

# DNA Representation

Bharti Rajendra Kumar

*B.T. Kumaon Institute of Technology, Dwarahat, Almora, Uttarakhand,  
India*

## 1. Introduction

The term DNA sequencing refers to methods for determining the order of the nucleotides bases adenine, guanine, cytosine and thymine in a molecule of DNA. The first DNA sequence were obtained by academic researchers, using laboratories methods based on 2- dimensional chromatography in the early 1970s. By the development of dye based sequencing method with automated analysis, DNA sequencing has become easier and faster. The knowledge of DNA sequences of genes and other parts of the genome of organisms has become indispensable for basic research studying biological processes, as well as in applied fields such as diagnostic or forensic research.

DNA is the information store that ultimately dictates the structure of every gene product, delineates every part of the organisms. The order of the bases along DNA contains the complete set of instructions that make up the genetic inheritance.

The rapid speed of sequencing attained with modern DNA sequencing technology has been instrumental in the sequencing of the human genome, in the human genome project.

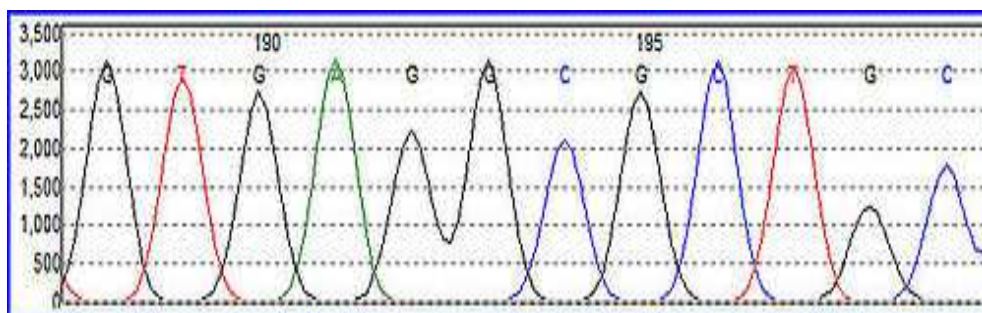


Fig. 1. DNA Sequence Trace

DNA can be sequenced by a chemical procedure that breaks a terminally labelled DNA molecule partially at each repetition of a base. The length of the labelled fragments then identify the position of that base. We describe reactions that cleave DNA preferentially at guanines, at adenines, at cytosine and thymines equally, and at cytosine alone. When the product of these four reactions are resolved by size, by electrophoresis on a polyacrylamide gel, the DNA sequences can be read from the pattern of radioactive bands. The technique

will permit sequencing of at least 100 bases from the point of labelling. The purine specific reagent is dimethyl sulphate; and the pyrimidine specific reagent is hydrazine.

In 1973, Gilbert and Maxam reported the sequence of 24 base pairs using a method known as wandering-spot analysis.

The chain termination method developed by Sanger and coworkers in 1975 owing to its relative ease and reliability.

In 1975 the first complete DNA genome to be sequenced is that of bacteriophage  $\phi$ X174.

By knowing the DNA sequence, the cause of the various diseases can be known. We can determine the sequence responsible for various diseases and can be treated with the help of Gene therapy.

DNA sequencing is very significant in research and forensic science. The main objective of DNA sequence generation method is to evaluate the sequencing with very high accuracy and reliability.

There are some common automated DNA sequencing problems :-

1. Failure of the DNA sequence reaction.
2. Mixed signal in the trace (multiple peaks).
3. Short read lengths and poor quality data.
4. Excessive free dye peaks "dye blobs" in the trace.
5. Primer dimer formation in sequence reaction
6. DNA polymerase slippage on the template mononucleotide regions.

So, we should have to do the sequencing in such a manner to avoid or minimize these problems.

DNA sequencing can solve a lot of problems and perform a lot of work for human welfare

A sequencing can be done by different methods :

1. Maxam – Gilbert sequencing
2. Chain-termination methods
3. Dye-terminator sequencing
4. Automation and sample preparation
5. Large scale sequencing strategies
6. New sequencing methods.

## 2. Maxam-Gilbert sequencing

In 1976-1977, Allan Maxam and Walter Gilbert developed a DNA sequencing method based on chemical modification of DNA and subsequent cleavage at specific bases.

The method requires radioactive labelling at one end and purification of the DNA fragment to be sequenced. Chemical treatment generates breaks at small proportions of one or two of the four nucleotides in each of four reactions (G, A+G, C, C+T). Thus a series of labelled fragments is generated, from the radiolabelled end to the first 'cut' site in each molecule. The fragments in the four reactions are arranged side by side in gel

electrophoresis for size separation. To visualize the fragments, the gel is exposed to X-ray film for autoradiography, yielding a series of dark bands each corresponding to a radiolabelled DNA fragment, from which the sequence may be inferred.

### 3. Chain-termination method

The chain terminator method is more efficient and uses fewer toxic chemicals and lower amount of radioactivity than the method of Maxam and Gilbert.

The key principle of the Sanger method was the use of dideoxynucleotide triphosphates (ddNTPs) as DNA chain terminators.

The chain termination method requires a single-stranded DNA template, a DNA primer, a DNA polymerase, radioactively or fluorescently labelled nucleotides, and modified nucleotides that terminate DNA strand elongation. The DNA sample is divided into four separate sequencing reactions, containing all four of the standard deoxynucleotides (dATP, dGTP, dCTP, dTTP) and the DNA polymerase. To each reaction is added only one of the four dideoxynucleotide (ddATP, ddGTP, ddCTP, ddTTP) which are the chain terminating nucleotides, lacking a 3'-OH group required for the formation of a phosphodiester bond between two nucleotides, thus terminating DNA strand extension and resulting in DNA fragments of varying length.

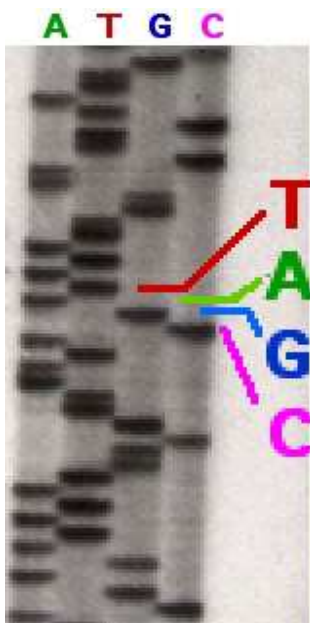


Fig. 2. Part of a radioactively labelled sequencing gel

The newly synthesized and labelled DNA fragments are heat denatured, and separated by size by gel electrophoresis on a denaturing polyacrylamide-urea gel with each of the four reactions run in one of the four individual lanes (lanes A, T, G, C), the DNA bands are then visualized by autoradiography or UV light, and the DNA sequence can be directly read off

the X-ray film or gel image. A dark band in a lane indicates a DNA fragment that is result of chain termination after incorporation of a dideoxynucleotide (ddATP, ddGTP, ddCTP, or ddTTP). The relative position of the different bands among the four lanes are then used to read (from bottom to top) the DNA sequence.

The technical variations of chain termination sequencing include tagging with nucleotides containing radioactive phosphorus for labelling, or using a primer labelled at the 5' end with a fluorescent dye. Dye- primer sequencing facilitates reading in an optical system for faster and more economical analysis and automation.

Chain termination methods have greatly simplified DNA sequencing. Limitations include non-specific binding of the primer to the DNA, affecting accurate read-out of the DNA sequence, and DNA secondary structures affecting the fidelity of the sequence.

### 3.1 Dye-terminator sequencing

Dye-terminator sequencing utilizes labelling of the chain terminator ddNTPs, which permits sequencing in a single reaction, rather than four reactions as in the labelled- primer method. In dye- terminator sequencing, each of the four dideoxynucleotide chain terminators is labelled with fluorescent dyes, each of which with different wavelengths of fluorescence and emission. Owing to its greater expediency and speed, dye terminator sequencing is now the mainstay in automated sequencing. Its limitation include dye effects due to differences in the incorporation of the dye-labelled chain terminators into the DNA fragment, resulting in unequal peak heights and shapes in the electronic DNA sequence trace chromatogram.

The common challenges of DNA sequencing include poor quality in the first 15-40 bases of the sequence and deteriorating quality of sequencing traces after 700-900 bases.

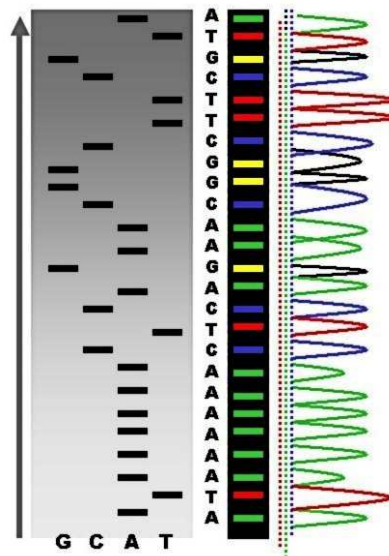


Fig. 3. Sequence ladder by radioactive sequencing compared to fluorescent peaks

### 3.2 Automation and sample preparation

Automated DNA sequencing instruments (DNA sequencers) can sequence up to 384 DNA samples in a single batch (run) in up to 24 runs a day. DNA sequencers carry out capillary electrophoresis for size separation, detection and recording of dye fluorescence, and data output as fluorescent peak trace chromatograms.

A number of commercial and non-commercial software packages can trim low-quality DNA traces automatically. These programmes score the quality of each peak and remove low-quality base peaks (generally located at the ends of the sequence).

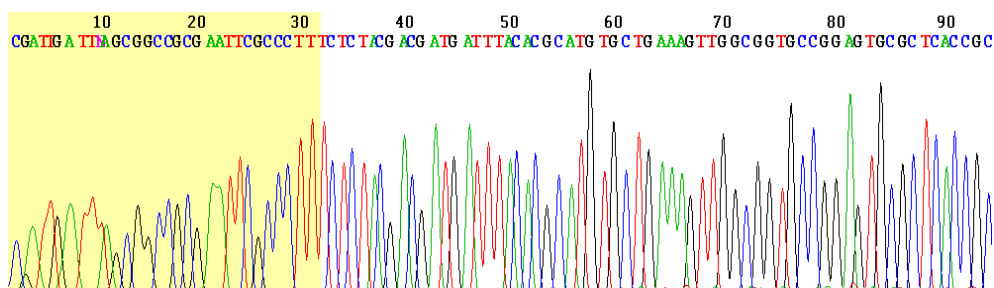


Fig. 4. View of the start of an example dye-terminator read

### 4. Large-scale sequencing strategies

Current methods can directly sequence only relative short (300-1000 nucleotides long) DNA fragments in a single reaction. The main obstacle to sequencing DNA fragments above this size limit is insufficient power of separation for resolving large DNA fragments that differ in length by only one nucleotide.

Large scale sequencing aims at sequencing very long DNA pieces, such as whole chromosomes. It consists of cutting (with restriction enzymes) or shearing (with mechanical forces) large DNA fragments into shorter DNA fragments. The fragmented DNA is cloned into a DNA vector, and amplified in *E. coli*. Short DNA fragments purified from individual bacterial colonies are individually sequenced and assembled electronically into one long, contiguous sequence. This method does not require any pre-existing information about the sequence of the DNA and is referred to as *de novo* sequencing. Gaps in the assembled sequence may be filled by primer walking. The different strategies have different tradeoffs in speed and accuracy.

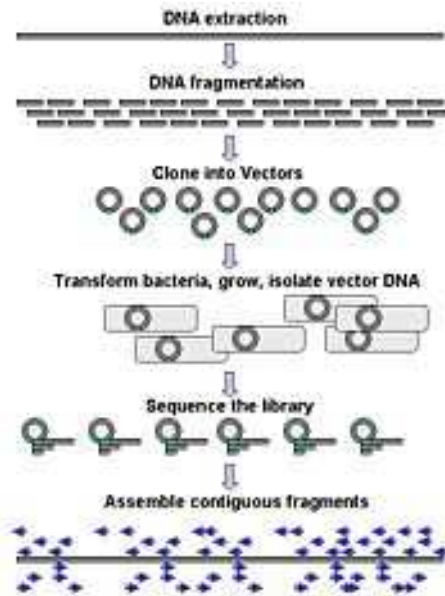


Fig. 5. Genomic DNA is fragmented into random pieces and cloned as a bacterial library. DNA from individual bacterial clones is sequenced and the sequence is assembled by using overlapping regions.

## 5. New sequencing methods

The high demand for low-cost sequencing has driven the development of high-throughput sequencing technologies that parallelize the sequencing process, producing thousands or millions of sequences at once. High-throughput sequencing technologies are intended to lower the cost of DNA sequencing.

Molecular detection methods are not sensitive enough for single molecule sequencing, so most approaches use an in vitro cloning step to amplify individual DNA molecules.

In microfluidic Sanger sequencing the entire thermocycling amplification of DNA fragments as well as their separation by electrophoresis is done on a single chip (approximately 100  $\mu\text{m}$  in diameter) thus reducing the reagent usage as well as cost. In some instances researchers have shown that they can increase the throughput of conventional sequencing through the use of microchips.

## 6. High throughput sequencing

The high demand for low-cost sequencing has driven the development of high-throughput sequencing technologies that parallelize the sequencing process, producing thousands or millions of sequences at once. High-throughput sequencing technologies are intended to lower the cost of DNA sequencing beyond what is possible with standard dye-terminator methods.

### **6.1 Lynx therapeutics' massively parallel signature sequencing (MPSS)**

The first of the "next-generation" sequencing technologies, MPSS was developed in the 1990s at Lynx Therapeutics, a company founded in 1992 by Sydney Brenner and Sam Eletr. MPSS is an ultra high throughput sequencing technology. When applied to expression profile, it reveal almost every transcript in the sample and provide its accurate expression level. MPSS was a bead-based method that used a complex approach of adapter ligation followed by adapter decoding, reading the sequence in increments of four nucleotides; this method made it susceptible to sequence-specific bias or loss of specific sequences. However, the essential properties of the MPSS output were typical of later "next-gen" data types, including hundreds of thousands of short DNA sequences. In the case of MPSS, these were typically used for sequencing cDNA for measurements of gene expression levels. Lynx Therapeutics merged with Solexa in 2004, and this company was later purchased by Illumina.

### **6.2 Polony sequencing**

It is an inexpensive but highly accurate multiplex sequencing technique that can be used to read millions of immobilized DNA sequences in parallel. This techniques was first developed by Dr. George Church in Harvard Medical college. It combined an in vitro paired-tag library with emulsion PCR, an automated microscope, and ligation-based sequencing chemistry to sequence an E. coli genome at an accuracy of > 99.9999% and a cost approximately 1/10 that of Sanger sequencing.

### **6.3 Pyrosequencing**

A parallelized version of pyrosequencing was developed by 454 Life Sciences, which has since been acquired by Roche Diagnostics. The method amplifies DNA inside water droplets in an oil solution (emulsion PCR), with each droplet containing a single DNA template attached to a single primer-coated bead that then forms a clonal colony. The sequencing machine contains many picolitre-volume wells each containing a single bead and sequencing enzymes. Pyrosequencing uses luciferase to generate light for detection of the individual nucleotides added to the nascent DNA, and the combined data are used to generate sequence read-outs. This technology provides intermediate read length and price per base compared to Sanger sequencing on one end and Solexa and SOLiD on the other.

### **6.4 Illumina (Solexa) sequencing**

Solexa developed a sequencing technology based on dye terminators. In this, DNA molecule are first attached to primers on a slide and amplified, this is known as bridge amplification. Unlike pyrosequencing, the DNA can only be extended one neucleotide at a time. A camera takes images of the fluorescently labeled nucleotides, then the dye along with the terminal 3' blocker is chemically removed from the DNA, allowing the next cycle.

### **6.5 SOLiD sequencing**

The technology for sequencing used in ABISolid sequencing is oligonucleotide ligation and detection. In this, a pool of all possible oligonucleotides of fixed length are labelled according to the sequenced position. This sequencing results to the sequences of quantities and lengths comparable to illumine sequencing.

## 6.6 DNA nanoball sequencing

It is high throughput sequencing technology that is used to determine the entire genomic sequence of an organisms. The method uses rolling circle replication to amplify fragments of genomic DNA molecules. This DNA sequencing allows large number of DNA nanoballs to be sequenced per run and at low reagent cost compared to other next generation sequencing platforms. However, only short sequences of DNA are determined from each DNA nanoball which makes mapping the short reads to a reference genome difficult. This technology has been used for multiple genome sequencing projects and is scheduled to be used for more.

## 6.7 Helioscope(TM) single molecule sequencing

Helioscope sequencing uses DNA fragments with added polyA tail adapters, which are attached to the flow cell surface. The next steps involve extension-based sequencing with cyclic washes of the flow cell with fluorescently labeled nucleotides. The reads are performed by the Helioscope sequencer. The reads are short, up to 55 bases per run, but recent improvement of the methodology allows more accurate reads of homopolymers and RNA sequencing.

## 6.8 Single molecule SMRT(TM) sequencing

SMRT sequencing is based on the sequencing by synthesis approach. The DNA is synthesised in so called zero-mode wave-guides (ZMWs) - small well-like containers with the capturing tools located at the bottom of the well. The sequencing is performed with use of unmodified polymerase and fluorescently labelled nucleotides flowing freely in the solution. The wells are constructed in a way that only the fluorescence occurring by the bottom of the well is detected. The fluorescent label is detached from the nucleotide at its incorporation into the DNA strand, leaving an unmodified DNA strand. The SMRT technology allows detection of nucleotide modifications. This happens through the observation of polymerase kinetics. This approach allows reads of 1000 nucleotides.

## 6.9 Single molecule real time (RNAP) sequencing

This method is based on RNA polymerase (RNAP), which is attached to a polystyrene bead, with distal end of sequenced DNA is attached to another bead, with both beads being placed in optical traps. RNAP motion during transcription brings the beads in closer and their relative distance changes, which can then be recorded at a single nucleotide resolution. The sequence is deduced based on the four readouts with lowered concentrations of each of the four nucleotide types.

## 7. Other sequencing technologies

Sequencing by hybridization is a non-enzymatic method that uses a DNA microarray. A single pool of DNA whose sequence is to be determined is fluorescently labelled and hybridized to an array containing known sequences. Strong hybridization signals from a given spot on the array identifies its sequence in the DNA being sequenced. Mass spectrometry may be used to determine mass differences between DNA fragments produced in chain-termination reactions.



Some important applications of DNA sequencing are :

1. To analyse any protein structure and function we must have the knowledge of its primary structure i.e its DNA sequence.
2. With its study we can understand the function of a specific sequence and the sequence responsible for any disease.
3. With the help of comparative DNA sequence study we can detect any mutation.
4. Kinship study.
5. DNA fingerprinting.
6. By knowing the whole genome sequence, Human genome project get completed.

The main problem with sequencing is its intactness. If we perform the sequencing of same sample with different methods the result may be different so we should have to do it in such a manner that atleast 40-50% sequence must be same of similar sample.

## 8. Benchmarks in DNA sequencing

- 1953 Discovery of the structure of the DNA double helix.
- 1972 Development of recombinant DNA technology, which permits isolation of defined fragments of DNA; prior to this, the only accessible samples for sequencing were from bacteriophage or virus DNA.
- 1975 The first complete DNA genome to be sequenced is that of bacteriophage  $\phi$ X174
- 1977 Allan Maxam and Walter Gilbert publish "DNA sequencing by chemical degradation". Fred Sanger, independently, publishes "DNA sequencing by enzymatic synthesis".
- 1980 Fred Sanger and Wally Gilbert receive the Nobel Prize in Chemistry
- EMBL-bank, the first nucleotide sequence repository, is started at the European Molecular Biology Laboratory
- 1982 Genbank starts as a public repository of DNA sequences.
- Andre Marion and Sam Eletr from Hewlett Packard start Applied Biosystems in May, which comes to dominate automated sequencing.
- Akiyoshi Wada proposes automated sequencing and gets support to build robots with help from Hitachi.
- 1984 Medical Research Council scientists decipher the complete DNA sequence of the Epstein-Barr virus, 170 kb.
- 1985 Kary Mullis and colleagues develop the polymerase chain reaction, a technique to replicate small fragments of DNA
- 1986 Leroy E. Hood's laboratory at the California Institute of Technology and Smith announce the first semi-automated DNA sequencing machine.
- 1987 Applied Biosystems markets first automated sequencing machine, the model ABI 370.
- Walter Gilbert leaves the U.S. National Research Council genome panel to start Genome Corp., with the goal of sequencing and commercializing the data.
- 1990 The U.S. National Institutes of Health (NIH) begins large-scale sequencing trials on *Mycoplasma capricolum*, *Escherichia coli*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae* (at 75 cents (US)/base).

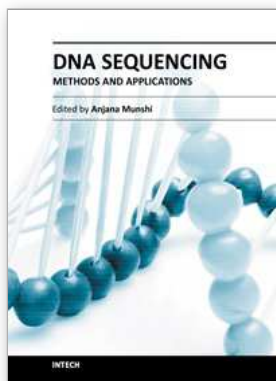
- Barry Karger (January), Lloyd Smith (August), and Norman Dovichi (September) publish on capillary electrophoresis.
- 1991 Craig Venter develops strategy to find expressed genes with ESTs (Expressed Sequence Tags).
- Uberbacher develops GRAIL, a gene-prediction program.
- 1992 Craig Venter leaves NIH to set up The Institute for Genomic Research (TIGR).
- William Haseltine heads Human Genome Sciences, to commercialize TIGR products.
- Wellcome Trust begins participation in the Human Genome Project.
- Simon et al. develop BACs (Bacterial Artificial Chromosomes) for cloning.
- First chromosome physical maps published:
  - Page et al. - Y chromosome;
  - Cohen et al. chromosome 21.
  - Lander - complete mouse genetic map;
  - Weissenbach - complete human genetic map.
- 1993 Wellcome Trust and MRC open Sanger Centre, near Cambridge, UK.
  - The GenBank database migrates from Los Alamos (DOE) to NCBI (NIH).
- 1995 Venter, Fraser and Smith publish first sequence of free-living organism, *Haemophilus influenzae* (genome size of 1.8 Mb).
  - Richard Mathies et al. publish on sequencing dyes (PNAS, May).
  - Michael Reeve and Carl Fuller, thermostable polymerase for sequencing.
- 1996 International HGP partners agree to release sequence data into public databases within 24 hours.
  - International consortium releases genome sequence of yeast *S. cerevisiae* (genome size of 12.1 Mb).
  - Yoshihide Hayashizaki's at RIKEN completes the first set of full-length mouse cDNAs.
  - ABI introduces a capillary electrophoresis system, the ABI310 sequence analyzer.
- 1997 Blattner, Plunkett et al. publish the sequence of *E. coli* (genome size of 5 Mb)
- 1998 Phil Green and Brent Ewing of Washington University publish "phred" for interpreting sequencer data (in use since '95).
  - Venter starts new company "Celera"; "will sequence HG in 3 yrs for \$300m."
  - Applied Biosystems introduces the 3700 capillary sequencing machine.
  - Wellcome Trust doubles support for the HGP to \$330 million for 1/3 of the sequencing.
  - NIH & DOE goal: "working draft" of the human genome by 2001.
  - Sulston, Waterston et al finish sequence of *C. elegans* (genome size of 97Mb).
- 1999 NIH moves up completion date for rough draft, to spring 2000.
  - NIH launches the mouse genome sequencing project.
  - First sequence of human chromosome 22 published.
- 2000 Celera and collaborators sequence fruit fly *Drosophila melanogaster* (genome size of 180Mb) - validation of Venter's shotgun method. HGP and Celera debate issues related to data release.
  - HGP consortium publishes sequence of chromosome 21.
  - HGP & Celera jointly announce working drafts of HG sequence, promise joint publication.

- Estimates for the number of genes in the human genome range from 35,000 to 120,000. International consortium completes first plant sequence, *Arabidopsis thaliana* (genome size of 125 Mb).
- 2001 HGP consortium publishes Human Genome Sequence draft in *Nature* (15 Feb).
  - Celera publishes the Human Genome sequence.
- 2005 420,000 VariantSeq human resequencing primer sequences published on new NCBI Probe database.
- 2007 For the first time, a set of closely related species (12 *Drosophilidae*) are sequenced, launching the era of phylogenomics.
  - Craig Venter publishes his full diploid genome: the first human genome to be sequenced completely.
- 2008 An international consortium launches The 1000 Genomes Project, aimed to study human genetic variability.
- 2008 Leiden University Medical Center scientists decipher the first complete DNA sequence of a woman.

## 9. References

- [1] Olsvik O, Wahlberg J, Petterson B, et al. (January 1993). "Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit B in *Vibrio cholerae* O1 strains". *J. Clin. Microbiol.* 31 (1): 22-5. PMC 262614. PMID 7678018.
- [2] Pettersson E, Lundeberg J, Ahmadian A (February 2009). "Generations of sequencing technologies". *Genomics* 93 (2): 105-11. doi:10.1016/j.ygeno.2008.10.003. PMID 18992322.
- [3] Min Jou W, Haegeman G, Ysebaert M, Fiers W (May 1972). "Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein". *Nature* 237 (5350): 82-8. Bibcode:1972Natur.237...82J. doi:10.1038/237082a0. PMID 4555447.
- [4] Fiers W, Contreras R, Duerinck F, et al (April 1976). "Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene". *Nature* 260 (5551): 500-7. Bibcode 1976 Natur.260..500F. doi:10.1038/260500a0. PMID 1264203.
- [5] Maxam AM, Gilbert W (February 1977). "A new method for sequencing DNA". *Proc. Natl. Acad. Sci. U.S.A.* 74 (2): 560-4. Bibcode 1977PNAS...74..560M. doi:10.1073/pnas.74.2.560. PMC 392330. PMID 265521.
- [6] Gilbert, W. DNA sequencing and gene structure. Nobel lecture, 8 December 1980.
- [7] Gilbert W, Maxam A (December 1973). "The Nucleotide Sequence of the lac Operator". *Proc. Natl. Acad. Sci. U.S.A.* 70 (12): 3581-4. Bibcode 1973PNAS...70.3581G. doi:10.1073/pnas.70.12.3581. PMC 427284. PMID 4587255.
- [8] Sanger F, Coulson AR (May 1975). "A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase". *J. Mol. Biol.* 94 (3): 441-8. doi:10.1016/0022-2836(75)90213-2. PMID 1100841.
- [9] Sanger F, Nicklen S, Coulson AR (December 1977). "DNA sequencing with chain-terminating inhibitors". *Proc. Natl. Acad. Sci. U.S.A.* 74 (12): 5463-7. Bibcode:1977PNAS...74.5463S. doi:10.1073/pnas.74.12.5463. PMC 431765. PMID 271968.
- [10] Sanger F. Determination of nucleotide sequences in DNA. Nobel lecture, 8 December 1980.

- [11] 10.Graziano Pesole; Cecilia Saccone (2003). Handbook of comparative genomics: principles and methodology. New York: Wiley-Liss. pp. 133. ISBN 0-471-39128-X.
- [12] Smith LM, Fung S, Hunkapiller MW, Hunkapiller TJ, Hood LE (April 1985). "The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis". *Nucleic Acids Res.* 13 (7): 2399–412. doi:10.1093/nar/13.7.2399. PMC 341163. PMID 4000959.
- [13] Base-calling for next-generation sequencing platforms – Brief Bioinform". Retrieved 2011-02-24.
- [14] Murphy, K.; Berg, K.; Eshleman, J. (2005). "Sequencing of genomic DNA by combined amplification and cycle sequencing reaction". *Clinical chemistry* 51 (1): 35–39. doi:10.1373/clinchem.2004.039164. PMID 15514094. edit
- [15] Sengupta, D. .; Cookson, B. . (2010). "SeqSharp: A general approach for improving cycle-sequencing that facilitates a robust one-step combined amplification and sequencing method". *The Journal of molecular diagnostics : JMD* 12 (3): 272–277. doi:10.2353/jmoldx.2010.090134. PMC 2860461. PMID 20203000. edit
- [16] Richard Williams, Sergio G Peisajovich, Oliver J Miller, Shlomo Magdassi, Dan S Tawfik, Andrew D Griffiths (2006). "Amplification of complex gene libraries by emulsion PCR". *Nature methods* 3 (7): 545–550. doi:10.1038/nmeth896. PMID 16791213.
- [17] Hall N (May 2007). "Advanced sequencing technologies and their wider impact in microbiology". *J. Exp. Biol.* 210 (Pt 9): 1518–25. doi:10.1242/jeb.001370. PMID 17449817.
- [18] Church GM (January 2006). "Genomes for all". *Sci. Am.* 294 (1): 46–54. doi:10.1038/scientificamerican0106-46. PMID 16468433.
- [19] Schuster, Stephan C. (2008). "Next-generation sequencing transforms today's biology". *Nature methods* (Nature Methods) 5 (1): 16–18. doi:10.1038/nmeth1156. PMID 18165802.
- [20] Brenner, Sidney; Johnson, M; Bridgham, J; Golda, G; Lloyd, DH; Johnson, D; Luo, S; McCurdy, S et al. (2000). "Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays". *Nature Biotechnology* (Nature Biotechnology) 18 (6): 630–634. doi:10.1038/76469. PMID 10835600.
- [21] Schuster SC (January 2008). "Next-generation sequencing transforms today's biology". *Nat. Methods* 5 (1): 16–8. doi:10.1038/nmeth1156. PMID 18165802.
- [22] Mardis ER (2008). "Next-generation DNA sequencing methods". *Annu Rev Genomics Hum Genet* 9: 387–402. doi:10.1146/annurev.genom.9.081307.164359. PMID 18576944.
- [23] Valouev A, Ichikawa J, Tonthat T, et al. (July 2008). "A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning". *Genome Res.* 18 (7): 1051–63. doi:10.1101/gr.076463.108. PMC 2493394. PMID 18477713.
- [24] Human Genome Sequencing Using Unchained Base Reads in Self-Assembling DNA Nanoarrays. Drmanac, R. et. al. *Science*, 2010, 327 (5961): 78-81,
- [25] Genome Sequencing on Nanoballs Porreca, JG. *Nature Biotechnology*, 2010, 28:(43-44)
- [26] Human Genome Sequencing Using Unchained Base Reads in Self-Assembling DNA Nanoarrays, Supplementary Material. Drmanac, R. et. al. *Science*, 2010, 327 (5961):78-81, Complete Genomics Press release, 2010
- [27] Hanna GJ, Johnson VA, Kuritzkes DR, et al (1 July 2000). "Comparison of Sequencing by Hybridization and Cycle Sequencing for Genotyping of Human Immunodeficiency Virus Type 1 Reverse Transcriptase". *J. Clin. Microbiol.* 38 (7): 2715–21. PMC 87006. PMID 10878069.



## **DNA Sequencing - Methods and Applications**

Edited by Dr. Anjana Munshi

ISBN 978-953-51-0564-0

Hard cover, 174 pages

**Publisher** InTech

**Published online** 20, April, 2012

**Published in print edition** April, 2012

This book illustrates methods of DNA sequencing and its application in plant, animal and medical sciences. It has two distinct sections. The one includes 2 chapters devoted to the DNA sequencing methods and the second includes 6 chapters focusing on various applications of this technology. The content of the articles presented in the book is guided by the knowledge and experience of the contributing authors. This book is intended to serve as an important resource and review to the researchers in the field of DNA sequencing.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Bharti Rajendra Kumar (2012). DNA Representation, DNA Sequencing - Methods and Applications, Dr. Anjana Munshi (Ed.), ISBN: 978-953-51-0564-0, InTech, Available from: <http://www.intechopen.com/books/dna-sequencing-methods-and-applications/dna-representation>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.