Rapid Identification of Monkeypox Virus using Tandem Repeats with Insertion, Deletion and SNPs

Xiangxue Guo [1, 2], Pengbo Liu [1, 2], Yu Li [1]

[1]Centers for Disease Control and Prevention, 1600 Clifton Rd NE, Atlanta, GA; [2]Eagle Medical Services, LLC (EMS)

INTRODUCTION

Slippage in tandem repeat regions occurs during DNA replication. In the human genome it happens about once per 1,000 generations; three orders of magnitude more frequently than point mutations. Thus, the copy number changes and genetic mutations in tandem repeat regions are very useful for genome evolution studies. Comparative analysis of tandem repeats with insertions, deletions and SNPs in pathogen sequences has become possible due to the large number of genomes available. In functional coding and regulation regions tandem repeats are notable as potential genetic markers with epidemiologic significance on host-pathogen interaction and disease diagnosis. Beside the traditionally used hemagglutinin gene, a three real-time PCR assay for DNA sequence SNPs has been developed to differentiate MPV clades (*ref1*). However, the 3-PCR sequencing method is still slow, costly and shows poor resolution.

PROJECT GOALS

To identify a tandem repeat region with genetic mutations in the MPV genome that can be used to rapidly cluster and identify a new pathogen's sub-group directly from raw NGS data or through a single PCR.

CONCLUSION

Of the 48 NCBI-available MPV genomes, eight tandem repeats were identified as having large statistical variability in their sequence patterns and copy number changes. However, when combining the consideration of their genetic mutations within each tandem repeat, a 9-nt tandem repeat rearrangement (mpvATATGATGG152k) was found to be highly representative of the different MPV groups: West and Central Africa-A, -B, and -C/d. Through looking at both copy number changes and the potential nucleotide insertions, deletions and SNPs within that tandem repeat region, a newly found MPV pathogen can be rapidly identified as a specific group of MPVs by sequencing the short local region with a single PCR. This method will be much faster, less costly and provide higher resolution than the current standard.

RESULTS

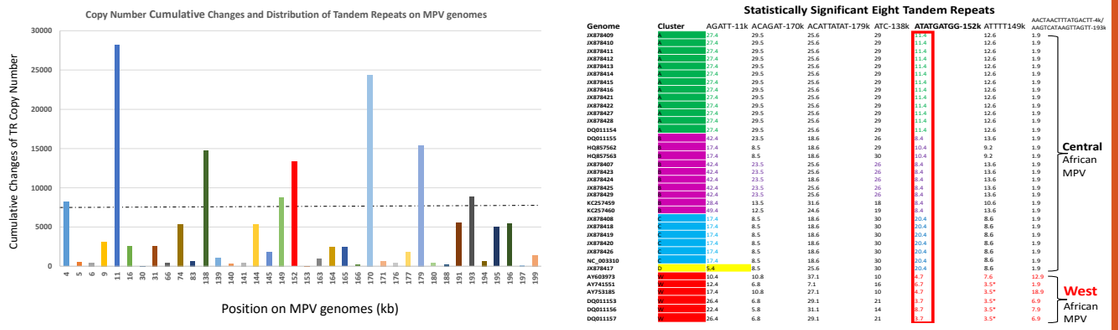

Figure 1. Tandem repeat distribution, cumulative changes (left) of copy number (right) in different MPVs
The dashed line indicates the cutoff for tandem repeats that exhibit large cumulative changes in copy number in MPVs.

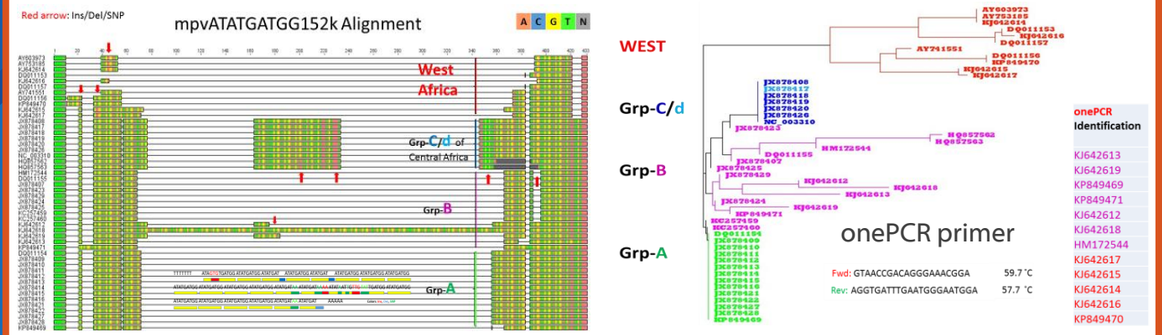

Figure 2. Alignment of ATATGATGG152k local region DNA (left) and clusters coincided with MPV groups(right)
These short local DNA sequences were aligned and placed in a neighbor-joining tree, where the clustering of the tandem repeat regions could be compared with the clustering of the whole genome.

FUTURE DIRECTION

- Identify more tandem repeat local regions with increased genetic complexity
- Applying the testing hypothesis on the classification of new MPV pathogens from CDC NGS-dataset
- Developing practical, simple and rapid single PCR identification methods for MPV and other pathogens

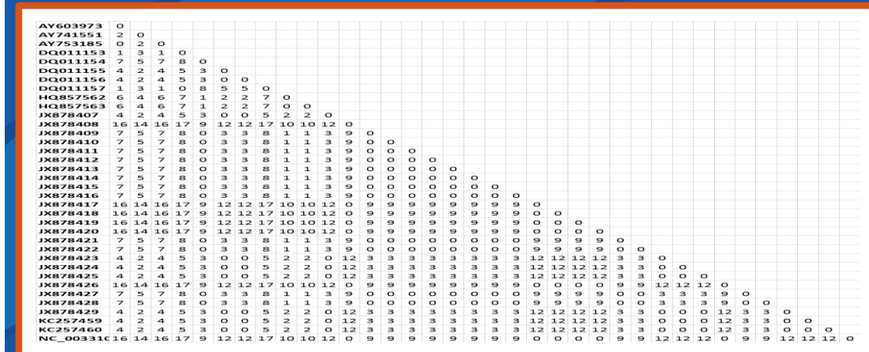

Table 1. Tandem repeat copy number changes of ATATGATGG152k among MPVs
The software *trf* (tandem repeat finder, *ref2*) was applied to identify the sequence pattern and copy number of tandem repeats in each MPV genome. The cumulative changes in copy numbers among MPVs in the dataset are compared and normalized with random nucleotide sequence(*ref3*).

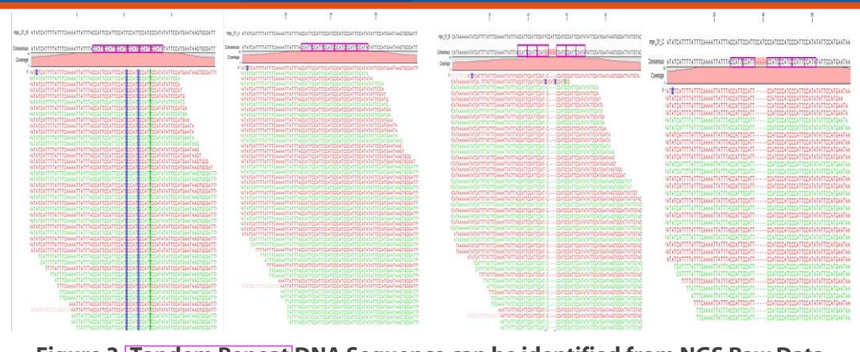

Figure 3. Tandem Repeat DNA Sequence can be identified from NGS Raw Data
Using CLC_Genomic workbench or Linux server, the same DNA consensus with tandem repeats can be obtained directly from the raw NGS data of a newly found sample. The consensus remains regardless of the MPV reference sequence applied. This shows that the process is not dependent on the sub-type reference.

ACKNOWLEDGEMENTS

This work is supported by the CDC's Office of Advanced Molecular Detection (OAMD)

REFERENCE

1. Y. Li, *Journal of Virological Methods* (2010) 169: 223-227
2. G. Benson, *Nucleic Acids Research* (1999) 27(2): 573-580
3. J. Mrazek, *Proc Natl Acad Sci USA* (2007) 104(20): 8472-7

National Center for Emerging and Zoonotic Infectious Diseases

Office of Infectious Diseases, Center of Disease Control and Prevention

www.cdc.gov | Contact CDC at: 1-800-CDC-INFO or www.cdc.gov/info
The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.