RESEARCH



Rapid, user-friendly, cost-effective DNA and library Preparation methods for wholegenome sequencing of bacteria with varying cell wall composition and GC content using minimal DNA on the illumina platform



Bertrand Bearzatto¹, Jean-François Durant¹, Jérôme Ambroise¹ and Jean-Luc Gala^{1*}

Abstract

Background Whole-genome sequencing using high-throughput sequencing is essential for identifying and characterising chromosomes and plasmids in nosocomial and environmental bacterial pathogens, including those with bioterrorism potential. To expedite outbreaks investigations, including accidental or intentional bacterial release, without compromising sequencing quality, we evaluated a more time-efficient, user-friendly, and cost-effective approach, using minimal DNA (~ 1 ng) from a single bacterial colony. Four DNA extraction methods were compared: the automated nucleic acid extractor (EZ1 Advanced, Qiagen) with or without DNA purification using AMPure[®] beads (EZ1 vs. EZ1-AMP), and two rapid and inexpensive methods: heat shock lysis (HS), and glass bead disruption (GBD). Additionally, we evaluated four library preparation kits: Illumina DNA Prep (DN), Illumina Nextera XT (XT), Roche KAPA HyperPlus (KP), and NEBNext[®] Ultra[™] II FS DNA Library Prep Kit (NN).

Results Whole-genome sequencing performance was evaluated on *Bacillus cereus* (*B. cereus*), *Staphylococcus epidermidis* (*S. epidermidis*), and *Enterobacter cloacae* (*E. cloacae*) ATCC strains. Key performance indicators included sequencing depth evenness across chromosome and plasmids (accounting for GC bias), genome assembly quality measured by contig number, N50, genome fraction, and percentage of mismatches. Key performance indicators confirmed that DNA and library preparation methods significantly influenced WGS quality. GBD enabled efficient sequencing across all three bacterial species, while HS proved inadequate for spore-forming bacteria *B. cereus*. DN, KP, and NN produced high-quality results with low GC bias, whereas XT exhibited significant GC bias and lower quality for bacteria with low GC content.

Conclusions This study highlights the importance of selecting suitable DNA and sequencing library preparation methods based on bacterial cell wall composition and GC content for optimal HTS outcomes.

*Correspondence: Jean-Luc Gala jean-luc.gala@uclouvain.be

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Keywords ATCC bacteria, *Bacillus*, Library preparation kits, Extraction methods, Whole genome sequencing, Plasmid, Sequencing depth, CBRN biothreats

Background

In case of infection, accurate identification and characterisation of microorganisms is critical for providing effective patient care and ensuring adherence to appropriate biosafety procedures and guidelines. Over the past decade, whole-genome sequencing (WGS) using highthroughput sequencing (HTS) has emerged as a key tool in clinical diagnostic laboratories. It enables rapid and precise nosocomial pathogen identification and characterisation, antimicrobial resistance profiling, and epidemiological investigations [1, 2]. The technology's high resolution and rapidly decreasing costs have expanded its applications beyond clinical microbiology, making it indispensable in public health microbiology laboratories for surveillance, outbreak investigations, and transmission tracking. Furthermore, HTS has proven useful in zoonotic surveillance [3], CBRN biothreat monitoring [4], and food safety by detecting foodborne pathogens and reducing foodborne disease outbreaks [5].

HTS technology enables the sequencing of multiple pathogen genomes in a single run. However, the quality of sequencing data heavily depends on preparatory steps such as DNA extraction and library preparation. Traditional bacterial genome sequencing methods rely on complex kit-based DNA extraction from liquid cultures. These methods are intended to eliminate potential confounding factors, such as ethylenediaminetetraacetic acid (EDTA), that may interfere with subsequent processes. To address the need for rapid and resource-efficient processing, novel DNA preparation methods such as Heat Shock lysis (HS) [6-8] and mechanical lysis using glass beads disruption (GBD), also known as bead beating [9] have been introduced. In HTS workflows, the method used for preparing a library also has a significant impact on the quality of WGS results [10, 11]. Several library preparation kits have been developed that use enzymaticbased methods such as transposases or endonucleases for fragmentation. Each method has distinct advantages, depending on the application and sequencing objectives.

This study aims to identify the optimal combination of DNA preparation and library preparation methods for cost-effective and accurate sequencing of bacterial chromosomes and plasmids. To achieve this, WGS was carried out on single colonies from three representative bacterial species: *B. cereus*, *S. epidermidis*, and *E. cloacae. B. cereus*, a spore-forming, heat-tolerant Gram-positive bacterium with a thick, highly resistant cell wall, was selected due to its known challenges in DNA extraction. Additionally, species from the *Bacillus* and *Staphylococcus* genera, including, *B. cereus* and *S. epidermidis*, are frequently used as simulants for validating biorisk assessment in a context of CBRN threat monitoring and surveillance, including scenarios involving lethal biological agents [12–14]. Four DNA preparation methods were compared: DNA extraction from liquid cultures using a Qiagen kit (EZ1) with or without additional purification with AMPure[®] XP beads (EZ1-AMP). Additionally, two rapid DNA preparation methods were compared: HS and GBD. For each DNA preparation method, we tested four library preparation kits, starting with a minimal DNA input of 1 ng. These kits included two enzymatic DNA fragmentation kits (Kapa[™] HyperPlus Kit (KP) and NEB-Next[®] Ultra[™] II FS DNA Library Prep Kit (NN)) and two tagmentase-based kits Nextera (XT DNA Library Prep Kit (XT) and Illumina DNA Prep Kit) (DP).

Methods

Reference bacterial species

For this study, three biosafety level 1 (BSL-1) ATCC strains of Gram-positive and Gram-negative bacteria with varying GC content and genome sizes were selected: *Bacillus cereus ATCC 14,579* (35% GC), *Staphylococcus epidermidis ATCC 12,228* (32% GC) and *Enterobacter cloacae ATCC* (54% GC) (Table 1).

Bacterial cultures and DNA extraction

B. cereus and *E. cloacae* were grown on LB-agar medium, while *S. epidermidis* was cultured on TSA medium, all overnight at 37 °C. Petri dishes with bacterial colonies were then kept at 4 °C.

Four methods of DNA preparation or extraction were performed on single bacterial colonies (Table 2). Two methods involved rapid bacterial DNA lysis performed without prior DNA extraction: HS and GBD. Additionally, two purified DNA preparations were obtained using an automatic nucleic acid extractor (EZ1 Advanced, Qiagen, Hilden, Germany). These were processed either as a one-step process (EZ1) or as a two-step process (EZ1-AMP), which included additional purification with AMPure[®] beads.

Heat shock bacterial Lysis

A single colony was resuspended in a tube with 100 μ L of molecular grade water. The suspension was heated in a water bath at 100 °C for 10 min and then immediately placed on ice. Subsequently, the suspension was centrifuged at 18,500 x g for 5 min, and the supernatant containing the lysate was collected and stored at -20 °C until use.

Table 1 Key bacterial features of species used in the study

Bacterial Strain	ATCC number	ATCC Genomic Element	ATCC size	ATCC GC content (%)	NCBI Reference Sequence	NCBI Genomic Element
B. cereus	14,579	Chromosome	5 416 622	35.3	NC_004722.1	Chromosome
		Plasmid 1	15,200	38.0	NC_004721	pBClin15
E. cloacae	13,047	Chromosome	5 317 012	54.8	NC_014121.1	Chromosome
		Plasmid 1	199 566	52.5	NC_014107	pECL_A
		Plasmid 2	86 314	47.4	NC_014108	pECL_B
		Plasmid 3	5 1 2 8	37.0	No equivalence	No equivalence
S. epidermidis	12,228	Chromosome	2 503 245	32.1	NC_004461	Chromosome
		Plasmid 1	27,017	29.1	NC_005004	pSE-12228-05
		Plasmid 2	21,979	28.9	NC_005005	pSE-12228-04
		Plasmid 3	8007	35.6	NC_005006	pSE-12228-03
		Plasmid 4	6585	31.5	NC_005003	pSE-12228-06
		Plasmid 5	4679	31.8	NC_005007	pSE-12228-02
		Plasmid 6	4439	30.1	NC_005008	pSE-12228-01

Table 2 DNA Preparation methods

Library Prepara- tion Kits	HS	GBD	EZ1	EZ1-AMP	Standard DNA Prepara- tion Time*
Bacterial Liquid Culture	NA	NA	Overnight	Overnight	Overnight
Assay time (min/ number of samples)	20 / 24	30- 60 / 6	60–90 / 6	150–170 / 24	30–60 / 24

Caption: DNA preparations were initiated either directly from a colony culture (NA) or after overnight liquid culture at 37 °C. Assay time reflects the total duration required to obtain DNA ready for library preparation, starting from a bacterial colony. *Standard DNA preparation time based on literature [13]

Bacterial Lysis using glass bead disruption

Mechanical disruption with glass beads was used to extract DNA as previously described by Köser et al., 2014 [15]. Briefly, a single colony was resuspended in a tube containing a volume of 8.33 μ L of glass beads (425–600 μ m) and 25 μ L of DNA-free water for a final bead-towater ratio of 1:3. The sample was vortexed at speed 6 on a Vortex Multi Reax (Heidolph Instruments, Schwabach, Germany) for 5 min and subsequently centrifuged in a benchtop centrifuge for 2 min at 18,500 x g. Finally, 10 μ L of the supernatant was collected and stored at -20 °C until use.

Bacterial purification using automatic nucleic acid EZ1 extractor

From a single colony, an overnight liquid culture at 37 °C was performed on LB broth for *B. cereus* and *E. cloacae*, and in Tryptic Soy Broth for *S. epidermidis*. The culture was then centrifuged at 1,900 x g at room temperature for 10 min, and the supernatant was discarded. For Grampositive bacteria, the pellet was resuspended in a solution containing 170 μ L of buffer G2 (Qiagen, Hilden, Germany), 20 μ L of lysozyme 50 mg/ml (Roche Diagnostics GmbH, Mannheim, Germany) and 10 μ l of lysostaphin

(5 mg/ml) (ProSpec-Tany TechnoGene Ltd., Ness-Ziona, Israel), and incubated at 37 °C for 30 min with shaking at 300 rpm. For Gram-negative bacteria, the pellet was resuspended in 200 μ L of buffer G2. The lysate was then purified on the EZ1 Advanced XL apparatus (Qiagen, Hilden, Germany) using the pre-programmed DNA Bacteria Card protocol and the EZ1 DNA Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The elution volume was set at 100 μ L, and the eluate was stored at -20 °C until use.

AMPure® bead purification

Fifty μ L of DNA extracted according to the EZ1 protocol was purified at a ratio of 1.8:1 of AMPure[®] beads (Agencourt[®] AMPure[®], Beckman Coulter, MA, USA) relative to DNA, aiming to remove fragments smaller than 100 bp. The elution volume was adjusted at 50 μ L, and the eluate was stored at -20 °C until use.

DNA quantification and Dilution

DNA quantification was performed using the Qubit 1X dsDNA High Sensitivity Assay Kit (Invitrogen Life Sciences, Merelbeke, Belgium) on a Qubit 4 Fluorometer (Invitrogen Life Sciences, Merelbeke, Belgium), following the manufacturer's instructions with a 2 μ L sample volume. Based on the DNA concentration stocks, dilutions were made in DNA-free Tris-HCl 10 mM to adjust the working solution to 0.333 ng/ μ L. Four aliquots of each DNA solution were prepared for use in library preparation and stored at -20 °C.

Library Preparation for WGS

Four library preparation kits using different fragmentation methods were used for this study: Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) (XT), DNA Prep Kit (Illumina, San Diego, CA, USA) (DP), KAPA HyperPlus Kit (Roche Diagnostics GmbH, Mannheim, Germany) (KP), and NEBNext[®] Ultra[™] II FS DNA Library Prep Kit (Bioké, Leiden, The Netherlands) (NN). Each library was prepared using a minimal starting quantity of DNA (1 ng), with amplification conditions adapted according to the manufacturer's instructions (Table 3).

The library preparations and WGS sequencing were carried out four times for each DNA sample. The quantity and quality of each individual library were assessed using the Qubit 1X dsDNA High Sensitivity Assay Kit on a Qubit 4 Fluorometer and a High Sensitivity DNA Assay kit on a 2100 Expert Bioanalyzer apparatus, following the manufacturer's instructions. Subsequently, the libraries were equimolarly pooled, and the quantity and the quality of each pool were assessed again using the Qubit and Bioanalyzer, respectively. Finally, each pool was loaded onto a MiSeq for a paired-end 2×300 bp sequencing run using the MiSeq reagent kit V3 (600 cycles).

Bioinformatic analysis

To evaluate the impact of different DNA preparation methods and library preparation kits on WGS results, we compared key performance indicators (KPIs) from two bioinformatics pipelines: reference-based mapping and *de novo* assembly. The KPIs included sequencing depth evenness across chromosomes and plasmids (accounting for GC bias), the number of contigs, N50, genome fraction, and percentage of mismatches.

Reference-based mapping

Raw sequence reads were aligned to the ATCC reference genomes of the three target bacteria (Table 1) using Minimap2 v.2.17. The resulting SAM files were sorted and indexed using SAMtools v.1.6. Sequencing depth was calculated with GenomicAlignments v.1.34.1. The sequencing depth of plasmids was normalised using the average chromosomal sequencing depth as a baseline. To assess potential GC content bias, the genome was segmented into 200 bp fragments. For each segment, the local-to-average sequencing depth ratio was calculated and plotted against the corresponding GC content.

De Novo assembly

After down-sampling reads to 150.000/sample, filtered FASTQ files were assembled into contigs using SPAdes v.4.0.0. Contigs shorter than 1000 bp were discarded.to enhance assembly quality. The quality of the assemblies (i.e. N50, number of contigs, genome fraction compared to ATCC reference genomes, and the percentage of mismatches between the assemblies and the ATCC reference genomes) was evaluated using QUAST 5.0.2.

We used the ggplot2 R package to visualise and compare quality metrics across different DNA extraction and library preparation conditions.

Results

Four DNA preparation methods, (EZ1, EZ1-AMP, GBD, and HS) were tested, generating the sufficient minimal genomic DNA (1 ng) required for preparing sequencing libraries for the three bacterial strains assessed in this study (Table 4). Due to the small amount of starting DNA, additional amplification cycles were needed to produce sufficient material for library construction. In line with the library preparation kit recommendations, the final library underwent 9 to 14 amplification cycles. Sufficient final library material for MiSeq sequencing was successfully generated for S. epidermidis and E. cloacae using any DNA preparation method and library preparation kit. However, for *B. cereus*, adequate final library material was only obtained with DNA isolated using the EZ1, EZ1-AMP, and GBD methods. The HS method did not yield sufficient final library material for B. cereus, regardless of the library preparation kit used (Table 4). Since each library preparation required a different number of PCR amplification cycles, which can introduce PCR duplicates, we calculated the percentage of duplicates generated. The rate of PCR duplicates across all four library preparations was extremely low, ranging from 0 to 0.3%.

Table 3 Lib	prary Preparation	kits and fragm	entation methods

Library preparation kits	Acronym	Company*	Fragmentation Methods	Amplification cycles (PCR)	Average Library Preparation Time (hours)
Illumina DNA Prep	DP	Illumina	Tagmentation by transposome	12	4
Nextera XT DNA Library Prep	XT	Illumina	Tagmentation by transposome	12	5,5
KAPA HyperPlus	KP	Roche	Enzymes	14	3
NEBNext® Ultra™ II FS DNA Library Prep	NN	New England Biolabs	Enzymes	9	2,5

Caption: Fragmentation methods and protocol adaptations based on the manufacturer's instructions for the number of PCR cycles using 1 ng of DNA. Average completion time reflects the time required to prepare a library ready for sequencing, starting from DNA. *Details of the companies are provided in the Material and Methods section

Bacterial Species	DNA Preparation Method	DNA amount (ng)	Library Preparation Outcome from four trials			
			DP	KP	NN	ХТ
B. cereus ATCC 14,579	EZ1	1	4/4	4/4	4/4	4/4
	EZ1-AMP	1	4/4	4/4	4/4	4/4
	GBD	1	4/4	4/4	4/4	4/4
	HS	1	0/4	0/4	0/4	0/4
E. cloacae ATCC 13,047	EZ1	1	3/4	4/4	4/4	4/4
	EZ1-AMP	1	4/4	4/4	4/4	4/4
	GBD	1	4/4	4/4	4/4	4/4
	HS	1	2/4	4/4	4/4	4/4
S. epidermidis ATCC 12,228	EZ1	1	4/4	4/4	4/4	4/4
	EZ1-AMP	1	4/4	4/4	4/4	4/4
	GBD	1	4/4	4/4	4/4	4/4
	HS	1	3/4	4/4	4/4	4/4

Table 4 Outcomes of DNA and library preparations

Caption: DNA was obtained for ATTC bacterial species using four DNA preparation methods (EZ1, EZ1-AMP, GBD, and HS). The library preparations were performed four times for each DNA purification method using four library preparation kits (DP, KP, NN, and XT). The numbers in the library preparation outcome columns represent the number of successful library preparations (out of four attempts) that yielded enough final libraries for sequencing in a single MiSeq run (minimum 1 nM of final libraries for each sample)



Fig. 1 Relative sequencing depth of plasmids to chromosomes. Caption: This panel shows the relative sequencing depth of plasmid to chromosome for *E. cloacae* (a) and *S. epidermidis* (b). Data were obtained using four DNA preparation methods (EZ1, EZ1-AMP, GBD, and HS) and four library preparation kits (DP, KP, NN, XT)

The quality of the sequencing data was compared across the different DNA and library preparation methods using indicators from two bioinformatics pipelines: read alignment against reference genomes, and *de novo* assembly.

Assessing WGS quality by aligning with a reference genome

The generated reads from each of the three bacteria were analysed and compared to a reference genome (Table 1). The alignment analysis enabled the determination of the mean sequencing depth for both chromosomes and plasmids.

To evaluate the relative sequencing depth of plasmids, we calculated the ratio of the mean sequencing depth on plasmids to that on the chromosome (Fig. 1). This ratio is crucial for ensuring that plasmids are deeply sequenced without consuming an excessive number of sequencing reads, which could reduce chromosomal sequencing depth and coverage. The three ATCC strains in this study carried plasmids with varying numbers and GC content, which can affect sequencing quality (Table 1). E. cloacae ATCC 13,047 showed comparable sequencing depth for both the pECL_A (199.6 kb, 52% GC) and pECL_B (84.6 kb, 47.4%) plasmids relative to the chromosome (Fig. 1a). However, the third plasmid (5.1 kb, 37.0% GC, no NCBI reference), which had a lower GC content, exhibited reduced sequencing depth when using XT library preparation kit compared to the other three kits (Fig. 1a). In contrast, for S. epidermidis ATCC 1228, the relative sequencing depth on its six plasmids was 3 to 500 times higher than that of the chromosome (Fig. 1b).

No sequencing results were obtained from the plasmid typically found in *B. cereus* ATCC 14,579 (data not shown). PCRs targeting two plasmid loci in *B. cereus* ATCC 14,579 confirmed the absence of gene amplification, supporting the hypothesis that the plasmid was lost in the strain used in this study.

The relationship between sequencing depth and GC content was assessed under all conditions (Gram-positive and Gram-negative bacteria, DNA preparation method, and library preparation kit) to identify potential GC content-associated bias. Bacterial genomes were segmented into 200 bp fragments, and the GC content of each segment was calculated. The local-to-average sequencing depth ratio showed that the GC content did not affect sequencing depth for three library preparation kits (DP, KP, and NN), regardless of the bacteria strain or the DNA preparation method (Fig. 2) (Supplementary figure S1). The XT kit was the only library preparation kit where sequencing depth was influenced by GC content, with GC-rich regions being more deeply sequenced than regions with low GC content (Fig. 2) (Supplementary figure S1).

Assessing WGS quality using de Novo assembly analysis

To evaluate the ability to reconstruct a whole genome from sequencing reads, we conducted a second analysis using *de novo* genome assembly. The number of contigs was unaffected by the DNA preparation method when libraries were prepared with the DP, KP, and NN kits (Fig. 3a-c). However, libraries generated using the XT kit produced more contigs (Fig. 3a-c). Notably, when sufficient DNA was available for *E. cloacae* (Fig. 3b) and *S. epidermidis* (Fig. 3c) prepared with the HS method, we observed a noticeable increase in the number of contigs when the DP kit was used for library preparation. The N50 values were evaluated and found to be lower for genome assemblies from libraries prepared with the XT kit compared to those from the other library preparation kits (Fig. 4a-c). For the DP, KP and NN kits, N50 values across all bacteria were similar regardless of the DNA extraction method used, except for *E. cloacae* (Fig. 4b) and *S. epidermidis* (Fig. 4c) when the HS method was followed by the tagmentase-based DP library preparation.

Assemblies of *B. cereus, E. cloacae*, and *S. epidermidis* using the EZ1, EZ1-AMP, and GBD DNA preparation methods, combined with the DP, KP, and NN kits, achieved high genome fraction (>95%) compared to the ATCC reference genomes (Fig. 5a-c). In contrast, assemblies obtained using the HS DNA preparation method and/or the XT kit exhibited lower genome fractions, ranging from 60 to 95%.

Except for the HS method, assemblies for each bacterium generated very low mismatch percentages compared to the ATCC reference genomes, with particularly high base-calling performance observed using the NN library preparation kit (Supplementary figure S2).

Discussion

Our investigation aimed to determine the optimal combination of DNA preparation and library preparation for sequencing the bacterial genome and plasmid DNA when working with limiting amounts of DNA. The study focused on three bacteria with varying average GC contents: *B. cereus, S. epidermidis,* and *E. cloacae,* which represent both Gram-positive and Gram-negative types.

The impact of DNA extraction method on WGS has already been studied by comparing various commercial DNA preparation kits [16]. Among the commercial DNA preparation protocols used in this study, EZ1 and EZ1-AMP using solid-phase reversible immobilisation (SPRI) beads yielded similar results, with little effect on library preparation efficacy and sequencing quality. Among the two rapid DNA preparation techniques, DNA prepared by mechanical disruption of bacteria using GBD provided results comparable to other DNA purification methods. However, the boiling extraction method (HS) had some limitations. Although this method produced enough genomic DNA (1 ng) for preparing the library of all tested bacteria, no final library could be obtained from B. cereus DNA, regardless of the library preparation kit used.

Numerous studies have demonstrated the efficacy of boiling extraction in retrieving DNA from both Grampositive and Gram-negative strains for quantitative PCR tests [6, 17, 18]. However, *B. cereus*, a food-borne pathogenic bacterium that is commonly found as a contaminant in food and dairy products, has a high tolerance to heat [19, 20] and used as biological simulant for *B. anthracis* [12, 13]. This feature may be attributed



Fig. 2 Sequencing depth bias as a function of GC content (%). Caption: This panel shows the sequencing depth bias as a function of GC content (%) for *S. epidermidis*, using DP, KP, NN, and XT library preparation kits

to its cell wall composition, which is typical of sporeforming Gram-positive bacteria. It contains peptidoglycan, which provides greater rigidity than the cell walls of other Gram-positive bacteria such as *S. epidermidis* or the Gram-negative *E. cloacae*. The latter two bacteria produced better sequencing results, due to their cell walls being more easily broken down by boiling. Although this boiling extraction method is simple to use and sufficient from preparing for PCR testing, it should not be used for Gram-positive spore-forming bacteria such as *B. cereus*.

The WGS analysis showed that plasmid sequencing depth in *S. epidermidis* exceeded that of the chromosome. This well-known phenomenon, most likely caused by the plasmid's higher copy number, does not always occur. In our study, plasmid sequencing depth in *E. cloacae* was equal to or slightly lower than chromosome sequencing depth, while no plasmid sequencing was found in *B.*



Fig. 3 WGS quality evaluation using *de novo* assembly analysis: number of contigs larger than 1000 bp. Caption: This panel shows the number of contigs larger than 1000 bp generated for *B. cereus* (a), *E. cloacae* (b), and *S. epidermidis* (c) under various DNA preparation methods (EZ1, EZ1-AMP, GBD, HS) and library preparation kits (DP, KP, NN, and XT). Small dots represent replicates, while larger dots indicate means

cereus. There are several explanations for low plasmid sequencing depth or the absence of plasmid sequencing. In some bacteria, plasmids can exist in low copy numbers [21], restricting their sequencing depth. Although plasmids can benefit their bacterial hosts, replication and gene expression through host machinery are considered

metabolic burdens [22]. Consequently, in the absence of selection for plasmid-encoded traits, cells lacking plasmids outcompete cells carrying plasmids, leading to plasmid silencing over time [23]. Plasmid loss is particularly common in *Bacillus* bacteria [24–26]. Despite observing variability in plasmid sequencing depth across the three



Fig. 4 WGS quality evaluation using *de novo* assembly: N50 values. Caption: This panel shows the N50 values for *B. cereus* (a), *E. cloacae* (b), and *S. epider-midis* (c) under various DNA preparation methods (EZ1, EZ1-AMP, GBD, HS) and library preparation kits (DP, KP, NN, and XT). Small dots represent replicates, while larger dots indicate means

bacterial species, this variability did not correlate with differences in DNA extraction methods or library preparation technique except for one *E. cloacae* plasmid with a low GC content, which showed reduced sequencing depth when using the Nextera XT library preparation kit.

It is generally accepted that GC bias in high-throughput sequencing data complicates genome assembly, with problems becoming increasingly severe outside the 45–65% GC range. This causes low sequencing depth in GC-poor sequences, with genomic windows containing 30% GC having less sequencing depth than windows



Fig. 5 Genome fractions of *de novo* assemblies. Caption: This panel shows the genome fraction of *de novo* assemblies for *B. cereus* (a), *E. cloacae* (b), and *S. epidermidis* (c) compared to the ATCC reference genomes under various DNA preparation methods (EZ1, EZ1-AMP, GBD, HS) and library preparation kits (DP, KP, NN, and XT). Small dots represent replicates, while larger dots indicate means

close to 50% GC content [27, 28]. Among traditional library preparation kits, XT is particularly sensitive to GC content [29]. The XT sequencing depth bias problem is especially severe across the entire genome for bacteria with low GC-content [30]. This GC bias can be attributed

to the binding motif of Nextera XT tagmentase, which is dependent on G and C base content [30].

In our study, the apparent similarity in GC content between *S. epidermidis* (32.07%) and *B. cereus* (35.29%) represents genome-wide averages, although local variations in GC content exist within each genome. Our

Table 5 Cost per sample (€; excl. VAT)) for DNA and library Preparation

DNA Pu- rification	Price/extraction	Library Preparation including indexes					
Method		Price/sample					
		DP	КР	NN	ХТ		
		39,20	46,05	29,91	40,35		
EZ1	12,08	51,28	58,13	41,99	52,43		
EZ1-AMP	13,22	52,42	59,27	43,13	53,57		
GBD	0,11	39,31	46,16	30,02	40,46		
HS	0,00	39,20	46,05	29,91	40,35		

analyses focused on 200-bp bacterial genome segment with highly variable GC content, ranging from 15 to 55% for S. epidermidis (Fig. 2), 20 to 60% for B. cereus (Supplementary figure S1a), and up to 75% for certain segments of *E. cloacae* (Supplementary figure S1b). This wide range of intragenomic GC variations strengthens the conclusions of our study regrading sequencing GC bias and its applicability to regions with GC content between 20 and 60%. Although plasmids with varying GC content were expected to similarly assess sequencing biases, their variable copy numbers could complicate the interpretation of general trends. Nonetheless, plasmid sequencing remains crucial for applications such as surveillance of antimicrobial resistance and bacterial virulence [1-3], CBRN biothreat monitoring in genetic engineering and synthetic biology [31, 32], and food safety [5].

Our results confirmed that the XT kit produced lowerquality genome assemblies compared to enzymatic fragmentation kits such as KP and NN. Recent studies further support this observation, showing that the Collibri EZ kit, which also uses enzymatic fragmentation, outperforms Nextera XT in terms of genome assembly quality [33]. Additionally, genome assembly from DP kit after HS-based DNA extraction yielded similar results to those obtained with the XT kit for S. epidermidis and E. cloacae. Although previous studies have shown that the DP kit generally provides superior performance metrics including assembly quality, compared to XT [10, 34, 35], the choice of DNA preparation remains crucial when using tagmentase-based preparation kits. Lower sequencing depth in some regions, leading to poor genome coverage, may prevent the detection of single-nucleotide polymorphisms and genomic regions with functional or phylogenetical significance. Attempts to reduce gaps or low-sequencing depth in some genomic regions by obtaining more sequence reads increase sequencing costs and may limit the effectiveness of genomic analyses, especially those involving a large number of samples [36]. In this study, we intentionally reduced the number of reads per sample to 150.000, creating a constrained sequencing scenario that simulates a minimal sequencing depth (approximately 16 to 35X). This approach allowed us to evaluate the effectiveness of genomic analyses, particularly in determining the threshold for complete coverage, defined as a genome fraction of 100%, even under conditions of potentially low sequencing depth in certain genomic regions. In this study, we showed that low-cost DNA preparations, such as GBD, combined with the DP, KP, and NN library preparation kits, produce sequencing results comparable to those obtained with more expensive kit-based DNA preparation methods (Table 5). The combination of rapid GBD DNA preparation with the NN library preparation kit yielded excellent sequencing results for S. epidermidis, E. cloacae, and B. cereus, while being the most cost-effective combination. Although not tested in our study, several reports have shown that reducing reaction volumes for library preparation and omitting certain clean-up steps can significantly lower the cost of sample preparation for sequencing [37-39]. Reducing overall sequencing costs could expand the use of HTS in applications where time-efficiency, userfriendliness, and cost-effectiveness are critical. These include nosocomial pathogen surveillance in hospitals, outbreak investigation and transmission monitoring by public health microbiology laboratories, and European security efforts, where rapid and unambiguous genomic and plasmid characterisation of Bacillus species, including B. anthracis, is critical in cases of accidental or intentional bacterial release.

One limitation of our study is that library preparations were performed using DNA extracted from a single bacterial colony across four replicates. While this approach enabled us to assess technical reproducibility of the library preparation process, it does not account for potential biological variability, such as colony size or age differences, which can influence DNA yields and potentially impact the pipeline's broader applicability [40].

Conclusion

In conclusion, based on ATCC reference strains and two bioinformatics pipelines, our KPIs demonstrate that DNA and library preparation methods significantly affect the quality and efficiency of bacterial chromosome and plasmid sequencing. Rapid, GBD-based DNA preparation method is effective for library preparation and performs comparably to automated robotic extraction. However, the HS method is ineffective against sporeforming Gram-positive B. cereus and unsuitable for de novo assembly when using tagmentase-based preparation kits. Nonetheless, Both GBD and HS, as low-cost DNA preparation methods, remain viable options for resource-constrained and mobile laboratories. Notably, the DP kit and endonuclease-based methods consistently yield high-quality results with minimal GC bias, whereas the XT kit shows significant GC bias and lower quality in low-GC-content bacteria. These findings emphasise the

importance of selecting appropriate preparation methods, particularly for bacteria with different cell wall compositions and GC content.

Abbreviations

Whole-genome sequencing
High-throughput sequencing
EZ1 Advanced Nucleic Acid Extractor (Qiagen)
AMPure® beads
Heat shock
Glass bead disruption
Illumina DNA Prep
Illumina Nextera XT
Roche KAPA HyperPlus
NEBNext® Ultra™ II FS DNA Library Prep Kit
American Type Culture Collection
Polymerase Chain Reaction
Ethylenediaminetetraacetic acid
Luria-Bertani (agar or broth)
Tryptic Soy Agar
Tryptic Soy Broth
Solid Phase Reversible Immobilisation
Base pairs
Quality Control
Bacillus cereus ATCC 14579
Staphylococcus epidermidis ATCC 12228
Enterobacter cloacae ATCC 13047

Supplementary Information

The online version contains supplementary material available at https://doi.or g/10.1186/s12864-025-11598-7.

ĺ	Supplementary Material 1
	Supplementary Material 2
	Supplementary Material 3
	<

Acknowledgements

We extend our gratitude to Béatrice Sulka (Belgian Defence Laboratories Department) for setting up the heat shock and the glass-bead disruption method on plasmid-containing Bacillus species, and to Michèle Bouyer (Belgian Defence Laboratories Department) for her invaluable assistance in culturing bacterial isolates in the early phase of this study.

Author contributions

B.B., J-F.D., J.A., and J-L.G. contributed to the conceptualization, methodology, validation, writing (original draft and review & editing), and visualization of this study. J.A. also conducted the formal analysis and data curation. J-L.G. provided resources, supervision, project administration, and funding acquisition.

Funding

This study was funded by the project MOBVEC (Mobile Bio-Lab to support first response in Arbovirus outbreaks), grant 101099283 from HORIZON-EIC-2022-PATHFINDEROPEN-01, as well as the project TeamUp Holistic capability and technology evaluation and co-creation framework for upskilled first responders and enhanced CBRN-E response), grant 101121167 from HORIZON-CL3-2022-DRS-01-09, and the project eNOVATION (Extended Network of CBRN Training Centres for Innovation), grant 101168349 HORIZON-CL3-2023-DRS-01-04. The funder had no role in the study design, collection, analysis and interpretation of data, preparation of the manuscript, or the decision to submit the paper for publication.

Data availability

The datasets generated and/or analysed during the current study are available in the European Nucleotide Archive repository, with the accession number: PRJEB81290https://www.ebi.ac.uk/ena/browser/view/PRJEB81290.

Declarations

Ethics approval and consent to participate Not Applicable.

Consent for publication Not Applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Centre for Applied Molecular Technologies (CTMA), Institute of Experimental and Clinical Research (IREC), Université catholique de Louvain (UCLouvain), Avenue Hippocrate 54/B1.54.01, Woluwe-Saint-Lambert 1200, Belgium

Received: 26 September 2024 / Accepted: 11 April 2025 Published online: 23 April 2025

References

- Hilt EE, Ferrieri P. Next generation and other sequencing technologies in 1. diagnostic microbiology and infectious diseases. Genes. 2022;13(9):1566.
- Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, García-Cobos S, Kooistra-Smid AM, Raangs EC, Rosema S, Veloo AC, et al. Application of next generation sequencing in clinical microbiology and infection prevention. J Biotechnol. 2017;243:16-24.
- Suminda GGD, Bhandari S, Won Y, Goutam U, Kanth Pulicherla K, Son YO, 3 Ghosh M. High-throughput sequencing technologies in the detection of livestock pathogens, diagnosis, and zoonotic surveillance. Comput Struct Biotechnol J. 2022;20:5378-92.
- 4. Minogue TD, Koehler JW, Stefan CP, Conrad TA. Next-Generation sequencing for biodefense: biothreat detection, forensics, and the clinic. Clin Chem. 2019;65(3):383-92
- 5 Sekse C, Holst-Jensen A, Dobrindt U, Johannessen GS, Li W, Spilsberg B, Shi J. High throughput sequencing for detection of foodborne pathogens. Front Microbiol 2017, 8:2029.
- Dashti AA, Jadaon MM, Abdulsamad AM, Dashti HM. Heat treatment of bac-6. teria: a simple method of DNA extraction for molecular techniques. Kuwait Med J. 2009;41(2):117-22.
- Dimitrakopoulou M-E, Stavrou V, Kotsalou C, Vantarakis A. Boiling extraction method vs commercial kits for bacterial DNA isolation from food samples. J Food Sci Nutr Res. 2020;3(4):311-9.
- 8. Koentjoro MP, Donastin A, Prasetyo EN. A simple method of Dna extraction of mycobacterium tuberculosis from sputum cultures for sequencing analysis. Afr J Infect Dis. 2021;15(2):19-22.
- 9. Köser CU, Fraser LJ, Ioannou A, Becq J, Ellington MJ, Holden MTG, Reuter S, Török ME, Bentley SD, Parkhill J, et al. Rapid single-colony wholegenome sequencing of bacterial pathogens. J Antimicrob Chemother. 2013:69(5):1275-81.
- 10. Seth-Smith HMB, Bonfiglio F, Cuénod A, Reist J, Egli A, Wüthrich D. Evaluation of rapid library Preparation protocols for whole genome sequencing based outbreak investigation. Front Public Health. 2019;7:241.
- 11. Huptas C, Scherer S, Wenning M. Optimized illumina PCR-free library Preparation for bacterial whole genome sequencing and analysis of factors influencing de Novo assembly. BMC Res Notes. 2016;9:269.
- 12. Greenberg DL, Busch JD, Keim P, Wagner DM. Identifying experimental surrogates for Bacillus anthracis spores: a review. Invest Genet. 2010;1:1-12.
- 13. Mondange L, Tessier É, Tournier JN. Pathogenic Bacilli as an emerging biothreat?? Pathogens 2022, 11(10).
- 14. Chiera S, Bosco F, Mollea C, Piscitello A, Sethi R, Nollo G, Caola I, Tessarolo F. Staphylococcus epidermidis is a safer surrogate of Staphylococcus aureus in testing bacterial filtration efficiency of face masks. Sci Rep. 2023;13(1):21807.
- 15. Koser CU, Fraser LJ, Ioannou A, Becq J, Ellington MJ, Holden MT, Reuter S, Torok ME, Bentley SD, Parkhill J, et al. Rapid single-colony wholegenome sequencing of bacterial pathogens. J Antimicrob Chemother. 2014;69(5):1275-81.
- 16. Nouws S, Bogaerts B, Verhaegen B, Denayer S, Piérard D, Marchal K, Roosens NHC, Vanneste K, De Keersmaecker SCJ. Impact of DNA extraction on whole

genome sequencing analysis for characterization and relatedness of Shiga toxin-producing Escherichia coli isolates. Sci Rep. 2020;10(1):14649.

- Jr R, Tamanini R, Soares B, Oliveira A, Silva F, Silva F, Augusto N, Beloti V. Efficiency of boiling and four other methods for genomic DNA extraction of deteriorating spore-forming bacteria from milk. Semina: Ciências Agrárias. 2016;37:3069–78.
- Ripon MK, Hasan M, Ahasan MM, Alam M, Kabir S, COMPARISON OF THREE DIFFERENT METHODS OF GENOMIC DNA EXTRACTION FROM GRAM POSI-TIVE AND GRAM NEGATIVE BACTERIA. 2011, 2:55–60.
- Periago PM, van Schaik W, Abee T, Wouters JA. Identification of proteins involved in the heat stress response of Bacillus cereus ATCC 14579. Appl Environ Microbiol. 2002;68(7):3486–95.
- Rossi EM, Mahl SC, Spaniol AC, Honorato JFB, Rocha T. Evaluating the thermoresistance of Bacillus cereus strains isolated from wheat flour. Res Soc Dev. 2021;10(6):e2510615268–2510615268.
- Jesus TF, Ribeiro-Gonçalves B, Silva DN, Bortolaia V, Ramirez M, Carriço JA. Plasmid ATLAS: plasmid visual analytics and identification in high-throughput sequencing data. Nucleic Acids Res. 2018;47(D1):D188–94.
- 22. San Millan A, MacLean RC. Fitness costs of plasmids: a limit to plasmid transmission. Microbiol Spectr 2017, 5(5).
- De Gelder L, Ponciano JM, Joyce P, Top EM. Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. Microbiol (Reading). 2007;153(Pt 2):452–63.
- Fasanella A, Losito S, Trotta T, Adone R, Massa S, Ciuchini F, Chiocco D. Detection of anthrax vaccine virulence factors by polymerase chain reaction. Vaccine. 2001;19(30):4214–8.
- Nair K, Al-Thani R, Jaoua S. Bacillus thuringiensis strain QBT220 pBtoxis plasmid structural instability enhances δ-endotoxins synthesis and bioinsecticidal activity. Ecotoxicol Environ Saf. 2021;228:112975.
- Braun P, Grass G, Aceti A, Serrecchia L, Affuso A, Marino L, Grimaldi S, Pagano S, Hanczaruk M, Georgi E, et al. Microevolution of Anthrax from a young ancestor (M.A.Y.A.) suggests a Soil-Borne life cycle of Bacillus anthracis. PLoS ONE. 2015;10(8):e0135346.
- Browne PD, Nielsen TK, Kot W, Aggerholm A, Gilbert MTP, Puetz L, Rasmussen M, Zervas A, Hansen LH. GC bias affects genomic and metagenomic reconstructions, underrepresenting GC-poor organisms. GigaScience. 2020;9(2):qiaa008.
- Chen YC, Liu T, Yu CH, Chiang TY, Hwang CC. Effects of GC bias in nextgeneration-sequencing data on de Novo genome assembly. PLoS ONE. 2013;8(4):e62856.
- Sato MP, Ogura Y, Nakamura K, Nishida R, Gotoh Y, Hayashi M, Hisatsune J, Sugai M, Takehiko I, Hayashi T. Comparison of the sequencing bias of currently available library Preparation kits for illumina sequencing of bacterial genomes and metagenomes. DNA Res. 2019;26(5):391–8.

- Segerman B, Ástvaldsson Á, Mustafa L, Skarin J, Skarin H. The efficiency of nextera XT tagmentation depends on G and C bases in the binding motif leading to uneven coverage in bacterial species with low and neutral GCcontent. Front Microbiol. 2022;13:944770.
- 31. Lebeda FJ, Wolinsky M, Lefkowitz EJ. Information Resources and Database Development for Defense Against Biological Weapons. In: 2005; 2005.
- 32. Mo W, Vaiana CA, Myers CJ. The need for adaptability in detection, characterization, and attribution of biosecurity threats. Nat Commun. 2024;15(1):10699.
- Nassar-Míguez L, Flores-Cruz A, Atmetlla-Salazar I, Campos-Sánchez R. Comparison of Nextera XT and Collibri ES library preparation kits: from wet lab to bioinformatics analysis. In: 2022 IEEE 4th International Conference on BioInspired Processing (BIP): 2022: IEEE; 2022: 1–5.
- Haendiges J, Jinneman K, Gonzalez-Escalona N. Choice of library Preparation affects sequence quality, genome assembly, and precise in Silico prediction of virulence genes in Shiga toxin-producing Escherichia coli. PLoS ONE. 2021;16(3):e0242294.
- Bruinsma S, Burgess J, Schlingman D, Czyz A, Morrell N, Ballenger C, Meinholz H, Brady L, Khanna A, Freeberg L, et al. Bead-linked transposomes enable a normalization-free workflow for NGS library Preparation. BMC Genomics. 2018;19(1):722.
- Tyler AD, Christianson S, Knox NC, Mabon P, Wolfe J, Van Domselaar G, Graham MR, Sharma MK. Comparison of sample Preparation methods used for the Next-Generation sequencing of Mycobacterium tuberculosis. PLoS ONE. 2016;11(2):e0148676.
- 37. Li H, Wu K, Ruan C, Pan J, Wang Y, Long H. Cost-reduction strategies in massive genomics experiments. Mar Life Sci Technol. 2019;1(1):15–21.
- Gaio D, Anantanawat K, To J, Liu M, Monahan L, Darling AE. Hackflex: lowcost, high-throughput, illumina nextera flex library construction. Microb Genom 2022, 8(1).
- Jones A, Stanley D, Ferguson S, Schwessinger B, Borevitz J, Warthmann N. Cost-conscious generation of multiplexed short-read DNA libraries for whole-genome sequencing. PLoS ONE. 2023;18(1):e0280004.
- Gimonet J, Portmann AC, Fournier C, Baert L. Optimization of subculture and DNA extraction steps within the whole genome sequencing workflow for source tracking of Salmonella enterica and Listeria monocytogenes. J Microbiol Methods. 2018;151:66–8.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.