



# Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality

# 15

Nidhi Gupta and Vijay K. Verma

## Abstract

Next-generation sequencing has the ability to revolutionize almost all fields of biological science. It has drastically reduced the cost of sequencing. This allows us to study the whole genome or part of the genome to understand how the cellular functions are governed by the genetic code. The data obtained in huge quantity from sequencing upon analysis gives an insight to understand the mechanism of pathogen biology, virulence, and phenomenon of bacterial resistance, which helps in investigating the outbreak. This ultimately helps in the development of therapies for public health welfare against human pathogen and diagnostic reagents for the screening. This chapter includes the basic of Sanger's method of DNA sequencing and next-generation sequencing, different available platforms for sequencing with their advantages, and limitations and their chemistry with an overview of downstream data analysis. Furthermore, the breadth of applications of high-throughput NGS technology for human health has been discussed.

## Keywords

Pyrosequencing · Roche 454 · ABISOLiD · Nonopore

---

N. Gupta

Department of Biochemistry, Central University of Rajasthan, Kishangarh, Rajasthan, India

V. K. Verma (✉)

Department of Microbiology, Central University of Rajasthan, Kishangarh, Rajasthan, India

e-mail: [vijayverma@curaj.ac.in](mailto:vijayverma@curaj.ac.in)

© Springer Nature Singapore Pte Ltd. 2019

P. K. Arora (ed.), *Microbial Technology for the Welfare of Society*,

Microorganisms for Sustainability 17,

[https://doi.org/10.1007/978-981-13-8844-6\\_15](https://doi.org/10.1007/978-981-13-8844-6_15)

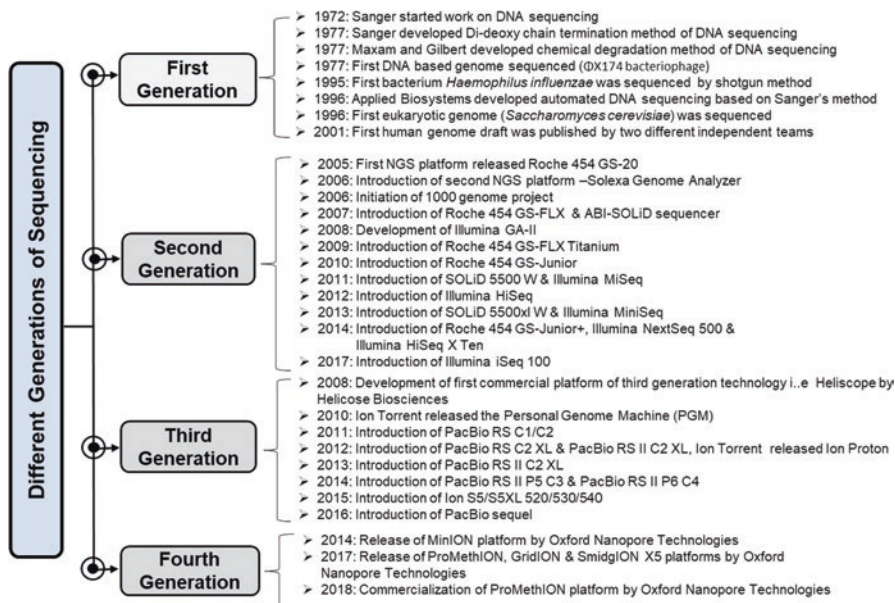
313

## 15.1 Introduction and History of Sequencing

Earlier, Sanger's method of DNA sequencing was the only easy and popular method to determine the sequence of DNA molecule. On top of everything, this dideoxy chain termination method was diligently used for more than three decades from its discovery. Since the start of the twenty-first century, high-throughput sequencing technology has made an impact on the genomics research because it enabled genome-wide sequencing and screening far easier, inexpensive, and reproducible with lesser need of manpower (Metzker 2010). Using NGS, the genome of the bacteria can be sequenced in a single run. Sequencing of the DNA allows us to provide the basic information, i.e., the sequence/order of the nucleotides. After recognizing that this could be a magic tool to understand the gene sequences and location of the gene regulatory molecules, a UK-based team led by Dr. Frederick Sanger initiated work on DNA sequencing in 1972. He finally developed the "dideoxy chain termination" method for DNA sequencing and also published his work in 1977 (Sanger et al. 1977a). The method was based upon the base-specific termination of the growing chain. DNA polymerase adds the nucleotides on the chain, but upon the incorporation of dideoxynucleotide, the chain terminates due to its inability of the phosphodiester bond formation between newly coming nucleotide and dideoxynucleotide. In the same year 1977, a US team led by Maxam and Gilbert also published a chemical-based sequencing method in which the sequencing of DNA was dependent on chemical cleavage protocol (Maxam and Gilbert 1977). This method was based upon the use of harmful chemicals such as dimethyl sulfate (DMS) and hydrazine and hence was not that popular among molecular biologists. In contrast, the Sanger's method was accepted widely because of its easy protocol and use of lesser harmful reagents (Obenrader 2003). However, both Fredrick Sanger and Walter Gilbert were awarded with the Nobel Prize in chemistry in 1980 for their DNA sequencing method discovery. Using Sanger's method, bacteriophage  $\Phi$ X174 genome of 5386 bp was sequenced, and it was the beginning of the full DNA genome sequencing (Sanger et al. 1977b).

Sanger's method was the method of choice and used for the genome sequencing of a number of organisms. *Haemophilus influenzae* was the first bacterium of genome size 1,830,140 bp sequenced using an approach, namely, shotgun, in 1995 (Fleischmann et al. 1995). Soon after this, *Saccharomyces cerevisiae* was the first eukaryotic genome of size 12,156,677 bp sequenced in 1996 (Goffeau et al. 1996). Major breakthrough happened when human genome (14.8 billion bp) was sequenced and got published in 2001. The human genome was sequenced using two different approaches by two different independent teams. Using shotgun approach (where the genome got fragmented randomly), Dr. Craig Venter's team from a company called Celera Genomics published its dataset in 2001 (Venter et al. 2001), whereas Francis Collins group from the National Human Genome Research Institute (NHGRI, NIH) used the BAC contigs (bactigs) approach for the mapping of sequence.

The major limitation of the Sanger's method was that it could sequence very less number of DNA sequences in one go and also the cost per base was very high. Apparently, the high rise in the complex diseases and their relation with the



**Fig. 15.1** Sequencing events, developments, and introduction of different generations' sequencing platforms in chronological order

mutations/changes in the genome demand a far-flung knowledge of the genome sequence. Thus, it is required to sequence the genome of huge number of individuals as well as other organisms for the diagnosis and treatment in short period of time with low cost. This prompted the need of high-throughput sequencing technologies, which can provide information at substantially lower cost. Further, the development of high-throughput next-generation sequencing technology has proven that it can generate enormous data (millions of sequences) cost effectively and rapidly.

Here we have shown a timeline of sequencing events and introduction of platforms of different generations' sequencing technologies (Fig. 15.1).

## 15.2 Different Generations of Sequencing

It is very prominent that for almost four decades, the Sanger's method was considered the gold standard because there was not a considerable development in DNA sequencing techniques in those years. In the last almost 15 years (from 2005), there are massive changes in the generation of sequencing technology. In short, changes in generation means the change in chemistry and platform. The widely accepted method of sequencing, i.e., Sanger's method of DNA sequencing, covered the first-generation sequencing along with Maxam-Gilbert method. Both methods were capable of sequencing about 1 kilobase length DNA fragment at one shot. In the process of analyzing longer sequences, researchers used the "shotgun technique"

where overlapping sequences were cloned and sequenced separately followed by assembly into contigs (Anderson 1981). Sanger's method is one of the best methods for years to come for the sequencing of gene cloned in heterologous system because of its precision, robustness, and ease of use.

Scientists presented a next-generation sequencing technology, which comes under second-generation sequencing technology that includes Roche 454, Illumina Solexa, and ABI-SOLiD and has transformed the field of omics. This technology was able to produce enormous amount of data at very economic cost and expeditiously. Moreover, this technology is very rapid than traditional method that whole genome of small organisms can be sequenced in a single day. However, in recent years, Illumina platforms have contributed very much to the second-generation sequencers and hence are considered to be one of the best platform providers.

The third-generation sequencing that is also known as next next-generation sequencing refers to those technologies which do not depend on the PCR amplification of DNA molecule. Thus, the problem related to biasness through PCR amplification and dephasing was ruled out. Platforms of this generation, which include Helicos and PacBio, are capable of sequencing single molecule.

There is a platform called Ion Torrent that has been kept between second and third generation because it is based upon first "post light sequencing" technology. This does not require fluorescence or luminescence. Nanopore sequencers offered by Oxford Nanopore Technologies (ONT), namely, GridION and MinION, lie under fourth generation. These platforms are based on different chemistry from third-generation sequencers.

Note that Roche 454 platform was commercially introduced first, but nowadays it is not available in the market. This shows that the changes made in this field in the past 12–15 years are very rapid.

---

### 15.3 Comparison of First-Generation and Second-Generation Sequencing Principle

First generation includes two separate methods, namely, Sanger and Maxam Gilbert method. Both methods were equally accepted in the beginning, but Sanger's method was extensively used for the routine sequencing purposes, which was based on the chain termination method. In the chain termination method, upon incorporation of the dideoxynucleotides (ddNTPs), the growing chain terminates. Fragments of varied length (length varied by single nucleotide) of the DNA molecule then run on the traditional slab gels and pattern of bands obtained for the sequence determination. Subsequently, radiolabeling is replaced with fluorescently labeled (automated method of sequencing) ddNTPs, and the laser light at different wavelengths does the sequence determination (Smith et al. 1986). This method can generate a maximum read length that ranges from 800 to 1000 bp. In this method, only one fragment can be sequenced in one capillary, which means the output of one run is the length of the sequenced fragment.

Conversely, the principle of second-generation sequencers or next-generation sequencers is based on clonal amplification of DNA molecule where billions of different DNA fragments get sequenced at the same time in parallel fashion and generate enormous data. For sequencing of the whole genome of an organism, random fragmentation is done at particular size range and then fragments ligated to the oligonucleotide adaptors, which are platform specific, followed by independent parallel sequencing. Parallel analysis increases the sequencing speed. NGS offers the capability to produce massive volume of data from a single run at a very low cost in a very short duration without the need of fragment cloning strategy, which is generally used in conventional method. As we can see, there is a vast difference in the cost per genome sequencing in 2001 (\$100M) as compared to the cost in 2017 (<\$1K) (data from NHGRI genome sequencing program). Complete draft sequence of the human genome with the help of automated Sanger's method was published in 2001, which was the outcome of 13 years of rigorous efforts of international project of \$2.7 billion. On the contrary, using NGS platform, the whole human genome can be sequenced in a week for a few thousand dollars (Gullapalli et al. 2012).

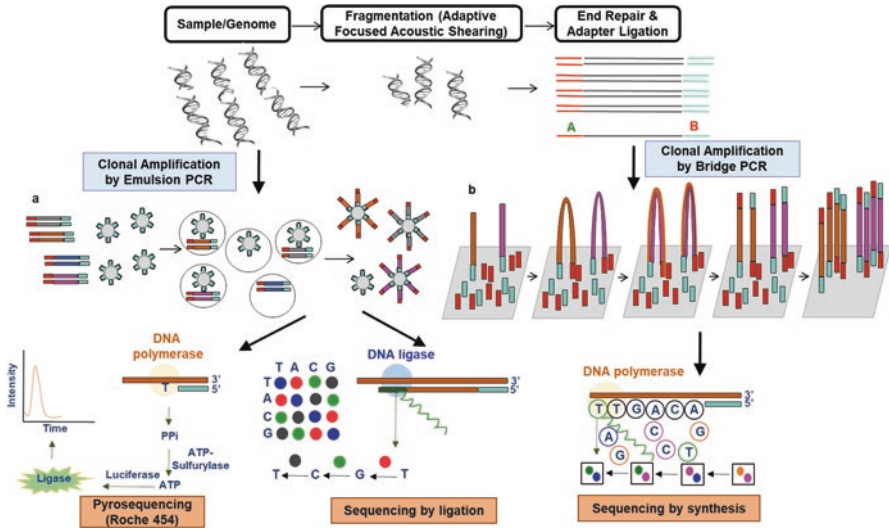
---

## 15.4 Second-Generation/Next-Generation Sequencing Technologies

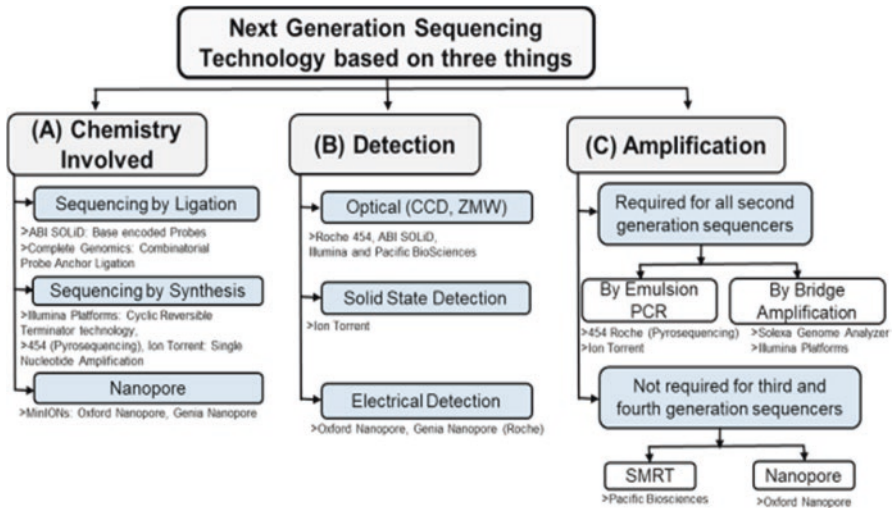
NGS and high-throughput sequencing generally denote to technologies that permit the millions of sequencing reactions in parallel on the same solid surface which may be beads or glass slide. This does not require the physical separation of reaction in different well or in lane/tube but spatially separated. Hence, thousands of million different reactions ensue simultaneously, because of which there is a dramatic decrease in the labor input as compared to other conventional methods and the huge reduction in cost per se. The path involves several commercial NGS platforms that are based on different technologies but typically follow a general pattern or steps. General steps involved in the DNA sequencing using NGS are (i) library preparation (random fragmentation of genome, ligation with appropriate adaptors), (ii) amplification of library, and (iii) sequencing using different approaches. The basic steps involved are presented in a flowchart in Fig. 15.2. The generated results differ with respect to read length, quality of the data, and quantity of the data based upon the platform used. Classification of different sequencing technologies based upon the type of technology, chemistry, detection system used, and method of amplification used in different generations of sequencing platforms is described in Fig. 15.3. Here in the following section, we will discuss about the current sequencing technologies, their principle, and their advantages and limitation.

### 15.4.1 Pyrosequencing Technology

Nyren's group pioneered this sequencing by synthesis (SBS) approach technology in 1993, where DNA sequencing is based upon detection of released pyrophosphate



**Fig. 15.2** Schematic representation of the basic steps involved in DNA sequencing using different NGS platforms



**Fig. 15.3** Classification of the next generation sequencing (NGS) on the basis of technology, chemistry, detection system and clonal application

(PPI) molecule during addition of nucleotide by DNA polymerase (Ronaghi et al. 1996). The speed of the reaction is very fast as it just takes 3–4 s at room temperature to complete the reaction from addition of nucleotide to chemiluminescent detection. Later, 454 Life Sciences (a USA-based biotechnology company which was later acquired by Roche) took over this technology and commercialized it with some



modifications. Pyrosequencing uses the enzyme luciferase from *Photinus pyralis* (American firefly) and recombinant ATP sulfurylase from *Saccharomyces cerevisiae* (Karamohamed et al. 1999). Two different approaches are available in pyrosequencing: (i) solid phase (Ronaghi et al. 1996) where DNA gets immobilized (based on three-enzyme cascade method with washing step) and (ii) liquid phase (Ronaghi et al. 1998) in which a nucleotide-degrading enzyme, pyrase, has been added (based on four-enzyme cascade method without washing step) which excludes the requirement of solid support because of which reaction can be performed in a single tube.

#### 15.4.2 Roche 454 (GS FLX plus)

In 2007, 454 Life Sciences, a pyrosequencing technology-based sequencer, was taken over by Roche and later known as Roche 454. Steps involved in pyrosequencing are fragmentation of the nucleic acid to be sequenced followed by synthesis of template strand with the assistance of polymerase enzyme. On incorporation of new nucleotide by polymerase, the pyrophosphate molecule is released. This pyrophosphate molecule converts the ADP to ATP in the presence of the enzyme ATP sulfurylase. ATP molecule supplies the energy for the oxidation of luciferin, which subsequently emits light and recorded by charge-coupled device (CCD). The identity of nucleotides added (all four dNTPs added in the reaction separately) into the reaction for polymerization is known to help in the determination of template sequence.

Pyrosequencing-based platform uses a massively parallel system for sequencing in picolitre volumes in microfluidic format. In brief, the methodology of sequencing is the fragmentation of DNA (~800 bp) using spray method (nebulizer), adaptors ligation to the fragmented DNA, library preparation followed by attachment of library to beads (DNA capture beads, which allows attachment of one fragment per bead). The beads make individual compartments, which are usually referred to as microreactors/microvesicles. Clonal amplification takes place in those compartments by emulsion PCR; subsequently, emulsion gets broken and beads attached clonally amplified DNA becomes enriched in microreactors/microvesicles (Margulies et al. 2005). All clonally amplified DNA-bound beads are individually loaded on picotitre plate (PTP; imprinted on the fiber-optic slide), which contains approx.  $3.4 \times 10^9$  wells of ~55  $\mu\text{m}$  in depth and 75 picolitre calculated size. The slide/plate comprising the picolitre-sized wells mounted in a flow cell, which forms the channel for the flow of the sequencing reagents above the wells. The base of the plate is connected with the imaging device called as CCD that captures the emitted light and provides the results in the form of flowgrams. Genome sequencer GS FLX produces nearly 450 MB data from a single run whereas new genome sequencer (GS FLX plus) can produce 700 MB data from a single run in ~10 h with an approximate cost of 5K to 7K USD.

**Advantage** This is a fast (700 MB data in a day), accurate (~99.9% after filter) and reliable technology for high throughput real-time sequencing. The Roche 454 is an upgraded platform, which can give read length >700 bp. Further, the technology

does not require labeled nucleotides and primers followed by gel electrophoresis and suitable for *de novo* sequencing as well as confirmatory sequencing (Ronaghi 2001). The technology provides flexibility in primer designing since it starts sequencing immediately downstream of primer sequence without keeping gap of 30–40 bp.

**Limitation** The platform sure has some limitation. The main constraint is the problem in sequencing same nucleotide repeat (>8 bp), that is, homopolymer sequencing (Mardis 2008). Another, it is relatively costlier than other NGS technologies.

---

## 15.5 Reversible Terminator Technology

This technology also depends on the sequencing by synthesis (SBS) strategy. Dr. Jingyue Ju was the first person who described reversible terminator sequencing technology (Li et al. 2003). The basic difference between traditional sequencing and this is that the traditional sequencing uses ddNTPs to irreversibly terminate the extension of the primer whereas reversible sequencing technology employs modified analogue of nucleotide to terminate the extension of primer reversibly (Guo et al. 2010). In the past decade, numerous reversible terminators have identified based upon reversible blocking groups and can be categorized into two types. One category belongs to 3'-O-blocked reversible terminators and other is 3' unblocked reversible terminator. Illumina Solexa commercializes this technology because of its comprehensive acceptance in second-generation sequencers (Bentley et al. 2008).

### 15.5.1 Illumina Solexa

As stated above, this is the popular NGS platform of second generation. David Klenerman and Shankar Balasubramanian gave the idea for only one DNA molecule sequencing attached to microsphere with the foundation of Solexa in 1998. The “Solexa Genome Analyzer” system, which came in 2006 and later, it was acquired by Illumina for the sequencing of clonally amplified DNA (Voelkerding et al. 2009).

A flow cell, which is used by Illumina Solexa, is made up of optical transparent slide having eight lanes on its surface. Oligonucleotide anchors are immobilized to the surface of the flow cell. In brief, the methodology of the sequencing is the fragmentation of template DNA, end repairing of fragments (blunting and 5' end phosphorylation). Adenylation of 3' ends by the addition of single 'A' nucleotide to facilitate ligation with the oligonucleotide adapters carrying a 'T' overhang at 3'. Since ligated adapters are complementary to the flow cell, anchors thus get hybridized. The DNA template attached to the anchors of the flow cell relies for the cluster generation by “bridge amplification” in contrast to emulsion PCR (Adessi et al. 2000). Further, DNA fragment makes an arc and hybridizes with its distal end to neighboring anchor oligonucleotide to its complementary part. Because of clonal



amplification, each template generates thousands of copies (clusters) of same template DNA and subsequently millions of separate (unique) clusters are generated on the single flow cell followed by addition of DNA polymerase, and four different fluorescent-labeled reversible terminators aid in sequencing the millions of clusters in parallel fashion onto the flow cell. Polymerization terminates because of the incorporation of fluorescent-labeled reversible terminators (ddNTPs) and incorporated nucleotide is identified by the fluorescence captured (Guo et al. 2008). By the enzymatic cleavage, the fluorescence label gets cleaved, which permits the incorporation of next nucleotide ([www.illumina.com](http://www.illumina.com)). Recently, several technical improvements are happening rapidly in this sequencing technology including library preparation method, which involves fragmentation of the DNA in acceptable sizes by Covaris (Adaptive Focused Acoustic wave) sonication method and improved adapter ligation efficiency, etc.

Illumina platforms dominate in high throughput sequencing market. Currently, Illumina is producing a series of platforms (MiSeq, HiSeq series and NextSeq series). Different platforms are augmented for throughput and turnaround time of the run. Most recognized platforms are MiSeq and series of HiSeq platforms. The MiSeq is a personal tabletop sequencer marketed in 2011, where a run can be completed in as fast as in four hours for the targeted bacterial sequencing. On the contrary, HiSeq 2500 is applicable for high throughput sequencing like 1 TB data from a single run in 5–6 days. New model of HiSeq platform, that is, HiSeq 2500, can also be run in fast mode. However, run will not be cost effective and will sequence 30X human genome in approximately 27 h. In the beginning of 2014, Illumina launched another two NextSeq 500 and HiSeq X Ten. NextSeq 500 is similar to MiSeq made for individual labs. HiSeq X Ten platform works as whole genome sequencer at population scale. Presently, Illumina only supports sample of human for whole genome sequencing on HiSeq X Ten. MiniSeq platform from Illumina also came into the market in 2016. Most recently released platforms from Illumina are HiSeq 3000 and HiSeq 4000 based on the patterned flow cell technology. Their data output and run time lie between HiSeq X Ten and HiSeq 2500 (Reuter et al. 2015).

The latest machine from Illumina launched at the end of 2017 is HiSeq 100. This is the smallest and most inexpensive sequencer in the portfolio of Illumina with a maximum data output of 1.2 GB and 4 million reads per run, with run time ranging from 9 to 17.5 h ([www.illumina.com/iseq](http://www.illumina.com/iseq)).

**Advantage** First and foremost, the technology provides high throughput data in a very short period of time with very low amount of sample per run (Buermans and den Dunnen 2014). The newer platforms of Illumina like HiSeq 2500, HiSeq 2000 and MiSeq generate more data (up to 600 GB) at low cost per base. Using the platform based on reversible terminator technology, 1 TB data can be generated in a single day. Another big advantage is the longer read lengths, that is, 300 bp paired-end sequencing in Illumina (MiSeq) platform is now possible, which was earlier 25 bp single-end reads by Solexa. Besides this, Illumina platforms provide 99.9% accuracy of the sequencing data (Morey et al. 2013). Because of the presence of

blocking group, the addition of only one nucleotide per cycle facilitates the sequencing of homopolymeric regions efficiently (Mardis 2013).

**Limitation** Major limitation is guanine and cytosine (GC) biasness, which gets introduced during bridge amplification (Mardis 2013). Another concern is dephasing, which means the different copies of DNA in a cluster get out of sync (inconsistent). In other words, inappropriate deblocking of nucleotide results in varying length fragments in a cluster. This decreases the accuracy in base calling at 3' end of the DNA fragments, especially in invert repeat sequence (Nakamura et al. 2011).

### 15.5.2 Sequencing by Ligation Technology

This is a DNA sequencing technique, which determines the DNA sequence by utilizing the mismatch sensitivity of DNA ligase enzyme (Ho et al. 2011). Applied Biosystems, USA marketed this technology in 2008. The platforms of this technology rely upon oligonucleotide probes of variable lengths, labeled with different fluorescence tag liable to the nucleotide to be sequenced.

### 15.5.3 ABI-SOLiD

The expansion of SOLiD is small oligonucleotide ligation and detection system. The technology was invented in 2005 by George Church. Later in 2008, it was further upgraded and marketed by Applied Biosystems (Voelkerding et al. 2009), which is now acquired by Life Technologies. The sequencing reaction can be divided into five broad steps: (1) preparation of DNA library (2) clonal amplification in microreactors by emulsion PCR (3) attachment of the beads (4) sequencing and (5) resetting of primer. In brief, the methodology for sequencing using this technology includes fragmentation of DNA, attachment of fragments to the beads and clonal amplification of fragments attached to the beads by emulsion PCR. Following this, adapter sequences in amplified fragments hybridized to the specific primers. This facilitates the ligation of octamer (eight base pair) interrogation probe (fluorescently labeled) by offering 5'  $-PO_4$  group in place of 3'  $-OH$  group. The first two bases of interrogation probes are specific while the other six are degenerate. Set of four different fluorescent-tagged probes (interrogation probe), having one out of sixteen possible combinations (e.g., AC, AT, AG, CG, TC, GT and TT) of two base (specific) at the end compete to ligate with the primer. After ligation, fluorescence is imaged that is equivalent to the interrogation probe ligated. For another round of cycle, 5'  $-PO_4$  group regenerated by the deletion of fluorescence label of the attached interrogation probe. Further, steps of the previous cycle are repeated after injection of set of four different fluorescent-tagged probes. Generally after seven cycles of ligation, template is reset to the n-1 position of complementary primer for another round of ligation. This procedure is repeated every time with a consecutive offset like n-1, n-2 and so on of a new primer. A single run on SOLiD 5500 platform takes approx. 6–7 days to complete

and produce 120–240 GB data with 75 bases read length while SOLiD 4 platform generates 100 GB data.

**Advantage** This technology offers the highest accuracy of ~99.99% (Voelkerding et al. 2009) because each nucleotide sequenced two times, thus, there is pretty less chances of miscall from two contiguous colors.

**Limitation** One of the limitations is time taken in one run is so long (6–7 days) along with production of less data as compared to Illumina platforms. Another limitation is that if we compare with other second-generation methods (sequencing by synthesis), the analysis of the data is complex in this technology, which impedes the marketing of the platform. *De novo* sequencing is another limitation in using this technology.

---

## 15.6 Third-Generation Sequencing Technologies

Second-generation technology platforms were the most widely used platforms. But the major limitation of them was the occurrence of biasness because of PCR amplification step. On the contrary, third-generation sequencing technologies do not require amplification step and are capable of sequencing single DNA molecule in real time. These platforms have the capability to provide single run at very low cost as well as made the preparation of sample easier. Further, third-generation platforms produce generally longer read of about some kilobase length, which resolve the problem of assembling the reads.

Dr. Stephen Quake's team established the first single-molecule sequencing technology (SMT) (Braslavsky et al. 2003), which was further commercialized by Helicos Biosciences.

### 15.6.1 Single-Molecule Real-Time Sequencing

Nanofluidics Incorporation pioneered single-molecule real-time sequencing (SMRT). This technology for sequencing is based upon two key inventions: phospholinked nucleotides and zero-mode waveguides (ZMW). The key feature of ZMW is that it only permits light to illuminate the bottom of a well where the immobilized template and DNA polymerase are present.

#### 15.6.1.1 Pacific Biosciences (PacBio)

The PacBio depends on the SMRT technology, which was commercialized in 2010 by Pacific Biosciences (Roberts et al. 2013). This allows us to know about the synthesis of DNA in real time. This is possible because of the presence of zero-mode waveguide (ZMW) holes where DNA synthesis takes place.

Methodology in brief, single stranded adapters (hairpin adapters) ligated to the fragmented DNA, which is then known as capped template. Here to increase the accuracy, a strand displacing DNA polymerase is used to sequence the same template several times (Travers et al. 2010). DNA polymerase and template immobilize at the base of the ZMW where DNA synthesis takes place (Levene et al. 2003). Approximately 75K ZMWs/SMRT cells are present allowing 75K single-molecule reactions in parallel fashion. There is concept of physics, which does not allow the laser light (600 nm) to pass completely through the ZMWs (because of zepto-litre holes) and it decays exponentially after entering into ZMW. Thus, the laser light only illuminates the 30 nm of the hole. Phospholinked nucleotides of all four types subsequently pass over the ZMWs. The nucleotides get excited and fluoresce when reaching the base because laser cannot penetrate up within the hole. Therefore, nucleotides cannot be fluoresced till they are present on the surface of the hole. Hence, polymerization reaction takes place continuously; fluorescent signals can be detected in real time so the sequence can be read (Eid et al. 2009). RS II platform of Pacific Biosciences, which is commercially available, was released in 2010, whereas PacBio Sequel is the latest platform released by Pacific BioSciences in autumn 2015 in collaboration with Roche Diagnostics for the development of clinical grade sequencer for diagnostics. RS II platform is the foremost platform that offers sequencing read length >20 Kb and PacBio Sequel is analogous to its former. The new platform PacBio sequel generates almost seven times (~365,000) more number of reads than RS II (55,000). The driving force for the long read length is single-molecule real-time technology in combination with zero-mode waveguides.

**Advantage** The reaction can be monitored in real time is the biggest advantage per se, which permits to gather the data related to base composition or sequence of the DNA template as well as the enzyme kinetics. Difference in enzyme kinetics provides us the clue about different modifications present in the DNA like methylation (6-methyladenine, 5-methylcytosine) (Flusberg et al. 2010; Fang et al. 2012). With the help of the identification of this modification sites genomewide, the approach can also be used to identify the potential modification present in different genetic diseases. Additional advantage includes longer reads and unbiased data. Moreover, using the SMRT approach, not only DNA, ribosome can also be observed at single-molecule resolution (Uemura et al. 2010). *De novo* sequencing can be easily performed because of the longer read length. Hence, short read allows error in the assembly of fragments and formation of scaffold in repeat and GC-rich regions (Bahassi el and Stambrook 2014).

**Limitation** The cost per base sequencing being relatively high and the lower throughput are disadvantages, which limit its use in maximum genomewide studies. Higher error rate (insertion and deletion) is the main limitation of the technology (Mardis 2011). Along with this, the data generated per run is very less as compared to second-generation platforms.

## 15.6.2 True Single-Molecule Real-Time Sequencing

This technique also provides us the sequence of single molecule of template DNA. Hence, evade the requirement of clonal amplification and library preparation (Harris et al. 2008). It is also known as single-molecule fluorescent sequencing. In this method, virtual terminators were employed for the fluorescent detection of nucleotide. The virtual terminators were introduced in 2009 (Bowers et al. 2009) for the third-generation sequencing technology. Their working methodology is similar to the second-generation reversible terminators. However, different fluorescent dyes and different blocking groups express different features rendering to structure and nucleotide binding region of them. Generally, virtual terminators are made up of free 3' -OH group, which interacts with DNA polymerase and fluorescent molecule bounded with the linker group (Korlach et al. 2010).

### 15.6.2.1 Helicos Biosciences (HeliScope)

This is the first commercial platform of third generation, which has again revolutionized the DNA resequencing technology in 2008. Helicos Biosciences Corporation, Cambridge MA, USA has launched the HeliScope platform and produces  $3 \times 10^7$  reads/channel in a channel slide format (Metzker 2010). In this method, the template to be sequenced is fragmented and polyadenylated at the 3' end by terminal transferase. The flow cell, which is coated with the oligo-dT containing primers gets hybridized to poly-A tail present in the fragments. To avoid Poly-A tail sequencing present in the fragments, virtual terminators of nucleotides other than dTTP are added in the initial sequencing step. The principle/chemistry of the sequencing is same as for sequencing by reversible terminator. Likewise, in the cyclic extension manner template, molecules get sequenced and imaged using CCD camera after cleavage of blocking group and fluorescent dye (Thompson and Steinmann 2010). Approximately 35 GB data can be generated from a single run on this platform.

**Advantage** It does not require clonal amplification step, which resulted in reduction of biasness. So, it is a beneficial substitute for applications that are mostly affected by PCR biasness like RNASeq. This is the first sequencing method able to sequence every single nucleotide of each DNA molecule from the fossils, which provide the information regarding DNA damage (Krause et al. 2010). Furthermore, this requires very less concentration of template molecule.

**Limitation** Despite the advantages, there are some limitations as well, like shorter reads and high cost, because of the repetitive sequencing to get the most accurate data by reducing error rates. Shorter reads make it difficult to be used for the *de novo* sequencing. Also, the technology does not generate the paired end sequences, which could help in the orientation and location determination of contigs for assembling of the data. In 2012, Helicos Biosciences announced the impoverishment. However, SeqLLCompany (Boston MA, USA) provides the services for DNA and RNA sequencing using this technology.

### 15.6.3 Ion Semiconductor Sequencing

Basically, this is an extension of the pyrosequencing technology as described by Ansorge in 2010. This technology uses a chip (semiconductor), which is fabricated with millions of micro wells. These wells capture the release of proton ( $H^+$ ) during the sequencing followed by change in pH. Proton detected by the technology is the product other than PPI molecule released during polymerization. This is an amalgamation of semiconductor technique (digital) and chemistry, which allows the expression of chemical signals into digital data (to determine the base call/sequence). DNA Electronics in London licensed the principle of this technique, that is, the detection of proton ( $H^+$ ).

This is a first commercial sequencing technology, which does not demand for costly optics, lasers and different fluorescently labeled nucleotides for complex sequencing chemistries (Ansorge 2016).

#### 15.6.3.1 Ion Torrent

Ion Torrent released the Personal Genome Machine (PGM), a compact benchtop platform in late 2010 that was later acquired by Life Technologies, Carlsbad, USA. This machine uses the high-density arrays and generates the data usually 10–20,000 MB per run of up to 400 bp read length in 2–7 hours. It is according to the chip used and also based on application purpose. The basic methodology is almost similar to other NGS technology, which involves fragmentation and ligation of adapters. Hybridization to the complementary sequences (primers) bounded with the beads followed by emulsion PCR. After clonal amplification, beads flooded over the semiconductor chip. Each bead goes to the individual well present on the chip flowed by the floating of nucleotides sequentially. On incorporation of each nucleotide by DNA polymerase, a proton ( $H^+$ ) is released, which results in change of the pH. This change in pH determines the base sequence by changing the chemical signal into digital signal. In homopolymers (more than one same nucleotide incorporated simultaneously), intensity of signal gets high, which is parallel to the pyrosequencing technology (Quail et al. 2012).

In the third quarter of 2012, Ion Torrent released its advanced and bigger platform named Ion Proton. This new platform play an important role in the sequencing of whole genomes, transcriptomes and exomes. The data output was up to 10,000 MB with 200 bases read length in a very short duration, that is, two to four hours. The platform has a number of applications including *de novo* sequencing, ChIP sequencing, sequencing analysis of the methylation in DNA, small RNA sequencing and gene expression analysis. Subsequently, other versions, namely, Ion S5 and Ion S5 XL, came with their broad range of applications having both low throughput and high throughput. In reference of throughput, these platforms can be compared to Illumina HiSeq platforms.

**Advantage** These platforms require very low input of (DNA or RNA) concentration (~10 ng) for the identification of mutations and expression profile. Also, the technology has simplified the analysis of the sequencing data because of the new



Ion Reporter Software. There are plug-ins and operating software available for data analysis from amplicon sequencing, microbial sequencing, etc. They are widely accepted because of the reasonable cost, though they generate shorter reads than some other platforms like PacBio. In simpler words, they are affordable, rapid (run completes in 2–4 h) and simple and so are suitable for laboratories.

**Limitation** Major limitation is lack of coverage in sequencing of genomes that contain very high-AT content using Ion Torrent Personal Genome Machine (Ballester et al. 2016). Another difficulty in homopolymer sequencing is stretches of more than six same nucleotides trigger deletion and insertion error rate (~1%) (Reuter et al. 2015).

---

## 15.7 Fourth-Generation Sequencing Technologies

This generation of the sequencing has the ability to in situ (perform sequencing directly in the cell) sequence the fixed tissue and cells by using second-generation methodology (Mignardi and Nilsson 2014). Targeted and untargeted methods developed for in situ RNA sequencing are based on the principle of ligation chemistry. Further, it has the quality to sequence the entire human genome rapidly and authentically at very low cost, that is, <\$1000. Thus, fourth generation has come up with the use in numerous applications like validation of biomarkers and transcriptomic analysis. A group led by Church overcame the limitation present in the Ke's method where they gave an idea of partition sequencing to reduce read density. With this approach, determination of expression of large number of genes in the cell is possible in parallel fashion for several types of RNA. Example: mRNA, rRNA, anti-sense RNA and non-coding RNA (Mignardi and Nilsson 2014). Using in situ sequencing method, it is possible to screen the whole cell population with the resolution of single cell.

Fourth-generation platforms based on recent technique, namely, "Spatial Transcriptomics," are in their infantile stage (Stahl et al. 2016). Again, this technique is also based upon NGS chemistry for the sequencing. This technique offers the simultaneous visualization and quantitative analysis of the transcriptome (gene expression data) in the fixed tissues. Nanopore-based sequencing method is also available to sequence nucleic acid inexpensively in a short duration.

### 15.7.1 Nanopore Technology

Though this idea of nanopore sensors based sequencing is very old, it was first envisaged by David Deamer in 1989. This portable technology emerged from coulter counter and ion channels. On the supply of the voltage, particles of smaller size than the pore size circulated across the pore. The read lengths of > 150 Kb can be attained. Now, many companies have offered the strategies for nanopore-based sequencing. One is NanoTag sequencing by Genia where DNA strand gets excised

in monomers followed by their channeling one after another, across a nanopore. Another is strand sequencing by Oxford Nanopore; here, whole single strand of DNA passes through a nanopore, which allows the pulling base by base in only one direction (ratcheted) via nanopore. Till now, Oxford Nanopore Technology is the most successful technology for strand sequencing by nanopore.

### 15.7.1.1 Oxford Nanopore (MinION)

MinION is the first nanopore device for sequencing. Oxford Nanopore Technologies, UK, licensed it in 2007 and commercialized in May 2014. A flow cell is present at the core of this device in which 2048 individual nanopores are present. They are divided into four groups of 512 nanopores in each group and controlled by application-specific integrated circuit (ASIC). Brief methodology for sequencing involves ligation of adapters to the fragments at each end. Adapters enable capture of the fragment and polymerase binding at the 5' ends of the fragments. Additionally, these adapters concentrate the DNA fragments closer to the nanopore, which enhances the rate of fragment capture thousandfold. Also, these hairpin-like adapters allow adjoined sequencing of two complementary strands by covalently attaching the strands to each other. On translocation of a fragment through nanopore, the polymerase processes along the template strand and the process repeats for complementary strand. The sensor identifies the change in ionic charge when fragments move through the nanopore. The change in the ionic charge or characteristic disruption in current is divided into separate events, which ensure associated duration, mean amplitude and variance. The series of events is finally interpreted using computer software/graphical models (e.g., MinKNOW) to identify the nucleotide sequence. Finally, the information collected from the template and complementary strands is merged to generate the "2D read." Another available method for the library preparation does not involve the hairpin adapters to covalently connect two strands of the fragments. This method generates the "1D reads"; so in this, nanopore reads only template strand. However, this allows high throughput but slightly less accuracy in data in comparison to 2D data (Jain et al. 2018).

**Advantage** The key advantage of this technology is that the device is able to produce long read length > 882 Kb (Quick et al. 2017; Jain et al. 2018). The ultra long read length provides the comfort in data alignment and assembly, which lowers the computational burden. Another advantage is that it is a portable device, which also provides us chance to see the data in real time. MinION is an economic and high throughput device for sequencing of nucleic acid. The biggest advantage of its portability and mini size is that it has given a chance to look for the opportunity of life in outer space (Castro-Wallace et al. 2017). This is the first DNA sequencing platform used in the cosmos. Major properties of MinION assists in rapid surveillance of epidemics like Ebola virus and Zika virus. Nanopore sequencers can also detect cytosine modifications in the native DNA (Rand et al. 2017).

**Limitation** One major drawback of the nanopore technology is the higher error rate. Recently in the mid of 2016, Oxford Nanopore Technologies launched a newer version of MinION. This platform is based on the newer chemistry known as R9 (R stands for reader) and providing the lesser error rates (<https://nanoporetech.com/about-us/news/update-new-r9-nanopore-faster-more-accurate-sequencing-and-new-ten-minute-preparation>). However, it is also not up to the mark for frontline applications. Currently, the latest version of the technology i.e. R9.4 is getting used in the flow cells of the MinION platform. Since the improvements in next-generation technologies are taking place very quickly, the limitations of this generation will also be taken care very soon.

Following are the features why Nanopore sequencers are more suitable for sequencing:

The biggest quality of them is no requirement of fluorescent labeling of the nucleotides in sequencing and also provides longer read length. Identification of nucleotide is based upon the chemical or electronic structure. Compact size (four-inch-long device) of the machine offers the in-field/natural environment experiments possible and it has lower cost of sequencing per run including higher throughput. This MinION platform has the capability to break the set market of \$1000 target fixed by the NHGRI, USA.

#### 15.7.1.2 Oxford Nanopore (ProMethION)

In the beginning of year 2017, Oxford Nanopore Technologies delivered its new highest throughput sequencing benchtop platform ProMethION to the laboratories, which has been commercially available in May 2018. Here, up to 48 flow cells can run independently, each consisting of 3000 channels (nanopores). Oxford Nanopore Technologies promised that ProMethION would perform even better than the best platform by Illumina. They have also assumed the data it will generate per run will be approximately 11 TB when the manufacturing was underway. Currently in 2018, it generates approximately 2 TB in 48 h.

Now this platform has been placed in many sites in many countries proving that results continue to rise. In June 2018, more than 100 GB from the individual flow cell of ProMethION at University of Aalborg was first achieved. In the University of Birmingham, at the time of writing of this chapter (first week of August 2018), ProMethION benchtop platform was able to break the record by producing >130 GB data/flow cell (<https://nanoporetech.com/about-us/news/promethion-wild-2-data-yield-continues-climb>).

**Advantage** The platform offers on-demand sequencing. It means the researcher can start and stop the run as and when required or utilize more than one flow cell for single experiment for high throughput and faster speed. Individual flow cells of 3000 nanopores can be used and it provides almost six times more data compared to MinION and GridION. Currently, ProMethION beta system module is in use that allows 192 different libraries within the whole device.

### 15.7.1.3 Oxford Nanopore (GridION X5)

In early 2017, Oxford Nanopore Technologies released another platform, namely, GridION X5. This is a grid collection of five units of MinION with built-in computing software for the base calling. This allows five queries simultaneously or individually at a time depending upon the requirement of the researcher.

This platform also uses the same core technology and is useful in generating huge data (~35 GB in 2017) of long read length along with immediate access to the data in real time like MinION and ProMethION (<https://nanoporetech.com/products/gridion>). Library preparation is very easy and fast and almost the same that of MinION, ProMethION and GridION. It requires very less concentration of sample (femtogram for >40 Kb DNA) and also has versatile and complete range of cDNA and gDNA library preparation kits. There are two methods for library preparation based on amplification of library. PCR amplification-based preparation requires when the starting amount of DNA is low. If we use 20 GridION platforms at a time, then it can sequence the whole human genome in only 15 min with relatively lower cost. A compact microfluidic device, VolTrax (programmable Hand Off preparation of Sample), is made available by Oxford Nanopore Technologies for the automation of library preparation (Leggett and Clark 2017) that is made up of USB-powered base. A consumable cartridge can be placed onto this base with an array of fluid comprising pixels on the surface. The software controls movement of drops of fluid on the surface of the cartridge. In addition, Oxford Nanopore Technologies is in the making of other protocols and customized user protocols. Recently, a new protocol has been developed for direct RNA sequencing, which is the most awaited protocol. The initial versions are not as precise as those of DNA but there is high hope to see the potential method shortly for direct RNA sequencing.

### 15.7.1.4 Oxford Nanopore (SmidgION X5)

In late 2017, Oxford Nanopore Technologies released an even smaller platform than MinION that can be attached with the mobile phone, SmidgION. This device is also based on the same technology like MinION and ProMethION. This is a very small 128-nanopore channel flow cell platform for sequencing of clinical, environmental and ecological samples. It is useful in the monitoring of the outbreaks (pathogens) remotely (<https://nanoporetech.com/products/smidgion>).

Other than Oxford Nanopore Technologies, Hitachi (Goto et al. 2016) and Genia (maintained by Roche) are among the companies who are working on the biological nanopore technologies. But, there is no company till now that has launched its platform in the market. This Oxford Nanopore Technologies is competing with both long-standing PacBio longer read and Illumina's shorter read technologies. On the other hand, Oxford Nanopore Technologies is providing platform at almost no cost or very low cost. Laboratories have to just pay for the consumables.

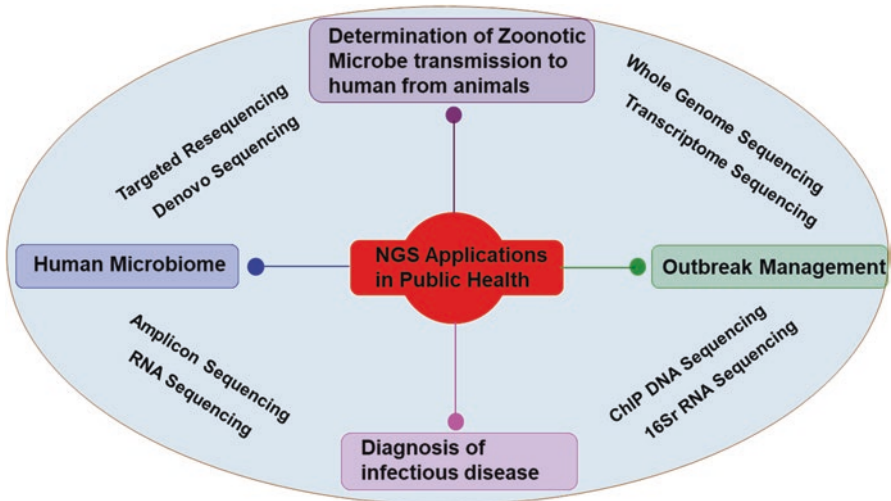
Table 15.1 provides the overview and characteristics of commonly used next-generation sequencing platforms.

**Table 15.1** Overview and characteristics of new and commonly used next generation sequencing platforms

Company	Platform	Maximum Read Length	Throughput/run (Gb)	Run time	Error rate (%)	Error type
Illumina	MiSeq (Benchtop)	2 × 300	0.3–15	5–55 h	0.1	Mismatch
	HiSeq (Capital)	2 × 150	125–700	<1–4 days	0.1	Mismatch
	MiniSeq (Benchtop)	2 × 150	0.5–7.5	4–24 days	1	Mismatch
	NextSeq (Benchtop)	2 × 150	20–120	12–32 hr	1	Mismatch
	iSeq100 (Benchtop)	2 × 150	0.2–1.2	9–17.5 hr	<1	NA
Thermofisher (Ion Torrent)	PGM (Benchtop)	200–400	0.03–2	2–8 hr	1	Insertion/deletion
	Ion S5 (Benchtop)	200–400	0.5–15	2–4 hr	1	Insertion/deletion
	Ion S5 XL (Benchtop)	200–400	0.5–15	<24 hr	1	Insertion/deletion
Pacific BioSciences	RS II (Capital)	>20,000	0.5–1	30 min–24 hr	15	Insertion/deletion
	Sequel (Capital)	>20,000	0.6–1.25	30 min–6 hr	NA	NA
Oxford Nanopore	MinION (Portable)	230,000–300,000	21–42	1 min–48 hr	12	Insertion/deletion
	ProMethION (Benchtop)	>200,000	3–6 Tb (>130 Gb/flowcell)	1 min–64 hr	NA	NA

## 15.8 Applications of NGS to Address Public Health

To advance public health by unlocking the power of genome, we have just started utilizing high throughput sequencing technology, that is, next-generation sequencing. Schematic representation in Fig. 15.4 is illustrating that all the applications that address the public health utilize or take advantage of the different methods of high throughput sequencing technology like whole genome sequencing, targeted sequencing, ChIP sequencing, RNA sequencing, whole exome sequencing, transcriptome sequencing and amplicon sequencing. NGS is exhibiting a broad impact in public health welfare and clinical laboratories. This new high throughput technology holds remarkable promises with a wide range of applications in exploring biological questions, which includes management and surveillance of outbreaks, study of human microbiome to investigate the infectious organism/polymicrobial infections and taxonomic identification of microbiomes, diagnosis of infectious disease and investigation of zoonotic microbes transmission to humans from animals and so on. In this chapter, some key areas of NGS applications related to public health are summarized:



**Fig. 15.4** Applications of Next-Generation Sequencing (NGS) in Public Health

### 15.8.1 Outbreak Management

“An outbreak anywhere is a risk everywhere”—Dr. Frieden. Outbreaks can be stressful for individuals and public. However, traditional epidemiology generally catches the source of an outbreak, for example, by case control studies (King et al. 2012). For the past so many decades, laboratory investigation played a significant role in investigation and management of outbreak (Sabat et al. 2013). Now, Whole Genome Sequencing (WGS)-based typing is encouraging the employment of next-generation sequencing for investigations of public health. WGS is very useful in the detection of outbreak and its management locally and globally and also in the monitoring of evolution of multidrug resistance pathogens (Albiger et al. 2016). First application of WGS in public health was to dissect the epidemiological connections in hospital-acquired infections, for example, bacterial (*Acinetobacterbaumannii*) outbreak in 2010 in a hospital in Birmingham, UK (Lewis et al. 2010). In a very short time, several studies have shown that WGS has taken charge for interpretation and stopping of the transmission pathways of pathogens in hospital outbreaks. Few examples of characterization of newly emerging pathogens, which helped to stop the transmission/spread between patients at the same center and inter health care centers transmission, are Methicillin Resistance *Staphylococcus aureus* (MRSA), carbapenem resistant *Klebsiellapneumoniae* (Harris et al. 2010; Snitkin et al. 2012) and early detection of *K. pneumoniae* high-risk clone (HiRiC) (Zhou et al. 2016).

A large outbreak of highly virulent Shiga toxin producing *Escherichia coli* (STEC) was also characterized by WGS. Upon characterization, WGS has the ability to reveal about isolates like species, strain, virulence, antibiotic resistance and much more information from the genome other than phylogenetic information to manage the case and outbreak. Investigation of foodborne disease and its outbreak



management is the most important area for public health welfare, and WGS promises to identify those bacteria. According to a WHO survey, approximately 1900 million people get infected with foodborne pathogens every year and out of them, a big proportion (7,15,000) are not able to survive (2007–2015, Food Borne Disease Burden Epidemiology Reference Group). Great improvement has been found in the outbreak detected and outbreak solved (management) of listeriosis by traditional gold standard method in countries like France and US after implementation of WGS. Information about drug resistance and virulence characters can be taken into account for the clinical practice but more correlation among genotype and phenotype is required. Further, NGS data mining may disclose new targets, which may help in the investigation of outbreak by highly clonal pathogens. Remarkably, a major downside of NGS is that there is no standard guideline from controlling agencies for the sharing of data.

### 15.8.2 Human Microbiome

Human microbiome is one of the important players to affect the immunity of the host and metabolic functions that are not determined by human genome. Microbes, mainly bacteria, are both closest relative and enemy of our body. Yeast, single-celled eukaryotes, helminths and some viruses are also associated to our body. There are organisms in our microbiota, which cannot be cultured and identified. Some very well-known spots are found for the colonization in the human body like stomach, vagina, colon, skin, esophagus, oral cavity, nose and hair. In 2008, National Institute of Health (NIH), USA has started two high-profile human microbiome projects of international level, namely, Human Microbiome Project (HMP) and Metagenomics of the Human Intestinal Tract (MetaHit) on the foundation of NGS as a tool. These projects were initiated to isolate and characterize microbes present in the healthy individuals and diseased and to develop the new methods for computational analysis of sequenced genomes. In the MetaHit program, from the sequencing of gut microbiome of 124 healthy adults having obesity, it was found that there are >1000 bacterial species present in the human gut microbiota. This revolutionary project was based on the *de novo* assembly of short reads from human microbiome datasets (Qin et al. 2010; Arumugam et al. 2011).

The new sequencing technologies like NGS are facilitating researcher community to analyze the world of different microbial populations in varied environments and in human body from wider and deeper viewpoints. The ongoing advances in sequencing platforms are not only supporting the characterization of whole genome of microbes but also are a valuable tool for the taxonomic identification of microbiomes, which are an inhabitant of particular niche. This enables us to detect polymicrobial infections and colonization at better perspective. Also, identification of new species using sequencing of metagenomes by implementing whole genome sequencing method or 16S rRNA gene amplicon sequencing from a mixed population. A new approach, that is, metatranscriptomic method of sequencing, has made the contribution in functional analysis of the interactions among different microbes of a

single microbiome. In case of whole genome sequencing, we get longer read length, which enables us in better assembling of genome from diverse organisms. By using reference sequences or denovo clustering, taxonomy profiling can also be performed. Nowadays, it is very much noticeable that imbalance of gut microbiome (also known as dysbiosis) is intensely connected with the immune disorder development and/or improper metabolic functioning (Kim et al. 2015). Here, we can say that microbiomics has its use in a very controlled manner at the point of care. For instance, if a person is suffering from meningitis and doctors/clinicians are not able to identify (pathogen eluding the clinicians like criminal eluding the police), then microbiomics plays pivotal role for the identification of the responsible pathogen, be it bacteria or amoeba. This identification helps the clinicians to precisely medicate (precision medicine) the patient with choice of antibiotics. In other words, genomics in these cases is life saving.

### 15.8.3 Diagnosis of Infectious Disease

Recently, the application of next-generation sequencing has also approached the field of infectious disease diagnosis like any other medical field to empower the public health globally. Earlier, the origin of the disease causing microbe and its diagnosis was only based upon the evidence of the existence of a given pathogen in a particular given sample. This standard time-taking method of culture-based identification/detection is still being used. However, this classical approach has many limitations, which include problem in cultivation of certain species of microbes like viruses and some other pathogens, which are hard to grow in culture. Further, culturing of microbes is time consuming and too expensive. Nucleic acid based diagnostic methods, for example, Polymerase Chain Reaction (PCR), have progressively replaced the culturing method. PCR-based method is cost effective, sensitive and specific but the main limitation is that it demands prior hypothetical knowledge. It can only identify the conserved targets of the pathogen and cannot distinguish between genotypes; this again restricts the detection of new emerging pathogens (Lecuit and Eloit 2014). Other methods for diagnosis were also developed, which include multiplex PCR assays, enzyme-linked immunoassays and pulse field gel electrophoresis (PFGE) to widen the pathogen detection sensitivity and specificity. But, these conventional methods also do not prove to be much useful because of insensitivity for clinical diagnosis.

On the other hand, NGS, which has revolutionized the diagnostic field, is different from other diagnostic assays because it does not require prior hypothetical knowledge. And now it has been commonly used to diagnose and discover the novel pathogens, for example, bacteria, fungi, virus or parasite (Frey et al. 2014). The field of “diagnosis genomics” or “pathogenomics” has unveiled emerging and re-emerging pathogens that translate the genomic technologies into methods for diagnosis. This new method has facilitated high-resolution mapping of the genetic

determinants in microorganisms that uphold the pathogenicity. Whole genome sequencing and targeted amplicon sequencing of rRNA genes have come up as favored technologies for microbial identification from primary human specimens and to analyze the dynamic genomes with a high moldability that is a must for pathogens to cling to life in arduous environment (Edwards and Rohwer 2005; Weinstock 2012). Analysis of genomes using whole genome sequencing delivers high-resolution information to differentiate microbial strains that possess difference of as low as one Single Nucleotide Polymorphisms (SNP), thereby it can replace other multiple tests. For the detection of pathogenic microorganism from patients having suspected infections from uncultivable microbe and or not possible to diagnose by standard diagnostic method, whole genome sequencing method is an ideal option. In such cases, specimen directly from the patient can be sequenced. Here is an example of the strong ability of the whole genome sequencing method in diagnosis of a pediatric instance of alymphocytosis or severe combined immunodeficiency (SCID) and recurrent meningoencephalitis, where this whole genome sequencing together with a dedicated bioinformatics software (Naccache et al. 2014) diagnosed the *Leptospirasantarosai* in cerebrospinal fluid (CSF) (total DNA from CSF) using Illumina platform within 48 hours' time from the sample collection, where 475 sequence reads were aligned with the pathogen out of 3,063,784 total reads (Wilson et al. 2014). Similarly, there are a number of cases where whole genome sequencing diagnosed the pathogens from the uncultivable sample in patients having infectious syndrome of uncertain cause.

For the analysis of heterogeneity or identification of microbial species in a given medical sample, targeted amplicon sequencing of 16S rRNA genes by NGS is also a method of choice. Several studies have been performed in relation with the diagnosis of patient sample as well as healthy sample to identify the heterogeneity among the patient. Because from the studies it is now known that during progression of disease, bacterial diversity decreases, which may be due to the increased antibiotic exposure to the pathogen in the patients (Morgan et al. 2012).

Further, RNA sequencing technology, which is in its infancy, also holds the promise for its applicability in diagnosis of infectious disease causing organism. Based on the performed studies as of now, this is clear that next-generation sequencing is progressively working on its way for routine diagnostic purpose in the foundation for public health and clinical laboratories. The approach is applicable for diagnosis of all kinds of microorganisms (virus, fungi, bacteria and or parasite/eukaryotic organism), which participate in the infection process. But there are a number of challenges that still obstruct the widespread use of next-generation sequencing in the diagnosis of infectious disease. First and foremost, requirement is the development and improvement in the software that is required for the analysis of the sequencing data. However, now there are several open source pipelines for the diagnosis of pathogen by NGS available. But the limitation is again that they require a substantial knowledge of bioinformatics/bioinformatics expertise that is generally not available in clinical health laboratories.

### 15.8.4 Determination/Investigation of Zoonotic Microbes Transmission to Humans from Animals

Zoonotic diseases are a threat to public health, which is only recently acknowledged though the transmission of zoonotic microbial (infections that animals spread to humans) agents is on rise from last so many decades. From the studies it has been shown that since the Second World War, annually one pathogenic disease comes to light globally and developing countries like India significantly carry the ball.

A research group has reported that three out of four emerging pathogens to human are of zoonotic origin (Taylor et al. 2001). It has recently become apparent that zoonotic diseases involve worldwide devastating diseases like Ebola virus infection, bird flu (highly pathogenic avian influenza), severe acute respiratory syndrome (SARS) and so many other (Heymann and Dar 2014). These newly apparent threats for public health are linked with considerable economic cost, which includes direct or indirect impact on our healthcare system.

Next-generation sequencing (NGS) is revealing more understanding on the transmission of zoonotic microbes and the method of choice nowadays (Chatterjee et al. 2017). Earlier studies in this area were done by serotyping (Tenover et al. 1997). Recently, studies are based on newer methods like pulse field gel electrophoresis (PFGE) or multilocus variable number tandem repeat (MLVNTR) analysis to identify the specific species of microbe in animals and humans (Sabat et al. 2013). Nonetheless, still so much is left to understood, mainly frequency of transmission (number of contacts needed for the transmission from animals), risk factor linked with the acquiring of zoonotic agent and how the transfer of pathogen from animal to human is affected by antibiotic use in animal. For all these topics, NGS shows a newer perspective. This is also unraveling the difference in previously indistinguishable strains of human and animal (Harrison et al. 2013). Besides, NGS also permits a wide range analysis of how the use of antibiotics changes the specific microbiome and effects on interspecies transmission. Current projects are based on whole genome sequencing methods that will help to understand the dynamics and mechanism of transmission of pathogen between animals, human and environment. A recent study by a veterinary research group has revealed the presence of *mcr-I* gene (using whole genome sequencing method) in three *E. coli* strains isolated from poultry meat. Although this gene is not carried by any human strain, two of the three strains were related to ST117 (avian pathogenic *E. coli*), which is a common strain between human and poultry animal. This represents a potential concern in public health (Kluytmans-van den Bergh et al. 2016).

According to a latest published study in 2018, the next-generation sequencing technique has been used to diagnosed the neurobrucellosis from cerebrospinal fluid (Fan et al. 2018). Neurobrucellosis is the condition of brucellosis where central nervous system gets involved and it is very common zoonotic disease globally. Its diagnosis is challenging with clinical indications it shows because of their non-specificity and low sensitivity of routine culture test. This group has shown the command of NGS for diagnosis from CSF together with bioinformatics analysis.

Using whole genome sequencing on Illumina HiSeq platform, molecular epidemiology of the related isolates of methicillin-resistant *Staphylococcus aureus* (MRSA) has been investigated. Also, it was revealed that human and animal isolates of the same farm were only varied by few number of Single Nucleotide Polymorphisms (SNP). This analysis supports the possibility of zoonotic transmission of MRSA isolates. The study further shows that *mecC*-MRSA ST130 isolates can transmit between animal and human (Harrison et al. 2013). And this *mecC* gene is liable for the resistance to the penicillin like antibiotic methicillin. This study emphasizes the role of farm animal (livestock) as a likely reservoir of antibiotic resistant pathogen.

A bacterium, *Corynebacterium ulcerans*, that causes diphtheria-like infection in humans and present in pets was thought to work as reservoir for zoonotic transmission. Furthermore, the reports reveal that this new bacteria is now playing the lead role to cause diphtheria in past years in several economically developed countries. Here also, next-generation sequencing approach enables us to identify the novel virulence genes rapidly acquired by *C. ulcerans* and a putative pathogenicity island, which possesses the diphtheria toxin gene. This rapid acquisition of genes changes the virulence of the strain even in single round of zoonotic transmission. During genomewide sequencing/SNP profiling of pair of patient and domestic animal companion, it was revealed that there is very less or almost no difference between their profiles. This supports the idea that *C. ulcerans* encounter zoonotic transmission between human and animal. In addition to this, these results demonstrate that NGS helps in improving the phylogenetic and epidemiological studies by giving insights between closely relative isolates.

---

## 15.9 Future Perspectives

Next-generation sequencing is making possible the term “One test fits for all” as clinical laboratories, public health laboratories and researchers are progressively embracing it. Further, this technology has revolutionized each and every field of medical science and life sciences imparting numerous benefits in terms of massive parallel sequencing. Earlier, high cost of sequencing was also the barrier and now the reduced cost by several folds is enormously attracting the researchers and making it feasible to plan their research based on sequencing.

In time to come, sequencing of individual genomes of importance living in different conditions, having different nutritional intake and/or under different treatment conditions will pave the way for disease control and its prevention thereby facilitates the social security. It is thought that genome sequencing of livestock will enable scientific community for more precise identification of genetic markers of important traits. However, the information obtained from NGS data will also bring advantage to the medical field with respect to better diagnostics and therapeutics. Also, the data generated from NGS will help agricultural society to improve the breed of dairy cattle and beef cattle. Earlier, the sequencing of animal was used to be done for its use as model system to study the human/public health issues.

Further, sequencing of human microbiomes and parasites of agricultural animals can benefit in the development of therapeutics and new vaccines for social welfare. Cell-free fetus DNA sequencing (which simply requires only blood sample of six-week pregnant mother) is showing the new opportunities for prenatal diagnosis, which ends the risk for the fetus.

In the end, these rapidly growing new technologies are accelerating the process of drug discovery and personalized medicine for public health welfare. Now, it is not surprising to say that only our imagination can put the limits of what is possible to be done by next-generation sequencing technologies!

---

## References

- Adessi C, Matton G, Ayala G et al (2000) Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Res* 28:E87
- Albiger B, Leitmeyer K, Struelens M et al (2016) European Centre for Disease Prevention and Control. ECDC roadmap for integration of molecular and genomic typing into European-level surveillance and epidemic preparedness, Version 2.1, 2016–19. ECDC, Stockholm
- Anderson S (1981) Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Res* 9:3015–3027
- Ansorge W (2016) Next generation DNA sequencing (II): techniques, applications. *Next Gene Seq App*. <https://doi.org/10.4172/2469-9853.1000S1-005>
- Arumugam M, Raes J, Pelletier E et al (2011) Enterotypes of the human gut microbiome. *Nature* 473:174–180
- Bahassi el M, Stambrook PJ (2014) Next-generation sequencing technologies: breaking the sound barrier of human genetics. *Mutagenesis* 29:303–310
- Ballester LY, Luthra R, Kanagal-Shamanna R, Singh RR (2016) Advances in clinical next-generation sequencing: target enrichment and sequencing technologies. *Expert Rev Mol Diagn* 16:357–372
- Bentley DR, Balasubramanian S, Swerdlow HP et al (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53–59
- Bowers J, Mitchell J, Beer E et al (2009) Virtual terminator nucleotides for next-generation DNA sequencing. *Nat Methods* 6:593–595
- Braslavsky I, Hebert B, Kartalov E, Quake SR (2003) Sequence information can be obtained from single DNA molecules. *Proc Natl Acad Sci U S A* 100:3960–3964
- Buermans HP, den Dunnen JT (2014) Next generation sequencing technology: advances and applications. *Biochim Biophys Acta* 1842:1932–1941
- Castro-Wallace SL, Chiu CY, John KK et al (2017) Nanopore DNA sequencing and genome assembly on the International Space Station. *Sci Rep* 7:18022
- Chatterjee P, Sarma N, Hansda S (2017) Tropical diseases on insurgence: clinician's perspective. *Indian J Dermatol* 62:468–477
- Edwards RA, Rohwer F (2005) Viral metagenomics. *Nat Rev Microbiol* 3:504–510
- Eid J, Fehr A, Gray J et al (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138
- Fan S, Qiao X, Liu L et al (2018) Next-generation sequencing of cerebrospinal fluid for the diagnosis of neurocysticercosis. *Front Neurol* 9:471
- Fang G, Munera D, Friedman DI et al (2012) Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat Biotechnol* 30:1232–1239
- Fleischmann RD, Adams MD, White O et al (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496–512



- Flusberg BA, Webster DR, Lee JH et al (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 7:461–465
- Frey KG, Herrera-Galeano JE, Redden CL et al (2014) Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *BMC Genomics* 15:96
- Goffeau A, Barrell BG, Bussey H et al (1996) Life with 6000 genes. *Science* 274(546):563–567
- Goto Y, Yanagi I, Matsui K, Yokoi T, Takeda K (2016) Integrated solid-state nanopore platform for nanopore fabrication via dielectric breakdown, DNA-speed deceleration and noise reduction. *Sci Rep* 6:31324
- Gullapalli RR, Desai KV, Santana-Santos L, Kant JA, Becich MJ (2012) Next generation sequencing in clinical medicine: challenges and lessons for pathology and biomedical informatics. *J Pathol Inform* 3:40
- Guo J, Xu N, Li Z et al (2008) Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proc Natl Acad Sci U S A* 105:9145–9150
- Guo J, Yu L, Turro NJ, Ju J (2010) An integrated system for DNA sequencing by synthesis using novel nucleotide analogues. *Acc Chem Res* 43:551–563
- Harris TD, Buzby PR, Babcock H et al (2008) Single-molecule DNA sequencing of a viral genome. *Science* 320:106–109
- Harris SR, Feil EJ, Holden MT et al (2010) Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327:469–474
- Harrison EM, Paterson GK, Holden MT et al (2013) Whole genome sequencing identifies zoonotic transmission of MRSA isolates with the novel *mecA* homologue *mecC*. *EMBO Mol Med* 5:509–515
- Heymann DL, Dar OA (2014) Prevention is better than cure for emerging infectious diseases. *BMJ* 348:g1499
- Ho A, Murphy M, Wilson S, Atlas SR, Edwards JS (2011) Sequencing by ligation variation with endonuclease V digestion and deoxyinosine-containing query oligonucleotides. *BMC Genomics* 12:598
- Jain M, Koren S, Miga KH et al (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* 36:338–345
- Karamohamed S, Nilsson J, Nourizad K, Ronaghi M, Pettersson B, Nyren P (1999) Production, purification, and luminometric analysis of recombinant *Saccharomyces cerevisiae* MET3 adenosine triphosphate sulfurylase expressed in *Escherichia coli*. *Protein Expr Purif* 15:381–388
- Kim Y, Koh I, Rho M (2015) Deciphering the human microbiome using next-generation sequencing data and bioinformatics approaches. *Methods* 79–80:52–59
- King LA, Nogareda F, Weill FX et al (2012) Outbreak of Shiga toxin-producing *Escherichia coli* O104:H4 associated with organic fenugreek sprouts, France, June 2011. *Clin Infect Dis* 54:1588–1594
- Kluytmans-van den Bergh MF, Huizinga P, Bonten MJ et al (2016) Presence of *mcr-1*-positive Enterobacteriaceae in retail chicken meat but not in humans in the Netherlands since 2009. *Euro Surveill* 21:30149
- Korlach J, Bjornson KP, Chaudhuri BP et al (2010) Real-time DNA sequencing from single polymerase molecules. *Methods Enzymol* 472:431–455
- Krause J, Fu Q, Good JM et al (2010) The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464:894–897
- Lecuit M, Eloit M (2014) The diagnosis of infectious diseases by whole genome next generation sequencing: a new era is opening. *Front Cell Infect Microbiol* 4:25
- Leggett RM, Clark MD (2017) A world of opportunities with nanopore sequencing. *J Exp Bot* 68:5419–5429
- Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW (2003) Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 299:682–686

- Lewis T, Loman NJ, Bingle L et al (2010) High-throughput whole-genome sequencing to dissect the epidemiology of *Acinetobacter baumannii* isolates from a hospital outbreak. *J Hosp Infect* 75:37–41
- Li Z, Bai X, Ruparel H, Kim S, Turro NJ, Ju J (2003) A photocleavable fluorescent nucleotide for DNA sequencing and analysis. *Proc Natl Acad Sci U S A* 100:414–419
- Mardis ER (2008) Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 9:387–402
- Mardis ER (2011) A decade's perspective on DNA sequencing technology. *Nature* 470:198–203
- Mardis ER (2013) Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto, Calif)* 6:287–303
- Margulies M, Egholm M, Altman WE et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380
- Maxam AM, Gilbert W (1977) A new method for sequencing DNA. *Proc Natl Acad Sci U S A* 74:560–564
- Metzker ML (2010) Sequencing technologies - the next generation. *Nat Rev Genet* 11:31–46
- Mignardi M, Nilsson M (2014) Fourth-generation sequencing in the cell and the clinic. *Genome Med* 6:31
- Morey M, Fernandez-Marmiesse A, Castineiras D, Fraga JM, Couce ML, Cocho JA (2013) A glimpse into past, present, and future DNA sequencing. *Mol Genet Metab* 110:3–24
- Morgan XC, Tickle TL, Sokol H et al (2012) Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 13:R79
- Naccache SN, Federman S, Veeraraghavan N et al (2014) A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res* 24:1180–1192
- Nakamura K, Oshima T, Morimoto T et al (2011) Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res* 39:e90
- Obenrader S (2003) The sanger method, Dept. of biology, Davidson college. Retrieved from [http://www.bio.davidson.edu/Courses/Molbio/MolStudents/spring2003/Obenrader/sanger\\_method\\_page.htm](http://www.bio.davidson.edu/Courses/Molbio/MolStudents/spring2003/Obenrader/sanger_method_page.htm)
- Qin J, Li R, Raes J et al (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59–65
- Quail MA, Smith M, Coupland P et al (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13:341
- Quick J, Grubaugh ND, Pullan ST et al (2017) Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 12:1261–1276
- Rand AC, Jain M, Eizenga JM et al (2017) Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods* 14:411–413
- Reuter JA, Spacek DV, Snyder MP (2015) High-throughput sequencing technologies. *Mol Cell* 58:586–597
- Roberts RJ, Carneiro MO, Schatz MC (2013) The advantages of SMRT sequencing. *Genome Biol* 14:405
- Ronaghi M (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Res* 11:3–11
- Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P (1996) Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* 242:84–89
- Ronaghi M, Uhlen M, Nyren P (1998) A sequencing method based on real-time pyrophosphate. *Science* 281(363):365
- Sabat AJ, Budimir A, Nashev D et al (2013) Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveill* 18:20380
- Sanger F, Air GM, Barrell BG et al (1977a) Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 265:687–695
- Sanger F, Nicklen S, Coulson AR (1977b) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74:5463–5467

- Smith LM, Sanders JZ, Kaiser RJ et al (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* 321:674–679
- Snitkin ES, Zelazny AM, Thomas PJ et al (2012) Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Sci Transl Med* 4:148ra116
- Stahl PL, Salmen F, Vickovic S et al (2016) Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353:78–82
- Taylor LH, Latham SM, Woolhouse ME (2001) Risk factors for human disease emergence. *Philos Trans R Soc Lond Ser B Biol Sci* 356:983–989
- Tenover FC, Arbeit RD, Goering RV (1997) How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. Molecular Typing Working Group of the Society for Healthcare Epidemiology of America. *Infect Control Hosp Epidemiol* 18:426–439
- Thompson JF, Steinmann KE (2010) Single molecule sequencing with a HeliScope genetic analysis system. *Curr Protoc Mol Biol* 7:Unit7 10
- Travers KJ, Chin CS, Turner SW et al (2010) A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res* 38(15):e159. <https://doi.org/10.1093/nar/gkq543>
- Uemura S, Aitken CE, Korlach J, Flusberg BA, Turner SW, Puglisi JD (2010) Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature* 464:1012–1017
- Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. *Science* 291:1304–1351
- Voelkerding KV, Dames SA, Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. *Clin Chem* 55:641–658
- Weinstock GM (2012) Genomic approaches to studying the human microbiota. *Nature* 489:250–256
- Wilson MR, Naccache SN, Samayoa E et al (2014) Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N Engl J Med* 370:2408–2417
- Zhou K, Lokate M, Deurenberg RH et al (2016) Use of whole-genome sequencing to trace, control and characterize the regional expansion of extended-spectrum beta-lactamase producing ST15 *Klebsiella pneumoniae*. *Sci Rep* 6:20840