medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Multi-site validation of an amplicon-based sequencing approach for human monkeypox virus

Nicholas F.G. Chen[1*], Chrispin Chaguza[1*], Luc Gagne[2*], Matthew Doucette[2], Sandra Smole[2], Erika Buzby[2], Joshua Hall[2], Stephanie Ash[2], Rachel Harrington[2], Seana Cofsky[2], Selina Clancy[2], Curtis J. Kapsak[3], Joel Sevinsky[3], Kevin Libuit[3], Daniel J. Park[4], Peera Hemarajata[5], Jacob M. Garrigues[5], Nicole M. Green[5], Sean Sierra-Patev[6], Kristin Carpenter-Azevedo[6], Richard C. Huard[6], Claire Pearson[7], Kutluhan Incekara[7], Christina Nishimura[7], Jian Ping Huang[7], Emily Gagnon[7], Ethan Reever[7], Jafar Razeq[7], Anthony Muyombwe[7], Vítor Borges[8], Rita Ferreira[8], Daniel Sobral[8], Silvia Duarte[9], Daniela Santos[9], Luís Vieira[9], João Paulo Gomes[8,10], Carly Aquino[11], Isabella M. Savino[11], Karinda Felton[11], Moneeb Bajwa[11], Nyjil Hayward[11], Holly Miller[11], Allison Naumann[11], Ria Allman[11], Neel Greer[11], Amary Fall[12], Heba H. Mostafa[12], Martin P. McHugh[13,14], Daniel M. Maloney[13,15], Rebecca Dewar[13], Juliet Kenicer[13], Abby Parker[13], Katharine Mathers[13], Jonathan Wild[13], Seb Cotton[13], Kate E. Templeton[13], George Churchwell[16], Philip A. Lee[16], Maria Pedrosa[16], Brenna McGruder[16], Sarah Schmedes[16], Matthew R. Plumb[17], Xiong Wang[17], Regina Bones Barcellos[18], Fernanda M.S. Godinho[18], Richard Steiner Salvato[18], Aimee Ceniseros[19], Mallery I. Breban[1], Nathan D. Grubaugh[1,20], Glen R. Gallagher[2,6#@], Chantal B.F. Vogels[1#@]

[1] Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, USA
[2] Massachusetts Department of Public Health, Boston, MA, USA
[3] Theiagen Genomics, Highlands Ranch, CO, USA
[4] Broad Institute, Cambridge, MA, USA
[5] Los Angeles County Public Health Laboratories, Downey, CA, USA
[6] Rhode Island Department of Health, Rhode Island State Health Laboratory, Providence, RI, USA
[7] Connecticut Department of Public Health, Hartford, CT, USA
[8] Genomics and Bioinformatics Unit, Department of Infectious Diseases, National Institute of Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal
[9] Technology and Innovation Unit, Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal
[10] Faculty of Veterinary Medicine, Lusófona University, Lisbon, Portugal
[11] Delaware Public Health Laboratory, Smyrna, DE, USA
[12] Johns Hopkins School of Medicine, Baltimore, MD, USA
[13] Viral Genotyping Reference Laboratory Edinburgh, NHS Lothian, Royal Infirmary of Edinburgh, Edinburgh, UK
[14] School of Medicine, University of St Andrews, St Andrews, UK
[15] Institute of Ecology and Evolution, University of Edinburgh, Edinburgh, UK
[16] Florida Department of Health, Bureau of Public Health Laboratories, Jacksonville, FL, USA
[17] Minnesota Department of Health, Public Health Laboratory, St. Paul, MN, USA
[18] Centro Estadual de Vigilância em Saúde, Secretaria Estadual da Saúde do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil
[19] Idaho Bureau of Laboratories, Boise, ID, USA
[20] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

* Co-first authors
# Co-senior authors
@ Corresponding authors

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

The 2022 multi-country monkeypox outbreak concurrent with the ongoing COVID-19 pandemic has further highlighted the need for genomic surveillance and pathogen whole genome sequencing. While metagenomic sequencing approaches have been used to sequence many of the early human monkeypox virus infections, these methods are resource intensive and require samples with high viral DNA concentrations. Given the atypical clinical presentation of cases associated with the current outbreak and uncertainty regarding viral load across both the course of infection and anatomical body sites, there is an urgent need for a more sensitive and broadly applicable sequencing approach. Amplicon-based sequencing (PrimalSeq) was initially developed for sequencing of Zika virus, and later adapted as the main sequencing approach for SARS-CoV-2. Here, we used PrimalScheme to develop a primer scheme for human monkeypox virus that can be used with many sequencing and bioinformatics pipelines implemented during the COVID-19 pandemic. We sequenced clinical samples that tested presumptive positive for monkeypox virus with amplicon-based and metagenomic sequencing approaches. Upon comparison, we found notably higher genome coverage across the virus genome, particularly in higher PCR cycle threshold (lower DNA titer) samples, with minimal amplicon drop-outs, in using the amplicon-based sequencing approach. By sending out primer pool aliquots to laboratories across the United States and internationally, we validated the primer scheme in 12 public health laboratories with their established Illumina or Oxford Nanopore Technologies sequencing workflows and with different sample types across a range of Ct values. Our findings suggest that amplicon-based sequencing increases the success rate across a wider range of viral DNA concentrations, with the PCR Ct value threshold at which laboratories were able to achieve 80% genome coverage at 10X ranging between Ct 25-33. Therefore, it increases the number of samples where near-complete genomes can be generated and it provides a cost-effective and widely applicable alternative to metagenomics for continued human monkeypox virus genomic surveillance. Importantly, we show that the human monkeypox virus primer scheme can be used with currently implemented amplicon-based SARS-CoV-2 sequencing workflows, with minimal change to the protocol.

Introduction

The integration of pathogen whole genome sequencing with public health surveillance provides a powerful tool to inform outbreak control [1,2]. While the feasibility of real-time genomic surveillance was demonstrated during the 2013-2016 Ebola outbreak [3], the SARS-CoV-2 pandemic has launched a revolution in viral genomics [4]. To date, more than 13 million SARS-CoV-2 genomes across 103 countries have been sequenced and shared publicly [5], furthering our understanding of viral transmission and evolution. As the COVID-19 pandemic remains ongoing, the recent spread of human monkeypox virus outside of endemic areas has provided a new target for genomic surveillance efforts [6].

Monkeypox is a zoonotic DNA virus of the *Orthopox* genus endemic to Western and Central Africa [7]. The virus consists of 3 major clades (I, IIa, and IIb), with a subgroup of clade IIb being referred to as human monkeypox virus due to direct transmission from human to human [8]. Initially, rare outside endemic countries beyond imported cases, human monkeypox virus has emerged as a global threat. It was first detected in the United Kingdom on May 7, 2022, quickly spreading to other continents through travel-related infections and sometimes unknown transmission chains [9]. As of October 3, 2022, 67,739 cases of monkeypox across 99 non-endemic countries have been reported, often with atypical clinical presentations [6,10]. The lack of consistent clinical presentations, unknown transmission dynamics, and uncertainty surrounding animal reservoirs highlight the importance of establishing rapid genomic surveillance networks.

Much of the early human monkeypox virus sequencing during the 2022 global outbreak has been accomplished via a metagenomics approach [9], which employs sequencing of the total nucleic acid present in a sample. While considered the gold standard for untargeted sequencing, metagenomics incurs a high

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

resource cost, requires experienced sequencing and bioinformatics teams, is computationally demanding, and relies on samples with high viral concentrations relative to background nucleic acids [11]. The initial metagenomic approaches included hybrid assemblies of both long and short read sequencing to allow for high confidence and polishing of the early genomes. Although this was important to establish reference genomes, it comes at the cost of being able to sequence larger numbers of specimens [12]. Highly multiplexed amplicon-based sequencing (PrimalSeq) was initially developed to generate greater coverage and depth for Zika virus genomes [13], and was later adapted as the main sequencing approach during the SARS-CoV-2 pandemic for both Illumina and Oxford Nanopore Technologies sequencing platforms [14]. Here, we developed a primer scheme for human monkeypox virus using PrimalScheme for use with amplicon-based sequencing workflows widely implemented during the COVID-19 pandemic. We validated the primer scheme in 12 public health laboratories, and showed that the amplicon-based sequencing approach can be implemented in different Illumina and Oxford Nanopore Technologies sequencing workflows. We show a consistently high percent genome coverage at 10X across a range of PCR cycle-threshold (Ct) values, with different workflows and sequencing platforms. Amplicon-based sequencing provides a more sensitive, lower cost, and higher throughput strategy that can be plugged into currently established genomic infrastructure, which will help to improve genomic surveillance of human monkeypox virus.

Results

In May 2022, a growing cluster of monkeypox cases in humans was reported outside its endemic region [6,10]. Difficulties in obtaining sufficient coverage with metagenomic sequencing approaches led us to develop a primer scheme for use with amplicon-based sequencing approaches. Given that many of the early B.1 outbreak clade genomes had low coverage, we used the closely related pre-outbreak A.1 clade genome (GenBank accession: MT903345) as a reference for the primer scheme. The primer scheme, designed using PrimalScheme [15], consists of 163 primer pairs with an amplicon length ranging between 1597 and 2497 bp (average length of 1977 bp; **Supplementary Table 1**). For the initial validation, we sequenced 10 clinical specimens with a range of PCR Ct values with both amplicon-based and metagenomic sequencing approaches at the Massachusetts State Public Health Laboratory (MASPHL). Clinical specimens ranged in Ct value from 15.0 (highest DNA concentration) to 34.6 (lowest DNA concentration). We found comparable genome coverage between amplicon and metagenomic sequencing with low Ct (<18) samples, but a dramatic increase in genome coverage with amplicon sequencing in higher Ct samples (>18; **Figure 1**). These findings show that amplicon-based sequencing approaches can help to improve coverage of human monkeypox virus genomes from clinical specimens, particularly at higher Ct values.

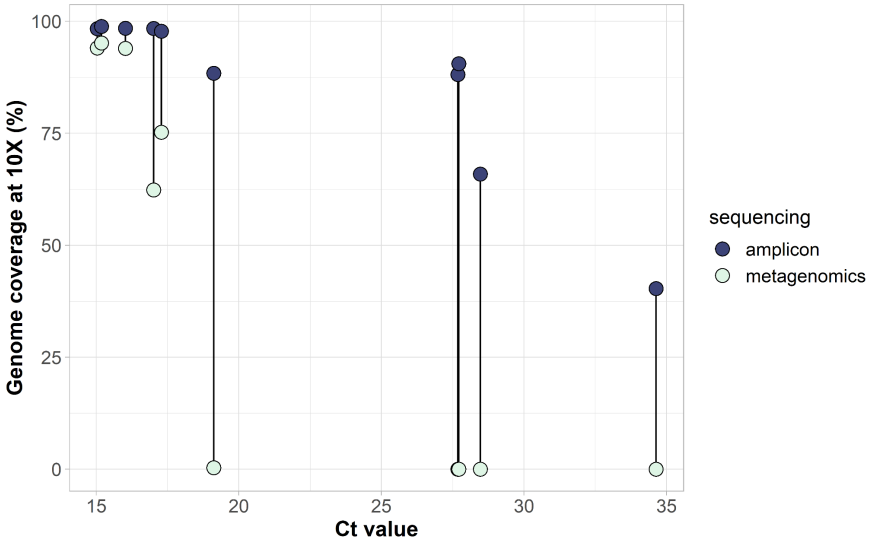


Figure 1: Comparison of percent genome coverage at 10X of clinical specimens sequenced with amplicon-based and metagenomic sequencing approaches. DNA was extracted from 10 clinical samples manually extracted with the

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

QIAamp DSP DNA Blood Mini kit and PCR cycle threshold (Ct) values were determined with the non-variola *Orthopox* real-time PCR assay. Libraries were prepared with amplicon-based and metagenomic sequencing approaches and sequenced on the Illumina MiSeq (2x150 bp) with a targeted 0.5-1 million reads per library for amplicon-based sequencing and 1.5-3 million reads per library for metagenomic sequencing. For amplicon-based sequencing, consensus genomes were generated at a coverage of 10X and percent genome coverage as compared to the reference genome (MT903345) was determined using TheiaCoV_Illumina_PE Workflow Series on Terra.bio. For metagenomic sequencing, genomes were generated using the Broad Institute's viral-pipelines workflows on Terra.bio using both the assemble_refbased and assemble_denovo workflows.

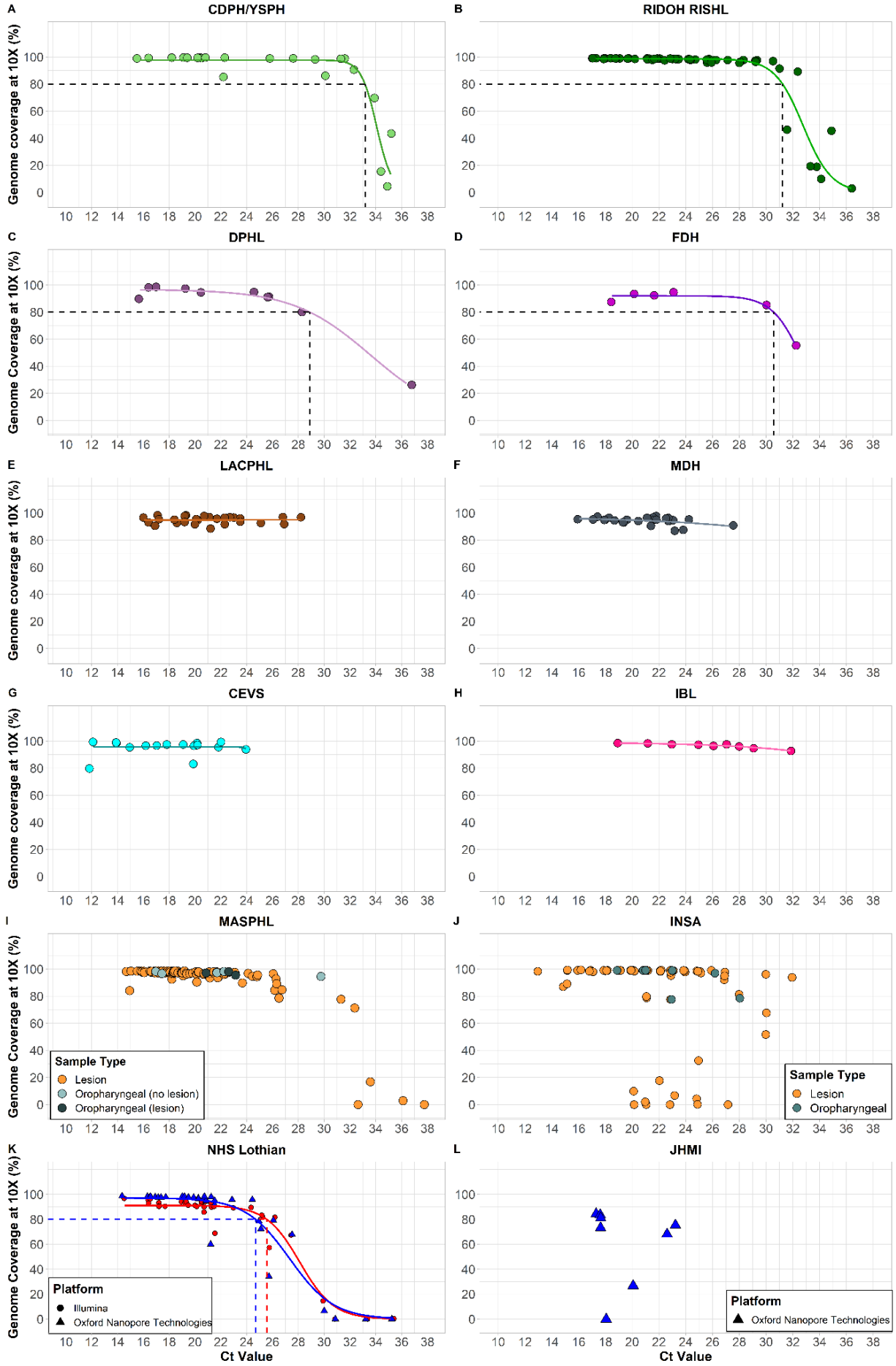
After the initial validation, we immediately made our primer scheme and protocols publicly available and we shipped primer pool aliquots to laboratories across the United States and internationally, to support their sequencing efforts [16]. This resulted in a collaboration between 12 public health laboratories to validate the amplicon-based sequencing approach across different workflows and sequencing platforms (**Figure 2**). Each laboratory used the human monkeypox virus primers with their established SARS-CoV-2 amplicon-based sequencing workflow and sequenced their samples on Illumina or Oxford Nanopore Technologies platforms. We unified the data analysis for all laboratories, by running the same bioinformatics pipeline (iVar) to generate consensus genomes and to determine percent genome coverage at 10X depth (i.e. at least 10 sequencing reads aligned to a genome position). Overall, we found that amplicon-based sequencing resulted in consistently high genome coverage (>80%), with decreasing coverage at higher Ct values (**Figure 3**). By using logistic regression analysis, we determined the PCR Ct value threshold at which each laboratory was able to achieve 80% genome coverage at 10X, which ranged between Ct 24.7 to 33.2 **(Figure 3A-D, K)**. For some labs we were not able to determine this threshold, because (1) the range in Ct values was too narrow to see the decrease at higher Ct values (**Figure 3E-H**), (2) there was insufficient data at higher Ct values (>28) relative to lower Ct values (<28) preventing fitting of a regression curve (**Figure 3I**), or (3) there was high variation in the coverage at lower Ct values (**Figure 3J, L**). Currently, we are further investigating why some samples with low Ct values failed to achieve high genome coverage, but this was likely due to technical error during library preparation or sequencing, and not due to the primer scheme itself as such patterns would have been consistent across all the laboratories. Most samples were sequenced on Illumina platforms, as this was the most commonly used sequencing platform in laboratories that performed the validation. Based on our previous experience with amplicon-based sequencing of viruses such as Zika and SARS-CoV-2, we expected that the primer scheme could also be used with Oxford Nanopore Technologies workflows. The NHS Lothian laboratory had the ability to prepare libraries with both Illumina and Oxford Nanopore technologies workflows and sequenced on the Illumina MiSeq and Oxford Nanopore Technologies GridION. This pairwise comparison showed that genome coverage at 10X was comparable between both workflows (**Figure 3K**). Thus, our multi-site validation shows that the human monkeypox primer scheme can be successfully used across Illumina and Oxford Nanopore Technologies workflows and sequencing platforms, with high genome coverage (>80%) at a Ct range of 24.7 to 33.2 (average: Ct of 29.0).

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 2. Geographical distribution of public health laboratories that validated the human monkeypox virus primer scheme with their established amplicon-based sequencing workflows. Public health laboratories contributing data to this study include: Connecticut Department of Public Health (CDPH), Centro Estadual de Vigilância em Saúde (CEVS), Delaware Public Health Lab (DPHL), Florida Department of Health (FDH), Idaho Bureau of Laboratories (IBL), Johns Hopkins Medical Institutions (JHMI), Los Angeles County Public Health Lab (LACPHL), Massachusetts State Public Health Laboratory (MASPHL), Minnesota Department of Health (MDH), National Health Service Lothian (NHS Lothian), National Institute of Health Dr. Ricardo Jorge (INSA), and Rhode Island Department of Health/ Rhode Island State Health Laboratory (RIDOH RISHL).

Currently, the U.S. Food & Drug Administration (FDA) recommends using lesion swabs for clinical diagnostics [17]. As a result, the most common sample type that we used for sequencing was lesion swabs. However, several studies have shown that human monkeypox virus can also be detected in other sample types such as the throat or oropharyngeal swabs, saliva, feces, urine, and semen [18,19]. Two laboratories had access to these alternate sample types and sequenced oropharyngeal swabs with or without the presence of lesions in the throat. Although the total number of samples included in this study is low, we show that near-complete genomes can be sequenced from oropharyngeal swabs, in the presence and absence of lesions (**Figure 3I-J**). Our findings provide evidence that oropharyngeal swabs could serve as an alternative sample type for sequencing.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 3: Percent genome coverage at 10X for clinical specimens sequenced with the amplicon-based sequencing approach. A. Lesion swabs (n=22) obtained from 12 individuals through the CDPH and sequenced by the Yale School of Public Health (YSPH). Data are fitted with a logistic regression line and the dashed line corresponds to 80% genome coverage at a threshold of Ct=32.0. **B.** Dry vesicle swabs (n=56) from various anatomical sites and sequenced by the RIDOH RISHL. Data are fitted with a logistic regression line and the dashed line corresponds to 80% genome coverage at a threshold of Ct=30.6. **C.** Lesion swabs (n=10) sequenced by the DPHL. Data are fitted with a logistic regression line and the dashed line corresponds to 80% genome coverage at a threshold of Ct=28.4. **D.** Lesion swabs (n=6) sequenced by the FDH. Data are fitted with a logistic regression line and the dashed line corresponds to 80% genome coverage at a threshold of Ct=30.6. **E.** Lesion swabs (n=27) sequenced by the LACPHL. Data are fitted with a logistic regression line. **F.** Lesion swabs (n=25) from various anatomical sites and sequenced by the MDH. Data are fitted with a logistic regression line. **G.** Clinical specimens (n=19) consisting of lesion swabs and crusts of healing lesions sequenced by the CEVS. Data are fitted with a logistic regression line. **H.** Lesion swabs (n=9) sequenced by IBL. Data are fitted with a logistic regression line. **I.** Clinical specimens (n=123) consisting of lesion swabs, oropharyngeal swabs in the absence of lesions, and oropharyngeal swabs in the presence of lesions sequenced by the MASPHL. **J.** Clinical specimens (n=78) consisting of lesion swabs from various anatomical sites as well as oropharyngeal swabs sequenced by INSA. **K.** Vesicle swabs (n=34) from various anatomical sites tested in parallel on Illumina and ONT sequencing platforms by the NHS. Data from both sequencing platforms are fitted with logistic regression lines and the dashed lines correspond to 80% genome coverage at a threshold of Ct=25.0 on the Illumina platform and Ct=24.4 on the ONT platform. **L.** Lesion swabs (n=8) sequenced on the ONT platform by JHMI.

Due to variability in the number of samples that were pooled in sequencing runs, there was some level of variation in the number of sequencing reads generated for each sample by the different labs. This may explain some of the differences in the Ct value threshold to generate a coverage of ~80%. To further investigate the combined effects of Ct value and number of sequencing reads on coverage, we down-sampled data from Connecticut (CDPH/YSPH) to 0.5, 1, 1.5, and 2 million sequencing reads per sample. We used the Connecticut data for down-sampling because samples span the full range of Ct values from ~15-35 and we sequenced libraries on the Illumina NovaSeq with a relatively high number of ~12 million raw reads per sample (range: ~0.4 to ~20 million reads per sample). We found that down-sampling the total number of raw reads from 2 million to 0.5 million resulted in a 7.2% (range: 0-26.1%) decrease in percent genome coverage at 10X (**Figure 4**). To achieve an overall high genome coverage of more than 80% at 10X, we recommend generating at least 1 million sequencing reads per sample, if resources allow. Further increasing the number of sequencing reads to 2 million helps to generate higher coverage for samples with relatively high Ct values >29, but only small differences in coverage at 10X of on average 2.0% were observed with samples that had Ct values <29. Thus, we recommend selecting samples for sequencing with a Ct value below 29 and generating at least 1 million reads per sample, to maximize genome coverage when resources are limited.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

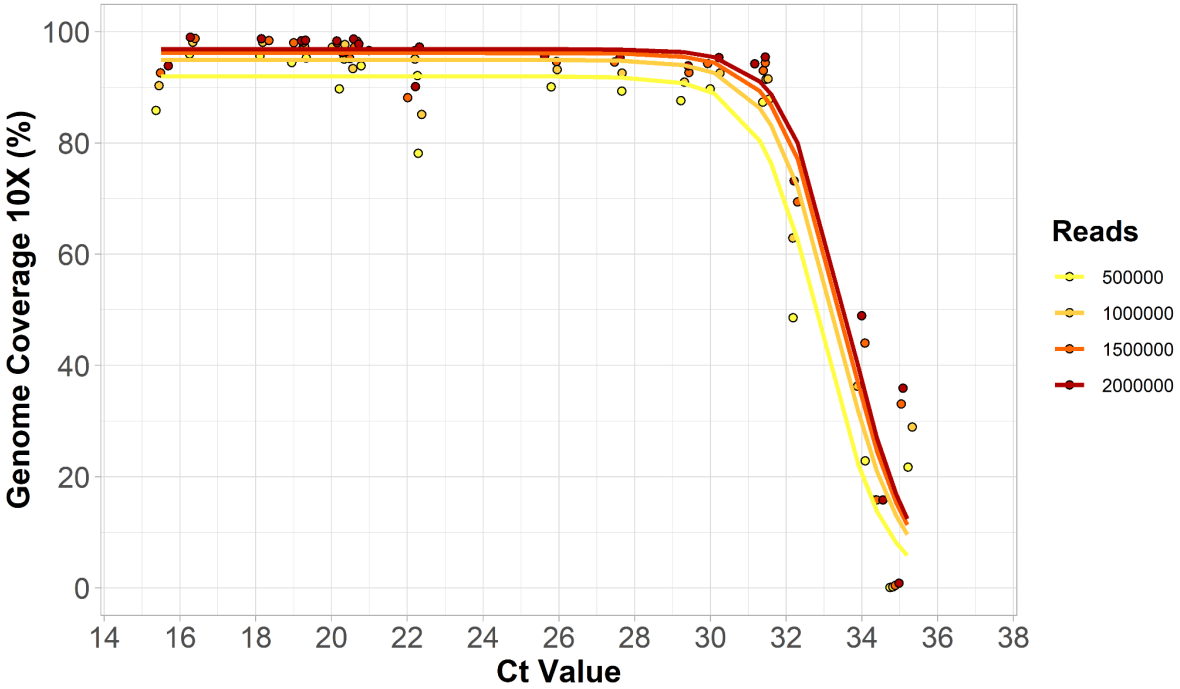


Figure 4. Percent genome coverage at 10X mapped read depth for 22 clinical specimens after down-sampling to a specific number of sequencing reads. To further investigate the combined effects of Ct value and number of sequencing reads per sample, we down-sampled the CDPH/YSPH data to 2, 1.5, 1, and 0.5 million reads per sample. We used logistic analysis to plot regression lines indicating the decrease in percent genome coverage with higher Ct values.

To further investigate the differences in coverage depth for the down-sampled data, we determined the depth of coverage at each nucleotide position across the genome. This analysis again confirmed that increasing Ct values and decreasing sequencing reads per sample result in more regions of the genomes with coverage <10X depth (**Figure 5**). Moreover, it showed variation in depth of coverage across the genome. To further investigate which specific amplicons consistently had low coverage, we determined the depth of coverage for non-overlapping regions of amplicons in our dataset that was down-sampled to 1 million reads/sample and for samples with a Ct <29. By excluding genomic regions covered by overlapping amplicons, we found that amplicons 75 and 188 consistently had a depth of coverage <10X across all 13 samples included in the analysis. Additionally, the mean depth of coverage for amplicons 26, 28, 56, 59, 60, 74, 75, and 96 was also below <10X. To understand the cause for the lower coverage, we investigated whether there were any mismatches in the primers for these amplicons. This revealed a single mismatch with the MPXV_11_RIGHT primer, whereas no mismatches were present in any of the other primers. This suggests that differences in PCR efficiency may explain the lower coverage across these amplicons. Further optimization would be required to improve uniformity in PCR efficiency across the primer pairs in the current primer scheme. This may also help to reduce the number of sequencing reads required to reach a minimum coverage of 10X across the entire genome.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

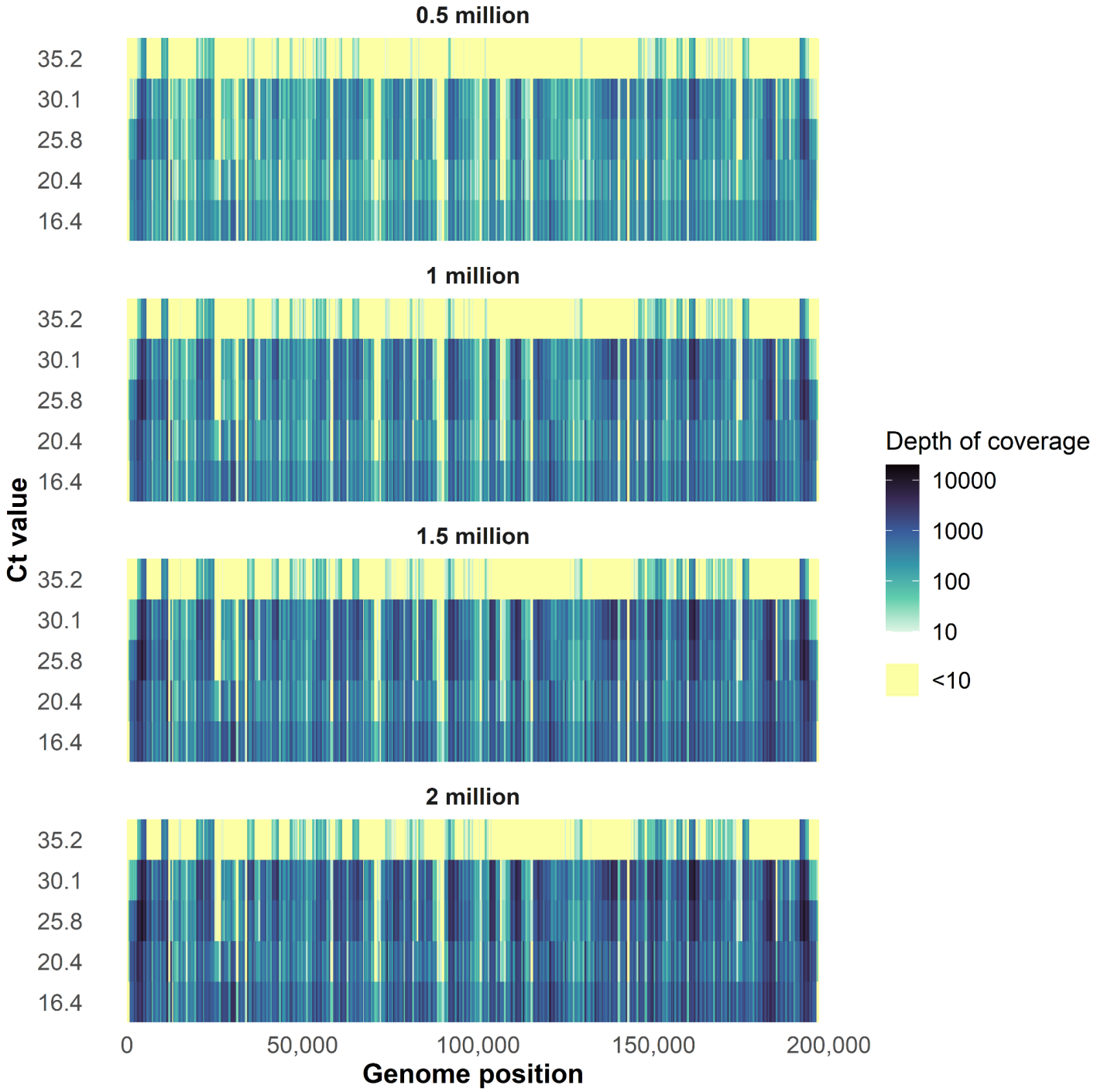


Figure 5: Depth of coverage by genome position for samples representing a range of Ct values and down-sampled to different numbers of sequencing reads. We determined the depth of coverage at each nucleotide position for selected samples that represented a range in Ct values from 16.4-35.2, and for which the number of raw sequencing reads was down-sampled to 2, 1.5, 1, and 0.5 million sequencing reads. Each row represents a single specimen, ranked by Ct value from high (low DNA titer) to low (high DNA titer). Highlighted in yellow are positions of the genome with a depth of coverage below 10X.

Discussion

We developed and validated an amplicon-based sequencing approach for human monkeypox virus to provide a more sensitive, lower cost, and higher throughput alternative to metagenomic sequencing. We used PrimalScheme to develop primers for human monkeypox virus based on a pre-outbreak A.1 lineage genome

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(GenBank accession MT903345) and validated the primer scheme with clinical specimens across 12 public health laboratories. Our validation showed greater breadth and depth of genome coverage when using the amplicon-based sequencing approach as compared to metagenomics, particularly for specimens with lower DNA concentrations. We identified Ct value and number of sequencing reads as the two main factors that influence percent genome coverage. Based on our findings we made the following recommendations for amplicon-based sequencing of human monkeypox virus:

1) Utilize existing amplicon-based SARS-CoV-2 sequencing infrastructure
2) Prioritize samples with a Ct value <29, if resources are limited
3) Generate at least 1 million sequencing reads per sample

Importantly, our human monkeypox virus primer scheme can be used with currently implemented SARS-CoV-2 sequencing workflows and Illumina or Oxford Nanopore Technologies sequencing platforms. This will enable laboratories that are currently using amplicon-based sequencing approaches for SARS-CoV-2 to expand their portfolio by including human monkeypox virus, with minimal change to their overall sequencing and bioinformatics workflow.

The comparatively steep drop-off in genome coverage with higher PCR Ct samples seen with metagenomics reinforces a limitation of this sequencing approach found in other studies that used clinical metagenomics for pathogen sequencing [20]. This lack of sensitivity presents a potential challenge to genomic surveillance, as human monkeypox virus DNA concentrations fluctuate throughout the course of infection and across specimen types [21]. We show that high genome coverage can be achieved with different workflows, sequencing platforms, and sample types such as lesion and oropharyngeal swabs. The higher sensitivity seen in amplicon-based sequencing allows for sequencing of a larger variety of sample types across a wider range of Ct values compared to metagenomic sequencing.

Given the large number (163) of amplicons needed to span the entire human monkeypox virus genome, there was an increased potential for amplicon drop-outs and a resulting reduction in genome coverage. Despite being an inherent risk of amplicon sequencing, here we identified only two amplicons (75 and 118) with consistent low coverage <10X, and an additional 8 amplicons with overall lower coverage. This lower coverage is likely the result of lower amplification efficiency, and may be improved by further optimization to achieve high PCR efficiency across all the primer sets. As a double-stranded DNA virus with genetic proof-reading mechanisms, the monkeypox virus has a comparatively slower evolutionary rate than single-stranded RNA viruses [22]. Less genetic diversity over time translates to fewer differences between the reference genome used to generate the primer scheme and genomes that are associated with the 2022 outbreak, decreasing the risk for amplicon drop-out. As a result, the primer scheme may not have to be updated as frequently as, for example, the primer scheme used to sequence SARS-CoV-2 [14].

There were several limitations in this study. First, we designed the primer scheme to match the current outbreak human monkeypox viruses (clade IIb). The scheme has not been validated for sequencing of viruses belonging to the other endemic monkeypox virus clades (i.e. clades I and IIa) [8], and higher divergence will likely result in lower genome coverage. Future development is needed to expand the scheme to cover all monkeypox virus clades. Second, the large number of primers included in the scheme resulted in differences in PCR efficiency between amplicons. By down-sampling the sequencing data, we show that a higher number of sequencing reads can overcome these challenges. Further optimization by changing primer concentrations may help to reduce the number of raw sequencing reads needed to reach at least 10X coverage depth across the genome. Third, although results between laboratories were consistent, we observed some variability in the relation between Ct value and genome coverage. Differences in the used qPCR assays and sequencing workflows can likely explain this variability. Fourth, there were some challenges with the more variable ends of

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the genome that contain inverted terminal repeats. This resulted in multiple binding sites for primers spanning these regions. In addition, this can create challenges when mapping reads to the inverted terminal repeats for both Illumina and Oxford Nanopore Technologies data, which may result in the misalignment of reads. This warrants careful interpretation of sequencing results in these genome regions. Lastly, Orthopox viruses are known to have genomic rearrangements (translocations, duplications, and inversions) especially in the inverted terminal repeats [23], which may not be well identified by amplicon-based sequencing. Although we recently identified a 600 bp deletion in the inverted terminal repeats that affects commonly used real-time PCR assays [24], amplicon-based sequencing may not be able to identify other genomic rearrangements that may have epidemiological or clinical significance. We recommend periodic long read metagenomic sequencing to supplement large-scale amplicon-based sequencing as an additional strategy to improve genomic surveillance of human monkeypox virus.

Through this study, we have shown that amplicon-based sequencing can increase the sensitivity, breadth, and depth of human monkeypox virus genome coverage with low DNA concentration specimens. This finding has significant implications for genomic surveillance of human monkeypox virus. In conclusion, our amplicon-based sequencing approach has helped laboratories worldwide to quickly adapt their existing sequencing workflows in response to the global outbreak of human monkeypox virus through genomic surveillance.

Methods

Ethics statement

As part of this study, we sequenced remnant clinical samples that tested presumptive positive for monkeypox virus. Ethical oversight for each institution is indicated in **Table 1**. All data were de-identified prior to sharing and sample codes as included in the manuscript are not known outside the research groups.

Table 1: Ethical oversight

Remnant clinical samples	State/Country	Institutional oversight	Decision	Protocol ID
Lesion swabs	Brazil	Centro Estadual de Vigilância em Saúde	Waived	SES-RS Nº 357/2021 and SES-RS Nº 849/202
Lesion swabs	Connecticut, US	Connecticut Department of Public Health	Waived (deemed public health surveillance)	N/A
Lesion swabs	Connecticut, US	Institutional Review Board from the Yale University Human Research Protection Program	Waived	2000033281
Lesion swabs	Delaware, US	Delaware Public Health Lab	Waived (deemed public health surveillance)	N/A
Lesion swabs	Florida, US	Florida Department of Health	Waived (deemed public health surveillance)	N/A
Swabs	Idaho, US	Idaho Bureau of Laboratories	Waived (deemed public health surveillance)	N/A
Swabs	Maryland, US	Johns Hopkins Medical Institutional Review Board	Approved	IRB00317591
Lesion swabs	California, US	Los Angeles County Public Health Lab	Waived (deemed public health surveillance)	N/A
Lesion and oropharyngeal swabs	Massachusetts, US	Massachusetts Department of Public Health, Massachusetts State Public Health Laboratory	Approved	1917413
Swabs	Minnesota, US	Minnesota Department of Health	Waived (deemed public health surveillance)	N/A

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Lesion and oropharyngeal swabs	Portugal	National Institute of Health	Waived (deemed public health surveillance)	Technical Orientation nº04/2022
Swabs	United Kingdom	NHS Lothian BioResource	Approved	20/ES/0061
Dry vesicle swabs	Rhode Island, US	Rhode Island Department of Health, Rhode Island State Health Laboratories	Waived (deemed public health surveillance)	216-RICR-30-05-1

Primer scheme design

The human monkeypox primer scheme (v1) was designed with PrimalScheme [15] using a pre-outbreak clade IIb reference genome (GenBank accession: MT903345) belonging to the A.1 lineage, following the newly proposed monkeypox virus naming system [8]. The primer scheme comprises a total of 163 primer pairs with an amplicon length ranging between 1597 and 2497 bp (average length of 1977 bp; **Supplementary Table 1**).

Validation & data analysis

We made a detailed protocol publicly available and shared pooled primer aliquots with laboratories across the United States and internationally to validate the human monkeypox primer scheme across different sequencing workflows and platforms [16]. Each laboratory sequenced samples by adapting their SARS-CoV-2 amplicon sequencing workflows, performed their own bioinformatics, and submitted sequencing data and consensus genomes to public repositories such as NCBI or GISAID (**Supplementary Table 2**). By sharing de-identified sequencing data (e.g. coded human-depleted raw reads or BAM files containing mapped reads) and metadata (e.g. sample codes, sample types, and Ct values), we ran the same bioinformatics pipeline on all data to determine breadth and depth of coverage for each genome under standardized conditions.

Each laboratory shared either de-identified human-depleted raw sequencing reads in FASTQ or BAM files containing mapped reads. We then extracted mapped reads from the BAM using bamToFastq or “BEDtools bamtofastq” (version 2.30.0) [25] into FASTQ files before downstream analysis. To start the analysis, we re-mapped the reads to the human monkeypox reference genome (GenBank accession: MT903345) using BWA-MEM (version 0.7.17-r1188) [26]. The generated BAM mapping files were sorted using SAMtools (version 1.6) [27] and then used as input to iVar (version 1.3.1) [28] to trim primer sequences from reads. The trimmed BAM files were then used to generate consensus sequences using iVar with a minimum read depth of 10. We also calculated the per-base read coverage depth using genomeCoverageBed or “BEDTools genomecov” (version 2.30.0) [25]. To calculate the percent genome coverage for each sample, we generated a pairwise whole-genome sequence alignment of each sample against the human monkeypox virus reference genome (GenBank accession: MT903345) using Nextalign (version 1.4.1) [29]. We then calculated the percentage of alignment positions, excluding ambiguous nucleotides and deletions, using BioPython (version 1.7.9) [30].

To investigate the effect of the number of sequenced reads on the percent genome coverage, we generated down-sampled sequence reads for the CDPH/YSPH samples. We used the CDPH/YSPH samples because they were sequenced at sufficiently high depth to allow down-sampling at different total read counts, namely 0.5, 1, 1.5, and 2 million reads per sample. We first generated interleaved sequencing read files from the paired-end sequencing reads using seqfu (version 1.14.0) [31] and then randomly downsampled the reads at the specified total read count thresholds using “seqtk sample” (version 1.3-r106) [32]. Similarly, we generated the consensus whole-genome sequences and calculated the percentage genome coverage as described in the previous paragraph.

All further data analysis and plotting were performed using R statistical software v4.2.0 [33] using the ggplot2 v3.3.6 [34], dplyr v1.0.9 [35], tidyr v1.2.0 [36], and cowplot v1.1.1 [37] packages. Initially, we removed sample data submitted from partnering labs with not detected (ND) Ct values due to the inability to determine the

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

relationship between Ct value and genome coverage for these data points. Then we subsetted each dataset to a coverage depth of 10X and fitted a logistic function specified as follows: y=a / (1 + exp(-b * (x - c))), where a, b, and c are parameters and 'x' is the PCR Ct value and 'y' is the percentage genome coverage or completeness. We used the 'curve_fit' function in the numpy Python package to fit the model [38]. To interpolate the Ct value corresponding to an 80% threshold Ct value, we rearranged the logistic equation to estimate 'x' given the estimated parameters a, b, and c, and genome coverage or 'y' of 80% at each read coverage depth.

Centro Estadual de Vigilância em Saúde (CEVS)

Clinical specimens

A total of 19 clinical specimens were included in this validation consisting of lesion swabs and the crusts of healing lesions. DNA was extracted from clinical specimens using the Invitrogen PureLink Viral RNA/DNA Mini kit and tested with a clade-specific PCR assay on the Bio-Rad CFX opus 96 instrument.

Amplicon-based sequencing

Libraries were prepared for sequencing using the Illumina DNA prep kit. Pooled libraries were sequenced on the Illumina MiSeq v3 (paired-end 150), with 2 million reads per library. Primers were trimmed and consensus genomes were generated at a minimum depth of coverage of 10X using the ViralFlow pipeline [39].

Connecticut Department of Public Health/ Yale School of Public Health (CDPH/YSPH)

Clinical specimens

A total of 22 clinical specimens were included in this validation consisting of swabs from lesions of 12 individuals. DNA was extracted from clinical specimens via the Roche MagNA pure 96 kit or manual extraction and tested with the non-variola *Orthopox* real-time PCR assay on the Bio-Rad CFX96 instrument [40].

Amplicon-based sequencing

Libraries were prepared for sequencing using the Illumina COVIDSeq test (RUO version). Pooled libraries were sequenced on the Illumina NovaSeq (paired-end 150), with on average 12 million reads per library (range: ~0.4 to ~20 million reads per library). Primers were trimmed and consensus genomes were generated at a minimum depth of coverage of 10X using iVar v 1.3.1 [28].

Delaware Public Health Lab (DPHL)

Clinical specimens

A total of 10 clinical specimens were included in this validation consisting of dry swabs from lesion sites. DNA was manually extracted from clinical specimens using the QIAGEN QIAamp DSP DNA Blood Mini Kit. Amplification was achieved by using the Perfecta Multiplex qPCR SuperMix Low Rox PCR assay in conjunction with CDC issued Non-variola Orthopoxvirus Real-Time PCR Primer and Probe Set on the Applied Biosystems 7500 Fast Real-Time PCR instrument.

Amplicon-based sequencing

Libraries were prepared for sequencing using the Illumina DNA prep kit. Pooled libraries were sequenced on the Illumina Miseq (paired-end 150), with about 1 million to 2.5 million reads per library. Primers were trimmed and consensus genomes were generated at a minimum depth of coverage of 10X via BWA Version 0.7.17-r1188 and iVar version 1.3.1.

Florida Department of Health (FDH)

Clinical specimens

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

A total of 6 clinical specimens were included in this validation consisting of lesion swabs. DNA was extracted from clinical specimens using the Qiagen QIAamp DSP DNA Blood Mini Kit and tested with the non-variola *Orthopox* real-time PCR assay on the BioRad T100 instrument [40]

Amplicon-based sequencing

Libraries were prepared for sequencing using the Illumina Nextera v2 kit. Pooled libraries were sequenced on the Illumina iSeq 100 v2 (paired-end 150), with 1-2.4 million reads per library. Primers were trimmed and consensus genomes were generated at a minimum depth of coverage of 10X using the flaq_mpx pipeline [41].

Idaho Bureau of Laboratories (IBL)

Clinical specimens

A total of 9 lesion swab clinical specimens were included in this validation. DNA was extracted from clinical specimens using the QIAGEN QIAamp DSP DNA Blood Mini kit and tested with Perfecta Multiplex qPCR SuperMix Low Rox PCR assay in conjunction with the CDC FDA-approved Non-variola Orthopoxvirus (VAC1) assay on the Applied Biosystems® 7500 Fast Dx Real-Time PCR instrument.

Amplicon-based sequencing

Libraries were prepared for sequencing using the Illumina DNA prep kit. Pooled libraries were sequenced on the Illumina MiSeq, with 0.8-1.2 million reads per library. Primers were trimmed and consensus genomes were generated at a minimum depth of coverage of 10X using the CLC Genomics Workbench 22.0.2 [42].

Johns Hopkins Medical Institutions (JHMI)

Clinical specimens

A total of 8 clinical specimens were included in this validation consisting of lesion swabs. DNA was extracted from clinical specimens using the bioMérieux easyMag and tested with a LDT PCR assay that adopted the primer and probe sequences of the non-variola orthopoxvirus, modified from Li, et al [43]. Total reaction volume for the real-time PCR was 20 µL (5 µL of template and 15 µL master mix). The master mix contained 5 µL TaqPath 1-Step RT-qPCR Master Mix (Applied Biosystems, A15299, Waltham, MA), 4 µL water, and 1 µL of each primer (10 nm) and the probe (5 nm) in addition to 1 µL of each primer (10 nm) and the probe (5 nm) for the RNAseP internal control target. Real-time PCR was performed using Prism 7500 Detection System (Applied Biosystems) and the following cycling conditions: 1 cycle at 95.0°C for 2 minutes and 40 cycles at 95.0°C for 3 seconds and 60.0°C for 31 seconds.

Amplicon-based sequencing

Libraries were prepared for sequencing using NEBNext ARTIC reagents for SARS-CoV-2 sequencing. Pooled libraries were sequenced on the Oxford Nanopore Technologies GridIon. Primers were trimmed and consensus genomes were generated at a minimum depth of coverage of 10X using the ARTIC bioinformatics pipeline [44].

Los Angeles County Public Health Lab (LACPHL)

Clinical specimens

A total of 27 remnant specimens were included in this validation consisting of lesion swabs. DNA was extracted using the Qiagen EZ1&2 DNA Tissue Kit from specimens that were previously tested with the non-variola *Orthopoxvirus* real-time PCR assay on the ABI 7500 FastDx instrument following Centers for Disease Control and Prevention Laboratory Response Network protocols [40].

Amplicon-based sequencing

Libraries were prepared for sequencing using the Illumina DNA prep kit. Pooled libraries were sequenced on the Illumina MiSeq v2 (paired-end 150), with 0.5-1 million reads per library. Primers were trimmed and

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

consensus genomes were generated at a minimum depth of coverage of 10X via the TheiaCoV_Illumina_PE Workflow Series on Terra.bio [45].

Massachusetts State Public Health Laboratory (MASPHL)

Clinical specimens

A total of 133 clinical specimens were included in this validation consisting of both lesion swabs and oropharyngeal swabs with and without detection of lesions in the throat. DNA was extracted from clinical specimens using the QIAamp DSP DNA Blood Mini kit and tested with the non-variola *Orthopox* real-time PCR assay on the ABI 7500 Fast Real Time PCR instrument [40].

Amplicon-based sequencing

Libraries were prepared for sequencing using the Illumina DNA prep kit. Initial library clean-up in the Illumina DNA prep protocol was done by following recommendations for standard DNA input, but later optimization showed improved coverage when following the recommendations for small PCR amplicon input. Pooled libraries were sequenced on the Illumina MiSeq (paired-end 150), with a target of 0.5-1 million reads per library. Primers were trimmed and consensus genomes were generated at a minimum depth of coverage of 10X via the TheiaCoV_Illumina_PE Workflow Series on Terra.bio. TheiaCoV_Illumina_PE was originally written to enable genomic characterization of SARS-CoV-2 specimens from Illumina paired-end amplicon read data. Modifications to TheiaCoV_Illumina_PE were made to to support MPXV genomic characterization; these modifications accomodated the use of a MPXV reference sequence and primer scheme for consensus genome assembly. These updates were made available in the TheiaCov v2.2.0 release [45].

Metagenomic sequencing

Samples were initially sequenced using a metagenomics approach after preparation with the Illumina DNA prep kit. Samples were first quantified using the Qubit 1x dsDNA High Sensitivity kit to determine concentration. Initial loading volume was calculated for each sample to fall in the recommended range of 100 to 500 ng. Each sample was then run in duplicate through the Illumina DNA Prep kit for library preparation following manufacturer's protocol. After Flex Amplification PCR, the libraries were cleaned using the protocol option for standard DNA input. Post-library purification, each sample library was run on the Tapestation D1000 to determine the average peak size. Samples were also again quantified using the Qubit dsDNA High Sensitivity kit to best normalize the samples when pooling the libraries. Samples were then pooled, denatured, and diluted to 10 pM before being spiked with 5% PhiX. The diluted pooled libraries were then loaded and run on an Illumina v.2 300 cycle MiSeq cartridge, with a target of 1.5-3 million reads per library. Genomes were generated using the Broad Institute's viral-pipelines workflows on Terra.bio using both the assemble_refbased and assemble_denovo workflows. The assemble_refbased workflow aligned reads against the MA001 (ON563414.3) genome for consensus generation. The assemble_denovo workflow scaffolded de novo SPAdes contigs against both the MA001 (ON563414.3) and RefSeq (NC_063383.1) references, followed by read based polishing.

Minnesota Department of Health (MDH)

Clinical specimens

A total of 25 clinical specimens were included in this validation consisting of lesion swabs from various anatomical sites. DNA was extracted from clinical specimens using the QIAamp DSP DNA Blood Mini kit and tested with a non-variola *Orthopox* real-time PCR assay on the ABI 7500 Fast Real Time PCR instrument [40]

Amplicon-based sequencing

Libraries were prepared for sequencing using the Illumina DNA prep kit. Pooled libraries were sequenced on the Illumina MiSeq v2 (paired-end 250), with target of 500,000 reads per library. Primers were trimmed and consensus genomes were generated at a minimum depth of coverage of 10X using nf-core/viralrecon.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

National Institute of Health Dr. Ricardo Jorge (INSA)

Clinical specimens

A total of 78 clinical specimens were included in this validation consisting of lesion swabs from various anatomical sites and oropharyngeal swabs. DNA was extracted from clinical specimens using the MagMAX Viral/Pathogen Nucleic Acid Isolation kit and tested with a real-time PCR assay on the CFX Opus real-time PCR system [46–48].

Amplicon-based sequencing

Libraries were prepared for sequencing using the Illumina Nextera XT library prep kit. Pooled libraries were sequenced on the Illumina NextSeq 550, with 2 million reads per library. Reads were trimmed and consensus genomes were generated at a minimum depth of coverage of 30X using INSaFLU (https://insaflu.insa.pt/; v1.5.2) pipeline [49].

National Health Service Lothian (NHS Lothian)

Clinical specimens

A total of 34 clinical specimens were included in this validation consisting of vesicle swabs and swabs from various anatomical sites. DNA was extracted from clinical specimens using the Biomerieux Nuclisens EasyMag kit and tested with the clade-specific real-time PCR assay on the Applied Biosystems 7500 Fast Real-Time PCR instrument [50].

Amplicon-based sequencing

Illumina libraries were prepared for sequencing using the Illumina COVIDSeq test (RUO version). Pooled libraries were sequenced on the Illumina MiSeq - Micro v2 reagent kit, with 400,000 reads per library. Primers were trimmed and consensus genomes were generated at a minimum depth of coverage of 5X using the Public Health Wales Nextflow nCoV-2019 pipeline which utilizes iVar [51]. Whole genome PCR amplicons were also used to prepare Nanopore libraries using the Artic LoCost method [52], substituting Blunt TA Ligase with NEBNext Ultra II Ligation Mastermix for barcode ligation. Libraries were pooled on a single R9.4.1 flowcell and sequenced with a GridION (Oxford Nanopore Technologies) running live High Accuracy basecalling in MinKnow v21.11.6, aiming for 100,000 reads per library. Consensus genomes were generated at a minimum depth of coverage of 20x with the Artic field bioinformatics pipeline v1.2.1 and variants called with Nanopolish [51].

Rhode Island Department of Health/ Rhode Island State Health Laboratory (RIDOH RISHL)

Clinical specimens

A total of 56 clinical specimens were included in this validation consisting of dry vesicle swabs from various anatomical sites. DNA was extracted from clinical specimens using the Qiagen QIAmp DSP Blood mini kit and tested with the non-variola *Orthopox* real-time PCR assay on the Applied biosystems 7500 instrument.

Amplicon-based sequencing

Libraries were prepared for sequencing using the Illumina DNA prep kit. Pooled libraries were sequenced on the Illumina MiSeq (paired-end 150), with 1.2 million reads per library. Primers were trimmed and consensus genomes were generated at a minimum depth of coverage of 10X using BBMap version 38.96, iVar version 1.3.1, and Samtools version 1.6.

References

1. Grubaugh ND, Ladner JT, Lemey P, Pybus OG, Rambaut A, Holmes EC, et al. Tracking virus outbreaks in the twenty-first century. Nat Microbiol. 2019;4: 10–19.
2. Gardy J, Loman NJ, Rambaut A. Real-time digital pathogen surveillance - the time is now. Genome Biol.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2015;16: 155.
3. Gire SK, Goba A, Andersen KG, Sealfon RSG, Park DJ, Kanneh L, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science. 2014;345: 1369–1372.
4. Chen Z, Azman AS, Chen X, Zou J, Tian Y, Sun R, et al. Global landscape of SARS-CoV-2 genomic surveillance and data sharing. Nat Genet. 2022;54: 499–507.
5. GISAID. EpiCoV. 2022. https://www.gisaid.org/
6. Otu A, Ebenso B, Walley J, Barceló JM, Ochu CL. Global human monkeypox outbreak: atypical presentation demanding urgent public health action. Lancet Microbe. 2022;3: E554–E555.
7. CDC. About Monkeypox. 2022. https://www.cdc.gov/poxvirus/monkeypox/about.html
8. Happi C, Adetifa I, Mbala P, Njouom R, Nakoune E, Happi A, et al. Urgent need for a non-discriminatory and non-stigmatizing nomenclature for monkeypox virus. PLoS Biol. 2022;20: e3001769.
9. Isidro J, Borges V, Pinto M, Sobral D, Santos JD, Nunes A, et al. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. Nat Med. 2022;28: 1569–1572.
10. CDC. 2022 Monkeypox Outbreak Global Map. 2022. https://www.cdc.gov/poxvirus/monkeypox/response/2022/world-map.html
11. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol. 2017;35: 833–844.
12. Gigante CM, Korber B, Seabolt MH, Wilkins K, Davidson W, Rao AK, et al. Multiple lineages of Monkeypox virus detected in the United States, 2021- 2022. bioRxiv. 2022. doi:10.1101/2022.06.10.495526
13. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017;12: 1261–1276.
14. Artic Network. 2020. https://artic.network/ncov-2019
15. Smith A. PrimalScheme: a tool for designing primer panels for multiplex PCR. https://github.com/aresti/primalscheme
16. Chen NFG, Gagne L, Doucette M, Smole S, Buzby E, Hall J, et al. Monkeypox virus multiplexed PCR amplicon sequencing (PrimalSeq) V.2. 2022. doi:10.17504/protocols.io.5qpvob1nbl4o/v2
17. US Food & Drug Administration. For Monkeypox Testing, Use Lesion Swab Samples to Avoid False Results: FDA Safety Communication. In: U.S. Food and Drug Administration. 2022. https://www.fda.gov/medical-devices/safety-communications/monkeypox-testing-use-lesion-swab-samples-avoid-false-results-fda-safety-communication
18. Thornhill JP, Barkati S, Walmsley S, Rockstroh J, Antinori A, Harrison LB, et al. Monkeypox Virus Infection in Humans across 16 Countries - April-June 2022. N Engl J Med. 2022;387: 679–691.
19. Peiró-Mestres A, Fuertes I, Camprubí-Ferrer D, Marcos MÁ, Vilella A, Navarro M, et al. Frequent detection of monkeypox virus DNA in saliva, semen, and other clinical samples from 12 patients, Barcelona, Spain, May to June 2022. Euro Surveill. 2022;27. doi:10.2807/1560-7917.ES.2022.27.28.2200503
20. Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet. 2019;20: 341–355.
21. Adler H, Gould S, Hine P, Snell LB, Wong W, Houlihan CF, et al. Clinical features and management of human monkeypox: a retrospective observational study in the UK. Lancet Infect Dis. 2022;22: 1153–1162.
22. Peck KM, Lauring AS. Complexities of Viral Mutation Rates. J Virol. 2018;92. doi:10.1128/JVI.01031-17
23. Gigante CM, Plumb M, Ruprecht A, Zhao H, Wicker V, Wilkins K, et al. Genomic deletions and rearrangements in monkeypox virus from the 2022 outbreak, USA. bioRxiv. 2022. p. 2022.09.16.508251. doi:10.1101/2022.09.16.508251
24. CDC. Lab Alert: MPXV TNF Receptor Gene Deletion May Lead to False Negative Results with Some MPXV Specific LDTs. 2022. https://www.cdc.gov/locs/2022/09-02-2022-lab-alert-MPXV_TNF_Receptor_Gene_Deletion_May_Lead_False_Negative_Results_Some_MPXV_Specific_LDTs.html
25. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841–842.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

26. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760.
27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079.
28. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. 2019;20: 8.
29. Neher Lab. nextalign: Viral genome reference alignment. 2021. Available: https://github.com/neherlab/nextalign
30. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009;25: 1422–1423.
31. Telatin A, Fariselli P, Birolo G. SeqFu: A Suite of Utilities for the Robust and Reproducible Manipulation of Sequence Files. Bioengineering (Basel). 2021;8. doi:10.3390/bioengineering8050059
32. Li H. seqtk. https://github.com/lh3/seqtk
33. R Core Team. The R Project for Statistical Computing. In: R: A language and environment for statistical computing. https://www.r-project.org/
34. Wickham H. ggplot2. Springer New York; 2009.
35. Wickham H, François R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/web/packages/dplyr/index.html
36. Wickham H, Girlich M. tidyr: Tidy Messy Data. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/web/packages/tidyr/index.html
37. Wilke CO. cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2.” https://CRAN.R-project.org/package=cowplot
38. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585: 357–362.
39. Dezordi FZ, Neto AM da S, Campos T de L, Jeronimo PMC, Aksenen CF, Almeida SP, et al. ViralFlow: A Versatile Automated Workflow for SARS-CoV-2 Genome Assembly, Lineage Assignment, Mutations and Intrahost Variant Detection. Viruses. 2022;14: 217.
40. Aden TA, Blevins P, York SW, Rager S, Balachandran D, Hutson CL, et al. Rapid Diagnostic Testing for Response to the Monkeypox Outbreak - Laboratory Response Network, United States, May 17-June 30, 2022. MMWR Morb Mortal Wkly Rep. 2022;71: 904–907.
41. Bphl FL. flaq_mpx: FL BPHL’s Monkeypox analysis pipeline for clinical specimens. https://github.com/BPHL-Molecular/flaq_mpx
42. QIAGEN CLC Genomics Workbench 22.0.2. Available: https://digitalinsights.qiagen.com/
43. Li Y, Olson VA, Laue T, Laker MT, Damon IK. Detection of monkeypox virus with real-time PCR assays. J Clin Virol. 2006;36: 194–203.
44. Loman N, Rowe W, Andrew R. ARTIC Network nCoV-2019 novel coronavirus bioinformatics protocol. 2020 [cited 17 Jan 2022]. Available: https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html
45. Libuit K, Petit RA III, Ambrosio F, Kapsak C, Smith E, Sevinsky J, et al. Public Health Viral Genomics: Bioinformatics workflows for genomic characterization, submission preparation, and genomic epidemiology of viral pathogens, especially SARS-CoV-2. (Version 2.2.0). 2022. Available: https://github.com/theiagen/public_health_viral_genomics
46. Nitsche A, Ellerbrok H, Pauli G. Detection of orthopoxvirus DNA by real-time PCR and identification of variola virus DNA by melting analysis. J Clin Microbiol. 2004;42: 1207–1213.
47. Kurth A, Achenbach J, Miller L, Mackay IM, Pauli G, Nitsche A. Orthopoxvirus detection in environmental specimens during suspected bioterror attacks: inhibitory influences of common household products. Appl Environ Microbiol. 2008;74: 32–37.
48. Shchelkunov SN, Shcherbakov DN, Maksyutov RA, Gavrilova EV. Species-specific identification of variola, monkeypox, cowpox, and vaccinia viruses by multiplex real-time PCR assay. J Virol Methods. 2011;175:

medRxiv preprint doi: https://doi.org/10.1101/2022.10.14.22280783; this version posted October 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

163–169.
49. Borges V, Pinheiro M, Pechirra P, Guiomar R, Gomes JP. INSaFLU: an automated open web-based bioinformatics suite “from-reads” for influenza whole-genome-sequencing-based surveillance. Genome Med. 2018;10: 46.
50. Li Y, Zhao H, Wilkins K, Hughes C, Damon IK. Real-time PCR assays for the specific detection of monkeypox virus West African and Congo Basin strain DNA. J Virol Methods. 2010;169: 223–227.
51. Connor Lab. ncov2019-artic-nf: A Nextflow pipeline for running the ARTIC network’s fieldbioinformatics tools, with a focus on ncov2019. https://github.com/connor-lab/ncov2019-artic-nf
52. Tyson JR, Phillip J, Stoddart D, Sparks N, Wickenhagen A, Hall G, et al. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. Cold Spring Harbor Laboratory. 2020. p. 2020.09.04.283077. doi:10.1101/2020.09.04.283077

Supplementary data

Supplementary Table 1: Human monkeypox virus primer scheme (v1). Primer positions are determined based on alignment to the MT903345 reference genome.

Supplementary Table 2: Source data for samples included in this validation study. Listed are institute, specimen code, sample type, Ct value, sequencing platform, percent genome coverage at 10X, and accession numbers.

Acknowledgements

We thank Cornelius Roemer for help with the logistic regression analysis. This publication was made possible by CTSA Grant Number UL1 TR001863 from the National Center for Advancing Translational Science (NCATS), a component of the National Institutes of Health (NIH) awarded to CBFV. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH. INSA was partially funded by the HERA project (Grant/2021/PHF/23776) supported by the European Commission through the European Centre for Disease Control.

Data availability

Genomic data are available on NCBI Sequence Read Archive (SRA), GenBank, and GISAID (see accession numbers in **Supplementary Table 2**). All other data are included in the manuscript and the supplementary files.

Competing interests

NDG is a consultant for Tempus Labs and the National Basketball Association for work related to COVID-19. All other authors declare no competing interests.

Author contributions

NFGC, CC, LG, NDG, GRG, and CBFV designed the study; all authors contributed to data/sample collection, sequencing, or bioinformatic analysis; NFGC, CC, LG, MD, DJP, NDG, GRG, CBFV analyzed the validation data; NFGC, CC, and CBFV drafted the manuscript; all authors reviewed and approved the manuscript.