Expansion of TC-Containing Tandem Repeats in Monkeypox Virus: A Putative Defense Mechanism Against APOBEC-Mediated Mutagenesis

Sabancı Üniversitesi
SMTB
Cancer Genomics Lab



BACKGROUND

Monkeypox virus (MPXV) is an enveloped double-stranded DNA virus of the *Orthopoxvirus* genus that causes mpox (Fig.1). During the 2022 global outbreak, MPXV showed an unusual burst of >50 APOBEC-signature mutations (Fig.2), contrasting with its typical evolutionary rate of only ~1–2 mutations per year. APOBEC3 enzymes are components of the host innate antiviral defense, restricting viruses by editing their genomes. They act on single-stranded DNA, preferentially targeting TC motifs and deaminating C→U, which gives rise predominantly to C→T (and complementary G→A) mutations.

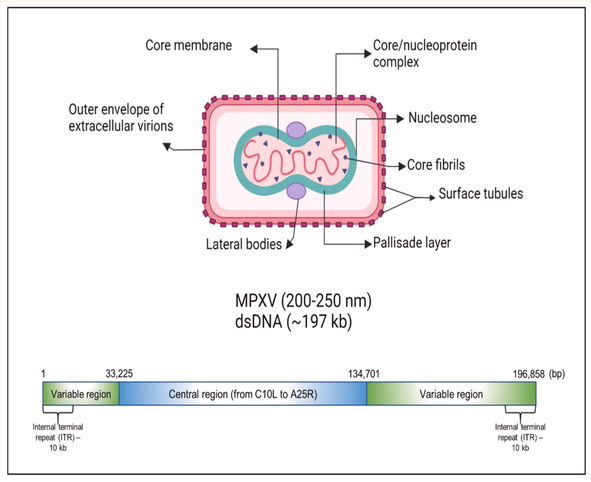


Fig. 1. Structure and genome organization of monkeypox virus (MPXV).

Вирус оспы обезьян (MPXV) – оболочечный двуцепочечный ДНК-вирус рода Orthopoxvirus семейства Poxviridae (Fig.1). Во время глобальной вспышки 2022 года в геномах MPXV было выявлено необычно большое число – более 50 – мутаций с мутационной подписью APOBEC (Fig.2), что резко превышает обычную скорость накопления мутаций у этого вируса – около 1-2 мутаций в год. APOBEC3 – семейство ферментов врождённой противовирусной защиты, ограничивающих размножение вирусов путём редактирования их геномов. APOBEC3 действует преимущественно на одноцепочечную ДНК, дезаминируя цитозин (C→U) главным образом в TC-контексте, что приводит к характерным заменам C→T и комплементарным G→A.

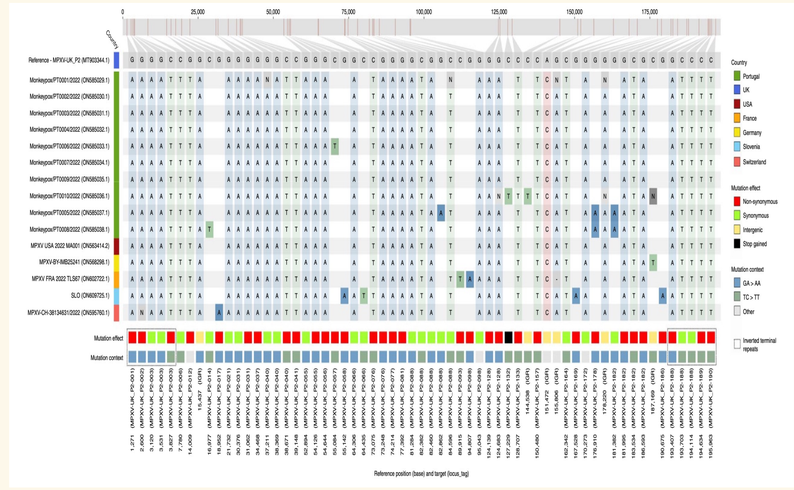


Fig. 2. Distribution of mutations across MPXV genomes from the 2022 outbreak, highlighting APOBEC-associated TC→TT and GA→AA substitutions.

HYPOTHESIS

We found expansions of TC-containing repeats in several MPXV strains, which may act as APOBEC decoys, potentially reducing APOBEC-induced mutagenesis in functional genomic regions.

Мы обнаружили экспансию TC-содержащих повторов в нескольких штаммах MPXV, которые могут служить мишенями-приманками для APOBEC, потенциально снижая APOBEC-индуцированный мутагенез в функционально значимых участках генома.

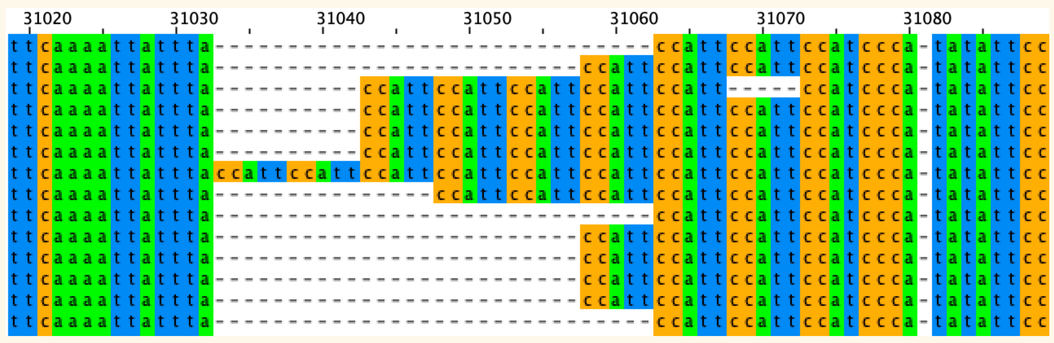


Figure 3. Multiple sequence alignment illustrating the expansion of TC-containing tandem repeats in selected monkeypox virus (MPXV) strains.

DATA & METHODS

- 196 MPXV genomes sampled in Nigeria and Cameroon between 2018 and 2023 were obtained from the dataset of Parker et al. (2025) (Fig. 4).
- A custom python-script was developed to identify genomic regions containing TC-rich tandem repeats and characterize repeat expansions.
- A phylogenetic tree of MPXV strains was reconstructed using FastTree.
- Protein structure models were generated using AlphaFold.

- 196 геномов MPXV, полученных из образцов, собранных в Нигерии и Камеруне в период с 2018 по 2023 год, были взяты из набора данных Parker et al. (2025).
- Для выявления участков генома, содержащих богатые TC тандемные повторы, и характеристики экспансии повторов был разработан собственный скрипт на языке Python.
- Филогенетическое дерево штаммов MPXV было реконструировано с использованием FastTree.
- Модели структуры белков были построены с использованием AlphaFold.

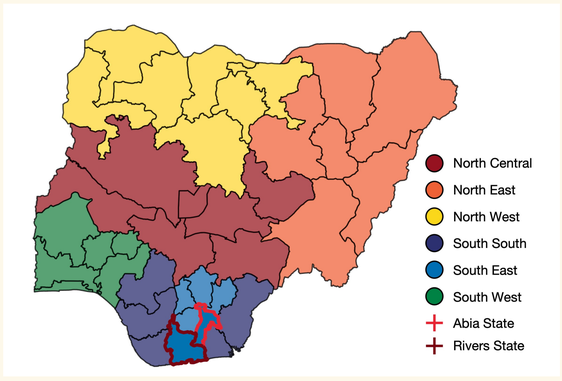
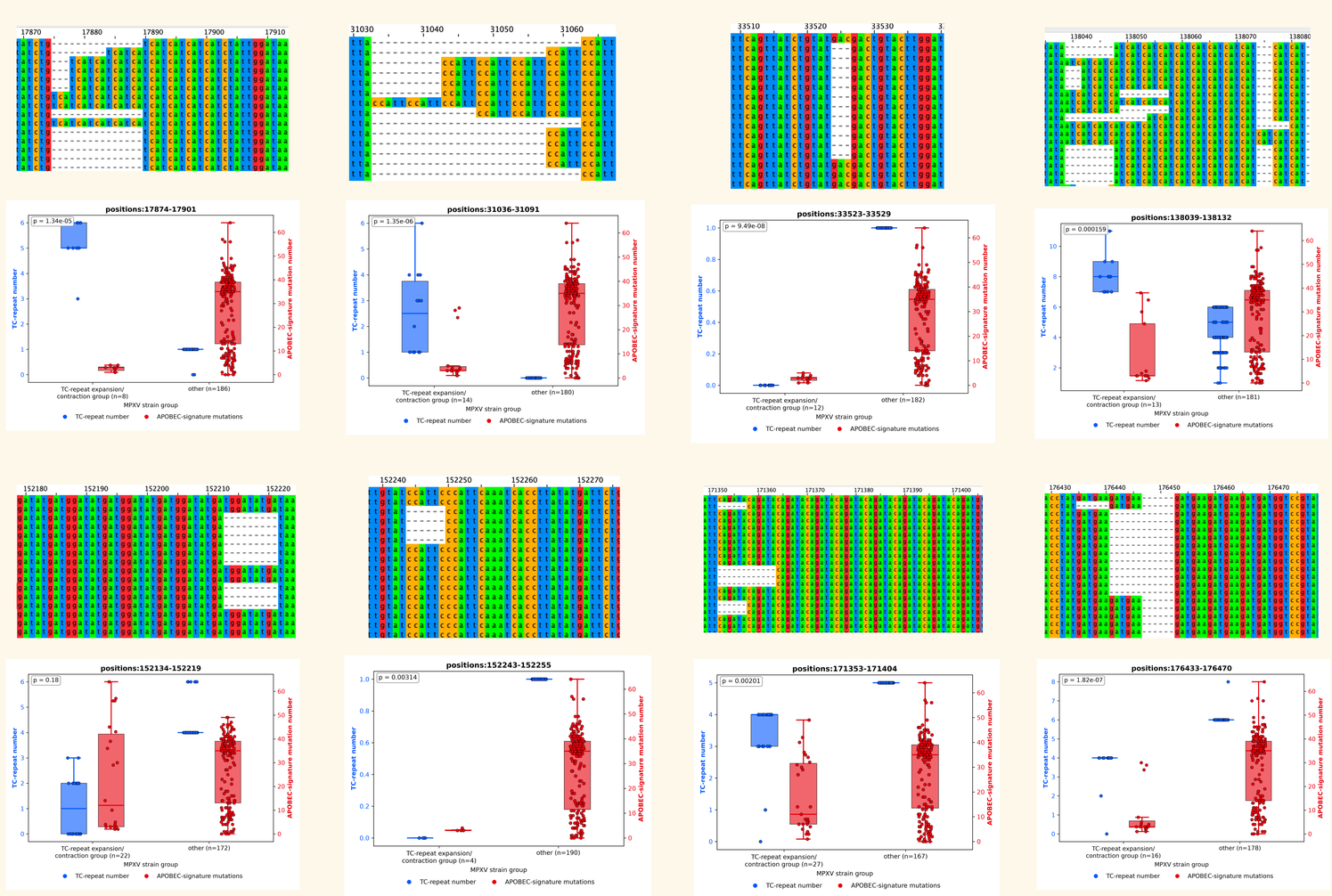


Figure 4. Geographic distribution of MPXV sampling in Nigeria (adapted from Parker et al., 2025), with MPXV strains analyzed in our study selected from the genomic dataset reported in this work.

RESULTS

- A custom script was developed to identify genomic regions containing TC-rich tandem repeats and characterize changes in repeat copy number across MPXV strains.
- For each repeat-containing region, strains were divided into two groups: strains that gained or lost repeat units (TC-repeat altered) and all other strains.
- For each region, we compared the number of APOBEC-signature mutations between the TC-repeat altered and other strains to test whether changes in TC-containing repeats were associated with differences in APOBEC mutagenesis.

- Был разработан скрипт для выявления участков генома, содержащих TC-богатые тандемные повторы, и анализа изменений числа копий повторов у различных штаммов MPXV.
- Для каждой области, содержащей повторы, штаммы были разделены на две группы: штаммы, которые приобрели или потеряли повторы (TC-repeat altered), и все другие штаммы.
- Для каждой области мы сравнили количество мутаций с APOBEC-подписью между этими двумя группами, чтобы определить, связаны ли изменения TC-содержащих повторов с различиями в уровне APOBEC-индуцированного мутагенеза.



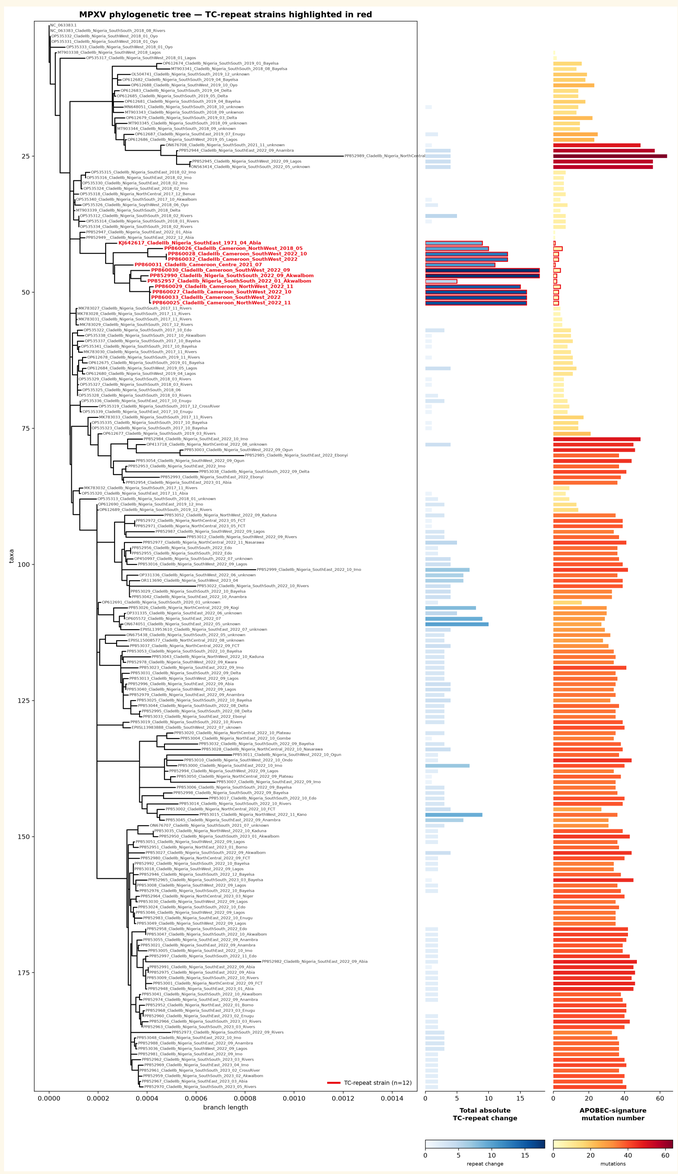


Figure 4. Comparison of APOBEC-signature mutation burden between TC-repeat altered and other strains across individual TC-containing repeat regions. For each region, strains were classified according to changes in repeat copy number, and the number of APOBEC-signature mutations was compared between the two groups.

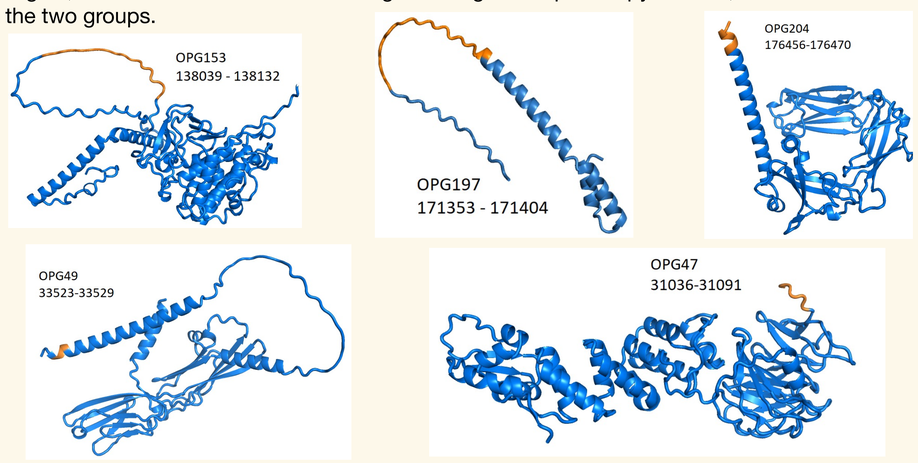


Figure 5. Genomic and structural localization of regions with changes in TC-containing tandem repeat copy number. Three regions are located in intergenic regions, whereas the others fall within protein-coding sequences and predominantly map to protein loops, where repeat-length changes are unlikely to substantially destabilize protein structure.

CONCLUSIONS

- Изменения числа TC-содержащих тандемных повторов были обнаружены в небольшой группе из 12 близкородственных штаммов MPXV, что указывает на то, что увеличение или уменьшение числа повторов произошло в пределах определенной вирусной линии.
- Штаммы с измененным числом TC-повторов характеризовались меньшим количеством мутаций с APOBEC-подписью по сравнению с штаммами со стабильным числом повторов.
- Увеличение числа TC-содержащих повторов потенциально может создавать дополнительные мишени для APOBEC, выступая в роли «приманки» и тем самым снижая APOBEC-индуцированный мутагенез в других участках вирусного генома.
- Необходим дальнейший анализ, чтобы определить, являются ли изменения числа TC-повторов адаптивным ответом на APOBEC-опосредованную противовирусную защиту хозяина или отражают другой эволюционный процесс.

- Changes in TC-containing tandem repeats were identified in a small group of 12 closely related MPXV strains, suggesting that repeat expansion/contraction occurred within a specific viral lineage.
- Strains with altered TC-repeat copy number showed a lower burden of APOBEC-signature mutations compared with closely related strains with stable repeat numbers.
- Expansion of TC-containing repeats could potentially provide additional APOBEC target sites, acting as a “decoy” that may reduce APOBEC-induced mutagenesis elsewhere in the viral genome.
- Further analysis is required to determine whether TC-repeat changes represent an adaptive response to APOBEC-mediated host defense or reflect another evolutionary process.