

Whole genome sequences currently provide the highest resolution for typing bacterial pathogens. The implementation of next generation sequencing (NGS) in routine clinical microbiology laboratories provides the foundation to analyze bacteria with high resolution, reproducibility and accuracy. Decreasing costs and increasing ease of implementation through increasingly flexible platform options, means that more laboratories will seek this technology over time. Whole genome sequencing (WGS) has shown its value in molecular epidemiology, from seminal papers on MRSA and Mycobacterium tuberculosis helping to trace and resolve epidemics, to implementation in routine laboratories, and local molecular epidemiological studies. Methods of analysis range from determination of multi-locus sequence type (MLST; low resolution) through core genome MLST (cgMLST; high resolution) to whole genome phylogenies based on single nucleotide polymorphisms (SNPs; highest resolution). Using WGS in outbreak detection ideally takes account of all mutations and genomic variability in order to fully resolve outbreak scenarios and transmission chains. Factors encoded within the genomes, such as antimicrobial resistance (AMR) and virulence factors, can also be determined from good quality assemblies. Quality assurance, backward compatibility, communication between experts in different fields, and reporting to clinicians are issues currently being addressed. Behind all these analyses lies the all-important data. Several technologies have been used over the past decade for WGS: Ion Torrent PGM, Roche 454, PacBio and most recently Oxford Nanopore Technologies. But it is predominantly data from Illumina machines, from the MiniSeq, MiSeq, NextSeq, or HiSeq platforms, that is used for molecular epidemiology or bacterial genomics, as evidenced by the vast amounts of Illumina data deposited in databases (>90% at the Short Read Archive). Prior to the sequencing step, DNA libraries need to be made, protocols for which can vary greatly. Given the relatively high cost of library preparation compared to sequencing, and the time required to perform it, library preparation is a critical and rate-limiting step. Although many aspects of WGS can be optimized for routine diagnostic microbiology, to date few studies have addressed the data quality produced by different library methods. The most popular and implementable library protocols use proprietary transposases to cleave the DNA and ligate the adapters in one step, a method which is rapid but dependent on the DNA/enzyme concentration ratio, and is subject to sequence bias. The impact of this bias on the %G+C rich Mycobacterium tuberculosis genome has been explored, and the TruSeq (Illumina) method, involving mechanical shearing of DNA, was found to be superior to the enzymatic Nextera XT (Illumina). On the AT-rich Plasmodium falciparum genome, Nextera was again found to give highly biased results. With QIAseq FX, Qiagen have recently released a library preparation protocol that is based on fully enzymatic fragmentation (nuclease). The advantage of this approach is that the efficiency of the fragmentation is not as strongly affected by %G+C-content as the transposase from the Nextera XT approach. As QIAseq FX uses only an enzyme and not a whole complex, the adaptor ligation must then be applied in a separate step (QIAseq FX DNA Library Handbook). Another recent launch, Nextera Flex (Illumina) is also a transposome based library preparation kit, promising consistent yield and fragment size, and less sequence bias.