
Sequencing and Mapping

CAP 5937-01 Bioinformatics

Fall 2004

Amar Mukherjee

Biology Background

- The genome of an organism is the complete set of the DNA sequences that constitutes its total genetic information content in a cell. This information is wrapped up in a set of **chromosomes** in the cell.
 - The eukaryotes also have an additional set of **extrachromosomal** genes. These are located outside the nucleus of the cell within the energy producing organelles called **mitochondria**.
-

Biology Background

- For plants and algae, there are genes located in the **chloroplasts**. By the word genome, we usually mean the **nuclear genome**.
 - For prokaryotic cell, the genome is a circular DNA molecule.
 - For eukaryotes, like human, the genome consists of a set of linear DNA molecules contained in different chromosomes.
-

Biology Background

- In most eukaryotes, there are two copies of each chromosome, and hence two copies of each gene. This is called the **diploid** complement.
 - The nucleus of a **haploid** contains only one copy of each chromosome, found only in reproductive cells. The number of chromosomes in a genome is characteristic of a given species.
-

Example

Organism	Genome Size(kb)	No. of Chromosomes	Avg. no. of DNA/chromosome
Prokaryotes			
E.Coli	4 000	1	4000
Eukaroytes			
Yeast	20 000	16	1250
Fruit Fly	165 000	4	41 250
Human	3 200 000	23	130 000
Mouse	3 454 200		
Maize	15000 000	10	1 500 000
Salamander	90 000 000	12	7 500 000
Puffer Fish	375 000		

Biology Background

- Obviously, genome size does not predict the complexity of the organism and also there is no direct correlation between the **genome size** and the **number of chromosomes**.
- It is generally true that it takes more genes to make the species more complex but there are also other factors.
- About 2-3% of the human nuclear genome actually takes part in the production of proteins.
- Even if we ignore the **introns**, apparently 70 to 80% of the genome is unused! This paradox may be due to the existence of highly repetitive DNAs.