Lesson Prepared Under MHRD project " National Mission on Education Through ICT"

Discipline: Botany
Paper: Plant Biotechnology

National Coordinator: Prof. S.C. Bhatla

Lesson: DNA Sequencing

Lesson Developer: Vinee Khanna
Department/College: Department of Genetics, University of Delhi, South Campus

Lesson Reviewer: Dr. Rahul Kumar
Repository of Tomato Genomic Resources
Department of Plant Sciences, School of Life Sciences
University of Hyderabad

Language Editor: Namrata Dhaka
Department/College: Department of Genetics, University of Delhi, South Campus

Lesson Editor: Dr Rama Sisodia, Fellow in Botany ILLL

Chapter: DNA Sequencing

Table of Contents

Chapter: DNA Sequencing

- **Advantages**
- **Applications of NGS**

- **Summary**
- **Exercise/Practice**
- **Glossary**
- **References**
- **Further reading**

Learning outcomes

After reading this chapter, the reader should be able to understand the following:

- Genome sequencing and its basics.
- The principle behind DNA sequencing.
- Various methods of sequencing, including Next generation sequencing (NGS).
- Applications of sequencing.

Introduction

DNA sequencing tells us about the precise sequence of nucleotides in the sample of DNA. The oldest method of sequencing is Sanger's method, which was first introduced in the year 1977. There have been many modifications in this method since then and many technologies have been introduced to improve the quality, time, length and cost of reads. The latest technology introgression in this field has been that of Next Generation Sequencing (NGS), which is a rapid, high throughput method. We shall be discussing various methods of sequencing in detail.

What is a genome?

It is a list of instructions which encodes the formation of DNA which in turn constitutes an organism. For example, the human genome is made of 3 billion bases of DNA which are arranged into 24 chromosomes.

Video:http://www.dnalc.org/view/16812-Animation-39-A-genome-is-an-entire-set-of-genes-.html (cc)

The DNA molecule is made of a sugar backbone and 3'-5' OH (hydroxyl) residues. The sugar backbone is made up by nucleotides. These nucleotides consist of a five carbon sugar deoxyribose, a nitrogenous base and a phosphate group. The nitrogenous bases may be Adenine, Cytosine, Guanine or Thymine. The 3'-5' OH group is called the deoxy- group and is essential for the chain elongation. In DNA sequencing these bases are read in a DNA fragment.

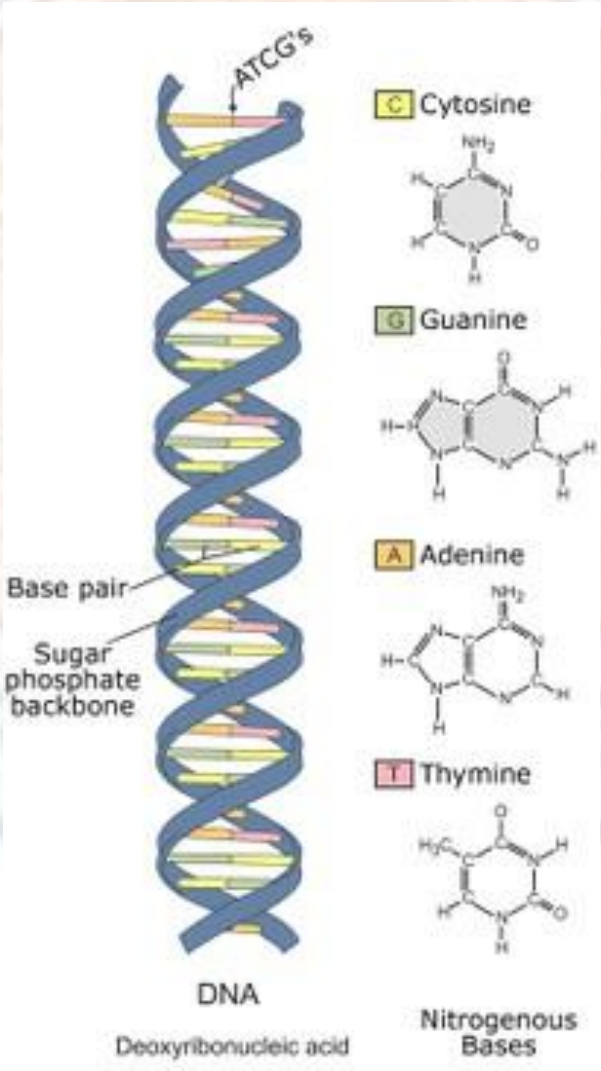


Figure: Illustration of structure of a DNA molecule and its constituent bases.

Source: http://www2.le.ac.uk/departments/genetics/vgec/schoolscolleges/topics/dna-genes-chromosomes (cc)

Video: http://www.dnalc.org/view/15922-Early-DNA-sequencing.html(cc)

Methods of sequencing

DNA can be sequenced using various methods as shown in the figure ahead. Sanger's method and Maxam and Gilbert's method are the first methods to be used for DNA sequencing and they have been variously modified into other methods of sequencing.

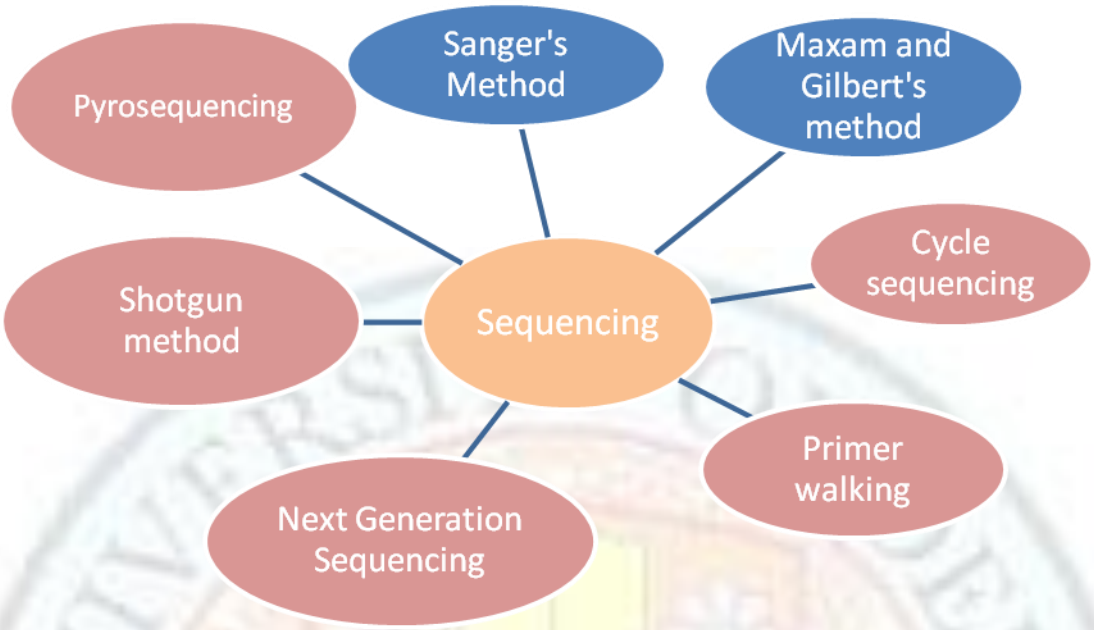


Figure: The different methods of sequencing.

Source: Author

Sanger's method

Video: Introduction to sanger sequencing methodology

This method was described by Fred Sanger in 1977. In this technique, the DNA is sequenced using an enzymatic method which polymerizes the DNA fragments complimentary to the DNA of interest. P^{32} is used to label the synthetically designed primer that binds to the DNA template at a known sequence. The synthesis occurrs with DNA polymerases and dNTPs (deoxynucleotide triphosphate) until a ddNTP (dideoxynucleotide triphosphate) is incorporated which terminates the reaction due to the absence of the deoxy- group.

This is carried out in four reaction tubes containing the four nucleotides (A, T, G, and C) in the dideoxy- form. The starting point of synthesis is same but the 3' end is specific to the ddNTP attached. The fragments are run on a denaturing polyacrylamide gel on four different lanes. The gel pattern specifies the chain termination site and the sequence can be read on an autoradiograph.

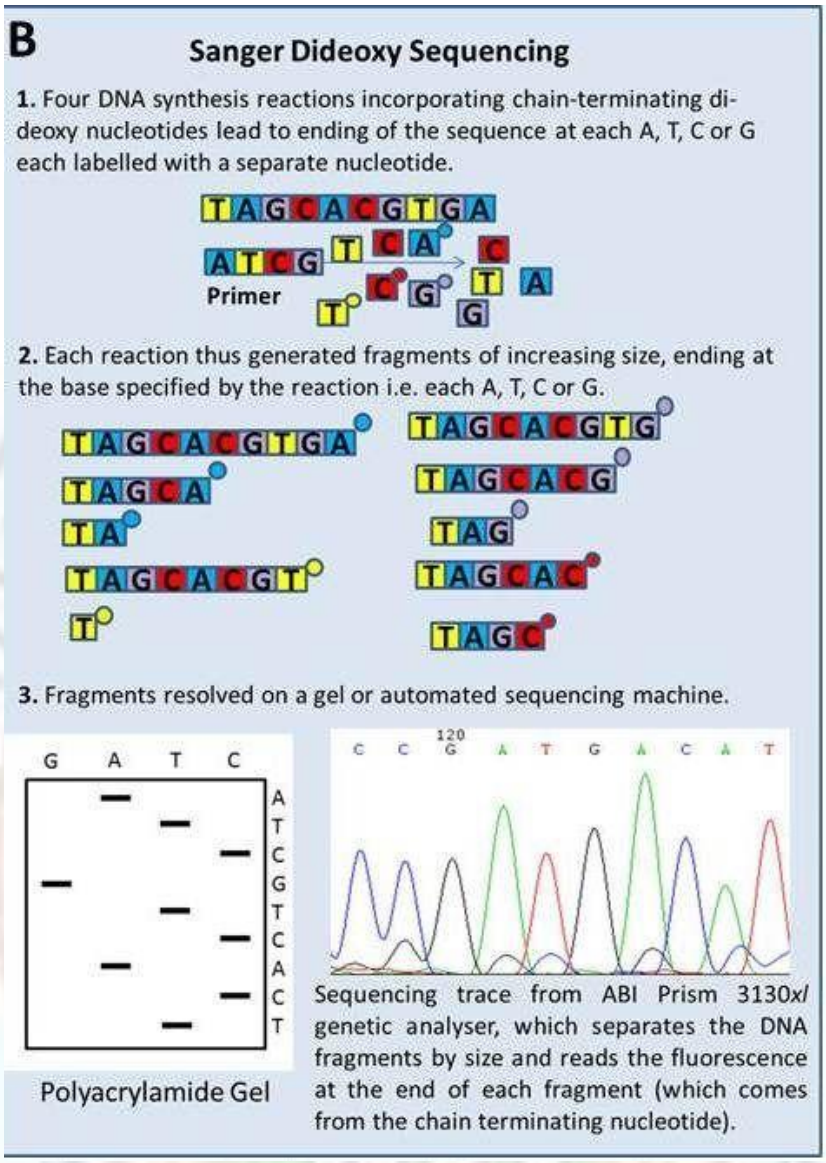


Figure: Sanger sequencing method.

Source: http://www.oxbridgebiotech.com/review/research-and-policy/whats-so-special-about-next-generation-sequencing/ (cc)

File: Sanger_DNA_sequencing.swf

Animation: Sanger sequencing strategy

Fred Sanger

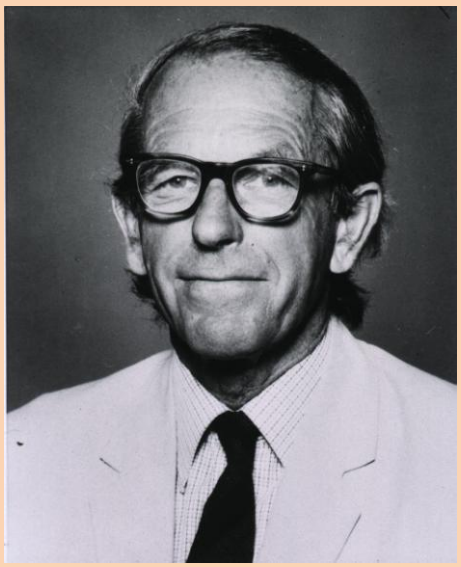


Figure: Photograph of Fred Sanger.

Source: http://en.wikipedia.org/wiki/Frederick_Sanger (cc)

He got his first Nobel prize for chemistry on the structure of protein deciphering. On a similar note he observed that genes and DNA that make the protein should also be ordered.

In 1980 he shared his second Nobel prize for chemistry with Maxam Gilbert for his contribution in deciphering the base sequence in the nucleic acids.

Video: http://www.dnalc.org/view/16036-Fred-Sanger-1975.html(cc)

Cycle sequencing

It is a modification of the Sanger sequencing method as it involves a polymerase stable at high temperatures. The reaction includes the DNA template, polymerase, and ddNTPs and is carried out in a PCR machine. The reaction mixture is then loaded onto a sequencer which detects the specific signals associated with the incorporated ddNTPs. As the enzyme is stable over high temperatures the reaction can be repeated again.

Video: http://www.dnalc.org/view/15923-Cycle-sequencing.html(cc)

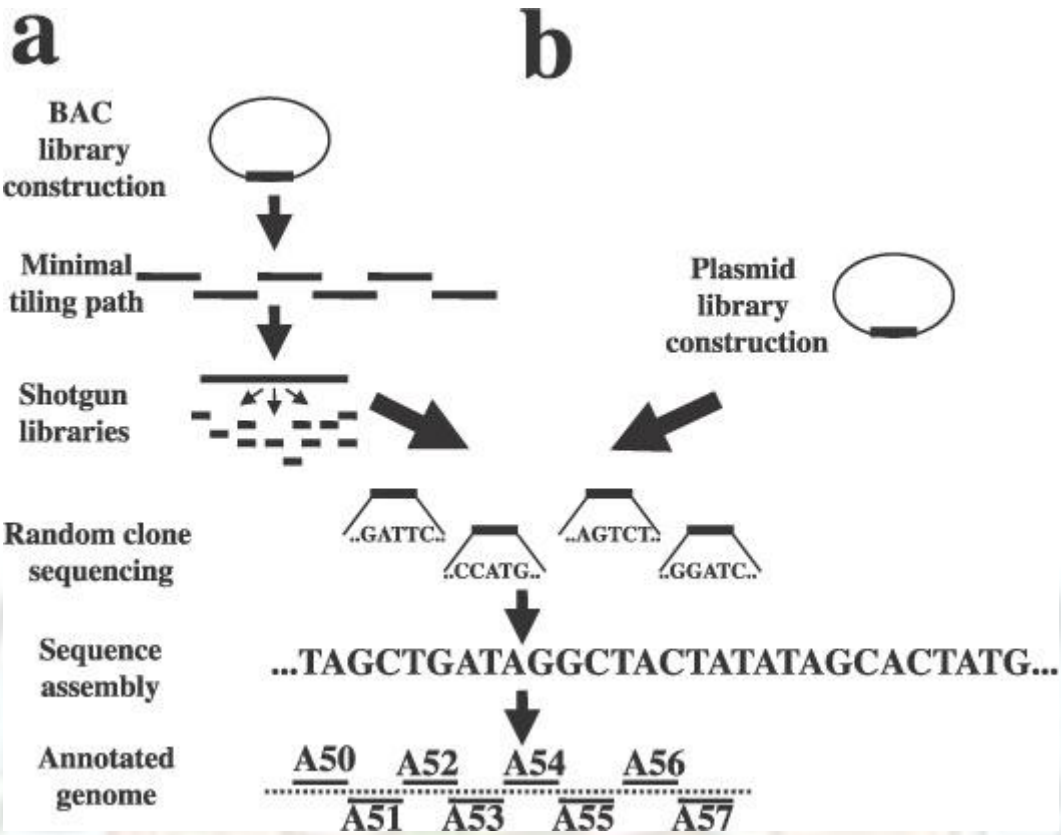


Figure: Whole genome sequencing a. Clone-by-clone approach; b. Shot-gun approach.

Source:http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0716-97602002000300013

(cc)

Maxam and Gilbert's method of sequencing

This method is also known as chemical degradation. It involves following main steps:

- Modification of a base.
- Its removal.
- Subsequent cleavage at that site.
- The pre labeled fragments are then analyzed on an autoradiograph.

The process involves the labeling of the DNA strand to be sequenced. The DNA is denatured and is divided into four reactions each having a different chemical treatment. The chemical modifies a base (A, T, G, and C), the modified base is removed and the fragment is cleaved into two at that position. The fragments are run on a gel in four different lanes. The chemical can be specific for A and G or only G. Similarly for C and T or only T. Thus the

fragments are separated on the basis of their size and the bases identified on the basis of the bands shown in the autoradiograph.

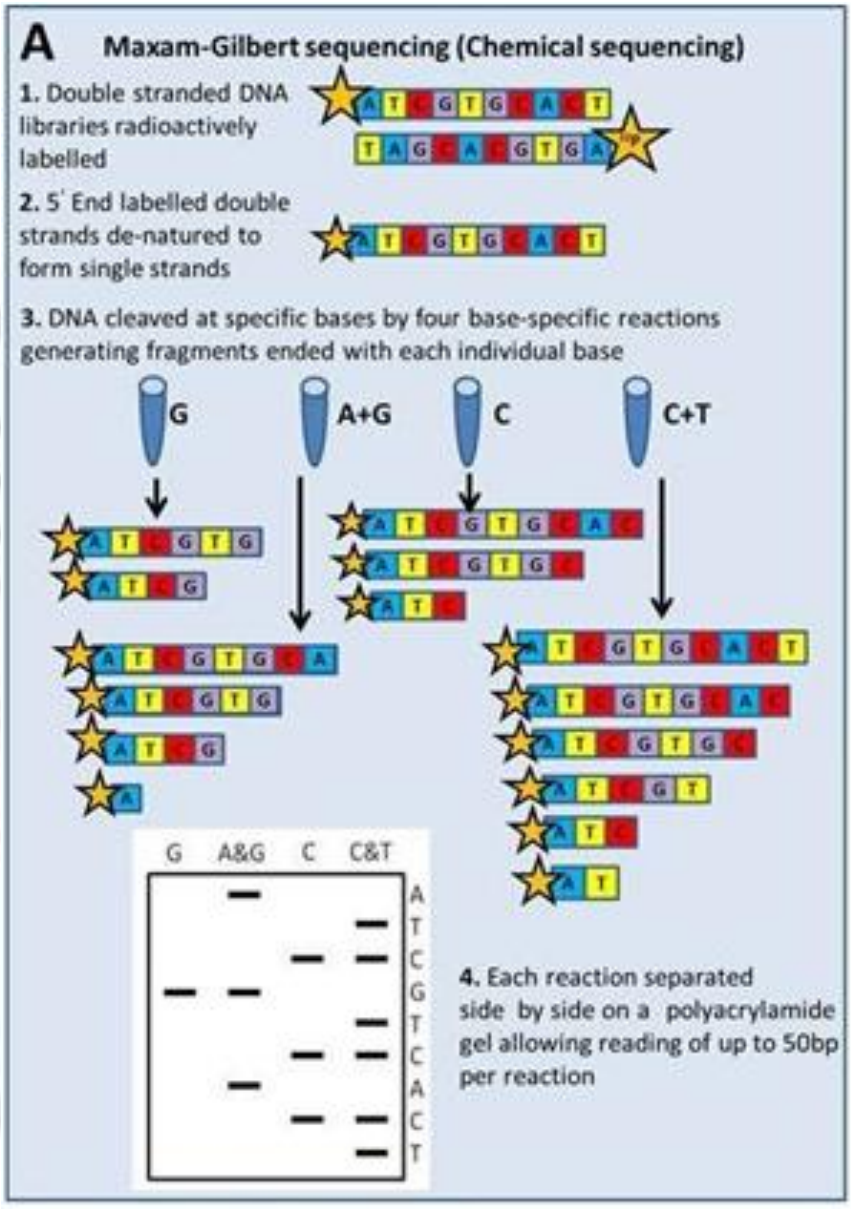


Figure: Maxam Gilbert sequencing method.

Source:http://www.oxbridgebiotech.com/review/research-and-policy/whats-so-special-about-next-generation-sequencing/ (cc)

Shotgun approach

It is also called random sequencing as there is no specific order of sequencing. The DNA is randomly fragmented and fragments are cloned into a vector. The different positive clones are sequenced and then assembled. The assembly of sequence is automated using assembly software.

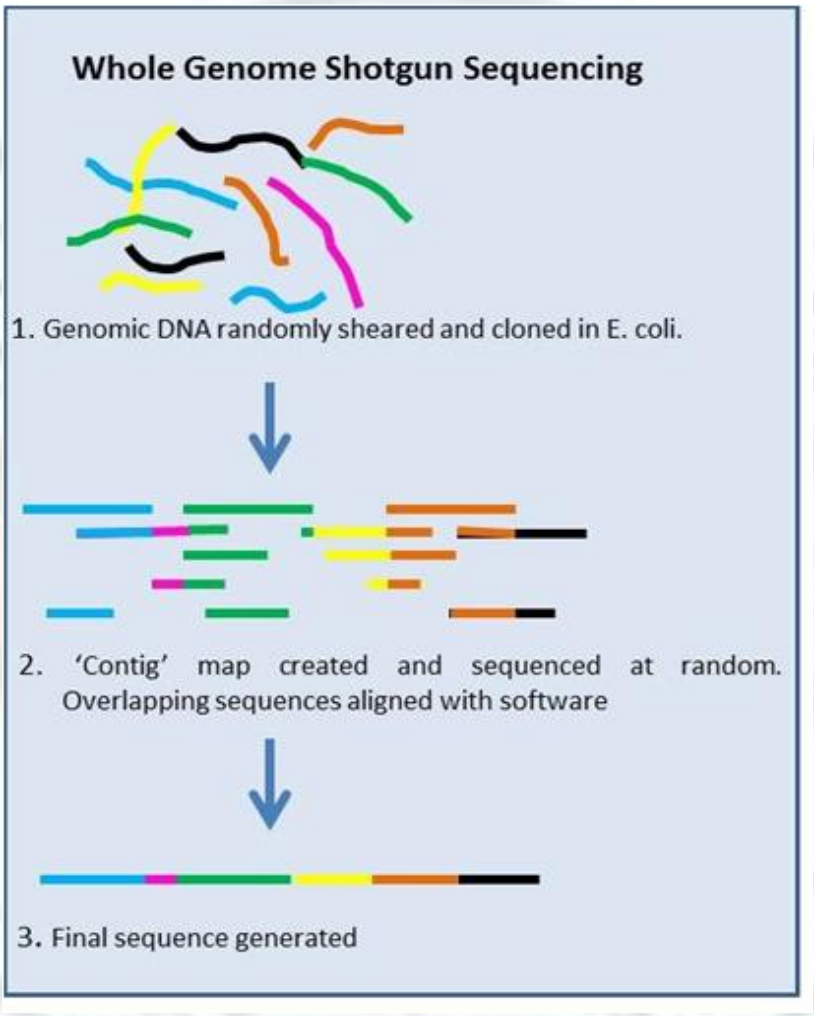


Figure: Shotgun approach of sequencing.

Source: http://www.oxbridgebiotech.com/review/research-and-policy/whats-so-special-about-next-generation-sequencing/ (cc)

Video: http://www.dnalc.org/view/15537-Shotgun-sequencing-and-dealing-with-repeat-sections-3D-animation-with-basic-narration.html(cc)

Advantages

- The method is optimized and maximally automated.
- Universal fluorescent labeled primers are available.

Disadvantages

- There is a possibility of obtaining gaps which can only be filled by direct sequencing.
- It has high redundancy (same fragment can be sequenced 5-6 times).

Primer walking

It is a direct approach where the labeled primer is designed from a known part of DNA sequence. The DNA fragment is cloned in a vector and sequenced. The second set of primer is designed for the new sequence generated and in the same direction.

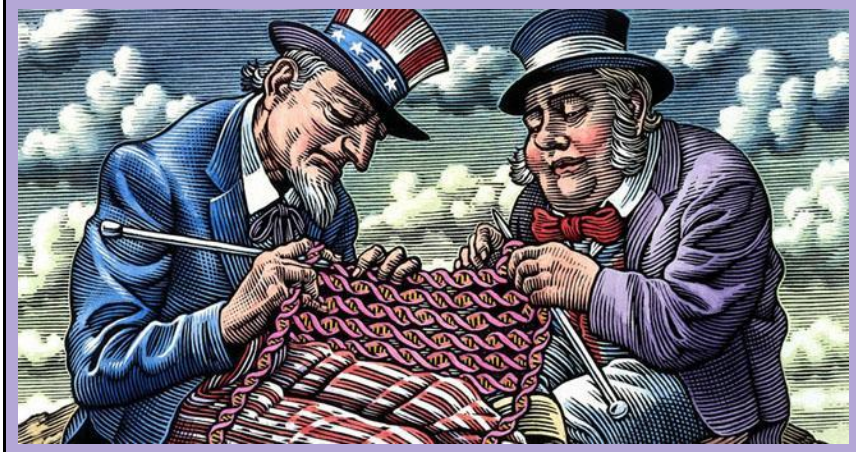
Advantages

- It has low redundancy.

Disadvantages

- It is costly and time taking.
- The number of primers to be designed varies with the size of fragment to be sequenced.

The human genome project



This figure has shown the American symbol Uncle Sam and the British symbol John Bull knitting strands of DNA in friendly collaboration. In 2000 the completion of the rough draft of human genome was announced by the American president and the British Prime Minister simultaneously.

Video: http://www.dnalc.org/view/15910-Sequencing-head-to-toe.html (cc)

Pyrosequencing

This method is based on the release of PPi during the DNA polymerization reaction.

$$(DNA)_n + dNTP \xrightarrow{\text{DNA polymerase}} (DNA)_{n+1} + PPi$$

$$PPi + APS \xrightarrow{\text{ATP sulphurylase}} ATP + SO_4^{-2}$$

$$ATP + \text{luciferin} + O_2 \xrightarrow{\text{luciferase}} AMP + PPi + \text{oxyluciferin} + CO_2 + h\nu$$

The DNA is denatured into ssDNA (single stranded DNA) and is added to a mixture containing DNA polymerase, adenosinesulphate, ATP sulphurylase, luciferin and luciferase.

As and when a nucleotide is added PPi is released. The PPi released is detected by the amount of light emitted by luciferase and corresponds to the nucleotide added in the reaction.

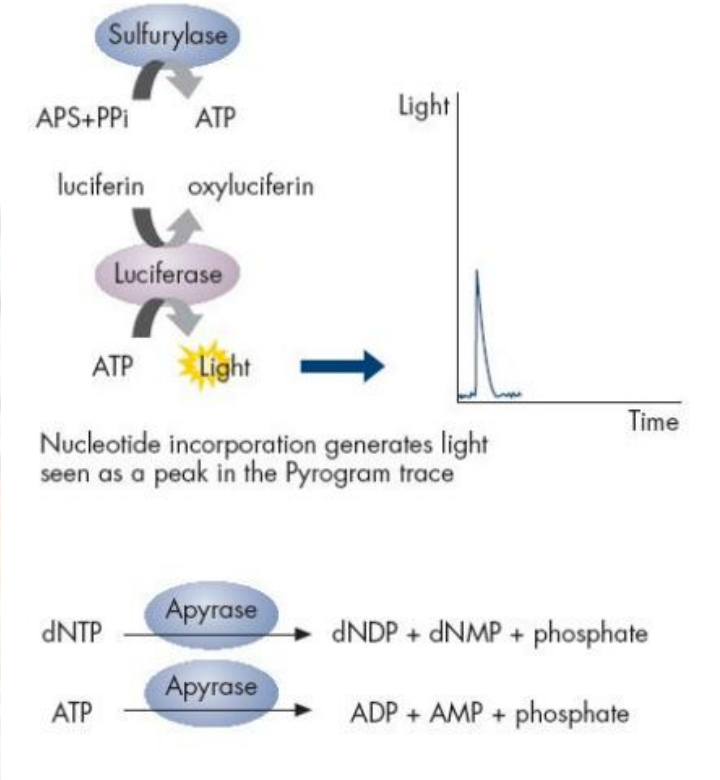


Figure: Principle of pyrosequencing.

Source: http://www.oxbridgebiotech.com/review/research-and-policy/whats-so-special-about-next-generation-sequencing/ (cc)

Advantages

- It overcomes the need of labeling primers, probes or DNA.
- Real time detection.
- Reaction can take place at room temperature and physiological pH.
- Multiple samples can be processed.
- Short DNA fragments can be sequenced.

Disadvantages

- Difficult to detect number of nucleotides in a homopolymeric reaction.
- Repeated washing off the nucleotides decreases the signal intensity.
- PPi contamination increases the background signal intensity.

- The incorporation fidelity of DNA polymerase is not high.
- For GC rich templates stability should be more.

Applications

- Secondary structure analysis.
- Detection of single nucleotide polymorphism.
- Detection of mutation.
- De novo sequencing of short DNA strands.

Single molecule sequencing using exonuclease

This method involves labeling the template DNA, cleaving the DNA by 3' exonuclease which releases one base at a time and the simultaneous detection of the cleaved base.

The four nucleotides of the template DNA strand are differently fluorescent labeled. The DNA is attached to a microsphere and denatured. The exonuclease enzyme cleaves the nucleotides sequentially at the 3' end. The cleaved nucleotide is detected by a focus laser beam.

Next Generation Sequencing (NGS)

With the advent of Sanger's method of sequencing deciphering the DNA and its pattern had become easy. But this had various disadvantages:

- The speed was less. Therefore, Human genome project took 10 years for completion.
- Low throughput.
- Low resolution.

To overcome these problems a high throughput method was introduced with the NGS. The principle is similar to the Sanger's method as the DNA is fragmented and is sequenced. But the sequence reaction is set up in large numbers parallel. The reads obtained from the reaction are then assembled and compared to a reference genome (resequencing) or a new sequence is generated (*de novo* sequencing).

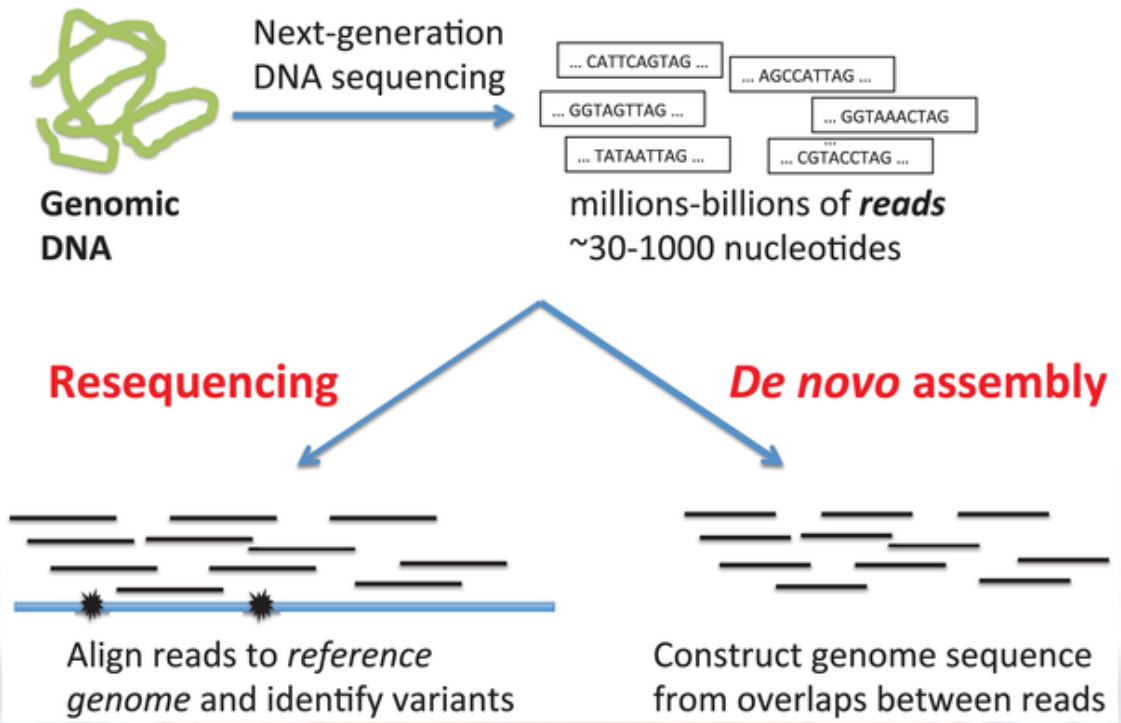
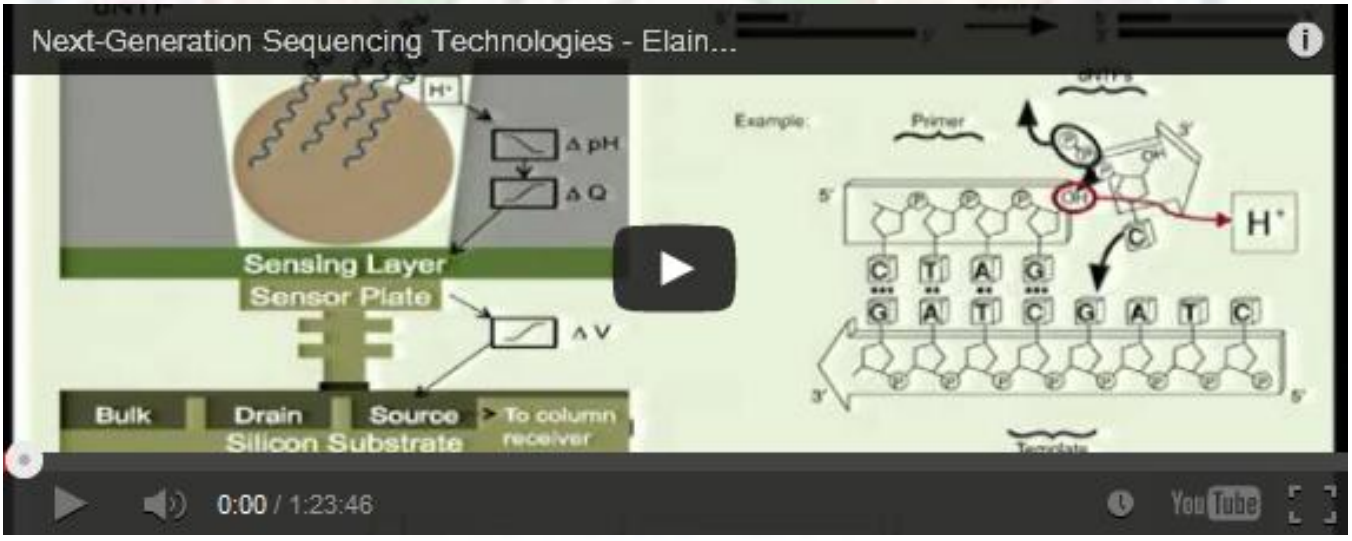


Figure: Methodolgy of NGS.

Source:http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1002821 (cc)



Video: Next Generation Sequencing.

Source:https://www.youtube.com/watch?v=P_MIF6zUeKko&feature=player_embedded (cc)

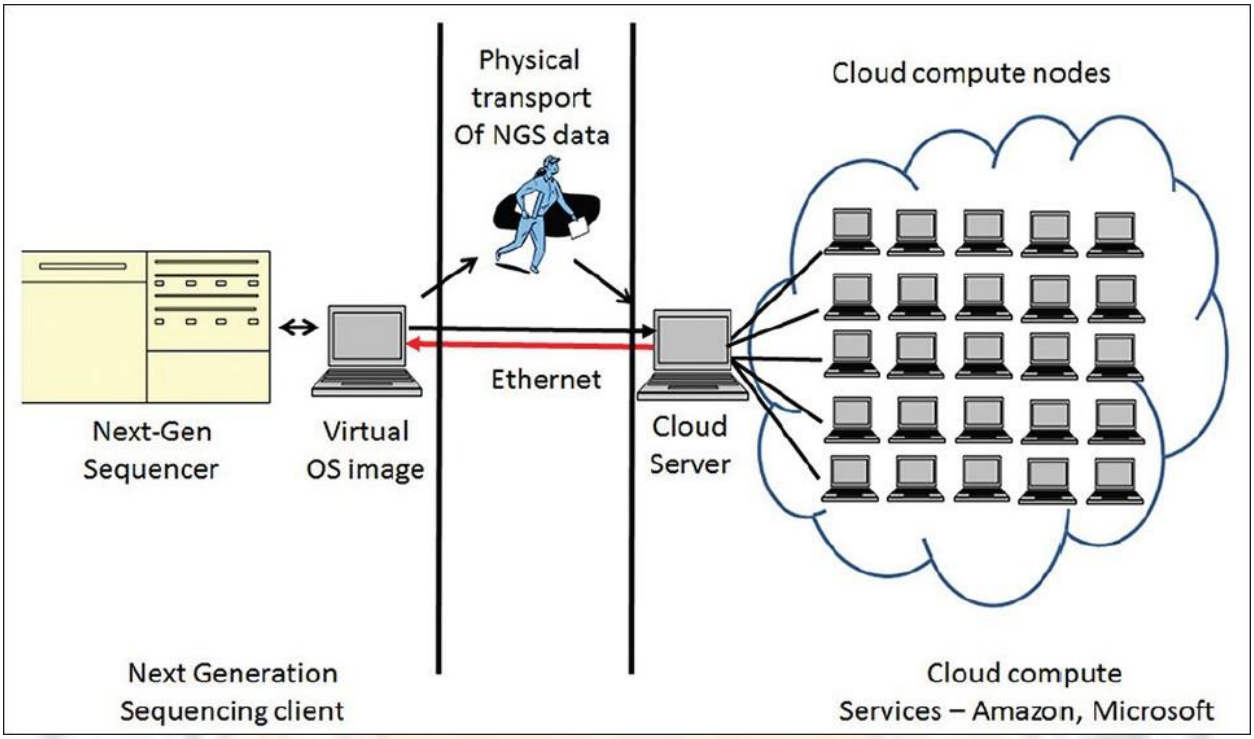


Figure: Analysis of NGS data using cloud computing.

Source:http://www.jpathinformatics.org/article.asp?issn=2153-3539;year=2012;volume=3;issue=1;spage=40;epage=40;aulast=Gullapalli (cc)

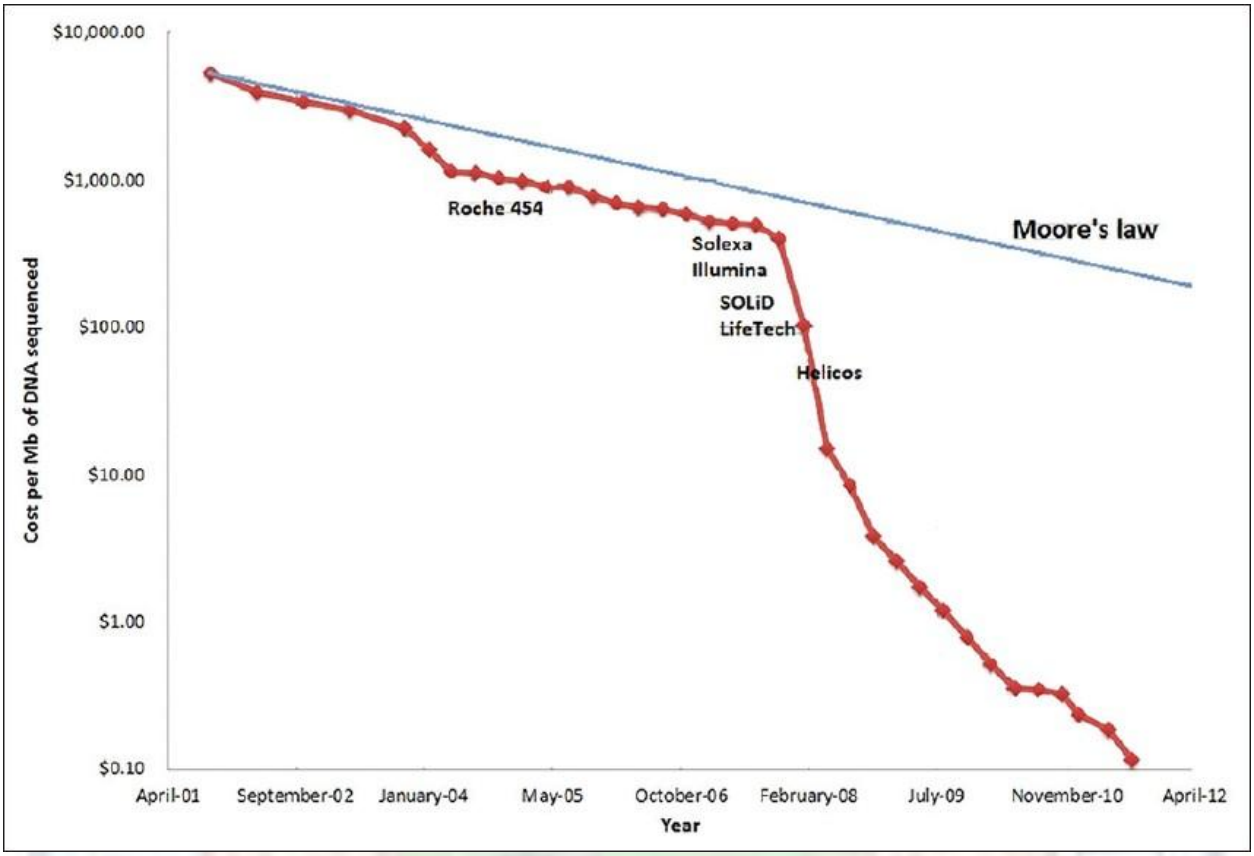


Figure: The graph showing cost per megabase of DNA in the last 10 years. It also shows the time of introduction of various sequencing methods.

Source:http://www.jpathinformatics.org/article.asp?issn=2153-3539;year=2012;volume=3;issue=1;spage=40;epage=40;aulast=Gullapalli (cc)

Animation: Methodology involved in next generation sequencing

Source: http://www.yourgenome.org/downloads/animations.shtml

Sequencers in NGS

1. Roche/454 FLX Pyrosequencer

This instrument applies the basic method of pyrosequencing. The polymerization of DNA leads to release of PPi which in turn activates enzyme luciferase which emits light. The amount of light emitted is proportional to the nucleotides added. The DNA is fragmented and attached to adaptors. The mixture is added to agarose beads which carry complimentary sequences to the 454 adaptors. The mixture is subject to PCR to amplify the fragments and the solution is then loaded onto the picotiter plate containing enzyme linked beads. This is a flow cell and is attached to a detector to monitor the light emitted in response to the polymerization reaction.

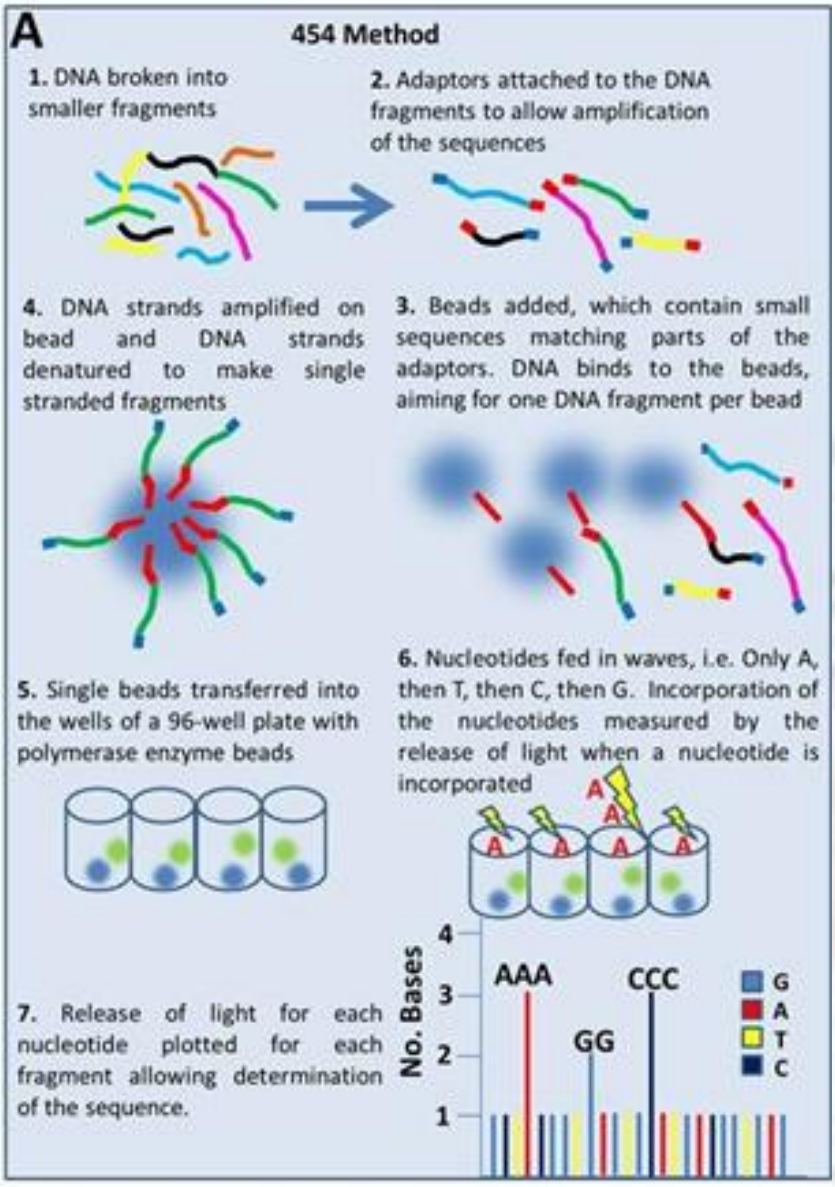


Figure: 454 method of NGS.

Source: http://www.oxbridgebiotech.com/review/research-and-policy/whats-so-special-about-next-generation-sequencing/ (CC)

2. Illumina Genome Analyzer

It is based on the sequencing by synthesis principle. The DNA is fragmented and attached to the adaptors. The DNA-adaptor complex is attached to the surface and labeled nucleotides with blocked OH group are released. The DNA complex is now denatured to made single stranded. The imaging detects the nucleotide incorporated. This step is repeated with

consecutive washings to remove the earlier nucleotides. Thus a cycle of events can be manually programmed to analyze the sequence of the template DNA

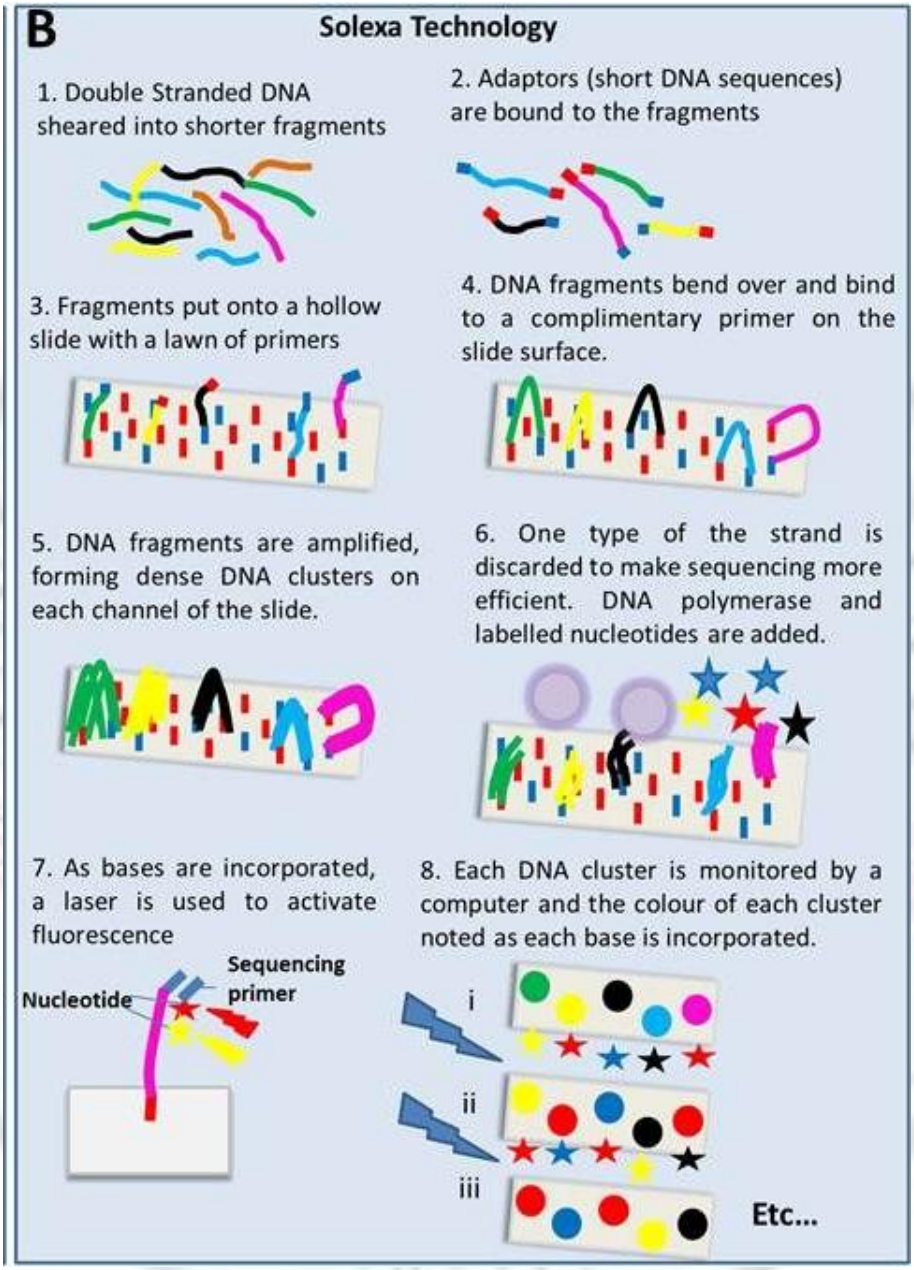


Figure: Steps in Solexa sequencing.

Source: http://www.oxbridgebiotech.com/review/research-and-policy/whats-so-special-about-next-generation-sequencing/ (CC)

3. Applied Biosystems SOLiDTM Sequencer

This instrument is based on ligation dependant sequencing. The DNA template is fragmented and attached to beads. The amplification is achieved by PCR. The beads are attached to a solid support and primers for the adaptor sequences are attached. The four labeled nucleotides are release in a dibase probe fashion. A series of ligation reaction occurs. Five sets of primers are used deleting the first base in every new set. Multiple cycles are performed depending on the sequence length. Each base is read twice based on the principle of two base encoding.

Advantages of NGS

1. High throughput

The multiplex sampling provides for screening a large number of samples in lesser time. The process is scalable and can be adjusted according to the need of a short or long sequence.

2. Tunable resolution

The resolution is described by the coverage. The coverage for a sample means the number of reads in a given cycle. The coverage and resolution can be adjusted according to the need of the scientist. It can be used to amplify a particular region for higher resolution or taking a larger area with lower resolution.

3. Wide range

The sensitivity of the method is high. It can be used to detect gene expression microarray data with high precision.

4. Universal tool

It is a universal tool for research and the data generated can be used for various purposes in biology.

5. Varied applications

The NGS can be used to generate DNA data that can be used for analysis in various research and medical field. It can be used to study the genome, transcriptome, small RNAome and the epigenome of different organisms.

Applications of NGS

NGS is being routinely used in research and disease diagnosis. Different platforms of sequencing are preferred for different objectives. The figures given ahead summarize the use of NGS in various diseases and a comparison of different NGS platforms with respect to different applications.

High-end sequencing- Platform†	Sequencing chemistry	Read lengths/ through put	Run time	Template prep	Application
Roche 454 -Titanium FLX	Pyrosequencing	400 bp 400 Mb/run	10 hours	Emulsion PCR	Denovo WGS of microbes, pathogen discovery, Exome seq
Illumina/Solexa -HiSeq 2000	Reversible terminator chemistry	2×100bp 600 GB/ run (dual cell)	11.5 days	Solid-phase	Human WGS, exome seq, RNA-seq, Methylation
ABI/LifeTechnology-SOLiD 5550XL	Sequencing by ligation	2×60bp 15 GB/day	8 days	Emulsion PCR	Human WGS, exome seq, RNA-seq, Methylation
HelicosBiotechnologies	Reversible Terminator chemistry	25-55 bp 28 GB/run (avg)	>1 GB/hour	Single molecule	Human WGS, exome seq, RNA-seq, Methylation
Roche 454- GS Junior	Pyrosequencing	400 bp 50 Mb/run	10 hours	Emulsion PCR	Denovo WGS of microbes, pathogen discovery, Exome seq
Illumina/Solexa- MiSeq	Reversible terminator chemistry	2×150bp 1.0-1.4 Gb	26 hours	Solid-phase	Microbial discovery, Exome seq, Targeted capture
ABI/ Lifetechnology- Iontorrent	H+ Ion sensitive transistor	320 Mb/run	8 hours*	Emulsion PCR	Microbial discovery, Exome seq, Targeted capture

*Sample preparation – 6 hours, sequencing time – 2 hours, †Data shown here represent the highest figures currently available on the company website and is highly likely to change by the time this article is published

Figure: Popular NGS methods available.

Source:http://www.jpathinformatics.org/article.asp?issn=2153-3539;year=2012;volume=3;issue=1;spage=40;epage=40;aulast=Gullapalli (cc)

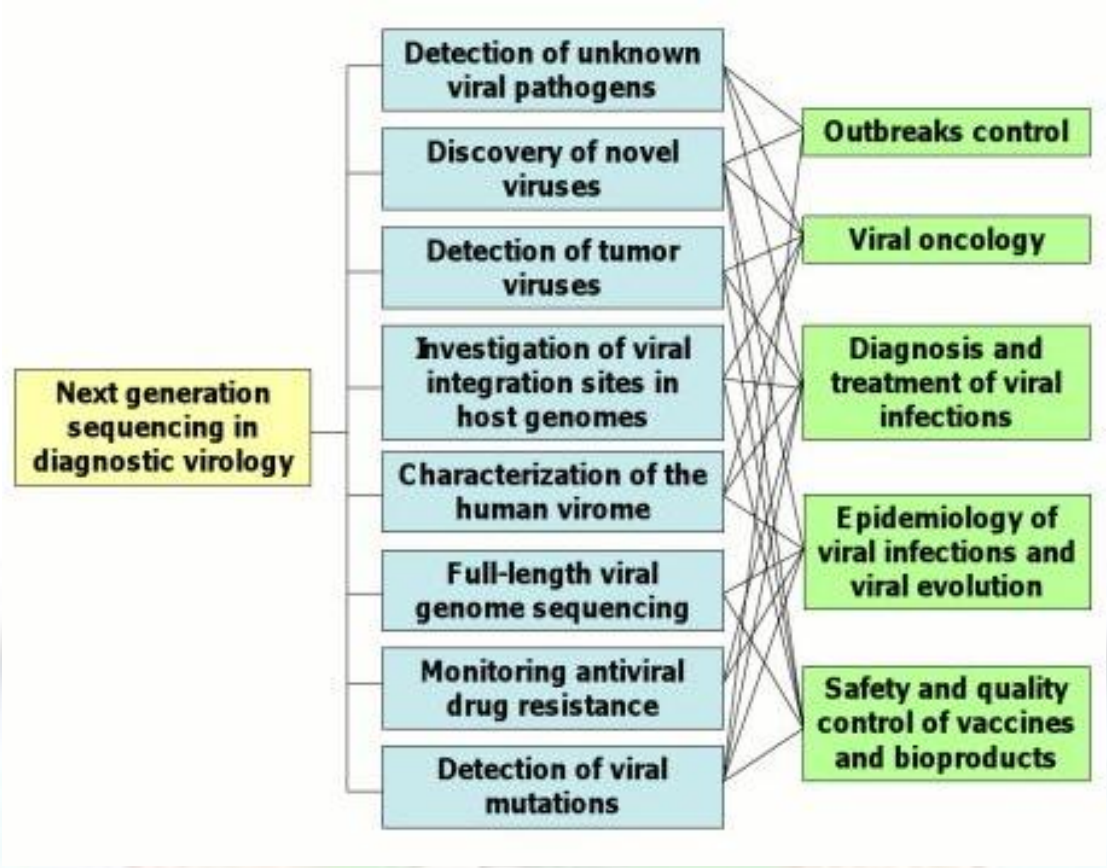


Figure: Application of NGS in diagnostic virology

Source: http://www.mdpi.com/1422-0067/12/11/7861 (cc)

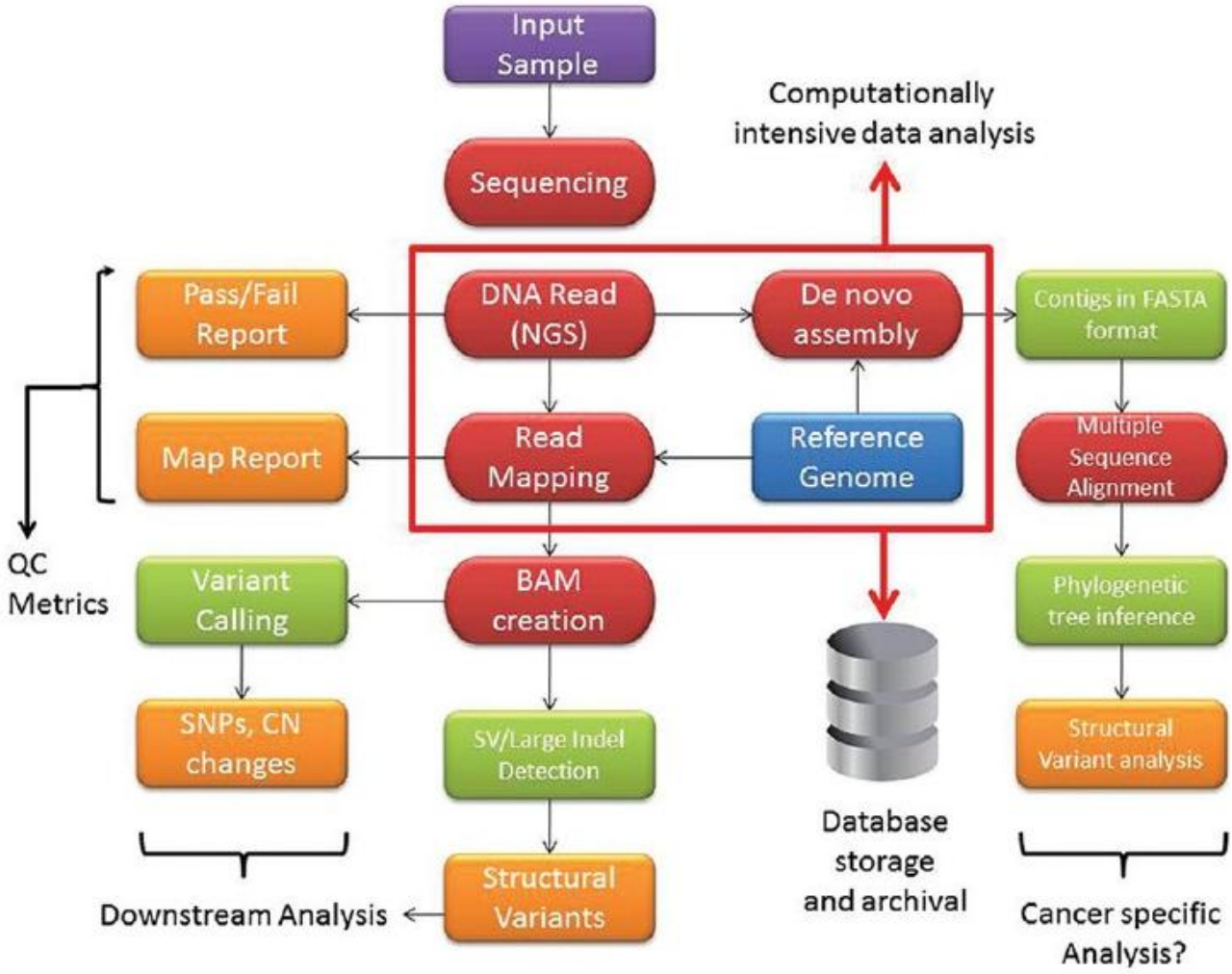


Figure: Application of NGS in cancer diagnosis

Source:http://www.jpathinformatics.org/article.asp?issn=2153-3539;year=2012;volume=3;issue=1;spage=40;epage=40;aulast=Gullapalli (cc)

HTS applications		Study highlights	Pathogens/Sample	Platform
Culture-dependent	Bacterial genomic epidemiology	feasibility study in a hospital context: improving genetic resolution over common genotyping strategies	*S. aureus*/clinical samples *C. difficile*/fecal samples	Illumina MiSeq
		pilot study: investigating an outbreak and current limitations for routine use	multidrug-resistant *E. coli* /rectal swab	PGM Ion Torrent
		WGS data exploring MLST: toward a standardized analysis	*C. jejuni* and *C. coli*	Illumina HiSeq 2000
		WGS to rapidly highlight antibiotic resistance determinants	*A. baumannii* /tracheal samples	454-Titanium and Solid version 4
	pathogen evolution	high-resolution genotyping by HTS allowing new insights about an emerging pathogen	methicillin-resistant *S. aureus* /clinical isolates	Illumina GA IIx
		Recombination-filtered core genome to understand pathogen adaptation	*E. faecium*/isolates from hospitalized patients	Illumina GA IIx
Culture-independent	Community profiling	proof-of-principle: metagenomics data could be integrated in a diagnosis of cystic fibrosis	airway microbiota in cystic fibrosis /mucolysed sputa	PGM Ion Torrent
		large-scale study monitoring resistance genes in human gut microbiota	gut microbiota	Illumina GA IIx
	Clinical metagenomics and pathogen discovery	a metagenomics approach to avoid pathogen culture and isolation	Shiga-toxigenic *E. coli* /stool samples	Illumina HiSeq 2500 and MiSeq
		an unbiased method to detect viral pathogens	viral pathogens/ nasopharyngeal samples	Illumina GA IIx
	Single-cell microbiology	first evidence of a genome capture from a single cell in a clinical context	*P. gingivalis* /sink drain	Illumina GA IIx
		Immunomagnetic separation for targeted bacterial enrichment with multiple displacement amplification	*C. trachomatis* /cervical or vaginal swab	Illumina GA IIx and HiSeq

Figure: Application of NGS in clinical diagnostics.

Source: http://www.mdpi.com/2076-0817/3/2/258 (cc)

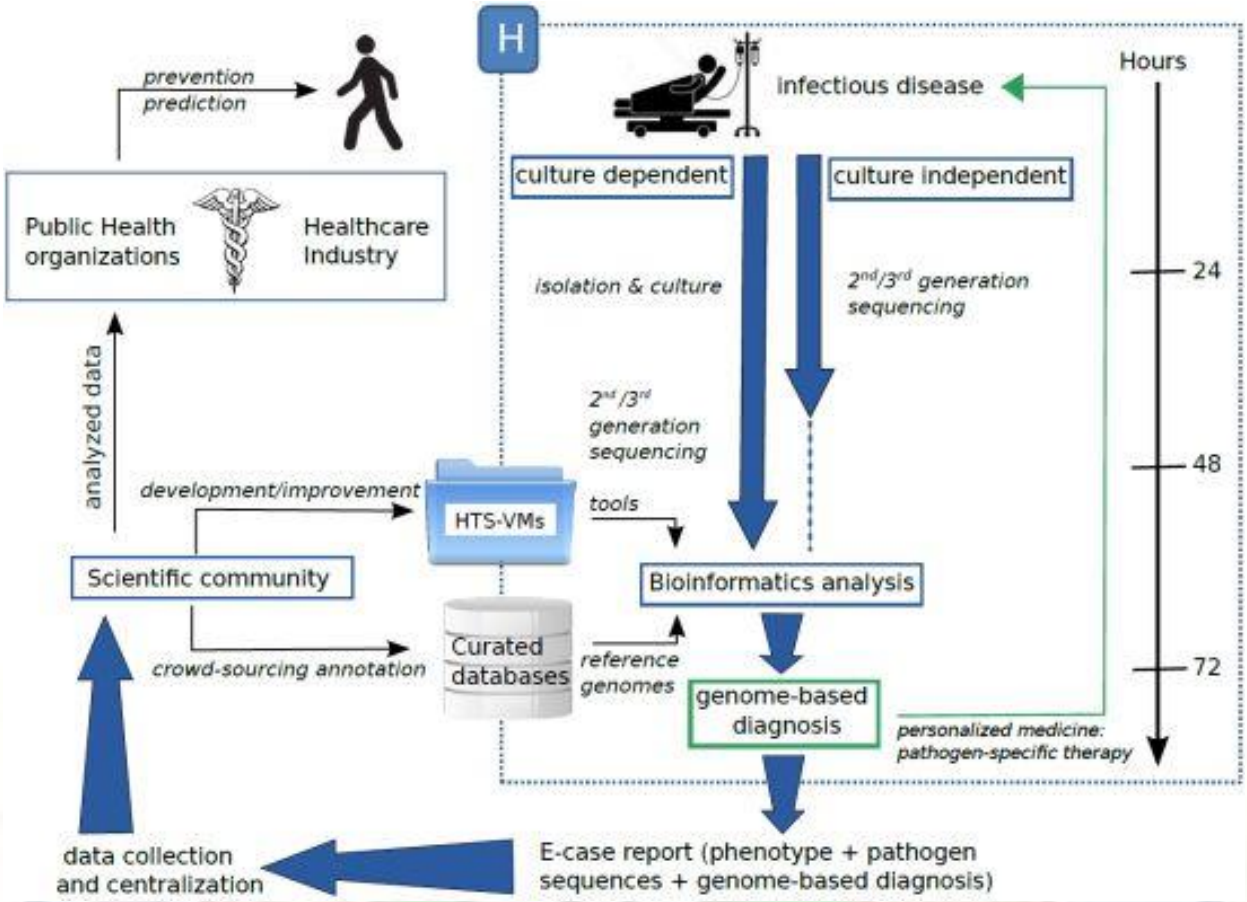


Figure: A virtual machine project for global data analysis in hospitals.

Source: http://www.mdpi.com/2076-0817/3/2/258 (cc)

Nanopore sequencing

It's the latest technique for real time sequencing of DNA. The DNA strand to be sequence is passed through a protein and is sequenced in real time.

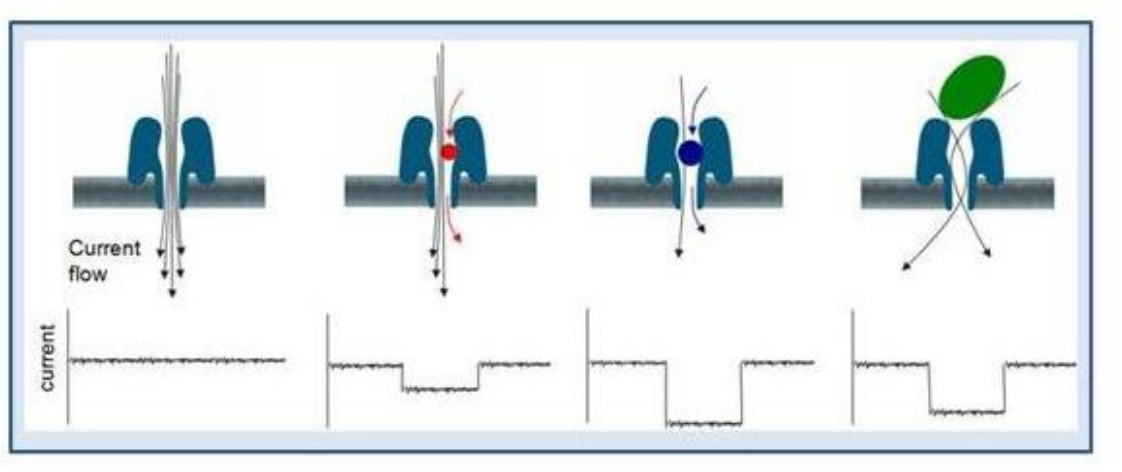


Figure: Protein nanopores are set in an electrically resistant membrane bilayer. An ionic current is passed through the nanopores, and if an analyte passes through the pore or near its aperture, a characteristic disruption of current is created . By measuring that current, it is possible to identify the molecule in question. During sequencing, for example, a DNA strand is fed through the nanopore by an enzyme and each of the four standard DNA bases G, A, T and C can be identified.

Source:http://www.oxbridgebiotech.com/review/research-and-policy/whats-so-special-about-next-generation-sequencing/ (cc)

Nanoball Sequencing method

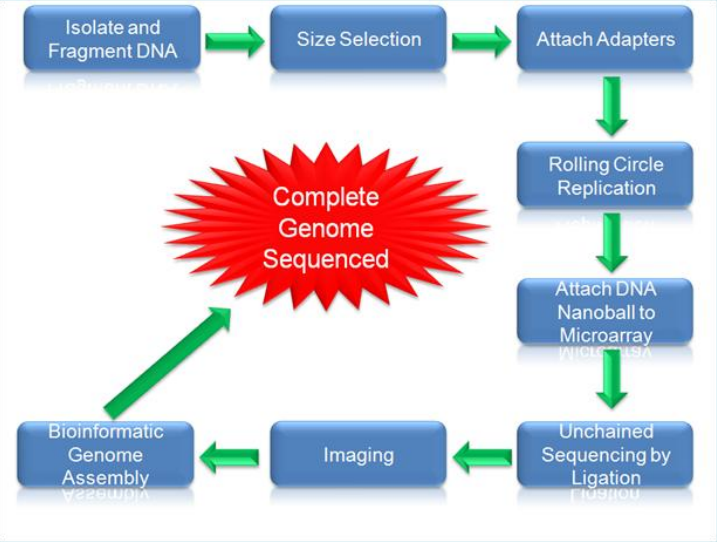


Figure: Nanoball sequencing method

Source: http://en.wikipedia.org/wiki/DNA_nanoball_sequencing

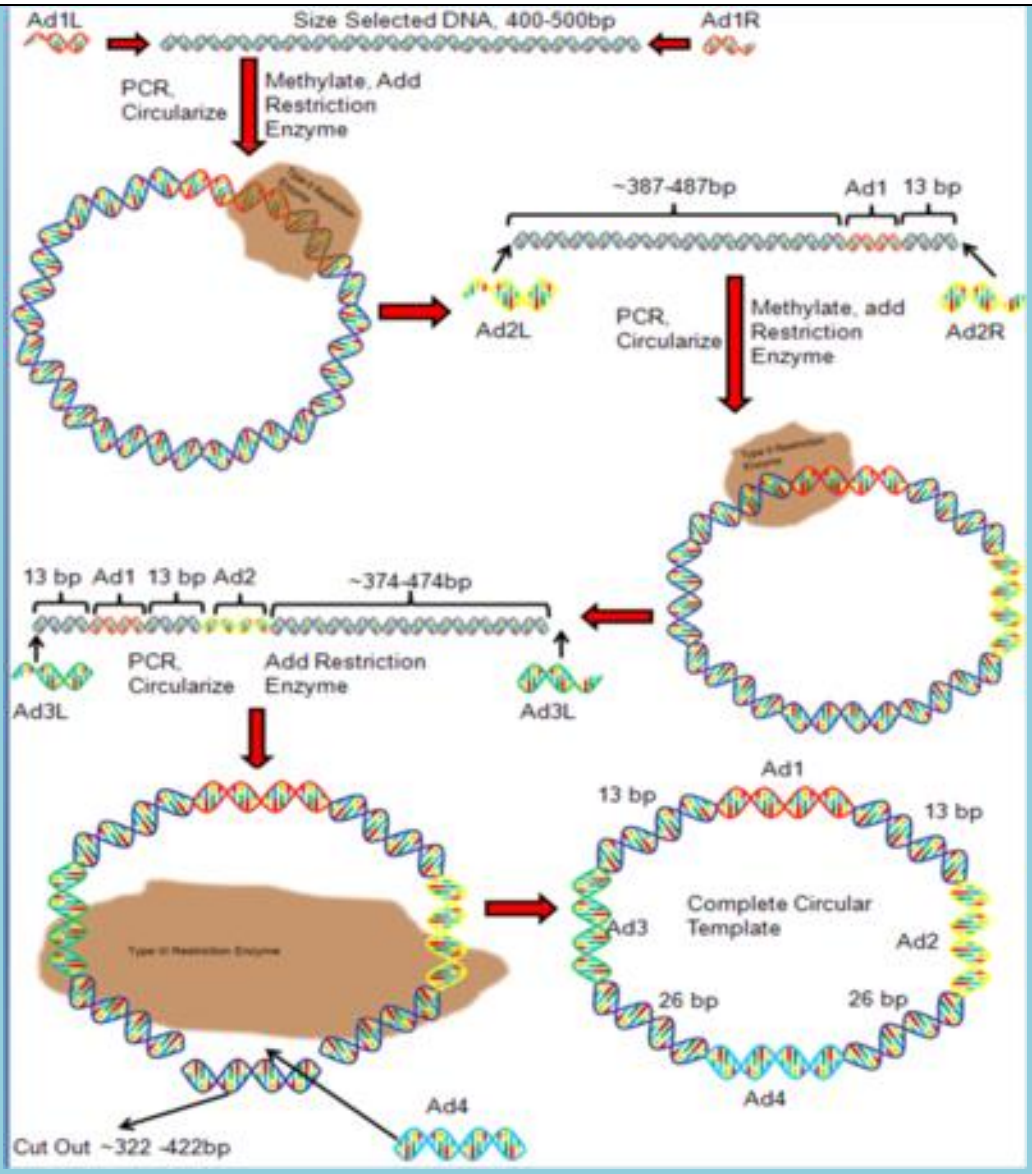


Figure: The construction of nanoball for sequencing

Source: http://en.wikipedia.org/wiki/DNA_nanoball_sequencing (cc)

This method allows sequencing of multiple nanoballs at a time. This method has a limitation of reading small lengths of DNA.

It involves DNA fragments of 500-600bp in length to be attached to adaptor sequences. The nanoballs are then passed through a sequencing flow cell and the data is analyzed.

An interactive sequencing game:

http://www.dnalc.org/view/15891-DNA-sequencing-game-interactive-2D-animation.html
(cc)

Summary

Video Lecture: http://media.hhmi.org/ibio/weissman/weissman_1.mp4

To summarize the chapter we would conclude that Next Generation Sequencing though based on the Sanger's method is a high throughput method of sequencing. It involves less time and is cost effective.

The sequencing principle involves the chain elongation of DNA by adding modified or labeled dNTPs/ddNTPs which can either terminate the reaction or can be detected.

The early method of Sanger's sequencing had the following disadvantages:

- Low throughput
- Time taking
- Costly

These were overcome by the new Next Generation Sequencing which also gave the advantage of:

- Wider range
- Versatile applications

Exercise

- What is the principle of Sanger's sequencing method?
- What are the advantages of Next Generation Sequencing?
- What is meant by chemical degradation in Maxam Gilbert's method of DNA sequencing?
- What is the principle of pyrosequencing?
- What is the difference between Sanger's method and Maxam and Gilbert's method of sequencing?
- What are the different methods of sequencing?
- What are the applications of Next Generation Sequencing?

Glossary

Chain termination: This method involves addition of a nucleotide lacking the 3'OH group (dideoxy) which cannot bind to an additional nucleotide thus terminating the chain elongation.

Complimentary DNA: The chain of nucleotides that can form a double stranded structure by base pairing A-T G-C.

DNA deoxyribonucleic acid: It's a biomolecule that constitutes the gene and the genetic material of an organism.

dNTPs: deoxynucleotide triphosphate molecules used for DNA synthesis.

DdNTP: dideoxy form of sugar which is not capable of chain elongation due to the blockage of OH group.

Exonuclease: enzyme that cleaves DNA strand base by base.

Fragmented DNA: The DNA is large in size and cannot be downstreamed. It is cut at specific sites using restriction endonucleases.

Next generation sequencing: high throughput method of DNA sequencing.

Restriction endonuclease: A proteins that identifies the DNA sequence and cuts the DNA at that specified point.

Resequencing: Sequencing of a Modified DNA sample and its comparison to a reference genome eg mutational studies.

RNA ribonucleic acid: It's a biomolecule that forms the basis of the expression of a phenotype of an organism.

References

Green, Michael R., and Joseph Sambrook. *Molecular cloning: a laboratory manual*. New York: Cold Spring Harbor Laboratory Press, 2012.

Mardis, Elaine R. "Next-generation DNA sequencing methods." *Annu. Rev. Genomics Hum. Genet.* 9 (2008): 387-402.

Shendure, Jay, and Hanlee Ji. "Next-generation DNA sequencing." *Nature biotechnology* 26.10 (2008): 1135-1145.

http://res.illumina.com/documents/products/illumina_sequencing_introduction.pdf

http://web.stanford.edu/class/cs262/notes/lecture11.pdf

http://fas.org/sgp/othergov/doe/lanl/pubs/00326703.pdf

Web Links

https://www.nanoporetech.com/technology/analytes-and-applications-dna-rna-proteins/dna-an-introduction-to-nanopore-sequencing

https://www.nanoporetech.com/technology/the-minion-device-a-miniaturised-sensing-system/the-minion-device-a-miniaturised-sensing-system

http://res.illumina.com/documents/products/illumina_sequencing_introduction.pdf

http://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-next-generation-dna-sequencing/illumina-

http://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-next-generation-dna-sequencing/454-seque

http://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-next-generation-dna-sequencing/ion-torre

http://www.wellcomecollection.org/explore/life-genes--you/topics/genetics/articles/human-genome-quick-facts.aspx

http://dnaftb.org/sitemap.html
http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1002821