iScience

Article

Molecular evolution of 2022 multi-country outbreak-causing monkeypox virus Clade IIb

Perumal
Arumugam
Desingu, Tirutani
Parthasarathi
Rubeni, K.
Nagarajan,
Nagalingam Ravi
Sundaresan

padesingu@gmail.com (P.A.D.)
rsundaresan@iisc.ac.in (N.R.S.)

Highlights

Clade IIb Mpoxv viruses
have specific genetic
changes compared to the
other two Clades

Clade I and Clade IIb
viruses' specific genetic
variation observed in
multiple genes

Clade IIa and Clade IIb
viruses' specific genetic
variation observed in
multiple genes

Genetic variations of
genomic gaps, frameshift,
nonsense mutations, and
tandem repeats

Article

Molecular evolution of 2022 multi-country outbreak-causing monkeypox virus Clade IIb

Perumal Arumugam Desingu,[1,4,*] Tirutani Parthasarathi Rubeni,[1] K. Nagarajan,[2,3] and Nagalingam Ravi Sundaresan[1,*]

SUMMARY

The monkeypox virus (Mpoxv) Clade IIb viruses that caused an outbreak in 2017–18 in Nigeria and its genetically related viruses have been detected in many countries and caused multi-country outbreak in 2022. Since the pandemic-causing Mpoxv Clade IIb viruses are closely related to Clade IIa viruses which mostly cause endemic, the Clade IIb Mpoxv might have certain specific genetic variations that are still largely unknown. Here, we have systematically analyzed genetic alterations in different clades of Mpox viruses. The results suggest that the Mpoxv Clade IIb have genetic variations in terms of genomic gaps, frameshift mutations, in-frame nonsense mutations, amino acid tandem repeats, and APOBEC3 mutations. Further, we observed specific genetic variations in the multiple genes specific for Clade I and Clade IIb, and exclusive genetic variations for Clade IIa and Clade IIb. Collectively, findings shed light on the evolution and genetic variations in the outbreak of 2022 causing Mpoxv Clade IIb.

INTRODUCTION

Human monkeypox is a zoonotic disease caused by the monkeypox virus (Mpoxv), a double-stranded DNA virus belonging to orthopoxvirus, closely related to the variola virus, cowpox virus (CPXV) and vaccinia virus (VACV).[1–5] The first human case of monkeypox was identified in the Democratic Republic of the Congo in 1970, Mpoxv evolved into Central African (CA-Mpoxv/Clade I) and West African (WA-Mpoxv/Clade IIa) clades[1–5] and caused outbreaks in West and Central Africa until 2003.[6–8] The Clade I Mpoxv viruses caused periodic outbreaks in Central African countries in 1981–86, 1996–97, 2003, 2005–07, and 2013.[9,10] Similarly, the Clade IIa Mpoxv viruses caused a short chain of outbreaks between 1970 and 1981 in West African countries, including Cote d'Ivoire, Liberia, Nigeria, and Sierra Leone.[6] In the USA, the human outbreak of the Clade IIa Mpoxv virus was observed in 2003, spread from an infected prairie dog traveled from Ghana.[6–8] The Clade IIb viruses caused the outbreak in Nigeria in 2017-18 and further spread to countries such as the United Kingdom,[11] Israel,[8] and Singapore.[12] The Clade IIb Mpoxv viruses exhibit 99.6% and 99.7% nucleotide identity with Clade I and Clade IIa viruses, respectively, at the whole genome level.[13] Furthermore, the Clade IIb Mpoxv viruses are phylogenetically related to Clade IIa viruses at the level of nucleotide identity, suggesting possible micro-evolution in Mpox viruses.[4,13,14]

The evolution and genetic diversity of poxviruses occur through homologous and nonhomologous recombination, gene duplications, gene gain, and loss.[15,16] While gene loss occurs on multiple independent occasions, gene inactivation occurs through frameshift mutations, nonsense mutation, within gene deletions, larger genomic region deletions, and complete deletion.[15–17] The gene deletion mainly occurs at both terminals of the genome, whereas frameshift mutations occur more frequently in genes located in the central region.[15,16] Notably, 2,502 accessory genes are inactivated in orthopoxviruses, of which 795 by frameshift mutations, 947 due to within gene deletions, and 759 through complete deletion.[15] Studies indicate that gene gain by horizontal gene transfer or recombination and gene loss/inactivation through frameshift mutations or genomic region deletion determine host range, evasion of the host immunity, pathogenicity, and virulence of Mpoxv.[15,18] Since three clades of Mpoxv display differences in disease manifestation, geographical spread, and transmission patterns, these clades might have marked genetic variations including gene gain and gene loss/inactivation through frameshift mutations or genomic region deletion.

Previous studies indicate that Mpox viral proteins, B10R and B14R are truncated when compared to Clade I, while D14L protein is absent in Clade IIa.[7,19] Notably, a few Clade I Mpoxv viruses have been found to lose the MPV-Z-N2R/OMCP protein during evolution, associated with better fitness in humans.[20] In addition, APOBEC3 mutations were enriched in Mpoxv Clade IIb.1 lineages responsible for the pandemic in 2022 compared to Mpoxv Clade IIb A lineages that caused the outbreak in Nigeria during 2017–18.[4,21] Therefore, the APOBEC3 mutations are considered to as the dominant driver of adaptive microevolution in humans.[4,21,22] However, the APOBEC3 mutations comparison between the Mpoxv clades is not fully understood. Our previous study found that point mutations accumulated in Mpoxv Clade IIb A, A.1, A.1.1, A.2, and B.1 lineages from 2017 to 2022 may be due to selection pressure.[13] Interestingly, a few virus sequences with gene duplication

[1]Department of Microbiology and Cell Biology, Indian Institute of Science, Bengaluru 560012, India
[2]Department of Veterinary Pathology, Madras Veterinary College, Vepery, Chennai 600007, Tamil Nadu
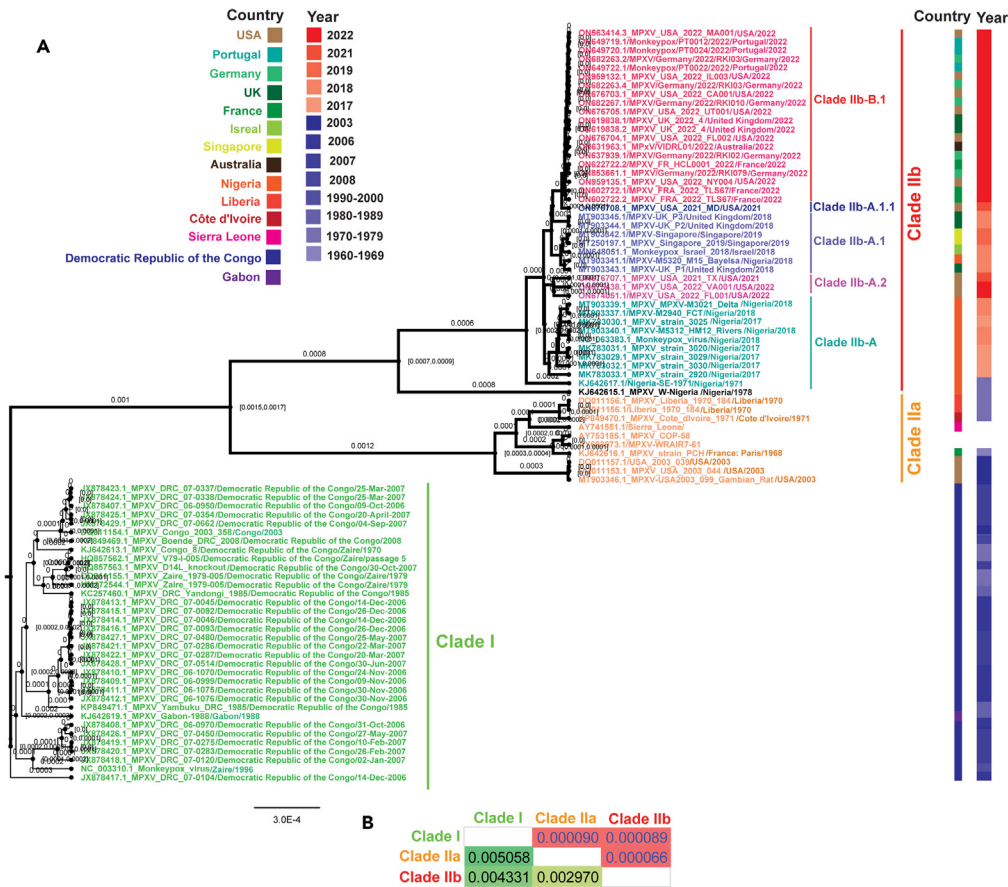[3]Veterinary and Animal Sciences University (TANUVAS)
[4]Lead contact
*Correspondence: padesingu@gmail.com (P.A.D.), rsundaresan@iisc.ac.in (N.R.S.)
https://doi.org/10.1016/j.isci.2023.108601

Figure 1. Complete genome nucleotide sequence level genetic phylogenetic in the Mpoxv

(A) Whole genome nucleotide sequences based on phylogenetic analysis without alignment curation/trimming of sequences separated them into Clade I, Clade IIa, Clade IIb-A, A.1, A.1.1, A.2, and B.1 lineages.

(B) NBGMD analysis based on whole genome nucleotide sequences revealed less than 0.5% nucleotide diversity among clades of the Mpox virus. The standard error estimated in the NBGMD analysis for the Mpox virus was displayed above the diagonal.

and deletion by translocation of the genomic region have also been detected in the 2022 monkeypox outbreak.[23] Similarly, some viral sequences with the deletion of 2.3–15 kb and with more complex genomic rearrangements also were detected in the 2022 Mpoxv outbreak.[24] In this line, a few viral genomes with ITR regions expanded from 6.4 kb to 24.6 kb, and extensive gene duplication and loss were also observed in the 2022 outbreak.[25] Also, some viral genomes with frameshift mutation, deletion mutation, and nonsense mutations to interrupt ORFs have been documented in the 2022 Mpoxv outbreak.[26,27] However, genomic regions, frameshift mutations, in-frame nonsense mutations, and amino acid tandem repeats specific to different clades of Mpoxv are largely unknown. Generally, the alignment curation/trimming methods used in the genome alignment of Mpox viruses are expected to remove the gaps/genetically altered sites such as gene gain and gene loss/inactivation in different clades of Mpox viruses. This could have resulted in the inability to recognize the critical genetic alterations in the genome of Mpox viruses that are expected to provide cues, such as the genomic variations responsible for the multi-country outbreak in 2022.

In the present study, we have systematically analyzed genetic alterations in different clades of Mpox viruses without applying the alignment curation/trimming methods. The results suggest the evolution of Mpoxv clade-specific genomic regions, frameshift mutations, in-frame nonsense mutations, amino acid tandem repeats, and APOBEC3 mutations in Mpoxv Clade IIb viruses.

RESULTS

Clade-specific genomic regions are found in the Mpoxv

Previously the phylogenetic analyses were performed for the Mpoxv clades after alignment curation/trimming.[4,21] In this study, we performed phylogenetic analysis on complete genome sequences of Clade I, Clade IIa, Clade IIb A, A.1, A.1.1, A.2 lineages, and representative complete genome sequences of Clade IIb B.1 available in NCBI public database without alignment curation/trimming. In this analysis, the sequences were divided into clades and lineages (Figure 1A) and net between group mean distance (NBGMD) analysis revealed less than 0.5% nucleotide diversity among Clade I, Clade IIa, and Clade IIb viruses at the whole genome level (Figure 1B). Next, we performed the SimPlot analysis
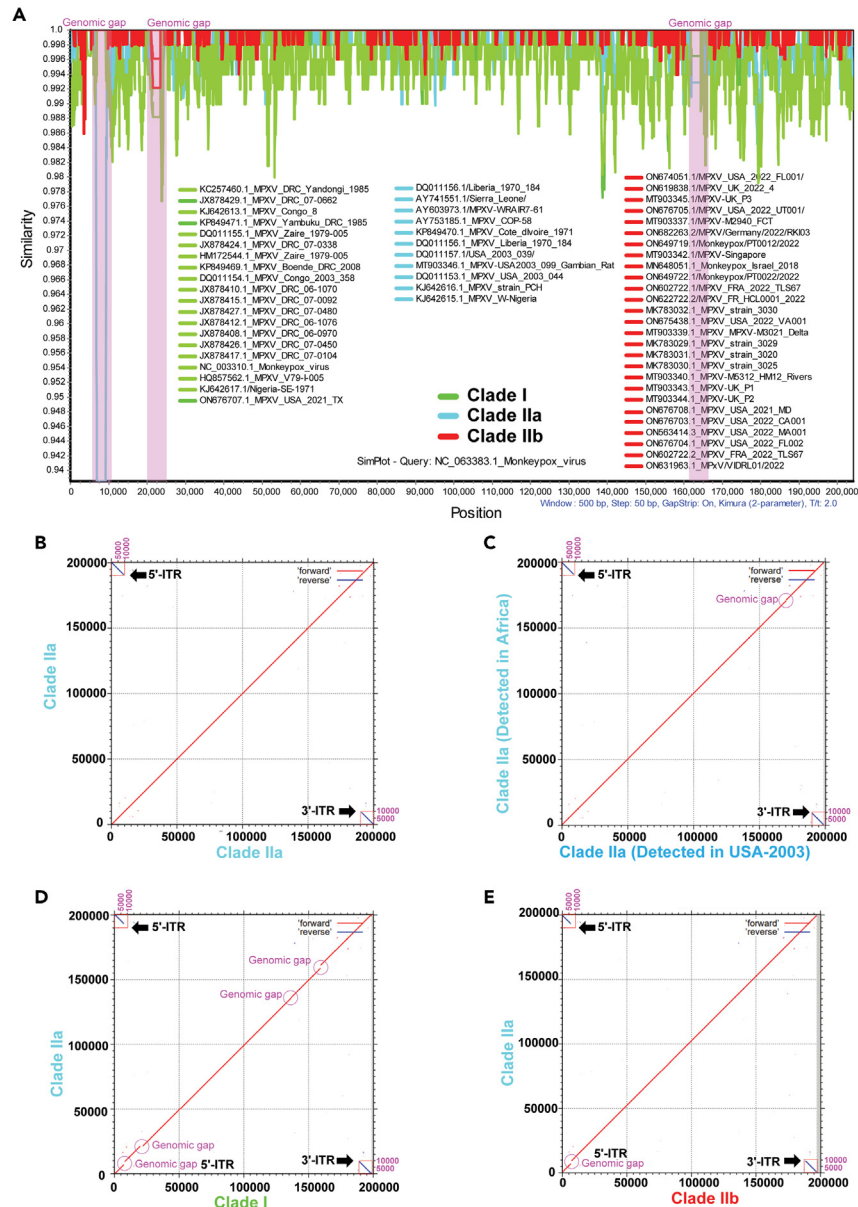
Figure 2. Complete genome nucleotide sequence level genetic diversity in the Mpoxv

(A) The SimPlot analysis depicts the presence of three distinct genomic gaps among clades of the Mpox virus at the whole genome nucleotide sequence levels. The NC_063383.1/Clade IIb is used as a reference query sequence.

(B–E) The Plot LAST hits analysis shows differences in genomic gaps and ITR length between clades of the Mpox virus, and Mpoxv Clade IIa viruses were used as a reference. The inserted boxes in the corners depict the length of the ITR regions in both the ends of the Mpoxv genome and the circle highlighting the break in the red line, indicating the genomic gap/break.

to determine the genomic regions that have the highest genetic diversity among these three clades and found genomic gaps in three clades of Mpoxv (Figure 2A). The LAST hits plot and alignment analysis revealed a genomic gap (170,751–171,230 bp as per DQ011156.1/Liberia-1970-184) in Clade IIa viruses identified in the USA in 2003 when compared to Clade IIa viruses detected in Africa (Figures 2B and 2C). Similarly, four genomic gaps ([1] 7,047–9,357 bp as per DQ011156.1/Liberia-1970-18, [2] 18,952–20,904 bp as per NC_003310.1, [3] 135,913–136,366 bp as per DQ011156.1, and [4] 158,782–161,027 bp as per DQ011156.1) were observed in Clade I viruses as compared to Clade IIa viruses (Figure 2D). Also, a genomic gap (7,047–9,357 bp as per DQ011156.1/Liberia-1970-18) in Clade IIb viruses was noted when compared to Clade IIa viruses (Figure 2E). Notably, similar genomic gaps (7,047–9,357 bp as per DQ011156.1/Liberia-1970-18) are present in Clade I and Clade IIb viruses (Figures 2B–2E). We used the Highlighter tool to determine the nucleotide mismatches and the genomic gaps in the Mpoxv clades.[28] The results suggested two genomic gaps, 18,952–20,904 bp (as per NC_003310.1) and 49,960–50,212 bp (as per NC_003310.1) in Clade IIa and
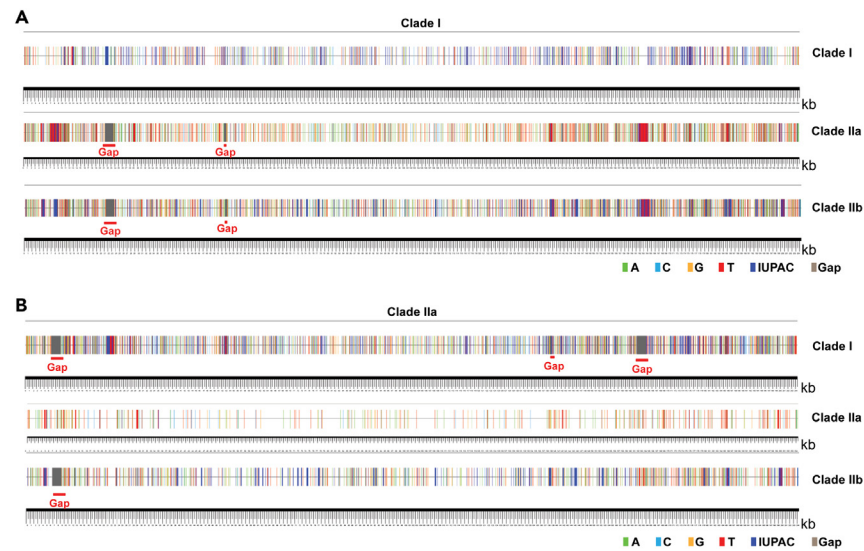
Figure 3. Clade-specific genomic regions among the Mpoxv clades

(A) Nucleotide mismatches Highlighter analysis revealed two genomic gaps in Clade IIa and Clade IIb viruses compared to Clade I. The Mpoxv Clade I viruses were used as a reference. Detail illustrated in Figure S1A.

(B) Nucleotide mismatch analysis indicated three and one genomic gaps in Clade I and Clade IIb, respectively, compared to Clade IIa viruses. The Mpoxv Clade IIa viruses were used as a reference. Detail illustrated in Figure S1B.

Clade IIb viruses when compared to Clade I viruses (Figures 3A and S1A). Similarly, another genomic gap of 7,047–9,357 bp (as per DQ011156.1/Liberia-1970-184) was found in Clade I and Clade IIb viruses in comparison with Clade IIa viruses (Figures 3B and S1B). Moreover, two genomic gaps, 135,913–136,366 bp (as per DQ011156.1/Liberia-1970-184) and 158,782–161,027 bp (as per DQ011156.1/Liberia-1970-184) were observed in Clade I viruses compared to Clade IIa viruses (Figures 3B and S1B). Overall, these findings indicate that genomic gaps exist among the three clades of Mpoxv.

Previous studies have used the predicted ORFs in MT903344.1_Mpoxv-UK_P2 virus or NC_063383.1 virus to identify the nucleotide or amino acid differences between coding regions in different clades/lineages of Mpoxv viruses.[4,5,13] In this study, we determined consensus ORFs from all these predicted ORFs of Mpoxv available in the NCBI database and determined the genes in different clades of Mpoxv (Figure 4A; Figures S2–S5; Data S1, S2, S3, and S4; Table S1) and their frequencies (Figures 4B and S2–S5; Data S1, S2, S3, and S4; Table S1). We also identified/annotated 31 unnamed proteins in VACV-WR and VACV-Cop viruses and named these proteins as MV proteins as stated in the Virus Pathogen Resource (ViPR) online resource tool (Figures 4A, 4B, and S5).

Genetic variations in the Mpoxv Clade IIb viruses compared to Clade I and Clade IIa

Our findings indicate specific genomic gaps and differences in the ORFs in the three clades of Mpoxv. Therefore, we compared the ORFs of Mpoxv Clade IIb viruses with Clade I and Clade IIa to identify the genetic variations. Our analysis found frameshift mutations specific to the Clade IIb viruses in several genes. The C11L gene, which codes for a 36 kDa major membrane protein, is truncated in Clade IIb viruses due to a frame-shift mutation (Figures 5A and S6). Specifically, this frameshift mutation results in an alternate start site of translation resulting in the truncation of the first 26 amino acids in the N-terminal of C11L in Clade IIb viruses (Figures 5A, 5B, and S6). Similarly, the B14R, a gene that codes for a secreted IL-1b binding protein also has frameshift mutation which results in premature termination leading to the loss of 146 amino acids in the C-terminal of B14R in Clade IIb viruses (Figures 5C and S7). Moreover, the C-terminal of B14R differs among the clades of Mpoxv (Figure 5C). The B14R differs by 36 amino acids at the N-terminal and 127 amino acids at the C-terminal in Clade IIa viruses when compared to Clade I viruses (Figures 5C, 5D, and S7). We also found that D18L which codes for a kelch-like protein, initiates translation in the alternate start sites in Clade I, Clade IIa, and Clade IIb viruses due to frameshift mutations (Figures 5E, 5F, and S8). Interestingly, the D18L gene frameshift is followed by the within-gene deletion in the genomic region which is absent in Clade II viruses (18,952–20,904 bp as per NC_003310.1/Clade I) (Figure 4A). Similarly, the B1R gene which codes for BTB-Kelch-domain protein also has a frameshift mutation, followed by the within-gene deletion in Clade I viruses (158,782–161,027 bp as per DQ011156.1/Liberia-1970-184) (Figure 4A). The B1R gene initiates translation at a different start codon and position in Clade I and Clade IIb viruses, which results in the generation of different fragments of the BTB-Kelch-domain protein (Figures 5G, 5H, and S9). We also noted that D4L, B7R, and C12L genes are truncated by in-frame nonsense mutations in Clade IIb viruses (Figures 5I–5K, S10, and S11). In further analysis, we found specific copy number variations in amino acid tandem repeats of Clade IIb viruses. Specifically, the B16R which codes for an IFN-alpha/beta receptor glycoprotein contains two copies of "MK" amino acid tandem repeats in Clade I and Clade IIa viruses, whereas Clade IIb viruses have four copies (Figures 5L and S11). In addition, Clade I, Clade IIa, and Clade IIb viruses have I-MV178 (127 amino acids), IIa-MV179 (134 amino acids), and IIb-MV27 (76 amino acids) proteins, respectively, which contain the IL1-beta inhibitor domain, as per our
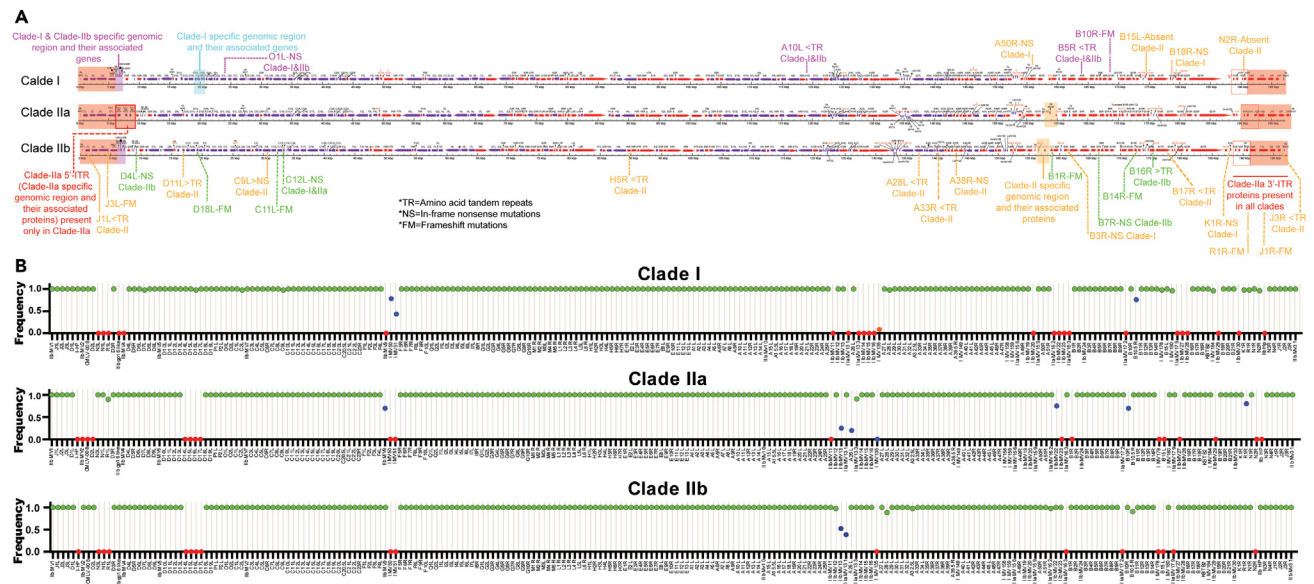
Figure 4. Consensus ORFs for the different clades of Mpoxv

(A) The graphical figure illustrates the Mpoxv clades-specific genomic regions and their associated genes after determining consensus ORFs for the different clades of Mpoxv. Further, the figure also illustrates the Mpoxv clades-specific coding region genetic diversity in terms of frameshift mutations, in-frame nonsense mutations, and amino acid tandem repeats.

(B) The frequency graph depicts the frequency levels of different genes in different clades of Mpoxv. This analysis was performed using almost all high-quality complete genome sequences of Clade I (n = 46), Clade IIa (n = 11), Clade IIb (n = 45), Clade IIb-A (n = 11), and Clade IIb-A.1(n = 9), Clade IIb-A.1.1(n = 1), Clade IIb-A.2 lineages (n = 3), and representative complete genome sequences of Clade IIb-B.1 lineage (n = 21), also, our previous study reported that the 963 high-quality complete genome sequences in the Clade IIb-B.1 lineage (n = 963) with identical ORFs.[13] The ORF annotation and frequency details are provided in Figures S2–S5; Data S1, S2, S3, and S4.

annotations (Figures 4A and 4B; Data S4). Two uncharacterized proteins, IIb-MV23 (72 amino acids) and IIb-MV29 (58 amino acids) are also found uniquely in Clade IIb viruses, which are absent in other clades (Figures 4A and 4B; Data S4). These findings indicate that Clade IIb Mpoxv viruses have specific genetic changes when compared to the other two Clades.
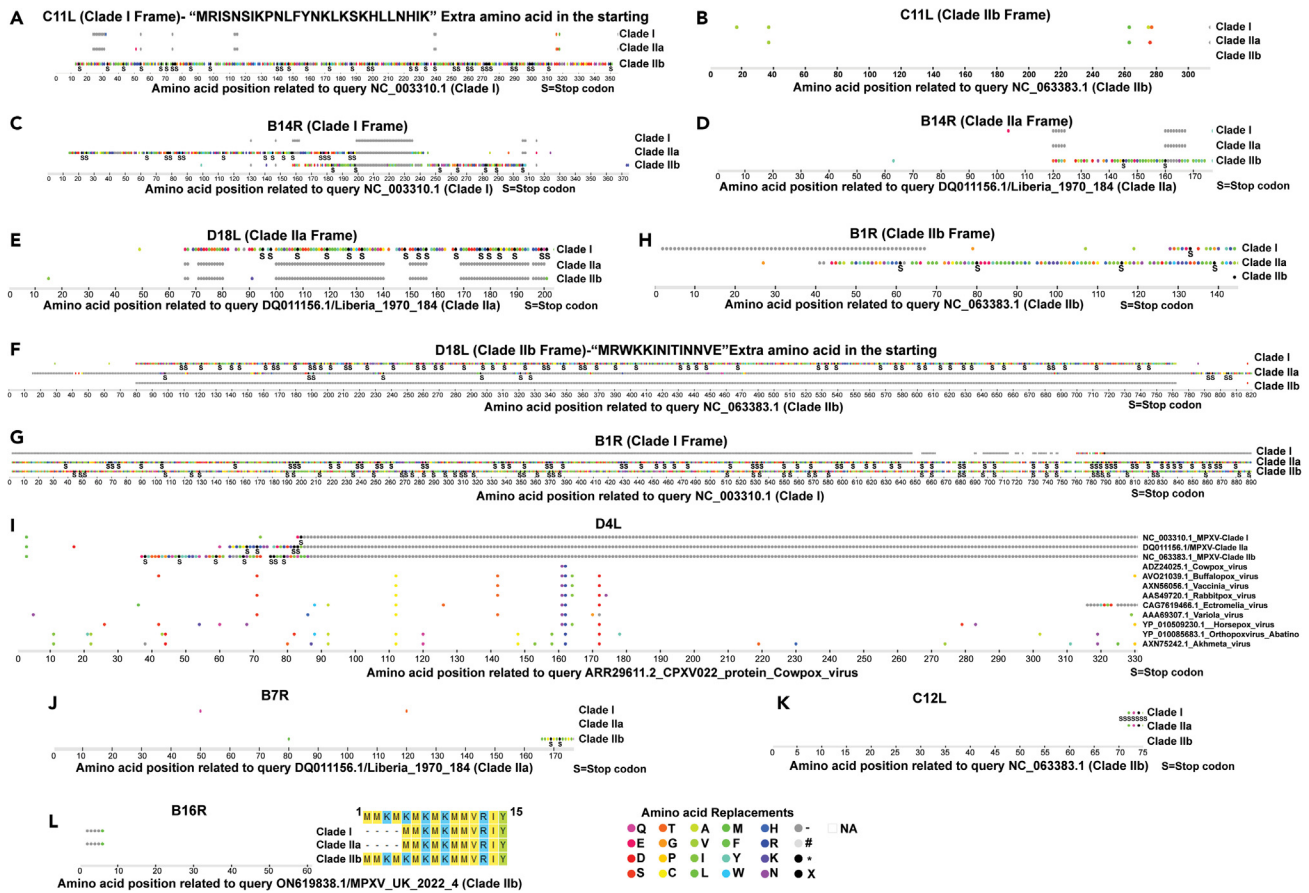
Genetic variations found in Clade I and Clade IIb viruses compared to Clade IIa viruses

Our NCBI database genomic organization analysis revealed that the Clade IIa have longer ITR regions (~9,357 bp) when compared to the ITR regions of Clade I and Clade IIb viruses (~6,400 bp). However, our LAST hits plot and nucleotide mismatches highlighter analyses revealed that 3′-ITR of all three clades of Mpoxv are nearly similar in length (~9,357 bp) whereas the 5′-ITR of Clades I and IIb are shorter (~6,400 bp) due to the genomic deletion (Figures 2D, 2E, 6A, 6B, and S12). Curiously, the genomic gap of 7,047–9,357 bp (as per DQ011156.1/Liberia-1970-184) present in Clade I and Clade IIb viruses is located in the 5′-ITR regions of Clade IIa viruses (Figures 4A, 6B, and S12). This deletion resulted in the loss of R1L, N1L, and N3L in Clades I and IIb of Mpoxv (Figures 6C, 6D, and S13). Our study also reveals that R1R, N1R, and N3R proteins are expressed in two copies in Clade IIa viruses, whereas in Clade I and Clade IIb viruses, in a single copy (Figures 6C–6G, S13, and S14). The nucleotide mismatch analysis revealed that Clade I and Clade IIb viruses have nucleotide sequences coding for other genes instead of N1L, N3L, and R1L (Figure 6B). Therefore, we were interested in determining the genes present in the genomic regions of Clade I and Clade IIb viruses. This analysis revealed the presence of D2L, CMLV-008, and IIb-MV2 genes in Clade I and Clade IIb viruses (Figures 6H, 6I, S15, and S16), instead of N1L, N3L, and R1L genes in Clade IIa.

Our further analysis found that the B10R gene is 103 amino acids truncated in the N-terminal in Clade IIa viruses due to a frameshift mutation (Figures 7A, 7B, and S17). Similarly, we also found an in-frame nonsense mutation in the O1L gene in Clade I and Clade IIb viruses resulting in premature termination (Figures 7C and S18). Moreover, the IIa-MV163 (113AA) and IIa-MV172 (73AA) genes are absent in Clade I and Clade IIb (Figures 4A and 4B). In addition, the A10L protein contains "SN" amino acid tandem repeats of 5–9 copy numbers in Clade IIa viruses, whereas two copy numbers in Clade I and Clade IIb viruses (Figures 7D and S18). Similarly, the B5R of Clade IIa viruses contains three copy numbers of "NSS" repeats, and Clade I and Clade IIb viruses contain two copies (Figures 7E and S19). These findings indicate that Clade I and Clade IIb Mpoxv viruses have genetic variations different from the Clade IIa Mpoxv viruses.

### Genetic variations found in Clade IIa and Clade IIb viruses compared to Clade I viruses

Previous studies have found that the B10R and B14R proteins are truncated, and the D14L protein is absent in Clade IIa viruses compared to Clade I Mpoxv viruses.[7,19] Our nucleotide mismatch highlighter analysis revealed two genomic regions 18,952–20,904 bp (as per

Figure 5. Genetic variations found in the Mpoxv Clade IIb viruses

(A and B) The frameshift mutation in C11L causes genetic diversity in the Clade IIb viruses. According to the Clade I frame, the C11L gene is truncated in Clade-IIb viruses (A), whereas this gene is shorter in the Clade IIb frame and expressed in all clades (B). Details of individual strains are provided in Figure S6; Data S5.

(C and D) The frameshift mutation mediated genetic diversity in the Clade IIb viruses B14R gene. The Clade IIb viruses truncated in both Clade I and Clade IIa frames, but Clade IIa viruses truncate only in the Clade I frame. Detailed in Figure S7; Data S5.

(E and F) The frameshift mutation in D18L causes genetic diversity in the Clade IIb viruses. If the D18L gene initiates translation at the frame of Clade IIa viruses, this gene is truncated in Clade I viruses (E), whereas in the Clade IIb frame, Clade I and Clade IIa virus D18L is truncated (F). Detailed in Figure S8; Data S5.

(G and H) The frameshift mutation mediated genetic diversity in the Clade IIb viruses B1R gene. According to the Clade I frame, the B1R gene is truncated in Clade IIa and Clade IIb viruses (G), whereas this gene in the Clade IIb frame, Clade I, and Clade IIa virus B1R is truncated (H). Detailed in Figure S9; Data S5.

(I) The D4L present in the Clade I viruses are truncated versions of D4L in the other orthopoxvirus, and this gene is further truncated in the Clade IIa and Clade IIb viruses. Detailed in Figure S10; Data S5.

(J) An in-frame nonsense mutation is observed in the B7R gene of the Clade IIb viruses. Detailed in Figure S11; Data S5.

(K) An in-frame nonsense mutation is observed in the C12L gene of the Clade I and Clade IIa viruses. Detailed in Figure S11; Data S5.

(L) The Clade IIb viruses display more copy numbers of the amino acid tandem repeats compared to Clade I and Clade IIa viruses in the B16R protein. Detailed in Figure S11; Data S5.

NC_003310.1) (Figures 8A and S20) and 49,960–50,212 bp (as per NC_003310.1) found Clade I viruses, and absent in Clade II viruses (Figures 8B and S20). Furthermore, our analysis revealed that the 18,952–20,904 bp region of these two genomic regions contains genes such as D14L, D15L, D16L, and D17L, but the 49,960–50,212 bp region is a non-coding region (Figures 8C, 8D, and S21). Moreover, the D14L gene in the Clade I is a fragmented gene in other orthopoxviruses, but this gene is absent in the Clade II viruses (Figures 8C, 8D, and S21). Furthermore, our nucleotide mismatch highlighter analysis revealed that two genomic regions, 135,913–136,366 bp (Figures 8E and S22) and 158,782–161,027 bp (as per DQ011156.1/Liberia-1970-184) (Figures 8F and S22), are found in Clade II viruses, and absent in Clade I viruses. Further, our analysis revealed that among these two genomic regions, the 135,913–136,366 bp region is a non-coding region, and the 158,782–161,027 bp region contains genes IIa-MV163 (Figures 8G and S23) and IIb-MV23 (Figures 8H and S23). Furthermore, our analysis revealed that the IIa-MV163 gene is a fragment of the kelch-like protein in other orthopoxviruses and is only expressed in Clade IIa viruses (Figures 8G and S23), while the IIb-MV23 gene is only expressed in Clade IIb viruses (Figures 8H and S23). However, the functional importance of these genomic regions and their associated proteins are not fully understood.
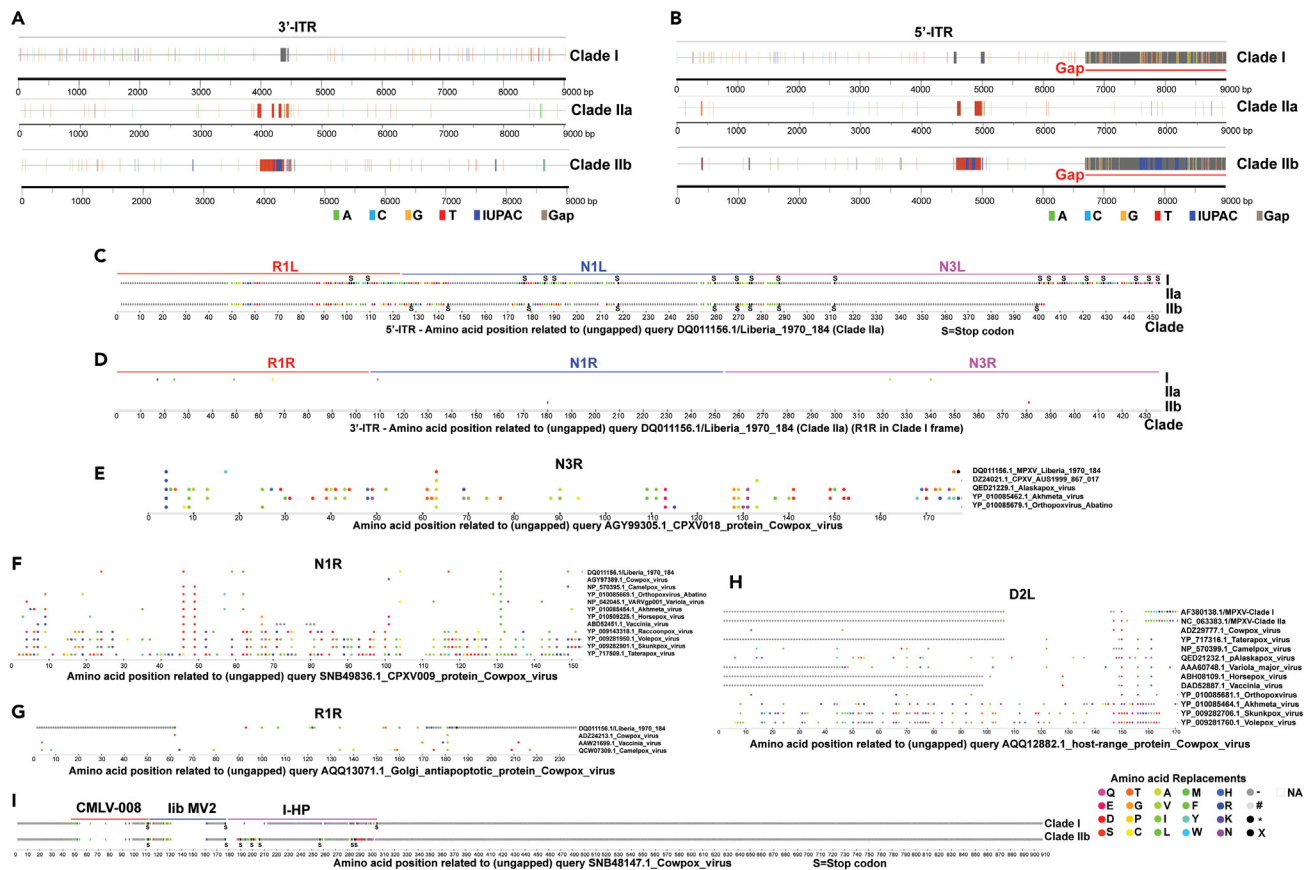
**Figure 6. The unique genetic regions in the Clade I and Clade IIb viruses**

(A) The Clade IIa viruses 3'-ITR regions nucleotide mismatch analysis indicates minimal genetic diversity and without genomic gaps among the Mpoxv clades. The Mpoxv Clade IIa was used as a reference. Details of individual strains are provided in Figure S12A.

(B) The Clade IIa viruses 5'-ITR regions nucleotide mismatch analysis indicates the presence of a genomic gap of ~2,325 bp in Clade I and Clad IIb viruses compared to Clade IIa. Detailed in Figure S12B.

(C) The 5'-ITR regions' genomic gap is associated with N3L, N1L, and R1L genes in Clade IIa viruses, which are absent in Clade I and Clade IIb viruses in the 5'-ITR region. Detailed in Figure S13A; Data S6.

(D) In the 3'-ITR region, N3R, N1R, and R1R genes are present in all the clades of Mpoxv. Detailed in Figure S13B; Data S6.

(E) The N3R gene present in Mpoxv is similar to the N3R gene present in the other orthopoxviruses. Detailed in Figure S14; Data S6.

(F) The N1R gene present in Mpoxv is similar to the N1R gene present in the other orthopoxviruses. Detailed in Figure S14; Data S6.

(G) The R1R in the Mpox viruses is the truncated version of R1R in the other orthopoxvirus. Detailed in Figure S14; Data S6.

(H) The D2L present in the Mpoxv Clade I and Clade IIb is the truncated version of D2L in the other orthopoxvirus. Detailed in Figure S15; Data S6.

(I) The CMLV-008, IIb-MV2, and I-HP genes in the Mpoxv Clade I and Clade IIb are the truncated and fragmented version of the single ankyrin repeat-containing gene in the other orthopoxvirus. Detailed in Figure S16; Data S6.

We found frameshift mutations in R1R, and J1R/J3L genes in Clade I viruses, which resulted in truncations (Figures 6D, 9A–9E, and S24–S26). Our further analysis revealed that the N2R gene is expressed only in Clade I viruses (Figure 5F). Interestingly, the loss of the MPV-Z-N2R/OMCP gene in some of the Clade I Mpoxv has been associated with human-to-human transmission.[20] We also found in-frame nonsense mutations in genes such as C9L, B15L, A38R B3R, B18R, K1R, and A50R in Clade I viruses (Figures 9G–9M, and S27–S29). Further, we found variations in copy numbers of amino acid tandem repeats in proteins such as D11L, B17R, A28L, J1L/J3R, A33R, and H5R between Clade I and Clade II viruses (Figures 10A–10G, and S30–S32). We also found that the IIb-MV9, IIb-MV22, IIb-MV28, IIa-MV162, IIb-MV30 (116AA), Gp16-like and IIb-MV4 proteins are present in Clade IIa/Clade IIb viruses, but absent in Clade I viruses (Figures 4A and 4B). These findings reveal unique genetic variations between Clade I and II Mpoxv.

Analysis of APOBEC3 mutations in the clades of Mpoxv

Comparisons within clades of Mpoxv have shown that APOBEC3 mutations were abundant in Clade IIb viruses, suggesting that Clade IIb viruses have undergone microevolution for human adaptation.[4,21,22] However, APOBEC3 mutations between clades of Mpoxv are largely unknown. Our findings revealed that Clade I and Clade IIb viruses had enhanced nucleotide mismatches in genes transcribed in both forward
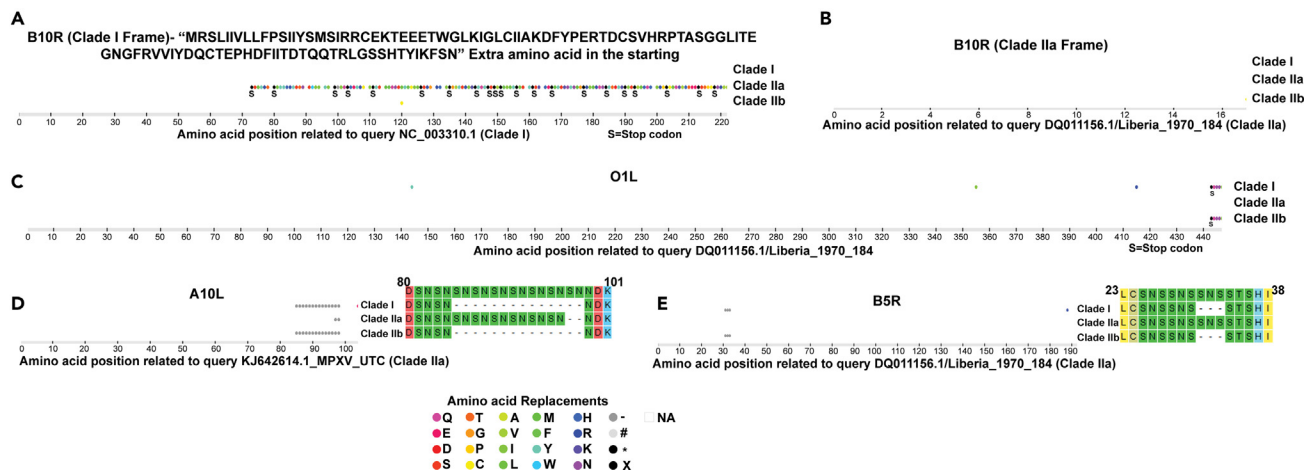
Figure 7. The coding region genetic variations found in the Clade I and Clade IIb

(A and B) The frameshift mutation mediated genetic diversity in the Clade I and Clade IIb viruses B10R gene. If the B10R gene initiates translation at the frame of Clade I viruses, this gene is truncated in Clade IIa viruses (A), whereas in the Clade IIa frame, a shorter gene is present in all the Mpoxv clades (B). Detailed in Figure S17; Data S6.

(C) An in-frame nonsense mutation is observed in the O1L gene of the Clade I and Clade IIb viruses. Detailed in Figure S18; Data S6.

(D) The Clade I and Clade IIb viruses display lesser copy numbers of the amino acid tandem repeats compared to Clade IIa viruses in the A10L protein. Detailed in Figure S18; Data S6.
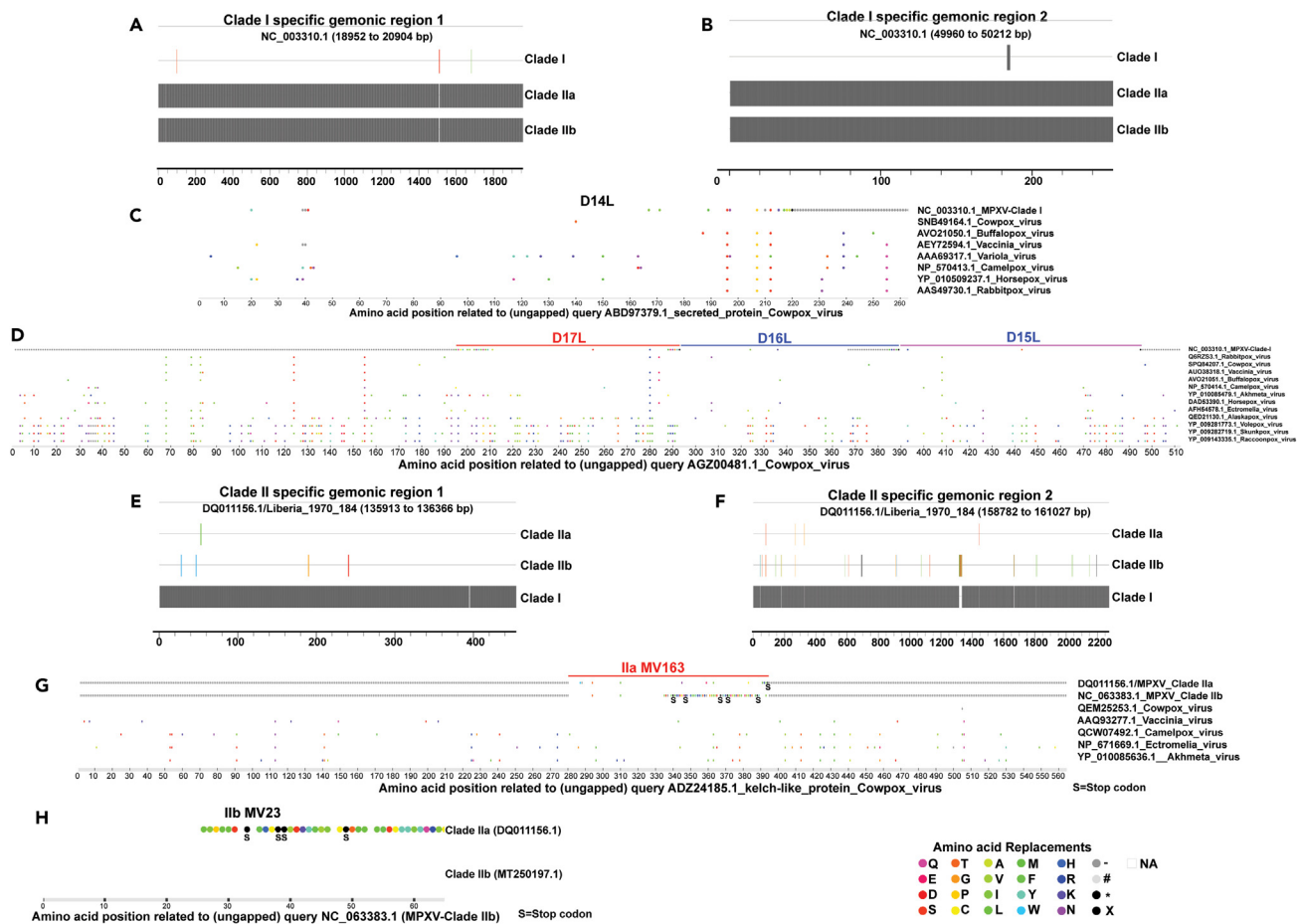
(E) The Clade I and Clade IIb viruses display lesser copy numbers of the amino acid tandem repeats compared to Clade IIa viruses in the B5R protein. Detailed in Figure S19; Data S6.

and reverse directions compared to Clade IIa viruses (Figures 11A, 11B, and S33). Notably, most nucleotide mismatches are transition (Figures 11C, 11D, and S34) and silent mutations (Figures 11E, 11F, and S35). The Clade I and Clade IIb viruses showed abundant APOBEC3 mutations, especially the GA-to-AA mutation in genes transcribed in forward and reverse directions compared to Clade IIa viruses (Figures 12A–12H, S36 and S37). The Clade I and Clade IIb viruses enriched with A-to-G mutations compared to Clade IIa viruses (Figures 12C–12H). Furthermore, dN/dS analysis revealed the existence of negative selection pressure in genes transcribed in forward and reverse directions in Clade I and Clade IIb viruses (Figures 12I and 12J), indicating the presence of negative selection. We also found that the Clade I and Clade IIb viruses had more transition mutations and APOBEC3 mutations compared to Clade I viruses at the whole genome level, and GA-to-AA mutation was abundant among these APOBEC3 mutations (Figures 12K–12O, S38, and S39). These findings indicate that Clade I/Clade IIb viruses accumulated APOBEC3 mutations during evolution.

DISCUSSION

This study sheds light on the genetic variations in different clades of Mpox viruses with different types of disease manifestation, unveils the genetic variations in the Clade IIb viruses, and also discloses the combinations of shared genetic variation with other clades of viruses. The Clade IIb viruses caused outbreaks in Nigeria in 2017–18,[8] and the virus was transmitted by travelers to Asian and European countries in 2017-18.[11,12] However, Clade IIb viruses causing outbreaks in non-endemic countries from 2022 without an epidemiological link to Africa. Although the Clade I viruses have been endemic in Africa since the first detection in the 1970s,[29] Clade IIb viruses have been documented outside Africa since 2017.[4,21] Studies indicate that the accumulation of APOBEC3 mutations and point mutations within Clade IIb viruses from 2017 to 2022 may be responsible for the 2022 outbreak.[4,21] Our work uncovered several unique genetic variations in Clade IIb viruses, which could be responsible for the pandemic outbreak of 2022.

A previous study found that D10L, D14L, B10R, B14R, and B19R as probable virulence genes in Mpoxv, and D14L could be a leading candidate.[7,19] Interestingly, the vaccinia virus strain lacking the B14R ortholog exhibited changes in virulence.[30,31] Our analysis revealed that genetic variations are found in several genes of Clade IIb viruses, including the C11L, D18L, B1R, B7R, B14R, D4L, IIb-MV27, B16R, and uncharacterized genes such as IIb-MV23, and IIb-MV29. We found that C11L, a major membrane protein is truncated in 26 amino acids in Clade IIb viruses by frameshift mutation. We also observed the truncation of B14R in Clade IIb viruses by frameshift mutation, whereas the truncation pattern is different from the Clade IIa. Vaccinia viruses express C2 and F3 which are BTB-Kelch proteins expressed in early infection, known to affect the outcome of infection.[32] Similarly, ectromelia virus, a mousepox virus, also contains a family of four BTB/kelch proteins, and interestingly, EVM150, one of the four BTB/kelch proteins is an inhibitor of the cellular NF-κB pathway, an important component of the antiviral response.[33] In our work, we found two Kelch-domain-containing genes (D18L and B1R) that exhibit frameshift mutations in Clade IIb viruses. Furthermore, although there are no experimental studies of whether proteins are translated in the alternate frame adopted by frameshift mutations in poxviruses to the author's best knowledge, previous studies using viral genomic data analysis have clarified that frameshift mutations play an

Figure 8. The unique genetic regions in the Clade IIa and Clade IIb viruses

(A) The nucleotide mismatch analysis displays the lack of the genomic region 18,952 to 20,904 bp (as per NC_003310.1/Clade I) in the Clade IIa and Clade IIb compared to Clade I viruses. Details of individual strains are provided in Figure S20A.

(B) The nucleotide mismatch analysis displays the lack of the genomic region 49,960 to 50,212 bp) (as per NC_003310.1/Clade I) in the Clade IIa and Clade IIb viruses compared to Clade I viruses. The Mpoxv Clade I viruses were used as a reference. Detailed in Figure S20B.

(C) The genomic region 18,952 to 20,904 bp (as per NC_003310.1/Clade I) contains the D14L gene and is a truncated version of the D14L gene present in other orthopoxviruses. Detailed in Figure S21; Data S7.

(D) The genomic region 18,952 to 20,904 bp (as per NC_003310.1/Clade I) contains the D15L, D16L, and D17L genes, which are a truncated fragmented version of the single Kelch-repeat containing gene in other orthopoxviruses. Detailed in Figure S21; Data S7.

(E) The nucleotide mismatch analysis displays the presence of the extra genomic region 135,913 to 136,366 bp (as per DQ011156.1/Liberia_1970_184/Clade IIa) in the Clade IIa and Clade IIb viruses compared to Clade I viruses. Detailed in Figure S22A.

(F) The nucleotide mismatch analysis displays the presence of the extra genomic region 158,782 to 161,027 bp (as per DQ011156.1/Liberia_1970_184/Clade IIa) in the Clade IIa and Clade IIb viruses compared to Clade I viruses. Detailed in Figure S22B.

(G) The genomic region 158,782 to 161,027 bp (as per DQ011156.1/Liberia_1970_184/Clade IIa) contains the IIa-MV163 gene only in Clade IIa and is a truncated version of the gene present in other orthopoxviruses. Detailed in Figure S23; Data S7.

(H) The genomic region 158,782 to 161,027 bp (as per DQ011156.1/Liberia_1970_184/Clade IIa) contains the IIb-MV23 gene only in Clade IIb and is a truncated version of the gene present in other orthopoxviruses. Detailed in Figure S23; Data S7.

important role in the evolution of poxviruses.[15,18,34] We believe that these genetic variations observed in the Clade IIb viruses might be involved in the pathogenesis.

Poxviruses have identical and reverse complementary ITR regions at the end of the genome.[35,36] Whereas genes acquired/transferred in one ITR region are often duplicated in the next ITR region through replication-based mechanisms.[35] Notably, the variations in gene-copy numbers are reported to regulate virus replication[35,37,38] and virus pathogenesis.[16,20,37] However, a portion of 3′-ITR deletion was reported in the Clade I and is linked with fitness in humans.[20] In the present study, we observed that the Clade I and Clade IIb viruses have unique genetic variations such as a shorter ITR region associated with the reduced copy numbers of R1R, N1R, and N3R genes. Similarly, variations are also noted in the genomic region and their associated genes, especially in D2L. Further, Clade I and Clade IIb viruses have unique genetic
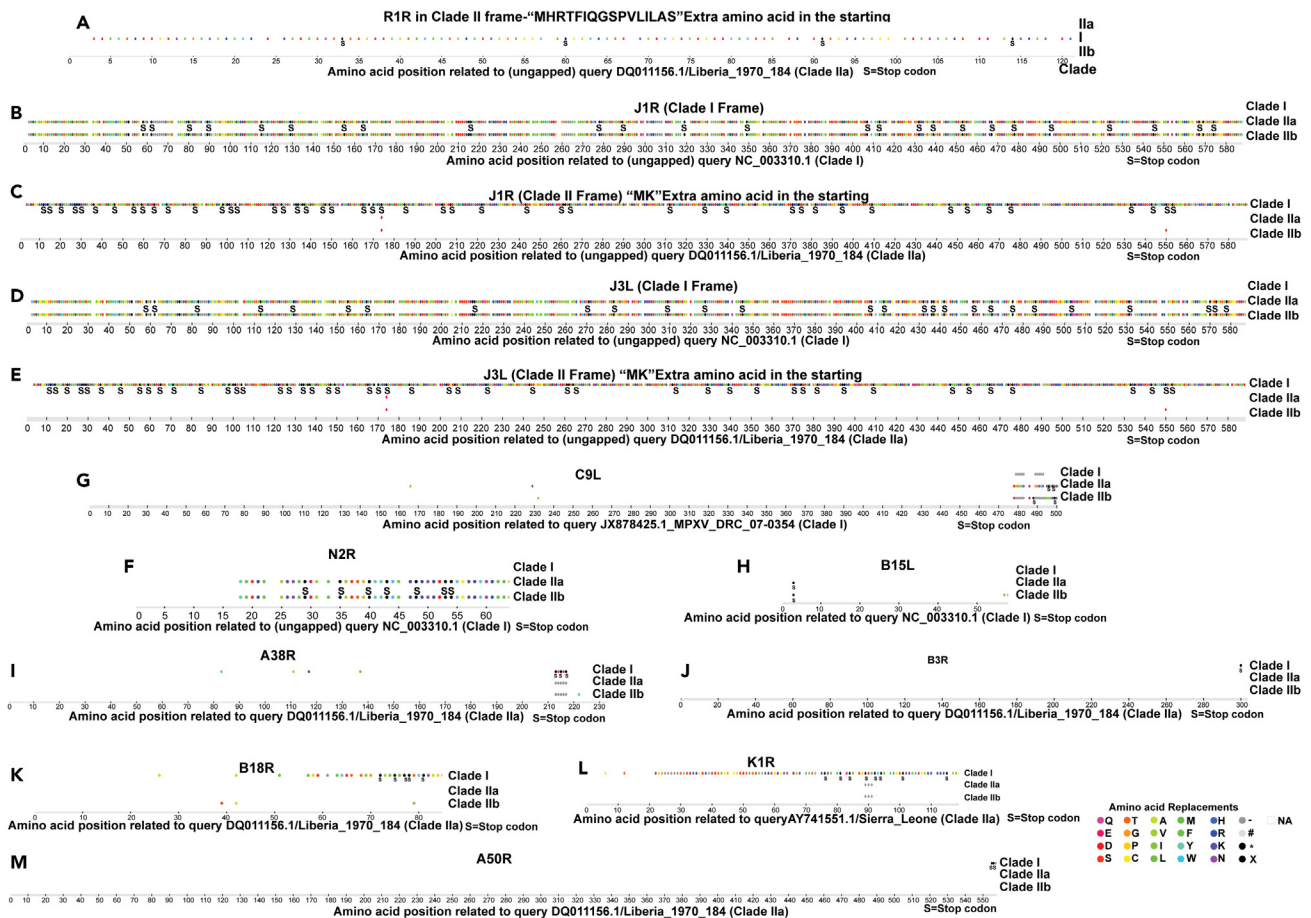
Figure 9. Frameshift mutations, and in-frame nonsense mutations found in the Clade IIa and Clade IIb viruses

(A) In the Clade II viruses frame, the R1R gene is truncated in Clade I viruses. Detailed in Figure S24; Data S7.

(B and C) The frameshift mutation mediated genetic diversity in the Clade II viruses J1R gene. If the J1R gene initiates translation at the frame of Clade I viruses, this gene is truncated in Clade II viruses (B), whereas in the Clade II frame, it is truncated in Clade I (C). Detailed in Figure S25; Data S7.

(D and E) The frameshift mutation mediated genetic diversity in the Clade II viruses J3L gene. If the J3L gene initiates translation at the frame of Clade I viruses, this gene is truncated in Clade II viruses (D), whereas in the Clade II frame, it is truncated in Clade I (E). Detailed in Figure S26; Data S7.

(F–H) An in-frame nonsense mutation is observed in the N2R (F), C9L (G), and B15L (H) genes of the Clade IIa and Clade IIb viruses. Detailed in Figure S27; Data S7.

(I and J) An in-frame nonsense mutation is observed in the A38R (I) and B3R (J) genes of the Clade I viruses. Detailed in Figure S28; Data S7.

(K–M) An in-frame nonsense mutation is observed in the B18R (K), K1R (L), and A50R (M) genes of the Clade I viruses. Details of individual strains are provided in Figure S29; Data S8.

variations in genes such as B10R, O1L, A10L, B5R, IIa-MV163 and IIa-MV172. Studies have observed that B10R gene truncation in Clade IIa viruses compared to Clade I viruses, suggesting to change in virulence Mpoxv.[13,19] Similarly, the genomic deletions were documented in vaccinia viruses and correlated with the host range and disease outcome.[39] Previously, functionally active protein truncations were also reported in poxviruses[39,40] and modified vaccinia virus Ankara (MVA) when passaged in chick embryo fibroblasts.[41–44] Therefore, we speculate that these genomic deletions and genetic variations found in Clade I and Clade IIb viruses might be involved in determining the host range and transmission of the virus.

Our study also observed that the Clade IIa and Clade IIb viruses have unique genetic variations such as genomic region variations, frameshift mutations, and copy number variations. The Clade II Mpoxv has unique genetic variations such as two genomic region variations and their associated genes IIa-V163 (kelch-like) and IIb-MV23, and the absence/loss of two genomic region variations and their associated genes D14L (C3b/C4b-binding) and kelch-like proteins protein's fragments of D15L, D16L, and D17L. Furthermore, Clade IIa and Clade IIb viruses have unique genetic variations in genes such as NMDA receptor-like protein (R1R), ankyrin-like (J1R/J3L), Kelch-like protein (C9L), A38R (IEV but not CEV envelope protein), B3R (serine; Ser/Thr Kinase), B18R (kelch-like protein), A28L (the major component of IMV surface tubules), B17R (ankyrin-like), H5R (virosome-associated), and MV9 (virosome component proteins), B15L (IL-1 beta inhibitor), D11L (TLR/IFN-beta signaling inhibitor), J1L/J3R (chemokine binding), IIa-MV162 (TLR/IL-1 signaling inhibitor), IIb-MV30 (116AA) (TNF-alpha-receptor-like
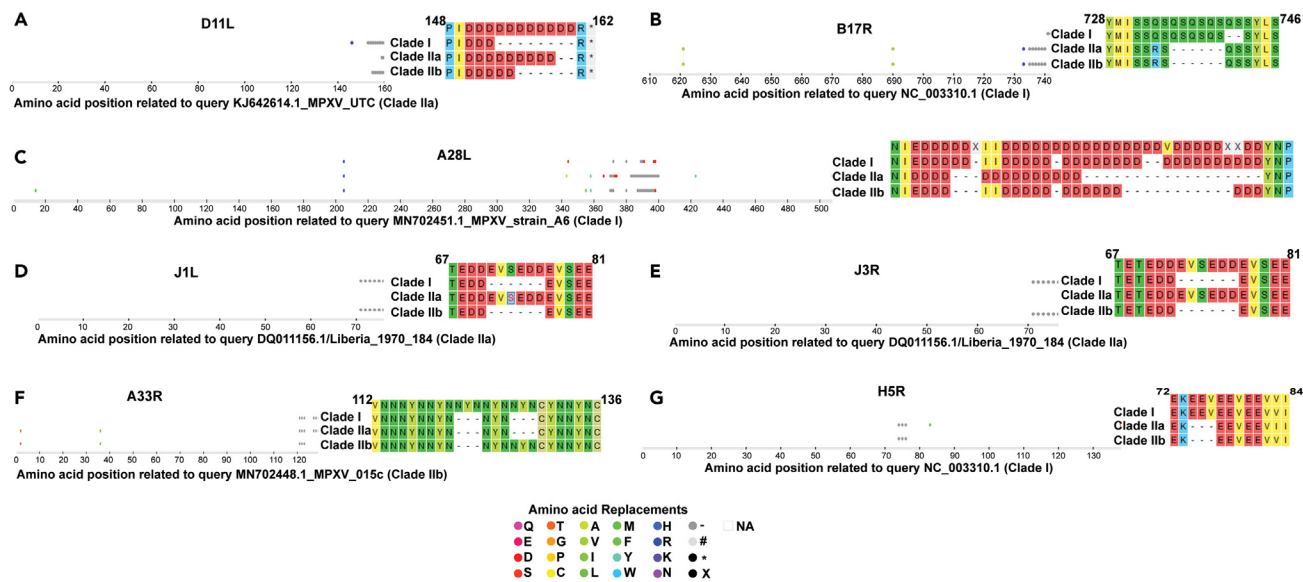
Figure 10. Amino acid tandem repeats mutations found in the Clade IIa and Clade IIb viruses

(A–G) The amino acid tandem repeats copy number differences were observed between Clade I and Clade II viruses in D11L (A), B17R (B), A28L (C), J1L/J3R (D and E), A33R (F), and H5R (G). Detailed in Figures S30–S32; Data S8.

protein), Gp16-like (IL-1 receptor antagonist) and IIb-MV4 (IL-1 receptor antagonist) genes. Previous study has also observed the large genomic region deletion in the Mpoxv and experimentally proven that the complement control protein (CCP) plays an essential role in the host immune response mediated virus pathogenesis of Mpoxv.[45] Previous works also found that the loss of the MPV-Z-N2R/OMCP gene in some of the Mpoxv is associated with fitness in humans.[20] The present study also identified genomic region variations and their associated, and also multiple gene genetic variations between the Clade II and Clade I viruses. The functional importance of these genetic variations is fully understood, and we speculate that the genetic variation might contribute to disease outcomes.

The amino acid tandem repeats are abundant in proteins of eukaryotes, these repeats are acquired through replication slippage and are helpful in mutation, genetic variability, and adaptive processes/evolutions, and it has been reported that these repeats are abundant in alternatively spliced genes.[46,47] Also, reports suggest that amino acid tandem repeats are also involved in protein function and folding.[48,49] It is noteworthy that amino acid tandem repeats are not generally found in viruses, but poxviruses contain amino acid repeats such as ankyrin and kelch repeats.[50,51] Previous studies also identified the tandem repeats A26L, B16R, and B19R proteins of Mpoxv and suggested the functional role of the repeats.[52] In the present study, we observed that the amino acid tandem repeats copy number difference between the clades of Mpoxv in the B16R, A10L, B5R, D11L, B17R, A28L, J1L, J3R, A33R, and H5L proteins. The functional importance of the amino acid tandem repeats in these proteins is not fully understood. We believe that these amino acid tandem repeats may have a role in the disease outcome, which needs to be explored.

Previous studies have identified the APOBEC3 mutations within the clades of Mpoxv,[4,21] whereas in the present study we observed the APOBEC3 mutations between the clades of Mpoxv. Our data suggest that the enrichment of APOBEC3 and other mutations in Clade I and Clade IIb viruses is capable of human-to-human transmission when compared to Clade IIa viruses with limited human-to-human transmission. These findings suggest that Clade I and Clade IIb viruses might have undergone adaptive microevolution for adaptation in humans during the process of intermittent or continuous cryptic spread of infection in humans. Similarly, our findings suggest that APOBEC3 GA-to-AA mutations are more enriched in Clade IIb viruses than Clade I viruses, suggesting that Clade IIb viruses might have undergone extensive adaptive microevolution than Clade I viruses for effective human adaptation. Further, the presence of negative selection in the Mpoxv indicates the process of continuous evolution through purifying selection.

The combination of loss/gain/inactivation of many genes in addition to the loss/gain/inactivation of individual genes, can often determine the host range, evasion of the host immunity, pathogenicity, and virulence of poxviruses.[15,18] Therefore, it is likely to be realized that the combined genetic variations in the various genes mentioned above have contributed to shaping the evolution of the virus to emerge into a pandemic outbreak-causing virus. Our findings indicate that the pandemic outbreak might have evolved as a combination of genetic variations in several genes, and the genetic variations might have been accumulated in a continuous evolutionary process. We believe that our findings will help in understanding the evolutionary process of Mpox viruses in the ongoing outbreak and are expected to help in tracking the spread and transmission of Mpoxv, designing proper diagnostics, developing anti-virals, and also to predict the future direction of the Mpoxv evolution.

Limitations of the study

The present study is a molecular evolution and genetic diversity study based on Mpoxv complete genome sequences available in the NCBI database. This analysis was performed using almost all high-quality complete genome sequences of Clade I, Clade IIa, Clade IIb, Clade IIb-A,
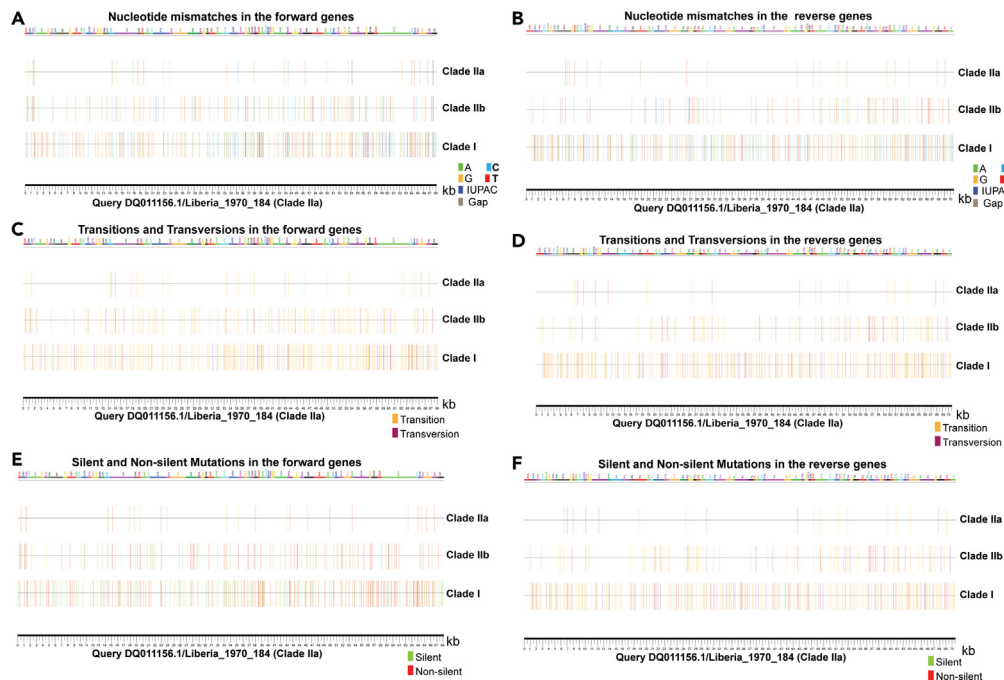
Figure 11. Genetic variations are found in the genes transcribed in forward and reverse directions

(A and B) The nucleotide mismatch analysis displays the presence of abundant mismatches in the Clade I and Clade IIb viruses genes that transcribed in the forward (A) and reverses (B) directions compared to Clade IIa viruses. The Clade IIa viruses were used as a reference. Details of individual strains are provided in Figure S33.

(C and D) The nucleotide transitions and transversions analysis exhibited abundant transitions in the Clade I and Clade IIb virus genes transcribed in the forward (C) and reverse (D) directions compared to Clade IIa viruses. The Mpoxv Clade IIa viruses were used as a reference. Detailed in Figure S34.

(E and F) The silent and non-silent mutations analysis showed the enhancement of silent mutations in the Clade I and Clade IIb virus genes that were transcribed in the forward (E) and reversed (F) directions compared to Clade IIa viruses. The Mpoxv Clade IIa viruses were used as a reference. Detailed in Figure S35.

and Clade IIb-A.1, Clade IIb-A.1.1, Clade IIb-A.2 lineages, and representative complete genome sequences of Clade IIb-B.1 lineage (n = 963) with identical ORFs.[13] Also, few genetically diversified viral genomic sequences have recently been identified in the Clade IIb-B.1 lineage, which were not analyzed in the present study. Furthermore, in this present study frameshift mutations or frameshift mutation results in an alternate start site of translations have been detected, whereas the frame in which proteins are translated can only be determined through experimental studies. Therefore, future experimental studies are required to validate the findings of this study.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Sequence alignment
  - The LAST plot hits in the MAFFT alignment
  - Phylogenetic analysis
  - Net between group mean distance (NBGMD) analysis
  - SimPlot analysis
  - Determining consensus ORFs for the different clades of mpoxv
  - Measurement of nucleotide and amino acid mismatch, transition and transversion, and silent and non-silent mutation analysis for the mpoxv
  - Measuring APOBEC3 motif mutations and dN/dS ratio in the mpoxv
- QUANTIFICATION AND STATISTICAL ANALYSIS
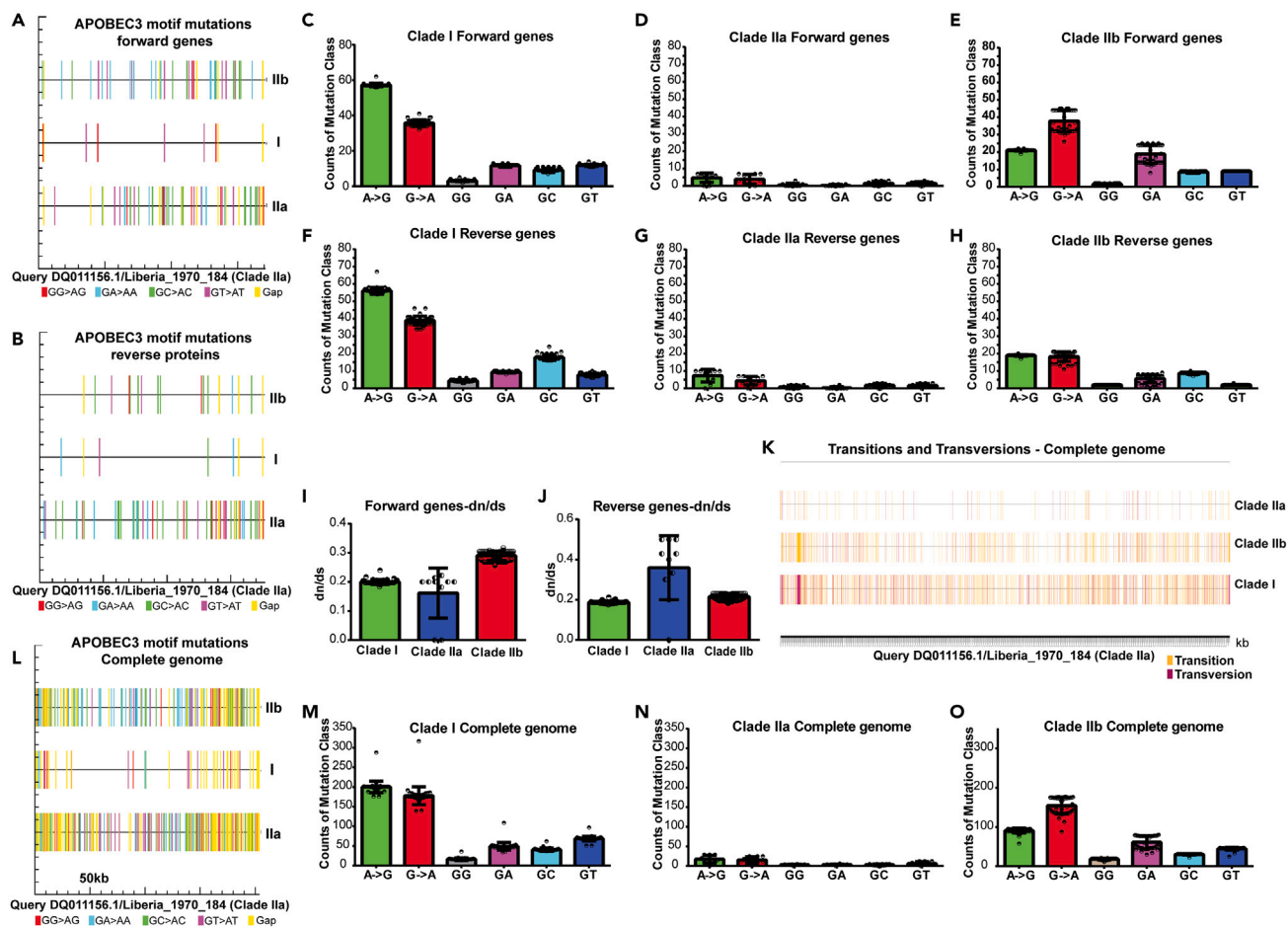
Figure 12. APOBEC3 mutations among the Mpoxv clades

(A and B) The APOBEC3 mutations analysis exposed the enrichment of APOBEC3 mutations in the Clade I and Clade IIb virus genes that were transcribed in the forward (A) and reverse (B) directions compared to Clade IIa viruses. The Mpoxv Clade IIa viruses were used as a reference. Detailed in Figures S36 and S37; Data S9 and S10.

(C–E) The graphical representation displaying different types of APOBEC3 and non-APOBEC3 motif mutations in the Clade I (C), Clade IIa (D), and Clade IIb (E) virus genes that transcribed in the forward direction. The Mpoxv Clade IIa viruses were used as a reference. Detailed in Data S9.

(F–H) The graphical representation showing different types of APOBEC3 and non-APOBEC3 motif mutations in the Clade I (F), Clade IIa (G), and Clade IIb (H) virus genes that transcribed in the reverse direction. The Mpoxv Clade IIa viruses were used as a reference. Detailed in Data S10.

(I and J) The graphical representation of the dN/dS ratio among different Mpoxv clades virus genes that transcribed in the forward (I) and reversed (J) directions. The Mpoxv Clade IIa viruses were used as a reference. Detailed in Data S11 and S12.

(K) The nucleotide transitions and transversions analysis exhibited abundant mutations in the Clade I and Clade IIb viruses at the whole genome levels compared to Clade IIa viruses. The Mpoxv Clade IIa viruses were used as a reference. Detailed in Figure S38.

(L) The APOBEC3 mutations analysis exposed the enrichment of APOBEC3 mutations in the Clade I and Clade IIb viruses at the whole genome levels compared to Clade IIa viruses. The Mpoxv Clade IIa viruses were used as a reference. Detailed in Figure S39; Data S13.

(M–O) The graphical representation showing different types of APOBEC3 and non-APOBEC3 motif mutations in the Clade I (M), Clade IIa (N), and Clade IIb (O) virus genes at the whole genome levels compared to Clade IIa viruses. The Mpoxv Clade IIa viruses were used as a reference. Detailed in Data S13.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108601.

AUTHOR'S CONTRIBUTIONS

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Brown, K., and Leggat, P.A. (2016). Human Monkeypox: Current State of Knowledge and Implications for the Future. Trop. Med. Infect. Dis. 1, 8.
2. Reynolds, M.G., Davidson, W.B., Curns, A.T., Conover, C.S., Huhn, G., Davis, J.P., Wegner, M., Croft, D.R., Newman, A., Obiesie, N.N., et al. (2007). Spectrum of infection and risk factors for human monkeypox, United States, 2003. Emerg. Infect. Dis. 13, 1332–1339.
3. Hutin, Y.J., Williams, R.J., Malfait, P., Pebody, R., Loparev, V.N., Ropp, S.L., Rodriguez, M., Knight, J.C., Tshioko, F.K., Khan, A.S., et al. (2001). Outbreak of human monkeypox, Democratic Republic of Congo, 1996 to 1997. Emerg. Infect. Dis. 7, 434–438.
4. Isidro, J., Borges, V., Pinto, M., Sobral, D., Santos, J.D., Nunes, A., Mixão, V., Ferreira, R., Santos, D., Duarte, S., et al. (2022). Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. Nat. Med. 28, 1569–1572.
5. Happi, C., Adetifa, I., Mbala, P., Njouom, R., Nakoune, E., Happi, A., Ndodo, N., Ayansola, O., Mboowa, G., Bedford, T., et al. (2022). Urgent need for a non-discriminatory and non-stigmatizing nomenclature for monkeypox virus. PLoS Biol. 20, e3001769.
6. Rezza, G. (2019). Emergence of human monkeypox in west Africa. Lancet Infect. Dis. 19, 797–799.
7. Likos, A.M., Sammons, S.A., Olson, V.A., Frace, A.M., Li, Y., Olsen-Rasmussen, M., Davidson, W., Galloway, R., Khristova, M.L., Reynolds, M.G., et al. (2005). A tale of two clades: monkeypox viruses. J. Gen. Virol. 86, 2661–2672.
8. Yinka-Ogunleye, A., Aruna, O., Dalhat, M., Ogoina, D., McCollum, A., Disu, Y., Mamadu, I., Akinpelu, A., Ahmad, A., Burga, J., et al. (2019). Outbreak of human monkeypox in Nigeria in 2017-18: a clinical and epidemiological report. Lancet Infect. Dis. 19, 872–879.
9. Rimoin, A.W., Mulembakani, P.M., Johnston, S.C., Lloyd Smith, J.O., Kisalu, N.K., Kinkela, T.L., Blumberg, S., Thomassen, H.A., Pike, B.L., Fair, J.N., et al. (2010). Major increase in human monkeypox incidence 30 years after smallpox vaccination campaigns cease in the Democratic Republic of Congo. Proc. Natl. Acad. Sci. USA 107, 16262–16267.
10. Formenty, P., Muntasir, M.O., Damon, I., Chowdhary, V., Opoka, M.L., Monimart, C., Mutasim, E.M., Manuguerra, J.C., Davidson, W.B., Karem, K.L., et al. (2010). Human monkeypox outbreak caused by novel virus belonging to Congo Basin clade, Sudan, 2005. Emerg. Infect. Dis. 16, 1539–1545.
11. Mauldin, M.R., McCollum, A.M., Nakazawa, Y.J., Mandra, A., Whitehouse, E.R., Davidson, W., Zhao, H., Gao, J., Li, Y., Doty, J., et al. (2022). Exportation of Monkeypox Virus From the African Continent. J. Infect. Dis. 225, 1367–1376.
12. Ng, O.T., Lee, V., Marimuthu, K., Vasoo, S., Chan, G., Lin, R.T.P., and Leo, Y.S. (2019). A case of imported Monkeypox in Singapore. Lancet Infect. Dis. 19, 1166.
13. Desingu, P.A., Rubeni, T.P., and Sundaresan, N.R. (2022). Evolution of monkeypox virus from 2017 to 2022: In the light of point mutations. Front. Microbiol. 13, 1037598.
14. Happi, C., Adetifa, I., Mbala, P., Njouom, R., Nakoune, E., Happi, A., Ndodo, N., Ayansola, O., Mboowa, G., Bedford, T., et al. (2022). Urgent need for a non-discriminatory and non-stigmatizing nomenclature for monkeypox virus. https://virological.org/t/urgent-need-for-a-non-discriminatory-and-non-stigmatizing-nomenclature-for-monkeypox-virus/853.
15. Senkevich, T.G., Yutin, N., Wolf, Y.I., Koonin, E.V., and Moss, B. (2021). Ancient Gene Capture and Recent Gene Loss Shape the Evolution of Orthopoxvirus-Host Interaction Genes. mBio 12, e0149521.
16. Brennan, G., Stoian, A.M.M., Yu, H., Rahman, M.J., Banerjee, S., Stroup, J.N., Park, C., Tazi, L., and Rothenburg, S. (2022). Molecular mechanisms of poxvirus evolution. mBio 14, e0152622.
17. Shchelkunov, S.N. (2009). How long ago did smallpox virus emerge? Arch. Virol. 154, 1865–1871.
18. Senkevich, T.G., Zhivkoplias, E.K., Weisberg, A.S., and Moss, B. (2020). Inactivation of Genes by Frameshift Mutations Provides Rapid Adaptation of an Attenuated Vaccinia Virus. J. Virol. 94, e01053-20.
19. Chen, N., Li, G., Liszewski, M.K., Atkinson, J.P., Jahrling, P.B., Feng, Z., Schriewer, J., Buck, C., Wang, C., Lefkowitz, E.J., et al. (2005). Virulence differences between monkeypox virus isolates from West Africa and the Congo basin. Virology 340, 46–63.
20. Kugelman, J.R., Johnston, S.C., Mulembakani, P.M., Kisalu, N., Lee, M.S., Koroleva, G., McCarthy, S.E., Gestole, M.C., Wolfe, N.D., Fair, J.N., et al. (2014). Genomic variability of monkeypox virus among humans, Democratic Republic of the Congo. Emerg. Infect. Dis. 20, 232–239.
21. Gigante, C.M., Korber, B., Seabolt, M.H., Wilkins, K., Davidson, W., Rao, A.K., Zhao, H., Smith, T.G., Hughes, C.M., Minhaj, F., et al. (2022). Multiple lineages of monkeypox virus detected in the United States, 2021-2022. Science 378, 560–565.
22. Wang, L., Shang, J., Weng, S., Aliyari, S.R., Ji, C., Cheng, G., and Wu, A. (2022). Genomic annotation and molecular evolution of monkeypox virus outbreak in 2022. J. Med. Virol. 95, e28036.
23. Jones, T.C., Schneider, J., Mühlemann, B., Veith, T., Beheim-Schwarzbach, J., Tesch, J., Schmidt, M.L., Walper, F., Bleicker, T., Isner, C., et al. (2022). Genetic variability, including gene duplication and deletion, in early sequences from the 2022 European monkeypox outbreak. Preprint at bioRxiv. https://doi.org/10.1101/2022.07.23.501239.
24. Gigante, C.M., Plumb, M., Ruprecht, A., Zhao, H., Wicker, V., Wilkins, K., Matheny, A., Khan, T., Davidson, W., Sheth, M., et al. (2022). Genomic deletions and rearrangements in monkeypox virus from the 2022 outbreak, USA. Preprint at bioRxiv. https://doi.org/10.1101/2022.09.16.508251.

25. Brinkmann, A., Kohl, C., Pape, K., Bourquain, D., Thürmer, A., Michel, J., Schaade, L., and Nitsche, A. (2023). Extensive ITR expansion of the 2022 Mpox virus genome through gene duplication and gene loss. Virus Gene. *59*, 532–540.

26. Sereewit, J., Lieberman, N.A.P., Xie, H., Bakhash, S.A.K.M., Nunley, B.E., Chung, B., Mills, M.G., Roychoudhury, P., and Greninger, A.L. (2022). ORF-Interrupting Mutations in Monkeypox Virus Genomes from Washington and Ohio, 2022. Viruses *14*, 2393.

27. Garrigues, J.M., Hemarajata, P., Lucero, B., Alarcon, J., Ransohoff, H., Marutani, A.N., Kim, M., Marlowe, E.M., Realegeno, S.E., Kagan, R.M., et al. (2022). Identification of Human Monkeypox Virus Genome Deletions That Impact Diagnostic Assays. J. Clin. Microbiol. *60*, e0165522.

28. Keele, B.F., Giorgi, E.E., Salazar-Gonzalez, J.F., Decker, J.M., Pham, K.T., Salazar, M.G., Sun, C., Grayson, T., Wang, S., Li, H., et al. (2008). Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc. Natl. Acad. Sci. USA *105*, 7552–7557.

29. Hakim, M.S., and Widyaningsih, S.A. (2023). The recent re-emergence of human monkeypox: Would it become endemic beyond Africa? J. Infect. Public Health *16*, 332–340.

30. Spriggs, M.K., Hruby, D.E., Maliszewski, C.R., Pickup, D.J., Sims, J.E., Buller, R.M., and VanSlyke, J. (1992). Vaccinia and cowpox viruses encode a novel secreted interleukin-1-binding protein. Cell *71*, 145–152.

31. Alcamí, A., and Smith, G.L. (1992). A soluble receptor for interleukin-1 beta encoded by vaccinia virus: a novel mechanism of virus modulation of the host response to infection. Cell *71*, 153–167.

32. Zhang, R.Y., Pallett, M.A., French, J., Ren, H., and Smith, G.L. (2022). Vaccinia virus BTB-Kelch proteins C2 and F3 inhibit NF-kappaB activation. J. Gen. Virol. *103*, 001786.

33. Wang, Q., Burles, K., Couturier, B., Randall, C.M.H., Shisler, J., and Barry, M. (2014). Ectromelia virus encodes a BTB/kelch protein, EVM150, that inhibits NF-kappaB signaling. J. Virol. *88*, 4853–4865.

34. Sprygin, A., Babin, Y., Pestova, Y., Kononova, S., Wallace, D.B., Van Schalkwyk, A., Byadovskaya, O., Diev, V., Lozovoy, D., and Kononov, A. (2018). Analysis and insights into recombination signals in lumpy skin disease virus recovered in the field. PLoS One *13*, e0207480.

35. Fixsen, S.M., Cone, K.R., Goldstein, S.A., Sasani, T.A., Quinlan, A.R., Rothenburg, S., and Elde, N.C. (2022). Poxviruses capture host genes by LINE-1 retrotransposition. Elife *11*, e63332.

36. Wang, L., Shang, J., Weng, S., Aliyari, S.R., Ji, C., Cheng, G., and Wu, A. (2023). Genomic annotation and molecular evolution of monkeypox virus outbreak in 2022. J. Med. Virol. *95*, e28036.

37. Elde, N.C., Child, S.J., Eickbush, M.T., Kitzman, J.O., Rogers, K.S., Shendure, J., Geballe, A.P., and Malik, H.S. (2012). Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. Cell *150*, 831–841.

38. Brennan, G., Kitzman, J.O., Rothenburg, S., Shendure, J., and Geballe, A.P. (2014). Adaptive gene amplification as an intermediate step in the expansion of virus host range. PLoS Pathog. *10*, e1004002.

39. Mühlemann, B., Vinner, L., Margaryan, A., Wilhelmson, H., de la Fuente Castro, C., Allentoft, M.E., de Barros Damgaard, P., Hansen, A.J., Holtsmark Nielsen, S., Strand, L.M., et al. (2020). Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. Science *369*, eaaw8977.

40. Hendrickson, R.C., Wang, C., Hatcher, E.L., and Lefkowitz, E.J. (2010). Orthopoxvirus genome evolution: the role of gene loss. Viruses *2*, 1933–1967.

41. Wyatt, L.S., Carroll, M.W., Czerny, C.P., Merchlinsky, M., Sisler, J.R., and Moss, B. (1998). Marker rescue of the host range restriction defects of modified vaccinia virus Ankara. Virology *251*, 334–342.

42. Liu, R., Mendez-Rios, J.D., Peng, C., Xiao, W., Weisberg, A.S., Wyatt, L.S., and Moss, B. (2019). SPI-1 is a missing host-range factor required for replication of the attenuated modified vaccinia Ankara (MVA) vaccine vector in human cells. PLoS Pathog. *15*, e1007710.

43. Peng, C., and Moss, B. (2020). Repair of a previously uncharacterized second host-range gene contributes to full replication of modified vaccinia virus Ankara (MVA) in human cells. Proc. Natl. Acad. Sci. USA *117*, 3759–3767.

44. Sutter, G. (2020). A vital gene for modified vaccinia virus Ankara replication in human cells. Proc. Natl. Acad. Sci. USA *117*, 6289–6291.

45. Hudson, P.N., Self, J., Weiss, S., Braden, Z., Xiao, Y., Girgis, N.M., Emerson, G., Hughes, C., Sammons, S.A., Isaacs, S.N., et al. (2012). Elucidating the role of the complement control protein in monkeypox pathogenicity. PLoS One *7*, e35086.

46. Mularoni, L., Ledda, A., Toll-Riera, M., and Albà, M.M. (2010). Natural selection drives the accumulation of amino acid tandem repeats in human proteins. Genome Res. *20*, 745–754.

47. Haerty, W., and Golding, G.B. (2010). Genome-wide evidence for selection acting on single amino acid repeats. Genome Res. *20*, 755–760.

48. Luo, H., and Nijveen, H. (2014). Understanding and identifying amino acid repeats. Brief. Bioinform. *15*, 582–591.

49. Perez-Riba, A., Synakewicz, M., and Itzhaki, L.S. (2018). Folding cooperativity and allosteric function in the tandem-repeat protein class. Philos. Trans. R. Soc. Lond. B Biol. Sci. *373*, 20170188.

50. Balinsky, C.A., Delhon, G., Afonso, C.L., Risatti, G.R., Borca, M.V., French, R.A., Tulman, E.R., Geary, S.J., and Rock, D.L. (2007). Sheeppox virus kelch-like gene SPPV-019 affects virus virulence. J. Virol. *81*, 11392–11401.

51. Sonnberg, S., Seet, B.T., Pawson, T., Fleming, S.B., and Mercer, A.A. (2008). Poxvirus ankyrin repeat proteins are a unique class of F-box proteins that associate with cellular SCF1 ubiquitin ligase complexes. Proc. Natl. Acad. Sci. USA *105*, 10955–10960.

52. Monzón, S., Sarai, V., Negredo, A., Patiño-Galindo, J.A., Vidal-Freire, S., Zaballos, A., Orviz, E., Ayerdi, O., Muñoz-García, A., Delgado-Iribarren, A., et al. (2022). Changes in a new type of genomic accordion may open the pallets to increased monkeypox transmissibility. Preprint at bioRxiv. https://doi.org/10.1101/2022.09.30.510261.

53. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol E *30*, 772–780.

54. Katoh, K., Rozewicki, J., and Yamada, K.D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. *20*, 1160–1166.

55. Kuraku, S., Zmasek, C.M., Nishimura, O., and Katoh, K. (2013). aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res. *41*, W22–W28.

56. Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol *4*, vey016.

57. Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol E *33*, 1870–1874.

58. Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., and Ray, S.C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. *73*, 152–160.

59. Rose, P.P., and Korber, B.T. (2000). Detecting hypermutations in viral sequences with an emphasis on G–> A hypermutation. Bioinformatics *16*, 400–401.

60. Korber-Irrgang, B. (2000). HIV Signature and Sequence Variation Analysis. Computational Analysis of HIV Molecular Sequences. In Chapter 4, G. Allen Rodrigo and G.H. Learn, eds. (Kluwer Academic Publishers), pp. 55–72.

61. Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., and Gascuel, O. (2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. Nucleic Acids Res. *47*, W260–W265.

62. Mareuil, F., Doppelt-Azeroual, O., and Ménager, H. (2017). A public Galaxy platform at Pasteur used as an execution engine for web services. F1000Res. *10*.

63. Katoh, K., and Frith, M.C. (2012). Adding unaligned sequences into an existing alignment using MAFFT and LAST. Bioinformatics *28*, 3144–3146.

64. Desingu, P.A., Nagarajan, K., and Dhama, K. (2022). SARS-CoV-2 gained a novel spike protein S1-N-Terminal Domain (S1-NTD). Environ. Res. *211*, 113047.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
MAFFT 7.407_1	Katoh, K. and Standley, D. M.[53]	https://ngphylogeny.fr/tools/tool/268/form
MAFFT version 7	https://mafft.cbrc.jp/alignment/server/ [54,55]	https://mafft.cbrc.jp/alignment/server/
BEAST.v1.10.4	https://beast.community/2018-11-14.10.4_released.html [56]	https://beast.community/2018-11-14.10.4_released.html
FigTree v1.4.4	https://beast.community/figtree [56]	https://beast.community/figtree
MEGA7	Kumar et al.[57]	https://www.megasoftware.net/
SimPlot 3.5.1	Lole et al.[58]	https://sray.med.som.jhmi.edu/SCRoftware/SimPlot/#downloads
NCBI ORFfinder	https://www.ncbi.nlm.nih.gov/orffinder/	https://www.ncbi.nlm.nih.gov/orffinder/
Highlighter tool	Highlighter tool[28]	https://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter_top.html
Variant Visualizer tools	https://www.hiv.lanl.gov/content/index	https://www.hiv.lanl.gov/content/sequence/VVISUALIZER/vvisualizer.html
Hypermut 2.0	Hypermut 2.0[59]	https://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html
SNAP v2.1.1	SNAP v2.1.1[60]	https://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Perumal Arumugam Desingu (padesingu@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- This paper does not generate original data
- This paper does not report the original code
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request

METHOD DETAILS

Sequence alignment

In the present study, the monkeypox virus's complete genome, different genomic regions, and gene nucleotide sequences were retrieved from the NCBI public database and aligned in the MAFFT 7.407_1 multiple alignment program with the parameters of Gap extend penalty of 0.123 and Gap opening penalty was fixed as 1.53.[53,61,62] This analysis was performed using almost all high-quality complete genome sequences of Clade-I (n = 46), Clade-IIa (n = 11), Clade-IIb (n = 45), Clade-IIb-A (n = 11), and Clade-IIb-A.1(n = 9), Clade-IIb-A.1.1(n = 1), Clade-IIb-A.2 lineages (n = 3), and representative complete genome sequences of Clade-IIb-B.1 lineage (n = 21), also, our previous study reported that the 963 high-quality complete genome sequences in the Clade-IIb-B.1 lineage (n = 963) with identical ORFs.[13]

The LAST plot hits in the MAFFT alignment

The LAST Plot hits and alignment for the complete genome nucleotide sequences of monkeypox viruses have been performed using MAFFT version 7 online server with the Score = 39 (E = $8.4e^{-11}$) (https://mafft.cbrc.jp/alignment/server/).[54,63] This analysis was carried out with almost all high-quality complete genome sequences of Clade-I (n = 46), Clade-IIa (n = 11), Clade-IIb (n = 45), Clade-IIb-A (n = 11), and Clade-IIb-A.1(n = 9), Clade-IIb-A.1.1(n = 1), Clade-IIb-A.2 lineages (n = 3), and representative complete genome sequences of Clade-IIb-B.1 lineage (n = 21), details of sequences provided in Figure 1A.

Phylogenetic analysis

In the present study, the complete genome nucleotide sequences of monkeypox viruses were first aligned using MAFFT 7.407_1, MAFFT aligned sequences were used to generate the BEAST XML file in BEAUti v1.10.4 with the MCMC method. Next, the generated BEAST XML file was used to run the BEAST v1.10.4, and the tree file was generated using the HKY substitution model, strict clock, coalescent constant size, random starting tree, length of chain 10000000, echo state every 1000, and log parameter every 1000. Then, the generated tree file was used to run the TreeAnnotator v1.10.4 with the maximum clade credibility (MCC) tree. Finally, the was visualized in the FigTree v1.4.4 with the tMRCA (HPD95%).

Net between group mean distance (NBGMD) analysis

In this study, the complete genome nucleotide sequences of the Mpox virus were first aligned in MAFFT 7.407_1, and then the NBGMD analysis was performed in MEGA7.[57] The NBGMD analysis was carried out using the Kimura 2-parameter model with the Transitions plus Transversions substitution, gamma distribution (shape parameter = 5), pairwise deletion of gaps/missing data were deleted pairwise, and standard errors were estimated using the bootstrap test of 1000 replicates. The measure standard error was presented above the diagonal in the NBGMD analysis result tables. This analysis was carried out with almost all high-quality complete genome sequences of Clade-I (n = 46), Clade-IIa (n = 11), Clade-IIb (n = 45), Clade-IIb-A (n = 11), and Clade-IIb-A.1(n = 9), Clade-IIb-A.1.1(n = 1), Clade-IIb-A.2 lineages (n = 3), and representative complete genome sequences of Clade-IIb-B.1 lineage (n = 21), details of sequences provided in Figure 1A.

SimPlot analysis

SimPlot 3.5.1[64] was utilized to measure the percent identity/similarities among the query of Mpox viruses and reference sequences of Mpoxv. In this study, the complete genome nucleotide sequences of Mpox viruses were first aligned in MAFFT 7.407_1 before they were exported to SimPlot 3.5.1 for the subsequent analysis. In the present study, SimPlot analysis of Mpox viruses was carried out with the Kimura (2-parameter) method by 500 base pairs of the window at a 50 base-pair step, and NCBI reference sequence NC_063383.1 was used as the query sequence.

Determining consensus ORFs for the different clades of mpoxv

First, the ORF differences in predicting the ORFs of individual strains in Clade-I, Clade-IIa, and Clade-IIb using different methods by different research groups at different times were identified using the information derived from the Virus Pathogen Resource (ViPR) online resource tool. Following this, we determined whether the different ORFs in individual strains within each clade were present in other viruses in that particular clade by retrieving sequences from other viruses aligned with the nucleotide sequences associated with these ORFs. For this, first, the complete genome nucleotide sequences of Mpox viruses were aligned in MAFFT 7.407_1[53,61,62]; from these aligned sequences, the aligned nucleotide sequences related to each ORF were retrieved, and the ORFs were determined within and between the Mpoxv-Clades using MEGA7[57] and NCBI ORFfinder tool with different ORF lengths of nucleotides (https://www.ncbi.nlm.nih.gov/orffinder/). Further, the details of ORF and their VACV-WR, VACV-Cop, OPG, and MPXVgp naming as well as functional annotations are provided in Table S1.

Measurement of nucleotide and amino acid mismatch, transition and transversion, and silent and non-silent mutation analysis for the mpoxv

The nucleotide and amino acid mismatch, Transition and transversion, and Silent and non-silent mutations among the Mpox viruses complete genome sequences, ITR regions, different clade-specific regions, and genes that are transcribed in the forward and reverse directions using the Highlighter tool[28] with or without similarity sorting of the sequences and treating the gaps as a character (https://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter_top.html). Further, the nucleotide and amino acid mismatch were also visualized using the Variant Visualizer tools (https://www.hiv.lanl.gov/content/index) https://www.hiv.lanl.gov/content/sequence/VVISUALIZER/vvisualizer.html.

Measuring APOBEC3 motif mutations and dN/dS ratio in the mpoxv

The APOBEC3 motif mutations in the Mpox virus's complete genome sequences and genes that are transcribed in the forward and reverse directions were measured by utilizing the Hypermut 2.0 tool with customized options[59] (https://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html). Fisher Exact p-value was used as a statistical analysis. The reference sequences used in the analysis were displayed in the respective figures. Clade IIa virus sequence DQ011156.1/Liberia_1970_184 was used as the query sequence. Next, the dN/dS ratio in the Mpox virus's genes that are transcribed in the forward and reverse directions was calculated using the SNAP v2.1.1[60] (https://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html). Clade IIa virus sequence DQ011156.1/Liberia_1970_184 was used as the query sequence. The reference sequences used in the analysis were displayed in the supplementary data related to the respective figures.

QUANTIFICATION AND STATISTICAL ANALYSIS

In phylogenetic analysis, we detected tMRCA (HPD95%) in BEAST v1.10.4 and visualized in FigTree v1.4.4. The genetic diversity and standard errors in NBGMD analysis were estimated using Kimura 2-parameter model in MEGA7. Further, detecting APOBEC3 motif mutations, Fisher Exact p-value was used as a statistical analysis.