Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 8, Issue, 01, pp.24836-24845, January, 2016

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

A REVIEW ON DNA SEQUENCING – YESTERDAY, TODAY AND TOMORROW

***Adhav Rahul, Sonawane Hiralal and Deokule Subhash**

Department of Botany, Savitribai Phule Pune University, Pune - 411007, India

ARTICLE INFO

Article History:
Received 24th October, 2015
Received in revised form
10th November, 2015
Accepted 25th December, 2015
Published online 31st January, 2016

Key words:

Next generation sequencing, Roche/454 FLX Pyrosequencer, Illumina Genome Analyzer, ABI Solid Sequencer.

ABSTRACT

For the past 30 years, the Sanger method has been the dominant approach for DNA sequencing. In the new era of high-throughput sequencing the genomic analysis now referred as next-generation sequencing (NGS). Next generation sequencing (NGS) technologies have opened fascinating opportunities for the analysis of sequences. During the last few years, NGS methods have become widely available and cost effective. They can be applied to a wide variety of biological fields. In this review, fundamental principles of commercially available NGS platforms are discussed. Although differences in their working and sequencing qualities, through cycles of polymerase-mediated nucleotide extensions in one approach, through successive oligonucleotide ligations, sequence outputs in the range of hundreds of megabases to gigabases are now obtained routinely..In this review, the impact of NGS on basic research, bioinformatics considerations, and translation of this technology and a view into future technologies, including various sequencing technologies are highlighted.

Citation: **Adhav Rahul, Sonawane Hiralal and Deokule Subhash, 2016.** "A review on dna sequencing – yesterday, today and tomorrow", *International Journal of Current Research,* 8, (01), 24836-24845.

INTRODUCTION

DNA sequencing techniques are key tools in many fields. A large number of different sciences are receiving the benefits of these techniques, ranging from archaeology, anthropology, genetics, biotechnology, molecular biology, forensic sciences, and others. A silent and remarkable revolution is under way in many disciplines; DNA sequencing is promoting new discoveries that are revolutionizing the conceptual foundations of many fields. At the same time new and very important issues are emerging with these developments, such as bioethical questions and questions related to public health and safety (Lilian *et al*., 2002). The discovery of the double helix DNA by Watson and Crick (1953) was base of molecular biology. Prior to the development of rapid DNA sequencing methods by Sanger in the early 1970s, Gilbert and Maxam (1977) were used wandering-spot analysis method. In 1973 they reported sequence of 24 base pairs (Gilbert and Maxam, 1977). The chain-termination method was developed by Sanger and coworkers in 1977, which is relatively easy and reliable (Sanger and Coulson, 1975). The first complete genome of Bacterium (Bacteriophage MS2) was sequenced by Fiers (Sanger *et al*., 1977). Knowledge of the deoxyribonucleic acid (DNA) sequence is known in most of the living organisms, has revolutionized biology and driven a massive acceleration in research and development (Shendure and Ji 2008). The most commonly used approach for DNA sequence is the chain termination methodology developed by Fredrick Sanger in the 1970. Since the late 1990's, researchers were revisited the concept of DNA sequencing, and maximize the efficiency of sequence production. These efforts have yielded a number of step-changing approaches that have been rapidly commercialized today (Shendure and Ji 2008). DNA sequencing includes several methods that are used for determining the order of the nucleotide bases viz., adenine, guanine, cytosine and thymine. The Knowledge of DNA sequences has become indispensable for basic biological research, other research branches.DNA sequencing technology is useful in numerous applied fields; such as diagnostic, biotechnology, forensic etc.(http://en.wikipedia.org/wiki/DNA_sequencing).

Since 2005, these new sequencing machines are delivering significantly cheaper and faster sequence generation into a market that demonstrates massive demand (Schuster and Stephan, 2008). Few years ago the Sanger enzymatic dideoxy methods (Sanger *et al*., 1977) and the Maxam and Gilbert chemical degradation method were used (Maxam and Gilbert 1977). The two laboratories where the first automated DNA sequencers were produced, Leroy Hood at Caltech (Smith *et al*., 1986), which commercialized by Applied Biosystems,

****Corresponding author: Adhav Rahul***
Department of Botany, Savitribai Phule Pune University, Pune - 411007, India.

and Wilhelm Ansorge at the European Molecular Biology Laboratory EMBL and commercialized by Pharmacia-Amersham, later General Electric (GE) Healthcare (Ansorge *et al*., 1986, 1987). Recently, there has been a fundamental shift away from the application of automated Sanger sequencing for genome analysis. The limitations of automated Sanger sequencing showed a need for new and improved technologies for sequencing large numbers of genomes (Metzker, 2010). The automated Sanger method is considered as a 'first-generation' technology, and newer methods are referred to as next-generation sequencing (NGS) (Metzker, 2010). These newer technologies constitutes various strategies that rely on a combination of template preparation, sequencing and imaging, genome alignment and assembly methods. The major advance offered by NGS is the ability to produce an enormous volume of data cheaply — in some cases in excess of one billion short reads per instrument run. This feature expands the realm of experimentation beyond just determining the order of bases (Metzker, 2010).

Next-generation is high-throughput DNA sequencing techniques, which are opening fascinating new opportunities in biomedicine (Schuster *et al*., 2008). The large-scale DNA sequencing technique is without gels, extending primers in 'sequencing-by-synthesis, addition and detection of the incorporated base', proposing and describing the use of the so-called 'reversible terminators' for speed and efficiency (Ansorge and Heidelberg, 1991). The first step of the technique consisted in detecting the next added fluorescently labelled base (reversible terminator) in the growing DNA chain by means of a sensitive CCD camera. This was performed on a large number of DNA samples in parallel, attached either to a planar support or to beads, on DNA chips, minimizing reaction volumes in aminiaturised micro system. In the next step the terminator was converted to a standard nucleotide and the dye removed from it. This cycle and the process were repeated to determine the next base in the sequence (Ansorge, 2009). The principle described in the patent application is in part very similar to that used today in the so-called next-generation devices, with many additional original developments commercialized by Illumina-Solexa, Helicos and other companies. The relative affordability of next generation sequencing (NGS) is the great opportunities and challenges for biological research. The rapid speed of sequencing attained with modern DNA sequencing technology has been used for the genome sequencing e.g. human genome project and other related projects (Ansorge, 2009). Recently, complete DNA sequences of many animal, plant, and microbial genomes were done by modern DNA sequencing technology (Shendure and Ji 2008).

Methods for DNA sequencing

1. Chain-termination methods

The chain-terminator method is more efficient and uses fewer toxic chemicals and lower amounts of radioactivity (Sanger and Coulson, 1975); in this method dideoxy nucleotide triphosphates (ddNTPs) was used as DNA chain terminators and DNA sample was divided into four separate sequencing reactions, containing all four of the standard deoxynucleotides (dATP, dGTP, dCTP and dTTP) with DNA polymerase. In each reaction only one of the four deoxynucleotides (ddATP, ddGTP, ddCTP, or ddTTP) was added. The labelled DNA fragments are heat denatured and separated by gel electrophoresis on a denaturing polyacrylamide gel; the DNA bands were visualized by autoradiography or UV light. The dark bands correspond to DNA fragments of different lengths, indicates a DNA fragment that is the result of chain termination after incorporation of a dideoxynucleotide (ddATP, ddGTP, ddCTP, or ddTTP) (Sanger, Nicklen and Coulson, 1977).

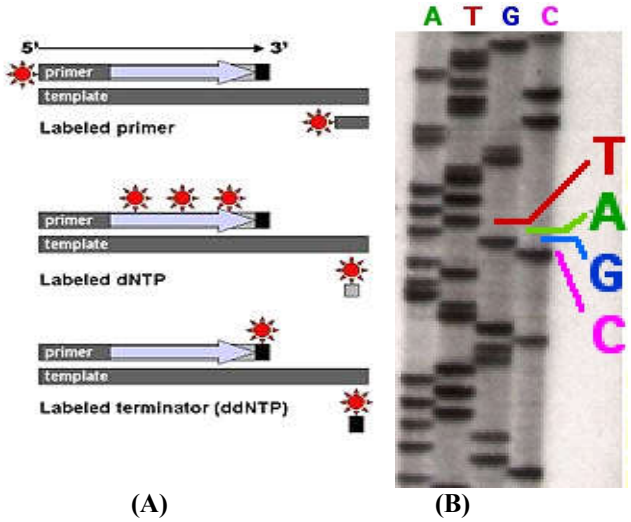


Fig. 1 (A). Labeled DNA fragment, (B). Part of a radioactively labeled sequencing gel (Smith *et al*., 1985; http://en.wikipedia.org/wiki/DNA_sequencing)

2. Maxam-Gilbert sequencing

In 1976–1977, Maxam and Gilbert developed a DNA sequencing method based on chemical modification of DNA and subsequent cleavage at specific bases (Maxam and Gilbert, 1977), it become more popular than Sanger method. However, Maxam-Gilbert sequencing has fallen out of favor due to its technical complexity prohibiting its use in standard molecular biology kits, extensive use of hazardous chemicals, and difficulties with scale-up (Graziano and Cecilia, 2003). This method requires radioactive labeling at one 5' end of the DNA (gamma-^{32}P ATP) and purified DNA fragment. Chemical treatment breaks nucleotide bases in each of four reactions (G, A+G, C, and C+ T). The fragments in the four reactions are electrophorized in denaturing acrylamide gels for size separation. The dark bands each corresponding to a radiolabelled DNA fragment.

3. Automated DNA Sequencing

Automated DNA-sequencer is modified Sangar DNA sequencing method. By using automated DNA-sequencer, we can sequence up to 384 DNA samples (http://en.wikipedia.org/wiki/DNA sequencing). In a single run using capillary electrophoresis for size separation, detection and recording of dye fluorescence (Smith *et al*., 1986) and data output as fluorescent peak trace chromatograms. The accuracy of such algorithms is below visual examination by a human

operator, but sufficient for automated processing of large sequence data sets (Olsvik *et al*., 1993).

4. High-throughput sequencing

The high-throughput DNA sequencing techniques are opening fascinating new opportunities in biomedicine, and which were selected by Nature scientific as the method of the year in 2007. (Schuster *et al*., 2008) The high demand for low-cost sequencing has driven the development of high-throughput sequencing technologies that producing thousands or millions of sequences at once. High-throughput sequencing technologies are intended to lower the cost of DNA sequencing methods .The following methods are used for High-throughput sequencing:

i. Lynx 'Therapeutics' Massively Parallel Signature Sequencing (MPSS) (Brenner, *et al*., 2000)
ii. Polony sequencing (Shendure, *et al*., 2005)
iii. 454 pyro-sequencing
iv. SOLID sequencing (Mardis, 2008)
v. Ion semiconductor sequencing
vi. DNA nanoball sequencing (Porreca, 2010)
vii. Single Molecule SMRT(TM) sequencing (Quake *et al*., 2003)
viii. Single Molecule real time (RNAP) sequencing
ix. Nanopore DNA sequencing

Next Generation Sequencing Technology

Next Generation Sequencing Technology is the cheaper, faster and more accurate sequencing from the next wave of sequencing technology; it is different technology to the current generation sequencing machines. There are three main companies developing these sequencing systems and the first to become commercially available viz., Roche, Illumina. NGS technique includes number of steps: template preparation, sequencing and imaging, and data analysis, it shows unique combination of specific protocols distinguishes one technology from another and determines the type of data produced from each platform (Metzker, 2010).

Table 1. Comparative account of three most popular DNA sequencers

SOLiD 5500xl	Illumina HiSeq 2000	Roche 454 GS FLX+
MP 2x60bp, PE 75bp x 35bp, Fragment 75bp	1X36bp, 2x50bp, 2x100bp, 2x150bp	Up to 1000bp
Up to 99.99% acc. 20-30 Gbase/day	80% > Q30, 2x100 55 Gbase/day	99.997% accuracy 1M reads, 700bp; 0.7 Gbase/day
MP 2x60bp in 14 (18 with ECC) days. leading to 210 Gbase or 2 35x coverage human genomes	2x100bp in 11 days. 2x150 in 16 days leading to 600 Gbase or 6 30x coverage human genomes	Homopolymer issue
Cost $6,000K 2x60 dual flowcells $600K instrument $29/Gbase	Cost ~$28K 2x100 dual flowcell $750K instrument $46/Gbase $4500 30x service	Cost ~$14K for run $500K instrument $20K/Gbase

There are two methods used in preparing templates for NGS reactions: clonally amplified templates originating from single DNA molecules, and single DNA molecule templates. For the sequencing by synthesis, cyclic reversible termination (CRT), single-nucleotide addition (SNA) and real-time sequencing methods were used. The imaging methods are coupled with these sequencing strategies range from measuring bioluminescent signals to four-color imaging of single molecular events (Pop and Salzberg, 2008).

(A)Roche 454 GS FLX **(B) Illumina HiSeq 2000**

(C) SOLiD 5500xl

Fig. 2. The popular Next-Generation Sequencers available in the market

Next generation sequencing methods

1. Roche/454 FLX Pyrosequencer

The pyrophosphate detection technology was described in 1985 (Nyren, 1985), and using this principle in a new method for DNA sequencing was reported in 1988 (Hyman, 1988). This next-generation sequencer was the first to achieve commercial introduction in market (Margulies *et al*., 2006) and uses an alternative sequencing technology known as pyrosequencing (Ronaghi *et al*., 1996). In pyrosequencing, each incorporation of a nucleotide by DNA polymerase results in the release of pyrophosphate, which initiates a series of downstream reactions that ultimately produce light by the firefly enzyme (Nyren and Lundin, 1985). In this system DNA fragments are ligated with specific adapters that cause the

binding of one fragment to a bead. The library fragments are mixed with a population of agarose beads whose surfaces carry oligonucleotides complementary to the 454-specific adapter sequences on the fragment library, so each bead is associated with a single fragment. The PCR is carried out for fragment amplification, with water droplets containing one bead and PCR reagents immersed in oil. Each of these fragment:bead complexes is isolated into individual oil: water micelles that also contain PCR reactants. The amplification is obtained with sufficient light signal intensity for reliable detection in the sequencing-by-synthesis reaction steps. In PCR amplification cycles after denaturation, each bead with its one amplified fragment is placed at the top end of an etched fiber in an optical fiber chip, created from glass fiber bundles. The individual glass fibers are excellent light guides, with the other end facing a sensitive CCD camera, enabling positional detection of emitted light. Each bead thus sits on an addressable position in the light guide chip, containing several hundred thousand fibres with attached beads (Shendure and Hanlee, 2008).

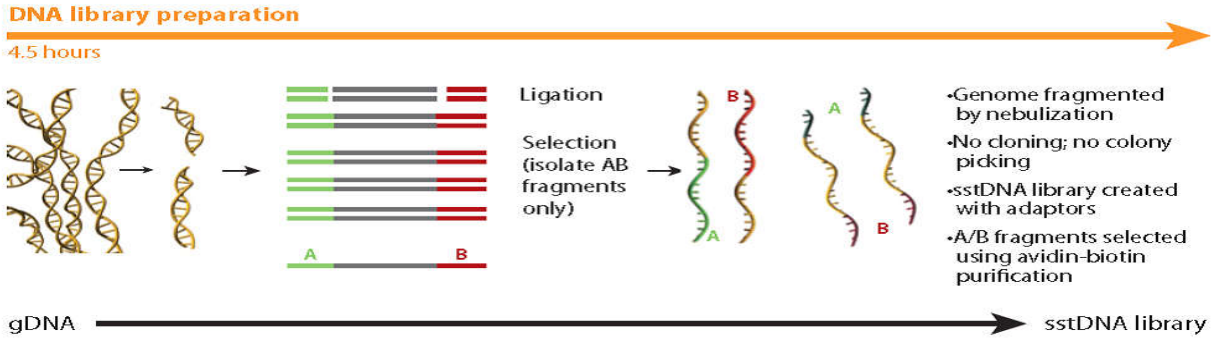
Thermal cycling (emulsion PCR) of the micelles produces approximately one million copies of each DNA fragment on the surface of each bead (Dressman, 2003). These amplified single molecules are then sequenced. Enzyme containing beads that catalyze the downstream pyrosequencing reaction steps are then added to the PTP and the mixture is centrifuged to surround the agarose beads. On instrument, the PTP acts as a flow cell into which each pure nucleotide solution is introduced in a stepwise fashion, with an imaging step after each nucleotide incorporation step. The PTP is seated opposite a CCD camera that records the light emitted at each bead.

The first four nucleotides (TCGA) on the adapter fragment adjacent to the sequencing primer added in library construction correspond to the sequential flow of nucleotides into the flow cell. The method has recently increased the achieved reading length to the 400–500 base range, with paired-end reads, and as such is being applied to genome (bacterial, animal, human) sequencing (Shendure and Hanlee ,2008). The next upgrade 454 FLX Titanium will quintuple the data output from 100 Mb to about 500 Mb, and the new picotiter plate in the device uses smaller beads.

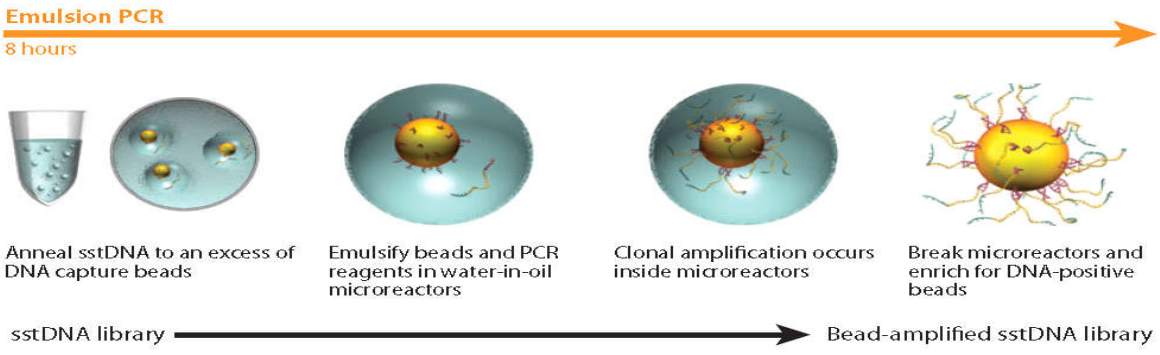
2. Illumina Genome Analyzer

The Solexa sequencing method was commercialised in 2006, with Illumina acquiring Solexa in early 2007. This method is based on sequencing-by-synthesis chemistry, with novel reversible terminator nucleotides for the four bases each labelled with a different fluorescent dye, and a special DNA polymerase enzyme able to incorporate them. The single molecule amplification step for the Illumina Genome Analyzer starts with an Illumina-specific adapter library. DNA fragments are ligated at both ends to adapters and after denaturation, immobilization the surface of the support is coated densely with the adapters and the complementary adapters. Each single-stranded fragment was immobilised at one end on the surface. The flow cell is an 8-channel sealed glass micro fabricated device that allows bridge amplification of fragments on its surface, the uses of DNA polymerase to produce multiple DNA copies or clusters, each represented the single molecule to initiate the cluster amplification (Dressman, 2003). The bridge structure was hybridised with its free end to the complementary adapter on the surface of the support.

Fig. 3 (A) Roche 454 Pyrosequencer DNA library preparation (Nyren, 1985; Hyman, 1988)

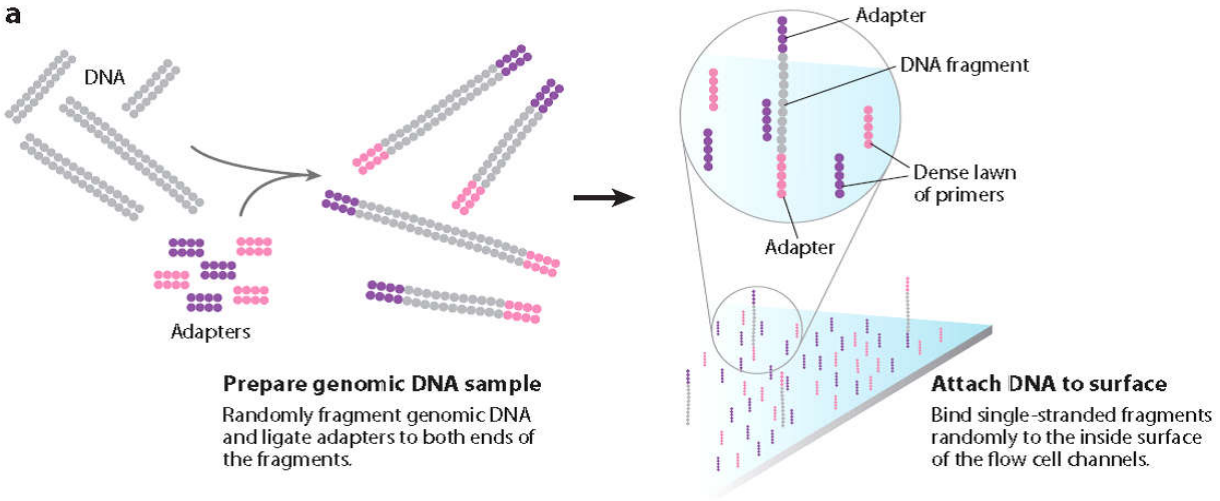


(B) Amplification of DNA

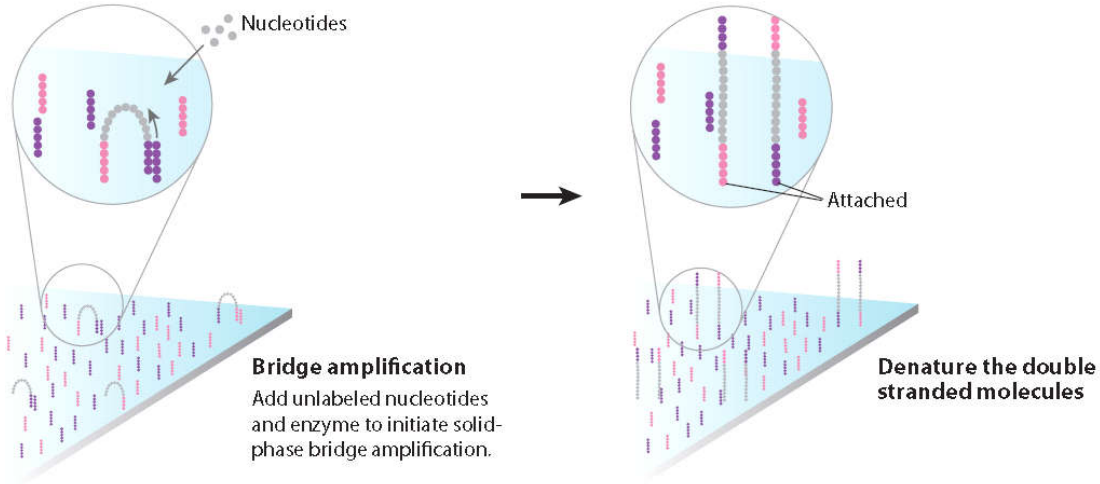


(c) Sequencing of DNA

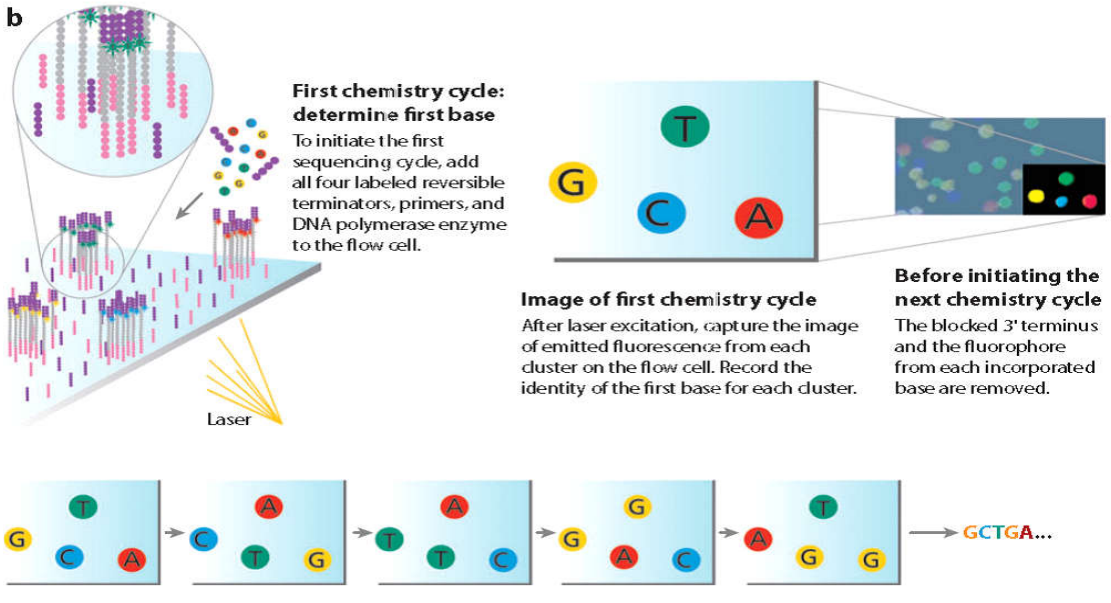
Fig. 4. (A) Preparation of genomic DNA sample (Elaine, Mardis 2008)



(B) Amplification of DNA (Bridge PCR)



(C) Sequencing of DNA



Each cluster contains approximately one million copies of the original fragment, which is sufficient for reporting incorporated bases at the required signal intensity for detection during sequencing. The PCR amplification needed sufficient light signal intensity for reliable detection of the added bases. After several PCR cycles, random clusters of about 1000 copies of single-stranded DNA fragments are created on the surface. The DNA synthesis is supplied onto the surface and contains primers, four reversible terminator nucleotides each labeled with a different fluorescent dye and the DNA polymerase. After incorporation into the DNA strand, the terminator nucleotide, as well as its position on the support surface, is detected and identified via its fluorescent dye by the CCD camera. The sequence read length achieved about 35 nucleotides (Elaine, 2008). The upgraded Genome Analyzer gives triples output compared to the previous Genome Analyzer. The new instrument triples the output per paired-end run from 1 to 3 Gb. The system generates at least 1.5 Gb of single-read data per run and at least 3 Gb of data in a paired-end run, recording data from more than 50 million reads per flow cell. The run time for a 36 cycle run was decreased to two days for a single-read run, and four days for a paired-end run (Elaine, 2008).

3. Applied Biosystems SOLiD TM Sequencer

The ABI SOLiD sequencing model is chemistry based upon ligation. this adapter-ligated fragment library similar to those of the other next-generation. In this technique, DNA fragments are ligated to adapters then bound to beads and the water droplet in the oil emulsion for the amplification only one fragment bound per bead and DNA fragments on the beads are amplified by the emulsion PCR for sequencing (Dressman, 2003; Shendure, *et al*., 2005; McKernan, 2006; Housby, *et al*., 1998). In the PCR, the beads are deposited onto a glass support surface after the denaturation of DNA.

Then in first step primer was hybridised to the adapter. The resulting mixture of oligonucleotide octamers is hybridised to the DNA fragments and ligation mixture added. In these octamers was characterised by one of four fluorescent labels at the end. The detection of the fluorescence from the label, bases 4 and 5 in the sequence are thus determined and then after the fifth base, removing the fluorescent label, then hybridisation and ligation cycles are repeated, this time determining bases 9 and 10 in the sequence; in the subsequent cycle bases 14 and 15 are determined, and so on. Read lengths of sequences are between 25–35 bp, and each sequencing run yields between 2–4 Gb of DNA sequence data. Recently, updated version (SOLiD 2.0) was lunched, which may increase the output of the instrument from 3 to 10 Gb per run in 4.5 days from 8.5 days on the existing machine (Elaine, 2008).

Applications of next generation sequencing

1). Plant breeding

The application of NGS in plant breeding will yield reference genomes for crop species including those with very large genomes (e.g. wheat). Re-sequencing of cultivated and wild variants will provide insights into domestication and adaptation to changing environments (e.g. climate change). Re-sequencing of multiple crop cultivars will facilitate the dissection of the genetic architecture of quantitative phenotypic traits; this will allow marker-assisted selection on a genome scale, an approached dubbed 'genomics assisted breeding'. It will also facilitate the incorporation of useful allelic variants from landraces and ecotypes into domestic varieties, thus helping to maintain the influx of novel genetic variation into elite breeding material. NGS-enabled predictive breeding is expected to advance efforts to adapt crops for changing environments. The large-scale genotyping of thousands of individual plants is already a possibility for commercial selective breeding programmes, and the computational intensity of such work can be expected to increase significantly. There is, therefore, scope for significant socio-economic impacts from NGS-powered advances in crop breeding.

2). Plant-microbe interactions

Metagenomic studies will make an important contribution to understanding how plants and microbes influence rhizosphere composition, how microbial community composition affects microbial pathogenesis (e.g. soil-borne pathogens) and how new and emerging crop diseases originate. The ability to sequence, rapidly, the genomes of microbial pathogens affecting agricultural production will help to identify outbreak strains more quickly and enable prompt intervention.

3). Animal breeding

Similar to plant breeding, the key opportunities lie in the identification of genomic variation and its exploitation in selective breeding for improved productivity and health traits. It could also speed up disease diagnosis and the development of vaccines. Successful deployment could lead to significant economic and societal impact.

4). Pathogens

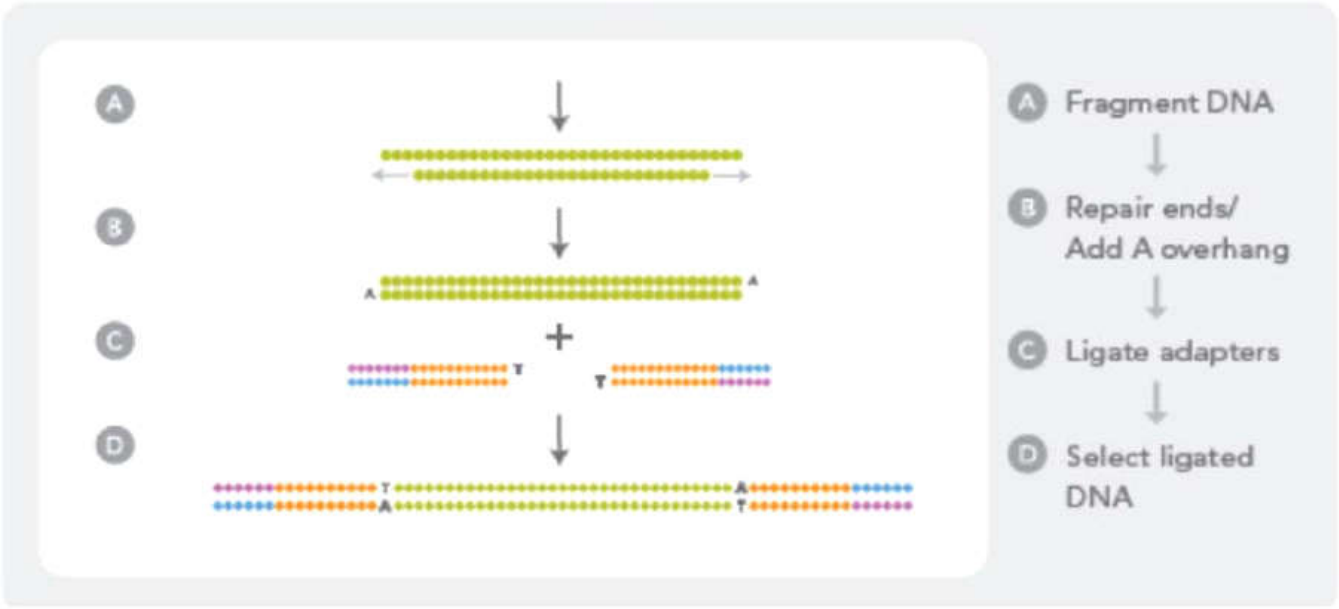
NGS-scaled methods identification and comparison of genetic differences between related pathogens will help researchers to elucidate how a pathogen adapts to new host ranges or environments. These studies will inform risk assessments for new and emerging diseases. Pathogen genome and population genome information may be used to map, understand and control important human, animal and plant pathogens.

Within five years, targeted intervention may become feasible for serious bacterial infections; it is anticipated that such techniques will first applied to clinical applications, but that they will also become available to veterinary medicine.

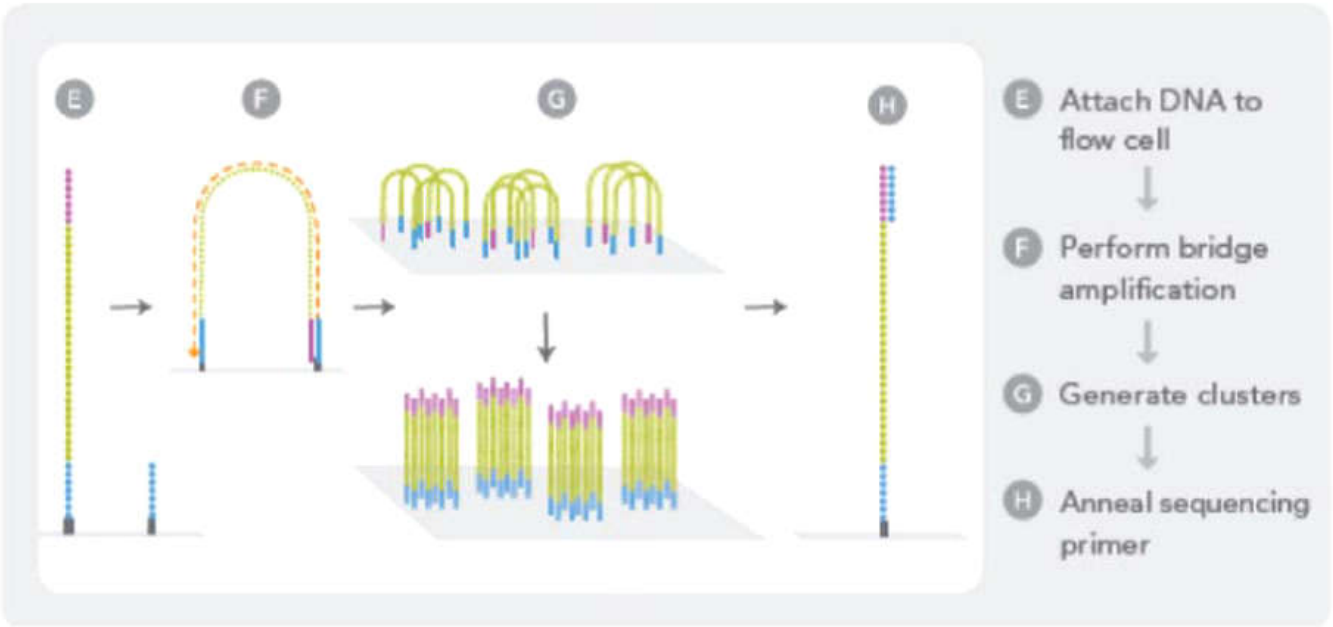
5). Industrial biotechnology

Metagenomics studies are expected to make a significant contribution to industrial biotechnology, through identification of metabolic pathways and novel secondary metabolites (e.g. antimicrobials) with applications in drug discovery and biopharmaceutical development.

Fig. 5. Applied Biosystems SOLiD TM Sequencer (Shendure *et al*., 2005; McKernan, 2006; Housby, *et al*., 1998) (A) Preparation of DNA Sample for sequencing



(B) Bridge amplification, Primer annealing and sequencing



6). **Personalised medicine and diagnostics**

NGS could lead to the routine sequencing of individuals (e.g. at birth, or at medical need) and thereby usher in an era of personalised medicine. Genetic mapping and dissection of complex diseases could improve the ability of medical practitioners to predict the propensity of individuals to develop specific diseases (i.e. 'risk genes') and lead to development of earlier, more accurate and more sensitive disease diagnostics. Knowledge of an individual patient's genome sequence could inform treatment regimes, identifying appropriate drug targets and susceptibilities and enabling higher efficacy drug selection. Of particular note, identifying the complex genetic determinants and corresponding efficacious drug targets for individual cancer cases has the potential to revolutionize diagnosis, intervention and prognosis.

7). Respondents also taxonomy and systems biology as two broad research methodologies that would be considerably benefited by application of NGS approaches. In both cases, access to large, accurate NGS generated datasets would enable deeper, more comprehensive studies to be undertaken. In the case of systems biology, in particular, the scales of investigation enabled by NGS could be considered integral to the high-throughput investigative modes that will be necessary to make significant future progress.

8).**De novo sequencing**

NGS technologies are currently being used to sequence the small, simpler genomes of prokaryotes. There is an expansion into understudied species and generation of additional reference genomes. The use of NGS in the de novo sequencing of eukaryotes is currently expanding, with uptake in cDNA sequencing studies and more limited application in whole genome approaches (e.g. some application in sheep and wheat). Looking ahead, it is expected that NGS deployment will become common-place, yielding the genomes of many more prokaryotic and eukaryotic species of strategic importance. Respondents anticipated that both large-scale sequencing centers and small scale laboratories will be engaged in de novo sequencing

9).Gene expression and epigenetics analysis tools

NGS is a popular method for high throughput analysis of gene expression and is in the process of replacing DNA microarrays as the method of choice. Advantages of using NGS (termed RNA-seq) are:(a) it can be used with any organism as it is not restricted to the probe set on the microarray; (b) it offers greater sensitivity and dynamic range; (c) it can assess the levels of non-coding / antisense transcripts, and; (d) it detects sequence and splice differences. NGS is also replacing arrays in chromatin immune precipitation studies (termed ChIPseq), a method used in epigenetics to study DNA-binding proteins, histone modifications and nucleosomes. The UK is considered a global leader in epigenetics and a recent BBSRC/MRC investment (£1.65M) at the Babraham Institute will deploy NGS to investigate the epigenetic basis of normal development and healthy ageing. NGS is also being used to profile small RNA populations (sRNA-seq) that influence the transcription and translation of protein coding RNAs, and to identify DNA methylation through bisulphite NGS.

10). The past several years have seen an accelerating flurry of publications in which next generation sequencing is applied for a variety of goals Important applications include: (i) full-genome resequencing or more targeted discovery of mutations or polymorphisms (ii) mapping of structural rearrangements, which may include copy number variation, balanced translocation breakpoints and chromosomal inversions; (iii) 'RNA-Seq', analogous to expressed sequence tags (EST) or serial analysis of gene expression (SAGE), where shotgun libraries derived from mRNA or small RNAs are deeply sequenced; the counts corresponding to individual species can be used for quantification over a broad dynamic range, and the sequences themselves can be used for annotation (e.g., splice junctions and transcript boundaries); (iv) large-scale analysis of DNA methylation, by deep sequencing of bisulfite-treated DNA; (v) 'ChIP-Seq', or genome-wide mapping of DNA-protein interactions, by deep sequencing of DNA fragments pulled down by chromatin immune precipitation. Over the next few years, the list of applications will undoubtedly grow, as will the sophistication with which existing applications are carried out.

REFERENCES

Ansorge, J. 2009. Next-generation DNA sequencing techniques. *New Biotechnology* ! Volume 25, Number 4.

Ansorge, W. B Sproat, J Stegemann, C Schwager, and M Zenke 1987. Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis. *Nucleic Acids Res*., 15, 4593–4602.

Ansorge, W. Sproat BS, Stegemann J, and Schwager C. 1986. A non-radioactive automated method for DNA sequence determination. *J. Biochem. Biophys. Methods*, 13, 315–323.

Bentley, D. 2008. Accurate whole human genome sequencing using reversible terminator. chemistry. *Nature*., 456: 53–59

Braslavsky, I. Hebert, B. Kartalov, E. and Quake R. 2003. Sequence Information can be obtained from single DNA molecules. Proc. Natl. Acad. Sci. USA., 100: 3960– 3964.

Brenner, Sidney, Johnson M. Bridgham, J. Golda, G. Lloyd, H. Johnson, D. Luo, S. and McCurdy, S. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnol*., 18 (6): 630–634.PMID 1074580.

Brian, O. 2011. Massively Parallel Sequencing experimental designs and computational Challenges.(academy.nescent. org/wg/academy/images/b/b1/Lecture_2_NGS_Experiment al_Designs_and_Infrastructure)

Church, G. 2006."Genomes for all". *Sci. Am*., 294 (1): 46–54.

Dressman, D. Yan, H. Traverso, G. Kinzler, K. and Vogelstein, B. 2003. Transforming single DNA molecules into fluorescent magnetic particlesfor detection and enumeration of gnetic variations. *Proc. Natl. Acad. Sci*. USA 100: 8817–8822.

Drmanac, R. 2010. Human genome sequencing using unchained base reads in self assembling DNA nanoarrys. *Science*, 327 (5961): 78-81.

Edwards, A. Voss, H. Rice, P. Civitello, A. Stegemann, J. Schwager, C. Zimmerman, J. Erfle, H. Caskey, T. and Ansorge, W. 1990. "Automated DNA sequencing of the human HPRT locus". Genomics 6 (4): 593–608.doi:10.1016/0888-7543(90)90493-E.PMID 2341149.

Edwards, J. Ruparel, H. and Ju, J. 2005. Mass-spectrometry DNA sequencing. *Mutation Research,* 573 (1–2): 3–12. PMID 15829234.

Elaine, R. 2008. Next-Generation DNA Sequencing Methods. *Annu. Rev. Genomics Hum. Genet*, 9:387–402

Erlich, Y. Mitra, P. Delabastide, M. McCombie, W. and Hannon, G. 2008. Alta-Cyclic: a self- optimizing base caller for next-generation seq. *Nature Meth*, 5, 679–682.

Fedurco, M. Romieu, A. Williams, S. Lawrence, I. and Turcatti, G. 2006. BTA, a novel reagent for DNA attachment onglass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res*., 34, e22.

Fiers, W. Contreras, R. Duerinck, F. 1976. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene". *Nature,* 260 (5551): 500 PMID 126403.

Flusberg, B. 2010. Direct detection of DNA methylation during single-molecule, real time sequencing. *Nat. Methods,* 7, 461–465.

Franca, L. Carrilho, E. Kist, T. 2002. A review of DNA sequencing techniques. *Q. Rev. Biophys*., 35:169-200.

Gilbert, W. Maxam, A. 1973. The nucleotide sequence of the lac operator. Proc. Natl. Acad. Sci. U.S.A. 70(12): 3581–PMC 427284.PMID 428755.

Graziano, P. Cecilia, S. 2003. Hand book of comparative genomics Principles and Methodology Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Hall, N. 2007. Advanced sequencing technologies and their wider impact in microbiology. *J. Exp. Biol*., 210 (Pt 9): 1518–25.PMID 17449817.

Hanna, G. Johnson, V. Kuritzkes, D. 2000. Comparison of sequencing by hybridization and cycle sequencing for genotyping of human immunodeficiency virus 1 reverse transcriptase. *J. Clin. Microbiol*., 38 (7): 2715–21. PMC 87006. PMID 10878069.

Harris, T. 2008. Single-molecule DNA sequencing of a viral genome. *Science*, 320:106– 109.

Homer, N. 2010. A survey of sequence alignment algorithms for next generation sequencing. *Brief Bioinform*, 11:473-483.

Housby, J. and Southern, E. 1998. Fidelity of DNA ligation: a novel experimental approach based on the polymerisation of libraries of oligonucleotides. *Nucleic Acids Res*., 26: 4259–4266.

http://en.wikipedia.org/wiki/DNA sequencing.

Hyman, E. 1988. A new method of sequencing DNA. *Anal. Biochem*., 174: 423–436.

Kim, J. 2007. Polony multiplex analysis of gene expression (PMAGE) in mouse Hypertrophic cardiomyopathy. *Science*, 316: 1481–1484.

Leamon, J. 2003. A massively parallel Pico Titer Plate based platform for discrete picoliter- Scale polymerase chain reactions. *Electrophoresis*, 24:3769–3777.

Lilian, T. Emanuel, C. and Tarso, B. 2002. A review of DNA sequencing techniques. Cambridge University Press, *Quarterly Reviews of Biophysics*, 35, 2 (2002),pp.169–200.

Mardis, E. 2008. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*, 9: 387–402. PMID 18571739.

Margulies, M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437: 376–380

Margulies, M. Egholm, M. Altman, E. 2005. *Nature*, 437 (7057): 376–PMC 1464427. PMID 1605462.

Maxam, M. Gilbert, W. 1977. A new method for DNA sequencing. Proc. Natl. Acad. Sci. U.S.A. 74 (2): 560–PMC 392330. PMID 2632530.

McCarthy, A. 2010. Third generation DNA sequencing: Pacific biosciences' single molecule realtime technology. *Chem. Biol*., 17:675–676.

McKernan, K. Blanchard, A. Kotler, L. and Costa, G. 2006. Reagent, Methods and libraries for bead-based sequencing. US patent application. Life Technologies Corporation Pub. No.: US 2011/0077169 A1.

Metzker M. L. 2010. Sequencing technologies - the next generation. *Nat Rev Genet*., 11:31-46.

Min, J. Haegeman, G. Ysebaert, M. Fiers, W. 1972. Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature*, 237 (5350): 82–PMID 4555447

Munro, D. and Harris, T. 2010. Third-generation sequencing fireworks at MarcoIsl and. *Nat. Biotechnol*., 28: 426–428.

Murphy, K. Berg, K. and Eshleman, J. 2005. Sequencing of genomic DNA by combined amplification and cycle sequencing reaction. *Clinical Chemistry*, 51 (1): 35– PMID 15514093.

Nyren, P. and Lundin, A. 1985. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Anal. Biochem.*, 151: 504–509.

Olsvik, O. 1993. Use of automated sequencing of polymerase chain reaction generated amplicons to identify three types of cholera toxin sub unit B " *J. Clin. Microbiol.,* 31 (1): 22–5 PMC 262614. PMID 7678018.

Pettersson, E. Lundeberg. J. and Ahmadian, A. 2009. Generations of sequencing technologies. *Genomics*, 93 (2): 105–11 PMID 18992322.

Porreca, J. 2010. Genome sequences of nanoballs. *Nature, Biotechnol*., 28:(43-44).

Richard, W. Sergio, G. Oliver, J. Shlomo, M. Dan, S. and Andrew, D. 2006. Amplification of complex gene libraries by emulsion PCR. *Nature Methods*, 3(7): 545–550. PMID 16791312.

Ronaghi, M. Karamohamed, S. Pettersson, B. Uhlen, M. and Nyren, P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem*., 242: 84–89.

Rusk, N. 2011. Torrents of sequence. *Nat Meth*, 8(1): 44-44.

Sanger, F. 1980. Determination of nucleotide DNA sequences. Nobel lecture.

Sanger, F. Air, G. Barrell, B. 1977. Nucleotide sequence of Bacteriophage phi X174 DNA. *Nature,* 265 (5596): 687–95.PMID 830838.

Sanger, F. and Coulson, A. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol*., 94 (3): 441–8. PMID 110048.

Sanger, F. Nicklen. S. and Coulson, A. 1977. DNA sequencing with chain terminating inhibitor" Proc. Natl. Acad. Sci. *U.S.A*. 74 (12): 5463–PMC 431765. PMID 432729.

Schadt, E. Turner, S. and Kasarskis, S. 2010. A window into third-generation sequencing. *Human Molecular Genetics*, 19 (R2): R227–40.PMID 20858600.

Schuster, S. 2008. Method of the year,next-generation DNA sequencing. Functional genomics and medical applications. *Nat. Methods*, 5: 11–21.

Schuster, S. 2008. Next-generation sequencing transforms today's biology. *Nat. Methods*, 5, (1):16–8PMID 14165802.

Schuster, S. and Stephan, C. 2008. Next-generation sequencing transforms today's biology. *Nature Methods*, "5(1): 16–18.

Sengupta D, Cookson, B. 2010. The Journal of molecular diagnostics *: JMD* 12 (3): 272–PMC 2860461 PMID 2020300.

Shendure, J. and Ji, H. 2008. Next-generation DNA sequencing. *Nature Biotechnology*, 26, 1135-1145

Shendure, J. Porreca, G. Reppas, N. Lin, X. McCutcheon, J. Rosenbaum, A. Wang, M. and Zhang, K. 2005. Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science*, 309 (5741): 1728–32.

Smith, L. Fung, S. Hunkapiller, M. Hunkapiller, and T. Hood, L. 1985. *Nucleic Acids Res*., 13 (7): 2399–412 PMC 341163. PMID 4000969.

Smith, L. Sanders, J. and Kaiser, R. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature*, 321(6071): 674–9. PMID 3713851.

The harvard nanopore group, 2009. Mcb.harvard.edu.

Turcatti, G. Romieu, A. Fedurco, M. and Tairi, A. 2008. A new class of cleavable fluorescent cleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Res*., 36, e25.

Valouev, A. Ichikawa, J. and Tonthat, T. 2008. *Genome Res*., 18(7):

Visigen Biotechnologies, 2009. Inc- Technology Overview. Visigenbio.com.

Watson, J. and Crick, F. 1953. The structure of DNA. Cold Spring Harb. Symp. Quant. Biol. 18: 123–31. PMID 13138976.

Xu, M. Fujita, D. and Hanagata, N. 2009. Perspectives and challenges of emerging single-molecule DNA sequencing technologies. *Small* 5 (23): 2638–49 PMID 19904762.

Ying-Ja, C. Eric, E. and Xiaohua H. 2010. DNA sequencing by denaturation: experimental proof of concept with an integrated fluidic device. Lab on Chip 10 (10): 1153–1159.

****