bioRxiv preprint doi: https://doi.org/10.1101/2022.06.01.494368; this version posted June 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genomic characterization of the recent monkeypox outbreak

Federico M. Giorgi, Daniele Pozzobon, Antonio Di Meglio, Daniele Mercatelli

Department of Pharmacy and Biotechnology, University of Bologna, Via Selmi 3, 40126 Bologna, Italy

Federico M. Giorgi: 0000-0002-7325-9908

Daniele Pozzobon: 0000-0003-0817-9247

Antonio di Meglio: 0000-0002-0397-2249

Daniele Mercatelli: 0000-0003-3228-0580

Abstract

The Monkeypox virus is the etiological cause of a recent multi-country outbreak, with nearly one hundred distinct cases detected outside the endemic areas of Africa in May 2022. In this article, we analyze the sequences of two full genomes of Monkeypox virus from Portugal and Belgium, and compare them with all available Monkeypox sequences, annotated by year and geographic origin, as well as related Cowpox and Variola (smallpox) virus sequences. Our results show that the recent outbreak is most likely originating from the West African clade of Monkeypox, with >99% sequence identity with sequences derived from historical and recent cases, dating from 1971 to 2017. We also analyze specific mutations occurring in viral proteins A42R (for which a crystal structure is available) and H3L (an important epitope in host immune recognition), highlighting specific amino acids varying between the current outbreak, previous Monkeypox and Cowpox sequences and the historical Variola virus.

Introduction

Monkeypox is an infectious viral disease affecting humans with symptoms similar to smallpox, including fever, muscle pains and blisters. Monkeypox and smallpox are caused by the Monkeypox virus and the Variola virus, respectively, both members of the Orthopoxvirus family of double-stranded DNA viruses. However, while the Variola virus has been eradicated worldwide in 1980 [1], the Monkeypox virus is still endemic in sub-Saharan Africa [2].

In May 2022, nearly a hundred independent cases of monkeypox have been reported in 12 non-African countries, namely Australia, Belgium Canada, France, Germany, Italy, Netherlands, Portugal, Spain, Sweden, United Kingdom and United States of America [3,4]. While the main source of contagion in the endemic areas is believed to be zoonotic, i.e. deriving from the contact with infected animals in the wild [5], human-to-human transmission has been confirmed in cases from the United Kingdom in 2018 [6] and also in the recent multi-country outbreak [7]. Previous

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.01.494368; this version posted June 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

monkeypox cases in countries outside the African continent have been reported since 2003 [5], but never with case numbers as high as the recent outbreak [8,9].

From the genetic point of view, there are two distinct clades of the Monkeypox virus, namely the West African clade and the Central African (or Congolese) clade [10], with similar clinical and pathological features [11]. The Monkeypox virus is generally considered closely related to the Variola and Cowpox viruses, both in etiological aspects and genetic measurements [12–14]. This similarity is reflected also in immunological contexts, for example it has been reported that smallpox vaccination offers 85% protection against monkeypox in humans [15]. The reference Monkeypox virus, available on NCBI with accession code NC_003010.1 and derived from the sample Zaire-96-I-16 [14], carries a 196,858 nucleotide-long double-stranded DNA linear genome, containing 191 non-overlapping genes [16] (**Figure 1 A**), and is a representative of the Central African clade.

As shown by the recent worldwide pandemic caused by the SARS-CoV-2[17], a strict monitoring of the genomic features of the Monkeypox virus is of paramount importance, as well as the need for continued genomic surveillance on the evolution of this potential threat. In this article, we will analyze the feature of the Monkeypox virus genome, including two full sequences from the recent outbreak of May 2022, and establish its relationship with historical sequences from Monkeypox viruses, as well as Cowpox viruses and the Variola virus.

Results

We retrieved two draft full-genome Monkeypox virus sequences from recent (May 2022) cases of monkeypox in Lisbon, Portugal [18] and Antwerpen, Belgium [19], named respectively PT0001 and ITM001 by the original authors. Using BLAST [20] with the PT0001 Monkeypox genome sequence as query, we retrieved the 100 most similar sequences according to BLAST bitscore. These were formed by 72 Monkeypox sequences and 28 Cowpox sequences, dating back to samples retrieved from 1971. We added to this selection the two reference sequences of Central African Monkeypox virus NC_003310.1 [14] and the Indian Variola virus NC_001611.1 [21]. The total of analyzed genomes is 104 and is available, in FASTA format, as **Supplementary File 1**.

Phylogenetic analysis of Monkeypox virus genome sequences shows a clear distinction between Central African sequences (including NC_003310.1) and West African sequences. The two sequences from the recent cases in Portugal and Belgium appear highly related to each other, and robustly located within the West African clade (**Figure 1 B**).

Beyond the Monkeypox sequences, Cowpox genomes constitute a clearly different genetic group, highlighted by both phylogenetic analysis (**Supplementary File 2**) and by clustering based on sequence similarity (**Figure 2 A**). The Variola virus appears genetically distinct from both Monkeypox and Cowpox viruses (**Figure 2 A**). Nevertheless, the Variola virus genome is highly similar, in terms of genome-wide sequence identity, to the Monkeypox virus (**Figure 2 B**), with 96.6% DNA identity with the Central African Monkeypox (sequence NC_003310.1), 96.2% with the West African Monkeypox (representative sequence from Nigeria MT903340.1), 96.8% with the recent Monkeypox case from Belgium (sequence ITM001), and 95.7% with the recent Monkeypox case from Portugal (sequence PT0001). Overall, all West African Monkeypox, Central African Monkeypox, Cowpox and Variola virus genome sequences analyzed possess a sequence identity above 95% (**Figure 2 B**).

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.01.494368; this version posted June 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We proceeded then to analyze specific sequences within the genomes of Monkeypox and related viruses, focusing on proteins of relevance. Initially, we searched the Protein Data Bank [22] for protein structures, and found a single entry for a Monkeypox protein: the structure, inferred from crystallography, of the A42R protein from the Central African Monkeypox strain Zaire-96-I-16, represented in our genome analysis by reference sequence NC_003310.1 [23]. A42R is a short (~130 aminoacid long) protein homologous to eukaryote profilins, actin-binding proteins involved in cytoskeletal structure and function, such as membrane trafficking [24]. Our analysis showed a 100% identity between Monkeypox sequences from the recent outbreak PT0001 and ITM001 and the other Monkeypox A42R proteins; only 4 amino acids differ between Monkeypox A42R sequences and other closely related Orthopoxvirus, namely Lysin 16, which is replaced by an Asparagin in Variola virus (**Figure 2 C**), and Valine 48, Histidine 55 and Phenylalanine 66, which are Isoleucine, Arginine and Tyrosine in Cowpox virus (**Figure 2 C**). All these varying amino acids are located either outside or on the border of structural alpha-helices and beta-sheets and are therefore unlikely to cause a major structural rearrangement of this protein.

We then focused our analysis on H3L, an Orthopoxvirus 324 amino acid long glycosiltransferase located on the viral envelope, which plays a role in the attachment to the human target cell and in the general poxvirus entry [25]. We considered H3L due to its pivotal immunological importance, since it offers one of the main epitopes recognized by the host immune system [26]; studies showed how antibodies elicited by smallpox vaccines are binding H3 proteins [27], with specific recognition of H3L as an antigen by immune CD8+ T cells [27]. As with A42R, we extracted H3L proteins between the two recent Monkeypox genome sequences (PT0001 and ITM001) and compared them with representative database sequences for West African Monkeypox (MT903340.1), Central African Monkeypox (NC_003310.1), Cowpox (KC813504.1) and Variola (NC_001611.1) viruses. Multiple sequence alignment shows a high similarity between H3L sequences, however with a selection of amino acids diverging between major clades (**Figure 3 A**). Phylogenetic analysis based on this single protein confirms what previously inferred from genome-wide analysis, specifically a high similarity between the recent Belgium and Portugal Monkeypox sequences, and their likely origin from the West African clade (**Figure 3 B**). H3L proteins from Central African Monkeypox, Cowpox and Variola viruses are more distantly related to the recent outbreak. 21 amino acids differ between the reference Monkeypox and the reference Smallpox (Variola) virus in H3L, suggesting a higher variance for this protein, with respect to the core protein A42R (**Figure 3 C**). The sequence of this important protein epitope has indeed remained identical in the available sequences form the recent 2022 outbreak (PT0001 and ITM001) when compared to the West African sequences, represented here by the MT903340.1 sequences, retrieved from a 2017 Nigerian sample [8].

Discussion

Our preliminary analysis shows that sequences retrieved from the recent multi-country Monkeypox outbreak, started in May 2022, are likely derived from the West African Monkeypox clade, both from genome-wide (**Figure 1 B**) and protein-specific (**Figure 3 B**) phylogenetic analyses. Our analyses on specific Monkeypox proteins, A42R (selected due to the availability of a crystal structure) and H3L (selected due to its centrality in immune recognition) show in fact complete identity between the recent outbreak and West African Monkeypox sequences (**Figure 2 C** and **Figure 3 C**). Genome-wide sequence identity between all Monkeypox viruses is well above 98% (**Figure 2 B**) and above 95% when considering also the more distantly related Cowpox and Variola viruses. A higher variability (21 amino acids over 324 amino acids, 6.5% of the whole protein

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.01.494368; this version posted June 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sequence) is found between the Variola and the Monkeypox virus H3L protein (**Figure 3**). All these results suggest a low variability within the Orthopoxvirus family, but at the same time the differences between the Variola and Monkeypox viruses in H3L, a host recognition protein, may suggest host affinity and adaption of these viruses. While both pathogens are able to infect different species, Variola virus is an ancient actor in human history and it is considered highly specific to human hosts [28], while Monkeypox has been diagnosed in humans only since 1970 [29].

The roles of H3 in virus/host recognition and in the viral entry process, as well as its high immunogenic potential, allow to observe functional analogies with the Spike protein in coronavirus infections [30]. In the case of Monkeypox H3 can therefore, similarly to Spike, constitute an ideal target for vaccine development, resembling the recent success of Spike mRNA-based vaccines against COVID-19 [31]. The high similarity between H3 in Variola and Monkeypox viruses are one of the likely molecular causes for the efficacy of smallpox vaccines in protecting against monkeypox [15].

While the multi-country monkeypox case number of May 2022 remains the largest ever occurred for this disease outside the endemic regions in Africa [9], it must be noted that thankfully no death related to this outbreak has been reported, and that it is unlikely that this disease will reach pandemic status [32]. Moreover, as shown in our work here, there has been virtually no genetic divergence from previously known endemic Monkeypox viruses (**Figure 1 B**). However, a close and continuous inspection on the evolution of this protein is of paramount importance to monitor the progression of monkeypox, during this outbreak and in the future years, in order to maintain the capability of developing efficacious vaccines.

Methods

Pairwise sequence analysis, including sequence identity calculation, was performed using BLASTN version 2.9.0+ [20]. Sequences were retrieved from the NCBI database [33] and, for the recent outbreak, from the virological.org database [18,19].

Genome-wide multiple sequence alignment was performed using Parsnp version 1.2 [34] with default parameters, using the RefSeq Monkeypox sequence NC_003310.1 as reference. Protein multiple sequence alignment was performed using Muscle with default parameters [35]. Multiple sequence alignment visualization was performed using Jalview version 2.11.2.0 [36].

Phylogenetic trees were constructed using the Neighbor-Joining method [37]. Tree evaluation was performed with 1000 bootstrap replicates [38]. Sequence distances in phylogenetic trees were computed using the Poisson method [39]. All phylogenetic analysis was performed using the MEGA tool, version 11 [40].

Protein structure visualization was performed using the PDB web tools [22]. The structure of Monkeypox protein A42R was obtained from PDB entry 4QWO [23]. All reported amino acid numbers are based on the reference sequence of Central African Monkeypox NC_003310.1 [14].

All sequence preprocessing and the heatmap plot in **Figure 2 A** has been performed using the R statistical software [41].

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.01.494368; this version posted June 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends

Figure 1

A) Schematic representation of the structure of the Monkeypox virus. Reported genome information is based on the reference sequence NC_003310.1, associated with the Central African Monkeypox virus sample Zaire-96-I-96. B) Phylogenetic tree of Monkeypox viruses. Phylogenetic analysis was based on a core alignment including all available Monkeypox viral sequences from 1971 to May 2022. The percentage of bootstraps supporting each branch is reported. The ruler in the phylogenetic tree reports the number of nucleotide substitutions per site. A likely origin from the West African clade is supported for the recent Belgium and Portugal Monkeypox sequences (highlighted in green).

Figure 2

A) Sequence similarity heat map of Monkeypox, Cowpox, and Variola viruses, based on sequence identity. The color intensity in the heatmap reflects the percentage of identity, calculated via the BLASTN algorithm, between genomes in a scale ranging from blue (95% identity) to red (100% identity). The position of the Variola virus is indicated by a black arrow. B) Percentage sequence identity highlight between the Variola virus genome and representative Monkeypox or Cowpox virus genomes. A sequence identity >95% between all highlighted genomes was detected, confirming high sequence similarity between these viruses. C) 3D representation of the A42R Profilin-like Protein from Monkeypox Virus Zaire-96-I-16 (PDB entry 4QWO). The location of the 4 amino acids differing between Monkeypox and other closely related Orthopoxvirus A42R sequences are indicated by red circles.

Figure 3

A) Multiple sequence alignment of the H3L protein encoded by poxvirus. The two Monkeypox virus sequences from recent cases of monkeypox (PT0001 and ITM001) were compared with representative sequences of West African (MT903340.1) and Central African Monkeypox (NC_003310.1), Cowpox (KC813504.1), and Variola (NC_001611.1) viruses. B) Phylogenetic tree based on H3L protein sequences in the previously indicated representative genomes. The percentage of bootstraps supporting each branch is reported, from a total of 1000 bootstraps. The ruler in the phylogenetic tree reports the number of amino acid substitutions per site. C) Coordinates of amino acids differing in the H3L protein sequence between Monkeypox reference NC_003310.1 and other selected Orthopoxviruses. Amino acids corresponding to the reference NC_003310.1 H3L sequence are highlighted in green.

Supplementary files

Supplementary 1: FASTA files. Top 100 BLAST hits. Date reported is collection date. When not specified (e.g., HQ857562.1), the date of submission or publication is indicated.

Supplementary 2: tree with cowpox and monkeypox

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.01.494368; this version posted June 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgments

We sincerely thank the teams of Philippe Selhorts (https://virological.org/t/belgian-case-of-monkeypox-virus-linked-to-outbreak-in-portugal/801) and Joao Paulo Gomes (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) for providing the draft genome sequences of the two recent cases of Monkeypox virus infection, available on the aggregator site virological.org. We also thank Rossana Bartone, Angela Tomasini and Giulia Minarini for their administrative support.

Funding

This research was funded by the CARISBO Foundation, under the Bando Ricerca Medica e Alta Tecnologia 2021 (project 2021.0167); Italian Ministry of University and Research, under the PON "Ricerca e Innovazione" 2014–2020 program; CINECA consortium, project HP10CJH90B.

References

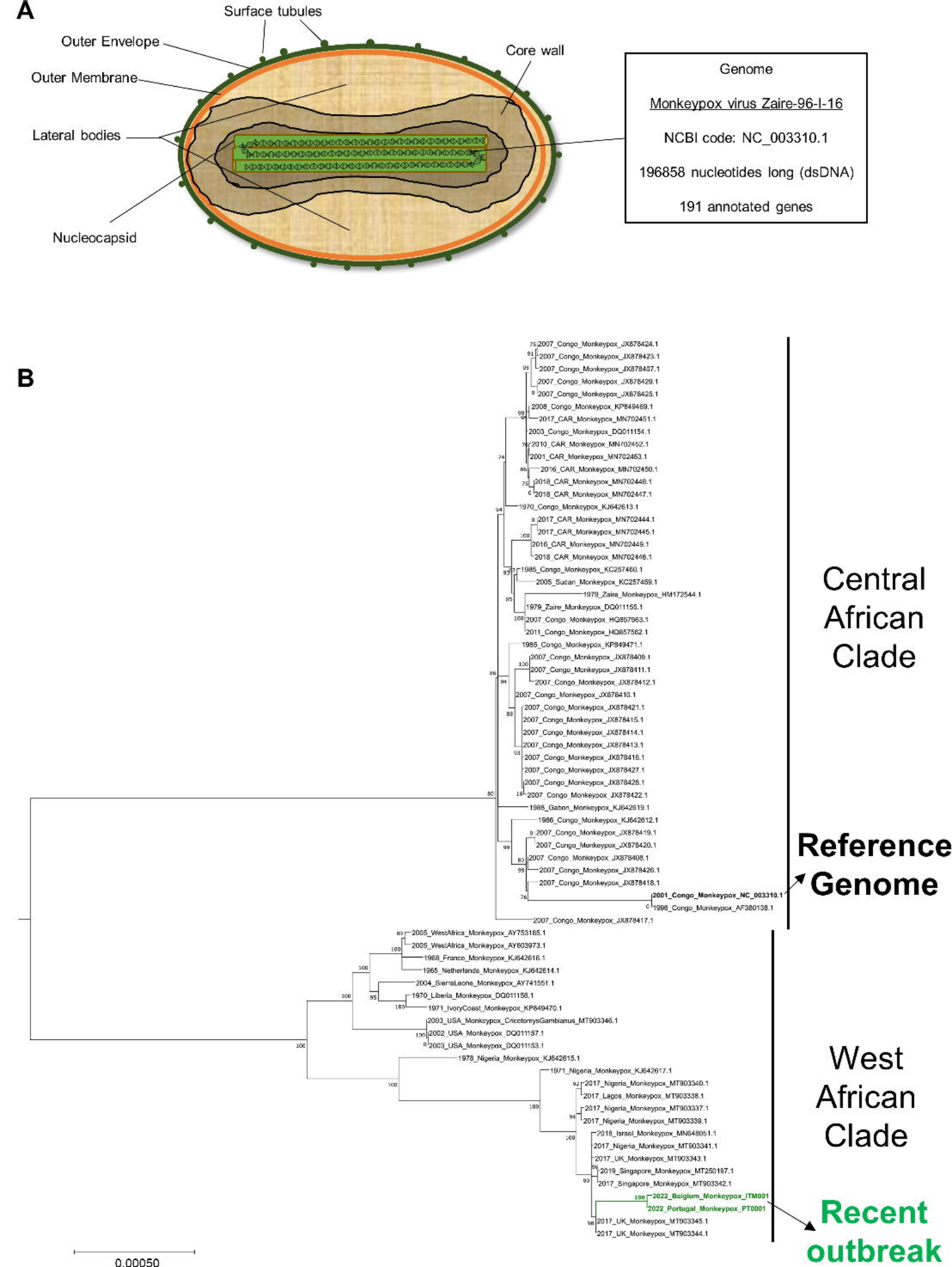
1. Henderson, D.A. The Eradication of Smallpox – An Overview of the Past, Present, and Future. *Vaccine* **2011**, *29*, D7–D9, doi:10.1016/j.vaccine.2011.06.080.
2. McCollum, A.M.; Damon, I.K. Human Monkeypox. *Clinical Infectious Diseases* **2014**, *58*, 260–267, doi:10.1093/cid/cit703.
3. Multi-Country Monkeypox Outbreak in Non-Endemic Countries Available online: https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON385 (accessed on 31 May 2022).
4. Mahase, E. Seven Monkeypox Cases Are Confirmed in England. *BMJ* **2022**, *377*, o1239, doi:10.1136/bmj.o1239.
5. Bunge, E.M.; Hoet, B.; Chen, L.; Lienert, F.; Weidenthaler, H.; Baer, L.R.; Steffen, R. The Changing Epidemiology of Human Monkeypox—A Potential Threat? A Systematic Review. *PLOS Neglected Tropical Diseases* **2022**, *16*, e0010141, doi:10.1371/journal.pntd.0010141.
6. Vaughan, A.; Aarons, E.; Astbury, J.; Brooks, T.; Chand, M.; Flegg, P.; Hardman, A.; Harper, N.; Jarvis, R.; Mawdsley, S.; et al. Human-to-Human Transmission of Monkeypox Virus, United Kingdom, October 2018. *Emerg. Infect. Dis.* **2020**, *26*, 782–785, doi:10.3201/eid2604.191164.
7. Mahase, E. Monkeypox: What Do We Know about the Outbreaks in Europe and North America? *BMJ* **2022**, *377*, o1274, doi:10.1136/bmj.o1274.
8. Mauldin, M.R.; McCollum, A.M.; Nakazawa, Y.J.; Mandra, A.; Whitehouse, E.R.; Davidson, W.; Zhao, H.; Gao, J.; Li, Y.; Doty, J.; et al. Exportation of Monkeypox Virus From the African Continent. *J Infect Dis* **2022**, *225*, 1367–1376, doi:10.1093/infdis/jiaa559.
9. Velavan, T.P.; Meyer, C.G. Monkeypox 2022 Outbreak: An Update. *Trop Med Int Health* **2022**, doi:10.1111/tmi.13785.
10. Karumathil, S.; Raveendran, N.T.; Ganesh, D.; Kumar NS, S.; Nair, R.R.; Dirisala, V.R. Evolution of Synonymous Codon Usage Bias in West African and Central African Strains of Monkeypox Virus. *Evol Bioinform Online* **2018**, *14*, 1176934318761368, doi:10.1177/1176934318761368.
11. Chen, N.; Li, G.; Liszewski, M.K.; Atkinson, J.P.; Jahrling, P.B.; Feng, Z.; Schriewer, J.; Buck, C.; Wang, C.; Lefkowitz, E.J.; et al. Virulence Differences between Monkeypox Virus Isolates from West Africa and the Congo Basin. *Virology* **2005**, *340*, 46–63, doi:10.1016/j.virol.2005.05.030.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.01.494368; this version posted June 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

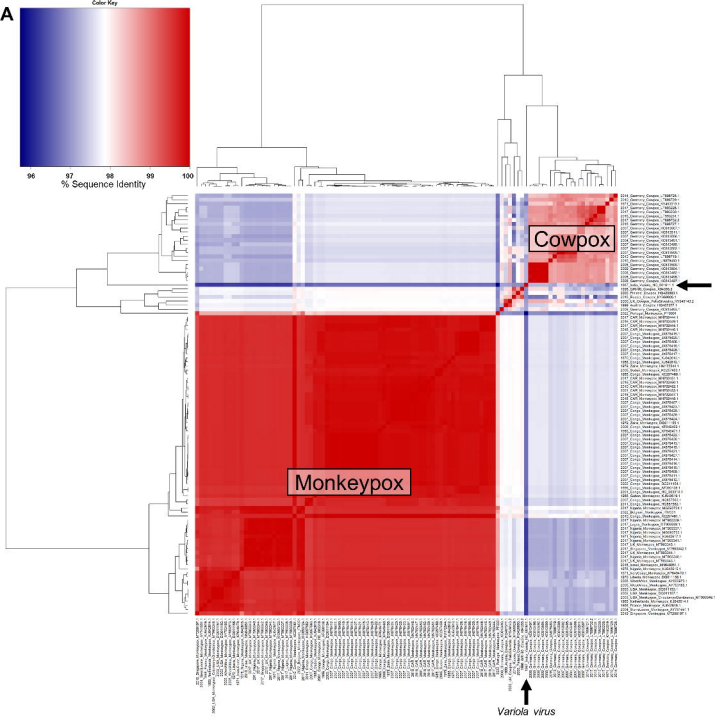
12. Douglass, N.; Dumbell, K. Independent Evolution of Monkeypox and Variola Viruses. *J Virol* **1992**, *66*, 7565–7567, doi:10.1128/JVI.66.12.7565-7567.1992.
13. Shchelkunov, S.N.; Safronov, P.F.; Totmenin, A.V.; Petrov, N.A.; Ryazankina, O.I.; Gutorov, V.V.; Kotwal, G.J. The Genomic Sequence Analysis of the Left and Right Species-Specific Terminal Region of a Cowpox Virus Strain Reveals Unique Sequences and a Cluster of Intact ORFs for Immunomodulatory and Host Range Proteins. *Virology* **1998**, *243*, 432–460, doi:10.1006/viro.1998.9039.
14. Shchelkunov, S.N.; Totmenin, A.V.; Babkin, I.V.; Safronov, P.F.; Ryazankina, O.I.; Petrov, N.A.; Gutorov, V.V.; Uvarova, E.A.; Mikheev, M.V.; Sisler, J.R.; et al. Human Monkeypox and Smallpox Viruses: Genomic Comparison. *FEBS Lett* **2001**, *509*, 66–70, doi:10.1016/s0014-5793(01)03144-1.
15. Fine, P.E.; Jezek, Z.; Grab, B.; Dixon, H. The Transmission Potential of Monkeypox Virus in Human Populations. *Int J Epidemiol* **1988**, *17*, 643–650, doi:10.1093/ije/17.3.643.
16. Kugelman, J.R.; Johnston, S.C.; Mulembakani, P.M.; Kisalu, N.; Lee, M.S.; Koroleva, G.; McCarthy, S.E.; Gestole, M.C.; Wolfe, N.D.; Fair, J.N.; et al. Genomic Variability of Monkeypox Virus among Humans, Democratic Republic of the Congo. *Emerg. Infect. Dis.* **2014**, *20*, doi:10.3201/eid2002.130118.
17. Chen, Z.; Azman, A.S.; Chen, X.; Zou, J.; Tian, Y.; Sun, R.; Xu, X.; Wu, Y.; Lu, W.; Ge, S.; et al. Global Landscape of SARS-CoV-2 Genomic Surveillance and Data Sharing. *Nat Genet* **2022**, *54*, 499–507, doi:10.1038/s41588-022-01033-y.
18. First Draft Genome Sequence of Monkeypox Virus Associated with the Suspected Multi-Country Outbreak, May 2022 (Confirmed Case in Portugal) - Monkeypox / Genome Reports - Virological Available online: 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 (accessed on 31 May 2022).
19. Belgian Case of Monkeypox Virus Linked to Outbreak in Portugal - Monkeypox / Genome Reports Available online: https://virological.org/t/belgian-case-of-monkeypox-virus-linked-to-outbreak-in-portugal/801 (accessed on 31 May 2022).
20. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J Mol Biol* **1990**, *215*, 403–410, doi:10.1016/S0022-2836(05)80360-2.
21. Shchelkunov, S.N.; Totmenin, A.V.; Sandakhchiev, L.S. Analysis of the Nucleotide Sequence of 23.8 Kbp from the Left Terminus of the Genome of Variola Major Virus Strain India-1967. *Virus Res* **1996**, *40*, 169–183, doi:10.1016/0168-1702(95)01269-9.
22. Burley, S.K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chen, L.; Crichlow, G.V.; Christie, C.H.; Dalenberg, K.; Di Costanzo, L.; Duarte, J.M.; et al. RCSB Protein Data Bank: Powerful New Tools for Exploring 3D Structures of Biological Macromolecules for Basic and Applied Research and Education in Fundamental Biology, Biomedicine, Biotechnology, Bioengineering and Energy Sciences. *Nucleic Acids Research* **2021**, *49*, D437–D451, doi:10.1093/nar/gkaa1038.
23. Bank, R.P.D. RCSB PDB - 4QWO: 1.52 Angstrom Crystal Structure of A42R Profilin-like Protein from Monkeypox Virus Zaire-96-I-16 Available online: https://www.rcsb.org/structure/4QWO (accessed on 31 May 2022).
24. Gupta, A.; Sabarinathan, R.; Bala, P.; Donipadi, V.; Vashisht, D.; Katika, M.R.; Kandakatla, M.; Mitra, D.; Dalal, A.; Bashyam, M.D. A Comprehensive Profile of Genomic Variations in the SARS-CoV-2 Isolates from the State of Telangana, India. *J Gen Virol* **2021**, *102*, 001562, doi:10.1099/jgv.0.001562.
25. Moss, B. Poxvirus Entry and Membrane Fusion. *Virology* **2006**, *344*, 48–54, doi:10.1016/j.virol.2005.09.037.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.01.494368; this version posted June 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

26. Singh, K.; Gittis, A.G.; Gitti, R.K.; Ostazeski, S.A.; Su, H.-P.; Garboczi, D.N. The Vaccinia Virus H3 Envelope Protein, a Major Target of Neutralizing Antibodies, Exhibits a Glycosyltransferase Fold and Binds UDP-Glucose. *J Virol* **2016**, *90*, 5020–5030, doi:10.1128/JVI.02933-15.
27. Drexler, I.; Staib, C.; Kastenmuller, W.; Stevanović, S.; Schmidt, B.; Lemonnier, F.A.; Rammensee, H.-G.; Busch, D.H.; Bernhard, H.; Erfle, V.; et al. Identification of Vaccinia Virus Epitope-Specific HLA-A*0201-Restricted T Cells and Comparative Analysis of Smallpox Vaccines. *Proc Natl Acad Sci U S A* **2003**, *100*, 217–222, doi:10.1073/pnas.262668999.
28. Riedel, S. Edward Jenner and the History of Smallpox and Vaccination. *Proc (Bayl Univ Med Cent)* **2005**, *18*, 21–25.
29. Breman, J.G.; Kalisa-Ruti, null; Steniowski, M.V.; Zanotto, E.; Gromyko, A.I.; Arita, I. Human Monkeypox, 1970-79. *Bull World Health Organ* **1980**, *58*, 165–182.
30. Liu, Z.; Xiao, X.; Wei, X.; Li, J.; Yang, J.; Tan, H.; Zhu, J.; Zhang, Q.; Wu, J.; Liu, L. Composition and Divergence of Coronavirus Spike Proteins and Host ACE2 Receptors Predict Potential Intermediate Hosts of SARS-CoV-2. *Journal of Medical Virology* **2020**, *92*, 595–601, doi:10.1002/jmv.25726.
31. Bettini, E.; Locci, M. SARS-CoV-2 MRNA Vaccines: Immunological Mechanism and Beyond. *Vaccines* **2021**, *9*, 147, doi:10.3390/vaccines9020147.
32. Zumla, A.; Valdoleiros, S.R.; Haider, N.; Asogun, D.; Ntoumi, F.; Petersen, E.; Kock, R. Monkeypox Outbreaks Outside Endemic Regions: Scientific and Social Priorities. *The Lancet Infectious Diseases* **2022**, *0*, doi:10.1016/S1473-3099(22)00354-1.
33. Gibney, G.; Baxevanis, A.D. Searching NCBI Databases Using Entrez. *Current Protocols in Bioinformatics* **2011**, *34*, 1.3.1-1.3.25, doi:10.1002/0471250953.bi0103s34.
34. Treangen, T.J.; Ondov, B.D.; Koren, S.; Phillippy, A.M. The Harvest Suite for Rapid Core-Genome Alignment and Visualization of Thousands of Intraspecific Microbial Genomes. *Genome Biol* **2014**, *15*, 524, doi:10.1186/s13059-014-0524-x.
35. Edgar, R.C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res* **2004**, *32*, 1792–1797, doi:10.1093/nar/gkh340.
36. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics* **2009**, *25*, 1189–1191, doi:10.1093/bioinformatics/btp033.
37. Saitou, N.; Nei, M. The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol Biol Evol* **1987**, *4*, 406–425, doi:10.1093/oxfordjournals.molbev.a040454.
38. Felsenstein, J. CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP. *Evolution* **1985**, *39*, 783–791, doi:10.1111/j.1558-5646.1985.tb00420.x.
39. Zuckerkandl, E.; Pauling, L. Evolutionary Divergence and Convergence in Proteins. In *Evolving Genes and Proteins*; Bryson, V., Vogel, H.J., Eds.; Academic Press, 1965; pp. 97–166 ISBN 978-1-4832-2734-4.
40. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* **2021**, *38*, 3022–3027, doi:10.1093/molbev/msab120.
41. Giorgi, F.M.; Ceraolo, C.; Mercatelli, D. The R Language: An Engine for Bioinformatics and Data Science. *Life (Basel)* **2022**, *12*, 648, doi:10.3390/life12050648.


A
Surface tubules
Outer Envelope
Core wall
Outer Membrane
Lateral bodies
Nucleocapsid
Genome
Monkeypox virus Zaire-96-I-16
NCBI code: NC_003310.1
196858 nucleotides long (dsDNA)
191 annotated genes
B
Central African Clade
Reference Genome
West African Clade
Recent outbreak
2001_Congo_Monkeypox_NC_003310.1
2022_Belgium_Monkeypox_ITM001
2022_Portugal_Monkeypox_PT0001
0.00050

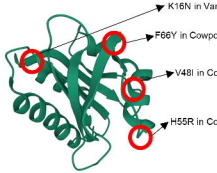
A



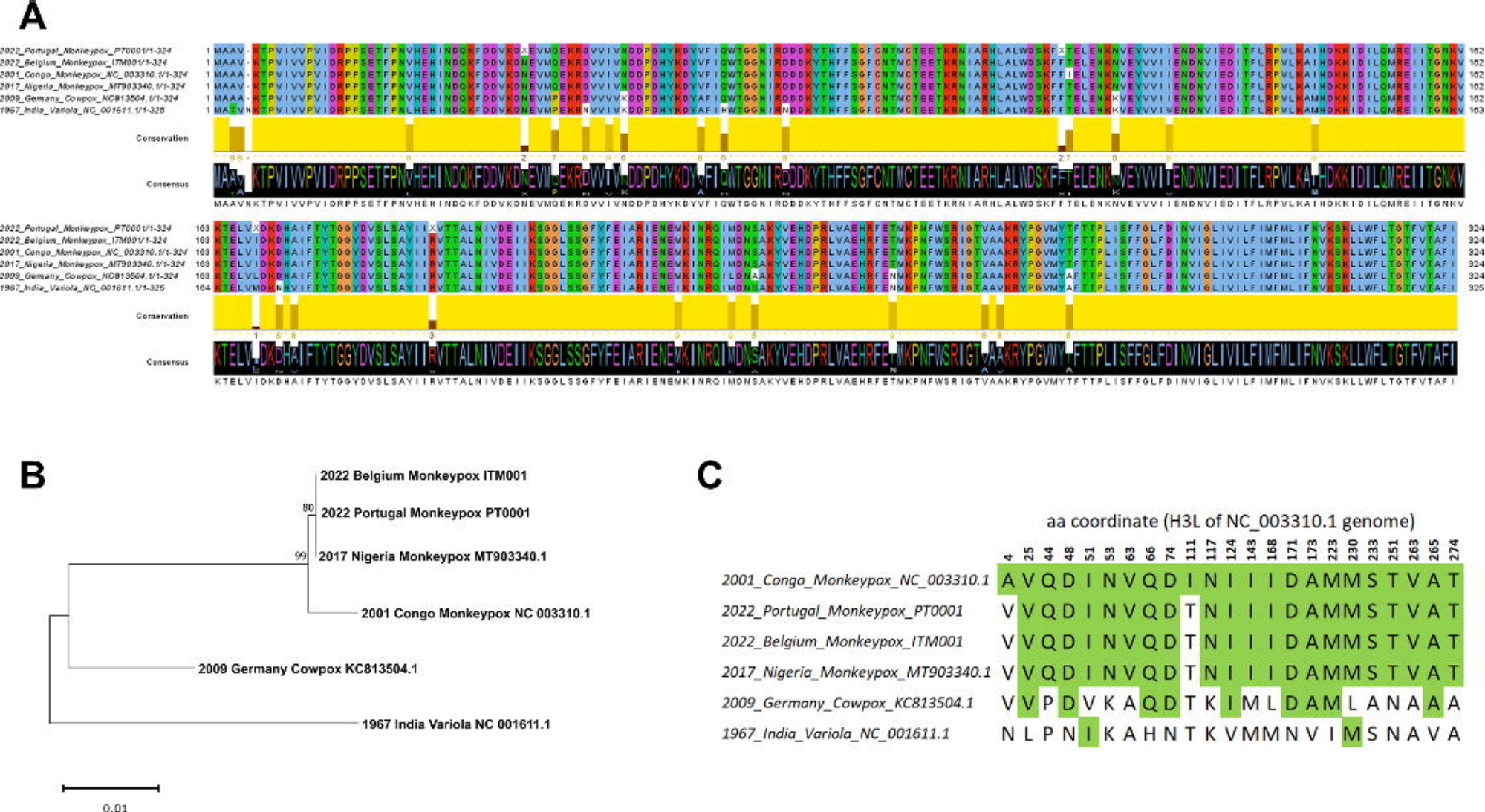
B

	2022_Portugal_Monkeypox_PT0001	2022_Belgium_Monkeypox_ITM001	2001_Congo_Monkeypox_NC_003310.1	2017_Nigeria_Monkeypox_MT903340.1	2009_Germany_Cowpox_KC813504.1	1967_India_Variola_NC_001611.1
2022_Portugal_Monkeypox_PT0001	100	98.608	98.274	98.637	96.907	95.737
2022_Belgium_Monkeypox_ITM001	98.608	100	99.554	99.905	97.735	96.792
2001_Congo_Monkeypox_NC_003310.1	98.274	99.554	100	99.574	97.636	96.621
2017_Nigeria_Monkeypox_MT903340.1	98.637	99.905	99.574	100	97.171	96.243
2009_Germany_Cowpox_KC813504.1	96.907	97.735	97.636	97.171	100	97.258
1967_India_Variola_NC_001611.1	95.737	96.792	96.621	96.243	97.258	100

C



Monkeypox NC_003310.1
A42R protein
PDB entry 4QWO


A
2022_Portugal_Monkeypox_PT0001/1-324
2022_Belgium_Monkeypox_ITM001/1-324
2001_Congo_Monkeypox_NC_003310.1/1-324
2017_Nigeria_Monkeypox_MT903340.1/1-324
2009_Germany_Cowpox_KC813504.1/1-324
1967_India_Variola_NC_001611.1/1-325
Conservation
Consensus
B
2022 Belgium Monkeypox ITM001
2022 Portugal Monkeypox PT0001
2017 Nigeria Monkeypox MT903340.1
2001 Congo Monkeypox NC 003310.1
2009 Germany Cowpox KC813504.1
1967 India Variola NC 001611.1
80
99
0.01
C
aa coordinate (H3L of NC_003310.1 genome)
4 25 44 48 51 53 63 66 74 111 117 124 143 168 171 173 223 230 233 251 263 265 274
2001_Congo_Monkeypox_NC_003310.1 A V Q D I N V Q D I N I I I D A M M S T V A T
2022_Portugal_Monkeypox_PT0001 V V Q D I N V Q D T N I I I D A M M S T V A T
2022_Belgium_Monkeypox_ITM001 V V Q D I N V Q D T N I I I D A M M S T V A T
2017_Nigeria_Monkeypox_MT903340.1 V V Q D I N V Q D T N I I I D A M M S T V A T
2009_Germany_Cowpox_KC813504.1 V V P D V K A Q D T K I M L D A M L A N A A A
1967_India_Variola_NC_001611.1 N L P N I K A H N T K V M M N V I M S N A V A