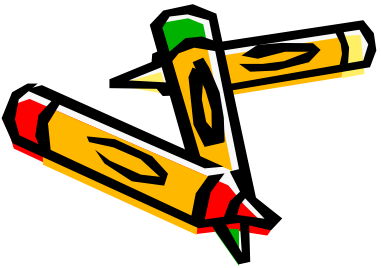
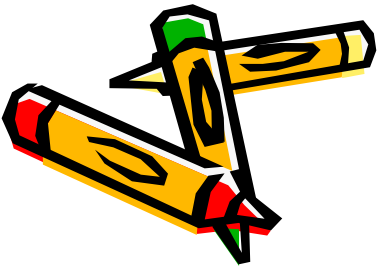
Genome

- The "molecules of heredity" (genetic information) are present in every cell of all living organisms.
- Each molecule is called a "chromosome".
- The genome is the set of all chromosomes.

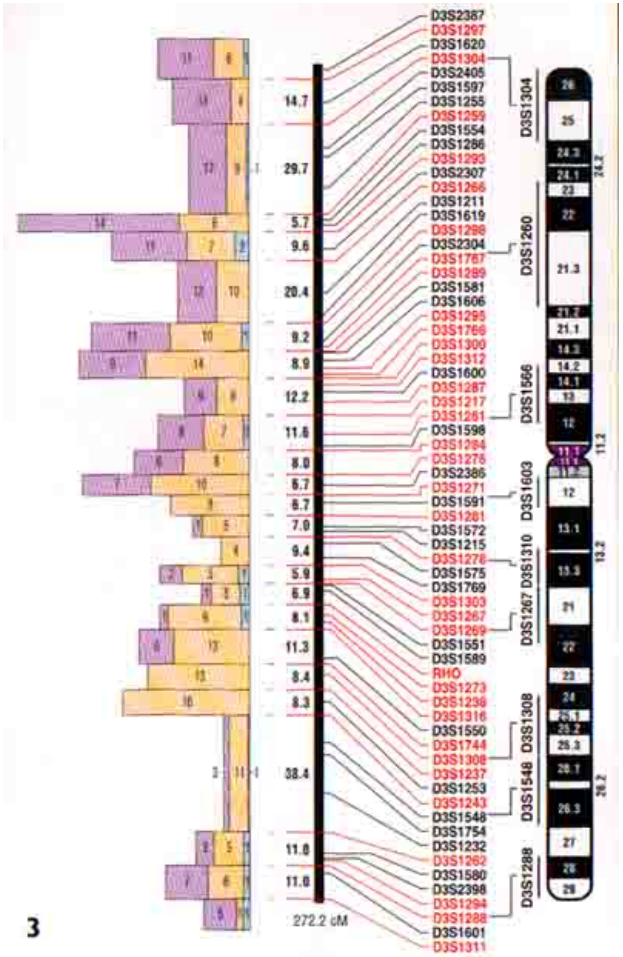


Chromosome

- Chromosomes are made of DNA (deoxyribonucleic acid, 去氧核糖核酸), a double stranded molecule in which each strand is a long succession (sequence) of nucleotides.
- A – T, G - C

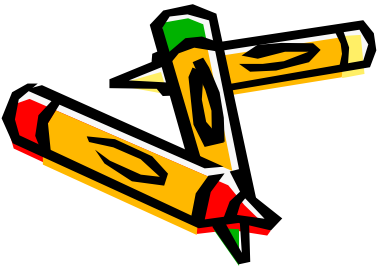


Human Chromoson 3



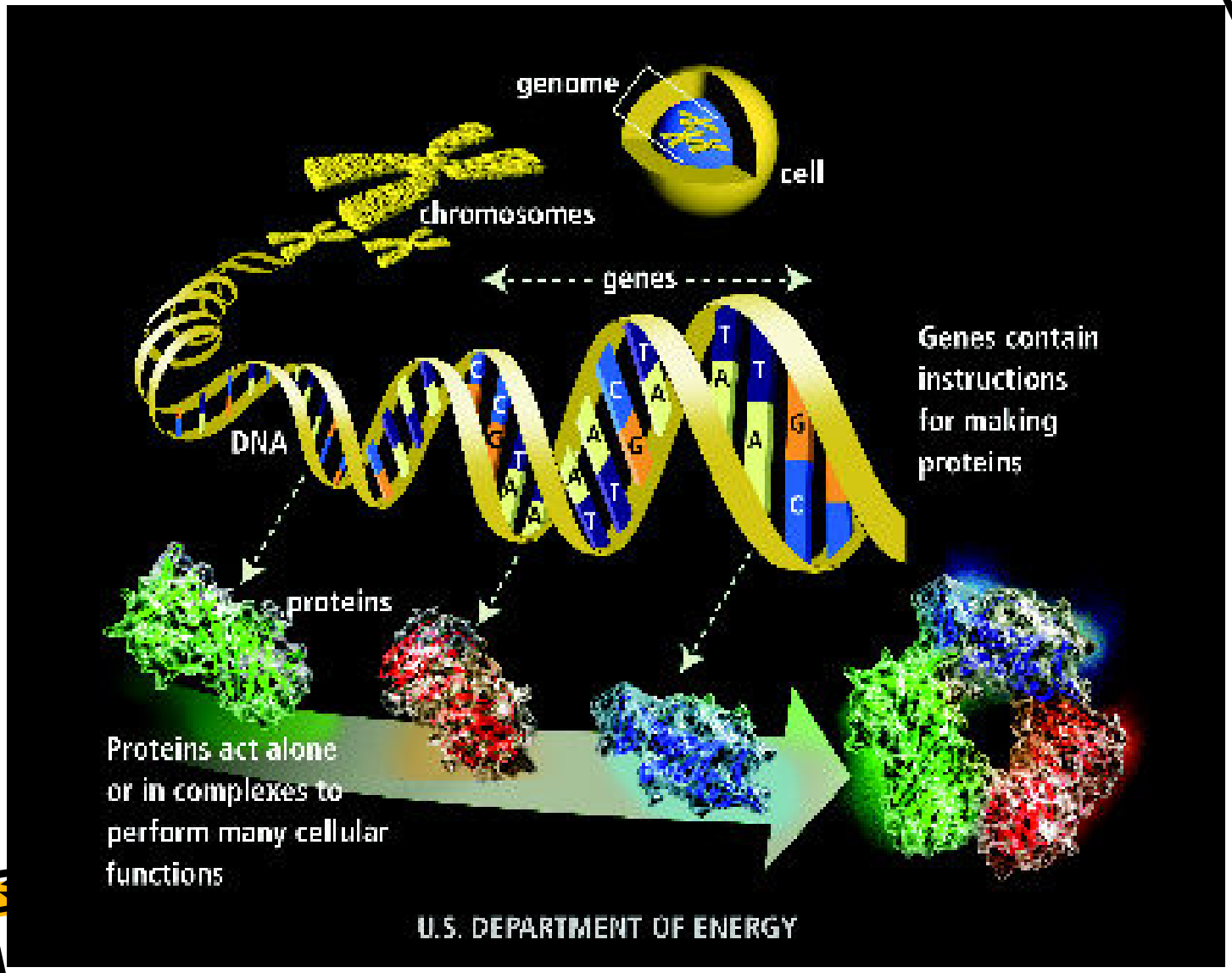
去氧核糖核酸

- DNA: Deoxyribonucleic Acid
- Nucleotide (核苷酸)
  - Adenine (A) : 腺嘌呤
  - Thymine (T) : 胸腺嘧啶
  - Guanine (G) : 鳥嘌呤
  - Cytosine (C) : 胞嘧啶



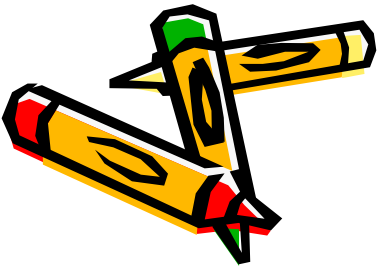
Coupled

- A on a strand is always coupled with a T on the other strand, and a C on a strand is always coupled with a G on the other strand.
- Those two strands are said to be complimentary.
- A DNA molecule is usually represented as a single sequence.


genome
cell
chromosomes
genes
DNA
Genes contain instructions for making proteins
proteins
Proteins act alone or in complexes to perform many cellular functions
U.S. DEPARTMENT OF ENERGY

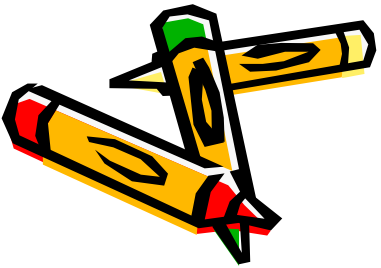
Genes

- A "special" sequence of DNA is called a gene.
- Genes contain instructions of making proteins.
- Proteins act alone or in complexes to perform many cellular functions.



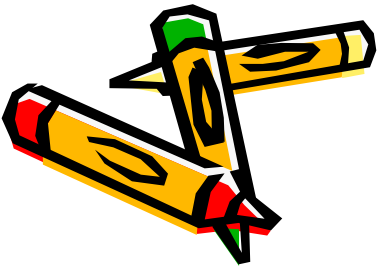
Inside Proteins (蛋白質)

- The twenty *amino acids* commonly found in proteins:
- A Ala Alanine
- C Cys Cysteine
- D Asp Aspartic Acid
- E Glu Glutamic Acid



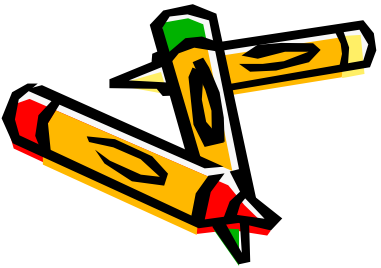
Continued …

- F Phe Phenylalanine
- G Gly Glycine
- H His Histidine
- I Ile Isoleucine
- K Lys Lysine
- L Leu Leucine
- M Met Methionine
- N Asn Asparagine
- P Pro Proline



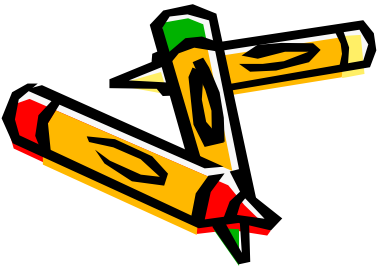
Continued …

- Q Gln Glutamine
- R Arg Arginine
- S Ser Sreine
- T Thr Threonine
- V Val Valine
- W Trp Trptophan
- Y Tyr Tyrosine



About Proteins

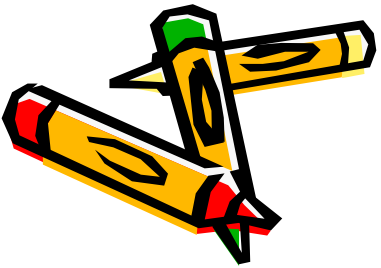
- Proteins are large biomolecules, or macromolecules, consisting of one or more long chains (sequences) of amino acid residues.
- Proteins perform a vast array of functions within organisms: catalyzing metabolic reactions (新陳代謝), DNA replication and responding to stimuli.



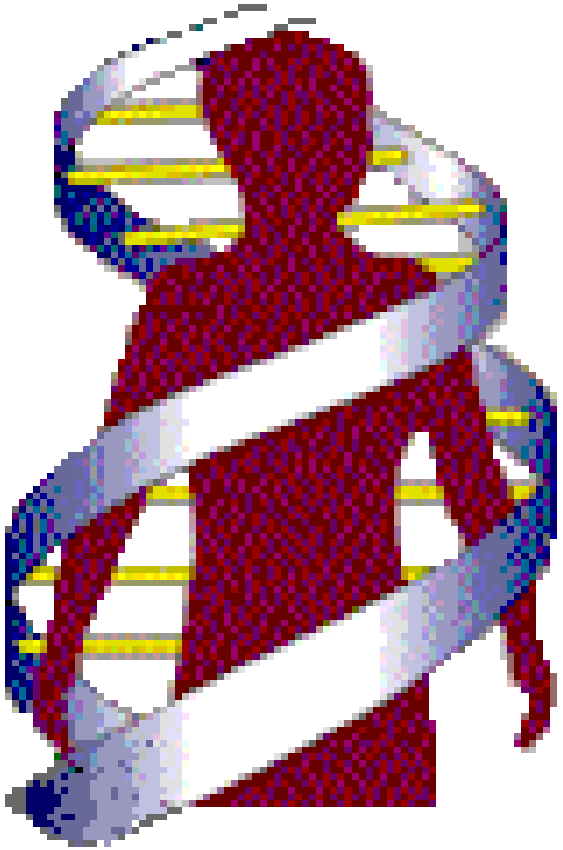
More functions

- Proteins provide structure to cells and organisms, and transporting molecules from one location to another.

(*) Proteins differ from one another primarily in their sequence of amino acids of their genes results in protein folding into a specific 3-dim. Structure that determines its activity.

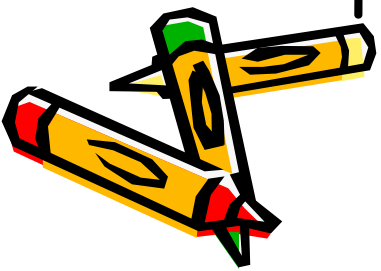


Human Genome Project



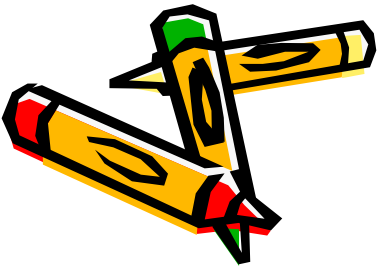
Goals

- Identify all the approximate 30,000 genes in human DNA.
- Determine the sequences of all 3 billion chemical base pairs that make up the human DNA.
- Store this sequence (databases).
- Improve tools for data analysis.



Important Issues

- Develop new technologies in sequencing the DNA sequences.
- Transfer related technologies to the "private" sector, and address the ethical, legal, and issues (social) that may arise from HGP.



Milestones

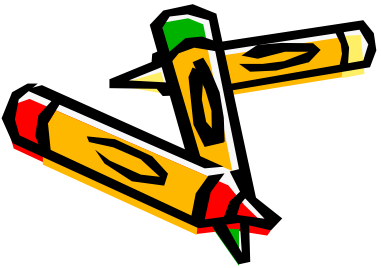
- In 1990, the project was initiated jointly by U. S. Department of Energy and the National Institute of Health.
- A working draft was completed of the entire human genome in 2000 and the analyses are published one year later.
- In April 2003, HGP sequencing is completed and the project is declared finished two years ahead of schedule.

How does the human genome stack up?

Organism	Genome Size (Base pairs)	Estimated Genes
Human (人類)	3 billion	30,000
Laboratory mouse (白老鼠)	2.6 billion	30,000
Mustard weed (*A. thaliana*)	100 million	25,000
Roundworm (*C. elegans*)	97 million	19,000
Fruit fly (果蠅)	137 million	13,000
Yeast (酵母菌)	12.1 million	6,000
Bacterium (大腸桿菌)	4.6 million	3,200
Human immunodeficiency virus (HIV)	9700	9

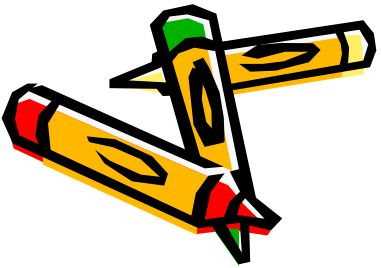
主要課題

- 序列組合
- 序列分析
- 基因認定
- 生物資訊資料庫、種族樹建構、蛋白質三維結構推測…。



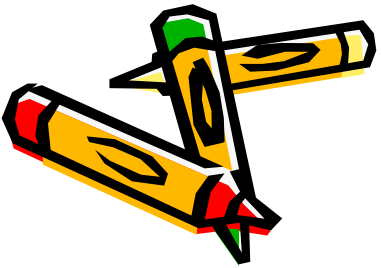
序列比對探究

- 說明
  - 找序列中「相似」及「相異」的部份。
  - 為何要比對序列？
  - 為何要使用電腦比對？
  - 如何使用電腦比對？
  - 困難點：序列型態多樣、需建構不同的資料結構及**演算法**。**如何讀取序列也是一項挑戰!**



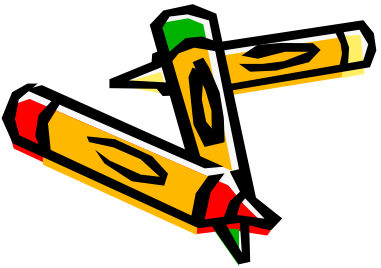
Challenges

- A sequence may evolve (演化) by modifying its organization at a larger scale.
- These large-scale mutations (突變) are called structural variations or rearrangements in the terms of modeling.



Types of rearrangements

1. Deletions: A segment of the genome is lost.
2. Transpositions: A segment of the genome moves to another location.
3. Inversions or reversals: A segment of the genome is reversed and the strands are exchanged.



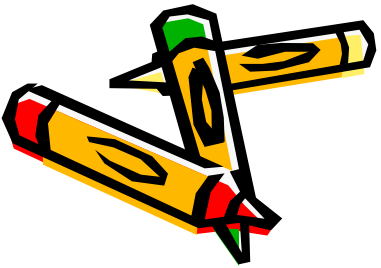
Continued …

4. Duplications

(a) Tandem duplications: insert the copy next to the original.

(b) Retro-transpositions: insert a copy of a gene at an arbitrary location in the genome.

(c) Whole genome duplications: Copy either the whole genome or some of its chromosomes.



More

5. Reciprocal translocation: A segment of a chromosome that contains a telomere is exchanged with a segment of another chromosome that also contains a telomere.

Telomere: **端粒**是真核生物染色體末端的DNA重複序列，作用是保持染色體的完整性和控制細胞分裂週期。 由於DNA複製的機制，每次染色體複製後，延遲股上的染色體末端必無法被複製。因此，真核生物在染色體末端演化出**端粒**以作為可被重複遺棄的部分。

More

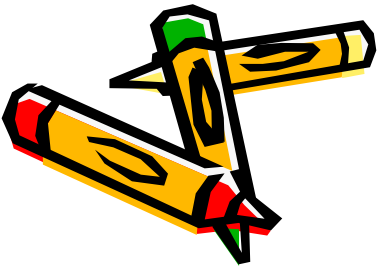
6. Fusion: Two chromosomes are joined into one.

7. Fission: One chromosome splits into two (inverse of fusion).

8. Horizontal transfer: A segment of the genome is copied from one genome to another. (This is common mainly in unicelluar organisms.)

Notices

- All the operations act on a segment at the level of DNA segments rather on nucleotides.
- This is why a genome is often represented by a sequence of segments that are found in an almost identical state in several species, not cut by rearrangements.



Homologous (同族)

- Two segments are said to be homologous if they derive from a common ancestor and are distinguished by a replication event (they end up in two different genomes) or a duplication event (they both belong to the same genome).
- Genes are often taken as those homologous segments due to their functional utility (rarely cut by rearrangement).