Recombination shapes 2022 monkeypox outbreak

Ting-Yu Yeh[1,6], Gregory P. Contreras[1], Ying-Chieh Su[2,3], Shang-Lin Hsieh[4], Dylan A. Lewis[5]

[1]Auxergen Inc., Columbus Center, Baltimore, MD 21202, USA

[2]Department of Thoracic Surgery, Chi-Mei Medical Center, Tainan, Taiwan

[3]Deapartment of Biotechnology and Bioindustry Sciences, National Cheng Kung University, Tainan, Taiwan

[4]MacKay Memorial Hospital, Taipei City, Taiwan

[5]Monte Vista High School, Danville, California 94526, USA

[6]Corresponding author; yehty@auxergen.com

Key words: Monkeypox, tandem repeats, polymorphism, site-frequency spectrum, linkage disequilibrium, recombination

Word count: 1848 words (text), 2142 words (including Materials and Methods)

Summary

Background

On July 23, 2022, the WHO declared monkeypox outbreak a global health emergency. Here we analyze monkeypox viral sequences during 2022 pandemic to investigate whether the virus is adapting for better survival and transmission among the human population.

Methods

By studying tandem repeats (TRs) and linkage disequilibrium (LD), we analyzed 415 monkeypox viral sequences from January 1 to July 20, 2022 worldwide.

Finding

The 2022 monkeypox viral population has diverged into 4 lineages and 11 subgroups based on various TRs and their copy numbers. LD analysis also shows that virus has evolved into 3 new lineages. We identify 8 new recombinants (six from Slovenia, one from Australia, one from Italy) using TR analysis and 3 recombinants (two from Germany, one from Spain) using LD analysis.

Conclusion

Our results indicate that the monkeypox genome is evolving and expanding quickly during the 2022 pandemic. We conclude that in combination with genomic surveillance, TR analysis, as well as LD analysis, is a useful tool to monitor and track phylogenetic dynamics and recombination of monkeypox transmission.

Introduction

The 2022 monkeypox outbreak represents the first time this disease has spread widely beyond Central and West Africa. Initially identified in the United Kingdom in May 2022, monkeypox case numbers quickly increased in Europe, North and South America, Asia, Africa, and Oceania. On July 23, 2022, the WHO declared monkeypox outbreak a global health emergency. On August 12, 2022, a total of 35,032 cases were confirmed in nearly 80 countries (Monkeypox data explorer. *Our World in Data*).

Most viral sequences from the 2022 monkeypox outbreak belong to the B.1 clade (except two cases from the United States). In this study, we analyze monkeypox viral sequences during 2022 pandemic to investigate whether the virus is adapting for better survival and transmission among the human population.

In a rapidly evolving poxvirus, adaptation is simultaneously driven by two mechanisms: recombination (gene copy number variation, fast) and single nucleotide variants (SNVs, slow) at the same loci (Sasani, et al., 2018). Recombination generates new phenotypes with greatly altered disease potential which are better suited to viral survival. It has been shown that vaccinia viral DNA is swapped back and forth ~18 times per genome in a single round of infection (Qin and Evans, 2014) to make recombinant phenotypes. Recombination in poxvirus genomes has been commonly detected by selection or screening in laboratory animals or cell culture for more than 60 years (Fenner & Comben, 1958). Gershon et al. described recombination of four capripoxvirus isolates during natural virus transmission by analyzing physical maps of the viral genome (Gershon et al., 1989). However, very little is known about poxvirus recombination in nature due to its relatively large genome size and lack of genomic surveillance data, which is now becoming available.

Our efforts focused on discovering the variability of the monkeypox genome via recombination to determine the potential risk of new viral strains as they emerge. Tandem repeats (TRs) were first identified within the inverted terminal repeat of vaccinia virus DNA with a 70-base-pair sequence arranged in two blocks of 13 and 17 copies, respectively (Wittek and Moss, 1982). Other poxvirus TRs are small pieces of DNA sequences (3-25 nucleotides), and their sequences and copy numbers vary among different poxvirus family members or isolates (Kugelman et al., 2014). The insertions and deletions of TRs are common events among these poxviruses. (Coulson and Upton, 2011). Therefore, these TRs exhibit high rates of variation and they represent a target for poxvirus gene truncation and variation (Kugalman et al., 2014; Hatcher et al., 2015). Here we report the first evidence of natural recombination of monkeypox virus by analyzing TRs. We also used linkage disequilibrium, a well-known SNVs-based analysis, to detect new lineages and recombination in monkeypox genome of 2022 pandemics.

Results

Tandem repeats (TRs) as novel genetic markers of monkeypox virus to detect new lineages

To determine the genomic diversity in monkeypox genomes in the 2022 outbreak, we first searched TRs in 415 available sequences (B.1 clade) worldwide from January 1 to July 20, using the Tandem Repeat Finder algorithm (Benson, 1999). The advantage of this method is that it eliminates the bias caused by sequence alignment error, especially in low complexity sequences like TRs. Based on the criteria of the score (>100) and length (>7 base pair, bp), we identified 6 TRs with variations in their copy numbers (Figure 1A and 1B). TR A/E have identical 16 bp sequences of inverted repeats (5'-TAACTCTAACTTATGACT-3' and 5'-AGTCATAAGTTAGTTA-3') at both ends of monkeypox virus genome (Figure 1B). The viral populations were further categorized into six groups based on TR numbers (TRNs) of TRA/E (Figure 1C). 378 cases (90.6%) were associated with TRN=7.9 versus 14 cases of TRN=15.9

(3.4%), which were collected in the USA (ON959133, ON959134, ON959135, ON959136, ON954773, ON959131, ON959132), Belgium (ONON622712, ON622713, ON880419, ON880420, ON880421, ON880422) and the Czech Republic (ON983168). One case of TRN=5.6, TRN=3.6, and TRN=2.6 was found in United Kingdom (ON 619837, ON619835, ON022171). There were 21 cases with different TR numbers between TR A and E ("mismatch" in Figure 1C). This result shows that genome diversity can be grouped by TR polymorphism among monkeypox virus population in 2022 pandemic.

TR B, C, D, and F are direct repeats and located at either the intergenic regions (TR B, C, D) or 3' inverted terminal repeat (TR F) (Figure 1C, detail information of TR B, C, D in Supplementary Figure 2-4). Each TR C and F contain one and three copies of 9 bp sequence (5'-TATGATGGA-3'), respectively. While the majority of viral sequences contain TR F (TRN=3.5, 97.1%), 50.8% and 48.2% of total samples either have TR C with TRN=9.7 and TRN=7.7, respectively (Figure 1D). Based on TR C/F pattern, the viral populations can be classified into 4 linages (M, 210 cases; U, 193 cases; I , 7 cases; and one uncategorized, Figure 1D). Further, in combination with TRNs of TR C/F and TR A/E, we are able to categorize viral populations into 11 subgroups. TR D is a 9 base-pair sequence (5'-ATATCATT) with more various copy numbers, ranging from TRN=2 to 54.6 (Figure 1E). Interestingly, the sequences with high TRNs of TR D (TRN>30) also contain higher TRNs of TR A/E (TRN>15.9, 20 cases) in the group U, indicating that lineage U has more TR diversity than the others.

Taken together, our data demonstrate that TRs diverged frequently during natural transmission within the B.1 clade and that the virus is evolving as its population expands (Zeng's E = -1.65, Achaz's Y=-2.52, p< 0.001).

TR analysis identifies recombination of monkeypox genome in natural transmission

Poxvirus recombination between two co-infecting parental viruses generates genetic diversity (Ball, 1987; Evans et al., 1988; Sasani, et al, 2018). Monkeypox recombination in natural infection has not been reported to date. Using TR polymorphism, we identified 8 genomes with recombinant crossovers (Figure 1F). Case FVG-ITA-01 (ON755039) in Italy may be generated from parental sequences from the group I and M. Case VIDRL01 (ON631963) in Australia comes from parental sequence of the group M and U (Figure 1D), as well as six cases of Slovenia cases (ON838178, ON631241, ON609725, ON754985, ON754986, ON754987). This is the first report of recombination of monkeypox in natural transmission to our knowledge. Our results also suggests that six Slovenia cases may have evolved into a new lineage.

Linkage disequilibrium (LD) analysis shows new lineages and recombination

We then employed single nucleotide polymorphism (SNPs) analysis using the DNASP v6 algorithm to detect the occurrence of recombination (Rozas's Za=0.0005, $p<0.05$) (Rozas et al., 2011). We also used Haploview algorithm to visualize the patterns of linkage disequilibrium (LD) between variants with minor alleles in at least two monkeypox isolates and to detect the possible recombination (Figure 2A). In the absence of evolutionary forces or natural selection, *D*', the normalized coefficient of LD, converges to zero along the time axis at a rate depending on the magnitude of the recombination rate between the two loci. Since most SNPs were at very low frequencies, many SNP pairs had low values of squared coefficient of correlation (r^2) and the log of the odds (LOD) (Fig. 2D and 2E).

The LD analysis reveals five SNP pairs located at C22736T/G74357A, G34305A/G148421A, G34305A/G189246A, G148421A/G189246A, and G186153A/C188379T with the high log of odds (>10) and strong evidence of LD ($\chi 2$ test, $P<0.0001$), in which the upper 95% confidence bound of *D*' is above 0.98 and the lower bound is above 0.7 (Figure 2B and 2C). C22736T/G74357A SNPs are present in 28 cases, including 25 Germany, one Austria, one UK and one Portugal

cases. G186153A/C188379T SNP pairs has 8 German cases. There are 14 Canadian cases containing G34305A/G148421A/G189246A SNPs (Figure S1). Our results suggest that 11.6% of virus populations have evolved into at least three new lineages.

Moreover, the upper 95% confidence bound of *D*' for SNP pairs C25641T/C70777T and G5592A/G78031A was 0.34 and 0.87, respectively, showing strong evidence of recombination. This result suggests that two Germany cases (ON959149 and ON637939) and one Spain case (ON720849) already gained their mutations via recombination (Figure 2B).

Discussion

In this study, we show the first report of natural recombination of monkeypox virus genome based on: TR (SNP-independent analysis) and LD (SNP-dependent analysis). Kugelman et al. has investigated monkeypox genome diversity from 60 human samples collected in Democratic Republic of the Congo from 2005 through 2007 based on 4 regions with TRs (Kugelman et al., 2014). Virus populations in 2022 pandemics can be categorized into 4 lineages with 11 different subgroups in clade B.1. Like Kugelman et al.'s studies, we found that none of the TR of monkeypox strains were located at the protein coding region. However, all of their 60 samples (2005-2007) had identical right and left TRs. In contrast, we have detected 21 genomes (5.1%) with mismatch TR A/E during the 2022 pandemic to date (Figure 1C).

LD analysis also detected three new lineages (G22736T/G74357A, G186153A/G188379T, G34305A/G148421A/G189246A), suggesting that monkeypox virus has diverged during 2022 pandemics. Based on neutrality test, directional selection appears to not be significant yet (normalized Wu and Fay's *DH*= -0.74, $p > 0.05$), consistent with the idea that SNVs' mutation rates are generally slower than recombination in poxvirus evolution (Coulson and Upton, 2011).

It has been shown TRs with the diverse length (54, 70, 125 bp) near the ends of vaccinia virus genome can provide a novel marker to detect unequal crossing over and compare the relatedness of poxviruses. (Baroudy and Moss, 1982). However, the TR diversity has been underappreciated in poxvirus study. Our results also indicate that TR analysis is well-suited for detecting poxvirus recombination events. In the case of poxviruses, insertion, deletion or recombination may make the construction of phylogenetic trees problematic if only based on the SNP algorithm (Coulson and Upton, 2011). Our results also demonstrated that TR and LD analysis can detect different recombination events (Figure 1F versus 2C). In combination with genomic surveillance, TR and LD analysis are both useful tools to monitor and track phylogenetic dynamics and genetic epidemiology of monkeypox transmission.

TR insertions in the promoter and 3'-untranslated regions of monkeypox virus may also have influence on gene expression and regulation (Supplementary Materials, Figure S2-S5). It has been reported that the 3'-to-5' exonuclease activity of viral DNA polymerase plays an essential role in promoting extraordinarily high levels of genetic recombination in vaccinia infection (Gammon and Evans, 2009). Recombination is involved in vaccinia virus adaptation to counteract the interferon-induced antiviral host cell response mediated by the double-stranded RNA-dependent protein kinase (PKR). One of the weak PKR inhibitors, vaccinia K3L, undergo recurrent gene amplification with a beneficial SNV (His47Arg) via active recombination driven by selection (Elde, et al., 2012; Sasani et al., 2018). This event is mediated by vaccinia RNA polymerase and leads to rapid homogenization of K3L gene arrays (Cone et al., 2017). It is worth noting that natural monkeypox C3L (vaccinia K3L homologue) is truncated at amino acid 43 by the stop codon and loss of binding site to PKRs and eIF2B (Figure S5). It is unclear whether monkeypox virulence is changed due to the loss of the anti-interferon activity.

Materials and methods

Sequence information

Genomic sequences of monkeypox virus (B.1 clade) from January 1 to July 20, 2022 (N=415) were obtained from NCBI database and available on Figshare (DOI: 10.6084/m9.figshare.20486313). FASTA files of viral sequences were downloaded and analyzed by Tandem Repeat Finder Version 4.09 (Benson, 1999). Only score more than 100 by Tandem Repeat Finder were considered TRs. Our TR analysis did not include small TRs (<6 bp, microsatellites), because microsatellites comprise 24% of the nucleotide sequence of these poxvirus genomes on average (Hatcher et al., 2015). Neutrality tests are carried out as described in Supplementary data.

The linkage disequilibrium (LD) analysis

We first searched the loci via LD analysis of DNASP v6, which used the Bonferroni correction for the $\chi 2$ test to determine if the obtained LD between two loci was statistically significant (P< 0.001). Nucleotide polymorphism at 28 SNPs in monkeypox sequences was converted by SNP_tools plug-in in Microsoft Excel to create the input files. HaploView software, version 4.1 (Broad Institute, Cambridge, USA) was used to measure and plot the normalized values (D') of the coefficient of linkage disequilibrium (D) (Barret et al., 2005). D' is defined by dividing D with $D_{\max}$, where $D_{\max}$ is the theoretical maximum difference between the observed and expected haplotype frequencies. HaploView was also used to calculate the squared coefficient of correlation (r^2) and the log of the odds (LOD) of there being a disequilibrium between two loci. To detect recombination, we used HaploView to plot 95% confidence bounds for D'. Pairs are in strong LD if the upper 95% confidence bound for D' is above 0.98 and the lower bound is above 0.7 (no recombination). Strong evidence for recombination is defined if pairs for which the upper confidence bound of D' is less than 0.9.

Footnote:

The methods of neutrality tests, discussion of potential influences of TRs on monkeypox genes, and supplemental figures are available in Supplemental Materials.

Declaration of interests

T.-Y. Yeh and G. Contreras are founders of Auxergen Inc. Y.-C Su and S.-L. Hsieh are stockholders of Auxergen Inc. D.A. Lewis declares no competing interests.

Funding

No funding

Ethical approval

None declared.

References

Ball LA. High-frequency homologous recombination in vaccinia virus DNA. *J. Virol.* 1987; **61**:1788–1795

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; **21**:263–5.

Baroudy BM, Moss B. Sequence homologies of diverse length tandem repetitions near ends of vaccinia virus genome suggest unequal crossing over. *Nucleic Acids Res* 1982; **10**:5673-5679.

Benson G, Tandem repeat finder: a program to analyze DNA sequences. *Nucleic Acids Res* 1999; **27**:573-80.

Cone KR, Kronenberg ZN, Yandell M, Elde NC. Emergence of a Viral RNA Polymerase Variant during Gene Copy Number Amplification Promotes Rapid Evolution of Vaccinia Virus. *J Virol* 2017; **91**: e01428-16.

Coulson D, Upton C. Characterization of indels in poxvirus genomes. *Virus Genes* 2011; **42**:171-177.

Elde NC, Child SJ, Eickbush MT, Kitzman JO, Rogers KS, Shendure J, Geballe AP, Malik HS. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. *Cell*. 2012; **150**: 831–841.

Evans DH, Stuart D, McFadden G. High levels of genetic recombination among cotransfected plasmid DNAs in poxvirus-infected mammalian cells. *J. Virol.* 1988; **62**:367–375

Fenner, F and Comben, BM (1958). Genetic studies with mammalian poxviruses. I. Demonstration of recombination between two strains of vaccinia virus. *Virology* 1958; **5**:530-548.

Gammon DB, Evans DH. The 3′-to-5′ exonuclease activity of vaccinia virus DNA polymerase is essential and plays a role in promoting virus genetic recombination. *J. Virol. 2009*; **83**:4236–4250.

Gershon PD, Kitching RP, Hammond JM, Black DN. Poxvirus genetic recombination during natural virus transmission. *J Gen Virol* 1989; **70**:485-489.

Kugelman JR, Johnston SC, Mulembakani PM. et al Genomic Variability of Monkeypox Virus among Humans, Democratic Republic of the Congo. *Emerg Infect Dis.* 2014; **20**: 232–239.

Hatcher EL, Wang C, Lefkowitz EJ. Genome Variability and Gene Content in Chordopoxviruses: Dependence on Microsatellites. *Viruses* 2015; **7**: 2126–2146.

Sasani FA, Cone KR, Quinlan AR. Elde, NC. Long read sequencing reveals poxvirus evolution through rapid homogenization of gene arrays. *eLife* 2018; **7**:e35453.

Qin L, Evans DH. Genome scale patterns of recombination between coinfecting vaccinia viruses. *J Virol.* 2014; **88**:5277-86.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol Biol Evol* 2017; **34**:3299-3302.

Wittek R, Moss B. Tandem repeats within the inverted terminal repetition of vaccinia virus DNA. *Cell* 1980; **21**:277-84.

Figure legend

Figure 1. TR analysis of genetic diversity and recombination of monkeypox genome in 2022 pandemics. (A to E) Characterization of tandem repeats (TRs) of monkeypox genome in 2022 pandemics. (A) The map of monkeypox genome. The location of TR A to E are indicated. (B) Summary of TR sequences and locations in this study. (C) The sequence variation of TR A/E among monkeypox sequences (B.1 clade) from January 1 to July 20, 2022 (*N*=415). Upper: TR numbers (TRNs), bottom: case numbers. "Mismatch" indicates that TRNs are different between TR A and E. (D) Sequence variation was represented by TR C/F and TRA/E. TRN=9.7 and TRN=7.7 of TR C are labeled as blue and red line, respectively. Group M, U, and I are named based on the representative sequences: MUW1527495 (M, ON019276) or USA_2022_CA002 (U, ON954773), and IRBA22-14 (I, ON755040). (E) Distribution of TR D based on TR numbers. (F) Graphical representation of recombination based on TR polymorphism. Each box (TR A, C, D, E, F) contains individual TRNs with different colors associated with the parental viral sequences: IRBA22-14 (I, ON755040), MUW1527495 (M, ON019276) or USA_2022_CA002 (U, ON954773). Here "parental" only refers to their phylogenic relationship among the sequences, not to the routes of transmission. Recombination is defined when a new virus gains a TR with different TRNs, which leads to the mosaic pattern of the genome. Asterisks include six Slovenia cases that may have evolved into a new lineage.

Figure 2. Linkage disequilibrium (LD) between SNPs in the monkeypox viruses. (A) LD plot of any two SNP pairs among the 28 sites that have minor alleles in at least two strains. Each box represents a pair of SNPs. The number at the top shows the coordinate of sites in the monkeypox genome. Color in the square is given by standard (*D*'/LOD). *D*', normalized the coefficient of linkage disequilibrium. LOD, the log of the odds. Color scheme: bright red: *D*' = 1, LOD ≥ 2; blue: *D*' = 1, LOD < 2, shades of pink/red: *D*' < 1, LOD ≥ 2; white: *D*' < 1, LOD < 2. (B) Summary of monkeypox SNP pairs with strong LD (χ^2 test, P< 0.001). r^2, squared coefficient of correlation. (C) Haplotype blocks organization of HaploView to display the

confidence bounds color scheme in (A). Strong evidence of recombination is defined if pairs for which the upper confidence bound of the *D*' is less than 0.9 (white squares). Pairs are in strong LD if the upper 95% confidence bound of the *D*' is above 0.98 and the lower bound is above 0.7 (black squares). (D) The r^2 of each pair of SNPs (y-axis) against the genomic distance between that pair (x-axis). (E) The LOD of each pair of SNPs (y-axis) against the genomic distance between that pair (x-axis).

Figure 1.

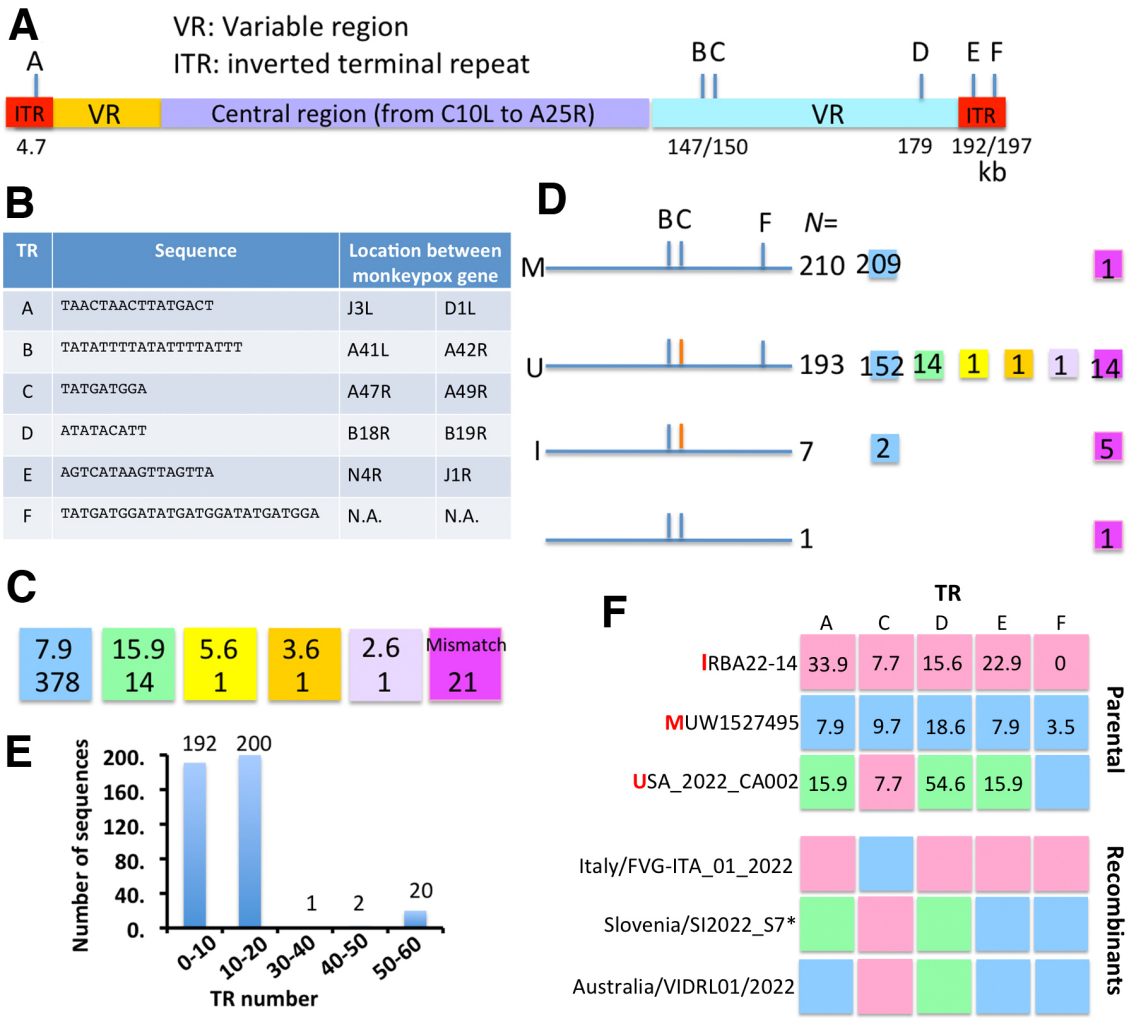
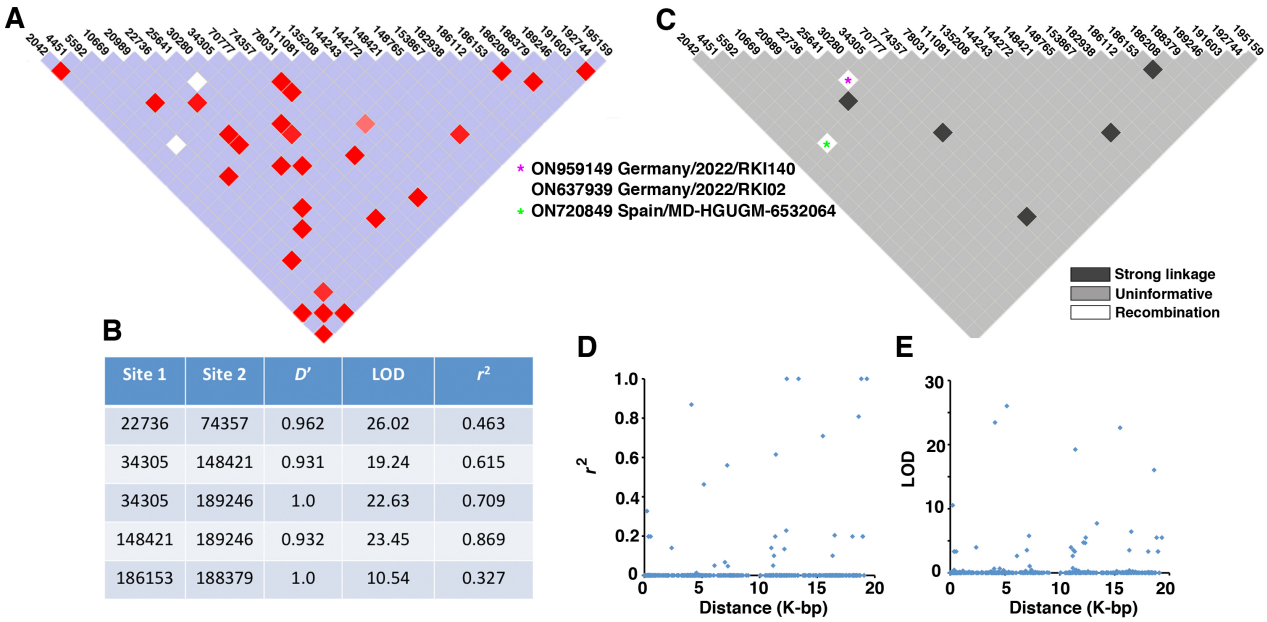

A
VR: Variable region
ITR: inverted terminal repeat
A
BC
D
E F
ITR
VR
Central region (from C10L to A25R)
VR
ITR
4.7
147/150
179
192/197
kb
B
TR | Sequence | Location between monkeypox gene
A | TAACTAACTTATGACT | J3L | D1L
B | TATATTTTATATTTTATTT | A41L | A42R
C | TATGATGGA | A47R | A49R
D | ATATACATT | B18R | B19R
E | AGTCATAAGTTAGTTA | N4R | J1R
F | TATGATGGATATGATGGATATGATGGA | N.A. | N.A.
C
7.9 378
15.9 14
5.6 1
3.6 1
2.6 1
Mismatch 21
E
Number of sequences
200.
160.
120.
80.
40.
0.
192
200
1
2
20
0-10
10-20
30-40
40-50
50-60
TR number
D
BC
F
N=
M
210
209
1
U
193
152
14
1
1
1
14
I
7
2
5
1
1
F
TR
A
C
D
E
F
IRBA22-14
33.9
7.7
15.6
22.9
0
MUW1527495
7.9
9.7
18.6
7.9
3.5
USA_2022_CA002
15.9
7.7
54.6
15.9
Parental
Italy/FVG-ITA_01_2022
Slovenia/SI2022_S7*
Australia/VIDRL01/2022
Recombinants

Figure 2



B

Site 1	Site 2	*D'*	LOD	r^2
22736	74357	0.962	26.02	0.463
34305	148421	0.931	19.24	0.615
34305	189246	1.0	22.63	0.709
148421	189246	0.932	23.45	0.869
186153	188379	1.0	10.54	0.327

Supplemental Materials

Neutrality tests

Sequences are aligned using MAFFT 7 software (Katoh and Standley, 2013; Kuraku et al, 2013) and converted by MEGA 7 (Kuma et al., 2016). Neutrality tests within monkeypox populations were analyzed using the polymorphism of the full monkeypox genome based on the site-frequency spectrum: (1) Zeng's *E* test, (2) Achaz's *Y*, (3) normalized Wu and Fay's *DH* test, and using DNASP v6 software with the camelpox virus (MZ300860) as the outgroup sequence (Rozas, et al., 2017). *DH* test is sensitive to the changes in high-frequency SNPs and primarily affected by directional selection and no other driving forces, such as demographic expansion (Zeng et al., 2006). Zeng's *E* test contrasts high and low frequency regions of frequency spectrum, so it can detect demographic expansion after a sweep (Zeng et al., 2006). To avoid sequence errors that could reject neutrality, Achaz's *Y* test employs the SNP number and frequency spectrum but ignores singletons, which are the class of polymorphisms that have the highest impact on several neutrality tests (such as Tajima's *D*), (Achaz, 2008; Achaz, 2009). Recombination and the linkage disequilibrium analysis were carried out by DNASP v6 (Hudson and Kaplan, 1985; Rozas et al., 2001). Statistical significance of observed values of all tests was obtained by coalescent simulation with free recombination after 1,000 replicates. The probability for each test statistic was calculated as the frequency of replicates with a value lower than the observed statistic (two-tailed test) by DNASP v6 (Rozas, et al., 2017).

Potential influences of TRs on monkeypox genes

TR C is located at the intergenic region between A47R (IL-1/Toll-like receptor signaling inhibitor) and A49R (thymidylate kinase) (Figure S2A). Both A47R and A49R are the earliest expressed genes after virus infection (Assarsson et al., 2008; Yang et al, 2015). Vaccinia V46R (monkeypox A47R homologue) is one of

determinant of viral virulence. (Stack, et al., 2005) Vaccinia mRNAs of early genes terminate within around 50 nucleotides after a UUUUUNU sequence and are subsequently polyadenylated (Yuen and Moss, 1987). Insertion of TR C is located at 14 bp downstream of the stop codon of monkeypox A47R (Figure S2B). Since TR C is also 4 bp downstream of 3' termination sequence (UUUUUAU) of monkeypox virus, this suggests that A47R transcripts may be variable in the 3'-untranslated region between TRN=9.7 versus TRN=7.7. Whether or not the properties (degradation, location et al.) of A47R mRNA transcripts are altered require further investigation.

TR B contains a 19 bp AT-rich sequence (5'-TATATTTTATATTTTATTT-3'), and its TRN is 3.4 in 60.7% of total cases. Since sequencing information near 39.3% of TR B insertion sites is ambiguous, TR B is not included in our analysis. TR B is located at the intergenic region between A41L (chemokine-binding protein) and A42R (profilin-like protein) (Figure S3A). Since A41L and A42R transcribe in the opposite direction, this intergenic region is very likely as the promoter of A41L and A42R gene. A42R promoter contains a 5′-poly(A) leader with TAAAT and AAAAAA element, the feature of intermediate gene promoter (Yang et al., 2011; Yang et al., 2012; Yang et al., 2013) (Figure S3C). TR B is inserted at -77 bp upstream of the start codon of A42R gene, and this TR is not present in vaccinia virus genome (Figure S3B).

TR D is located at -28 bp upstream of the start codon of B19R (IFNa/b receptor glycoprotein) in the promoter region, and its copy numbers are variable (Figure S4A). Poor alignment indicates that the intergenic B18R/B19R region is very different between vaccinia and monkeypox (Figure S3B). Vaccinia B19R also has both TAAAT (-27 to -23) and AAAAAAA (-41 to -35) motif in its promoter (Yang et al., 2012), but these two elements appear to be absent in monkeypox B19R promoter (Figure S3B). Monkeypox virus has 85 bp sequence upstream of the start codon of B19R, and this 85 bp sequence does not exist in vaccinia (Figure S4C and S4D). These results suggest that gene expression of monkeypox B19R may be regulated differently from its vaccinia homologue (Figure S4B). It has been

reported that the upstream core elements of both poxvirus intermediate and late promoters are clearly AT-rich sequences and regulated by TATA-binding proteins (Knutson et al., 2006). We speculate that insertion of TATA-like sequences like TR B or D may also influence gene expression of A42R and B19R.

References

Achaz G. Testing for neutrality in samples with sequencing errors. *Genetics* 2008; **179**:1409–1424.

Achaz G. Frequency spectrum neutrality tests: one for all and all for one. *Genetics* 2009; **183**: 249–258,

Assarsson E, Greenbaum JA, Sundström M et al. Kinetic analysis of a complete poxvirus transcriptome reveals an immediate–early class of genes. *Proc Natl Acad Sci USA.* 2008;**105**:2140–2145.

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; **21**:263–5.

Benson G, Tandem repeat finder: a program to analyze DNA sequences. *Nucleic Acids Res* 1999; **27**:573-80.

Chen B, Wilkening S, Drechsel M, Hemminki K, SNP_tools: A compact tool package for analysis and conversion of genotype data for MS-Excel. *BMC Res Notes* 2009; **2**:214.

Cone KR, Kronenberg ZN, Yandell M, Elde NC. Emergence of a Viral RNA Polymerase Variant during Gene Copy Number Amplification Promotes Rapid Evolution of Vaccinia Virus. *J Virol* 2017; **91**: e01428-16.

Elde NC, Child SJ, Eickbush MT, Kitzman JO, Rogers KS, Shendure J, Geballe AP, Malik HS. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. *Cell*. 2012; **150**: 831–841.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. *Science*. 2002; **296**:2225–2229.

Gammon DB, Evans DH. The 3′-to-5′ exonuclease activity of vaccinia virus DNA polymerase is essential and plays a role in promoting virus genetic recombination. *J. Virol. 2009*; **83**:4236–4250.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013; **30**:772-80.

Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis Version 7.0 for bigger datasets. *Mol Biol Evol.* 2016; **33**:1870–4.

Kuraku S, Zmasek CM, Nishimura O, Katoh K. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res.* 2013;**41**, W22-8.

Hudson RR, Kaplan NL. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 1985;**111**:147-64.

Knutson BA, Liu X, Oh J, Broyles SS. Vaccinia virus intermediate and late promoter elements are targeted by the TATA-binding protein. *J Virol* 2006; **80**:6784-6793.

Sasani FA, Cone KR, Quinlan AR. Elde, NC. Long read sequencing reveals poxvirus evolution through rapid homogenization of gene arrays. *eLife* 2018; **7**:e35453.

Rozas J, Gullaud M, Blandin G, and Aguade M DNA variation at the rp49 gene region of *Drosophila simulans*: evolutionary inferences from an unusual haplotype structure. *Genetics* 2001; **158**,1147-1155.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol Biol Evol* 2017; **34**:3299-3302.

Stack J, Haga IR, Schröder M, et al., Vaccinia virus protein A46R targets multiple Toll-like–interleukin-1 receptor adaptors and contributes to virulence. *J Exp Med.* 2005; **201**: 1007–1018.

Yang Z, Cao S, Martens CA, Porcella SF, Xie Z, Ma M, Shen B, Moss B. Deciphering Poxvirus Gene Expression by RNA Sequencing and Ribosome Profiling. *J. Virol.* 2015; **89**: 6874–6886.

Yang Z, Martens CA, Bruno DP Porcella SF, Moss B. Pervasive Initiation and 3′-End Formation of Poxvirus Postreplicative RNAs. *J Biol Chem.* 2012; **287**: 31050–31060.

Yang Z, Reynolds SE, Martens CA, Bruno DP, Porcella SF, Moss B. Expression Profiling of the Intermediate and Late Stages of Poxvirus Replication. *J. Virol.* 2011; **85**: 9899–9908.

Yang Z, Maruri-Avidal L, Sisler J, Stuart CA, Moss B. Cascade regulation of vaccinia virus gene expression is modulated by multistage promoters. *Virology* 2013; **447**:213-20.

Yuen L, Moss B. Oligonucleotide sequence signaling transcriptional termination of vaccinia virus early genes. *Proc. Natl. Acad. Sci. U.S.A.* 1987; **84**: 6417–6421.

Zeng, K, Fu, YX, Shi S, Wu CI. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* 2006; **174:** 1431-1439.

Figure S1. Sequences that contain SNP pairs with strong linkage disequilibrium, which the upper 95% confidence bound of the *D*' is above 0.98 and the lower bound is above 0.7.

SNPS	Sequences
C22736T/G74357A	OP019276.1 Monkeypox virus isolate hMPXV/Austria/MUW1527495/2022, complete genome
C22736T/G74357A	ON755240.1 Monkeypox virus isolate MPXV/Germany/2022/RKI034
C22736T/G74357A	ON853650.1 Monkeypox virus isolate MPXV/Germany/2022/RKI069, complete genome
C22736T/G74357A	ON813264.2 Monkeypox virus isolate MPXV/Germany/2022/RKI064, complete genome
C22736T/G74357A	ON813266.2 Monkeypox virus isolate MPXV/Germany/2022/RKI056, complete genome
C22736T/G74357A	ON929084.1 Monkeypox virus isolate MPXV/Germany/2022/RKI129, complete genome
C22736T/G74357A	ON929059.1 Monkeypox virus isolate MPXV/Germany/2022/RKI072, complete genome
C22736T/G74357A	ON929080.1 Monkeypox virus isolate MPXV/Germany/2022/RKI125, complete genome
C22736T/G74357A	ON929085.1 Monkeypox virus isolate MPXV/Germany/2022/RKI130, complete genome
C22736T/G74357A	OP018591.1 Monkeypox virus isolate MPXV/Germany/2022/RKI183, complete genome
C22736T/G74357A	OP018599.1 Monkeypox virus isolate MPXV/Germany/2022/RKI174, partial genome
C22736T/G74357A	ON959168.1 Monkeypox virus isolate MPXV/Germany/2022/RKI154, complete genome
C22736T/G74357A	ON959176.1 Monkeypox virus isolate MPXV/Germany/2022/RKI163, complete genome
C22736T/G74357A	OP018606.1 Monkeypox virus isolate MPXV/Germany/2022/RKI184, complete genome
C22736T/G74357A	ON755232.1 Monkeypox virus isolate MPXV/Germany/2022/RKI026
C22736T/G74357A	ON853655.1 Monkeypox virus isolate MPXV/Germany/2022/RKI088, complete genome
C22736T/G74357A	ON853660.1 Monkeypox virus isolate MPXV/Germany/2022/RKI078, complete genome
C22736T/G74357A	ON853672.1 Monkeypox virus isolate MPXV/Germany/2022/RKI093, complete genome
C22736T/G74357A	ON853662.1 Monkeypox virus isolate MPXV/Germany/2022/RKI080, partial genome
C22736T/G74357A	ON853659.1 Monkeypox virus isolate MPXV/Germany/2022/RKI075, partial genome
C22736T/G74357A	ON813257.2 Monkeypox virus isolate MPXV/Germany/2022/RKI060, partial genome
C22736T/G74357A	ON808413.1 Monkeypox virus isolate CVR_MPXV1a, complete genome
C22736T/G74357A	ON853677.1 Monkeypox virus isolate MPXV/Germany/2022/RKI099, partial genome
C22736T/G74357A	ON694331.1 Monkeypox virus isolate MPXV/Germany/2022/RKI015, complete genome
C22736T/G74357A	ON694339.1 Monkeypox virus isolate MPXV/Germany/2022/RKI016, complete genome
C22736T/G74357A	OP018589.1 Monkeypox virus isolate MPXV/Germany/2022/RKI180, complete genome
C22736T/G74357A	ON637938.1 Monkeypox virus isolate MPXV/Germany/2022/RKI01, complete genome
C22736T/G74357A	ON843163.1 Monkeypox virus isolate Monkeypox/PT0048/2022, partial genome

G186153A/C188379T	ON755244.1 Monkeypox virus isolate MPXV/Germany/2022/RKI038
G186153A/C188379T	ON813258.2 Monkeypox virus isolate MPXV/Germany/2022/RKI054, complete genome
G186153A/C188379T	ON813262.2 Monkeypox virus isolate MPXV/Germany/2022/RKI062, complete genome
G186153A/C188379T	OP018588.1 Monkeypox virus isolate MPXV/Germany/2022/RKI169, complete genome
G186153A/C188379T	ON959151.1 Monkeypox virus isolate MPXV/Germany/2022/RKI160, complete genome
G186153A/C188379T	ON929075.1 Monkeypox virus isolate MPXV/Germany/2022/RKI120, complete genome
G186153A/C188379T	ON853658.1 Monkeypox virus isolate MPXV/Germany/2022/RKI072, partial genome
G186153A/C188379T	ON853649.1 Monkeypox virus isolate MPXV/Germany/2022/RKI068, complete genome
G34305A/G148421A/ G189246A	ON803413.1 Monkeypox virus isolate MpxV/human/CAN/UN-NML-2928/2022, complete genome
G34305A/G148421A/G189246A	ON803414.1 Monkeypox virus isolate MpxV/human/CAN/UN-NML-2930/2022, complete genome
G34305A/G148421A/G189246A	ON803415.1 Monkeypox virus isolate MpxV/human/CAN/UN-NML-2936/2022, partial genome
G34305A/G148421A/G189246A	ON880505.1 Monkeypox virus isolate MpxV/human/CAN/UN-NML-2906/2022, partial genome
G34305A/G148421A/G189246A	ON880507.1 Monkeypox virus isolate MpxV/human/CAN/UN-NML-3071/2022, partial genome
G34305A/G148421A/G189246A	ON880510.1 Monkeypox virus isolate MpxV/human/CAN/UN-NML-3193/2022, partial genome
G34305A/G148421A/G189246A	ON880513.1 Monkeypox virus isolate MpxV/human/CAN/UN-NML-3407/2022, complete genome
G34305A/G148421A/G189246A	ON880514.1 Monkeypox virus isolate MpxV/human/CAN/UN-NML-3412/2022, partial genome
G34305A/G148421A/G189246A	ON880515.1 Monkeypox virus isolate MpxV/human/CAN/UN-NML-3467/2022, partial genome
G34305A/G148421A/G189246A	ON983163.1 Monkeypox virus isolate MpxV/human/CAN/UN-NML-3532/2022, complete genome
G34305A/G148421A/G189246A	ON983167.1 Monkeypox virus isolate MpxV/human/CAN/UN-NML-3725/2022, complete genome
G34305A/G148421A/G189246A	ON983159.1 Monkeypox virus isolate MpxV/human/CAN/UN-NML-3293/2022, partial genome
G34305A/G148421A/G189246A	ON983164.1 Monkeypox virus isolate MpxV/human/CAN/UN-NML-3558/2022, partial genome
G34305A/G148421A/G189246A	ON754989.2 Monkeypox virus isolate mpx_2022_Canada_AB2, partial genome

Figure S2. TR C in the intergenic region of monkeypox A47R and A49R. (A) The map of TR C, A47R, and A49R gene. Arrows indicate the gene direction. (B) The location of TR C at the 3'-untranslated region of A47R transcript. The 3' termination UUUUUNU sequence for adding poly(A) tail is boxed in red.

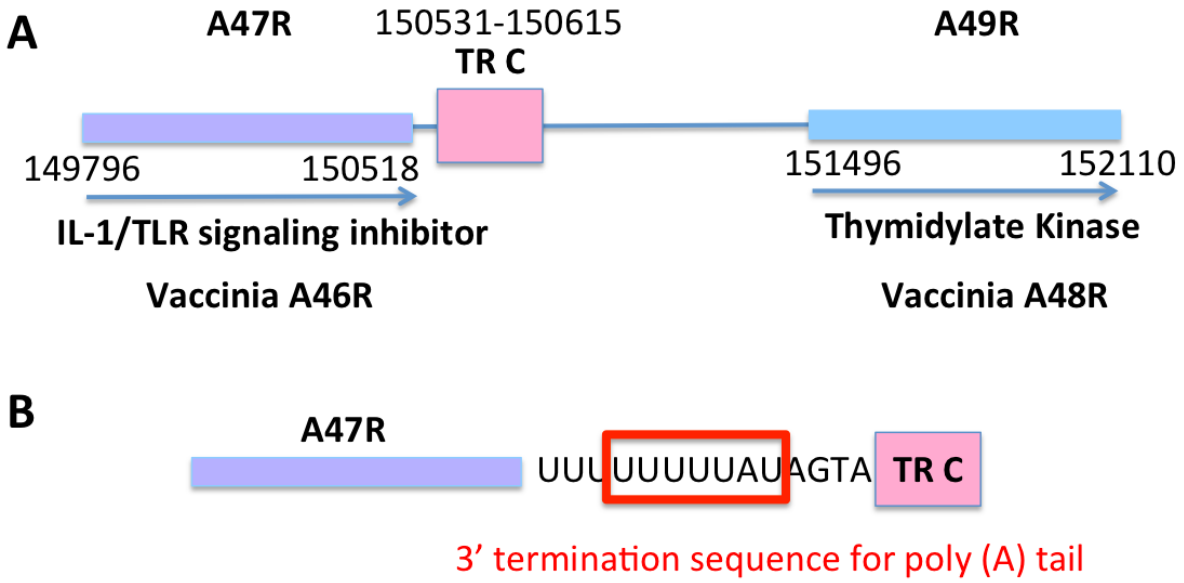


Figure S3. TR B in the intergenic region of monkeypox A41L and A42R. (A) The map of TR B, A41L, and A42R gene. Arrows indicate the gene direction. Sequence B and C are shown in (B) and (C). (B) Sequence of TR B insertion in monkeypox genome as in (A) compared to vaccinia sequence. (C) Alignment of monkeypox and vaccinia in A42R promoter. TAAAT and AAAAAA element are boxed in red.

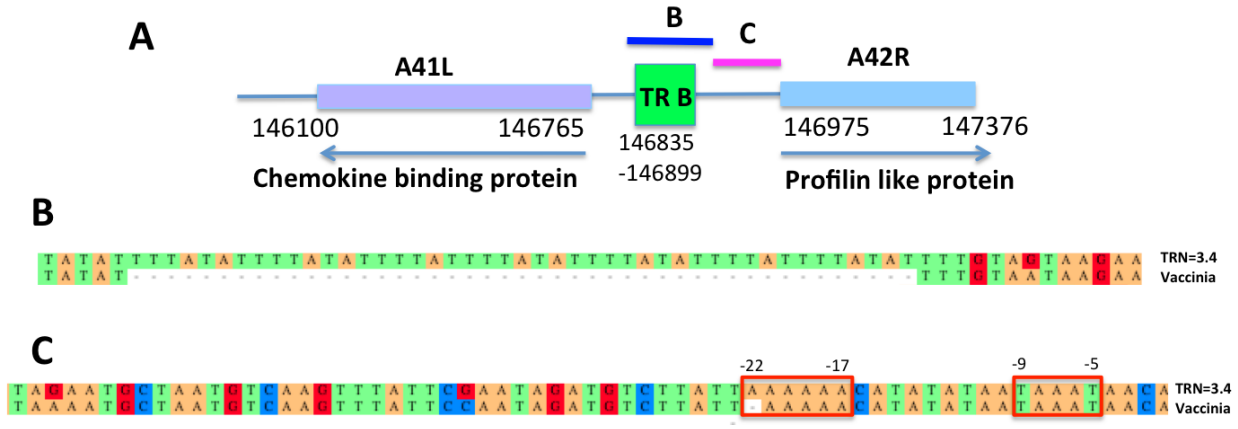


Figure S4. TR D in the intergenic region of monkeypox B18R and B19R. (A) (A) The map of TR D, B18R, and B19R gene. Arrows indicate the gene direction. Sequence B, C and D are shown in (B), (C) and (D). (B) Alignment of monkeypox sequence B with vaccinia B19R promoter, including TAAAT and AAAAAAA element only found in vaccinia. (C and D) Sequences of B19R promoter of monkeypox. Sequence of TR D insertion in monkeypox genome as in (A).

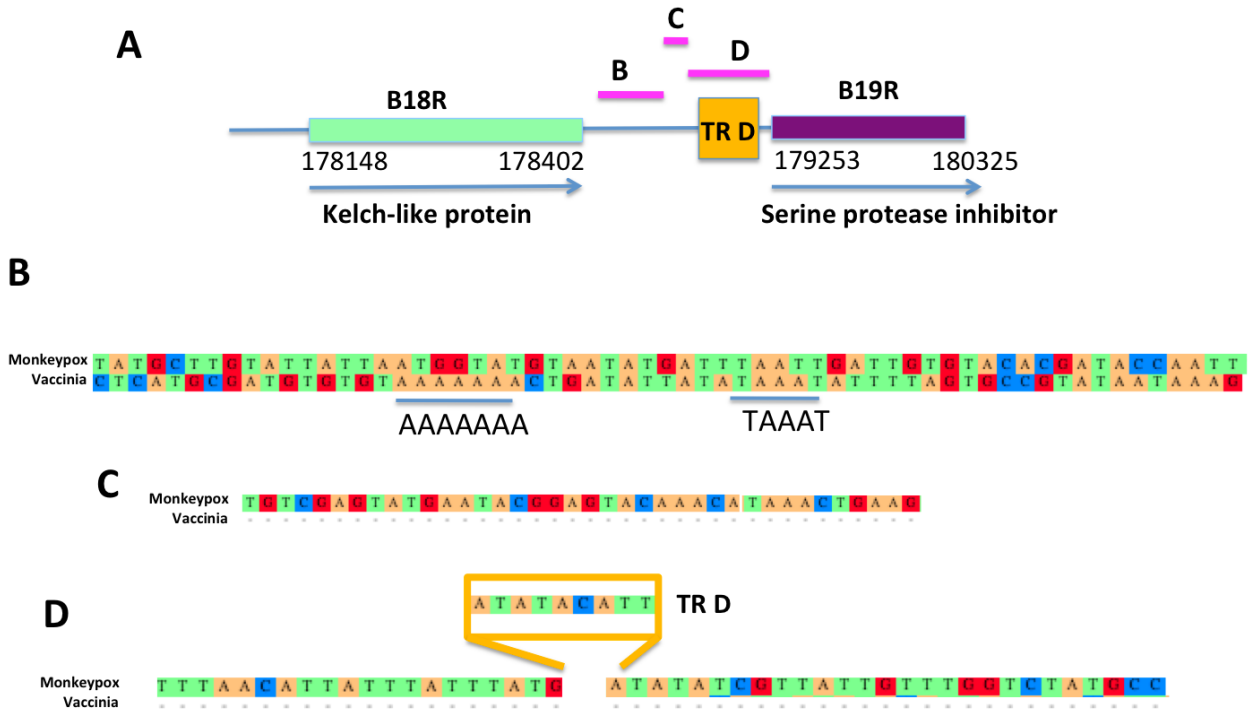


Figure S5. Comparison of vaccinia K3L and monkeypox C3L sequence. The stop codon was introduced at bp 130-132, causing C3L a truncated protein compared to K3L.

Vaccinia K3L	112	GCT	ATC	TTG	GCA	GAG	AGT	GTT	AAG	ATG	CAT	ATG	144
	38	A	I	L	A	E	S	V	K	M	H	M	48
Monkeypox C3L	112	GCT	ATC	TTG	GCA	GAG	TGT	TAA	GAG	ATG	CAT	ATG	144
	38	A	I	L	A	E	C	*					43