CrossMark

A crash course in sequencing for a microbiologist

Aleksandra Kozińska[1] · Paulina Seweryn[2] · Izabela Sitkiewicz[1]

Abstract

For the last 40 years, "Sanger sequencing" allowed to unveil crucial secrets of life. However, this method of sequencing has been time-consuming, laborious and remains expensive even today. Human Genome Project was a huge impulse to improve sequencing technologies, and unprecedented financial and human effort prompted the development of cheaper high-throughput technologies and strategies called next-generation sequencing (NGS) or whole genome sequencing (WGS). This review will discuss applications of high-throughput methods to study bacteria in a much broader context than simply their genomes. The major goal of next-generation sequencing for a microbiologist is not really resolving another circular genomic sequence. NGS started its infancy from basic structural and functional genomics, to mature into the molecular taxonomy, phylogenetic and advanced comparative genomics. Today, the use of NGS expended capabilities of diagnostic microbiology and epidemiology. The use of RNA sequencing techniques allows studying in detail the complex regulatory processes in the bacterial cells. Finally, NGS is a key technique to study the organization of the bacterial life—from complex communities to single cells. The major challenge in understanding genomic and transcriptomic data lies today in combining it with other sources of global data such as proteome and metabolome, which hopefully will lead to the reconstruction of regulatory networks within bacterial cells that allow communicating with the environment (signalome and interactome) and virtual cell reconstruction.

Keywords Sanger sequencing · Next-generation sequencing · Structural genomics · Transcriptomics · Phylogenetic analysis · Microbiome

Introduction

A little over 40 years ago, Fred Sanger published his paper describing dideoxy chain terminator sequencing (Sanger et al. 1977). This first genome sequencing of bacteriophage φX174 was performed using long sequencing gels and radioactively labeled nucleotides. In addition, gels needed exposure to photographic film and manual sequence reading. Only later development of fluorescence-based methods and automated instrumentation allowed to increase speed and the scale of dideoxy chain terminator sequencing (Smith et al. 1986).

For over 30 years, "Sanger sequencing" allowed to unveil crucial secrets of life. The technology was used to sequence in

1981 first part of the human genome—mitochondrial DNA (Anderson et al. 1981)—and years later first complete bacterial genome of *Haemophilus influenzae* (Fleischmann et al. 1995) A decade between the mid-1990s and mid-2000s gave us genomic sequences of other bacterial species, and we saw the rise of comparative genomics and metagenomics (for a review, see Loman and Pallen (2015)). Those early years were crucial not only for gaining basic knowledge about bacterial genomes, but also to develop sequencing strategies and lay the foundation for future microbial genomics and transcriptomics. Sequencing of bacterial genomes also prompted the development of bioinformatics and software required to handle a large amount of data.

Although dideoxy chain terminator sequencing was absolutely critical in the development of modern microbiology, from molecular to environmental sciences, the method itself has been relatively troublesome. Sequencing was time-consuming and laborious, especially in the years before the introduction of cycle sequencing (Murray 1989), and remains expensive even today. Human Genome Project was a huge impulse to improve sequencing technologies, and unprecedented financial and human effort prompted the development of cheaper, less time-consuming, and high-throughput technologies and strategies.

Communicated by: Agnieszka Szalewska-Palasz

✉ Izabela Sitkiewicz
  i.sitkiewicz@nil.gov.pl

1   Department of Drug Biotechnology and Bioinformatics, National Medicines Institute, Chelmska 30/34, 00-725 Warszawa, Poland

2   Department of Microbiology and Antibiotics, National Medicines Institute, Chelmska 30/34, 00-725 Warszawa, Poland

This review will discuss some applications of high-throughput methods to study bacteria in a much broader context than simply their genomes. It could be even stated that nowadays the major goal of next-generation sequencing is not really resolving another circular genomic sequence; NGS is simply another tool that can be used to characterize bacteria in a very broad context. We can use NGS in detailed investigations and large-scale epidemiological studies. However, the major influence of affordable high-throughput methods on modern microbiology is a change of the way we think about the experimental approach (Fig. 1). The way we test hypotheses, plan experiments, and make conclusions, have changed. Instead of simple reasoning that leads from single observations of biological phenomena to generating a model of activity (Fig. 1a), multiple high-throughput methods, including genomic studies, combined with data mining, allow drawing a broad picture (Fig. 1b).

Technologies

Because a number of excellent comparisons of next-generation sequencing technologies are available (Buermans and den Dunnen 2014; Goodwin et al. 2016; Heather and Chain 2016; Levy and Myers 2016; Metzker 2010), this review will rather concentrate on applications of those sequencing methods in microbiology, than the description of technologies and economic aspects of such experiments. In general, over the last decade, we could observe the constant development of methods and changes in sequencing chemistry that produced simplified, more efficient instruments that allow to cheaply generate a tremendous amount of data. Costs of sequencing dropped dramatically with the development of new sequencing technologies. Within the last 10 years, a cost to determine a sequence of a human exome was lowered over 15,000 times, from about 15 million USD to below 1000 dollars (https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/). It makes the technology usable in multiple applications that require high-volume output. As Michael L. Metzker wrote in his review of sequencing technologies "… the potential of NGS is akin to the early days of PCR, with one's imagination being the primary limitation to its use…" (Metzker 2010). And we would like it to be a take-home message for all the readers of this review. It is not the technology we should focus on, but we should rather expand our research and gather more complex data that can be used to answer multiple questions at the same time. It is only our imagination that can lead us to scientific discoveries.

First attempts to modify and make sequencing technologies more flexible started in the late 1980s (Hultman et al. 1989) and continued through the whole decade of 1990s (Nyren et al. 1993). In the early 1990s, multiple

scientists from large academic hubs such as Cambridge or Harvard University started developing new sequencing methods based on different principles than Sanger sequencing. The early technologies included polony sequencing (Shendure et al. 2005), pyrosequencing (Margulies et al. 2005), and early single molecule sequencing (Helicos BioSciences). This whole scientific movement promoted the development of new technologies and, as a consequence, made the new technologies available on the market. In recent years, more sequencing technologies such as single molecule real-time sequencing or nanopore sequencing have been developed (Table 1) and are currently on the market or pre-market stage. There are new platforms from "old" players like Qiagen (GeneReader), Roche (Roche Genia), or Illumina (NovaSeq), but also completely new technologies are being developed. New sequencing technologies utilize multiple approaches like quantum (http://quantumbiosystems.com) or microdroplet/microfluidic sequencing (http://www.base4.co.uk).

High-throughput methods for a microbiologist

The infancy and childhood—structural and functional genomics

Early microbial sequencing projects tried to simply answer the questions about the nucleotide sequence and attempted to elucidate gene function, predominantly based on homology to known genes. Hence, the first attempts can be rather simply described as structural and early functional genomic projects. The biggest problem, however, for microbial and structural genomics since those early attempts is genome annotation. Often, open reading frames were misidentified, not much has been also known about basic transcription and translation in non-model microorganisms. Only recent use of NGS in transcriptomic approaches (see below) allows to answer those early questions about the basic biology of microorganisms.

Annotation and assigning function based only on homology is often misleading as genes of the similar sequence can have completely different functions in various microorganisms. It is especially true for regulatory genes; their specificity and biological function can be rarely predicted based on the sequence alone. Therefore, proper annotation of at least model organisms should include functional analysis. So far, the best microbial annotation is available for *Escherichia coli* K12 MG1655 in EcoCyc (https://ecocyc.org/) database. It includes over 85% of genes with experimentally confirmed function (Karp et al. 2007; Keseler et al. 2017). EcoCyc also contains a reconstruction of metabolic networks, cellular trafficking, and regulatory processes. Decades of research on *E. coli* generated combined knowledge that can be further utilized from modeling and network prediction using

Fig. 1 Differences of experimental approach between classical (**a**) and experiments involving high-throughput analyses such as NGS (**b**)

bioinformatic tools (Feist et al. 2009). Because multiple bacterial species are not as well studied as *E. coli*, they usually contain a large percentage of genes of unknown function encoded by their genome. Their annotation is predominantly based on gene ontology, but also advanced algorithms to reconstruct metabolic pathways and assign gene function based

Table 1 Comparison of sequencing technologies available currently on the market, based on Liu et al. (2012), Quail et al. (2012), Pevsner (2015), Levy and Myers (2016)) and information from system manufacturers (https://www.thermofisher.com, https://nanoporetech.com, http://sequencing.roche.com, https://sequencing.roche.com, https://www.illumina.com/, https://www.pacb.com/)

Method	Instrument manufacturer	Read length	Accuracy (single read not consensus)	Reads per run
Chain termination (Sanger sequencing)	ThermoFisher	400 to 900 bp	99.9%	N/A
Ion semiconductor (ion torrent sequencing)	ThermoFisher	Up to 600 bp	98%	Up to 80 million
Nanopore sequencing	Oxford Nanopore	Depends on the library preparation, not the device	~92–97%	Dependent on read length selected by user
Pyrosequencing	Roche	700 bp	99.9%	1 Million
Sequencing by synthesis	Illumina	50–500 bp depending on the instrument	99.9%	1 Million–3 billion depending on the instrument
Sequencing by ligation (SOLiD sequencing)	ThermoFisher	50 + 35 or 50 + 50 bp	99.9%	1.2 to 1.4 Billion
Single-molecule real-time sequencing	Pacific Biosciences	On average 14,000 bp maximum read length > 40,000 bases	87% Single-read accuracy	50,000 per SMRT cell, or 500–1000 megabases

Method	Time per run (not including library preparation)	Cost per 1 million bases (in US$)	Advantages	Disadvantages
Chain termination (Sanger sequencing)	20 min to 3 h	$2400	Useful for many smaller applications	Very expensive
Ion semiconductor (ion torrent sequencing)	2 h	$1	Less expensive NGS equipment. Fast	Homopolymer errors
Nanopore sequencing	Data streamed in real time. No fixed running time	$500–999 per flow cell	Long individual reads. Portable (palm-sized)	Lower throughput than other machines, low single read accuracy
Pyrosequencing	24 h	$10	Long reads. Fast	Runs are expensive. Homopolymer errors
Sequencing by synthesis	1 to 11 days, depending upon sequencer and specified read length	$0.05 to $0.15	High output. Low costs of sequencing	Expensive equipment
Sequencing by ligation (SOLiD sequencing)	1 to 2 weeks	$0.13	Low cost per base	Slower than other methods. Problems with palindromic sequences
Single-molecule real-time sequencing	30 min to 4 h	$0.13–$0.60	Fast. Long reads	Moderate throughput. Equipment is very expensive

on functional reconstruction are used. Examples of such tools that allow predicting gene function with high probability are RAST (Aziz et al. 2008) or Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al. 2016). In the opinion of many microbiologists that analyze genomic data, all the early sequencing projects require extensive re-working, corrections, and additional functional information.

Teenage years and early adulthood—taxonomic and phylogenetic studies beyond comparative genomics

Those early sequencing projects quickly expanded into comparative genomic field as the number of sequenced genomes grew. Initially, as a consequence of a lack of a large number of fully sequenced genomes, only comparisons between species were performed (for a review, see Loman and Pallen (2015)). Later with expanding the number of genomic sequences, first intra-species comparisons were performed, such as comparison of two *Helicobacter pylori* strains (Alm et al. 1999). Further accumulation of sequence data reflecting the bacterial diversity within the species allowed to define genomic elements that shape the bacterial community and changed the perspective on the definition of bacterial species (Gupta 2016; Caro-Quintero and Konstantinidis 2012; Gao and Gupta 2012) (Fig. 2).

The decreasing cost of sequencing inevitably changed microbial taxonomy and allowed to sequence and compare thousands of genomic sequences available for single bacterial species (Nasser et al. 2014; Long et al. 2017).

Since the genome itself is no longer a sole goal and costs of NGS are much lower, multiple techniques requiring sequencing are increasingly switching to NGS. A big field that utilizes sequencing techniques is microbial diagnostics. For years, typing techniques to determine unique strain characteristics included restriction fragment analysis (RFLP-PFGE), analysis of polymorphic sites such as MLVA, SNP detection by

Fig. 2 A schematic structure of a pan-genome that includes genes shared by all strains/isolates, genes shared by all genomes (core genome), and strain-specific genes that are present only in individual

sequencing, high-resolution melting curves, detection of virulence, and antibiotic resistance genes (for a review of the typing methods, see Li et al. (2009)). Even a few years ago, NGS was too expensive to replace those methods for routine diagnostics in the microbiology laboratory. Today, many larger laboratories switch to NGS and other molecular diagnostic methods to increase the speed of proper diagnosis (Opota et al. 2015). Many companies produce user-friendly software to speed up typing and increase the sensitivity of detection dedicated to microbiology research (http://www.ridom.de/seqsphere/). It allows to detect all polymorphisms, structural changes, and mobile genetic elements and reconstruct outputs of traditional typing methods such as RFLP-PFGE solely based on the genomic sequence. Such a detailed analysis can be used for advanced correlation studies such as GWAS (genome-wide association studies) to detect a connection between polymorphisms and disease manifestation or severity (Chen and Shapiro 2015). Diagnostic microbiology also utilizes amplicon-based profiling that allows to sequence selected amplicons such as regions encoding 16S rRNA that are used for species identification. Kits and protocols for 16S sequencing are commercially available and easy to apply. It also allows to perform virtual equivalents of other PCR typing methods such as MLVA-MLVF, detect certain virulence traits, or analyze highly polymorphic loci using user-designed amplicons.

Diagnostic laboratories routinely assay for antibiotic resistance of bacterial strains isolated from patients. The major disadvantage of standard procedure to determine minimal inhibitory concentration is the length of the procedure. It usually requires pure culture that can take over 24 h and growth test that takes from 16 to 24 h. DNA-based methods to detect determinants of antibiotic resistance are much faster, but they do not give much information about MIC value. Recently, NGS sequencing was employed for in silico prediction of MICs for selected antibiotics based on the genomic sequence (Nguyen et al. 2018). MICs can be predicted based on cell properties such as the presence of a certain combination of resistance determinants and membrane transporters. The ability to predict MICs based on genomic data completely changes the outlook on the modern diagnostics microbiology.

Another important application of NGS is tracing epidemics in real time. It can be used as a clinical tool to trace transfers of bacteria in the hospital settings as part of epidemiological investigations (Long et al. 2014). To date, confirmation that epidemic strains isolated from infected subjects are clones of the same strain required laborious investigation. For example, Pulse Net (https://www.cdc.gov/pulsenet/index.html)—a network dedicated to the detection of foodborne infections causing outbreaks, for years—have used RFLP-PFGE to establish a connection between strains. Recently, PulseNet introduced sequencing methods to establish clonal outbreaks. Improvement of portable instruments such as MinION
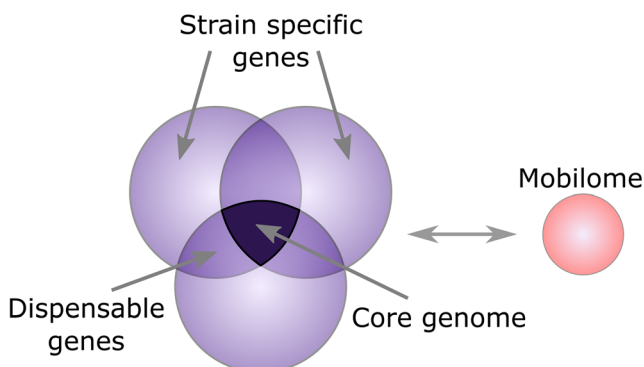
enables rapid in situ diagnostics and creates a stream of genomic data that can reveal critical epidemiological aspects of an outbreak or epidemics dynamics. By coupling sequencing to an enhanced surveillance and response platform, a more anticipatory approach to outbreak prevention and control can be applied (Gardy and Loman 2017).

Sequencing data are used by global networks such as the global microbial identifier (GMI) (http://www.globalmicrobialidentifier.org/). GMI is focused on the improvement of a global system of DNA genome databases for microbial and infectious disease identification and diagnostics. One of the goals of the network is early detection of outbreaks and tracking outbreaks in real time. Institutions involved in this project include Food and Drug Administration (FDA), World Health Organization (WHO), Food and Agriculture Organization (FAO), World Organization of Animal Health (OIE), and National Center for Biological Information (NCBI), public health agencies from multiple European countries, Thailand, Singapore, Japan, China.

Next level of comparative genomics utilizes large datasets generated using NGS allow to perform sophisticated comparative evolutionary analyses of multiple layered networks such as epistasis analysis detecting interactions and co-evolution of virulence or resistance traits in bacteria (Skwark et al. 2017). However, NGS technologies are increasingly more often utilized for various analyses of gene expression.

Maturity—transcriptomics and microbiome analysis

Transcriptomics started in the early 1980s from quantitative and qualitative studies on single transcripts using northern analysis and later making cDNA libraries and studying expressed sequence tags (EST). From the early on, researchers interested in studying the transcriptome focused rather on eukaryotic than bacterial genes. A massive effort was put into sequencing cDNA libraries and later development of techniques such as SAGE (serial analysis of gene expression) (Velculescu et al. 1995). The next big step in the transcriptome analysis was the development and improvement of quantitative PCR (see references in Kralik and Ricchi (2017)) and later microarrays (Davenport et al. 2017). Today, most of these techniques are currently being replaced by RNA sequencing technologies. Quantitative PCR, and microarrays are rather used for the detection of genes or mutations, rather than transcriptomic studies. Multiple transcriptomic studies in bacteria changed the outlook on the gene regulation in bacteria (see below) and prompted, in turn, massive changes in our understanding of basic biological processes in bacteria.

The improvement of RNA sequencing technologies expands the capabilities of the method. Today, sequencing requires small amount of initial material and allows to investigate the coordination of gene expression in vivo, during the

infection (Beres et al. 2016), or easily correlate in vivo and in vitro investigations to study the role of particular regulators in virulence (Eraso et al. 2016; Calfee et al. 2017). Extensive genome sequencing followed by transcriptomic studies allowed the identification of massive antisense transcription in bacteria, as well as the presence of multiple small regulatory RNAs that play important role in the regulation of virulence (Georg and Hess 2018; Figueroa-Bossi and Bossi 2018).

RNA sequencing allows to expand studies on basic bacterial biology such as transcription, translation, and the interaction of basic molecules in the living cells (Fig. 3) and link multiple levels of information such as genomic and transcriptomic data.

Another direction in microbiology possibly thanks to NGS is a complex analysis of microbial communities. It started as simply as attempts to quantitatively assess composition and proportions of species within various microbiomes. The preferred method for such analysis has been 16S amplicon sequencing as a relatively inexpensive way to establish community composition. Similar studies utilizing sequencing of certain regions such as ITS2 and/or 18S RNA were also conducted to establish human mycobiome (Nash et al. 2017).

In recent few years, with lower sequencing costs, the scale of microbiome sequencing could be increased to the whole metagenome instead of 16S amplicon sequencing. It allows for example to study the metabolic properties of the bacterial community or co-evolution of the host and bacteria (Davenport et al. 2017; Peisl et al. 2017). This approach allows to characterize unique bacterial communities and bacteria that are unable to grow in laboratory conditions, which often leads to discoveries of new active molecules such as antibiotics (Hover et al. 2018; Charlop-Powers et al. 2016; Ling et al. 2015). The availability of the whole genome sequences of the microbiome prompted also the rapid development of microbiome metatranscriptomics, especially for the microbiota and gut interactions (Lavelle and Sokol 2018; Bashiardes et al. 2016).

The opposite trend in sequencing is the increase of sequencing sensitivity that allows to sequence material isolated from single cells (Gawad et al. 2016). Single cell genomics allows understanding unique components of the complex microbial ecosystems. In addition, single-cell microorganism sequencing has enabled the genome assembly of new phyla and is beginning to provide new biological insights into the microbial world we had no idea exists. The ability to sequence DNA from single cells also allows studying transcriptome on the single-cell level.

Summary

Within the last 2–3 years, we have witnessed a rapid increase in quantity and quality in genomic and transcriptomic research that expands into other "-omics" fields. Technological
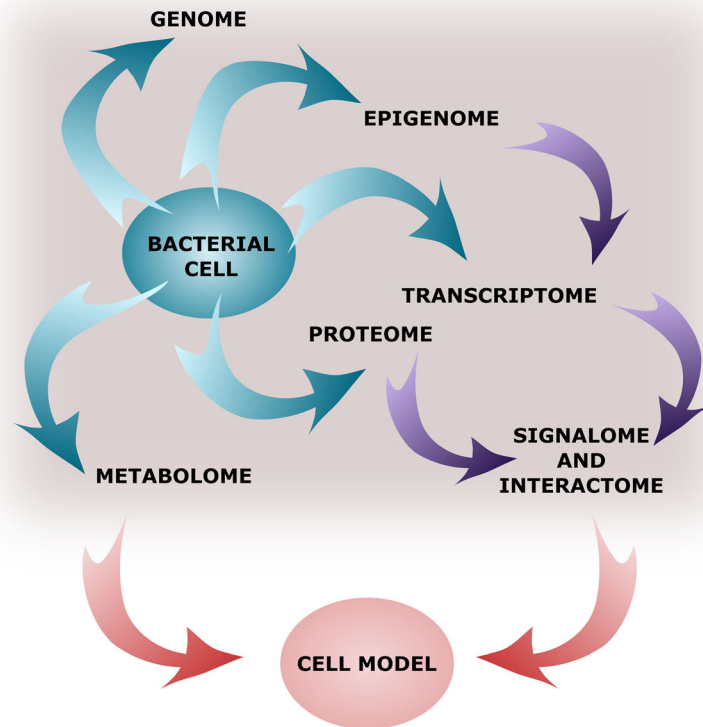
changes allowed the expansion of our knowledge and changed views on bacterial genetics and biology. Our views, especially on the mechanisms of gene expression in bacteria, changed. Even operon concept is no longer considered "simple" as a vast array of alternative regulatory mechanisms has been detected. Multiple "eukaryotic" regulatory mechanisms such as alternate transcripts, non-coding RNAs, overlapping UTRs, leaderless mRNA, riboswitches, antisense RNA, regulation by genome structure, or epigenetic modifications are being detected in bacteria (for a review, see Güell et al. (2011) and

Sánchez-Romero et al. (2015). The new discoveries and downstream studies are in the vast majority possibly thanks to the development of the new sequencing techniques.

The major challenge in understanding genomic and transcriptomic data lies today in combining it with other sources of global data such as proteome and metabolome, which hopefully will lead to the reconstruction of regulatory networks within bacterial cells that allow communicating with the environment (signalome and interactome) and virtual cell reconstruction (Fig. 3).

Fig. 4 Overview of the different steps involved in the use of NGS technologies for data gathering and utilization. After Angers-Loustau et al. (2018), modified

The NGS techniques involve multiple layers of sample collection, preparation, and analysis (Fig. 4). The use of the sequencing data goes beyond hardcore basic science, but is also a base for the translational research and can be used as a diagnostic tool, in treatment risk assessment or clinical interventions.

Compliance with ethical standards

References

Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, Smith DR et al (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen helicobacter pylori. Nature 397(6715):176–180. https://doi.org/10.1038/16495

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290(5806):457–465 http://www.ncbi.nlm.nih.gov/pubmed/7219534

Angers-Loustau A, Petrillo M, Bengtsson-Palme J, Berendonk T, Blais B, Chan K-G, Coque TM et al (2018) The challenges of designing a benchmark strategy for bioinformatics pipelines in the identification of antimicrobial resistance determinants using next generation sequencing technologies. F1000Res 7:459. https://doi.org/10.12688/f1000research.14509.1

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K et al (2008) The RAST server: rapid annotations using subsystems technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75

Bashiardes S, Zilberman-Schapira G, Elinav E (2016) Use of metatranscriptomics in microbiome research. Bioinf Biol Insights 10:BBI.S34610. https://doi.org/10.4137/BBI.S34610

Beres SB, Kachroo P, Nasser W, Olsen RJ, Zhu L, Flores AR, de la Riva I et al (2016) Transcriptome remodeling contributes to epidemic disease caused by the human pathogen Streptococcus pyogenes. MBio 7(3):1–14. https://doi.org/10.1128/mBio.00403-16

Buermans HPJ, den Dunnen JT (2014) Next generation sequencing technology: advances and applications. Biochim Biophys Acta Mol basis Dis 1842(10):1932–1941. https://doi.org/10.1016/j.bbadis.2014.06.015

Calfee G, Danger JL, Jain I, Miller EW, Sarkar P, Tjaden B, Kreikemeyer B, Sumby P (2017) Identification and characterization of serotype-specific variation in group a streptococcus pilus expression. Edited by Nancy E. Freitag. Infect Immun 86(2). https://doi.org/10.1128/IAI.00792-17

Caro-Quintero A, Konstantinidis KT (2012) Bacterial species may exist, metagenomics reveal. Environ Microbiol 14(2):347–355. https://doi.org/10.1111/j.1462-2920.2011.02668.x

Charlop-Powers Z, Pregitzer CC, Lemetre C, Ternei MA, Maniko J, Hover BM, Calle PY et al (2016) Urban park soil microbiomes are a rich reservoir of natural product biosynthetic diversity. Proc Natl Acad Sci 113(51):14811–14816. https://doi.org/10.1073/pnas.1615581113

Chen PE, Shapiro BJ (2015) The advent of genome-wide association studies for bacteria. Curr Opin Microbiol 25:17–24. https://doi.org/10.1016/j.mib.2015.03.002

Davenport ER, Sanders JG, Song SJ, Amato KR, Clark AG, Knight R (2017) The human microbiome in evolution. BMC Biol 15(1):127. https://doi.org/10.1186/s12915-017-0454-7

Eraso JM, Olsen RJ, Beres SB, Kachroo P, Porter AR, Nasser W, Bernard PE, Deleo FR, Musser M (2016) Genomic landscape of intrahost variation in group a streptococcus: repeated and abundant mutational inactivation of the FabT gene encoding a regulator of fatty acid synthesis. Infect Immun 84(12):3268–3281. https://doi.org/10.1128/IAI.00608-16

Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ (2009) Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol 7(2):129–143. https://doi.org/10.1038/nrmicro1949

Figueroa-Bossi N, Bossi L (2018) Sponges and predators in the small RNA world. Microbiol Spectr 6(4). https://doi.org/10.1128/microbiolspec.RWR-0021-2018

Fleischmann R, Adams M, White O, Clayton R, Kirkness E, Kerlavage A, Bult C et al (1995) Whole-genome random sequencing and assembly of Haemophilus Influenzae Rd. Science 269(5223):496–512. https://doi.org/10.1126/science.7542800

Gao B, Gupta RS (2012) Microbial systematics in the post-genomics era. Antonie Van Leeuwenhoek 101(1):45–54. https://doi.org/10.1007/s10482-011-9663-1

Gardy JL, Loman NJ (2017) Towards a genomics-informed, real-time, global pathogen surveillance system. Nat Rev Genet 19(1):9–20. https://doi.org/10.1038/nrg.2017.88

Gawad C, Koh W, Quake SR (2016) Single-cell genome sequencing: current state of the science. Nat Rev Genet 17(3):175–188. https://doi.org/10.1038/nrg.2015.16

Georg J, Hess WR (2018) Widespread antisense transcription in prokaryotes. Microbiol Spectr 6(4). https://doi.org/10.1128/microbiolspec.RWR-0029-2018

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17(6):333–351. https://doi.org/10.1038/nrg.2016.49

Güell M, Yus E, Lluch-Senar M, Serrano L (2011) Bacterial transcriptomics: what is beyond the RNA Horiz-Ome? Nat Rev Microbiol 9(9):658–669. https://doi.org/10.1038/nrmicro2620

Gupta RS (2016) Impact of genomics on the understanding of microbial evolution and classification: the importance of Darwin's views on classification. FEMS Microbiol Rev 40(4):520–553. https://doi.org/10.1093/femsre/fuw011

Heather JM, Chain B (2016) The sequence of sequencers: the history of sequencing DNA. Genomics 107(1):1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

Hover BM, Kim S-H, Katz M, Charlop-Powers Z, Owen JG, Ternei MA, Maniko J et al (2018) Culture-independent discovery of the Malacidins as calcium-dependent antibiotics with activity against multidrug-resistant gram-positive pathogens. Nat Microbiol 3(4): 415–422. https://doi.org/10.1038/s41564-018-0110-1

Hultman T, Ståhl S, Hornes E, Uhlén M (1989) Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as Solid support. Nucleic Acids Res 17(13):4937–4946 http://www.ncbi.nlm.nih.gov/pubmed/2668874

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44(D1):D457–D462. https://doi.org/10.1093/nar/gkv1070

Karp PD, Keseler IM, Shearer A, Latendresse M, Krummenacker M, Paley SM, Paulsen I et al (2007) Multidimensional annotation of the Escherichia Coli K-12 genome. Nucleic Acids Res 35(22): 7577–7590. https://doi.org/10.1093/nar/gkm740

Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, Fulcher C et al (2017) The EcoCyc database: reflecting new knowledge about Escherichia coli K-12. Nucleic Acids Res 45(D1):D543–D550. https://doi.org/10.1093/nar/gkw1003

Kralik P, Ricchi M (2017) A basic guide to real time PCR in microbial diagnostics: definitions, parameters, and everything. Front Microbiol 8. https://doi.org/10.3389/fmicb.2017.00108

Lavelle A, Sokol H (2018) Gut microbiota: beyond metagenomics, metatranscriptomics illuminates microbiome functionality in IBD. Nat Rev Gastroenterol Hepatol 15(4):193–194. https://doi.org/10.1038/nrgastro.2018.15

Levy SE, Myers RM (2016) Advancements in next-generation sequencing. Annu Rev Genomics Hum Genet 17(1):95–115. https://doi.org/10.1146/annurev-genom-083115-022413

Li W, Raoult D, Fournier P-E (2009) Bacterial strain typing in the genomic era. FEMS Microbiol Rev 33(5):892–916. https://doi.org/10.1111/j.1574-6976.2009.00182.x

Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, Mueller A et al (2015) A new antibiotic kills pathogens without detectable resistance. Nature 517(7535):455–459. https://doi.org/10.1038/nature14098

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M (2012) Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012:251364. https://doi.org/10.1155/2012/251364

Loman NJ, Pallen MJ (2015) Twenty years of bacterial genome sequencing. Nat Rev Microbiol 13(12):787–794. https://doi.org/10.1038/nrmicro3565

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

Long SW, Olsen RJ, Eagar TN, Beres SB, Zhao P, Davis JJ, Brettin T, Xia F, Musser JM (2017) Population genomic analysis of 1,777 extended-spectrum beta-lactamase-producing klebsiella pneumoniae isolates, Houston, Texas: unexpected abundance of clonal group 307. MBio 8(3). https://doi.org/10.1128/mBio.00489-17

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057):376–380. https://doi.org/10.1038/nature03959

Metzker ML (2010) Sequencing technologies—the next generation. Nat Rev Genet 11(1):31–46. https://doi.org/10.1038/nrg2626

Murray V (1989) Improved double-stranded DNA sequencing using the linear polymerase chain reaction. Nucleic Acids Res 17(21):8889 http://www.ncbi.nlm.nih.gov/pubmed/2587244

Nash AK, Auchtung TA, Wong MC, Smith DP, Gesell JR, Ross MC, Stewart CJ et al (2017) The gut mycobiome of the human microbiome project healthy cohort. Microbiome 5(1):153. https://doi.org/10.1186/s40168-017-0373-4

Nasser W, Beres SB, Olsen RJ, Dean MA, Rice KA, Wesley Long S, Kristinsson KG et al (2014) Evolutionary pathway to increased virulence and epidemic group a Streptococcus disease derived from 3, 615 genome sequences. Proc Natl Acad Sci U S A 111(17):E1768–E1776. https://doi.org/10.1073/pnas.1403138111

Nguyen M, Brettin T, Wesley Long S, Musser JM, Olsen RJ, Olson R, Shukla M et al (2018) Developing an in silico minimum inhibitory concentration panel test for Klebsiella pneumoniae. Sci Rep 8(1): 421. https://doi.org/10.1038/s41598-017-18972-w

Nyren P, Pettersson B, Uhlen M (1993) Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. Anal Biochem 208(1):171–175. https://doi.org/10.1006/abio.1993.1024

Opota O, Jaton K, Greub G (2015) Microbial diagnosis of bloodstream infection: towards molecular diagnosis directly from blood. Clin Microbiol Infect 21(4):323–331. https://doi.org/10.1016/j.cmi.2015.02.005.

Peisl BYL, Schymanski EL, Wilmes P (2017) Dark matter in host-microbiome metabolomics: tackling the unknowns—a review. Anal Chim Acta. https://doi.org/10.1016/j.aca.2017.12.034

Pevsner J (2015) Bioinformatics and functional genomics, 3rd edn. Wiley, USA

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

Sánchez-Romero MA, Cota I, Casadesús J (2015) DNA methylation in bacteria: from the methyl group to the methylome. Curr Opin Microbiol 25(June):9–16. https://doi.org/10.1016/j.mib.2015.03.004.

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes JC, Hutchison CA, Slocombe PM, Smith M (1977) Nucleotide sequence of bacteriophage ΦX174 DNA. Nature 265(5596):687–695. https://doi.org/10.1038/265687a0

Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. Science (New York, NY) 309(5741):1728–1732. https://doi.org/10.1126/science.1117389

Skwark MJ, Croucher NJ, Puranen S, Chewapreecha C, Pesonen M, Xu YY, Turner P et al (2017) Interacting networks of resistance, virulence and core machinery genes identified by genome-wide epistasis analysis. PLoS Genet 13(2):e1006508. https://doi.org/10.1371/journal.pgen.1006508

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE (1986) Fluorescence detection in automated DNA sequence analysis. Nature 321(6071):674–679. https://doi.org/10.1038/321674a0

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science (New York, NY) 270(5235):484–487 https://doi.org/10.1126/science.270.5235.484