Inter-EURLs Working Group on NGS (NEXT GENERATION SEQUENCING)

Foreword

The working group (WG) has been established by the European Commission with the aim to promote the use of next generation sequencing (NGS) and in particular whole genome sequencing (WGS) across the networks of the European Union Reference Laboratories (EURLs), build WGS capacity within the European Union (EU) and ensure liaison between the EURLs, European Food Safety Authority (EFSA) and European European Centre for Disease Prevention and Control (ECDC) activities concerning the WGS mandate sent by the Commission. The WG includes all the EURLs operating in the field of the microbiological contamination of food and feed. The present document represents a deliverable of the WG and is meant to be dispatched to the respective networks of the National Reference Laboratories (NRLs).

Guidance document for WGS-Benchmarking

Maroua SAYEB, EURL *Listeria monocytogenes*

Anses Laboratory for food safety, Maisons-Alfort, France

Date: 08 March 2021

Version 01

Table of contents

1 Introduction

Whole genome sequencing (WGS) is increasingly used in diagnostics and surveillance of pathogenic microorganisms. Common applications for WGS is to use the data for the identification of pathogenic bacteria, antimicrobial resistance (AMR), virulence gene detection and during outbreak investigations. This guidance document presents how to benchmark the different WGS steps. A checklist was drafted for bacterial genome analysis using WGS technology that sets standards for the analytical wet lab (bench) process and bioinformatics analyses, also called ”dry lab (bench)”. In other words, bacterial WGS includes 2 steps:

(1) Analytical wet lab analyses;

(2) Bioinformatics analyses of reads obtained from sequencing.

The wet bench component generally includes any or all of the following steps: handling of isolates, DNA extraction, DNA fragmentation, barcoding (molecular indexing) of DNA fragments, enrichment of gene targets within certain panels, adapter ligation, amplification, library preparation, flow cell loading, and generation of sequence reads. Generation of reads is almost entirely automated and the output consists of millions to billions of short sequences (reads). The following dry bench workflow consists in intensive computational and bioinformatics analyses that use a variety of algorithms that, for instance, (i) map and align the short reads to a linear reference bacterial genome and/or (ii) perform *de novo* assembly.

The EURLs WG on NGS approached the analytic wet bench process and the bioinformatics analyses (‘‘dry bench’’) as two discrete processes requiring separate considerations for standards.

The EURLs WG have shared a survey to their NRLs network to ask them for different parameters used in the wet and dry lab part of WGS in their laboratories. These parameters are used in this guidance to present the different steps of WGS.

2 Wet-Bench

A detailed documentation of the wet bench process is a critical part to ensure good quality in WGS. All standard operating protocols of DNA preparation, fragmentation, library preparation, barcoding (molecular indexing), sample pooling, and generation of reads must be documented in order to trace back each step and subsequent manipulations. This includes documentation of all methods and reagents, as well as instruments and instrument software version used throughout the wet bench process. In addition, controls are required and need to be described. Microbiological laboratories that analyse different species of pathogenic bacteria (*e.g. Bacillus*, *Campylobacter*, *E. coli*, *Listeria monocytogenes*, *Salmonella*, *Staphylococcus, …*) should have adapted standard operating procedures (SOPs) for DNA extraction for each bacterial species. The reagents and protocols used for pooled analysis of isolates must be detailed and should include the sequence information of the barcodes used for each sample. Metrics and quality control parameters used to assess run performance should also be documented. Commonly used metrics include the fraction of bases over a specific quality, average of reads mapped along the targeted region (*i.e.* depth of coverage (X)) and proportion of the targeted region presenting at least 1X of mapped reads (*i.e.* breadth of coverage (%)). The laboratory must establish and document acceptance and rejection criteria for the wet-bench process from sample preparation to sequencing.

A panel of sequenced (reference) strains are important for comparing the quality assurance (QA) and quality control (QC) for wet-lab steps. QC should be performed during the preanalytical steps (DNA isolation and library preparation), analytical steps (quality metrics of the sequencing run), and postanalytical steps (data analysis) of WGS. Five QC checkpoints can be defined throughout the WGS process: DNA template QC, library QC, sequencing run QC, raw data QC and data analysis QC.

2.1 Input DNA, DNA library quality and quantity metrics

2.1.1 Quality of the input DNA

Quality of the input DNA for all compared isolates should be assessed by the ratio between the absorbance at 260 nm and at 280 nm.

The value of the 260/280 absorbance ratio must be >1.7 (Lucena-Aguilar, Sanchez-Lopez *et al.* 2016). Alternatively, at least agarose gel or capillary electrophoresis should be performed to check pureness of the extracted DNA.

2.1.2 Quantity of the input DNA

Quantity of the input DNA for all compared isolates should be measured using a fluorometric system (e,g, the Qubit fluorometer, the Quantus fluorometer) and the corresponding reagents kit.

DNA concentration should fulfill the requirements of the library preparation kit in use (generally ≥1ng/µl; Lucena-Aguilar *et al.*, 2016)

If the abovementioned quality and quantity parameters are not fulfilled, the DNA isolation step should be repeated and troubleshooted.

2.2 DNA library quality and quantity metrics

2.2.1 DNA library size distribution

DNA library size distribution for included isolates is measured using a capillary electrophoresis fluorimetric instrument (for example the BioAnalyzer instrument or Tape station and the corresponding reagents kits). The included isolates should preferably have the following variety (if present within the run): different species, Gram-positive and Gram-negative bacteria, and species with different GC content.

The average size of the library should be compliant with what is requested for the planned sequencing run (e.g within the range of 300bp-3kb).

2.2.2 DNA library concentration

DNA library concentration should be measured using a fluorometric system (*e.g.* the Qubit fluorometer, the Quantus fluorometer) and the corresponding reagents kit.

DNA library concentration should be ≥ 1nM for Illumina sequencers (Hussing, Kampmann *et al.* 2018).

If the abovementioned quality parameters are not fulfilled, repeat and troubleshoot the fragmentation/amplification/post-PCR cleanup steps of library preparation (Smits, 2019). In some instances, a BioAnalyzer run failure leads to missing or shifted library peaks, in that case repeat the BioAnalyzer run. Correct localization of the peaks of the DNA ladder and the expected peaks of lower and upper marker in the run of each sample should be checked to identify possible shift in the molecular weight of the library.

2.3 Quality metrics of the sequencing reads

Quality metrics of included strains should preferably (?) meet the following parameters [preliminary quality thresholds]:

- Percent of bases with quality score >Q30 for the run must be ≥ 50%;
- Q30 score for generated genome sequences must be ≥75% for at least 80bp of the read length;
- Average depth of coverage of ≥ 30X across the whole genome is recommended but is depending with the software used.

If above-mentioned quality parameters are not fulfilled, repeat and troubleshoot the library preparation and/or library pooling/loading (Arnold, Edwards *et al.* 2018).

2.3.1 Negative controls to test for contamination

Negative control of sequencing process represents an index combination which does not correspond to any strain in the current sequencing run but matches one of the index combinations used in the previous sequencing run (Arnold, Edwards *et al.* 2018). If the negative control generates reads, this indicates a possibility of carry-over contamination with the library fragments generated in the previous run.

In case ofcontamination , and if using MiSeq instrument, wash it with 0.01% sodium hypochlorite and clean all working surfaces with 10% bleach. Similar cleaning procedures might also be applied to other sequencing platforms, under guidance of specific technical assistance.

2.4 Comparison between sequencing platforms

Platform accuracy refers to the accuracy of individual base calling in a bacterial genome. To benchmark the different sequencing platforms it is possible to compare base calling results with a reference sequence. In addition, validated quality parameters and their ranges to provide high platform accuracy could be used. The preliminary quality thresholds were adjusted based on validation data to match the most stringent values of quality parameters, which were detected during the validation (±5%). In several cases, the threshold was kept at the level that was even more stringent than any of the detected values.

2.4.1 Quality parameters affecting platform accuracy

The quality parameters of the sequencing data affecting platform accuracy were identified as well as the quality parameters thresholds, which provide ≥ 90% accuracy of base calling. The following types of errors affect the different platform sequencing: a) sequence errors introduced by DNA library preparation technique (e.g. amplification-introduced errors); and b) base calling accuracy of the sequencer. A set of quality parameters to account for corresponding types of errors was established.

Sequence errors introduced by DNA library preparation technique

The first type of sequencing errors which is introduced by PCR errors during library amplification is stochastic and independently performed library preps are not likely to have the same errors. High depth and good uniformity of coverage reduce the impact of sequencing errors. For that reason, the thresholds for depth and uniformity of coverage providing accurate base calling were determined empirically during this validation, resulting in the demand that:

Average depth of coverage must be ≥ 15X across genome.

The minimum coverage of 15X was achieved for targeted areas used in gene-specific analysis: MLST scheme genes. If the minimum coverage threshold is not achieved for targeted areas, an alternate method such as Sanger sequencing should be used for sequencing a given genome region. Uniformity of coverage: >50% of positions on the target (coding sequence) should have a coverage ≥10X and >70% of positions should have coverage ≥5X.

Base calling accuracy of the sequencer

The second type of errors is determined by the accuracy of base calling of the sequencer. The parameter used to estimate the accuracy of the base calling by the platform is the Phred quality score, which reflects a probability of incorrect base calling.

Example for Illumina MiSeq sequencing platform:

For an Illumina sequencing platform, the Phred score of Q30 is generally used and it corresponds to a probability of one incorrect base calling in 1,000 (Sato, Ogura *et al.* 2019). The MiSeq sequencer specifications cited on the internet site of the manufacturer suggest the following base calling accuracy, measured by the Phred quality score (Q score): > 70% bases in 300bp-long fragments should have Phred score higher than Q30, while it is noted that "actual performance parameters may vary based on sample type, sample quality and clusters passing filter".

The EURL-*Lm* evaluated the following quality metrics, to account for base calling accuracy from the sequencer:

- Accuracy of base calling: Sequencing reads must have ≥ Q30 for more than 75% bases for at least 86bp of the read length. The average read length after trimming and discarding the base pairs with quality score <Q30 should be >109bp.

As an example from the EURL-Lm experience, the sequencer run metrics were assessed to establish the optimal performance of the MiSeq platform:

- Percent of bases with the quality score >Q30 for the run sequencer base calling accuracy metrics should be > 57% for the 600 cycles MiSeq reagents.
- Cluster density for the run- density of clusters formed by clonally amplified library fragments on the flow cell surface should be >800 K/mm2, Maximum 1700K/mm2, to preferably obtain cluster density within the range of 800-1100 K/mm2.
- Cluster passing filter of the run – percentage of clusters that pass quality filter for the purity of the signal should be >72%.

2.4.2 Accuracy of base calling against a reference genome

The accuracy of the platform can be assessed by determining the closeness of agreement between base calling made by the platform sequencer (measured value) and a NCBI reference sequence (the ‘true value’). Reference sequences should preferably be available for all the compared genomes of the dataset to determine the different platform accuracies by mapping generated reads to the corresponding reference sequence and identifying Single Nucleotide Polymorphisms (SNPs). Real SNPs are expected due to the possibility of mutations accumulated in the reference genomes during cultivation. This would result in pairwise SNPs differences between the reference genome and the reads generated by the platform (i.e. validation sequences). For this reason, when comparing validation sequences against a reference genome, the within- and between-run triplicate sequences of validation strains have to be taken into account.

Example for the SNP mapping for genomes of Listeria monocytogenes:

The EURL-*Lm* compared reference genomes and reads generated during 5 independent library preparations. A SNP detected between the reference genome and validation sequences were considered a sequencing error only when the SNP was detected in less than all 5 replicates. If a SNP detected between the reference genome and validation sequences was identical in all 5 validation replicates, this SNP was not considered a sequencing error, and instead was considered a possible mutation in the reference genome.

3 Dry-bench

Modern epidemiology of foodborne bacterial pathogens relies increasingly on WGS techniques. As opposed to profiling techniques such as pulsed-field gel electrophoresis, WGS requires a variety of computational methods, often called bioinformatics pipelines. Comparing thousands of genomes/sequences across an entire species requires a fast method with coarse resolution; however, capturing the fine details of highly related isolates requires computationally heavy and sophisticated algorithms. Three applications of WGS covered in this document include:

- *In silico* whole- or core-genome Multilocus Sequence Typing (wg/cgMLST);
- Detection of genes (e.g., antimicrobial resistance (AMR) genes, virulence genes);
- Genotyping using high quality SNPs (hqSNPs).

Most bacterial investigations employing WGS depend on the ability to identify an outbreak clade whose inter-genomic distances are less than an empirically determined threshold. Indeed, thresholds of pairwise SNP differences can help to distinguish between genomes associated and unassociated to an outbreak genome of interest. Matrices of distances and phylogenetic trees are the outputs to compare the pipelines. This document presents tools for comparing different pipelines to accurately identify outbreak clusters. For more information about cluster analysis, please read the Guidance document for cluster analysis of whole genome sequence data, also produced within the Inter EURL WGS WG.

Currently in the US PulseNet system, operated by US CDC, wg/cgMLST are the preferred primary methods for WGS foodborne cluster detection and outbreak investigation due to their ability to name standardized genomic profiles, the use of a central database, and their ability to be run in a graphical user interface. However, creating functional wg/cgMLST schemes require extended up-front developments and subject-matter expertise. These approaches greatly increase the discriminatory power over traditional MLST and are being adopted by PulseNet International as one of the main methodologies for food-borne bacterial typing and molecular surveillance. There are three publicly available online databases that facilitate gene-by-gene analysis for an increasing number of bacterial species and host schemes (i.e. defined set of loci to be used in MLST or wg/cgMLST): PubMLST, which among others hosts a scheme for *Campylobacter* spp; Pasteur Institute, which host a *Listeria monocytogenes* scheme ; and Enterobase, which hosts schemes for *E. coli* and*Salmonella* . Several other schemes are being developed. This kind of analysis can be also performed with pipelines that can be downloaded and run locally, such as “Genomic Profiler” or “chewBBACA”.

Whole genome single nucleotide polymorphisms (wgSNPs), which detects Single Nucleotide Polymorphisms (SNPs) on whole genome sequences and performs cluster analyses on the resulting wgSNP matrix. Phylogenetic methods exploiting nucleotide resolution variation (SNPs) between bacterial isolates can be used to elucidate then relatedness and ancestry of strains under robust evolutionary models and provide a framework to explore the genetic diversity. In the SNP-based method, single nucleotide changes were used to infer phylogenetic relatedness.

To estimate the concordance between the different bioinformatics approaches (analysis of SNPs in the whole genome and cg/wgMLST) representing the two most commoly used methods for cluster analysis during outbreak investigation. Alleles in conventional MLST and loci from wg- and cgMLST are compared for each genome against a database of known and labelled alleles. In many wg/cgMLST pipelines, a single nucleotidic difference or any insertion or deletion in the sequence of a locus results in a different allele. For other pipelines based on the comparison of the translated coding sequences corresponding to the loci in the scheme (e.g. chewBBACA), only the nucleotide differences resulting in different translated proteins would result in calling different alleles. With wg- and cgMLST, thousands of loci are compared and their distances are used to generate a phylogenomic reconstruction usually with either the unweighted-pair-groupmethod- with-arithmetic-mean (UPGMA) or neighbor-joining (NJ) algorithms. Concerning the SNP-based method, pairwise SNP differences or concatenated SNPs may be used to infer distance-based or character-based phylogenomic relatedness, respectively.

One has to determine what to consider as a single test for each specific assay performed in the laboratory. Assay accuracy could be measured only for the validation strains which present reference genomes available from public DNA databases like NCBI. The platform accuracy and assay accuracy are interconnected, but it is important to distinguish them in a WGS benchmarking. For the assay accuracy, one could focus only on areas of several housekeeping genes or the percentage of genes correctly identified within the cgMLST, while the platform accuracy to generate a correct base calling across of the genome. The high quality SNP genotyping across the genome can be used as a main assay to validate the platform accuracy, since it allows validating the accuracy of base calling throughout the genome. Even though ultimately the accuracy of a single base call made by the platform has to be evaluated, WGS assays may tolerate a certain error rate of the platform and still can yield accurate results as long as the assay was validated with a given platform. This is especially true when it is possible to reach a decent depth of coverage in a particular area of the targeted genome. Indeed, erroneous base calls can be excluded during *in silico* trimming steps. As previously indicated, the SNP analysis or wg/cgMLST should be done with at least at 30X depth of genome coverage. The

EURL-*Lm* determined optimal depth of coverage to be ≥ 60X based on accuracy of SNP detection at various simulated genome coverages when the sequences were produced on a MiSeq? The EURL-*Campylobacter* determined the optimal depth of coverage to be ≥ 70X for cgMLST analysis for *Campylobacter* using sequences generated on a MiSeq with read-length 2x75. However, 15X coverage threshold was sufficient for other WGS assays (MLST, AMR genes detection) and 15X was determined as the minimum acceptance criteria for raw data in order to be considered for the mentioned types of analyses. If presence/absence of certain genes is a key diagnostic feature, the corresponding WGS assay should be added to the validation panel. Validation of the specific assays allows determination of the threshold for the base calling accuracy of the platform.

3.1 Accuracy of WGS-derived MLST assay

MLST is a method of bacterial genotyping based on sequencing of 6-to-7 housekeeping genes. Sequence variation (or alleles) of those genes are used to assess genetic relatedness between the isolates. A combination of known alleles allows assignment of a corresponding sequence type number. There are two options for the single test definition in the case of MLST: 1) to consider the final sequence type result as a single test result; or 2) to evaluate the result of each allele identification separately and to consider each allele call as a separate test. It is reasonable to consider the detection of each of the multiple genetic determinants as a separate test, especially when sequence variations change the final results. For instance, any sequence variation in the MLST alleles will lead to the change of the allele identification number and will result in a new sequence type. The definition of the correct result for MLST corresponds to a correct identification of each of the MLST alleles in the validation sequence. As an example, fifteen validation sequences and their corresponding reference sequences for the same strains of *Listeria monocytogenes* were analysed by MLST. MLST profiles were identified from raw readsusing the CGE tool included in the Core lab analysis pipeline. Accuracy is represented by percentage of agreement among the alleles detected in validation genomes compared to reference sequences. For example, in *L. monocytogenes* analysis for all validation isolates, each of the 7 housekeeping genes used in the typing scheme were identified correctly, resulting in 100% allele identification accuracy.

3.2 Accuracy of genotyping assay

Accuracy of the genotyping assay is the ability of the assay to correctly determine genetic relatedness between the isolates. High-quality SNP genotyping is based on mapping of the reads against a reference genome, which is followed by genome-wide SNP calling against the reference genome. The identified SNPs are used to build a phylogenetic tree illustrating the genetic relatedness between tested isolates. The topology of the tree reflects genetic distances between isolates. Short branches refer to closely related isolates and long branches refer to diverging unrelated isolates. Preferably add at least 5 strains of the same species to build a tree in order to assess reproducibility. To assess accuracy of genotyping assay, phylogenetic trees should be built using reference sequences and validated sequences. Then, trees should be compared together based on topology and cluster patterns.

3.2.1 Topological similarity between reference tree and validation tree

Tree agreement can be statistically measured using R software, which provided a percentage of topological similarity between two trees. See bellow and present Reference to the study and report.

3.2.2 Comparison of clustering pattern of validation tree and reference tree

Clustering pattern upon the phylogenetic comparison of validation sequences must match the pattern generated from the clustering of reference sequences. In other words, conclusions made about the relatedness of the isolates drawn from validation tree and reference tree should be the same.

3.3 Comparaison of phylogenetic tree and metrics distance for the dry-lab part

Each of the WGS dry-lab pipelines produces an output that could be used for interpreting the relationship between genomes in various forms such as distance matrices, phylogenetic tree and dendrograms; however, they have different underlying algorithms and output formats. For example, each SNP pipeline uses different alignments software and SNP-callers producing different formats to describe their SNP datasets. Algorithms implemented in wg/cgMLST and SNP workflows are different. On the one hand, a SNP might be located in an intergenic region, yielding zero allelic differences by wg/cgMLST; on the other hand, many SNPs might be located on a single gene, yielding to the collapse of multiple SNPs into a single allelic differencein wg/cgMLST. A classic tree comparison method is the Robinson-Foulds metric, sometimes called the symmetric difference metric, where the number of internal branches that exist in one tree but not the other are counted (Robbins, Devare et al. 1981). Another tree comparison metric is the Kuhner-Felsenstein score, sometimes called “branch score” which is similar to Robinson-Foulds but calculates the Euclidean distance between each branch’s length. Both Robinson-Foulds and Kuhner-Felsenstein metrics were implemented in the Phylip package in the program treedist and in some R libraries such as ape, pegas, ade4, phytools, phangorn and dendextend. Both of these classical metrics rely on unrooted trees, and small differences between two trees can artificially magnify the distance between two trees. A more robust tree metric, the Kendall-Colijn, accounts for both tree topology and branch length (Kendall and Colijn, 2015). The Kendall-Colijn metric compares two rooted trees using Euclidean distances from tip to root with a correclation coefficient R to give more weight to either topology (R = 0) or branch length (R = 1). Phylogenetic workflows could also be compared based on their matrix of pairwise distances. For instance, the Mantel test uses a generalized regression approach to identify correlations between two distance matrices. Therefore, if the genome distances from one workflow vs. another workflow are consistently high and correlated, the Mantel test will yield a high correlation coefficient.

3.4 Detection of specific genes

To estimate the accuracy of the gene detection assay, different tools can be used. Genome annotation is the process of identifying the location and biological role of genetic features present in a DNA sequence. It is typically the first step applied after assembly of a draft genome. This process involves software pipelines that use multiple external feature prediction algorithms allowing the identification of genetic features such as protein coding sequences, transfer RNA genes (tRNAs), ribosomal RNA genes (rRNAs) and occasionally higher-order features such as CRISPR elements. An important use of draft genomes is the gene-by-gene approach, which extends the concept of classical MLST to incorporating a few discriminatory genes to a much larger number of targets comprising the core genome harboured by a given species (core genome (cg) MLST), or alternatively the whole (core and accessory) genome for a given species (whole genome (wg) MLST).

4 Validation stage

4.1 WGS repeatability and reproducibility.

Repeatability (within-run precision) should be assessed as the concordance of the assay results and quality metrics obtained for a sample tested multiple times within the same sequencing run. Identical results should be obtained from same dataset, at least twice with identical:

- Computer/IT infrastructure;
- Version of the software;
- Options/parameters.

The interpretation of the results should not change: no significant difference should be observed while repeating the WGS workflow in the same laboratory, with the same operators using the same instrument.

Reproducibility (between-run precision) should be assessed as the consistency of the assay results and quality metrics for the same sample sequenced on different occasions. As an example, at EURL-*Lm* 34 validation strains were sequenced three times in the same sequencing run (for repeatability) and three times in different runs (for reproducibility). For within-run replicates, one DNA extract was used, but independent library preparations were done, with the final genomes being included in a single sequencing run. Therefore, for each sample, 3 within-run replicates and 3 between-run replicates were made, and the total number of repeated results was 5. All quality parameters (depth of coverage, uniformity of coverage, and accuracy of base calling [*Q* score], etc.) did not change significantly for within- and between-run replicates, as determined by a two-tailed *t* test. For the quality values for all sequenced strains, the reproducibility and repeatability of the WGS assay were evaluated with two methods: (i) evaluation of base calling reproducibility and repeatability per replicate and (ii) evaluation of base calling reproducibility and repeatability relative to genome size. Comparable results should be obtained from the same dataset, at least twice with different computers/IT infrastructure with identical:

- Version of the software;
- Options/parameters.

The interpretation of the results should not change: no significant difference should be obtained while reproducing the WGS workflow in different laboratories, with different operators, or different instruments. Minor differences are expected, caused by methodological steps but should not be considered significant.

EURL CPS European Union Reference Laboratory for Coagulase Positive Staphylococci http://eurl-staphylococci.anses.fr
European Union Reference Laboratory Foodborne Viruses
European Union Reference Laboratory Antimicrobial Resistance
European Union Reference Laboratory for Parasites
EURL Salmonella
EURL Lm European Union Reference Laboratory for Listeria monocytogenes http://eurl-listeria.anses.fr
EURL Campylobacter
EU-RL VTEC

4.2 WGS sensitivity and specificity

The sensitivity of WGS should be assessed as (i) analytical sensitivity (minimum coverage that allows accurate SNPs or wg/cgMLST analysis) and (ii) diagnostic sensitivity (the likelihood that a WGS assay will detect a sequence variation when it is present) (this value reflects the false-negative rate of the assay).

The specificity of WGS should be determined as (i) analytical specificity (the ability of an assay to detect only the intended target in the presence of potentially cross-reacting nucleotide sequences) and (ii) diagnostic specificity (the probability that a WGS assay will not detect any sequence variation when none is present, this value reflects the false-positive rate of the assay).

4.3 Validation of WGS workflow

In order to validate your WGS workflow with this guidance document you need to constitute a dataset for your targeted bacterium, virus or parasite, then you will be able to benchmark your workflow evaluating the performance criteria proposed in the current guidance document: stability, repeatability, and reproducibility. This guidance document contributes as a first step to harmonise methods and thus to obtain reliable results.

5 References

Arnold, C., K. Edwards, M. Desai, S. Platt, J. Green and D. Conway (2018). "Setup, Validation, and Quality Control of a Centralized Whole-Genome-Sequencing Laboratory: Lessons Learned." J Clin Microbiol **56**(8).

Endrullat, C., J. Glokler, P. Franke and M. Frohme (2016). "Standardization and quality management in next-generation sequencing." Appl Transl Genom **10**: 2-9.

Hussing, C., M. L. Kampmann, H. S. Mogensen, C. Borsting and N. Morling (2018). "Quantification of massively parallel sequencing libraries - a comparative study of eight methods." Sci Rep **8**(1): 1110.

Lucena-Aguilar, G., A. M. Sanchez-Lopez, C. Barberan-Aceituno, J. A. Carrillo-Avila, J. A. Lopez-Guerrero and R. Aguilar-Quesada (2016). "DNA Source Selection for Downstream Applications Based on DNA Quality Indicators Analysis." Biopreserv Biobank **14**(4): 264-270.

Robbins, K. C., S. G. Devare and S. A. Aaronson (1981). "Molecular cloning of integrated simian sarcoma virus: genome organization of infectious DNA clones." Proc Natl Acad Sci U S A **78**(5): 2918-2922.

Sato, M. P., Y. Ogura, K. Nakamura, R. Nishida, Y. Gotoh, M. Hayashi, J. Hisatsune, M. Sugai, I. Takehiko and T. Hayashi (2019). "Comparison of the sequencing bias of currently available library preparation kits for Illumina sequencing of bacterial genomes and metagenomes." DNA Res **26**(5): 391-398.

Smits, T. H. M. (2019). "The importance of genome sequence quality to microbial comparative genomics." BMC Genomics **20**(1): 662.