Check for updates

OPEN

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

Joana Isidro[1,6], Vítor Borges[1,6], Miguel Pinto[1,6], 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] and João Paulo Gomes [1,5] ✉

The largest monkeypox virus (MPXV) outbreak described so far in non-endemic countries was identified in May 2022 (refs. [1–6]). In this study, 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 double-stranded DNA virus.

Monkeypox is a rare zoonotic disease that is caused by the MPXV from the Orthopoxvirus genus, which includes the variola virus, the causative agent of smallpox[1–3]. With an incubation period of 5–21 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 harbor the virus, leading to occasional spill-over events to humans[1–3]. 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[1–4]. We are now facing the first multi-country outbreak without known epidemiological links to West or Central Africa[1], with more than 2,500 confirmed cases reported worldwide as of 18 June 2022 (refs. [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 patients with monkeypox (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 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.

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 20 May 2022 by Portugal[8], as well as on additional sequences released in the National Center for Biotechnology Information (NCBI) before 27 May 2022, with 15 sequences in total (most of them from Portugal) (Supplementary Tables 1 and 2). The rapid integration of the first sequence into the global MPXV genetic diversity (Fig. 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 clades 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 >10% CFR[10,11]. All outbreak MPXV strains sequenced so far tightly cluster together (Fig. 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 (ref. [13]) (Fig. 1). Given these findings and the MPXV historical epidemiology (rare cases in non-endemic
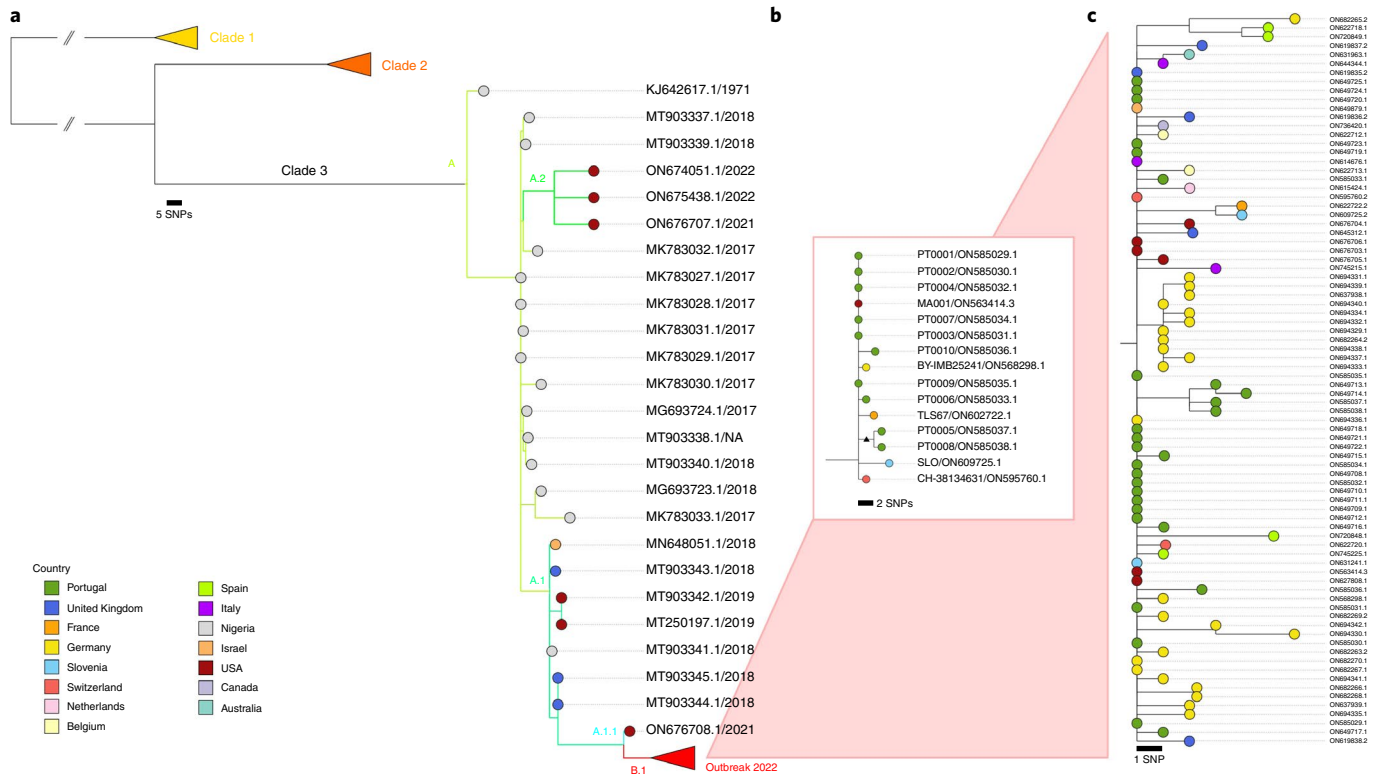
Fig. 1 | Phylogenetic analysis of MPXV viral sequences associated with the 2022 worldwide outbreak. a, MPXV global phylogeny showing that the 2022 outbreak cluster (lineage B.1) belongs to clade 3. Clade and lineage are designated according to the nomenclature proposed by Happi et al.[9]. **b**, Genetic diversity within the outbreak cluster, including the 15 sequences analyzed in this study (released in the NCBI before 27 May 2022). The deletion symbol (Δ) denotes a large deletion (11,335–12,247 in the MPXV-UK_P2-010 gene) shared by sequences segregating in a small subcluster. **c**, Outbreak phylogenetic tree updated with sequences available in the NCBI as of 15 June 2022 (provided during revision for more updated contextualization). The list of the sequences used in these phylogenetic analyses is detailed in Supplementary Table 2, and the alignments are provided as Supplementary Data.

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 an MPXV sequence from a 2021 travel-associated case from Nigeria to the United States (USA_2021_MD; accession no. ON676708)[14] phylogenetically placed between 2018–2019 and 2022 sequences (Fig. 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 (for example, after the reported 2018–2019 importations). Silent human-to-human transmission (for example, due to underdiagnosis) seems less likely considering the known disease characteristics of the affected individuals, usually involving localized or generalized 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 a 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) (for example, 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 3-week period[1] after a first report on 7 May 2022 by the United Kingdom[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 United Kingdom on 3–4 May 2022 (refs. [1,3])), this scenario can be discarded as the earliest symptom onset dates for confirmed cases in Portugal and in the United Kingdom were in late April[15,16].

Notably, the 2022 MPXV diverges from the related 2018–2019 viruses by a mean of 50 single-nucleotide polymorphisms (SNPs) (Figs. 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 genome per year)[17]. Such a divergent branch might represent accelerated evolution. Of note, among the 46 SNPs (24 non-synonymous, 18 synonymous and four intergenic) (Supplementary Table 3) separating the 2022 MPXV outbreak virus from the reference sequence (MPXV-UK_P2, 2018; GenBank accession no. MT903344.1), three amino acid changes (D209N, P722S and 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, ten synonymous and two intergenic) and 15 (nine non-synonymous and 16 synonymous) being GA > AA and TC > TT nucleotide replacements, respectively (Fig. 2 and Supplementary Table 3). A tool (https://github.com/insapathogenomics/mutation_profile) was built to rapidly screen these and other mutation profiles. The observed (hyper)mutation signature might suggest the potential action of apolipoprotein B mRNA-editing catalytic polypeptide-like 3
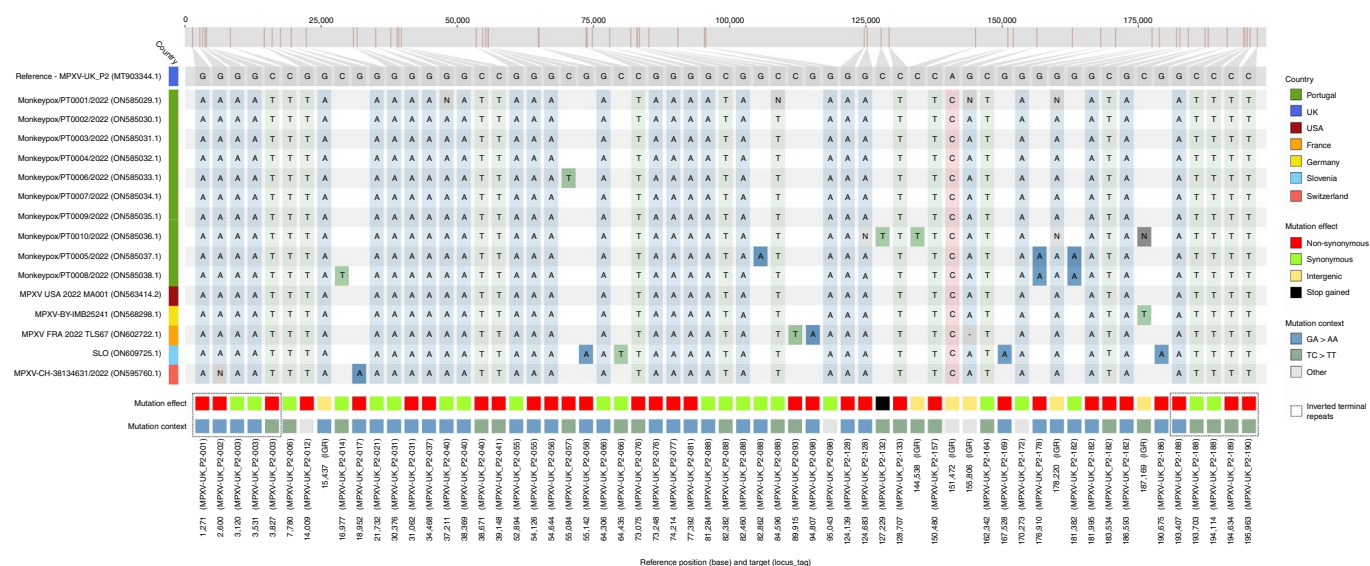
Fig. 2 | SNPs characterizing the 2022 MPXV outbreak variant. Light-colored mutations represent the SNPs separating the MPXV 2022 outbreak cluster from the MPXV_UK_P2 (MT903344.1) reference sequence (Supplementary Table 3). Dark-colored mutations represent the genetic diversity within the outbreak cluster (Supplementary Table 4).

(APOBEC3) enzymes in the viral genome editing[20]. Also, MPXV is 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 (for example, 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 (for example, 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 pre-cancerous 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 MPXV evolution augments the uncertainty regarding the 2022 outbreak origins and introductions, in addition to the complexity of the epidemiological context. This raises the need for future studies focusing on the weight of APOBEC3 in MPXV diversification. In particular, functional studies assessing whether this mutational driver triggers MPXV adaptive evolution toward altered phenotypic features, such as enhanced transmissibility, are warranted.

Further phylogenomic analysis revealed the first signs of microevolution of this virus during human-to-human transmission. Among the 15 outbreak sequences analyzed here, we detected the emergence of 15 SNPs (eight non-synonymous, four synonymous, two intergenic and one stop gained) (Fig. 2 and Supplementary Table 4). Notably, all SNPs also follow the same mutational bias, including eight GA > AA (six non-synonymous and two synonymous) and seven TC > TT (two non-synonymous, two synonymous, one stop gained and two intergenic) nucleotide replacements.

This further suggests a continuous action of APOBEC3 during MPXV evolution. Among the seven phylogenetic branches directly descendant from the most recent ancestor of the MPXV outbreak strain (Fig. 1), we identified a subcluster (supported by two SNPs) of two sequences (PT0005 and PT0008, each with an additional SNP) that also share a 913-bp 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 reveal 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 3–5) found to be targeted during the 2022 outbreak MPXV 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 MPXV outbreak strain and shed light on potential mechanisms and targets of human adaptation. The observed accelerated evolution of this human MPXV, 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 double-stranded DNA virus. Together with the adopted strategy of real-time data sharing, this study may help guide novel outbreak control measures and subsequent research directions.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-022-01907-y.

Received: 27 May 2022; Accepted: 20 June 2022;
Published online: 24 June 2022

References

1. European Centre for Disease Prevention and Control. Risk assessment: monkeypox multi-country outbreak. https://www.ecdc.europa.eu/en/publications-data/risk-assessment-monkeypox-multi-country-outbreak (2022).
2. World Health Organization. Monkeypox fact sheet. 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**, e0010141 (2022).
4. UK Health Security Agency. Guidance: monkeypox: background information. https://www.gov.uk/guidance/monkeypox (2022).
5. European Centre for Disease Prevention and Control. Epidemiological update: monkeypox multi-country outbreak. 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.* https://doi.org/10.1016/S1473-3099(22)00359-0 (2022).
7. World Health Organization. Guidelines vaccines and immunization for monkeypox: interim guidance, 14 June 2022. 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**, 2168–2184 (2015).
12. Mauldin, M. R. et al. Exportation of monkeypox virus from the African continent. *J. Infect. Dis.* **225**, 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. Preprint at https://www.biorxiv.org/content/10.1101/2022.06.10.495526v1 (2022).
15. Perez Duque, M. et al. Ongoing monkeypox virus outbreak, Portugal, 29 April to 23 May 2022. *Eur. Surveill.* **27**, 2200424 (2022).
16. Vivancos, R. et al. Monkeypox Incident Management Team. Community transmission of monkeypox in the United Kingdom, April to May 2022. *Eur. Surveill.* **27**, 2200422 (2022).
17. Firth, C. et al. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol. Biol. Evol.* **27**, 2038–2051 (2010).
18. Hammarlund, E. et al. Multiple diagnostic techniques identify previously vaccinated individuals with protective immunity against monkeypox. *Nat. Med.* **11**, 1005–1011 (2005).
19. O'Toole, A. & Rambaut, A. Initial observations about putative APOBEC3 deaminase editing driving short-term evolution of MPXV since 2017. https://virological.org/t/initial-observations-about-putative-apobec3-deaminase-editing-driving-short-term-evolution-of-mpxv-since-2017/830
20. Pecori, R. et al. Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination. *Nat. Rev. Genet.* **7**, 1–14 (2022).
21. Sadeghpour, S. et al. Human APOBEC3 variations and viral infection. *Viruses* **13**, 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**, 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 Comput. Biol.* **13**, 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**, 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**, 230–233 (2008).
26. Mühlemann, B. et al. Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. *Science* **369**, eaaw8977 (2020).
27. Kugelman, J. R. et al. Genomic variability of monkeypox virus among humans, Democratic Republic of the Congo. *Emerg. Infect. Dis.* **20**, 232–239 (2014).
28. Rehm, K. E. et al. Vaccinia virus A35R inhibits MHC class II antigen presentation. *Virology* **397**, 176–186 (2010).
29. Perdiguero, B. & Esteban, M. The interferon system and vaccinia virus evasion mechanisms. *J. Interferon Cytokine Res.* **29**, 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**, 10162–10167 (2000).

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 technical orientation no. 004/2022 of 31 May 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 MPXV by real-time polymerase chain reaction (PCR) targeting the *rpo18* gene[31], on a CFX Opus Real-Time PCR System (Bio-Rad). The first set of samples ($n = 2$, received 10–11 May 2022) were subjected to DNA extraction using the QIamp DNA Mini Kit (Qiagen) before library preparation using the Rapid Barcoding Sequencing Kit (SQK-RBK004) and shotgun metagenomics sequencing on an Oxford Nanopore Technologies (ONT) MinION apparatus for 18 hours. We obtained the first draft genome sequence (Monkeypox/PT0001/2022) covering ~92% of the reference sequence, with a mean depth of coverage of seven-fold throughout the genome.

For the second set of samples ($n = 13$, received 16 May 2022), before DNA extraction a pre-treatment was performed by sonication (S30 Elmasonic) in a bath for 20 seconds, twice, followed by a DNase/RNase (1:10 solution of 18.5 mg of DNAse (Sigma-Aldrich) 400 Kunitz mg^{-1} + 52.14 mg of RNAse AppliChem 100.8 Kunitz mg^{-1} in HBSS 1×) treatment (20 minutes at 37 °C, 15 minutes at 65 °C and 1 minute on ice), to deplete host DNA. The DNA samples were then subjected to Nextera XT library preparation and subsequent shotgun metagenomics by paired-end sequencing (2 × 150 bp) on an Illumina NextSeq 2000 apparatus, with ~80 million total reads per sample. Mean depth of coverage throughout the monkeypox genome ranged from 38× to 508× (mean 201×). Seven out of the 13 samples were also subjected to ONT MinION sequencing, as previously described. Of note, although DNase/RNase treatment has been shown to perform well, results with the present sample set showed that we experienced wet lab technical issues affecting host depletion efficiency. 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/). In brief, Illumina reads were quality processed using FastQC version 0.11.5 and Trimmomatic version 0.27 and mapped using Snippy version 3.2, and ONT reads were quality processed using NanoStat version 1.4.0 and NanoFilt version 2.6.0 and mapped using medaka version 1.2.1. All genome sequences were further compared to de novo assemblies obtained using SPAdes version 3.11.1 for Illumina (to investigate the presence of large insertions/deletions or rearrangements), and all detected mutations were carefully inspected using 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 (Fig. 1a) was conducted upon core SNP alignment (1,057 variant positions) retrieved from a rapid alignment (using parsnp version 1.2) of the newly sequenced genomes with publicly available genomes (Supplementary Table 2), with reference genome Zaire-96-I-16 (RefSeq accession no. NC_003310.1) set as an outgroup (clade 1). Fine-tune phylogeny of the 2022 outbreak-related genomes (Fig. 1b) was performed by aligning novel genomes ($n = 15$; Supplementary Table 2) with mafft version 7.487, followed by manual alignment curation and maximum likelihood phylogenetic tree construction using MEGA version 10 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 15 June 2022 (Fig. 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 3–5, 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 iSNVs

displaying intra-sample frequency between 1% and 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—that is, 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 1,000-fold, respectively. Non-synonymous iSNVs were inspected in Integrative Genomics Viewer before validation (https://insaflu.insa.pt/)[33].

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

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Monkeypox reads mapping to the reference sequence MPXV-UK_P2 (GenBank accession no. MT903344.1) were deposited in the European Nucleotide Archive (BioProject accession no. PRJEB53055). Assembled consensus sequences were deposited in the National Center for Biotechnology Information under accession nos. 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

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**, 1207–1213 (2004).
32. Rotmistrovsky, K. & Agarwala, R. BMTagger: best match tagger for removing human reads from metagenomics datasets. https://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 Med.* **10**, 46 (2018).

Acknowledgements

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

The authors declare no competing interests.

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

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