PLOS One

LAB PROTOCOL

High-performance protocol for ultra-short DNA sequencing using Oxford Nanopore Technology (ONT)

Lukas Žemaitis[1☯], **Rūta Palepšienė**[1,2*☯], **Simonas Juzėnas**[1], **Gediminas Alzbutas**[1,3], **Pierre-Yves Burgi**[4], **Thomas Heinis**[5], **Jérôme Charmet**[6], **Silvia Angeloni Suter**[6], **Martin Jost**[7], **Renaldas Raišutis**[2], **Friedrich Simmel**[8], **Ignas Galminas**[1,9]

1 Department of DNA data storage, Genomika, Kaunas, Lithuania, **2** Ultrasound Research Institute, Kaunas University of Technology, Kaunas, Lithuania, **3** Institute for Digestive Research, Academy of Medicine, Lithuanian University of Health Sciences, Kaunas, Lithuania, **4** Department of Information and Communication Systems and Technologies, University of Geneva, Geneva, Switzerland, **5** Department of Computing, Imperial College London, London, United Kingdom, **6** Western Switzerland, Neuchâtel, Switzerland, **7** MABEAL GmbH, Graz, Austria, **8** Department of Bioscience, TU Munich, School of Natural Sciences, Garching, Germany, **9** Faculty of Natural Sciences, Vytautas Magnus University, Kaunas, Lithuania

☯ These authors contributed equally to this work.
* rutap@genomika.lt

Citation: Žemaitis L, Palepšienė R, Juzėnas S, Alzbutas G, Burgi P-Y, Heinis T, et al. (2025) High-performance protocol for ultra-short DNA sequencing using Oxford Nanopore Technology (ONT). PLoS One 20(4): e0318040. https://doi.org/10.1371/journal.pone.0318040

Editor: Isabelle Chemin, Centre de Recherche en Cancerologie de Lyon, FRANCE

Received: December 12, 2024

Accepted: March 27, 2025

Published: April 29, 2025

Data availability statement: The data and code associated with this study are available on the GitHub repository (https://github.com/genomika-lt/high-performance) and archived on Zenodo (https://zenodo.org/records/14244834).

Abstract

In recent years, Oxford Nanopore Technologies (ONT) has gained substantial attention across various domains of nucleic acid research, owing to its unique advantages over other sequencing platforms. Originally developed for long-read sequencing, ONT technology has evolved, with recent advancements enhancing its applicability beyond long reads to include short, synthetic DNA-based applications. However, sequencing short DNA fragments with nanopore technology often results in lower data quality, likely due to the absence of protocols optimised for these fragment sizes. To address this challenge, we refined the standard ONT library preparation protocol to improve its performance for ultra-short DNA targets. By utilising the same core reagents required for conventional ONT workflows, we introduced targeted alterations to enhance compatibility with shorter fragment lengths. We then benchmarked these adjustments against libraries prepared using the standard ONT protocol. Here, we present a comprehensive, step-by-step protocol that is accessible to researchers of various technical expertise, facilitating high-quality sequencing of ultra-short DNA fragments. This protocol represents a significant improvement in sequencing quality for short DNA sequences using ONT technology, broadening the range of possible applications.

Funding: This research was funded by the European Innovation Council Pathfinder Program, under the project DNAMIC (DNA Microfactory for Autonomous Archiving), Grant Agreement No. 101115389. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Oxford Nanopore Technologies has revolutionized sequencing by offering solutions that combine ease of operation, portability, accessibility, and rapid data acquisition. Unlike traditional sequencing technologies, the ONT platform does not require elaborate sample preparation or complex sequencing infrastructure, allowing researchers to analyse DNA in real-time and in various settings [1]. These appealing features have made nanopore sequencing increasingly popular across multiple domains. Beyond its widespread use in genomics and life sciences, there is growing interest in the application of nanopore sequencing in emerging fields such as DNA data storage [2,3], DNA origami [4], and nanomaterials science [5], where DNA is utilized for both data encoding and as a structural component.

These emerging applications rely on short (< 100 bp) or ultra-short (20–40 bp) [6] synthetic DNA, a constraint imposed by the phosphoramidite chemistry typically employed in oligonucleotide synthesis [7]. However, the ONT long-read sequencing platform encounters challenges when sequencing these oligonucleotides, often leading to substantial underutilisation of flow cells with reduced sequencing output and quality [8]. To counter these limitations, some strategies have focused on sample barcoding and multiplexing to maximise flow cell utilisation [9] while others have attempted to concatenate ultra-short fragments into longer stretches, improving both efficiency and read quality [10]. Nevertheless, these strategies introduce additional steps, complicating and extending the library preparation process, and certain applications require native sequencing rather than concatenated forms. These challenges highlight the need for further optimisations of the current protocols to make ONT more effective for applications involving short oligonucleotide sequencing.

Here, we introduce an optimised, step-by-step protocol for sequencing DNA fragments as short as 40 bp using ONT. To develop this protocol, we systematically evaluated various modifications to the standard ONT protocols and identified six key steps that significantly enhance ONT performance for sequencing ultra-short DNA. We then benchmarked our protocol against the standard ONT protocol, demonstrating that by refining the standard Ligation Sequencing Kit V14, our approach achieves over ten times the sequencing output without requiring additional reagents beyond those used in standard library preparation protocol.

The high reproducibility and streamlined nature of our protocol make it a compelling alternative to standard ONT protocol, particularly for applications involving short DNA fragments. Previously, the limited output for short fragments posed significant challenges, making the technology inefficient and costly. Our approach, however, transforms this paradigm, offering a solution that overcomes these limitations. What is more, due to its simplicity and low complexity, this protocol can be seamlessly integrated into workflows across various laboratories, regardless of the technical background of the users. This broad applicability supports expanding ONT into research fields where the sequencing of ultra-short DNA fragments has traditionally posed challenges. By offering a significant boost in sequencing output and accuracy for short DNA sequences, this protocol has the potential to improve ONT usability and integrate it into new scientific domains, enhancing the quality of research outcomes.

Materials and methods

DNA preparation

All experiments were performed using ultra-short DNA sequences of 40 bp. These sequences were synthesised as complementary ssDNA oligonucleotides using an on-site DNA synthesiser (Kilobaser) [11] and subsequently dissolved in 20 μL of Milli-Q water (Thermo Fisher Scientific). Concentration measurements were conducted using a Qubit 4 Fluorometer with the Qubit ssDNA assay kit (Thermo Fisher Scientific). Sequences used in the study were:

40_TOP 5' – CTAACCCGACCTAATATGAACTCCTGCCTACAATGACCCT – 3'

40_BOT 5' – AGGGTCATTGTAGGCAGGAGTTCATATTAGGTCGGGTTAG – 3'.

To obtain dsDNA fragments, 23 μM of each oligonucleotide was combined with 1 μL of T4 ligase buffer (10x) (Thermo Fisher Scientific) and Milli-Q water, resulting in a final reaction volume of 10 μL. The reaction mixture was denatured at 95°C for 5 minutes, followed by gradual cooling at a rate of 5°C/min until reaching 25°C to promote annealing. After the reaction, the concentration of the dsDNA duplex was measured using the Qubit HS dsDNA assay kit (Thermo Fisher Scientific).

Library preparation for ONT

Library preparation was performed using the Ligation Sequencing Kit with V14 chemistry (SQK-LSK114, ONT). Pre-annealed 40 bp fragments were subjected to library preparation by following the standard ONT protocol according to the manufacturer's guidelines [12] (ACDE_9163_v114_revS_29Jun2022 protocol version) or a High-performance protocol, which is being proposed in this study. The protocol described in this peer-reviewed article is published on protocols.io (dx.doi.org/10.17504/protocols.io.dm6gp973jvzp/v1) and is included for printing purposes as S1 File.

In brief, an increased DNA input of 250 fmol of dsDNA duplex was phosphorylated at the 5' end and adenylated at the 3' end, following the manufacturer's instructions. Without purification, ligation adapters were then attached to the processed samples using the Quick T4 DNA Ligation Module (New England Biolabs) for extended reaction time (20 mins). Following ligation, the samples were purified using AMPure XP beads with an increased beads-to-DNA ratio of 1.8x and subsequently eluted.

Samples prepared using the standard ONT protocol exhibited library concentrations below measurable limits; therefore, 10 μL of the prepared library was used for sequencing. In case samples prepared with the High-performance protocol, 8–10 μL (40 ± 3 fmol) of the DNA library was loaded onto the flow cell to ensure adequate pore occupancy. This was followed by a 10-minute incubation period prior to initiating the sequencing experiment.

Sequencing

Sequencing was conducted using R10.4.1 Flongle flow cells (ONT) and a MinION Mk1B sequencing device (ONT), adhering to the default parameters specified for the ligation sequencing kit. The sequencing runs were performed for a duration of 1.5 hours, with raw signal data collected in POD5 format and subjected to basecalling using Dorado (v0.7.2) basecaller with simplex v5.0 SUP (dna_r10.4.1_e8.2_400bps_sup@v5.0.0) basecalling model from ONT. The resulting sequences in FASTQ files were categorised into "passed" or "failed" based on their average quality scores (q-scores), which were calculated by employing fx2tab from SeqKit (v2.8.2) [13]. Only reads with q-scores above or equal to the threshold of 9, which is the default value for MinKNOW software, were included in subsequent mapping and clustering analyses.

Bioinformatic analysis

Mapping. Reads that passed the q-score threshold were aligned to a reference sequence (40_TOP) using the LAST (v1542) [14] aligner with local alignment mode and default parameters. To extract the number of mismatches (NM) and the

CIGAR string, the alignments in MAF were converted to SAM format using maf-convert from the LAST toolkit. Sequences with 3 or fewer NM and an aligned length of 37–40 bases to the reference were considered mapped. To obtain error profiles from the mapped reads, a custom script was used to parse the CIGAR string to calculate substitution, deletion and insertion rates per position relative to the reference sequence, excluding hard-clipped bases.

Clustering. Raw sequencing reads with the adapters and an average quality score ≤9 or lengths shorter than 50 bp or longer than 300 bp were discarded. High-quality reads were dereplicated using VSEARCH (v2.22.1) [15]. The dereplicated reads were clustered using Swarm (v3.1.5) [16] with a local clustering distance threshold of d = 10. Clustered sequences were aligned using MAFFT (v7.525) [17]. Alignments were trimmed using TrimAl (v1.5.rev0) [18]. After trimming, sequences were reclustered using Swarm with the same parameters (d = 10), and representative sequences were selected. Representative sequences were compared against reference sequences using BLASTn (v2.15.0+) [19] with the blastn-short task. Data analysis and visualisation were performed using custom R scripts in R (v4.3.3). The analysis pipeline was managed using Snakemake (v8.11.3) [20].

Statistical analysis

Statistical analysis was performed using R (v4.4.0) [21] and ggplot2 (v3.5.1) [22] was used for visualization. Data are presented as the mean ± standard error (standard deviation divided by the square root of the sample size) or as the median, including the first and third quartiles and a 1.5 × interquartile range. Statistical significance between the two sample groups was evaluated using an unpaired Student's t-test with a "greater" one-sided alternative hypothesis.

Expected results

Recent advancements in Oxford Nanopore Technologies suggest that it is possible to sequence DNA fragments as short as 20 bp. Yet, previous studies have highlighted a notable decline in ONT performance with short DNA fragments [8,10] which may be further aggravated when targeting ultra-short DNA. Therefore, we selected an ultra-short DNA fragment of 40 bp and aimed to address key questions: i) evaluating the performance of ONT for ultra-short DNA fragments; ii) determining how adjustments of the current protocols might enhance sequencing quality for such fragments.

This article describes a refined approach, termed the High-performance protocol, which was rigorously evaluated through benchmarking against the standard ONT library preparation protocol, focusing on key sequencing performance metrics, including sequencing throughput, quality, mapping and error rates. The study design, alongside the methodological distinctions between the protocols, is outlined in Fig 1A.

Our primary goal was to determine whether overall sequencing quality (generated output and read quality) varies between protocols. Analysis of fragment length distributions revealed a peak at approximately 85 bp for both the standard and High-performance protocols, aligning with the intended length of the annealed DNA duplex in combination with attached ONT sequencing adapters (Fig 1B). This consistency shows that both protocols successfully prepared libraries of the intended fragment size. A notable distinction between protocols emerged in output efficiency: the High-performance protocol yielded approximately 13 times more reads than the standard protocol (Fig 1C). Specifically, the High-performance protocol produced an average of 33252 raw reads, compared with 2551 raw reads for the standard protocol over the same sequencing period, underscoring its superior throughput.

The obtained raw reads were then classified into high-quality and low-quality categories based on quality scores, as described in the Methods and materials section. It was observed that both protocols predominantly yielded high-quality reads. For the High-performance protocol approximately 77% the total reads (25669 out of 33252) were classified as a high-quality, while for the standard protocol, high-quality reads composed approximately 68% (1740 out of 2551) (Fig 1D). The statistically significant difference ($p = 0.0286$) in the number of high-quality reads between the protocols primarily reflects the substantially higher initial read count generated by the High-performance protocol. However, these results also indicate that the High-performance protocol can achieve a marked increase in read counts without compromising read

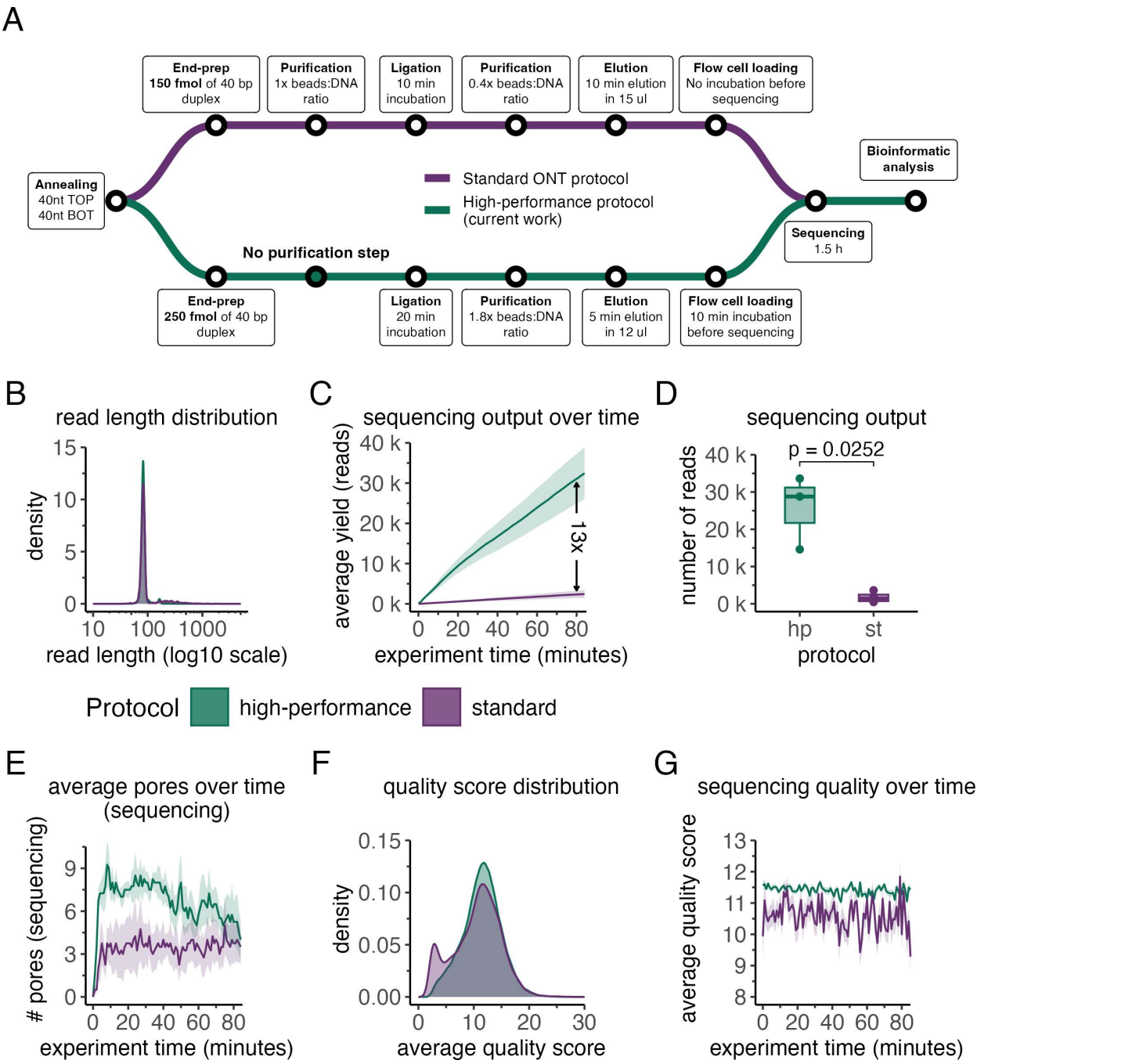

Fig 1. The High-performance enhances sequencing efficiency compared to standard library preparation in ONT sequencing. A) Detailed schematics of the experimental design and the key differences in library preparation methods to illustrate the workflow and methodology. B) Sequencing read length distributions for libraries using standard (purple) and High-performance (green) protocols. C) Average cumulative throughput (reads) and standard error (SE; y-axis) versus experiment duration (sequencing) (x-axis). D) Number of "passed" reads (q-score ≥ 9; y-axis) generated using High-performance and standard library preparation protocols (x-axis). The boxplots show the median (centre line), first and third quartiles (lower and upper hinges), and 1.5 × interquartile range (lower and upper whiskers). Statistical significance between the protocols was evaluated using an unpaired Student's t-test with a one-sided "greater" alternative hypothesis. E) Number of pores in the sequencing state (y-axis) throughout the experiment duration (x-axis). The line graph shows the mean ± SE of pore numbers in the sequencing state for both High-performance and standard libraries. F) Q-score distributions of libraries prepared by standard (purple) and High-performance (green) protocols. G) Average q-scores with SE of sequencing reads (y-axis) over the experiment time (x-axis) generated by both the High-performance and standard libraries. The results represent three (N = 3) and four (N = 4) independent sequencing experiments by using the High-performance and standard ONT protocol, respectively.

https://doi.org/10.1371/journal.pone.0318040.g001

quality. To investigate the underlying cause of the substantially higher read count observed with the High-performance protocol, we hypothesised that it may be due to the amount of DNA input loaded onto the flow cell. As outlined in the Materials and methods section, the precise quantification of DNA was not feasible for the standard ONT protocol due to the low concentration. In contrast, the High-performance protocol consistently produced purified DNA libraries with measurable concentrations, of which ~40 fmol, were loaded for sequencing. The higher concentration of DNA molecules would be expected to increase pore occupancy, thus generating a higher read count [23]. To test this hypothesis, we monitored the number of actively sequencing pores throughout the experiment (Fig 1E). Consistent with our hypothesis, the High-performance protocol maintained, on average, more than twice the number of pores in the active state as compared to the standard protocol. In addition to the higher DNA concentration, the High-performance protocol includes an extra 10-minute incubation period before sequencing. This step may provide more time for tether molecules to bring DNA closer to the membrane, improving the DNA capture rate by the nanopore [24].

As the quality score of reads is as critical a metric as overall output, we conducted a more detailed analysis. A comparison of the average read quality scores between protocols revealed that both protocols produced reads with similar quality scores, centred around 11 (q-score = 11) (Fig 1F). However, the distribution pattern of read quality differed. The High-performance protocol generated a relatively low proportion of reads with extreme quality values ($q \leq 5$ or $q \geq 15$), with most reads displaying intermediate quality, as expected for such short fragments. In contrast, the standard ONT protocol demonstrated a marked increase in reads with low quality scores of $q \leq 5$. This discrepancy may be partially attributed to the generation of reads with highly variable quality scores over the course of the experiment (Fig 1G) and to sequencing artifacts, which are unexpectedly long reads exhibiting low sequence diversity (S1 Fig). This was mostly observed in samples prepared using standard ONT protocol while High-performance protocol maintained relatively stable quality scores and read length throughout the sequencing period. The factors contributing to this variability in samples prepared according to the standard protocol are not yet fully understood and will require further study.

Previous studies have indicated that ONT typically generates error rates of approximately 3–10% in case of long reads (>1 kb) [25,26]. This error rate is expected to increase with ultra-short DNA fragments, potentially due to the rapid translocation speed of DNA through the nanopore which can hinder the precise interpretation of signals and accurate base identification during basecalling [1]. Consequently, in the next phase of this study, we evaluated the sequencing accuracy and compared potential differences between samples prepared using the standard and High-performance protocols.

To conduct this analysis, we aligned high-quality ($q\text{-score} \geq 9$) reads to a reference sequence and assessed accuracy from multiple perspectives. As a first step, we examined the average length of the mapped reads (Fig 2A). Notably, the High-performance protocol produced a substantially higher number of full-length reads compared to the standard protocol, which generated relatively fewer full-length reads. This difference suggests that the High-performance protocol may provide better coverage of the target sequence, improving downstream analyses.

Further distinctions between the protocols became evident when quantifying the total number of mapped reads (Fig 2B). The High-performance protocol yielded significantly more mapped reads than the standard protocol ($p = 0.003$), averaging 16320 reads compared to 874. This difference largely reflects the higher total number of high-quality reads generated by the High-performance protocol. However, when we evaluated the mapping fraction (i.e., the proportion of mapped reads relative to the total number of high-quality reads), additional factors emerged. The samples prepared by the High-performance protocol achieved a notably higher mapping fraction of approximately 60%, compared to only about 45% for the standard protocol, indicating significantly improved mapping efficiency. To test If this discrepancy might be related to overall lower quality scores in reads generated from samples prepared with the standard ONT protocol (as indicated in Fig 1F and G) we analysed mapping quality over all 85-minute sequencing periods for each protocol (Fig 2D). In accordance with our observation, mapping quality scores differed significantly between protocols: the High-performance protocol maintained consistently high mapping quality, while the standard protocol exhibited fluctuations of greater amplitude. Positional error rate profiles were also compared between protocols (Fig 2E), revealing similar overall error rates,

Fig 2. Sequencing reads generated by the High-performance protocol yield a higher proportion of correct, full-length sequences. A) Probability densities (y-axis) of mapped sequence lengths relative to the reference length (x-axis) for libraries prepared with both the standard protocol (purple) and the High-performance protocol (green). B) Number of reads (y-axis) mapped to the reference sequence with 3 or fewer mismatches for data generated by High-performance and standard library preparation protocols (x-axis). The bar plot displays the mean ± SE in read numbers. C) Fraction of mapped reads relative to the total number of passed reads (y-axis) for sequencing libraries prepared using High-performance and standard library preparation protocols (x-axis). The boxplots display the median (centre line), first and third quartiles (lower and upper hinges), and 1.5 × interquartile range (lower and upper whiskers). D) Average mapping-quality (MAPQ) values ± SE of passed sequencing reads (y-axis) over the experiment duration (x-axis) generated by both the High-performance and standard libraries. E) Positional fractions of deletions, insertions and mismatches (y-axis) with respect to the reference sequence (x-axis) for libraries prepared by High-performance (green) and standard (purple) library preparation protocols. The line graph displays the mean values for the fraction of total passed reads for the High-performance and standard libraries. The results represent three (N = 3) and four (N = 4) independent sequencing experiments by using the High-performance and standard ONT protocol, respectively.

https://doi.org/10.1371/journal.pone.0318040.g002

with only a slightly higher error rate observed in reads generated by the standard protocol. Finally, we investigated the structural composition of the reads using sequence clustering analysis. As presented in Fig 3, reads from both protocols exhibited the expected structure, with the target sequence centred and flanked by adapter fragments. In the case of samples prepared using the High-performance protocol (Fig 3A), the largest cluster comprised 43% of the total reads (32936 reads), while the second-largest cluster accounted for 27% (21130 reads). The largest cluster contained the correct insert in the forward orientation (plus strand), whereas the second-largest cluster corresponded to the same insert just in reverse complement orientation (minus strand). The data obtained from the libraries prepared using the standard protocol (Fig 3B) revealed a similar clustering pattern, but with reduced structural homogeneity: the largest cluster presented 35% of the total reads (2361 reads), corresponding to the plus strand, while the second-largest cluster accounted for 20% (1309 reads) and corresponded to the minus strand. These results highlight the structural homogeneity differences between the two protocols.

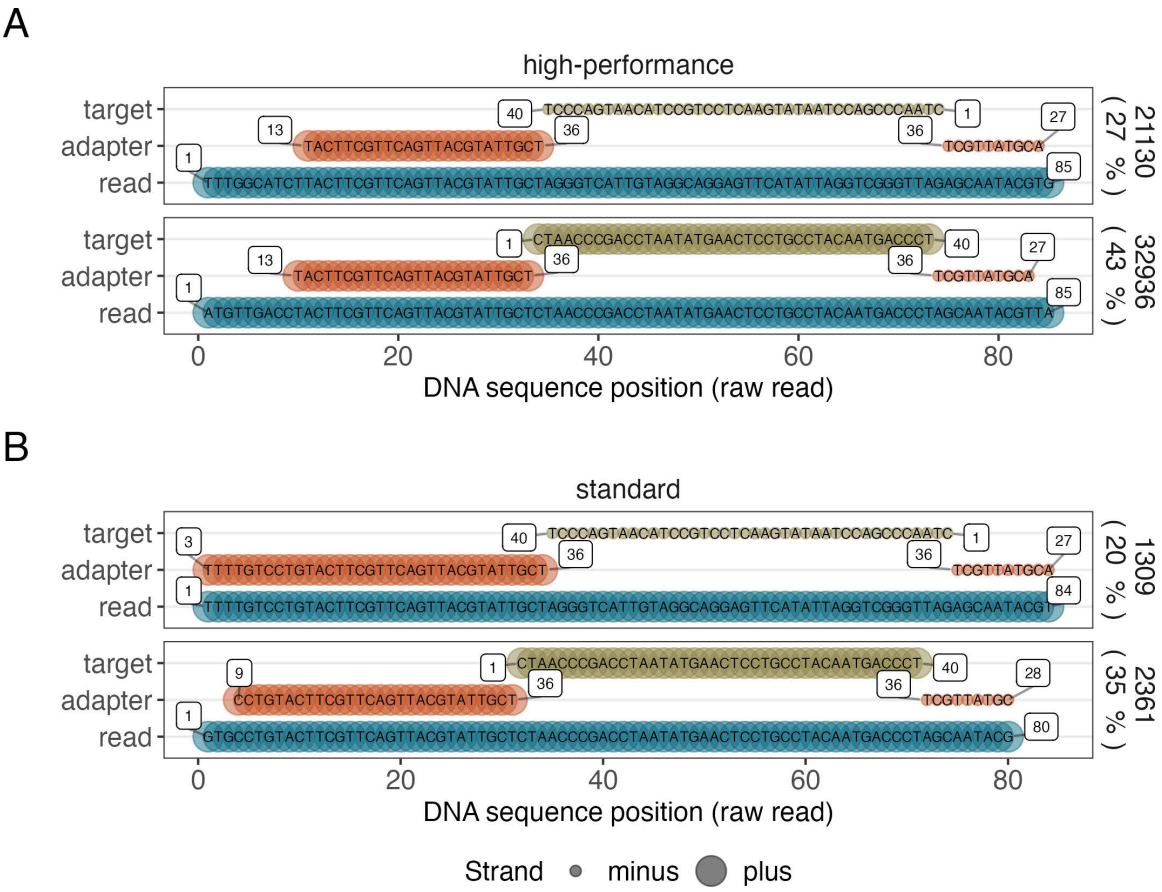

Fig 3. Clustering analysis reveals that High-performance protocol yields a higher fraction of reads containing the correct sequence. The graph represents the top two most abundant consensus sequences in libraries prepared by High-performance (A) and standard (B) library preparation protocols. Reads were aligned with the target sequence (green) flanked by adapter fragments (dark and light orange). Numbers within the bars indicate specific sequence positions. The proportions of reads in the largest and second-largest clusters are displayed to the right of each panel, alongside the total read counts for each cluster. The x-axis represents the DNA sequence position of raw reads, while strand orientation (plus or minus) is indicated by line thickness.This clustering distribution suggests that the High-performance protocol was more effective in preserving the structural integrity of the target sequence across its full length, as evidenced by the higher proportion of reads assigned to the dominant cluster. The increased homogeneity observed with the High-performance protocol likely results from its ability to generate a higher proportion of full-length reads (Fig 2A), which contributes to the observed clustering patterns.

https://doi.org/10.1371/journal.pone.0318040.g003

Conclusions

This study presents a refined High-performance protocol, specifically designed to optimise sequencing of ultra-short DNA fragments. Relying on the same conventional reagents as those used in standard ONT protocols, the High-performance protocol delivers over a tenfold increase in sequencing yield, while maintaining stable sequencing quality and improving mapping efficiency for fragments as short as 40 bp.

The protocol is also applicable to longer DNA fragments (> 40 bp), although with possibly reduced efficiency due to its underlying principles such as the limited effect of increased bead concentration on DNA recovery as fragment length increases. Further research is needed to evaluate its performance across different fragment lengths comprehensively.

Overall, this study highlights the advantages of the High-performance protocol, demonstrating its potential for a wide range of applications previously considered incompatible with nanopore sequencing technology. This streamlined

approach expands the capabilities of ONT sequencing, opening new possibilities for its use in fields focused on short and ultra-short DNA analysis.

Supporting information

S1 File. Step-by-step High-performance protocol, also available on Protocols.io. dx.doi.org/10.17504/protocols.io.dm6gp973jvzp/v1.
(PDF)

S1 Fig. The High-performance protocol produces fewer sequencing artifacts than the standard library preparation method. A) Average read length of sequences produced over time are more homogenous for the libraries prepared by High-performance protocol. The line graph shows the average read length with standard error (SE) for libraries prepared using the standard and High-performance protocols. B) As shown in the boxplots, libraries prepared using the standard (st) library preparation protocol have a higher fraction of reads with a length of 1 kb or more compared to those prepared using the High-performance (hp) protocol. Of note, sequencing artifacts in this case refer to unexpectedly long reads exhibiting low sequence diversity. The results represent three (N = 3) and four (N = 4) independent sequencing experiments by using the High-performance and standard ONT protocol, respectively.
(TIF)

Author contributions

Conceptualization: Lukas Žemaitis, Rūta Palepšienė, Simonas Juzėnas, Gediminas Alzbutas, Ignas Galminas.

Data curation: Simonas Juzėnas, Gediminas Alzbutas.

Formal analysis: Simonas Juzėnas, Gediminas Alzbutas.

Funding acquisition: Lukas Žemaitis, Pierre-Yves Burgi, Thomas Heinis, Jérôme Charmet, Silvia Angeloni Suter, Martin Jost, Renaldas Raišutis, Friedrich Simmel, Ignas Galminas.

Investigation: Lukas Žemaitis, Rūta Palepšienė.

Methodology: Lukas Žemaitis, Rūta Palepšienė.

Project administration: Ignas Galminas.

Resources: Lukas Žemaitis, Ignas Galminas.

Software: Simonas Juzėnas, Gediminas Alzbutas.

Supervision: Lukas Žemaitis, Ignas Galminas.

Validation: Rūta Palepšienė.

Visualization: Simonas Juzėnas.

Writing – original draft: Rūta Palepšienė, Simonas Juzėnas, Gediminas Alzbutas.

Writing – review & editing: Lukas Žemaitis, Rūta Palepšienė, Simonas Juzėnas, Gediminas Alzbutas, Pierre-Yves Burgi, Thomas Heinis, Jérôme Charmet, Ignas Galminas.

References

1. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol. 2021;39(11):1348–65. https://doi.org/10.1038/s41587-021-01108-x PMID: 34750572
2. Lopez R, Chen Y-J, Dumas Ang S, Yekhanin S, Makarychev K, Racz MZ, et al. DNA assembly for nanopore data storage readout. Nat Commun. 2019;10(1):2933. https://doi.org/10.1038/s41467-019-10978-4 PMID: 31270330

3. Yan Y, Pinnamaneni N, Chalapati S, Crosbie C, Appuswamy R. Scaling logical density of DNA storage with enzymatically-ligated composite motifs. Sci Rep. 2023;13(1):15978. https://doi.org/10.1038/s41598-023-43172-0 PMID: 37749195
4. Barati Farimani A, Dibaeinia P, Aluru NR. DNA Origami-Graphene Hybrid Nanopore for DNA Detection. ACS Appl Mater Interfaces. 2017;9(1):92–100. https://doi.org/10.1021/acsami.6b11001 PMID: 28004567
5. Abedini-Nassab R. Nanotechnology and Nanopore Sequencing. Recent Pat Nanotechnol. 2017;11(1):34–41. https://doi.org/10.2174/1872210510666160602152913 PMID: 27262629
6. Chaisson MJ, Brinza D, Pevzner PA. De novo fragment assembly with short mate-paired reads: Does the read length matter?. Genome Res. 2009;19(2):336–46. https://doi.org/10.1101/gr.079053.108 PMID: 19056694
7. Hoose A, Vellacott R, Storch M, Freemont PS, Ryadnov MG. DNA synthesis technologies to close the gene writing gap. Nat Rev Chem. 2023;7(3):144–61. https://doi.org/10.1038/s41570-022-00456-9 PMID: 36714378
8. Chan AP, Choi Y, Brinkac LM, Krishnakumar R, DePew J, Kim M, et al. Multidrug resistant pathogens respond differently to the presence of co-pathogen, commensal, probiotic and host cells. Sci Rep. 2018;8(1):8656. https://doi.org/10.1038/s41598-018-26738-1 PMID: 29872152
9. Lau BT, Almeda A, Schauer M, McNamara M, Bai X, Meng Q, et al. Single-molecule methylation profiles of cell-free DNA in cancer with nanopore sequencing. Genome Med. 2023;15(1):33. https://doi.org/10.1186/s13073-023-01178-3 PMID: 37138315
10. Wilson BD, Eisenstein M, Soh HT. High-fidelity nanopore sequencing of ultra-short DNA targets. Anal Chem. 2019;91(10):6783–9. https://doi.org/10.1021/acs.analchem.9b00856 PMID: 31038923
11. Kilobaser - Personal DNA/RNA Synthesizer. MABEAL GmbH. [cited 2024 Dec 10]. Avaliable from https://kilobaser.com/dna-and-rna-synthesizer
12. Oxford Nanopore Technologies. Ligation sequencing amplicons V14 (SQK-LSK114).
13. Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One. 2016;11(10):e0163962. https://doi.org/10.1371/journal.pone.0163962 PMID: 27706213
14. Habegger L, Sboner A, Gianoulis TA, Rozowsky J, Agarwal A, Snyder M, et al. RSEQtools: a modular framework to analyze RNA-Seq data using compact, anonymized data summaries. Bioinformatics. 2011;27(2):281–3. https://doi.org/10.1093/bioinformatics/btq643 PMID: 21134889
15. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016;4:e2584. https://doi.org/10.7717/peerj.2584 PMID: 27781170
16. Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. Swarm: robust and fast clustering method for amplicon-based studies. PeerJ. 2014;2:e593. https://doi.org/10.7717/peerj.593 PMID: 25276506
17. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80. https://doi.org/10.1093/molbev/mst010 PMID: 23329690
18. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3. https://doi.org/10.1093/bioinformatics/btp348 PMID: 19505945
19. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712
20. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics. 2012;28(19):2520–2. https://doi.org/10.1093/bioinformatics/bts480 PMID: 22908215
21. R Core Team. R: A Language and Environment for Statistical ## Computing. 2023
22. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2016
23. MacKenzie M, Argyropoulos C. An introduction to nanopore sequencing: past, present, and future considerations. Micromachines (Basel). 2023;14(2):459. https://doi.org/10.3390/mi14020459 PMID: 36838159
24. Vilgis S, Deigner HP. Sequencing in Precision Medicine. In: Precision Medicine. Elsevier; 2018. p. 79–101.
25. Delahaye C, Nicolas J. Sequencing DNA with nanopores: Troubles and biases. PLoS One. 2021;16(10):e0257521. https://doi.org/10.1371/journal.pone.0257521 PMID: 34597327
26. Sabary O, Yucovich A, Shapira G, Yaakobi E. Reconstruction algorithms for DNA-storage systems. Sci Rep. 2024;14(1):1951. https://doi.org/10.1038/s41598-024-51730-3 PMID: 38263421