nature medicine

https://doi.org/10.1038/s41591-022-01907-y

Accelerated Article Preview

Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus

Received: 27 May 2022

Accepted: 20 June 2022

Accelerated Article Preview
Published online: 24 June 2022

Cite this article as: Isidro, J. . et al. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. *Nature Medicine* https://doi.org/10.1038/s41591-022-01907-y (2021).

Joana Isidro, Vítor Borges, Miguel Pinto, Daniel Sobral, João Dourado Santos, Alexandra Nunes, Verónica Mixão, Rita Ferreira, Daniela Santos, Silvia Duarte, Luís Vieira, Maria José Borrego, Sofia Núncio, Isabel Lopes de Carvalho, Ana Pelerito, Rita Cordeiro & João Paulo Gomes

Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus

Joana Isidro[1,*], Vítor Borges[1,*], Miguel Pinto[1,*], Daniel Sobral[1], João Dourado Santos[1], Alexandra Nunes[1], Verónica Mixão[1], Rita Ferreira[1], Daniela Santos[2], Silvia Duarte[2], Luís Vieira[2], Maria José Borrego[3], Sofia Núncio[4], Isabel Lopes de Carvalho[4], Ana Pelerito[4], Rita Cordeiro[4], João Paulo Gomes[1, 5, †].

[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 Response and Biopreparedness Unit, Department of Infectious Diseases, National Institute of Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal
[5] Faculty of Veterinary Medicine, Lusófona University, Lisbon, Portugal

*These authors contributed equally to this work.

†E-mail: j.paulo.gomes@insa.min-saude.pt

Abstract / Introductory paragraph

The largest monkeypox virus (MPXV) outbreak described so far in non-endemic countries was identified in May 2022[1-6]. Here, shotgun metagenomics allowed the rapid reconstruction and phylogenomic characterization of the first MPXV outbreak genome sequences, showing that this MPXV belongs to clade 3 and that the outbreak most likely has a single origin. Although 2022 MPXV (lineage B.1) clustered with 2018-2019 cases linked to an endemic country, it segregates in a divergent phylogenetic branch, likely reflecting continuous accelerated evolution. An in-depth mutational analysis suggests the action of host APOBEC3 in viral evolution as well as signs of potential MPXV human adaptation in ongoing microevolution. Our findings also indicate that genome sequencing may provide resolution to track the spread and transmission of this presumably slow-evolving dsDNA virus.

Main

Monkeypox is a rare zoonotic disease that is caused by the monkeypox virus (MPXV) from the Orthopoxvirus genus, which includes the variola virus, the causative agent of smallpox[1,2,3]. With an incubation period from 5 to 21 days, human disease typically begins with fever, myalgia, fatigue and headache, often followed by maculopapular rash at the site of primary infection that can spread to other parts of the body[1]. Although the natural reservoir of MPXV remains unknown, animals such as rodents and non-human primates may harbour the virus, leading to occasional spill-over events to humans[1,2,3]. MPXV is endemic in West and Central African countries and the rare reports outside these regions have been associated with imports from those endemic countries[1,2,3,4]. We are now facing the first multi-country outbreak without known epidemiological links to West or Central Africa[1], with more than 2500 confirmed cases reported worldwide, as of June 18[th 5,6], since the first confirmed case on 7 May 2022, in the United Kingdom[4]. Several measures are being recommended by international health authorities to contain MPXV transmission[1], including the use of vaccines for selected close contacts of monkeypox patients (post-exposure) and for groups at risk of occupational exposure to monkeypox (pre-exposure)[7]. The virus can be

transmitted from human to human by close contact with lesions, body fluids, respiratory droplets and contaminated materials[1,3], but the current epidemiological context poses some degree of uncertainty about the viral transmission dynamics and outbreak magnitude.

International sequencing efforts immediately began to characterize the outbreak-causing MPXV, in order to identify its origin and track its dissemination. Genome data will also inform about the virus evolutionary trajectory, genetic diversity and phenotypic characteristics with relevance for guiding diagnostics, prophylaxis and research. Here, we report the rapid application of high-throughput shotgun metagenomics to reconstruct the first genome sequences of the MPXV associated with the 2022 MPXV outbreak, providing valuable genomic and phylogenetic data on this emerging threat.

In order to rapidly get the first insights on phylogenetic placement and evolutionary trends of the 2022 outbreak-causing MPXV we focused our analysis on a first outbreak-related MPXV genome sequence, publicly released on May 20th, 2022 by Portugal[8], as well as on additional sequences released on NCBI before the 27th of May 2022, with 15 sequences in total (most of them from Portugal) (Supplementary Table 1 and 2). The rapid integration of the first sequence into the global MPXV genetic diversity (Figure 1) confirmed that the 2022 outbreak virus belongs to the MPXV clade 3 (within the formerly designated "West African" clade, which also includes clade 2)[9]. MPXV from clade 2 and 3 are most commonly reported from western Cameroon to Sierra Leone and usually carries a <1% case-fatality ratio (CFR), in contrast with viruses from the clade 1 (formerly designated as "Central African" or "Congo Basin" clade)[9], which are considered more virulent with a CFR >10%[10,11]. All outbreak MPXV strains sequenced so far tightly cluster together (Figure 1), suggesting that the ongoing outbreak has a single origin. The 2022 outbreak cluster (lineage B.1)[9] forms a divergent branch descendant from a branch with viruses (lineage A.1)[9] associated with the exportation of MPXV in 2018 and 2019 from an endemic country (Nigeria) to the United Kingdom, Israel and Singapore[12,13], with genetic linkage to a large outbreak occurring in Nigeria in 2017-2018[13] (Figure 1). Given these findings and the MPXV historical epidemiology (rare cases in non-endemic countries), it is likely that the emergence of the 2022 outbreak resulted from importation(s) of this MPXV from an endemic country, with the MPXV detected in 2022 potentially representing the continuous circulation and evolution of the virus that caused the 2017-2018 Nigeria outbreak. The recent release of a MPVX sequence from a 2021 travel-associated case from Nigeria to the USA (USA_2021_MD; accession ON676708)[14] phylogenetically placed between 2018-2019 and 2022 sequences (Figure 1) is aligned with such hypothesis. We cannot however exclude the hypothesis of a prolonged period of cryptic dissemination in humans or animals in a non-endemic country (e.g., after the reported 2018-2019 importations). Silent human-to-human transmission (e.g., due to underdiagnosis) seems less likely considering the known disease characteristics of the affected individuals, usually involving localised or generalised skin lesions[1]. Cryptic transmission in an animal host in a non-endemic country followed by a recent spill-over event is another hypothesis, even though, again, this would be somehow surprising as such scenario has never been reported. Altogether, current data points for a scenario of more than one introduction from a single origin, with superspreader event(s) (e.g., saunas used for sexual encounters) and travel abroad likely triggering the rapid worldwide dissemination[15,16]. Considering the expected incubation period of 5-21 days[3], limited sampling (including limited viral genotyping data for the first confirmed cases in 2022) and the fact that multiple cases were confirmed in several countries in a three week period[1] following a first report on May 7th by the UK[1], the identification of the index cases associated with such presumable several introductions can be challenging. For example, although the first confirmed case has been hypothesized as the index of the outbreak (due to travel from Nigeria to the UK on 3-4 May 2022[1,3]), this scenario can be discarded as the earliest symptom onset dates for confirmed cases in Portugal and in the UK were in late April[15,16].

Notably, the 2022 MPXV diverges from the related 2018-2019 viruses by a mean of 50 SNPs (Figure 1 and 2), which is far more (roughly 6-12 fold more) than one would expect considering previous estimates of the substitution rate for Orthopoxviruses (1-2 substitutions per site per year)[17]. Such a divergent branch might represent accelerated evolution. Of note, among the 46 SNPs (24 non-synonymous, 18 synonymous, 4 intergenic) (Supplementary Table 3) separating the 2022 MPXV outbreak virus from the reference sequence (MPXV-UK_P2, 2018; GenBank accession #MT903344.1), 3 amino acid changes (D209N, P722S, M1741I) occurred in the immunogenic surface glycoprotein B21 (MPXV-UK_P2-182)[18]. Serological studies have previously indicated that the monkeypox B21 protein might be an important antibody target with several key immunodominant epitopes[18]. As discussed previously[19], fine inspection of the mutation profile of those 46 SNPs further revealed a strong mutational bias, with

26 (14 non-synonymous, 10 synonymous, 2 intergenic) and 15 (9 non-synonymous, 16 synonymous) being GA>AA and TC>TT nucleotide replacements, respectively (Figure 2, Supplementary Table 3). A tool (https://github.com/insapathogenomics/mutation_profile) was built to rapidly screen these and others mutations profiles. The observed (hyper)mutation signature might suggest the potential action of apolipoprotein B mRNA-editing catalytic polypeptide-like 3 (APOBEC3) enzymes in the viral genome editing[20]. Also, MPXV are A:T rich, so a mutation bias leading to further incorporation of A/T suggests the action of a non-random mutational driver, such as APOBEC3. In fact, APOBEC3 enzymes can be upregulated in response to viral infection, being capable of inhibiting a wide range of viruses by introducing mutations through deaminase and deaminase-independent mechanisms[20,21]. In some circumstances (e.g., lower levels of deamination), APOBEC3-mediating mutations might not completely disrupt the virus, thus increasing the likelihood of producing hyper-mutated (but viable) variants with altered characteristics (e.g. HIV immune escape variants)[20,22]. The repertoire and level of APOBEC3 enzymes depend on the host species/tissue and different enzymes display different preferences for the nucleotide or motif (such as dinucleotides or tetranucleotides) to be mutated[20,23,24]. For instance, the GA>AA and TC>TT nucleotide replacements observed in the 2022 outbreak MPXV were also found to be the preferred mutational pattern of human APOBEC3A enzymes (expressed in keratinocytes and skin) during genetic editing of human papillomavirus (HPV) in HPV1a plantar warts and HPV16 precancerous cervical biopsies[25]. Whether the excess of mutations seen in the 2022 MPXV is a direct consequence of APOBEC3-mediated genome editing in the human host cannot be discerned at this stage. Also, the putative APOBEC3 effect on MPVX evolution augments the uncertainty regarding the 2022 outbreak origins and introductions, in addition to the complexity of the epidemiological context. This raises the need on future studies focusing on the weight of APOBEC3 in MPVX diversification. In particular, functional studies assessing whether this mutational driver triggers MPVX adaptive evolution towards altered phenotypic features, such as enhanced transmissibility, are warrant.

Further phylogenomic analysis revealed the first signs of microevolution of this virus during human-to-human transmission. Among the 15 outbreak sequences analysed here, we detected the emergence of 15 SNPs (8 non-synonymous, 4 synonymous, 2 intergenic and 1 stop gained) (Figure 2; Supplementary Table 4). Notably, all SNPs also follow the same mutational bias, including eight GA>AA (6 non-synonymous, 2 synonymous) and seven TC>TT (2 non-synonymous, 2 synonymous, 1 stop gained and 2 intergenic) nucleotide replacements. This further suggests a continuous action of APOBEC3 during MPXV evolution. Among the 7 phylogenetic branches directly descendant from the most recent ancestor of the MPXV outbreak strain (Figure 1), we identified a sub-cluster (supported by 2 SNPs) of two sequences (PT0005 and PT0008, each with an additional SNP) that also share a 913bp frameshift deletion in a gene coding for an Ankyrin/Host Range (MPXV-UK_P2-010). Although gene loss events are not unexpected for orthopoxviruses (for example, variola virus has most likely undergone reductive evolution[26]), these were previously observed in the context of endemic MPXV circulation in Central Africa, being hypothesized to correlate with human-to-human transmission[27].

Our data reveals additional clues of ongoing viral evolution and potential human adaptation. Most emerging SNPs in sequences from Portugal were not 100% fixed in the viral population (frequencies: 75%-95%), supporting the existence of viral intra-patient population diversity. Further inspection of minor intra-patient single nucleotide variants (iSNVs) in Illumina samples led to the validation of 11 non-synonymous minor iSNVs (across five samples), again most with the “APOBEC3 signature” (Supplementary Table 5). Notably, among the targeted viral transcripts, we highlight a few encoded proteins that are known to interact with host immune system, such as an MHC class II antigen presentation inhibitor[28], an IFN-alpha/beta receptor glycoprotein[29] and IL-1/TLR signaling inhibitor[30]. These and other proteins (Supplementary tables 3, 4 and 5) found to be targeted during the 2022 outbreak MPVX divergence and microevolution might constitute priority targets for future functional studies aiming to assess their potential role in adaptation.

In summary, our genomic and phylogenomic data provide insights into the evolutionary trajectory of the 2022 MPVX outbreak strain, and sheds light on potential mechanisms and targets of human adaptation. The observed accelerated evolution of this human MPVX, potentially driven by the APOBEC3 action, suggests that viral genome sequencing might provide sufficient resolution to track the transmission dynamics and outbreak spread, which seemed to be challenging for a presumably slow-evolving dsDNA virus. Together with the adopted strategy of real-time data sharing, this study may help guide novel outbreak control measures and subsequent research directions.

Acknowledgements

We thank Prof. Andrew Rambaut from the University of Edinburgh for launching the discussion on MPVX genomics at Virological.org, with emphasis for his first insights on potential APOBEC3-mediated genome editing. We also gratefully acknowledge the authors, originating and submitting laboratories of the genetic sequences released in GenBank. The acquisition of WGS-associated equipment used in this study (including the Illumina NextSeq 2000) was funded by the HERA project (Grant/2021/PHF/23776) supported by the European Commission through the European Centre for Disease Control, and partially funded by the GenomePT project (POCI-01-0145-FEDER-022184), supported by COMPETE 2020 – Operational Programme for Competitiveness and Internationalisation (POCI), Lisboa Portugal Regional Operational Programme (Lisboa2020), Algarve Portugal Regional Operational Programme (CRESC Algarve2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF), and by the Portuguese Science and Technology Foundation (FCT). This study was also supported by the ERINHA-Advance project (funding from the European Union's Horizon 2020 Research & Innovation program, grant agreement No. 824061) and also benefited from co-funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 773830: One Health European Joint Programme (EJP), in particular by co-funding JSD and VM post-doc fellowships and INSaFLU development. We also thank Miguel Pinheiro (iBiMED/Universidade de Aveiro) for his continuous support on updating INSaFLU platform and the Infraestrutura Nacional de Computação Distribuída (INCD) for providing computational resources for testing it. INCD was funded by FCT and FEDER under the project 22153-01/SAICT/2016. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author Contributions

J.I., V.B., M.P. and J.P.G, contributed to study design. R.C., A.P., I.L.C, S.N., M.J.B. and R.F. contributed to the collection and processing of clinical specimens. J.I., V.B., M.P., A.N., D.S., V.M. and J.D.S performed bioinformatics analysis. J.I., D.S., S.D., and L.V performed wet-lab sequencing procedures. J.I., V.B., M.P. and J.P.G were the main contributors for manuscript writing. All authors critically reviewed the manuscript for intellectual content, approved the final version of the manuscript for submission and agreed to be accountable for all aspects of the work

Competing interests

All the other authors have no competing interests.

Figures

Figure 1

Phylogenetic analysis of MPXV viral sequences associated with the 2022 worldwide outbreak. A. MPVX global phylogeny showing that the 2022 outbreak cluster (lineage B.1) belongs to clade 3. Clade and lineage are designated according to the nomenclature proposed by Happi et al., 2022[9]; B. Genetic diversity within the outbreak cluster including the 15 sequences analysed in this study (released on NCBI before May 27th of 2022). The deletion symbol (Δ) denotes a large deletion (11335-12247 in MPXV-UK_P2-010 gene) shared by sequences segregating in a small sub-cluster; C. Outbreak phylogenetic tree updated with sequences available in NCBI as of June 15th 2022 (provided during revision for more updated contextualization). The list of the sequences used in these phylogenetic analyses are detailed in Supplementary Table 2 and the alignments are provided as Supplementary Data.

Figure 2

SNPs characterizing the 2022 MPXV outbreak variant. Light coloured mutations represent the SNPs separating the MPXV 2022 outbreak cluster from MPXV_UK_P2 (MT903344.1) reference sequence (Supplementary Table 3). Dark coloured mutations represent the genetic diversity within the outbreak cluster (Supplementary Table 4).

References

1. European Centre for Disease Prevention and Control. Monkeypox multi-country outbreak – 23 May 2022.ECDC: Stockholm https://www.ecdc.europa.eu/en/publications-data/risk-assessment-monkeypox-multi-country-outbreak (2022).
2. World Health Organization. Monkeypox fact sheet. WHO: Geneva https://www.who.int/news-room/fact-sheets/detail/monkeypox (2019).
3. Bunge, E. et al. The changing epidemiology of human monkeypox-A potential threat? A systematic review. PLoS Negl Trop Dis, 16(2), e0010141 (2022).
4. Government of the United Kingdom. Infectious disease guidance: monkeypox. https://www.gov.uk/guidance/monkeypox (2022).
5. ECDC, Epidemiological update: Monkeypox multi-country outbreak - 25 May 2022 ECDC: Stockholm https://www.ecdc.europa.eu/en/news-events/epidemiological-update-monkeypox-multi-country-outbreak (2022)
6. Kraemer, M.U.G. et al. Tracking the 2022 monkeypox outbreak with epidemiological data in real-time. Lancet Infect Dis 2022, S1473-3099(22)00359-0 (2022).
7. WHO. Guidelines Vaccines and immunization for monkeypox. WHO: Geneva https://www.who.int/publications/i/item/who-mpx-immunization-2022.1 (2022).
8. Isidro, J. et al. First draft genome sequence of Monkeypox virus associated with the suspected multi-country outbreak, May 2022 (confirmed case in Portugal). https://virological.org/t/first-draft-genome-sequence-of-monkeypox-virus-associated-with-the-suspected-multi-country-outbreak-may-2022-confirmed-case-in-portugal/799 (2022).
9. Happi, C. et al. Urgent need for a non-discriminatory and non-stigmatizing nomenclature for monkeypox virus. https://virological.org/t/urgent-need-for-a-non-discriminatory-and-non-stigmatizing-nomenclature-for-monkeypox-virus/853 (2022).
10. Likos, A.M. et al. A tale of two clades: monkeypox viruses. J Gen Virol 86, 2661-2672 (2005).
11. Nakazawa, Y. et al. A phylogeographic investigation of African monkeypox. Viruses 7(4), 2168-2184 (2015).
12. Mauldin, M.R. et al. Exportation of Monkeypox Virus From the African Continent. J Infect Dis 225(8),1367-1376 (2022).
13. Yinka-Ogunleye, A. et al. Outbreak of human monkeypox in Nigeria in 2017–18: a clinical and epidemiological report. Lancet Infect Dis 19, 872–879 (2019).
14. Gigante, C.M. et al. Multiple lineages of Monkeypox virus detected in the United States, 2021- 2022. bioRxiv 2022.06.10.495526 (2022).
15. Perez Duque, M, et al. Ongoing monkeypox virus outbreak, Portugal, 29 April to 23 May 2022. Euro Surveill 27(22):2200424 (2022).
16. Vivancos, R. et al. Monkeypox Incident Management Team. Community transmission of monkeypox in the United Kingdom, April to May 2022. Euro Surveill 27(22):2200422 (2022).
17. Firth, C. et al. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. Mol Biol Evol 27(9), 2038-2051 (2010).
18. Hammarlund, E. et al. Multiple diagnostic techniques identify previously vaccinated individuals with protective immunity against monkeypox. Nat Med 11(9), 1005–1011 (2005).
19. Rambaut, A. Discussion of on-going MPXV genome sequencing. https://virological.org/t/discussion-of-on-going-mpxv-genome-sequencing/802 (2022).
20. Pecori, R. et al. Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination. Nat Rev Genet 7, 1–14 (2022).

Methods

This research complies with all relevant ethical regulations. The Portuguese NIH (INSA) is the national reference laboratory, being the Portuguese laboratory authorized by the General Directorate of Health (through the Technical orientation nº 004/2022 of May 31st 2022) to process the samples for identification and characterization of MPXV. All samples subjected to viral genetic characterization are processed in an anonymized fashion.

DNA Extraction and shotgun metagenomics sequencing

All clinical samples (lesions and vesicle swabs) were received by the Emergency Response and Biopreparedness Unit at INSA, and screened for Monkeypox virus by real-time PCR targeting the *rpo18* gene[31], on an CFX Opus Real-Time PCR System (Biorad). The first set of samples (n=2, received May 10-11th) were subjected to DNA

extraction using the QIamp DNA Mini Kit (Qiagen), prior to library preparation using the Rapid Barcoding Sequencing kit (SQK-RBK004) and shotgun metagenomics sequencing on an Oxford Nanopore (ONT) MinION apparatus, for 18h. We obtained the first draft genome sequence (Monkeypox/PT0001/2022) covering ~92% of the reference sequence, with a mean depth of coverage of 7-fold throughout the genome.

For the second set of samples (n=13, received May 16th), prior to DNA extraction, a pre-treatment was performed by sonication (S30 Elmasonic) in bath 20s, twice, followed by a DNase/RNase (1:10 solution of 18.5mg DNAse Sigma 400 KUnitz/mg + 52,14mg RNAse Applichem 100,8 K Units/mg in HBSS 1x) treatment (20min at 37ºC, 15min at 65ºC and 1min on ice), in order to deplete host DNA. The DNA samples were then subjected to Nextera XT library preparation, and subsequent shotgun metagenomics by paired-end sequencing (2x150 bp) on an Illumina NextSeq 2000 apparatus, with ~80M total reads per sample. Mean depth of coverage throughout Monkeypox genome ranged from 38x to 508x (mean of 201x). Seven out of the thirteen samples were also subjected to ONT MinION sequencing, as previously described. Of note, although DNase/RNase treatment have shown to perform well, results with the present sample set showed that we have experienced wet-lab technical issues affecting host-depletion effeciency. Samples details are presented in Supplementary Table 1.

Genome assembly

Reads were human-depleted using BMTagger[32] and subsequently mapped to the reference genome MPXV-UK_P2 (MT903344.1; also being used as reference in the newest Monkeypox nextstrain build https://nextstrain.org/monkeypox) using the INSaFLU pipeline[33] (https://insaflu.insa.pt/). Briefly, Illumina reads were quality processed using FastQC v0.11.5 and Trimmomatic v0.27 and mapped using Snippy v3.2, and ONT reads were quality processed using NanoStat v1.4.0 and NanoFilt v2.6.0, and mapped using medaka v1.2.1. All genome sequences were further compared with *de novo* assemblies obtained using SPAdes v3.11.1 for Illumina (to investigate the presence of large insertions/deletions or rearrangements) and all detected mutations were carefully inspected using the Integrative Genomics Viewer software. Particularly, to characterize the mutational profile of the large inverted terminal repeats, Illumina reads were independently mapped against each of the terminal repeats of MPXV-UK_P2. As this confirmed that both terminal repeat regions were identical within each genome, these regions were manually joined to each end of the final genome sequences.

Phylogenetics

A draft phylogenetic analysis for clade positioning (Figure 1A) was conducted upon core SNP alignment (1057 variant positions) retrieved from a rapid alignment (using parsnp v1.2) of the newly sequenced genomes with publicly available genomes (Supplementary Table 2), with reference genome Zaire-96-I-16 (RefSeq accession #NC_003310.1) set as an outgroup (clade 1). Fine-tune phylogeny of the 2022 outbreak-related genomes (Figure 1B) was performed by aligning novel genomes (n=15, Supplementary Table 2) with mafft v7.487, followed by manual alignment curation and Maximum Likelihood phylogenetic tree construction using MEGA v10 software. Of note, as one of the genomes (MPXV-CH-38134631/2022 - #ON595760.1) presented an excess of mutations in its terminal regions (known to be error prone during sequencing), its positions were masked for phylogenetic analysis. The updated outbreak 2022 phylogenetic tree with sequences available in NCBI as of June 15th 2022 (Figure 1C) was built similarly, with further masking (both outbreak alignments are available as Supplementary Data). Phylogenetic data visualization was performed with Microreact (https://microreact.org/).

Microevolution analysis

Rapid extraction and/or visualization of variant sites from sequence alignments was performed using ReporTree (https://github.com/insapathogenomics/ReporTree), snipit (https://github.com/aineniamh/snipit), and NextClade (https://clades.nextstrain.org/). A python script (https://github.com/insapathogenomics/mutation_profile) was developed to rapidly obtain the sequence context flanking all detected SNPs (including the outbreak cluster-defining mutations, “intra-cluster” mutations and intra-host minor variants, described in Supplementary Tables 3, 4 and 5,

respectively) and screen whether they follow signatures potentially compatible with APOBEC3-mediated viral genome editing (namely GA>AA and TC>TT replacements). Analysis of minor intra-host single nucleotide variants (iSNVs) - SNV displaying intra-sample frequency between 1- 50% was performed using the pipeline implemented in INSaFLU (min-alternate-count set to 10), with the minimum "allele" frequency being contingent on the depth of coverage of each processed site, i.e., the identification of iSNV sites at frequencies of 10%, 2% and 1% is only allowed for sites with depth of coverage of at least 100-fold, 500-fold and 1000-fold, respectively. Non-synonymous iSNVs were inspected in IGV before validation (https://insaflu.insa.pt/)[33].

Statistical analyses

No specific advanced statistical methods were required for the data analysis of the present study.

Data availability

Monkeypox reads mapping to the reference sequence MPXV-UK_P2 (GenBank accession #MT903344.1) were deposited in the European Nucleotide Archive (ENA) (BioProject accession no. PRJEB53055; www.ebi.ac.uk/ena/data/view/PRJEB53055). Assembled consensus sequences were deposited in the National Center for Biotechnology Information (NCBI) under the accession numbers ON585029-ON585038. All accession numbers are included in Supplementary Table 1.

Code availability

A python script (https://github.com/insapathogenomics/mutation_profile) was developed to rapidly obtain the sequence context flanking all detected SNPs and screen whether they follow signatures potentially compatible with APOBEC3-mediated viral genome editing.

References

21. Sadeghpour, S. et al. Human APOBEC3 Variations and Viral Infection. Viruses 13(7), 1366 (2021).

22. Giorgi, E.E. et al. Estimating the Timing of Early Simian-Human Immunodeficiency Virus Infections: a Comparison between Poisson Fitter and BEAST. mBio. 11(2):e00324-20 (2020).

23. Chen, J. & MacCarthy, T. The preferred nucleotide contexts of the AID/APOBEC cytidine deaminases have differential effects when mutating retrotransposon and virus sequences compared to host genes. PLoS computational biology 13(3), e1005471 (2017).

24. Soday, L. et al. Quantitative Temporal Proteomic Analysis of Vaccinia Virus Infection Reveals Regulation of Histone Deacetylases by an Interferon Antagonist. Cell Rep 27(6), 1920-1933 (2019).

25. Vartanian, J.P. et al. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. Science 320(5873), 230–233 (2008).

26. Mühlemann, B. et al. Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. Science 369(6502):eaaw8977 (2020).

27. Kugelman, J.R. et al. Genomic variability of monkeypox virus among humans, Democratic Republic of the Congo. Emerg Infect Dis 20(2), 232-239 (2014).

28. Rehm, K.E. et al. Vaccinia virus A35R inhibits MHC class II antigen presentation. Virology 397(1), 176-186 (2010).

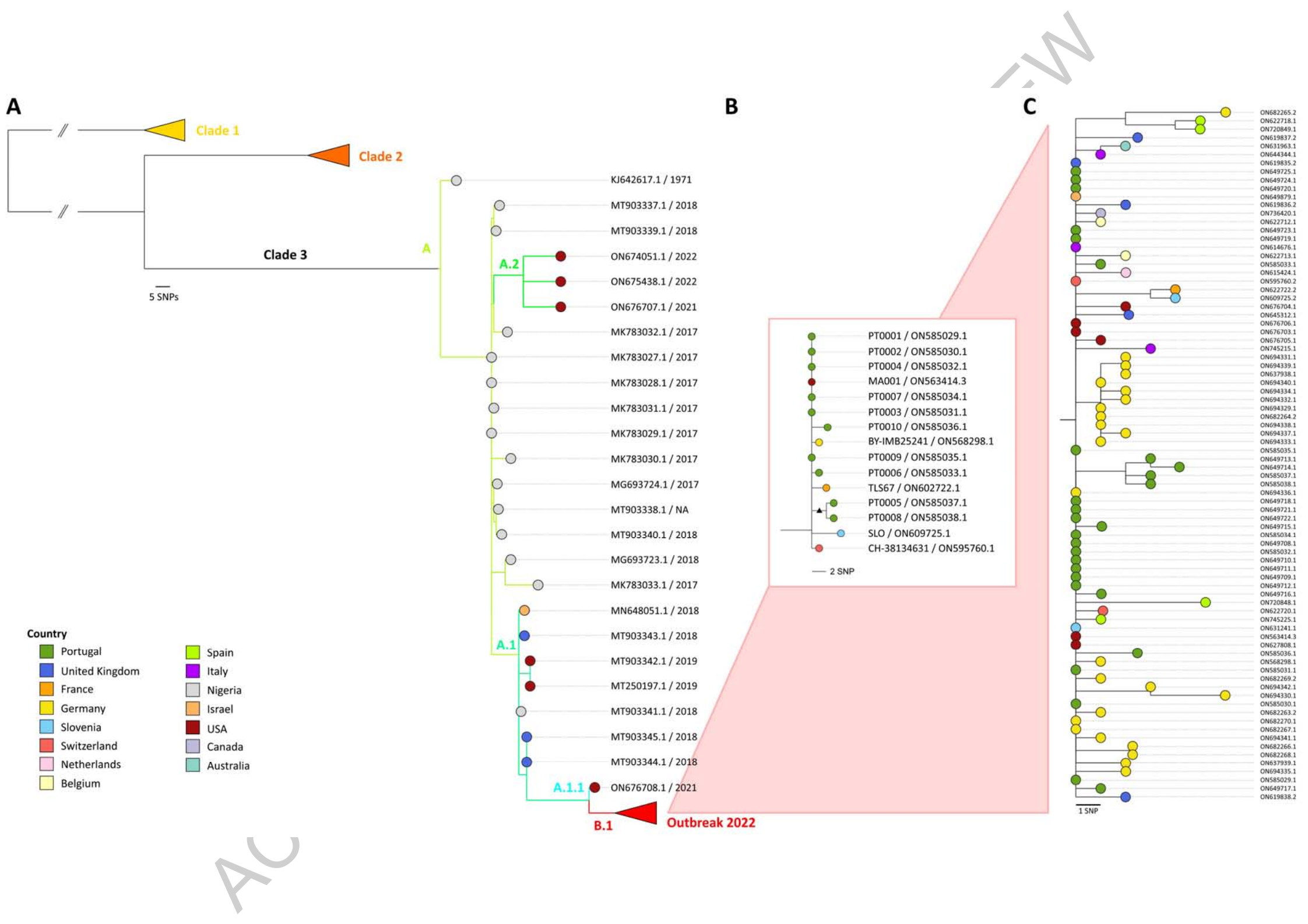
29. Perdiguero, B. and Esteban, M. The interferon system and vaccinia virus evasion mechanisms. J Interferon Cytokine Res 29(9), 581-598 (2009).

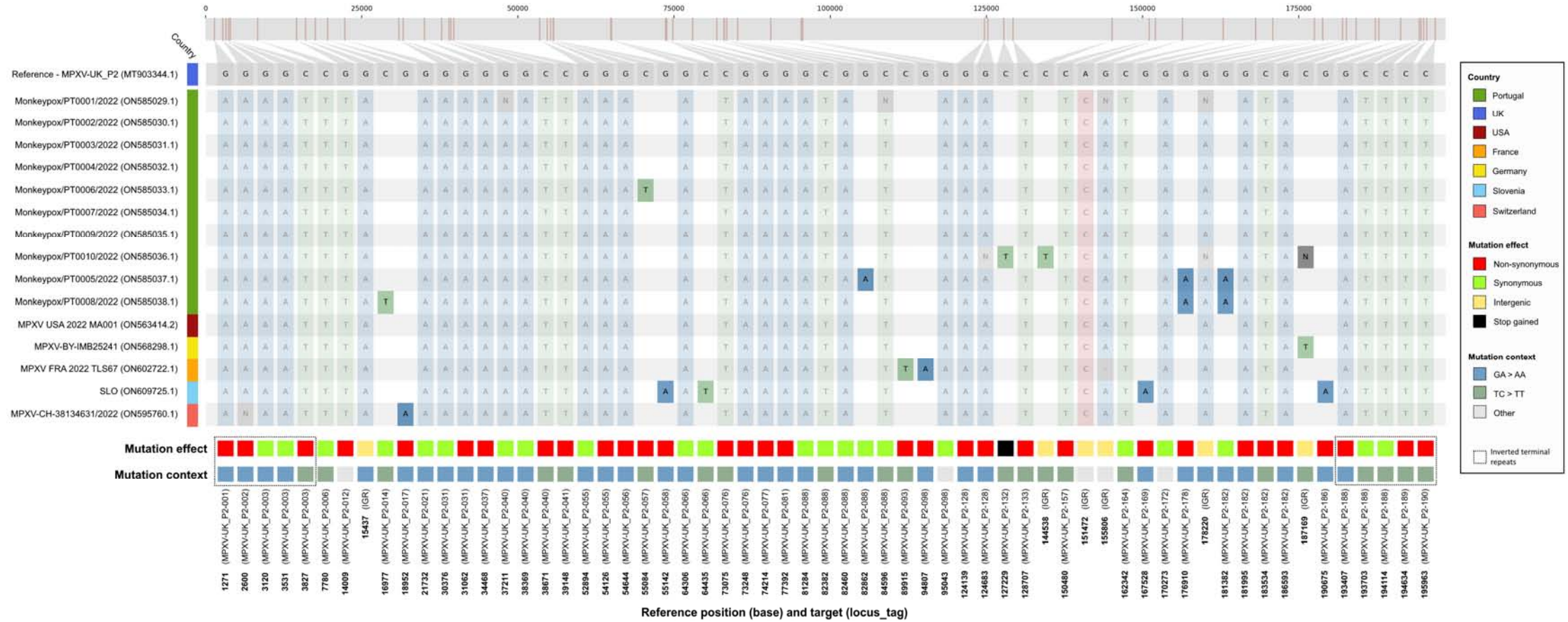
30. Bowie, A. et al. A46R and A52R from vaccinia virus are antagonists of host IL-1 and toll-like receptor signaling. Proc Natl Acad Sci USA 97(18), 10162-10167 (2000).

31. Nitsche, A. et al. Detection of orthopoxvirus DNA by real-time PCR and identification of variola virus DNA by melting analysis. J Clin Microbiol 42(3), 1207-1213 (2004).

32. Rotmistrovsky, K. & Agarwala, R. BMTagger: Best Match Tagger for Removing Human Reads from Metagenomics Datasets ftp://ftp.ncbi.nlm.nih.gov/pub/agarwala/bmtagger/ (2011).

33. Borges, V. et al. INSaFLU: an automated open web-based bioinformatics suite "from-reads" for influenza whole-genome-sequencing-based surveillance. Genome medicine 10(1), 46 (2018).


A
Clade 1
Clade 2
Clade 3
5 SNPs
A
A.2
A.1
A.1.1
B.1
KJ642617.1 / 1971
MT903337.1 / 2018
MT903339.1 / 2018
ON674051.1 / 2022
ON675438.1 / 2022
ON676707.1 / 2021
MK783032.1 / 2017
MK783027.1 / 2017
MK783028.1 / 2017
MK783031.1 / 2017
MK783029.1 / 2017
MK783030.1 / 2017
MG693724.1 / 2017
MT903338.1 / NA
MT903340.1 / 2018
MG693723.1 / 2018
MK783033.1 / 2017
MN648051.1 / 2018
MT903343.1 / 2018
MT903342.1 / 2019
MT250197.1 / 2019
MT903341.1 / 2018
MT903345.1 / 2018
MT903344.1 / 2018
ON676708.1 / 2021
Outbreak 2022
Country
Portugal
United Kingdom
France
Germany
Slovenia
Switzerland
Netherlands
Belgium
Spain
Italy
Nigeria
Israel
USA
Canada
Australia
B
PT0001 / ON585029.1
PT0002 / ON585030.1
PT0004 / ON585032.1
MA001 / ON563414.3
PT0007 / ON585034.1
PT0003 / ON585031.1
PT0010 / ON585036.1
BY-IMB25241 / ON568298.1
PT0009 / ON585035.1
PT0006 / ON585033.1
TLS67 / ON602722.1
PT0005 / ON585037.1
PT0008 / ON585038.1
SLO / ON609725.1
CH-38134631 / ON595760.1
2 SNP
C
1 SNP



nature research

Corresponding author(s): João Paulo Gomes

Last updated by author(s): Jun 18, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Not applicable
Data analysis	INSaFLU v.1.5.2 (https://insaflu.insa.pt/; code: https://github.com/INSaFLU/INSaFLU), including FastQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), Trimmomatic v0.27 (http://www.usadellab.org/cms/index.php?page=trimmomatic), NanoStat v1.4.0 (https://github.com/wdecoster/nanostat), NanoFilt v.2.6.0 (https://github.com/wdecoster/nanofilt), RabbitQC 0.0.1 (https://github.com/ZekunYin/RabbitQC), SPAdes v3.11.1 (http://cab.spbu.ru/software/spades/), Snippy v.3.2-dev (https://github.com/tseemann/snippy, sligthly modified in https://github.com/INSaFLU/INSaFLU), Medaka v1.2.1 (https://nanoporetech.github.io/medaka/), msa_masker (https://github.com/rfm-targa/BioinfUtils/blob/master/msa_masker.py), Integrative Genomics Viewer (http://software.broadinstitute.org/software/igv/), getCoveragev1.1 (https://github.com/monsanto-pinheiro/getCoverage), ReporTree (https://github.com/insapathogenomics/ReporTree); get_mutation_profile (https://github.com/insapathogenomics/mutation_profile); MAFFT v7.487 (https://mafft.cbrc.jp/alignment/software/); MEGA v10 (https://www.megasoftware.net/); snipit (https://github.com/aineniamh/snipit), Microreact (https://microreact.org/), Nextclade 2.1.0 (https://clades.nextstrain.org/).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Monkeypox reads mapping to the reference sequence MPXV-UK_P2 (MT903344.1) were deposited in the European Nucleotide Archive (ENA) (BioProject accession no. PRJEB53055; www.ebi.ac.uk/ena/data/view/PRJEB53055). Reads maping against the human genome were removed before submission. Assembled consensus sequences were deposited in the National Center for Biotechnology Information (NCBI) under the accession numbers ON585029-ON585038. All accession numbers are included in Supplementary Table 1.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	To define the sample size we opted for sequencing the viral genomes from all the samples that were available at the Ref Lab at the time we performed the wet-lab procedures. The goal was to provide a first robust snapshot on phylogenomics of the 2022 Monkeypox virus.
Data exclusions	Samples presenting a very low viral load (defined by the real-time PCR threshold cycle values) were excluded.
Replication	Not applicable. Shotgun metagenomics sequencing was performed for each specimen, with robustness of sequence data being ensured by the applied depth of coverage and data analysis.
Randomization	As described above, all available suitable samples were processed at that time.
Blinding	Blinding was not relevant/applicable. We analyzed all available samples strictly for genomic purposes and the data is anonymized.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging