Heliyon 5 (2019) e02745

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.heliyon.com

Research article

Application of different DNA extraction procedures, library preparation protocols and sequencing platforms: impact on sequencing results

F. Pasquali [a], I. Do Valle [b], F. Palma [a], D. Remondini [c], G. Manfreda [a], G. Castellani [c], R.S. Hendriksen [d], A. De Cesare [a,*]

[a] Department of Food and Agricultural Sciences, Alma Mater Studiorum-University of Bologna, via del Florio 2, Ozzano dell'Emilia, 40064 Italy
[b] Department of Physics, Northeastern University, 360 Huntington Avenue, Boston, MA, 02115-5000, USA
[c] Department of Physics and Astronomy, Alma Mater Studiorum-University of Bologna, viale Berti Pichat 6/2, 40127, Bologna, Italy
[d] Technical University of Denmark, Kemitorvet, Kgs. Lyngby, 2800, Denmark

ARTICLE INFO

Keywords:
Bioinformatics
Genetics
Microbiology
Molecular biology
De novo assembly statistics
SNPs calling
Antimicrobial resistance genes
Whole genome sequencing

ABSTRACT

In this study three DNA extraction procedures, two library preparation protocols and two sequencing platforms were applied to analyse six bacterial cultures and their corresponding DNA obtained as part of a proficiency test. The impact of each variable on sequencing results was assessed using the following parameters: reads quality, assembly and alignment statistics; number of single nucleotide polymorphisms (SNPs), detected applying assembly- and alignment-based strategies; antimicrobial resistance genes (ARGs), identified on *de novo* assemblies of all sequenced genomes. The investigated nucleic acid extraction procedures, library preparation kits and sequencing platforms do not significantly affect *de novo* assembly statistics and number of SNPs and ARGs. The only exception was observed for two duplicates, which were associated to one PCR-based library preparation kit. Results from this comparative study can support researchers in the choice toward the available pre-sequencing and sequencing options, and might suggest further comparisons to be performed.

1. Introduction

The advent of next generation sequencing (NGS) has allowed an explosion in sequencing of individual genomes. NGS methods, also referred to as deep, high-throughput sequencing (HTS) or massively parallel sequencing, comprise several sequencing technologies that have succeeded the traditional dideoxynucleoside chain termination (i.e., Sanger) method (Loman and Pallen, 2015; Besser et al., 2018). The advent of NGS technologies has revolutionised molecular microbiology by making readily available sequences of the whole genome (often referred as whole genome sequencing, WGS) of pathogens of clinical or public health importance (Bertelli and Greub, 2013). Microbial WGS data can be easily handled and used to elucidate phylogenetic relationships between isolates belonging to disease-causing lineages, in order to enhance their traceability and monitoring over time (Taboada et al., 2017). One of the great advantages of WGS data is their possible use for performing multiple secondary analyses, such as virulence gene detection, antibiotic resistance gene profiling, synteny comparisons, mobile genetic element identification and geographic attribution (Ronholm et al., 2016). In addition, several specialized software tools are currently available to generate *in silico* results for traditional subtyping methods from genomic sequences, allowing for efficient comparisons with historical databases (Ronholm et al., 2016).

A widespread implementation of NGS/WGS technologies is taking place in clinical and public health laboratories, although there are still limitations in resources and infrastructures (Kwong et al., 2015) as well as in the interpretation of results (Edwards and Holt, 2013). However, understanding and assessing the robustness of results collected applying different genomic approaches is essential to simplify, harmonise and standardise the "wet-lab" (e.g. DNA extraction and libraries preparation protocols) and "dry-lab" (e.g. bioinformatics pipelines for sequencing data analysis) workflows (Moran-Gilad et al., 2015). A national working group on laboratory medicine convened by the US Centers for Disease Control and Prevention (Gargis et al., 2012) has outlined some principles and guidelines for the validation of NGS workflows for clinical testing. The workgroup emphasised the need for adequate validations and quality controls parameters, as well as the use of reference materials and performance of independent proficiency tests (PTs). In agreement with such recommendations, the Global Microbial Identifier (GMI - http://www.globalmicrobialidentifier.org/) executed a pilot PT scheme with the

* Corresponding author.
E-mail address: alessandra.decesare@unibo.it (A. De Cesare).

https://doi.org/10.1016/j.heliyon.2019.e02745
Received 9 November 2018; Received in revised form 1 April 2019; Accepted 25 October 2019

main objective to ensure harmonisation and standardisation in whole genome sequencing protocols and computational analyses of sequencing data, enabling data comparability for the GMI initiative. A further objective was to assess and improve the amount of sequencing data linked to relevant metadata, which are uploaded to public databases such as NCBI, EBI and DDBJ (Wielinga et al., 2008).

The PT consisted of two wet-lab and one dry-lab component(s) targeting microorganisms relevant for public health, specifically represented by *Salmonella enterica*, *Escherichia coli* and *Staphylococcus aureus*. The wet-lab components aimed at assessing the laboratories ability to perform DNA extraction, libraries preparation and sequencing procedures. The dry component aimed at evaluating the laboratories ability to analyse a whole genome sequencing dataset and distinguish between clonally related and sporadically occurring genomes.

There are significant differences in both NGS data generation (i.e. sequencing methods, specimen preparation, run throughput and hands-on time between different sequencing platforms) and data analysis processes (i.e. possible bioinformatics tools for cluster analysis) (Moran-Gilad et al., 2015; Taboada et al., 2017). In addition, the amount and quality of sequencing data enabling pathogen characterisation (i.e. genome 'coverage') and the choice of the optimal software tools for outbreak investigations remains the subject of debate (Bertelli and Greub, 2013; Gullapalli et al., 2012; Deurenberg et al., 2017). As also observed by Quail et al. (2012), such variables may have technology (e.g. read error rates) and coverage (read depths) specific effects affecting the detection of genomic variants (e.g. percentage of correct SNP called). Thus, laboratories are expected to balance pathogen genome characteristics, instrument throughput, accuracy of variant-calling algorithms and cost of sequencing runs (Kwong et al., 2015).

So far, many efforts have been undertaken to improve existing library preparation procedures for paired-end genome sequencing (Aird et al., 2011; Kozarewa et al., 2009; Oyola et al., 2012; Quail et al., 2008). Currently, the Illumina TruSeq® DNA library preparation kit represents one of the most widely used solutions for the generation of paired-end genome sequencing libraries. It includes genomic DNA shearing by adaptive focused acoustics, which leads to random fragmentation of DNA in contrast to the more directed fragmentation via enzymatic digestion. Unbiased shearing of DNA reduces the unevenness of sequencing depth across sequenced genomes (Oyola et al., 2012; Tyler et al., 2016). Additionally, the use of magnetic beads for DNA clean-up and size selection is much less prone to contamination compared to traditional gel-based systems. One of the main drawbacks of this protocol is the need for very high amounts of starting material (1–4 μg total DNA). Moreover, in the current design this library preparation procedure is directed to the generation of libraries with only two rather short average fragment lengths (350 or 550 bps) (Huptas et al., 2016). The Nextera XT DNA Library preparation kit is suitable to prepare sequencing-ready libraries from small genomes, like bacteria genomes, starting from a much lower DNA amount (i.e., 1 ng of input DNA). Using a single tagmentation enzymatic reaction, sample DNA is simultaneously fragmented and tagged with adapters. Then a limited-cycle PCR amplifies tagged DNA and adds sequencing indexes.

Similar to pre-sequencing protocols (wet-lab), a jungle of open-source and Web- or Windows-based software are currently available for processing of post-sequencing data. Bioinformatics software can be in the form of stand alone software or pipelines to carry out automated analysis starting from raw short-or long-reads. Reads can be assembled into longer continuous stretches of sequences (contigs) or mapped against reference genomes or genomic regions of interest for variant calling and gene-by-gene analysis (i.e. extended multilocus sequence typing (MLST) based on WGS) (Carriço et al., 2018; Lynch et al., 2016; Taboada et al., 2017; Nadon et al., 2017). However, turning sequencing reads into meaningful biological data is not trivial, and each analytical step (i.e. genome assembly, reads alignment and variant analysis) can have a considerable effect on outputs (Olson et al., 2015). The choice of most appropriate computational approach depends on several factors, including the availability of complete reference genomes(Ekblom and Wolf, 2014). Therefore, an extensive knowledge of alternative approaches and their strategies is needed to estimate relative advantages and disadvantages (Carriço et al., 2018; Taboada et al., 2017).

The aim of this study was to compare the sequencing outputs achieved testing six bacterial cultures and six extracted DNA samples received as part of a GMI PT 2015 and analysed using three different DNA extraction methods, two library preparation strategies and two sequencing platforms. Several parameters have been quantified to compare the sequencing outputs, such as (1) reads quality, assembly and alignment statistics; (2) differences in single nucleotide polymorphisms (SNPs) detected applying an assembly-based and an alignment-based strategy; (3) antimicrobial resistance genes identified on *de novo* assemblies of all sequenced genomes.

2. Materials and methods

The experimental design followed in this study is summarized in Table 1.

2.1. Bacterial strains and DNA samples

Six bacterial cultures (herein referred as BACT) and six DNA samples (herein referred as DNA) were provided to the University of Bologna within the GMI PT 2015 (Table 2).

The bacterial cultures were shipped lyophilised as KwikStik's (Microbiologics, St Cloud, Minnesota). They were represented by two *Salmonella enterica* strains, labelled as BACT1 and BACT2; two *Escherichia coli* strains, labelled as BACT3 and BACT4; two *Staphylococcus aureus* strains, labelled as BACT5 and BACT6. Upon arrival, the KwikStik's were streaked onto Brain Heart Infusion Agar (BHA, Oxoid, Milan, Italy) and incubated at 37 °C overnight.

The DNA samples of the six strains, labelled as DNA1 to DNA6, were shipped as dried samples in vials containing a minimum of 2 μg of DNA using a DNA stabilizing agent (DNAstable® Plus, Biomatrica). On arrival, the DNA samples were re-suspended in 100 μl of sterile water (Molecular biology reagent, W4502, Sigma-Aldrich, Milan, Italy) and mixed by gentle vortexing for 2 min. Rehydrated samples were then stored at room temperature (15–25 °C).

2.2. DNA extraction and quantification

For DNA extraction, all bacterial cultures (Table 2) were incubated in Brain Heart Infusion at 37 °C overnight. At the end of the incubation period, 500 μl of broth were centrifuged at 5000 x g for 10'. Pellets obtained from cultures BACT1 to BACT4 were suspended in 180 μl of ATL buffer and submitted to DNA extraction using the DNeasy Blood & Tissue Kit (Qiagen, Milan, Italy). Pellets obtained from cultures BACT5 and BACT6 were suspended in 100 μl of R4 buffer and submitted to DNA extraction using the ChargeSwitch® gDNA Mini Bacteria Kit (Invitrogen, Milan, Italy). The DNA extraction from samples DNA1 to DNA6 was performed using the Easy-DNA™ Kit (Invitrogen). At the end of the DNA extractions, samples were measured on a biospectrometer (BioSpectrometer, Eppendorf, Milan, Italy; QuBit) to assess DNA quantity and quality.

2.3. Library preparations and quantification

Total DNA extracted from each individual bacterial culture (BACT1 to BACT6) as well as DNA samples (DNA1 to DNA6) were divided into two aliquots of 10 μl each. Five μl containing 1 ng of input DNA from one of the two aliquots were enzymatically fragmented and tagged with sequencing adapters using Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA). Furthermore, 11 μl containing 1 ng of input DNA from the second aliquot were mechanically fragmented and submitted to library preparation performing the TruSeq Nano DNA Library

Table 1
Experimental plan detailing DNA extraction procedures, library preparation protocol and sequencing strategies applied to the six bacterial cultures (BACT) and the six DNA (DNA) samples tested.

	BACT 1-4			BACT 5-6			DNA 1-6		
DNA extraction	DNeasy Blood & Tissue Kit			ChargeSwitch® gDNA Mini Bacteria Kit			Easy-DNA™ Kit		
Library preparation	Nextera XT DNA Library Preparation Kit		TruSeq Nano DNA Library Prep	Nextera XT DNA Library Preparation Kit		TruSeq Nano DNA Library Prep	Nextera XT DNA Library Preparation Kit		TruSeq Nano DNA Library Prep
Sequencing	MiSeq (2 × 250 bp)	HiScanSQ (2 × 100 bp	MiSeq (2 × 250 bp)	MiSeq (2 × 250 bp)	HiScanSQ (2 × 100 bp)	MiSeq (2 × 250 bp)	MiSeq (2 × 250 bp)	HiScanSQ (2 × 100 bp)	MiSeq (2 × 250 bp)

Prep (Illumina). At the end of each protocol, all libraries were quantified by using the Qubit® fluorimeter (Thermofisher, Milan, Italy) and the Fragment analyzer (Advanced Analytical) (Heidelberg, Germany) and then normalized.

2.4. Sequencing

Libraries prepared using Nextera XT DNA Library Preparation Kit were sequenced using both the MiSeq sequencing platform (Illumina) and the HiScanSQ sequencer (Illumina). Libraries prepared using TruSeq Nano DNA Library Prep were sequenced with the MiSeq platform only. MiSeq sequencing for both libraries strategies was carried out using the MiSeq Reagent Kit v2 500 Cycles (2 × 250 bp read length) (Illumina). After sequencing, short-reads obtained from pooled libraries were de-multiplexed using the on-board MiSeq Reporter software (v2.3.32) available with the sequencing platform. Shotgun whole genome sequencing in the HiScanSQ sequencer (Illumina) was carried out using the TruSeq SBS Kit v3-HS (200-cycles) (2 × 100 bp read length).

2.5. Sequencing data analysis

The overall sequencing reads quality was assessed using FastQC, a toolbox for displaying sequence statistics and quality control for high throughput sequence data. Low-quality reads and Illumina adapters were filtered using Trimmomatic (Bolger et al., 2014), a flexible command-line based tool for trimming of pair-end reads, with default parameters. Two strategies for polymorphism detection in the sequencing data were followed: one assembly-based and the other alignment-based. In the assembly-based strategy, filtered raw reads were assembled into contigs using Velvet (Zerbino and Birney, 2008) with the *k*-mer corresponding to the 60% of the read length. From MUMmer alignment tool (Kurtz et al., 2004), nucmer and show-snps were used for aligning contigs against appropriate reference genomes (Table 1) and for SNPs detection, respectively. In the alignment-based strategy, reads were aligned to the reference genomes (Table 1) using the Burrows-Wheeler Aligner (BWA) software package (http://bio-bwa.sourceforge.net/) (Li and Durbin, 2009). Then, SortSam, MarkDuplicates and BuildBamIndex Java command line based tools for manipulating HTS data and formats from Picard package (https://github.com/broadinstitute/picard) were applied for sorting reads in the alignment files, marking PCR duplicate reads and creating the index file. Finally, HaplotypeCaller, SelectVariants, and VariantFiltration tools from the GATK package were applied in accordance with the best practice described by Van der Auwera et al. (2013) for SNPs detection via local re-assembly of haplotypes and filter variant calls. Parameters for SNPs filtration were the followings: quality by depth <2; fisher strand bias >60; mapping quality <40; ReadPosRankSum < -8; and MQRankSum < -12.5. In order to test the impact of the different wet-lab protocols on the quality of *de novo* assembly in terms of accessory genes, the Web-based service ResFinder (Zankari et al., 2012) was used to identify antimicrobial resistance genes (ARGs) on *de novo* assemblies obtained in this study.

Table 2
Bacterial strains (BACT) and DNA samples (DNA) tested as part of the GMI PT 2015

Bacterial strain	Sample ID	Ref. genome sequence
Salmonella enterica	BACT1/BACT2	STA00025
E. coli	BACT3/BACT4	EC002143
S. aureus	BACT5/BACT6	SAH596
DNA sample		
Salmonella enterica	DNA1/DNA2	STA00025
E. coli	DNA3/DNA4	EC002143
S. aureus	DNA5/DNA6	SAH596

2.6. Statistical analysis

The outputs of all sequencing data were compared based on several parameters in relation to reads quality (e.g. total sequencing base count and coverage), assembly (e.g. total length, N50 values and contig sizes) and alignment (e.g. number of mapped, unmapped, and duplicate reads). In addition, number of SNPs in both assembly-based and alignment-based strategies and the distribution of the variant-level parameters DP and QD for the alignment-based SNPs were statistically assessed. Moreover, ANOVA two-way tests were performed to evaluate the impact of different DNA extraction procedures, library preparation protocols, sequencing platforms and their interaction on the evaluated features. The statistical analyses were performed by using the statsmodels Python package.

3. Results

3.1. Read lengths and coverages obtained using different DNA extraction procedures, library preparation protocols and sequencing platforms

The MiSeq sequencing outputs from BACT and DNA samples expressed in millions base count (Mbp) and fold sequencing coverage were calculated and reported in Fig. 1 and Fig. 2, respectively. For the BACT samples, higher values of base count and coverage were obtained using the Nextera XT DNA Library Preparation Kit in comparison to the TruSeq Nano DNA Library Prep (529.89 vs 375.06 Mbp; 138.02 vs 94.63 fold) (Table 3).

On the contrary, for DNA samples higher base count and sequencing coverage values were achieved preparing libraries using TruSeq Nano DNA Library Prep in comparison to Nextera XT DNA Library Preparation Kit (482.28 vs 368.45 Mbp; 117 vs 89.59 fold) (Table 2). The two-way ANOVA analysis revealed that changes in base count and coverage were not significantly affected by the library preparation kit ($P = 0.73$ and 0.71, respectively) as well as by DNA extraction procedures ($P = 0.65$ and 0.56, respectively) with reference to the DNA extraction protocols applied both on BACT and DNA samples. Nevertheless, the interaction between the library preparation kits and the DNA extraction protocols significantly affected the base count ($P = 0.03$), although it did not affect sequencing coverage ($P = 0.12$).

With the exception of BACT5 and BACT6, the sequencing outputs in term of base count (Fig. 3) and fold sequencing coverage (Fig. 4) obtained for BACT samples using Nextera XT DNA Library Preparation Kit were higher by sequencing with the HiScanSQ rather than the MiSeq platform (663.01 vs 529.89 Mbp; 155.25 vs 138.02 fold) (Table 2). The same

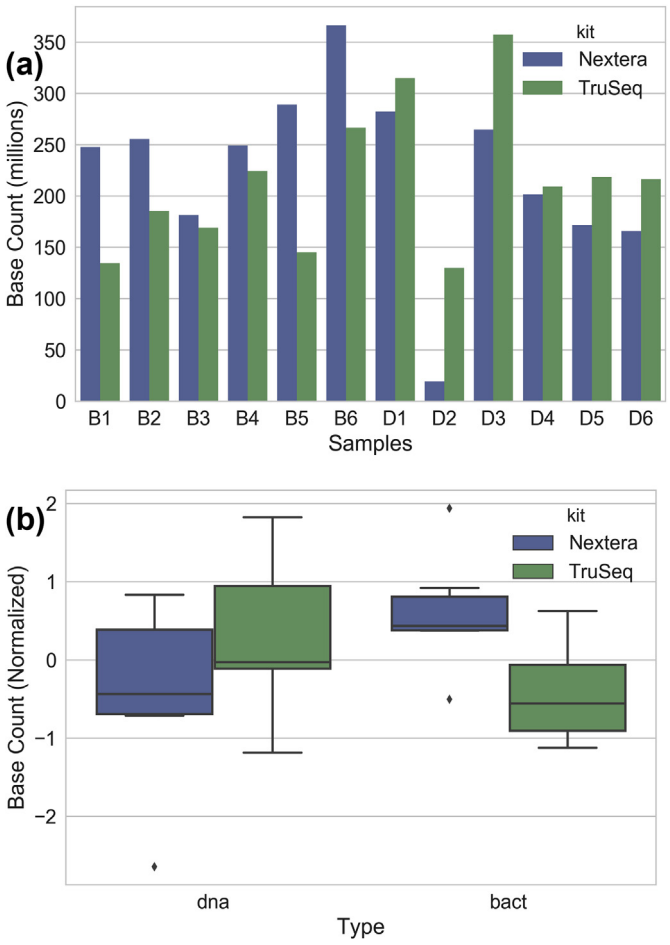

Fig. 1. Barplot showing the base counts (a) and boxplot showing the normalized (Z-scores) base counts (b) of samples (MiSeq sequencing) prepared by two different library kits (Nextera XT and TruSeq) and two DNA extraction methods (DNA vs BACT).

results applied to samples DNA4, DNA5 and DNA6, whereas MiSeq platform provided better performances for DNA1, DNA2 and DNA3 (Table 3). Overall, the two-way ANOVA analysis did not reveal significant differences between base counts and coverage values obtained using the two sequencing platforms, the two extraction procedures and their interactions (P = 0.06, 0.29 and 0.56 and P = 0.14, 0.43 and 0.37, respectively).

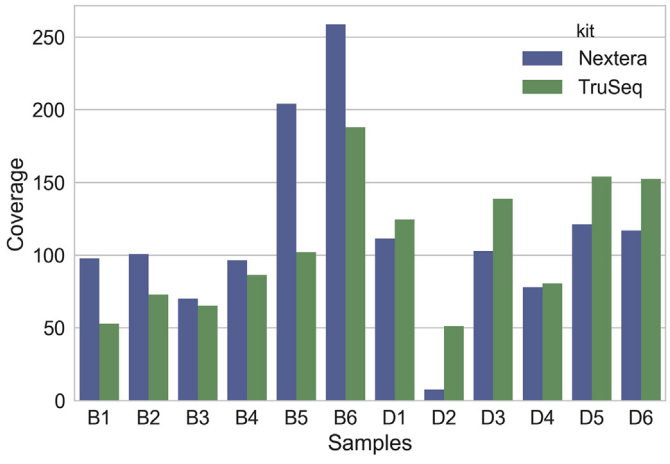

Fig. 2. Barplot showing the coverage of samples (MiSeq sequencing) prepared by NexteraXT and TruSeq, and different extraction procedures (DNA vs BACT).

3.2. Contigs mean size and N50 values of de novo assemblies obtained using different DNA extraction procedures, library preparation protocols and sequencing platforms

For BACT samples, the contig mean sizes ranged between 9.39 and 207.6 kb, whereas N50 values ranged between 18.47 to 4969.86 kb. For DNA samples, the contig mean sizes ranged between 9.20 and 92.80 kb, whereas N50 values ranged between 17 to 5101.23 kb. The only exception was sample DNA2, which showed poor assembly parameters values (contig mean size and N50 ranging from 0.23 and 0.50 kb). Overall, contig mean sizes and N50 values were not significantly affected by the library preparation kit (contig mean size P = 0.72; N50 P = 0.74), DNA extraction protocol (contig mean size P = 0.27; N50 P = 0.93) or their interactions (contig mean size P = 0.92; N50 P = 0.76) (Table 2).

Furthermore, the same parameters were not affected by the sequencing platform (contig mean size P = 0.39; N50 P = 0.60), DNA extraction protocols (contig mean size P = 0.27; N50 P = 0.56) or their interactions (contig mean size P = 0.91; N50 P = 0.70) (Table 2).

3.3. Assembly- and alignment-based SNPs calling obtained using different DNA extraction procedures, library preparation protocols and sequencing platforms

A similar number of SNPs was detected by both the assembly- and alignment-based strategy for DNA and BACT samples prepared using different library preparation procedures and sequencing platforms (Table 3). Except for BACT1, DNA1 and in the assembly-based results for DNA2 sample, a SNPs count ranging from 10^4 to 10^5 was obtained using both sequencing platforms (Table 3). Excluding the DNA2 sample, similar levels of reads duplicates were observed within BACT and DNA samples data for each adopted procedure (standard deviation ranging between 0.3 and 1.68). Even though, the application of different library preparation kits and sequencing platforms led to relevant differences in the duplicates percentages (Table 3). The lowest numbers of PCR duplicates were obtained by MiSeq sequencing in comparison to HiScanSQ platform using both library kits, in particular the TruSeq Nano DNA Library Prep (Table 3). Overall, the number of duplicate reads was significantly affected both by the library preparation kits (P = 6.4e-4) and by the sequencing platform (P = 4.0e-6) (Fig. 5).

3.4. Antimicrobial resistance gene identification on de novo assemblies obtained using different DNA extraction procedures, library preparation protocols and sequencing platforms

In order to assess the quality of *de novo* assemblies based on accessory gene comparison, antimicrobial resistance genes (ARGs) were identified in all samples. Genomes of BACT1-DNA1 (*Salmonella enterica*), BACT4-DNA4 (*Escherichia coli*) and BACT6-DNA6 (*Staphylococcus aureus*) harbored no ARGs. All the other genomes were positive for genes associated to resistance to aminoglycosides, beta-lactams, fluoroquinolones, macrolides, phenicols, sulfonamides, tetracyclines, trimethoprim (Table 4). ARGs patterns were the same in all genomes belonging to the same strain, irrespective of the different tested protocols, except for DNA2 genomes (NXT-MiSeq, TS-MiSeq and NXT-HiScanSQ), BACT2 (TS-MiSeq) and BACT3 (NXT-MiSeq).

Regarding the three genomes obtained from DNA2, no genes or a lower number related to antimicrobial resistance were detected. This result confirmed that the initial DNA extract for this sample was degraded, as already suggested by the very low-quality parameters of *de novo* assemblies. On the other hand, one (e.g. *tetA*) or two (e.g. *sul3* and $bla_{\text{CTX-M-1}}$) genes were missing in BACT2 (TS-MiSeq) and BACT3 (NXT-MiSEq), respectively. In these particular cases, although the slightly lower quality of the assembly, the absence of the ARGs might be related to their localization on the tip of contigs. Despite the missing genes, the predicted phenotype of antimicrobial susceptibility might not change since additional genes associated to conferring resistance to the same antimicrobial class have been identified.

Table 3
Base count, coverage, contig mean size, N50 values, assembly- and alignment-based single nucleotide polymorphisms (SNPs) obtained for bacterial cultures and percentage of PCR duplicates for BACT samples, labelled as BACT1-BACT6, and DNA samples, labelled as DNA1-DNA6, by using Nextera XT DNA Library Preparation Kit and TruSeq Nano DNA Library Prep before sequencing in the MiSeq and HiScanSQ platforms.

	NexteraXT		TruSeq		NexteraXT		TruSeq
	MiSeq	HiScanSQ	MiSeq		MiSeq	HiScanSQ	MiSeq
Base count (Mbp)							
BACT 1	495.6	698.76	268.95	DNA1	565.02	532.49	630.57
BACT 2	511.45	860.69	370.95	DNA 2	38.50	24.26	259.88
BACT 3	362.73	724.68	338.24	DNA 3	529.49	457.44	715.07
BACT 4	498.53	701.67	448.82	DNA 4	402.94	1200.64	418.39
BACT 5	578.4	463.21	289.84	DNA 4	343.44	699.33	436.98
BACT 6	732.69	529.06	533.59	DNA 6	331.36	778.87	432.83
Mean	529.89	663.01	375.06	Mean	368.45	615.5	482.28
sd	121.59	148.33	100.46	sd	187.95	389.24	163.84
Coverage							
BACT 1	97.75	137.33	52.85	DNA1	111.65	104.68	124.70
BACT 2	100.97	169.16	72.91	DNA 2	7.60	4.79	51.11
BACT 3	70.14	139.80	65.25	DNA 3	102.92	88.28	138.94
BACT 4	96.48	135.36	86.57	DNA 4	77.90	231.63	80.68
BACT 5	204.16	163.33	102.16	DNA 4	121.19	246.57	154.02
BACT 6	258.67	186.55	188.08	DNA 6	116.92	274.61	152.56
Mean	138.02	155.25	94.63	Mean	89.69	158.42	117.00
sd	75.17	20.94	48.85	sd	43.05	107.73	42.03
N50							
BACT 1	26.65	4784.21	1674.25	DNA1	4780.86	4782.83	222.59
BACT 2	4086.51	4969.86	2816.93	DNA 2	0.55	0.23	23.30
BACT 3	18.47	46.84	75.94	DNA 3	34.59	51.74	4132.07
BACT 4	3797.37	48.52	2991.98	DNA 4	5085.87	2604.86	5101.23
BACT 5	2886.12	2851.16	115.56	DNA 4	17.01	52.82	115.56
BACT 6	2687.45	44.89	2686.40	DNA 6	2491.99	48.83	2666.79
Mean	2250.42	2124.24	1726.84	Mean	2068.47	1256.88	2043.59
sd	1804.77	2393.71	1344.18	sd	2419.19	2009.47	2245.62
Contig mean size (kb)							
BACT 1	14.14	74.66	76.86	DNA1	92.81	76.95	36.73
BACT 2	38.81	86.90	59.51	DNA 2	0.56	0.24	16.65
BACT 3	8.99	17.45	21.39	DNA 3	13.38	17.78	62.71
BACT 4	28.60	9.40	37.83	DNA 4	31.26	26.06	37.58
BACT 5	100.40	59.11	27.44	DNA 4	9.20	14.60	27.70
BACT 6	180.03	17.53	207.61	DNA 6	88.41	17.52	89.33
Mean	61.82	44.17	71.77	Mean	39.27	25.52	45.11
sd	66.57	33.50	69.69	sd	41.03	26.55	26.47
Number SNPs/kb (number tot SNPs) Assembly-based							
BACT 1	0.15 (744)	0.14 (737)	0.15 (745)	DNA1	0.14 (733)	0.15 (760)	0.15 (744)
BACT 2	7.77 (39517)	7.85 (39954)	7.83 (39826)	DNA 2	0.02 (106)	0.20 (1006)	0.00 (0)
BACT 3	10.20 (52865)	10.38 (53809)	10.35 (53645)	DNA 3	10.31 (53451)	10.39 (53845)	10.37 (53734)
BACT 4	18.54 (96091)	19.01 (98531)	18.95 (98246)	DNA 4	18.91 (98021)	18.92 (98076)	18.92 (98102)
BACT 5	18.80 (53321)	18.81 (53348)	18.78 (53268)	DNA 4	17.90 (50781)	18.80 (53324)	18.80 (53334)
BACT 6	6.19 (17559)	6.25 (17731)	6.23 (17672)	DNA 6	6.15 (17445)	6.25 (17734)	6.23 (17672)
Mean	10.27 (43349)	10.40 (44018)	10.38 (43900)	Mean	8.90 (36756)	9.11 (37457)	9.07 (37264)
sd	7.30 (33065)	7.39 (33868)	7.38 (33772)	sd	8.33 (38059)	8.48 (38103)	8.52 (38306)
Number SNPs/kb (number tot SNPs) Alignment-based, after filtering							
BACT 1	0.34 (1716)	0.31 (1571)	0.31 (1577)	DNA1	0.33 (1665)	0.31 (1596)	0.34 (1740)
BACT 2	8.02 (40812)	7.98 (40595)	7.91 (40262)	DNA 2	6.13 (31210)	4.37 (22246)	4.28 (21777)
BACT 3	10.53 (54581)	10.44 (54130)	10.43 (54070)	DNA 3	10.56 (54737)	10.44 (54144)	10.55 (54707)
BACT 4	19.25 (99767)	19.15 (99288)	19.07 (98830)	DNA 4	19.28 (99946)	19.09 (98956)	19.12 (99089)
BACT 5	16.97 (48141)	17.35 (49216)	16.71 (47396)	DNA 4	17.21 (48811)	17.30 (49071)	16.42 (46562)
BACT 6	6.06 (17186)	6.10 (17303)	6.02 (17065)	DNA 6	6.10 (17294)	6.09 (17260)	6.01 (17042)
Mean	10.19 (43700)	10.22 (43683)	10.07 (43200)	Mean	9.93 (42277)	9.6 (40545)	9.45 (40152)
sd	7.02 (33928)	7.08 (33795)	6.94 (33629)	sd	7.24 (34408)	7.42 (34823)	7.27 (34853)
PCR duplicates percentage							
BACT 1	5.807	12.547	0.763	DNA1	6.476	10.529	4.888
BACT 2	5.657	11.838	0.968	DNA 2	2.366	2.752	0.728
BACT 3	4.543	12.267	1.093	DNA 3	6.742	9.560	5.509
BACT 4	4.724	10.034	1.298	DNA 4	3.241	10.693	3.241
BACT 5	4.996	8.720	1.580	DNA 4	3.863	10.920	3.863
BACT 6	5.894	10.018	1.541	DNA 6	3.555	12.322	3.555
Mean	5.270	10.904	1.207	Mean	4.374	9.463	3.631
sd	0.58	1.53	0.32455006	sd	1.80395778	3.405668018	1.65816361

4. Discussion

Whole Genome Sequencing (WGS) is quickly moving from proof-of-concept research into routine clinical and public health use (Crisan et al., 2018). In particular, a rapid shift towards implementation of this technologies for routine use in national public health reference laboratories across the EU/EEA countries has been determined (Revez et al., 2017). This has led to an exponential increase of NGS protocols and pipelines, which have raised the need to validate and compare available methods in order to promote standardization, both in the wet- and dry-lab parts. Few studies focusing on a limited number of bacterial isolates/species and analytical methods (Harris et al., 2013; Quail et al.,

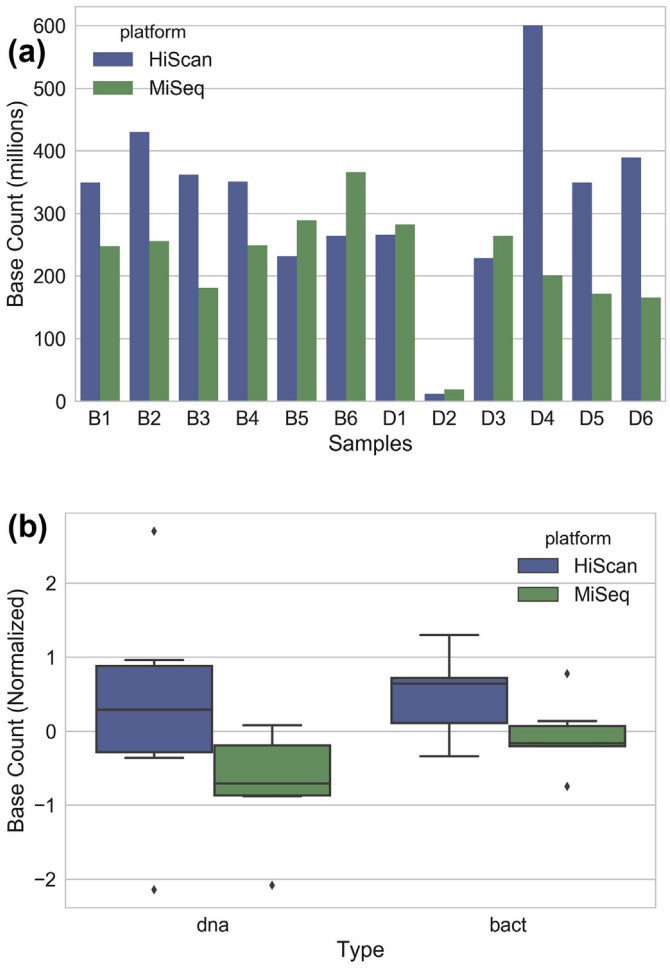

Fig. 3. Barplot showing the base counts (a) and boxplot showing the normalized (Z-scores) base counts (b) of samples sequenced by two platforms (HiScanSQ and MiSeq) and two DNA extraction procedures (DNA vs BACT).

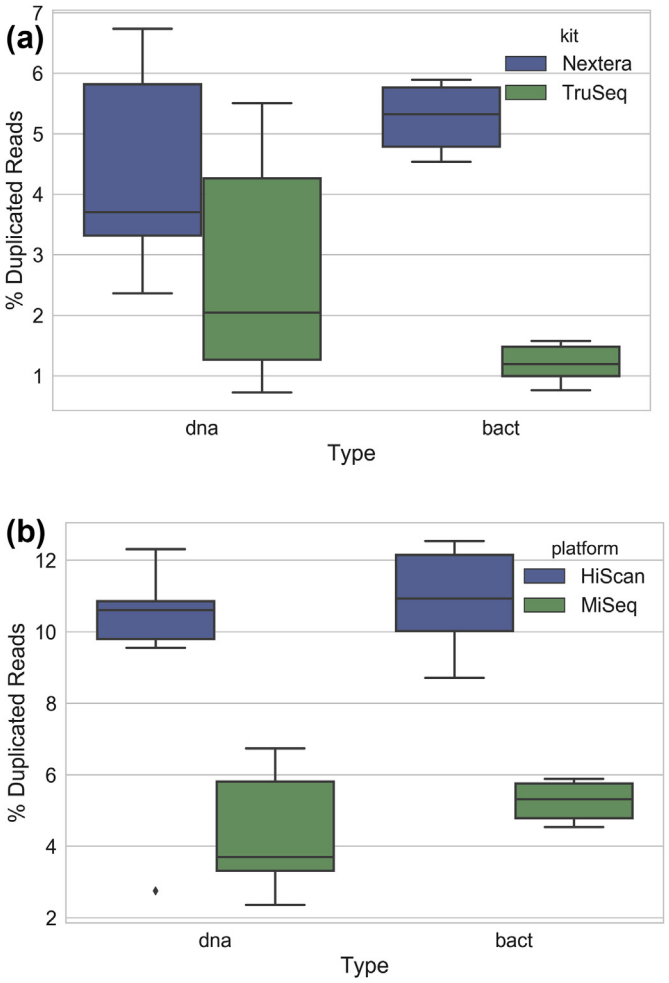

Fig. 5. Boxplots showing the percentage of duplicated reads in samples prepared by two extraction procedures (DNA vs BACT), comparing the library preparation kits (a) and DNA sequencing platforms (b).

2012) have been performed to assess and validate the comparability of WGS data generated using different platforms (Loman et al., 2012) with different error profiles based on the library preparation methods (Besser et al., 2018). The present study was designed to compare the quality of sequencing outputs achieved for six bacterial cultures and six DNA samples obtained as part of the proficiency test organised by the Global

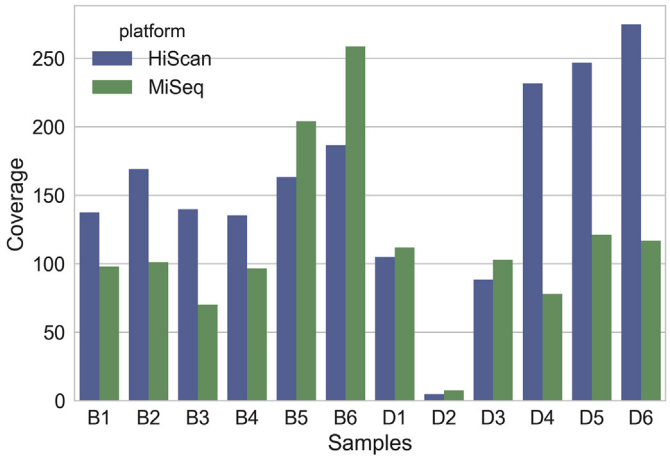

Fig. 4. Barplot showing the coverage of samples sequenced by HiScanSQ and MiSeq, and different extraction procedures (DNA vs BACT).

Microbial Identifier in 2015 using three different protocols for DNA extraction, two library preparation kits and two NGS Illumina sequencing platforms. Finally, yet importantly, assembly- and alignment-based strategies for SNPs calling were compared. We evaluated different parameters to cover different research questions related to bacterial genome sequencing. For example, detection of new genes or genomic rearrangements (e.g. duplications and deletions) require good genome reconstructions (assemblies) with high values of parameters like N50, total length, and contig size. If the interest relies on genotyping and describing genetic diversity of closely related bacterial strains, one should consider alignment-based and variant level parameters, since they reflect the sensitivity and specificity for detection of single nucleotide variants using the sequencing data. Overall, the sequenced genomes included in this study showed similar sequencing and assembly parameters statistics with reference to sequencing outputs, in millions base count, fold sequencing coverage, contig mean sizes and N50 values. Draft genomes of isolate DNA2 represent an exception, suggesting a significant degradation of the starting extracted DNA. None of the aforementioned statistics were significantly affected by the sequencing platform, the extraction protocol or the library preparation kit applied in this study. Nevertheless, a significant difference was observed on the number of duplicate reads in relation to both library preparation kits and sequencing platforms. Duplicates are considered as a concern because they can lead to false positive variant calls (Ebbert et al., 2016). Nevertheless, the study of Ebbert et al. confirmed that removing duplicates may not affect the variant calling output. Although few samples were included

Table 4
Antimicrobial resistance genes detected by ResFinder. Only antimicrobial classes for which genes were detected are reported.

Sample	Protocol	Antimicrobial resistance genes per antibiotic class							
		Aminoglycosides	b-lactams	Fluoro-quinolones	Macrolides	Phenicols	Sulfonamides	Tetracyclines	Trimethoprim
BACT1	NXT-MiSeq	-	-	-	-	-	-	-	-
	TS-MiSeq	-	-	-	-	-	-	-	-
	NXT-HiScanSQ	-	-	-	-	-	-	-	-
BACT2	NXT-MiSeq	*aac(6′)-Iic, strA, strB*	*blaSHV-12, blaCTX-M-15, blaTEM-1B*	*QnrB2*	*ere(A)*	*floR*	*sul2, sul1*	*tet(A), tet(D)*	*dfrA18*
	TS-MiSeq	*aac(6′)-Iic, strA, strB*	*blaSHV-12, blaCTX-M-15, blaTEM-1B*	*QnrB2*	*ere(A)*	*floR*	*sul2, sul1*	*tet(D)*	*dfrA18*
	NXT-HiScanSQ	*aac(6′)-Iic, strA, strB*	*blaSHV-12, blaCTX-M-15, blaTEM-1B*	*QnrB2*	*ere(A)*	*floR*	*sul2, sul1*	*tet(D),tet(A)*	*dfrA18*
BACT3	NXT-MiSeq	*aadA2, aadA1, strA, strB*	*blaTEM-1B*	-	*mph(B)*	*cmlA1, catA1*	*sul2, sul1*	*tet(A)*	*dfrA1, dfrA12*
	TS-MiSeq	*aadA2, aadA1, strA, strB*	*blaTEM-1B, blaCTX-M-1*	-	*mph(B)*	*cmlA1, catA1*	*sul3, sul2, sul1*	*tet(A)*	*dfrA1, dfrA12*
	NXT-HiScanSQ	*aadA2, aadA1, strA, strB*	*blaTEM-1B, blaCTX-M-1*	-	*mph(B)*	*cmlA1, catA1*	*sul3, sul2, sul1*	*tet(A)*	*dfrA1, dfrA12*
BACT4	NXT-MiSeq	-	-	-	-	-	-	-	-
	TS-MiSeq	-	-	-	-	-	-	-	-
	NXT-HiScanSQ	-	-	-	-	-	-	-	-
BACT5	NXT-MiSeq	*str*	*mecA, blaZ*	-	-	-	-	-	-
	TS-MiSeq	*str*	*mecA, blaZ*	-	-	-	-	-	-
	NXT-HiScanSQ	*str*	*mecA, blaZ*	-	-	-	-	-	-
BACT6	NXT-MiSeq	-	-	-	-	-	-	-	-
	TS-MiSeq	-	-	-	-	-	-	-	-
	NXT-HiScanSQ	-	-	-	-	-	-	-	-
DNA1	NXT-MiSEq	-	-	-	-	-	-	-	-
	TS-MiSeq	-	-	-	-	-	-	-	-
	NXT-HiScanSQ	-	-	-	-	-	-	-	-
DNA2	NXT-MiSeq	*aac(6′)-Iic, strA, strB*	*blaCTX-M-3*	*QnrB2*	*ere(A)*	*floR*	*sul2, sul1*	*tet(D)*	*dfrA18*
	TS-MiSeq	-	-	-	-	-	-	-	-
	NXT-HiScanSQ	*aac(6′)-Iic, strA, strB*	-	-	-	-	*sul2, sul1*	-	*dfrA18*
DNA3	NXT-MiSeq	*aadA2, aadA1, strA, strB*	*blaTEM-1B, blaCTX-M-1*	-	*mph(B)*	*cmlA1, catA1*	*sul3, sul2, sul1*	*tet(A)*	*dfrA1, dfrA12*
	TS-MiSeq	*aadA2, aadA1, strA, strB*	*blaTEM-1B, blaCTX-M-1*	-	*mph(B)*	*cmlA1, catA1*	*sul3, sul2, sul1*	*tet(A)*	*dfrA1, dfrA12*
	NXT-HiScanSQ	*aadA2, aadA1, strA, strB*	*blaTEM-1B, blaCTX-M-1*	-	*mph(B)*	*cmlA1, catA1*	*sul3, sul2, sul1*	*tet(A)*	*dfrA1, dfrA12*
DNA4	NXT-MiSeq	-	-	-	-	-	-	-	-
	TS-MiSeq	-	-	-	-	-	-	-	-
	NXT-HiScanSQ	-	-	-	-	-	-	-	-
DNA5	NXT-MiSeq	*str*	*mecA, blaZ*	-	-	-	-	-	-
	TS-MiSeq	*str*	*mecA, blaZ*	-	-	-	-	-	-
	NXT-HiScanSQ	*str*	*mecA, blaZ*	-	-	-	-	-	-
DNA6	NXT-MiSeq	-	-	-	-	-	-	-	-
	TS-MiSeq	-	-	-	-	-	-	-	-
	NXT-HiScanSQ	-	-	-	-	-	-	-	-

in this study, as in other proficiency tests (Mellmann et al., 2017; Timme et al., 2018), the results show that a significant lower number of duplicates was associated to the TruSeq library preparation kit and to the MiSeq sequencing platform, suggesting the possibility to achieve a better sequencing performance when applying these strategies in WGS-based study.

Recently enhanced and newly developed bioinformatics tools, such as SPAdes (Bankevich et al., 2012) and SKESA (Souvorov et al., 2018), are showing great improvements on the state-of-the-art of genome assemblers. However, a latest comprehensive study on *de novo* genome assemblers performance identified Velvet as one of the best tool for assembly of paired-end prokaryotic data in comparison to other tools (i.e. Abyss, Edena, SGA, Ray, SSAKE, Perga), showing a mean N50 value close to 25 kb (Khan et al., 2018). This value is in the range (from 17 to 5085 kb) of the N50 calculated on *de novo* assemblies for the samples from this study, suggesting the high reliability of *de novo* assemblies obtained with the selected WGS approaches. The only exception was DNA2 sample for which poor assembly parameters were calculated for all related *de novo* assemblies irrespective of the DNA extraction procedures, library preparation kits or sequencing platforms. This observation highlighted that the poor quality of the starting DNA affected all downstream applications.

Comparative bioinformatics analysis of SNPs and antimicrobial resistance genes showed congruent results for all the sequencing outputs, regardless of the adopted pre-sequencing strategy. In case of comparable reliabilities, two crucial points to take into consideration are cost and time when choosing one approach or the other. Comparing the two DNA extraction procedures, similar costs (around 5 €) and time (approximately half a day for 24 samples) were observed. On the contrary, significant differences can be registered by comparing DNA library preparation kits and sequencing platforms. In particular, although the time of library preparation is similar (1 day), TruSeq kits are four times more expensive than Nextera XT ones. Huge differences were found also for sequencing kits on both cost and time. The sequencing kit for HiScanSQ platform is approx. 142 € per sample and the run lasts in 11 days. The sequencing kit for MiSeq is 68 € and the run lasts in 2 days. Moreover, even though the output should be higher using HiScanSQ (1E11 bases) than MiSeq (8,5E09 bases), the results of this study show that the less expensive Illumina sequencing platforms may deliver data with comparable quality and performance. Addressing the time of analysis, a useful parameter for the comparison of technologies is the time complexity linked to the rate of growth of time strictly correlated to the amount of input (samples or data). Both library preparation kits enable the preparation of up to 24 libraries thus showing similar time complexities. The difference is more related to the platform of choice. Although the number of samples to be sequenced on the same run is strictly linked to the expected final coverage, as a rule of thumbs considering an expected coverage of 80X for a genome of 5 Mb (*E. coli*), 250 samples can be included in a single run in HiScanSQ vs only 21 on MiSeq. This means that in case of 250 samples, 11 days would be required by HiscanSQ versus approx. 19 days by MiSeq. Therefore, HiScanSQ is particularly time saving when a higher number of samples needs to be sequenced.

In accordance to the conclusion of Harris et al. (2013), along with performance, costs and time effectiveness have to be taken into account when choosing a specific sequencing protocol. However, it is worth to notice that many efforts have been recently made in long-read sequencing (e.g. nanopore and single-molecule real-time sequencing), a more expensive technology which appears to bring advantages for both *de novo* and alignment-based genome assembly as well as for identifying large structural variation (Wick et al., 2017). Despite the benefits of generating high quality complete assemblies, a much higher per-base error rate (5–15% vs <1%) has been observed analyzing long-reads in comparison to Illumina reads (Wick et al., 2017; Besser et al., 2018). However, novel solutions for SNP calling and haplotype assembly for long-reads generating high quality data comparable to Illumina reads alternatives are being developed (Guo et al., 2018).

In conclusion, in the present study, based on 12 DNA samples obtained as part of the proficiency test organised by the Global Microbial Identifier in 2015, different DNA extraction procedures, library preparation protocols and sequencing platforms did not show statistically significant effects on the quality parameters of sequenced reads and *de novo* assemblies as well as on SNPs or ARGs detection. The only exceptions were the significant lower number of duplicates associated to the TruSeq library preparation kit and to the MiSeq sequencing platform. Our results highlight that laboratories have to make multiple remarks when implementing NGS technologies based on the number of samples to be processed and the quality of the output sequencing, especially in order to avoid the use of expensive and time-consuming protocols, which can lead to higher coverage values or longer assemblies but that ultimately show similar performances in terms of variant detection and gene calling, as compared to cheaper alternatives.

Declarations

Author contribution statement

A. De Cesare, F. Pasquali: Conceived and designed the experiments; Wrote the paper.

F. Palma: Performed the experiments; Wrote the paper.

D. Remondini, I. Do Valle, G. Castellani: Analyzed and interpreted the data.

G. Manfreda, R. Hendriksen: Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported by H2020 project named COMPARE, Grant 643476.

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

Aird, D., Ross, M.G., Chen, W.S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., Gnirke, A., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 12 (2), R18.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477.

Bertelli, C., Greub, G., 2013. Rapid bacterial genome sequencing: methods and applications in clinical microbiology. Clin. Microbiol. Infect. 19, 803–813.

Besser, J., Carleton, H.A., Gerner-Smidt, P., Lindsey, R.L., Trees, E., 2018. Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clin. Microbiol. Infect. 24 (4), 335–341.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Carriço, J.A., Rossi, M., Moran-Gilad, J., Van Domselaar, G., Ramirez, M., 2018. A primer on microbial bioinformatics for non bioinformaticians. Clin. Microbiol. Infect. 24, 342–349.

Crisan, A., McKee, G., Munzner, T., Gardy, J., 2018. Evidence-based design and evolution of a whole genome sequencing for clinical report for the reference microbiology laboratory. PeerJ 6, e4218.

Deurenberg, R.H., Bathoorn, E., Chlebowicz, M.A., Couto, N., Ferdous, M., García-Cobos, S., Kooistra-Smid, A.M.D., Raangs, E.C., Rosema, S., Veloo, A.C.M., Zhou, K., 2017. Application of next generation sequencing in clinical microbiology and infection prevention. J. Biotechnol. 243, 16–24.

Ebbert, M.T., Wadsworth, M.E., Staley, L.A., Hoyt, K.L., Pickett, B., Miller, J., Duce, J., 2016. Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. BMC Bioinf. 17 (7), 239.

Edwards, D.J., Holt, K.E., 2013. Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. Microb. Inf. Exp. 3 (1), 2.

Ekblom, R., Wolf, J.B., 2014. A field guide to whole-genome sequencing, assembly and annotation. Evolut. Appl. 7 (9), 1026–1042.

Gargis, A.S., Kalman, L., Berry, M.W., Bick, D.P., Dimmock, D.P., Hambuch, T., et al., 2012. Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat. Biotechnol. 30, 1033–1036.

Gullapalli, R.R., Desai, K.V., Santana-Santos, L., Kant, J.A., Becich, M.J., 2012. Next generation sequencing in clinical medicine: challenges and lessons for pathology and biomedical informatics. J. Pathol. Inform. 3, 40.

Guo, F., Wang, D., Wang, L., 2018. Progressive approach for SNP calling and haplotype assembly using single molecular sequencing data. Bioinformation 34 (12), 2012–2018.

Harris, S.R., Török, M.E., Cartwright, E.J., Quail, M.A., Peacock, S.J., Parkhill, J., 2013. Read and assembly metrics inconsequential for clinical utility of whole-genome sequencing in mapping outbreaks. Nat. Biotechnol. 31 (7), 592.

Huptas, C., Scherer, S., Wenning, M., 2016. Optimized Illumina PCR-free library preparation for bacterial whole genome sequencing and analysis of factors influencing de novo assembly. BMC Res. Notes 9 (1), 269.

Khan, A.R., Pervez, M.T., Babar, M.E., Naveed, N., Shoaib, M., 2018. A comprehensive study of de novo genome assemblers: current challenges and future prospective. Evol. Bioinform. Online 14, 1176934318758650.

Kozarewa, I., Ning, Z., Quail, M.A., Sanders, M.J., Berriman, M., Turner, D.J., 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G + C)-biased genomes. Nat. Methods 6 (4), 291–295.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L., 2004. Versatile and open software for comparing large genomes. Genome Biol. 5 (2), R12.

Kwong, J.C., McCallum, N., Sintchenko, V., Howden, B.P., 2015. Whole genome sequencing in clinical and public health microbiology. For. Pathol. 47 (3), 199–210.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.
Loman, N.J., Pallen, M.J., 2015. Twenty years of bacterial genome sequencing. Nat. Rev. Microbiol. 13 (12), 787.
Loman, N.J., Misra, R.V., Dallman, T.J., Constantinidou, C., Gharbia, S.E., Wain, J., Pallen, M.J., 2012. Performance comparison of benchtop high-throughput sequencing platforms. Nat. Biotechnol. 30 (5), 434.
Lynch, T., Petkau, A., Knox, N., Graham, M., Domselaar, G.V., 2016. A primer on infectious disease bacterial genomics. Clin. Microbiol. Rev. 29, 881–913.
Mellmann, A., Andersen, P.S., Bletz, S., Friedrich, A.W., Kohl, T.A., Lilje, B., et al., 2017. High interlaboratory reproducibility and accuracy of next-generation-sequencing-based bacterial genotyping in a ring trial. J. Clin. Microbiol. 55, 908–913.
Moran-Gilad, J., Sintchenko, V., Pedersen, S.K., Wolfgang, W.J., Pettengill, J., Strain, E., Hendriksen, R.S., 2015. Proficiency testing for bacterial whole genome sequencing: an end-user survey of current capabilities, requirements and priorities. BMC Infect. Dis. 15 (1), 174.
Nadon, C., Van Walle, I., Gerner-Smidt, P., Campos, J., Chinen, I., Concepcion-Acevedo, J., Gilpin, B., Smith, A.M., Kam, K.M., Perez, E., Trees, E., 2017. PulseNet International: vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. Eurosurv 22 (23).
Olson, N.D., Lund, S.P., Colman, R.E., Foster, J.T., Sahl, J.W., Schupp, J.M., Keim, P., Morrow, J.B., Salit, M.L., Zook, J.M., 2015. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. Front. Genet. 6, 235.
Oyola, S.O., Otto, T.D., Gu, Y., Maslen, G., Manske, M., Campino, S., Turner, D.J., Macinnis, B., Kwiatkowski, D.P., Swerdlow, H.P., et al., 2012. Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. BMC Genomics 13, 15.
Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., Gu, Y., 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13 (1), 341.
Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J., Durbin, R., Swerdlow, H., Turner, D.J., 2008. A large genome center's improvements to the Illumina sequencing system. Nat. Methods 5 (12), 1005–1010.
Revez, J., Espinosa, L., Albiger, B., Leitmeyer, K.C., Struelens, M.J., ECDC National Microbiology Focal Points and Experts Group, 2017. Survey on the use of whole-genome sequencing for infectious diseases surveillance: rapid expansion of european national capacities, 2015-2016. Front. Public Health 5, 347.
Ronholm, J., Nasheri, N., Petronella, N., Pagotto, F., 2016. Navigating microbiological food safety in the era of whole-genome sequencing. Clin. Microbiol. Rev. 29, 837–857.
Souvorov, A., Agarwala, R., Lipman, D.J., 2018. SKESA: strategic k-mer extension for scrupulous assemblies. Genome Biol. 19 (1), 153.
Taboada, E.N., Graham, M.R., Carriço, J.A., Van Domselaar, G., 2017. Food safety in the age of next generation sequencing, bioinformatics, and open data access. Front. Microbiol. 8, 909.
Timme, R.E., Rand, H., Sanchez Leon, M., Hoffmann, M., Strain, E., Allard, M., Roberson, D., Baugher, J.D., 2018. GenomeTrakr proficiency testing for foodborne pathogen surveillance: an exercise from 2015. Microb. Genom. 4 (7).
Tyler, A.D., Christianson, S., Knox, N.C., Mabon, P., Wolfe, J., Van Domselaar, G., Graham, M.R., Sharma, M.K., 2016. Comparison of sample preparation methods used for the next-generation sequencing of Mycobacterium tuberculosis. PLoS One 11 (2), e0148676.
Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K., Altshuler, D., Gabriel, S., DePristo, M., 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Cur. Prot. Bioinf. 43 (1), 11.10.
Wick, R.R., Judd, L.M., Gorrie, C.L., Holt, K.E., 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. 13 (6), e1005595.
Wielinga, P.R., de Vries, A., van der Goot, T.H., Mank, T., Mars, M.H., Kortbeek, L.M., van der Giessen, J.W., 2008. Molecular epidemiology of Cryptosporidium in humans and cattle in The Netherlands. Int. J. Parasitol. 38 (7), 809–817.
Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. 67 (11), 2640–2644.
Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.