Human Genome Project :→

- Genetic make-up of an organism or an individual lies in the DNA sequences. If two individuals differ, then their DNA sequence should also be different, at least at some places. These assumptions led to the quest of finding out the complete DNA sequence of human genome.

- A very ambitious project of sequencing human genome was launched in the year 1990.

- HGP is defined as mega project due to
(i) Huge cost estimated to be 9 billion US dollars.
(ii) Very large number of base pairs to be identified and sequenced
(iii) If the obtained sequence were to be stored in typed form in books, and if each page of the book contained 1000 letters and each book contained 1000 pages, than 3300 such books would be required to store the information of DNA sequence from the single human cell.

(iv) The enormous data expected to be generated also necessitated the use of high speed Computational devices for data storage and retrieval and analysis.

(v) The science of Bioinformatics also developed during this period and helped HGP.

** Bioinformatics is a new field in which computer hardware and software technologies are developed and used together, store, analyse and disseminate biological data, images and other information.

* Goals of HGP :→ Some are as follows.

(i) Identification of all approx 20,000 - 25000 genes in human

(ii) Determine the sequences of the 3 billion chemical bP that makes up human DNA

(iii) Store this information in data base.

(iv) Improve tool for data analysis

(v) Transfer related technologies to other sectors, such as industries

(vi) Address the ethical, legal and social issues

(ELSI) that may arise from project.

- The HGP was 13-year project co-ordinated by the U.S. Department of Energy and the National Institute of Health.
- During early yeas of the HGP, the Welcome Trust (U.K.) became a major partner.
- Additional contribution came from Japan, france, Germany, china and others.
- The Project was completed in 2003.
- Besides providing clues to understanding human biology, learning about non-human organisms DNA sequences can lead to an understanding of their natural capibities that can be applied towards solving challenges in health care, agriculture, energy production, environmental remediation.

★★★

- Many non-human model organisms, such as bacteria, yeast, Caenorhabditis elegans (a free living non-pathogenic nematode), Drosophila (the fruit fly), Plant (rice and Arobidopsis, etc. have also been sequenced.

Methodologies :-

* There are two approaches for sequencing human genome.

(i) ESTs (Expressed sequence Tags) :- Identify all genes that are expressed as RNA and sequence the same

(ii) Sequence Annotation :- It took blind approach of simply sequencing the whole set of genome that contained all the coding and non-coding sequences, and later assigning different regions in the sequence with functions.

* In HGP, sequence annotation has been carried out which involve following steps

(i) The whole DNA of the cell is isolated and broken randomly into fragments. (b/c there is technical limitition in sequencing very long piece of DNA).

(ii) These fragments are inserted into specialised vectors called $\boxed{\text{BAC}}$ (Bacterial artificial chromosome) and $\boxed{\text{YAC}}$ (Yeast artificial chromosome)

(iii) These fragments are cloned in suitable host like Bacteria and Yeast result into amplification of them.

(iv) The fragments were sequenced using automated DNA sequencers that worked on the principle of a method developed by Frederick Sanger. [Sanger is also credited for developing method for determination of amino acid sequence in proteins. Sanger was Double Nobel laureate]

(v) These sequences were then arranged on the basis of some overlapping regions. This required generation of overlapping fragments for sequencing.

(vi) Computer based programmes were used to allign the sequences.

(vii) These sequence were then annotated and were assigned to each chromosome. the sequence of chromosome-1 was completed only in May 2006 This was last chromosome out of 24 to be sequenced.

***
(viii) with the help of polymorphism in microsatellites and restriction endonuclease recognition sites, the genetic and physical maps of the genome have also been prepared.

Salient features of Human Genome :→

1. The human genome contains 3164.7 million bP
2. The average gene consists of 3000 bases, but the size vary greatly, with the largest known human gene being dystrophin (Duchene Muscular Dystrophy on X-chromosome) at 2.4 million bP.
3. The total number of genes is estimated at 30,000 – much lower than previous estimates of 80,000 to 1,40,000 genes.
4. All most all (99.9%) nucleotide bases are exactly the same in all human beings.
5. Less than 2% of the genome codes for proteins
6. The functions are unknown for over 50% of the discovered genes.
7. Repeated sequences make up very large portion of human genome.
8. Repeated sequences are stretches of DNA sequences that are repeated many times, sometimes hundred to thousand times. They are thought to have no direct coding functions, but they shed light on chromosome structure, dynamics and evolution.
9. Chromosome-1 has maximum number of genes (2968), and the chromosome 'Y' has minimum

number of genes (231).

10. Scientists have identified about 1.4 million locations where single base DNA differences (SNPs - single nucleotide polymorphism, pronounced as 'snips') occur in humans. This information will be helpful in finding chromosomal locations for disease-associated sequences and tracing human history.

Applications & Future challanges :→

- * Deriving meaningful knowledge from the DNA sequences will define research through the coming decades leading to our understanding of biological systems.
- * The above enormous task require the expertise and creativity of tens of thousands of scientists from varied disciplines in both public and private sectors worldwide.
- * One of the greatest impacts of having the HG sequence may well be enabling a radically new approach to biological research.

- In the past, researchers studied one or few genes at a time. With whole-genome sequences and new-throughput technologies, we can approach questions systematically and on a much broader scale.
- They can study all the genes in a genome. for example all the transcripts in a perticular tissue or organ or tumor, or how tens of thousands of genes and Proteins work together in interconnected networks to orchestrate the chemistry of life.