

Evolutionary potential of the monkeypox genome arising from interactions with human APOBEC3 enzymes

Brenda Delamonica,^{1,†} Liliana Davalos,^{2,3,§} Mani Larijani,^{4,5} Simon J Anthony,⁵ Jia Liu,⁶ and Thomas MacCarthy^{1,7,†}*

¹Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794, USA, ²Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794, USA, ³Consortium for Inter-Disciplinary Environmental Research, Stony Brook University, Stony Brook, NY 11794, USA, ⁴Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC V5A 1S6, Canada, ⁵Department of Pathology, Microbiology, and Immunology, University of California-Davis School of Veterinary Medicine, Davis, CA 95616, USA, ⁶Department of Microbiology and Immunology, University of Arkansas for Medical Sciences (UAMS), Little Rock, AR 72205, USA and ⁷Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY 11794, USA

[†]Nominated for communication with the editorial office.

[†]<https://orcid.org/0009-0008-8430-9238>

[§]<https://orcid.org/0000-0002-4327-7697>

*Corresponding author: E-mail: Thomas.maccarthy@stonybrook.edu

Abstract

APOBEC3, an enzyme subfamily that plays a role in virus restriction by generating mutations at particular DNA motifs or mutational ‘hotspots’, can drive viral mutagenesis with host-specific preferential hotspot mutations contributing to pathogen variation. While previous analysis of viral genomes from the 2022 Mpox (formerly Monkeypox) disease outbreak has shown a high frequency of C>T mutations at TC motifs, suggesting recent mutations are human APOBEC3-mediated, how emerging monkeypox virus (MPXV) strains will evolve as a consequence of APOBEC3-mediated mutations remains unknown. By measuring hotspot under-representation, depletion at synonymous sites, and a combination of the two, we analyzed APOBEC3-driven evolution in human poxvirus genomes, finding varying hotspot under-representation patterns. While the native poxvirus molluscum contagiosum exhibits a signature consistent with extensive coevolution with human APOBEC3, including depletion of TC hotspots, variola virus shows an intermediate effect consistent with ongoing evolution at the time of eradication. MPXV, likely the result of recent zoonosis, showed many genes with more TC hotspots than expected by chance (over-representation) and fewer GC hotspots than expected (under-representation). These results suggest the MPXV genome: (1) may have evolved in a host with a particular APOBEC GC hotspot preference, (2) has inverted terminal repeat (ITR) regions—which may be exposed to APOBEC3 for longer during viral replication—and longer genes likely to evolve faster, and therefore (3) has a heightened potential for future human APOBEC3-mediated evolution as the virus spreads in the human population. Our predictions of MPXV mutational potential can both help guide future vaccine development and identification of putative drug targets and add urgency to the task of containing human Mpox disease transmission and uncovering the ecology of the virus in its reservoir host.

Keywords: APOBEC; evolution; monkeypox; virus; computational biology.

Introduction

Monkeypox (renamed to ‘mpox’ by the WHO in late 2022), the disease caused by the monkeypox virus (MPXV), was first found to infect humans in 1970 in the Democratic Republic of Congo and has been mostly contained within West Africa, with local outbreaks continuing to occur since 2017–2018 (Mauldin et al. 2022). Nevertheless, in 2022 the global outbreak of Mpox rose to widespread concern, causing the WHO to designate it as a global health emergency in July 2022 (World Health Organization 2023). While earlier outbreaks were exported by travelers or zoonotic spillover from endemic countries in Africa, recent studies indicated the 2022 outbreak in non-endemic countries was due to regional human–human transmission within the community (Gigante et al. 2022). This suggests the virus has and will continue

to spread and evolve in the human population (Roper et al. 2023). It has become important to characterize the mutational potential of the emergent MPXV, especially as it relates to the activity of human APOBEC3 enzymes.

The APOBEC3 enzymes, of which there are seven in humans, act as part of the innate immune system, playing a role in restriction of viruses and transposable elements. APOBEC3s cause mutations both in single-stranded DNA (ssDNA) and in RNA in the case of some family members (Jalili et al. 2020), and have been found to drive restriction of certain viruses even in the absence of deaminase potential. Although poxviruses, specifically vaccinia virus (VACV), were not thought to be affected by APOBEC3s during short-term viral infection in cell culture (Kremer et al. 2006), recent studies have identified a high frequency of C to T mutations in newly sequenced genomes of MPXV (Isidro et al. 2022) in

a short period of time. This is alarming because DNA viruses generally mutate slower than RNA viruses (Sanjuán and Domingo-Calap 2016). For example, Isidro et al. (2022) compared the MPXV genome MPXV-UK_P2 (see accession number in Supplementary Table 1) from a 2017 to 2018 outbreak to MPXV_USA_2022_MA001 from the multinational MPXV outbreak of May 2022. There were 46 single nucleotide mutations of which 41 (89%) occurred in the context of a TC hotspot (Isidro et al. 2022), consistent with APOBEC activity, specifically APOBEC3 (any except APOBEC3G) which have a preference for TC hotspots. These single nucleotide mutations most likely occurred as the virus circulated and evolved in humans following a 2017–2018 outbreak, exposing the virus to human APOBEC3 (Isidro et al. 2022). Most APOBEC3s are found in the cytoplasm, except for APOBEC3B which is exclusively located in the nucleus (Bogerd et al. 2006; Stenglein, Matsuo, and Harris 2008; Pak et al. 2011; Salamango et al. 2018) and APOBEC3A which locates to both nucleus and cytoplasm (Lackey et al. 2012, 2013). Thus, because it replicates in the cytoplasm, the MPXV genome could be exposed to any APOBEC3 with the possible exception of APOBEC3B. The tissue tropism of MPXV is extensive, and the virus is capable of replicating in most mammalian cells (McFadden 2005). One possible trajectory after human-to-human transmission is that MPXV first replicates at the site of entry, sometimes through respiratory droplets in the lungs and oropharyngeal mucosa, or through skin to skin contact, then spreads to local lymph nodes (Kaler et al. 2022). According to the Genotype-Tissue Expression (GTEx) Portal, the top three highest median bulk tissue expression sites for APOBEC3A are whole blood, spleen, and lungs. As the spleen is a major secondary lymphoid organ, it is possible that other lymphoid organs have similar expression levels. Of the remaining APOBEC3s, APOBEC3D, F, and H spleen expression is also in the top three tissues, whereas expression of APOBEC3C is much more diffuse across many cell types with the exception of brain tissue. Furthermore, APOBEC expression levels are likely to be increased further as a result of cytokine stimulation during infection (Koning et al. 2009; Mehta et al. 2012). Given the variety of APOBEC expression sites, and the overlap with the tissue tropism of MPXV infection, there are many opportunities for interaction. APOBEC mutations may be deleterious to the targeted virus, but may also cause diversification and thus accelerate viral evolution, as has been suggested for both HIV (Monajemi et al. 2014) and SARS-CoV2 (Ratcliff and Simmonds 2021; Simmonds, Ansari, and Pichlmair 2021).

Previous studies by ourselves (Chen, MacCarthy, and Matsen 2017) and others (Poulain et al. 2020) have shown that viruses evolving with possible APOBEC targeting often have a statistically significant reduction of hotspot motifs in their genomes. In our previous analyses of two human gamma-herpesviruses (Epstein-Barr virus, EBV, and Kaposi-sarcoma virus) and an alpha-herpesvirus (herpes simplex virus, HSV1), we used the cytidine deaminase under-representation reporter (CDUR) package (Shapiro, Meier, and MacCarthy 2018) to evaluate the APOBEC3 TC hotspot. Briefly, the package evaluates hotspot under-representation in coding sequences by comparing the number of observed hotspot motifs (e.g. TC) against a null distribution derived from sequences having the same amino acid sequence but with shuffled codons. Although several shuffling algorithms are available within CDUR, both here and in previous work we have used the ‘n3’ algorithm that generates permutations of the nucleotides at the third codon position, while preserving both the amino acid sequence and the GC content, which can greatly affect estimates if not controlled for (Chen, MacCarthy, and Matsen

2017). The null distribution is thus assumed to represent the distribution of hotspot motifs which could have occurred and, thus, selection against hotspot motifs is expected to result in those motifs being under-represented with respect to the null model. More formally, the percentage of shuffled sequences with fewer hotspot motifs than the original sequence is calculated. If this estimated P-value falls on the left tail of the null distribution (e.g. $P < 0.05$), it indicates hotspot motif under-representation, while falling on the right tail indicates over-representation (e.g. $P > 0.95$). CDUR also quantifies the impact of hotspot mutations on amino acid changes (amino acid replacement potential) by simulating C>T mutations at hotspots and counting the number of these that are nonsynonymous. The number of nonsynonymous mutations is then divided by the number of hotspot motifs. Using this measure, the original sequence is again compared with the null distribution derived from the shuffled sequences. Here, if the original sequence falls on the left tail of the null distribution (e.g. $P < 0.05$), then amino acid changes are less likely than expected (nonsynonymous change under-representation), whereas the right tail of the distribution (e.g. $P > 0.95$) indicates more amino acid changes than expected (nonsynonymous change over-representation). In previous work, we found TC hotspot under-representation in many herpesvirus genes (Chen, MacCarthy, and Matsen 2017). Furthermore, simulated C>T mutations at TC hotspots in the same genes caused amino acid changes which suggests a prolonged coevolution between the APOBEC3 enzymes and these viruses (Martinez et al. 2019). We concluded that these genes had evolved under selection by APOBEC3A or B, leading to a depletion of TC hotspots, particularly in synonymous sites.

Although APOBEC3s have already been strongly implicated in the MPXV outbreak of 2022 (Gigante et al. 2022; Isidro et al. 2022; Dobrovolná et al. 2023; Forni et al. 2023), the question of how emerging MPXV strains will continue to undergo further APOBEC3-mediated mutations remains open. Thus, we have evaluated the potential for future evolution of the MPXV genome mediated by APOBEC3. In particular, we measured and compared the APOBEC3 hotspot under-representation and amino acid replacement potential across various human poxvirus genomes to evaluate the evolutionary potential of MPXV genes under APOBEC3 selection. Interestingly, our analysis of MPXV showed many genes had more TC hotspots than expected by chance (over-representation) and fewer GC hotspots than expected (under-representation) suggesting some mechanism which may have targeted GC hotspots, such as prior adaptation to a GC APOBEC targeting motif in a former host species. In some viruses, it has been found that certain viral proteins may block APOBECs activity, such as Vif in human immunodeficiency virus, HIV-1, which targets APOBEC3G (Stopak et al. 2003) or BORF2 in EBV which binds to APOBEC3B (Cheng et al. 2019), although a recent study found that the homolog of BORF2 in cytomegalovirus did not inhibit APOBEC3B (Fanunza et al. 2023). Many gene gain and loss events have been observed as part of the adaptive evolution of poxviruses, especially in the genus of Orthopoxvirus, which includes VACV and variola viruses (VARVs) (McLysaght, Baldi, and Gaut 2003; Hughes and Friedman 2005; Hendrickson et al. 2010). VACVs, for example, have been seen to increase fitness relatively quickly *in vitro* through a ‘gene accordion’ model (Elde et al. 2012). While gene duplication and loss in MPXV genomes may be a major component of host evasion (Brinkmann et al. 2022) and inverted repeats may be mutational hotspots in MPXV (Dobrovolná et al. 2023), we focus part of our analysis on the TC hotspot representation and amino acid replacement of inverted terminal repeats (ITRs), since our analysis

suggests these genes may have a higher probability of being exposed to APOBEC3 during viral replication. Our results suggest that the prior evolution of the virus has led to a greatly increased present potential for APOBEC3-mediated mutations, assuming that other selective forces do not counteract this potential. If Mpox continues to spread in the human population then the abundance of TC hotspots may be targeted and may drive evolutionary selection of mutants with fewer hotspots.

Results

Comparison of human poxviruses suggests that a subset of monkeypox virus genes will be under greater APOBEC3-mediated selection

We previously developed the CDUR package (Shapiro, Meier, and MacCarthy 2018). As mentioned above, our previous analyses of herpesviruses suggested that viral genes under APOBEC3 selection will evolve toward hotspot under-representation together with over-representation of nonsynonymous changes for C>T hotspot mutations (a consequence of hotspot depletion biased towards synonymous sites). This pattern of evolution defines a ‘footprint’ of APOBEC3 selection.

As shown in Fig. 1, analysis of poxvirus genomes using CDUR revealed both TC hotspot under-representation and amino acid replacements in four human poxvirus genomes (NCBI RefSeq versions): (1) molluscum contagiosum virus (MCV), a naturally circulating poxvirus that has unique tissue tropism for human epidermis, (2) MPXV, (3) VACV, a well-characterized poxvirus also used for the smallpox vaccine and (4) VARV, also known as smallpox. Starting with MCV, similar to viruses such as HSV1 that have coevolved extensively with humans, and have been under APOBEC3 selection in particular, we found that MCV (Fig. 1A) has a high proportion of genes with both TC hotspot under-representation and high amino acid replacement. In contrast, MPXV genes (Fig. 1B) show no evidence for TC hotspot under-representation and indeed, many MPXV genes have a significant over-representation, or lack of depletion, of TC hotspots (dots close to the right edge in Fig. 1B). As evident from the distribution of genes along the y-axis in Fig. 1B, MPXV genes also show no evidence of selection for or against amino acid replacements in hotspots. These results are consistent with MPXV not yet having evolved extensively in humans, and in particular, not under selection by human APOBEC3. The results for VACV and VARV are shown in Fig. 1C and 1D. Previous phylogenetic analyses have shown that VACV is closely related to MPXV (Gubser et al. 2004). Although the origins of MPXV and VACV are both unclear and unlikely to be the same, the CDUR results for the two viruses are consistent with a lack of extensive coevolution with their human hosts and demonstrate a significant over-representation of TC hotspots for many genes. The etiology of MCV is distinct from that of MPXV or VARV. In contrast to VACV, the natural hosts of VARV are humans, so more coevolution with human APOBEC3 hotspots is expected in Fig. 1D. While VARV does have significantly lower hotspot under-representation ($P=0.018$, t-test) and higher amino acid replacement ($P=0.048$, t-test) than MPXV, it does not have the same ‘footprint’ of APOBEC3 selection as MCV. The lack of a more pronounced under-representation of TC hotspots in VARV may be due to the slow mutation rate of DNA viruses (Sanjuán et al. 2010; Sanjuán and Domingo-Calap 2016) or that any ongoing coevolution ended once VARV was eradicated. New strains of MPXV are clearly affected by human APOBEC3 and have high mutation rates (Gigante et al. 2022) so we expect genes to evolve toward the top left corner of the CDUR plot over time,

much like most MCV genes, despite the differences between MCV and VARV.

To test our hypothesis that TC motifs in viral genomes may be depleted over evolutionary time in human hosts, we performed our CDUR analysis on a well annotated VARV strain from the 17th century (VD21, accession number KY358055.1) and compared it to the NCBI reference sequence VARV which was collected in 1967 (Duggan et al. 2016; Smithson, Imbery, and Upton 2017; Mühlemann et al. 2020). Although the overall difference between the two genomes per gene is modest (Supplementary Table 9 for full list), the average shifts between genes tend slightly toward the top left corner, as expected (Supplementary Fig. 1). Thus, the TC hotspot representation difference (VD21 minus VARV RefSeq) is positive and significant ($P=0.039$, paired t-test), suggesting that TC hotspots tend to deplete over time. The amino acid replacement difference is negative, meaning this value increases, again as expected, although the change is not significant ($P=0.862$, paired t-test).

Many monkeypox virus genes show over-representation of TC hotspots, especially in repeat regions

Under the assumption that MPXV will continue the early trend of selection of synonymous APOBEC3 TC hotspot mutations (TC>TT), eventually leading to a profile similar to MCV (Fig. 1A), we combined both hotspot under-representation and amino acid replacement measures to define a simple measure of mutation potential as the Euclidean distance of each gene to the top left corner of Fig. 1B (i.e. hotspot under-representation = 0, amino-acid replacement = 1). We calculated this measure for MPXV genes using the relatively well annotated RefSeq genome. Table 1 shows the top 15 MPXV genes from Fig. 1B ordered by distance to the top left corner, with genes at the top of the table having the highest distance, i.e. the highest hotspot representation and lowest amino acid replacement score (for a full list of genes see Supplementary Table 2). Many genes near the top of the list participate in immune evasion. For example, at the top of the list is a secreted chemokine binding protein (OPG001) which in VACV blocks the interaction of chemokines with cellular receptors (Alcamí et al. 1998). Based on the list of poxvirus genes annotated as being involved in ‘host immune modulation’ in the ViralZone database (Hulo et al. 2011), we found other genes within the top 15 (not counting repeated ITR genes), including: OPG029, a homolog of VACV C6, which inhibits induction of interferons (IFN- α) (Stuart et al. 2016) and IFN- β (Unterholzner et al. 2011), OPG002/CrmB which binds to host tumor necrosis factor (Gileva et al. 2006), and OPG019 which in VACV binds to host epidermal growth factor receptor to stimulate pro-viral cellular proliferation around infected cells. Interestingly, these four genes are also in the repeat regions of the genome. Among the bottom 15 genes (Supplementary Table 2) we only found one gene, OPG176, an inhibitor of the TLR4 signaling pathway in the ‘host immune modulation’ category, that has acquired a C>T mutation in the recent outbreak. While there was some enrichment (i.e. relative abundance of TC hotspots), at the extreme top of the list, the overall ranking of the annotated ‘host immune modulation’ genes was not significantly high or low (Wilcoxon one-sample test, $P=0.2935$).

Previous analysis suggests that APOBEC3 causes mutations predominantly during viral genome replication (Suspène et al. 2006; Willems and Gillet 2015). For example, in the case of HIV, there is a high correlation between the gradient of APOBEC3G

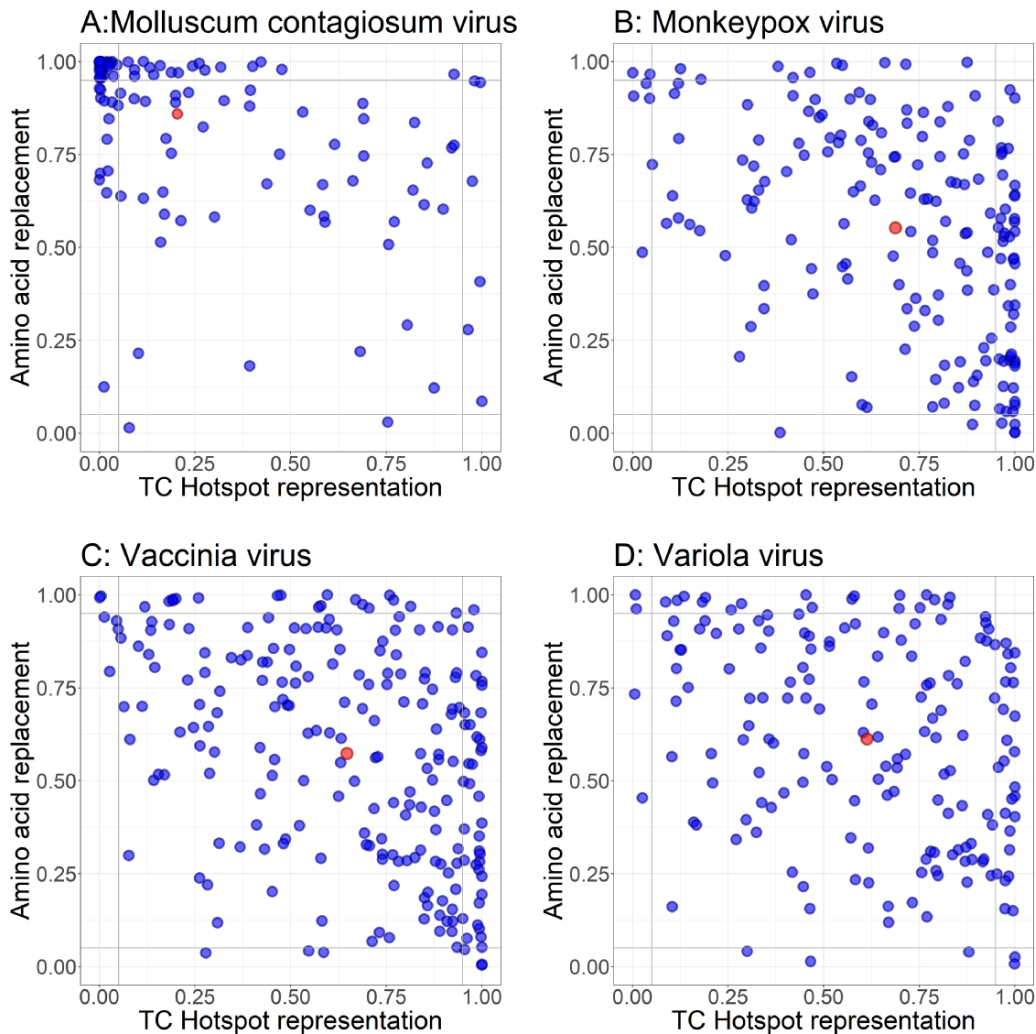


Figure 1. Plots of CDUR statistics (CDUR plots) for TC hotspot under representation (horizontal axis) and amino acid replacement (vertical axis) for genes of human poxviruses. (A) Molluscum contagiosum virus (MCV), (B) Monkeypox virus (MPXV), (C) Vaccinia virus (VACV), and (D) Variola virus (VARV). Red dot indicates the mean.

targeting and the amount of time the DNA minus strand remains single stranded (Willems and Gillet 2015). More generally, APOBEC3 may target regions close to origins of replication more efficiently since the ssDNA will often be exposed for longer there, during replication initiation. A previous study in VACV identified the ITR region, at the ends of the genome, as dominant origins of replication (Yang et al. 2011). We hypothesize that if ITR genes have an over-representation of TC hotspots they may be more susceptible to APOBEC3 mutations than other genes with high mutation potential since ITRs can potentially be exposed to APOBEC3s for longer during viral replication. We found that six of the 15 MPXV genes in Table 1 that are predicted to be most vulnerable to APOBEC3-mediated mutations, also appeared close to one of the chromosome ends (within 20 kb of either end), including within the annotated MPXV ITRs. Indeed, among the 15 genes, three out of four unique genes (six of a total of eight ITR genes) are within the top 10th percentile of all genes, with OPG001 at the top, OPG002 at the 5th percentile and OPG003 at the 8th percentile. Furthermore, these three distinct ITR genes (OPG001-3) have a hotspot representation value of close to 1 from CDUR, suggesting that genes within the MPXV ITRs are more likely to be targeted

as the virus evolves in humans. Similarly, the values for amino acid replacement are all less than 0.25, suggesting they are less vulnerable to APOBEC3-mediated mutations causing amino acid changes. Of note are the ITR genes OPG001 and OPG019 which have amino acid replacement values less than 0.05 making them even less vulnerable to amino acid changes. Considering positions further away from the chromosome ends, we found only a weak negative correlation between closeness to the chromosome ends (distance in nt of each gene midpoint to the closest chromosome end) and the distance from the top left corner of the CDUR plot for TC hotspots ($r = -0.14$, $P = 5.9 \times 10^{-2}$). These results suggest that genes closer to chromosome ends may evolve quicker in the future via APOBEC3-induced mutations and that this effect will be strongest for genes in the ITRs. As with HIV, this effect may be a result of ITR genes at chromosome ends having a longer exposure time of ssDNA during replication than other genes within the chromosome (Willems and Gillet 2015). This finding is consistent with previous publications which found higher gene gain and loss activity at the ends (Brinkmann et al. 2022) and a higher frequency of recent mutations in the MPXV inverted repeat regions (Dobrovolná et al. 2023).

Table 1. Top 15 genes in monkeypox virus genome sorted by CDUR TC motif result distance from the point (0,1).

Gene	Protein	TC hotspot under-representation	TC hotspot amino acid replacement	Distance from top left	Repeats	Mutation count
OPG001 ^a	Chemokine binding protein	1	0	1.413	Repeat region	1
OPG001 (repeat)	Chemokine binding protein	1	0	1.412	Repeat region	1
OPG100	IMV membrane protein J1	1	0.02	1.397	CDS	0
OPG138	A12 protein	0.966	0.03	1.371	CDS	0
OPG079	DNA binding phosphoprotein 2	0.995	0.06	1.369	CDS	0
OPG002 ^a	Crm B secreted TNF alpha receptor like protein	1	0.08	1.362	Repeat region	1
OPG029 ^a	Bcl 2 like protein	0.978	0.06	1.357	CDS	0
OPG002 (repeat)	Crm B secreted TNF alpha receptor like protein	1	0.09	1.355	Repeat region	1
OPG076	MV membrane EFC component	0.96	0.07	1.339	CDS	0
OPG153	Orthopoxvirus A26L A30L protein	0.996	0.12	1.328	CDS	0
OPG019 ^a	EGF like domain protein	0.889	0.02	1.32	CDS	1
opg205 ^a	Ankyrin repeat protein 44	0.97	0.13	1.306	CDS	0
OPG003 ^a	Ankyrin repeat protein 25	1	0.18	1.293	Repeat region	3
OPG003 (repeat)	Ankyrin repeat protein 25	1	0.19	1.288	Repeat region	3
OPG163	MHC class II antigen presentation inhibitor	0.895	0.08	1.287	CDS	0

^a= genes within 20kb of end.

Distance From top left = Euclidean distance of hotspot representation and amino acid replacement from point (0,1) of CDUR plot.

Monkeypox virus genes exhibit under-representation of GC hotspots

Our previous work (Chen, MacCarthy, and Matsen 2017; Martinez et al. 2019) has shown that under-representation statistics for different hotspots will often be negatively correlated if those hotspots are mutually exclusive. For example, strong under-representation of APOBEC3A/B TC hotspots may coincide with an over-representation of APOBEC3G CC hotspots, because any remaining C sites must be either GC, CC, or AC. Because in the case of MPXV we observed a moderate over-representation of TC hotspots, we evaluated the alternative hotspots GC, CC, or AC, as shown in Fig. 2. We found that for GC hotspots (Fig. 2A), there is a strong effect for both CDUR measures (hotspot under-representation and amino acid replacement over-representation). Also, the under-representation measures for GC and TC hotspots are significantly negatively correlated in MPXV genes (Pearson $r = -0.35$, $P = 1.9 \times 10^{-6}$, Supplementary Table 3). Similarly, the amino acid replacement measures for GC and TC hotspots are also negatively correlated ($r = -0.2$, $P = 8.8 \times 10^{-3}$, Supplementary Table 3). Conversely, the CC hotspot results demonstrate a low amino acid replacement (Fig. 2B), suggesting that CC mutations are less likely to cause an amino acid change. As expected, amino acid replacement for GC and CC hotspots are negatively correlated ($r = -0.14$, $P = 6.1 \times 10^2$, Supplementary Table 3). In Table 2, MPXV genes are ranked from shortest to greatest distance from the top left corner of GC (rather than TC) hotspot under-representation and amino acid replacement (i.e. hotspot under-representation = 0, amino-acid replacement = 1, see Supplementary Table 4 for the full list). The GC results are roughly the opposite of those for TC (Table 1), where the same three distinct genes, OPG002, OPG003, and OPG001, are within the bottom 2 per cent, 7 per cent, and 8 per cent of genes respectively. GC hotspot under-representation for the three ITR genes does not exceed 0.002, while the amino acid replacement ranges between 0.996 and 1. An under-representation of GC hotspots may indicate prior exposure to nonhuman APOBECs with different hotspot

preferences than those of human APOBECs (see Discussion). These findings collectively support several of our hypotheses. To summarize, if the previous host of MPXV had an APOBEC that favors GC hotspots, and, because the ends of the viral genomes are exposed with higher chance during viral replication, these areas may have been targeted for GC depletion more than other locations of the genome. Given the negative correlation between GC and TC hotspots, this could have resulted in TC over-representation at the ITRs of the genome, allowing a subset of genes to be more susceptible to mutation in humans with APOBEC3s that target TC hotspots.

This analysis focused on -1 nucleotide motifs since most human APOBEC3s have a dominant preference at the -1 nt position. However, there are exceptions, such as AID (hotspot WRC, $W = A/T$, $R = A/G$) which includes GC as a subset. To explore the relevance of the -2-nucleotide position, we evaluated the four NGC hotspots and found a somewhat similar trend towards GC amongst the three DGC hotspots (-2 position: A, T, or G), whereas CGC did not show any apparent trend (Supplementary Fig. 2). We conclude there is no strong preference for a -2 nucleotide in this case.

Gene length and the potential for APOBEC3-mediated mutations

We also identified gene length as a factor for potential evolvability with human APOBEC3 (TC hotspots). Again, we used the distance metric (Supplementary Table 2, column 'distance from left upper') as a proxy for evolutionary potential and compared it to gene length (log10-transformed), as shown in Fig. 3A1, finding a significant positive correlation ($r = 0.3$, $P = 4.9 \times 10^{-5}$, Supplementary Table 5). This is primarily due to the strong correlation between gene length and TC hotspot representation ($r = 0.42$, $P = 1.1 \times 10^{-8}$, Fig. 3B1, Supplementary Table 5) as there is no correlation between gene length and amino acid replacement ($r = 0$, $P = 9.8 \times 10^{-1}$, Fig. 3C1, Supplementary Table 5). Genes of length greater than

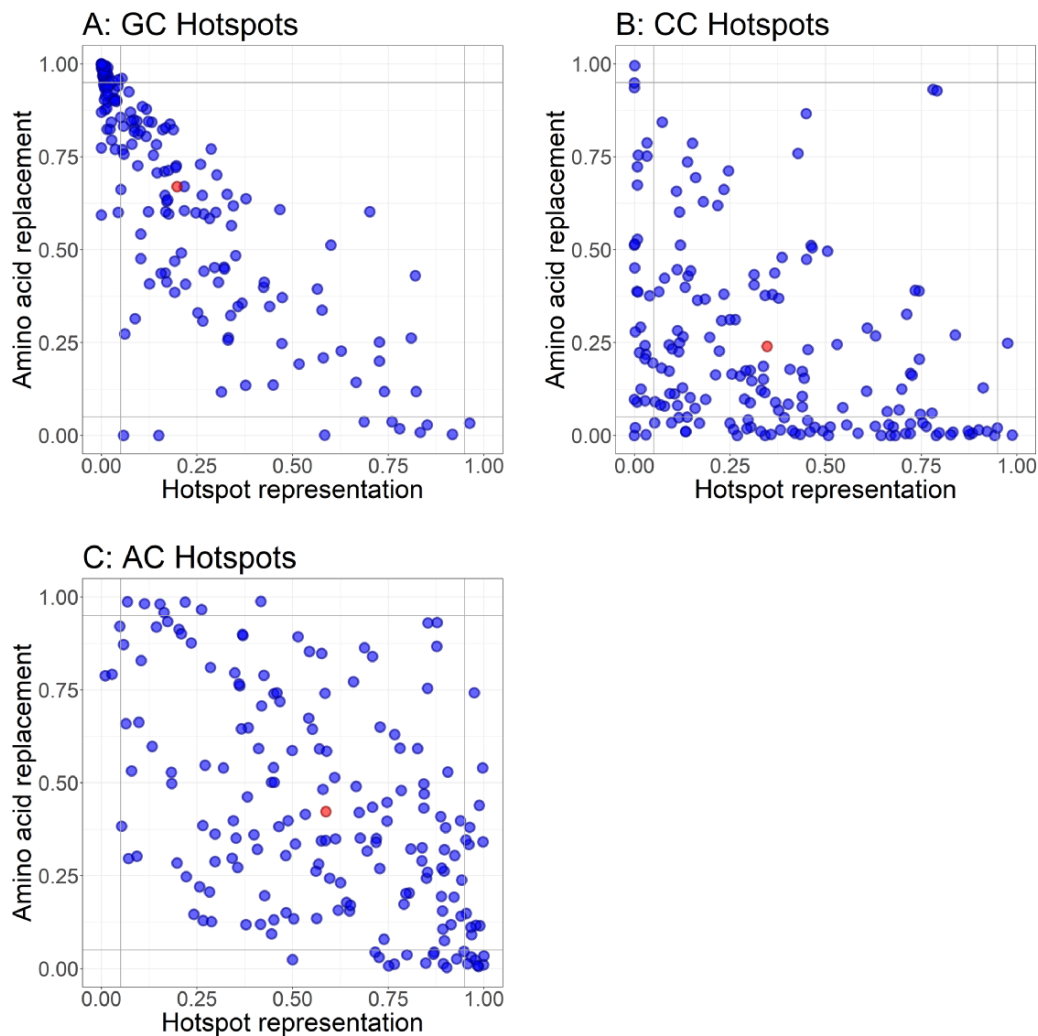


Figure 2. CDUR plots for non-TC hotspots in Monkeypox virus, showing, for each gene, the hotspot under representation (horizontal axis) and amino acid replacement (vertical axis). Each plot shows results for a different hotspot: (A) GC hotspots, (B) CC hotspots, and (C) AC hotspots. Red dot indicates the mean.

1.8kb (\log_{10} : ~ 3.25) had very high TC hotspot over-representation with the exception of only two outliers, OPG210, a T-cell suppressor involved in host immune modulation, and OPG025, which also has a host immune modulation function. Eleven of the 21 longest genes are involved in genome replication, 5 genes are involved in host immune modulation, and 4 in virion assembly or budding (ViralZone and SIB Swiss Institute of Bioinformatics 2023). Thus, longer genes may be associated with a higher likelihood of APOBEC targeted mutations (see Discussion).

As mentioned above, the underlying reason for high TC hotspot over-representation may partially be the inverse relationship with GC hotspot measures (as CDUR measures should not otherwise be affected by gene length). The reason for the minimal preference for TC over-representation when compared to CC and AC (Fig. 2) is, however, unclear. When we compared gene length with the combined measure ('distance from left upper') hotspots, we found a significant negative correlation, as expected ($r = -0.41$, $P = 1.8 \times 10^{-8}$, Fig. 3A2, Supplementary Table 5). We found a negative correlation between GC hotspot under-representation and gene length ($r = -0.27$, $P = 2.6 \times 10^{-4}$, Fig. 3B2, Supplementary Table 5), and we noticed that the five longest genes all had significant GC hotspot under-representation. Meanwhile amino acid replacement had a positive correlation with gene length

($r = 0.44$, $P = 8.2 \times 10^{-10}$, Fig. 3C2, Supplementary Table 5). Since the current MPXV strain may have originated from a previous host with APOBECs that favored GC hotspots, future mutations within humans may occur preferentially in larger genes, particularly those with high TC over-representation. As human APOBEC3s cause new C>T mutations, the number of TC hotspots are expected to greatly reduce, as observed in MCV. To get a better understanding of the trajectory of MPXV, we considered the relationship between TC hotspots and gene length in MCV. Here, there is a weak negative correlation between gene length and the distance from the left upper corner of the CDUR plot ($r = -0.18$, $P = 1.9 \times 10^{-2}$, Supplementary Table 6). Repeating the analysis with the two CDUR measures separately, we found TC hotspot representation and gene length have a very low negative correlation ($r = -0.04$, $P = 6.3 \times 10^{-1}$, Supplementary Table 6), while amino acid replacement and gene length have a relatively strong positive correlation ($r = 0.35$, $P = 3.7 \times 10^{-6}$, Supplementary Table 6).

Mutations in UKP2 and future evolution

A recent study (Isidro et al. 2022) described mutations in the 2022 outbreak using the MPXV-UK_P2 genome (see Supplementary Table 1 for accession number) as a putative ancestor for the viruses sequenced thus far. The NCBI RefSeq MPXV and

Table 2. Bottom 15 genes in monkeypox virus genome sorted by CDUR GC motif result distance from the point (0,1).

Gene	Protein	GC hotspot under-representation	GC hotspot amino acid replacement	Distance from top left	Repeats	Mutation count
OPG002	Crm B secreted TNF alpha receptor like protein	0	1	0	Repeat region	1
OPG080	Ribonucleoside diphosphate reductase 2	0	1	0	CDS	0
OPG105	DNA dependent RNA polymerase subunit rpo147	0	1	0	CDS	4
OPG002	Crm B secreted TNF alpha receptor like protein	0	1	0	Repeat region	1
OPG098	Nucleic acid binding protein VP8 L4R	0	0.999	0.001	CDS	1
OPG117	NTPase 1	0	0.999	0.001	CDS	0
OPG200	Bcl 2 like protein	0.001	0.999	0.001	CDS	0
OPG054	Serine threonine protein kinase	0	0.998	0.002	CDS	0
OPG189	Ankyrin repeat protein 25	0	0.998	0.002	CDS	0
OPG198	Ser thr kinase	0	0.998	0.002	CDS	1
OPG003	Ankyrin repeat protein 25	0	0.998	0.002	Repeat region	3
OPG003	Ankyrin repeat protein 25	0.001	0.998	0.002	Repeat region	3
OPG129	Virion core protein P4b	0	0.997	0.003	CDS	0
OPG001	Chemokine binding protein	0.002	0.997	0.004	Repeat region	1
OPG001	Chemokine binding protein	0.002	0.996	0.004	Repeat region	1

Distance From top left = Euclidean distance of hotspot representation and amino acid replacement from point (0,1) of CDUR plot.

MPXV-UK_P2 were both exported from Nigeria in 2017 to 2018 and are closely related (Isidro et al. 2022). Comparing the two genomes, MPXV-UK_P2 has 11 insertions of varying lengths (ranging from 114 to 2,319 nt) and a total of 17 single nucleotide mutations with respect to RefSeq MPXV. Despite these differences, the two genomes are still very closely related, with an overall BLAST alignment similarity of 99.98 per cent.

As mentioned above, previous studies reported a large proportion of TC>TT mutations ascribed to APOBEC3 (Isidro et al. 2022; Dobrovolná et al. 2023; Forni et al. 2023). Isidro et al. reported a total of 46 CT>TT or GA>AA mutations across 27 different genes. The new mutations occurred in three of the four distinct ITR genes, as shown in Supplementary Table 7, which is again ordered by column 'distance to left upper'. In particular, the genes with locus tags MPXV-UK_P2_001/190, MPXV-UK_P2_002/189, and MPXV-UK_P2_003/188 align most closely with MPXV genes OPG001, OPG002, and OPG003, and had 1, 1, and 3 mutations in each gene, respectively. These three ITR genes are all within the 10th percentile of greatest TC distance from the left corner. This gives further evidence that the ITR genes are particularly susceptible to APOBEC3 mutations.

While we did not find any statistical significance between mutation count (for the new mutations) and distance from the left upper corner for TC hotspots, the mutation count was significantly correlated with log10 gene length ($r = 0.70$, $P = 4.2 \times 10^{-4}$, Supplementary Table 8), as one would expect, since longer genes are more likely to be mutated. The low correlation between the CDUR measures and mutation count may be primarily due to the few mutations detected, usually one per mutated gene, and at most 4. In plotting the new mutations for MPXV-UK_P2, an association between either amino acid replacement or TC hotspot under-representation and new mutations (Fig. 4A, compare red, mutated genes, with blue, non-mutated genes) is not immediately obvious. In contrast, the moving average of mutations over a position window of 1k nt across the genome does show two of the highest peaks at the ends, where the ITRs are located (Fig. 4B). While

mutation frequency (see Methods) was only marginally associated with TC hotspot under-representation ($r = 0.15$, $P = 0.0451$, Supplementary Table 8), amino acid replacement was not ($r = -0.11$, $P = 0.129$, Supplementary Table 8). When we compared the combined measure ('distance from left upper') between genes that did have a mutation against those genes that did not, we did find a significant difference (Welch two sample t-test, $P = 0.046$). We expect this difference to become stronger as more mutations are reported including for the individual CDUR measures, amino acid replacement, and hotspot under-representation. In summary, the trends observed for new (2022 outbreak) mutations suggest that TC hotspot under-representation and amino acid replacement will be important factors for predicting ongoing mutations of MPXV as the virus continues to circulate among humans, and that this effect is especially strong for ITR genes.

Discussion

As expected for a case of recent zoonosis, MPXV genes do not exhibit signs of coevolution with human APOBEC3s, such as APOBEC3A/B/H (TC hotspots) or APOBEC3G (CC hotspots). Our results show that MPXV genes have both under-representation and high sensitivity (to amino acid changes) for the motif GC. In part because GC and TC are mutually exclusive di-nucleotides, there is a negative correlation between the GC and TC hotspot measures (for both under-representation and amino acid replacement), which may explain why many MPXV genes have more TC hotspots than expected. These features may have arisen due to virus evolution in the previous host if that host had an APOBEC (or APOBEC-like) protein with a preference for GC hotspot mutations. Interestingly, we found that GC hotspot under-representation and amino acid replacement was particularly high for genes in the terminal repeat regions, which contain the most origins of replication and thus presumably expose more ssDNA to potential APOBEC mutations. Thus, we speculate that ITR genes may also have

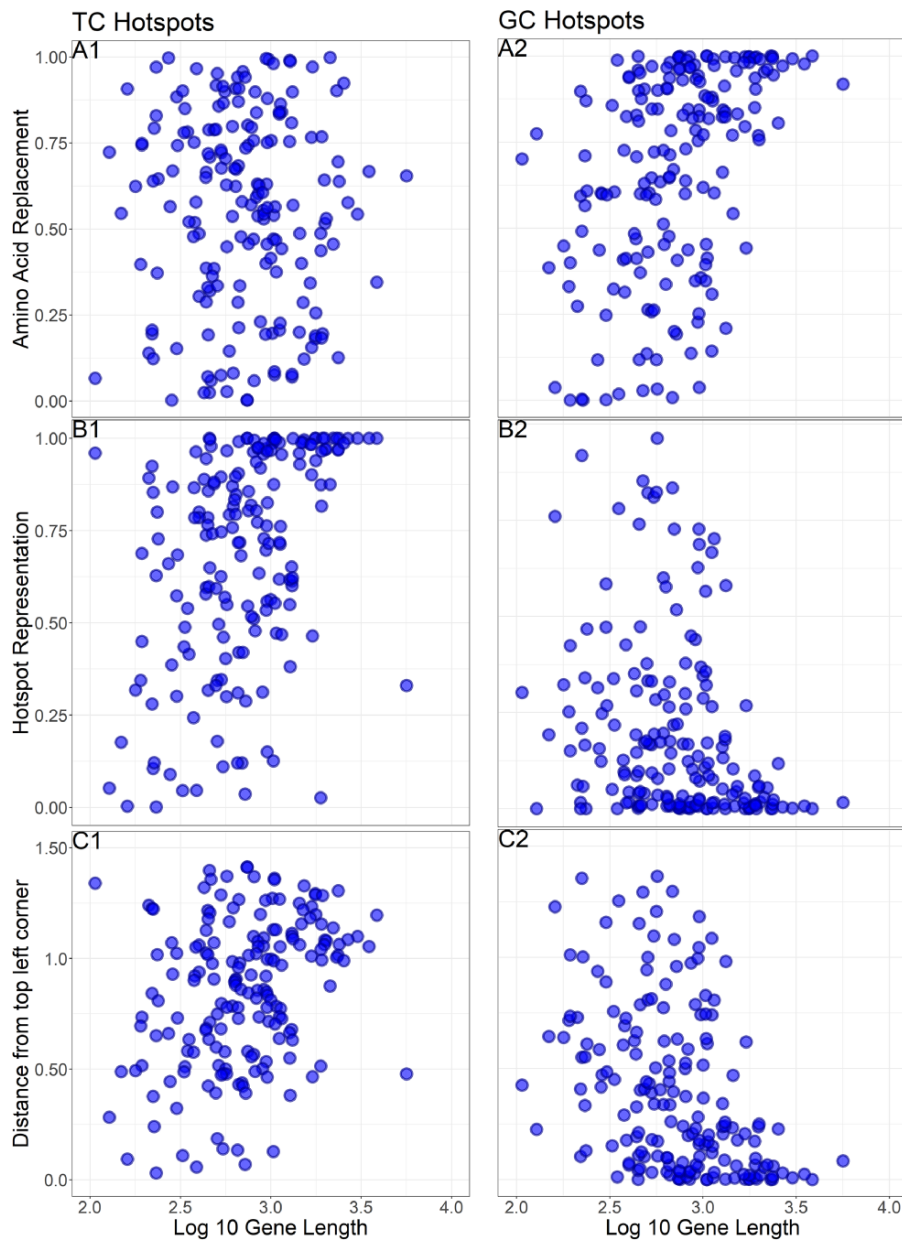


Figure 3. On the left are plots of Monkeypox gene length (log10) vs (A1) the CDUR plot ‘distance from the top left corner’ measure of evolutionary potential for TC hotspots, (B1) TC hotspot under representation, and (C1) amino acid replacement. On the right are equivalent plots for GC hotspots, showing Monkeypox gene length (log10) vs (A2) the CDUR plot ‘distance from the top left corner’, (B2) GC hotspot under representation, and (C2) amino acid replacement.

been among the most mutated, or selected, in the putative previous host as a consequence of greater APOBEC exposure. These genes were also predicted to have the highest potential for future mutation in human hosts in response to human APOBEC3 activity.

Regardless of hotspot targeting, heightened APOBEC activity would be expected to reduce the overall GC (G+C) content of a viral genome, particularly in the short-term following spillover. Yet, despite the apparent footprint of APOBEC3 on the evolution of MCV, the overall GC content of the MCV genome (Fig. 5A) is relatively high, at ~64 per cent (mean per coding sequence—see Methods). In the longer term, other selection agents, such as transcriptional efficiency, which appears to be positively correlated with GC content (Mordstein et al. 2021), may play a larger role in changing GC content. The case of MCV suggests that codons

and/or amino acids can evolve so as to reduce APOBEC3 hotspots independently of GC content. In contrast to MCV, VACV (Fig. 5B), VARV (Fig. 5C), and MPXV strains (Fig. 5D–E) have a relatively low GC content of ~34 per cent. Although the discrepancy with MCV is not easily explained, it is known that certain viral families have large differences of GC content between viral species, Poxviridae being one of them (Mordstein et al. 2021). VACV and MPXV may have evolved a low GC content due to selective pressures in their previous host(s). For MPXV in particular, characterization of these previous selective processes will have to wait until the previous host(s) is identified. Unfortunately, APOBEC preferences have been characterized only for a few species, so this feature alone is unlikely to help identify the previous host (Münk et al. 2008). MPXV is assumed to have been transmitted relatively recently to humans

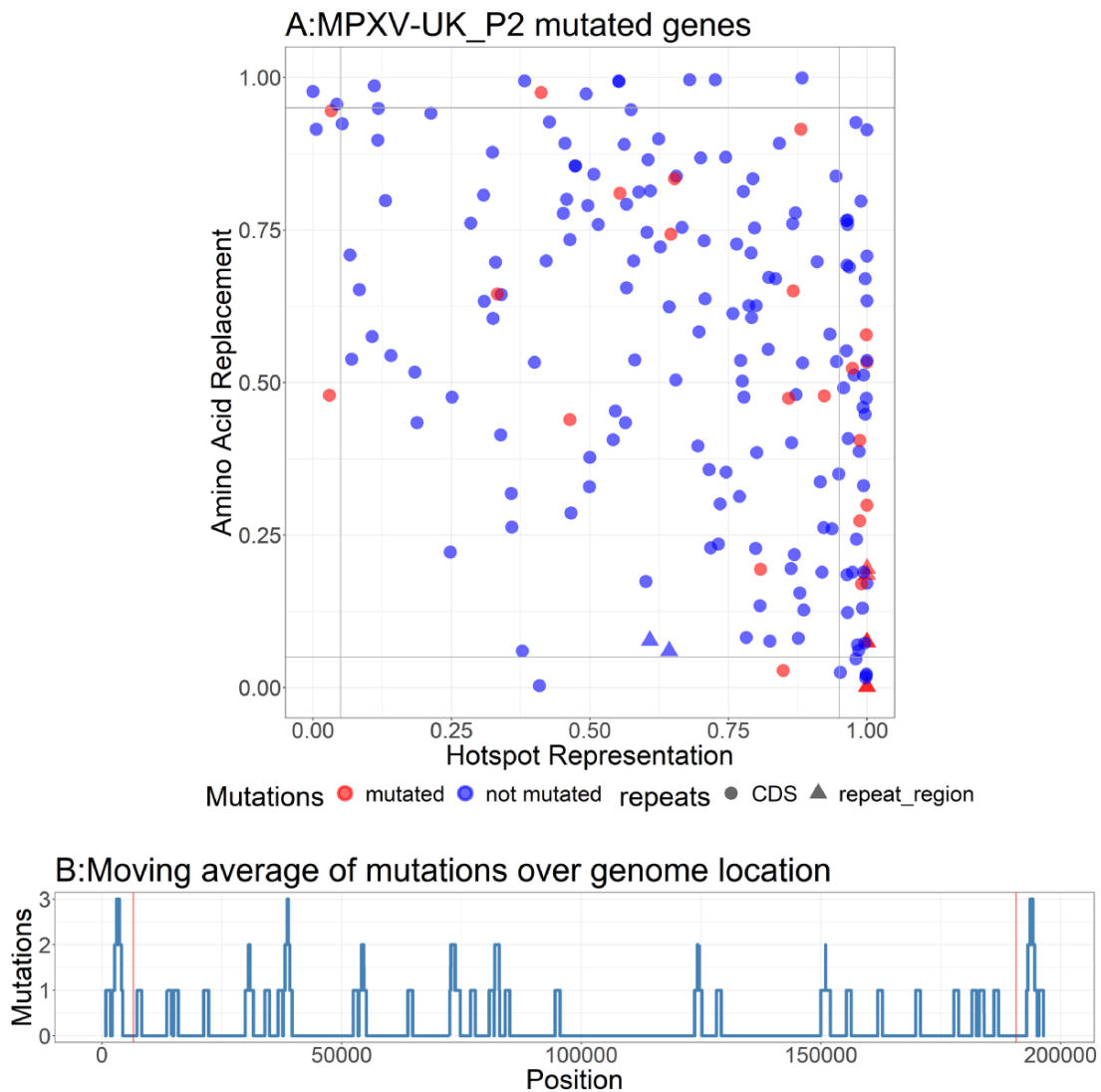


Figure 4. (A) CDUR plot of TC hotspot under representation (horizontal axis) and amino acid replacement (vertical axis) for MPXV-UK_P2. Triangles are genes within the repeat regions (ITR) and red points are mutated genes. (B) Rolling average of mutations from MPXV-UK_P2 over nucleotide position with a window of 1k nb. Red lines indicate end of repeat regions.

from a previous, possibly rodent, host (Mauldin et al. 2022; Ghazy et al. 2023). The mouse APOBEC3 homolog (mA3) has been found to restrict retroviruses and inhibit their replication, while certain mouse viruses may have evolved mechanisms which counteract mA3 (Salas-Briceno, Zhao, and Ross 2020). In the case of mouse mammary tumor virus (MMTV), viral replication is inhibited by mA3, but G>A mutations in the integrated proviral genome are rare (MacMillan, Kohli, and Ross 2013), suggesting that some mA3-mediated inhibition occurs via a deamination-independent mechanism, at least in some retroviruses. The biochemistry of deamination-independent mechanisms is still poorly understood, but it is likely that adaptation of virus genomes to deamination-independent mechanisms will lead to quite different outcomes in terms of di-nucleotide distribution, most obviously because the hotspots themselves are not directly depleted. Alternatively, the virus may have evolved in a previous organism to accumulate TC motifs in subregions of the genomes as an immune escape response. For example, Martinez et al. found that EBV had an over-representation of CCC motifs in particular subregions which, as the authors argued, may be acting as a ‘decoy’ mechanism

to attract AID toward those sites making it unlikely to mutate, therefore limiting mutations in more critical subregions of the genome (Martinez et al. 2019). Similarly, Monajemi et al. observed HIV adaptations *ex vivo* and found an abundance of APOBEC3G/F hotspots in sequences which encode certain common HIV T cell epitopes. These APOBEC-mediated mutations favored CTL escape, diminishing HIV specific T cell response (Monajemi et al. 2014). As these examples illustrate, biased di-nucleotide distributions in certain MPXV genes may be a signature of genetic robustness in response to previous natural selection by host APOBECs.

In primates including humans, the APOBEC3 cytidine deaminases appear to be under strong diversifying selection, which has led to the expansion and specialization (Münk, Willemsen, and Bravo 2012) of the seven genes in the subfamily. Previous work has shown evidence of viral genome evolution in response to human APOBEC3s, including DNA viruses (herpesvirus, papillomavirus, polyomavirus, and others), both endogenous and exogenous retroviruses and RNA viruses (coronavirus) (Willems and Gillet 2015). A key feature affecting the viral evolution process is the amount of time the host and viral species have had to

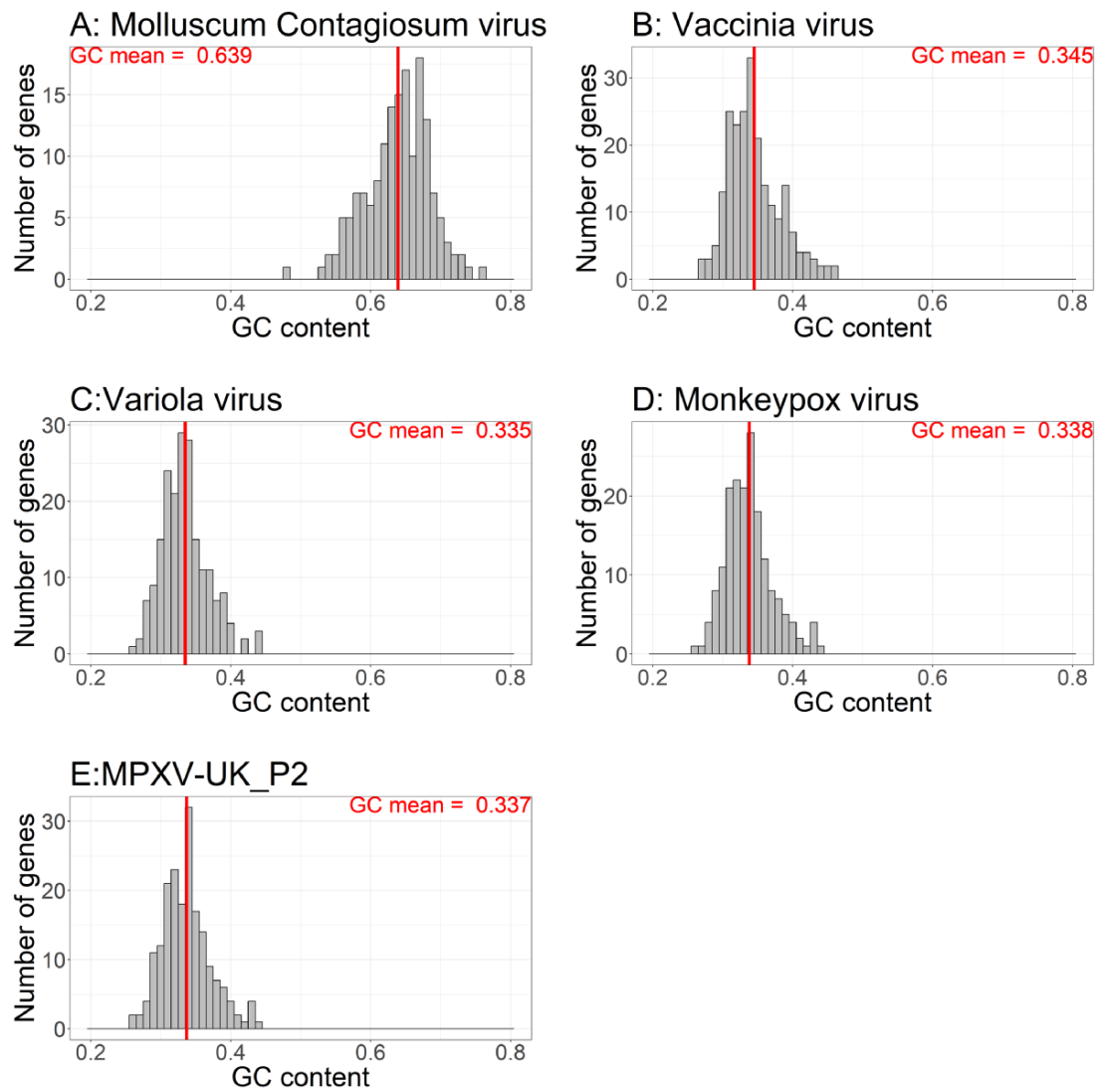


Figure 5. G+C content histograms and averages of (A) Molluscum contagiosum virus, (B) Vaccinia virus, (C) Variola virus, (D) Monkeypox virus, and (E) MPXV-UK_P2. Red lines indicate the average G+C content.

coevolve. Well-established native species of virus in humans, such as papillomaviruses, polyomaviruses and herpesviruses contain a genomic ‘footprint’, most obviously TC hotspot depletion, indicative of extensive coevolution (Poulain et al. 2020; Shapiro, Krug, and MacCarthy 2021). Our analyses reveal signatures of coevolution with human APOBEC3s (having TC hotspot preferences) for MCV, a native human poxvirus (Fig. 1A), but not for the recently emerged MPXV. Given that poxviruses are DNA viruses that replicate in the cytoplasm, they might also be expected to be exposed to APOBEC3G, which has a CC (or CCC) hotspot preference, but we found no evidence for this in MPXV (Fig. 2B) or VARV (Supplementary Fig. 3), which are both very closely related with very similar hotspot distributions. However, as the virus circulates in humans, APOBEC3G may also deplete CC hotspots much like MCV (Fig. 6). In contrast to well-established native species, viral species that have recently emerged in new host species might be expected to accumulate APOBEC3 mediated mutations as they evolve, which appears to have happened in the case of SARS-CoV-2 (Ratcliff and Simmonds 2021; Simmonds, Ansari, and Pichlmair 2021).

We found an association between gene length and the potential for APOBEC3-mediated mutations, despite the method correcting for length by comparing with reshuffled versions of the same sequence. The simplest explanation may be that accumulation of more mutations increases the probability of one of these mutations being functionally deleterious. At the level of transcription, genes with longer transcripts may be more vulnerable to APOBEC mutations due to the presence of multiple polymerase molecules interfering with one another during transcription, which may in turn increase the exposure time of ssDNA. Recent work also suggests that ssDNA secondary structure is an important factor for deamination and can partly compensate for generating high affinity substrates even when no hotspot is present (Langenbucher et al. 2021). Thus, although hotspot representation and amino acid replacement are dominant factors, gene length, and secondary ssDNA structure may also combine in nontrivial ways to further determine APOBEC mutability. We propose to investigate in the future the impact of ssDNA secondary structure on our existing hotspot enrichment measures.

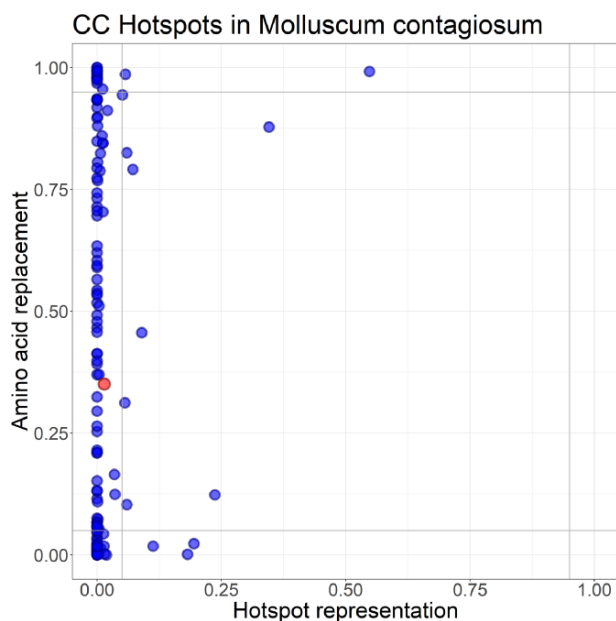


Figure 6. Plot of CDUR statistics for CC hotspot under representation (horizontal axis) and amino acid replacement (vertical axis) for genes of *Molluscum contagiosum*. Red dot indicates the mean.

Further experimental work will be required to validate APOBEC3-mediated restriction of MPXV using cell-based assays. For example, with deep sequencing we will be able to verify whether the genes we have identified are most susceptible to APOBEC3-mediated mutations, which allow us to confirm our findings that ITRs and/or gene length are important attributes in APOBEC3 mutation potential. Other future studies could test if any of the mutations in genes we have identified change the efficacy of existing drugs or vaccines for VARV which have been used therapeutically during the 2022 Mpox outbreak (Russo et al. 2020; Carvalho 2022; Matias et al. 2022). Treating Mpox with Brincidofovir remains an option. The protection of the current vaccine toward MPXV that was developed against VARV is imperfect. Tecovirimat, a pan-Orthopoxvirus inhibitor drug used for VARV, is still in clinical trials for use against MPXV and its effectiveness is not yet complete for evaluation at the time of writing (Carvalho 2022; Matias et al. 2022; National Institute of Allergy and Infectious Diseases (NIAID) 2023). Tecovirimat targets the homolog of the VACV F13 gene, which produces a Palmitoylated EEV membrane protein in MPXV (Duraffour et al. 2008; Frenois-Veyrat et al. 2022). Interestingly, this is one of the genes that underwent a mutation from MPXV-UK_P2 to MPXV_USA2022_MA001 (Supplementary Table 7). Gigante et al. evaluated this mutation using an infectivity assay and found that the median EC50 for attenuation of the new virus was more than twofold higher, even though the difference was not significant. Since the authors used an *in vitro* assay, the *in vivo* efficacy may turn out to be lower. While APOBEC3 may also cause other functional mutations, the correspondence between therapeutic targets and mutations can be another line of information on future treatments. Biochemical characterization of APOBECs may reveal specific targeting preferences that could be used to narrow down or identify the possible reservoir species.

In our analysis, MPXV genes predicted to mutate with highest frequency were predominantly those that have acquired most observed mutations during the current (2022) outbreak. These results suggest that MPXV, as it adapts to the human host

APOBEC3s, may acquire more mutations. At the time of writing the 2022 outbreak appears to be declining, which should reduce diversification, but obviously future outbreaks that again increase diversity are possible. Also, further mutations may produce less benign strains than those currently circulating given that selection acts primarily at the level of transmissibility rather than virulence (Wertheim 2022). While the case for containment has been made on broad evolutionary grounds before (Johnson et al. 2022), our analyses outline the mutational landscape of coevolution and have been partially validated by observed mutations.

In conclusion, we show patterns of mutation across three poxviruses with divergent host ecologies related to their exposure to human APOBEC3 to different degrees. While enrichment for TC hotspots in MPXV both reflects its recent emergence in humans and perhaps past adaptation in its natural host, mutations detected so far are consistent in their location and composition with mutations preferentially induced by human APOBEC3A/B/H. Although concerns surrounding MPXV appear to have abated (Burki 2023) in high income countries, this pandemic continues to unfold elsewhere and its natural reservoir remains unknown. Given the great potential for further adaptation for improved transmissibility, aided in part by mutations induced by human APOBEC3, our analyses add urgency to the task of containing human MPXV transmission and uncovering the ecology of the virus in its reservoir host.

Methods

Data

All genome data were downloaded from NCBI using the Accession numbers in Supplementary Table 1. The tissue tropism data described in the main text were obtained from the Genotype-Tissue Expression (GTEx) Portal (<https://www.gtexportal.org/home/index.html>) on 13 June 2023.

CDUR plots

All CDUR statistics data (Figs 1, 2, 4, 6 and Supplementary Figs 1, 2, and 3; Tables 1 and 2, and Supplementary Tables 2, 4, 7, and 9) were produced by running CDUR against the coding sequences of the NCBI genomes listed in Supplementary Table 1. CDUR is available for download on <https://gitlab.com/maccarthyslab/CDUR>. The Supplementary data files contain the values 'below_X', hotspot representation for mutational motif X, and 'repTrFrac_belowX', amino acid replacement for hotspot X, for each gene in the given genome. Figs 1, 2, 4, and 6 plot 'repTrFrac_belowX' against 'below_X' using the R ggplot package.

GC content plots

The GC content of the four pox viruses was calculated by using the EMBOS function geecee which computes the fraction of G+C bases found in a given sequence by finding the mean value, dividing the sum of G and C bases with the length of the entire input sequence. We ran geecee for the coding sequences of the NCBI genomes listed in Supplementary Table 1, in turn calculating the average G and C count over each genome (<https://emboss.sourceforge.net/apps/cvs/emboss/apps/geecee.html>).

Tables

From the CDUR results, we calculated 'distance from the top left corner' as the Euclidean distance between the point (0,1) and the location of the amino acid replacement and hotspot representation obtained from CDUR for each gene. Table 1 shows the 15

MPXV genes with CDUR results for the hotspot TC that are farthest away from the top left corner (0,1). Table 2 shows the 15 genes with CDUR results corresponding to GC that are closest to the top left corner (0,1).

In the Supplementary Tables S2, 4, and 7, we provide all measures including metadata from the NCBI genome (gene id, locus_tag, dbxref, protein id, location start and end, gbkey) as well as calculated measures for gene length and distance from chromosome ends. The repeat region was extracted from the NCBI annotations for MPXV. Gene length was calculated from the annotated NCBI data, subtracting the location end with the gene location start. Distance from chromosome end was the minimum value between the median of the gene (mid value between location start and location end) subtracted from either genome end. The columns UKP2_id in Supplementary Tables 2 and 4 and MPXV_id in Supplementary Table 7 were found by using BLAST to align the MPXV Reference genome with the MPXV-UK_P2 genome, providing a one-to-one alignment between the two genomes. The 'Mutation count' column was taken from Supplementary Table 3 of Isidro et al. (2022) and mapped to the corresponding UKP2 genes. We calculated the mutation frequency as mutation count divided by gene length.

Statistical analysis

Duplicate genes in MPXV (specifically the four repeated ITR genes) were removed, leaving a single copy, to conduct the statistical analysis in Supplementary Tables 3 and 6, to avoid double-counting. In Supplementary Tables 3, 5, 6 and 8, Pearson correlations were calculated using the built in cor() function in R and P values were obtained using the R linear model fit function lm(). To evaluate MPXV-UK_P2 mutations, we estimated P-values and correlations by removing genes without mutations. Welch two sample t-tests were obtained using the t.test() function in R. For the comparison of VD21 and VARV only paired genes were used to conduct a Welch two sample t-test, and any unmatched genes between the genomes were not included.

Supplementary data

Supplementary data is available at *Virus Evolution* Journal online.

Funding

The study was supported by grants NIH R01AI132507 to T.M. and NFRFE-2021-00933 (Canada) to M.L., L.D., and T.M. The study was supported by NIH R01AI139106 to J.L., a fund from the River Valley Ovarian Cancer Coalition, and a start-up by UAMS Department of Microbiology and Immunology to J.L.

Conflict of interest: None declared.

References

- Alcamí, A. et al. (1998) 'Blockade of Chemokine Activity by a Soluble Chemokine Binding Protein from Vaccinia Virus', *The Journal of Immunology*, 160: 624–33.
- Bogerd, H. P. et al. (2006) 'APOBEC3A and APOBEC3B are Potent Inhibitors of LTR-retrotransposon Function in Human Cells', *Nucleic Acids Research*, 34: 89–95.
- Brinkmann, A. et al. (2022) 'Possible Adaptation of the 2022 Monkeypox Virus to the Human Host through Gene Duplication and Loss', *bioRxiv*, 2022.10.21.512875.
- Burki, T. (2023) 'The End of the Mpox Pandemic?', *The Lancet Infectious Diseases*, 23: 159–60.
- Carvalho, T. (2022) 'The Unknown Efficacy of Tecovirimat against Monkeypox', *Nature Medicine*, 28: 2224–5.
- Cheng, A. Z. et al. (2019) 'Epstein–Barr Virus BORF2 Inhibits Cellular APOBEC3B to Preserve Viral Genome Integrity', *Nature Microbiology*, 4: 78–88.
- Chen, J., MacCarthy, T., and Matsen, F. A. (2017) 'The Preferred Nucleotide Contexts of the AID/APOBEC Cytidine Deaminases Have Differential Effects When Mutating Retrotransposon and Virus Sequences Compared to Host Genes', *PLOS Computational Biology*, 13: e1005471.
- Dobrovolná, M. et al. (2023) 'Inverted Repeats in the Monkeypox Virus Genome are Hot Spots for Mutation', *Journal of Medical Virology*, 95: e28322.
- Duggan, A. T. et al. (2016) '17th Century Variola Virus Reveals the Recent History of Smallpox', *Current Biology*, 26: 3407–12.
- Duraffour, S. et al. (2008) 'Specific Targeting of the F13L Protein by ST-246 Affects Orthopoxvirus Production Differently', *Antiviral Therapy*, 13: 977–90.
- Elde, N. C. et al. (2012) 'Poxviruses Deploy Genomic Accordions to Adapt Rapidly against Host Antiviral Defenses', *Cell*, 150: 831–41.
- Fanunza, E. et al. (2023) 'Human Cytomegalovirus Mediates APOBEC3B Relocalization Early during Infection through a Ribonucleotide Reductase-independent Mechanism', *bioRxiv: The Preprint Server for Biology*, 2023.01.30.526383.
- Forni, D. et al. (2023) 'An APOBEC3 Mutational Signature in the Genomes of Human-Infecting Orthopoxviruses', *mSphere*, 8: e0006223.
- Frenois-Veyrat, G. et al. (2022) 'Tecovirimat Is Effective against Human Monkeypox Virus in Vitro at Nanomolar Concentrations', *Nature Microbiology*, 7: 1951–5.
- Ghazy, R. M. et al. (2023) 'How Can Imported Monkeypox Break the Borders? A Rapid Systematic Review', *Comparative Immunology, Microbiology and Infectious Diseases*, 92: 101923.
- Gigante, C. M. et al. (2022) 'Multiple Lineages of Monkeypox Virus Detected in the United States, 2021–2022', *Science*, 378: 560–5.
- Gileva, I. P. et al. (2006) 'Properties of the Recombinant TNF-binding Proteins from Variola, Monkeypox, and Cowpox Viruses are Different', *Biochimica Et Biophysica Acta (BBA)—Proteins and Proteomics*, 1764: 1710–8.
- Gubser, C. et al. (2004) 'Poxvirus Genomes: A Phylogenetic Analysis', *Journal of General Virology*, 85: 105–17.
- Hendrickson, R. C. et al. (2010) 'Orthopoxvirus Genome Evolution: The Role of Gene Loss', *Viruses*, 2: 1933–67.
- Hughes, A. L., and Friedman, R. (2005) 'Poxvirus Genome Evolution by Gene Gain and Loss', *Molecular Phylogenetics and Evolution*, 35: 186–95.
- Hulo, C. et al. (2011) 'ViralZone: A Knowledge Resource to Understand Virus Diversity', *Nucleic Acids Research*, 39: D576–582.
- Isidro, J. et al. (2022) 'Phylogenomic Characterization and Signs of Microevolution in the 2022 Multi-country Outbreak of Monkeypox Virus', *Nat Med*, 28: 1569–72.
- Jalili, P. et al. (2020) 'Quantification of Ongoing APOBEC3A Activity in Tumor Cells by Monitoring RNA Editing at Hotspots', *Nature Communications*, 11: 2971.
- Johnson, P. L. F. et al. (2022) 'Evolutionary Consequences of Delaying Intervention for Monkeypox', *The Lancet*, 400: 1191–3.
- Kaler, J. et al. (2022) 'Monkeypox: A Comprehensive Review of Transmission, Pathogenesis, and Manifestation', *Cureus*, 14: e26531.
- Koning, F. A. et al. (2009) 'Defining APOBEC3 Expression Patterns in Human Tissues and Hematopoietic Cell Subsets', *Journal of Virology*, 83: 9474–85.
- Kremer, M. et al. (2006) 'Vaccinia Virus Replication Is Not Affected by APOBEC3 Family Members', *Virology Journal*, 3: 86.

- Lackey, L. et al. (2012) 'APOBEC3B and AID Have Similar Nuclear Import Mechanisms', *Journal of Molecular Biology*, 419: 301–14.
- et al. (2013) 'Subcellular Localization of the APOBEC3 Proteins during Mitosis and Implications for Genomic DNA Deamination', *Cell Cycle*, 12: 762–72.
- Langenbucher, A. et al. (2021) 'An Extended APOBEC3A Mutation Signature in Cancer', *Nature Communications*, 12: 1602.
- MacMillan, A. L., Kohli, R. M., and Ross, S. R. (2013) 'APOBEC3 Inhibition of Mouse Mammary Tumor Virus Infection: The Role of Cytidine Deamination versus Inhibition of Reverse Transcription', *Journal of Virology*, 87: 4808–17.
- Martinez, T. et al. (2019) 'Evolutionary Effects of the AID/APOBEC Family of Mutagenic Enzymes on Human Gamma-herpesviruses', *Virus Evolution*, 5: vey040.
- Matias, W. R. et al. (2022) 'Tecovirimat for the Treatment of Human Monkeypox: An Initial Series from Massachusetts, United States', *Open Forum Infectious Diseases*, 9: ofac377.
- Mauldin, M. R. et al. (2022) 'Exportation of Monkeypox Virus from the African Continent', *The Journal of Infectious Diseases*, 225: 1367–76.
- McFadden, G. (2005) 'Poxvirus Tropism', *Nature Reviews Microbiology*, 3: 201–13.
- McLysaght, A., Baldi, P. F., and Gaut, B. S. (2003) 'Extensive Gene Gain Associated with Adaptive Evolution of Poxviruses', *Proceedings of the National Academy of Sciences*, 100: 15655–60.
- Mehta, H. V. et al. (2012) 'IFN α and LPS Up-regulate APOBEC3 (A3) mRNA through Different Signaling Pathways', *The Journal of Immunology*, 189: 4088–103.
- Monajemi, M. et al. (2014) 'Positioning of APOBEC3G/F Mutational Hotspots in the Human Immunodeficiency Virus Genome Favors Reduced Recognition by CD8+ T Cells', *PLOS ONE*, 9: e93428.
- Mordstein, C. et al. (2021) 'Transcription, mRNA Export, and Immune Evasion Shape the Codon Usage of Viruses', *Genome Biology and Evolution*, 13: evab106.
- Mühlemann, B. et al. (2020) 'Diverse Variola Virus (Smallpox) Strains Were Widespread in Northern Europe in the Viking Age', *Science*, 369: eaaw8977.
- Münk, C. et al. (2008) 'Functions, Structure, and Read-through Alternative Splicing of Feline APOBEC3 Genes', *Genome Biology*, 9: R48.
- Münk, C., Willemsen, A., and Bravo, I. G. (2012) 'An Ancient History of Gene Duplications, Fusions and Losses in the Evolution of APOBEC3 Mutators in Mammals', *BMC Evolutionary Biology*, 12: 71.
- National Institute of Allergy and Infectious Diseases (NIAID). (2023) *A Randomized, Placebo-Controlled, Double-Blinded Trial of the Safety and Efficacy of Tecovirimat for the Treatment of Human Monkeypox Virus Disease*. clinicaltrials.gov, Report No.: NCT05534984. <<https://clinicaltrials.gov/ct2/show/NCT05534984>> accessed 2 Jun 2023.
- Pak, V. et al. (2011) 'The Role of Amino-Terminal Sequences in Cellular Localization and Antiviral Activity of APOBEC3B ∇ ', *Journal of Virology*, 85: 8538–47.
- Poulain, F. et al. (2020) 'Footprint of the Host Restriction Factors APOBEC3 on the Genome of Human Viruses', *PLOS Pathogens*, 16: e1008718.
- Ratcliff, J., and Simmonds, P. (2021) '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.
- Roper, R. L. et al. (2023) 'Monkeypox (Mpox) Requires Continued Surveillance, Vaccines, Therapeutics and Mitigating Strategies', *Vaccine*, 41: 3171–7.
- Russo, A. T. et al. (2020) 'Co-administration of Tecovirimat and ACAM2000TM in Non-human Primates: Effect of Tecovirimat Treatment on ACAM2000 Immunogenicity and Efficacy versus Lethal Monkeypox Virus Challenge', *Vaccine*, 38: 644–54.
- Salamango, D.J. et al. (2018) 'APOBEC3B Nuclear Localization Requires Two Distinct N-Terminal Domain Surfaces', *Journal of Molecular Biology*, 430: 2695–708.
- Salas-Briceno, K., Zhao, W., and Ross, S. R. (2020) 'Mouse APOBEC3 Restriction of Retroviruses', *Viruses*, 12: 1217.
- Sanjuán, R. et al. (2010) 'Viral Mutation Rates', *Journal of Virology*, 84: 9733–48.
- Sanjuán, R., and Domingo-Calap, P. (2016) 'Mechanisms of Viral Mutation', *Cellular and Molecular Life Sciences*, 73: 4433–48.
- Shapiro, M., Krug, L. T., and MacCarthy, T. (2021) 'Mutational Pressure by Host APOBEC3s More Strongly Affects Genes Expressed Early in the Lytic Phase of Herpes Simplex Virus-1 (HSV-1) and Human Polyomavirus (Hpyv) Infection', *PLOS Pathogens*, 17: e1009560.
- Shapiro, M., Meier, S., and MacCarthy, T. (2018) 'Correction To: The Cytidine Deaminase Under-representation Reporter (CDUR) as a Tool to Study Evolution of Sequences under Deaminase Mutational Pressure', *BMC Bioinformatics*, 19: 256.
- Simmonds, P., Ansari, M. A., and Pichlmair, A. (2021) 'Extensive C->U Transition Biases in the Genomes of a Wide Range of Mammalian RNA Viruses; Potential Associations with Transcriptional Mutations, Damage- or Host-mediated Editing of Viral RNA', *PLOS Pathogens*, 17: e1009596.
- Smithson, C., Imbery, J., and Upton, C. (2017) 'Re-Assembly and Analysis of an Ancient Variola Virus Genome', *Viruses*, 9: 253.
- Stenglein, M. D., Matsuo, H., and Harris, R. S. (2008) 'Two Regions within the Amino-terminal Half of APOBEC3G Cooperate to Determine Cytoplasmic Localization', *Journal of Virology*, 82: 9591–9.
- Stopak, K. et al. (2003) 'HIV-1 Vif Blocks the Antiviral Activity of APOBEC3G by Impairing Both Its Translation and Intracellular Stability', *Molecular Cell*, 12: 591–601.
- Stuart, J. H. et al. (2016) 'Vaccinia Virus Protein C6 Inhibits Type I IFN Signalling in the Nucleus and Binds to the Transactivation Domain of STAT2', *PLoS Pathogens*, 12: e1005955.
- Suspène, R. et al. (2006) 'Twin Gradients in APOBEC3 Edited HIV-1 DNA Reflect the Dynamics of Lentiviral Replication', *Nucleic Acids Research*, 34: 4677–84.
- Unterholzner, L. et al. (2011) 'Vaccinia Virus Protein C6 Is a Virulence Factor that Binds TBK-1 Adaptor Proteins and Inhibits Activation of IRF3 and IRF7', *PLoS Pathogens*, 7: e1002247.
- ViralZone and SIB Swiss Institute of Bioinformatics (2023), *Monkeypoxvirus genome*. <<https://viralzone.expasy.org/9959>> accessed Jan 2023.
- Wertheim, J. O. (2022) 'When Viruses Become More Virulent', *Science*, 375: 493–4.
- Willems, L., and Gillet, N. A. (2015) 'APOBEC3 Interference during Replication of Viral Genomes', *Viruses*, 7: 2999–3018.
- World Health Organization (WHO) (2023), *Monkeypox*. <<https://www.who.int/news-room/fact-sheets/detail/monkeypox>> accessed Jan 2023.
- Yang, Z. et al. (2011) 'Expression Profiling of the Intermediate and Late Stages of Poxvirus Replication', *Journal of Virology*, 85: 9899–908.