



Coding-Complete Genome Sequences for Two Confirmed Monkeypox Cases in South Africa 2022

Wai Yin Chan,^{a,b} Phillip Senzo Mtshali,^a Antoinette Grobbelaar,^a  Naazneen Moolla,^{a,c} Thabo Mohale,^a Michelle Lowe,^a Morne Du Plessis,^a  Arshad Ismail,^{a,d}  Jacqueline Weyer^{a,c,e}

^aNational Institute for Communicable Diseases, National Health Laboratory Service, Johannesburg, South Africa

^bDepartment of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa

^cDepartment of Medical Virology, Faculty of Health Sciences, University of Pretoria, Tshwane, South Africa

^dDepartment of Biochemistry and Microbiology, Faculty of Science, Engineering and Agriculture, University of Venda, Thohoyandou, South Africa

^eDepartment of Microbiology and Infectious Diseases, Faculty of Health Sciences, University of Witwatersrand, Johannesburg, South Africa

ABSTRACT The coding-complete genome sequences of monkeypox virus (MPXV) were obtained from skin lesion swabs from two human cases detected in South Africa in June 2022. Sequence analyses indicated that the genetic sequences of the viruses associated with these two cases were related most closely to the genetic sequences of other MPXVs reported during the 2022 multicountry outbreak and belong to the monkeypox hMPXV-1 clade (previously West Africa clade) and B.1 lineage.

The monkeypox virus (MPXV) is a zoonotic virus belonging to the *Orthopoxvirus* genus. The virus is associated with a yet-to-be-determined natural animal host from the Western and Central African regions, with human cases rarely reported (1, 2). MPXV is classified into two clades, namely, I and II (previously known as the Central Africa and West Africa clades, respectively), with 3 sublineages, including Ia, IIa, and IIb (3, 4). Since May 2022, an unprecedented outbreak of monkeypox with human-to-human transmission and multicountry spread has been reported (5), as a result of a new lineage (B.1) which diverged from clade IIb (6).

Here, we report the coding-complete genome sequences of two confirmed monkeypox cases detected in South Africa in June 2022. These sequences were generated using a metagenomics approach. The first case involved a 30-year-old male residing in the Gauteng Province of South Africa (sample reference NICD-SVPL223), and the second case involved a 32-year-old male residing in the Western Cape Province (sample reference NICD-SVPL232). Both cases reported no travel history prior to becoming ill but did report contact with individuals that did travel abroad and may have served as possible sources of exposure.

This study was approved by the Human Research Ethics Committee of the University of the Witwatersrand Johannesburg South Africa (protocol number: M210752). Nucleic acids were isolated from swabs collected from skin lesions using the QIAamp viral RNA mini kit (Qiagen, Germany) according to the manufacturer's instructions. DNA was quantified with the Qubit double-stranded DNA (dsDNA) high-sensitivity assay (Life Technologies, USA) on the Qubit 3.0 instrument (Life Technologies) according to the manufacturer's instructions. Shotgun metagenomics sequencing was performed using 11 to 14 ng of the extracted DNA. Multiplexed paired-end libraries were prepared using the Illumina Nextera DNA flex preparation kit followed by sequencing (2 × 150 bp) on a NextSeq 550 instrument (Illumina, Inc., USA). Low-quality and adaptor sequence regions were trimmed using trimGalore v0.6.2 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Host sequence reads were removed using Bowtie 2 v2.4.2 (7) against a human reference genome (GCA_000001405). The remaining reads were mapped to the reference genome MPXP_UK_P2 (NCBI accession number MT903344) with BWA-MEM v0.7.17 (8) followed by variant calling with LoFreq v2.1.3.1 (9)

Editor Jelle Matthijnssens, KU Leuven

Copyright © 2022 Chan et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Jacqueline Weyer, jacquelinew@nicd.ac.za.

The authors declare no conflict of interest.

Received 3 August 2022

Accepted 20 October 2022

Published 10 November 2022

TABLE 1 Next-generation sequence data mapping statistics of the genomes of MPXV

Sample	Total no. of reads	No. of reads passing QC ^a	No. of reads mapped to MPXV reference	Sequencing coverage (X)	Total genome size (bp)	GC content (%)
NICD-SVPL223	169,216,239	169,039,316	560,807	391.433	197,213	32.99
NICD-SVPL232	53,526,611	53,164,490	28,670	20.1	197,126	33

^a QC, quality control.

and bcltools v1.10.2-2. The consensus sequence was generated with iVar v1.0 (10), using the output of the SAMtools v1.15 (11) mpileup-aligned BAM output file with a consensus frequency threshold of 0.6. Gene annotations and open reading frames were predicted using Genome Annotation Transfer Utility (GATU) (12) with MPXP_UK_P2 as the reference. The remaining reads were also subject to *de novo* assembly with SPAdes v3.14.1 (13) and the Galaxy monkeypox workflow (<https://galaxyproject.org/projects/mpxv/>) (14) as additional verification of the assembly result. Default parameters were used unless otherwise stated.

The coding-complete genomes consist of double-stranded DNA with a genome size of 197 kb (Table 1) and a genome coverage of 99.9% when aligned to the reference genome

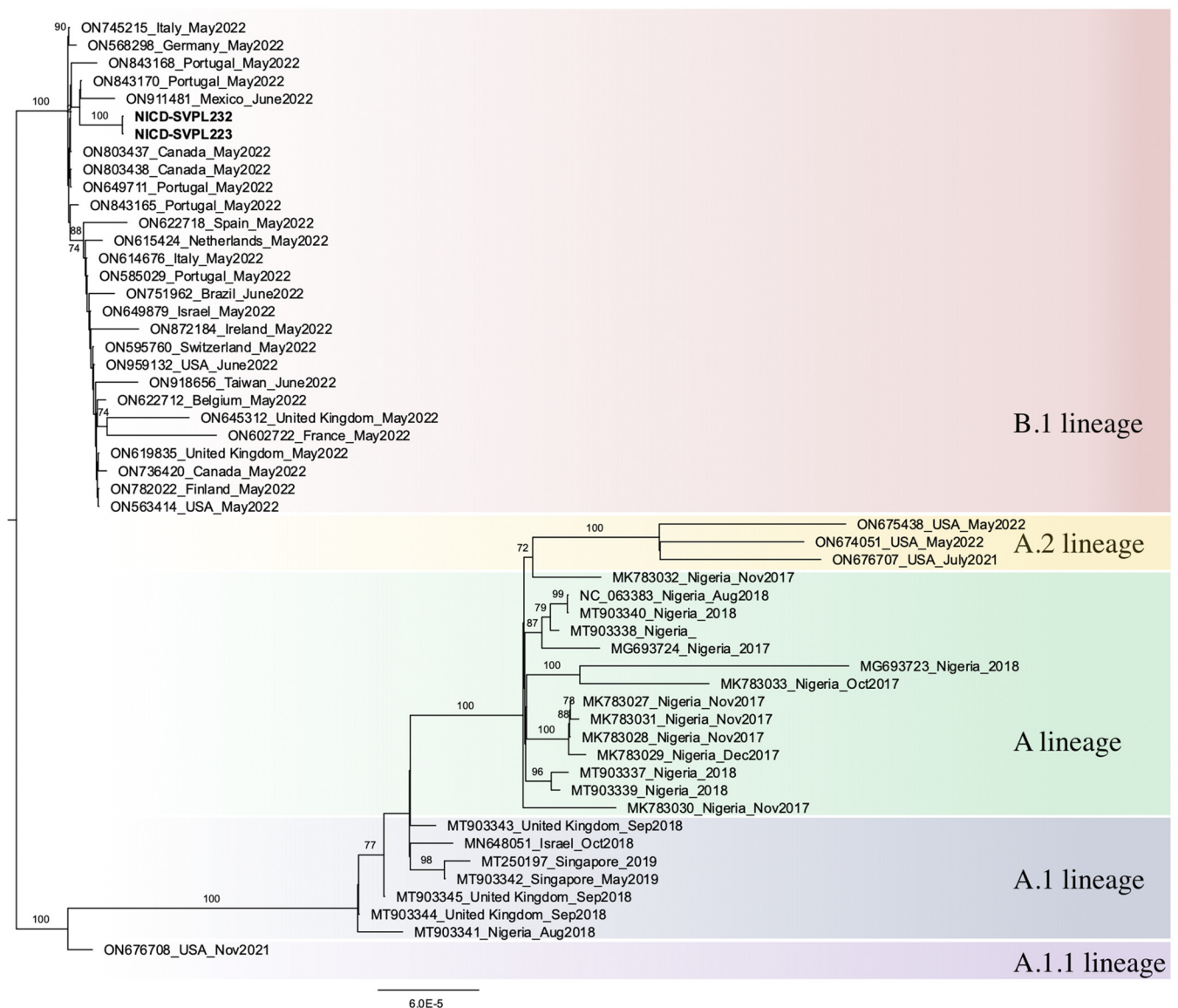


FIG 1 Maximum likelihood tree of the hMPXV-1 clade. Both NICD-SVPL223 and SVPL232 grouped in the hMPXV-1 clade, B.1 lineage. Alignment of the MPXV genomes and the consensus sequences of NICD-SVPL223 and NICD-SVPL232 were generated using MAFFT v7.505 (16) and were curated. The phylogenetics tree was constructed using IQ-TREE2 v2.0.3 (17) using the best-fit model from ModelFinder and a 1,000 ultrafast bootstrap.

sequence. The differences between the two reported assembly sequences to the reference genome are represented by 37 single nucleotide mutations (23 nonsynonymous and 14 synonymous mutations). The commonest nucleotide mutations detected were G > A (23/37), C > T (12/37), GA>AA APOBEC3-like (14/37), and TC>TT (9/37). Publicly available MPXV genome sequences with near-complete to complete status were collected from NCBI (15) for phylogenetic analysis, which indicates that the South African-associated MPXV cases belonged to the B.1 lineage of the hMPXV-1 and that they are related most closely to other MPXV genetic sequences reported during the multicountry outbreak of 2022 (Fig. 1).

Data availability. The genome sequences for NICD-SVPL223 and NICD-SVPL232 were submitted to NCBI GenBank (accession numbers [ON918611](https://doi.org/10.1093/nar/gks918) and [ON927248](https://doi.org/10.1093/nar/gks918)). Raw sequence data have also been deposited in NCBI (BioProject accession number [PRJNA856120](https://doi.org/10.1093/nar/gks918) and SRA accession number [SRR19995508](https://doi.org/10.1093/nar/gks918) and [SRR19995509](https://doi.org/10.1093/nar/gks918)).

ACKNOWLEDGMENTS

We acknowledge Peter van Heusden from the South African National Bioinformatics Institute (SANBI) and Mushal Allam from the Department of Genetics and Genomics College of Medicine and Health Sciences at UAEU for their advice on the methodology.

REFERENCES

- Ligon BL. 2004. Monkeypox: a review of the history and emergence in the Western hemisphere. *Semin Pediatr Infect Dis* 15:280–287. <https://doi.org/10.1053/j.spid.2004.09.001>.
- Durski KN, McCollum AM, Nakazawa Y, Petersen BW, Reynolds MG, Briand S, Djingarey MH, Olson V, Damon IK, Khalakdina A. 2018. Emergence of monkeypox—West and Central Africa. *MMWR Morb Mortal Wkly Rep* 67:306–310. <https://doi.org/10.15585/mmwr.mm6710a5>.
- Likos AM, Sammons SA, Olson VA, Frace AM, Li Y, Olsen-Rasmussen M, Davidson W, Galloway R, Khristova ML, Reynolds MG, Zhao H, Carroll DS, Curns A, Formenty P, Esposito JJ, Regnery RL, Damon IK. 2005. A tale of two clades: monkeypox viruses. *J Gen Virol* 86:2661–2672. <https://doi.org/10.1099/vir.0.81215-0>.
- Happi C, Adetifa I, Mbala P, Njouou R, Nakoune E, Happi A, Ndodo N, Ayansola O, Mboowa G, Bedford T, Neher RA, Roemer C, Hodcroft E, Tegally H, O'Toole A, Rambaut A, Pybus O, Kraemer MUG, Wilkinson E, Isidro J, Borges V, Pinto M, Gomes JP, Freitas L, Resende PC, Lee RTC, Maurer-Stroh S, Baxter C, Lessells R, O'Connell AE, Kebede Y, Tessema SK, de Oliveira T. 2022. Urgent need for a non-discriminatory and non-stigmatizing nomenclature for monkeypox virus. *PLoS Biol* 20:e3001769. <https://doi.org/10.1371/journal.pbio.3001769>.
- Kraemer MUG, Tegally H, Pigott DM, Dasgupta A, Sheldon J, Wilkinson E, Schultheiss M, Han A, Oglia M, Marks S, Kanner J, O'Brien K, Dandamudi S, Rader B, Sewalk K, Bento AI, Scarpino SV, de Oliveira T, Bogoch II, Katz R, Brownstein JS. 2022. Tracking the 2022 monkeypox outbreak with epidemiological data in real-time. *Lancet Infect Dis* 22:941–942. [https://doi.org/10.1016/S1473-3099\(22\)00359-0](https://doi.org/10.1016/S1473-3099(22)00359-0).
- Khosravi E, Keikha M. 2022. B.1 as a new human monkeypox sublineage that linked with the monkeypox virus (MPXV) 2022 outbreak—correspondence. *Int J Surg* 105:106872. <https://doi.org/10.1016/j.ijssu.2022.106872>.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, Khor CC, Petric R, Hibberd ML, Nagarajan N. 2012. LoFreq: a sequence-quality aware ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res* 40:11189–11201. <https://doi.org/10.1093/nar/gks918>.
- Castellano S, Cestari F, Faglioni G, Tenedini E, Marino M, Artuso L, Manfredini R, Luppi M, Trenti T, Tagliafico E. 2021. iVar: an interpretation-oriented tool to manage the update and revision of variant annotation and classification. *Genes* 12:384. <https://doi.org/10.3390/genes12030384>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Tcherepanov V, Ehlers A, Upton C. 2006. Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. *BMC Genomics* 7:150. <https://doi.org/10.1186/1471-2164-7-150>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Jalili V, Afgan E, Gu Q, Clements D, Blankenberg D, Goetsch J, Taylor J, Nekrutenko A. 2020. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res* 48:W395–W402. <https://doi.org/10.1093/nar/gkaa434>.
- Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST. 2022. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 50:D20–D26. <https://doi.org/10.1093/nar/gkab1112>.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>.