Title

APOBEC3 deaminase editing in mpox virus as evidence for sustained human transmission since at least 2016.

Journal

Science, 382(6670)

Authors

OToole, Áine
Neher, Richard
Ndodo, Nnaemeka
et al.

Publication Date

2023-11-03

DOI

10.1126/science.adg8116

Peer reviewed

HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2024 February 21.

Published in final edited form as:

Science. 2023 November 03; 382(6670): 595–600. doi:10.1126/science.adg8116.

APOBEC3 deaminase editing in mpox virus as evidence for sustained human transmission since at least 2016

Áine O'Toole[1,*], **Richard A. Neher**[2], **Nnaemeka Ndodo**[3], **Vitor Borges**[4], **Ben Gannon**[5], **João Paulo Gomes**[4,6], **Natalie Groves**[7], **David J King**[8], **Daniel Maloney**[1], **Philippe Lemey**[9], **Kuiama Lewandowski**[5], **Nicholas Loman**[7,10], **Richard Myers**[7], **Ifeanyi F. Omah**[1,11], **Marc A. Suchard**[12], **Michael Worobey**[13], **Meera Chand**[7], **Chikwe Ihekweazu**[3], **David Ulaeto**[7,†], **Ifedayo Adetifa**[3,†], **Andrew Rambaut**[1,*,†]

[1]Institute of Ecology & Evolution, University of Edinburgh; Edinburgh, EH9 3FL, United Kingdom.

[2]Biozentrum, University of Basel and Swiss Institute of Bioinformatics; Basel, Switzerland.

[3]Nigeria Centers for Disease Control and Prevention; Abuja, Nigeria.

[4]National Institute of Health Doutor Ricardo Jorge (INSA); Lisbon, Portugal.

[5]UK Health Security Agency, Porton Down; Salisbury, United Kingdom.

[6]Veterinary and Animal Research Centre (CECAV), Faculty of Veterinary Medicine, Lusófona University, Lisbon, Portugal.

[7]UK Health Security Agency; London, E14 5EA, United Kingdom.

[8]CBR Division, Defence Science and Technology Laboratory; Salisbury SP4 0JQ, United Kingdom.

[9]Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven; Leuven, Belgium.

[10]University of Birmingham; Birmingham, United Kingdom.

[11]Department of Parasitology and Entomology, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.

[12]Department of Biostatistics, Fielding School of Public Health, University of California; Los Angeles, California, USA.

[13]Department of Ecology and Evolutionary Biology, University of Arizona; Tucson, Arizona, USA.

[*]Corresponding authors. aine.otoole@ed.ac.uk, a.rambaut@ed.ac.uk.
[†]These authors are joint-senior

Author contributions:
Conceptualization: ÁOT, AR
Methodology: AOT, RN, AR
Investigation: NN, VB, BG, JPG, NG, DK, DM, KL, RM, IFO, MW
Visualization: ÁOT, RN, AR
Funding acquisition: AR, PL, MS, MC, CI, DU, IA
Supervision: DU, IA, AR
Writing – original draft: ÁOT, RN, PL, MS, MW, AR
Writing – review & editing: ÁOT, RN, NN, VB, BG, JPG, NG, DK, DM, PL, KL, NL, RM, IFO, MS, MW, MC, CI, DU, IA, AR

Competing interests: Authors declare they have no competing interests.

Abstract

Historically, mpox has been characterised as an endemic zoonotic disease that transmits through contact with the reservoir rodent host in West and Central Africa. However, in May 2022, human cases of mpox were detected spreading internationally beyond countries with known endemic reservoirs. When the first cases from 2022 were sequenced, they shared 42 nucleotide differences from the closest mpox virus (MPXV) previously sampled. Nearly all these mutations are characteristic of the action of APOBEC3 deaminases; host-enzymes with antiviral function. Assuming APOBEC3-editing is characteristic of human MPXV infection, we develop a dual process phylogenetic molecular clock that — inferring a rate of ~6 APOBEC3 mutations per year — estimates MPXV has been circulating in humans since 2016. These observations of sustained MPXV transmission present a fundamental shift to the perceived paradigm of MPXV epidemiology as a zoonosis and highlight the need for revising public health messaging around MPXV as well as outbreak management and control.

One-Sentence Summary:

Using signatures of editing by APOBEC3 deaminase, an anti-viral host enzyme, we estimate that mpox virus (MPXV) has been transmitting in the human population since 2016, which presents a fundamental shift to the paradigm of MPXV epidemiology.

Since 2017, the Nigeria Centre for Disease Control has been reporting cases of MPXV (mpox virus) infection in humans (Fig. S1)(1). MPXV, a DNA virus in the genus *Orthopoxvirus*; Family *Poxviridae*, is often described as being endemic in West and Central Africa as a zoonotic disease that transmits through contact with the rodent reservoir host. Since the first human cases were observed in the 1970s, MPXV infections have been predominantly associated with infants and children (2–4). However, of the cases observed in Nigeria since 2017 very few have been in children, with the virus mainly affecting adults aged 20 to 50 (79%), 27% of which were in women (5). Genome sequencing of viruses revealed enough genetic diversity between cases that distinct zoonotic events were not ruled out. In May 2022, cases of MPXV infection were detected spreading widely across Europe and subsequently across the globe. The first MPXV genome sequences from these 2022 cases showed they had descended from the clade characterised by cases diagnosed in Nigeria and Israel, Singapore, and the UK (6) associated with travel from Nigeria (Fig. S2; Table S1, in bold). These early 2022 genomes are indicated as a triangle within Clade IIb in Fig. 1A and represent lineage B.1 as per the nomenclature proposed by Happi et al. (7). Isidro et al., (6) noticed that sequences within lineage B.1 shared 42 single nucleotide differences from the closest earlier MPXV genomes from 2018. From a 2017 outbreak of MPXV in chimpanzees, the evolutionary rate of MPXV was estimated to be 1.9×10^{-6} substitutions per site per year ($1.2 – 2.7\times10^{-6}$) translating to ~1 nucleotide change every 3 years (8). 42 substitutions in the space of 3–4 years is an unexpectedly large number.

Under the paradigm that MPXV is a zoonotic virus with limited human to human transmission, one interpretation of this long branch might be that it represents adaptation to humans facilitating the sustained transmission that is now observed. However, as we show here, and as was quickly seen when the first genomes from 2022 were sequenced,

it is clear that these mutations are not the result of errors by the virus's replication machinery and occur at much higher rate than would be expected for an orthopoxvirus (6, 9). Specifically, the majority of observed nucleotide changes appear to be of a particular type – a dinucleotide change from TC→TT or its reverse complement, GA→AA (9, 10). This particular mutation is characteristic of the action of the APOBEC3 (Apolipoprotein B mRNA editing enzyme, catalytic polypeptide 3) family of cytosine deaminases. These act on single stranded DNA (ssDNA) to deaminate cytosine to uracil causing a G→A mutation in the complementary strand when it is synthesised. Most human APOBEC3 molecules have a strong bias towards deaminating 5'TC dinucleotides and APOBEC3-driven deamination has been demonstrated with many DNA viruses and retroviruses; (11–18). Furthermore, a recent study has specifically demonstrated APOBEC3F editing in cell culture and during human MPXV infection (19).

We assess the extent to which APOBEC3 has acted on MPXV and explore whether this is the source of the elevated mutation rate observed since 2017. We also explore the evolutionary consequences of this mechanism driving evolution in MPXV and model the distinct processes underpinning the evolution of MPXV in the human population.

APOBEC3 editing as a signature of MPXV evolution in the human population

The known diversity of MPXV is decomposed into three major clades; Clades I, IIa and IIb (Fig. 1A; Happi et al., (7)). Clade I represents MPXV sampled in Central Africa and Clade IIa is composed of viruses from human and non-human animal samples taken in or connected to West Africa. Both of these clades include virus genomes spanning from the 1970s to present day, although the majority of samples were collected within the last 20 years (Fig. S3). Clade IIb has an early sample taken in 1971 (Genbank accession KJ642617) but most of the sequences in Clade IIb are more recent virus genomes from 2017–2022 that (7) have labelled as hMPXV-1 (Fig. 1A, labelled phylogeny Fig. S2). Within the recent diversity of Clade IIb (indicated as the darker box within IIb in Fig. 1A), we catalogued transmitted mutations that occurred between 2017 and 2022 with only a single representative from the 2022 global lineage B.1 (n=44, Fig. S2)(20).

Within MPXV Clade IIb, we observe rates of molecular evolution far greater than that expected for double-stranded DNA viruses and indeed that observed in Clades I and IIa of MPXV (8) and see this excess accumulation of mutation in samples as early as 2017. The great majority of these mutations are of the type G→A or C→T (90.8%) (Fig. 1B), regardless of sample host species (Fig. S4). Comparing MPXV Clade IIb with Clade I and IIa emphasises that this pattern is not seen outside of Clade IIb (Fig. 1B, labelled phylogenies Fig. S5), nor is it seen when looking at reconstructed mutations within a phylogeny of 46 Variola Virus (VARV) genomes, the human virus responsible for smallpox (Fig. S6). For the other MPXV clades, APOBEC-type mutations are observed at between 8 and 13% frequency, which fits with the expected proportion under standard models of nucleotide evolution (21, 22) (Fig. 1B). Strikingly, the heptamers of C→T and G→A mutations that occurred across the Clade IIb phylogeny show this is a specific enrichment

of APOBEC3-type dimer mutations of the type TC→TT and GA→AA (Fig. 1C). Similarly, this enrichment of TC→TT and GA→AA mutations is observed within the B.1 lineage (Fig. S7), where 84.8% of observed SNPs are consistent with APOBEC3 editing (Fig. S8). Observed heptamers around the observed C→T and G→A mutations show a striking enrichment in TC and GA target sites in the genomes sampled from 2017–2022 in contrast with the rest of MPXV diversity (Fig. 1C), and this enrichment is also reflected within lineage B.1 (Fig. S9).

Our analysis highlights that evolution within Clade IIb prior to the emergence of lineage B.1 mirrors that within lineage B.1, but is distinct from MPXV Clade I or IIa. Since 2022, the B.1 lineage has been sampled and sequenced internationally in the global epidemic of MPXV. Lineage B.1 is known to be circulating by sustained human-to-human transmission and as such, mutations that have accumulated in B.1 can be considered characteristic of this mode of transmission. We suggest that the APOBEC3-driven evolution of recent Clade IIb MPXV is a signature of a switch to sustained transmission within the human population. Within the B.1 lineage, believed to be entirely the result of human infection and transmission, we continue to see the same pattern of predominantly APOBEC3 mutations accumulating at a similar rate to that seen in A lineage genomes since 2017. It is unlikely that, by chance, MPXV evolved to become susceptible to APOBEC3 action within the putative rodent reservoir prior to the emergence of cases and to retain that susceptibility to human APOBEC3 molecules once transmitting in humans. Given that all human cases sequenced since 2017 share substantial numbers of APOBEC3 mutations, including nine on the stem branch leading to hMPXV-1 it is very unlikely these represent multiple zoonotic introductions. APOBEC3 genes emerged in placental mammals from a duplication of the ancestral AID gene and have a dynamic recent evolutionary past, with gene duplication and loss across phyla (23, 24). APOBEC3 genes in primates have undergone recent expansion, with primate genomes now having 7 paralogs of 3 ancestral genes (25, 26). Rodents, the reservoir of MPXV, have only a single functional APOBEC3 protein resulting from gene loss and fusion events (25). Rodent APOBEC3 has been shown to be expressed preferentially in spleen and bone marrow with limited expression observed in other tissues (27, 28).

APOBEC3 has a limited repertoire to generate variation in the MPXV genome

If we assume this observed evolution within hMPXV-1 is APOBEC3-driven, this may have implications for its sustained transmission in the human population. Considering all GA and TC dimer sites in the Clade II reference genome (Accession number NC_063383) – i.e. those that could be the target of APOBEC3 editing but had not been by that point – we assessed what amino acid changes a deamination mutation at these sites would bring about (Fig. 2). Of the 23,718 such dimers, 61.6% (14,618) would produce amino acid replacements, 21% (4990) would be synonymous, 2.9% (692) would induce stop codons, and 14.4% (3418) would occur outside of coding regions. For the Clade IIb genomes, of the 633 mutations at these dimers that did occur, 38.7% (245) were amino acid replacements and 35.7% (226) were synonymous, 4.7% (30) were nonsense, and a further 132 APOBEC3

mutations were in intergenic regions (20.8%). The probability of getting 226 or greater synonymous mutations out of 663 under a simple binomial distribution with 21.0% chance of a context being synonymous is $P=7.6\times10^{-18}$. We do not see the same enrichment for synonymous mutations in the mutations that are not APOBEC3-like, although the quantity of these mutations is considerably lower (Fig. S10). There are also more mutations outside of protein coding regions than we would expect based on the location of target dimers (probability of 4.5×10^{-6} of getting at least 132 non-coding mutations given only 14.4% of targets are in these regions). This supports the hypothesis that what we are observing are the residual least harmful APOBEC3 mutations after natural selection has eliminated those with substantial fitness costs to the virus. By comparing the density of observed C→T and G→A mutations and the density of APOBEC3 target sites (TC or GA dinucleotides) across the reference genome, we see that the distribution of mutations is not simply a product of the availability of target sites (Kolmogorov-Smirnov test statistic=0.07, P-value=0.0002; Fig. S11A–B). When considering synonymous and non-synonymous APOBEC3-like mutations separately, there is a significant difference between the density of target sites across the MPXV genome and the distribution of observed APOBEC3-like synonymous and non-synonymous mutations respectively (Fig. S11C–D).

The 'repertoire' of mutations that APOBEC3 is able to provide as genetic variation on which natural selection can act is severely restricted. Only a limited number of dinucleotide contexts are present, and the repertoire of amino acid changes that APOBEC3 editing can induce is also limited (Fig. 2C, Fig. S12). Only 13 different amino acid replacements are possible, and three that give rise to stop codons, and they are not reversible by the same mechanism. This means that given the restricted set of positions at which these mutations occur and the limited amino acid changes they can result in, the elevated rate won't necessarily facilitate adaptation of the virus.

APOBEC3 hypermutation is a host-mediated antiviral mechanism. These molecules act as the viral genome is being replicated and single strands are exposed. During repeated rounds of replication either strand can be deaminated leading to both C→T and G→A changes on the positive strand as seen here. Thus, it is likely that the genomes that are extensively mutated by APOBEC3 will simply not be viable and will not be transmitted further. MPXV replicates in the cytoplasm likely by means of rolling-circle amplification (29) and this facilitates extensive continuous genome replication (30). The high processivity of this mechanism efficiently produces high copy numbers of MPXV genome molecules in the cell, potentially saturating the APOBEC3 enzyme action if the concentration of MPXV DNA molecules is high enough. This likely means that many MPXV genomes are unaffected by APOBEC3 action. Occasionally however a genome, modestly mutated by APOBEC3, may remain viable and be transmitted. We see this in the enrichment of observed synonymous and intergenic mutations relative to available targets in the MPXV genome. Given the non-reversibility of the APOBEC3 action, sustained evolution within the human population may result in a depletion of lower-consequence target sites (i.e., synonymous or conservative amino acid changes) and thus expedite a decrease in fitness of MPXV over time. This could be both through a reduction in the number of viable offspring viruses produced by infected cells and as a result the accumulation of moderately deleterious mutations by genetic drift (i.e., mutation load). However, the timescale on which this might happen is uncertain and

other evolutionary forces such as recombination may act to restore fitness, and we do not address this further in this study. A further uncertainty arises when considering the variable repertoire of genes associated with virus infectivity or host immune modulation in poxvirus genomes. Mutations that alter or abrogate the function of these genes may have little direct effect on virus replication machinery, but may disrupt the virus/host interaction. There is precedent for the naturally occurring inactivation of genes in VARV contributing to host specificity, and consequently the loss of function of some MPXV genes through APOBEC3 activity may potentially have adaptive value for the virus as it replicates and transmits in a new host (31, 32).

Even if the mutations that accumulate through this process are simply the neutral residue of a sub-optimal antiviral host defence, they have produced sufficient variability for the phylogenetic analysis of the epidemic over the short term. The initial lineages proposed by Happi et al (7) have now expanded with the 2022 epidemic B.1 lineage now encompassing 17 sublineages at time of writing (33). The rapid and temporally linear accumulation of mutations means that genomic epidemiological models and tools (34, 35), usually employed for RNA viruses, may also have utility to hMPXV-1.

The linear accumulation of APOBEC3-type mutations since the emergence and spread of MPXV in humans

Since 2017 the genomes thus far sampled from Clade IIb have accumulated APOBEC3-type single nucleotide mutations approximately linearly over time (Fig. 3A–B, labelled phylogeny in Fig. S13). We applied Bayesian regression analysis on the root-to-tip plot of sequences in Fig. 3A, which includes one representative B.1 genome, and also separately on the B.1 lineage (B.1 phylogeny in Fig. S1)(20). To ensure the elevated temporal signal is unique to APOBEC3 data, we show combinations of APOBEC3 and non-APOBEC3 mutations on Clade IIb data in Fig. S14. The estimated rate of accumulation was 6.18 per year (95% credible intervals of 5.20, 7.16). For the B.1 lineage, the rate was 5.93 per year (2.95, 8.92) – suggesting that despite widespread and rapid transmission within MSM networks, the rate of accumulation of APOBEC3 mutations remained the same as the rest of the Clade IIb. It is notable that the regression line for B.1 lies substantially above that for the rest of Clade IIb suggesting that this lineage accumulated more mutations than expected prior to the emergence of B.1. However most of these mutations are also present in the genome from Maryland, USA (Accession number: ON676708) from November 2021(10), indicating they arose and circulated for some months prior to the B.1 epidemic (Fig. S15–S16). Extrapolating back to when the APOBEC3-type mutations started to accumulate provides an estimate of when the first APOBEC3 mutations occurred in the stem of the branch leading to the 2017 epidemic. If we assume all these mutations are due to APOBEC3 in humans then we estimate this date of emergence to be 20-Jun-2015. However, we expect a few APOBEC3-like mutations to actually be due to replication errors during the earlier pre-emergence epoch. The number of mutations we ascribe to this period will affect our estimate linearly – i.e., if 3 mutations were not due to APOBEC3 then the estimate would shift to 14-Dec-2015.

To accommodate this uncertainty in our estimates we have developed a more explicit model of APOBEC3-mediated evolution in the BEAST software package (20, 34). We estimate the action of APOBEC3 on the MPXV population is driving evolution at a rate ~28 times faster than that of the background evolutionary rate (Fig. S17). Gigante et al (10) also described an elevated overall rate of evolution in the A lineage but did not decompose the APOBEC3 and non-APOBEC3 contribution to this. The time of the most recent common ancestor of the post 2017 genomes is estimated to be 23-Feb-2016 (28-Jun-2015, 28-Sep-2016) with the transition to sustained human to human transmission to be 14-Sep-2015 (21-Aug-2014, 31-Jul-2016; Fig. 3C, Fig. S17). Unlike the assumption in Fig. 3B that all APOBEC3 mutations occurred post-emergence, the BEAST analysis estimates the transition point integrating over all possibilities. This allows for the fact that a few of the APOBEC3-like mutations may actually be due to replication errors in the earlier evolutionary epoch and this might explain the slightly more recent date. We also see evidence of exponential growth in the number of infections in the epidemic prior to the emergence of lineage B.1 in 2022 (Fig. 3D), despite the decline in cases reported in 2020 (Fig. S1C), albeit the growth rate is relatively slow reinforcing the indication from the demographics of the cases that this is not a generalised epidemic.

Implications for the global public health response to mpox cases

Since the identification of the B.1 lineage, a number of countries have reported other lineages that lie outside the diversity of B.1, including USA, UK, Portugal, India and Thailand. In almost all instances these cases are reported as having a history of international travel. The lineages these are placed in (designated as A.2.1, A.2.2, A.2.3 and A.3) can all be phylogenetically traced back to the epidemic in Nigeria (Fig. 3A). This suggests that at least one instance of sustained human to human transmission is still ongoing outside of the recognised MSM networks that were the focus of the 2022 global epidemic. Stopping transmission in these communities, whilst necessary, will not be sufficient to eliminate the virus as a human epidemic. There are large portions of the globe without the surveillance to detect MPXV cases and if sustained human to human transmission has been ongoing since 2015–2016 it is plausible there are other populations that are currently enduring epidemics.

Historically, mpox was considered a zoonotic disease and cases have been treated as independent spillover events with low levels of circulation in the human population. Thus far, this continues to be an accurate characterization of Clade I in Central Africa. For Clade IIb, whilst some non-B.1 lineage cases may be new zoonotic infections, the majority of cases since 2016 are likely the result of human-to-human transmission. Although the B.1 lineage across the world is now diminished – though not yet eradicated – the human epidemic from which it arose continues unabated. It is critical that global public health affords MPXV cases in countries that are historically considered to have endemic reservoir species equal attention and concern to those elsewhere. Surveillance needs to be global if MPXV is to be eliminated from the human population and then prevented from reemerging.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We would like to thank those involved in sequencing MPXV genome sequences and shared data online rapidly, facilitating rapid public health responses, and also those who have continued to contribute to open science by sharing their data in open access platforms such as Genbank. We have provided all data and scripts related to this manuscript on Zenodo (40).

Funding:

Wellcome Trust ARTIC (Collaborators Award 206298/Z/17/Z, ARTIC network) (ÁOT, PL, MAS, AR)

European Research Council (grant agreement no. 725422 – ReservoirDOCS) (PL, MAS, AR).

National Institutes of Health (R01 AI153044) (PL, MAS, AR).

David and Lucile Packard Foundation (MW).

Research Foundation, Flanders – Fonds voor Wetenschappelijk Onderzoek - Vlaanderen, G066215N, G0D5117N and G0B9317N (PL).

HORIZON 2020 EU grant 874850 MOOD (PL).

HERA project (Grant/2021/PHF/23776) supported by the European Commission through the European Centre for Disease Control and Prevention (VB and JPG).

Data and materials availability:

Data, code, and materials used in the analysis are available from Zenodo (40). Tables S1 and S2 include details of accession numbers and author lists for all source genome sequence data used. All except a single genome sequence were sourced from Genbank on NCBI. Data partitions of the alignments have been included in the XML files on Zenodo and GitHub for ease of reproduction. A single sequence was sourced from GISAID (gisaid.org) and as such was removed from the XMLs. Supplementary tables include the GISAID identifier, which can be used to independently source the data and instructions for constructing the alignment and addition to the xml are included.

References and Notes

1. NCDC, "An Update of Monkeypox Outbreak in Nigeria" (Nigeria Centre for Disease Control, 2017), (available at https://ncdc.gov.ng/diseases/sitreps/?cat=8).
2. Rimoin AW, Mulembakani PM, Johnston SC, Lloyd Smith JO, Kisalu NK, Kinkela TL, Blumberg S, Thomassen HA, Pike BL, Fair JN, Wolfe ND, Shongo RL, Graham BS, Formenty P, Okitolonda E, Hensley LE, Meyer H, Wright LL, Muyembe J-J, Major increase in human monkeypox incidence 30 years after smallpox vaccination campaigns cease in the Democratic Republic of Congo. Proc. Natl. Acad. Sci. U. S. A. 107, 16262–16267 (2010). [PubMed: 20805472]
3. Baker TD, "W.H.O., The Global Eradication of Smallpox. Final Report of The Global Commission for the Certification of Smallpox Eradication. Geneva, World Health Organization, 1980, 122 pp., Sfr. 11,-" in Clio Medica. Acta Academiae Internationalis Historiae Medicinae, Vol. 17 (Brill, 1982), pp. 268–269.
4. Jezek Z, Nakano JH, Arita I, Mutombo M, Szczeniowski M, Dunn C, Serological survey for human monkeypox infections in a selected population in Zaire. J. Trop. Med. Hyg. 90, 31–38 (1987). [PubMed: 3029395]
5. Alakunle E, Moens U, Nchinda G, Okeke MI, Monkeypox Virus in Nigeria: Infection Biology, Epidemiology, and Evolution. Viruses. 12 (2020), doi:10.3390/v12111257.
6. Isidro J, Borges V, Pinto M, Sobral D, Santos JD, Nunes A, Mixão V, Ferreira R, Santos D, Duarte S, Vieira L, Borrego MJ, Núncio S, de Carvalho IL, Pelerito A, Cordeiro R, Gomes JP,

Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. Nat. Med. 28, 1569–1572 (2022). [PubMed: 35750157]

7. Happi C, Adetifa I, Mbala P, Njouom R, Nakoune E, Happi A, Ndodo N, Ayansola O, Mboowa G, Bedford T, Neher RA, Roemer C, Hodcroft E, Tegally H, O'Toole Á, Rambaut A, Pybus O, Kraemer MUG, Wilkinson E, Isidro J, Borges V, Pinto M, Gomes JP, Freitas L, Resende PC, Lee RTC, Maurer-Stroh S, Baxter C, Lessells R, Ogwell AE, Kebede Y, Tessema SK, de Oliveira T, Urgent need for a non-discriminatory and non-stigmatizing nomenclature for monkeypox virus. PLoS Biol. 20, e3001769 (2022). [PubMed: 35998195]
8. Patrono LV, Pléh K, Samuni L, Ulrich M, Röthemeier C, Sachse A, Muschter S, Nitsche A, Couacy-Hymann E, Boesch C, Wittig RM, Calvignac-Spencer S, Leendertz FH, Monkeypox virus emergence in wild chimpanzees reveals distinct clinical outcomes and viral diversity. Nat Microbiol. 5, 955–965 (2020). [PubMed: 32341480]
9. O'Toole Á, Rambaut A, Initial observations about putative APOBEC3 deaminase editing driving short-term evolution of MPXV since 2017. Virological (2022), (available at https://virological.org/t/830).
10. Gigante CM, Korber B, Seabolt MH, Wilkins K, Davidson W, Rao AK, Zhao H, Smith TG, Hughes CM, Minhaj F, Waltenburg MA, Theiler J, Smole S, Gallagher GR, Blythe D, Myers R, Schulte J, Stringer J, Lee P, Mendoza RM, Griffin-Thomas LA, Crain J, Murray J, Atkinson A, Gonzalez AH, Nash J, Batra D, Damon I, McQuiston J, Hutson CL, McCollum AM, Li Y, Multiple lineages of monkeypox virus detected in the United States, 2021–2022. Science. 378, 560–565 (2022). [PubMed: 36264825]
11. Suspène R, Henry M, Guillot S, Wain-Hobson S, Vartanian J-P, Recovery of APOBEC3-edited human immunodeficiency virus G->A hypermutants by differential DNA denaturation PCR. J. Gen. Virol. 86, 125–129 (2005). [PubMed: 15604439]
12. Chen H, Lilley CE, Yu Q, Lee DV, Chou J, Narvaiza I, Landau NR, Weitzman MD, APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. Curr. Biol. 16, 480–485 (2006). [PubMed: 16527742]
13. Vartanian J-P, Guétard D, Henry M, Wain-Hobson S, Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. Science. 320, 230–233 (2008). [PubMed: 18403710]
14. Vartanian J-P, Henry M, Marchio A, Suspène R, Aynaud M-M, Guétard D, Cervantes-Gonzalez M, Battiston C, Mazzaferro V, Pineau P, Dejean A, Wain-Hobson S, Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. PLoS Pathog. 6, e1000928 (2010). [PubMed: 20523896]
15. Tsuge M, Noguchi C, Akiyama R, Matsushita M, Kunihiro K, Tanaka S, Abe H, Mitsui F, Kitamura S, Hatakeyama T, Kimura T, Miki D, Hiraga N, Imamura M, Takahashi S, Hayses CN, Chayama K, G to A hypermutation of TT virus. Virus Res. 149, 211–216 (2010). [PubMed: 20138932]
16. Suspène R, Aynaud M-M, Guétard D, Henry M, Eckhoff G, Marchio A, Pineau P, Dejean A, Vartanian J-P, Wain-Hobson S, Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. Proc. Natl. Acad. Sci. U. S. A. 108, 4858–4863 (2011). [PubMed: 21368204]
17. Peng Z-G, Zhao Z-Y, Li Y-P, Wang Y-P, Hao L-H, Fan B, Li Y-H, Wang Y-M, Shan Y-Q, Han Y-X, Zhu Y-P, Li J-R, You X-F, Li Z-R, Jiang J-D, Host apolipoprotein B messenger RNA-editing enzyme catalytic polypeptide-like 3G is an innate defensive factor and drug target against hepatitis C virus. Hepatology. 53, 1080–1089 (2011). [PubMed: 21480314]
18. Fehrholz M, Kendl S, Prifert C, Weissbrich B, Lemon K, Rennick L, Duprex PW, Rima BK, Koning FA, Holmes RK, Malim MH, Schneider-Schaulies J, The innate antiviral factor APOBEC3G targets replication of measles, mumps and respiratory syncytial viruses. J. Gen. Virol. 93, 565–576 (2012). [PubMed: 22170635]
19. Suspène R, Raymond KA, Boutin L, Guillier S, Lemoine F, Ferraris O, Tournier J-N, Iseni F, Simon-Lorière E, Vartanian J-P, APOBEC3F is a mutational driver of the human Monkeypox virus identified in the 2022 outbreak. J. Infect. Dis. (2023), doi:10.1093/infdis/jiad165.
20. See Materials and Methods.
21. Hasegawa M, Kishino H, Yano T, Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160–174 (1985). [PubMed: 3934395]

22. Kimura M, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120 (1980). [PubMed: 7463489]
23. Conticello SG, Thomas CJF, Petersen-Mahrt SK, Neuberger MS, Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol. Biol. Evol. 22, 367–377 (2005). [PubMed: 15496550]
24. Krishnan A, Iyer LM, Holland SJ, Boehm T, Aravind L, Diversification of AID/APOBEC-like deaminases in metazoa: multiplicity of clades and widespread roles in immunity. Proc. Natl. Acad. Sci. U. S. A. 115, E3201–E3210 (2018). [PubMed: 29555751]
25. Münk C, Willemsen A, Bravo IG, An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. BMC Evol. Biol. 12, 71 (2012). [PubMed: 22640020]
26. Yang L, Emerman M, Malik HS, McLaughlin Jnr RN, Retrocopying expands the functional repertoire of APOBEC3 antiviral proteins in primates. Elife. 9 (2020), doi:10.7554/eLife.58436.
27. Mikl MC, Watt IN, Lu M, Reik W, Davies SL, Neuberger MS, Rada C, Mice deficient in APOBEC2 and APOBEC3. Mol. Cell. Biol. 25, 7270–7277 (2005). [PubMed: 16055735]
28. Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, Wiltshire T, Orth AP, Vega RG, Sapinoso LM, Moqrich A, Patapoutian A, Hampton GM, Schultz PG, Hogenesch JB, Large-scale analysis of the human and mouse transcriptomes. Proc. Natl. Acad. Sci. U. S. A. 99, 4465–4470 (2002). [PubMed: 11904358]
29. Moss B, Poxvirus DNA replication. Cold Spring Harb. Perspect. Biol. 5 (2013), doi:10.1101/cshperspect.a010199.
30. Peng Q, Xie Y, Kuai L, Wang H, Qi J, Gao GF, Shi Y, Structure of monkeypox virus DNA polymerase holoenzyme. Science. 379, 100–105 (2023). [PubMed: 36520947]
31. Hendrickson RC, Wang C, Hatcher EL, Lefkowitz EJ, Orthopoxvirus genome evolution: the role of gene loss. Viruses. 2, 1933–1967 (2010). [PubMed: 21994715]
32. Mühlemann B, Vinner L, Margaryan A, Wilhelmson H, Castro C. de la Fuente, Allentoft ME, Damgaard P. de Barros, Hansen AJ, Nielsen S. Holtsmark, Strand LM, Bill J, Buzhilova A, Pushkina T, Falys C, Khartanovich V, Moiseyev V, Jørkov MLS, Sørensen P. Østergaard, Magnusson Y, Gustin I, Schroeder H, Sutter G, Smith GL, Drosten C, Fouchier RAM, Smith DJ, Willerslev E, Jones TC, Sikora M, Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. Science. 369 (2020), doi:10.1126/science.aaw8977.
33. GitHub - mpxv-lineages/lineage-designation: Official place for proposals and details around monkeypox virus lineage designations. GitHub, (available at https://github.com/mpxv-lineages/lineage-designation).
34. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A, Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 4, vey016 (2018). [PubMed: 29942656]
35. Sagulenko P, Puller V, Neher RA, TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol. 4, vex042 (2018). [PubMed: 29340210]
36. Li H, Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 34, 3094–3100 (2018). [PubMed: 29750242]
37. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Haeseler A. von, Lanfear R, Corrigendum to: IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. 37, 2461 (2020). [PubMed: 32556291]
38. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA, Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol. Biol. Evol. 30, 713–724 (2013). [PubMed: 23180580]
39. McElreath R, Statistical Rethinking: A Bayesian Course with Examples in R and STAN (CRC Press, 2020).
40. Zenodo - hmpxv/apobec3: All scripts and data related to this manuscript available at 10.5281/zenodo.8234483

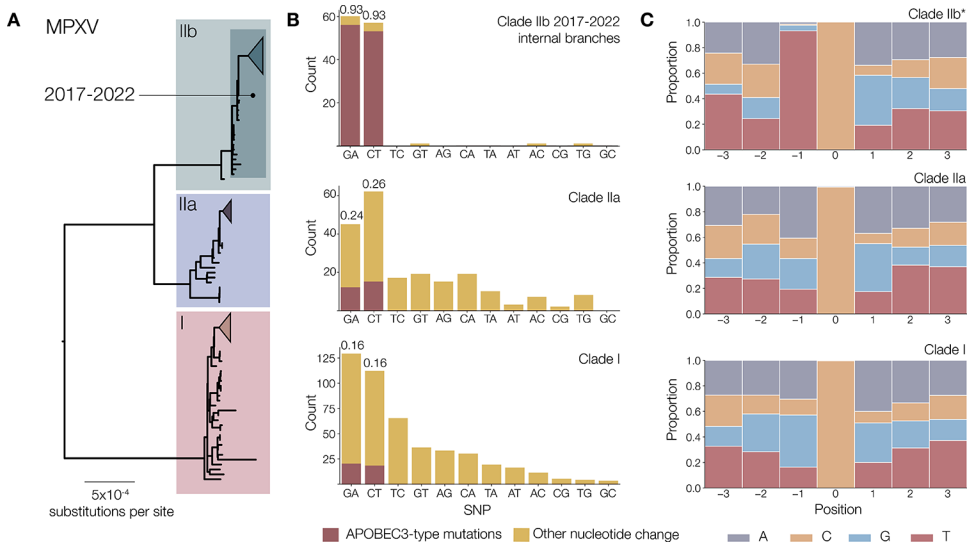

Fig. 1. Specific enrichment of APOBEC3-type mutations in MPXV samples collected since 2017.

(A) MPXV genetic diversity is categorised into Clade I (predominantly sequences from the DRC), Clade IIa (predominantly West African sequences) and Clade IIb. Within Clade IIb is a sub-clade of genomes sampled from 2017–2022 that show distinct mutational patterns to the other two clades. **(B)** We catalogue single nucleotide mutations across the phylogenies of Clade IIb, Clade IIa and Clade I (top to bottom). For Clade IIb, we include samples from 2017–2022 and only a single representative of the global lineage B.1. Of 120 reconstructed mutations that occurred on internal branches of the Clade IIb phylogeny (so are observed transmitted mutations), 109 are consistent with APOBEC3 editing (90.8% of mutations). Individual proportions of G→A and C→T mutations shown above the respective bars. Ancestral state reconstruction performed across Clade IIa and Clade I does not produce the same enrichment of mutations consistent with APOBEC3 editing, with only 27 of 207 observed mutations (13%) and 38 of 463 Clade I mutations (8%) fitting the dinucleotide pattern. **(C)** Observed heptamers of C→T or G→A mutated sites of Clade IIb, IIa and I phylogenies (top to bottom). Heptamers associated with G→A mutations have been reverse-complemented to reflect deamination on the negative strand. For Clade IIb, most C→T mutations are present in a TC dimer context, consistent with APOBEC3 editing (107 of 115 mutations, or 93%). However, the same is not seen for Clades IIa and I, in which 29 of 149 (19%) mutations and 42 of 256 (16%) have the dinucleotide context of APOBEC3 respectively, which is what we would predict under standard models of nucleotide evolution. *Only mutations occurring on internal branches of the Clade IIb phylogeny included.

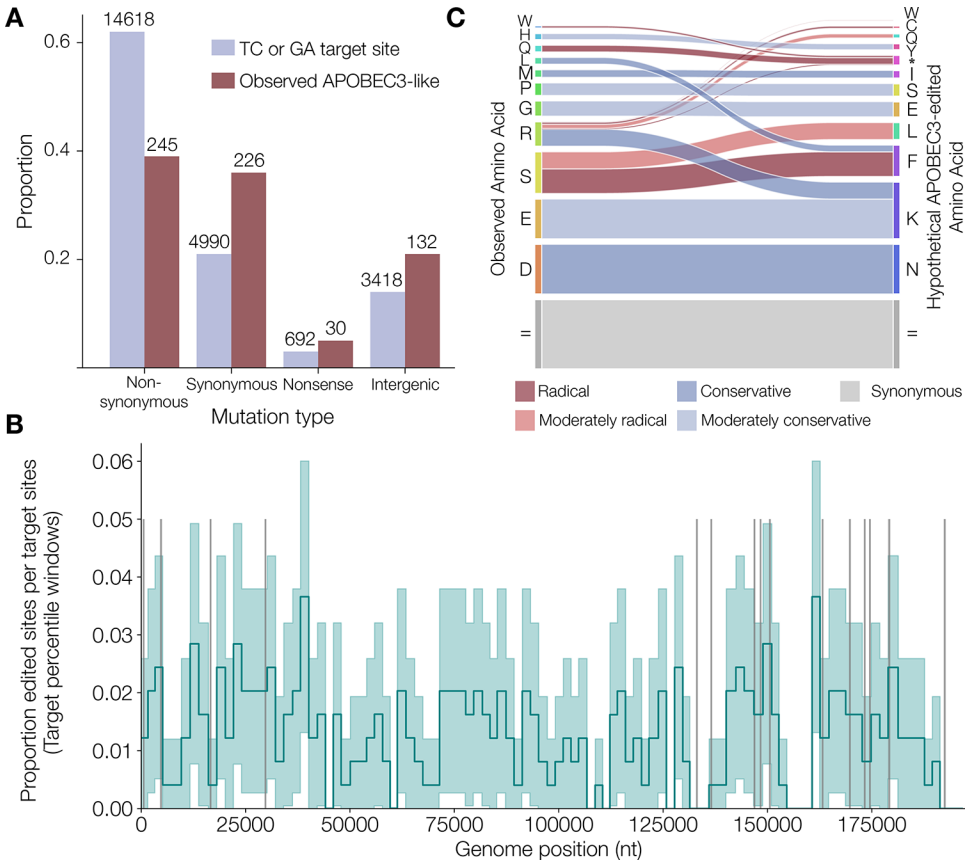

Fig. 2. Observed APOBEC3-type mutations are not merely a product of available target sites.

(A). Consequence of hypothetical APOBEC3 mutations at target dimer site (either the C in the TC target site or G in the GA target site) in the coding regions of the NCBI reference MPXV genome for Clade II (Accession NC_063383) and those observed APOBEC3 mutations across the coding regions of the Clade IIb phylogeny (not including the outgroup branch leading to the 1971 genome sequence). These are categorised into non-synonymous (altered amino acid), synonymous (amino acid remaining unchanged), nonsense (editing producing a stop codon) and intergenic (not present in a coding sequence). **(B)** The proportion of target sites edited for each target site percentile window across the MPXV genome. The teal shaded regions represent the binomial confidence interval around observations. Masked regions indicated by vertical grey bars. Observed edits include data from the Clade IIb phylogeny, with a single representative of lineage B.1 and not including the branch leading to the outgroup 1971 genome sequence. **(C)** Hypothetical amino acid changes for codons overlapping with TC and GA target sites in a reference MPXV genome (Genbank accession number: NC_063383) if APOBEC3 edited those dimers to TT and AA. Amino acid changes are coloured by Grantham Score (0–50 conservative: dark blue; 51–100 moderately conservative: light blue; 101–150 moderately radical: light red; >150 radical: dark red; synonymous: grey).

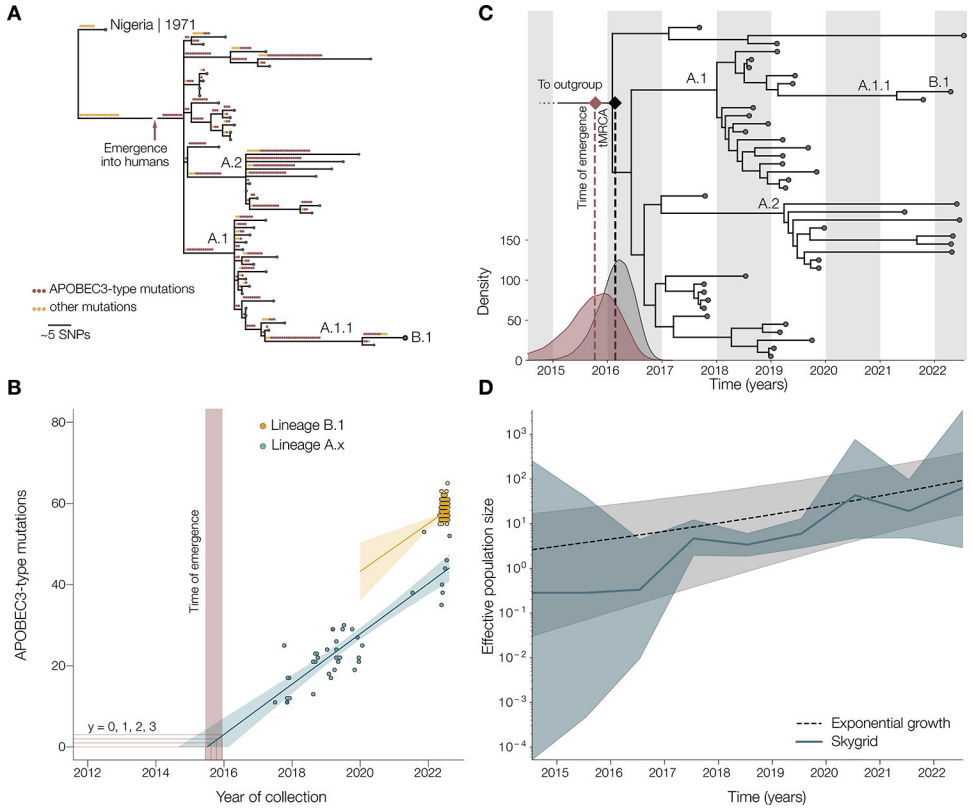

Fig. 3. Estimating the time of MPXV emergence into the human population from the accumulation of APOBEC3-type mutations.

(A) MPXV genomes sampled from human infections from 2017–2022, with an outgroup sequence from an outbreak in Nigeria in 1971 (n=44 including outgroup). Lineages indicated as per nomenclature proposed by Happi et al., (7). Mutations along each branch are indicated with circles coloured by whether it is putatively APOBEC3 edited (TC→TT and GA→AA; red) or whether it is another mutation type (yellow). The break in the basal branch illustrates the assumption made in the regression model in panel B, that all APOBEC3 mutations occurred after emerging into the human population, however we do not know the precise distribution of red or yellow mutations. **(B)** APOBEC3 mutations from MPXV genomes sampled since 2017. The reconstructed most recent common ancestor (MRCA) of the panel A phylogeny is used as the root in the root-to-tip plot and the y-intercept is used as a proxy for time of emergence which is inferred by fitting a Bayesian regression to the sequence dates from panel A. Intersects with y=1, 2, & 3 are also shown as it is likely that a small number of the APOBEC3-type mutations are actually earlier replication errors and not induced by APOBEC3. **(C)** Maximum Clade Credibility (MCC) phylogeny of MPXV Clade IIb with absolute time shown on the X-axis. We separated the alignment into an APOBEC3 and a non-APOBEC3 partition and modelled the substitution process in each independently. We used an epoch model with two outgroup sequences (not shown in panel) representing the first epoch and hMPXV-1 ingroup sequences representing MPXV post-emergence into the human population with an exponential growth model. The probability density distributions show the estimated time of the most recent common ancestor (tMRCA) of the ingroup as well as the estimated transition time that represents the

time of emergence into the human population. **(D)** Estimated effective population size of the outbreak using a non-parametric coalescent Skygrid model with 11 change points over a period of 8.5 years. This reconstruction falls within the bounds of the exponential growth model estimated from the second epoch in panel C, suggesting that the MPXV population has been exponentially growing since at least 2016.