



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



ELSEVIER

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Whole-genome sequencing of SARS-CoV-2 reveals diverse mutations in circulating Alpha and Delta variants during the first, second, and third waves of COVID-19 in South Kivu, east of the Democratic Republic of the Congo

Patrick Bisimwa Ntagereka^{1,*}, Samuel O. Oyola², Simon Patrick Baenyi¹, Gilbert Kibet Rono², Ahadi Bwihangane Birindwa¹, Dieudonné Wasso Shukuru¹, Tshass Chasinga Baharanyi³, Théophile Mitima Kashosi¹, Jean-Paul Cikwanine Buhendwa³, Parvine Basimane Bisimwa³, Aline Byabene Kusinza³, Rodrigue Ayagirwe Basengere¹, Denis Mukwege^{3,4}

¹ Université Evangélique en Afrique, Molecular Biology Laboratory, Bukavu, Democratic Republic of the Congo

² International Livestock Research Institute (ILRI), Nairobi, Kenya

³ Université Evangélique en Afrique, Faculty of Medicine, Bukavu, Democratic Republic of the Congo

⁴ Panzi General Referral Hospital, Gynecology and General Surgery, Bukavu, Democratic Republic of the Congo

ARTICLE INFO

Article history:

Received 4 March 2022

Revised 16 May 2022

Accepted 17 May 2022

Keywords:

COVID-19
Genetic evolution
Pandemic
SARS-CoV-2
South Kivu
Variants

ABSTRACT

Objectives: We used whole-genome sequencing of SARS-CoV-2 to identify variants circulating in the Democratic Republic of the Congo and obtain molecular information useful for diagnosis, improving treatment, and general pandemic control strategies.

Methods: A total of 74 SARS-CoV-2 isolates were sequenced using Oxford Nanopore platforms. Generated reads were processed to obtain consensus genome sequences. Sequences with more than 80% genome coverage were used for variant calling, phylogenetic analysis, and classification using Pangolin lineage annotation nomenclature.

Results: Phylogenetic analysis based on Pangolin classification clustered South Kivu sequences into seven lineages (A.23.1, B.1.1.6, B.1.214, B.1.617.2, B.1.351, C.16, and P.1). The Delta (B.1.617.2) variant was the most dominant and responsible for outbreaks during the third wave. Based on the Wuhan reference genome, 289 distinct mutations were detected, including 141 missenses, 123 synonymous, and 25 insertions/deletions when our isolates were mapped to the Wuhan reference strain. Most of these point mutations were located within the coding sequences of the SARS-CoV-2 genome that includes spike, ORF1ab, ORF3, and nucleocapsid protein genes. The most common mutation was D614G (1841A>G) observed in 61 sequences, followed by L4715L (14143 C>T) found in 60 sequences.

Conclusion: Our findings highlight multiple introductions of SARS-CoV-2 into South Kivu through different sources and subsequent circulation of variants in the province. These results emphasize the importance of timely monitoring of genetic variation and its effect on disease severity. This work set a foundation for the use of genomic surveillance as a tool for future global pandemic management and control.

© 2022 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

* Corresponding author: Patrick Bisimwa Ntagereka, Université Evangélique en Afrique, Molecular Biology Laboratory, PO Box 3323, Bukavu, Democratic Republic of the Congo

E-mail address: patrick.ntagereka@gmail.com (P.B. Ntagereka).

Introduction

SARS-CoV-2 is the causative agent of COVID-19 that emerged from China in December 2019 and has spread worldwide, including in the Democratic Republic of the Congo (DRC), causing a

global pandemic of unprecedented proportions in the recent past. The adequate clinical course of SARS-CoV-2 infection affects the lungs causing a influenza-like syndrome with clinical signs such as a cough, fever, and in more severe cases, difficulty in breathing (Zou et al., 2020; Ruzzenenti et al., 2021).

Although the pandemic continues, until now (February 8, 2022), this virus has infected more than 400 million people, with more than 5.7 million deaths worldwide (<https://www.worldometers.info/coronavirus/>). SARS-CoV-2 belongs to the *Coronaviridae* family, the *Betacoronavirus* genus, and it is among the largest single-stranded positive-sense RNA viruses. SARS-CoV-2 genome is ~30 kilobase in length, consisting of 10 open reading frames (ORFs) that code for ORF1ab polyproteins, spike (S), envelope (E), membrane (M), nucleocapsid (N), and other accessory proteins (Tai et al., 2020; Lu et al., 2020; Bar-On et al., 2020). Among the genes, S glycoprotein plays essential roles in virus attachment, fusion, and entry into the host cell by binding with angiotensin-converting enzyme 2 receptor (Tai et al., 2020) and eliciting protective humoral and cell-mediated immune responses in hosts during infection (Amanat et al., 2020; Belouzard et al., 2012).

To date, SARS-CoV-2 isolates have been classified into 11 clades based on the global initiative on sharing all influenza data (GI-SAID) nomenclature. These clades include L (the clade to which the Wuhan reference strain belongs), S, V, G, GH, GK, GR, GV, O, S, and V (Hamed et al., 2021).

The first COVID-19 case was reported in the DRC on March 10, 2020, in Kinshasa, the capital city (Butera et al., 2021). Since then, the disease has spread rapidly across the country, including South Kivu province, and as expected, outbreak dynamics differ among provinces.

Subsequently, a countrywide partial lockdown that led to schools and international boundary closures was imposed by the government from April 2020 to August 2020. Despite the enforcement of these preventive measures, the number of COVID-19 cases and deaths continued to increase. This led to the first wave of local transmission between March 2020 and August 2020. The partial lockdown was lifted when schools and other activities resumed in September 2020. However, in December 2020, another wave of COVID-19 infection hit the country, peaking in February–March 2021. As a result, new containment measures such as curfew, the closing of schools, ban on social gatherings, mandatory wearing of face masks, and proof of a negative polymerase chain reaction (PCR) test for international travelers upon arrival were enforced.

Similarly, from June 2021 to September 2021, an exponential increase of COVID-19 cases was reported across the country; this culminated in the third wave of infection. The situation deteriorated very fast, particularly in South Kivu province, where cases quadrupled, compared with previous waves where an average of 30 cases were reported per week. In the third wave, up to 320 infection cases and ~10 deaths per week were reported. This placed South Kivu among the top four provinces that were most affected by the COVID-19 pandemic. A total of 2712 infection cases and 58 deaths were recorded in South Kivu by January 14, 2022, of 83,194 cases and 1086 deaths reported countrywide. Variability was observed within and between waves in the number of infected people and the severity of the disease in affected patients. This may indicate variant strains that may be in circulation.

Genomic surveillance is an important tool for global epidemic control by enabling monitoring, evolution, tracking the spread, and investigating multiple introductions of the virus (Fraser et al., 2009; Gardy and Loman, 2018; Holshue et al., 2020). Whole-genome sequencing has proven to be a critical epidemiological tool

for understanding SARS-CoV-2, transmission dynamics, and circulating variants.

To date, few studies on genomic characterization of SARS-CoV-2 circulating in the Congolese population have been reported, and the little available information is restricted to Kinshasa, the capital city of the country. However, no information is available on SARS-CoV-2 variants circulating in the eastern part of the DRC.

South Kivu province is of special interest owing to its geographical location. It borders several countries, including Burundi, Rwanda, and Tanzania. The province consists of a mixed population from different countries bordering the DRC and may favor the introduction and circulation of different SARS-CoV-2 lineages. This study reports the complete genome sequences and phylogenetic analysis of 74 SARS-CoV-2 isolates from the first, second, and third waves of the epidemic in South Kivu province, eastern DRC.

Methods

Sample collection

A total of 130 nasopharyngeal swab samples were selected from SARS-CoV-2 positive samples (from the first, second, and third waves) at the Molecular Biology Laboratory of the Université Evangélique en Afrique, DRC. Testing was performed following a quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) method. The samples were collected in different hospitals in Bukavu city from April 2020 to September 2021 and from travelers crossing different border points between Rwanda and DRC during the same period. The nasopharyngeal samples were collected in a 5-ml tube containing 2 ml of sample storage buffer (Ensure, Biotechnology) composed of 0.9% of normal saline raisin. The sample selection criterion for whole-genome sequencing was based on high viral load predicted by the low cycle threshold (Ct) values of ≤ 25 on real-time RT-PCR analysis.

RNA extraction and complementary DNA synthesis

The viral RNA was extracted from nasopharyngeal swabs using the QIAamp Viral RNA Mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The complementary DNA (cDNA) library was constructed using FIREScript® RT cDNA synthesis kit (Solis, Biodyne, Italy) following the manufacturer's instructions. Briefly, 1 μ l of the total RNA was added into 5 μ l of the 5 \times cDNA synthesis mix containing 1 μ l of SOLIScript Reverse Transcriptase, 0.5 μ l of RiboGrip RNase Inhibitor, 2 μ l of 10 \times RT Reaction Buffer with 10-mM dithiothreitol, 1 μ l of random hexamers primers, and 0.5 μ l dNTP mix. The volume was brought up to 20 μ l with nuclease-free water.

The reactions were incubated for 10 minutes at 25°C, then 30 minutes at 60°C for reverse transcription, and 5 minutes at 85°C for enzyme activation. These steps were carried out in a thermocycler machine (ProFlex, PCR system, Applied Biosystems, Waltham, Massachusetts). The cDNA samples obtained were stored at -20°C until further use.

Next-generation sequencing and genome assembly

Library preparation and high throughput genomic sequencing was performed at the ILRI Genomic Platform (Nairobi, Kenya). SARS-CoV-2 whole-genome sequencing was carried out as described previously (Gohl et al., 2020). Briefly, viral RNA extraction was performed using TANBead® Maelstrom 9600 (Taiwan Advanced Nanotech Inc, Taiwan) automated nucleic acid extractor

following the manufacturer's directions. RNA was eluted in 60- μ l buffer and stored in RNase-free Eppendorf tubes. The ARTIC SARS-CoV-2 Library Prep Kit (ONT®) was used following workflows for Oxford Nanopore Technology (ONT) (Grubaugh et al., 2019). ONT sequencing was done on the MinION platform. Base-calling, demultiplexing, and adapter trimming was performed using Guppy version 5.0.11, and FASTQ outputs were used for downstream analysis. From the 130 sequenced samples, only 74 were considered for the downstream analysis, and 56 sequences were removed as we could assemble only 70% of the SARS-CoV-2 reference genome.

Variant calling and lineage/clade assignment were carried out using the singularity container of the nf-core/viralrecon version 2.2: an analysis pipeline for assembly and intra-host/low-frequency variant calling for viral samples (<https://github.com/nf-core/viralrecon/releases>). Read mapping and consensus generation were done by Minimap2, and medaka was used for mutation calling. The SnpEff was used for mutation annotation. Pangolin UShER was used further downstream to analyze consensus sequences for lineage assignments based on parsimony and Nextclade (Hadfield et al., 2018) for clade specification. The sequence statistics generated from this study are presented in Supplementary Material Table 1.

Phylogenetic analysis

A total of 399 SARS-CoV-2 reference sequences subsampled by Nextstrain were retrieved from the National Center for Biotechnology Information and included in the analysis. These genome sequences were filtered to remove sequences with no metadata and those with greater than 1% of ambiguous bases and partial genomes of less than 80% genome coverage. The subsampling of the reference genomes was done using the Nextstrain command-line Augur filter tool based on an already subsampled dataset of 2867 that were evenly distributed across time and geography. Our subsampling ensured that we include only genomes with key metadata, such as Pangolin lineages, sampled before August 31, 2021, from a human host and further probabilistic sampling to minimize any bias. The tree was generated through nextclades' command-line workflow and visualized using a local Auspice installation. The nexus tree was downloaded and further edited in the FigTree tool. Phylogenetic relationships were inferred using IQ-TREE version 1.6.8 with 1000 bootstrap iterations. The tree was further visualized with FigTree version 1.4.4. SARS-CoV-2 Pango lineages were assigned using Pangolin 3.1.14 (pangoLEARN version 2021-10-13). Nextclade CLI 1.4.2 tool (Aksamentov and Nehler, 2021) was used to assign SARS-CoV-2 "Nextstrain clades" to the genomes based on predetermined mutations.

Results

Clinical characteristics of the 74 tested patients and Pango lineage distribution

A total of 57 samples (77%) were from male patients within the age range of 31 to 50 years (51.3%). Most patients presented severe symptoms at admission (fever, headache, cough, difficulty in breathing, dyspnea) ($n = 38$, 51.3%) followed by asymptomatic cases ($n = 31$, 41.9%). Diabetes was the major chronic comorbidity presenting in 33 patients (44.6%), followed by hypertension in 21 patients (28.3%) (Table 1).

Based on the identified mutation profiles, the Nextclade analysis clustered the sequences in four main Pango lineages. The B.617.2 was the most prevalent (34 sequences), followed by B.1.214.2 with 29 sequences, then A.23.1 having nine sequences. The B.1351 lineage had only two sequences. Most sequences belonging to the Pango lineage B.617.2 were from male patients

($n = 26$, 74.5%) within the age range of 31 to 50 years ($n = 19$, 55.9%). Most patients in this category presented severe symptoms ($n = 20$, 58.8%), with 14 (41.2%) having diabetes, and most were admitted to hospitals ($n = 17$, 50%). Similarly, B.1.214.2 lineage comprised 19 (65.5%) samples from female patients and 10 (34.5%) samples from male patients. Most were young patients within the age range of 31 to 50 years ($n = 18$, 62%), more than half were asymptomatic ($n = 16$, 55.2%), and more than half ($n = 15$, 51.2%) had diabetes. Interestingly, 20 (68.9%) patients were admitted to hospitals. Demographic and epidemiological characteristics of patients infected with SARS-CoV-2 categorized by the Pango nomenclature are listed in Table 1.

Phylogenetic analysis

Phylogenetic analysis revealed the presence of two major SARS-CoV-2 variants of concern circulating in South Kivu, DRC. The two main variants were Delta (B.1.617.2) with its two sublineages (AY.33 and AY.39) detected in 36 samples from the third wave and Beta (B.1.351) identified in two sequences from the first and second wave. The lineages mentioned in this line are not considered variants of interest by the World Health Organization. In addition, two variants of interest, including A.23.1 and B.1.214, were identified. According to the Pangolin website report, the B.1.214 was primarily associated with the DRC and neighboring countries, whereas A.23.1 was responsible for most of the infections in Uganda. The finding most likely indicates local and imported transmission events of SARS-CoV-2 in the South Kivu metropolitan area (Figure 1). Furthermore, one sequence classified as C.16 was detected in an infected patient in November 2020 during the second wave.

Pango lineage distribution

Pangolin analysis revealed that most sequences ($n = 26$, 35.1%) belonged to Pango lineage B.1.214.2, which was the major lineage responsible for epidemics during the first and second waves of infection in the DRC. The second most common Pango lineage was B.1.617.2, found in 19 sequences (25.7%), and its subvariant AY.33 was identified in 10 sequences (13.5%), constituting the major drivers of COVID-19 epidemic during the third wave of infection. In addition, lineage B.1 was detected in 3 (4%) samples, whereas B.1.223 and C.16 were detected in one patient each and were associated with epidemics during the first wave. Pango lineages assigned to the sequences and countries where most of the lineages have been reported are listed in Table 2. In addition, the Pangolin lineages and the corresponding date/time of isolation are presented in Figure 2.

Mutational dimension

Our analysis identified a total of 289 distinct mutations in our data. These mutations included 141 missense mutations, 123 synonymous mutations, and 25 insertions and deletions. Most mutations were in the ORF1ab gene (170), followed by the S gene with 49 mutations, the N gene with 22 mutations, and the ORF10 gene had only two mutations (Table 3). Of the missense mutations, 71 were identified in ORF1ab, the longest ORF of SARS-CoV-2 covering roughly two-thirds of the entire genome. The S gene had 31 missense mutations. Similarly, most (90) synonymous mutations were found in the ORF1ab, followed by 12 synonymous mutations located in the S region.

Mutations with a frequency of more than 10 sequences are listed in Table 4. The most frequent mutations identified were the missense mutation D614G in the S gene found in 61 sequences and the synonymous mutation L4715L in the ORF1ab gene observed in 60 sequences. In addition, a synonymous mutation F53F in the N

Table 1
Demography and clinical characteristics of patients according to Pangolin nomenclature.

Variables	Modality	No. n = 74	Pangolin lineages			
			A.23.1 n = 9	B.1.214.2 n = 29	B.1351 n = 2	B.1.617.2 n = 34
Sex	Female	17 (23)	3 (33.3)	19 (65.5)	0 (0)	8 (23.5)
	Male	57 (77)	6 (66.7)	10 (34.5)	2 (100)	26 (74.5)
Age (year)	10–30	12 (16.2)	2 (22.2)	5 (17.2)	1 (50)	11 (32.4)
	31–50	38 (51.3)	6 (66.7)	18 (62)	1 (50)	19 (55.9)
	51–70	16 (21.6)	1 (1.1)	4 (13.8)	0 (0)	3 (8.8)
	>70	8 (10.8)	0 (0)	2 (6.9)	0 (0)	1 (2.9)
	Patient condition	31 (41.9)	3 (33.3)	16 (55.2)	0 (0)	20 (58.8)
Chronic comorbidity	Asymptomatic					
	Mild	5 (6.7)	2 (22.2)	1 (3.5)	0 (0)	2 (5.9)
	Severe	38 (51.3)	4 (44.4)	12 (41.4)	2 (100)	12 (35.3)
Sampling site	Hypertension	21 (28.3)	1 (1.1)	6 (20.7)	0 (0)	8 (23.5)
	Diabetes	33 (44.6)	4 (44.5)	15 (51.2)	1 (50)	14 (41.2)
	Cardiovascular disease	9 (12.1)	1 (1.1)	2 (6.9)	1 (50)	2 (5.9)
	Liver disease	1 (1.3)	2 (22.2)	3 (10.3)	0 (0)	3 (8.8)
	Others	10 (13.5)	1 (1.1)	2 (6.9)	0 (0)	7 (20.6)
Sampling site	Hospitals	26 (35.1)	5 (55.6)	20 (68.9)	2 (100)	17 (50)
	Ruzizi boarder	39 (52.7)	2 (22.2)	6 (20.7)	0 (0)	9 (26.5)
	Community	9 (12.1)	2 (22.2)	3 (10.3)	0 (0)	8 (23.5)

Table 2
Distribution of SARS-CoV-2 Pango lineages among our sequences.

Pango lineage	Sequences, n (%)	Most common country	Wave
A.23.1	5 (6.7)	Canada, UK, Uganda, Rwanda, USA	2
AY.33	10 (13.5)	Germany, Denmark, Belgium, Sweden, USA	3
AY.39	2 (2.7)	USA, UK, Denmark, Sweden	3
B.1	3 (4)	USA, UK, Turkey, France, Canada	1
B.1.2	1 (1.3)	USA, Canada, Mexico, Germany, Denmark	2
B.1.214	1 (1.3)	DRC, Germany, Uganda, USA, Republic of the Congo	2
B.1.214.2	26 (35.1)	Belgium, Switzerland, France, USA, Republic of the Congo	1,2
B.1.214.3	2 (2.7)	Belgium, Switzerland, France, Mayotte, Angola	2
B.1.223	1 (1.3)	UK, Spain, USA, Norway, Netherlands	1
B.351	2 (2.7)	South Africa, Philippines, USA, Sweden, Germany	2
B.1.617.2	19 (25.7)	USA, UK, Denmark, Germany, India	3
C.16	1 (1.3)	France, Angola, Portugal, USA, Switzerland	1

DRC, Democratic Republic of the Congo; UK, United Kingdom; USA, United States of America.

Table 3
Number of mutations observed in coding regions.

Genome region	Mutation types			Total
	Missense	Synonymous	Insertion/deletion	
ORF1ab	71	90	9	170
S	31	12	6	49
E	2	1	1	4
ORF3a	13	2	1	16
ORF7a	4	2	0	6
ORF8	3	2	4	9
ORF10	1	1	0	2
M	4	7	0	11
N	12	6	4	22
Total	141	123	25	289

gene was found in 32 sequences, and missense mutations S26L and T120I in ORF3a and ORF7a, respectively, were found in 27 sequences. A missense mutation I82T and a synonymous mutation V75V (225C>T) were the most common in genes M and E, respectively.

Discussion

A genome constitutes the molecular architecture of any living organism and encodes its phenotypic and genotypic makeup.

Mutations in the SARS-CoV-2 genome play a vital role in the virus genome transmission, replication efficiency, pathogenicity, and virulence of its human host (Maitra et al., 2020). This study describes molecular characteristics and analysis of the mutation pattern of full-length SARS-CoV-2 strains responsible for COVID-19 epidemics in the South Kivu province, DRC, from December 2020 to July 2021. We have elucidated the circulating variants and possible resultant effects of mutations identified in infected individuals.

In this study, Pango analysis of 74 samples clustered the SARS-CoV-2 sequences in four main Pangolin lineages, with the most prevalent being the B.1.617.2 primarily identified in young male individuals, as well as A.23.1 and B.1.351 (variant of concern). Previous studies conducted in Rwanda and Uganda (Bugembe et al., 2021) reported the sublineage A.23.1 that is currently dominating and detection of lineage B.1.315 among incoming travelers.

Our analyses show that most tested samples were from patients with diabetes mellitus, and about a quarter was from hypertensive individuals. These findings corroborate previous studies suggesting that comorbidities play a significant role in COVID-19 disease severity (Goyal et al., 2020; Richardson et al., 2020). The B.1.617.2 was primarily detected in patients with severe symptoms, and most were admitted to hospitals. This finding is in line with previous studies that demonstrated that patients with the Delta variant of SARS-CoV-2 are more than twice as likely to be admitted

Table 4
Mutations with a recurrence of more than 10 sequences.

Genome region	Nucleus position and change	AA position and change	Type of mutation	Number of samples	
5'UTR	25C>T			68	
	56G>T			30	
ORF1ab	2772C>T	F924F	Synonymous	42	
	3916G>T	A1306S	Missense	28	
	14143C>T	L4715L	Synonymous	60	
	6137C>T	P2046L	Missense	20	
	6276C>T	H2092H	Synonymous	15	
	6859C>T	P2287S	Missense	30	
	8721C>T	D2907D	Synonymous	28	
	8788G>T	V2930L	Missense	28	
	9764C>T	T3255I	Missense	20	
	10712C>T	A3571V	Missense	18	
	10936A>G	T3646A	Missense	17	
	11067A>G	V3689V	Synonymous	25	
	15186G>A	A5062A	Synonymous	28	
	16201C>T	H5401Y	Missense	31	
	18955C>T	L6319L	Synonymous	28	
	9764C>T	T3255I	Missense	24	
	9387G>T	M3119I	Missense	11	
	540C>T	G180G	Synonymous	26	
	4192A>G	I1398V	Missense	26	
	4626C>T	T1542T	Synonymous	30	
	5642C>T	T1881I	Missense	28	
	11685C>T	H3895H	Synonymous	31	
	12047C>T	A4016V	Missense	31	
	12911C>T	T4304I	Missense	29	
	14922C>T	L4974L	Synonymous	32	
	16745C>T	S5782F	Missense	31	
	2220C>T	I740I	Synonymous	29	
	6276C>T	H2092H	Synonymous	15	
	S gene	1841A>G	D614G	Missense	61
		1355T>G	L452R	Missense	17
		1433C>A	T478K	Missense	17
		2042C>G	P681R	Missense	31
		2848G>A	D950N	Missense	31
1240C>A		Q414K	Missense	24	
N gene	2147C>T	T716I	Missense	30	
	188A>G	D63G	Missense	23	
	608G>T	R203K	Missense	29	
	643G>T	G215C	Missense	27	
	1129G>T	D377Y	Missense	28	
	159C>T	F53F	Synonymous	32	
ORF3a	573C>T	R191R	Synonymous	27	
	614C>T	T205I	Missense	28	
	77C>T	S26L	Missense	27	
	514G>T	G172C	Missense	19	
	ORF7a	245T>C	V82A	Missense	21
		359C>T	T120I	Missense	27
183T>C		T61T	Synonymous	26	
M gene	245T>C	I82T	Missense	31	
E gene	225C>T	V75V	Synonymous	14	
3'UTR					
	4358G>T			33	

UTR, untranslated region.

to hospital than those infected with the Alpha or Gamma variants (Wise, 2021; Twohig et al., 2022).

The Pango lineage classification was used to identify the major lineages circulating in South Kivu because this nomenclature system is designed to combine both genetic and geographical information about SARS-CoV-2 dynamics.

Thus, the comparisons based on phylogenetic analysis entailed 399 genomes retrieved from the GISAID database. Our analysis revealed the circulation of two main variants that include Beta

(B.1.351) and Delta (B.1.617.2) variants, with its two sublineages (AY.33 and AY.39) responsible for most COVID-19 outbreaks during the third wave (May to July end, 2021). The third wave was characterized by an increase in COVID-19 cases and deaths in the DRC, particularly in South Kivu, compared with the previous waves, where an average of 320 cases and 10 deaths were recorded daily. The Delta variant reportedly originated from the state of Maharashtra in late 2020 and spread throughout India (Milcochova et al., 2021).

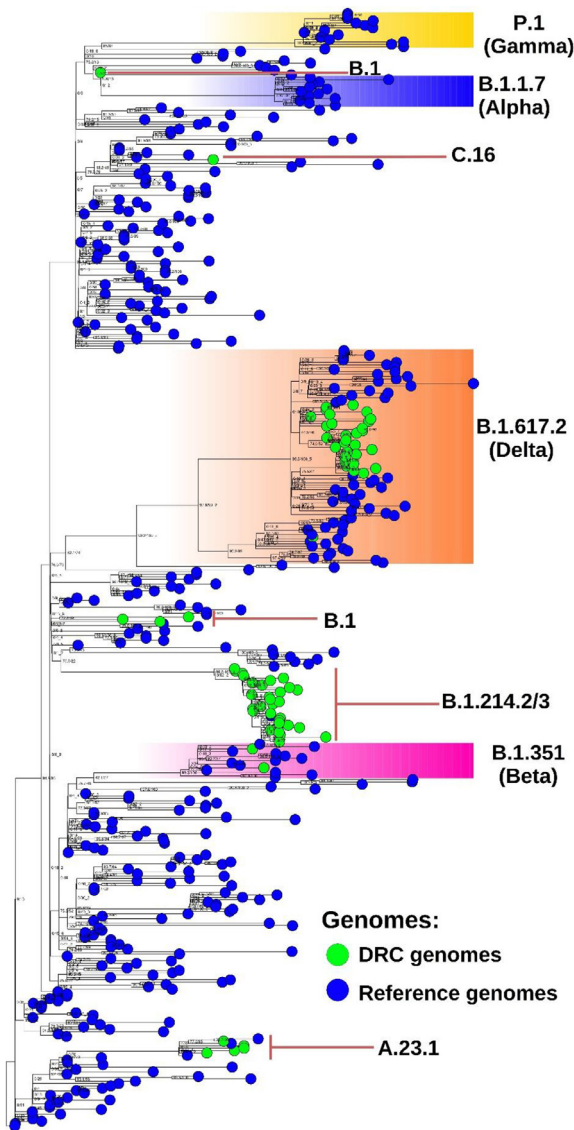


Figure 1. Phylogenetic tree comparing the 74 Congolese SARS-CoV-2 sequences with 399 previously published whole-genome sequences. Sequences are represented as rectangular tips, with genomes generated from this study in dark-green color and reference genomes in dark blue. The internal nodes display bootstrap values generated in IQ-TREE v2. 2.0 based on ultra-fast bootstrapping (UFBoot 10,000 replicates) and single-branch support (SH-aLRT): (UFBoot%/SH-aLRT%). The phylogenetic tree was generated with Wuhan/WH01 as the root sequence. The lineages (variants) A.23.1, B.1.1.7, B.1.214.*, B.1.617.2, B.1.351, C.16, and P.1 are labeled. Delta sublineages are highlighted with light-brown background, the Gamma in yellow background, and the Beta sublineages are highlighted with a pink background. Branch length indicates the number of mutations in reference to Wuhan-Hu-1/2019 genome. DRC, Democratic Republic of the Congo.

The Delta variant spread throughout the DRC and led to more severe disease symptoms in the local population with many hospitalization cases. This is in line with previous studies showing the Delta variant to be the most dominant strain worldwide, more transmissible, and presenting a higher risk of hospitalization and mortality (Callaway, 2021). The Delta variant was declared a variant of concern because of this characteristic .

In addition, the Beta (B.1.351) variant was also responsible for outbreaks during the first and second waves in South Kivu. The presence of the Beta variant was first reported in India in December 2020 before it spread in Africa, particularly in South Africa, with 8,664,388 cases reported in the continent. This variant was associated with in-hospital mortality in the second wave

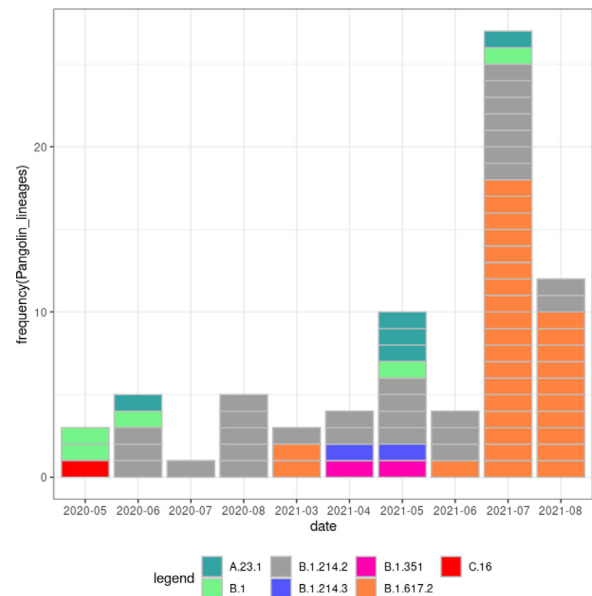


Figure 2. Frequency of Pangolin lineages between May 2020 and August 2021 in eastern Democratic Republic of the Congo.

in South Africa (Tegally et al., 2021). Beta is also a variant of concern and a highly transmissible SARS-CoV-2 variant (Yadav et al., 2021).

Two other variants of interest, such as A.23.1 and B.1.214, were also identified in this study. These variants were responsible for most outbreaks in neighboring countries such as the Republic of Rwanda () and Uganda (Bugembe et al., 2021), emphasizing the role of neighboring countries in the spread of SARS-CoV-2. Several Delta variant infection cases were from travelers crossing the Ruzizi border, the main point of entry in South Kivu. These findings suggest multiple introductions of SARS-CoV-2 strains into South Kivu, eastern DRC, mainly through international travelers, and therefore require critical surveillance in the disease management strategies.

An increased frequency of mutations in the SARS-CoV-2 genome has been reported throughout the pandemic, with more than 3000 unique point mutations discovered worldwide (Flores-Alanis et al., 2021). This study showed that in the South Kivu population, ORF1ab polyprotein, S glycoprotein, N phosphoprotein, ORF3a, ORF7a, M glycoprotein, and E protein were the mutation targets. The frequent mutations in these regions may be contributing to the adaptability of SARS-CoV-2 in the human host and may lead to observed variations in the pathogenicity of the virus. This agrees with a previous study reporting recurrent mutations in the same SARS-CoV-2 genome targets in the African population (Omotoso et al., 2021). In addition, a similar study has shown that most mutations are found within the ORF1ab, S, and N genes of the SARS-CoV-2 genome (Wang et al., 2020).

Our study shows a higher number of missense mutations than synonymous mutations, and half of them were located in the ORF1ab gene followed by the S gene (49 missense mutations). A similar observation was reported in studies conducted in Colombia and Egypt (Bermúdez-Forero et al., 2021; Sahin et al., 2021; Zekri et al., 2021; Maitra et al., 2020). Interestingly, the L4715L mutation found in 60 SARS-CoV-2 sequences from this study corroborates with previous findings reporting the P4715L mutation as a potential antiviral drug target (Pachetti et al., 2020).

Mutation analysis further showed many samples with D614G mutation in the S glycoprotein (1841A>G), flagging this position as a SARS-CoV-2 mutation hotspot. A previous study has reported

the same mutation in a high number of sequences in Turkey, and it has gradually become dominant in the world (Sahin et al., 2021). In addition, the D614G mutation in the SARS-CoV-2 S protein has been associated with increased infectivity and transmissibility (Zhang et al., 2020).

N gene encoding the nucleocapsid protein plays an important role in RNA packaging and the release of viral particles (Zeng et al., 2020). The most frequent mutation in this gene reported globally is the R203K, but it was shown not to affect the protein function. However, the R203K mutation was not found in our study; instead, we identified the R203M (608G>T) mutation, whose functional impact is still unknown.

Several insertions/deletion mutations were in some of the sequences in this study, even though further confirmation is required. However, these mutations would imply modification of the gene (protein) of concern and, therefore, may affect the adaptability and or pathogenicity of the virus. Other point mutations that have not been reported before were also identified in many sequences. We can speculate a correlation between these genetic changes and virus adaptation because these mutations are never fixed in the population.

Conclusion

This study analyzed SARS-CoV-2 whole genomes from 74 Congolese samples during the first, second, and third waves. Two lineages, including Delta (B.1.617.2) and Beta (B.1.351), were identified to be circulating with several mutations in coding genes. Delta was the most abundant of these lineages, especially during the third wave, causing most COVID-19 deaths in South Kivu from May 2021 to the end of July 2021. We cannot assume the possibility that the simultaneous circulation of these two SARS-CoV-2 variants could have exacerbated the transmissibility potential of the virus and thus caused devastating infections in this densely populated region. Our results further highlight a need for regional collaborations for a durable response toward combating COVID-19.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding sources

This work was fully supported by a grant from the German Institute for Medical Mission (DIFÄM), through the project Université Evangélique en Afrique (UEA)- DIFÄM titled “Whole-Genome sequencing of SARS-CoV-2” under the grant no. 2021-651-CD.

Ethical approval

The study protocol was approved by the Interdisciplinary Center of Ethical Research (CIRE) of the Université Evangélique en Afrique (UEA), Ref: CNES 001/DPSK/115PP/2021. Informed consent was obtained from all participants after being signed, and the confidentiality of the data was ensured.

Acknowledgments

We acknowledge the Université Evangélique en Afrique (UEA) through the Bioscience unit and German Institute for Medical Mission (DIFÄM) for the financial support. Our sincere appreciation goes further to the Provincial Ministry of Health and all the health structure leaders to facilitate access to samples and data. We also appreciate the lab technicians of the Molecular Biology laboratory of UEA for their technical support.

Author contributions

Patrick Bisimwa Ntagereka: study's design, data collection, drafting the manuscript; Samuel O. Oyola: sequence analysis and drafting the manuscript; Simon Baenyi Patrick: data collection and reviewing the manuscript; Gilbert Kibet Rono: study design and sequence analysis; Ahadi Bwihangane Birindwa: study design and reviewing the manuscript; Dieudonné Wasso Shukuru: study conceptualization and data analysis; Tshass Chasinga Baharanyi: drafting the initial manuscript and reviewing the manuscript; Théophile Mitima Kashosi: data collection, drafting the manuscript; Jean-Paul Chikwanine Buhendwa: study design, and data collection; Parvine Basimane Bisimwa: study design and reviewing the manuscript; Aline Byabene Kusinza: data collection design and reviewing the manuscript; Rodrigue Ayagirwe Basengere: data collection supervision and reviewing the manuscript; Denis Mukwege: study supervision and reviewing the manuscript. All authors approved the manuscript.

Data and materials availability

The sequences of complete viral genomes are available online through GISAID (<http://gisaid.org>) and can be found in the Supplementary Material Table 2. The datasets generated and analyzed during this study are also available from the corresponding author on request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijid.2022.05.041.

References

- Aksamentov I, Neher R, Nextclade. <https://clades.nextstrain.org>, XXX (accessed 13 September 2021).
- Bar-On YM, Flamholz A, Phillips R, Milo R. SARS-CoV-2 (COVID-19) by the numbers. *Elife* 2020;9:e57309 doi:10.7554/eLife.57309.
- Belouzard S, Millet JK, Licitra BN, Whittaker GR. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses* 2012;4:1011–33.
- Bermúdez-Forero M, Soto-Viáfara JA, Gardeazábal-Acuña PA, Anzola-Samudio DA, García-Otálora MA, et al. Effect of the first year of COVID-19 pandemic on the collection and use of blood components in Colombia monitored through the national haemovigilance system. *Transfus Med* 2021;31:421–30. doi:10.1111/tme.12828.
- Bugembe DL, Phan MV, Ssewanyana I, Semanda P, Nansumba H, Dhaala B, Nabadda S, O'Toole AN, Rambaut A, Kaleebu P, Cotten M. A SARS-CoV-2 lineage A variant (A. 23.1) with altered spike has emerged and is dominating the current Uganda epidemic. medRxiv. 2021. <https://www.medrxiv.org/content/10.1101/2021.02.08.21251393v2> (accessed on 23 March 2021).
- Butera Y, Mukantwari E, Artesi M, Umuringa JD, ÁN O'Toole, Hill V, et al. Genomic sequencing of SARS-CoV-2 in Rwanda reveals the importance of incoming travelers on lineage diversity. *Nat Commun* 2020;12:5705.
- Callaway E. Delta coronavirus variant: scientists brace for impact. *Nature* 2021;595(7865):17–18 PMID:34158664. doi:10.1038/d41586-021-01696-3.
- Flores-Alanis A, Cruz-Rangel A, Rodríguez-Gómez F, et al. Molecular epidemiology surveillance of SARS-CoV-2: mutations and genetic diversity one year after emerging. *Pathogens* 2021;10:184.
- Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, et al. Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* 2009;324:1557–61.
- Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet* 2018;19:9–20.
- Gohl DM, Garbe J, Grady P, et al. A rapid, cost-effective tiled amplicon method for sequencing SARS-CoV-2. *BMC Genomics* 2020;21:863.
- Goyal P, Choi JJ, Pinheiro LC, Schenck EJ, Chen R, Jabri A, et al. Clinical characteristics of Covid-19 in New York City. *N Engl J Med* 2020;382:2372–4.
- Grubaugh ND, Gangavarapu K, Quick J, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol* 2019;20(8).
- Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018;34:4121–3.
- Hamed SM, Elkhatib WF, Khairalla AS, Noreddin AM. Global dynamics of SARS-CoV-2 clades and their relation to COVID-19 epidemiology. *Sci Rep* 2021;11:8435.
- Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First case of 2019 novel coronavirus in the United States. *N Engl J Med* 2020;382:929–36.

- Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2020;395:565–74.
- Maitra A, Sarkar MC, Raheja H, Biswas NK, Chakraborti S, Singh AK, et al. Mutations in SARS-CoV-2 viral RNA identified in Eastern India: possible implications for the ongoing outbreak in India and impact on viral structure and host susceptibility. *J Biosci* 2020;45:76.
- Mlcochova P, Kemp SA, Dhar MS, Papa G, Meng B, Ferreira IA, et al. SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* 2021;599:114–19.
- Omotoso OE, Babalola AD, Matareek A. Mutational hotspots and conserved domains of SARS-CoV-2 genome in African population. *Beni Suef Univ J Basic Appl Sci* 2021;10:11.
- Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med* 2020;18:179.
- Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. *JAMA* 2020;323:2052–9.
- Ruzzenenti G, Maloberti A, Giani V, Biolcati M, Leidi F, Monticelli M, et al. Covid and cardiovascular diseases: direct and indirect damages and future perspective. *High Blood Press Cardiovasc Prev* 2021;28:439–45.
- Sahin E, Bozdayi G, Yigit S, Muftah H, Dizbay M, Tunccan OG, Fidan I, Caglar K. Genomic characterization of SARS-CoV-2 isolates from patients in Turkey reveals the presence of novel mutations in spike and nsp12 proteins. *J Med Virol* 2021;93:6016–26.
- Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, et al. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immunol* 2020;17:613–20.
- Tegally H, Wilkinson E, Christian L, et al. Rapid replacement of the Beta variant by the Delta variant in South Africa. medRxiv. 2021. <https://www.medrxiv.org/content/10.1101/2021.09.23.21264018v1> (accessed 17, September 2021).
- Twohig KA, Nyberg T, Zaidi A, et al. Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. *Lancet Infect Dis* 2022;22:35–42.
- Wang R, Chen J, Gao K, Hozumi Y, Yin C, Wei Wei G, et al. Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants. *Commun Biol* 2021;4:228. doi:10.1038/s42003-021-01754-6.
- Wise J. Covid-19: delta variant doubles risk of hospital admission compared with alpha variant, study shows. *BMJ* 2021;374:n2152.
- Yadav PD, P Sarkale, A Razdan, N Gupta, Nyayanit DN, Sahay RN, Potdar V, Patil DY, Baradkar S, Kumar A, Aggarwal N, Shete AM, Kaur H, Mohandas S, et al. Isolation and characterization of SARS-CoV-2 Beta variant from UAE travelers. *J Infect Public Health* 2022;15:182–6.
- Zekri AN, Bahnasy AA, Hafez MM, Hassan ZK, Ahmed OS, Soliman HK, El-Sisi ER, Dine MHSE, Solimane MS, Latife LSA, Seadawy MG, Elsafty AS, Abouelhoda M. Characterization of the SARS-CoV-2 genomes in Egypt in first and second waves of infection. *Sci Rep* 2021;11:21632.
- Zhang L, Jackson CB, Mou H, Ojha A, Peng H, Quinlan BD, Rangarajan ES, Pan A, Vanderheiden A, Suthar MS, Li W, Izard T, Rader C, Farzan M, Choe H. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat Commun* 2020;11:6013.
- Zeng W, Liu G, Ma H, Zhao D, Yang Y, Liu M, et al. Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochem Biophys Res Commun* 2020;527(3):618–23 30. doi:10.1016/j.bbrc.2020.04.136.
- Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, Yu J, Kang M, Song Y, Xia J, Guo Q, Song T, He J, Yen HL, Peiris M, Wu J. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N Engl J Med* 2020;382:1177–9.