

DNA isolation methods for Nanopore sequencing of the *Streptococcus mitis* genome

David Pinzauti, Francesco Iannelli, Gianni Pozzi and Francesco Santoro*

Abstract

Streptococcus mitis is a Gram-positive bacterium, member of the oral commensal microbiota, which can occasionally be the etiologic agent of diseases such as infective endocarditis, bacteraemia and septicaemia. The highly recombinogenic and repetitive nature of the *S. mitis* genome impairs the assembly of a complete genome relying only on short sequencing reads. Oxford Nanopore sequencing can overcome this limitation by generating long reads, enabling the resolution of genomic repeated regions and the assembly of a complete genome sequence. Since the output of a Nanopore sequencing run is strongly influenced by genomic DNA quality and molecular weight, the DNA isolation is the crucial step for an optimal sequencing run. In the present work, we have set up and compared three DNA isolation methods on two *S. mitis* strains, evaluating their capability of preserving genomic DNA integrity and purity. Sequencing of DNA isolated with a mechanical lysis-based method, despite being cheaper and quicker, did not generate ultra-long reads (maximum read length of 59516 bases) and did not allow the assembly of a circular complete genome. Two methods based on enzymatic lysis of the bacterial cell wall, followed by either (i) a modified CTAB DNA isolation procedure, or (ii) a DNA purification after osmotic lysis of the protoplasts allowed the sequencing of ultra-long reads up to 107294 and 181199 bases in length, respectively. The reconstruction of a circular complete genome was possible sequencing DNAs isolated using the enzymatic lysis-based methods.

DATA SUMMARY

All Nanopore sequencing data used in this study have been uploaded to the sequence read archive (SRA): SRX13372482, SRX13372483, SRX13372484, SRX13372479, SRX13372480, SRX13372481. *Streptococcus mitis* S022-V3-A4 and S022-V7-A3 complete genomes are available at the NCBI with accession numbers CP047883.1 and CP067992.1.

INTRODUCTION

The ‘mitis group’ of Streptococci comprises 20 species of human-associated Streptococci, including the pathogen *Streptococcus pneumoniae*. The phylogenetic relationships among *Streptococcus oralis*, *Streptococcus mitis* and *Streptococcus pneumoniae* are complex and cannot be completely untangled using standard biochemical approaches or sequencing marker genes such as *sodA* or the 16S rRNA gene [1]. *Streptococcus mitis*, formerly known as *Streptococcus mitior*, is a Gram-positive bacterium, member of the oral commensal microbiota, which can occasionally be the etiologic agent of human infectious diseases [2–4]. The *S. mitis* genome is between 1.8 and 2.1 Mb in length containing up to 2.277 open reading frames, and has a median G+C content of 40%. To date (December 2021), 170 *S. mitis* genomes are deposited in GenBank (<https://www.ncbi.nlm.nih.gov/genome/genomes/530>), of which only ten are complete. The 170 *S. mitis* genomes were downloaded and analysed with the speciator tool built in the PathogenWatch website (<https://pathogen.watch/>), which uses the Mash algorithm [5] to query a reference genome database and perform species assignment. The analysis confirmed the species assignment for 156 genomes (91.76%), of those 125 had a Mash distance (roughly corresponding to the average nucleotide identity) of 0, while 31 had a Mash distance between 0 and 0.05.

Received 16 September 2021; Accepted 15 December 2021; Published 16 February 2022

Author affiliations: ¹Laboratory of Molecular Microbiology and Biotechnology (LAMMB), Department of Medical Biotechnologies, University of Siena, Siena, Italy.

***Correspondence:** Francesco Santoro, santorof@unisi.it

Keywords: genomic DNA extraction; high molecular weight DNA; Gram positive; *Streptococcus mitis*; Oxford Nanopore sequencing.

Abbreviations: BUSCO, Benchmarking Universal Single-Copy Orthologs; CTAB, cetyl trimethyl ammonium bromide; HMW, high molecular weight; ONT, Oxford Nanopore Technologies.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files.

000764 © 2022 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution NonCommercial License.

Impact Statement

The possibility of achieving long sequencing reads makes Oxford Nanopore technology an essential tool for bacterial genomics. In fact, long reads have the capability of spanning long repetitive elements, solving genomic complexities and allowing complete genome assembly. Preservation of genomic DNA integrity, during DNA isolation and sequencing library preparation, is crucial to obtain ultra-long sequence reads. The isolation of high molecular weight DNA from Gram-positive bacteria can be particularly challenging. Here we set up three DNA isolation methods, which were validated for the sequencing of the highly recombinogenic *S. mitis* genome. Sequencing of DNA, isolated using a method based on cell-wall enzymatic lysis followed by protoplast osmotic lysis, produces multiple ultra-long reads, up to 181 199 bases in length. These methods can be readily applied for isolation of high molecular weight genomic DNA from difficult to lyse Gram-positive bacteria.

The remaining 14 genomes were incorrectly classified as *S. mitis* and were assigned to *S. oralis* ($n=4$), *S. pseudopneumoniae* ($n=4$), *S. pneumoniae* ($n=3$), *S. timonensis* ($n=1$), generically to *Streptococcus sp* ($n=1$), or could not be classified ($n=1$). This analysis confirmed the difficulties in taxonomic assignments within the 'mitis group' of Streptococci, and calls for a better, genome-based classification. Unfortunately, bacterial species such as *S. mitis*, *Lactobacillus crispatus* [6] or *Bordetella pertussis* [7] have highly recombinogenic and repetitive genomes, which impair the assembly of a complete genome sequence. Furthermore, the presence of more than one copy of the same insertion sequences and/or of the ribosomal RNA genes in bacterial genomes, makes difficult the assembly of complete genome sequences [8]. In fact, to date (December 2021) only 7.12 % (24884 out of 349934) of the bacterial genomes available in the GenBank database are complete. Oxford Nanopore technology can generate long and ultra-long sequencing reads, which can be as long as the DNA template used for sequencing library preparation. Long reads enable resolution of genomic complexities and allow a complete genome assembly. Since Nanopore sequencing is strongly influenced by genomic DNA quality and molecular weight, the DNA isolation step remains the first challenge for an optimal sequencing run. In the present study we have set up and compared three different DNA extraction methods, evaluating their ability to isolate pure high molecular weight DNA from two *S. mitis* strains prior to Oxford Nanopore sequencing. These methods can be readily applied for isolation of high molecular weight genomic DNA from difficult to lyse Gram positive bacteria allowing the assembly of complete genomes.

THEORY AND IMPLEMENTATION

METHODS

DNA isolation methods

S022-V3-A4 and S022-V7-A3 bacterial strains were isolated from the saliva of a healthy subject treated with minocycline during the ANTIRESEDEV project and identified as *Streptococcus mitis* by 16S rRNA gene sequencing [9]. Frozen starter cultures were diluted 50-fold in 10 ml of Tryptic Soy Broth (TSB) and incubated at 37 °C in ambient air. Bacterial growth was monitored until an $OD_{590}=1.0$, corresponding approximately to 3×10^7 c.f.u. ml⁻¹. Bacterial cultures were centrifuged at 6600 g for 5 min and the supernatants were discarded. Genomic DNAs were isolated using three different methods designed as cetyl trimethyl ammonium bromide (CTAB), Raffinose and TissueLyser. CTAB and Raffinose methods rely on enzymatic lysis of the bacterial cell wall, while TissueLyser is based on mechanical lysis. A schematic representation of the isolation methods is reported in Fig. 1.

The CTAB method was adapted from Current Protocols in Molecular Biology [10]. The collected bacterial pellet was resuspended in 14.8 ml of sterile TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0). Pre-digestion of the bacterial cell wall was carried out for 60 min (min) at 37 °C in 2.6 mg ml⁻¹ lysozyme (Sigma Aldrich). Bacterial cells were lysed for 30 min in 0.5 % sodium dodecyl sulphate (SDS) and 0.1 mg ml⁻¹ proteinase K. Proteins and polysaccharides were precipitated incubating for 10 min at 65 °C in 0.5 M NaCl and CTAB/NaCl (10 % CTAB, 0.7 M NaCl). DNA purification was carried out two times in an equal volume of chloroform/isoamyl alcohol (24 : 1 v/v) mixing well by inversion. The solution was centrifuged at 6600 g for 15 min, the supernatant was recovered and transferred into a fresh tube. DNA precipitation was obtained incubating for 30 min at -20 °C in 0.6 volumes of ice-cold isopropanol. DNA precipitate was pelleted by centrifuging at 6600 g for 15 min, let dry at room temperature (RT), then the pellet was resuspended in 100 µl of saline solution (0.9 % NaCl) and stored at 4 °C.

In the Raffinose method, bacterial lysis relies on osmotic lysis of the protoplasts while proteins and polysaccharides precipitation was obtained in absence of CTAB. The bacterial pellet was washed with 10 ml of sterile TE buffer and resuspended in 7.5 ml Raffinose buffer (50 mM Tris pH 8, 5 mM EDTA, 20 % Raffinose). Pre-digestion of the bacterial cell wall was carried out for 60 min at 37 °C in 2.6 mg ml⁻¹ lysozyme. The solution was centrifuged at 6600 g for 5 min, the supernatant was discarded. The pellet was resuspended in 8 ml of sterile water, 400 µl of 10 % SDS (final concentration of 0.5%) and 80 µl of 10 mg ml⁻¹ proteinase K (final concentration of 0.1 mg ml⁻¹) were added, incubating for 30 min at 37 °C to induce protoplasts osmotic lysis. Proteins and

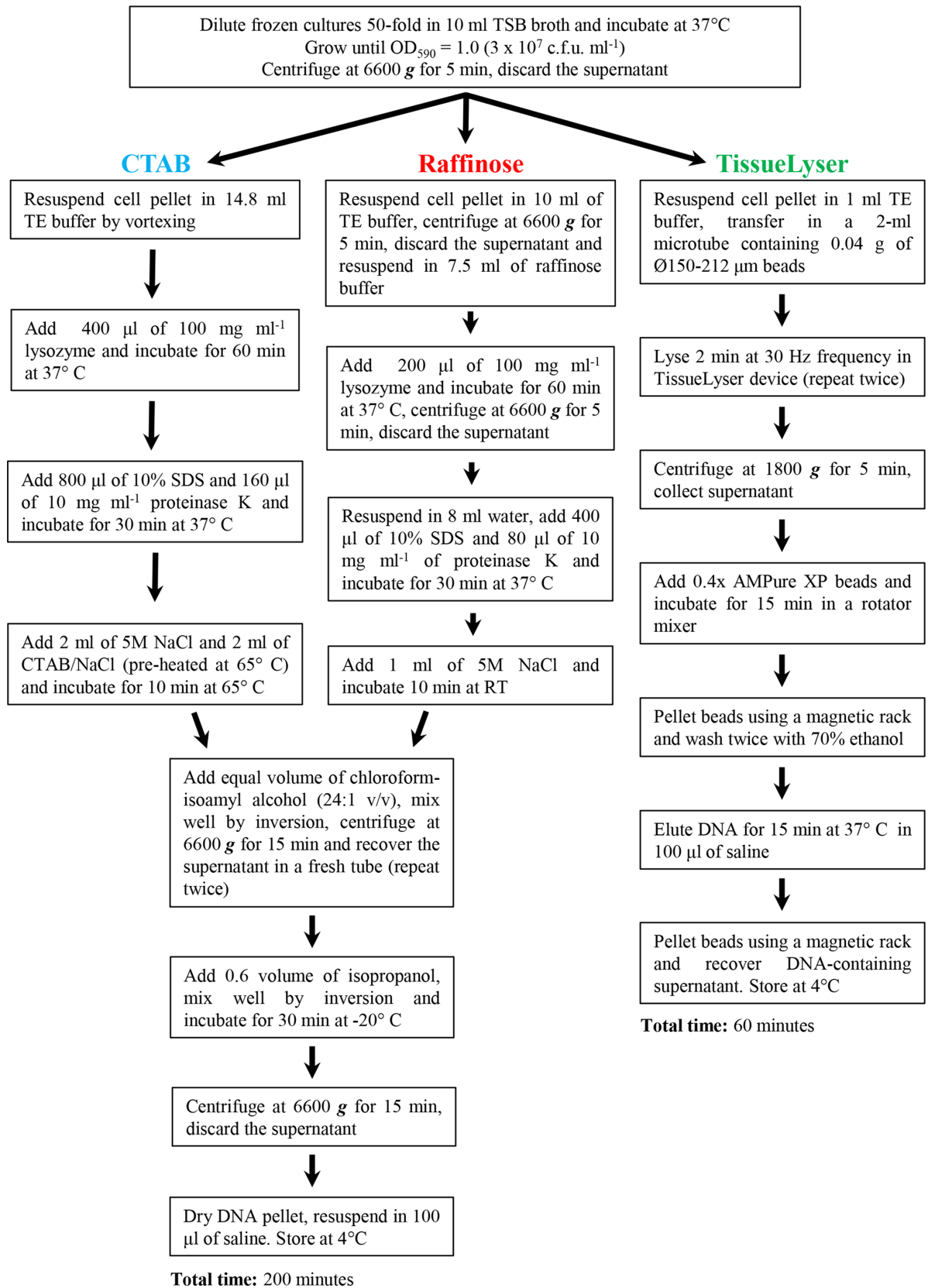


Fig. 1. Schematic representation of the DNA isolation methods.

Table 1. Quantitative and qualitative analysis of *S. mitis* genomic DNAs

| Method | DNA concentration* (ng μl^{-1}) | | DNA absorbance* | | | |
|-------------|---|--------------------|-------------------|-----------------|-------------------|-----------------|
| | S022-V3-A4 | S022-V7-A3 | A_{260}/A_{280} | | A_{260}/A_{230} | |
| | S022-V3-A4 | S022-V7-A3 | S022-V3-A4 | S022-V7-A3 | S022-V3-A4 | S022-V7-A3 |
| CTAB | 177.13 \pm 80.1 | 355 \pm 221.73 | 1.96 \pm 0.07 | 1.79 \pm 0.17 | 2.13 \pm 0.04 | 1.97 \pm 0.10 |
| Raffinose | 158.4 \pm 43.4 | 196.66 \pm 42.72 | 1.92 \pm 0.05 | 1.94 \pm 0.07 | 2.04 \pm 0.07 | 2.09 \pm 0.02 |
| TissueLyser | 48.08 \pm 45.74 | 25.74 \pm 16.29 | 1.96 \pm 0.13 | 2.03 \pm 0.13 | 2.00 \pm 0.13 | 2.11 \pm 0.05 |

*Values represent the average of at least three independent DNA isolation experiments or measurements. For Oxford Nanopore sequencing acceptable DNA absorbance ratios are as follows: $A_{260}/A_{280}=1.8-2.0$ and $A_{260}/A_{230}=2.0-2.2$.

polysaccharides were precipitated incubating for 10 min at RT in 0.5 M NaCl. DNA purification was carried out as described in the CTAB method.

The TissueLyser protocol relies on glass beads to mechanically disrupt the bacterial cell wall. The bacterial pellet was resuspended in 1 ml of sterile TE buffer and transferred into a fresh 2 ml microtube containing 0.04 g of sterile glass beads (Sigma-Aldrich, $\text{O}150-212 \mu\text{m}$). Bacterial cells were lysed with two passages of 2 min at 30 Hz frequency in the TissueLyser device (Qiagen). The solution was centrifuged for 5 min at 1800 g, recovering the supernatant. DNA purification was performed by adding $0.4 \times$ AMPure XP beads (Beckman Coulter) and incubating for 15 min in a rotator mixer. Beads were pelleted using a magnetic rack (NimaGen), washed twice with freshly prepared 70 % ethanol, and resuspended in 100 μl of saline incubating for 15 min at 37 $^{\circ}\text{C}$. Beads were pelleted again in the magnetic rack, the eluate containing the DNA was recovered and stored at 4 $^{\circ}\text{C}$.

DNA quantification

Extraction methods were compared in terms of genomic DNA quantity, quality and molecular weight. Extraction yields were measured using Qubit dsDNA BR assay kit (Thermo Fisher Scientific) in a Qubit 2.0 Fluorometer device (Invitrogen). The DNA concentration and the whole DNA amount recovered were measured for each extraction protocol. Genomic DNA purity grade was determined using a NanoPhotometer device (IMPLEN), evaluating the DNA absorbance ratios 260 nm/280 nm and 260 nm/230 nm. DNA integrity was determined using an agarose gel electrophoretic assay. Samples were loaded into a 0.8 % agarose gel and ran for 4 h at 3 V/cm in $0.5 \times$ Tris-Borate-EDTA buffer.

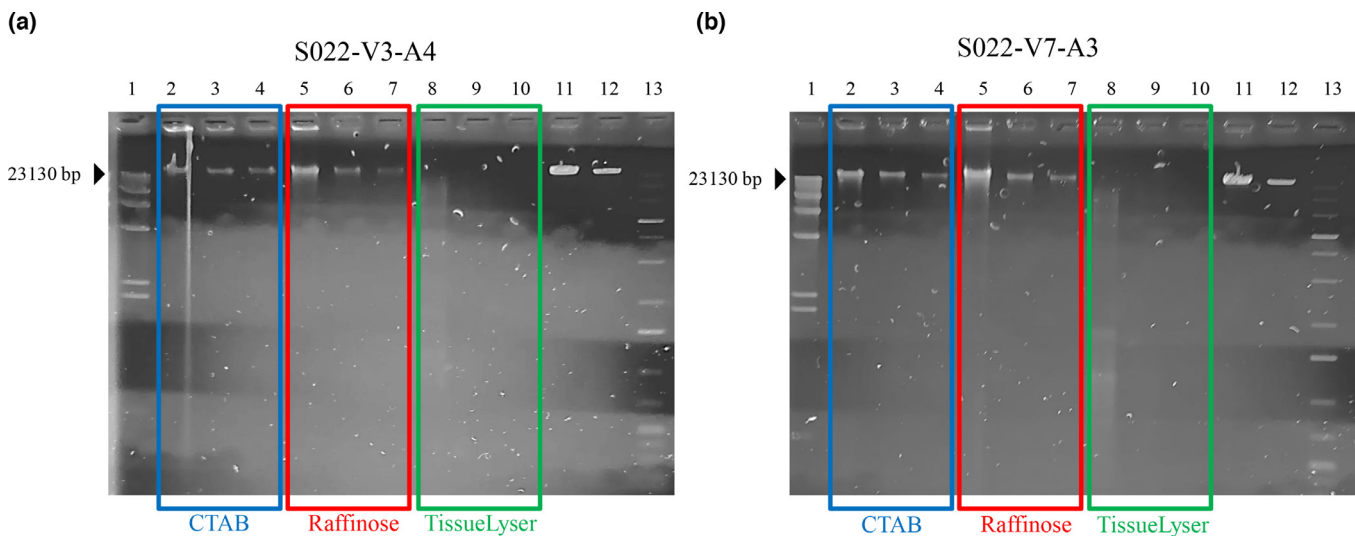


Fig. 2. Agarose gel electrophoretic analysis of *S. mitis* genomic DNAs. (a) Representative DNAs extracted from S022-V3-A4 strain, (b) representative DNAs extracted from S022-V7-A3 strain. Lane 1, Lambda (λ) DNA/HindIII (100 ng). Genomic DNA purified with the CTAB method: lane 2, (1 μl); lane 3, (1/10 μl); lane 4, (1/20 μl). Genomic DNA purified with the Raffinose method: lane 5, (1 μl); lane 6, (1/10 μl); lane 7, (1/20 μl). Genomic DNA purified with the TissueLyser method: lane 8, (1 μl); lane 9, (1/10 μl); lane 10, (1/20 μl); lane 11, λ DNA (100 ng); lane 12, λ DNA (30 ng); lane 13, GeneRuler 1 kb Plus DNA Ladder (100 ng).

Table 2. *S. mitis* genome sequencing readout

| Sequencing output | DNA extraction method | | | | | |
|--------------------|-----------------------|------------|------------|------------|-------------|------------|
| | CTAB | | Raffinose | | TissueLyser | |
| | S022-V3-A4 | S022-V7-A3 | S022-V3-A4 | S022-V7-A3 | S022-V3-A4 | S022-V7-A3 |
| Mean read length | 4476.0 | 4216.3 | 5338.5 | 4270.9 | 5121.9 | 4128.9 |
| Median read length | 2552 | 2171 | 2519 | 2047 | 4695 | 3455 |
| Mean read quality | 11.3 | 11.2 | 11.3 | 11.2 | 11.2 | 11.1 |
| Number of reads | 18485 | 17778 | 18920 | 24227 | 19109 | 57825 |
| Read length N50 | 7608 | 7700 | 10610 | 8371 | 6651 | 5317 |
| Total bases | 82739476 | 74956919 | 101004905 | 103470132 | 97875297 | 238752790 |
| Longest reads: | | | | | | |
| 1 | 82241 | 107,294 | 181199 | 124694 | 34966 | 59516 |
| 2 | 73834 | 93,505 | 130,614 | 105568 | 31951 | 52114 |
| 3 | 73597 | 84226 | 118691 | 103501 | 31676 | 47645 |
| 4 | 70310 | 82270 | 116448 | 97398 | 31098 | 38005 |
| 5 | 68235 | 80707 | 114758 | 92372 | 27268 | 36981 |

Oxford Nanopore library preparation

Isolated DNA samples were used for sequencing library preparation employing the Ligation Sequencing kit (SQK-LSK108) and the Expansion Barcoding kit (EXP-NB103), manufactured by Oxford Nanopore Technologies (ONT). A unique barcode sequence was associated to each DNA sample, enabling sample multiplexing. Sequencing libraries were prepared following the manufacturer's instructions (<https://nanoporetech.com/>), though introducing an initial size-selection step to reduce small DNA fragments' contamination. Wide bore pipette tips were used and vortexing was avoided trying to further reduce DNA shearing. Briefly, for each extraction protocol 2.5 µg of genomic DNA were size-selected using 0.7 × AMPure XP beads. The reaction was incubated for 15 min in a rotator mixer and pelleted on a magnetic rack. The bead pellet was washed twice using freshly prepared 70 % ethanol and eluted in 40 µl of saline. Size-selected samples were end-repaired using FFPE and Ultra II End-Prep kits (New England BioLabs, NEB). Barcodes were ligated using Blunt TA Master Mix (NEB) and mixed together into an equimolar pool of 1.2 µg DNA in 50 µl of saline. Adapter proteins were ligated using Quick T4 Ligase (NEB), for 30 min at room temperature. Finally, 520 ng of sequencing library was loaded on a new R9.4 flowcell (ONT). The sequencing run was performed using a MinION device (ONT).

Data analysis

Oxford Nanopore native fast5 files were basecalled and demultiplexed using the stand-alone Guppy module v. 5.0.16 (High Accuracy mode, quality threshold >7). Sequencing readouts were analysed using NanoStat (v. 1.38.0) [11], while read-length distributions were plotted using the R-package ggplot2 (v. 2.2.1) [12]. Nanopore reads were filtered using Filtlong (v. 0.2.0) (<https://github.com/rrwick/Filtlong>) to exclude reads smaller than 3000 bases (--min_length 3000) and the worst 5 % of reads (--keep_percent 95). The *de novo* genome assembly was performed with Flye (v. 2.9) [13] assembler and polished with Medaka (v. 1.4.4) (<https://github.com/nanoporetech/medaka>). A hybrid assembly was generated with Unicycler (v. 0.4.7) [14] using Nanopore and Illumina reads. Assembly quality and completeness were evaluated comparing predicted genes with the Benchmarking Universal Single-Copy Orthologs (BUSCO v. 5.2.2) tool [15], using the 'genome' running mode and lactobacillales_odb10 database. All tools were run using default parameters unless otherwise specified.

RESULTS

Three genomic DNA isolation methods were set up and compared on two *S. mitis* strains, namely S022-V3-A4 and S022-V7-A3. The capability of isolating high molecular weight DNA without compromising the purity was evaluated. Then, we used the purified DNAs as templates in whole-genome sequencing experiments to assess the ability to generate long and ultra-long Nanopore reads suitable for *de novo* genome assembly. One method based on mechanical lysis of bacterial cells using the Qiagen TissueLyser device was designed as TissueLyser. Two methods based on enzymatic lysis of the bacterial cell wall, followed by either (i) a modified CTAB DNA extraction procedure, or (ii) a DNA purification after osmotic lysis of the protoplasts were designed as CTAB and Raffinose, respectively.

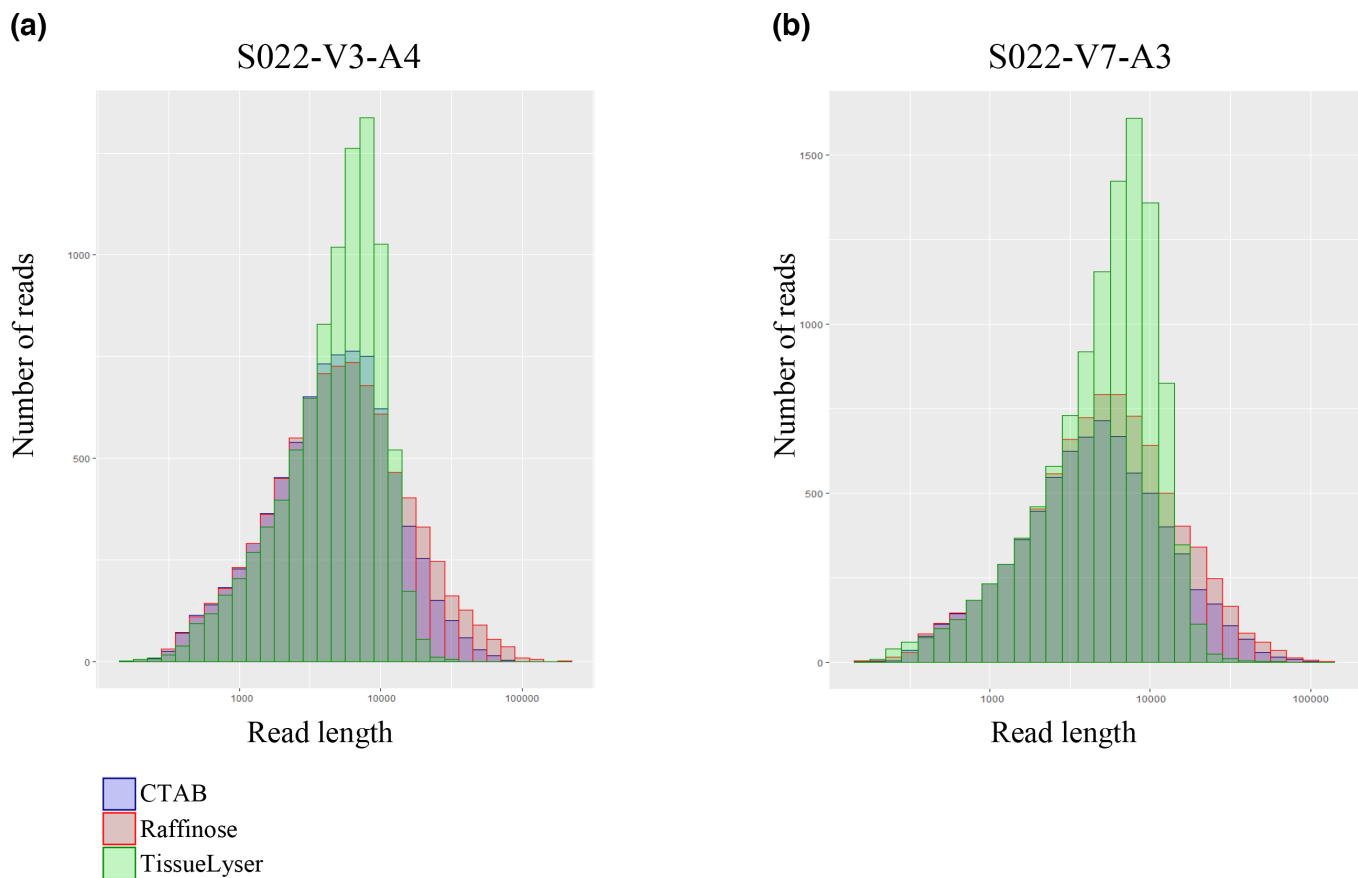


Fig. 3. Read-length distribution of *S. mitis* genome sequencing experiments. (a) S022-V3-A4 strain, (b) S022-V7-A3 strain. Read-length values are plotted on the X-axis, while the Y-axis reports the number of reads. Reads obtained sequencing the DNAs isolated using the three methods are represented with different colours: blue for CTAB, red for Raffinose and green for TissueLyser.

Genomic DNA quantitative and qualitative analysis

To determine the concentration, yield, purity and integrity of *S. mitis* genomic DNAs, fluorometric, spectrophotometric and agarose gel electrophoretic analysis were carried out. Quantitative fluorometric analysis, obtained with Qubit Fluorometer, showed that the two enzymatic lysis-based methods are more efficient than the mechanical lysis-based method allowing to achieve a total DNA yield up to 56.2 μg (Table 1). Qualitative spectrophotometric analysis showed that both the DNA absorbance ratios were within acceptable ranges for all methods. Agarose gel electrophoresis analysis showed that enzymatic lysis-based methods were able to preserve the integrity of genomic DNAs, which could be detected as two sharp high molecular weight (HMW) bands, of which one is retained in the loading well and the other has an estimated MW of about 30 kb (Fig. 2). The TissueLyser method yielded a smear in the lanes with no HMW band indicating DNA fragmentation.

Genome DNA sequencing

As a final readout for the three DNA extraction methods' comparison, three representative DNA samples obtained with each method were mixed and used as templates in whole-genome sequencing experiments (Table 2). Libraries prepared with the different DNA samples were run in the same flow cell; the sequencing run was stopped after 4 h, generating a total of ~ 700 Mb. All sequencing experiments yielded a $50 \times$ genome coverage, with the exception of the S022-V7-A3 sequencing on the templates obtained with the CTAB method, which roughly achieved a $37 \times$ coverage. Statistical analysis suggested a similar mean read length across the three methods, while the TissueLyser generated a higher median length possibly due to two size selection steps performed during the DNA isolation and before library preparation. Read-length N50 values suggested that both enzymatic lysis-based methods were able to preserve DNA integrity generating longer reads. The highest N50 values were achieved sequencing DNAs extracted with the Raffinose method, which generated multiple ultra-long reads (>100 kb). Read-length distribution showed that sequencing of DNAs extracted with the TissueLyser method, generated more reads of about 10 kb in length compared to DNAs extracted with the two enzymatic-based methods (Fig. 3). A tail of longer reads can be observed for the sequencing of DNAs obtained with the CTAB and Raffinose methods. *De novo* genome assembly showed that the long reads obtained sequencing Raffinose DNAs solved *S. mitis* genomic

Table 3. *S. mitis* genomes' assembly and data availability

| DNA extraction method | Length* | | BUSCO† | | Sequence read archive no. | | Genome GenBank no.‡ | |
|-----------------------|------------|------------|---|--|---------------------------|-------------|---------------------|------------|
| | S022-V3-A4 | S022-V7-A3 | S022-V3-A4 | S022-V7-A3 | S022-V3-A4 | S022-V7-A3 | S022-V3-A4 | S022-V7-A3 |
| CTAB | 2094027 bp | 2033697 bp | C:370 (92%), F:25 (6.2%), M:7 (1.8%) | C:356 (88.6%), F:36 (9%), M:10 (2.4%) | SRX13372482 | SRX13372479 | CP047883.1 | CP067992.1 |
| Raffinose | 2087264 bp | 2033595 bp | C:377 (93.8%), F:15 (3.7%), M:10 (2.5%) | C:372 (92.5%), F:23 (5.7%), M:7 (1.8%) | SRX13372483 | SRX13372480 | | |
| TissueLyser | 2083754 bp | 2032458 bp | C:381 (94.8%), F:19 (4.7%), M:2 (0.5%) | C:389 (96.8%), F:9 (2.2%), M:4 (1%) | SRX13372484 | SRX13372481 | | |

*Five linear contigs were obtained assembling reads obtained sequencing DNAs extracted with the CTAB and the TissueLyser method for *S. mitis* S022-V3-A4 strain, while DNAs extracted with the TissueLyser method generate a single, linear contig for S022-V7-A3 strain.

†Assembly completeness was evaluated with BUSCO v. 5.2.2 using the lactobacillales_odb10 database. A total of 402 BUSCO genes were identified, the number of complete and single-copy (C) fragmented (F), and missing (M) genes in each assembly is reported with their relative abundance in brackets.

‡The deposited genome sequences were assembled combining Illumina reads and Nanopore reads, obtained using DNAs isolated with the three different methods.

complexities achieving a complete genome for both strains (Table 3). Sequencing of the CTAB DNAs was capable of generating a complete genome for S022-V7-A3, but not for S022-V3-A4 (five contigs). Incomplete genomes were generated assembling reads obtained sequencing the TissueLyser DNAs: (i) an incomplete, fragmented assembly (five contigs) of 2083754 bp was obtained for the *S. mitis* S022-V3-A4 strain where the longest contig is 1941254 bp; (ii) a 2032458 bp incomplete, single assembly was obtained for the *S. mitis* S022-V7-A3 strain. A total of 402 BUSCO genes, contained in the lactobacillales_odb10 database, were used to assess genome quality and completeness from nanopore-only assemblies (Table 3). BUSCO identified on average 375.5 ± 5.5 (93.4 ± 1.4 %) and 372.5 ± 16.5 (92.7 ± 4.1 %) complete, single-copy genes for S022-V3-A4 and S022-V7-A3, respectively. A total of 20 ± 5 (4.95 ± 1.25 %) and 22.5 ± 13.5 (5.6 ± 3.4 %) fragmented genes, while 6 ± 4 (1.15 ± 0.65 %) and 7 ± 3 (1.7 ± 0.7 %) missed genes were identified, respectively, for S022-V3-A4 and S022-V7-A3 strains. BUSCO analysis of GenBank *S. mitis* deposited genomes showed scores >99 % for all but eight genomes, suggesting that nanopore reads alone do not allow yet perfect bacterial genome assemblies. To further improve the assembly quality, the genomic architecture of both strains was finally reconstructed based on a hybrid assembly approach, using both Nanopore and Illumina reads: (i) representative samples were also sequenced with the Illumina technology at MicrobesNG (Birmingham, UK) (sequence data not shown); (ii) Nanopore reads obtained sequencing DNAs extracted with the three different methods were merged together achieving an overall $150 \times$ genome coverage (~ 300 Mb) for each strain. S022-V3-A4 genome was assembled into a complete chromosome of 2086958 bp, while S022-V7-A3 genome was assembled into a complete chromosome of 2033396 bp. Both final genomes were assigned to *S. mitis* with a Mash distance <0.05 and reached a BUSCO score of 100 %.

Conclusions

In the present study, to overcome difficulties in complete-genome assembly of the highly recombinogenic and repetitive *S. mitis* genome, we have set up and compared three DNA isolation methods. The methods aimed to isolate a pure high molecular weight DNA from *S. mitis* to be used as a template in Oxford Nanopore sequencing experiments. We found that the two enzymatic lysis-based methods (CTAB and Raffinose) were able to preserve genomic DNA integrity and purity, while the mechanical lysis-based method (TissueLyser) was not able to isolate high molecular weight DNA. Oxford Nanopore sequencing showed that: (i) the DNA isolated with a mechanical lysis-based method, did not generate ultra-long reads failing in complete genome assembly; (ii) the DNAs isolated using the enzymatic lysis-based methods generated ultra-long reads, up to 181199 bases, achieving circular complete genome assembly. Despite the limited sequencing coverage ($\sim 50 \times$) achieved and the overall low read accuracy (mean quality 11, 92 % accuracy), the evaluation of BUSCO scores suggested that it was possible to achieve complete, good-quality genomes solely relying on Nanopore reads.

Funding information

This work was supported in part from the Italian Ministry of Education, University and Research (MIUR-Italy) under grant number 20177J5Y3P (call 'Progetti di Ricerca di Rilevante Interesse Nazionale – Bando 2017'), and in part from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement 241446 (project ANTIRESDDEV).

Acknowledgements

Illumina genome sequencing was provided by MicrobesNG (<http://www.microbesng.uk>).

Author contributions

This study was conceptualized by D.P., F.I., G.P. and F.S. Data collection and curation was undertaken by D.P. and F.S. Sample collection and laboratory testing was undertaken by F.S. and D.P. Sequencing and computational resources were supported by D.P. Formal genomic data analysis was undertaken by D.P., F.I. and F.S. The original draft was written by D.P. The final manuscript was revised by all authors.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. Velsko IM, Perez MS, Richards VP. Resolving phylogenetic relationships for *Streptococcus mitis* and *Streptococcus oralis* through Core- and pan-genome analyses. *Genome Biol Evol* 2019;11:1077–1087.
2. Mitchell J. *Streptococcus mitis*: walking the line between commensalism and pathogenesis. *Mol Oral Microbiol* 2011;26:89–98.
3. Kilian M, Riley DR, Jensen A, Brüggemann H, Tettelin H. Parallel evolution of *Streptococcus pneumoniae* and *Streptococcus mitis* to pathogenic and mutualistic lifestyles. *mBio* 2014;5:e01490-14.
4. Skov Sørensen UB, Yao K, Yang Y, Tettelin H, Kilian M. Capsular polysaccharide expression in commensal *Streptococcus* species: genetic and antigenic similarities to *Streptococcus pneumoniae*. *mBio* 2016;7:e01844-16.
5. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* 2016;17:132.
6. Teodori L, Colombini L, Cuppone AM, Lazzeri E, Pinzauti D, et al. Complete genome sequence of *Lactobacillus crispatus* type strain ATCC 33820. *Microbiol Resour Announc* 2021;10:e0063421.
7. Ring N, Abrahams JS, Jain M, Olsen H, Preston A, et al. Resolving the complex *Bordetella pertussis* genome using barcoded nanopore sequencing. *Microb Genom* 2018;4:11.
8. Koren S, Phillippy AM. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr Opin Microbiol* 2015;23:110–120.
9. Zaura E, Brandt BW, Teixeira de Mattos MJ, Buijs MJ, Caspers MPM, et al. Same exposure but two radically different responses to antibiotics: resilience of the salivary microbiome versus long-term microbial shifts in feces. *mBio* 2015;6:e01693-15.
10. Wilson K. Preparation of genomic DNA from bacteria. *Curr Protoc Mol Biol* 2001;Chapter 2:Unit 2.4.
11. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 2018;34:2666–2669.
12. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016. <https://ggplot2.tidyverse.org>
13. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
14. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
15. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 2021;38:4647–4654.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.