

1 **Genomic surveillance reveals the spread patterns of SARS-CoV-2 in coastal Kenya during** 2 **the first two waves**

3
4 Charles N. Agoti^{1,2,#}, Lynette Isabella Ochola-Oyier¹, Khadija Said Mohammed¹, Arnold W.
5 Lambisia¹, Zaydah R. de Laurent¹, John M. Morobe¹, Maureen W. Mburu¹, Donwilliams O.
6 Omuoyo¹, Edidah M. Ongera¹, Leonard Ndwiga¹, Eric Maitha³, Benson Kitole³, Thani
7 Suleiman³, Mohamed Mwakinangu³, John Nyambu³, John Otieno³, Barke Salim³, Jennifer
8 Musyoki¹, Nickson Murunga¹, Edward Otieno¹, John Kiiru³, Kadondi Kasera³, Patrick Amoth³,
9 Mercy Mwangangi³, Rashid Aman³, Samson Kinyanjui^{1,2,4}, George Warimwe^{1,4}, My Phan⁵,
10 Ambrose Agweyu¹, Matthew Cotten^{5,6}, Edwine Barasa¹, Benjamin Tsofa¹, D. James Nokes^{1,7},
11 Philip Bejon^{1,4}, George Githinji¹

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/ 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**

24 Dr Charles N. Agoti (Email: cnyaigoti@kemri-wellcome.org)

25 26 **ABSTRACT**

27 The transmission networks of SARS-CoV-2 in sub-Saharan Africa remain poorly understood.
28 We analyzed 684 genomes from samples collected across six counties in coastal Kenya during
29 the first two waves (March 2020 - February 2021). Up to 32 Pango lineages were detected in the
30 local sample with six accounting for 88.0% of the sequenced infections: B.1 (60.4%), B.1.1
31 (8.9%), B.1.549 (7.9%), B.1.530 (6.4%), N.8 (4.4%) and A (3.1%). In a contemporaneous global
32 sample, 571 lineages were identified, 247 for Africa and 88 for East Africa. We detected 262
33 location transition events comprising: 64 viral imports into Coastal Kenya; 26 viral exports from
34 coastal Kenya; and 172 inter-county import/export events. Most international viral imports
35 (61%) and exports (88%) occurred through Mombasa, a key coastal touristic and commercial
36 center; and many occurred prior to June 2020, when stringent local COVID-19 restriction
37 measures were enforced. After this period, local transmission dominated, and distinct local
38 phylogenies were seen. Our analysis supports moving control strategies from a focus on
39 international travel to local transmission.

40

41 INTRODUCTION

42 By 15th June 2021, at least 175 million cases of coronavirus disease 2019 (COVID-19) and > 3.7
43 million associated deaths were reported worldwide (1). By the same date, Kenya reported
44 175,176 laboratory-confirmed cases of severe acute respiratory syndrome coronavirus 2 (SARS-
45 CoV-2) and 3,396 COVID-19 associated deaths (2). Serological surveys indicated that Kenya's
46 COVID-19 epidemic had progressed further than could be discerned from the limited laboratory
47 testing case reports that were available during the period (3, 4). At the end of Kenya's first
48 wave (August 2020), nationwide anti-SARS-CoV-2 IgG prevalence based on analysis of blood
49 donor samples was estimated to be 9% (5).

50
51 By June 2021, Kenya had experienced three waves of SARS-CoV-2 infections. The first wave
52 peaked in July-August of 2020, the second in November 2020 and the third in March-April 2021
53 (6). Despite this progression, the local SARS-CoV-2 spread patterns remain poorly understood.
54 The analysis presented here examined genome sequences from the first two waves in Coastal
55 Kenya. Overall, the documented SARS-CoV-2 infections in Kenya during these early two
56 waves were concentrated in cities, especially Nairobi (~42%), the capital, and Mombasa (~8%),
57 the country's second largest city on the Indian ocean coast. In addition to Mombasa, there are
58 five other counties that make up Kenya's Coastal region namely, Kilifi, Kwale, Taita Taveta,
59 Tana River and Lamu. The region is a major tourism destination accessible through multiple
60 airports, seaports and land border entry points from Tanzania through Kwale and Taita Taveta
61 counties (7).

62
63 Throughout the COVID-19 pandemic, genomic surveillance has been critical for tracking the
64 spread of SARS-CoV-2 and investigating viral evolution (8-11). In Kenya, genome sequencing
65 started soon after the identification of the initial case on 13th March 2020 and has remained a key
66 tool informing public health response decisions. We have previously reported on the genomic
67 analysis of SARS-CoV-2 in Coastal Kenya during the early phase of the pandemic (up to July
68 2020) revealing multiple lineage introductions but with mainly lineage B.1 establishing local
69 transmission (7). Here we report the genetic composition of the subsequent two waves (up to
70 February 2021) with the aim of documenting patterns of viral introductions, evolution and spread
71 across the six Coastal counties of Kenya.

72 73 RESULTS

74 *Infection waves in Coastal Kenya counties*

75 The first cases of COVID-19 in Coastal Kenya were reported in Mombasa and Kilifi counties in
76 March 2020, **Figure 1A**. The timing of the first infection peaks across the six counties differed,
77 with Mombasa showing the earliest surge. By 26th February 2021, Mombasa, Lamu and Taita
78 Taveta had observed two major infection peaks while Kilifi, Kwale and Tana River had
79 experienced only a single peak, with the latter two documenting very few cases, **Figure 1A**.
80 During this period, the Kenya Ministry of Health (MoH) reported a total of 12,307 laboratory

81 confirmed SARS-CoV-2 cases for the six Coastal counties: Mombasa (n=8,450, 68.7%), Kilifi
 82 (n=2,458, 20.0%), Taita Taveta (n=567, 4.6%), Kwale (n=437, 3.6%), Lamu (n=309, 2.5%) and
 83 Tana River (n=85, 0.7%), **S1 Figure**.

84

85 ***SARS-CoV-2 testing and sequencing at KWTRP***

86 The KEMRI-Wellcome Trust Research Programme (KWTRP) was designated as the government
 87 SARS-CoV-2 testing center for Coastal region soon after identification of the first case in Kenya.

88 Between 17th March 2020 and 26th February 2021, we tested 82,716

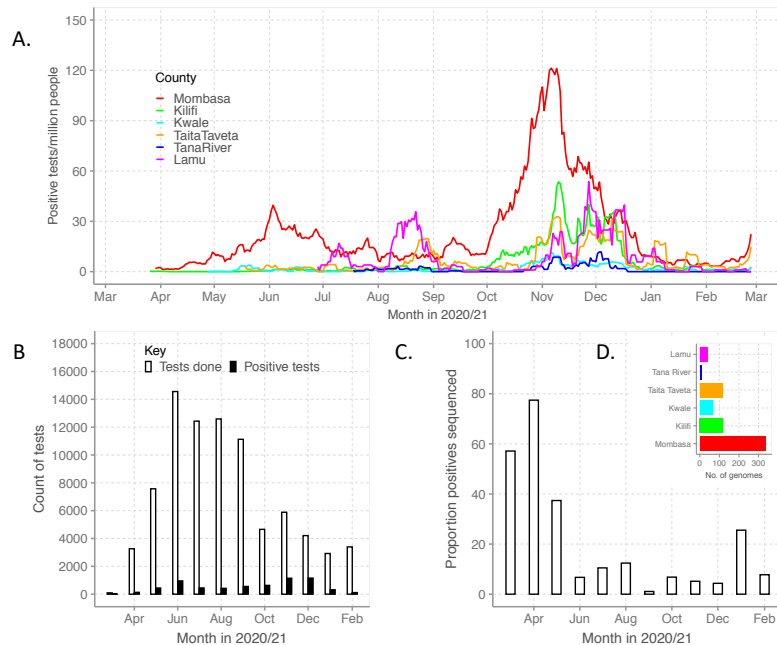
89 nasopharyngeal/oropharyngeal (NP/OP) swab samples from the six Coastal counties and 6,307
 90 (7.6%) of these were determined as SARS-CoV-2 positive (approximately half of the total

91 reported by MoH), distributed by month as in **Figure 1B**. Overall, a total of 3,137 (6.8%) tests

92 at KWTRP were positive from samples taken in Mombasa, 1,142 (8.8%) for Kilifi, 657 (4.5%)

93 for Taita Taveta, 349 (12.7%) for Lamu, 437 (7.9%) for Kwale and 105 (12.0%) for Tana River.

94



95 **Figure 1. SARS-CoV-2 testing and sequencing on the Kenyan Coast.** Panel A, epidemic
 96 curves for each of the six Coastal Kenya counties derived from the daily positive case numbers
 97 reported by the Ministry of Health. Panel B, monthly count of SARS-CoV-2 RT-PCR tests
 98 undertaken at the KEMRI-Wellcome Trust Research Programme (KWTRP) and those positive
 99 during the study period. Panel C, monthly proportion of samples sequenced from total SARS-
 100 CoV-2 positives detected at KWTRP. Panel D, Distribution of the sequenced 684 samples across
 101 the 6 Coastal Kenya counties.
 102

103

104 We sequenced the genomes of 684 (10.8%) RT-PCR positive samples as follows: Mombasa: 337
105 (10.7%), Kilifi: 118 (8.1%), Taita Taveta: 115 (13.4%), Kwale: 65 (14.9%), Lamu: 40 (11.5%)
106 and Tana River: 9 (8.5%), **Figure 1C and 1D**. The sequences were spread across the study
107 period with 272 (39.8%) from the introduction phase (17th March to 20th May 2020 when
108 nationally <50 new cases were reported per day), 169 (24.7%) from wave one (21st May-15th
109 September 2020) and 243 (35.5%) from wave two (16th September 2020 to 26th February 2021),
110 **S1 Figure**. Overall, our sequencing rate corresponded to approximately one sequence for every
111 18 confirmed cases.

112
113

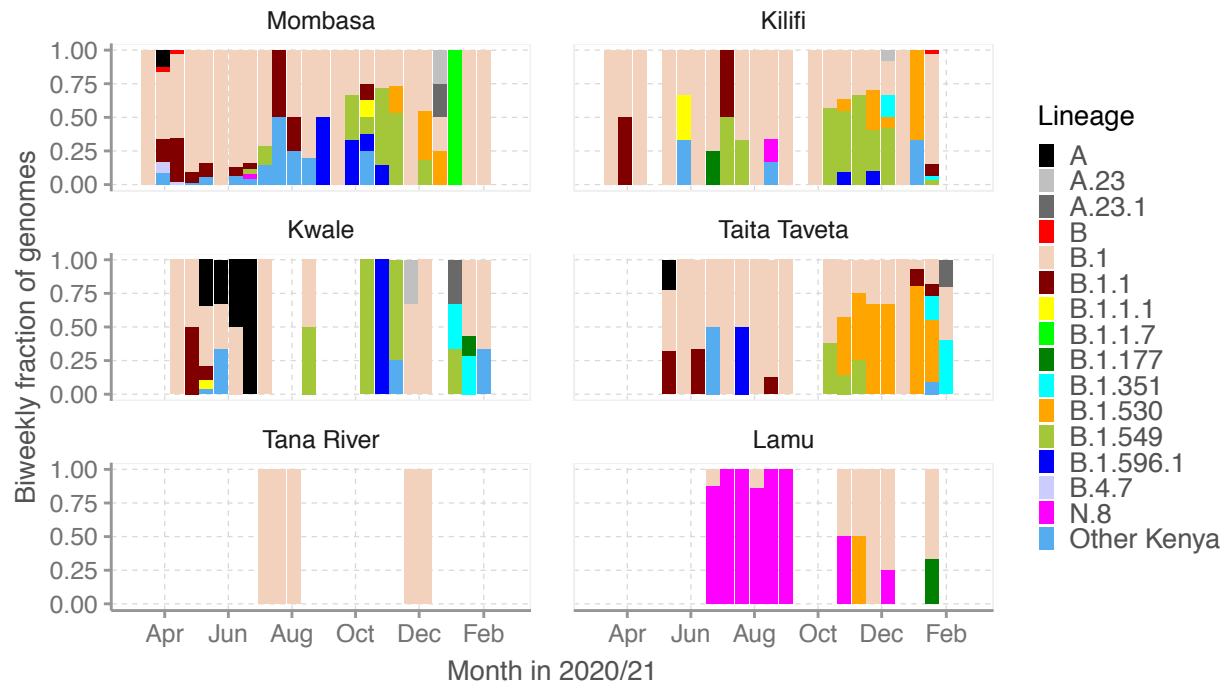
114 *Lineage dynamics in Coastal Kenya*

115 The 684 genomes were classified into 32 Pango lineages, including four first identified in Kenya
116 (N.8, B.1.530, B.1.549 and B.1.596.1), **S1 Table**. During the introductions phase we detected 12
117 lineages while during wave one we detected 16 lineages (11 of which were being identified
118 locally for the first time) and during wave two we detected 18 lineages (nine of which were
119 being identified locally for the first time). Thirteen lineages (40.6%) were identified in three or
120 more samples with the top six accounting for 88.0% of the sequenced infections: B.1 (n=413,
121 60.4%, European lineage which predominated in the Northern Italy outbreak early in 2020),
122 B.1.1 (n=61, 8.9%, European lineage), B.1.549 (n=54, 7.9%, Kenyan lineage), B.1.530 (n=44,
123 6.4%, Kenyan lineage), N.8 (n=30, 4.4%, Kenyan lineage) and A (n=21, 3.1%, lineage root to
124 the pandemic first observed in China).

125

126 Lineage B.1 was the first to be detected in the region and comprised the initial cases identified in
127 Mombasa, Kilifi, Kwale and Tana River, **Figure 2**. This lineage comprised 71.3% (n=194) of the
128 viruses we sequenced from the introduction phase, 65.7% (n=111) from wave one and 44.4%
129 (n=108) from wave two. Overall, B.1 accounted for most of the virus sequences detected in all
130 counties except Lamu where lineage N.8 (alias lineage B.1.1.33.8) was most frequent (70% of
131 the cases, first detected there on 23rd June 2020). Only two additional sequenced cases of
132 Lineage N.8 occurred outside Lamu County (one in Mombasa and another in Kilifi).

133



134

135 **Figure 2. Biweekly distribution of the common Pango lineages identified on the Kenyan**
 136 **coast from March 2020 to February 2021 across the six counties presented as proportions.**

137 Lineages detected in less than three cases were put together and referred to as “other lineages”.
 138 This group included 18 lineages namely: A.25, B.1.1.254, B.1.179, B.1.378, B.1.393, B.1.1.111,
 139 B.1.1.33, B.1.1.485, B.1.1.7, B.1.1.70, B.1.153, B.1.340, B.1.450, B.1.526, B.1.535, B.1.553,
 140 B.1.558 and B.4 (see details in S1 Table).

141

142 Although lineage A was also detected in Mombasa and Taita Taveta, most of its cases occurred
 143 in Kwale where it comprised 20% of all cases identified in the county, **Figure 2.** Lineage B.1.1
 144 was mostly observed in Mombasa with limited number of cases observed in Kilifi, Kwale and
 145 Taita Taveta.

146

147 During the wave two period in Coastal Kenya, there was an emergence of three lineages first
 148 identified in Kenya: (a) lineage B.1.549 (first observed 26th June 2020) which was responsible
 149 for a significant number of the sequenced infections in Kilifi (n=24), Mombasa (n=18), Taita-
 150 Taveta (n=5), and Kwale (n=7), (b) lineage B.1.596.1 (first observed 01st August 2020) and
 151 detected in Mombasa, Kilifi, Kwale and Taita-Taveta and (c) lineage B.1.530 (first observed
 152 28th October 2020) that predominated infections in Taita-Taveta (n=28) with a fewer cases
 153 observed in Mombasa (n=8), Kilifi (n=7) and Lamu (n=1).

154

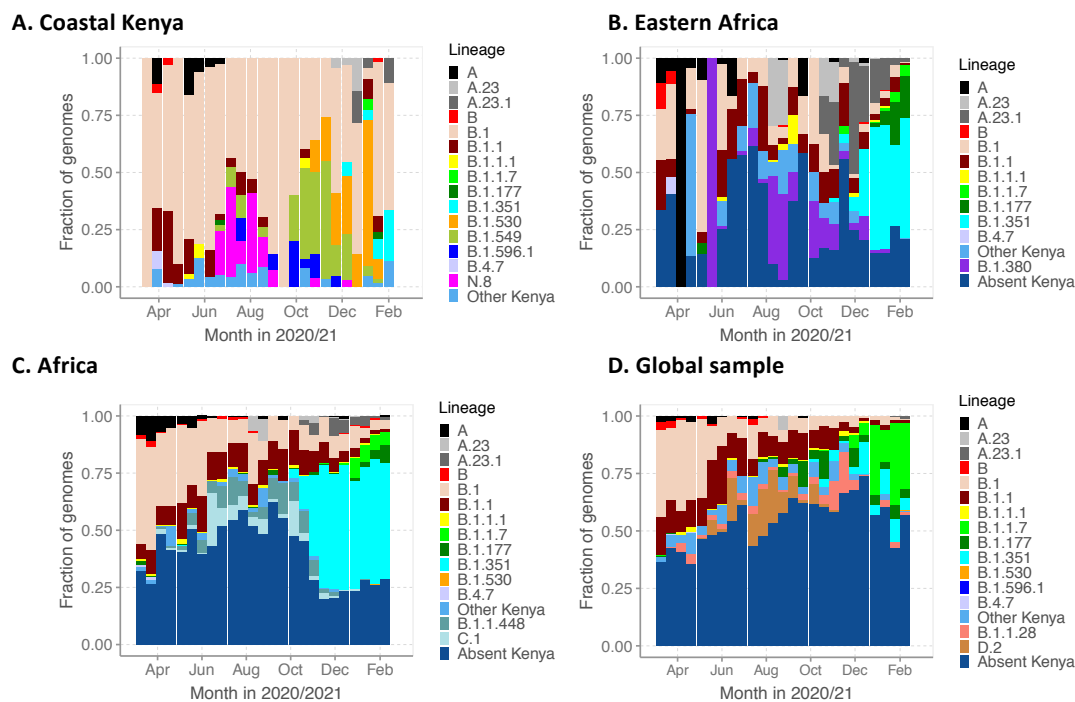
155 Two known variants of concern (VOC) were detected during the study period: the lineage
 156 B.1.351 (the Beta variant in 10 samples; and the lineage B.1.1.7 (the Alpha variant in one sample

157 (12). The earliest B.1.351 infections were identified in international travelers from South Africa
 158 in mid-December 2020 while the earliest detected B.1.1.7 case was in second week of January
 159 2020 in a local with no history of recent international travel. We also detected three cases each of
 160 A.23 and A.23.1 after the peak of wave two (both lineages first identified in Uganda, and are
 161 considered variants of interest (VOI) in the region (13)).

162

163 **Lineage dynamics with a widening scale of observation.**

164 We compared the temporal prevalence patterns of the lineages identified in Coastal Kenya to
 165 Eastern Africa, Africa and a global sub-sample, **Figure 3**. In these comparative genome sets, we
 166 identified 88 Pango lineages for Eastern Africa, 247 for Africa and 571 for the global sub-
 167 sample.



168
 169 **Figure 3. Lineage dynamics at widening scales of observation.** Panel A, Biweekly distribution
 170 of detected lineages in Coastal Kenya from the sequenced 684 genomes. Panel B, biweekly
 171 distribution of detected lineages in Eastern Africa from 2,316 contemporaneous genomes from
 172 10 countries whose contemporaneous data were available in GISAID. Lineages not observed in
 173 the Coastal Kenya dataset have been lumped together as “Absent Kenya” but with the those
 174 appearing among top six and not observed in Kenya indicated just below the other Kenya group.
 175 Panel C, weekly distribution of detected lineages in Africa from 12,506 contemporaneous
 176 genomes from 39 countries that were available in GISAID. Panel D, weekly distribution of
 177 detected lineages in a global sub-sample of 9,989 contemporaneous genomes from 112 countries
 178 that were compiled from GISAID (see detail in methods section).

179
180 In the Eastern Africa sample (n=2,315), all the top lineages were also observed in the Coastal
181 Kenya data (B.1.351, A.23.1, B.1, B.1.1, B.1.177) except lineage B.1.380 (n=128, a Rwanda
182 lineage), **Figure 3B**. In the Africa sample (n=12,506), four of the six top lineages (B.1, B.1.1,
183 B.1.351 and B.1.1.7) were also identified in Coastal Kenya. The undetected two lineages were
184 B.1.1.448 (n=392) and C.1 (n=330) both of which were mainly observed South Africa samples.
185 In the global sub-sample (n=8,909) four of the top six lineages (B.1, B.1.1, B.1.1.7 and B.1.177)
186 where observed in the samples from Coastal Kenya. The unobserved lineages were D.2 (Alias of
187 B.1.1.25.2, an Australian lineage) and B.1.1.28 (a Brazilian lineage).

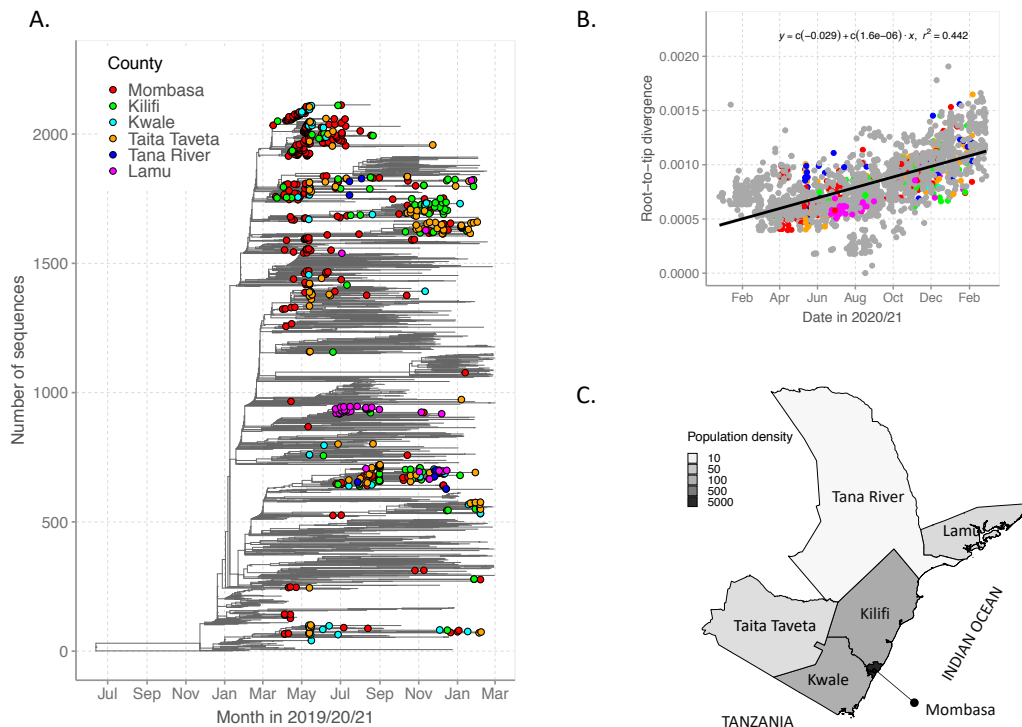
188
189 At the time of the second wave in Kenya, the VOC (B.1.1.7 and B.1.351) were already
190 widespread across Eastern Africa and Africa but there were only sporadic detections our dataset
191 from Coastal Kenya. Although in the comparison data lineage B.1 occurred in substantial
192 proportions across the different scales early in the pandemic, its prevalence diminished faster
193 overtime in the outside Kenya sample sets when compared to the Coastal Kenya dataset set.
194 Approximately 60% of the lineages comprising infections globally were not seen in the Coastal
195 Kenya samples.

196

197 ***SARS-CoV-2 diversity in Coastal Kenya***

198 When we reconstructed time-resolved ML phylogeny of the Coastal Kenya genomes while
199 including a global reference set of genomes (n=1500) we observed that (i) the Coastal Kenya
200 genomes were represented across many but not all of the major phylogenetic clusters observed
201 globally, ii) some of the coast clusters expanded after introduction whereas others did not
202 (observed as singletons) and iii) all counties appeared to have had multiple variants introduced
203 with some clusters comprising genomes detected across multiple counties, **Figure 4A**. There
204 was considerable correlation of the root-to-tip genetic distance and the date of sampling
205 ($r^2=0.442$), **Figure 4B**.

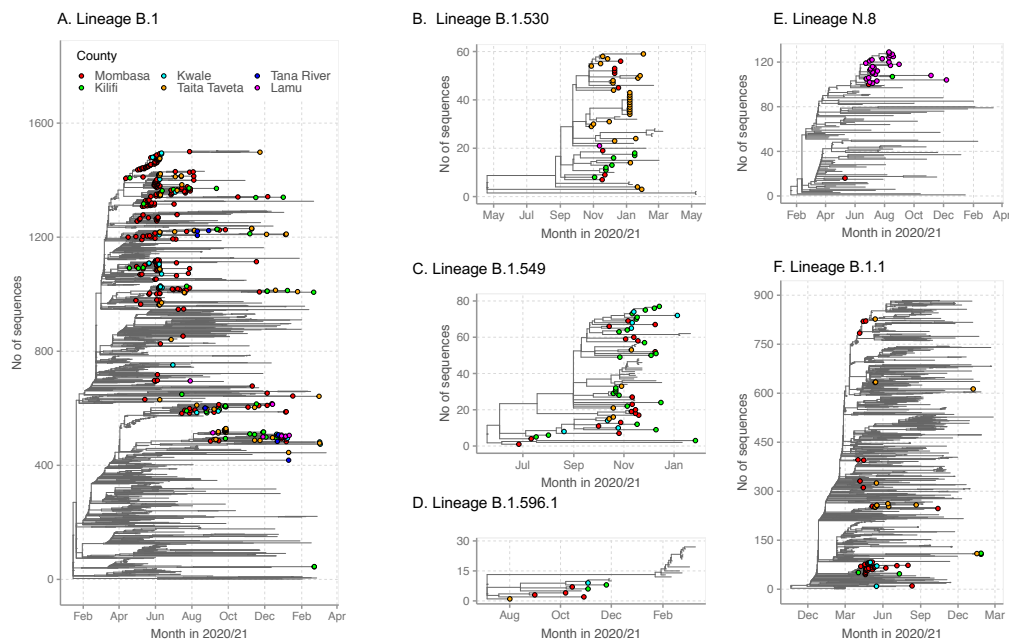
206



207
208 **Figure 4. SARS-CoV-2 diversity in Coastal Kenya.** Panel A, Time-resolved global phylogeny
209 that combined 639 Coastal Kenya SARS-CoV-2 genomes and 1474 global reference sequences.
210 Panel B, Roo-to-tip regression analysis coastal Kenya (colored by county) and global (shown in
211 grey) sequences with determination of co-efficient (r^2). Panel C, map of the six study counties
212 colored scaled by the population density (number of persons per square kilometer).

213
214 For detailed investigation into the local SARS-CoV-2 genetic diversity, we reconstructed time-
215 resolved lineage specific phylogenetic trees are shown in **Figure 5**. Viruses occurring within
216 lineage B.1 and lineage B.1.1 showed significant genetic divergence and formed multiple
217 phylogenetic clusters interspersed with other global sequences, a feature indicative of multiple
218 viral introductions into a geographic area. Three lineages first identified in Kenya (B.1.530,
219 B.1.549 and B.1.596.1) were found to (a) possess significant diversity consistent with widescale
220 spread within Kenya (**Figure 5B-D**), (b) formed multiple county-specific sub-clusters and (c)
221 show local sequences interspersed with global comparison genomes from the same lineage
222 implying potentially export (or import) events, **Figure 5C**. However, the picture painted by
223 lineage N.8 was different. This lineage was mainly detected in Lamu forming a single
224 monophyletic group (**Figure 5E**) when co-analysed with its precursor lineage B.1.1.33, an
225 observation consistent with a single introduction then expansion.

226

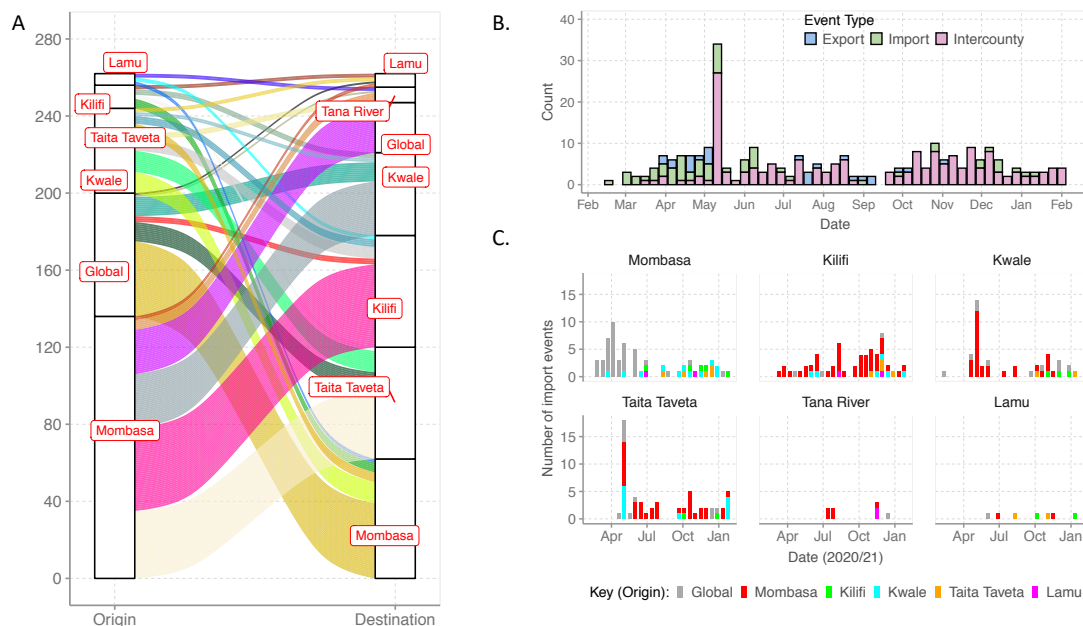


227
 228 **Figure 5.** Time-resolved lineage-specific phylogenetic trees for the commonly detected lineages
 229 in Coastal Kenya. The Coastal Kenya genomes are indicated with filled circles colored by
 230 county. Panel A, phylogeny for lineage B.1 that combined 369 Coastal Kenya genomes and 1134
 231 global reference sequences. Panel B, phylogeny for lineage B.1.530 that combined 44 Coastal
 232 Kenya genomes and 15 global sequences belonging also to B.1.530. Panel C, phylogeny for
 233 lineage B.1.549 that combined 54 Coastal Kenya genomes and 23 sequences from other
 234 countries of lineage B.1.549. Panel D, phylogeny for lineage B.1.596.1 combines 8 Coastal
 235 Kenya genomes and 19 global reference sequences. Panel E, phylogeny for lineage N.8 that
 236 combined 30 Coastal Kenya genomes of lineage N.8, 1 Coastal Kenya genome of lineage
 237 B.1.1.33 and 99 global sequences of lineage B.1.1.33. Panel F, phylogeny for lineage B.1.1 that
 238 combined 60 Coastal Kenya genomes and 824 global sequences belonging to lineage B.1.1.

239
 240 ***Viral imports and export from Coastal Kenya***

241 We used ancestral location state reconstruction of the dated phylogeny (**Figure 4**) to infer the
 242 number of viral import and exports (see details in the methods (14)). In total, between March
 243 2020 and February 2021, we detected 262 location transition events, 64 of which were viral
 244 import events into Coastal Kenya from outside and 26 were viral exports events from Coastal
 245 Kenya to outside populations, **Figure 6A-C**. A total of 172 import/export events were detected
 246 between the six Coastal counties. Virus imports into the region occurred through Mombasa
 247 (n=39, 61%), Kwale (n=10, 16%), Taita Taveta (n=10, 16%), Kilifi (n=3, 5%) and Lamu (n=1,
 248 2%). Virus export from the region occurred through Mombasa (n=23, 88%) and Kwale (12%).
 249 Of the 64 detected viral imports from outside the Coastal Kenya region, 40 (63%) occurred
 250 during the introductions phase, 12 (19%) occurred during wave one and another 12 (19%) during

251 wave two. Of the 26 detected viral exports, 13 (50%) occurred during introductions phase, 10
 252 (38%) during wave one and three (12%) during wave two, **Figure 6B and D**.
 253



254
 255 **Figure 6. Viral imports and export from Coastal Kenya.** Panel A, the number of imports and
 256 exports of viral lineages from Coastal Kenya shown as an alluvium plot. Panel B, biweekly
 257 location transition events observed in the Coastal Kenya and global genomes shown as either
 258 viral export, import or inter-county transmission. Panel C, Biweekly number of import events
 259 into the individual Coastal counties and their origins.

260
 261 ***Amino acid changes in the Coastal Kenya genomes***

262 We scanned the entire SARS-CoV-2 genome for amino acid changes from the Wuhan reference
 263 genome in the 684 Coastal Kenya sequences. Below we summarize changes that were identified
 264 in ≥ 5 genomes. We observed 20 changes in the spike (S) protein meeting the set criterion, 15 in
 265 the nucleocapsid (N) protein, 14 in ORF3a, 13 in ORF1b, nine in the ORF1a, seven in ORF8,
 266 five in ORF6, and one in the envelope (E) protein. No codon changes, meeting of criterion was
 267 observed in the matrix (M) protein, ORF7a, ORF7b or ORF10. The most common amino acid
 268 changes observed in the Coastal Kenya genomes were S: D614G (94.4%), ORF1b: P314L
 269 (92.9%), ORF3a: Q57H (15.5%), ORF1a: T265I (15.4%) and N: R195K (12.9%), **S2 Figure**.
 270 Several lineage-defining mutations were observed for the Kenya-specific lineages as summarized
 271 in **S3 Figure**.

272
 273 **DISCUSSION**

274 We present evidence of continuous introductions of SARS-CoV-2 into Coastal Kenya
275 documenting at least 64 independent viral introduction events into the region during the first year
276 of the pandemic. Multiple variant introductions were observed even at the individual counties
277 level, with extensive inter-county transmission events after the early phase period, re-affirming
278 that Coastal Kenya populations are quite linked epidemiologically. Strikingly, many of the
279 imports and exports from the region occurred through Mombasa, a major commercial, industrial
280 and tourist destination in the region, before dispersing to the other counties. This observation
281 highlights Mombasa's central role as a gateway of viral variants into the region and the
282 significance of enhanced and persistent surveillance in this city for early detection of introduced
283 of variants of concern.

284
285 The analyzed counties had both differences and similarities in SARS-CoV-2 transmission
286 patterns. While Mombasa, Taita Taveta and Lamu counties had experienced two waves of
287 infections by February 2021, Kilifi, Tana River and Kwale counties had experienced only a
288 single major wave. This may have been brought about by differences in population density,
289 accessibility, and applied infection control measures. For instance, Mombasa is densely
290 populated (~5,604.64 persons/sq.km, 2019 census), has a large seaport and an international
291 airport while Lamu has a sparsely populated mainland and over 65 islands (total population
292 143,920, population density, 21.84 persons/sq.km). Taita Taveta and Kwale counties are
293 primarily rural counties bordering Tanzania which had a largely uncontrolled epidemic (15).
294 Tana River is the most remote and sparsely populated (8.9 persons/sq.km) county in the region
295 (**Figure 4C**).

296
297 During the early months of the pandemic several Pango lineages were introduced into the region
298 (A, B.1, B.1.1, B.1.340, B.1.535, B.4 and B.4.7). However, it is lineage B.1 which has a
299 European origin, that became predominant. This lineage has the D614G change in its spike
300 protein considered to enhance fitness (16) and this may have boosted its local transmission and
301 early dominance. Lineage N.8 was specific to Lamu County and B.1.530 to Taita Taveta.
302 Lineage N.8 has not been observed outside Kenya and its precursor lineage (B.1.1.33) was
303 observed earlier in South America, mainly Brazil. Lineage N.8 may have arisen from a single
304 introduction event into Lamu county, and then remained localized to Lamu. The N.8 lineage has
305 seven characteristic lineage defining mutations including S: D614G and R203K, G204R and
306 I292T on the nucleocapsid protein.

307
308 Kilifi county, which neighbors Mombasa to the North (~1.4 million with a population density of
309 117.67 persons/sq.km, 2019 census) observed its first peak of infections during the second
310 national wave. The Kilifi first wave was comprised of mainly three lineages: B.1 (53.5%),
311 B.1.530 (8.1%) and B.1.549 (25.6%). Lineages B.1.530 and B.1.549, first identified in Kenya,
312 and may have arisen from local evolution of B.1. Lineage B.1.530 has six characteristic
313 mutations including spike P681H change adjacent to the biologically important furin cleavage

314 site while lineage B.1.549 has seven characteristic mutations, five occurring in the ORF1a or
315 ORF1b. Both two lineages have now been observed in a few other countries albeit in small
316 numbers: seven countries for B.1.530 (Rwanda, Netherlands, Germany, Denmark, USA, Japan
317 and Australia) and three countries for B.1.549 (England, USA, Canada).

318
319 Near the tail end of the national wave two, we detected only sporadic cases of the VOC B.1.351
320 and B.1.1.7. Lineage B.1.351 was first to be observed, initially in Kilifi in two asymptomatic
321 international travelers in mid-December 2020. Lineage B.1.1.7 was subsequently detected in a
322 local who presented to a Mombasa clinic in the second week of January 2021. In the subsequent
323 weeks up to the end of the period covered by this analysis (February 2021), no additional B.1.1.7
324 was detected unlike lineage B.1.351 which continued to be detected sporadically in January and
325 February 2021, especially in truckers screened at Tanzania border points of entry in Taita Taveta
326 and Kwale counties.

327
328 When we compared the local lineage patterns to the global patterns, the VOC were already
329 extensively spread across Eastern Africa (B.1.351 VOC), Africa (B.1.351 VOC) and worldwide
330 (B.1.1.7 VOC) in the last quarter of 2020. This implies that a lag occurred in their arrival into
331 Coastal Kenya perhaps due to public health measures that have remained in place during the
332 period especially at the international borders. It is also notable several lineages were detected
333 globally (>500) but only a small fraction (<10%) of this were documented locally.

334
335 Our study contributes to the limited but growing literature illuminating SARS-CoV-2
336 transmission patterns in Africa (11, 13, 17-19). The patterns revealed have potential to inform
337 pandemic mitigation strategies. The key limitations of our analysis include, first, the samples we
338 sequenced were only those available to us through the rapid response teams (RRTs) whose case
339 identification protocols were altered at the different study phases following guidance from the
340 MoH. Second, we only sequenced ~10% of the positive samples identified in our laboratory, the
341 majority from the introduction phase and were prioritizing samples with a Ct value of <30.0.
342 Third, the sampling across the six Coastal counties was not uniform, probably in part due to
343 varied distance from our testing centre located in Kilifi County but also the total number of
344 positive cases varied between the counties.

345
346 In conclusion, we show that the first two SARS-CoV-2 waves in Coastal Kenya observed
347 transmission of both newly introduced variants and potentially locally evolved variants. New
348 variant introductions appeared to mainly occur through Mombasa city. Strikingly, only a limited
349 number of the many introduced variants progressed to transmit extensively perhaps due to
350 ongoing public health interventions e.g., screening at ports of entry, case isolation and
351 quarantining of contacts of cases. Unlike in the global contemporaneous sample, we did not find
352 evidence of extensive local transmission of the global VOC during wave two. Thus, we infer that
353 it is more likely that the relaxation of some of the interventions (e.g., reopening of learning

354 institutions, airspace, bars and restaurants) that drove the second wave of infections. Overall, our
355 study shows the importance of detailed local genomic surveillance even in remote and under-
356 resourced settings to understand origins and spread patterns of SARS-CoV-2 to optimize local
357 interventions.

358

359 **MATERIALS AND METHODS**

360 ***Ethical statement***

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

367

368 ***Study period and population***

369 This study examined samples collected between 17th March 2020 to 26th February 2021 from six
370 counties in Coastal Kenya i.e Mombasa, Kilifi, Kwale, Tana River, Taita Taveta and Lamu. We
371 divided the study period into three phases based on trends in daily reported COVID-19 national
372 case numbers by the MoH (2): (a) Introduction phase – 17th March-20th May, 2020 (b) Wave one
373 – 21st May-15th September 2020 and (c) Wave two – 16th September to 28th February 2021, **S1**
374 **Figure**. We considered the transition from introduction phase to Wave as the timepoint when the
375 national daily positives exceeded 50 and the transition from Wave one to Wave two as the
376 timepoint when a consistent renewed rise of national daily number of positives started after
377 Wave one peak.

378

379 ***Patient samples***

380 This study analyzed SARS-CoV-2 positive NP/OP swab samples that were collected by the MoH
381 County Department of Health RRTs, across all the six counties of Coastal Kenya. SARS-CoV-2
382 positive diagnosis was made at the KWTRP in Kilifi County (20). The RRTs delivered the
383 NP/OP swabs to KWTRP laboratories within 48 hours of collection in cool boxes with ice packs.
384 The samples were from persons of any age and sampling followed the MOH eligibility criteria
385 that were revised from time to time (7). Persons sampled included those with acute respiratory
386 symptoms, those with a recent history of travel to the early COVID-19 hotspots, contacts of
387 confirmed cases, persons presenting at international border points seeking entry into Kenya,
388 among other criteria.

389

390 ***SARS-CoV-2 detection at KWTRP***

391 Viral RNA was extracted from the NP/OP samples using any one of seven commercial kits that
392 were available namely, QIAamp Viral RNA Mini Kit, RNeasy ® QIAcube ® HT Kit,
393 QIASYMPHONY ® RNA Kit, T0IANamp Virus RNA Kit, Da An Gene Nucleic acid Isolation

394 and Purification Kit, SPIN X Extraction and RADI COVID-19 detection Kit. The extracts were
395 tested for the presence of SARS-CoV-2 nucleic acid following different protocols depending on
396 which one was available of the following 7 kits/protocols: 1) the Berlin (Charité) primer-probe
397 set (targeting envelope (E) gene, nucleocapsid (N) or RNA-dependent RNA-polymerase
398 (RdRp)), 2) European Virus Archive – GLOBAL (EVA-g) (targeting E or RdRp genes), 3) Da
399 An Gene Co. detection Kit (targeting N or ORF1ab), 4) BGI RT-PCR kit (targeting ORF1ab), 5)
400 Sansure Biotech Novel Coronavirus (2019-nCoV) Nucleic Acid Diagnostic real-time RT-PCR
401 kit or 6) Standard M kit (targeting E and ORF1ab) and 7) TIB MOLBIOL kit (targeting E gene).
402 Protocol-specific recommended cycle threshold cut-offs were followed in defining SARS-CoV-2
403 positives.

404

405 ***SARS-CoV-2 genome sequencing***

406 Only samples that had a PCR cycle threshold value of <30.0 were targeted for whole genome
407 sequencing (7). Viral RNA from positive samples was reextracted using QIAamp Viral RNA
408 Mini kit following the manufacturer's instructions and reverse transcribed using LunaScript®
409 RT SuperMix Kit. The cDNA was amplified using Q5® Hot Start High-Fidelity 2x Mastermix
410 along with the ARTIC nCoV-2019 version 3 primers. The PCR products were run on an 1.5%
411 agarose gel and for samples whose SARS-CoV-2 amplification was considered successful were
412 purified using Agencourt AMPure XP beads and taken forward for library preparation.
413 Sequencing libraries were constructed using Oxford Nanopore Technology (ONT) ligation
414 sequencing kit and the ONT Native Barcoding Expansion kit as described in the ARTIC
415 protocol (21). Every MinION (Mk1B) run comprised 23 samples and one negative (no-template)
416 control.

417

418 ***SARS-CoV-2 genome assembly***

419 Following MinION sequencing, the FAST5 files were base-called using the OTN's software
420 Guppy v3.5-4.2. Consensus SARS-CoV-2 sequences were derived from the reads using the
421 ARTIC bioinformatics pipeline. A threshold of $\times 20$ coverage was required for a base to be
422 included in the consensus genome otherwise it was masked with an N. Only complete or near-
423 complete genomes with $>80\%$ coverage and $\times 20$ or more read depth were taken forward for
424 phylogenetic analysis.

425

426 ***Lineage assignment***

427 The consensus genomes were assigned into Pango lineages via the Phylogenetic Assignment of
428 named Global Lineages (PANGOLIN) software suite (Pangolin v2.4.2, Pango v1.1.23) (22). The
429 amino acid changes profiles of the Coastal Kenya genomes from the reference strains was
430 investigated using the nextclade tool v0.14.2 (23).

431

432 ***Global contextual sequences***

433 We prepared three sets of contextual sequences from GISAID deposited as of May 2021.
434 Selected sequences were those with non-ambiguous sampling date, sampled during the period
435 covered by our study and after alignment possessed <1500 ambiguous nucleotides i.e showed
436 >95% genome completeness.

- 437 a) **Set 1:** We compiled an African SARS-CoV-2 genome set which we assigned Pango
438 lineages and used to compare the continental lineage temporal patterns to the Coastal
439 Kenya lineage distribution. A total of 12,542 genomes were retrieved for this purpose on
440 the 11th of May 2021. From these we also created an Eastern Africa subset which
441 comprised of 2,316 genomes from 10 countries, namely, Zimbabwe, Zambia, Uganda,
442 Rwanda, Reunion (overseas territory), Mozambique, Malawi, Madagascar, Ethiopia and
443 Comoros.
- 444 b) **Set 2:** We compiled a global reference set of 9,989 genomes collected across all the six
445 inhabited continents between 24th December 2019 and 28th February 2021. Between 23
446 and 899 genomes were selected for each included month (mean of 667 and median of
447 701) randomly selected across all continents from across 112 countries. These genomes
448 were used to infer the global temporal patterns of the lineages observed in Coastal Kenya.
449 A subset of these genomes (n=1500) was combined with the Kenyan genomes to infer
450 their global phylogenetic context.
- 451 c) **Set 3:** We compiled a genotype reference for the top six observed lineages in Coastal
452 Kenya. For lineages B.1. and B.1.1 which >5000 genomes exist in GISAID were
453 prepared we used the sub-sample from Set 2 above. For lineages B.1.530, B.1.549,
454 B.1.596.1 and N.8 we retrieved all genomes assigned these lineages available from
455 GISAID to infer their global phylogenetic context.

456 457 *Phylogenetic analysis*

458 Multiple sequence alignments were prepared in Nextalign v 0.1.6 software and using the initial
459 Wuhan sequence (Accession number: NC_045512) as the reference. The alignment was
460 manually inspected in SEAVIEW v4.6.4 to spot any obvious misalignments. Quick non-
461 bootstrapped neighbor joining trees were created in SEAVIEW to identify any aberrant
462 sequences that were henceforth discarded.

463
464 We reconstructed maximum likelihood (ML) phylogenies using IQTREE v2.1.3. The software
465 initiates tree reconstruction after assessment and selection of the best model of nucleotide
466 substitution for the alignment. The ML trees were linked to the various metadata in R
467 programming software v4.0.2 and visualized using R ggTree v2.4.2. TempEst v1.5.3 was used to
468 assess the presence of a molecular clock signal in analysed data and linear regression of root-to-
469 tip genetic distances against sampling dates plotted.

470 471 *Import/export analysis*

472 The global ML tree topology was used to estimate the number of viral transmission events
473 between Coastal Kenya and the rest of the world as in (17). The software TreeTime was used to
474 transform the ML tree topology into a dated tree assuming a constant genomic evolutionary rate
475 of SARS-CoV-2 of 8.0×10^{-4} nucleotide substitutions per site per year (14). Outlier sequences
476 were identified by TreeTime and excluded during this process. A migration model was
477 henceforth fitted to the resulting time-scaled phylogenetic tree from TreeTime, mapping the
478 location status of the genomes from the six counties at both the tips and internal nodes. Using
479 the date and location annotated tree topology, we counted the number of transitions between and
480 within Coastal Kenya counties and the rest of the world and plotted this using ggplot2 v3.3.3.

481

482 ***Epidemiological data***

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

487

488 ***Kenya COVID-19 response***

489 We derived status of Kenya government COVID-19 interventions using data from Our World in
490 Data database that has calculated the Oxford Stringency Index (SI). These SI estimations are
491 composite measure based on nine response indicators implemented by governments including
492 school closures; workplace closures; cancellation of public events; restrictions on public
493 gatherings; closures of public transport; stay-at-home requirements; public information
494 campaigns; restrictions on internal movements; and international travel controls, rescaled to a
495 value 0-100, with 100 being strictest (24). The Kenya government revised the policy
496 interventions at approximately monthly intervals (6) and the changes in SI overtime are shown
497 in **S1 Figure**.

498

499 **Acknowledgements**

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

508

509 ***Members of COVID-19 Testing Team at KWTRP***

510 Agnes Mutiso, Alfred Mwanu, Angela Karani, Bonface M. Gichuki, Boniface Karia, Brian
511 Bartilol, Brian Tawa, Calleb Odundo, Caroline Ngetsa, Clement Lewa, Daisy Mugo, David
512 Amadi, David Ileri, Debra Riako, Domtila Kimani, Edwin Machanja, Elijah Gicheru, Elisha
513 Omer, Faith Gambo, Horace Gumba, Isaac Musungu, James Chemweno, Janet Thoya, Jedida
514 Mwacharo, John Gitonga, Johnstone Makale, Justine Getonto, Kelly Ominde, Kelvias Keter,
515 Lydia Nyamako, Margaret Nunah, Martin Mutunga, Metrine Tendwa, Moses Mosobo, Nelson
516 Ouma, Nicole Achieng, Patience Kiyuka, Perpetual Wanjiku, Peter Mwaura, Rita Warui,
517 Robinson Cheruiyot, Salim Mwarumba, Shaban Mwangi, Shadrack Mutua, Susan Njuguna,
518 Victor Osoti, Wesley Cheruiyot, Wilfred Nyamu, Wilson Gumbi and Yiakon Sein

519

520 **Data availability**

521 Coastal Kenya genomes have been deposited to both GISAID (Accession numbers:
522 EPI_ISL_1039223-29; EPI_ISL1440075-109, IPI_ISL_457845-931; EPI_ISL_568695-872;
523 EPI_ISL806548-717; EPI_ISL_855494-548; EPI_ISL_968807-9245).

524

525 **Funding**

526 This work was supported by the National Institute for Health Research (NIHR) (project
527 references 17/63/82 (PI Prof. James Nokes) and 16/136/33 (PI Prof. Mark Woolhouse) using
528 UK aid from the UK Government to support global health research, The UK Foreign,
529 Commonwealth and Development Office and Wellcome Trust (grant# 220985). The views
530 expressed in this publication are those of the author (s) and not necessarily those of NIHR, the
531 Department of Health and Social Care, Foreign Commonwealth and Development Office,
532 Wellcome Trust or the UK government. Some members of COVID-19 Testing Team at
533 KWTRP were supported by funding received by Dr Marta Maia (BOHEMIA study funded
534 UNITAID), Dr Francis Ndungu (Senior Fellowship and Research and Innovation Action (RIA)
535 grants from EDCTP) and Prof. Anthony Scott (PCIVS grant from GAVI).

536

537 **References**

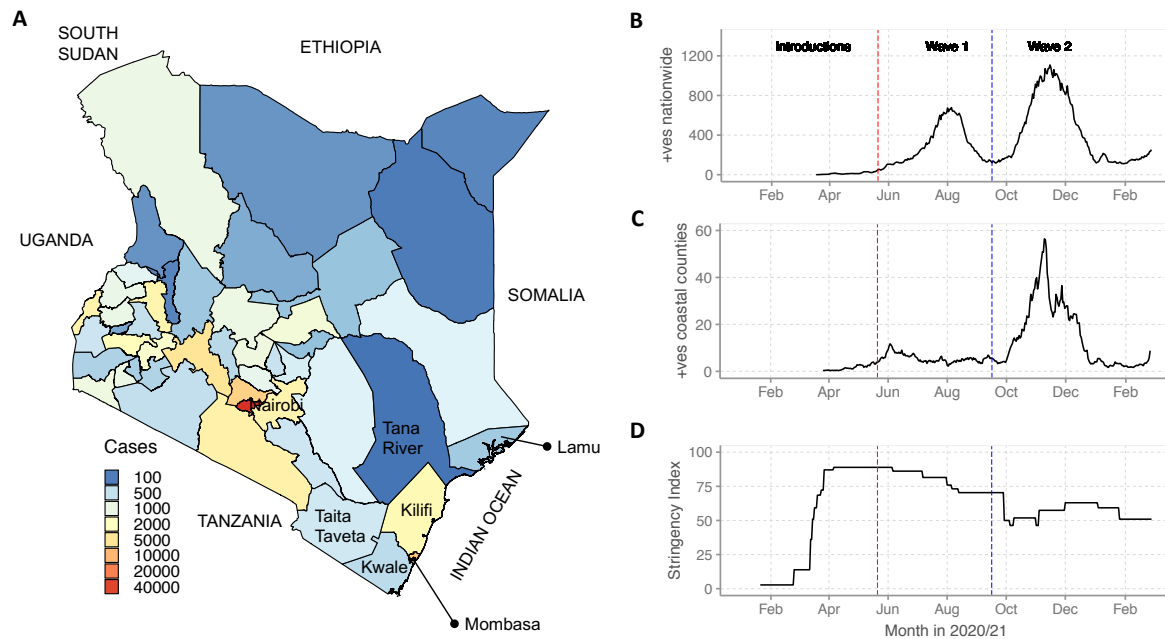
- 538 1. Dong E, Du H, Gardner L. 2020. An interactive web-based dashboard to track COVID-
539 19 in real time. *Lancet Infect Dis* 20:533-534.
- 540 2. MOH. 2021. Update on COVID-19 in Kenya doi:[https://www.health.go.ke/press-](https://www.health.go.ke/press-releases/)
541 [releases/](https://www.health.go.ke/press-releases/). MOH.
- 542 3. Uyoga S, Adetifa IMO, Karanja HK, Nyagwange J, Tuju J, Wanjiku P, Aman R,
543 Mwangangi M, Amoth P, Kasera K, Ng'ang'a W, Rombo C, Yegon C, Kithi K,
544 Odhiambo E, Rotich T, Orgut I, Kihara S, Otiende M, Bottomley C, Mupe ZN, Kagucia
545 EW, Gallagher KE, Etyang A, Voller S, Gitonga JN, Mugo D, Agoti CN, Otieno E,
546 Ndwiga L, Lambe T, Wright D, Barasa E, Tsofa B, Bejon P, Ochola-Oyier LI, Agweyu
547 A, Scott JAG, Warimwe GM. 2021. Seroprevalence of anti-SARS-CoV-2 IgG antibodies
548 in Kenyan blood donors. *Science* 371:79-82.

- 549 4. Etyang AO, Lucinde R, Karanja H, Kalu C, Mugo D, Nyagwange J, Gitonga J, Tuju J,
550 Wanjiku P, Karani A, Mutua S, Maroko H, Nzomo E, Maitha E, Kamuri E, Kaugiria T,
551 Weru J, Ochola LB, Kilimo N, Charo S, Emukule N, Moracha W, Mukabi D, Okuku R,
552 Ogutu M, Angujo B, Otiende M, Bottomley C, Otiemo E, Ndwiga L, Nyaguara A, Voller
553 S, Agoti C, Nokes DJ, Ochola-Oyier LI, Aman R, Amoth P, Mwangangi M, Kasera K,
554 Ng'ang'a W, Adetifa I, Kagucia EW, Gallagher K, Uyoga S, Tsofa B, Barasa E, Bejon P,
555 Scott JAG, Agweyu A, Warimwe G. 2021. Seroprevalence of Antibodies to SARS-CoV-
556 2 among Health Care Workers in Kenya. *Clin Infect Dis* doi:10.1093/cid/ciab346.
- 557 5. Adetifa IMO, Uyoga S, Gitonga JN, Mugo D, Otiende M, Nyagwange J, Karanja HK,
558 Tuju J, Wanjiku P, Aman R, Mwangangi M, Amoth P, Kasera K, Ng'ang'a W, Rombo C,
559 Yegon C, Kithi K, Odhiambo E, Rotich T, Orgut I, Kihara S, Bottomley C, Kagucia EW,
560 Gallagher KE, Etyang A, Voller S, Lambe T, Wright D, Barasa E, Tsofa B, Bejon P,
561 Ochola-Oyier LI, Agweyu A, Scott JAG, Warimwe GM. 2021. Temporal trends of
562 SARS-CoV-2 seroprevalence during the first wave of the COVID-19 epidemic in Kenya.
563 *Nature Communications* 12:3966.
- 564 6. Brand SPC, Ojal J, Aziza R, Were V, Okiro EA, Kombe IK, Mburu C, Ogero M,
565 Agweyu A, Warimwe GM, Nyagwange J, Karanja H, Gitonga JN, Mugo D, Uyoga S,
566 Adetifa IMO, Scott JAG, Otiemo E, Murunga N, Otiende M, Ochola-Oyier LI, Agoti CN,
567 Githinji G, Kasera K, Amoth P, Mwangangi M, Aman R, Ng'ang'a W, Tsofa B, Bejon P,
568 Keeling MJ, Nokes DJ, Barasa E. 2021. COVID-19 Transmission Dynamics Underlying
569 Epidemic Waves in Kenya. *medRxiv*
570 doi:10.1101/2021.06.17.21259100:2021.06.17.21259100.
- 571 7. Githinji G, de Laurent ZR, Mohammed KS, Omuoyo DO, Macharia PM, Morobe JM,
572 Otiemo E, Kinyanjui SM, Agweyu A, Maitha E, Kitole B, Suleiman T, Mwakinangu M,
573 Nyambu J, Otiemo J, Salim B, Kasera K, Kiiru J, Aman R, Barasa E, Warimwe G, Bejon
574 P, Tsofa B, Ochola-Oyier LI, Nokes DJ, Agoti CN. 2020. Tracking the introduction and
575 spread of SARS-CoV-2 in coastal Kenya. *medRxiv*
576 doi:10.1101/2020.10.05.20206730:2020.10.05.20206730.
- 577 8. Geoghegan JL, Ren X, Storey M, Hadfield J, Jelley L, Jefferies S, Sherwood J, Paine S,
578 Huang S, Douglas J, Mendes FK, Sporle A, Baker MG, Murdoch DR, French N,
579 Simpson CR, Welch D, Drummond AJ, Holmes EC, Duchêne S, de Ligt J. 2020.
580 Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in
581 Aotearoa New Zealand. *Nature Communications* 11:6351.
- 582 9. Worobey M, Pekar J, Larsen BB, Nelson MI, Hill V, Joy JB, Rambaut A, Suchard MA,
583 Wertheim JO, Lemey P. 2020. The emergence of SARS-CoV-2 in Europe and North
584 America. *Science* 370:564-570.
- 585 10. Oude Munnink BB, Nieuwenhuijse DF, Stein M, O'Toole Á, Haverkate M, Mollers M,
586 Kamga SK, Schapendonk C, Pronk M, Lexmond P, van der Linden A, Bestebroer T,
587 Chestakova I, Overmars RJ, van Nieuwkoop S, Molenkamp R, van der Eijk AA,
588 GeurtsvanKessel C, Vennema H, Meijer A, Rambaut A, van Dissel J, Sikkema RS,

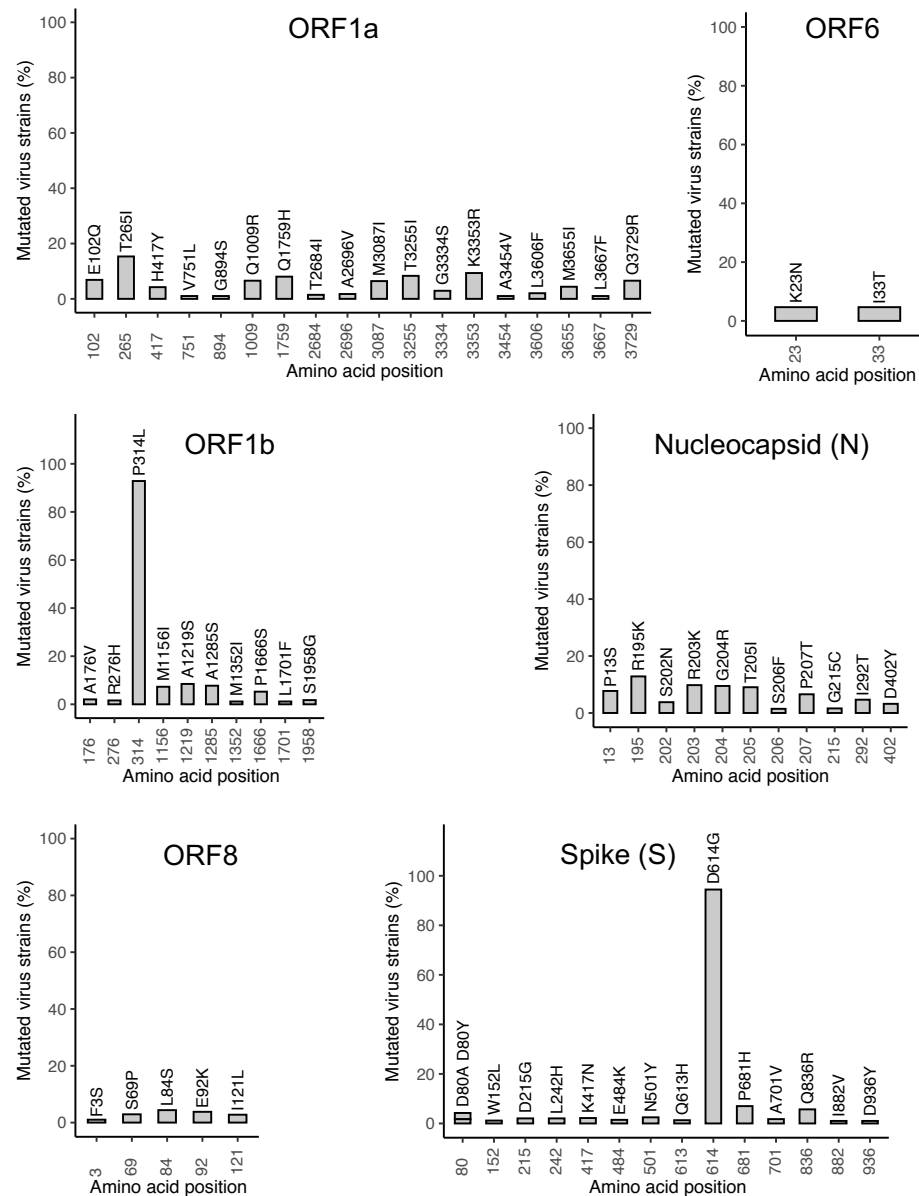
- 589 Timen A, Koopmans M. 2020. Rapid SARS-CoV-2 whole-genome sequencing and
590 analysis for informed public health decision-making in the Netherlands. *Nat Med*
591 26:1405-1410.
- 592 11. Bugembe DL, Kayiwa J, Phan MVT, Tushabe P, Balinandi S, Dhaala B, Lexow J,
593 Mwebesa H, Aceng J, Kyobe H, Ssemwanga D, Lutwama J, Kaleebu P, Cotten M. 2020.
594 Main Routes of Entry and Genomic Diversity of SARS-CoV-2, Uganda. *Emerg Infect*
595 *Dis* 26:2411-2415.
- 596 12. Konings F, Perkins MD, Kuhn JH, Pallen MJ, Alm EJ, Archer BN, Barakat A, Bedford
597 T, Bhiman JN, Caly L, Carter LL, Cullinane A, de Oliveira T, Druce J, El Masry I, Evans
598 R, Gao GF, Gorbalenya AE, Hamblion E, Herring BL, Hodcroft E, Holmes EC, Kakkar
599 M, Khare S, Koopmans MPG, Korber B, Leite J, MacCannell D, Marklewitz M, Maurer-
600 Stroh S, Rico JAM, Munster VJ, Neher R, Munnink BO, Pavlin BI, Peiris M, Poon L,
601 Pybus O, Rambaut A, Resende P, Subissi L, Thiel V, Tong S, van der Werf S, von
602 Gottberg A, Ziebuhr J, Van Kerkhove MD. 2021. SARS-CoV-2 Variants of Interest and
603 Concern naming scheme conducive for global discourse. *Nature Microbiology*
604 doi:10.1038/s41564-021-00932-w.
- 605 13. Bugembe DL, Phan MVT, Ssewanyana I, Semanda P, Nansumba H, Dhaala B, Nabadda
606 S, O'Toole AN, Rambaut A, Kaleebu P, Cotten M. 2021. Emergence and spread of a
607 SARS-CoV-2 lineage A variant (A.23.1) with altered spike protein in Uganda. *Nature*
608 *Microbiology* doi:10.1038/s41564-021-00933-9.
- 609 14. Sagulenko P, Puller V, Neher RA. 2018. TreeTime: Maximum-likelihood phylodynamic
610 analysis. *Virus Evol* 4:vex042.
- 611 15. Rice BL, Annapragada A, Baker RE, Bruijning M, Dotse-Gborgbortsi W, Mensah K,
612 Miller IF, Motaze NV, Raheerinandrasana A, Rajeev M, Rakotonirina J, Ramiadantsoa T,
613 Rasambainarivo F, Yu W, Grenfell BT, Tatem AJ, Metcalf CJE. 2021. Variation in
614 SARS-CoV-2 outbreaks across sub-Saharan Africa. *Nature Medicine* 27:447-453.
- 615 16. Baric RS. 2020. Emergence of a Highly Fit SARS-CoV-2 Variant. *New England Journal*
616 *of Medicine* 383:2684-2686.
- 617 17. Wilkinson E, Giovanetti M, Tegally H, San JE, Lessells R, Cuadros D, Martin DP, Zekri
618 A-RN, Sangare AK, Ouedraogo A-S, Sesay AK, Hammami A, Amuri AA, Sayed A,
619 Rebai A, Elargoubi A, Trotter AJ, Keita AK, Sall AA, Kone A, Souissi A, Gutierrez AV,
620 Page AJ, Iranzadeh A, Lambisia A, Sylverken A, Ibrahimi A, Dhaala B, Kouriba B,
621 Kleinhans B, Brook C, Williamson C, Pratt CB, Akoua-Koffi CG, Agoti CN, Morang'a
622 CM, Nokes DJ, Bridges DJ, Bugembe DL, Baker D, Doolabh D, Ssemwanga D,
623 Tshabuila D, Bassirou D, Amuzu DSY, Goedhals D, Maruapula D, Foster-Nyarko E,
624 Lusamaki EK, Simulundu E, et al. 2021. A year of genomic surveillance reveals how the
625 SARS-CoV-2 pandemic unfolded in Africa. *medRxiv*
626 doi:10.1101/2021.05.12.21257080:2021.05.12.21257080.
- 627 18. Tegally H, Wilkinson E, Lessells RJ, Giandhari J, Pillay S, Msomi N, Mlisana K,
628 Bhiman JN, von Gottberg A, Walaza S, Fonseca V, Allam M, Ismail A, Glass AJ,

- 629 Engelbrecht S, Van Zyl G, Preiser W, Williamson C, Petruccione F, Sigal A, Gazy I,
630 Hardie D, Hsiao NY, Martin D, York D, Goedhals D, San EJ, Giovanetti M, Lourenço J,
631 Alcantara LCJ, de Oliveira T. 2021. Sixteen novel lineages of SARS-CoV-2 in South
632 Africa. *Nat Med* 27:440-446.
- 633 19. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, Doolabh D,
634 Pillay S, San EJ, Msomi N, Mlisana K, von Gottberg A, Walaza S, Allam M, Ismail A,
635 Mohale T, Glass AJ, Engelbrecht S, Van Zyl G, Preiser W, Petruccione F, Sigal A,
636 Hardie D, Marais G, Hsiao NY, Korsman S, Davies MA, Tyers L, Mudau I, York D,
637 Maslo C, Goedhals D, Abrahams S, Laguda-Akingba O, Alisoltani-Dehkordi A, Godzik
638 A, Wibmer CK, Sewell BT, Lourenço J, Alcantara LCJ, Kosakovsky Pond SL, Weaver S,
639 Martin D, Lessells RJ, Bhiman JN, Williamson C, de Oliveira T. 2021. Detection of a
640 SARS-CoV-2 variant of concern in South Africa. *Nature* 592:438-443.
- 641 20. Agoti C, Mutunga M, Lambisia A, Kimani D, Cheruiyot R, Kiyuka P, Lewa C, Gicheru
642 E, Tendwa M, Said Mohammed K, Osoti V, Makale J, Tawa B, Odundo C, Cheruiyot W,
643 Nyamu W, Gumbi W, Mwacharo J, Nyamako L, Otieno E, Amadi D, Thoya J, Karani A,
644 Mugo D, Musyoki J, Gumba H, Mwarumba S, M. Gichuki B, Njuguna S, Riako D,
645 Mutua S, Gitonga J, Sein Y, Bartilol B, Mwangi S, O. Omuoyo D, M. Morobe J, de
646 Laurent Z, Bejon P, Ochola-Oyier L, Tsofa B. 2021. Pooled testing conserves SARS-
647 CoV-2 laboratory resources and improves test turn-around time: experience on the
648 Kenyan Coast [version 2; peer review: 3 approved]. *Wellcome Open Research* 5.
- 649 21. Tyson JR, James P, Stoddart D, Sparks N, Wickenhagen A, Hall G, Choi JH, Lapointe H,
650 Kamelian K, Smith AD, Prystajeky N, Goodfellow I, Wilson SJ, Harrigan R, Snutch TP,
651 Loman NJ, Quick J. 2020. Improvements to the ARTIC multiplex PCR method for
652 SARS-CoV-2 genome sequencing using nanopore. *bioRxiv*
653 doi:10.1101/2020.09.04.283077.
- 654 22. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, du Plessis L, Pybus
655 OG. 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist
656 genomic epidemiology. *Nat Microbiol* 5:1403-1407.
- 657 23. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P,
658 Bedford T, Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution.
659 *Bioinformatics* 34:4121-4123.
- 660 24. Hale T, Angrist N, Boby T, Cameron-Blake E, Hallas L, Kira B, Majumdar S, Petherick
661 A, Phillips T, Tatlow H, Webster S. December 2020 Noam Angrist. Variation in
662 government responses to COVID-19.
663 doi:[https://www.bsg.ox.ac.uk/research/publications/variation-government-responses-](https://www.bsg.ox.ac.uk/research/publications/variation-government-responses-covid-19)
664 [covid-19](https://www.bsg.ox.ac.uk/research/publications/variation-government-responses-covid-19)doi:[https://www.bsg.ox.ac.uk/research/publications/variation-government-](https://www.bsg.ox.ac.uk/research/publications/variation-government-responses-covid-19)
665 [responses-covid-19](https://www.bsg.ox.ac.uk/research/publications/variation-government-responses-covid-19).

666
667 **SUPPLEMENTAL MATERIAL**

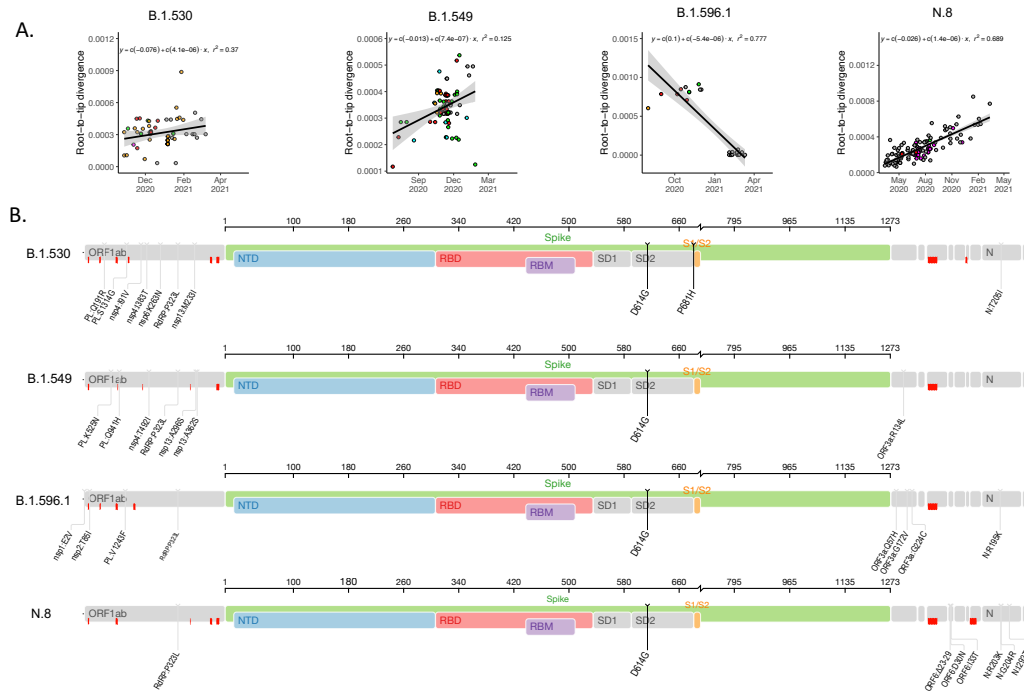


668
669 **S1 Figure.** The SARS-CoV-2 epidemic in Kenya. Panel A, Map of Kenya showing county
670 administrative boundaries and neighboring countries. The map is colored by the number of cases
671 documented in the counties as at of 26th February 2021. Panel, B. Reported daily new cases in
672 Kenya from March 2020 to February 2021 showing the first two national SARS-CoV-2 infection
673 waves. Panel C, daily cases reported across coastal Kenya during the study period. Panel D,
674 Kenya government intervention levels during the study period as measured by the Oxford
675 stringency index.



676
 677 **S2 Figure.** Position of non-synonymous mutations in the 684 genomes from coastal Kenya
 678 compare with the Wuhan reference genome (Accession number NC_045512). Only changes
 679 occurring in >1% of genomes are shown. The percentage frequency of the changes in the 684
 680 genomes is shown on the y-axis.

681



682
683 **S3 Figure. Genomic characteristics of the four lineages first identified in Kenya.** Panel A,
684 root-to-tip regression plots for the four lineages first identified in Kenya. Panel B, Genome maps
685 of the “Kenya” lineages where the spike region is shown in detail and in color and the rest of the
686 genome is shown in grey color.
687

688 **S1 Table. Lineages observed in Coastal Kenya, their global history and VOC/VOI status**

Lineage	Frequency (%)	Mombasa	Kilifi	Kwale	Taita Taveta	Tana River	Lamu	Earliest date (Global)	Total Global #	Description
A	21 (3.1)	3	-	13	5	-	-	2019-12-30	2930	Root of the pandemic lies within lineage A.
A.23	3 (0.4)	1	1	1	-	-	-	2020-08-14	97	Ugandan lineage
A.23.1	3 (0.4)	1	-	1	1	-	-	2020-10-21	957	International lineage
A.25	2 (0.3)	2	-	-	-	-	-	2020-06-22	34	Ugandan lineage
B	4 (0.6)	3	1	-	-	-	-	2019-12-24	5214	Second major haplotype
B.1	413 (60.4)	241	69	28	56	9	10	2020-01-24	75776	A large European lineage
B.1.1	61 (8.9)	39	5	4	13	-	-	2020-02-15	46205	European lineage
B.1.1.1	4 (0.6)	1	1	2	-	-	-	2020-03-04	2529	England
B.1.1.111	1 (0.2)	-	-	1	-	-	-	2020-04-04	59	African lineage (Zambia, Zimbabwe)
B.1.1.254	2 (0.3)	1	-	1	-	-	-	2020-03-14	76	Mostly South Africa
B.1.1.33	1 (0.2)	1	-	-	-	-	-	2020-03-01	1830	Brazilian lineage
B.1.1.485	1 (0.2)	-	-	-	1	-	-	2020-04-09	86	Ghanian Lineage
B.1.1.7	1 (0.2)	1	-	-	-	-	-	2020-02-07	694561	Alpha variant of concern, UK first identified
B.1.1.70	1 (0.2)	-	-	-	1	-	-	2020-03-06	2596	European Lineage
B.1.153	1 (0.2)	-	-	1	-	-	-	2020-01-24	267	English lineage
B.1.177	3 (0.4)	-	1	1	-	-	1	2020-02-14	68271	Major mostly European lineage
B.1.179	2 (0.3)	2	-	-	-	-	-	2020-03-09	256	Denmark
B.1.340	1 (0.2)	1	-	-	-	-	-	2020-04-06	136	USA lineage
B.1.351	10 (1.5)	-	3	3	4	-	-	2020-05-11	18425	Beta variant of concern
B.1.378	2 (0.3)	2	-	-	-	-	-	2020-03-14	190	USA
B.1.393	2 (0.3)	1	1	-	-	-	-	2020-07-04	32	Ugandan lineage
B.1.450	1 (0.2)	-	1	-	-	-	-	2020-03-14	126	USA (TX)
B.1.526	1 (0.2)	-	-	1	-	-	-	2020-11-23	18387	Iota variant of interest
B.1.530	44 (6.4)	8	7	-	28	-	1	2020-10-11	60	Kenyan
B.1.535	1 (0.2)	1	-	-	-	-	-	2020-03-25	36	Australian
B.1.549	54 (7.9)	18	24	7	5	-	-	2020-06-26	76	Kenya + England lineage
B.1.553	1 (0.2)	-	1	-	-	-	-	-	-	-
B.1.558	1 (0.2)	1	-	-	-	-	-	2020-04-06	167	USA/Mexico lineage
B.1.596.1	8 (1.2)	4	2	1	1	-	-	2020-08-03	29	Kenyan lineage
B.4	1 (0.2)	1	-	-	-	-	-	2020-01-18	374	Iranian lineage and global export
B.4.7	3 (0.4)	3	-	-	-	-	-	2020-03-14	68	Africa and UAE lineage
N.8	30 (4.4)	1	1	-	-	-	28	2020-06-23	19	Alias of B.1.1.33.8, Kenyan lineage

689