

Global genomic surveillance of monkeypox virus

Received: 6 March 2024

Accepted: 21 October 2024

Published online: 23 October 2024

 Check for updates

James R. Otieno ^{1,21}, Christopher Ruis^{1,2,3,4,21}, Anyebe B. Onoja ¹, Krutika Kuppalli^{1,5,6}, Ana Hoxha ¹, Andreas Nitsche⁷, Annika Brinkmann⁷, Janine Michel ⁷, Placide Mbala-Kingebeni ^{8,9}, Daniel Mukadi-Bamuleka ^{9,10}, Muntasir Mohammed Osman¹¹, Hanadi Hussein¹¹, Muhammad Ali Raja ¹², Richard Fotsing¹³, Belinda L. Herring¹⁴, Mory Keita¹⁴, Jairo Mendez Rico¹⁵, Lionel Gresh ¹⁵, Amal Barakat¹⁶, Victoria Katawera¹⁷, Karen Nahapetyan¹⁸, Dhamari Naidoo¹⁹, R. Andres Floto ^{2,3,20}, Jane Cunningham¹, Maria D. Van Kerkhove ¹, Rosamund F. Lewis ¹ & Lorenzo Subissi ¹ ✉

Monkeypox virus (MPXV) is endemic in western and Central Africa, and in May 2022, a clade IIb lineage (B.1) caused a global outbreak outside Africa, resulting in its detection in 116 countries and territories. To understand the global phylogenetics of MPXV, we analyzed all available MPXV sequences, including 10,670 sequences from 65 countries collected between 1958 and 2024. Our analysis reveals high mobility of clade I viruses within Central Africa, sustained human-to-human transmission of clade IIb lineage A viruses within the Eastern Mediterranean region and distinct mutational signatures that can distinguish sustained human-to-human from animal-to-animal transmission. Moreover, distinct clade I sequences from Sudan suggest local MPXV circulation in areas of eastern Africa over the past four decades. Our study underscores the importance of genomic surveillance in tracking spatiotemporal dynamics of MPXV clades and the need to strengthen such surveillance, including in some parts of eastern Africa.

Mpox, formerly known as monkeypox, is a disease that is caused by the monkeypox virus (MPXV). MPXV is a member of the Orthopoxvirus genus, which also includes the variola virus, the causative agent of smallpox¹. In humans, mpox can be associated with a range of clinical symptoms, but classically presents with a short febrile prodromal phase, which lasts 1–5 days, followed by the appearance of a skin and/or mucosal rash, which might include single or multiple lesions^{2–4}. The incubation period of mpox has historically ranged from 4 to 14 days (ref. 5). MPXV is divided genetically into two clades—clade I (formerly known as Congo Basin clade) and clade II (formerly known as West African clade); clade II is further classified into subclades IIa and IIb (ref. 6). Clade I and subclade IIa circulate endemically within as yet unknown animal reservoirs, potentially including rodents and nonhuman primates, and human cases are mostly the result of spillover from these reservoirs^{1,7,8}. Historical surveillance has not

been sufficient to identify the frequency of spillover. In 2022, mpox epidemiology shifted with the emergence of a new lineage—clade IIb—that spread worldwide through human-to-human transmission, and based on the vast number of sequences from this outbreak, it was inferred that clade IIb has circulated continually within humans since at least 2016⁹. Mpox human-to-human transmission primarily occurs through direct contact with infected lesions or bodily fluids, which includes sexual contact, but transmission can also occur through contact with fomites¹⁰.

In May 2022, a novel lineage of clade IIb, termed B.1, emerged and spread globally, establishing efficient local transmission within many countries with no previous history of mpox transmission. As of 12 August 2024, the multicountry outbreak has been associated with 99,176 cases and 208 fatalities from 116 countries, areas and territories, representing a case fatality ratio (CFR) of 0.21%. The outbreak is

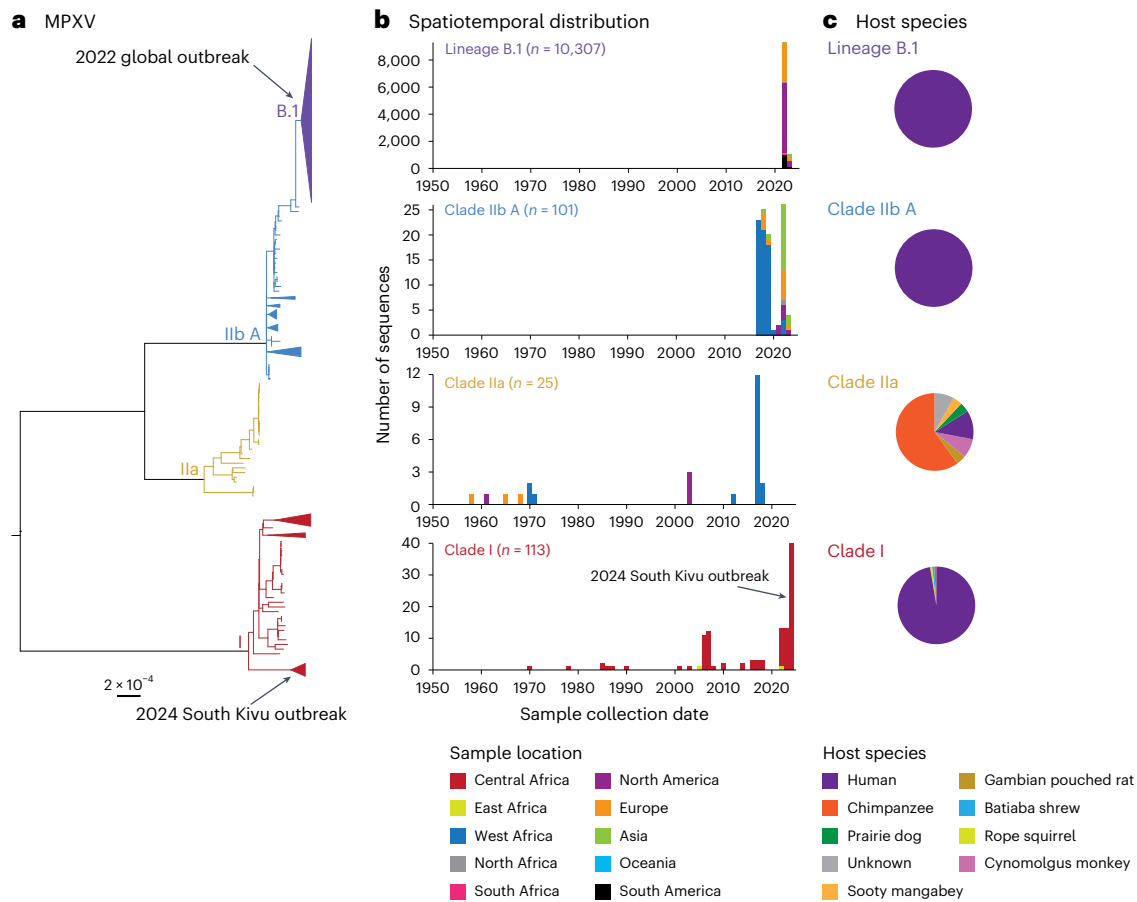


Fig. 1 | Spatiotemporal and host species distributions of MPXV sequences (worldwide, 1958–2024). **a**, Maximum likelihood phylogenetic tree highlighting the major clades of MPXV. The branches are colored by clade. Lineage B.1 clusters within clade I Ib and caused the 2022 global MPXV outbreak; this lineage is therefore separated from the remainder of clade I Ib. The scale bar shows

the expected number of nucleotide substitutions per site. A subset of clades are collapsed for clarity. **b**, The temporal and regional distribution of MPXV sequences is shown for each clade. The *n* numbers show the total number of sequences from the clade. **c**, Distributions of the number of sequences from each host species.

primarily driven by sexual transmission among males who self-identify as men who have sex with men, with 7% of cases requiring hospitalization. Other groups at higher risk of hospitalization include female cases, those younger than 5 years of age or greater than 65 years of age, and the immunosuppressed (either due to being HIV positive or from other immunocompromising conditions)¹¹. In response to the global outbreak, a number of countries have started to establish mpox surveillance programs.

In addition to the global outbreak, the detection of mpox is spreading to new African countries such as Burundi, Rwanda, Uganda and Kenya, which led to the declaration of a second Public Health Emergency of International Concern in August 2024 by the World Health Organization (WHO)¹². In 2024, and as of 14 September 2024, a total of 21,835 suspected mpox cases were reported by the Democratic Republic of the Congo, including 5,160 laboratory-confirmed cases (sample testing rate: 46%), with a positivity rate of 51.3% (ref. 12). This is substantially higher than those of previous years. This recent increase, along with the newly documented sexual transmission recorded in March 2023 in Kwango province and then in September 2023 in South Kivu province, which later drove the spread to Burundi, Rwanda, Uganda and Kenya, confirms the growing importance of human-to-human transmission, including through sexual contact, in Central and eastern Africa^{13–15}.

The global mpox surveillance that was quickly established in 2022 provided a platform for the generation of genome sequencing data. These data have been useful in characterizing MPXV evolution,

understanding the origins of emerging lineages and monitoring local and global spread. Previous studies have shown that clade I Ib exhibits a higher substitution rate than other Orthopoxvirus variants^{9,16–18}. This appears to be due to elevated TC>TT mutations (which represent C mutating to T with an upstream T nucleotide and also include the reverse GA>AA mutations) driven by human apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3 (APOBEC-3) proteins causing cytosine deamination in the viral genome^{9,16–18}. This mutational signature has enabled inference that clade I Ib is transmitting from human to human.

In addition to characterizing mutations and mutational processes, global genomic surveillance can also enable monitoring of the integrity and stability of the MPXV genomic termini, which can rearrange driving gene duplication or gene loss¹⁹. These rearrangements are also drivers of poxvirus evolution and host adaptation, and are therefore important to monitor^{20,21}.

Mpox surveillance programs and their associated genomic strategies globally remain critical to the understanding of the disease and the characterization of the virus evolutionary trajectory and genetic diversity, and provide potential insights into the associated phenotype. Ultimately, these data support the deployment of suitable countermeasures (diagnostics, therapeutics and vaccines), leveraging efforts garnered from smallpox interventions before eradication and preparedness in the years since, as well as advance research and development for countermeasures. Here we report a global analysis of the publicly available MPXV genomic sequence data, which provide key insights

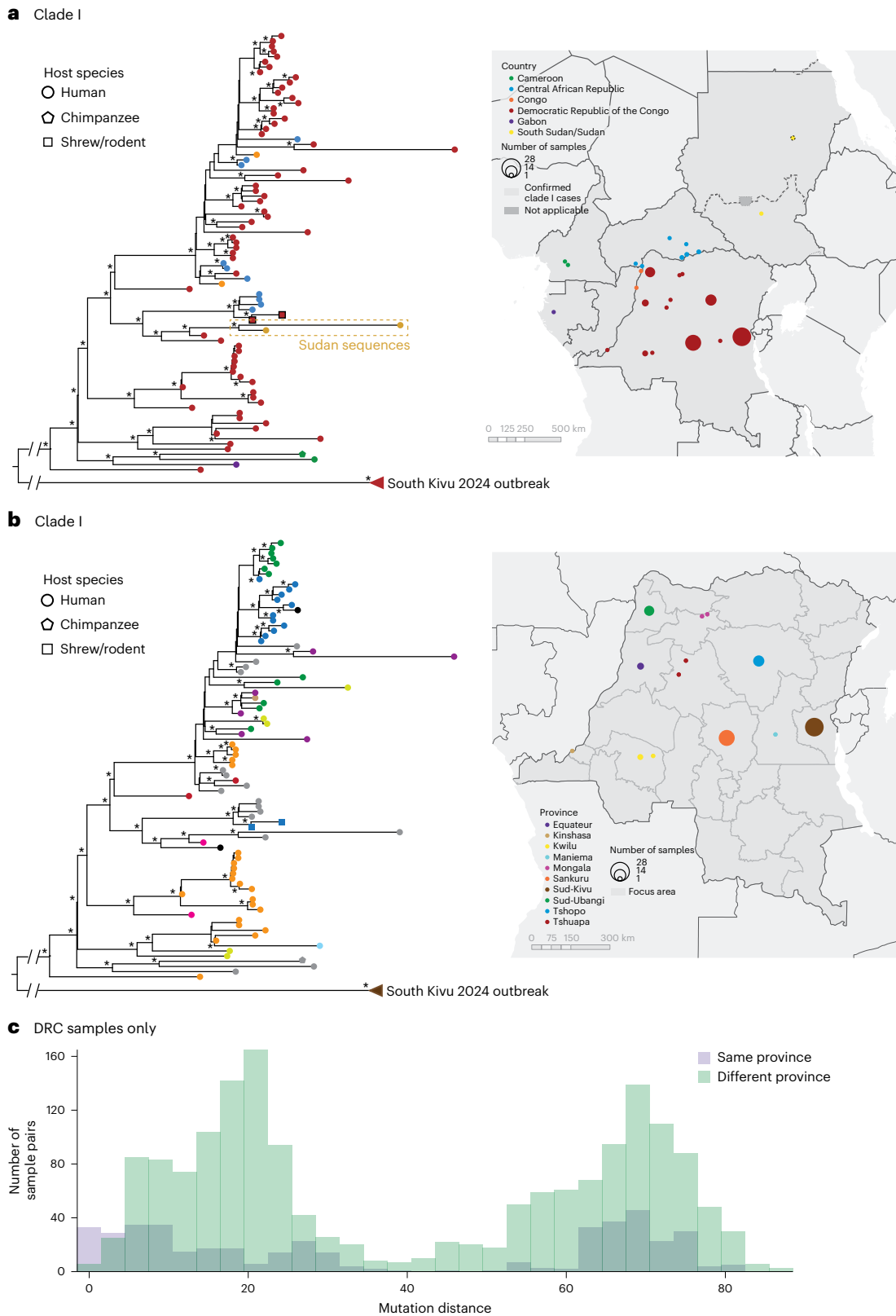


Fig. 2 | Regular international and inter-province transmission of clade I. **a,b**, Maximum likelihood phylogenetic tree of 113 high-quality clade I sequences. **a**, The tips are colored by country to match the map, and the shapes show the host species from which the sequence was isolated. The Sudan sequence cluster is highlighted. The 2024 South Kivu outbreak clade has been collapsed for clarity. The asterisks show phylogenetic nodes with bootstrap support of 70 or higher. The map shows sampling locations with points proportional to the number of sequences from the location. **b**, The tips are colored by province within the

Democratic Republic of the Congo to match the map. Tips collected outside of the Democratic Republic of the Congo are colored gray and tips sampled within the Democratic Republic of the Congo but without a recorded province are colored black. **c**, We calculated the mutation distance between all possible pairs of clade I sequences (that is, each sequence was compared against all other sequences); the number of mutations is plotted stratified by whether the pairs are from the same (purple) or different (green) provinces.

into the spatiotemporal spread, host species range and evolution of this ongoing threat.

Results

We collected all available MPXV sequences and filtered them to retain 10,546 high-quality sequences from 64 countries (Methods). Of these sequences, 6,585 were extracted from GenBank (62%) and 3,914 sequences from the Global Initiative for Sharing of All Influenza Data (GISAID; 37%). Owing to its distinct epidemiology and sequence diversity, we divided clade IIb into the A sublineages and the B.1 lineage for the analyses below, and refer to these groupings as clade IIb A and lineage B.1, respectively. As expected, the majority (97.7%) of the available MPXV sequences cluster within lineage B.1 (Extended Data Fig. 1a), representing intensive genomic sequencing efforts during the global outbreak. Clades I, IIa and IIb A have been sequenced far less often (Extended Data Fig. 1a). Correspondingly, the majority (98.6%) of MPXV sequences were collected from 2022 to 2024, with limited historical surveillance resulting in many years between 1958 and 2015 with no sequences in global databases (Extended Data Fig. 1b). Nonetheless, we observe differences in the temporal, spatial and host species distributions of the major MPXV clades and discuss them below.

MPXV clades show distinct temporal sampling distributions

Subclade IIa and clade I were the first to be detected, in 1958 and 1970, respectively (Fig. 1 and Extended Data Fig. 1c)^{22,23}. Sporadic detection of both clades has continued through to recent years (Fig. 1 and Extended Data Fig. 1c), showing continued circulation within the animal reservoir. Clade I continued to be detected in the Democratic Republic of the Congo and Sudan during the lineage B.1 outbreak between 2022 and 2024 (Fig. 1), with the Democratic Republic of the Congo cases in 2024 being associated with a novel divergent lineage showing signatures of human-to-human transmission within South Kivu^{14,15}. Clade IIa has not been observed since 2018 (Fig. 1). Except for a sample from 1971, all clade IIb genomes were sampled from 2017 to 2023. The 1971 sample probably reflects the fact that also clade IIb originated in an animal reservoir. Clade IIb A was first detected in Nigeria in 2017 and has continued circulating through human-to-human transmission to at least 2023⁸, while the descendent lineage B.1 was first detected in 2022 (Fig. 1 and Extended Data Fig. 1c).

MPXV sampling shows distinct host species patterns

Although it probably originated from an animal reservoir, clade IIb circulates via human-to-human transmission²⁴, and correspondingly, both clade IIb A and B.1 lineages have, to date, been sampled exclusively in humans (Fig. 1). While clades I and IIa both circulate within poorly understood animal reservoirs, the hosts from which they have been sampled are markedly different. The majority of clade I sequences (96%) have been sampled from humans, with single samples from an outbreak in captive chimpanzees²⁵, and from wild shrew (*Crocidura littoralis*) and rope squirrel (*Funisciurus anerythrus*)²⁶. Conversely, only 12% (3 of 25) of clade IIa samples with recorded host species were collected in humans (Fig. 1), two in 1970 and one in 2003. Clade IIa has been isolated most often in chimpanzees (60%, Fig. 1), although these samples are mostly from a single study within Tai National Park in the Ivory Coast⁷, and chimpanzees are probably a spillover host rather than a reservoir host. Clade IIa has additionally been isolated from a wild sooty mangabey, imported cynomolgus monkeys in the USA and Denmark, and a prairie dog during the 2003 USA outbreak (Fig. 1)²⁷.

Geographical distribution of MPXV clades

Clade I has mostly been isolated from the Congo Basin area, where it has been sampled in the Democratic Republic of the Congo, the Republic of the Congo, Central African Republic, Cameroon and Gabon (Fig. 2 and Extended Data Fig. 2). Sampling locations are broadly spread around these countries. Sequences from individual countries and provinces

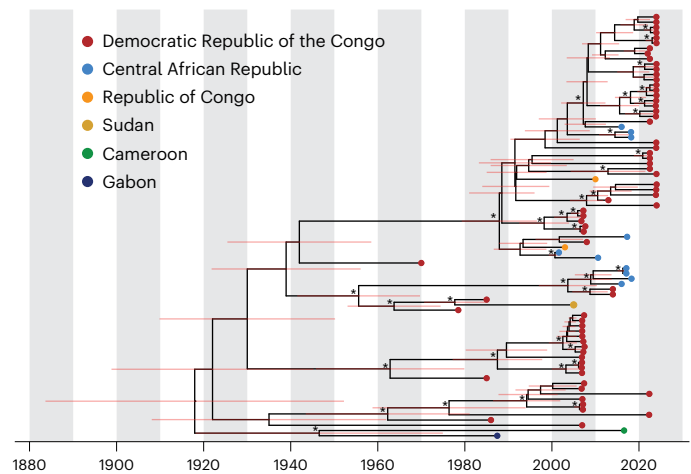


Fig. 3 | Temporal evolutionary history of clade I. The temporal maximum clade credibility phylogenetic tree is shown. The tips are colored by country of isolation. The red bars show the 95% HPD on the date of the corresponding node. The asterisks show nodes with posterior support of 70 or higher.

often do not cluster within the phylogenetic tree (Fig. 2), showing multiple introductions of clade I into local geographical regions.

Having first confirmed the presence of a temporal signal (Methods), we sought to identify the timescale of these virus movements by reconstructing a temporal phylogenetic tree (Fig. 3). We found that the most recent common ancestor of clade I (excluding the 2024 South Kivu outbreak that has a different substitution rate; Methods)⁹ occurred in approximately 1917 (95% highest probability density (HPD) 1880–1949). We observe frequent international and inter-province transmission over the past several decades (Figs. 2 and 3). For example, a clade sampled in Sud-Ubangi, Equateur and Kinshasa in 2023–2024 coalesces to a common ancestor in 2007 (95% HPD 2003–2011), supporting recent virus movement within the animal reservoir (Figs. 2b and 3). Furthermore, we observe highly similar distributions of genetic relatedness between clade I samples from the same and different provinces within the Democratic Republic of the Congo (Fig. 2c), further highlighting the regular movement of viruses between geographical locations.

Clade I has been isolated outside of the Congo Basin in Sudan in 2005 (in a region that is now part of South Sudan) and 2022 (Fig. 2a). The two Sudan sequences cluster in the phylogenetic tree and share an ~10.5 Kb duplication (Extended Data Fig. 3). We estimate that the Sudan sequences diverged from their closest sampled relative (a Democratic Republic of the Congo sample from 1985) in 1978 (95% HPD 1969–1984); this lineage has therefore been sampled only in Sudan over roughly 46 years. It is likely that this lineage has circulated in the animal reservoir during this period as it shows 8.5% TC>TT mutations, highly similar to that expected in animals (8%) but far lower than expected from evolution in humans (85%)⁹. This is confirmed by the geographical spread (6 states in western, southern and eastern Sudan) of the 18 mpox cases that were laboratory confirmed in Sudan in 2022 (Extended Data Fig. 4).

Clades IIa and IIb A both circulate in West Africa and have been exported to other regions (Figs. 1 and 4). Clade IIa has not been observed outside West Africa since an outbreak in the USA in 2003 (Fig. 1). Samples of clade IIa from West Africa remain sparse with single sequences from Liberia and Sierra Leone from human cases in 1970 and two closely related clusters of sequences from chimpanzees in the Ivory Coast collected from 2017 to 2018 (Fig. 4a)⁷. We therefore currently lack the resolution to examine spatial transmission patterns in more detail for clade IIa.

While clade IIb A initially spread in Nigeria, it was then exported to other countries in Europe, Asia, North America and, more recently, North Africa (Figs. 1 and 4b). Genetic sequences from multiple

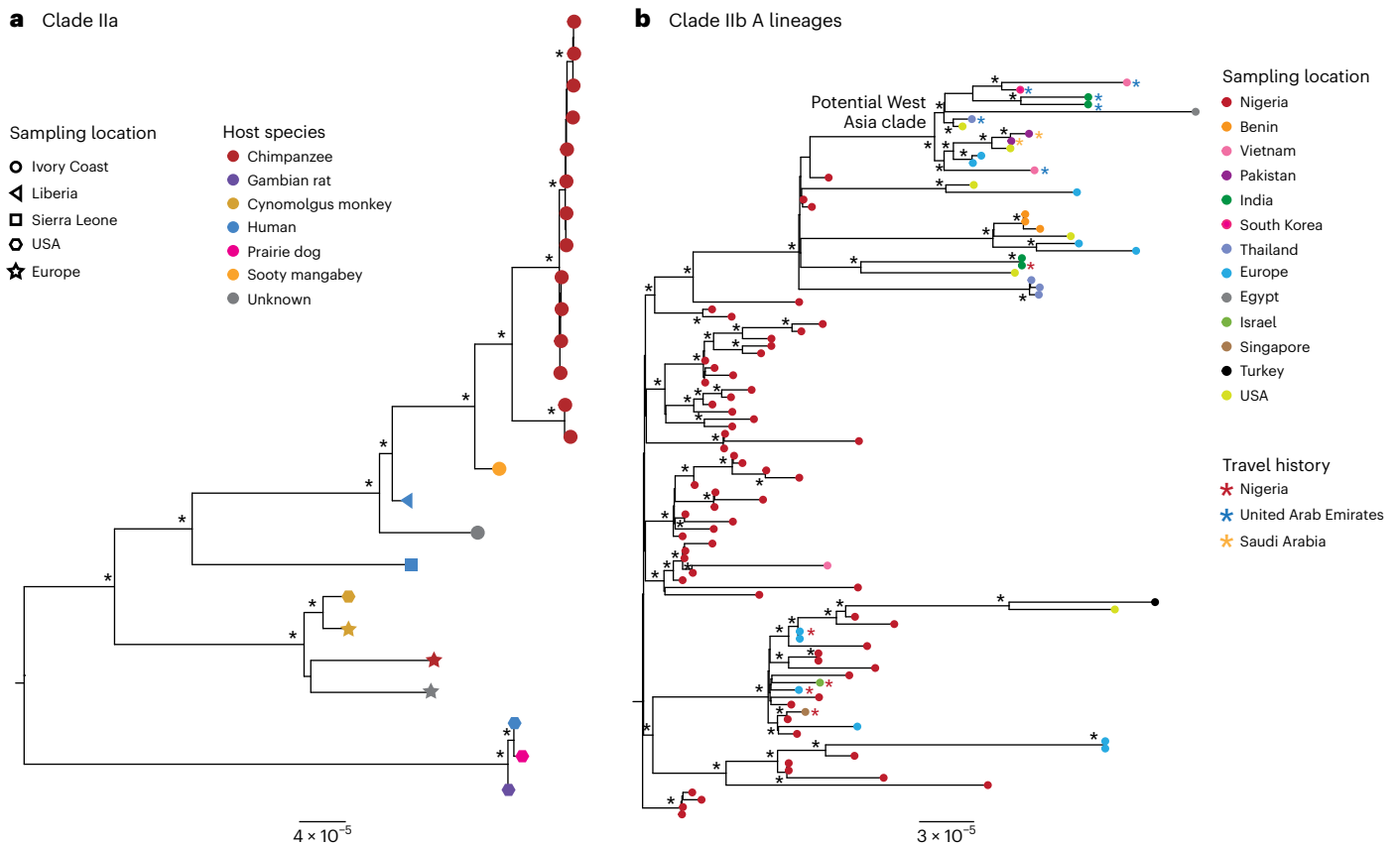


Fig. 4 | Spatial distributions of clades IIa and IIb A. **a**, Maximum likelihood phylogenetic tree of 25 high-quality clade IIa isolates. The tips are colored by host species and the shapes show sampling locations. The asterisks show nodes with bootstrap support of 70 or higher, and the scale bar shows the expected number of nucleotide substitutions per site. **b**, Maximum likelihood phylogenetic tree of 101 clade IIb A sequences. The tips are colored by country of collection. The red,

blue and orange asterisks show samples with travel history. The potential Eastern Mediterranean clade containing samples with travel history to the United Arab Emirates and Saudi Arabia is highlighted. The black asterisks show nodes with bootstrap support of 70 or above, and the scale bar shows the expected number of nucleotide substitutions per site.

individuals infected with clade IIb A from the UK, Israel, Singapore and India have travel history to Nigeria (Fig. 4b), supporting infection in endemic regions of Nigeria and subsequent export. In addition, eight sequences from India, South Korea, Vietnam and Thailand were isolated from travellers returning from the United Arab Emirates or Saudi Arabia; these sequences cluster within a single phylogenetic lineage that also includes sequences from the USA, UK, Slovenia and Egypt for which no information regarding recent travel is recorded (Fig. 4b). This is consistent with sustained circulation of a lineage of clade IIb A in the Eastern Mediterranean region; no sequences are currently available from the Eastern Mediterranean to allow further examination.

MPXV mutational spectra are influenced by transmission route

Previous studies have shown that mutational spectra of MPXV lineages that are transmitted from human to human show a high proportion of TC>TT mutations, implicating APOBEC-3 as a major driver of mutagenesis in human MPXV infections⁹. We calculated complete single-base substitution (SBS) mutational spectra (incorporating all possible nucleotide contexts for each mutation type) for the dominant clades of MPXV and corrected these for genomic composition (Fig. 5 and Methods). We find similar mutational patterns to previous studies^{9,18}; for example, the spectrum of clade IIb (both A and B.1 lineages) is dominated by C>T mutations in which the C is preceded by a T, with preference for A or G following the substitution site (Fig. 5).

The clade I mutational spectrum shows that mutational processes within the animal reservoir are also dominated by C>T mutations (66% of the mutational burden accounting for genome composition), but with different contextual preferences to evolution within humans (Fig. 5). In clade I, C>T mutations occur most commonly in which the C is preceded by G (57% of the C>T mutational burden, Fig. 5b) and are most frequent in G[C>T]A and G[C>T]G contexts (Fig. 5a). C>A (which includes G>T mutations) is the second most common mutation in clade I (17% of the total mutational burden, Fig. 5). We observe contextual preferences within C>A mutations, with mutations in AC and GC contexts being less common (Fig. 5). As reactive oxygen species cause G>T (ref. 28), they are a potential driver of these mutations.

The clade IIa mutation spectrum contains 222 mutations, which may be too low to examine mutational patterns in detail²⁹. We therefore used the clade I and lineage B.1 mutational spectra as references and investigated whether either could explain the clade IIa mutational spectrum. We found that the clade IIa mutational patterns could have been generated by the clade I spectrum but not by the lineage B.1 spectrum (Fig. 5c). This suggests that clade IIa shows similar mutational processes as clade I, consistent with a lack of human APOBEC-3 activity and therefore both clades consistently circulating outside of humans.

Discussion

Implementing measures to control mpox outbreaks will require an in-depth understanding of how the virus is transmitting within human populations and among animals across different spatial scales. Here we carried out an in-depth analysis of spatiotemporal and host species

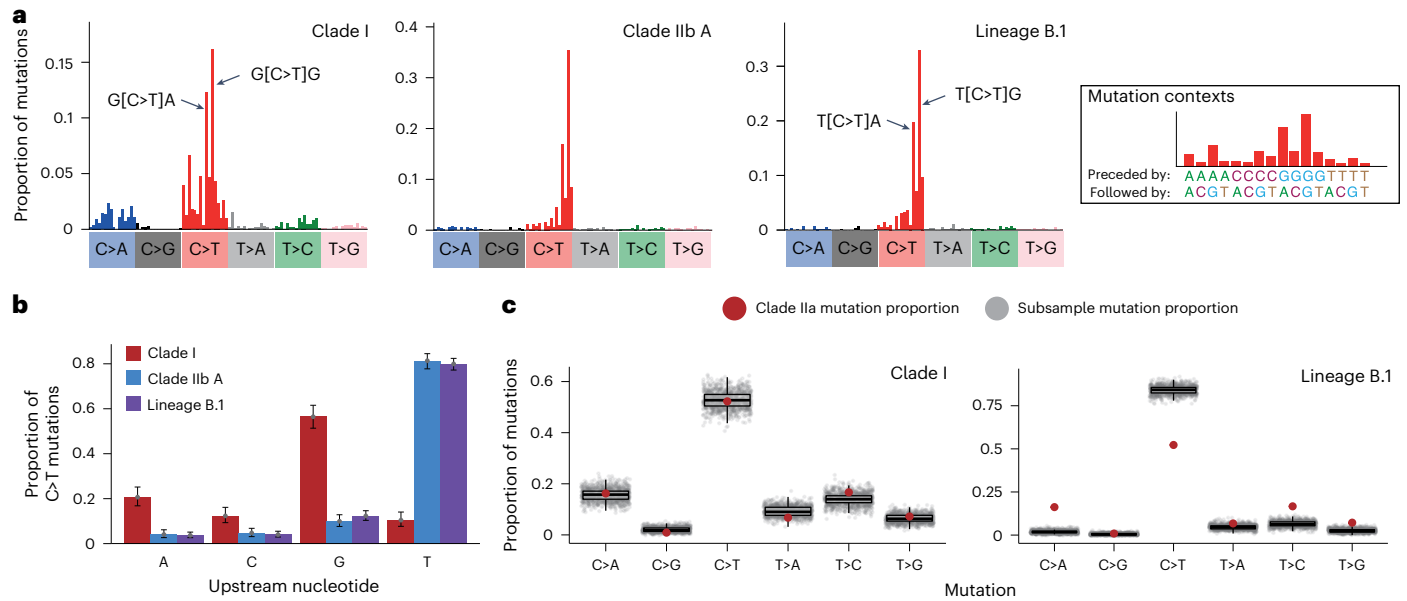


Fig. 5 | Mutational spectra differ between major MPXV clades. a, SBS mutational spectra for clade I, clade IIb A and lineage B.1 (clade IIa was not included because of insufficient mutations; Methods). SBS spectra show the proportion of mutations of each mutation type within each surrounding nucleotide context; contexts for an example mutation type are shown in the right-hand panel. The two most prevalent contextual mutations are highlighted for clade I and lineage B.1. Mutational spectra are corrected for genome composition (Methods). Symmetrical mutations (for example, C>T and G>A) are combined as MPXV is a DNA pathogen²⁹. **b**, The proportion of C>T mutations with each nucleotide upstream is shown for each clade. Each bar and dot show the proportion of C>T mutations that occur in the corresponding context within the clade (that is, $n = 1$ in each case). The error bars represent the Wilson score interval calculated using the corresponding proportion and number of sampled

C>T mutations. **c**, To examine the potential for the clade IIa mutational spectrum to have been generated by the mutational spectra of clade I and lineage B.1, we compared the proportion of each mutation type in the clade IIa spectrum with that in 1,000 subsamplings of the other clade spectrum to the number of mutations in the clade IIa spectrum (Methods). Each gray point represents the mutation type proportion in one subsample of the respective mutational spectrum while each red point shows the mutation type proportion in clade IIa. The clade IIa mutation type proportions are within that expected from clade I but often outside that expected from lineage B.1. The boxplot center lines show median values; the upper and lower bounds show the 25th and 75th quantiles, respectively; the upper and lower whiskers show the largest and smallest values within 1.5 times the interquartile range above the 75th percentile and below the 25th percentile, respectively.

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

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

While clades I and IIa both circulate within animals and spillover into humans, we observe highly different distributions of host species among sampled genetic sequences, with clade I mostly being sampled from humans and clade IIa from animals (Fig. 1c). This difference in host sampling may be driven by (1) distinct abilities of the clades to infect humans, (2) differential disease severity in humans and/or animals altering the likelihood of case detection, (3) different contacts between humans and sampled animals and animal reservoirs (which may differ for the two clades) and/or (4) different surveillance and sampling strategies in humans and animals within the affected countries. Identifying

the driver(s) of differential host sampling will require stronger surveillance in humans and animals, and linkage with sample metadata to determine likely routes of infection.

While clade I has mostly been detected within the Congo Basin, our data highlight the potential for widespread endemic and/or enzootic circulation of this clade within Sudan (former Sudan, now divided into Sudan and South Sudan). We identified a clade I lineage that has been sampled only from Sudan over roughly 45 years (Figs. 2 and 3). This lineage has probably circulated within the animal reservoir (owing to a lack of APOBEC-3-like mutations) and may have circulated continually in Sudan following introduction at any point during this time period. Alternatively, this lineage could have been introduced multiple times from an unsampled region either shortly before being sampled in Sudan or with some level of local transmission. Distinguishing between these possibilities will require additional sequencing data from Sudan. However, our results, combined with recent epidemiological data, highlight the potential for large numbers of MPXV cases in Sudan³⁰. Following the declaration of the first Public Health Emergency of International Concern in July 2022, the Federal Ministry of Health of Sudan, with support from the WHO, started mpox surveillance. As a consequence, suspected mpox cases were reported from 17 states with 42 affected localities, of which some states host refugees and internally displaced persons. Over 40% of the suspected cases were children under the age of five³⁰. Of the suspected cases, only a portion was tested and a total of 18 cases were laboratory confirmed from 6 states and 9 localities, including 1 death (CFR 5.8%)³¹.

Our data, combined with the recent emergence of a novel human transmissible clade I lineage in South Kivu^{14,15}, where mpox cases have not previously been detected (except for a few cases in 2011 and 2012),

highlight the potential for an increased risk of international spread of clade I MPXV. Recent studies highlighted that the affected population in South Kivu was composed of young adults, of which half were females and 30% were sex workers, with over 80% of patients reporting recent visits to bars for (professional) sexual interactions^{14,15,32}. This represents a significant shift in the historical epidemiology of mpox in the Democratic Republic of the Congo, which involves children <15 years of age as the main affected age group^{14,33}. As a comparison, in the years preceding the eradication of smallpox (1956–1971), the maximum number of smallpox cases reported by the Democratic Republic of the Congo (at that time Zaire) Ministry of Health to the WHO was 5,523 cases including 710 deaths in 1963³⁴, which is a quarter of the suspected mpox cases reported by the Democratic Republic of the Congo in 2024³³. Concerningly, a recent study has reported that 8 of 14 pregnant women with mpox had fetal loss³². Considering the high mobility of the population in South Kivu, cross-border transmission is likely to occur, as highlighted by the introductions of clade Ib to North Kivu³⁵ and then Burundi, Uganda, Rwanda and Kenya. Further studies to better understand transmission patterns in these settings (including whether enzooticity has been established) and the geographical distribution of mpox in eastern Africa are urgently needed.

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

Local human-to-human transmission chains were established in many countries across all six WHO regions during the 2022 MPXV global outbreak³⁶. Local transmission of the ancestral clade IIb A lineages has also been identified in some cases outside of West Africa³⁷. We were here able to infer local transmission of clade IIb A MPXV in the Eastern Mediterranean through travel data associated with sequences from other countries, despite sequences from that region being unavailable for analysis. This highlights the importance of associating detailed metadata with genetic sequences where possible.

Mutational signatures have provided major insights into MPXV and can be used to identify lineages that are transmitting from human to human and to infer outbreak origin dates^{9,18}. Here we calculated and compared in-depth mutational spectra of the major MPXV clades, showing that C>T mutations are most common within both human and animal hosts, but differences in contextual preferences exist between species (Fig. 5). The C>T mutations observed within clade I occur most commonly where the C is preceded by G (Fig. 5b) and are most frequent in G[C>T]A and G[C>T]G contexts (Fig. 5a). These mutations are unlikely to be the result of spontaneous deamination of cytosine as this has a strong preference for CG>TG contexts in human DNA³⁸. The C>T mutations may therefore be the result of polymerase errors during genome replication, the action of alternative APOBEC enzymes within the animal reservoir and/or additional mutagens. The clade IIb spectra show similar enrichment of G over A and C as the nucleotide preceding C>T mutations (Fig. 5b), which may suggest that the GC>GT mutations are driven by a non-host species factor, but this will require additional future work to untangle. Our analyses suggest that the ratio between GC>GT mutations and TC>TT mutations is a reliable marker to distinguish human-to-human transmission from transmission in animals and is therefore potentially able to identify sustained human outbreaks from transmission from the reservoir.

The global outbreak has provided critical insights into the epidemiology of mpox in humans. However, it is unknown whether the characteristics of lineage B.1 were acquired following adaptation in humans, or whether they may be generalizable across MPXV clades. This can only be revealed with stronger surveillance, sharing of sequences and continued genotypic and functional characterization of the differences between clades and lineages³⁹. Such characterization would shed light into the drivers of epidemiological differences between clades, but this work is technically challenging and, currently, few laboratories worldwide have such capacity. However, a small number of studies have found that the apparent difference in morbidity and mortality between clades I and II is probably driven by multiple proteins present in clade I but absent in clade II^{40,41}. One particular area of interest is gene duplication, such as that found in clade I sequences from Sudan. Such duplications have been described for clade IIb sequences from the 2022 outbreak, when they were assumed to be involved in immune evasion and host range^{20,42,43}. Gene duplication and loss in the MPXV terminal regions are also considered drivers of poxvirus evolution and adaptation to the host^{20,21}. Analyses comparing strains with and without deletions will be essential to uncover their functional consequences and understand and forecast the epidemiology of lineages showing such changes. In addition, deletions may lead to diagnostic failure, especially for nucleic acid amplification tests that target less conserved genes, such as some of the clade-specific PCRs, which are designed to distinguish clades. To date, there have been two reported examples of such diagnostic failure episodes, one for a variant of MPXV clade IIb detected in the USA that did not spread widely⁴⁴, and one for the clade I lineage currently circulating in South Kivu¹⁴. This clearly highlights the critical importance of a strategic genomic surveillance system where mpox circulates.

As highlighted in the standing recommendations for mpox issued by the director general of the WHO, it is critical that countries have national mpox strategic plans integrated into broader health systems, and that capacities that have been built in resource-limited settings and among marginalized groups should be sustained⁴⁵. Without surveillance, no genomic sequence data can be generated and no virological characterization of circulating clades and lineages can be done. In this regard, more laboratories should engage in virological characterization of MPXV clades and lineages. Furthermore, countries are strongly encouraged to continue documenting and making sequences publicly available, prioritizing specimens for both targeted sequencing (for example, of imported cases, the first few cases of local emergence and cases with divergent demographic or clinical profiles) and representative sequencing. This will enable the tracking of virus circulation and evolution over time. If we want to prevent the next mpox global outbreak, it is now time to strengthen mpox surveillance, including in Africa, focusing on populations at highest risk and ensuring integration with existing systems to ensure comprehensive and seamless delivery of care.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-024-03370-3>.

References

1. Bunge, E. M. et al. The changing epidemiology of human monkeypox—a potential threat? A systematic review. *PLoS Negl. Trop. Dis.* **16**, e0010141 (2022).
2. Damon, I. K. in *Goldman's Cecil Medicine*, 24th edn, Vol. 2 (eds. Goldman, L. & Schafer, A. I.) 2117–2121 (Elsevier, 2012); <https://doi.org/10.1016/B978-1-4377-1604-7.00380-8>

3. Damon, I. K. Status of human monkeypox: clinical disease, epidemiology and research. *Vaccine* **29**, D54–D59 (2011).
4. Adler, H. et al. Clinical features and management of human monkeypox: a retrospective observational study in the UK. *Lancet Infect. Dis.* **22**, 1153–1162 (2022).
5. Nolen, L. D. et al. Extended human-to-human transmission during a monkeypox outbreak in the Democratic Republic of the Congo. *Emerg. Infect. Dis.* **22**, 1014–1021 (2016).
6. Ulaeto, D. et al. New nomenclature for mpox (monkeypox) and monkeypox virus clades. *Lancet Infect. Dis.* **23**, 273–275 (2023).
7. Patrono, L. V. et al. Monkeypox virus emergence in wild chimpanzees reveals distinct clinical outcomes and viral diversity. *Nat. Microbiol.* **5**, 955–965 (2020).
8. Yinka-Ogunleye, A. et al. Outbreak of human monkeypox in Nigeria in 2017–18: a clinical and epidemiological report. *Lancet Infect. Dis.* **19**, 872–879 (2019).
9. O’Toole, Á. et al. APOBEC3 deaminase editing in mpox virus as evidence for sustained human transmission since at least 2016. *Science* **382**, 595–600 (2023).
10. Gessain, A., Nakoune, E. & Yazdanpanah, Y. Monkeypox. *N. Engl. J. Med.* **387**, 1783–1793 (2022).
11. Laurenson-Schafer, H. et al. Description of the first global outbreak of mpox: an analysis of global surveillance data. *Lancet Glob. Health* **11**, e1012–e1023 (2023).
12. World Health Organization Regional Office for Africa. *Mpox in the WHO African Region, Regional Mpox Bulletin* (WHO, 2024).
13. Kibungu, E. M. et al. Clade I-associated mpox cases associated with sexual contact, the Democratic Republic of the Congo. *Emerg. Infect. Dis.* **30**, 172–176 (2024).
14. Masirika, L. M. et al. Ongoing mpox outbreak in Kamituga, South Kivu province, associated with monkeypox virus of a novel clade I sub-lineage, Democratic Republic of the Congo, 2024. *Euro Surveill.* **29**, 2400106 (2024).
15. Vakaniaki, E. H. et al. Sustained human outbreak of a new MPXV clade I lineage in the eastern Democratic Republic of the Congo. *Nat. Med.* **30**, 2791–2795 (2024).
16. Isidro, J. et al. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. *Nat. Med.* **28**, 1569–1572 (2022).
17. Gigante, C. M. et al. Multiple lineages of monkeypox virus detected in the United States, 2021–2022. *Science* **378**, 560–565 (2022).
18. Ndodo, N. et al. Distinct monkeypox virus lineages co-circulating in humans before 2022. *Nat. Med.* **29**, 2317–2324 (2023).
19. Brinkmann, A. et al. Extensive ITR expansion of the 2022 Mpox virus genome through gene duplication and gene loss. *Virus Genes* **59**, 532–540 (2023).
20. Esteban, D. J. & Hutchinson, A. P. Genes in the terminal regions of orthopoxvirus genomes experience adaptive molecular evolution. *BMC Genomics* **12**, 261 (2011).
21. Hughes, A. L. & Friedman, R. Poxvirus genome evolution by gene gain and loss. *Mol. Phylogenet. Evol.* **35**, 186–195 (2005).
22. von Magnus, P., Andersen, E. K., Petersen, K. B. & Birch-Andersen, A. A pox-like disease in cynomolgus monkeys. *Acta Pathol. Microbiol. Scand.* **46**, 156–176 (1959).
23. Ladnyj, I. D., Ziegler, P. & Kima, E. A human infection caused by monkeypox virus in Basankusu Territory, Democratic Republic of the Congo. *Bull. World Health Organ.* **46**, 593–597 (1972).
24. Kraemer, M. U. G. et al. Tracking the 2022 monkeypox outbreak with epidemiological data in real-time. *Lancet Infect. Dis.* **22**, 941–942 (2022).
25. Brien, S. C. et al. Clinical manifestations of an outbreak of monkeypox virus in captive chimpanzees in Cameroon, 2016. *J. Infect. Dis.* **229**, S275–S284 (2024).
26. Falendysz, E. A. et al. Characterization of monkeypox virus infection in African rope squirrels (*Funisciurus* sp.). *PLoS Negl. Trop. Dis.* **11**, e0005809 (2017).
27. Centers for Disease Control and Prevention (CDC). Update: multistate outbreak of monkeypox—Illinois, Indiana, Kansas, Missouri, Ohio, and Wisconsin, 2003. *MMWR Morb. Mortal. Wkly. Rep.* **52**, 616–618 (2003).
28. Zou, X. et al. A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nat. Cancer* **2**, 643–657 (2021).
29. Ruis, C., Tonkin-Hill, G., Floto, R. A. & Parkhill, J. Calculating and applying pathogen mutational spectra using MutTui. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.06.15.545111> (2023).
30. Izzoddeen, A., Elbadri, O., Nageeb Abdalla, M., Magbol, M. & Osman, M. Monkeypox pandemic in Sudan, surveillance epidemiologic report, 2022. *BMC Public Health* **24**, 2457 (2024).
31. World Health Organization. *Multi-country Outbreak of Monkeypox External Situation Report #8* (WHO, 2022).
32. Murhula, L. et al. Mapping and sequencing of cases from an ongoing outbreak of clade Ib monkeypox virus in South Kivu, Eastern Democratic Republic of the Congo between September 2023 to June 2024. Preprint at *medRxiv* <https://doi.org/10.1101/2024.09.18.24313835> (2024).
33. World Health Organization Mpox—Democratic Republic of the Congo. *Disease Outbreak News* (17 June 2024).
34. Fenner, F., Henderson, D. A., Arita, I., Jezek, Z. & Ladnyj, I. D. *Smallpox and Its Eradication* (World Health Organization, 1988).
35. Mukadi-Bamuleka, D. et al. First imported cases of MPXV clade Ib in Goma, Democratic Republic of the Congo: implications for global surveillance and transmission dynamics. Preprint at *medRxiv* <https://doi.org/10.1101/2024.09.12.24313188> (2024).
36. World Health Organization. *Multi-country Outbreak of Mpox, External Situation Report #35* (WHO, 2024).
37. Shete, A. M. et al. Genome characterization of monkeypox cases detected in India: identification of three sub clusters among A.2 lineage. *J. Infect.* **86**, 66–117 (2023).
38. Nik-Zainal, S. et al. Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
39. Ulaeto, D. O., Dunning, J. & Carroll, M. W. Evolutionary implications of human transmission of monkeypox: the importance of sequencing multiple lesions. *Lancet Microbe* **3**, e639–e640 (2022).
40. Hudson, P. N. et al. Elucidating the role of the complement control protein in monkeypox pathogenicity. *PLoS ONE* **7**, e35086 (2012).
41. Hutson, C. L. et al. Comparison of monkeypox virus clade kinetics and pathology within the prairie dog animal model using a serial sacrifice study design. *Biomed. Res. Int.* **2015**, 965710 (2015).
42. Moss, B. & Shisler, J. L. Immunology 101 at poxvirus U: immune evasion genes. *Semin. Immunol.* **13**, 59–66 (2001).
43. Alcamí, A. & Koszinowski, U. H. Viral mechanisms of immune evasion. *Trends Microbiol.* **8**, 410–418 (2000).
44. Garrigues, J. M. et al. Identification of human monkeypox virus genome deletions that impact diagnostic assays. *J. Clin. Microbiol.* **60**, e0165522 (2022).
45. World Health Organization. *Standing Recommendations for Mpox Issued by the Director-General of the World Health Organization (WHO) in Accordance with the International Health Regulations (2005) (IHR)* (WHO, 2023).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in

the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

¹World Health Organization, Geneva, Switzerland. ²Victor Phillip Dahdaleh Heart and Lung Research Institute, University of Cambridge, Cambridge, UK. ³Cambridge Centre for AI in Medicine, University of Cambridge, Cambridge, UK. ⁴Department of Veterinary Medicine, University of Cambridge, Cambridge, UK. ⁵Department of Medicine, University of Texas Southwestern, Dallas, TX, USA. ⁶Peter O'Donnell Jr School of Public Health, University of Texas Southwestern, Dallas, TX, USA. ⁷ZBS—Highly Pathogenic Viruses, German Consultant Laboratory for Poxviruses, WHO Collaborating Centre for Emerging Infections and Biological Threats, Robert Koch Institute, Berlin, Germany. ⁸National Institute for Biomedical Research, Kinshasa, Democratic Republic of the Congo. ⁹University of Kinshasa, Kinshasa, Democratic Republic of the Congo. ¹⁰National Institute for Biomedical Research, Goma, Democratic Republic of the Congo. ¹¹Health Emergencies and Epidemics Control, Federal Ministry of Health, Khartoum, Republic of the Sudan. ¹²World Health Organization Country Office, Port Sudan, Republic of the Sudan. ¹³World Health Organization Country Office, Kinshasa, Democratic Republic of the Congo. ¹⁴World Health Organization Regional Office for Africa, Brazzaville, Republic of the Congo. ¹⁵World Health Organization Regional Office for the Americas, Washington DC, USA. ¹⁶World Health Organization Regional Office for the Eastern Mediterranean, Cairo, Egypt. ¹⁷World Health Organization Regional Office for the Western Pacific, Manila, Philippines. ¹⁸World Health Organization Regional Office for Europe, Copenhagen, Denmark. ¹⁹World Health Organization Regional Office for South-East Asia, New Delhi, India. ²⁰Cambridge Centre for Lung Infection, Papworth Hospital, Cambridge, UK. ²¹These authors contributed equally: James R. Otieno, Christopher Ruis. ✉ e-mail: subissil@who.int

Methods

Dataset assembly and filtering

We aimed to collate a dataset containing all available high-quality MPXV genetic sequences. To do this, we initially downloaded all MPXV nucleotide sequences from GenBank (identified as sequences containing at least one of the search terms ‘monkeypox’, ‘mpox’, ‘MPXV’ and ‘MPV’, $n = 7,252$) and the GISAID EpiPox database ($n = 8,843$) as of 17 February 2024. We then combined these datasets and filtered out duplicate and lower-quality sequences. To do this, we initially removed sequences containing fewer than 30,000 nucleotides (nt), with this cutoff chosen to retain historical MPXV sequences that were ~32,000 nt in length. We next discarded sequences with >20% indeterminate bases (Ns) and assigned clades and (for clade IIb) Pango lineages using Nextclade⁴⁶. Sequences that could not be assigned a clade were assumed to be low quality and were excluded from further analysis. Based on the clade and lineage assignments from Nextclade, we divided the sequence dataset into four groups: clade I, clade IIa, clade IIb A (containing sequences from the A sublineages within clade IIb but not those within lineage B.1 and its descendent lineages) and lineage B.1 (containing sequences from lineage B.1 and its descendent lineages).

We identified sequences that were duplicated between GenBank and GISAID initially by identifying sequences with the same sample name, country and collection date. After removing one of each of these sequence pairs, we carried out an additional phylogenetic screen for duplicate sequences in the clade I, clade IIa and clade IIb A datasets. Sequences within each of the four datasets were aligned using squirrel v0.1 (<https://github.com/aineniambh/squirrel>)⁹, specifying the clade to which the sequences belong. We then reconstructed a phylogenetic tree for each clade using IQ-TREE v2.1.3 (ref. 47), using a Jukes-Cantor (JC) model of nucleotide substitution. We identified closely related pairs of sequences in the resulting phylogenetic trees and checked their sequence names, countries and collection dates to determine whether they might be duplicates. Where the sequence names and collection dates were similar (that is, the sequence names contained shared elements and the collection dates were the same to the most accurate level possible), we retained one of the sequences.

We carried out a further quality control check by analyzing the root-to-tip distances of sequences compared with their collection date⁴⁸. For clade I, clade IIa and clade IIb A, we aligned sequences with squirrel v0.1 and reconstructed maximum likelihood phylogenetic trees using IQ-TREE v2.1.3 as above but including an outlier sequence (outlier accession numbers [KJ642617.1](#) for clade I, [KJ642616.1](#) for clade IIa and clade IIb A) that was used to root the tree. We then identified sequences that were clear outliers in a root-to-tip plot of the rooted tree in TempEst v1.5.3 (ref. 48) and removed these from further analyses. Owing to the large size of the lineage B.1 dataset ($n = 10,369$ sequences before filtering), instead of reconstructing a phylogenetic tree, we carried out a root-to-tip-like analysis by comparing the collection date with the mutation distance to sample [ON676708.1](#), which clusters immediately upstream of lineage B.1. We aligned the lineage B.1 dataset and [ON676708.1](#) using squirrel v0.1 as above and calculated the number of mutations between each sequence and [ON676708.1](#). This showed a strong correlation (Extended Data Fig. 5), so we generated a linear model between sample collection date and this mutation distance and removed samples whose residual within the linear model is more than five times the median absolute deviation away from the median residual (Extended Data Fig. 5). These filtering steps resulted in final datasets of 113 clade I sequences, 25 clade IIa sequences, 101 clade IIb A sequences and 10,307 lineage B.1 sequences. The majority of these sequences contain close to the complete genome (Extended Data Fig. 6 and Supplementary Tables 1 and 2). In addition, we have added 47 sequences from a GitHub directory of a recent paper that describes highly divergent viruses¹⁵.

We identified the spatiotemporal and host species distributions of sequences within each dataset using location, collection data and

host species metadata associated with the sequence accession on either GenBank or GISAID. Where these metadata were missing, we attempted to identify them within original publications.

Reconstruction of phylogenetic trees

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

To examine phylogenetic relationships within each clade, we aligned sequences within the respective dataset using squirrel v0.1 with clade set to clade I for the clade I dataset and set to clade II for the remaining datasets. Final maximum likelihood phylogenetic trees were reconstructed for each dataset using IQ-TREE v2.1.3 as above. Topological robustness was assessed using 1,000 bootstrap replicates. Travel histories for clade IIb A were identified in GISAID metadata and from examination of original publications.

Phylogenetic trees were visualized using FigTree v1.4.4 and ggtree v3.0.2 (ref. 49).

Reconstruction of the temporal history of clade I

We aimed to reconstruct the temporal history of clade I. The recently identified outbreak in South Kivu contains evidence of APOBEC-3 mutagenesis¹⁵, which increases the substitution rate⁹ and may make the application of a single clock model unreliable. We therefore did not include the sequences from this outbreak in the temporal reconstruction.

Methods to infer temporal history are only valid if there is a temporal signal within the dataset⁴⁸. We assessed temporal signal using root-to-tip randomization in which we compared the R2 correlation between the sample collection date and root-to-tip distance with that in 1,000 randomizations of collection dates. This supported the presence of a temporal signal ($P < 0.001$). We therefore reconstructed the temporal history of clade I using BEAST v2.6.6 using the JC69 model of nucleotide substitution. We used a relaxed log-normal clock model with a log-normal prior on the substitution rate with mean 1.9×10^{-6} (chosen to match the estimated slope in TempEst) and standard deviation 0.5. Population history was modeled using a coalescent constant population prior. Four independent runs were carried out for 150,000,000 total MCMC steps. Convergence was assessed using Tracer v1.7 (refs. 49,50) and all effective sample size values were above 550. From each run, 10% burnin was removed before the runs were combined and the maximum clade credibility tree was identified and annotated using TreeAnnotator.

Identification of genome rearrangements

Genome rearrangements in the Sudan clade I sequences were identified as described previously¹⁹.

Calculation of mutational spectra

We calculated an SBS mutational spectrum for each of the major MPXV clades using the sequence alignments and maximum likelihood phylogenetic trees generated above and containing all high-quality sequences. Phylogenetic trees were outgroup rooted using the outgroups described above to enable the direction of each mutation (that is, the ancestral and mutated nucleotides) to be robustly identified. The outgroup was removed before spectrum calculation. We reconstructed mutational spectra using MutTui v2.0.2 (<https://github.com/chrisruis/MutTui>)²⁹. We rescaled the resulting mutational spectra using MutTui v2.0.2 to account for the number of A, C, G and T nucleotides and the distribution of nucleotide triplets across the genome²⁹.

It has previously been suggested that a dataset requires at least 300–600 mutations for the mutational spectrum to be accurately estimated²⁹. The clade I, clade IIb A and lineage B.1 datasets each contain more than 600 mutations. However, the clade IIa spectrum contains 222 mutations so we did not attempt to examine the detailed contextual patterns in these mutations. To estimate whether the clade IIa mutational spectrum could have been generated by the spectrum of clade I or clade IIb, we compared the mutation type proportions in the clade IIa spectrum with those in 1,000 random downsamplings of the clade I and lineage B.1 mutational spectra to 222 mutations.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Sequence alignments (excluding GISAID sequences), BEAST log files and mutational spectra are available via GitHub at https://github.com/chrisruis/Mpox_global. GenBank sample accessions and metadata are listed in Supplementary Table 1. GISAID sample accessions and acknowledgements are listed in Supplementary Table 2.

Code availability

The code for calculating mutational spectra is available via GitHub at <https://github.com/chrisruis/MutTui>.

References

46. Aksamentov, I., Roemer, C., Hodcroft, E. & Neher, R. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *J. Open Source Softw.* **6**, 3773 (2021).
47. Minh, B. Q. et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
48. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
49. Xu, S. et al. Ggtree: a serialized data object for visualization of a phylogenetic tree and annotation data. *iMeta* **1**, e56 (2022).
50. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).

Acknowledgements

We thank all data contributors, that is, the authors and their originating laboratories responsible for obtaining the specimens, and their submitting laboratories for generating and sharing the genetic sequence and metadata on which this study is based. We thank Y. Izawa (WHO) for generating the maps. C.R. and R.A.F. were supported by a Fondation Botnar Research Award (program grant number 6063 to R.A.F.), the UK Cystic Fibrosis Trust (Innovation Hub Award 001 to R.A.F.) and The Wellcome Trust (Investigator award 107032/Z/15/Z to R.A.F.). The findings and conclusions of this report are those of the author(s) and do not necessarily represent the views of the funding agencies.

Author contributions

J.R.O. and L.S. conceived and planned the study. J.R.O., C.R., A.B.O., K.K., A.H., A.N., A. Brinkmann, J.M., P.M.-K., D.M.-B., M.M.O., H.H., M.A.R., R.F., B.L.H., M.K., J.M.R., L.G., A. Barakat, V.K., K.N., D.N., R.A.F., J.C., M.D.V.K., R.F.L. and L.S. analyzed, processed, visualized and interpreted the data. J.R.O., C.R. and L.S. wrote the first draft of the paper. All authors contributed in the editing and interpretation of the results.

Competing interests

The authors declare no competing interests.

Additional information

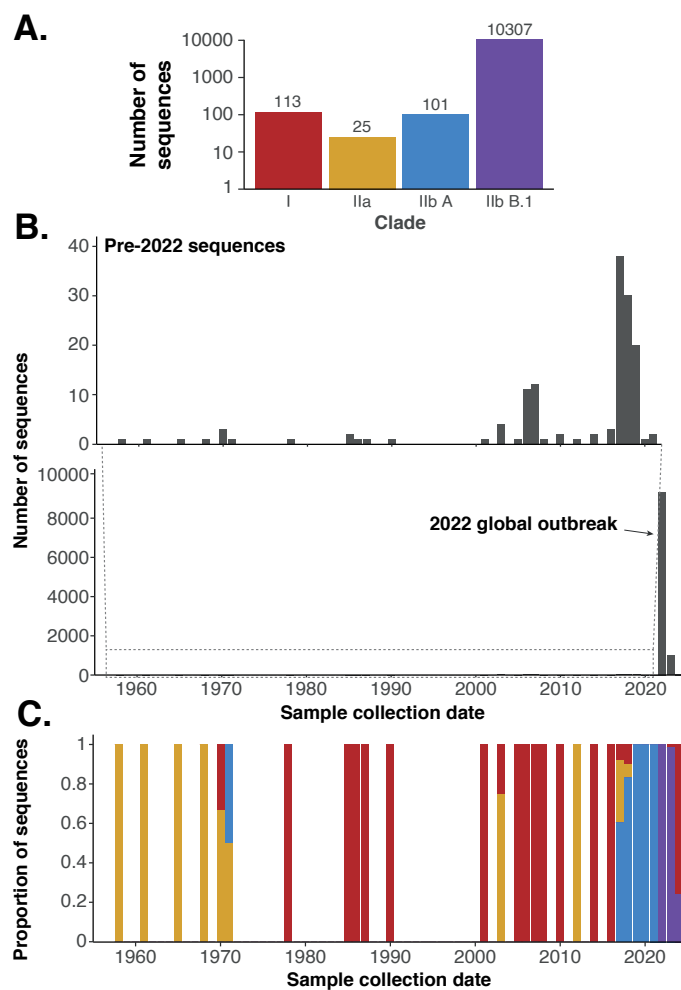
Extended data is available for this paper at <https://doi.org/10.1038/s41591-024-03370-3>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41591-024-03370-3>.

Correspondence and requests for materials should be addressed to Lorenzo Subissi.

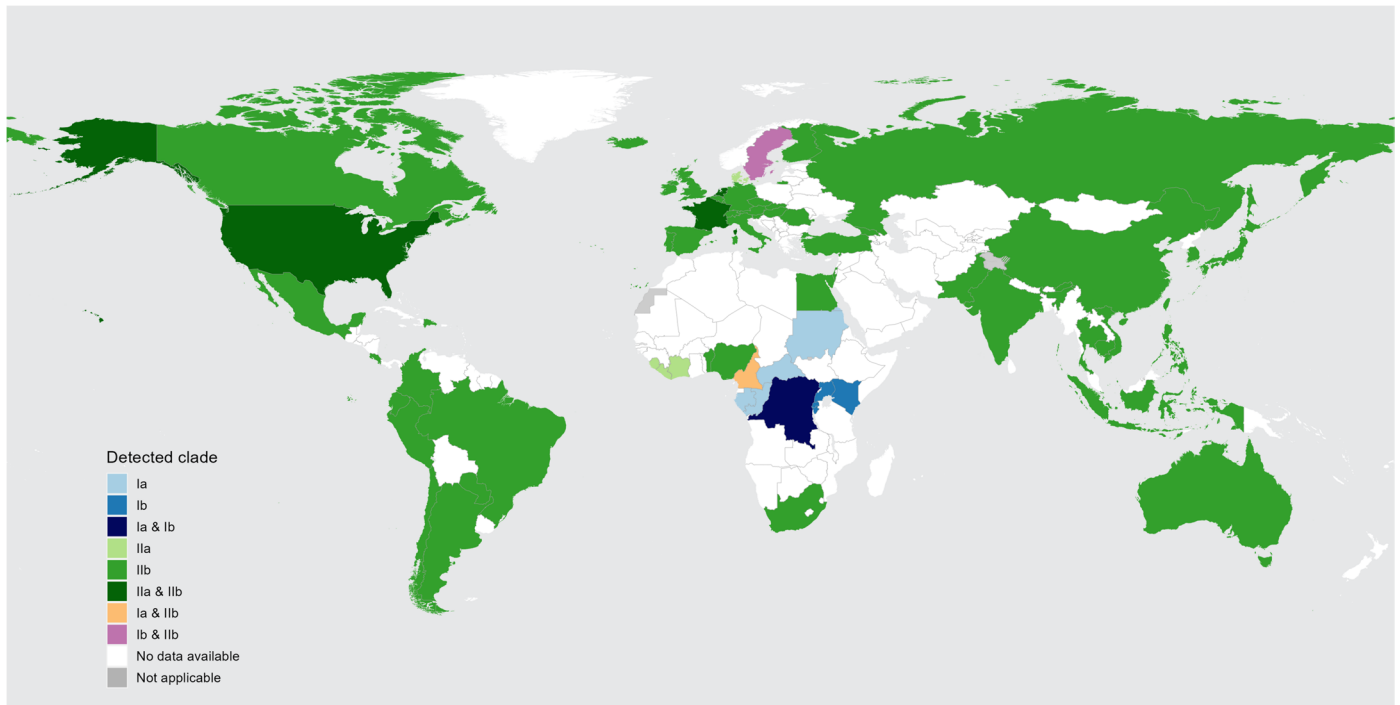
Peer review information *Nature Medicine* thanks David Ulaeto and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Alison Farrell and Joao Monteiro, in collaboration with the *Nature Medicine* team.

Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | 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

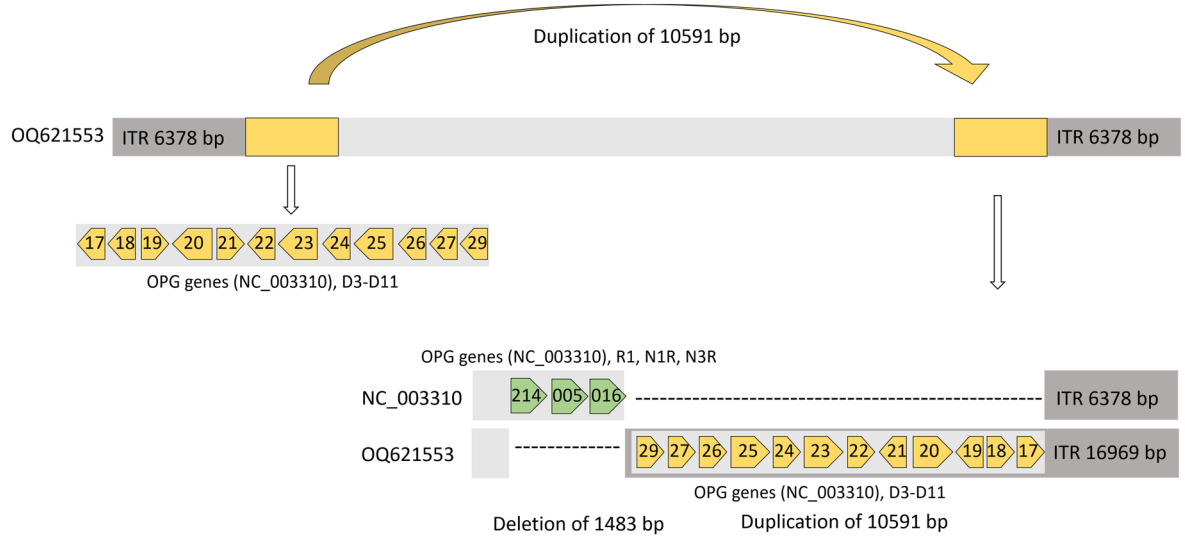
the large number of sequences in 2022 compared to other years, we extract the years before 2022 in the upper panel. (c) The proportion of sequences belonging to each major MPXV clade in each year is shown, colours match those in panel A.



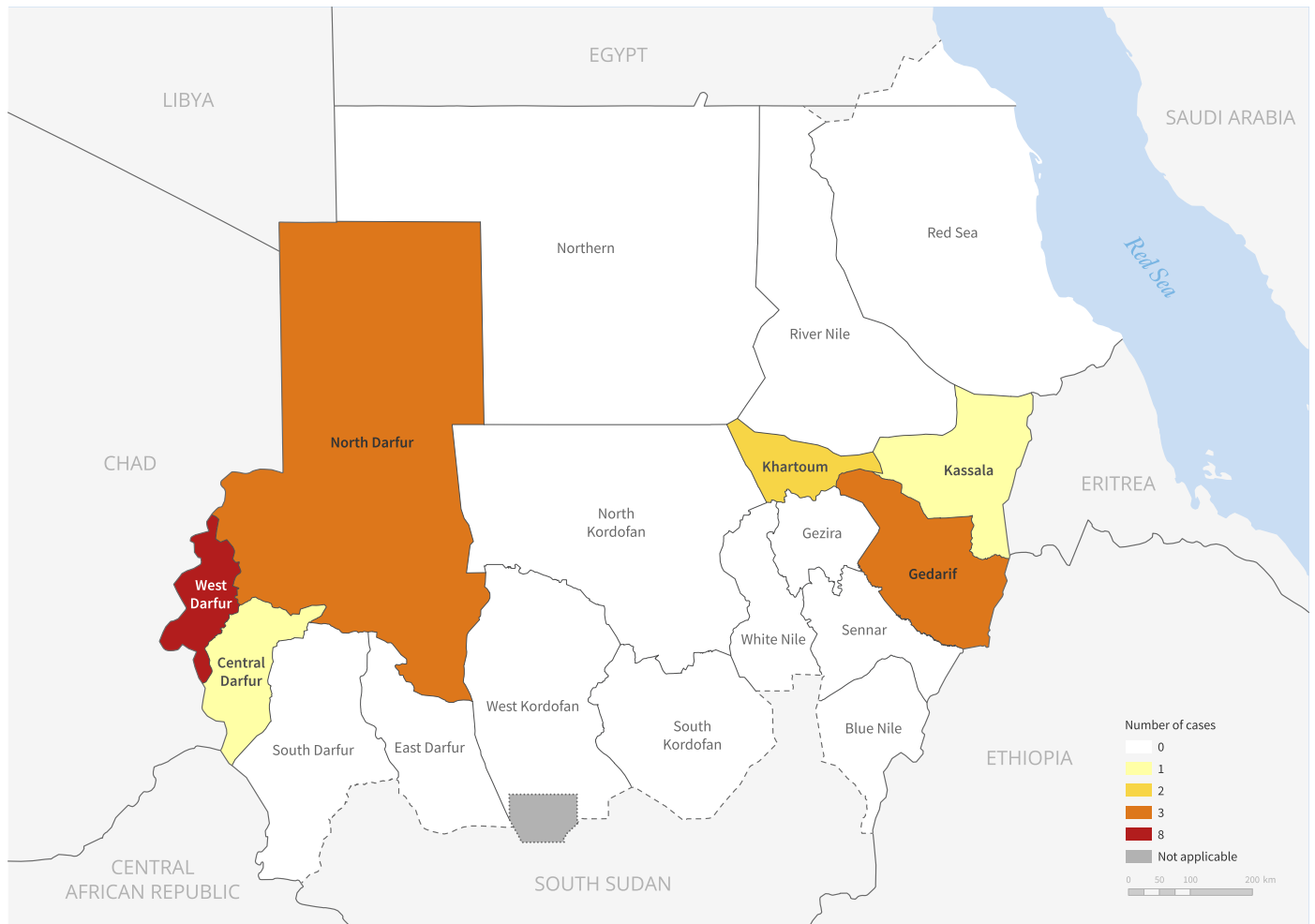
The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

Data Source: World Health Organization, GISAID and GenBank
Map Production: WHO Health Emergencies Programme
© WHO 2024. All rights reserved.

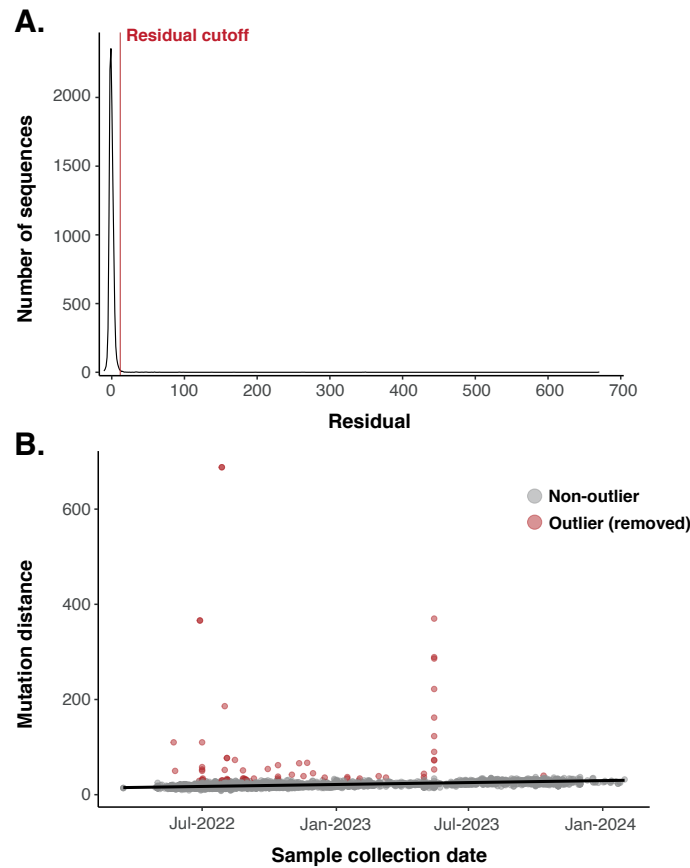
Extended Data Fig. 2 | Geographic distribution of MPXV clades across WHO regions. The map was generated from 10,670 sequences obtained from GenBank and GISAID, 1958- February 2024. Clade I is red, clade IIa is green, and clade IIb is blue.



Extended Data Fig. 3 | Duplication of the genome from Sudan 2005 and 2022 (KC257459, OQ621553). A 10591 bp region directly downstream the left ITR is duplicated to the right site of the genome, resulting in ITRs of 16969 bp and a genome length of 204,808 bp). Three genes (R1, N1R and N3R) are deleted at the site of duplication.

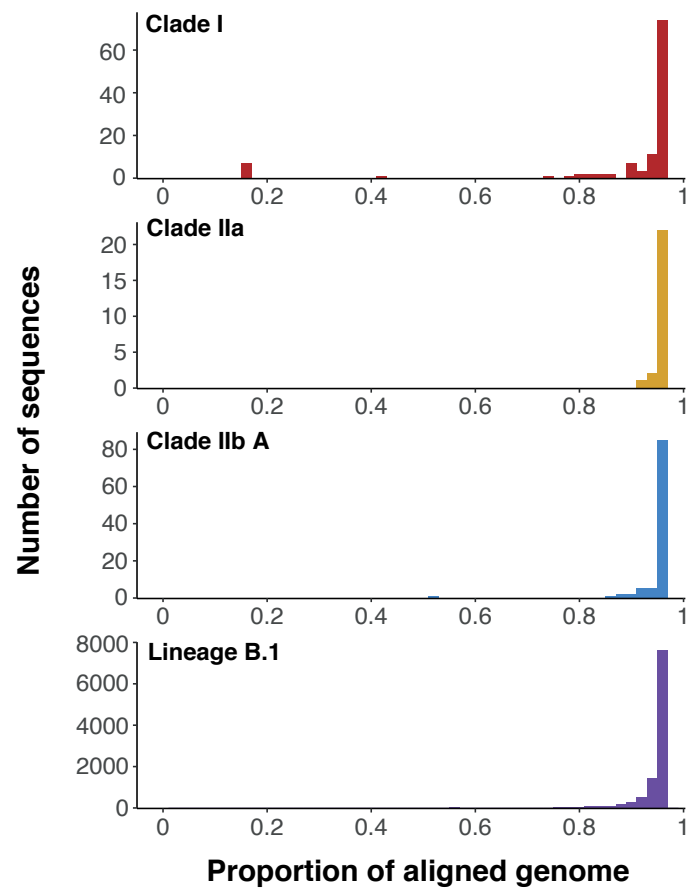


Extended Data Fig. 4 | Geographical distribution of laboratory confirmed cases of mpox in Sudan in 2022. Confirmed cases were reported in Sudan states neighbouring Chad, Central African Republic, Ethiopia and Eritrea, while cases in 2005 were reported from a region that now belong to South Sudan.



Extended Data Fig. 5 | 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.



Extended Data Fig. 6 | 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 alignment⁹, no sequence will contain the complete genome.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection No software was used for data collection. Genetic sequences were downloaded from GISAID and GenBank databases.

Data analysis MPXV clades and lineages were assigned using nextclade. Sequences within each dataset were aligned using squirrel v0.1. Phylogenetic trees were reconstructed using IQ-TREE v2.1.3 employing a JC model of nucleotide substitution. Root-to-tip identification of outliers was carried out using TempEst v1.5.3. Phylogenetic trees were visualised using FigTree v1.4.4 and ggtree v3.0.2. Reconstruction of the clade I temporal tree was carried out with BEAST v2.6.6. Convergence was assessed using Tracer v1.7. The maximum clade credibility tree was identified using TreeAnnotator v2.6.4. Mutational spectra were calculated and rescaled using MutTui v2.0.2.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Sequence alignments (excluding GISAID sequences), BEAST log files and mutational spectra are available at https://github.com/chrisruis/Mpox_global. GenBank sample accessions and metadata are listed in Supplementary Table 1. GISAID sample accessions and acknowledgements are listed in Supplementary Table 2.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Not applicable
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was undertaken. All available monkeypox virus sequences were downloaded from GenBank and GISAID.
Data exclusions	First, sequences shorter than 30,000 nucleotides (nt) were excluded and this minimum sequence length was based on the observation that the majority of MPXV sequences prior to the 2022 outbreak were ~32,000 nt in length and assumed to have been sufficient to inform genomic epidemiology at the time. Secondly, we discarded sequences with >20% indeterminate bases (Ns), non-human sample sequences, and sequences known to fail Nextclade clade assignment. Thirdly, sequence duplicates were removed by country and collection date.
Replication	We used all available genetic sequences from each clade.
Randomization	Sequences were divided into clades based on phylogenetic relationships.
Blinding	All samples were coded by unique lab identification and therefore all associated metadata are anonymised.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

- n/a Involved in the study
- Antibodies
- Eukaryotic cell lines
- Palaeontology and archaeology
- Animals and other organisms
- Clinical data
- Dual use research of concern
- Plants

- n/a Involved in the study
- ChIP-seq
- Flow cytometry
- MRI-based neuroimaging

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.