

# Investigating hemmaglutin gene of Monkeypox virus for mutations in the recent outbreak - 2022

Cristian Silva<sup>1</sup>

<sup>1</sup>DreamSpace Academy Bio Lab, Batticaloa

## ABSTRACT

The current (2022) Monkeypox virus outbreak was declared a Public Health Emergency of International Concern (PHEIC). Mutations in the monkeypox virus can be contributing factor to its recent outbreak. In this project, we investigate Hemmaglutin gene which plays a role in virion internalization for mutations.

Keywords: Monkeypox, Hemmaglutin, Glycoprotein, Comparative genomics

## INTRODUCTION

Monkeypox virus causes smallpox-like disease but symptoms are mild compared to smallpox. It is a viral zoonotic disease endemic to Africa. Monkeypox virus has a 5-21 day incubation period and the disease lasts for 2 to 4 weeks followed by spontaneous recovery (Ala'A. B et al. (2022)). Monkeypox virus mainly has two strains central African and west African. The virus belongs to the family Poxviridae, genus Orthopoxvirus. Known pathogens like Monkeypox virus, vaccinia virus, cowpox virus and variola virus all belong to the genus Orthopoxvirus (Kugelman et al. (2014)).

The monkeypox virus strain that originates from the Congo basin of Africa is now considered clade 1 and the west African strain is considered clade 2. Clade 2 has two subclades which are 2a and 2b. The recent outbreak (2022) emerged from the 2b subclade. A lineage is closely genetically related variants of the same virus which has a common ancestor. To this date (29.08.2022), there are 10 lineages from the 2b subclade which are most likely to increase in number.

Viral immunomodulatory proteins play an important role in disease transmission and virulence. In this project, we are investigating the hemmaglutin gene mutations in monkeypox viral lineages. Hemmaglutin protein is a major glycoprotein that facilitates virion internalization by host cells (Uniprot - P04664). Single point or multipoint mutations in the gene can result enhance virulence or hinder virulence.

In this project, Hemmaglutin gene sequences of B.1, B.1.1, B.1.2, B.1.3, B.1.4, and B.1.5 lineages were compared with the reference genome (NC\_063383.1) Hemmaglutin gene sequence.

## SOFTWARE

To run the analysis below mentioned software/packages was installed.

MAFFT v7.245, Python 3.9, Entrez direct 13.9, Conda 4.13.0, Biopython 1.79

## METHODOLOGY

### Data retrieval

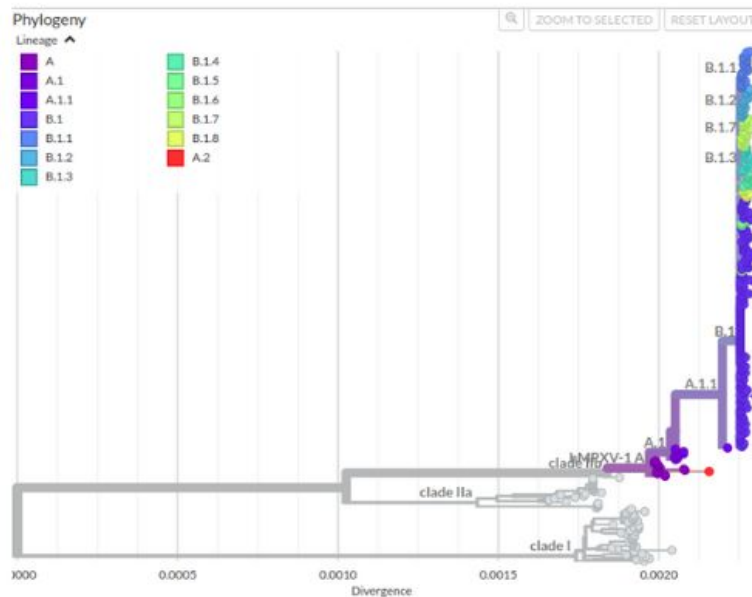
nextstrain.org website was referred to identify recently published genomes of different lineages. Once accession numbers were obtained data was retrieved through NCBI Entrez Direct (EDirect) through the Unix command terminal.

### Write python and Shell script for analysis

In this project, 3 python functions were written to obtain an aligned gene sequence of hemmaglutin gene from each genome, to find nucleotide mutations among comparing DNA sequences, and to find amino acid mutations among comparing amino acid sequences. Biopython package was used to import genomes

## Genomic epidemiology of monkeypox virus

Built with nextstrain/monkeypox. Maintained by Nextstrain team. Enabled by data from GenBank.  
Showing 434 of 434 genomes.



**Figure 1.** Lineages of monkeypox virus.

to the pipeline and converted to 'Seq' objects at the beginning of the analysis. The jupyter notebook was uploaded to GitHub.

### Nucleotide sequence alignment

MAFFT was used to align the genomes to the reference genome of monkeypox virus.

### Find mutations in DNA sequences

Aligned nucleotide sequences were compared to find mutations. DNA sequences were passed as arguments for the previously written function 'get\_mutations'.

### Amino acid sequence alignment

Sequences which has nucleotide differences were translated to amino acid sequences and again aligned against reference genome hemmaglutin protein amino acid sequences.

### Find mutations in amino acid sequences

Aligned amino acid sequences were compared to find mutations. Amino acid sequences were passed as arguments for the previously written function 'get\_aa\_mutations'.

## RESULTS

Retrieved gene sequences of hemmaglutin genes of the 2022 monkeypox outbreak were conserved. Hemmaglutin gene sequence of 6 isolates from the 2022 outbreak was compared with the sequence of the monkeypox reference genome, the HA gene (Table 01).

A Single nucleotide change was observed at 248 positions of OP185714.1 genome hemmaglutin gene compared to the reference genome, hemmaglutin gene. The reference genome had Thymine but the OP185714.1 genome had Cytosine. Amino acid sequences of all six HA sequences were identical. All six amino acid sequences had neither gaps nor insertion/deletions.

Accession number	Lineage	Nucleotide differences compared to reference genome
OP150926.1	B.1	0
OP215252.1	B.1.1	0
OP215274.1	B.1.2	0
OP215250.1	B.1.3	0
OP185714.1	B.1.4	1
OP215225.1	B.1.5	0

**Table 1.** Nucleotide changes compared to reference genome.

## DISCUSSION

Viruses undergo subtle genetic changes through mutations. However, it can be deleterious, neutral or sometimes favourable for the virus. RNA viruses have a high rate of mutation occurrence compared to DNA viruses. Further, viruses could have major genetic changes through recombination (Fleischmann Jr (1996)). In human pathogenic viruses, the APOBEC3 enzyme is also playing a role in inducing viral mutations. APOBEC enzymes are essential components of the human innate immune system and are able to inhibit a wide range of viruses. However, these enzymes can accidentally induce mutations in viruses that enhance viral fitness, host range and virulence (Sadeghpour et al. (2021)). Recent studies suggest, that APOBEC3 is acting in various hosts in low-level accumulating mutations in a series of infections contributing to an increased number of lineages (Gigante et al. (2022)).

This project, only focused on Hemmagglutinin gene mutations in monkeypox viruses however while analysing the gene for mutations, the author was able to build the bioinformatics pipeline for analysing any gene of interest for mutations by comparing to reference genome or comparing to other lineages. Further analysis can be done for immunomodulatory genes like BR-203, BR-209, and COP-C3L.

According to the bioinformatics analysis, hemmagglutinin gene doesn't show any mutations except B.1.4 compared to the reference genome. However, the identified mutation was a single nucleotide change and in protein translation, the nucleotide change didn't contribute to an amino acid change which concludes all the analysed viruses have the same amino acid sequence as the reference amino acid sequence.

Further, the development of a bioinformatics pipeline could be done and automation and development of a web app is a possibility. Results can be enhanced by adding more lineages to the analysis and increasing the number of genomes analysed.

## ACKNOWLEDGMENTS

This project was done at DreamSpace Bio Lab, the first community bio lab in Sri Lanka and funded by DreamSpace Foundation.

## REFERENCES

- Ala'A. B, A.-T., Albakri, R., and Alabsi, S. (2022). The outbreak of human monkeypox in 2022: A changing epidemiology or an impending aftereffect of smallpox eradication. *Front. Trop. Dis.* 3: 951380. doi: 10.3389.
- Fleischmann Jr, W. R. (1996). Viral genetics. *Medical Microbiology*. 4th edition.
- Gigante, C. M., Korber, B., Seabolt, M. H., Wilkins, K., Davidson, W., Rao, A. K., Zhao, H., Hughes, C. M., Minhaj, F., Waltenburg, M. A., et al. (2022). Multiple lineages of monkeypox virus detected in the united states, 2021-2022. *bioRxiv*.
- 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. *Emerging infectious diseases*, 20(2):232.
- Sadeghpour, S., Khodaei, S., Rahnama, M., Rahimi, H., and Ebrahimi, D. (2021). Human apobec3 variations and viral infection. *Viruses*, 13(7):1366.