Genome Editing and Engineering
Prof. Utpal Bora
Department of Bioscience and Bioengineering
Indian Institute of Technology, Guwahati

Module - 01
Introduction to genetics and genetic engineering
Lecture - 01
Introduction: Genes and Genome Organization

Welcome to my course on Genome Editing and Engineering. Today we are going to have a lecture on the Introduction of Genes and Genome Organization.

(Refer Slide Time: 00:48)

"My time will come,"

Father of Genetics
Mendel through his breeding experiments with pea proposed that there were 2 factors for each basic trait and that 1 factor was inherited from each parent.

These inheritable factors are what we know as genes today.

On February 8, 1865, Mendel presented his work to the Brunn Society for Natural Science and the next year his paper, "Experiments on Plant Hybridization," was published but was forgotten by the scientific community until rediscovered by DeVries, Correns and Tschermak independently in 1900.

Let us start with the work of Mendel, who is considered as the father of genetics. Mendel through his breeding experiments with pea proposed that there were 2 factors for each basic trait and that 1 factor was inherited from each parent. These inheritable factors are what we know today as genes.

On February 8, 19 1865 Mendel presented his work to the Brunn Society for Natural Science and the next year his paper, “Experiments on Plant Hybridization” was published. But was forgotten by the scientific community until rediscovered by DeVries Correns and Tschermak independently in 1900. Mendel died without his work being recognized, but he had the confidence that people will understand the value of his work in future.

(Refer Slide Time: 01:47)

The word "gene" was coined much later in the early 20th century, by the Danish botanist Wilhelm Ludvig Johannsen (1909).

It rapidly became fundamental to the then new science of genetics, and Mendel's factors were started to be called as genes.

Johannsen also introduced the concepts of phenotype and genotype.

Wilhelm Ludvig Johannsen,
Wikimedia commons.
Lisens: Falt i det fri (Public domain)

M1L1 3

The term "gene" was, however, coined much later in the early 20th century by the Danish botanist Wilhelm Ludvig Johannsen in 1909. It rapidly became fundamental to the then new science of genetics, and Mendel's factors were started to be called as genes. Interestingly Johannsen also introduced the concept of phenotype and genotype.

(Refer Slide Time: 02:14)

Etymology
From the Greek "gen" (gamma, epsilon, nu), gene (gamma, epsilon, nu, epsilon) = create/creation, birth,

Gene

Gene is the basic unit of heredity passed from parent to child. Genes are made up of sequences of DNA and are arranged, one after another, at specific locations on chromosomes in the nucleus of cells.

They contain information for making specific proteins that lead to the expression of a particular physical characteristic or trait, such as hair color or eye color, or to a particular function in a cell.

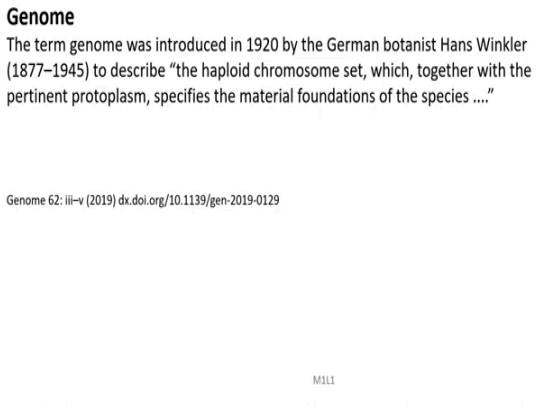
Adapted from National Cancer Institute
https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/gene

M1L1 4

The origin of the word gene is in the Greek word 'gen' which means to create or creation or birth. The National Cancer Institute defines gene as follows. Gene is a basic unit of heredity passed from parent to child. Genes are made up of the sequences of DNA and are arranged

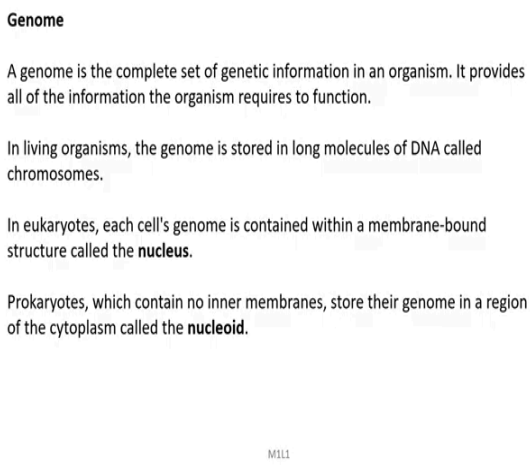
one after another at specific locations on chromosomes in the nucleus of cells. They contain information for making specific proteins that lead to the expression of a particular physical characteristics or trait, such as hair color or eye color or to a particular function in a cell.

(Refer Slide Time: 03:08)



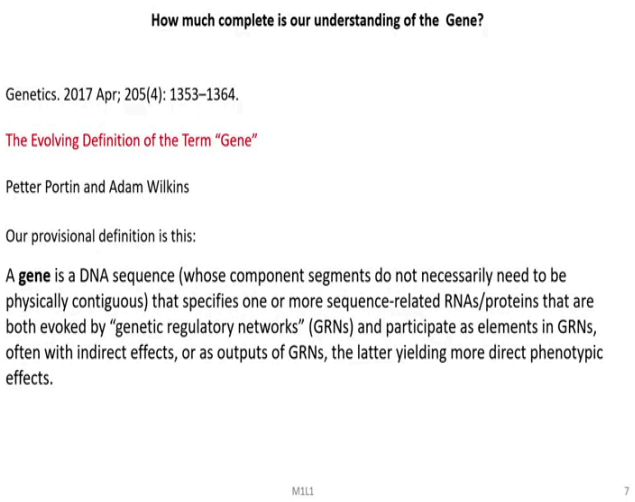
Another term that is associated with heredity and genes is genome. What is genome? The term genome was introduced in 1920 by the German botanist Hans Winkler to describe the haploid chromosome set which together with the pertinent protoplasm specifies the material foundations of the species. We humans are deploid in nature. The half set of the diploid chromosomes are what we call as the haploid chromosome set.

(Refer Slide Time: 03:48)



This genome is the complete set of genetic information in an organism. It provides all of the information the organism requires to function. In living organisms, the genome is stored in long molecules of DNA and we call them as chromosomes. In eukaryotes, each cells genome is contained within a membrane bound structure called the nucleus. However, the prokaryotes which do not contain any inner membranes, they store their genome in reason of the cytoplasm called as the nucleoid.

(Refer Slide Time: 04:32)

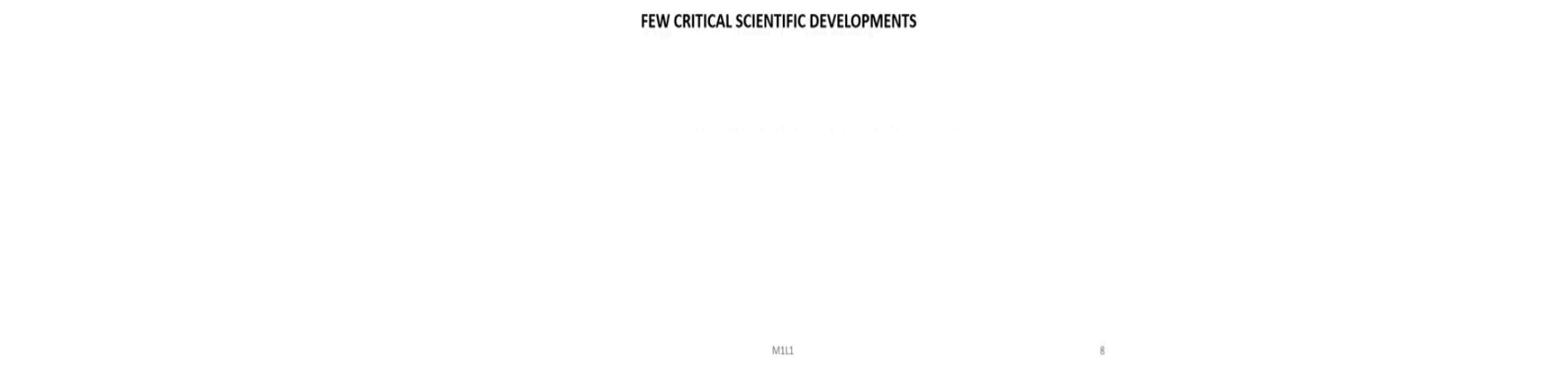


With so many advances since the days of Mendel, how much do we actually know about the gene? Although there have been many discoveries, inventions and advances in the understanding of the gene. We still have many things unexplored regarding the nature of the gene and this paper by Petter Portin Adam Wilkins deals with an interesting topic the evolving definition of the term gene.

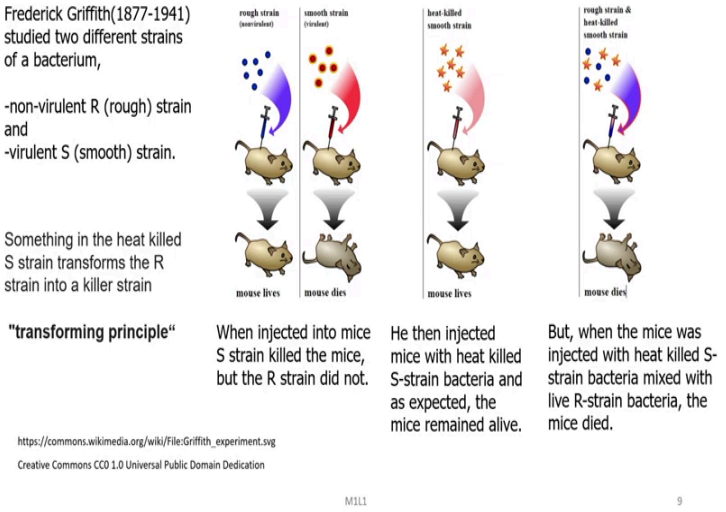
So, our concept of the gene has been changing from time immemorial and for this reason Portin and Wilkins provided a provisional definition of gene. According to them, a gene is a DNA sequence whose component segments do not necessarily need to be physically contiguous. This is very very important point. The genes may be split they may a single gene may not stay together as a piece of information.

This information can remain in various space. And it specifies one or more sequence related RNAs or proteins that are both evoked by genetic regulatory networks and participate as elements in GRNs. Often with indirect effects or as outputs of GRNs, the letter yielding more direct phenotypic effect.

(Refer Slide Time: 06:15)



(Refer Slide Time: 06:24)



Let us now discuss about some of the important scientific developments in the field of genetics. Let us discuss about first the work of Frederick Griffith, who studied two different strains of a bacterium, a non virulent strain and a virulent strain. The non virulent strain called the R strain or the rough strain did not cause any disease while the virulent strain called the S strain causes disease in mice.

So, when injected into mice the S strain kill the mice, but the R strain did not kill the mice. This was a very simple experiment, but very interesting. Later on he injected mice with heat killed S strain bacteria. Since, the virulent strain as is now heat killed as expected it could not cause disease and kill the mice.

So, the mice was alive at the end of the experiment. What he did after this was something very very innovative. He injected the mice with heat killed S strain bacteria which cannot kill the mice, but then along with it he mixed the alive R strained bacteria which is harmless. So, both things were supposed to save the mice, not kill it. But the result was something very very interesting, the mice got killed.

So, there must be something in the heat killed S strain which transforms the R strain into a killer strain. That was the observation by Frederick Griffith and he named these as the transforming principle.

(Refer Slide Time: 08:25)

Maclyn McCarty with Francis Crick and James D Watson -

10.1371 journal.pbio.0030341.g001-O.jpg
Licensed under the Creative Commons Attribution 2.5 Generic

1944: DNA is "Transforming Principle"

Avery, MacLeod and McCarty identified DNA as the "transforming principle" postulated by Griffith, while studying *Streptococcus pneumoniae*, bacteria that can cause pneumonia.

M1L1 10

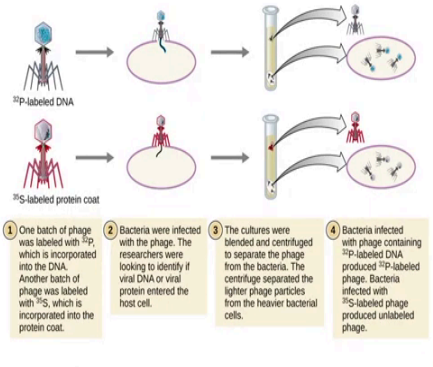
This interesting work let made other scientists to verify what these transforming principle is? what is it is nature? So, in 1944 Avery MacLeod and McCarty identified that DNA is the transforming principle as postulated by Griffith and they established these with the study of streptococcus pneumonia that causes bacteria… in this picture you can see Maclyn Maccarty with Francis Crick and James D Watson about whom we will discuss in the next few slides

(Refer Slide Time: 09:07)

1951-1952,
Alfred Hershey and Martha Chase

It was however the Hershey-Chase experiment, called the Waring Blender experiment, which provided the concrete evidence that genes were made of DNA.

Hershey shared the 1969 Nobel Prize in Physiology or Medicine with Max Delbrück and Salvador Luria for their "discoveries concerning the genetic structure of viruses".



Images/ https://cnx.org/contents/5CvTdmJL@4.4

Image of Hershey and Chase by Rohit Kumar Sengupta licensed under the Creative Commons Attribution-Share Alike 4.0 International

M1L1 11

This was followed by other interesting experiments. For example, the experiments by Hershey and Chase called the Waring Blender experiment provided the concrete evidence

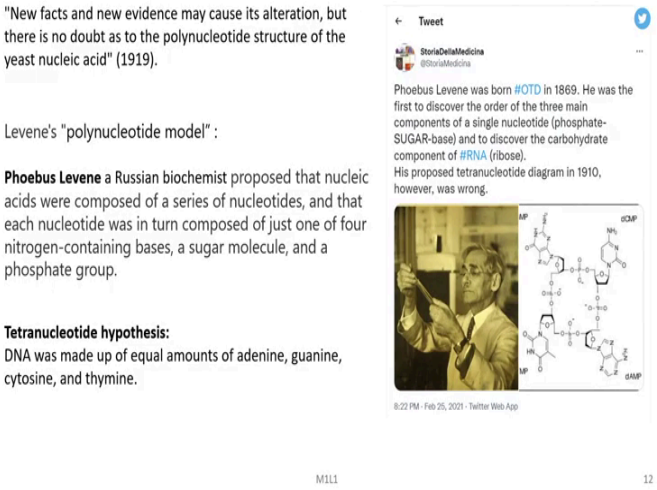
that genes were made up of DNA. For this work in 1969 Hershey shared the Nobel prize in Physiology or Medicine with Max Delbruck and Luria for their discoveries concerning the genetic structure of viruses. Now if you look into this diagram, you can see the work carried out by Hershey and Chase.

So, in step one they labeled the DNA with 32P and another batch was labeled with 32S. So, 32P labels the DNA while 35S sulfur labels the protein coat. So, one match of the phage was level with 32P which is incorporated into the DNA. While another batch of phage was labeled with 32S which is incorporated into the protein coat.

Next the bacteria were infected with the labeled phage and they look to identify if the viral DNA or viral protein enters the host cell. The cultures were blended and centrifuged to separate the phage from the bacteria. The centrifuge separated the lighter phage particles from the heavier bacterial cells as you can see in this picture.

The bacteria infected with phage containing 32P labelled DNA produced 32P labeled phage. While the bacteria infected with 35S labeled phage produced unlabeled phage which means that DNA is the material that was inherited into the next progeny and not protein. So, these firmly established that the genes are made up of DNA.

(Refer Slide Time: 11:38)



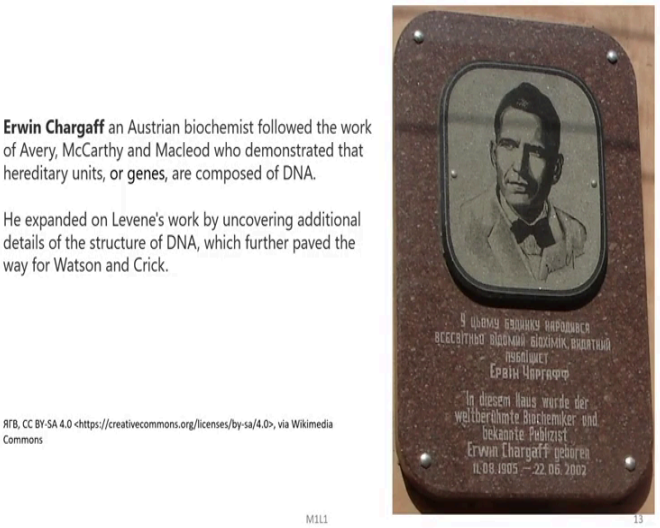
It was much later, Levene came into the picture and he put forward a famous model called the polynucleotide model and he in fact, told with high confidence that new facts and new

evidence may cause its alteration, but there is no doubt as to the polynucleotide structure of the yeast nucleic acid.

Phoebus Levene was a Russian biochemist, who proposed that nucleic acids were composed of a series of nucleotides and that each nucleotide was in turn composed of just one of four nitrogen containing bases, a sugar molecule and a phosphate group.

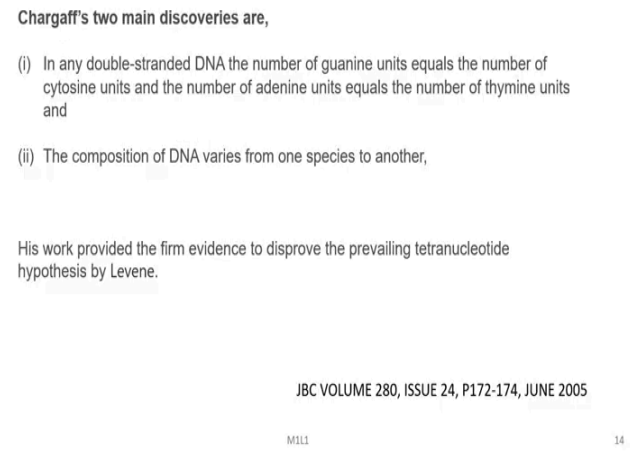
In fact, he was the first to discover the order of the three main components of a single nucleotide and to discovered a carbohydrate component of RNA. Another of his work and hypothesis known as the tetranucleotide hypothesis; however, turned out to be wrong where he stated that DNA was made up of equal amounts of adenine, guanine, cytosine and thymine. It is important to know about the tetranucleotide hypothesis at this stage because soon we will know the exact composition of a DNA molecule by the work of Erwin Chargaff.

(Refer Slide Time: 13:10)



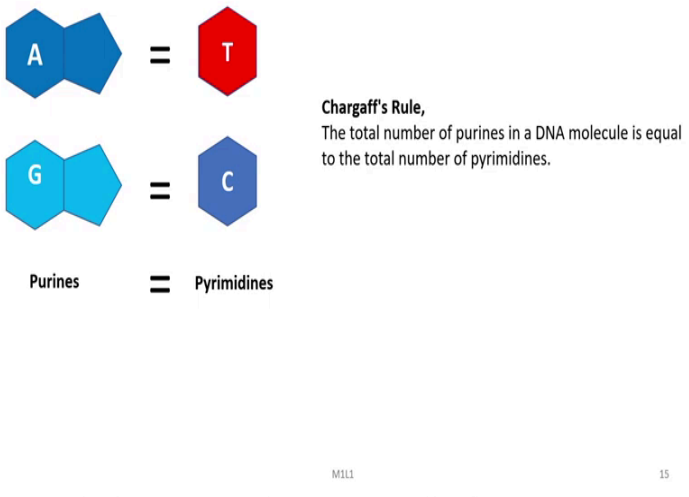
Erwin Chargaff was an Austrian biochemist who followed the work of Avery, Mccarty and Macleod and he demonstrated that hereditary units or genes are composed of DNA. He expanded on Levenes work by uncovering additional details of the structure of DNA which further paved the way for Watson and Crick model.

(Refer Slide Time: 13:41)



The two main discoveries done by Chargaff are; 1, In any double stranded DNA the number of guanine units equals the number of cytosine units and a number of adenine units equals the number of thymine units and the composition of DNA varies from one species to another. Chargaff's work provided the firm evidence to disprove the prevailing tetranucleotide hypothesis by Levene, which we discussed prior to this slide.

(Refer Slide Time: 14:15)



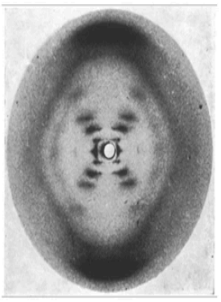
Now, what is the Chargaff's Rule? As you can see here on the left side are the purines and on the right side are the pyramidines and from the figure you can see purines A is equal to T and

G is equal to C. So, Chargaff's rules is quite simple, which states that the total number of purines in a DNA molecule is equal to the total number of pyramidines. This is one of the fundamental principles we have to remember whenever we study about the structure and function of DNA molecule.

(Refer Slide Time: 14:49)

Let us now focus on the work by Rosalind Franklin and others. James Watson, Francis Crick, Maurice Wilkins and Rosalind Franklin, they all played a critical role in discovering the structure of DNA. So, this is one of the most famous photographs in molecular biology. This is an X ray diffraction photo of DNA, which is taken by Wilkins and Franklin and this served as a key line of evidence in figuring out the structure of DNA.

(Refer Slide Time: 15:36)



DNA
Av Rosalind Franklin.
Lisens: Falt i det fri (Public domain)
https://snl.no/Rosalind_Franklin

An X-ray diffraction photo of DNA taken by Wilkins and Franklin which served as a key line of evidence in figuring out the structure of DNA. The x-shaped pattern in the image strongly suggested a helical form and other details of the structure.

M1L1 17

You can see here the x-shaped pattern in the image which strongly suggested a helical form and other details of the structure. This picture of DNA was taken by Rosalind Franklin. Let us discuss about the Watson and Crick model based on the X-ray photo taken by Rosalind and Wilkins.

(Refer Slide Time: 16:07)

Watson and Crick Model

Deoxyribonucleic Acid is a double-stranded, helical molecule consisting of two sugar-phosphate backbones on the outside, held together by hydrogen bonds between pairs of nitrogenous bases on the inside.

The bases are of four types: adenine (A), cytosine (C), guanine (G), and thymine (T)

Pairing always occurs between A & T, and C & G.

Thus either strand contained all the information necessary to make a new copy of the entire molecule.

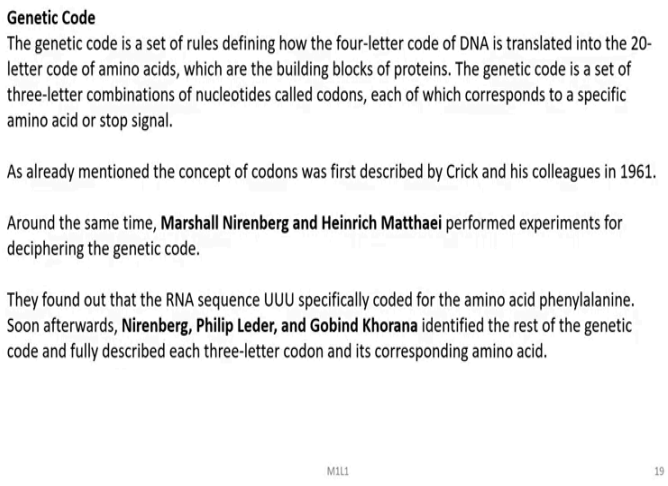
Watson and Crick figured that the aperiodic order of bases might provide a "genetic code".

M1L1 18

According to Watson and Crick model, deoxyribonucleic acid is a double stranded, helical molecule consisting of two sugar phosphate backbones on the outside which are held together by hydrogen bonds between pairs of nitrogenous bases on the inside. The bases are of four

types; adenine, cytosine, guanine and thymine. Pairing occurs always between adenine and thymine and cytosine and guanine. Thus either stand contain all the information necessary to make a new copy of the entire molecule. And they figured out that the aperiodic order of the basis might provide a genetic code which took a couple of years to be proven.

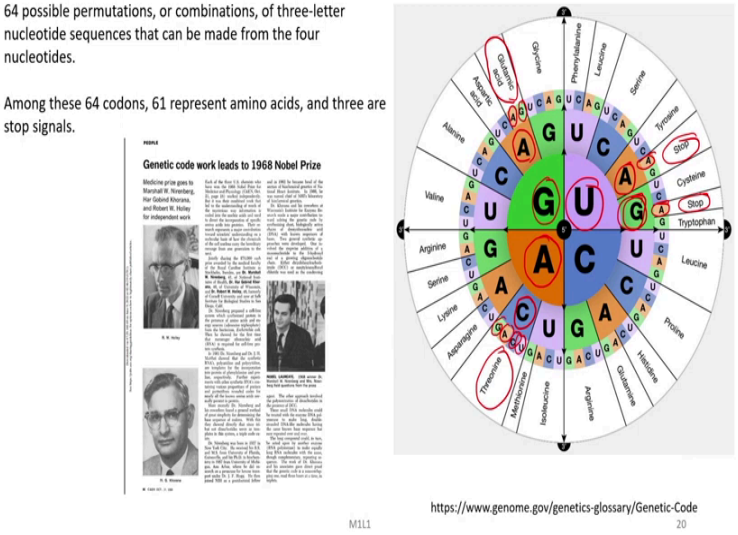
(Refer Slide Time: 17:06)



Now, what is this generic code? The genetic code is a set of rules defining how the four letter code of DNA is translated into the 20 letter code of amino acids, which are building blocks of proteins. The generic code is a set of three letter combinations of nucleotides which we call as codons, each of which correspond to a specific amino acid or a stop signal.

As already mentioned earlier, the concept of codons was first described by Crick and his colleagues in 1961. Around the same time Marshall Nirenberg and Heinrich Matthaei performed experiments for deciphering the generic code. They found out that the RNA sequence UUU specifically coded for the amino acid phenylalanine. And soon afterwards Nirenberg, Philip Leder and Har Gobind Khorana identified the rest of the generic code and fully described each three letter codon and its corresponding amino acid.

(Refer Slide Time: 18:27)



The 64 possible permutations or combinations of a three letter nucleotide sequence can be made from the four nucleotides. And you can see in this diagram, circular diagram, if you for example, the first letter is G followed by A then G, it will give represent glutamic acid. Again if the first letter is G, second letter is A and the third letter is A, that will also represent glutamic acid, which means there is some kind of redundancy in the codon dictionary. This is very very important to understand that one amino acid may have more than one codon.

For example here Threonine can have one, two, three four different kind of codons, but what is interesting to observe over here; the first letter would be always A second letter would be always C, but the third letter there is a variation almost all the four bases are present over there.

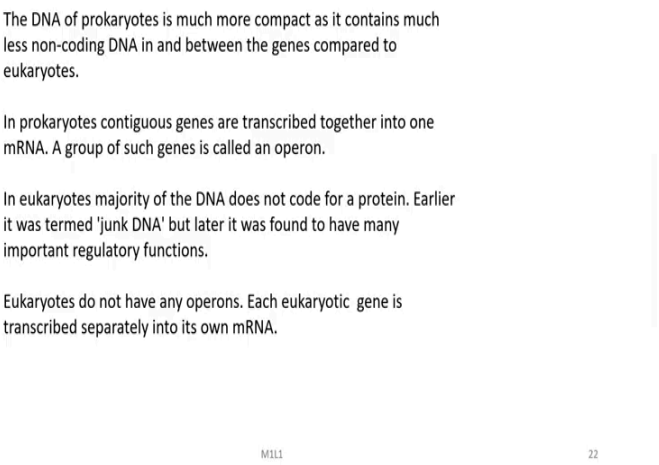
So, this is the redundancy in the codon dictionary. Now there are other codons which do not represent any amino acids. For example, if the first letter is G second, sorry. The first letter is U second letter is G and the third letter is A, it will signal a stop signal or a stop codon, it do not represent any kind of an amino acid. Similarly UAA also represents a stop codon or a null codon which do not represent any amino acids.

Now for this work on the genetic code, the Nobel prize is awarded in 1988 1968 to Nirenberg, Hargabin Khorana and Holley for their independent establishment of this codon dictionary.

(Refer Slide Time: 21:00)

With this basic ideas about the genetic material now let us try to understand how this genetic material is organized inside a living cell. So, we discuss about the genome organization now.

(Refer Slide Time: 21:22)

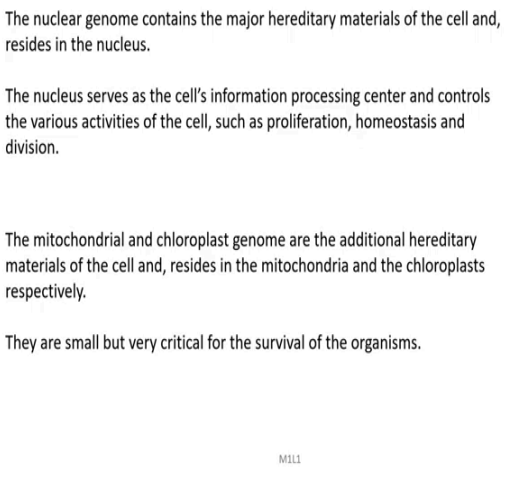


The DNA of prokaryotes is much more compact as it contains much less non coding DNA in and between genes compared to eukaryotes. It need to be mentioned over here that higher organisms have very large genomes, but only a fraction of these genome code for proteins or RNA or tRNA which means only a fraction of these proteins contain codons or sequences for tRNAs, rRNAs.

The remaining large majority of the sequences are known as junk DNA and earlier they were thought to be of no use, but over the years it has been found out the junk DNA are equally important for the function and health of the organism.

So, in prokaryotes the genes are contiguous and they are transcribed together in one mRNA. A group of such genes are called as operons. In eukaryotes a majority of the DNA does not code for the protein as already discussed and this was known as junk DNA and recently many functions have started emerging and many of this play regulatory role. Eukaryotes do not have any operons. Each eukaryotic gene is transcribed separately into its own mRNA.

(Refer Slide Time: 23:16)



The nuclear genome contains the major hereditary material of the cell and resides in the nucleus. This is what happens in case of eukaryotes. The nucleus serves as the cells information a processing center and it controls the various activities of the cell such as proliferation, homeostasis and division. In addition to these eukaryotic cells have other compartments where the genetic information is stored and these are the mitochondrial DNA and the chloroplast DNA.

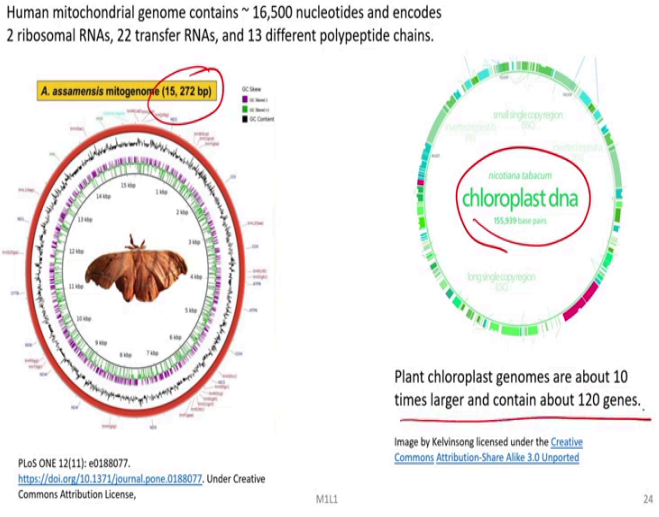
They constitute the mitochondrial and chloroplast genome respectively and these are additional hereditary materials of the cell and as I already told you they resides in the mitochondria and the chloroplast respectively. It need to be mentioned that chloroplasts are available in plants only.

So, organisms like humans have two important genomes; one is the nuclear genome and the other is the mitochondrial genome. While plants have three important genomes; the first one is the nuclear genome second one is in the mitochondrial genome and third one is the chloroplast genome.

The mitochondrial genome and the chloroplast genome are very small compared to the nuclear genome, but they are very very critical for the survival of the organisms. The mitochondria is considered the powerhouse of all organisms, without it the organism cannot survive.

Similarly the chloroplast fix carbon dioxide and synthesize sugar by harvesting solar energy. Without the presence of the chloroplast genome, these whole creation actually will fall down. So, in spite of their very tiny size they are very very critical in the survival of the organism and in fact, the entire ecosystem.

(Refer Slide Time: 25:38)

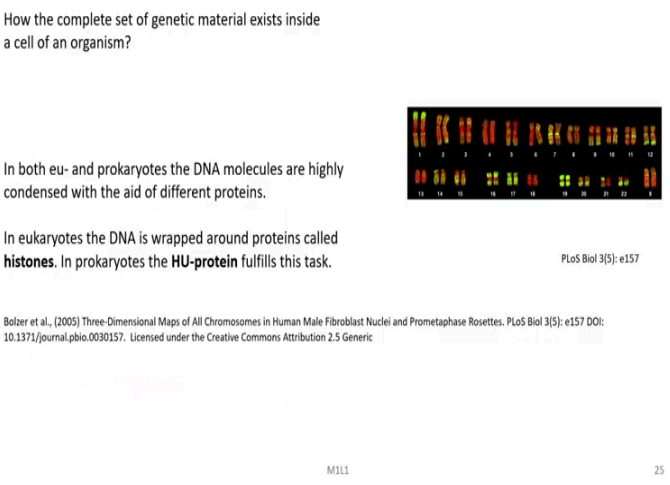


Now if we look into the human mitochondrial genome, it contains roughly around 16500 nucleotides and it encodes 2 ribosomal RNAs, 22 transfer RNAs and 13 different polypeptide chains. This is the map of a mitochondrial genome of antheraea assamensis.

As you can see on the left side and these contains around 15272 nucleotides and it encodes similar number of ribosomal RNAs transfer RNAs and various polypeptide chains. If you look into the right side we can see the chloroplast DNA of nicotiana tobacum which contains

around 155.939 base pairs. So, these plant chloroplast genomes are about 10 times larger and contains about 120 genes.

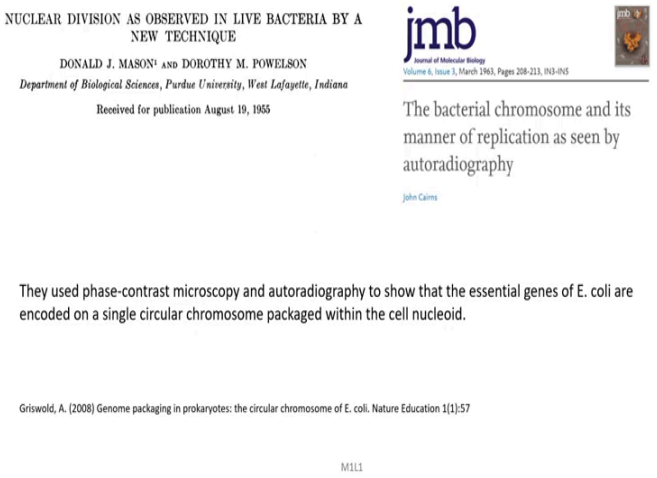
(Refer Slide Time: 26:54)



How the complete set of genetic material exist inside the cell of an organism is the next topic that we are going to discuss. We now know the genome is not kept in one place, the larger part of the genome is kept in the nucleus while there are smaller parts which may be kept in mitochondria and another organelle called as the chloroplast. Now how these material exist? For example, in the nucleus. In both eukaryotes and prokaryotes the DNA molecules are highly condensed with the aid of different proteins. In eukaryotes, the DNA is wrapped around proteins called histones.

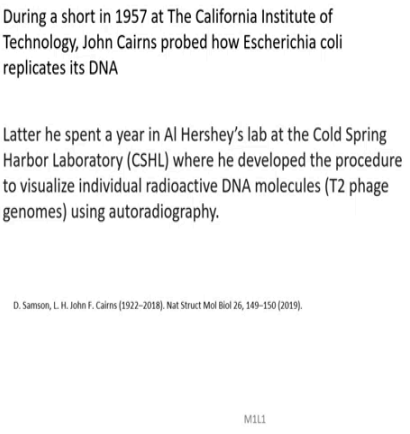
In prokaryotes, the place of histones is taken up by hue proteins in a similar fashion.

(Refer Slide Time: 27:59)



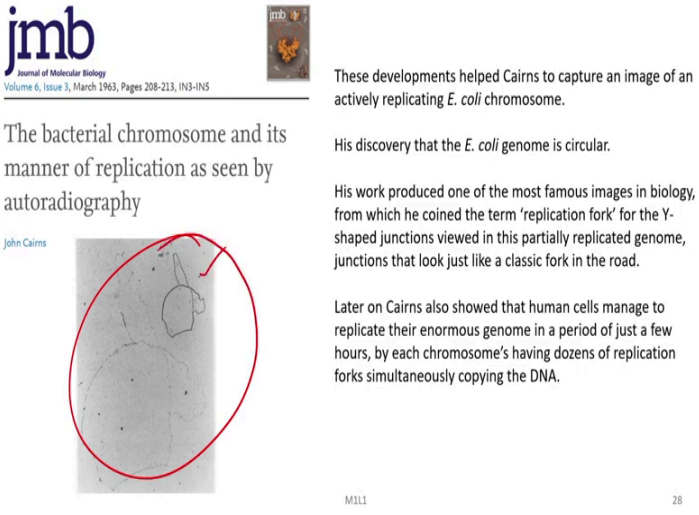
These are the two important papers regarding the nuclear division as observed in live bacteria by new technique by Mason and M. Powelson. And another paper by John Cairns, which spoke about the bacterial chromosome in its manner of replication as seen by autoradiography. Both these groups used phase contrast microscopy and auto radiography to show that the essential genes of Escherichia coli are encoded on a single circular chromosome packaged within the cell nucleoid.

(Refer Slide Time: 28:47)



During a short break in 1957 at the California Institute of Technology John Cairns probed how Escherichia coli replicates it's DNA. Latter he spent another year in Hershey's lab at the cold spring harbor laboratory, where he developed the procedure to visualize the individual radioactive DNA molecules of T2 phage genomes using auto radiography.

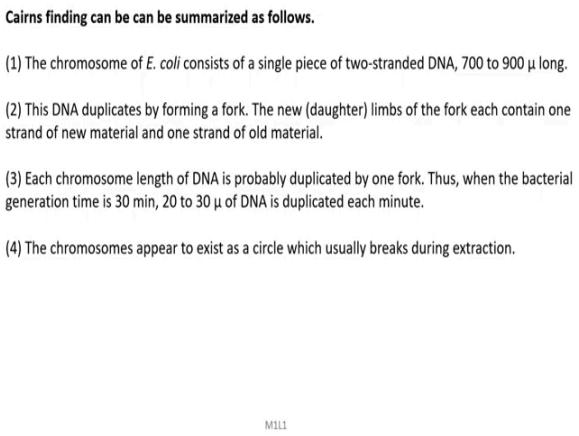
(Refer Slide Time: 29:19)



This is the bacterial chromosome as observed by John Cairns and it is one of the very famous photographs in biology. His work in the two laboratories helped him to capture an image of an actively replicating E. coli chromosome as you can see. These developments help cairns to capture an image of an actively replicating E. coli chromosome. He discovered that the E. coli genome is circular.

His work produced one of the most famous images in biology, from which it term the coin a coined term replication fork for the Y shaped junctions viewed in this partially replicated genome. Later on cairns also showed that human cells manage to replicate their enormous genome in a period of just a few hours by each chromosomes having dozens of replication forks simultaneously copying the DNA. So, this is the famous Y fork as you can see over here in this image.

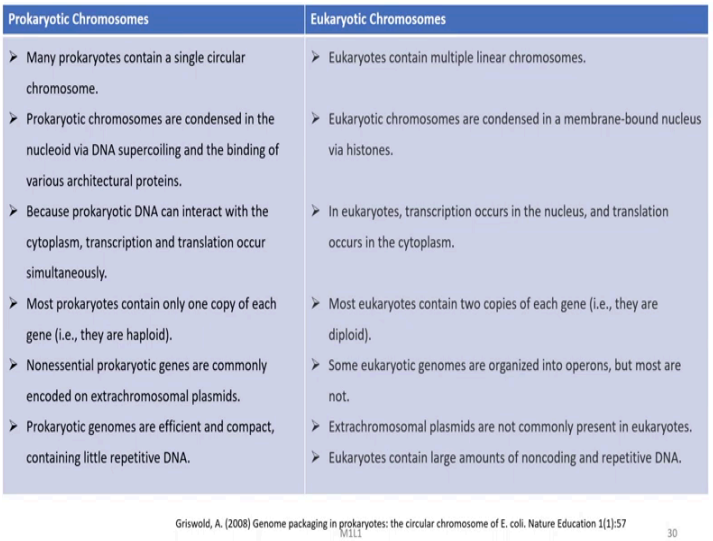
(Refer Slide Time: 30:33)



Overall, Cairns findings can be summarized as follows. Number 1, the chromosome of E. coli consist of a single piece of two stranded DNA, which are 700 to 900 microns long. Number 2, this DNA duplicates by forming a fork. The new daughter, limbs of the fork each contain one strand of new material and one strand of the old DNA material.

Thirdly each chromosome length of DNA is probably duplicated by one fork. Thus when the bacterial generation time is 30 minute, 20 to 30 microns of DNA is duplicated each minute. Lastly, the chromosome appears to exist as a circle which usually breaks during extraction.

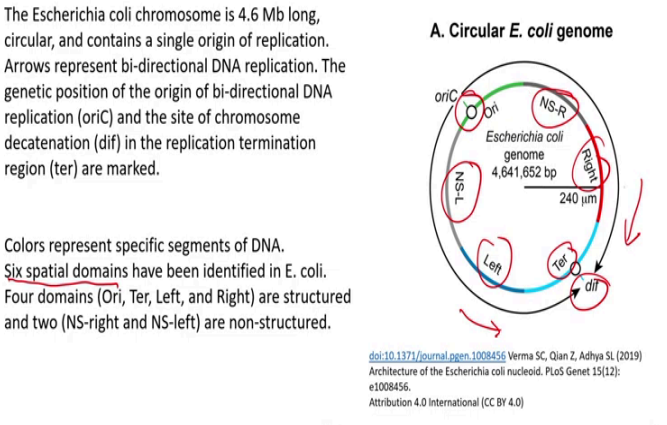
(Refer Slide Time: 31:40)



Now, let us compare the various features between prokaryotic chromosomes and eukaryotic chromosomes. Many prokaryotes contain a single circular chromosome, while eukaryotes contain multiple linear chromosomes. Prokaryotic chromosomes are condensed in the nucleoid via DNA supercoiling and the binding of various architectural proteins. Eukaryotic chromosomes are condensed in a membrane bound nucleus via histones. Because prokaryotic DNA can interact with the cytoplasm, transcription and translation occur simultaneously; however, in eukaryotes this cannot happen.

The transcription occurs in the nucleus and the translation occurs in the cytoplasm. Most prokaryotes contain only one copy of each gene that is they are haploid, but most eukaryotes contain two copies of each gene that is they are deployed. Non-essential prokaryotic genes are commonly encoded on extrachromosomal plasmids. Some eukaryotic genomes are organized into operons, but most are not. Prokaryotic genomes are efficient and compact containing little repetitive DNA. Extrachromosomal plasmids are not commonly present in eukaryotes and eukaryotes contain large amounts of non-coding and repetitive DNA.

(Refer Slide Time: 33:36)

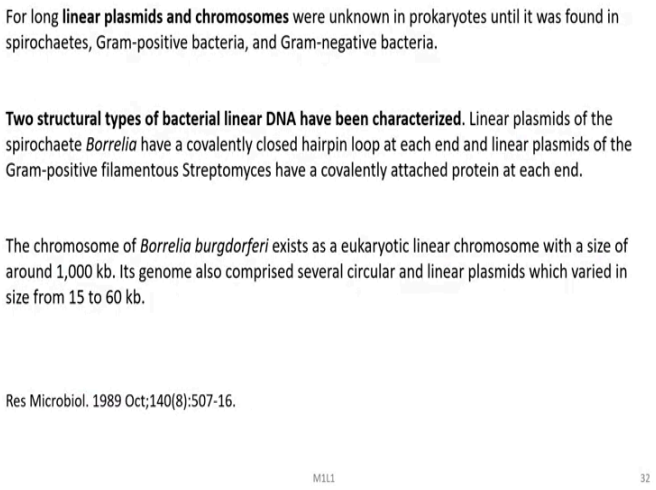


Let us now look into the chromosome of the prokaryotic organism Escherichia coli. The Escherichia coli chromosome is 4.6 megabases long. It is circular in shape and contains a single origin of replication. The arrows represent the bidirectional replication of DNA. The genetic position of the origin of bidirectional DNA replication oriC and the site of

chromosome decatenation dif in this replication termination region Ter are shown in this figure.

These colors represent specific segment of DNA. There is six special domains which have been identified in E. coli. The four domains Ori, Ter, Left and Right are structured and the two NS-right and NS-left are non-structured. So, these are some of the important features of the E. coli chromosome.

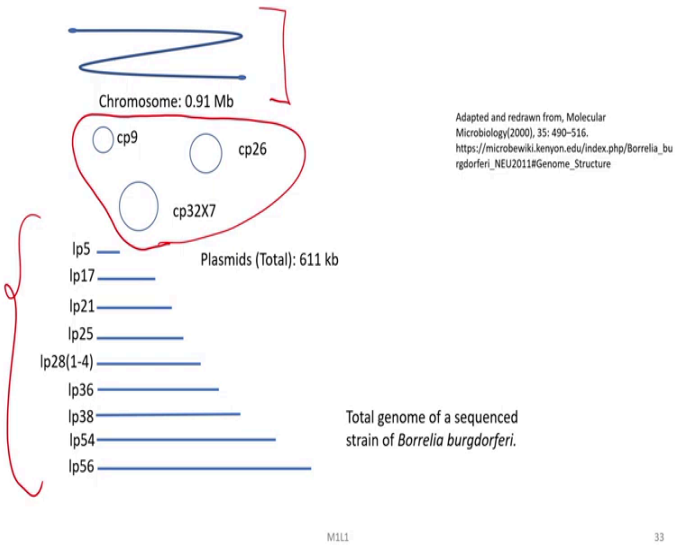
(Refer Slide Time: 35:05)



For long linear plasmids and chromosomes were unknown in prokaryotes and it was believed that all their genetic material are arranged in a circular chromosome. But later on it was found in spirochetes, Gram-positive-bacteria and Gram-negative–bacteria, that plasmids could be linear as well. Two structural types of bacterial linear DNA has been characterized. The linear plasmids of the spirocheate Borrelia have a covalently closed hairpin loop at each end.

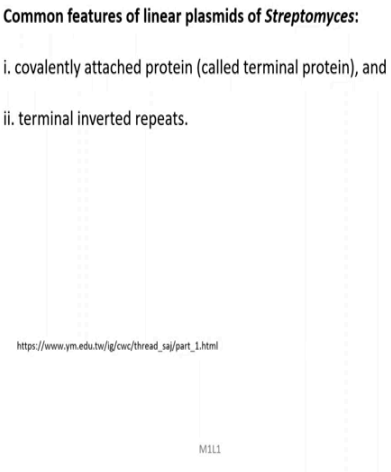
And a linear plasmid of the Gram-positive filamentous Streptomyces have a covalently attached protein at each end. The chromosome of Borrelia exists as a eukaryotic linear chromosome with a size of around 1000 kb. Its genomes also comprised of several circular and linear plasmids which varied in size from 15 to 1600 kb.

(Refer Slide Time: 36:22)



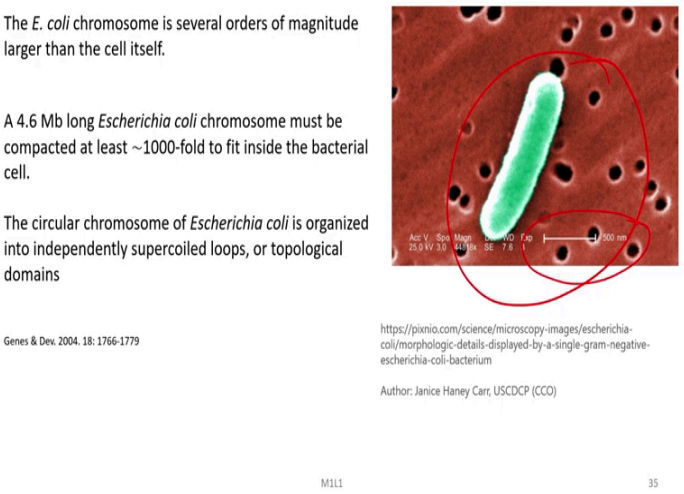
So, this is the map or of the total genome of Borrelia. You can see here a linear chromosome, which is around 0.91 mb, then you can see several other linear plasmids over here with variable sizes. And then apart from these, it also contains some circular small genomes. So, overall the total genome of Borrelia is very very interesting having linear plasmids a linear chromosome and circular plasmid.

(Refer Slide Time: 37:12)



Similarly, the common features of linear plasmid of streptomyces are it has a covalently attached protein called terminal protein and their terminal inverted repeats.

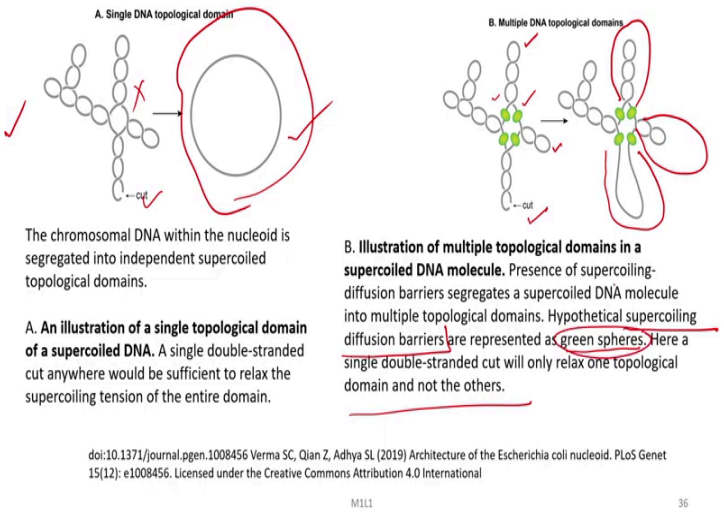
(Refer Slide Time: 37:21)



Now, let us go into the features of the E. coli chromosome once more. The E. coli chromosome is several orders of magnitude larger than the cell itself. You can see here in this figure, the shape of a E.coli and here is the scale and you can make some comparison over here, the length and breadth of this particular organism. Now into this small sized E. coli, a 4.6 m b long E. coli chromosome must be compacted and packaged.

So, this is done by compacting it at least 1000 fold to fit inside the bacterial cell. The circular chromosome of E. coli is organized into independently supercoiled loop or topological domains in order to fit into this small sized organism. Next we are going to discuss about this organization of supercoiled loops and topological domains.

(Refer Slide Time: 38:38)



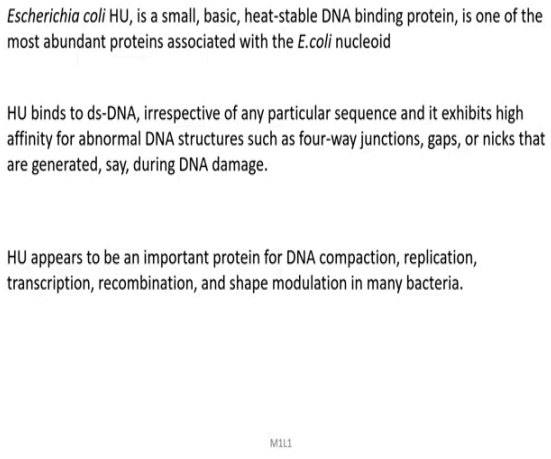
So, you can see here on the left side the single DNA topological domain. Here the chromosomal DNA within the nucleoid is segregated into independent supercoiled topological domains. And you can see from this illustration of a single topological domain of a supercoiled DNA. A single double stranded cut anywhere would be sufficient to relax the supercoiling tension of the entire domain and transform it into a open structure like this. However, in figure B we can see a multiple DNA topological domains.

And the difference between the two figures is actually in the presence of certain factors which are colored green over here as you can see which are absent in figure A. We will discuss what this green elements are.

So, this is an illustration of multiple topological domain in a supercoiled DNA molecule. Here the presence of supercoiling diffusion barriers segregates a supercoiled DNA molecule into multiple topological domains unlike in the first case. Hypothetical supercoiling diffusion barriers are represented as the green spheres which are speaking of.

Here a single double stranded cut will only relax one topological domain and not the other. So, here due to this cut we get only relaxation in one particular domain. So, if there would be a cut in the next domain here we will have another relaxation over here and similarly another over here if there is a cut in the third topological domain. So, I think this discussion makes clear the various multiple DNA topological domains and the supercoiling diffusion barriers offered by the green spheres in this structures.

(Refer Slide Time: 41:14)

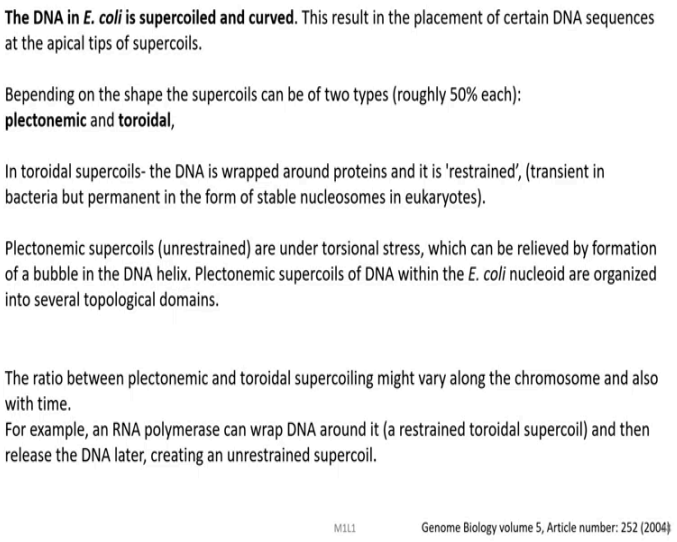


E. coli HU protein is a small basic heat stable DNA binding protein and is one of the most abundant proteins associated with the E. coli nucleoid. HU binds to double stranded DNA irrespective of any particular sequence. It is not a sequence specific binder. And it exhibits high affinity of abnormal DNA structures such as four way junctions, gaps or nicks that generated for example, during DNA damage.

HU appears to be an important protein for DNA compaction, replication, transcription, recombination and shape modulation in many bacteria. The DNA in E. coli is supercoiled and carved and we have seen this in our earlier discussion. This result in the placement of certain DNA sequences at the apical tips of supercoils.

Depending on the shape the supercoils can be of two types; plectonemic and toroidal. They are present in equal amounts.

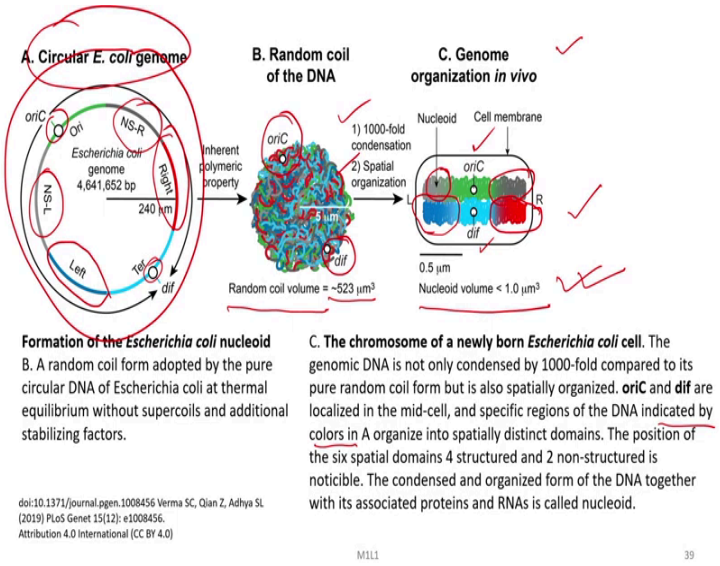
(Refer Slide Time: 42:53)



In toroidals supercoils, the DNA is wrapped around proteins and it is restrained, transient in bacteria, but permanent in the form of stable nucleosomes in eukaryotes. Plectonemic supercoils which are unrestrained are under torsional stress which can be relieved by formation of a bubble in the DNA helix. Plectonemic supercoils of DNA within the E. coli nucleoid are organized into several topological domains.

The ratio between plectonemic and toroidal supercoiling might vary along the chromosomes and also with time although this may be ideally represent in equal amounts. For example, an RNA polymerase can wrap DNA around it, a restrained toroidal supercoil and then release the DNA letter creating an unrestrained supercoil.

(Refer Slide Time: 43:50)



Formation of the *Escherichia coli* nucleoid
B. A random coil form adopted by the pure circular DNA of Escherichia coli at thermal equilibrium without supercoils and additional stabilizing factors.

C. **The chromosome of a newly born *Escherichia coli* cell**. The genomic DNA is not only condensed by 1000-fold compared to its pure random coil form but is also spatially organized. **oriC** and **dif** are localized in the mid-cell, and specific regions of the DNA indicated by colors in A organize into spatially distinct domains. The position of the six spatial domains 4 structured and 2 non-structured is noticible. The condensed and organized form of the DNA together with its associated proteins and RNAs is called nucleoid.

doi:10.1371/journal.pgen.1008456 Verma SC, Qian Z, Adhya SL (2019) PLoS Genet 15(12): e1008456.

So, in figure A, we can see the circular E. coli genome and we studied about it is origin of replication and then the termination sequence, the left right NS-L and NS-R elements. So, these entire chromosome get transformed into a random coil structure. So, the formation of these E. coli nucleoid takes place in a stepwise manner.

So, in B we can see a random coil form adopted by the pure circular DNA of E. coli at thermal equilibrium without supercoils and additional stabilizing factors. In C and you can see here the position of oriC here and the dif sequence over here. This random coil volume is roughly around 523 cubic microns mere micrometers.

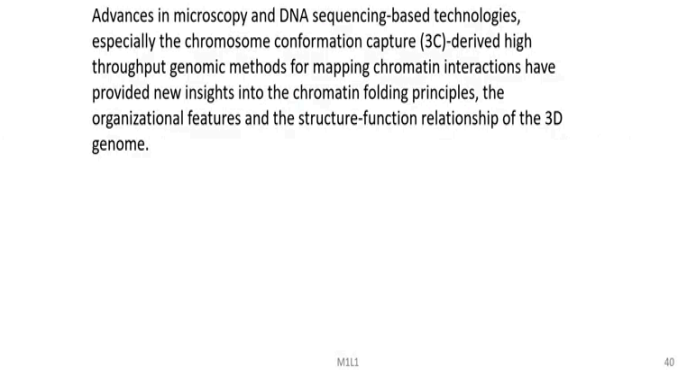
So, in figure C we see the genome organization in vivo. This is exactly how the genome is organized inside a E. coli cell. This circular chromosome in a finally, organized into a typical structure as shown in figure C. So, in C we can see the chromosome of a newly born E. coli cell the GNA genomic DNA is not only condensed by 1000 fold compared to its pure random coil, but it is also specially organized.

So, there are two things happening here a 1000 fold condensation and a spatial organization that is the arrangement of these in space. And here also you can see that oriC and dif are localized in the mid cell and specific regions of the DNA indicated by the colors in A. For example, this is the right turn red that is located in the right extreme. And the left turn left sequences represented by blue are present in the left extreme.

The position of the six spatial domains 4 structured and 2 non-structured is easily noticeable over here. The condensed and organized form of the DNA together with its associated proteins and RNA is called as the nucleoid. So, this is the nucleoid and these are the NS-L here and this is the NS-R here. So, this is how the 6 spatial domains in a E. coli chromosome are finally, getting arranged in the nucleoid.

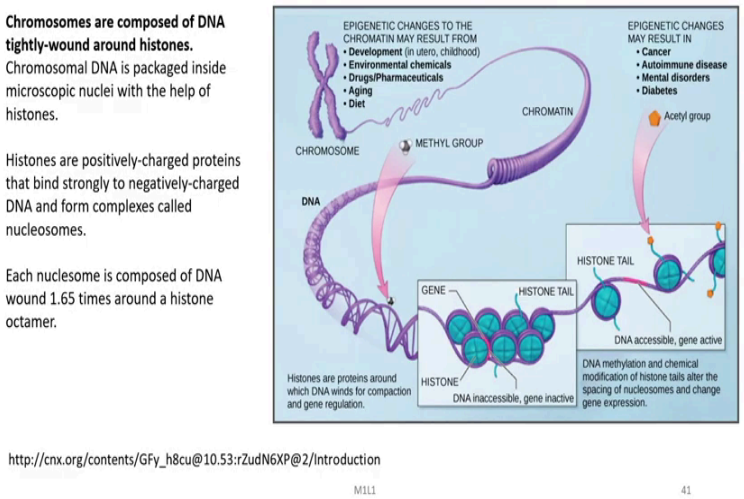
And you can see over here the reduction in the volume from random coil which was around roughly 523 cubic microns to less than 1 micron. So, this is a 1000 fold condensation and quite unremarkable arrangement.

(Refer Slide Time: 47:44)



The advances in microscopy and DNA sequencing based technologies, especially the chromosome conformation capture derived high throughput genomic methods for mapping promoting interactions have provided us new insights in the chromatic following principles the organizational features and the structure function relationship of the 3D genome.

(Refer Slide Time: 48:10)



Chromosomes are composed of DNA tightly wound around histones. So, let us now see the structure of the eukaryotic chromosomes. So, you can see here starting from the histone proteins into which the DNA gets wrapped and then these are condensed into the next level of structure.

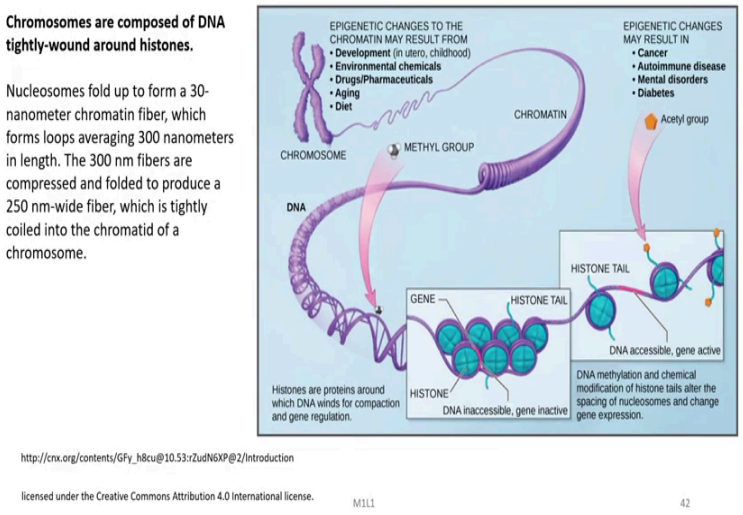
So here, you can see this DNA and this gene which is tightly wrapped in histone and also packed is inaccessible and these makes the gene inactive. The gene is active only when it is exposed and as in this particular case.

So, here DNA methylation and chemical modification of the histone tails after this alter the spacing of nucleosome and change the gene expression. Now these histone proteins around which DNA winds up for compaction and gene regulation are a kind of a hurdle when you want to express certain gene. So, the regulatory mechanism in DNA unwinding helps us in gene expression.

So, those are important points to be noted that as the genome gets organized and tightly packaged the gene expression becomes difficult. It is very very interesting how the cell manages to express his genes in spite of the highly organized chromosome, chromosomal structural elements. So, this small discussion probably gives you some idea how in spite of this packaging the gene expression is possible. So, chromosomal DNA is packaged inside the microscopic nuclei with the help of histones.

This is one thing we have to remember all the time. The histones are positively charged proteins then bind strongly to negatively charged DNA and form complex is called nucleosomes. Each nucleosome is composed of DNA which is wrapped or wound 1.6 times around a histone octamer.

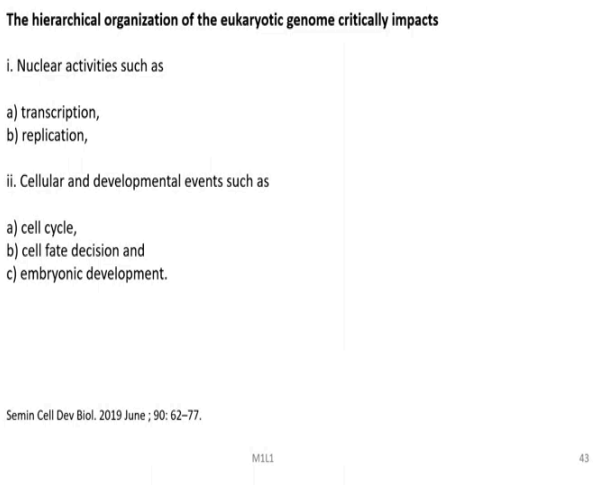
(Refer Slide Time: 50:57)



The chromosomes are composed of DNA tightly wound around histones. Chromosomes fold up to form a 30 nanometer chromatin fiber which forms loops averaging 300 nanometers in length. The 300 nanometer fibers are compressed and folded to produce a 250 nanometer wide fiber which is tightly coiled into the chromatid of a chromosome. In this figure you can see various other important features.

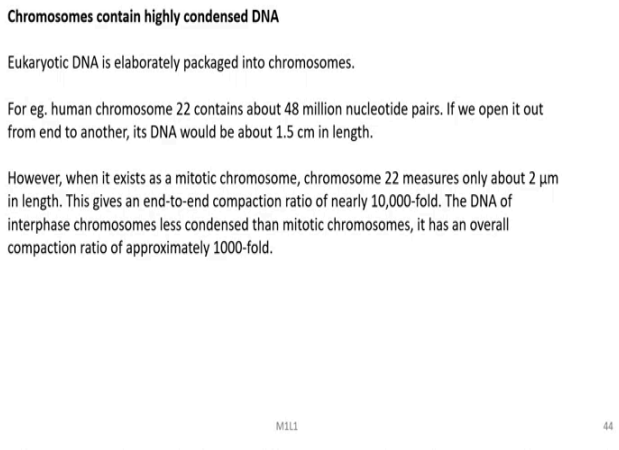
For example, you can see these big chromosome which is constituted by chromatins and these chromatins are made up of DNA, which have wound around a histone proteins and this is in brief the hierarchical structure of genes and chromosomes.

(Refer Slide Time: 52:08)



The hierarchical organization of the eukaryotic genome critically impacts as I have already discussed various nuclear activities such as transcription and replication. And you now know why that is due to the high order of packaging. It also impacts cellular and developmental events such as cell cycle, cell fate decision and embryonic development.

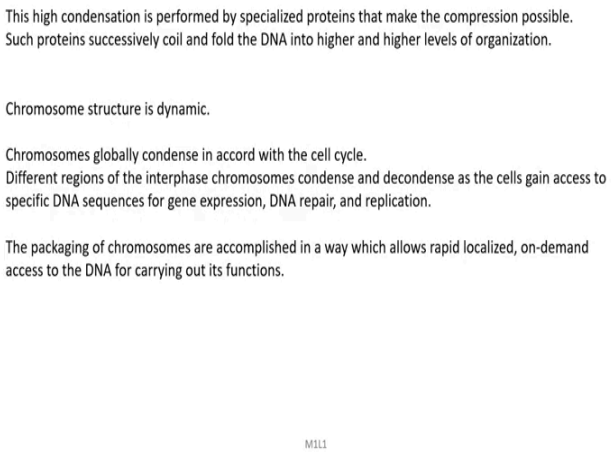
(Refer Slide Time: 52:38)



These chromosomes contain highly condensed DNA. Chromosomes contain highly condensed DNA. Eukaryotic DNA is elaborately packaged into chromosomes.

For example human chromosome 22 contains about 48 million nucleotide pairs, if you open it out from end to end its DNA would be about 1.5 centimeter in length. However, when it exist as a mitotic chromosome, chromosome 22 measures only about 2 microns in length. This gives an end to end compaction ratio of nearly 10000 fold. The DNA of interface chromosomes are less condensed than mitotic chromosomes and it has an overall compaction ratio of approximately 1000 fold.

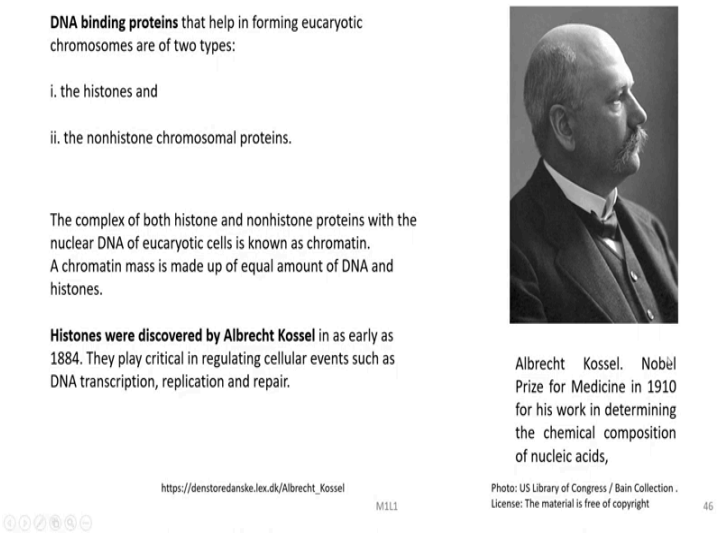
(Refer Slide Time: 53:31)



This high condensation is performed by specialized proteins that make the compression possible such proteins successively coil and fold the DNA into higher and higher levels of organization. We have to understand that the chromosome structure is not static, it is dynamic. Chromosomes globally condense in accord with the phases of the cell cycle. Different regions of the interface chromosomes condense and de-condense as the cells gain access to specific DNA sequences for gene expression, DNA repair and replication.

The packaging of chromosomes accomplish in a way which allows rapid localization on demand access to the DNA for carrying out functions. Without these flexibility, the organism will be as good as dead.

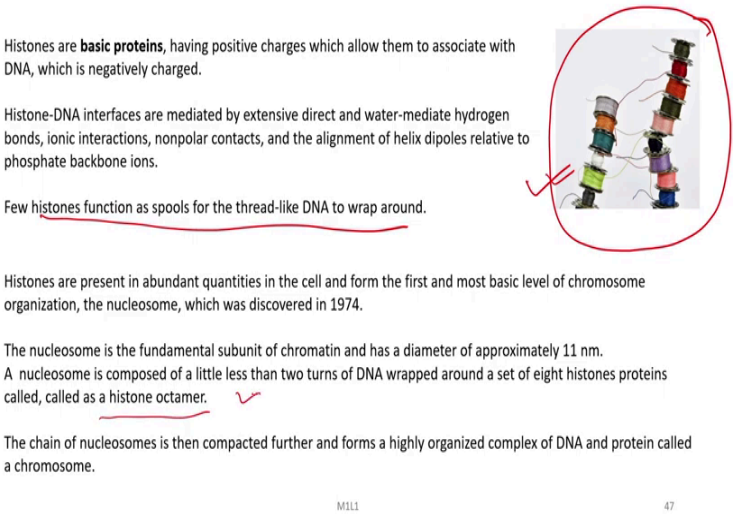
(Refer Slide Time: 54:35)



For this DNA condensation and packaging, we have to know that these proteins must be having some kind of DNA binding properties. DNA binding proteins that help in forming eukaryotic chromosomes are of two types; the first one are the histones and the second one are the non-histone chromosomal proteins. The complex of both histone and non-histone proteins with the nuclear DNA of eukaryotic cells is known as a chromatin. A chromatin mass is made up of equal amount of DNA in histones. The histones were discovered by Albrecht Kossel in as early as 1884.

They play a critical role in regulating cellular events such as DNA transcription, replication and repair. Albrecht Kossel was awarded the Nobel prize for medicine in 1910 for his work in determining the chemical composition of nucleic acids.

(Refer Slide Time: 55:48)

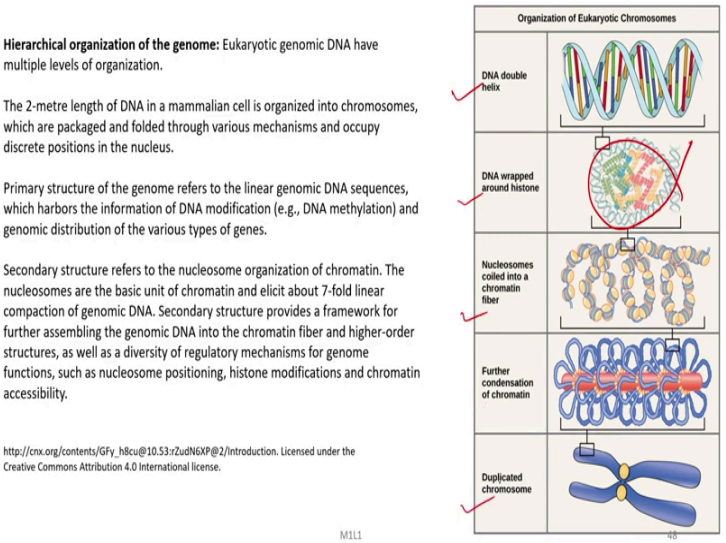


These histones are basic proteins as already told they have positive charges which allow them to associate with DNA which is negatively charged. This picture on the left is of some threads wrapped on bobbins and many of you may have seen it.

So, this would help us in understanding the arrangement of DNA on histone proteins. The histone DNA interfaces are mediated by extensive direct and water mediated hydrogen bonds, ionic interactions, non polar context and the alignment of helix dipoles relative to phosphate backbone ions. Few histones functions as spools for the DNA to wrap around similar to these threads on bobbins. Histones are present in abundant quantities in the cell and form the first and most basic levels of chromosome organization the nucleosome.

The nucleosome is the fundamental subunit of chromatin and has a diameter of approximately 11 nanometers. A nucleosome is composed of a little less than two turns of DNA wrapped around a set of eight histone proteins called as a histone octamer. So, basically the DNA structure is a string on beads and not a bead on string. The scene of nucleosome is then compacted further and forms a highly organized complexed DNA and protein call as a chromosomes.

(Refer Slide Time: 57:43)

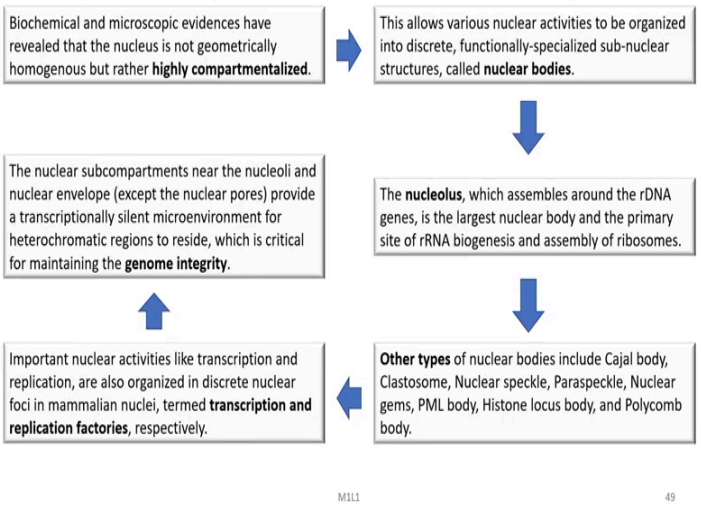


Let us examine the hierarchical organization of the genome. Eukaryotic genomic DNA have multiple levels of organization the 2 meter length of DNA in a mammalian cell is organized into chromosomes, which are packaged and folded through various mechanisms and occupy discrete positions in the nucleus. The primary structure of the genome refers to the linear genomic sequences which harbors the information of DNA modification example DNA methylation and genomic distribution of the various types of genes.

The secondary structure refers to the nucleosome organization of chromatin. The nucleosomes are the basic unit of chromatin and elicit about 7 fold linear compaction of genomic DNA. Secondary structures provide a framework for further assembling the genomic DNA into the chromatin fiber and higher order structures as well as a diversity of regulatory mechanisms for genome functions such as nucleosome positioning, histone modifications and chromatin assembly.

So, here you can see in the pictorial representation, the various levels of organization of the eukaryotic chromosome. You have a DNA double helix structure over here. This DNA wraps around the histones and the histone core here which is around 1.65 turns. Further these nucleosomes are coiled into a chromatin fiber and results in the next level of the structure. These are further condensed and they form the chromatin and finally, they gives rise to the large chromosomes.

(Refer Slide Time: 59:50)

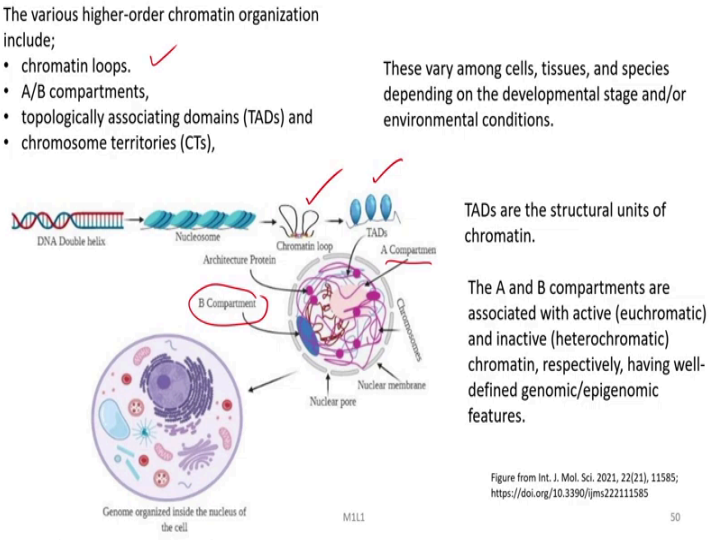


The biochemical and microscopic evidences have revealed that the nucleus is not geometrically homogeneous, but rather highly compartmentalized. We need to remember these to understand the next higher level of genome organization. So, the nucleus is highly compartmentalized. This compartmentalization allows various nucleic activities nuclear activities to be organized into discrete functionally specialized subcellular structures called nuclear bodies.

The nucleolus which assembles around the rDNA genes. The nucleolus which assembles around the rDNA genes is the largest nuclear body and the primary site of rRNA biogenesis, an assembly of ribosomes. There are other types of nuclear bodies like Cajal body, Clastosome, Nuclear speckle, Paraspeckle, Nuclear gems, PML body, Histone locus body and Polycomb body and each of them have special activities inside them.

We are not going to discuss those in large detail over here. But what we need to understand is that important nuclear activities like transcription and replication are also organized in discrete nuclear foci in mammalian nuclei and these are termed as transcription and replication factories respectively. The nuclear compartments near the nucleoli and nuclear envelope except the nucleopores provide a transcriptionally silent microenvironment for heterochromatic regions to reside which is critical for maintaining the genome integrity.

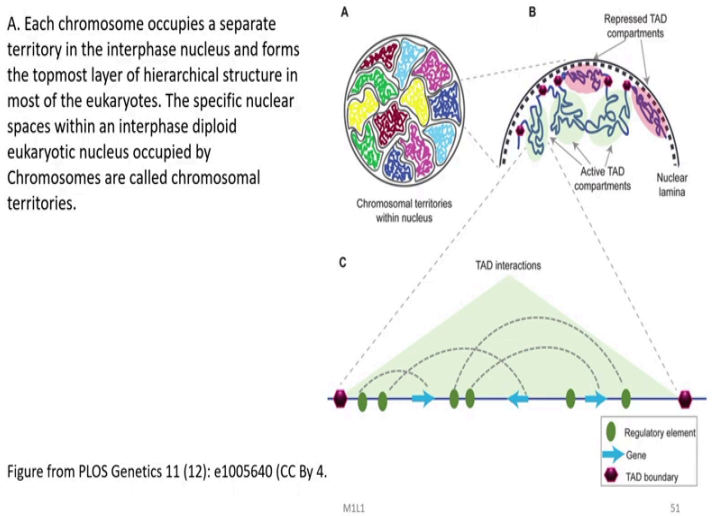
(Refer Slide Time: 61:57)



The various higher order chromatin organization include a chromatin loops A and B compartments, topologically associated associating domains and chromosome territories. These vary among cells tissues and species depending on the developmental stage and or environmental conditions thats are the structural units of the chromatin.

The A and B compartments are associated active and inactive chromatin. The active chromatin are the eukaryotic chromatin and the inactive chromatin are the heterochromatic chromatin. And they have well defined genomic and epigenomic features.

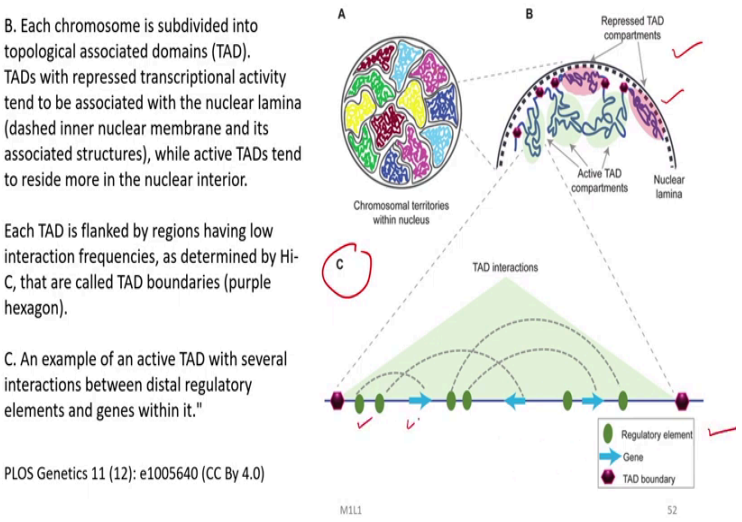
(Refer Slide Time: 63:01)



We need to understand that the nucleus is highly compartmentalized and each chromosome occupies a separate territory in the interface nucleus and forms the topmost layer of hierarchical structure in most of the eukaryotes. So, you can see here the chromosome territories within the nucleus and each chromosome occupies a particular space in their colored accordingly, the green represents green colors represent some chromosomes, yellow and blue and so on.

So, they are not organized in a random way, they are always arranged and positioned in a particular place in the nucleus. The specific nucleus species within an interface deployed eukaryotic nucleus occupied by chromosomes are called as nuclear territories. In the way certain animals in the jungles have defined territories, chromosomes also have defined territories inside the nucleus. We need to remember this for our next level of discussion.

(Refer Slide Time: 64:27)



Each chromosome is subdivided into topological associated domains. TADs with repressed transcriptional activity tend to be associated with the nuclear lamina. This is the nuclear lamina and these TADs which are repressed for their transcriptional activity prefer to stay near to this nuclear lamina while the active TAD tend to reside more into the nuclear interior. So, these are the active TADs and these are not near the nuclear lamina, they are more into the nuclear interior spaces.

Each TAD is flanked by regions having low interaction frequencies as determined by Hi-C and these are called as TAD boundaries and you can see these in the purple hexagons in

figure C we can see the example of an active TAD with several interactions between distal regulatory elements and genes within it. So, these are the regulatory elements and these are the genes and these are the TAD boundaries we had discussed prior to this. So, with this, we come to end of our lecture.

Thank you for your attention.