Molecular and phylogenetic characterization of the monkeypox outbreak in the South of Spain

Carlos S. Casimiro-Soriguer[1,2+], Javier Perez-Florido[1,2,+], Maria Lara[1], Pedro Camacho-Martinez[3], Laura Merino-Diaz[3], Inmaculada Pupo-Ledo[3], Adolfo de Salazar[4,5,6], Ana Fuentes[4,6], Laura Viñuela[4,5], Natalia Chueca[4,6], Luis Martinez-Martinez[4,7,8,9], Nicola Lorusso[10], The Andalusian genomic surveillance network[**], Jose A Lepe[2,3,4,*], Joaquín Dopazo[1,2,11,*], Federico Garcia[4,5,6,*]

1 Computational Medicine Platform, Andalusian Public Foundation Progress and Health-FPS, Sevilla, 41013, Spain.

2 Institute of Biomedicine of Seville, IBiS, University Hospital Virgen del Rocío/CSIC/University of Seville, Spain.

3 Servicio de Microbiología. Unidad Clínica Enfermedades Infecciosas, Microbiología y Medicina Preventiva. Hospital Universitario Virgen del Rocío. 41013. Sevilla. Spain.

4 Centro de Investigación Biomédica en Red en Enfermedades Infecciosas (CIBERINFEC), ISCIII, Madrid, Spain

5 Servicio de Microbiología. Hospital Universitario San Cecilio. 18016, Granada, Spain.

6 Instituto de Investigación Biosanitaria, ibs.GRANADA, 18012, Granada, Spain

7 Unit of Microbiology, University Hospital Reina Sofia, Cordoba, Spain

8 Departamento de Química Agrícola, Edafología y Microbiología, Universidad de Córdoba, Córdoba, Spain.

9 Instituto Maimónides de Investigación Biomédica de Córdoba (IMIBIC), Córdoba, Spain.

10 Dirección General de Salud Pública. Consejería de Salud y Familias. Junta de Andalucía. 41020, Sevilla, Spain

11 The members of the network are listed under Collaborators

12 FPS/ELIXIR-ES, Andalusian Public Foundation Progress and Health-FPS. 41013. Sevilla. Spain

+ These authors contributed equally to this work and share first authorship.

* These authors contributed equally to this work and share last authorship.

** The Andalusian Genomic Surveillance Network: Teresa Cabezas, Hospital Torrecárdenas (Almería); Pilar Luzón, Hospital Poniente (Almería); Carolina Roldan, Hospital Jaén (Jaén); Purificación Cantudo, Hospital Linares (Jaén); Fernando Fernández, Hospital Costa del Sol (Málaga); Isabel Viciana, Hospital Virgen de la Victoria (Málaga); Federico Acosta, Hospital Antequera (Málaga); Inmaculada De Toro, Hospital Carlos Haya (Málaga); Maria del Mar Gallardo-Garcia, Hospital Motril (Granada); Francisco Franco Alvarez de Luna, Hospital Juan Ramon Jimenez (Huelva); Miguel Angel Castaño, Hospital Infanta Elena (Huelva); Gema Varo, Hospital de Riotinto (Huelva); Manuel Rodriguez-Iglesias, Hospital Puerta del Mar (Cádiz); Maria Carmen Martinez-Rubio, Hospital de Puerto Real (Cádiz); Maria Dolores Prieto, Hospital de Jerez (Cádiz); Berta Becerril, Hospital Punta Europa (Cádiz); Ainara Rodriguez, Hospital la Línea (Cádiz); Félix Gascón, Hospital Pozoblanco (Córdoba); Alejandro Peña, Hospital Infanta Margarita (Córdoba); Mónica Chaves, Hospital San Juan de Dios (Córdoba); Samuel Bernal, Hospital de Valme (Sevilla); Álvaro Pascual, Hospital Virgen Macarena (Sevilla); Esther Roldán, Hospital La Merced (Sevilla)

Abstract

Until the May 2022 Monkeypox outbreak, which spread rapidly to many non-endemic countries, the virus was considered a viral zoonosis limited to some African countries. The Andalusian circuit of genomic surveillance was rapidly applied to characterize the Monkeypox outbreak in the South of Spain. Whole genome sequencing was used to obtain the genomic profiles of samples collected across the south of Spain, representative of all the provinces of Andalusia. Phylogenetic analysis was used to study the relationship of the isolates and the available sequences of the 2022 outbreak. Whole genome sequencing of a total of 160 monkeypox viruses from the different provinces that reported cases were obtained. Interestingly, we report the sequences of monkeypox viruses obtained from two patients who died. While one of the isolates bore no noteworthy mutations that explain a potential heightened virulence, in another patient the second consecutive genome sequence, performed after the administration of tecovirimat, uncovered a mutation within the A0A7H0DN30 gene, known to be a prime target for tecovirimat in its Vaccinia counterpart. In general, a low number of mutations were observed in the sequences reported, which were very similar to the reference of the 2022 outbreak (OX044336), as expected from a DNA virus. The samples likely correspond to several introductions of the circulating monkeypox viruses from the last outbreak. The virus sequenced from one of the two patients that died presented a mutation in a gene that bears potential connections to drug resistance. This mutation was absent in the initial sequencing prior to treatment.

Introduction

Monkeypox (MPXV) is a viral zoonosis endemic in some West and Central African countries and with few cases outside Africa. In May 2022, an unexpectedly large MPXV clade B.1 outbreak affecting a considerable number of non-endemic countries was reported [1, 2]. After the first autochthonous cases were reported in the UK on May 13th and in Spain on May 17th, a rapid spread to more than 14,000 cases were reported in more than 60 countries only in the first two months of the outbreak, that summed up to more than 88,000 cases worldwide as of June 2023 [3]. Although the incidence has drastically reduced in 2023 [3], the observed simultaneous MPXV incidence in different countries due to a rapid cross-border transmission [4] poses a real threat that must be addressed by robust public health surveillance and control measures [5]. Moreover, further research is required to delve deeper into the origins of the recent outbreak, investigating potential factors such as animal reservoirs, human behavior, or viral mutations that might be driving its occurrence. [5]. In this context, genomic monitoring of the MPXV samples sequenced from the epidemiologic surveillance in Andalusia results crucial from the epidemiological point of view, that assigns clearly the Andalusian sequences to the circulating clade. Moreover, whole-genome virus sequencing has also a relevant role in monitoring polymorphisms, as well as in detecting gene losses based on possible intragenic frameshifts or premature stop codons that could appear locally and might be relevant as virulence or enhanced transmissibility determinants.

Results

The Andalusian outbreak

The phylogeny of the 2022 MPXV outbreak in the context of the rest of available MPXV sequences, as displayed in the Genomic Surveillance Circuit of Andalusia [6, 7], is depicted in Figure 1. The specific features of the sequences of this outbreak and the apparently fast evolution with respect to previous outbreaks has already been discussed [8]. As expected from a DNA virus, with a relatively low introduction into the general population, the samples isolated in Andalusia have only a few mutations with respect to the rest of MPXV isolates in the outbreak. Andalusian isolates are scattered across the outbreak branch in the phylogeny and are related to sequences from other countries, suggesting different introductions of the virus in Andalusia.

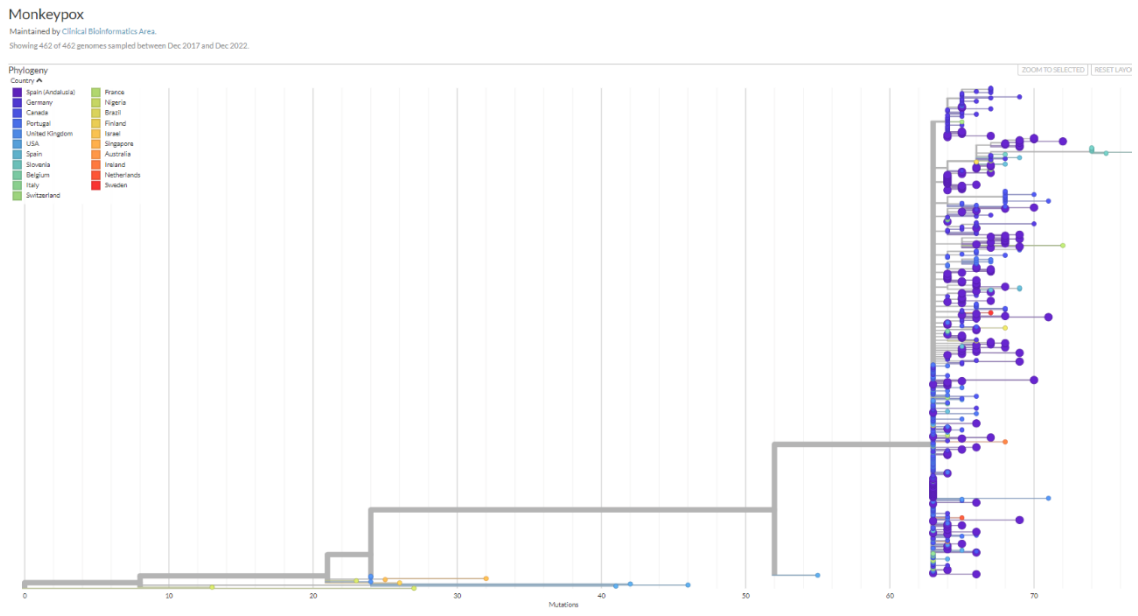


Figure 1. Phylogeny obtained with Nexstrain of the 2022 MPXV outbreak along with the rest of MPXV sequences available.

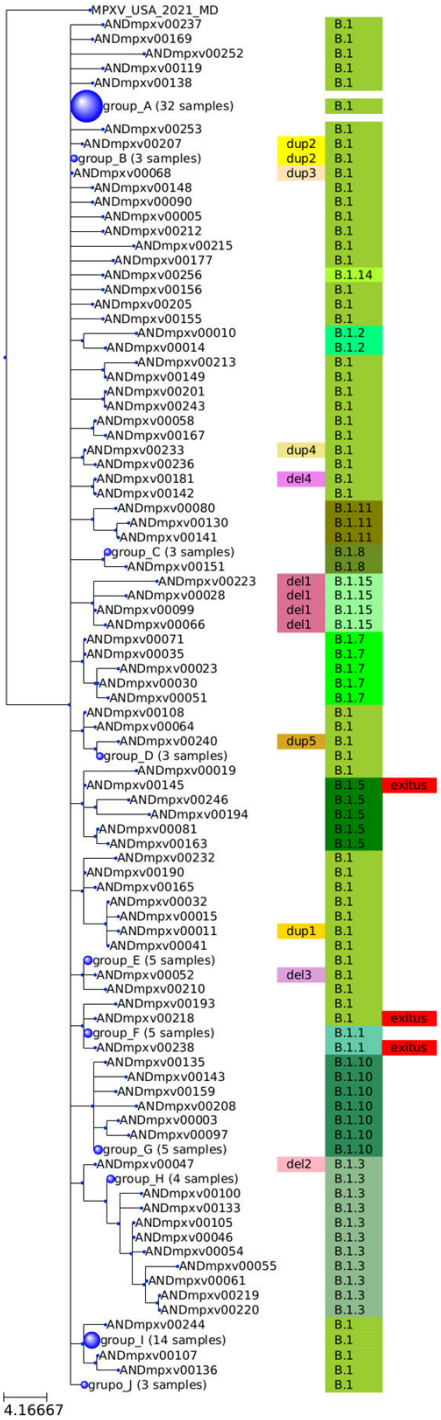
Mutational spectrum of the Andalusian outbreak

Figure 2 portrays a detail of the phylogeny of the MPXV isolated in Andalusia (See Supplementary Figure S1 for a more detailed picture). In addition, it is worth analyzing in detail some non-synonymous mutations that appear specifically in the Andalusian samples. Supplementary Table S2 shows the non-synonymous mutations found in the samples under study (Supplementary Table S1) with respect to the reference ON563414 with the genomic coordinates of NC_063383 [9]. The most common mutation occurs in 14 Andalusian MPXV isolates in the protein *A0A7H0DNG7*, which belongs to the Bcl-2-like protein family, which function as immunomodulators to evade the host innate immune response through the inhibition of apoptosis or blocking the activation of pro-inflammatory transcription factors [10]. Another frequent mutation shared by 13 Andalusian MPXV occurs in *A0A7H0DN47,* a transmembrane protein of unknown function. Other proteins that have been found mutated in 11 MPXV isolates are *A0A7H0DN82*, an envelope protein which has been described as a late gene transcription factor VLTF-4 [11] and *M1LBQ5*, shared by 11 Andalusian MPXV isolates, which is part of a large complex required for early virion morphogenesis [12]. Also, *A0A7H0DN66*, a component of the entry fusion complex (EFC), which consists of 11 proteins and mediates entry of the virion core into the host cytoplasm and *A0A7H0DNF5*, a soluble interferon-gamma receptor-like protein, were found mutated in 8 and 6 Andalusian MPXV isolates respectively. Thus, some differences in the immune response or in the viral replication could characterize currently circulating Andalusian isolates. There are also 96 more mutations, most of them private mutations of specific MPXV isolates and a few of them shared by up to 5 isolates as much (See Supplementary Table S2). Among them, it is worth mentioning those found in genes *A0A7H0DNG4* (in

129 one isolate) and *A0A7H0DNG6* (in four isolates), which were previously identified as

130 virulence genes B19R and B21R, respectively, by comparing isolates of two outbreaks in

131 Nigeria with different mortality rates [11].



132

133 Figure 2. Summarized phylogeny of the Andalusian isolates, with identical sequences grouped in nodes

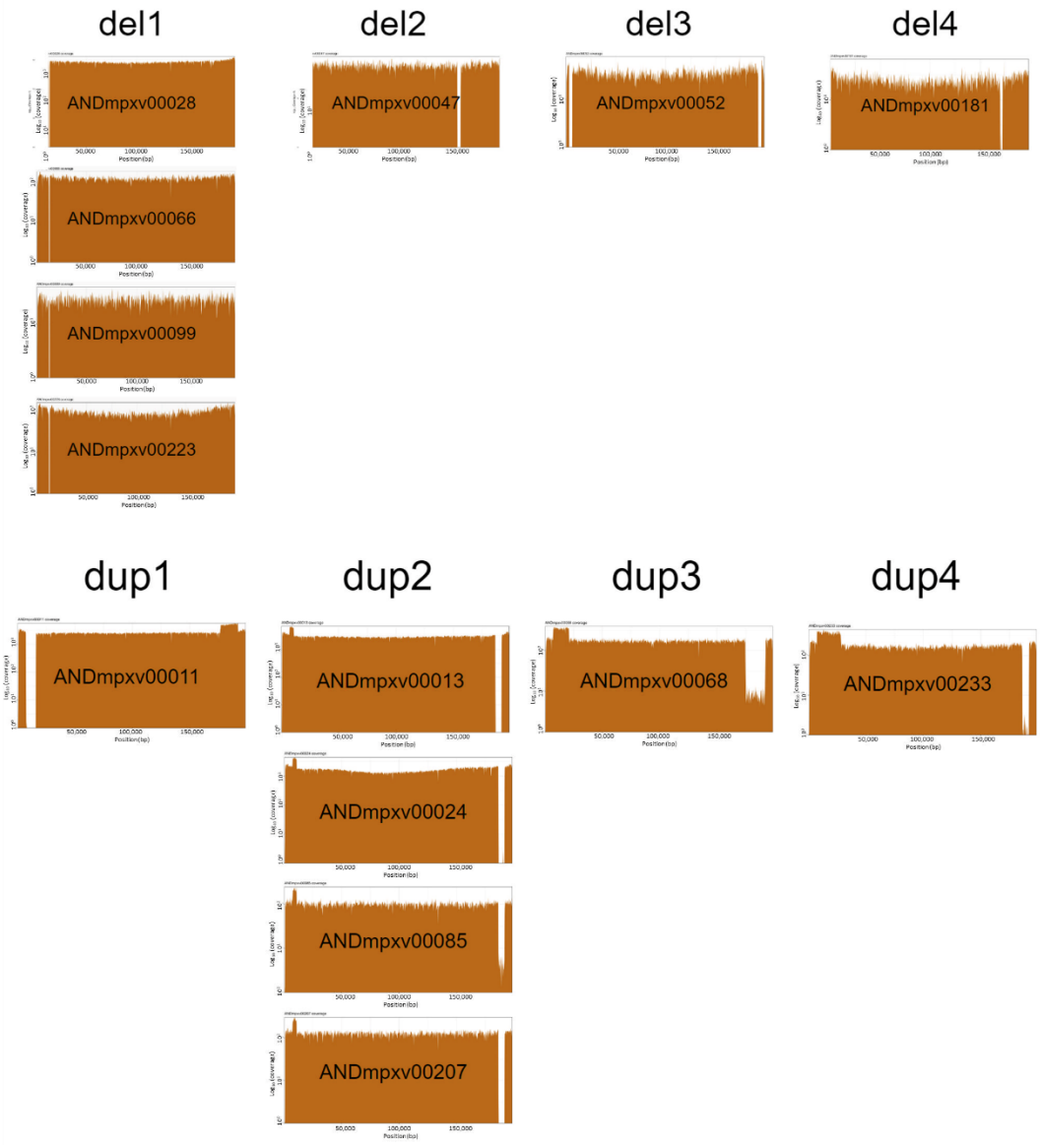
134 with a size proportional to the number of corresponding isolates. Each sequence is labeled with its clade

(B.1 and derived clades), and, in case of harboring a structural variation, it is also labeled. The three sequences, corresponding to the two deceased patients are labeled as well. The sequences in the nodes are: group_A: ANDmpxv00247, ANDmpxv00216, ANDmpxv00029, ANDmpxv00128, ANDmpxv00152, ANDmpxv00103, ANDmpxv00245, ANDmpxv00249, ANDmpxv00161, ANDmpxv00222, ANDmpxv00171, ANDmpxv00248, ANDmpxv00110, ANDmpxv00007, ANDmpxv00157, ANDmpxv00084, ANDmpxv00016, ANDmpxv00164, ANDmpxv00217, ANDmpxv00255, ANDmpxv00178, ANDmpxv00025, ANDmpxv00139, ANDmpxv00184, ANDmpxv00188, ANDmpxv00033, ANDmpxv00027, ANDmpxv00096, ANDmpxv00250, ANDmpxv00140, ANDmpxv00241, ANDmpxv00060; group_B: ANDmpxv00024, ANDmpxv00013, ANDmpxv00085; group_C: ANDmpxv00196, ANDmpxv00235, ANDmpxv00214; group_D: ANDmpxv00132, ANDmpxv00067, ANDmpxv00146]; group_E: ANDmpxv00075, ANDmpxv00022, ANDmpxv00048, ANDmpxv00076, ANDmpxv00123], group_F: ANDmpxv00124, ANDmpxv00154, ANDmpxv00153, ANDmpxv00021, ANDmpxv00254; group_G: ANDmpxv00251, ANDmpxv00056, ANDmpxv00095, ANDmpxv00121, ANDmpxv00020; group_H: ANDmpxv00069, ANDmpxv00162, ANDmpxv00137, ANDmpxv00017; group_I: ANDmpxv00018, ANDmpxv00074, ANDmpxv00077, ANDmpxv00126, ANDmpxv00092, ANDmpxv00242, ANDmpxv00089, ANDmpxv00098, ANDmpxv00008, ANDmpxv00172, ANDmpxv00031, ANDmpxv00147, ANDmpxv00150, ANDmpxv00012; group_J: ANDmpxv00175, ANDmpxv00102, ANDmpxv00053.

Structural variation spectrum of the Andalusian outbreak

Among the 160 sequenced samples analyzed, a total of 15 isolates displayed structural variations (see Figure 3). Within them, 7 isolates exhibited distinct types of deletions, with sizes ranging from 912 to 6472 base pairs. Notably, the most prevalent deletion type, observed in 4 isolates, involved the partial deletion of the *A0A7H0DMZ9* protein, specifically affecting the region spanning 12143-13055 nucleotides (see del1 label in Figure 2). This protein mediates the ubiquitination and subsequent proteasomal degradation of NF-kappa-B by targeting NF-kappa-B RELA subunit to the SCF E3 ligase

161 complex. Ubiquitination and proteasomal degradation are cellular mechanisms that are
162 known to be targeted or modulated by some viruses to manipulate host cell signaling
163 pathways, evade immune responses, or regulate viral protein stability [13]. Other 8
164 isolates bear genomic rearrangement in which one terminal part of the genome was
165 deleted and replaced by an inverted duplication of the other end of the genome. Four
166 of these isolates exhibit a deletion of the 3' region and an inverted duplication of the 5'
167 region. This rearrangement causes the partial loss of the *A0A7H0DNG6* protein,
168 mentioned above as related to virulence.



169

Figure 3. Coverage plots representing the different structural variants found.

All the structural variants seem to have appeared de novo at different points of the monkeypox phylogeny, and only in the case of del1 and dup2 (see Figure 2) seems to have configured clusters of monkeypox isolates sharing the specific variant.

Monkeypox virus from deceased patients

Here we describe the cases of two deceased patients, represented by the sequence ANDmpxv00145, isolated from a patient with no comorbidities, and the sequences ANDmpxv00218 and ANDmpxv00238, which are two consecutive sequencing samples from an immunocompromised patient.

In the first case, the isolate ANDmpxv00145 did not present any remarkable mutation that justifies a higher virulence and, actually, it is identical to the reference sequence, except for a nucleotide synonymous mutation (C70780T) in the gene *A0A7H0DN61*, a putative nuclease. Considering both the synonymous nature of the mutation and the functional role of the implicated gene, it is improbable that this mutation alone could contribute to increased virulence in the isolate. Moreover, this synonymous mutation is also present in other two Andalusian MPXV isolates (ANDmpxv00081 and ANDmpxv00163), as well as in isolates from other countries such as Portugal, Italy, Switzerland, and Germany, all of which belong to the same clade as ANDmpxv00145. It has also been found in another Andalusian isolate (ANDmpxv00019) as a private mutation (Figure 2). To the best of our knowledge, no instances of mortality or severe complications have been documented in association with any of these isolates, thus substantiating the neutral nature of this mutation.

On the other hand, the second patient, who exhibited immunocompromised status, underwent sequencing on two occasions. During the initial sequencing, prior to the administration of tecovirimat antiviral treatment, two mutations were identified: OPG094:R194H and OPG205:E452K. The first mutation was observed in the A0A7H0DN66 gene, and has also been found in several Andalusian isolates with no apparent pathological phenotype. However, in the second sequencing conducted after tecovirimat treatment, the OPG205:E452K mutation was no longer detected, while the OPG057:A290V mutation emerged. Interestingly, this latter mutation was observed in the A0A7H0DN30 gene, whose homologous gene in Vaccinia, that present a high similarity of 99.46%, has been recognized as a target for tecovirimat [14].

On the other hand, no deletions or rearrangements of any kind of structural variation were found in any of the isolates from both deceased patients.

Conclusions

The genomic changes detected in this study are important in assessing the microevolution of the circulating virus, although the functional impact of these mutations is still difficult to assess in the general context of virus circulation. In conclusion, the genomic surveillance platform currently running in Andalusia, created as a response of the COVID-19 pandemic, has enabled an extremely rapid response to monitor the spread and evolution of MPXV in the region, and to contribute to national and international genomic surveillance, in order to provide data and knowledge to monitor MPXV epidemics. Specific genomic surveillance criteria must be established at the national and international levels to optimize resources and increase the usefulness of its results for the control of MPXV transmission.

Methods

Samples

Since the MPXV appeared for the first time in the South of Spain (Andalusia), in May 26th, a few days later than the first report in Spain [15], until the end of December a total of 160 MPXV complete genomes were sequenced in the Genomic Surveillance Circuit of Andalusia [6, 7]. Supplementary Table S1 lists the viral genomes sequenced. The isolates were evenly sampled across all Andalusia.

Sequencing method

Prior to DNA extraction from ulcerative lesion samples using the QIAamp DNA kit (Qiagen), sonication and DNAse/RNAse treatment was performed [8]. Subsequently, shotgun metagenomics was performed. In brief, DNA libraries were prepared using the Illumina DNA Prep kit (Illumina, San Diego, CA, USA) and IDT for Illumina DNA/RNA UD Indexes sets (Ilumina). The quality of the libraries was validated by Qubit 4 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Sequencing was performed on Nextseq 550/1000 (Illumina).

Data processing

Sequencing data were analyzed using in-house scripts and the nf-core/viralrecon pipeline software [16], version 2.4.1. Briefly, after read quality filtering, sequences for each sample are aligned to the high quality MPXV isolate OX044336.2 [17] related to the 2022 outbreak using bowtie2 algorithm [18]. Genomic variants were identified through iVar software [19], using a minimum allele frequency threshold of 0.25 for calling variants and a filtering step to keep variants with a minimum allele frequency threshold

of 0.75. Using the set of high confidence variants and the OX044336.2 genome, a consensus genome per sample was finally built using bcftools [20].

Phylogenetic analysis

Phylogenetic analysis was carried out on the obtained MPXV genomes in the context of a world-wide representative set of MPXV genomes available in NCBI virus [21] and virological.org using the Augur application [22], whose functionality relies on the IQ-Tree software [23]. The MAFFT program [24], was utilized for the multiple alignment, using the isolate MPXV-M5312_HM12_Rivers (NC_063383.1) as reference. A maximum likelihood method with a general time reversible model with unequal rates and unequal base frequencies [25] was used to reconstruct the viral phylogeny. The results can be viewed in the Nextstrain Auspice [26] local server, which is now part of the Genomic Surveillance Circuit of Andalusia [27].

The amino acid substitutions for each of the 160 samples sequenced in Andalusia (Supplementary Table S1) were obtained using the nextclade web application [28]. Specifically using as pathogen reference "Human Monkeypox Clade B.1" (ON563414), which belongs to the same clade as the Andalusian samples but in the coordinates of the isolate NC_063383.

Structural variations

To identify samples that could harbor large structural variants or deletions, analyses of coverage plots were conducted. Specifically, samples exhibiting regions of low coverage or regions with high coverage in the coverage plots were located [29].

To confirm potential structural variants, reads from the samples displaying non-homogeneous patterns in the coverage plots were aligned using the bwa program [30],

employing the "-a" argument to retain all alignments. The sample OX044336.2 was used as the reference in this process.

The determination of the breakpoints for deletions and the insertion points for rearrangements was achieved by selecting reads that did not have an appropriate insert size. This indicated that the mate pairs mapped to both sides of a deletion or different regions in the case of duplication/rearrangement. To filter and obtain these reads, the samtools application [31] was used with the "-F14" filter, which eliminates reads that have not mapped and properly mapped reads. Furthermore, to confirm the start and end points, only reads with chimeric alignments, where a portion of the read aligned to one genomic region and another portion aligned to a different region, were retained. This selection was made using the flag "2048" to filter the reads.

Finally, in IGV [32], it was confirmed that the accumulation of reads with inappropriate insert sizes and chimeric reads corresponded to the positions where the coverage plot exhibited changes.

Data availability

The 160 MPXV genomes are available at the European Nucleotide Archive (ENA) repository under the accession number PRJEB55075. Supplementary Table S1 provides also each individual sample ENA IDs.

Acknowledgements

This work was supported by Spanish Ministry of Science and Innovation (grants PID2020-117979RB-I00 and FJC2021-046546-I), the Instituto de Salud Carlos III (ISCIII), co-funded with European Regional Development Funds (ERDF) (grant IMP/00019), it has also been funded by Consejería de Salud y Consumo, Junta de Andalucía (grants COVID-0012-

2020), and by grant ELIXIR-CONVERGE - Connect and align ELIXIR Nodes to deliver sustainable FAIR lifescience data management services (AMD-871075-16), funded by EU – H2020.

References

1. Bunge EM, Hoet B, Chen L, Lienert F, Weidenthaler H, Baer LR, et al. The changing epidemiology of human monkeypox—A potential threat? A systematic review. PLoS neglected tropical diseases. 2022;16(2):e0010141.

2. Kraemer MU, Tegally H, Pigott DM, Dasgupta A, Sheldon J, Wilkinson E, et al. Tracking the 2022 monkeypox outbreak with epidemiological data in real-time. The Lancet infectious diseases. 2022.

3. Mathieu E, Dattani S, Ritchie H, Roser M. OurWorldInData. Monkeypox Data Explorer 2020 [cited 2022 20/07/]. Available from: https://ourworldindata.org/monkeypox.

4. Lai C-C, Hsu C-K, Yen M-Y, Lee P-I, Ko W-C, Hsueh P-R. Monkeypox: an emerging global threat during the COVID-19 pandemic. Journal of Microbiology, Immunology and Infection. 2022.

5. Kapoor S, Varadharajan A. Monkeypox 2022: Emerging zoonotic epidemic threat, future implications, and way ahead. Journal of Public Health and Primary Care. 2023;4(1):8.

6. SARS-CoV-2 whole genome sequencing circuit in Andalusia 2021 [22 November 2022]. Available from: https://www.clinbioinfosspa.es/COVID_circuit/.

7. Dopazo J, Maya-Miles D, García F, Lorusso N, Calleja MÁ, Pareja MJ, et al. Implementing Personalized Medicine in COVID-19 in Andalusia: An Opportunity to Transform the Healthcare System. Journal of Personalized Medicine. 2021;11(6):475.

8. Isidro J, Borges V, Pinto M, Sobral D, Santos JD, Nunes A, et al. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. Nature Medicine. 2022:1-.

9. Mauldin MR, McCollum AM, Nakazawa YJ, Mandra A, Whitehouse ER, Davidson W, et al. Exportation of monkeypox virus from the African continent. The Journal of infectious diseases. 2022;225(8):1367-76.

10. Franklin E, Khan AR. Poxvirus antagonism of innate immunity by Bcl-2 fold proteins. Journal of Structural Biology. 2013;181(1):1-10.

11. Faye O, Pratt CB, Faye M, Fall G, Chitty JA, Diagne MM, et al. Genomic characterisation of human monkeypox virus in Nigeria. The Lancet infectious diseases. 2018;18(3):246. Epub 2018/01/24. doi: 10.1016/s1473-3099(18)30043-4. PubMed PMID: 29361427.

12. Chen N, Li G, Liszewski MK, Atkinson JP, Jahrling PB, Feng Z, et al. Virulence differences between monkeypox virus isolates from West Africa and the Congo basin. Virology. 2005;340(1):46-63. Epub 2005/07/19. doi: 10.1016/j.virol.2005.05.030. PubMed PMID: 16023693.

13. Luo H. Interplay between the virus and the ubiquitin-proteasome system: molecular mechanism of viral pathogenesis. Current opinion in virology. 2016;17:1-10. Epub 2015/10/02. doi: 10.1016/j.coviro.2015.09.005. PubMed PMID: 26426962; PubMed Central PMCID: PMCPMC7102833.

14. Duraffour S, Lorenzo MM, Zöller G, Topalis D, Grosenbach D, Hruby DE, et al. ST-246 is a key antiviral to inhibit the viral F13L phospholipase, one of the essential proteins for orthopoxvirus wrapping. Journal of Antimicrobial Chemotherapy. 2015;70(5):1367-80. doi: 10.1093/jac/dku545.

15. Martínez JI, Montalbán EG, Bueno SJ, Martínez FM, Juliá AN, Díaz JS, et al. Monkeypox outbreak predominantly affecting men who have sex with men, Madrid, Spain, 26 April to 16 June 2022. Eurosurveillance. 2022;27(27):2200471.

16. Patel H, Varona S, Monzón S, Espinosa-Carrasco J, Heuer ML, Gabernet G, et al. Nf-Core/Viral recon v2.4.1 - Plastered Magnesium Marmoset (2.4.1) 2021 [22/11/2022]. Available from: https://doi.org/10.5281/zenodo.6320980.

17. Monzon S, Varona S, Negredo A, Patiño-Galindo JA, Vidal-Freire S, Zaballos A, et al. Changes in a new type of genomic accordion may open the pallets to increased monkeypox transmissibility. bioRxiv. 2022:2022.09. 30.510261.

18. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9. Epub 2012/03/06. doi: 10.1038/nmeth.1923. PubMed PMID: 22388286; PubMed Central PMCID: PMC3322381.

19. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome biology. 2019;20(1):1-19.

20. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021;10(2):giab008.

21. Brister JR, Ako-Adjei D, Bao Y, Blinkova O. NCBI viral genomes resource. Nucleic acids research. 2015;43(D1):D571-D7.

22. Huddleston J, Hadfield J, Sibley TR, Lee J, Fay K, Ilcisin M, et al. Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens. Journal of Open Source Software. 2021;6(57):2906.

23. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular biology and evolution. 2020;37(5):1530-4.

24. Katoh K, Standley DM. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. Bioinformatics. 2016;32(13):1933-42.

25. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures on mathematics in the life sciences. 1986;17(2):57-86.

26. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018;34(23):4121-3.

27. Monkeypox: Genomic Surveillance Circuit of Andalusia 2022 [cited 2022 27/07/2022]. Available from: http://nextstrain.clinbioinfosspa.es/monkeypox.

28. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. Nextclade: clade assignment, mutation calling and quality control for viral genomes. Journal of Open Source Software. 2021;6(67):3773.

29. Gigante CM, Plumb M, Ruprecht A, Zhao H, Wicker V, Wilkins K, et al. Genomic deletions and rearrangements in monkeypox virus from the 2022 outbreak, USA. bioRxiv. 2022:2022.09. 16.508251.

30. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25(14):1754-60. doi: 10.1093/bioinformatics/btp324.

31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. Epub 2009/06/10. doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PubMed Central PMCID: PMC2723002.

32. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics. 2012;14(2):178-92. doi: 10.1093/bib/bbs017.

Supporting information

Supplementary Table S1. Monkeypox samples sequenced in this study, along with their collection date, origin and individual ENA IDs, belonging to the collective accession number PRJEB55075.

Supplementary Table S2. Non-synonymous mutations found in the samples under study with respect to the reference ON563414 in NC_063383 coordinates.

Supplementary Figure 1. Whole phylogeny of the Andalusian isolates, with identical sequences grouped in nodes with a size proportional to the number of corresponding isolates. Each sequence is labeled with its clade (B.1 and derived clades), and, in case of harboring a structural variation, it is also labeled. The three sequences, corresponding to the two deceased patients are labeled as well.

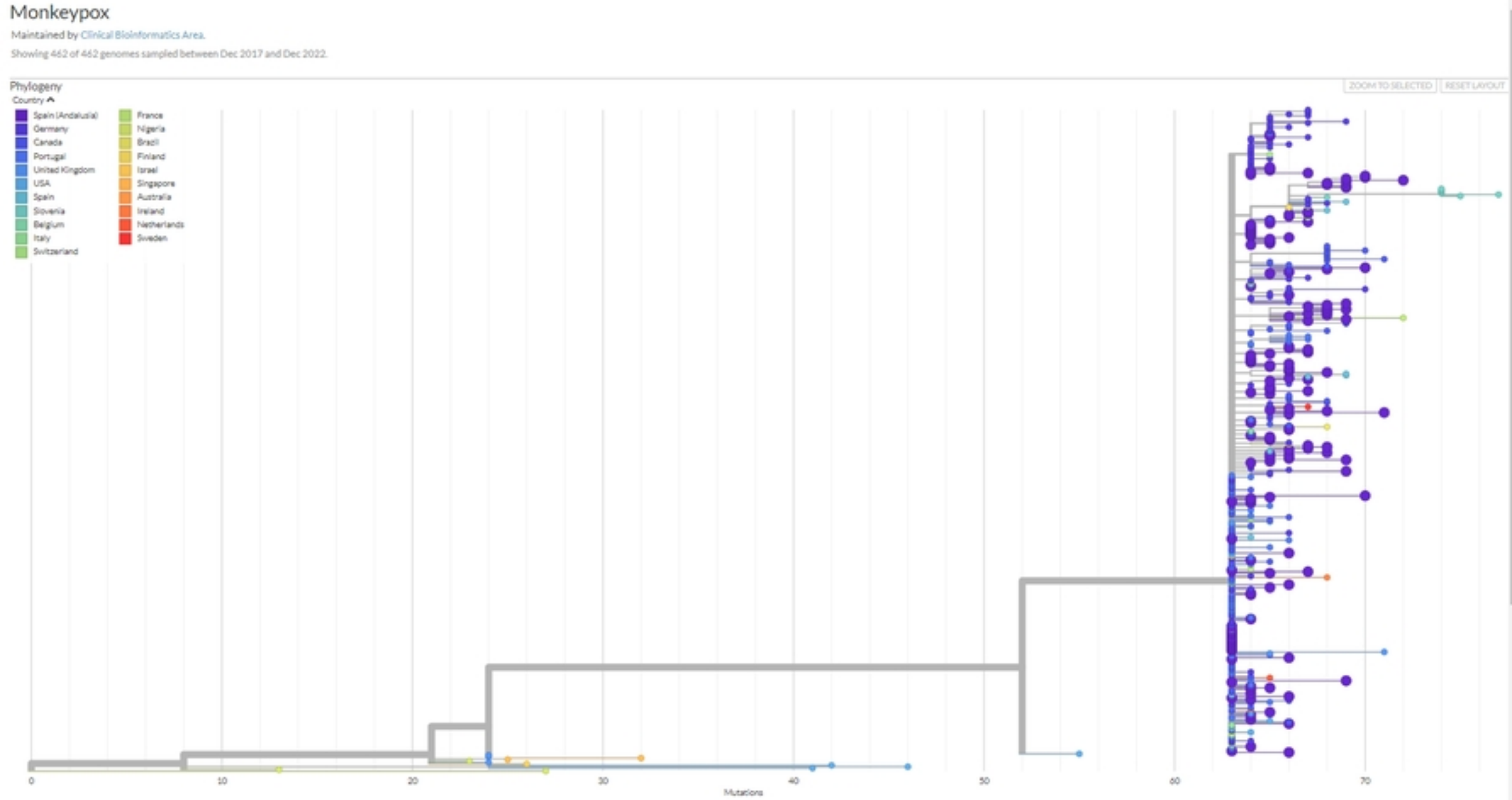

Monkeypox
Maintained by Clinical Bioinformatics Area.
Showing 462 of 462 genomes sampled between Dec 2017 and Dec 2022.
Phylogeny
ZOOM TO SELECTED
RESET LAYOUT
Country
Spain (Andalusia)
Germany
Canada
Portugal
United Kingdom
USA
Spain
Slovenia
Belgium
Italy
Switzerland
France
Nigeria
Brazil
Finland
Israel
Singapore
Australia
Ireland
Netherlands
Sweden
0
10
20
30
40
50
60
70
Mutations

Figure 1

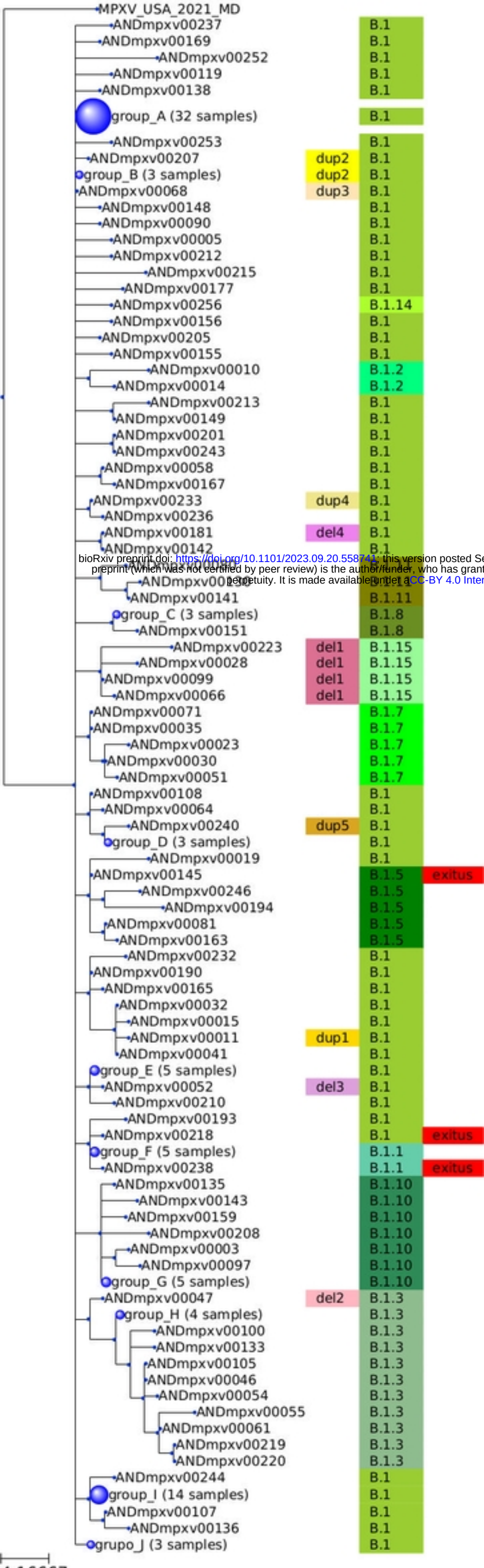


Figure 2

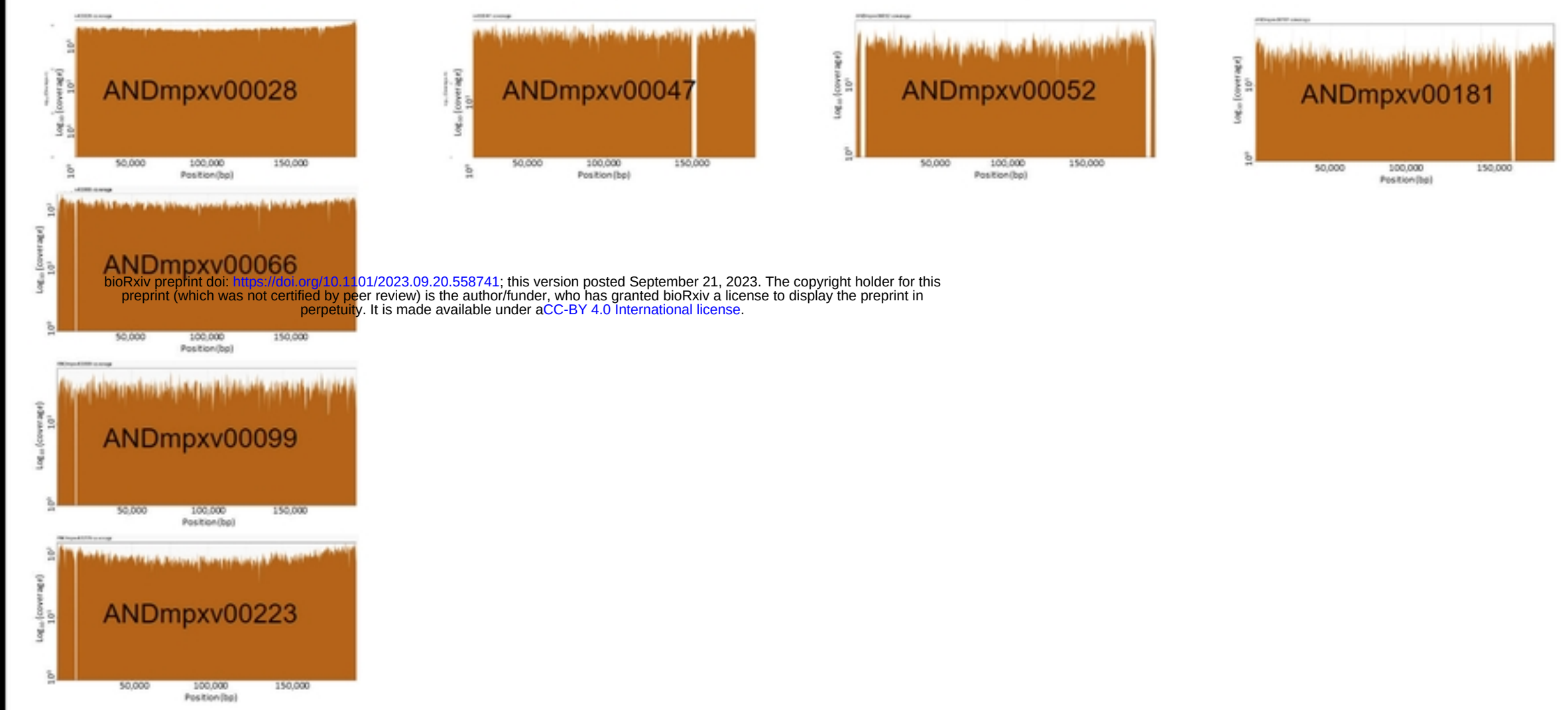


Figure 3