

# A draft of the first genome sequence of Monkeypox virus associated with the multi-country outbreak in May 2022 from the Canary Islands, Spain

jlorsal

Jun '22

## Authors

Julia Alcoba-Florez<sup>1</sup>, Adrián Muñoz-Barrera<sup>2</sup>, Laura Ciuffreda<sup>3</sup>, Héctor Rodríguez-Pérez<sup>3</sup>, Luis A. Rubio-Rodríguez<sup>2</sup>, Helena Gil-Campesino<sup>1</sup>, Diego García-Martínez de Artola<sup>1</sup>, Antonio Íñigo-Campos<sup>2</sup>, Oscar Díez-Gil<sup>1</sup>, Rafaela González-Montelongo<sup>2</sup>, Agustín Valenzuela-Fernández<sup>4</sup>, José M. Lorenzo-Salazar<sup>2</sup>, Carlos Flores<sup>2,3,5,6</sup>

## Affiliations

<sup>1</sup>Servicio de Microbiología, Hospital Universitario Ntra. Sra. de Candelaria, 38010 Santa Cruz de Tenerife, Spain

<sup>2</sup>Genomics Division, Instituto Tecnológico y de Energías Renovables, 38600 Santa Cruz de Tenerife, Spain

<sup>3</sup>Fundación Canaria Instituto de Investigación Sanitaria de Canarias at the Research Unit, Hospital Universitario Ntra. Sra. de Candelaria, 38010 Santa Cruz de Tenerife, Spain

<sup>4</sup>Laboratorio de Inmunología Celular y Viral, Unidad de Farmacología, Facultad de Medicina, Universidad de La Laguna, 38200 San Cristóbal de La Laguna, Spain

<sup>5</sup>CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, 28029 Madrid, Spain

<sup>6</sup>Facultad de Ciencias de la Salud, Universidad Fernando Pessoa Canarias, 35450 Las Palmas de Gran Canaria, Spain

## Introduction

Monkeypox virus (MPXV) is a zoonotic Orthopoxvirus (OPV) (family *Poxviridae*) [1,2], endemic to West and Central Africa [3,4]. MPXV has been described in humans in Central and Western Africa (occurring mainly in tropical forest areas of Central Africa), as well as in other parts of the world [5-10]. Around the 13 May 2022, MPXV cases were reported in several countries and WHO declared community transmission of the virus [11]. Most reported cases so far have been presented through sexual health or other health services and have involved mainly men who have sex with men [12,13]. By early June 2022, 129 viral genomes had been deposited at GISAID with 46 SNPs shared by all these sequences and differing from the viral genome sequences from the 2018-2019 MPXV outbreak [14]. Preliminary data from polymerase chain reaction (PCR) assays indicate that these MPXV strains detected in Europe and other non-endemic areas belong to the West African clade [15].

In Europe, several cases of MPXV infection have been associated with outbreaks in the Canary Islands and Spain [15]. A few sequences of the cases in Spain have been reported [16]. A total of 15 positive cases have been confirmed in the Canary Islands so far (10 more cases are under investigation, including 5 suspects and 5 probable cases as of June 13 2022)

[Skip to main content](#) the draft sequences of the first MPVX viral genome isolated in the

Canary Islands the 31 May 2022 from a male adult patient with one week-onset mild symptoms (fever, odynophagia) and presenting at ER but not necessitating hospital admission.

## Materials and Methods

### ***DNA extraction and PCR testing***

Viral DNA was extracted at the Hospital Universitario Ntra. Sra. de Candelaria (HUNSC) from five samples (nasopharyngeal swab, lesion crust, and vesicles) from the same patient using the eMAG system (Biomerieux) following manufacturer's instructions. Virus inactivation was conducted under a biosafety class II cabinet (TELSTAR bio-II-A), following ECDC procedures [18]. Diagnosis of MPXV infection was confirmed using the LightMix Modular Orthopox (Roche) and a real-time PCR assay described elsewhere [19].

### ***Sequencing***

Five independent DNA dual index libraries (one for each sample) were processed at ITER with Nextera XT DNA Library Preparation Kit (Illumina Inc.), following the manufacturer's recommendations with manual library normalization, and pooled prior to sequencing. The quality of the libraries was assessed with a D1000 ScreenTape kit on the 4200 TapeStation System (Agilent). Library concentrations ranged from 7.4 to 10.4 nM, and showed a fragmentation profile ranging from 721 to 808 bp. The mean fragment size for the sequencing pool was 677 bp as measured with a D1000 High Sensitivity ScreenTape kit. Paired-end sequences were obtained on a MiSeq Sequencing System (Illumina Inc.), using the reagent kit v3 chemistry with 150 cycles and an expected throughput of 3.3-3.8 Gb. The pool concentration was 15 pM, and 5% of PhiX Control V3 was used as the internal control.

In addition, DNA libraries for nanopore sequencing were prepared from the sample with the highest yield (taken from a skin lesion exudate) using the Rapid Barcoding kit (SQK-RBK004) from Oxford Nanopore Technologies (ONT). To increase the quantity of the starting material, the protocol used 30 to 45 ng of the DNA extract in 7.5 µl for 12 independently barcoded libraries that were pooled in order to obtain the maximum yield from the run. The pooled libraries were loaded onto an R9.4.1 flow cell and were run in a MinION (ONT) for 42 hours. Basecalling of raw ONT signal data as well as demultiplexing and adapter trimming was carried out using Guppy v.6.0.7 with default parameters and the high-accuracy basecall model.

### ***Bioinformatic and phylogenetic analyses***

As the first step, the individual demultiplexed FASTQ pair of Illumina files were interleaved with **BBMap** (Reformat tool) and then merged into a single interleaved FASTQ file. Then two different bioinformatic tools were tested to identify and remove human reads: **Kraken 2** and **NCBI SRA Human Scrubber** v.1.0.2021\_05\_05.

The remaining interleaved paired-end reads were subjected to different bioinformatic procedures to obtain draft sequences.

On the one hand, a reference-based analysis was conducted with Illumina unclassified reads that were mapped to the MPVX viral genome MPXV-UK\_P2 (GenBank **MT903344.1**) by means of **minimap2** v.2.24-r1122. At this stage, duplicate metrics from **PICARD** v.2.18.7, and coverage metrics from **SAMtools** v.1.6 and **mosdepth** v.0.3.3 were obtained from the remaining interleaved paired-end reads.

Variant calling was carried out with **iVar** v1.3.1 and **LoFreq** v.2.1.5 using default parameters against the **MT903344.1** genome. The consensus sequence in FASTA format corresponding to **Skip to main content**

the Illumina sequencing experiment was obtained by piping a [SAMtools](#) v.1.6 pileup with [iVar](#) v.1.3.1 consensus as described elsewhere [20].

On the other hand, a hybrid *de novo* assembly was obtained combining the filtered Illumina and ONT reads using an in-house script based on the [Unicycler](#) v.0.5.0 assembler. Quality control parameters for the draft assembly were obtained with [QUAST](#) v.5.0.2 using the [MT903344.1](#) as the reference genome. [Bandage](#) v.0.9.0 was used to visualize the resulting contigs in the assembly. A refined version of the hybrid *de novo* assembly was obtained after running [Kraken 2](#) v.2.1.2 with [PlusPF](#) database to remove non-viral assembled contigs.

The Illumina-derived consensus sequence was aligned with 137 MPXV sequences downloaded from NCBI GenBank ([table 1](#)) using [MAFFT](#) v.7.505. A phylogenetic analysis was performed using both [IQ-TREE](#) v.2.2.0.3 with the K3Pu+F+I model as best-predicted model and default parameters, and [Nextstrain monkeypox](#).

## Results

The Illumina sequencing run produced 3.88 Gb and 25.5 M reads in total. A mapping of 101,814 and 100,897 Illumina reads were obtained using [minimap2](#) on the [NCBI SRA Human Scrubber](#) (mean depth: 38.3x) and [Kraken 2](#) (mean depth: 38.1x), respectively, thus providing equivalent results. [mosdepth](#) showed that 99% of the MPXV genome was covered  $\geq 1x$ , while a fraction of 85% of the genome was covered at  $\geq 10x$ . [PICARD](#) estimated as few as 2.81% of duplicated reads. The resulting consensus FASTA for Illumina-only reads provided a near-fully complete viral genome (99.91%) against the [MT903344.1](#) reference. This genome has been released through the GenBank repository with accession ([ON782055](#)) and at our [GitHub project repository](#).

The ONT run provided 1.98 Gb and a total of 1.38 M reads, ranging from 499 to 101,895 bp in length, a mean length of 1,432 bases. ONT sequencing provided 2,246 non-human mapping reads after filtering with [Kraken 2](#), thus, a theoretical viral genome depth of 14.9x.

A hybrid *de novo* assembly based on Illumina and ONT [Kraken 2](#)-filtered reads performed with Unicycler provided four contigs. Contigs 1 and 2 expanded 186,315+4,703 bases (191,018 bp total) and mapped to [Monkeypox virus Zaire-96-I-16](#), whereas contigs 3 and 4 expanded 10,530 bases but did not map to MPXV. A consensus sequence from the hybrid *de novo* assembly and the [MT903344.1](#) MPXV reference genome has been deposited in GenBank with accession ([ON782054](#)) and is also available through our [GitHub project repository](#).

A preliminary phylogenetic analysis (Fig. 1) shows that the draft MPXV genome of the patient belongs to the so-called West African clade or B.1 [21,22]. In addition, the closest sequences are related to the Slovenian-MPXV GenBank-released genomes, contributing further evidence of community spread in the present worldwide MPXV outbreak.

[Skip to main content](#)


[nextstrain\\_monkeypox\\_ByDivergence\\_General\\_ed](#)

**Figure 1.** A phylogenetic tree depicting the draft MPXV sequence isolated on May 31, 2022 from a patient from the Canary Islands along with NCBI GenBank publicly available sequences computed by a [Nextstrain monkeypox](#) local instance.

### Data Availability

A table ([table 1](#)) with the acknowledgement and accession numbers used for phylogenetic analysis shown in Fig. 1 is also available in: [GitHub - genomicsITER/monkeypox: Monkeypox public repository](#). A FASTA file with multiple sequences of MPXV from NCBI GenBank used in the Multiple Sample Alignment step with MAFFT or Nextstrain-monkeypox is also available in: [GitHub - genomicsITER/monkeypox: Monkeypox public repository](#)

### Funding

This study has been funded by Cabildo Insular de Tenerife (CGIEU0000219140 and “*Apuestas científicas del ITER para colaborar en la lucha contra la COVID-19*”), Instituto de Salud Carlos III (FI18/00230) cofunded by European Union (ERDF) “*A way of making Europe*”, and by the agreement with Instituto Tecnológico y de Energías Renovables (ITER) to strengthen scientific and technological education, training, research, development and innovation in Genomics, Personalized Medicine and Biotechnology (OA17/008).

### References

1. Esposito JJ, Fenner F, Poxvirus. Fields Virology. Lippincott Williams and Wilkins; NY, USA: 2001. pp. 2885-2921.
2. Ryan, K.J. and Ray, C.G., Eds. (2004) Sherris Medical Microbiology. 4th Edition, pp. 525–28, McGraw-Hill, New York.
3. Ligon BL. Monkeypox: a review of the history and emergence in the Western hemisphere. Semin Pediatr Infect Dis. 2004;15(4):280-287. doi:10.1053/j.spid.2004.09.001
4. Jezek Z, Fenner F. Human monkeypox. In: Melnick JL, editor. Monographs in virology, Vol. 17. Basel: Karger; 1988, p. 1-140.
5. Parker S, Nuara A, Buller RM, Schultz DA. Human monkeypox: an emerging zoonotic disease. Future Microbiol. 2007;2(1):17-34.
6. Gispen R, Brand-Saathof BB, Hekker AC. Monkeypox-specific antibodies in human and simian sera from the Ivory Coast and Nigeria. Bull World Health Organ. 1976;53(4):355-360.
7. Jezek Z, Fenner F. Human Monkeypox. Karger, Basel, Switzerland: 1988; Hutin YJ, [Skip to main content](#) [ait P, et al. Outbreak of human monkeypox, Democratic Republic of](#)

- Congo, 1996 to 1997. *Emerg. Infect. Dis.* 2001;7(3):434-438.
8. Di Giulio DB, Eckburg PB. Human monkeypox: an emerging zoonosis. *Lancet Infect Dis.* 2004;4:15-25.
  9. Parker S, Buller RM. A review of experimental and natural infections of animals with monkeypox virus between 1958 and 2012. *Future Virol.* 2013;8(2):129-157. doi:10.2217/fvl.12.130
  10. Reed KD, Melski JW, Graham MB, et al. The detection of monkeypox in humans in the Western Hemisphere. *N Engl J Med.* 2004;350(4):342-350. doi:10.1056/NEJMoa032299
  11. World Health Organization (21 May 2022). Disease Outbreak News; Multi-country monkeypox outbreak in non-endemic countries. Available at: <https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON385>
  12. World Health Organization (29 May 2022). Disease Outbreak News; Multi-country monkeypox outbreak in non-endemic countries. Available at: <https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON388>
  13. World Health Organization (4 June 2022). Disease Outbreak News; Multi-country monkeypox outbreak in non-endemic countries: Update. Available at: <https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON390>
  14. Multi-country outbreak of Monkeypox virus: genetic divergence and first signs of microevolution. *Virological, Monkeypox, Genome Reports*, <https://pando.tools/t/multi-country-outbreak-of-monkeypox-virus-genetic-divergence-and-first-signs-of-microevolution/806>
  15. Illumina whole-genome sequence of Monkeypox virus in a patient travelling from the Canary Islands to France. *Virological, Monkeypox, Genome Reports*. <https://pando.tools/t/illumina-whole-genome-sequence-of-monkeypox-virus-in-a-patient-travelling-from-the-canary-islands-to-france/829>
  16. UPDATE: Two draft genomes from Madrid, Spain, of the Monkeypox virus 2022 outbreak. *Virological, Monkeypox, Genome Reports*, <https://pando.tools/t/update-two-draft-genomes-from-madrid-spain-of-the-monkeypox-virus-2022-outbreak/848>
  17. Health related news from the Government of the Canary Islands: <https://www3.gobiernodecanarias.org/noticias/sanidad-contabiliza-tres-casos-confirmados-y-tres-en-estudio-de-viruela-del-mono-desde-el-viernes/>
  18. ECDC, Factsheet for health professionals on monkeypox. <https://www.ecdc.europa.eu/en/all-topics-z/monkeypox/factsheet-health-professionals>
  19. Li Y, Zhao H, Wilkins K, Hughes C, Damon IK. Real-time PCR assays for the specific detection of monkeypox virus West African and Congo Basin strain DNA. *J Virol Methods.* 2010;169(1):223-227. doi:10.1016/j.jviromet.2010.07.012
  20. First French draft genome sequence of Monkeypox virus, may 2022. *Virological, Monkeypox*, <https://pando.tools/t/first-french-draft-genome-sequence-of-monkeypox-virus-may-2022/819>
  21. Urgent need for a non-discriminatory and non-stigmatizing nomenclature for monkeypox virus. *Virological, Monkeypox*, <https://pando.tools/t/urgent-need-for-a-non-discriminatory-and-non-stigmatizing-nomenclature-for-monkeypox-virus/853>
  22. Rename monkeypox strains to remove geographic stigma, researchers say. Proposal would avoid references to West African and Congo. 10.1126/science.add4325. <https://www.science.org/content/article/rename-monkeypox-remove-geographic-stigma-researchers-say>

Hi, could you confirm that ON782055 and ON782054 are from the same sample but were sequenced and assembled differently?

---

jlorsal

Jun '22

Hi Yu,

Yes, they belong to the same patient but were obtained as described in the post. One of the accessions corresponds to a FASTA consensus sequence obtained from Illumina-MiSeq and mapped to the reference; and the other accession shows the FASTA sequence resulting from a hybrid *de novo* assembly combining Illumina-MiSeq short-reads and MinION long-reads.

The full pipeline describing the steps is shown here:

 [MPXV\\_pipeline\\_v1](#)

We have more details here: [HUNSC-ITER MPXV GitHub repository](#)

---

YuLi

Jun '22

Thank you!

There are 5 SNPs over 5 kb sequences at the right terminal end of the genomes between ON782055 and ON782054. are those differences real or due to different sequencing method or assembling?

---

jlorsal

Jun '22

Hi Yu,

Yes, there are five SNPs after position 193,439 seen in the short-reads based sequence

[Skip to main content](#) seen in the hybrid *de novo* assembly (ON782054.1), as depicted here in the [snipit](#) shot:



 [HUNSC\\_ITER\\_MPXV01\\_snp\\_plot](#)

And yes, it is a consequence of the *de novo* assembly since the small contig of MPXV assembled in this region has not a good mapping quality (though it blasted to MPXV), and the MPXV reference used for the consensus. In addition, you can see that our deposited sequences are incomplete genomes at the right terminal end because of the reference used to map (in the case of short-reads) and to hybrid *de novo* assemble the sequences (MT903344). However, the coverage at this terminal is quite good for the short-reads derived sequence to perform a variant calling (work in progress).