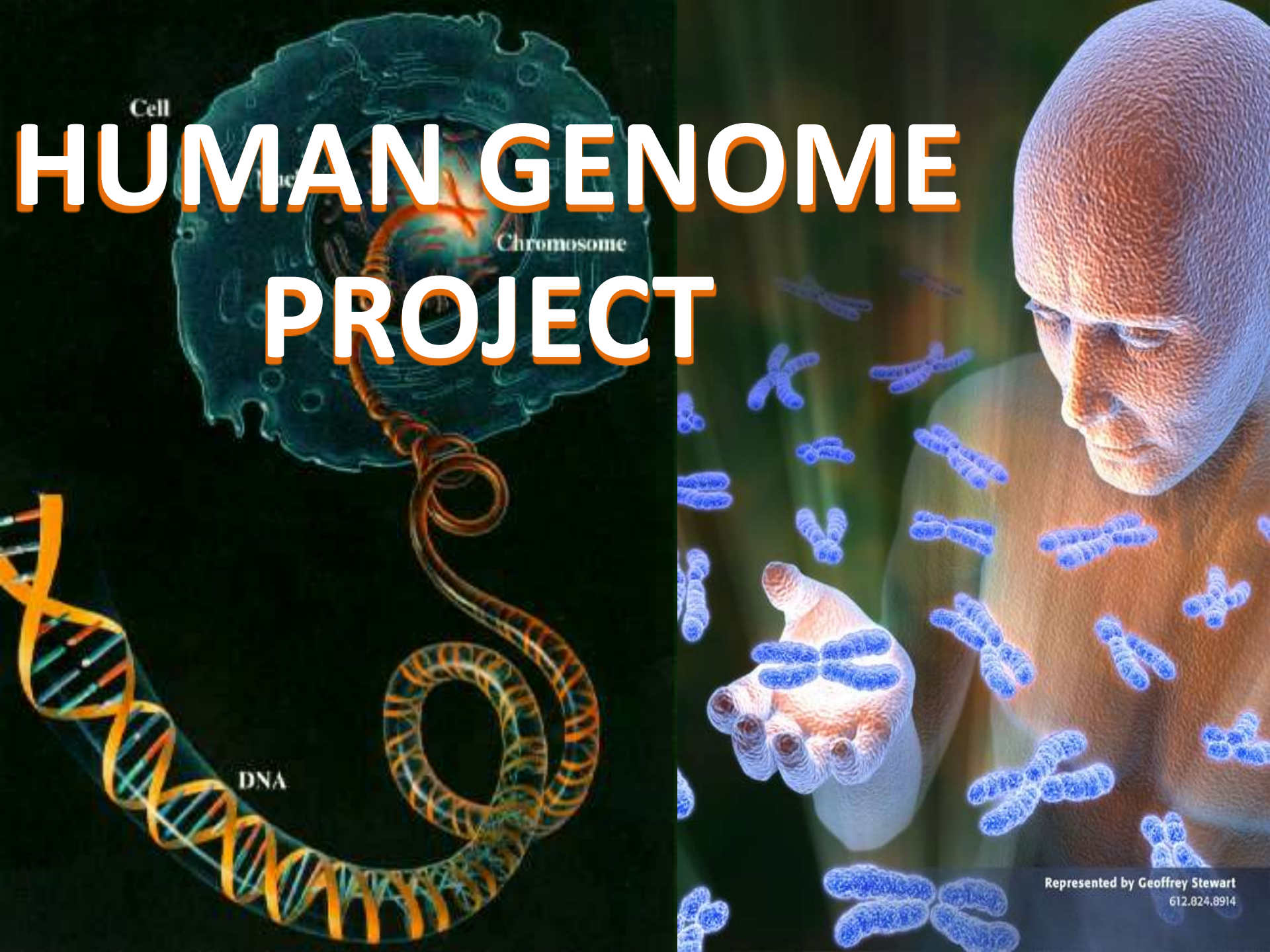


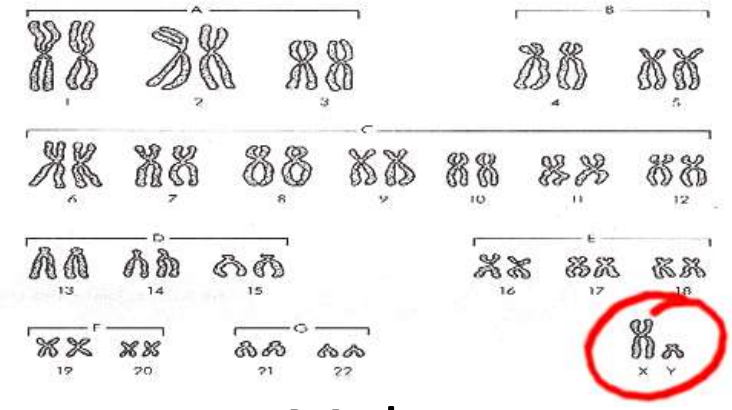
HUMAN GENOME PROJECT

At a glance

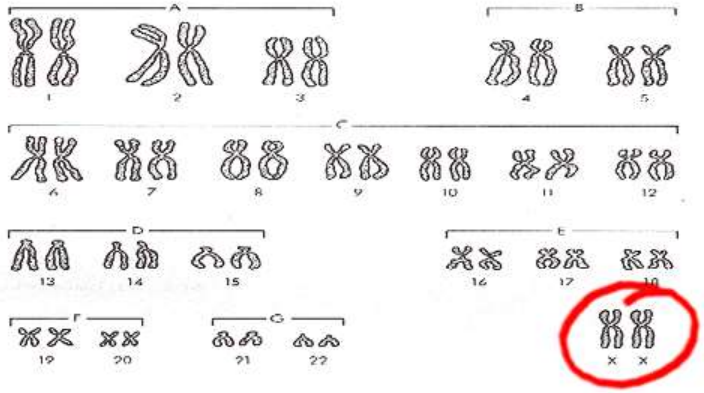
- The Human Genome
  - Salient Features of Human Genome
- What was Human Genome Project(HGP)
  - Milestones
  - Goals of Human Genome Project
  - Issues of concern
  - Future Challenges
- Vectors for Large-Scale Genome Project
  - Yeasts artificial chromosomes
  - Bacterial artificial chromosome
  - Difference between YAC and BAC

The Human Genome

- The **human genome** **is the complete set of *genetic information* for humans** (*Homo sapiens*).
- The human genome is by far **the most complex and largest genome**.
- Its size spans a length of about **6 feet of DNA, containing more than 30,000 genes**.
- The DNA material is organized into a haploid chromosomal set of **22 (autosome) and one sex chromosome (X or Y)**.

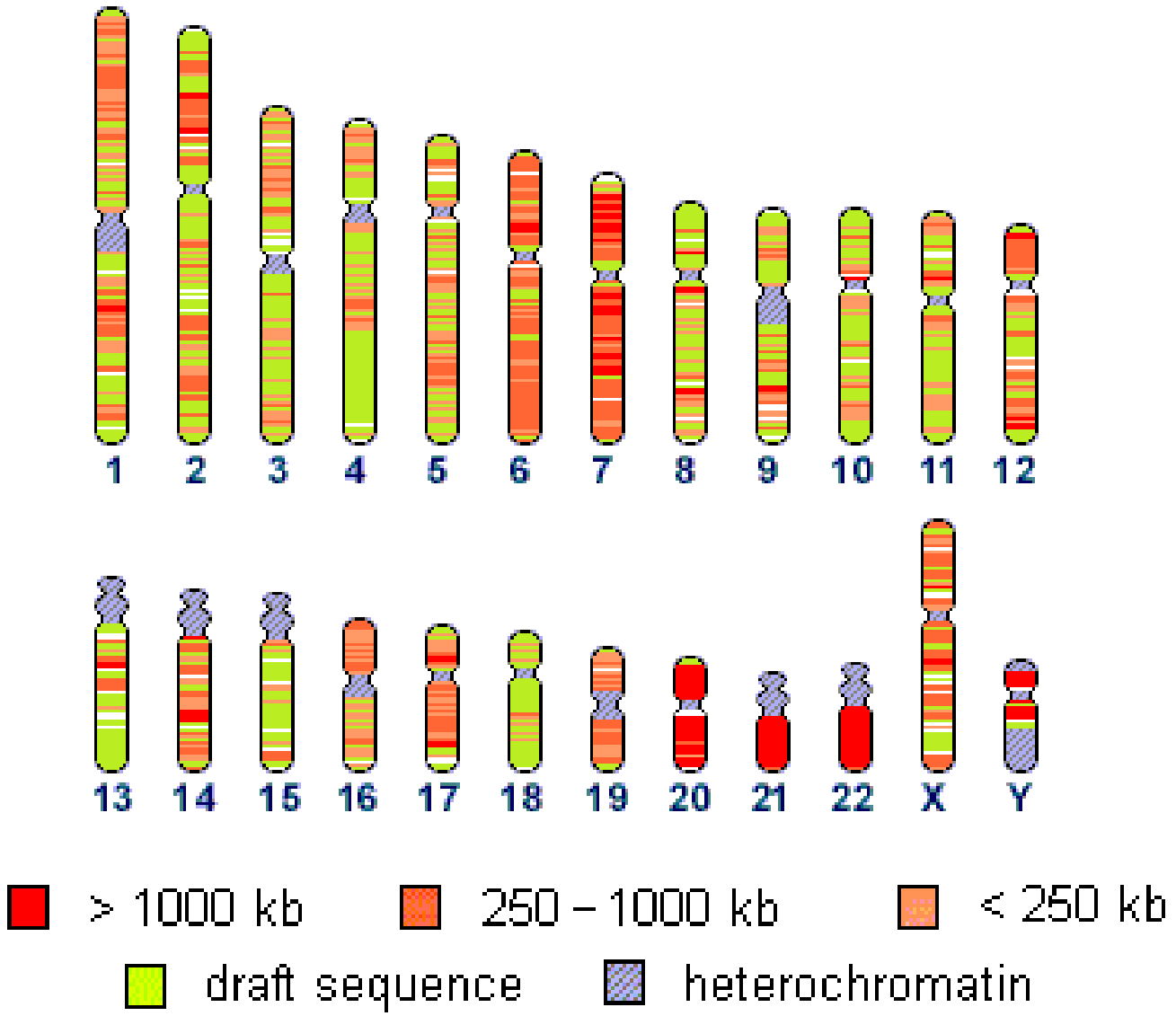


Male



Female

Human Genome Sequencing 2/11/2001
22 autosome + 2 sex chromosomes



From NCBI

Salient Features of Human Genome:

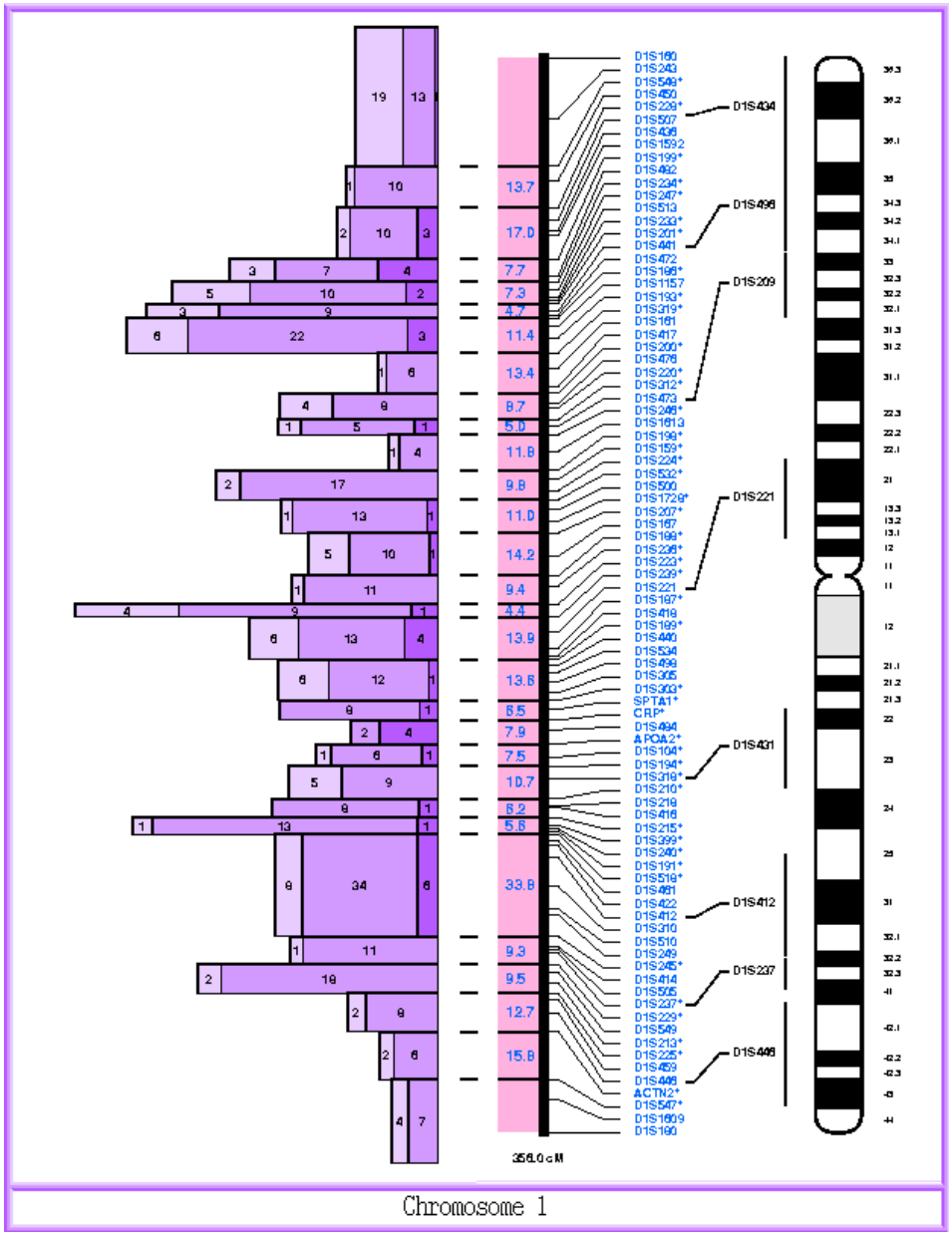
- Human genome consists the information of **24 chromosomes** (22 autosome + X chromosome + one Y chromosome); in *Homo sapiens* 2n = 2x = 46
- The human genome contains **over 3 billion nucleotide pairs**.
- Human genome is estimated to have **about 30,000 genes** .
- Average gene consists **of 3000 bases**. But sizes of genes vary greatly, with the largest known human gene encoding dystrophin containing 2.5 million base pairs.
- Only about **3 %of the genome** encodes amino acid sequences of polypeptides and **rest of it junk** (**repetitive DNA**).
- The **functions are unknown** for over **50%** of the discovered genes.

Continue……

- The repetitive sequences makeup very large portion of human genome. Repetitive sequences have no direct coding function but they shed light on the chromosome structure, dynamics and evolution.
- Chromosome **1 has most genes** (**2968**) and **Y chromosome has the lowest** (**231**).
- Almost all nucleotide bases are exactly the same in all people. Genome sequences of different individuals differ for **less than 0.2% of base pairs**.

  Most of these differences occur in the form of single base differences in the sequence. These single base differences are called **single nucleotide polymorphisms (SNPs).** One SNP occurs at every ~ 1,000 bp of human genome. About 85% of all differences in human DNAs are due to SNPs.

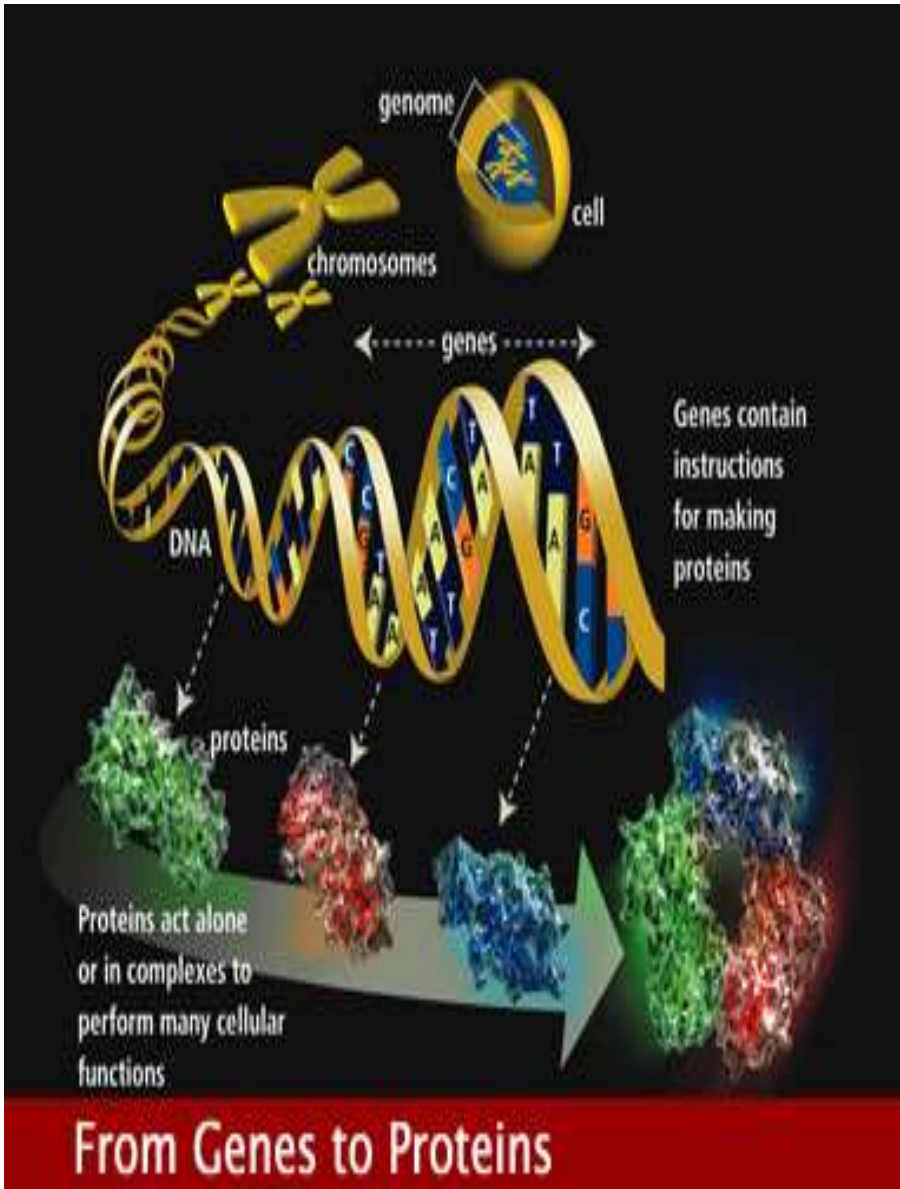
Human Chromosome 1 Genetic Map



What was Human Genome Project(HGP)

- The Human Genome Project was an international research effort to determine the **sequence of the human genome** and identify the **genes that it contains.**

- The US Human Genome Project is a 13 year effort, which is coordinated by the
  - **Department of Energy (DOE) and**
  - **National Institutes of Health (NIH).**



Milestones

1986 The birth of the Human Genome Project.

1990 Project initiated as joint effort of US Department of Energy and the National Institute of Health.

1994 Genetic Privacy Act: to regulate collection, analysis, storage and use of DNA samples and genetic information is proposed.

1996 Welcome Trust joins the project.

1998 **Celera Genomics** (a private company founded by Craig Venter) formed to sequence much of the human genome in 3 years.

1999 Completion of the sequence of Chromosome 22-the first human chromosome to be sequenced.

2000 Completion of the working draft of the entire human genome.

2001 Analysis of the working draft are published.

2003 **HGP sequencing is completed and Project is declared finished two years ahead of schedule.**

Goals of Human Genome Project

1. To identify all the genes in human DNA.
2. To develop a genetic linkage map of human genome.
3. To obtain a physical map of human genome.
4. To develop technology for the management of human genome information.
5. To know the function of genes.
6. Determine the sequences of the 3 billion chemical base pairs that make up human DNA.
7. Store this information in public databases.
8. Develop tools for data analysis.
9. Transfer related technologies to the private sectors.

ISSUES OF CONCERN

Ethical, Legal and Social issues of the Human Genome Project

- Fairness in the use of genetic information.
- Privacy and confidentiality of genetic information.
- Psychological impact, stigmatization, and discrimination.
- Reproductive issues.
- Clinical issues.
- Uncertainties associated with gene tests for susceptibilities and complex conditions.
- Fairness in access to advanced genomic technologies.
- Conceptual and philosophical implications.
- Health and environmental issues.
- Commercialization of products.
- Education, Standards, and Quality control.
- Patent issues.

Future Challenges: What We Still Don't Know

1. Gene number, exact locations, and functions
2. Gene regulation
3. Chromosomal structure and organization
4. Non-coding DNA types, amount, distribution, information content, and functions
5. Coordination of gene expression, protein synthesis, Proteomes and post-translational events
6. Predicted vs experimentally determined gene function
7. Evolutionary conservation among organisms
8. Disease-susceptibility prediction based on gene sequence variation
9. Genes involved in complex traits and multigene diseases
10. Developmental genetics, genomics

Vectors for Large-Scale Genome Project

- A vector is a DNA molecule that has the ability to replicate in an appropriate host cell, and into which the DNA insert is integrated for cloning.
- A vector must have a origin of DNA replication (ori).
- The vector is a vehicle or carrier which is used for cloning foreign DNA in bacteria.
- For genome sequencing, first DNA fragments of the genome must be cloned in appropriate vectors. Two of the most popular vector:

    1. Yeast artificial chromosomes (**YACs**) and
    2. Bacterial artificial chromosomes (**BACs**)

Yeasts artificial chromosomes (YACs):

Yeast artificial chromosomes (YACs) are genetically engineered chromosomes derived from the DNA of the yeast, *Saccharomyces cerevisiae*, which is then ligated into a bacterial plasmid.

- **YACs** were very useful in mapping the human genome because **they could accommodate hundreds of thousands of kilo bases each.**
- YACs containing a mega base or more are known as "**mega YACs**."

A YAC can be considered as self replicating element, because it includes **three specific DNA sequences**:

1. **TEL:** The **telomere** which is located at each chromosome end, protects the chromosome's ends from degradation by nucleases.
2. **CEN:** The **centromere** which is the attachment site for mitotic spindle fibers and necessary for segregation of sister chromotids to opposite poles of the dividing yeast cell. The centromere is placed in adjacent to the left telomere, and a huge piece of human (or any other) DNA can be placed in between the centromere and the right telomere.

Continue……

3. ORI: Replication origin sequences which are specific DNA sequences that allow the DNA replication.

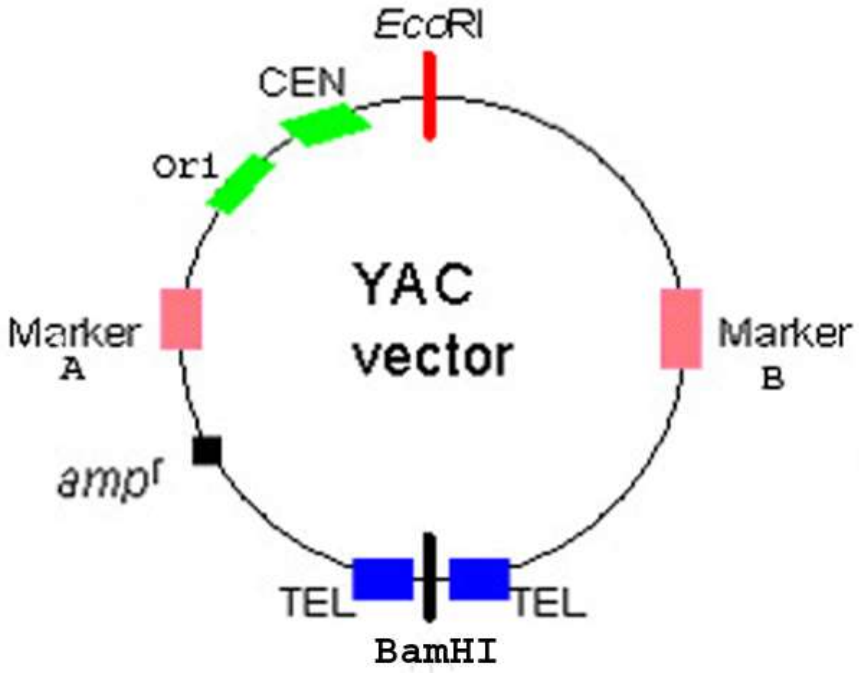


Fig: YAC Vector

It also contains few other specific sequences like:

- **Selectable markers (A** and **B)** that allow the easy isolation of yeast cells that have taken up the artificial chromosome.
- **Recognition site** for the two **restriction enzymes** EcoRI and BamHI**.**

Cloning genomic DNA into a YAC

1. **Genomic DNA** is partially digested with a **restriction enzyme (EcoRI)**.

2. The **YAC** is digested by the two restriction enzymes **EcoRI and BamHI.**

3. Those two elements **recombine at the EcoRI sites of YAC** and are covalently linked by the DNA ligase.

4. A recombinant YAC vector, a yeast artificial chromosome with genomic DNA inserted, is produced. Then YACs vector can be introduced into yeast cells and generated an unlimited number of copies.

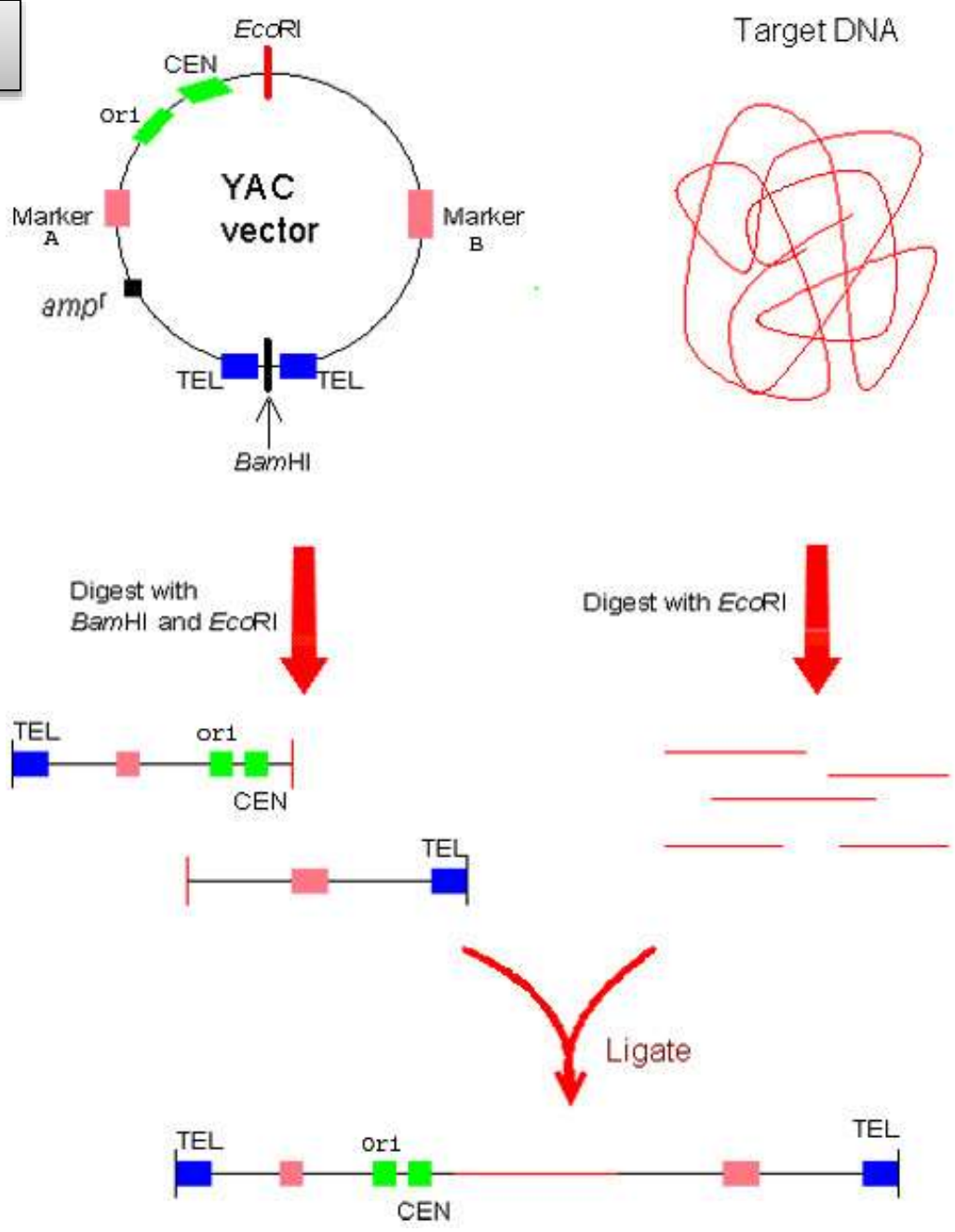


Fig: **Cloning of genomic DNA into a YAC**

Bacterial artificial chromosome (BAC)

A bacterial artificial chromosome (BAC) is an engineered DNA molecule, used to clone DNA segment in bacterial cells (*E. coli*).

It is based on a well-known natural **F plasmid** (inhabits *E. coli* cells). This plasmid allows conjugation between bacterial cells.

- Segments of an organism's DNA, ranging from **150 to about 300 kilo base** pairs, can be inserted into BACs.
- These vectors are able to maintain in **stable state *in vivo* and *in vitro*.**
- Their copy number is about **two per cell**.
- Extensively **used in analysis of large genomes** but the main disadvantage of BAC vectors is some what **laborious construction** of BAC libraries.

Common gene components

Bacterial artificial chromosome is another cloning vector system in *E.coli* (**pBAC108L**), developed by Melsimon and his colleagues in 1992, have

- ***Hin*dIII and *Bam*HI**: the cloning sites
- **CmR**: the **chloramphenicol resistance gene**, used as a selection tool.
- ***oriS*:** the origin of replication
- ***repE***: for plasmid replication and regulation of copy number.
- ***ParA* and *ParB***: the genes governing partition of plasmids to daughter cells during division and ensures stable maintenance of the BAC.

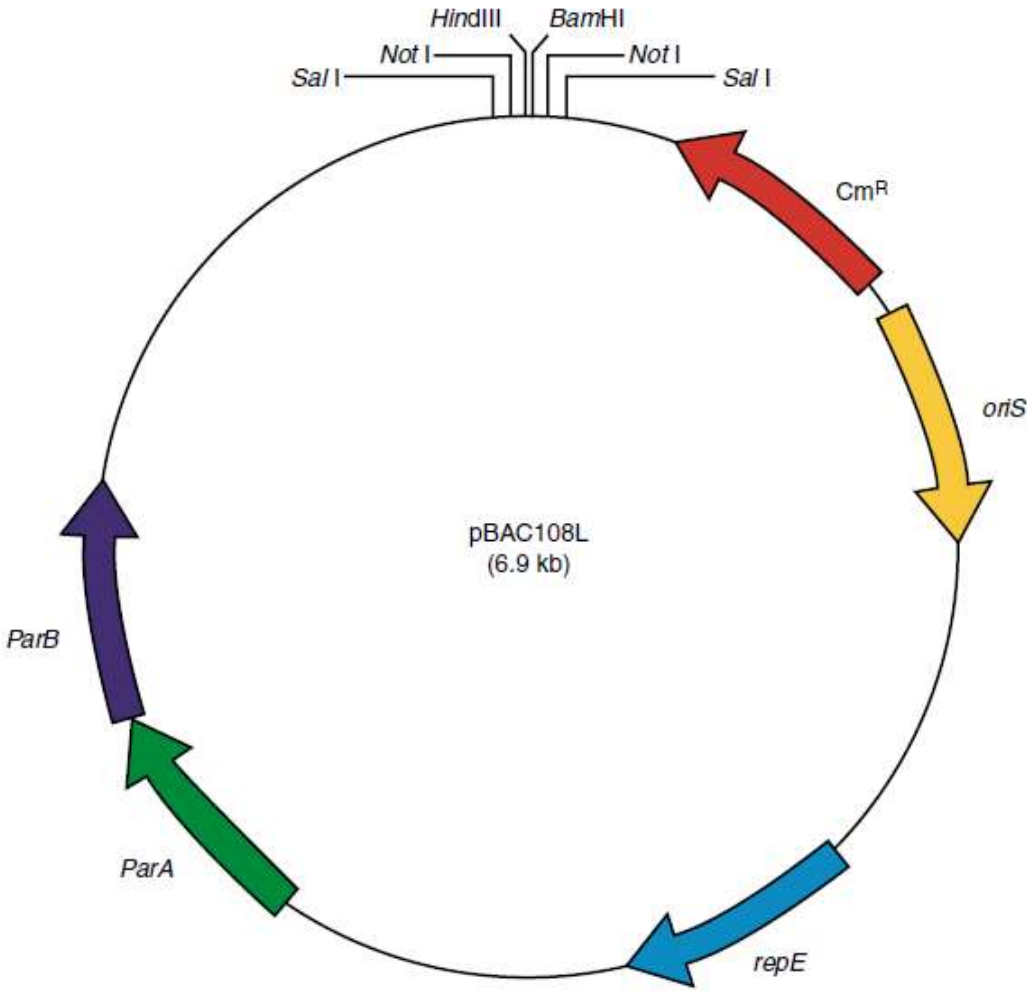


Fig: Map of the BAC vector, pBAC108L

In some conjugation events, the F-plasmid itself is transferred from a donor **F^+ cell** to a recipient **F^- cell**, converting the letter to an F^+ cell.

In other events, a small piece of host DNA is transferred as an insert in the F (which is called an F' plasmid if it has an insert of foreign DNA). And in still other events, the F' plasmid inserts into the host chromosome and mobilizes the whole chromosome to pass from the donor cell to the recipient cell. Thus, because the E. coli chromosome contains over 4 million bp, the F plasmid can obviously accommodate a large insert of DNA.

Cloning genomic DNA into a BAC

1. Genomic DNA is isolated from a desired source and used restriction enzymes to cleave the target DNA into fragments.

2. The BAC is digested by restriction enzymes in the cloning sites ***Hin*dIII and *Bam*HI**.

3. Those two elements recombine by the DNA ligase and attach into a host bacterium.

4. As the bacterial cells grow and divide, they amplify the BAC DNA, which can then be isolated and used in sequencing DNA.

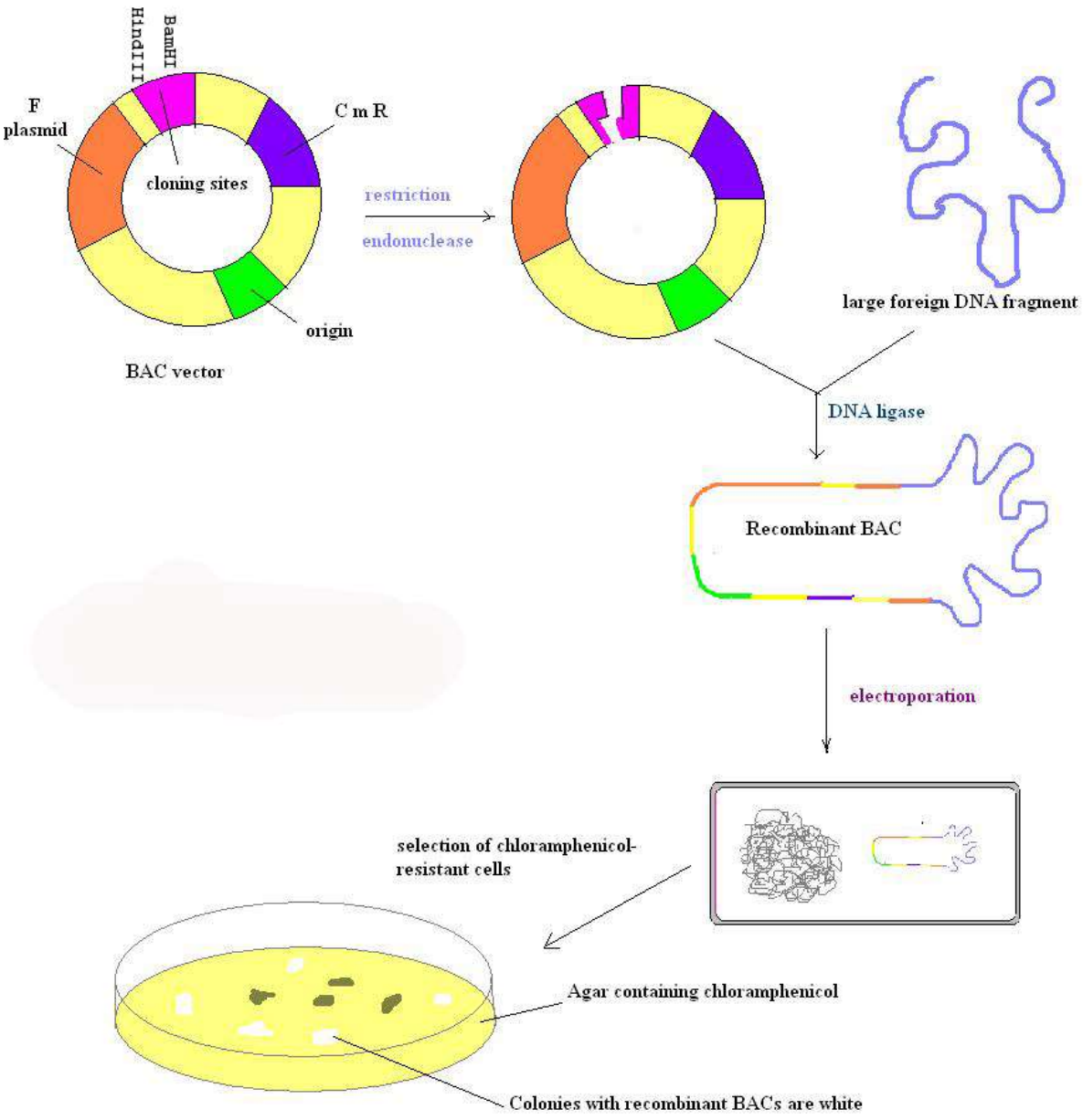


Fig: BAC as a Cloning vector

Difference between YAC and BAC
as vector of genome sequencing

Yeast artificial chromosomes (YACs)	Bacterial artificial chromosome (BAC)
1. Yeast artificial chromosomes (YACs) are genetically engineered chromosomes derived from the DNA of the yeast, *Saccharomyces cerevisiae*.	1. A bacterial artificial chromosome (BAC) is an engineered DNA molecule, used to clone DNA segment in bacterial cells (*E. coli*).
2. YAC's are used for cloning very large (1000-2000kb) DNA segments.	2. These vectors are used to clone the DNA inserts up to 300kb.
3. They are inefficient.	3. They are inefficient.
4. Unlike BAC library, it is not so hard to construct YAC library.	4. It is very hard to construct BAC library.
5. They are unstable.	5. They are more stable.
6. They tend to contain scrambled inserts, i.e. composites of DNA fragments from more than one site.	6. They contain pure inserts.
7. The linear YACs, which tend to break under shearing forces.	7. The circular, super coiled BACs resist breakage.
8. They are hard to isolate from yeast cells.	8. They are easy to isolate.

References

- Weaver RF 2005. Molecular Biology. McGraw-Hill International edition, NY.
- Gardner EJ, MJ Simmons and DP Snustad 1991. Principles of Genetics. John Wiley and Sons Inc, NY.
- Gupta, P.K. 2007. Genetics. Rastogi Publications, Meerut.
- Singh, B.D. 2007. Fundamentals of Genetics. Kalyani Publishers, Ludhiana.
- **Internet**


THANKS
TO
ALL