Human genome project (HGP)



Cracking down The life's code

Dr. Anshumala Chaturvedi

Assistant Professor

Zoology Department

Career College Bhopal,

sanshu47@gmail.com

Content

- Introduction of HGP
- Goals of HGP
- Estimate and reality of HGP
- Methodology
- Sequencing
- Maps created
- Application

Introduction of HGP

- Genome A total number of genes present on a haploid set of chromosome.
- The human genome project was an international scientific research project that aimed to determine the complete sequence of nucleotide base pairs that make up human DNA and all the genes it contains.
- It remains the world's largest collaborative biological project.
- The idea was picked up by the US government in 1984 when the planning started, the project was formally launched in 1990 and was declared complete in 2003.

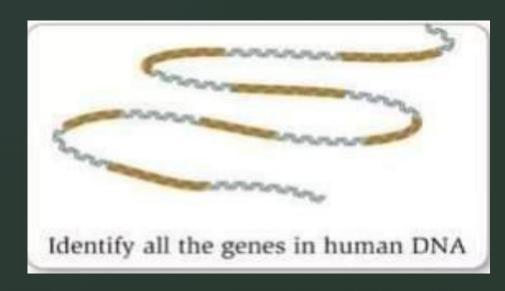
- The genome of any given individual is unique; mapping the human genome involved sequencing the genomes of a small number of individuals and then assembling these together to get a complete sequence of each chromosome.
- The finished human genome is thus a mosaic, not representing anyone individual.

- In 1997 United States established the National human genome research Institute (NHGRI)
- The HGP was an international research groups from six countries - USA, UK, France, Germany, Japan and China and several laboratories and a large number of scientist and technicians from various disciplines.
- HGP provide a new area for research work And develop a new branch called <u>Bioinformatics</u> where statistics and data were maintained.

Goals of HGP

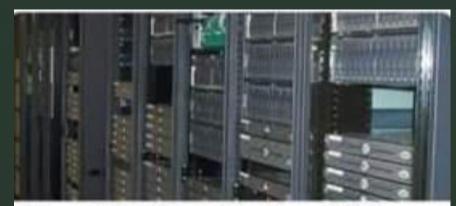
- To identify and map all the gene in the human DNA from a physical and functional stand point.
- To determine the sequence of the three billion chemical base pairs that make up the human DNA.
- To store these informations in databases.
- Also to sequence the genomes of other organisms that are important in medical research such as mouse, drosophila etc,.

Human genome project: Goals



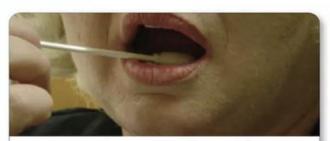


Determine the sequences of the 3 billion nucleotides that make up human DNA

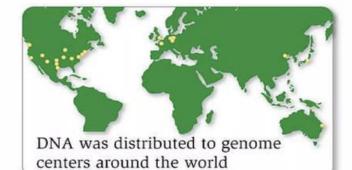


Store this information in publicly accessible databases

The Human Genome Project: How it was done



DNA samples were collected from several anonymous volunteers





Researchers performed mapping, sequencing, and analysis using specialized equipment



Sequences were placed into a central database





Image credit: NCBI/NLM/NIH

Estimate of HGP

- 1. It was 15 years plan.
- 2. They assume that human genome contain 80,000 to 1,00,000 gene.
- 3. Approximately 3 billion Nucleotides Were present.

Reality of HGP

- 1. human genome contain 20,000 to 25,000 gene
- 2. project was completed before time (2003).

Methodology

Two major approach:

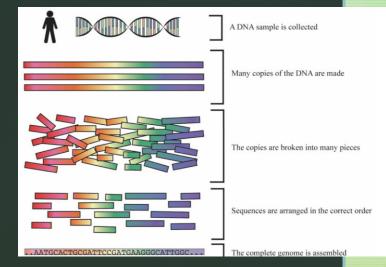
- identify all genes that are expressed as RNA That means all the mRNA work as a functional gene and expressed gene is known as <u>EST</u> (Expressed sequence tag's)
- Blind approach That means simply sequencing the whole set of genome. It is a time taking process because sequence annotation is must here

How to do sequence annotations?

Take a Long strand of DNA break them into small fragments then clone them in a suitable vector (*bacterial artificial chromosome and yeast* artificial chromosome) the introduce the vector inside the host (*bacteria and yeast*) and create a multiple copies.

Sequencing

- Automated DNA sequencer The method given by
 <u>Frederick sanger.</u> He design machine and technique to sequence hole DNA easily.
- He also determine the sequence of amino acids in protein.
- Overlapping sequence DNA sequence get highly coiled with each other, so to solve the overlap sequence into a straight line it required a computer based program (PCR polymerase chain reaction) by Kary Mullis.



- Sequence of chromosome no.1 was completed in may 2006. Its sequence will complete lastly .
- Sequencing of 22 autosomes, X+Y chromosome will be completed in 2003, only chromosome no.1 took longer time to complete.
- Last chromosome sequence is Y chromosome.

Maps created

- Another challenge was assigning the genetic and physical maps on the genomes. This was generated using information on polymorphism of restriction endonuclease recognition sites, and some repetitive DNA sequences known as microsatellites (1-6repeats).
- The average gene we get is minimum 3000bases.
- Largest human gene is called <u>Dystrophin</u> and it is around 2.4 million bases.
- 99.9% of nucleotides sequence of every human being is same only
 0.1% difference will be seen .

- 50% genes functions are undiscovered till now.
- Less than 2% of gene were coded for protein.
- Chromosome no.1 carry maximum no. Of gene and that is 2,968
- Chromosome Y carry least no. Of gene that is 231 only.
- There are 1.4 million locations in the genome were only single base pair differes (SNP's) single nucleotide polymorphism.

1. Genetic linkage map

Arrangement or distance between the Gene and the marker (a fix point)

2. Physical linkage map

It detect the actual location of the gene with no function.

- Scientists sequence different kind of organisms :
- 1. Arabidopsis thaliana plant
- 2. Saccharomyces cerevisiae yeast
- 3. Caenor habditis elegans worm
- 4. Mus musculus mouse
- 5. Mycobacterium genitalium (having smallest genome) 569 kilo base pair genome
- 6. Hemophilus influnzae 1830 kbp genome
- 7. Methanococcus janaschii 1660 kbp genome
- Yeast having 16 chromosome and having 12500 kbp genome.

Application

- The sequencing of the human genome holds benefits for many fields from molecular medicine to human evolution
- Help in identifying disease causing genes.
- Identification of mutations linked to different form of cancer.
- The sequence of DNA age store In data base available to anyone on the internet
- Benefited the advancement of forensic science
- Will allow for advance in agriculture through genetic modification to yield healthier and more disease resistance crops.

