Received: 6 November 2023 | Accepted: 5 April 2024

DOI: 10.1002/jmv.29610

RESEARCH ARTICLE

JOURNAL OF MEDICAL VIROLOGY WILEY

Nanopore adaptive sampling of a metagenomic sample derived from a human monkeypox case

Charlotte Hewel[1] | Hanno Schmidt[2,3] | Stefan Runkel[2,4] | Wolfgang Kohnen[2,5] | Susann Schweiger-Seemann[1,2] | André Michel[2,6] | Sven-Ernö Bikar[2,7] | Bettina Lieb[7] | Bodo Plachter[2,3] | Thomas Hankeln[2,8] | Matthias Linke[1,2] | Susanne Gerber[1,2]

[1]Institute of Human Genetics, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

[2]SARS-CoV-2 Sequencing Consortium Mainz, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

[3]Institute for Virology and Research Center for Immunotherapy (FZI), University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

[4]Transfusion Unit & Test Center, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

[5]Department of Hygiene and Infection Prevention, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

[6]Medical Management Department, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

[7]StarSEQ GmbH, Mainz, Germany

[8]Faculty of Biology, Institute of Organismic and Molecular Evolution, Molecular Genetics & Genome Analysis, Johannes Gutenberg University of Mainz, Mainz, Germany

Correspondence
Charlotte Hewel and Matthias Linke, Institute of Human Genetics, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany.
Email: chhewel@uni-mainz.de and matthias.linke@unimedizin-mainz.de

Abstract

In 2022, a series of human monkeypox cases in multiple countries led to the largest and most widespread outbreak outside the known endemic areas. Setup of proper genomic surveillance is of utmost importance to control such outbreaks. To this end, we performed Nanopore (PromethION P24) and Illumina (NextSeq. 2000) Whole Genome Sequencing (WGS) of a monkeypox sample. Adaptive sampling was applied for in silico depletion of the human host genome, allowing for the enrichment of low abundance viral DNA without a priori knowledge of sample composition. Nanopore sequencing allowed for high viral genome coverage, tracking of sample composition during sequencing, strain determination, and preliminary assessment of mutational pattern. In addition to that, only Nanopore data allowed us to resolve the entire monkeypox virus genome, with respect to two structural variants belonging to the genes OPG015 and OPG208. These SVs in important host range genes seem stable throughout the outbreak and are frequently misassembled and/or misannotated due to the prevalence of short read sequencing or short read first assembly. Ideally, standalone standard Illumina sequencing should not be used for Monkeypox WGS and de novo assembly, since it will obfuscate the structure of the genome, which has an impact on the quality and completeness of the genomes deposited in public databases and thus possibly on the ability to evaluate the complete genetic reason for the host range change of monkeypox in the current pandemic.

KEYWORDS

engineering and technology, evolution, genetic mapping, genetics, mutation/mutation rate, poxvirus, research and analysis methods, virus classification

M. Linke and S. Gerber are joint senior authors.

1 | INTRODUCTION

Monkeypox is a zoonotic disease caused by the enveloped monkeypox virus (MPXV) which contains double-stranded DNA with a genome size of 197 kb harboring around 200 genes. MPXV belongs to the *Orthopoxvirus* genus of the Poxviridae family. Three phylogenetically distinct clades of MPXV are known (I, IIa, IIb). MPXV of the 2022 outbreak belong to Clade IIb,[1] with a single origin potentially in the endemic region of Nigeria.[2] So far, more than 93 000 confirmed cases and over 170 fatalities in more than 100 locations were counted in the 2022 MPXV outbreak.[3] The natural reservoir of this most recent outbreak remains unknown, but MPXV is frequently found among rodents in Africa, even in 120-year-old historic specimen.[4] Human-to-human transmission of MPXV is most prevalent in Clade IIb and occurs primarily via direct contact with infected skin lesions and/or other body fluids and recently contaminated objects.[5,6] Epidemiological and clinical features of the current outbreak indicate it is sexually transmitted, which has not been observed before.[7] The cessation of smallpox vaccinations, providing some cross-protection against monkeypox, has been discussed as one factor that has led to significantly increased case numbers over the last 13 years, culminating in the outbreak of 2022. Taken together, monkeypox is gradually evolving to become of global public health relevance. Surveillance and detection programs are essential tools for understanding the continuously changing epidemiology of this resurging disease.

Integrating high-throughput sequencing metagenomic workflows into pathogen diagnostics has the power to disrupt traditional procedures for infectious diseases since turnaround time is dramatically reduced by waiving the cultivation of suspected pathogens. Moreover, it provides higher taxonomical resolution of microorganism kingdoms, even of those species that cannot be cultivated.[8] PCR has become a gold standard for detection of viruses. Highly specific PCR assays for differentiating closely related viral species are increasingly difficult to maintain due to the mutability of a viral genome and the entirety of human viruses has yet to be revealed. The applicability of clinical metagenomics has been evaluated in viral pathogen detection, demonstrating high potential benefits here as well.[9] Despite the immense potential of high-throughput sequencing metagenomic workflows, there are also crucial limitations. Highly problematic is the high host-to-pathogen nucleic acid fraction in clinical samples with extreme values of more than 99%.[10,11] This unfavorable ratio can be overcome by complex and time-consuming laboratory methods, like differential lysis or enzymatic and immunomagnetic protocols. These solutions in turn can disturb the metagenomic composition of the sample.[12–15] The limited read length of the current sequencing platforms also hampers the assignment of important genes (e.g., antimicrobial resistance) to the respective species and the alignment of reads to repetitive regions and structural variants (SVs, genomic aberrations ≥50 bp) to cover and resolve the whole genome of pathogens.[16] Nanopore sequencing bears the potential to overcome these limitations.[17] Due to its unrivaled read length, it can span complex genomic regions and resolve them unambiguously. Furthermore, it is technically possible to reverse the voltage across nanopores to enable selective sequencing of target fragments, based on real-time assessment of a small initial part of their sequence. This strategy, known as adaptive sampling,[18] enables depletion of unwanted host-derived nucleic acid fragments[19] and avoids rather complex wet lab sample purifications. The high efficiency of human host depletion in clinical samples via this approach has been shown by Marquet et al.[10] Without changing the microbial composition, the authors showed a 1.7-fold increase in sequencing depth for samples with high levels of human DNA contamination. While working with relatively short reads because of a PCR-based sequencing library preparation, the authors outlined that adaptive sampling's decision-making would be more efficient with longer reads, further increasing overall sequencing depth. Using a PCR-free library preparation of a clinical MPXV sample, we aimed here to evaluate the potential of adaptive sampling towards a clinical routine in diagnosing culture-free metagenomic samples.

2 | MATERIALS AND METHODS

2.1 | Ethical statement

Ethical approval was provided by the local Ethics Committee of the Landesaerztekammer Rheinland-Pfalz, Germany in accordance with the Declaration of Helsinki. Data was collected in the course of regular clinical diagnostics and could therefore be used for publication in accordance with internal data protection regulations (University Medical Center of the Johannes Gutenberg University Mainz).

2.2 | Sample acquisition, DNA isolation, and qPCR testing

DNA extraction of a MPXV sample was performed from material of a single pustular lesion using the QIAamp DNA Mini kit, according to the manufacturer's instructions. Extracted DNA was quantified using a Qubit dsDNA HS Assay (Invitrogen) following the manufacturer's protocol and stored at −20°C until use. Initial testing for MPXV nucleic acid via qPCR analysis was positive (Ct value: 21) and performed on a NeuMoDx 288 Molecular System (Molecular Systems).

2.3 | Metagenomic sequencing and bioinformatic processing

2.3.1 | Nanopore runs

Two ligation sequencing libraries were prepared from 1000 ng DNA, one using the Ligation Sequencing Kit 10 (SQK-LSK110) and one using the Ligation Sequencing Kit 14 (SQK-LSK114), respectively (ONT, Oxford Nanopore Technologies Ltd., Oxford, UK) following the manufacturer's protocol. The clean-up step after adapter ligation was

intended to size-select fragments and was done with Short Fragment Buffer. Each library was loaded on a single flow cell (R9.4.1 and R10.4.1) and sequenced on a PromethION 24 device within 72 h. MinKNOW (v22.12.05) was used to supervise the initial sequencing run, including adaptive sampling on half a flow cell with depletion of human DNA by setting human genome build hg38 as input reference, the other half was used as a normal sequencing control. For the second run with Kit 14 the entire flow cell was used for adaptive sampling. NanoComp (v1.12.0; https://github.com/wdecoster/nanocomp) was used to determine effectiveness of the adaptive sequencing by comparing the respective subsets of the *sequencing_summary. txt file. After run completion, samples were re-basecalled and aligned to MPXV reference genome MT903344, while filtering out nonmapping reads using Guppy (v6.1.5) with the SUP model on a NVIDIA A100 DGX-station. All fastq reads were used as input for final species determination using Centrifuge (v1.0.4-beta with the default h + p + v + c index; Song et al. 2016; https://ccb.jhu.edu/software/centrifuge/) and visualized via Pavian (v1.0; https://github.com/fbreitwieser/pavian). The filtered reads were further used to obtain read quality metrics via pomoxis, counterr and NanoComp. Moreover, the filtered reads were re-formatted using bam2fq and used as input for the wf-mpx (v0.0.6; https://github.com/epi2me-labs/wf-mpx) pipeline for initial assessment and generation of a draft consensus with ON563414.3 as reference. For all final assembly related steps, only data from ONTs most recent pore and chemistry (FLO-PRO114M, Kit 14) was used, because read depth was sufficient and Kit 14 has superior read quality in homopolymer stretches, leading to better analysis results. Then, to match the Illumina assembly, the Kit 14 data was realigned against NC_063383.1 and the reads were rebasecalled with Dorado 0.3.2. Subsequently medaka and sniffles2 were used to call variants and a consensus was generated with bcftools. Final differences between the assemblies were assessed with compare_assemblies. py and IGV. Raw reads were realigned to the final assembly to spot check for coverage deviations. Initial lineage assignment was performed via NextClade (v2.12.0; https://clades.nextstrain.org).

3 | OTHER DATA SETS

For annotation evidence and comparison on RNA level *fastq and *bam file data was downloaded from PRJEB56841. Realignment for the RNASeq reads to the final ONT assembly was performed via the original mapping parameters -ax splice -Y -C5 --cs --MD -un -G 10000 using minimap2. DNA level data for additional monkeypox samples was also obtained from SRA and the accession numbers can be found in Supplement1-4_MPX. xlsx.

3.1 | Illumina run

Automated purification of gDNA from the monkeypox isolate was performed using the Maxwell® CSC 48 instrument with the Maxwell® RSC Blood DNA Kit according to Maxwell® Application Note PA473 (Promega).

After NGS library preparation according to the NEBNext Ultra II FS DNA Library Prep Kit protocol for Illumina (NEB, New England Biolabs) 2 × 150 bp paired-end sequencing was performed on the NextSeq. 2000 platform (Illumina, ILM) using P3 reagents. The genome sequence was reconstructed from 151 mio demultiplexed FASTQ reads using the Robert Koch Institute analysis pipeline nCoV_minipipe aka CovPipe v3.1.0 (https://gitlab.com/RKIBioinformaticsPipelines/ncov_minipipe), initially optimized for SARS-CoV-2 data and applicable to other virus data. Clipping of adapters was activated and reads below a Phred score of 20 or a length of 30 nt were removed. Reads were mapped to the clade B1 monkeypox reference genome (MPXV_USA_2022_MA001, ON563414) and variants with a total coverage of at least 20 reads and a support of at least 10 reads and 90% of the total reads were detected. Generated consensus sequence positions below a coverage of 20 reads were masked and nonunique positions were represented according to IUPAC nomenclature.

4 | RESULTS

4.1 | Efficiency of adaptive sampling and lineage assignment of MPXV genome

To establish third generation metagenomic analysis in a sequencing-based surveillance consortium and to characterize a monkeypox sample from the 2022 outbreak, material from a skin lesion swap was prepared and sequenced via ONT and ILM technologies. For the ONT samples, two single flow cells using Kit 10 and Kit 14 were loaded on a P24. During the run, adaptive sampling was activated for 1/2 half of the flow cell using Kit 10 and for the whole flow cell using Kit 14.

The nanopore runs generated 72.31 Gb (Kit 10) and 35.75 Gb (Kit 14) of data after 72 h of sequencing, consisting of 83.04 M and 22.32 M reads (see Figure S1). As evaluated by MinKNOW, N50 for a completed read was at 1.49 kb and 2.22 kb. For the Illumina run 151 M reads with a length of 150 bp were generated. The average coverage of NC_063383.1 was 821× and 521× for ONT and 159× for ILM.

Run time with the Illumina settings used in this study is around 48 h. The sequence quality and general error pattern was assessed with pomoxis and counterr (see Figures S2 and S3). For Nanopore reads especially in longer homopolymeric regions, errors become frequent and deletions start to be the predominant type of artifacts. Kit 14 chemistry and duplex basecalling show improvements with regard to this known issue. Overall Illumina sequencing has the best raw read and homopolymer quality. ONTs Kit 10 performs worse and in our initial assembly done with wf-mpx there were four unexpected frameshifts that were likely caused by subpar accuracy compared to the newer nanopore kit and to Illumina sequencing. However, Kit 14 and the Illumina sequencing did not show these unexpected frameshifts. Furthermore, both nanopore kits performed better at resolving long repetitive stretches of the genome as can be seen by the unequal number of SVs reported between ONT (2) and Illumina (0).

Metagenomic composition was queried throughout the first nanopore (Kit 10) run using Centrifuge. In fact, MPXV could be detected among the first 4000 reads that were generated. After rebasecalling, 99.6% of reads mapped to chordate sequences, and the remaining reads mostly to the MPXV reference (Figure 1, Table S2).

Adaptive sampling was active on half of the flow cell for the entire duration of the Kit 10 nanopore run. Adaptive sampling is not available for Illumina devices. The rationale was to deplete reads originating from human DNA and thereby to passively increase coverage on the MPXV target. The N50 of an adaptively sampled read was evaluated at 744 bp by MinKNOW. Compared to the other half of the flowcell, the enrichment via adaptive sampling was approximately 2× for reads mapping to the MPXV reference (Figure 2).

Lineage assignment for both the ONT and Illumina consensus was performed by NextClade and revealed that the MPXV sample belongs to lineage B.1 and Clade IIb, frequently observed during the 2022 outbreak (Table S3).

Three notable nucleotide substitutions remained: C101074T, G138682A, C189167T, of which only G138682A leads to an amino acid substitution OPG156:S243F.

The most striking nucleotide difference between the MPXV genome in this study and the MPXV reference were the three SVs that could only be resolved with long read sequencing (see Figure 3 and Table S4).

Two of them are mirror copies of each other and located within the Inverted Terminal Repeat (ITR) regions of the monkeypox genome. The SVs are 128 bp long in relation to NC_063383.1 and mostly contain copies of the AACTAACTTATGACTT repeat unit, which brings up the total copies of this repeat from 7 to 13. The last SV starts at position 179077, has a length of 324 bp and is composed of CATTATATA repeats. That equals 35 additional copies of the repeat bringing the total copy number from $n = 18$ to $n = 53$ repeats in the genome sequenced in this study.

Next, we reannotated the SV parts specifically by downloading and realigning publicly available long read RNAseq data from a bio sample isolated during the recent outbreak in Prague.[20] This confirmed the expression of the two 128 bp SVs as a part of OPG015 and the expression of the 324 bp SV as a part of OPG208 (Figure 4).

This highlights the importance of multiomic datasets and periodic re-sequencing to generate and maintain the highest level of genome annotation.

FIGURE 1 Sankey plot of metagenomic reads after exclusion of reads below the quality threshold and exclusion of human reads (Kit 10 data).

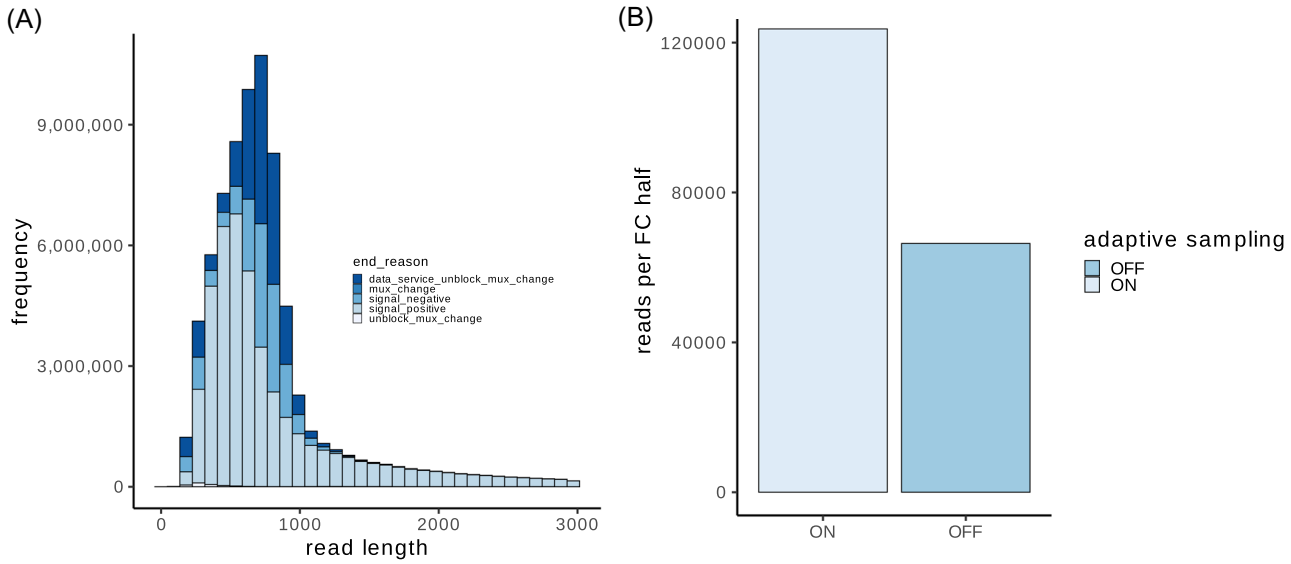

FIGURE 2 Effect of adaptive sampling decisions on half of the PromethION flow cell. (A) Read end decision as evaluated by MinKNOW. (B) Number of reads mapping against the MPVX reference split by flow cell half (light blue: adaptive sampling = ON, darker blue: adaptive sampling = OFF).

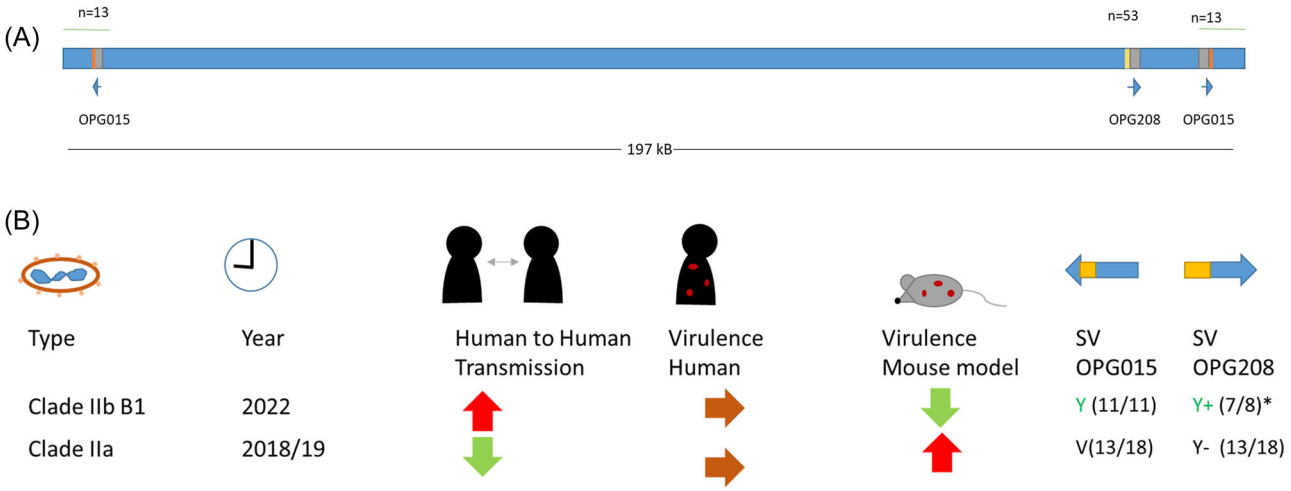

FIGURE 3 Variation in the SVs overlapping OPG015 and OPG208 (A) Schematic representation of the monkeypox genome and structural variations that were sequenced in this study. Gray boxes are the approximate locations of OPG015 and OPG208. The blue arrows represent the direction of the genes and the orange and yellow boxes mark the relative position of the SV at either the end or start of the gene. *N* = 13 means the new assembly has 13 copies of the AACTAACTTATGACTT repeat, *n* = 53 means the new assembly has 53 copies of the CATTATATA repeat. Green horizontal lines mark the ITR. (B) The structural variations and associated metadata of the additional monkeypox data from the SRA. Red arrow means increase, for example the human-to-human transmission in the B1 lineage from 2022 is increased, likewise the green arrow pointing downwards means decreased. The brown arrow pointing to the side means no change. The green Y means the repeat is present and stable the number in the brackets designates the samples, resp. (11/11) means 11 samples were downloaded and 11 samples harbored the repeat expansion. Y+: the repeat is present, mostly stable and always longer than in Clade IIa. * the three UK genomes had to be excluded from this one SV due to library prep issues. V: repeat copies in a higher number than the reference genome have been found but the repeat count was very variable. Y-: the OPG208 elongation was also found for clade IIa genomes, but it was always shorter than in the B1 lineage. 5 of the 18 genomes could not be assessed due to data quality.

Then we were curious how conserved these SVs were in the monkeypox clades, starting with the B1 lineage reference. As can be seen from Figure 5, there seem to be repeat copy number deviations in OPG015 and OPG208 that are not reflected in the final reference assembly/annotation (ON563414.3), but the read length made it hard to fully resolve them. Thus, we focused entirely on available long read DNA sequencing for the rest of the comparison (see Figure 3B). All samples downloaded–where data quality permitted SV estimation–show insertions overlapping the SVs compared to the reference NC_063383.1. If we further divide the samples into Clade

FIGURE 4 RNAseq overlapping the SVs close to OPG015 and OPG208 to provide transcript level evidence for the gene annotations (A) RNA data aligned to ON563414, showing an insertion in the transcripts from OPG015, which corresponds to SV1. (B) RNA data for OPG208, which shows an insertion corresponding to SV2.

IIa and Clade IIb B1, it seems that SV1 OPG015 varied a lot more in Clade IIa, whereas the length seems rather conserved at around 128 bp in all recent outbreak genomes that could be analyzed. SV2 in OPG208 is always shorter in preoutbreak genomes. The samples from the current outbreak are the only ones, where lengths of 300+ bp have been recorded, meaning the SV has apparently grown in size to 50+ repeat copies between the 2022 outbreak and the one before.

5 | DISCUSSION

The present study seeked to characterize the performance of metagenomic nanopore sequencing on a skin lesion sample from a patient presenting with monkeypox. The sample had a qPCR Ct of 21 and consequently a high viral load, thus the conclusions for ultra-low and fragmented coverage scenarios are limited. Host depletion via adaptive sampling led to a successful, albeit moderate, enrichment of ca. 2X for the MPXV fraction. Metagenomic composition revealed that this enrichment of MPXV sequences was achieved despite a massive excess of human DNA (>99.5%), exceeding previously reported ratios.[10] In general, depletion should result in a significantly more favorable human host to pathogen nucleic acid ratio, which would of course increase the taxonomical resolution of micro-organism and speed up clinical management of acute infections. A variety of ONT's adaptive sampling strategies have been previously demonstrated[21] and could be applied for human host depletion in metagenomic studies.[10,19] Working with clinical samples characterized by high amounts of host DNA, Marquet et al.[10] demonstrated a 1.70-fold increase in total sequencing depth after human host depletion. Cheng et al.[19] could increase microbial sequence yield at least eightfold while adapting human host depletion in another clinical setting. In principle, the method works well, but is not without limitations, such as increased pore blockage. Another known issue of human host depletion via ONT's adaptive sampling option in MinKNOW is the usage of its general-purpose alignment program minimap2 (Li et al. 2018). Previous studies showed that around 25% of human reads could not be rejected accurately by this software, thus wasting sequencing resources.[10,21] Addressing this problem,

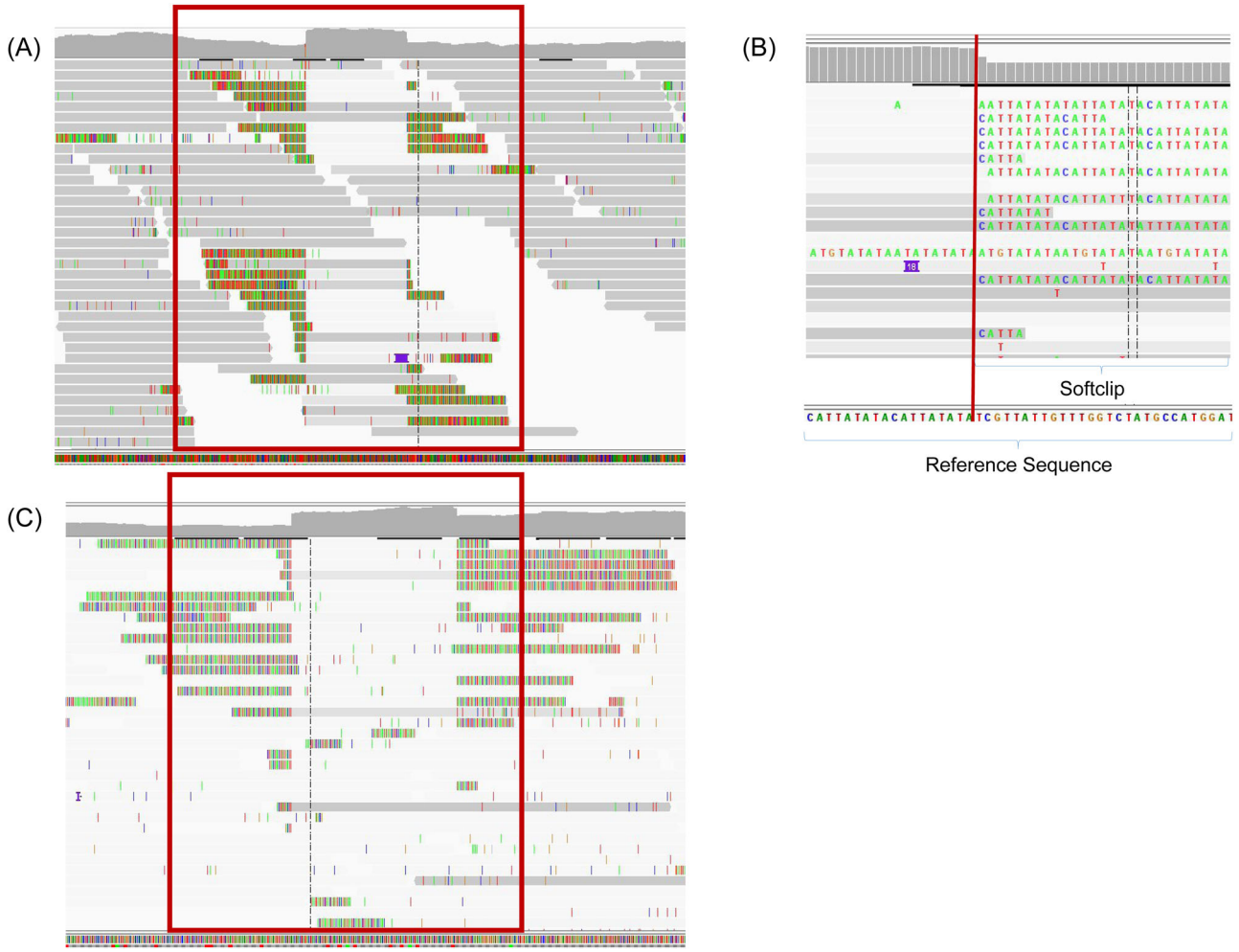

FIGURE 5 Realigned reads from SRR20698670 against their assembly ON563414.3 (A) the region containing SV2 and OPG208 is shown, (B) Zoom in to the right side of Figure 5A, on the top is the gray coverage bar, below the reads, which are colored gray when they match the reference and colored in individual basepairs when they stop matching the reference, we see that the reads have CATTATATA repeats below the reference sequence (C) the region overlapping OPG015 and SV1 is zoomed in. In both cases we see a sharp coverage spike and partially softclipped reads. The softclips are matching the respective repeat units, whose copies give rise to the SVs.

ReadBouncer was introduced by Ulrich, Lutfi, Rutzen and Renard[22] as an improved classification tool. It shows a higher read sensitivity than other state-of-the-art classification tools for adaptive sampling, while retaining a high specificity. A different potential limitation is the fixed target reference, i.e., the adaptive sampling reference is chosen at the beginning and stays constant during the run. The tool BOSS-RUNS, a further development in this field, can generate dynamically updated decision strategies where coverage is tracked throughout the run and the adaptive sampling target list is kept updated based on the coverage resp. species abundance.[23] Lastly, adaptive sampling is also reliant on read length and increasing the read length will improve abundance of target species.[11] Both approaches, improvements in the laboratory workflow leading to longer reads, as well as the afore mentioned and upcoming bioinformatic adaptations will make adaptive sampling even more attractive in the future.

One of the complications that afflict a subset of monkeypox patients is bacterial superinfection of the skin lesions.[24] Metagenomic sampling with real-time species determination has the advantage to give an overview of bacterial and viral sample composition within minutes after sequencing starts, thus either confirming or refuting the presence of the target pathogen and/or additional pathogens. In this study, we could confirm MPXV as the predominant pathogen in data generated already 10 min after sequencing started.

Unlike RNA viruses, the double-stranded DNA of orthopoxviruses is very stable and their DNA polymerase has proofreading exonuclease activity, resulting in a low mutation rate. MPXV of clade IIb most probably originated from clade II in a 2017 outbreak in Nigeria.[2] Phylogenetically, these two clades differ by a mean of 50 SNPs, which is roughly 6–12-fold more than previous estimations of the substitution rate for Orthopoxviruses (1–2 substitutions per genome per year).[2,25] In the present study, we can confirm that the underlying MPXV strain belongs to clade IIb and thus confirm the already known genome sequence differences between clade II and IIb.[2] Clade IIb exhibits molecular signatures that can been associated with the

potential action of apolipoprotein B mRNA-editing catalytic polypeptide-like 3 (APOBEC3) enzymes via viral genome editing.[26] Human APOBEC3 proteins introduce GA > AA and TC > TT replacements into viral genomes and are thus part of the cellular defense mechanisms.[27] Mutation patterns of this type are indicative of a large amount of human-to-human transmission. With respect to the ITRs at the ends of the MPXV genomic sequence, it has been previously reported that only long read Nanopore reads were able to cover the ITRs completely.[28] Since third generation sequencing data has grown in recent years, insertions corresponding to SV1 of this study have frequently been found in the genomes of the 2022 outbreak. Due to lack of data collection and data processing standardization, the true frequency remains likely underestimated. This calls for caution when re-analyzing publicly deposited MPXV genome assemblies for repeat content and repeat expansion pattern, since we found this class of genomic event seems to cluster more with regard to sequencing technology and assembly method rather than true repeat content of the genome. Moreover, this type of SV has been collectively described as intergenic.[29,30] In contrast, we believe there is probable cause that this may be an artifact due to the main annotation method chosen, which seems to be an annotation transfer of gene features from a genome without the SV and therefore re-mapping of a shorter gene model sans SV. Overlapping expression data from the current outbreak, shows SV1 expressed as part of OPG015.[20] OPG015 (or D1L) is frequently implicated as a gene participating in immune invasion and influencing monkeypox host range.[31,32] Regarding the CATTATATA SV overlapping OPG208, Chen et al.[33] were the first to describe a variation in repeat content between 37 copies in WRAIR-61, 27 copies in COP-58 and 16 in ZAI-96, although the functional effect of the 5' putatively untranslated mRNA remained unclear. In contrast to this, Monzón et al. and this study have found 53+ copies for recent genomes. OPG208 (or SPI-1) is a key host range factor that enables replication in human cells, which has been proven in loss of function, gain of function and related experiments.[34,35]

One remaining key question is how the variations mentioned above affect the MPXV transmissibility, pathogenicity, and adaption to the human host via functional studies and re-annotation.

Conclusively, metagenomic sequencing of a MPXV sample via adaptive sampling on a high-throughput PromethION-type flow cell allows coverage of the target species to be increased in addition to rapid, real-time metagenomic classification and lineage determination.

ACKNOWLEDGMENTS

Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study is available upon request. The data It is currently being processed to be in accordance with public databases.

ORCID

Matthias Linke http://orcid.org/0000-0002-7494-9812

REFERENCES

1. Gigante CM, Korber B, Seabolt MH, et al. Multiple lineages of monkeypox virus detected in the United States, 2021-2022. *Science*. 2022;378(6619):560-565.
2. Isidro J, Borges V, Pinto M, et al. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. *Nat Med*. 2022;28(8):1569-1572. doi:10.1038/s41591-022-01907-y
3. Mathieu E, Spooner F, Dattani S, Ritchie H, Roser M 2022. Mpox (monkeypox). Published online at OurWorldInData.org. Retrieved on March 16th 2024 from: https://ourworldindata.org/monkeypox
4. Tiee MS, Harrigan RJ, Thomassen HA, Smith TB. Ghosts of infections past: using archival samples to understand a century of monkeypox virus prevalence among host communities across space and time. *R Soc Open Sci*. 2018;5:171089.
5. Fleischauer AT, Kile JC, Davidson M, et al. Evaluation of human-to-human transmission of monkeypox from infected patients to health care workers. *Clin Infect Dis*. 2005;40(5):689-694. doi:10.1086/427805
6. Hutson CL, Gallardo-Romero N, Carroll DS, et al. Transmissibility of the monkeypox virus clades via respiratory transmission: investigation using the prairie dog-monkeypox virus challenge system. *PLoS One*. 2013;8(2):e55488. doi:10.1371/journal.pone.0055488
7. Bragazzi NL, Kong JD, Wu J. Is monkeypox a new, emerging sexually transmitted disease? A rapid review of the literature. *J Med Virol*. 2023;95(1):e28145. doi:10.1002/jmv.28145
8. Simner PJ, Miller S, Carroll KC. Understanding the promises and hurdles of metagenomic next-generation sequencing as a diagnostic tool for infectious diseases. *Clin Infect Dis*. 2018;66:778-788.
9. Kiselev D, Matsvay A, Abramov I, Dedkov V, Shipulin G, Khafizov K. Current trends in diagnostics of viral infections of unknown etiology. *Viruses*. 2020;12(2):211. doi:10.3390/v12020211
10. Marquet M, Zöllkau J, Pastuschek J, et al. Evaluation of microbiome enrichment and host DNA depletion in human vaginal samples using Oxford nanopore's adaptive sequencing. *Sci Rep*. 2022;12(1):4000.
11. Martin S, Heavens D, Lan Y, Horsfield S, Clark MD, Leggett RM. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biol*. 2022;23(1):11.
12. Ferretti P, Farina S, Cristofolini M, Girolomoni G, Tett A, Segata N. Experimental metagenomics and ribosomal profiling of the human skin microbiome. *Exp Dermatol*. 2017;26:211-219.
13. Horz H-P, Scheer S, Huenger F, Vianna ME, Conrads G. Selective isolation of bacterial DNA from human clinical specimens. *J Microbiol Meth*. 2008;72:98-102.
14. Liu G, Weston CQ, Pham LK, et al. Epigenetic segregation of microbial genomes from complex samples using restriction endonucleases hpaii and McrB. *PLoS One*. 2016;11:e0146064.
15. Oechslin CP, Lenz N, Liechti N, et al. Limited correlation of shotgun metagenomics following host depletion and routine diagnostics for viruses and bacteria in low concentrated surrogate and clinical samples. *Front Cell Infect Microbiol*. 2018;8:375.
16. Yang Y, Che Y, Liu L, et al. Rapid absolute quantification of pathogens and ARGs by nanopore sequencing. *Sci Total Environ*. 2022;809:152190.
17. Liu L, Yang Y, Deng Y, Zhang T. Nanopore long-read-only metagenomics enables complete and high-quality genome reconstruction from mock and complex metagenomes. *Microbiome*. 2022;10(1):209.
18. Loose M, Malla S, Stout M. Real-time selective sequencing using nanopore technology. *Nat Methods*. 2016;13:751-754.
19. Cheng H, Sun Y, Yang Q, et al. A rapid bacterial pathogen and antimicrobial resistance diagnosis workflow using Oxford nanopore adaptive sequencing method. *Brief Bioinform*. 2022;23(6):bbac453.

20. Kakuk B, Dörmő Á, Csabai Z, et al. In-depth temporal transcriptome profiling of monkeypox and host cells using nanopore sequencing. *Sci Data*. 2023;10(1):262.
21. Payne A, Holmes N, Clarke T, Munro R, Debebe BJ, Loose M. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nature Biotechnol*. 2021;39(4):442-450. doi:10.1038/s41587-020-00746-x
22. Ulrich JU, Lutfi A, Rutzen K, Renard BY. ReadBouncer: precise and scalable adaptive sampling for nanopore sequencing. *Bioinformatics*. 2022;38(suppl 1):i153-i160.
23. Weilguny L, De Maio N, Munro R, et al. Dynamic, adaptive sampling during nanopore sequencing using Bayesian experimental design. *Nat Biotechnol*. 2023;41:1018-1025.
24. de Sousa D, Frade J, Patrocínio J, Borges-Costa J, Filipe P. Monkeypox infection and bacterial cellulitis: a complication to look for. *Int J Infect Dis*. 2022;123:180-182.
25. Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. Using timestructured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol Biol Evol*. 2010;27(9):2038-2051.
26. Pecori R, Di Giorgio S, Paulo Lorenzo J, Nina Papavasiliou F. Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination. *Nat Rev Genet*. 2022;23(8):505-518. doi:10.1038/s41576-022-00459-8
27. Sadeghpour S, Khodaee S, Rahnama M, Rahimi H, Ebrahimi D. Human APOBEC3 variations and viral infection. *Viruses*. 2021;13(7):1366.
28. Vandenbogaert M, Kwasiborski A, Gonofio E, et al. Nanopore sequencing of a monkeypox virus strain isolated from a pustular lesion in the Central African Republic. *Sci Rep*. 2022;12(1):10768.
29. Desingu PA, Nagarajan K, Sundaresan NR. Unique tandem repeats in the inverted terminal repeat regions of monkeypox viruses. microbiology. *Spectrum*. 2023;11(2):e03199-22.
30. Monzon S, Varona S, Negredo A, et al. Changes in a new type of genomic accordion may open the pallets to increased monkeypox transmissibility. *bioRxiv*. 2022;20:09. doi:10.1101/2022.09.30.510261
31. Bratke KA, McLysaght A, Rothenburg S. A survey of host range genes in poxvirus genomes. *Infect Genet Evol*. 2013;14:406-425.
32. Lum FM, Torres-Ruesta A, Tay MZ, et al. Monkeypox: disease epidemiology, host immunity and clinical interventions. *Nat Rev Immunol*. 2022;22(10):597-613.
33. Chen N, Li G, Liszewski MK, et al. Virulence differences between monkeypox virus isolates from West Africa and the Congo basin. *Virology*. 2005;340(1):46-63.
34. Liu R, Mendez-Rios JD, Peng C, et al. SPI-1 is a missing host-range factor required for replication of the attenuated modified vaccinia Ankara (MVA) vaccine vector in human cells. *PLoS Pathog*. 2019;15(5):e1007710.
35. Zhu J, Gao X, Li Y, et al. Human FAM111A inhibits vaccinia virus replication by degrading viral protein I3 and is antagonized by poxvirus host range factor SPI-1. *Proc Nat Acad Sci*. 2023;120(35):e2304242120.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hewel C, Schmidt H, Runkel S, et al. Nanopore adaptive sampling of a metagenomic sample derived from a human monkeypox case. *J Med Virol*. 2024;96:e29610. doi:10.1002/jmv.29610