

1           **Emergence and transmission dynamics of the FY.4 Omicron variant in Kenya**

2  
3   **Authors:**

4 Sebastian Musundi<sup>1</sup>, Mike J. Mwangi<sup>1</sup>, Arnold W. Lambisia<sup>1</sup>, John Mwita Morobe<sup>1</sup>, Nickson  
5 Murunga<sup>1</sup>, Edidah Moraa <sup>1</sup>, Leonard Ndwiga, Robinson Cheruyoit<sup>1</sup>, Jennifer Musyoki<sup>1</sup>, Martin  
6 Mutunga<sup>1</sup>, Laura M Guzman-Rincon<sup>1,2</sup>, Charles Sande<sup>1</sup>, Joseph Mwangangi<sup>1</sup>, Philip Bejon<sup>1</sup>,  
7 Lynette Isabella Ochola-Oyier<sup>1</sup>, D James Nokes<sup>1,2</sup>, Charles N. Agoti<sup>1</sup>, Joyce Nyiro<sup>1</sup>, George  
8 Githinji<sup>1#</sup>

9  
10  
11   **Affiliations:**

12   <sup>1</sup> KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya.

13   <sup>2</sup> University of Warwick, Coventry, UK.

14  
15  
16   **#Corresponding author:** [ggithinji@kemri-wellcome.org](mailto:ggithinji@kemri-wellcome.org)

17  
18  
19   **Target journal:** Virus evolution

20  
21  
22   **Keywords:** FY.4, Omicron, Kenya, Phylogenetics

23 **Abstract (words=175)**

24 The recombinant FY.4 SARS-CoV-2 variant was first reported in Kenya in March 2023 and was  
25 the dominant circulating variant between April and July 2023. The variant was characterised by  
26 two important mutations: Y451H in the receptor binding domain of the spike protein and P42L in  
27 open reading frame 3a. Using phylogenetics and phylodynamic approaches, we investigated the  
28 emergence and spread of the FY.4 in Kenya and the rest of the world. Our findings suggest FY.4  
29 circulated early in Kenya before export to North America and Europe. Early circulation of FY.4  
30 in Kenya was predominantly observed in the coastal part of the country and the estimated time to  
31 the most recent common ancestor suggests FY.4 circulated as early as December 2022. The  
32 collected genomic and epidemiological data show that the FY.4 variant led to a large local outbreak  
33 in Kenya and resulted in localised outbreaks in Europe, North America and Asia-pacific. These  
34 findings underscore the importance of sustained genomic surveillance especially in under sampled  
35 regions in deepening our understanding of the evolution and spread of SARS-CoV-2 variants.

36

## 37 **Introduction**

38 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in Wuhan China in late  
39 2019 (Lu et al., 2020; Zhu et al., 2020). Over the past five years, SARS-CoV-2 has evolved into  
40 several lineages, some of which are more transmissible than others and reduce effectiveness to  
41 public health interventions such as therapeutics, diagnostics and vaccines (Centre for Disease  
42 Control and Prevention, 2021). The World Health Organization (WHO) designated these lineages  
43 as variant of concern (VOC) (<https://www.who.int/activities/tracking-SARS-CoV-2-variants>).  
44 The VOCs include Alpha (B.1.17) (Hill et al., 2022), Beta (B.1.35) (Tegally et al., 2021), Gamma  
45 (P.1) (Faria et al., 2021), Delta (B.1.617.2) (Cherian et al., 2021) and Omicron (Viana et al., 2022).  
46 The Omicron variant was first reported on 24th November 2021 in South Africa and Botswana  
47 and carried over 30 mutations in the spike protein compared to the Wuhan-Hu-1 strain (Tegally et  
48 al., 2022; Viana et al., 2022). The presence of these mutations increased transmissibility and  
49 reduced neutralization by anti-SARS-CoV-2 antibodies (Planas et al., 2022). While the origin of  
50 the Omicron remains debatable, possible hypotheses explaining its emergence include extreme  
51 under-sampling resulting in undetected lineages, zoonotic spillover from animals that transmit  
52 SARS-CoV-2 without an intermediate host like the white-tailed deer and farmed mink and  
53 persistence of the virus in chronically infected individuals (Markov et al., 2023).

54  
55 The Omicron variant was initially classified into three lineages BA.1, BA.2 and BA.3 before the  
56 emergence of BA.4 and BA.5 (Tegally et al., 2022). Unlike previous waves of VOCs which were  
57 driven by few circulating variants (Li et al., 2021), the emergence of Omicron led to convergent  
58 evolution and diversification (Andreis et al., 2023; Ito et al., 2023), resulting in multiple co-  
59 circulating lineages and an increased probability of recombination events. In mid-August 2022,  
60 the XBB recombinant was reported in India, Singapore and other parts of Asia (Wang et al., 2023).  
61 XBB emerged from two BA.2 omicron variants BM.1.1.1 (BA.2.75.3.1.1.1) and BJ.1  
62 (BA.2.10.1.1) with breakpoints located in the receptor-binding domain of spike protein (Tamura  
63 et al., 2023). XBB showed increased immune resistance, fusogenicity and binding affinity to  
64 angiotensin-converting enzyme 2 (ACE2), the receptor mediating viral infection compared to the  
65 parental strains (Tamura et al., 2023). Since then, multiple XBB variants such as XBB.1.5 (Parums,  
66 2023), and XBB.1.16 (Looi, 2023) emerged and spread across the globe.

67

68 The FY.4 (XBB.1.22.1.4) omicron variant became the dominant lineage in Kenya between March  
69 to July 2023 in Kenya following its initial detection on 10<sup>th</sup> March 2023 (Mwanga et al., 2023).  
70 Globally, the first FY.4 sample outside Kenya was reported first on 3<sup>rd</sup> April 2023 in Germany  
71 (Khare et al., 2021). In this work, we aimed to understand the origin and transmission of the FY.4  
72 Omicron sub-variant using genomic data. FY.4 circulated as early as December 2022 and was  
73 observed in Kenya first before being exported to other regions of the world. These findings suggest  
74 FY.4 potentially originated from Kenya and underscores the importance of genomic surveillance  
75 in tracking SARS-CoV-2 variants.

76

77

## 78 **Methods**

### 79 **Data sources**

80 SARS-CoV-2 FY.4 samples (n=73) used for analysis in this study were generated at Kenya  
81 Medical Research Institute Wellcome Trust Research Programme (KWTRP) as previously  
82 described (Mwanga et al., 2023). An additional (n=139) FY.4 genomes sequenced within Kenya  
83 and across the globe were retrieved from Global Initiative for Sharing all Influenza Data (GISAID).  
84 Consensus genomes for FY.4 lineages and XBB were retrieved from  
85 (<https://github.com/corneliusroemer/pango-sequences>) and aligned using MAFFT. Amino acid  
86 mutations relative to the Wuhan-HU-1 and XBB and were plotted using SNIPit (O'Toole et al.,  
87 2024).

88

### 89 **Time-scaled analysis of the FY.4 variant**

90 We compiled a dataset of circulating SARS-CoV-2 sequences in Kenya between October 2022  
91 and October 2023 (n=1,078). Out of these 73 genomes were sequenced at KWTRP and an  
92 additional (n=139) sequenced locally in Kenya, making a total of 212 Kenyan FY.4 sequences.  
93 We analysed this dataset against a background of FY.4 global sequences retrieved from GISAID  
94 (n=755). The global dataset from where the sequences were obtained are summarized in  
95 Supplementary Table 1. Sequences with less than 70% genome coverage (n=1), incomplete  
96 collection dates (n=11), and whose lineage were misclassified based on NextClade version 3.8.2  
97 (n=5) were excluded from further analysis. The remaining sequences were aligned against the  
98 SARS-CoV-2 (XBB) reference (a Wuhan-Hu-1/2019 reference (Accession number MN908947.3))

99 containing XBB single nucleotide polymorphisms) using MAFFT version 6.240 (Kato &  
100 Standley, 2013). The alignment was visualized using AliView version 1.28 (Larsson, 2014) and  
101 the 5' and 3' ends containing the untranslated regions were manually trimmed. A maximum  
102 likelihood (ML) tree was generated using IQ-TREE version 2.3.3 with 1000 bootstrap values using  
103 the General Time Reversible (GTR) model (Nguyen et al., 2015). The maximum likelihood tree  
104 was visualized using the R package “ggtree” and low-quality sequences (n=16) dropped from the  
105 tree using the R package “ape.” A time-resolved maximum likelihood tree was then generated  
106 using Treetime (Sagulenko et al., 2018).

107

### 108 **Phylogeography analysis**

109 The presence of a molecular signal was evaluated using TempEst version 1.5.3 (Rambaut et al.,  
110 2016) prior to Bayesian analysis using BEAST (Drummond & Rambaut, 2007). To identify the  
111 clades circulating within Kenya, a discrete phylogeography approach utilizing two locations;  
112 “Kenya” and “Others” was run in BEAST 1.10.4. This approach utilized the time-resolved  
113 phylogeny as the starting tree to reduce computational time needed to generate transmission  
114 information (Bollen et al., 2021). Bayesian inference was run using BEAST v1.10.4 for  $6 \times 10^5$   
115 Monte Carlo Markov Chain (MCMC) and sampling every 1000 step. Log files were checked for  
116 convergence and mixing using Tracer v1.7.1 (Rambaut et al., 2018). After discarding 10% as burn-  
117 in, a maximum clade credibility tree (MCC) generated using TreeAnnotator version 1.10.4  
118 (Suchard et al., 2018). The overall number of transitions between the two locations was counted  
119 in BEAST package Babel utilizing sampled trees.

120

121 To explore the local transmission of the FY.4 variant within Kenya, the relaxed random walk  
122 (RRW) diffusion model which allows for dispersal velocity in the tree to vary but remain the same  
123 in the branches was inferred using the longitude and latitudes of sample locations as continuous  
124 traits (Lemey et al., 2010; Pybus et al., 2012). Since the RRW model does not allow samples to  
125 have the same geographical coordinates, we selected the centroid location of each administrative  
126 point and added random jitter of 0.05 to each tip making each location distinct. A MCMC chain  
127 was run in duplicate for 500 million iterations sampled after every 50,000 steps with its mixing  
128 properties checked in Tracer v1.7.1 to ensure the effective sample size (ESS) is  $>200$ . The R  
129 package “seraphim” used to extract spatiotemporal information embedded in the posterior,

130 maximum clade credibility trees and visualize the dispersal of FY.4 within Kenyan counties  
131 (Dellicour et al., 2016).

132  
133 Next, we carried out a discrete phylogeographic analysis to infer the global origin and movement  
134 of FY.4 across different regions across the globe. We aggregated the country of sampling into four  
135 regions including Kenya (n=206), North America (n=494), Europe (n=140) and Asia-Pacific (Asia  
136 and Oceania) (n=91). One sample from South America and two samples from Uganda were  
137 excluded to minimize biased posterior distributions due to unequal sample size. In addition, we  
138 analyzed samples collected from March 2023 considering the first case was reported on 10<sup>th</sup> March  
139 2023. By focusing on this timeline, we aimed to capture the spread of FY.4 from the time it was  
140 initially identified providing a clear evolutionary trajectory of the geographic spread of FY.4 while  
141 in circulation. During this study, three additional FY.4 sequences were recently deposited in  
142 GISAID from Nakuru, Kenya with collection dates ranging between 11<sup>th</sup> -14<sup>th</sup> November 2022.  
143 Although these sequences predate the first reported case suggesting earlier circulation, we  
144 excluded these samples due to insufficient information from very few samples collected in  
145 November 2022 and absence of samples collected between November 2022 and March 2023.  
146 Since a large number of sequences originated from the USA, down sampling was carried out based  
147 on time, pangolin lineage and region resulting in a total of 511 sequences distributed as follows:  
148 Kenya (n=94), North America (n=233), Europe (n=107) and Asia-Pacific (n=77). Ancestral  
149 locations reconstruction of the FY.4 variant and asymmetric viral exchanges between regions was  
150 estimated utilizing a non-continuous time Markov chain model. The Bayesian stochastic search  
151 variable selection was used to infer non-zero migration rates and identify the best-supported  
152 transitions rates using a bayes factor test. In addition, Markov jump counting was used to estimate  
153 the number of transitions across the different regions. The time-scaled phylogeny was created  
154 using the HKY substitution model with a gamma distributed rate for invariant sites, an uncorrelated  
155 relaxed molecular clock with a log normal prior and Bayesian skygrid with 52 change points over  
156 a period of one year. The analysis was run in duplicate on BEAST v1.10.4 using 200 million  
157 MCMC steps with sampling after every 20000 steps. Mixing and convergence properties were  
158 assessed using Tracer v.1.71 ensuring the ESS was >200. After discarding the 10% burn-in, the  
159 MCC tree was constructed using TreeAnnotator version 1.10 and the number of  
160 transitions/Markov jumps estimated using TreeMarkovJumpHistoryAnalyzer (Lemey et al., 2021).

161

## 162 **Results**

163

### 164 **FY.4 circulated in Kenya before global expansion and underwent rapid evolution over a** 165 **short period of time.**

166 FY.4 variant became the dominant circulating lineage in Kenyan between March and July 2023  
167 before eventual replacement with other Omicron subvariants (Figure 1A). The period coincides  
168 with the high positivity rate observed from samples collected from health facilities within the Kilifi  
169 Health Demographic Surveillance System (KHDSS) (Supplementary Figure 1). On the global  
170 scale, FY.4 was first detected in Africa before its eventual detection in other continents across the  
171 globe (Figure 1B). In Africa, FY.4 circulation peaked between April and May 2023 while in North  
172 America and Europe the peak was observed in July 2023 (Figure 1B). In Africa, sublineage FY.4.1  
173 was the dominant circulating lineage and co-circulated with its expanded lineage FY.4.1.2. North  
174 America initially saw FY.4.1 and FY.4.2 in April 2023 followed by FY.4.1.2 before the co-  
175 circulation of FY.4.2 and FY.4.1.1 in the later stages. FY.4.1 and FY.4.1.2 were the main  
176 circulating lineages in Europe from April 2023 and while the same lineages were initially detected  
177 in Asia from May 2023, FY.4.2 became the main circulating lineage thereafter from June to  
178 September 2023 (Figure 1B).

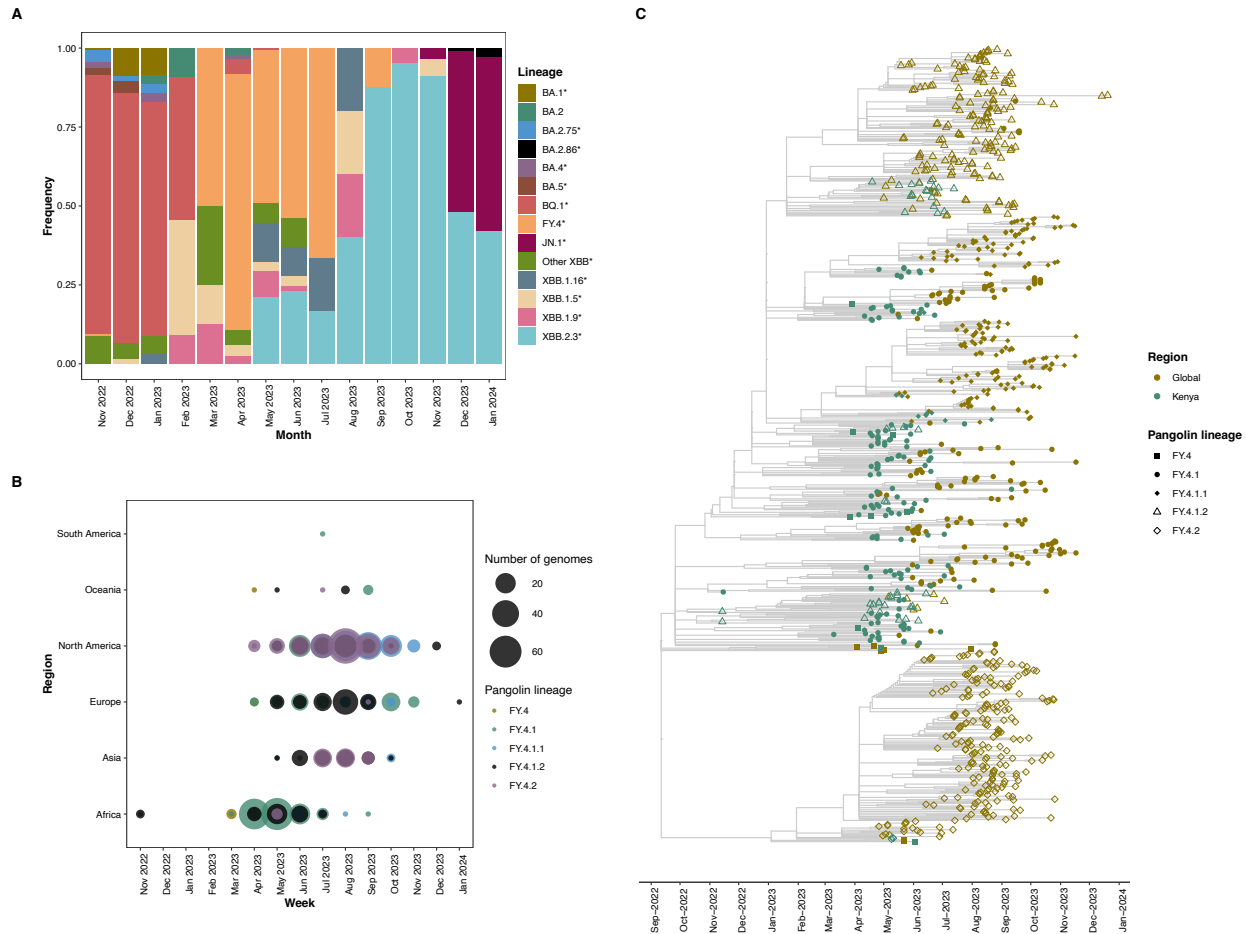
179

180 Phylogenetic analysis of the circulating strains in Kenya (n=1078) between October 2022 and  
181 October 2023, showed that FY.4 was closely related to other XBB variants and more specifically  
182 XBB.1.16 and XBB.1.9 (Supplementary Figure 2). As expected, FY.4 was also present in the same  
183 clade as BA.2 and BA.2.75 which has previously been proposed as the main parental lineages for  
184 the emergence of the recombinant XBB. In addition to the two lineage-defining mutations-Y451H  
185 in the spike protein and P42L in ORF3a- expanded sub-lineages exhibited notable mutations.  
186 FY.4.1 contained S494P in the receptor binding domain of spike glycoprotein while FY.4.1.1  
187 included S704L in subdomain 2. The FY.4.1.2 sub-lineage was characterized by S2926F in ORF1a  
188 and FY.4.2 displayed mutations Y2171F in ORF1a and V2287I in ORF1b.

189

190 A time-resolved tree ML tree containing Kenyan (n=207) FY.4 and global sequences (n=738)  
191 revealed a short period between the expansion of FY.4.1 to FY.4.1.1 and FY.4.1.2 suggesting a  
192 rapid diversification process (Figure 1C). Following its expansion, FY.4.1 persisted for 36 weeks

193 in circulation together with its derived lineages. In contrast, FY.4.2, which also emerged from  
 194 FY.4, did not expand to additional lineages, likely reflecting the different selective pressures acting  
 195 on individual FY.4 lineages. Analysis of the temporal spread revealed FY.4 circulated early in  
 196 Kenya compared to other regions of the globe (Figure 1C). Additionally, Kenyan FY.4 sequences  
 197 displayed significant genetic diversity by mapping to multiple lineages and formed clusters  
 198 providing evidence for local transmission ahead of introductions to other regions (Figure 1C).  
 199



200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211

**Figure 1. Circulation of the FY.4 omicron subvariant.** A) Proportion of SARS-CoV-2 lineages circulating in Kenya from September 2022 to January 2024. FY.4 circulated in Kenya predominantly from March to September 2023. B) Global distribution of circulating FY.4 variants from November 2022 to January 2024. The individual colours represent the FY.4 sub-lineages while the size of the circle denotes the number of genomes deposited in GISAID per month. C) Time-resolved phylogenetic tree showing the temporal clustering of FY.4 between Kenya and global sequences. The colours represent the location the sample was collected from, and the shape represent the FY.4 pangolin sub-lineages.

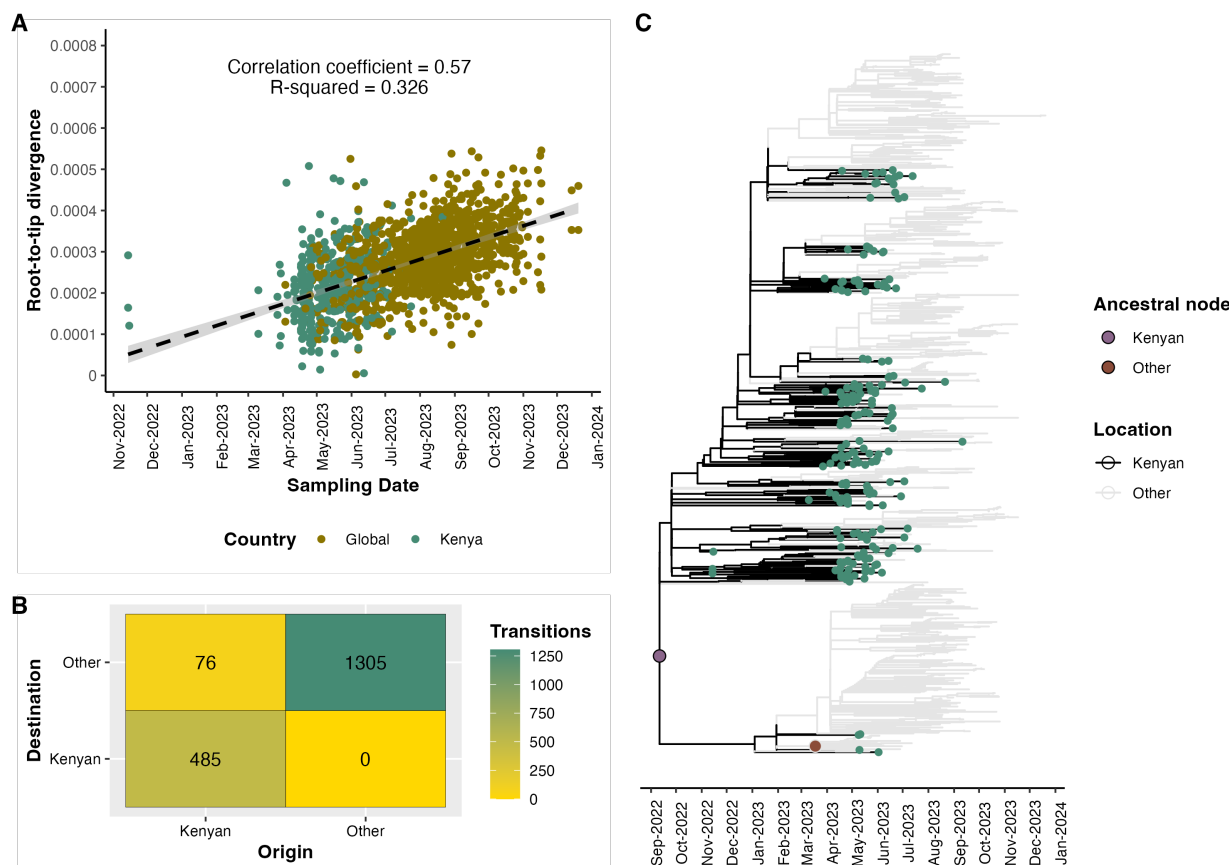


212 **The FY.4 variant originated in Kenya.**

213 The presence of a temporal signal was confirmed by the linear relationship between genetic  
214 divergence and sampling dates using TempEst version 1.5.3 ( $R^2=0.326$ ) (Figure 2A).  
215 Subsequently, we carried out a discrete phylogeographic analysis to identify the potential origin  
216 and the number of import and export events to Kenya using the FY.4 variant using genomic data.  
217 Based on this, we observed no evidence of independent introductions of FY.4 in Kenya (Figure  
218 2B) based on the 95% highest posterior density (HPD) ([0-1]) However, we identified at least 75  
219 transmission events from Kenya to the rest of the globe (95% HPD interval = [73-79]). Local  
220 transmission dynamics in Kenya, comprised of over 485 transition events between counties (95%  
221 HPD interval= [480-489]) (Figure 2B). Evidence of spread is supported by export and imports  
222 events between the counties. Global spread was observed with support for multiple export events  
223 (1,304 transition events (95% HPD interval = [1298-1312]) across the globe showing FY.4 arrived  
224 and transmitted across multiple regions.

225  
226 Ancestral reconstruction analysis revealed the presence of two most common recent ancestors. The  
227 first node infers Kenyan ancestry and included all Kenyan sequences alongside global sequences  
228 forming the largest clade with 915 sequences (Figure 2C). The genetic diversity of the Kenyan  
229 sequences in this node supports strong local transmission within Kenya while the presence of both  
230 Kenyan and global sequences in the same node suggests possible worldwide dissemination of the  
231 FY.4 variant. In contrast, the second ancestral node was associated with fewer sequences (n=19)  
232 and likely represented a more localized transmission route. Notably, the second ancestral node did  
233 not include Kenyan samples which might suggest limited or no transmission to or from Kenya or  
234 a pathway that completely bypassed Kenya (Figure 2C).

235



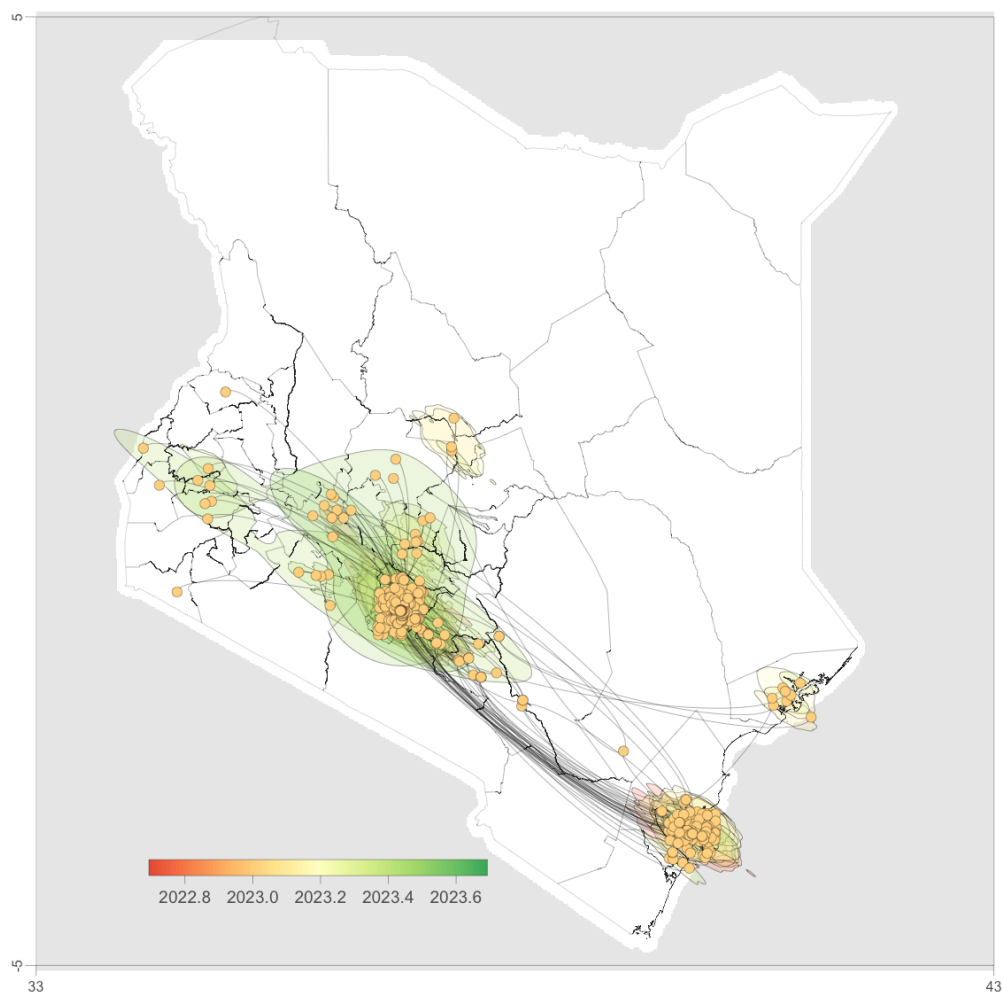
236

237 **Figure 2. Preliminary phylogeography discrete trait analysis.** A) Root-to-tip regression from  
238 TempEst analysis showing the presence of a temporal signal ( $R^2 = 0.326$ , correlation coefficient =  
239 0.57). The FY.4 sequences are colored according to the locations (Kenya or Global). The  
240 regression line represents the estimate mean evolutionary rate with error buffers in grey showing  
241 the 90% confidence intervals. B). The number of transitions was derived from the MCC tree by  
242 counting changes between “Kenyan” and “Other” locations. Transitions were counted when the  
243 location of the internal node changed from “Kenya” to “Others” or vice versa or when maintained  
244 in the same position. C) Preliminary discrete trait analysis identified two ancestral clades  
245 associated with spread of FY.4 using a fixed time-resolved tree. The green tips represent individual  
246 Kenya samples, black the most probable location for Kenyan sequences and background grey the  
247 most probable location for global sequences. Two ancestral nodes are filled by the plum and  
248 salmon colours respectively. The “Other” ancestral node occurs primarily on the FY.4.2 sub-  
249 lineage which was not largely observed in Kenya but was predominant in North America.

250

251 We extracted the ancestral clade that inferred Kenyan ancestry and examined virus dispersal within  
252 Kenya. The subsequent continuous phylogeography of patterns of dispersal within Kenya  
253 supported the evidence that FY.4 was the dominant circulating lineage in coastal region from  
254 March 2023. By May-June 2023, most cases were observed centrally in the capital city (Figure 3).

255 Most peripheral samples also directly connected to either the coastal region or centrally in the  
256 capital city which were the predominant sampling sites. This pattern likely suggests a limited  
257 geographical range of sampling in Kenya, potentially leading to an underestimate of the circulation  
258 of FY.4 in other parts of Kenya.

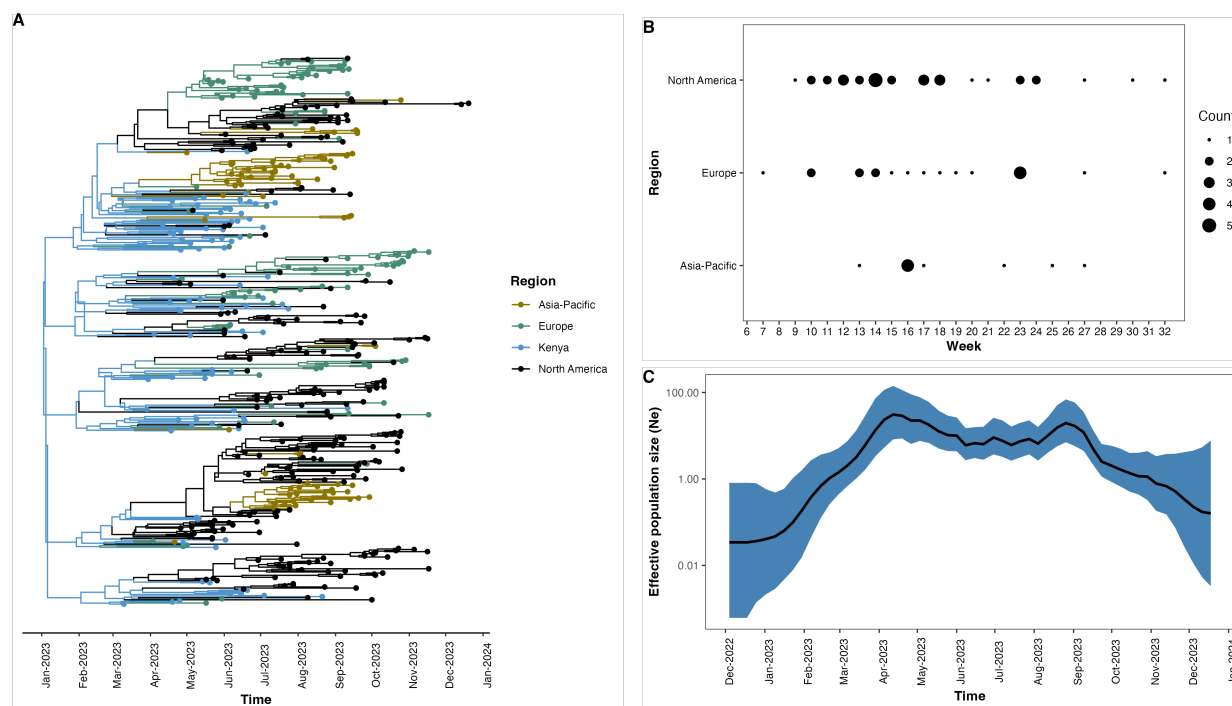


259  
260 **Figure 3. Dispersal of FY.4 across Kenya over time based on 1000 subsampled trees from a**  
261 **continuous phylogeographic posterior distribution.** The nodes of the MCC tree are color-coded  
262 based on the time of occurrence. The 80% HPD regions are displayed in successive layers with  
263 the colors reflecting corresponding time periods for virus spread.

264  
265 Phylogeographic reconstruction implied that FY.4 may have emerged in Kenya in early January  
266 (mean tMRCA 3<sup>rd</sup> January, 95% HPD 2<sup>nd</sup> December 2022- 1<sup>st</sup> February 2023) (Figure 4A).  
267 Thereafter, FY.4 was first exported to Europe then to North America in week 7 and 9 respectively

268 and later exported to Asia-Pacific around week 13 (Figure 4B). Out of the 60 export events from  
269 Kenya, most were to North America (n=32), followed by Europe (n=19) and lastly Asia-Pacific  
270 (n=9) (Figure 4B). There were no inferred introductions from the globe into to Kenya during this  
271 period. An exponential increase in the effective population size was observed between January  
272 and April 2023 followed by a minor decline and another slight exponential increase from August  
273 to September before a decline to January 2024 (Figure 4C).

274



275

276

277 **Figure 4. Bayesian phylogeographic reconstruction of FY.4.** A) Time-resolved maximum  
278 credibility clade tree with branches coloured by inferred geographic location. B) A summary of  
279 the number of Markov Jumps from Kenya to other regions stratified by epidemiological weeks. C)  
280 A Bayesian Skyline plot describing the inferred change in the effective population size of FY.4  
281 infections over time.

282

283

## 284 Discussion

285 The FY.4 variant increased in circulation in Kenya between March and July 2023 accounting for  
286 increased cases and hospitalization (Mwanga et al., 2023). This variant was also observed in other  
287 regions across the globe (Figure 1C). Given the early isolation of FY.4 and increased transmission  
288 intensity in Kenya, we applied genomic and epidemiological data to investigate the emergence and

289 transmission dynamics of FY.4 variant from sequenced samples in Kenya in addition to those  
290 collected from the globe.

291  
292 The FY.4 lineage of SARS-COV-2 is characterized by two notable mutations namely the Y451H  
293 in the receptor binding domain (RBD) of the spike protein, whose functional implication is unclear  
294 and the P42L in the ORF3a, with potential for contribution to the loss of recognition of T-cell  
295 epitope (de Silva et al., 2021). Additionally, the S494P present in the RBD of FY.4.1 enhances  
296 binding affinity to ACE2 receptor increasing transmissibility (Chakraborty, 2021) and potentially  
297 contributing to the increased number of observed cases. Neutralization assays against multiple  
298 circulating Omicron variants including FY.4 in Kenya between March 2023-March 2024 have  
299 provided evidence for decline in naturally acquired and vaccine-mediated antibody responses  
300 implying that the Kenyan population was still susceptible to infections caused by emerging  
301 Omicron subvariants. (Lugano et al., unpub. data doi.org/10.1101/2024.06.26.24309525). Here,  
302 we used an established health facilities surveillance platform (Nyiro et al., 2018), and combined  
303 this with further local and global genomic data to determine the origin and describe the  
304 transmission dynamics of this variant.

305  
306 The phylogenetic and phylogeographic reconstruction suggests that Kenya was the potential origin  
307 of the FY.4 variant (Figure 2B). Phylogenetic and phylogenomic analysis provides evidence for  
308 multiple exportation events from Kenya, primarily to North America and Europe (Figure 3B)  
309 which occurred between March and July 2023. Previously we have observed multiple  
310 introductions of ancestral strains (Githinji et al., 2021) and variants of concern (Agoti et al., 2022;  
311 Githinji et al., unpub. data doi.org/10.1101/2022.10.26.22281446). The Bayesian skyline plot  
312 showed that the effective population size corresponded with increase in the number of cases and  
313 potentially indicate missed cases during the outbreak period. A peak in the effective population  
314 size was observe between August and September 2023 suggesting a surge in cases outside of  
315 Kenya and this coincides with a rise in the FY.4 genome sequences deposited from North America  
316 and Europe (Figure 3C). This suggests ongoing transmission events between Kenya and the rest  
317 of the world following the initial emergence of FY.4.

318

319 A limitation with this study is that the true number of FY.4 derived COVID-19 cases in Kenya,  
320 African and South America is likely underestimated given the limited number of tests at the period,  
321 under-surveillance and the limited genomic surveillance in the country. The biased sampling is  
322 underscored by clustered outbreaks in areas with higher sampling rates.

323

### 324 **Conclusion**

325 Genomic surveillance is essential in identifying emerging variants of SARS-CoV-2. In this study,  
326 the use of phylogenetics and phylodynamic approaches provided insights to the possible origin  
327 and dispersal patterns of the FY.4 variant The study reinforces the need for increased sequencing  
328 capacity especially in under-sampled geographic regions especially in studying the evolution of  
329 SARS-CoV-2.

330

331

### 332 **Data Availability**

333 The raw data files and consensus genome sequences obtained in this study were submitted to both  
334 GISAID and GenBank databases and the accession numbers prepared for deposition in Harvard  
335 DataVerse (<https://doi.org/10.7910/DVN/VPWUXN>). For more detailed information beyond the  
336 metadata used in the paper, there is a process of managed access requiring submission of a request  
337 form for consideration by our Data Governance Committee.

338

### 339 **Acknowledgments**

340 We thank the members of our field study team involved in collecting the samples, the members of  
341 the KWTRP COVID-19 testing team, PEO research group members that undertook real-time RT-  
342 PCR processing. This work was support by multiple funding sources that included the New Variant  
343 Assessment Platform (NVAP), Wellcome grants (220985/Z/20/Z and 226002/A/22/Z). The  
344 Rockerfeller Foundation subaward (OXFFDG01) and the Department of Health and Social Care  
345 grant (project references 17/63/82 and 16/136/33). The views expressed in this publication are  
346 those of the author (s) and not necessarily those of the Department of Health and Social Care,  
347 Foreign Commonwealth and Development Office, Wellcome Trust or the UK government.

348

349 This manuscript was published with permission from the Director General

350

351 **Role of funders**

352 The funders of the study had no role in study design, data collection, data analysis, data  
353 interpretation or writing of the report.

354

355 **Author's contributions**

356 The project was conceived and designed by GG; Project supervision and funding was obtained by  
357 GG, CNA, LI, JN, DJN; Laboratory processing of the samples was conducted by MM, AWL, JM,  
358 EM, LK, RC, JM, MM, LG; Data management and analysis was carried out by NM,SM,LN.  
359 SM,GG wrote the initial draft and GG, CNA, PB, reviewed the manuscript and produced the final  
360 draft. All authors contributed to and reviewed the final draft.

361 **References**

- 362 Agoti, C. N., Ochola-Oyier, L. I., Dellicour, S., Mohammed, K. S., Lambisia, A. W., de  
363 Laurent, Z. R., Morobe, J. M., Mburu, M. W., Omuoyo, D. O., Ongera, E. M., Ndwiga,  
364 L., Maitha, E., Kitole, B., Suleiman, T., Mwakingu, M., Nyambu, J. K., Otieno, J.,  
365 Salim, B., Musyoki, J., ... Githinji, G. (2022). Transmission networks of SARS-CoV-2  
366 in Coastal Kenya during the first two waves: A retrospective genomic study. *ELife*, *11*.  
367 <https://doi.org/10.7554/eLife.71703>
- 368 Andreis, T. F., Cantarelli, V. V., da Silva, M. B., Helfer, M. S., Brust, F. R., Zavascki, A.  
369 P., Westercamp, M., Tashayev, V., Donadel, M., Magleby, R., Petersen, E., & Mahon,  
370 G. (2023). Substantial Diversity in Cocirculating Omicron Lineages in Hospital  
371 Setting, Porto Alegre, Brazil. *Emerging Infectious Diseases*, *29*(12).  
372 <https://doi.org/10.3201/eid2912.230880>
- 373 Bollen, N., Artesi, M., Durkin, K., Hong, S. L., Potter, B., Boujemla, B., Vanmechelen, B.,  
374 Martí-Carreras, J., Wawina-Bokalanga, T., Meex, C., Bontems, S., Hayette, M. P.,  
375 André, E., Maes, P., Bours, V., Baele, G., & Dellicour, S. (2021). Exploiting genomic  
376 surveillance to map the spatio-temporal dispersal of SARS-CoV-2 spike mutations in  
377 Belgium across 2020. *Scientific Reports*, *11*(1). [https://doi.org/10.1038/s41598-021-](https://doi.org/10.1038/s41598-021-97667-9)  
378 [97667-9](https://doi.org/10.1038/s41598-021-97667-9)
- 379 Center for Disease Control and Prevention. (2021). *SARS-CoV-2 Variant Classifications*  
380 *and Definitions*. <https://stacks.cdc.gov/view/cdc/105817>
- 381 Chakraborty, S. (2021). Evolutionary and structural analysis elucidates mutations on SARS-  
382 CoV2 spike protein with altered human ACE2 binding affinity. *Biochemical and*  
383 *Biophysical Research Communications*, *534*.  
384 <https://doi.org/10.1016/j.bbrc.2020.11.075>
- 385 Cherian, S., Potdar, V., Jadhav, S., Yadav, P., Gupta, N., Das, M., Rakshit, P., Singh, S.,  
386 Abraham, P., & Panda, S. (2021). Sars-cov-2 spike mutations, I452R, T478K, E484Q and  
387 P681R, in the second wave of COVID-19 in Maharashtra, India. *Microorganisms*, *9*(7).  
388 <https://doi.org/10.3390/microorganisms9071542>
- 389 de Silva, T. I., Liu, G., Lindsey, B. B., Dong, D., Moore, S. C., Hsu, N. S., Shah, D.,  
390 Wellington, D., Mentzer, A. J., Angyal, A., Brown, R., Parker, M. D., Ying, Z., Yao,  
391 X., Turtle, L., Dunachie, S., Aanensen, D. M., Abudahab, K., Adams, H., ... Dong, T.  
392 (2021). The impact of viral mutations on recognition by SARS-CoV-2 specific T cells.  
393 *IScience*, *24*(11). <https://doi.org/10.1016/j.isci.2021.103353>
- 394 Dellicour, S., Rose, R., Faria, N. R., Lemey, P., & Pybus, O. G. (2016). SERAPHIM:  
395 Studying environmental rasters and phylogenetically informed movements.  
396 *Bioinformatics*, *32*(20). <https://doi.org/10.1093/bioinformatics/btw384>



- 397 Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by  
398 sampling trees. *BMC Evolutionary Biology*, 7(1). [https://doi.org/10.1186/1471-2148-7-](https://doi.org/10.1186/1471-2148-7-214)  
399 214
- 400 Faria, N. R., Mellan, T. A., Whittaker, C., Claro, I. M., Candido, D. D. S., Mishra, S.,  
401 Crispim, M. A. E., Sales, F. C. S., Hawryluk, I., McCrone, J. T., Hulswit, R. J. G.,  
402 Franco, L. A. M., Ramundo, M. S., De Jesus, J. G., Andrade, P. S., Coletti, T. M.,  
403 Ferreira, G. M., Silva, C. A. M., Manuli, E. R., ... Sabino, E. C. (2021). Genomics and  
404 epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*, 372(6544).  
405 <https://doi.org/10.1126/science.abh2644>
- 406 Githinji, G., de Laurent, Z. R., Mohammed, K. S., Omuoyo, D. O., Macharia, P. M.,  
407 Morobe, J. M., Otieno, E., Kinyanjui, S. M., Agweyu, A., Maitha, E., Kitole, B.,  
408 Suleiman, T., Mwakinangu, M., Nyambu, J., Otieno, J., Salim, B., Kasera, K., Kiiru, J.,  
409 Aman, R., ... Agoti, C. N. (2021). Tracking the introduction and spread of SARS-  
410 CoV-2 in coastal Kenya. *Nature Communications*, 12(1).  
411 <https://doi.org/10.1038/s41467-021-25137-x>
- 412 Githinji, G., Lambisia, A., Omah, I., medRxiv, A. O.-, & 2022, undefined. (n.d.). The  
413 genomic epidemiology of SARS-CoV-2 variants of concern in Kenya. *Medrxiv.OrgG*  
414 *Githinji, AW Lambisia, I Omah, A O'Toole, KS Mohamed, ZR de Laurent, TO*  
415 *MakorimedRxiv, 2022•medrxiv.Org*. Retrieved October 15, 2024, from  
416 <https://www.medrxiv.org/content/10.1101/2022.10.26.22281446.abstract>
- 417 Hill, V., Plessis, L. Du, Peacock, T. P., Aggarwal, D., Colquhoun, R., Carabelli, A. M.,  
418 Ellaby, N., Gallagher, E., Groves, N., Jackson, B., McCrone, J. T., O'Toole, Á., Price,  
419 A., Sanderson, T., Scher, E., Southgate, J., Volz, E., Barclay, W. S., Barrett, J. C., ...  
420 Rambaut, A. (2022). The origins and molecular evolution of SARS-CoV-2 lineage  
421 B.1.1.7 in the UK. *Virus Evolution*, 8(2). <https://doi.org/10.1093/ve/veac080>
- 422 Ito, J., Suzuki, R., Uriu, K., Itakura, Y., Zahradnik, J., Kimura, K. T., Deguchi, S., Wang,  
423 L., Lytras, S., Tamura, T., Kida, I., Nasser, H., Shofa, M., Begum, M. M., Tsuda, M.,  
424 Oda, Y., Suzuki, T., Sasaki, J., Sasaki-Tabata, K., ... Sato, K. (2023). Convergent  
425 evolution of SARS-CoV-2 Omicron subvariants leading to the emergence of BQ.1.1  
426 variant. *Nature Communications*, 14(1). <https://doi.org/10.1038/s41467-023-38188-z>
- 427 Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software  
428 version 7: Improvements in performance and usability. *Molecular Biology and*  
429 *Evolution*, 30(4). <https://doi.org/10.1093/molbev/mst010>
- 430 Khare, S., Gurry, C., Freitas, L., Schultz, M. B., Bach, G., Diallo, A., Akite, N., Ho, J., Lee,  
431 R. T. C., Yeo, W., Team, G. C. C., & Maurer-Stroh, S. (2021). GISAID's Role in  
432 Pandemic Response. *China CDC Weekly*, 3(49).  
433 <https://doi.org/10.46234/ccdcw2021.255>

- 434 Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large  
435 datasets. *Bioinformatics*, 30(22). <https://doi.org/10.1093/bioinformatics/btu531>
- 436 Lemey, P., Rambaut, A., Welch, J. J., & Suchard, M. A. (2010). Phylogeography takes a  
437 relaxed random walk in continuous space and time. *Molecular Biology and Evolution*,  
438 27(8). <https://doi.org/10.1093/molbev/msq067>
- 439 Lemey, P., Ruktanonchai, N., Hong, S. L., Colizza, V., Poletto, C., Van den Broeck, F.,  
440 Gill, M. S., Ji, X., Levasseur, A., Oude Munnink, B. B., Koopmans, M., Sadilek, A.,  
441 Lai, S., Tatem, A. J., Baele, G., Suchard, M. A., & Dellicour, S. (2021). Untangling  
442 introductions and persistence in COVID-19 resurgence in Europe. *Nature*, 595(7869).  
443 <https://doi.org/10.1038/s41586-021-03754-2>
- 444 Li, J., Lai, S., Gao, G. F., & Shi, W. (2021). The emergence, genomic diversity and global  
445 spread of SARS-CoV-2. In *Nature* (Vol. 600, Issue 7889).  
446 <https://doi.org/10.1038/s41586-021-04188-6>
- 447 Looi, M. K. (2023). What do we know about the Arcturus XBB.1.16 subvariant? *BMJ*  
448 (*Clinical Research Ed.*), 381. <https://doi.org/10.1136/bmj.p1074>
- 449 Lu, J., du Plessis, L., Liu, Z., Hill, V., Kang, M., Lin, H., Sun, J., François, S., Kraemer, M.  
450 U. G., Faria, N. R., McCrone, J. T., Peng, J., Xiong, Q., Yuan, R., Zeng, L., Zhou, P.,  
451 Liang, C., Yi, L., Liu, J., ... Ke, C. (2020). Genomic Epidemiology of SARS-CoV-2 in  
452 Guangdong Province, China. *Cell*, 181(5). <https://doi.org/10.1016/j.cell.2020.04.023>
- 453 Lugano, D., Kutima, B., Kimani, M., Sigilai, A., Gitonga, J., Karani, A., Akech, D., Karia,  
454 B., Ziraba, A. K., Maina, A., Omuoyo, D., Mugo, D., Lucinde, R., Newman, J., Nduati,  
455 E., Githinji, G., Agoti, C. N., Bejon, P., Anthony, J. G., ... Nyagwange, J. (2023).  
456 Evaluation of population immunity against SARS-CoV-2 variants, EG. 5.1, FY. 4, BA.  
457 2.86, JN. 1, and JN. 1.4, using samples from two health demographic surveillance.  
458 *Medrxiv.OrgDI* <https://doi.org/10.1101/2024.06.26.24309525>
- 459 Markov, P. V., Ghafari, M., Beer, M., Lythgoe, K., Simmonds, P., Stilianakis, N. I., &  
460 Katzourakis, A. (2023). The evolution of SARS-CoV-2. In *Nature Reviews*  
461 *Microbiology* (Vol. 21, Issue 6). <https://doi.org/10.1038/s41579-023-00878-2>
- 462 Mwanga, M. J., Lambisia, A. W., Morobe, J. M., Murunga, N., Moraa, E., Ndwiga, L.,  
463 Cheruiyot, R., Musyoki, J., Mutunga, M., Guzman-Rincon, L. M., Sande, C.,  
464 Mwangangi, J., Bejon, P., Ochola-Oyier, L. I., Nokes, D. J., Agoti, C. N., Nyiro, J., &  
465 Githinji, G. (2023). New SARS-CoV-2 Omicron Variant with Spike Protein Mutation  
466 Y451H, Kilifi, Kenya, March-May 2023. *Emerging Infectious Diseases*, 29(11).  
467 <https://doi.org/10.3201/eid2911.230894>

- 468 Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast  
469 and effective stochastic algorithm for estimating maximum-likelihood phylogenies.  
470 *Molecular Biology and Evolution*, *32*(1). <https://doi.org/10.1093/molbev/msu300>
- 471 Nyiro, J. U., Munywoki, P., Kamau, E., Agoti, C., Gichuki, A., Etyang, T., Otieno, G., &  
472 James Nokes, D. (2018). Surveillance of respiratory viruses in the outpatient setting in  
473 rural coastal Kenya: Baseline epidemiological observations. *Wellcome Open Research*,  
474 *3*. <https://doi.org/10.12688/wellcomeopenres.14662.1>
- 475 O'Toole, Á., Aziz, A., & Maloney, D. (2024). Publication-ready single nucleotide  
476 polymorphism visualization with snipit. *Bioinformatics*, *40*(8).  
477 <https://doi.org/10.1093/BIOINFORMATICS/BTAE510>
- 478 Parums, D. V. (2023). Editorial: The XBB.1.5 ('Kraken') Subvariant of Omicron SARS-  
479 CoV-2 and its Rapid Global Spread. In *Medical Science Monitor* (Vol. 29).  
480 <https://doi.org/10.12659/MSM.939580>
- 481 Planas, D., Saunders, N., Maes, P., Guivel-Benhassine, F., Planchais, C., Buchrieser, J.,  
482 Bolland, W. H., Porrot, F., Staropoli, I., Lemoine, F., Péré, H., Veyer, D., Puech, J.,  
483 Rodary, J., Baele, G., Dellicour, S., Raymenants, J., Gorissen, S., Geenen, C., ...  
484 Schwartz, O. (2022). Considerable escape of SARS-CoV-2 Omicron to antibody  
485 neutralization. *Nature*, *602*(7898). <https://doi.org/10.1038/s41586-021-04389-z>
- 486 Pybus, O. G., Suchard, M. A., Lemey, P., Bernardin, F. J., Rambaut, A., Crawford, F. W.,  
487 Gray, R. R., Arinaminpathy, N., Stramer, S. L., Busch, M. P., & Delwart, E. L. (2012).  
488 Unifying the spatial epidemiology and molecular evolution of emerging epidemics.  
489 *Proceedings of the National Academy of Sciences of the United States of America*,  
490 *109*(37). <https://doi.org/10.1073/pnas.1206598109>
- 491 Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior  
492 summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5).  
493 <https://doi.org/10.1093/sysbio/syy032>
- 494 Rambaut, A., Lam, T. T., Carvalho, L. M., & Pybus, O. G. (2016). Exploring the temporal  
495 structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus*  
496 *Evolution*, *2*(1). <https://doi.org/10.1093/ve/vew007>
- 497 Sagulenko, P., Puller, V., & Neher, R. A. (2018). TreeTime: Maximum-likelihood  
498 phylodynamic analysis. *Virus Evolution*, *4*(1). <https://doi.org/10.1093/ve/vex042>
- 499 Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A.  
500 (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10.  
501 *Virus Evolution*, *4*(1). <https://doi.org/10.1093/ve/vey016>

- 502 Tamura, T., Ito, J., Uriu, K., Zahradnik, J., Kida, I., Anraku, Y., Nasser, H., Shofa, M., Oda,  
503 Y., Lytras, S., Nao, N., Itakura, Y., Deguchi, S., Suzuki, R., Wang, L., Begum, M. M.,  
504 Kita, S., Yajima, H., Sasaki, J., ... Sato, K. (2023). Virological characteristics of the  
505 SARS-CoV-2 XBB variant derived from recombination of two Omicron subvariants.  
506 *Nature Communications*, 14(1). <https://doi.org/10.1038/s41467-023-38435-3>
- 507 Tegally, H., Moir, M., Everatt, J., Giovanetti, M., Scheepers, C., Wilkinson, E.,  
508 Subramoney, K., Makatini, Z., Moyo, S., Amoako, D. G., Baxter, C., Althaus, C. L.,  
509 Anyaneji, U. J., Kekana, D., Viana, R., Giandhari, J., Lessells, R. J., Maponga, T.,  
510 Maruapula, D., ... de Oliveira, T. (2022). Emergence of SARS-CoV-2 Omicron  
511 lineages BA.4 and BA.5 in South Africa. *Nature Medicine*, 28(9).  
512 <https://doi.org/10.1038/s41591-022-01911-2>
- 513 Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J.,  
514 Doolabh, D., Pillay, S., San, E. J., Msomi, N., Mlisana, K., von Gottberg, A., Walaza,  
515 S., Allam, M., Ismail, A., Mohale, T., Glass, A. J., Engelbrecht, S., Van Zyl, G., ... de  
516 Oliveira, T. (2021). Detection of a SARS-CoV-2 variant of concern in South Africa.  
517 *Nature*, 592(7854). <https://doi.org/10.1038/s41586-021-03402-9>
- 518 Viana, R., Moyo, S., Amoako, D. G., Tegally, H., Scheepers, C., Althaus, C. L., Anyaneji,  
519 U. J., Bester, P. A., Boni, M. F., Chand, M., Choga, W. T., Colquhoun, R., Davids, M.,  
520 Deforche, K., Doolabh, D., du Plessis, L., Engelbrecht, S., Everatt, J., Giandhari, J., ...  
521 de Oliveira, T. (2022). Rapid epidemic expansion of the SARS-CoV-2 Omicron  
522 variant in southern Africa. *Nature*, 603(7902). [https://doi.org/10.1038/s41586-022-](https://doi.org/10.1038/s41586-022-04411-y)  
523 [04411-y](https://doi.org/10.1038/s41586-022-04411-y)
- 524 Wang, Q., Iketani, S., Li, Z., Liu, L., Guo, Y., Huang, Y., Bowen, A. D., Liu, M., Wang,  
525 M., Yu, J., Valdez, R., Luring, A. S., Sheng, Z., Wang, H. H., Gordon, A., Liu, L., &  
526 Ho, D. D. (2023). Alarming antibody evasion properties of rising SARS-CoV-2 BQ  
527 and XBB subvariants. *Cell*, 186(2). <https://doi.org/10.1016/j.cell.2022.12.018>
- 528 Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu,  
529 R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., & Tan, W.  
530 (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New*  
531 *England Journal of Medicine*, 382(8). <https://doi.org/10.1056/nejmoa2001017>
- 532

