7

Bacterial Genomes

PIERRE-EDOUARD FOURNIER | DIDIER RAOULT

KEY CONCEPTS

- With an increased output and lower cost, genomic sequencing has a clear impact on clinical microbiology.
- More than 49 000 bacterial genomes are currently available, including those of all major human pathogens.
- Clinical applications of bacterial genomics include the design of PCR and genotyping assays, the detection of virulence and antibiotic resistance markers, the development of culture media, serology assays and vaccines.
- Real-time genomics may help decipher infectious disease outbreaks within a few days.
- The data analysis capacity and cost remain major drawbacks.

With more than 49 000 bacterial genome sequences currently available, including those from all significant human pathogens, genomics has a significant impact on clinical microbiology and infectious diseases by enabling the development of improved diagnostic, genotyping, taxonomic, antibiotic and virulence marker detection tools as well as development of new culture media or vaccines. This chapter summarizes the current achievements in bacterial genomics relevant to medical microbiology.

Introduction

The genomic sequencing era may be divided into two periods (Figure 7-1). In the first decade, from 1995, when the sequencing of the *Haemophilus influenzae* genome was performed[1] to 2005, sequencing relied on the classic Sanger method, was time- and money-consuming and was reserved to a limited number of sequencing centers worldwide. Fewer than 300 bacterial genomes were sequenced during this period (Figure 7-1). Since 2005, the development of new and high-throughput sequencing methods,[2] together with a steep decrease of the sequencers' and reagents' cost enabling many laboratories to develop their own sequencing projects, led to a striking increase in the number of sequenced genomes, approaching 6000 for the year 2013 alone.

The tremendous source of information provided by genome sequences revolutionized basic aspects of microbiology. In particular, genome sizes of bacteria range from 139 kb for *Candidatus* Tremblaya princeps to 14 782 kb for *Sorangium cellulosum* (http://genomesonline.org/).

Genome analysis, also known as genome mining or *in silico* analysis, currently constitutes an irreplaceable research tool for various aspects of microbiology. In particular, the availability of genomes from virtually all bacterial human pathogens has opened perspectives in the fields of diagnosis, epidemiology, pathophysiology and treatment.

A major advantage of genome sequences over phenotypic methods is that data can rapidly be shared among scientists worldwide by being deposited in online databases and thus are easily comparable among laboratories. The main three databases are the National Center for

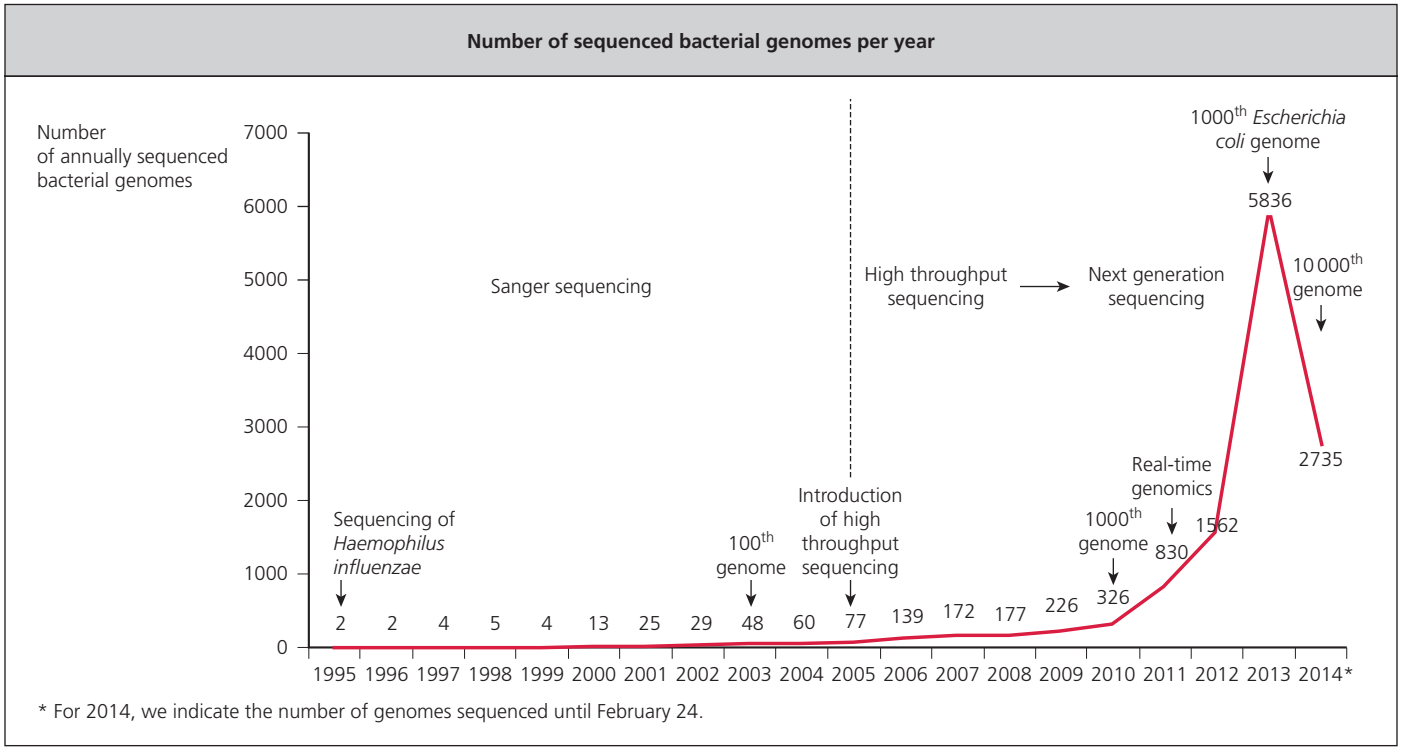

Figure 7-1 Number of sequenced bacterial genomes per year.

Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov/), the DNA Data Bank of Japan (DDBJ, www.ddbj.nig.ac.jp/) and the European Bioinformatics Institute (EBI, www.ebi.ac.uk/). In addition to offering complete microbial genome sequences with links to corresponding publications, these databases provide online tools for analyzing genome sequences. As of February 24, 2014, 12 272 genome sequences from 2897 bacterial species are available online (www.genomesonline.org/, https://gold.jgi.doe.gov). For some species, several genomes have been sequenced. For 31 species, more than 50 genomes are available, including 16 species for which more than 100 genomes have been sequenced, the species holding the record being *Escherichia coli*, with 1261 currently available genomes. Sequenced genomes include the most significant human bacterial pathogens, covering all the phylogenetic domains of bacteria. In addition, more than 27 000 sequencing projects are ongoing (www.genomesonline.org/). Moreover, new sequencing technologies are making possible the sequencing of random community DNA and single cells of bacteria without the need for cloning or cultivation.

There are multiple applications for genomics in clinical microbiology:

- Real-time genomics may be used to investigate infectious disease outbreaks.
- Bacterial genomes may be used as target sources for molecular detection, identification or genotyping.
- The gene content, obtained by comparison to databases such as Clusters of Orthologous Groups (www.ncbi.nlm.nih.gov/COG/) or Kyoto Encyclopedia of Genes and Genomes (www.genome.ad.jp/kegg/), may be searched for specific phenotypic traits such as virulence or antibiotic resistance markers, or deficient metabolic pathways enabling design of improved culture media.
- Antigenic epitopes detected in the deduced proteome may be used for serologic applications, development of monoclonal antibodies or development of vaccines (Figure 7-2).
- Taxonomic description of new bacterial species.

Sequencing Strategies

Since 2005, the introduction of high-throughput sequencing techniques has revolutionized genomics and metagenomics, and totally superseded conventional capillary sequencing (Sanger sequencing). Over recent years, the landscape of new high-throughput sequencing methods, termed 'next-generation sequencing (NGS)' has changed, some technologies being progressively abandoned, as is the case for the SOLID (Life Technologies), and others emerging, such as single cell sequencing (Table 7-1). Currently, two major NGS platforms are used, including the 454/Roche and Illumina/Solexa platforms (Table 7-1).

BENCHTOP SEQUENCERS

Following the introduction of ultrahigh throughput sequencers, Roche, Illumina/Solexa and Life Technologies introduced more rapid and affordable benchtop, or real-time high-throughput sequencers. These systems enable sequencing bacterial genomes within 1–2 days.[3] Benchtop sequencers were recently and efficiently used for whole genome sequencing (WGS). The GS Junior benchtop sequencer is also being extensively used in microbial WGS.[4]

SINGLE-CELL GENOME SEQUENCING

Thanks to the development of technologies that rely on sequencing single stretches of DNA molecules rather than using polymerase chain reaction (PCR) amplification of DNA fragments, it is now possible to access the genomic content of a unique cell. Three single-cell sequencers are currently available, including the Heliscope (Helicos Biosciences Corp), the Single Molecule Real Time (SMRT) sequencer (Pacific Biosciences) and the Oxford Nanopore DNA sequencer (Oxford Nanopore Technologies). These technologies, although different in their whole-genome amplification method,[5] are fast (~200–400 bases/sec.), require minimal sample preparation and produce long read length (up to 10 000 bp reads). In addition, such a strategy is especially valuable for micro-organisms that cannot be cultivated. Single-cell sequencing was successfully used for *E. coli* and *Bacillus subtilis*. It may also be of particular interest for complex floras, notably when combined with cell sorting methods like fluorescence-activated cell sorting (FACS) or microfluidic chips.[6] The sequence from the genome of a *Porphyromonas gingivalis* strain present in a hospital environmental sample was obtained by single-cell genomics, demonstrating the ability of this strategy to precisely analyze the genetic variations of micro-organisms at the strain level.

Analysis of Genome Sequences

DATA ANALYSIS

Despite its advantages, high-throughput sequencing is facing new challenges. The huge amounts of produced sequences generated a need for high performance computational hardware and expertise. In particular, the challenge of assembling genomes by combining millions of reads obtained by various sequencing methods, although theoretically complementary, is not, as yet, solved, in particular when read lengths are smaller than 100 bp. Another major issue of high throughput genomics is sequence annotation. Both the availability of automated and high throughput annotation tools and the quality of the annotation produced are also crucial. To date, there is no consensus on annotation method and standardization of genomic data. Such an effort will be necessary to facilitate future studies.

PANGENOMICS

The pangenome, made of a core genome (genes shared by all strains), a dispensable genome (genes shared by some, but not all, strains) and strain-specific genes, represents the complete gene pool of a species. The pangenome size of bacteria may differ greatly among species and

TABLE 7-1 Currently Available Sequencing Methods

	Sanger Di-deoxy Nucleotide Sequencing	Pyrosequencing	Sequencing by Synthesis	Ligation-Based Sequencing
Platform	Capillary sequencers	Roche (454) GS-FLX+	Illumina HiSeq 2500	Applied Biosystems SOLID
Mb/run	0.44	700	10^6	3000
Time/run	7 hours	23 hours	6 days	5 days
Read length (bp)	650–800	700	2x125	35
Limits	Cost	Limited paired-end and targeted sequencing, difficult sequence assembly, especially to disambiguate repeat regions		
	Need for high DNA quantity Cloning step	Low sensitivity in homopolymer sequencing Misincorporation of excess nucleotides Beads with mixed templates Redundancy		Specific sequence format

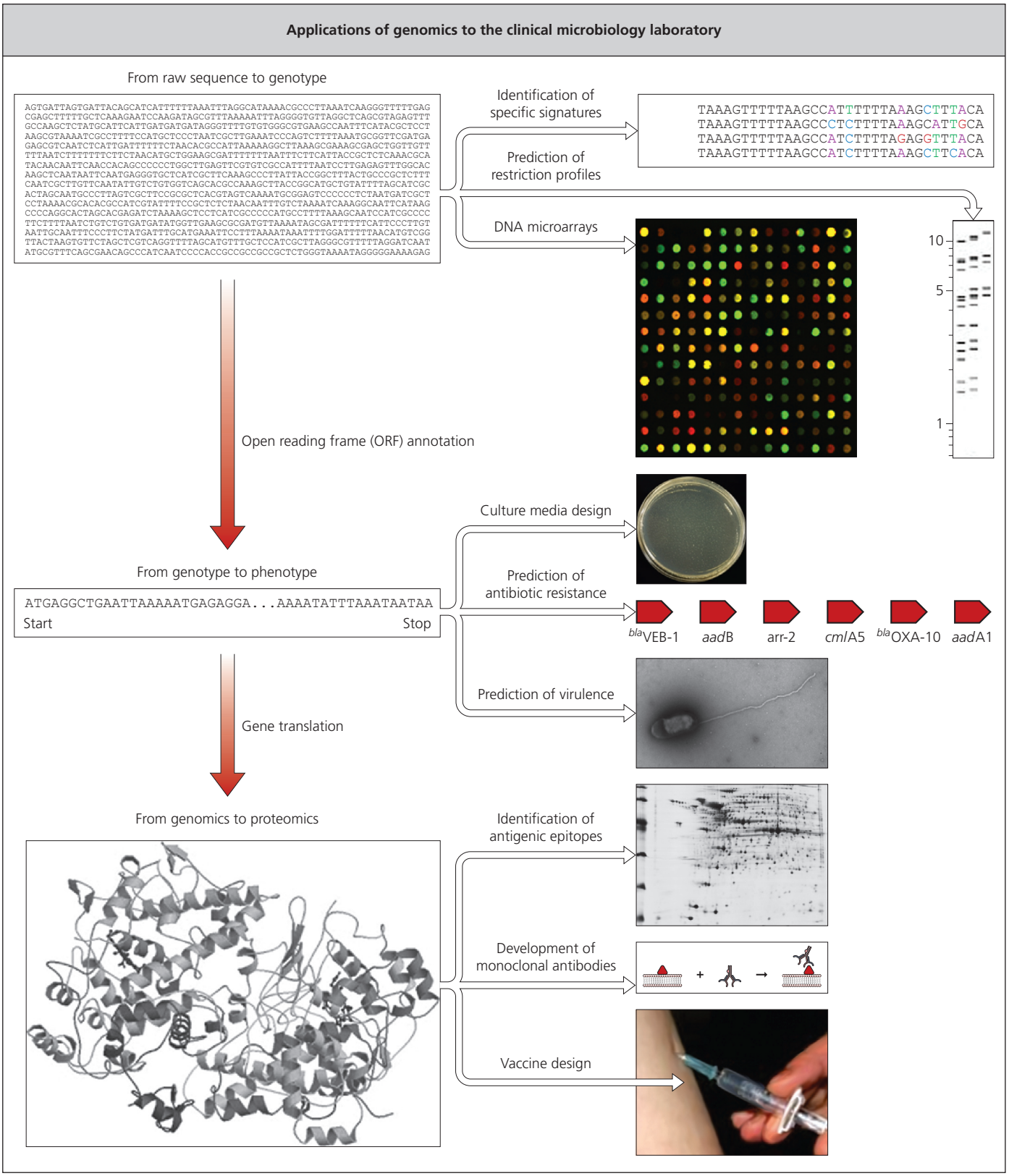

Figure 7-2 Applications of genomics to the clinical microbiology laboratory.

genera, in part because of the frequency of homologous recombination and lateral gene transfer. Therefore, one genome per species often underestimates the genetic complexity at the species level. For *E. coli*, *Streptococcus agalactiae* and *Strep. pyogenes* it was estimated that an average 441, 33 and 27 new genes, respectively, would be unveiled by every new sequenced genome. Such intraspecies variations might be linked to niche adaptation. Bacteria living in niches with limited access to the global microbial gene pool, such as *Bacillus anthracis*, *Chlamydia trachomatis* and *Mycobacterium tuberculosis*, have a much smaller intraspecies genomic diversity.[7] The most extreme example of genome homogeneity is *Buchnera aphidicola*, which experienced neither genome rearrangement nor gene duplication or transfer.

METAGENOMICS

Deciphering complex floras by identifying the DNAs from the various micro-organisms present (termed metagenomics, see Chapter 8) has provided a unique access to the microbial species and genes present in these environments. Similar to genome sequencing, metagenomics has benefited from NGS methods, which enable sequencing of random community DNA without the need for cloning or cultivation. By analogy, studies of all copies of 16S rDNA in a polymicrobial specimen have also been named metagenomic studies. Metagenomic studies have demonstrated that a great proportion of the biologic diversity in the bacterial world remains unexplored.[8] To date, the metagenomic approach has widely been used to study the human microbiome, in particular the oral, skin,[9] vaginal and lower intestinal floras,[8] the latter being by far the most studied. In the gastrointestinal tract, 395 distinct bacteria were demonstrated; 62% of these were not previously known and 80% were uncultivated. In addition, microbial genes outnumbered human genes 100 times, suggesting that the human microbiome may play a major role in human physiology. Unexpectedly, this study demonstrated great intersubject variability and significant differences between stool and mucosal community compositions. Metagenomic studies have also demonstrated that the composition of the human gut ecosystem is influenced by several factors, including age, geographic origin, environment, dietary habits, antibiotics and probiotics.[10] In addition, links between microbiota composition imbalance and various diseases such as irritable bowel syndrome, cord colitis syndrome,[11] Crohn's disease, necrotizing enterocolitis, polyposis or colorectal cancer, type II diabetes and obesity[12] were suggested. In the latter disease, both the gut microbiota richness and its composition might play a causative role.[13] Several human infections have also benefited from metagenomic studies, notably bacterial vaginosis, brain abscess,[14] cystic fibrosis, dental infections, diarrhea and pneumonia.

Using Genome Sequences

REAL-TIME GENOMICS

The development of NGS benchtop sequencers has made genome sequencing compatible with the routine clinical microbiology workflow. Such a strategy enables, within a few hours and for a few hundred dollars, exhaustive access to the genotype,[16] virulence markers and antibiotic resistance repertoire. Real-time genomics has notably been used to investigate hospital outbreaks of *Acinetobacter baumannii*,[17] *Clostridium difficile*, *E. coli* or *Staphylococcus aureus* infections, a large community-acquired outbreak of *E. coli* O104:H4 infections in Germany, the spread of multiresistant *Staph. aureus* in cystic fibrosis patients, and to identify the virulence determinants of a *Staph. epidermidis* strain that had caused a native valve endocarditis[18] or a chronic osteitis-causing *Clostridium tetani*. These findings demonstrated how rapid and precise sequencing could transform patient management or improve hospital infection control in routine clinical practice.

DESIGN OF MOLECULAR ASSAYS FOR DETECTION OF BACTERIAL PATHOGENS

By giving access to the complete genetic repertoire of a bacterium, genomics enabled a rational selection of PCR targets according to the desired objective (genus-, species-, subspecies- or strain-specific).[19] Genome sequences may also be used to design multiplex PCR assays enabling simultaneous detection and discrimination of various micro-organisms, as was the case for members of the *M. tuberculosis* complex.

In addition, the study of genomic sequences enabled the optimization of the sensitivity of detection, either by selecting a gene or fragment of noncoding DNA present as several copies in the genome or by designing nested PCR assays with a reduced risk of contamination, such as 'suicide PCR'.

MOLECULAR GENOTYPING

Molecular typing methods have largely superseded phenotypic methods. In a similar fashion as described above, genomic sequences may be a source of genotyping targets. Various genotyping methods, classified as nonsequence-based and sequence-based, have been designed by using genomic sequences.

- Nonsequence-based methods rely on the *in silico* design of macro-restriction profiles for rare cutter enzymes that may serve for restriction fragment length polymorphism-based assays, on genomic screening for single nucleotide polymorphisms or tandem repeats.[20] Such strategies were used for typing *Yersinia pestis*, *Salmonella typhimurium* or *Coxiella burnetii* isolates.
- Several sequence-based genotyping methods have been designed on the basis of genome sequences, including multilocus sequence typing (MLST) for the study of multidrug resistant *E. coli* or *Streptococcus pneumoniae*, and multispacer typing (MST) for typing *Y. pestis*, *Coxiella burnetii*, *Bartonella* or *Rickettsia* isolates.[21] However, bacterial whole-genome sequencing, by giving access to the whole genetic content of a strain, is the most discriminatory sequence-based genotyping method, as was demonstrated by the genome-based typing of *M. tuberculosis* or *Staph. aureus* clinical isolates.[22] However, due to its cost, even if it has decreased steeply since 1995, genome sequencing is not, as yet, a routine tool in most clinical microbiology laboratories.
- DNA microarrays are an intermediate between nonsequence-based and sequence-based methods, as they enable detection of subtle strain- or species-specific differences by simultaneous comparison of strains at the whole genome level. Such an application was used to discriminate among environmental *Legionella pneumophila* strains. Another advantage of microarrays is that their interpretation may be automated. DNA microarrays may also be used to detect and identify micro-organisms in complex floras.[20]

TAXONO-GENOMICS

The current strategy for classifying bacterial isolates is named polyphasic taxonomy. It combines phenotypic, chemotaxonomic and genotypic criteria. However, the currently used genotypic criteria, including DNA–DNA hybridization (DDH), 16S rRNA similarity and genomic G+C content, suffer from several drawbacks, notably a lack of inter-laboratory and inter-assay reproducibility for DDH, and inter-generic variations for 16S rRNA similarity and G+C content, which limit their reliability. By providing the total genetic background of a strain, WGS represents a new approach of taxonomy by taking into consideration the entire genomic content.[15] Various approaches to incorporate genomic data in taxonomy have been described, either based on partial or whole genomic data. An example of use of partial genomic information is the PhyloPhlAn method that uses sequences from 400 proteins. Among the taxonomic tools developed using whole genomic data, the determination of the average nucleotide identity (ANI) has been the most used. This parameter was proposed to replace DDH results. ANI was used to characterize new species of *Burkholderia*, *Geobacter* and *Vibrio* as well as the new genus *Sphaerochaeta* and the new class Dehalococcoidetes. Another slightly different method is the determination of the average genomic identity of orthologous gene sequences (AGIOS) that was used in a polyphasic strategy together with phenotypic criteria.[23,24] The latter method enabled the validation

of *Alistipes timonensis* sp. nov., *Anaerococcus senegalensis* sp. nov., *Brevibacillus massiliensis* sp. nov., *Brevibacterium senegalense* sp. nov., *Enterobacter massiliensis* sp. nov., *Herbaspirillum massiliense* sp. nov. and *Senegalimassilia anaerobia* gen. nov., sp. nov. However, currently, there is no consensus on the method that should be used to define a species.

Phenotype Prediction

DEVELOPMENT OF SPECIFIC CULTURE MEDIA

Although cultivation of the causative agent remains a priority during the diagnosis of bacterial infections, many fastidious species cannot be isolated in routine laboratory conditions. Thus, genomic sequencing constitutes a unique approach to identify incomplete metabolic pathways as well as the essential nutrients that a bacterium is unable to produce. It is then theoretically possible to compose specific media by incorporating metabolites that bacteria cannot produce. The first 'noncultivable' pathogenic bacterium whose genome analysis permitted axenic culture was *Tropheryma whipplei*, the causative agent of Whipple's disease.[25] A similar approach was used for *Coxiella burnetii*. However, the counter-example of *M. leprae* that cannot be grown axenically, despite the identification of many important lacking metabolic activities, highlighted the fact that the genomic identification of deficient metabolic pathways may not always provide all the clues to the growth of fastidious bacteria.

DETECTION OF RESISTANCE TO ANTIMICROBIALS

The rapid increase and spread of multidrug resistant (MDR) bacteria have become a major public health problem worldwide. Genome sequencing has the potential to identify the various genetic resistance determinants of a given bacterial strain, as was done for an MDR *Acinetobacter baumannii* that caused a nationwide outbreak of nosocomial infections in France.[26] Alternatively, genome sequencing may help understand the dynamics of resistance spread in a bacterial species, as was the case for *Staph. aureus*. However, genome sequencing has demonstrated that genetic resistance determinants, once thought to have mainly arisen under the pressure of antibiotics, may have emerged from ancient or environmental sources. In addition, genomic findings may enable the development of PCR assays specifically targeting resistance-causing genes or mutations that may serve as routine detection tools.

Another situation demonstrating the value of genome sequencing is during identification of resistance mechanisms in fastidious bacteria, for which phenotypic testing of antibiotic resistance is difficult or impossible, as in *T. whipplei* where the genome revealed the presence of mutated *gyrA* and *parC* genes, which explains resistance to fluoroquinolones. Genome sequencing may also allow the identification of the mechanism of action and target genes of new antimicrobial compounds.

IDENTIFICATION OF VIRULENCE FACTORS

Virulence genes are potential targets for risk assessment and intervention strategies. The identification of virulence genes may lead to the development of rapid screening tests in order to proceed with effective isolation measures in hospitalized patients or delay hospitalization after carriage decontamination. Comparative sequence analyses provide insight into pathogenic mechanisms of bacteria, allowing identification of known virulence proteins with conserved sequences or motifs, as well as putative new virulence proteins. However, recent studies have suggested that genome reduction, rather than gene gain, may also confer virulence, possibly by loss of regulation.[27]

In addition to allowing a better understanding of bacterial pathogenesis, identification of virulence factors in genomes may also allow the design of new potential antimicrobials in addition to allowing the identification of antimicrobial targets.

PROTEOME PREDICTION

Development of Serologic Tools

Genome analysis offers the possibility of identifying all putative protein-encoding genes of a given bacterium. This exhaustive approach may be completed by expression of the corresponding proteome, testing immunoreactive characteristics of selected proteins, and use of the best antigens for the development of serologic tools. This strategy allowed the identification of a representative panel of antigens for *Treponema pallidum* and *M. leprae*. Conversely, the genome may serve to identify antigens that have been detected within the proteome of a bacterium by immunoblotting and mass spectrometry. This strategy was used for *T. whipplei* for which 17 proteins specifically reacting with patients' antibodies were identified.

Vaccine Design

In a similar fashion to the development of serologic assays, a complete bacterial genome sequence offers the possibility of rational selection of vaccine candidates among its complete antigenic repertoire, notably those that are surface-exposed and conserved among strains.[28] This strategy, named reverse vaccinology, may be completed by functional immunomics for optimal epitope prediction and may result in DNA vaccines. It has been used successfully to identify potential vaccine targets for *B. anthracis*, *Brucella* species, *Chlamydia pneumoniae*, *Leptospira interrogans*, *M. tuberculosis*, *Neisseria meningitidis*, *Porphyromonas gingivalis*, *Rickettsia prowazekii*, *Strep. agalactiae*, *Strep. pneumoniae* and *Strep. pyogenes*. However, the major drawback of reverse vaccinology for vaccine development is that the strain under investigation does not represent the genetic diversity of its species. This risk has been highlighted by the comparison of genome sequences from several strains of *Strep. agalactiae*. In this species, the core genome is constituted by only approximately 80% of genes, with each new genome exhibiting ~18% new genes. This finding motivated the design of a 'universal anti-*S. agalactiae*' vaccine made up of four antigens, none of which was present in all strains but the combination of which was protective against all strains.[29]

This example, together with accumulating evidence that a single genomic sequence may not be sufficient to represent the variability of bacterial populations within a species, support the use of genome sequences from multiple strains of a species to identify an efficient vaccine formulation.

Conclusions and Perspectives

In 1995, the outcome of bacterial genome sequencing promised breakthroughs in microbiology and infectious disease research. Since then, almost 12 300 bacterial genomes have been sequenced. The multiplication of genome sequencing projects, together with extensive metagenomic studies, will provide a much more complete picture of the bacterial world. The potential of this inestimable source of information has already allowed scientists to reconsider the fields of bacterial virulence, host–bacteria interactions, microbiologic diagnosis and human microbial ecology. However, despite the many advances permitted by NGS, we are facing new challenges, including the need to develop improved assembly, annotation and analysis programs able to handle the huge amounts of sequence data produced by new sequencing technologies, the continuously growing number of genomic sequences, and the complexity of sequences in metagenomic studies. Whether single cell genomics will help solve some of these issues remains uncertain. Finally, there are not current minimum criteria for genome sequence quality, the majority being unfinished, or 'draft' or 'dirty' genomes,[30] and such standards may be useful in order to obtain reliable genome comparisons.

References available online at expertconsult.com.

KEY REFERENCES

Aagaard K., Petrosino J., Keitel W., et al.: The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. *FASEB J* 2013; 27(3):1012-1022.

Blainey P.C.: The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiol Rev* 2013; 37(3): 407-427.

Didelot X., Bowden R., Wilson D.J., et al.: Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet* 2012; 13(9):601-612.

Fleischmann R.D., Adams M.D., White O., et al.: Whole-genome random sequençing and assembly of *Haemophilus influenzae* Rd. *Science* 1995; 269:496-512.

Lagier J.C., Armougom F., Million M., et al.: Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect* 2012; 18(12):1185-1193.

Ley R.E., Turnbaugh P.J., Klein S., et al.: Microbial ecology: human gut microbes associated with obesity. *Nature* 2006; 444(7122):1022-1023.

Li W., Raoult D., Fournier P.E.: Bacterial strain typing in the genomic era. *FEMS Microbiol Rev* 2009; 33(5):892-916.

Ramasamy D., Mishra A.K., Lagier J.C., et al.: A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *Int J Syst Evol Microbiol* 2014; 64(Pt 2):384-391.

REFERENCES

1. Fleischmann R.D., Adams M.D., White O., et al.: Whole-genome random sequençing and assembly of *Haemophilus influenzae* Rd. *Science* 1995; 269:496-512.
2. Hall N.: Advanced sequencing technologies and their wider impact in microbiology. *J Exp Biol* 2007; 210:1518-1525.
3. Loman N.J., Misra R.V., Dallman T.J., et al.: Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 2012; 30(5):434-439.
4. Didelot X., Nell S., Yang I., et al.: Genomic evolution and transmission of *Helicobacter pylori* in two South African families. *Proc Natl Acad Sci USA* 2013; 110(34):13880-13885.
5. Blainey P.C.: The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiol Rev* 2013; 37(3):407-427.
6. Lasken R.S.: Genomic sequencing of uncultured microorganisms from single cells. *Nat Rev Microbiol* 2012; 10(9):631-640.
7. Mira A., Martin-Cuadrado A.B., D'Auria G., et al.: The bacterial pan-genome: a new paradigm in microbiology. *Int Microbiol* 2010; 13(2):45-57.
8. Aagaard K., Petrosino J., Keitel W., et al.: The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. *FASEB J* 2013; 27(3):1012-1022.
9. Gao Z., Tseng C.H., Pei Z., et al.: Molecular analysis of human forearm superficial skin bacterial biota. *Proc Natl Acad Sci USA* 2007; 104(8):2927-2932.
10. Lagier J.C., Armougom F., Million M., et al.: Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect* 2012; 18(12):1185-1193.
11. Bhatt A.S., Freeman S.S., Herrera A.F., et al.: Sequence-based discovery of *Bradyrhizobium enterica* in cord colitis syndrome. *N Engl J Med* 2013; 369(6):517-528.
12. Ley R.E., Turnbaugh P.J., Klein S., et al.: Microbial ecology: human gut microbes associated with obesity. *Nature* 2006; 444(7122):1022-1023.
13. Le Chatelier E., Nielsen T., Qin J., et al.: Richness of human gut microbiome correlates with metabolic markers. *Nature* 2013; 500(7464):541-546.
14. Al Masalma M., Armougom F., Scheld W.M., et al.: The expansion of the microbiological spectrum of brain abscesses with use of multiple 16S ribosomal DNA sequencing. *Clin Infect Dis* 2009; 48(9):1169-1178.
15. Didelot X., Bowden R., Wilson D.J., et al.: Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet* 2012; 13(9):601-612.
16. Beres S.B., Carroll R.K., Shea P.R., et al.: Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proc Natl Acad Sci USA* 2010; 107(9):4371-4376.
17. Lewis T., Loman N.J., Bingle L., et al.: High-throughput whole-genome sequencing to dissect the epidemiology of *Acinetobacter baumannii* isolates from a hospital outbreak. *J Hosp Infect* 2010; 75(1):37-41.
18. Fournier P.E., Gouriet F., Gimenez G., et al.: Deciphering genomic virulence traits of a *Staphylococcus epidermidis* strain causing native-valve endocarditis. *J Clin Microbiol* 2013; 51(5):1617-1621.
19. Li W., Raoult D., Fournier P.E.: Bacterial strain typing in the genomic era. *FEMS Microbiol Rev* 2009; 33(5): 892-916.
20. Fournier P.E., Drancourt M., Raoult D.: Bacterial genome sequencing and its use in infectious diseases. *Lancet Infect Dis* 2007; 7(11):711-723.
21. Drancourt M., Roux V., Dang L.V., et al.: Genotyping, Orientalis-like *Yersinia pestis*, and plague pandemics. *Emerg Infect Dis* 2004; 10:1585-1592.
22. Koser C.U., Holden M.T., Ellington M.J., et al.: Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med* 2012; 366(24):2267-2275.
23. Sentausa E., Fournier P.E.: Advantages and limitations of genomics in prokaryotic taxonomy. *Clin Microbiol Infect* 2013; 19(9):790-795.
24. Ramasamy D., Mishra A.K., Lagier J.C., et al.: A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *Int J Syst Evol Microbiol* 2014; 64(Pt 2):384-391.
25. Renesto P., Crapoulet N., Ogata H., et al.: Genome-based design of a cell-free culture medium for *Tropheryma whipplei*. *Lancet* 2003; 362(9382):447-449.
26. Fournier P.E., Vallenet D., Barbe V., et al.: Comparative genomics of multidrug resistance in *Acinetobacter baumannii*. *PLoS Genet* 2006; 2(1):e7.
27. Merhej V., Georgiades K., Raoult D.: Postgenomic analysis of bacterial pathogens repertoire reveals genome reduction rather than virulence factors. *Brief Funct Genomics* 2013; 12(4):291-304.
28. Seib K.L., Zhao X., Rappuoli R.: Developing vaccines in the era of genomics: a decade of reverse vaccinology. *Clin Microbiol Infect* 2012; 18(Suppl. 5):109-116.
29. Maione D., Margarit I., Rinaudo C.D., et al.: Identification of a universal Group B streptococcus vaccine by multiple genome screen. *Science* 2005; 309(5731): 148-150.
30. Fournier P.E., Raoult D.: Prospects for the future using genomics and proteomics in clinical microbiology. *Annu Rev Microbiol* 2011; 65:169-188.