Department of BSBE
Indian Institute Of Technology Guwahati

Genome organisation
Genome size and complexcity

Sanjukta Patra

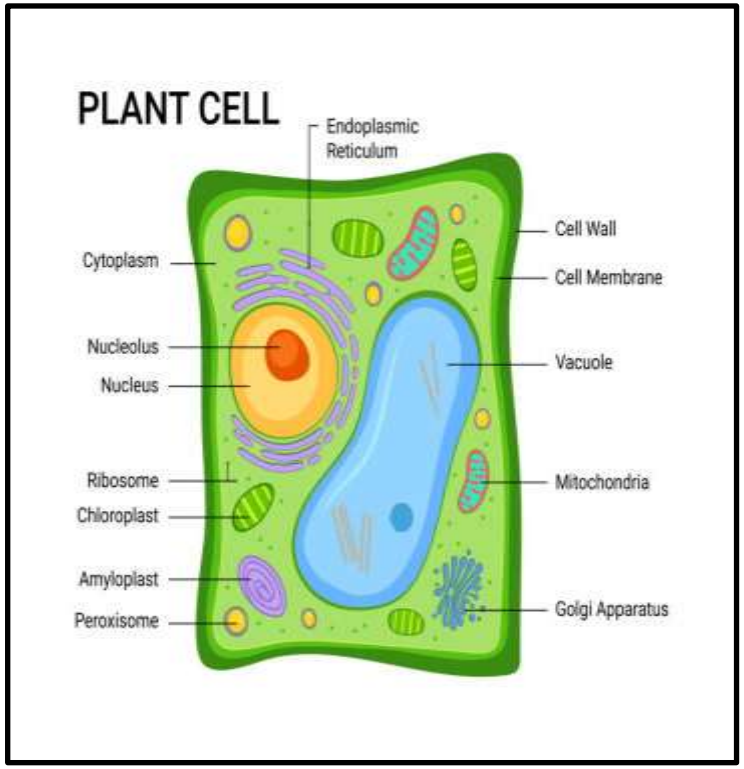
BT 207 – Genetic Engineering

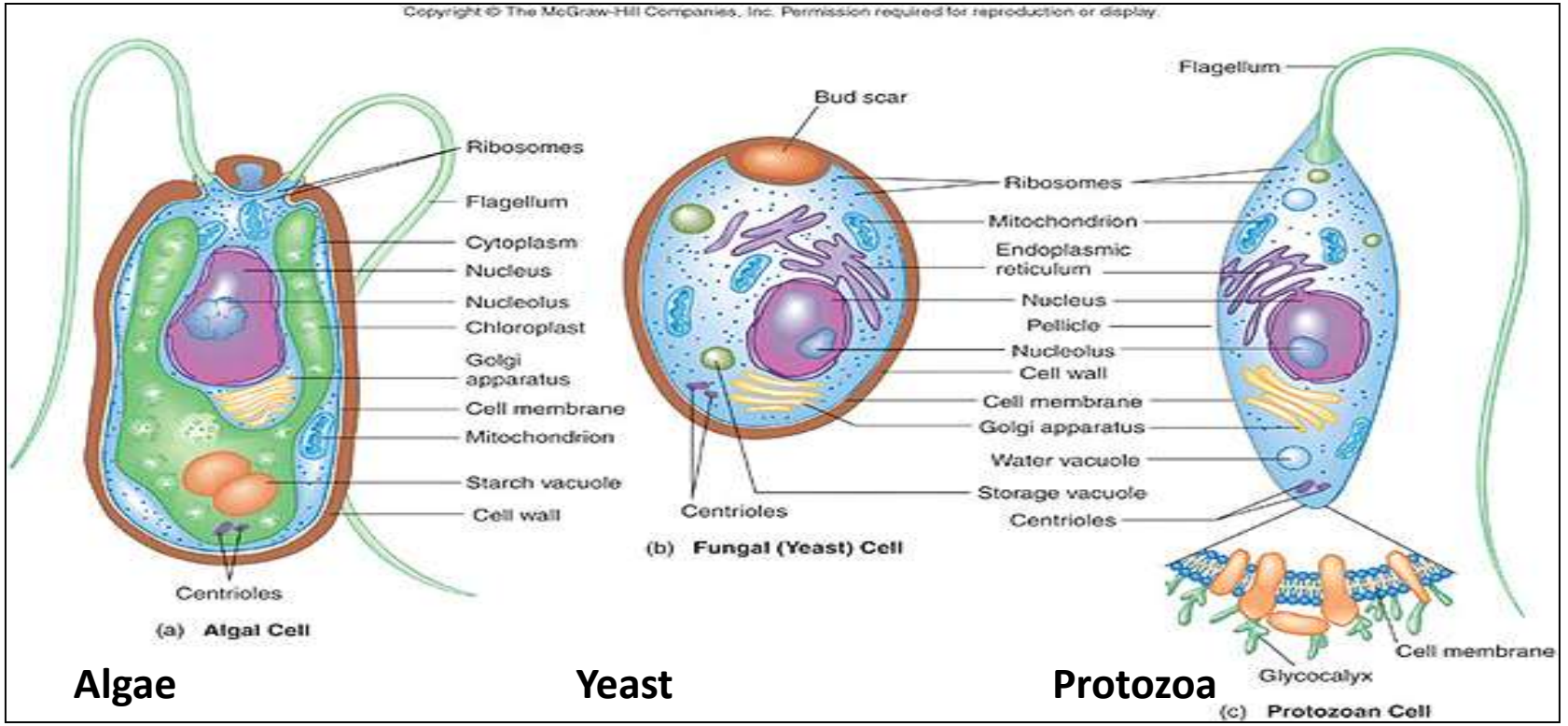
Jan -May 2023

- What is genome?
- Where is the genome present in a cell?
- What are the Genome types?
- The genome of different organisms.

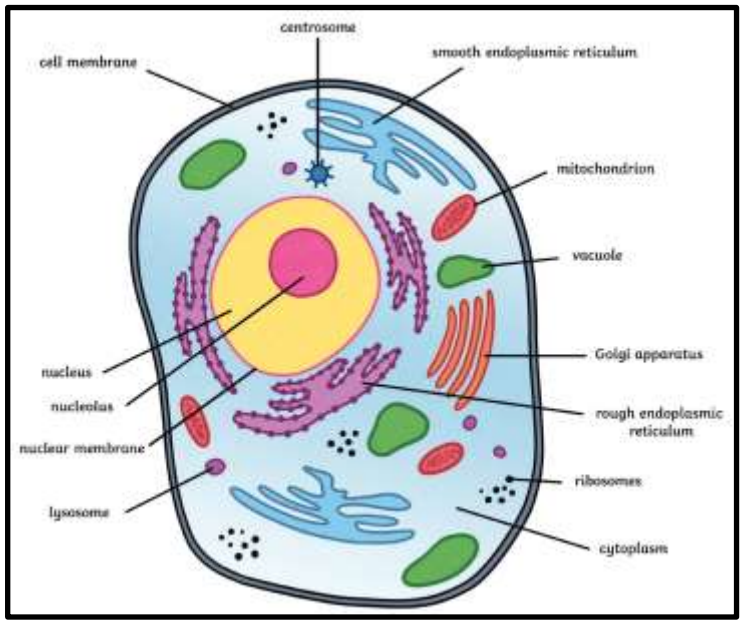
Genomes

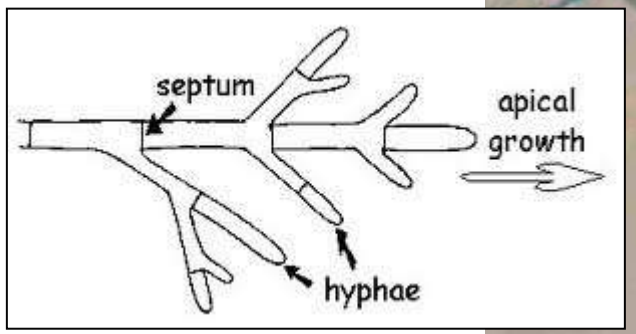
- The word "genome," coined by German botanist Hans Winkler in 1920, was derived simply by combining *gene* and the final syllable of *chromosome.*
- The genome includes both the genes and non-coding sequences of DNA.
- An organism's **genome** is defined as the entire collection of genes and all other functional and non-functional DNA sequence in a haploid set of chromosomes.
- It includes structural genes, regulatory genes and non-functional nucleotide sequence.
- The genome is the ultimate source of information about an organism.

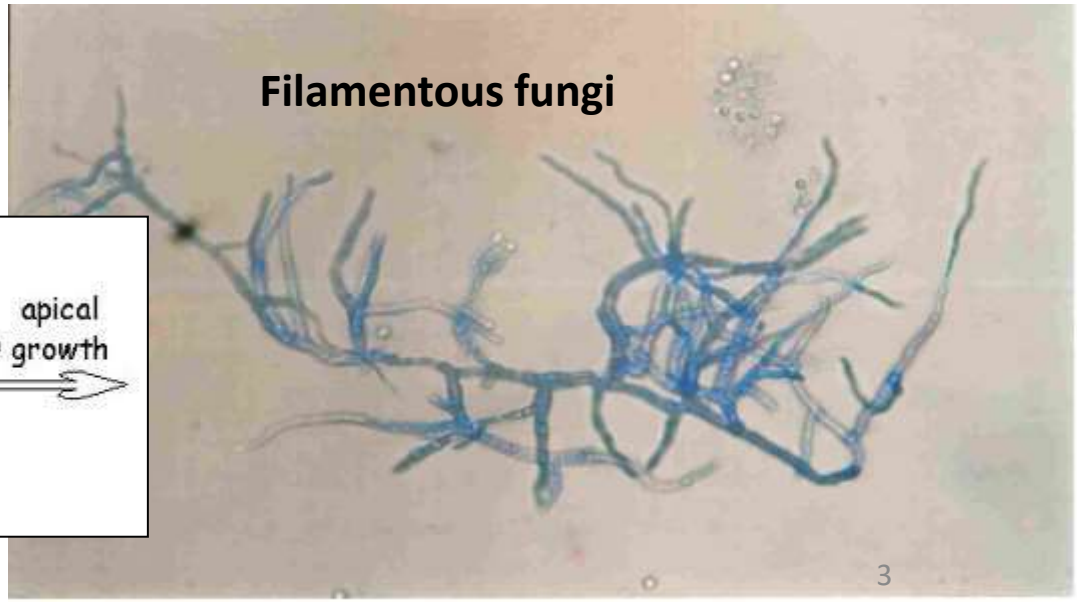




Algae **Yeast** **Protozoa**







"Genes" are units of genetic information present on the DNA in the chromosomes and chromatin.

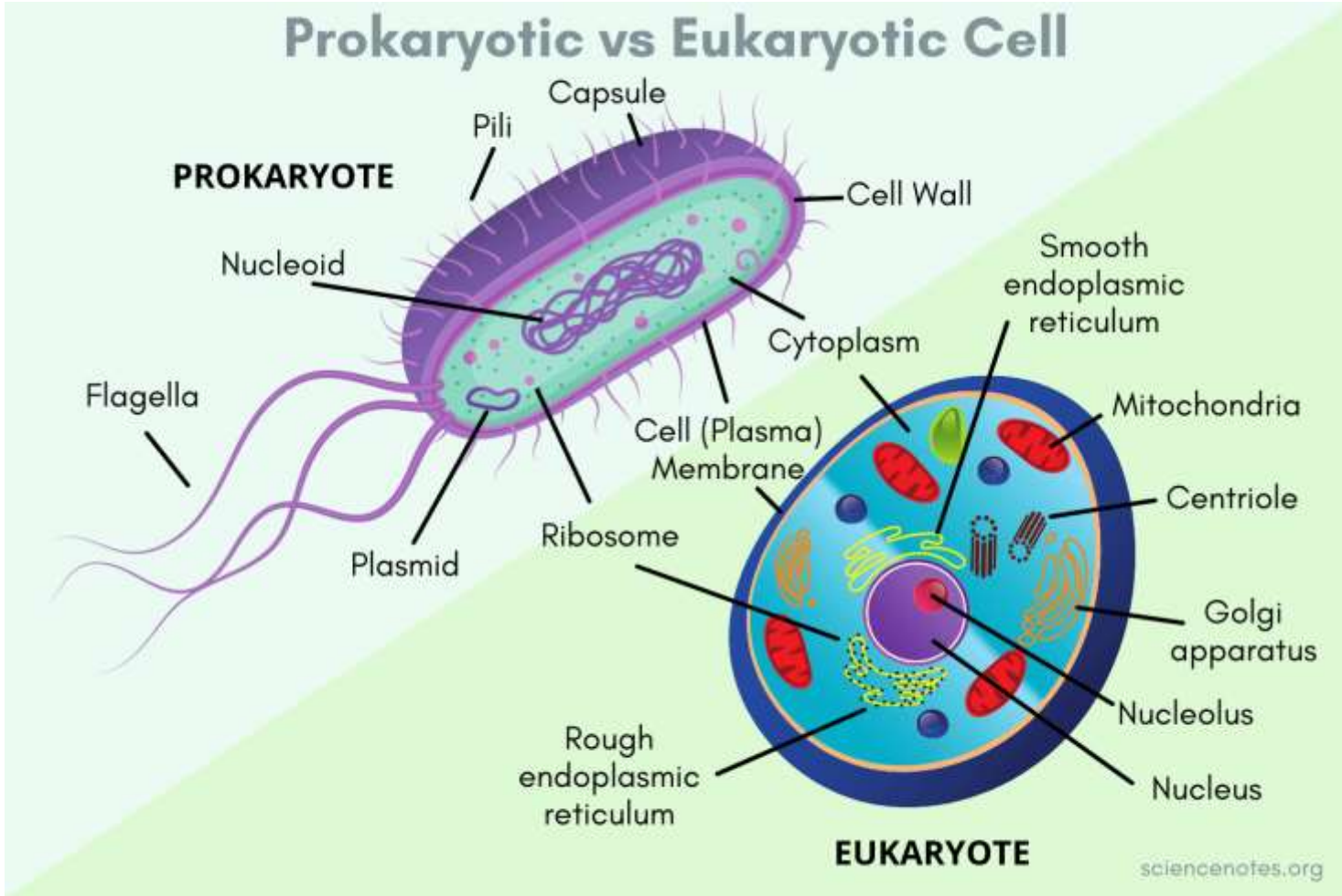
"Genome" is the entirety of an organism's hereditary organization. It is encoded either in DNA, or for many types of viruses, in RNA.

Can you localise genome in a cell?

Hoechst and SYTO® series dyes include a number of nuclear stains for live cells.

Phase contrast microscope – Live cell Nucleus

Fluorescent microscope – Fluorescent dyes

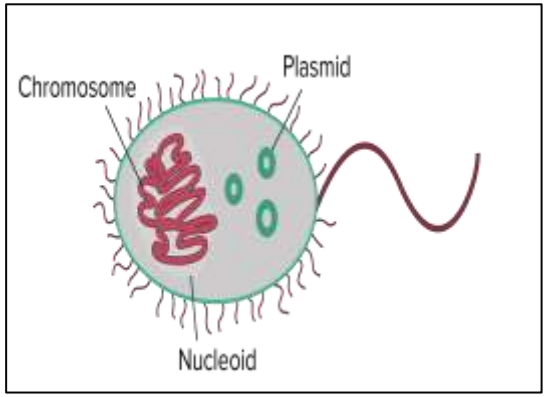

Prokaryotic vs Eukaryotic Cell
Capsule
Pili
PROKARYOTE
Cell Wall
Nucleoid
Smooth endoplasmic reticulum
Cytoplasm
Flagella
Mitochondria
Cell (Plasma) Membrane
Centriole
Ribosome
Plasmid
Golgi apparatus
Nucleolus
Rough endoplasmic reticulum
Nucleus
EUKARYOTE
sciencenotes.org

Difference between prokaryotes and eukaryotes

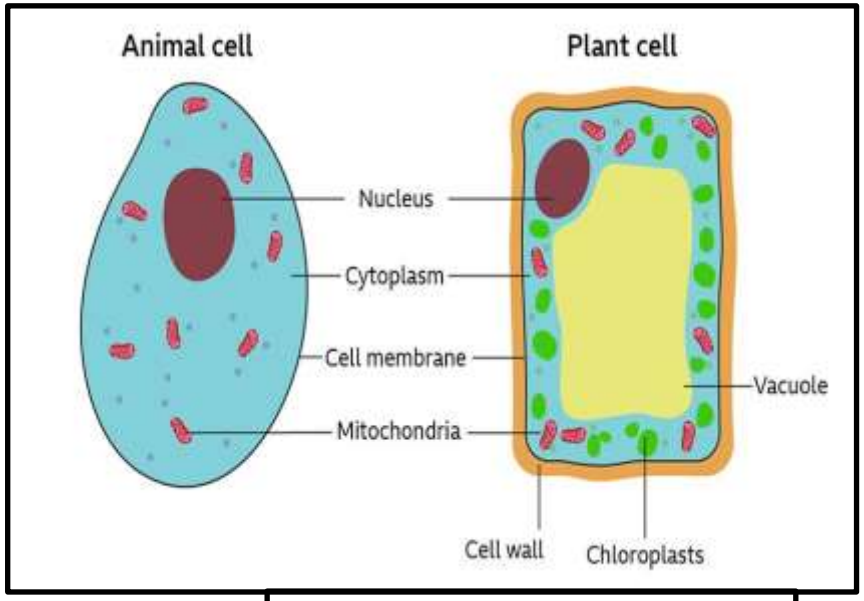
Characteristic	Prokaryotes	Eukaryotes
Size of cell	Typically 0.2-2.0 m icro m in diameter	Typically 10-100 micro m in diameter
Nucleus	No nuclear membrane or nucleoli (nucleoid)	True nucleus, consisting of nuclear membrane & nucleoli
Membrane-enclosed organelles	Absent	Present; examples include lysosomes, Golgi complex, endoplasmic reticulum, mitochondria & chloroplasts
Flagella	Consist of two protein building blocks	Complex; consist of multiple microtubules
Glycocalyx	Present as a capsule or slime layer	Present in some cells that lack a cell wall
Cell wall	Usually present; chemically complex (typical bacterial cell wall includes peptidoglycan)	When present, chemically simple
Plasma membrane	No carbohydrates and generally lacks sterols	Sterols and carbohydrates that serve as receptors present
Cytoplasm	No cytosketeton or cytoplasmic streaming	Cytoskeleton; cytoplasmic streaming
Ribosomes	Smaller size (70S)	Larger size (80S); smaller size (70S) in organelles
Chromosome (DNA) arrangement	Single circular chromosome; lacks histones	Multiple linear chromosomes with histones
Cell division	Binary fission	Mitosis
Sexual reproduction	No meiosis; transfer of DNA fragments only (conjugation)	Involves Meiosis

Genomes

- Genome - describes all of the genetic material of the cell.
- Genome - entire sequence of nucleotides in the DNA that is in all of the chromosomes of a cell.
- Nuclear genome – Genome of nucleus
- Plasmids – Bacteria and yeasts
- Eukaryotic cells also have organelles like mitochondria and chloroplasts that have their own DNA.
- These are referred to as the **mitochondrial or chloroplast genomes** to distinguish them from the nuclear genome.



Prokaryotic cell



Eukaryotic cell

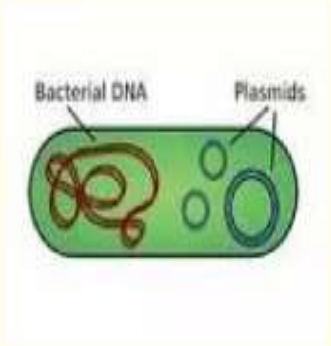
Yeasts are Eukaryotic cells with Nucleus, Mitochondria and plasmids

Why genome needs to be organised?

Prokaryotic genome organisation

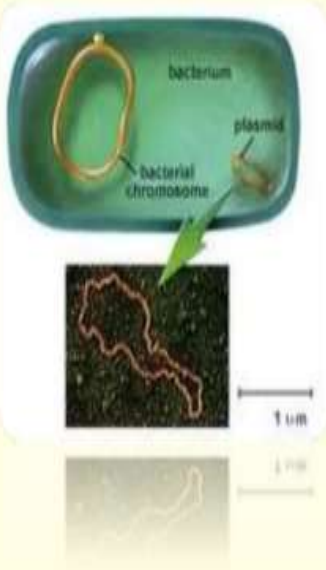
Eukaryotic genome organisation

- Prokaryotic genomes generally contain one large circular piece of DNA referred to as a "chromosome" (not a true chromosome in the eukaryotic sense).
- Some bacteria have linear "chromosomes".
- Many bacteria have small circular DNA structures called plasmids which can be swapped between neighbors and across bacterial species.

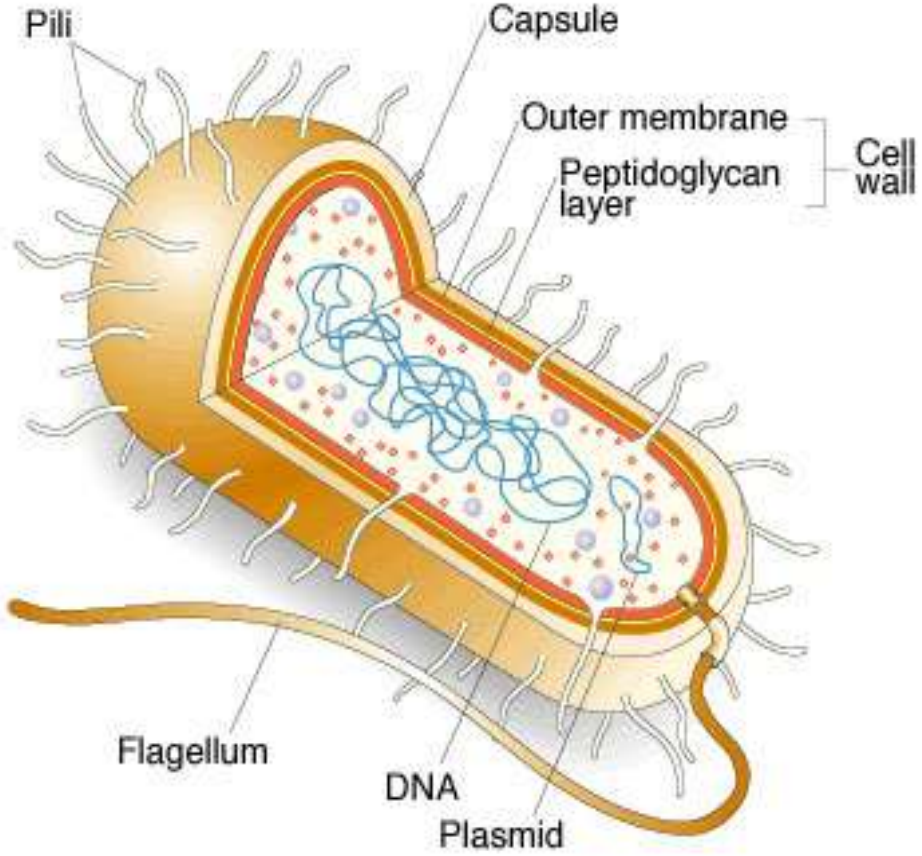


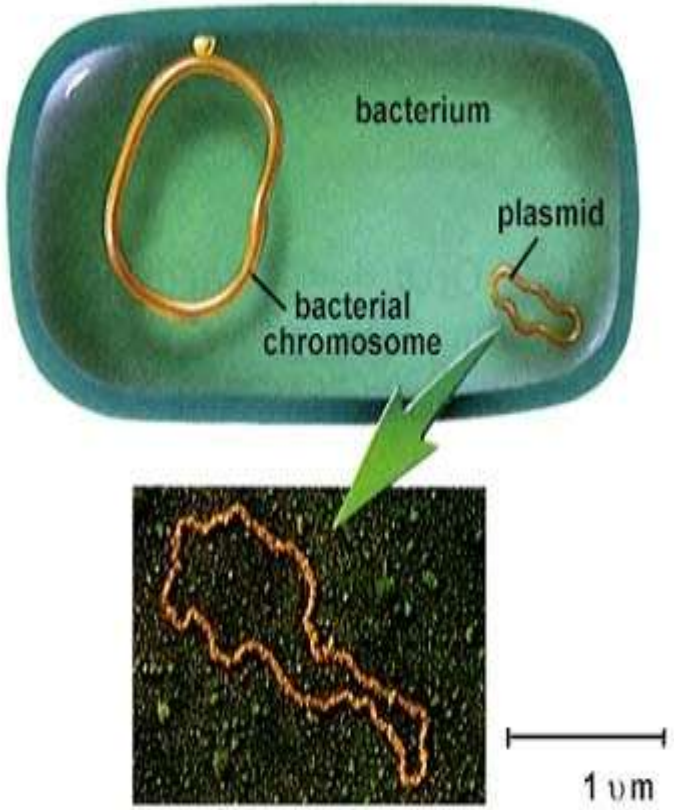
Plasmid

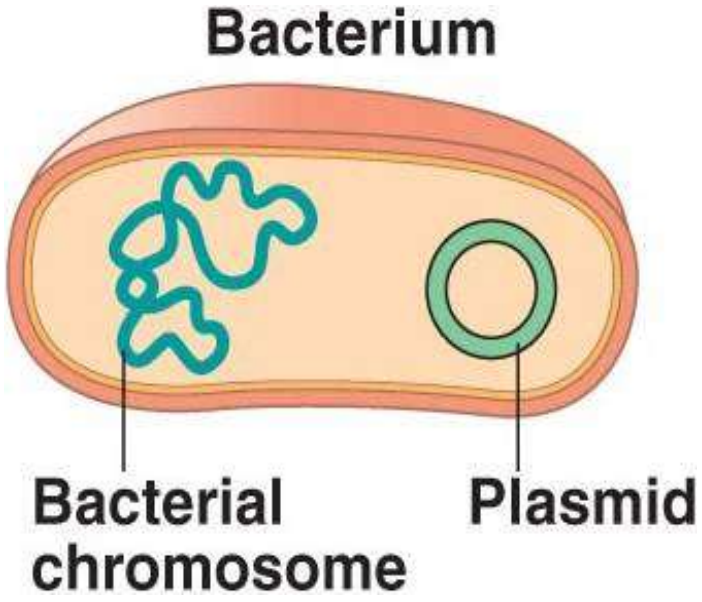
- The term *plasmid* was first introduced by the American molecular biologist Joshua Lederberg in 1952.
- A plasmid is separate from, and can replicate independently of, the chromosomal DNA.
- Plasmid size varies from 1 to over 1,000 (kbp).

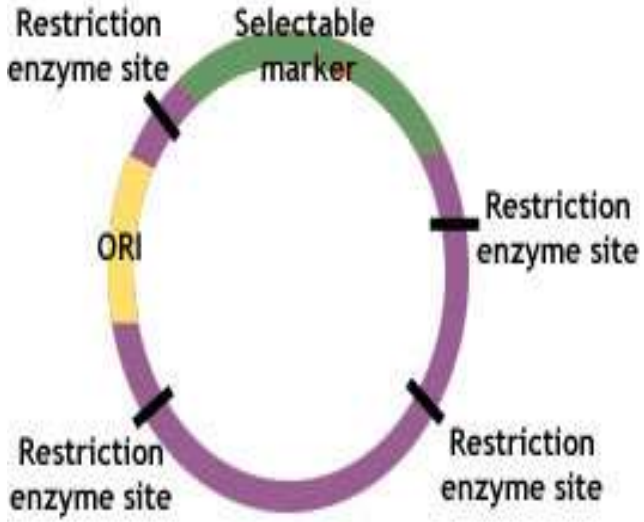


Bacterial genetic material





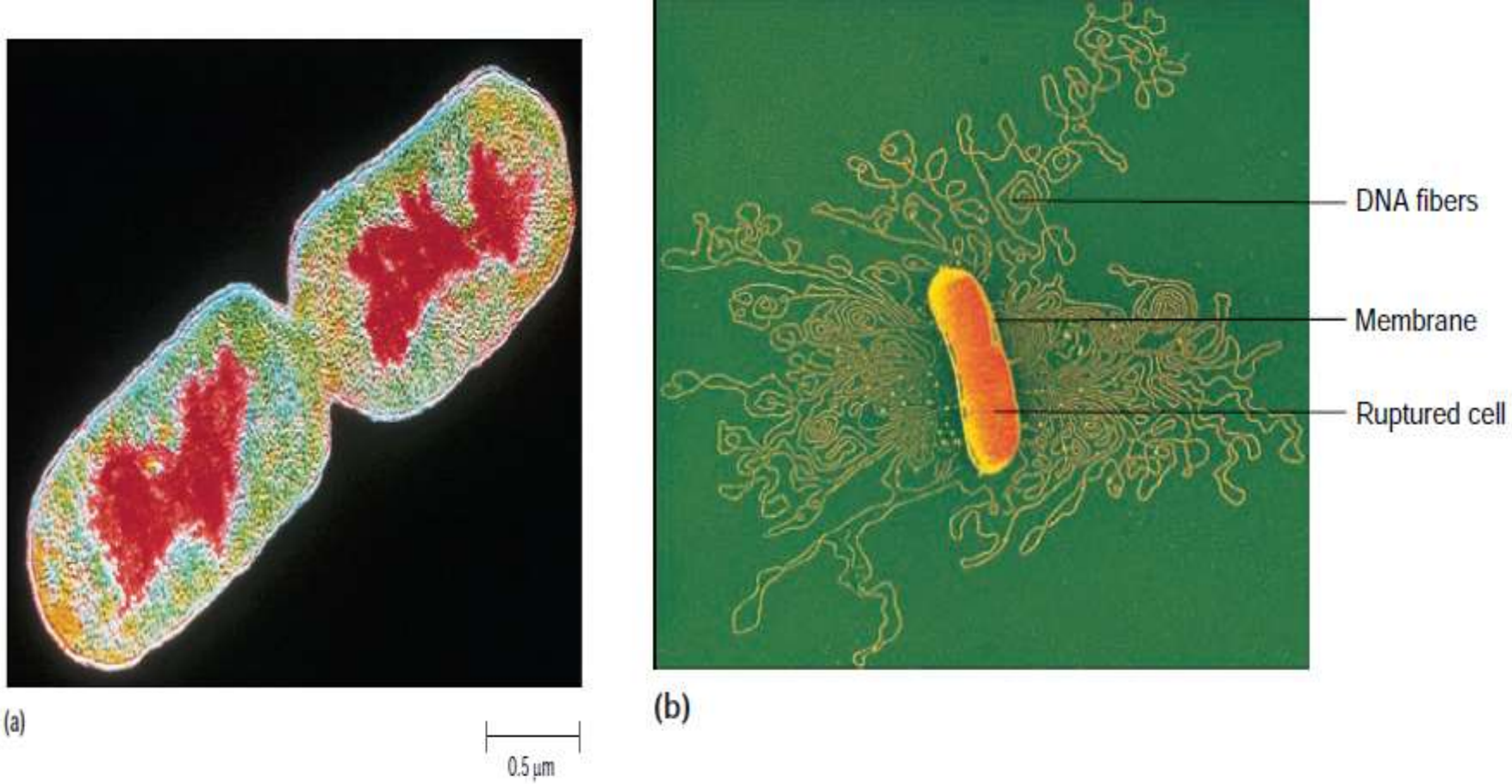




- Genomic DNA
- Plasmids

Most bacterial chromosomes contain a circular DNA molecule - there are no free ends to the DNA

The circularity of the bacterial chromosome was demonstrated by electron microscopy in both Gram negative bacteria (*Escherichia coli*) and Gram positive bacteria (*Bacillus subtilis*)



Procaryotic Nucleoids and Chromosomes.
Procaryotic chromosomes are located in the nucleoid, an area in the cytoplasm.
(a) A color-enhanced transmission electron micrograph of a thin section of a dividing *E. coli cell. The red areas are* the nucleoids present in the two daughter cells. **(b) Chromosome** released from a gently lysed *E. coli cell. Note how tightly packaged* the DNA must be inside the cell.

The Nucleoid

- Prokaryotes lack a membrane-bound nucleus. The prokaryotic

chromosome is located in an irregularly shaped region called the

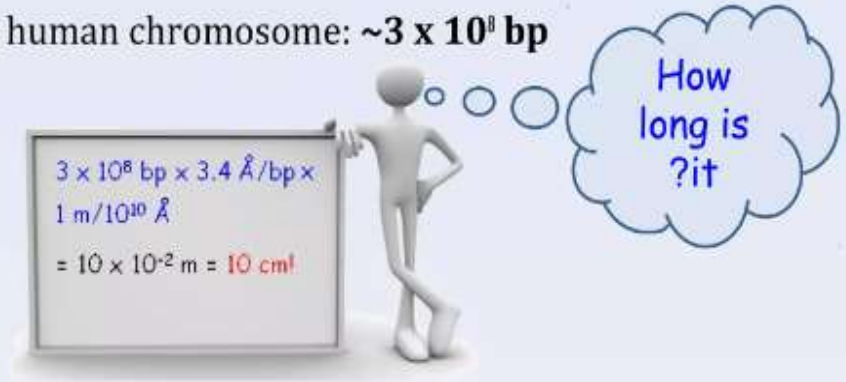
nucleoid

The nucleoid has a fibrous appearance in electron micrographs; the fibers are probably DNA

- Usually prokaryotes contain a single circle of double-stranded **deoxyribonucleic acid (DNA)** but some have a linear DNA
- Chemical analysis of purified nucleoids reveal that they are composed of about 60% DNA, 30% RNA, and 10% protein by weight
- How the genome is organised?

A severe problem of packaging

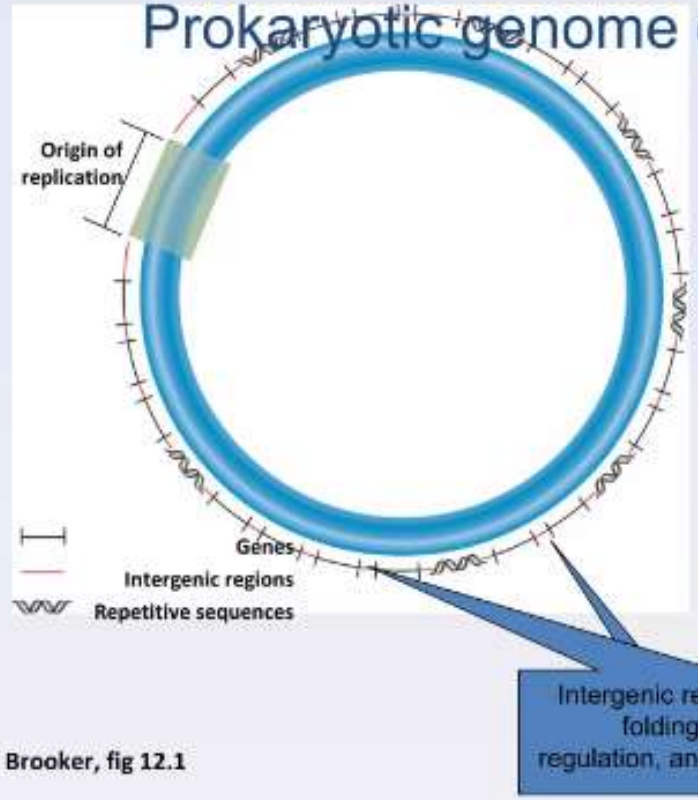
1. Largest human chromosome: **~3 x 10^8 bp**



2. A typical cell = 10 μm = 10 x 10^{-6} m
3. Therefore the DNA must be compacted ~10^4-fold.

This is like fitting an 11-mile-long string into a 6-foot box !

Prokaryotic genome characteristics



Most, but not all, bacterial species contain circular chromosomal DNA

A typical chromosome is a few million base pairs in length

Most bacterial species contain a single type of chromosome, but it may be present in multiple copies

A few thousand different genes are interspersed throughout the chromosome

Brooker, fig 12.1

- Most prokaryotes do not have histones (with the exception of those species in the domain Archaea).
- Thus, one way prokaryotes compress their DNA into smaller spaces is through supercoiling.
- Like twisting a rubber band so that it forms tiny coils. Now when twisted again it further, so that the original coils fold over one another and form a condensed ball.
- When this type of twisting happens to a bacterial genome, it is known as supercoiling.
- Genomes can be negatively supercoiled, meaning that the DNA is twisted in the opposite direction of the double helix, or positively supercoiled, meaning that the DNA is twisted in the same direction as the double helix.
- **Most bacterial genomes are negatively supercoiled during normal growth.**

Proteins Involved in Supercoiling

During the 1980s and 1990s, researchers discovered that multiple proteins act together to fold and condense prokaryotic DNA.

One protein called HU (Heat unstable), which is the most abundant protein in the nucleoid, works with an enzyme called topoisomerase I **to bind DNA and introduce sharp bends in the chromosome, generating the tension necessary for negative supercoiling.**

Recent studies have also shown that other proteins, including **integration host factor (IHF), can bind to specific sequences within the genome and introduce additional bends.**

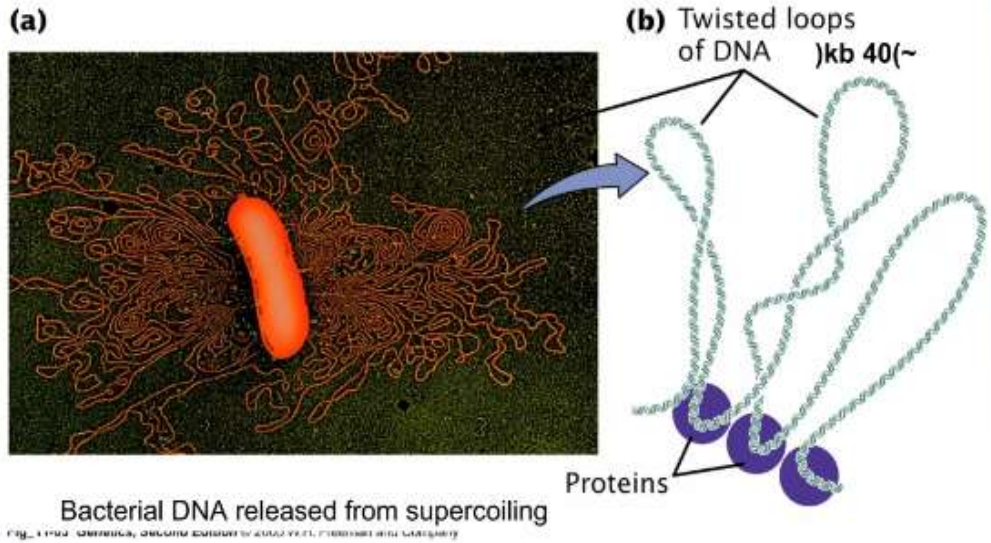
The folded DNA is then organized into a variety of conformations that are supercoiled and **wound around tetramers of the HU protein**, much like eukaryotic chromosomes are wrapped around histones.

Once the prokaryotic genome has been condensed, **DNA topoisomerase I, DNA gyrase**, and other proteins help maintain the supercoils.

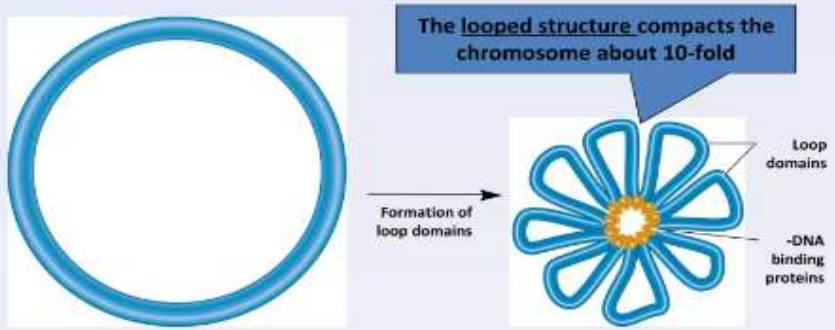
One of these maintenance proteins, H-NS (histone-like nucleoid structuring), plays an active role in transcription by modulating the expression of the genes involved in the response to environmental stimuli.

Another maintenance protein, factor for inversion stimulation (FIS- factor for inversion stimulation), is abundant during exponential growth and regulates the expression of more than 231 genes, including DNA topoisomerase I.

Bacterial chromosome is normally supercoiled

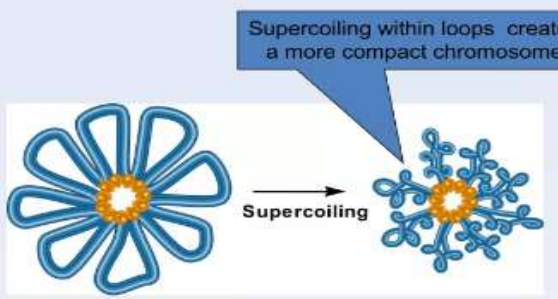


To fit within bacterial cell, the chromosome must be compacted ~1000-fold



Brooker, Fig 12.3

DNA supercoiling is a second important way to compact the bacterial chromosome

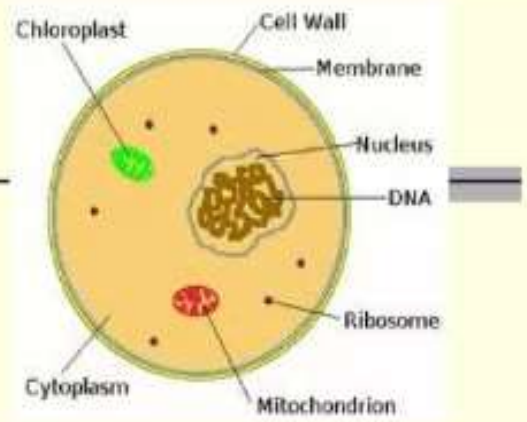


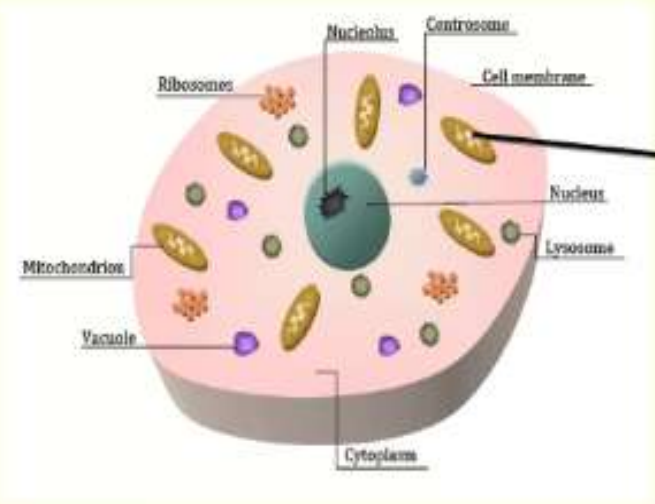
Brooker, Fig 12.3 -- illustration of DNA supercoiling

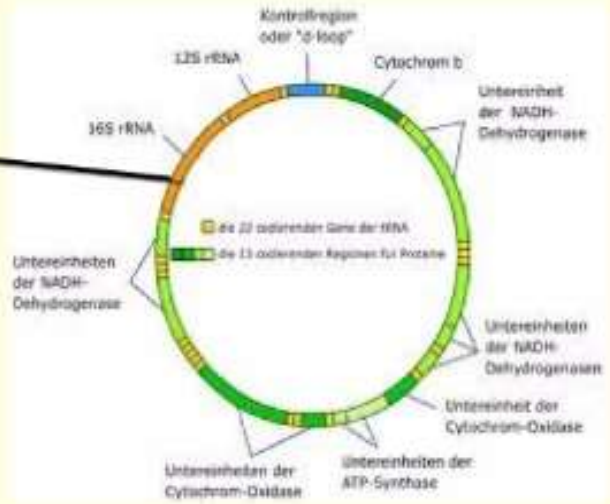
Which genes are present where?

Eukaryotic Genome

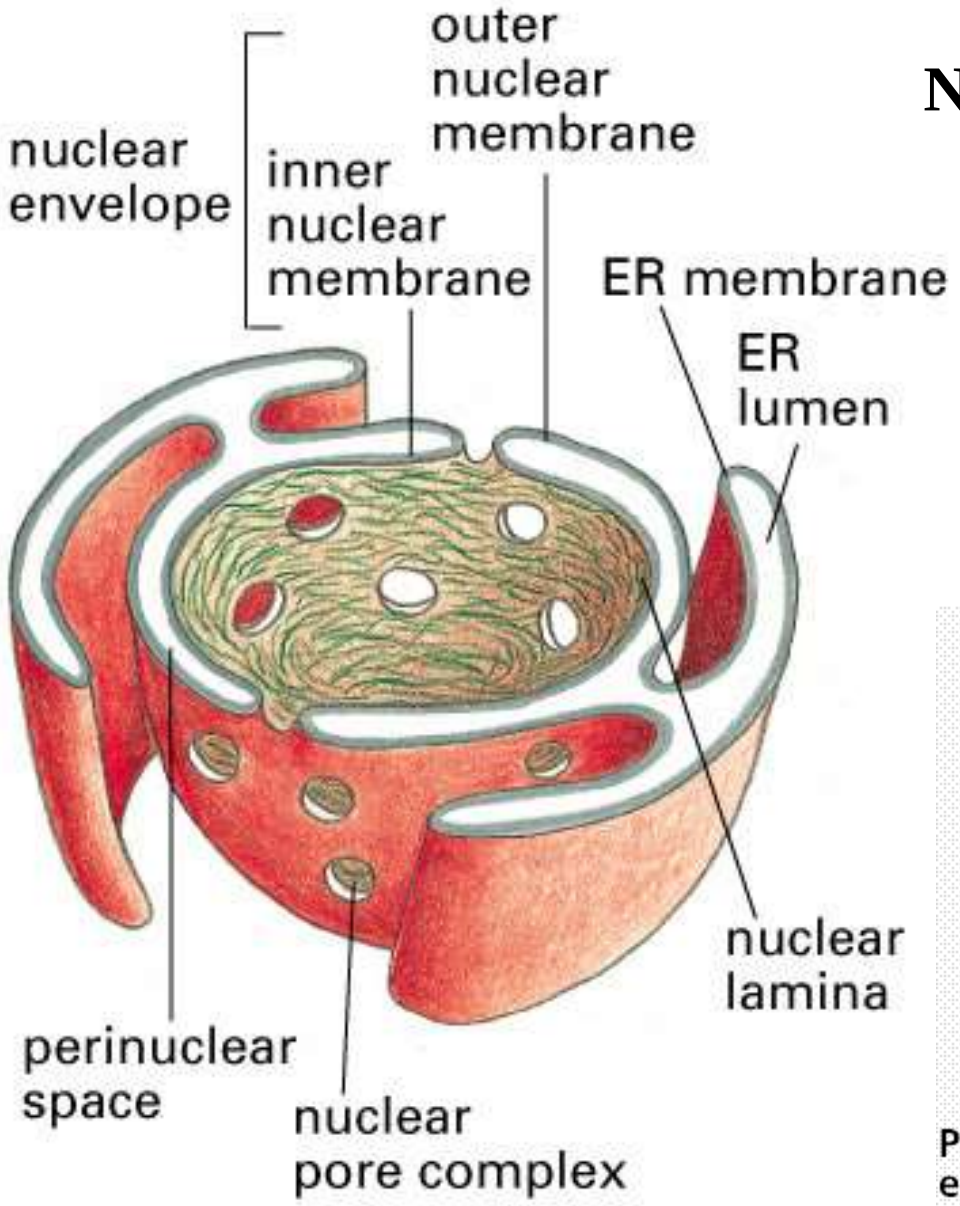
- Nuclear
- Mitochondrial
- Plasmids (in yeast)
- Chloroplasts (in Plants)







NUCLEUS



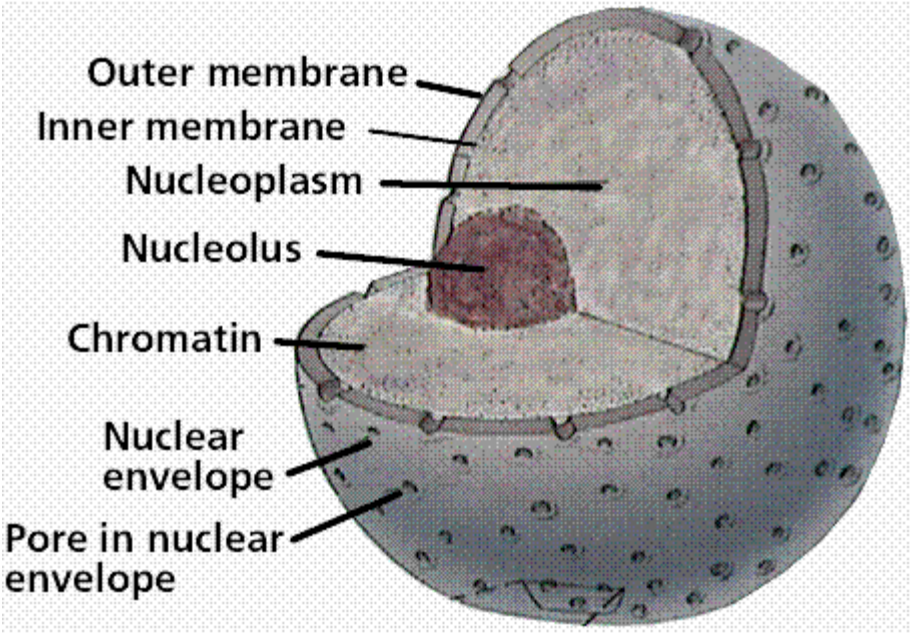
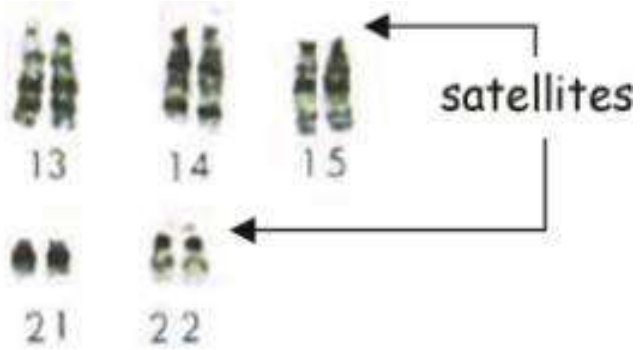


Figure 12–9. Molecular Biology of the Cell, 4th Edition.

The Nucleolus

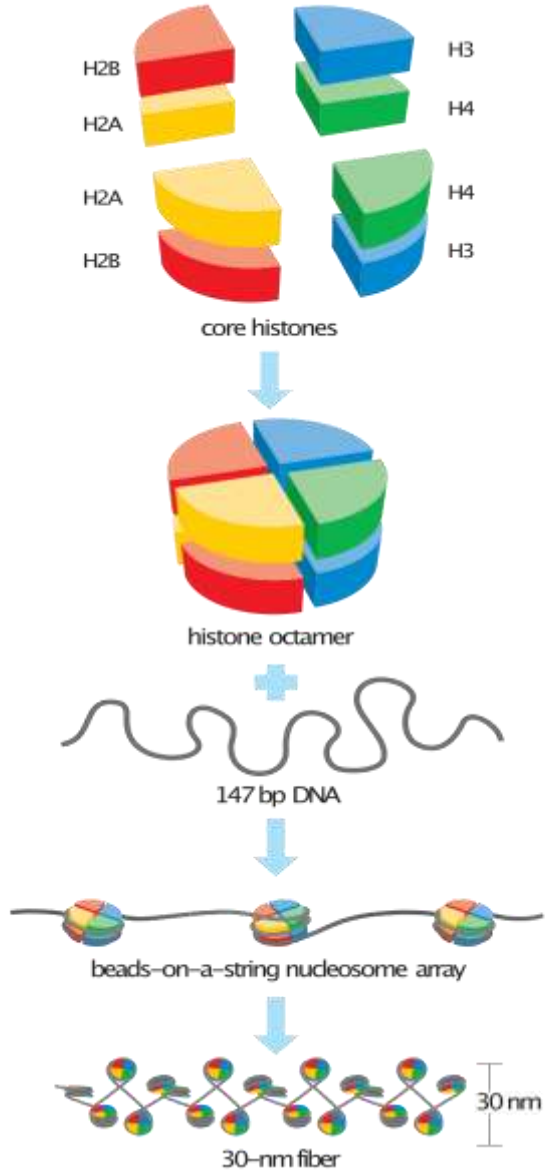
- Consists of two parts:
- fibrillar part - consists of chromatin:
- (DNA transcribing r-RNA)



- granular part – consists of ribonucleoprotein particles (r-RNA + proteins)

- The nucleolar DNA
- The genes for 5.8S, 18S and 28S r-RNA form a gene cluster
- they are all transcribed together forming a r-RNA complex of 45S
- these genes are present on the nucleolus organiser regions on the satellites of the acrocentric chromosomes 13, 14, 15, 21 and 22

- A **nucleosome** is the basic structural unit of DNA packaging in eukaryotes. The structure of a nucleosome consists of a segment of DNA wound around eight histone proteins and resembles thread wrapped around a spool. The nucleosome is the fundamental subunit of chromatin. Each nucleosome is composed of a little less than two turns of DNA wrapped around a set of eight proteins called histones, which are known as a histone octamer. Each histone octamer is composed of two copies each of the histone proteins H2A, H2B, H3, and H4.
- DNA must be compacted into nucleosomes to fit within the cell nucleus. In addition to nucleosome wrapping, eukaryotic chromatin is further compacted by being folded into a series of more complex structures, eventually forming a chromosome. Each human cell contains about 30 million nucleosomes.
- Nucleosomes are thought to carry epigenetically inherited information in the form of covalent modifications of their core histones. Nucleosome positions in the genome are not random, and it is important to know where each nucleosome is located because this determines the accessibility of the DNA to regulatory proteins.

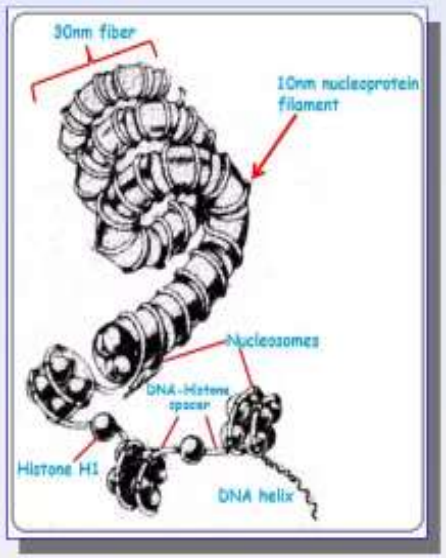


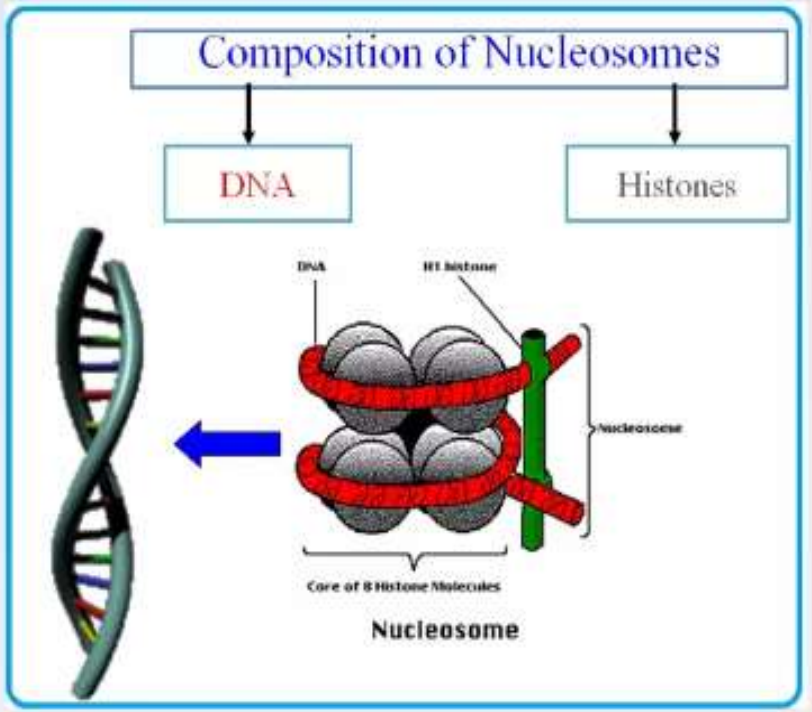
- Nucleosomes were first observed as particles in the electron microscope by Don and Ada Olins in 1974, and their existence and structure (as histone octamers surrounded by approximately 200 base pairs of DNA) were proposed by Roger Kornberg.
- The nucleosome core particle consists of approximately 146 base pairs (bp) of DNA wrapped in 1.67 left-handed superhelical turns around a histone octamer, consisting of 2 copies each of the core histones H2A, H2B, H3, and H4.
- Core particles are connected by stretches of linker DNA, which can be up to about 80 bp long. Technically, a nucleosome is defined as the core particle plus one of these linker regions.
- Linker histones such as H1 and its isoforms are involved in chromatin compaction and sit at the base of the nucleosome near the DNA entry and exit binding to the linker region of the DNA.
- Non-condensed nucleosomes without the linker histone resemble "beads on a string of DNA" under an electron microscope.
- In contrast to most eukaryotic cells, mature sperm cells largely use protamines to package their genomic DNA, most likely to achieve an even higher packaging ratio. Histone equivalents and a simplified chromatin structure have also been found in Archaea, suggesting that eukaryotes are not the only organisms that use nucleosomes.

How r they organised

Nucleosomes: The Basic Units of DNA Condensation

- The material of chromosomes, both protein and DNA, is often referred to as **chromatin.** The protein component is about equal in mass to the DNA component.
- Histones constitute the largest protein component of chromatin, are highly conserved, basic proteins that assemble into octameric complexes containing two each of four different histone subunits.
- DNA wraps around the histones to form condensed **nucleosomes.**

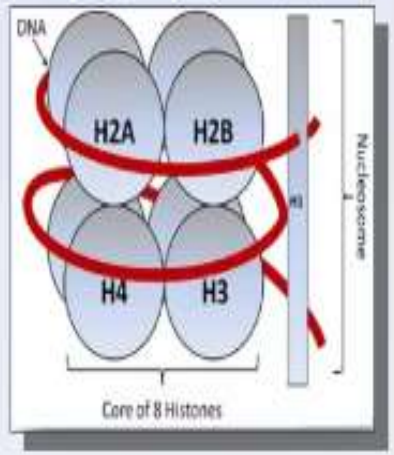




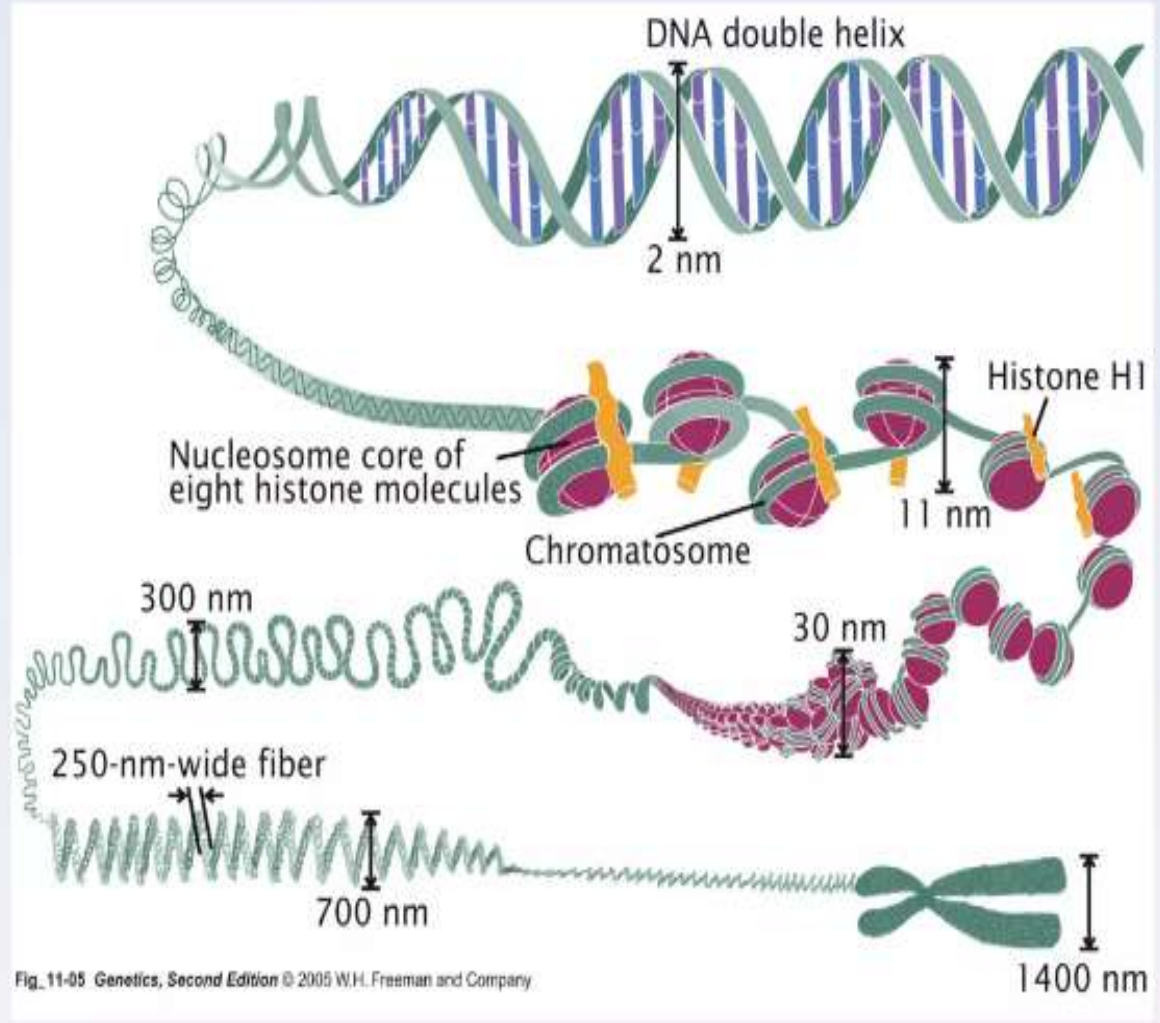
Nucleosome (10 nm diameter):

- 8 histones in bead & 1 outside.
- Each bead: is surrounded by 140 bpDNA and there are 60 bp in the linker region.
- Space between beads is about 14 nm.

- Histones are rich in the basic amino acids arginine and lysine, which together make up about 25% of the amino acid residues in any given histone protein.
- Histone proteins are highly conserved among eukaryotic cells.
- Histones H3 and H4 are nearly identical in all eukaryotes, suggesting strict conservation of their functions.
- Histones H1, H2A, and H2B show less sequence similarity, but on the whole, they are more conserved than other types of proteins.
- Salt bridges between positively charged histones and negatively charges DNA play a major role in stabilizing DNA-histone complex



Levels of DNA Packaging in Eukaryotes

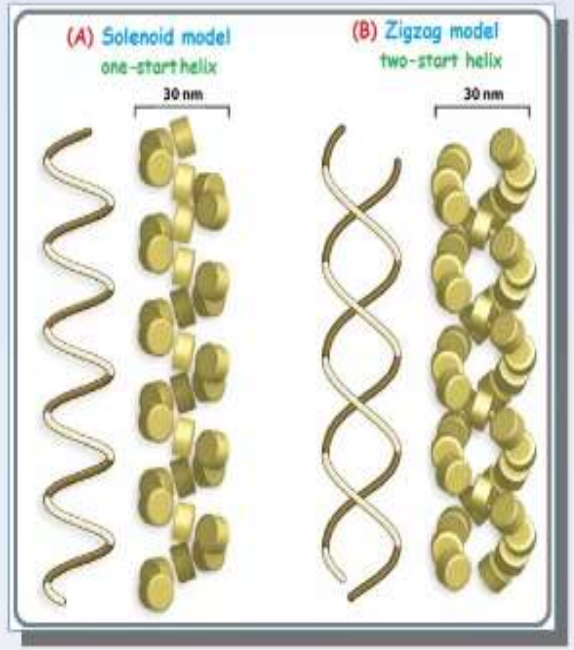


Chromosomes Condense into a Compact Chromatin Filament

- Nucleosomes condense into a compact filament with a width of about 30 nm, referred to as the 30 nm filament.
- H1 promotes condensation into the 30 nm filament, but it is not essential for forming the filament.
- In contrast, the N-terminal tails of the core histones are absolutely required, suggesting that the tails provide important nucleosome-nucleosome contacts needed for 30 nm filament formation.
- There are two most widely accepted models for nucleosome arrangement in the 30 nm filament:

i. Solenoid model

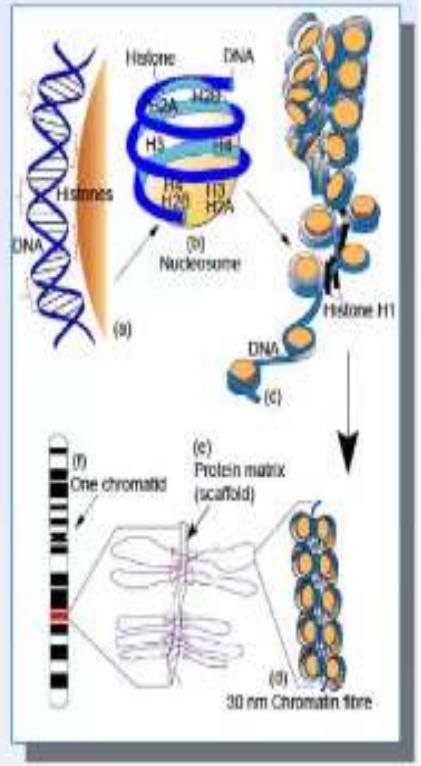
ii. Zigzag model



.The 30 nm filament, a higher-order organization of nucleosomes
The compact filament is formed by the tight packing of nucleosomes. Two proposed .models of filament structure are (A) the solenoid model and (B) the zigzag model
.]Source: (a) Barbara Hamkalo, Department of Molecular Biology and Biochemistry University of California, Irvine[

Higher-Order Chromosome Structure Involves Loops and Coils

- Inside chromosomes, DNA is much more highly condensed than in the 30 nm filament.
- 30 nm filaments is appear to be organized in loops estimated at 40 to 100 kbp long.
- **chromosomal scaffold:** Proteinaceous residue after extraction of histones from chromosomes, comprised mainly of Structural maintenance of chromosomes (SMC) proteins.
- Regions of the DNA interact with chromosomal scaffold proteins to give a protein core with DNA loops sticking out of it.
- This protein core then coils up to further package the DNA into the chromatids that are visible by light microscopy in metaphase.



Summary of Chromosome Folding

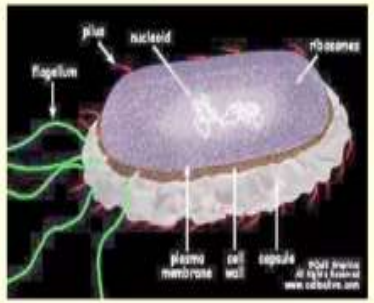
Level of folding	Consists of	Base pair/turn
DNA double helix	Nucleotides	10
Nucleosomes	bp each 200	100
Nanometer fiber 30	Nucleosomes /turn 6	1,200
Loops	Solenoids/loop 50	60,000
Miniband	loops 18	1,080,000
Chromatid	minibands 1,000,000	

Prokaryotes and Eukaryotes genome

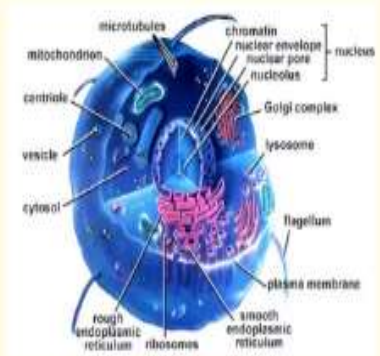
Prokaryotes	Eukaryotes
Single cell	Single or multi cell
No nucleus	Nucleus
One piece of circular DNA	Chromosomes
No mRNA post transcriptional modification	Exons/Introns splicing

Prokaryotic and Eukaryotic Cells
Chromosomal differences

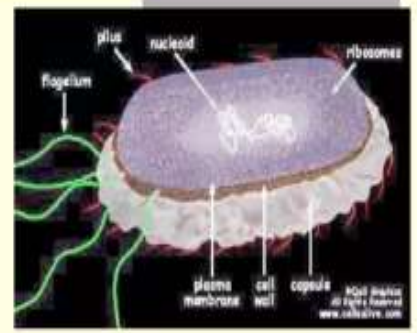
Prokaryotes



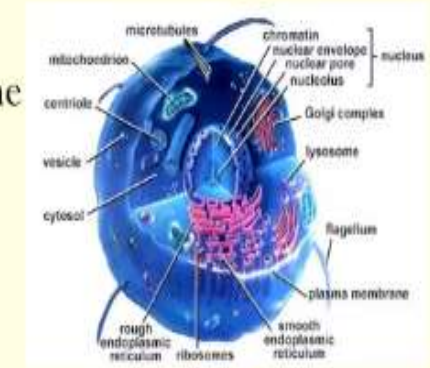
- The genome of E.coli contains amount of $4X10^6$ base pairs
- > 90% of DNA encode protein
- Lacks a membrane-bound nucleus.
  - Circular DNA and supercoiled domain
- Histones not present



Eukaryotes

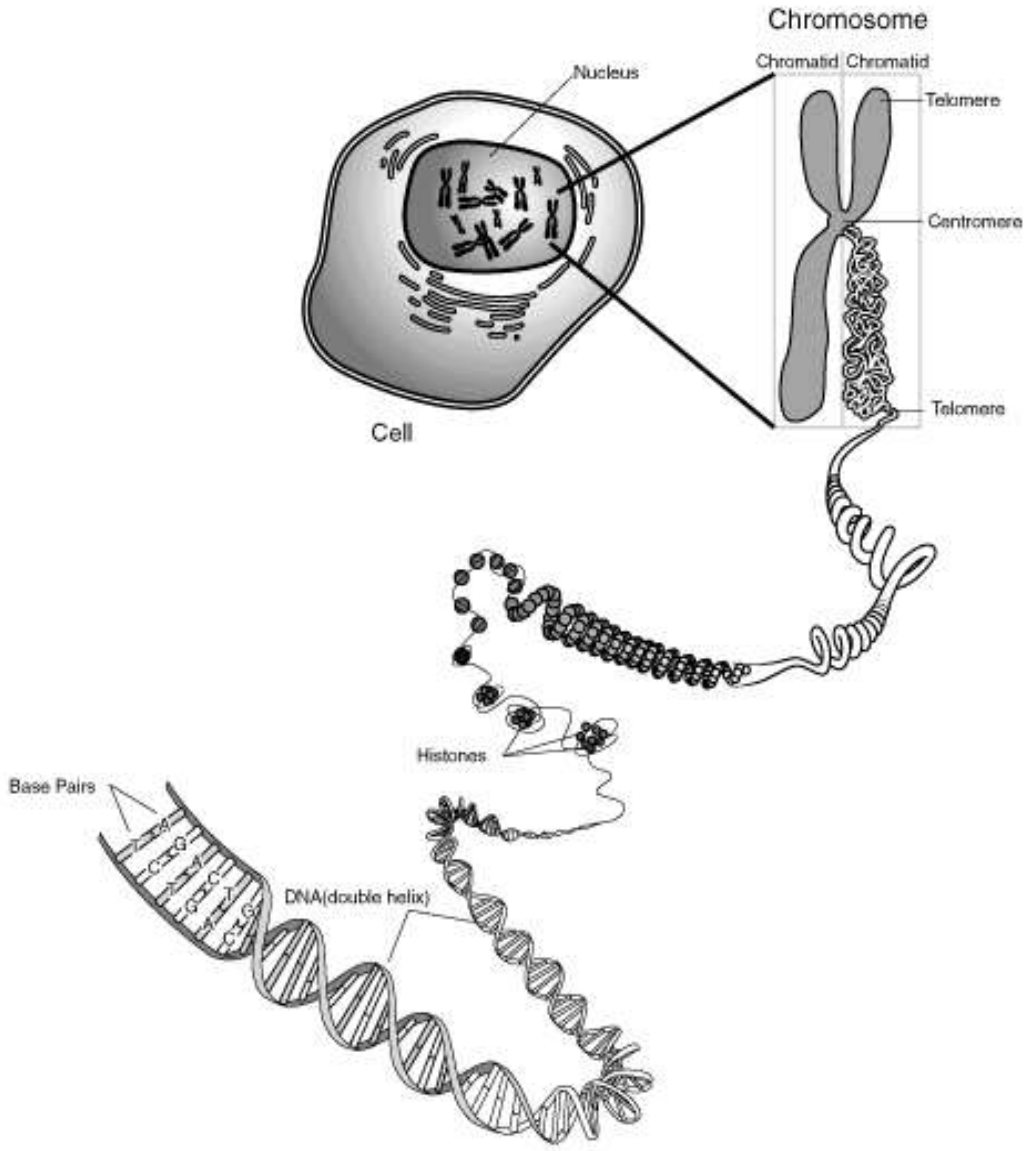


- The genome of yeast cells contains $1.35x10^7$ base pairs
- A small fraction of the total DNA encodes protein.
  - Many repeats of non-coding sequences
- All chromosomes are contained in a membrane bound nucleus
  - DNA is divided between two or more chromosomes
- A set of five histones
  - DNA packaging and gene expression regulation



Genome size, number of genes

- Bacterial genome size range:
  - 0.58 million bp (Mb), 467 genes (*Mycoplasma genitalium*)
  - 4.64 Mb, 4289 genes (*Escherichia coli*)
- Yeast *S. cerevisiae*: 12 Mb, 6241 genes
  - Only 2.6 X that of *E. coli*.
- *Caenorhabditis elegans*: 97 Mb; 18,424 genes
- *Drosophila melanogaster*: 180 Mb; 13,601 genes
  - ~120 Mb euchromatic (sequenced)
- *Homo sapiens*: ~3200 Mb; ~21,000 genes



Compared to bacterial genomes, vertebrate genomes:

- Are MUCH bigger
- Have a complex gene structure
- A majority of the DNA does NOT code for protein
- Are subject to wide-spread methylation, with multiple roles (not fully understood)
- Sustain continued “assault” by transposable elements
- Endogenous viruses
- DNA transposons
- Retro-transposons (move via RNA)
- Almost every genomic feature varies along chromosomes
- Gene density
- G+C content
- Repeat content
- Rates of nucleotide substitutions, insertions, deletions

Matters of size

- A common assumption about genomes would be that if genes specify proteins, then the more proteins an organism made, the more genes it would need to have, and thus, the larger its genome would be.
- Comparison of various genomes shows, surprisingly, that there is not necessarily a direct relationship between the complexity of an organism and the size of its genome.
- To understand how this could be true, it is necessary to recognize that while genes are made up of DNA, all DNA does not consist of genes.
- In the human genome, less than 3% of the total DNA seems to be the sort of coding sequence that directs the synthesis of proteins.
- For many years, non-coding DNA in genomes was believed to be useless, and was described as “junk DNA” although it was perplexing that there seemed to be so much “useless” sequence.
- Recent discoveries have, however, demonstrated that much of this so-called junk DNA may play important roles in evolution, as well as in regulation of gene expression.

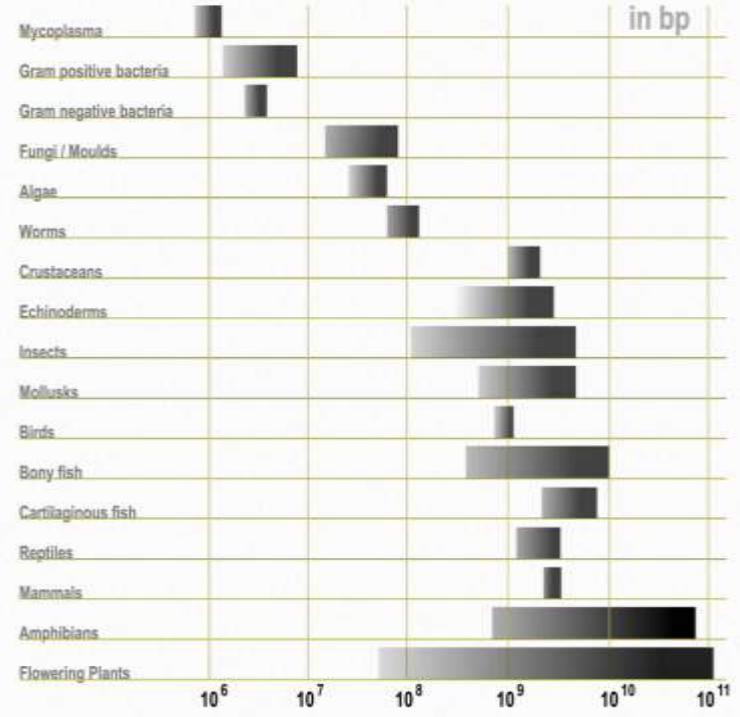


Figure. Sizes of various genomes

- The number of genomes sequenced in their entirety is now in the thousands and includes organisms ranging from bacteria to mammals.
- The first complete genome to be sequenced was that of the bacterium *Haemophilus influenzae*, in 1995.
- The first eukaryotic genome sequence, that of the yeast *Saccharomyces cerevisiae*, followed in 1996.
- The genome sequence for the bacterium *Escherichia coli* became available in 1997 .
- The much larger effort directed at the human genome was also accelerating.

Genome size and complexity

Genome size and complexity

- The genome of an organism is the whole DNA content of its cells, including genes and intergenic regions.
- In prokaryotes (Archaea and Bacteria) there is, in general, a linear relationship between genome size and the number of genes.
- The smallest genomes are found in symbionts and parasites, as they undergo a gene degradation process during adaptation to their new lifestyle.
- However, in eukaryotes there is no correlation between genome size and the complexity of the organism. This is known as the C-value paradox.
- The largest genome is found in an amoeba, a one-cell organism, with 686,000 Mb, 200 fold larger than the human genome and 20,000 fold larger than the one found in yeast.
- Now we know that most excess DNA is repetitive DNA, apparently lacking a function (selfish DNA) and whose possible role in genome evolution is still unknown.

- **GENOME RANGES AND SIZES**
- Figure represents the range of sizes of genome found in the three domains of life: bacteria, archaea and eukaryotes.
- It seems clear that prokaryotes are, in general, smaller than eukaryotes, with the exception of some large-sized bacteria and some very small-sized eukaryotes. Let us see the data in more detail.

PROKARYOTES: BACTERIA AND ARCHAEA

- According to the data published so far, the size varies from 0.58 megabases (1 megabase (Mb) is one million base pairs (bp)) in the intracellular pathogen *Mycoplasma genitalium*, to more than 10 Mb in several species of *Cyanobacteria*, with the exception of *Bacillus megaterium*, which has a genome of 30 Mb.
- The second smallest genome ever published is that of *Buchnera* sp. APS, endosymbiont of the cereal aphid *Acyrthosiphon pisum*, with a size of 641 kb.

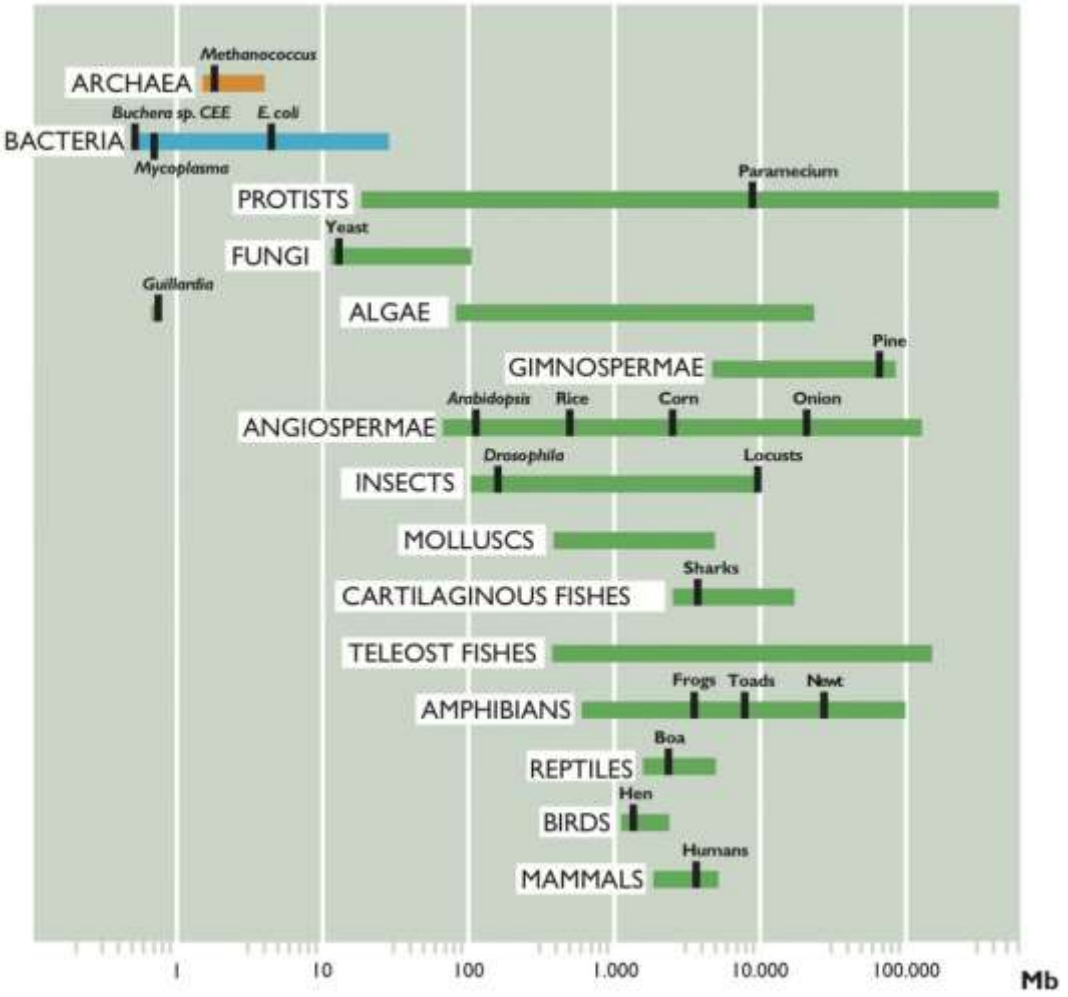


Fig. Range of genome size in organisms of the three domains of life.

- Recently, six genomes smaller than even those of Mycoplasma has been characterised, the smallest of all being that of *Buchnera* sp. CCE, endosymbiont of the aphid *Cinara cedri*, with a size of 0.45 Mb.
- Most genomes are less than 5 Mb in size, as shown in Table.

Organism				
Common name or class	Scientific name	Genome size (Mb)	Number of genes	Gene density (genes/Mb)
Eukaryotes				
Baker's yeast	Saccharomyces cerevisiae	12	6,241	480
Nematode	Caenorhabditis elegans	97	18,424	190
Cruciferous	Arabidopsis thaliana	125	25,498	204
Fruit fly	Drosophila melanogaster	180	13,601	75
Pufferfish	Fugu rubripes	400	35,000	100
Rice	Oryza sativa	450		
Sea urchin	Strongylocentrotus purpuratus	900	27,350	30
Maize	Zea mays	2,400		
Human	Homo sapiens	3,400	35,000	10
Onion	Allium cep	18,000		
Amoeba	Amoeba dubia	686,000		
Archaea				
Crenarchaeota	Aeropyrum pernix	1.55	1,522	981
Euryarchaeota	Methanococcus jannaschii	1.66	1,715	1033
Euryarchaeota	Archaeoglobus	2.18	2,420	1110
Bacteria				
Proteobacteria	Buchnera sp. CCE	0.45		
Gram positive	Mycoplama genitalium	0.58	479	831
Proteobacteria	Buchnera sp. APS	0.64	564	881
Gram negative	Haemophilus influenzae	1.8	1,727	959
Cyanobacteria	Synechocystis sp.	3.6	3,168	880
Gram positive	Bacillus subtilis	4.2	4,100	976
Proteobacteria	Escherichia coli	4.6	4,288	932

Table. Genome size, gene number and gene density.

- Is there a relationship between genome size and number of genes?
- The size of the prokaryotic gene is uniform, about 900 to 1000 bp.
- Therefore, one can estimate the gene density at each sequenced genome.
- As seen in Table, gene density is more or less constant, both in bacteria and archaea.
- We can conclude that, at least in prokaryotes, genomes have a higher gene density. The number of genes reflects the life-type.
- Thus, smaller bacteria are specialists, such as obligate parasites and endosymbionts, and larger bacteria are generalists, and may even have a certain degree of development, such as sporulation in *Bacillus*.

- EUKARYOTES: C-VALUE PARADOX
- Genome size in eukaryotes is defined as the **C-value or amount of DNA per haploid genome,** such as that which exists in the nucleus of a spermatozoon. It is called C, for constant or characteristic, to indicate the fact that size is practically constant within a species.
- Referring back to Figure, we see that, in general, eukaryotes have larger genomes than prokaryotes, except for some endosymbiont or parasitic green algae, which have very small genomes.
- Specifically, the smallest eukaryotic genome ever sequenced is that of *Guillardia theta*, a symbiont red algae, of only 0.55 Mb.
- We can also see in the figure that there is a wide range of sizes, much greater than that of prokaryotes, more than 80,000-fold larger, from organisms such as yeast (1.2 Mb) to the amoeba (686,000 Mb).
- As we can observe, unicellular protists such as amoebae show the greatest variation in C-values (23.5 Mb to 686.000 Mb, with a ratio of 29,191 between the largest and the smallest), while mammals, birds and reptiles show less variation in the size of their genome (a ratio of only 4, 1 and 4, respectively).

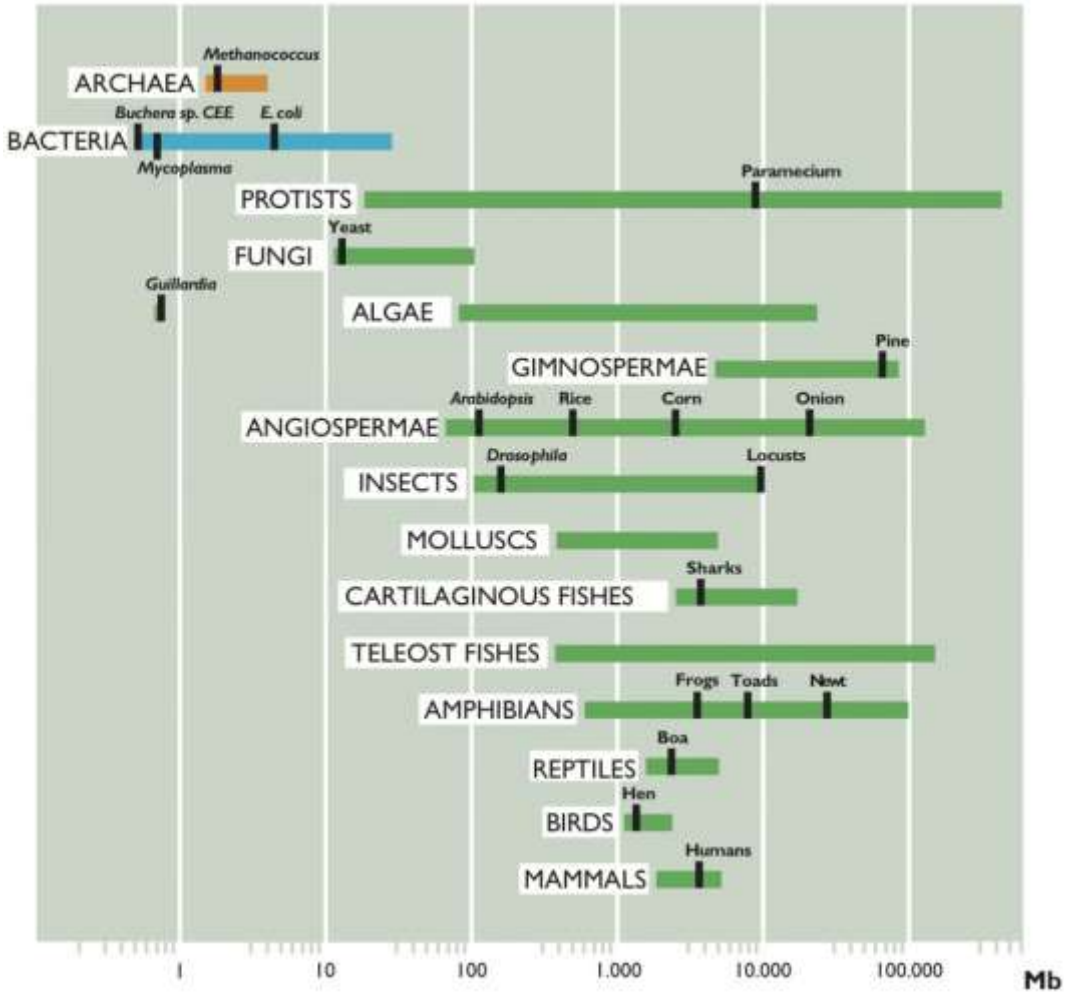


Fig Range of genome size in organisms of the three domains of life.

- Furthermore, the large variation in genome sizes between eukaryotic species does not seem to have a relationship with either the complexity of the organism or the number of genes they contain.
- For example, amoebae, which have the largest genomes, have 200 times more DNA than humans (3,400 Mb) and it is clear that an amoeba cannot be more complex than a human.
- Moreover, it would be expected that mammals, more complex organisms, present larger genomes. However, many other organisms, such as fish, amphibians or plants, have much larger genomes.
- Even when we compare the sizes between organisms that appear similar in terms of complexity, there are also wide differences in their C-values.
- To give some examples, flies and locusts, onions and lilies, etc. have considerable variations in the sizes of their genomes.
- Amphibians as a group have variations of up to 91 times and it is hard to believe that this may reflect variations of nearly 100 times the number of genes necessary to give rise to the corresponding amphibians, or that onions need 200 times more DNA than rice.
- Figure shows some living beings with size proportional to the size of their genome and needs no further explanation.

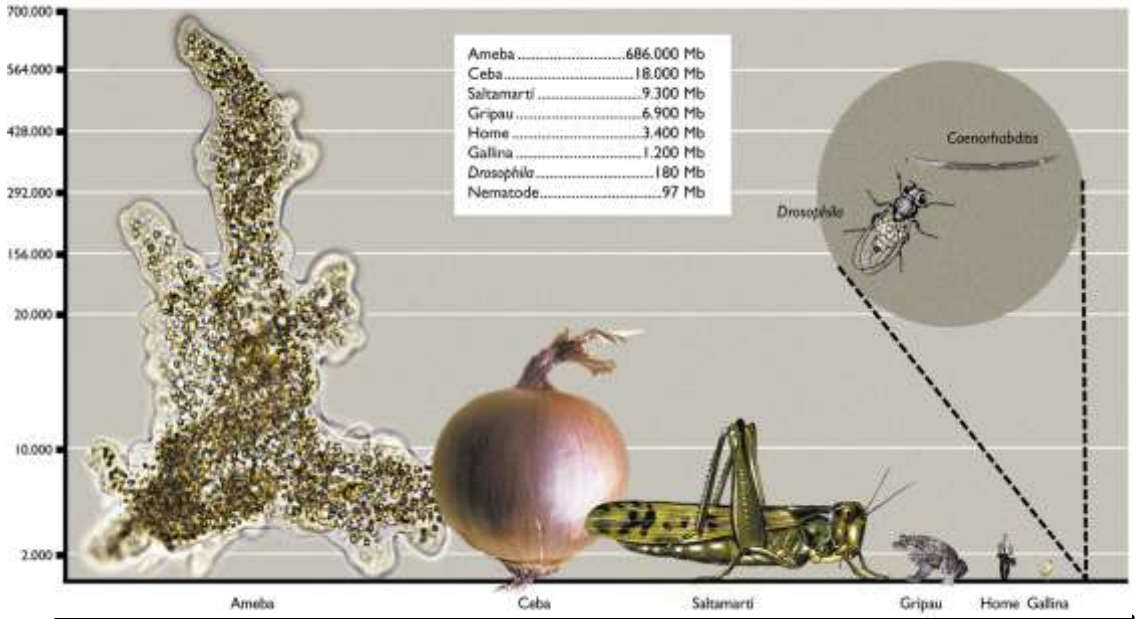


Figure Genome size in some living beings. The height of the drawings is proportional to the size of their genome. The specimens are amoebae, onions, grasshoppers, toads, humans, hens, Drosophila and Caenorhabditis, a nematode worm.

- The mismatch between the C-values and the presumed amount of genetic information contained within the genomes was called C-value paradox. Since we cannot assume that a species possesses less DNA than the quantity required to specify its vital functions, we have to explain why many species contain this amount of excess DNA.

GENE DNA OR NON-GENE DNA

- The first question that has to be clarified is whether there is a correlation between genome size and the number of genes. That is, are the differences in genome sizes due to gene or non-gene DNA?
- We have known since the late 60s that the eukaryotic genome is composed of a large amount of repetitive DNA. Moreover, since the late 70s we have known that genes are interrupted by non-coding sequences, introns, which must be removed before the ribosome synthesizes protein.
- We are talking in both cases about a seemingly superfluous DNA, which contributes to the wide variation in C-values and therefore explains the apparent paradox.
- The size and number of introns vary widely along the evolutionary scale, mammals being the ones with the highest number and larger size.
- Repetitive DNA also varies between organisms. Traditionally this DNA is classified as: highly repetitive, with sequences such as microsatellites and minisatellites; and moderately repetitive, where transposable elements, the sequences that constitute the clearest example of selfish DNA, are found.

Number of genes and complexity of the organism

- As sequences of whole genomes are completed, we will know with more or less accuracy the number of genes derived from these sequences, since what we had so far were indirect estimates.
- However, some data is proving to be surprising because, in some cases, there appears to be a clear correlation between the number of genes and the complexity of the organism.
- The nematode worm *C. elegans* has 18,000 genes (Table 1), about 5,000 more than *Drosophila*, a more complex organism.
- Man has only twice as many genes as *C. elegans* (estimates indicated about 100,000).
- We are also in beginning to understand these data. There are mechanisms in higher eukaryotes that are able to expand the proteome. That is, from the same DNA sequence, they can obtain more than one protein.
- Large introns found in mammals can, in many cases, hide information that cannot be inferred only with the DNA sequence.
- So in short more complex organisms have more gene functions.

Global genome initiative

- The Global Genome Initiative, a collaborative effort to sequence at least one species from each of the 9,500 described invertebrate, vertebrate, and plant families is one of many such ventures.
- The information from these various efforts is collected in enormous online repositories, so that it is freely available to scientists.
- As the sequence databases compile ever more information, the fields of computational biology and bioinformatics have arisen, to analyse and organize the data in a way that helps biologists understand what the information in DNA means in the cellular context.

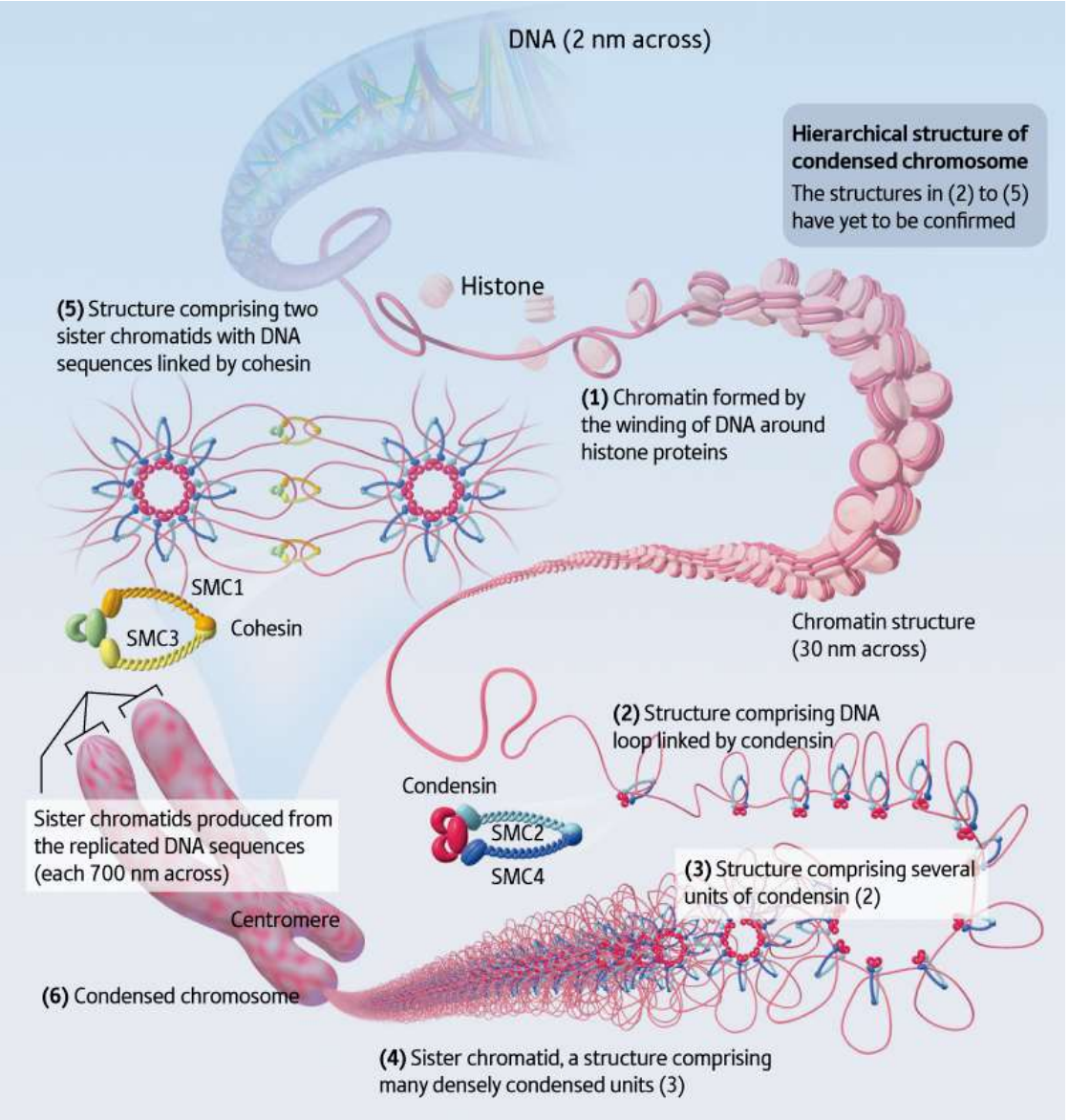


Figure 2. Condensation of DNA into chromatin and chromosome

Department of BSBE
Indian Institute Of Technology Guwahati

Isolation of genes
&
Primer design

Dr. Sanjukta Patra

BT 207

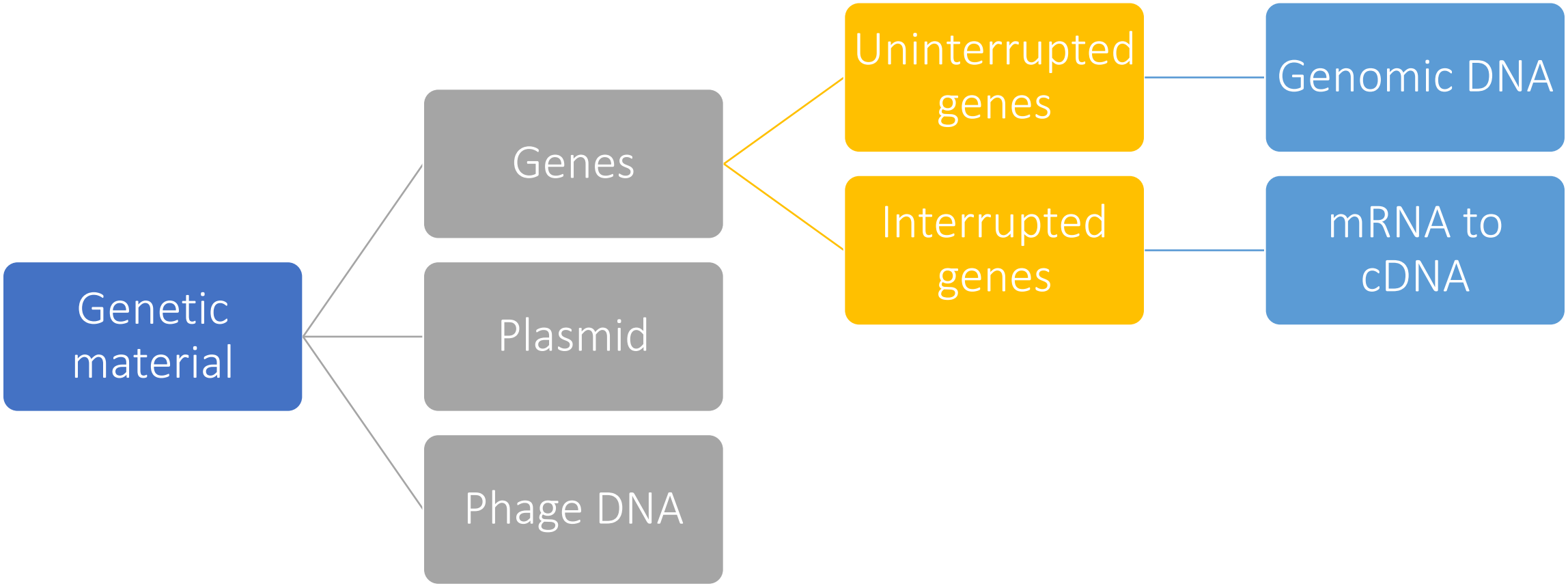
Genetic Engineering

Jan-April 2022

STEPS for cloning

- Isolation of the template
- Know the gene
- Design the primer
- Go for amplification
- Check for size
- Confirmation of target gene
- Restriction digestion and Ligation into a vector
- Screening of colony
- Sequencing and clone confirmation

Forms of Genetic material in cells

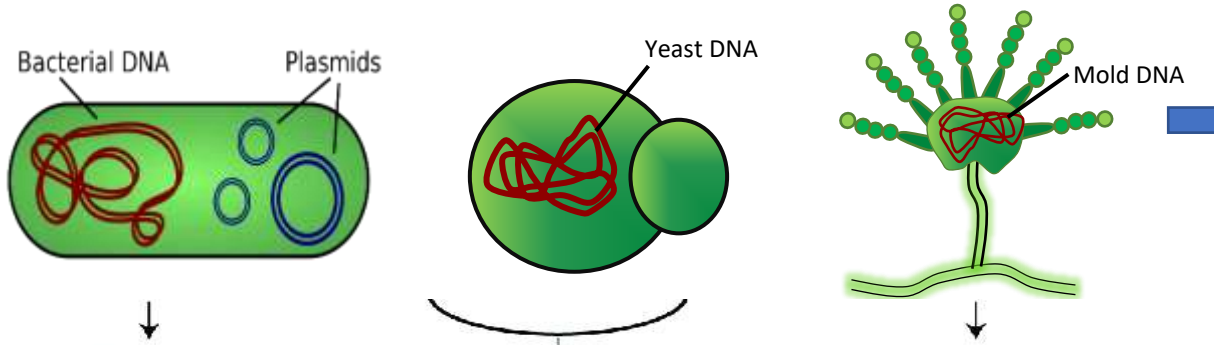


Genomic DNA and Plasmid isolation

1. Isolation of the template

Genomic DNA is the “instruction booklet” for the cell- WHY?

Colonies on agar plate



Cell lysis by lysozyme

Cell break with glass beads in **GX buffer**

Cell break by crushing

+ GX buffer
+ Proteinase K

+ Proteinase K

+ GX buffer
+ Proteinase K

GX buffer: Extraction buffer with detergents & salts to maintain pH

Incubation at 60°C for 30 min

Neutralization with a 3 M sodium acetate solution

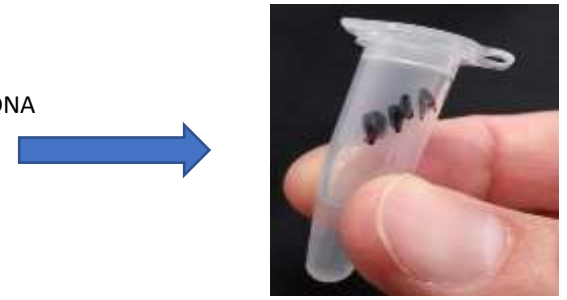
Centrifugation and collection of the supernatant

DNA/RNA precipitation in cold isopropanol

Washing the pellet with ethanol of 70%

DNA sample treated with RNase

Genomic DNA product



PCR-ready template

Genomics is divided into two basic areas

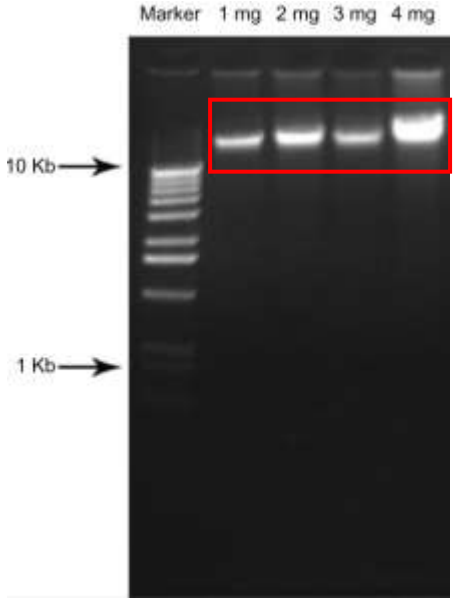
structural genomics, characterizing the physical nature of whole genomes

functional genomics, characterizing the transcriptome (the entire range of transcripts produced by a given organism) and the proteome (the entire array of encoded proteins)

Human Genomic DNA has:
Protein-Coding Genes
Regulatory DNA Sequences
Genes for Non-coding RNA
Introns

RNase A treatment is used for **the removal of RNA from genomic DNA samples**. RNase A cleaves the phosphodiester bond between the 5'-ribose of a nucleotide and the phosphate group attached to the 3'- ribose of an adjacent pyrimidine nucleotide!

Traditional agarose gels are most effective at the visualization of genomic DNA generally ranging in sizes of >10kb upto 25 kb.

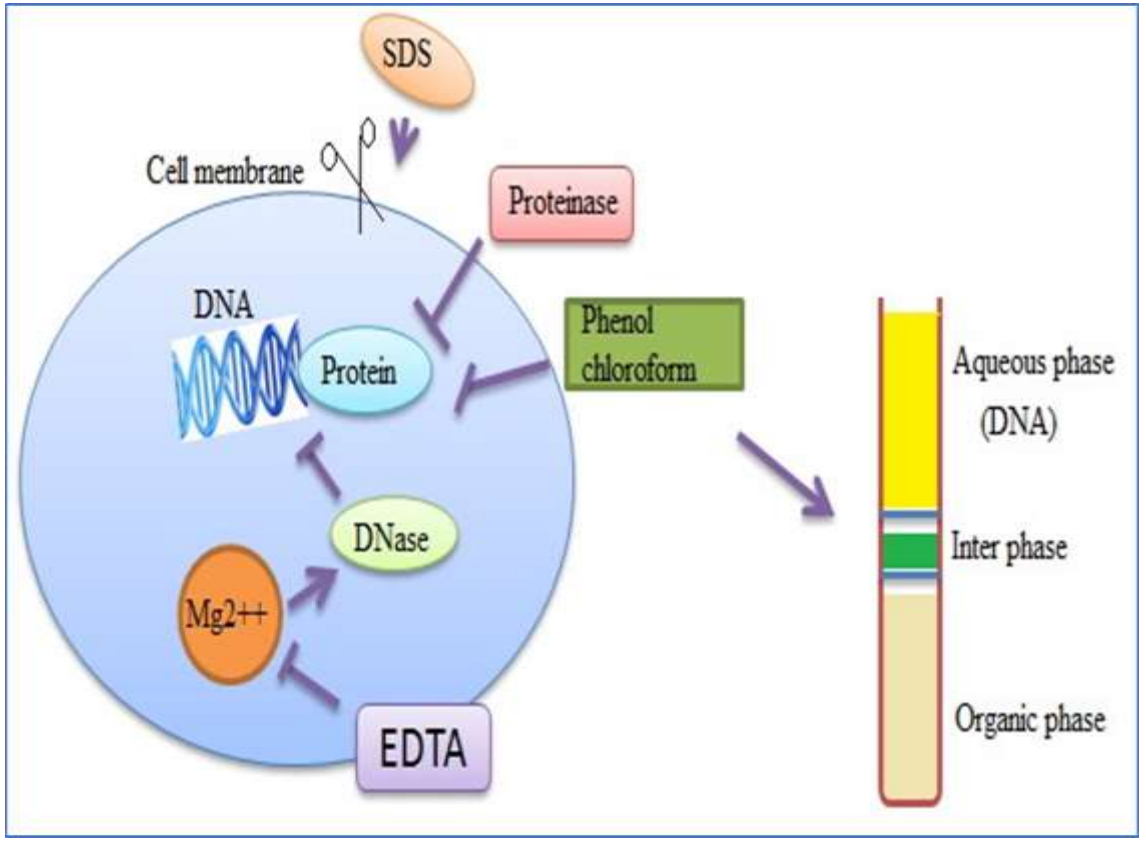


Genomic DNA on agarose-gel electrophoresis

Genomic DNA isolation

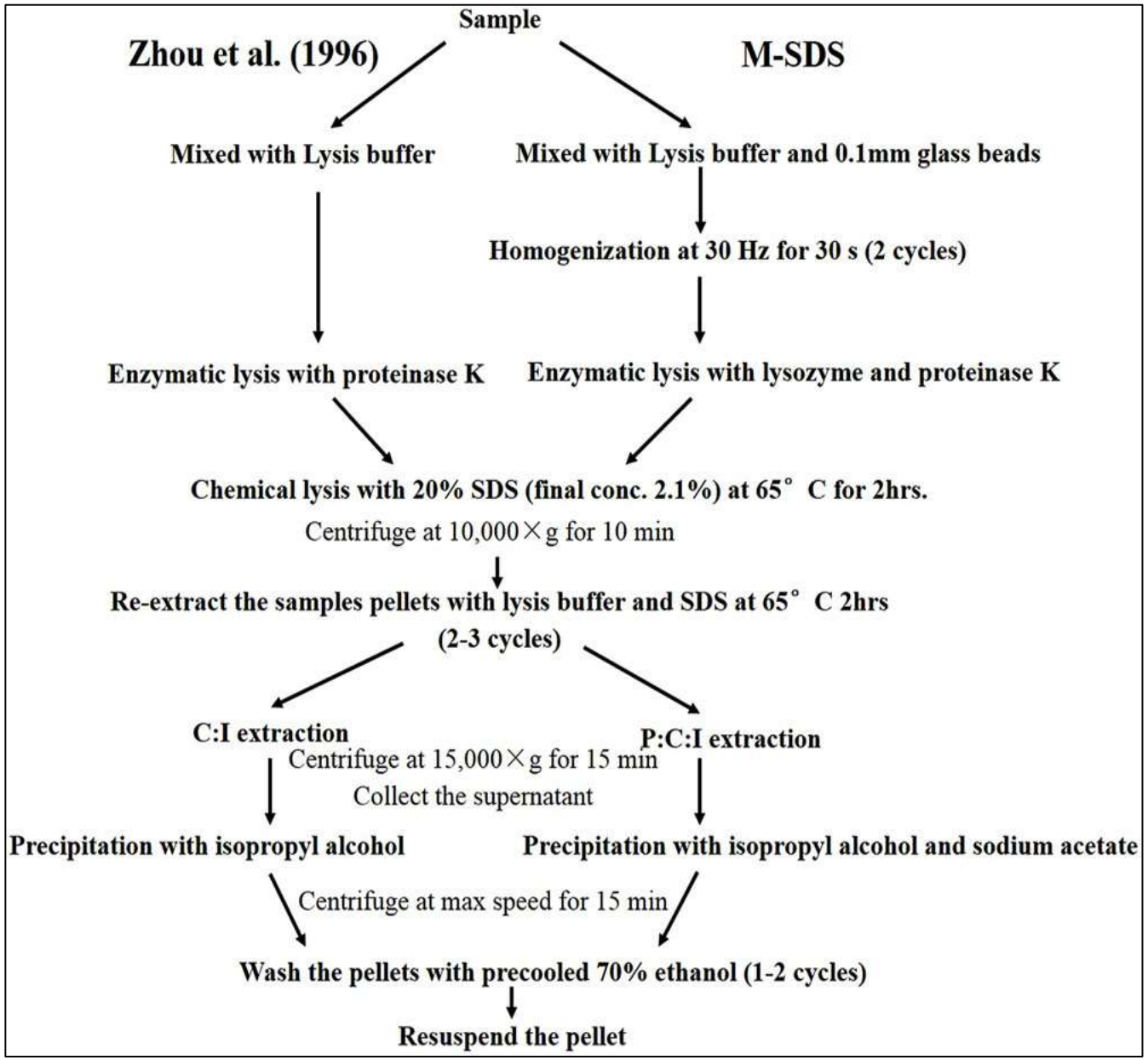
Basic aims:

- To lyse cell walls/cell membrane, nuclear membrane (bacteria, fungi, plant, etc.)
- To remove proteins
- To remove RNA
- Inactivate DNase
- To collect DNA and preserve it from breakage



Schematic diagram showing the principle of isolation of genomic DNA from bacteria.

Basic Protocols for Genomic DNA isolation



- Lysozyme acts on Gram positive cell walls by disintegrating peptidoglycan layer.
- It can not penetrate gram negative cell walls as it can not penetrate lipopolysaccharide layer.

Isolation of bacterial genomic DNA

The genomic DNA isolation needs to separate total DNA from RNA, protein, lipid, etc.

Initially the cell membranes must be disrupted in order to release the DNA in the extraction buffer.

Lysozyme acts on Peptidoglycan of Gram +ve cell wall.

SDS (sodium dodecyl sulphate) is used to disrupt the cell membrane.

Once cell is disrupted, the endogenous nucleases tend to cause extensive hydrolysis.

DNA can be protected from endogenous nucleases by chelating Mg^{2++} ions using EDTA. Mg^{2++} ion is considered as a necessary cofactor for action of most of the nucleases.

Nucleoprotein interactions are disrupted with SDS, phenol or proteinase K.

Proteinase K enzyme is used to degrade the proteins in the disrupted cell soup.

Phenol and chloroform are used to denature and separate proteins from DNA.

Chloroform is also a protein denaturant, which stabilizes the rather unstable boundary between an aqueous phase and pure phenol layer.

The denatured proteins form a layer at the interface between the aqueous and the organic phases which are removed by centrifugation.

DNA released from disrupted cells is precipitated by cold absolute ethanol or isopropanol.

Plasmid isolation
Alkaline Lysis Method

- Alkaline lysis method is one of the most commonly used method for lysis bacterial cells prior to plasmid purification. It has four basic steps :-

- 1. Resuspension : Harvested bacterial cells are resuspended by using solution I contains EDTA (ethylene diamine tetra-acetic acid) and Tris-CL.

  EDTA – chelates the magnesium and calcium ions

  Tris-CL – maintains pH.

- 2 . LYSIS : Cells are lysis with alkaline solution II contains NaOH and SDS (sodium dodecyl sulfate).

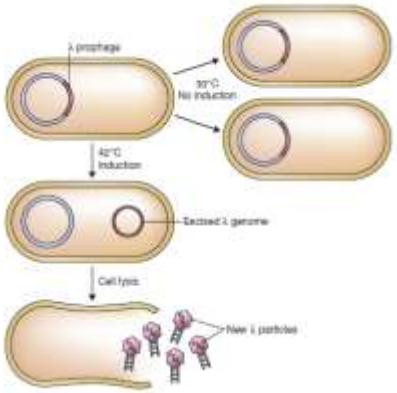
  NaOH -- denatures the chromosomal and plasmid DNAs as well as proteins.

  SDS -- solubilizes the phospholipids and protein components of the cell membrane, leading to lysis and release of the cell membrane.

- 3. NEUTRALIZATION : The lysate is neutralized by addition of solution III of acidic potassium acetate. The high salt concentration causes potassium dodecyl sulfate (KDS) to precipitate and denatured proteins, chromosomal DNA and cellular debris are co-precipitated in insoluble.

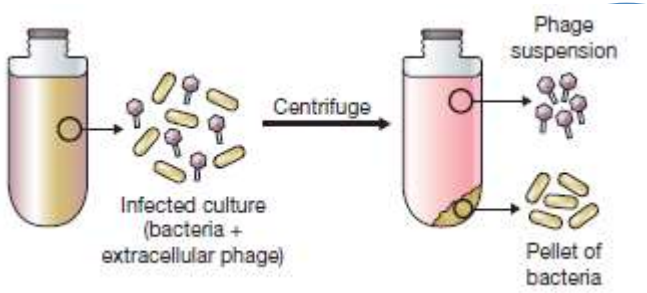
- 4. CLEARING OF LYSATES : Precipitated debris is removed by either high speed centrifugation or filtration, producing cleared lysates

Isolation and Purification of DNA from λ bacteriophages



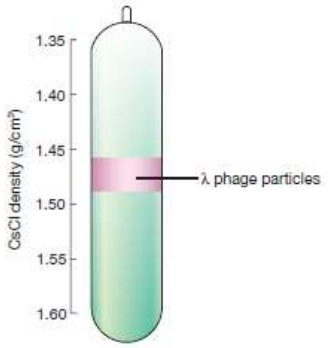
Growth of bacterial (λ infected) cultures to obtain a high λ titer

- Large volume cultures (500-1000) ml
- the culture must be induced, so that all the bacteria enter the lytic phase of the infection cycle, resulting in cell death and release of λ particles into the medium.
- the stage at which the cells are infected is the key to obtaining a high titer (not too low or high bacterial cell density)



Collection of phages from an infected culture

- The culture is centrifuged, the bacteria are pelleted, leaving the phage particles in suspension
- Collection of phages is therefore usually achieved by precipitation with polyethylene glycol (PEG)

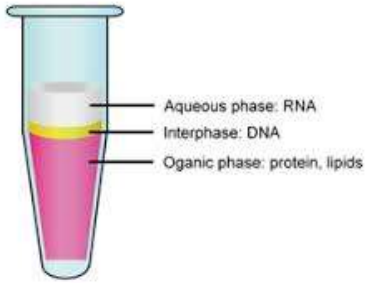


Purification of DNA from λ phage particles

- By CsCl density gradient centrifugation to remove bacterial debris, possibly including unwanted cellular DNA, which may exist in PEG precipitate.
- The λ particles band in a CsCl gradient at 1.45–1.50 g/cm3

mRNA EXTRACTION

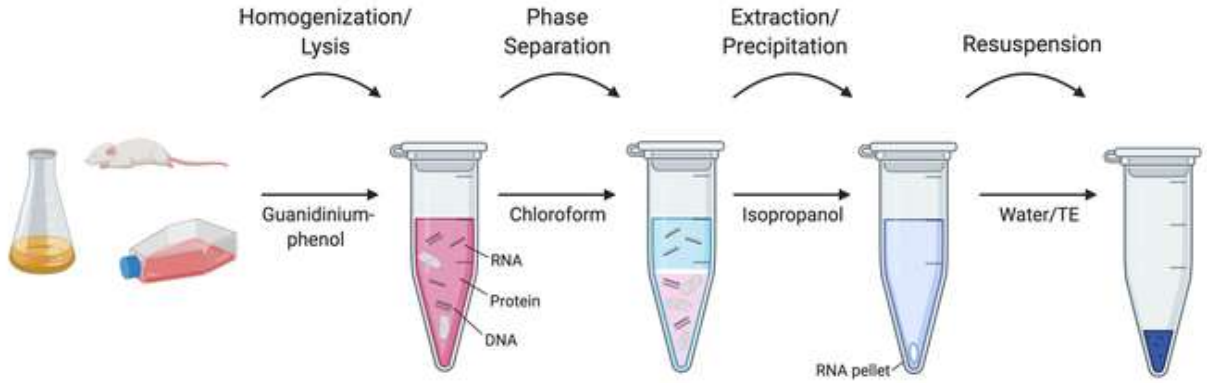
mRNA extraction is important to research and industry settings. In research, mRNA offers important insight into **what proteins are being translated or how much transcript is being produced by the cell**.

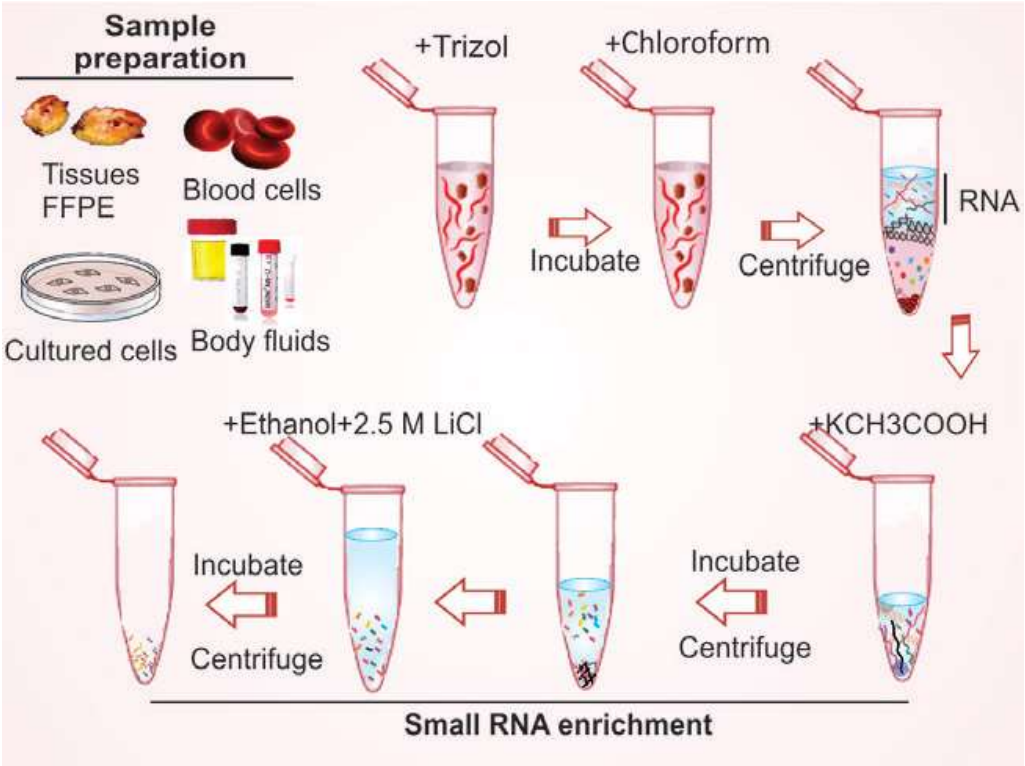


The TRI in TRIzol stands for Total RNA Isolation

TRIzol Reagent is a ready to use mixture of phenol, guanidine isothiocyanate, red dye and other proprietary components that can be used to isolate total RNA. DNA and proteins can be recovered with sequential precipitation from the organic phase.

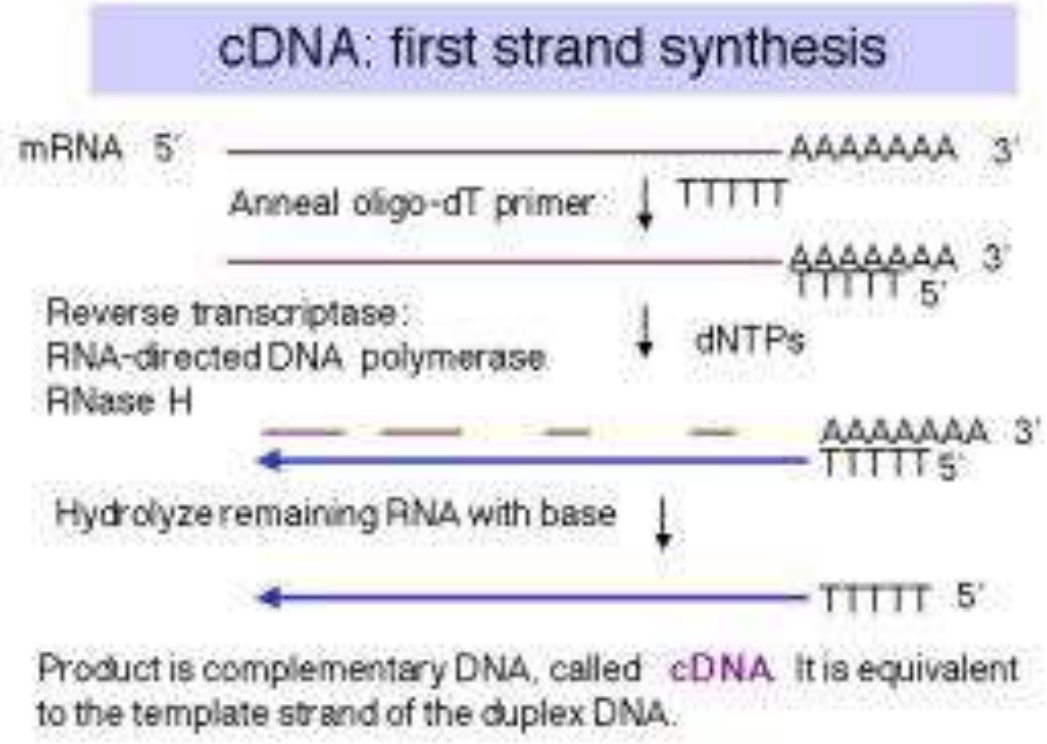
TRIzol was developed by Piotr Chomczynski. The red dye allows easy detection of the organic phase and is non-interactive with nucleic acids.



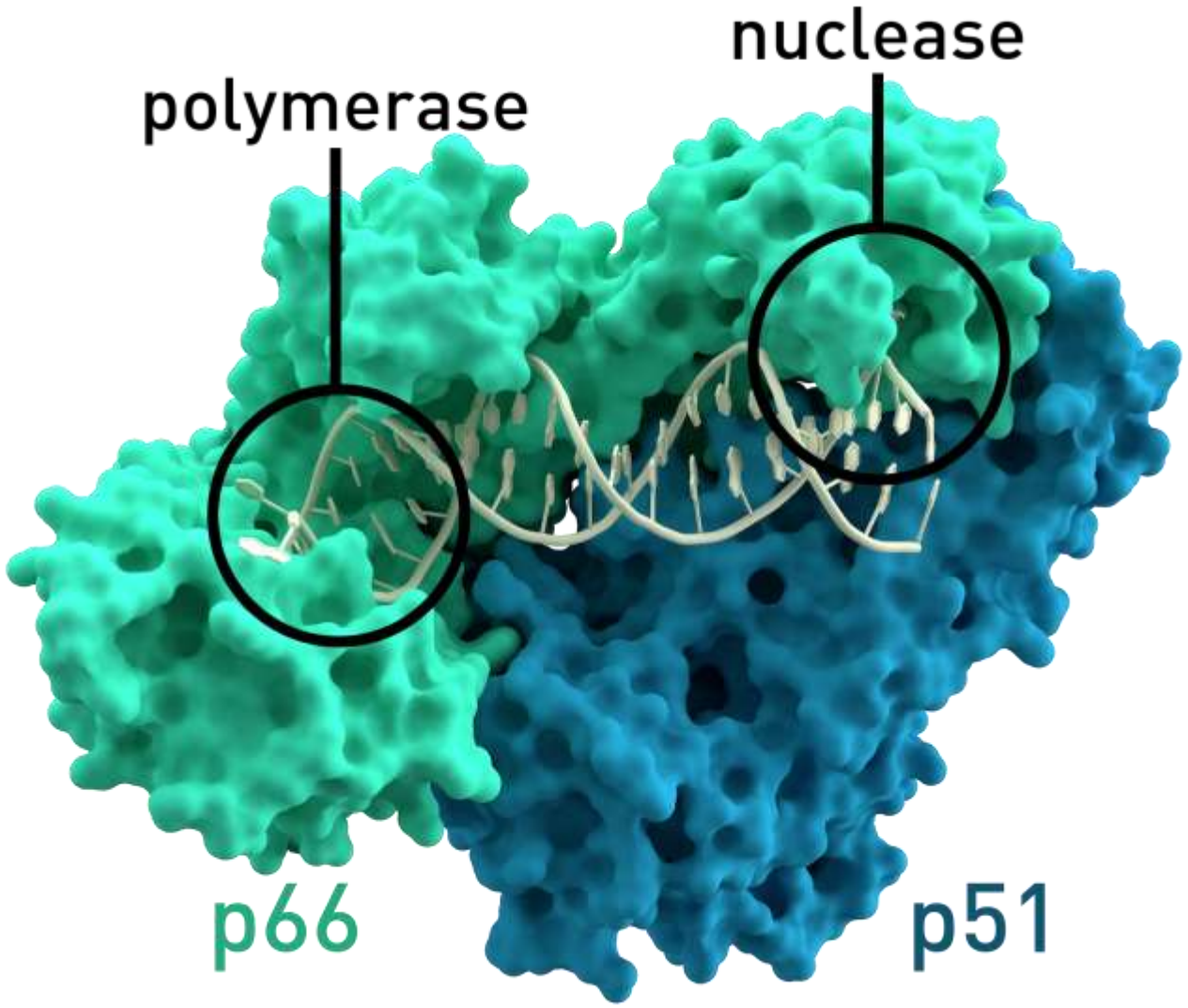


- Clinically it has become vastly important with its role in the SARS-CoV-2 vaccine. Synthetic mRNA production requires a purification step to get the final mRNA product out of the solution in which it was transcribed.
- **mRNA only accounts for 5% of the RNA in the cell** so it is important to have a technique which will specifically purify this type of RNA.
- RNA is also very sensitive to RNase contamination, which is found all over your skin as an antimicrobial. To avoid contamination it is helpful to have an efficient and simple method for mRNA extraction.
- A common method for mRNA extraction is the **use of magnetic beads.** The beads utilize the poly-A tail on mRNA which makes it unique from other RNA. This is especially helpful in separating it from **rRNA** which is much more abundant.
- TWO types are there: 1 and 2 (next slide……)

mRNA to cDNA



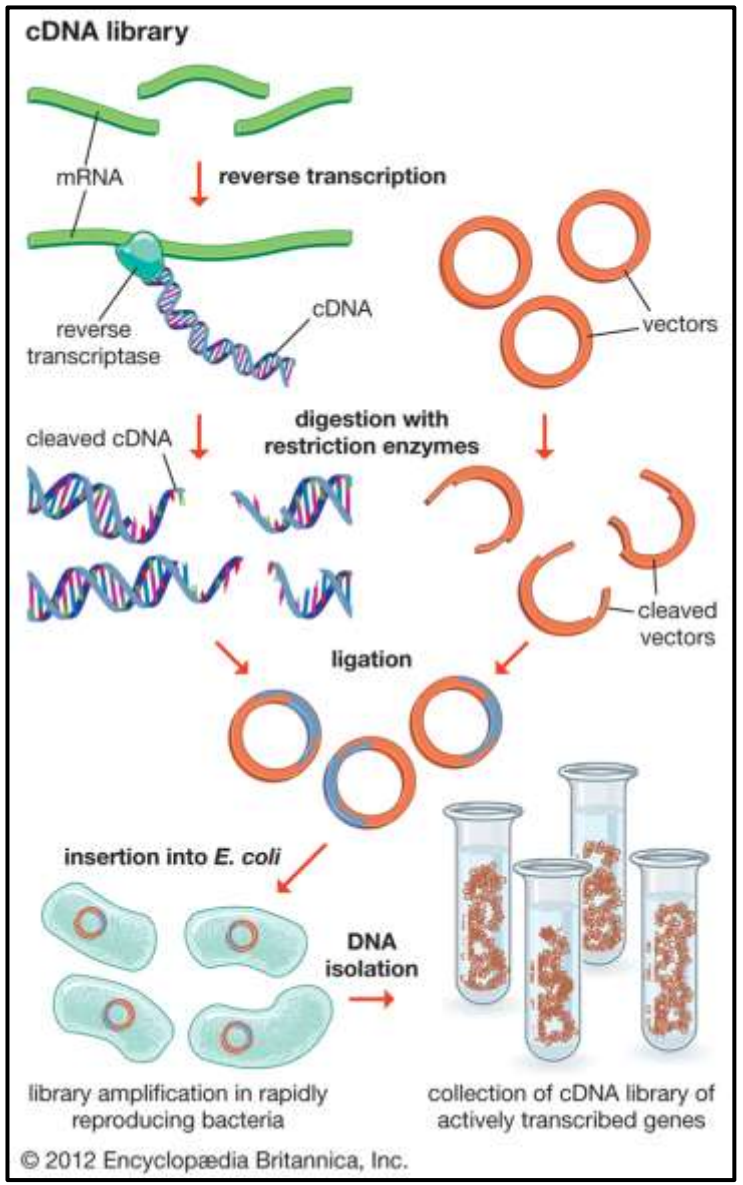
- Reverse transcriptases are used by viruses such as HIV and hepatitis B to replicate their genomes
- By retrotransposon mobile genetic elements to proliferate within the host genome
- By eukaryotic cells to extend the telomeres at the ends of their linear chromosomes.



Crystallographic structure of HIV-1 reverse transcriptase

Reverse transcriptase is commonly used in research to apply the polymerase chain reaction technique to RNA in a technique called reverse transcription polymerase chain reaction (RT-PCR)

Reverse transcriptase is used also to create cDNA libraries from mRNA



Oligo (dT) primers

- Typically, Oligo (dT) primers are a string of 12-20 deoxythymidines that anneal to poly(A) tails of eukaryotic mRNAs. These are ideal for constructing cDNA libraries and recommended with reverse transcriptases.

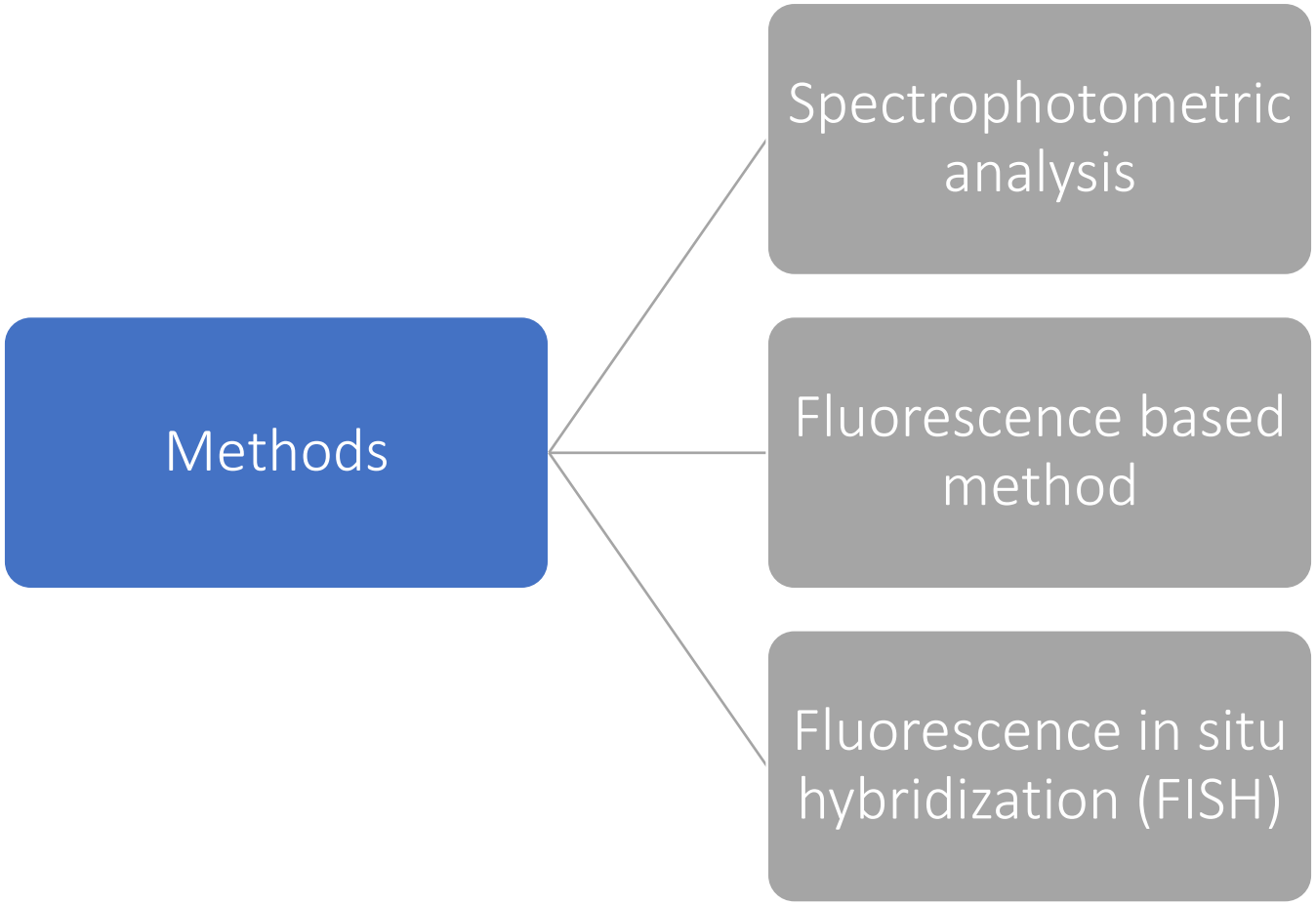
Random primers and hexamers

- These are short oligodeoxyribonucleotides with random base sequences (usually [d(N)6]) and are commonly referred to as random primers or hexamers. These are typically used to prime mRNAs with or without poly(A) for cDNA synthesis. These primers are suitable for DNA synthesis using Klenow fragments with DNA templates or for cDNA synthesis using reverse transcriptase with mRNA templates.

Sequencing primers

- Short RNA or DNA oligonucleotides are often used as primers in PCR reactions to amplify a specific target or to directly sequence a specific genetic region of interest (Sanger sequencing)

Quantitation and quality check of nucleic acids



DNA Purity and Yield

Using spectrophotometer:

DNA concentration - measuring absorbance at 260nm, adjusting the A_{260} measurement for turbidity (measured by absorbance at 320nm), multiplying by the dilution factor, and using the relationship that an A_{260} of 1.0 = 50µg/ml pure dsDNA.

A260 dsDNA of 1.0 = 50 µg/ml

A260 ssDNA of 1.0 = 33 µg/ml

A260 ssRNA of 1.0 = 40 µg/ml

Concentration (µg/ml) = (A_{260} – A_{320}) × dilution factor × 50µg/ml

Total yield is obtained by multiplying the DNA concentration by the final total purified sample volume.

DNA yield (µg) = DNA concentration × total sample volume (ml)

- However, DNA is not the only molecule that can absorb UV light at 260nm.
- RNA also has a great absorbance at 260nm.
- The aromatic amino acids present in protein absorb at 280nm, both contaminants, if present in the DNA solution, will contribute to the total measurement at 260nm.
- This means that if the A260 number is used for calculation of yield, the DNA quantity may be overestimated.

DNA Purity and Yield

To evaluate DNA purity, measure absorbance from 230nm to 320nm to detect other possible contaminants.

The most common purity calculation is the ratio of the absorbance at 260nm divided by the reading at 280nm.

Good-quality DNA will have an A260/A280 ratio of 1.7–2.0.

A reading of 1.6 does not render the DNA suitable for any application.

Lower ratios indicate more contaminants are present.

The ratio can be calculated after correcting for turbidity (absorbance at 320nm).

DNA purity (A260/A280) = (A260 – A320) ÷ (A280 – A320)

Strong absorbance around 230nm can indicate that organic compounds or chaotropic salts are present in the purified DNA.

A ratio of 260nm to 230nm can help evaluate the level of salt carryover in the purified DNA.

The lower the ratio, the greater the amount of thiocyanate salt is present, for example.

As a guideline, the A260/A230 is best if greater than 1.5.

A reading at 320nm will indicate if there is turbidity in the solution, another indication of possible contamination.

Therefore, taking a spectrum of readings from 230nm to 320nm is most informative.

- The concentration of unknown samples is calculated based on comparison to a standard curve generated from samples of known DNA concentration.
- Genomic, fragment and plasmid DNA will each require their own standard curves and these standard curves cannot be used interchangeably. As with absorbance methods, dilution factor must be taken into account when calculating DNA concentration from fluorescence values.
- Materials required for fluorescence methods are: a fluorescent DNA binding dye, a fluorometer to detect the dyes, and appropriate DNA standards. Depending on the dye selected, size qualifications may apply, and the limit of detection may vary.
- The usual caveats for handling fluorescent compounds also apply—photobleaching and quenching will affect the signal.

Spectrophotometric analysis

Purity

A_{260}/A_{280}

- 1.85-1.88 for pure dsDNA
- 2.1 for pure RNA
- Less than 1.8 (protein or phenol contamination)
- More than 1.88 for DNA (RNA contamination)

A_{260}/A_{230}

- 2.3–2.4 for pure ds DNA
- 2.1–2.3 for RNA
- Less than 1.8 (contamination by chaotropic salts such as guanidine thiocyanate (GTC) and guanidine hydrochloride (GuHCl), EDTA, non-ionic detergents like Triton™ X-100 and Tween® 20, proteins, and phenol)

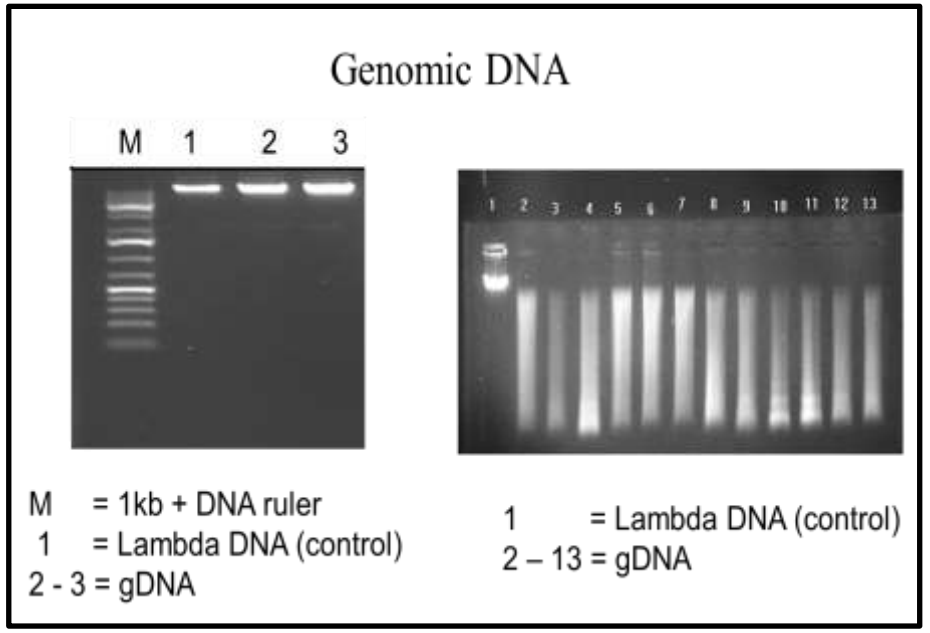
Fluorescence Methods

- The widespread availability of fluorometers and fluorescent DNA-binding dyes make fluorescence measurement another popular option for determining of DNA yield and concentration.
- Fluorescence methods are more sensitive than absorbance, particularly for low-concentration samples, and the use of DNA-binding dyes allows more specific measurement of DNA than spectrophotometric methods allows.
- Hoechst bisbenzimidazole dyes, PicoGreen and QuantiFluor dsDNA dyes selectively bind double-stranded DNA.
- The availability of single-tube and microplate fluorometers gives flexibility for reading samples in PCR tubes, cuvettes or multiwell plates and makes fluorescence measurement a convenient modern alternative to the more traditional absorbance methods.
- Fluorescence measurements are set at excitation and emission values that vary depending on the dye chosen.

- The concentration of unknown samples is calculated based on comparison to a standard curve generated from samples of known DNA concentration.
- Genomic, fragment and plasmid DNA will each require their own standard curves and these standard curves cannot be used interchangeably.
- As with absorbance methods, dilution factor must be taken into account when calculating DNA concentration from fluorescence values.
- Materials required for fluorescence methods are: a fluorescent DNA binding dye, a fluorometer to detect the dyes, and appropriate DNA standards. Depending on the dye selected, size qualifications may apply, and the limit of detection may vary.
- The usual caveats for handling fluorescent compounds also apply—photobleaching and quenching will affect the signal.

Agarose Gel Electrophoresis

- Agarose gel electrophoresis is another way to quickly estimate DNA concentration.
- DNA sample and an intercalating DNA dye along with appropriately sized DNA standards are required. A sample of the isolated DNA is loaded into a well of the agarose gel and then exposed to an electric field.
- Since small DNA fragments migrate faster, the DNA is separated by size. The percentage of agarose in the gel will determine what size range of DNA will be resolved with the greatest clarity.
- Concentration and yield can be determined after gel electrophoresis is completed by comparing the sample DNA intensity to that of a DNA quantitation standard.
- For example, if a 2µl sample of undiluted DNA loaded on the gel has the same approximate intensity as the 100ng standard, then the solution concentration is 50ng/µl (100ng divided by 2µl).
- Standards used for quantitation should be labeled as such and be the same size as the sample DNA being analyzed.
- In order to visualize the DNA in the agarose gel, staining with an intercalating dye such as ethidium bromide or SYBR Green is required.



Questions