Molecular Ecology Resources

WILEY

MOLECULAR ECOLOGY RESOURCES

RESOURCE ARTICLE OPEN ACCESS

Evaluating the Benefits and Limits of Multiple Displacement Amplification With Whole-Genome Oxford Nanopore Sequencing

Fiifi Agyabeng-Dadzie[1] | Megan S. Beaudry[2] | Alex Deyanov[3] | Haley Slanis[3] | Minh Q. Duong[3] | Randi Turner[4,5] | Asis Khan[5] | Cesar A. Arias[3,6,7] | Jessica C. Kissinger[1,4,8] | Travis C. Glenn[1,2,8] | Rodrigo de Paula Baptista[3,4,6,7,8]

[1]Department of Genetics, University of Georgia, Athens, Georgia, USA | [2]Department of Environmental Health Science, University of Georgia, Athens, Georgia, USA | [3]Center for Infectious Disease, Houston Methodist Research Institute, Houston, Texas, USA | [4]Center for Tropical and Emerging Global Diseases, University of Georgia, Athens, Georgia, USA | [5]USA Department of Agriculture, Agricultural Research Service, Beltsville Agricultural Research Service, Animal Parasitic Disease Laboratory, Beltsville, Maryland, USA | [6]Division of Infectious Diseases and Department of Medicine, Houston Methodist Hospital, Houston, Texas, USA | [7]Department of Medicine, Weill Cornell Medical College, New York, New York, USA | [8]Institute of Bioinformatics, University of Georgia, Athens, Georgia, USA

Correspondence: Rodrigo de Paula Baptista (rdepaulabaptista@houstonmethodist.org)

Received: 14 August 2024 | **Revised:** 4 February 2025 | **Accepted:** 17 February 2025

Handling Editor: Joanna Kelley

Funding: This work was funded in part by the National Institute for Allergy and Infectious Diseases, NIAID, R01 AI148667 to T.C.G. and J.C.K., NIH T32GM142623 to F.A.-D. The funders played no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Keywords: Apicomplexa | Enterobacterales | Enterococcaceae | infectious diseases | low-abundance DNA | rare DNA

ABSTRACT

Multiple displacement amplification (MDA) outperforms conventional PCR in long fragment and whole-genome amplification, making it attractive to couple MDA with long-read sequencing of samples with limited quantities of DNA to obtain improved genome assemblies. Here, we explore the efficacy and limits of MDA for efficient low-cost genome sequence assembly using Oxford Nanopore Technologies (ONTs) rapid library preparations and minION sequencing. We successfully generated almost complete genome sequences for all organisms examined, including Gram-positive (*Staphylococcus aureus*, *Enterococcus faecium*) and Gram-negative (*Escherichia coli*) prokaryotes and one challenging eukaryotic pathogen (*Cryptosporidium* spp) representing a broad spectrum of critical infectious disease pathogens. High-quality data from those samples were generated starting with only 0.025 ng of total DNA. Controlled sheared DNA samples exhibited a distinct pattern of size increase after MDA, which may be associated with the amplification of long, low-abundance fragments present in the assay, as well as generating concatemeric sequences during amplification. To address concatemers, we developed a computational pipeline (CADECT: Concatemer Detection Tool) to identify and remove putative concatemeric sequences. This study highlights the efficacy of MDA in generating high-quality genome assemblies from limited amounts of input DNA. Also, the CADECT pipeline effectively mitigated the impact of concatemeric sequences, enabling the assembly of contiguous sequences even in cases where the input genomic DNA was degraded. These results have significant implications for the study of organisms that are challenging to culture in vitro, such as *Cryptosporidium*, and for expediting critical results in clinical settings with limited quantities of available genomic DNA.

1 | Introduction

The advent of next-generation sequencing technologies has revolutionized genomics research by enabling the rapid and cost-effective generation of vast amounts of sequencing data (Slatko et al. 2018; Hu et al. 2021). Among these technologies, Oxford Nanopore Sequencing (ONT) stands out due to its ability to provide long-read sequencing data in real-time, with lower instrument costs when compared to the other major commercial long-read sequencing platform, PacBio (Pacbio 2022). ONT sequencing has been used for numerous applications, including de novo genome assembly, metagenomics and pathogen detection. However, ONT sequencing library preparations typically still require higher quality and higher quantities of DNA inputs than may be available for many projects. ONT rapid library preparations usually require at least 50–200 ng of input DNA for the older and newer kits per sample. However, more input DNA is required when pooling fewer than eight barcoded samples, as a minimum of 400 ng is recommended for loading onto a MinION flow cell. Many samples have also degraded DNA with fragments that are too short for ONT rapid library preparation or have limited utility in assembly. Removing these small fragments to optimize read length and assembly processes further reduces available DNA quantity. This poses challenges when working with samples with limited quantity and/or degraded DNA (Delahaye and Nicolas 2021). For this reason, alternative library preparation or sequencing techniques, including short-read sequencers (e.g., Illumina, Element Biosciences AVITI), are often preferred for handling samples with low molecular weight and/or low quantities of DNA.

To overcome these limitations, multiple displacement amplification (MDA) has emerged as a valuable and highly efficient method for amplifying small quantities of DNA. MDA has significant advantages over conventional PCR-based techniques such as random or degenerate oligonucleotide-primed PCR (DOP-PCR) and primer extension preamplification (PEP) (Hou et al. 2015). MDA advantages include reduced waste of rare samples since we can use lower amounts of DNA, isothermal amplification for efficiency, heightened sensitivity in detecting low amounts of DNA inputs, minimized coverage bias and error rates when compared to the other PCR-based methods and amplification of longer DNA fragments (Dean et al. 2002; Fullwood et al. 2008). MDA uses Phi29 DNA polymerase, which has a displacement activity that enables the isothermal amplification of DNA with high fidelity and exponential amplification of DNA molecules (Dean et al. 2002). This technique has been successfully applied in various genomic studies, including single-cell sequencing (Troell et al. 2016), ancient DNA analysis and microbiome studies (Binga et al. 2008; Lasken 2009). While a protocol for MDA with ONT ligation sequencing kits is available using the repli-G kit (Qiagen, Germany), the potential of MDA's application with ONT rapid kits, which offer faster processing times and less cost (FFPE DNA repair is not needed in the rapid kit protocol), though it yields relatively smaller fragments compared to ligation kits, has not been investigated. Consequently, MDA's potential limitations and impacts on whole-genome assembly in this context remain relatively unexplored.

The use of MDA combined with ONT sequencing has the potential to unlock genomic insights for organisms that are small to microscopic, especially those that are difficult or impossible to culture in vitro (e.g., *Cryptosporidium* species, *Mycobacterium leprae* and *Treponema pallidum*). Furthermore, clinical samples and isolates with limiting amounts of DNA pose a challenge for rapid and accurate genome sequence analysis, especially in urgent clinical situations where timely results are crucial. Working with degraded DNA samples becomes an issue since it could limit the sequence genomic coverage and assembly (Ceccherini et al. 2003; Dabney and Meyer 2019). While MDA effectively amplifies low DNA inputs, it presents significant challenges when applied to severely degraded DNA samples (Dabney and Meyer 2019). Specifically, MDA can negatively impact genomic analysis through (i) reduced efficiency due to DNA breaks or lesions, leading to incomplete amplification; (ii) coverage bias resulting in uneven genomic representation; (iii) potential contaminant interference with the amplification reaction (Wang et al. 2004); and (iv) the yield from MDA decreases as the length of the input DNA decreases (Lage et al. 2003).

Beyond these initial limitations, MDA introduces additional complexities in genomic sequencing, including (i) nonspecific amplification from primer dimer formation and template switching, (ii) chimeric DNA rearrangements and (iii) inherent genomic representation amplification bias that can compromise the accuracy and comprehensiveness of molecular analysis (Binga et al. 2008).

Some studies show that chimeric reads are usually inverted chimeras or direct chimeras, but it was previously observed that most of the detected MDA chimeric sequences (85%) are inverted chimeras, such as inverted sequences with intervening deletions, which can be caused by template switching (Lasken and Stockwell 2007; Lu et al. 2023). These chimeric sequences are known to affect genome sequencing because they can be considered amplification artefacts, which cannot be used for genome assembly (Lu et al. 2023). Studies suggest that chimerism in MDA sequencing data is a significant concern that is gaining attention, particularly with the rise of single-cell studies (Hard et al. 2023).

To address the challenges associated with artifactual concatemeric sequences generated during MDA, we developed a novel bioinformatic tool called CADECT (Concatemer Detection Tool), which is made available at https://github.com/rpbap/CADECT. This tool enabled the identification and removal of putative inverted chimeric concatemers, thus improving the accuracy and contiguity of the genome assembly.

Our study aims to provide valuable insights into the use of MDA for whole-genome ONT sequencing, particularly for low molecular weight and/or low quantities of DNA samples, highlighting its potential as a powerful method to obtain high-quality long-read sequencing data. We assessed the MDA advantages, constraints and effectiveness for whole-genome assembly in microbial organisms with genome sizes <10 Mb. This is especially significant for infectious disease agents, where obtaining enough DNA can be challenging. Overall, our study underscores the potential of MDA in enabling high-quality long-read sequencing from challenging low-concentration DNA samples, emphasizing its importance in various genomic research and clinical applications.

2 | Results

2.1 | MDA Efficiency Across Different Species

Our whole-genome amplification (WGA), using MDA, results reveal that in each sample type tested, we find an overall fold change of >500× in comparison to the input sample DNA (Table 1). Following amplification, approximately 1.5 μg of the product was debranched using T7 endonuclease prior to library preparation for ONT sequencing. Usually, in our assays, we obtained an ~45% recovery after this step, with the loss mostly attributed to the bead purification process (Table 1). Although a large amount of DNA is lost during the DNA purification step post T7 endonuclease reaction, an overall fold change of around 100× is observed compared to the MDA DNA input. The variation observed in the T7 recovery is affected by pipetting and the Ampure XP magnetic bead ratio that is set based on the DNA integrity from the QC. In this manuscript, we used a fixed ratio for all samples.

We successfully obtained contiguous and sometimes even chromosomal-level assemblies from the samples analyzed in this study, starting with MDA DNA inputs much lower than Oxford Nanopore's recommended minimum of 50 ng and 200 ng for the rapid barcode kit (RBK) for old and new library prep kits, respectively (Table S1).

For certain samples, such as *E. faecium*, we observed that achieving improved contiguity required generating higher depth coverage during the sequencing. Our results indicate that, for this organism, reaching depths beyond 70× allowed us to attain a chromosomal-level assembly with only 2.5 ng of starting total DNA (Table S2). In comparison, increased sequencing depth on samples that started with less than 0.001 ng of input into the MDA did not enhance contiguity. Combining separate MDA amplifications of the same limited input samples did improve the final genome coverage because the random nature of the initial templates and amplification process. Thus, multiple independent MDAs appears to be advantageous because it could randomly amplify by chance different regions that are beneficial for the genome assembly.

2.2 | MDA Using Serially Diluted Samples

Serial dilutions of a single *E. faecium* sample reveal successful DNA amplification even at a low initial DNA amount of 2.5E-5 ng (Figure 1A). The MDA technique imposes a size limit on its amplified products, with an average product length of 10–12 kb (Dean et al. 2002). The debranching step prior to library prep also leads to a reduction in the mean sequence read sizes (Table S3). Post MDA, the average size of the reads observed in our study was within the range of 2–3 kb. In contrast, standard Oxford Nanopore Technologies (ONT) assays without amplification (i.e., ONT RBK), which includes a transposase step that simultaneously cleaves template molecules and attaches tags to the cleaved ends, are expected to generate DNA fragments ranging from 5 to 20 kb. When assessing genome coverage from the tested samples such as *E. faecium*, we observed that DNA inputs below 0.025 ng result in incomplete coverage of certain genomic regions that are distributed randomly across the genomes tested (Figure 1B). The same is expected for the other samples (Troell et al. 2016; Chen et al. 2024; Penumarthi et al. 2024).

2.3 | MDA Results in an Unexpected Size Increase From Fragmented DNA Samples

Following controlled enzymatic fragmentation using a dsDNA fragmentase (to simulate degraded input DNA) and MDA according to the repli-G manufacturer protocol, we observed an unexpected size-increase distribution of fragments (Figure 2A). Indeed, for all samples, except *Cryptosporidium*, the size distribution post-MDA was nearly identical for intact and fragmented input DNA. Subsequent analysis of the ONT sequencing results revealed the existence of longer read lengths than those present in the same sample without MDA amplification (Figure 2B).

When analyzing the sequence content of the run (all reads generated), two distinct types of reads were identified. Some represented potentially low-abundance longer reads that escaped fragmentation during the enzyme incubation and were subsequently amplified. The other reads were primarily chimeric concatemers, likely generated through template switching of short fragments during MDA (Figure 3A). While the occurrence of concatemers in MDA assays has been reported previously (Paul and Apgar 2005; Lu et al. 2023), they are typically present in low amounts after sequencing. In our case, the fragmentation process seemed to enhance the prevalence of these chimeric reads in our ONT sequencing. As expected, assembly of the data revealed that the presence of these chimeric/concatemer regions significantly impacted genome assembly, resulting in a bubble

TABLE 1 | Observed amplification yield increase by sample type.

		MDA input (ng)	MDA output (ng)	T7 output (ng)	T7 recovery (%)	Estimated DNA output fold increase
Gram-positive	*S. aureus*	2.5	1500	894	59.6	357.6×
	E. faecium	2.5	1430	749	52.4	299.6×
Gram-negative	*E. coli*	5.0	8360	552	36.8	110.0×
Eukaryotic Pathogen (*Cryptosporidium* ssp.)		2.5	1976	555	44.1	222.0×
Positive control (Calf thymus)		5.0	3800	620	41.33	124.0×

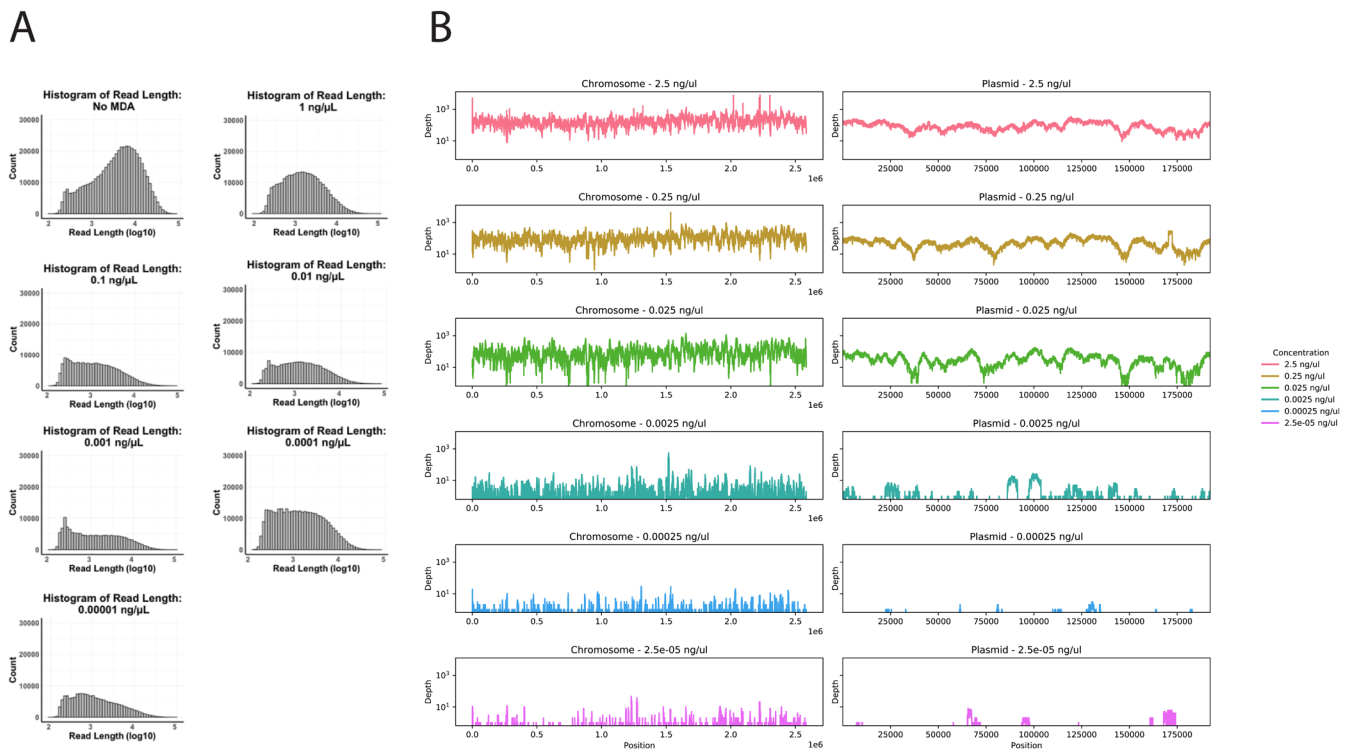

FIGURE 1 | Serial dilution test with *E. faecium* sample DNA. Panel (A) depicts the distribution and counts (*y*-axis) of read lengths (*x*-axis) across all diluted samples sequenced after MDA. The control sample had 400 ng used to see how a normal run in a sample without MDA would look like on read length distribution for comparison with the MDA diluted samples. Panel (B) illustrates the breadth of coverage of the chromosomal and plasmid regions across all diluted and subsequently amplified samples, with read depth on the *y*-axis and genome position on the *x*-axis.

fragmentation effect across the entire genome and affecting contiguity (Figure 3B).

2.4 | Concatemer Detection Tool

To identify and eliminate potential concatemers generated by MDA, we designed a concatemer detection tool specifically tailored for raw ONT reads called CADECT. This tool enables the differentiation of putative concatemeric chimeric reads from non-concatemeric ones. To achieve this, the process involves dividing each long-read sequence into multiple fragments using a sliding window approach and then aligning these fragments with one another using Biopython global pairwise2 (pairwise2.align.globalxx). The underlying hypothesis is that the presence of a concatemer would result in certain windows aligning with each other (presence of alignments with global alignment coverage greater than 0.7), thereby confirming the existence of a potential concatemer or tandem repeat within the sequenced read. No overlapping detection under the same parameters as mentioned above would be identified as non-concatemeric. Reads with lengths less than twice the given window size are categorized and stored as short reads. Additionally, it incorporates a size selection mechanism to isolate longer reads, thereby streamlining the genome assembly process and enhancing contiguity.

Following evaluation of the CADECT pipeline on fragmented DNA and a comparative analysis of results pre- and post-amplification assays (Table S1), we confirmed that the final genome assembly exhibited significantly reduced fragmentation. The integration of MDA and CADECT proved to be effective, particularly in handling challenging, low-quantity DNA samples. This combination facilitated the generation of nearly complete genome assemblies with depths above 70× (Figure 4; Table S4).

Overall, when comparing the data before and after CADECT using default parameters with a 500 base window size, we observe that it is a stringent process, which separates putative concatemers and shorter reads and tends to affect the average final depth of the final input. Specifically, in the case of *S. aureus*, we note that for high-quality intact amplified DNA, the detection of putative chimeras and size selection decreases coverage by 40%, whereas for amplified fragmented samples, it decreases coverage by 50% (Table S5). Due to size selection, the effect on depth is more pronounced for fragmented samples.

3 | Discussion

Our study demonstrates that MDA offers a promising solution for amplifying low amounts of DNA from precious samples for ONT sequence generation with the ONT rapid barcode sequencing Library kit (RBK). In this study, we demonstrated key points: (A) Using our method, we can successfully sequence samples with DNA inputs as low as 0.025 pg., corresponding to approximately 1.5 times the DNA content of a single *Cryptosporidium*

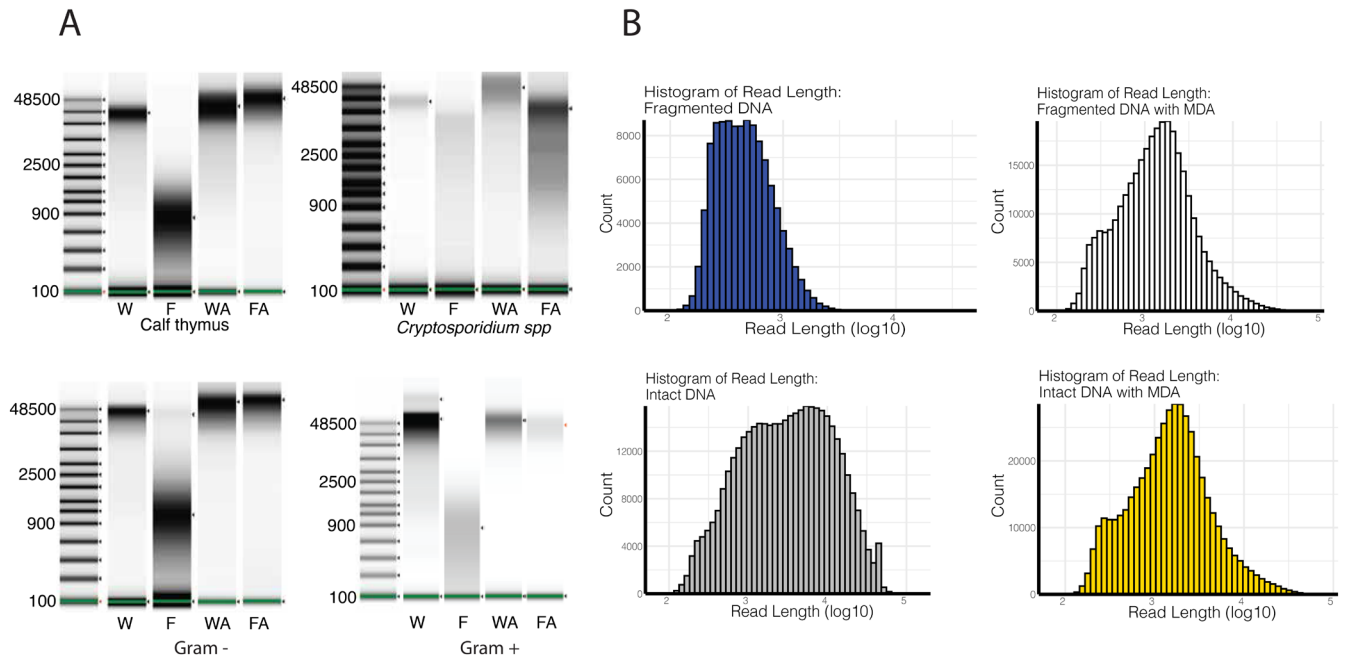

FIGURE 2 | DNA fragment and read size range pre- and post-multiple displacement amplification using size-controlled fragmented DNA. (A) TapeStation results for different organism sets. F = Fragmented DNA; W = whole intact DNA; WA and FA = after amplification. Uncropped TapeStation results are in Figure S1; (B) ONT sequencing read length results obtained before and after amplification for *S. aureus*.

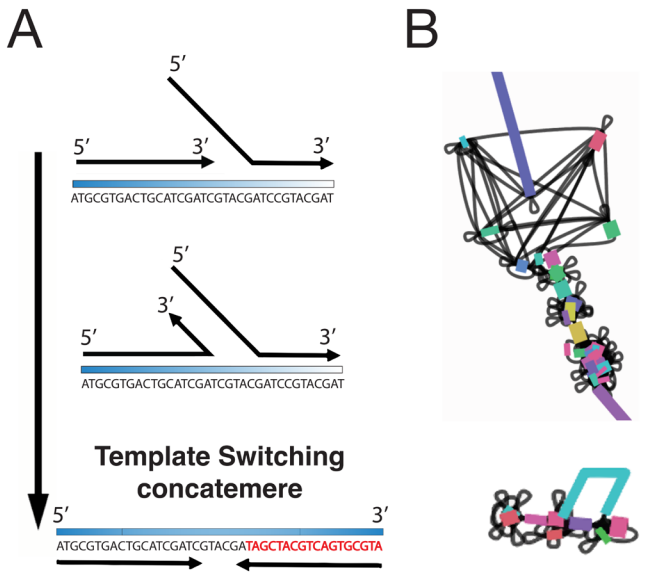

FIGURE 3 | Impact of MDA-generated concatemers on the genome assembly. (A) Concatemers generated by template switching; (B) Graph representation of the effect of concatemers on genome assembly (bubble fragmentation effect). Visualization made using BANDAGE v.9.0.

oocyst (genome size ~9 Mb; Table 2). For bacteria with genome sizes ranging from 3 to 7 Mb, this DNA amount would represent contributions from more cells. However, as shown in Figure 1 for bacteria, achieving whole genome and plasmid assemblies with whole-genome coverage typically requires approximately 2.5 pg. of input DNA, equivalent to more haploid genome equivalents. This demonstrates that while our approach has high sensitivity, generating complete genome assemblies is more reliable when working with DNA from more cells. Single-oocyst sequencing minimizes the variability associated with bulk sequencing approaches, providing more consistent genomic insights. Nonetheless, for applications requiring complete assemblies, starting with higher DNA concentrations is advisable to ensure robust results. (C) MDA using ONT RBK kits shows a good option to decrease the library prep time and cost, generating closed bacterial genomes and chromosomal-level genomes with a limited amount of DNA input. Here, we have only examined organisms with genome sizes < 10 Mb. Larger genome sizes will require additional starting material, and smaller genome sizes should have success with even less input DNA.

We were able to obtain whole-genome sequences at the chromosomal level for almost all tested organisms when generating depth coverages > 70×. This indicates that we can overcome the challenge of the relatively shorter reads that MDA with T7 debranching generates compared to reads generated from higher DNA input without amplification. Although the V14 RBK kits for R10.4.1 flow cells were not available at the time experiments were conducted, we have subsequently used the updated ONT V14 chemistry and confirmed that the protocol yields reproducible results, consistent with our earlier findings (data not shown). It is essential to highlight that highly complex regions, such as repetitive regions with tandem repeats larger than the window size used for CADECT detection, might be identified as concatemeric reads. This occurs because the tool detects repeat overlaps, which can lead to their exclusion before the assembly process, potentially causing some regions of the genome to remain fragmented. Thus, the judicious use of CADECT, including optimizing the window size, is warranted depending on the size and structure of repeats in the genome of the organisms being sequenced and assembled.

At a lower amount of initial DNA, we observed that the amplification appears to be random, exhibiting no apparent coverage

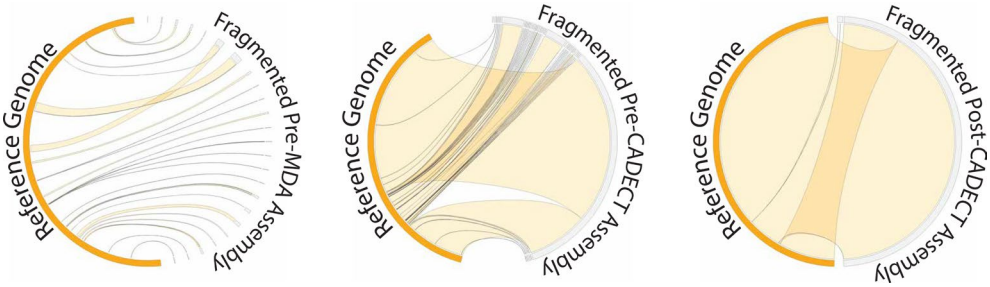

FIGURE 4 | Circos plot illustrating a synteny comparison between the reference *S. aureus* ATCC-29213 genome sequence and pre- and post-amplification genome assemblies. The Circos plots contrast the assemblies resulting from the amplified fragmented DNA before and after CADECT processing. A comparison between the reference genome and (A) genomic assembly of fragmented DNA sample without MDA, (B) genomic assembly of fragmented DNA sample with MDA before CADECT and (C) genomic assembly of fragmented DNA sample with MDA after CADECT. Grey lines are potential secondary alignments (multicopy region, repeats or concatemers).

TABLE 2 | Estimated DNA concentration in a single cell of the organisms studied in this project.

Sample	Estimated genome size (Mb)	Amt of DNA/ cell (fg)
E. coli	5.00	5.43
S. aureus	2.81	3.03
E. faecium	2.91	3.14
Cryptosporidium. ssp. (oocysts)*	9.2	39.70
Cryptosporidium ssp. (sporozoite)		9.90

*One *Cryptosporidium* ssp. oocyst contains four haploid sporozoites.

amplification bias across the genome (Figure 1). At extremely low DNA amounts, achieving full coverage of the target genome may be challenging, but the method remains valuable for potential taxon identification and may prove effective for the identification of plasmids as well (Figure 1). While the effectiveness in metagenomic samples requires further evaluation, there is promise in using this approach for taxon identification. Extremely low-abundance samples tend to produce patchy sequence information. Thus, although extremely low-concentration samples provide valuable sequence information, they also lack coverage in many regions, which impacts the assembly process and the ability to produce full genomic information. Combining multiple MDA replicates is likely to increase the chances of amplifying more regions and thus will be more likely to enhance genome coverage versus deeper sequencing.

In the case of sheared DNA, the higher impact on the depth after CADECT is primarily related to loss in the size selection pipeline (Table S5). However, our concatemer detection tool, CADECT, effectively identified and removed several concatemers, facilitating the assembly and yielding good results. This highlights the importance of bioinformatic tools in overcoming challenges associated with amplification artefacts, thus improving the accuracy of genome assembly.

It is worth noting that the CADECT pipeline will remove a significant number of reads, which will impact depth for an optimal genome assembly. If there is no sufficient coverage obtained post-CADECT run, an alternative is to merge the reads identified as short by the program with the non-concatemeric reads. As observed previously, the chimeric rate produced by MDA is positively associated with the mean read length (Lu et al. 2023), indicating a decreased likelihood of chimeric reads in this short dataset. Consequently, this dataset is less likely to negatively impact the assembly process. In more complex genome sequences that are rich in repeats, further investigation is required to address these regions effectively and be able to distinguish concatemers from genuine repetitive patterns within the genome. As a solution, the CADECT pipeline generates a separate concatemer fastq file. This file includes putative concatemeric regions as well as true repeats. For example, in highly repetitive genomes such as trypanosomatids with an ~50% genomic repeat content (El-Sayed et al. 2005), CADECT would detect a high number of reads containing real tandem repeats in the genome as putative concatemers, which would result in a higher impact on coverage depth loss and also impact the genome content of the organisms used for assembly. To mitigate this, we recommend incorporating a repeat identification step into the pipeline, such as using RepeatModeler (Flynn et al. 2020) trained with the organism of interest on the putative concatemer generated sequence file from CADECT. This additional step would enhance the recovery of information and data for the subsequent assembly process.

Moving forward, it is crucial to continue exploring the potential of MDA in various biological contexts and optimize the amplification protocol to minimize potential amplification biases and errors. Additionally, considering the clinical applications of MDA, further research and development of rapid and reliable sequencing approaches are necessary to unlock its full potential in diagnosing and monitoring infectious diseases and other clinical applications.

4 | Materials and Methods

4.1 | Sample Collection and Preparation

A 100-ng DNA sample of *Cryptosporidium meleagridis* isolate TU1867 was obtained from BEI Resources (NR-2521) (Manassas, VA). DNA samples from cultured *Staphylococcus aureus* ATCC-29213, *Enterococcus faecium* TX-1330, and *Escherichia coli* strain K12, which were available in our laboratory, were used

for testing. The bacterial DNA samples were prepared for downstream processing using the QIAGEN QIAamp DNA Mini Kit with lysozyme for Gram-positive samples and buffer ATL (tissue lysis buffer) for Gram-negatives. To assess sequence integrity, an *S. aureus* DNA sample aliquot was sheared using NEBNext dsDNA Fragmentase for 15 min to generate fragments approximately 1000 bp in size. All DNA samples were quality controlled using a TapeStation (Agilent Technologies, Santa Clara, CA) and Qubit (ThermoFisher Scientific, Waltham, MA). In addition, we conducted serial dilutions on all samples to assess the limit of detection for amplification in the assay. The dilutions ranged from 2.5E-5 ng to 2.5 ng, allowing us to determine the minimum concentration at which successful amplification could be achieved.

4.2 | Multiple Displacement Amplification (MDA)

Prior to MDA, the concentration of DNA was obtained using a Qubit fluorometer dsDNA high-sensitivity assay kit (ThermoFisher, Waltham, MA). For the *C. meleagridis* DNA, three different amounts were used as input for whole-genome amplification (i.e., 2, 5 and 10 ng) in a final volume of 5 μL. For the bacterial samples, MDA was performed on 2.5 ng of fragmented intact *S. aureus DNA* as well as serial dilutions of DNA from *E. faecium* ranging from 2.5 ng to 2.5E-5 ng. 400 ng of non-amplified DNA was used as an input control for the ONT rapid kit library preparation (Oxford Nanopore Technologies, Oxford, United Kingdom).

MDA was performed using the Qiagen repli-G kit (CAT #150023, Qiagen, Hilden, Germany), following the manufacturer's instructions. Following this, concentrations of DNA were obtained using a Qubit fluorometer dsDNA high-sensitivity assay kit (ThermoFisher, Waltham, MA).

For T7 Endonuclease I debranching, up to 42 μL (i.e., all product from the MDA reaction) of MDA DNA was used as input (Catalogue #M0302, New England BioLabs, Ipswich, MA). After scaling up the reaction to accommodate a 42-μL input, all reaction components were added following the manufacturer's guidelines and incubated for 15 min at 37°C in a BioRad T100 thermocycler (BioRad, Hercules, CA). The incubated reaction was brought up to a final volume of 50 μL using TE buffer pH 8. AMPure XP beads (CAT# A63880) were prepared ahead of time following the manufacturer's instructions, and 35 μL of beads were added to the reaction and mixed thoroughly. The bead-reaction mixture was placed on a rotator mixer (e.g., Hula mixer) for 10 min at room temperature. Following this, the bead-reaction mixture was spun down and placed on a magnet until the eluate was clear and colourless. With the bead-reaction mixture on the magnet, the clear supernatant was pipetted off and 200 μL of freshly prepared 70% ethanol was carefully added not to disturb the pellet (i.e., wash step). The wash step was repeated one time, for a total of two washes. After removing the supernatant from the second wash, 49 μL of water was used to resuspend the pellet, which was immediately incubated for 1 min at 50°C in a BioRad T100 thermocycler, followed by 5 min at room temperature. The bead-reaction mixture was placed back on the magnet, and 49 μL of the elute was transferred to a sterile 1.5-mL tube. Concentrations of DNA were obtained using a Qubit fluorometer dsDNA high-sensitivity assay kit (ThermoFisher, Waltham, MA). A summarised diagram of the protocol is provided (Figure S2).

4.3 | Whole-Genome Sequencing and Assembly

Sequencing of the amplified DNA samples was performed using the ONT SQK-RBK110.96 kit for library preparation R9.4.1 MinION flow cells (Oxford Nanopore Technologies, Oxford, UK). The amplified DNA samples were prepared according to the kit instructions and loaded onto the flow cell for sequencing following the manufacturer's instructions. Sequencing was carried out in Mk1B and GridION MK1 devices for 72 h, and the resulted fast5 files were base-called using guppy v6.3.7 using the high accuracy model (dna_r9.4.1_450bps_hac). By sequencing for 72 h, we aim to recover as much data as possible. The real-time sequencing capability of ONT allowed us to access data as it is generated, which is particularly advantageous in urgent cases like clinical assays. When it is necessary to pellet cultured bacteria to obtain sufficient DNA, this approach enables sequencing without requiring prolonged cultivation periods. Furthermore, we are able to commence data analysis as soon as the sequencing reaches the desired depth, as monitored by the MinKNOW ONT software.

Flye 2.9 (Kolmogorov et al. 2019) was used for assembly. For samples with > 100× coverage, the '—asm-coverage 100' parameter was used to improve assembly and facilitate the assembler's performance. We then used Nextpolish 1.4.1 (Hu et al. 2020) to increase the overall base call quality of the genome and facilitate further quality control analysis such as Benchmarking Universal Single-Copy Orthologs (BUSCO) scores and better gene annotation. Illumina sequencing was not used here because the objective of this research was to determine how MDA would affect long-read generation.

4.4 | Putative Concatemer Detection in Intact Versus Fragmented Amplified Samples

To examine the impact of fragmentation on MDA products, we treated DNA aliquots with NEBNext dsDNA Fragmentase (CAT# M0348S) at 37°C for 16 min, creating fragments between 500 and 1000 bp. This enzyme-based method induces DNA shearing, generating fragments of specified sizes in a time-dependent manner. The process provides random fragmentation similar to mechanical methods. Both fragmented and non-fragmented (high molecular weight DNA) samples were sequenced as described above.

The CADECT tool (https://github.com/rpbap/CADECT) was developed in-house and was used for the detection and removal of putative concatemeric chimeric sequences in the ONT amplified reads. CADECT splits all reads into separate files and performs sliding windows with a user-defined preferred size and gap between windows. For ONT amplified reads, a window size of ≥ 500 bp with no overlaps was used (e.g., −w 500 and -s 500). Reads generating less than one window (< 1 kb in size in the 500 bp window example) were skipped, and their IDs were stored in the short.txt output file. Fragment windows from reads with more than two windows were aligned using nucmer from mummer 4 (Marcais et al. 2018), and reads with overlaps were reported in the stats file, with their IDs stored in the concat_ID output file. Statistics including the total number of reads, number of putative concatemers, number of reads with no concatemer detection and overlap frequency were recorded in the stats.txt

output file. Fastq/Fasta files containing the characterized reads were generated for further analysis.

These methods were employed to investigate the benefits and limitations of MDA in whole-genome ONT, focusing on low-concentration DNA samples.

Author Contributions

F.A.-D., M.S.B., A.D., H.S., M.Q.D. and R.P.B. performed the research. M.S.B., J.C.K., T.C.G. and R.P.B. conceived the research. F.A.-D., M.S.B., A.D., M.Q.D. and R.P.B. analysed the results. F.A.-D., M.S.B., A.D., R.T., A.K., C.A.A., J.C.K., T.C.G. and R.P.B. contributed to writing the manuscript. C.A.A., J.C.K., T.C.G. and R.P.B. obtained funding. All authors reviewed and approved the manuscript.

Acknowledgements

This work was funded in part by the National Institute for Allergy and Infectious Diseases, NIAID, R01 AI148667 to T.C.G. and J.C.K., NIH T32GM142623 to F.A.-D. The funders played no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The raw sequence data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession numbers PRJNA1063853 and PRJNA1022047. The assembled genomes in this project are preliminary drafts and are currently unavailable at this stage due to the scope of this project. They were generated solely using ONT reads and have not undergone polishing with Illumina short-read data. Additionally, they have not been checked for potentially contaminating 'leftover' contigs. The raw data are accessible for reproduction purposes, and the final, polished and decontaminated assemblies will be made available in subsequent publications.

References

Binga, E. K., R. S. Lasken, and J. D. Neufeld. 2008. "Something From (Almost) Nothing: The Impact of Multiple Displacement Amplification on Microbial Ecology." *ISME Journal* 2: 233–241.

Ceccherini, M., J. Pote, E. Kay, et al. 2003. "Degradation and Transformability of DNA From Transgenic Leaves." *Applied and Environmental Microbiology* 69: 673–678.

Chen, Y., J. Huang, H. Qin, et al. 2024. "Chromosome-Level Genome Assembly of *Cryptosporidium parvum* by Long-Read Sequencing of Ten Oocysts." *Scientific Data* 11: 1287.

Dabney, J., and M. Meyer. 2019. "Extraction of Highly Degraded DNA From Ancient Bones and Teeth." *Methods in Molecular Biology* 1963: 25–29.

Dean, F. B., S. Hosono, L. Fang, et al. 2002. "Comprehensive Human Genome Amplification Using Multiple Displacement Amplification." *Proceedings of the National Academy of Sciences of the United States of America* 99: 5261–5266.

Delahaye, C., and J. Nicolas. 2021. "Sequencing DNA With Nanopores: Troubles and Biases." *PLoS One* 16: e0257521.

El-Sayed, N. M., P. J. Myler, D. C. Bartholomeu, et al. 2005. "The Genome Sequence of Trypanosoma Cruzi, Etiologic Agent of Chagas Disease." *Science* 309: 409–415.

Flynn, J. M., R. Hubley, C. Goubert, et al. 2020. "RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families." *Proceedings of the National Academy of Sciences of the United States of America* 117: 9451–9457.

Fullwood, M. J., J. J. Tan, P. W. Ng, et al. 2008. "The Use of Multiple Displacement Amplification to Amplify Complex DNA Libraries." *Nucleic Acids Research* 36: e32.

Hard, J., J. E. Mold, J. Eisfeldt, et al. 2023. "Long-Read Whole-Genome Analysis of Human Single Cells." *Nature Communications* 14: 5164.

Hou, Y., K. Wu, X. Shi, et al. 2015. "Comparison of Variations Detection Between Whole-Genome Amplification Methods Used in Single-Cell Resequencing." *GigaScience* 4: 37.

Hu, J., J. Fan, Z. Sun, and S. Liu. 2020. "NextPolish: A Fast and Efficient Genome Polishing Tool for Long-Read Assembly." *Bioinformatics* 36: 2253–2255.

Hu, T., N. Chitnis, D. Monos, and A. Dinh. 2021. "Next-Generation Sequencing Technologies: An Overview." *Human Immunology* 82: 801–811.

Kolmogorov, M., J. Yuan, Y. Lin, and P. A. Pevzner. 2019. "Assembly of Long, Error-Prone Reads Using Repeat Graphs." *Nature Biotechnology* 37: 540–546.

Lage, J. M., J. H. Leamon, T. Pejovic, et al. 2003. "Whole Genome Analysis of Genetic Alterations in Small DNA Samples Using Hyperbranched Strand Displacement Amplification and Array-CGH." *Genome Research* 13: 294–307.

Lasken, R. S. 2009. "Genomic DNA Amplification by the Multiple Displacement Amplification (MDA) Method." *Biochemical Society Transactions* 37: 450–453.

Lasken, R. S., and T. B. Stockwell. 2007. "Mechanism of Chimera Formation During the Multiple Displacement Amplification Reaction." *BMC Biotechnology* 7: 19.

Lu, N., Y. Qiao, Z. Lu, and J. Tu. 2023. "Chimera: The Spoiler in Multiple Displacement Amplification." *Computational and Structural Biotechnology Journal* 21: 1688–1696.

Marcais, G., A. L. Delcher, A. M. Phillippy, R. Coston, S. L. Salzberg, and A. Zimin. 2018. "MUMmer4: A Fast and Versatile Genome Alignment System." *PLoS Computational Biology* 14: e1005944.

Pacbio. 2022. "Technical Note Preparing DNA for PacBio HiFi Sequencing—Extraction and Quality Control."

Paul, P., and J. Apgar. 2005. "Single-Molecule Dilution and Multiple Displacement Amplification for Molecular Haplotyping." *BioTechniques* 38: 553–554.

Penumarthi, L. R., R. P. Baptista, M. S. Beaudry, T. C. Glenn, and J. C. Kissinger. 2024. "A New Chromosome-Level Genome Assembly and Annotation of *Cryptosporidium meleagridis*." *Scientific Data* 11, no. 1: 1388. https://doi.org/10.1101/2024.02.16.580748.

Slatko, B. E., A. F. Gardner, and F. M. Ausubel. 2018. "Overview of Next-Generation Sequencing Technologies." *Current Protocols in Molecular Biology* 122: e59.

Troell, K., B. Hallstrom, A. M. Divne, et al. 2016. "Cryptosporidium as a Testbed for Single Cell Genome Characterization of Unicellular Eukaryotes." *BMC Genomics* 17: 471.

Wang, G., E. Maher, C. Brennan, et al. 2004. "DNA Amplification Method Tolerant to Sample Degradation." *Genome Research* 14: 2357–2366.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.