Genome sequencing in clinical microbiology

Jacqueline Z-M Chan, Mark J Pallen, Beryl Oppenheim & Chrystala Constantinidou

High-throughput sequencing is being adopted to characterize microbial pathogens and track their evolution and spread.

The remit of the clinical microbiology laboratory encompasses analysis of samples, detection and identification of microorganisms associated with infection, determination of antibiotic susceptibilities and monitoring of pathogen spread. Although several new molecular approaches, such as nucleic acid amplification, mass spectrometry and sequence typing, have been adopted to varying extents, many of the techniques in use today would have been familiar to a nineteenth-century microbiologist. Now, as shown in a spate of recent papers (**Table 1**), clinical microbiology is on the cusp of a major transformation driven by high-throughput genome sequencing, used as an adjunct, or even an alternative, to traditional diagnostic approaches (**Fig. 1**). In addition, sequencing technology is shedding light on the unseen world of unculturable microorganisms and enhancing our ability to track the evolution and spread of pathogens and resistance genes in our hospitals and communities.

The pace of technological development in sequencing is astonishing, with sequencing capability doubling every 6–9 months[1]. Three major platforms currently compete on the world stage: Illumina (GA IIx, MiSeq, HiSeq), Roche (454 GS-FLX+, 454 GS-Junior) and Life Technologies (SOLiD, PGM, Proton). All three have advantages and disadvantages in terms of read length, accuracy, run-time, workflow, throughput and cost[1]. Single-molecule sequencing technologies (Pacific Biosciences, Oxford Nanopore), which provide multikilobase reads, are poised to change the sequencing landscape, not only because of their advantages in genome assembly but because of their cheap and easy, library-free sample preparation protocols. Hybrid approaches that combine short-read, second-generation technologies and long-read data from third-generation platforms are gaining in popularity, with recent publications by Bashir *et al.*[2] and Koren *et al.*[3]. Both groups found that high-fidelity short reads serve to correct the high-error sequence data obtained on a PacBio RS platform. Such efforts bring us closer to the ultimate aim of one-contig, one-chromosome genome assemblies, which would enable rapid and definitive microbial identification and characterization.

All authors are at the Institute of Microbiology and Infection, University of Birmingham, UK. Mark J. Pallen, Beryl Oppenheim and Chrystala Constantinidou are also at the NIHR Surgical Reconstruction and Microbiology Research Centre, Queen Elizabeth Hospital, Birmingham, UK.
e-mail: m.pallen@bham.ac.uk

Outbreak investigation

Bacterial genomic epidemiology has, until recently, concentrated on the identification of single-nucleotide polymorphisms (SNPs). This approach has been applied to track the spread of pathogens worldwide[4–9], within communities[10,11], between and within hospitals[12–14] and even between patients in adjacent beds[15]. Most recently, Eyre *et al.*[13] applied SNP-based epidemiology in real time and showed that integration of benchtop sequencing into routine clinical practice can result in enhanced infection control. McAdam *et al.*[12] tracked the emergence and transmission of a major lineage of methicillin-resistant *Staphylococcus aureus* (MRSA), using a high-resolution phylogenomic approach to reveal the time frame for the emergence and expansion of MRSA pandemics and to identify the genetic components that facilitate adaptation of clones to a hospital environment.

In another recent example of the power of genomic epidemiology, Baker *et al.*[10] have shown how high-throughput sequencing can be combined with Google maps to replicate the classical epidemiology of John Snow (founder of epidemiology; he identified a water pump as the source of the London 1854 cholera outbreak), but with a twenty-first century twist.

SNP typing is no longer the only genomic epidemiology tool available. For example, Bashir *et al.*[2] carried out epidemiological studies on the basis of genome structural features. In particular, their assembly showed that the genome of a Haitian strain of *Vibrio cholerae* contains two tandem copies of the toxin-linked cryptic genetic element, in common with many other *V. cholerae* isolates from the seventh cholera pandemic, but also has a particular transposase gene in this region, which had previously been seen only in an isolate from Nepal[2].

Sequence-based patient management

Since the time of Koch and Pasteur, the established paradigm in clinical microbiology has been that most infections are caused by single microbial agents that can be isolated in pure culture in the laboratory. However, these assumptions have been challenged in recent years, first because many potential pathogens are difficult or impossible to grow, and second because microorganisms almost never exist in pure culture but instead reside in complex microbial communities including biofilms[16–18]. Indeed, in some cases it seems that a community of microorganisms rather than one culprit species should be treated as a disease- or health-promoting entity[19–23]. High-throughput sequencing can form the basis of culture-independent approaches for cataloguing and understanding the complex microbial communities that are associated with health and disease in humans. In this context, the term metagenome is applied to the total genetic material present in a complex microbial community.

The stage is now set for metagenome-based patient management. The microbiology of chronic wounds, which represents a growing problem in modern medicine, provides an obvious early application for clinical metagenomics[24]. Tuttle *et al.*[25] carried out a longitudinal study of the microbiota of wounds by conventional culturing and high-throughput sequencing of bacterial molecular barcodes amplified from specimens. Sequence-based diagnosis, but not culturing of organisms, revealed that slow-to-heal wounds were characterized by higher bacterial abundance and diversity and that, in wounds that did heal, there was an increase in the ratio of *Pseudomonadaceae* to *Actinomycetales* compared with nonhealing wounds. A similar study by Wolcott *et al.*[26] of surgical site infections identified two previously uncharacterized *Bacteroidales* that predominate in chronic wounds.

High-throughput sequencing has also been applied to cDNA samples from the nasopharynx to detect viral pathogens in patients with acute lower respiratory tract infections[27,28]. Greninger *et al.*[27] noted that a streamlined metagenomic detection strategy could, in the near future, replace the multiple diagnostic tests currently required to identify an outbreak of a novel pathogen. And in another study, sequence-based, culture-independent approaches shed light on the microbiology of

Table 1 Selected bacterial genomic epidemiology reports 2010–2012

Use of HTS		Platform	Description	Reference
Tracking pathogen spread	Globally	Illumina GA IIx	High-throughput sequencing provided a global snapshot of the epidemiology and microevolution of 63 MRSA ST239 isolates. A robust and precise phylogenetic tree based on core-genome SNPs showed striking consistency with geographic source. More than 25% of the homoplasies detected were related to antibiotics currently in use.	4
		Illumina GA IIx	A comparative genomics study highlighted the genetic relatedness of 24 Nepalese *V. cholerae* isolates to three previously sequenced isolates from the 2010 Haitian outbreak.	5
		PacBio RS DNA Sequencer	The application of single-molecule sequencing technology to the high-profile German *Escherichia coli* outbreak yielded crucial insights into strain origins and pathogenicity.	7
		Ion Torrent PGM	Open-source genomics approach to public health management used rapid benchtop DNA sequencing technology, open-source data release and prompt crowd-sourced data analyses to reveal the lineage and virulence factors of the *E. coli* German outbreak strain in less than a week.	8
		PacBio RS DNA Sequencer	*V. cholerae* strains from the 2010 Haitian outbreak were sequenced on the PacBio platform and compared with sequences of other outbreaks. The Haitian strains were more closely related to South Asian strains than to South American strains.	9
		Illumina GA IIx	From SNP analysis and temporal and geographical distribution, compelling evidence was presented for repeated radiations of the El Tor biotype from the Bay of Bengal.	6
	Nationally within the community	Illumina GA IIx	Social network analysis integrated with genome sequencing is used to identify tuberculosis outbreaks and putative transmission events within a Canadian city.	11
		Illumina GA IIx	Classical epidemiology combined with high-throughput sequencing enabled characterization of endemic typhoid fever transmission in Nepal. Geographical clustering of cases unrelated to population density; *Salmonella enterica* serovar Typhi infections limited to specific locations, *Salmonella enterica* serovar Paratyphi A cases widely disseminated probably through contaminated ground water.	10
	Between hospitals	Illumina GA IIx & Roche 454 GS FLX	An extensive study documented evidence for the emergence of the EMRSA-16 lineage 35 years ago and its subsequent transmission within, and adaptation to, the hospital environment in the UK.	12
	Within hospitals	Illumina MiSeq	A retrospective study of an MRSA outbreak in a neonatal ward showed that some isolates sharing identical antibiograms were not associated with the outbreak and that an infection originally misidentified as community-associated was in fact acquired on the ward.	14
		Illumina MiSeq	A pioneering proof-of-principle study showed the integration of benchtop sequencing into routine clinical practice. Analysis of two MRSA clusters, where genomic and epidemiological evidence were in agreement, led to enhanced infection control. Analysis of two putative *Clostridium difficile* clusters suggested undetected community transmission events.	13
	Between patients (in adjacent beds)	Roche 454 GS FLX	High-throughput sequencing shed light on a small-scale outbreak, allowing discrimination between alternative epidemiological hypotheses.	15
Evolution of microbial pathogens		Illumina GA IIx	More than 100 isolates of *Burkholderia dolosa* obtained from 14 individuals over 16 years were sequenced. Evidence of parallel adaptive evolution was found in 17 bacterial genes responsible for antibiotic resistance and bacterial membrane composition.	29
		Illumina GA IIx & Roche 454 GS FLX	Examination using HTS of the evolution of the *Streptococcus pneumoniae* PMEN1 multidrug-resistance lineage over 24 years. The authors concluded that this highly recombinogenic pathogen can adapt in response to clinical interventions over remarkably short time scales.	30
Metagenomic medicine		Roche 454 GS FLX	16S rRNA pyrosequencing compared the metagenome of wound debridement samples from ten patients. Wounds that had not healed at 6 months had a significantly higher bacterial abundance and diversity.	25
		Roche 454 GS FLX	Bacterial diversity in surgical site infections was examined by 16S amplicon sequencing. Unlike previously published results, anaerobic bacilli seemed to predominate in these infections.	26
HTS-derived diagnosis		Illumina GA IIx	Comparison of two metagenomic-based diagnostic methods—Virochip and deep sequencing, for the identification of H1N1 influenza A virus from nasopharyngeal swabs. Detection by sequencing was possible at titers near the limits of detection for RT-PCR. In addition, when the sequence data from all samples were combined, 90% of the viral genome could be assembled *de novo*.	27
		Illumina GA IIx	The metagenome of nasopharyngeal aspirate samples from patients with acute lower respiratory tract infections was examined by deep sequencing. The authors identified seven known respiratory viral agents including a case of co-infection that would have been misdiagnosed by conventional PCR assays.	28

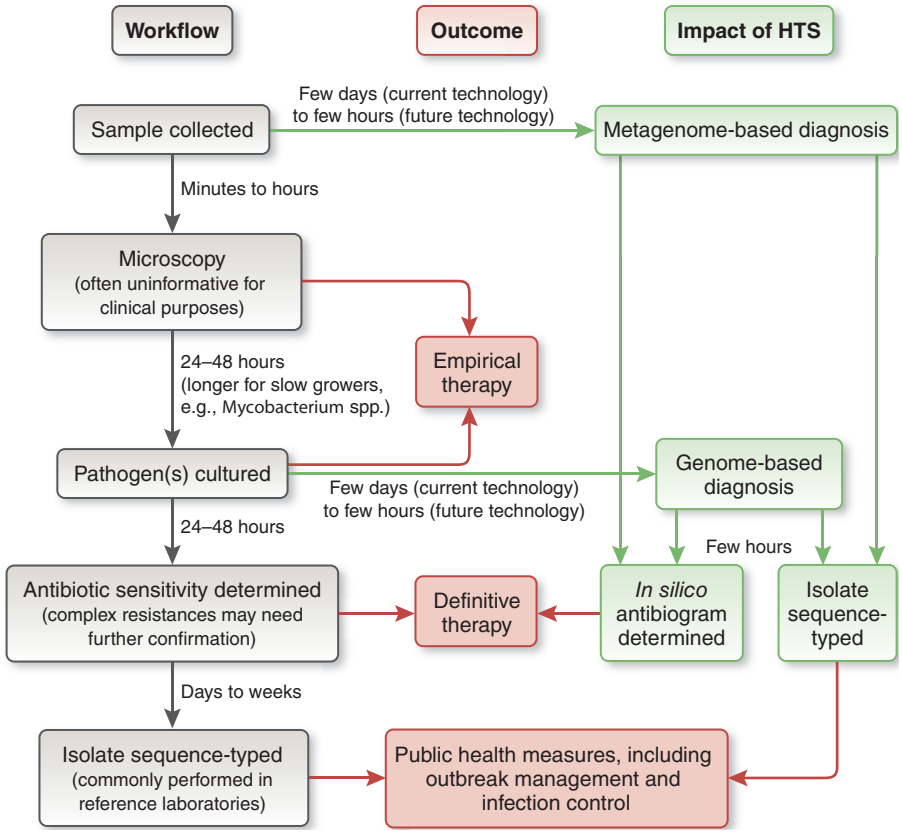

Figure 1 Clinical microbiological laboratory workflow, now and in the future. High-throughput sequencing (HTS) has the potential to replace or complement standard clinical microbiology practice, such as PCR diagnostics, mass spectrometry and vaccine design. The transition is likely to be gradual given the many challenges.

patients receiving stem-cell transplants, showing that massive increases in the proportion of enterococci in the gut were predictive of subsequent bloodstream infection with vancomycin-resistant enterococci[23]. The authors suggested that monitoring of gut microbiota by high-throughput sequencing could help to identify patient subgroups at increased risk of developing infections and guide clinical interventions, such as administration of probiotics, that might prevent or attenuate infection.

From sequence to antibiogram

High-throughput sequencing has already contributed to the identification and characterization of resistance determinants. For example, Lieberman *et al.*[29] used the Illumina platform to sequence whole genomes of more than 100 isolates of *Burkholderia dolosa* obtained from 14 individuals during a 16-year period. They found evidence of parallel adaptive evolution in 17 bacterial genes responsible for antibiotic resistance and bacterial membrane composition. Croucher *et al.*[30] examined the evolution of the *Streptococcus pneumoniae* PMEN1 multidrug-resistance lineage over 24 years and concluded that this highly recombinogenic pathogen can adapt in response to clinical interventions during remarkably short time scales. Hornsey *et al.*[31] identified the genomic changes associated with loss of sensitivity to tigecycline in multidrug-resistant *Acinetobacter baumannii*. Changes included a single nonsynonymous SNP in a gene encoding a signal transduction histidine kinase, AdeS, implicated in the upregulation of the AdeABC efflux system. Whole-genome sequencing also identified a disruption in the DNA mismatch repair gene *mutS*, which the authors suggest could have led to an increase in the rate of accumulation of SNPs. Feng *et al.*[32] applied high-throughput sequencing to identify novel resistance mutations in genes involved in linezolid resistance. In addition to a previously described resistance mutation in the 23S RNA gene, three new mutations were experimentally confirmed to be involved in resistance, with another 28 implicated but not verified. More recently, Kos *et al.*[33] have used Illumina sequencing to compare the genomes of 12 vancomycin-resistant isolates of *Staphylococcus aureus*, showing that acquisition of Tn1546 is responsible for resistance.

How easy will it be to predict antibiotic resistance and virulence phenotypes from genome sequences? A recent publication by Köser *et al.*[14] provides grounds for optimism in reporting complete concordance between genotypic and phenotypic evidence for antibiotic resistance in MRSA isolates from a neonatal ward.

Moving sequencing into routine use

The future organization of clinical microbiology laboratories has been the subject of much debate[34,35]. The introduction of automation and molecular biology techniques have decreased the time to diagnosis and improved standardization[36], but can the current format cope with the threat of bioterrorism and the rapid spread of multidrug-resistant pathogens mediated by global travel? Would the integration of high-throughput sequencing into clinical workflows solve these problems (**Fig. 1**)? Perhaps it might, but clearly there are many challenges to be overcome before sequencing can become a routine technology for the diagnosis of infection.

These challenges are biological, technical and organizational. In terms of biology, we have yet to document the mode and tempo of genomic evolution at all epidemiologically relevant time scales and in all clinically relevant taxa. Similarly, despite tremendous progress with the human microbiome project[37–40], we have not catalogued and sequenced representatives of all the bacterial species associated with human health and disease. In some cases, we may have to accept that whole-genome sequencing provides no more information than conventional approaches to molecular typing and that we may not always be able to predict clinically relevant phenotypes from genome sequences.

What has become clear is that methods of DNA extraction can have a substantial effect on results obtained[27,41], so among the technical challenges is the need to devise a standard protocol for sample preparation that can differentiate DNA from living and dead cells[42,43]. Further to this, standard operating procedures using approved reagents on approved instruments should be developed for all clinical microbiology applications of high-throughput sequencing. The Human Microbiome Project has laid the groundwork for such initiatives, but we face the additional challenge that any accepted procedures will rapidly become outdated in the face of continuing improvements in sequencing technologies.

There are also many organizational, social and cultural challenges. Laboratory staff will require training to operate high-throughput sequencing instruments. Most importantly,

analytical pipelines and databases will have to be created to parse sequencing data into information that can be readily interpreted by clinicians. Turnaround times and sample costs will have to match those of current techniques—in this context, analysis rather than generation of sequence data is likely to become the limiting factor[1]. Managers of diagnostic and reference laboratories will have to be persuaded that, in most cases, culture-based approaches can and should be replaced with sequence-based approaches, with culture-based methods restricted to specialized applications.

Adding to this chorus of caveats, there are ongoing technological issues that remain to be tackled. Compared with the nineteenth century method of culture on laboratory media, the cost of genome sequencing remains prohibitive. This cost is compounded when one considers that most samples processed in diagnostic laboratories show no significant growth or are negative, meaning that large amounts of sequencing might be done without producing clinically relevant information. Another stumbling block is how to process the large amount of sequencing data that will be amassed. Results obtained from different platforms are often not compatible or comparable, which complicates data sharing. Bioinformatic analyses will therefore have to be rigorously tested and standardized. We must also address the question of where and how long data should be stored and whether storage poses any ethical issues. Finally, we are still heavily reliant on culture-based techniques to obtain the samples required for high-throughput sequencing, and there is still some way to go before sequencing becomes fully independent of culturing.

Caveats aside, however, it no longer seems to be a question of 'if' high-throughput sequencing will ever enter the clinical microbiology laboratory but of 'when'. This means that clinical microbiologists need to prepare for a diagnostic paradigm in which Pasteur and Koch make way for Watson and Crick.

COMPETING FINANCIAL INTERESTS
The authors declare competing financial interests: details are available in the online version of the paper.

1. Loman, N.J. *et al. Nat. Rev. Microbiol.* **10**, 599–606 (2012).
2. Bashir, A. *et al. Nat. Biotechnol.* **30**, 701–707 (2012).
3. Koren, S. *et al. Nat. Biotechnol.* **30**, 693–700 (2012).
4. Harris, S.R. *et al. Science* **327**, 469–474 (2010).
5. Hendriksen, R.S. *et al. mBio* **2**, e00157–11 (2011).
6. Mutreja, A. *et al. Nature* **477**, 462–465 (2011).
7. Rasko, D.A. *et al. N. Engl. J. Med.* **365**, 709–717 (2011).
8. Rohde, H. *et al. N. Engl. J. Med.* **365**, 718–724 (2011).
9. Chin, C.-S. *et al. N. Engl. J. Med.* **364**, 33–42 (2011).
10. Baker, S. *et al. Open Biol.* **1**, 110008 (2011).
11. Gardy, J.L. *et al. N. Engl. J. Med.* **364**, 730–739 (2011).
12. McAdam, P.R. *et al. Proc. Natl. Acad. Sci. USA* **109**, 9107–9112 (2012).
13. Eyre, D.W. *et al. BMJ Open* **2**, e001124 (2012).
14. Köser, C.U. *et al. N. Engl. J. Med.* **366**, 2267–2275 (2012).
15. Lewis, T. *et al. J. Hosp. Infect.* **75**, 37–41 (2010).
16. Nadell, C.D., Xavier, J.B. & Foster, K.R. *FEMS Microbiol. Rev.* **33**, 206–224 (2009).
17. Rogers, G.B., Carroll, M.P. & Bruce, K.D. *J. Med. Microbiol.* **58**, 1401–1418 (2009).
18. Hall-Stoodley, L., Costerton, J.W. & Stoodley, P. *Nat. Rev. Microbiol.* **2**, 95–108 (2004).
19. Cho, I. & Blaser, M.J. *Nat. Rev. Genet.* **13**, 260–270 (2012).
20. Rogers, G.B. *et al. Gut Microbes* **1**, 85–93 (2010).
21. Stecher, B. & Hardt, W.-D. *Trends Microbiol.* **16**, 107–114 (2008).
22. de Sablet, T. *et al. Infect. Immun.* **77**, 783–790 (2009).
23. Ubeda, C. *et al. J. Clin. Invest.* **120**, 4332–4341 (2010).
24. Sen, C.K. *et al. Wound Repair Regen.* **17**, 763–771 (2009).
25. Tuttle, M.S. *et al. J. Clin. Microbiol.* **49**, 3812–3819 (2011).
26. Wolcott, R.D., Gontcharova, V., Sun, Y., Zischakau, A. & Dowd, S.E. *J. Wound Care* **18**, 317–323 (2009).
27. Greninger, A.L. *et al. PLoS ONE* **5**, e13381 (2010).
28. Yang, J. *et al. J. Clin. Microbiol.* **49**, 3463–3469 (2011).
29. Lieberman, T.D. *et al. Nat. Genet.* **43**, 1275–1280 (2011).
30. Croucher, N.J. *et al. Science* **331**, 430–434 (2011).
31. Hornsey, M. *et al. J. Antimicrob. Chemother.* **66**, 1499–1503 (2011).
32. Feng, J. *et al. Genome Res.* **19**, 1214–1223 (2009).
33. Kos, V.N. *et al. mBio* **3**, e00112–12 (2012).
34. Fournier, P.-E. & Raoult, D. *Annu. Rev. Microbiol.* **65**, 169–188 (2011).
35. Pallen, M.J., Loman, N.J. & Penn, C.W. *Curr. Opin. Microbiol.* **13**, 625–631 (2010).
36. Raoult, D., Fournier, P.E. & Drancourt, M. *Nat. Rev. Microbiol.* **2**, 151–159 (2004).
37. The NIH HMP Working Group *et al. Genome Res.* **19**, 2317–2323 (2009).
38. Li, K., Bihan, M., Yooseph, S. & Methé, B.A. *PLoS ONE* **7**, e32118 (2012).
39. Wylie, K.M. *et al. PLoS ONE* **7**, e35294 (2012).
40. The Human Microbiome Project Consortium. *Nature* **486**, 207–214 (2012).
41. Sergeant, M.J., Constantinidou, C., Cogan, T., Penn, C.W. & Pallen, M.J. *PLoS ONE* **7**, e38094 (2012).
42. Azhikina, T.L. *et al. Biotechniques* **48**, 139–144 (2010).
43. Nocker, A. & Camper, A.K. *FEMS Microbiol. Lett.* **291**, 137–142 (2009).