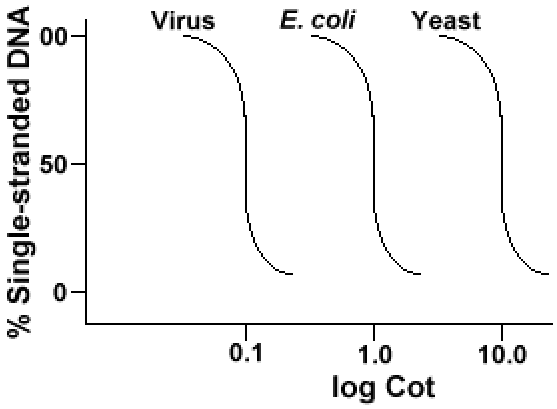
I. Genome size, composition and complexity

The total amount of DNA in a genome (C) can vary enormously, and is not always related to the number of genes or complexity of organism (**C value paradox**).

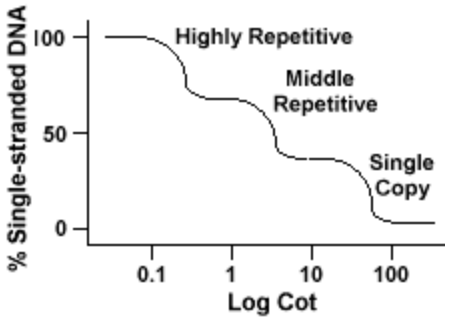
Species	DNA bp (C)	Genes	% Coding
phage λ	48,502	67	84
E. coli	4,639,221	4289	85
S. cerevisiae	14,213,386	6241	70
D. melanogaster	180,000,000	18,000	5
Grasshopper	10,000,000,000		
H. sapiens	3,000,000,000	35,000	2-5
Newt	19,000,000,000		~1
Lungfish	140,000,000,000		<1

Reassociation kinetics can be used to characterize DNA complexity: DNA is sheared into pieces of a few hundred bp, heated to denature into single strands, then allowed to renature during cooling. **C_0t** (usually pronounced like the word cot) curve = % single-stranded as a function of initial concentration of DNA (C_0) multiplied by time (t).



The rate of renaturation is related to the sequence complexity. Let's say we start with 1 ng (nanogram = 10^{-9} gram = 10^{12} bp) of DNA in 1 ml. For a virus with a genome size of 100,000 bp, that's 10^7 genomes, so each piece is present 10^7 times. But for a human with a genome size of 3 x 10^9 bp, that's 300 genomes, so each piece is present only 300 times in our 1 ng. Thus, each fragment of DNA will find a complementary strand much more quickly if we have 1 ng of viral DNA than if we have 1 ng of human DNA.

A C_0t cure of mammalian genomic DNA typically has three humps, corresponding to sequences that are present many, many times in the genome (**highly repetitive**), many times (**middle repetitive**), and once (**single copy** or unique).



Most genes are **single copy** – there is one sequence for that gene in each haploid genome.

Middle repetive sequences include transposable elements and genes clusters;
rRNA and histone genes occur in 100s – 1000s of copies;
tRNAs are represented multiple times in the genome.
minisatellites = sequences 15-100 bp long repeated tandemly many times; the number of repeats varies between individuals, leading to the name **variable number tandem repeats** (**VNTR**); basis of DNA fingerprinting in humans
microsatellites = repeats of mono-, di-, or tri-nucleotide sequences

Highly repetitive sequences are primarily **satellite DNA**, short sequences tandemly repeated many thousands of times (*e.g.*, **AAGATAAGATAAGATAAGATAAGATAAGATAAGA**) most satellite DNA is located around centromeres.

Species	% high	% middle	% unique
E. coli	0	2	98
S. cerevisiae	0	10	80
D. melanogaster	25	15	60
H. sapiens	8	45	47
maize	40	20	40

II. Physical structure of chromosomes

chromatin = complex of DNA and protein in the nucleus; allows compaction of DNA, additional control of gene expression;

nucleosome = DNA wrapped around a histone octamer, which has two copies of each of two heterodimers (H2A and H2B) and (H3 and H4);
146 bp around the octamer (not 196 as the text says on p. 541);
linker DNA associated with histone H1
30 nm solenoid = spiral of nucleosomes, 6 per turn.
chromatin loops = about 300 nm width, attached at base by **scaffold proteins**.

During mitosis, the chromatin is **condensed**; most interphase chromatin is **decondensed**.
Some regions remain highly condensed throughout the cell cycle = **heterochromatin**;
composed primarily of satellite DNA, many transposable elements, rare genes;
centromeres and telomeres of higher eukaryotes are embedded in heterochromatin;
Drosophila and mammalian *Y* chromosome is entirely heterochromatic.
Most genes are found in **euchromatin**; if moved into heterochromatin by chromosome rearrangement, they are not expressed = **position effects**.

Differences in compaction and composition along a chromosome results in different, reproducible properties when stained with DNA dyes = **chromosome banding**; allows physical mapping of genes along chromosomes.

III. Genetic structure of chromosomes

You don't need to know these numbers or descriptions. This is just to give you an idea of what chromosomes look like in various organisms.

E. coli: 4.6 Mb, circular
transposable elements = insertion sequences (IS) and transposons
genes packed closely together, into operons, no introns

S. cerevisiae: 13 Mb, 16 chromosomes (0.23 Mb – 2 Mb each)
transposable elements = *Ty* elements (viral retrotransposons)
very few introns
centromere = about 125 bp of specific sequence
telomeres = $(TG_{1\text{-}3})$ x 100-300

D. melanogaster: 140 Mb, 4 chromosomes (5 – 69 Mb each)
transposable elements: >50 kinds, each in 0-100 copies per genome;
throughout DNA, but concentrated near euchromatin-heterochromatin boundary.
many introns (sometimes even genes within introns of other genes)
large variability in gene density;
centromeres = 200-300 kb (no specific sequence, probably specific chromatin structure)
within Mb of satellite sequences;
telomeres = arrays of retrotransposons

H. sapiens: 2900 Mb, 23 chromosomes (28 Mb (*Y*) to 363 Mb (*1*))
many interspersed repeats, especially LINES, SINES, and retrotransposons
repetitive DNA makes up 45% of the draft sequence – over 1,000,000 *Alu* sequences
DNA transposable elements extinct for 50 Myr, only "fossils" remain
many introns, some very large
large variability in gene density;
centromeres = 200-300 kb (no specific sequence, probably specific chromatin structure)
within Mb of satellite sequences;
telomeres = TTAGGG x 300-1500