Diagnostic Microbiology and Infectious Disease 95 (2019) 28–33

Contents lists available at ScienceDirect

Diagnostic Microbiology and Infectious Disease

journal homepage: www.elsevier.com/locate/diagmicrobio

In-depth comparison of library pooling strategies for multiplexing bacterial species in NGS

B.H. Muller [a,*], P. Mollon [a], E. Santiago-Allexant [a], F. Javerliat [b], G. Kaneko [a]

[a] bioMérieux, Data Analytics, Marcy l'Etoile F-69280, France
[b] bioMérieux, R&D Microbiology, La Balme Les Grottes F-38390, France

ARTICLE INFO

Article history:
Received 8 March 2019
Received in revised form 26 April 2019
Accepted 27 April 2019
Available online 22 May 2019

Keywords:
Genomic investigation
Whole-genome sequencing
Bacterial genome size polymorphism
Coverage depth
Multi-species multiplexing
Library pooling

ABSTRACT

For bacterial genome sequencing, libraries from different strains are usually multiplexed in a single run. Normalized libraries are most often pooled in equal volumes, as recommended by next-generation sequencing platform manufacturers. This equal-volume strategy is well suited for multiplexing isolates from the same species. However, for runs involving multiple microbial species, an equimolar library pooling is more adapted because of the variation in bacterial genome size. To demonstrate its utility in clinical microbiology, we compared both equal-volume and equimolar strategies using a menu comprising 13 bacterial species involved in healthcare-associated infections. We show that equimolar pooling limits the retesting risk due to insufficient coverage depth, particularly when interspecies genome size difference is more than 2-fold. The use of this alternative strategy for multiplexing pathogenic bacteria should lead to more cost effective whole-genome sequencing applications in clinical microbiology.

1. Introduction

Bacterial genomes vary in size extensively, sometimes even within a species (Bobay and Ochman 2017). As of today, the largest bacterial complete genome in GenBank is the one for the *Sorangium cellulosum* strain So0157-2 at 14,782,125 bp (Han et al. 2013), while the smallest one is described for the *Candidatus Nasuia deltocephalinicola* strain NAS-ALF at only 112,091 bp (Bennett and Moran 2013), more than 2 orders of magnitude smaller. Focusing on the 10 major bacterial species involved in global healthcare-associated infections (HAIs) (Mirande et al. 2018), median genome sizes range from 2.9 megabase pairs (Mbp) for *Staphylococcus aureus* to 6.5 Mbp for *Pseudomonas aeruginosa*.

Whole-genome sequencing (WGS) has become an indispensable tool of modern microbiology (Land et al. 2015). The most recent generation of DNA sequencing platforms can generate accurate whole-genome sequences for a broad range of bacterial species in less than a day (Reuter et al. 2013). WGS has many practical applications, from genome-scale metabolic modeling (Zhu et al. 2018) to fine-scale resolution of bioforensic data (Kalinowski et al. 2018), biosurveillance (Timme et al. 2017), and investigations into infectious disease epidemiology (Parkhill and Wren 2011; Rouard et al. 2019). WGS is particularly attractive as it can be applied simultaneously to large numbers of bacterial isolates of any species (Kanamori et al. 2017) with no need for organism-specific reagents. WGS also allows for surveillance of antibiotic resistance (Snitkin et al. 2012; Reuter et al. 2013; Peacock 2014; Rouard et al. 2018) and possibly even for the detection of virulence factors (Durand et al. 2018). In addition, the resulting data are portable and can be compared with past and future data sets to facilitate historical and geographical tracing of bacterial strains (Coll et al. 2017; Snitkin et al. 2017).

There are several next-generation sequencing (NGS) technologies available for WGS. Third-generation technologies, such as those developed by Pacific Biosciences (PacBio) (Rhoads and Au 2015) or Oxford Nanopore Technologies (ONT) (Tyler et al. 2018), are capable of sequencing single molecules and produce long reads (1–100 kb) in real time (Ameur et al. 2019). Long-read DNA sequencing, which facilitates *de novo* assembly, is particularly suited for elucidation of novel complex genomes containing long repeated sequences. However, some limitations (i.e., high cost per base for PacBio or high error rate for ONT) explain why the use of third-generation sequencing is not yet as widespread in routine applications as second-generation technologies.

Abbreviations: bp, base pair; EqSa, equivalent *Staphylococcus aureus*; gDNA, genomic DNA; HAI, healthcare-associated infection; kb, kilobase (i.e., 1000 bases); Mbp, megabase pairs (i.e., one million base pairs); NGS, next-generation sequencing; WGS, whole-genome sequencing.

* Corresponding author at: bioMérieux, Data Analytics, Chemin de l'Orme, Marcy l'Etoile F-69280, France. Tel.: +33-478-872-378.
E-mail addresses: bruno.muller@biomerieux.com (B.H. Muller), patrick.mollon@biomerieux.com (P. Mollon), emmanuelle.santiago@biomerieux.com (E. Santiago-Allexant), fabien.javerliat@biomerieux.com (F. Javerliat), gael.kaneko@biomerieux.com (G. Kaneko).

https://doi.org/10.1016/j.diagmicrobio.2019.04.014

The latter perform huge numbers of parallel sequencing reactions on a micrometer scale (Heather and Chain 2016) and produce from millions to billions of short reads (50–600 bases) per run. Whole-genome shotgun approaches require NGS library preparation including genomic DNA (gDNA) fragmentation, adapter and barcode/index addition, and PCR amplification. Large numbers of libraries can be pooled and sequenced simultaneously during a single NGS run. For this purpose, sample libraries must be normalized at the same DNA concentration. As recommended by NGS platform manufacturers, normalized libraries are then pooled in equal volumes (i.e., the amount of DNA is the same across all libraries) before sequencing. This equal-volume library pooling strategy is hereafter referred to as the traditional strategy.

Data generated during a WGS run may, however, be useless for some bioinformatics pipelines if coverage depth is insufficient. Such cases require retesting and have an economic impact on routine sequencing because of costs and loss of time. In particular, given the range of bacterial genome size, the retesting risk caused by insufficient depth could be significant when multiplexing multiple bacterial species with widely differing genome lengths. An alternative library pooling strategy was already used (Salipante et al. 2015) to enhance the probability that all sequenced samples achieve sufficient coverage depth. This alternative strategy was based on equimolar sample representation in the library pool. Here we provide an in-depth comparison with the traditional strategy for a menu comprising 13 bacterial HAI-related species and confirm its advantages, especially for genomic epidemiology. The evaluation was conducted on a second-generation platform with direct applicability to other pathogen bacterial species and platforms for WGS applications in clinical microbiology.

2. Material and methods

2.1. Strains and culture conditions

Strains were selected from the internal bioMérieux collection and involved 13 bacterial species (Table 1) (BIOMÉRIEUX EPISEQ© CS 2019). All strains were described in Additional file 1: Table S1. As mentioned in Table S1, some strains belong to a type culture collection (i.e., ATCC or NCTC). When the genome had been sequenced before, the GenBank accession number is indicated. All strains were grown at 37 °C on Columbia agar with 5% sheep blood (bioMérieux, Marcy l'Etoile, France) in an aerobic atmosphere for 12–24 h, except for *C. difficile* which was cultured for 24–48 h in an anaerobic environment using GENbag anaer (bioMérieux).

2.2. Description of equimolar library pooling strategy (R3)

Genome size equivalence for each species was determined relative to the calibrator species with the smallest genome size. The genome size equivalence for a species is calculated as the genome size ratio between this species (numerator) and the calibrator species (denominator). In the present work, *S. aureus* was selected as the calibrator, and the equivalent *S. aureus* (EqSa) genome size was determined for each species (see Table 1). Equimolarity was then implemented from the EqSa genome size by defining the normalized library volume (see Table 1) to add in the pool using the following formula: EqSa genome size × 5 μL (i.e., calibrator species library volume). Once both parameters (genome size equivalence and normalized library volume) had been set for a given species menu, the composition (i.e., the number of samples) remained the only variable element to be determined for each sequencing run according to the combination of species. In the present study (see Table 2), the run composition was determined with EqSa genome sizes of each species and according to a maximum of 30 EqSa samples per run (see subsection 2.4).

2.3. Whole-genome sequencing

gDNA was extracted from fresh culture using the UltraClean® Microbial DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA). For gDNA quality control, quantitation was performed by the fluorometric method using the Qubit® dsDNA BR Assay kit (Life Technologies, Carlsbad, CA), and purity was assessed by UV absorbance using a DS-11FX spectrophotometer (DeNovix, Wilmington, DE). Library preparation was performed from 1 ng of gDNA extract using Nextera® XT DNA library preparation kit (Illumina, San Diego, CA) and Nextera® XT index kit (Illumina). Library profiles were defined on 2100 Bioanalyzer System (Agilent Technologies, Santa Clara, CA) using a High Sensitivity DNA kit (Agilent). Library normalization was performed following the manufacturer's instruction. Normalized-library pooling was performed according to the specified strategy (traditional R0 or alternative R3). Sequencing run was performed using MiSeq® Reagent kit v3 (600-cycle) on a MiSeq® instrument (Illumina) to generate 2 × 200 base paired-end reads. Raw sequence data of each sample from the current study have been deposited in the Sequence Read Archive under BioProject ID PRJNA490820.

2.4. Sequencing run specificities

The read length was systematically limited at 200 bases (400 cycles) to reduce the total run time by approximately 15 h (40 h for 400 cycles vs. 55 h

Table 1
Description of the equimolar genomic library pooling strategy (R3) used in the study for multi-species multiplexing.

Species	Genomes [a]	Genome size (Mbp)			Multiplexing strategy R3	
		Minimal	Median	Maximal	EqSa genome size [b]	Library volume for pooling [c]
Staphylococcus aureus	5251	2.6	2.9	3.1	NA	5.0 μL
Enterococcus faecium	323	2.4	2.9	3.4	1.2	6.0 μL
Enterococcus faecalis	437	2.6	3.0	3.5	1.2	6.0 μL
Acinetobacter baumannii	1108	3.5	4.0	4.5	1.6	8.0 μL
Clostridioides difficile[d]	261	3.6	4.1	4.5	1.6	8.0 μL
Enterobacter cloacae	305	4.2	4.9	5.6	1.9	9.5 μL
Klebsiella aerogenes[e]	75	4.2	5.0	5.7	2.0	10.0 μL
Serratia marcescens	72	4.9	5.2	5.7	2.0	10.0 μL
Escherichia coli	3485	4.1	5.2	6.0	2.1	10.5 μL
Klebsiella pneumoniae	802	4.7	5.7	6.5	2.2	11.0 μL
Klebsiella oxytoca	72	5.1	6.0	6.7	2.3	11.5 μL
Pseudomonas aeruginosa	1122	5.5	6.5	7.7	2.7	13.5 μL
Burkholderia cepacia	11	5.9	8.0	8.9	3.1	15.5 μL

EqSa: equivalent *S. aureus*; **NA**: not applicable.
[a] Genome number selected to calculate the genome sizes (see Table S4 for more information).
[b] Calculated EqSa genome size (strategy R3): maximal_species/median_Saureus.
[c] Calculated normalized-library volume for pooling: eq_Saureus_genome_size × Saureus_lib_volume.
[d] Formerly known as *Clostridium difficile*.
[e] Formerly known as *Enterobacter aerogenes*.

Table 2
Composition of sequencing runs performed with the equimolar (R3) or equal-volume (R0) library pooling strategies.

Runs	Species	Strain ID	EqSa genome size [a]	Strategies	
				R3 [a]	R0 [b]
				Library volume for pooling	Library volume for pooling
Multiplexing 2 species (_2sp_)	*Staphylococcus aureus*	BS87	1.0	5.0 µL	5.0 µL
	Staphylococcus aureus	BS297	1.0	5.0 µL	5.0 µL
	Staphylococcus aureus	BS358	1.0	5.0 µL	5.0 µL
	Staphylococcus aureus	BS359	1.0	5.0 µL	5.0 µL
	Staphylococcus aureus	BS361	1.0	5.0 µL	5.0 µL
	Staphylococcus aureus	BS362	1.0	5.0 µL	5.0 µL
	Staphylococcus aureus	BS363	1.0	5.0 µL	5.0 µL
	Staphylococcus aureus	BS364	1.0	5.0 µL	5.0 µL
	Pseudomonas aeruginosa	BS340	2.7	13.5 µL	5.0 µL
	Pseudomonas aeruginosa	BS341	2.7	13.5 µL	5.0 µL
	Pseudomonas aeruginosa	BS342	2.7	13.5 µL	5.0 µL
	Pseudomonas aeruginosa	BS343	2.7	13.5 µL	5.0 µL
	Pseudomonas aeruginosa	BS344	2.7	13.5 µL	5.0 µL
	Pseudomonas aeruginosa	BS345	2.7	13.5 µL	5.0 µL
	Pseudomonas aeruginosa	BS693	2.7	13.5 µL	5.0 µL
	Pseudomonas aeruginosa	BS694	2.7	13.5 µL	5.0 µL
	Total per run		**29.6**	**148.0 µL**	**80.0 µL**
Multiplexing 13 species (_13sp_)	*Staphylococcus aureus*	BS87	1.0	5.0 µL	5.0 µL
	Staphylococcus aureus	BS297	1.0	5.0 µL	5.0 µL
	Enterococcus faecium	BS772	1.2	6.0 µL	5.0 µL
	Enterococcus faecium	BS773	1.2	6.0 µL	5.0 µL
	Enterococcus faecalis	BS352	1.2	6.0 µL	5.0 µL
	Acinetobacter baumannii	BS350	1.6	8.0 µL	5.0 µL
	Clostridioides difficile[c]	BS355	1.6	8.0 µL	5.0 µL
	Enterobacter cloacae	BS349	1.9	9.5 µL	5.0 µL
	Klebsiella aerogenes[d]	BS348	2.0	10.0 µL	5.0 µL
	Serratia marcescens	BS771	2.0	10.0 µL	5.0 µL
	Escherichia coli	BS346	2.1	10.5 µL	5.0 µL
	Klebsiella pneumoniae	BS334	2.2	11.0 µL	5.0 µL
	Klebsiella oxytoca	BS338	2.3	11.5 µL	5.0 µL
	Pseudomonas aeruginosa	BS339	2.7	13.5 µL	5.0 µL
	Pseudomonas aeruginosa	BS693	2.7	13.5 µL	5.0 µL
	Burkholderia cepacia	BS351	3.1	15.5 µL	5.0 µL
	Total per run		**29.8**	**149.0 µL**	**80.0 µL**

The equivalent *S. aureus* (EqSa) genome size for each species (see Table 1) was used to define the number of samples to include in the run so as to be the closest to 30 EqSa.
EqSa: equivalent *S. aureus*.
[a] Equimolar library pooling strategy (R3).
[b] Equal-volume library pooling strategy (R0).
[c] Formerly known as *Clostridium difficile*.
[d] Formerly known as *Enterobacter aerogenes*.

for 600). This reduction was advantageous for routine use considering result delivery time and day shift. Metrics (Additional file 2: Table S2) from all MiSeq 2 × 200 bp runs were within the manufacturer's specifications.

According to Illumina specifications for MiSeq sequencing with Reagent kit v3 (i.e., 50 million paired-end reads passing filter for the maximal cluster density of 1400 k/mm^2), a single run of 2 × 200 base paired-end reads theoretically generates 5 billion bases (5 Gb) per DNA strand. To obtain a minimal depth per sample of 45× (i.e., required for our bioinformatics pipeline) and using *S. aureus* minimal (2.6 Mbp), median (2.9 Mbp), or maximal (3.1 Mbp) genome size, the theoretical maximum number of *S. aureus* samples per run was $n = 42$, $n = 38$, or $n = 35$, respectively. In this study, considering that it was suitable to add a safety margin because cluster density is not systematically maximal (i.e., typically 900–1200 k/mm^2), the maximum number of samples was set at 30 *S. aureus* (or equivalent) in a single MiSeq run.

2.5. Sequencing run composition

2.5.1. Runs multiplexing 2 species

Two runs multiplexing strains of *S. aureus* and *P. aeruginosa* were performed using either equal-volume R0 (2sp_R0) or equimolar R3 (2sp_R3) strategy. The same normalized libraries, corresponding to 8 different strains per species and a total of 29.6 EqSa samples (Table 2), were sequenced in each run using 2 different MiSeq sequencers.

2.5.2. Runs multiplexing 13 species

Three runs multiplexing strains belonging to 13 bacterial species were performed. The same gDNA extracts were used for all library preparations (i.e., 3 sets independently prepared). All library preparations and sequencing runs were performed on different days. To respect a maximum of 30 EqSa samples per run, 16 samples were sequenced, corresponding to 1 strain per species except for *S. aureus*, *E. faecium*, and *P. aeruginosa* (2 strains); each run included a total of 29.8 EqSa samples (Table 2). One run (13sp_R0) was performed using the traditional pooling strategy. To evaluate reproducibility, the run using the alternative R3 was done twice (13sp_R3a and 13sp_R3b) on the same MiSeq sequencer (i.e., only flow cells and reagents were different) with libraries prepared and normalized separately.

2.6. Genome size determination for the bacterial species

For each bacterial species the minimal, median and maximal genome sizes were determined using publicly available genome sequences from the National Center for Biotechnology Information (NCBI, version of September 21, 2015). Genomes with size <Q1−[3×(Q3−Q1)] or >Q3+[3×(Q3−Q1)] were excluded (Q1 and Q3 are the first and third quartiles, respectively). The number of genomes (i.e., total, excluded, and selected) for each species was described in Additional file 3: Table S3.

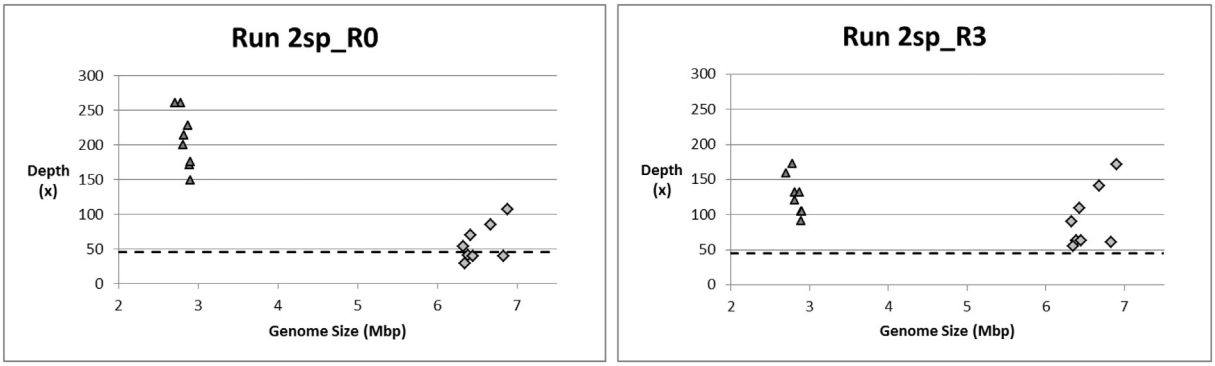

Fig. 1. Genome size and coverage depth of samples sequenced in 2-species runs. The same *S. aureus* (▲) and *P. aeruginosa* (♦) normalized libraries were pooled according to either equal-volume (Run 2sp_R0) or equimolar (Run 2sp_R3) strategy. Dotted line indicates the minimum depth (45×) required. Genome size and coverage depth values are detailed in Table S1.

2.7. Data analysis

De novo assembly (reference-free approach) was performed using the A5-MiSeq open-source pipeline (Coil et al. 2015) especially to determine coverage depth and genome size of each sequence. Read cleaning consisted in removing the sequencing adapters using Trimmomatic (Bolger et al. 2014). Then, the reads were filtered and trimmed further according to quality and length criteria (Additional file 4: Table S4) using Trimmomatic and the preprocess function of String Graph Assembler (SGA) (Simpson and Durbin 2012). Finally, SGA was used to correct errors in the reads by a k-mer frequency-based method. After being quality filtered and error corrected, the reads were assembled by the IDBA-UD500 with k-mer lengths from 35 to read length minus 1 (Peng et al. 2013). The reads were then mapped against the assembly using BWA-MEM (Li 2014) in order to polish the contigs at every position where base-calls differed between the mapping and the assembly.

3. Results

A first experiment was performed to assess the impact of the alternative equimolar library pooling strategy (hereafter referred to as R3) on coverage depth when the difference in genome size is more than 2-fold. Two sequencing runs multiplexing strains from 2 species were performed using either traditional R0 or alternative R3 strategy. De novo assemblies were generated from raw data of both runs to determine coverage depth and genome size of each sample (Fig. 1 and Additional file 5: Table S5). Using R0, 2 clusters corresponding to both bacterial species were clearly distinct. Depths were between 150× and 260× for *S. aureus* samples but lower than 110× for all *P. aeruginosa* samples. Among *P. aeruginosa* samples, 4 (BS340-BS343) even generated a depth (28×–40×) lower than the minimum 45× required, for which a retest is warranted. For R3, more homogeneous coverage depths were observed between samples of both species, with approximately 90×–170× for *S. aureus* and 55×–170× for *P. aeruginosa*. Depth was

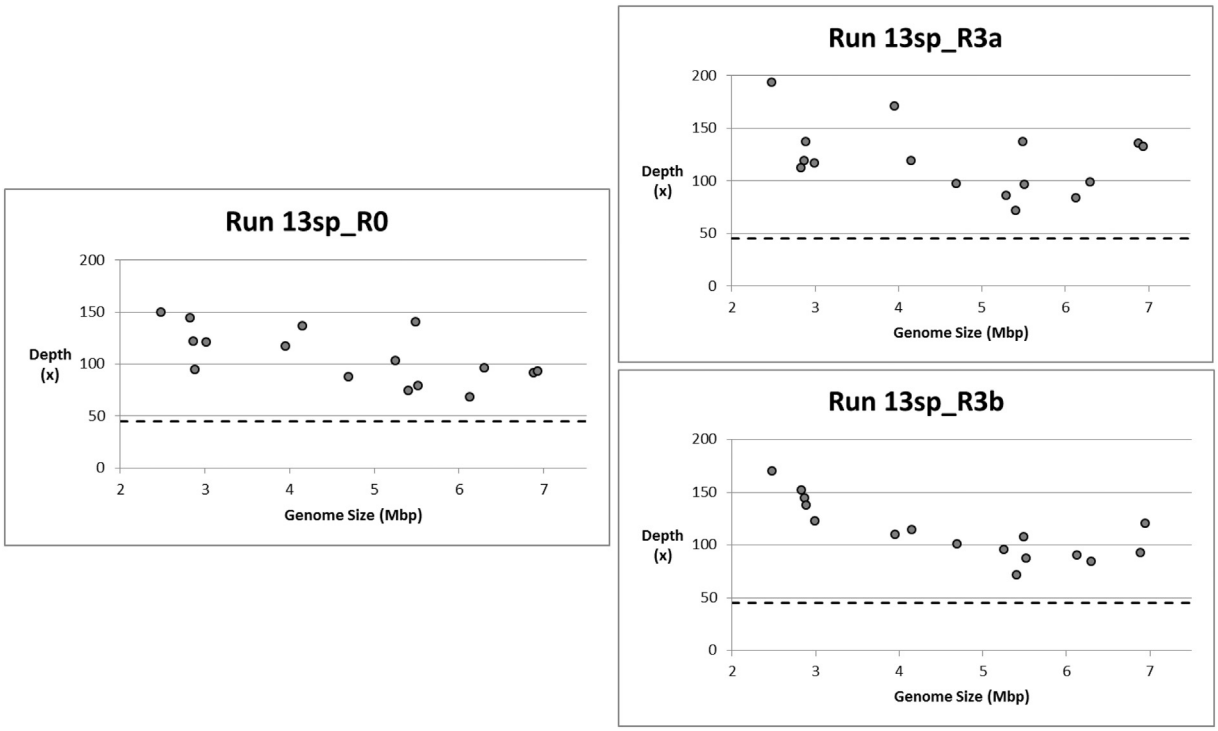

Fig. 2. Genome size and coverage depth of samples sequenced in 13-species runs. Three normalized library sets, independently prepared on different days, were pooled according to either equal-volume R0 (Run 13sp_R0) or equimolar R3 (Runs 13sp_R3a and 13sp_R3b) strategy. The dotted line indicates the minimum depth (45×) required. Genome size and coverage depth values were detailed in Table S1.

systematically higher than 45×, especially for the 4 samples (BS340-BS343) with the lowest depths between 55× and 63×. So upon using R3, not a single sample needed retesting, which makes it a more attractive option than R0.

Multiplexing strains belonging to 13 bacterial species was performed using both the R0 and R3 strategies. Cluster density was slightly higher for run R3a (1263 k/mm^2) than for run R3b (1007 k/mm^2). De novo assembly was performed from raw data of all 3 runs, and coverage depth and genome size of each sample were determined as before (Fig. 2 and Additional file 5: Table S5). Depths obtained using R0 were between 68× (BS338) and 150× (BS772). For both R3 runs, depths ranged from 71× (BS771) to either 193× or 170× (BS772, in R3a and R3b, respectively). Variability between the 2 R3 runs is consistent with that observed in our hands for independently and manually prepared libraries, with samples having similar coverage depths (see Table S1: BS297, BS771, BS355) and others having significant different depths (see Table S1: BS693, BS350). However, depth of all sequenced samples was higher than the required minimum (i.e., 45×, see section 2.4) for both strategies.

4. Discussion

For multiplexing in a single sequencing run, NGS platform manufacturers classically recommend an equal-volume library pooling strategy. Even if this strategy is well suited for multiplexing bacterial isolates from the same species, equimolar library pooling is more adapted to multispecies multiplexing. This makes it highly attractive for genomic investigations involving several pathogenic bacteria, such as those conducted to track the spread of multidrug-resistant organisms in healthcare settings (Kanamori et al. 2017): this is why we compared both strategies using HAI-related bacterial species.

In this study, the required minimum sequencing depth had been set to 45× to ensure reliable downstream analyses by our bioinformatic pipeline. To obtain such depth, the maximal sample number has been set at 30 EqSa per MiSeq run. However, for other bioinformatic pipelines requiring different depth values, the maximal sample number could vary inversely to the minimum depth.

To reduce the insufficient depth-based retesting risk in a multispecies NGS run, we used the equimolar strategy (R3). R3 is most effective when the interspecies genome size difference is large. As an example, R3 multiplexing of 8 *S. aureus* and 8 *P. aeruginosa* generated a depth that was always higher than 45×, while for R0, 25% of the samples had an insufficient depth. When 13 species were multiplexed, no retesting was needed whatever the library pooling strategy used. For such runs, the size of the genomes ranged between 3 and 7 Mbp including intermediate sizes (i.e., 4–5 and 5–6 Mbp), and the sample distribution within the different-sized segments was relatively homogeneous (e.g., 3 to 5 samples per size segment). Consequently, the relative molar differences were smaller in the R0 run with 13 species than with 2 species, which could explain the absence for the need to retest. A complementary experiment (data not shown), performed with strains belonging to 6 species (*S. aureus*, *E. faecalis*, *C. difficile*, *K. aerogenes*, *K. oxytoca*, and *B. cepacia*) and a strict equimolar strategy, tends to confirm this since it showed a more homogeneous coverage depth (65×–155×) than R0 (50×–220×). In addition, considering all R3 runs (i.e., with 13 or 2 species), the alternative strategy has no negative impact at all, so it could be systematically used for library pooling of any multispecies runs.

R3 reduces but does not fully eliminate the retesting risk, notably because interspecies genome size is not the only factor involved: other experimental factors such as library preparation and/or cluster generation efficiency come into play. Depth differences we observed for some genomes between 13-species R3 runs are a good example of variability linked to library preparation. Two independently prepared sets of libraries from the same gDNA samples were sequenced on the same MiSeq sequencer. Even if manual operations are probably a crucial parameter to consider, the Nextera XT library prep kit using an engineered transposase for in-solution tagmentation could also generate significant variability. Note that Illumina recently commercialized a new version of library prep kit (e.g., Nextera DNA Flex), which uses on-bead tagmentation to minimize bias and should provide more robust and reproducible sequencing data.

R3 could potentially be used with other NGS platforms or technologies. As a recently described example, PacBio long-read NGS libraries pooled at equimolar representation resulted in relatively even depth (137–224×) of all genomes belonging to 8 bacterial species with genome sizes ranging from 2.7 to 5.7 Mbp (Heiner et al. 2018).

R3 might as well be used for multiplexing other species of bacteria with large genome size variability. To calculate genome size equivalence factors, a value (e.g., minimum, median, and/or maximum) must be chosen to summarize genome size distribution. This choice should be made according to the tradeoff desired between benefit (i.e., reduce retesting risk) and cost (i.e., maximize sample number per run).

5. Conclusions

Using a menu comprising 13 clinically relevant bacterial species, equal-volume and equimolar library pooling strategies were compared for multiplexing samples in a single NGS run. Because of bacterial genome size variability, the risk for retesting due to insufficient coverage depth can be significant when equal-volume strategy is used. Conversely, use of equimolar strategy reduces this risk and controls the associated costs, particularly when interspecies genome size difference is more than 2-fold. Whatever NGS library preparation method, platform, and technology, the equimolar strategy should lead to more cost-effective WGS applications in clinical microbiology.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.diagmicrobio.2019.04.014.

Acknowledgments

We gratefully acknowledge Alex van Belkum and Bruno Lacroix for reviewing the manuscript.

Declarations

Funding

This work was supported by the bioMérieux company (Marcy l'Etoile, France).

Competing interests

The authors declare being employees of bioMérieux S.A.

Availability of data and material

The datasets generated and analyzed during the current study are available in the NCBI Sequence Read Archive repository (http://www.ncbi.nlm.nih.gov/bioproject/490820).

Authors' contributions

BHM, PM, ESA, FV, and GK contributed to the conception and the design of the work. BHM, PM, ESA, and GK contributed to the acquisition, the analysis, and the interpretation of the data. BHM wrote the article. All authors read and approved the final manuscript.

References

Ameur A, Kloosterman WP, Hestand MS. Single-molecule sequencing: towards clinical applications. Trends Biotechnol 2019;37(1):72–85. https://doi.org/10.1016/j.tibtech.2018.07.013.

Bennett GM, Moran NA. Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect. Genome Biol Evol 2013;5(9):1675–88. https://doi.org/10.1093/gbe/evt118.

BIOMÉRIEUX EPISEQ© CS for clinical pathogens outbreak monitoring. 2019. Available from: https://www.biomerieux-episeq.com. [Accessed 06 March 2019].

Bobay LM, Ochman H. The evolution of bacterial genome architecture. Front Genet 2017;8(72). https://doi.org/10.3389/fgene.2017.00072.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 2015;31(4):587–9. https://doi.org/10.1093/bioinformatics/btu661.

Coll F, Harrison EM, Toleman MS, Reuter S, Raven KE, Blane B, et al. Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community. Sci Transl Med 2017;9(413). https://doi.org/10.1126/scitranslmed.aak9745.

Durand G, Javerliat F, Bes M, Veyrieras J-B, Guigon G, Mugnier N, et al. Routine whole-genome sequencing for outbreak investigations of Staphylococcus aureus in a National Reference Center. Front Microbiol 2018;9(511). https://doi.org/10.3389/fmicb.2018.00511.

Han K, Li ZF, Peng R, Zhu LP, Zhou T, Wang LG, et al. Extraordinary expansion of a Sorangium cellulosum genome from an alkaline milieu. Sci Rep 2013;3(2101). https://doi.org/10.1038/srep02101.

Heather JM, Chain B. The sequence of sequencers: the history of sequencing DNA. Genomics 2016;107(1):1–8. https://doi.org/10.1016/j.ygeno.2015.11.003.

Heiner C, Kim M, Ferrao H, Wallace VJ, Lambert C, Fedak R, et al. Single chromosomal genome assemblies on the sequel system with circulomics high molecular weight DNA extraction for microbes. Poster MBP LB13. ASM Microbe 2018. Atlanta (GA, USA).

Kalinowski J, Ahrens B, Al-Dilaimi A, Winkler A, Wibberg D, Schleenbecker U, et al. Isolation and whole genome analysis of endospore-forming bacteria from heroin. Forensic Sci Int Genet 2018;32:1–6. https://doi.org/10.1016/j.fsigen.2017.10.001.

Kanamori H, Parobek CM, Juliano JJ, van Duin D, Cairns BA, Weber DJ, et al. A prolonged outbreak of KPC-3-producing Enterobacter cloacae and Klebsiella pneumoniae driven by multiple mechanisms of resistance transmission at a large academic burn center. Antimicrob Agents Chemother 2017;61(2), e01516. https://doi.org/10.1128/AAC.01516-16.

Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, et al. Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics 2015;15(2):141–61. https://doi.org/10.1007/s10142-015-0433-4.

Li H. Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics 2014;30(20):2843–51. https://doi.org/10.1093/bioinformatics/btu356.

Mirande C, Bizine I, Giannetti A, Picot N, van Belkum A. Epidemiological aspects of healthcare-associated infections and microbial genomics. Eur J Clin Microbiol Infect Dis 2018;37:823–31. https://doi.org/10.1007/s10096-017-3170-x.

Parkhill J, Wren BW. Bacterial epidemiology and biology—lessons from genome sequencing. Genome Biol 2011;12(10):230. https://doi.org/10.1186/gb-2011-12-10-230.

Peacock S. Health care: bring microbial sequencing to hospitals. Nature 2014;509(7502):557–9. https://doi.org/10.1038/509557a.

Peng Y, Leung HC, Yiu SM, Lv MJ, Zhu XG, Chin FY. IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. Bioinformatics 2013;29(13):i326–34. https://doi.org/10.1093/bioinformatics/btt219.

Reuter S, Ellington MJ, Cartwright EJ, Koser CU, Torok ME, Gouliouris T, et al. Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology. JAMA Intern Med 2013;173(15):1397–404. https://doi.org/10.1001/jamainternmed.2013.7734.

Rhoads A, Au KF. PacBio sequencing and its applications. Genomics Proteomics Bioinformatics 2015;13(5):278–89. https://doi.org/10.1016/j.gpb.2015.08.002.

Rouard C, Garnier F, Leraut J, Lepainteur M, Rahajamananav L, Languepin J, et al. Emergence and within-host genetic evolution of methicillin-resistant Staphylococcus aureus resistant to linezolid in a cystic fibrosis patient. Antimicrob Agents Chemother 2018;62(12). https://doi.org/10.1128/AAC.00720-18.

Rouard C, Bourgeois-Nicolaos N, Rahajamanana L, Romain O, Pouga L, Derouin V, et al. Evaluation of an "all-in-one" seven-day WGS solution in the investigation of a Staphylococcus aureus outbreak in a neonatal intensive care unit. J Hosp Infect 2019. https://doi.org/10.1016/j.jhin.2019.01.029.

Salipante SJ, SenGupta DJ, Cummings LA, Land TA, Hoogestraat DR, Cookson BT. Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. J Clin Microbiol 2015;53(4):1072–9. https://doi.org/10.1128/JCM.03385-14.

Simpson JT, Durbin R. Efficient de novo assembly of large genomes using compressed data structures. Genome Res 2012;22(3):549–56. https://doi.org/10.1101/gr.126953.111.

Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Group NCSP, Henderson DK, et al. Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. Sci Transl Med 2012;4(148):148ra16. https://doi.org/10.1126/scitranslmed.3004129.

Snitkin ES, Won S, Pirani A, Lapp Z, Weinstein RA, Lolans K, et al. Integrated genomic and interfacility patient-transfer data reveal the transmission pathways of multidrug-resistant Klebsiella pneumoniae in a regional outbreak. Sci Transl Med 2017;9(417). https://doi.org/10.1126/scitranslmed.aan0093.

Timme RE, Rand H, Shumway M, Trees EK, Simmons M, Agarwala R, et al. Benchmark datasets for phylogenomic pipeline validation, applications for foodborne pathogen surveillance. PeerJ 2017;5, e3893. https://doi.org/10.7717/peerj.3893.

Tyler AD, Mataseje L, Urfano CJ, Schmidt L, Antonation KS, Mulvey MR, et al. Evaluation of Oxford Nanopore's MinION sequencing device for microbial whole genome sequencing applications. Sci Rep 2018;8(1), 10931. https://doi.org/10.1038/s41598-018-29334-5.

Zhu Y, Czauderna T, Zhao J, Klapperstueck M, Maifiah MHM, Han ML, et al. Genome-scale metabolic modeling of responses to polymyxins in *Pseudomonas aeruginosa*. Gigascience 2018;7(4):giy021. https://doi.org/10.1093/gigascience/giy021.