

Genomic epidemiology uncovers the timing and origin of the emergence of mpox in humans

Edyth Parker^{*,1,2+}, Ifeanyi F. Omah^{*,3,4}, Patrick Varilly^{*,5}, Andrew Magee⁶, Akeemat Opeyemi Ayinla¹, Ayotunde E. Sijuwola¹, Muhammad I. Ahmed¹, Oludayo O. Ope-ewe¹, Olusola Akinola Ogunsanya¹, Alhaji Olono¹, Philomena Eromon¹, Christopher H Tomkins-Tinch⁵, James Richard Otieno⁷, Olusola Akanbi⁸, Abiodun Egwuenu⁸, Odianoson Ehiakhamen⁸, Chimaobi Chukwu⁸, Kabiru Suleiman⁸, Afolabi Akinpelu⁸, Adama Ahmad⁸, Khadijah Isa Imam⁸, Richard Ojedele⁸, Victor Oripenaye⁸, Kenneth Ikeata⁸, Sophiyah Adelakun⁸, Babatunde Olajumoke⁸, Delia Doreen Djuicy⁹, Loique Landry Messanga Essengue⁹, Moïse Henri Moumbeket Yifomnjou⁹, Mark Zeller², Karthik Gangavarapu², Áine O'Toole³, Daniel J Park⁵, Gerald Mboowa¹⁰, Sofonias Kifle Tessema¹⁰, Yenew Kebede Tebeje¹⁰, Onikepe Folarin^{1,11}, Anise Happi¹, Philippe Lemey¹², Marc A Suchard^{6,13,14}, Kristian G. Andersen^{2,15}, Pardis Sabeti^{5,16}, Andrew Rambaut³, Richard Njourn⁹, Chikwe Ihekweazu⁸, Idriss Jide⁸, Ifedayo Adetifa⁸, Christian T Happi^{1,11,16+}

*These authors contributed equally to this work

+ Correspondences: Christian T Happi (happic@run.edu.ng), Edyth Parker (edythp@run.edu.ng)

1. African Center of Excellence for Genomics of Infectious Diseases, Redeemer's University, Ede, Osun State, Nigeria
2. Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA, USA
3. Institute of Ecology and Evolution, University of Edinburgh, The King's Buildings, Edinburgh EH9 3FL, UK.
4. Department of Parasitology and Entomology, Nnamdi Azikiwe University, Awka, Nigeria.
5. The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
6. Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA
7. Theiagen Genomics, Highlands Ranch, CO, USA
8. Nigeria Centre for Disease Control and Prevention, Abuja, Nigeria
9. Virology Service, Centre Pasteur du Cameroun, 451 Rue 2005, Yaounde 2, P.O. Box 1274
10. Africa Centres for Disease Control and Prevention, P.O. Box 3243, Addis Ababa, Ethiopia
11. Department of Biological Sciences, Faculty of Natural Sciences, Redeemer's University, Ede, Osun State, Nigeria
12. Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium
13. Department of Biomathematics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA.
14. Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA 90095, USA
15. Scripps Research Translational Institute, La Jolla, CA 92037, USA.
16. Department of Immunology and Infectious Diseases, Harvard T H Chan School of Public Health, Boston, MA 02115

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

1 **Five years before the 2022–2023 global mpox outbreak Nigeria reported its first**
2 **cases in nearly 40 years, with the ongoing epidemic since driven by sustained**
3 **human-to-human transmission. However, limited genomic data has left**
4 **questions about the timing and origin of the mpox virus' (MPXV) emergence.**
5 **Here we generated 112 MPXV genomes from Nigeria from 2021-2023. We identify**
6 **the closest zoonotic outgroup to the human epidemic in southern Nigeria, and**
7 **estimate that the lineage transmitting from human-to-human emerged around**
8 **July 2014, circulating cryptically until detected in September 2017. The epidemic**
9 **originated in Southern Nigeria, particularly Rivers State, which also acted as a**
10 **persistent and dominant source of viral dissemination to other states. We show**
11 **that APOBEC3 activity increased MPXV's evolutionary rate twenty-fold during**
12 **human-to-human transmission. We also show how Delphy, a tool for near-real-**
13 **time Bayesian phylogenetics, can aid rapid outbreak analytics. Our study sheds**
14 **light on MPXV's establishment in West Africa before the 2022–2023 global**
15 **outbreak and highlights the need for improved pathogen surveillance and**
16 **response.**
17

18 Mpox is a viral zoonosis caused by the mpox virus (MPXV) of the genus *Orthopoxvirus*.
19 MPXV diversity is partitioned into two major clades: Clade 1, which is endemic in an
20 as-yet unknown non-human animal reservoir in Central Africa, and Clade II with
21 subclades Clade IIa and Clade IIb, similarly endemic in Western Africa.^{1,2} From the
22 first recognized human case in 1970 in the Democratic Republic of Congo (DRC) until
23 2017 there have only been occasional outbreaks and sporadic cases with limited
24 human-to-human transmission in endemic regions.^{3–5}
25

26 In September 2017, Nigeria reported its first mpox outbreak in nearly 40 years, with a
27 larger second wave of cases towards 2022.^{5,6} The outbreak had distinct
28 epidemiological characteristics not typically seen with MPXV zoonotic infections.
29 There was a marked demographic shift to more urban, adult individuals of 30 to 40
30 years of age, compared to more typical cases reported in children in rural
31 communities.^{6,7}
32

33 In May 2022, approximately five years after the 2017 outbreak in Nigeria, a Clade IIb
34 lineage termed B.1 rapidly disseminated around the world to cause the global mpox
35 outbreak.⁸ The apparent human-to-human transmission of the B.1 lineage raised the
36 possibility of a new transmission route for MPXV. B.1 showed significant divergence
37 from the closest Clade IIb genome sampled in Nigeria in 2018, with an evolutionary
38 rate significantly elevated above the expected rate for Orthopoxviruses.⁹ The global
39 outbreak was characterised by enrichment of mutations in a specific dinucleotide
40 context (TC→TT or the reverse complement GA→AA) associated with the cytosine
41 deaminase activity of the APOBEC3 (apolipoprotein B mRNA editing enzyme, catalytic
42 polypeptide 3) host antiviral mechanism.^{10,11} This mutational signature has not been
43 observed in sequences from zoonotic infections, strongly suggesting that APOBEC3
44 genomic editing was a characteristic feature of sustained transmission in the human
45 population.¹¹

46
47 In light of this new evolutionary dynamic, several studies confirmed that the ongoing
48 mpox epidemic in Nigeria was driven by sustained human-to-human transmission, not
49 independent zoonotic infections or self-limiting transmission chains.^{11,12} It is estimated
50 that MPXV emerged in the human population in Nigeria in 2016, circulating and
51 diversifying cryptically into multiple distinct lineages.^{11,12} The diversity circulating
52 during the human outbreak in Nigeria, from which the B.1 global lineage descended,
53 is referred to as hMPXV-1, as designated by Happi *et al.*¹³

54
55 However, the limited availability of high-quality genomes and their restricted
56 geographic representation leaves several uncertainties around the transmission
57 dynamics of hMPXV-1 during the ongoing human epidemic in Nigeria. The genomic

58 data supports a single zoonotic origin for hMPXV-1.¹¹ However, no close zoonotic
59 outgroup has been identified to date, and estimates of the timing of hMPXV-1's
60 emergence in the human population is based on limited genomic data. The
61 evolutionary dynamics of hMPXV-1 after the initial spillover remains largely unknown,
62 with a limited genomic picture already revealing multiple distinct lineages co-
63 circulating.¹² Importantly, it is not known whether all cases in Nigeria are part of the
64 human epidemic, or whether there are ongoing spillover events resulting in
65 independent onward transmission. It is also not clear what the drivers of this sustained
66 human transmission are, and whether there are clear spatiotemporal spread patterns
67 seeding and reseeding local epidemics across Nigerian states.

68

69 To address the open questions around the origin, spread, and evolution of the recent
70 mpox outbreak, we formed a Pan-African consortium to collate the largest MPXV
71 dataset from Africa to date. In this study, we generated 112 genomes sampled from
72 Nigeria from 2021 to 2023, including the period before the emergence of the B.1 global
73 outbreak. We re-estimate the date of MPXV's transition to sustained human-to-human
74 transmission using a custom phylodynamic model. With the widest geographic
75 representation sampled in Nigeria yet, we reconstruct the spatiotemporal spread of
76 hMPXV-1 across Nigeria and quantify the drivers of hMPXV-1 transmission.

77 **Results:**

78 **Distinct hMPXV-1 lineages co-circulate during the** 79 **ongoing human epidemic in Nigeria**

80
81 To investigate whether all of our mpox samples were part of the human epidemic, as
82 well as the extent of hMPXV-1's cryptic diversification across the epidemic, we
83 generated 112 MPXV genomes sampled across Nigeria from July 2021 to May 2023.
84 The median coverage of the sequences generated ranged from 30 to 7900-x. Our
85 sequences were predominantly sampled from the South South, South East, and South
86 West regions of Nigeria (Figure 1 A, B). The South region was the epicentre of the
87 epidemic from 2017 onwards, with southern states reporting the earliest cases (Figure
88 1 A, B). From 2022 onwards, both the northern and southern regions experienced a
89 substantial resurgence of cases after a period of low incidence. Our dataset largely
90 encompasses the resurgence from 2022 onwards (Figure 1A). All of our sequences
91 belonged to Clade IIb in the global phylogeny (Figure 1C).

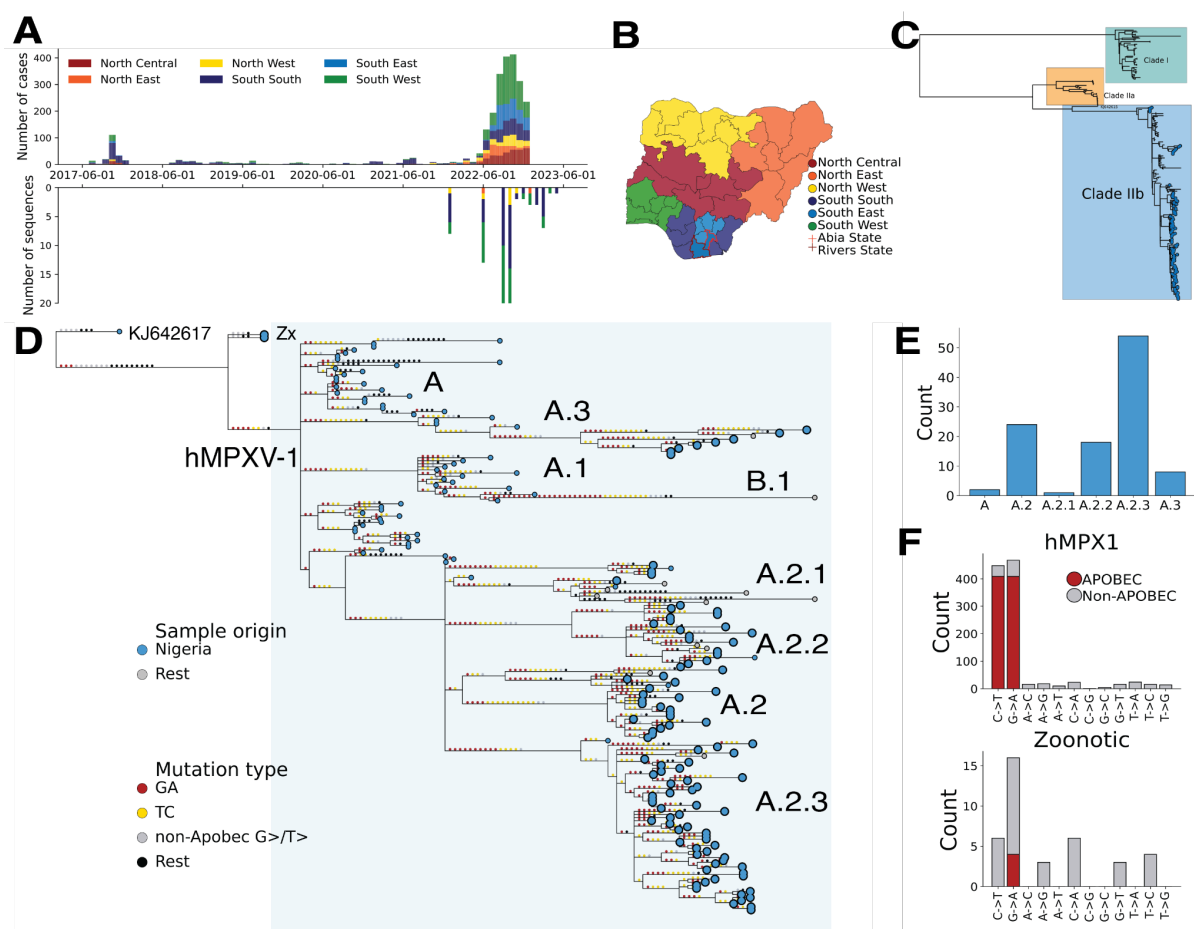
92
93 To investigate hMPXV-1's cryptic diversification with our expanded dataset, we
94 reconstructed the Clade IIb phylogeny with our 105 non-B.1 novel genomes and all
95 available Clade IIb sequences.¹² Within hMPXV-1, distinct lineages are designated
96 according to a system similar to the SARS-CoV-2 Pango nomenclature to track the
97 lineages circulating within the outbreak.^{13,14} Under this nomenclature hMPXV-1 is
98 referred to as Lineage A, with its descendants designated as e.g. A.1 and A.2 and
99 second subdivision descendants designated as e.g. A.1.1. We found that 105 of our
100 sequences were interspersed throughout six divergent, co-circulating sub-lineages of

101 hMPXV-1 (Lineage A) (Figure 1D, E). Lineage A represents the sustained human-to-
102 human transmission reported in Nigeria from 2017 onwards, as well as travel-
103 associated cases and the B.1 global outbreak.^{11–13} The majority of our sequences
104 belonged to sub-lineage A.2.3 (Figure 1E). We did not find any geographic structure
105 in hMPXV-1, indicating that the sub-lineages were not confined to specific geographic
106 subpopulations. We also identified three B.1 sequences, representing re-introductions
107 from the global epidemic into Nigeria. Notably, none of our sequences were from
108 lineage A.1.1, from which the B.1 global outbreak lineage descended (Figure 1D).

109
110 We investigated whether all of our sequences formed part of the ongoing human
111 epidemic or whether any of them represented zoonotic transmission events by
112 quantifying the APOBEC3 mutational bias characteristic of human-to-human
113 transmission across our phylogeny.^{11,15,16} We performed ancestral state
114 reconstruction across our Clade IIb phylogeny to map single nucleotide polymorphism
115 (SNPs) to their relevant branches. We annotated APOBEC3 characteristic mutations
116 i.e. C→T or G→A in the correct dimer context along branches (Figure 1D) and
117 calculated their relative proportion across internal branches (Figure 1F).

118
119 We observed a significant proportion of mutations consistent with APOBEC3 activity
120 across the internal and terminal branches of hMPXV-1, including our sequences
121 (Figure 1D, F). Approximately 74% of reconstructed SNPs in hMPXV-1 were indicative
122 of APOBEC3 editing, consistent with previous work.¹¹ In contrast, only 9% of
123 reconstructed SNPs across the zoonotic parts of the tree (excluding the subtree
124 annotated in blue) were APOBEC3-type mutations (Figure 1F). This APOBEC3
125 enrichment indicates that 105 of our 109 non-B.1 sequences form part of the sustained

126 human-to-human transmission underlying the Nigerian epidemic. Notably, two
 127 sequences did not show APOBEC3 enrichment, indicating these represent zoonotic
 128 transmission events (outgroup “Zx” in Figure 1D). Taken together, these findings
 129 suggest that hMPXV-1 has cryptically circulated and diversified, largely driven by
 130 APOBEC3 activity, in the human population without geographic restrictions in Nigeria
 131 for a prolonged period after initial spillover.^{11,12}



132
 133 **Figure 1: A)** Epidemiological incidence of mpxv cases in Nigeria coloured by geopolitical region (top
 134 panel), relative to our genomic dataset’s temporal and geographic distribution (bottom panel). **B)**
 135 Geopolitical regions of Nigeria, with Abia and Rivers State highlighted with red borders **C)** Global MPXV
 136 phylogeny of Clade I, Clade IIa and Clade IIb. Our Clade IIb sequences are annotated with tip points
 137 **D)** Clade IIb phylogeny with reconstructed SNPs mapped onto branches. APOBEC3 mutations along
 138 the branches are annotated in yellow and red, with the remainder in grey and black. The hMPXV-1
 139 Clade (Lineage A) is annotated and highlighted in the light blue box, with the lineage annotation in text.
 140 Our new zoonotic outgroup sequences are annotated as “Zx”. Our sequences are highlighted as
 141 enlarged tips relative to the background tips. **E)** Lineage distribution of our Clade IIb sequences. **F)** The
 142 number of overall reconstructed SNPs that APOBEC3 substitutions accounts for in the hMPXV-1

143 subtree (highlighted and annotated in blue in Figure D) and the Zoonotic outgroup (KJ642617 and Zx
144 annotated in Figure D).

145 **Closest zoonotic outgroup to hMPXV-1 circulated in** 146 **southern Nigeria**

147 We explored whether the two zoonotic sequences we identified that lack APOBEC3
148 signal might represent a closer zoonotic outgroup to hMPXV-1 than the current
149 outgroup (KJ642617), sampled in 1971 (Figure 1D). We found that they form a sister
150 lineage to hMPXV-1 ("Zx" in Figure 1D), breaking up the long stem branch from
151 KJ642617 to hMPXV-1. As our new sequences share a common ancestor with
152 hMPXV-1, they represent the closest zoonotic outgroup to hMPXV-1 (Figure 1D).¹⁷
153 The Zx sequences reduced the branch from a zoonotic outgroup to hMPXV-1 from 27
154 to 8 SNPs. Our new Zx sequences were sampled in Abia State in southern Nigeria,
155 where KJ642617 was also sampled in 1971. Abia State provides suitable ecological
156 conditions to host reservoir populations where zoonotic precursors may have
157 circulated.¹⁷

158
159 A previous study found that the genomic data supported a single zoonotic origin for
160 hMPXV-1, as all of the sequences from the human epidemic shared a substantial
161 number of APOBEC3 mutations that occurred along the stem branch from KJ642617
162 and hMPXV-1's zoonotic common ancestor.¹¹ With our new zoonotic outgroup, we
163 investigated the number of shared APOBEC3 mutations along the stem branch. We
164 found that all of the diversified hMPXV-1 sub-lineages shared six APOBEC3 mutations
165 (Figure 1D). This supports findings that it is unlikely that the emergence of hMPXV-1
166 in the human population resulted from repeated independent zoonotic events.¹¹

167 **The mpox epidemic has grown exponentially, but**
168 **slowly, since emergence in 2014**

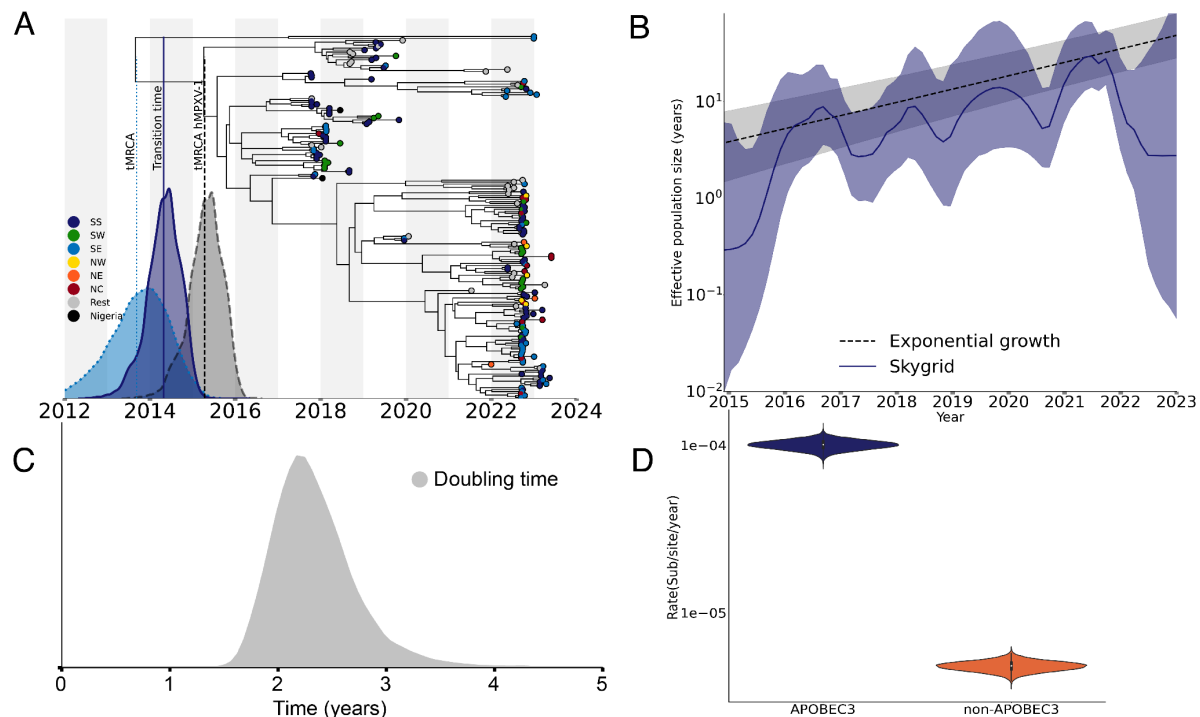
169 Our new zoonotic outgroup breaks up the long stem branch from hMPXV-1's most
170 recent common ancestor (MRCA) in non-human animals. This more recent divergence
171 of our zoonotic outgroup and hMPXV-1 places a tighter bound on when hMPXV-1
172 could have emerged in the human population. To estimate the timing of MPXV's initial
173 emergence in the human population with our newly identified zoonotic outgroup and
174 our expanded dataset, we adopted the two-epoch model of O'Toole *et al.* implemented
175 in the BEAST software package.¹¹ The model explicitly accounts for APOBEC3-
176 mediated evolution by allowing for the transition from a background evolutionary rate
177 driven by the polymerase error rate to an APOBEC3 driven rate across the tree in a
178 partitioned alignment (see Methods).

179
180 We estimated that the transition to sustained human transmission, representing the
181 time of emergence in the human population, occurred in mid July 2014 (95% HPD 6
182 October 2013 - 23 February 2015) (Figure 2A). Our estimate is approximately 14
183 months earlier than previous reports, though the credible intervals overlap (vs 14
184 September 2015, 95% HPD 21 August 2014 to 31 July 2016).¹¹ We estimated that
185 time to the most recent common ancestor or tMRCA of hMPXV-1 was mid July 2015
186 (median time to the most recent common ancestor or tMRCA, 95% HPD 8 November
187 2014 - 22 February 2016), representing the time at which hMPXV-1 started to diversify
188 in humans (Figure 2A). Our estimate is approximately eight months earlier than
189 previous, though credible intervals overlap (vs 23 February 2016, 95% HPD 28 June
190 2015, 28 September 2016).¹¹ Collectively, our results indicate that hMPXV-1 circulated

191 cryptically in the human population in Nigeria for approximately three years before
192 detection during the late 2017 wave of cases, and more than seven years before
193 disseminating globally during the B.1 outbreak in 2022. This prolonged period of
194 cryptic human-to-human transmission likely explains the diversification of hMPXV-1
195 into the distinct sub-lineages we observed (Figure 1D).

196

197 After the initial surge of cases in 2017 there was a period of low reported incidence
198 between 2018-2022, including the early years of the COVID-19 pandemic (Figure 1A).
199 This low incidence, especially from 2020-2021, may be attributable to underreporting
200 associated with the impact of COVID-19 on health systems, as well as COVID-19
201 related public health interventions. To investigate whether the mpox epidemic had
202 grown continuously since emergence or whether there was a real decline in cases
203 during the period of low reported incidence, we allowed for a two-phase coalescent
204 model in our epoch-model. Under the model, the tree from the MRCA(hMPXV-1)
205 onwards was modelled with an exponential growth model, and the earlier zoonotic
206 phases were modelled under a constant-population size model. We also estimated the
207 population dynamics of hMPXV-1 alone under a non-parametric Skygrid model. We
208 found evidence of exponential growth following hMPXV-1's emergence under both
209 models (Figure 2B), as previously reported.¹¹ This indicates that the epidemic was
210 growing exponentially even during periods of low reported incidence (Figure 1A) with
211 an estimated doubling time of approximately two and a half years (Figure 2C). This
212 exponential, but slow growth rate suggests that the epidemic has not spread to the
213 general population, but is concentrated in a more restricted subpopulation.¹⁸



214 **Figure 2 A).** Bayesian maximum clade credibility (MCC) tree of Clade IIb indicating the time of
 215 emergence of hMPXV-1 into the human population. The distributions indicate the 95% HPD for 1) the
 216 tMRCA of the closest zoonotic outgroup of the hMPXV-1 2) the time of transition to sustained human-
 217 human transmission and 3) the tMRCA of hMPXV-1. SS: South South; SW: South West; SE: South
 218 East; NW: North West; NE: North East; NC: North Central. **B)** The effective population size of the
 219 epidemic in Nigeria under a Skygrid and exponential coalescent model **C)** The posterior distribution of
 220 the estimated doubling time of the epidemic since emergence **D)** Estimates of the APOBEC3 and Non-
 221 APOBEC3 clock rates.
 222

223 **APOBEC3 activity increased the evolutionary rate 20-** 224 **fold during sustained human transmission**

225 APOBEC3-mediated genomic editing has significantly elevated the evolutionary rate
 226 of hMPXV-1 above the expected rate for double-stranded DNA viruses.¹¹ To
 227 investigate how much higher the APOBEC3-mediated evolutionary rate is relative to
 228 the polymerase-driven rate in the context of human-to-human transmission, we
 229 employed the partitioned two-epoch model. We estimated an APOBEC3 clock rate of
 230 1×10^{-4} substitutions per site per year (subs/site/year) (95% HPD 8.8×10^{-5} - 1.14×10^{-4})
 231 ⁴), and a background evolutionary rate of 4.8×10^{-6} subs/site/year (95% HPD 4.2×10^{-6}
 232 - 5.5×10^{-6}) (Figure 2D). This indicates that APOBEC3 activity increased the rate ~ 20

233 times higher relative to the background evolutionary rate during the human epidemic
234 in Nigeria.

235 **Southern Nigeria was an early and persistent source** 236 **of hMPXV-1 dissemination from emergence onwards**

237 Nigeria's southern states were the early epicentre for the mpox epidemic, and reported
238 the highest number of cases throughout (Figure 1A). Despite the clustering of the
239 earliest cases in South South states, the geographic origin of hMPXV-1 remains
240 unknown. The northern states only reported a significant number of cases after the
241 resurgence in 2022 (Figure 1A). However, it is not known whether there was
242 widespread, under-ascertained and unsampled transmission outside of the southern
243 states before the resurgence in 2022. It is also not clear which states contributed to
244 interstate viral dissemination and how these patterns may have shifted across the
245 different epidemic phases. To better understand the spatiotemporal spread of hMPXV-
246 1 within Nigeria, we used discrete and continuous phylogeographic reconstructions on
247 a state and regional level. We found that both the reconstructions support that hMPXV-
248 1 likely originated in Rivers State in the South South (Figure 3A, Posterior = 0.91,
249 Extended Data Figure 1). The MRCA of hMPXV-1 circulated in Rivers State in July
250 2015 (tMRCA, 95% HPD 8 November 2014 - 22 February 2016) (Figure 3A). Notably,
251 this is consistent with our sampling of the closest zoonotic outgroup to hMPXV-1 in the
252 neighbouring southern Abia state (Figure 1D).

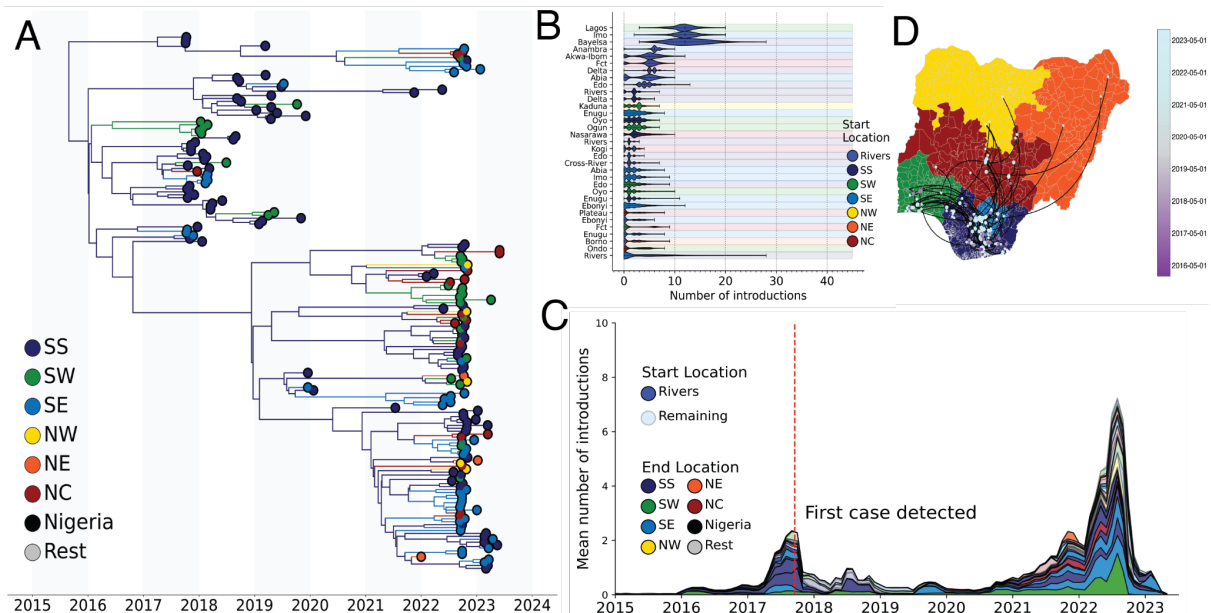
253
254 We found that Rivers State was the primary source of interstate viral exports across
255 the epidemic, with an estimated 72 introductions originating in Rivers (95% HPD: 62-
256 80) (Figure 3 A, B, Extended Data Figure 1). The highest number of viral exports from

257 Rivers spread to other South South states as well as the South West, followed by the
258 South East (Figure 3B, Extended Data Figure 2). Overall, neighbouring Imo and Lagos
259 in the South West had the highest number of introductions from Rivers, followed by
260 neighbouring Bayelsa. The remainder of the South South states as well as the South
261 East and South West states were all equivalently the second highest source of viral
262 exports overall.

263
264 We found that all introductions in the early epidemic originated in Rivers State (Figure
265 3C). Viral spread from Rivers into neighbouring South South states such as Bayelsa
266 and Imo, as well as Lagos in the South West, occurred as early as 2016 (Figure 3C,
267 Extended Data Figure 3). This is consistent with the epidemiological data that confirms
268 southern Nigeria as the epicentre of the early epidemic, with the first case reported in
269 Bayelsa on 11 September 2017 (Figure 1A). All save one sampled introductions into
270 Northern states occurred in the later phase of the epidemic, after the resurgence
271 towards 2022 (Figure 3C, Figure 1A). This spatiotemporal pattern was consistent
272 across our discrete and continuous phylogeographic reconstructions on both a
273 regional and state level (Figure 3D, Extended Data Figures 1-3).

274
275 To investigate how widespread early hMPXV-1 transmission was across Nigeria by
276 the time the outbreak was declared on 22 September 2017, we performed a
277 continuous phylogeographic reconstruction. We found that the virus had spread more
278 than 500 km beyond Rivers State into Bayelsa, Imo, Delta, Edo, FCT, and Lagos
279 before the first case was detected on 11 September 2017 (Figure 3D). Collectively,
280 this suggests that the human epidemic originated in Rivers State, with early spread of

281 the virus to neighbouring South South and South East states and Lagos before
 282 outbreak declaration, and with delayed dispersal to the north.



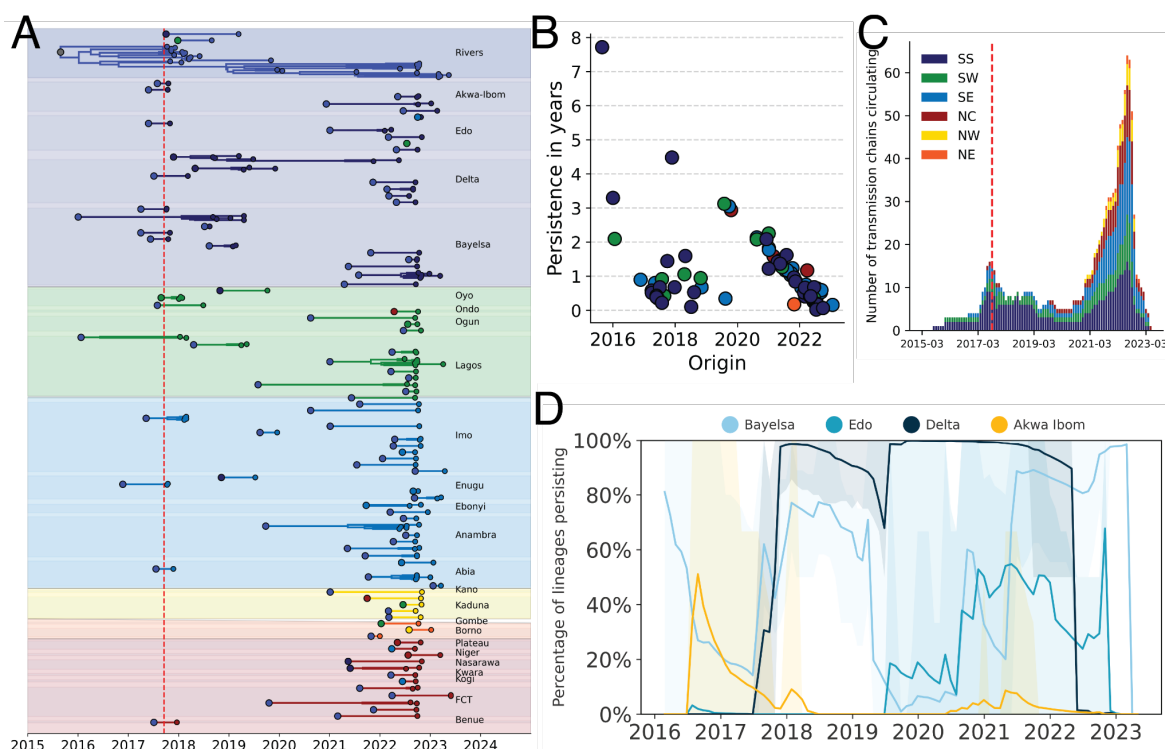
283
 284 **Figure 3: A)** Phylogeographic reconstruction of the spatiotemporal spread of Clade IIb in Nigeria. The
 285 branches of the Maximum Clade Credibility tree (MCC) are coloured by source region, as per legend.
 286 SS: South South; SW: South West; SE: South East; NW: North West; NE: North East; NC: North
 287 Central. **B)** Total number of introductions by state from each start location, annotated as per legend in
 288 colour, to each end location on the y-axis. The regions of the end location state on the y-axis are
 289 highlighted in colour in the plot background as per the legend. **C)** The distribution of the number of
 290 introductions across time by state. The end location state is coloured by region, as per legend. The start
 291 location is highlighted by transparency: all introductions originating from Rivers State are presented
 292 with no transparency, whereas introductions originating from other states are more transparent. **D)**
 293 Continuous phylogeography of hMPXV-1 spatiotemporal spread across Nigeria, with timing of viral
 294 dissemination highlighted by colour range as per legend.
 295

296 It is not clear whether cases across Nigerian states were continuously seeded by
 297 interstate introductions from Rivers and the south, or whether there were locally
 298 persistent transmission chains in other regions sustaining local epidemics. Towards
 299 understanding the respective contribution of persistence and introductions, we
 300 investigated the persistence of transmission chains in each state. We found that
 301 hMPXV-1 has persistently circulated in Rivers from emergence onwards (Figure 4A,
 302 B). hMPXV-1 diversified in Rivers State for more than two years before the first case
 303 was reported in Bayelsa on 11 September 2017 (Figure 4A, Extended Data Figure

304 4).^{5,6} When the outbreak was declared, transmission chains had already been
305 established in 15 states outside of Rivers (Figure 4A). Delta and Bayelsa in the South
306 South had the second longest persistence of a transmission chain at approximately
307 four and a half and three years respectively, followed by the earliest chain established
308 in Lagos at three years (Figure 4 A,B). Outside of Rivers State and the early chains in
309 Bayelsa and Lagos, the longest persistence was estimated for lineages introduced
310 during the period of low reported incidence in 2018-2021, when sampling was sparse
311 (Figure 4B, 1A).¹⁹

312
313 We found that persistent transmission was the primary driver of the epidemic in Rivers
314 State, relative to repeat introductions (Figure 4A, C). However, the percentage of
315 transmission chains persistently circulating was dynamic over time in other South
316 South states (Figure 4 D). Transmission chains seeded by Rivers State early in the
317 epidemic in the South South, South West and South East only persisted locally for
318 less than two years, excluding the first transmission chains established in Bayelsa and
319 Lagos (Figure 4 A, B). There was a significant increase in the number of transmission
320 chains circulating during the resurgence of cases towards 2022 (Figure 4C). This is
321 consistent with the increased viral exports during this period (Figure 3C), seeding new
322 transmission chains that drove local surges across Nigeria (Figure 1A, Figure 4A).
323 Transmission chains from the later stage of the epidemic were predominantly
324 introduced from Rivers, and persisted for less than two years (Figure 4A). However, it
325 is unclear whether this pattern persists past the end of our sampling frame. There was
326 no evidence for significant persistence in Northern states prior to the later phase of the
327 epidemic, which is consistent with the low reported incidence and delayed viral spread

328 observed (Figure 4 A, 3D, Figure 1A). Altogether, this further supports that Rivers
 329 State acted as the persistent source for the epidemic, while local epidemics in other
 330 states were largely driven by repeat introductions.
 331
 332 To account for uneven sampling across states, we also performed our
 333 phylogeographic analyses at the regional level. We observed a consistent pattern to
 334 our state-level analyses: early and predominant spread from the South South, with
 335 initial spread to the South East and South West. There was strong evidence of
 336 persistent circulation in the South South, with local epidemics in all other regions
 337 driven by repeated introduction from the South South (Extended Data figures 1-4).



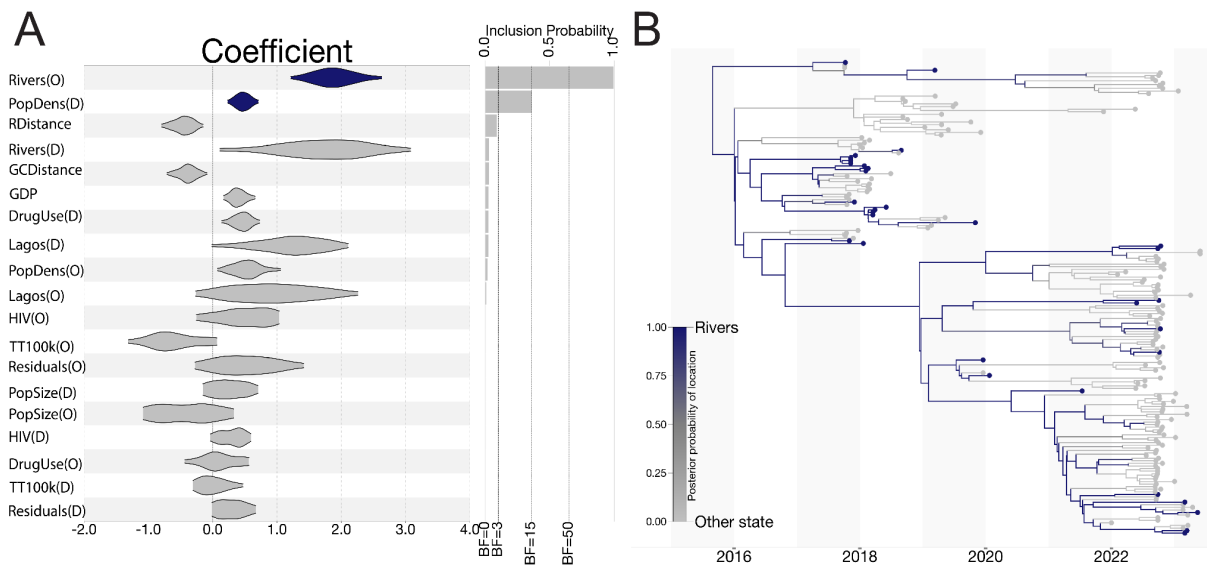
338
 339 **Figure 4:** **A)** Persistence of transmission chains across all Nigerian states sampled. Individual chains
 340 are coloured by region, with the boundary of each individual state highlighted by a filled background
 341 and annotated in text on the right. The start of each transmission chain is coloured by its state of origin.
 342 The red line indicates the date of report for the first case in Bayelsa on 17 September 2017. **B)** The
 343 persistence in years of each transmission chain across its time of origin, coloured by region. **C)** The
 344 number of transmission chains circulating across all regions across time, calculated by a month-sliding
 345 window. The red line indicates the date the outbreak was declared on 22 September 2017. SS: South
 346 South; SW: South West; SE: South East; NW: North West; NE: North East; NC: North Central. **D)** The
 347 percentage of transmission chains persisting across time for South South states excluding Rivers.

348 **Rivers State origin was associated with hMPXV-1** 349 **dispersal**

350 Our phylogeographic reconstructions consistently support a spatiotemporal pattern of
351 early viral spread between and then from southern states, with hMPXV-1 persistently
352 circulating in the South South region. To identify potential drivers of this pattern, we
353 used a phylogeographic generalised linear model that integrates covariates of spatial
354 spread to determine what factors were associated with hMPXV-1 dispersal. We
355 incorporated covariates in our model including epidemiological, demographic,
356 geographic and economic variables, as well as location-specific predictors that capture
357 a pairwise binary transition across different states (see Extended Data table 1).

358

359 Of the 19 covariates analysed, we found that the main predictor that positively affected
360 viral dispersal was whether the lineage originated in Rivers State ($BF > 50$) (Figure 5A,
361 B). There was no support for the residual covariates that assessed the deviations of
362 sampling numbers relative to epidemiological cases (Figure 5A). This suggests that
363 the out-of-Rivers pattern is robust and was not due to sampling heterogeneity across
364 locations. This finding supports our previous analyses highlighting Rivers' early and
365 dominant role in the spread of hMPXV-1 from emergence onwards (Figure 5B). The
366 population density in the destination was also positively associated with hMPXV-1
367 dispersal ($BF > 15$), which is consistent with the principles of a gravity model in
368 epidemiology (Figure 5A).^{20,21}



369 **Figure 5: A)** The GLM coefficients of covariates of spatial spread and their associated Bayes Factors.
370 Significant covariates are highlighted in purple. **B)** Phylogeographic reconstruction of the migratory
371 pattern of hMPXV-1 from Rivers to other Nigerian states.
372
373

374 Informing real-time public health decision making 375 with Delphy

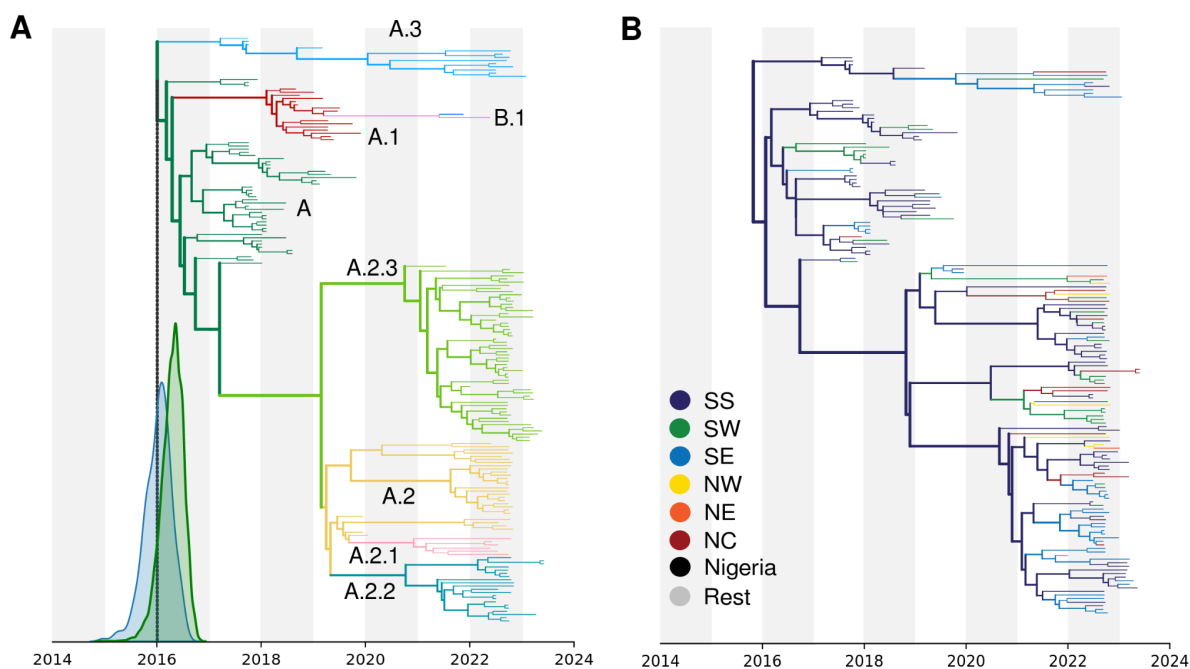
376 Effective outbreak responses rely on rapid action and real-time information to guide
377 decision making. To provide public health professionals with a broadly accessible tool
378 for near-real-time Bayesian phylogenetics, we have developed a new tool named
379 Delphy.²² Central to Delphy's speed is a reframing of Bayesian phylogenetics that
380 exploits the characteristics of typical genomic epidemiology datasets to make it both
381 faster and more scalable (see ²², Methods and SI). Delphy is accessible as a client-
382 side web application at <https://delphy.fathom.info>, which provides an intuitive interface
383 for analysing the entire posterior distribution of trees.²³

384
385 To demonstrate Delphy's utility in outbreak response analytics, we adapted it to
386 analyse our mpox sequences from the human epidemic in Nigeria. For this analysis,
387 we assume that any zoonotic spillover event precedes the MRCA of hMPXV-1 (see
388 Methods). We found that Delphy produced consistent estimates with our analogous

389 BEAST runs, with credible intervals overlapping (Figure 6A, 2A). The tMRCA estimate
390 of the hMPXV-1 clade from Delphy was 18 Jan 2016 (95% HPD 17 Jun 2015 to 1 July
391 2016), compared to the BEAST estimate of 15 July 2015 (95% HPD 8 November 2014
392 - 22 February 2016). The estimates for the APOBEC and polymerase-driven
393 substitution rate as well as the doubling time of hMPXV-1 were also consistent: for the
394 polymerase, it was 8.6×10^{-6} substitutions per site per year (95% HPD 7.4×10^{-6} -
395 9.9×10^{-6}) with Delphy and 9.7×10^{-6} substitutions per site per year (95% HPD 8.2×10^{-6} -
396 1.1×10^{-5}) for BEAST; for the APOBEC-driven substitution rate the estimates were
397 1.1×10^{-4} substitutions per site per year (95% HPD 1×10^{-4} - 1.3×10^{-4}) for Delphy and
398 1×10^{-4} substitutions per site per year (95% HPD 8.8×10^{-5} - 1.14×10^{-4}) for BEAST; the
399 doubling time estimates were 2 years (95% HPD 1.6 - 2.5 years) for Delphy and 2.2
400 years (95% HPD 1.7 - 2.9) for BEAST.

401
402 Delphy can also be used to reconstruct simple estimates of tip-associated metadata
403 at the inner nodes of a Maximum Clade Credibility tree (MCC) with a simple parsimony
404 algorithm. We found that the simplified parsimony approach recovered spatial
405 dynamics consistent with the phylogeographic patterns reconstructed in the BEAST
406 analyses (Figure 6A vs 1D, Figure 6B vs Figure 3B, Extended Data Figure 5).

407
408 Delphy allows public health professionals to approximate the type of outbreak
409 analytics demonstrated in this paper in an accessible client-side web application in
410 near real-time. The Delphy run was completed in 30 minutes on a standard 4-core
411 2020 laptop without a GPU, relative to the 8 hour run time of the BEAST analyses on
412 a standard M1 Macstudio.



413

414 **Figure 6: A)** Delphy MCC for hMPXV-1 samples, coloured by nomenclature¹³ (hosted at
415 <https://delphy.fathom.info/?mpox2024-post-spillover-with-metadata.dphy>); compare to Figs 2A and 1D.
416 The tMRCAs distribution calculated by Delphy (blue) overlaps with that of an analogous BEAST run on
417 the same sequences (green); the run for Figure 2A also includes the remaining zoonotic sequences.
418 **B)** Delphy MCC for hMPXV-1 sequences from Nigeria only (hosted at
419 <https://delphy.fathom.info/?mpox2024-post-spillover-nigeria-only-final-with-metadata.dphy>), coloured
420 by geographical origin at the tips and propagated to inner nodes via naive parsimony; compare to Fig
421 3A.
422

423 Discussion

424 Consensus evidence now strongly supports that MPXV Clade IIb is no longer solely a
425 zoonotic disease but is now sustained in a human subpopulation in Nigeria, where it
426 has been circulating cryptically for nearly a decade. However, the drivers behind this
427 emergence in humans remains uncertain.¹⁷ It is not clear why this cross-species
428 transmission event did not simply result in a sporadic case or a self-limiting
429 transmission chain as all previous zoonotic MPXV infections. Nonetheless, it is likely
430 that the spillover event occurred in a more connected, mobile subpopulation in Nigeria

431 with more probable onward transmission driven by behavioural or demographic
432 factors.¹⁷

433 The epidemic in Nigeria has already resulted in the global B.1 outbreak, as well as
434 several self-limiting exports of distinct Clade IIb sub-lineages detected in countries
435 with higher levels of surveillance.¹⁶ It is possible that there is related or independent
436 cryptic sustained human-to-human transmission in other regional countries with less
437 robust surveillance systems, including for Clade IIa and Clade I.²⁴ However, we did
438 not detect any evidence of sustained human-to-human transmission in Cameroon.¹⁷ It
439 is highly likely that viral export and potential seeding of outbreaks, both regional and
440 intercontinental, will continue as the virus circulates in the human population. It is also
441 possible that sustained circulation will result in viral adaptation through APOBEC3-
442 accelerated evolution across larger chains of infection.²⁵ It is therefore vital that the
443 Nigerian epidemic is resolved or reduced by targeted public health interventions. Our
444 study has provided critical insights to facilitate strategic public health interventions.
445 Across our phylogeographic reconstructions, we found that Rivers State and other
446 South South states served as early, dominant and persistent sources of viral export.
447 Interventions should accordingly be targeted to these regions. Notably, we find
448 evidence of prolonged cryptic circulation and geographic expansion before detection
449 in these regions, emphasising the need for enhanced surveillance and improved
450 diagnostic and surveillance infrastructure in these regions. We also estimated a
451 relatively slow doubling rate, suggesting that the virus is circulating in a restricted
452 subpopulation.²⁵ Enhanced surveillance of cases is required to characterise the

453 underlying transmission network and associated risk factors, allowing for targeted
454 interventions before the epidemic becomes more generalised.

455 However, our results should be interpreted within the limits of our sample. Our sample
456 represents all available and viable samples from routine surveillance, but were not
457 systematically selected for spatiotemporal representativeness. However, our sample
458 represents 4.2% of all suspected mpox cases in Nigeria from 2017 onwards. Our
459 samples are predominantly collected from the South South and South East regions,
460 and do not include the period closer to the estimated emergence or the start of the
461 epidemic in 2017. However, as Southern regions represented the highest number of
462 cases throughout the epidemic from 2017 to the 2022 resurgence, it is unlikely that
463 the geographic distribution of samples represents a strong sampling bias.

464
465 Controlling the ongoing mpox epidemic in Nigeria is impeded by inequities of access
466 to resources such as diagnostics, vaccines and therapeutics. After the WHO declared
467 mpox a Public Health Emergency of International Concern, high income countries
468 increased and deployed their stockpiles of smallpox vaccines, including third
469 generation vaccines.^{26,27} However, these resources have not been made available to
470 nations in Africa, despite the historic increased incidence, morbidity and mortality of
471 the disease on the continent.²⁸⁻³⁰ Tecovirimat, a smallpox antiviral effective against
472 mpox, has also not been made accessible for African nations.²⁸ Without access to
473 therapeutics and vaccines, transmission cannot be reduced in either the sustained
474 human epidemic or in populations at high risk for recurrent spillovers from the
475 reservoir. Ongoing zoonotic and human transmission in Africa does not only increase
476 the probability of re-emergence, future global epidemics and potential viral adaptation

477 but, vitally, results in preventable morbidity and mortality in endemic countries. The
478 global community can no longer afford to neglect mpox in Africa or perpetuate
479 inequities in therapeutic access in our vulnerably connected world.

480

481 **Methods:**

482 **Ethics declaration**

483 No ethical approval was required for this study as it is based on data from Nigeria's
484 national surveillance program, collected by the Nigeria Centre for Disease Control.
485 Under the program, individual written or oral informed consent was obtained from all
486 suspected mpox cases. Informed consent for children was obtained from their parents
487 or recognized guardians.

488 **Sampling**

489 Samples were collected by laboratory personnel and Local Government Area (LGA)
490 Disease Surveillance and Notification Officers (DSNOs) equipped with appropriate
491 personal protective equipment (PPE), adhering to the guidelines outlined in the Nigeria
492 Centre for Disease Control and Prevention (NCDC) National Monkeypox Public Health
493 Response Guidelines.³¹ Samples comprised: swabs from the exudate of vesicular or
494 pustular lesions, lesion crusts obtained during the acute rash phase, whole blood
495 collected in ethylenediaminetetraacetic acid (EDTA) or plain/non-anticoagulated
496 tubes. All samples were labelled with case information and stored at 2-8°C during

497 transport to either the NCDC National Reference Laboratory (Gaduwa-Abuja) or the
498 Central Public Health Laboratory (Yaba-Lagos). On arrival, the crusts and swabs were
499 eluted, while the serum/plasma was separated from the red blood cells. Subsequently,
500 these components were stored at ultralow temperatures of $\leq -70^{\circ}\text{C}$ at the NCDC
501 biorepository.

502 **Genome Sequencing**

503 Enrichment bead-linked transposomes was used to tagment the extracted DNA and
504 enriched using the Illumina- rna -prep enrichment with the VSP panel. Libraries were
505 quantified using dsDNA BR Assay, normalized to a concentration of 0.6nM and
506 sequenced on the Illumina NovaSeq 6000 platform with a read length of 151 base pair
507 paired end at the African Centre of Excellence for Genomics of Infectious Diseases
508 (ACEGID), based at Redeemer's University, Ede, Nigeria.

509 **Genome Assembly**

510 We performed initial *de novo* assembly with the viral-ngs pipeline, followed by
511 reference based assembly with an in-house pipeline.³² Briefly, we mapped reads
512 against a Clade IIb reference genome (NC_063383, an early hMPXV-1 genome from
513 Nigeria) with *bwa-mem*³³, and called consensus using samtools³⁴ and iVar.³⁵

514 **Genomic dataset curation**

515 We combined our 112 genomes with all high-quality, publicly available Clade IIb MPXV
516 genomes from Genbank (as of August 2023). This included all Clade IIb genomes from
517 2017-2022 sampled in Nigeria, sequences from non-endemic countries with a travel
518 history to Nigeria and two older sequences sampled in Nigeria in 1971 and 1978
519 (accession numbers KJ642617 and KJ642615, respectively). We included a single
520 representative of the global outbreak lineage B.1, as it was not our primary focus. In
521 total, our dataset consists of 202 sequences.

522 **Phylogenetic Analysis**

523 We aligned our dataset to the Clade IIb reference genome (NC_063383) using the
524 'squirrel' package (<https://github.com/aineniamh/squirrel>) developed by O'Toole *et*
525 *al.*¹¹ The alignment was trimmed, and the 3' terminal repeat region and regions of
526 repetition or low complexity were masked.

527

528 We investigated the preliminary placement of our sequences in a phylogeny of all
529 available mpox genome sequences from Genbank across clades. We reconstructed
530 the complete MPXV phylogeny with IQ-TREE v2.0, under the Jukes-Cantor
531 substitution model.³⁶ We identified three B.1 lineage sequences in our dataset. To
532 investigate whether these sequences represented re-importations to Nigeria, we
533 reconstructed a phylogeny with 769 B.1 genomes from Genbank. We confirmed the
534 sequences represented re-importations of B.1 into Nigeria and excluded them from
535 subsequent analyses (data not shown).

536

537 We reconstructed a phylogeny for Clade IIb alone under the same parameters as the
538 global phylogeny. We rooted our phylogeny to KJ642615, a sequence sampled in
539 Nigeria in 1978, as it is notably divergent from the remaining Clade IIb diversity. We
540 collapsed all zero branch lengths. We performed ancestral state reconstruction with
541 IQ-TREE2 across the Clade IIb phylogeny.³⁶ We mapped all nucleotide mutations that
542 occurred across the phylogeny to internal branches using tree traversal, excluding
543 missing data. Additionally, we catalogued the dimer genomic context of all C→T or
544 G→A mutations, as described by O’Toole *et al.*¹¹ We classified our sequences into
545 lineages under the nomenclature developed in Happi *et al.*¹³ using Nextclade.³⁷

546 **Modelling APOBEC3-mediated evolution**

547 We adopted a similar approach described by O’Toole *et al.*¹¹ to analyse the
548 evolutionary dynamics of hMPXV-1 in the software package BEAST³⁸ with the
549 BEAGLE high-performance computing library.³⁹ First, we partitioned the Clade IIb
550 alignment into two distinct partitions. The first partition comprised sites with potential
551 APOBEC3 modifications (specifically C→T and G→A substitutions in the dinucleotide
552 context TC and GA), along with target sites (e.g. C and G) that were conserved. In this
553 partition, we masked all other sites as ambiguous nucleotides. This partition
554 represents APOBEC3 mutations relative to the target APOBEC3 sites. The second
555 partition inversely contained sites with the APOBEC3 target sites masked. The
556 APOBEC3 alignment comprised 24 680 unmasked sites, whereas the non-APOBEC3
557 alignment comprised 172 529 sites. We used the standard nucleotide GTR+G
558 substitution model with four distinct rate categories for the non-APOBEC3 partition.

559 For the APOBEC3 partition, we developed a substitution process where we categorise
560 the nucleotides as modified (T) and unmodified (C). We used a two-state continuous-
561 time Markov chain with an asymmetric rate to permit C→T mutations but not the
562 reverse.

563
564 We used a two-epoch model to estimate the time of MPXV's emergence in the human
565 population. Under this model, the evolutionary rate transitions from the background
566 rate (i.e. non-APOBEC3 rate, driven by polymerase error rate) to the APOBEC3 rate
567 at a specific time point t_p for the APOBEC3 partition. We parameterized this transition
568 time as $t_p = t_{MRCA}(\text{Lineage A}) + x$, where x is a free parameter in BEAST representing
569 the pre-sampled transmission history before the most recent common ancestor
570 (MRCA) of sampled Lineage A viruses.¹¹ We incorporated a local clock to scale the
571 mutation proportion attributed to APOBEC3 activity across the branches up to the
572 transition time. We allowed the non-APOBEC3 partition to evolve under the
573 background evolutionary rate across the entire phylogeny.

574
575 We additionally used a two-phase coalescent model: the tree from the MRCA (Lineage
576 A) onward was modelled with an exponential growth model, with the earlier phase
577 modelled as a constant-population size coalescent model. We estimated the virus'
578 doubling time relative to the transition time. The doubling time is expressed as
579 $\log(2)/\text{growth rate}$, with the growth rate estimated from the exponential model. We
580 further investigated Lineage A's epidemic growth pattern using a non-parametric
581 coalescent Skygrid model with 12 change points over 10 years, excluding the zoonotic
582 portion of the tree. For each model, we ran two independent chains of 100 million
583 states to ensure convergence, discarding the initial 10% of each chain as burn-in. The

584 chains were then combined with LogCombiner. For all subsequent analyses, we
585 assessed convergence using Tracer, and constructed a maximum clade credibility
586 (MCC) tree in TreeAnnotator 1.10.⁴⁰

587 **Geographic history of hMPXV-1 in Nigeria**

588 **Discrete phylogeographic analysis**

589 To investigate the spread of hMPXV-1 across Nigeria, we reconstructed the timing
590 and pattern of geographic transitions across Nigerian states under an asymmetric
591 discrete trait analyses.⁴¹ We used Bayesian stochastic search variable selection
592 (BSSVS) to infer non-zero migration rates and identify statistically supported migration
593 routes. We used a non-parametric skygrid coalescent tree prior, with 12 change points
594 distributed over 10 years as described above.⁴² We combined two independent MCMC
595 runs of 50 million states each, sampling every 2000 states and discarding the
596 respective initial 10% of trees as burn-in. We confirmed all ESS values are above 200.
597
598 We used a Markov jump counting procedure to investigate the timing and origin of
599 geographic transitions, or Markov jumps, across the full posterior to account for
600 uncertainty in phylogeographic reconstruction.⁴³ We used the
601 TreeMarkovJumpHistoryAnalyzer from the pre-release version of BEAST 1.10.5 to
602 obtain the Markov jumps from posterior tree distributions.¹⁹ Using the tree distribution
603 annotated with Markov jumps, we performed the persistence analysis on a month-to-
604 month interval to calculate the percentage of lineages that persisted in their ancestral

605 state for each Nigerian state and region.¹⁹ We used the PersistenceSummarizer from
606 the pre-release version of BEAST 1.10.5.

607
608 We also performed all phylogeographic analyses on a regional level, to account for
609 the uneven distribution of sequences across Nigerian states. We categorised the
610 states into the six geopolitical zones of Nigeria.⁴⁴ We have a limited number of
611 sequences from Northern Nigeria, which had very low epidemic incidence from 2017
612 - 2022. To account for this, we combined the North West and North East zones into a
613 single North category. All trees were visualised using baltic
614 (<https://github.com/evogytis/baltic>).

615 **Continuous phylogeographic analysis**

616 We performed a continuous phylogeographic analysis to quantify the dispersal of
617 hMPXV-1 across Nigerian states. We assigned each sequence a latitude and
618 longitude that matched the local government area or village of collection. We used the
619 two-epoch and skygrid coalescent model described above, with a Cauchy distribution
620 to model the among-branch heterogeneity in dispersal velocity.⁴⁵ We ran two
621 independent MCMC chains of 50 million states, sampling every 2000 states. We
622 combined the chains after discarding 10% of the states as burn-in.

623 **Discrete phylogeographic analysis using a sparse General Linear** 624 **Model (GLM)**

625 To investigate the drivers of the transmission dynamics of the hMPXV-1 epidemic, we
626 used a sparse generalised linear model (GLM) in a continuous-time Markov chain

627 (CTMC) diffusion framework.⁴⁶ We considered 19 covariates in the model, including
628 epidemiological case counts relative to the numbers of sampled genomes per
629 locations, demographic data, and geographic and economic factors (Extended Data
630 Table 1). Besides the location-specific predictors, which were encoded as binary
631 variables to reflect migration patterns, we log-transformed and standardised all other
632 predictors. This standardisation involved adding a pseudo-count to each entry to
633 ensure a robust analysis.¹⁹ We ran two independent MCMC chains of 50 million
634 iterations, sampling every 2000 iterations. We combined the resulting posterior
635 distributions after removing the initial 10% as burn-in.

636 **Covariate collation**

637 Covariates (Extended Data table 1) included in the GLM were collected from the
638 following sources: Epidemiological data was obtained from the Nigerian CDC;
639 Economic covariates were sourced from Okeowo et al.⁴⁷; Population covariates were
640 sourced from the Bulletin of the National Bureau of Statistics⁴⁴³; HIV prevalence was
641 obtained from the PEPFAR program⁴⁸; Drug use statistics were obtained from the
642 National Bureau of Statistics.⁴⁹ The distance by road travel between each states was
643 calculated from Google Maps.

644 **Geographical metadata**

645 Administrative level 2 (admin2) metadata for the sampling location of sequences in the
646 dataset were mapped to official admin2 as found in the Global Administrative
647 Database (GADM, <https://gadm.org>).

648 **Delphy**

649 We are preparing a separate, detailed paper on Delphy.²² Here, we present only the
650 essential details to motivate the public health surveillance application we describe in
651 the main text. Delphy is a reframing of Bayesian phylogenetics in terms of Explicit
652 Mutation-Annotated Trees (EMATs), which exploits the near-parsimonious nature of
653 trees over genomic epidemiology datasets with limited genetic diversity to gain
654 significant efficiencies. See Supplementary Information for a more precise
655 formulation.

656

657 In the version of Delphy adapted for mpox, we classify each site as having APOBEC3
658 context according to the sequence of the first sample in the input file (with missing
659 sites replaced by the consensus across all input sequences). A site has APOBEC3
660 context if it and the preceding site form a **TC** or a **TT** dimer, or if it and its subsequent
661 site form a **GA** or **AA** dimer. For sites without APOBEC3 context, we use a transition
662 rate matrix for a Jukes-Cantor evolutionary model with mutation rate μ and no site-rate
663 heterogeneity:

$$Q_0 = \mu \begin{pmatrix} -1 & 1/3 & 1/3 & 1/3 \\ 1/3 & -1 & 1/3 & 1/3 \\ 1/3 & 1/3 & -1 & 1/3 \\ 1/3 & 1/3 & 1/3 & -1 \end{pmatrix}.$$

664

665

666 In sites with APOBEC3 context, we add transitions from C->T and G->A at a rate μ^* ,
667 as follows:

$$Q_1 = \mu \begin{pmatrix} -1 & 1/3 & 1/3 & 1/3 \\ 1/3 & -1 & 1/3 & 1/3 \\ 1/3 & 1/3 & -1 & 1/3 \\ 1/3 & 1/3 & 1/3 & -1 \end{pmatrix} + \mu^* \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & +2 \\ +2 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

668

669

670 The factor of 2 is chosen to match the conventions used in the BEAST runs to define
671 the APOBEC3 rate (the effective rate at which transitions would be observed in a
672 sequence of APOBEC3 sites that are 50% C and 50% T).

673

674 The above setup retains the essence of that of the BEAST runs but differs minimally
675 in its details. By deciding on the APOBEC context using the first sequence, we avoid
676 having to previously estimate an ML tree and extract its root sequence (the assumption
677 is that the APOBEC3 context of the vast majority sites is stable throughout the entire
678 tree), and allows us to handle arbitrary new datasets, possibly aligned to a different
679 reference, without complications. By preserving the polymerase mutation mechanism
680 in sites with APOBEC3 context (i.e., the first term in the definition of Q_1), we avoid
681 having to identify only the sites that have not yet mutated (i.e., only the **TC** and **GA**
682 sites) and avoid technical difficulties with completely irreversible transition matrices.
683 For simplicity, we also use a simple Jukes-Cantor model with no site rate
684 heterogeneity instead of a more sophisticated GTR model with 4 gamma rate
685 categories. Delphy does not yet implement tip-date sampling, so tips with uncertain
686 dates are imputed to the middle of the date range. Since Delphy does not yet
687 implement a Skygrid population model, we model the viral population as growing
688 strictly exponentially (see Figure~2B). Finally, by design, we have not implemented a
689 transition between a pre-spillover and post-spillover portion of the tree, as our interest
690 is in public health response to a growing outbreak that is known to be spreading

691 through humans. We suspect this is the dominant factor in the discrepancies between
692 the tMRCAs and mutation rates of the main BEAST run in Figures 2 and 3 and the
693 Delphy runs. The long branches near the top of the tree extending to the late 19th
694 century likely exhibit purifying selection that reduces their observed substitution rate¹⁷,
695 which depresses the estimate of the polymerase-driven mutation rate and pushes
696 tMRCAs towards the past.

697

698 **Data availability**

699 All sequences are available on Genbank under Accession numbers PP852943 -
700 PP853055. All other data are available at [https://github.com/andersen-lab/hMPXV-](https://github.com/andersen-lab/hMPXV-1_Nigeria)
701 [1_Nigeria](https://github.com/andersen-lab/hMPXV-1_Nigeria), or upon request.

702 **Code availability**

703 All code to run the analyses is available in [https://github.com/andersen-lab/hMPXV-](https://github.com/andersen-lab/hMPXV-1_Nigeria)
704 [1_Nigeria](https://github.com/andersen-lab/hMPXV-1_Nigeria). Delphy's code is available at <https://github.com/broadinstitute/delphy> and
705 <https://github.com/fathominfo/delphy-web>.

706 **References**

- 707 1. Nakazawa, Y. *et al.* A phylogeographic investigation of African monkeypox. *Viruses* **7**, 2168–
708 2184 (2015).
- 709 2. Doty, J. B. *et al.* Assessing Monkeypox Virus Prevalence in Small Mammals at the Human–
710 Animal Interface in the Democratic Republic of the Congo. *Viruses* **9**, 283 (2017).
- 711 3. Breman, J. G. *et al.* Human monkeypox, 1970-79. *Bull. World Health Organ.* **58**, 165–182 (1980).
- 712 4. McCollum, A. M. & Damon, I. K. Human monkeypox. *Clin. Infect. Dis.* **58**, 260–267 (2014).
- 713 5. Yinka-Ogunleye, A. *et al.* Reemergence of Human Monkeypox in Nigeria, 2017. *Emerg. Infect.*
714 *Dis.* **24**, 1149–1151 (2018).
- 715 6. Yinka-Ogunleye, A. *et al.* Outbreak of human monkeypox in Nigeria in 2017–18: a clinical and
716 epidemiological report. *Lancet Infect. Dis.* **19**, 872–879 (2019).
- 717 7. Hutin, Y. J. *et al.* Outbreak of human monkeypox, Democratic Republic of Congo, 1996 to 1997.
718 *Emerg. Infect. Dis.* **7**, 434–438 (2001).
- 719 8. Borges, V. *et al.* Viral genetic clustering and transmission dynamics of the 2022 mpox outbreak
720 in Portugal. *Nat. Med.* **29**, 2509–2517 (2023).
- 721 9. Firth, C. *et al.* Using time-structured data to estimate evolutionary rates of double-stranded
722 DNA viruses. *Mol. Biol. Evol.* **27**, 2038–2051 (2010).
- 723 10. Suspène, R. *et al.* APOBEC3F Is a Mutational Driver of the Human Monkeypox Virus Identified in
724 the 2022 Outbreak. *J. Infect. Dis.* **228**, 1421–1429 (2023).
- 725 11. O’Toole, Á. *et al.* APOBEC3 deaminase editing in mpox virus as evidence for sustained human
726 transmission since at least 2016. *Science* **382**, 595–600 (2023).
- 727 12. Ndodo, N. *et al.* Distinct monkeypox virus lineages co-circulating in humans before 2022. *Nat.*
728 *Med.* **29**, 2317–2324 (2023).

- 729 13. Happi, C. *et al.* Urgent need for a non-discriminatory and non-stigmatizing nomenclature for
730 monkeypox virus. *PLoS Biol.* **20**, e3001769 (2022).
- 731 14. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic
732 epidemiology. *Nat Microbiol* **5**, 1403–1407 (2020).
- 733 15. Isidro, J. *et al.* Phylogenomic characterization and signs of microevolution in the 2022 multi-
734 country outbreak of monkeypox virus. *Nat. Med.* **28**, 1569–1572 (2022).
- 735 16. Gigante, C. M. *et al.* Multiple lineages of monkeypox virus detected in the United States, 2021-
736 2022. *Science* **378**, 560–565 (2022).
- 737 17. Djuicy, D. D. Molecular epidemiology of recurrent zoonotic transmission of mpox virus in West
738 Africa. Preprint at <https://doi.org/10.5281/ZENODO.11653044> (2024).
- 739 18. Fraser, C. *et al.* Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* **324**,
740 1557–1561 (2009).
- 741 19. Lemey, P. *et al.* Untangling introductions and persistence in COVID-19 resurgence in Europe.
742 *Nature* **595**, 713–717 (2021).
- 743 20. Viboud, C. *et al.* Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*
744 **312**, 447–451 (2006).
- 745 21. Dudas, G. *et al.* Virus genomes reveal factors that spread and sustained the Ebola epidemic.
746 *Nature* **544**, 309–315 (2017).
- 747 22. Varilly, P. *Broadinstitute/delphy: v0.995*. doi:10.5281/zenodo.12073532.
- 748 23. Schifferli, M. *Fathominfo/delphy-Web: v0.0.995*. (Zenodo, 2024).
749 doi:10.5281/ZENODO.11998924.
- 750 24. World Health Organization. Mpox (monkeypox)-Democratic Republic of the Congo. 2023– 11–
751 23][2023– 12– 18]. <https://www.who.int>.
- 752 25. Johnson, P. L. F. *et al.* Evolutionary consequences of delaying intervention for monkeypox. *The*

- 753 *Lancet* vol. 400 1191–1193 (2022).
- 754 26. Kozlov, M. Monkeypox vaccination begins - can the global outbreaks be contained? *Nature* **606**,
755 444–445 (2022).
- 756 27. Mitjà, O. *et al.* Monkeypox. *Lancet* **401**, 60–74 (2023).
- 757 28. Kozlov, M. WHO may soon end mpox emergency - but outbreaks rage in Africa. *Nature* **614**,
758 600–601 (2023).
- 759 29. Huhn, G. D. *et al.* Clinical characteristics of human monkeypox, and risk factors for severe
760 disease. *Clin. Infect. Dis.* **41**, 1742–1751 (2005).
- 761 30. Kozlov, M. Monkeypox in Africa: the science the world ignored. *Nature* vol. 607 17–18 (2022).
- 762 31. Nigeria Centre for Disease Control. National monkeypox public health response guidelines.
763 Preprint at https://ncdc.gov.ng/themes/common/docs/protocols/96_1577798337.pdf.
- 764 32. Park, D. *et al.* *Broadinstitute/viral-Ngs: v1.19.2*. (Zenodo, 2018). doi:10.5281/zenodo.1167849.
- 765 33. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform.
766 *Bioinformatics* **25**, 1754–1760 (2009).
- 767 34. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079
768 (2009).
- 769 35. Grubaugh, N. D. *et al.* An amplicon-based sequencing framework for accurately measuring
770 intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* **20**, 8 (2019).
- 771 36. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in
772 the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
- 773 37. Aksamentov, I., Roemer, C., Hodcroft, E. & Neher, R. Nextclade: clade assignment, mutation
774 calling and quality control for viral genomes. *J. Open Source Softw.* **6**, 3773 (2021).
- 775 38. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST
776 1.10. *Virus Evol.* **4**, vey016 (2018).

- 777 39. Ayres, D. L. *et al.* BEAGLE 3: Improved Performance, Scaling, and Usability for a High-
778 Performance Computing Library for Statistical Phylogenetics. *Syst. Biol.* **68**, 1052–1061 (2019).
- 779 40. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in
780 Bayesian phylogenetics using tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
- 781 41. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its
782 roots. *PLoS Comput. Biol.* **5**, e1000520 (2009).
- 783 42. Gill, M. S. *et al.* Improving Bayesian population dynamics inference: a coalescent-based model
784 for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013).
- 785 43. Minin, V. N. & Suchard, M. A. Counting labeled transitions in continuous-time Markov models of
786 evolution. *J. Math. Biol.* **56**, 391–412 (2008).
- 787 44. National Bureau of Statistics. Demographic Statistics Bulletin 2020. Preprint at
788 <https://nigerianstat.gov.ng/elibrary/read/1241121>.
- 789 45. Lemey, P., Rambaut, A., Welch, J. J. & Suchard, M. A. Phylogeography takes a relaxed random
790 walk in continuous space and time. *Mol. Biol. Evol.* **27**, 1877–1885 (2010).
- 791 46. Lemey, P. *et al.* Unifying viral genetics and human transportation data to predict the global
792 transmission dynamics of human influenza H3N2. *PLoS Pathog.* **10**, e1003932 (2014).
- 793 47. BudgIT. State of States 2022. Preprint at [https://yourbudgit.com/wp-](https://yourbudgit.com/wp-content/uploads/2022/10/2022-State-of-States-Report.pdf)
794 [content/uploads/2022/10/2022-State-of-States-Report.pdf](https://yourbudgit.com/wp-content/uploads/2022/10/2022-State-of-States-Report.pdf).
- 795 48. PEPFAR. 2019 Country Operational Plan Guidance for all PEPFAR Countries. Preprint at
796 [https://www.state.gov/wp-content/uploads/2019/08/PEPFAR-Fiscal-Year-2019-Country-](https://www.state.gov/wp-content/uploads/2019/08/PEPFAR-Fiscal-Year-2019-Country-Operational-Plan-Guidance.pdf)
797 [Operational-Plan-Guidance.pdf](https://www.state.gov/wp-content/uploads/2019/08/PEPFAR-Fiscal-Year-2019-Country-Operational-Plan-Guidance.pdf).
- 798 49. National Bureau of Statistics. Drug Use in Nigeria. Preprint at
799 https://www.nigerianstat.gov.ng/pdfuploads/Drug_Use_in_Nigeria_2018.pdf.

800 **Acknowledgements**

801 This work is made possible by support from Flu Lab and a cohort of generous donors
802 through TED's Audacious Project, including the ELMA Foundation, MacKenzie Scott,
803 the Skoll Foundation, and Open Philanthropy. This work was supported by grants from
804 the National Institute of Allergy and Infectious Diseases grants U01HG007480
805 (H3Africa), U54HG007480 (H3Africa), U01AI151812 (WARN-ID), U19AI135995
806 (CViSB), U19AI110818 (GCID), R01AI153044 and R01AI162611.). This work was
807 also supported by the World Bank grants projects ACE-019 and ACE-IMPACT; The
808 Rockefeller Foundation (Grant #2021 HTH); The Africa CDC through the African
809 Society of Laboratory Medicine [ASLM] (Grant #INV018978), and the Science for
810 Africa Foundation. Ifeanyi Omah is supported by the Wellcome Trust Hosts, Pathogens
811 & Global Health program [Wellcome Trust, Grant number 218471/Z/19/Z] in partnership
812 with Tackling infectious Disease to Benefit Africa, TIBA. PL is supported by the European
813 Union's Horizon 2020 project MOOD (grant agreement no. 874850) and by the Research
814 Foundation - Flanders ('Fonds voor Wetenschappelijk Onderzoek - Vlaanderen',
815 G0D5117N, G005323N and G051322N). We thank Advanced Micro Devices, Inc. for
816 the donation of massively parallel computing hardware.

817 **Author Contributions**

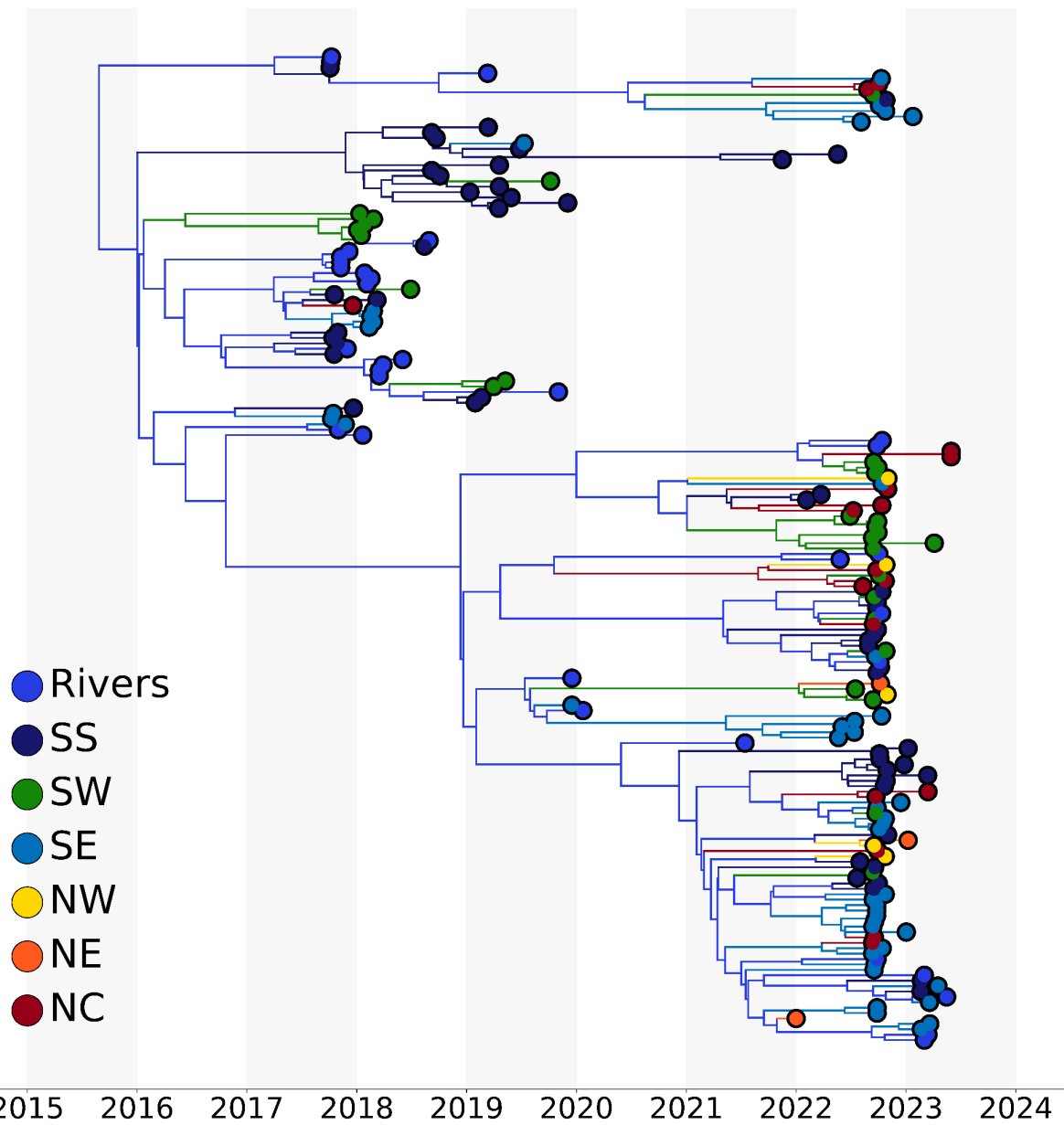
818 C.H., P.S., R.N., I.J., I.A., conceptualized the study. E.P., I.F.O., P.V., A.M, A.O.T.,
819 P.L., M.A.S., K.G.A., A.R., C.H. contributed the methodology. P.V., K.G., A.O.T.,
820 P.L., M.A.S., A.R. provided software. E.P., I.F.O., P.V., A.M. performed formal
821 analysis. E.P., I.F.O., P.V., A.M., A.O.A, A.E.S, M.I.A., O.O.E, O.A.O., A.O., P.E.,
822 O.A., A.E., O.E., C.C, K.S., A.Akinpelu , A.Ahmad, K.I.I., D.D.D., L.L.M.E, M.H.M.Y,
823 F.M.F.C, M.Z., A.R. conducted the investigation. A.O.A, A.E.S, M.I.A., O.O.E,
824 O.A.O, A.O., P.E., O.A., A.E., O.E., C.C, K.S., A.Akinpelu , A.Ahmad, K.I.I.,
825 D.D.D., L.L.M.E, M.H.M.Y, F.M.F.C, D.J.P, P.L., G.M., S.K.T., Y.K.T., O.F., A.H.,
826 M.A.S., K.G.A., A.R., R.N., C.I., I.J., I.A., provided resources. E.P., I.F.O., C.T.T,
827 J.R.O., A.R., K.S. curated data. E.P., I.F.O., P.V. wrote the original draft of the
828 manuscript. All authors reviewed and edited the manuscript. E.P., I.F.O., P.V.
829 performed visualization. C.H., R.N., C.I., I.J., I.A., P.S., A.R., K.G.A supervised the
830 study. E.P., O.F., C.H., A.E. undertook project administration. C.H., P.S., I.A., K.G.A.
831 acquired funding.

832 **Competing interest declaration**

833 MAS receives grants and contracts from the U.S. Food & Drug Administration, the U.S.
834 Department of Veterans Affairs and Johnson & Johnson all outside the scope of this work.

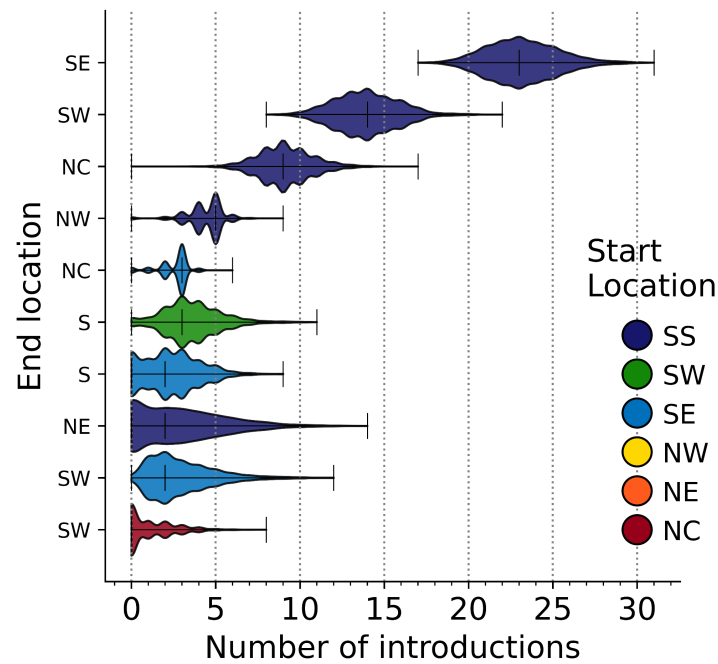
835

836 Extended Data



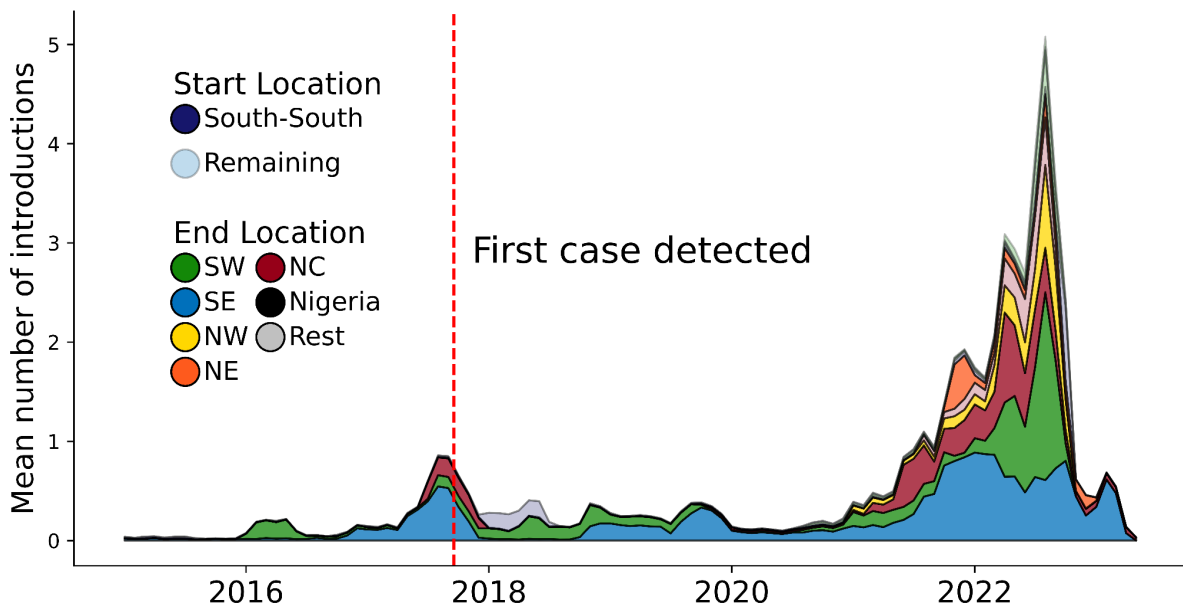
837

838 **Extended Data Figure 1:** Phylogeographic analyses of Clade IIb in Nigeria. The branches of the MCC
839 are coloured by source state, as per legend. Non-Rivers state were grouped by region. SS: South
840 South; SW: South West; SE: South East; NW: North West; NE: North East; NC: North Central.
841



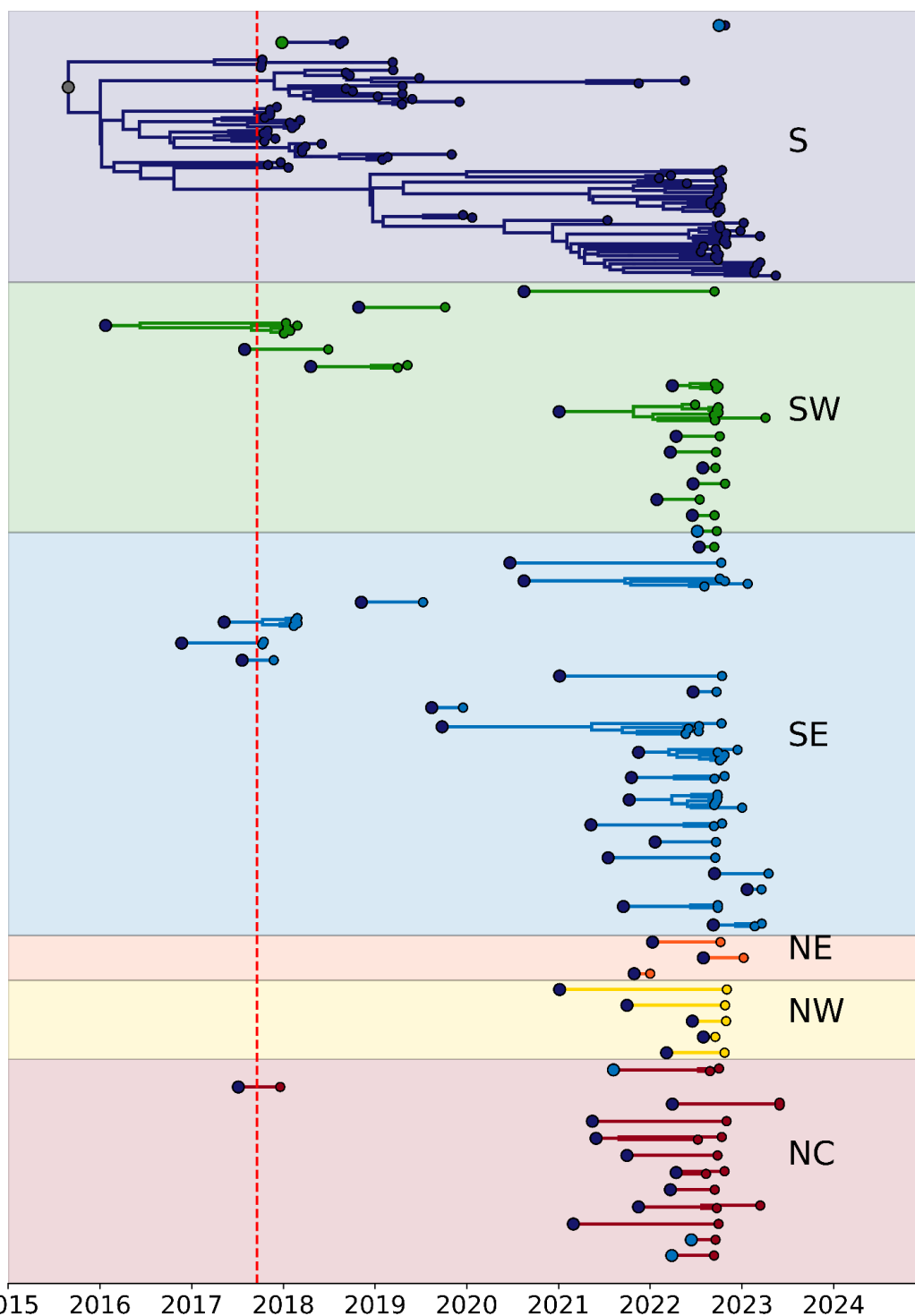
842

843 **Extended Data Figure 2:** Total number of introductions by region from each start region, annotated as
 844 per legend in colour, to each end location on the y-axis.
 845



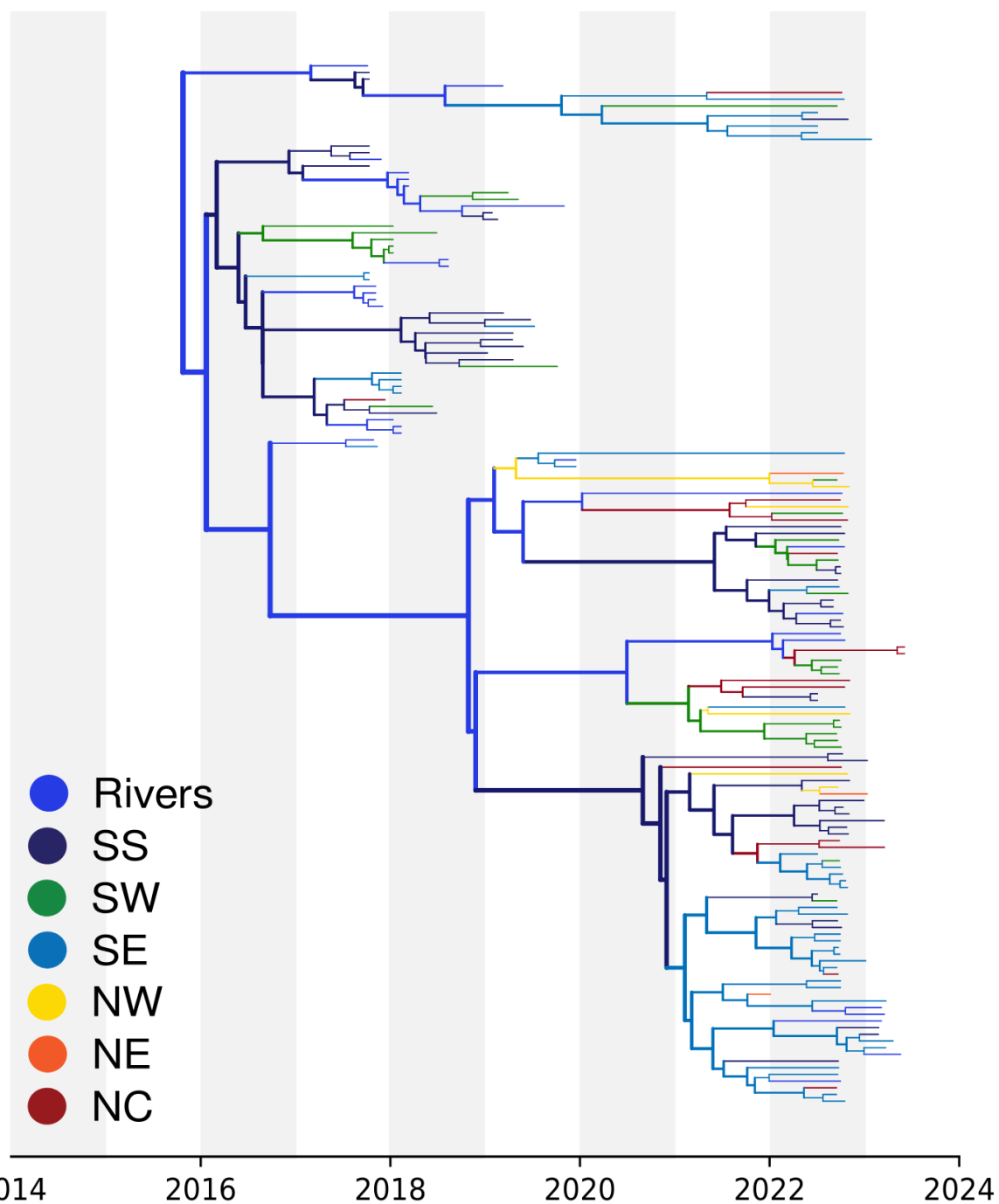
846

847 **Extended Data Figure 3:** The distribution of the number of introductions across time by region. The
 848 end location state is coloured by region, as per legend. The start location is highlighted by transparency:
 849 all introductions originating from the South-South region are presented with no transparency, whereas
 850 introductions originating from other regions are transparent.
 851



852

853 **Extended Data Figure 4:** Persistence of transmission chains across all regions, as annotated in text.
854 The start of each transmission chain is coloured by its region of origin. The red line indicates the date
855 of report for the first case in Bayelsa.



856
857
858
859
860

Extended Data Figure 5: Delphy analog to Extended Data Figure 1. Although Delphy's naive parsimony propagation of geographical labels on the MCC is particularly crude, the basic outline of the geographical spread is clearly recovered.

861

862 Extended Data Table 1: Covariates for GLM analyses

Covariates Type	Abbreviation	Covariates description	863
Administrative	Lagos(O)	Lagos as origin of infection	864
Administrative	Lagos(D)	Lagos as Destination of infection	865
Administrative	Rivers(O)	Rivers as origin of infection	866
Administrative	Rivers(D)	Rivers as Destination of infection	867
Demographic	Pop Size(O)	Origin population size, log-transformed, standardised	868
Demographic	Pop Size(D)	Destination population size, log-transformed, standardised	
Demographic	Pop Dens(O)	Origin population density, log-transformed, standardised	
Demographic	Pop Dens(D)	Destination population density, log-transformed, standardised	
Demographic	Time100k (O)	Average log-transformed travel time in minutes to the closest city with a population of over 100,000 from the origin location, standardised.	
Demographic	Time 100k (D)	Average log-transformed travel time in minutes to the closest city with a population of over 100,000 from the destination location, standardised.	
Economics	GDP	Economic output, log-transformed, standardised	
Epidemiology	Residuals(O)	Sampled genomes relative to the incidence case in Origin log-transformed, standardise	
Epidemiology	Residuals(D)	Sampled genomes relative to the incidence case in Destination log-transformed, standardise	
Epidemiology	HIV(O)	HIV cases at Origin log-transformed, standardise	
Epidemiology	HIV(D)	HIV cases at Destination log-transformed, standardise	
Epidemiology	Drug use(O)	Drug use at the origin log-transformed, standardise	
Epidemiology	Drug use(D)	Destination Drug use, log-transformed, standardise	
Geography	GCDistances	Great circle distances between the locations' population centroids, standardise log-transformed,	
Geography	RDistances	Road distance by driving, log-transformed, standardise	