REVIEW

# Next-generation and whole-genome sequencing in the diagnostic clinical microbiology laboratory

W. M. Dunne Jr · L. F. Westblade · B. Ford

**Abstract** The identification and/or prediction of the antimicrobial resistance of microorganisms in clinical materials solely by molecular means in the diagnostic microbiology laboratory is not novel. However, the ability to sequence multitudes of bacterial genomes and deliver and interpret the resultant sequence information in near "real-time" is the basis of next-generation sequencing (NGS) technologies. There have been numerous applications and successes of NGS applications in the clinical and public health domain. However, none have, as yet, delivered perhaps the most sought after application, i.e., the generation of microbial sequence data for "real-time" patient management. In this review, we discuss the use of NGS and whole-genome sequencing (WGS) of microorganisms as a logical next step for the routine diagnosis of infection and the prediction of antimicrobial susceptibility in the clinical microbiology laboratory.

Review Editor: Alex van Belkum, no conflict of interest.

W. M. Dunne Jr (✉)
bioMérieux, Inc.,
100 Rodolphe Street,
Durham, NC 27712, USA
e-mail: william.dunne@biomerieux.com

W. M. Dunne Jr · L. F. Westblade · B. Ford
Department of Pathology and Immunology, Washington
University School of Medicine,
St. Louis, MO, USA

W. M. Dunne Jr
Department of Pediatrics, Duke University School of Medicine,
Durham, NC, USA

## Introduction

The detection of the antimicrobial resistance and/or identification of microorganisms by molecular means is nothing new to the field of clinical microbiology and continues to evolve, albeit less rapidly than some might have predicted [1]. As an example of this evolution, a protocol for the detection of methicillin resistance among staphylococci using a polymerase chain reaction (PCR) assay combined with agarose gel electrophoresis of amplified nucleic acid product was reported in 1991 by Murakami et al. [2]. In this study, the absence of an amplifiable *mecA* gene product among *Staphylococcus aureus* isolates had 100 % correlation with methicillin susceptibility by phenotypic testing. The clinical utility of this approach soon became obvious and was transformed 10 years later into a rapid "real-time" PCR assay for the simultaneous determination of *mecA* status and *S. aureus* species identification [3]. A commercial version of this genotypic screening approach was cleared by the U.S. Food and Drug Administration (FDA) shortly thereafter [4] and within a second 10-year period, eight automated or semi-automated assays were released for the detection of methicillin-resistant *S. aureus* (MRSA) colonization of at-risk patients and/or direct identification from positive blood cultures ([5], plus NucliSENS EasyQ® MRSA in May of 2011). Despite improvements in processing and speed, the information generated by some of these targeted molecular assays was, at times, misleading. Desjardins et al. [5] noticed a high percentage of false-positive results after the implementation of one MRSA screening assay that were likely due to the "kick-out" of the *mecA* gene, but the retention of the target amplification site at the *orfX* gene–staphylococcal cassette chromosome *mec* (SCC*mec*) element intersection. An additional demonstration of the limitation of such targeted molecular testing was

provided by Shore et al. [6], who identified two clonal complex (CC) 130 MRSA isolates that were falsely negative by an MRSA screening assay, owing to a novel SCC*mec* XI element with a highly divergent *mecA* gene. These authors utilized whole-genome sequencing (WGS) to elucidate the root cause of the false-negative results.

## Next-generation sequencing

Similar to the progression of single-target or multiplex amplification assays described above, the technological wherewithal to sequence large stretches of DNA/RNA and to patch those sequences together into complete genomes or transcriptomes of prokaryotic and eukaryotic organisms has advanced rapidly following the publication of the drafts of the first microbial [7] and human genomes [8]. The term "next-generation sequencing" (NGS) refers to those strategies that have supplemented or supplanted the Sanger dideoxy chain termination sequencing method used nearly exclusively through 2004 [9, 10]. While a detailed review of the technology that comprises NGS methods is beyond the scope of this review, a comparison of the salient characteristics that differ between methods is shown in Table 1. In addition, several excellent reviews of this topic are provided within and are highly recommended [10–13].

The massive sequence output, cost per base, size of microbial genomes, and the ability to generate large quantities of microbial DNA/RNA starting material makes NGS an attractive option to current single-target or multiplex amplification methods for the detection of multiple resistance determinants, virulence factors, or epidemiological markers in a single sequencing run. But is this the proverbial use of an elephant gun to kill a mouse?

## Potential for the routine use of NGS technologies in the clinical microbiology laboratory

If we assume that NGS will become increasingly affordable, rapid, and simple to use (i.e., DNA/RNA in; aligned, assembled, annotated, and interpreted genome/transcriptome out), and that technologies and databases will evolve to the extent that WGS will be highly reproducible and reliable, how could this technology be put to its best use in the clinical microbiology laboratory? Before we speculate, let's try to estimate the cost of WGS in the current market. Bacterial genomes range in size from 0.5 to 10 Mb, but the genomes that most interest clinical microbiologists at the moment (e.g., common human pathogens such as *Escherichia coli*, *Pseudomonas aeruginosa*, and *S. aureus*) reside somewhere in the 2–5-Mb range [14]. It has been estimated that the cost of sequencing using second-generation technology ranges from $1 to $60/Mb [9–13]. Using these figures, one could estimate the cost of sequencing a single bacterial genome to range from $2 to $300, depending on the technology employed. This does not include capital outlay (which, ranging from $100 K to $1,350 K, is not trivial) and overhead costs including labor, whereas service agreements vary greatly. In reality, using second-generation sequencing instrumentation and 100× coverage of a 4-Mb-sized genome, the output of an eight-lane, two flow cell Illumina HiSeq2000 instrument is around 1,600 genomes over a 10-day period at a cost of around $25.00 per genome; however, that is only the sequencing costs. When the sample preparation steps are added, e.g., DNA/RNA extraction, fragment preparation, library construction, ligation of barcoded adapters to allow for the pooling of DNA fragments, multiplex sequencing, repeat runs, sequence quality assessment, deconvoluting, eliminating bad sequences, and construction of high-quality genomes (which might require re-

**Table 1** Macroscopic comparison of next-generation technologies (data extrapolated from [9–12, 68])

| Generation[a] | Chemistry | Platform | Throughput per run (bases) | Error rate (expressed as a Phred score) | Read length (bp) |
|---|---|---|---|---|---|
| Second | Pyrosequencing | Life Sciences 454, e.g., GS FLX Titanium XLR70 | 450 Mb | Q30[b] | 450 |
| Second | Dye termination/synthesis | Illumina/Solexa, e.g., HiSeq 2500 | 120,000 Mb | Q15[b] | 2×150 paired |
| Second | Ligation | AB SOLiD, e.g., 5500xl | 20,000 Mb | Q27[b] | 50 |
| Third | Semiconductor | Ion Torrent, e.g., Ion PGM | 10–100 Mb | Q15–Q20 est. | 100 |
| Third | Direct detection | Pacific Biosciences RS | 75 Mb, estimate | Unknown | 1,500 est. |

[a] With "first generation" single-chain sequencing (Sanger dideoxy chain termination sequencing) as reference, "second generation" sequencing analyzes an ensemble of DNA molecules simultaneously by wash and scan techniques, and "third generation" sequencing interrogates molecules without the need to halt between read steps [12]

[b] Phred score, $Q = -10 \log_{10}(P(\geq n/s))$, where $s$ is the observed signal and $n$ is the length of the homopolymer that produced the signal [69, 70], serves here as a common-denominator indicator of sequence quality that is familiar from low-throughput dideoxy or Sanger sequencing. A score of 20 is equivalent to an error rate of 1 in 100 bases. Scores for the Roche, Illumina, and Applied Biosystems platforms can, therefore, be estimated from data published by Suzuki et al. [68], but this estimate should be considered as approximate

sequencing or complicated gap-closing procedures), the estimate is in the range of $200 to $400 per genome (George Weinstock, Washington University School of Medicine, St. Louis, MO, USA, personal communication). The cost will undoubtedly continue to drop due to improved techniques and competition among sequencer manufacturers, so it is reasonable to assume that, in the very near future, WGS of microorganisms will be affordable and reasonably rapid. Now, back to the original question: how can we put this technology to work for improved patient care and clinical outcomes?

## Virtual resistance testing

One of the prime uses of WGS would undoubtedly include the virtual resistance testing of bacteria and viruses [15], both on primary patient encounters and throughout the duration of antimicrobial therapy. As output and cost continue to drop, fungi and parasites of clinical importance might be added to the list. The clinical interpretation of potential antimicrobial resistance secondary to the identification of fully characterized resistance-associated sequences (point mutations, indels, or putative open reading frames [ORFs]) would need to be extremely conservative at first. For example, an organism would be reported as "potentially resistant" to a specific antimicrobial agent based solely on a "hit" for a verified resistance sequence. The beauty of this lies in the ability to distinguish among hundreds of different antibiotic-modifying enzymes (e.g., β-lactamases or aminoglycoside-modifying enzymes), each of which might have subtle or distinct substrate profiles that could be used to fine-tune appropriate antimicrobial therapy [16]. Obviously, this approach in the absence of any phenotypic support for the functionality of a resistance gene or mutation could lead to major errors (reported as resistant, actually susceptible) and preclude the use of potentially useful therapies, as would be the case with *mecA*-positive strains of *S. aureus* that are susceptible to oxacillin secondary to mutations in the *fem* gene family [17]. Conversely, and perhaps more importantly, as it would lead to the generation of very major errors, the absence of an identifiable gene or mutation sequence does not guarantee susceptibility. This potential failure stems from NGS' dependence on annotation and interpretation using prior knowledge about how genetic sequences translate to resistance phenotypes. Indeed, novel resistance mechanisms or combinatorial factors that require two or more separate genetic alterations for the expression of resistance could be very difficult to predict a priori by sequence alone without performing comparisons of an isolate's genome "before" and "after" therapy. Such a comparative WGS approach was used to identify the resistance determinants associated with a strain of *Acinetobacter baumannii* that had acquired tigecycline resistance during therapy [18]. The authors determined that resistance was associated with 18 single nucleotide polymorphisms (SNPs) and three deletions between susceptible and resistant strains. Therefore, the reliability of WGS as a means of predicting antimicrobial susceptibility is critically dependent upon the availability of a current and curated database of reference sequences, e.g., http://ardb.cbcb.umd.edu/index.html and http://img.jgi.doe.gov (see also [19]). While it would be desirable to have databases such as these openly available, it is likely that such sequence databanks will require a licensing or user fee or perhaps be accompanied by promotional advertisements! Until the odds of accurately predicting antimicrobial resistance based upon the identification of a species-specific sequence are known, confirmation using conventional phenotypic antimicrobial susceptibility testing (AST) would be required. In fact, it is likely that phenotypic methods will continue to be used, at least for the foreseeable future, to screen microbial isolates for unrecognized resistance patterns, and, thus, their mechanisms of resistance, before gene-level inquiry is pursued.

Alternatively, the need for phenotypic verification of the genotype might be partly circumvented by transcriptome analysis [15], thus, providing direct evidence of functional resistance rather than using gene identification as a proxy. An example of this approach was provided by Feng et al. [20] for linezolid resistance in *Streptococcus pneumoniae*, a phenotype which required the overexpression of proteins and enzymes involved in sugar metabolism rather than a defined resistance mutation or genetic locus. For second-generation sequencing systems, this would require the conversion of mRNA to cDNA with amplification prior to sequencing, while third-generation systems hold the promise of direct sequencing of isolated mRNA [9–13]. A novel protocol for the isolation and enrichment of bacterial mRNA from total cellular RNA, including the eukaryotic ribosomal fraction, for this very purpose has been described [21]. Transcriptome analysis would also require exposure of an organism to subinhibitory concentrations of an antimicrobial agent and the subsequent sequencing of mRNA transcripts to identify upregulated resistance determinants and/or SOS genes to predict or determine drug susceptibility. One possible pitfall of using SOS transcripts as a congener of drug susceptibility would be the presence of small heteroresistant subpopulations that would be undetected by this strategy. However, heteroresistant populations could be selected upon prolonged antibiotic exposure with conventional AST.

## Taxonomy and epidemiology

The routine use of WGS would likely open a whole new door in terms of taxonomic classification and identification

of novel bacterial species and subspecies. In addition to resistance prediction, the expanding collection of completely sequenced genomes could provide for the epidemiological typing of microorganisms at the ultimate level of resolution—essentially generating whole-genome SNP analysis [15, 22]. The use of WGS for epidemiological purposes has been demonstrated with clinical isolates of multidrug-resistant strains of *A. baumannii* that could not be differentiated using standard epidemiologic tools, such as pulsed-field gel electrophoresis and variable number tandem repeat analyses [18]. Further, WGS would also permit previously characterized virulence genes to be detected and identified simultaneously. The expanded use of WGS for microorganisms has the potential to produce the equivalent of a HapMap for microbes (http://hapmap.ncbi.nlm.nih.gov/). When combined with data generated by the human HapMap, the risk for serious sequelae when individuals of a particular haplotype become infected by specific strains of an organism carrying characterized virulence factors could begin to be assessed. This is akin to the scenario described for *Helicobacter pylori* virulence genes *cagA*, *vacA*, and *babA2* versus human cytokine polymorphisms when evaluating the risk for the development of precancerous gastric lesions [23]. Another promising application of NGS would be in urinary tract infection (UTI), where genetically well-determined bacterial virulence factors define host recognition and response during infection [24], the probability of recurrent disease [25], and progression of uncomplicated UTI to bacteremia [26]. Simultaneous genotyping of host and microbe in this case could potentially define treatment tailored to the patient's potential for severe and recurrent disease versus the likelihood of experiencing a single episode of uncomplicated cystitis.

Obviously, much of this technology would have broad application in the public health domain, providing that funding is maintained at a level equivalent to that in the private sector.

The availability of WGS, obtained using NGS technologies, as a means of supplanting all existing forms of molecular epidemiology typing methods and, at the same time, providing real-time, longitudinal analysis of outbreaks in progress on any scale has been referred to as "Public Health 2.0" by Pallen and Loman [27]. Prominent examples of such "Public Health 2.0" endeavors include the elucidation of the origin of the *Vibrio cholerae* isolate that devastated Haiti after the 2010 earthquake [28], where it was revealed that the epidemic isolate was closely related to isolates from Asia rather than circulating South American isolates and likely introduced by human activity. The sequence of the *E. coli* O104:H4 isolate, the etiologic agent of the 2011 foodborne outbreak in Germany, which saw the release of the organism's genome before a contaminated food source was appreciated, provides another example [29]. This

analysis rapidly revealed, via a reductionist multilocus sequence typing approach that examined only seven common housekeeping genes, that the outbreak strain was a hybrid of enteroaggregative and enterohemorrhagic *E. coli* that had acquired AAF/I fimbriae and the CTX-M-15 β-lactamase relative to a non-outbreak strain from 2001. A more in-depth analysis resulted from "crowd sourcing" the analysis of the genome to the public, dubbed the "*E. coli* O104:H4 Genome Analysis Crowd-Sourcing Consortium", which has now publicly annotated and commented on the genomes of ten outbreak strains [30, 31].

But, as Pallen and Loman wisely point out, universal use of WGS in the public health arena will require a level of standardization and consistency that has not yet been attained, as well as better control of a host of other variables, such as the recognition of sequence differences among heterogeneous sub-populations within outbreak-associated organisms that could generate confounding epidemiological findings (e.g., isoniazid or rifampin heteroresistance among isolates of *Mycobacterium tuberculosis*) [27]. And while it is clear that WGS will not be a panacea for all of the nuances facing public health microbiologists, it may well be that NGS may find an initial home in public health microbiology laboratories, which have more centralized resources and a keener interest in epidemiology, where it will serve as another weapon in the arsenal that may advance the productivity of public health diagnostic laboratories beyond their current means.

Of course, risk evaluation is not limited to bacterial infection. Barzon et al. [32] have developed a method for the direct identification of human papillomavirus (HPV) genotypes in cervical samples using NGS for both single- and multiple-genotype infections. NGS provided a sensitivity of up to 100 genome equivalents/μL of cytology sample for high-risk strains HPV16 and 18, and had a sensitivity of 100 % for the detection of HPV DNA in cervical samples compared to conventional PCR using consensus primers. Further, multiple infections could be identified when present at 1 % of the total genome equivalents. Compared to Sanger sequencing, 454 NGS was 100 % sensitive and specific for the detection of HPV in clinical samples and identified more samples with mixed HPV infection. By contrast, the agreement between NGS and the INNO-LiPA HPV Genotyping Extra kit (Innogenetics) was genotype-dependent [32].

## Microbiome analysis

One of the most powerful applications of NGS has been its role in understanding the complex biodiversity of the human microbiome: the plethora of microorganisms colonizing every "nook and cranny" of the human body. While we are only beginning to appreciate what constitutes the normal or

abnormal microbial constituents of the gut [33–38], lower and upper respiratory tract [39, 40], vagina [41–44], skin [45–48], urinary tract [49–51], gingiva [52, 53], and wounds [54, 55], it is not difficult to envision that human microbiome analysis as a means of evaluating and stratifying different disease processes could become part of mainstream clinical management. Indeed, gut microbiome analysis already shows promise in classifying disease processes such as inflammatory bowel disease into more specific entities, such as Crohn's disease or idiopathic bowel syndrome [56–59], and as more detailed information is filtered from so-called "altered microbiome" studies, it should be possible to winnow disease-associated changes down to one or two highly predictive species. Indeed, the "alpha-bug" theory postulates that there is an association between the development of colon cancer in mice, and potentially in humans, to gut over-colonization with and exposure to enterotoxigenic *Bacteroides fragilis* [60]. A further promise of human microbiome studies are their ability to register the microbial diversity of the human microbiome during and in response to clinical therapy, so that the resultant information can be used to both predict prognosis and alter unfavorable clinical outcomes. In a seminal study, Ubeda et al. assessed the potential of cataloging the human gut microbiome by NGS in the clinical management of allogeneic hematopoietic stem cell transplant (allo-HSCT) patients. The authors observed that post-antimicrobial overpopulation of the intestinal microbiome with vancomycin-resistant enterococci (VRE) rendered this patient population at higher risk of developing bacterial sepsis compared to allo-HSCT patients whose intestinal microbiome was not enriched with VRE. Thus, at least in this highly fragile patient population, the definition of the composition of the intestinal microbiome affords some clinical utility, especially as these patients might serve as candidates for the administration of probiotic microbes in an attempt to prevent bacterial sepsis [61].

## Challenges

Every new technology and its implementation into a specific environment is accompanied by a plethora of challenges, and this will certainly be true for the inclusion of NGS technologies in the clinical microbiology laboratory. A notable challenge is the development of highly reproducible NGS sequencing technology/software platforms that can provide clinically actionable reports to physicians at a pace equal to, if not vastly greater than, the culture-based methods used at present. This, coupled with the construction and maintenance of readily accessible databases of reference microbial genomes up-to-date with contemporary and historical resistance determinants (e.g., discrete genes, indels, SNPs, and promoter/regulator mutations), is critical to the

success of NGS technologies in the clinical microbiological laboratory. Also essential to the reproducible identification of the genetic elements and mutations listed above are the resultant sequence "read lengths". While monolithic "read lengths" are reported by manufacturers, this is typically a rule-of-thumb truncation of longer individual reads that are low quality on the their distal ends but may, nevertheless, be "piled" in non-truncated form to increase the likelihood that a contiguous sequence is mapped correctly. In cases where only short reads are obtained, the sequencing of clinical isolates using reference genomes as templates would be complicated by difficulties calling SNPs, indels, and copy number variation [62].

An additional, yet not to be overlooked, challenge is the ability of laboratory directors [63] and personnel to collect, manage, analyze, and interpret the colossal amounts of bioinformatic data that will be generated by NGS runs. The treatment of such data may require skills that are well beyond those of the current workforce. Thus, as a profession, we as clinical microbiologists must ensure that the emerging workforce are endowed with the skills to face the challenges of NGS in the clinical microbiology laboratory, and that the proficiency of this workforce is maintained.

On the regulatory level, the clinical validation of molecular assays [64] is a relatively common activity, however, the mechanism(s) by which to validate clinical NGS-based assays is as yet undefined, particularly as concepts such as limit of detection and normal ranges are difficult to define for microbiota. Thus, as with training of the workforce, we as a profession must ensure that we play an active role in defining the mechanism(s) by which NGS-based assays may become validated and verified for clinical use. Any working assay must be validated on many levels: sample and library preparation, sequence generation and subsequent automated annotation and interpretation (which are highly dependent on the platform and analysis pipeline used [Table 1]), and distillation of the data into a reportable result that carries enough weight to influence clinical outcomes. Initially, the trading of split samples between institutions with different approaches to similar tests may serve as sufficient proficiency testing but, in the long term, national and international regulatory bodies will need to set clear standards.

Finally, classical indicators of a disordered microbiota such as "clue cells", gingivitis, and diarrhea may be more familiar to physicians and deemed more worthy paying for over reports generated by WGS analyses. This problem may be mitigated by the existence of problems where a great deal of sequence information can be distilled into a well-established risk analysis from the microbial side (e.g., HPV [32, 65] or hepatitis C virus [66] typing) or the perspective of the host (e.g., carrier testing for an array of diseases such as cystic fibrosis at birth [67]).

## Summary

As the cost and complexity of sequencing platforms declines and access to curated databases containing validated sets of resistance and virulence markers become readily available, the routine use of NGS/WGS technology in the diagnostic clinical microbiology laboratory will, in our opinion, become increasingly commonplace. We also strongly believe that sequence-based diagnostics will not likely replace conventional microbiological methods, including growth-based assays, any time soon. There will be a continued need to validate novel resistance mechanisms versus standardized in vitro AST, characterize previously unrecognized agents by microscopic and colonial morphology and metabolic profiles, establish animal models of infection, and establish parallel, low-cost methods of identification. All of these alternatives will work in a supplementary fashion to provide a blend of complexity suited to the needs of individual laboratories. Regardless of the mix of technologies used in the future, it is likely that sequence-based diagnostics will be a major part of the blend. Clearly, it is an exciting time for clinical microbiology.

## References

1. Dunne WM Jr, Pinckard JK, Hooper LV (2002) Clinical microbiology in the year 2025. J Clin Microbiol 40:3889–3893
2. Murakami K, Minamide W, Wada K, Nakamura E, Teraoka H, Watanabe S (1991) Identification of methicillin-resistant strains of staphylococci by polymerase chain reaction. J Clin Microbiol 29:2240–2244
3. Reischl U, Linde HJ, Metz M, Leppmeier B, Lehn N (2000) Rapid identification of methicillin-resistant Staphylococcus aureus and simultaneous species confirmation using real-time fluorescence PCR. J Clin Microbiol 38:2429–2433
4. Warren DK, Liao RS, Merz LR, Eveland M, Dunne WM Jr (2004) Detection of methicillin-resistant Staphylococcus aureus directly from nasal swab specimens by a real-time PCR assay. J Clin Microbiol 42:5578–5581
5. Desjardins M, Guibord C, Lalonde B, Toye B, Ramotar K (2006) Evaluation of the IDI-MRSA assay for detection of methicillin-resistant Staphylococcus aureus from nasal and rectal specimens pooled in a selective broth. J Clin Microbiol 44:1219–1223
6. Shore AC, Deasy EC, Slickers P, Brennan G, O'Connell B, Monecke S, Ehricht R, Coleman DC (2011) Detection of staphylococcal cassette chromosome mec type XI carrying highly divergent mecA, mecI, mecR1, blaZ, and ccr genes in human clinical isolates of clonal complex 130 methicillin-resistant Staphylococcus aureus. Antimicrob Agents Chemother 55:3765–3773
7. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb J-F, Dougherty BA, Merrick JM, McKenney K, Sutton G, Fitzhugh W, Fields C, Gocyne JD, Scott J, Shirley R, Liu L-I, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Hedblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrmann JL, Geoghagen NSM, Gnehm CL, McDonald LA, Small KV, Fraser CM, Smith HO, Venter JC (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 269:496–512
8. McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M, Wylie K, Mardis ER, Wilson RK, Fulton R, Kucaba TA, Wagner-McPherson C, Barbazuk WB, Gregory SG, Humphray SJ, French L, Evans RS, Bethel G, Whittaker A, Holden JL, McCann OT, Dunham A, Soderlund C, Scott CE, Bentley DR, Schuler G, Chen HC, Jang W, Green ED, Idol JR, Maduro VV, Montgomery KT, Lee E, Miller A, Emerling S, Kucherlapati, Gibbs R, Scherer S, Gorrell JH, Sodergren E, Clerc-Blankenburg K, Tabor P, Naylor S, Garcia D, de Jong PJ, Catanese JJ, Nowak N, Osoegawa K, Qin S, Rowen L, Madan A, Dors M, Hood L, Trask B, Friedman C, Massa H, Cheung VG, Kirsch IR, Reid T, Yonescu R, Weissenbach J, Bruls T, Heilig R, Branscomb E, Olsen A, Doggett N, Cheng JF, Hawkins T, Myers RM, Shang J, Ramirez L, Schmutz J, Velasquez O, Dixon K, Stone NE, Cox DR, Haussler D, Kent WJ, Furey T, Rogic S, Kennedy S, Jones S, Rosenthal A, Wen G, Schilhabel M, Gloeckner G, Nyakatura G, Siebert R, Schlegelberger B, Korenberg J, Chen XN, Fujiyama A, Hattori M, Toyoda A, Yada T, Park HS, Sakaki Y, Shimizu N, Asakawa S, Kawasaki K, Sasaki T, Shintani A, Shimizu A, Shibuya K, Kudoh J, Minoshima S, Ramser J, Seranski P, Hoff C, Poustka A, Reinhardt R, Lehrach H (2001) A physical map of the human genome. Nature 409:934–941
9. Mardis ER (2011) A decade's perspective on DNA sequencing technology. Nature 470:198–203
10. Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387–402
11. Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26:1135–1145
12. Schadt EE, Turner S, Kasarskis A (2010) A window into third-generation sequencing. Hum Mol Genet 19:R227–R240
13. Su Z, Ning B, Fang H, Hong H, Perkins R, Tong W, Shi L (2011) Next-generation sequencing and its applications in molecular diagnostics. Expert Rev Mol Diagn 11:333–343
14. Ochman H, Davalos LM (2006) The nature and dynamics of bacterial genomes. Science 311:1730–1733
15. Pallen MJ, Loman NJ, Penn CW (2010) High-throughput sequencing and clinical microbiology: progress, opportunities and challenges. Curr Opin Microbiol 13:625–631
16. Ledeboer NA, Hodinka RL (2011) Molecular detection of resistance determinants. J Clin Microbiol 49:S20–S24
17. Giannouli S, Labrou M, Kyritsis A, Ikonomidis A, Pournaras S, Stathopoulos C, Tsakris A (2010) Detection of mutations in the FemXAB protein family in oxacillin-susceptible mecA-positive Staphylococcus aureus clinical isolates. J Antimicrob Chemother 65:626–633
18. Hornsey M, Loman N, Wareham DW, Ellington MJ, Pallen MJ, Turton JF, Underwood A, Gaulton T, Thomas CP, Doumith M, Livermore DM, Woodford N (2011) Whole-genome comparison of two Acinetobacter baumannii isolates from a single patient, where resistance developed during tigecycline therapy. J Antimicrob Chemother 66:1499–1503
19. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC (2012) IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res 40:D115–D122
20. Feng J, Billal DS, Lupien A, Racine G, Winstall E, Légaré D, Leprohon P, Ouellette M (2011) Proteomic and transcriptomic analysis of linezolid resistance in Streptococcus pneumoniae. J Proteome Res 10:4439–4452

21. Chen Z, Duan X (2011) Ribosomal RNA depletion for massively parallel bacterial RNA-sequencing applications. Methods Mol Biol 733:93–103

22. Ward TJ, Ducey TF, Usgaard T, Dunn KA, Bielawski JP (2008) Multilocus genotyping assays for single nucleotide polymorphism-based subtyping of Listeria monocytogenes isolates. Appl Environ Microbiol 74:7629–7642

23. Rad R, Dossumbekova A, Neu B, Lang R, Bauer S, Saur D, Gerhard M, Prinz C (2004) Cytokine gene polymorphisms influence mucosal cytokine expression, gastric inflammation, and host specific colonisation during Helicobacter pylori infection. Gut 53:1082–1089

24. Johnson JR (2003) Microbial virulence determinants and the pathogenesis of urinary tract infection. Infect Dis Clin North Am 17:261–278, viii

25. Foxman B, Zhang L, Tallman P, Palin K, Rode C, Bloch C, Gillespie B, Marrs CF (1995) Virulence characteristics of Escherichia coli causing first urinary tract infection predict risk of second infection. J Infect Dis 172:1536–1541

26. Marschall J, Zhang L, Foxman B, Warren DK, Henderson JP (2012) Both host and pathogen factors predispose to Escherichia coli urinary-source bacteremia in hospitalized patients. Clin Infect Dis (in press)

27. Pallen MJ, Loman NJ (2011) Are diagnostic and public health bacteriology ready to become branches of genomic medicine? Genome Med 3:53

28. Chin C-S, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, Bullard J, Webster DR, Kasarskis A, Peluso P, Paxinos EE, Yamaichi Y, Calderwood SB, Mekalanos JJ, Schadt EE, Waldor MK (2011) The origin of the Haitian cholera outbreak strain. N Engl J Med 364:33–42

29. Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Karch H (2011) Prospective genomic characterization of the German enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology. PLoS One 6:e22751

30. Rohde H, Qin J, Cui Y, Li D, Loman NJ, Hentschke M, Chen W, Pu F, Peng Y, Li J, Xi F, Li S, Li Y, Zhang Z, Yang X, Zhao M, Wang P, Guan Y, Cen Z, Zhao X, Christner M, Kobbe R, Loos S, Oh J, Yang L, Danchin A, Gao GF, Song Y, Li Y, Yang H, Wang J, Xu J, Pallen MJ, Wang J, Aepfelbacher M, Yang R (2011) Open-source genomic analysis of Shiga-toxin-producing E. coli O104: H4. N Engl J Med 365:718–724

31. E. coli O104:H4 Genome Analysis Crowdsourcing. GitHub. https://github.com/ehec-outbreak-crowdsourced/BGI-data-analysis/wiki

32. Barzon L, Militello V, Lavezzo E, Franchin E, Peta E, Squarzon L, Trevisan M, Pagni S, Dal Bello F, Toppo S, Palù G (2011) Human papillomavirus genotyping by 454 next generation sequencing technology. J Clin Virol 52:93–97

33. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA (2005) Diversity of the human intestinal microbial flora. Science 308:1635–1638

34. Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A, Takami H, Morita H, Sharma VK, Srivastava TP, Taylor TD, Noguchi H, Mori H, Ogura Y, Ehrlich DS, Itoh K, Takagi T, Sakaki Y, Hayashi T, Hattori M (2007) Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. DNA Res 14:169–181

35. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI (2008) A core gut microbiome in obese and lean twins. Nature 457:480–484

36. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI (2009) The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci Transl Med 1:6ra14

37. Turnbaugh PJ, Quince C, Faith JJ, McHardy AC, Yatsunenko T, Niazi F, Affourtit J, Egholm M, Henrissat B, Knight R, Gordon JI (2010) Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. Proc Natl Acad Sci USA 107:7503–7508

38. Greenblum S, Turnbaugh PJ, Borenstein E (2012) Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. Proc Natl Acad Sci USA 109:594–599

39. Nakamura S, Yang C-S, Sakon N, Ueda M, Tougan T, Yamashita A, Goto N, Takahashi K, Yasunaga T, Ikuta K, Mizutani T, Okamoto Y, Tagami M, Morita R, Maeda N, Kawai J, Hayashizaki Y, Nagai Y, Horii T, Iida T, Nakaya T (2009) Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PLoS One 4:e4219

40. Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J, Tammadoni S, Nosrat B, Conrad D, Rohwer F (2009) Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. PLoS One 4: e7370

41. Hyman RW, Herndon CN, Jiang H, Palm C, Fukushima M, Bernstein D, Vo KC, Zelenko Z, Davis RW, Giudice LC (2012) The dynamics of the vaginal microbiome during infertility therapy with in vitro fertilization-embryo transfer. J Assist Reprod Genet 29:105–115

42. Oakley BB, Fiedler TL, Marrazzo JM, Fredricks DN (2008) Diversity of human vaginal bacterial communities and associations with clinically defined bacterial vaginosis. Appl Environ Microbiol 74:4898–4909

43. Fredricks DN, Fiedler TL, Marrazzo JM (2005) Molecular identification of bacteria associated with bacterial vaginosis. N Engl J Med 353:1899–1911

44. Hummelen R, Fernandes AD, Macklaim JM, Dickson RJ, Changalucha J, Gloor GB, Reid G (2010) Deep sequencing of the vaginal microbiota of women with HIV. PLoS One 5: e12078

45. Fahlén A, Engstrand L, Baker BS, Powles A, Fry L (2012) Comparison of bacterial microbiota in skin biopsies from normal and psoriatic skin. Arch Dermatol Res 304:15–22

46. Kong HH, Oh J, Deming C, Conlan S, Grice EA, Beatson MA, Nomicos E, Polley EC, Komarow HD; NISC Comparative Sequence Program, Murray PR, Turner ML, Segre JA (2012) Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. Genome Res 22:850–859

47. Capone KA, Dowd SE, Stamatas GN, Nikolovski J (2011) Diversity of the human skin microbiome early in life. J Invest Dermatol 131:2026–2032

48. Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC; NISC Comparative Sequencing Program, Bouffard GG, Blakesley RW, Murray PR, Green ED, Turner ML, Segre JA (2009) Topographical and temporal diversity of the human skin microbiome. Science 324:1190–1192

49. Wolfe AJ, Toh E, Shibata N, Rong R, Kenton K, Fitzgerald M, Mueller ER, Schreckenberger P, Dong Q, Nelson DE, Brubaker L (2012) Evidence of uncultivated bacteria in the adult female bladder. J Clin Microbiol 50:1376–1383

50. Nelson DE, Van Der Pol B, Dong Q, Revanna KV, Fan B, Easwaran S, Sodergren E, Weinstock GM, Diao L, Fortenberry JD (2010) Characteristic male urine microbiomes associate with asymptomatic sexually transmitted infection. PLoS One 5:e14116

51. Dong Q, Nelson DE, Toh E, Diao L, Gao X, Fortenberry JD, Van Der Pol B (2011) The microbial communities in male first catch urine are highly similar to those in paired urethral swab specimens. PLoS One 6:e19709

52. Lazarevic V, Whiteson K, Huse S, Hernandez D, Farinelli L, Østerås M, Schrenzel J, François P (2009) Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. J Microbiol Methods 79:266–271

53. Lazarevic V, Whiteson K, Hernandez D, François P, Schrenzel J (2010) Study of inter- and intra-individual variations in the salivary microbiota. BMC Genomics 11:523

54. Wolcott RD, Gontcharova V, Sun Y, Dowd SE (2009) Evaluation of the bacterial diversity among and within individual venous leg ulcers using bacterial tag-encoded FLX and titanium amplicon pyrosequencing and metagenomic approaches. BMC Microbiol 9:226

55. Dowd SE, Sun Y, Secor PR, Rhoads DD, Wolcott BM, James GA, Wolcott RD (2008) Survey of bacterial diversity in chronic wounds using pyrosequencing, DGGE, and full ribosome shotgun sequencing. BMC Microbiol 8:43

56. Pflughoeft KJ, Versalovic J (2012) Human microbiome in health and disease. Annu Rev Pathol 7:99–122

57. Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, Nalin R, Jarrin C, Chardon P, Marteau P, Roca J, Dore J (2006) Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. Gut 55:205–211

58. Manges AR, Labbe A, Loo VG, Atherton JK, Behr MA, Masson L, Tellis PA, Brousseau R (2010) Comparative metagenomic study of alterations to the intestinal microbiota and risk of nosocomial Clostridum difficile-associated disease. J Infect Dis 202:1877–1884

59. Turnbaugh PJ, Bäckhed F, Fulton L, Gordon JI (2008) Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host Microbe 3:213–223

60. Sears CL, Pardoll DM (2011) Perspective: alpha-bugs, their microbial partners, and the link to colon cancer. J Infect Dis 203:306–311

61. Ubeda C, Taur Y, Jenq RR, Equinda MJ, Son T, Samstein M, Viale A, Socci ND, van den Brink MRM, Kamboj M, Pamer EG (2010) Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. J Clin Invest 120:4332–4341

62. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498

63. Thomson RB Jr, Doern GV (2011) What will the role of the clinical microbiology laboratory director be in 2015? J Clin Microbiol 49:S68–S71

64. Jennings L, Van Deerlin VM, Gulley ML (2009) Recommended principles and practices for validating clinical molecular pathology tests. Arch Pathol Lab Med 133:743–755

65. Kong CS, Narasimhan B, Cao H, Kwok S, Erickson JP, Koong A, Pourmand N, Le Q-T (2009) The relationship between human papillomavirus status and other molecular prognostic markers in head and neck squamous cell carcinomas. Int J Radiat Oncol Biol Phys 74:553–561

66. Elahi E, Pourmand N, Chaung R, Rofoogaran A, Boisver J, Samimi-Rad K, Davis RW, Ronaghi M (2003) Determination of hepatitis C virus genotype by pyrosequencing. J Virol Methods 109:171–176

67. Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, Langley RJ, Zhang L, Lee CC, Schilkey FD, Sheth V, Woodward JE, Peckham HE, Schroth GP, Kim RW, Kingsmore SF (2011) Carrier testing for severe childhood recessive diseases by next-generation sequencing. Sci Transl Med 3:65ra4

68. Suzuki S, Ono N, Furusawa C, Ying B-W, Yomo T (2011) Comparison of sequence reads obtained from three next-generation sequencing platforms. PLoS One 6:e19534

69. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175–185

70. Richterich P (1998) Estimation of errors in "raw" DNA sequences: a validation study. Genome Res 8:251–259