Jul 03, 2019

A step-by-step beginner's protocol for whole genome sequencing of human bacterial pathogens

Sanjay Gautam[1], Rajendra KC[1], Kelvin WC Leong[2], Micheál Mac Aogáin[3], Ronan F. O'Toole[1]

[1]School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, Australia, [2]School of Molecular Sciences, College of Science, Health and Engineering, La Trobe University, Australia, [3]Department of Clinical Microbiology, School of Medicine, Trinity College Dublin, Ireland

Working dx.doi.org/10.17504/protocols.io.4xxgxpn

Sanjay Gautam

ABSTRACT

Bacterial whole genome sequencing (WGS) is becoming a widely-used technique in research, clinical diagnostic, and public health laboratories. It enables high resolution characterization of bacterial pathogens in terms of properties that include antibiotic resistance, molecular epidemiology, and virulence. The introduction of next-generation sequencing instrumentation has made WGS attainable in terms of costs. However, the lack of a beginner's protocol for WGS still represents a barrier to its adoption in some settings. Here, we present detailed step-by-step methods for obtaining WGS data from a range of different bacteria (Gram-positive, Gram-negative, and acid-fast) using the Illumina platform. Modifications have been performed with respect to DNA extraction and library normalization to maximize the output from the laboratory consumables invested. The protocol represents a simplified and reproducible method for producing high quality sequencing data. The key advantages of this protocol include simplicity of the protocol for users with no prior genome sequencing experience and reproducibility of the protocol across a wide range of bacteria.

EXTERNAL LINK

http://www.jbmethods.org/jbm/article/view/276/247

276-3464-1-PB.pdf

GUIDELINES

Background

Using Sanger sequencing, the Human Genome Project expended approximately USD $2.7 billion and took more than 10 years to pro- duce the first human genome sequence. Today, a human genome can be sequenced in a matter of days for less than USD $1000 on a single next-generation sequencing (NGS) machine. This step change in through- put and per-base cost has transformed the use of DNA sequencing in biomedical research and is being translated in an expanding number of ways into medicine. NGS is increasingly being applied to understand- ing and managing infectious diseases. This includes the sequencing of microbial genomes for the purposes of laboratory identification of infectious agents [1], detection of antibiotic resistance markers [2], and the public health surveillance of epidemiological clusters and outbreaks [3]. Examples include its deployment in public health surveillance and control of community cases of Escherichia coli [4], Campylobacter jejuni [5], Legionella pneumophila [6] and Mycobacterium tuberculo- sis [7] disease, or global and regional epidemics caused by influenza [8], Ebola [9], and Zika [10] viruses. It has also been utilised to track the source and spread of healthcare-associated infections caused by Staphylococcus aureus [11], Pseudomonas aeruginosa [12], Acineto- bacter baumannii [13], and Enterococcus faecium [14] in order to guide infection prevention and control in hospitals.

In addition to its whole genome (WGS), whole exome (WES), transcriptome (RNA-Seq), bisulphite methylome, and metagenomic sequencing capabilities, NGS can be directed to the detection of specific genes or mutations associated with human disease through targeted-panel amplicon screening. However, barriers remain with regard to establishing NGS in a laboratory for the first time and this hinders its uptake in clinical microbiology and other settings. One of these challenges is the lack of a simplified step-by-step protocol that can be picked up by laboratory personnel with no prior training or experience in NGS and used to gen- erate reliable, high quality sequence data. Illumina dye-sequencing is currently considered the gold standard internationally in terms of read depth and base-calling accuracy, genome coverage, scalability, and the range of sequencing applications it delivers.

In this work, we produced an easy-to-follow, step-by-step NGS protocol with consistent genome coverage and average read depth that

was applicable to a range of bacterial pathogens i.e., Gram-positive van- comycin-resistant Enterococcus faecium, Gram-negative non-typeable Haemophilus influenzae, and acid-fast high-GC content Mycobacterium tuberculosis. This protocol can be used to generate Illumina-based WGS data for clinical isolates of bacterial pathogens of importance to human health.

Figure 1 is the graphical summary of the process of obtaining whole genome sequence data from bacterial culture. This wet laboratory procedure generated FastQ reads from the sequencer within three days of start. We modified a number of the DNA extraction steps to obtain a sufficient quantity of contamination free template. Similarly, we replaced library normalization plates and Nextera XT tagment amplicon (NTA) plates with conventional polymerase chain reaction (PCR) tubes which may represent a cost-effective alternative. In addition, we have recommended the use of equal DNA concentrations of each library during library normalization to ensure better coverage and minimize bias. Simplification of bacterial NGS may assist in its uptake by beginner users.

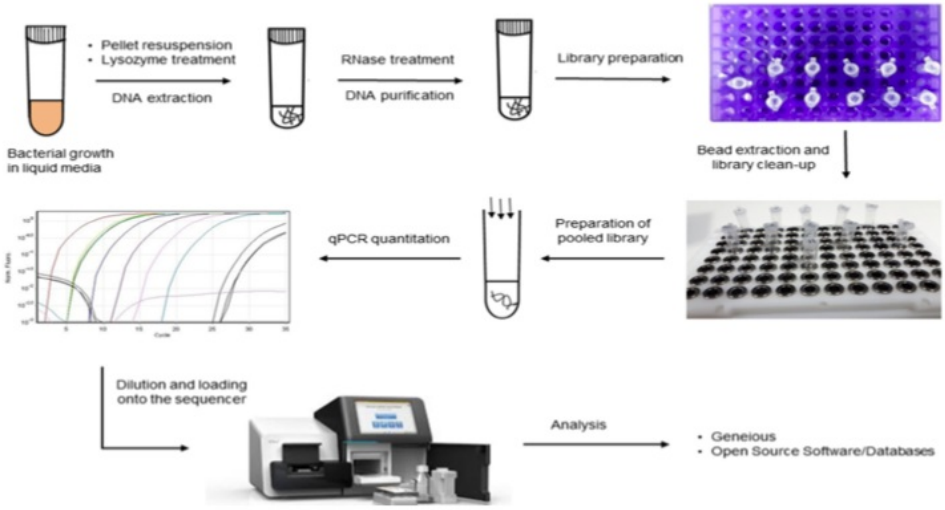

Figure 1. Graphical summary of the process of obtaining whole genome sequence data from a bacterial culture.

Anticipated Results

Table 1. Percentage sequence coverage and mean read depth for each of the sequenced genomes with respect to reference strains.

Sample	Reference sequence coverage (%)	Mean read depth
VRE1	90.98	197.4
VRE2	89.4	178.4
VRE3	90.96	168
NtHi1	93.8	263.5
NtHi2	86.5	61.3
NtHi3	89.1	124.9
MTBC1	96.4	104.4
MTBC2	96.8	100.4
MTBC3	96.8	68.4

Coverage refers to the percentage of reference genome bases covered by mapped sequence reads. Mean read depth indicates the mean number of times each base is mapped by a sequence read. Reference genomes used were *E. faecium* ST18 DO (TX16) (accession number NC_017960), *Haemophilus influenzae* 86-028NP (nontypeable) (accession number NC_007146), and *Mycobacterium tuberculosis* H37Rv (accession number NC000962). VRE, vancomycin resistant *Enterococcus faecium*; NtHi, non-typeable Haemophilus influenza; MTBC, *Mycobacterium tuberculosis* complex.

Anticipated Results

A consensus sequence was generated for each of the isolates analysed in Geneious. The Geneious report provided information on the percentage coverage of test sequence to the reference genome and the mean read depth (**Table 1**). Each contiguous sequence is viewable in Geneious and can be analysed for coverage with respect to the reference genome. Quality control checks of raw sequence data were also performed using FastQC [22]. This freely-available software provided information re- garding per base sequence content and quality, per base and sequence GC content, and highlighted the parameters of the sequence quality.

Initial typing analysis

We used open source databases to analyze the sequence data. For example, Geneious mapped contiguous sequences were imported into PubMLST (https://pubmlst.org/) for sequence typing of *Haemophilus influenzae* and vancomycin-resistant *Enterococcus faecium*. This can also be achieved using raw fastq reads in the MLST profiling tool from the Center for Genomic Epidemiology (CGE) database (http://www. genomicepidemiology.org/). The Resfinder tool (https://cge.cbs.dtu. dk/services/ResFinder/) was used to identify acquired antimicrobial resistance genes from raw fastq files. For example, PubMLST typing classified NTHi 1 as sequence type 46 and Resfinder did not detect the presence of any antimicrobial resistance determining mutations.

Mycobacterium tuberculosis complex raw fastq.gz files were uploaded to the TGS-TB database (https://gph.niid.go.jp/tgs-tb/) to predict drug susceptibility, in *silico* spoligotype, lineage type, and phylogenetic classification. This database also enabled detection of IS6110 insertion sites, and 43 loci for variable number tandem repeat (VNTR) typing. The drug resistance profile of the MTBC isolates were further confirmed using PhyResSE database (http://phyresse.org/). For example, TGS-TB identified MTBC1 as a drug susceptible *Mycobacterium bovis* isolate.

Coverage refers to the percentage of reference genome bases covered by mapped sequence reads. Mean read depth indicates the mean number of times each base is mapped by a sequence read. Reference genomes used were E. *faecium* ST18 DO (TX16) (accession number NC_017960), *Haemophilus influenzae* 86-028NP (nontypeable) (accession number NC_007146), and *Mycobacterium tuberculosis* H37Rv (accession number NC000962). VRE, van- comycin resistant *Enterococcus faecium*; NtHi, non-typeable Haemophilus influenza; MTBC, *Mycobacterium tuberculosis* complex.

Troubleshooting

Table 2. Troubleshooting table.

Step #	Problems	Causes	Suggestions
23	Low concentration of AMPure XP bead captured purified products	Bead clean-up affects the quality and quantity of amplified libraries that will be included in down-stream sequencing process	Make sure AMPure XP beads are held at room temperature for 20 min before starting the clean-up process. Furthermore, ensure that 80% ethanol is freshly prepared
35	Variation in the concentration of amplified library	The concentration of input DNA used for library preparation affects the final yield of genomic data	Measure the concentration of input DNA using Qubit fluorometer rather than a nanodrop and make the ap-propriate dilution for a DNA concentration of 0.2 ng/µl
56	Poor sequencing results	The choice of forward and reverse index primer set affects the sequencing of libraries prepared	Avoid primer combinations S502 with N705 /706, and S503 with N701/702
56	Poor sequencing results	Effective denaturation of pooled library not achieved after bead clean-up	Ensure that the NaOH is freshly prepared at the correct concentration
56	Poor sequencing results	Repeated thawing and freezing of the pooled library reduces the quality of sequence reads generated	Before preparing the pooled library for loading onto MiSeq, ensure that the machine has already been appropriately cleaned after the previous run and has sufficient storage space (at least 25 GB)

Troubleshooting

Possible problems and their troubleshooting solutions are listed in **Table 2**. There are a number of limitations associated with the protocol that should be noted. These include: effective results with the protocol are reliant on the efficacy of the extraction procedure in producing a sufficient quantity of genomic DNA; analysis of sequences generated on an Illumina platform can be affected by the presence of highly repetitive regions; and depending on the output information sought, genome assembly can be influenced by the reference genome selected for the mapping of reads. Nevertheless, the protocol was effective in generating high quality sequencing data for the range of bacterial species tested.

Acknowledgments

This research was supported by funding from the Royal Hobart Hospital Research Foundation (17-104) and the Tasmanian Community Fund (36Medium00014).

MATERIALS TEXT

Reagents

- Lysozyme (VWR, Australia, Cat.# 0663-10G)
- Ethanol, Pure (Sigma-Aldrich, Australia, Cat. # E7023)
- 2-Propanol (Sigma-Aldrich, Australia, Cat. # I8912)
- Phosphate Buffered Saline (Gibco™- Thermo Fisher Scientific, UK, Cat. # 10010023)
- Ultrapure™ DNase/RNase Free Distilled Water (Invitrogen, Australia Cat. # 10977-015)
- DNeasy® Blood and Tissue Kit (Qiagen, Germany, Cat. # 69504)
- High Pure PCR Template Preparation Kit (Roche, Germany, Cat. # 11796828001)
- Qubit™ dsDNA HS (High Sensitivity) Assay Kit (Invitrogen, Australia, Cat. # Q32851)
- Nextera® DNA Library Preparation Kit (Illumina, USA, Cat. # FC-121-1030)
- Nextera® XT Library Preparation Kit (Illumina, USA, Cat. # FC-131-1024)
- Nextera® XT Index Kit (Illumina, USA, Cat. # FC-131-1001)
- Miseq Reagent Kit v2 (300 cycles) (Illumina, USA, Cat. # MS-102-2002)
- KAPA Library Quantification Kit (Illumina, USA, Cat. # 07960140001)
- Agencourt® AMPure XP beads (Beckman Coulter, USA, Cat. # A63880)

Recipes

- Qubit working solution: dilute Qubit dsDNA HS Reagent 1:200 in Qubit dsDNA HS buffer.
  For *n* samples, prepare *n* × **200 µl** working solution.

- 80% ethanol: add **2 ml** absolute ethanol into **8 ml** distilled water.

- 0.2 M NaOH: weigh **0.04 g** of NaOH pellet and dissolve it into **5 ml** distilled water.

Equipment

- Qubit™ assay tubes (Life-technologies, USA, Cat. # Q32856)
- PCR tubes (Molecular Bioproducts, USA, Cat. # MBP3412)
- Qubit® 2.0 Fluorometer (Invitrogen, Australia, Cat. # Q32866)
- Agencourt Magnetic stand (Beckman Coulter, USA, Cat. # A32782)
- Applied Biosystems® Veriti 96-Well thermal cycler (Thermo Fisher Scientific, USA)
- Rotor-Gene 6000 real-time thermocycler (Corbett Research, Australia)

SAFETY WARNINGS

See SDS (Safety Data Sheet) for safety warnings and hazards.

All bacterial cultures should be treated as potentially pathogenic to the laboratory worker and colleagues. Therefore, the use of appropriate aseptic techniques, and the wearing of appropriate personal protective equipment are strongly recommended to maintain acceptable work health and safety standards and minimise exposure to harmful agents.

Extraction of bacterial genomic DNA

1 Pellet the liquid culture media (**200 µl**) by centrifuging at **8000 x g** for **00:08:00** in a sterile microfuge tube.

CRITICAL STEP: All bacterial cultures should be treated as potentially pathogenic to the laboratory worker and colleagues. Therefore, the use of appropriate aseptic techniques, and the wearing of appropriate personal protective equipment are strongly recommended to maintain acceptable work health and safety standards and minimise exposure to harmful agents.

2 Resuspend the pellet in **600 µl** phosphate-buffered saline (1x) until the absorbance at 600 nm (A_{600}) is between 1.0 and 2.0.

2.1 Lyse the cells by adding **30 µl** lysozyme (50 mg/ml), vortex, and incubate at **37 °C** for **01:00:00** .

3 Follow the DNeasy® Blood and Tissue Kit Quick-start protocol to extract the DNA.

4 Elute the DNA in **100 µl** volume.

4.1 Treat it with **2 µl** RNase (100 mg/ml) (Qiagen, Hilden, Germany) and incubate at **Room temperature** for **01:00:00** .

5 Purify RNase-treated DNA using the High Pure PCR Template Preparation Kit.

TIP: Perform only 4 DNA spin-wash steps instead of 9 recommended steps. Pre-incubate the elution buffer in a heat block set at **70 °C**.

6 Add **100 µl** of binding buffer to RNase treated DNA and incubate at **70 °C** for **00:10:00** .

7 Add **50 µl** of 2-Propanol and transfer the content to a Roche spin column and spin at **8000 x g** for **00:01:00** .

8 Discard the flow through and insert the spin column into a new collection tube.

9 Wash by adding **500 µl** wash buffer and spin at **8000 x g** for **00:01:00** .

10 Discard the flow through and insert the spin column into a new collection tube.

11 Perform a final spin at **8000 x g** for **00:01:00** .

12 Finally, insert the column into a 1.5 ml sterile microfuge tube, add **50 µl** of pre-heated elution buffer and spin at **8000 x g** for **00:01:00** to elute the purified DNA for next generation sequencing.

CRITICAL STEP: For next generation sequencing, contaminant-free, high-molecular weight DNA with an absorbance (260 nm/280 nm) ratio between 1.8 to 2.0 is considered a high-quality template DNA.

Quantification of bacterial genomic DNA

13 Dispense **190 µl** and **198 µl** of Qubit working solution in standard and sample tubes, respectively.

14 Add **10 µl** standards (1 and 2) and **2 µl** of sample in separate Qubit assay tubes.

15 Vortex the mixture for **00:00:03** and incubate at **Room temperature** for **00:02:00** before taking the reading.

16 Adjust the DNA concentration of each sample to 0.2 ng/µl by diluting with a required volume of distilled water.

CRITICAL STEP: The use of an accurate concentration of DNA is crucial for bacterial DNA genomic library preparation.

Tagmentation and PCR amplification of bacterial genomic DNA

17 **TIP:** For all of the methods below, the recommended 96-well TYC plate can be replaced with 0.2 ml thin wall clear, flat capped PCR tubes. In addition, multichannel pipettes and the high-speed micro plate shaker can be replaced with single channel pipettes and a bench top centrifuge, respectively.

Nextera XT tagment amplicon construction

18 In a PCR tube, add **5 µl** tagmentation DNA buffer and **2.5 µl** amplification tagmentation mix to **2.5 µl** (0.2 ng/µl) input DNA.

19 Briefly vortexed the content and transfer to the thermocycler programmed for one step at **55 °C** for **00:05:00** with heated lid, followed by a hold at **10 °C** for a volume of **10 µl**.

Neutralization of Nextera XT tagment amplicon

20 Immediately after reaching the hold temperature of **10 °C** in the above step, neutralize NTA by adding **2.5 µl** neutralization tagmentation buffer and incubate at **Room temperature** for **00:05:00** .

PCR amplification

21 For amplification, add **7.5 µl** Nextera® PCR mastermix and **2.5 µl** of each index primer, 1 and 2, to a tube containing neutralized NTA.

CRITICAL STEP: Primer combinations, S502 with N705 /706 and S503 with N701/702 should be avoided. Avoid any repeated combinations and carefully note the primers used for each sample.

22 Gently pipette the content and perform a quick spin.

23 Proceed to amplification in a thermocycler programmed for a working volume of **25 µl** with the following settings and a heated lid:

Cyles	Temperature (°C)	Time
1	72	3 min
1	95	30 s
12	95	10 s
12	55	30 s
12	72	30 s
1	72	5 min
1	10	hold

The amplified, tagmented library can be stored at **4 °C** overnight for PCR clean-up the next day.

Cleaning up the PCR product

24 **NOTE:** Bring AMPure XP beads to room temperature (for **00:20:00**).

CRITICAL STEP: Prepare fresh 80% (v/v) ethanol and 0.2 M NaOH.

24.1 To **22.5 µl** of PCR product, add **11.25 µl** of vortexed (**00:00:30**) AMPure XP beads and mix by pipetting (10 times).

25 Incubate at room temperature for **00:05:00**.

26 Place the tube on a magnetic stand for **00:02:00**.

27 While leaving the PCR tubes on the magnetic stand, carefully aspirate the supernatant.

CRITICAL STEP: Do not aspirate beads. If aspirated, redo steps 24 and 25.

28 Add **100 µl** of 80% ethanol and leave on the stand in the magnetic stand for **00:00:30**.

CRITICAL STEP: Do not resuspend the beads.

29 Aspirate out the supernatant carefully.

30 Add **100 µl** of 80% ethanol and leave on the stand in the magnetic stand for **00:00:30** .

31 Aspirate out the supernatant carefully.

32 Remove the tube from magnetic stand and allow to air dry in a tube stand for approximately **00:05:00** .

CRITICAL STEP: Visually check for cracks as over drying the beads will significantly reduce elution efficiency.

33 Add **26.15 µl** of resuspension buffer and gently pipette 20 times to mix.

34 Incubate the tubes at room temperature for **00:02:00** and then place on a magnetic stand for **00:02:00** (until the supernatant cleared).

35 Transfer the supernatant (**25 µl**) to a new PCR tube.

NOTE: The final supernatant can be stored at −15°C to −20°C for up to 1 week but we recommend proceeding to library normalization immediately.

Library normalization

36 Perform the Qubit DNA quantification method as described above to determine the genomic DNA concentration in cleaned up product.

37 Pool the genomic DNA from all of the tubes.

NOTE: Sample with the lowest DNA concentration can be used in a volume of **10 µl** as the reference to prepare a library pool using the formula:

Volume required (V2) = Concentration original (S1) x Volume total (V1 = 10 µl) / Concentration required (S2).

38

To *x* µl of library pool, add *x* µl of freshly prepared 0.2 molar NaOH (final concentration 0.1 molar) and incubate for **00:05:00** at **Room temperature**.

39 To the NaOH treated suspension add an equal volume (2*x* µl) of LNS1. Label the tube as pooled amplified library (PAL).

TIPS: In this modified step, normalize the library by using LNS1 (Library Normalisation Storage Buffer 1) only.

40 Dilute the PAL to 1:1000 by adding **1 µl** of NaOH-LNS1 treated suspension to **999 µl** of ultrapure distilled water.

41 Use KAPA library quantification kit (No ROX) to check the concentration of diluted pooled library in a real time PCR system using the following set up:

Cyles	Temperature (°C)	Time
1	95	10 min
40	95	10 s
40	60	30 s

NOTE: Include a set of six DNA standards (with concentrations ranging from 20 pM to 0.0002 pM), three sets of negative control (ultrapure distilled water), and three sets of the DNA library in the qPCR run.

42 Determine the concentration of DNA in the pooled library by the standard curve method and calculate concentration in picomolar (pM) for each tube.

NOTE: To calculate the original concentration of the pooled library we applied the formula:

[Average sample concentration (in pM) . insert size standards (452 bp) . dilution factor (1000)] divisible by the Insert size of pooled library (500 BP)

For example, for a qPCR determined concentration of 2.36 pM in a 1:1000 dilution of the pooled library, the library DNA concentration will be:

(2.36 pM *x* 452 bp *x* 1000) / 500 bp = 2133.44 pM

The value obtained from the calculation represents the concentration of DNA in the pooled library.

43 To estimate the dilution factor required to achieve a final library concentration of **15 Picomolar (pM)** in a **600 µl** volume use the formula:

Volume required = (Concentration $_{required}$ × Volume $_{total}$) / Concentration $_{original}$
= (15 pM × 600 µl) / 2133.44 pM
= 4.22 µl

NOTE: Therefore, **4.22 µl** is added to **595.78 µl** of HT buffer to produce a final concentration of **15 Picomolar (pM)** , in a final volume of **600 µl**.

The diluted library is then ready to be heat denatured and loaded into the MiSeq reagent cartridge.

Preparing pooled library for loading onto MiSeq

44 Thaw the PAL at room temperature and mix by pipetting up and down (5 times) followed by brief centrifugation.

45 Based on the library concentration example above, transfer **595.78 µl** of HT buffer to a **1.5 ml** diluted amplified library (DAL) tube containing **4.22 µl** PAL.

46 Mix using a pipette (5 times).

47 Vortex the DAL tube at top speed, centrifuge briefly , and incubate exactly for **00:02:00** at **96 °C** ± **2 °C**.

48 Immediately transfer the DAL tube to ice for at least **00:05:00** or until loading.

CRITICAL STEP: Put the Illumina MiSeq sequencer through a short wash cycle to avoid cross-contamination of the DAL from previous usage.

49 Thaw the MiSeq reagent cartridge at **Room temperature**.

50 Generate a MiSeq sample sheet using the Illumina Experiment Manager. See step 22 (**go to step #22**) to identify primer sets for each sample.

51 Use the following configuration to set up the Miseq machine. Generate FASTQ workflow; FASTQ Only application; NexteraXT assay; 151 insert reads; assignment of the samples with a unique identifier and index-pair combination.

CRITICAL STEP: Rinse the flow cell with MilliQ water and remove traces of water using soft tissue paper before inserting into the machine.

52 Transfer the entire **600 µl** of DAL to the “Load here” well of the MiSeq reagent cartridge.

53 Following the setup procedure of the Illumina Experiment Manager, insert the cartridge into MiSeq instrument for sequencing to commence.

TIPS: The raw FastQ sequence reads from whole-genome sequencing can be stored on the local computer as well as on the Illumina BaseSpace server (https://basespace.illumina.com/) for further analysis.

Bioinformatic analyses

54 **NOTE:** The selection of bioinformatics software for the analysis of WGS data will be determined by the objective of the study. Here, we used Geneious 9.1.8 (Biomatters Ltd.), a desktop software to analyse our sequence data. Geneious was used to map the Fastq sequence reads to a publicly available reference genome for each species as follows:

55 Download Geneious.

56 Go to File | Import | From File. Import raw-read files (Sample_xx_R1.fastq.gz and Sample_xx_R2.fastq. gz) into Geneious.

57 Download the Reference Genome from the NCBI database. For example, *Enterococcus faecium* NC_017960. In the Left panel | Go to NCBI | Nucleotide.

57.1 Enter NC_017960 | Click Search.

57.2 Once the genome has been found, click Download Full Sequence(s) .

57.3 Download the NC_017960 reference genome (The icon changes to a green circular genome when completed).

57.4 Drag and drop the NC_017960 reference genome into the working folder.

58 Mapping the isolate sequence to the reference genome Hold CTRL and select both R1 and R2 raw read files (imported), and the reference genome (NC_017960) (downloaded).

58.1 Click Align | Assemble|Map to Reference.

58.2 Check the settings

58.3 Reference Sequence = NC_017960

58.4 Mapper = Bowtie2–fast and accurate read mapper

58.5 Trim Before Mapping = Do not trim

58.6 Results: Select all options

58.7 Results | Save consensus sequences | Options

58.8 Threshold = Highest Quality

Threshold for sequences without quality = 95%

58.9

58.10 No coverage call = '–'

59 When mapping to reference is complete, a new folder will be created containing four files:

59.1 Assembly Report

59.2 Consensus

59.3 Contig

59.4 Unused Reads

60 **NOTE:** Setting may vary depending on objective of analyses and quality of fastq reads.

NOTE: We also used open source databases, for example, TGS-TB, PhyResSe and the Center for Genomic Epidemiology's [20] ResFinder and VirulenceFinde, to further analyse the whole genome sequence data of our selection of bacterial pathogens. These freely-available databases enable the acquisition of information on bacterial pathogens that included genotype and phylogeny, antibiotic-resistance mutations, and the presence of known virulence genes.