

# 1 Global genomic surveillance of monkeypox virus

## 2 Authors

3 James R. Otieno<sup>1\*</sup>, Christopher Ruis<sup>1,2,3,4\*</sup>, Bernard A. Onoja<sup>1</sup>, Krutika Kuppalli<sup>1</sup>, Ana Hoxha<sup>1</sup>,  
4 Andreas Nitsche<sup>5</sup>, Annika Brinkmann<sup>5</sup>, Janine Michel<sup>5</sup>, Placide Mbala-Kisengeni<sup>6,7</sup>, Daniel  
5 Mukadi-Bamuleka<sup>7,8</sup>, Muntasir Mohammed Osman<sup>9</sup>, Hanadi Elawad Hussein<sup>9</sup>,  
6 Muhammad Ali Raja<sup>10</sup>, Richard Fotsing<sup>11</sup>, Belinda L. Herring<sup>12</sup>, Mory Keita<sup>12</sup>, Jairo Mendez  
7 Rico<sup>13</sup>, Lionel Gresh<sup>13</sup>, Amal Barakat<sup>14</sup>, Victoria Katawera<sup>15</sup>, Karen Nahapetyan<sup>16</sup>, Dhamari  
8 Naidoo<sup>17</sup>, R. Andres Floto<sup>2,3,18</sup>, Jane Cunningham<sup>1</sup>, Maria D. Van Kerkhove<sup>1</sup>, Rosamund  
9 Lewis<sup>1</sup>, Lorenzo Subissi<sup>1#</sup>

10

11 # Corresponding author

12 \*contributed equally

13 1 World Health Organization, Geneva, Switzerland

14 2 Victor Phillip Dahdaleh Heart & Lung Research Institute, University of Cambridge,  
15 Cambridge, UK

16 3 Cambridge Centre for AI in Medicine, University of Cambridge, Cambridge, UK

17 4 Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

18 5 Robert Koch Institute, Berlin, Germany

19 6 National Institute for Biomedical Research, Kinshasa, Democratic Republic of the Congo

20 7 University of Kinshasa, Kinshasa, Democratic Republic of the Congo

21 8 National Institute for Biomedical Research, Goma, Democratic Republic of the Congo

22 9 Health Emergencies and Epidemics Control, Federal Ministry of Health, Khartoum, Sudan

23 10 World Health Organization Country Office, Khartoum, Sudan

24 11 World Health Organization Country Office, Kinshasa, Democratic Republic of the Congo

25 12 World Health Organization Regional Office for Africa, Brazzaville, Republic of the Congo

26 13 World Health Organization Regional Office for the Americas, Washington, United States of  
27 America

28 14 World Health Organization Regional Office for the Eastern Mediterranean, Cairo, Egypt

29 15 World Health Organization Regional Office for the Western Pacific, Manila, the Philippines

30 16 World Health Organization Regional Office for Europe, Copenhagen, Denmark

31 17 World Health Organization Regional Office for South East Asia, New Delhi, India

32 18 Cambridge Centre for Lung Infection, Papworth Hospital, Cambridge, UK

33

## 34 Abstract

35 Monkeypox virus (MPXV) is endemic in Western and Central Africa and, in May 2022, a clade  
36 IIb lineage (B.1) caused a global outbreak outside Africa, resulting in its detection in 117  
37 countries/territories. To understand the global phylogenetics of MPXV, we carried out the first  
38 analysis of all available MPXV sequences, including 10,670 sequences from 65 countries  
39 collected between 1958 and 2024. Our analysis reveals high mobility of clade I viruses within  
40 Central Africa, sustained human-to-human transmission of clade IIb lineage A viruses within  
41 the Eastern Mediterranean region, and distinct mutational signatures that can distinguish

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

42 sustained human-to-human from animal-to-animal transmission. Moreover, distinct clade I  
43 sequences from Sudan suggest local MPXV circulation in areas of Eastern Africa over the past  
44 four decades. Our study underscores the importance of genomic surveillance in tracking  
45 spatiotemporal dynamics of MXPV clades and the need to strengthen such surveillance,  
46 including in some parts of Eastern Africa.

47

## 48 Introduction

49

50 Mpox, formerly known as monkeypox, is a disease that is caused by the *monkeypox virus*  
51 (MPXV). MPXV is a member of the *orthopoxvirus* genus, which also includes the *variola virus*,  
52 the causative agent of smallpox.<sup>1</sup> In humans, mpox can be associated with a range of clinical  
53 symptoms, but classically presents with a short febrile prodromal phase, which lasts 1-5 days,  
54 followed by the appearance of skin and/or mucosal rash, which might include single or  
55 multiple lesions.<sup>2-4</sup> The incubation period of mpox has historically ranged from 5 to 21 days.<sup>5</sup>  
56 MPXV is divided genetically into two clades — clade I (formerly known as Congo Basin clade)  
57 and clade II (formerly known as West African clade), which is further classified into subclades  
58 IIa and IIb.<sup>6</sup> Clade I and subclade IIa circulate endemically within as yet unknown animal  
59 reservoirs, potentially including rodents and non-human primates, and human cases are  
60 mostly the result of spill-over from these reservoirs.<sup>1,7,8</sup> Historical surveillance has not been  
61 sufficient to identify the frequency of spillover. In 2022, mpox epidemiology shifted with  
62 emergence of a new human-adapted lineage – clade IIb - that spread worldwide, and based  
63 on the vast number of sequences from this outbreak, it was inferred that clade IIb has  
64 circulated continually within humans since at least 2016.<sup>9</sup> Mpox human-to-human  
65 transmission primarily occurs through direct contact with infected lesions or bodily fluids,  
66 which includes sexual contact, but transmission can also occur through contact with  
67 fomites.<sup>10</sup>

68

69 In May 2022, a novel lineage of clade IIb termed B.1 emerged and spread globally, establishing  
70 efficient local transmission within many countries with no previous history of mpox  
71 transmission. As of 30 April 2024, the multi-country outbreak has been associated with 97,208  
72 cases and 186 fatalities from 117 countries/areas/territories, representing a case fatality ratio  
73 (CFR) of 0.19%.<sup>11</sup> The outbreak is primarily driven by sexual transmission among males who  
74 self-identify as men who have sex with men, with 7% of cases requiring hospitalization. Other  
75 groups at higher risk of hospitalization include female cases, those younger than 5 years of  
76 age or greater than 65 years of age, and the immunosuppressed (either due to being HIV  
77 positive or from other immunocompromising conditions).<sup>12</sup> In response to the global  
78 outbreak, a number of countries have started to establish mpox surveillance programs.

79

80 As well as the global outbreak, detection of mpox is increasing within endemic regions. For  
81 example, in 2023, a total of 14 626 mpox suspected cases and suspected 654 deaths (CFR

82 4.5%) were reported in the DRC, representing the highest figures in the recorded in the  
83 country and the highest among countries in the WHO African Region.<sup>13</sup> In 2024, a total of 7  
84 851 mpox suspected cases were reported as of 26 May, including 384 suspected deaths (CFR  
85 4.9%).<sup>13</sup> This is substantially higher than previous years. This recent increase, along with the  
86 newly documented sexual transmission recorded in March and then July-September 2023 in  
87 Kwango and South Kivu provinces, respectively, confirm the growing importance of human-  
88 to-human transmission, including through sexual contact, in the DRC.<sup>14–16</sup> Furthermore, this  
89 highlights the importance of MPXV as an emerging human pathogen.

90  
91 The global mpox surveillance that was quickly established in 2022 provided a platform for the  
92 generation of genome sequencing data. This data has been useful to characterize MPXV  
93 evolution, understand the origins of emerging lineages and monitor local and global spread.  
94 Previous studies have shown that clade IIb exhibits a higher substitution rate than other  
95 *orthopoxvirus* variants.<sup>9,17–19</sup> This appears to be due to elevated TC>TT mutations (which  
96 represents C mutating to T with an upstream T nucleotide and also includes the reverse  
97 GA>AA mutations) driven by human APOBEC-3 enzymes causing cytosine deamination in the  
98 viral genome.<sup>9,17–19</sup> This mutational signature has enabled inference that clade IIb is a human-  
99 adapted lineage.

100 Global genomic surveillance can also enable monitoring of the integrity and stability of the  
101 MPXV genomic termini, which can rearrange driving gene duplication or gene loss,<sup>20</sup> which  
102 are also drivers of poxvirus evolution and adaption to the host, and are therefore important  
103 to monitor.<sup>21,22</sup>

104 Mpox surveillance programs and their associated genomic strategies globally remain critical  
105 to understand the disease and characterize the virus evolutionary trajectory and genetic  
106 diversity, as well as potential insights into the associated phenotype. Ultimately, this data  
107 supports the deployment of suitable countermeasures (diagnostics, therapeutics, and  
108 vaccines), leveraging efforts garnered from smallpox interventions prior to eradication and  
109 preparedness in the years since, as well as advance research and development for  
110 countermeasures. Here, we report a global analysis of the publicly available MPXV genomic  
111 sequence data, which provide key insights into the spatiotemporal spread, host species range  
112 and evolution of this ongoing threat.

## 113 Results

114 We collected all available MXPV sequences and filtered these to retain 10 546 high quality  
115 sequences from 64 countries (see **Methods**). Of those, 6 585 sequences were extracted from  
116 GenBank (62%) and 3 914 sequences from GISAID (37%). Due to their distinct epidemiology  
117 and sequence diversity, we divided clade IIb into the A sublineages and the B.1 lineage for the  
118 analyses below, and refer to these groupings as clade IIb A and lineage B.1, respectively. As  
119 expected, the majority (97.7%) of the available MPXV sequences cluster within lineage B.1

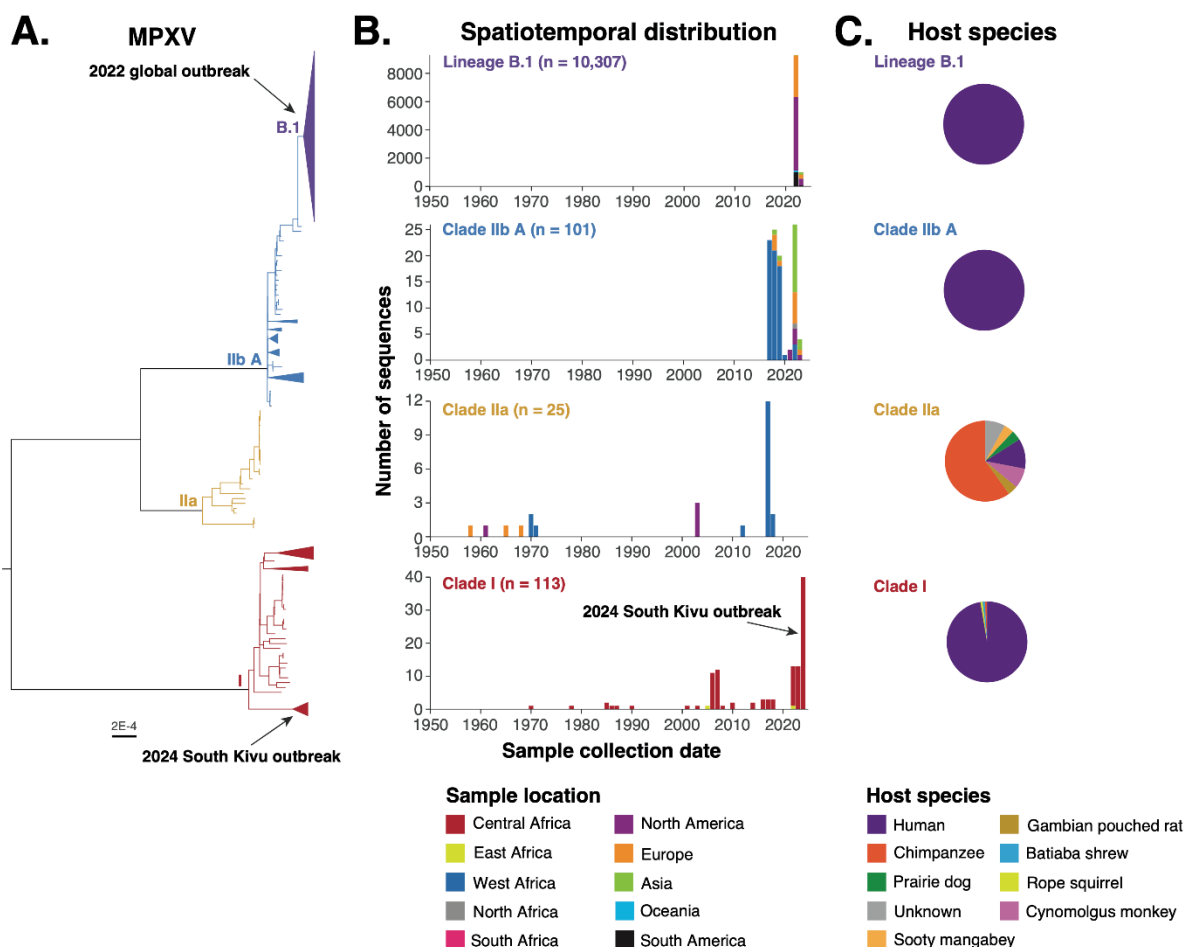
120 (Figure S1a), representing intensive genomic sequencing efforts during the global outbreak.  
 121 Clades I, IIa and IIb A have been sequenced far less often (Figure S1a). Correspondingly, the  
 122 majority (98.6%) of MPXV sequences were collected from 2022-2024, with limited historical  
 123 surveillance resulting in many years between 1958 and 2015 with no sequences in global  
 124 databases. (Figure S1b). Nonetheless, we observe differences in the temporal, spatial and host  
 125 species distributions of the major MPXV clades and discuss these below.

126

### 127 MPXV clades exhibit distinct temporal sampling distributions

128 Subclade IIa and clade I were the first to be detected, in 1958 and 1970, respectively (Figures  
 129 1, S1c).<sup>23,24</sup> Sporadic detection of both of these clades has continued through to recent years  
 130 (Figures 1, S1c), showing continued circulation within the animal reservoir. Clade I continued  
 131 to be detected in DRC and Sudan during the lineage B.1 outbreak between 2022 and 2024  
 132 (Figure 1), with DRC cases in 2024 being associated with a novel divergent lineage showing  
 133 signatures of human-to-human transmission within South Kivu.<sup>15,25</sup> Clade IIa has not been  
 134 observed since 2018 (Figure 1). Except for a sample from 1971, all clade IIb genomes were  
 135 sampled from 2017-2023. The 1971 sample likely reflects the fact that also clade IIb originated  
 136 in an animal reservoir. Clade IIb A was first detected in Nigeria in 2017 and has continued  
 137 circulating through human-to-human transmission to at least 2023,<sup>8</sup> while the descendent  
 138 lineage B.1 was first detected in 2022 (Figures 1, S1c).

139



140

141 **Figure 1. Spatiotemporal and host species distributions of MPXV sequences.** (A) Maximum  
142 likelihood phylogenetic tree highlighting the major clades of MPXV. Branches are coloured by  
143 clade. Lineage B.1 clusters within clade IIb and caused the 2022 global MPXV outbreak; this  
144 lineage is therefore separated from the remainder of clade IIb. Scale bar shows the expected  
145 number of nucleotide substitutions per site. A subset of clades are collapsed for clarity. (B)  
146 The temporal and regional distribution of MPXV sequences is shown for each clade. The n  
147 numbers show the total number of sequences from the clade. (C) Distributions of the number  
148 of sequences from each host species.

149

### 150 **MPXV sampling shows distinct host species patterns**

151 Although it likely originated from an animal reservoir, clade IIb circulates via human-to-human  
152 transmission<sup>26</sup> and correspondingly both clade IIb A and B.1 lineages have, to date, been  
153 sampled exclusively in humans (**Figure 1**). While clades I and IIa both circulate within poorly  
154 understood animal reservoirs, the hosts from which they have been sampled are markedly  
155 different. The majority of clade I sequences (96%) have been sampled from humans, with  
156 single samples from an outbreak in captive chimpanzees,<sup>27</sup> and from wild shrew (*Crocidura*  
157 *littoralis*) and rope squirrel (*Funisciurus Anerythrus*).<sup>28</sup> Conversely, only 12% (3/25) of clade IIa  
158 samples with recorded host species were collected in humans (**Figure 1**), two in 1970 and one  
159 in 2003. Clade IIa has been isolated most often in chimpanzees (60%, **Figure 1**), although these  
160 samples are mostly from a single study within Taï National Park in Cote d'Ivoire,<sup>7</sup> and  
161 chimpanzees are likely a spillover host rather than a reservoir host. Clade IIa has additionally  
162 been isolated from a wild sooty mangabey, imported cynomolgus monkeys in USA and  
163 Denmark, and a prairie dog during the 2003 USA outbreak (**Figure 1**).<sup>29</sup>

164

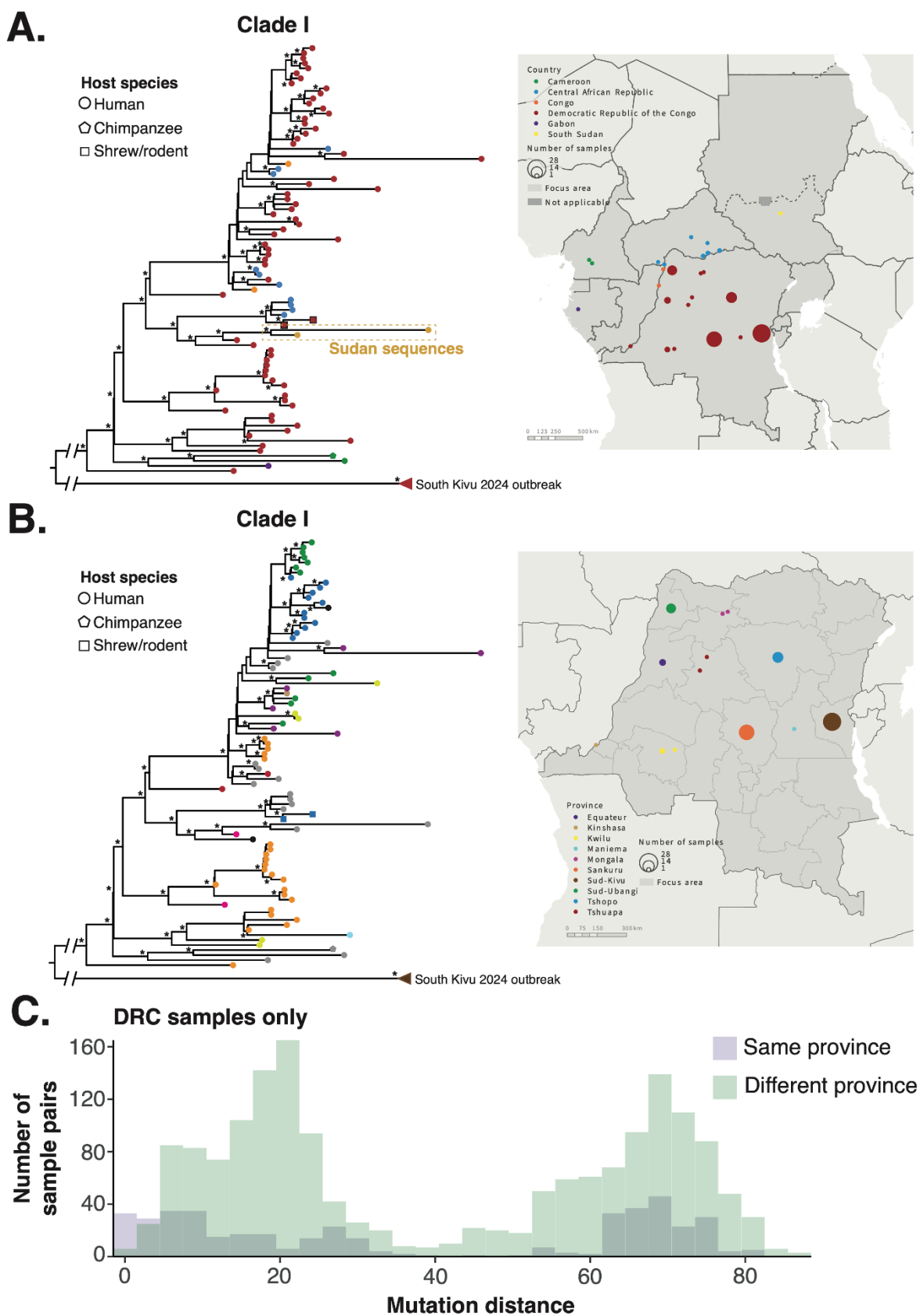
### 165 **Geographical distribution of MPXV clades**

166 Clade I has mostly been isolated from the Congo Basin area, where it has been sampled in the  
167 DRC, the Republic of the Congo, Central African Republic, Cameroon and Gabon (**Figures 2,**  
168 **S2**). Sampling locations are broadly spread around these countries. Sequences from individual  
169 countries and provinces often do not cluster within the phylogenetic tree (**Figure 2**),  
170 demonstrating multiple introductions of clade I into local geographical regions.

171

172 Having first confirmed the presence of a temporal signal (see **Methods**), we sought to identify  
173 the timescale of these virus movements by reconstructing a temporal phylogenetic tree  
174 (**Figure 3**). We found that the most recent common ancestor of clade I (excluding the 2024  
175 South Kivu outbreak which has a different substitution rate, see **Methods**)<sup>9</sup> occurred in  
176 approximately 1917 (95% highest probability density (HPD) 1880-1949). We observe frequent  
177 international and inter-province transmission over the past several decades (**Figures 2, 3**). For  
178 example, a clade sampled in Sud-Ubangi, Equateur and Kinshasa in 2023-2024 coalesces to a  
179 common ancestor in 2007 (95% HPD 2003-2011), supporting recent virus movement within  
180 the animal reservoir. Furthermore, we observe highly similar distributions of genetic  
181 relatedness between clade I samples from the same and different provinces within DRC

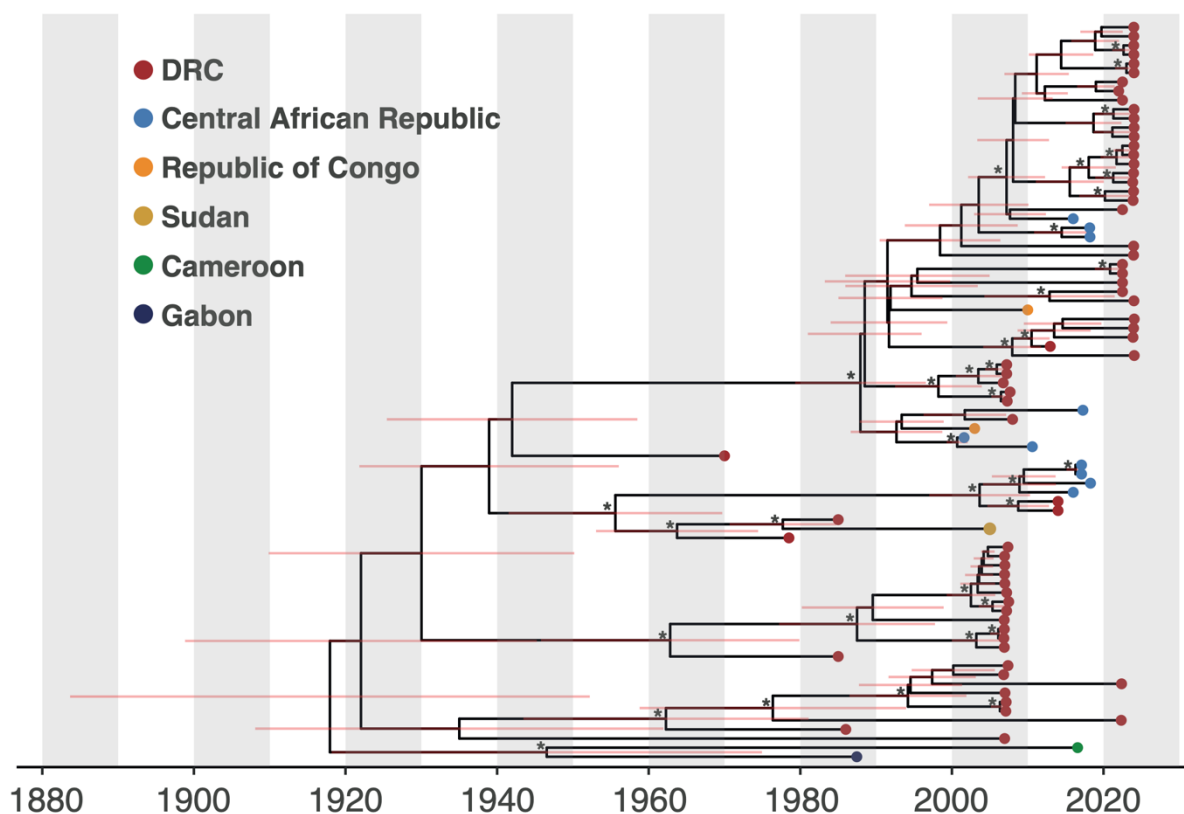
182 (Figure 2C), further highlighting the regular movement of viruses between geographical  
 183 locations.  
 184



185



186 **Figure 2. Regular international and inter-province transmission of clade I.** (A, B) Maximum  
187 likelihood phylogenetic tree of 113 high-quality clade I sequences. (A) Tips are coloured by  
188 country to match the map and shapes show the host species from which the sequence was  
189 isolated. The Sudan sequences cluster is highlighted. The 2024 South Kivu outbreak clade has  
190 been collapsed for clarity. Asterisks show phylogenetic nodes with bootstrap support of 70 or  
191 higher. The map shows sampling locations with points proportional to the number of  
192 sequences from the location. (B) Tips are coloured by province within DRC to match the map.  
193 Tips collected outside DRC are coloured grey and tips sampled within DRC but without a  
194 recorded province are coloured black. (C) Mutation distance between all pairs of clade I  
195 samples stratified by whether the pairs are from the same (*purple*) or different (*green*)  
196 provinces.  
197



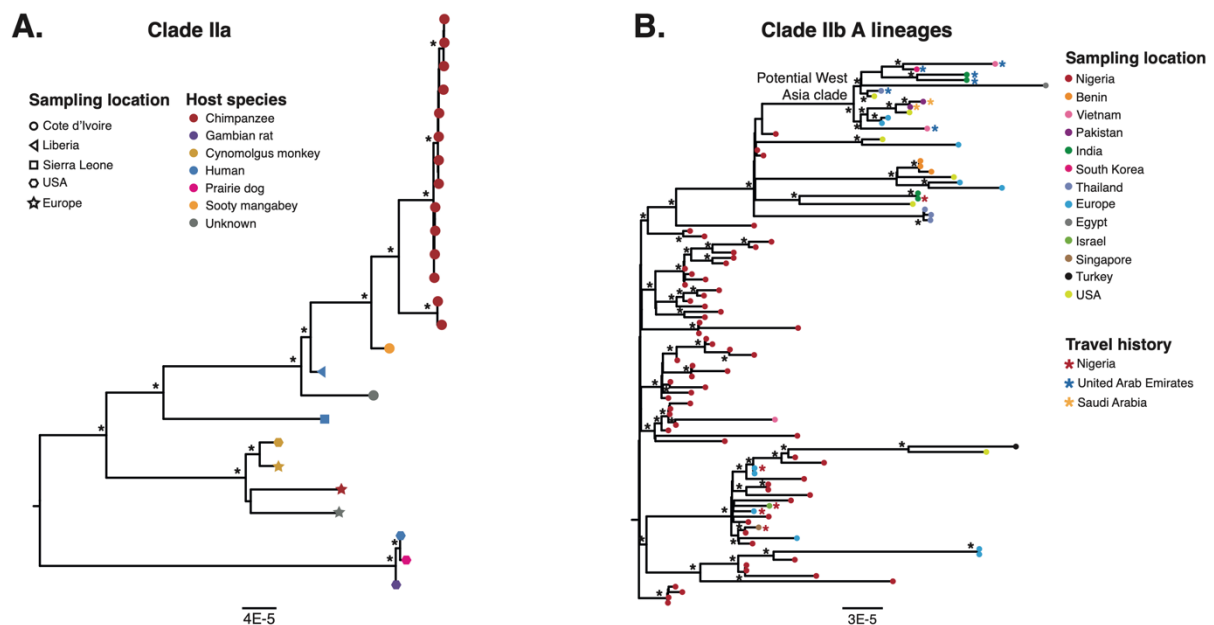
198 **Figure 3. Temporal evolutionary history of clade I.** The temporal maximum clade credibility  
199 phylogenetic tree is shown. Tips are coloured by country of isolation. Red bars show the 95%  
200 HPD on the date of the corresponding node. Asterisks show nodes with posterior support of  
201 70 or higher.  
202

203  
204 Clade I has been isolated outside of the Congo Basin in Sudan in 2005 (in a region that is now  
205 part of South Sudan) and 2022 (**Figure 2A**). The two Sudan sequences cluster in the  
206 phylogenetic tree and share a ~10.5Kb duplication (**Figure S3**). We estimate that the Sudan  
207 sequences diverged from their closest sampled relative (a DRC sample from 1985) in 1978  
208 (95% HPD 1969 – 1984); this lineage has therefore been sampled only in Sudan over roughly  
209 46 years. It is likely that this lineage has circulated in the animal reservoir during this period

210 as it exhibits 8.5% TC>TT mutations, highly similar to that expected in animals (8%) but far  
211 lower than expected from evolution in humans (85%).<sup>9</sup> This is confirmed by the geographical  
212 spread (6 states in western, southern and eastern Sudan) of the 18 mpox cases that were  
213 laboratory confirmed in Sudan 2022 (**Figure S4**).

214  
215 Clades IIa and IIb A both circulate in West Africa and have been exported to other regions  
216 (**Figures 1, 4**). Clade IIa has not been observed outside West Africa since an outbreak in USA  
217 in 2003 (**Figure 1**). Samples of clade IIa from West Africa remain sparse with single sequences  
218 from Liberia and Sierra Leone from human cases in 1970, and two closely related clusters of  
219 sequences from chimpanzees in Cote d'Ivoire collected from 2017-2018 (**Figure 4A**).<sup>7</sup> We  
220 therefore currently lack the resolution to examine spatial transmission patterns in more detail  
221 for clade IIa.

222



223  
224 **Figure 4. Spatial distributions of clades IIa and IIb A. (A)** Maximum likelihood phylogenetic  
225 tree of 25 high quality clade IIa isolates. Tips are coloured by host species and shapes show  
226 sampling locations. Asterisks show nodes with bootstrap support of 70 or higher and the scale  
227 bar shows the expected number of nucleotide substitutions per site. **(B)** Maximum likelihood  
228 phylogenetic tree of 101 clade IIb A sequences. Tips are coloured by country of collection.  
229 Red, blue and orange asterisks show samples with travel history. The potential Eastern  
230 Mediterranean clade containing samples with travel history to United Arab Emirates and Saudi  
231 Arabia is highlighted. Black asterisks so nodes with bootstrap support of 70 or above and the  
232 scale bar shows the expected number of nucleotide substitutions per site.

233  
234 While clade IIb A initially spread in Nigeria, it was then exported to other countries in Europe,  
235 Asia, North America and, more recently, North Africa (**Figures 1, 4B**). Genetic sequences from  
236 multiple individuals infected with clade IIb A from United Kingdom, Israel, Singapore and India  
237 have travel history to Nigeria (**Figure 4B**), supporting infection in endemic regions of Nigeria



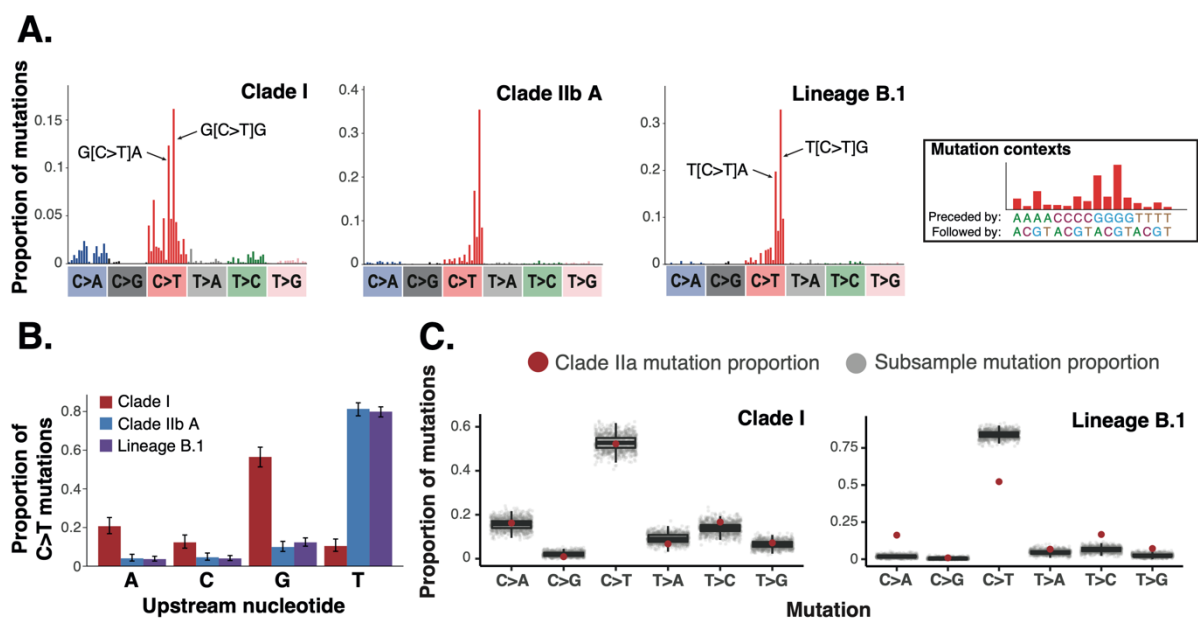
238 and subsequent export. Additionally, eight sequences from India, South Korea, Vietnam and  
239 Thailand were isolated from travellers returning from United Arab Emirates or Saudi Arabia;  
240 these sequences cluster within a single phylogenetic lineage that also includes sequences  
241 from USA, UK, Slovenia and Egypt for which no information regarding recent travel is recorded  
242 (Figure 4B). This is consistent with sustained circulation of a lineage of clade IIb A in the  
243 Eastern Mediterranean region; no sequences are currently available from Eastern  
244 Mediterranean to examine this further.

245

## 246 MPXV mutational spectra are influenced by transmission route

247 Previous studies have demonstrated that mutational spectra of MPXV lineages that transmit  
248 from human-to-human exhibit a high proportion of TC>TT mutations, implicating APOBEC-3  
249 as a major driver of mutagenesis in human MPXV infections.<sup>9</sup> We, for the first time, calculated  
250 complete single base substitution (SBS) mutational spectra for the dominant clades of MPXV  
251 and corrected these for genomic composition (Figure 5, Methods). Similar to previous  
252 studies,<sup>9,19</sup> we find the spectrum of clade IIb (both A and B.1 lineages) to be dominated by C>T  
253 mutations where the C is preceded by a T, with preference for A or G following the substitution  
254 site (Figure 5).

255



256

257 **Figure 5. Mutational spectra differ between major MPXV clades.** (A) Single base substitution  
258 (SBS) mutational spectra for clade I, clade IIb A and lineage B.1 (clade IIa not included because  
259 of insufficient mutations, see Methods). SBS spectra show the proportion of mutations of  
260 each mutation type within each surrounding nucleotide contexts; contexts for an example  
261 mutation type are shown in the right-hand panel. The two most prevalent contextual  
262 mutations are highlighted for clade I and lineage B.1. Mutational spectra are corrected for  
263 genome composition (see Methods). Symmetrical mutations (for example C>T and G>A) are  
264 combined as MPXV is a DNA pathogen.<sup>30</sup> (B) The proportion of C>T mutations with each  
265 nucleotide upstream is shown for each clade. (C) To examine the potential for the clade IIa

266 mutational spectrum to have been generated by the mutational spectra of clade I and lineage  
267 B.1, we compared the proportion of each mutation type in the clade IIa spectrum with that in  
268 1000 subsamplings of the other clade spectrum to the number of mutations in the clade IIa  
269 spectrum (see **Methods**). Each grey point represents the mutation type proportion in one  
270 subsample of the respective mutational spectrum while the red point shows the mutation  
271 type proportion in clade IIa. The clade IIa mutation type proportions are within that expected  
272 from clade I but often outside that expected from lineage B.1. Boxplot centre lines show  
273 median value; upper and lower bounds show the 25th and 75th quantile, respectively; upper  
274 and lower whiskers show the largest and smallest values within 1.5 times the interquartile  
275 range above the 75th percentile and below the 25th percentile, respectively.

276  
277 The clade I mutational spectrum shows that mutational processes within the animal reservoir  
278 are also dominated by C>T mutations (66% of mutational burden accounting for genome  
279 composition), but with different contextual preferences to evolution within humans (**Figure**  
280 **5**). In clade I, C>T mutations occur most commonly where the C is preceded by G (57% of the  
281 C>T mutational burden, **Figure 5B**) and are most frequent in G[C>T]A and G[C>T]G contexts  
282 (**Figure 5A**). The C>T mutations are unlikely to be the result of spontaneous deamination of  
283 cytosine as this has a strong preference for CG>TG contexts in human DNA.<sup>31</sup> The C>T  
284 mutations may therefore be the result of polymerase errors during genome replication, action  
285 of alternative APOBEC enzymes within the animal reservoir and/or additional mutagens. The  
286 clade IIb spectra exhibit similar enrichment of G over A and C as the nucleotide preceding C>T  
287 mutations (**Figure 5B**), which may suggest the GC>GT mutations are driven by a non-host  
288 species factor, but this will require additional future work to untangle.

289  
290 C>A (which includes G>T mutations) is the second most common mutation in clade I (17% of  
291 the total mutational burden, **Figure 5**); reactive oxygen species are a potential driver of these  
292 mutations.<sup>32</sup>

293  
294 The clade IIa mutation spectrum contains 222 mutations, which may be too low to examine  
295 mutational patterns in detail.<sup>30</sup> We therefore used the clade I and lineage B.1 mutational  
296 spectra as references and investigated whether either could explain the clade IIa mutational  
297 spectrum. We found that the clade IIa mutational patterns could have been generated by the  
298 clade I spectrum but not by the lineage B.1 spectrum (**Figure 5C**). This suggests that clade IIa  
299 exhibits similar mutational processes as clade I, consistent with a lack of human APOBEC-3  
300 activity and therefore both clades consistently circulating outside of humans.

301

## 302 Discussion

303 Implementing measures to control MPXV cases will require in depth understanding of how  
304 the virus is transmitting within humans and animals across different spatial scales. We here  
305 carried out an in-depth analysis of spatiotemporal and host species patterns across all major

306 MPXV clades. Our analysis revealed large differences in the spatiotemporal and host species  
307 patterns of the major clades. We identified regular international and inter-province  
308 transmission of clade I, including its likely circulation in animal reservoirs and/or endemicity  
309 in parts of eastern Africa, inferred transmission of clade IIb A in Eastern Mediterranean and  
310 showed that MPXV mutational patterns are associated with transmission route.

311  
312 Our analysis provides the first evidence of regular recent transmission of clade I amongst  
313 countries and provinces within Central Africa (**Figures 2, 3**). We also found co-circulation of  
314 multiple clade I lineages within individual countries and provinces (**Figures 2, 3**). This  
315 demonstrates that high MPXV diversity is maintained within the clade I animal reservoir(s),  
316 suggesting high MPXV prevalence and thereby highlighting the potential for frequent  
317 spillover when humans interact with reservoir species. Our results also show that the animal  
318 reservoir is mobile. Understanding the speed and dynamics of virus movements may help to  
319 pinpoint possible reservoir species, as well as source locations, but this will require  
320 increased sequencing of epidemiologically representative samples in future.

321  
322 While clades I and IIa both circulate within animals and spillover into humans, we observe  
323 highly different distributions of host species amongst sampled genetic sequences, with clade  
324 I mostly being sampled from humans and clade IIa from animals (**Figure 1C**). It is unclear  
325 whether this difference in host sampling is driven by distinct abilities of the clades to infect  
326 humans, differential disease severity in humans and/or animals altering likelihood of case  
327 detection, different contacts between humans/sampled animals and animal reservoirs  
328 (which may differ for the two clades) and/or different surveillance and sampling strategies in  
329 humans and animals within the affected countries. Identifying the driver(s) of differential  
330 host sampling will require stronger surveillance in humans and animals, and linkage with  
331 sample metadata to determine likely routes of infection.

332  
333 While clade I has mostly been detected within the Congo Basin, our data highlight the  
334 potential for widespread endemic and/or enzootic circulation of this clade within Sudan  
335 (former Sudan, now divided into Sudan and South Sudan). We identified a clade I lineage  
336 that has been sampled only from Sudan over roughly 45 years (**Figures 2, 3**). This lineage has  
337 likely circulated within the animal reservoir (due to a lack of APOBEC3-like mutations) and  
338 may have circulated continually in Sudan following introduction at any point during this time  
339 period or could have been introduced multiple times from an unsampled region either  
340 shortly before being sampled in Sudan or with some level of local transmission.

341 Distinguishing between these possibilities will require additional sequencing data from  
342 Sudan. However, our results combined with recent epidemiological data highlight the  
343 potential for large numbers of MPXV cases in Sudan: following the declaration of the Public  
344 Health Emergency of International Concern in July 2022, the Federal Ministry of Health of  
345 Sudan, with support from WHO, started mpox surveillance and as a consequence, suspected  
346 mpox cases were reported from 12 states and 32 localities in Sudan, of which 5 states host

347 refugees and internally displaced persons, and with >40% of suspected cases being children  
348 under five.<sup>33</sup> Of the suspected cases, only a portion was tested and a total of 18 cases were  
349 laboratory confirmed from 6 states and 9 localities, and including one death (CFR 5.8%).<sup>33</sup>  
350

351 Our data combined with the recent emergence of a novel human transmissible clade I  
352 lineage in South Kivu,<sup>25</sup> where mpox cases have not previously been detected (except few  
353 cases in 2011/2012), highlights the potential for an increased risk of international spread of  
354 clade I MPXV. A recent study highlighted that the affected population in South Kivu had a  
355 median age of 22 years, 52% were females and 30% were sex workers, which represent a  
356 significant shift in the historical mpox epidemiology in the DRC, which involves children <15  
357 years as the main affected age group.<sup>13,25</sup> As a comparison, in the years preceding the  
358 eradication of smallpox (1956-1971), the maximum number of smallpox cases reported by  
359 the DRC (at that time Zaire) Ministry of Health to the WHO was 5,523 cases including 710  
360 deaths in 1963,<sup>34</sup> which is a third of the suspected mpox cases reported by the DRC in  
361 2023.<sup>13</sup> Both Sudan and DRC have reported cases in provinces that directly border with other  
362 African countries such as Burundi, Uganda, Rwanda, Angola, South Sudan, and Chad, and in  
363 some instances, such as for South Kivu province (DRC), at-risk populations are known to  
364 include regular cross-border commuters. Further studies to better understand transmission  
365 patterns in these settings (including if enzooticity has been established) and the  
366 geographical distribution of mpox in Eastern Africa are urgently needed.

367  
368 The co-circulation of multiple clade I lineages in individual provinces combined with regular  
369 geographical movements and the potential for an East Africa clade I lineage suggests there is  
370 high prevalence and diversity of clade I within the animal reservoir(s). It is currently unclear  
371 whether this genotypic diversity is associated with phenotypic diversity. However, this high  
372 prevalence is likely to make control of clade I within the animal reservoir highly challenging.  
373 Prevention of human cases is therefore likely to require interventions at the human-animal  
374 interface and rapid detection and cessation of human-to-human transmission chains. Our  
375 results therefore underpin the importance of further studies to understand how humans  
376 become infected with clade I viruses and studies carrying out functional characterisation of  
377 diverse clade I viruses.

378  
379 Local human-to-human transmission chains were established in many countries across all six  
380 WHO regions during the 2022 MPXV global outbreak.<sup>11</sup> Local transmission of the ancestral  
381 clade IIb A lineages has also been identified in some cases outside of West Africa.<sup>35</sup> We were  
382 here able to infer local transmission of clade IIb A MPXV in Eastern Mediterranean through  
383 travel data associated with sequences from other countries, despite sequences from that  
384 region being unavailable for analysis. This highlights the importance of associating detailed  
385 metadata with genetic sequences where possible.  
386

387 Mutational signatures have provided major insights into MPXV and can be used to identify  
388 lineages that are transmitting from human-to-human and to infer outbreak origin dates.<sup>9,19</sup>  
389 We here calculated the most in depth mutational spectra of MPXV to date and compared  
390 these across spectra, showing that C>T mutations are most common within both human and  
391 animal hosts, but differences in contextual preferences exist between species (**Figure 5**).  
392 Identifying the drivers of spectrum differences will require future work, but our analyses  
393 suggest that the ratio between GC>GT mutations and TC>TT mutations is a reliable marker  
394 to distinguish human-to-human transmission from transmission in animals, therefore  
395 potentially being able identify sustained human outbreaks from transmission from the  
396 reservoir.

397  
398 The global outbreak has provided critical insights into the epidemiology of mpox in humans.  
399 However, it is unknown whether characteristics of lineage B.1 were acquired following  
400 adaptation in humans, or whether they may be generalisable across MPXV clades. This can  
401 only be revealed with stronger surveillance, sharing of sequences and continued genotypic  
402 and functional characterization of the differences between clades and lineages.<sup>36</sup> Such  
403 characterization would shed light into the drivers of epidemiological differences between  
404 clades, but this work is technically challenging and currently few laboratories worldwide have  
405 such capacity. However, a small number of studies have found that the apparent difference  
406 in morbidity and mortality between clades I and II is likely driven by multiple proteins present  
407 in clade I but absent in clade II.<sup>37,38</sup> One particular area of interest is gene duplication, such as  
408 that found in clade I sequences from Sudan. Such duplications have been described for clade  
409 IIb sequences from the 2022 outbreak, when they were assumed to be involved in immune  
410 evasion and host range.<sup>21,39,40</sup> Gene duplication and loss in the MPXV terminal regions are also  
411 considered drivers of poxvirus evolution and adaption to the host.<sup>21,22</sup> Analyses comparing  
412 strains with and without deletions will be essential to uncover their functional consequences  
413 and understand/forecast the epidemiology of lineages exhibiting such changes. In addition,  
414 deletions may lead to diagnostic failure, especially for nucleic acid amplification tests that  
415 target less conserved genes, such as some of the clade-specific PCRs, which are designed to  
416 distinguish clades. To date, there have been two reported examples of such diagnostic failure  
417 episodes, one for a variant of MPXV clade IIb detected in the US that did not spread widely,  
418 <sup>41</sup> and one for the clade I lineage currently circulating in South Kivu.<sup>15</sup> This clearly highlights  
419 the critical importance of a strategic genomic surveillance system where mpox circulates.

420 As highlighted in the standing recommendations for mpox issued by the Director General of  
421 the WHO, it is critical that countries have national mpox strategic plans integrated into broader  
422 health systems, and that capacities that have been built in resource-limited settings and  
423 among marginalized groups should be sustained.<sup>42</sup> Without surveillance, no genomic  
424 sequence data can be generated, and no virological characterization of circulating clades and  
425 lineages can be done. In this regard, more laboratories should engage in virological  
426 characterization of MPXV clades and lineages. Furthermore, countries are strongly

427 encouraged to continue documenting and making sequences publicly available, prioritising  
428 specimens for both targeted sequencing (for example of imported cases, the first few cases of  
429 local emergence and cases with divergent demographic or clinical profiles) and representative  
430 sequencing. This will enable tracking of virus circulation and evolution over time. If we want  
431 to prevent the next mpox global outbreak, it is now time to strengthen mpox surveillance,  
432 including in Africa, focusing on populations at highest risk and ensuring integration with  
433 existing systems to ensure comprehensive and seamless delivery of care.  
434

### 435 [Competing interests](#)

436 No competing interests declared

### 437 [Data availability statement](#)

438 All sequence accessions and inferred phylogenetic trees are available to be publicly shared  
439 without any restrictions.

### 440 [Disclaimer](#)

441 The findings and conclusions in this report are those of the author(s) and do not necessarily  
442 represent the views of the funding agencies

### 443 [References](#)

444

445 1. Bunge, E. M. *et al.* The changing epidemiology of human monkeypox-A potential threat?

446 A systematic review. *PLoS Negl Trop Dis* **16**, e0010141 (2022).

447 2. Damon, I. K. Smallpox, Monkeypox, and Other Poxvirus Infections. in *Goldman's Cecil*

448 *Medicine* 2117–2121 (Elsevier, 2012). doi:10.1016/B978-1-4377-1604-7.00380-8.

449 3. Damon, I. K. Status of human monkeypox: clinical disease, epidemiology and research.

450 *Vaccine* **29 Suppl 4**, D54-59 (2011).

451 4. Adler, H. *et al.* Clinical features and management of human monkeypox: a retrospective

452 observational study in the UK. *Lancet Infect Dis* **22**, 1153–1162 (2022).

453 5. World Health Organization. Mpox (monkeypox). (2023) doi:[https://www.who.int/news-](https://www.who.int/news-room/fact-sheets/detail/monkeypox)

454 [room/fact-sheets/detail/monkeypox](https://www.who.int/news-room/fact-sheets/detail/monkeypox).



- 455 6. Ulaeto, D. *et al.* New nomenclature for mpox (monkeypox) and monkeypox virus clades.  
456 *The Lancet Infectious Diseases* **23**, 273–275 (2023).
- 457 7. Patrono, L. V. *et al.* Monkeypox virus emergence in wild chimpanzees reveals distinct  
458 clinical outcomes and viral diversity. *Nat Microbiol* **5**, 955–965 (2020).
- 459 8. Yinka-Ogunleye, A. *et al.* Outbreak of human monkeypox in Nigeria in 2017-18: a clinical  
460 and epidemiological report. *Lancet Infect Dis* **19**, 872–879 (2019).
- 461 9. O’Toole, Á. *et al.* APOBEC3 deaminase editing in mpox virus as evidence for sustained  
462 human transmission since at least 2016. *Science* **382**, 595–600 (2023).
- 463 10. Gessain, A., Nakoune, E. & Yazdanpanah, Y. Monkeypox. *N Engl J Med* **387**, 1783–1793  
464 (2022).
- 465 11. World Health Organisation. Multi-country outbreak of mpox, External situation report  
466 #33. (2024).
- 467 12. Laurenson-Schafer, H. *et al.* Description of the first global outbreak of mpox: an analysis  
468 of global surveillance data. *Lancet Glob Health* **11**, e1012–e1023 (2023).
- 469 13. World Health Organization. Mpox - Democratic Republic of the Congo, Disease Outbreak  
470 News. (2024).
- 471 14. Kibungu, E. M. *et al.* Clade I–Associated Mpox Cases Associated with Sexual Contact, the  
472 Democratic Republic of the Congo. *Emerg. Infect. Dis.* **30**, (2024).
- 473 15. Masirika, L. M. *et al.* Ongoing mpox outbreak in Kamituga, South Kivu province,  
474 associated with monkeypox virus of a novel Clade I sub-lineage, Democratic Republic of  
475 the Congo, 2024. *Eurosurveillance* **29**, (2024).
- 476 16. Vakaniaki, E. H. *et al.* Sustained Human Outbreak of a New MPXV Clade I Lineage in the  
477 Eastern Democratic Republic of the Congo. *Nat Med* (2024) doi:10.1038/s41591-024-  
478 03130-3.

- 479 17. Isidro, J. *et al.* Phylogenomic characterization and signs of microevolution in the 2022  
480 multi-country outbreak of monkeypox virus. *Nat Med* (2022) doi:10.1038/s41591-022-  
481 01907-y.
- 482 18. Gigante, C. M. *et al.* Multiple lineages of monkeypox virus detected in the United States,  
483 2021–2022. *Science* **378**, 560–565 (2022).
- 484 19. Ndodo, N. *et al.* Distinct monkeypox virus lineages co-circulating in humans before 2022.  
485 *Nat Med* **29**, 2317–2324 (2023).
- 486 20. Brinkmann, A. *et al.* Extensive ITR expansion of the 2022 Mpox virus genome through  
487 gene duplication and gene loss. *Virus Genes* **59**, 532–540 (2023).
- 488 21. Esteban, D. J. & Hutchinson, A. P. Genes in the terminal regions of orthopoxvirus  
489 genomes experience adaptive molecular evolution. *BMC Genomics* **12**, 261 (2011).
- 490 22. Hughes, A. L. & Friedman, R. Poxvirus genome evolution by gene gain and loss.  
491 *Molecular Phylogenetics and Evolution* **35**, 186–195 (2005).
- 492 23. Magnus, P. von, Andersen, E. K., Petersen, K. B. & Birch-Andersen, A. A POX-LIKE DISEASE  
493 IN CYNOMOLGUS MONKEYS. *Acta Pathologica Microbiologica Scandinavica* **46**, 156–176  
494 (1959).
- 495 24. Ladnyj, I. D., Ziegler, P. & Kima, E. A human infection caused by monkeypox virus in  
496 Basankusu Territory, Democratic Republic of the Congo. *Bull World Health Organ* **46**,  
497 593–597 (1972).
- 498 25. Vakaniaki, E. H. *et al.* Sustained Human Outbreak of a New MPXV Clade I Lineage in  
499 Eastern Democratic Republic of the Congo. Preprint at  
500 <https://doi.org/10.1101/2024.04.12.24305195> (2024).
- 501 26. Kraemer, M. U. G. *et al.* Tracking the 2022 monkeypox outbreak with epidemiological  
502 data in real-time. *The Lancet Infectious Diseases* **22**, 941–942 (2022).

- 503 27. Brien, S. C. *et al.* Clinical Manifestations of an Outbreak of Monkeypox Virus in Captive  
504 Chimpanzees in Cameroon, 2016. *J Infect Dis* **229**, S275–S284 (2024).
- 505 28. Falendysz, E. A. *et al.* Characterization of Monkeypox virus infection in African rope  
506 squirrels (*Funisciurus* sp.). *PLoS Negl Trop Dis* **11**, e0005809 (2017).
- 507 29. Centers for Disease Control and Prevention (CDC). Update: multistate outbreak of  
508 monkeypox--Illinois, Indiana, Kansas, Missouri, Ohio, and Wisconsin, 2003. *MMWR Morb*  
509 *Mortal Wkly Rep* **52**, 616–618 (2003).
- 510 30. Ruis, C., Tonkin-Hill, G., Floto, R. A. & Parkhill, J. Calculating and applying pathogen  
511 mutational spectra using MutTui. Preprint at  
512 <https://doi.org/10.1101/2023.06.15.545111> (2023).
- 513 31. Nik-Zainal, S. *et al.* Mutational Processes Molding the Genomes of 21 Breast Cancers.  
514 *Cell* **149**, 979–993 (2012).
- 515 32. Zou, X. *et al.* A systematic CRISPR screen defines mutational mechanisms underpinning  
516 signatures caused by replication errors and endogenous DNA damage. *Nat Cancer* **2**,  
517 643–657 (2021).
- 518 33. World Health Organization. Multi-country outbreak of monkeypox, External situation  
519 report #8. (2022).
- 520 34. Fenner, F., Henderson, D. A., Arita, I., Jezek, Z. & Ladnyi, I. D. *Smallpox and Its*  
521 *Eradication*. (World Health Organization, 1988).
- 522 35. Shete, A. M. *et al.* Genome characterization of monkeypox cases detected in India:  
523 Identification of three sub clusters among A.2 lineage. *Journal of Infection* **86**, 66–117  
524 (2023).

- 525 36. Ulaeto, D. O., Dunning, J. & Carroll, M. W. Evolutionary implications of human  
526 transmission of monkeypox: the importance of sequencing multiple lesions. *The Lancet*  
527 *Microbe* **3**, e639–e640 (2022).
- 528 37. Hudson, P. N. *et al.* Elucidating the Role of the Complement Control Protein in  
529 Monkeypox Pathogenicity. *PLoS ONE* **7**, e35086 (2012).
- 530 38. Hutson, C. L. *et al.* Comparison of Monkeypox Virus Clade Kinetics and Pathology within  
531 the Prairie Dog Animal Model Using a Serial Sacrifice Study Design. *Biomed Res Int* **2015**,  
532 965710 (2015).
- 533 39. Moss, B. & Shisler, J. L. Immunology 101 at poxvirus U: Immune evasion genes. *Seminars*  
534 *in Immunology* **13**, 59–66 (2001).
- 535 40. Alcamí, A. & Koszinowski, U. H. Viral mechanisms of immune evasion. *Trends in*  
536 *Microbiology* **8**, 410–418 (2000).
- 537 41. Garrigues, J. M. *et al.* Identification of Human Monkeypox Virus Genome Deletions That  
538 Impact Diagnostic Assays. *J Clin Microbiol* **60**, e0165522 (2022).
- 539 42. World Health Organization. Standing recommendations for mpox issued by the Director-  
540 General of the World Health Organization (WHO) in accordance with the International  
541 Health Regulations (2005) (IHR). (2023).
- 542 43. Aksamentov, I., Roemer, C., Hodcroft, E. & Neher, R. Nextclade: clade assignment,  
543 mutation calling and quality control for viral genomes. *JOSS* **6**, 3773 (2021).
- 544 44. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic  
545 Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530–1534 (2020).
- 546 45. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure  
547 of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* **2**,  
548 vew007 (2016).

549 46. Guangchuang Yu [Aut, C. ggtree. Bioconductor

550 <https://doi.org/10.18129/B9.BIOC.GGTREE> (2017).

551 47. Xu, S. *et al.* *Ggtree* : A serialized data object for visualization of a phylogenetic tree and

552 annotation data. *iMeta* **1**, e56 (2022).

## 553 Online Methods

### 554 Dataset assembly and filtering

555 We aimed to collate a dataset containing all available high quality MPXV genetic sequences.  
556 To do this, we initially downloaded all MPXV nucleotide sequences from GenBank (identified  
557 as sequences containing at least one of the search terms “monkeypox”, “mpox”, “MPXV” and  
558 “MPV”,  $n=7,252$ ) and the Global Initiative for Sharing of All Influenza Data (GISAID) EpiPox  
559 database ( $n=8,843$ ) as of 17 February 2024. We then combined these datasets and filtered to  
560 remove duplicate and lower quality sequences. To do this, we initially removed sequences  
561 containing fewer than 30,000 nucleotides (nt), with this cutoff chosen to retain historical  
562 MPXV sequences that were  $\sim 32,000$  nt in length. We next discarded sequences with  $>20\%$   
563 indeterminate bases (Ns) and assigned clades and (for clade IIb) Pango lineages using  
564 Nextclade.<sup>43</sup> Sequences that could not be assigned a clade were assumed to be low quality  
565 and were excluded from further analysis. Based on the clade and lineage assignments from  
566 Nextclade, we divided the sequence dataset into four groups: clade I, clade IIa, clade IIb A  
567 (containing sequences from the A sublineages within clade IIb but not those within lineage  
568 B.1 and its descendent lineages) and lineage B.1 (containing sequences from lineage B.1 and  
569 its descendent lineages).

570  
571 We identified sequences that were duplicated between GenBank and GISAID initially by  
572 identifying sequences with the same sample name, country, and collection date. After  
573 removing one of each of these sequence pairs, we carried out an additional phylogenetic  
574 screen for duplicate sequences in the clade I, clade IIa and clade IIb A datasets. Sequences  
575 within each of the four datasets were aligned using squirrel v0.1  
576 (<https://github.com/aineniambh/squirrel>),<sup>9</sup> specifying the clade to which the sequences  
577 belong. We then reconstructed a phylogenetic tree for each clade using IQ-TREE v2.1.3,  
578 <sup>44</sup>employing a JC model of nucleotide substitution. We identified closely related pairs of  
579 sequences in the resulting phylogenetic trees and checked their sequence names, countries  
580 and collection dates to determine whether they might be duplicates. Where the sequence  
581 names and collection dates were similar (i.e. the sequence names contained shared elements  
582 and the collection dates were the same to the most accurate level possible), we retained one  
583 of the sequences.

584  
585 We carried out a further quality control check by analysing the root-to-tip distances of  
586 sequences compared to their collection date.<sup>45</sup> For clade I, clade IIa and clade IIb A, we aligned  
587 sequences with squirrel v0.1 and reconstructed maximum likelihood phylogenetic trees using  
588 IQ-TREE v2.1.3 as above but including an outlier sequence (outlier accession numbers  
589 KJ642617.1 for clade I, KJ642616.1 for clade IIa and clade IIb A) that was used to root the tree.  
590 We then identified sequences that were clear outliers in a root-to-tip plot of the rooted tree

591 in TempEst v1.5.3,<sup>45</sup> and removed these from further analyses. Due to the large size of the  
592 lineage B.1 dataset (n = 10369 sequences prior to filtering), instead of reconstructing a  
593 phylogenetic tree, we carried out a root-to-tip-like analysis by comparing collection date with  
594 mutation distance to sample ON676708.1, which clusters immediately upstream of lineage  
595 B.1. We aligned the lineage B.1 dataset and ON676708.1 using squirrel v0.1 as above and  
596 calculated the number of mutations between each sequence and ON676708.1. This showed  
597 a strong correlation (**Figure S5**) so we generated a linear model between sample collection  
598 date and this mutation distance and removed samples whose residual within the linear model  
599 is more than five times the median absolute deviation away from the median residual (**Figure**  
600 **S5**). These filtering steps resulted in final datasets of 113 clade I sequences, 25 clade IIa  
601 sequences, 101 clade IIb A sequences and 10,307 lineage B.1 sequences. The majority of these  
602 sequences contain close to the complete genome (**Figure S6, Tables S1, S2**). In addition, we  
603 have added 47 sequences from a GitHub directory of a recent paper that describes highly  
604 divergent viruses.<sup>16</sup>

605

606 We identified the spatiotemporal and host species distributions of sequences within each  
607 dataset using location, collection data and host species metadata associated with the  
608 sequence accession on either GenBank or GISAID. Where this metadata was missing, we  
609 attempted to identify it within original publications.

610

#### 611 [Reconstruction of phylogenetic trees](#)

612 To visualise the phylogenetic relationships between the clades, we reconstructed a tree  
613 containing all sequences from clades I, IIa and IIb A and the sublineage references for each of  
614 the sublineages within B.1 (accession numbers obtained from [https://github.com/mpxv-  
615 lineages/lineage-designation/blob/master/auto-generated/lineages.md](https://github.com/mpxv-lineages/lineage-designation/blob/master/auto-generated/lineages.md)). We aligned these  
616 sequences using squirrel v0.1 with clade set to clade II and reconstructed a maximum  
617 likelihood phylogenetic tree as above.

618

619 To examine phylogenetic relationships within each clade, we aligned sequences within the  
620 respective dataset using squirrel v0.1 with clade set to clade I for the clade I dataset and set  
621 to clade II for the remaining datasets. Final maximum likelihood phylogenetic trees were  
622 reconstructed for each dataset using IQ-TREE v2.1.3 as above. Topological robustness was  
623 assessed using 1000 bootstrap replicates. Travel histories for Clade IIb A were identified in  
624 GISAID metadata and from examination of original publications.

625

626 Phylogenetic trees were visualised using [FigTree v1.4.4](#) and [ggtree v3.0.2](#).<sup>46</sup>

627

#### 628 [Reconstruction of the temporal history of Clade I](#)

629 We aimed to reconstruct the temporal history of clade I. The recently identified outbreak in  
630 South Kivu contains evidence of APOBEC3 mutagenesis,<sup>16</sup> which increases the substitution  
631 rate<sup>9</sup> and may make the application of single clock model unreliable. We therefore did not  
632 include the sequences from this outbreak in the temporal reconstruction.

633

634 Methods to infer temporal history are only valid if there is a temporal signal within the  
635 dataset.<sup>45</sup> We assessed temporal signal using root-to-tip randomisation where we compared  
636 the R2 correlation between sample collection date and root-to-tip distance with that in 1000  
637 randomisations of collection dates. This supported the presence of a temporal signal ( $P <$



638 0.001). We therefore reconstructed the temporal history of clade I using BEAST v2.6.6  
639 employing the JC69 model of nucleotide substitution. We used a relaxed log-normal clock  
640 model with a log-normal prior on the substitution rate with mean 1.9E-6 (chosen to match  
641 the estimated slope in TempEst) and standard deviation 0.5. Population history was modelled  
642 using a coalescent constant population prior. Four independent runs were carried out for  
643 150,000,000 total MCMC steps. Convergence was assessed using Tracer v1.7 [REF] and all ESS  
644 values were above 550. 10% burnin was removed from each run before the runs were  
645 combined and the maximum clade credibility tree identified and annotated using  
646 TreeAnnotator.

647

#### 648 [Identification of genomic rearrangements](#)

649 Genome rearrangements in the Sudan clade I sequences were identified as described in  
650 Brinkmann et al.<sup>20</sup>

651

#### 652 [Calculation of mutational spectra](#)

653 We calculated a single base substitution mutational spectrum for each of the major MPXV  
654 clades using the sequence alignments and maximum likelihood phylogenetic trees generated  
655 above and containing all high-quality sequences. Phylogenetic trees were outgroup rooted  
656 using the outgroups described above to enable the direction of each mutation (i.e. the  
657 ancestral and mutated nucleotides) to be robustly identified. The outgroup was removed prior  
658 to spectrum calculation. We reconstructed mutational spectra using MutTui v2.0.2  
659 (<https://github.com/chrisruis/MutTui>).<sup>30</sup> We rescaled the resulting mutational spectra using  
660 MutTui v2.0.2 to account for the number of A, C, G and T nucleotides and the distribution of  
661 nucleotide triplets across the genome.<sup>30</sup>

662

663 It has previously been suggested that a dataset requires at least 300-600 mutations for the  
664 mutational spectrum to be accurately estimated.<sup>30</sup> The clade I, clade IIb A and lineage B.1  
665 datasets each contain more than 600 mutations. However, the clade IIa spectrum contains  
666 222 mutations so we did not attempt to examine the detailed contextual patterns in these  
667 mutations. To estimate whether the clade IIa mutational spectrum could have been generated  
668 by the spectrum of clade I or clade IIb, we compared the mutation type proportions in the  
669 clade IIa spectrum with those in 1000 random downsamplings of the clade I and lineage B.1  
670 mutational spectra to 222 mutations.

671

#### 672 [Data availability](#)

673 Sequence alignments (excluding GISAID sequences), BEAST log files and mutational spectra  
674 are available at [https://github.com/chrisruis/Mpox\\_global](https://github.com/chrisruis/Mpox_global). GenBank sample accessions and  
675 metadata are listed in **Table S1**. GISAID sample accessions and acknowledgements are listed  
676 in **Table S2**.

677

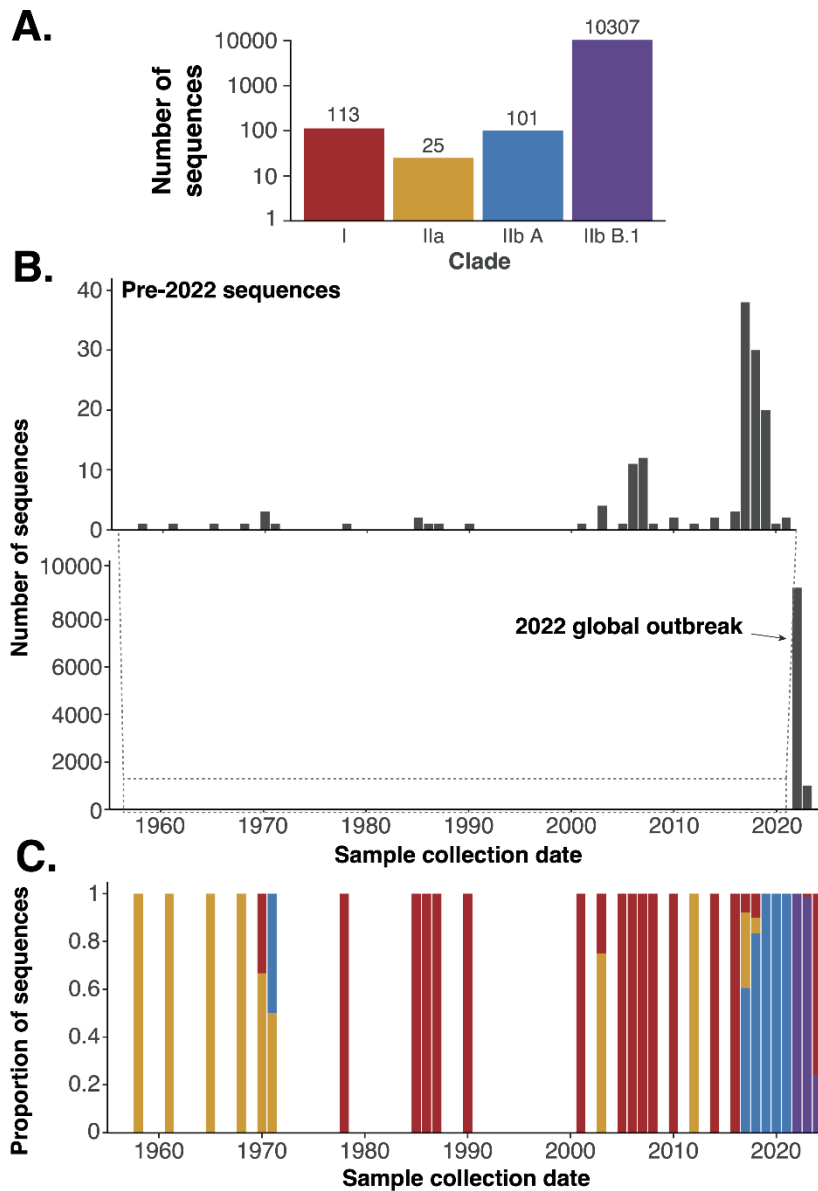
678

679 17

680

681 Supplementary figures

682



683

684

685

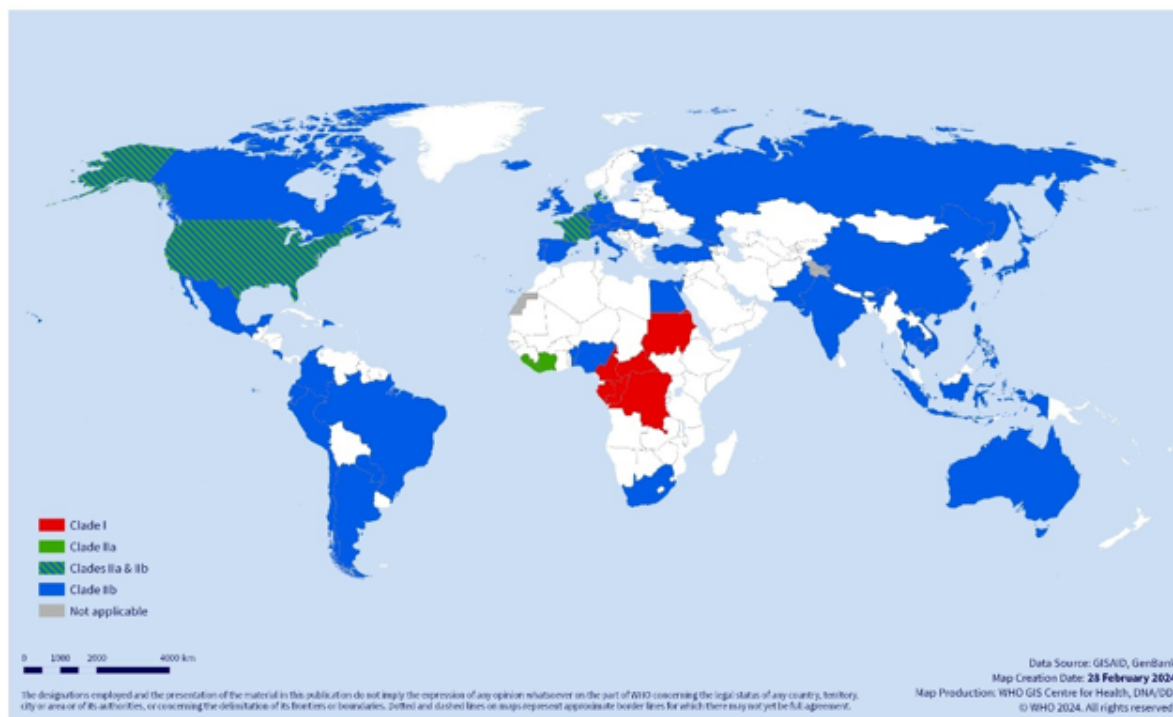
686

687

688

689

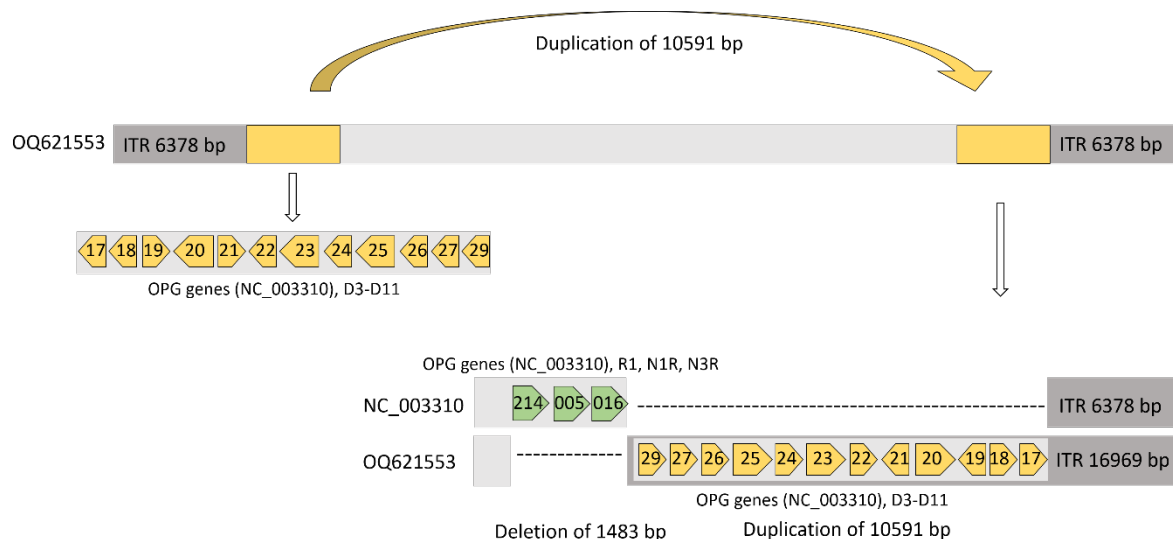
**Figure S1. Distribution of MXPV genetic sequences.** (A) The number of high-quality MXPV genetic sequences (see **Methods**) is shown for each major clade. (B) The number of MXPV sequences from each year is shown. Due to the large number of sequences in 2022 compared to other years, we extract the years before 2022 in the upper panel. (C) The proportion of sequences belonging to each major MPXV clade in each year is shown, colours match those in panel A.



690

691 **Figure S2: Geographic distribution of MPXV clades across WHO regions.** The map was  
 692 generated from 10,670 sequences obtained from GenBank and GISAID, 1958- February  
 693 2024. Clade I is red, clade IIa is green, and clade IIb is blue.

694



695

696 **Figure S3: Duplication of the genome from Sudan 2005 and 2022 (KC257459, OQ621553).** A  
 697 10591 bp region directly downstream the left ITR is duplicated to the right site of the genome,  
 698 resulting in ITRs of 16969 bp and a genome length of 204,808 bp). Three genes (R1, N1R and  
 699 N3R) are deleted at the site of duplication.

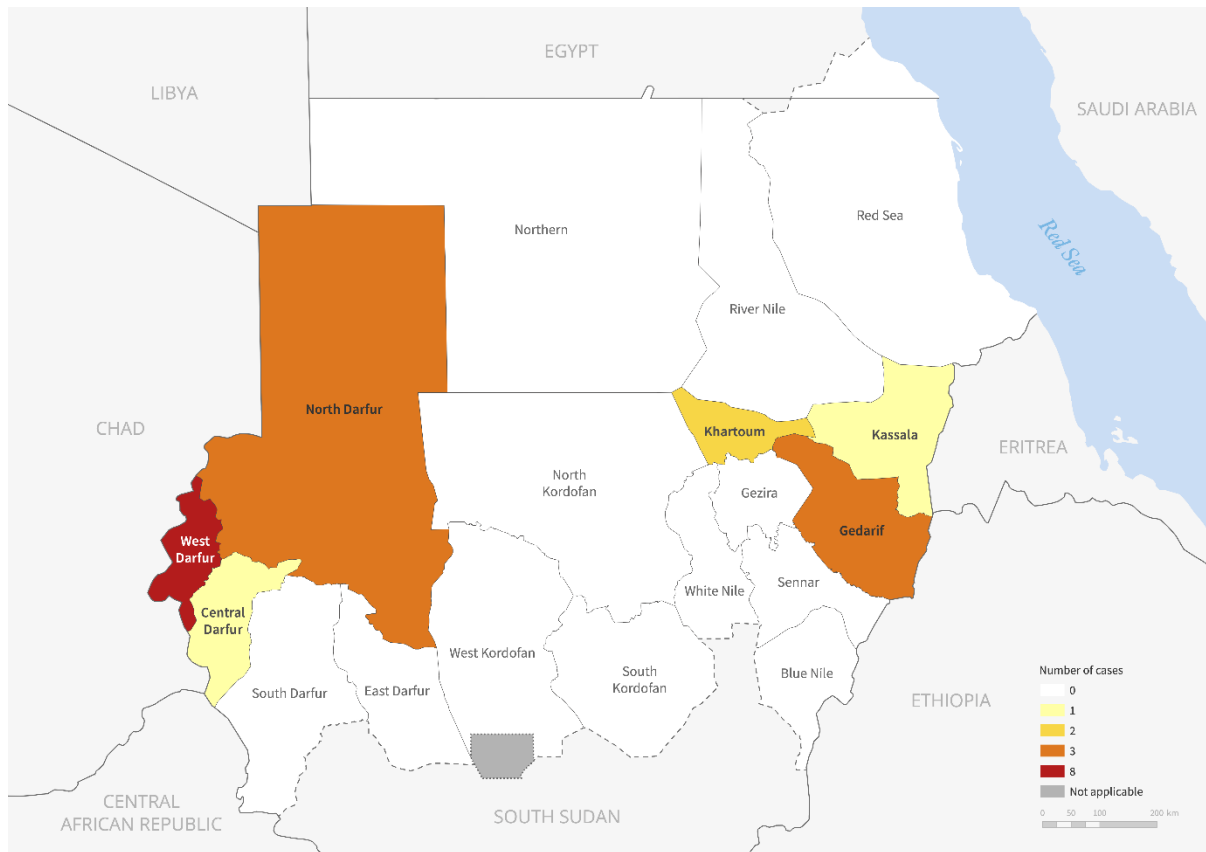
700

701

702

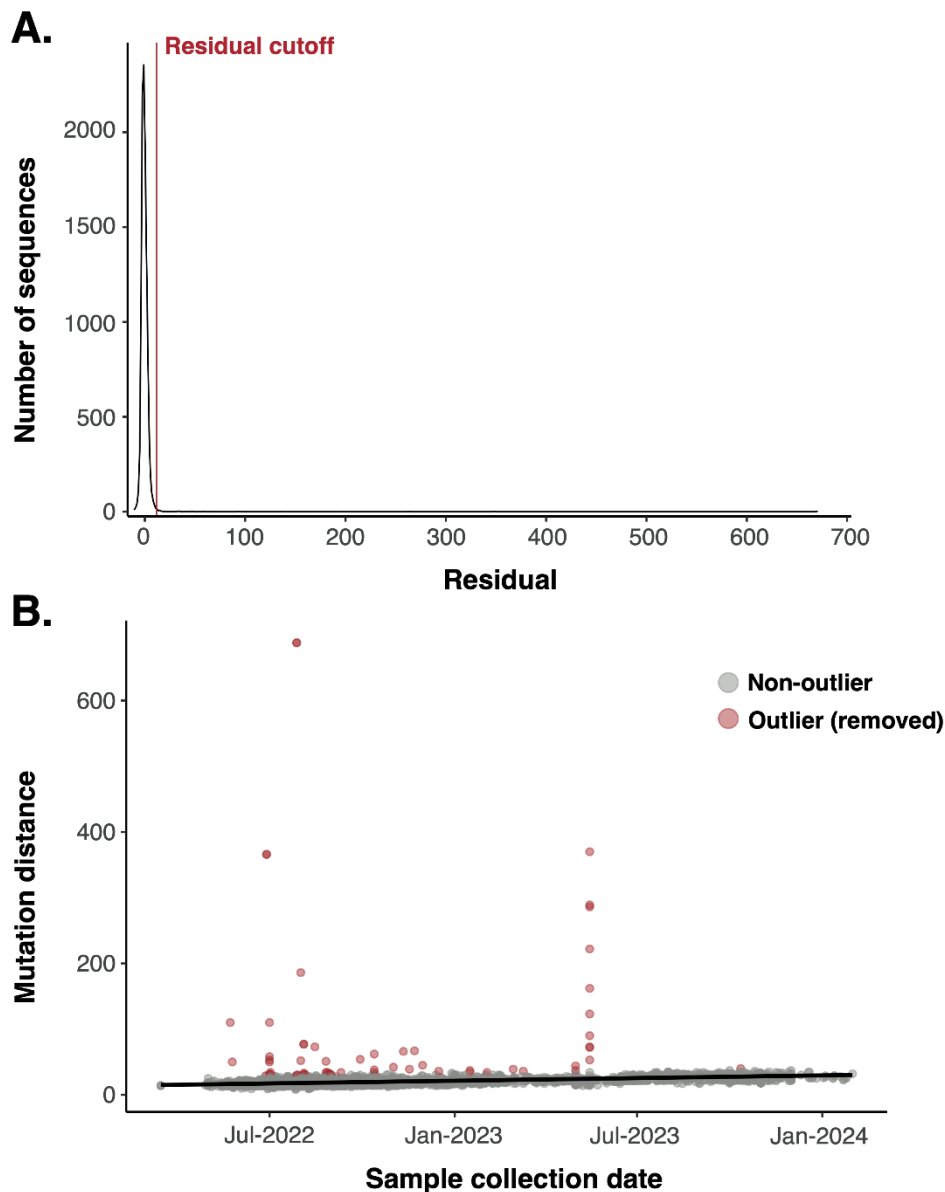
703

704



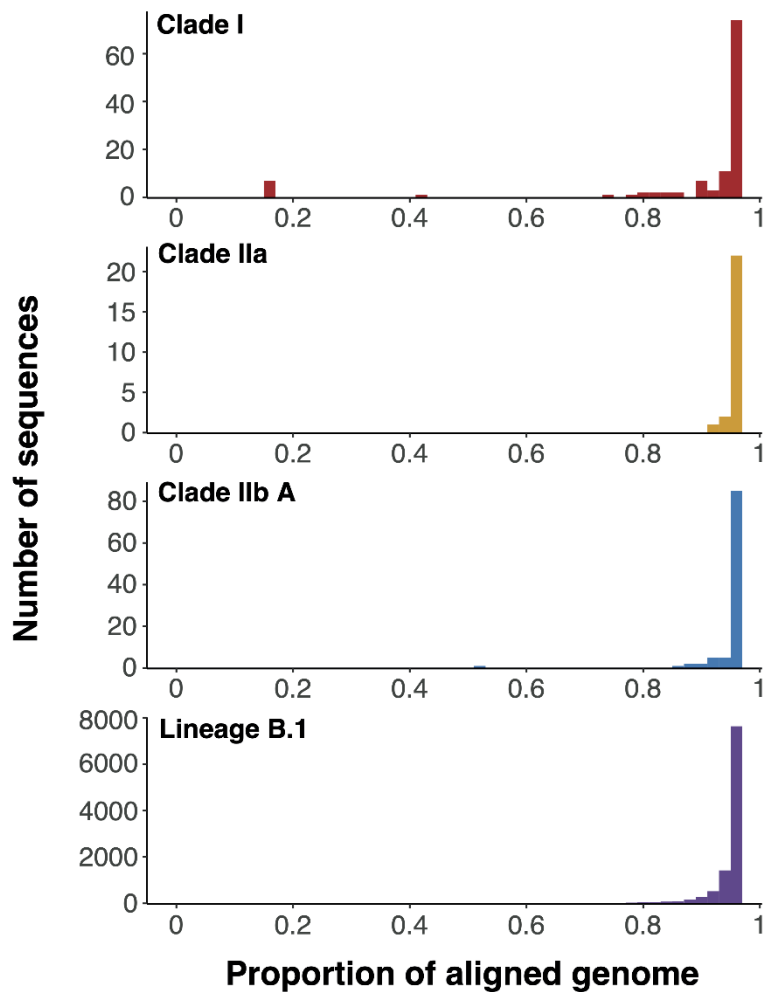
705  
706  
707  
708  
709  
710  
711

**Figure S4. Geographical distribution of laboratory confirmed cases of mpox in Sudan in 2022.** Confirmed cases were reported in Sudan states neighbouring Chad, Central African Republic, Ethiopia and Eritrea, while cases in 2005 were reported from a region that now belong to South Sudan.



712  
713  
714  
715  
716  
717  
718  
719  
720

**Figure S5. Identification of lineage B.1 outliers.** (A) The distribution of residuals around the best fit line shown in B. The red vertical line shows five median absolute deviations above the median residual which was used as a cutoff to identify outliers. (B) The collection date of each lineage B.1 sequence is plotted against the number of mutations between the sequence and sample ON676708.1 which clusters immediately upstream of lineage B.1. Sequences shown in red contain greater diversity than expected given their sampling data so were excluded from further analyses.



721

722

**Figure S6. Most included MPXV sequences contain close to the complete genome.** The proportion of the total genome covered is plotted for each sequence retained after filtering, split by major clade. Note that due to masking of ITR regions during alignment,<sup>9</sup> no sequence will contain the complete genome.

725

726

727

**Table S1. GenBank sample accessions and metadata**

728

729

**Table S2. GISAID sample accessions and acknowledgements**

730

731

732