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

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

[1]School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, TAS 7001, Australia
[2]Department of Clinical Microbiology, School of Medicine, Trinity College Dublin, Dublin D08 W9RT, Ireland
[3]School of Molecular Sciences, College of Science, Health and Engineering, La Trobe University, Victoria 3690, Australia

*Corresponding Author: Ronan F. O'Toole, Email: r.otoole@latrobe.edu.au

Competing interests: The authors have declared that no competing interests exist.

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.

Keywords: whole genome sequencing, *Enterococcus faecium*, *Haemophilus influenzae*, *Mycobacterium tuberculosis*

BACKGROUND

Using Sanger sequencing, the Human Genome Project expended approximately USD $2.7 billion and took more than 10 years to produce 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 throughput 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 understanding 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 tuberculosis* [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], *Acinetobacter 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 generate 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.
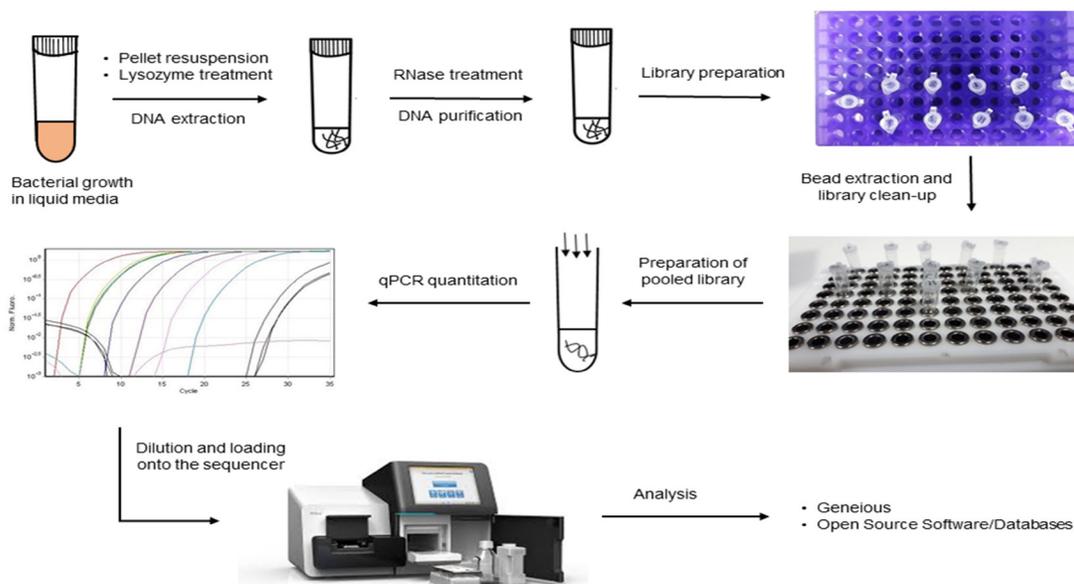
Figure 1. Graphical summary of the process of obtaining whole genome sequence data from a bacterial culture.

MATERIALS

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)

Extraction of bacterial genomic DNA

1. Pellet the liquid culture media (200 µl) by centrifuging at 8000 *g* for 8 min 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 (1×) until the absorbance at 600 nm (A_{600}) is between 1.0 and 2.0. Lyse the cells by adding 30 µl lysozyme (50 mg/ml), vortex, and incubate at 37°C for 1 h.

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

4. Elute the DNA in 100 µl volume and treat it with 2 µl RNase (100 mg/ml) (Qiagen, Hilden, Germany) and incubate at room temperature for 1 h.

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 10 min.

7. Add 50 µl of 2-Propanol and transfer the content to a Roche spin column and spin at 8000 *g* for 1 min.

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 *g* for 1 min.

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

11. Perform a final spin at 8000 *g* for 1 min.

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 *g* for 1 min 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 [15].

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 3 s and incubate at room temperature for 2 min 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

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

17. 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.

18. Briefly vortexed the content and transfer to the thermocycler programmed for one step at 55°C for 5 min with heated lid, followed by a hold at 10°C for a volume of 10 µl.

Neutralization of Nextera XT tagment amplicon

19. 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 5 min.

PCR amplification

20. 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.

21. Gently pipette the content and perform a quick spin.

22. Proceed to amplification in a thermocycler programmed for a working volume of 25 µl with the following settings: heated lid, initial cycle at 72°C for 3 min followed by 95°C for 30 s and 12 cycles of (95°C for 10 s, 55°C for 30 s and 72°C for 30 s) with a final run at 72°C for 5 min followed by a hold temperature of 10°C. The amplified, tagmented library can be stored at 2 to 4°C overnight for PCR clean-up the next day.

Cleaning up the PCR product

   NOTE: Bring AMPure XP beads to room temperature (for 20 min).

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

23. To 22.5 µl of PCR product, add 11.25 µl of vortexed (30 s) AMPure XP beads and mix by pipetting (10 times).

24. Incubate at room temperature for 5 min.

25. Place the tube on a magnetic stand for 2 min.

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

    CRITICAL STEP: Do not aspirate beads. If aspirated, redo steps 25 and 26.

27. Add 100 µl of 80% ethanol and leave on the stand in the magnetic stand for 30 s.

    CRITICAL STEP: Do not resuspend the beads.

28. Aspirate out the supernatant carefully.

29. Add 100 µl of 80% ethanol and leave on the stand in the magnetic stand for 30 s.

30. Aspirate out the supernatant carefully.

31. Remove the tube from magnetic stand and allow to air dry in a tube stand for approximately 5 min.

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

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

33. Incubate the tubes at room temperature for 2 min and then place on a magnetic stand for 2 min (until the supernatant cleared).

34. 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

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

36. 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) × Volume $_{total}$ (V1 = 10 µl) / Concentration $_{required}$ (S2).

37. To × µl of library pool, add × µl of freshly prepared 0.2 molar NaOH (final concentration 0.1 molar) and incubate for 5 min at room temperature.

38. To the NaOH treated suspension add an equal volume (2× 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.

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

40. 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: a hot start run at 95°C for 10 min followed by 40 cycles of (95°C for 10 s and 60°C for 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.

41. 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 × 452 bp × 1000) / 500 bp = 2133.44 pM

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

42. To estimate the dilution factor required to achieve a final library concentration of 15 pM in a 600 µl volume use the formula:

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

NOTE: Therefore, 4.22 µl is added to 595.78 µl of HT buffer to produce a final concentration of 15 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

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

44. 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.

45. Mix using a pipette (5 times).

46. Vortex the DAL tube at top speed, centrifuge briefly, and incubate exactly for 2 min at 96°C ± 2°C.

47. Immediately transfer the DAL tube to ice for at least 5 min 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.

48. Thaw the MiSeq reagent cartridge at room temperature [16].

49. Generate a MiSeq sample sheet using the Illumina Experiment Manager. See step 21 to identify primer sets for each sample.

50. 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.

51. Transfer the entire 600 µl of DAL to the "Load here" well of the MiSeq reagent cartridge.

52. 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

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.) [17], 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:

53. Download Geneious from https://www.geneious.com/.

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

55. Download the Reference Genome from the NCBI database. For example, *Enterococcus faecium* NC_017960.

    55.1. In the Left panel | Go to NCBI | Nucleotide.

    55.2. Enter NC_017960 | Click Search.

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

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

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

56. Mapping the isolate sequence to the reference genome

    56.1. Hold CTRL and select both R1 and R2 raw read files (imported), and the reference genome (NC_017960) (downloaded).

    56.2. Click Align | Assemble|Map to Reference.

    56.3. Check the settings

        56.3.1. Reference Sequence = NC_017960

        56.3.2. Mapper = Bowtie2–fast and accurate read mapper

        56.3.3. Trim Before Mapping = Do not trim

        56.3.4. Results: Select all options

        56.3.5. Results | Save consensus sequences | Options

        56.3.6. Threshold = Highest Quality

        56.3.7. Threshold for sequences without quality = 95%

        56.3.8. No coverage call = '–'

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

56.4.1. Assembly Report

56.4.2. Consensus

56.4.3. Contig

56.4.4. Unused Reads

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 [18], PhyResSe [19] and the Center for Genomic Epidemiology's [20] ResFinder and VirulenceFinder [21], 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.

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 regarding 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.

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.

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.

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 downstream 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 appropriate 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)

References

1. Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, et al. (2015) Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. Lancet Infect Dis 15: 1193-1202. doi: 10.1016/S1473-3099(15)00062-6. PMID: 26116186

2. Takiff HE, Feo O (2015) Clinical value of whole-genome sequencing of Mycobacterium tuberculosis. Lancet Infect Dis 15: 1077-1090. doi: 10.1016/S1473-3099(15)00071-7. PMID: 26277037

3. Harris SR, Cartwright EJP, Török ME, Holden MTG, Brown NM, et al. (2012) Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. Lancet Infect Dis 13: 130-136. doi: 10.1016/S1473-3099(12)70268-2. PMID: 23158674

4. Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, et al. (2012) Genomic epidemiology of the Escherichia coli O104:H4 outbreaks in Europe, 2011. Proc Natl Acad Sci USA 109: 3065-3070. doi: 10.1073/pnas.1121491109. PMID: 22315421

5. Llarena A, Taboada E, Rossi M (2017) Whole-Genome Sequencing in Epidemiology of Campylobacter jejuni Infections. J Clin Microbiol 55: 1269-1275. doi: 10.1128/JCM.00017-17. PMID: 28249998

6. Reuter S, Harrison TG, Köser CU, Ellington MJ, Smith GP, et al. (2013) A pilot study of rapid whole-genome sequencing for the investigation of a Legionella outbreak. BMJ Open 3: e002175. doi: 10.1136/bmjopen-2012-002175. PMID: 23306006

7. Gautam SS, Mac Aogáin M, Cooley LA, Haug G, Fyfe JA, et al. (2018) Molecular epidemiology of tuberculosis in Tasmania and genomic characterisation of its first known multi-drug resistant case. PLoS One 13: e0192351. doi: 10.1371/journal.pone.0192351. PMID: 29466411

8. McGinnis J, Laplante J, Shudt M, George KS (2016) Next generation sequencing for whole genome analysis and surveillance of influenza A viruses. J Clin Virol 79: 44-50. doi: 10.1016/j.jcv.2016.03.005. PMID: 27085509

9. Park DJ, Dudas G, Wohl S, Goba A, Whitmer SLM, et al. (2015) Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone. Cell 161: 1516-1526. doi: 10.1016/j.cell.2015.06.007. PMID: 26091036

10. Metsky HC, Matranga CB, Wohl S, Schaffner SF, Freije CA, et al. (2017) Zika virus evolution and spread in the Americas. Nature 546: 411-415. doi: 10.1038/nature22402. PMID: 28538734

11. Köser CU, Holden MTG, Ellington MJ, Cartwright EJP, Brown NM, et al. (2012) Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl J Med 366: 2267-2275. doi: 10.1056/NEJMoa1109910. PMID: 22693998

12. Halachev MR, Chan JZ, Constantinidou CI, Cumley N, Bradley C, et al. (2014) Genomic epidemiology of a protracted hospital outbreak caused by multidrug-resistant Acinetobacter baumannii in Birmingham, England. Genome Med 6: 70. doi: 10.1186/s13073-014-0070-x. PMID: 25414729

13. Davis RJ, Jensen SO, Van Hal S, Espedido B, Gordon A, et al. (2015) Whole Genome Sequencing in Real-Time Investigation and Management of a Pseudomonas aeruginosa Outbreak on a Neonatal Intensive Care Unit. Infect Control Hosp Epidemiol 36: 1058-1064. doi: 10.1017/ice.2015.133. PMID: 26050101

14. Leong KWC, Cooley LA, Anderson TL, Gautam SS, McEwan B, et al. (2018) Emergence of Vancomycin-Resistant Enterococcus faecium at an Australian Hospital: A Whole Genome Sequencing Analysis. Sci Rep 8: 6274. doi: 10.1038/s41598-018-24614-6. PMID: 29674657

15. Healey A, Furtado A, Cooper T, Henry RJ (2014) Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. Plant Methods 10: 21-8. doi: 10.1186/1746-4811-10-21. PMID: 25053969

16. Assad NA, Balmes J, Mehta S, Cheema U, Sood A (2015) Chronic obstructive pulmonary disease secondary to household air pollution. Semin Respir Crit Care Med 36: 408-421. doi: 10.1055/s-0035-1554846. PMID: 26024348

17. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647-1649. doi: 10.1093/bioinformatics/bts199. PMID: 22543367

18. Sekizuka T, Yamashita A, Murase Y, Iwamoto T, Mitarai S, et al. (2015) TGS-TB: Total Genotyping Solution for Mycobacterium tuberculosis Using Short-Read Whole-Genome Sequencing. PLoS One 10: 2015-10. doi: 10.1371/journal. pone.0142951. PMID: 26565975

19. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, et al. (2015) PhyResSE: a Web Tool Delineating Mycobacterium tuberculosis Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data. J Clin Microbiol 53: 1908-1914. doi: 10.1128/JCM.00025-15. PMID: 25854485

20. Thomsen MCF, Ahrenfeldt J, Cisneros JLB, Jurtz V, Larsen MV, et al. (2016) A Bacterial Analysis Platform: An Integrated System for Analysing Bacterial Whole Genome Sequencing Data for Clinical Diagnostics and Surveillance. PLoS One 11: e0157718. doi: 10.1371/journal.pone.0157718. PMID: 27327771

21. Kleinheinz KA, Joensen KG, Larsen MV (2014) Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and E. coli virulence genes in bacteriophage and prophage nucleotide sequences. Bacteriophage 4: e27943. doi: 10.4161/bact.27943. PMID: 24575358

22. Babraham Bioinformatics (2010) FastQC: a quality control tool for high throughput sequence data. [cited October 4, 2018]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc