

Putative APOBEC3 deaminase editing in MPXV as evidence for sustained human transmission since at least 2016.

Áine O'Toole^{1,\$}

Richard A. Neher²

Nnaemeka Ndodo³

Vitor Borges⁴

Ben Gannon⁵

João Paulo Gomes⁴

Natalie Groves⁶

David J King⁷

Daniel Maloney¹

Philippe Lemey⁸

Kuiama Lewandowski⁵

Nicholas Loman^{6,9}

Richard Myers⁶

Marc A. Suchard¹⁰

Michael Worobey¹¹

Meera Chand⁶

Chikwe Ihekweazu³

David Ulaeto^{7,†}

Ifedayo Adetifa^{3,†}

Andrew Rambaut^{1,†,\$}

Affiliations

1. Institute of Ecology & Evolution, University of Edinburgh, Edinburgh, EH9 3FL, United Kingdom
2. Biozentrum, University of Basel and Swiss Institute of Bioinformatics, Switzerland
3. Nigeria Centers for Disease Control and Prevention, Abuja, Nigeria.
4. National Institute of Health Doutor Ricardo Jorge, Lisbon, Portugal
5. UK Health Security Agency, Porton Down, Salisbury, United Kingdom

6. UK Health Security Agency, London, SW1P 3HX, United Kingdom
7. Defence Science and Technology Laboratory (DSTL), Chemical, Biological and Radiological Division, Porton Down, Salisbury, United Kingdom
8. Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven – University of Leuven, Belgium.
9. University of Birmingham, Birmingham, United Kingdom
10. Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, USA
11. Department of Ecology and Evolutionary Biology, University of Arizona, USA

[†]these authors are joint-senior

[§]correspondence should be addressed to aine.otoole@ed.ac.uk or a.rambaut@ed.ac.uk

Keywords: Mpox, MPXV, APOBEC, evolution, phylogenetics

Abstract

Mpox is often described as being endemic in West and Central Africa as a zoonotic disease that transmits through contact with the reservoir rodent host, likely a species of African squirrel. In May 2022, human cases of Mpox were detected spreading internationally beyond countries with known endemic reservoirs. At time of writing, 84,700 confirmed cases have been reported in 110 countries. When the first cases from 2022 were sequenced, it was seen that they shared 42 single nucleotide differences from the closest mpox virus (MPXV) genome sampled in 2018. This number of changes within 3-4 years is unexpectedly large and points to a much greater evolutionary rate than expected for a poxvirus. Strikingly, most nucleotide changes are of a specific type – a dinucleotide change from TC->TT or its reverse complement GA->AA. This mutation type is characteristic of the action of APOBEC3 deaminases; host-enzymes with reported antiviral function. Analysis of MPXV genomes sampled from 2017 to 2022 showed further evidence of TC->TT mutation pattern enrichment, with 93% of transmitted single nucleotide mutations since 2017 consistent with APOBEC3 editing. Assuming APOBEC-editing is characteristic of MPXV infection in human hosts, we propose an APOBEC clock that – at a rate of ~6 APOBEC3 mutations per year – estimates MPXV has been circulating in humans since 2016. This evolutionary pattern of host-enzyme editing has implications for the longer-term fitness of the virus in this epidemic as such mechanisms are primarily

antiviral in function, but in the context of a poxvirus also provide a source of variation that may conceivably facilitate adaptation.

Introduction

Since 2017, the Nigeria Centre for Disease Control has been reporting cases of MPXV (mpox virus) infection in humans (1). MPXV is often described as being endemic in West and Central Africa as a zoonotic disease that transmits through contact with the reservoir rodent host. The virus is divided into two clades with isolates from Central Africa falling into Clade I and isolates from West Africa predominantly falling into Clade II. Whilst MPXV has a broad host range and can readily infect multiple rodent and primate species in the laboratory and in nature (2–8), African rodents remain the most likely reservoir hosts, specifically rope and sun squirrels (genera *Funisciurus* and *Heliosciurus*), African dormice (genus *Graphiurus*) and pouched rats (genus *Cricetomys*), native to West and Central Africa. MPXV-specific antibodies have been detected in squirrel species in a number of surveys that sampled wildlife from areas of reported human MPXV cases (3–5) and in 1985 an infected rope squirrel was associated with a zoonotic outbreak in the Democratic Republic of Congo (9). Pouched rats and dormice housed in proximity to North American prairie dogs were the likely cause of a widespread zoonotic MPXV outbreak in the USA in 2003 (10). Ecological niche modelling predicted a dramatic range expansion of the rope squirrel with climate change by 2050, however this is not common to all putative host species as the range of the pouched rat (*Cricetomys gambianus*) is predicted to contract in response to climate change (11). Furthermore, continued forest clearing in favour of agricultural fields and forests may actually increase habitat suitable for reservoir hosts, such as the rope squirrel, and precipitate more frequent human-reservoir interactions (3, 11).

Since the first human cases were observed in the 1970s MPXV infections have been predominantly associated with infants and children, with a marked increase in case counts since the 1980s, likely due in part to the end of smallpox vaccination after eradication (12–14). The smallpox vaccine offers protection against MPXV, however with the eradication of variola virus in 1977, protection has been waning in the population. Prior to 2017, the majority of cases have been reported in DRC, as zoonoses initiating in school-age boys with Clade I viruses (12). The outbreak observed in Nigeria from 2017, from which the global outbreak has grown, has involved Clade II viruses (8, 15). The

demographic mainly affected in Nigeria has been men aged 25 to 40, with only ~25% of cases involving women, and very few cases in children (8).

Between 2018 and March 2022 there have been a limited number of cases in Israel (MN648051), Singapore (MT250197), USA (ON676708), and the United Kingdom (UK; MT903344, MT903343, OL504741) associated with travel from Nigeria, with one case of onward transmission in the UK (MT903345). However, in May 2022, cases of MPXV infection were detected spreading widely across Europe and subsequently across the globe. At time of writing, 84,700 confirmed cases have been reported in 110 countries (16), predominantly in men who have sex with men (MSM) sexual contact networks. The rapid growth of cases in these networks has prompted concern that the 2022 epidemic represents a transition to a more transmissible form of the virus.

In the wake of the SARS-CoV-2 pandemic, the expanded capacity to perform viral genome sequencing enabled rapid generation of MPXV genome data, and the legacy of open-data sharing practices led to their quick dissemination to public repositories. The first MPXV genome sequences from 2022 cases showed these viruses had descended from the clade sampled in 2017-2019 from cases diagnosed in Israel, Nigeria, Singapore, and the UK (17). These early 2022 genomes are indicated as a triangle within Clade IIb in Figure 1a and represent lineage B.1 as per the nomenclature proposed by Happi et al. (15). Isidro et al., (17) noticed that sequences within lineage B.1 shared 42 single nucleotide differences from the closest earlier MPXV genomes from 2018 (specifically UK_P2 and UK_P3; accession numbers: MT903344, MT903345). Later, an MPXV sample from 2021 was retrospectively sequenced (Accession number: ON676708) (18) and split this long branch to lineage B.1 (Figure 4).

The long term evolutionary rate of the related variola virus (VARV; the smallpox virus) has previously been estimated to be about 9×10^{-6} (posterior mean, with 95% Bayesian credible intervals of $7.8 - 10.2 \times 10^{-6}$) substitutions per site per year (19) translating into about 1-2 nucleotide changes per year for a nearly 200,000 nucleotide genome. During 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 ($1.2 - 2.7 \times 10^{-6}$), on the same order of magnitude as the VARV estimate (~1 nucleotide change every 3 years) (20). This makes 42 substitutions in the space of 3-4 years an unexpectedly large number. Furthermore, across the phylogeny of 45 MPXV genome sequences collected since 2017 (Figure 4A) and within the 769 genomes available within lineage B.1 (Supplementary Figure 1), we see accumulation

of an excess of mutations since 2017. If MPXV is a zoonotic virus with limited human to human transmission, this long branch could be evidence of adaptation to humans facilitating the sustained transmission that is now observed.

Strikingly, 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 (21). This specific 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, with the exception being APOBEC3G which prefers 5'CC dinucleotides (22). Recent work also highlighted the importance of ssDNA secondary structure flanking the target site for APOBEC3 editing (23).

APOBEC-type enzymes are a large and diverse family of deaminase proteins with the ability to bind RNA and ssDNA. Reported functions range from tissue-specific RNA diversification with APOBEC1, somatic hypermutation in the adaptive immune system with Activation-Induced cytidine Deaminase (AID) (24–26), skeletal and muscle cell differentiation with APOBEC2, and restriction of mobile elements and antiviral action in hosts with APOBEC3-type enzymes (24, 27). APOBEC3-driven deamination has been demonstrated with many DNA viruses and retroviruses; (28–35). Vaccinia virus (VACV), an Orthopoxvirus closely related to MPXV, was subjected to ABOBEC3 *in vitro* and reportedly showed no growth disadvantage (36). Although this study did not involve confirmatory sequencing to confirm that no APOBEC3 editing had taken place.

Whether a virus is subject to APOBEC3-induced mutation will be a function of the availability of appropriate viral target ssDNA molecules during virus replication, co-localisation of virus replication with APOBEC3, and sufficient stoichiometry of enzyme to virus genetic material for an observable amount of editing to occur but insufficient to control the virus completely. Poxviruses are large DNA viruses with complex replication cycles that transiently expose ssDNA during replication. They replicate exclusively in the cytoplasm (37, 38) where a number of members of the APOBEC3 enzyme family localise in addition to many other host antiviral factors. Poxviruses have evolved various antiviral factor antagonists and some have dedicated up to 50% of their genomic capacity to immune modulatory mechanisms (39, 40). For observable levels of APOBEC3 editing to

occur, it may come down to a balance between the level of host enzyme readily available to bind target virus ssDNA and the rate of virus DNA amplification during replication.

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 (41, 42). APOBEC3 genes in primates have undergone recent expansion, with primate genomes now having 7 paralogs of 3 ancestral genes (27, 43). Rodents, the reservoir of MPXV, have only a single functional APOBEC3 protein resulting from gene loss and fusion events (27). This is thought to be reflected in the decreased mobile element expansion in primates relative to rodents (44, 45). Rodent APOBEC3 has been shown to be expressed preferentially in spleen and bone marrow with limited expression observed in other tissues (46, 47).

Nomenclature

A recently published paper (15) proposed an updated, systematic, nomenclature for the phylogenetic structure of MPXV to replace previous geographical clade names. In this nomenclature the 2022 global epidemic lies in Clade IIb with all genomically characterised cases since 2017 designated with lineage labels similar to SARS-CoV-2. The recent human epidemic, provisionally named hMPXV-1 by Happi *et al.* (15), comprises a hierarchy of lineages starting with 'A' currently represented by genomes from 2017 cases in Nigeria. The bulk of 2022 genomes are part of a lineage denoted 'B.1' but other lineages ('A.2' and 'A.3') have also been reported in the USA, UK and Portugal, in most cases in individuals with international travel history. We follow this nomenclature here.

Methods

We compiled high-quality MPXV genomes from human and non-human animal outbreaks sampled from as early as 1965 and up to the current 2022 outbreak. The dataset consisted of 94 genomes with representatives from Clades I, IIa and IIb. A second dataset comprises all Clade IIb genomes from 2017-2022 but with a single representative of lineage B.1 of hMPXV-1 and two earlier sequences from Nigeria from 1971 and 1978 (accession numbers KJ642617 and KJ642615, respectively). The genome from 1978 (accession KJ642615) was used as an outgroup, due to its greater divergence despite being sampled more recently, used to root the tree and then not analysed further. A separate analysis was run specifically on lineage B.1 where all high-quality B.1 genomes with an exact date of collection on Genbank were downloaded on the 2022-08-22. The 2021 genome from Maryland, USA (lineage A.1.1; accession number: ON676708; Ref (18)

was used as an outgroup to root the B.1 tree (n=769). Accession numbers and acknowledgments for genome sequences used in this study are provided in Supplementary Table 1.

All the MPXV genome datasets were aligned against the Clade II reference genome (which is an early hMPXV-1 genome from Nigeria, accession: NC_063383) using minimap2 v2.17 (48) to generate a single coordinate system. Genome sequences were trimmed from position 190,788 to the end, corresponding to the 3' terminal repeat region of the MPXV genome. A series of repetitive or low-complexity regions were masked from the alignment (Supplementary Table 3). For convenience, this alignment and extraction pipeline is available as an installable tool at <https://github.com/aineniamh/squirrel>.

The full MPXV phylogeny of Clade I, IIa and IIb was estimated using maximum likelihood in IQ-TREE v2.0 using a Jukes-Cantor substitution model (49) and subsequently midpoint rooted. For the ancestral reconstruction analysis, we estimated phylogenies for each of Clade I, Clade IIa and Clade IIb. The MPXV Clade IIb phylogeny was estimated with the same parameters, but with the outgroup specified as the 1978 Nigerian MPXV genome sequence (Accession KJ642615). Similarly, we estimated phylogenies independently for the hMPXV-1 coding sequence alignment and for Clade I and Clade II of MPXV using the above parameters with genome sequence KJ642617 as an outgroup.

We performed the ancestral state reconstruction using IQ-TREE2 on all individual Clade phylogenies of MPXV. We parsed out all reconstructed sites that vary unambiguously across the phylogeny (i.e. we exclude missing data) and mapped the single nucleotide mutations that occurred to the relevant branch of the phylogeny. Using the reconstructed node states we also catalogued the dimer and heptamer context of all C->T or G->A mutations that occurred across the phylogenies. Scripts are available from <http://github.org/hmpxv/apobec3/>.

We compared each recent Clade IIb sequence with the reconstructed most recent common ancestor of the Clade IIb phylogeny (Figure 4) and collating the mutations down each branch from root-to-tip and their dinucleotide context. Using only APOBEC3-type mutations, we constructed an APOBEC3 root-to-tip linear regression for Clade IIb and similarly for lineage B.1 (phylogeny in Supplementary Figure 1). We performed a linear regression of the number of APOBEC3 mutations from the root of the tree to each tip against the date of collection of the tip. The slope is an estimate of the APOBEC3-specific

mutation rate and assuming that APOBEC3 mutations are characteristic of infection in the human host, we use the X-intercept of this regression to estimate the date of emergence into a human host. For more details and the code, see Supplementary Info. This approach does not adjust for phylogenetic correlation and, in particular, the upward shift in B.1 is the result of a single shared branch with a higher than average number of mutations, but one that is not necessarily beyond expectations.

To investigate the target site context of mutations occurring in the hMPXV-1 phylogeny, we extracted the relevant nucleotide heptamers for all C->T or G->A mutations that occurred across the MPXV Clade IIb phylogeny (260 of 296 mutations) and for G->A mutations we took the reverse complement of the nucleotide sequences to normalise for strand. We also calculated the heptamer context for just the backbone of the phylogeny (i.e. the branches leading to the 2022 outbreak), which included 55 of 57 mutations on the respective branches. Using the reconstructed mutations and branch states in the phylogeny, we collected the ancestral dimer context for mutations that were either G->A or C->T to assess which mutation occurred in a context consistent with APOBEC3 editing.

Results

The diversity of MPXV can be decomposed into three major clades; Clades I, IIa and IIb (Figure 1A; Happi et al., (15)). 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. Clade IIb has an early sample taken in 1971 (KJ642617) but most of the sequences in Clade IIb are more recent virus genomes from 2017-2022 that (15) have labelled as hMPXV-1 (Figure 1A). Within the recent diversity of Clade IIb (indicated as the darker box within IIb in Figure 1A), we catalogued transmitted mutations that occurred between 2017 and 2022 with only a single representative from the 2022 global lineage B.1. The great majority of these mutations are of the type G->A or C->T (90.8%) (Figure 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 (Figure 1C). Similarly, this enrichment of TC->TT and GA->AA mutations is observed within the B.1 lineage (Supplementary Figure 2).

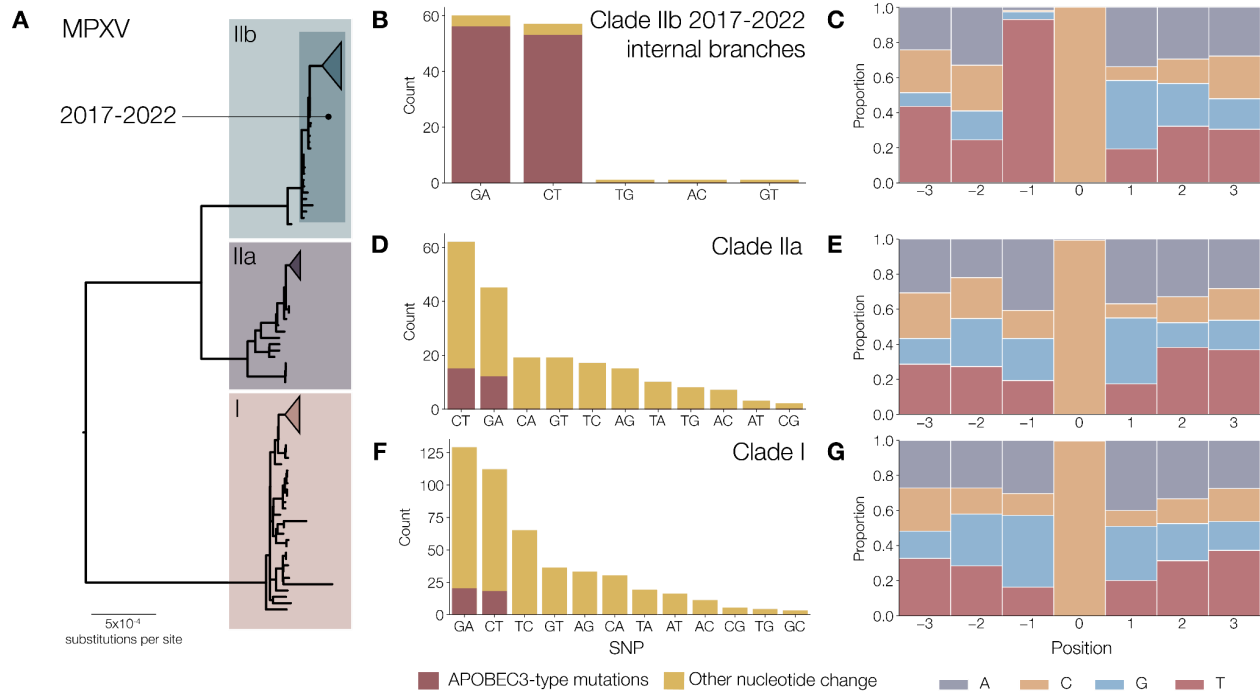


Figure 1 | A) MPXV phylogenetic tree highlighting Clade I (predominantly sequences from the DRC), Clade IIa (predominantly West African sequences) and Clade IIb. Within Clade IIb is a sub-clade of genomes sampled from 2017-2022 that show distinct mutational patterns to the other two clades. B) Of 120 reconstructed single nucleotide 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). C) Observed heptamers of C->T or G->A mutated sites in internal branches of the IIb phylogeny. Heptamers associated with G->A mutations have been reverse-complemented to reflect deamination on the negative strand. Most C->T mutations are present in a TC dimer context, consistent with APOBEC3 editing (107 of 115 mutations, or 93%). D) Ancestral state reconstruction performed across Clade IIa does not produce the same enrichment of mutations consistent with APOBEC3 editing, with only 27 of 207 observed mutations (13%) fitting the dinucleotide pattern. E) Observed heptamers around C->T mutations (and reverse-complemented G->A mutations) in the Clade IIa phylogeny. 29 of 149 (19%) mutations have the dinucleotide context of APOBEC3. F) Only 38 of 463 Clade I mutations (8%) are consistent with APOBEC3 editing. G) Observed heptamers in C->T mutations (and reverse-complemented G->A mutations) in the Clade I phylogeny respectively. 42 of 256 (16%) normalised C->T or G->A mutations are in a TC or GA dinucleotide context, which is what we would predict under standard models of nucleotide evolution.

Comparing MPXV Clade IIb with Clade Ia and IIa emphasises that this pattern is not seen outside of Clade IIb (Figure 1D-G). 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 (50, 51) (Figure 1D & F). 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 (Figure 1E & G).

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 looked at what the effect of a deamination mutation at these sites would be in terms of amino acid changes (Figure 4). 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 $P=7.6 \times 10^{-18}$. 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 (Supplementary Figure 3). 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 merely a product of the availability of target sites (Kolgorov-Smirnov test statistic=0.07, P-value=0.0002 ; Figure 2B-C). When considering synonymous and non-synonymous APOBEC3-like mutations separately, we observe the difference between target and observed APOBEC3-like mutations for both cases (Supplementary Figure 4).

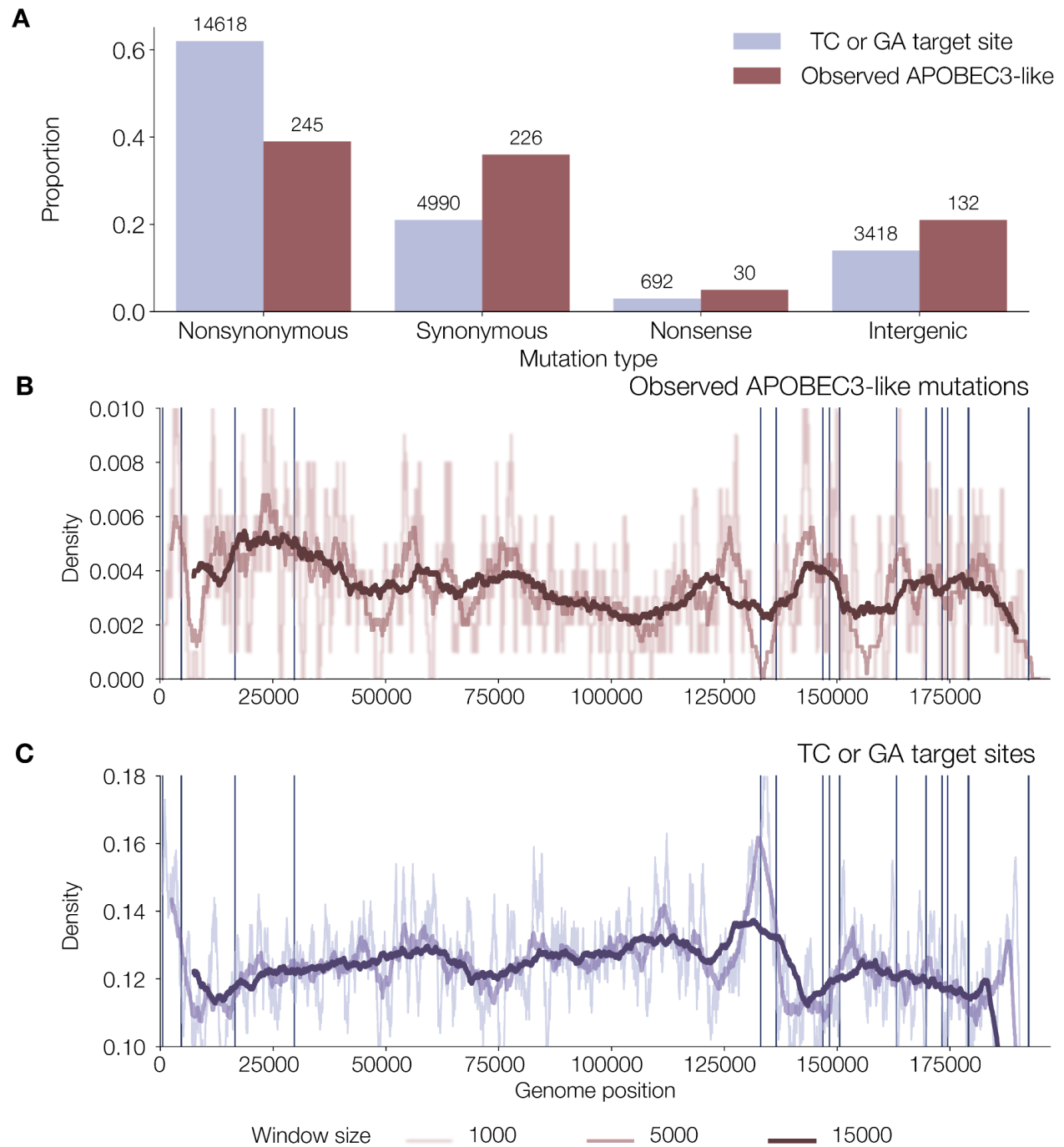


Figure 2 | A) Consequence of hypothetical APOBEC3 mutations at target dimer site (either the C in the TC target site or G in the GA target site) in currently unmutated sites in the coding regions of the NCBI reference MPXV genome (Accession NC_063383) and those observed APOBEC3 mutations across the coding regions of the Clade IIb phylogeny (including genomes in Figure 4 and Supplementary Figure 1, except 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 density (moving average linear convolution with window

sizes of 1000, 5000 and 15000) of observed APOBEC-consistent mutations and C) APOBEC-target dimer sites (GA and TC) in a masked Clade IIb reference genome (Accession NC_063383). Repetitive/ low complexity masked regions in the alignment indicated by vertical bars.

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 amino acid changes that APOBEC3 editing can induce is also limited (Figure 3). 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, we don't expect the elevated rate to change the rate of adaptation of the virus substantially.

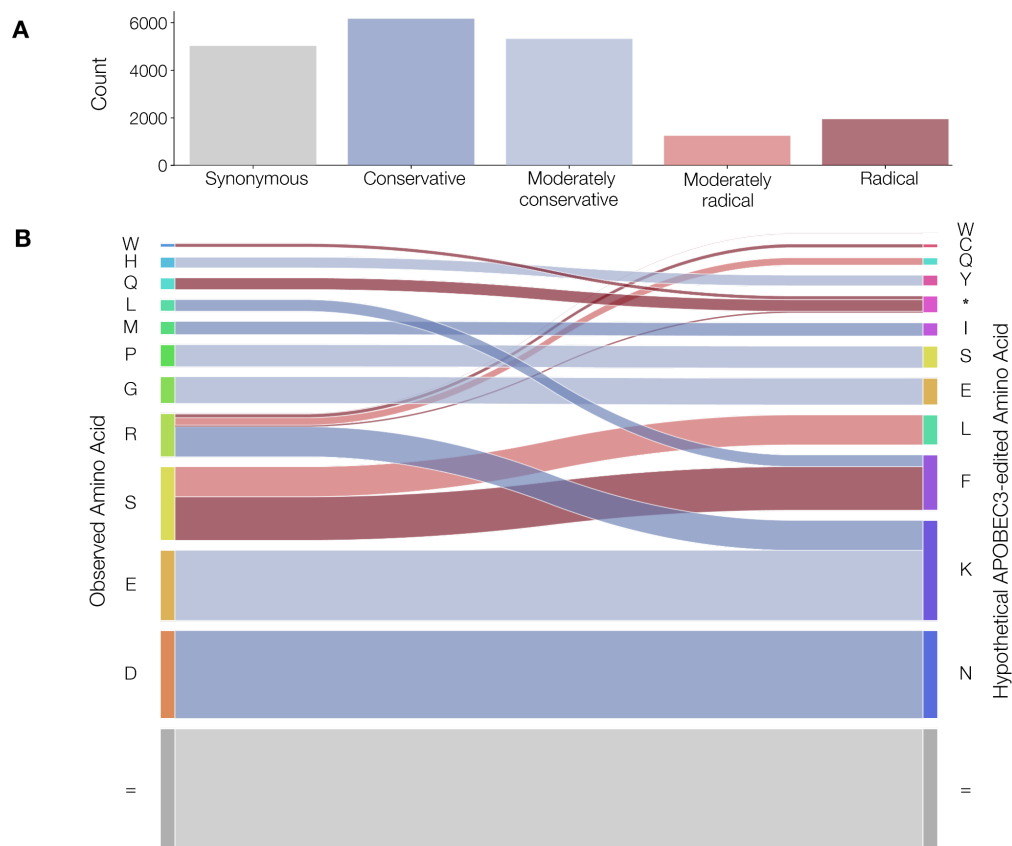


Figure 3 | Amino acid mutations at TC and GA dimer sites in a reference MPXV genome (accession number NC_063383) that could occur through APOBEC3 editing. A) Barplot of amino acid changes categorised by Grantham Score (0-50 conservative, 51-100 moderately conservative, 101-150 moderately radical, >150 radical). B) Hypothetical amino acid changes for codons overlapping with TC and GA target sites if APOBEC3 edited those dimers to TT and AA.

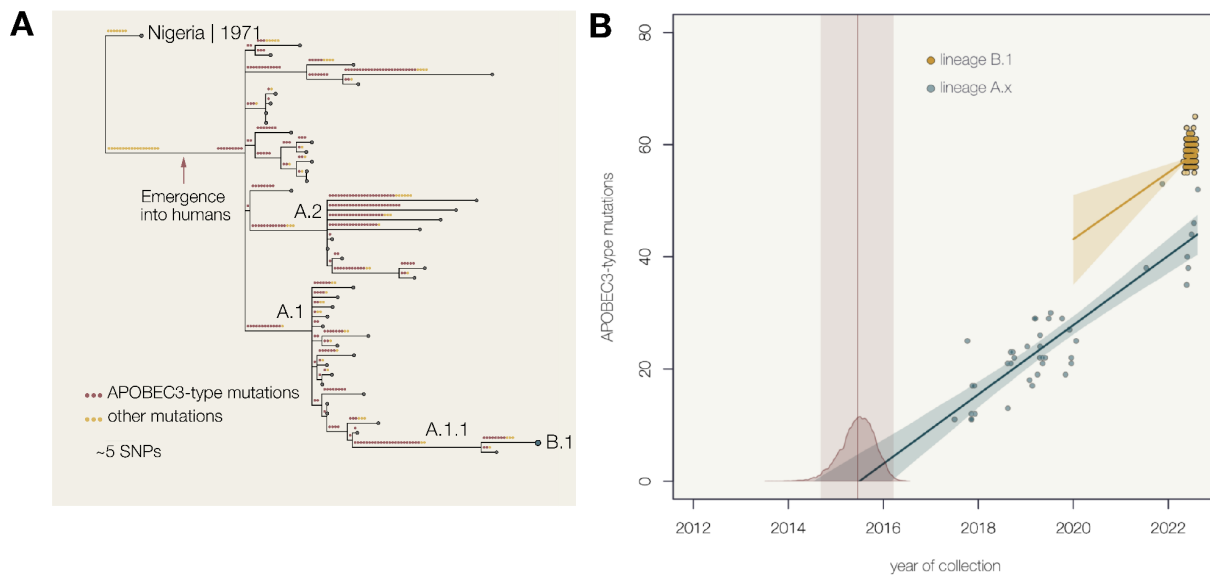


Figure 4 | A) A phylogenetic tree of MPXV genomes sampled from human infections from 2017-2022, with an outgroup sequence from an outbreak in Nigeria in 1971 ($n=45$ including outgroups). Lineages indicated as per nomenclature proposed by Happi et al., (15). Single nucleotide mutations along each branch are indicated with circles coloured by whether a mutation is putatively APOBEC3 edited (TC->TT and GA->AA; red) or whether it is another mutation type (yellow). B) Root to tip plot of 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. We performed Bayesian regression analysis on sequences from panel A, which includes one representative B.1 genome, and also separately on the B.1 lineage using the reconstructed MRCA (B.1 phylogeny shown in Supplementary Figure 1).

Since 2017 the genomes thus far sampled from Clade IIb have accumulated APOBEC3-type single nucleotide mutations approximately linearly over time (Figure 4b). 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 from November 2021 (Accession number: ON676708), indicating they arose and circulated for some months prior to the B.1 epidemic.

Extrapolating back to number-of-suspected-APOBEC3-mutations (y-axis) = 0 provides an estimate of when the first APOBEC3 mutations occurred in the stem of the branch leading to the 2017 epidemic. We estimate this to be 20-Jun-2015 (95% credible intervals of 6-Oct-2014, 3-Mar-2016) which represents the likely time range in which the initial jump to sustained human to human transmission occurred.

Discussion

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 (20). The recent diversity in Clade IIb overwhelmingly consists of TC->TT and GA->AA nucleotide changes, consistent with APOBEC3 enzyme editing, and a recent study has demonstrated APOBEC3F editing during human MPXV infection (52). 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. 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.

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.

If we assume this observed evolution within hMPXV-1 is APOBEC3-driven, this may have implications for its sustained transmission in the human population. 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. The extensive rounds of genome replication that MPXV undertakes in the cytoplasm likely means that most genomes are not affected by APOBEC3 action. Occasionally however a genome, only 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. However the timescale on which this might happen is uncertain and other evolutionary forces such as recombination may act to restore fitness. A further uncertainty arises from the number of genes associated with 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 such genes in VARV, and consequently the loss of function of some MPXV genes through APOBEC3 activity may potentially have adaptive value for the virus (53, 54).

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 (15) have now expanded with the 2022 epidemic B.1 lineage now encompassing 12 sublineages [github.com/mpxv-lineages/lineage-designation]. The rapid and temporally linear accumulation of mutations means that genomic epidemiological models and tools (55, 56), usually employed for RNA viruses, are also applicable to hMPXV-1.

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. All of these lineages (designated as A.2.1, A.2.2, A.2.3 and A.3) can be phylogenetically traced back to the epidemic in Nigeria (Figure 4a). This suggests that sustained human to human transmission is still ongoing outside of recognised MSM networks and stopping transmission in the MSM 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 highly plausible there are other populations that are currently enduring epidemics.

Historically, MPX was considered a zoonotic disease and cases have been treated as independent spillover events with low levels of circulation in the human population. This assumption has informed outbreak control measures as low levels of spillover at the human/reservoir interface would suggest more targeted interventions were necessary rather than a large-scale vaccination campaign. In the light of our observations, this historic assumption may no longer be valid when considering MPX cases and future outbreak control measures should treat any novel zoonotic outbreak of MPXV as having the potential for sustained human-to-human transmission, regardless of whether infections are spreading internationally or within a country that has an endemic reservoir host.

Funding Statement

ÁOT, PL, MAS and AR are supported by Wellcome Trust ARTIC (Collaborators Award 206298/Z/17/Z ARTIC network) & supplement. PL, MAS and AR acknowledge support from the European Research Council (grant agreement no. 725422 – ReservoirDOCS) and National Institutes of Health (R01 AI153044). MW was supported by the David and Lucile Packard Foundation. PL acknowledges support from the Research Foundation - Flanders (Fonds voor Wetenschappelijk Onderzoek - Vlaanderen, G066215N, G0D5117N and G0B9317N) and by HORIZON 2020 EU grant 874850 MOOD.

References

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. S. S. Marennikova, E. M. Seluhina, Susceptibility of some rodent species to monkeypox virus, and course of the infection. *Bull. World Health Organ.* **53**, 13–20 (1976).
3. L. Khodakevich, M. Szczeniowski, Manbu-ma-Disu, Z. Jezek, S. Marennikova, J. Nakano, D. Messinger, The role of squirrels in sustaining monkeypox virus transmission. *Trop. Geogr. Med.* **39**, 115–122 (1987).
4. L. Khodakevich, Z. Jezek, D. Messinger, Monkeypox virus: ecology and public health significance. *Bull. World Health Organ.* **66**, 747–752 (1988).

5. K. Brown, P. A. Leggat, Human Monkeypox: Current State of Knowledge and Implications for the Future. *Trop Med Infect Dis.* 1 (2016), doi:10.3390/tropicalmed1010008.
6. C. L. Hutson, I. K. Damon, Monkeypox virus infections in small animal models for evaluation of anti-poxvirus agents. *Viruses.* 2, 2763–2776 (2010).
7. S. Parker, R. M. Buller, A review of experimental and natural infections of animals with monkeypox virus between 1958 and 2012. *Future Virol.* 8, 129–157 (2013).
8. E. Alakunle, U. Moens, G. Nchinda, M. I. Okeke, Monkeypox Virus in Nigeria: Infection Biology, Epidemiology, and Evolution. *Viruses.* 12 (2020), doi:10.3390/v12111257.
9. L. Khodakevich, Z. Jezek, K. Kinzanzka, Isolation of monkeypox virus from wild squirrel infected in nature. *Lancet.* 1, 98–99 (1986).
10. Centers for Disease Control and Prevention (CDC), Multistate outbreak of monkeypox--Illinois, Indiana, and Wisconsin, 2003. *MMWR Morb. Mortal. Wkly. Rep.* 52, 537–540 (2003).
11. H. A. Thomassen, T. Fuller, S. Asefi-Najafabady, J. A. G. Shiplacoff, P. M. Mulembakani, S. Blumberg, S. C. Johnston, N. K. Kisalu, T. L. Kinkela, J. N. Fair, N. D. Wolfe, R. L. Shongo, M. LeBreton, H. Meyer, L. L. Wright, J.-J. Muyembe, W. Buermann, E. Okitolonda, L. E. Hensley, J. O. Lloyd-Smith, T. B. Smith, A. W. Rimoin, Pathogen-host associations and predicted range shifts of human monkeypox in response to climate change in central Africa. *PLoS One.* 8, e66071 (2013).
12. A. W. Rimoin, P. M. Mulembakani, S. C. Johnston, J. O. Lloyd Smith, N. K. Kisalu, T. L. Kinkela, S. Blumberg, H. A. Thomassen, B. L. Pike, J. N. Fair, N. D. Wolfe, R. L. Shongo, B. S. Graham, P. Formenty, E. Okitolonda, L. E. Hensley, H. Meyer, L. L. Wright, J.-J. Muyembe, 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).
13. T. D. Baker, "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.
14. Z. Jezek, J. H. Nakano, I. Arita, M. Mutombo, M. Szczeniowski, C. Dunn, Serological survey for human monkeypox infections in a selected population in Zaire. *J. Trop. Med. Hyg.* 90, 31–38 (1987).
15. C. Happi, I. Adetifa, P. Mbala, R. Njouom, E. Nakoune, A. Happi, N. Ndodo, O. Ayansola, G. Mboowa, T. Bedford, R. A. Neher, C. Roemer, E. Hodcroft, H. Tegally, Á. O'Toole, A. Rambaut, O. Pybus, M. U. G. Kraemer, E. Wilkinson, J. Isidro, V. Borges, M. Pinto, J. P. Gomes, L. Freitas, P. C. Resende, R. T. C. Lee, S. Maurer-Stroh, C. Baxter, R. Lessells, A. E. Ogbwell, Y. Kebede, S. K. Tessema, T. de Oliveira, Urgent need for a non-discriminatory and non-stigmatizing nomenclature for monkeypox virus. *PLoS Biol.* 20, e3001769 (2022).
16. WHO, "Multi-country outbreak of monkeypox, External situation report" (2022), (available at

<https://www.who.int/publications/m/item/multi-country-outbreak-of-monkeypox--external-situation-report--7---5-october-2022>).

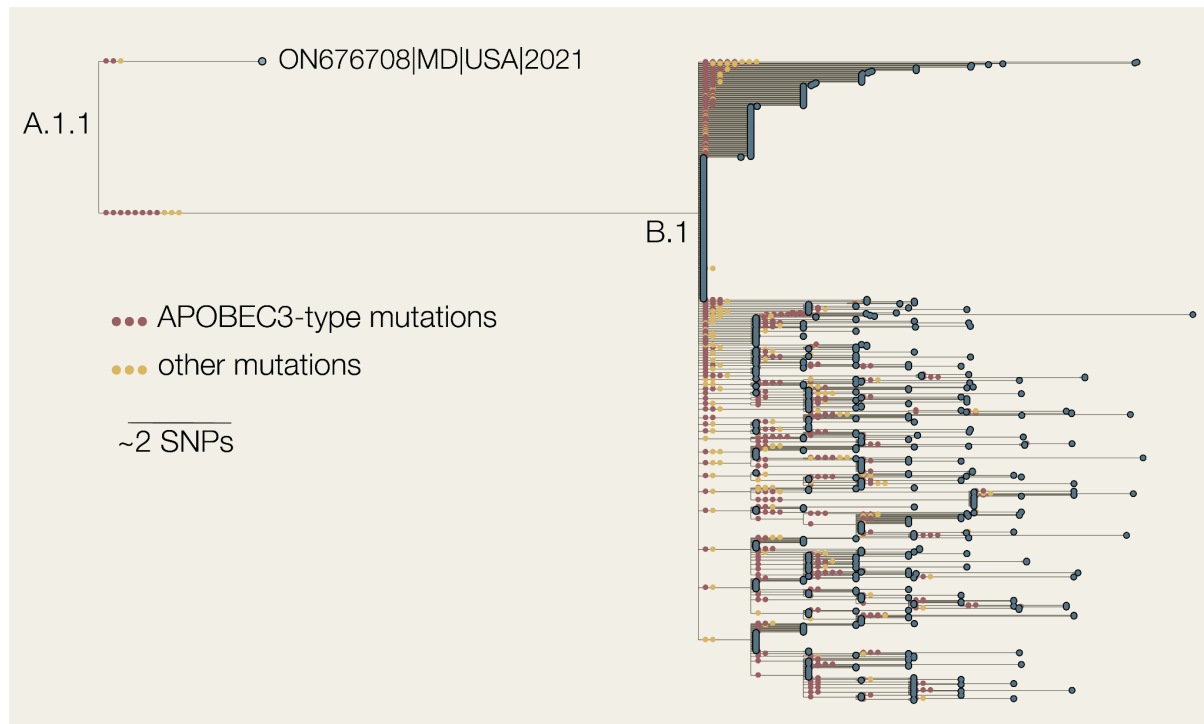
17. J. Isidro, V. Borges, M. Pinto, D. Sobral, J. D. Santos, A. Nunes, V. Mixão, R. Ferreira, D. Santos, S. Duarte, L. Vieira, M. J. Borrego, S. Núncio, I. L. de Carvalho, A. Pelerito, R. Cordeiro, J. P. Gomes, Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. *Nat. Med.* **28**, 1569–1572 (2022).
18. C. M. Gigante, B. Korber, M. H. Seabolt, K. Wilkins, W. Davidson, A. K. Rao, H. Zhao, C. M. Hughes, F. Minhaj, M. A. Waltenburg, J. Theiler, S. Smole, G. R. Gallagher, D. Blythe, R. Myers, J. Schulte, J. Stringer, P. Lee, R. M. Mendoza, L. A. Griffin-Thomas, J. Crain, J. Murray, A. Atkinson, A. H. Gonzalez, J. Nash, D. Batra, I. Damon, J. McQuiston, C. L. Hutson, A. M. McCollum, Y. Li, Multiple lineages of Monkeypox virus detected in the United States, 2021–2022. *bioRxiv* (2022), p. 2022.06.10.495526.
19. C. Firth, A. Kitchen, B. Shapiro, M. a. Suchard, E. C. Holmes, A. Rambaut, Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol. Biol. Evol.* **27**, 2038–2051 (2010).
20. L. V. Patrono, K. Pléh, L. Samuni, M. Ulrich, C. Röthemeier, A. Sachse, S. Muschter, A. Nitsche, E. Couacy-Hymann, C. Boesch, R. M. Wittig, S. Calvignac-Spencer, F. H. Leendertz, Monkeypox virus emergence in wild chimpanzees reveals distinct clinical outcomes and viral diversity. *Nat Microbiol.* **5**, 955–965 (2020).
21. O’toole Á & Rambaut, Initial observations about putative APOBEC3 deaminase editing driving short-term evolution of MPXV since 2017. *Virological* (2022), (available at <https://virological.org/t/830>).
22. Q. Yu, R. König, S. Pillai, K. Chiles, M. Kearney, S. Palmer, D. Richman, J. M. Coffin, N. R. Landau, Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nat. Struct. Mol. Biol.* **11**, 435–442 (2004).
23. Y. Z. McDaniel, D. Wang, R. P. Love, M. B. Adolph, N. Mohammadzadeh, L. Chelico, L. M. Mansky, Deamination hotspots among APOBEC3 family members are defined by both target site sequence context and ssDNA secondary structure. *Nucleic Acids Res.* **48**, 1353–1371 (2020).
24. J. Ratcliff, P. Simmonds, Potential APOBEC-mediated RNA editing of the genomes of SARS-CoV-2 and other coronaviruses and its impact on their longer term evolution. *Virology.* **556**, 62–72 (2021).
25. I. B. Rogozin, L. M. Iyer, L. Liang, G. V. Glazko, V. G. Liston, Y. I. Pavlov, L. Aravind, Z. Pancer, Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat. Immunol.* **8**, 647–656 (2007).
26. M. F. Flajnik, M. Kasahara, Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat. Rev. Genet.* **11**, 47–59 (2010).
27. C. Münk, A. Willemsen, I. G. Bravo, An ancient history of gene duplications, fusions and losses

- in the evolution of APOBEC3 mutators in mammals. *BMC Evol. Biol.* **12**, 71 (2012).
28. R. Suspène, M. Henry, S. Guillot, S. Wain-Hobson, J.-P. Vartanian, Recovery of APOBEC3-edited human immunodeficiency virus G->A hypermutants by differential DNA denaturation PCR. *J. Gen. Virol.* **86**, 125–129 (2005).
 29. H. Chen, C. E. Lilley, Q. Yu, D. V. Lee, J. Chou, I. Narvaiza, N. R. Landau, M. D. Weitzman, APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr. Biol.* **16**, 480–485 (2006).
 30. J.-P. Vartanian, D. Guétard, M. Henry, S. Wain-Hobson, Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science*. **320**, 230–233 (2008).
 31. J.-P. Vartanian, M. Henry, A. Marchio, R. Suspène, M.-M. Aynaud, D. Guétard, M. Cervantes-Gonzalez, C. Battiston, V. Mazzaferro, P. Pineau, A. Dejean, S. Wain-Hobson, Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. *PLoS Pathog.* **6**, e1000928 (2010).
 32. M. Tsuge, C. Noguchi, R. Akiyama, M. Matsushita, K. Kunihiro, S. Tanaka, H. Abe, F. Mitsui, S. Kitamura, T. Hatakeyama, T. Kimura, D. Miki, N. Hiraga, M. Imamura, S. Takahashi, C. N. Hayses, K. Chayama, G to A hypermutation of TT virus. *Virus Res.* **149**, 211–216 (2010).
 33. R. Suspène, M.-M. Aynaud, D. Guétard, M. Henry, G. Eckhoff, A. Marchio, P. Pineau, A. Dejean, J.-P. Vartanian, S. Wain-Hobson, 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).
 34. Z.-G. Peng, Z.-Y. Zhao, Y.-P. Li, Y.-P. Wang, L.-H. Hao, B. Fan, Y.-H. Li, Y.-M. Wang, Y.-Q. Shan, Y.-X. Han, Y.-P. Zhu, J.-R. Li, X.-F. You, Z.-R. Li, J.-D. Jiang, 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).
 35. M. Fehrholz, S. Kendl, C. Prifert, B. Weissbrich, K. Lemon, L. Rennick, P. W. Duprex, B. K. Rima, F. A. Koning, R. K. Holmes, M. H. Malim, J. Schneider-Schaulies, The innate antiviral factor APOBEC3G targets replication of measles, mumps and respiratory syncytial viruses. *J. Gen. Virol.* **93**, 565–576 (2012).
 36. M. Kremer, Y. Suezter, Y. Martinez-Fernandez, C. Münk, G. Sutter, B. S. Schnierle, Vaccinia virus replication is not affected by APOBEC3 family members. *Viol. J.* **3**, 86 (2006).
 37. B. Moss, Poxvirus DNA replication. *Cold Spring Harb. Perspect. Biol.* **5** (2013), doi:10.1101/cshperspect.a010199.
 38. B. Moss, Poxviridae: the viruses and their replication. *Fields Virology*, 4th DM Knipe & PM Howley (2001).
 39. S. R. Bidgood, J. Mercer, Cloak and Dagger: Alternative Immune Evasion and Modulation Strategies of Poxviruses. *Viruses*. **7**, 4800–4825 (2015).

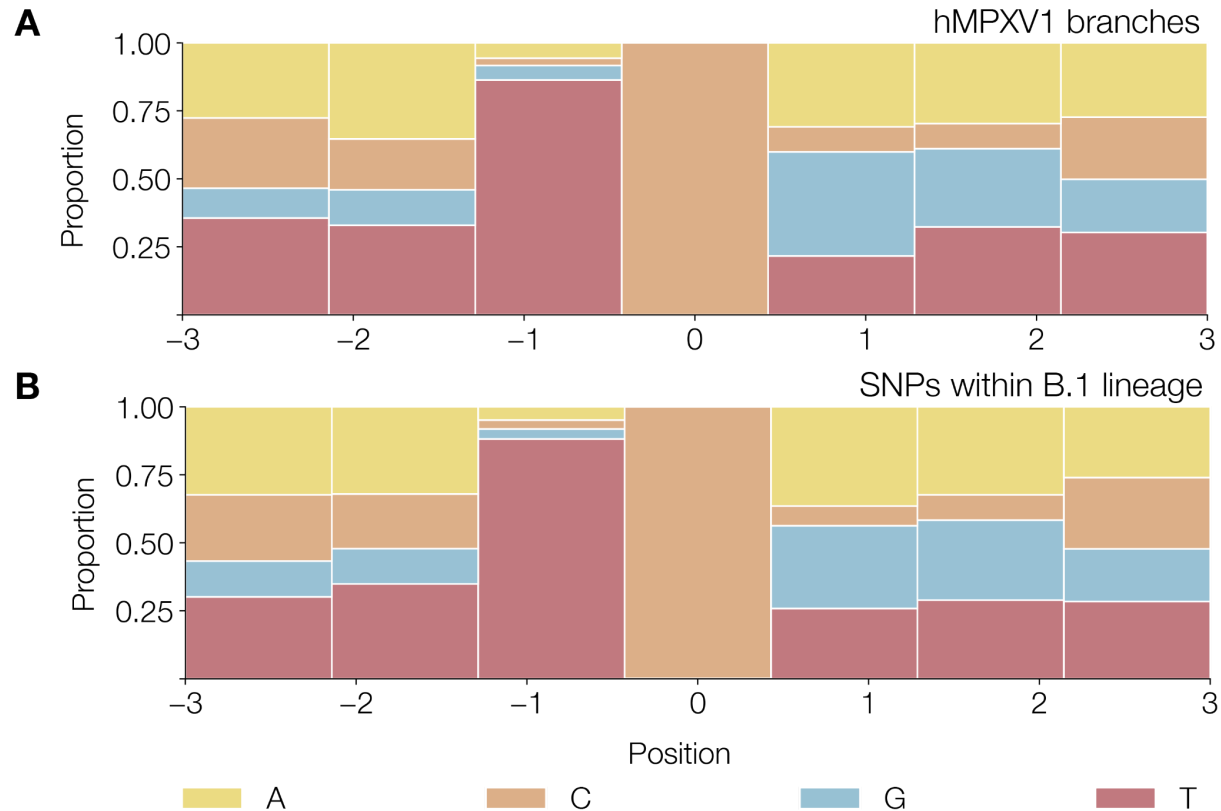
40. G. L. Smith, C. T. O. Benfield, C. Maluquer de Motes, M. Mazzon, S. W. J. Ember, B. J. Ferguson, R. P. Sumner, Vaccinia virus immune evasion: mechanisms, virulence and immunogenicity. *J. Gen. Virol.* **94**, 2367–2392 (2013).
41. S. G. Conticello, C. J. F. Thomas, S. K. Petersen-Mahrt, M. S. Neuberger, Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol. Biol. Evol.* **22**, 367–377 (2005).
42. A. Krishnan, L. M. Iyer, S. J. Holland, T. Boehm, L. Aravind, 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).
43. L. Yang, M. Emerman, H. S. Malik, R. N. McLaughlin Jr, Retrocopying expands the functional repertoire of APOBEC3 antiviral proteins in primates. *Elife.* **9** (2020), doi:10.7554/eLife.58436.
44. F. Anwar, M. P. Davenport, D. Ebrahimi, Footprint of APOBEC3 on the genome of human retroelements. *J. Virol.* **87**, 8195–8204 (2013).
45. R. S. LaRue, S. R. Jónsson, K. A. T. Silverstein, M. Lajoie, D. Bertrand, N. El-Mabrouk, I. Hötzel, V. Andrésdóttir, T. P. L. Smith, R. S. Harris, The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. *BMC Mol. Biol.* **9**, 104 (2008).
46. M. C. Mikl, I. N. Watt, M. Lu, W. Reik, S. L. Davies, M. S. Neuberger, C. Rada, Mice deficient in APOBEC2 and APOBEC3. *Mol. Cell. Biol.* **25**, 7270–7277 (2005).
47. A. I. Su, M. P. Cooke, K. A. Ching, Y. Hakak, J. R. Walker, T. Wiltshire, A. P. Orth, R. G. Vega, L. M. Sapinoso, A. Moqrich, A. Patapoutian, G. M. Hampton, P. G. Schultz, J. B. Hogenesch, Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 4465–4470 (2002).
48. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* **34**, 3094–3100 (2018).
49. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, R. Lanfear, Corrigendum to: IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 2461 (2020).
50. M. Hasegawa, H. Kishino, T. Yano, Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).
51. M. Kimura, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120 (1980).
52. R. Suspène, K. A. Raymond, L. Boutin, S. Guillier, F. Lemoine, O. Ferraris, J.-N. Tournier, F. Iseni, E. Simon-Lorière, J.-P. Vartanian, APOBEC3F is the main source of editing identified during the 2022 outbreak of human monkeypox virus. *bioRxiv* (2023), p. 2023.01.06.522979.
53. R. C. Hendrickson, C. Wang, E. L. Hatcher, E. J. Lefkowitz, Orthopoxvirus genome evolution:

the role of gene loss. *Viruses*. 2, 1933–1967 (2010).

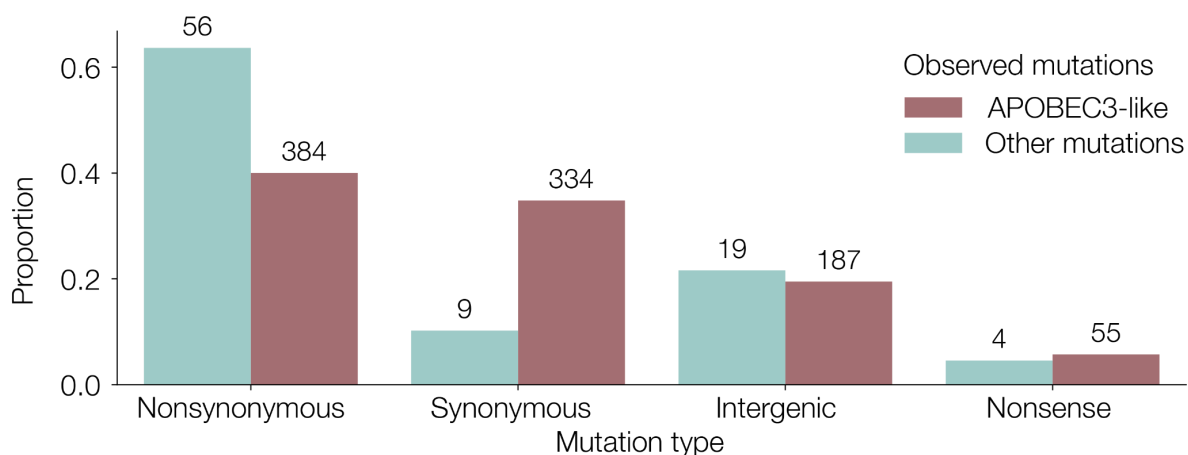
54. B. Mühlemann, L. Vinner, A. Margaryan, H. Wilhelmson, C. de la Fuente Castro, M. E. Allentoft, P. de Barros Damgaard, A. J. Hansen, S. Holtsmark Nielsen, L. M. Strand, J. Bill, A. Buzhilova, T. Pushkina, C. Falys, V. Khartanovich, V. Moiseyev, M. L. S. Jørkov, P. Østergaard Sørensen, Y. Magnusson, I. Gustin, H. Schroeder, G. Sutter, G. L. Smith, C. Drosten, R. A. M. Fouchier, D. J. Smith, E. Willerslev, T. C. Jones, M. Sikora, Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. *Science*. 369 (2020), doi:10.1126/science.aaw8977.
55. M. A. Suchard, P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, A. Rambaut, Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 4, vey016 (2018).
56. P. Sagulenko, V. Puller, R. A. Neher, TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol.* 4, vex042 (2018).



Supplementary Figure 1 | Phylogeny containing all high-quality hMPXV1 B.1 genomes shared on Genbank as of 2021-08-22, with the 2021 Maryland sample (ON676708) used as an outgroup (n=769).

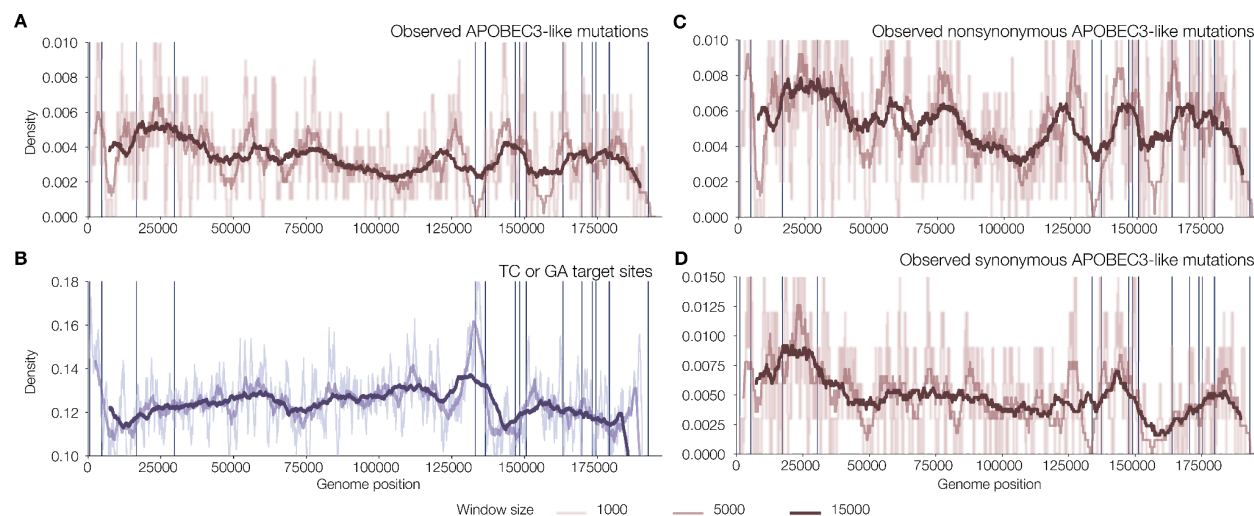


Supplementary Figure 2 | A. Observed heptamers of C->T or G->A mutated sites in 337 of 383 unambiguous mutations that occur along branches in the MPXV phylogeny in Figure 4A, excluding the mutations that occur on the 1971 branch. Heptamers associated with G->A mutations have been reverse-complemented to reflect deamination on the negative strand. B. Observed heptamers of C->T or G->A mutations in the B.1 lineage of Clade IIb (677 of 836 mutations), which we know is being transmitted human-to-human.



Supplementary Figure 3 | Observed mutations across the Clade IIb phylogeny (including genomes in Figure 4 and Supplementary Figure 1, except 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), and coloured by whether they have occurred in an APOBEC3-like context or not.



Supplementary Figure 4 | The density (moving average linear convolution with window sizes of 1000, 2500 and 15000) of A) observed APOBEC-like mutations across the MPXV genome, B) available APOBEC3 target sites in the Clade IIb reference genome (Accession NC_063383), C) observed APOBEC-like mutations that result in synonymous changes and D) observed APOBEC-like mutations that result in nonsynonymous changes. We see the same difference between target density for both observed synonymous and non-synonymous distributions (Kolgorov-Smirnov test statistic synonymous vs target = 0.1133, P-value=0.0003; Kolgorov-Smirnov test statistic non-synonymous vs target = 0.067, P-value=0.059). Repetitive masked regions indicated by vertical bars.

Supplementary Table 1: Authors, Institute and Accession numbers for sequences in analysis of hMPXV1.

| Authors | Institute | Accessions |
|---|---|----------------------|
| Nakazawa Y, Mauldin MR, Emerson GL, Reynolds MG, Lash RR, Gao J, Zhao H, Li Y, Muyembe JJ, Kingebeni PM, Wemakoy O, Malekani J, Karem KL, Damon IK, Carroll DS. | Poxvirus Program, Centers for Disease Control and Prevention, 1600 Clifton Rd. NE, Atlanta, GA 30333, USA | KJ642615 KJ642617 |
| Faye O, Pratt CB, Faye M, Fall G, Chitty JA, Diagne MM, Wiley MR, Yinka-Ogunleye AF, Aruna S, Etebu EN, Aworabhi N, Ogoina D, Numbere W, Mba N, Palacios G, Sall AA, Ihekweazu C. | Center for Genome Sciences, USAMRIID, 1425 Porter Street, Fort Detrick, Frederick, MD 21701, USA | MG693724 |

| | | |
|---|---|---|
| Yinka-Ogunleye A, Aruna O, Dalhat M, Ogoina D, McCollum A, Disu Y, Mamadu I, Akinpelu A, Ahmad A, Burga J, Ndoreraho A, Nkunuzimana E, Manneh L, Mohammed A, Adeoye O, Tom-Aba D, Silenou B, Ipadeola O, Saleh M, Adeyemo A, Nwadiutor I, Aworabhi N, Uke P, John D, Wakama P, Reynolds M, Mauldin MR, Doty J, Wilkins K, Musa J, Khalakdina A, Adedeji A, Mba N, Ojo O, Krause G, Ihekweazu C | NCEZID/DHCPP/PRB, CDC, 1600 Clifton Rd, Atlanta, GA 30333, USA | MK783027 MK783028 MK783029 MK783031 MK783030 MK783032 MK783033 |
| Cohen Gihon,I., Israeli,O., Shifman,O., Erez,N., Melamed,S., Paran,N., Beth-Din,A. and Zvi,A. | Institute for Biological Research, Reuven 24, Ness Ziona 74100, Israel | MN648051 |
| Yong SEF, Ng OT, Ho ZJM, Mak TM, Marimuthu K, Vasoo S, Yeo TW, Ng YK, Cui L, Ferdous Z, Chia PY, Aw BJW, Manaus CM, Low CKK, Chan G, Peh X, Lim PL, Chow LPA, Chan M, Lee VJM, Lin RTP, Heng MKD, Leo YS | National Public Health Laboratory, National Centre for Infectious Diseases, 16 Jln Tan Tock Seng, Singapore, Singapore 308442, Singapore | MT250197 |
| Mauldin MR, McCollum AM, Nakazawa YJ, Mandra A, Whitehouse ER, Davidson W, Zhao H, Gao J, Li Y, Doty J, Yinka-Ogunleye A, Akinpelu A, Aruna O, Naidoo D, Lewandowski K, Afrough B, Graham V, Aarons E, Hewson R, Vipond R, Dunning J, Chand M, Brown C, Cohen-Gihon I, Erez N, Shifman O, Israeli O, Sharon M, Schwartz E, Beth-Din A, Zvi A, Mak TM, Ng YK, Cui L, Lin RTP, Olson VA, Brooks T, Paran N, Ihekweazu C, Reynolds MG. | NCEZID/DHCPP/PRB, CDC, 1600 Clifton Rd, Atlanta, GA 30333, USA | MT903342 MT903338 MT903339 MT903341 MT903343 MT903344 MT903345 NC_063383 |
| Gigante,C.M., Myers,R., Seabolt,M.H., Wilkins,K., McCollum,A., Hutson,C., Davidson,W., Rao,A., Blythe,D. and Li,Y. | Division of High-Consequence Pathogens and Pathology, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30329, USA | ON563414 ON674051 ON675438 ON676707 ON676708 |
| Pragya Yadav, Rima Sahay, Anita Aich Shete, Sreelekshmy Mohandas, Priya Abraham | Indian Council of Medical Research-National Institute of Virology, Microbial Containment Complex, 130/1 ,Sus Road, Pashan, Pune 411021, India | EPI_ISL_13953610 EPI_ISL_13953611 |
| Pilailuk Okada; Siripaporn Phuygun; Nuttida Thongpramul; Thanutsapa Thanadachakul; Kazuhisa Okada; | Bangkok Hospital Phuket, 2/1 Hongyok Utis Road, Muang District, Phuket, 83000, Thailand National | EPI_ISL_13983888 EPI_ISL_14295679 |

| | | |
|--|--|--|
| Archawin Rojanawiwat; Chakkarat Pitayawonganon; Supakit Sirlak | Institute of Health, Department of Medical Sciences, Ministry of Public Health, Thailand, 88/7 Tiwanon Road, Amphoe Muang, Nonthaburi 11000, Thailand | |
| Pilailuk Okada; Siripaporn Phuygun; Nuttida Thongpramul; Thanutsapa Thanadachakul; Kazuhisa Okada; Archawin Rojanawiwat; Chakkarat Pitayawonganon; Supakit Sirlak | Thai Red Cross Emerging Infectious Diseases Clinical Center and Faculty of Medicine, Chulalongkorn University | EPI_ISL_14011193 |
| Ndodo,N., Ashcroft,J., Lewandowski,K., Yinka-Ogunleye,A., Chukwu,C., Ahmad,A., King,D., Akinpelu,A., Maluquer de Motes,C., Ribeca,P., Summer,R.P., Rambaut,A., Chester,M., Maishman,T., Babatunde,O., Mba,N., Babatunde,O., Aruna,O., Pullan,S.T., Gannon,B., Brown,C., Ihekweazu,C., Adetifa,I. and Ulaeto,D.O. | Nigeria Centre for Disease Control, National Reference Laboratory, Gaduwa, Federal Capital Territory, Abuja Defence Science and Technology Laboratory, Porton Down, Salisbury SP4 0JQ, Wiltshire | OP612674 OP612675 OP612676 OP612677 OP612678 OP612679 OP612680 OP612681 OP612682 OP612683 OP612684 OP612685 OP612686 OP612687 OP612688 OP612689 OP612690 OP612691 |

Supplementary Table 2: Authors, institute and accession numbers for sequences in Supplementary Figure 1.

| Authors | Institute | Accession numbers |
|---|--|--|
| Zakotnik,S., Vljaj,D., Suljic,A., Zorec,T.M., Skubic,C., Rozman,D., Korva,M., Poljak,M. and Avsic Zupanc,T. | Laboratory for Diagnostics of Zoonoses and WHO Centre, Institute of Microbiology and Immunology, Faculty of Medicine, University of Ljubljana, Zaloska 4, Ljubljana 1000, Slovenia | ON609725;ON631241;ON754987;ON754986;ON754985;ON754984;ON838178 |
| Groves,N., Osman,K.L., Lewandowski,K.S., Pullan,S.T., Carter,D.P., Crook,J.M., Myers,R., Vipond,R. and Chand,M. | Research and Evaluation, UKHSA, Porton Down, Salisbury, Wiltshire SP4 0JG, UK | OP205140;OP205139;OP205138;OP205137;OP205136;OP205135;OP205134;OP205133;OP205132;OP205131;OP205130;OP205129;OP205128;OP205127;OP205126;OP205125;OP205124;OP205123;OP205122;OP205121;OP205120;OP205119;OP205118;OP205117;OP205116;OP205115;OP205114;OP205113;OP205112;OP205111;OP205110;OP205109;OP205108;OP205107;OP205106;OP205105;OP205104;OP205103;OP205102;OP205101;OP20 |

| Authors | Institute | Accession numbers |
|--|--|---|
| | | 5100;OP205099;OP205098;OP205097;OP205096;OP205095;OP205094;OP205093;OP205092;OP205091;OP205090;OP205089;OP205088;OP205087;OP205086;OP205085;OP205084;OP205083;OP205082;OP205081;OP205080;OP205079;OP205078;OP205077;OP205076;OP205075;OP205074;OP205073;OP205072;OP205071;OP205070;OP205069;OP205068;OP205067;OP022170;OP022171;ON619838;ON619837;ON619836;ON619835 |
| Chmel,M., Pajer,P., Nagy,A., Zlamal,M., Jirincova,H., Dresler,J. and Bartos,O. | Military Health Institute in Prague, Military Health Institute, U Vojenske nemocnice 1200/1, Prague 16200, Czech Republic | ON983168 |
| Israeli,O., Guedj-Dana,Y., Lazar,S., Shifman,O., Erez,N., Weiss,S., Paran,N., Israely,T., Schuster,O., Zvi,A., Beth-Din,A. and Cohen Gihon,I. | Biochemistry and Molecular Genetics, Israel Institute for Biological Research, Reuven 24, Ness Ziona 7410001, Israel | ON649879 |
| Licastro,D., DeGasperi,M., Negri,C., Piscianz,E., Koncan,R., Dal Monego,S., Segat,L. and D'Agaro,P. | Genomics and Epigenomics, AREA Science Park, km 163,5 Strada Statale 14, Basovizza, Friuli Venezia Giulia 34149, Italy | ON644344 |
| Hammerschlag,Y., MacLeod,G., Papadakis,G., Adnan Sanchez,A., Druce,J., Savic,I., Taiaroa,G., Roberts,J., Caly,L., Williamson,D.A., Cheng,A.C. and McMahon,J.H. | Victorian Infectious Diseases Reference Laboratory Doherty Institute 792 Elizabeth Street Melbourne Victoria 3000 Australia | ON631963 |
| Rhie,G.-E. | Division of High-risk Pathogens Korea Disease Control and Prevention Agency 187 Osongsaengmyeong2-ro Osong-eup Heungdeok-gu Cheongju Chungcheongbuk-do 28159 Republic of Korea | OP204857 |
| De Baetselier,I., Van Dijck,C., Kenyon,C., Coppens,J., Smet,H., de Block,T., Coppens,S., Vanroye,F., Bugert,J., Girt,P., Liesenborghs,L., Selhorst,P., Arien,K., Van den Bossche,D., Florence,E., Rezende,A.M., Vercauteren,K. and Van Esbroeck,M. | Department of Clinical Sciences Institute of Tropical Medicine Nationalestraat 155 Antwerp Antwerp 2000 Belgium | ON950045;OP144217;OP144216;OP144215;OP144214;OP144213;OP144212;OP144211 |
| Sereewit,J., Xie,H., Roychoudhury,P. and Greninger,A.L. | Laboratory Medicine UW Virology 1616 Eastlake Ave East Suite 320 Seattle WA 98102 USA | OP055801;OP123055;OP123054;OP123053;OP123052;OP123051;OP123050;OP123049;OP123048;OP123047;OP123046;OP123045;OP123044;OP123043;OP123042;OP123041;OP123040;OP055809;OP055808;OP055807;OP055806;OP055805;OP055804;OP055803;OP055802;OP055800;OP257267; |

| Authors | Institute | Accession numbers |
|---|---|---|
| | | OP257266;OP257265;OP257264;OP257263;OP257262;OP257261;OP257260;OP257259;OP257258;OP257257;OP257256;OP257255;OP257254;OP257253;OP257252;OP257251;OP257250;OP257249;OP257248;OP257247;OP257246;OP257245;OP257244;OP257243;OP184765;OP184764;OP184763;OP184762;OP184761;OP184760;OP169346;OP169345;OP169344;OP169343;OP169342;OP169341;OP169340;OP169339;OP169338;OP169337;OP169336 |
| Gigante,C.M., Hughes,S., Seabolt,M.H., Zhao,H., Wilkins,K., Respress,J., Howard,D., Batra,D., McCollum,A., Hutson,C., Davidson,W., Rao,A., Baumgartner,J. and Li,Y. | Division of High-Consequence Pathogens and Pathology, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30329, USA | OP185716;OP185715;OP185714;OP185713;OP225967;OP225965;OP185712;OP225969;OP225966;OP225964;OP225962;OP185711;OP185710;OP225960;OP185709;ON959136;ON959135;ON959134;ON959133;ON959132;ON959131;ON954773;ON676708;ON676706;ON676705;ON676704;ON676703;ON563414;OP225968;OP225963;OP225961;OP185717;OP150927;OP150926;OP150925;OP150924;OP150923 |
| Duggan,A., Hole,D., Knox,N., Yadav,C., Haidl,E., Chapel,M., Domselaar,G.V., Jolly,G., Audet,J., Fernando,L., Antonation,K., Safronetz,D., Hagan,M., Griffiths,E., Leung,A., Graham,M., Peters,G., Go,A., Laminman,V., Kaplen,B., Eshaghi,A., Gubbay,J.B., Hasso,M., Marchand-Austin,A., Olsha,R. and Patel,S.N. | Public Health Agency of Canada, National Microbiology Laboratory, 1015 Arlington Street, Winnipeg, MB R3E 3R2, Canada | OP226139;ON880519;ON880549;ON880548;ON880547;ON880546;ON880545;ON880544;ON880543;ON880542;ON880541;ON880540;ON880539;ON880538;ON880537;ON880536;ON880535;ON880534;ON880533;ON880532;ON880531;ON880530;ON880529;ON880528;ON880527;ON880526;ON880525;ON880524;ON880523;ON880522;ON880521;ON880520;OP062236;OP062234;OP062235;OP062233;OP062232;OP062231;OP062230;OP062229;OP013017;OP013016;OP013015;OP013014;OP013013;OP013012;OP013011;OP013010;OP013009;OP013008;OP013007;OP013006;OP013005;OP013004;OP013003;OP013002;OP013001;ON983167;ON983166;ON983165;ON983164;ON983163;ON983162;ON983161;ON983160;ON983159;ON880518;ON880517;ON880516;ON880515;ON880514;ON880513;ON880512;ON880511;ON880510;ON880509;ON880508;ON880507;ON880506;ON880505 |
| Brinkmann,A., Kohl,C., Uddin,S., Pape,K., Schrick,L., Michel,J., Pfaefflin,F., Schaade,L. and Nitsche,A. | Centre for Biological Threats, Highly Pathogenic Viruses, Robert Koch Institute, Seestr 10, Berlin, Berlin 13353, Germany | OP263636;OP263635;OP263634;OP263633;OP263632;OP263631;OP263630;OP263629;OP263628;OP263627;OP263626;ON853682;ON853681;ON85 |

Authors

Institute

Accession numbers

3680;ON853679;ON853678;ON853677;ON853676;ON853675;ON853674;ON853673;ON853672;ON853671;ON853670;ON853669;ON853668;ON853667;ON853666;ON853665;ON853664;ON853663;ON853662;ON853661;ON853660;ON853659;ON853658;ON853657;ON853656;ON813267;ON813266;ON813265;ON813264;ON813263;ON813262;ON813261;ON813260;ON813259;ON813258;ON755255;ON755251;ON755249;ON755239;ON755238;ON694341;ON682268;ON755256;ON755254;ON755253;ON755252;ON755250;ON755248;ON755247;ON755246;ON755245;ON755244;ON755243;ON755242;ON755241;ON755240;ON755237;ON755236;ON755235;ON755234;ON755233;ON755232;ON755231;ON682270;ON682269;ON682267;ON682266;ON694342;ON694340;ON694339;ON694338;ON694337;ON694336;OP215288;OP215287;OP215286;OP215285;OP215284;OP215283;OP215282;OP215281;OP215280;OP215279;OP215278;OP215277;OP215276;OP215275;OP215274;OP215273;OP215268;OP215267;OP215266;OP215265;OP215264;OP215263;OP215262;OP215261;OP215260;OP215259;OP215258;OP215257;OP215256;OP215255;OP215254;OP215253;OP215252;OP215251;OP215250;OP215249;OP215248;OP215247;OP215246;OP215245;OP215244;OP215243;OP215242;OP215241;OP215240;OP215239;OP215238;OP215237;OP215236;OP215235;OP215234;OP215233;OP215232;OP215231;OP215230;OP018607;OP018606;OP018605;OP018604;OP018603;OP018602;OP018601;OP018600;OP018599;OP018598;OP018597;OP018596;OP018595;OP018594;OP018593;OP018592;ON959177;ON959176;ON959175;ON959174;ON959173;ON959172;ON959171;ON959170;ON959169;ON959168;ON959167;ON959166;ON959165;ON959164;ON959163;ON959162;ON959161;ON959160;ON959159;ON959158;ON959157;ON959156;ON959155;ON959154;ON959153;ON959152;ON929091;ON929090;ON929089;ON929088;ON929087;ON929086;ON929085;ON929084;ON929083;ON929082;ON929081;ON929080;ON929079;ON929078;ON929

| Authors | Institute | Accession numbers |
|---|---|--|
| | | 077;ON929076;ON929075;ON929074;ON929073;ON929072;ON929071;ON929070;ON929069;ON929068;ON929067;ON929066;ON929065;ON929064;ON929063;ON929062;ON929061;ON929060;OP215272;OP215271;OP215270;OP215269;OP215229;OP215228;OP215227;OP215226;OP215225;OP215224;OP215223;OP215222;OP215221;OP215220;OP215219;OP215218;OP215217;OP215216;OP215215;OP215214;OP215213;OP215212;OP215211;OP215210;OP215209;OP215208;OP215207;OP018591;OP018590;OP018589;OP018588;ON959151;ON959150;ON959149;ON929059;ON929058;ON929057;OP215206;OP215205;OP215204;ON853655;ON853654;ON853653;ON853652;ON853651;ON853650;ON853649;ON813253;ON813252;ON694331;ON694330;ON694329;ON637938;ON813257;ON813256;ON813255;ON813254;ON682265;ON682264;ON694335;ON694334;ON694333;ON694332;ON637939;ON813251;ON682263 |
| Laubscher,F., Chudzinski,V., Schibler,M., Kaiser,L. and Renzoni,A. | Laboratory of Virology, University Hospitals of Geneva, Rue Gabrielle-Perret-Gentil 4, Geneva, Geneva 1205, Switzerland | ON622720;ON595760 |
| Knox,N., Hole,D., Duggan,A., Yadav,C., Haidl,E., Chapel,M., Graham,M., Domselaar,G.V., Jolly,G., Audet,J., Fernando,L., Antonation,K., Hagan,M., Griffiths,E., Leung,A., Safronetz,D., Eshaghi,A., Gubbay,J.B., Hasso,M., Marchand-Austin,A., Olsha,R. and Patel,S.N. | Public Health Agency of Canada, National Microbiology Laboratory, 1015 Arlington Street, Winnipeg, MB R3E 3R2, Canada | ON803444;ON803443;ON803442;ON803441;ON803440;ON803439;ON803438;ON803437;ON803436;ON803435;ON803434;ON803433;ON803432;ON803431;ON803430;ON803429;ON803428;ON803427;ON803426;ON803425;ON803424;ON803423;ON803422;ON803421;ON803420;ON803419;ON803418;ON803417;ON803416;ON803415;ON803414;ON803413 |
| Palomino-Cabrera,R., Penas-Utrilla,D., Buenestado-Serrano,S., Perez-Lago,L., Herranz Martin,M., Veintimilla,C., Catalan,P., Munoz,P. and Garcia de Viedma,D. | Microbial Genomics, HGUGM, Doctor Esquerdo 46, Madrid, Madrid 28007, Spain | ON927243 |
| Lin,J.-H., Chiu,S.-C., Huang,H.-I., Huang,W.-L., Fann,W.-B., Hsieh,P.-Y. and Yang,J.-Y. | Center of Diagnostics and Vaccine Development Centers for Disease Control Taiwan No. 161 Kuyang St. Taipei 115161 Taiwan | ON918656 |
| Coletti,T.M., Ghilardi,F., Khan,M.J., Claro,I.M., Valenca,I.N., Faria,N.R. and Sabino,E.C. | School of Public Health, Imperial College London, South Kensington Campus, London, United Kingdom SW7 2AZ, United Kingdom | ON880413 |

| Authors | Institute | Accession numbers |
|--|---|----------------------------------|
| Alcoba-Florez,J., Munoz-Barrera,A., Ciuffreda,L., Rodriguez-Perez,H., Rubio-Rodriguez,L.A., Gil-Campesino,H., Garcia-Martinez de Artola,D., Inigo-Campos,A., Díez-Gil,O., Gonzalez-Montelongo,R., Valenzuela-Fernandez,A., Lorenzo-Salazar,J.M. and Flores,C. | Genomics Division, Instituto Tecnologico y de Energias Renovables (ITER) | ON782055;ON782054 |
| Jarjaval,F., Nolent,F., Criqui,A., Chapus,C., Lamer,O., Ferraris,O. and Gorge,O. | D.DNRBC/D.2MI/U.BACT, IRBA, 1, place Generale Valerie Andre, Bretigny/Orge, IdF 91223, FRANCE | ON755040;ON755039 |
| Claro,I.M., de Lima,E.L., Romano,C.M., Candido,D.S., Lindoso,J.A.L., Barra,L.A.C., Borges,L.M.S., Medeiros,L.A., Tomishige,M.Y.S., Ramundo,M.S., Moutinho,T., da Silva,A.J.D., Rodrigues,C.C.M., de Azevedo,L.C.F., Villas-Boas,L.S., da Silva,C.A.M., Coletti,T.M., O'Toole,A., Quick,J., Loman,N., Rambaut,A., Faria,N.R., Figueiredo-Mello,C. and Sabino,E.C. | School of Public Health Imperial College London South Kensington Campus London United Kingdom SW7 2AZ United Kingdom | ON751962 |
| Buenestado-Serrano,S., Palomino-Cabrera,R., Penas-Utrilla,D., Rodriguez-Grande,J., Sola Campoy,P., Perez-Lago,L., Rodriguez-Grande,C., Herranz Martin,M., Suarez,J., Catalan,P., Munoz,P. and Garcia de Viedma,D. | Microbial Genomics, HGUGM, Doctor Esquerdo 46, Madrid, Madrid 28007, Spain | ON745225;ON720849;ON720848 |
| Science Park | Science Park of Comenius University of Bratislava, Ilkovicova 8, 841 04 Bratislava, Slovakia | OX261748.1;OX261749.1;OX248696.1 |
| Alcolea-Medina,A., Charalampous,T., Snell,L.B., Batra,R. and Edgeworth,J.D. | Centre for Clinical Infection & Diagnostics Research, King's College London, St Thomas Hospital, London SE1 7EH, UK | ON645312 |
| Martinez-Puchol,S., Coello,A., Bordoy,A.E., Soler,L., Panisello,D., Gonzalez-Gomez,S., Clara,G., Paris de Leon,A., Not,A., Hernandez,A., Bofill-Mas,S., Saludes,V., Blanco,I., Martro,E. and Cardona,P.-J. | Microbiology, Hospital Universitari Germans Trias i Pujol, Canyet s/n, Badalona, Catalonia 08916, Spain | ON622718 |
| Oude Munnink,B.B., Boter,M., Wellers,B., Molenkamp,R., Sikkema,R.S. and Koopmans,M. | Public Health Virology Erasmus Medical Centre 's-Gravendijkwal 230 Rotterdam 3015 CE The Netherlands | ON615424 |
| Gruber,C.E.M., Rueca,M., Gramigna,G., Carletti,F., Butera,O., Fabeni,L., Specchiarello,E., Meschi,S., Colavita,F., Minosse,C., Francalancia,M., Lapa,D., | Laboratory of Virology INMI Lazzaro Spallanzani IRCCS via Portuense 292 Rome 00149 Italy | ON614676 |

| Authors | Institute | Accession numbers |
|---|--|--|
| Garbuglia,A.R. and Giombini,E. | | |
| Antwerpen,M.H., Lang,D., Zange,S., Walter,M.C. and Woelfel,R. | Microbiol Genomics and Bioinformatics Bundeswehr Institute of Microbiology Neuherbergstr. 11 Munich 80937 Germany | ON568298 |
| Tegomoh,B., Cross,S.T., Chapman,R.C., Bernhard,K., McCutchen,E.L., Fauver,J.R., Pratt,C.B., Warden,D.E., Iwen,P.C., Donahue,M. and Wiley,M.R. | MPXV Nebraska Pathogen Genomics Consortium. Environmental Agricultural and Occupational Health University of Nebraska Medical Center 984388 Nebraska Medical Center Omaha NE 68198-4388 USA | OP171925;OP171924;OP171923;OP1 71922;OP171921;OP171920;OP1719 19 |
| Isidro,J., Borges,V., Pinto,M., Sobral,D., Santos,J., Nunes,A., Mixao,V., Ferreira,R., Santos,D., Duarte,S., Vieira,L., Borrego,M.J., Nuncio,S., Lopes de Carvalho,I., Pelerito,A., Cordeiro,R. and Gomes,J.P. | Department of Infectious Diseases, National Institute of Health Doutor Ricardo Jorge, Portugal (INSA | OP245306.1;OP245306.1;OP245307. 1;OP245308.1;OP245309.1;OP24531 0.2;OP245311.1;OP245312.1;OP2453 13.1;OP245314.1;OP245315.1;OP245 316.1;OP245317.1;OP245318.1;OP24 5319.1;OP245320.1;OP245321.1;OP2 45322.1;OP245323.1;OP245324.1;OP 245325.1;OP245326.1;OP245327.1;O P245328.1;OP245329.1;OP245330.1; OP245331.1;OP245332.1;OP245333. 1;OP245334.1;OP245335.1;OP24533 6.1;OP245337.1;OP245338.1;OP2453 39.1;OP245340.1;OP245341.1;OP245 342.1;OP245343.1;OP245344.2;OP24 5345.1;OP245346.1;OP245347.1;OP2 45348.1;OP245349.1;OP245350.1;OP 245351.2;OP245352.2;OP245353.1;O P245354.1;OP245355.1;OP245356.1; OP245357.1;OP245358.1;OP245359. 1;OP245360.1;OP245361.1;ON84316 3.1;ON843164.1;ON843165.1;ON8431 66.1;ON843167.1;ON843168.1;ON843 169.1;ON843170.1;ON843171.1;ON84 3172.1;ON843173.1;ON843174.1;ON8 43175.1;ON843176.1;ON843177.1;ON 843178.1;ON843179.1;ON843180.1;O N843181.1;ON843182.1;ON649708.1; ON649709.1;ON649710.1;ON649711. 1;ON649712.1;ON649713.1;ON64971 4.1;ON649715.1;ON649716.1;ON6497 17.1;ON649718.1;ON649719.1;ON649 720.1;ON649721.1;ON649722.1;ON64 9723.1;ON649724.1;ON649725.1;ON5 85029.1;ON585030.1;ON585031.1;ON 585032.1;ON585033.1;ON585034.1;O N585035.1;ON585036.1;ON585037.1; ON585038.1 |
| Tutu van Furth AM, van der Kuip M, van Els AL, Fievez LC, van Rijkevorschel GG, van den Ouden A, Jonges M and Welkers MR. | Medical Microbiology & Infection Prevention, Amsterdam Medical Centres location AMC, Meibergdreef 9, Amsterdam 1105AZ, Netherlands | OP160532.1 |

| Authors | Institute | Accession numbers |
|--|--|--|
| Camp,J.V., Redlberger-Fritz,M. and Aberle,S.W. | Center for Virology, Medical University of Vienna, Kinderspitalgasse 15, Vienna 1090, Austria | OP133004.1;OP133005.1;OP133006.1;OP133007.1;OP019275.1;OP019276.1;OP019277.1 |
| Chan,W.Y., Mtshali,P.S., Grobbelaar,A., Moolla,N., Mohale,T., Lowe,M., Du Plessis,M., Ismail,A. and Weyer,J. | Sequencing core facility, National Institute for Communicable Disease, 1 Modderfontein Road, Johannesburg, Gauteng 2131, South Africa | ON918611.1;ON927248.1 |
| Kant,R., Smura,T., Vauhkonen,H., Vapalahti,O. and Sironen,T. | Department of Virology, Faculty of Medicine, University of Helsinki, Hartmaninkatu 3, Helsinki, Helsinki 00014, Finland | ON959143.1;ON782021.1;ON782022.1 |
| Vanmechelen,B., Wawina-Bokalanga,T., Logist,A.-S., Sinnesael,R., Ysebaert,L., Verlinden,J., Van Holm,B., Bloemen,M. and Maes,P. | Microbiology Immunology and Transplantation KU Leuven Rega Institute Herestraat 49 box 1040 Leuven BE3000 Belgium | ON880419.1;ON622713.1;ON880420.1;ON880421.1;ON880422.1;ON622712.1 |
| Fletcher,N., Gonzalez,G., Meredith,L., Purves,K., Carr,M., Dean,J., Keogan,B., Crowley,B., Lyons,F., O'Reilly,S., Gautier,V., Mallon,P., Gordon,S., Connell,J. and De Gascun,C.F. | National Virus Reference Laboratory University College Dublin Belfield campus Dublin 4 Co. Dublin D04V1W8 Ireland | ON872184.1 |
| de la Hoz-Sanchez,B., Lopez-Ortiz,M., Gutierrez-Arroyo,A., Roces-Alvarez,P., Lazaro-Peona,F., Dahdouh,E., Bloise,I., Garcia-Rodriguez,J. and Mingorance,J. | Microbiology, Hospital Universitario La Paz, Paseo de la Castellana, 261, Madrid, Madrid 28046, Spain | ON838939.1;ON838940.1 |
| Filipe,A., Tong,L., Vattipally,S.B., Maclean,A., Gunson,R., Holden,M.T.G., Barr,D., Ho,A., Palmarini,M., Rambaut,A., Robertson,D.L. and Thomson,E.C. | Viral Genomics and Bioinformatics MRC University of Glasgow Centre for Virus Research 464 Bearsden Road Glasgow G61 1QH United Kingdom | ON808413.1;ON808414.1;ON808415.1;ON808416.1;ON808417.1 |
| Kufner,V., Ziltener,G., Zaheri,M., Schmutz,S., Audige,A., Bernasconi,O., Steiner,K., Huder,J., Shah,C., Capaul,R., Bloembergen,G., Boeni,J., Huber,M. and Trkola,A. | Institute of Medical Virology, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland | ON792320.1;ON792321.1;ON792322.1 |
| Rueca,M., Giombini,E., Gruber,C.E.M., Gramigna,G., Mazzotta,V., Carletti,F., Lapa,D., Pittalis,S., Puro,V., Fabeni,L., Butera,O., Colavita,F., Meschi,S., Matusali,G., Specchiarello,E., Vairo,F., Vaia,F., Nicastrì,E., Antinori,A., Girardi,E. and Maggi,F. | Laboratory of Virology, INMI Lazzaro Spallanzani IRCCS, via Portuense 292, Rome 00149, Italy | ON780016.1 |
| Gramigna,G., Giombini,E., Gruber,C.E.M., Rueca,M., Carletti,F., Cicalini,S., Lapa,D., Puro,V., Marani,A., Fabeni,L., Butera,O., Colavita,F., Meschi,S., Matusali,G., Rivano Capparuccia,M., Specchiarello,E., Vairo,F., Vaia,F., Nicastrì,E., Antinori,A., Girardi,E. and Maggi,F. | Laboratory of Virology, INMI Lazzaro Spallanzani IRCCS, via Portuense 292, Rome 00149, Italy | ON780017.1 |

| Authors | Institute | Accession numbers |
|--|---|-----------------------|
| Croxen,M., Deo,A., Dieu,P., Dong,X., Gill,K., Granger,D., Ferrato,C., Ikkurti,V., Kanji,J., Koleva,P., Li,V., Lloyd,C., Lynch,T., Ma,R., Pabbaraju,K., Rotich,S., Sergeant,H., Shideler,S., Skitsko,T., Shokopoles,S., Tipples,G., Thayer,J. and Wong,A. | Provincial Public Health Laboratory - Research Unit Alberta Precision Laboratories 3030 Hospital Drive NW Calgary Alberta T2N4W4 Canada | ON754989.2;ON736420.2 |
| Giombini,E., Gruber,C.E.M., Rueca,M., Gramigna,G., Vltá,S., Carletti,F., D'Abramo,A., Lapa,D., Puro,V., Fabeni,L., Butera,O., Colavita,F., Meschi,S., Matusali,G., Specchiarello,E., Vairo,F., Vaia,F., Garbuglia,A.R., Nicastri,E., Antinori,A., Glardi,E. and Maggi,F. | Laboratory of Virology, INMI Lazzaro Spallanzani IRCCS, via Portuense 292, Rome 00149, Italy | ON745215.1 |
| Young,E.L., Hergert,J. and Oakeson,K.F. | Department of Health, Utah Public Health Laboratory, 4431 2700 W, SLC, UT 84129, USA | ON627808.1 |
| genEPII sequencing platform | Virology GENomique EPIdemiologique des maladies Infectieuses 103 grande rue Croix Rousse Lyon 69004 France | ON622722; |

Supplementary Table 3: Regions masked in alignment.

| Name | Type | Minimum | Maximum | Length |
|-------------------|--------------|---------|---------|--------|
| repetitive region | misc_feature | 16609 | 16622 | 14 |
| repetitive region | misc_feature | 29749 | 29762 | 14 |
| homopolymeric run | misc_feature | 148334 | 148351 | 18 |
| repetitive region | misc_feature | 163190 | 163214 | 25 |
| repetitive region | misc_feature | 174518 | 174545 | 28 |
| repetitive region | misc_feature | 173274 | 173317 | 44 |
| repetitive region | misc_feature | 169721 | 169774 | 54 |
| repetitive region | misc_feature | 136513 | 136569 | 57 |

| | | | | |
|----------------------|--------------|--------|--------|-----|
| repetitive region | misc_feature | 146848 | 146912 | 65 |
| repetitive region | misc_feature | 150542 | 150628 | 87 |
| repetitive region | misc_feature | 192404 | 192529 | 126 |
| repetitive region | misc_feature | 4680 | 4806 | 127 |
| repetitive region | misc_feature | 179057 | 179245 | 189 |
| homopolymeric run | misc_feature | 592 | 596 | 5 |
| homopolymeric run | misc_feature | 133071 | 133115 | 45 |

Supplementary Methods

The model for the linear regression of number of APOBEC3-type mutations against time was as follows:

$$m_i \sim \text{Normal}(\mu, \sigma)$$

$$\mu = \alpha + \beta (t_i - t_{\text{mean}})$$

$$\alpha \sim \text{Normal}(11, 100)$$

$$\beta \sim \text{Lognormal}(0, 1) \text{ and}$$

$$\sigma \sim \text{Normal}(0, 20),$$

where m_i is the number of mutations for genome i ,

α is the y-intercept and has a prior centred on the minimum number of mutations observed over all genomes, β is a strictly-positive evolutionary rate per year, t_i is the time of collection of the sample and σ is the model error standard deviation.

The model was fitted to the data in a Bayesian framework using the quadratic approximation implemented in the *rethinking* package in R (McElreath, 2020). Posterior estimates and 97% highest posterior densities of the parameters for the lineage A data are α : 25.24 (23.80, 26.66), β : 6.18 (5.19, 7.16), σ : 4.76 (3.74, 5.78) and for the lineage B.1 data are α : 57.67 (57.49, 57.85), β : 5.93 (2.95, 8.91), σ : 1.77 (1.64, 1.90). An R script performing this analysis and generating the graphic seen in Figure 4b is available at <http://github.org/hmpxv/apobec3/>

McElreath (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (2nd edition). CRC Press.