/DOEW90015J>.
- [37]. *National Wastewater Surveillance System*. (2022, May 16). Centers for Disease Control and Prevention. <https://www.cdc.gov/healthywater/surveillance/wastewater-surveillance/wastewater-surveillance.html>
- [38]. Neuman BW, Adair BD, Yoshioka C, Quispe JD, Orca G, Kuhn P, Milligan RA, Yeager M, Buchmeier MJ. 2006. Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. *J Virol* *80*:7918 –7928. <https://doi.org/10.1128/JVI.00645-06>.
- [39]. Neuman BW, Kiss G, Kunding AH, Bhella D, Baksh MF, Connelly S, Droese B, Klaus JP, Makino S, Sawicki SG, Siddell SG, Stamou DG, Wilson IA, Kuhn P, Buchmeier MJ. 2011. A structural analysis of M protein in coronavirus assembly and morphology. *J Struct Biol* *174*:11–22. <https://doi.org/10.1016/j.jsb.2010.11.021>.
- [40]. Nghiem, L. D., Morgan, B., Donner, E., & Short, M. D. (2020). The COVID-19 pandemic: considerations for the waste and wastewater services sector. *Case studies in chemical and environmental engineering*, *1*, 100006.
- [41]. Noorimotlagh, Z., Mirzaee, S. A., Kalantar, M., Barati, B., Fard, M. E., & Fard, N. K. (2021). The SARS-CoV-2 (COVID-19) pandemic in hospital: An insight into environmental surfaces contamination, disinfectants' efficiency, and estimation of plastic waste production. *Environmental Research*, *202*, 111809.
- [42]. Patel, M., Chaubey, A. K., Pittman Jr, C. U., Mlsna, T., & Mohan, D. (2021). Coronavirus (SARS-CoV-2) in the environment: Occurrence, persistence, analysis in aquatic systems and possible management. *Science of The Total Environment*, *765*, 142698.
- [43]. Popa, A., Genger, J. W., Nicholson, M. D., Penz, T., Schmid, D., Aberle, S. W., ... & Bergthaler, A. (2020). Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Science translational medicine*, *12*(573), eabe2555.
- [44]. Rambaut, A. (2010) FigTree v1.3.1. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. <http://tree.bio.ed.ac.uk/software/figtree/>
- [45]. Randazzo, W., Truchado, P., Cuevas-Ferrando, E., Simón, P., Allende, A., & Sánchez, G. (2020). SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. *Water research*, *181*, 115942.
- [46]. Singh, J., Pandit, P., McArthur, A. G., Banerjee, A., & Mossman, K. (2021). Evolutionary trajectory of SARS-CoV-2 and emerging variants. *Virology Journal*, *18*(1). <https://doi.org/10.1186/s12985-021-01633-w> *Low income - World Bank Data*. (2022). World Bank. <https://data.worldbank.org/country/XM>
- [47]. *U.S. Relations With Austria*. (2021, April 14). United States Department of State. <https://www.state.gov/u-s-relations-with-austria/#:%7E:text=Austria%20was%20one%20of%20the,new%20jobs%20in%20both%20countries.>