

Human Genome Project- A Review

Roushan T¹, Ahmed D², Ali MR³

Abstract

Biotechnology and genetic engineering has advanced medical science so that we are now blessed with treatments of numerous diseases with the gene therapy. Mapping and sequencing of human genome was started in 1989 and ended in 2003. The genome project aimed to find out the function of every gene in human body and replace any gene that is mutated and causes disease. This is a review of the process by which the actual sequencing was done.

Key words- Human genome, Genetic engineering

Introduction

Molecular biology has advanced a lot during the past few decades. Among some mentionable achievements of molecular biology, biotechnology and medical science, mapping and sequencing human genome is a major breakthrough. Following the discovery of double helix structure of deoxyribonucleic acid (DNA) by Watson and Crick in 1953, genetic researches were carried out throughout the world. It triggered interest of scientists towards the physical and chemical structure of DNA, bioinformatics and genetic engineering. Human genome project (HGP) is one of the most challenging quests undertaken by scientists. Human genome project aims to determine the complete sequence of 3.1 billion DNA subunits (bases), identify all human genes and make them accessible for further biological study.

What is Human Genome Project: Every cell in our body has built-in programs which control its development and functions. This program or piece of information is stored in units called genes, which are organized in groups one after another in our chromosomes. "In eukaryotic cells, a structure in the nucleus consisting of chromatin and carrying the genetic information for the cell" best describes what a chromosome is. Humans contain 46 chromosomes in its nucleus of every cell.¹ Chromosomes comprise of deoxyribonucleic acid (DNA). "Each DNA strand consists of a linear sequence of four bases- guanine (G), cytosine (C), adenine (A) and thymine (T) – covalently linked by phosphate bonds. The sequence of one strand of double-stranded DNA determines

the sequence of the opposite strand because the helix is held together by hydrogen bonds between adenine and thymine or guanine and cytosine".²

To understand Human Genome Project, one must have a clear idea about genome. According to Dorland's Dictionary, 2009,³ genome is the entirety of the genetic information encoded by the nucleotides of an organism, cell, organelle or virus. The order of the nucleotide bases in a genome is determined by the DNA sequence. The scientific research, under which approximately 20,000-25,000 genes of human genome are mapped and sequenced, is known as the human genome project (HGP).

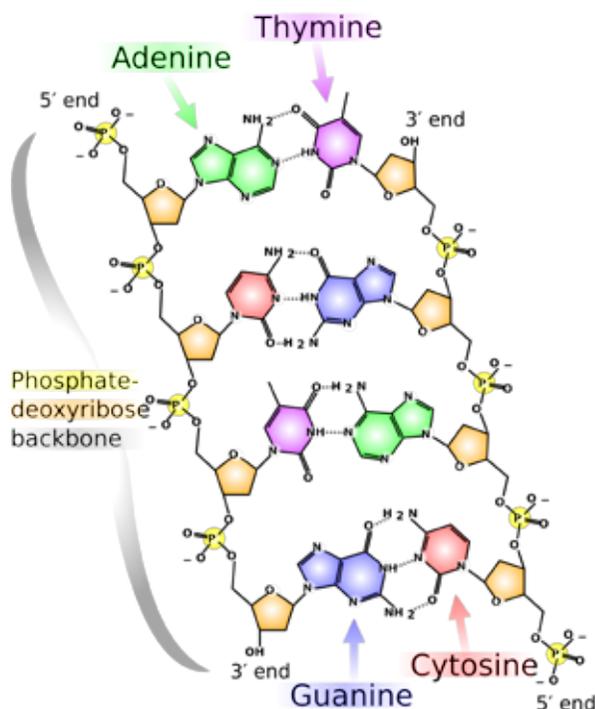


Fig-1: The chemical Structure of a Deoxyribonucleic Acid (DNA)

Historic Background

The initiative of HGP was taken by the US Department of Energy and the National Institute of Health (NIH) in 1989. NIH provided three million dollars to initiate the project; mostly it was a government sponsored research. In 1998 Craig Venter a former researcher of NIH during early 1990s and his firm, Celera Genomics launched a similar type of project which was privately funded. Later on this private and the public projects were collaborated together. Research centers and universities of Germany, Japan, China, France, UK, and USA also had enormous contribution to the HGP.

Human Genome Project⁴ officially began in 1990. It was planned to finish in 2005, a fifteen year long scheme. In 2000 a draft of the sequenced genome was presented. However

1. Tanvir Roushan, B. Sc in CSE
BRAC University, Mohakhali, Dhaka
2. Dr. Dilruba Ahmed
Assistant Professor
Department of Community Medicine
Medical College for Women & Hospital, Uttara, Dhaka
3. Corresponding author :
Professor Dr. Md. Roushan Ali
Professor & Head
Department of Medicine
Medical College for Woman and Hospital, Uttara, Dhaka

in April 2003 the complete human genome was presented and announced to have come to an end two years ahead of schedule, with high degree of accuracy.

It was possible to complete this huge task two years earlier than planned because of the tremendous advancement of technology. These huge analyzed data were stored in super computers and data centers called “GeneBank”. By the improvement of tools and know-how, more and more bases were being sequenced and analyzed in lesser time. Fewer protein coding was done than expected. It was seen that 1.5% of DNA were code DNA, and the rest, called junk DNA, had nothing to sequence in them.



Fig-2: Gene Bank located in Aleppo, Syria, holds approximately 25% of the CGIAR’s total germplasm collections. HGP used similar Gene Banks

Process of DNA Sequencing

Now the question arises, how this sequencing was done, which procedure did the scientists choose to proceed? The process of polymerase chain reaction (PCR) and automated sequencing via computers are most widely used for analysis of genes. The process is described in brief. Chromosomes are broken into shorter pieces with small number of base pairs in each piece. Fragments in a set are separated by gel electrophoresis.

Initialization step: It consists of heating the reaction to activate the DNA polymerase.

Denaturation step: Here temperature is lowered to break the hydrogen bond between the base pairs, and DNA synthesis starts.

Elongation step: In this step optimum temperature is maintained suitable for the DNA polymerase to synthesize a new complementary strand.

After these separating suitable DNA strand base at the end of each fragment is identified. This recreates the original sequence of As, Ts, Cs and Gs. Then computers are used to make long continuous stretches from short sequences that are analyzed for errors, gene-coding regions, and other characteristics.

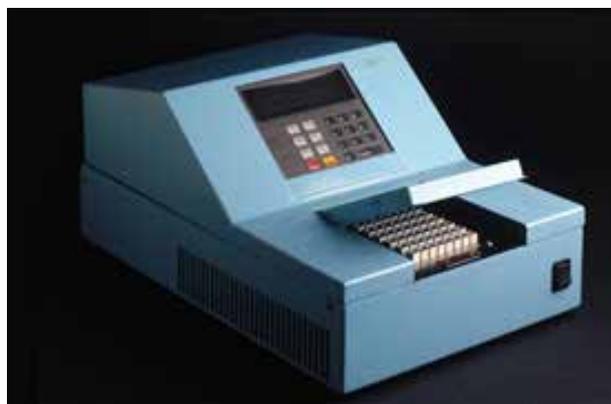


Fig-3: PCR machine automates the process of making large quantities of DNA from a tiny starting sample. HGP was carried out through these PCR machines

Algorithm

Needleman-Wunsch algorithm was used to sequence data from the DNAs.⁵ “It is an example of dynamic programming (DP), and was the first application of DP to biological sequence comparison (Needleman-Wunsch algorithm, 2012). The genome project used a “divide and conquer” strategy. It is a parallel algorithm which brought down an polynomial n^2 algorithm with complexity of $O(N \times M)$ where N is the length and M is depth of a 2D array, to a reduced time linear algorithm with complexity of $O(N+M)$.

As time went on, cost reduced and more genes were sequenced. The following graph shows that clearly

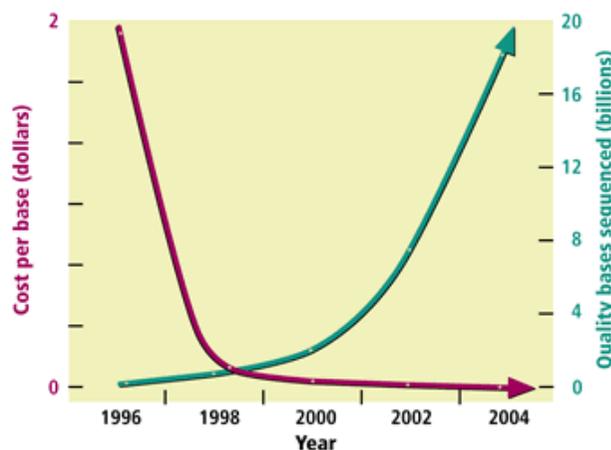
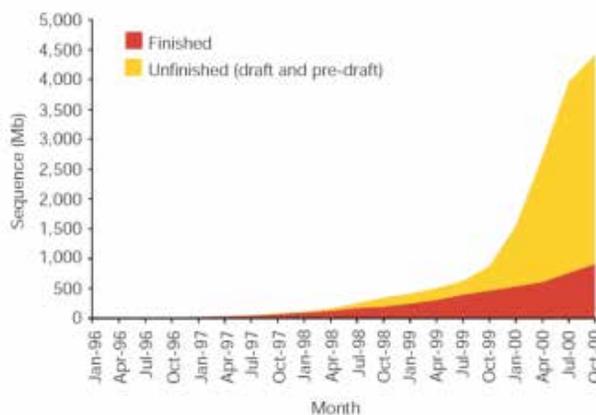


Fig-4: DNA Sequencing vs. Time Graph and Sequencing Cost per Base vs. Quality Bases Sequence

Now that the research is finished post HGP research is going on. The genome sequence contains cure of almost all diseases. Till now only a few gene positions are known, the genome project aims to find out the function of every gene in human body and replace any gene that is mutated and causes disease.

Future: The gene therapy can be a great field of application of the data found from the HGP. The sequences of the genome will be useful to treat any disease. However wrong placement of gene therapy vectors will have a huge impact and might even cause outbreak of new untreatable diseases. A case can be taken into account. Experimental gene therapy patient Rhys Evans had SCID-X1 (Severe Combined Immunodeficiency) disease which was caused due to mutation in his genes interleukin-2 receptor subunit gamma or IL2RG. Patients with this disease cannot fight infections. Rhys and eleven other children after treatment with gene therapy repaired their immune system and lived well.⁶ Despite all the success and potentials of gene therapy, it is not used worldwide. It has not been approved for commercial use. There are other setbacks also, which slowed down the progress of gene therapy. "Death of 18 year old Jesse Gelsinger was a major setback. Jesse was being treated for Ornithine Transcarbamylase Deficiency (OTCD). He died from multiple organ failure four days after starting the treatment."⁷

Researchers and scientists are now looking ahead to add the 47th chromosome into humans which would not affect the other 46 chromosomes' functions, but will stay inside human cells to replace any gene that is mutated. Similarly researchers are working on creating a way to cure Human Immunodeficiency Virus Acquired Immune Deficiency Syndrome (HIV AIDS). These are all possible since the complete mapping of the human genome is done.

Ethical Issues

There are ethical issues along with social concerns. The question remains, who will benefit from the genome's revealed treasures. The whole process took a lot of money and effort. And the ongoing post researches are also costly. The goal of these experiments is to treat human beings. But we cannot expect treatment would come easy and cheap. They will always be beyond the reach of maximum population. Investing on clinically proven treatment procedures would

have helped more people to be treated rather than the human genome project. If all the genome functions are known, researchers might go on to make super humans by gene therapy, which again knocks the ethical wall. Potential benefits must be weighed across its potential risks, including ecological and social repercussions.

At the end I would like to remind all about the hard work of many scientists for years, by which the mysteries of life is unveiled and the field of bioinformatics and genetics took a quantum leap. Though human genome project was completed in 2003, it still has a long way to reach its ultimate goal.

References

1. About the Human Genome Project. (2011). In *Genomics.energy.gov, Human Genome Project Information*. Retrieved on February 4, 2012, from http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml
2. Davidson, S. Molecular and genetics factors in disease. In Colledge, N. R., Walker, B. R. & Ralstor, S. H. (Eds.). *Davidson's Principles & Practice of Medicine*. (21st ed.).bChurchill: Livingstone Elsevier. 2010;40
3. Elsevier. *Dorland's Pocket Medical Dictionary* (28th ed.). Delhi: Thomson Press India Ltd. 2009.
4. Human Genome Project. (2012). In *Wikipedia, the Free Encyclopedia*. Retrieved February 5, 2012, from http://en.wikipedia.org/w/index.php?title=Human_Genome_Project&oldid=476999778
5. Parallel Needleman-Wunsch Algorithm for Grid, by Tahir Naveed1, Imtaz Saeed Siddiqui2, Shaftab Ahmed. Retrieved from: <http://www.cloudbus.org/~alchemi/files/Parallel%20Needleman%20Algo.pdf>,
6. Evers C. A. & Starr L. Studying and Manipulating Genomes. *Biology Concepts and Applications*. (6th ed.). India: Baba Barkha Nath Printers, Haryana. 2007;230-233
7. Gene Therapy. (2011). In *Genomics.energy.gov, Human Genome Project Information*. Retrieved on February 5, 2012, from http://www.ornl.gov/sci/techresources/Human_Genome/medicine/genetherapy.shtml