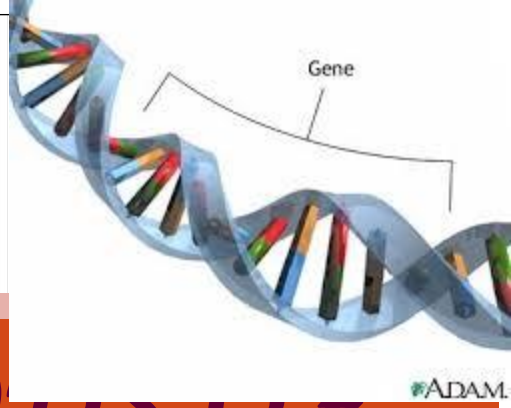
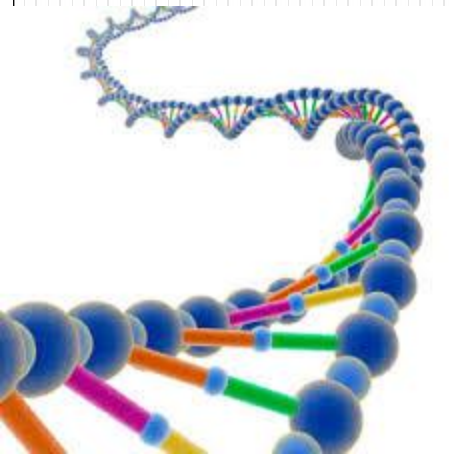


DIFFERENT METHODS OF GENE SEQUENCING



DURGESH SIROHI

CONTENTS

- Introduction
- DNA Sequencing
- What is a Gene Sequencing?
- History
- Principle
- Different methods of Gene Sequencing

1. Classical methods of Gene Sequencing

i. Maxam and Gilbert's Chemical Degradation Method
ii. Sanger's Dideoxynucleotide Synthetic Method

2. Advanced methods

i. Birdge PCR
ii. Shotgun sequencing

3. New generation methods of Gene Sequencing

i. Sequencing by Hybridization
ii. Sequencing by Synthesis
iii. Pyrosequencing
iv. Nanopore DNA Sequencing
v. Jon Torrent Semiconductor Sequencing
vi. Illumina (Solexa) Sequencing
vii. Transmission Electron Microscopy for DNA Sequencing

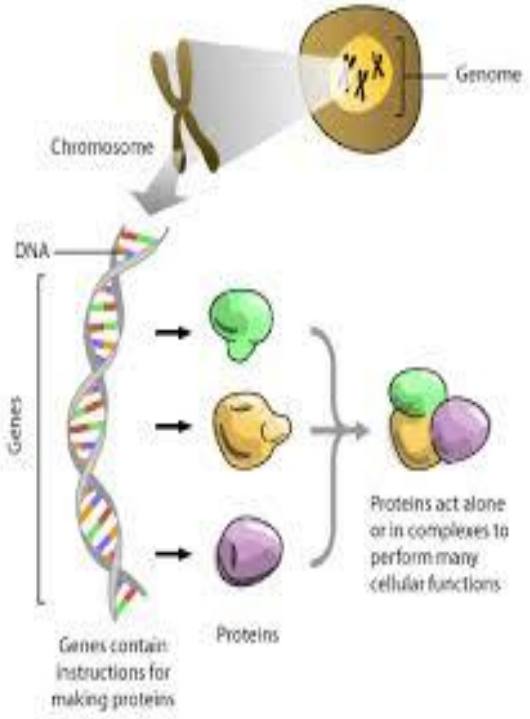
- Uses of Sequencing
- Conclusion
- References

Introduction

Gene is a hereditary unit consisting of a sequence of DNA that occupies a specific location on a chromosome and determines a particular characteristic in a organism. Genes undergo mutation when there DNA sequence changes.

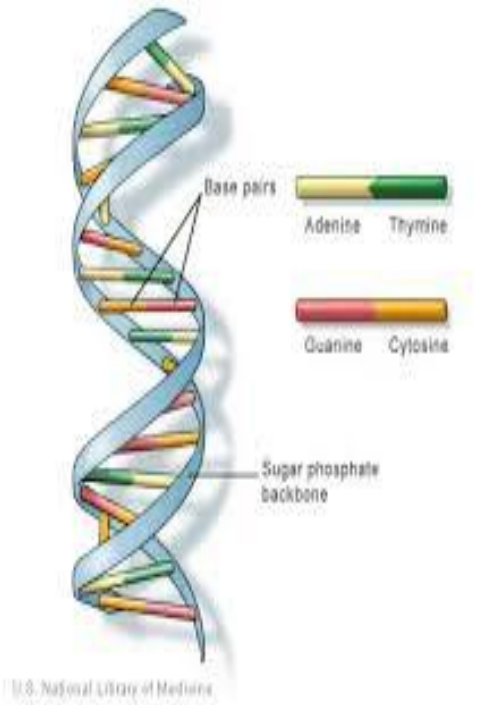
Gene or DNA sequencing is the determination of the order of bases in the sample of DNA. Probably the most important technique available to the molecular biologist is DNA sequencing, by which the precise order of nucleotides in a piece of DNA can be determined.

Gene sequencing technologies are needed to study the structure of cloned genes and whole genomes. It is therefore necessary that we have high throughput low-cost sequencing technologies for basic and applied research.



DNA sequencing

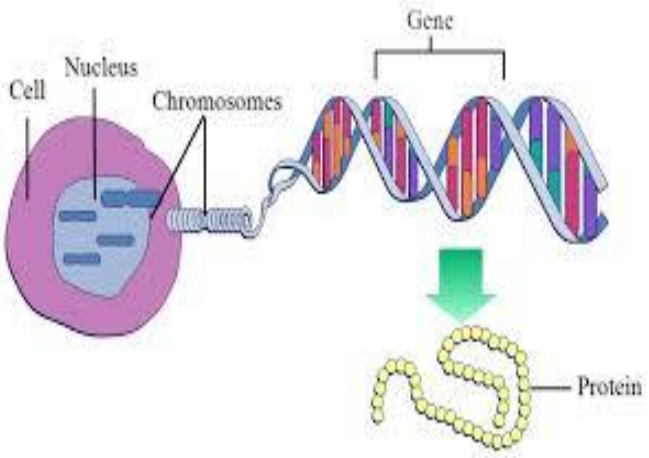
DNA Sequencing is the process of determining of the precise order of nucleotides with in a DNA molecule. It includes any method or technology that is used to determine the order of four bases –adenine, guanine, cytosine and thymine-in a strand of DNA.



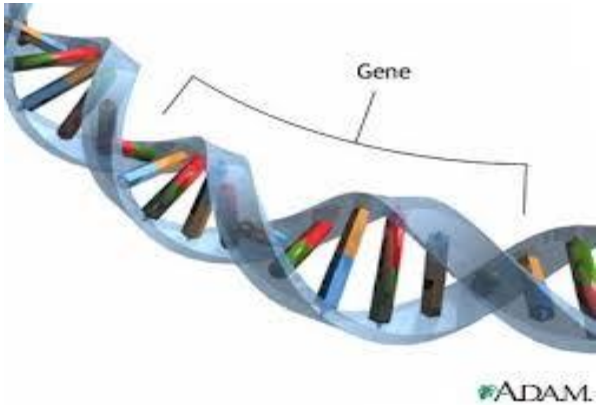
http://en.wikipedia.org/wiki/DNA Sequencing

iat is a Gene Sequencing?

- *Gene Sequencing is determining the exact sequence of nucleotide or bases in strand of DNA to better understand the behaviour of a gene.*
- **Gene Sequencing = DNA Sequencing = Nucleotide Sequencing = Base Sequencing**
- *Gene Sequencing is the determination of the order of nucleotides or bases in the sample of DNA. It is the reading of the genetic code.*



However, not all DNA sequences are genes(i.e.coding regions) as there may, depending on the organism and the source of the DNA sample, also be promoters, tandem repeats, introns, etc.



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

History

Though the structure of DNA was established as a double helix in 1953, several decades would pass before fragments of DNA could reliably analyzed for their sequence in the laboratory. RNA sequencing was one of the earliest forms of nucleotide sequencing. The major landmark of RNA sequencing is the sequence of the first complete genome of Bacteriophage MS2, identified and published by Walter Fiers and his coworkers at the University of Ghent in 1972 and 1976.

Several notable advancements in DNA sequencing were made during the 1970s. Frederick Sanger developed rapid DNA sequencing methods at the MRC Centre , Cambridge, UK and published a method for "DNA sequencing with chain- terminating inhibitors" in 1977. Walter Gilbert and Allan Maxam at Havard also developed sequencing methods, by including one for "DNA sequencing by chemical degradation".

Several new methods for DNA sequencing were developed in the mid to late 1990s . These techniques comprise the first of the "next-generation" sequencing methods. In 1996, Pal Nyren and his student Mostafa Ronaghi at the Royal Institute of Technology in Stockholm published their method of pyrosequencing.

Principle of Gene Sequencing

- The process of determining the order of the nucleotide bases along a DNA strand is called sequencing.
- In 1977, twenty-four years after the discovery of the structure of DNA, two separate methods for sequencing DNA or gene was developed: the chain termination method and the chemical degradation method.
- Both the method are equally important to begin with, but for many reasons, **the chain termination method** is the method more commonly used today. This method is based on the **principle that single-stranded DNA molecules that differ in length by just a single nucleotide can be separated from one another using polyacrylamide gel electrophoresis.**

- The DNA to be sequenced, called a **template DNA,** is first prepared as a single-stranded DNA. Next a short nucleotide is annealed or joined, to the same position an each template strand. The oligonucleotide act as a primer for the synthesis of new DNA strand that will be complimentary to the template DNA. This techniques requires the four nucleotide-specific reactions—one each four G,A,C and T-be performed on four identical samples of DNA.
- The four sequencing reactions require the addition of all the components necessary to synthesize and label new DNA, including:

- A DNA template;
- A primer tagged with a mildly radioactive molecule
- DNA polymerase—an enzyme that derives the synthesis of DNA
- Four deoxynucleotides (G, A, C,T); and
- One dideoxynucleotide, either ddG, ddA, ddC, or ddT

Different Methods of Gene Sequencing

- Classical/Basic Methods
- Advanced Methods
- New Generation Sequencing Methods

Classical/Basic Methods

Once a gene or a DNA fragment has been cloned its further study involves gene sequencing. The technique of DNA sequencing was very laborious, till 1977, when a breakthrough was made in DNA sequencing methods. Two different methods for determination of DNA sequences was published in **1977**, although later one of them proposed by Frederick Sanger became poplular and used everywhere.

- Maxam and Gilbert's Chemical Degradation Method
- Sanger's Dideoxynucleotide Synthetic Method

Maxam and Gilbert's Chemical Degradation Method.

Maxam-Gilbert sequencing(chemical cleavage method using double stranded(ds) DNA)

The chemical degradation, in which sequence of a ds DNA molecule is determined by treatment with chemicals that cut the molecule at specific nucleotide positions.

Allan and **Walter Gilbert** published a DNA sequencing method in 1977 based on chemical modification of DNA subsequent cleavage at specific bases.. In this method following steps involved:-

i. Label the 3' ends of DNA with ^{32}P.

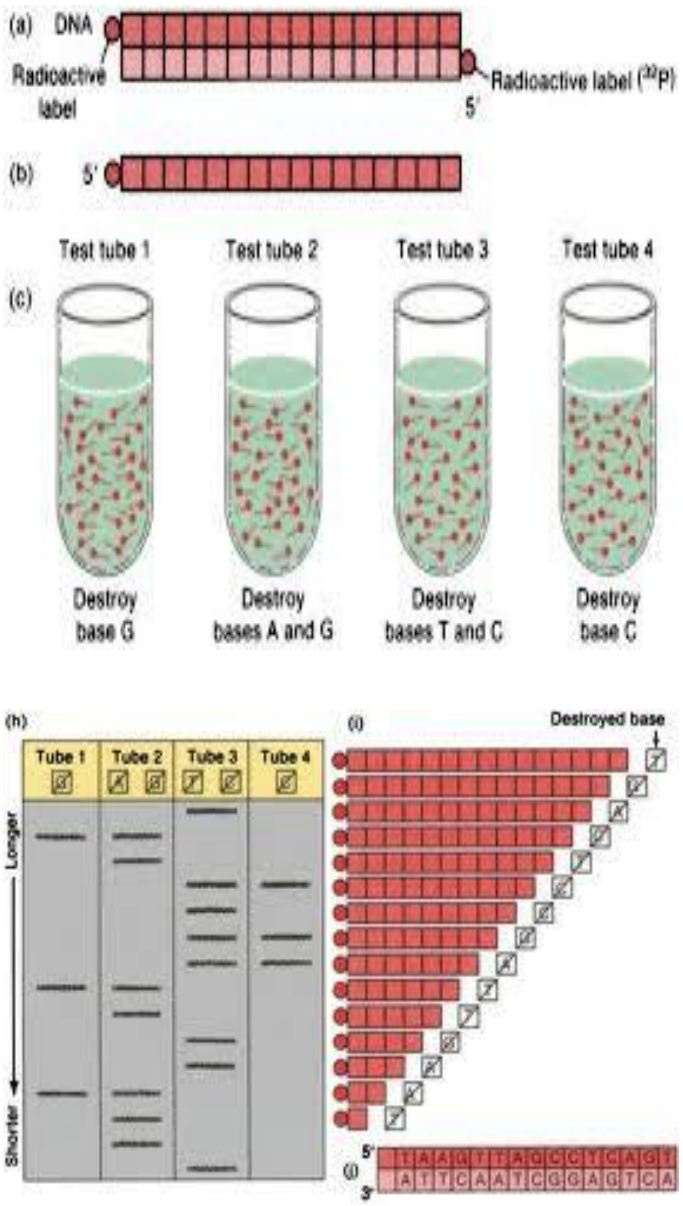
ii. Separate the two strands.

iii. Divide the mixture in four samples, each treated with a different reagent having the property of destroying either only G, or only C, or 'A and G' or 'T and C'. The concentration of reagent is so adjusted that 50% of target base is destroyed, so that fragments of different sizes having ^{32}P are produced.

iv. Electrophorese each of the four samples in the four different lanes of the gel.

v. Autoradiograph the gel and determine the sequence from positions of bands in four lanes.

Gupta, P.K.(2010). *"Plant Biotechnology"*, 1st Edition. Rastogi Publications, Meerut.



Different steps involved in Maxam and Gilbert's method of a DNA sequencing

Sanger's Dideoxynucleotide Synthetic Method

Sanger-Coulson sequencing(chain termination method using single-stranded (ss) DNA)

The chain termination method (Sanger et al.,1977), in which the sequence of a ss DNA molecule is determined by enzymatic synthesis of complementary polynucleotide chains , these chains terminating at specific nucleotide positions.

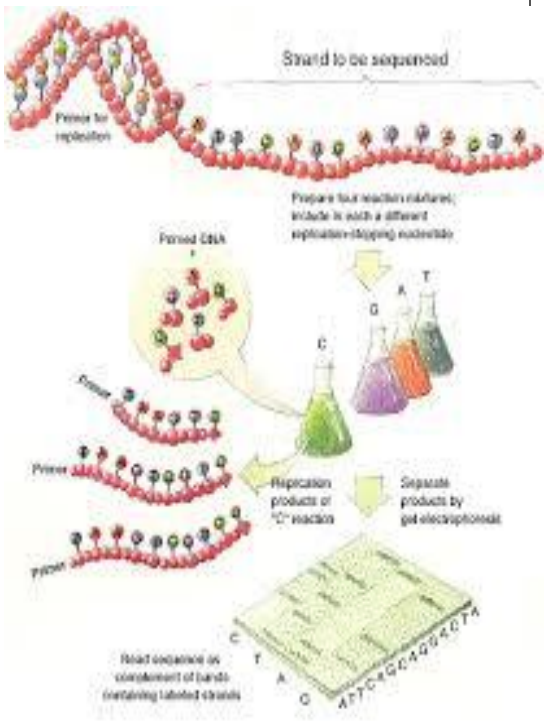
It is based on the principle that ss DNA molecules that differ in length by just a single nucleotide can be separated from one another by polyacrylamide gel electrophoresis.

i. The first step is to anneal a short oligonucleotide to the same position on each molecule, this oligonucleotide acting as a primer for synthesis of a new DNA strand that is complementary to the template.

ii. The strand synthesis reaction, which is catalyzed by a DNA polymerase enzyme and requires the four deoxyribonucleotide triphosphates (dNTPs-dATP, dCTP, dGTP, dTTP) and as substrates, would normally continue untill thousands nucleotides had been polymerized.

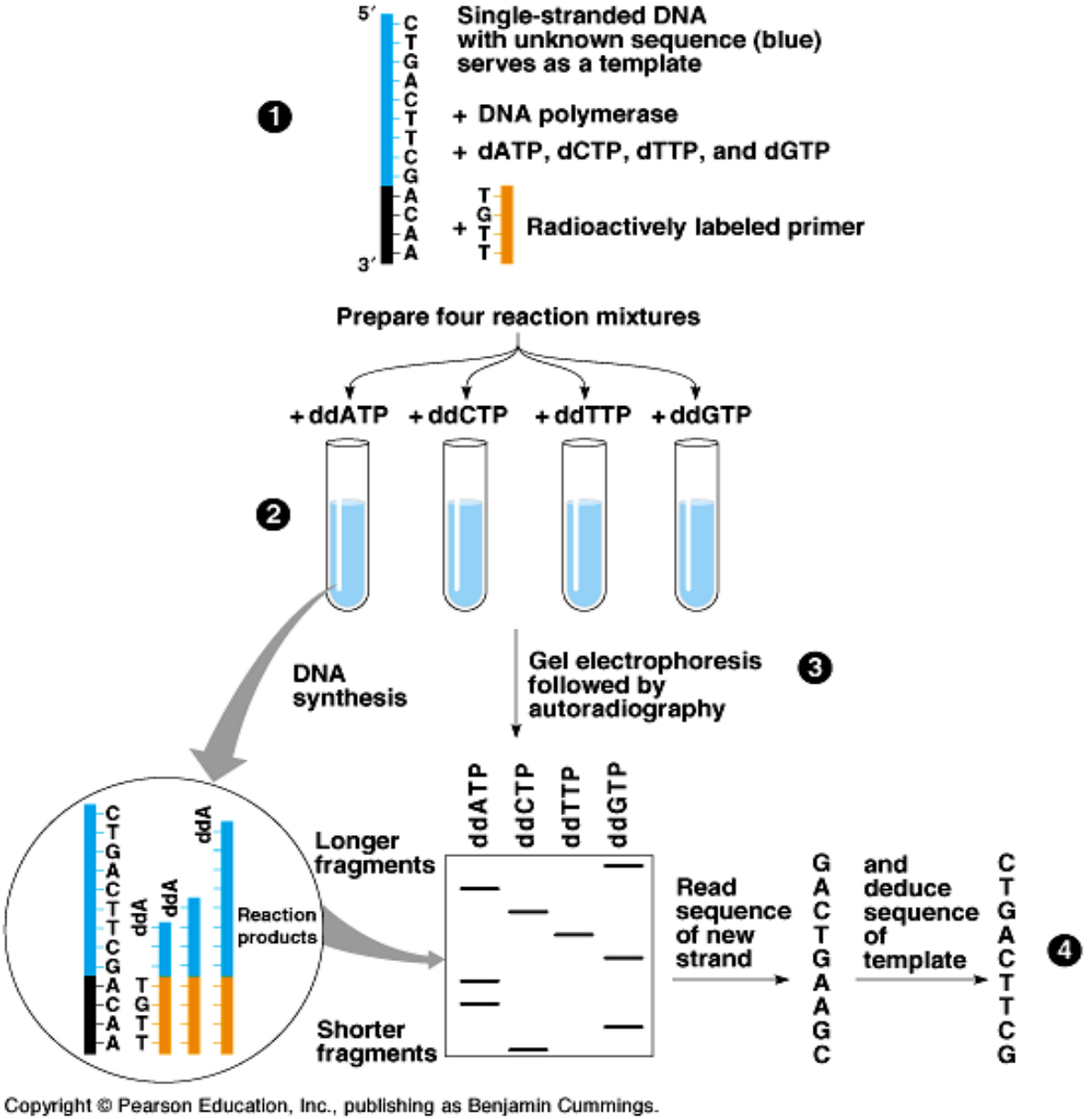
iii. The polymerase enzyme does not discriminate between deoxy-and dideoxynucleotides, but once incoporated a dideoxynucleotides blocks the further elongation because it lacks the 3'-hydroxyl group. Because the normal deoxynucleotides are also present in larger amounts, the strand synthesis does not always terminate close to the primer. The result is a set of new molecules, all of different lengths that is present at the equivalent positon in a template DNA.

iv. To work out the DNA sequence all that we have to do is to identify dideoxynucleotides at the end of each terminated molecule. The mixture is loaded into a well of polyacrylamide slab gel and electrophoresis carried out to separate the molecules according to their lengths.Than, the molecules are run past a fluorescent detector. The detector therefore determines if each molecule ends in an A, C, G or T. The sequence can be printed out by examination by the operator.

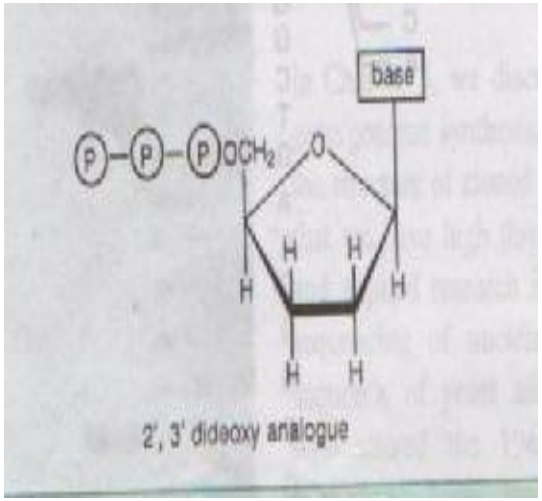


http://en.wikipedia.org/wiki/DNA Sequencing

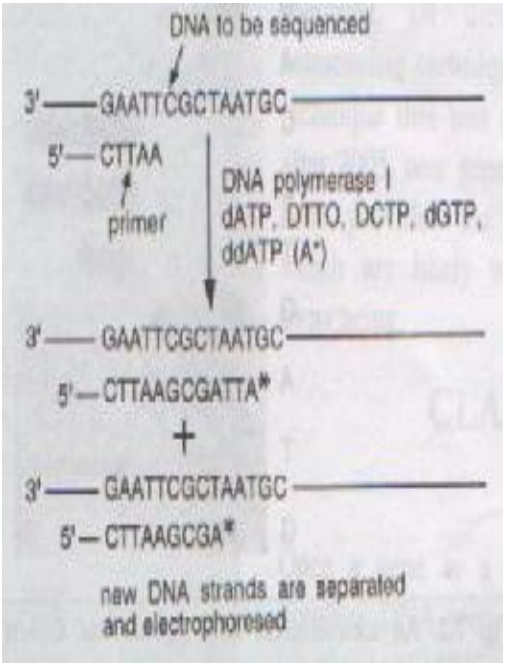
Chain Termination Method



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



A representative dideoxynucleotide used for chain termination in Sanger's method for DNA sequencing.



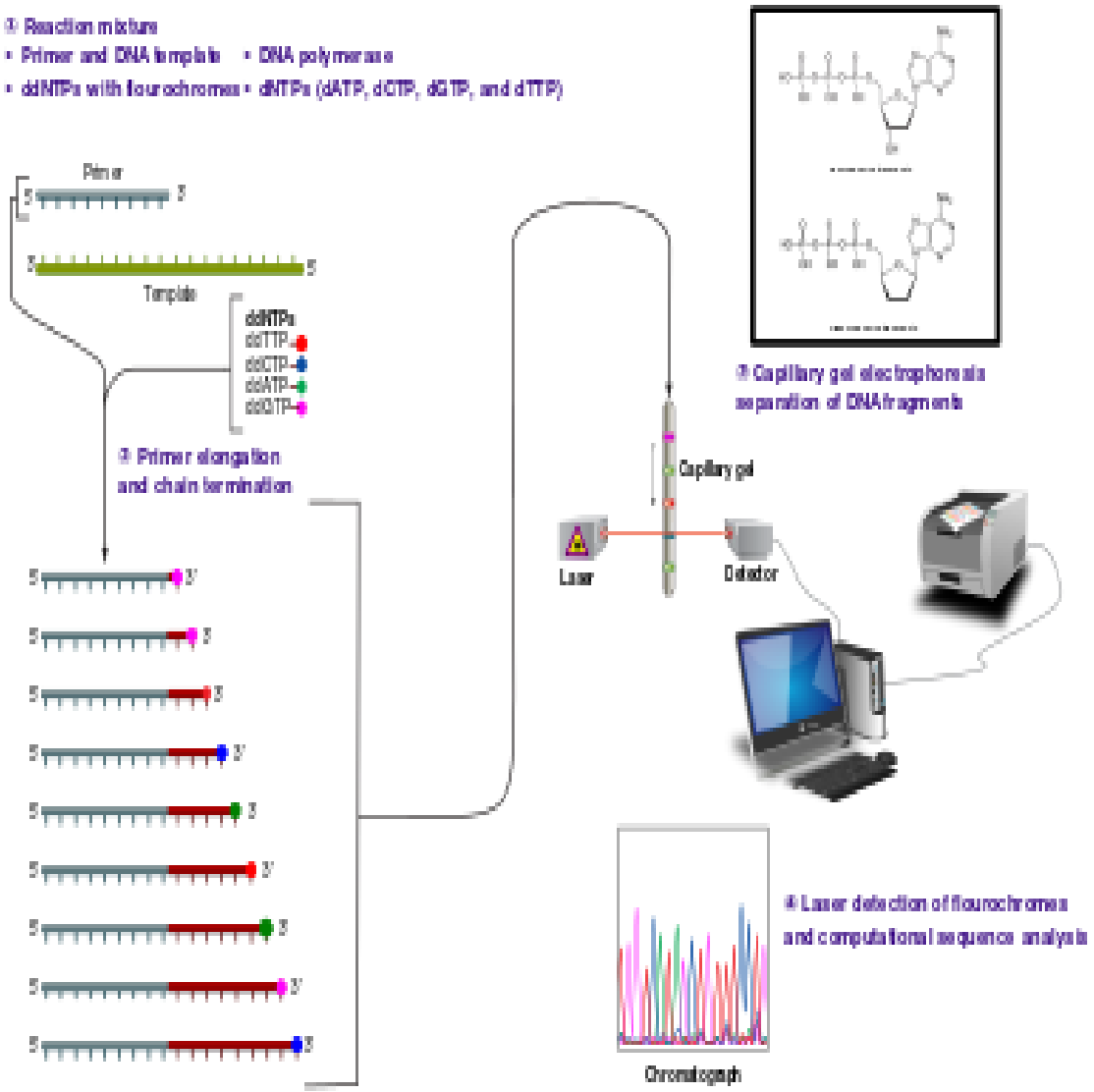
Strategy used in chain termination using dideoxy analogue of a base.

Gupta, P.K.(2010). *"Plant Biotechnology"*, 1st Edition. Rastogi Publications, Meerut.

Automatic DNA Sequencing

- Starting in 1986, several variant of the dideoxy method were developed which allow the production of **automatic sequencers.**
- In it, initially a different fluorescent dye was tagged to the oligonucleotide primer in each of the four reaction tubes.
- The four reaction mixtures were pooled and electrophoresed together in a single polyacrylamide gel tube.A high sensitivity fluorescence detector placed near the bottom of the tube.
- The sequence was determined from the temporal order of peaks corresponding to four different dyes.

Automatic DNA Sequencing



http://en.wikipedia.org/wiki/DNA *Sequencing*

- ***Direct DNA Sequencing using PCR***

  (Also called Ligation Mediated PCR = LMPCR)

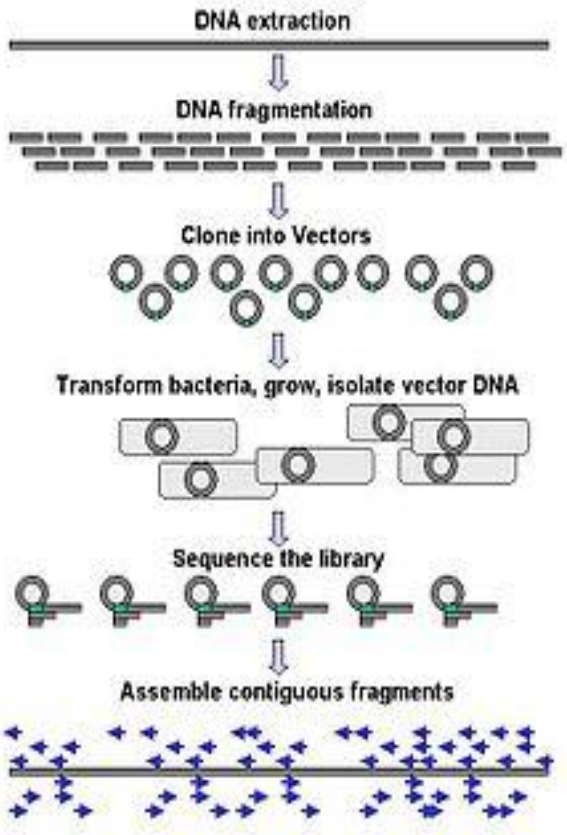
Polymerase Chain Reaction has also been used for sequencing the amplified DNA product. This method of DNA sequencing is faster and more reliable and can utilize either the whole genomic DNA or cloned fragments for sequencing a particular DNA segment. The DNA sequencing using PCR involves two steps:

i. Generation of sequencing templates using PCR and
ii. Sequencing of PCR products with the thermolabile DNA polymerase.

Advanced Methods

Large-scale sequencing often aims at sequencing very long DNA pieces, such as whole chromosomes, although large-scale sequencing can also be used to generate very large numbers of short sequences. For longer targets such as chromosomes, common approaches consist of cutting (with restriction enzymes) or shearing (with mechanical forces) large DNA fragments into shorter DNA fragments.

Gaps in the assembled sequence may be filled by primer walking. The different strategies have different trade offs in speed and accuracy; shotgun methods are often used for sequencing large genomes, but its assembly is complex and difficult, particularly with sequence repeats often causing gaps in genome assembly.



- Shotgun Sequencing
- Bridge PCR

Shotgun Sequencing

- Shotgun sequencing is a sequencing method designed for analysis of DNA sequences longer than 1000 base pairs, upto and including entire chromosomes. This method require the target DNA to be broken into random fragments. After sequencing into individual fragments, the sequences can be reassembled on the basis of their overlapping regions.

Bridge PCR

Another method for in vitro clonal amplification is bridge PCR, in which fragments are amplified upon primers attached to the solid surface and form "DNA colonies" or "DNA clusters" . This method used in the Illumina Genome Analyzer sequencers. Single -molecule method such, as that developed by Stephen Quake's laboratory are an exception: they use bright fluorophores and laser excitation to detect base addition events from individual DNA molecules fixed to a surface, eliminating the need for molecular amplification.

New Generation Sequencing Methods

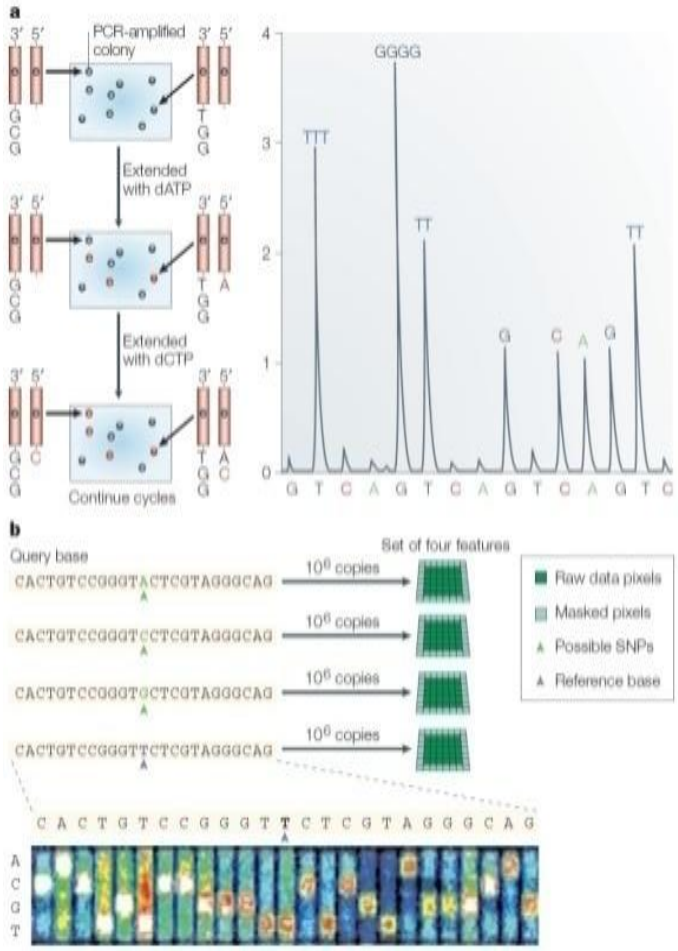
- Sequencing by Hybridization(SBH)
- Sequencing by Synthesis(SBS)
- Sequencing by Ligation(SBL)
- Pyrosequencing
- Nanopore DNA Sequencing
- Ion Torrent semiConductor Sequencing
- Illumina(Solexa) Sequencing
- Transmission Electron Microscopy for DNA Sequencing

Sequencing by Hybridization(SBH)

Sequencing by Hybridization is based on the principle that differential hybridization of oligonucleotide probes, each due to mismatch of a single base, can be used to decode the target DNA sequence. There at least two alternative approaches for SBH:

i. In the first approach, genomic DNA to be sequenced is first immobilized on a membrane and is then serially hybridized with short oligonucleotide probes of known sequences.

ii. In the second approach, genomic DNA to be sequenced is hybridized to microfabricated tilling arrays immobilized oligonucleotide, with ~100,000 copies of each individual of feature.



a) Schematic of cyclic-array sequencing-by-synthesis methods (for example, fluorescent in situ sequencing, pyrosequencing, or single-molecule methods b) Sequencing by hybridization. To resequence a given base, four features are present on the microarray, each identical except for a different nucleotide at the query position (the central base of 25-bp oligonucleotides).

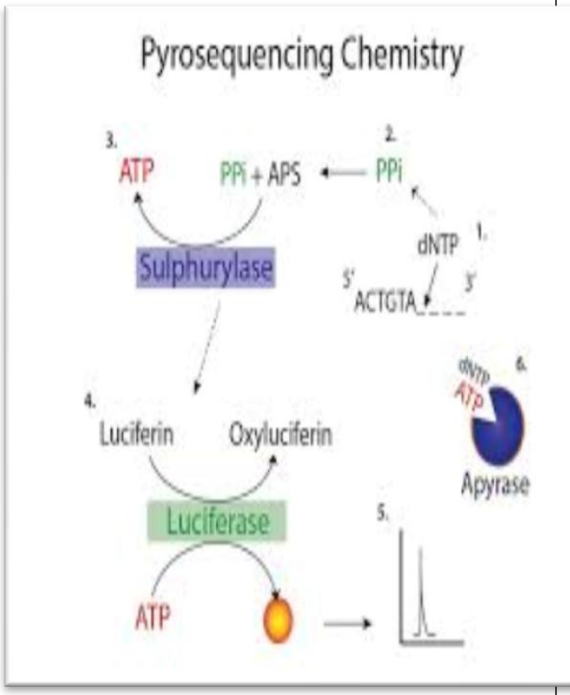
Sequencing by Synthesis(SBS)

The SBS technology encompasses many different DNA polymerase-dependent strategies, so that the term SBS has become increasinly ambiguous; therefore, the following classification of DNA polymerase-dependent strategies may prove useful:

i. Sequencing by Sanger's method
ii. Sequencing by 'single nucleotide addition'(SNA)
iii. Sequencing by 'cyclic reversible termination'
iv. FISSEQ or polony sequencing

Pyrosequencing

- Pyrosequencing is the method of gene sequencing based on the "Sequencing by synthesis" principle. The technique was developed by **Mostafa Ronaghi and Pal Nyren at the Royal Institute of Technology in Stockholm in 1996.** It differs from Sanger sequencing, in that it relies on the detection of pyrophosphate release on nucleotide incorporation rather than the chain termination with dideoxynucleotides.
- Pyrosequencing is the important type of is the method of gene sequencing methodology that is in use today. It does not require electrophoresis or any other fragment separation procedure and so more rapid.
- **Pyrosequencing involves detection of pulses of chemiluminiscence .** Pyrosequencing requires a preparation of identical ss DNA molecules as the starting material. These are obtained by alkali denaturation of PCR products or, more rarely, recombinant plasmid molecules. After attachment of the primer, template is copied by a DNA polymerase in a straight-forward manner without added dideoxynucleotides. As the new strand is being made, the order in which the deoxynucleotides are incorporated is detected , so the sequence can be "read" as the reaction proceeds.
- Pyrosequencing employs a series of four enzymes to accurately detect nucleic acid sequences during the synthesis.
- Pyrosequencing has the potential advantages of accuracy, flexibitity, parallel processing and can be easily automated.



Step 1

A sequencing primer is hybridized to a single stranded, PCR amplified, DNA template, and incubated with the enzymes, DNA polymerase, ATP sulfurylase, luciferase and apyrase, and the substrates, adenosine 5´ phosphosulfate (APS) and luciferin.

Step 2

The first of four deoxyribonucleotide triphosphates (dNTP) is added to the reaction. DNA polymerase catalyzes the incorporation of the deoxyribonucleotide triphosphate into the DNA strand, if it is complementary to the base in the template strand. Each incorporation event is accompanied by release of pyrophosphate (PPi) in a quantity equimolar to the amount of incorporated nucleotide.

Step 3

ATP sulfurylase quantitatively converts PPi to ATP in the presence of adenosine 5´ phosphosulfate (APS). This ATP drives the luciferase mediated conversion of luciferin to oxyluciferin that generates visible light in amounts that are proportional to the amount of ATP. The light produced in the luciferase-catalyzed reaction is detected by a charge coupled device (CCD) camera and seen as a peak in a Pyrogram™. The height of each peak (light signal) is proportional to the number of nucleotides incorporated.

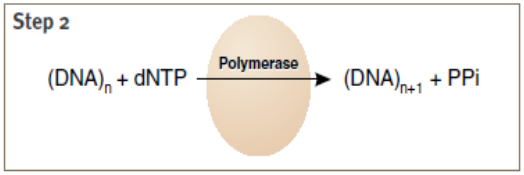
Step 4

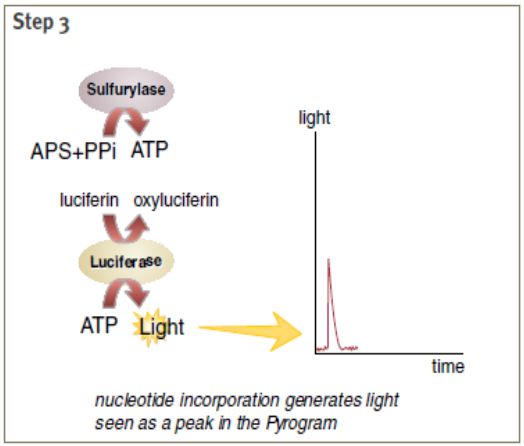
Apyrase, a nucleotide degrading enzyme, continuously degrades ATP and unincorporated dNTPs. This switches off the light and regenerates the reaction solution. The next dNTP is then added.

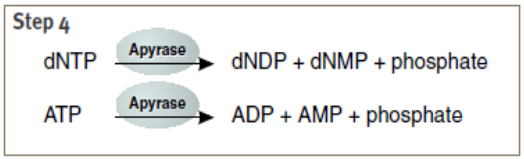
Step 5

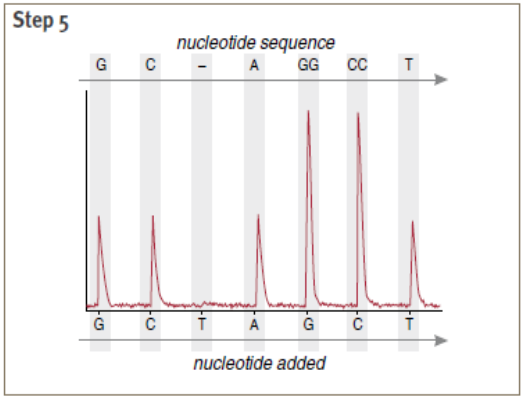
Addition of dNTPs is performed one at a time. It should be noted that deoxyadenosine alfa-thio triphosphate (dATPαS) is used as a substitute for the natural deoxyadenosine triphosphate (dATP) since it is efficiently used by the DNA polymerase, but not recognized by the luciferase.

As the process continues, the complementary DNA strand is built up and the nucleotide sequence is determined from the signal peaks in the Pyrogram.









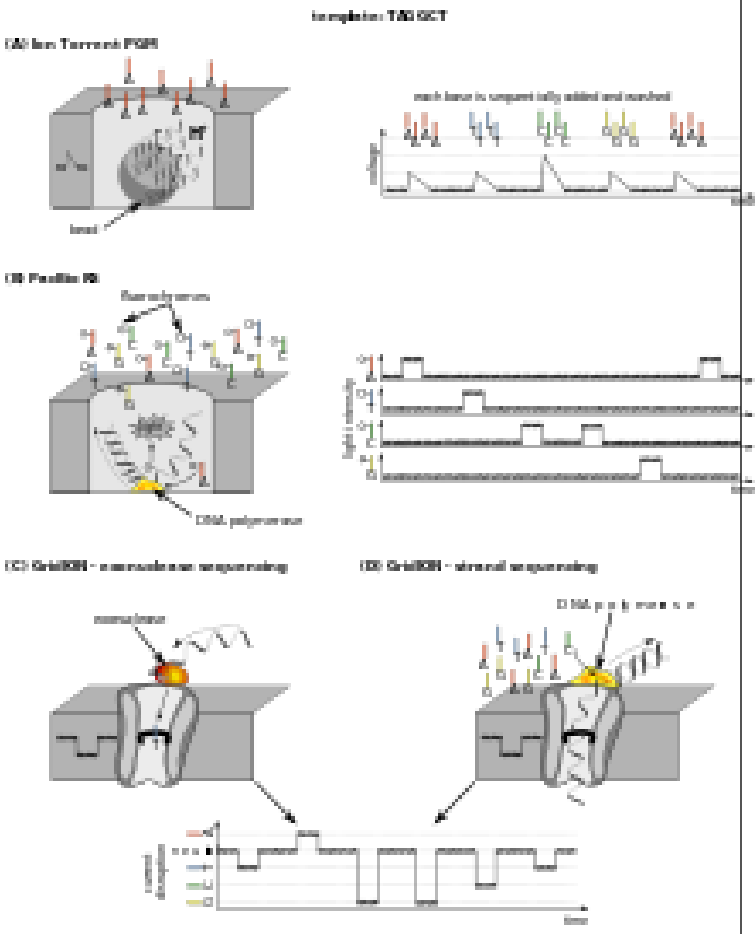
Nanopore Sequencing

- This method is based on the readout of electrical signals occurring at nucleotides passing by alpha-hemolysin pores covalently bound with cyclodextrin. The DNA passes through the nanopore changes its ion current. This change is dependent on the shape, size and length of the DNAs sequence. Each type of the nucleotide blocks the ion flow through the pore for a different period of time,
- Two main areas of nanopore sequencing in development are –

1. Solid state nanopore sequencing
2. Protein based nanopore sequencing

Ion Torrent semiconductor sequencing

- IN 2010, Ion Torrent announced a commercialization of their sequencing technology, which makes the use of simple chemistry with semiconductor technology.
- This made it possible to use for the first time chemistry with information technology for DNA sequencing.
- For this purpose Semiconductor Sequencing Chips are desinged, manufactured and packaged like any other semiconductor chips.
- The machine will measure the change in pH as each nucleotide was incorporated and convert this information change in voltage so that within 4 seconds , 240 data points will be created. For each well, 100-200 bases will be sequenced with in an hour.
- This method of sequencing is based on the detection of hydrogen ions that are released during the polymerisation of DNA.



Sequencing of theTAGGGCT template with Ion Torrent, PacBioRS and GridION

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

Illumina(Solexa) Sequencing

- Solexa, now part of Illumina, developed a sequencing method based on reversible dye-terminators technology, and engineered polymerases, that is developed internally. The terminated chemistry was developed internally at Solexa and the concept of the Solexa system was invented by **Balasubramanian and Klennerman** from Cambridge University's chemistry department.
- In this method, DNA molecules and primers are first attached on a slide and amplified with polymerase so that local clonal DNA colonies, later coined "DNA clusters", are formed. To determine the sequence four types of reversible terminator bases (RT-bases) are added and incorporated nucleotides are washed away.
- Decoupling the enzymatic reaction and the image capture allows for optimal throughput and theoretically unlimited sequencing capacity. With an optimal configuration, the ultimately reachable instrument throughput is thus dictated solely by the analog-to-digital conversion rate of the camera, multiplied by the number of cameras and divided by the number of pixels per DNA colony required for visualizing them optimally.

An Illumina HiSeq 2500 sequencer

An Illumina MiSeq sequencer

http://en.wikipedia.org/wiki/DNA *Sequencing*

Transmission Electron Microscopy for DNA Sequencing

- A novel SMS platform, where DNA sequence is read directly with the help of Transmission Electron Microscope is also being developed by a company named ZS Genetics ('ZSG') based in North Reading, MA(USA).
- The technology involves the linearization of DNA molecule, followed by synthesis of a complimentary strand, where three of the four bases are labelled, the fourth base remaining unlabelled .
- When this complimentary strand is observed under TEM, the bases can be discriminated.
- The company has already released the images of a 23-kilobase piece of DNA, and claim to be able to produce reads of 1.7 billion base pairs.

Uses/Applications of Gene sequencing

- DNA sequencing may be used to determine the sequence of individual genes, larger genetic regions (i.e. clusters of genes or operons), ful chromosomes or entire genomes. Sequencing provides the order of individual nucleotides in DNA or RNA (commonly represented as A, G, T, and U) isolated from cells of animals, plants, bacteria, archaea, or virtually any other source of genetic information.
- Molecular biology - studying the genome itself, how proteins are made, what proteins are made, identifying new genes and associations w diseases and phenotypes, and identifying potential drug targets.
- Evolutionary biology - studying how different organisms are related and how they evolved.
- Metagenomics - Identifying species present in a body of water, sewage, dirt, debris filtred from the air, or swab samples of organisms. Hel in ecology, epidemiology, microbiome research, and other fields.
- Less-precise information is produced by non-sequencing techniques like DNA fingerprinting. This information may be easier to obtain and
- Detect the presence of known genes for medical purposes.
- Forensic identification.
- Parental testing.

Computational Challenges

The sequencing technologies described here produce the raw data that need to be assembled into longer sequences such as complete genomes(sequence assembly). There are many computational challenges to achieve this, such as the evaluation of raw sequences data which is done by program and algorithms as the Phred and Phrap. Other challenges have to deal with repetitive sequences that often prevent complete genome assemblies because they occur in many places of the genomes.

As a consequences, many sequences may not be assinged to a particular chromosomes. The production of raw sequence data in the only beginning of its detailed bioinformatical analysis.

Conclusion

- As the previous sections illustrate, sequencing technologies both old and new have brought us information about many genomes. In the Beginning of 1970s, the Sanger process made it possible for researchers to sequence stretches of DNA at speeds never before possible. Further refinement and automation of this process continued to increase sequencing rates, thereby allowing researchers to reach major milestones in the Human Genome Project. Today, newer pyrosequencing methods have drastically cut the cost of sequencing and may eventually allow every person the possibility of personalized genome information. Being able to read how our genes are expressed offers the promise of advanced medical treatments, but it will certainly require considerable work to generate, understand, organize, and apply this massive amount of data to human disease.

References

- Brown, T.A. (2010). *"Gene Cloning and DNA Analysis", 6th* Edition. Wiley-Blackwell, John Wiley & Sons, Ltd., Publication, UK. pp. 165-173.
- Glick, B.R. and J.J. Paternack (2001). *"Molecular Biotechnology"*, 2nd Edition. American Society for Microbiology, Washington, U.S.A. pp. 89-91.
- Gupta, P.K. (2010). *"Plant Biotechnology"*, 1st Edition. Rastogi Publication, Merrut. pp. 96-112.
- Singh, B.D. (2010). *"Biotechnology: Expanding Horizons", 3rd* Edition. Kalyani Publications, New Delhi.
- http://en.wikipedia.org/wiki/DNA Sequencing

THANK YOU