

# 1 **Transmission networks of SARS-CoV-2 in coastal Kenya during the first two waves: a** 2 **retrospective genomic study**

3  
4 Charles N. Agoti<sup>1,2,#</sup>, Lynette Isabella Ochola-Oyier<sup>1</sup>, Khadija Said Mohammed<sup>1</sup>, Arnold W.  
5 Lambisia<sup>1</sup>, Zaydah R. de Laurent<sup>1</sup>, John M. Morobe<sup>1</sup>, Maureen W. Mburu<sup>1</sup>, Donwilliams O.  
6 Omuoyo<sup>1</sup>, Edidah M. Onger<sup>1</sup>, Leonard Ndwiga<sup>1</sup>, Eric Maitha<sup>3</sup>, Benson Kitole<sup>3</sup>, Thani  
7 Suleiman<sup>3</sup>, Mohamed Mwakinangu<sup>3</sup>, John Nyambu<sup>3</sup>, John Otieno<sup>3</sup>, Barke Salim<sup>3</sup>, Jennifer  
8 Musyoki<sup>1</sup>, Nickson Murunga<sup>1</sup>, Edward Otieno<sup>1</sup>, John Kiiru<sup>3</sup>, Kadondi Kasera<sup>3</sup>, Patrick  
9 Amoth<sup>3</sup>, Mercy Mwangangi<sup>3</sup>, Rashid Aman<sup>3</sup>, Samson Kinyanjui<sup>1,2,4</sup>, George Warimwe<sup>1,4</sup>, My  
10 Phan<sup>5</sup>, Ambrose Agweyu<sup>1</sup>, Matthew Cotten<sup>5,6</sup>, Edwine Barasa<sup>1</sup>, Benjamin Tsofa<sup>1</sup>, D. James  
11 Nokes<sup>1,7</sup>, Philip Bejon<sup>1,4</sup>, George Githinji<sup>1</sup>

## 12 13 **Affiliations**

- 14 1. Kenya Medical Research Institute (KEMRI)-Wellcome Trust Research Programme  
15 (KWTRP), Kilifi, Kenya.
- 16 2. Pwani University, Kilifi, Kenya.
- 17 3. Ministry of Health, Kenya.
- 18 4. Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.
- 19 5. Medical Research Centre (MRC)/ Uganda Virus Research Institute, Entebbe, Uganda.
- 20 6. MRC-University of Glasgow Centre for Virus Research, Glasgow, UK.
- 21 7. University of Warwick, Coventry, United Kingdom.

22  
23 **#Corresponding author:** Email: [cnyaigoti@kemri-wellcome.org](mailto:cnyaigoti@kemri-wellcome.org)

## 24 25 **ABSTRACT**

26 **Background:** The transmission networks of SARS-CoV-2 in sub-Saharan Africa remain  
27 poorly understood.

28 **Methods:** We undertook phylogenetic analysis of 747 SARS-CoV-2 positive samples  
29 collected across six counties in coastal Kenya during the first two waves (March 2020 -  
30 February 2021). Viral imports and exports from the region were inferred using ancestral state  
31 reconstruction (ASR) approach.

32 **Results:** The genomes were classified into 35 Pango lineages, six of which accounted for 79%  
33 of the sequenced infections: B.1 (49%), B.1.535 (11%), B.1.530 (6%), B.1.549 (4%), B.1.333  
34 (4%) and B.1.1 (4%). Four identified lineages were Kenya specific. In a contemporaneous  
35 global subsample, 990 lineages were documented, 261 for Africa and 97 for Eastern Africa.  
36 ASR analysis identified >300 virus location transition events during the period, these  
37 comprising: 69 viral imports into Coastal Kenya; 93 viral exports from coastal Kenya; and 191  
38 inter-county import/export events. Most international viral imports (58%) and exports (92%)  
39 occurred through Mombasa City, a key touristic and commercial Coastal Kenya center; and  
40 many occurred prior to June 2020, when stringent local COVID-19 restriction measures were  
41 enforced. After this period, local virus transmission dominated, and distinct local phylogenies  
42 were seen.

43 **Conclusions:** Our analysis supports moving control strategies from a focus on international  
44 travel to local transmission.

45 **Funding:** This work was funded by Wellcome (grant#: 220985) and the National Institute for  
46 Health Research (NIHR), project references: 17/63/and 16/136/33 using UK aid from the UK  
47 Government to support global health research, The UK Foreign, Commonwealth and  
48 Development Office.

49  
50 **INTRODUCTION** This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

51 By 31<sup>st</sup> August 2021, at least 218 million cases of COVID-19 and > 4.5 million associated  
52 deaths were reported worldwide. By the same date, Kenya, an East Africa nation with a  
53 population of ~50 million had reported 235,863 cases of SARS-CoV-2 and 4,726 associated  
54 deaths (MOH, 2021). Serological surveys indicated that Kenya's COVID-19 epidemic had  
55 progressed further than could be discerned from the limited laboratory testing case reports that  
56 were available during the period (Etyang et al., 2021; Uyoga et al., 2021a). In the period after  
57 the peak of Kenya's second wave (January-March 2021), nationwide anti-SARS-CoV-2 IgG  
58 prevalence based on analysis of blood donor samples was estimated to be 48.5% (Adetifa et  
59 al., 2021; Uyoga et al., 2021b).

60  
61 In August 2021, 18 months after the first report of SARS-CoV-2 occurrence in Kenya, the  
62 country had experienced four major waves of SARS-CoV-2 infections (Brand et al., 2021).  
63 Despite this progression, the local SARS-CoV-2 spread patterns, like in many sub-Saharan  
64 Africa settings, remained poorly understood (Wilkinson et al.). The analysis presented here  
65 examined genome sequences from the first two waves in Coastal Kenya. Overall, the  
66 documented SARS-CoV-2 infections in the country during these early waves were  
67 concentrated in cities, especially Nairobi (~42%), the capital, and Mombasa (~8%), the  
68 country's second largest city located on the Indian ocean coast (**Figure 1A-C**). In addition to  
69 Mombasa city, there are five other counties that make up Kenya's Coastal region (Kilifi,  
70 Kwale, Taita Taveta, Tana River and Lamu) which has a total population size of ~4.3 million.

71  
72 Throughout the COVID-19 pandemic, genomic surveillance has been critical for tracking the  
73 spread of SARS-CoV-2 and investigating its evolutionary changes (Bugembe et al., 2020;  
74 Geoghegan et al., 2020; Oude Munnink et al., 2020; Worobey et al., 2020). In-country SARS-  
75 CoV-2 genome sequencing started in Kenya soon after the identification of the initial case and  
76 has remained a key tool informing public health response and tracking the progress of the local  
77 epidemic. We have previously reported on the genomic analysis of SARS-CoV-2 in Coastal  
78 Kenya during the early phase of the pandemic (up to July 2020) revealing introduction multiple  
79 lineages but with mainly Pango lineage B.1 establishing local transmission (Githinji et al.,  
80 2021). Here we investigated the lineage composition and diversity of the subsequent two  
81 waves (up to February 2021) and examine the patterns of further viral introductions, local  
82 evolution and spread of SARS-CoV-2 across the six Coastal counties of Kenya.

83

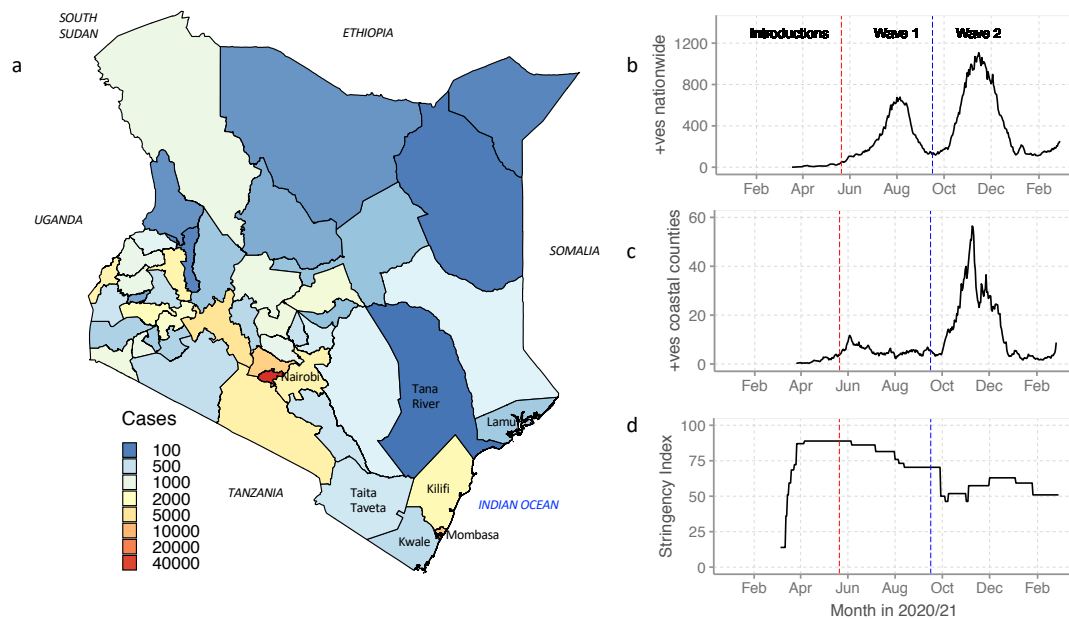
## 84 **RESULTS**

85

### 86 ***Infection waves in Coastal Kenya counties***

87 The first cases of COVID-19 in Coastal Kenya were reported in Mombasa and Kilifi counties  
88 in March 2020. This was after initial detection in Nairobi City on 13<sup>th</sup> March 2020. Following  
89 this, several COVID-19 related containment measures were put in place in March /April 2020  
90 to curtail the spread of the virus including ban of movement into and out of infected areas. The  
91 national Oxford Stringency index in Kenya in April 2020 was at 89 (Brand et al., 2021), **Figure**  
92 **1D**. However, two months later (June 2020), a steady removal of these restriction measures  
93 started leading to a SI of 75 by August 2020 and a SI <50 by October 2020. At the peak of the  
94 second wave, some of the restriction measures were reinstated increasing the SI to >=50,  
95 **Figure 1D**.

96



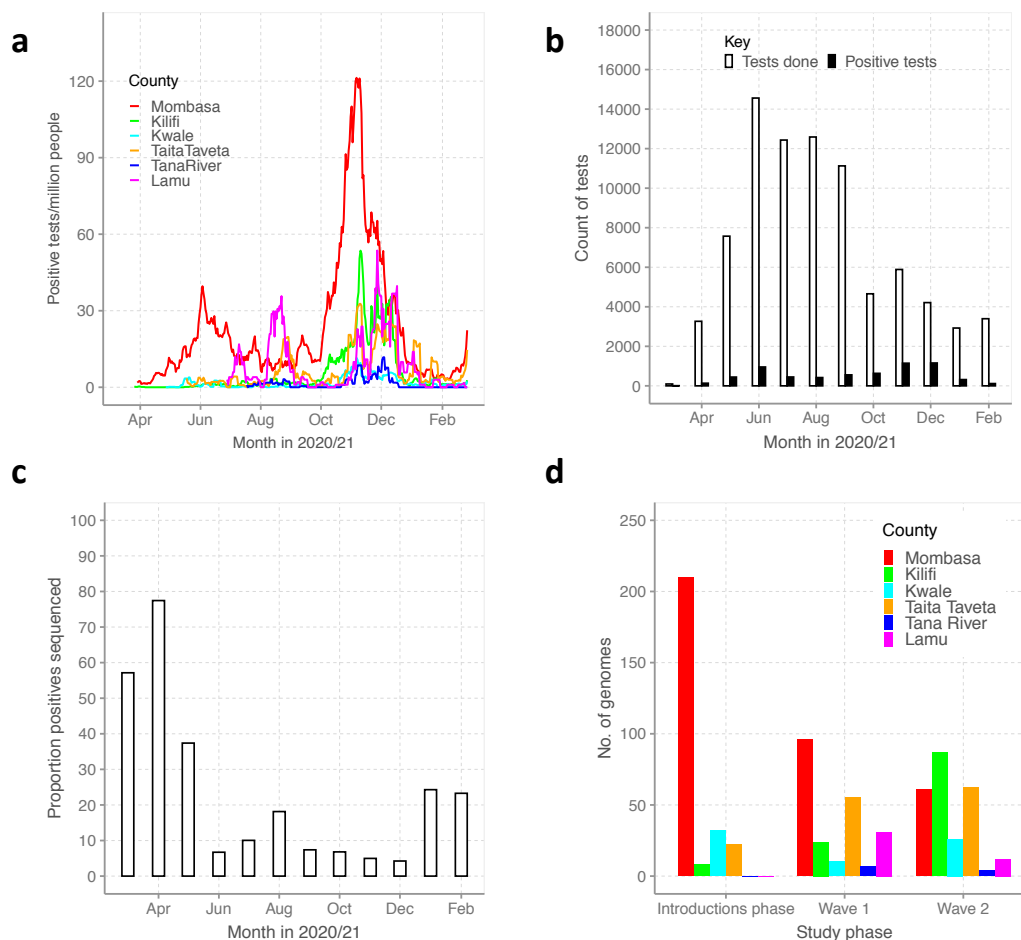
97  
 98 **Figure 1. The SARS-CoV-2 epidemic in Kenya.** Panel A, map showing administrative  
 99 boundaries of the 47 counties in Kenya and the neighboring countries. The map is colored by  
 100 the cumulative number of SARS-CoV-2 cases documented in each county as of 26<sup>th</sup> February  
 101 2021. Panel B, reported daily new cases, 7-day-rolling average, in Kenya from March 2020 to  
 102 February 2021 showing the first two national SARS-CoV-2 infection waves. Panel C, reported  
 103 daily cases, 7-day-rolling average, across coastal Kenya during the study period. Panel D,  
 104 Kenya government COVID-19 intervention levels during the study period as summarized by  
 105 the Oxford stringency index (abbreviated SI, see details in materials and methods section).

106  
 107 Mombasa county endured the earliest surge of SARS-CoV-2 cases in coastal Kenya with its  
 108 first peak occurring in June 2020, **Figure 2A**. This was earlier than the first national peak that  
 109 occurred in July-August of 2020. By 26<sup>th</sup> February 2021 (the most recent date included in our  
 110 analysis) three of the six coastal Kenya counties (i.e. Mombasa, Lamu and Taita Taveta) had  
 111 experienced at least two major infection peaks while the other three (i.e. Kilifi, Kwale and Tana  
 112 River) had experienced only a single peak. During this period covered, the Kenya Ministry of  
 113 Health (MoH) reported 12,307 laboratory confirmed SARS-CoV-2 cases for all the six Coastal  
 114 counties: Mombasa (n=8,450, 68.7%), Kilifi (n=2,458, 20.0%), Taita Taveta (n=567, 4.6%),  
 115 Kwale (n=437, 3.6%), Lamu (n=309, 2.5%) and Tana River (n=85, 0.7%), **Figure 2A**.

116  
 117  
 118 **Testing and sequencing at KWTRP**

119 The flow of samples analyzed in this study is provided in **Supplementary-Figure 1**. Between  
 120 17<sup>th</sup> March 2020 and 26<sup>th</sup> February 2021, the KEMRI-Wellcome Trust Research Programme  
 121 (KWTRP) tested a total of 82,716 nasopharyngeal/oropharyngeal (NP/OP) swab samples from  
 122 the six Coastal counties. Of these, 6,307 (7.6%) were determined as SARS-CoV-2 positive  
 123 (approximately half of the total reported by MoH), distributed by month as in **Figure 2B**.  
 124 Overall, a total of 3,137 (6.8%) tests at KWTRP were positive from samples taken in Mombasa,  
 125 1,142 (8.8%) for Kilifi, 657 (4.5%) for Taita Taveta, 349 (12.7%) for Lamu, 437 (7.9%) for  
 126 Kwale and 105 (12.0%) for Tana River.

127



128  
 129 **Figure 2. SARS-CoV-2 identified cases on the Kenyan Coast and testing and sequencing**  
 130 **at KEMRI-Wellcome Trust Research Programme (KWTRP).** Panel A, epidemic curves  
 131 for each of the six Coastal Kenya counties derived from the daily positive case numbers, 7-  
 132 day-rolling average, reported by the Ministry of Health. Panel B, monthly count of SARS-  
 133 CoV-2 RT-PCR tests undertaken at the KWTRP and those positive during the study period.  
 134 Panel C, monthly proportion of samples sequenced from total SARS-CoV-2 positives detected  
 135 at KWTRP. Panel D, Distribution of the sequenced 747 samples by study period and across the  
 136 6 Coastal Kenya counties.

137  
 138 Of the identified 3,137 RT-PCR positive samples, 747 (23.8%) were genome sequenced,  
 139 distributed as follows; Mombasa: 367 (11.7%), Kilifi: 119 (10.4%), Taita Taveta: 139 (21.2%),  
 140 Kwale: 68 (15.6%), Lamu: 43 (12.4%) and Tana River: 11 (10.5%), **Figure 2C**. The multiple  
 141 sequence alignment of these coastal Kenya genomes and observed differences from the  
 142 Wuhan-Hu-1 isolate is shown in **Supplementary-Figure 2**. The sequences were spread across  
 143 the study period with 272 (36.4%) from the introduction phase (17<sup>th</sup> March to 20<sup>th</sup> May 2020  
 144 when nationally <50 new cases were reported per day), 223 (29.9%) from wave one (21<sup>st</sup> May-  
 145 15<sup>th</sup> September 2020) and 252 (33.7%) from wave two (16<sup>th</sup> September 2020 to 26<sup>th</sup> February  
 146 2021), **Figure 1**. Overall, our sequencing rate corresponded to approximately one sequence for  
 147 every 16 confirmed cases.  
 148

149 Most of the sequenced isolates were from males (n=443, 59.3%), persons aged between 30-60  
150 years (n=420, 56.2%) and persons of Kenyan nationality (n=509, 68.1%), **Table 1**. The second  
151 largest nationality of the individuals providing sequenced samples was Tanzania (n=34)  
152 followed by South Africa (n=5). A total of 110 samples (14.7%) had been collected from  
153 persons with a recent international travel history (<14 days) but a considerable proportion of  
154 the sequenced samples were obtained from persons whose travel history information was  
155 missing (n=373, 49.9%).

156

### 157 *Lineage dynamics in Coastal Kenya*

158 As a reference, we adopted the PANGOLIN (Phylogenetic Assignment of Named Global  
159 LINeages) lineage nomenclature (Rambaut et al., 2020). The 747 genomes were classified into  
160 35 Pango lineages, including four first identified in Kenya (N.8, B.1.530, B.1.549 and  
161 B.1.596.1), **Table 2 and Figure 3A**. During the introductions phase we detected 14 lineages  
162 while during wave one we detected 21 lineages (10 of which were being identified locally for  
163 the first time) and during wave two we detected 19 lineages (seven of which were being  
164 identified locally for the first time), **Figure 3B**. Seventeen lineages (48.6%) were identified in  
165 three or more samples with the top six lineages accounting for 78.8% of the sequenced  
166 infections; B.1 (n=369, 49.4%, European lineage which predominated in the Northern Italy  
167 outbreak early in 2020), B.1.535 (n=82, Australian lineage), B.1.530 (n=42, 5.6%, Kenyan  
168 lineage), B.1.530 (n=42, 6%, Kenyan lineage), B.1.549 (n=33, 4.4%, Kenyan lineage), B.1.333  
169 (n=32, 4.3%, Washington USA lineage), B.1.1 (n=31, 4.1%, European lineage).

170

171 **Table 1:** Demographics of the sequenced SARS-CoV-2 cases in coastal Kenya

<b>Characteristic</b>	<b>Total cases (%)</b>	<b>Introductions phase</b>	<b>Wave 1</b>	<b>Wave 2</b>
<b>Age category</b>				
0-9	16 (2.1)	5	6	5
10-19	40 (5.4)	11	10	19
20-29	139 (18.6)	52	39	48
30-39	195 (26.1)	75	47	73
40-49	137 (18.3)	48	40	49
50-59	88 (11.8)	45	12	31
60-69	31 (4.1)	17	6	8
70-79	12 (1.6)	4	3	5
80+	4 (0.5)	3	-	1
Missing	85 (11.4)	12	60	13
<b>Gender</b>				
Female	203 (27.2)	81	43	79
Male	443 (59.3)	167	118	158
Missing	101 (13.5)	24	62	15
<b>Nationality</b>				
Kenyan	509 (68.1)	174	138	197
Tanzania	34 (4.5)	18	7	9
South Africa	5 (0.1)	0	0	5
Other <sup>#</sup>	20 (2.7)	3	4	13
Missing	179 (24.0)	77	74	28
<b>Travel History<sup>‡</sup></b>				
Yes	110 (14.7)	55	28	27
No	264 (35.3)	108	81	75
Missing	373 (49.9)	109	114	150

172 <sup>#</sup> The other nationalities included Ethiopia (4), Rwanda(2), Sweden (2), USA (2), Britain (2), India (2), Pakistan (1), Burundi (1), Poland (1), Eritrea (1),  
 173 Uganda (1) and Zambia (1) ;<sup>‡</sup> Defined as having moved into Kenya in the previous 14 days or sampled at a point of entry (POE) into Kenya.

174

**Table 2.** Lineages observed in coastal Kenya, their county distribution, global history and VOC/VOI status

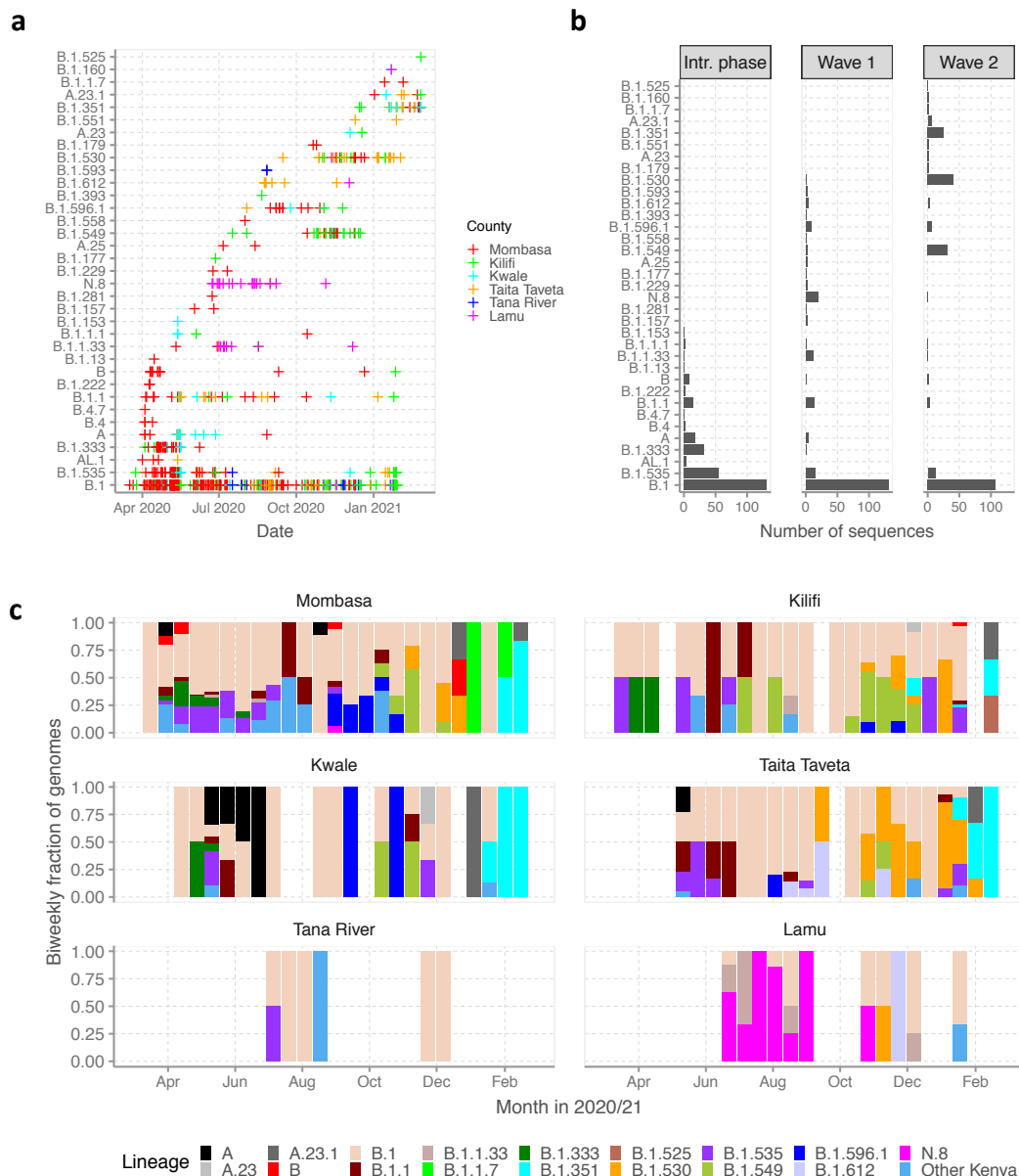
Lineage	Frequency (%)	Mombasa	Kilifi	Kwale	Taita Taveta	Tana River	Lamu	Earliest date	Number assigned	Description
<b>A</b>	22 ( 2.9 )	4	-	13	5	-	-	2019-12-30	2164	Root of the pandemic lies within lineage A. Many sequences originating from China
<b>A.23</b>	2 ( 0.3 )	-	1	1	-	-	-	2020-06-08	99	Ugandan lineage
<b>A.23.1</b>	6 ( 0.8 )	2	1	1	2	-	-	2020-10-21	1109	International lineage with a number of variants of potential biological concern, including 681R
<b>A.25</b>	2 ( 0.3 )	2	-	-	-	-	-	2020-06-08	46	Ugandan lineage
<b>AL.1</b>	4 ( 0.5 )	3	-	-	1	-	-	2020-02-23	1268	Alias of B.1.1.231.1, Canada lineage
<b>B</b>	11 ( 1.5 )	10	1	-	-	-	-	2019-12-24	5408	Second major haplotype (and first to be discovered)
<b>B.1</b>	369 ( 49.4 )	201	65	19	65	8	11	2020-01-24	71452	A large European lineage the origin of which roughly corresponds to the Northern Italian outbreak early in 2020.
<b>B.1.1</b>	31 ( 4.1 )	12	3	4	12	-	-	2020-02-16	42839	European lineage with 3 clear SNPs `28881GA`, `28882GA`, `28883GC`
<b>B.1.1.1</b>	4 ( 0.5 )	1	1	2	-	-	-	2020-03-02	2796	England
<b>B.1.1.33</b>	13 ( 1.7 )	2	1	-	-	-	10	2020-03-01	2039	Brazilian
<b>B.1.1.7</b>	2 ( 0.3 )	2	-	-	-	-	-	2020-02-07	986233	UK lineage of concern, associated with the N501Y mutation.
<b>B.1.13</b>	1 ( 0.1 )	1	-	-	-	-	-	2020-03-09	398	UK lineage
<b>B.1.153</b>	1 ( 0.1 )	-	-	1	-	-	-	2020-03-01	773	English lineage with some Australian, New Zealand and European sequences
<b>B.1.157</b>	2 ( 0.3 )	2	-	-	-	-	-	2020-02-03	324	UK/ Spanish lineage
<b>B.1.160</b>	2 ( 0.3 )	-	-	1	-	-	1	2020-02-06	24207	Large European lineage, sequences from BeNeLuX, Denmark, Switzerland, Hungary and UK.
<b>B.1.177</b>	1 ( 0.1 )	-	1	-	-	-	-	2020-02-14	69319	Major mostly European lineage, result of opening borders summer 2020
<b>B.1.179</b>	2 ( 0.3 )	2	-	-	-	-	-	2020-03-09	231	Denmark
<b>B.1.222</b>	2 ( 0.3 )	2	-	-	-	-	-	2020-02-24	562	Scottish lineage
<b>B.1.229</b>	2 ( 0.3 )	2	-	-	-	-	-	2020-03-19	104	UK lineage
<b>B.1.281</b>	1 ( 0.1 )	1	-	-	-	-	-	2020-04-08	41	Bahrain lineage
<b>B.1.333</b>	32 ( 4.3 )	27	2	3	-	-	-	2020-03-03	188	USA lineage WA
<b>B.1.351</b>	25 ( 3.3 )	6	4	8	7	-	-	2020-03-27	25467	Lineage of concern detected in South Africa
<b>B.1.393</b>	1 ( 0.1 )	-	1	-	-	-	-	2020-05-29	37	Uganda
<b>B.1.525</b>	1 ( 0.1 )	-	1	-	-	-	-	2020-12-11	6729	International lineage with E484K, del69-70 among other defining mutations.
<b>B.1.530</b>	42 ( 5.6 )	8	7	-	26	-	1	2020-10-20	54	Kenya
<b>B.1.535</b>	82 ( 11.0 )	50	11	10	10	1	-	2020-02-03	2573	Australia

<b>B.1.549</b>	33 ( 4.4 )	11	17	3	2	-	-	2020-07-17	51	Kenya + England
<b>B.1.551</b>	2 ( 0.3 )	-	-	-	2	-	-	2020-07-31	751	USA lineage
<b>B.1.558</b>	1 ( 0.1 )	1	-	-	-	-	-	2020-04-06	190	USA/ Mexico lineage
<b>B.1.593</b>	2 ( 0.3 )	-	-	-	-	2	-	2020-08-27	90	USA lineage
<b>B.1.596.1</b>	15 ( 2.0 )	10	2	2	1	-	-	2020-09-07	54	Kenyan lineage
<b>B.1.612</b>	7 ( 0.9 )	-	-	-	6	-	1	2020-03-11	380	USA lineage, split out from B.1.361
<b>B.4</b>	3 ( 0.4 )	3	-	-	-	-	-	2020-01-18	376	Iranian lineage and global exports
<b>B.4.7</b>	1 ( 0.1 )	1	-	-	-	-	-	2020-03-14	68	Africa and UAE
<b>N.8</b>	20 ( 2.7 )	1	-	-	-	-	19	2020-06-23	14	Alias of B.1.1.33.8, Kenyan

177  
178  
179



180 Lineage B.1 was the first to be detected in the region and comprised the initial cases identified  
 181 in Mombasa, Kilifi, Kwale and Tana River, **Figure 3A and C**. This lineage comprised 48.2%  
 182 (n=131) of the viruses we sequenced from the introduction phase, 58.7% (n=131) from wave  
 183 one and 42.5% (n=107) from wave two. Overall, B.1 accounted for most of the virus sequences  
 184 detected in all counties except Lamu where lineages B.1.1.33 and N.8 (alias lineage B.1.1.33.8)  
 185 were most frequent (together accounting for 67.5% of Lamu cases, especially wave one, first  
 186 detected there on 23<sup>rd</sup> June 2020), **Table 2**. Only three additional sequenced cases of lineage  
 187 B.1.1.33 occurred outside Lamu County (two Mombasa and one in Kilifi) and only one  
 188 additional sequenced case of lineage N.8 occurred outside Lamu County (in Mombasa).  
 189



190 **Figure 3. Lineage introductions and temporal dynamics in Coastal Kenya.** Panel A,  
 191 SARS-CoV-2 Pango lineages in the sequenced 747 coastal Kenya samples and timing of  
 192 detections. Panel B, Lineage detections by study phase. Intr. phase refers to introduction phase.  
 193 Panel C, Biweekly distribution of the common lineages identified across the six counties  
 194 presented as proportions. Lineages detected in less than three cases were put together and  
 195

196 referred to as “other lineages”. This group included 18 lineages namely: A.25, AL.1, B.1.1.1,  
197 B.1.13, B.1.153, B.1.157, B.1.160, B.1.177, B.1.179, B.1.222, B.1.229, B.1.281, B.1.393,  
198 B.1.551, B.1.558, B.1.593, B.4 and B.4.7.

199

200 During wave one, 18 cases of lineage A were observed. Most (56%) of these were identified  
201 in Kwale county (one of the Kenya counties bordering Tanzania, **Figure 1A**), and the lineage  
202 comprised 19% of all cases identified Kwale county, **Figure 3C**. Only two other counties  
203 observed lineage A cases: Mombasa (n=4) and Taita Taveta (n=5). Lineage B.1.535, which  
204 was the second most common lineage in the region was mostly observed in during wave one  
205 and predominantly in Mombasa County with limited number of cases observed in Kilifi, Kwale  
206 and Taita Taveta, **Table 2**.

207

208 During the wave two period, there was an emergence of three lineages first identified in Kenya:  
209 (a) lineage B.1.549 (first observed 17<sup>th</sup> July 2020) which was responsible for a significant  
210 number of the sequenced infections in Kilifi (n=17), Mombasa (n=11), Taita Taveta (n=2), and  
211 Kwale (n=3), (b) lineage B.1.596.1 (first observed 03<sup>rd</sup> August 2020) and detected in  
212 Mombasa, Kilifi, Kwale and Taita-Taveta and (c) lineage B.1.530 (first observed 15<sup>th</sup>  
213 September 2020) that predominated infections in Taita Taveta (n=26) with a fewer cases  
214 observed in Mombasa (n=8), Kilifi (n=7) and Lamu (n=1), **Figure 3A and B**.

215

216 Two variants of concern (VOC) were detected during the study period: Beta (i.e. lineage  
217 B.1.351 in 25 samples); and Alpha (i.e lineage B.1.1.7 in two samples (Konings et al., 2021)).  
218 The earliest B.1.351 infections were identified in international travelers from South Africa in  
219 mid-December 2020 while the earliest detected B.1.1.7 case was in second week of January  
220 2020 in a local with no history of recent international travel. We also detected two cases of  
221 A.23 and six cases of A.23.1 after the peak of wave two (both lineages first identified in  
222 Uganda, and A.23.1 was considered a variant of interest (VOI) in the region in early 2021  
223 (Bugembe et al., 2021a)).

224

### 225 *Lineage dynamics with a widening scale of observation.*

226 We compared the temporal prevalence patterns of the lineages identified in Coastal Kenya to  
227 Eastern Africa, Africa and a global sub-sample, **Figure 4**. In these comparative genome sets,  
228 we identified 97 Pango lineages for Eastern Africa, 261 for Africa and 990 for the global sub-  
229 sample. Note that by May 2021, a total of 1,293 Pango lineages has been assigned worldwide  
230 since the beginning of the pandemic (O’Toole et al., 2021) indicating very high  
231 representativeness in our global subsample.

232

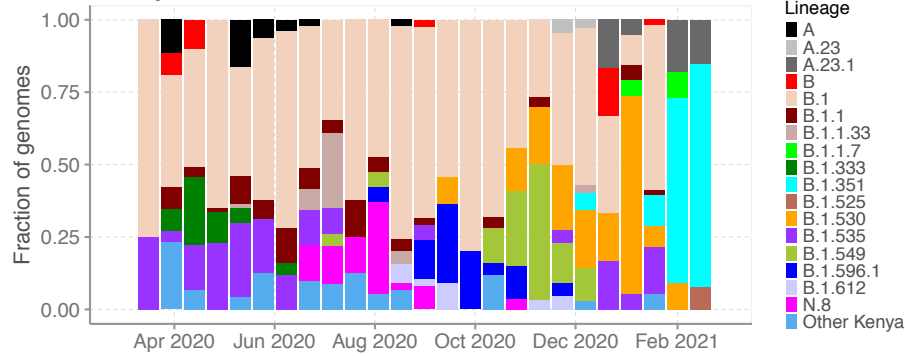
233 In our Eastern Africa sample (n=2,261), the top lineages were also observed in the Coastal  
234 Kenya data (B.1.351, A.23.1, B.1.1 and B.1) except lineage B.1.351.2 (n=180, sub-lineage of  
235 B.1.351 circulating in Mayotte) and B.1.380 (n=128, a Rwanda lineage (Butera et al., 2021)),  
236 **Figure 4B**. In the Africa sample (n=12,191), four of the six top lineages (B.1, B.1.1, B.1.351  
237 and B.1.1.7) were also identified in Coastal Kenya. The undetected two lineages were  
238 B.1.1.448 (n=392) mainly observed South Africa samples, and B.1.351.2, **Figure 4C**. In the  
239 global sub-sample (n=24,511) four of the top six lineages (B.1, B.1.1, B.1.1.7 and B.1.351)  
240 were observed in the samples from Coastal Kenya. The unobserved lineages were D.2 (Alias  
241 of B.1.1.25.2, an Australian lineage) and B.1.2 (a USA lineage), **Figure 4D**.

242

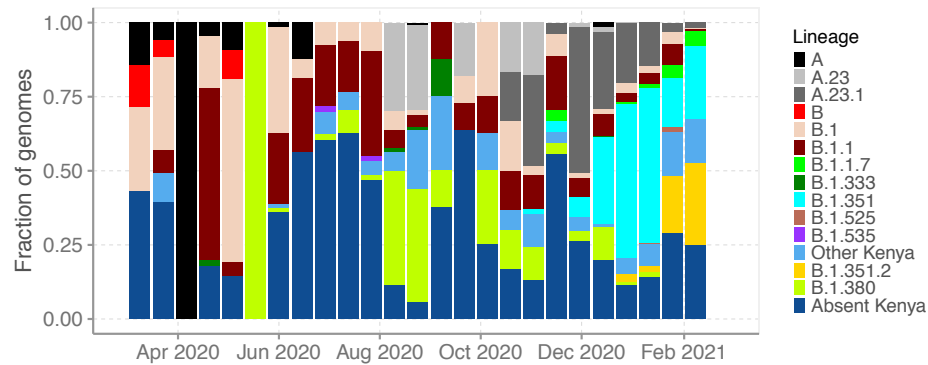
243 At the time of the second wave in Kenya, the Alpha and Beta VOC were already widespread  
244 across Eastern Africa and Africa but there were only sporadic detections from our Coastal  
245 Kenya dataset. Although in the comparison data lineage B.1 occurred in substantial proportions

246 across the different scales early in the pandemic, its prevalence diminished faster overtime in  
 247 the outside Coastal Kenya sample sets when compared to the Coastal Kenya dataset set.  
 248 Greater than 95% of the lineages comprising infections globally were not seen in the Coastal  
 249 Kenya samples.  
 250

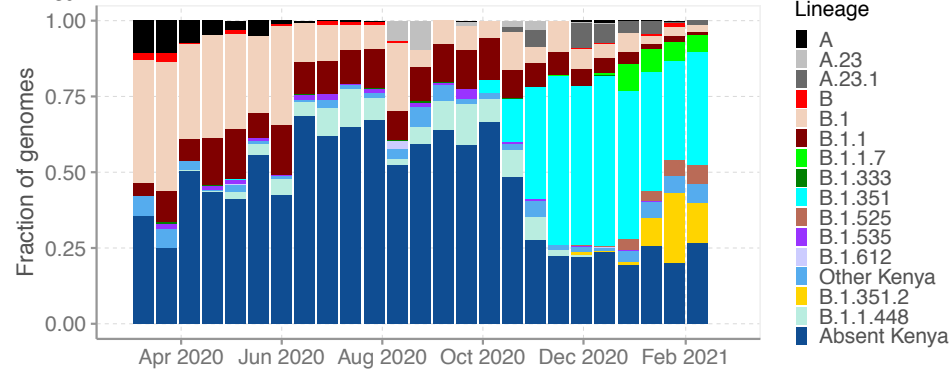
**a. Coastal Kenya**



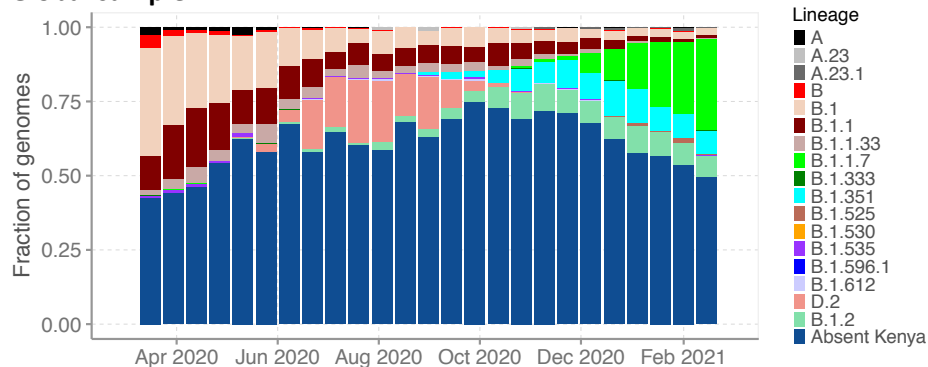
**b. Eastern Africa**



**c. Africa**



**d. Global sample**



251

252 **Figure 4. Lineage dynamics at widening scales of observation.** Panel A, Biweekly  
253 distribution of detected lineages in Coastal Kenya from the sequenced 747 genomes. Panel B,  
254 biweekly distribution of detected lineages in Eastern Africa from 2,261 contemporaneous  
255 genomes from 10 countries whose contemporaneous data were available in GISAID. Lineages  
256 not observed in the Coastal Kenya dataset have been lumped together as “Absent Kenya” but  
257 with the those appearing among top six and not observed in Kenya indicated just below the  
258 other Kenya group. Panel C, weekly distribution of detected lineages in Africa from 12,191  
259 contemporaneous genomes from 39 countries that were available in GISAID. Panel D, weekly  
260 distribution of detected lineages in a global sub-sample of 25,511 contemporaneous genomes  
261 from 139 countries that were compiled from GISAID (see detail in methods section).

262  
263

#### 264 ***SARS-CoV-2 diversity in Coastal Kenya***

265 We constructed both a mutational maximum likelihood (ML) (i.e. non-time resolved) and a  
266 time-resolved phylogeny of the Coastal Kenya genomes both while including a random global  
267 reference set of genomes (n=4000) shown in **Supplementary-Figure 3** and **Figure 5A**,  
268 respectively. These analysis showed that (i) the Coastal Kenya genomes were represented  
269 across many but not all of the major phylogenetic clusters observed globally, ii) some of the  
270 coastal Kenya clusters expanded after introduction whereas others did not (observed as  
271 singletons) and iii) all counties appeared to have had multiple lineage/variant introductions  
272 with some clusters comprising genomes detected across multiple counties, **Figure 5A**. There  
273 was considerable correlation of the root-to-tip genetic distance and the date of sampling dates  
274 ( $r^2=0.49$ ), **Figure 5B**.

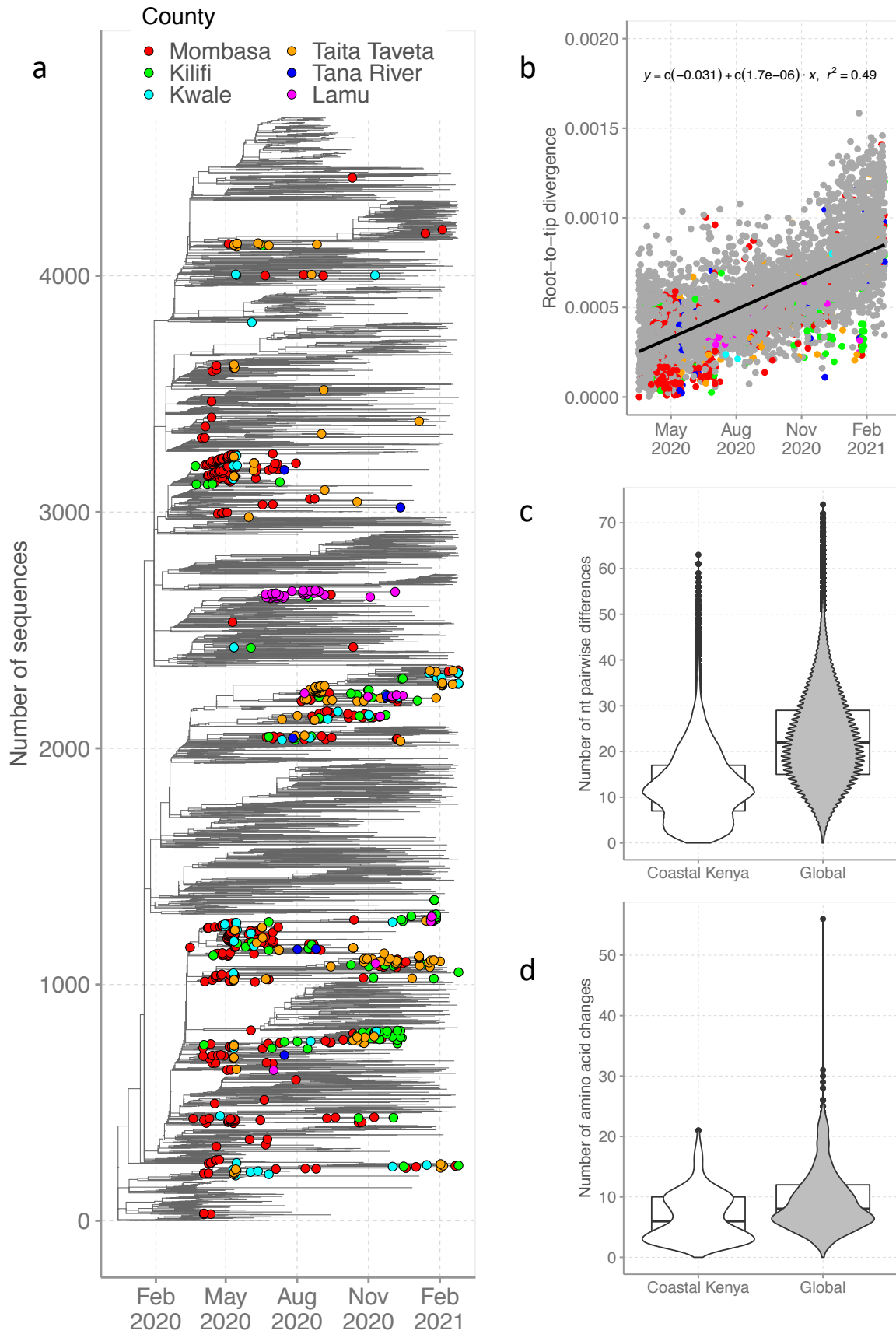
275

276 The coastal Kenya genomes had a fewer pairwise nucleotide differences (median 12,  
277 interquartile range (IQR): 7-17) compared to the global subsampled comparison dataset median  
278 22, IQR: 15-29), **Figure 5C**. A similar difference was observed at amino acid level, with the  
279 coastal Kenya genomes showing a median of six amino acid substitutions (IQR: 3-10) while  
280 the global dataset showed a median of eight amino acid substitutions (IQR: 6-12), **Figure 5D**.

281

282 For detailed investigation into the local SARS-CoV-2 genetic diversity, we reconstructed time-  
283 resolved lineage specific phylogenetic trees are shown in **Figure 6**. The three lineages first  
284 identified in Kenya (B.1.530, B.1.549 and B.1.596.1) were found to (a) possess significant  
285 diversity consistent with widescale spread within Kenya (**Figure 6C-E**), (b) formed multiple  
286 county-specific sub-clusters and (c) show local sequences interspersed with global comparison  
287 genomes from the same lineage implying potentially export (or import) events, **Figure 6D**.  
288 However, the picture painted by lineage N.8 was different. This lineage was mainly detected  
289 in Lamu forming a single monophyletic group (**Figure 6H**) when co-analyzed with its  
290 precursor lineage B.1.1.33, an observation consistent with a single introduction as B.1.1.33  
291 then local evolution expansion.

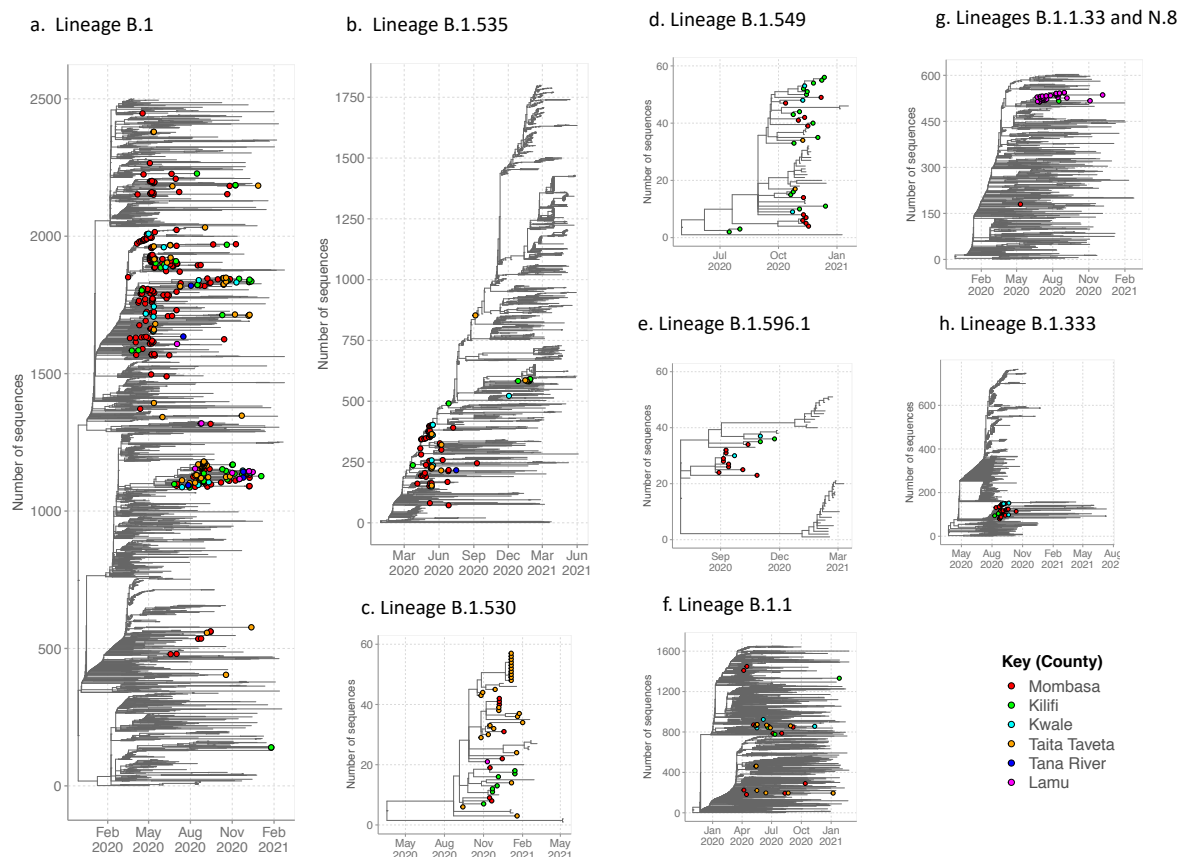
292



293  
294

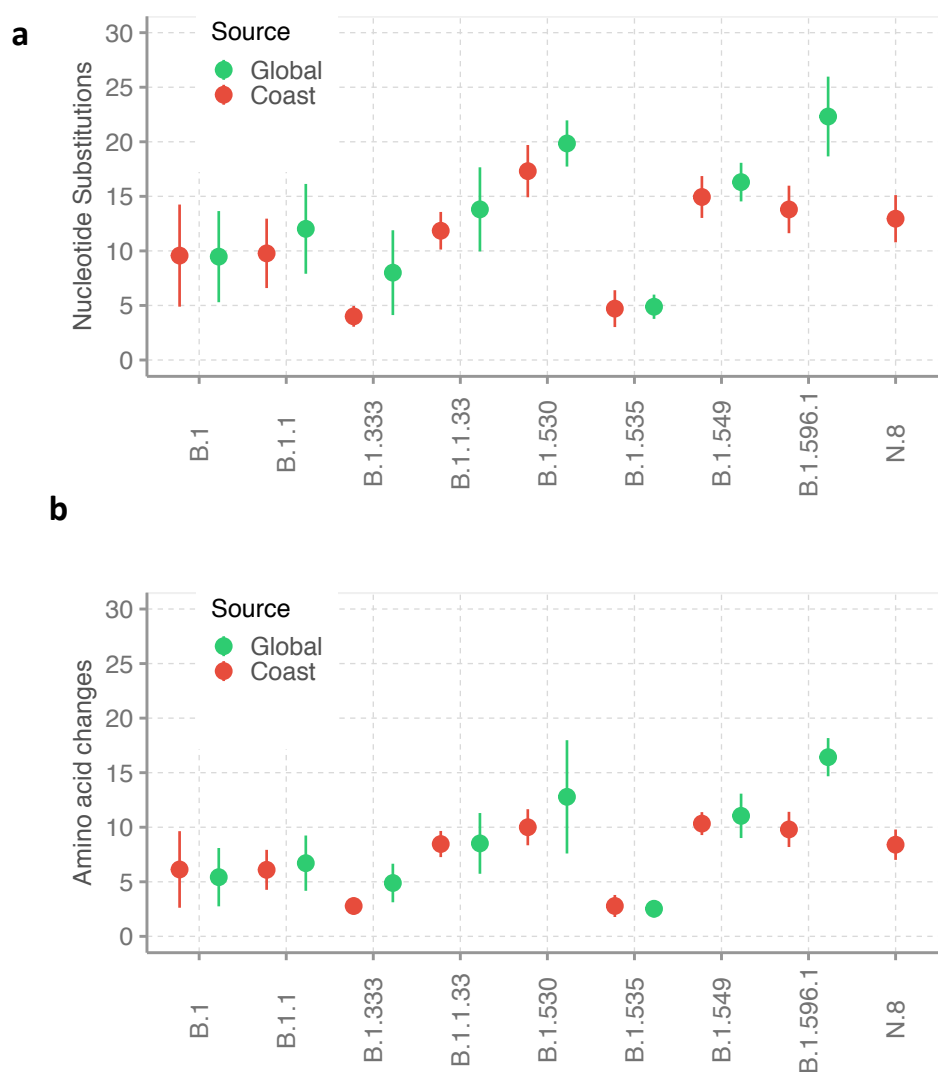
**Figure 5. SARS-CoV-2 diversity in Coastal Kenya. Panel A, Time-resolved global**

295 phylogeny that combined 732 Coastal Kenya SARS-CoV-2 genomes and 3,958 global  
 296 reference sequences (after removal of sequences not fitting molecular clock-like evolution  
 297 using Tree-Time). Panel B, Root-to-tip regression analysis coastal Kenya (colored by county)  
 298 and global (shown in grey) sequences with determination of correlation co-efficient ( $r^2$ ). Panel  
 299 C, comparison of pairwise nucleotide differences of the 732 coastal Kenya genomes included  
 300 on the phylogenetic tree and the 3,958 global genomes. Panel D, comparison of the number of  
 301 amino acid substitutions of the coastal Kenya genomes versus the global dataset relative to the  
 302 Wuhan-Hu-1 reference genome.  
 303



304 **Figure 6. Time-resolved lineage-specific phylogenetic trees for the commonly detected**  
 305 **lineages in Coastal Kenya.** The Coastal Kenya genomes are indicated with filled circles  
 306 colored by county. Panel A, phylogeny for lineage B.1 that combined 369 Coastal Kenya  
 307 genomes and 2,319 global reference sequences. Panel B, phylogeny for lineage B.1.535 that  
 308 combined 82 Coastal Kenya genomes and 1,867 global sequences belonging also to B.1.535.  
 309 Panel C, phylogeny for lineage B.1.530 that combined 42 Coastal Kenya genomes and 15  
 310 global sequences belonging also to B.1.530. Panel D, phylogeny for lineage B.1.549 that  
 311 combined 33 Coastal Kenya genomes and 23 sequences from other countries of lineage B.1.  
 312 549. Panel E, phylogeny for lineage B.1.596.1 combines 15 Coastal Kenya genomes and 36  
 313 global reference sequences. Panel F, phylogeny for lineage B.1.1 that combined 31 Coastal  
 314 Kenya genomes and 1,629 global sequences belonging to lineage B.1.1. Panel G, phylogeny  
 315 for lineage N.8 that combined 20 Coastal Kenya genomes of lineage N.8, 13 Coastal Kenya  
 316 genomes of lineage B.1.1.33 and 570 global sequences of lineage B.1.1.33. Panel H, phylogeny  
 317 for lineage B.1.333 combines 32 Coastal Kenya genomes and 753 global reference sequences.  
 318  
 319  
 320

321 In **Figure 7**, we present a comparison of both the number of nucleotide and amino acid  
322 substitutions for each major lineage between the coastal Kenya genomes and the global  
323 subsample. The divergence of the lineages from the original Wuhan sequence at the global  
324 level was closely mirrored at the local coastal Kenya level. The most divergent lineages among  
325 the subset we examined were three of the Kenya specific lineages (B.1.530, B.1.549 and  
326 B.1.596.1). The least genetically divergence lineages from the original Wuhan SARS-CoV-2  
327 strain were viruses occurring within lineage B.1.333 and lineage B.1.535, **Figure 7**.  
328

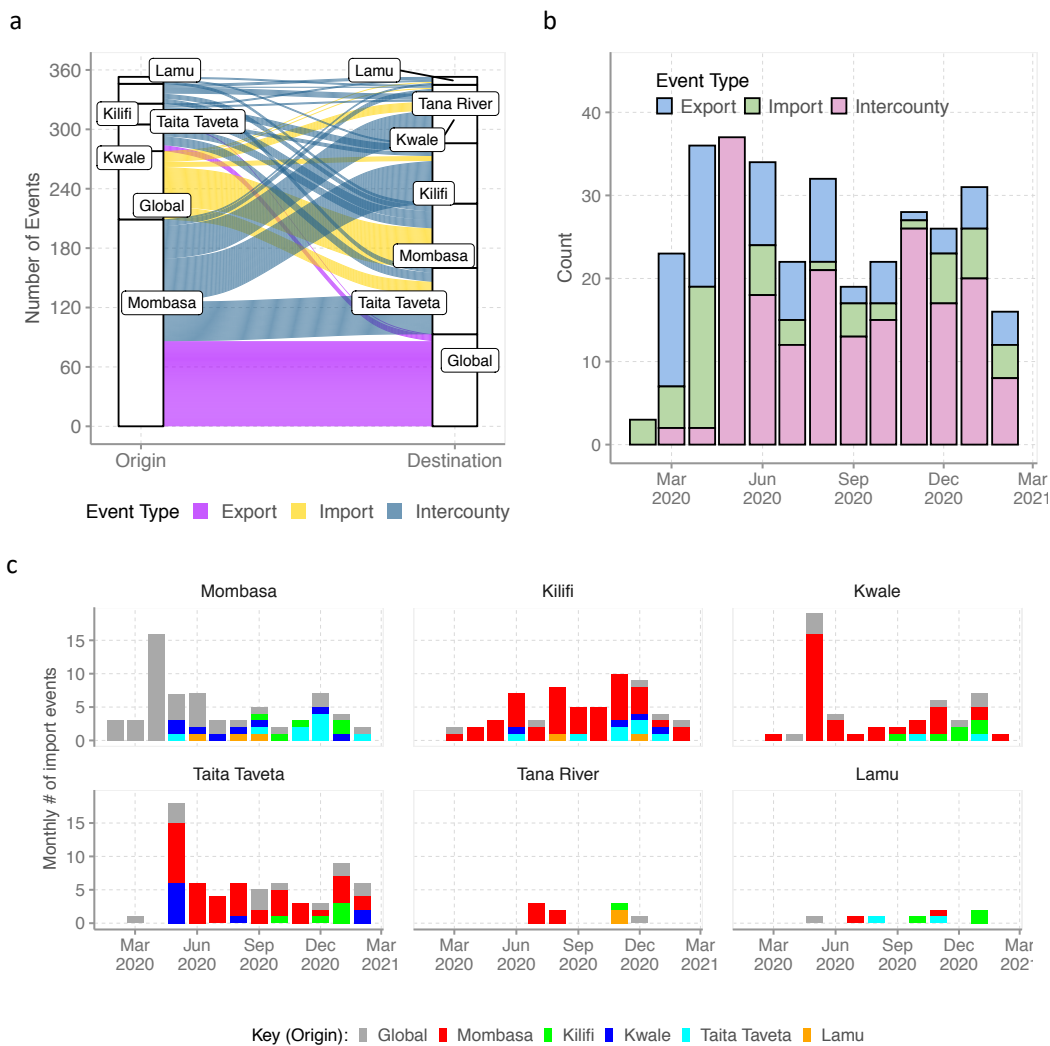


329 **Figure 7. Level of genetic divergence of the major detected lineages in coastal Kenya.** The  
330 sequences were first aligned to the 2019 Wuhan-Hu-1 isolate and the number of changes  
331 compared to isolates across the globe of the same lineage. Panel A shows the divergence at  
332 nucleotide sequence level. Panel B shows the divergence at amino acid sequence level  
333 focusing on synonymous substitutions.  
334

### 335 *Viral imports and export from Coastal Kenya*

336 We used ancestral location state reconstruction of the dated phylogeny (**Figure 5A**) to infer the  
337 number of viral import and exports (Sagulenko et al., 2018). In total, between March 2020 and  
338 February 2021, we detected 357 location transition events within and between the six counties  
339

340 and the outside world. Sixty-nine of these were viral import events into Coastal Kenya from  
 341 outside and 93 were viral exports events from Coastal Kenya to outside populations, **Figure**  
 342 **8A-C**. A total of 191 import/export events were detected between the six Coastal counties.



343 **Figure 8. Viral imports and export from Coastal Kenya.** Panel A, an alluvium plot showing  
 344 the number of imports and exports of viral lineages from Coastal Kenya shown as an alluvium  
 345 plot. Panel B, monthly location transition events observed in the Coastal Kenya and global  
 346 genomes shown as either viral export, import or inter-county transmission. Panel C, monthly  
 347 number of import events into the individual Coastal counties and their origins.  
 348  
 349

350 Virus imports into the region occurred through Mombasa (n=40, 60%), Kwale (n=9, 13%),  
 351 Taita Taveta (n=13, 19%), Kilifi (n=5, 7%) and Lamu (n=1, 1%) while virus export from the  
 352 region occurred through Mombasa (n=86, 41%), Kwale (n=6, 6%) and Taita Taveta (n=1, 1%).  
 353 Of the 69 detected viral imports, 34 (49%) occurred during the introductions phase, 16 (23%)  
 354 occurred during wave one and 19 (28%) during wave two while of the 93 detected viral exports,  
 355 43 (46%) occurred during introductions phase, 31 (33%) during wave one and 19 (20%) during  
 356 wave two, **Figure 8B and 8D**.  
 357

358 The analysis revealed the top six country sources of importation into coastal Kenya as; USA,  
 359 (n=19, 31%), England (n=10, 16%), Uganda (n=6, 10%), Japan (n=5, 8%) and Mozambique



360 (n=3, 5%), **Supplementary-Figure 4 and Supplementary-Figure 5**. On the other hand, the  
361 top six country destinations of viral export from coastal Kenya were USA (n=22, 17%),  
362 Australia (n=16, 13%), India (n=14, 11%), Uganda (n=8, 6%), England (n=7, 6%) and Scotland  
363 (n=4, 3%).

364  
365 To infer the sensitivity of our analysis, we repeated the import/export analysis using a (a)  
366 different global sub-sample (**Supplementary-Figure 4**) and (b) with a normalized subsample  
367 of the coastal Kenya genomes accounting for total reported infections per county,  
368 **Supplementary-Figure 5**. The reanalysis came to roughly similar conclusions as above and  
369 the results are presented in **Supplementary-Table 1**.

### 370 371 *Amino acid changes in the Coastal Kenya genomes*

372 We scanned the entire SARS-CoV-2 genome for amino acid changes from the Wuhan  
373 reference genome in the 747 Coastal Kenya sequences. Below we summarize changes that  
374 were identified in  $\geq 5$  genomes. We observed 20 changes in the spike (S) protein meeting the  
375 set criterion, 15 in the nucleocapsid (N) protein, 14 in ORF3a, 13 in ORF1b, nine in the ORF1a,  
376 seven in ORF8, five in ORF6, and one in the envelope (E) protein. No codon changes, meeting  
377 of criterion was observed in the matrix (M) protein, ORF7a, ORF7b or ORF10. The most  
378 common amino acid changes observed in the Coastal Kenya genomes were S: D614G (94.4%),  
379 ORF1b: P314L (92.9%), ORF3a: Q57H (15.5%), ORF1a: T265I (15.4%) and N: R195K  
380 (12.9%), **Supplementary-Figure 6**. Several lineage-defining mutations were observed for the  
381 Kenya-specific lineages as summarized in **Supplementary-Figure 7**.

382

## 383 **DISCUSSION**

384 We present evidence of continuous additional introductions of SARS-CoV-2 into Coastal  
385 Kenya during the first year of the pandemic documenting at least 69 independent viral  
386 introduction events into the region. Multiple introductions were observed even at county level,  
387 with extensive inter-county spread observed after the early phase period. Strikingly, many of  
388 the imports and exports from the region occurred through the city of Mombasa, a major  
389 commercial, industrial and tourist destination in the region. This observation highlights  
390 Mombasa's central role as a gateway for virus entry into coastal Kenya and the significance  
391 continuous surveillance in this county for early detection of variants in coastal Kenya  
392 populations.

393

394 During the early phase of the pandemic several Pango lineages were introduced into the region  
395 (A, B, AL.1, B.1, B.1.1, B.1.1.1, B.1.1.33, B.1.13, B.1.153, b.1.222, B.1.333, B.1.535, B.4 and  
396 B.4.7). However, it is lineage B.1 which has a European origin, that became predominant,  
397 48.2% of the cases (Githinji et al., 2020). This lineage had the D614G change in the spike  
398 protein which has been found to enhance viral fitness (Baric, 2020) and this may have boosted  
399 its local transmission and early dominance. Notably also the import/export analysis found that  
400 the majority of viral imports to coastal Kenya had an Origin in USA, England and neighboring  
401 Uganda. Viral exports from coastal Kenya happened to USA, Uganda, Australia among other  
402 destinations.

403

404 During the national wave one, the Coastal region observed SARS-CoV-2 transmission mainly  
405 in counties of Mombasa, Taita Taveta and Lamu. The key lineages observed then were B.1  
406 (58.7%), N.8 (8.517%), B.1.535 (6.7%), B.1.1 (5.8%). Notably, lineage N.8 was specific to  
407 Lamu County. Lineage N.8 has not been observed outside Kenya and its precursor lineage

408 (B.1.1.33) was observed earlier in South America, mainly Brazil. Lineage N.8 may have arisen  
409 from an introduction event of B.1.1.33 into Lamu county, and then local evolution and  
410 localization to Lamu. The N.8 lineage has seven characteristic lineage defining mutations  
411 including S: D614G and R203K, G204R and I292T on the nucleocapsid protein.

412  
413 The national wave two started in October 2020 and led to significant viral transmission in all  
414 the six coastal counties. In Kilifi, Tana River and Kwale counties, this was the first significant  
415 peak. This wave two was majorly comprised mainly three lineages: B.1 (42%), B.1.530 (16%),  
416 and B.1.549 (12%). Lineages B.1.530 and B.1.549, first identified in Kenya, and may have  
417 arisen from local evolution of lineage B.1. Lineage B.1.530 has six characteristic mutations  
418 including spike P681H change adjacent to the biologically important furin cleavage site while  
419 lineage B.1.549 has seven characteristic mutations, five occurring in the ORF1a or ORF1b.  
420 Both two lineages have now been observed in a few other countries albeit in small numbers:  
421 seven countries for B.1.530 (Rwanda, Netherlands, Germany, Denmark, USA, Japan and  
422 Australia) and three countries for B.1.549 (England, USA, Canada).

423  
424 Near the tail end of the national wave two, we detected cases of the Beta VOC (B.1.351) and  
425 the Alpha VOC (B.1.1.7). Beta was first to be observed, initially in Kilifi in two asymptomatic  
426 international travelers in mid-December 2020. Alpha was subsequently detected in a local who  
427 presented to a Mombasa clinic in the second week of January 2021. In the subsequent weeks  
428 up to the end of the period covered by this analysis (February 2021), only one additional Alpha  
429 case was detected in our sequenced samples unlike lineage Beta which continued to be detected  
430 sporadically in January and February 2021 especially in the counties of Taita Taveta and Kwale  
431 which border Tanzania.

432  
433 When we compared the local lineage patterns to the global patterns, it was notable that although  
434 several lineages were detected globally (>900), only a small fraction (<10%) of this were  
435 documented locally (O'Toole et al., 2021). The VOC were already extensively spread across  
436 Eastern Africa (Beta VOC), Africa (Beta VOC) and worldwide (Alpha VOC) in the last quarter  
437 of 2020 unlike the findings here for coastal Kenya. This implies that a slight lag occurred in the  
438 VOC arrival and large-scale spread in Coastal Kenya perhaps due to remoteness and public  
439 health measures that have remained in place during the period.

440  
441 While Mombasa, Taita Taveta and Lamu counties had experienced two waves of infections by  
442 February 2021, Kilifi, Tana River and Kwale counties had experienced only a single major  
443 wave. This may have been brought about by differences in population size, density, and  
444 accessibility for these different counties. For instance, Mombasa is densely populated  
445 (~5,604.64 persons/sq.km, 2019 census), has a large seaport and an international airport while  
446 Lamu has a sparsely populated mainland and with several islands (total population 143,920,  
447 population density, 21.84 persons/sq.km). Taita Taveta and Kwale counties are primarily rural  
448 counties bordering Tanzania which had a largely uncontrolled SARS-CoV-2 epidemic during  
449 the study period (Rice et al., 2021). Tana River is the most remote and sparsely populated (8.9  
450 persons/sq.km) county in the region.

451  
452 Our study contributes to the limited but growing literature illuminating SARS-CoV-2  
453 transmission patterns in sub-Saharan Africa (Bugembe et al., 2021a; Bugembe et al., 2021b;  
454 Tegally et al., 2021a; Tegally et al., 2021b; Wilkinson et al.). The patterns revealed here inform  
455 on frequency, sources and routes of introductions of SARS-CoV-2 into Coastal Kenya and  
456 have implications on future control strategies. The key limitations of our analysis include, first,  
457 the samples we sequenced were only those available to us through the rapid response teams

458 (RRTs) whose case identification protocols were altered at the different study pandemic  
459 phases. Second, we only sequenced ~20% of the positive samples identified in our laboratory,  
460 the majority from the introduction phase and were prioritizing samples with a Ct value of  
461 <30.0. Third, the sampling across the six Coastal counties was not uniform, probably in part  
462 due to varied distance from our testing centre located in Kilifi County.

463

464 In conclusion, we show that the first two SARS-CoV-2 waves in Coastal Kenya observed  
465 transmission of both newly introduced and potentially locally evolved variants. New variant  
466 introductions were mainly through Mombasa city and originated USA, England, Uganda  
467 among other predicated sources. Only a limited number of the many introduced variants  
468 progressed to transmit extensively. Unlike in the global contemporaneous sample, we did not  
469 find evidence of extensive local transmission of the global VOC during wave two. Thus, we  
470 infer that it is more likely that the relaxation of some of the interventions (e.g., reopening of  
471 learning institutions, airspace, bars and restaurants) that drove the second wave of infections.  
472 Our study shows the importance of detailed local genomic surveillance even in remote and  
473 under-resourced settings to understand spread patterns of SARS-CoV-2 to optimize local  
474 interventions.

475

## 476 **MATERIALS AND METHODS**

### 477 *Ethical statement*

478 Samples analysed here were collected under the MoH protocols as part of the national response  
479 to the COVID-19 pandemic. The whole genome sequencing study protocol was reviewed and  
480 approved by the Scientific and Ethics Review Committee (SERU), Kenya Medical Research  
481 Institute (KEMRI), Nairobi, Kenya (SERU #4035). Individual patient consent was not required  
482 by the committee for the use of these samples for genomic surveillance to inform public health  
483 response.

484

### 485 *Study period and population*

486 This study examined samples collected between 17<sup>th</sup> March 2020 to 26<sup>th</sup> February 2021 from  
487 six counties in Coastal Kenya i.e. Mombasa, Kilifi, Kwale, Tana River, Taita Taveta and Lamu.  
488 We divided the study period into three phases based on trends in daily reported COVID-19  
489 national case numbers by the MoH (MOH, 2021): (a) Introduction phase – 17<sup>th</sup> March-20<sup>th</sup>  
490 May, 2020 (b) Wave one – 21<sup>st</sup> May-15<sup>th</sup> September 2020 and (c) Wave two – 16<sup>th</sup> September  
491 to 28<sup>th</sup> February 2021, **Figure 1**. We considered the transition from introduction phase to Wave  
492 one as the timepoint when the national daily positives exceeded 50 and the transition from  
493 Wave one to Wave two as the timepoint when a consistent renewed rise of national daily  
494 number of positives started after Wave one peak.

495

### 496 *Patient samples*

497 This study analyzed SARS-CoV-2 positive NP/OP swab samples that were collected by the  
498 MoH County Department of Health RRTs, across all the six counties of Coastal Kenya. SARS-  
499 CoV-2 positive diagnosis was made at the KWTRP in Kilifi County (Agoti et al., 2021). The  
500 RRTs delivered the NP/OP swabs to KWTRP laboratories within 48 hours of collection in cool  
501 boxes with ice packs. The samples were from persons of any age and sampling followed the  
502 MOH eligibility criteria that were revised from time to time (Githinji et al., 2020). Persons  
503 sampled included those with acute respiratory symptoms, those with a recent history of travel  
504 to the early COVID-19 hotspots, contacts of confirmed cases, persons presenting at  
505 international border points seeking entry into Kenya, among other criteria.

506

### 507 *SARS-CoV-2 detection at KWTRP*

508 Viral RNA was extracted from the NP/OP samples using any one of seven commercial kits that  
509 were available namely, QIAamp Viral RNA Mini Kit, RNeasy ® QIAcube ® HT Kit,  
510 QIASYMPHONY ® RNA Kit, TOIANamp Virus RNA Kit, Da An Gene Nucleic acid Isolation  
511 and Purification Kit, SPIN X Extraction and RADI COVID-19 detection Kit. The extracts were  
512 tested for the presence of SARS-CoV-2 nucleic acid following different protocols depending  
513 on which one was available of the following 7 kits/protocols: 1) the Berlin (Charité) primer-  
514 probe set (targeting envelope (E) gene, nucleocapsid (N) or RNA-dependent RNA-polymerase  
515 (RdRp)), 2) European Virus Archive – GLOBAL (EVA-g) (targeting E or RdRp genes), 3) Da  
516 An Gene Co. detection Kit (targeting N or ORF1ab), 4) BGI RT-PCR kit (targeting ORF1ab),  
517 5) Sansure Biotech Novel Coronavirus (2019-nCoV) Nucleic Acid Diagnostic real-time RT-  
518 PCR kit or 6) Standard M kit (targeting E and ORF1ab) and 7) TIB MOLBIOL kit (targeting  
519 E gene). Protocol-specific recommended cycle threshold cut-offs were followed in defining  
520 SARS-CoV-2 positives.

521

### 522 ***SARS-CoV-2 genome sequencing***

523 Only samples that had a RT-PCR cycle threshold value of <30.0 were targeted for whole  
524 genome sequencing (Githinji et al., 2020), **Supplementary-Figure 1** . Viral RNA from  
525 positive samples was reextracted using QIAamp Viral RNA Mini kit following the  
526 manufacturer’s instructions and reverse transcribed using LunaScript® RT SuperMix Kit. The  
527 cDNA was amplified using Q5® Hot Start High-Fidelity 2x Mastermix along with the ARTIC  
528 nCoV-2019 version 3 primers. The PCR products were run on a 1.5% agarose gel and for  
529 samples whose SARS-CoV-2 amplification was considered successful were purified using  
530 Agencourt AMPure XP beads and taken forward for library preparation.

531

532 Sequencing libraries were constructed using Oxford Nanopore Technology (ONT) ligation  
533 sequencing kit and the ONT Native Barcoding Expansion kit as described in the ARTIC  
534 protocol (Tyson et al., 2020). Every MinION (Mk1B) run comprised 23 samples and one  
535 negative (no-template) control.

536

### 537 ***SARS-CoV-2 genome assembly***

538 Following MinION sequencing, the FAST5 files were base-called and demultiplexed using the  
539 OTN’s software Guppy v3.5-4.2. For base calling, high accuracy mode was used. Consensus  
540 SARS-CoV-2 sequences were derived from the reads using the ARTIC bioinformatics pipeline  
541 (<https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>; last accessed 2021-08-  
542 03). A threshold of ×20 read depth was required for a base to be included in the consensus  
543 genome otherwise it was masked with an N. Only complete or near-complete genomes N count  
544 < 5,980 (i.e >80% coverage) were taken forward for phylogenetic analysis.

545

### 546 ***Lineage assignment and calling of amino acid changes***

547 The consensus genomes were assigned into lineages using the PANGOLIN nomenclature  
548 (Rambaut et al., 2020). We used the command line pangolin v3.1.10 software to assign Pango  
549 lineages to each consensus sequence with pango v1.2.46 and pangoLEARN model v2021-07-  
550 28 (O’Toole et al., 2021). Contextual information about lineages was obtained from the  
551 Lineage list available at [https://cov-lineages.org/lineage\\_list.html](https://cov-lineages.org/lineage_list.html) (last accessed 2021-08-04).  
552 Variants of concern (VOC) and variants of interest (VOI) were designated based on the WHO  
553 framework as of 31<sup>st</sup> May 2021 ([https://www.who.int/en/activities/tracking-SARS-CoV-2-  
554 variants/](https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/)).

555

556 The amino acid changes in the Coastal Kenya genomes from the reference strain Wuhan-Hu-  
557 1/2019 was investigated using the nextclade tool v0.14.2 (Hadfield et al., 2018):

558 <https://clades.nextstrain.org/>; last accessed 2021-08-03. Mutations in the Kenyan lineages were  
559 visualized using the tool available on Stanford University CORONAVIRUS ANTIVIRAL &  
560 RESISTANCE Database webpage : <https://covdb.stanford.edu/page/mutation-viewer/>; last  
561 accessed 2021-08-03).

562

### 563 **Global contextual sequences**

564 We prepared three sets of contextual sequences from GISAID database  
565 (<https://www.gisaid.org/>). Selected sequences were those with non-ambiguous sampling date,  
566 sampled during the period covered by our study (March 2020 – February 2021) and after  
567 alignment possessed <5,980 ambiguous (N) nucleotides i.e. showed >80% genome  
568 completeness.

- 569 a) **Set 1:** We compiled an African SARS-CoV-2 genome set which we assigned Pango  
570 lineages and used to compare the continental lineage temporal patterns to the Coastal  
571 Kenya lineage distribution. A total of 12,191 genomes were included for this purpose  
572 downloaded on the 11<sup>th</sup> of May 2021. From these we also created an Eastern Africa  
573 subset which comprised of 2,261 genomes from 10 countries, namely, Zimbabwe,  
574 Zambia, Uganda, Rwanda, Reunion (a France overseas territory), Mozambique,  
575 Malawi, Madagascar, Ethiopia and Comoros.
- 576 b) **Set 2:** We compiled a global reference set of 24,511 genomes collected across all the  
577 six inhabited continents between 17<sup>th</sup> March 2020 and 27<sup>th</sup> February 2021. The  
578 genomes were selected from all available genomes in GISAID database using the R  
579 randomization command: *sample\_n()*. Between 1916 and 2174 genomes were selected  
580 for each included month (mean of 2042 and median of 2049) randomly selected across  
581 all continents from across 164 countries. These genomes were used to infer the global  
582 temporal patterns of the lineages observed in Coastal Kenya. A subset of these genomes  
583 (n=4000) was combined with the Kenyan genomes to infer their global phylogenetic  
584 context and in the import/export analysis described below.
- 585 c) **Set 3:** We compiled a genotype reference for the top six observed lineages in Coastal  
586 Kenya. For lineages B.1. and B.1.1 which >5000 genomes exist in GISAID were  
587 prepared we used the sub-sample from Set 2 above. For lineages B.1.530, B.1.549,  
588 B.1.596.1 and N.8 we retrieved all genomes assigned these lineages available from  
589 GISAID to infer their global phylogenetic context.

590

### 591 **Phylogenetic analysis**

592 Multiple sequence alignments were prepared in Nextalign v 0.1.6 software and using the initial  
593 Wuhan sequence (Accession number: NC\_045512) as the reference, command:

594

```
595 nextalign -r NC_045512.fasta -i input.fasta
```

596

597 The alignment was manually inspected in AliView v1.21 to spot any obvious misalignments.  
598 Quick non-bootstrapped neighbor joining trees were created in SEAVIEW v4.6.4 to identify  
599 any aberrant sequences that were henceforth discarded. We reconstructed maximum likelihood  
600 (ML) phylogenies using IQTREE v2.1.3, command:

601

```
602 ./iqtree2 -s input.aligned.fasta -nt 4
```

603

604 The software initiates tree reconstruction after assessment and selection of the best model of  
605 nucleotide substitution for the alignment using ModelFinder. For example for the large ML  
606 phylogeny presented in Figure 5A, the Best-fit mode was determined as GTR+F+R3.

607

608 The ML trees were linked to the various metadata (lineage, county, source etc) in R  
609 programming software v4.0.2 and visualized using R ggTree v2.4.2. TempEst v1.5.3 was used  
610 to assess the presence of a molecular clock signal in analysed data and linear regression of root-  
611 to-tip genetic distances against sampling dates plotted in RStudio and the correlation  
612 coefficient assessed.

613

### 614 ***Import/export analysis***

615 The global ML tree topology (**Figure 5**) was used to estimate the number of viral transmission  
616 events between Coastal Kenya and the rest of the world as in (Wilkinson et al.). The software  
617 TreeTime was used to transform the ML tree topology into a dated tree assuming a constant  
618 genomic evolutionary rate of SARS-CoV-2 of  $8.4 \times 10^{-4}$  nucleotide substitutions per site per  
619 year (Sagulenko et al., 2018), command:

620

```
621 treetime --tre input.aligned.fasta.treefile --aln input.aligned.fasta
```

```
622 --clock-rate 0.00084 --dates dates.csv
```

623

624 Outlier sequences were identified by TreeTime and excluded during this process. A migration  
625 model was henceforth fitted to the resulting time-scaled phylogenetic tree from TreeTime,  
626 mapping the location status of the genomes from the six counties at both the tips and internal  
627 nodes, command:

628

```
629 treetime migration --tree input.tree.nwk --states migration.csv --attribute Location
```

630

631 Using the date and location annotated tree topology, we counted the number of transitions  
632 between and within Coastal Kenya counties and the rest of the world using the python script  
633 developed and described by the team at KRISP ([https://github.com/krisp-kwazulu-  
634 natal/SARSCoV2\\_South\\_Africa\\_major\\_lineages/tree/main/Phylogenetics](https://github.com/krisp-kwazulu-natal/SARSCoV2_South_Africa_major_lineages/tree/main/Phylogenetics) ; last accessed  
635 2020-08-04) and plotted this using ggplot2 v3.3.3. This analysis was repeated with different  
636 sub-sample of the global background data and with a down sample of the coastal Kenya  
637 genomes that was normalized spatial-temporally ( using a maximum of 15 samples per county  
638 per month).

639

640

### 641 ***Epidemiological data***

642 The Kenya daily case data for the period between March 2020 and February 2021 was  
643 downloaded from Our World in Data database  
644 (<https://ourworldindata.org/coronavirus/country/kenya>). Metadata for the Coastal Kenya  
645 samples was compiled from the MoH case investigation forms delivered alongside the samples.

646

### 647 ***Kenya COVID-19 response***

648 We derived status of Kenya government COVID-19 interventions using data available from  
649 Our World in Data database that has calculated the Oxford Stringency Index (SI). These SI  
650 estimations are composite measure based on nine response indicators implemented by  
651 governments rescaled to a value 0-100, with 100 being strictest (Hale et al., Noam Angrist).  
652 The indicators are: (a) School closures, (b) Workplace closures, (c) Cancellation of public  
653 events, (d) Restrictions on public gatherings, (e) Closures of public transport, (f) Stay-at-home  
654 requirements, (g) Public information campaigns, (h) Restrictions on internal movements and  
655 (d) International travel controls. These are The Kenya government revised the policy  
656 interventions at approximately monthly intervals (Brand et al., 2021) and the changes in SI  
657 overtime are shown in **Figure 1D**.

658

## 659 **Statistical analysis**

660 Statistical data analyses were performed in R v 4.0.5. Summary statistics including proportions,  
661 means, median and ranges were calculated and provided where applicable. T-test was used to  
662 evaluate a statistical difference in means of the pairwise nucleotide or amino acid differences  
663 between the coastal and global sample. The *lm* function in R was used to fit a linear regression  
664 model evaluating the relationship between sampling dates and root-to-tip genetic distance in  
665 the ML phylogeny. The goodness of fit was inferred from the correlation coefficient.

666

## 667 **Data availability**

668 The data and scripts used to generate the figures shown in the manuscript are hosted on Harvard  
669 dataverse: DOI: <https://doi.org/10.7910/DVN/4ZZYIM>.

670

671 The Coastal Kenya genomes are available in the GISAID database (Accession numbers:  
672 EPI\_ISL\_1039223-29; EPI\_ISL1440075-109, IPI\_ISL\_457845-931; EPI\_ISL\_568695-872;  
673 EPI\_ISL806548-717; EPI\_ISL\_855494-548; EPI\_ISL\_968807-9245).

674

## 675 **Acknowledgements**

676 We thank (a) the members of the six Coastal counties of Kenya RRTs for collecting the samples  
677 analysed here; (b) the members of the COVID-19 KWTRP Testing Team who tirelessly  
678 analysed the samples received at KWTRP to identify positives (see full list of members below);  
679 (c) the KWTRP data entry team, (d) Laboratories that have shared sequence data on GISAID  
680 that we included as comparison data in our analysis (see list in appendix); (e) the KRISP team  
681 in South Africa for sharing the scripts we used in the import/export analysis and AFRICA-  
682 CDC for facilitating Africa genomics training. This paper is published with permission of the  
683 Director of KEMRI.

684

### 685 ***Members of COVID-19 Testing Team at KWTRP***

686 Agnes Mutiso, Alfred Mwanzu, Angela Karani, Bonface M. Gichuki, Boniface Karia, Brian  
687 Bartilol, Brian Tawa, Calleb Odundo, Caroline Ngetsa, Clement Lewa, Daisy Mugo, David  
688 Amadi, David Ileri, Debra Riako, Domtila Kimani, Edwin Machanja, Elijah Gicherua, Elisha  
689 Omer, Faith Gambo, Horace Gumba, Isaac Musungu, James Chemweno, Janet Thoya, Jedida  
690 Mwacharo, John Gitonga, Johnstone Makale, Justine Getonto, Kelly Ominde, Kelvias Keter,  
691 Lydia Nyamako, Margaret Nunah, Martin Mutunga, Metrine Tendwa, Moses Mosobo, Nelson  
692 Ouma, Nicole Achieng, Patience Kiyuka, Perpetual Wanjiku, Peter Mwaura, Rita Warui,  
693 Robinson Cheruiyot, Salim Mwarumba, Shaban Mwangi, Shadrack Mutua, Sharon Owuor,  
694 Susan Njuguna, Victor Osoi, Wesley Cheruiyot, Wilfred Nyamu, Wilson Gumbi and Yiakon  
695 Sein.

696

## 697 **Funding**

698 This work was supported by the National Institute for Health Research (NIHR) (project  
699 references 17/63/82 (PI Prof. James Nokes) and 16/136/33 (PI Prof. Mark Woolhouse) using  
700 UK aid from the UK Government to support global health research, The UK Foreign,  
701 Commonwealth and Development Office and Wellcome Trust (grant# 220985). The views  
702 expressed in this publication are those of the author (s) and not necessarily those of NIHR, the

703 Department of Health and Social Care, Foreign Commonwealth and Development Office,  
704 Wellcome Trust or the UK government. Some members of COVID-19 Testing Team at  
705 KWTRP were supported by funding received by Dr Marta Maia (BOHEMIA study funded  
706 UNITAID), Dr Francis Ndungu (Senior Fellowship and Research and Innovation Action (RIA)  
707 grants from EDCTP) and Prof. Anthony Scott (PCIVS grant from GAVI), and IAVI (USAID  
708 grant number: AID-OAA-A-16-00032).

709  
710

## 711 REFERENCES

- 712 Adetifa, I.M.O., Uyoga, S., Gitonga, J.N., Mugo, D., Otiende, M., Nyagwange, J., Karanja, H.K.,  
713 Tuju, J., Wanjiku, P., Aman, R., Mwangangi, M., Amoth, P., Kasera, K., Ng'ang'a, W.,  
714 Rombo, C., Yegon, C., Kithi, K., Odhiambo, E., Rotich, T., Orgut, I., Kihara, S., Bottomley,  
715 C., Kagucia, E.W., Gallagher, K.E., Etyang, A., Voller, S., Lambe, T., Wright, D., Barasa, E.,  
716 Tsofa, B., Bejon, P., Ochola-Oyier, L.I., Agweyu, A., Scott, J.A.G., Warimwe, G.M., 2021.  
717 Temporal trends of SARS-CoV-2 seroprevalence during the first wave of the COVID-19  
718 epidemic in Kenya. *Nature Communications* 12(1), 3966.
- 719 Agoti, C., Mutunga, M., Lambisia, A., Kimani, D., Cheruiyot, R., Kiyuka, P., Lewa, C., Gicheru, E.,  
720 Tendwa, M., Said Mohammed, K., Osoti, V., Makale, J., Tawa, B., Odundo, C., Cheruiyot,  
721 W., Nyamu, W., Gumbi, W., Mwacharo, J., Nyamako, L., Otieno, E., Amadi, D., Thoya, J.,  
722 Karani, A., Mugo, D., Musyoki, J., Gumba, H., Mwarumba, S., M. Gichuki, B., Njuguna, S.,  
723 Riako, D., Mutua, S., Gitonga, J., Sein, Y., Bartilol, B., Mwangi, S., O. Omuoyo, D., M.  
724 Morobe, J., de Laurent, Z., Bejon, P., Ochola-Oyier, L., Tsofa, B., 2021. Pooled testing  
725 conserves SARS-CoV-2 laboratory resources and improves test turn-around time: experience  
726 on the Kenyan Coast [version 2; peer review: 3 approved]. *Wellcome Open Research* 5(186).
- 727 Baric, R.S., 2020. Emergence of a Highly Fit SARS-CoV-2 Variant. *New England Journal of*  
728 *Medicine* 383(27), 2684-2686.
- 729 Brand, S.P.C., Ojal, J., Aziza, R., Were, V., Okiro, E.A., Kombe, I.K., Mburu, C., Ogero, M.,  
730 Agweyu, A., Warimwe, G.M., Nyagwange, J., Karanja, H., Gitonga, J.N., Mugo, D., Uyoga,  
731 S., Adetifa, I.M.O., Scott, J.A.G., Otieno, E., Murunga, N., Otiende, M., Ochola-Oyier, L.I.,  
732 Agoti, C.N., Githinji, G., Kasera, K., Amoth, P., Mwangangi, M., Aman, R., Ng'ang'a, W.,  
733 Tsofa, B., Bejon, P., Keeling, M.J., Nokes, D.J., Barasa, E., 2021. COVID-19 Transmission  
734 Dynamics Underlying Epidemic Waves in Kenya. *medRxiv*, 2021.2006.2017.21259100.
- 735 Bugembe, D.L., Kayiwa, J., Phan, M.V.T., Tushabe, P., Balinandi, S., Dhaala, B., Lexow, J.,  
736 Mwebesa, H., Aceng, J., Kyobe, H., Ssemwanga, D., Lutwama, J., Kaleebu, P., Cotten, M.,  
737 2020. Main Routes of Entry and Genomic Diversity of SARS-CoV-2, Uganda. *Emerg Infect*  
738 *Dis* 26(10), 2411-2415.
- 739 Bugembe, D.L., Phan, M.V.T., Ssewanyana, I., Semanda, P., Nansumba, H., Dhaala, B., Nabadda, S.,  
740 O'Toole, Á.N., Rambaut, A., Kaleebu, P., Cotten, M., 2021a. Emergence and spread of a  
741 SARS-CoV-2 lineage A variant (A.23.1) with altered spike protein in Uganda. *Nature*  
742 *Microbiology*.
- 743 Bugembe, D.L., V.T.Phan, M., Ssewanyana, I., Semanda, P., Nansumba, H., Dhaala, B., Nabadda, S.,  
744 O'Toole, Á.N., Rambaut, A., Kaleebu, P., Cotten, M., 2021b. A SARS-CoV-2 lineage A  
745 variant (A.23.1) with altered spike has emerged and is dominating the current Uganda  
746 epidemic. *medRxiv*, 2021.2002.2008.21251393.
- 747 Butera, Y., Mukantwari, E., Artesi, M., D'Arc Umuringa, J., O'Toole, Á.N., Hill, V., Rooke, S.,  
748 Hong, S.L., Dellicour, S., Majyambere, O., Bontems, S., Boujemla, B., Quick, J., Resende,  
749 P.C., Loman, N., Umumararungu, E., Kabanda, A., Murindahabi, M.M., Tuyisenge, P.,  
750 Gashegu, M., Rwabihama, J.P., Sindayiheba, R., Gikic, D., Souopgui, J., Ndifon, W.,  
751 Rutayisire, R., Gatare, S., Mpunga, T., Ngamije, D., Bours, V., Rambaut, A., Nsanzimana, S.,  
752 Baele, G., Durkin, K., Mutesa, L., Rujeni, N., 2021. Genomic Sequencing of SARS-CoV-2 in  
753 Rwanda: evolution and regional dynamics. *medRxiv*, 2021.2004.2002.21254839.
- 754 Etyang, A.O., Lucinde, R., Karanja, H., Kalu, C., Mugo, D., Nyagwange, J., Gitonga, J., Tuju, J.,  
755 Wanjiku, P., Karani, A., Mutua, S., Maroko, H., Nzomo, E., Maitha, E., Kamuri, E., Kaugiria,  
756 T., Weru, J., Ochola, L.B., Kilimo, N., Charo, S., Emukule, N., Moracha, W., Mukabi, D.,



- 757 Okuku, R., Ogutu, M., Angujo, B., Otiende, M., Bottomley, C., Otieno, E., Ndwiga, L.,  
758 Nyaguara, A., Voller, S., Agoti, C., Nokes, D.J., Ochola-Oyier, L.I., Aman, R., Amoth, P.,  
759 Mwangangi, M., Kasera, K., Ng'ang'a, W., Adetifa, I., Kagucia, E.W., Gallagher, K., Uyoga,  
760 S., Tsofa, B., Barasa, E., Bejon, P., Scott, J.A.G., Agweyu, A., Warimwe, G., 2021.  
761 Seroprevalence of Antibodies to SARS-CoV-2 among Health Care Workers in Kenya. *Clin*  
762 *Infect Dis.*
- 763 Geoghegan, J.L., Ren, X., Storey, M., Hadfield, J., Jelley, L., Jefferies, S., Sherwood, J., Paine, S.,  
764 Huang, S., Douglas, J., Mendes, F.K., Sporle, A., Baker, M.G., Murdoch, D.R., French, N.,  
765 Simpson, C.R., Welch, D., Drummond, A.J., Holmes, E.C., Duchêne, S., de Ligt, J., 2020.  
766 Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in  
767 Aotearoa New Zealand. *Nature Communications* 11(1), 6351.
- 768 Githinji, G., de Laurent, Z.R., Mohammed, K.S., Omuoyo, D.O., Macharia, P.M., Morobe, J.M.,  
769 Otieno, E., Kinyanjui, S.M., Agweyu, A., Maitha, E., Kitole, B., Suleiman, T., Mwakingu,  
770 M., Nyambu, J., Otieno, J., Salim, B., Kasera, K., Kiiru, J., Aman, R., Barasa, E., Warimwe,  
771 G., Bejon, P., Tsofa, B., Ochola-Oyier, L.I., Nokes, D.J., Agoti, C.N., 2020. Tracking the  
772 introduction and spread of SARS-CoV-2 in coastal Kenya. *medRxiv*,  
773 2020.2010.2005.20206730.
- 774 Githinji, G., de Laurent, Z.R., Mohammed, K.S., Omuoyo, D.O., Macharia, P.M., Morobe, J.M.,  
775 Otieno, E., Kinyanjui, S.M., Agweyu, A., Maitha, E., Kitole, B., Suleiman, T., Mwakingu,  
776 M., Nyambu, J., Otieno, J., Salim, B., Kasera, K., Kiiru, J., Aman, R., Barasa, E., Warimwe,  
777 G., Bejon, P., Tsofa, B., Ochola-Oyier, L.I., Nokes, D.J., Agoti, C.N., 2021. Tracking the  
778 introduction and spread of SARS-CoV-2 in coastal Kenya. *Nature Communications* 12(1),  
779 4809.
- 780 Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford,  
781 T., Neher, R.A., 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*  
782 34(23), 4121-4123.
- 783 Hale, T., Angrist, N., Boby, T., Cameron-Blake, E., Hallas, L., Kira, B., Majumdar, S., Petherick, A.,  
784 Phillips, T., Tatlow, H., Webster, S., Noam Angrist. Variation in government responses to  
785 COVID-19. BSG-WP-2020/032 vols. University of Oxford.
- 786 Konings, F., Perkins, M.D., Kuhn, J.H., Pallen, M.J., Alm, E.J., Archer, B.N., Barakat, A., Bedford,  
787 T., Bhiman, J.N., Caly, L., Carter, L.L., Cullinane, A., de Oliveira, T., Druce, J., El Masry, I.,  
788 Evans, R., Gao, G.F., Gorbalenya, A.E., Hamblion, E., Herring, B.L., Hodcroft, E., Holmes,  
789 E.C., Kakkar, M., Khare, S., Koopmans, M.P.G., Korber, B., Leite, J., MacCannell, D.,  
790 Marklewitz, M., Maurer-Stroh, S., Rico, J.A.M., Munster, V.J., Neher, R., Munnink, B.O.,  
791 Pavlin, B.I., Peiris, M., Poon, L., Pybus, O., Rambaut, A., Resende, P., Subissi, L., Thiel, V.,  
792 Tong, S., van der Werf, S., von Gottberg, A., Ziebuhr, J., Van Kerkhove, M.D., 2021. SARS-  
793 CoV-2 Variants of Interest and Concern naming scheme conducive for global discourse.  
794 *Nature Microbiology*.
- 795 MOH, 2021. Update on COVID-19 in Kenya. In: coronavirus, N.E.C.o. (Ed.). MOH.
- 796 O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J.T., Colquhoun, R., Ruis, C.,  
797 Abu-Dahab, K., Taylor, B., Yeats, C., du Plessis, L., Maloney, D., Medd, N., Attwood, S.W.,  
798 Aanensen, D.M., Holmes, E.C., Pybus, O.G., Rambaut, A., 2021. Assignment of  
799 epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evolution*.
- 800 Oude Munnink, B.B., Nieuwenhuijse, D.F., Stein, M., O'Toole, Á., Haverkate, M., Mollers, M.,  
801 Kamga, S.K., Schapendonk, C., Pronk, M., Lexmond, P., van der Linden, A., Bestebroer, T.,  
802 Chestakova, I., Overmars, R.J., van Nieuwkoop, S., Molenkamp, R., van der Eijk, A.A.,  
803 GeurtsvanKessel, C., Vennema, H., Meijer, A., Rambaut, A., van Dissel, J., Sikkema, R.S.,  
804 Timen, A., Koopmans, M., 2020. Rapid SARS-CoV-2 whole-genome sequencing and  
805 analysis for informed public health decision-making in the Netherlands. *Nat Med* 26(9),  
806 1405-1410.
- 807 Rambaut, A., Holmes, E.C., O'Toole, Á., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., Pybus,  
808 O.G., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic  
809 epidemiology. *Nat Microbiol* 5(11), 1403-1407.
- 810 Rice, B.L., Annapragada, A., Baker, R.E., Bruijning, M., Dotse-Gborgbortsi, W., Mensah, K., Miller,  
811 I.F., Motaze, N.V., Raheerinandrasana, A., Rajeev, M., Rakotonirina, J., Ramiadantsoa, T.,

- 812 Rasambainarivo, F., Yu, W., Grenfell, B.T., Tatem, A.J., Metcalf, C.J.E., 2021. Variation in  
813 SARS-CoV-2 outbreaks across sub-Saharan Africa. *Nature Medicine* 27(3), 447-453.
- 814 Sagulenko, P., Puller, V., Neher, R.A., 2018. TreeTime: Maximum-likelihood phylodynamic analysis.  
815 *Virus Evol* 4(1), vex042.
- 816 Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D.,  
817 Pillay, S., San, E.J., Msomi, N., Mlisana, K., von Gottberg, A., Walaza, S., Allam, M., Ismail,  
818 A., Mohale, T., Glass, A.J., Engelbrecht, S., Van Zyl, G., Preiser, W., Petruccione, F., Sigal,  
819 A., Hardie, D., Marais, G., Hsiao, N.Y., Korsman, S., Davies, M.A., Tyers, L., Mudau, I.,  
820 York, D., Maslo, C., Goedhals, D., Abrahams, S., Laguda-Akingba, O., Alisoltani-Dehkordi,  
821 A., Godzik, A., Wibmer, C.K., Sewell, B.T., Lourenço, J., Alcantara, L.C.J., Kosakovsky  
822 Pond, S.L., Weaver, S., Martin, D., Lessells, R.J., Bhiman, J.N., Williamson, C., de Oliveira,  
823 T., 2021a. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* 592(7854),  
824 438-443.
- 825 Tegally, H., Wilkinson, E., Lessells, R.J., Giandhari, J., Pillay, S., Msomi, N., Mlisana, K., Bhiman,  
826 J.N., von Gottberg, A., Walaza, S., Fonseca, V., Allam, M., Ismail, A., Glass, A.J.,  
827 Engelbrecht, S., Van Zyl, G., Preiser, W., Williamson, C., Petruccione, F., Sigal, A., Gazy, I.,  
828 Hardie, D., Hsiao, N.Y., Martin, D., York, D., Goedhals, D., San, E.J., Giovanetti, M.,  
829 Lourenço, J., Alcantara, L.C.J., de Oliveira, T., 2021b. Sixteen novel lineages of SARS-CoV-  
830 2 in South Africa. *Nat Med* 27(3), 440-446.
- 831 Tyson, J.R., James, P., Stoddart, D., Sparks, N., Wickenhagen, A., Hall, G., Choi, J.H., Lapointe, H.,  
832 Kamelian, K., Smith, A.D., Prystajecy, N., Goodfellow, I., Wilson, S.J., Harrigan, R.,  
833 Snutch, T.P., Loman, N.J., Quick, J., 2020. Improvements to the ARTIC multiplex PCR  
834 method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv*.
- 835 Uyoga, S., Adetifa, I.M.O., Karanja, H.K., Nyagwange, J., Tuju, J., Wanjiku, P., Aman, R.,  
836 Mwangangi, M., Amoth, P., Kasera, K., Ng'ang'a, W., Rombo, C., Yegon, C., Kithi, K.,  
837 Odhiambo, E., Rotich, T., Orgut, I., Kihara, S., Otiende, M., Bottomley, C., Mupe, Z.N.,  
838 Kagucia, E.W., Gallagher, K.E., Etyang, A., Voller, S., Gitonga, J.N., Mugo, D., Agoti, C.N.,  
839 Otieno, E., Ndwiga, L., Lambe, T., Wright, D., Barasa, E., Tsofa, B., Bejon, P., Ochola-  
840 Oyier, L.I., Agweyu, A., Scott, J.A.G., Warimwe, G.M., 2021a. Seroprevalence of anti-  
841 SARS-CoV-2 IgG antibodies in Kenyan blood donors. *Science* 371(6524), 79-82.
- 842 Uyoga, S., Adetifa, I.M.O., Otiende, M., Yegon, C., Agweyu, A., Warimwe, G.M., Scott, J.A.G.,  
843 2021b. Prevalence of SARS-CoV-2 Antibodies From a National Serosurveillance of Kenyan  
844 Blood Donors, January-March 2021. *Jama*.
- 845 Wilkinson, E., Giovanetti, M., Tegally, H., San, J.E., Lessells, R., Cuadros, D., Martin, D.P.,  
846 Rasmussen, D.A., Zekri, A.-R.N., Sangare, A.K., Ouedraogo, A.-S., Sesay, A.K., Priscilla,  
847 A., Kemi, A.-S., Olubusuyi, A.M., Oluwapelumi, A.O.O., Hammami, A., Amuri, A.A.,  
848 Sayed, A., Ouma, A.E.O., Elargoubi, A., Anthony, A.N., Victoria, A.F., Kazeem, A., George,  
849 A., Trotter, A.J., Yahaya, A.A., Keita, A.K., Diallo, A., Kone, A., Souissi, A., Chtourou, A.,  
850 Gutierrez, A.V., Page, A.J., Vinze, A., Iranzadeh, A., Lambisia, A., Ismail, A., Rosemary, A.,  
851 Sylverken, A., Femi, A., Ibrahim, A., Marycelin, B., Oderinde, B.S., Bolajoko, B., Dhaala,  
852 B., Herring, B.L., Njanpop-Lafourcade, B.-M., Kleinhans, B., McInnis, B., Tegomoh, B.,  
853 Brook, C., Pratt, C.B., Scheepers, C., Akoua-Koffi, C.G., Agoti, C.N., Peyrefitte, C.,  
854 Daubenberger, C., Morang'a, C.M., Nokes, D.J., Amoako, D.G., Bugembe, D.L., Park, D.,  
855 Baker, D., Doolabh, D., Ssemwanga, D., Tshiabuila, D., Bassirou, D., Amuzu, D.S.Y.,  
856 Goedhals, D., Omuoyo, D.O., Maruapula, D., Foster-Nyarko, E., Lusamaki, E.K., Simulundu,  
857 E., Ong'era, E.M., Ngabana, E.N., Shumba, E., Fahime, E.E., Lokilo, E., Mukantwari, E.,  
858 Philomena, E., Belarbi, E., Simon-Lorriere, E., Anoh, E.A., Leendertz, F., Ajili, F., Enoch,  
859 F.O., Wasfi, F., Abdelmoula, F., Mosha, F.S., Takawira, F.T., Derrar, F., Bouzid, F.,  
860 Onikepe, F., Adeola, F., Muyembe, F.M., Tanser, F., Dratibi, F.A., Mbunsu, G.K., et al., A  
861 year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa.  
862 *Science* 0(0), eabj4336.
- 863 Worobey, M., Pekar, J., Larsen, B.B., Nelson, M.I., Hill, V., Joy, J.B., Rambaut, A., Suchard, M.A.,  
864 Wertheim, J.O., Lemey, P., 2020. The emergence of SARS-CoV-2 in Europe and North  
865 America. *Science* 370(6516), 564-570.

866

Supplementary Material for  
**Transmission networks of SARS-CoV-2 in coastal Kenya during the first two waves: a retrospective genomic study**

Agoti et al, 2021

**Corresponding author:**

Dr Charles N. Agoti

Email : [cnyaigoti@kemri-wellcome.org](mailto:cnyaigoti@kemri-wellcome.org)

**Contents**

[Supplementary Figure 1-7](#)

[Supplementary Table 1](#)

**Additional available material relating to this manuscript**

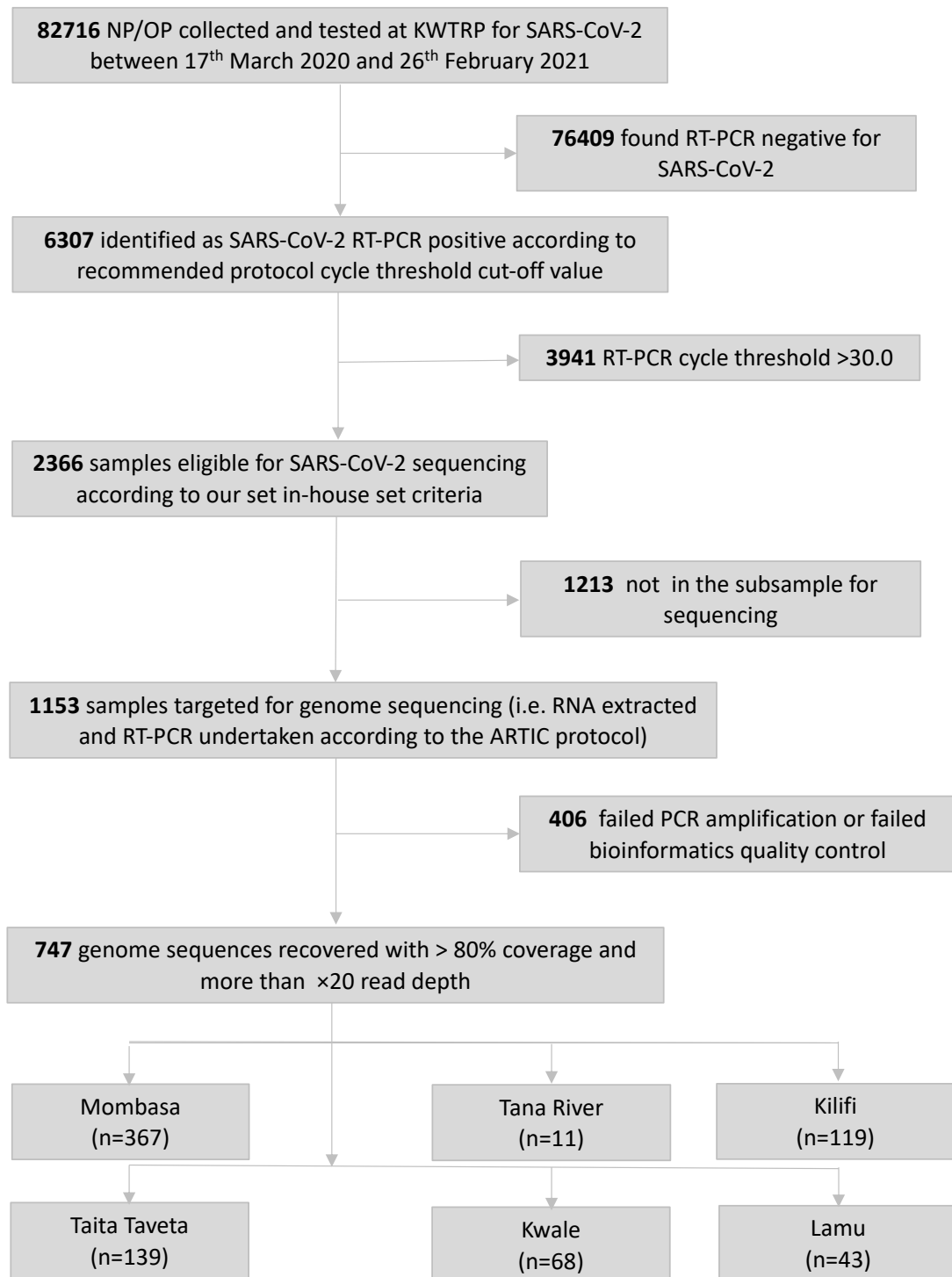
[STROBE checklist](#)

[Transparent Reporting File](#)

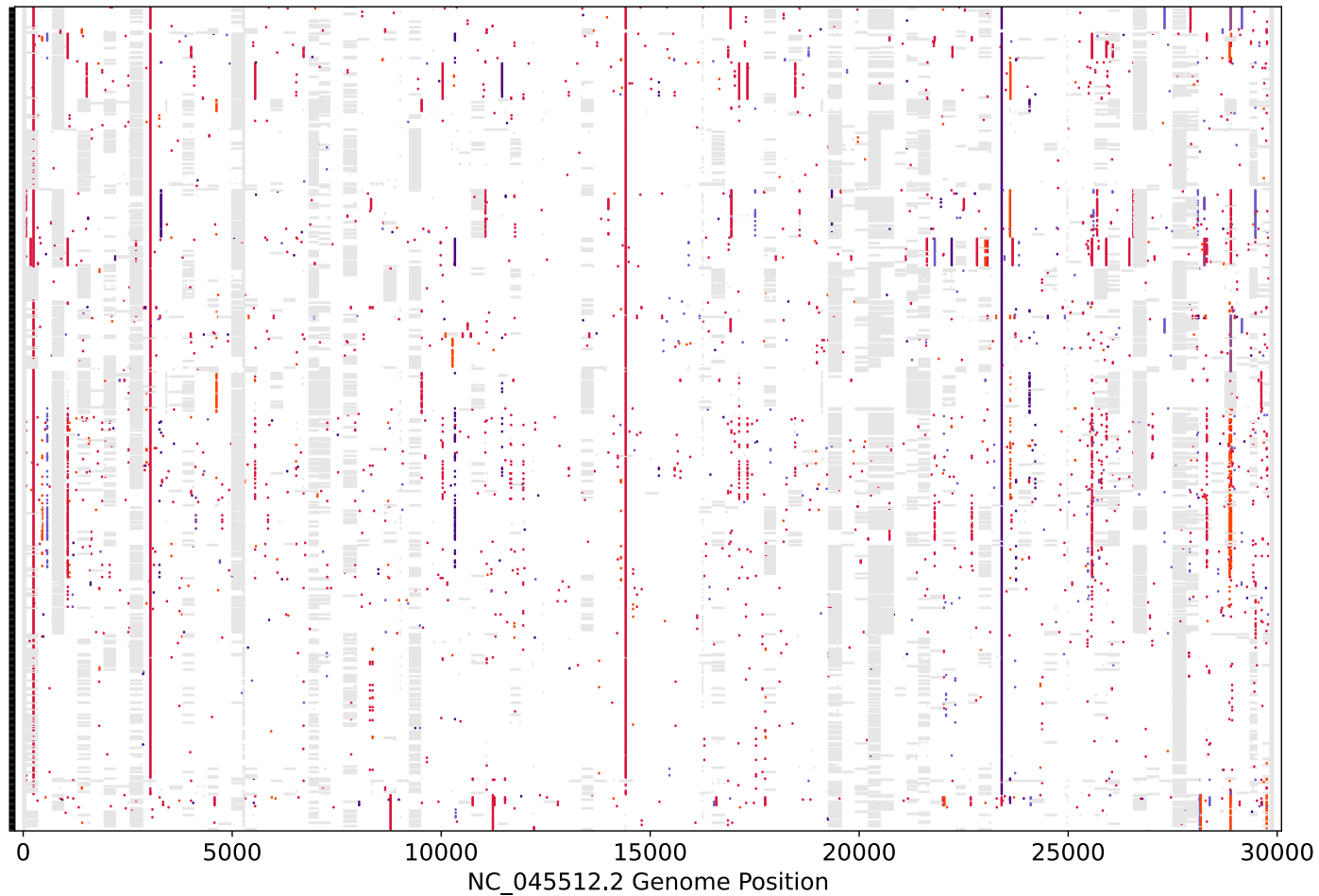
[Source code for figures presented in the manuscript](#)

[Accession numbers of coastal Kenya genomes presented in this study](#)

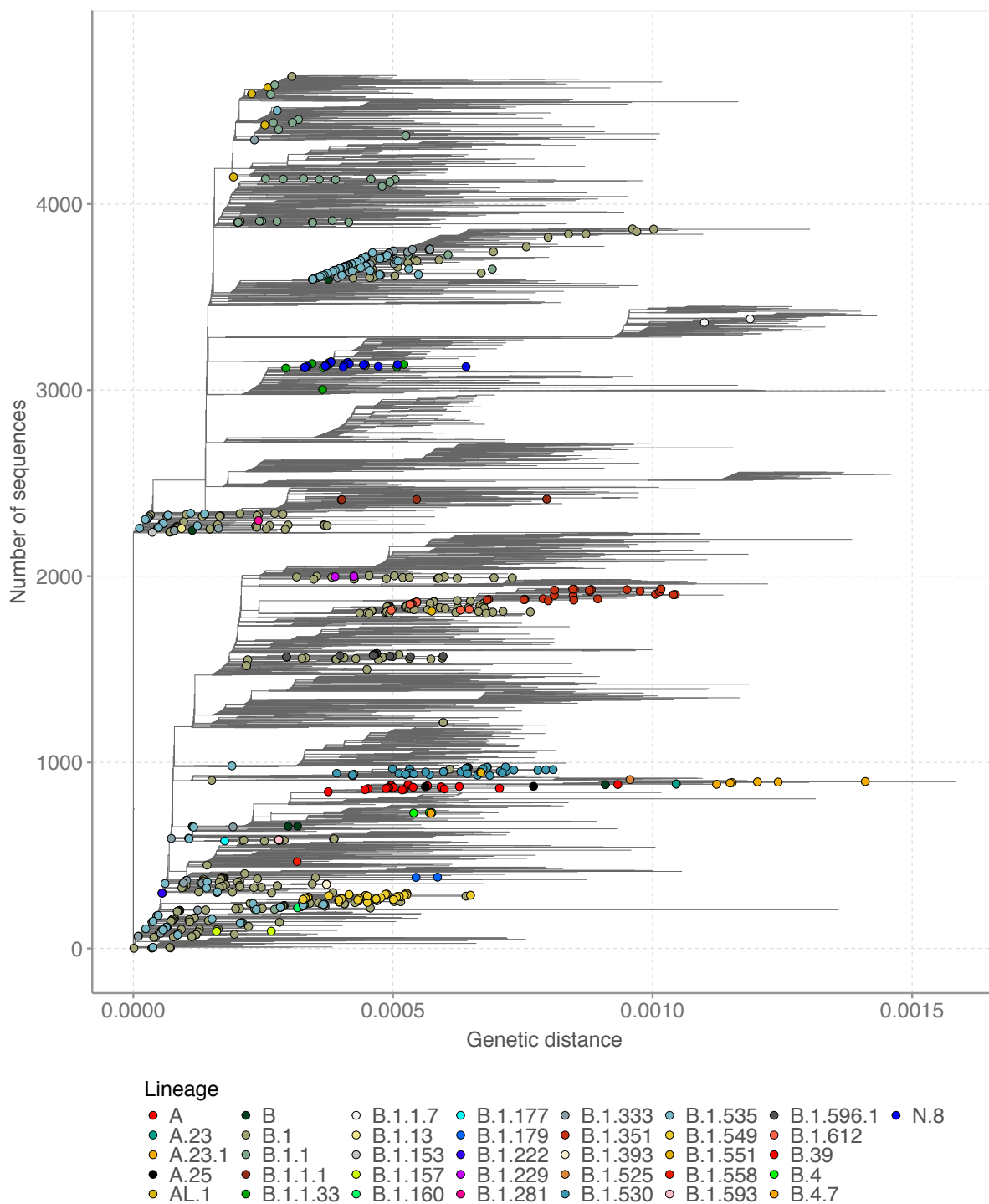
[Acknowledgements GISAID depositors](#)



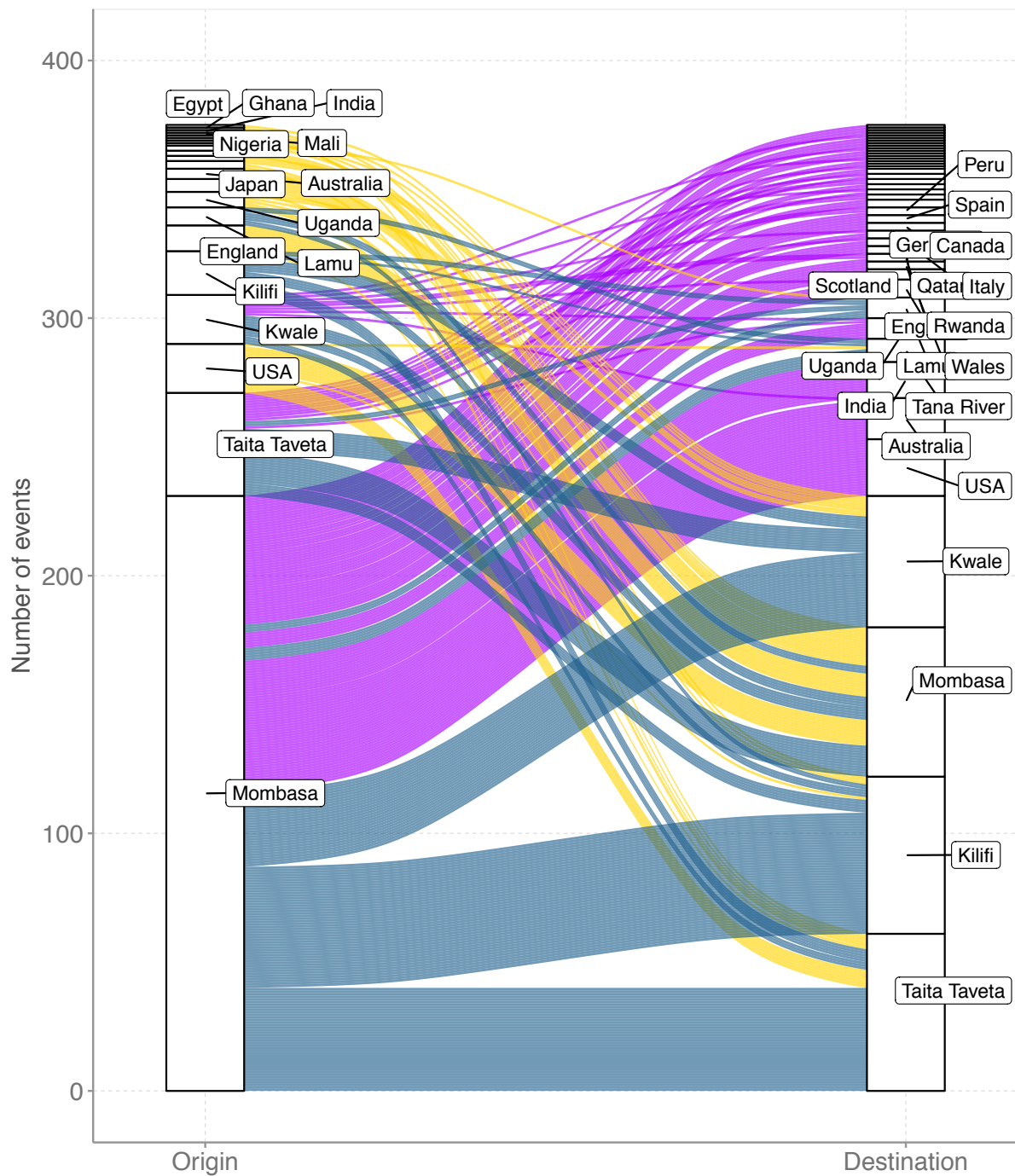
**Supplementary-Figure 1. Flow of samples analysed in this study.**



**Supplementary-Figure 2.** Nucleotide differences in SARS-CoV-2 genome sequences from the coastal Kenya samples compared to the reference genome Wuhan-Hu-1 (Accession #: NC\_045512.2). Vertical coloured bars show the nucleotide differences from the reference genome; orange is a change to A, crimson is a change to T, indigo a change to G, and slate blue a change to C. Grey is a nucleotide deletion or an area not sequenced in our laboratory due to amplicon drop-off or inconclusive assembly.

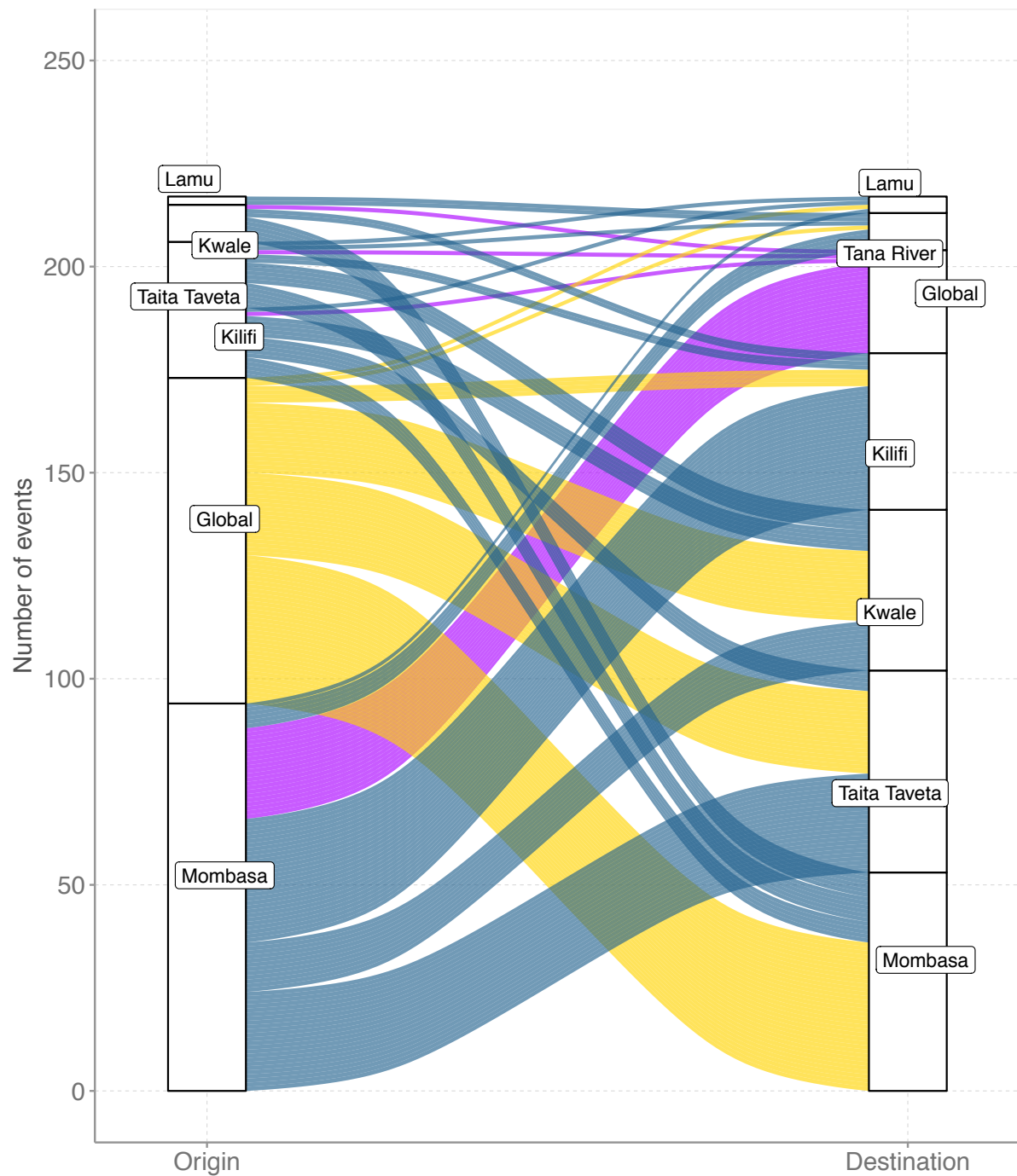


**Supplementary-Figure 3. Mutation-resolved global phylogenetic tree showing the genetic closeness of the sequences from coastal Kenya while including a representative global sub-sample.** The tree uses the same data as Figure 5A in the main test (time-resolved tree) by combining 3,958 global sequences and 732 coastal Kenya sequences. The coastal Kenya sequences are indicated by a filled circle symbol, coloured by the assigned Pango lineage.



Key (Event Type): ■ Export ■ Import ■ Inter-county

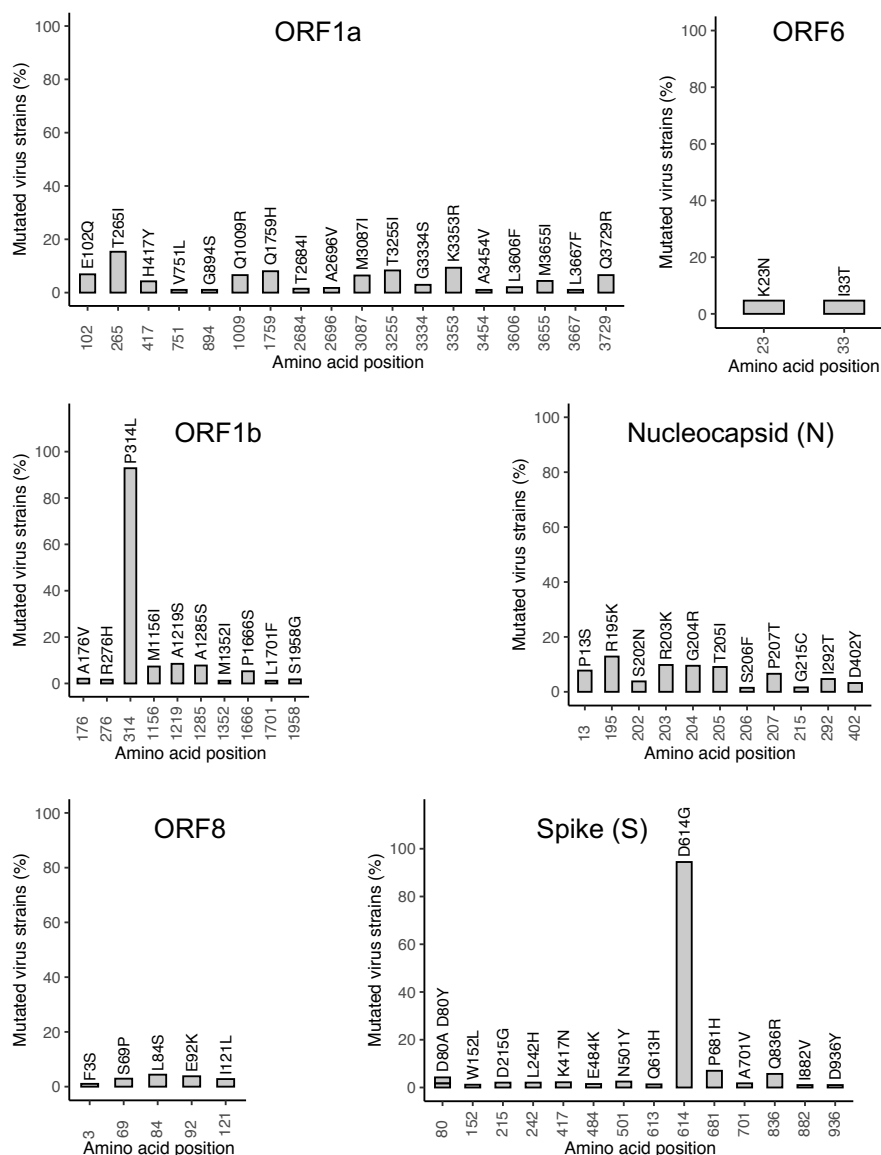
**Supplementary-Figure 4. An alluvium plot showing the number of viral imports and exports from coastal Kenya over the study period.** This figure, unlike Figure 8A, provides the details of the countries from which import to or exports from coastal Kenya occurred. Exports from coastal Kenya are depicted by a purple colour, imports to coastal Kenya are depicted by yellow colour, within coastal Kenya intercounty spread is indicated by the cadet blue colour.



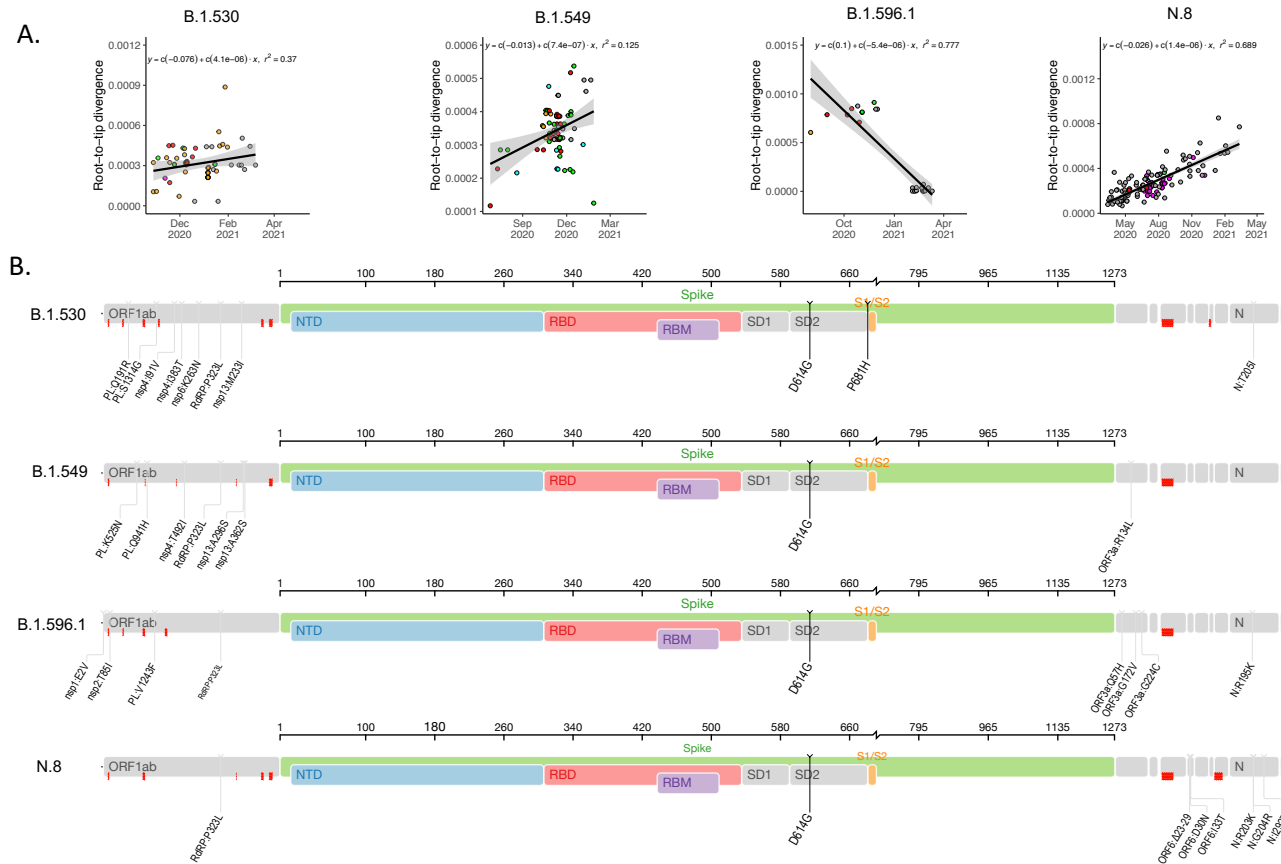
Key (Event Type): ■ Export ■ Import ■ Inter-county

**Supplementary-Figure 5. An alluvium plot showing the number of viral imports and exports from coastal Kenya over the study period using a subsample of the coastal Kenya genomes (n=360).** The sub-sample was randomly selected aiming to be spatial-temporally representative with up to 15 genomes per county per month over the study period. Exports are depicted by purple colour, imports are depicted by yellow colour, within coastal Kenya intercounty spread is indicated by the cadet blue colour.





**Supplementary-Figure 6. Position of non-synonymous mutations in the 747 genomes from coastal Kenya compare with the Wuhan-Hu-1 reference genome (Accession number NC\_045512). Only changes occurring in >1% of genomes are shown. The percentage frequency of the changes in the 747 genomes is shown on the y-axis.**



**Supplementary-Figure 7. Genomic variation characteristics of the four lineages first identified in Kenya.** Panel A, root-to-tip regression plots for the four lineages first identified in Kenya. Panel B, Genome maps of the “Kenya” lineages where the spike region is shown in detail and in colour and the rest of the genome is shown in grey colour. Red marks indicate where sequencing of Kenya strains resulted in ambiguous nucleotides. Plots generated using tool available from Stanford University Web-page Coronavirus antiviral and resistance database <https://covdb.stanford.edu/page/mutation-viewer/>: last accessed on 2021-08-03.

**Supplementary-Table 1.** Summary output from separate runs of the import/export ancestral state reconstruction (ASR) analysis

<b>Characteristic</b>	<b>Run 1<sup>#</sup></b>	<b>Run 2<sup>‡</sup></b>	<b>Run 3<sup>§</sup></b>
<b>Total Location transition events</b>	357	341	217
<b>International imports</b>			
<b>Total</b>	69	65	79
<b>Through Mombasa</b>	40	38	36
<b>Through Kwale</b>	9	5	17
<b>Through Taita Taveta</b>	13	13	20
<b>Through Kilifi</b>	5	5	17
<b>Through Lamu</b>	1	1	1
<b>International exports</b>			
<b>Total</b>	93	87	25
<b>Through Mombasa</b>	86	81	22
<b>Through Kwale</b>	6	3	1
<b>Through Taita Taveta</b>	1	3	1
<b>Through Kilifi</b>	-	-	1
<b>Inter-county transmission events</b>	193	189	113

<sup>#</sup>Import/export analysis results from ASR combining 732 genome sequences from coastal Kenya that passed the molecular clock quality control and a random 3,958 global (outside Kenya) genomes.

<sup>‡</sup>Import/export analysis results from ASR combining 728 genomes from coastal Kenya passing the molecular clock quality control and 3,968 global (outside Kenya) genomes.

<sup>§</sup> Import/export analysis results from ASR combining 338 genome sequences from coastal Kenya that passed the molecular clock quality control and a random 3979 global (outside Kenya) genomes. The coastal Kenya sub-sample used for this run 3 was randomly selected aiming to be spatial-temporally representative with up to 15 genomes per county per month over the study period.