а								
Consensus:	174095 TACATGTGTTTTAGA 5' flanking region	174105 <mark>ATATA</mark>	174115 TATATA'	174125 TATATATATATA TATATATATATA	174135 TATATATATA	174145 ATATATATATA	174155 <b>ATATATGGG</b> 3'	174165 CAAAACATATAA flanking region
353R_High_Quality_Genome/1-198547	TACATGTGTTTTAGA	<mark>ATA</mark> TA	Τ <mark>ΑΤΑΤΑΤΑ</mark>	ΤΑΤΑΤΑΤΑΤΑΤΑ	ТАТАТАТАТА	TATATATATA	ATATATGGG	СААААСАТАТАА
353R_consensus_High_Quality_Cenome/1-198547 ON563414.2_MPXV_USA_2022_MA001/1-197124 353R_ON563414.2_MPXV_USA_2022_MA001/1-197133 ON568298.1_MPXV_BY_IMB25241/1-197378 353R_ON568298.1_MPXV_BY_IMB25241/1-197377 NC_063383.1_MPXV_NG_2018/1-197209 353R_NC_063383.1_MPXV_NG_2018/1-197199	TACATGTGTTTTAGA TACATGTGTTTTAGA TACATGTGTTTTAGA TACATGTGTTTTAGA TACATGTGTTTTAGA TACATGTGTTTTAGA TACATGTGTTTTAGA	ATATA ATA ATA ATA ATA	TATATATA TATATATA TATATATA TATATATA TATATATA TATATATA ATATATA ATATATA	TATATATATATAT TATATATATATAT TATATATATAT	ATATATATATA ATATATATATA ATATATATATA ATATATATATA ATATATATATA ATATATATATA ATATATATATATA ATATATATATATATA	ATATATATAT ATATATATATAT ATATATATATA ATATATATATA ATATATATATA ATATATATATA ATATATATATA ATATATATATA	ATATATGGG 'ATATATGGG 'ATATATGGG' 'ATATATGGG 'ATATATGGG 'ATATATGGG 'ATATATGGG	СААААСАТАТАА СААААСАТАТАА СААААСАТАТАА СААААСАТАТАА СААААСАТАТАА СААААСАТАТАА СААААСАТАТАА
b								

	137346	137356	137366	137376	137386	137396	137406	137416
Consensus:	AGGTTTTGGATTATA	ATCATCATCA	TCATCATCAT	CATCATCATC.	A <mark>TCATCAT</mark> CA	T <mark>CTATGATAT</mark>	CATCATCTT	<b>CGATATTTATTT</b>
	5' flanking region							3' flanking region
353R_High_Quality/1-198547	AGGTTTTTGGATTATA	ATCATCATCA	TCATCATCAT	CATCATCATC	ATCATCATCA	ICTATGATAT	CATCATCTT	CGATATTTATTT
NC_063383.1/1-197209	AGGTTTTTGGATTATA	TCATCATCA	TCATCATCAT	CATCATCATC	ATCATCATCA	ICTATGATAT	CATCATCTT	CGATATTTATTT
353R_nova_assembly/1-190773	AGGTTTTGGATTATA	ATCATCATCA	TCATCATCAT	CATCATCATC	ATCATCATCA'	ICTATGATAT	CATCATCTT	CGATATTTATTT
353R_miseq_assembly/1-192918	AGGTTTTTGGATTATA	ATCATCATCA	TCATCATCAT	CATCATCATC	ATCATCATCA'	ICTATGATAT	CATCATCTT	CGATATTTATTT
353R_nanopore_flye_pilon/1-198541	AGGTTTTTGGATTATA	ATCATCATCA	TCATCATCAT	CATCATCATC	ATCATCATCA'	ICTATGATAT	CATCATCTT	CGATATTTATTT

Figure S1. Examples for read mapping artefacts and correction in monkeypox virus (MPXV) genome low-complexity regions (LCRs) a LCR2 alignment highlighting differences compared with various consensus sequences. b LCR7 alignment demonstrating identical results obtained using three sequencing platforms compared to the subclade IIb lineage A MPXV reference isolate MPXV-M5312\_HM12\_Rivers sequence.



#### Figure S2. Phylogenetic analysis of monkeypox virus (MPXV)

a Phylogenetic maximum-likelihood (ML) tree showing monkeypox virus (MPXV) subclade IIb single-nucleotide polymorphism (SNP) clustering. Bootstrap supports >60 are indicated by labels with their number of supports. b Haplotype network showing SNP differences among samples included in the phylogenetic tree. Details on groups can be found in Supplementary Data 4.



#### Figure S2. Phylogenetic analysis of monkeypox virus (MPXV)

a Phylogenetic maximum-likelihood (ML) tree showing monkeypox virus (MPXV) subclade IIb single-nucleotide polymorphism (SNP) clustering. Bootstrap supports >60 are indicated by labels with their number of supports. b Haplotype network showing SNP differences among samples included in the phylogenetic tree. Details on groups can be found in Supplementary Data 4.



Figure S3. Conservation and variation in proteins encoded by orthologous poxvirus gene (OPG) 208 a MetaLogo visualization of conserved and varying amino-acid residues in OPG-encoded proteins among monkeypox virus (MPXV) clade I, subclade IIa, and subclade IIb, with homologous and nonhomologous sites highlighted; b Entropy heatmap; c Entropy analysis by site



Figure S4. Conservation and variation in proteins encoded by orthologous poxvirus gene (OPG) 153 a Entropy heatmap. CMLV, camelpox virus; VARV, variola virus; VACV, vaccinia virus; CPXV, cowpox virus; MPXV, monkeypox virus. b Clustering result of sequence logo groups. MPXV, monkeypox virus; CPXV, cowpox virus; VACV, vaccinia virus; VARV, variola virus; CMLV, camelpox virus.



# Figure S5. Low-complexity regions (LCRs) in coding areas evolve mostly as changes of nucleotide triplets (codons), a pattern significantly different from LCRs in non-coding areas

a Proportion of variants in LCRs (inside versus outside coding areas) that do not cause frameshifts. b Proportion of LCRs (inside versus outside coding areas) that include stop codons.



#### Figure S6. Characterization of monkeypox virus (MPXV) promoters

a Logo plot displaying the frequencies of nucleotides that form the promoter sequences from the top 10 genes most expressed early (above) and late during infection (below). b Distribution (top 10 early genes) of distances of the 3' of the promoter to the initial codon (note that the promoter of late genes includes the first alternative start codon [ATG]). c 5' end of OPG204, including its promoter aligned to the top 10 early genes (black box) and the LCR21 repeat region, that contains its ATG (red box). d 5' end of OPG208, including its promoter aligned to the top 10 late genes (black box) and the ATG followed by the first two repeats of LCR3 (red box).

## **SUPPLEMENTARY NOTE 1**

Supplementary Note 1. Analysis parameters for validation of short tandem repeats (STRs) in National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA)

#### A. Short tandem repeats curation

Short Tandem Repeats areas were manually inspected in assemblies from sample 353R. Alignments show for all identified LCRs:

- 1. NC\_063363.1 reference genome
- 2. 353\_R sample de novo assembly from MiSeq reads 2x 300
- 3. 353\_R sample de novo assembly from NovaSeq 2x150 reads
- 4. 353\_R sample de novo assembly from Nanopore reads.
- 5. High quality genome

### LCR1:



## LCR2:



#### LCR3:

. NC\_063383.1/1-197209 NC\_063383.1/1-197209 353\_R\_nova\_assembly/1-190773 353\_R\_miseq\_assembly/1-192918 353\_R\_nanopore\_flye\_pilon/1-198541 ata fa catta fa ta catta fa fa catta fa a ta fa catta fa fa catta fa c 353\_R\_High\_Quality/1-198556 Cons TA TA TACA T TA TA CA T TA TA CA T TA TA TACA T TA TA TACA T TA TA TACA T TA TA TACA T TA TACA T TA TACA T TA TA ATTATTTATTATGA Occupancy ..... .... . . ! . . . . acat . . . . . . . . . TACA . . . . . . . . . a la la catta la catta la catta la catta la catta catta catta catta la catta catta catta catta catta la catta cat 50256 100266 100256 100256 100276 100260 100276 100266 100366 100366 100366 100366 100366 100366 100366 100366 atacat atacat atacat a ta ta catta ta ta catta ta catta ta catta ta catta ta ta catta ta catta ta catta ta catta ta catta ta catta a ta ta catta ta ta catta ta catta ta catta ta catta ta catta catta catta catta catta catta catta catta catta c 

_	18	03	46						1	80	03	56							180	03	66						18	30	37	6					1	80	38	6					1	180	039	96					1	80	40	6					18	80
а	t	a	a	1	a	t	t	а	t a	ı t	a	1	a	t	t	а	t	а	tε	1 0	a	t	t	a t	a	t	a	С	а	t t	a	i t	а	t	a	a	t	t	a	t a	ı t	а	c	a t	t	а	t a	a t	а	t	Сļ	j t		а	t t	g	t		g	g
a	t	a	a	<b>1</b> 0	a	t	t	а	t a	ı t	a	L (	а	t	t	а	t	а	t a	1 0	a	t.	t	a t	a	t	а	С	а	t t	a	ı t	а	t	a	a	t	t	a	t a	ı t	а	с	a t	t	а	t į	a t	а		C (	j t			t t		t			g
a	t	a	ta	1	a	t	t	а	t a	ı t	a	L (	а	t	t	а	t	а	t a	1 0	a	t.	t į	a t	a	t	а	С	а	t t	a	ı,t	a	t	a	a	t	t	a	t a	ı,t	а	С	a t	t	а	t į	a)t	a		C (	1 t			t t		t			ç
a	t i	a	a	1	a	t	t	а	t a	1 1	a	1	a	t	t	а	t	а	tε	1 0	a	t.	t	a t	a	t	а	С	а	tt	8	1 t	а	t	a	a	t	t	a	t a	۱t	а	c	a t	t	а	t i	1	а		Сļ	j t			t t		t			g
а	t i	a	a	1 0	a	t	t	a	t e	1	a	1 0	a	t	t	а	t	а	tε	1 0	a	t	t	a t	a	t	a	С	а	t t	8	i t	а	t	a	a	t.	t	a	t a	i t	а	C	a t	t	а	t i	a t	a	t	C (	j t	t	a	t t	g	t	t t	g	g
A	Т	Α.	ΤA	١C	: A	Т	Т	A	T/	1	ΤÆ	١C	A	١T	Т	A	Т	A	T/	١C	A	Т	T.	A 1	r A	۱T	٢A	С	A	T.	F/	١	A	T.	AC	C A	ιT	Т	A'	T/	1	A	С	A 1	ГΤ	A	T/	٩.1	٢A	Т	С	G 1	Т	A	TΊ	G	T	TΊ	٢G	0

# <u>LCR4:</u>

	193783	193793	193803	193813	193823	193833	193843	193853	193863	193873	193883	193893
NC_063383.11.197209 353_R_nova_assembly/1-190773 353_R_miseq_assembly/1-192819 353_R_nanopore_flye_pilon/1-198541 353_R_High_Quality/1-198556	cttgaaatgagttaa cttgaaatgagttaa cttgaaatgagttaa cttgaaatgagttaa cttgaaatgagttaa	agt <mark>c</mark> ataagtt agtcataagtt agtcataagtt	agttaagtca agttaagtca agttaagtca	ataagttagtt ataagttagtt ataagttagtt	aagt <mark>c</mark> ataa aagtcataa aagtcataa	g t t a g t t a a g t g t t a g t t a a g t g t t a g t t a a g t	cataagttagt cataagttagt cataagttagt	taagtcataa taagtcataa taagtcataa	gttagttaa gttagttaa gttagttaa	g t cataagtta g t cataagtta g t cataagtta	agttaagtca agttaagtca agttaagtca	taa <mark>g</mark> tt taagtt taagtt
Consensus												
	C T TG A A A TG AG T T A A T	TAG TC A TA AG T I	AG T TAAG TC /	A TAAG T TAG T T	AAGTCATAA	G T T A G T T A A G T (	CATAAGTTAGI	TAAGTCATAA	G T T A G T T A A (	G TC A TA AG T T	AG T T A A G T C A	TAAGTT
Occupancy												
193903 193913 E official offic	193923 193933 Diticit and it in ditina dit diticit and it in ditina dit diticit and it in ditica dit	193943 Cataa gttagtta Cataa gttagtta Cataa gttagtta Cataa gttagtta	193953 a gt cata a gt ta a gt cata a gt ta a gt cata a gt ta a gt cata a gt ta	193963 19 Official official a official a	13973 193 g ttag ttaag g ttag ttaag g ttag ttaag g ttag tta	cataagttagtta cataagttagtta cataagttagtta cataagttagtta cataagttagtta cataagttagtta	194003 a a g t cata a g t t a a g t cata a g t t a g t cata a g t t a g t cata a g t t a g t cata a g t t a	194013 Igitaagicata Igitaagicata Igitaagicata Igitaagicata Igitaagicata	194023 ag t t ag t t a ag ag t t ag t t a ag ag t t ag t t ag ag t t ag t t ag ag t t ag t t ag	194033 t cataag t tag t t cataag t tag t	94043 194 tatagtctaa tatagtctaa tatagtctaa tatagtctaa tatagtctaa tatagtctaa	053 cactt cactt cactt cactt cactt
AG T T A AG T C A T A AG T T AG T T A A	IG TC A TAAG T TAG T TAAG T	CATAAGTTAGTTA	AGTCATAAGTTA	G T T A A G T C A T A A	G T T A G T T A A G T	ICA TAAG T TAG T TA	AAGTCATAAGTT	AG T T A AG T C A T A	AG T T AG T T A AG	TC A TA AG T TAG T	TTA TAG TC TAA	CACTT

<u>LCR5:</u>		100001	400004		400054
	133911	133921	133931	133941	133951
NC_063383.1/1-197211 353_R_nova_assembly/1-190773 353_R_miseq_assembly/1-192819 353_R_nanopore_flye_pilon/1-198541 353_R_High_Quality/1-198556	ctaggacggcgatt ctaggacggcgatt ctaggacggcgatt ctaggacggcgatt ctaggacggcgatt ctaggacggcgatt	<mark>c</mark> ttttttttt cttttttttt c-ttttttttt ct ct ct		tttttcgaat tttttcgaat tttttcgaat tttttcgaat tttttcgaat tttttcgaat	t ca cggggt a t ca cggggt a
Consensus	CTAGGACGGCGATT	стттттттт		TTTTCGAAI	TCACGGGGTA
Occupancy					

# <u>LCR6:</u>



A G G T T T T G G A T T A T C



#### LCR8:



#### LCR10:



### LCR11:



### LCR12:

			2	293	57				293	367				1	293	877				
NC_063383.1/1-197209	tgata	act	t t i	t t	сt	a g	a a	aa	a a	a a	a a	t g	a a	t	t g	a	t g	a	t a	t a
353_R_nova_assembly/1-190773	tgata	acti	t t i	t t	c t	a g	a a	aa	a a	ı a i	a a	t g	a a		t g	а	t g	a	t a	t a
353_R_miseq_assembly/1-192918	tgata	acti	t t i	t t	c t	a g	a a	aa	a a	ı a i	a a	t g	a a		t g	а	t g	а	t a	t a
353_R_nanopore_flye_pilon/1-198541	tgata	acti	t t i	t t	c t	a g	a a	aa	a a	ı a i	a a	t g	a a		t g	а	t g	a	t a	t a
353_R_High_Quality/1-198556	tgata	acti	t t :	t t	c t	a g	a a	a a	a a	ı a i	a a	t g	a a	t	t g	а	t g	a	t a	t a
Consensus																				
	TGAT	AC T	TΤ	ТΤ	СТ	AG	6 A A	AAA	AA	۱A	ΑA	ΤG	AA	ιT	ΤG	βA	ΤG	A	ΤA	ΤA
Occupancy																				

#### LCR13:

	· .	76912	76922	76932
NC_063383.1/1-197209 353_R_nova_assembly/1-190773 353_R_miseq_assembly/1-192918 353_R_nanopore_flye_pilon/1-198541	g c t c c t g g c t c c t g g c t c c t g g c t c c t g	gtagcatac gtagcatac gtagcatac gtagcatac gtagcatac	ttttttta ttttttttta ttttttttta tttttttt	iccttgagaatatc iccttgagaatatc iccttgagaatatc iccttgagaatatc
353_R_High_Quality/1-198556	gctcctg	gtagcatac	ttttttt <mark>a</mark>	ccttgagaatatc
Consensus				
	GC TCC TC	G TAGC A TAC 1	ΤΤΤΤΤΤΤΤΑ	CC T TG AG A A T A TC
Occupancy				

#### LCR14:



#### LCR15:



#### LCR16:

![](_page_12_Figure_2.jpeg)

#### LCR18:

	166924	4 166934	166944	166954	166964
NC_063383.1/1-197209	atataaatccgtt	t a a a a t a a t t a a t	taattaataatt	a a t a a <mark>c g a a</mark> i	caagtatcaaa
353_R_nova_assembly/1-190773	atataaatccgtt	<mark>taa</mark> aataattaat	t <mark>aattaataa</mark> tt	aataa <mark>cgaa</mark>	caagtat caaa
353_R_miseq_assembly/1-192918	atataaatccgtt	taa aataattaa t	taattaa taa tt	aataa <mark>cgaa</mark>	caagtatcaaa
353_R_nanopore_flye_pilon/1-198541	atataaatccgtt	taa aataattaa t	taattaa taa tt	aataa <mark>cgaa</mark>	caagtatcaaa
353_R_High_Quality/1-198556	atataaatccgtt	<mark>taa</mark> aataattaat	taattaataatt	aataa cgaa	caagtatcaaa
Consensus					
	ATATAAATCCGT	TAAAATAATTAA	TAATTAATAATT	AATAACGAA	CAAGTATCAAA
_					
Occupancy					

## LCR19:

	170551	170561	170571	170581	170591	170601	170611	170621
NC_063383.1/1-197209	cttactactaacatg g	att <mark>c</mark> a <mark>g</mark> ataca	agatacagata	a cagata caga	itaca <mark>g</mark> ataca	g a ta ca ga ta c	<mark>a g</mark> a tg ta ga a	gatatcatg
353_R_nova_assembly/1-190773	cttactactaacatg g	att <mark>cag</mark> ataca	a <mark>g</mark> a ta ca <mark>g</mark> a ta	a <mark>c a g</mark> a t a c a g a	ıtaca <mark>g</mark> ataca	g a ta ca <mark>g</mark> a ta c	<mark>a ga t</mark> g ta ga a	gatatcatg
353_R_miseq_assembly/1-192918	cttactactaacatgg	att ca <mark>g</mark> ataca	a gata ca gata	a ca gata ca ga	ita ca <mark>g</mark> ata ca	g a ta ca g a ta c	<mark>a g a t</mark> g tagaa	gatatcatg
353_R_nanopore_flye_pilon/1-198541	cttactactaacatgg	attcagataca	agatacagata	a ca ga ta ca ga	ita cagata ca	gatacagatac	agatgtagaa	gatatcatg
353_R_Hign_Quality/1-198556	cttactactaacatgg	att <mark>cag</mark> ataco	agatacagata	a cagata caga	itacagataca	g a ta ca g a ta c	agatgtagaa	gatatcatg
Consensus								
	C T T A C T A C T A A C A T G G	A T TC AG A T AC /	AGATACAGAT	ACAGATACAGA	A TAC AG A TAC A	G A T A C A G A T A C	AGATGTAGAA	GATATCATG
Occupancy								

#### LCR20:

![](_page_13_Figure_1.jpeg)

175

Occupancy

#### B. Nanopore samples analysis downloaded from SRA

Monkeypox samples with nanopore data available in SRA (August 10, 2022) were downloaded from NCBI webpage (https://www.ncbi.nlm.nih.gov/sra) resulting in 35 samples. Next, reads were de novo assembled using flye assembler v2.9-b1768 (Kolmogorov et al. 2019) with default parameters in nanopore raw reads mode, and analyzed using a modified version of strsearch (https://github.com/BU-ISCIII/MPXstreveal). Reads spanning the defined LCR3, LCR1 and LCR4 regions were identified. Those with both flanking reads with 0 mismatches were collected, aligned and generated a consensus sequence. Potential number of repeats of selected LCRs in each sample were inspected by comparing the assembled genome and the strsearch result according to Supplementary data 6.

The "de novo" assembly method only resulted in the resolution of a few complete genomes. To identify the number of LCRs repeats, we mostly used the method spanning the flanking regions. Although the results differ in some cases between bioinformatic methods, and we cannot validate the exact numbers of repeats experimentally since we do not have access to the samples, the values obtained show clear differences in the tendency of the number of repeats. For clarity, the most conservative number was selected, considering the biology of MPXV replication, the solution that was supported by the highest number of genomic data, and the selection of the solution that underestimated variation.