Rapid single-colony whole-genome sequencing of bacterial pathogens

**Claudio U. Köser[1,2]†, Louise J. Fraser[3]†, Avgousta Ioannou[3], Jennifer Becq[3], Matthew J. Ellington[2],
Matthew T. G. Holden[4], Sandra Reuter[4], M. Estée Török[1,2,5], Stephen D. Bentley[4], Julian Parkhill[4],
Niall A. Gormley[3], Geoffrey P. Smith[3] and Sharon J. Peacock[1,2,4,5]***

[1]*Department of Medicine, University of Cambridge, Cambridge, UK;* [2]*Clinical Microbiology and Public Health Laboratory, Public Health
England, Cambridge, UK;* [3]*Illumina United Kingdom, Great Chesterford, UK;* [4]*Wellcome Trust Sanger Institute, Hinxton, UK;* [5]*Cambridge
University Hospitals NHS Foundation Trust, Cambridge, UK*

*Corresponding author. E-mail: sjp97@medschl.cam.ac.uk
†Contributed equally.

Objectives: As a result of the introduction of rapid benchtop sequencers, the time required to subculture a bacterial pathogen to extract sufficient DNA for library preparation can now exceed the time to sequence said DNA. We have eliminated this rate-limiting step by developing a protocol to generate DNA libraries for whole-genome sequencing directly from single bacterial colonies grown on primary culture plates.

Methods: We developed our protocol using single colonies of 17 bacterial pathogens responsible for severe human infection that were grown using standard diagnostic media and incubation conditions. We then applied this method to four clinical scenarios that currently require time-consuming reference laboratory tests: full identification and genotyping of salmonellae; identification of bla_{NDM-1}, a highly transmissible carbapenemase resistance gene, in *Klebsiella pneumoniae*; detection of genes encoding staphylococcal toxins associated with specific disease syndromes; and monitoring of vaccine targets to detect vaccine escape in *Neisseria meningitidis*.

Results: We validated our single-colony whole-genome sequencing protocol for all 40 combinations of pathogen and selective, non-selective or indicator media tested in this study. Moreover, we demonstrated the clinical value of this method compared with current reference laboratory tests.

Conclusions: This advance will facilitate the implementation of whole-genome sequencing into diagnostic and public health microbiology.

Keywords: infectious diseases, antibiotic resistance, typing

Introduction

Standard methods used in diagnostic and reference microbiology laboratories to characterize bacterial pathogens can be time-consuming and complex.[1,2] Once a bacterial pathogen is isolated on a primary culture plate, it has to be identified to provide an early prediction of the potential disease progression and to identify antibiotics to which the pathogen is intrinsically resistant. This is followed by antibiotic susceptibility testing to identify acquired drug resistances and, in some cases, the elucidation of significant resistance mechanisms. The detection of virulence determinants such as toxins or antigens can also inform clinical management.[1,2] Finally, bacterial typing may be required for a limited number of isolates to support infection control and outbreak investigations.[1,2]

In contrast to this multifaceted testing algorithm, rapid benchtop sequencers can provide multiple pieces of clinically relevant information in a single process.[3–11] The starting material for bacterial whole-genome sequencing (WGS) is typically purified DNA extracted from liquid culture. In a clinical setting, however, it is important to sequence directly from a single bacterial colony on primary isolation plates.[2,9,12] First, this avoids the delays and costs that result from subculture and DNA extraction.[4,5,13] This is vital to direct early infection control interventions to contain outbreaks and, in the case of foodborne outbreaks, to maximize the possibility of recovering organisms from the presumed source.[7,8,14] Second, it minimizes the chance of contamination (with either a different species or a different variant of the same species) associated with subculture or picking multiple colonies from a primary isolation plate.[7] Third, it reduces the chance of introducing genomic changes, such as the loss of unstable plasmids.[15,16] Our aim was therefore to develop a simple protocol to enable WGS directly from a single bacterial colony and to validate it using a broad range of pathogenic bacteria.

Materials and methods

Ethical approval

Ethical approval was not required as clinical isolates were collected, processed and stored as part of routine clinical care at the Cambridge Public Health England Clinical Microbiology and Public Health Laboratory. Clinical isolates were anonymized prior to DNA extraction and sequencing. Research and development (R&D) approval for WGS of anonymized bacterial isolates was granted by the Cambridge University Hospitals NHS Foundation Trust R&D Department.

Reference species and clinical isolates

Details of the 17 reference and 2 clinical isolates, which encompass a diverse range of bacterial species with different modes of transmission that occur both in the hospital and community setting, are shown in Table 1 and Tables S1 and S2 (both available as Supplementary data at *JAC* Online).

Culture conditions

All bacterial isolates were grown from frozen stocks (stored at −80°C) as specified in Table S3 (available as Supplementary data at *JAC* Online).

DNA extraction, library preparation and WGS

Mechanical lysis with 425–600 micron glass beads (Sigma–Aldrich, St Louis, MO, USA) was used to extract DNA (Figure S1, available as Supplementary data at *JAC* Online). Libraries were prepared using one of two kits, depending on the colony size. The standard Nextera kit (Illumina, San Diego, CA, USA),[17] which requires 50 ng of DNA in 25 μL of UltraPure DNase RNase-free distilled (UP) water (Invitrogen, Paisley, UK), was used for species that formed large colonies (defined as a diameter of ≥1.5 mm), and the Nextera XT kit (Illumina), which needs only 1 ng of DNA in 5 μL, was used

for colonies with a smaller diameter (Table S4, available as Supplementary data at *JAC* Online). To obtain sufficiently concentrated DNA after lysis, the amount of UP water used for the lysis step was varied depending on the pathogen. For the Nextera kit, a single colony, or part of a single colony if the colony in question was very large (i.e. several millimetres) was picked with a 1 μL loop and resuspended in 40 μL of UP water in a 1.5 mL screw-capped Eppendorf tube containing glass beads [at a ratio of 1:3 of beads relative to water to minimize the amount of 'dead' volume (beads were transferred using the cap of a 0.2 mL PCR tube)]. The sample was vortexed at speed 6 on a Vortex-Genie with a microtube foam insert (Scientific Instruments, Bohemia, NY, USA) for 5 min and subsequently centrifuged in a benchtop centrifuge for 2 min at full speed. Twenty-five microlitres of the supernatant was removed and its DNA concentration was quantified using the Qubit dsDNA HS Assay Kit (Invitrogen, Eugene, OR, USA) and diluted with UP water to obtain 50 ng of DNA in 25 μL.

For small colonies processed with the Nextera XT kit (Table S4, available as Supplementary data at *JAC* Online), the process was similar. A single colony was resuspended in 25 μL of UP water with beads, then vortexed and centrifuged as described earlier, except that only 10 μL of supernatant was removed. DNA was quantified and diluted to obtain 1 ng in 5 μL of UP water. For *Clostridium difficile* 630 and *Streptococcus pneumoniae* TIGR4, which yielded the lowest amounts of DNA, the initial resuspension volume was reduced to 15 μL to obtain sufficiently concentrated DNA (7 μL of the lysed supernatant were removed in this case after lysis). Care was taken to dilute the DNA accurately, which meant that if the supernatant was too concentrated (i.e. if <0.5 μL of supernatant would have had to be transferred) an additional dilution step was done to ensure that 1 ng of DNA could be pipetted accurately.

Both the diluted and undiluted supernatants were kept on ice and the DNA was immediately processed with the respective Nextera kit. The tagmentation was also set up on ice. The manufacturer's protocol was used with the following exceptions for the standard Nextera kit. Given that the solution with the tagmented DNA probably contained viable bacteria, it was cleaned up with individual capped DNA Clean & Concentrator columns (Zymo Research, Irvine, CA, USA) rather than the 96-well version of the

Table 1. Overview of the 17 reference strains used in this study, which cause significant disease in the UK and globally;[a] in addition, we studied two clinical isolates (one *Salmonella* Enteritidis and one *K. pneumoniae*)

Species	Gram classification	Chromosome size (Mb)	GC content (%)	No. of plasmids[b]
Acinetobacter baumannii ATCC 17978	−	3.98	39	2
Campylobacter jejuni subsp. *jejuni* NCTC 11168	−	1.64	31	
Clostridium difficile 630	+	4.29	29	1
Enterobacter cloacae subsp. *cloacae* type strain ATCC 13047	−	5.31	55	2
Enterococcus faecalis V583	+	3.22	38	3
Escherichia coli str. K-12 substr. MG1655	−	4.64	51	
Haemophilus influenzae Rd KW20	−	1.83	38	
Klebsiella pneumoniae subsp. *pneumoniae* MGH 78578	−	5.32	57	5
Legionella pneumophila subsp. *pneumophila* str. Philadelphia 1	−	3.40	38	
Neisseria meningitidis serogroup B strain MC58	−	2.27	52	
Pseudomonas aeruginosa PAO1	−	6.26	67	
Salmonella enterica subsp. *enterica* serovar Enteritidis str. P125109	−	4.69	52	1
Shigella sonnei 53G	−	4.99	51	4
Staphylococcus aureus HO 5096 0412	+	2.83	33	1
Streptococcus agalactiae 2603V/R	+	2.16	36	
Streptococcus pneumoniae TIGR4	+	2.16	40	
Streptococcus pyogenes M1	+	1.85	39	

[a]Tables S1 and S2, available as Supplementary data at *JAC* Online.
[b]The sizes and GC contents of the plasmids can be found in Table S2 (available as Supplementary data at *JAC* Online).

kit. Eppendorf tubes (1.5 mL) were used to mix the binding buffer and for the elution step. Moreover, the Zymo column was placed onto a new 2 mL collection tube (Qiagen, Hilden, Germany) after each spin and the old tube discarded. For both Nextera kits, capped 0.2 mL PCR strip tubes were used for the tagmentation and amplification steps instead of 96-well plates. Rather than cleaning up 50 μL of PCR product as specified in the protocol, 40 μL of product were added to 10 μL of UP water for this step to compensate for evaporation and pipetting errors during the PCR amplification step.

The final libraries were quantified as described earlier and 1 μL of the undiluted library was analysed on a 2100 Bioanalyzer using a High Sensitivity DNA chip (Agilent Technologies, Waldbronn, Germany; Table S4, available as Supplementary data at *JAC* Online). Sequencing was done on the MiSeq sequencing platform (Illumina) using 150 bp paired-end reads as previously described.[5] The sequence data were deposited at the European Nucleotide Archive (Table S2, available as Supplementary data at *JAC* Online).

The clinical New Delhi metallo-β-lactamase (NDM)-1-producing *Klebsiella pneumoniae* was highly mucoid and care was taken to transfer biomass from the centre of the colony into 40 μL of UP water. After vortexing and centrifugation, 10 μL of the supernatant were removed slowly to minimize carryover of extracellular components. The supernatant was cleaned up using the DNA Clean & Concentrator kit as described earlier and the DNA was eluted in 10 μL of UP water, which was then processed with the Nextera XT kit.

Where appropriate, the aforementioned steps were done in a Microflow Class I Advanced Biosafety Cabinet (Bioquell UK, West Portway, UK).

Sterility tests

To check for bacterial cell death during the single-colony library preparation, three libraries for each of the 40 combinations of pathogen and growth conditions (Tables S3 and S4, available as Supplementary data at *JAC* Online) were either incubated in 10 mL of brain heart infusion broth (Public Health England Media Services, London, UK) for 5 days in air at 36°C (except for *S. pneumoniae* and *Neisseria meningitidis*, which were incubated with 5% CO_2), or on solid medium for the following: *Clostridium difficile* was incubated anaerobically on Brazier's CCEY agar and fastidious anaerobe agar with horse blood (Oxoid, Basingstoke, UK) for 5 days at 36°C; *Haemophilus influenzae* Rd KW20 was incubated on Columbia agar with chocolate horse blood for 5 days at 36°C with 5% CO_2; *Campylobacter jejuni* was incubated on Columbia blood agar with horse blood in a microaerophilic environment at 42°C for 10 days; and *Legionella pneumophila* was incubated on *Legionella* BMPA selective agar for 10 days at 36°C with 5% CO_2. More details about the culture media can be found in Table S3 (available as Supplementary data at *JAC* Online).

Analysis of reference genomes

The sequence data of the 17 reference isolates were analysed using the default settings of the 'resequencing' workflow in the Illumina Sequence Integration Software (ISIS) version 2.0, which is installed by default on the MiSeq instrument. The software and documentation of the latest version of the software are available on the Illumina customer support web site (http://support.illumina.com/sequencing/sequencing_instruments/miseq/downloads.ilmn).

The following steps are automatically performed by the software: demultiplexing; FASTQ generation and adapter trimming; alignment of the reads against the reference genome [by concatenating chromosomes and plasmids when relevant (Table S2, available as Supplementary data at *JAC* Online)]; and variant calling. ISIS uses the Burrows–Wheeler Aligner[18] for paired-reads alignment with its default parameters, apart from read trimming of bases with <Q15. ISIS uses the GATK UnifiedGenotyper[19,20] for calling both single-nucleotide polymorphisms (SNPs) and insertions/deletions (indels) using its default parameters, except for also outputting variants with phred-scaled confidence between 10 and 30 flags as LowQual.

SAMTools[21] was used to calculate the following metrics for Table S5 (available as Supplementary data at *JAC* Online): the proportion of unaligned reads per sample; the average depth of coverage per reference sequence; the proportion of reference positions that were not covered by at least one read; the proportion of reference positions that had >10× coverage; and the SNPs and indels that passed filtering with GATK caller and had a variant base frequency of 100%.

Retrospective analysis of clinical isolates

To confirm the genetic basis of carbapenem resistance in the clinical *K. pneumoniae* isolate, we used the same set of carbapenemases tested by the Public Health England Antimicrobial Resistance and Healthcare Associated Infections Reference Unit (AMRHAI, London, UK). Moreover, we included *bla*GES-1 (Table S6, available as Supplementary data at *JAC* Online). The genome of the isolate in question was assembled *de novo* using Velvet[22] and reference resistance genes were then mapped onto the assembly using SMALT software (http://www.sanger.ac.uk/resources/software/smalt/), allowing the same gene to map multiple times to the assembly using a cut-off for detection of 90% DNA sequence identity. To further assess whether the candidate genes identified from the assembly were present, the raw sequencing reads were mapped to each candidate gene, allowing the assessment of coverage and SNP/indel variation in the isolate of interest. The *bla*NDM-1 hit was confirmed using a BLAST search, and no amino acid changes compared with *bla*NDM-1 were found. The same approach was used for the detection of toxin genes in methicillin-resistant *Staphylococcus aureus* (MRSA), as previously described.[5]

For the analysis of the clinical *Salmonella* isolate, we constructed a maximum-likelihood tree of core genome SNPs using RAxML[23] and Fast tree.[24] The core genome was identified using comparative genomic analysis, and excluded mobile genetic elements and genomic island regions from the *Salmonella* Enteritidis P4 chromosome.[25] The paired-end reads from the study isolate, and DNA sequence from other previously sequenced *Salmonella* isolates (Table S7, available as Supplementary data at *JAC* Online), were mapped against the chromosome of *Salmonella* Enteritidis PT4.[25] SNPs were identified as previously described.[26]

Results and discussion

Assay development and validation

We initially explored whether the observation by Adey *et al.*,[17] who found that it is possible to generate a sequencing library directly from *Escherichia coli* colonies, could be reproduced with other Gram-negative bacteria as well as Gram-positive bacteria such as MRSA. We found that simply picking a colony of MRSA into a PCR tube containing sterile water and vortexing it before adding Nextera reagents generated successful genomic libraries. By contrast, the use of lysozyme, which is often used to lyse Gram-positive bacteria, inhibited the formation of DNA libraries (data not shown). In our final protocol we therefore opted for mechanical lysis using glass beads, which can be used for any bacterial species, followed by a dilution step of the supernatant to control the insert size of the final libraries [by normalizing the amount of DNA relative to the number of transposons (Figure S1, available as Supplementary data at *JAC* Online)].

We evaluated our protocol using 17 bacterial pathogens that are designated as 'alert organisms' by the UK Department of Health or are subject to surveillance in England [Table 1 and Table S1 (available as Supplementary data at *JAC* Online)] and internationally. We selected one representative isolate for each species using the criteria that it had been sequenced previously and represented a clinically

relevant strain. For example, for *N. meningitidis* we used the serogroup B strain MC58 because this is the most prevalent serogroup in the UK and was used to develop the 4CmenB vaccine.[27] The genomes of the 17 bacterial species had a range of sizes and GC contents, and up to five plasmids [Table 1 and Table S2 (available as Supplementary data at *JAC* Online)].

To ensure that our DNA extraction protocol was geared for use in routine diagnostic laboratory practice, we grew the reference isolates using the media and incubation conditions specified in the standard operating procedures for the isolation of each species from clinical specimens at the Cambridge Public Health England Clinical Microbiology and Public Health Laboratory (Table S3, available as Supplementary data at *JAC* Online).[1]

Irrespective of the species, colony size or culture medium, we found that suspending all or part of a single colony in water followed by vortexing with glass beads for 5 min yielded sufficient DNA to successfully prepare standard Nextera or Nextera XT libraries, as judged by their concentration and size distribution (Table S4, available as Supplementary data at *JAC* Online). Importantly, this method did not require any additional clean-up steps prior to library preparation, with the exception of a highly mucoid clinical isolate of *K. pneumoniae* (whether this also applies to other
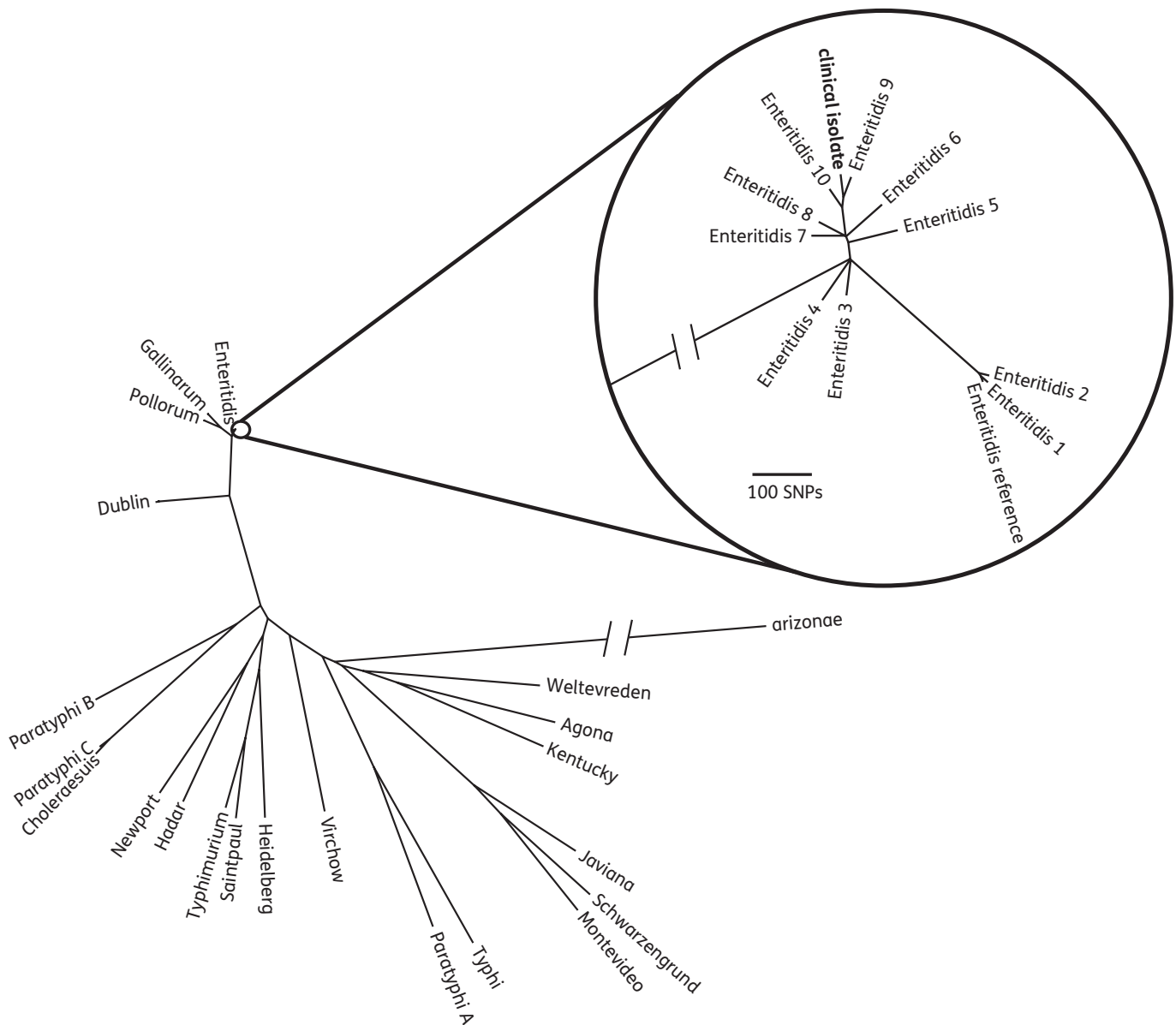

Figure 1. Unrooted maximum-likelihood tree of the clinical *Salmonella* isolate in the context of 21 different *S. enterica* serovars (e.g. Dublin) and *S. enterica* subsp. *arizonae* (with *Salmonella* Enteritidis PT4 as reference[25]). This not only showed that the clinical isolate belonged to the serovar Enteritidis, which matched the reference laboratory results, but simultaneously provided genome-level discrimination between isolates (i.e. the study isolate was only distantly related to the other previously sequenced Enteritidis isolates). Thus single-colony WGS of salmonellae could not only replace serotyping but all currently used epidemiological techniques.

mucoid isolates remains to be determined). Moreover, the libraries were found to be sterile, allowing subsequent handling in a Containment Level 1 laboratory (please refer to the Supplementary data for a more detailed discussion of this aspect).

We further confirmed the quality of the single-colony protocol by sequencing one library for each of the 17 reference isolates using the Illumina MiSeq platform. On average, 94% of each chromosome and plasmid had coverage of >10×, which represents the minimum commonly used to call SNPs (Table S5, available as Supplementary data at *JAC* Online).[15] Those plasmids that showed <90% coverage had poor mapping due to the presence of homologous regions elsewhere in the genome.[28] A single plasmid from *Shigella sonnei* 53G was not covered, but further investigation determined that this plasmid was not present in our stock of this isolate. The differences (SNPs and indels) that we found relative to the published references probably represent errors in the reference genomes or minor differences between variants of the reference strains used in different laboratories (Table S5, available as Supplementary data at *JAC* Online).[29,30]

Application to four clinical scenarios

To illustrate the potential value of rapid single-colony WGS for diagnostic purposes, we applied our protocol to four clinical scenarios in which bacterial isolates are often sent to a reference laboratory: (i) full identification of a clinical *Salmonella* isolate; (ii) verification of carbapenem resistance in a *K. pneumoniae* isolate; (iii) detection of toxin genes in MRSA; and (iv) the investigation of antigens in the vaccine strain of *N. meningitidis*. Each can result in long delays compared with diagnostic functions that are performed locally.

Bacterial pathogens are usually identified at a species or genus level, which can be achieved in laboratories using biochemical tests or matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy (MALDI-TOF).[2] For some organisms, however, this level of discrimination is insufficient for clinical management. Salmonellae, for example, are most often differentiated into serovars at reference laboratories using several agglutination reactions.[31] This extra step discriminates between those species associated with typhoid fever versus non-typhoidal gastroenteritis, and sometimes gives clues as to the possible source (e.g. if a rare serovar associated with pets such as reptiles is identified[32]). We randomly selected a clinical isolate of *Salmonella* cultured from faeces in our diagnostic laboratory, which had been identified by the reference laboratory using agglutination methods as *Salmonella enterica* subsp. *enterica* serovar Enteritidis. WGS was performed using the single-colony protocol, and the sequence placed into a phylogenetic tree containing several previously sequenced *S. enterica* serovars (Figure 1). We confirmed from the position in the tree that the test isolate was serovar Enteritidis, and simultaneously provided genome-level discrimination between isolates.

For antimicrobial susceptibility testing, a single inexpensive front-line assay (e.g. a disc diffusion test or automated liquid-culture susceptibility assay) is usually able to detect the most common phenotypic resistance patterns within 24 h.[2,5] However, to confirm or refute certain significant resistance mechanisms (e.g. the presence of the gene encoding NDM, a carbapenemase that is highly transmissible and therefore demands enhanced infection control procedures) requires the use of one or more secondary assays in a local laboratory, which have to be complemented with further time-consuming reference laboratory tests.[9,33] By contrast, single-colony WGS enabled the gold-standard molecular detection of the bla_{NDM-1} gene in a clinical isolate of *K. pneumoniae* that was isolated from a perinephric abscess (Figure 2a).[32,34] This agreed with the PCR results obtained by the AMRHAI, where this strain had been analysed after first isolation.
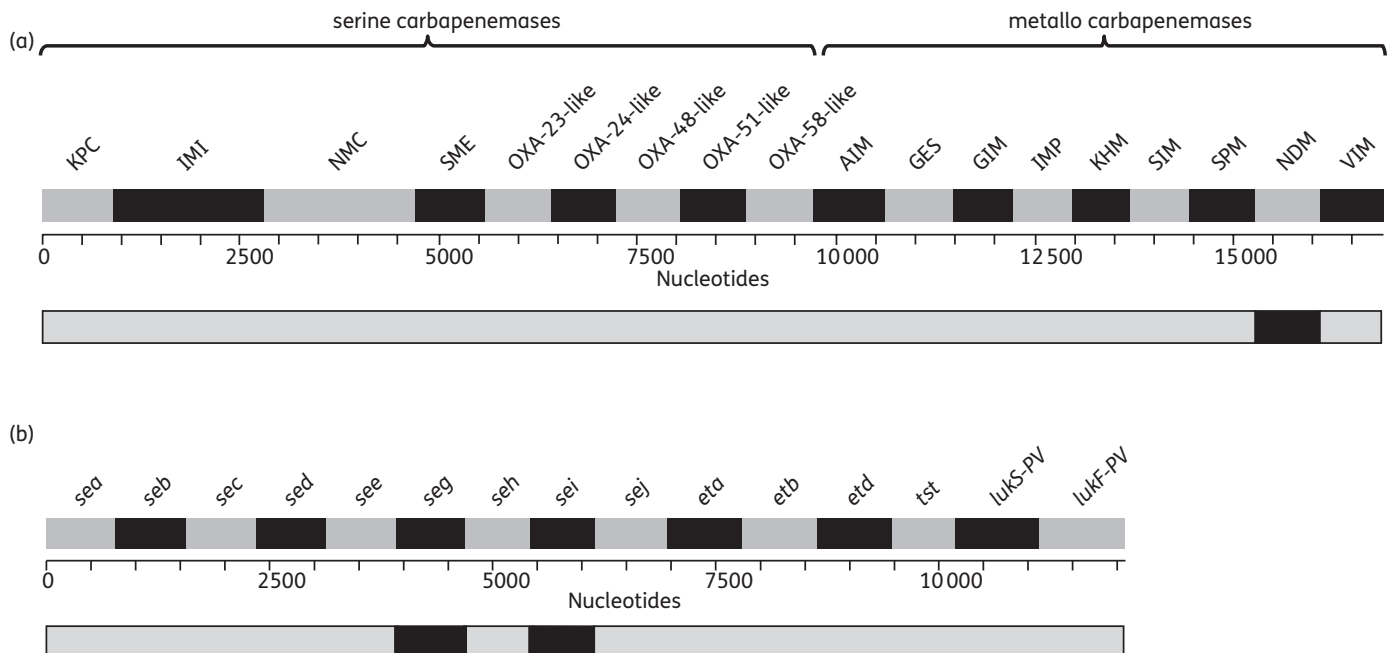

Figure 2. Heat maps showing the presence (black) or absence (grey) of resistance or virulence determinants. (a) Carbapenemase genes for the clinical *K. pneumoniae* isolate. (b) Toxome for MRSA reference strain.

Detection of bacterial genes encoding virulence determinants is not undertaken commonly, but the AMRHAI offers detection of a panel of 14 *S. aureus* toxin genes using multiplex PCRs in cases where toxic shock syndrome is suspected in a patient. We found that the equivalent information could be derived from our MRSA reference strain using the single-colony protocol (Figure 2b). We also had full coverage of all six genes encoding antigens used in the 4CmenB vaccine (fHbp, GNA2091, NHBA, GNA1030, NadA and PorA[27]) in the WGS data of the *N. meningitidis* reference strain (data not shown), which could be used to monitor the effects of the introduction of 4CmenB on these antigens.[35]

Concluding remarks

Next-generation sequencing is possible directly from clinical samples.[36,37] Yet in light of the low rate of positivity of most clinical samples, and the fact that culture and PCR are usually more sensitive than WGS, this approach is not cost-effective in most clinical scenarios.[2] As a result, a primary culture step will remain routine in diagnostic bacteriology in the foreseeable future.[2,38] Our sample preparation method does, however, eliminate the need for a secondary culture and manual DNA extraction with a kit for 17 of the most clinically important bacterial pathogens, provided that a single colony is recovered on the primary isolation plate (i.e. primary plates that show evidence of mixed, overlapping growth would still need to be subcultured in order to recover pure, single colonies for phenotypic susceptibility testing and WGS). This represents a major advance given that subculture can exceed the time required for library preparation and WGS using rapid benchtop sequencers.[2,9,12]

Once the analysis of sequence data has been fully automated to generate reports suitable for clinicians, and the cost of sequencing has fallen further, we predict that rapid single-colony WGS will solve one of the major problems facing diagnostic bacteriology.[1,2,9,39,40] Because the four clinical scenarios addressed in this paper are relatively rare in the UK, the costs for reagents, equipment and staff for the associated tests are often prohibitive for local diagnostic laboratories.[9] By contrast, combining different bacterial species in the same sequence run will probably make single-colony WGS cost-effective in the future. As a result, local laboratories will be able to provide many current reference laboratory tests, which would avoid the delays associated with shipping.[2,9] In fact, WGS can address multiple clinically relevant questions for many pathogens, as illustrated in this proof-of-principle study. For example, we showed that WGS of salmonellae can replace serotyping as well as current typing techniques such as multilocus sequence typing, variable-number tandem repeat analysis, PFGE and microarrays.[14,31,39] Similarly, the WGS data from the NDM-1-producing *K. pneumoniae* that we used to confirm the carbapenem resistance mechanism could equally have been used to track onward transmission[9,16,41] or to monitor the genetic basis of resistance to other antibiotics, thereby providing a powerful surveillance tool.[2,5,42–44]

Acknowledgements

Funding

Transparency declarations

L. J. F., A. I., J. B., N. A. G. and G. P. S. are employees and shareholders of Illumina Inc. M. J. E. and M. E. T. were funded to attend conferences by Bruker Daltonics and Illumina Inc., respectively. M. E. T. has received speaker's honoraria and book royalties from Oxford University Press. J. P. has received funding for travel and accommodation from Pacific Biosciences Inc. and Illumina Inc. S. J. P. is a consultant for Pfizer Inc. and received funding for travel and accommodation from Illumina Inc. All other authors: none to declare.

Supplementary data

Supplementary data, including Figure S1 and Tables S1 to S7, are available at *JAC* Online (http://jac.oxfordjournals.org/).

References

1 Didelot X, Bowden R, Wilson DJ *et al*. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet* 2012; **13**: 601–12.

2 Köser CU, Ellington MJ, Cartwright EJ *et al*. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog* 2012; **8**: e1002824.

3 Rohde H, Qin J, Cui Y *et al*. Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. *N Engl J Med* 2011; **365**: 718–24.

4 Eyre DW, Golubchik T, Gordon NC *et al*. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. *BMJ Open* 2012; **2**: e001124.

5 Köser CU, Holden MT, Ellington MJ *et al*. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med* 2012; **366**: 2267–75.

6 Vogel U, Szczepanowski R, Claus H *et al*. Ion torrent personal genome machine sequencing for genomic typing of *Neisseria meningitidis* for rapid determination of multiple layers of typing information. *J Clin Microbiol* 2012; **50**: 1889–94.

7 Harris SR, Cartwright EJ, Török ME *et al*. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect Dis* 2013; **13**: 130–6.

8 McDonnell J, Dallman T, Atkin S *et al*. Retrospective analysis of whole genome sequencing compared to prospective typing data in further informing the epidemiological investigation of an outbreak of *Shigella sonnei* in the UK. *Epidemiol Infect* 2013; **141**: 2568–75.

9 Reuter S, Ellington MJ, Cartwright EJ *et al*. Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology. *JAMA Intern Med* 2013; **173**: 1397–404.

10 Reuter S, Harrison TG, Köser CU *et al*. A pilot study of rapid whole-genome sequencing for the investigation of a *Legionella* outbreak. *BMJ Open* 2013; **3**: e002175.

11 Sherry NL, Porter JL, Seemann T *et al*. Outbreak investigation using high-throughput genome sequencing within a diagnostic microbiology laboratory. *J Clin Microbiol* 2013; **51**: 1396–401.

12 Harris SR, Török ME, Cartwright EJ *et al*. Read and assembly metrics inconsequential for clinical utility of whole-genome sequencing in mapping outbreaks. *Nat Biotechnol* 2013; **31**: 592–4.

13 Eyre DW, Cule ML, Wilson DJ *et al*. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N Engl J Med* 2013; **369**: 1195–205.

14 Elkins CA, Kotewicz ML, Jackson SA *et al*. Genomic paradigms for food-borne enteric pathogen analysis at the US FDA: case studies highlighting method utility, integration and resolution. *Food Addit Contam Part A* 2013; **30**: 1422–36.

15 Holt KE, Baker S, Weill FX *et al*. *Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. *Nat Genet* 2012; **44**: 1056–9.

16 López-Camacho E, Gómez-Gil R, Tobes R *et al*. Genomic analysis of the emergence and evolution of multidrug resistance during a *Klebsiella pneumoniae* outbreak including carbapenem and colistin resistance. *J Antimicrob Chemother* 2014; **69**: 632–6.

17 Adey A, Morrison HG, Asan *et al*. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. *Genome Biol* 2010; **11**: R119.

18 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754–60.

19 McKenna A, Hanna M, Banks E *et al*. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–303.

20 DePristo MA, Banks E, Poplin R *et al*. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43**: 491–8.

21 Li H, Handsaker B, Wysoker A *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–9.

22 Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; **18**: 821–9.

23 Rokas A. Phylogenetic analysis of protein sequence data using the Randomized Axelerated Maximum Likelihood (RAXML) Program. *Curr Protoc Mol Biol* 2011; **96**: 19.11.1–14.

24 Price MN, Dehal PS, Arkin AP. FastTree 2-approximately maximum-likelihood trees for large alignments. *PLoS One* 2010; **5**: e9490.

25 Thomson NR, Clayton DJ, Windhorst D *et al*. Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res* 2008; **18**: 1624–37.

26 Croucher NJ, Harris SR, Fraser C *et al*. Rapid pneumococcal evolution in response to clinical interventions. *Science* 2011; **331**: 430–4.

27 Bai X, Findlow J, Borrow R. Recombinant protein meningococcal serogroup B vaccine combined with outer membrane vesicles. *Expert Opin Biol Ther* 2011; **11**: 969–85.

28 Cahill MJ, Köser CU, Ross NE *et al*. Read length and repeat resolution: exploring prokaryote genomes using next-generation sequencing technologies. *PLoS One* 2010; **5**: e11518.

29 Köser CU, Niemann S, Summers DK *et al*. Overview of errors in the reference sequence and annotation of *Mycobacterium tuberculosis* H37Rv, and variation amongst its isolates. *Infect Genet Evol* 2012; **12**: 807–10.

30 Ribeiro FJ, Przybylski D, Yin S *et al*. Finished bacterial genomes from shotgun sequence data. *Genome Res* 2012; **22**: 2270–7.

31 Achtman M, Wain J, Weill FX *et al*. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS Pathog* 2012; **8**: e1002776.

32 Bertrand S, Rimhanen-Finne R, Weill FX *et al*. *Salmonella* infections associated with reptiles: the current situation in Europe. *Euro Surveill* 2008; **13**: pii=18902.

33 Nordmann P, Poirel L, Walsh TR *et al*. The emerging NDM carbapenemases. *Trends Microbiol* 2011; **19**: 588–95.

34 Livermore DM, Andrews JM, Hawkey PM *et al*. Are susceptibility tests enough, or should laboratories still seek ESBLs and carbapenemases directly? *J Antimicrob Chemother* 2012; **67**: 1569–77.

35 Gill CJ. Novel assessment of a novel meningitis B vaccine. *Lancet Infect Dis* 2013; **13**: 381–2.

36 Loman NJ, Constantinidou C, Christner M *et al*. A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104:H4. *JAMA* 2013; **309**: 1502–10.

37 Hasman H, Saputra D, Sicheritz-Ponten T *et al*. Rapid whole genome sequencing for the detection and characterization of microorganisms directly from clinical samples. *J Clin Microbiol* 2013; doi:10.1128/JCM.02452-13.

38 Köser CU, Bryant JM, Becq J *et al*. Whole-genome sequencing for rapid susceptibility testing of *M. tuberculosis*. *N Engl J Med* 2013; **369**: 290–2.

39 Struelens MJ, Brisse S. From molecular to genomic epidemiology: transforming surveillance and control of infectious diseases. *Euro Surveill* 2013; **18**: pii=20386.

40 Larsen MV. Internet-based solutions for analysis of next-generation sequence data. *J Clin Microbiol* 2013; **51**: 3162.

41 Snitkin ES, Zelazny AM, Thomas PJ *et al*. Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Sci Transl Med* 2012; **4**: 148ra16.

42 Aarestrup FM, Brown EW, Detter C *et al*. Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. *Emerg Infect Dis* 2012; **18**: e1.

43 Zankari E, Hasman H, Kaas RS *et al*. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. *J Antimicrob Chemother* 2013; **68**: 771–7.

44 Stoesser N, Batty EM, Eyre DW *et al*. Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data. *J Antimicrob Chemother* 2013; **68**: 2234–44.