Changes in a new type of genomic accordion may open the pallets to increased monkeypox transmissibility

Sara Monzón[1#], Sarai Varona[1#], Anabel Negredo[2,3#], Juan Angel Patiño-Galindo[4], Santiago Vidal-Freire[4], Angel Zaballos[5], Eva Orviz[6], Oskar Ayerdi[6], Ana Muñoz-García[6], Alberto Delgado-Iribarren[6], Vicente Estrada[3,6], Cristina García[2], Francisca Molero[2], Patricia Sánchez[2,3], Montserrat Torres[2], Ana Vázquez[2,7], Juan-Carlos Galán[7,8], Ignacio Torres[9], Manuel Causse del Río[10], Laura Merino[11], Marcos López[12], Alicia Galar[13], Laura Cardeñoso[14], Almudena Gutiérrez[15], Cristina Loras[16], Isabel Escribano[17], Marta Elena Alvarez-Argüelles[18], Leticia del Río[19], María Simón[20], MªAngeles Meléndez[21], Juan Camacho[2], Laura Herrero[2], Pilar Jiménez Sancho[5], Maria Luisa Navarro-Rico[5], Jens H. Kuhn[22], Mariano Sanchez-Lockhart[23], Nicholas Di Paola[23], Jeffrey R. Kugelman[23], Elaina Giannetti[4], Susana Guerra[4,24,25], Adolfo García-Sastre[4,24,26,27,28], Gustavo Palacios[4,24&*], Isabel Cuesta[1&], and Maripaz P. Sánchez-Seco[2,3&]

[1]Unidad de Bioinformática, Unidades Centrales Científico Técnicas, Instituto de Salud Carlos III, 28029 Madrid, Spain

[2]Centro Nacional de Microbiología, Instituto de Salud Carlos III, 28029 Madrid, Spain

[3]Centro de Investigación Biomédica en Red de Enfermedades Infecciosas (CIBERINFEC), Instituto de Salud Carlos III, 28029 Madrid, Spain

[4]Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[5]Unidad de Genómica, Unidades Centrales Científico Técnicas, Instituto de Salud Carlos III, 28029 Madrid, Spain

[6]Centro Sanitario Sandoval, Hospital Clínico San Carlos, 28040 Madrid, Spain

[7]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Instituto de Salud Carlos III, 28029 Madrid, Spain

[8]Servicio de Microbiología, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), 28034 Madrid, Spain

[9]Servicio de Microbiología, Hospital Clínico Universitario, Instituto de Investigación INCLIVA, 46010 Valencia, Spain

[10]Unidad de Microbiología, Hospital Universitario Reina Sofía, Instituto Maimónides de Investigación Biomédica de Córdoba, 14004 Córdoba, Spain

[11]Unidad Clínico de Enfermedades Infecciosas, Microbiología y Medicina Preventiva, Hospital Universitario Virgen del Rocío, 41013 Sevilla, Spain

[12]Servicio de Microbiología y Parasitología, Hospital Universitario Puerta de Hierro Majadahonda, 28222 Madrid, Spain

[13]Servicio de Microbiología Clínico y Enfermedades Infecciosas, Hospital General Universitario Gregorio Marañón, 28007 Madrid, Spain

[14]Servicio de Microbiología, Instituto de Investigación Sanitaria, Hospital Universitario de la Princesa, 28006 Madrid, Spain

[15]Servicio de Microbiología y Parasitología Clínica, Hospital Universitario La Paz, 28046 Madrid, Spain

[16]Servicio de Microbiología, Hospital General y Universitario, 13005 Ciudad Real, Spain

[17]Hospital General Universitario Dr. Balmis, 03010 Alicante, Spain

[18]Servicio de Microbiología, Hospital Universitario Central, 33006 Asturias, Spain

[19]Hospital Quironsalud Torrevieja, 03184 Alicante, Spain

[20]Servicio de Microbiología, Hospital Central de la Defensa "Gómez Ulla", 28947 Madrid, Spain

[21]Servicio de Microbiología y Parasitología, Hospital Universitario 12 de Octubre, 28041 Madrid, Spain

[22]Integrated Research Facility at Fort Detrick, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Fort Detrick, Frederick, MD 21702, USA

[23]United States Army Research Institute for Infectious Disease, Fort Detrick, Frederick, MD 21702, USA

[24]Global Health Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[25]Departmento de Medicina Preventiva, Salud Publica y Microbiología, Universidad Autónoma de Madrid, 28029 Madrid, Spain

[26]Department of Medicine, Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[27]The Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[28]Department of Pathology, Molecular and Cell-Based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[#]These authors contributed equally to this work.

[&]These senior authors contributed equally to this work.

Correspondence

[*]Gustavo Palacios: gustavo.palacios@mssm.edu

SUMMARY

The currently expanding monkeypox epidemic is caused by a subclade IIb descendant of a monkeypox virus (MPXV) lineage traced back to Nigeria in 1971. In contrast to monkeypox cases caused by clade I and subclade IIa MPXV, the prognosis of current cases is generally favorable, but person-to-person transmission is much more efficient. MPXV evolution is driven by selective pressure from hosts and loss of virus–host interacting genes. However, there is no satisfactory genetic explanation using single-nucleotide polymorphisms (SNPs) for the observed increased MPXV transmissibility. We hypothesized that key genomic changes may occur in the genome's low-complexity regions (LCRs), which are highly challenging to sequence and have been dismissed as uninformative. Using a combination of highly sensitive techniques, we determined a first high-quality MPXV genome sequence of a representative of the current epidemic with LCRs resolved at unprecedented accuracy. This effort revealed significant variation in short-tandem repeats within LCRs. We demonstrate that LCR entropy in the MPXV genome is significantly higher than that of SNPs and that LCRs are not randomly distributed. *In silico* analyses indicate that expression, translation, stability, or function of MPXV orthologous poxvirus genes (OPGs) 153, 204, and 208 could be affected in a manner consistent with the established "genomic accordion" evolutionary strategies of orthopoxviruses. Consequently, we posit that genomic studies focusing on phenotypic MPXV clade-/subclade-/lineage-/strain differences should change their focus to the study of LCR variability instead of SNP variability.

Keywords

Chordopoxvirinae, genomic accordion, low-complexity region, molecular epidemiology, monkeypox, MPX, MPXV, OPG, orthopoxvirus, *Poxviridae*, short tandem repeat, STR, transmission, genomics, bioinformatics, computational biology

INTRODUCTION

Monkeypox virus (MPXV) is a double-stranded DNA virus that belongs to genus *Orthopoxvirus* (varidnavirian *Nucleocytoviricota*: *Poxviridae*: *Chordopoxvirinae*) along with other human viruses, such as vaccinia virus (VACV) and variola virus (VARV) (International Committee on Taxonomy of Viruses, 2022). First encountered in 1958 in crab-eating macaques imported to Belgium (Magnus et al., 2009), MPXV has caused sporadic disease outbreaks in humans since the 1970s in Eastern, Middle, and Western Africa, totaling approximately 25,000 cases (case fatality rate 1–10%) (Beer and Rao, 2019), and also sporadic disease outbreaks among wild monkeys and apes (Patrono et al., 2020; Radonić et al., 2014). Exposure to MPXV animal reservoirs, in particular rope squirrels and sun squirrels, is a significant risk factor of human infections (Khodakevich et al., 1988).

The human disease caused by MPXV is designated as monkeypox in the World Health Organization (WHO) International Classification of Diseases, Eleventh Revision (ICD-11; code 1E71) (World Health Organization, 2022a). Phylogenetically, historic MPXV isolates cluster into two clades (Likos et al., 2005), designated I and II (Happi et al., 2022; World Health Organization, 2022b). Clade I viruses are considered more virulent and transmissible than clade II viruses (Damon, 2011; Likos *et al.*, 2005; World Health Organization, 2022b).

Since May 2022, multiple European countries have reported a continuously increasing number of MPXV infections and associated disease, including clusters of cases associated with potential superspreading events in Belgium, Spain, and the United Kingdom (UK). As of September 30, 2022, a total of 68,428 cases had been reported in 106 countries/territories/areas in all six WHO regions. Of these, 67,739 were in 99 countries that have not reported MPXV infections prior to 2022 (Centers for Disease Control and Prevention, 2022). This rapid increase in infections prompted WHO to declare this epidemic

a Public Health Emergency of International Concern (PHEIC) (Nuzzo et al., 2022). The viruses of the 2022 epidemic belong to subclade IIb (Antinori et al., 2022; Nextstrain, 2022; Vivancos et al., 2022), a line of descent of MPXV that had been circulating in Nigeria, likely since 1971 (Faye et al., 2018).

The clinical presentation monkeypox caused by clade I or IIa includes fever, headache, lymphadenopathy, and/or malaise, followed by a characteristic rash that progresses centrifugally from maculopapules via vesicles and pustules to crusts that may occur on the face, body, mucous membranes, palms of the hands, and soles of the feet (Ježek et al., 1987). The clinical presentation of the current MPXV subclade IIb infections diverges from classical monkeypox by having a good prognosis, self-limiting but infectious skin lesions (typically emerging at and restricted to the genital, perineal/perianal, and/or peri-oral areas) before the development of fever, lymphadenopathy, and malaise. Generalized disease usually manifests with a rash that has not been widely observed in the current outbreak. Human-to-human transmission is substantially higher in subclade IIb MPXV-associated outbreaks than in clade I and clade IIa MPXV (Bunge et al., 2022; Otu et al., 2022; Thornhill et al., 2022; Ulaeto et al., 2022; Vusirikala et al., 2022). The R_0 for MXPV IIb among men who have sex with men (MSM) is higher than 1. Transmission may be catalyzed by decreasing protection associated with the VARV/smallpox vaccination campaign that ended in 1980 (Grant et al., 2020; Rimoin et al., 2010). Furthermore, a change in transmission route may be the cause of the difference in clinical presentation and pathogenesis as was shown in animal models (Reynolds et al., 2006).

Orthopoxvirus infections are classified as systemic or localized illnesses. Localized usually means that signs are restricted to the site of MPXV entry, as described in the 2022 outbreak. The involved orthopoxvirus, its route of entry, and the immune status of the host are usually the only determinants of generalized or localized infection. Different mechanisms

of virion entry and egress, as well as virus-encoded host factors, are the main biological determinants (Liu et al., 2019; McFadden, 2005; Moss, 2006; 2016; Roberts and Smith, 2008). Changes in the genome of the current MPXV variant, such as gene loss (Kugelman et al., 2014), may explain both trends.

The MPXV genome is a linear, ≈197-kb-long double-stranded DNA with covalently closed hairpin ends. The genome's densely packed orthologous poxvirus genes (OPGs) (Senkevich et al., 2021) are distributed over a central conserved region ("core") and flanking terminal regions, each of which ends in identical but oppositely oriented ≈6.4-kb-long terminal inverted repetitions (ITRs). Roughly 193 open reading frames (ORFs) encode proteins with ≥60 amino-acid residues. "Housekeeping" proteins involved in MPXV transcription, replication, and virion assembly are encoded by OPGs located in the central conserved region, whereas proteins involved in host range and pathogenesis are mostly encoded by OPGs located in the terminal regions. Like all orthopoxvirus genomes, the MPXV genome contains numerous tandem repeats in the ITRs as well as nucleotide homopolymers all over the genome (Moss and Smith, 2021; Shchelkunov et al., 2002; Wittek and Moss, 1980). However, we also observed other similar structures through the MPXV genome in the form of short tandem repeats (STRs). Moreover, initial observations appear to indicate that these STRs (which may consist of dinucleotide, trinucleotide, or more complex palindromic repeats) are localized in areas where more variation is observed suggesting a crucial role in the MXPV biology and evolution.

Orthopoxviruses rapidly acquire higher fitness by massive gene amplification (genome expansion) when encountering severe bottlenecks *in vitro*. This amplification, akin to gene reduplication in organismal evolution, enables gene copies to accumulate mutations, potentially resulting in new protein variants that can overcome the bottlenecks. Subsequent gene copy reduction (genome contraction) offsets the costs associated with increasing

genome length, thereby retaining the adaptive mutations (Elde et al., 2012). Orthopoxviruses also rapidly adapt to selective pressures by single-nucleotide insertions (genome expansion) or deletions (genome contractions) within poly-A or poly-T stretches, resulting in easily reversible gene-inactivating or re-activating frameshifts (Senkevich et al., 2020). These rhythmic genome expansions and contractions are referred to as "genomic accordions" at the gene and base level (Elde *et al.*, 2012). Given the overall conservation of STRs in orthopoxvirus genomes, we hypothesized that their variation could be a third type of genomic accordion and that, overall, this type of adaptation (which we designate here as low-complexity regions [LCRs]), rather than single-nucleotide polymorphisms (SNPs), could be the key to understanding the unusual epidemiology of 2022 subclade IIb MPXV.

RESULTS

De novo assembly of subclade II lineage B.1 monkeypox virus (MPXV) genome sequence 353R

Using a template-based mapping approach, shotgun metagenomic short-read-based sequencing of nucleic acids in vesicular lesion swabs from Spanish monkeypox patients resulted in the determination of 48 MPXV consensus genome sequences with at least 10X read depth. A median of 39,697,742 high-quality reads per swab (maximum = 111,030,976; minimum = 7,780,032) were obtained using a NovaSeq 6000 Illumina sequencer. Although 98.12% of the reads were assigned as being of human origin, a median of 74,085 MPXV reads (maximum = 27,516,891; minimum = 30,854) sufficed to cover >99% of the genome (**Table S1**).

Read mapping indicated that, as expected, LCRs of the MPXV genome were mostly unresolved. More importantly, those results were biased by the MPXV genome sequence used for mapping (subclade II lineage A MPXV isolate MPXV-M5312_HM12_Rivers):

LCRs were resolved by the used reference-mapping software tools "following" the sequence provided in the reference genome instead of reporting the actual sequence (**Figure S1A**). To determine the actual LCR sequence, we explored different assembly strategies generally used for resolving eukaryotic genomes, which mostly combine different sequencing technologies. To increase the chances of success, we applied these technologies to a monkeypox patient sample with a high proportion of high-quality viral reads (swab 353R). The *de novo* assembly obtained from NovaSeq (2x150-bp pair-ended reads), MiSeq (2x300-bp pair-ended reads), and Nanopore sequencing generated 3, 2, and 1 contigs belonging to MPXV, covering 97%, 97%, and 101% of the MPXV-M5312_HM12_Rivers sequence, respectively (**Figure 1**).

Characterization and validation of non-randomly distributed low-complexity regions (LCRs) in monkeypox virus (MPXV) genome sequence 353R

We applied a systematic approach for LCR discovery to the MPXV 353R sequence that resulted in the identification of 21 LCRs (13 STRs, 8 homopolymers; **Tables 1** and **S2**). Two pairs of LCRs (1/4 and 10/11) are located in the ITRs and are identical copies in reverse-complementary form. Consequently, we moved forward with 19 LCRs.

In general, LCRs were resolved using the assembly obtained from single-molecule sequencing and further validated using short-read sequencing since most sequences were 13–67 bp long and therefore were covered by reads from each side or flanking region without mismatches (**Figure S1B**; **File S1)**. All LCRs (except pair 1/4 and 3) were validated this way. LCR pair 1/4 (256 bp) and LCR3 (468 bp) were only resolved with single-molecule sequencing reads due to their lengths (**Table 2**).

LCR3 contains a complex tandem repeat with the sequence ATAT $[ACATTATAT]_n$. Our analysis indicated n=52. No publicly available MPXV genome sequence contains a tandem repeat of similar length. However, applying the analysis to 35 publicly available

MPXV National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) datasets of single-molecule raw reads allowed us to resolve the LCR3 of some (**Figure 2A).** Fifteen datasets revealed supporting long reads that include both LCR3 flanking regions. Interestingly, four subclade IIb lineage B.1 MPXV sequences associated with the 2022 monkeypox epidemic and available in SRA have n=54–62 repeats in LCR3. Their number of repeats separate these sequences from 2018–2019 subclade IIb lineage A sequences that have n=12–42 repeats in LCR3, indicating LCR3 as a region of genomic instability and high variability.

LCR pair 1/4 contains a tandem repeat with the sequence $[AACTAACTTATGACTT]_n$. Our analysis indicated n=16. Instead, the sequences of the subclade IIb lineage B.1 MPXV isolate MPXV_USA_2022_MA001 and lineage A reference isolate MPXV-M5312_HM12_Rivers LCR pair 1/4 have n=8 (**Table 3**). Inclusion of the NCBI SRA datasets into the analysis confirmed the n=16 value (**Figure 2B)**. In addition, the analysis revealed subclade IIb lineage-specific repeat differences in LCR pair 1/4. Lineage A.1 viruses are polymorphic, having 14 repeats (n=1 sequences), 16 (n=3), 17 (n=7), and 19 (n=1); lineage A.2 viruses have n=23–26 repeats; lineage A viruses have n=32, 43, 53, or 71 repeats. In contrast, lineage B1 viruses consistently have n=16. While LCR3 appears to have "increased" in length since the spillover, LCR pair 1/4 appears to be decreasing in length, thus behaving like an "accordion" over time.

The subclade II lineage B.1 353R and MPXV_USA_2022_MA001 genome sequences have the same 67 SNPs called against the subclade II lineage A reference isolate MPXV-M5312_HM12_Rivers genome sequence. Additionally, the 353R sequence has two additional paired SNPs in the left and right ITRs (5,595G→A; 191,615C→T compared with the MPXV-M5312_HM12_Rivers sequence) that result in the introduction of a stop codon in OPG015. We observed this variation in only two other patients among our sample set. The

353R and MPXV_USA_2022_MA001 sequences also differ by two number of insertion (ins) or deletion (del) of bases (indels) at positions 133,077 and 173,273, respectively, which correspond to differences in LCR2 and LCR5, respectively. As a result of the resolution LCRs, sequence 353R differs by 1,342 bp in genome length compared with the MPXV_USA_2022_MA001 sequence and 1,338 bp compared with the MPXV-M5312_HM12_Rivers sequence. Most of the variation is due to differences in the length of LCR pair 1/4 and LCR3, along with minor length differences in LCR2, LCR5, and LCR pair 10/11 (**Table 3**). In general, the number of repeats (n) found with the hybrid assembly approach we used here doubled the length of LCRs.

Based on the higher resolution of the MPXV 353R genome sequence, in particular regarding LCRs, we propose this sequence as the new MPXV high-quality genome (HQG) reference sequence according to the sequence quality standards defined in (Ladner et al., 2014).

Low-complexity regions (LCRs) are non-randomly distributed in the monkeypox virus (MPXV) genome

We compared the distribution of LCRs between the different major functional protein OPG groups following the classification of (Senkevich *et al.*, 2021). Differences between functional groups were statistically significant (Kruskal–Wallis test, χ^2 p-value <0.001). Pairwise analysis demonstrated that the functional group "core" (orthopoxvirus genomic central conserved region) includes LCRs at a significantly lower frequency (multiple pairwise-comparison Wilcoxon test) than functional groups "ANK/PRANC" (false discovery rate [FDR]-corrected p-value <0.0001), "Bcl-2 domain" (corrected p-value = 0.04), and "accessory" (FDR-corrected p-value <0.0001) (**Figure 3A**). These analyses indicate that

LCRs in orthopoxvirus genomes are non-randomly distributed and that there is a significant purifying selection force against introducing LCRs in central conserved region areas.

Next, we compared the degree of diversity among the 21 identified LCRs with the observed SNP variability that had been the focus of the field. In the 353R HQG sequence, LCRs 2, 5, 7, 10, 11, and 21 had intra-host genetic diversity, with entropy values that ranged from 0.18 (LCR7) to 1.66 (LCR2), with an average of 0.81 and a standard deviation (SD) of 0.64 among them (**Table 2**). Only five nucleotide positions (1,285; 6,412; 88,807; 133,894; and 145,431) had intra-host genetic diversity at the SNP level. The entropy values ranged from 0.17 (position 133,894) to 0.69 (position 6,412), with an average of 0.38 and a SD=0.21 among them. Interestingly, a student's t-test revealed a significantly higher level of diversity in LCRs than in SNPs (p-value = 0.021; **Figure 3B**).

Then, we characterized, collected, and compared the allele frequencies for all LCRs from all dataset samples (**Table S3**) applying the filters described above. Our analyses revealed that the average inter-sample Euclidean distances at LCRs ranged from 0.05 (LCR21) to 0.73 (LCR2) (**Figure 3C**). We found statistically significant differences between LCRs (Kruskal–Wallis χ^2 p-value <0.001). More specifically, multiple pairwise comparison Wilcoxon test results showed that all LCRs have significantly different levels of inter-sample distances (FDR-corrected p-values < 0.001), except in case of the LCR10 versus LCR11 (FDR-corrected p-value = 0.48) and LCR2 versus LCR5 (FDR-corrected p-value = 0.25) (**Figure 3C**). Average distances in SNPs were 0.0018–0.4168. Our randomization tests revealed that all LCRs have a significantly higher level of inter-sample diversity than the SNPs (all FDR-corrected p-values <0.05) (**Figure 3C**). These analyses uncovered that most of the variability in the orthopoxvirus genome is located in LCR. Consequently, we posit that studies focusing on phenotypic MPXV (and likely other orthopoxvirus) clade-/subclade-

/lineage-/strain differences due to genomic sequence variation should change their focus to the study of LCR variability instead of SNP variability.

Low-complexity regions (LCRs) might be more phylogenetically informative than single-nucleotide polymorphisms (SNPs) for inter-host sequence analysis

Analysis of only two monkeypox patient samples (353R and 349R) resulted in sufficient sequence coverage information to enable allele frequency comparison in most LCRs (**Figure 4A**). Their side-by-side comparison revealed differences in allele frequency in some of them (LCR2, LCR5, and LCR pair 10/11) (**Figure 4B**). The sequence coverage achieved with the remaining samples only enabled to unequivocally resolve an LCR subset (i.e., covering both flanking regions: LCRs 2, 5, 7, 8, 9, and pair 10/11). LCR8 and LCR9 were identical across the entire sample set. However, LCR7 and LCR pair 10/11 had considerable intra-host variation, as well as differences in the preponderant allele (LCR pair 10/11) between samples (**Figure 4C**).

Phylogenetic (**Figure S2A**) and haplotype network (**Figure S2B**) analyses of the sequences determined from the monkeypox patient samples yielded limited information regarding the outbreak. Most sequences were highly similar and are, therefore, part of the basal ancestral MPXV subclade IIb lineage B1 node. Some sequences formed supported clusters: Sequences clustered into groups (**Table S4**):

- group 1 (lineage B.1): sequences from patients 395, 399, and 441 and Floridian MPXV isolate MPXV_USA_2022_FL002 (GenBank #ON676704);
- group 2 (lineage B.1): sequences from patients 347, 352R, 353R, 416, and Spanish MPXV isolate MPXV/ES0001/HUGTiP/2022 (GenBank #ON622718); all share a stop-codon mutation in OPG015;

- group 3 (lineage B.1.3): sequences from patients 22,369 formed with Slovenian MPXV isolate SLO (GenBank #ON609725.2), French isolate MPXV_FR_HCL0001_2022 (GenBank #ON622722), and 38 other sequences worldwide; defined by NBT03_gp174 mutation G190,660A, resulting in an R84K amino-acid residue change;
- group 4 (lineage B.1): sequences from patients 417 and 2,437;
- group 5 (lineage B.1.1): sequences from patients 698; 1,300; 2,388; 2,428; German MPXV isolate MPXV/Germany/2022/RKI01 (GenBank #ON637938.1) and 97 other sequences worldwide; defined by OPG094 mutation G74,360A resulting in a R194H amino-acid residue change); and
- group 6 (lineage B.1): sequences from patients 2,309 and 2,317.

Only one epidemiological link among group samples was identified; patients 395 and 399 came from sexual partners who attended events in Portugal and Spain.

In summary, at least at this time in the monkeypox epidemic, there appears to be limited value in full-genome SNP characterization for transmission analysis.

Conservation and variation in proteins encoded by orthologous poxvirus gene (OPG) and codon usage analysis in OPG low-complexity regions (LCRs)

Analyses showed that LCRs were associated with intra-host and inter-sample variation. Although most LCRs that showed variability in our sample set (pair 1/4, 2, 3, 5, 7, pair 10/11, and 21), three (3, 7, and 21) are located in regions that, considering orthopoxvirus evolutionary history, are associated with virulence or transmission (Chen et al., 2005; Kastenmayer et al., 2014). Noteworthy, three of the 21 highly repetitive areas identified in our intra-host variation analysis (those of LCRs 5, 6, and 7) are located in a defined central conserved region of the orthopoxvirus genome between positions 130,000 and 138,000

(**Figure 1**). This region contains OPG152 (which is truncated in the MPXV genome), OPG153 (directly affected by LCR7), and OPG154. LCR7 is the only STR that is located at the center of a functional ORF). In contrast, LCR3 and LCR21 are situated in the promoter/start area, potentially modifying the ORF start site. The repeat area of LCR7 encodes a poly-D homopolymer in a nonstructured region of OPG153 (**Figure 5A**). The changes we uncovered result in the insertion of two isoleucyls. This change resembles the primary structure found in clade I viruses. In contrast, pre-2017 African subclade IIa viruses lack such insertions.

Another region of potential functional impact is the area between 170,000 to 180,000 that includes LCRs 19, 20, 2, 21 and 3) (**Figure 1**). The LCR3 repeat $[CATTATATA]_n$ is located 21 bp upstream of the putative translation start site of OPG208. Importantly, a methionyl codon is located immediately upstream of LCR3. The usage of this start codon would result in the introduction of an Ile-Ile-Tyr repeat. This codon has a medium to low probability of being used as start codon in the cognate mRNA (T base in position -3), compared to a "strong" Kozak sequence of the downstream putative start codon. Nevertheless, LCR3 remained in-frame in all clade II MPXV samples, indicating selective pressure to maintain the possibility of alternative start translation (**Figure 5B**). Interestingly, LCR3 is not located in-frame in most clade I viruses. This may be significant because OPG208 is a member of a set of genes most likely responsible for increased virulence of clade I MPXV compared to classical (pre-epidemic) clade IIa MPXV (Chen *et al.*, 2005). The LCR3 tandem repeat $[CATTATATA]_n$ is present with n=52, 54, and 62 copies in epidemic subclade IIb lineage B viruses (**Figure 2A**), whereas it is n=7, 37, and 27 in subclade IIa MPXV isolates Sierra Leone (GenBank #AY741551), MPXV-WRAIR7-61 (GenBank #AY603973), and MPXV-COP-58 (GenBank #AY753185) sequences, respectively (Chen *et al.*, 2005), as well as n=16 for clade I MPXV isolate Zaire-96-I-16 (GenBank #AF380138).

Interestingly, all publicly available subclade II lineage A single-molecule long-read sequence data imply a repeat n<40 (**Figure 2A**). Alternatively, the LCR3 repeat sequence could also alter promoter function. Interestingly, the repeat sequence also introduces codons with a low usage ratio that are not be optimized for expression in primates. The codon triplet ATA, which encodes isoleucyl, has a rare codon usage of 0.17 (Stothard, 2000) (**Figure 5B**).

Similarly, the STR downstream of LCR21 introduces a methionyl codon upstream of the putative start codon for OPG204 (**Figure 5C**). Kozak sequence analysis revealed a medium to high probability for translation compared with the putative start codon (**Figure 5C**).

The remaining LCRs (2, pair 4/1, and pair 10/11) are located downstream of known ORFs; thus, their variation is less likely to be associated with a change in phenotype.

DISCUSSION

MPXV subclade IIb traces back to a human MPXV infection that likely occurred after spillover from a local animal reservoir in Ihie, Abia State, Nigeria, in 1971. An additional 10 human infections with MPXV of this lineage were detected through 1978, when this lineage seemed to have disappeared. However, in 2017, it reemerged in Yenagoa, Bayelsa State, Nigeria (Faye *et al.*, 2018). Since then, hundreds of monkeypox cases have been reported and MPXV belonging to the IIb lineage has been sampled in several countries. But, there were no secondary cases prior to the 2022 epidemic (Cohen-Gihon et al., 2020; Ng et al., 2019; Vaughan et al., 2018; Yong et al., 2020).

Subclade IIb viruses cause monkeypox that presents differently than the classical disease caused by clade I and subclade IIa viruses, i.e., subclade IIb infections are associated with higher prevalence among adults rather than adolescents, are predominant in the MSM community, and are efficiently transmitted human-to-human from localized infectious skin lesions rather than requiring disseminated infection (Bunge *et al.*, 2022; Otu *et al.*, 2022; Thornhill *et al.*, 2022; Ulaeto *et al.*, 2022; Vusirikala *et al.*, 2022).

Comparative genomics demonstrated obvious relationships between orthopoxvirus genotype and phenotype, driven by selective pressure from hosts (Baroudy and Moss, 1982; Chen *et al.*, 2005; Esposito et al., 2006; Gubser et al., 2004; Hendrickson et al., 2010; Kugelman *et al.*, 2014; Shchelkunov, 2012; Shchelkunov et al., 2001). Consequently, it was expected that increased MPXV genotype IIb human-to-human transmission would go hand in hand with genotypic changes. However, since orthopoxviruses genomes are organized in redundant ways (Bratke and McLysaght, 2008; Elde *et al.*, 2012; McLysaght et al., 2003; Senkevich *et al.*, 2020), genotypic changes were expected to be modulating, rather than radical.

Thus far, genomic characterization of the 2022 epidemic has focused on describing its evolutionary history and tracking MPXV introductions into western countries. The 2022 MPXV cluster diverges from its predecessor viruses by an average of 50 SNPs. Of these, the majority (n=24) are non-synonymous mutations with a second minority subset of synonymous mutations (n=18) and a few intergenic differences (n=4) (Isidro et al., 2022). A strong mutational bias was mainly attributed to the potential action of apolipoprotein B mRNA-editing catalytic polypeptide-like 3 (APOBEC3) enzymes (O'Tool and Rambaut, 2022). Genetic variation, including deletion of immunomodulatory genes, also occurred (Jones et al., 2022). MPXV sublineages mostly represent very small variations, usually characterized by one or two SNP differences to basal nodes in phylogenetic analyses (Nextstrain, 2022). Four of the MPXV genome sequences determined in this study can be assigned to global lineage B.1.1, one sequence could be assigned to global lineage B.1.3, and the remaining new sequences belong to lineage B1. We detected additional clusters that were also defined by few SNPs, but only in one case could we identify an epidemiological link. Thus, there appears to be a limited relationship between SNPs and epidemiology, which might hint to sequencing errors. Consequently, MPXV genomic epidemiology might need a change of focus.

Our analyses located considerable MPXV genomic variability in areas previously considered of poor informative value, i.e., in LCRs. Because LCR entropy is significantly higher than that of SNPs and LCRs are not randomly distributed in defined coding areas in the genome, person-to-person transmission-associated changes were observed in the immunomodulatory region (Kugelman *et al.*, 2014), and genomic accordions are a rapid path for adaptation of orthopoxvirus during serial passaging (Elde *et al.*, 2012; Senkevich *et al.*, 2020), we posit that LCR changes might be associated with MPXV transmissibility differences over time.

Eight LCRs had evident signs of intra-host and inter-sample variation (pair 1/4, 2, 3, 5, 6, 7, pair 10/11, 21). Five of them (5, 6, 7, 3, and 21) were co-located in two areas of the MPXV genome: base pairs 130,000–135,000 (5, 6, and 7) which are in the central conserved region of the OPXV genome in which most "housekeeping" genes are located; and base pairs 170,000–180,000 (3 and 21), which are located in the immunomodulatory area (**Figure 1**). Three of those LCRs are located inside the putative translated regions of genes OPG153, OPG204, and OPG208. Changes in OPG204 and OPG208 are located near the N-terminal region and might involve modulating the expression of translation. The changes observed in OPG153 stand out as they are located inside a region that is under high selective pressure for transmission in a "housekeeping" orthologous poxvirus gene, which is involved in virion attachment and egress (Senkevich *et al.*, 2021). Thus, we urge that these areas be scrutinized for changes that might affect the MPXV interactome.

The OPG153 repeat results in a poly-Asp amino-acid homopolymer string (**Figure 5A**); the LCR repeat in OPG208 results in an Ile-Ile-Tyr repeat (**Figure 5B**); and the N-terminal domain variation in OPG204 results in a Met-Lys repeat (**Figure 5C**). Self-association guided by stretches of single amino-acid residue repeats may lead to the formation of aggregates (Oma et al., 2004). Many human diseases are associated with detrimental effects of homopolymers (Gatchel and Zoghbi, 2005; Shoubridge et al., 2007). Expansion or contraction of the repeat may increase self-attraction and trigger disease. These homopolymers also regulate the activity of transcription factors (Gemayel et al., 2015) or direct proteins to different cellular compartments (Oma *et al.*, 2004; Salichs et al., 2009). Modulation of ORF translation via MPXV LCR-like repeats also has been described for various microbes. For instance, the functionality of baker's yeast proteins Flo1p and Flo11p is proportionally modulated by the repeat length of their N-terminal regions. A relatively small change in the number of tandem repeats is crucial for yeast adaptation to a new

environment (Fidalgo et al., 2006; Verstrepen et al., 2005). Another example is glutamic acid-rich protein (GARP) of plasmodia, which contains repetitive sequences that direct the protein to the periphery of the infected erythrocyte. At least nine other exported plasmodium proteins target the periphery of the erythrocyte using this strategy. Interestingly, the lengths of the tandem repeats vary among plasmodium strains (Davies et al., 2017; Davies et al., 2016).

Protein translation rates are in part regulated by the availability of mRNA codon-cognate aminoacylated tRNAs. Homopolymers composed of codons for rare tRNAs directly diminish translation via tRNA depletion. The non-optimal tyrosyl codons in MPXV LCRs 3 and 21 suggest such translation modulation for OPG208 and OPG204, respectively. The protein encoded by OPG208, B19R (Cop-K2L, SPI-1) is a serine protease inhibitor-like protein that functions as an inhibitor of apoptosis in VACV-infected cells (Kettle et al., 1995; Kotwal and Moss, 1989), which could prevent VACV proliferation and protect nearby cells (Brooks et al., 1995; Jorgensen et al., 2017). Consequently, MPXV B19R is now considered a potential MPXV virulence marker (Chen *et al.*, 2005). The protein encoded by OPG204, B16R, is a secreted decoy receptor for interferon type I (Colamonici et al., 1995; Hernáez et al., 2018). We did not observe any repeat number changes OPG204-associated LCR21, but SNPs in clade I, subclade IIa, and subclade IIb result in alternative translational start sites followed by a, suggesting these SNPs could have direct effects on OPG204 translation.

The most intriguing finding from our dataset involves LCR7 in OPG153. The OPG153 expression product (A26L) attaches orthopoxvirus particles to laminin (Chiu et al., 2007) and regulates orthopoxvirus particle egress (Howard et al., 2008; Kastenmayer *et al.*, 2014; Liu et al., 2014; Ulaeto et al., 1996), thereby modulating key steps in the virus lifecycle. OPG153 is unique as it is the central conserved region gene that has been “lost” the most times during orthopoxvirus evolution (Senkevich *et al.*, 2021). Inactivation of OPG153

genes by frameshift mutations occurs rapidly in experimental orthopoxvirus evolution models (Senkevich *et al.*, 2020), resulting increased virus replication levels, changes in particle morphogenesis, decreased particle-to-PFU ratios, and differences in pathogenesis (Kastenmayer *et al.*, 2014; Senkevich *et al.*, 2020). Finally, A26L is the main target of the host antibody response to orthopoxvirus infection (Keasey et al., 2010; Pugh et al., 2016). Thus, any genomic change that modulates OPG153 is likely of significance. LCR7 encodes a poly-D non-structured region that is conserved among orthopoxviruses A26Ls; however, its length is highly variable. In mammals, poly-D stretches appear to provide functionality to asporin, a small leucin rich repeat proteoglycan that also possesses a unique stretch of aspartyls at its N terminus (Henry et al., 2001), associated with calcium-binding (Zhu et al., 2018). Interestingly, orthopoxviruses that form a dense protein matrix within the cytoplasm called A-type inclusions (ATIs), such as MPXV, generically have very long poly-D stretches in this region, whereas orthopoxviruses that do not form ATIs, such as VACV and VARV, have reduced their LCR7 poly-D stretches to four Ds. Even among MPXV clades, patterns are observable. Subclade IIa viruses have an extended 21 amino-acid residue poly-D stretch, whereas clade I and subclade IIb viruses have poly-D stretches with two inserted isoleucyls. Intriguingly, both insertions result from the incorporation of the same "ATCATA" nucleotide insertion in the "GAT" repetitive stretch.

In summary, our findings expand the concept of genome accordions as a simple and recurrent mechanism of adaptation on a genomic scale in orthopoxvirus evolution. A consequence of this broadening is the recognition that MPXV genome LCRs might hold the key to improved understanding of current monkeypox epidemiology and clinical presentation. A new standardized approach to generate and analyze sequencing data via prioritizing LCR characterization and subsequent functional mechanism-of-action studies are warranted.

MATERIALS AND METHODS

Study population

This study includes confirmed human monkeypox cases diagnosed from May 18 to July 14, 2022, at the Centro Nacional de Microbiología (CNM), Instituto de Salud Carlos III, Madrid, Spain. The study was performed as part of the public health response to the current monkeypox epidemic by the Spanish Ministry of Health. Sample information is listed in **Tables S1** and **S5**.

The samples used in this work were obtained in the context of the Microbiological Surveillance and Diagnosis Program for the Monkeypox Outbreak of the Centro Nacional de Microbiología, Instituto de Salud Carlos III. The study was based on routine testing, did not involve any additional sampling or tests and stored RNA extracts were used, so specific ethical approval was not required for this study. All sequenced viruses corresponded to those to patients that gave consent to be analyzed for diagnosis or surveillance purposes.

Study sample processing

Swabs of vesicular lesions from study patients in viral transport media were sent refrigerated to CNM. Nucleic acids were extracted at CNM using either QIAamp MinElute Virus Spin (DNA) or QIAamp Viral RNA Mini kits (Qiagen, Germantown, MD, USA) according to the manufacturer's recommendations. Inactivation of samples was conducted in a certified class II biological safety cabinet in a biosafety level (BSL) 2 laboratory using BSL-3 best practices with appropriate personal protective equipment.

Monkeypox virus (MPXV) laboratory confirmation

MPXV detection by PCR in a sample was considered laboratory confirmation and resulted in inclusion of the swab in the study. A previously described orthopoxvirus-generic real-time

PCR (qPCR) was used for screening (Fedele et al., 2006). A previously described conventional validated nested PCR targeting OPG002 (encoding a TFN receptor) was used for results confirmation (Sánchez-Seco et al., 2006).

MPXV genome sequencing

Sequencing libraries were prepared with a tagmentation-based Illumina DNA Prep kit (Illumina, San Diego, CA, USA) and run in a NovaSeq 6000 SP Reagent Kit (Illumina) flow cell using 2x150 paired-end sequencing. To improve assembly quality, the library from swab 353R, an unpassaged vesicular fluid from a confirmed case, was also run in a MiSeq Reagent Kit v3 (Illumina) flow cell using 2 x 300 paired-end sequencing. Additionally, sample 353R was also analyzed by single-molecule methods using Nanopore sequencing (Oxford Nanopore Technologies, Oxford, UK). For Nanopore sequencing, 210 ng of DNA was extracted from swab 353R and used to prepare a sequence library with a Rapid Sequencing Kit (Oxford Nanopore Technologies); the library was analyzed in an FLO-MIN106D (Oxford Nanopore Technologies) flow cell for 25 h. The process rendered 1.12 Gb of filter-passed bases.

Bioinformatics

***De novo* assembly and annotation of subclade II lineage B.1 MPXV genome sequence 353R.** Due to the high yield of MPXV genomic material in a preparatory run, swab 353R was selected as source material for the determination of an MPXV high-quality genome (HQG) sequence. Single-molecule long-sequencing reads were preprocessed using Porechop v0.3.2pre (Wick et al., 2017) with default parameters. Reads were *de novo* assembled using Flye v2.9-b1768 (Kolmogorov et al., 2019) in single-molecule sequencing raw read mode with default parameters, resulting in one MPXV contig of 198,254 bp. Short 2x150

554 sequencing reads were mapped with Bowtie2 v2.4.4 (Langmead and Salzberg, 2012) against
555 the selected contig, and resulting BAM files were used to correct the assembly using Pilon
556 v1.24 (Walker et al., 2014). At this intermediate step, this corrected sequence was used as a
557 reference in the nf-core/viralrecon v2.4.1 pipeline (Patel et al., 2022) for mapping and
558 consensus generation with short sequencing reads. The allele frequency threshold of 0.5 was
559 used for including variant positions in the corrected contig.
560 Short MiSeq 2x300 and NovaSeq 2x150 sequencing reads were also assembled *de*
561 *novo* using the nf-core/viralrecon v2.4.1 pipeline, written in Nextflow (Di Tommaso et al.,
562 2017) in collaboration between the nf-core community (Ewels et al., 2020) and the Unidad de
563 Bioinformática, Instituto de Salud Carlos III, Madrid, Spain (https://github.com/BU-ISCIII).
564 FASTQ files containing raw reads were quality controlled using FASTQC v0.11.9 (Andrews,
565 2010). Raw reads were trimmed using fastp v0.23.2 (Chen et al., 2018). The sliding-window
566 quality-filtering approach was performed, scanning the read with a 4-base-wide sliding
567 window and cutting 3' and 5' base ends when average quality per base dropped below a
568 Qphred33 of 20. Reads shorter than 50 bp and reads with more than 10% read quality under
569 Qphred 20 were removed. Host genome reads were removed via kmer-based mapping of the
570 trimmed reads against the human genome reference sequence GRCh38
571 (https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000001405.26/) using Kraken 2
572 v2.1.2 (Wood et al., 2019). The remaining non-host reads were assembled using SPADES
573 v3.15.3 (Antipov et al., 2016; Prjibelski et al., 2020) in rnaviral mode. A fully ordered MPXV
574 genome sequence was generated using ABACAS v1.3.1 (Assefa et al., 2009), based on the
575 MPXV isolate MPXV_USA_2022_MA001 (Nextstrain subclade IIb lineage B.1) sequence
576 (GenBank #ON563414.3) (Gigante et al., 2022). The independently obtained *de novo*
577 assemblies and reference-based consensus genomes obtained from swab 353R were aligned

using MAFFT v7.475 (Katoh et al., 2019) and visually inspected for variation using Jalview v2.11.0 (Waterhouse et al., 2009).

Systematic identification of low-complexity regions in orthopoxvirus genomes. Detection of short tandem repeats (STRs) in the HQG sequence and other orthopoxvirus genomes was performed with Tandem repeats finder (Benson, 1999), using default parameters. Briefly, the algorithm works without the need to specify either the pattern or its length. Tandem repeats are identified considering percent identity and frequency of insertion (ins) or deletion (del) of bases (indels) between adjacent pattern copies, using statistically based recognition criteria. Since Tandem repeats finder does not detect single-nucleotide repeats, we developed an R script to systematically identify homopolymers of at least 9 nucleotide residues in all available orthopoxvirus genome sequences. STRs and homopolymers were annotated as low-complexity regions (LCRs).

Curation of low-complexity regions in the MPXV high-quality virus genome sequence. We curated LCRs in the HQG sequence using a modified version of STRsearch (Wang et al., 2020). Once provided with identifying flanking regions, STRsearch performed a profile analysis of STRs in massively parallel sequencing data. To ensure high-quality characterization of LCR alleles, we modified the script (https://github.com/BU-ISCIII/MPXstreveal) to complement reverse reads that map against the reverse genome strand according to their BAM flags. In addition, output was modified to add information later accessed by a custom Python script to select only reads containing both LCR flanking regions. All LCRs in the HQG sequence were manually validated using STRsearch results and *de novo* assemblies obtained from all sequencing approaches. When an LCR was only resolved by single-molecule long-sequencing technologies (LCR pair 1/4 and LCR3), we also analyzed publicly available data by downloading all single-molecule long-sequencing data from the National Center for Biotechnology Information (NCBI) Sequence Read Archive

(SRA) (https://www.ncbi.nlm.nih.gov/sra) as of August 10, 2022, and analyzed the data according to **File S1**.

Final MPXV high-quality virus genome sequence assembly. The consensus genome constructed with the nf-core/viralrecon v2.4.1 pipeline using the corrected *de novo* contig as stated above, along with the resulting curated and validated consensus LCRs, were used to build the final HQG reference sequence using a custom Python script. The resulting HQG is available from the European Nucleotide Archive (#OX044336.2).

Generation of MPXV high-quality virus genome reference-based consensus sequence for all other samples. For the remaining specimens, sequencing reads were analyzed for MPXV genome sequence determination using the nf-core/viralrecon v2.4.1 pipeline. Trimmed reads were mapped with Bowtie2 v2.4.4 against the HQG sequence and the sequence of subclade II lineage A MPXV isolate M5312_HM12_Rivers (GenBank #MT903340.1) (Mauldin et al., 2022). Picard v2.26.10 (The Broad Institute, 2018) and SAMtools v1.14 (Li et al., 2009) were used to generate MPXV genome mapping statistics. iVar v1.3.1 (Grubaugh et al., 2019), which calls for low-frequency and high-frequency variants, was used for variant calling. Variants with an allele frequency higher than 75% were kept to be included in the consensus genome sequence. BCFtools v1.14 (Danecek et al., 2021) was used to obtain the MPXV genome sequence consensus with filtered variants and masked genomic regions with coverage values lower than 10X. All variants, included or not, in the consensus genome sequence, were annotated using SnpEff v5.0e (Cingolani et al., 2012b), and SnpSift v4.3 (Cingolani et al., 2012a). Final summary reports were created using MultiQC v.1.11 (Ewels et al., 2016). Consensus genome sequences were analyzed with Nextclade v2.4.1 (Aksamentov et al., 2021) using the "MPXV (All clades)" dataset (timestamp 2022-08-19T12:00:00Z). Raw reads and consensus genomes are available from the European

Nucleotide Archive (#ERS12168855–ERS12168865, #ERS12168867, #ERS12168868, #ERS13490510–ERS13490543).

Intra-host and inter-host allele frequency analyses. Intra-host genetic entropy (defined as -sum(Xi*log(Xi)), in which Xi denotes each of the allele frequencies in a position) was calculated according to the single-nucleotide polymorphisms (SNP) frequencies of each position along the genome using nf-core/viralrecon v2.4.1 pipeline results. Similarly, genetic entropy for each LCR was calculated considering the frequencies of repeat lengths.

LCR intra-host and inter-host variations in the sample set were analyzed using the modified version of STRsearch. As a filter for quality for this analysis, STRsearch results (**Table S5**) were filtered, keeping alleles with at least 10 reads spanning the region and allele frequency above 0.03. Quality control and allele frequency graphs were created using a customized R script.

Pairwise genetic distances between samples were calculated as Euclidean distances (defined as /X-Y/=sqrt(sum(xi-yi)^2), in which xi and yi are the allele frequencies of sample X and Y at a given position, respectively), thus accounting for the major and minor alleles at each analyzed position. Distances were calculated individually for each variable LCR (STRs 2, 5, 7, 10, 11, and 21) and for each of all 5,422 SNPs showing inter-sample variability (compared to MPXV-M5312_HM12_Rivers). The distributions of inter-sample distances were compared between LCRs using a Kruskal–Wallis test (χ^2 *p*-values) followed by multiple pairwise-comparison between groups (Wilcoxon test), with *p*-values subjected to the false discovery rate (FDR) correction. A randomization test was used to test whether inter-sample variability in LCRs is higher than that in SNPs: first, the average Euclidean distance for each LCR and each SNP position was calculated; then, the average value of each LCR was compared to a random sample of 1,000 values from the distribution of mean distances from

651 the SNPs along the genome. The *p*-value was calculated from the percentage of times that the
652 mean of the LCR was higher than the randomly taken values from the SNPs.
653 ***Phylogenetic analysis of the MPXV central conserved region.*** Variant calling and SNP
654 matrix generation was performed using Snippy v4.4.5 (Seeman, 2015), including sequence
655 samples and representative MPXV genome sequences downloaded from GenBank (**Table**
656 **S5**). The SNP matrix with both invariant and variant sites was used for phylogenetic analysis
657 using IQ-Tree 2 v. 2.1.4-beta (Minh et al., 2020) via predicted model K3Pu+F+I and 1,000
658 bootstrap replicates. A phylogenetic tree was visualized and annotated using iTOL v6.5.8
659 (Letunic and Bork, 2021). The SNP matrix was also used for generating the haplotype
660 network using PopArt v1.7 (Bandelt et al., 1999).
661 ***Selected MPXV ORF analysis.*** Representative orthopoxvirus genomes (Senkevich *et al.*,
662 2021) were downloaded from GenBank together with the consensus genome sequences from
663 the specimens analyzed in this study (**Table S5**). MPXV genomes were assigned to clades
664 and lineages according to the most recent nomenclature recommendations according to
665 Nextstrain (Nextstrain, 2022) using Nextclade v2.4.1. Annotations from RefSeq
666 #NC_063383.1 (subclade II lineage A MPXV virus isolate MPXV-M5312_HM12_Rivers)
667 GFF file were transferred to all FASTA genome sequences using Liftoff v1.6.3 (Shumate and
668 Salzberg, 2021). OPG153 was extracted using AGAT v0.9.1 (Dainat et al., 2022) and multi-
669 FASTA files were generated for each group and gene. OPG204 and OPG208 alternative
670 annotation start site ORFs were re-annotated in Geneious Prime (Biomatters, San Diego, CA,
671 USA), and extracted as new alignments. We used MUSCLE v3.8.1551 for aligning each
672 multi-FASTA file and Jalview v2.11.0 for inspecting and editing the alignments. Finally,
673 MetaLogo v1.1.2 (Chen et al., 2022) was used for creating and aligning the sequence logos
674 for each orthopoxvirus group of the OPG153/LCR7, OPG204/LCR21, and OPG208/LCR3
675 areas.

Comparison of LCR frequencies in protein functional groups. The potential biological impact of LCRs was evaluated by mapping the frequency and location of STRs and homopolymers in the orthopoxvirus genome and considering the biological function of the affected genes. The frequency of inclusion of LCRs between distinct functional groups of genes was compared as previously described (Senkevich *et al.*, 2021). Orthopoxviruses (n=231, Akhmeta virus [AKMV]: n=6 sequences; alaskapox virus [AKPV]: n=1; cowpox virus [CPXV]: n=82; ectromelia virus [ECTV]: n=5; MPXV: n=62; VACV: n=18; VARV: n=57) include 216 functionally annotated OPGs classified in 5 categories ("Housekeeping genes/Core" ANK/PRANC family, Bcl-2 domain family, PIE family, and "Accessory/Other" [e.g., virus-host interacting genes]). The frequency was calculated after normalizing count numbers with the sample size of the OPG alignment. Statistical analysis of the significance of differences was performed by applying a Kruskal–Wallis test (χ^2 *p*-values) followed by a non-parametric multiple pairwise comparison between groups (Wilcoxon test), with *p*-values subjected to FDR correction.

690 **ACKNOWLEDGMENTS**

691 We would like to thank the work of the Rapid Response Unit of the National Center for
692 Microbiology, especially MªJosé Buitrago, and Cristobal Belda, ISCIII General Director. We
693 also thank Anya Crane (Integrated Research Facility at Fort Detrick, National Institute of
694 Allergy and Infectious Diseases, National Institutes of Health) for critically editing the
695 manuscript and Jiro Wada (Integrated Research Facility at Fort Detrick, National Institute of
696 Allergy and Infectious Diseases, National Institutes of Health) for helping with figure
697 preparation.

698 The work for this study at Instituto de Salud Carlos III was partially funded by Acción
699 Estratégica "Impacto clínico y microbiológico del brote por el virus de la viruela del mono en
700 pacientes en España (2022): proyecto multicéntrico MONKPOX-ESP22" (CIBERINFEC).
701 The work for this study at the Department of Microbiology, Icahn School of Medicine at
702 Mount Sinai as part of Global Health Emerging Pathogen Institute activities was funded by
703 institutional funds (G.P.) from the Department of Microbiology, Icahn School of Medicine at
704 Mount Sinai in support of Global Health Emerging Pathogen Institute activities.

705 This work was also supported in part through Laulima Government Solutions, LLC,
706 prime contract with the U.S. National Institute of Allergy and Infectious Diseases (NIAID)
707 under Contract No. HHSN272201800013C. J.H.K. performed this work as an employee of
708 Tunnell Government Services (TGS), a subcontractor of Laulima Government Solutions,
709 LLC, under Contract No. HHSN272201800013C.

710 Opinions, interpretations, conclusions, and recommendations are those of the authors
711 and are not necessarily endorsed by the U.S. Army.

712 The views and conclusions contained in this document are those of the authors and
713 should not be interpreted as necessarily representing the official policies, either expressed or
714 implied, of the U.S. Department of Health and Human Services or of the institutions and

715 companies affiliated with the authors, nor does mention of trade names, commercial products,
716 or organizations imply endorsement by the U.S. Government.
717

718 **AUTHOR CONTRIBUTIONS**

719 Conceptualization, S. M., S.V., A. N., I. J., M.S.L., A.G.S., I.C., M.S.S., G.P.
720 Methodology, S.M., S.V., A.N., J.A.P.G., S.V.F., A.Z., J. H. K., M.S.L., N.D., J.R.K., E.G.,
721 S.G., G.P.
722 Investigation, S.M., S.V., A.N., J.A.P.G., S.V.F., A.Z., E.O., O.A., A.M.G., A.D.I., V.E.,
723 C.G., F.M., P.S., M.T., A.V., J.C.G., I.T., M.C.R., L.M., M.L., A.G., L.C., A.G., J.C., L.H.,
724 P.J.S., M.L.N.R., I.J., M.E.A.A., C.L., L.R., I.E., M.S., M.A.M., J.H.K., M.S.L., N.D.P.,
725 J.R.K., E.G., S.G., A.G.S., I.C., M.S.S., G.P.
726 Formal analysis, S.M., S.V., A.N., J.A.P.G., S.V.F., J.H.K., M.S.L., N.D.P., J.R.K., E.G.,
727 I.C., M.S.S, G.P.
728 Writing – original draft, S.M., S.V., A.N., G.P.
729 Writing – review & editing, S. M., S.V., A. N., A.G.S., I.C., M.S.S., G.P.
730 Visualization, S.M., S.V., A.N., G.P.
731 Supervision, A.G.S., I.C., M.S.S., G.P.
732 Resources, A.G.S., I.C., M.S.S., G.P.
733 Funding Acquisition, A.G.S., I.C., M.S.S., G.P.
734

735 **DECLARATION OF INTERESTS**

736 A.G.-S. has consulting agreements for the following companies involving cash and/or stock:
737 Castlevax, Amovir, Vivaldi Biosciences, Contrafect, 7Hills Pharma, Avimex, Vaxalto,
738 Pagoda, Accurius, Esperovax, Farmak, Applied Biological Laboratories, Pharmamar, Paratus,
739 CureLab Oncology, CureLab Veterinary, Synairgen, and Pfizer, outside of the reported work.

A.G.-S. has been an invited speaker in meeting events organized by Seqirus, Janssen, Abbott, and Astrazeneca. A.G.-S. is inventor on patents and patent applications on the use of antivirals and vaccines for the treatment and prevention of virus infections and cancer, owned by the Icahn School of Medicine at Mount Sinai, New York, outside of the reported work. The authors declare no competing interests.

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Gustavo Palacios (gustavo.palacios@mssm.edu).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

All scripts and codes used for this study can be found at github repository: (https://github.com/BU-ISCIII/MPXstreveal).

FIGURE LEGENDS

Figure 1. *De novo* assembly of subclade II lineage B.1 monkeypox virus (MPXV) genome sequence 353R

A visual representation of the fully annotated MPXV isolate 353R genome (based on the subclade II lineage A reference isolate MPXV-M5312_HM12_Rivers genome sequence annotation). Shown are (from the outside to the inside): high-quality genome (HQG) hybrid assembly (wide outer light blue ring); sequencing coverage distribution graph (thin ragged line [green: ≥10,000x, 99.42%; black: 1,000x–10,000, 0.28%; orange: <1,000x-10, 0.1%; red: <10-0, 0.2%]); orthologous poxvirus gene (OPG) annotations according to the standardized nomenclature (Senkevich *et al.*, 2021) (lettering and shaded boxes [orange: ANK/PRANC (N-terminal ankyrin protein with PRANC domain) inverted terminal repetition [ITR] regions; gold: Bcl-2 domain; blue: BTB/Kelch domains; green: housekeeping; purple: other; pink: TNFR and/or PIE domains); contigs from NovaSeq, MiSeq, and Nanopore sequencing (wide inner gray rings). Additionally, radial lines and shading that originate in the center and reach outward on the white background indicate low-complexity regions (LCRs; royal blue) and areas with more change (light blue).

Figure 2. Characterization and validation of non-randomly distributed low-complexity regions (LCRs) in monkeypox virus (MPXV) genome sequence 353R

(A) LCR3 sequence validation using MPXV 353R Nanopore sequencing data and 15 additional raw data sequencing reads downloaded from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA). (B) LCR pair 1/4 sequence validation using MPXV 353R Nanopore sequencing data and 20 additional raw data sequencing reads downloaded from NCBI SRA. Detailed information on the represented materials, along with originator and epidemiological data, is provided in **Table S6**.

Figure 3. Low-complexity regions (LCRs) are non-randomly distributed in the monkeypox virus (MPXV) genome

(A) Frequencies (mean +/- standard error) at which LCRs occur in orthologous poxvirus genes (OPGs) of different functional groups. Shown are functional classes in which pairwise comparisons had a significantly different frequency than other groups (multiple pairwise Wilcoxon test, false discovery rate [FDR] corrected *p*-value <0.05). (B) Entropy value distribution for short tandem repeats (STRs) in LCRs (left) and single-nucleotide polymorphisms (SNPs; right). (C) Distributions of the pairwise inter-sample Euclidean distances for each STR in LCRs (**Table S3**). SNPs in the boxplot represent the distribution of average Euclidean distances of each variable position along the MPXV genome.

Figure 4. Low-complexity regions (LCRs) might be more phylogenetically informative than single-nucleotide polymorphisms (SNPs) for inter-host sequence analysis

Monkeypox virus (MPXV) population genomics within the biological specimen (intra-host) and across different specimens (inter-host). (A) Panel shading indicates the number of reads supporting each LCR for each sample. Only paired reads that include a perfect match to both flanking regions were counted; the gradient shows the maximum value in black and the minimum value (n=1) in the lightest blue. Samples without coverage are indicated in gray. (B) Comparison of LCR allele frequency for samples 353R and 349R. Only LCRs with at least 10 supporting paired reads including both flanking regions were counted; only alleles with a frequency of 0.03 or higher were considered. The gradient shows the maximum value in black and the minimum value (n=0.03) in the lightest blue. (C) Comparison of LCR allele frequency in all samples for LCRs with good coverage (7, pair 10/11, 12, 13, 14, 19, 20, and 21). Only LCRs with at least 10 supporting paired reads including both flanking regions were

807 counted; only alleles with a frequency of 0.03 or higher were considered. The gradient shows
808 the maximum value in black and the minimum value (n=0.03) in the lightest blue.
809
810 **Figure 5. Conservation and variation in proteins encoded by orthologous poxvirus gene**
811 **(OPG) and codon usage analysis in OPG low-complexity regions (LCRs)**
812 MetaLogo visualization of conserved and varying amino-acid residues in OPG-encoded
813 proteins among monkeypox virus (MPXV) clade I, subclade IIa, and subclade IIb, with
814 homologous and nonhomologous sites highlighted. (A) OPG153/ LCR7-derived variability.
815 CMLV, camelpox virus; VACV, vaccinia virus; VARV, variola virus; CPXV, cowpox virus;
816 MPXV, monkeypox virus. (B) OPG204/LCR21-derived variability; and (C) OPG208/LCR3-
817 derived variability. Visualizations created at Biorender.com and Geneious version 2022.2
818 created by Biomatters.

TABLE LEGENDS

Table 1. Low-complexity regions (LCRs) in monkeypox virus (MPXV) genome sequence 353R

Short tandem repeats (STRs) are described using nucleotide base-pair coordinates with reference to the high-quality genome (HQG) sequence (ENA Accession #OX044336). Listed are the number of repeat units, description of the sequence (with $_n$ = number of repeats for this particular genome), identification of the nearest annotated orthologous poxvirus gene (OPG), type of LCR (STR or homopolymer), position of the LCR to the nearest gene, and distance of the LCR to the nearest gene. OPG notations follow the standardized nomenclature (Senkevich *et al.*, 2021); vaccinia virus (VACV) Copenhagen strain and classical VACV gene notations are shown in addition to enable comparisons.

Table 2. Low-complexity region (LCR) validation and entropy-level intra-host analysis in monkeypox virus (MPXV) genome sequence 353R

Listed are the type and number of supporting reads for each LCR. Definitions of quality: Yes, LCR is found entirely in the assembly in one contig; no, LCR is not assembled with the reported method. All LCRs with entropy levels above 0.15 are shaded in gray. OPG, orthologous poxvirus gene.

Table 3. Comparison of monkeypox virus (MPXV) genome sequence 353R with reference sequences

LCR repetitions for each genome are indicated. Discrepant number (n) of LCR repeats are shaded in gray. indels, number of insertion (ins) or deletion (del) of bases; LCR, low-complexity region; SNP, single-nucleotide polymorphism.

841 **SUPPLEMENTAL FIGURE LEGENDS**

842 **Figure S1. Examples for read mapping artefacts and correction in monkeypox virus (MPXV)**
843 **genome low-complexity regions (LCRs)**

844 (A) LCR2 alignment highlighting differences compared with various consensus sequences. (B) LCR7
845 alignment highlighting differences of results obtained using three sequencing platforms compared to
846 the subclade II lineage A monkeypox reference isolate MPXV-M5312_HM12_Rivers sequence.

847

848 **Figure S2. Phylogenetic analysis of monkeypox virus (MPXV)**

849 (A) Phylogenetic maximum-likelihood (ML) tree showing monkeypox virus (MPXV) subclade IIb
850 single-nucleotide polymorphism (SNP) clustering. Bootstrap supports >60 are indicated by labels
851 with their number of supports. (B) Haplotype network showing SNP differences among samples
852 included in the phylogenetic tree. Details on groups can be found in **Table S4**.

853

854 **Figure S3. Conservation and variation in proteins encoded by orthologous poxvirus gene**
855 **(OPG) 208.** (A) MetaLogo visualization of conserved and varying amino-acid residues in OPG-
856 encoded proteins among monkeypox virus (MPXV) clade I, subclade IIa, and subclade IIb, with
857 homologous and nonhomologous sites highlighted; (B) Entropy heatmap; (C) Entropy analysis by
858 site.

859

860 **Figure S4. Conservation and variation in proteins encoded by orthologous poxvirus gene**
861 **(OPG) 153.**

862 (A) Entropy heatmap. CMLV, camelpox virus; VARV, variola virus; VACV, vaccinia virus; CPXV,
863 cowpox virus; MPXV, monkeypox virus. (B) . MPXV, monkeypox virus; CPXV, cowpox virus;
864 VACV, vaccinia virus; VARV, variola virus; CMLV, camelpox virus.

SUPPLEMENTAL TABLE LEGENDS

Table S1. NovaSeq sequencing quality control values

Table S2. Low-complexity regions (LCRs) in the monkeypox virus (MPXV) high-quality genome (HQG) sequence 353R

Listed are annotated positions (according to the reference MPXV-M5312_HM12_Rivers isolate genome sequence), sequence, and flanking regions of each area, as described previously (Phillips et al., 2018). ID, identification; STR, short tandem repeats.

Table S3. Short tandem repeats (STRs) in low-complexity regions (LCRs)

Detailed information on STRs of the entire dataset used for analysis in **Figure 2**.

Table S4. Phylogenetic analysis

Detailed information on groups depicted in **Figure S2**. SNP, single nucleotide polymorphism.

Table S5. Genomes used in the study

GI, genome identifier; NA, not applicable

Table S6. Characterization and validation of non-randomly distributed low-complexity regions (LCRs) in the monkeypox virus (MPXV) 353R genome sequence

Detailed information on the represented materials, along with originator and epidemiological data, used for analysis in **Figure 2**. SRA, Sequence Read Archive (SRA); ID, identification; QC, quality control.

888 **SUPPLEMENTAL FILE LEGENDS**

889 **File S1. Analysis parameters for validation of short tandem repeats (STRs) in National Center**

890 **for Biotechnology Information (NCBI) Sequence Read Archive (SRA)**

REFERENCES

Aksamentov, I., Roemer, C., Hodcroft, E.B., and Neher, R.A. (2021). Nextclade: clade assignment, mutation calling and quality control for viral genomes. J Open Source Softw *6*, 3773. 10.21105/joss.03773.

Andrews, S. (2010). FASTQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Antinori, A., Mazzotta, V., Vita, S., Carletti, F., Tacconi, D., Lapini, L.E., D'Abramo, A., Cicalini, S., Lapa, D., Pittalis, S., et al. (2022). Epidemiological, clinical and virological characteristics of four cases of monkeypox support transmission through sexual contact, Italy, May 2022. Euro Surveill *27*, 2200421. 10.2807/1560-7917.ES.2022.27.22.2200421.

Antipov, D., Korobeynikov, A., McLean, J.S., and Pevzner, P.A. (2016). hybridSPAdes: an algorithm for hybrid assembly of short and long reads. Bioinformatics *32*, 1009-1015. 10.1093/bioinformatics/btv688.

Assefa, S., Keane, T.M., Otto, T.D., Newbold, C., and Berriman, M. (2009). ABACAS: algorithm-based automatic contiguation of assembled sequences. Bioinformatics *25*, 1968-1969. 10.1093/bioinformatics/btp347.

Bandelt, H.-J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol *16*, 37-48. 10.1093/oxfordjournals.molbev.a026036.

Baroudy, B.M., and Moss, B. (1982). Sequence homologies of diverse length tandem repetitions near ends of vaccinia virus genome suggest unequal crossing over. Nucleic Acids Res *10*, 5673-5679. 10.1093/nar/10.18.5673.

Beer, E.M., and Rao, V.B. (2019). A systematic review of the epidemiology of human monkeypox outbreaks and implications for outbreak strategy. PLoS Negl Trop Dis *13*, e0007791. 10.1371/journal.pntd.0007791.

915 Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res
916 *27*, 573-580. 10.1093/nar/27.2.573.
917 Bratke, K.A., and McLysaght, A. (2008). Identification of multiple independent horizontal gene
918 transfers into poxviruses using a comparative genomics approach. BMC Evol Biol *8*, 67.
919 10.1186/1471-2148-8-67.
920 Brooks, M.A., Ali, A.N., Turner, P.C., and Moyer, R.W. (1995). A rabbitpox virus serpin gene
921 controls host range by inhibiting apoptosis in restrictive cells. J Virol *69*, 7688-7698.
922 10.1128/JVI.69.12.7688-7698.1995.
923 Bunge, E.M., Hoet, B., Chen, L., Lienert, F., Weidenthaler, H., Baer, L.R., and Steffen, R. (2022).
924 The changing epidemiology of human monkeypox—A potential threat? A systematic review. PLoS
925 Negl Trop Dis *16*, e0010141. 10.1371/journal.pntd.0010141.
926 Centers for Disease Control and Prevention (2022). 2022 Monkeypox Outbreak Global Map.
927 https://www.cdc.gov/poxvirus/monkeypox/response/2022/world-map.html.
928 Chen, N., Li, G., Liszewski, M.K., Atkinson, J.P., Jahrling, P.B., Feng, Z., Schriewer, J., Buck, C.,
929 Wang, C., Lefkowitz, E.J., et al. (2005). Virulence differences between monkeypox virus isolates
930 from West Africa and the Congo basin. Virology *340*, 46-63. 10.1016/j.virol.2005.05.030.
931 Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor.
932 Bioinformatics *34*, i884-i890. 10.1093/bioinformatics/bty560.
933 Chen, Y., He, Z., Men, Y., Dong, G., Hu, S., and Ying, X. (2022). MetaLogo: a heterogeneity-aware
934 sequence logo generator and aligner. Brief Bioinform *23*, 1-7. 10.1093/bib/bbab591.
935 Chiu, W.-L., Lin, C.-L., Yang, M.-H., Tzou, D.-L., and Chang, W. (2007). Vaccinia virus 4c (A26L)
936 protein on intracellular mature virus binds to the extracellular cellular matrix laminin. J Virol *81*,
937 2149-2157. 10.1128/JVI.02302-06.

938 Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M., and Lu, X. (2012a).
939 Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new
940 program, SnpSift. Front Genet *3*, 35. 10.3389/fgene.2012.00035.
941 Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and
942 Ruden, D.M. (2012b). A program for annotating and predicting the effects of single nucleotide
943 polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; *iso*-2; *iso*-3.
944 Fly (Austin) *6*, 80-92. 10.4161/fly.19695.
945 Cohen-Gihon, I., Israeli, O., Shifman, O., Erez, N., Melamed, S., Paran, N., Beth-Din, A., and Zvi, A.
946 (2020). Identification and whole-genome sequencing of a monkeypox virus strain isolated in Israel.
947 Microbiol Resour Announc *9*, e01524-01519. 10.1128/MRA.01524-19.
948 Colamonici, O.R., Domanski, P., Sweitzer, S.M., Larner, A., and Buller, R.M. (1995). Vaccinia virus
949 B18R gene encodes a type I interferon-binding protein that blocks interferon α transmembrane
950 signaling. J Biol Chem *270*, 15974-15978. 10.1074/jbc.270.27.15974.
951 Dainat, J., Hereñú, D., LucileSol, and pascal-git (2022). NBISweden/AGAT: AGAT-v0.8.1.
952 https://zenodo.org/record/5834795#.YzJOCzTMJyE. 10.5281/zenodo.5834795.
953 Damon, I.K. (2011). Status of human monkeypox: clinical disease, epidemiology and research.
954 Vaccine *29 Suppl 4*, D54-59. 10.1016/j.vaccine.2011.04.014.
955 Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane,
956 T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools.
957 GigaScience *10*, giab008. 10.1093/gigascience/giab008.
958 Davies, H.M., Nofal, S.D., McLaughlin, E.J., and Osborne, A.R. (2017). Repetitive sequences in
959 malaria parasite proteins. FEMS Microbiol Rev *41*, 923-940. 10.1093/femsre/fux046.
960 Davies, H.M., Thalassinos, K., and Osborne, A.R. (2016). Expansion of lysine-rich repeats in
961 *Plasmodium* proteins generates novel localization sequences that target the periphery of the host
962 erythrocyte. J Biol Chem *291*, 26188-26207. 10.1074/jbc.M116.761213.

963 Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., and Notredame, C. (2017).
964 Nextflow enables reproducible computational workflows. Nat Biotechnol *35*, 316-319.
965 10.1038/nbt.3820.
966 Elde, N.C., Child, S.J., Eickbush, M.T., Kitzman, J.O., Rogers, K.S., Shendure, J., Geballe, A.P., and
967 Malik, H.S. (2012). Poxviruses deploy genomic accordions to adapt rapidly against host antiviral
968 defenses. Cell *150*, 831-841. 10.1016/j.cell.2012.05.049.
969 Esposito, J.J., Sammons, S.A., Frace, A.M., Osborne, J.D., Olsen-Rasmussen, M., Zhang, M., Govil,
970 D., Damon, I.K., Kline, R., Laker, M., et al. (2006). Genome sequence diversity and clues to the
971 evolution of variola (smallpox) virus. Science *313*, 807-812. 10.1126/science.1125134.
972 Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results
973 for multiple tools and samples in a single report. Bioinformatics *32*, 3047-3048.
974 10.1093/bioinformatics/btw354.
975 Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso,
976 P., and Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines.
977 Nat Biotechnol *38*, 276-278. 10.1038/s41587-020-0439-x.
978 Faye, O., Pratt, C.B., Faye, M., Fall, G., Chitty, J.A., Diagne, M.M., Wiley, M.R., Yinka-Ogunleye,
979 A.F., Aruna, S., Etebu, E.N., et al. (2018). Genomic characterisation of human monkeypox virus in
980 Nigeria. Lancet Infect Dis *18*, 246. 10.1016/S1473-3099(18)30043-4.
981 Fedele, C.G., Negredo, A., Molero, F., Sánchez-Seco, M.P., and Tenorio, A. (2006). Use of internally
982 controlled real-time genome amplification for detection of variola virus and other orthopoxviruses
983 infecting humans. J Clin Microbiol *44*, 4464-4470. 10.1128/JCM.00276-06.
984 Fidalgo, M., Barrales, R.R., Ibeas, J.I., and Jimenez, J. (2006). Adaptive evolution by mutations in
985 the *FLO11* gene. Proc Natl Acad Sci U S A *103*, 11228-11233. 10.1073/pnas.0601713103.
986 Gatchel, J.R., and Zoghbi, H.Y. (2005). Diseases of unstable repeat expansion: mechanisms and
987 common principles. Nat Rev Genet *6*, 743-755. 10.1038/nrg1691.

988 Gemayel, R., Chavali, S., Pougach, K., Legendre, M., Zhu, B., Boeynaems, S., van der Zande, E.,
989 Gevaert, K., Rousseau, F., Schymkowitz, J., et al. (2015). Variable glutamine-rich repeats modulate
990 transcription factor activity. Mol Cell *59*, 615-627. 10.1016/j.molcel.2015.07.003.
991 Gigante, C.M., Korber, B., Seabolt, M.H., Wilkins, K., Davidson, W., Rao, A.K., Zhao, H., Hughes,
992 C.M., Minhaj, F., Waltenburg, M.A., et al. (2022). Multiple lineages of *Monkeypox virus* detected in
993 the United States, 2021- 2022. bioRxiv, 2022.2006.2010.495526. 10.1101/2022.06.10.495526.
994 Grant, R., Nguyen, L.L., and Breban, R. (2020). Modelling human-to-human transmission of
995 monkeypox. Bull World Health Organ *98*, 638-640. 10.2471/blt.19.242347.
996 Grubaugh, N.D., Gangavarapu, K., Quick, J., Matteson, N.L., De Jesus, J.G., Main, B.J., Tan, A.L.,
997 Paul, L.M., Brackney, D.E., Grewal, S., et al. (2019). An amplicon-based sequencing framework for
998 accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol *20*, 8.
999 10.1186/s13059-018-1618-7.
1000 Gubser, C., Hué, S., Kellam, P., and Smith, G.L. (2004). Poxvirus genomes: a phylogenetic analysis.
1001 J Gen Virol *85*, 105-117. 10.1099/vir.0.19565-0.
1002 Happi, C., Adetifa, I., Mbala, P., Njouom, R., Nakoune, E., Happi, A., Ndodo, N., Ayansola, O.,
1003 Mboowa, G., Bedford, T., et al. (2022). Urgent need for a non-discriminatory and non-stigmatizing
1004 nomenclature for monkeypox virus. PLoS Biol *20*, e3001769. 10.1371/journal.pbio.3001769.
1005 Hendrickson, R.C., Wang, C., Hatcher, E.L., and Lefkowitz, E.J. (2010). Orthopoxvirus genome
1006 evolution: the role of gene loss. Viruses *2*, 1933-1967. 10.3390/v2091933.
1007 Henry, S.P., Takanosu, M., Boyd, T.C., Mayne, P.M., Eberspaecher, H., Zhou, W., de Crombrugghe,
1008 B., Höök, M., and Mayne, R. (2001). Expression pattern and gene characterization of *Asporin*: a
1009 newly discovered member of the leucine-rich repeat protein family. J Biol Chem *276*, 12212-12221.
1010 10.1074/jbc.M011290200.

1011 Hernáez, B., Alonso-Lobo, J.M., Montanuy, I., Fischer, C., Sauer, S., Sigal, L., Sevilla, N., and
1012 Alcamí, A. (2018). A virus-encoded type I interferon decoy receptor enables evasion of host
1013 immunity through cell-surface binding. Nat Commun *9*, 5440. 10.1038/s41467-018-07772-z.
1014 Howard, A.R., Senkevich, T.G., and Moss, B. (2008). Vaccinia virus A26 and A27 proteins form a
1015 stable complex tethered to mature virions by association with the A17 transmembrane protein. J Virol
1016 *82*, 12384-12391. 10.1128/JVI.01524-08.
1017 International Committee on Taxonomy of Viruses (2022). Current ICTV Taxonomy Release. Current
1018 ICTV Taxonomy Release. https://ictv.global/taxonomy.
1019 Isidro, J., Borges, V., Pinto, M., Sobral, D., Santos, J.D., Nunes, A., Mixão, V., Ferreira, R., Santos,
1020 D., Duarte, S., et al. (2022). Phylogenomic characterization and signs of microevolution in the 2022
1021 multi-country outbreak of monkeypox virus. Nat Med *28*, 1569-1572. 10.1038/s41591-022-01907-y.
1022 Jones, T.C., Schneider, J., Mühlemann, B., Veith, T., Beheim-Schwarzbach, J., Tesch, J., Schmidt,
1023 M.L., Walper, F., Bleicker, T., Isner, C., et al. (2022). Genetic variability, including gene duplication
1024 and deletion, in early sequences from the 2022 European monkeypox outbreak. bioRxiv,
1025 2022.2007.2023.501239. 10.1101/2022.07.23.501239.
1026 Jorgensen, I., Rayamajhi, M., and Miao, E.A. (2017). Programmed cell death as a defence against
1027 infection. Nat Rev Immunol *17*, 151-164. 10.1038/nri.2016.147.
1028 Kastenmayer, R.J., Maruri-Avidal, L., Americo, J.L., Earl, P.L., Weisberg, A.S., and Moss, B.
1029 (2014). Elimination of A-type inclusion formation enhances cowpox virus replication in mice:
1030 implications for orthopoxvirus evolution. Virology *452-453*, 59-66. 10.1016/j.virol.2013.12.030.
1031 Katoh, K., Rozewicki, J., and Yamada, K.D. (2019). MAFFT online service: multiple sequence
1032 alignment, interactive sequence choice and visualization. Brief Bioinform *20*, 1160-1166.
1033 10.1093/bib/bbx108.
1034 Keasey, S., Pugh, C., Tikhonov, A., Chen, G., Schweitzer, B., Nalca, A., and Ulrich, R.G. (2010).
1035 Proteomic basis of the antibody response to monkeypox virus infection examined in cynomolgus

1036 macaques and a comparison to human smallpox vaccination. PLoS One *5*, e15547.
1037 10.1371/journal.pone.0015547.
1038 Kettle, S., Blake, N.W., Law, K.M., and Smith, G.L. (1995). Vaccinia virus serpins B13R (SPI-2)
1039 and B22R (SPI-1) encode M_r 38.5 and 40K, intracellular polypeptides that do not affect virus
1040 virulence in a murine intranasal model. Virology *206*, 136-147. 10.1016/s0042-6822(95)80028-x.
1041 Khodakevich, L., Ježek, Z., and Messinger, D. (1988). Monkeypox virus: ecology and public health
1042 significance. Bull World Health Organ *66*, 747-752.
1043 Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads
1044 using repeat graphs. Nat Biotechnol *37*, 540-546. 10.1038/s41587-019-0072-8.
1045 Kotwal, G.J., and Moss, B. (1989). Vaccinia virus encodes two proteins that are structurally related
1046 to members of the plasma serine protease inhibitor superfamily. J Virol *63*, 600-606.
1047 10.1128/JVI.63.2.600-606.1989.
1048 Kugelman, J.R., Johnston, S.C., Mulembakani, P.M., Kisalu, N., Lee, M.S., Koroleva, G., McCarthy,
1049 S.E., Gestole, M.C., Wolfe, N.D., Fair, J.N., et al. (2014). Genomic variability of monkeypox virus
1050 among humans, Democratic Republic of the Congo. Emerg Infect Dis *20*, 232-239.
1051 10.3201/eid2002.130118.
1052 Ladner, J.T., Beitzel, B., Chain, P.S.G., Davenport, M.G., Donaldson, E.F., Frieman, M., Kugelman,
1053 J.R., Kuhn, J.H., O'Rear, J., Sabeti, P.C., et al. (2014). Standards for sequencing viral genomes in the
1054 era of high-throughput sequencing. mBio *5*, e01360-01314. 10.1128/mBio.01360-14.
1055 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*,
1056 357-359. 10.1038/nmeth.1923.
1057 Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic
1058 tree display and annotation. Nucleic Acids Res *49*, W293-W296. 10.1093/nar/gkab301.
1059 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
1060 Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence

1061 Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
1062 10.1093/bioinformatics/btp352.
1063 Likos, A.M., Sammons, S.A., Olson, V.A., Frace, A.M., Li, Y., Olsen-Rasmussen, M., Davidson, W.,
1064 Galloway, R., Khristova, M.L., Reynolds, M.G., et al. (2005). A tale of two clades: monkeypox
1065 viruses. J Gen Virol *86*, 2661-2672. 10.1099/vir.0.81215-0.
1066 Liu, L., Cooper, T., Howley, P.M., and Hayball, J.D. (2014). From crescent to mature virion:
1067 vaccinia virus assembly and maturation. Viruses *6*, 3787-3808. 10.3390/v6103787.
1068 Liu, R., Mendez-Rios, J.D., Peng, C., Xiao, W., Weisberg, A.S., Wyatt, L.S., and Moss, B. (2019).
1069 SPI-1 is a missing host-range factor required for replication of the attenuated modified vaccinia
1070 Ankara (MVA) vaccine vector in human cells. PLoS Pathog *15*, e1007710.
1071 10.1371/journal.ppat.1007710.
1072 Magnus, P.v., Andersen, E.K., Petersen, K.B., and Birch-Andersen, A. (2009). A pox-like disease in
1073 cynomolgus monkeys. Acta Pathol Microbiol Scand *46*, 156-176. 10.1111/j.1699-
1074 0463.1959.tb00328.x.
1075 Mauldin, M.R., McCollum, A.M., Nakazawa, Y.J., Mandra, A., Whitehouse, E.R., Davidson, W.,
1076 Zhao, H., Gao, J., Li, Y., Doty, J., et al. (2022). Exportation of monkeypox virus from the African
1077 continent. J Infect Dis *225*, 1367-1376. 10.1093/infdis/jiaa559.
1078 McFadden, G. (2005). Poxvirus tropism. Nat Rev Microbiol *3*, 201-213. 10.1038/nrmicro1099.
1079 McLysaght, A., Baldi, P.F., and Gaut, B.S. (2003). Extensive gene gain associated with adaptive
1080 evolution of poxviruses. Proc Natl Acad Sci U S A *100*, 15655-15660. 10.1073/pnas.2136653100.
1081 Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and
1082 Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the
1083 genomic era. Mol Biol Evol *37*, 1530-1534. 10.1093/molbev/msaa015.
1084 Moss, B. (2006). Poxvirus entry and membrane fusion. Virology *344*, 48-54.
1085 10.1016/j.virol.2005.09.037.

1086 Moss, B. (2016). Membrane fusion during poxvirus entry. Semin Cell Dev Biol *60*, 89-96.
1087 10.1016/j.semcdb.2016.07.015.
1088 Moss, B., and Smith, G.L. (2021). *Poxviridae*: the viruses and their replication. In Fields virology.
1089 Volume 2: DNA viruses P.M. Howley, D.M. Knipe, B.A. Damania, and J.I. Cohen, eds. (Wolters
1090 Kluwer), pp. 573-613.
1091 Nextstrain (2022). Genomic epidemiology of monkeypox virus.
1092 https://nextstrain.org/monkeypox/hmpxv1.
1093 Ng, O.T., Lee, V., Marimuthu, K., Vasoo, S., Chan, G., Lin, R.T.P., and Leo, Y.S. (2019). A case of
1094 imported monkeypox in Singapore. Lancet Infect Dis *19*, 1166. 10.1016/S1473-3099(19)30537-7.
1095 Nuzzo, J.B., Borio, L.L., and Gostin, L.O. (2022). The WHO declaration of monkeypox as a Global
1096 Public Health Emergency. JAMA *328*, 615-617. 10.1001/jama.2022.12513.
1097 O'Tool, Á., and Rambaut, A. (2022). Initial observations about putative APOBEC3 deaminase
1098 editing driving short-term evolution of MPXV since 2017. https://virological.org/t/initial-
1099 observations-about-putative-apobec3-deaminase-editing-driving-short-term-evolution-of-mpxv-
1100 since-2017/830.
1101 Oma, Y., Kino, Y., Sasagawa, N., and Ishiura, S. (2004). Intracellular localization of homopolymeric
1102 amino acid-containing proteins expressed in mammalian cells. J Biol Chem *279*, 21217-21222.
1103 10.1074/jbc.M309887200.
1104 Otu, A., Ebenso, B., Walley, J., Barceló, J.M., and Ochu, C.L. (2022). Global human monkeypox
1105 outbreak: atypical presentation demanding urgent public health action. Lancet Microbe *3*, e554-e555.
1106 10.1016/S2666-5247(22)00153-7.
1107 Patel, H., Varona, S., Monzón, S., Espinosa-Carrasco, J., Heuer, M.L., Underwood, A., Gabernet, G.,
1108 nf-core bot, Ewels, P., MiguelJulia, et al. (2022). nf-core/viralrecon: nf-core/viralrecon v2.4.1 -
1109 Plastered Magnesium Marmoset. https://zenodo.org/record/6320980#.YzJOTzTMJyE.
1110 10.5281/zenodo.6320980.

1111 Patrono, L.V., Pléh, K., Samuni, L., Ulrich, M., Röthemeier, C., Sachse, A., Muschter, S., Nitsche,
1112 A., Couacy-Hymann, E., Boesch, C., et al. (2020). Monkeypox virus emergence in wild chimpanzees
1113 reveals distinct clinical outcomes and viral diversity. Nat Microbiol *5*, 955-965. 10.1038/s41564-020-
1114 0706-0.
1115 Phillips, C., Gettings, K.B., King, J.L., Ballard, D., Bodner, M., Borsuk, L., and Parson, W. (2018).
1116 "The devil's in the detail": release of an expanded, enhanced and dynamically revised forensic STR
1117 Sequence Guide. Forensic Sci Int Genet *34*, 162-169. 10.1016/j.fsigen.2018.02.017.
1118 Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., and Korobeynikov, A. (2020). Using SPAdes
1119 De Novo Assembler. Curr Protoc Bioinformatics *70*, e102. 10.1002/cpbi.102.
1120 Pugh, C., Brown, E.S., Quinn, X., Korman, L., Dyas, B.K., Ulrich, R.G., and Pittman, P.R. (2016).
1121 Povidone iodine ointment application to the vaccination site does not alter immunoglobulin G
1122 antibody response to smallpox vaccine. Viral Immunol *29*, 361-366. 10.1089/vim.2016.0025.
1123 Radonić, A., Metzger, S., Dabrowski, P.W., Couacy-Hymann, E., Schuenadel, L., Kurth, A., Mätz-
1124 Rensing, K., Boesch, C., Leendertz, F.H., and Nitsche, A. (2014). Fatal monkeypox in wild-living
1125 sooty mangabey, Côte d'Ivoire, 2012. Emerg Infect Dis *20*, 1009-1011. 10.3201/eid2006.13-1329.
1126 Reynolds, M.G., Yorita, K.L., Kuehnert, M.J., Davidson, W.B., Huhn, G.D., Holman, R.C., and
1127 Damon, I.K. (2006). Clinical manifestations of human monkeypox influenced by route of infection. J
1128 Infect Dis *194*, 773-780. 10.1086/505880.
1129 Rimoin, A.W., Mulembakani, P.M., Johnston, S.C., Lloyd Smith, J.O., Kisalu, N.K., Kinkela, T.L.,
1130 Blumberg, S., Thomassen, H.A., Pike, B.L., Fair, J.N., et al. (2010). Major increase in human
1131 monkeypox incidence 30 years after smallpox vaccination campaigns cease in the Democratic
1132 Republic of Congo. Proc Natl Acad Sci U S A *107*, 16262-16267. 10.1073/pnas.1005769107.
1133 Roberts, K.L., and Smith, G.L. (2008). Vaccinia virus morphogenesis and dissemination. Trends
1134 Microbiol *16*, 472-479. 10.1016/j.tim.2008.07.009.

1135 Salichs, E., Ledda, A., Mularoni, L., Albà, M.M., and de la Luna, S. (2009). Genome-wide analysis
1136 of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles
1137 compartment. PLoS Genet *5*, e1000397. 10.1371/journal.pgen.1000397.
1138 Sánchez-Seco, M.P., Hernández, L., Eiros, J.M., Negredo, A., Fedele, G., and Tenorio, A. (2006).
1139 Detection and identification of orthopoxviruses using a generic nested PCR followed by sequencing.
1140 Br J Biomed Sci *63*, 79-85. 10.1080/09674845.2006.11732725.
1141 Seeman, T. (2015). Snippy: rapid haploid variant calling and core genome alignment.
1142 https://github.com/tseemann/snippy.
1143 Senkevich, T.G., Yutin, N., Wolf, Y.I., Koonin, E.V., and Moss, B. (2021). Ancient gene capture and
1144 recent gene loss shape the evolution of orthopoxvirus-host interaction genes. mBio *12*, e0149521.
1145 10.1128/mBio.01495-21.
1146 Senkevich, T.G., Zhivkoplias, E.K., Weisberg, A.S., and Moss, B. (2020). Inactivation of genes by
1147 frameshift mutations provides rapid adaptation of an attenuated vaccinia virus. J Virol *94*, e01053-
1148 01020. 10.1128/JVI.01053-20.
1149 Shchelkunov, S.N. (2012). Orthopoxvirus genes that mediate disease virulence and host tropism. Adv
1150 Virol *2012*, 524743. 10.1155/2012/524743.
1151 Shchelkunov, S.N., Totmenin, A.V., Babkin, I.V., Safronov, P.F., Ryazankina, O.I., Petrov, N.A.,
1152 Gutorov, V.V., Uvarova, E.A., Mikheev, M.V., Sisler, J.R., et al. (2001). Human monkeypox and
1153 smallpox viruses: genomic comparison. FEBS Lett *509*, 66-70. 10.1016/s0014-5793(01)03144-1.
1154 Shchelkunov, S.N., Totmenin, A.V., Safronov, P.F., Mikheev, M.V., Gutorov, V.V., Ryazankina,
1155 O.I., Petrov, N.A., Babkin, I.V., Uvarova, E.A., Sandakhchiev, L.S., et al. (2002). Analysis of the
1156 monkeypox virus genome. Virology *297*, 172-194. 10.1006/viro.2002.1446.
1157 Shoubridge, C., Cloosterman, D., Parkinson-Lawerence, E., Brooks, D., and Gécz, J. (2007).
1158 Molecular pathology of expanded polyalanine tract mutations in the Aristaless-related homeobox
1159 gene. Genomics *90*, 59-71. 10.1016/j.ygeno.2007.03.005.

1160 Shumate, A., and Salzberg, S.L. (2021). Liftoff: accurate mapping of gene annotations.
1161 Bioinformatics *37*, 1639-1643. 10.1093/bioinformatics/btaa1016.
1162 Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and
1163 formatting protein and DNA sequences. Biotechniques *28*, 1102, 1104. 10.2144/00286ir01.
1164 The Broad Institute (2018). Picard coomand-line tools. https://broadinstitute.github.io/picard/.
1165 Thornhill, J.P., Barkati, S., Walmsley, S., Rockstroh, J., Antinori, A., Harrison, L.B., Palich, R., Nori,
1166 A., Reeves, I., Habibi, M.S., et al. (2022). Monkeypox virus infection in humans across 16 countries
1167 — April–June 2022. N Engl J Med *387*, 679-691. 10.1056/NEJMoa2207323.
1168 Ulaeto, D., Grosenbach, D., and Hruby, D.E. (1996). The vaccinia virus 4c and A-type inclusion
1169 proteins are specific markers for the intracellular mature virus particle. J Virol *70*, 3372-3377.
1170 10.1128/JVI.70.6.3372-3377.1996.
1171 Ulaeto, D.O., Dunning, J., and Carroll, M.W. (2022). Evolutionary implications of human
1172 transmission of monkeypox: the importance of sequencing multiple lesions. Lancet Microbe *3*, e639-
1173 e640. 10.1016/S2666-5247(22)00194-X.
1174 Vaughan, A., Aarons, E., Astbury, J., Balasegaram, S., Beadsworth, M., Beck, C.R., Chand, M.,
1175 O'Connor, C., Dunning, J., Ghebrehewet, S., et al. (2018). Two cases of monkeypox imported to the
1176 United Kingdom, September 2018. Euro Surveill *23*, 1800509. 10.2807/1560-
1177 7917.ES.2018.23.38.1800509.
1178 Verstrepen, K.J., Jansen, A., Lewitter, F., and Fink, G.R. (2005). Intragenic tandem repeats generate
1179 functional variability. Nat Genet *37*, 986-990. 10.1038/ng1618.
1180 Vivancos, R., Anderson, C., Blomquist, P., Balasegaram, S., Bell, A., Bishop, L., Brown, C.S.,
1181 Chow, Y., Edeghere, O., Florence, I., et al. (2022). Community transmission of monkeypox in the
1182 United Kingdom, April to May 2022. Euro Surveill *27*, 2200422. 10.2807/1560-
1183 7917.ES.2022.27.22.2200422.

1184 Vusirikala, A., Charles, H., Balasegaram, S., Macdonald, N., Kumar, D., Barker-Burnside, C.,
1185 Cumiskey, K., Dickinson, M., Watson, M., Olufon, O., et al. (2022). Epidemiology of early
1186 monkeypox virus transmission in sexual networks of gay and bisexual men, England, 2022. Emerg
1187 Infect Dis *28*, 2082-2086. 10.3201/eid2810.220960.
1188 Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q.,
1189 Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive
1190 microbial variant detection and genome assembly improvement. PLoS One *9*, e112963.
1191 10.1371/journal.pone.0112963.
1192 Wang, D., Tao, R., Li, Z., Pan, D., Wang, Z., Li, C., and Shi, Y. (2020). STRsearch: a new pipeline
1193 for targeted profiling of short tandem repeats in massively parallel sequencing data. Hereditas *157*, 8.
1194 10.1186/s41065-020-00120-6.
1195 Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview
1196 Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-
1197 1191. 10.1093/bioinformatics/btp033.
1198 Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017). Completing bacterial genome
1199 assemblies with multiplex MinION sequencing. Microb Genom *3*, e000132. 10.1099/mgen.0.000132.
1200 Wittek, R., and Moss, B. (1980). Tandem repeats within the inverted terminal repetition of vaccinia
1201 virus DNA. Cell *21*, 277-284. 10.1016/0092-8674(80)90135-x.
1202 Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2.
1203 Genome Biol *20*, 257. 10.1186/s13059-019-1891-0.
1204 World Health Organization (2022a). ICD-11. International Classification of Diseases 11th Revision.
1205 https://icd.who.int/en/.
1206 World Health Organization (2022b). Monkeypox: experts give virus variants new names.
1207 https://www.who.int/news/item/12-08-2022-monkeypox--experts-give-virus-variants-new-names.

1208 Yong, S.E.F., Ng, O.T., Ho, Z.J.M., Mak, T.M., Marimuthu, K., Vasoo, S., Yeo, T.W., Ng, Y.K.,
1209 Cui, L., Ferdous, Z., et al. (2020). Imported monkeypox, Singapore. Emerg Infect Dis *26*, 1826-1830.
1210 10.3201/eid2608.191387.
1211 Zhu, X., Jiang, L., Lu, Y., Wang, C., Zhou, S., Wang, H., and Tian, T. (2018). Association of aspartic
1212 acid repeat polymorphism in the asporin gene with osteoarthritis of knee, hip, and hand: a PRISMA-
1213 compliant meta-analysis. Medicine (Baltimore) *97*, e0200. 10.1097/MD.0000000000010200.
1214

Name	Location start[a]	Location end[b]	Repeat unit[c]	Pattern[d]	Nearest OPG[e]	Type of LCR[f]	Relative position to the OPG[g]	Distance in bp[h]	Copenhagen notation[i]	Vaccinia virus (VACV) notation[j]	Comments
LCR1	5,369	5,624	16	$[AACTAACTTATGACTT]_n$	OPG003 (ITR)	STR	Downstream	72	Cop-C19L	NA	
LCR1					OPG015 (ITR)	STR	Upstream	35	CPXV-017	NA	
LCR2	174,063	174,112	2	$[ATAT]_n$	NA	STR	Downstream	46	Cop-B16R	B14R	
LCR3	179,872	180,345	9	ATAT $[ACATTATAT]_n$	OPG208	STR	ATG Start/Promoter	21	Cop-K2L	B19R	SPI-1 apoptosis inhibition
LCR4	193,504	193,759	16	$[AAGTCATAAGTTAGTT]_n$	OPG003 (ITR)	STR	Downstream	72	Cop-C19L	NA	
LCR4					OPG015 (LITR)	STR	Upstream	35	CPXV-017	NA	
LCR5	133,895	133,918	1	$[T]_n$	MPXVgp137	homopolymer	Upstream	889	Cop-A25L	A27L	Fragmented gene area
LCR6	133,980	133,989	10	$[CAATCTTTCT]_n$	MPXVgp137	STR	Upstream	818	Cop-A25L	A27L	
LCR7	137,319	137,375	3	$[ATC]_n$	OPG153	STR	Inside ORF	NA	Cop-A28L	A26L	Attachment MVs/laminin
LCR8	147,655	147,718	5+7	$[ATATTTT]_n$ $[ATTTT]_n$ $[ATATTTT]_n$ $[ATTTT]_n$ $[ATATTTT]_n$ $[ATTTT]_n$ $[ATATTTT]_n$	OPG171	STR	Upstream	75	Cop-A42R	A42R	
LCR8					OPG170	STR	Upstream	70	Cop-A41L	A41L	
LCR9	151,350	151,417	9	$[TATGAAG]_n$ $[GATATGAT]_n$ $[GATATGATG]_n$ $[GATATGAT]_n$	OPG176	STR	Upstream	12	Cop-A46R	A47R	
LCR10	197,830	197,842	1	$[T]_n$	OPG001 (ITR)	homopolymer	Downstream	225	NA	NA	
LCR11	1,286	1,298	1	$[T]_n$	OPG001 (ITR)	homopolymer	Downstream	225	NA	NA	
LCR12	29,326	29,364	1	$[A]_n$	OPG044	homopolymer	Inside ORF	NA	Cop-K7R	B15R	C-terminal position
LCR13	76,896	76,904	1	$[T]_n$	OPG097	homopolymer	Inside ORF	NA	Cop-L3L/L4R	L3L/L4R	
LCR14	81,658	81,666	1	$[T]_n$	OPG104	homopolymer	Inside ORF	NA	Cop-J5L	L5L	Essential for viral replication
LCR15	140,911	140,977	9	$[ATAACAATT]_n$ $[ATAATTGTT]_n$ $[ATAATAATT]_n$ $[ATAATTGTT]_n$	OPG159	STR	Inside ORF	NA	Cop-A31L	A33L	PKR inhibitor candidate? / C-terminal position
LCR16	153,457	153,465	1	$[A]_n$	OPG180	homopolymer	Upstream	15	Cop-A50R	A50R	
LCR17	163,979	164,003	4	$[TAAC]_n$	OPG188	STR	Downstream	82	Cop-B2R	B4R	
LCR18	166,865	166,920	7	$[AATAATT]_n$	OPG190	STR	Downstream	15	Cop-B5R	B6R	
LCR19	170,508	170,563	6	$[GATACA]_n$	OPG197	STR	Inside ORF	NA	Cop-B11R	B11R	Hypothetical protein
LCR20	172,868	172,876	1	$[T]_n$	OPG199	homopolymer	Downstream	56	Cop-K2L	SPI-2/B12R	
LCR21	175,299	175,357	6	$[GATGAA]_n$	OPG204	STR	ATG Start/Promoter	NA	Cop-B19R	B16R	Alternative ATG repeat start

a Nucleotide base coordinate in reference HQG (Genbank# OX044336)

b Nucleotide base coordinate in reference HQG (Genbank# OX044336)

c Number of repeat units in the HQG (Genbank# OX044336)

d Description of the pattern of the LCR in representative MPXV, in which $_n$ is the number of repeats for this particular genome

e Identification according to Senkevich et al. of nearest identified gene. New Notation

f Type of LCR: short tandem repeats or homopolymer

g Position of the LCR to the neareast gene

h Distance of the LCR to the nearest gene

i Notation of the gene in the VACV Copenhagen strain

j Notation of the gene in the VACV Western Reserve strain

Name	Repeat Unit	Pattern HQ	Number of	Nearest Gene	Variation	Type of	Entropy	Resolved	Nanopore	MiSeq	NovaSeq	#	#
LCR4	16	TAGTCATAAGTTAGTT [AAGTCATAAGTTAGTT]$_{15}$	16	OPG003 (ITR)	NR	Length	NA	No[&]	Yes	No	No	NA	NA
LCR3	9	ATAT [ACATTATAT]$_{52}$	52	OPG208	Yes	Length	NA	Yes	Yes	No	No	NA	NA
LCR1	16	[AACTAACTTATGACTT]$_{15}$ AACTAACTTATGACTA	16	OPG003 (ITR)	NR	Length	NA	No[&]	Yes	No	No	NA	NA
LCR2	2	[AT]$_{25}$	25	NA	Yes	Length	1.66	No	Yes	Yes	Yes	768	90
LCR5	1	[T]$_{24}$	24	OPG152	Yes	Length	1.535	Yes	No	No	Yes	NA	112
LCR10	1	[T]$_{13}$	13	OPG001 (ITR)	Yes	Length	0.63	No[&]	Yes	No	No	6,561	11,945
LCR11	1	[T]$_{13}$	13	OPG001 (ITR)	Yes	Length	0.627	No[&]	Yes	Yes	Yes	6,448	11,589
LCR21	6	[GATGAA]$_4$ GATGA	4.5	OPG204	Yes	Mutation	0.207	Yes	Yes	Yes	Yes	6,578	6,661
LCR7	3	[ATC]$_{14}$ TATGAT [ATC]$_3$	19	OPG153	Yes	Length	0.181	Yes	Yes	Yes	Yes	4,541	6,607
LCR9	9	[TATGAAG]$_1$ [GATATGAT]$_1$ [GATATGATG]$_5$ [GATATGAT]$_1$	8	OPG176	No	NA	0	Yes	Yes	Yes	Yes	5,208	5,737
LCR8	5+7	[ATATTTT]$_1$ [ATTTT]$_1$ [ATATTTT]$_3$ [ATTTT]$_1$ [ATATTTT]$_2$ [ATTTT]$_1$ [ATATTTT]$_1$	10	OPG171	No	NA	0	Yes	Yes	Yes	Yes	6,581	6,790
LCR6	10	[CAATCTTTCT]$_1$	1	OPG152	Yes*	NA	0	No*	Yes	Yes	Yes	4,884	12,930
LCR20	1	[T]$_9$	9	OPG199	No	NA	0	Yes	Yes	Yes	Yes	10,106	13,315
LCR19	6	GATTCA [GATACA]$_8$ GAT	9.3	OPG197	No	NA	0	Yes	Yes	Yes	yes	4,119	4,685
LCR18	7	[AATAATT]$_3$ AATAA	3	OPG190	No	NA	0	Yes	Yes	Yes	Yes	9,755	11,838
LCR17	4	[TAAC]$_6$ T	6.1	OPG188	No	NA	0	Yes	Yes	Yes	Yes	7,388	9,474
LCR16	1	[A]$_9$	9	OPG180	No	NA	0	Yes	Yes	Yes	Yes	10,340	16,044
LCR15	9	[ATAACAATT]$_4$ [ATAATTGTT]$_1$ [ATAATAATT]$_1$ [ATAATTGTT]$_1$	7	OPG159	No	NA	0	Yes	Yes	Yes	Yes	7,067	6,569
LCR14	1	[T]$_9$	9	OPG104	No	NA	0	Yes	Yes	Yes	Yes	7,819	12,521
LCR13	1	[T]$_9$	9	OPG097/098	No	NA	0	Yes	Yes	Yes	Yes	7,480	12,126
LCR12	1	[A]$_9$	9	OPG044	No	NA	0	Yes	Yes	Yes	Yes	9,789	13,592

* LCR6 is a 10-bp repeat that was reported early in the outbreak as an insertion (https://virological.org/t/first-german-genome-sequence-of-monkeypox-virus-associated-to-multi-country-outbreak-in-may-2022/812). In our dataset, we have not seen any variation in this area.

& LCR pairs 1/4 and 10/11 are located in ITRs. Given that no read covering this area reached a unique are outside of the ITRs, we cannot technically state that we solved the repeat. Nonethless, the ITRs should be identical based on the know poxvirid replication mode.

	MPXV-M5312_HM12_Rivers	MPXV_USA_2022_MA001	353R
Genome length	197209	197205	198547
SNPs*	NA	67	69
Indels*	NA	10del 7ins	11del 6 ins
Homopolymeric sites**	408	405	399
Unique SNPs	NA	0	2
LCR characterization			
LCR pair 1/4	8	8	16
LCR2	22	24	25
LCR3	18	16	52
LCR5	25	28	24
LCR6	2	1	1
LCR7	19	17.6	17.6
LCR8	10	10	10
LCR9	8	6	6
LCR pair 10/11	17	14	13
LCR12	9	9	9
LCR13	9	9	9
LCR14	9	9	9
LCR15	7	7	7
LCR16	9	9	9
LCR17	6.1	6.1	6.1
LCR18	3.5	3.5	3.5
LCR19	9.3	9.3	9.3
LCR20	9	9	9
LCR21	4.5	4.5	4.5
*SNPs and indels vs MPXV-M5312_HM12_Rivers			
** Homopolymers >8 nt in length			

Figure 1

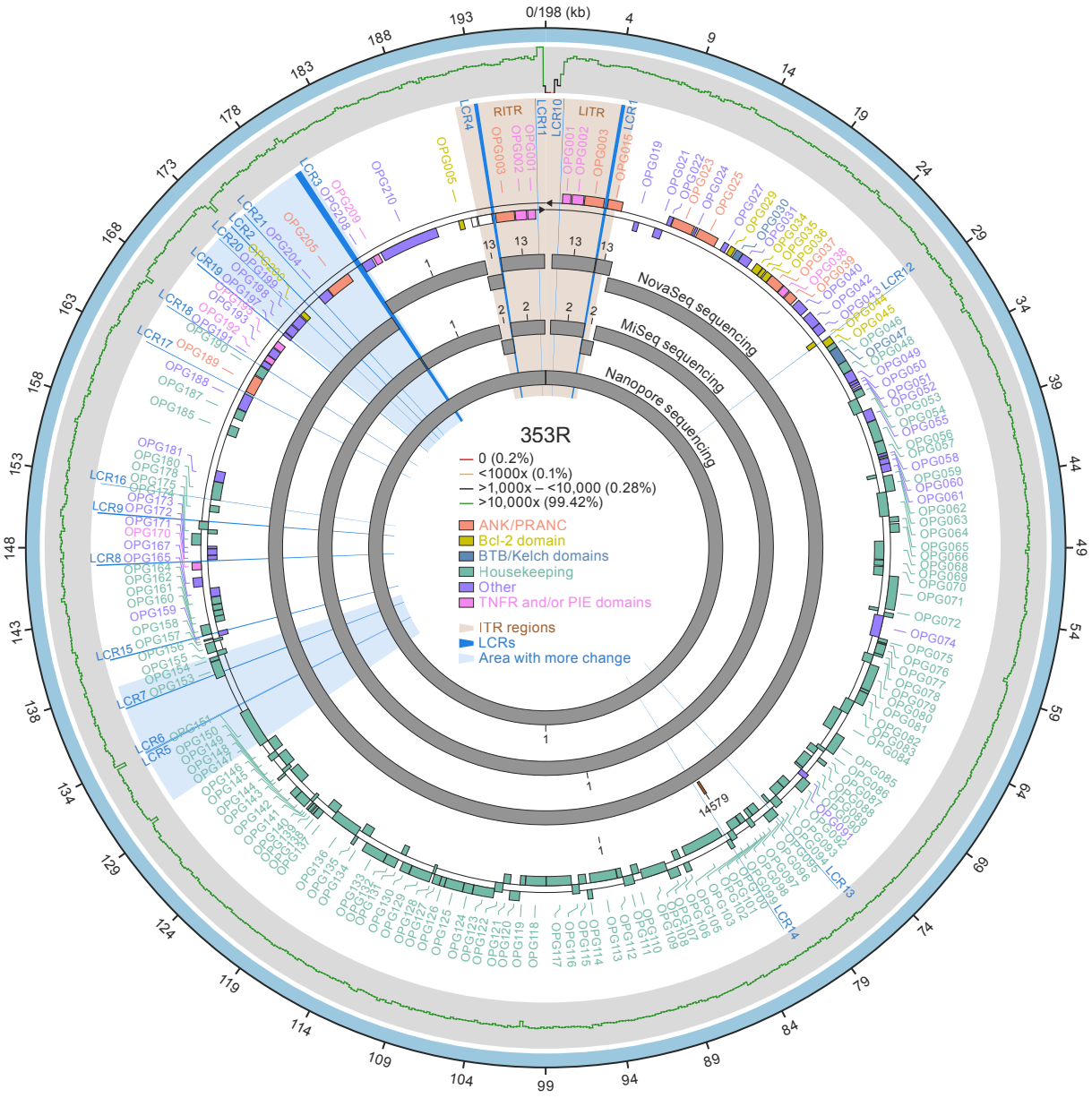

0/198 (kb)
RITR
LITR
NovaSeq sequencing
MiSeq sequencing
Nanopore sequencing
353R
0 (0.2%)
<1000x (0.1%)
>1,000x – <10,000 (0.28%)
>10,000x (99.42%)
ANK/PRANC
Bcl-2 domain
BTB/Kelch domains
Housekeeping
Other
TNFR and/or PIE domains
ITR regions
LCRs
Area with more change
14579

Figure 2

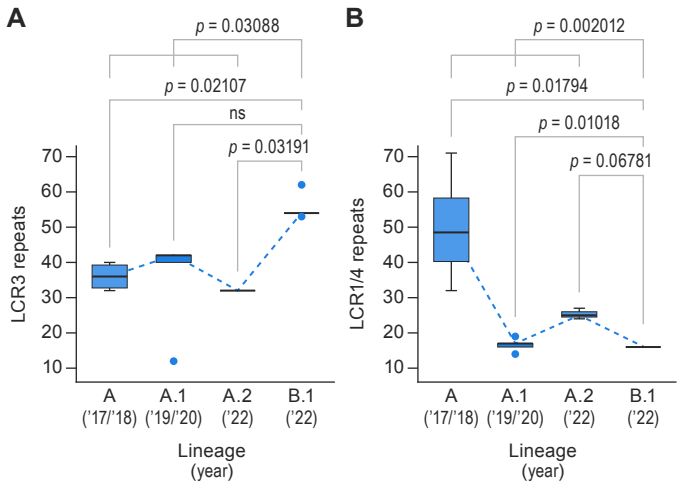

A
p = 0.03088
p = 0.02107
ns
p = 0.03191
LCR3 repeats
70
60
50
40
30
20
10
A ('17/'18)
A.1 ('19/'20)
A.2 ('22)
B.1 ('22)
Lineage (year)
B
p = 0.002012
p = 0.01794
p = 0.01018
p = 0.06781
LCR1/4 repeats
70
60
50
40
30
20
10
A ('17/'18)
A.1 ('19/'20)
A.2 ('22)
B.1 ('22)
Lineage (year)

Figure 3

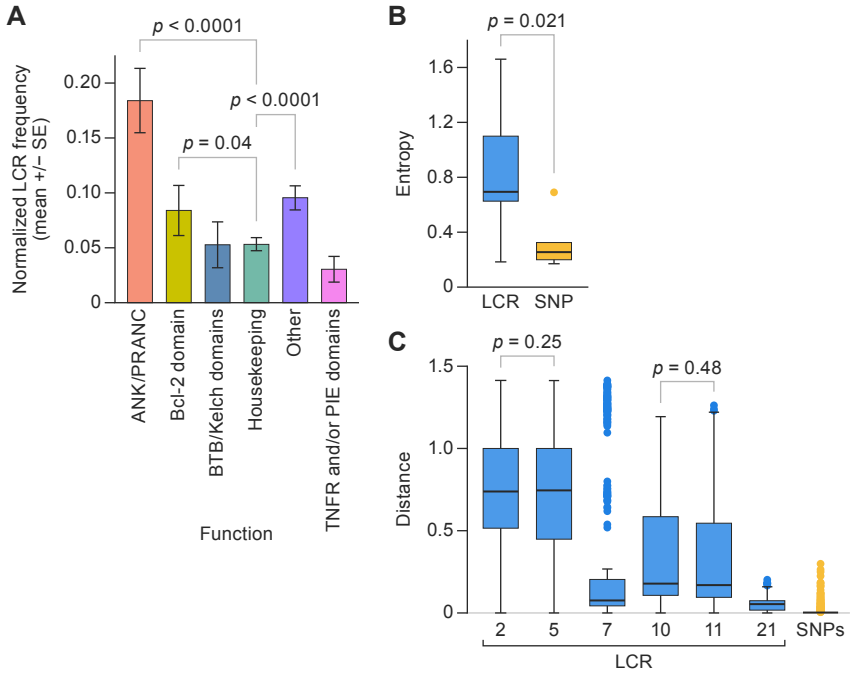

A
p < 0.0001
p < 0.0001
p = 0.04
Normalized LCR frequency (mean +/− SE)
0.20
0.15
0.10
0.05
0
ANK/PRANC
Bcl-2 domain
BTB/Kelch domains
Housekeeping
Other
TNFR and/or PIE domains
Function
B
p = 0.021
Entropy
1.6
1.2
0.8
0.4
0
LCR
SNP
C
p = 0.25
p = 0.48
Distance
1.5
1.0
0.5
0
2
5
7
10
11
21
SNPs
LCR

Figure 4

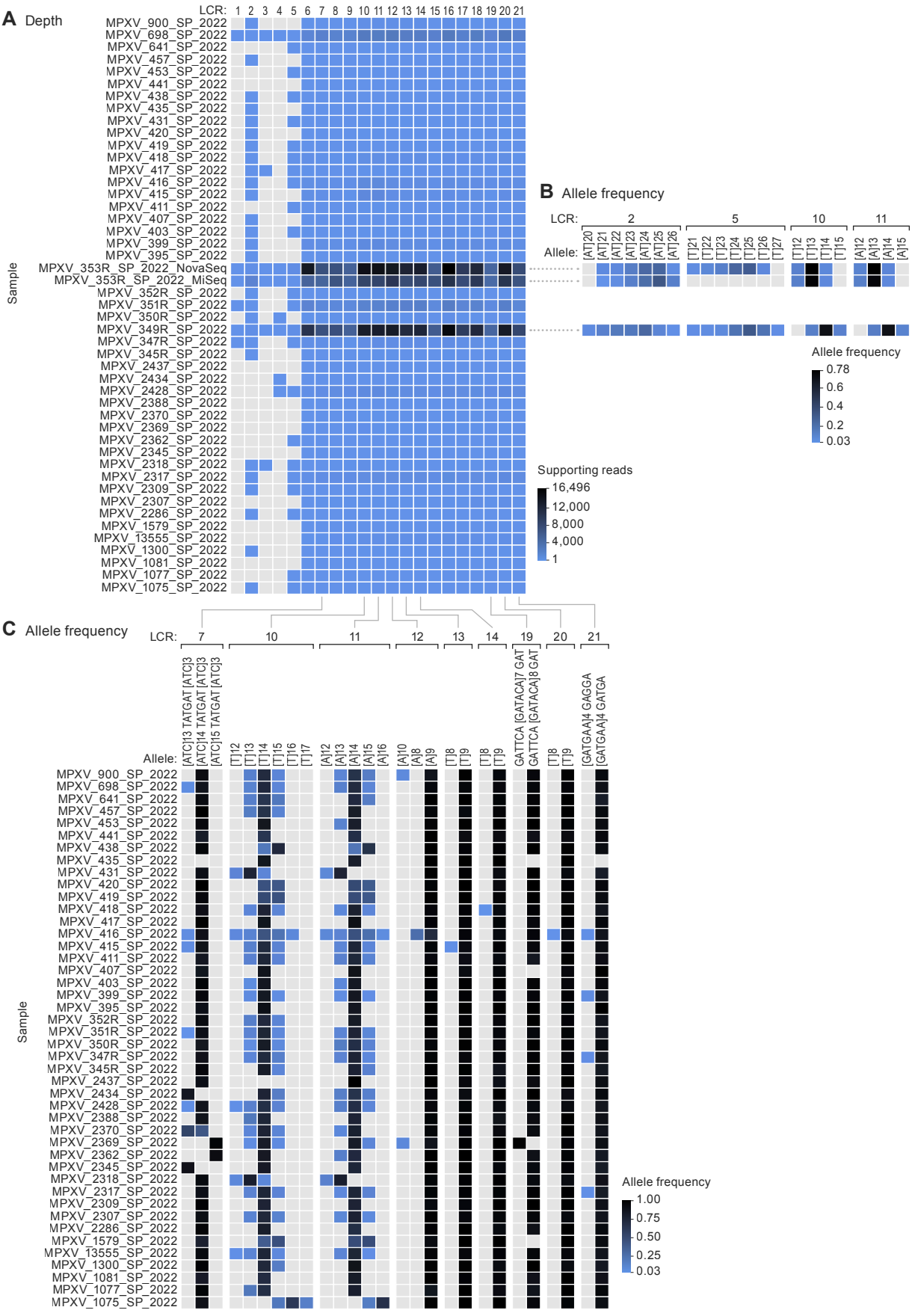
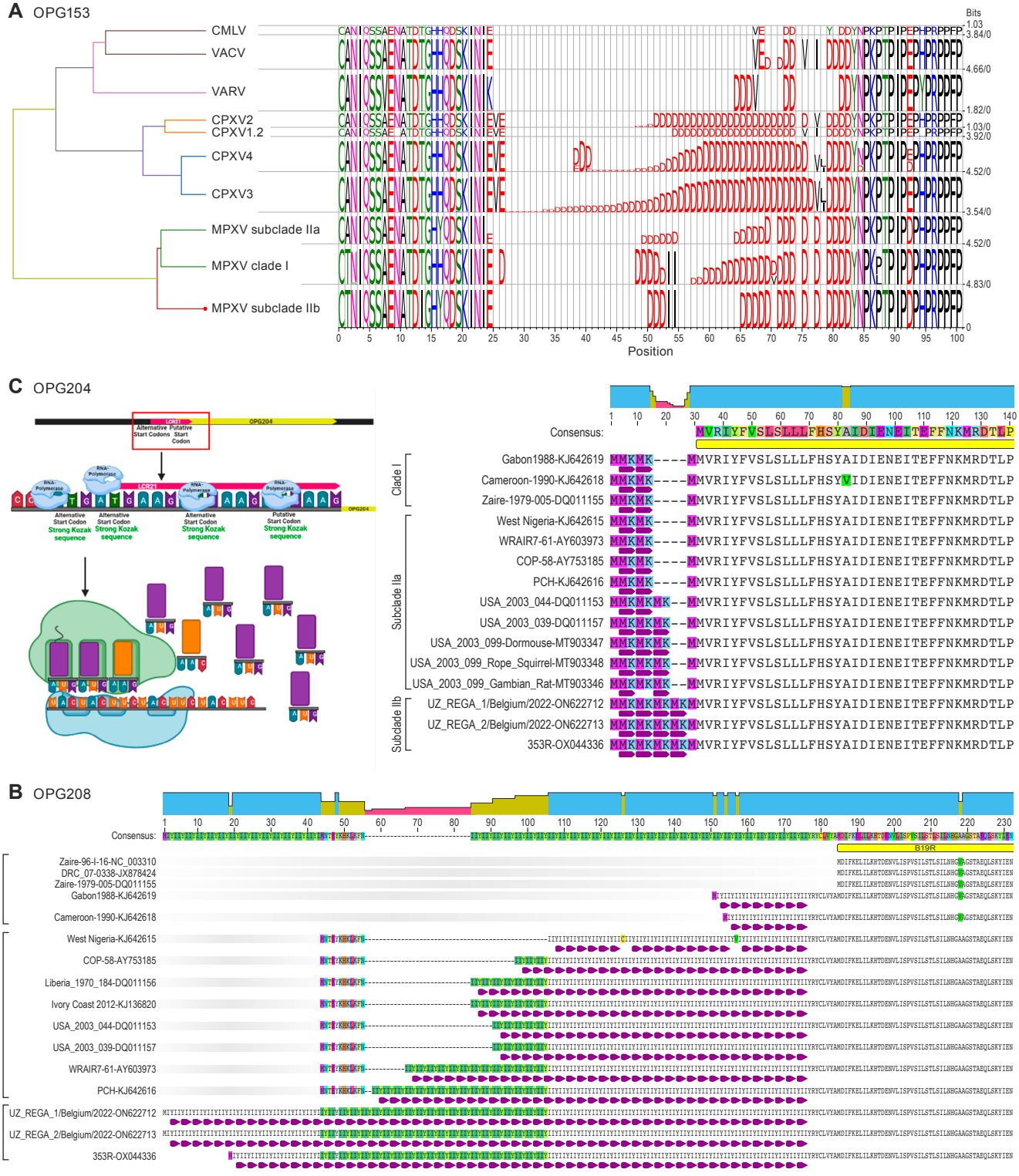

A Depth
LCR: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Sample
MPXV_900_SP_2022
MPXV_698_SP_2022
MPXV_641_SP_2022
MPXV_457_SP_2022
MPXV_453_SP_2022
MPXV_441_SP_2022
MPXV_438_SP_2022
MPXV_435_SP_2022
MPXV_431_SP_2022
MPXV_420_SP_2022
MPXV_419_SP_2022
MPXV_418_SP_2022
MPXV_417_SP_2022
MPXV_416_SP_2022
MPXV_415_SP_2022
MPXV_411_SP_2022
MPXV_407_SP_2022
MPXV_403_SP_2022
MPXV_399_SP_2022
MPXV_395_SP_2022
MPXV_353R_SP_2022_NovaSeq
MPXV_353R_SP_2022_MiSeq
MPXV_352R_SP_2022
MPXV_351R_SP_2022
MPXV_350R_SP_2022
MPXV_349R_SP_2022
MPXV_347R_SP_2022
MPXV_345R_SP_2022
MPXV_2437_SP_2022
MPXV_2434_SP_2022
MPXV_2428_SP_2022
MPXV_2388_SP_2022
MPXV_2370_SP_2022
MPXV_2369_SP_2022
MPXV_2362_SP_2022
MPXV_2345_SP_2022
MPXV_2318_SP_2022
MPXV_2317_SP_2022
MPXV_2309_SP_2022
MPXV_2307_SP_2022
MPXV_2286_SP_2022
MPXV_1579_SP_2022
MPXV_13555_SP_2022
MPXV_1300_SP_2022
MPXV_1081_SP_2022
MPXV_1077_SP_2022
MPXV_1075_SP_2022
Supporting reads
16,496
12,000
8,000
4,000
1
B Allele frequency
LCR: 2 5 10 11
Allele: [AT]20 [AT]21 [AT]22 [AT]23 [AT]24 [AT]25 [AT]26 [T]21 [T]22 [T]23 [T]24 [T]25 [T]26 [T]27 [T]12 [T]13 [T]14 [T]15 [A]12 [A]13 [A]14 [A]15
Allele frequency
0.78
0.6
0.4
0.2
0.03
C Allele frequency
LCR: 7 10 11 12 13 14 19 20 21
Allele: [ATC]13 TATGAT [ATC]3 [ATC]14 TATGAT [ATC]3 [ATC]15 TATGAT [ATC]3 [T]12 [T]13 [T]14 [T]15 [T]16 [T]17 [A]12 [A]13 [A]14 [A]15 [A]16 [A]10 [A]8 [A]9 [T]8 [T]9 [T]8 [T]9 GATTCA [GATACA]7 GAT GATTCA [GATACA]8 GAT [T]8 [T]9 [GATGAA]4 GAGGA [GATGAA]4 GATGA
Sample
Allele frequency
1.00
0.75
0.50
0.25
0.03

Figure 5


A OPG153
CMLV
VACV
VARV
CPXV2
CPXV1.2
CPXV4
CPXV3
MPXV subclade IIa
MPXV clade I
MPXV subclade IIb
Bits
Position
C OPG204
Consensus:
Clade I
Subclade IIa
Subclade IIb
Gabon1988-KJ642619
Cameroon-1990-KJ642618
Zaire-1979-005-DQ011155
West Nigeria-KJ642615
WRAIR7-61-AY603973
COP-58-AY753185
PCH-KJ642616
USA_2003_044-DQ011153
USA_2003_039-DQ011157
USA_2003_099-Dormouse-MT903347
USA_2003_099_Rope_Squirrel-MT903348
USA_2003_099_Gambian_Rat-MT903346
UZ_REGA_1/Belgium/2022-ON622712
UZ_REGA_2/Belgium/2022-ON622713
353R-OX044336
B OPG208
Consensus:
B19R
Clade I
Subclade IIa
Subclade IIb
Zaire-96-I-16-NC_003310
DRC_07-0338-JX878424
Zaire-1979-005-DQ011155
Gabon1988-KJ642619
Cameroon-1990-KJ642618
West Nigeria-KJ642615
COP-58-AY753185
Liberia_1970_184-DQ011156
Ivory Coast 2012-KJ136820
USA_2003_044-DQ011153
USA_2003_039-DQ011157
WRAIR7-61-AY603973
PCH-KJ642616
UZ_REGA_1/Belgium/2022-ON622712
UZ_REGA_2/Belgium/2022-ON622713
353R-OX044336