Virologica Sinica 39 (2024) 134–143

Contents lists available at ScienceDirect

Virologica Sinica

journal homepage: www.keaipublishing.com/en/journals/virologica-sinica
www.virosin.org

Research Article

Rapid identification of full-length genome and tracing variations of monkeypox virus in clinical specimens based on mNGS and amplicon sequencing

Changcheng Wu [a,1], Ruhan A [a,1], Sheng Ye [b,1], Fei Ye [a], Weibang Huo [a], Roujian Lu [a], Yue Tang [c,h], Jianwei Yang [c], Xuehong Meng [d], Yun Tang [b], Shuang Chen [b], Li Zhao [a], Baoying Huang [a], Zhongxian Zhang [a,e], Yuda Chen [a,e], Dongfang Li [f], Wenling Wang [a], Ke-jia Shan [g], Jian Lu [g,*], Wenjie Tan [a,e,*]

[a] NHC Key Laboratory of Biosafety, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, 100052, China
[b] Chongqing Municipal Key Laboratory for High Pathogenic Microbes, Chongqing Center for Disease Control and Prevention, Chongqing, 400042, China
[c] MGI, BGI-Shenzhen, Shenzhen, 518083, China
[d] Thermo Fisher Scientific, Beijing, 100013, China
[e] School of Public Health, Baotou Medical College, Baotou, 014030, China
[f] BGI PathoGenesis Pharmaceutical Technology, Shenzhen, 518000, China
[g] State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, 100871, China
[h] College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China

ARTICLE INFO

Keywords:
Monkeypox virus (MPXV)
Metagenomic
Next generation sequencing
Amplicon
Third-generation sequencing

ABSTRACT

The monkeypox virus (MPXV) has triggered a current outbreak globally. Genome sequencing of MPXV and rapid tracing of genetic variants will benefit disease diagnosis and control. It is a significant challenge but necessary to optimize the strategy and application of rapid full-length genome identification and to track variations of MPXV in clinical specimens with low viral loads, as it is one of the DNA viruses with the largest genome and the most AT-biased, and has a significant number of tandem repeats. Here we evaluated the performance of metagenomic and amplicon sequencing techniques, and three sequencing platforms in MPXV genome sequencing based on multiple clinical specimens of five mpox cases in Chinese mainland. We rapidly identified the full-length genome of MPXV with the assembly of accurate tandem repeats in multiple clinical specimens. Amplicon sequencing enables cost-effective and rapid sequencing of clinical specimens to obtain high-quality MPXV genomes. Third-generation sequencing facilitates the assembly of the terminal tandem repeat regions in the monkeypox virus genome and corrects a common misassembly in published sequences. Besides, several intra-host single nucleotide variations were identified in the first imported mpox case. This study offers an evaluation of various strategies aimed at identifying the complete genome of MPXV in clinical specimens. The findings of this study will significantly enhance the surveillance of MPXV.

1. Introduction

The monkeypox virus (MPXV) is a member of the *Orthopoxviridae* family and causes disease in humans and wild animals. In 1958, MPXV was first detected in monkeys. The first confirmed human case of MPXV infection was reported in the 1970s (Ladnyj et al., 1972). The geographical distribution of early reported cases shows that MPXVs have been classified into two distinct groups (Bunge et al., 2022). The clade distributed in central Africa (Congo Basin) is named CB (or Clade I), and the other located in west Africa is called WA (or Clade II). The fatality rate of the CB clade (10.6%) is higher than the WA clade (3.6%) (Bunge et al., 2022). Having been endemic in Africa for about fifty years, MPXV is now exported to other continents. In 2003, an outbreak of MPXV was caused by imported wildlife in the United States. Since 2018, sporadic cases have been reported in the United States, Israel, Singapore, and the United Kingdom. MPXV is now back in the spotlight worldwide (Tan and Gao, 2022). In May 2022, a new MPXV outbreak occurred in Europe. Almost all the sequenced MPXVs originated from the same source, mainly

* Corresponding authors.
E-mail addresses: luj@pku.edu.cn (J. Lu), tanwj@ivdc.chinacdc.cn (W. Tan).
[1] Changcheng Wu, Ruhan A, and Sheng Ye contributed equally to this work.

https://doi.org/10.1016/j.virs.2023.12.002
Received 23 September 2023; Accepted 4 December 2023
Available online 7 December 2023

close to the WA clade (named B.1 lineage) (Aljabali et al., 2022; Anwar and Waris, 2022; Bunge et al., 2022; Hraib et al., 2022; Isidro et al., 2022). Besides, an independent outbreak (A.2 lineage) was reported in the United States during the same period (Gigante et al., 2022).

Nucleic acid amplification and serology tests can rapidly diagnose MPXV infection cases at a low cost, and we have also developed specific detection primers for this major outbreak strain B.1 (Huo et al., 2022). However, these assays cannot be used to monitor the sources, lineage, transmission, and genomic variations of MPXV. Besides, Wu et al. found variations in the target regions of the monkeypox virus detection primers published by the CDC (Wu F. et al., 2023). These newly emerging variations could lead to an underestimation of the viral load in samples and even result in misdiagnosis. To better carry out MPXV prevention and treatment, it is necessary to develop cost-effective and high-quality whole-genome sequencing strategies for MPXV.

MPXV is a double-stranded DNA virus with a length of 197 kb and consists of about two hundred genes. The frequency of A/T is twofold larger than C/G. Tandem repeats at the end of the genome exist in almost all orthopoxviruses (Shchelkunov et al., 2002). MPXV has a pair of inverted terminal repeats (ITR, about 7 kb), which contain a small number of open reading frames and elements essential for viral genome replication (Shchelkunov et al., 2002). The major strategies applied for MPXV genome sequencing are metagenomic and amplicon sequencing (Gorgé et al., 2022; Martínez-Puchol et al., 2022; Chen N. et al., 2023). The former starts with DNA from clinical specimens and constructs the sequencing library directly. MPXV accounts for only a low fraction of the total DNA (less than 1%) of clinical specimens. Therefore, in order to sequence the MPXV genome, the virus must first be isolated from clinical samples and cultured before undergoing genome sequencing (Fuchs et al., 2022). Virus isolation is time-consuming and requires strict conditions. Instead, amplicons can be used to enrich or amplify viral nucleic acids (Charre et al., 2020; Gohl et al., 2020; Chen N. et al., 2023). The next- and third-generation sequencing platforms have been applied to MPXV genome sequencing. The next-generation sequencing primarily uses optical or other technologies and sequences a cluster of fragments of the same source with high accuracy but short read lengths, including Illumina, MGI, and Ion torrent sequencing platforms. The third-generation sequencing is less affected by base composition bias and can directly sequence long fragments without amplification, including platforms from Pacific Biosciences (PacBio) and Oxford/Nanopore Technologies. Recent advancements in MPXV genome sequencing were also discussed in our recently published review (Chen Y. et al., 2023).

In this study, we utilized metagenomic and amplicon sequencing methods to accurately identify the complete genome of MPXV. Amplicon sequencing, known for its cost-effectiveness and efficiency, enabled rapid sequencing of clinical specimens, thus facilitating the acquisition of high-quality MPXV genome sequences. Leveraging third-generation sequencing technology enhanced the assembly of terminal tandem repeat regions while rectifying a recurring assembly error commonly observed in published genomes. Notably, the integration of amplicon sequencing with third-generation single-molecule real-time sequencing allowed for the rapid acquisition of high-quality full-length genomes of the monkeypox virus. Additionally, we successfully conducted mutation analysis and performed preliminary tracing of the viral genome for the first reported imported monkeypox case on Chinese mainland, as well as a case detected in early June 2023. Moreover, we identified several intra-host single nucleotide variants (iSNVs) of MPXV in blister fluid swab specimens, along with various co-existing microorganisms in the imported monkeypox case.

2. Materials and methods

2.1. Clinical specimen collection

The first imported mpox case in Chinese mainland was identified in Chongqing in September 2022 (Zhao et al., 2022). On September 16, we collected specimens of the case. In June 2023, numerous cases of mpox were reported in China, and a total of six samples were collected from four cases. The information of all specimens was provided in Table 1.

2.2. DNA extraction

Clinical specimen inactivation was performed in the Biosafety-Level-3 (BLS-3) laboratory of the National Institute for Viral Disease Control and Prevention in the Chinese Center for Disease Control and Prevention. According to the manufacturer's instructions, DNA was extracted using QIAamp DNA Mini Kit (Qiagen, USA).

2.3. Library construction and sequencing

2.3.1. Nanopore amplicon sequencing

DNA was subjected to multiplex PCR. The amplicon primers were modified based on Chen et al. with three pairs of redesigned primers to prevent amplicon drop-out (Chen N. et al., 2023). Reactions were generated using 2× Q5-High Fidelity master mix (NEB) with 3.6 μL of 10 μmol/L primer pool, and the conditions were as follows: after 3 min initial denaturation at 98 °C, followed by 35 cycles of 15 s at 98 °C and 5 min at 63 °C. The PCR products were purified using AMPure XP beads (Beckman Coulter), and the concentration and amplicon size was measured using the Qubit 3.0 (Thermo Scientific, USA) and Qsep100 Analyzer (BiOptic). Library preparation for Nanopore sequencing was conducted using Ligation Sequencing 1D (SQK-LSK109 Oxford/Nanopore Technologies), according to the manufacturer's instructions. About 300 fmol purified amplicons were treated with NEBNext Ultra II End repair and dA-tailing Module (New England Biolabs, USA), then purified using 0.5 volume of AMPure XP beads. The AMX sequencing adapter was attached using Quick T4 DNA ligase (NEB). Each library was loaded on an R9.4.1 flow cell (FLO-MIN106) and sequenced by GridION.

Table 1
Clinical specimen information.

Case No.	Sources	Sample types	Ct	Third-generation sequencing platform	Next-generation sequencing platform				
				Nanopore	DNBSEQ-G99		DNBSEQ-G400	Nextseq-2000	Ion torrent/Genexus
1	Chongqing	Oropharyngeal swabs	27.8	Amplicon	Amplicon	mNGS	mNGS	mNGS	/
	Chongqing	Blister fluid	20.7	Amplicon	Amplicon	mNGS	mNGS	/	/
	Chongqing	Nasal swabs	31	mNGS	/	/	/	/	/
	Chongqing	Blood	33.7		Amplicon				
	Chongqing	Cultured Vero cells	20.5	/	/	/	/	/	Amplicon
2	Beijing	Oropharyngeal swabs	28.5	/	Amplicon	/	/	/	/
	Beijing	Blister fluid	21.3	/	Amplicon	/	/	/	/
3	Guangdong	Blister fluid	22.5	/	Amplicon	/	/	/	/
4	Inner Mongolia	Oropharyngeal swabs	31.2	/	Amplicon	/	/	/	/
	Inner Mongolia	Blister fluid	21.5	/	Amplicon	/	/	/	/
5	Shanghai	Blister fluid	20.2	/	Amplicon	/	mNGS	/	/

2.3.2. DNBSEQ-G99 amplicon sequencing

The enrichment of monkey pox viral full-length genome used ATOPlex MPXV panel (MGI, China), PCR primers designed against the monkeypox reference (NC_063383.1) using the ATOPlex platform (https://atoplex.mgi-tech.com/) which contained 739 pairs of PCR primers with amplicon sizes ranging from 313 to 395 base pairs (bp). According to the manufacturer's instructions, 10 μL volume of extracted DNA was subjected to whole genome amplification using the ATOPlex Multiplex PCR Amplification kit (MGI, China). The PCR cycling conditions consisted of 5 min at 37 °C, 5 min at 95 °C, 37 cycles of 20 s at 95 °C, 1 min at 60 °C, 1 min at 58 °C, followed by 20 s at 72 °C, and a final extension of 1 min at 72 °C. The PCR product with 50 μL was purified using MGIEasy DNA Clean Beads (MGI, China) and quantified using a Qubit 3.0 (Invitrogen, USA). The resulting PCR product of 200 ng prepared as an amplicon sequencing library using MGIEasy Fast PCR-Free FS DNA Library Prep Kit (MGI, China) and conducted following steps including enzyme fragmentation, end repair, adapter ligation and library purification. The libraries were quantified using a Qubit 3.0 (Invitrogen, USA) and normalized. Subsequently, the normalized libraries were circularized using the DNBSEQ OneStep DNB Make Reagent Kit (MGI, China) and generated DNA nanoballs (DNBs). Finally, the DNB libraries were sequenced on a DNBSEQ-G99 platform with SE100 + 10+10 according to the manufacturer's standard protocol.

2.3.3. Metagenomic next-generation genome sequencing

Extracted monkeypox genomic DNA was dissolved in 1× TE buffer (Ambion, USA), the integrity (OD260/280 = 1.8–2.0) and quality of gDNA was measured using an Agilent 2100 BioAnalyzer (Agilent, USA) and Qubit 3.0 with Qubit dsDNA HS Assay Kit (Invitrogen, USA). The library preparation for NGS sequencing included: shearing of input gDNA by fragmentase, end repair, adapter ligation, PCR and purification. Fragmentation mixture was prepared on ice, consists of fragmentase and FS buffer all from MGIEasy Fast FS DNA Library Prep Module (940-000031-00, MGI). About 2.5 ng gDNA with 40 μL 1× TE buffer added into 15 μL fragmentation mixture, then transferred to a thermocycler to conduct reaction as followed: 1 min at 4 °C, followed by 8 min at 30 °C to break gDNA, and 20 min at 65 °C for enzyme inactivation. Sheared DNA was purified using En-Beads that mixed by MGIEasy DNA Clean Beads and 1× Elute Enhancer (MGI, China). Adapter ligation was conducted using Ad Ligase, Ligation Enhancer and Fast Ligation Buffer (MGI, China) according to the manufacturer's instructions. PCR reactions were generated using 38 μL ligation product, 12 μL UDB PCR Primer and 50 μL PCR Enzyme mix (MGI, China) of 100 μL pool in total, and the conditions consisted of 3 min initial denaturation at 95 °C, followed by 18 cycles of 20 s at 98 °C, 15 s at 60 °C and 30 s at 72 °C, and a final extension of 10 min at 72 °C. The PCR product were purified using DNA clean beads, concentration and libraries fragment sizes distribution were measured using the Qubit 3.0 and Qsep100 Analyzer (BiOptic). Finally, the normalized libraries were circularized using the DNBSEQ OneStep DNB Make Reagent Kit (MGI, China) and generated DNA nanoballs (DNBs). The prepared DNB were loaded onto patterned nanoarrays and sequenced on DNBSEQ-G99 with PE100 + 10+10 Bioanalysis By Sequencing (BBS) model and DNBSEQ-G400 with PE100 + 10+10 according to the manufacturer's standard protocol.

2.3.4. Ion torrent/Genexus amplicon sequencing

The Ion AmpliSeq MonkeyPox Panel was designed by Ion AmpliSeq technology with two primer pools, the average amplicon length is 125–275 bp, and the total amplicons are 1,609. After quantification by qPCR, a volume of 25 μL of a tenfold dilution of each sample was distributed in an Applied Biosystems MicroAmp EnduraPlate Optical 96-Well Clear Reaction Plate. The plate was sealed with a sheet of adhesive PCR plate foil and was placed on the Ion Torrent Genexus (GX) Integrated Sequencer (Thermo Fisher Scientific), which can sequence 3–4 samples per lane and 12–16 samples per GX5 chip (four lanes per chip) for the MPXV virus. The sequencing workflow was performed using the Ion AmpliSeq MonkeyPox Panel on GX, an automated library preparation, sequencing, analysis, and reporting technology.

2.4. mNGS analysis

Raw reads were processed using Fastp v0.23.2 (Chen et al., 2018) to remove adapters and low-quality bases. Clean reads were mapped to the human genome (hg38) to remove the host using bwa v0.7.17 (Li and Durbin, 2009). The non-host reads were then used to identify potential pathogens using the Pathogeny Fast Identification (PFI) V5.0 pipeline (MGI).

2.5. Virus genome assembly

2.5.1. Metagenomic sequencing datasets

Clean reads were mapped to the genome of MPXV (NC_063383.1) using bwa. Low-quality alignments with a high proportion of soft-clipped bases were filtered. GATK4 Mutect2 (McKenna et al., 2010) and BCFtools v1.9 (Danecek et al., 2021) were used to call and filter variants. Consensus sequences were generated by modifying the reference genome according to the identified variants.

2.5.2. Amplicon dataset analysis

Reads were analyzed as the methods described in Metagenomic analysis except for primer trimming for datasets generated by the next-generation sequencing platforms. A modified ARTIC v1.2.1 minion module was used for reads generated by Nanopore. Reads alignments were viewed with Integrative Genomics Viewer (IGV) (Robinson et al., 2011). The codes were provided in GitHub (https://github.com/BioWu/Monkeypox-genome-assembly).

2.6. Tandem repeats copy number determination

The high-quality MPXV genome sequences were retrieved from the National Center for Biotechnology Information (Sayers et al., 2021) and GISAID (Shu and McCauley, 2017) (https://www.gisaid.org, as of 2022/9/27). Tandem repeats were identified using TRF v4.09 (Benson, 1999). Third-generation sequencing reads were used to determine the copy number variations of the 16-bp TR. The following primers were used in Sanger sequencing: forward: 5'-CACATTTTAGTCGTTTATTC-3' and reverse: 5'-TCTATCCACCCACCTTTCTT-3'. The consensus sequence obtained from Sanger sequencing analysis was submitted to the Genbase with the assigned ID: C_AA044237.1.

2.7. Phylogenetic analysis of the sequenced MPXV genomes

Variations of downloaded MPXV and newly sequenced genomes were called using Nextclade v2.4.0 (Aksamentov et al., 2021). The mutations were annotated by SnpEff v5.0e (Cingolani et al., 2012) based on the reference genome. As the assembly quality of low complexity regions was low, only single base mutations were used to determine the similarity between sequences. To build the phylogenetic tree of MPXV, we sampled the downloaded sequences according to lineages. The sampled and newly assembled viral genomes were combined and aligned using Nextclade. The ITRs were masked, and the phylogenetic tree was built using iqtree2 v2.2.0 (-m GTR + F + I + G4) (Nguyen et al., 2015) and viewed using ggtree (Yu et al., 2017). The numbers of nonsynonymous (N) and synonymous (S) sites in MPXV were estimated using PAML v4 (Yang, 2007) based on the published MPXV genomes.

3. Results

3.1. Performance of multiple strategies applied for MPXV genome sequencing

MPXV genome sequencing mainly adopts the next-generation metagenomic sequencing (mNGS) strategy (Isidro et al., 2022; Chen Y. et al.,

2023). Amplicon approaches can effectively enrich target pathogenic nucleic acids and have been widely used in genome sequencing of SARS-CoV-2 (Charre et al., 2020; Gohl et al., 2020). However, their performance in MPXV genome sequencing among different clinical specimens has only recently been reported (Chen N. et al., 2023). This study collected swabs of blister fluid (BF), oropharyngeal (OS) and nasopharyngeal (NS), and serum (SE) of the first imported MPXV case in Chinese mainland (MPXV/ChinaCQ-202209) and four local mpox cases. mNGS and amplicon sequencing were applied in parallel using three sequencing platforms. Viral load was estimated using Real-time quantitative PCR (Tables 1 and 2). The blister fluid swabs had the lowest Ct value (20.76) and the highest viral load. The serum specimen carried the lowest viral load and was sequenced using the amplicon approach. The nasopharyngeal swab specimen was sequenced using mNGS. Amplicon and third-generation sequencing platform (Oxford/Nanopore) was adopted additionally in oropharyngeal swabs and BF specimens. To investigate the performance of different amplicon methods, Vero cells (Ct = 20.5), after infecting the BF swabs, were collected and sequenced using another amplicon in the Ion torrent/Genexus (GX) platform. The applied sequencing strategies for the imported and four local mpox cases are listed in Tables 2 and 3. All of them achieved high sequencing depth and coverage rates (Fig. 1A and B and Supplementary Fig. S1). They generated nearly perfect assemblies except for mNGS of the OS specimen (10 × coverage rate = 81.38%) and amplicon of the SE specimen (10 × coverage rate = 17.91%) from the imported mpox case.

Without PCR amplification, the proportion of reads mapped to MPXV genome was less than 0.03% in OS and about 0.4% in the BF swabs. However, nearly all reads were mapped to the MPXV genome in amplicon datasets except for the serum sample, suggesting that amplicon can enrich and amplify the virus DNA efficiently. Chen et al. also compared the two sequencing strategies and found that amplicon had better performance on clinical specimens with low MPXV load (Ct > 18) (Chen N. et al., 2023). Among the amplicon datasets, the OS specimen sequenced in Oxford/Nanopore (ONT) had the lowest 10 × coverage rate (98.85%), while the rate of the amplicon on DNBSEQ-G99 was 99.66% (Table 2). The major difference is that ONT-amplicon has larger amplified fragments than the latter (1,500~2,000 bp *vs* 200 bp). While in the NS specimen, the coverage rates of these two amplicon approaches are both larger than 99%. Therefore, an amplicon with shorter amplified fragments may be more suitable for samples with low viral load (*e.g.,* Ct > 27). The proportion of reads mapped to MPXV genome was very low in the serum amplicon dataset, and most of the genome was not sequenced. We suspected that the fragmentation of the MPXV genome in serum was severe. Therefore, these specimens should be avoided in MPXV genome sequencing.

Large-fragment amplicons had a slight decrease in amplification efficiency in some regions. Further, we found that most of the amplified fragments in regions with reduced amplification efficiency were longer than 2 kb. The genome coverage rate of the short-fragment amplicons is higher, but the initial investment cost is also higher, and it is susceptible to the influences of new mutations. Therefore, multiple factors such as efficiency, potential off-target possibility, and cost should be fully balanced when designing the pool of amplification primers. As only limited specimens were collected in this study, the proper Ct value for short- and long-fragment amplicons still need more studies.

Increasing sequencing depth could improve the quality of MPXV genome assembly (Fig. 1B; Table 2). However, the cost of the experiment increases in parallel. To explore the least required sequencing dataset of specimens with different Ct values, we randomly sampled reads and examined dynamic changes in genome coverage rates (10×). In the case of mNGS, 10 million (M) reads were enough to cover nearly the whole genome of MPXV at 10 × in BF (Ct < 21) and NS (Fig. 1C). While at least 100 M reads were required in the OS specimen (Ct = 27.81), and a whole sequencing run (chip) was needed for one sample in DNBSeq-G99 or Nextseq 2000. If applying amplicon sequencing, at most 5 M reads could cover nearly the whole genome (Fig. 1D). Half this many reads were enough for those high viral load samples. Long-fragment third-generation (Oxford/Nanopore) amplicon requires less than 1 M reads even in the OS swab with low viral load. We prepared cell culture with Ct close to the BF swabs and did MPXV genome sequencing on Ion Torrent Genexus (GX) Integrated Sequencer. The sequencing depth of GX is evenly distributed across the genome, and fewer reads were required (Fig. 1D).

Table 2
Summary of sequencing datasets from the first imported mpox case in Chongqing.

Methods	Platform	Sample	Ct	Clean reads (M)	MPXV-mapped reads (M)	Average depth	10X Coverage rate (%)	Cost	
								Time (hour)	Money[d] ($1000/sample)
Amplicon	ONT[a]	Oropharyngeal swabs	27.81	0.92	0.8	596	98.85	8–12	1
		Blister fluid swabs	20.76	0.86	0.8	669	99.34	8–12	1
	DNBSEQ-G99[b]	Oropharyngeal swabs	27.81	69.54	69.07	6,631	99.66	12	1
		Blister fluid swabs	20.76	66.34	66.06	18,062	99.74	12	1
	Ion torrent/GX	Cultured Vero cells	20.51	5.74	5.74[c]	4,000	99.61	15	0.5
mNGS	DNBSEQ-G99	Oropharyngeal swabs	27.81	131.18	0.03	15	81.38	12	2
		Blister fluid swabs	20.76	173.38	0.66	335	99.93	12	2
	DNBSEQ-G400[b]	Oropharyngeal swabs	27.81	1,167.57	0.32	162	99.91	48	2
		Blister fluid swabs	20.76	1,692.74	7.35	3724	99.99	48	2
	Nextseq-2000[b]	Nasal swabs	31.12	249.24	0.49	263	99.99	24	3

[a] Oxford/Nanopore Single Molecular Real-Time sequencing.
[b] Next Generation Sequencing.
[c] Only read mapped to MPXV were exported by the built-in analysis system.
[d] Cost estimation was based on this experiment, and simultaneous sequencing of multiple samples might reduce the cost.

Table 3
Summary of sequencing datasets from four local mpox cases.

Clinical specimens	Strategies	Clean reads (M)	MPXV-mapped reads (M)	Average depth	10× Coverage rate (%)
Blister fluid of the mpox case from Shanghai	mNGS	1,129.74	3.94	1,882	99.99
Blister fluid of the mpox case from Shanghai	Amplicon	35.04	31.72	8,026	99.30
Oropharyngeal swabs of the mpox case from Beijing	Amplicon	99.58	80.08	28,797	99.40
Blister fluid of the mpox case from Beijing	Amplicon	30.54	30.19	12,778	99.70
Oropharyngeal swabs of the mpox case from Inner Mongolia	Amplicon	48.69	25.48	10,307	99.30
Blister fluid of the mpox case from Inner Mongolia	Amplicon	11.85	11.45	4,657	99.40
Blister fluid of the mpox case from Guangdong	Amplicon	79.37	76.49	31,533	99.70

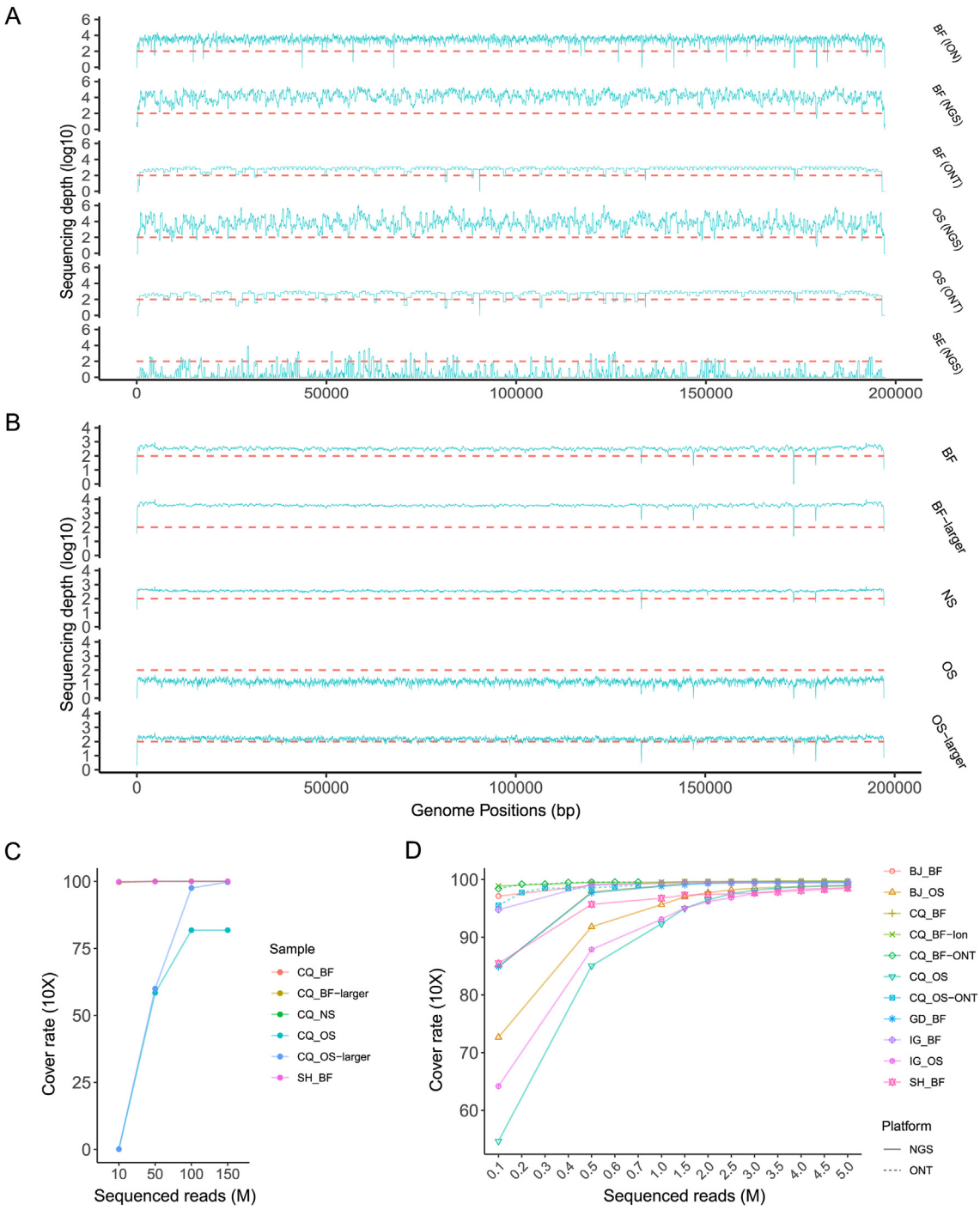


Fig. 1. Genome sequencing of MPXV using mNGS and amplicon. **A** Sequencing depth of MPXV from the first imported mpox case using amplicon strategies. NGS, Next Generation Sequencing; ONT, Oxford/Nanopore; ION, Ion torrent; BF, blister fluid swabs; CC, cultured cells; OS, oropharyngeal swabs; SE, serum; The red line indicates 100× sequencing depth. **B** Coverage of MPXV from the first imported mpox case using metagenomic sequencing strategies. NS, nasopharyngeal swab; BF/OS-larger, mNGS with larger sequencing depth. 10× genome coverage rate of MPXV with randomly sampled reads from mNGS (**C**) and amplicon (**D**) datasets of all the mpox cases. Amp, amplicon. BJ, Beijing; CQ, Chongqing; GD, Guangdong; IM, Inner mongolia; SH, Shanghai.

We also roughly estimated the time and cost required for these experiments (Tables 2 and 3). Although we attempted to include a larger number of clinical samples from monkeypox cases to objectively assess sequencing costs, this study ultimately only incorporated a limited number of samples. The amplicon sequencing cost per sample would decline hugely if more samples were sequenced in the same chip simultaneously. However, while mNGS is suitable for sequencing specimens with high viral loads such as blister fluid, its application to low viral load samples, such as oropharyngeal swabs, results in exorbitant costs. On the other hand, amplicon sequencing proves to be an effective approach for improving genome quality while significantly reducing expenses.

3.2. *Third-generation sequencing benefits the assembly of tandem repeats*

Third-generation sequencing platform generates longer reads than next-generation sequencing ones and has been proven to get the Telomere-

to-Telomere human genome (Nurk et al., 2022). Therefore, it can assist in assembling tandem repeat regions of the MPXV genome. The inverted tandem repeat (ITR) region at the dual ending of MPXV contains a complex region composed of multiple 16bp-TRs (left ITR: 5'-AACTAACTTAT-GACTT-3'; right ITR: 5'-AGTCATAAGTTAGTTA-3'). It locates between *J3L* and *D1L* in the left ITR and between *N3R* and *N4R* in the right ITR. Reads generated by the NGS platform are usually 100–150 bp. If the copy number of the 16bp-TR is larger than 8, the region would be collapsed during assembly, resulting in assembly errors (Fig. 2A). We analyzed published MPXV genomes and found that the copy number of the 16bp-TR was highly polymorphic. Most of them had only 7.9 copies of 16bp-TR, and the rest contained about 15.9 TRs (Fig. 2B). To clarify the accurate copy number of the 16bp-TR in the imported MPXV, we analyzed third generation sequencing datasets and identified an additional 128 bp insertion near the original 7.9 copies of 16bp-TR on the reference genome (Fig. 2C). And this insertion contains exactly 8 copies of 16bp-TR. The copy numbers of the 16-bp tandem repeat (TR) were also validated through Sanger sequencing. Therefore, we speculate that each ITR of the current outbreak of MPXV mostly contained 15.9 copies of 16bp-TR. These results suggest that third-generation sequencing is beneficial for resolving the assembly of complex tandem repeat regions. Low-complexity region (LCR) has been treated as a different type of genomic accordion resulting in gene length variation (Monzón et al., 2022), and has even been used to cluster MPXV (Yeh et al., 2022). Therefore, in subsequent studies, more attention should be paid to copy number polymorphisms of the 16bp-TR led by sequencing platforms or data analysis methods.

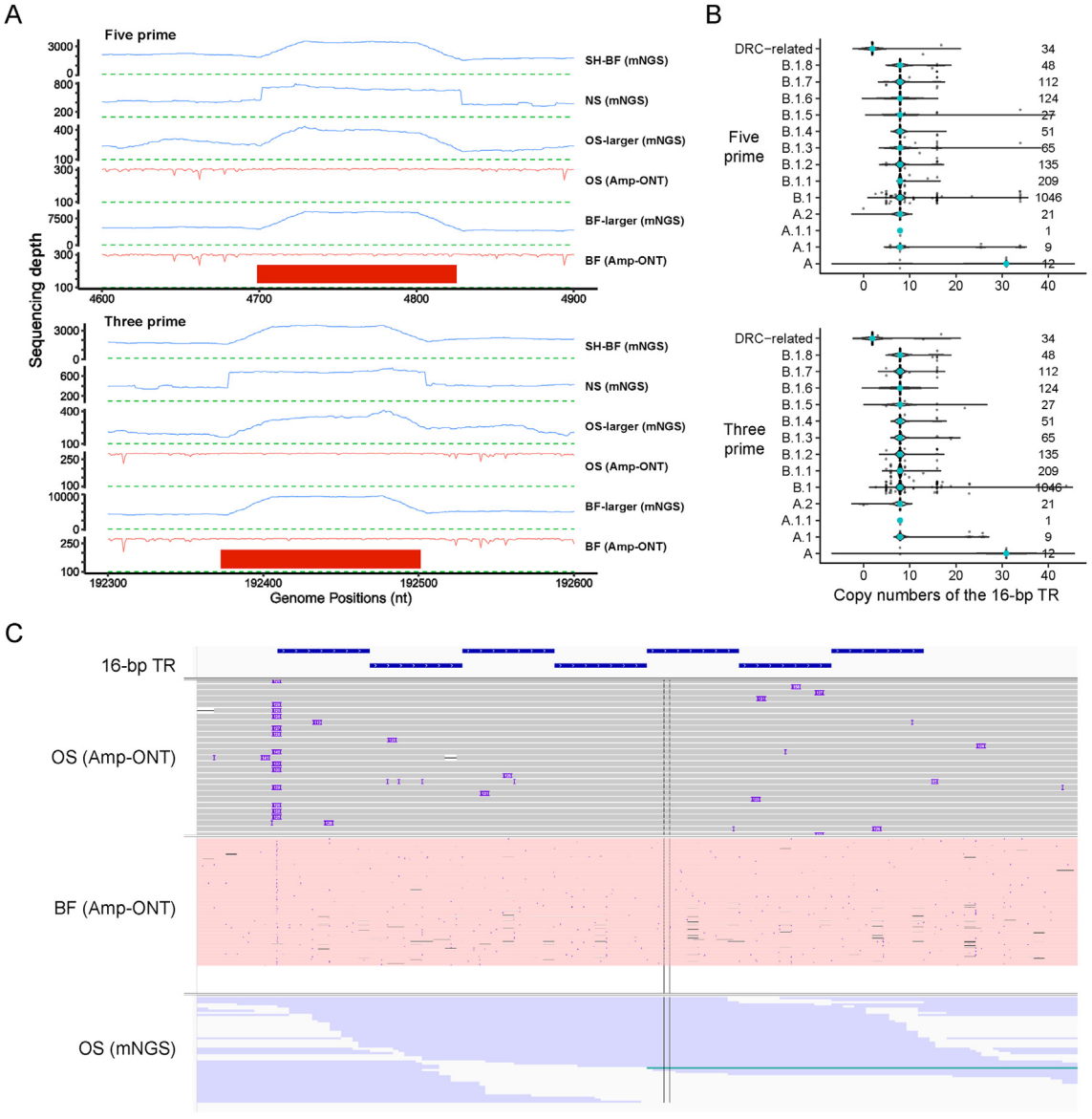


Fig. 2. Third-generation sequencing benefits the assembly of tandem repeats. **A** Sequencing depth of the regions, including 16bp-tandem repeats (TRs) at the dual ending of MPXV. BF, blister fluid swabs; OS, oropharyngeal swabs; NS, nasopharyngeal swabs; Amp, amplicon; mNGS, metagenomic Next Generation Sequencing. BF/OS-larger, mNGS with larger sequencing depth. Red rectangles, the regions including 16bp-TRs. All datasets except for SH-BF (blister fluid specimen from the mpox case in Shanghai) were obtained from the first import of mpox case. **B** Summary of the copy numbers of the 16bp-TR in published MPXV genomes. The median of copy numbers of the 16bp-TRs was highlighted in blue points. The numbers of MPXV genome sequences of each group were listed on the right. DRC, Democratic Republic of the Congo. **C** Read mapping results of the 16bp-TR region. ONT, Oxford/Nanopore. The purple rectangles are 128 bp insertion markers (8 copies of the 16-bp TR). Blue rectangles, the regions of 16bp-TRs.

3.3. Molecular tracing of the sequenced monkeypox virus genomes

Molecular tracing based on the genome of viruses has played an important role in the investigation of the transmission chain of the new outbreak. To clarify the source of MPXV from the imported and local mpox cases, we estimated nucleotide distances (*e.g.,* single nucleotide difference) between this newly assembled MPXV and those downloaded from the GISAID database and reconstructed the phylogenetic tree. The most closely related MPXV genome to MPXV/CHINACQ-20220 was from the one confirmed by Germany in June 2022 (GISAID: EPI_ISL_13889435). Relative to the EPI_ISL_13889435 variant, MPXV/CHINACQ-20220 has seven nucleotide differences (C16940T, C23105T, C54262T, C63921T, G120262A, C160702T, and C154855T). Both MPXV/CHINACQ-20220 and EPI_ISL_13889435 belonged to lineage B.1 of IIb-B, and they were closest to each other (Fig. 3). The MPXV/CHINACQ-20220 case had traveled to Germany recently and developed clinical signs consistent with MPXV infection when returning to China, supporting the notion that MPXV/CHINACQ-20220 was imported from Germany. The monkeypox virus circulating in Chinese mainland in June 2023 belongs to the C.1 lineage (also known as B.1.3.1) (Wu C. et al., 2023). These genomic sequences are most closely related to the reported ones from Japan, Portugal, and other locations (Fig. 3). These results indicate that our sequencing provides both the required repeatability and quality for accurate molecular tracing.

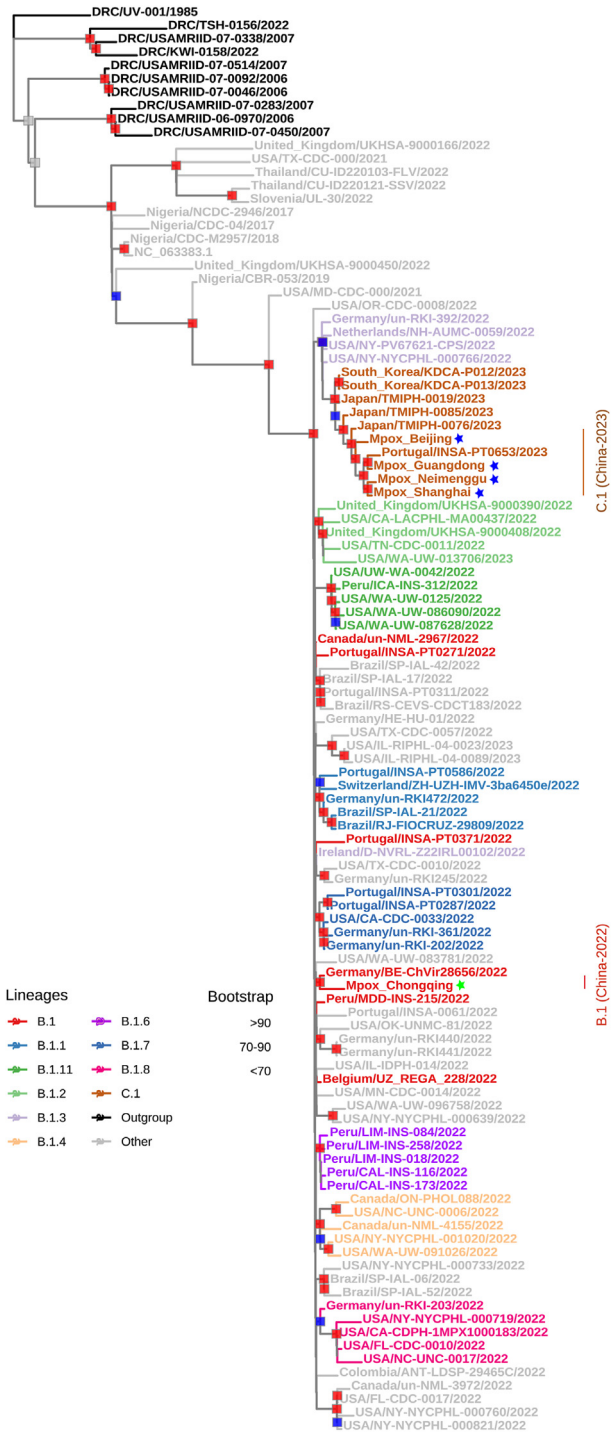


Fig. 3. Phylogenetic tree of MPXV. The green and blue five-pointed stars correspond respectively to the first imported case of monkeypox in Chongqing and four local cases of monkeypox in 2023. The virus from the first imported and four local mpox case and the closely related sequence from Germany and Japan, respectively.

3.4. Intra-host mutations of MPXV were identified in the imported mpox case

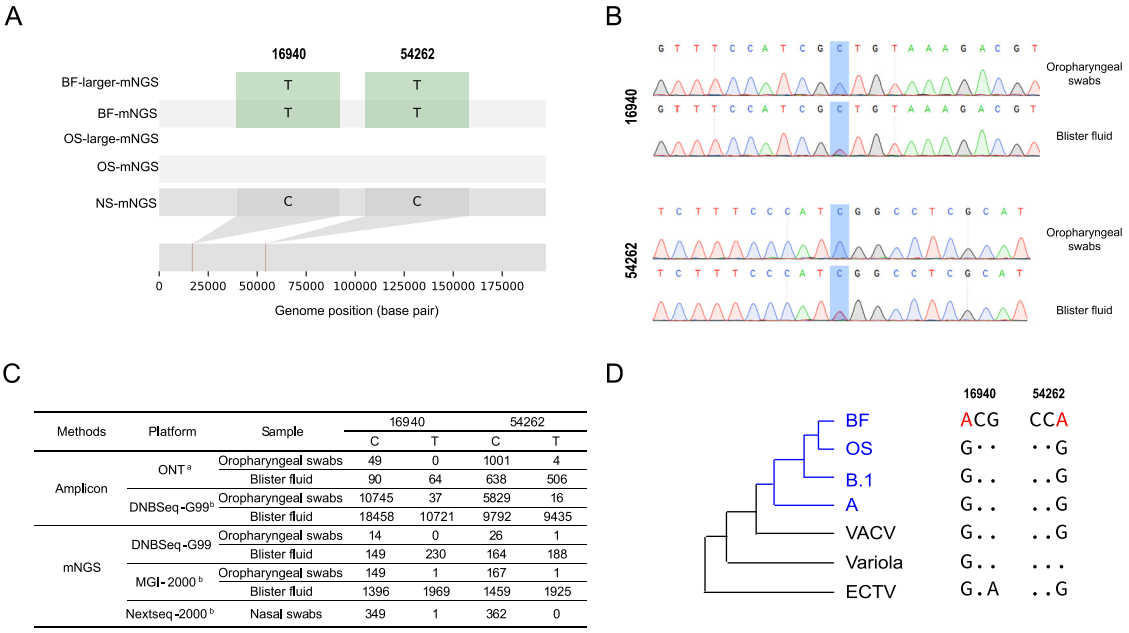
The MPXV genome mutation rate is approximately 6×10^{-5} substitutions per site per year (https://nextstrain.org/monkeypox/hmpxv1), substantially lower than single-stranded RNA viruses such as SARS-CoV-2 (9.69×10^{-4} substitutions per site per year) (Liu et al., 2020). Intra-host single nucleotide variations (iSNVs) were found in COVID-19 patients, providing important information for the genetic diversity and evolution selection of viruses in infected hosts (Lythgoe et al., 2021; Wang et al., 2021; Li et al., 2022; Pathak et al., 2022). The iSNVs could be fixed through random processes (*e.g.*, genetic drift) or natural selection. However, few studies focused on the characteristics of iSNVs in MPXV-infected cases. We analyzed the sequencing datasets of three clinical specimens and identified 73 homozygotic mutations (relative to the reference), including 36 nonsynonymous (N), 28 synonymous (S), and 9 intergenic variants. The observed N/S ratio is significantly lower than the expected N/S ratio under randomness (36/28 *vs* 126,244/39,292; $P = 0.0006$; Fisher's Exact Test), suggesting that MPXV genome undergo purifying selections.

Although heterozygotic mutation sites were not found in oropharyngeal and nasopharyngeal swabs, at least four heterozygotic mutation sites were in the blister fluid swabs (The list of SNVs was provided in Supplementary Tables S1 and S2). Two loci (C16940T and C54262T), with frequencies about 50%, were simultaneously supported by metagenomic, amplicon and verified using Sanger sequencing (Fig. 4A–C). C16940T has never been reported before, and C54262T was detected in only five published sequences. All of the five sequences belonged to the B.1 lineage and were uploaded by Canada, Germany, USA, and Slovakia. C54262T is the same as smallpox. C16940T exists in NBT03_gp012 (IFN-beta inhibitor) and introduces the A44T nonsynonymous mutation. C54262T locates in NBT03_gp050 (DNA polymerase, catalytic subunit) and is a synonymous mutation (CCG to CCA, Fig. 4D). We speculated that these heterozygotic mutations likely have occurred in this case recently. Their biological functions and adaptation to the host are still mysterious. As the number of iSNV is limited, whether purifying selection existed in these heterozygotic iSNVs is still unclear. These results also suggest that collecting clinical specimens from multiple loci of mpox cases is necessary for viral genome sequencing.

4. Discussion

4.1. Selecting appropriate sequencing strategies based on the viral load of clinical specimens

Based on various sequencing strategies and platforms, we performed whole-genome sequencing of monkeypox virus in nine nasal, oropharyngeal, and vesicle fluid swab samples, including the first imported case



Methods	Platform	Sample	16940		54262	
			C	T	C	T
Amplicon	ONT[a]	Oropharyngeal swabs	49	0	1001	4
		Blister fluid	90	64	638	506
	DNBSeq-G99[b]	Oropharyngeal swabs	10745	37	5829	16
		Blister fluid	18458	10721	9792	9435
mNGS	DNBSeq-G99	Oropharyngeal swabs	14	0	26	1
		Blister fluid	149	230	164	188
	MGI-2000[b]	Oropharyngeal swabs	149	1	167	1
		Blister fluid	1396	1969	1459	1925
	Nextseq-2000[b]	Nasal swabs	349	1	362	0

Fig. 4. Intra-host single nucleotide variants of MPXV from the Chongqing mpox case. **A** Base difference among multiple clinical specimens in two intra-host single nucleotide variants. BF, blister fluid swabs; OS, oropharyngeal swabs. NS, nasopharyngeal swabs. **B** Verification of two heterozygotic iSNVs using Sanger sequencing. **C** Raw read counts of two heterozygotic iSNVs. **D** Conservation of the codons, including these two iSNVs. A/B.1, lineages of MPXV; VACV, Vaccinia virus; Variola, smallpox; ECTV, Ectromelia virus. Two iSNVs are in red. Tips in blue are MPXV.

in Chinese mainland and four local mpox cases. The three major sequencing platforms are able to accurately identify genomic variation sites in the samples. Excluding the ITR region at both ends, the identification results of mutation sites are completely consistent. There were significant differences in viral load among different types of clinical samples, requiring the selection of suitable strategies based on sample type. For samples with high viral load, such as blister fluid, both metagenomic sequencing and amplicon sequencing could generate high-quality full-length viral genomes. However, for samples with low viral load like oropharyngeal and nasal swabs, direct metagenomic sequencing is not recommended due to the large amount of sequencing data required, which is highly cost-intensive. In addition, although the RT-qPCR results demonstrated the presence of monkeypox virus in the serum specimen, amplicon sequencing indicated that the sample may only contain short viral nucleic acid fragments. Therefore, it is recommended to prioritize the collection of swabs from lesion sites with higher viral load during clinical sample collection.

By analyzing the minimum amount of sequencing data required for different sequencing strategies, we found that amplicon enrichment generally requires less than 5 million sequencing reads. Therefore, when there are a large number of samples to be sequenced, amplicon sequencing can maximize the utilization of sequencing capacity and greatly reduce the sequencing cost per sample. Although mNGS showed slight disadvantages in the whole-genome sequencing of monkeypox virus, it can unbiasedly detect coexisting microorganisms at the sampling sites of the cases. For example, we detected various potential pathogenic or colonizing microorganisms in the clinical samples of the first imported mpox case (Supplementary Table S3). Most of the identified microorganisms are colonizing bacteria or contaminants. These results will contribute to the better diagnosis and treatment of the cases. It should be noted that the performance of mNGS sequencing requires the filtering of personal privacy information, such as host genomic data, in accordance with ethical and institutional regulatory rules.

4.2. Overlapping amplicons and new sequencing technologies enable rapid generation of high-quality MPXV genomes

In infectious disease prevention and control, especially for the identification of unknown pathogens, time is crucial. In this study, various new sequencing technologies combined with amplicon enrichment strategies were applied. Even for oral swab samples with low viral load (Ct = 27.81), we were able to obtain the complete viral genome sequence within 12 h. Long fragment amplicons (1.5–2 kb) combined with single-molecule real-time sequencing technology have the advantages of convenient library construction, short processing time, and real-time sequencing. It should be noted that this technology requires a relatively high total nucleic acid load and produces a relatively small amount of data per chip, making it generally suitable for fast sequencing after amplicon library preparation. Moreover, this study employed the DNBSEQ-G99 sequencing platform, developed by MGI, for the first time in preventing and controlling infectious diseases. This system can do the initial analysis within 2 h after running and complete paired-end 150 sequencing within 12 h. The application of new technologies greatly shortened the time required for sequencing MPXV.

4.3. Multiple-site sampling contributes to the monitoring of monkeypox virus microevolution and precise molecular tracing

The metagenomic sequencing of the first imported case and the first-generation sequencing validation supported the presence of iSNVs in the vesicle fluid sample. However, these several mutation sites were not identified in the oropharyngeal swab sample. Possible reasons for this situation include: (1) the case was infected with multiple monkeypox viruses; (2) the monkeypox virus underwent microevolution at the vesicle site. Mutation site analysis showed that the virus in the vesicle fluid sample contained all the mutation sites of the virus in the oropharyngeal sample. Therefore, we tend to agree with the second reason, that

is, the occurrence of microevolution in the monkeypox virus. Multiple studies have shown that the evolutionary rate of the B.1 branch of monkeypox virus exceeds the previous lineages, providing a theoretical basis for the appearance of the iSNVs. Therefore, in order to more comprehensively understand the diversity of monkeypox virus carried by cases and to provide early warning of genomic mutations, it is recommended to separately collect and sequence multiple affected areas of individual cases.

5. Conclusions

We conducted a comprehensive evaluation of metagenomic sequencing, amplicon sequencing, and multiple sequencing platforms for high-quality whole-genome sequencing of monkeypox virus, using real clinical samples from cases with varying viral loads and types. The results indicate that the currently available mainstream sequencing platforms can meet the requirements for whole-genome sequencing of MPXV. However, the choice of sequencing strategy should be carefully considered, considering factors such as sample quantity, expected duration and cost, and the need to focus on variations in complex repetitive regions at the terminus. It is important to note that due to the constraints of experimental costs and the scarcity of clinical samples, this study had a limited number of cases and samples. Future studies should conduct further evaluations and optimizations of sequencing strategies for a wider range of viral loads or sample types. This study provides valuable assistance in monitoring variations and issuing early warnings for significant pathogens, such as the monkeypox virus.

Data availability

All the sequenced datasets were deposited to GSA (Accession Number: CRA008627). The MPXV genome from the first imported mpox case was uploaded to GISAID (Accession Number: EPI_ISL_15293815). The sequence of 16bp-TR obtained from Sanger sequencing was submitted to the Genbase (assigned ID: C_AA044237.1). The sequences of the four local mpox cases are available from the corresponding author upon request.

Ethics statement

The informed consent has been obtained from all participants. The studies have been approved by the ethics committee of the National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention.

Author contributions

Changcheng Wu: formal analysis, writing-original draft, writing-review & editing. Ruhan A: data curation, formal analysis. Sheng Ye: data curation, formal analysis. Yun Tang: data curation, formal analysis. Shuang Chen: data curation, formal analysis. Yue Tang: data curation, formal analysis. Jianwei Yang: data curation, formal analysis. Xuehong Meng: data curation, formal analysis. Dongfang Li: data curation, formal analysis. Li Zhao: methodology. Baoying Huang: methodology. Roujian Lu: methodology. Wenling Wang: methodology. Yuda Chen: methodology. Zhongxian Zhang: methodology. Kejia Shan: methodology. Wenjie Tan: unding acquisition, investigation, writing-review & editing. Jian Lu: funding acquisition, investigation, writing-review & editing.

Conflict of interest

Yue Tang and Jianwei Yang are employed by MGI Tech Co., Ltd. Xuehong Meng is employed by Thermo Fisher Scientific Inc. Dongfang Li is employed by BGI PathoGenesis Pharmaceutical Technology Co., Ltd.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2022YFC2303401, 2022YFC2304100, 2016YFD0500301, 2021YFC0863300), the Beijing Science and Technology Plan (Z211100002521017), and the National Natural Science Foundation of China (82241080). We thank Wenbo Xu, Yanhai Wang, Hua Ling, and others for transferring clinical samples; Yong Zhang for supporting the sequencing platform of DNBSEQ-G400; Wei Fang, Yuan Zhang, Zhi Huang, Hsin Hua Yang, and Yue Han for supporting the sequencing platform of Ion Torrent/Genexus; Wansi Mao, Jie Chen, Peng Gao for technical supporting of data analysis of DNBSEQ-G99; Zhangping Tan and Minyue Wang for the help of sequencing of nasal swabs.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virs.2023.12.002.

References

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

Aljabali, A.A., Obeid, M.A., Nusair, M.B., Hmedat, A., Tambuwala, M.M., 2022. Monkeypox virus: an emerging epidemic. Microb. Pathog. 173, 105794.

Anwar, F., Waris, A., 2022. Monkeypox virus outbreak: a brief timeline. New Microb New Infect 48, 101004.

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580.

Bunge, E.M., Hoet, B., Chen, L., Lienert, F., Weidenthaler, H., Baer, L.R., Steffen, R., 2022. The changing epidemiology of human monkeypox-A potential threat? A systematic review. PLoS Neglected Trop. Dis. 16, e0010141.

Charre, C., Ginevra, C., Sabatier, M., Regue, H., Destras, G., Brun, S., Burfin, G., Scholtes, C., Morfin, F., Valette, M., Lina, B., Bal, A., Josset, L., 2020. Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation. Virus Evol 6, veaa075.

Chen, N.F.G., Chaguza, C., Gagne, L., Doucette, M., Smole, S., Buzby, E., Hall, J., Ash, S., Harrington, R., Cofsky, S., Clancy, S., Kapsak, C.J., Sevinsky, J., Libuit, K., Park, D.J., Hemarajata, P., Garrigues, J.M., Green, N.M., Sierra-Patev, S., Carpenter-Azevedo, K., Huard, R.C., Pearson, C., Incekara, K., Nishimura, C., Huang, J.P., Gagnon, E., Reever, E., Razeq, J., Muyombwe, A., Borges, V., Ferreira, R., Sobral, D., Duarte, S., Santos, D., Vieira, L., Gomes, J.P., Aquino, C., Savino, I.M., Felton, K., Bajwa, M., Hayward, N., Miller, H., Naumann, A., Allman, R., Greer, N., Fall, A., Mostafa, H.H., Mchugh, M.P., Maloney, D.M., Dewar, R., Kenicer, J., Parker, A., Mathers, K., Wild, J., Cotton, S., Templeton, K.E., Churchwell, G., Lee, P.A., Pedrosa, M., Mcgruder, B., Schmedes, S., Plumb, M.R., Wang, X., Barcellos, R.B., Godinho, F.M.S., Salvato, R.S., Ceniseros, A., Breban, M.I., Grubaugh, N.D., Gallagher, G.R., Vogels, C.B.F., 2023. Development of an amplicon-based sequencing approach in response to the global emergence of mpox. PLoS Biol. 21, e3002151.

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890.

Chen, Y., Wu, C., A, R, Zhao, L., Zhang, Z., Tan, W., 2023. Perspective on the application of genome sequencing for monkeypox virus surveillance. Virol. Sin. 38, 327–333.

Cingolani, P., Platts, A., Wang Le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6, 80–92.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., Mccarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. GigaScience 10, giab008.

Fuchs, J., Nekrutenko, A., Kohl, A.-K., Technau-Hafsi, K., Hornuß, D., Rieg, S., Jaki, L., Huzly, D., Maier, W., Grüning, B., Gutbrod, L., Falcone, V., Hengel, H., Panning, M., 2022. Travel-associated Monkeypox Virus Genomes from Two German Patients and of a Derived Virus Isolate All Closely Related to a US Sequence, 2022 [Online]. Available: https://virological.org/t/travel-associated-monkeypox-virus-genomes-from-two-german-patients-and-of-a-derived-virus-isolate-all-closely-related-to-a-us-sequence-2022/844. (Accessed 3 June 2022).

Gigante, C.M., Korber, B., Seabolt, M.H., Wilkins, K., Davidson, W., Rao, A.K., Zhao, H., Smith, T.G., Hughes, C.M., Minhaj, F., Waltenburg, M.A., Theiler, J., Smole, S., Gallagher, G.R., Blythe, D., Myers, R., Schulte, J., Stringer, J., Lee, P., Mendoza, R.M., Griffin-Thomas, L.A., Crain, J., Murray, J., Atkinson, A., Gonzalez, A.H., Nash, J., Batra, D., Damon, I., Mcquiston, J., Hutson, C.L., Mccollum, A.M., Li, Y., 2022. Multiple lineages of monkeypox virus detected in the United States, 2021-2022. Science 378, 560–565.

Gohl, D.M., Garbe, J., Grady, P., Daniel, J., Watson, R.H.B., Auch, B., Nelson, A., Yohe, S., Beckman, K.B., 2020. A rapid, cost-effective tailed amplicon method for sequencing SARS-CoV-2. BMC Genom. 21, 863.

Gorgé, O., Jarjaval, F., Nolent, F., Criqui, A., Lourenco, J., Badaoui, A., Khoury, R., Burrel, S.L.O., Chapus, C., Simon-Loriere, E., Tournier, J.N., Ferraris, O., 2022. MinIon-produced Illumina-Confirmed Genomes of the Two First Isolated Cases from France [Online]. Available: https://virological.org/t/minion-produced-illumina-confirmed-genomes-of-the-two-first-isolated-cases-from-france/868. (Accessed 17 June 2022).

Hraib, M., Jouni, S., Albitar, M.M., Alaidi, S., Alshehabi, Z., 2022. The outbreak of monkeypox 2022: an overview. Ann Med Surg 79, 104069.

Huo, S., Chen, Y., Lu, R., Zhang, Z., Zhang, G., Zhao, L., Deng, Y., Wu, C., Tan, W., 2022. Development of two multiplex real-time PCR assays for simultaneous detection and differentiation of monkeypox virus IIa, IIb, and I clades and the B.1 lineage. Biosaf Health 4, 392–398.

Isidro, J., Borges, V., Pinto, M., Sobral, D., Santos, J.D., Nunes, A., Mixao, V., Ferreira, R., Santos, D., Duarte, S., Vieira, L., Borrego, M.J., Nuncio, S., De Carvalho, I.L., Pelerito, A., Cordeiro, R., Gomes, J.P., 2022. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. Nat. Med. 28, 1569–1572.

Ladnyj, I.D., Ziegler, P., Kima, E., 1972. A human infection caused by monkeypox virus in Basankusu Territory, Democratic Republic of the Congo. Bull. World Health Organ. 46, 593–597.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li, J., Du, P., Yang, L., Zhang, J., Song, C., Chen, D., Song, Y., Ding, N., Hua, M., Han, K., Song, R., Xie, W., Chen, Z., Wang, X., Liu, J., Xu, Y., Gao, G., Wang, Q., Pu, L., Di, L., Li, J., Yue, J., Han, J., Zhao, X., Yan, Y., Yu, F., Wu, A.R., Zhang, F., Gao, Y.Q., Huang, Y., Wang, J., Zeng, H., Chen, C., 2022. Two-step fitness selection for intra-host variations in SARS-CoV-2. Cell Rep. 38, 110205.

Liu, Q., Zhao, S., Shi, C.M., Song, S., Zhu, S., Su, Y., Zhao, W., Li, M., Bao, Y., Xue, Y., Chen, H., 2020. Population genetics of SARS-CoV-2: disentangling effects of sampling bias and infection clusters. Dev. Reprod. Biol. 18, 640–647.

Lythgoe, K.A., Hall, M., Ferretti, L., De Cesare, M., Macintyre-Cockett, G., Trebes, A., Andersson, M., Otecko, N., Wise, E.L., Moore, N., Lynch, J., Kidd, S., Cortes, N., Mori, M., Williams, R., Vernet, G., Justice, A., Green, A., Nicholls, S.M., Ansari, M.A., Abeler-Dorner, L., Moore, C.E., Peto, T.E.A., Eyre, D.W., Shaw, R., Simmonds, P., Buck, D., Todd, J.A., , Oxford Virus Sequencing Analysis, G., Connor, T.R., Ashraf, S., Da Silva Filipe, A., Shepherd, J., Thomson, E.C., Consortium, C.-G.U., Bonsall, D., Fraser, C., Golubchik, T., 2021. SARS-CoV-2 within-host diversity and transmission. Science 372, eabg0821.

Martínez-Puchol, S., Coello, A., Bordoy, A.E., Soler, L., Panisello, D., González-Gómez, S., Clarà, G., León, A.P.D., Not, A., Hernández, Á., Bofill-Mas, S., Saludes, V., Martró, E., Cardona, P.J., 2022. Spanish Draft Genome Sequence of Monkeypox Virus Related to Multi-Country Outbreak (May 2022) [Online]. Available: https://virological.org/t/spanish-draft-genome-sequence-of-monkeypox-virus-related-to-multi-country-outbreak-may-2022/825. (Accessed 27 May 2022).

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., Depristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.

Monzón, S., Varona, S., Negredo, A., Patiño-Galindo, J.A., Vidal-Freire, S., Zaballos, A., Orviz, E., Ayerdi, O., Muñoz-García, A., Delgado-Iribarren, A., Estrada, V., García, C., Molero, F., Sánchez, P., Torres, M., Vázquez, A., Galán, J.-C., Torres, I., Causse Del Río, M., Merino, L., López, M., Galar, A., Cardeñoso, L., Gutiérrez, A., Loras, C., Escribano, I., Alvarez-Argüelles, M.E., Del Río, L., Simón, M., Meléndez, M.A., Camacho, J., Herrero, L., Sancho, P.J., Navarro-Rico, M.L., Kuhn, J.H., Sanchez-Lockhart, M., Paola, N.D., Kugelman, J.R., Giannetti, E., Guerra, S., García-Sastre, A., Palacios, G., Cuesta, I., Sánchez-Seco, M.P., 2022. Changes in a new type of genomic accordion may open the pallets to increased monkeypox transmissibility. bioRxiv. https://doi.org/10.1101/2022.09.30.510261, 2022.2009.2030.510261.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S.J., Diekhans, M., Logsdon, G.A., Alonge, M., Antonarakis, S.E., Borchers, M., Bouffard, G.G., Brooks, S.Y., Caldas, G.V., Chen, N.C., Cheng, H., Chin, C.S., Chow, W., De Lima, L.G., Dishuck, P.C., Durbin, R., Dvorkina, T., Fiddes, I.T., Formenti, G., Fulton, R.S., Fungtammasan, A., Garrison, E., Grady, P.G.S., Graves-Lindsay, T.A., Hall, I.M., Hansen, N.F., Hartley, G.A., Haukness, M., Howe, K., Hunkapiller, M.W., Jain, C., Jain, M., Jarvis, E.D., Kerpedjiev, P., Kirsche, M., Kolmogorov, M., Korlach, J., Kremitzki, M., Li, H., Maduro, V.V., Marschall, T., Mccartney, A.M., Mcdaniel, J., Miller, D.E., Mullikin, J.C., Myers, E.W., Olson, N.D., Paten, B., Peluso, P., Pevzner, P.A., Porubsky, D., Potapova, T., Rogaev, E.I., Rosenfeld, J.A., Salzberg, S.L., Schneider, V.A., Sedlazeck, F.J., Shafin, K., Shew, C.J., Shumate, A., Sims, Y., Smit, A.F.A., Soto, D.C., Sovic, I., Storer, J.M., Streets, A., Sullivan, B.A., Thibaud-Nissen, F., Torrance, J., Wagner, J., Walenz, B.P., Wenger, A., Wood, J.M.D., Xiao, C., Yan, S.M., Young, A.C., Zarate, S., Surti, U., Mccoy, R.C., Dennis, M.Y., Alexandrov, I.A., Gerton, J.L., O'neill, R.J., Timp, W., Zook, J.M., Schatz, M.C., Eichler, E.E., Miga, K.H., Phillippy, A.M., 2022. The complete sequence of a human genome. Science 376, 44–53.

Pathak, A.K., Mishra, G.P., Uppili, B., Walia, S., Fatihi, S., Abbas, T., Banu, S., Ghosh, A., Kanampalliwar, A., Jha, A., Fatma, S., Aggarwal, S., Dhar, M.S., Marwal, R., Radhakrishnan, V.S., Ponnusamy, K., Kabra, S., Rakshit, P., Bhoyar, R.C., Jain, A., Divakar, M.K., Imran, M., Faruq, M., Sowpati, D.T., Thukral, L., Raghav, S.K., Mukerji, M., 2022. Spatio-temporal dynamics of intra-host variability in SARS-CoV-2 genomes. Nucleic Acids Res. 50, 1551–1561.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26.

Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., Tse, T., Wang, J., Williams, R., Trawick, B.W., Pruitt, K.D., Sherry, S.T., 2021. Database resources of the national center for biotechnology information. Nucleic Acids Res. 50, D10–D17.

Shchelkunov, S.N., Totmenin, A.V., Safronov, P.F., Mikheev, M.V., Gutorov, V.V., Ryazankina, O.I., Petrov, N.A., Babkin, I.V., Uvarova, E.A., Sandakhchiev, L.S., Sisler, J.R., Esposito, J.J., Damon, I.K., Jahrling, P.B., Moss, B., 2002. Analysis of the monkeypox virus genome. Virology 297, 172–194.

Shu, Y.L., Mccauley, J., 2017. GISAID: global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 22, 2–4.

Tan, W., Gao, F.G., 2022. Perspectives: neglected zoonotic monkeypox in Africa but now back in the spotlight worldwide. China CDC Weekly 4, 2.

Wang, Y., Wang, D., Zhang, L., Sun, W., Zhang, Z., Chen, W., Zhu, A., Huang, Y., Xiao, F., Yao, J., Gan, M., Li, F., Luo, L., Huang, X., Zhang, Y., Wong, S.S., Cheng, X., Ji, J., Ou, Z., Xiao, M., Li, M., Li, J., Ren, P., Deng, Z., Zhong, H., Xu, X., Song, T., Mok, C.K.P., Peiris, M., Zhong, N., Zhao, J., Li, Y., Li, J., Zhao, J., 2021. Intra-host variation and evolutionary dynamics of SARS-CoV-2 populations in COVID-19 patients. Genome Med. 13, 30.

Wu, C., Cui, L., Pan, Y., Lv, Z., Yao, M., Wang, W., Ye, F., Huo, W., Zhao, L., Huang, B., Zhu, F., Lu, R., Deng, Y., Wang, W., Tan, W., 2023. Characterization of whole genomes from recently emerging Mpox cases in several regions of China, 2023 Sci. China Life Sci.. https://doi.org/10.1007/s11427-023-2485-8.

Wu, F., Oghuan, J., Gitter, A., Mena, K.D., Brown, E.L., 2023. Wide mismatches in the sequences of primers and probes for monkeypox virus diagnostic assays. J. Med. Virol. 95, e28395.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.

Yeh, T.Y., Hsieh, Z.Y., Feehley, M.C., Feehley, P.J., Contreras, G.P., Su, Y.C., Hsieh, S.L., Lewis, D.A., 2022. Recombination shapes 2022 monkeypox outbreak. Med. 3, 824–826.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.-Y., 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8, 28–36.

Zhao, H., Wang, W., Zhao, L., Ye, S., Song, J., Lu, R., Zong, H., Wu, C., Huang, W., Huang, B., Deng, Y., A, R, Xie, W., Qi, L., Xu, W., Lin, H., Tan, W., 2022. The first imported case of monkeypox in the mainland of China — Chongqing Municipality, China, September 16, 2022. China CDC Weekly 4, 853–854.