



PROGRESS IN DNA SEQUENCING

*^{1&2} **Bamanga R. A., ²Ja'afar J. N. and ²Gali A. I.**

¹Abertay University, Dundee, United Kingdom

²Department of Biotechnology, School of Life Sciences, Modibbo Adama University of Technology, P.M.B. 2076
Yola, Adamawa State

*Correspondence: rbamanga@mautech.edu.ng +2348184226759

ABSTRACT

The first human genome sequence took about a decade to complete and cost more than two billion dollars. This shows the major limitations of time and cost, and the development of recent technologies for DNA sequencing ultimately aimed at reducing these two factors. The major milestone of the HGP was the sequencing of the first billionth base out of the three billion base pair human genome. However, depending on the platform used in sequencing, the cost has drastically plummeted to about five thousand dollars and this is the work of a single day. The ultimate target of the HGP is to reach a one thousand dollar price mark to sequencing an entire human genome with the highest throughput, and this is slowly but steadily approaching, thanks to the refinements of existing methods, which are reducing the cost per base by the day. This review looks at the advancement of the DNA sequencing methods from the standard Sanger method, through to those applied in today's research and also focuses on the technologies that have evolved throughout the past three decades with a possible comparison between them and finally a look at some of the limitations of these technologies.

Keywords: Human genome project, DNA sequencing, Sanger method.

INTRODUCTION

Since the proposal of the three-dimensional double-stranded helical model and conformation of the DNA molecule by Watson and Crick in 1953, it is known that the genetic information of living things is hidden in the sequence and arrangement of four bases – adenine, guanine, cytosine and thymine (A, G, C and T respectively) (Merrill and Mazza, 2006). Interestingly enough, Watson and Crick stated in their work that the determination of the entire human genomic sequence would take centuries to complete, but here we are half a century later with the Human Genome Project (HGP) completed in 2003. However, the knowledge of the genes are yet to be fully understood. The actual sequencing began in 1977 with Fredrick Sanger when he developed an efficient method for the determination of the entire sequence of some small viral genomes (Sanger *et al.*, 1977). This led to a greater focus on the molecular genetics paradigm.

For a long time, how to sequence the genome has been only built on fragments sequencing. That is, the genomic DNA is sheared into small fragments, sequenced separately by different kinds of techniques and aligned by software (Zhao, 2010). Generally, the techniques for genome sequencing can be divided into two groups. The first is the traditional techniques, known as hierarchical and shotgun sequencing (Lander *et al.*, 2001) based on fragmentation and reassembly. It involves rearranging whole chromosome sequence from the sequences of several fragments each ranging between 500 and 1000bp in length (Gibson and Spencer, 2004). The second is the techniques invented during recent years and are referred to as "the second generation." They employ massively parallel approaches in sequencing several hundred thousand to millions of reads simultaneously

(Chan, 2009). They are able to sequence more quickly and at lower cost in terms of both money and labour (Smith *et al.*, 2008). These technologies include Solexa single molecule array (SMA) technology (Illumina), 454 genome sequencer system (Roche/GS), Accurate multiplex polony sequencing (SOLID), Sanger hybrid approach, fast DNA sequencing via transverse electronic transport (Nanopore) and the latest method called diploid genome reconstruction containing a key step called Gibbs sampling (Zhao, 2010). The use of these methods for SNP discovery is by far the most demanding application since it requires a high accuracy (Chan, 2009).

DNA sequencing is the determination of the number and arrangement of the nucleotide bases in the strands of the DNA molecule. The importance of DNA sequencing can be seen in the determination of the causes of diseases. For example, some human diseases can be caused by a silent mutation at a splice site, which implies that the resulting amino acid translated and coded for is the same but the spliceosome does not recognise the splice site because the nucleotide has changed on the DNA, leaving the intron not spliced out causing a defective protein such as in the case of the McArdle disease (Fernandez-Cadenas *et al.* 2003). A point mutation can also be detected using DNA sequencing methods to determine the genetic implications of other diseases like the glycogen storage deficiency type IV (GDS-IV) caused by a mutation in the glycogen branching enzyme (GBE) (Bao *et al.*, 1996), which leads to the classical and progressive form, liver cirrhosis. The missense and nonsense mutation in the skeletal muscle alpha-actin gene (ACTA 1) was also identified using the DNA sequencing methods (Sparrow *et al.*, 2003).

Another important application of DNA sequencing is seen today in forensic science with television shows like the popular Crime Scene Investigation (CSI) based on it helping the scientists solve crimes by the determination of the nucleotide sequence of a suspect's DNA sample and comparing it with the DNA sample retrieved from the crime scene. However, forensic science mostly uses an extra-nuclear genome, the mitochondrial DNA (mtDNA) (Wilson *et al.*, 1995) because of its desirable features such as high copy number, lack of recombination, and matrilineal inheritance, and used to analyse old bones, teeth, hair shafts, and other biological samples where nuclear DNA content is low (Budowle *et al.*, 2003).

The traditional (Sanger's chain termination) sequencing method

The principle involves DNA synthesis using deoxynucleotides (dNTPs) and dideoxynucleotides (ddNTPs) that result in the termination of the synthesis at specific nucleotides (Figure 1). The

process requires a primer, DNA polymerase, a template, a mixture of nucleotides and a detection system. Growing chains are terminated by the incorporation of the ddNTPs which do not permit chain elongation after their integration. The synthesised strand sizes are determined for each ddNTP using gel or capillary electrophoresis. The reason for the chain termination is that the 3'-OH group required for the formation of the phosphodiester bond is missing in the ddNTPs. Innis *et al.* (1988) supported this by stating that: "the basic procedure involves (i) hybridizing an oligonucleotide primer to a suitable single- or denatured double-stranded DNA template; (ii) extending the primer with DNA polymerase in four separate reaction mixtures, each containing one a labelled dNTP, a mixture of unlabeled dNTPs, and one chain-terminating ddNTP; (iii) resolving the four sets of reaction products on a high-resolution polyacrylamide/urea gel; and (iv) producing an autoradiographic image of the gel, which can be examined to infer the DNA sequence".

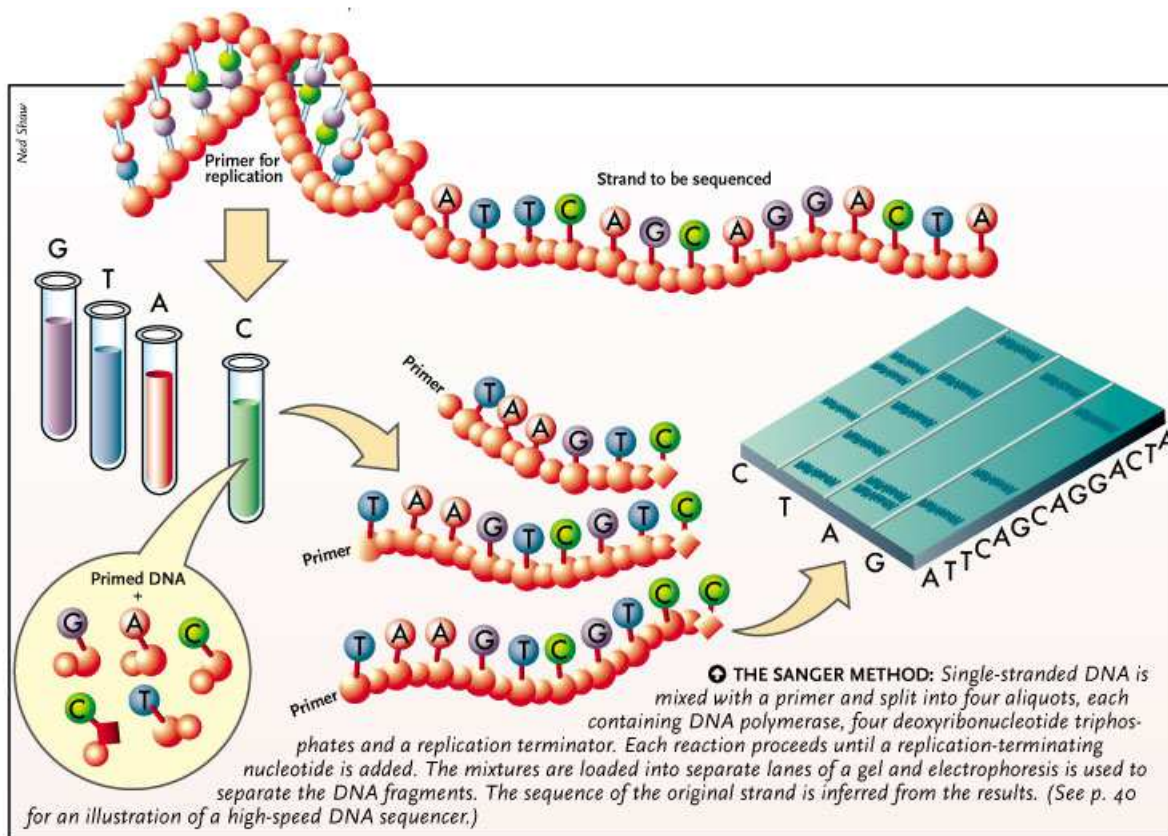


Figure 1: A representation of the Sanger chain termination sequencing method extracted. Source: Winnick (2004).

This is the standard and the most common method for the determination of the nucleotide sequence although modified in recent times. As stated by Misuzawa *et al.* (1986), using deoxy-7-deazaguanosine triphosphate (dc7GTP) in place of dGTP was found to be very useful because sequencing of a part of the human N-myc gene having 85% GC content is impossible by the original method using dGTP, because of compression of bands. Polymerase chain reaction (PCR) is a fundamental technique in

this process. Previously, the PCR procedure was very complicated because of the necessity to replenish the enzyme after each PCR cycle. This has been simplified, however, by the advent of the Taq polymerase, the highly thermostable DNA polymerase from *Thermus aquaticus*, which is ideal for both manual and automated DNA sequencing because it is fast, highly processed, has little or no 3'-exonuclease activity, and is active over a broad range of temperatures (Innis *et al.*, 1988).

From the Sanger method in 1977 until date, some innovative technologies have emerged with a great throughput in achieving the ultimate goal of the HGP. Some of these technologies are the pyrosequencing, reverse termination and sequencing by ligation. In order to further elaborate these three methods, an extensive review of Pettersson *et al* (2009) was done in which the paper described the principle involved in the methods, the instrumentation used in achieving the sequences and some of the limitations encountered.

**The second generation sequencing principles
Pyrosequencing**

Pyrosequencing is based on sequencing by synthesis, taking advantage the release of pyrophosphate whenever a nucleotide is incorporated in an open 3'-OH group of the DNA strand. This pyrophosphate is

then involved in a reaction that results in the release of ATP, which is used by luciferase which converts luciferin to oxyluciferin resulting in the emission of light which is collected and recorded as peaks (Figure 2). Pyrosequencing is a sequencing-by-synthesis method where a successful nucleotide incorporation event is detected as emitted photons (Ronaghi *et al.*, 1998). The addition of a nucleotide by a polymerase releases a pyrophosphate (ppi), which catalysed by ATP sulfurylase forms ATP by the use of adenosine phosphosulphate (APS). Subsequently, the enzyme luciferase (together with D-luciferin and oxygen) uses the newly formed ATP to produce light. Another enzyme, apyrase, is used for the degradation of unincorporated dNTPs as well as to stop the reaction by degrading ATP.

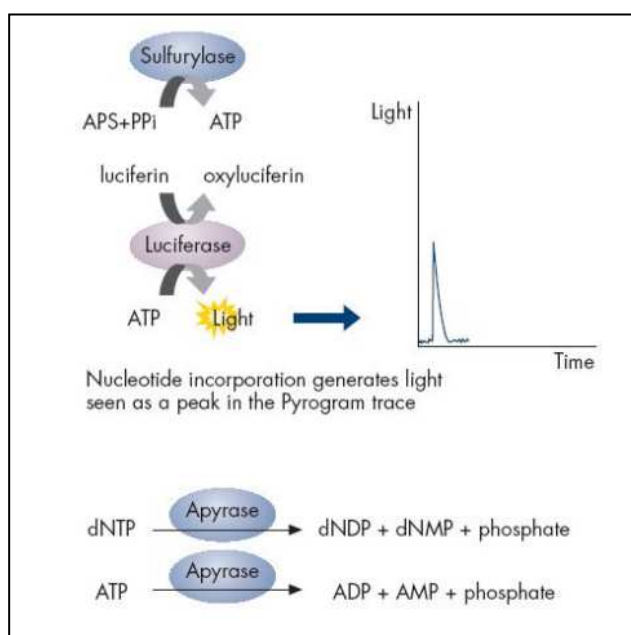


Figure 2: An illustration of the basic principles employed in pyrosequencing accessed from Qiagen sample and assay technologies. Source: Qiagen (2016)

Pettersson *et al* (2009) reported that the Genome Sequencer FLX developed by 454 Life Sciences and Roche, works as follows. The enzymes luciferase and ATP sulfurylase are immobilised on smaller beads surrounding the larger amplicon carrying beads; these larger amplicon carrying beads are products of an oil-aqueous based emulsion PCR carried out on a primer-coated bead in a PicoTiterPlate. All other reagents are supplied through a flow allowing reagents to diffuse to the templates in the PicoTiterPlate. Polymerase and one exclusive dNTP per cycle generate one or more incorporation events and the emitted light is proportional to the number of incorporated nucleotides (Pettersson *et al.*, 2009). The emitted light is detected by a CCD camera and after each round; the enzyme apyrase is flowed through in order to degrade excess nucleotides. They also stated that the washing procedure permits read lengths of over 400 bp. The limitation of this system however as explained by the paper, is that the incorporation of nucleotides in each cycle is not 100% complete the population of nucleotides that is not fully degraded by

the apyrase and can, therefore, be incorporated after the next nucleotide, and this will generate high levels of noise.

The pyrosequencing method can be used for single-nucleotide polymorphism (SNP) analysis and tag sequencing (up to 100 bases), as well as for whole-genome sequencing (Margulies *et al.*, 2005). Other applications of the method include analysis of allele frequency in pooled samples (Gruber *et al.* 2002), methylation analyses (Uhlmann *et al.*, 2002), molecular haplotyping (Ahmadian *et al.* 2000), sequencing of heteroplasmic DNA (Gorriely *et al.*, 2003) and forensic analyses (Andreasson *et al.*, 2002). In addition, insertion/deletions (indels), short tandem repeats, human leukocyte antigen (HLA) typing, gene copy number, allelic imbalance in RNA, and short sequencing stretches are among the numerous applications (Marsh, 2007). The assay is applicable to almost any source of DNA or RNA (e.g., blood, saliva, cell line, plasma, serum, tissue, formalin fixed and/or paraffin-embedded samples, and whole-genome-amplified DNA) (Marsh, 2007).

Reverse termination

In this method, the overall process involves fragmentation of genomic DNA followed by the addition of short adapters to the DNA fragments, then immobilization of these now modified DNA fragments to a flow cell, then bridge amplification to produce millions of distinct DNA clusters, then comes the important step of base-by-base sequencing-by-synthesis incorporation using fluorescent reversible terminators, and analysis. Pettersson *et al* (2009) explains that the Illumina 1G Genome Analyzer depends on clonal bridge amplification on a flow cell surface designed to present the DNA in a manner that facilitates access to enzymes whilst ensuring high stability of surface-bound template and low non-specific binding of fluorescently labelled nucleotides thereby producing 10 million single molecule clusters per square centimetre. The report also suggests the DNA fragments of choice are bridge-amplified using a solid-phase PCR on a surface generating spatially separated colonies of approximately thousand fragments each. A cyclic sequence incorporation procedure is performed using fluorescently labelled, reversibly terminating nucleotides. All four bases are added in each cycle and following incorporation and stringent washing procedures, the colour of each colony is detected. The dye is then removed and the termination reversed allowing for interrogation of the following base in each colony. However, the limitation of this process is that at 30–35 bases, the error rate is becoming high thereby limiting the read length.

Sequencing by ligation

This process is different from the above two and the Sanger method in that it uses a DNA ligase to identify specific nucleotides present in a given region on the DNA strand, and it does not require a DNA polymerase to synthesise a second strand. The process flow as described by Pettersson *et al* (2009) involves signature sequencing carried out by a restriction enzyme which mediates exposure of four nucleotides in each cycle followed by ligation of a probe and this process is repeated for 4–5 cycles, i.e., probing 16–20 bases in total. Then the use of 16 probes, each with a unique decoder binding site, to enable single dye detection.

The principle initiated the development of the commercial SOLiD system (Sequencing by Oligonucleotide Ligation and Detection) from Applied Biosystems, where clonal amplicons on 1 µm beads are generated by an emulsion PCR (similar to the Genome Sequencer FLX PCR), either from fragments (restriction enzyme products) or mate-paired libraries, the beads are then enriched, so that 80% of them generate signals, and attach on a glass surface to produce a very high-density random array. Sequencing-by-ligation is performed by ligating 3'-degenerated and 5'-labeled probes to the amplicons and detecting the colour (Pettersson *et al.*, 2009). Accuracy is improved by implementing a two-base encoding system that leads to an interrogation of each base twice. A sequencing run takes 6–10 days and the output is high, approximately 3–6 Gbp per run giving a read length of 25–35 bases per clonally amplified bead (Applied Biosystems).

In hierarchical sequencing method, the genomic DNA is sheared into pieces of about 150 kb and inserted into BAC vectors, then transformed into *E. coli* where they are replicated and stored. Each BAC fragment is isolated and fragmented randomly into smaller pieces and each of which is cloned into a plasmid and sequenced on both strands. These sequences are aligned so that identical sequences overlap. These contiguous pieces are then assembled into finished sequence once each strand is sequenced about four times to produce 8X coverage of data with high quality (Zhao, 2010). On the other hand, shotgun sequencing randomly shears genomic DNA into small pieces, which are cloned into plasmids and sequenced on both strands, thus eliminating the BAC step. The sequences obtained are aligned and assembled into a finished sequence (Venter *et al.*, 2001).

The second generation techniques include: Single Molecule Arrays (SMA)

In this method, a single tube reaction is carried out. Genomic DNA is fragmented and processed into ss-oligonucleotide fragments. Hundreds of millions of molecules are deposited and attached to discrete sites on a SMA. Complementary nucleotides base-pair to the first base of each oligonucleotide fragment and are added to the primer by the enzyme. Laser light excites the label on the incorporated nucleotides, and fluorescence is detected by a CCD camera that rapidly scans the entire array to identify the incorporated nucleotides on each fragment. This cycle of incorporation, detection and identification is repeated nearly 25 times to determine the first 25 bases in each fragment. These hundreds of millions of sequences are aligned and compared with the reference sequence (Smith, 2004).

454 Genome Sequencing

The genome is cut randomly into small fragments, denatured and adapters are added to each of them in order to link to the beads. One bead has its specific DNA fragment and then administered the emulsion-based PCR. PCR terminates at different sites to make every bead own different copies of the same template, in which PCR primers are biotinylated. The emulsion is washed, the DNA is denatured again, and then the beads are put with ssDNAs on a fiber-optic slide. Finally, sequencing is done by synthesis simultaneously in open wells of a fiber-optic slide using a modified pyrosequencing protocol that is designed to take advantage of the small scale of the wells, while the read length of this method reaches 100, which is somewhat larger than traditional pyrosequencing (Zhao, 2010).

Accurate Multiplex Polony Sequencing

It is a new DNA sequencing technology in which a commonly available, inexpensive epi-fluorescence microscope is converted to rapid non-electrophoretic DNA sequencing automation (Hutchison III, 2007). A cell-free, mate-paired library will provide the single DNA molecules to be amplified in parallel to 1µm beads by emulsion polymerase chain reaction. Millions of beads will be immobilised in a polyacrylamide gel and subjected to automated cycles of sequencing by ligation and four-color imaging (Zhao, 2010).

Sanger/Pyrosequencing Hybrid Approach

Combines both Sanger sequencing and 454 sequencing depending on the size of the genome of the organism. Initial 5X Sanger sequencing is carried out for all organism irrespective of size, then followed by 3X Sanger sequencing for genomes < 3mb, and two runs of 454 sequencing for those with > 3Mb genome (Zhao, 2010).

Nanopore Sequencing

Each DNA base is structurally and chemically different, thus creating its own distinct electronic signature. This method is based on the distributions of transverse electrical currents of single-stranded DNA while it translocates through a nanopore (Guiducci *et al.*, 2004). Previous attempts to sequence DNA using nanopores was not successful because the twisting and turning of the DNA strand introduced too much noise into the signal being recorded. This new method takes advantage of the electric field that drives the current perpendicular to the DNA strand to reduce the

structural fluctuations of DNA while it moves through the pore, thus minimising the noise (Ashkenasy *et al.*, 2005).

Diploid Genome Reconstruction

This method is based on the high genic variability in creatures. Other creatures with high genic variability are also suitable for this method, such as fish. The core step of this method is a program, called Gibbs sampling, which is often used for probabilistic inference, mainly for the copies of incomplete information. Kim *et al* (2007) found that this method could be used for sequencing those regions with high variability in homo-genome. However, this method is not suitable for human, because the variability of homo-genome is rather low comparatively (Zhao, 2010). On the other hand, this method confirmed that some short fragments of noncoding DNA were rather conservative, which offered an important and indirect evidence that junk DNA could have some unknown functions.

Table 1: Summary of Sequencing Methods

Sequencing Method	Merit(s)	Demerit(s)	Remark(s)
Hierarchical	Sequences are less likely to make mistakes when assembling the shotgun fragments into contigs as long as full chromosomes, because the chromosomal location for each BAC is known, and there are fewer random pieces to assemble (Zhao, 2010)	Time and expense (Green, 1997)	Reassembling of whole chromosome sequence from hundreds of thousands of fragments each typically between 500 and 1000 bp in length (Gibson and Spencer, 2004)
Shotgun	Faster and less expensive (Zhao, 2010)	More prone to errors due to incorrect assembly of finished sequence (Zhao, 2010)	Developed and preferred for prokaryotic genomes (because they are smaller in size and contain less repetitive DNA) (Venter <i>et al.</i> , 1998)
Single Molecule Array	Faster and cheaper than 454 sequencers (Smith <i>et al.</i> 2008); Possibility of creating arrays of very high density, around 10^8 sites/cm ² or more allowing massively parallel processing, because of only a single molecule at each site (Zhao, 2010)	Short read-outs are hard to re-assemble; PCR may introduce some copy errors and may be a barrier in decreasing the cost (van Oeveren and Janssen, 2009; Zhao, 2010); Produces very short reads as additional base calling at the end becomes very unreliable (Bentley, 2006)	Uses sequencing by synthesis, with DNA amplified on the surface of a flow cell resulting in a random array of dense clusters (Bentley, 2006); Has -5 to 40 as reliability index for its base call score; Developed to sequence clonally amplified single DNA molecules (Braslavsky <i>et al.</i> , 2003)
454 Genome Sequencing	Fast, ultra high throughput (up to 25 million bp/4hours) (Kaeller <i>et al.</i> , 2007)	Accuracy is 99%; Cannot resolve the problem of paired-end sequencing information; Has short read length that is not easily aligned without clustering (van Oeveren and Janssen, 2009; Zhao, 2010); Has a problem in accurately detecting the length of homopolymer regions (Margulies <i>et al.</i> , 2005)	Focuses on co-development of emulsion-based method to isolate and amplify DNA fragments <i>in vitro</i> , and of fabricated substrate and instrument that performs pyrosequencing in picoliter –sized wells (Thomas <i>et al.</i> , 2006; Margulies <i>et al.</i> , 2005; Dressmann <i>et al.</i> , 2003); Can be used for <i>de-novo</i> sequencing in addition to re-sequencing (Margulies <i>et al.</i> , 2005)
Polony Sequencing	Cost per base is one-ninth as much as that of conventional sequencing; Can read a slide with 10 million polonies in about 20 mins (Hall, 2007)	Single-pixel sequencing; Only for small genome-scale genome analysis (Shendure <i>et al.</i> , 2005)	Applied to re-sequence an evolved strain of <i>E.coli</i> at less than one error per million consensus bases (Dahl <i>et al.</i> , 2007)
Sanger/Hybrid Approach	Helpful for final draft genome in terms of coverage, reduction gaps, and poorly sequenced regions that degrade the value of an assembly, and leading to 25% reduction in cost compared to traditional Sanger only; Cheap and fast (Zhao, 2010)	Has low base call quality at the terminal ends of both 5' and 3' sides (Peterson <i>et al.</i> 2002; Ewing <i>et al.</i> 1998); Overall reduction of gaps is not sufficient to permit addition high throughput sequencing (Zhao, 2010)	The most cost effective for organisms with small genome size (<3 mb), and/or a small number of gaps and/or high levels of repetitive structure, including physical ends i.e using 8X Sanger sequencing (Zhao, 2010); Uses "Phred" score to assign a quality of base call score. Score of 20 is taken as reliable (van Oeveren and Janssen, 2009)
Nanopore Sequencing	Cheap and fast	Nanopore is yet to be made with the required configuration (Lagerqvist <i>et al.</i> , 2006)	Inventors used mathematical calculation and computer modelling of the motions and electrical fluctuations of DNA molecules to determine how to distinguish each of the four bases (Zwolak and Ventra, 2005)
Diploid Genome Reconstruction		Not suitable for human (Zhao, 2010)	

The past, present and future of DNA sequencing

With all these above-mentioned principles and technological innovations of DNA sequencing, we now look at the advancement in terms of the cost of sequencing. From the inception of the HGP, the initial sequencing and mapping of the human genome cost about \$3 billion, (Venter *et al.*, 2001; Lander *et al.*, 2001). The genome of Craig Venter determined in 2008 cost about \$70 million, sequencing a human genome using the Sanger method today would cost about \$10 million (Bentley, 2006), the 454 system sequences that at about \$1 million with a 20-fold reduction in time, while the Illumina is able to sequence a human genome with the 1G Analyzer for approximately \$100,000.

Since the ultimate goal of the HGP is to attain the \$1000 human genome mark, the future of DNA sequencing is being sought through an innovative technology called the Single Molecule Sequencing. This is a principle, in which a sequence by synthesis is done with the amplification step totally disregarded, and this has an advantage of drastically reducing the overall cost due a significant drop in reagent cost and there is an improved read length, since molecules are not getting out of phase (Pettersson *et al.*, 2009). Based on this process, a number of technological innovations have emerged with the goal of an increased throughput and reduced cost of sequencing two of them are described in Table 2.

Table 2: The description and possible comparison of two innovative technologies of DNA sequencing using the single molecule sequencing method showing the process workflow and the achievements.

Innovative technologies	The key workflow	Overall achievements
HeliScope™ by Helicos Bioscience	Single fragments are labelled with Cy3 for localisation of template strands on an array and a predefined, Cy5 labelled nucleotides are incorporated, detected by a fluorescent microscope and cleaved off in each cycle (Pettersson <i>et al.</i> , 2009).	Has a claimed throughput of 1.1 Gbp per day (Helicos Bioscience, 2016).
Single-molecule Real Time Sequencing-by-synthesis (SMRT™) by Pacific Biosciences	DNA sequencing technology enables, for the first time, the observation of natural DNA synthesis by a DNA polymerase as it occurs. The approach is based on eavesdropping on a single DNA polymerase molecule working in a continuous, processive manner (Pacific Biosciences, 2016).	Although it is still in the pipeline with no commercial instrument yet, it should be capable of generating 100 Gbp per hour (Pettersson <i>et al.</i> , 2009).

The SMRT™ in is based on a technology called the zero-mode wave (ZMW) (Pettersson *et al.* (2009) which are microfluidic wells of 70 nm in diameter. The ZMW contains a single polymerase macromolecule that is immobilised, with nucleotides that are fluorescently labelled at the triphosphates and a DNA strand which permits SMRT™ sequencing.

CONCLUSION

The HGP is very close to the target of the \$1000 genome sequence as seen by the recent advancement over the past three decades of the DNA sequencing methods dating back to when it first began with Fredrick Sanger in 1977. The journey so far has been achieved through certain improvements in technologies like the modification of the Sanger method which is based on the sequencing by synthesis (pyrosequencing and reverse termination methods), innovations in the sequencing by ligation process and a possible further research on the single molecule sequencing in the years to come. In conclusion, therefore, this review has described some of the key instrumentations of the different technologies involved in the DNA sequencing such as

the Genome Sequencer FLX developed by 454 Life Sciences and Roche, the Illumina 1G Genome Analyzer, and the commercial SOLiD system from Applied Biosystems with a brief on their throughput and cost of sequencing and some of their limitations. The future of the DNA sequencing and the HGP can ultimately stop asking the question of “can we get to the \$1000 mark” and start asking “what next” after that mark has been surpassed.

Author’s contributions

All authors contributed equally in the preparation and write-up of this review article.

Conflict of interest

There are no conflicts of interest among the authors in whatever form.

REFERENCES

Ahmadian, A., Lundeberg J., Nyrén, P., Uhlén, M and Ronaghi, M. (2000): Analysis of the p53 tumour suppressor gene by Pyrosequencing. *Biotechniques* 28:140-147.

Andreasson, H., Asp, A., Alderborn, A., Gyllensten, U and Allen, M. (2002): Mitochondrial sequence analysis for forensic identification using Pyrosequencing technology. *Biotechniques* 32:124-133.

- Ashkenasy N, Sanchez-Quesada, J., Bayley, H and Ghadiri, M. R. (2005): Recognising a single base in an individual DNA strand: a step toward DNA sequencing in nanopores. *Angew. Chem. Int. Ed. Engl.* 44(9):1401-1404.
- Bao, Y., Kishnani, P., Wu, J. Y. and Chen, Y. T. (1996): Hepatic and neuromuscular forms of glycogen storage disease type IV caused by mutations in the same glycogen-branching enzyme gene. *Journal of Clinical Investigations.* 97 (4), 921 – 948.
- Bentley, D. R. (2006): Whole-genome re-sequencing. *Current Opinion in Genetics & Development (Genomes and evolution)* 16(6):545-552
- Braslavsky, I. Hebert, B., Kartalov, E and Quake, S. R. (2003): Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Science* 100:3960-3964.
- Budowle, B., Allard, M. A., Wilson, M. R., and Chakraborty, R. (2003): Forensics and mitochondrial DNA: applications, debates and foundations. *Annual Review of Genomics and Human Genetics.* 4, 119 – 141.
- Chan, E. Y. (2009) Next-generation sequencing methods: impact of sequencing accuracy on SNP discovery. In: A.A. Komar (Editor) *Single Nucleotide Polymorphisms: Methods in Molecular Biology* Humana Press, New York, USA. 578:95-111.
- Dahl, F., Stenberg, J., Fredriksson, S., Welch, K., Zhang, M., Nilsson, M., Bicknell, D., Bodmer, W. F., Davis, R. W and Ji, H. (2007) Multi-gene amplification and massively parallel sequencing for cancer mutation discovery. *Proceedings of the National Academy of Science* 104:9387-9392.
- Dressman, D., Yan, H., Traverso, G., Kinzler, K. W and Vogelstein, B. (2003): Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Science* 100(15):8817-8822.
- Ewing, B., Hillier, L., Wendl, M. C and Green, P. (1998): Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research* 8:175–185.
- Fernandez-Cadenas, I., Andreu, A. L., Gamez, J., Gonzalo, R., Martin, M. A., Rubio, J. C., and Arenas, J. 2003. Splicing mosaic of the myophosphorylase gene due to a silent mutation in McArdle disease. *Neurology.* 61 (10), 1423 – 1434.
- Gibson, G and Spencer, V. (2004): *A primer of gene science.* Sinauer Associates, Incorporated, USA, pp.81-93.
- Goriely, A., McVean, G. A., Rojmyr, M., Ingemarsson, B and Wilkie, A. O. (2003): Evidence for selective advantage of pathogenic FGFR2 mutations in the male germ line. *Science* 301:643–646.
- Green, P. (1997): Against a whole-genome shotgun. *Genome Research* 7: 410-417.
- Gruber, J. D., Colligan, P. B and Wolford, J. K. (2002): Estimation of single nucleotide polymorphism allele frequency in DNA pools by using Pyrosequencing. *Hum. Genet.* 110:395-401.
- Guiducci, C., Stagni, C., Zuccheri, G., Bogliolo, A., Benini, L., Samorì, B and Riccò, B. (2004): DNA detection by integrable electronics. *Biosensors and Bioelectronics* 19(8):781-787.
- Hall, N. (2007): Advanced sequencing technologies and their wider impact in microbiology. *Journal of Experimental Biology.* 210:1518-1525.
- Helicos Bioscience Corporation. Available at www.helicosbio.com Accessed on 3rd August 2016.
- Hutchison III, C. A. (2007): DNA sequencing: bench to bedside and beyond. *Nucleic Acids Research.* 35(18):6227-6237.
- Innis, M. A., Myambo, B. K., Gelfand, D. H., and Brow, M. D. (1988): DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Procedure of National Academy of Science, Biochemistry* 85, 9463 – 9440.
- Kaeller, M., Lundeberg, J and Ahmadian, A. (2007): Arrayed identification of DNA signatures. *Expert Review of Molecular Diagnostics* 7(1):65- 76.
- Kim, J. H. Waterman, M. S and Li, L. M. (2007): Diploid genome reconstruction of *Ciona intestinalis* and comparative analysis with *Ciona savignyi*. *Genome Research.* 11:1101-1110.
- Lagerqvist, J., Zwolak, M and Di Ventra, M. (2006): Fast DNA sequencing via transverse electronic transport. *Nano Lett.* 6(4):779-782.
- Lander, E. Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczy, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T.,

- Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S and Chen, Y. J.(2001): Initial sequencing and analysis of the human genome (International human genome sequencing consortium) *Nature* 409:860-921.
- Margulies, M., Egholm, M., Altman, W., Attiya, S., Bader, J., Bemben, L., Berka, J., Braverman, M., Chen, Y., Chen, Z., Dewell, S., Du, L., Fierro, J., Gomes, X., Godwin, B., He, W., Helgesen, S., Ho, C., Irzyk, G., Jando, S., Alenquer, M., Jarvie, T., Jirage, K., Kim, J., Knight, J., Lanza, J., Leamon, J., Lefkowitz, S., Lei, M., Li, J., Lohman, K., Lu, H., Makhijani, V., McDade, K., McKenna, M., Myers, E., Nickerson, E., Nobile, J., Plant, R., Puc, B., Ronan, M., Roth, G., Sarkis, G., Simons, J., Simpson, J., Srinivasan, M., Tartaro, K., Tomasz, A., Vogt, K., Volkmer, G., Wang, S., Wang, Y., Weiner, M., Yu, P., Begley, R and Rothberg, J. (2005): Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057):376-380.
- Marsh, S (2007): Pyrosequencing® applications. In: S. Marsh (Editor). *Pyrosequencing® Protocols: Methods in Molecular Biology* Humana Press Inc, Totowa, New Jersey 373:15-23
- Merrill, S. A., and Mazza, A. M. (2006): *Reaping the benefits of genomic and proteomic research*. National academic press, Washington DC. USA.
- Misuzawa, S., Nishimura, S., and Seela, F. (1986): Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nucleic Acid Research*. 14 (3), 1319 – 1324.
- Pacific Biosciences. Available at www.pacificbiosciences.com Accessed on 3rd August, 2016.
- Peterson, D. G., Wessler, S. R and Paterson, A. H. (2002): Efficient capture of unique sequences from eukaryotic genomes. *Trends Genet*. 18:547–550.
- Pettersson, E., Ludeberg, J., and Ahmadian, A. (2009): Generations of sequencing technologies. *Genomics* 93, 105 – 111.
- Qiagen sample and assay technologies. Available at www.qiagen.com Accessed on 3rd August 2016.
- Ronaghi, M., Uhlen, M., and Nyren, P. (1998): A sequencing method based on real-time pyrophosphate. *Science*. 281, 363 – 365.
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977): DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Science* 74 (12), 5463 – 5467.
- Shendure, J., Gregory J. Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M., Wang, M. D., Zhang, K., Mitra, R. D and Church, G. M. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309(5741):1728-1732.
- Smith, A. D., Xuan, Z and Zhang, M. Q. (2008) Using quality scores and longer reads improves accuracy of solexa read mapping. *BMC Bioinformatics* 9:128
- Smith, T. (2004), Whole genome variation analysis using single molecule sequencing. *Drug Discovery Today: Targets* 3(3):112-116.
- Sparrow, J. C., Nowak, K. J., Durling, H. J., Beggs, A. H., Wallgren-Pettersson, C., Romero, N., Nonaka, I., and Laing, N. G. (2003): Muscle disease caused by mutations in the skeletal muscle alpha-actin gene (*ACTA 1*). *Neuromuscular disorder* 13, 519 – 531.
- Thomas, R. K. Nickerson, E., Simons, J. F., Jänne, P. A., Tengs, T., Yuza, Y., Garraway, L. A., LaFramboise, T., Lee, J. C., Shah, K., O'Neill, K., Sasaki, H., Lindeman, N., Wong, K., Borrás, A. M., Gutmann, E. J., Dragnev, K. H., DeBiasi, R., Chen, T., Glatt, K. A.,

Bajopas Volume 11 Number 1 June, 2018

- Greulich, H., Desany, B., Lubeski, C. K., Brockman, W., Alvarez, P., Hutchison, S. K., Leamon, J. H., Ronan, M. T., Turenchalk, G. S., Egholm, M., Sellers, W. R., Rothberg, J. M and Meyerson, M. (2006): Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picolitre reactor sequencing. *Nature Medicine* 12(7):852-855.
- Uhlmann, K., Brinckmann, A., Toliat, M. R., Ritter, H and Nurnberg, P. (2002) Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. *Electrophoresis* 23:4072-4079
- van Oeveren, J and Janssen, A (2009): Mining SNPs from DNA sequence data: Computational approaches to SNP discovery and analysis. In: A.A. Komar (Editor) *Single Nucleotide Polymorphisms: Methods in Molecular Biology* Humana Press, New York, USA. 578:73-94
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yoosheph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. (2001): The sequence of the human genome. *Science* 291, 1304 – 1351.
- Venter, J. C., Adams, M. D., Sutton, G. G., Kerlavage, A. R., Smith H. O and Hunkapiller, M. (1998): Shotgun sequencing of the human genome. *Science* 280(5369):1540-1542.
- Wilson, M. R., and DiZinno, J. A., Polanskey, D., Replogle, J., and Budowle, B. (1995): Validation of mitochondrial DNA sequencing for forensic casework analysis. *International Journal of Legal Medicine* 108, 68 – 74.
- Winnick, E. (2004): DNA sequencing industry sets its sights on the future. *The scientist* 18 (18), 44.
- Zhao, N. W. (2010): Recent progress in the methods of genome sequencing. *Brazilian Archives of Biology and Technology* 53(2):319-325
- Zwolak, M and Ventra, M. D. (2005): Electronic signature of DNA nucleotides via transverse transport. *Nano Lett.*, 5(3):421-424.