

Multi-country outbreak of Monkeypox virus: genetic divergence and first signs of microevolution

vborges

May 23

Joana Isidro¹, Vítor Borges¹, Miguel Pinto¹, Rita Ferreira¹, Daniel Sobral¹, Alexandra Nunes¹, João Dourado Santos¹, Verónica Mixão¹, Daniela Santos², Silvia Duarte², Luís Vieira², Maria José Borrego³, Sofia Núncio⁴, Ana Pelerito⁴, Rita Cordeiro⁴, João Paulo Gomes^{1,*}.

¹ Genomics and Bioinformatics Unit, Department of Infectious Diseases, National Institute of Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal

² Technology and Innovation Unit, Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal

³ National Reference Laboratory of Sexually Transmitted Infections, Department of Infectious Diseases, National Institute of Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal

⁴ Emergency and Biopreparedness Unit, Department of Infectious Diseases, National Institute of Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal

contact: j.paulo.gomes@insa.min-saude.pt

Following the ([First draft genome sequence of Monkeypox virus associated with the suspected multi-country outbreak, May 2022 \(confirmed case in Portugal\)](#)), we now release 9 additional genome sequences of Monkeypox virus causing a multi-country outbreak. These sequences were obtained from clinical specimens collected from 9 patients on May 15th and 17th, 2022 through high throughput shotgun metagenomics using Illumina technology (see details bellow), with depth of coverage throughout Monkeypox genome ranging from 38x to 508x (mean of 201x).

The rapid integration of the newly sequenced genomes into the Monkeypox genetic diversity, also including the sequence released by USA* (Gigante et al, [Monkeypox virus isolate MPXV_USA_2022_MA001, complete genome - Nucleotide - NCBI](#)), allow us to raise the following main observations:

- **The multi-country outbreak most likely has a single origin**, with all sequenced viruses released so far* tightly clustering together (Figure 1).
- **Confirmation of the phylogenetic placement unveiled by the first draft sequence Isidro et al**, : the outbreak virus belongs to the West African clade and is most closely related to viruses (based on available genome data) associated with the exportation of monkeypox virus from Nigeria to several countries in 2018 and 2019, namely the United Kingdom, Israel and Singapore (1, 2).

[Skip to main content](#)

- Still, **the outbreak virus diverges a mean of 50 SNPs from those 2018-2019 viruses** (46 SNPs from the closest reference MPXV_UK_P2, MT903344.1) ([↓ Table 1_2022-05-23.zip](#) (15.0 KB)), which is **far more than one would expect considering the estimated substitution rate for Orthopoxviruses** (3).
- As also mentioned by Rambaut ([Discussion of on-going MPXV genome sequencing](#)), one cannot discard the **hypothesis that the divergent branch results from an evolutionary jump** (leading to a hypermutated virus) caused by APOBEC3 editing (4)
- We have already detected the **first signs of microevolution within the outbreak cluster, namely the emergence of 7 SNPs** ([↓ Table 2_2022-05-23.zip](#) (10.9 KB)), leading to 3 descendant branches (Figure 1) including a further sub-cluster (supported by 2 SNPs) involving 2 sequences (PT0005 and PT0008). Notably, these two sequences **also share a 913bp frameshift deletion** in MPXV-UK_P2-010 gene coding for an Ankyrin/Host Range (Bang-D8L); D7L protein (MT903344.1 annotation). Gene loss events were already observed in the context endemic Monkeypox circulation in Central Africa, being hypothesized to correlate with human-to-human transmission (5).
- This microevolution scenario also suggests that **genome sequencing might provide enough resolution to track the virus dissemination in the context of the current outbreak** (which could seem implausible for a dsDNA virus).]

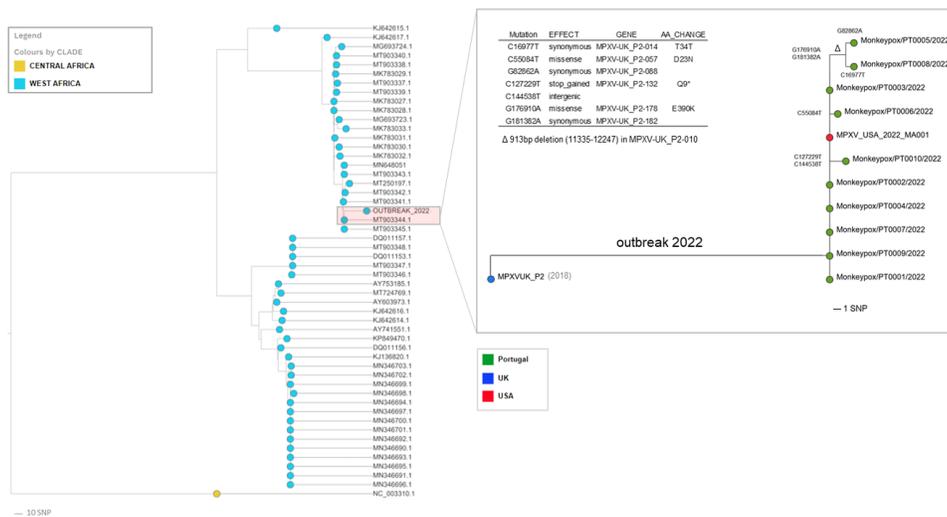


Figure 1. Draft phylogenetic analysis of Monkeypox viral sequences, **highlighting the diversity within the outbreak cluster.**

“* An additional genome ITM_MPX_1_Belgium was shared by Selhorst et al (ITM) ([Belgian case of Monkeypox virus linked to outbreak in Portugal](#)), but is not currently included in the tree as the sequence most likely has technical issue that might render phylogenetic misplacement.

All sequences from Portugal can be downloaded here

[↓ 2022-05-23_Monkeypox_PT_sequences.zip](#) (505.7 KB). This dataset includes a further curated version (v2) of the first sequence (Monkeypox/PT0001/2022) and 9 new sequences [Skip to main content](#))22- Monkeypox/PT0010/2022).

Brief description of the methods

In order to deplete host DNA, skin exudate samples were subjected to sonication and DNase/RNase cocktail treatment, prior to DNA extraction with the QIAamp DNA Mini kit (protocol for tissues) (Qiagen). After Nextera XT library preparation, shotgun metagenomics was performed by paired-end sequencing (2x150 bp) on an Illumina NextSeq 2000 apparatus, with about ~40M total reads per sample. Five out of the nine samples were also subjected to ONT MinION sequencing, with generated data being also used to curate the sequences.

Reads were human-depleted and subsequently mapped to the reference MPXV_UK_P2 (MT903344.1; also being used as reference in the newest Monkeypox nextstrain build [auspice](#)) using the INSaFLU platform (<https://insaflu.insa.pt/>). Genome sequences were further curated based on *de novo* assemblies and careful inspection of mutations. Submission of reads to ENA is ongoing.

1. Mauldin MR, McCollum AM, Nakazawa YJ, et al. *Exportation of Monkeypox Virus From the African Continent*. J Infect Dis. 2022; 225(8):1367-1376. doi: 10.1093/infdis/jiaa559.
2. Yinka-Ogunleye A, Aruna O, Dalhat M, et al. *Outbreak of human monkeypox in Nigeria in 2017–18: a clinical and epidemiological report*. Lancet Infect Dis 2019; 19:872–9. doi: 10.1016/S1473-3099(19)30294-4.
3. Firth C, Kitchen A, Shapiro B, et al. *Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses*. Mol Biol Evol. 2010; 27(9):2038-51. doi: 10.1093/molbev/msq088.
4. Pecori R, Di Giorgio S, Paulo Lorenzo J, Nina Papavasiliou F. *Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination*. Nat Rev Genet. 2022; 7:1–14. doi: 10.1038/s41576-022-00459-8.
5. Kugelman JR, Johnston SC, Mulembakani PM, et al. *Genomic variability of monkeypox virus among humans, Democratic Republic of the Congo*. Emerg Infect Dis. 2014; 20(2):232-9. doi: 10.3201/eid2002.130118.

[🔗 First German genome sequence of Monkeypox virus associated to multi-country outbre...](#)

[🔗 Illumina whole-genome sequence of Monkeypox virus in a patient travelling from the Ca...](#)

[🔗 Initial observations about putative APOBEC3 deaminase editing driving short-term evol...](#)

[🔗 Second draft genome from Spain of the Monkeypox virus 2022 outbreak](#)

[🔗 First 2022 monkeypox outbreak genome sequences from Slovenia](#)

[1 more](#)

last visit

[arambaut](#)  ARTIC Network

May 23

Thanks for posting your new genomes here. It is interesting that all of these additional SNPs

[Skip to main content](#) ^r TC → TT or GA → AA in dinucleotide context. This would support the idea

that most of the single nucleotide mutations we see in these genomes are the results of host enzyme editing rather than polymerase replication errors:

Discussion of on-going MPXV genome sequencing

For the sites where both MA001 and PT0001 differ from UK_P3 (and where ITM_MPX_1 has called a N) there is a very strong bias in what the difference is: G → A : 26 (14 non-synonymous, 12 synonymous) C → T: 13 (9 non-synonymous, 4 synonymous) G → T: 1 (non-synonymous) A → C: 1 (non-coding region) The G → A (and C→T on the other strand) could suggest APOBEC editing? Or some artefact of sequencing? It would be interesting to know why ITM called an N for these sites? Is it because they were amb...

vborges

May 23

Thanks. Indeed, there seems to be a pattern. The extra C187169T mutation in the (just) released sequence from Germany (MPXV-BY-IMB25241; [Monkeypox virus isolate MPXV-BY-IMB25241, complete genome - Nucleotide - NCBI](#)) is also a TC → TT.

vborges

May 24

The first 10 genome consensus sequences collected in Portugal are now also available at NCBI, under the following accession numbers. Read data was submitted to European Nucleotide Archive (ENA) and is in process to be released.

Sequence	NCBI accession number	ENA accession number
Monkeypox/PT0001/2022	ON585029	ERR9769166
Monkeypox/PT0002/2022	ON585030	ERR9769168
Monkeypox/PT0003/2022	ON585031	ERR9769169
Monkeypox/PT0004/2022	ON585032	ERR9769170
Monkeypox/PT0005/2022	ON585037	ERR9769171 / ERR9769167
Monkeypox/PT0006/2022	ON585033	ERR9769172
Monkeypox/PT0007/2022	ON585034	ERR9769173
Monkeypox/PT0008/2022	ON585038	ERR9769174
Monkeypox/PT0009/2022	ON585035	ERR9769175
Monkeypox/PT0010/2022	ON585036	ERR9769176

anekrut

May 30

[Skip to main content](#)

[@vborges](#) do you know when the raw read data will become available. The datasets are listed in NCBI and EBI but no actual data is available

vborges

May 30

[@anekrut](#) , we have submitted the reads to ENA 6 days ago, asking for an immediate release. Unfortunately, ENA is taking more than we would expect. We will contact ENA reinforcing if it is possible to accelerate the process.

vborges

May 30

update: ENA has just replied saying that they are facing some technical issues that prevented the browser from indexing new data. The issue is expected to be solved today or early tomorrow.

anekrut

May 30

Thank you! I think proving raw data is exceptionally important for verifying assemblies, assessing the extent of intra-host variation, detection of multiple infections etc. etc... Unfortunately for the most part it seems that MPX data will follow SARS-CoV-2 path with very little raw data making its way to the public domain. Thank you for making the difference!