International Conference on Sustainable Biotechnology (ICoSBi)
Available on-line at: https://sicash.sliit.lk/

Phylogenomic and Pangenome Analysis of Human Monkeypox Virus Strains in Recent Outbreaks Identifies Lineage-Specific Diversification

B. N. N. T. Batagoda[1]*, K. S. Senanayake[2]

[1]*Department of Interdisciplinary Studies, Faculty of Information Technology, University of Moratuwa, 10400, Sri Lanka*
[2]*Institute of Biochemistry, Molecular Biology and Biotechnology (IBMBB), University of Colombo, Cumaratunga Munidasa Mw, Colombo, 00300, Sri Lanka.*

Corresponding author*: nisharab@uom.lk

Abstract

Human monkeypox (Mpox) is a significant zoonotic disease that has seen a recent global resurgence, leading to a second WHO Public Health Emergency of International Concern (PHEIC) declaration in August 2024. This study investigated the genomic diversity and evolutionary dynamics of MPXV using 197 novel complete genomes sequenced between December 2023 and April 2025, alongside two reference sequences. Phylogenetic analysis revealed that recent global outbreak strains predominantly cluster within the West African lineage, distinct from older Central African isolates. The tight clustering of newer isolates, particularly from the USA and Australia, suggests low mutation rates and localised transmission following multiple independent introductions. Pangenome analysis comparing the two reference genomes identified 189 core genes, with clade-specific genes highlighting their evolutionary divergence. The pangenome of the novel isolates contained 238 genes, including 181 highly conserved core genes, and a notable presence of accessory and rare genes. These findings provide insights into Mpox's genomic stability and ongoing diversification as well as lead to identifying potential targets for antivirals and vaccines, emphasising the importance of continuous genomic surveillance.

Keywords: *Global outbreak, monkeypox, MPXV, pangenome, phylogenetic analysis*

https://doi.org/10.54389/EGSK4441

Introduction

Human monkeypox (Mpox) has become a serious viral infection that has threatened public health in recent years. It is caused by the monkeypox virus (MPXV), an Orthopoxvirus, whose primary hosts are small mammals, and the disease is endemic in Africa. In 1970, mpox was discovered in humans for the first time in the Democratic Republic of the Congo (DRC), formerly known as Zaire. Several countries have lately reported cases of MPXV, and now it is endemic in the Central African Republic, Cameroon, the Democratic Republic of the Congo, Ivory Coast, Nigeria, Gabon, Liberia, South Sudan, and Sierra Leone(Bass et al., 2013).

Since May 2022, the World Health Organization (WHO) has declared a Public Health Emergency of International Concern (PHEIC) due to the growing number of confirmed cases reported from several non-endemic countries (Nuzzo et al., 2022). That PHEIC was declared over in May 2023 after there had been a decline in global cases. In August 2024, the WHO declared Mpox as a PHEIC for the second time, considering the recent recurrence of Mpox cases in the Democratic Republic of the Congo and its spread to nearby countries (Zumla et al., 2025). As of 31st May 2025, 150,069 laboratory confirmed cases, and 343 deaths have been reported across 137 countries (WHO, 2025).

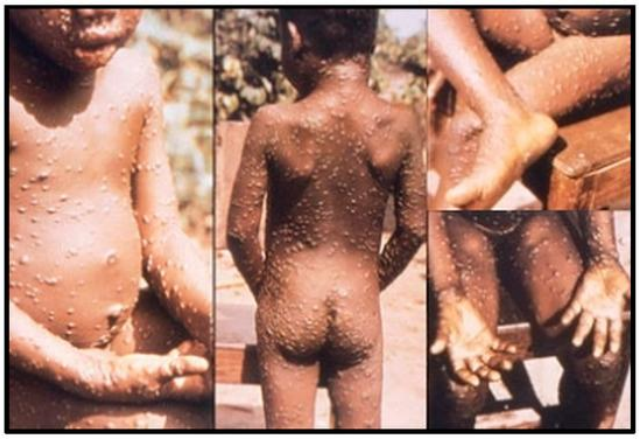

Figure 1: *Clinical picture of Monkeypox rash (Source: Brown and Leggat, 2016)*

Monkeypox virus, belonging to the Orthopoxvirus genus, shows clinical features resembling smallpox, (Figure 1), but generally milder. Since 90% of mpox patients suffered from enlarged lymph nodes following the infection, which are uncommon in smallpox, these enlarged lymph nodes are considered as a characteristic feature of mpox (Ahmed et al., 2023). Monkeypox virus spreads through both animal-to-human and human-to-human pathways. Humans may acquire the infection by handling infected animals or contaminated materials, particularly through exposure to body fluids such as blood, saliva or urine (Araf et al., 2024). Transmission between humans occurs mainly through close contact, including exposure to respiratory droplets, skin lesions or other bodily fluids of infected individuals. Human-to-human transmission through close contact with infected individuals was the primary source of the 2022 outbreak, which was primarily caused by the clade IIb version of the virus; the majority of cases were reported in men who had sexual contact with other men. The idea of sexual transmission has been supported by multiple studies that have found and isolated viral DNA in the semen of infected individuals, weeks after their infection (Antinori et al., 2022).

The Mpox virus is classified into two major genetic clades: Central African (Clade I) and West African (Clade II). Clade I, found predominantly in Central Africa and particularly in the Democratic Republic of the Congo (DRC), is linked with more severe disease and higher mortality rates of 4–11%. In contrast, Clade II, which remained largely restricted to West Africa until the 2022 global outbreak, is associated with milder symptoms and a lower fatality rate of under 4% (Bunge et al., 2022). Furthermore, Clade II is subdivided into two lineages, IIa and IIb (Happi et al., 2022). Subclade IIa, which includes West African isolates, is typically linked to milder disease, lower mortality and reduced transmissibility, whereas subclade IIb includes genomes from recent outbreaks, including the current global spread, with its clinical and epidemiological characteristics still under investigation. The Mpox virus genome is a double-stranded DNA molecule of about 196,858 nucleotides, encoding 191 non-overlapping genes, with reference sequences NC_003310.1 (Central African clade) and NC_063383.1 (West African clade) available in the NCBI database.

This study aims to investigate whether the recent surge in infections could be attributed to genomic changes in the virus, focusing on novel strains sequenced during 2023, 2024 and 2025. Specifically the study targeted to elucidate the genomic diversity and evolutionary trajectory of recent MPXV outbreak strains through phylogenetic analysis, to construct a pangenome for these strains to characterize the core and accessory genome; and to identify lineage-specific diversification and highly conserved genomic regions that may represent potential targets for the development of antiviral therapeutics and vaccines. The phylogenetic tree can be used to study viral evolution, transmission, and outbreak dynamics. The sequences are organized by genetic similarity and help to identify lineages, sublineages and variants and also allow identifying geographically or temporally related clusters. A pangenome can be constructed to understand the proteome of MPox variants. This allows identifying core proteins (present in all), accessory proteins (present in some) and the unique genes which may be strain-specific.

Materials and Methods

Complete MPXV genomes sequenced from 31st December 2023 to 12th April 2025 (197 sequences), were retrieved from the NCBI virus database (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/). All genome FASTA sequences were screened for the presence of ambiguous bases (N). A custom Bash script was used to remove any genome that contained ambiguous bases in the genome sequence. Only genomes with zero ambiguous bases were retained for downstream annotation and pangenome analysis to ensure the highest possible sequence quality.

The sequences used for the study, included 118 global samples from the USA, Germany, Australia, South Africa, China, Japan, Kenya, Nigeria, Thailand and the Democratic Republic of Congo. Phylogenetic analysis was performed using NCBI tree viewer incorporating two reference sequences: NC_003310 (Central African) and NC_063383 (West African).

Structural annotation of genomes of all the 118 sequences and the 2 reference sequences were done using MAKER Annotation pipeline (Cantarel et al., 2008). The *E. coli* expressed sequence tags (EST) was used as EST evidence and Uniprot/Swissprot was used as protein homology evidence. The Generic Feature Format (GFF) files generated by MAKER were used for pangenome analysis.

Pangenome analysis was conducted using Roary v3.13.0 (Page et al., 2015), a high-speed pipeline for large scale prokaryotic pan-genome analysis. Pangenome was constructed from the two Mpox reference genomes (NC_003310.1 and NC_063383.1) and the 118 novel sequences. The core genome was defined as genes present in 99–100% of isolates.

Results

The phylogenetic tree (Figure 2) is rooted with the earliest known strain (Zaire 1996 - NC_003310.1) at the base and producing two major clusters. Central African Lineage (Congo Basin) represents older and more diverse isolates. NC_063383.1 (Nigeria 2018) sequence clustered with the bulk of the other sequences, indicating that the majority of the genomes belong to the West African lineage (Clade II). The isolates form distinct clades, reflected both geographical origins and sampling times.

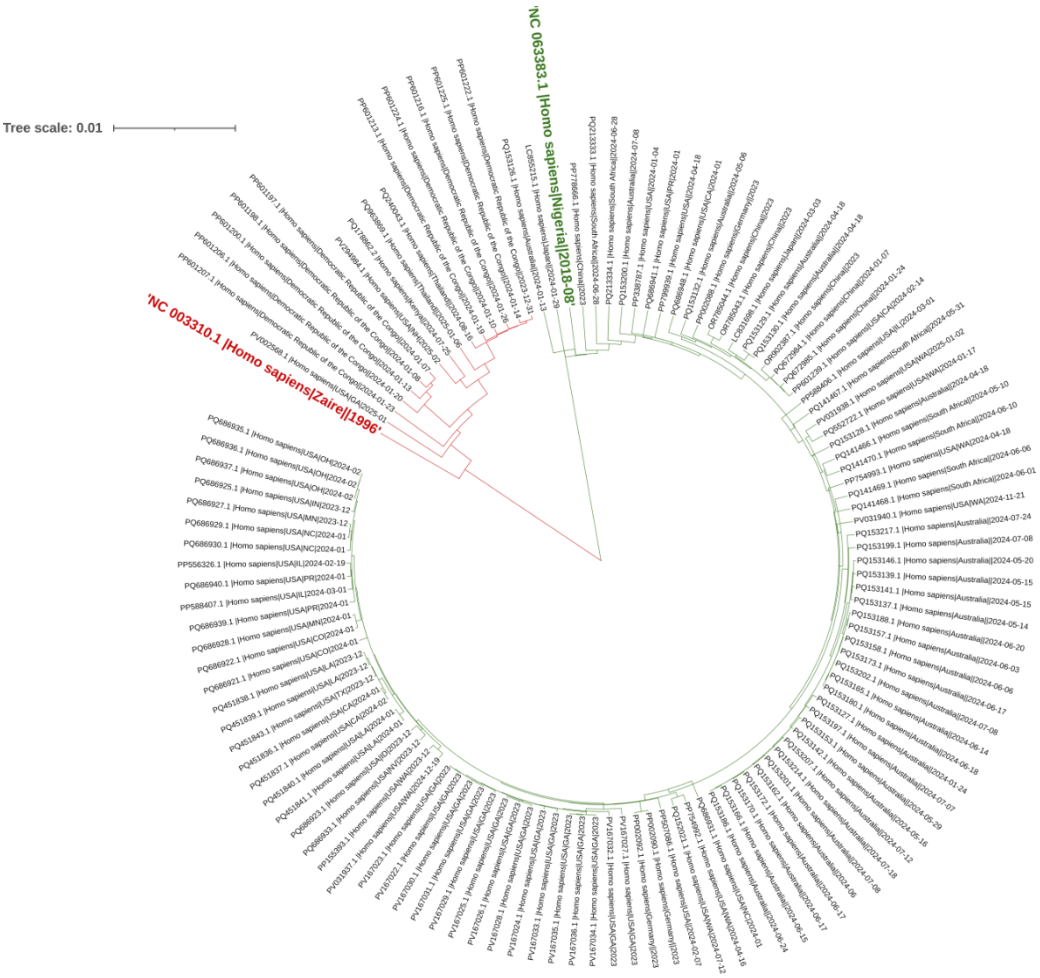

Figure 2: *Phylogenetic tree constructed using the novel MPox sequences and the two reference sequences: (NC_003310 and NC_063383). The tree is rooted with the earliest known strain (Zaire 1996 - NC_003310.1) at the base by producing two major clusters.*

Pangenome analysis between the two Mpox reference genomes (NC_003310.1 and NC_063383.1) resulted 189 core genes out of 200 total genes present in both genomes. A total of 11 shell genes (present in 15–94% of strains) were found out of which 5 genes were only found in the NC_003310.1 reference sequence representing Central African clade and 6 genes were only found in the NC_063383.1 reference sequence representing West African clade. Considering the pangenome of the novel 118 sequences, total 238 genes were found. This included 181 core genes (present in more than 98% of strains), 4 soft core genes (present in 95–98% of strains), 18 shell genes (present in 15–94% of strains) and 35 cloud genes (present in fewer than 15% of strains).

Discussion

According to the phylogenetic tree (Figure 2), NC_003310.1 (Zaire 1996) and its related sequences form a distinct cluster including sequences primarily from the Democratic Republic of the Congo (DRC). These isolates have longer branch lengths, indicating that they are genetically divergent from more recent outbreak strains.

The recent outbreak clade (West African lineage) was dominated by new isolates from USA, South Africa, Australia, and a few sequences from Germany, China, and Japan. The newer isolates (2024) formed tight clusters with short branches, indicating low mutation rates and possibly the ongoing transmission chains. The isolates from the USA showed strong clustering by state and sampling time, suggesting localized outbreaks or community transmission, with limited intermixing between states. The Australia Cluster showed tight clustering with isolates sequenced from May to July 2024. Isolates from South Africa, Germany, and China clustered within the broader global group. The phylogenetic tree provides evidence of multiple independent introductions of Mpox into various countries. Some countries, like the USA and Australia, exhibit patterns of localized transmission following these introductions.

The pangenome comparison between the two Mpox reference genomes (NC_003310.1 and NC_063383.1) revealed 189 core genes shared between them out of a total of 200 genes. Importantly, 5 genes were unique to the Central African clade, and 6 genes were limited to the West African clade, reflecting the distinct genetic makeup and evolutionary divergence of the two major Mpox lineages. The distinct proteins (shell genes) found in the two Mpox reference genomes highlight how each variant has uniquely adapted. The NC_003310.1 sequence has proteins that suppress the host's immune system and aid in viral DNA production. The NC_063383.1 possesses proteins that help with viral assembly and modify host immune cell function. These unique proteins identified in the two Mpox lineages may serve as potential candidates for drug targeting.

The pangenome of 118 novel Mpox genomes from 2024–2025, had a total gene count of 238, indicating the presence of additional genes not present in the reference strains, likely representing recent or strain-specific genetic changes. Among these, 181 genes were classified as core genes, which are highly conserved in the viral core proteome. These core genes likely encode essential functions and represent potential targets for broad-spectrum antiviral drugs or vaccines. The presence of 4 soft core genes, 18 shell genes and 35 cloud genes reflect variable regions of the Mpox genome, which may be under different evolutionary pressures or may arise from recombination, gene loss or horizontal gene acquisition events.

Conclusion

Integrated genomic and proteomic analyses of the Mpox virus reveal clear genomic diversity and evolutionary dynamics, distinguishing the two major clades (Central and West African), with a majority of recent global outbreak strains clustering within the West African lineage. Geographic and temporal sub-clustering in countries such as the USA and Australia indicates multiple introductions, while close genetic similarity among isolates reflects localized transmission and the slow evolutionary rate typical of DNA viruses. Pangenome analysis highlights divergence between clades, with conserved core genes and distinct lineage-specific genes, alongside novel accessory and cloud genes in 2024–2025 isolates that suggest ongoing diversification, possibly driven by host adaptation or selective pressures. The identification of 181 core genes underscores a stable proteomic backbone, offering conserved targets for antiviral and vaccine development, emphasizing the importance of sustained genomic surveillance to track viral evolution and emerging variants.

References

Ahmed, S. K., El-Kader, R. G. A., Abdulqadir, S. O., Abdullah, A. J., El-Shall, N. A., Chandran, D., Dey, A., Emran, T. B., & Dhama, K. (2023). Monkeypox clinical symptoms, pathology, and advances in management and treatment options: An update. *International Journal of Surgery*, *109*(9), 2837. https://doi.org/10.1097/JS9.0000000000000091

Antinori, A., Mazzotta, V., Vita, S., Carletti, F., Tacconi, D., Lapini, L. E., D'Abramo, A., Cicalini, S., Lapa, D., Pittalis, S., Puro, V., Rivano Capparuccia, M., Giombini, E., Gruber, C. E. M., Garbuglia, A. R., Marani, A., Vairo, F., Girardi, E., Vaia, F., … Vergori., A. (2022). Epidemiological, clinical and virological characteristics of four cases of monkeypox support transmission through sexual contact, Italy, May 2022. *Eurosurveillance*, *27*(22), 2200421. https://doi.org/10.2807/1560-7917.ES.2022.27.22.2200421

Araf, Y., Nipa, J. F., Naher, S., Maliha, S. T., Rahman, H., Arafat, K. I., Munif, M. R., Uddin, M. J., Jeba, N., Saha, S., Zhai, J., Hasan, S. M. N., Xue, M., Hossain, Md. G., & Zheng, C. (2024). Insights into the Transmission, Host Range, Genomics, Vaccination, and Current Epidemiology of the Monkeypox Virus. *Veterinary Medicine International*, *2024*(1), 8839830. https://doi.org/10.1155/2024/8839830

Bass, J., Tack, D. M., McCollum, A. M., Kabamba, J., Pakuta, E., Malekani, J., Nguete, B., Monroe, B. P., Doty, J. B., Karhemere, S., Damon, I. K., Balilo, M., Okitolonda, E., Shongo, R. L., & Reynolds, M. G. (2013). Enhancing health care worker ability to detect and care for patients with monkeypox in the Democratic Republic of the Congo. *International Health*, *5*(4), 237–243. https://doi.org/10.1093/inthealth/iht029

Brown, K., & Leggat, P. A. (2016). Human monkeypox: current state of knowledge and implications for the future. Tropical medicine and infectious disease, 1(1), 8.

Bunge, E. M., Hoet, B., Chen, L., Lienert, F., Weidenthaler, H., Baer, L. R., & Steffen, R. (2022). The changing epidemiology of human monkeypox—A potential threat? A systematic review. *PLOS Neglected Tropical Diseases*, *16*(2), e0010141. https://doi.org/10.1371/journal.pntd.0010141

Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A. S., & Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, *18*(1), 188–196. https://doi.org/10.1101/gr.6743907

Happi, C., Adetifa, I., Mbala, P., Njouom, R., Nakoune, E., Happi, A., Ndodo, N., Ayansola, O., Mboowa, G., Bedford, T., Neher, R. A., Roemer, C., Hodcroft, E., Tegally, H., O'Toole, Á., Rambaut, A., Pybus, O., Kraemer, M. U. G., Wilkinson, E., … Oliveira, T. de. (2022). Urgent need for a non-discriminatory and non-stigmatizing nomenclature for monkeypox virus. *PLOS Biology*, *20*(8), e3001769. https://doi.org/10.1371/journal.pbio.3001769

Nuzzo, J. B., Borio, L. L., & Gostin, L. O. (2022). The WHO Declaration of Monkeypox as a Global Public Health Emergency. *JAMA*, *328*(7), 615–617. https://doi.org/10.1001/jama.2022.12513

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics (Oxford, England)*, *31*(22), 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

World Health Organization. (2025). *Global mpox trends*.

Zumla, A., Rosenthal, P. J., Sam-Agudu, N. A., Ogoina, D., Mbala-Kingebeni, P., Ntoumi, F., Nakouné, E., Njouom, R., Ndembi, N., Mills, E. J., Muyembe-Tamfum, J.-J., & Nachega, J. B. (2025). The 2024 Public Health Emergency of International Concern: A Global Failure to Control Mpox. *The American Journal of Tropical Medicine and Hygiene*, *112*(1), 17–20. https://doi.org/10.4269/ajtmh.24-0606