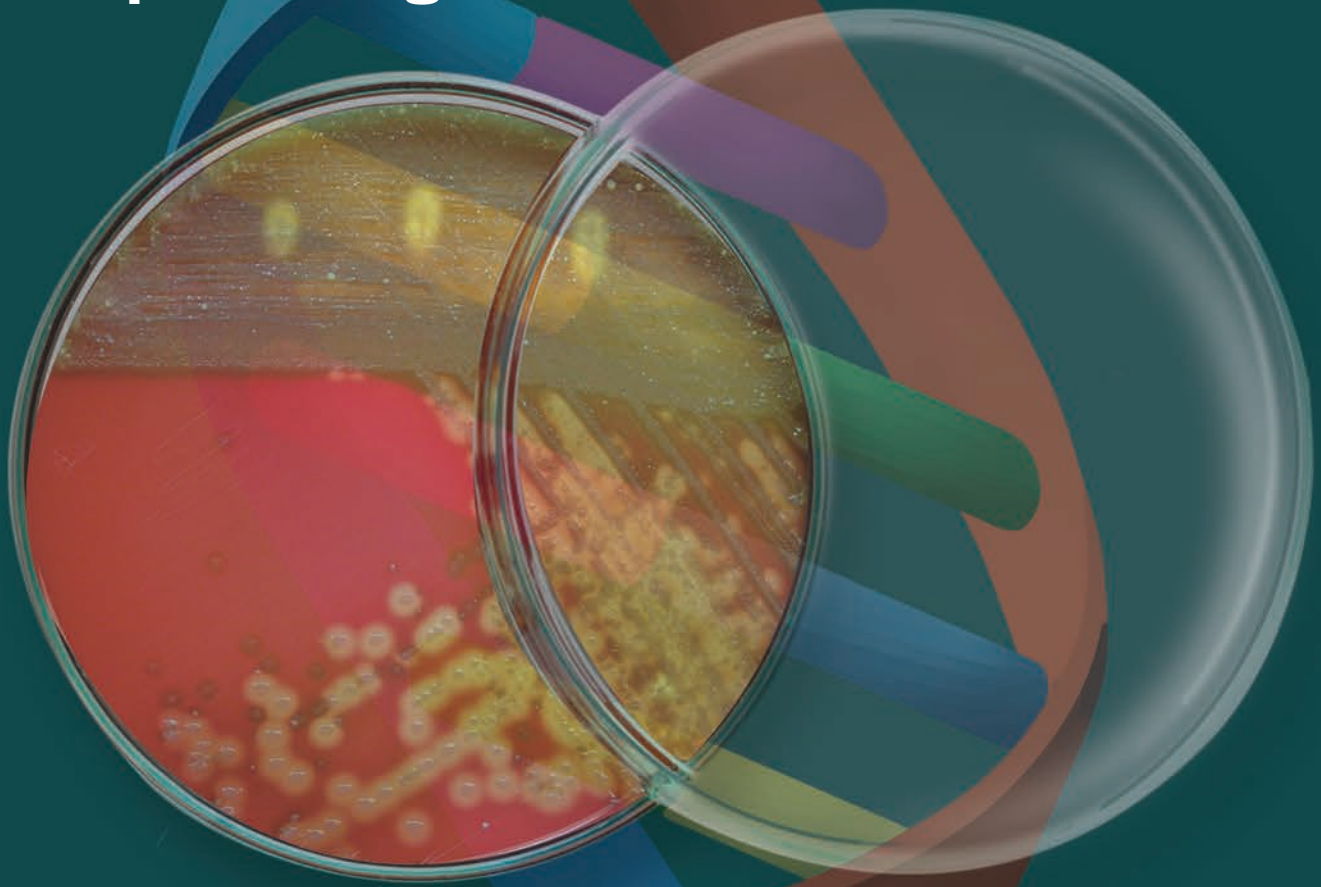
Applications of Clinical Microbial Next-Generation Sequencing



AMERICAN SOCIETY FOR MICROBIOLOGY

A report from the

AMERICAN ACADEMY OF MICROBIOLOGY

RECOGNIZING SCIENTIFIC EXCELLENCE

Applications of Clinical Microbial Next-Generation Sequencing

Report on an American Academy of Microbiology Colloquium held in Washington, DC, in April 2015.

The American Academy of Microbiology (Academy) is the honorific branch of the American Society for Microbiology (ASM), a nonprofit scientific society with nearly 40,000 members. Fellows of the Academy have been elected by their peers in recognition of their outstanding contributions to the microbial sciences. Through its colloquium program, the Academy draws on the expertise of these fellows to address critical issues in the microbial sciences.

This report is based on the deliberations of experts who gathered for two days to discuss a series of questions developed by the steering committee regarding the use of next-generation sequencing for faster detection and identification of clinical pathogens. This report has been reviewed by the majority of participants, and every effort has been made to ensure that the information is accurate and complete. The contents reflect the views of the participants and are not intended to reflect official positions of the Academy or ASM.

The Academy acknowledges Peggy A. Reel for her assistance in editing this report.

*Was not able to attend colloquium but reviewed the report as a participant

Summary

Summary

Next-generation sequencing (NGS) has the potential to dramatically revolutionize the clinical microbiology laboratory by replacing current time-consuming and labor-intensive techniques with a single, all-inclusive diagnostic test. Traditional methods for identifying organisms such as mycobacteria, some bacterial species, and fungi in particular are often slow, specialized, and organism specific. Culturing, Gram staining, and biochemical and molecular tests are traditional assays that consume the manpower of the clinical microbiology laboratory. From this battery of tests, relevant treatment guidance for the clinician is not always produced. This has been described elsewhere as a "diagnostic odyssey" or a guessing game for the diagnosis and identification of infectious diseases. Executing diverse clinical tests can waste precious time for a patient and might be the difference between life and death. In the groundbreaking neuroleptospirosis case described by Wilson *et al* (2014), 38 different diagnostic tests on various sample types (e.g., cerebrospinal fluid [CSF], brain, urine, stool, sputum, blood, serum, plasma, oropharyngeal/nasopharyngeal swab) were performed before the diagnosis was ultimately made with NGS. This technology can potentially expedite the turnaround time for a result and allow clinically actionable information to be obtained sooner than a traditional laboratory workup would allow. Data generated from an NGS assay would ideally provide a diagnostic and therapeutic decision in a clinically relevant time frame that would positively enhance patient care and outcome. The American Academy of Microbiology (Academy) conducted a two-day colloquium to review the potential future impact of NGS on clinical microbiology and how barriers for implementation can be overcome through suggested recommendations identified by invited colloquium experts. Main topics discussed at the event are highlighted in an *mBio* minireview article prepared by the NGS steering committee members.

Basic science and applied research coupled with emerging technologies has enabled NGS to transition into the diagnostic laboratory setting to provide clinically actionable results. Insights acquired from NGS methods can be exploited to improve our health as individuals and the greater public health. NGS is poised to broaden our understanding of how microbes interact in different ecosystems, as well as their functioning during health and disease in humans, animals, and environments, including both the built and natural (outdoor) settings. Food safety measures and product quality may also be advanced by NGS technology by allowing faster detection of contaminating pathogens that may arise during manufacturing or processing and subsequently trigger outbreak scenarios. Taken together, the functionalities of this sequencing tool can help achieve the main goal of the One Health Initiative, which is to obtain optimal health for people, animals, and various environments (***Figure 1***).

Moreover, NGS will contribute to the Precision Medicine Initiative, a bold and innovative approach for disease treatment, management, and prevention launched by President Obama in his 2015 State of the Union address. At the present time, NGS is being applied to precision medicine to help diagnose human genetic disorders, prenatal disorders, and cancers. For example, Worthey *et al* (2011) executed whole-exome sequencing, a type of NGS approach, to diagnose a rare immune defect that saved a 15-month-old patient's life. All standard diagnostic tests had been exhausted with no definitive cause for the patient's Crohn's disease-like illness that was coupled with recurrent episodes of sepsis and infection. Exome sequencing revealed a novel, hemizygous missense mutation in the X-linked inhibitor of apoptosis gene that atypically presented as severe gastrointestinal (GI) disease. This mutation directed clinical management to perform a successful allogeneic hematopoietic progenitor cell transplant that cured the patient's GI illness.

Next-generation sequencing (NGS) is a blanket term that collectively refers to high-throughput DNA sequencing strategies that can produce large amounts of genomic data in a single reaction by diverse methodologies. NGS is also referred to in the literature as "deep," "high throughput," or "massively parallel" sequencing.

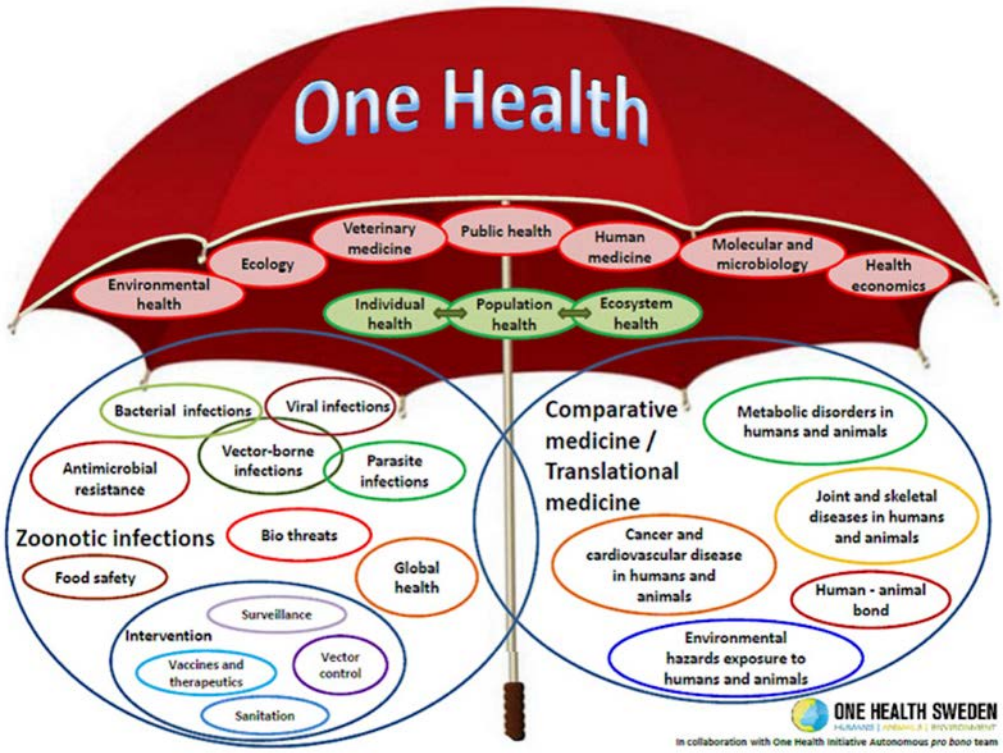


Figure 1. One Health Initiative: the interdisciplinary global health collaboration.
Next-generation sequencing (NGS) capabilities can support the One Health Initiative or the interconnectedness of humans, animals, and the environment. Knowledge gained from the power of NGS can help improve our overall well-being.

Precision medicine recognizes that a "one-size-fits-all" diagnostic and treatment approach is not successful for every patient due to individual variability and host characteristics, such as genetic composition, lifestyle practices, and exposure to different environments/agents. Another component that contributes to a person's unique biological makeup consists of the microbial communities that reside in niches throughout our bodies. Collectively known as the human microbiome, it is estimated that nearly 100 trillion microorganisms live within our gut, mouth, skin, and other mucosal surfaces. These microorganisms perform many beneficial functions that help maintain our health, such as by combatting offending pathogens that want to inflict harm and by synthesizing essential vitamins and minerals. Therefore, the scope of precision medicine encompasses the study of microbes by NGS for microbiome analyses and pathogen diagnostics. Precision medicine will integrate patient and family medical histories, clinical signs and symptoms of illness/disease, and genomic information whether human or microbial in order to create treatment pathways that are individualized and tailored for each patient. As an example, a customized treatment plan for a septic patient might be life-saving, as sepsis can be caused by a severe bacterial, viral, or fungal infection. It is thought that newer drugs for the treatment of sepsis have failed clinical trials because of the heterogeneity of sepsis as a clinical entity. Although these drugs showed promise on the bench, it is speculated that inherent variability among individuals and disease pathogenesis is the cause for poor therapeutic responses, thereby calling for a more individualized approach that precision medicine and NGS technologies bolster.

NGS offers great opportunities for advancing precision medicine in the clinical microbiology laboratory. With the hope of applying this technology for microbial organism and antibiotic/vir-

ulence marker identification for clinical diagnostics, patient care could be dramatically impacted; however, implementing NGS as a routine test in this setting faces significant infrastructure (i.e., computing resources) and bioinformatics challenges. Refining sequencing data to guide clinical decision-making is a complex bioinformatics task, as it involves formulation and interpretation of molecular data in the biological sphere in connection with treatment information in the clinical sphere. Without question, this is an exciting time for the field of clinical microbiology, given the empowering ability of NGS technologies to help us understand and treat infectious disease. This enthusiasm was expressed at the Academy's colloquium, at which invited participants were tasked to answer seven key questions and subquestions developed by the steering committee. Findings from the discussions are summarized in this report, along with real clinical case studies that used NGS for diagnosis and recommendations that address microbiological NGS challenges. The recommendations put forth by the colloquium participants were identified and categorized under the sections below.

Issues That Need To Be Addressed for the Implementation of NGS into Clinical Microbiology

The field of NGS for infectious disease diagnostics has progressed very slowly. To advance the transition of NGS technology into the clinical microbiology laboratory, colloquium attendees established main topic areas and listed suggestions on how to address each issue.

Performance Factors and Infrastructure

In order for NGS-based assays to become commonplace in the clinical microbiology laboratory, there is a need for the development of "turnkey" solutions for all phases of testing (e.g., sample preparation, sequencing, data analysis, and result interpretation). The ultimate goal of diagnostic NGS is to place a direct clinical specimen from any matrix into the NGS workflow and generate an actionable result within a reasonable time frame. Continued efforts for direct clinical sample sequencing should be pursued. (**Recommendation 2.1**)

It is recommended that a distinction be made between diagnostic clinical specificity/sensitivity and analytical specificity/sensitivity when discussing a clinical microbiological NGS test. The qualifiers of "diagnostic clinical" and "analytical" are not interchangeable, and confusion can arise when reporting a laboratory test result. More efforts are also needed to understand the mutation rates and population structure of commonly encountered clinical pathogens in relationship to their effects on NGS sensitivity and specificity as well the use of NGS for molecular epidemiology. (**Recommendation 2.2**)

To help minimize the cost and bulkiness of NGS hardware implementation, the utility of benchtop and point-of-care (field-able) sequencing platforms should be emphasized to clinical laboratories seeking to engage in this space. These sequencing systems consume less space and are generally less expensive than larger NGS platforms, and data analysis can be completed on a high-end desktop server or even a laptop. (**Recommendation 5.1**)

Standard Operating and Validation Procedures

A group of stakeholders (e.g., government agencies, clinical laboratory professionals, academia, industry) should be brought together to develop standardized reference materials and data sets that can be used for assay validation and quality control (QC) procedures. To make the validation of bioinformatics pipelines easier, publically hosted "digital" validation test sets could be purchased and evaluated by clinical laboratories. A set of reads that have a known answer could be downloaded and subjected to the lab's bioinformatics components. Also, fully characterized biological reference organisms will be needed to evaluate both the wet and dry NGS processes. (**Recommendation 4.1**)

It is recommended that different wet and dry bench NGS protocols be created for the detection of etiological agents such as bacteria, viruses, fungi, yeasts, and parasites. Although NGS has the potential to detect all pathogens in a clinical sample, specific protocols would help to

advance the transition of NGS into the clinical microbiology laboratory. There needs to be guidance on how to validate and perform QC procedures for these protocols as they pertain to the different pathogens (i.e., what are the unique/pathogen-specific QC metrics that must be considered to ensure the quality of NGS results). (**Recommendation 4.2**)

Interdisciplinary Teams and Education

To ensure successful utilization of an NGS result, a multidisciplinary team within the clinical or public health laboratory setting should be formed to include the expertise of clinical microbiologists, medical technologists, clinicians, infectious diseases physicians, pathologists, basic research scientists, software developers, and bioinformaticians. This collaborative effort will maximize the strength and interpretation of NGS data. (**Recommendation 4.3**)

Adoption of NGS into the clinical microbiology laboratory will require clinical microbiologists, medical technologists, and clinicians to receive training in molecular biology and bioinformatics. Future clinical microbiology and public health laboratory professionals will be required to be competent in the field of bioinformatics in order to effectively communicate with bioinformaticians. Beyond general programming skills and bioinformatics knowledge, there needs to be training on understanding and interpreting NGS results. It is recommended that bioinformatics be incorporated into the coursework of medical school students and clinical microbiology/pathology fellowship programs so the students gain familiarity with this diagnostic approach, which is likely to be used during their clinical practice. Exposure to informatics could even begin at the high school and undergraduate levels, since the basic principles are applicable to many fields. (**Recommendation 4.4**)

Professionals who will use NGS technology should work closely with software developers to create a proficient, streamlined, and more manageable analysis pipeline to provide a quicker return of complete diagnostic information. This collaboration will help in the development of more efficient and user-friendly software programs, interpretable analysis reports, and improved algorithms for genomic data analysis. (**Recommendation 6.1**)

Interpretation, Sharing, and Management of Data

To greatly assist in outbreak scenarios in both the hospital and the community setting, guidelines or models for responsible data sharing among institutions should be developed and endorsed by a consortium of relevant stakeholders. These models should encourage continual sharing of microbial genomic data and maximize public availability while balancing the need for patient privacy. This balance of sharing data and maintaining privacy is necessary for predictive outbreak detection to work; hence, the public health benefit of using NGS in the clinic can be gained. (**Recommendation 3.1**)

When NGS data are to be stored remotely such as on a server, in the cloud, or within a database, the information must protect patient privacy and be Health Insurance Portability and Accountability Act (HIPAA) compliant. (**Recommendation 5.3**)

Sequence data, including raw sequence reads, .FASTQ files, and the complete genomic sequence of the identified pathogen, are large and would consume considerable storage space for a hospital. The assembled sequence should be uploaded to an appropriate database, and only the clinically relevant result should be maintained in the patient's electronic medical file. (**Recommendation 5.4**)

Clinical Reporting and Billing

Some NGS-based assays, e.g., metagenomic assays, are capable of returning a complex set of results that require careful interpretation by the clinical microbiologist to determine what is clinically actionable and what should be included in the result report to ensure optimal patient care. For example, NGS results should assist the physician in determining what antimicrobial can be used, rather than what the organism is resistant to or what drug is not suitable for

treatment. The presence of the gene conferring antimicrobial resistance (AMR) is not evidence of its expression and hence AMR phenotype. There needs to be a way to phenotypically verify the genotypic result generated by NGS. Rapid phenotypic testing methods are currently under development. (**Recommendation 5.2**)

An information visualization style approach to conducting controlled user studies might help to determine which form of NGS clinical reporting is most effective for clinicians to make a diagnosis or initiate/change treatment for the patient. (**Recommendation 6.2**)

New regulatory guidelines and insurance reimbursement codes for the use of NGS testing in the clinical microbiology laboratory need to be developed. Insurance billing codes should be revised to enhance the transparency of molecular services that are performed. Additionally, payers such as CMS (Centers for Medicare & Medicaid Services) should review publications showcasing that NGS assays guided or improved diagnostic and therapeutic decisions that could not be made using current laboratory methods. (**Recommendation 6.3**)

Outcome analyses and clinical trials highlighting the success and cost savings of NGS for the diagnosis of infectious diseases are highly recommended and could serve as justification for reimbursement companies. Incentive and, more specifically, funding, which is not widely available, must be given to initiate such studies. Therefore, an advisory board composed of relevant stakeholders should be created to address this issue. (**Recommendation 6.4**)

Reference Databases

It is recommended that genomic sequences of emerging microbial pathogens be uploaded to a unified, public database as quickly as possible to allow for community engagement of the data analysis and use of those data to inform other clinical professionals of the pathogens they are encountering in their laboratories. If genomic sequences for high-priority pathogens are routinely deposited, NGS has the potential to serve as the new early warning system for outbreaks that may occur locally, nationally, or internationally. This tactic could help monitor the stability of the outbreak isolate's genome over time and determine if acquisition or removal of genomic information affects diagnostic and therapeutic decision-making. (**Recommendation 3.2**)

A crucial recommendation is the expansion of curated and regulatory-grade microbial sequence databases in the public domain. Genomic sequence submissions should include high-quality sequence data that are accurately annotated with metadata. These databases should not be a static collection of information but should allow for local, national, and international data exchanges that are in line with agreed standards. Additional databases are not needed, but existing databases should establish standardized quality metrics or curation strategies to promote confidence in clinical decision-making. (**Recommendation 7.1**)

It is recommended that sequencing efforts be focused on obtaining more pertinent whole genomes for pathogenic fungal, yeast, and parasitic species. (**Recommendation 7.2**)

Resistance genes should be annotated as a subset within an appropriate existing database. With new genetic mechanisms of resistance frequently arising, these databases would be ongoing projects requiring active curation and reannotation efforts. (**Recommendation 7.3**)

*One Health Initiative. [http://www.onehealthinitiative.com/about.php]. Accessed June 28, 2015.

Introduction

Sanger sequencing has been recognized as the reference standard for DNA sequencing over the past 37 years. Since its introduction in 1977 by Frederick Sanger and colleagues, dideoxynucleotide sequencing has dominated the sequencing landscape (1, 2, 3, 4, 5, 6, 7). Sanger biochemistry has been modified over three decades to yield read lengths of up to 1,000 bases, with raw reads obtaining accuracies as high as 99.999% (8, 9). This is a well-defined, mature chemistry that has laid the groundwork for the sequencing of genes and even whole genomes; however, it is limited in its throughput capacity (1). Despite the success of Sanger sequencing to produce early maps of the human genome, its slow pace ignited the demand for more robust DNA sequencing technologies that could generate large amounts of genomic data in a quicker and more affordable manner (9, 10, 11). In 2005, the GS-20 sequencing platform from 454 Life Sciences, the first non-Sanger-based sequencing system, was launched (12, 13, 14). This system executed a massively parallel pyrosequencing method that formed the underpinnings of a new wave of high-throughput genomic analysis known as next-generation sequencing (NGS). Subsequently, NGS technologies have revolutionized the field of genomics, enabling a comprehensive analysis of genomes, both human and microbial, in days rather than years and at a cost of thousands of dollars per sample rather than billions (8, 15, 16). In the following paragraphs, specific examples of the scientific impact or forthcoming advantages of NGS techniques are highlighted.

Currently, a major challenge of NGS is the application of its brute sequencing power for infectious disease diagnoses in the clinical microbiology laboratory (17). Infectious disease is one of the leading causes of mortality worldwide (18). Rapid detection of the causative agent is crucial for implementing appropriate therapeutic measures and improving the patient's standard of care (19). Testing pathways in the microbiology laboratory have changed little over the past 50 years, but regardless of established methodologies, complete diagnostic information is not always generated (19, 20, 21, 22). For example, when a 20-year-old woman returned to the United States after hiking in Western Australia, she presented with a fever, rash, headache, nausea, and muscle and joint pain. A wide range of tests for common infectious causes of acute febrile illness were completed but yielded negative results. While in Australia, the patient was warned of a Ross River virus outbreak, but diagnostic tests for this organism were not readily available. Metagenomic NGS, which tests for the entire spectrum of disease-causing organisms in a sample without requiring specific primers or probes, was conducted on the patient's blood sample. The sequencing unexpectedly revealed the presence of reads corresponding to human herpesvirus 7 (HHV-7), an infection typically seen during childhood. Metagenomic NGS was able to clearly demonstrate the presence of HHV-7 within 48 hours of receiving the specimen, and the patient recovered within 2 weeks (23). This example highlights how NGS can bypass many of the limitations of the current diagnostic scheme by allowing physicians to assess for multiple pathogens as part of the initial diagnostic evaluation, thereby avoiding multiple rounds of testing that look for progressively less-common pathogens (24, 25, 26, 27, 28). NGS technology is appealing to clinical microbiologists because of its increased availability, decreased cost per base, and capability to detect a broad range of pathogens (8, 25, 29, 30). NGS also has the potential to eliminate the multitude of microbiological tests that are currently conducted on clinical specimens (29, 30). In addition, the data generated using NGS could assist with the development of new or enhanced diagnostic assays, for instance, by providing sequence information that would allow for the design of improved, pathogen-specific DNA targets and primers used in multiplex assays (31). NGS has the capacity to profoundly change how infectious diseases are diagnosed, although adoption of the technology has been slower than some may have predicted, at least at the current time (22). With these technological sequencing advances come new challenges with methodologies, bioinformatics, clinical reporting (32), and databases (33), all factors that are addressed in upcoming sections. Despite these caveats, the clinical microbiology laboratory is still on the threshold of implementing NGS into routine practice (31).

Not only has NGS advanced the field of clinical microbiology, it has also been successfully applied to the field of human genetics and precision medicine (34). Common applications of human NGS include sequencing of multigene panels, noninvasive pre-natal testing for the detection of fetal chromosomal abnormalities, and exome and whole-genome sequencing (WGS) for the identification of common and rare genetic disorders, disease-specific variants, and cancer-specific alleles (34, 35, 36, 37, 38, 39, 40). Another useful application for NGS is human and microbe identification for forensic practices. The standard method for forensic analysis is DNA profiling by short tandem repeats (STRs) (14). Although STRs provide adequate discriminatory power when the origin of and relationship between DNA samples are being examined, a more

Case Study 1

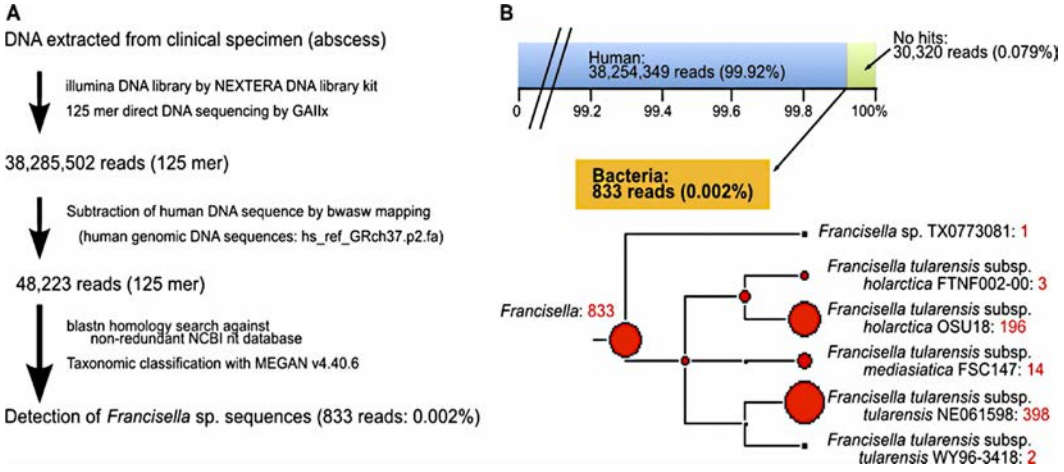
Deep sequencing differentiates between *Francisella tularensis* subspecies, a capability critical for biosecurity (42)

A 57-year-old male that resided in Fukushima Prefecture, Japan, presented to a clinic with a right thumb ulcer, swollen right axillary lymph node, and pain in the right wrist through elbow joint. Laboratory tests diagnosed the patient with cellulitis and lymphadenopathy, and treatment was given. However, 1 month later, he returned to the clinic with a swollen axillary node and redness and reported that he had skinned a hare. Pus from the enlarged axillary region was cultured but unrevealing. The pus was also examined for potential pathogens by unbiased deep sequencing, which yielded *Francisella* species-positive reads (Figure A). Whole-genome single-nucleotide variations (SNV) were used to extrapolate the subspecies and potential virulence of the *Francisella* species. 16S rRNA sequencing does not have this discriminatory power. Deep sequencing and SNV genotyping detected *F. tularensis* subsp. *holarctica* rather than the potential bioterrorism agent *F. tularensis* subsp. *tularensis* (Figure B). This crucial distinction prevented the need for patient isolation and use of a biocontainment facility when working with *F. tularensis* subsp. *holarctica*.

refined level of discrimination and phylogenetic analysis is given by NGS technologies (41). NGS has the potential to quickly and safely characterize microbes related to biocrimes and bioterrorist events (14). The Department of Homeland Security introduced the BioWatch program to function as an early detection mechanism for dangerous pathogens in public places and therefore mitigate the risk of biological threats. Pathogens themselves can act as bioweapons (e.g., anthrax), or specific toxins produced by pathogens can be extorted for use in biocrimes (e.g., ricin). [**See Case Study 1 for an example of how the resolution power of NGS is capable of differentiating between the potential bioterrorism agent *Francisella tularensis* (*F. tularensis*) subspecies *tularensis* and *F. tularensis* subspecies *holarctica* (42).**] The availability of high-quality sequence data can aid investigations or even expedite them, be used as evidence in legal cases, or guide the government's response in the case of a bioterrorist event (14, 43). Thus, NGS technologies have pertinent technical, governmental, and legal roles (42, 43, 44, 45).

Analysis of the human microbiome is yet another area where NGS is gaining traction (46). Comprehensive characterizations of microbial communities that comprise areas of the human body, such as the gut and skin, can be completed with NGS methodologies (47). Specifically with the gut microbiome, the majority of organisms are uncultivable, anaerobic species and hence could not be characterized before the introduction of high-throughput sequencing. NGS methods such as culture-free 16S rRNA gene and metagenomic sequencing have enabled the study of novel anaerobic and aerobic gut microorganisms. Metagenomic data have allowed for the development of microbiome-based research that have been applied to cystic fibrosis management, fecal transplant therapy, and bacterial vaginosis therapy (48, 49, 50). Microbiome studies have linked the bacterial metabolism of dietary phosphatidylcholine with an increased risk of major adverse cardiovascular events in humans (51). Other human metagenomic studies have demonstrated how specific alterations in the composition of gut microbiota can contribute to GI disease, obesity, and type II diabetes (49, 52). Taken together, NGS methodologies have expanded our knowledge of the complexity and composition of the human microbiome in relation to development, health, and disease (31, 46, 48, 49, 50, 51, 52, 53, 54, 55, 56).

In addition to the use of NGS in studying the microbial diversity of humans, NGS technologies have been used to investigate the microbial populations among different food ecosystems. NGS is making great headway in the food microbiology field, which investigates both the beneficial and harmful effects that microbes have on food safety and quality. NGS methods have been used to examine the microbial profiles of various foods to help optimize maturation or preservation practices, to detect unexpected microorganisms that cause spoilage, and to detect unwanted pathogens that cause illness (57). Although the U.S. food supply is among the safest in the world, sources of contamination can arise from any step within food production, such as during processing, packaging, transportation, and



Case study 1 (Figure) Comprehensive direct sequencing from fluid collected from the patient's axillary region. (**A**) Schematic representation of the analysis pipeline for the detection of pathogens based on comprehensive sequencing of human clinical specimens. After excluding human-derived DNA sequences using the BWA-SW read-mapping software program, the remaining short reads were subjected to a BLASTN homology search to detect potential pathogens. (**B**) Summary of homology search results for all obtained reads from the abscess specimen. The identified bacteria reads corresponded to *Francisella* sp., and the taxonomic classification suggests a *Francisella* subsp. group for every sequence read. The number in red indicates the assigned read counts for the respective bacteria species.

preparation. In a foodborne outbreak, NGS data can be leveraged to track the pathogen's source with the hope of stopping the outbreak in its tracks and protecting public health. The Center for Food Safety and Applied Nutrition (CFSAN) within the FDA uses WGS, a specific NGS approach, to (i) differentiate sources of contamination that may have occurred within the same outbreak, (ii) determine which ingredient in a multi-ingredient food contained the pathogen responsible for the outbreak, (iii) track the source of the contaminated ingredient, and (iv) execute food surveillance to prevent further disease. CFSAN's GenomeTrakr program, a component of an ongoing global surveillance project that uses WGS for the rapid detection of outbreaks of foodborne illness and pathogen traceback, is an example of the food surveillance function. For example, in March 2014, the FDA shut down a Kenton, Delaware, company due to a multistate *Listeria* outbreak that stemmed from contaminated Hispanic-style cheeses manufactured in this facility. WGS showed that the *Listeria* strains isolated from outbreak patients were highly related to the *Listeria* strains detected in the Hispanic-style cheese products (58, 59). More than 19,000 bacterial draft genomes have been deposited within GenomeTrakr and are publically available for comparison of outbreak isolates in real time (60, 61).

Up until this point, the applications of NGS that have been briefly reviewed involve the clinical and biomedical arenas, but NGS is an important tool that extends beyond these fields and is at the forefront of environmental microbiology and ecological science. Studies on the biodiversity of marine, freshwater, terrestrial, and agricultural ecosystems have revealed intriguing insights about the microbial inhabitants (62). For instance, NGS has provided a wealth of knowledge in determining plant viral diseases of unknown origin and other agricultural pathogens that have aided in pesticide optimization and efficacy (63). NGS can detect mutations causing pesticide resistance and, therefore, has the potential to aid in plant protection and fitness. Plant genomics has also unveiled the complexity and diversity of domesticated crop species. In turn, this diversification has been utilized to improve the health of crops and food safety (64).

Since its introduction in 2005, NGS technology has had a fundamental and far-reaching impact on many fields related to the biological sciences, including the medical, forensic, environmental, and

Statement of Task

The guessing game aspect of diagnosis or identification of infectious diseases can waste precious time for a patient. Clinicians are commonly forced to make an educated guess about therapy prior to knowing the infecting pathogen, and delays in microbe identification increase the risk of ineffective treatment and spread of infection. Rather than running a variety of tests to identify a pathogen, NGS offers a wide diagnostic repertoire that can identify the culprit no matter the organism—bacterium, virus, fungus, yeast, or parasite. NGS methodology has tremendous potential to impact patient care by helping clinicians tailor patient treatment, therefore reducing the usage of ineffective drugs and decreasing the selective pressure for resistance development. The colloquium will examine the core issues of NGS as a diagnostic tool for fast pathogen detection, including:

- The purpose and utility of NGS for the detection of bacteria; status of fungi, yeast, and parasite NGS applications
- Functionalities of NGS versus competing MALDI-TOF MS, PCR assays, and other relevant diagnostics (i.e., ELISA, Smarticles™); a comparison in turnaround time for results, sensitivity, and specificity among the technologies
- The impact of NGS if implemented in clinical/public health microbiology laboratories; the influence of NGS on outbreak investigations, unculturable organisms, metagenomics, and human microbiome studies
- The applicability of NGS to clinical, diagnostic, and reference laboratories, including an examination of feasibility into workflow, cost of the technology, efficiency of correct identification, and development of standard operating procedures
- Data management issues of NGS; pipeline for interpreting NGS data, how the data can be made clinically relevant or how they fit into standard of care, and database development to store genetic data
- Deterrents and shortcomings of NGS technology and the challenges of implementation; regulatory issues, recommendations on how to address barriers for implementation, and current clinical trials using NGS
- Standardization of NGS data and the creation and maintenance of a reference database

Case Study 2

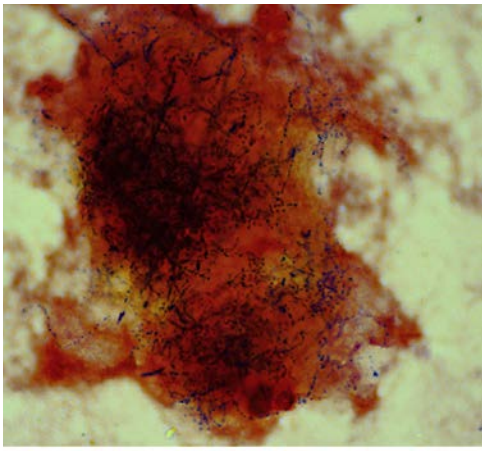
Mastoiditis caused by *Fusobacterium nucleatum-Actinomyces israelii* coinfection determined by NGS (68)

This case study demonstrates the use of NGS in the diagnosis of multiple organisms that caused aggressive mastoiditis in a 46-year-old developmentally delayed man. The use of conventional culture and molecular clinical techniques can be challenging when characterizing the organisms that comprise polymicrobial infections. In September 2013, the patient presented to a hospital with a swollen left mastoid, bleeding gingiva, and increased oral secretions. A computed tomography (CT) scan revealed abscess formation in the left temporalis muscle that involved a significant portion of the left temporalis bone. Surgical debridement of the affected area showed necrotic tissue, which was Gram stained and revealed rare branching Gram-positive rods that were suggestive of *Nocardia* or *Actinomyces* (**Figure**). However, 16S rRNA sequencing was positive for *Fusobacterium nucleatum* (*F. nucleatum*). Because of the diagnostic uncertainties of the patient's mastoid infection, NGS was performed on an Illumina MiSeq platform that specifically targeted bacterial 16S rRNA genes in a culture-independent fashion. Along with the detection of normal oral flora, deep sequencing reads detected *F. nucleatum* (78%) and *Actinomyces israelii* (0.13%). Both of these organisms are well-known infectious agents that are frequently coisolated from infection sites. This case report demonstrates that NGS can be used to evaluate the overall composition of a polymicrobial infection and is a powerful technique for resolving discrepancies produced by conventional laboratory methods.

agricultural disciplines (65). As NGS technologies continue to improve, it is anticipated that there will be additional innovative applications for them in the clinical and public health settings (66). The focus of this report is the implementation of NGS for routine microbiological use, a relatively new avenue for this technology that holds immense transformative potential for the detection, identification, and characterization of infectious agents in clinical and public health microbiology laboratories.

Approach to the Task

In April 2015, the Academy convened a group of 25 experts at ASM Headquarters to discuss the clinical microbial applications of NGS. A steering committee was appointed prior to the colloquium and was tasked with developing discussion questions that were to be posed at the event and with compiling a list of invitees that had expertise covering the scope of this topic. Invited colloquium attendees included representatives from government agencies, commercial industry, clinical laboratories, and academia, with backgrounds in clinical microbiology, diagnostic development/regulation, food microbiology, public health surveillance, DNA sequencing technologies, bioinformatics, computer science, and database formulation and management. Given that NGS is a field that is constantly evolving, the ASM and the Academy's Board of Governors sought to cast light on the current status of NGS technologies in clinical microbiology practice and to provide recommendations for implementation. Colloquium participants were assigned seven main discussion questions and subquestions that covered the broad categories explained in the statement of task. Participants were divided into working groups that consisted of a variety of expertises in order to answer the discussion questions. Groups reconvened for plenary sessions to review all answers. This report summarizes the discussions held during the colloquium's plenary sessions.



Case study 2 (Figure) Rare aggregate of branching, Gram-positive rods, characteristic of *Nocardia* and *Actinomyces*.
With further analysis, the *F. nucleatum* was determined to be the etiological agent of which culture did not detect.

Republished from reference 68 with permission.

Structure of the Report

The report is divided into seven sections based on the questions that were asked during the colloquium. These discussion questions were developed by the steering committee members prior to the event. Section 1 specifically focuses on the use of NGS for bacterial detection. Although NGS can detect other microbial pathogens, such as viruses, fungi, yeasts, and parasites, the colloquium underscored bacterial identification. Section 2 provides a comparison of NGS technologies to other types of diagnostic assays, including differences in turnaround time, specificity, and sensitivity. Section 3 elaborates on the impact that NGS could have on the clinical and public health microbiology fields, including the use of NGS for metagenomics and human microbiome studies and the identification of unculturable/difficult-to-culture organisms. Section 4 describes the necessary factors for incorporating NGS into the clinical microbiology laboratory workflow, including standard operating procedures, process validation, and reference materials for validation. This section also discusses the synergy that needs to be created among various groups of professionals to fully realize the benefits of NGS technologies. Section 5 examines the issues encompassing NGS data analysis, interpretation, management, storage, and archiving. Section 6 explains the downfalls of NGS, the challenges of implementation, and suggestions for overcoming these difficulties. Section 7 discusses the need for quality and curated microbial genomic reference sequences and metadata in public repositories. Numerous case studies that demonstrate the utility of NGS as infectious disease diagnostic tests are also highlighted throughout this report.

Section 1
Utility of NGS for Bacterial Pathogen Detection

When discussing the role of NGS from a pathogen-agnostic perspective, detection has a multitude of purposes and unfolds in the clinical, public health, agricultural, and environmental fields. NGS of complex samples or metagenomics allows for comprehensive pathogen detection without *a priori* knowledge of the target organism (4, 8, 10, 15, 25, 67). Unlike other techniques that can be used to identify microbial pathogens from clinical samples, metagenomics by NGS is not limited to known organismal sequences. [**See Case Study 2, which features the inclusivity of an NGS assay by showing that a mastoid infection was caused not only by *Actinomyces israelii* but also by *Fusobacterium nucleatum* (68).**] Essentially, NGS yields "detection with benefits," with those added benefits being organism differentiation, novel organism discovery, elucidation of virulence factors and resistance markers, and host characteristics in response to the offending microbe and administered therapy. Detection can also operate on a purpose-dependent level, such as for outbreak tracking and hospital infection control surveillance.

1a. Status of Virus, Fungus, Yeast, and Parasite NGS Applications

Discussions at the colloquium focused primarily on the use of NGS for the detection and identification of bacterial species, since bacteria are the most commonly encountered microorganisms in the clinical microbiology laboratory (69). A brief synopsis of the status of NGS with other kinds of microorganisms, including viruses, yeasts, parasites, and fungi, is described below.

Much of the pioneering work in establishing NGS as a method for pathogen detection was conducted by virologists. The initial studies using metagenomics by NGS on human clinical specimens unveiled a novel polyomavirus associated with Merkel cell carcinoma and a novel arenavirus that caused fatal febrile illness.). [**See Case Study 3, which demonstrates how RNA sequencing and ultimately protein homology were used to discover a new arenavirus that had been transmitted through solid organ transplantation (70, 71).**] Briese *et al* (72) used unbiased pyrosequencing of RNA extracts to evaluate an unexplained outbreak of extremely fatal hemorrhagic fever in southern Africa. Within 72 hours of receiving the clinical samples from outbreak victims, phylogenetic analyses showed the presence of a highly novel viral genetic lineage that was only distantly related to the Old World arenaviruses and was known as Lujo virus (LUJV). The work by Briese *et al* marks the first application of NGS for pathogen discovery associated with an outbreak of hemorrhagic fever caused by a genetically distinct virus. Other examples of NGS for identification of novel viruses in outbreak settings include the discovery of Bas-Congo virus (BASV), a novel rhabdovirus associated with a cluster of hemorrhagic fever cases in central Africa (73), and novel adenoviruses in an acute respiratory outbreak in a baboon colony with evidence of coincident human infection (74).

As evidenced by the abovementioned studies and many others (75, 76, 77, 78, 79), NGS technologies have made significant contributions to the clinical virology field, including diagnostics, discovery, pathogenesis, epidemiology, and genome sequencing, through metagenomic-based sequencing approaches (known as viromes) (13, 79). With viral metagenomics, either DNA or RNA may serve as the input nucleic acid for analysis. Although library preparation is more time-consuming for RNA than for DNA, in some cases, RNA sequencing is more beneficial than DNA sequencing. First, some viruses have an RNA genome which would not be detected if only DNA was extracted and used as the starting material, as exhibited by Case Study 3. Second, performing total RNA extraction/sequencing can capture both DNA and RNA expression, and hence, mRNA sequence can be translated into protein. Amino acid sequence tends to be more conserved than nucleotide sequence and therefore may yield more defined taxonomic information (70, 80, 81).

The use of NGS for virus detection has progressed tremendously since initial studies were published in 2008. In fact, clinical viral diagnostics are now more advanced than bacterial NGS diagnostics. Given the high specificity of viral sequence data, presently it is more practical for a viral diagnostic test to be CLIA (Clinical Laboratory Improvement Amendments of 1988) approved than a bacterial NGS assay. Applying NGS to bacterial analysis to make a specific diagnosis can be complicated when performing a metagenomic NGS application, simply because large amounts of commensal bacterial microorganisms can colonize the patient, making determination of the causative organism more difficult. Furthermore, it has been demonstrated that contamination of laboratory reagents that are used for NGS, such as DNA extraction kits or molecular biology-grade water (a common contaminant is *Bradyrhizobium* sp.), can significantly impact the results obtained using metagenomics (82). In contrast, while certain viruses can occasionally constitute part of the normal microbial flora (e.g., torque teno virus) (83) or be potential

Case Study 3

Application of metagenomic RNA sequencing discovers a novel arenavirus responsible for three cases of fatal febrile illness (70)

Arenaviruses are a species of enveloped, negative-strand RNA viruses with a bisegmented genome that commonly infect specific rodents. These rodents can transmit the arenavirus to humans via their excrement (e.g., inoculation with infected urine), aerosolized infectious particles, or contaminated food products. In this case, three Australian patients received either a liver or kidney transplant from a single donor who had died of a cerebral hemorrhage after a 3-month trip in southern Europe. Within 4 to 6 weeks after transplant, all three patients died of a fatal febrile illness with various degrees of encephalopathy. The battery of tests performed on the patients' samples included culture, PCR, serological assays, and oligonucleotide microarrays, which were specific for numerous bacterial, viral, and panmicrobial pathogens. When NGS technology was used, DNA sequencing reads were unrevealing of any candidate pathogens. However, RNA sequencing identified novel viral reads that were identical for all tested samples (i.e., tissue, blood, and CSF clinical specimens); hence, a single virus had been transmitted to all organ recipients. Only 14 of 103,632 sequence reads or a viral/human RNA read ratio of 0.0135% was sufficient to demonstrate the presence of a new Old World arenavirus. Exposure to and acquisition of this novel virus were thought to have occurred when the organ donor visited rural areas abroad. He subsequently transmitted the virus to three recipients of solid-organ transplants, which caused fatal disease and death. This case represents the first study that implemented NGS for new pathogen detection and discovery.

Case Study 4

NGS reveals a *Coccidioides immitis* cluster in three organ transplant patients (90)

Whole-genome sequence typing (WGST) is an application of NGS that was used in this case to identify multiple donor transplant-related coccidioidomycosis events that occurred in patients who resided in an area of the United States where *Coccidioides immitis* (*C. immitis*) was endemic. In regions of endemicity, it can be difficult to distinguish between donor-derived primary infections with *Coccidioides* and latent infections with this pathogenic fungus. Whole-genome sequencing and single-nucleotide polymorphism (SNP) analysis showed that the three *C. immitis* isolates collected from the organ recipient patients were nearly genetically identical, with only three SNP differences among them. Postmortem analysis of the donor's serum was positive for IgM antibodies to *Coccidioides* spp., therefore making the donor the common source for this cluster of transplant-related infections. WGST demonstrated the genotypic and epidemiologic relatedness of *C. immitis* clinical isolates among organ recipients and their donor.

laboratory contaminants (e. g. *Circoviridae*/parvovirus-like hybrid virus [NIH-CQV/PHV] and xenotropic murine leukemia virus-related virus [XMRV]) (84, 85), the high sequence specificities of viruses make them easier to identify, and potentially diagnose, from metagenomic data than bacteria. NGS can detect 95% of viral pathogens (13, 86). This high detection rate cannot be said for bacteria unless targeted methods such as 16S rRNA gene sequencing are used. The identification of reads corresponding to a virus is likely indicative of ongoing disease. For example, if influenza virus is detected in a bronchoalveolar lavage (BAL) fluid specimen of a patient with pneumonia, then the infection was likely caused by this virus. On the other hand, if *Staphylococcus aureus* is detected in a BAL specimen, it is not guaranteed that *S. aureus* is the cause of pneumonia. *S. aureus* could present as a colonizer or commensal oral flora, contamination of the laboratory environment or reagents, or a simple misidentification.

Despite the advancements made with NGS viral diagnostics, there are challenges that viral genera and other nonbacterial organisms face when subjected to the technology. Briefly, there are studies being conducted using NGS to detect viral quasispecies with human immunodeficiency virus (HIV), Ebola virus, and influenza virus (56, 79, 87, 88, 89). These small minor variant viral populations can be clinically important but are often challenging to detect, since viral nucleic acid integrity can be compromised during extraction. Furthermore, parasites, yeasts, and fungi are eukaryotic microorganisms that pose bioinformatics challenges. These organisms possess larger genomes than viruses and bacteria, and their reads can be difficult to distinguish from human reads. [**See Case Study 4 to understand how WGS was applied to detect a pathogenic fungus that infected three organ transplant patients (90).**] In addition, low-quality reference genomes can be contaminated by other viral, bacterial, fungal, or yeast sequences, and thus it can be taxing to determine if a top alignment hit is truly the correct pathogenic hit or just a database contaminant. More discovery efforts are necessary for parasite, yeast, and fungus NGS in addition to reference databases for these nonbacterial organisms (69, 91, 92, 93, 94). There are various medically important parasites, yeasts, and fungi that do not have high-quality reference genomes, and therefore NGS applications involving these organisms will have a tough time advancing without more concrete and curated reference databases.

As the Precision Medicine Initiative moves forward, NGS has the potential to be incorporated into routine clinical microbiology workups to direct diagnostic and therapeutic decision-making that is specific for each patient. Since this colloquium focused mainly on the clinical bacterial applications of NGS, participants identified five main areas that they believe could benefit from the capabilities of NGS. These include (i) clinical identification from primary samples or a pure culture, (ii) infection control actions, (iii) antimicrobial stewardship, (iv) outbreak investigation in community and hospital settings to guide measures for containment, and (v) pathogen discovery (1, 70, 95, 96, 97).

1b. NGS Platforms: Past and Present Commercial Entities

Competition among sequencing vendors has resulted in sustained improvements to NGS platforms from streamlined setup to sequencing chemistry adjustments and user-friendly data analysis (38). Several commercial NGS platforms are currently available in the sequencing marketplace, and newer systems are on the horizon (3, 7, 22, 25, 33, 38) (***Figure 2***). The basic functionalities of these platforms include WGS, whole-exome sequencing (WES), metagenomic sequencing, and targeted gene sequencing (***Table 1***) (1, 3, 31, 98). The optimal sequencing platform for a particular laboratory is highly dependent on the purpose for which it will be used. In the clinical microbiology laboratory, NGS platforms are primarily used for WGS and metagenomic sequencing, of which the latter task does not require knowledge of the possible causative agent. Targeted sequencing is also being used in the clinical laboratory if an etiological agent is suspected at the time of clinical sample collection. Regardless of the application, the capital outlay to purchase and maintain NGS equipment is considerable, and smaller hospitals and institutions may not have adequate funding, particularly without a well-established billing infrastructure in place (38).

Although these platforms differ in their sequencing chemistries and engineering principles, they all perform massively parallel sequencing which yields terabytes of data (13). This common technological feature involves the sequencing of either spatially separated, clonally amplified DNA or spatially separated, single DNA molecules on a flow cell (1, 3, 5, 8, 15, 22, 99). Hundreds of megabases (millions) to gigabases (billions) of DNA sequence data can be produced from a single NGS run, which is in stark contrast to the hundreds of bases produced from an individual sequencing reaction of a targeted region by the Sanger method. Also, the amount of input DNA is dependent on the NGS platform and the intended application, along with the library preparation protocol used to extract the DNA to be sequenced. All of these factors might affect expected results and data output (1).

For instance, the typical DNA requirement ranges for the Illumina MiSeq, Ion Torrent PGM, and PacBio RSII are 50 to 1,000 ng, 100 to 1,000 ng, and ~1 µg, respectively (100). Newer, single-molecule, real-time sequencing (e.g., PacBio) tends to require a larger quantity of purified DNA to yield optimal results (25). Furthermore, when using the NexteraXT library preparation kit for bacterial WGS on the Illumina MiSeq, 1 ng of input DNA is needed, yet when the TruSeq library preparation kit is applied for DNA extraction, 3 to 5 µg of input DNA is required for analysis.

There was a paradigm shift in sequencing design when the **Roche-454** system (previously 454 Life Sciences) entered the market in 2005. This technology involves the construction of adapter-ligated

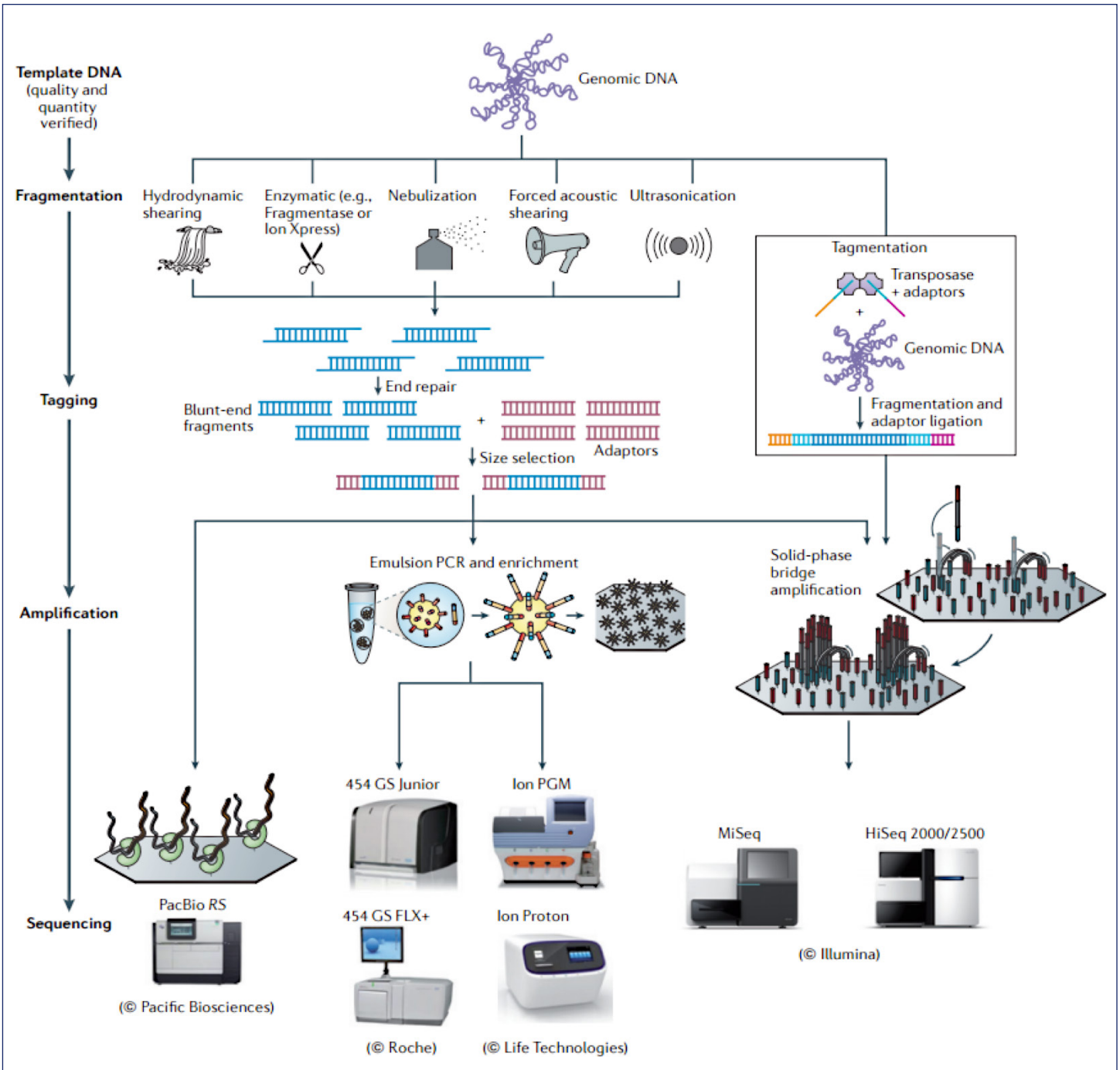


Figure 2. High-throughput sequencing platform examples.
The schematic shows the main high-throughput sequencing platforms available today, and the associated sample preparation and template amplification procedures. It is important to note that Roche will be ceasing operations of the 454 sequencing platforms by mid-2016. Illumina also provides other series of NGS platforms (i.e. NextSeq, HighSeqX™) beyond those mentioned here. See text for more details.

Republished from reference 38 with permission from the publisher.

Table 1. Characteristics of NGS platforms[a]

NGS platform	Sequencing chemistry	Instrument	Maximum read length	Purpose/main use
Roche-454[b]	Pyrosequencing; sequencing by synthesis	GS FLX+	1,000 bp	Will no longer be used by 2016; long reads made the platform well suited for *de novo* assembly and pathogen discovery
Illumina	Reversible terminator chemistry	NextSeq	300 bp	Current workhorse of NGS platforms; supplies bidirectional reads; platform can be used for pathogen discovery, exome sequencing, targeted sequencing; also overcomes homopolymeric regions
ABI SOLiD	Sequencing by ligation, oligonucleotide probe ligation	5500 SOLiD	75 bp	High sequence accuracy makes the platform equipped for genome resequencing and polymorphism analysis
Ion Torrent	H^+ ion-sensitive transistor	Ion Proton I	200 bp	Platform is suitable for small genome sequencing, exome sequencing, and targeted sequencing
HeliScope™[b]	Reversible terminator chemistry	HeliScope™ single-molecule sequencer	35 bp	Requires the addition of a poly(A) tail; platform is effective at sequencing native viral genomes and immune-precipitated methylated DNA; capable of sequencing small sample quantities; high platform costs and poor sales lead caused production to cease
PacBio	Real-time sequencing; phospholinked fluorescent nucleotides	PacBio RSII	50 kb	Platform provides long read sequencing and a low degree of bias; suitable for de *novo* assembly, targeted sequencing, and base modification detection
Oxford Nanopore	Real-time sequencing; electronic sensing or nanopore sequencing	MinION	>50 kb	A portable, USB-powered sequencer that is under development and has been used by participants in MAP[c] for *de novo* assembly and resequencing

[a]Adapted with permission from references 12, 34, and 99.

[b]Sequencing platform is no longer manufactured.

[c]MAP, MinION Access Programme (sequencing centers that were granted early access to the MinION).

DNA libraries, emulsion PCR, and pyrosequencing (**Figure 3**) (5, 15). Although the Roche-454 instrument had a higher throughput capacity and a lower sequencing cost per base than the Sanger methodology, interest in this next-generation technology among the scientific community was muted. Critics of the technology identified read length, fidelity, infrastructure requirements, and cost of operation as the main issues; however, these same concerns were also voiced when Sanger sequencing was introduced. Currently, the Roche-454 is considered a legacy platform and has paved the way for more advanced and higher-throughput sequencing platforms as described below (19). The Roche-454 systems, including the benchtop GS Junior and the high-end GS FLX+ sequencers, have been surpassed by other less-expensive, higher-throughput platforms with better scalability, and production will be phased out by 2016 (101).

The **Illumina** platforms employ bridge or cluster PCR to amplify adapter-ligated DNA libraries on the surface of a proprietary flow cell (5, 8). Illumina technology uses reversible dye terminator sequencing by synthesis chemistry that involves repetitive cycles of single base incorporation, imaging, and dye chemistry termination (***Figure 4***) (5). Sequencing by synthesis is the most widely adopted NGS technology. Illumina sequencers offer the option of paired-end sequencing that enables reads to be generated from both ends of a single clonal fragment. The forward and reverse reads are aligned as read pairs, which allows for more accurate read alignment and superior indel detection than that provided by a single end of the

pair. With paired-end reads, the distance between the forward and reverse reads is known and allows assembly algorithms to reconstruct the sequence for repetitive sections of DNA, such as homopolymeric or AT/GC-rich regions of the genome. In recent years, Illumina has dominated the sequencing market for both microorganisms and larger organisms because of the platform's high sequence throughput, low error rate, and low sequencing cost per base. Illumina has a line of machines serving a multitude of purposes and sequencing power on all scales (15). The Illumina sequencing platforms include the HiSeq X™, HiSeq, NextSeq, and MiSeq series that range in cost from nearly $10 million for the latest HiSeq X™ Ten to $99,000 for the smaller, benchtop MiSeq system (9, 100).

The **SOLiD** (Sequencing by Oligonucleotide Ligation Detection) platforms by Life Technologies, including the 5500 W series genetic analysis systems, use a sequencing approach known as oligonucleotide ligation detection. The sequencing by ligation method utilizes dinucleotide encoding, which measures every base twice. Because of a two-base sequencing method, the SOLiD sequencing systems can obtain base call accuracies as high as 99.99%. Although SOLiD sequencing is well suited for high-accuracy applications such as genome resequencing and polymorphism detection, these systems are fading platforms (***Figure 5***) (5). Of all available NGS platforms, the SOLiD sequencing method produces the shortest read lengths (35 to 50 bp) and has the longest run times (1 to 2 weeks).

Other popular components of the NGS portfolio are the benchtop **Ion Torrent** platforms by Life

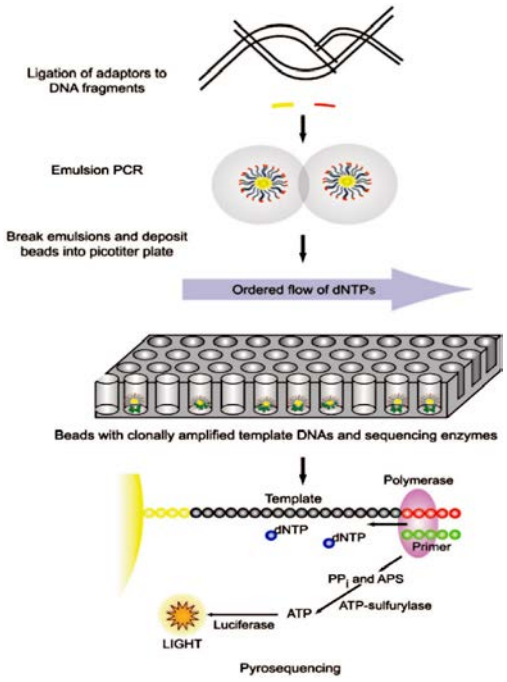


Figure 3. Roche 454 GS FLX sequencing.
Template DNA is fragmented, end-repaired, ligated to adapters, and clonally amplified by emulsion PCR. After amplification, the beads are deposited into picotiter-plate wells with sequencing enzymes. The picotiter plate functions as a flow cell where iterative pyrosequencing is performed. A nucleotide-incorporation event results in pyrophosphate (PP_i) release and well-localized luminescence. APS, adenosine 5'-phosphosulfate.

Republished from reference 5 with permission from the publisher.

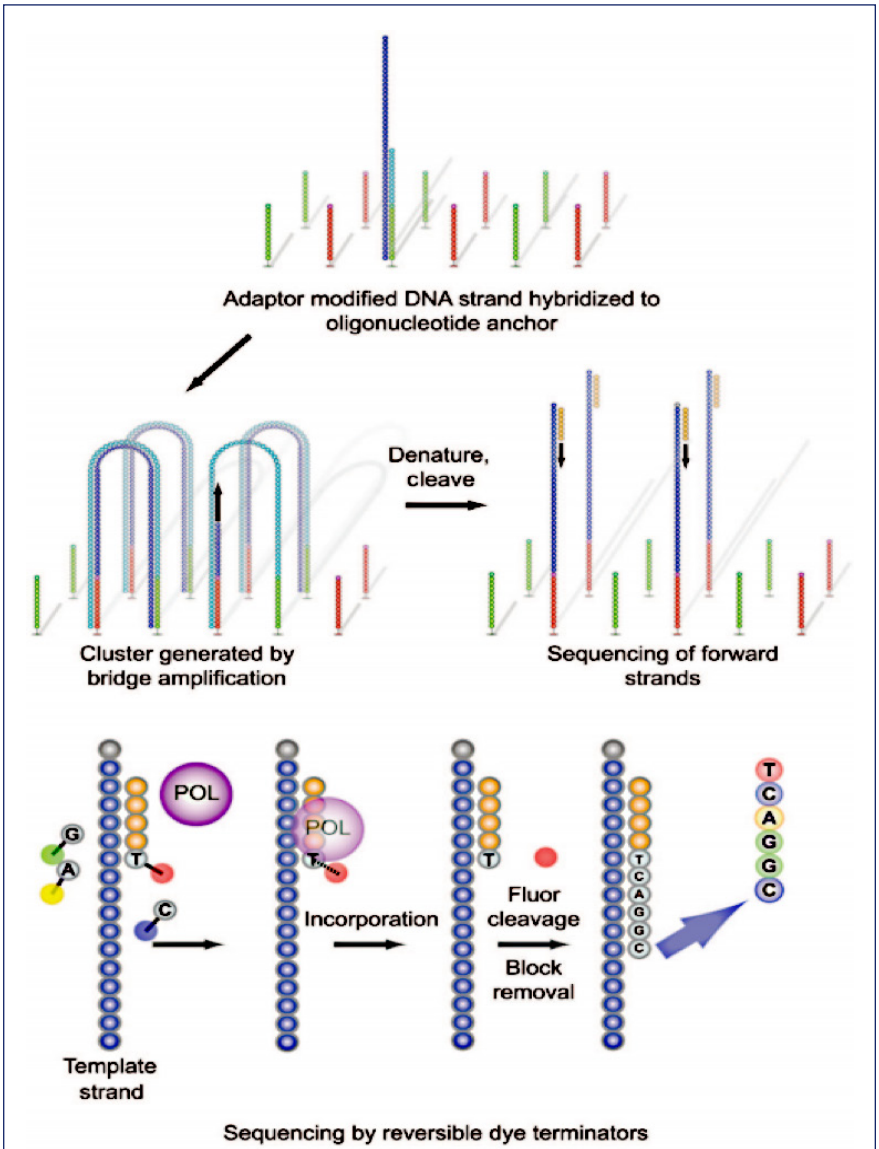


Figure 4. Illumina Genome Analyzer sequencing, the first of the Illumina platforms to be commercially launched.
Adapter-modified, single-stranded DNA is added to the flow cell and immobilized by hybridization. Bridge amplification generates clonally amplified clusters. Clusters are denatured and cleaved; sequencing is initiated with addition of primer, polymerase (POL) and 4 reversible dye terminators. Postincorporation fluorescence is recorded. The fluor and block are removed before the next synthesis cycle.

Republished from reference 5 with permission from the publisher.

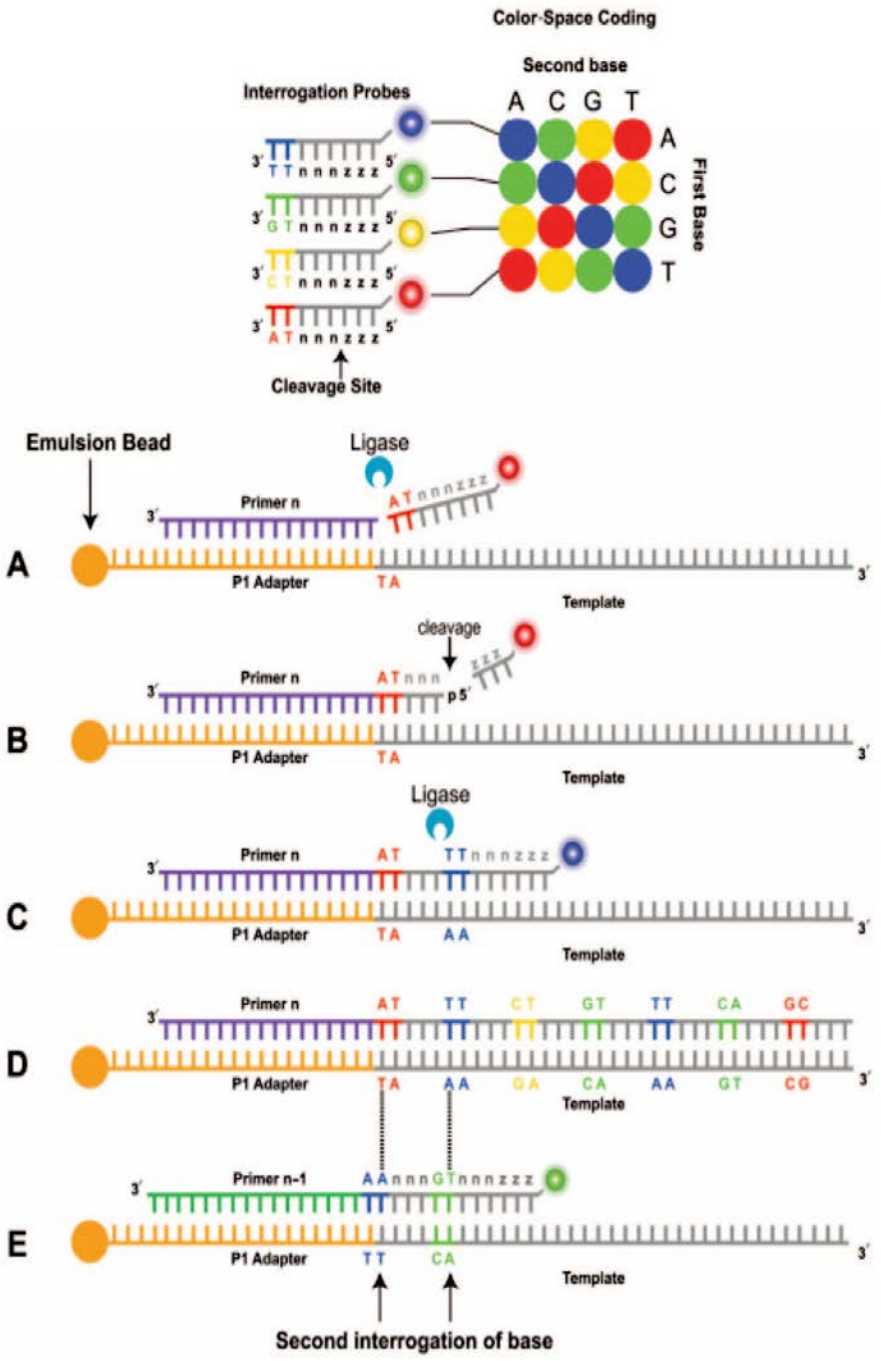


Figure 5. Applied Biosystems (Life Technologies) SOLiD sequencing by ligation.
Top: SOLiD color-space coding. Each interrogation probe is an octamer, which consists of (3'-to-5' direction) 2 probe-specific bases followed by 6 degenerate bases (nnnzzz) with one of 4 fluorescent labels linked to the 5' end. The 2 probe-specific bases consist of one of 16 possible 2-base combinations. Bottom: (**A**), The P1 adapter and template with annealed primer (n) is interrogated by probes representing the 16 possible 2-base combinations. In this example, the 2 specific bases complementary to the template are AT. (**B**), After annealing and ligation of the probe, fluorescence is recorded before cleavage of the last 3 degenerate probe bases. The 5' end of the cleaved probe is phosphorylated (not shown) before the second sequencing step. (**C**), Annealing and ligation of the next probe. (**D**), Complete extension of primer (n) through the first round consisting of 7 cycles of ligation. (**E**), The product extended from primer (n) is denatured from the adapter/template, and the second round of sequencing is performed with primer (n – 1). With the use of progressively offset primers, in this example (n – 1), adapter bases are sequenced, and this known sequence is used in conjunction with the color-space coding for determining the template sequence by deconvolution. In this technology, template bases are interrogated twice.

Technologies, including the Personal Genome Machine (PGM) and the Ion Proton. Although similar to other platforms in their sequencing by synthesis methodology and amplification by emulsion PCR, Ion Torrent platforms differ from other technologies in the detection step. The release of H^+ ions during base incorporation is measured instead of using fluorescence or chemiluminescence detection strategies, making the Ion Torrent sequencers very sensitive pH meters. Both the PGM and the Ion Proton have flexible reagent chips that generate different scales of sequencing output (10 Mb, 100 Mb, 1 Gb, 10 to 100 Gb) depending on the user's desired sequencing coverage (62). These platforms are popular because of the low sequencer costs ($80,000 to $145,000) and speed of sequencing. The entire workflow, from DNA extraction to annotation, can be completed in less than 24 hours, making the Ion Torrent sequencers most suitable for targeted sequencing or smaller genome sequencing projects (100).

Pacific Biosciences (PacBio) is an example of a sequencing instrument that delivers "single-molecule real-time sequencing" (38). Nucleotides are labeled with dye and continuously added to the growing DNA strand by use of a highly processive DNA polymerase. Attached to this enzyme is a waveguide detector that permits the constant monitoring of the incorporated dye-labeled nucleotides. The sequencing reaction takes place on zero-mode waveguide nanostructure arrays (102). Because of its long reads, PacBio is suited to a variety of applications such as *de novo* genome assembly and, hence, the characterization of new organisms along with targeted sequencing to detect base modifications (100). Consensus assembly and long reads are providing new reference sequences in regions of genomes which the short-read sequencers fail to uncover, and longer-read PCR sequencing shows promise for phasing alleles and in viral quasispecies. PacBio may yield long reads of ≥10 kb, but the 87% accuracy of individual reads and the substantial infrastructure requirements are downsides of this platform for clinical applications. The PacBio RSII is a high-end sequencing instrument that is priced at nearly $750,000, a cost that is beyond the reach of the average clinical laboratory. The input DNA requirement of approximately 1 µg is also significantly higher than that of other platforms, which need nanograms or even sub-nanogram quantities (100, 103). Additionally, this quantity may not be obtainable when analyzing clinical samples (25). However, it is well suited for the creation of high-quality reference genomes and produces the most accurate and complete genomes of the current NGS platforms (15).

The first in a new breed of ultrafast DNA sequencing technologies is **Oxford Nanopore,** with nanopore sequencing platforms. A single DNA molecule is guided through a protein nanopore, which results in changes in electrical current across a lipid membrane. Nanopore technology reads native DNA and is predicted to be able to sequence directly from clinical samples with a low abundance of DNA (15). Data generated from the MinION are streamed live to a computer via a USB connection and can potentially provide NGS at the bedside or in the field (104). With Nanopore sequencing, DNA sequence is produced in real time, which enables analysis to be conducted on a continuous stream of long reads (105). Since the MinION sequences a single DNA molecule per pore, it eliminates the library amplification phase and its associated biases, which is a common mechanism used by many NGS technologies. However, at this stage of its development, the MinION still requires the initial library preparation prior to DNA sequencing. The MinION has a low capital cost, and its flow cells that contain the nanopores are to be used only once, which eliminates the tribulations of installation and instrument maintenance. It is anticipated that the MinION will yield 150 Mb of sequence per hour with mean read lengths of ~2 kb. Despite the long reads, the MinION has an error rate of 15 to 22%, which currently limits its ability to compete with other NGS platforms for many applications, although improvements to the base-calling algorithm are yielding benefits (104, 106, 107, 108). So-called "2D" nanopore reads, where both strands of a read are processed by the nanopore, have error rates below 20% (104, 107). In addition to the MinION, the Oxford Nanopore

Text box 1. Common nomenclature used throughout the report

16S rRNA sequencing – the process by which the 16S ribosomal gene, a gene common to all bacterial species, is sequenced to help identify and compare bacteria present within a sample

Analytical sensitivity – the lowest concentration of a given substance in a biological sample that can be accurately measured by an assay

Analytical specificity – the ability of an assay to measure a particular organism or substance, rather than others, in a biological sample

Base calling – the process that converts raw signal into a nucleobase prediction and quality value; the process by which the order of bases in a template is determined during a sequencing reaction; a base-calling platform that is designed to reduce platform-specific issues should be used

Clinical diagnostic sensitivity – the percentage of samples that have a given pathogen and are identified by the assay as positive for the pathogen

Clinical diagnostic specificity – the percentage of samples that do not have a given pathogen and are identified by the assay as being negative for the pathogen

Contig (continguous sequences) – the assembly of several regions of a genome; the size and continuity of contigs directly influence gap formation, regions where no reads align

Coverage – the number of reads that cover a given base position; the overlapping of reads enables each section of the genome to be covered by multiple reads

***De novo* sequencing** – the assembly of sequence reads into a contiguous sequence, or into a scaffold without the use of a genomic reference or gold standard reference sequence; this method is commonly used for sequencing novel microbial genomes and epidemiological applications

Disruptive innovation – NGS represents a cost-effective platform technology that can be substituted for a range of traditional microbiological tests performed in the laboratory

Dry components – the steps involved in analyzing sequenced DNA; the bioinformatics process

Epigenetics – the study of heritable changes in gene regulation occurring without a change in the DNA sequence encompassing DNA methylation, small RNA-mediated regulation, and DNA/protein interactions

FASTQ – a type of data format applied to NGS that stores both the biological sequence and the corresponding quality scores associated with the generated sequence

Genome annotation – process of assigning biological information or function to the final assembled sequence

Genome assembly – process by which multiple, fragmented sequence reads generated from an NGS platform are assembled to reconstruct the original sequence; achieved by *de novo* assembly or with the use of a reference sequence (when a reference sequence is used, this is typically referred to as **genome alignment**, not assembly)

Indel – structural biology term referring to an insertion or deletion in a segment of DNA

Limit of detection – a measurement of analytical sensitivity for a specific target; the lowest concentration of target that can be sequenced reliably and distinguishably from negative specimens and that is consistently detected in ≥95% of specimen replicates

Metagenomics – the study of microbial communities from complex samples (i.e., environmental or clinical) that could help in novel organism discovery or exemplify the dynamics of a population under different conditions; all DNA content in a clinical sample is sequenced before bioinformatic analyses are used to filter out human and nonpathogenic organism DNA to identify the causative agent; also known as a random, unbiased, or shotgun sequencing approach

NGS (next-generation sequencing) – newer, non-Sanger-based methods of sequencing with higher-throughput capabilities that are produced at a lower cost per base than that by Sanger sequencing; these methods encompass "second-" and "third-generation sequencing"; also known as high-throughput sequencing and massively parallel sequencing

Phred score (Quality or Q score) – the metric used to assess base-calling accuracy that assigns a level of certainty; a measure of the quality of the identification of nucleobases generated from automated DNA sequencing

Reads – DNA sequences that range in length (small reads, 30 to 500 bp; long reads, >5 kb) and are generated from NGS platforms; each read represents a small fragment of the genome that includes the bacterial chromosome, phage sequences, and plasmid sequences

Reproducibility – assesses the variability when the same material is evaluated and multiple variables are introduced

Sanger sequencing – "first-generation sequencing" that uses capillary electrophoresis methods to sequence DNA fragments; method is low throughput and expensive

Unbiased sequencing – a sequencing approach that amplifies both host and microbial nucleic acids; also referred to as agnostic sequencing or metagenomics

Wet components – the sample preparation process for DNA sequencing; procedures completed on the benchside

Whole-exome sequencing – the sequencing of all protein-encoding regions (i.e., exons) within a genome

WGS (whole-genome sequencing) – construction of the complete nucleotide sequence of an organism

PromethION and GridION, which provide benchtop throughput analyses and scalable throughput using stackable desktop devices, respectively, are under development. Interestingly, Oxford Nanopore has promised a "pay-as-you-go" model for these devices, eliminating upfront infrastructure costs, which could hasten their clinical adoption. However, there is a lot of uncertainty around Nanopore technology. As with any emerging technology, it is difficult to separate promises from the true capabilities of the instrument (15, 22, 105).

There are many sequencing manufacturers that are competing for sequencing supremacy (109). As more companies enter the NGS race, it is probable that instrument purchase costs will decline and democratization of NGS will occur. Currently, the purchase of NGS instrumentation is a significant financial obligation and commitment (9). It is paramount to assess the needs and demands placed on a laboratory to determine which platform is most suitable and fits within the specified budget (25). If the goal is to incorporate NGS into routine use in the clinical microbiology laboratory, the machine must be able to handle the projected daily volume of isolates and to yield cost-effective and actionable results in a clinically relevant time frame (9, 10, 27, 28, 110).

1c. Main Steps of NGS Workflow

Even though the technical specifics of NGS technologies differ, there are several common steps that are shared among the majority of high-throughput sequencing methods, with the exception of single-molecule real-time NGS as described below (1, 5). Typically, the NGS workflow in a clinical laboratory includes sample collection/nucleic acid extraction, NGS library preparation, sequencing, data analysis, and data storage (7, 8, 25, 111). The clinical sample, e.g., swab, sputum, stool, urine, cerebrospinal fluid (CSF), or tissue, contains the nucleic acid (DNA/RNA) of interest and is placed in a transport device for delivery to the clinical microbiology laboratory. Clinical specimen types for NGS evaluation will differ depending on the patient's clinical syndrome but should be collected from appropriate sources during disease progression. The type of transport device, storage time, and temperature can influence the quantity and integrity of the nucleic acids present. Many transport mechanisms are not microbial DNA free, raising the possibility of false positives and the introduction of background contaminants.

Nucleic acids can be prepared from clinical samples by using a variety of methodologies, some of which are dependent on the NGS system being used. Suitable extraction methods are essential for a successful result and thereby help to lessen the introduction of biases and false negatives. Once the input nucleic acid (e.g., genomic DNA, reverse-transcribed RNA or cDNA, immunoprecipitated DNA) is obtained, the genetic material is fragmented by different methods, such as sonication, nebulization, or enzymatic digestions (5, 8, 112). These fragments are then ligated to platform-specific oligonucleotide adapters to create a library of overlapping sequences (7). The library is then hybridized to beads or a flow cell, which is followed by clonal amplification, such as emulsion PCR or bridge amplification. Not all platforms require the clonal amplification phase or preparation of a DNA library (8). Enrichment procedures can also be completed at this stage to help select for a specific type of DNA if an organism is suspected. Depending upon the NGS platform, the clonally amplified templates are sequenced by various chemistries, such as pyrosequencing, reversible dye terminators, oligonucleotide probe ligation, and phospholinked fluorescent nucleotides. Sequence data are then analyzed to determine the composition of the DNA sequences for pathogen identification. The final components of the NGS workflow are data release and dissemination of a clinically actionable report. If the patient's sequencing data would need to be reexamined, appropriate NGS analysis files should be stored or archived on- or off-site with patient privacy/confidentiality upheld (7, 10, 25, 28).

SECTION 1 FINDINGS AND RECOMMENDATIONS

Key Finding 1.1 The use of NGS for fungal, yeast, and parasitic detection has lagged behind bacterial and viral detection efforts. However, this is not to say that bacterial and viral diagnostic applications do not have challenges. The key issues surrounding viral NGS diagnostics are maintaining viral nucleic acid integrity during the extraction phase and utilizing an extraction method that is capable of isolating quasispecies that may be clinically important. Fungi, yeasts, and parasites possess large genomes that complicate data analysis, can be confused with human host reads, and may be present in low titers in the clinical specimen, all of which challenge the NGS process. There are limited databases with clinically important fungi, yeast, and parasite species, further causing these NGS applications to fall behind bacterial and viral NGS testing.

Section 2

Performance Characteristics of Competing Orthologous DNA Technologies

NGS has undeniably reinvigorated the DNA sequencing field, yet it faces competition from other molecular and immunological diagnostic tests that are currently on the market. Each technology, including NGS, has its own benefits and pitfalls. Nonetheless, a recurring theme with NGS, unlike with many of the molecular approaches presented in this section, is that prior knowledge of the suspected organism or genome annotation is not required for pathogen identification. Thereby, pathogen-specific primers are not needed to perform NGS, but high-quality microbial genomic reference sequences are necessary to successfully identify the pathogen (9, 27). Although the principal NGS paradigm is WGS, NGS technologies have additional functionalities, such as gene expression profiling by RNA sequencing (RNA-seq or transcriptome), epigenetics, mapping of DNA binding proteins and chromatin analysis, discovery of noncoding RNAs, exome sequencing, single-nucleotide-polymorphism (SNP) detection, and amplicon resequencing (5, 13, 99). Some of these applications are geared more toward research or human genetic testing laboratories rather than the clinical microbiology laboratory. Nevertheless, some of the available technologies that are currently competing with NGS are described in more detail below (***Table 2***).

The 16S rRNA gene is found in all bacterial genomes and is composed of both highly conserved and divergent regions. The dissimilar regions of the 16S rRNA genes create distinct microbial signatures that allow for molecular identification of bacteria (113). A combination of PCR amplification and sequence analysis of the 16S gene is the accepted reference standard for identifying unknown bacterial species for a single isolate (1, 114,

Table 2. Some technologies in competition with NGS[a]

Methodology[b]	Description
16S rRNA gene sequencing	Reference standard for identifying unknown bacterial species
Microarrays	Multiplex assays performed on a solid surface such as a glass slide or silicon film; requires a reference database or knowledge of genomic features
MALDI-TOF MS	Analyzes the patterns of biomolecules (i.e., proteins) produced, which is both cheap and rapid; requires a reference database
Multiplex PCR – BioFire (see below for description)	A direct, uncultured clinical sample is subjected to multiplex PCR, followed by melting curve analysis, which is able to detect multiple pathogens included on specific panels
Multiplex PCR – GenMark (see below for description)	Uses proprietary eSensor® technology (a combination of DNA hybridization and electrochemical detection) to detect a panel of 14 respiratory viruses
Singleplex PCR assays	Individual PCRs that amplify and detect genes that are pathogen specific
ELISA	A plate-based assay that has an antibody bound to the matrix that will bind antigen from the culture and is detected through a second antibody; "sandwich" assay; due to antibody specificity, ELISAs can target only a specific organism or group of related organisms
PCR-ESI/MS	Measures the mass/charge ratio of PCR-generated amplicons for a specific panel of organisms; not an all-inclusive test
Smarticles™	Technology can detect multidrug-resistant bacteria and provide susceptibility results within a single shift in the clinical laboratory

[a]This table does not represent all possible competitors of NGS.

[b]MALDI-TOF MS, matrix-assisted laser desorption ionization–time of flight mass spectrometry; ELISA, enzyme-linked immunosorbent assay; PCR-ESI/MS, PCR-electrospray ionization/mass spectrometry.

115). NGS, however, is the only method for use in mixed-sample sequencing. While whole-genome NGS has the ability to identify a bacterial species and/or subspecies by the use of additional gene markers, 16S rRNA sequencing may not be able to discriminate between organisms with closely related 16S rRNA genes (116, 117, 118). There is no accepted cutoff value of 16S rRNA sequence similarity for species definition. In addition, this method is highly dependent on the sequences provided in a database for correct bacterial identification. If a bacterial species is not listed or is incorrectly labeled in a database, the 16S rRNA sequence may not be able to correctly identify the organism (1).

Similar to 16S rRNA sequencing, the use of microarrays such as Affymetrix tiling arrays also requires knowledge of the query genome or genomic features. Therefore, discovery and metagenomic analyses cannot be performed with microarrays (119). As stated previously, metagenomic sequencing interrogates all DNA present in a sample at one time and does not target specific genes or pathogens (71). Another intrinsic limitation of microarrays is probe cross-hybridization to similar sequences within a genome. Cross-hybridization is not an issue for NGS, as single-nucleotide resolution can distinguish allelic differences within one nucleotide, provided there is sufficient read coverage (33). Hybridization arrays do not provide the richness of data that NGS can produce. The dynamic range and analytical sensitivity are scalable for NGS, a feature that is not applicable to microarrays. NGS measures digital sequencing read counts that can be adjusted based on optimal throughput; however, microarrays measure continuous signal, which limits the detection range due to signal saturation and noise. Even though it is highly probable that NGS platforms will outstrip the applications of microarrays, specific niches in the clinical microbiology laboratory will still be fulfilled by microarrays (10, 120).

MALDI-TOF MS (matrix-assisted laser desorption ionization–time of flight mass spectrometry) can rapidly identify organisms by comparing proteomic profiles of highly conserved proteins to a database of reference protein profiles (***Figure 6A***). Species-specific spectral signatures that can be used to identify microorganisms are produced (9) (***Figure 6B***). This technology is being readily adopted by clinical microbiology laboratories because of its ability to accurately identify a bacterial species from a pure culture within minutes at a running cost of approximately $1 per isolate (1, 121). To utilize the speed of MALDI-TOF MS, traditional laboratory workup to obtain a pure culture of the organism of interest is still required, which can take as little as 4 hours when "scum plates" are used but more commonly takes 2 to 3 days (or longer, depending on the organism). In addition, the $1 cost is somewhat misleading because it does not

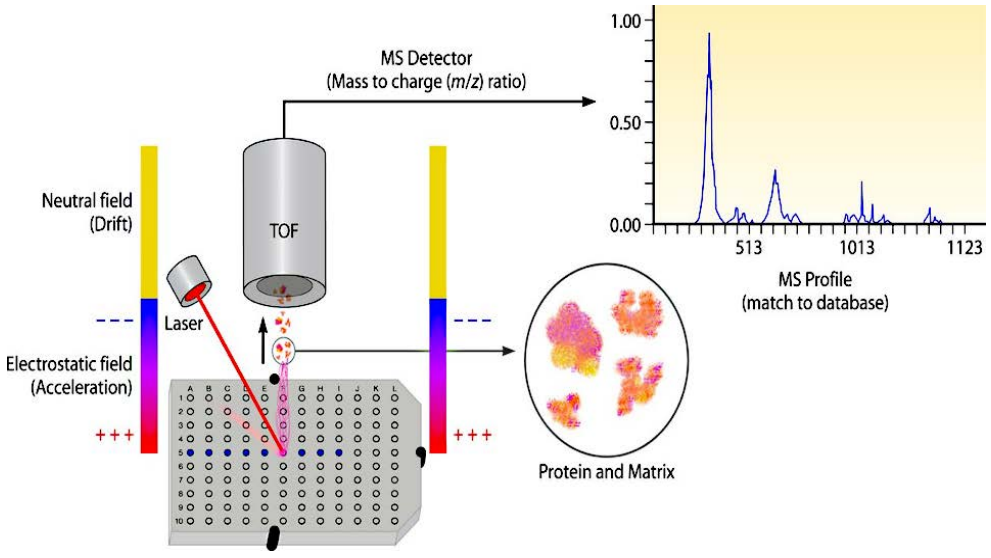


Figure 6A. General schematic for MS analysis of ionized microbiological isolates and clinical material. Once appropriately processed samples are added to the MALDI plate, overlaid with matrix, and dried, the sample is bombarded by the laser. This bombardment results in the sublimation and ionization of both the sample and matrix. These generated ions are separated based on their mass-to-charge ratio via a TOF tube, and a spectral representation of these ions is generated and analyzed by the MS software, generating an MS profile. This profile is subsequently compared to a database of reference MS spectra and matched to either identical or the most related spectra contained in the database, generating an identification for bacteria or yeast contained within the sample.

Republished from reference 124 with permission.

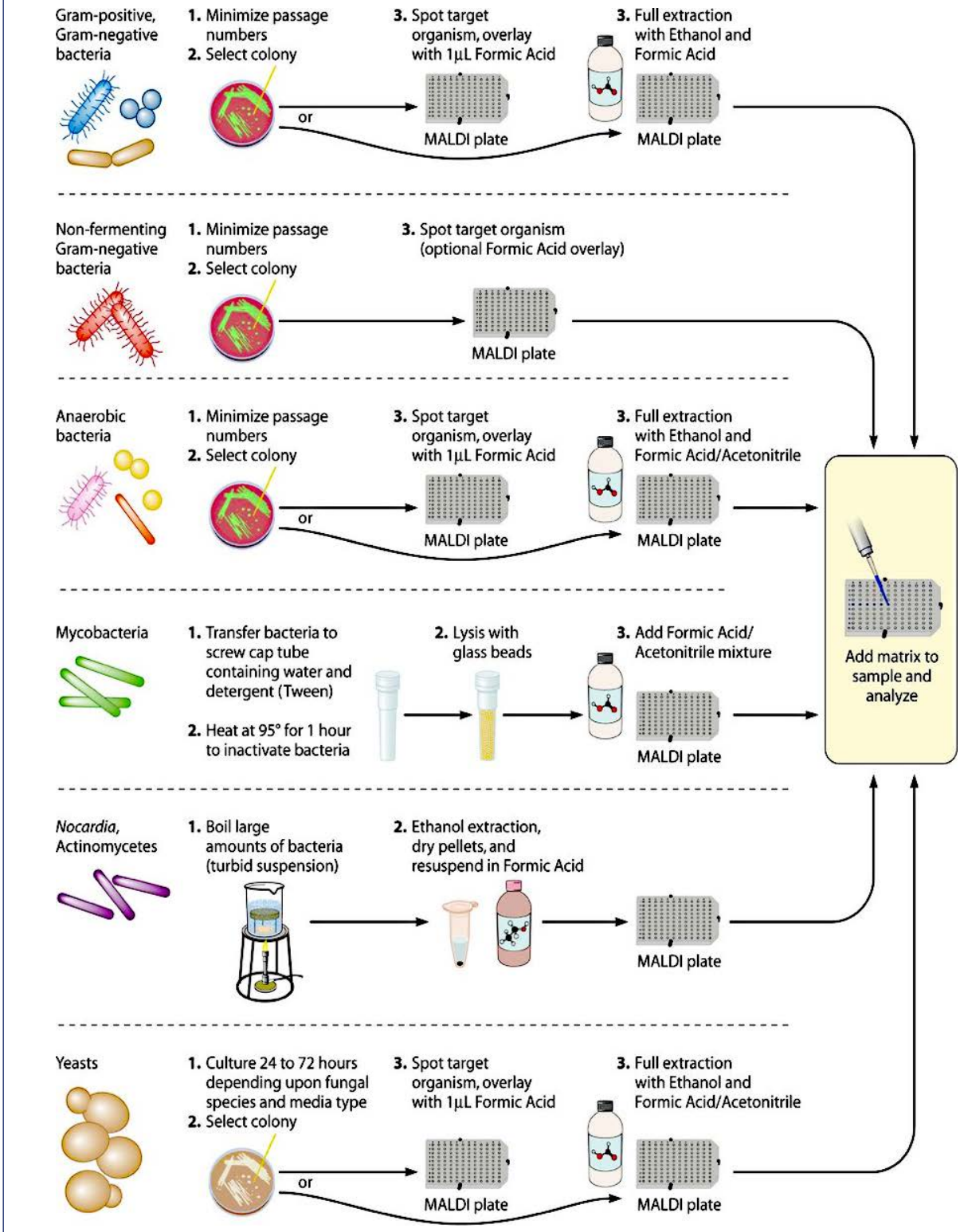


Figure 6B. Mechanisms of MALDI-TOF MS for the identification of bacterial species.
Additional suggestions for MALDI-TOF MS sample preparations for use with different classes of microbes. Different groups of microorganisms vary fundamentally in their cellular composition and architecture. These differences have been demonstrated to affect the quality of spectra generated during MS experiments and, thus, the accuracy of MALDI-TOF MS-derived identifications. As such, investigators from a number of independent studies have evaluated different methods for sample preparation of different groups of microorganisms, ranging directly from intact-cell to full-protein extraction-based methodologies. Results from these studies are summarized here. Proper biological safety precautions should be followed with respect to dangerous members of these groups of organisms.

Republished from reference 124 with permission.

factor in the cost of purchasing and operating the instrument but rather includes only preprocessing steps and is restricted to species representation. Although MALDI-TOF MS has relatively low operational costs, the resolution for differentiating closely related species is weak and similar to the level of 16S rRNA sequencing (95). [**See Case Study 5 to see how WGS analysis discovered a novel *Bacteroides* species that MALDI-TOF identified as *Bacteroides fragilis* (122).**] Additionally, MALDI-TOF MS cannot detect potentially new or novel species because of the reliance on the characterized reference database (123). NGS analysis offers *de novo* assembly which does not require foreknowledge of a sample's composition but needs very-high-quality sequence data (i.e., long reads and significant coverage). Taken as a whole, MALDI-TOF MS is considered useful for culture-based identification but involves large capital expenses and is limited in its potential to identify organisms from direct clinical specimens and in its capacity to provide information regarding antimicrobial resistance (AMR) or virulence (7, 124).

Multiplex PCR methods, especially instruments that have a rapid turnaround time such as the BioFire, are a source of competition for NGS assays. Currently, BioFire offers clinical diagnostic detection for more than 100 bacterial, viral, yeast, and parasitic pathogens. BioFire has three FDA-cleared panels for the detection of pathogens causing respiratory, gastrointestinal, or bloodstream infections. This technology applies multiplex PCR and melting curve analysis to an unprocessed sample, without need for culture. Results are delivered within an hour in a simple, easy-to-read format, but AMR profiles are not included in the data output (125). It may be necessary to perform susceptibility testing on specific isolates if this information is needed for clinical management. Hence, BioFire does not necessarily free a clinical laboratory from culturing. Other multiplex PCR methodologies discussed included the xTAG® and MultiCode®-RT Luminex systems, which employ specific panels for the detection of gastrointestinal pathogens, respiratory viruses, and herpes simplex viruses 1 and 2. In comparison to the quick turnaround time offered by BioFire panels, the abovementioned Luminex assays have slightly longer run times of 4 to 6 hours, which consumes an entire shift in the clinical laboratory (126, 127). Similar to BioFire, Luminex assays are not all-encompassing but use specific primer sets to detect the pathogens represented in the panel.

GenMark also provides multiplex molecular diagnostic testing using their innovative eSensor® technology that is a unique combination of competitive DNA hybridization and electrochemical detection. The respiratory virus panel offered by GenMark has been cleared by the FDA and is able to detect 14 respiratory virus types and subtypes in approximately 4.5 hours from a direct clinical specimen (128). In contrast to GenMark, Luminex, and BioFire multiplex PCR assays, there is no need to develop specific primers to amplify target sequences nor is there a need to continuously alter the primer design to detect new variants with NGS technology (129).

Furthermore, there are diagnostic singleplex PCR and real-time PCR assays that are available with turnaround times of 2 to 3 hours. These assays detect and amplify only a single target and therefore are extremely limited in utility compared to the multiplexing capabilities of NGS. Additionally, detection of conventional PCR products requires agarose gel electrophoresis, a technique that is laborious and not suitable for high throughput. Despite these drawbacks, singleplex PCR assays are easier to design and validate than NGS assays. Similar to singleplex assays, enzyme-linked immunosorbent assays (ELISAs) typically target a single organism or category of organisms due to the specificity of the antibody-antigen interaction. With the high-throughput and automated ELISA systems that are now available for the detection of particular pathogens, the turnaround time for a result is relatively quick (within 2 hours) (130). Generally, singleplex PCRs and ELISAs are not completed routinely in the clinical laboratory but can be batch processed. However, it is likely that NGS assays will initially enter the clinical landscape via a batch processing approach until all issues are addressed for immediate and daily use.

Other technologies that could possibly compete with NGS technologies in terms of fast microbial detection and identification include PCR-electrospray ionization/mass spectrometry (PCR-ESI/MS). This technology uses a semiquantitative analysis to measure the mass/charge ratio (*m*/*z*) of PCR amplicons generated from several microbe-associated loci. Although PCR-ESI/MS has been around for more than 10 years, the technology has not been commercially available. At the current time, clinical trials such as the RADICAL sepsis trial are being conducted to evaluate the methodology (131). In contrast to MALDI-TOF MS, which uses proteomic information for detection down to the species level, PCR-ESI/MS applies genetic information that discriminates down to the representative strain type. The utility of PCR-ESI/MS can potentially be harnessed for epidemiology and infection control and can help to identify unculturable organisms, discover new pathogens, and detect silent mutations, all characteristics that are not offered by MALDI-TOF MS. PCR-ESI/MS has the ability to directly detect microbial DNA from a clinical sample, and unlike most broad-range PCR approaches, this

Case Study 5

WGS identified a novel genomospecies of *Bacteroides* (122)

A 71-year-old man had traveled throughout India for 1 month and developed intense abdominal pain. He was diagnosed with metastatic colon adenocarcinoma and underwent chemotherapy as well as a right hemicolectomy and hepatectomy. Abdominal abscesses developed postoperatively, and along with blood cultures, were positive for *Bacteroides fragilis* (*B. fragilis*) with resistance to multiple drugs, including metronidazole and imipenem. Whole-genome sequencing (WGS) on the MiSeq platform was used to characterize the patient's clinical isolate. In comparison to the three completed *B. fragilis* reference genomes, the patient's isolate displayed a high level of sequence variation. Sequencing showed that the clinical isolate was a genomospecies distinct from *B. fragilis*, but 16S rRNA sequencing and MALDI-TOF MS identified the isolate as *B. fragilis*. As demonstrated by this case, WGS has the potential to enhance diagnostic accuracy in identifying clinical microbial isolates.

Case Study 6

Determination of an anthrax outbreak among European heroin users by WGS technologies (137)

To elucidate the outbreak origin of unusual cases of injectional anthrax in European heroin users, WGS and canonical SNP (canSNP) genotyping were utilized. *Bacillus anthracis* (*B. anthracis*) isolates were collected and sequenced during the outbreak and compared to an extensive collection of isolates from abroad. *B. anthracis* has a very low mutation rate, and therefore phylogenetic analysis using SNPs is an accurate method for examining patterns of relatedness among the global *B. anthracis* phylogenetic tree. The majority of the world's heroin supply originates in Afghanistan and is then trafficked into western Europe. It was speculated that the contamination of heroin with anthrax spores had occurred at the primary Afghanistan site. The genotypic results demonstrated that the anthrax outbreak that affected heroin users in Scotland, Germany, and England was due to a single *B. anthracis* strain that clustered more closely with strains from Turkey rather than Afghanistan. Therefore, the genomic analyses revealed that the heroin did not become contaminated in Afghanistan but was most likely introduced by accident into the supply along the trafficking route between 2009 and 2010.

technology readily detects polymicrobial infections (132). A PCR-ESI/MS instrument is a significant expenditure that also has high processing costs of $50 to $100 per sample, features that can hinder its full clinical value (132, 133).

GeneWEAVE has instituted Smarticles™ technology, another new class of molecular diagnostics that can detect multidrug-resistant organisms and determine antibiotic susceptibility directly from a patient sample within a single clinical laboratory shift. Smarticles™ is attempting to provide rule-in therapy for treatment guidance with a rapid turnaround time, a concept that is relatively premature with NGS technology. This diagnostic test is marketed as "sample in, susceptibility out" but can only detect a specific species, genus, or family of bacteria. Therefore, Smarticles™ is not an all-inclusive test and cannot detect genetic material across organismal boundaries like NGS methodology (134).

Compared to existing technologies, NGS generates a wealth of data and enables detection of a broader scope of targets (7, 25). These technologies can be applied to a variety of potential applications. For example, NGS facilitates more precise genotyping and allows for better characterization of organisms discovered in clinical samples. More recently, it has been found that RNA-seq analyses from NGS data may allow for host profiling in response to a specific type of infection. It is hypothesized that the host immune response may differentiate between various types of infections, such as viral versus bacterial infections. Host transcriptome biomarker discovery efforts for chronic diseases, biothreat agents, Ebola virus, and influenza virus are under way (see more in "Future Directions" below) (135, 136). The power of NGS may enable the identification of SNPs that could be helpful in determining strain relatedness. NGS could unveil resistance and/or virulence mutations in the pathogen that are clinically significant (28, 33). [**See Case Study 6 to understand how WGS coupled with SNP genotyping revealed an unexpected source of anthrax spores that contaminated a heroin supply line and resulted in peculiar cases of injectional anthrax (137).**]

2a. Sample to Result: NGS Turnaround Times

Rapid diagnostic testing on direct clinical samples has been documented using PCR- and MALDI-TOF MS-specific methods (123, 138). In addition, BioFire technologies can be applied directly to a clinical sample, eliminating the need for isolated and purified microorganisms. BioFire panels are equipped with the coveted "easy button," which allows for "sample in, answer out" testing. PCR-ESI/MS is a technology that can also be applied directly to a clinical specimen. Initially, the application of NGS required the isolation of a pure bacterial species from culture in order to deliver key diagnostic information; however, many groups ventured away from this approach and explored the use of NGS for direct detection from clinical samples (29). If feasible, direct clinical specimen sequencing or metagenomic sequencing could theoretically reduce turnaround times from days or weeks to only a few hours, making NGS a procedure that could be completed within an average clinical laboratory workday (***Figure 7***). Although studies that applied NGS to crude clinical specimens have been published, there are still hurdles, including contaminating normal human microbiota and low-copy-number pathogens that require further evaluation (29, 83, 95, 139, 140, 141). Because NGS technologies sequence both viable and nonviable organisms in a sample, more efforts are needed to establish a normal baseline versus contaminants versus infectious agents. Moreover, the application of NGS to determine a clinical answer does not require knowledge of the infecting pathogen(s). NGS can establish a cause of infection and provide a potential answer in cases where other technologies may not provide an actionable finding (21, 36, 70, 71, 96, 142).

In principle, NGS data should not only detect the invading pathogen but also predict phenotypic resistance through the examination of genetic determinants of AMR. Traditional antimicrobial susceptibility testing requires an extra day of laboratory workup, extending the turnaround time to 3 or more days. Genetic information obtained from an NGS approach would ideally prompt rapid antibiotic treatment decision-making for the clinician and the patient. However, NGS is not at the stage where phenotypic susceptibility data can be extracted regularly or reliably from the genotype, although recent works of Pecora *et al* (143) and Tyson *et al* (144) are optimistic about this possibility. Genotypic data do not necessarily correlate to a clinical phenotype, and some types of AMR have nothing to do with genotype, such as intrinsically resistant microorganisms. For example, most Gram-negative rod bacteria are intrinsically resistant to vancomycin because the large-polypeptide antibiotic cannot penetrate the outer membranes of these organisms. Furthermore, detection of antibiotic resistance genes in some organisms, such as *mecA* in methicillin-resistant *Staphylococcus aureus* (MRSA), is more reliable than resistance gene detection in more challenging organisms, such as *Pseudomonas aeruginosa*. Hence, standardized growth-based susceptibility testing and perhaps newer, rapid phenotypic testing methods will likely be necessary to confirm an NGS result in the foreseeable future. A transcriptome-proteome combination could also assist in extrapolating the genotypic and phenotypic connections. Overall,

NGS can provide more information than is achievable by other methods.

Sequencing run times continue to progressively decrease as technology evolves. The first bacterial genome to be fully sequenced using Sanger sequencing was *Haemophilus influenzae*, which required more than 1 year to complete at an estimated cost of $1 million (9, 24, 25, 31, 57, 145, 146). Now, depending on the platform used, an entire bacterial genome can be sequenced in less than 1 day at a cost of less than $100 (10, 31, 39). However, compared to existing molecular diagnostic assays that can take minutes to a few hours, current NGS tests are tremendously slow. Although NGS technologies have longer run times, the trade-off is the comprehensive genomic data that are produced. With this barrage of sequencing data comes the challenge of elucidating tangible information that would be desired by the clinician.

2b. Sensitivity and Specificity Differences between NGS and Other Available Technologies

Just as there are differences in the volume of genomic data generated, there are also differences in the quality of NGS data that are generated from the various platforms. The same confidence in base calling cannot be applied to all technologies. Performance metrics such as read length, accuracy, and sequence output (coverage) vary between platforms and dictate the type of applications that can be performed. Laboratories will have different uses for these platforms, and therefore a comparison of the base calling and error rates among the different NGS systems is not feasible or useful. Because there are systematic biases between NGS platforms, it is difficult to assess the difference in specificity and sensitivity of the platforms in their abilities to detect etiological agents when com-

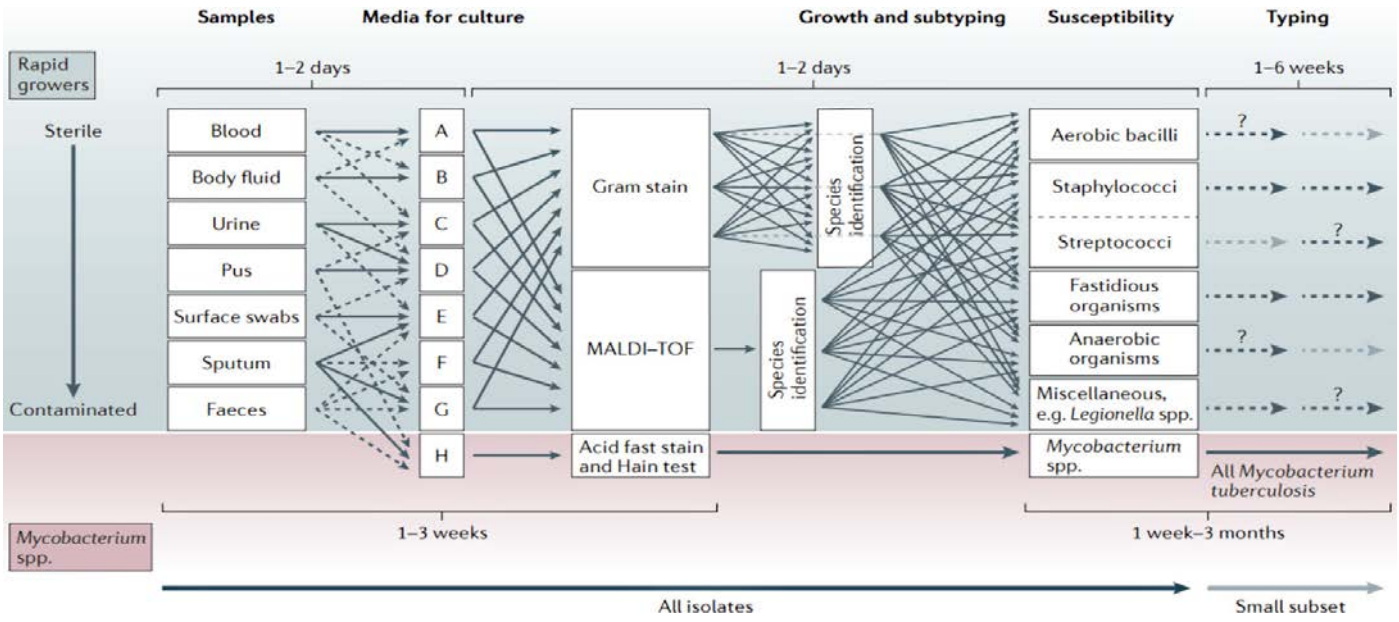


Figure 7. Principals of current processing of bacterial pathogens in the clinical microbiology laboratory.
A schematic representation of the current workflow for processing samples for bacterial pathogens is presented, showing high complexity and a typical timescale of a few weeks to a few months depending on the growth characteristics of the organism. The schematic is an approximation that highlights the principal steps in the workflow; it is not intended to be a comprehensive or precise description. Samples that are likely to be normally sterile are often cultured on a rich medium that will support the growth of any culturable organism. Samples that are contaminated with colonizing flora present a challenge for growing the infecting pathogen. Boxes A to H arbitrarily represent the many different media for culture. When an organism is growing, the morphological appearance and density of growth are properties that need specialist knowledge for deciding whether it is likely to be pathogenic. The likely pathogens are then processed through a complex pathway that has many contingencies to determine species and antimicrobial susceptibility. Broadly, there are two approaches. One approach uses matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) mass spectrometry for species identification before susceptibility testing is set up. The other uses Gram staining followed by biochemical testing to determine species; susceptibility testing is often set up simultaneously with doing biochemical tests. Categorization of pathogens into groups of species is needed to choose the appropriate susceptibility testing panel. Finally, depending on the species and perceived likelihood of an outbreak, a small subset of isolates may be chosen for further investigation using a wide range of typing tests that are often only provided by reference laboratories. The dashed lines and question marks are positioned arbitrarily to indicate that the further investigation is varied and happens in only a small number of cases.

Republished from reference 30 with permission from the publisher.

pared to the detection abilities of other DNA technologies. With each successive generation of sequencing platforms, performance continues to improve along with overall sensitivity. Better-targeted enrichment procedures are needed, and elimination of contaminating host DNA is critical for extracting higher-quality DNA. In general, there is a higher error rate associated with NGS than with traditional Sanger sequencing, mainly due to ambiguity in short read sequence alignment, resulting in inadequate coverage (14, 147). Plentiful short reads are generated, and to accurately assemble a sequence, analyzers rely on redundancy of sequence coverage from the short read structures, which can be problematic in areas with sequence repeats. Despite the huge potential of NGS, it is important to understand that the coverage and accuracy are not 100% and might lead to false positives, false negatives, and the misidentification of indels (112). Regardless of the flaws of NGS, it does have a higher analytical sensitivity for detecting low-frequency mutations than Sanger sequencing, whose sensitivity of detection is only 17 to 25% (148, 149). NGS technology is still maturing, and with additional fine-tuning of sequencing chemistries, optics, and processing algorithms, its accuracy will improve.

SECTION 2 FINDINGS AND RECOMMENDATIONS

Key Finding 2.1 Multiplex panels and MALDI-TOF MS are the main sources of competition for NGS, with MALDI-TOF MS being primarily for the identification of isolated, cultured colonies. Some multiplex PCR panels that are currently available can detect an array of pathogens without requiring a pure culture. MALDI-TOF MS has been incorporated into clinical laboratories, while NGS still remains more of a "service model." The turnaround time for MALDI-TOF MS is minutes once a pure culture has been obtained. MALDI-TOF MS has proven to be useful for the identification of organisms from pure culture with the assistance of a reference database. A key advantage of NGS is the option to perform *de novo* genome assembly, which does not require a genomic reference or prior knowledge of the target organism, but organism identification still requires a reference database.

Key Recommendation 2.1 In order for NGS-based assays to become commonplace in the clinical microbiology laboratory, there is a need for the development of "turnkey" solutions for all phases of testing (i.e., sample preparation, sequencing, data analysis, and result interpretation). The ultimate goal of diagnostic NGS is to place a direct clinical specimen from any matrix into the NGS workflow and generate an actionable result within a reasonable time frame. Continued efforts for direct clinical sample sequencing should be pursued.

Key Finding 2.2 There are platform-dependent errors associated with different sequencing chemistries, and thus, the base calling/error rate sensitivities and specificities of sequencing platforms cannot be directly compared to those of other types of DNA technologies.

Key Recommendation 2.2 It is recommended that a distinction be made between diagnostic clinical specificity/sensitivity and analytical specificity/sensitivity when discussing a clinical microbiological NGS test. The qualifiers of "diagnostic clinical" and "analytical" are not interchangeable, and confusion can arise when reporting a laboratory test result (27). More efforts are also needed to understand the mutation rates and population structures of commonly encountered clinical pathogens in relationship to their effects on NGS sensitivity and specificity.

Section 3
Impact of NGS on Clinical and Public Health Microbiology

Clinical and public health microbiology laboratories help to lessen the burden of infectious disease by detecting and characterizing pathogens in infected patients or those pathogens circulating in the community. Implementation of NGS in these environments has the power to inform clinical and public health decisions by determining the causative agent of infectious disease and/or the epidemiology and evolution of various infecting pathogens in the hospital or community settings (25). Although this technology has a multitude of benefits, it is not a first-line method in the majority of clinical and public health laboratories. At the current time, NGS is a method of last resort, deployed after all other standard diagnostic tests have been exhausted. [**Case Study 7 highlights an unusual presentation of progressive encephalitis that required the use of unbiased NGS because none of the conventional diagnostic assays revealed an infectious etiology (96).**]

Generally, the clinical laboratory workflow is divided into four cardinal stages, including pathogen detection, identification, drug susceptibility, and epidemiological typing (28, 30) (***Figure 7***). Bacterial and fungal isolates are detected through these steps, but virus detection relies on PCR-based assays that are species specific. Nonetheless, each step involves a range of specialized tests that must be performed individually on each isolated organism (95). As specimens enter the clinical laboratory workflow, there is a subsequent increase in the involvement of the hands-on technician at each successive step of the process (10, 27, 30). Additional challenges are posed by particular organisms, some of which may be of critical public health importance. For instance, *Mycobacterium tuberculosis* complex (MTBC) bacteria are extremely slow growing, and it may take weeks to 1 to 2 months to achieve susceptibility results. This extended turnaround time from clinical sample to result can delay appropriate treatment and negatively impact patient outcome (96). Many etiological agents, such as the *Cystoisospora* parasite, *Borrelia burgdorferi* (causative agent of Lyme disease), *Bartonella* species, *Mycobacterium leprae*, *Naegleria fowleri* (primary amebic meningoencephalitis), and HIV elude conventional testing altogether (21, 96, 129, 150, 151). Hence, NGS technologies can provide an alternative mechanism for the identification of unculturable or difficult-to-culture microorganisms, including the abovementioned organisms along with fastidious bacteria, anaerobes, and possible bioterrorism agents.

When it comes to patient care, time is critical. The quest to identify and diagnose an infection can waste precious time for a patient, as in the case of encephalitis. Up to 60% of acute encephalitis cases go undiagnosed due to a lack of assays that can test for the more than 100 etiological agents that can cause encephalitis. Thus, not having a definitive diagnosis instills apprehension and worry for the patient and family (96). Clinicians are commonly forced to make an educated guess about therapy prior to knowing the infecting pathogen, and delays in microbe identification increase the risk of ineffective treatment and spread of infection. Broad-spectrum empirical therapy is commonly administered prior to identification of the etiological agent. Such antimicrobials can cause "collateral damage" by eliminating helpful gut microbiota and subsequently giving rise to resistance development (116, 146). Ultimately, faster and more reliable detection methods are needed, of which NGS holds significant promise.

Rather than a variety of individual tests being required to identify a pathogen, NGS offers a wide diagnostic repertoire that has the capabilities of identifying the culprit no matter the organism—bacterium, virus, fungus, yeast, or parasite (116). Unbiased or agnostic NGS amplifies all nucleic acid present in a clinical sample, including both host and microbial genetic material, without requiring primers for targeted amplification (21, 71, 86, 129). NGS applications can potentially generate microbial sequence data for real-time patient management (***Figure 8***). As a result, NGS methodology has tremendous potential to impact patient care by helping clinicians customize and narrow patient treatment, therefore reducing the usage of ineffective drugs and decreasing the selective pressure for resistance development (116). A compelling and notable application of NGS is expedited turnaround time for difficult-to-culture organisms, such as HIV. Phenotypic and genotypic HIV-specific assays have low sensitivity and are not able to detect low-abundance variants within the viral quasispecies that may express resistance to antiretroviral drugs (13, 56). Because NGS provides a deep sequencing dive at levels far superior to those of Sanger sequencing, the low-level-resistance variants can be detected, which would allow one to monitor the development of resistance to determine at which point retroviral therapy would need to be changed (152). There is evidence that these populations can be detected at frequencies of less than 1%. Other studies have demonstrated

Case Study 7

Encephalitis caused by a neuroinvasive astrovirus infection detected using unbiased NGS (96)

Approximately 1 month following an allogeneic bone marrow transplant, a 42-year-old man developed lymphopenia, diarrhea, and aggressive sensorineural deafness resulting in bilateral hearing loss. The results of a brain magnetic resonance image (MRI) and enterovirus- and herpesvirus-specific PCR were negative. Microbiological testing on frontal lobe biopsy tissue was unrevealing. Because a viral etiological agent was suspected, the patient was treated with high-dose glucocorticoids and intravenous immunoglobulin. Due to the difficulty diagnosing the patient along with his progressive neurological deterioration, unbiased NGS was used on a CSF sample and brain biopsy tissue. Sequencing results displayed a neuroinvasive infection with an astrovirus belonging to the recently discovered VA/HMO clade that is distinct from human astrovirus clades 1 to 8 which are targeted by conventional reverse transcription-PCR assays. Despite the NGS diagnosis, treatment did not stop the progression of infection and the patient died. However, there is no approved therapy for astrovirus encephalitis. This report demonstrates that NGS has the capability of detecting all potential pathogens simultaneously.

the use of NGS in detecting mutations among variant viral populations of hepatitis B virus, hepatitis C virus, and influenza virus that confer resistance (13, 79, 129, 153).

3a. Use of NGS in Outbreak Scenarios

Despite concerted efforts to improve infection control practices, outbreaks can occur in the hospital setting. With the implementation of NGS, outbreak isolates may be sequenced in real time and guide infection control efforts and antibiotic stewardship (56). Vital to containing outbreaks is fast identification of the infecting pathogen and contaminated source (24). [**See Case Study 8 for a description of a MRSA hospital outbreak in a special-care baby unit and how NGS was successfully applied to determine the source (154).**] Because NGS holds promise for faster pathogen detection than traditional laboratory workup, this opens the opportunity for quicker intervention strategies, such as patient isolation, contact precautions, or decolonization. With the ability to extract vital pathogen information, NGS has direct clinical value and can contribute to individual and improved patient management through a more accurate diagnosis and tailored therapy. Ideally, these factors would decrease hospital stays, deter the development of AMR, and decrease mortality rates. NGS can also help promote infection control accountability within and between hospitals and thus help curtail the spread and infection rates of high-risk organisms in health care settings (15, 97, 154, 155). In an increasingly connected world, one could ideally determine if the same bacterial strain was responsible for an outbreak in geographically distinct health care facilities. A downfall of NGS is the lack of standardized guidelines or models endorsed by regulatory agencies for responsible data sharing. There is also hesitation from hospitals due to privacy issues that make open data sharing difficult. If the data are siloed and not shared, there will be limited public health benefit to the use of NGS in the clinic. Besides demonstrating transmission pathways of hospital-acquired pathogens, NGS can provide insight into how bacterial populations respond to drug treatment. Since NGS provides an in-depth interrogation of genomic sequences, genetic variation associated with bacterial virulence

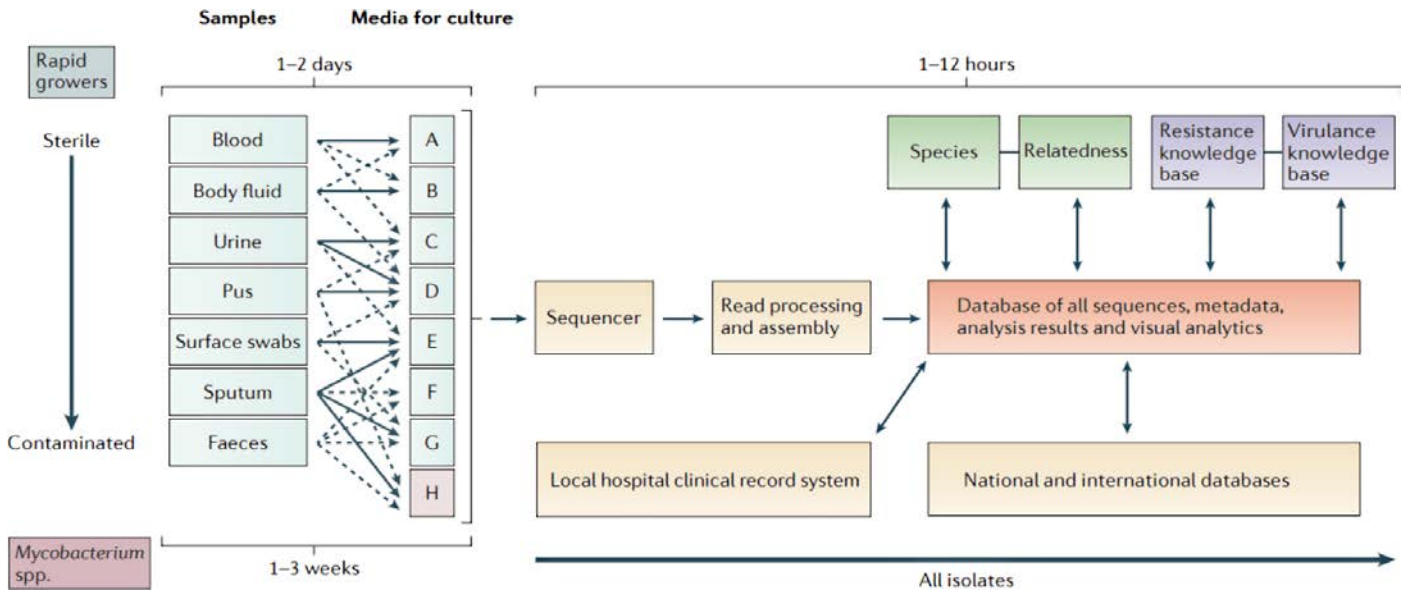


Figure 8 Hypothetical work flow based on whole genome sequencing.
A schematic of the clinical microbiology workflow that is anticipated after adoption of whole genome sequencing is displayed. The culture steps would be the same as those that are currently used in a routine microbiology laboratory. When a sample or likely pathogen is ready for sequencing, DNA will be extracted. This procedure is becoming simpler, as the input required for successful sequencing is reducing; it is now possible to use as little as 5 ng and to purify this in <30 minutes. For current bench-top machines, it can take as little as 2 hours to prepare the DNA for sequencing, and new platforms could enable sequencing without preparation. Therefore, bacterial genome sequencing in hours and possibly even minutes is a realistic prospect. After sequencing, the main processes for yielding information will be computational. The development of software and databases is a major challenge to overcome before pathogen sequencing can be deployed in clinical microbiology. Automated sequence assembly algorithms will be necessary to process the raw sequence data. This assembled sequence would then be analysed by modular software to determine species, relationship to other isolates of the same species, antimicrobial resistance profile and virulence gene content. Results of this analysis will be reported through hospital information systems. All of the results will also be used for outbreak detection and infectious diseases surveillance. These developments will require a new large database and other informatics technology and will take time to develop. In particular, it will need 'intelligent systems', which will incorporate elements of machine learning to allow automatic updating of key knowledge bases for species identification, antimicrobial resistance determination and virulence detection. Formal evaluation of such a solution will also need robust testing to ensure that it performs at least as well as current methods.

or antibiotic resistance could be determined (10, 28, 33, 156). Identifying these key biomarkers is advantageous for infection control teams to enhance their preparedness for current, emerging, or predicted infectious threats (6, 24, 97, 157).

NGS can not only help monitor, detect, and prevent hospital outbreaks, it can also initiate appropriate public health interventions when outbreaks occur in the community and abroad (158). [**Case Study 9 showcases the discriminatory power of NGS over pulsed-field gel electrophoresis (PFGE) in tracking a salmonellosis outbreak that afflicted more than 44 states (157).**] For example, NGS can provide insights into the source and mode of pathogen transmission associated with a foodborne disease outbreak and can subsequently stimulate a proper public health response in the community, such as a food recall (70, 159, 160, 161). Current bacterial typing methods used to examine isolates involved in outbreak situations include PFGE, multilocus sequence typing (MLST), and multilocus variable-number tandem-repeat analyses (MLVA). Roughly one-third of all outbreaks in the United States do not have a determined source because of the low resolution of PFGE, the gold standard method used for trace backs and source tracking (162, 163). These analyses target only selected regions of the microbe's genome, and thus sequencing a whole genome provides superior resolution and detailed linkages for isolates responsible for outbreaks (159, 164). Essentially, NGS provides confidence in the matches among relatively few isolates across time and space, which allows regulators to intervene early and provide links from past isolates from known foods and environmental samples (61, 158).

Outbreak investigators can unambiguously evaluate the relatedness of isolates, and with this information combined with an estimated mutation rate, their evolution from a common ancestor can be calculated (33). The isolate's evolution time and epidemiological data such as the patient's admittance date to the hospital can help predict if a transmission event occurred, which can then direct better targeting of infection control resources (144). NGS has the capacity to identify micro-evolutionary differences among clinical outbreak isolates and essentially "rule in" or "rule out" the links between isolates that would otherwise be indistinguishable by existing approaches (33, 143, 156, 162, 163, 165). Extensive mining of NGS data might reveal novel infectious agents and/or new targets to help bolster outbreak investigations involving highly clonal pathogens. Taken together, the rapid and open release of genomic data has the power to transform our response to outbreaks, especially when data are shared globally and in real time as with the GenomeTrakr project housed at the National Center for Biotechnology Information (NCBI, NLM, NIH) (http://www.ncbi.nlm.nih.gov/bioproject/183844) (105, 158, 166, 167).

3b. Impact of NGS on Metagenomics and Human Microbiome Studies

NGS also impacts human microbiome studies by characterizing the microbes that reside on the outer and inner surfaces of our bodies (168, 169). It is estimated that 90% of the human microbiome cannot be cultured using current technology (31, 56). 16S rRNA metagenomic sequencing, shotgun metagenomic sequencing, and microbial metatranscriptomics are NGS methods that have the

Case Study 8

WGS used to monitor the transmission of MRSA between the community and hospital (154)

WGS was used to retrospectively assess the accuracy of a MRSA infection control investigation at a special-care baby unit (SCBU) at a National Health Service Foundation Trust hospital. The original MRSA outbreak was evaluated using traditional epidemiological data and antibiogram profiles. Using this approach, 12 infants were shown to be colonized with MRSA over a 6-month time period, and a persistent outbreak was suspected but never confirmed. WGS identified MRSA carriage among 26 patients and demonstrated that transmission had occurred within the SCBU, the maternity ward in the hospital, and a wider community. A staff member was colonized with and transmitted the MRSA strain, which allowed the outbreak to persist during times of no known infection despite deep cleaning procedures. Had WGS been used rather than identification of MRSA based on case clustering, the outbreak could have been detected 6 months earlier and could have decreased morbidity and health care costs. This technology has the potential to help in implementing control strategies to curtail the transmission of MRSA and other pathogenic microbes between hospitals and the community.

Case Study 9

Tracking of a salmonellosis outbreak source using NGS (157)

Approximately 300 people from areas spanning 44 states and the District of Columbia became ill with salmonellosis stemming from ingestion of *Salmonella enterica (S. enterica)* serotype Montevideo. An epidemiologic investigation suggested that the cause of the outbreak was spiced salami produced at a New England processing facility. The manufacturer of the salami was also associated with a previous salmonellosis outbreak due to contaminated pistachios, but PFGE could not differentiate between the clinical *Salmonella* isolate from the spiced salami and the *Salmonella* that had originated from the pistachios. Therefore, the cause of the outbreak could have been ingredient supplies, the finished spiced salami product, or the pistachios. NGS was used to sequence 35 *S. enterica* serotype Montevideo genomes collected from the previously mentioned sources. The source of the outbreak was a red and black pepper rub used in the production of the spiced salami.

potential to investigate the human microbiome to determine the wealth of microbes that live with us either symbiotically or antagonistically (47). These NGS methods have the potential to evaluate the intrapersonal diversity of culturable and unculturable microorganisms and how their compositions change with various situations or diseases, such as admittance to the hospital, receiving a transplant (e.g., solid organ, bone marrow, fecal), and vaccination (47, 158, 170). The gene content of the microbes that inhabit specific human niches is approximately 100-fold higher than that of the human genome, encoding various physiological and biochemical functions that could contribute to health or disease states (13). If alterations observed in the microbiome are predictive of disease, perhaps interventions to prevent the onset of disease could occur (46). This available information could be distilled and, in turn, stimulate the development of new *in vitro* diagnostic devices such as a multiplex PCR assay. Although this technology has a multitude of benefits, one concern is the overaggressive interpretation of NGS data, keeping in mind that not every species detected is clinically significant. In general, interpreting microbiome studies is challenging because a standardized methodology has not been established.

Beyond the metagenomic studies of the human body, NGS can be used to research the microbial composition and complex biodiversity of different environments, such as soil, mines, oceans, built environments, and others. A minute fraction of microorganisms isolated from the environment have been able to grow successfully in pure culture. Thus, we lack a true understanding of the diversity that exists on Earth. Metagenomic NGS studies have the potential to discover unknown organisms, etiological agents, and potentially new and original pathogens (171, 172). For metagenomic investigations, DNA is extracted from environmental samples and direct NGS is applied, eliminating the need to culture the organisms beforehand (172). With the hopes of using NGS to discover novel species and subspecies, it is likely that a reevaluation of current taxonomic classification will need to occur (6). An examination of microbial populations and diversity in both clinical and environmental isolates can be performed faster and cheaper with NGS.

SECTION 3 FINDINGS AND RECOMMENDATIONS

Key Finding 3.1 NGS applications can generate microbial sequence data for real-time patient management and data sharing.

Key Recommendation 3.1 To greatly assist in outbreak scenarios in both the hospital and the community, guidelines or models for responsible data sharing among institutions should be developed and endorsed by a consortium of regulatory agencies. These models should encourage continual sharing of microbial genomic data and maximize public availability while balancing the need for patient privacy. This balance of sharing data and maintaining privacy is necessary for predictive outbreak detection to work; hence, the public health benefit of using NGS in the clinic can be gained.

Key Finding 3.2 PFGE, the gold standard method for bacterial typing and outbreak tracking, does not provide the level of resolution or the detailed linkages for outbreak isolates that NGS does.

Key Recommendation 3.2 It is recommended that genomic sequences of emerging microbial pathogens be uploaded to a unified, public database as quickly as possible to allow for community engagement of the data analysis and use of those data to inform other clinical professionals of the pathogens they are encountering in their laboratory. If genomic sequences for high-priority pathogens are routinely deposited, NGS has the potential to serve as the new early warning system for outbreaks that could occur locally, nationally, or internationally. This tactic could help monitor the stability of the outbreak isolate's genome over time and determine if acquisition or removal of genomic information affects diagnostic and therapeutic decision-making.

Section 4

Factors Involved with NGS Implementation into the Workflow of a Clinical Diagnostic Lab

4a. Cost of NGS Technology

Start-up costs and the necessary infrastructure (i.e., storage and retrieval services) to operate an NGS platform constitute a substantial in-house investment that can range from $80,000 to $10 million (5, 6, 24, 38, 173). These prices can easily extend beyond the reach of an average diagnostic laboratory. With the advent of benchtop NGS instrumentation, these sequencers will likely be a better fit for sequencing in the clinical laboratory environment and individual research labs, with their reduced platform size and cost (7, 25, 30). Sequencing systems such as the Illumina MiSeq and the Life Technologies Ion Torrent PGM deliver high-throughput sequencing on the benchtop. However, informatics constitutes a major portion of the NGS cost and is highly dependent on the throughput of the instrument. Sequencing platforms with higher-throughput abilities generally require large Unix-based servers with multiple, high-speed core processors and maintenance by trained information technology professionals. Smaller sequencing platforms with lower throughput can function on high-powered Windows or Unix-based desktop systems (147).

The cost of performing sequencing has decreased dramatically over the past 15 years and has even surpassed Moore's law (10, 174) (***Figure 9***). In contrast, bioinformatics remain a costly component of the analysis process. If an "easy button" or automated data analysis could be designed, then it is anticipated that bioinformatics costs will decline. There are also operational expenses associated with NGS instrumentation, including service contracts for machine maintenance, consumables for library preparation and sequencing, and the labor cost of the medical technologist performing the test (6). Ancillary equipment or robotics could be used to run various applications or to streamline processes, which can be major unforeseen costs (38). For example, it is highly recommended that the integrity and quantity of the DNA be evaluated following the library preparation, such as with a bioanalyzer and fluorometer, respectively. Devices

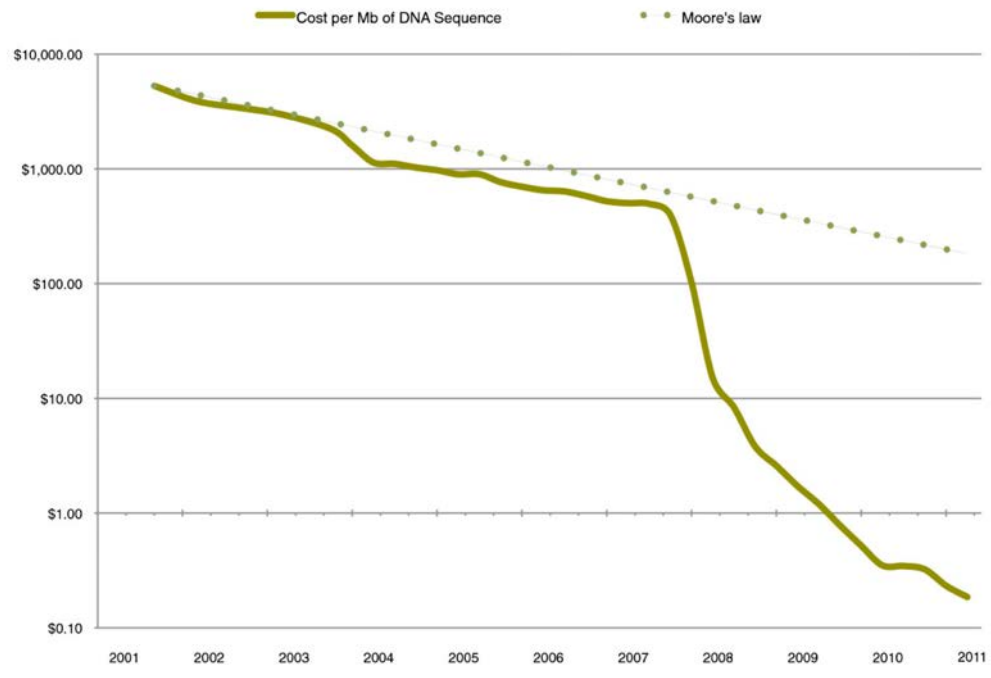


Figure 9. Dramatic decreases in the cost per megabase (Mb) of sequenced DNA compared to the expectation predicted by Moore's law.
The semi-log plot shows a dramatic reduction in the cost per megabase of DNA sequenced from 2001-2011. Sequencing costs have fallen dramatically since 2007 due to competition from multiple vendors. Cost was calculated in January of each year.

Republished from reference 16 with permission.

Is it feasible to routinely use NGS in the clinical microbiology laboratory?

Evaluation of WGS performed on every isolate that entered a clinical laboratory on an average workday

To analyze the feasibility of using WGS in the clinical laboratory, every isolate that was recovered from culture on a single day at Houston Methodist Hospital was sequenced. A total of 130 samples, including 107 aerobic cultures, 9 anaerobic cultures, and 14 acid-fast bacillus/mycology samples, were sequenced and worked up using conventional microbiological methods. One hundred fifteen isolates were correctly identified using WGS. It took approximately 12 hours to extract DNA and prepare the sequencing libraries, 39 hours to perform the sequencing, and 2 to 4 hours for *de novo* assembly of contigs. Isolates that were unable to be identified by sequencing were due to low read counts, insufficient sample preparation, or lack of a reference genome in the nucleotide collection (NT) database (part of the NCBI). This exercise confirmed that WGS can be implemented in the clinical microbiology laboratory to identify unknown organisms from cultured organisms (27).

used for enrichment procedures may also be vital for completion of an NGS assay and need to be factored into the total costs. Therefore, users need to consider the whole process, from sample to sequence interpretation, and the equipment necessary to maximize throughput and prevent sample backlogs (100). Despite these additional expenses, the sequencing cost per base is still significantly lower than that of Sanger sequencing (39). As competition among sequencing vendors intensifies, it is predicted that the price of sequencers, consumables, and analysis will steadily decline. NGS remains a relatively expensive technology compared to other diagnostic assays such as MALDI-TOF MS, which can be performed at a lower cost.

With appropriate scale and efficiency, the U.S. government instituted a goal for the sequencing industry to obtain a $1,000 human genome, which has now been accomplished with select sequencing platforms that are currently available (1, 2, 3, 10, 31). An average bacterial genome consists of approximately 2.8 million nucleotides. This size is less than one thousandth the size of a (haploid) human genome, which is composed of nearly 3 billion nucleotides (7, 9, 33). It is therefore predicted that a whole bacterial genome could be sequenced for $1 in the not too distant future (38, 174). Bacterial genomes are considerably smaller than eukaryotic genomes and house only one haploid chromosome. Despite a smaller genomic content, bacterial species are extremely diverse, with 10% to 40% of their genomes being dispensable due to mobile and accessory elements such as plasmids and transposons (15), which can be gained or lost over time due to selective pressures. In addition to the range of target genome sizes, factors such as genome complexity (e.g., high GC content, homopolymer regions), the anticipated level of completeness, and the availability of a reference strain for comparison could also possibly influence the price of a genome sequencing project as special protocols and/or additional technician time may be required.

NGS must appeal to hospital microbiology laboratories and public health laboratories, with cost being a significant factor impeding its widespread adoption. Depending on the volume of isolates encountered daily in a hospital laboratory, costs can be saved by multiplexing, indexing, or batching isolates with barcodes rather than performing sequencing on individual isolates as they come into the laboratory (6, 38, 43, 175). A unique index "barcode" sequence of 8 to 12 nucleotides that usually makes up part of a platform-specific adapter is added to each DNA fragment created during library preparation and is linked to the organism that is to be sequenced. These indexes allow for downstream *in silico* sorting of the sequences, enabling multiple isolates to be sequenced in one run, which maximizes investment in the instrument and is economically sound. However, with barcoding, there are problems with false positives, cross-contamination, and indice bleeding with off-the-shelf methods.

Nonetheless, there are differences in opinion as to how this technology should be marketed by the sequencing companies. It narrows down to a cost-versus-value analysis. From one perspective, NGS technologies offer a cost-effective catch-all for the diagnosis of infectious diseases. Operationally, NGS is cost efficient, because ideally no other tests would need to be performed once the assay has been optimized and validated. However, there are formidable hurdles to the development of an optimized, analytically and clinically relevant NGS assay. From a different perspective, NGS can be perceived as an expensive test but a value-added test that would fill the gaps from the diagnostics already conducted. With this viewpoint in mind, NGS is meant to enhance the standard of care when a clinical need is identified and results from other tests are elusive.

4b. Standard Operating Procedures for NGS Workflow

Genomic sequencing data generated in the clinical microbiology laboratory need to ultimately translate into meaningful information for patient care and intervention. In order to incorporate NGS into the standard of care, there is a need for consensus practice guidelines for different types of infections. Different protocols for both the wet and dry laboratory components of the process that are specific for the detection of etiological agents, including bacteria, viruses, fungi, yeasts, and parasites, should be created. All protocols must be optimized and internally validated prior to clinical implementation. During validation, specifications for essential performance characteristics (see below) need to be established (14, 37, 39, 176).

Before NGS is used routinely by the broader clinical and public health laboratory community, it is anticipated that NGS will first be adopted by larger academic medical centers and reference laboratories that have the significant financial support and computational infrastructure required for the use of NGS. Furthermore, it is expected that the first applications of NGS used regularly in the clinical setting will be limited to a small number of targeted assays and panels that are validated to perform those specific applications. The development of turnkey solutions, e.g., workflows that include automated sample preparation, standardized bioinformatics pipelines, and market authorization from FDA, would help promote the adoption of NGS in the clinical microbiology and public health laboratory space. NGS assays provide the user

with large amounts of data that require analysis by clinical laboratory professionals to identify and prioritize the findings that are clinically relevant. The development of fully automated sequence interpretation tools would allow for an easier transition into the diagnostic laboratory and could reduce bioinformatics analysis costs (7, 10, 28, 159, 160).

Standardization of NGS protocols and quality metrics is critical to ensure the validity of NGS test results (37, 39, 121). NGS consists of two major processes, the analytical "wet bench" component and the bioinformatics or "dry bench" component. Both processes require separate considerations for the development of standards (39). Any clinical laboratory test should establish analytical performance specifications (i.e., assay validation) of the assay for performance characteristics, which include accuracy, precision, analytic sensitivity and specificity, reproducibility/repeatability, and reportable range (limit of detection) (37, 43, 176). NGS-specific quality control (QC) metrics (e.g., depth of coverage, quality scores, etc.) should be established during test validation and be used as part of the standard operating procedures (SOPs) to monitor the performance of the assay and ensure accurate results (37, 177). In addition, laboratories developing high-complexity tests, like clinical applications that use NGS, should ensure that the tests are appropriate for the intended clinical use (i.e., establishing clinical validity). Quality assurance guidelines and standards for this technology need to be developed and continually refined with insight from clinical microbiologists (15). ***Figure 10*** emphasizes the key considerations in quality assessment for whole-genome NGS analyses (9).

There are a variety of NGS instruments, applications, sample preparation kits, and approaches to data analysis and result interpretation available; therefore, SOPs will not be the same for each lab using NGS. Nevertheless, clinical laboratories must assess their ability to detect analytes of interest and provide an independent verification of test performance compared to that of other laboratories that use the same or different methods

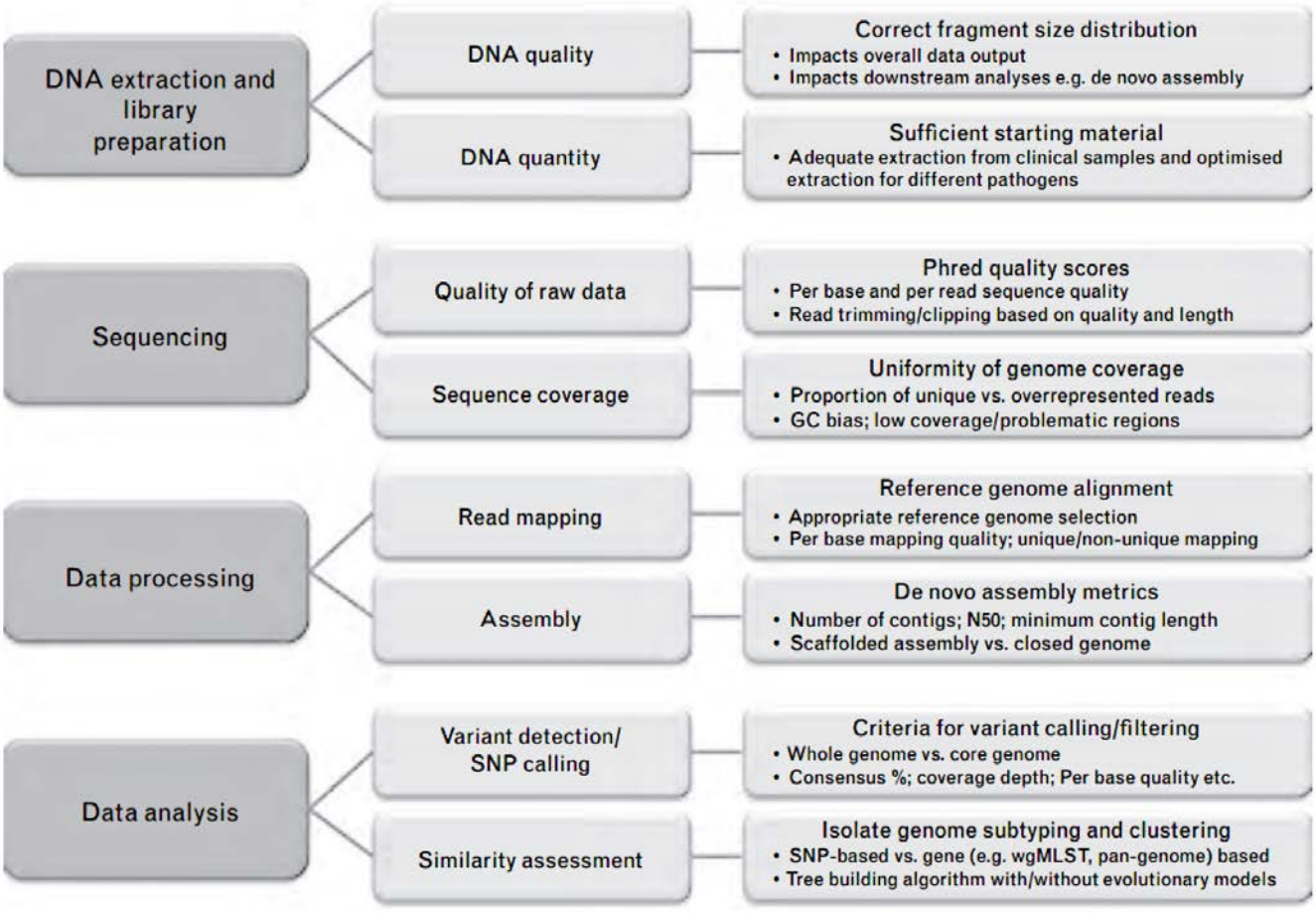


Figure 10. Quality assessment considerations for whole-genome NGS analyses.
Rigorous quality control processes and standardization of testing is required for NGS analyses. However, such processes and standards have yet to be determined. The above figure provides suggestions when assessing the quality of WGS analyses. Contigs, contiguous sequences; GC, genome coverage; SNP, single nucleotide polymorphism; wgMLST, whole-genome multi-locus sequence typing.

via proficiency testing or external quality assessment exercises. This requires the development of well-characterized proficiency testing materials that are suitable for the diverse range of NGS applications, as well as standard metrics to use as a comparator among laboratories (37). However, there is no current, formal proficiency testing program in place for NGS as applied in the clinical microbiology and public health setting. There is a need for the development of reference materials, such as well-characterized microbial genomic DNA, spiked matrices, and data sets, so that laboratories can run these reference materials for assay validation, performance assessment, and quality control. NIST (National Institute of Standards and Technology) has been granted FDA funding to develop standard microbial reference materials to be used when validating genomic sequencing diagnostics (111, 179). An international, inter-laboratory study involving five national metrology institutes assessed the variability in sequencing of the 16S rRNA gene from two certified genomic DNA references using different sequencing platforms. Participants were instructed to use the methods and protocols available in their laboratories. Although all data sets concurred for the biologically conserved regions, data sets varied in both read length and coverage, which impacted precision when examining the biologically variable positions of the 16S rRNA genes. Biases were also introduced by the individual laboratory's algorithm that aligned the reads to the reference genome (180). Therefore, addressing these inconsistency and reproducibility issues is critical before NGS can make its transition into the clinical laboratory. Allowances/provisions must be made so that SOPs for NGS technologies can evolve over time. Validation and proficiency testing is in its infancy for WGS, but some laboratories have shown reproducibility for resequencing pure culture isolates (158). As new devices and assays are marketed, there needs to be some malleability in proficiency testing procedures or a mechanism in place to change these operations in a relatively alacritous fashion. While it is anticipated that NGS technologies and applications will continue to advance and change over time, the quality management requirements for assay validation, quality control, and proficiency testing or external quality assessment will maintain a high standard (37).

Stakeholders from various sectors should collaborate to advance NGS guidance and standards. For example, the U.S. Centers for Disease Control and Prevention initiated the Next-Generation Sequencing: Standardization of Clinical Testing (Nex-StoCT I and Nex-StoCT II) working groups to develop guidelines for implementing NGS into clinical settings. The first of these guidelines concentrated on four main topics, including test validation, quality control, strategies for proficiency testing among laboratories, and the development and use of reference materials (37). The Nex-StoCT II informatics working group developed recommendations for the development and optimization of a clinical NGS bioinformatics analysis pipeline (181). Although the primary focus of the Nex-StoCT working groups was on human genetic NGS applications, many of the recommendations are appropriate for use in infectious disease testing and diagnosis. Furthermore, College of American Pathologists (CAP) formed an ad hoc committee referred to as the "NGS Work Group," which developed an 18-item accreditation checklist for NGS standardization in the clinical laboratory (39). A new International Organization for Standardization (ISO) genomics committee has been constituted recently, and global proficiency testing is being conducted by the Global Microbial Identifier (GMI) network (182). Having standards that are published and available for new pipelines and instrumentation is critical to assist laboratories with assay development and to maximize the positive impact NGS can have on clinical genomics and metagenomics.

4c. Formulation of a Team of Specialists for Clinical Translation of NGS Data

In order to shift NGS into the clinical arena, there needs to be open communication and partnerships among clinicians, clinical microbiologists, bioinformaticians, NGS platform technology experts, and software designers (10, 28, 173). Having different areas of expertise, these individuals would have distinct responsibilities, priorities, and ways by which they view and interpret NGS data. There is a pressing need to incorporate this technology into clinical fellowship training programs and pathology training programs. For example, with ASM's CPEP (Committee on Postgraduate Educational Programs) fellowship opportunities, students could be educated in both computational biology and clinical microbiology languages. Washington University School of Medicine has already implemented a Molecular Genetic Pathology fellowship program which functions as a collaborative effort among the Divisions of Anatomic and Molecular Pathology, Laboratory and Genomic Medicine, and Neuropathology. Many medical schools, such as the University of California—San Francisco, provide students with a medical and microbiological informatics lecture series to further prepare them for this era of genomic medicine. A common vernacular or lexicon encompassing both fields needs to be established for the next generation of clinical microbiologists, pathologists, and physicians (1, 31, 121). Since bench-level medical technologists perform the bulk of the work in the clinical laboratory, training opportunities for

these individuals must be made available as well. ASCP (American Society for Clinical Pathology) could offer bioinformatics certification exams like those presented for other subspecialties in order to spark interest. Master's level programs in NGS could be designed specifically for medical technologists in order to have adequately trained bench personnel for when NGS technology becomes a routine part of the clinical microbiology laboratory workflow.

Before NGS can reach its full potential in the clinical microbiology laboratory, validation and educational challenges need to be overcome. NGS technologies are not yet at the point where they simplify or accelerate the traditional, multi-step workflow that is currently carried out in the clinical microbiology laboratory. In the initial phases of clinical use, NGS data analysis and reporting will require a holistic combinatorial approach as depicted here.

SECTION 4 FINDINGS AND RECOMMENDATIONS

Key Finding 4.1 Wet and dry bench steps for NGS protocols as well as quality metrics need to be standardized if NGS is to be used routinely in the clinical microbiology laboratory. CLIA requires the development and validation of test performance for any technology in the laboratory.

Key Recommendation 4.1 A group of stakeholders (i.e., government agencies, clinical laboratory professionals, academia, industry) should be brought together to develop standardized reference materials and data sets that can be used for assay validation and quality control procedures. To make validation of bioinformatics pipelines easier, publically hosted "digital" validation test sets could be purchased and evaluated by clinical laboratories. A set of reads that have a known answer could be downloaded and subjected to the lab's bioinformatics components. Also, fully characterized biological reference organisms will be needed to evaluate both the wet and dry NGS processes.

Key Finding 4.2 Having consensus guidelines for all types of infections or clinical scenarios encountered in the clinical microbiology laboratory will help with incorporating NGS methods into the standard of care.

Key Recommendation 4.2 It is recommended that different wet and dry bench NGS protocols be created for the detection of etiological agents such as bacteria, viruses, fungi, yeasts, and parasites. Although NGS has the potential to detect all pathogens in a clinical sample, specific protocols would help to advance the transition of NGS into the clinical microbiology laboratory. There needs to be guidance for how to validate and perform QC procedures for these protocols as they pertain to the different pathogens (e.g., what are the unique/pathogen-specific QC metrics that need to be considered to ensure the quality of NGS results?).

Key Finding 4.3 Until NGS becomes widespread and routinely used in the clinical microbiology laboratory, data analysis will require a collaboration from various groups of people.

Key Recommendation 4.3 To ensure successful translation of an NGS result, a multidisciplinary team within the hospital or public health laboratory setting should be formed to include the expertise of clinical microbiologists, medical technologists, clinicians, infectious disease pathologists, basic research scientists, software developers, and bioinformaticians. This collaborative effort will maximize the strength and interpretation of NGS data.

Key Finding 4.4 Education of the clinical workforce is fundamental if NGS is to be used routinely in diagnostic microbiology practice.

Key Recommendation 4.4 Adoption of NGS into the clinical microbiology laboratory will require clinical microbiologists, medical technologists, and clinicians to receive training in molecular biology and bioinformatics. Competence in bioinformatics and programming software needs strengthening at many levels within the clinical workforce. Beyond general programming skills and bioinformatics knowledge, there needs to be training on understanding and interpreting NGS results. It is recommended that microbiological informatics be incorporated into the coursework of medical school students and clinical microbiology/pathology fellowship programs so these students gain familiarity with this diagnostic approach that is likely to be used during their clinical practice. Exposure to informatics might even begin at the high school and undergraduate levels, since the basic principles are applicable to many fields.

Section 5
Challenges with Analysis, Management, and Storage of NGS Data

5a. Software Choices for NGS Analysis

NGS platforms produce massive volumes of data at a high rate, a feature that is both a strength and a limitation of these technologies at the current time. Terabytes of data are typically generated from a single run. New analytic approaches that can mine NGS data sets cannot keep pace with the amount of data produced (5, 110, 183). As displayed in ***Figure 11***, NGS has shifted the sequencing workflow from less upstream sample preparation time to substantial downstream data analysis and management efforts (16). There is a need for advanced, efficient, and user-friendly computational architecture for NGS data analysis (6, 24, 110). Bioinformatics software is often freely available for download or can be purchased from commercial companies or designed by the user (9, 43). See ***Figure 12*** and supplemental tables in studies by Voelkerding *et al* (5) and Pabinger *et al* (184) for common examples of NGS software tools used for WGS and other applications. For a more detailed review of diagnostic software tools, see the article by Fricke and Rasko (185). This paper contains a list of seminal publications on bacterial diagnostic applications of WGS along with the analysis method used. The use of NGS analysis programs can be completed in-house within the microbiology laboratory, provided that the necessary hardware infrastructure and training are supplied. Analysis can also occur remotely with cloud-computing services where users rent computing power to evaluate NGS data (16). It is important to acknowledge that software for NGS interpretation is currently a moving target and will continue to change quickly, given that there are very few standard analysis practices at the moment.

Developed at the University of California—San Francisco (UCSF), sequence-based ultrarapid pathogen identification (SURPI) is a computational pipeline that can detect any pathogen in the GenBank reference database. This software program can be used on stand-alone and cloud-based servers. SURPI can provide a comprehen-

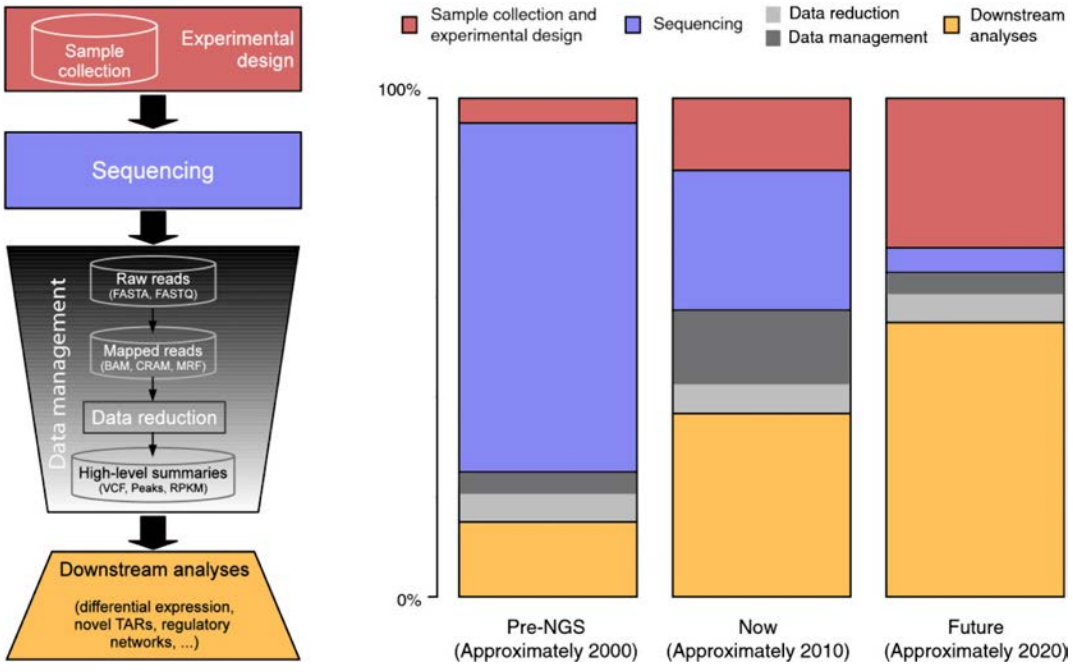


Figure 11. Contribution of different factors to the overall cost of a sequencing project across time.
Left: the four-step process: (i) experimental design and sample collection, (ii) sequencing, (iii) data reduction and management, and (iv) downstream analysis. **Right:** the changes over time of relative impact of these four components of a sequencing experiment. BAM, Binary Sequence Alignment/Map; BED, Browser Extensible Data; CRAM, compression algorithm; MRF, Mapped Read Format; NGS, next-generation sequencing; TAR, transcriptionally active region; VCF, Variant Call Format.

Republished from reference 16 with permission.

De novo assembly

- *De novo* assembly involves using computer algorithms to align overlapping WGS reads to form longer contiguous sequences known as contigs, and order the contigs into a framework of the sequenced genome (scaffolds). Velvet (https://www.Ebi.Ac.Uk/~zerbino/velvet/)[72] and SPAdes (http://bioinf.Spbau.Ru/spades)[73]are two of the more popular assemblers for Illumina short-reads, while Ion Torrent reads are better assembled using MIRA (http://www.Chevreux.Org/projects_mira.Html). Other commonly used assemblers include Newbler (http://swes.Cals.Arizona.Edu/maier_lab/kartchner/documentation/index.Php/home/docs/newbler) for 454 pyrosequencing reads, and the commercial CLC Genomics suite. Assemblers used for PacBio long reads include SPAdes, HGAP[74] and the Celera-MHAP assembler.[75]
- Contigs can be visualized in the Java-based program Mauve (http://gel.Ahabs.Wisc.Edu/mauve/), which can also order and orientate contigs to a reference genome. Alternatively, command-line tools such as MUMmer (http://mummer.Sourceforge.Net/) can be used to automate and batch this process as part of an assembly pipeline.

Annotation

- Genome annotation includes identification of DNA segments of known and probable open reading frames (ORF) that contain gene coding DNA, and matching the identified segments to a database of known gene sequences. Tools include the web-based RAST (http://www.nmpdr.org/FIG/wiki/view.cgi/FIG/RapidAnnotationServer), NCBI's Prokaryotic Genome Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/) or the command-line tool Prokka (http://www.vicbioinformatics.com/software.prokka.shtml) for automated genome annotation.

Genome visualization and comparison

- Once assembled and annotated, genomes can be viewed using a genome browser to display the structure and embedded genetic elements of a genome in a graphical format, and manipulate the genome sequence if required. The Wellcome Trust Sanger Institute's Artemis (http://www.sanger.ac.uk/resources/software/artemis/), and the commercially available Geneious Pro suite (http://www.geneious.com/) are examples of genome browsers.
- Visual comparisons of multiple genomes can also be made using the above utilities.

Alignment and read mapping

- Read mapping is the process of aligning reads to a reference, using a combination of local and global alignment. Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) and BWA (http://bio-bwa.sourceforge.net/) are two of the more popular short read alignment algorithms.[76]
- BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi), the most widely used utility for searching a sequence database, uses local alignment of sequence segments. BLAST can be run either as a web-based tool, or incorporated using a command line.
- Whole genome alignment is a computationally intensive process, but can be performed using Mauve or Mugsy/MUMmer (http://mugsy.sourceforge.net/).

SNP/variant calling

- Single nucleotide differences identified from aligning comparator sequences to a reference can be used to describe genetic relationships between isolates. Multiple tools are available,[77] and are frequently incorporated into more automated software packages.
- We use the Nesoni suite of tools (http://www.Vicbioinformatics.Com/software.Nesoni.Shtml) as well as SAMtools (http://samtools.Sourceforge.Net/), Freebayes (https://github.Com/ekg/freebayes) and Nucmer (part of MUMmer).

Phylogenetic analysis

- Phylogenetic trees can be used to analyze and visualize the SNP differences between isolates, although the true phylogeny of a group of isolates is never known. Popular methods include the simpler but rapid neighbor-joining method (most phylogenetic software), and the more complex maximum likelihood approach (RAxML https://github.com/stamatak/standard-RAxML, and https://PhyML http://atgc.lirmm.fr/phyml/). More recently, Bayesian approaches to estimating phylogenetic relationships have become popular as computation technology has improved. Examples include BEAST (http://beast2.org/), MrBayes (http://mrbayes.sourceforge.net/), and BAPS (http://www.helsinki.fi/bsg/software/).
- SplitsTree and FigTree are examples of phylogenetic software that can calculate neighbor-joining or display trees produced by other software.

Utilities for clinical microbiology

- Species identification can be performed on WGS data by either 16S characterization, or by identifying short strings of DNA used in genome assembly (k-mer identification). Both options can be performed on the Danish Center for Genomic Epidemiology Java-based website http://www.genomicepidemiology.org/
- A number of other clinically useful tools are available on this site, including ResFinder for the detection of antimicrobial resistance, and Multi-Locus Sequence Typing. Command-line based tools such as BLAST using de novo assemblies, or SRST2 (https://github.com/katholt/srst2)[78] which uses read-mapping on sequencing reads, are better suited to automation, batching of multiple sequence analyses, and incorporation into analysis pipelines.

Databases

- NCBI GenBank (http://www.ncbi.nlm.nih.gov/genbank) and Genome Bank (http://www.ncbi.nlm.nih.gov/genome)
- European Nucleotide Archive (http://www.ebi.ac.uk/ena)
- DNA Databank of Japan (http://www.ddbj.nig.ac.jp/)

Typing databases

- MLST database (http://www.mlst.net/databases/)

Antibiotic resistance gene databases

- ARG-ANNOT (http://en.mediterranee-infection.com/article.php?laref=283&titre=arg-annot-)
- ResFinder (https://cge.cbs.dtu.dk/services/data.php)

Multifunction bioinformatic suites

- Geneious Pro (http://www.geneious.com/)
- CLC Genomics (http://www.clcbio.com/products/clc-genomics-workbench/)
- Bionumerics (http://www.applied-maths.com/bionumerics)
- Nesoni (http://www.vicbioinformatics.com/software.nesoni.shtml)
- Harvest (https://github.com/marbl/harvest)
- Galaxy (http://galaxyproject.org/)

Figure 12. Examples of tools used for NGS bioinformatic analysis.
Some tools are not necessarily used in the clinical microbiology laboratory in order to deliver an actionable result. These tools include annotation, genome visualization and comparison, SNP/variant calling, and phylogenetic analysis.

sive identification of bacteria, viruses, fungi, yeasts, and parasites in a clinical sample within a clinically relevant time frame and is capable of diagnosing critically ill patients (21, 96). Illumina has also designed a suite of bioinformatics software that is integrated with each Illumina sequencer. BaseSpace is Illumina's genomic computing environment that carries out many NGS analysis methods, including WGS, *de novo* assembly, 16S metagenomics, RNA-seq, whole exome/enrichment, and targeted amplicon sequencing. NGS data storage, management, and sharing either through a cloud or on-site are other key features offered by BaseSpace (186). Furthermore, Torrent Suite™ is the data analysis software package that can be integrated with the Ion Torrent platforms. The Ion Reporter software is a component of the Torrent Suite™ that helps to streamline the informatics process by performing primary and secondary analysis and data transfer to a secure Amazon cloud (187).

Ultimately, the choice or design of software and computational analysis pipeline should align with the needs of the individual laboratory and the specific NGS application. Software programs can yield the same quality metrics but will likely differ in performance and potentially results. It is important to maintain records of software versions, dates of use, and changes to the software along with the parameter files used in the analysis phase to generate a complete audit trail (173, 188).

5b. Pipeline for NGS Interpretation

NGS data analysis is not a trivial task and is critical to any sequencing experiment (175). In the clinical laboratory, analysis will equate to a clinically actionable result that has the ability to impact patient care (9, 27, 110). There are three broad categories of NGS analysis tools, including base calling, alignment of the sequence to a reference sequence/database or *de novo* assembly, and annotation and variant detection (**Figure 13**). However, the data analysis process can be directed onto many different paths depending on the needs of the end user. In the microbiological laboratory, steps like full-genome assembly and annotation are not necessarily needed for clinically actionable data (111). Each step of NGS analysis employs a different tool or algorithm, which makes for a lengthy process. There are even differences in algorithms among platforms designed by the same company, such as the Illumina NextSeq versus MiSeq versus HiSeq.

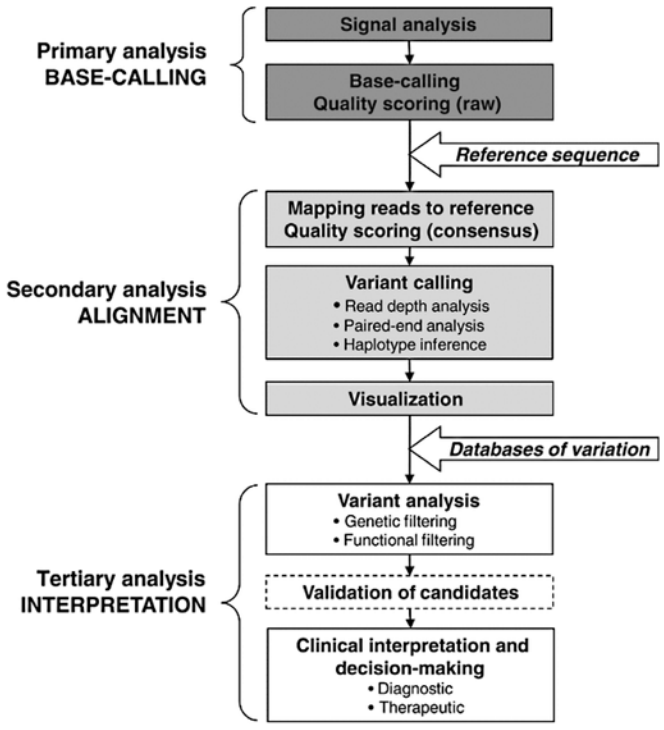


Figure 13. Outline of informatics pipeline for processing and analyzing data from massively parallel sequencing platforms.

The data analysis process can be broadly divided into three stages, including base calling, alignment, and interpretation. The variant calling and analysis steps of the procedure are not necessary for the clinical microbiology laboratory to perform. Data collection and analysis involve extensive use of various computational methods for converting raw data into sequence information and the application of bioinformatics techniques for the interpretation of that sequence.

Republished from reference 183 with permission from the publisher.

For nontechnical laboratory staff and health care professionals with no bioinformatics training, these tools are complicated, and thus results can remain cryptic without proper education. This issue has to be addressed before NGS technologies can be used routinely in the clinical laboratory.

Once the sequencing phase of the process is finished, raw sequencing reads are subjected to multiple steps beginning with preprocessing to remove adapter and low-quality sequences (38). Other analysis steps include base calling, assembly of the genome, alignment to known and curated genomic sequences, and annotation/comparative phylogenetics. Base calling is initially performed by proprietary software on the platform to account for technology-associated bias (1, 40). Each base call is assigned a level of certainty, similar to a Phred-like score that correlates to its predicted accuracy. Indications of mixed signals or other read errors are filtered or removed from analysis. The quality scores generated from each platform cannot be directly compared, because each platform uses a different algorithm (112). Following base calling, contigs are created by overlapping the numerous short reads (30 to 500 bp depending on the platform) or long reads (>5,000 bp) generated from the millions of sequencing read fragments of the target genome (9). Longer contigs provide a higher depth of sequence coverage that improves the accuracy and sensitivity of detecting pathogens (33). Bioinformatics software will align and compare the contigs against an available database of high-quality sequenced organisms (e.g., Kraken, SURPI, GOTTCHA, OneCodex, CLARK, LMAT, MG-RAST, MetaPhlan, MEGAN). The preferred program for assembling a genome depends on how the data are produced, either long or short read sequencing. For a more exhaustive list of genome assembly programs, see references 15, 189, and 190.

Genomes without a close relative in the reference database may not be detected using reference-based methods, necessitating *de novo* assembly for newly discovered genomes (15, 25, 38). *De novo* assembly is also useful for the discovery of mobile accessory elements, such as plasmids and phages, which may not be present in the reference genomes. Genomes that are assembled *de novo* do not use an existing reference scaffold and can be assembled by popular programs such as SPAdes, IDBA-UD, SOAPdenovo, MIRA (which can also be used for reference guided assembly), and Velvet (7, 15, 38, 190, 191, 192). Producing long overlapping reads is essential for *de novo* assembly since data are not aligned to a reference genome, and choice of assembly software is technology dependent. It is important to understand that a reference database is still necessary to accurately identify the organism even if *de novo* assembly is used. This may seem counterintuitive at first, but comparing long contigs against a reference database assists in the identification process and discovery of novel variants and genes.

After a set of ordered contigs, if not a whole genome, is obtained, annotation of the draft genome occurs. Most commonly used for bacterial genome annotation are Web-based tools such as the annotation tools RAST and Glimmer (38, 193). Ordered contigs are uploaded to RAST, and open reading frames that most likely represent genes are produced to create a highly annotated genome. The quality of the annotation process is largely dependent on the gene database that is applied. Besides RAST and Glimmer, there are other annotation tools available for users, including DIYA, Prokka, RATT, and BG-7. Additional information on annotation tools is presented in references 194 and 195. With annotated genomes, users are often interested in finding genes of biological importance,

such as genes conferring antibiotic resistance and virulence. Furthermore, comparative genome analysis is frequently exploited to align multiple genomes to determine regions of sequence homology or to identify unique regions of the genomes. Examples of software tools that enable annotated genomes to be compared and visualized include BRIG, Mauve, and ACT. Some of these alignment programs such as Mauve can identify SNPs that are useful in phylogenetic analyses and for establishing evolutionary relationships, tasks that are more suitable for public health surveillance and epidemiological studies than for the clinical laboratory (61, 193).

Because unbiased NGS generates sequence data for all types of DNA found in a clinical sample, the data can include organisms that do not prompt immediate action, such as normal microbiota components. With current microbiology practices, and more specifically, culture-based identification, the laboratory will report not only the etiological agent but also such commensal organisms. For example, the laboratory will report "normal respiratory microbiota" for sputum samples and "mixed microbiota" for urine specimens. In parallel to this approach, a result from an NGS-based test that is sent in a final report to the clinician should be similar and include normal microbiota if detected. Nonetheless, there is some skepticism associated with this tactic. Reporting all organisms could inadvertently introduce new organisms that are not necessarily actionable but could be acted upon by certain groups of physicians. This type of reporting could have a detrimental effect on antimicrobial stewardship moving forward, and therefore irrelevant information should be filtered before the report reaches the hands of the physician. For the immediate future, any prediction coming out of NGS and bioinformatics will likely be followed up by real phenotypic tests or some other confirmatory evidence to verify the prediction.

5c. Determining Clinical Relevancy of NGS by Distillation of Data: Integration into Care

To extrapolate clinical relevancy in a time-efficient manner, an ongoing dialogue between clinicians, clinical microbiologists, and the developers of NGS software is necessary. There should be a gradual and logical transition in the type of data transmitted to the clinician, from language with which they are familiar based on traditional tests and assays toward more complex analyses as training improves and as this dialogue continues. The tremendous amount of NGS output must be condensed into a reportable result, with unnecessary information removed or strongly qualified (6). There will still be uncertainty in the results generated from NGS, even with advancements in the technology and bioinformatics pipelines. The complexity of the data is substantial, and thus it is easy to cherry-pick an answer and make a false conclusion. Finding the most optimal way to communicate an uncertain result is important to relay to the clinician.

NGS methods can both detect and identify infectious agents in a single run. Ideally, information regarding antibiotic resistance mechanisms and virulence determinants should also be derived from the same NGS assay without having to perform additional testing (30). Predicting antibiotic resistance from genetic data is still in the early stages of development, with limited cases showing the direct genotypic-phenotypic link (10, 15, 27, 28, 30, 33, 143, 144, 196, 197). One study performed by Gordon *et al* (198) highlighted a genomic prediction tool for determining antimicrobial susceptibilities for *Staphylococcus aureus*. This genotypic tool was comparable to antimicrobial susceptibility testing methods with regards to specificity and sensitivity (198). Antibiotic resistance genes may hold predictive value; however, the genetic mechanisms responsible for a resistant phenotype are not always clear-cut. NGS can determine antibiotic resistance when known resistance genes or mutations are identified but cannot reliably demonstrate resistance when unknown or novel mechanisms of resistance are present in the organism (188). Antimicrobial resistance remains one of the three greatest threats to human health according to the World Health Organization (WHO) (199, 200, 201, 202, 203). Therefore, it is essential to detect resistance quickly and reliably, and in this aspect, NGS shows promise. Assessing virulence gene content from NGS data is also in the preliminary stages of development (30). In the foreseeable future, NGS will be a preferred method for studying virulence and antibiotic resistance mechanisms with the hope of generating knowledge to further diagnostics, therapeutics, and vaccine design (***Figure 14***) (1, 9, 28, 31, 56, 156, 164, 204).

5d. Storage of NGS Data and Security Concerns

Another principle challenge of NGS is downstream data management, storage, and sharing. Long-term storage of raw NGS data reads is not reasonable because of the large size of the data files (43). For example, storage of the reads for a single *Staphylococcus aureus* genome requires approximately 1 GB of space, a size that is comparable to the storage of 10 music albums in MP3 format (33). The 1 GB of storage space for the *S. aureus* genome is uncompressed, and therefore if the data were compressed to just the variants with respect to the reference (e.g., CRAM), the storage size would be much less. Following base calling, sequence data should be stored in platform-specific output files

such as .FASTQ files. Because each NGS platform delivers data in a slightly different format, this can cause difficulties when trying to compare data from various platforms and when uploading data into databases with multiple formats (38). A possible solution is to leverage cloud-based or public repositories, such as the NCBI Sequence Read Archive (SRA), which use standardized formats and offload storage costs. When regulations prohibit off-site storage, raw data and sample metadata (information on the clinical isolate) can be stored locally in a compressed format or replaced with a genome assembly or report of detected genes and variants. In fact, the NCBI, along with the FDA, CDC, and USDA, has already established a template (known as BioSample) for minimal information for pathogen metadata which could at least serve as a starting point for discussions of other organisms or for clinically relevant publicly available metadata (see https://submit.ncbi.nlm.nih.gov/biosample/template/?package=Pathogen.combined.1.0&action=definition). It should be emphasized that when the raw sequencing reads are submitted to the public archive, previous downstream analysis can be independently tested with new scientific algorithms, software, and protocols. Therefore, public repositories could be tools for the evaluation of new scientific understandings. Another information technology infrastructure demand includes long-term data archives, backup storage solutions, and retrieval facilities that would permit traceability to a patient's report (10). An important archival note involves recording the software version, database version, and analysis parameters so that if the data require reanalysis, users will know what was applied previously.

Because data generated from an NGS test is associated with a specific patient, laboratories must ensure that proper measures are in place to uphold patient confidentiality and privacy when these data are internally and externally stored

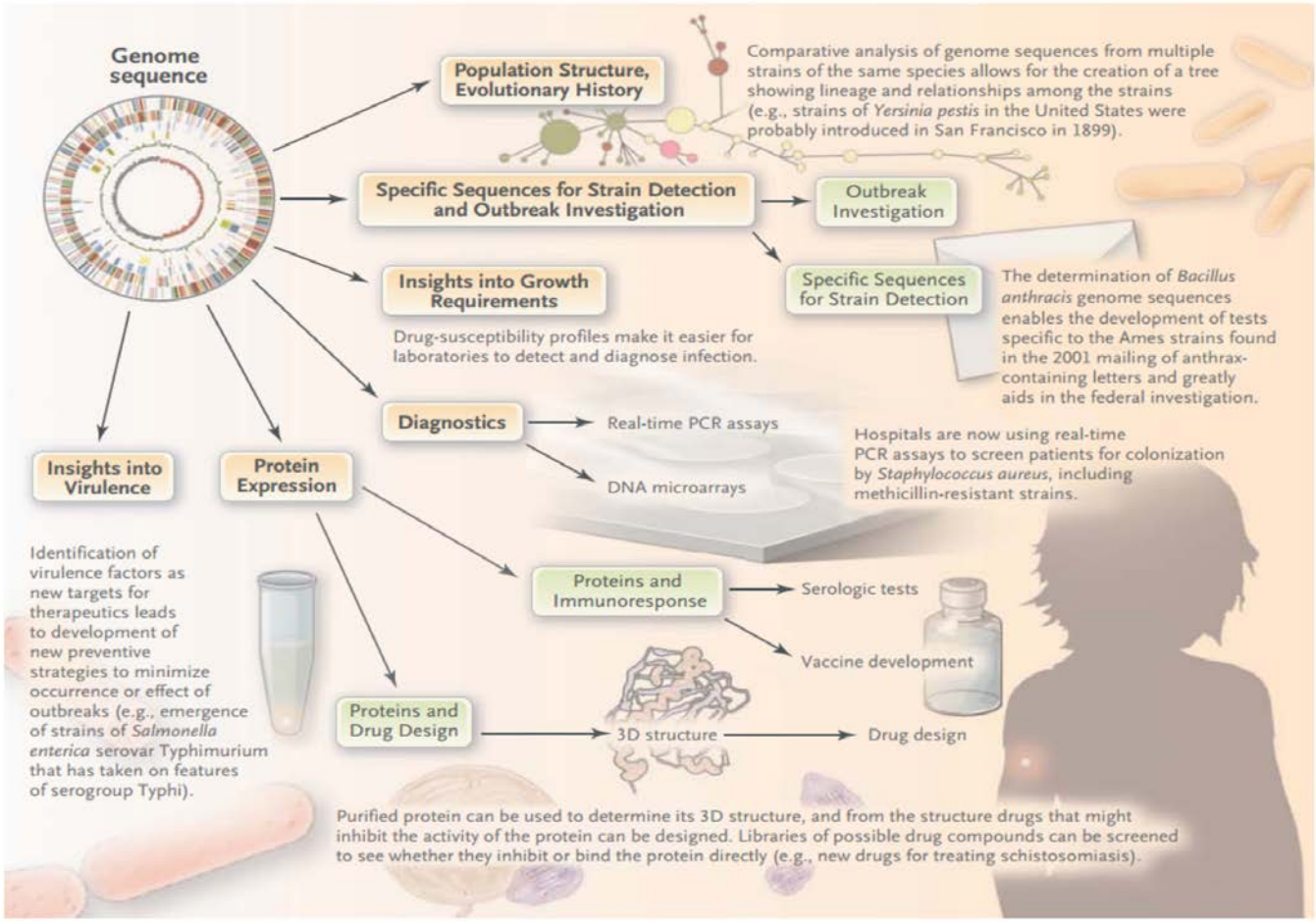


Figure 14. Use of microbial genomics for tool development.
A genome sequence can facilitate the development of a variety of tools and approaches for understanding, manipulating, and mitigating the overall effect of a microbe. The sequence provides insight into the population structure and evolutionary history of a microbe for epidemiologic investigation, information with which to develop new diagnostic tests and cultivation methods, new targets of drug development, and antigens for vaccine development.

Republished from reference 56 with permission from the publisher.

or transferred (129). Consequently, laboratories must maintain the standards required by the Health Insurance Portability and Accountability Act (HIPAA). Because in-house NGS computational infrastructure requires significant financial resources, a popular alternative is cloud computing (7, 38). These services are offered at a reasonable price by technology companies such as Google, Amazon, and Microsoft and are easily scalable depending on the sequencing needs (10). Analysis of data can occur in the cloud, and data can be shared through this platform, although this may be impaired by bandwidth and data transfer capabilities. However, storing patient data in the cloud is not permitted by some institutions, and compliance with HIPAA regulations varies with provider. Rigorous policies regarding the sending and receipt of a patient's clinical genomic data to cloud-based vendors or other health care institutions need to be enforced. Examples of these safety procedures include data encryption, secure data transfer, user authentication, and transmission tracking pathways (39).

SECTION 5 FINDINGS AND RECOMMENDATIONS

Key Finding 5.1 Establishing NGS capabilities in a clinical microbiology laboratory requires a significant financial outlay. Purchase of an NGS platform is only one component of the cost, and often underemphasized is the requirement for essential NGS hardware and accessory equipment. Bioinformatics expertise will be essential until the "easy button" is implemented.

Key Recommendation 5.1 To help minimize the cost and bulkiness of NGS hardware implementation, the utility of benchtop and point-of-care (field-able) sequencing platforms should be emphasized for clinical laboratories seeking to engage in this space. These sequencing systems consume less space and are generally less expensive than larger NGS platforms, and data analysis can be completed on a high-end desktop server or even a laptop.

Key Finding 5.2 NGS data need to be distilled into a clinically actionable result. This wealth of data can be made clinically relevant by identifying genotypic resistance mechanisms and genes encoding virulence factors. Deployment of NGS will require substantial validation of genotypic prediction of the AMR phenotype; however, there is limited precedent for this work.

Key Recommendation 5.2 Some NGS-based assays, e.g., metagenomic assays, are capable of returning a complex set of results that require careful interpretation by the clinical microbiologist to determine what is clinically actionable and what should be included in the results report to ensure optimal patient care. For example, NGS results should assist the physician in determining what antimicrobial can be used rather than what drugs the organism is resistant to or what drug is not suitable for treatment. The presence of the gene conferring AMR is not evidence of its expression and hence AMR phenotype. There needs to be a way to phenotypically verify the genotypic result generated by NGS. Rapid phenotypic testing methods are currently under development.

Key Finding 5.3 Concerns over patient privacy and confidentiality of NGS genomic data remain an issue.

Key Recommendation 5.3 When NGS data are to be stored remotely, such as on a server, in the cloud, or within a database, the information must protect patient privacy and be HIPAA compliant.

Key Finding 5.4 All of the files associated with a whole-genome NGS run represent an enormous amount of data requiring significant storage space.

Key Recommendation 5.4 The raw sequence reads, .FASTQ files, and the complete genomic sequence of the identified pathogen would consume considerable storage space for the hospital. The assembled sequence should be uploaded to an appropriate database, and only the clinically relevant result should be maintained in the patient's electronic medical file.

Section 6

Deterrents and Shortcomings of NGS Technology and Challenges for Implementation into the Clinical Microbiology Laboratory

The high cost of NGS instrumentation is a tremendous obstacle. Although NGS platforms continue to change and become more efficient, to a certain extent, this rate of change is hindering the adaptation of NGS technology in the clinical microbiology laboratory (121). It is necessary to reflect on the capabilities of the various platforms and adopt the equipment that best produces a clinically actionable result. All in all, this field is advancing at a remarkable rate which can, paradoxically, hinder acceptance of the technology. Continual NGS improvements make it arduous for clinical laboratories to lock down a platform and invest in the instrumentation. The constant altering and tweaking of NGS platforms and bioinformatics make it challenging to complete validation studies. Also, the clinical strains that will get validated will continue to evolve by acquiring or losing DNA. This necessitates continuous validation studies, which are not realistic (158).

Physically getting the computational infrastructure into the laboratory can be a costly obstacle as well. Substantial computing resources are required for successful operation of an NGS platform, including analysis software, short- and long-term data storage, backup solutions, and data retrieval capabilities. The entire laboratory may require updating with electrical connections, fiber for increased bandwidth, switching hubs, power supplies, generators, and air conditioning in order to support sequencing efforts and on-site computing and data centers (10).

The sheer amount of data generated from a single NGS experiment can be overwhelming and intimidating. The copious volume of genomic sequence that must be condensed into a meaningful result represents a formidable challenge. The primary bottleneck of NGS technologies that is preventing them from reaching their full clinical potential is complex data analysis (6, 7, 9, 10, 24, 25). Dehosting or filtering out human host reads from NGS data remains a challenge because of a lack of software tools that can perform this task. This "data jam" is also due in part to the insufficient training among the clinical laboratory workforce with respect to computational analysis tools and skills in interpreting large data sets (6, 24, 112). There is a lack of development of vital bioinformatics resources and pipelines for NGS interpretation. An example of a major resource that needs to be developed is accessible reference databases of curated microbial genomes (27). These databases should not be stagnant but constantly evolving as new, relevant organisms are being sequenced. True representation of diversity, including current, circulating organisms and older strains, needs to be considered when assembling these databases. It is important to understand the existing reposi-

For your information…

FDA-Approved Medical Devices
Those devices for which FDA has approved a pre-market approval (PMA) application prior to marketing. This approval process is generally reserved for high-risk medical devices and involves a more rigorous premarket review than the 510(k) pathway.

FDA-Cleared Medical Devices
Those devices that FDA has determined to be substantially equivalent to another legally marketed device. A premarket notification, referred to as a 510(k), must be submitted to FDA for clearance prior to marketing a device. A 510(k) is most often submitted by the medical device manufacturer.

FDA-Listed Medical Devices
A medical device is FDA-listed if the firm that manufactures or distributes the medical device has successfully completed an online listing for the device through the FDA Unified Registration and Listing System (FURLS).

Retrieved from http://www.fda.gov/AboutFDA/Transparency/Basics/ucm194468.htm

tories, both public and private, in order to prevent the creation of silos, or databases that contain the same organisms. The NCBI's GenBank database is the public repository for all types of genomic sequences and contains over 3,000 complete, annotated genomes and nearly 25,000 draft genomes that need annotation (9). The NCBI, the European Nucleotide Archive (ENA), and the DNA Data Bank of Japan (DDBJ) are collectively known as the International Nucleotide Sequence Database Collaboration (INSDC), an initiative that supports and endorses data sharing among the three databases (http://www.insdc.org/). In addition to the INSDC public repository, the GMI is yet another platform designed for the storage of WGS data. GMI is a global genomic epidemiological database that holds sequence information from all genera of microorganisms linked with important metadata (http://www.globalmicrobialidentifier.org/About-GMI). The goal of GMI is to aid in global surveillance of infectious diseases by serving as a resource in detecting outbreak isolates and emerging pathogens. Furthermore, the FDA developed a microbial genomic sequence reference database known as FDA-ARGOS to cover the diversity of circulating strains, including clinically and environmentally important microbes, for diagnostic and regulatory purposes (see Section 7A).

Another barrier to NGS usage in the clinical laboratory is the lack of a standardized report format because of issues with interfaces on the laboratory information management systems (LIMS). The unprecedented throughput and constant state of change associated with NGS create challenges for traditional LIMS. To support NGS genomics, LIMS must be scalable, adaptable to routine change and updates, and easily operated by nontechnical laboratory staff. Genomics-specific LIMS will capture data quality, streamline data tasks, eliminate bottlenecks, track key associations such as consumable lot numbers and SOPs, and perform other support activities as discussed elsewhere (205). These customized functions take significant time and money to develop. There are commercial and noncommercial, open-source LIMS available to track NGS samples from request submission to test result. Illumina offers three versions of GenoLogics NGS LIMS that enable users to preconfigure and track workflows, generate and manage reports, and receive instrument and third party software updates. It is a difficult task to find a framework that meets the demands of all laboratories, and hence, one user interface does not fit all. There needs to be a channel that allows for communication between the hospital databases and the sequencing databases to minimize errors in the process introduced by users. LIMS modules from different hospitals do not interact well with each other, making institutions that want to share data between systems concoct unique mechanisms for retrieval (206). Interconnectedness among LIMS is a hurdle that needs to be cleared and is vital for accessing, sharing, and leveraging the power of NGS data. Data sharing can aid in the creation of better software tools that would then allow the development of specific NGS applications.

6a. Regulatory Issues of an NGS Diagnostic Test

When any new technology, device, or method is implemented into a specific environment, there are challenges that have to be overcome. Incorporation of NGS technology in the clinical laboratory is no exception and has its own set of obstacles (37). In 1976, Congress enacted the Medical Device Amendments, which amended the Federal Food, Drug, and Cosmetic Act (the FD&C Act) to create a comprehensive system for the regulation of devices intended for use in humans. Under section 201(h) of the FD&C Act, "an instrument, apparatus, implement, machine, contrivance, implant, *in vitro* reagent, or other similar or related article, including any component, part, or accessory…intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals" is defined as a device. Under the FD&C Act, the U.S. Food and Drug Administration (FDA) ensures that medical devices, including *in vitro*[1] diagnostic (IVD) devices, provide a reasonable assurance of safety and effectiveness (176). The Center for Medicare and Medicaid Services (CMS) is responsible for regulating laboratories under the 1998 Clinical Laboratory Improvement Amendments (CLIA). CLIA governs the accreditation and certification process for laboratories (181, 207). In the United States, diagnostic tests, similar to other medical devices, undergo evaluation by the FDA to obtain marketing authorization. Determining the appropriate amount of regulatory oversight for NGS technology presents a new challenge. The FDA premarket review process has generally followed a "one test, one disease" paradigm that assesses a test's analytical and clinical performance. NGS technology is essentially an ultimate multiplex test that can detect many infectious agents.

[1] FDA regulations defines *in vitro* diagnostic products as "those reagents, instruments, and systems intended for use in the diagnosis of disease or other conditions, including a determination of the state of health, in order to cure, mitigate, treat, or prevent disease or its sequelae. Such products are intended for use in the collection, preparation, and examination of specimens taken from the human body. These products are devices as defined in section 201(h) of the Federal Food, Drug, and Cosmetic Act (the act), and may also be biological products subject to section 351 of the Public Health Service Act." 21 CFR 809.3(a).

In 2013, the FDA allowed marketing of the Illumina MiSeqDx instrument platform and the Illumina Universal kit, two devices that make up the first FDA-regulated test system that allows laboratories to develop and validate sequencing of any part of a patient's genome. This marketing authorization marked a significant milestone for this technology (96, 208). The FDA also cleared two devices that are used to detect DNA changes in the cystic fibrosis transmembrane conductance regulator (CFTR) gene, which can result in cystic fibrosis (CF), an inherited chronic disease that affects the lungs, pancreas, liver, intestines, and other organs of those who inherit a faulty CFTR gene from both parents (209). These two cleared devices are: Illumina MiSeqDx Cystic Fibrosis 139-Variant Assay and Illumina MiSeqDx Cystic Fibrosis Clinical Sequencing Assay. The Illumina MiSeqDx has not been granted marketing authorization by the FDA for microbial diagnostic use (http://www.accessdata.fda.gov/cdrh_docs/reviews/K123989.pdf). In addition, two other sequencing platforms were registered and listed with the FDA as *in vitro* diagnostic devices in 2014: Life Technologies Ion PGM and Vela Sentosa SQ301 (209, 210, 211). Taken together, establishing an FDA review process for an NGS platform or assay seeking marketing authorization that appropriately balances the interest in doing an adequate review against the interest in minimizing the information and effort involved in preparing a submission able to support a favorable decision has presented challenges, especially because the technology is changing at a rapid pace. By the time one test is cleared or approved, a new test will likely appear on the market that can perform that same test with a quicker turnaround time and at a lower cost. However, this situation is not unique to clinical microbial NGS technology.

The process of validating microbial NGS-based tests faces many obstacles. Establishment of clear validation standards by either national or international regulatory agencies could potentially help with the use of NGS tests in clinical laboratories. The standard reference materials being developed by NIST have the potential to address some of the challenges encountered during the validation process for microbial NGS-based tests (111). Similar to other medical devices, NGS testing is an inherently multi-analytical procedure for which the FDA, as part of its review, requires validation of all components of the system, including nucleic acid extraction, library preparation, sequence generation, sequence analysis, sequence interpretation, and formulation of a reportable result (7). This systems approach would benefit from continued collaboration from regulatory agencies such as the FDA and CMS as well as medical societies such as College of American Pathologists CAP. Another challenge to address is the regulatory and ethical issues that surface with the storage of patients' NGS test results. Such storage must adhere to privacy laws such as compliance with HIPAA requirements. With the capabilities of NGS methods when sequencing a human clinical specimen, it is also possible to uncover an unexpected finding such as a marker for an inheritable disease of which there are no guidelines for appropriate disclosure of such incidental data (212).

6b. Recommendations for Overcoming Barriers for NGS Implementation

Reimbursement codes need to be standardized or new codes that cover genomics-driven interventions or NGS assays should be created (31). Some suggestions include "sequencing," "microbiology genome diagnostics," and "metagenomic sequencing." It is important to engage insurance companies and inform them about the downstream cost savings of NGS analysis. Case studies and, ideally, clinical trials would be useful and encouraging tools that could be directed towards insurance companies. Success stories showing how an NGS test result decreased a patient's hospital stay or

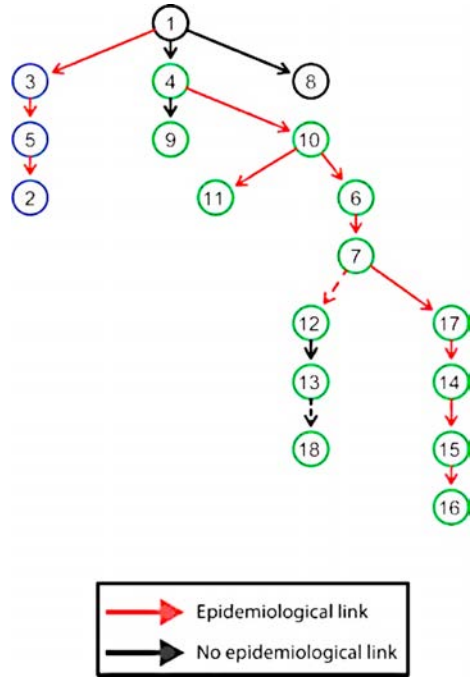
Case Study 10

WGS tracks a hospital outbreak of carbapenem-resistant *K. pneumoniae* (97)

Carbapenems are the most potent class of β-lactams, and with the emergence of carbapenem-resistant strains, treatment options have become extremely limited. The U.S. NIH Clinical Center had a carbapenem-resistant *K. pneumoniae* outbreak that afflicted 18 people and resulted in 11 deaths (***Figure***). WGS and epidemiological analysis demonstrated that the outbreak source stemmed from a single patient who had been discharged 3 weeks prior to the appearance of clinically relevant cases. The index patient independently transmitted the outbreak strain at least three times to create two main clusters of patients. *K. pneumoniae* is also an organism that can secretly colonize the gastrointestinal tract of patients without causing overt signs of infection. These patients act as silent transmission vectors, allowing an outbreak to develop covertly over a time period, as evidenced by this study. This case highlights how NGS can help to limit the nosocomial transmission of pathogens and thereby guide infection control efforts.

saved a patient's life because the pathogen was identified more quickly and therapy was adjusted appropriately could be highlighted (36). [**See Case Study 10, which demonstrates how WGS traced a hospital outbreak of a *Klebsiella pneumoniae* (*K. pneumoniae*) carbapenemase(KPC)-producing organism and hence provided actionable insights to decrease further intrahospital transmission (97).**] There needs to be an open dialogue between the scientific community and insurance companies as NGS progresses towards the infectious disease diagnostic space. Tactics to convey the value added of NGS to agencies or groups that deal with reimbursement need strengthening (10).

There are simple, cheap, and routine methodologies that work efficiently in the clinical laboratory at the current time. It is questioned that if these established methods are efficient, how is NGS better than what we have now? NGS signifies a value-added method that has the capabilities of detecting all pathogens from one test, exemplifying its power to transform the clinical microbiology field and potentially eliminate the need for a suite of conventional tests. A major hurdle is getting the medical community and hospital administrators to embrace NGS technology and to realize the clinical value it possesses over current tests performed in the microbiology laboratory.



Case Study 10 (Figure) Putative map of *K. pneumoniae* transmission during a hospital outbreak.

The transmission map was constructed with genetic and patient trace data. Nodes represent patients, and arrows indicate a transmission event directly or indirectly from one patient to another. Blue, cluster I; green, cluster II. Red arrows, opportunity for a direct transmission event from patients overlapping in the same ward before the recipient cultured positive; black arrows, transmission events that cannot be explained by patient overlap (may result from a more complicated transmission route or an intermediate patient or environmental source); dashed lines, at least one other equally parsimonious transmission link leading to the given patient exists.

Republished from reference 97.

Because traditional microbiological methods remain ingrained in diagnostic laboratories, hesitation and fear exist on the part of medical technologists, clinicians, and hospital administrators to learn the technology and terminology, which represents another barrier that is preventing acceptance of NGS (28, 33). This is not just a new technology but a new paradigm for how we conduct clinical microbiology, which can make these groups reluctant to engage NGS. Even though NGS is becoming widely available, there are a lack of education and sparse training opportunities, which inhibit the use of this advanced sequencing method (6, 24, 121). Courses in bioinformatics and data analytics could begin at the undergraduate level, with continued training in graduate and professional schools to help mold a new generation of scientists and physicians. Short courses, workshops, or CME (continuing medical education) credit could stimulate interest in NGS and help lessen anxiety for implementation while strengthening users' competence. It would also be beneficial for professional practice organizations to articulate and implement NGS language and evidence-based guidelines to bolster familiarity with this powerful sequencing method.

Numerous benchmarks must be met in order for NGS technologies to be incorporated into the workflow of the clinical microbiology laboratory. Many hurdles still need to be overcome before the medical community will embrace the capabilities of NGS technology and its deployment as an infectious disease diagnostic system (5, 22, 112).

6c. "Diagnostic" Clinical Trials Evaluating Microbial NGS

Several clinical trials have been launched to evaluate the ability of NGS-based tests to detect inherited genetic diseases or cancer alleles. Little progress has been made surrounding clinical trials for the diagnosis of infectious diseases. As shown on the ClinicalTrials.gov website (https://www.clinicaltrials.gov), there are clinical trials currently under way whose purpose is to evaluate the utility of NGS for the detection or characterization of microbes. However, these clinical trials are parallel to research studies in which patients are enrolled and provide consent for NGS analysis, including microbial NGS. Specifically, diagnostic infectious disease NGS clinical trials are minimal to nonexistent. The ClinicalTrials.gov website shows that an overseas

company, PathoQuest, is conducting a clinical trial for the evaluation of high-throughput sequencing for screening pathogens in immunocompromised patients with suspected infectious disease (Clinical Trials registration no. NCT02007642). The purpose of the study is to compare the high-throughput sequencing approach of PathoQuest to detect pathogens causing infectious disease to conventional diagnostic methods. Preliminary findings have not been declared on the website or published as of June 2015 (213).

For the appropriate implementation of NGS as a diagnostic tool, clinical trials are necessary to expose strengths and weaknesses of both current NGS applications and traditional clinical laboratory approaches. Although there were early successes of diagnostic NGS trials, financial resources were lacking to continue to perform the clinical trials necessary to generate sustained interest. Such interest relies in part on adoption of the technology as well as a true detailed understanding of the benefits and cost of use (7, 160). These elements are lacking right now for infectious disease NGS clinical trials. Because NGS is a multistep process, it will be necessary to perform clinical trials at all levels of validation (14). Moving forward, outcome studies will ultimately be needed to demonstrate how NGS information genuinely affects and impacts patient care, followed by a clinical trial evaluation (160). Currently, viable funding mechanisms to implement these types of clinical trials are not available.

SECTION 6 FINDINGS AND RECOMMENDATIONS

Key Finding 6.1 Data analysis remains a major bottleneck in the adoption of NGS. Genomic sequence data are able to be generated at orders of magnitude higher and at a lower cost per base than traditional Sanger sequencing, but interpretation of these data is not nearly as rapid. Analysis of NGS data is a time-consuming and cumbersome process that requires the data to be channeled through multiple software programs and subjected to many different algorithms of which the interfaces are not user-friendly.

Key Recommendation 6.1 Professionals who will use NGS technology should work closely with software developers to create a proficient, streamlined, and more manageable analysis pipeline to provide a quicker return of complete diagnostic information. This collaboration will help in the development of more efficient and user-friendly software programs and algorithms for genomic data analysis.

Key Finding 6.2 The unprecedented throughput and constant state of change associated with NGS create challenges involving the interfaces of traditional LIMS.

Key Recommendation 6.2 An information visualization style approach to conducting controlled user studies could help to determine which form of NGS clinical reporting is most effective for clinicians to make a diagnosis or initiate/change treatment for the patient.

Key Finding 6.3 Just like other clinical laboratory tests, an NGS diagnostic test will require a reimbursement code that provides an adequate description for the insurance company as to the type of test performed. It remains to be seen if insurance companies will pay for this genomic test, highlighting the need to educate these companies on the benefits of NGS applications for patient management and outcome.

Key Recommendation 6.3 New regulatory guidelines and insurance reimbursement codes for the use of NGS testing in the clinical microbiology laboratory need to be developed. Insurance billing codes should be revised to enhance the transparency of molecular services that are performed. Additionally, payers such as CMS should review publications showcasing that NGS assays guided or improved diagnostic and therapeutic decisions that could not be made using current laboratory methods.

Key Finding 6.4 Microbial diagnostic NGS clinical trials are not documented on the Clinical Trials.gov website (http://www.clinicaltrials.gov). It is possible that these types of trials are ongoing but are not publicized on this website. Clinical trials will need to be conducted at every step of the NGS workflow or conducted as part of a systems approach.

Key Recommendation 6.4 Outcome analyses and clinical trials highlighting the success and cost savings of NGS for the diagnosis of infectious diseases are highly recommended and could serve as justification for reimbursement companies. Incentive and, more specifically, funding, which are not widely available, must be given to initiate such studies. Therefore, an advisory board composed of relevant stakeholders should be created to address this issue.

Section 7
Establishment of a Curated Microbial Diagnostic NGS Database—A Monumental but Necessary Undertaking

7a. Database Creation and Maintenance of Whole-Genome Microbial Sequences

Efforts are under way at the FDA for the creation of a microbial reference database (i.e., FDA-ARGOS Public Regulatory-Grade Reference Database; NCBI BioProject 231221 accession no. PRJNA231221) populated with high-quality, accurately annotated genomic reference sequences that will serve as a reliable resource for NGS analysis, both for developers and clinical end users. The coordination, preservation, and updating of FDA-ARGOS are collaborative efforts among different agencies, including the FDA, NCBI, the Department of Defense (DOD) Medical Countermeasure Systems (MCS), the U.S. Army Medical Research Institute for Infectious Diseases (USAMRIID), the Defense Threat Reduction Agency (DTRA), and the Institute for Genomic Sciences at the University of Maryland. This public database is intended to provide the most accurate microbial genomic sequences (i) to inform and guide treatment decisions and (ii) to streamline the regulatory pathway for clearance and approval of NGS platforms for microbe identification by acting as a single-comparator method (111). With existing collections, there are notable gaps in deposited sequences, emphasizing the need for a broader diversity and depth of microbial submissions (27). It is important to have curated and complete reference databases in order to prevent false positives and public health scares, as evidenced by the high-profile "Anthrax in the Subway" article by Afshinnekoo *et al* (168). For many submissions in the NCBI BLAST databases, uploaded partial sequences can be contaminated, thereby complicating evaluations that are trying to search for novel microbes with low prevalence. Contaminated entries could prevent discovery of such microbes. Also, low sampling depth, common to *de novo* genome assembly, can result in incomplete and uneven read coverage for an organism of interest and consequently cause unsuspected contamination (80). The regulatory-grade sequences that will make up the FDA-ARGOS database aim to prevent contamination and serve as a major resource for diagnostic sequencing and ongoing surveillance studies.

Another FDA genomic database known as GenomeTrakr is building databases with the NCBI (see example under BioProject accession no. PRJNA183844) for foodborne pathogens for source tracking in real time (105). This database includes isolates from clinical, environmental, and food sources to embrace the One Health Initiative as well as to provide full public access to draft genomes circulating globally. Laboratories at numerous state and federal departments of health have their own GenomeTrakr bioprojects for the isolates that are important for those areas. The FDA NCBI system should work for any human pathogen, and early database building is beginning under NCBI's pathogen identification program, of which GenomeTrakr is one part for foodborne pathogens. Clinical and public health microbiology laboratories are on the front lines and are the first to encounter outbreak isolates and novel pathogens. These environments will play a key role in depositing these new genomes in the expanding knowledge base.

Assembly of reference databases will require expert consensus and feedback from the infectious disease community and the medical community to determine which organismal sequences are necessary to include. Organisms that have a high clinical index of reportability or are observed in 90% of cases in the laboratory should be included within the database along with rule-out organisms. However, there are also organisms that are not commonly seen in the laboratory but are clinically actionable. For example, *Leptospira* requires clinical intervention but is not regularly detected in the microbiology laboratory. Furthermore, a common skin contaminant, *Malassezia*, is not pathogenic for healthy individuals. In the context of an immunocompromised patient or neonate, *Malassezia* can act like a pathogen and cause sepsis. It is not necessary or cost-efficient to include all known environmental organisms, but a variety of these rare and unusual pathogens should be deposited in reference NGS microbial databases. It would be ideal to craft reference databases for each infectious disease etiological agent, including bacteria, viruses, fungi, yeasts, and parasites. The purpose of the databases is to capture the breadth of diversity for each infectious agent category, and thus more than one genome per species should be denoted. The databases should also be able to distinguish between pathogens and their near neighbors. Antibiotic resistance genes, phenotypic susceptibility data, and virulence factors could be represented as subsets within existing databases. For example, the

NCBI, as part of a multi-group effort, has established a template to capture antimicrobial susceptibility phenotype information for those organisms submitted to the BioSample database (http://www.ncbi.nlm.nih.gov/biosample/docs/antibiogram/). Additionally, it is imperative to "version" the databases when significant additions or updates are made. Finally, there needs to be a culture repository of these collected reference strains if a reexamination of the isolate is required.

For the greater good of public health, it is important to have NGS data publically available, filtered with metadata, so that patients are not identifiable. From a computational standpoint, the only way that this predictive strategy will work is if all of the genomic data are centralized. If sequence data are isolated in individual hospitals and not shared, this centralized approach will be nearly impossible to undertake (7). Individualized databases would represent only a single city's microbiome instead of that of the whole country or the whole world. Furthermore, the NIH issued a mandate in 2003 to promote data sharing from principal investigators requesting $500,000 or more in grant support. In January 2015, the NIH announced that all grantees are required to publicize large genomic data sets by the time of publication (166). One would imagine that this requirement would help tremendously in adding genomic sequence information to public repositories, but enforcement has been meek (214, 215).

7b. Representation of an NGS Diagnostic Test Result in a Patient's Medical Record

It was suggested that none of the actual genomic data should be incorporated into the patient's electronic medical record but rather the clinically actionable, interpreted result should be stated in the patient's file. The version of the software and the genomic reference sequence database should also be noted in the medical record. The suggestion that only the actionable result be stored is partly due to the challenges of laboratory management systems and hospital information systems, which largely cannot support the bioinformatics component of NGS. Hospitals can choose to store the bioinformatics data locally or to utilize off-site remote hosting like Amazon cloud support, as long as these systems uphold HIPAA compliance and security requirements. Clinical sequences could be deposited in the database, where personally identifiable information could be held locally and disconnected from the public data. All information in the reference database should be legally nondiscoverable. A sophisticated link from the patient's medical record to the reference database would benefit the broader community (30).

SECTION 7 FINDINGS AND RECOMMENDATIONS

Key Finding 7.1 There are many repositories of genomic sequence data, most of which are not publically available. It would be beneficial to assimilate concise reference databases for etiological agents such as bacteria, viruses, fungi, and parasites and include both commonly encountered pathogens and clinically actionable pathogens.

Key Recommendation 7.1 A crucial recommendation is the expansion of curated and regulatory-grade microbial sequence databases in the public domain. Genomic sequence submissions should include high-quality sequence data that are accurately annotated with metadata. These databases should not be a static collection of information but should allow for local, national, and international data exchange that is in line with agreed standards (28). Additional databases are not needed, but existing databases should establish quality metrics or curation strategies to promote confidence in clinical decision-making.

Key Finding 7.2 Reference databases should be maintained and updated with emerging clinically and environmentally important microbial genomic sequence entries.

Key Recommendation 7.2 It is recommended that sequencing efforts be focused on obtaining more pertinent whole genomes for pathogenic fungal, yeast, and parasitic species.

Key Finding 7.3 Correlating genomic DNA data with phenotypic antibiotic susceptibility remains in early stages. Antibiotic resistance genes and factors produced by the organism that cause resistance should be represented within databases.

Key Recommendation 7.3 Resistance genes should be annotated as a subset within an appropriate existing database. With new genetic mechanisms of resistance frequently arising, these databases would be ongoing projects requiring active curation and reannotation efforts.

Future directions

Text box 3.
Timeline for the invention of PCR

The conception of polymerase chain reaction (PCR) by Kary Mullis in 1987 marks one of the most important scientific advancements in the molecular biology field (221, 222). Amplification of a segment of DNA into billions of copies was groundbreaking and is a common technique used routinely in various laboratories today. Although Mullis is credited with developing PCR, the concept behind this technique had been described 17 years earlier by Har Gobind Khorana and colleagues in a process called repair replication for synthesizing short DNA duplexes and single-stranded DNA by polymerases (222). The discovery of the thermostable DNA polymerase by Erlich and colleagues in 1988 was the innovative landmark that sparked the beginning of PCR utility (222, 223, 224). PCR continued to evolve with improvements in polymerase stability and fidelity. The introduction of new PCR techniques such as real-time and digital PCR has ensured that PCR will remain a fundamental method across the sciences and clinical health. Taken together, this timeline of events for the discovery and implementation of PCR can be applied to the current status of NGS's transition into the clinical microbiology laboratory (***Figure 15***). This long, drawn-out process described throughout this report is not something that is unique for NGS technologies but in actuality coincides with the evolution of PCR testing (225). Ordinary regulation bureaucracy mandates this process along with continuous enhancements in the NGS technology.

- Microbial biomarkers, molecules produced by microorganisms during an infection that can be sensitively measured (e.g., endotoxin), are helpful in both diagnostic and therapeutic processes. Much research is currently focused on cancer biomarkers. There is also tremendous interest in identifying biomarkers for sepsis, a disease that is among the top 10 causes of mortality in the United States (133). The administration of correct antibiotics is key for the survival of septic patients, and yet there are few methods that are able to rapidly identify bloodstream pathogens. Identification of pathogen-specific biomarkers or signatures such as in sepsis cases can result in a more timely diagnosis and, subsequently, more targeted therapy. Responses to treatment can also be monitored by evaluating specific host biomarker levels (135, 216). It is highly unlikely that a single biomarker will correlate with a correct diagnosis and prognosis, but this is to be determined. The power of NGS can be leveraged to identify diagnostic biomarkers of infectious disease by global transcriptome profiling involving the pathogen during infection and the host response to the pathogen insult. Therefore, future efforts could be devoted to biomarker discovery by utilizing the genomic and transcriptomic data produced by NGS to provide the clinician with both diagnostic and prognostic information. There are studies investigating the biomarkers associated with infectious diseases such as dengue, sepsis, urinary tract infections, tuberculosis, Lyme disease, and Ebola, as well as chronic diseases such as chronic fatigue syndrome and lupus (217, 218, 219, 220).

  With metagenomic sequencing for diagnostics, potential biomarker data are not routinely examined, as human sequence reads are eliminated from analysis. Strategies to utilize these data should be developed to begin the formation of an integrative pipeline. However, to use the data generated for two different purposes would be highly dependent on the nucleic acid extraction. Evaluation of biomarker signatures would complement NGS and therefore provide a highly informative classification of disease.

- The potential of NGS might open the door for many discoveries that could help advance the infectious disease field and knowledge of human biology. NGS will drive the fundamental basic science behind these discoveries.

- To maximize the impact that NGS technologies could have on many fields, more basic and applied research efforts are needed to improve methods, analytics, and platforms.

- Training opportunities for those who want to be involved in cutting-edge bioinformatics technology, either as clinicians or Ph.D. scientists, should be made available. This type of training in clinical microbiology could assist in introducing genomic expertise to the hospital. Training should also be made available for bioinformaticians to learn clinical microbiology and medical terminology. The lack of educated individuals in the clinical bioinformatics field is associated with the lack of training opportunities.

- With human genetics, many laboratories are employing genetic counselors to help with appropriate lab utilization and also interpretation of the data. Do we need something similar, such as "microbial genetic counselors," to be housed in the clinical microbiology laboratory?

- The implementation of NGS technologies in the clinical microbiology lab has been an uphill battle. A topic of discussion at the colloquium was why it is taking so long for this technology to become routine use. It was suggested that data interpretation continues until a complete answer is obtained rather than what is specifically clinically relevant. A "complete answer" and a "clinically relevant answer" are not interchangeable terms. In the hunt to produce the most complete answer, we may be delaying the ability to produce an actionable answer.

- Approximately 70% of the population will not see a clinician when they are feeling ill. Therefore, surveillance could be done in the privacy of the home or crowd-sourcing could be attempted, as with the American Gut Project, the Harvard Personal Genome Project, and the GoViral Project. One would collect their specimen and send it for sequence analysis to determine their illness. The American Gut Project is the world's largest open-source, open-access study aimed at understanding the microbial diversity of the guts of people from the United States and globally (226). Founded in 2005, the goal of the Harvard Personal Genome Project is to gain insight into human genomes to advance human genetics, biology, and health by allowing participants to openly share their personal genomes (227). GoViral is a study by Boston Children's Hospital that collects viral diagnostic information from willing participants in the community to predict an individual's risk for developing influenza (228). The goal is to help track influenza and provide a risk assessment so people can alter their public behaviors accordingly to prevent illness. Recently launched in March 2015, Open Humans is an online portal, managed by researchers at New

York University, the University of California San Diego, and Harvard Medical School, that urges U.S. citizens to share their DNA and medical data for three research studies (229). Open-source data or a crowd-sourcing mechanism is a potential way to transition NGS into the clinical microbiology laboratory more quickly.

- A decade ago, such NGS data would have been unimaginable. With the arrival of NGS, one can only imagine what the next decade will unveil. Perhaps there will be futuristic approaches such as self-reporting and do-it-yourself sequencing. Instead of going to the doctor's office, an individual could analyze the clinical sample themselves.
- When a patient is admitted to the hospital, a sample could be sequenced to determine what organisms the patient is carrying. This type of information would help to distinguish between a colonizing organism present on admission and a hospital-acquired pathogen, which ultimately impacts insurance reimbursement. Approximately 2 million patients develop hospital-acquired infections (HAIs) annually, resulting in nearly 100,000 deaths. In the United States, HAIs are the sixth leading cause of death and account for $28 to $45 billion in extra hospital costs (230). Also, the completion of this preliminary screen evaluating microbiome or host immune status enables clinicians to see if an individual may be at risk for obtaining a hospital-acquired infection. For example, analyzing the gut microbiome could determine which patients may be more susceptible to a *Clostridium difficile* infection. NGS has the potential to decrease HAIs.

Figure 15. The evolution of the PCR machine or thermal cycler.

Republished from reference 225 with permission.

Conclusion

It has been nearly a decade since scientists George Church and Jonathan Rothberg of 454 Life Sciences (now part of Roche) ignited the genomic revolution with the advent of NGS. The sequencing field exploded with Church and Rothberg's publication highlighting the successful sequencing of a 580-kb bacterial genome in 4 hours (231). Since the introduction of NGS, there has been a doubling of sequence output nearly every 5 months, which has outpaced Moore's law and has caused sequencing costs to plummet (65, 174, 232, 233). At present, the genomics field is a steeplechase, with sequence manufacturers constantly pushing the level of innovation with platform modifications. With higher-throughput abilities and lower per base sequencing costs, NGS has become an attractive diagnostic and a state-of-the-art genomic technique. No field or industry has witnessed such a rapid escalation in processing speeds coupled with such a dramatic decrease in cost as genomics. The power of these technologies has been harnessed for applications in the medical, biomedical, basic, and life sciences, with many other uses on the horizon (112). NGS is a quickly evolving field with extensive cross talk among many disciplines. What is showcased in this document represents a snapshot of the current status and practice of NGS technology as a "disruptive innovation" in the clinical microbiology laboratory. NGS has enabled the clinical laboratory to expand its test menu from single-gene targets with Sanger sequencing to WGS, a sign of phenomenal progress (7, 8, 40).

Undoubtedly, NGS technologies are progressing and will eventually cross the divide between microbial research and the practice of diagnostic microbiology. NGS has made its mark on many fields and is anticipated to cause a real paradigm shift in how microbiology is performed when it is officially adopted as a clinical test (30). NGS platforms are highly versatile and offer a wide diagnostic repertoire that has the capabilities of replacing multiple clinical techniques (7, 8, 25). It is likely that a blend of NGS applications and traditional methodologies will be used in the near future. Perhaps NGS will not be the ultimate test but rather will serve as the underpinnings for more-precise molecular diagnostic assays and function as another tool in our clinical microbiology toolbox (6, 97).

To achieve the promise of an NGS-based diagnostic infectious disease test, clinical laboratories will need to overcome a number of operational, technical, regulatory, and strategic challenges (121). Overall, the complexity of an NGS-based clinical test is far greater than a Sanger sequencing-based test (39). Methods for collecting NGS sequence data could outstrip our capacity to adequately analyze the data (1). Although genomes can be fully sequenced in hours and for pennies per base, the daunting task of how to handle and exploit the overload of data calls for urgent attention. Much work still needs to be completed to establish a clinical diagnostic role for NGS, including user-friendly bioinformatics channels, a regulatory framework for approval, sufficient training, and thorough microbial databases. Currently there are incredible enthusiasm and interest in NGS capabilities. There needs to be a way to transform this zeal into practical, working solutions for mainstream diagnostic use.

The fields of public and clinical microbiology remain optimistic for the future of NGS to deliver tangible clinical benefits to improve patient care, patient outcomes, and public health. NGS would allow public health microbiology laboratories to become more proactive than reactive in identifying potential pathogenic species and lineages that could trigger major outbreaks (105). In an age in which new multidrug-resistant pathogens continue to emerge while the antimicrobial pipeline remains stagnant, we can be preemptive and use NGS as a valuable tool to monitor, detect, and control threatening agents locally, nationally, and internationally (7, 30, 61). Although NGS has been marketed for over a decade, widespread use of genomics in the clinical microbiology laboratory is not quite ready for prime time. Given adequate time to overcome the roadblocks, NGS will have profound implications in critical decision support systems and clinical care pathways. NGS represents a technological leap into the future of precision medicine by guiding preventative measures, diagnosis, and therapeutic options for the patient. It is without a doubt that cutting-edge NGS technologies will become a cornerstone of clinical care.

References

1. Rizzo JM, Buck MJ: **Key principles and clinical applications of "next-generation" DNA sequencing**. *Cancer Prevention Research* 2012, **5**(7):887-900. doi:10.1158/1940-6207.

2. Mardis ER: **The impact of next-generation sequencing technology on genetics**. *Trends in Genetics* 2008, **24**(3):133-141. doi:10.1016/j.tig.2007.12.007.

3. Mardis ER: **Next-generation DNA sequencing methods**. *Annual Review of Genomics and Human Genetics* 2008, **9**:387-402. doi:10.1146/annurev.genom.9.081307.164359.

4. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors**. *Proceedings of the National Academy of Sciences of the United States of America* 1977, **74**(12):5463-5467.

5. Voelkerding KV, Dames SA, Durtschi JD: **Next-generation sequencing: from basic research to diagnostics**. *Clinical Chemistry* 2009, **55**(4):641-658. doi:10.1373/clinchem.2008.112789.

6. Dunne WM, Jr., Westblade LF, Ford B: **Next-generation and whole-genome sequencing in the diagnostic clinical microbiology laboratory**. *European Journal of Clinical Microbiology & Infectious Diseases* 2012, **31**(8):1719-1726. doi:10.1007/s10096-012-1641-7.

7. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E: **ACMG clinical laboratory standards for next-generation sequencing**. *Genetics in Medicine* 2013, **15**(9):733-747. doi:10.1038/gim.2013.92.

8. Shendure J, Ji H: **Next-generation DNA sequencing**. *Nature Biotechnology* 2008, **26**(10):1135-1145. doi:10.1038/nbt1486.

9. Kwong JC, McCallum N, Sintchenko V, Howden BP: **Whole genome sequencing in clinical and public health microbiology**. *Pathology* 2015, **47**(3):199-210. doi:10.1097/PAT.0000000000000235.

10. Gullapalli RR, Desai KV, Santana-Santos L, Kant JA, Becich MJ: **Next generation sequencing in clinical medicine: challenges and lessons for pathology and biomedical informatics**. *Journal of Pathology Informatics* 2012, **3**:40. doi:10.4103/2153-3539.103013.

11. Diaz-Sanchez S, Hanning I, Pendleton S, D'Souza D: **Next-generation sequencing: the future of molecular genetics in poultry production and food safety**. *Poultry Science* 2013, **92**(2):562-572. doi:10.3382/ps.2012-02741.

12. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z *et al*: **Genome sequencing in open microfabricated high density picoliter reactors**. *Nature* 2005, **437**(7057):376-380. doi:10.3382/ps.2012-02741.

13. Barzon L, Lavezzo E, Militello V, Toppo S, Palù G: **Applications of next-generation sequencing technologies to diagnostic virology**. *International Journal of Molecular Sciences* 2011, **12**(11):7861-7884. doi:10.3390/ijms12117861.

14. Wilson MR, Allard MW, Brown EW: **The forensic analysis of foodborne bacterial pathogens in the age of whole-genome sequencing**. *Cladistics* 2013, **29**(4):449-461.

15. Koren S, Phillippy AM: **One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly**. *Current Opinion in Microbiology* 2015, **23**:110-120. doi:10.1016/j.mib.2014.11.014.

16. Sboner A, Mu X, Greenbaum D, Auerbach R, Gerstein M: **The real cost of sequencing: higher than you think!** *Genome Biology* 2011, **12**(8):125. doi:10.1186/gb-2011-12-8-125.

17. Ong FS, Lin JC, Das K, Grosu DS, Fan JB: **Translational utility of next-generation sequencing**. *Genomics* 2013, **102**(3):137-139. doi:10.1016/j.ygeno.2013.04.012.

18. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY *et al*: **Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010**. *The Lancet* 2012, **380**(9859):2095-2128. doi:10.1016/S0140-6736(12)61728-0.

19. Goldberg B, Sichtig H, Geyer C, Ledeboer N, Weinstock GM: **Making the leap from research laboratory to clinic: challenges and opportunities for next-generation sequencing in infectious disease diagnostics.** *mBio* 2015, **6**(6):e01888-15. doi:10.1128/mBio.01888-15.

20. Dickson RP, Erb-Downward JR, Prescott HC, Martinez FJ, Curtis JL, Lama VN, Huffnagle GB: **Analysis of culture-dependent versus culture-independent techniques for identification of bacteria in clinically obtained bronchoalveolar lavage fluid**. *Journal of Clinical Microbiology* 2014, **52**(10):3605-3613. doi:10.1128/JCM.01028-14.

21. Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, Salamat SM, Somasekar S, Federman S, Miller S *et al*: **Actionable diagnosis of neuroleptospirosis by next-generation sequencing**. *The New England Journal of Medicine* 2014, **370**(25):2408-2417. doi:10.1056/NEJMoa1401268.

22. Swartz A: **DNA Tests for Infectious Diseases**. In: *Genome*. Susan McClure; 2015. http://genomemag.com/dna-tests-for-infectious-diseases/#.VgFqB8tQZ1M.

23. Check W: **For viral diagnosis, metagenomics NGS**. In: *CAP Today*. vol. June 2015: 2015.

24. Sherry NL, Porter JL, Seemann T, Watkins A, Stinear TP, Howden BP: **Outbreak investigation using high-throughput genome sequencing within a diagnostic microbiology laboratory**. *Journal of Clinical Microbiology* 2013, **51**(5):1396-1401. doi:10.1128/JCM.03332-12.

25. Luheshi L, Raza S, Moorthie S, Hall A, Blackburn L, Rands C, Sagoo G, Chowdhury S, Kroese M, Burton H: ***Pathogen Genomics Into Practice***. PHG Foundation (2015). ISBN 978-1-907198-18-2.

26. Institute of Medicine, National Research Council: **Technologies to enable autonomous detection for BioWatch: ensuring timely and accurate information for public health officials:** Workshop Summary. Washington, DC: The National Academies Press; 2014.

27. Long SW, Williams D, Valson C, Cantu CC, Cernoch P, Musser JM, Olsen RJ: **A genomic day in the life of a clinical microbiology laboratory**. *Journal of Clinical Microbiology* 2013, **51**(4):1272-1277. doi:10.1128/JCM.03237-12.

28. Koser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, Holden MT, Dougan G, Bentley SD, Parkhill J *et al*: **Routine use of microbial whole genome sequencing in diagnostic and public health microbiology**. *PLoS Pathogens* 2012, **8**(8):e1002824. doi:10.1371/journal.ppat.1002824.

29. Koser CU, Fraser LJ, Ioannou A, Becq J, Ellington MJ, Holden MT, Reuter S, Torok ME, Bentley SD, Parkhill J *et al*: **Rapid single-colony whole-genome sequencing of bacterial pathogens**. *The Journal of Antimicrobial Chemotherapy* 2014, **69**(5):1275-1281. doi:10.1093/jac/dkt494.

30. Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW: **Transforming clinical microbiology with bacterial genome sequencing**. *Nature Reviews Genetics* 2012, **13**(9):601-612. doi:10.1038/nrg3226.

31. Olsen RJ, Long SW, Musser JM: **Bacterial genomics in infectious disease and the clinical pathology laboratory**. *Archives of Pathology & Laboratory Medicine* 2012, **136**(11):1414-1422. doi:10.5858/arpa.2012-0025-RA.

32. Saah AJ, Hoover DR: **"Sensitivity" and "specificity" reconsidered: the meaning of these terms in analytical and diagnostic settings**. *Annals of Internal Medicine* 1997, **126**(1):91-94.

33. Price JR, Didelot X, Crook DW, Llewelyn MJ, Paul J: **Whole genome sequencing in the prevention and control of Staphylococcus aureus infection**. *Journal of Hospital Infection* 2013, **83**(1):14-21. doi:10.1016/j.jhin.2012.10.003.

34. Collins FS, Varmus H: **A new initiative on precision medicine**. *New England Journal of Medicine* 2015, **372**(9):793-795. doi:10.1056/NEJMp1500523.

35. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Özen S, Sanjad S *et al*: **Genetic diagnosis by whole exome capture and massively parallel DNA sequencing**. *Proceedings of the National Academy of Sciences* 2009, **106**(45):19096-19101. doi:10.1073/pnas.0910672106.

36. Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, Serpe JM, Dasu T, Tschannen MR, Veith RL *et al*: **Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease**. *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 2011, **13**(3):255-262. doi:10.1097/GIM.0b013e3182088158.

37. Gargis AS, Kalman L, Berry MW, Bick DP: **Assuring the quality of next-generation sequencing in clinical laboratory practice**. *Nature Biotechnology* 2012, **30**(11):1033-1036. doi:10.1038/nbt.2403.

38. Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, Penn CW, Robinson ER, Pallen MJ: **High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity**. *Nature Reviews Microbiology* 2012, **10**(9):599-606. doi:10.1038/nrmicro2850.

39. Aziz N, Zhao Q, Bry L, Driscoll DK, Funke B, Gibson JS, Grody WW, Hegde MR, Hoeltge GA, Leonard DGB *et al*: **College of American Pathologists' laboratory standards for next-generation sequencing clinical tests**. *Archives of Pathology & Laboratory Medicine* 2015, **139**(4):481-493. doi:10.5858/arpa.2014-0250-CP.

40. Rehm HL: **Disease-targeted sequencing: a cornerstone in the clinic**. *Nature Reviews Genetics* 2013, **14**(4):295-300. doi:10.1038/nrg3463.

41. Berglund EC, Kiialainen A, Syvanen AC: **Next-generation sequencing technologies and applications for human genetic history and forensics**. *Investigative Genetics* 2011, **2**:23. doi:10.1186/2041-2223-2-23. doi:10.1186/2041-2223-2-23.

42. Kuroda M, Sekizuka T, Shinya F, Takeuchi F, Kanno T, Sata T, Asano S: **Detection of a possible bioterrorism agent, Francisella sp., in a clinical specimen by use of next-generation direct DNA sequencing**. *Journal of Clinical Microbiology* 2012, **50**(5):1810-1812. doi:10.1128/JCM.06715-11.

43. Budowle B, Connell ND, Bielecka-Oder A, Colwell RR, Corbett CR, Fletcher J, Forsman M, Kadavy DR, Markotic A, Morse SA *et al*: **Validation of high throughput sequencing and microbial forensics applications**. *Investigative Genetics* 2014, **5**:9. doi:10.1186/2041-2223-5-9.

44. Zinsstag J, Schelling E, Waltner-Toews D, Tanner M: **From "one medicine" to "one health" and systemic approaches to health and well-being**. *Preventive Veterinary Medicine* 2011, **101**(3-4):148-156. doi:10.1016/j.prevetmed.2010.07.003.

45. **One Health Initiative.** [http://www.onehealthinitiative.com/about.php]. Accessed June 28, 2015.

46. Cho I, Blaser MJ: **The human microbiome: at the interface of health and disease**. *Nature Reviews Genetics* 2012, **13**(4):260-270. doi:10.1038/nrg3182.

47. Lozupone CA, Stombaugh J, Gonzalez A, Ackermann G, Wendel D, Vázquez-Baeza Y, Jansson JK, Gordon JI, Knight R: **Meta-analyses of studies of the human microbiota**. *Genome Research* 2013, **23**(10):1704-1714. doi:10.1101/gr.151803.112.

48. Surette MG: **The cystic fibrosis lung microbiome**. *Annals of the American Thoracic Society* 2014, **11**(Suppl 1):S61-S65. doi:10.1513/AnnalsATS.201306-159MG.

49. Guinane CM, Cotter PD: **Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ**. *Therapeutic Advances in Gastroenterology* 2013, **6**(4):295-308. doi:10.1177/1756283X13482996.

50. Lambert JA, John S, Sobel JD, Akins RA: **Longitudinal analysis of vaginal microbiome dynamics in women with recurrent bacterial vaginosis: recognition of the conversion process**. *PLoS One* 2013, **8**(12):e82599. doi:10.1371/journal.pone.0082599.

51. Tang WHW, Wang Z, Levison BS, Koeth RA, Britt EB, Fu X, Wu Y, Hazen SL: **Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk**. *The New England Journal of Medicine* 2013, **368**(17):1575-1584. doi:10.1056/NEJMoa1109400.

52. Hansen T, Gobel R, Hansen T, Pedersen O: **The gut microbiome in cardio-metabolic health**. *Genome Medicine* 2015, **7**(1):33. doi:10.1186/s13073-015-0157-z.

53. Iskander KN, Osuchowski MF, Stearns-Kurosawa DJ, Kurosawa S, Stepien D, Valentine C, Remick DG: **Sepsis: multiple abnormalities, heterogeneous responses, and evolving understanding**. *Physiological Reviews* 2013, **93**(3):1247-1288. doi:10.1152/physrev.00037.2012.

54. Ulloa L, Brunner M, Ramos L, Deitch EA: **Scientific and clinical challenges in sepsis**. *Current Pharmaceutical Design* 2009, **15**(16):1918-1935.

55. Seam N, Suffredini AF: **Mechanisms of sepsis and insights from clinical trials**. *Drug Discovery Today Disease Mechanisms* 2007, **4**(2):83-93.

56. Relman DA: **Microbial genomics and infectious diseases**. *The New England Journal of Medicine* 2011, **365**(4):347-357. doi:10.1056/NEJMra1003071.

57. Solieri L, Dakal T, Giudici P: **Next-generation sequencing and its potential impact on food microbial genomics**. *Annals of Microbiology* 2013, **63**(1):21-37.

58. **News Briefs**. *Genetics in Medicine* 2014, **16**(8):569-570.

59. **FDA investigates presence of Listeria in some Hispanic-style cheeses.** [http://www.fda.gov/Food/RecallsOutbreaksEmergencies/Outbreaks/ucm386726.htm]. Accessed June 28, 2015.

60. **Whole genome sequencing program (WGS).** [http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/default.htm]. Accessed June 28, 2015.

61. Hoffmann M, Luo Y, Monday SR, Gonzalez-Escalona N, Ottesen AR, Muruvanda T, Wang C, Kastanis G, Keys C, Janies D *et al*: **Tracing origins of the Salmonella Bareilly strain causing a food-borne outbreak in the United States**. *Journal of Infectious Diseases* 2015. doi:10.1093/infdis/jiv297.

62. Shokralla S, Spall JL, Gibson JF, Hajibabaei M: **Next-generation sequencing technologies for environmental DNA research**. *Molecular Ecology* 2012, **21**(8):1794-1805. doi:10.1111/j.1365-294X.2012.05538.x.

63. Duhoux A, Carrère S, Gouzy J, Bonin L, Délye C: **RNA-Seq analysis of rye-grass transcriptomic response to an herbicide inhibiting acetolactate-synthase identifies transcripts linked to non-target-site-based resistance**. *Plant Molecular Biology* 2015, **87**(4-5):473-487. doi:10.1007/s11103-015-0292-3.

64. Varshney RK, May GD: **Next-generation sequencing technologies: opportunities and obligations in plant genomics**. *Briefings in Functional Genomics* 2012, **11**(1):1-2. doi:10.1093/bfgp/els001.

65. Morozova O, Marra MA: **Applications of next-generation sequencing technologies in functional genomics**. *Genomics* 2008, **92**(5):255-264. doi:10.1016/j.ygeno.2008.07.001.

66. Grada A, Weinbrecht K: **Next-generation sequencing: methodology and application**. *The Journal of Investigative Dermatology* 2013, **133**(8):e11. doi:10.1038/jid.2013.248.

67. Yongfeng H, Fan Y, Jie D, Jian Y, Ting Z, Lilian S, Jin Q: **Direct pathogen detection from swab samples using a new high-throughput sequencing technology**. *Clinical Microbiology and Infection* 2011, **17**(2):241-244. doi:10.1111/j.1469-0691.2010.03246.x.

68. Salipante SJ, Hoogestraat DR, Abbott AN, SenGupta DJ, Cummings LA, Butler-Wu SM, Stephens K, Cookson BT, Hoffman NG: **Coinfection of Fusobacterium nucleatum and Actinomyces israelii in mastoiditis diagnosed by next-generation DNA sequencing**. *Journal of Clinical Microbiology* 2014, **52**(5):1789-1792. doi:10.1128/JCM.03133-13.

69. Jones MG: **The first filamentous fungal genome sequences: Aspergillus leads the way for essential everyday resources or dusty museum specimens?** *Microbiology* 2007, **153**(1):1-6.

70. Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, Conlan S, Quan P-L, Hui J, Marshall J *et al*: **A new arenavirus in a cluster of fatal transplant-associated diseases**. *New England Journal of Medicine* 2008, **358**(10):991-998. doi:10.1056/NEJMoa073785.

71. Whitley R: **The new age of molecular diagnostics for microbial agents**. *New England Journal of Medicine* 2008, **358**(10):988-989. doi: 10.1056/NEJMp0708085.

72. Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G, Khristova ML, Weyer J, Swanepoel R, Egholm M *et al*: **Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa**. *PLoS Pathogens* 2009, **5**(5):e1000455. doi:10.1371/journal.ppat.1000455.

73. Grard G, Fair JN, Lee D, Slikas E, Steffen I, Muyembe J-J, Sittler T, Veeraraghavan N, Ruby JG, Wang C *et al*: **A novel rhabdovirus associated with acute hemorrhagic fever in Central Africa**. *PLoS Pathogens* 2012, **8**(9):e1002924. doi:10.1371/journal.ppat.1002924.

74. Chiu CY, Yagi S, Lu X, Yu G, Chen EC, Liu M, Dick EJ, Carey KD, Erdman DD, Leland MM *et al*: **A novel adenovirus species associated with an acute respiratory outbreak in a baboon colony and evidence of coincident human infection**. *mBio* 2013, **4**(2):e00084-13. doi:10.1128/mBio.00084-13.

75. Chiu CY: **Viral pathogen discovery**. *Current Opinion in Microbiology* 2013, **16**(4):468-478. doi:10.1016/j.mib.2013.05.001.

76. Quiñones-Mateu ME, Avila S, Reyes-Teran G, Martinez MA: **Deep sequencing: becoming a critical tool in clinical virology**. *Journal of Clinical Virology* 2014, **61**(1):9-19. doi:10.1016/j.jcv.2014.06.013.

77. Mokili JL, Rohwer F, Dutilh BE: **Metagenomics and future perspectives in virus discovery**. *Current Opinion in Virology* 2012, **2**(1):63-77. doi:10.1016/j.coviro.2011.12.004.

78. Wylie KM, Weinstock GM, Storch GA: **Virome genomics: a tool for defining the human virome**. *Current Opinion in Microbiology* 2013, **16**(4):479-484. doi:10.1016/j.mib.2013.04.006.

79. Capobianchi MR, Giombini E, Rozera G: **Next-generation sequencing technology in clinical virology**. *Clinical Microbiology and Infection* 2013, **19**(1):15-22. doi:10.1111/1469-0691.12056.

80. Laurence M, Hatzis C, Brash DE: **Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes**. *PLoS One* 2014, **9**(5):e97876. doi:10.1371/journal.pone.0097876.

81. Feng H, Shuda M, Chang Y, Moore PS: **Clonal integration of a polyomavirus in human Merkel cell carcinoma**. *Science* 2008, **319**(5866):1096-1100. doi:10.1126/science.1152586.

82. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW: **Reagent and laboratory contamination can critically impact sequence-based microbiome analyses**. *BMC Biology* 2014, **12**:87. doi:10.1186/s12915-014-0087-z.

83. Hino S, Miyata H: **Torque teno virus (TTV): current status**. *Reviews in Medical Virology* 2007, **17**(1):45-57.

84. Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, Aronsohn A, Hackett J, Delwart EL, Chiu CY: **The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns**. *Journal of Virology* 2013, **87**(22):11966-11977. doi:10.1128/JVI.02323-13.

85. Lee D, Das Gupta J, Gaughan C, Steffen I, Tang N, Luk K-C, Qiu X, Urisman A, Fischer N, Molinaro R *et al*: **In-depth investigation of archival and prospectively collected samples reveals no evidence for XMRV infection in prostate cancer**. *PLoS One* 2012, **7**(9):e44954. doi:10.1371/journal.pone.0044954.

86. Calistri A, Palù G: **Unbiased next-generation sequencing and new pathogen discovery: undeniable advantages and still-existing drawbacks**. *Clinical Infectious Diseases* 2015, **60**(6):889-91. doi:10.1093/cid/ciu913.

87. Shabman RS, Jabado OJ, Mire CE, Stockwell TB, Edwards M, Mahajan M, Geisbert TW, Basler CF: **Deep sequencing identifies noncanonical editing of Ebola and Marburg virus RNAs in infected cells**. *mBio* 2014, **5**(6):e02011-14. doi:10.1128/mBio.02011-14.

88. Ghedin E, Holmes EC, DePasse JV, Pinilla LT, Fitch A, Hamelin M-E, Papenburg J, Boivin G: **Presence of oseltamivir-resistant pandemic A/H1N1 minor variants before drug therapy with subsequent selection and transmission**. *The Journal of Infectious Diseases* 2012, **206**(10):1504-1511. doi:10.1093/infdis/jis571.

89. Rogers MB, Song T, Sebra R, Greenbaum BD, Hamelin M-E, Fitch A, Twaddle A, Cui L, Holmes EC, Boivin G *et al*: **Intrahost dynamics of antiviral resistance in influenza A virus reflect complex patterns of segment linkage, reassortment, and natural selection**. *mBio* 2015, **6**(2):e02464-14. doi:10.1128/mBio.02464-14.

90. Engelthaler DM, Chiller T, Schupp JA, Colvin J, Beckstrom-Sternberg SM, Driebe EM, Moses T, Tembe W, Sinari S, Beckstrom-Sternberg JS *et al*: **Next-generation sequencing of Coccidioides immitis isolated during cluster investigation**. *Emerging Infectious Diseases* 2011, **17**(2):227-232. doi:10.3201/eid1702.100620.

91. Cui L, Morris A, Ghedin E: **The human mycobiome in health and disease**. *Genome Medicine* 2013, **5**(7):63. doi:10.1186/gm467.

92. Robinson T, Campino SG, Auburn S, Assefa SA, Polley SD, Manske M, MacInnis B, Rockett KA, Maslen GL, Sanders M *et al*: **Drug-resistant genotypes and multi-clonality in Plasmodium falciparum analysed by direct genome sequencing from peripheral blood of malaria patients**. *PLoS One* 2011, **6**(8):e23204. doi:10.1371/journal.pone.0023204.

93. Winzeler EA: **Advances in parasite genomics: from sequences to regulatory networks**. *PLoS Pathogens* 2009, **5**(10):e1000649. doi:10.1371/journal.ppat.1000649.

94. Wong SS, Fung KS, Chau S, Poon RW, Wong SC, Yuen K-Y: **Molecular diagnosis in clinical parasitology: when and why?** *Experimental Biology and Medicine* 2014, **239**(11):1443-1460. doi:10.1177/1535370214523880.

95. Hasman H, Saputra D, Sicheritz-Ponten T, Lund O, Svendsen CA, Frimodt-Møller N, Aarestrup FM: **Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples**. *Journal of Clinical Microbiology* 2014, **52**(1):139-146. doi:10.1128/JCM.02452-13.

96. Naccache SN, Peggs KS, Mattes FM, Phadke R, Garson JA, Grant P, Samayoa E, Federman S, Miller S, Lunn MP *et al*: **Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing**. *Clinical Infectious Diseases* 2015, **60**(6):919-923. doi:10.1093/cid/ciu912.

97. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Henderson DK, Palmore TN, Segre JA: **Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing**. *Science Translational Medicine* 2012, **4**(148):148ra116. doi:10.1126/scitranslmed.3004129.

98. Bick D, Dimmock D: **Whole exome and whole genome sequencing**. *Current Opinion in Pediatrics* 2011, **23**(6):594-600. doi:10.1097/MOP.0b013e32834b20ec.

99. Chiu C, Miller S: **Next-generation sequencing**. *In* Persing DH et al. (ed), *Molecular Microbiology: Diagnostic Principles and Practice, Third Edition,* in press. ASM Press, Washington, DC.

100. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y: **A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers**. *BMC Genomics* 2012, **13**:341. doi:10.1186/1471-2164-13-341.

101. Perkel JM: **Next-generation DNA sequencing: 2015 update**. [http://www.biocompare.com/Editorial-Articles/171872-Next-Gen-DNA-Sequencing-2015-Update/]. Accessed June 28, 2015.

102. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B *et al*: **Real-time DNA sequencing from single polymerase molecules**. *Science* 2009, **323**(5910):133-138. doi:10.1126/science.1162986.

103. Ferrarini M, Moretto M, Ward J, Surbanovski N, Stevanovic V, Giongo L, Viola R, Cavalieri D, Velasco R, Cestaro A *et al*: **An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome**. *BMC Genomics* 2013, **14**:670. doi:10.1186/1471-2164-14-670.

104. Quick J, Quinlan AR, Loman NJ: **A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer**. *GigaScience* 2014, **3**:22. doi:10.1186/2047-217X-3-22.

105. Quick J, Ashton P, Calus S, Chatt C, Gossain S, Hawker J, Nair S, Neal K, Nye K, Peters T *et al*: **Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella**. *Genome Biology* 2015, **16**(1):114. doi:10.1186/s13059-015-0677-2.

106. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ: **Assessing the performance of the Oxford Nanopore Technologies MinION**. *Biomolecular Detection and Quantification* 2015, **3**:1-8.

107. Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M: **Improved data analysis for the MinION nanopore sequencer**. *Nature Methods* 2015, **12**(4):351-356. doi:10.1038/nmeth.3290.

108. Loman NJ, Quick J, Simpson JT: **A complete bacterial genome assembled de novo using only nanopore sequencing data**. *Nature Methods* 2015, **12**(8):733-735. doi:10.1038/nmeth.3444.

109. Eisenstein M: **Oxford Nanopore announcement sets sequencing sector abuzz**. *Nature Biotechnology* 2012, **30**(4):295-296. doi:10.1038/nbt0412-295.

110. Rusu LI, Wyres KL, Reumann M, Queiroz C, Bojovschi A, Conway T, Garg S, Edwards DJ, Hogg G, Holt KE: **A platform for leveraging next generation sequencing for routine microbiology and public health use**. *Health Information Science and Systems* 2015, **3**(Suppl 1):S7. doi:10.1186/2047-2501-3-S1-S7.

111. Sichtig H: **High-throughput sequencing technologies for microbial identification and detection of antimicrobial resistance markers**. In: *FDA Public Workshop on High Throughput Sequencing Devices in Microbial Diagnostics*. US Food and Drug Administration; White Oak Campus, MD; 2014:1-21.

112. Voelkerding KV, Dames S, Durtschi JD: **Next generation sequencing for clinical diagnostics—principles and application to targeted resequencing for hypertrophic cardiomyopathy: a paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology**. *The Journal of Molecular Diagnostics* 2010, **12**(5):539-551. doi:10.2353/jmoldx.2010.100043.

113. Trotha R, Hanck T, König W, König B: **Rapid ribosequencing—an effective diagnostic tool for detecting microbial infection**. *Infection* 2001, **29**(1):12-16.

114. Wellinghausen N, Kochem AJ, Disqué C, Mühl H, Gebert S, Winter J, Matten J, Sakka SG: **Diagnosis of bacteremia in whole-blood samples by use of a commercial universal 16S rRNA gene-based PCR and sequence analysis**. *Journal of Clinical Microbiology* 2009, **47**(9):2759-2765. doi:10.1128/JCM.00567-09.

115. Saito N, Hida A, Koide Y, Ooka T, Ichikawa Y, Shimizu J, Mukasa A, Nakatomi H, Hatakeyama S, Hayashi T *et al*: **Culture-negative brain abscess with Streptococcus intermedius infection with diagnosis established by direct nucleotide sequence analysis of the 16S ribosomal RNA gene**. *Internal Medicine* 2012, **51**(2):211-216.

116. Kommedal Ø, Wilhelmsen MT, Skrede S, Meisal R, Jakovljev A, Gaustad P, Hermansen NO, Vik-Mo E, Solheim O, Ambur OH *et al*: **Massive parallel sequencing provides new perspectives on bacterial brain abscesses**. *Journal of Clinical Microbiology* 2014, **52**(6):1990-1997. doi:10.1128/JCM.00346-14.

117. Mishra AK, Dufour H, Roche PH, Lonjon M, Raoult D, Fournier PE: **Molecular revolution in the diagnosis of microbial brain abscesses**. *European Journal of Clinical Microbiology & Infectious Diseases* 2014, **33**(12):2083-2093. doi:10.1007/s10096-014-2166-z.

118. Goyo D, Camacho A, Gómez C, de las Heras R, Otero JR, Chaves F: **False-positive PCR detection of Tropheryma whipplei in cerebrospinal fluid and biopsy samples from a child with chronic lymphocytic meningitis**. *Journal of Clinical Microbiology* 2009, **47**(11):3783-3784. doi:10.1128/JCM.00927-09.

119. Nakamura S, Yang C-S, Sakon N, Ueda M, Tougan T, Yamashita A, Goto N, Takahashi K, Yasunaga T, Ikuta K *et al*: **Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach**. *PLoS One* 2009, **4**(1):e4219. doi:10.1371/journal.pone.0004219.

120. Hurd PJ, Nelson CJ: **Advantages of next-generation sequencing versus the microarray in epigenetic research**. *Briefings in Functional Genomics & Proteomics* 2009, **8**(3):174-183. doi:10.1093/bfgp/elp013.

121. Wain J, Mavrogiorgou E: **Next-generation sequencing in clinical microbiology**. *Expert Review of Molecular Diagnostics* 2013, **13**(3):225-227. doi:10.1586/erm.13.8.

122. Salipante SJ, Kalapila A, Pottinger PS, Hoogestraat DR, Cummings L, Duchin JS, Sengupta DJ, Pergam SA, Cookson BT, Butler-Wu SM: **Characterization of a multidrug-resistant, novel Bacteroides genomospecies**. *Emerging Infectious Diseases* 2015, **21**(1):95-98. doi:10.3201/eid2101.140662.

123. Croxatto A, Prod'hom G, Greub G: **Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology**. *FEMS Microbiology Reviews* 2012, **36**(2):380-407. doi:10.1111/j.1574-6976.2011.00298.x.

124. Clark AE, Kaleta EJ, Arora A, Wolk DM: **Matrix-assisted laser desorption ionization–time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology.** *Clinical Microbiology Reviews* 2013, **26**(3):547-603. doi:10.1128/CMR.00072-12.

125. **The FilmArray System.** [http://filmarray.com/]. Accessed June 28, 2015.

126. Popowitch EB, O'Neill SS, Miller MB: **Comparison of the BioFire FilmArray RP, Genmark eSensor RVP, Luminex xTAG RVPv1, and Luminex xTAG RVP fast multiplex assays for detection of respiratory viruses**. *Journal of Clinical Microbiology* 2013, **51**(5):1528-1533. doi:10.1128/JCM.03368-12.

127. **Infectious Disease.** [https://www.luminexcorp.com/clinical/infectious-disease/]. Accessed August 12, 2015.

128. **eSensor Respiratory Viral Panel.** [http://www.genmarkdx.com/products/reagents/rvp.php]. Accessed June 28, 2015.

129. Prachayangprecha S, Schapendonk CME, Koopmans MP, Osterhaus ADME, Schürch AC, Pas SD, van der Eijk AA, Poovorawan Y, Haagmans BL, Smits SL: **Exploring the potential of next-generation sequencing in detection of respiratory viruses**. *Journal of Clinical Microbiology* 2014, **52**(10):3722-3730. doi:10.1128/JCM.01641-14.

130. Law JWF, Ab Mutalib NS, Chan KG, Lee LH: **Rapid methods for the detection of foodborne bacterial pathogens: principles, applications, advantages and limitations**. *Frontiers in Microbiology* 2014, **5**:770. doi:10.3389/fmicb.2014.00770.

131. Brealey D, Libert N, Pugin J, O'Dwyer M, Zacharowski K, Mikaszewska-Sokolewicz M, Maureau MP, Ecker DJ, Sampath R, Singer M *et al*: **RADICAL study: rapid diagnosis of suspected bloodstream infections from direct blood testing using PCR/ESI-MS**. *Critical Care* 2014, **18**(Suppl 2):P61.

132. Wolk DM, Kaleta EJ, Wysocki VH: **PCR–electrospray ionization mass spectrometry: the potential to change infectious disease diagnostics in clinical and public health laboratories**. *The Journal of Molecular Diagnostics* 2012, **14**(4):295-304. doi:10.1016/j.jmoldx.2012.02.005.

133. Kaleta EJ, Clark AE, Cherkaoui A, Wysocki VH, Ingram EL, Schrenzel J, Wolk DM: **Comparative analysis of PCR–electrospray ionization/mass spectrometry (MS) and MALDI-TOF/MS for the identification of bacteria and yeast from positive blood culture bottles**. *Clinical Chemistry* 2011, **57**(7):1057-1067. doi:10.1373/clinchem.2011.161968.

134. **Treatment Guidance: Quickly rule-in therapy for better patient outcomes.** [http://www.geneweave.com/clinical]. Accessed June 28, 2015.

135. Mejias A, Ramilo O: **Transcriptional profiling in infectious diseases: Ready for prime time?** *Journal of Infection* 2014, **68**:S94-S99. doi:10.1016/j.jinf.2013.09.018.

136. Sweeney TE, Shidham A, Wong HR, Khatri P: **A comprehensive time-course–based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set**. *Science Translational Medicine* 2015, **7**(287):287ra271. doi:10.1126/scitranslmed.aaa5993.

137. Price EP, Seymour ML, Sarovich DS, Latham J, Wolken SR, Mason J, Vincent G, Drees KP, Beckstrom-Sternberg SM, Phillippy AM *et al*: **Molecular epidemiologic investigation of an anthrax outbreak among heroin users, Europe**. *Emerging Infectious Disease Journal* 2012, **18**(8):1307-1313. doi:10.3201/eid1808.111343.

138. Cunningham SA, Sloan LM, Nyre LM, Vetter EA, Mandrekar J, Patel R: **Three-hour molecular detection of Campylobacter, Salmonella, Yersinia, and Shigella species in feces with accuracy as high as that of culture**. *Journal of Clinical Microbiology* 2010, **48**(8):2929-2933. doi:10.1128/JCM.00339-10.

139. Seth-Smith HMB, Harris SR, Skilton RJ, Radebe FM, Golparian D, Shipitsyna E, Duy PT, Scott P, Cutcliffe LT, O'Neill C *et al*: **Whole-genome sequences of Chlamydia trachomatis directly from clinical samples without culture**. *Genome Research* 2013, **23**(5):855-866. doi:10.1101/gr.150037.112.

140. Andersson P, Klein M, Lilliebridge RA, Giffard PM: **Sequences of multiple bacterial genomes and a Chlamydia trachomatis genotype from direct sequencing of DNA derived from a vaginal swab diagnostic specimen**. *Clinical Microbiology and Infection* 2013, **19**(9):E405-E408. doi:10.1111/1469-0691.12237.

141. Loman NJ, Constantinidou C, Christner M, Rohde H, Chan JZ, Quick J, Weir JC, Quince C, Smith GP, Betley JR, *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**(14):1502-1510. doi:10.1001/jama.2013.3231.

142. Brown PS, Pope CE, Marsh RL, Qin X, McNamara S, Gibson R, Burns JL, Deutsch G, Hoffman LR: **Directly sampling the lung of a young child with cystic fibrosis reveals diverse microbiota**. *Annals of the American Thoracic Society* 2014, **11**(7):1049-1055. doi:10.1513/AnnalsATS.201311-383OC.

143. Pecora ND, Li N, Allard M, Li C, Albano E, Delaney M, Dubois A, Onderdonk AB, Bry L: **Genomically informed surveillance for carbapenem-resistant Enterobacteriaceae in a health care system**. *mBio* 2015, **6**(4):e01030-15. doi:10.1128/mBio.01030-15.

144. Tyson GH, McDermott PF, Li C, Chen Y, Tadesse DA, Mukherjee S, Bodeis-Jones S, Kabera C, Gaines SA, Loneragan GH *et al*: **WGS accurately predicts antimicrobial resistance in Escherichia coli**. *Journal of Antimicrobial Chemotherapy* 2015, doi:10.1093/jac/dkv18.

145. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM *et al*: **Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.** *Science* 1995, **269**(5223):496-512.

146. Török ME, Peacock SJ: **Rapid whole-genome sequencing of bacterial pathogens in the clinical microbiology laboratory—pipe dream or reality?** *Journal of Antimicrobial Chemotherapy* 2012, **67**(10):2307-2308. doi:10.1093/jac/dks247.

147. Meldrum C, Doyle MA, Tothill RW: **Next-generation sequencing for cancer diagnostics: a practical perspective**. *The Clinical Biochemist Reviews* 2011, **32**(4):177-195.

148. Pant S, Weiner R, Marton MJ: **Navigating the rapids: the development of regulated next-generation sequencing-based clinical trial assays and companion diagnostics**. *Frontiers in Oncology* 2014, **4**:78. doi:10.3389/fonc.2014.00078.

149. Tsiatis AC, Norris-Kirby A, Rich RG, Hafez MJ, Gocke CD, Eshleman JR, Murphy KM: **Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications**. *The Journal of Molecular Diagnostics* 2010, **12**(4):425-432. doi:10.2353/jmoldx.2010.090188.

150. Slonczewski J, Foster J. **Microbiology: an evolving science**. New York: WW Norton, 2009: 700-702.

151. Mayer, PL, Larkin JA, Hennessy JM: **Amebic encephalitis**. *Surgical Neurology International* 2011, **2**:50. doi:10.4103/2152-7806.80115.

152. Gibson R, Schmotzer C, Quiñones-Mateu M: **Next-generation sequencing to help monitor patients infected with HIV: ready for clinical use?** *Current Infectious Disease Reports* 2014, **16**(4):401. doi:10.1007/s11908-014-0401-5.

153. McElroy K, Thomas T, Luciani F: **Deep sequencing of evolving pathogen populations: applications, errors, and bioinformatic solutions**. *Microbial Informatics and Experimentation* 2014, **4**:1. doi:10.1186/2042-5783-4-1.

154. Harris SR, Cartwright EJP, Török M, Holden MTG, Brown NM, Ogilvy-Stuart AL, Ellington MJ, Quail MA, Bentley SD, Parkhill J *et al*: **Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study**. *The Lancet Infectious Diseases* 2013, **13**(2):130-136. doi:10.1016/S1473-3099(12)70268-2.

155. Lewis T, Loman NJ, Bingle L, Jumaa P, Weinstock GM, Mortiboy D, Pallen MJ: **High-throughput whole-genome sequencing to dissect the epidemiology of Acinetobacter baumannii isolates from a hospital outbreak**. *The Journal of Hospital Infection* 2010, **75**(1):37-41. doi:10.1016/j.jhin.2010.01.012.

156. Price J, Gordon CN, Crook D, Llewelyn M, Paul J: **The usefulness of whole genome sequencing in the management of Staphylococcus aureus infections**. *Clinical Microbiology and Infection* 2013, **19**(9):784-789. doi:10.1111/1469-0691.12109.

157. Lienau EK, Strain E, Wang C, Zheng J, Ottesen AR, Keys CE, Hammack TS, Musser SM, Brown EW, Allard MW *et al*: **Identification of a salmonellosis outbreak by means of molecular sequencing**. *New England Journal of Medicine* 2011, **364**(10):981-982. doi:10.1056/NEJMc1100443.

158. Allard MW, Luo Y, Strain E, Li C, Keys CE, Son I, Stones R, Musser SM, Brown EW: **High resolution clustering of Salmonella enterica serovar Montevideo strains using a next-generation sequencing approach**. *BMC Genomics* 2012, **13**:32. doi:10.1186/1471-2164-13-32.

159. Le VTM, Diep BA: **Selected insights from application of whole genome sequencing for outbreak investigations**. *Current Opinion in Critical Care* 2013, **19**(5):432-439. doi:10.1097/MCC.0b013e3283636b8c.

160. Reuter S, Ellington MJ, Cartwright EJP, Köser CU, Török M, Gouliouris T, Harris SR, Brown NM, Holden MTG, Quail M *et al*: **Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology**. *JAMA Internal Medicine* 2013, **173**(15):1397-1404. doi:10.1001/jamainternmed.2013.7734.

161. Worby CJ, Lipsitch M, Hanage WP: **Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data**. *PLoS Computational Biology* 2014, **10**(3):e1003549. doi:10.1371/journal.pcbi.1003549.

162. Allard MW, Luo Y, Strain E, Pettengill J, Timme R, Wang C, Li C, Keys CE, Zheng J, Stones R *et al*: **On the evolutionary history, population genetics and diversity among isolates of Salmonella enteritidis PFGE pattern JEGX01.0004**. *PLoS One* 2013, **8**(1):e55254. doi:10.1371/journal.pone.0055254.

163. Long SW, Beres SB, Olsen RJ, Musser JM: **Absence of patient-to-patient intrahospital transmission of Staphylococcus aureus as determined by whole-genome sequencing**. *mBio* 2014, **5**(5):e01692-14. doi:10.1128/mBio.01692-14.

164. Grad YH, Lipsitch M: **Epidemiologic data and pathogen genome sequences: a powerful synergy for public health**. *Genome Biology* 2014, **15**(11):538. doi:10.1186/s13059-014-0538-4.

165. Salipante SJ, SenGupta DJ, Cummings LA, Land TA, Hoogestraat DR, Cookson BT: **Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology**. *Journal of Clinical Microbiology* 2015, **53**(4):1072-1079. doi:10.1128/JCM.03385-14.

166. Yozwiak NL, Schaffner SF, Sabeti PC: **Data sharing: make outbreak research open access.** *Nature* 2015, **518**(7540):477-479. doi:10.1038/518477a.

167. Schatz MC, Phillippy AM: **The rise of a digital immune system.** *GigaScience* 2012, **1**:4. doi:10.1186/2047-217X-1-4.

168. Afshinnekoo E, Meydan C, Chowdhury S, Jaroudi D, Boyer C, Bernstein N, Maritz JM, Reeves D, Gandara J, Chhangawala S *et al*: **Geospatial resolution of human and bacterial diversity with city-scale metagenomics**. *Cell Systems* 2015, **1**(1):72-87.

169. Check W: **NGS for determining the vaginal microbiome in clinical samples**. In: *CAP Today.* vol. June 2014; 2014.

170. Ubeda C, Taur Y, Jenq RR, Equinda MJ, Son T, Samstein M, Viale A, Socci ND, van den Brink MR, Kamboj M, Pamer EG: **Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans**. *The Journal of Clinical Investigation* 2010, **120**(12):4332-4341. doi:10.1172/JCI43918.

171. Strong MJ, Xu G, Morici L, Splinter Bon-Durant S, Baddoo M, Lin Z, Fewell C, Taylor CM, Flemington EK: **Microbial contamination in next generation sequencing: implications for sequence-based analysis of clinical samples**. *PLoS Pathogens* 2014, **10**(11):e1004437. doi:10.1371/journal.ppat.1004437.

172. Hutchison CA, 3rd: **DNA sequencing: bench to bedside and beyond**. *Nucleic Acids Research* 2007, **35**(18):6227-6237.

173. Vrijenhoek T, Kraaijeveld K, Elferink M, de Ligt J, Kranendonk E, Santen G, Nijman IJ, Butler D, Claes G, Costessi A *et al*: **Next-generation sequencing-based genome diagnostics across clinical genetics centers: implementation choices and their effects**. *European Journal of Human Genetics* 2015, **23**(9):1142–1150. doi:10.1038/ejhg.2014.279.

174. Hayden E: **Technology: The $1,000 genome.** *Nature* 2014, **507**(7492):294-295. doi:10.1038/507294a.

175. Petty TJ, Cordey S, Padioleau I, Docquier M, Turin L, Preynat-Seauve O, Zdobnov EM, Kaiser L: **Comprehensive human virus screening using high-throughput sequencing with a user-friendly representation of bioinformatics analysis: a pilot study**. *Journal of Clinical Microbiology* 2014, **52**(9):3351-3361. doi:10.1128/JCM.01389-14.

176. Burd EM: **Validation of laboratory-developed molecular assays for infectious diseases**. *Clinical Microbiology Reviews* 2010, **23**(3):550-576. doi:10.1128/CMR.00074-09.

177. Onsongo G, Erdmann J, Spears MD, Chilton J, Beckman KB, Hauge A, Yohe S, Schomaker M, Bower M, Silverstein KA *et al*: **Implementation of cloud based next generation sequencing data analysis in a clinical laboratory**. *BMC Research Notes* 2014, **7**:314. doi:10.1186/1756-0500-7-314.

178. **CLIA Overview.** [https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/Downloads/LDT-and-CLIA_FAQs.pdf]. Accessed August 14, 2015.

179. **Microbial Genomic Reference Materials.** [http://www.fda.gov/downloads/MedicalDevices/NewsEvents/WorkshopsConferences/UCM393056.pdf]. Accessed August 14, 2015.

180. Olson ND, Lund SP, Zook JM, Rojas-Cornejo F, Beck B, Foy C, Huggett J, Whale AS, Sui Z, Baoutina A *et al*: **International interlaboratory study comparing single organism 16S rRNA gene sequencing data: beyond consensus sequence comparisons**. *Biomolecular Detection and Quantification* 2015, **3**:17-24.

181. Gargis AS, Kalman L, Bick DP, da Silva C, Dimmock DP, Funke BH, Gowrisankar S, Hegde MR, Kulkarni S5, Mason CE *et al*: **Good laboratory practice for clinical next-generation sequencing informatics pipelines**. *Nature Biotechnology* 2015, **33**(7):689-693. doi:10.1038/nbt.3237.

182. **Global Microbial Identifier.** [http://www.globalmicrobialidentifier.org/Workgroups/About-the-GMI-Proficiency-Test-2015]. Accessed August 5, 2015.

183. Moorthie S, Hall A, Wright CF: **Informatics and clinical genome sequencing: opening the black box**. *Genetics in Medicine* 2013, **15**(3):165-171. doi:10.1038/gim.2012.116.

184. Pabinger S, Dander A, Fischer M, Snajder R, Sperk M, Efremova M, Krabichler B, Speicher MR, Zschocke J, Trajanoski Z: **A survey of tools for variant analysis of next-generation genome sequencing data**. *Briefings in Bioinformatics* 2013, **15**(2):256-278. doi:10.1093/bib/bbs086.

185. Fricke WF, Rasko DA: **Bacterial genome sequencing in the clinic: bioinformatic challenges and solutions**. *Nature Reviews Genetics* 2014, **15**(1):49-55. doi:10.1038/nrg3624.

186. **BaseSpace Cloud and OnSite Bioinformatics.** [http://www.illumina.com/informatics/research/sequencing-data-analysis-management/basespace.html]. Accessed June 28, 2015.

187. **Torrent Suite™ Software and Ion Torrent Storage™ Devices.** [http://www.thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-data-analysis-workflow/ion-torrent-suite-software.html?CID=WE18594]. Accessed June 28, 2015.

188. Wyres KL, Conway TC, Garg S, Queiroz C, Reumann M, Holt K, Rusu LI: **WGS analysis and interpretation in clinical and public health microbiology laboratories: what are the requirements and how do existing tools compare?** *Pathogens* 2014, **3**(2):437-458. doi:10.3390/pathogens3020437.

189. Nagarajan N, Pop M. **Sequence assembly demystified**. *Nature Reviews Genetics* 2013, **14**(3):157-167. doi:10.1038/nrg3367.

190. Magoc T, Pabinger S, Canzar S, Liu X, Su Q, Puiu D, Tallon LJ, Salzberg SL: **GAGE-B: an evaluation of genome assemblers for bacterial organisms**. *Bioinformatics* 2013, **29**(14):1718-1725. doi:10.1093/bioinformatics/btt273.

191. Chaisson MJ, Pevzner PA: **Short read fragment assembly of bacterial genomes**. *Genome Research* 2008, **18**(2):324-330.

192. Zerbino DR, Birney E: **Velvet: algorithms for de novo short read assembly using de Bruijn graphs**. *Genome Research* 2008, **18**(5):821-829. doi:10.1101/gr.074492.107.

193. Edwards D, Holt K: **Beginner's guide to comparative bacterial genome analysis using next-generation sequence data**. *Microbial Informatics and Experimentation* 2013, **3**(1):2. doi:10.1186/2042-5783-3-2.

194. Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, Sokolov A, Graim K, Funk C, Verspoor K, Ben-Hur A, *et al*: **A large-scale evaluation of computational protein function prediction**. *Nature Methods* 2013, **10**(3):221-227. doi:10.1038/nmeth.2340.

195. Seemann T: **Prokka: rapid prokaryotic genome annotation**. *Bioinformatics* 2014, **30**(14):2068-2069. doi:10.1093/bioinformatics/btu153.

196. Stoesser N, Batty EM, Eyre DW, Morgan M, Wyllie DH, Del Ojo Elias C, Johnson JR, Walker AS, Peto TEA, Crook DW: **Predicting antimicrobial susceptibilities for Escherichia coli and Klebsiella pneumoniae isolates using whole genomic sequence data**. *The Journal of Antimicrobial Chemotherapy* 2013, **68**(10):2234-2244. doi:10.1093/jac/dkt180.

197. Judge K, Harris SR, Reuter S, Parkhill J, Peacock SJ: **Early insights into the potential of the Oxford Nanopore MinION for the detection of antimicrobial resistance genes**. *Journal of Antimicrobial Chemotherapy* 2015. doi:10.1093/jac/dkv206.

198. Gordon NC, Price JR, Cole K, Everitt R, Morgan M, Finney J, Kearns AM, Pichon B, Young B, Wilson DJ *et al*: **Prediction of Staphylococcus aureus antimicrobial resistance by whole-genome sequencing**. *Journal of Clinical Microbiology* 2014, **52**(4):1182-1191. doi:10.1128/JCM.03117-13.

199. Infectious Diseases Society of America (IDSA), Spellberg B, Blaser M, Guidos RJ, Boucher HW, Bradley JS, Eisenstein BI, Gerding D, Lynfield R, Reller LB: **Combating antimicrobial resistance: policy recommendations to save lives**. *Clinical Infectious Diseases* 2011, **52**(Suppl 5):S397-S428. doi:10.1093/cid/cir153.

200. Walker B, Barrett S, Polasky S, Galaz V, Folke C, Engström G, Ackerman F, Arrow K, Carpenter S, Chopra K *et al*: **Looming global-scale failures and missing institutions**. *Science* 2009, **325**(5946):1345-1346. doi:10.1126/science.1175325.

201. The White House: **National Strategy for Combating Antibiotic-Resistant Bacteria. September 2014**; 2015:1-33.

202. The White House: **National Action Plan for Combating Antibiotic-Resistant Bacteria. March 2015**; 2015:1-62.

203. The President's Council of Advisors on Science and Technology (PCAST): **Report to the President on Combating Antibiotic Resistance. September 2014**; 2014:1-65.

204. Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, Bradley P, Iqbal Z, Feuerriegel S, Niehaus KE, Wilson DJ *et al*: **Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study**. *The Lancet Infectious Diseases* 2015, doi:10.1016/S1473-3099(15)00062-6.

205. Venco F, Vaskin Y, Ceol A, Muller H: **SMITH: a LIMS for handling next-generation sequencing workflows**. *BMC Bioinformatics* 2014, **15**(Suppl 14):S3. doi:10.1186/1471-2105-15-S14-S3.

206. Richter BG, Sexton DP: **Managing and analyzing next-generation sequence data**. *PLoS Computational Biology* 2009, **5**(6):e1000369. doi:10.1371/journal.pcbi.1000369.

207. Deverka PA, Dreyfus JC: **Clinical integration of next generation sequencing: coverage and reimbursement challenges**. *The Journal of Law, Medicine & Ethics* 2014, **42**(Suppl 1):22-41. doi:10.1111/jlme.12160.

208. Collins FS, Hamburg MA: **First FDA authorization for next-generation sequencer**. *New England Journal of Medicine* 2013, **369**(25):2369-2371. doi:10.1056/NEJMp1314561.

209. **FDA allows marketing of four "next generation" gene sequencing devices.** [http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm375742.htm]. Accessed August 17, 2015.

210. **Establishment Registration & Device Listing. Ion PGM™.** [http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfRL/rl.cfm?lid=427645&lpcd=PFF]. Accessed August 17, 2015.

211. **Establishment Registration & Device Listing. Sentosa SQ301.** [http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfRL/rl.cfm?lid=430009&lpcd=PFF]. Accessed August 17, 2015.

212. Javitt GH, Carner KS: **Regulation of next generation sequencing**. *The Journal of Law, Medicine & Ethics* 2014, **42**(Suppl 1):9-21. doi:10.1111/jlme.12159.

213. **Evaluation of high throughput sequencing for screening of pathogens in immunocompromised patients with suspected infectious disease. (PATHOQUEST1).** [https://www.clinicaltrials.gov/ct2/show/NCT02007642]. Accessed August 25, 2015.

214. Goben A, Salo D: **Federal research: data requirements set to change**. *College & Research Libraries News* 2013, **74**(8):421-425.

215. National Institutes of Health: **NIH genomic data sharing policy (GDS Policy)**. 2014:1-9.

216. Yang IS, Ryu C, Cho KJ, Kim JK, Ong SH, Mitchell WP, Kim BS, Oh HB, Kim KH: **IDBD: infectious disease biomarker database**. *Nucleic Acids Research* 2008, **36**(Database issue):D455-D460.

217. Shurtleff AC, Whitehouse CA, Ward MD, Cazares LH, Bavari S: **Pre-symptomatic diagnosis and treatment of filovirus diseases**. *Frontiers in Microbiology* 2015, **6**:108. doi:10.3389/fmicb.2015.00108.

218. Wu Q, Guan G, Liu Z, Li Y, Luo J, Yin H: **RNA-Seq-based analysis of changes in Borrelia burgdorferi gene expression linked to pathogenicity**. *Parasites & Vectors* 2015, **8**:155. doi:10.1186/s13071-014-0623-2.

219. Brenu EW, Ashton KJ, Batovska J, Staines DR, Marshall-Gradisnik SM: **High-throughput sequencing of plasma microRNA in chronic fatigue syndrome/myalgic encephalomyelitis**. *PLoS One* 2014, **9**(9):e102783. doi:10.1371/journal.pone.0102783.

220. Ellyard JI, Jerjen R, Martin JL, Lee AYS, Field MA, Jiang SH, Cappello J, Naumann SK, Andrews TD, Scott HS *et al*: **Identification of a pathogenic variant in TREX1 in early-onset cerebral systemic lupus erythematosus by whole-exome sequencing**. *Arthritis & Rheumatology* 2014, **66**(12):3382-3386.

221. Mullis KB, Faloona FA: **Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction**. *Methods in Enzymology,* 1987 **155:**335-350.

222. Jones D: **Milestone 11 Chain reaction.** *Nature Milestones DNA Technologies* 2007. doi:10.1038/nrg2251.

223. Saiki R, Gelfand D, Stoffel S, Scharf S, Higuchi R, Horn G, Mullis K, Erlich H: **Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase**. *Science* 1988, **239**(4839):487-491.

224. **The History of PCR**. [https://www.lifetechnologies.com/us/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/spotlight-articles/history-pcr.html]. Accessed June 28, 2015.

225. Campos J: **The cutting edges of contemporary diagnostics**. *Cultures* 2015, **2**(2):34-41.

226. **American Gut Project.** [http://americangut.org/]. Accessed June 28, 2015.

227. **Personal Genome Project: Harvard.** [http://www.personalgenomes.org/]. Accessed June 28, 2015.

228. **goVIRAL.** [https://www.goviralstudy.com/]. Accessed August 11, 2015.

229. Woolston C: **Website recruits people to share health data for studies.** *Nature News* 2015, **520**(7545):9.

230. Stone PW, Glied SA, McNair PD, Matthes N, Cohen B, Landers TF, Larson EL: **CMS changes in reimbursement for HAIs: setting a research agenda**. *Medical Care* 2010, **48**(5):433-439. doi:10.1097/MLR.0b013e3181d5fb3f.

231. Rothberg JM, Leamon JH: **The development and impact of 454 sequencing**. *Nature Biotechnology* 2008, **26**(10):1117-1124. doi:10.1038/nbt1485.

232. Baker M: **Next-generation sequencing: adjusting to data overload**. *Nature Methods* 2010, **7**(7):495-499.

233. Stein LD: **The case for cloud computing in genome informatics**. *Genome Biology* 2010, **11**(5):207. doi:10.1186/gb-2010-11-5-207.

Photo Credits:
p. 11: Tashatuvango / Shutterstock.com; p. 23: attem / Shutterstock.com; p.. 33: anyaivanova / Shutterstock; p. 37: T Nikita G. Sidorov / Shutterstock.com / ; p. 40: nicolas / istock.com

On the cover: Normal flora with *Streptococcus pyogenes*

Image retrieved from American Society of Microbiology (ASM)
DNA double helix artwork retrieved from chs-bio-vanea.blogspot.com

American Academy of Microbiology
1752 N Street, NW
Washington, DC 20036

academy.asm.org