C-value paradox

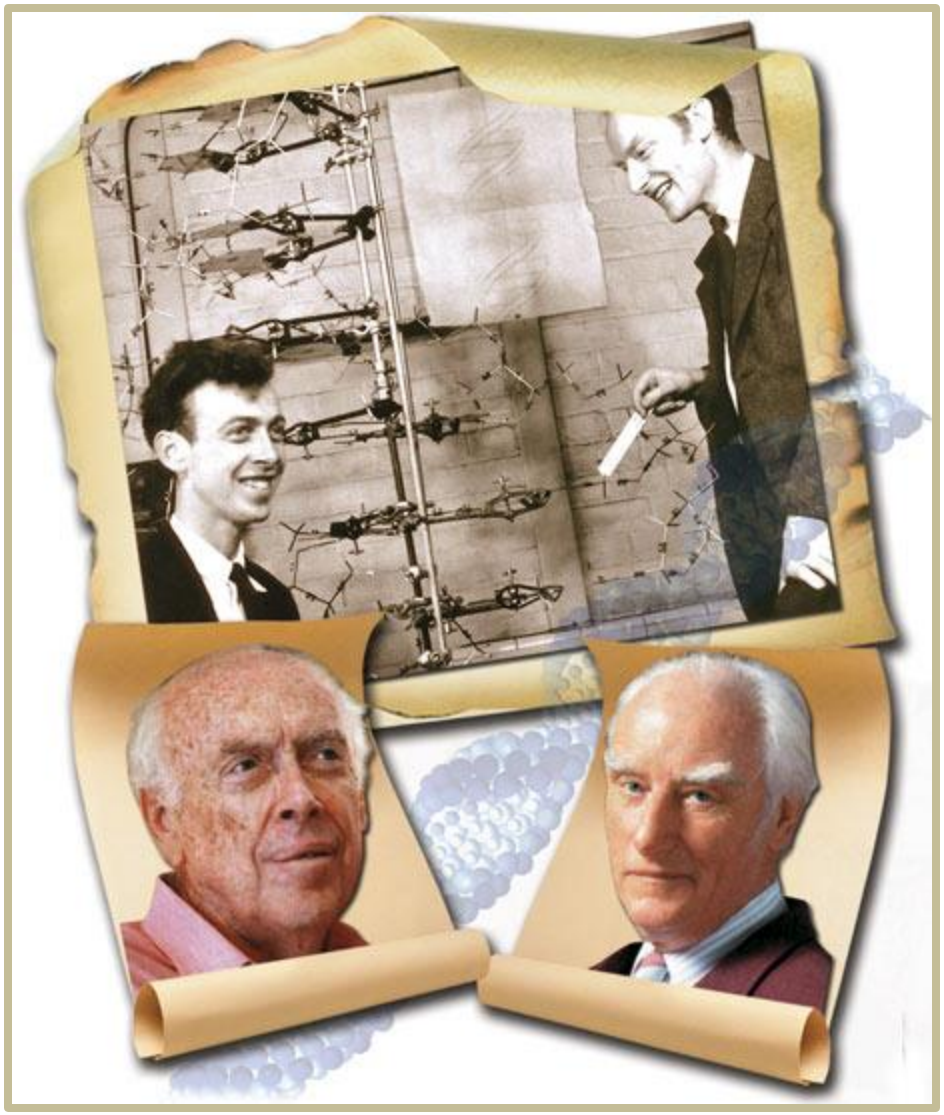
M.Sc. Microbiology, 2nd Semester

MCB 202 : Genetics and Gene regulation

Gr. A: Fundamental Genetics

by

Dr. Suman Kumar Halder



- Franklin is best known for her work on the X-ray diffraction images of DNA which led to the discovery of the DNA double helix.
- According to Francis Crick, her data was key to determining the structure to formulate Crick and Watson's 1953 model regarding the structure of DNA.
- Franklin's images of X-ray diffraction confirming the helical structure of DNA were shown to Watson without her approval or knowledge.
- This image and her accurate interpretation of the data provided valuable insight into the DNA structure, but Franklin's scientific contributions to the discovery of the double helix are often overlooked.
- Unpublished drafts of her papers show that she had independently determined the overall B-form of the DNA helix and the location of the phosphate groups
- However, her work was published third, in the series of three DNA *Nature* articles, led by the paper of Watson and Crick which only hinted at her contribution to their hypothesis.

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Van Ark, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

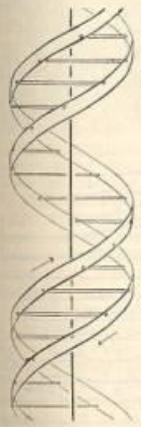
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

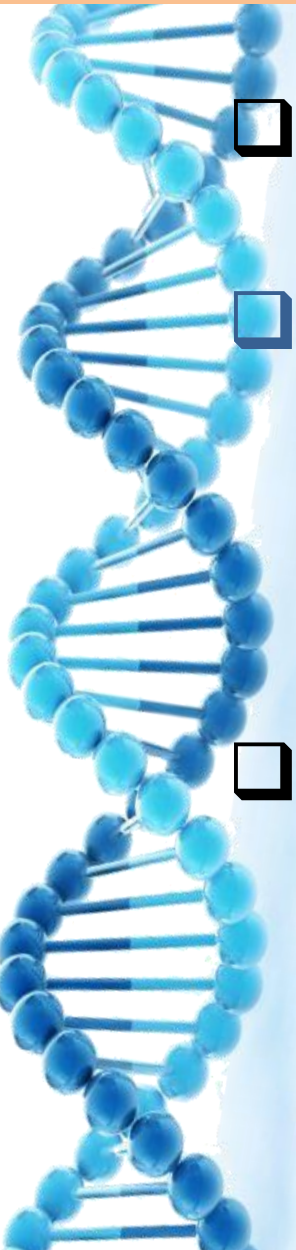
We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

- In 1952, Watson and Crick proposed that DNA is a double helix which is known to have alternative forms.

The Genome

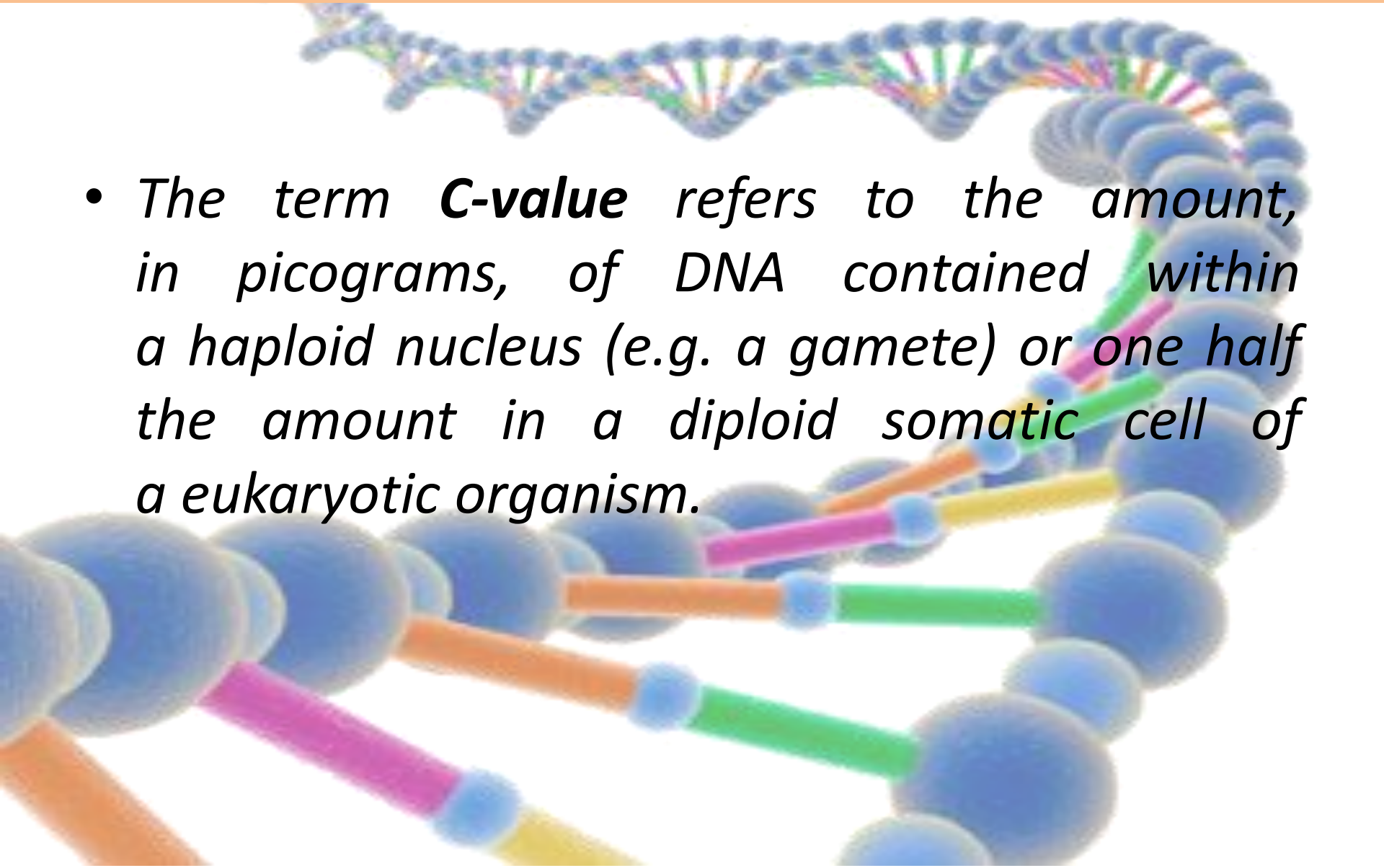
- The organization of total sum of genetic information (or genome) of an organism is in the form of double-stranded DNA.
- In many viruses and prokaryotes, the genome is a single linear or circular molecule. In eukaryotes, the nuclear genome consists of linear chromosomes (usually as a diploid set) and the mitochondrial and chloroplast (in plants) genomes are small circular DNA molecules.
- The DNA is converted from relaxed to supercoiled by DNA specific enzymes.
- In general, genome size increases with organism's complexity.

DNA per Cell

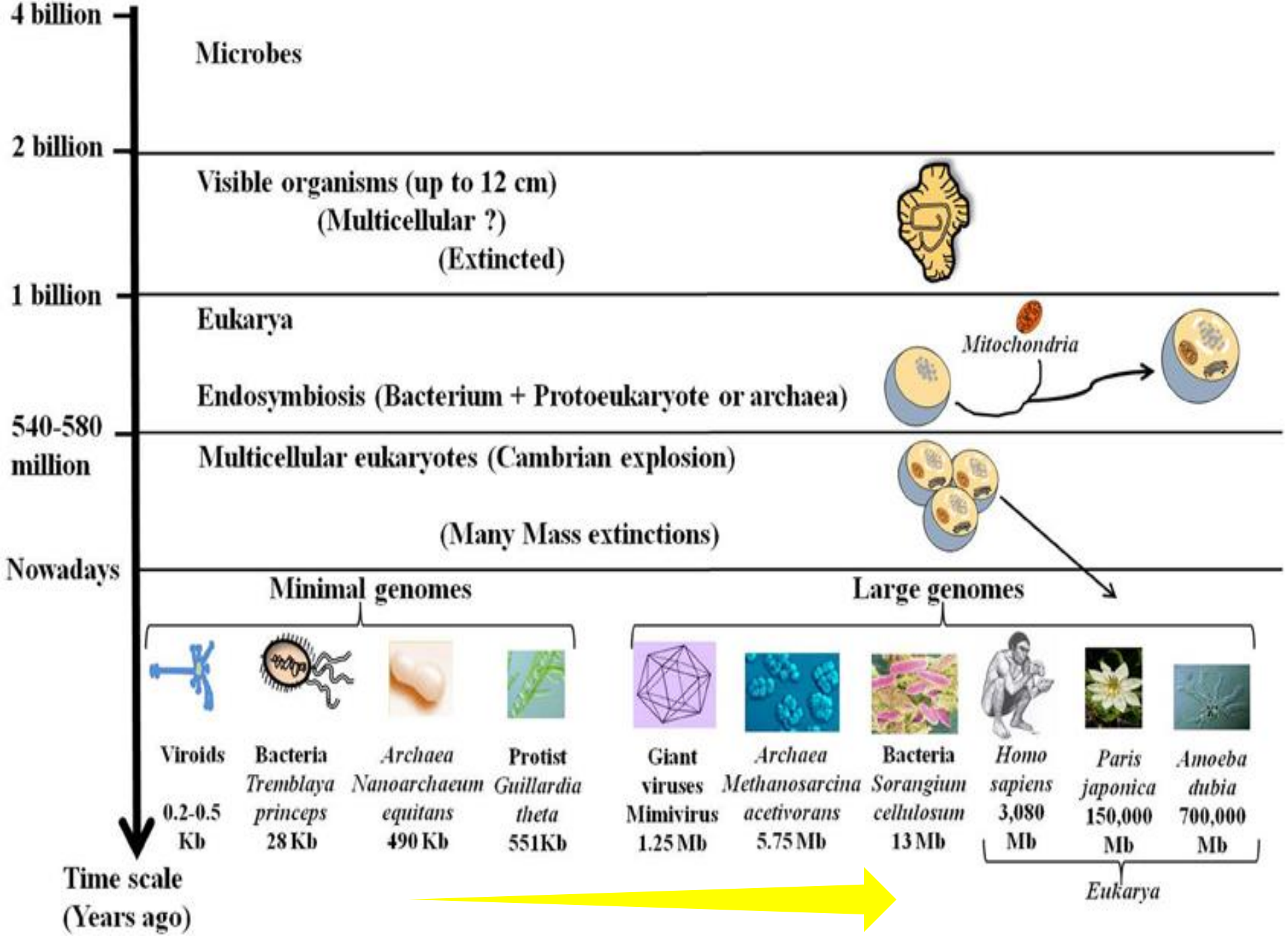


- Quantity of DNA in an organism per cell, in all cells, is always constant for a given species.
- All the organisms on this planet have its own genome whose size varies from one species to the other and no two species have the same amount of genome nor the same genomic value or character.
- At this point of time, nobody has found two different species, however close in phylogeny they may be, have the same size of the genome, even one finds such situations, and some of their sequence will be different.

C-Value



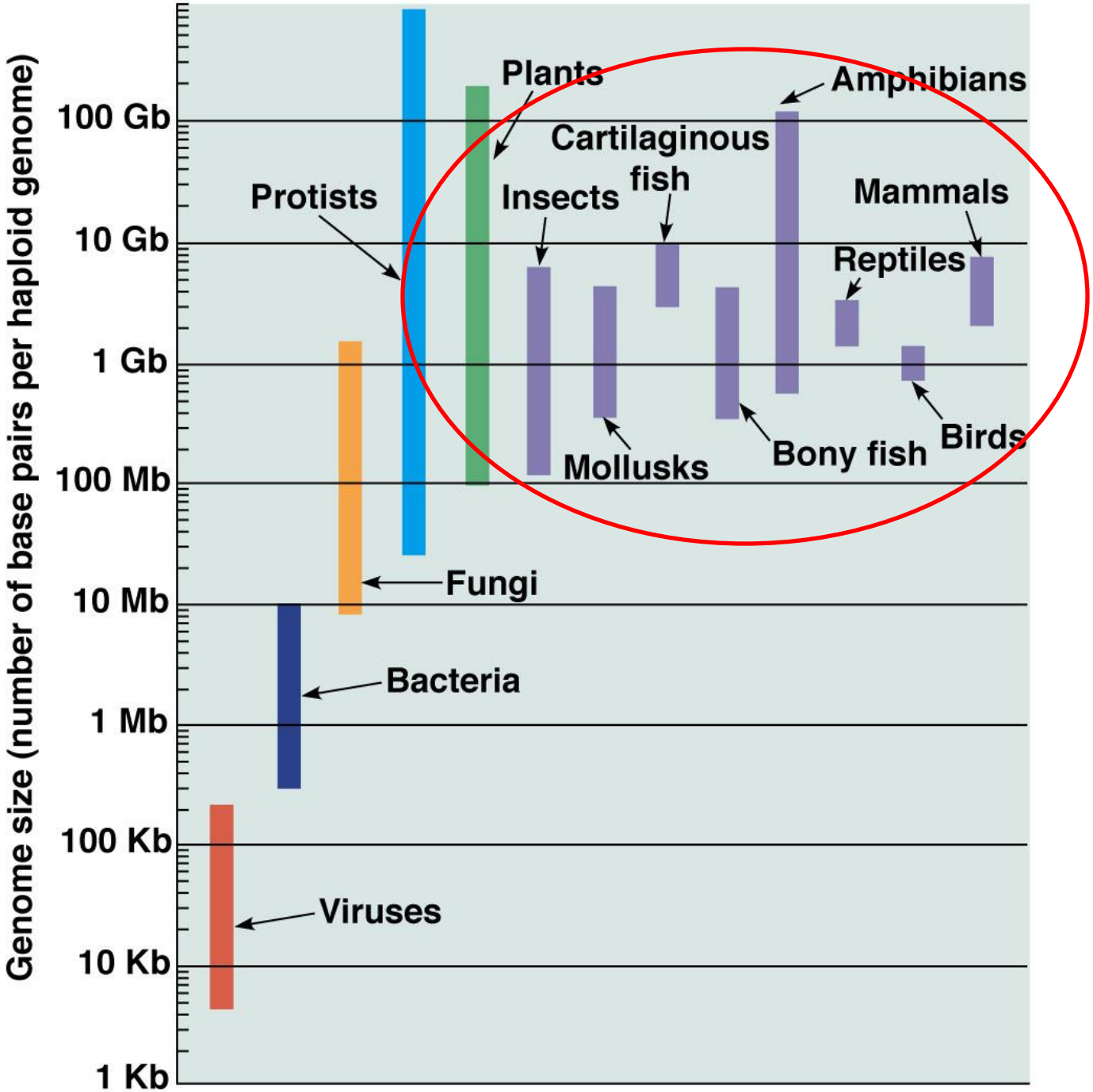
- *The term **C-value** refers to the amount, in picograms, of DNA contained within a haploid nucleus (e.g. a gamete) or one half the amount in a diploid somatic cell of a eukaryotic organism.*

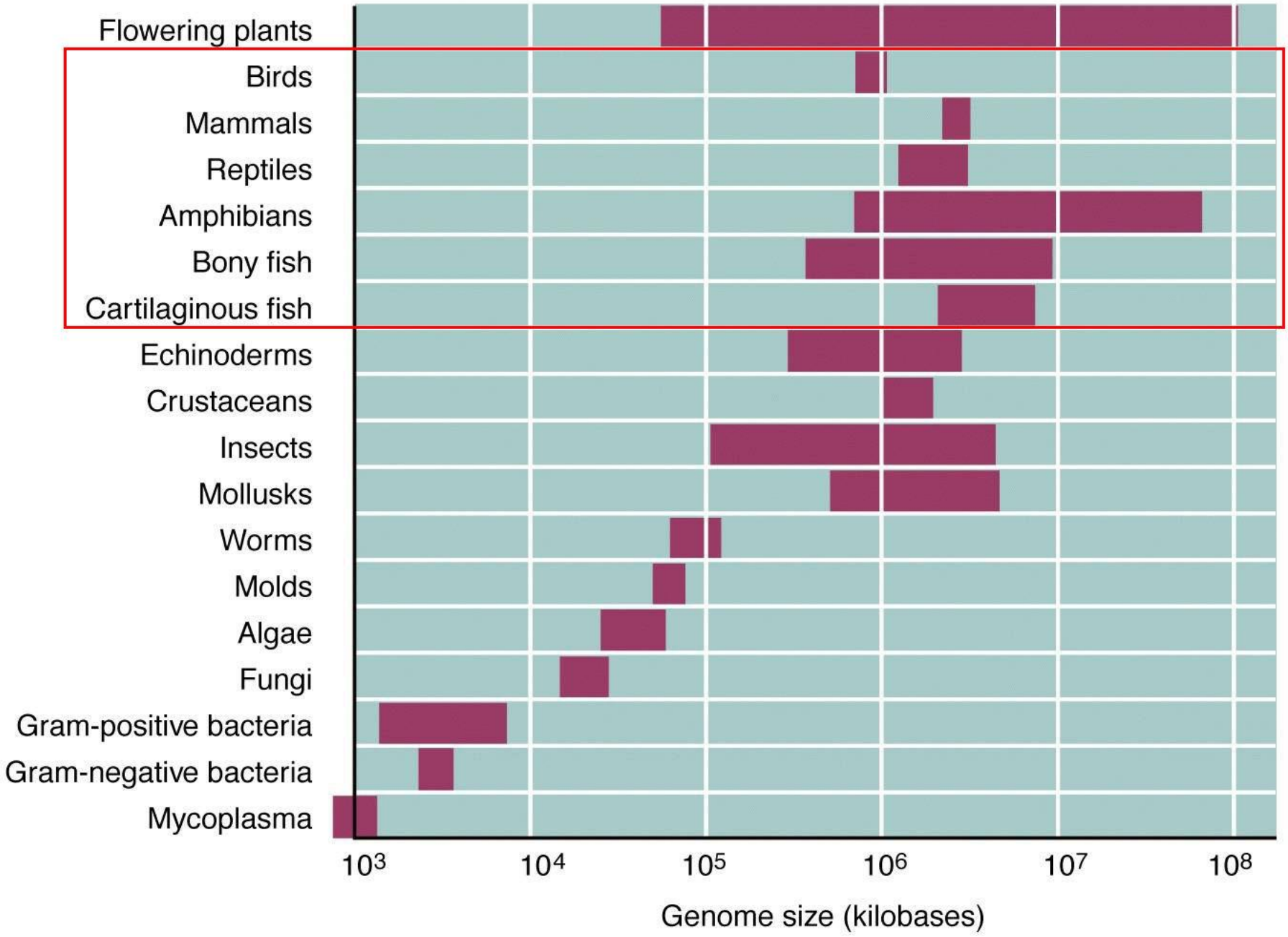

4 billion
Microbes
2 billion
Visible organisms (up to 12 cm)
(Multicellular ?)
(Extincted)
1 billion
Eukarya
Endosymbiosis (Bacterium + Protoeukaryote or archaea)
Mitochondria
540-580 million
Multicellular eukaryotes (Cambrian explosion)
(Many Mass extinctions)
Nowadays
Minimal genomes
Large genomes
Viroids
0.2-0.5 Kb
Bacteria
Tremblaya princeps
28 Kb
Archaea
Nanoarchaeum equitans
490 Kb
Protist
Guillardia theta
551Kb
Giant viruses
Mimivirus
1.25 Mb
Archaea
Methanosarcina acetivorans
5.75 Mb
Bacteria
Sorangium cellulosum
13 Mb
Homo sapiens
3,080 Mb
Paris japonica
150,000 Mb
Amoeba dubia
700,000 Mb
Eukarya
Time scale
(Years ago)

- This chart shows the range of **C-value** [genome size, measured as number of **Kbp of DNA**] for a variety of organisms. So-called "simple" prokaryotic organisms in general have less **DNA** per genome than do more "complex," eukaryotic organisms, such as plants and animals.
- **Genome size** (**C-value**) can be measured as either the **bp** or the **mass of DNA** in a *haploid* chromosome set, where **10^6 bp = 1 Mbp = 10^{-3} pg DNA**.
- The actual number of base pairs or mass of **DNA** in the cell of a *diploid* organism is therefore *twice* that shown here. For example, the human euchromatic genome is **3.2 x 10^9 bp**.

C-Value paradox

- **C-Value Paradox** refers to the observation that genome size does not uniformly increase with respect to perceived complexity of organisms, for example vertebrate with respect to invertebrate animals, or "lower" versus "higher" vertebrate animals. For examples that some Amphibians have more than 10-fold more **DNA** than do Mammals, including humans.




Flowering plants
Birds
Mammals
Reptiles
Amphibians
Bony fish
Cartilaginous fish
Echinoderms
Crustaceans
Insects
Mollusks
Worms
Molds
Algae
Fungi
Gram-positive bacteria
Gram-negative bacteria
Mycoplasma
10^3
10^4
10^5
10^6
10^7
10^8
Genome size (kilobases)

- The **C-value enigma** or **C-value paradox** is the complex puzzle surrounding the extensive variation in nuclear genome size among eukaryotic species. At the center of the C-value enigma is the observation that genome size does not correlate with organism complexity; for example, some single-celled protists have genomes much larger than that of humans

- The term "**C-value enigma**" represents an update of the more common but outdated term "C-value paradox“ (Thomas 1971), being ultimately derived from the term "C-value" (Swift 1950) in reference to haploid nuclear DNA contents. The term was coined by Canadian biologist Dr. T. Ryan Gregory of the University of Guelph in 2000/2001.

- There is in fact no "paradox." Evolution does not proceed in a linear manner, nor is there a linear succession of organisms from "lower" to "higher."
- Despite differences in **DNA** content, the number of genes in any vertebrate genome is roughly similar. Also, plant and amphibian genomes in particular are frequently **polyploid**, in which the chromosome number undergoes doubling to two-, four, or eight-fold without a radical change to the form of the organisms.

- However the discovery of a large amount of non-coding DNA lead to the concept of C-DNA value or C-Value paradox and variation is surprisingly so vast it is called C-DNA value paradox**.**
- The paradox or the enigma is between the C-value and the gene numbers. Elucidation of noncoding DNA and noncoding but functional RNA can resolve this.

- The discovery of non-coding DNA in the early 1970s resolved the main question of the C-value paradox: genome size does not reflect gene number in eukaryotes since most of their DNA is non-coding and therefore does not consist of genes.
- The human genome, for example, comprises less than 2% protein-coding regions, with the remainder being various types of non-coding DNA (especially transposable elements)