bioRxiv preprint doi: https://doi.org/10.1101/2022.07.30.502168; this version posted August 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A distinct phylogenetic cluster of Monkeypox genomes suggests an early and cryptic spread of the virus

Bani Jolly[1,2], Vinod Scaria[1,2,#]

[1]CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Mathura Road, New Delhi, India
[2]Academy for Scientific and Innovative Research, Human Resource Development Centre Campus, Ghaziabad, Uttar Pradesh, India

[#]Address for correspondence
Vinod Scaria, G. N. Ramachandran Knowledge Centre for Genome Informatics, CSIR-Institute of Genomics and Integrative Biology, Sukhdev Vihar, New Delhi, India, 110025, Email: vinods@igib.in,

Keywords: Monkeypox, 2022 outbreak, lineage A.2, lineage B.1

Abstract
Since its first reports in humans in 1970, monkeypox has been predominantly restricted to countries in Africa where the disease is endemic. Early in 2022, a large number of cases of the disease were reported from Europe and other countries in patients having no history of travel to regions where monkeypox is endemic. Amid a rise in cases, the availability of genome sequences of monkeypox virus isolates in the public domain provides an opportunity to understand the transmission and evolution of the virus. Here, we describe a distinct phylogenetic cluster of monkeypox virus (lineage A.2) using genome sequences available on GISAID. Lineage A.2 currently encompasses 9 genome sequences from 6 viral isolates collected from 3 countries and is distinctly different from the predominant lineage B.1 which is linked to the large European outbreak.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.30.502168; this version posted August 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Monkeypox, a rare viral infection, has been predominantly restricted to countries in Africa and was infrequently reported from regions where the disease is not endemic, until recently. Early in 2022, an outbreak of the disease has been reported and has been epidemiologically linked to potential superspreader events in Europe earlier in the year[1]. Over a short period, the disease has spread across 75 countries causing over 20,000 infections, as on 30 July 2022, with the World Health Organization subsequently declaring the outbreak as a public health emergency of international concern in July 2022[2]. Following the outbreak, the wide availability of monkeypox genome sequences in the public domain provided a unique opportunity to understand the genetic epidemiology as well as the evolution of the pathogen.

With heightened surveillance and molecular diagnosis across the globe, several genomes of the monkeypox virus are being deposited in public databases like GISAID[3]. As early epidemiologic studies linked the outbreak of monkeypox in 2022 to superspreader event(s) in Europe, it was therefore not surprising that a majority of the genomes available on GISAID cluster together. Further reports suggest that the early transmission in this outbreak was largely among gays, bisexuals and other men who have sex with men (MSM) with some exceptions[4,5]. However, the recently deposited genome sequences from the United States of America, Thailand and much more recently from India suggest a distinctly different phylogenetic cluster of genomes, classified by Nextclade as lineage A.2, in contrast to the cluster encompassing the majority of genomes (N=547) that are classified as lineage B.1 (**Figure 1A**)[6]. The A.2 lineage comprises 9 genomes from 6 unique clinical isolates, with the earliest genome belonging to this lineage (collected in July 2021) being deposited from the state of Texas, USA. The two genome isolates from India cluster closely with a genome isolate from Florida (hMpxV/USA/FL-DHCPPCDC-001/2022) on the phylogenetic tree, while the isolates from Texas, Virginia and Thailand mapped to separate sub-clusters (**Figure 1B**).

The genomes belonging to the A.2 lineage have 16 distinct genetic variations which are not found in other lineages, of which 9 are nonsynonymous, 3 are synonymous and 1 is a stopgain variation, along with a deletion of 3-amino acid deletion in the gene OPG174 (**Figure 1C**). Variants found at a minimum frequency of 90% in the respective lineages A.2 and B.1 were compared and are summarised in **Figure 1C**.

Albeit the limited number of sequences available for the A.2 lineage, we attempted to compute the time to the most recent common ancestor (tMRCA) for A.2 and the nucleotide substitution rates for lineages A.2 and B.1 using BEAST v1.10.4[7]. The tMRCA was calculated following a coalescent growth rate model with a strict molecular clock and the HKY+Γ substitution model. MCMC was run for 50 million steps and the initial 1% steps were discarded as burn-in. The tMRCA of the A.2 lineage was computed as 25 June 2021 (95% HPD 19 March 2021 to 1 July 2021). The A.2 lineage had a mean nucleotide substitution rate of 5.53×10^{-5} (95% HPD 3.39×10^{-5} to 7.46×10^{-5}) substitutions per base/year, suggesting a modest rate of substitution compared to the substitution

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.30.502168; this version posted August 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

rate of 1.13×10^{-4} (95% HPD 9.33×10^{-5} to 1.33×10^{-4}) substitutions per base/year for the larger B.1 lineage of genomes. Accelerated evolution of the B.1 lineage has been observed recently[8].

Limited demographic information could be linked to the members of the A.2 cluster and has been primarily compiled from the metadata associated with the genome sequences as well as reports in the public domain. The two genomes from Kerala, India were isolated from men who had a travel history to the United Arab Emirates, while the genome from Thailand was isolated from a male traveller from Nigeria. The genome from Texas, United States of America, the earliest in the cluster, was also isolated from a male traveller from Nigeria suggesting a wider geographic area with ongoing transmission of the virus beyond regions in Central Africa where the virus is endemic[9].

Put together, the evidence suggests that this unique and distinct phylogenetic cluster of genomes, therefore, represents sustained and previously uncharacterized human-human transmission events spanning multiple countries. The tMRCA dating to mid-2021 and the earliest genome dating to July 2021 suggests that this sustained transmission event possibly preceded the outbreak in 2022 in Europe and has remained largely undetected. The distinct genomic signatures suggest that this transmission chain may not be linked to the large outbreak of monkeypox which occurred in 2022 and has been potentially uncovered due to heightened awareness, surveillance and the wider availability of diagnostics.

This report, therefore, re-affirms the unique and significant value of genomic surveillance of emerging pathogens in uncovering potential new insights and leads for epidemiological investigations. The distinctive finding in this report may have a significant impact on public health policies, surveillance as well as public-health communication.

Declaration of Competing Interest

The authors report no potential conflicts of interest.

Acknowledgement

BJ acknowledges research fellowships from the Council of Scientific and Industrial Research (CSIR), India. The authors acknowledge the data deposited in the public domain by the investigators. A detailed list of acknowledgements is available at https://github.com/banijolly/Phylovis-MPX.

Funding

This work was supported by the Council of Scientific and Industrial Research (CSIR), India. The funders had no role in the analysis of data, preparation of the manuscript or decision to publish.

References

1. Investigation into monkeypox outbreak in England: technical briefing 1. GOV.UK. Available at https://www.gov.uk/government/publications/monkeypox-outbreak-technical-briefings/investigation-into-monkeypox-outbreak-in-england-technical-briefing-1. Accessed July 31, 2022
2. WHO Director-General declares the ongoing monkeypox outbreak a Public Health Emergency

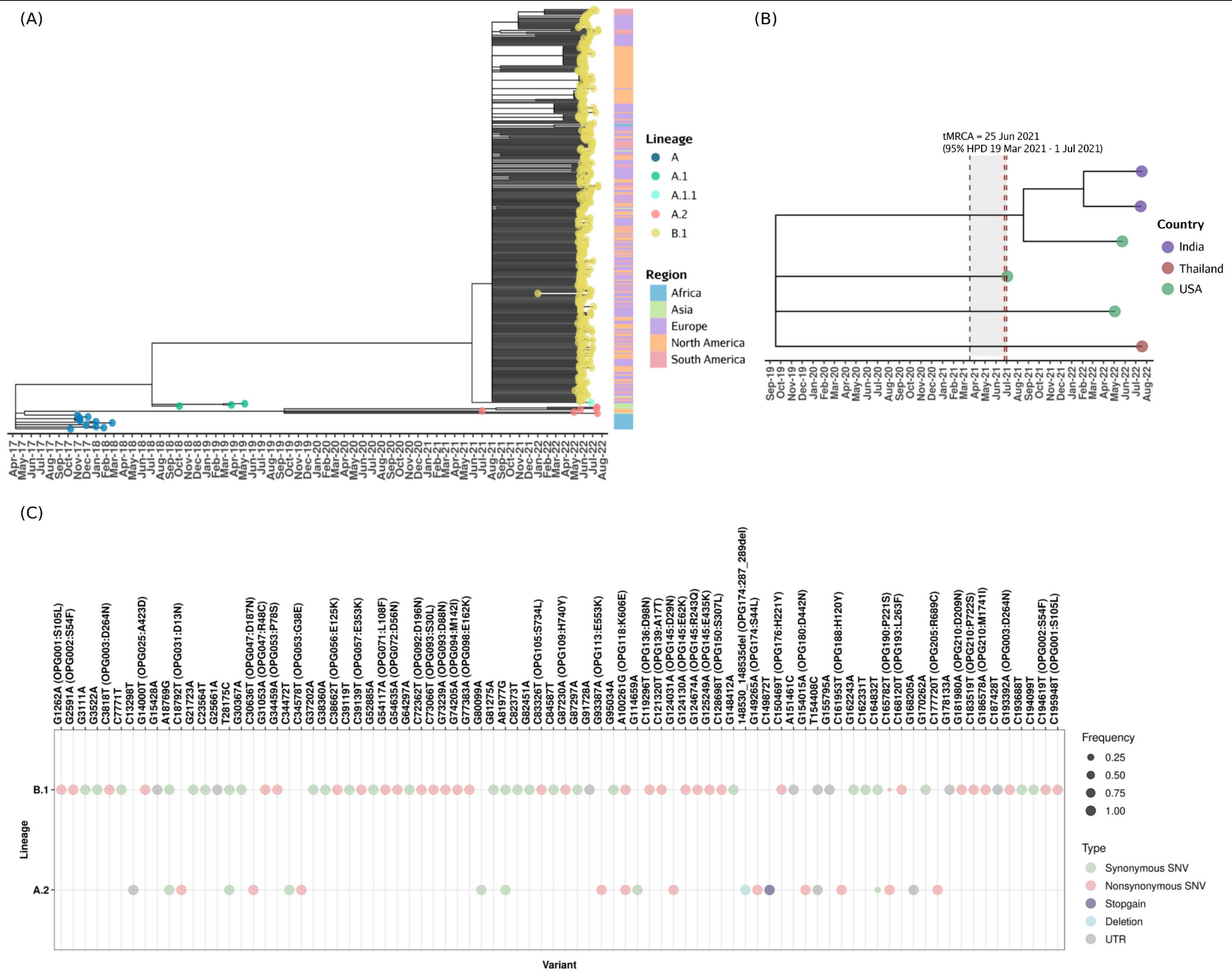
bioRxiv preprint doi: https://doi.org/10.1101/2022.07.30.502168; this version posted August 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of International Concern. Available at https://www.who.int/europe/news/item/23-07-2022-who-director-general-declares-the-ongoing-monkeypox-outbreak-a-public-health-event-of-international-concern. Accessed July 30, 2022

3. Shu Yuelong, McCauley John. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* 2017;**22**(13). Doi: 10.2807/1560-7917.ES.2017.22.13.30494.
4. World Health Organization. 2022 Monkeypox Outbreak: Global Trends. Available at https://worldhealthorg.shinyapps.io/mpx_global/. Accessed July 30, 2022
5. Thornhill John P, Barkati Sapha, Walmsley Sharon, Rockstroh Juergen, Antinori Andrea, Harrison Luke B, et al. Monkeypox Virus Infection in Humans across 16 Countries - April-June 2022. *N Engl J Med* 2022. Doi: 10.1056/NEJMoa2207323.
6. Hadfield James, Megill Colin, Bell Sidney M, Huddleston John, Potter Barney, Callender Charlton, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018;**34**(23):4121–3.
7. Suchard Marc A, Lemey Philippe, Baele Guy, Ayres Daniel L, Drummond Alexei J, Rambaut Andrew. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 2018;**4**(1):vey016.
8. Isidro Joana, Borges Vítor, Pinto Miguel, Sobral Daniel, Santos João Dourado, Nunes Alexandra, et al. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. *Nat Med* 2022:1–4.
9. Rao Agam K, Schulte Joann, Chen Tai-Ho, Hughes Christine M, Davidson Whitni, Neff Justin M, et al. Monkeypox in a Traveler Returning from Nigeria — Dallas, Texas, July 2021. *MMWR Surveill Summ* 2022;**71**(14):509.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.30.502168; this version posted August 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1. (A) Phylogenetic tree of genome isolates from GISAID belonging to the hMPXV-1 clade of monkeypox virus constructed using Nextstrain[6]. **(B)** Phylogenetic tree of the genomes belonging to A.2 monkeypox lineage. **(C)** Comparison of frequencies of variations present in >90% genomes of lineages A.2 (N=6) and B.1 (N=547).


(A)
Lineage
A
A.1
A.1.1
A.2
B.1
Region
Africa
Asia
Europe
North America
South America
(B)
tMRCA = 25 Jun 2021
(95% HPD 19 Mar 2021 - 1 Jul 2021)
Country
India
Thailand
USA
(C)
Lineage
B.1
A.2
Variant
Frequency
0.25
0.50
0.75
1.00
Type
Synonymous SNV
Nonsynonymous SNV
Stopgain
Deletion
UTR