1 - Chromosome Structure

The first reading assignment this semester examines the general features of the chromosomes found within viruses, prokaryotes (bacteria), and eukaryotes (fungi, protozoa, algae, plants, and animals).

Viral Chromosomes

The genetic material (**genome**) of viruses can be composed of either RNA or DNA; however, single virus type never has both DNA and RNA in the same virus particle.

The genomes of viruses can be in several forms:

- **Double-stranded DNA (dsDNA)**. A dsDNA genome contains two individual DNA strands held together by hydrogen bonds. Even though you will not be required to know examples of viruses with dsDNA genomes, several human pathogens have dsDNA genomes, including the smallpox virus and the herpes viruses.
- **Single-stranded DNA (ssDNA)**. The genomes of many viruses are composed of a single DNA strand. Parvovirus, which infects dogs and cats, has a single-stranded DNA genome.
- **Double-stranded RNA (dsRNA)**. A dsRNA genome contains two individual RNA strands held together by hydrogen bonds. Rotavirus, which causes severe diarrhea in humans, has a double-stranded RNA genome.
- **Single-stranded RNA (ssRNA)**. The genomes of many viruses are composed of a single RNA strand. Many disease-causing viruses, such as poliovirus, influenza virus, SARS-CoV-2 (causes COVID-19), and the human immunodeficiency virus (HIV) contain single-stranded RNA genomes.

The genomes of viruses can also be **circular** or **linear**. One way to determine if a viral genome is circular or linear is to isolate the viral genome and treat the genome with **nucleases**, enzymes that digest (cut) DNA or RNA. **Exonucleases** digest nucleic acids into nucleotides only if there is a free end; **endonucleases** cut DNA or RNA in the middle of a nucleic acid molecule. As a result, circular genomes are sensitive to endonucleases, while linear genomes are sensitive to both exonucleases and endonucleases.

The virus genome can be contained within one continuous nucleic acid molecule, or the viral genome can be divided into **segments**. The genome of the influenza virus, for example, contains eight linear ssRNA segments.

When the genome of a virus is located within a virus particle, the genome is inert, meaning that the genome is not copied and viral genes are not transcribed. A virus genome is copied and viral genes are transcribed only during the infection of a host cell.

Viral genomes can range from a few thousand base pairs to 250,000 base pairs in length. For comparison, the genome of the bacterium *E. coli* is 4 million base pairs in length, while the haploid human genome is 3 billion base pairs in length.

Key Questions

- What are the four major criteria used for classifying viral genomes?
- How can a scientist determine if a viral genome is linear or circular?

Bacterial Chromosomes

The genome within a bacterial cell is typically composed of a single **chromosome**. Bacteria are prokaryotic, and since prokaryotes do not contain nuclei, the bacterial chromosome is not contained within a nuclear membrane. Instead the bacterial chromosome is found in a region of the bacterial cytoplasm called the **nucleoid** (see **Figure 1.1**).

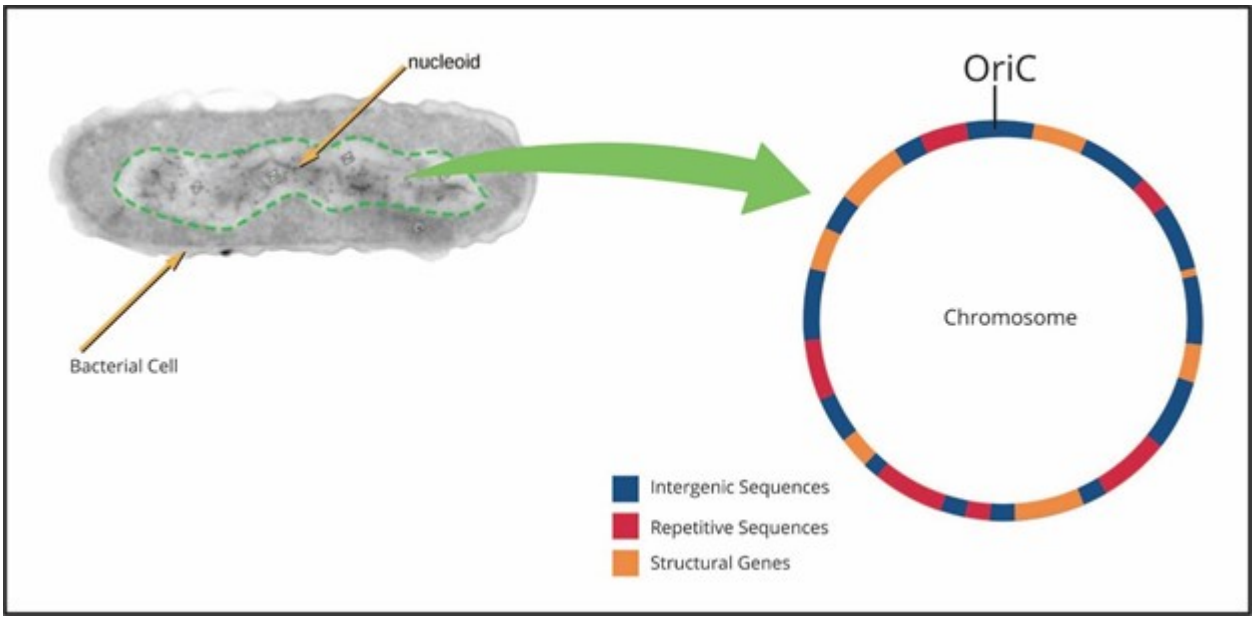


Figure 1.1 ***Bacterial Chromosome Structure*** *--- Prokaryote cell image adapted from OpenStax (access for free at https://openstax.org/books/biology-2e/pages/1-introduction) --- Image created by SL*

A bacterial chromosome has the following features:

- The bacterial chromosome is usually a single, circular double-stranded DNA molecule.
- The bacterial chromosome is usually 2 - 5 million base pairs (bp) in length.
- The bacterial chromosome contains 1,000 - 3,000 **structural genes**. These structural genes are transcribed and translated to make protein products.
- The bacterial chromosome has a single **origin of replication**. The origin of replication serves as the binding site for proteins involved in initiating DNA replication. The origin of replication in the bacterium *E. coli* is called ***oriC***.
- The bacterial chromosome includes **intergenic DNA sequences**. Intergenic sequences are located between structural genes and are not transcribed. Intergenic sequences serve as the binding sites for proteins that function to activate or deactivate structural genes.
- The bacterial chromosome contains **repetitive DNA sequences**. Repetitive sequences are repeats of a particular base pair sequence, are often found within the intergenic DNA sequences, and are involved in compacting the chromosome to fit into the nucleoid region of the bacterial cell.

Eukaryotic Chromosomes

The genome within a eukaryotic cell is subdivided into multiple chromosomes. Each eukaryotic chromosome is a single, linear double-stranded DNA molecule that is approximately 10–100 million base pairs (bp) in length (see **Figure 1.2**).

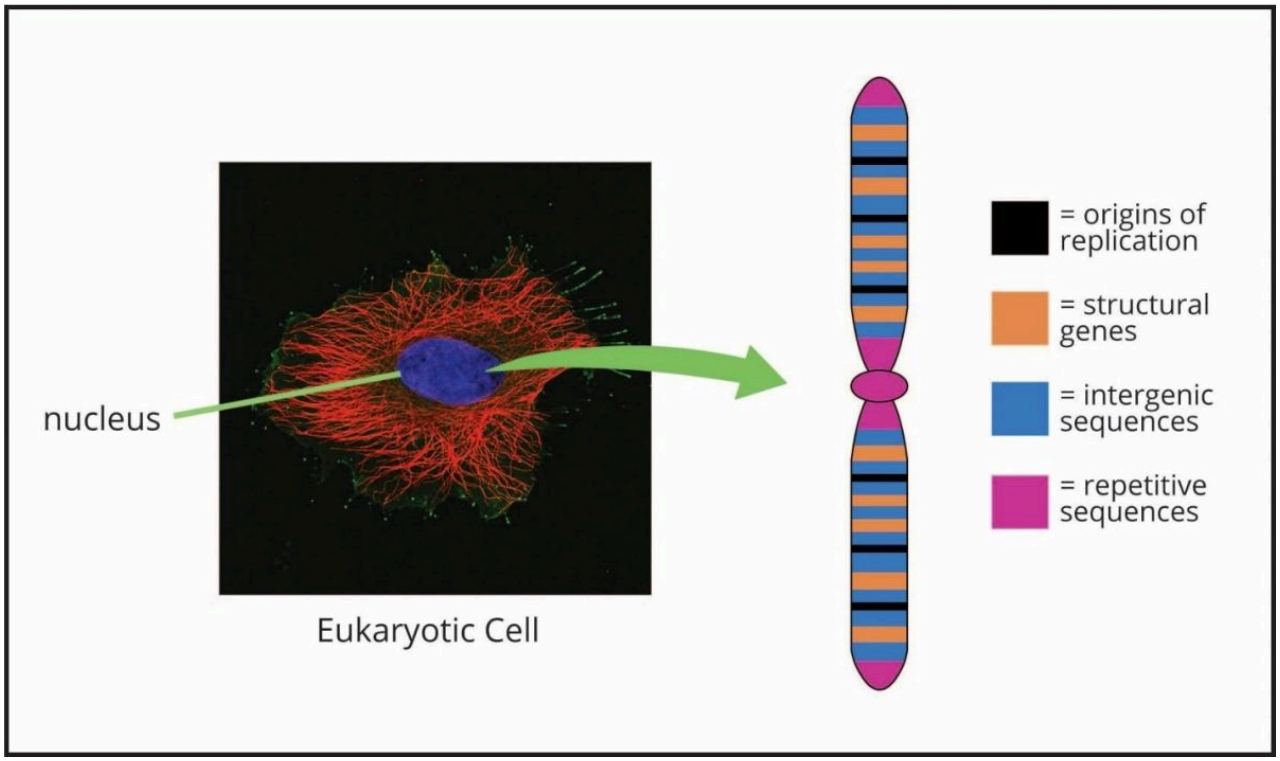


Figure 1.2 ***Eukaryotic Chromosome Structure --- Image created by SL***

A eukaryotic chromosome has several important features:

- **Origins of replication**. Eukaryotic chromosomes contain many origins of replication, spaced at approximately 100,000 base pair (bp) intervals along the chromosome. In the yeast *Saccharomyces cerevisiae*, each origin is called an **ARS element.**
- **Centromeres**. Each eukaryotic chromosome has a single centromere. Centromeres play a critical role in chromosome separation into daughter cells during mitosis and meiosis. A protein structure called the **kinetochore** covers the centromere DNA sequence. The kinetochore functions to link the centromere DNA to the microtubule spindle of the dividing cell, ensuring proper chromosome movement during mitosis and meiosis.
- **Telomeres**. Telomeres are the ends of eukaryotic chromosomes. Telomeres function to prevent chromosomes from sticking together (i.e., prevent translocations). Telomeres also protect the ends of chromosomes from exonucleases and prevent chromosome shortening during DNA replication.
- **Structural genes**. Several hundred to thousands of structural genes are found within a typical eukaryotic chromosome. Recall that structural genes encode protein products. Eukaryotic structural genes contain two types of DNA sequences: **exons** and **introns**. The exon sequences encode the amino acids within the protein product, while the intron sequences between exons do not code for the protein product.
- **Intergenic DNA sequences.** Intergenic sequences are located between structural genes and are not typically transcribed. Intergenic sequences include DNA sequences that serve as the binding sites for the proteins that function to activate or deactivate genes.
- **Repetitive DNA sequences**. Repetitive DNA sequences are repeats of the same DNA sequence and comprise approximately 60% of the human genome. Most of the repetitive DNA sequences do not encode protein products. Repetitive DNA sequences will be discussed in more detail below.
- **Heterochromatin**. Heterochromatin refers to regions along a chromosome that contain highly condensed DNA. These heterochromatin regions either lack structural genes altogether or contain structural genes that are not actively transcribed. The centromere and telomere regions of chromosomes are composed of heterochromatin.
- **Euchromatin**. Euchromatin refers to the loosely condensed regions along the chromosome. Many structural genes are located within euchromatin.

Key Questions

- How are prokaryotic and eukaryotic chromosomes similar?
- How are prokaryotic and eukaryotic chromosomes different?
- What is meant by the term *structural gene*?
- What is the difference between an exon and an intron?
- What are the functions of centromeres and telomeres?
- What is the difference between heterochromatin and euchromatin?

Repetitive Sequences in Eukaryotes

Some DNA sequences found within eukaryotic chromosomes are **unique DNA sequences.** Keep in mind that most eukaryotes are diploid, having two copies of each chromosome (i.e., a homologous chromosome pair). As a result, eukaryotes typically have two copies of each unique DNA sequence; one copy of the gene on each chromosome within a homologous chromosome pair. Most structural genes are examples of unique DNA sequences.

Eukaryotic genomes also contain **repetitive DNA sequences**. These repetitive DNA sequences include **moderately repetitive DNA sequences** and **highly repetitive DNA sequences.** Moderately repetitive sequences are present in a few hundred to a few thousand copies per genome. Highly repetitive sequences are present in tens of thousands to millions of copies per genome.

DNA Reassociation Experiments

How do we know that eukaryotic genomes have unique, moderately repetitive, and highly repetitive DNA sequences? Before scientists were able to determine the base pair sequence of a DNA molecule, **DNA reassociation experiments** were done to determine the overall composition of the genome, focusing on repetitive DNA sequences. In a typical DNA reassociation experiment, entire chromosomes are isolated and are mechanically sheared into fragments. The chromosome fragments are then denatured into single strands by increasing the temperature of the reaction. The reaction mixture is then cooled. As the reaction cools, single-stranded DNA molecules attempt to find each other and form hydrogen bonds to create double-stranded DNA molecules; different DNA fragments do so at different rates (see **Figure 1.3**). Think of it this way, a single-stranded DNA molecule will move around looking for its complement to reattach to base-for-base to form a double-stranded molecule. For highly or moderately repetitive DNA sequences, there are many single strands in the reaction with a complementary DNA sequence to choose from. As a result, highly and moderately repetitive sequences will find each other more rapidly than unique DNA sequences. The DNA reassociation experiment measures the amount of time it takes for single-stranded DNA to form double strands. DNA reassociation experiments showed that there are three populations of DNA: the DNA sequences that reassociated most rapidly were called highly repetitive, moderately repetitive DNA sequences reassociated next, and finally, unique DNA sequences had the slowest rate of reassociation.

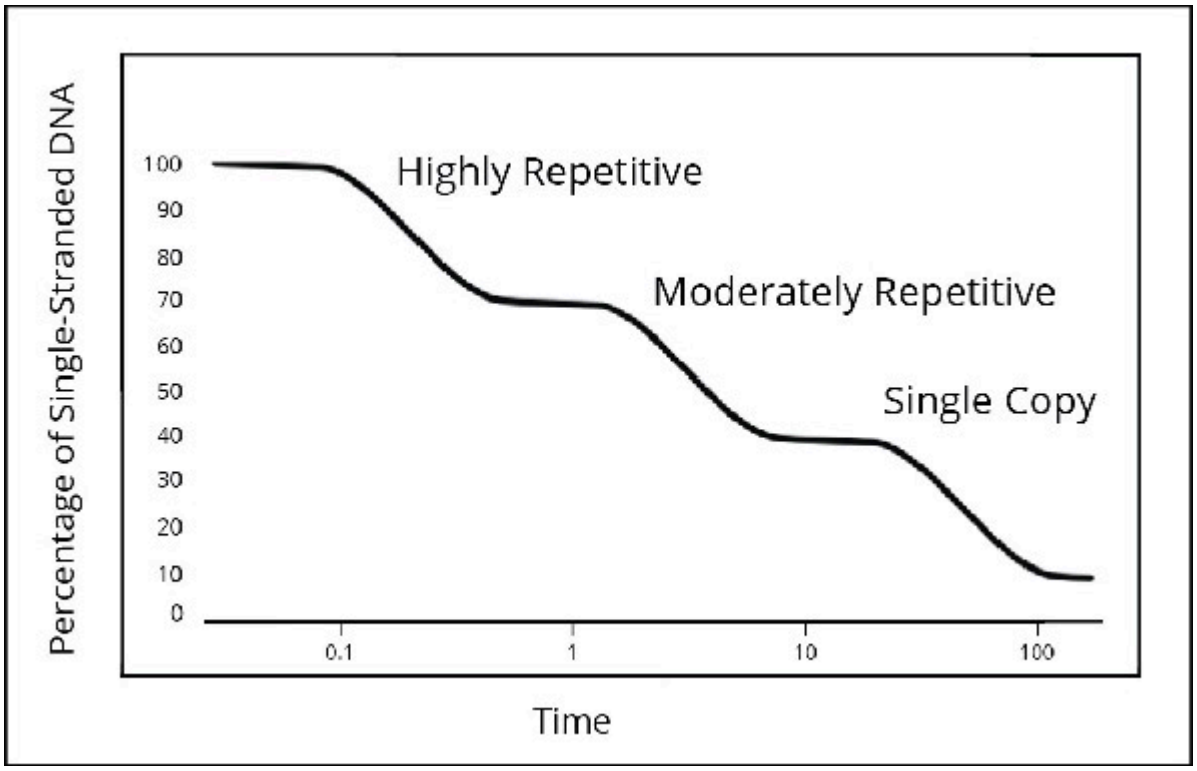


Figure 1.3 ***DNA Reassociation Experiment*** *--- Image created by SL*

Key Questions

- Describe how highly repetitive, moderately repetitive, and unique DNA sequences behave in a DNA reassociation experiment.

Moderately Repetitive Sequences

Moderately repetitive DNA sequences include some genes that produce products. For example, the genes that produce the **ribosomal RNA (rRNA)** components of ribosomes (see Part 11) and the genes that make **histone** proteins (see Part 2) are considered moderately repetitive DNA sequences.

Moderately repetitive DNA sequences also include sequences of unknown function. A good example of this type of moderately repetitive sequence is the **variable number tandem repeat (VNTR)** sequences. VNTRs are typically 15 to 100 base pairs long, are often located between genes, and are present in multiple copies repeated along the length of the chromosome. The number of VNTR repeats on each chromosome is unique to each individual. As a result, this variation in VNTR repeats is the basis of the forensics technique **DNA fingerprinting** (see **Figure 1.4**).

The **telomere repeat** sequences (see **figure 1.6**) are also moderately repetitive DNA sequences.

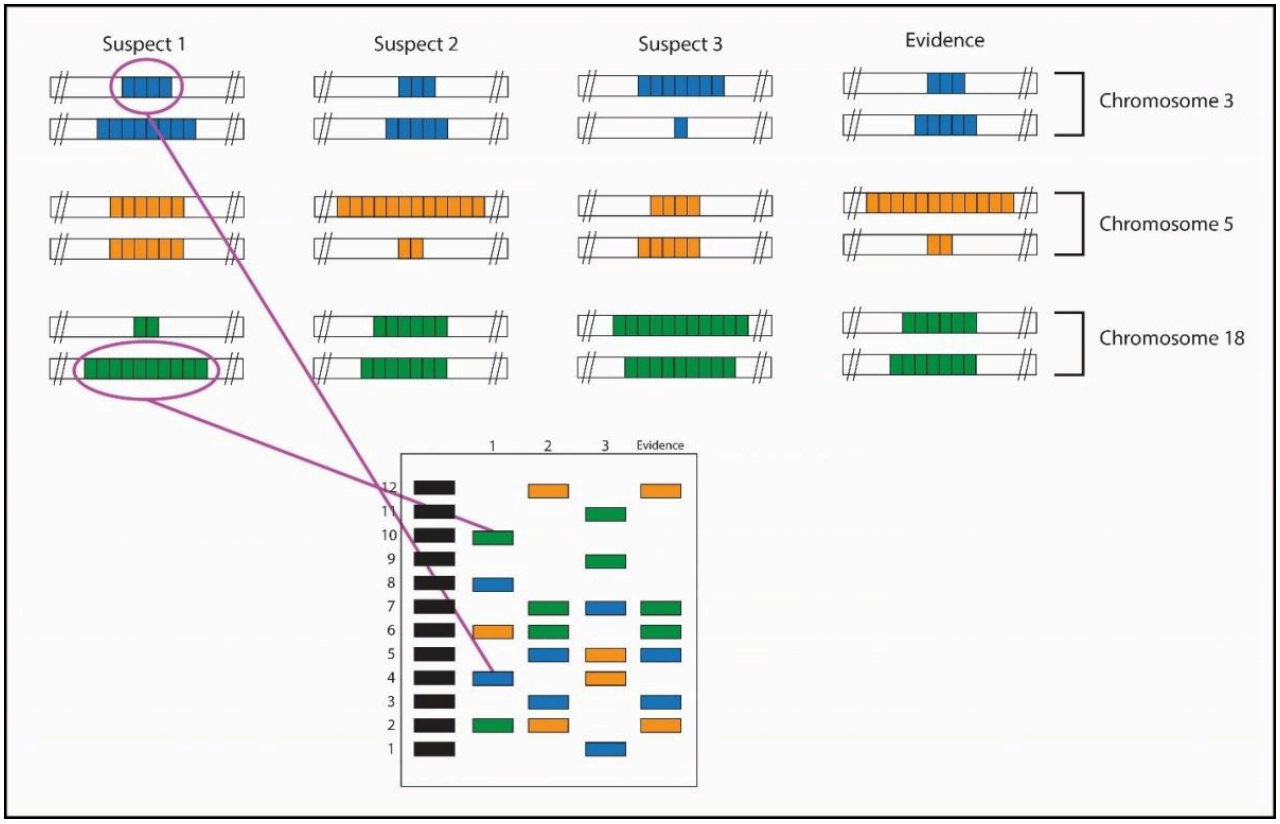


Figure 1.4 ***VNTRs can be used in forensics.*** *The VNTR repeats of three suspect individuals and a sample left at a crime scene (evidence) are shown. The number of boxes on each chromosome indicate the number of VNTR repeats. Recall that each person has two copies of each chromosome (i.e., a homologous chromosome pair). The bottom of the image represents the agarose gel electrophoresis technique (see Part 8) that separates DNA molecules by size; VNTRs with fewer repeats migrate farther through the gel (i.e., towards the bottom of the gel) than VNTRs with more repeats. The expected migration pattern of DNA molecules with 1-12 VNTR repeats is indicated on the left side of the gel in black. The results presented above shows that Suspect 2 had VNTR repeats that match the evidence left at a crime scene. --- Image created by SL*

Key Questions

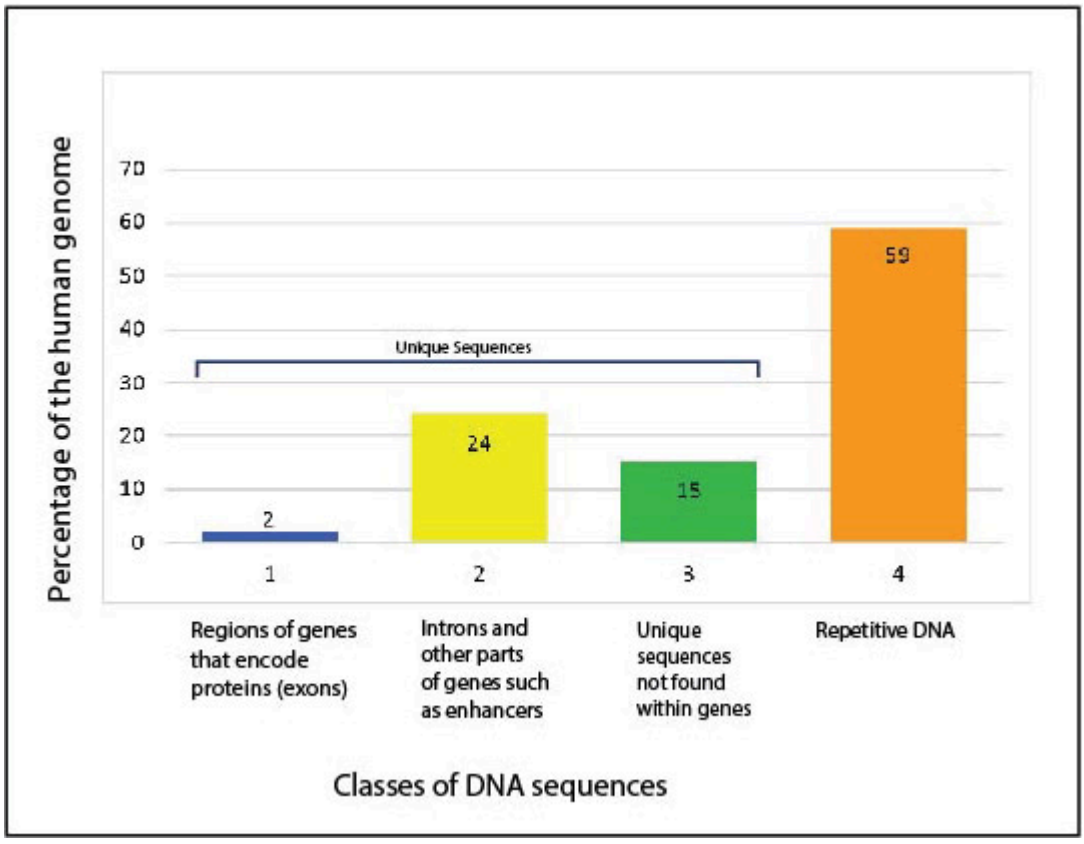
- What are four examples of moderately repetitive DNA sequences?
- Why are VNTRs well suited for forensics?

Highly Repetitive Sequence

The centromere region (**CEN** region) of the chromosome contains highly repetitive DNA sequences. In humans, the CEN region is approximately 10^6 base pairs (bp) long, consisting of a 170 bp **tandem repeat** (i.e., copies of the same 170 bp DNA sequence repeated many times in a row).

The ***Alu*** family of DNA sequences in humans is another example of a highly repetitive sequence. An individual *Alu* sequence within the human genome is only 300 bp long; however, there are so many copies of this *Alu* sequence scattered throughout the human genome that approximately 10% of the human genome is thought to be composed strictly of *Alu* sequences (see **Figure 1.5**). To put this into perspective, only 2% of the human genome is composed of structural genes that produce protein products. Some of these *Alu* sequences are particularly interesting because they have the potential to move from one location in the genome to another. DNA sequences that can move within the genome are called **transposable elements**.

Finally, the **heterochromatin** regions of a chromosome often contain highly repetitive DNA sequences.



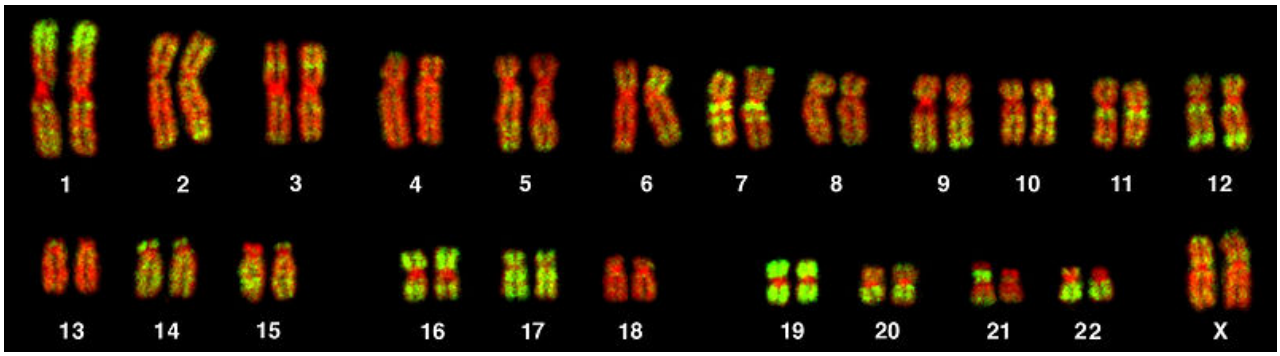


Figure 1.5 ***Top) Classes of DNA Sequence in the Human Genome.*** *Repetitive DNA sequences make up approximately 60% of the human genome, while the exon regions within structural genes compose only 2% of the genome. --- Image created by SL* ***Bottom) Alu Elements.*** *The Alu sequence elements are labeled with a fluorescent green tag. Notice that Alu sequences are found on all 46 chromosomes. --- Chromosomes Alu Fish by Andreas Bolzer is licensed under CC BY 2.5*

Key Questions

- What are three examples of highly repetitive DNA sequences?
- What makes transposable elements unique?

Telomeres

The telomeres of eukaryotic chromosomes have the following features:

- **Telomeres contain tandem repeat DNA sequences.** The tandem repeat DNA sequences within telomeres are 6–8 base pairs (bp) long; each tandem repeat contains multiple G and T nucleotides. For example, the telomere repeat sequence in humans is 5'-TTAGGG-3'. Depending on the eukaryotic species, each telomere may contain several hundred to several thousand tandem repeats of the same telomere DNA sequence (see **Figure 1.6**). Thus, the telomere repeat sequences are moderately repetitive.
- **Telomeres contain 3' single-stranded overhangs.** The 3' overhang is a single-stranded DNA sequence, containing multiple copies of the telomere repeat. The 3' overhang is typically 12–16 bp in length.

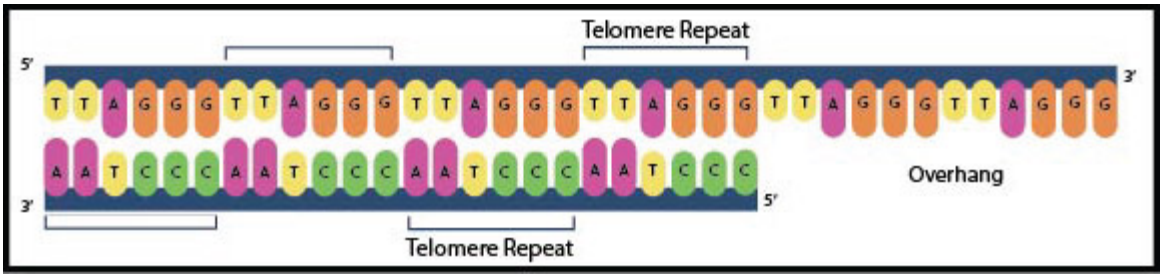


Figure 1.6 ***Telomere Structure ---*** *Image created by SL*

- **Telomere 3' single-stranded overhangs form loops**. The 3' single-stranded overhang within the telomere can turn back on itself to form a **t-loop** (see **Figure 1.7**). Within the t-loop, the 3' single-stranded overhang of the telomere invades another portion of the same chromosome and forms unusual hydrogen bonds between guanine (G) nitrogenous bases. These unusual hydrogen bonds involve four G bases, producing a **G quartet** structure. The t-loop is thought to be the actual structure that protects the eukaryotic chromosome from exonucleases.

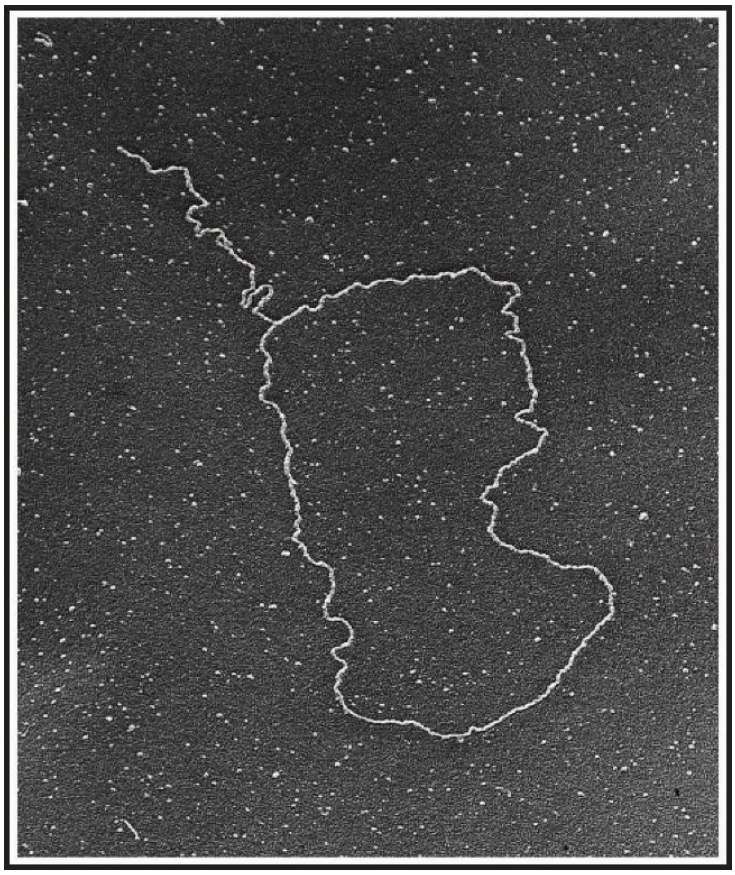


Figure 1.7 ***A telomere loop (t-loop) as visualized in the electron microscope.*** *--- Image courtesy of Dr. Jack Griffith*

Key Questions

- At what locations on a chromosome are you likely to find repetitive DNA sequences?
- What is the function of a t-loop?

Classifying Chromosomes

Eukaryotic chromosomes can be distinguished from each other in the microscope by the location of the centromere (see **Figure 1.8**), the size of the chromosome, and the banding patterns produced along the chromosome after staining with certain chemical dyes. The centromere separates the chromosome into halves (each half is called an arm); the shorter of the two chromosome arms is designated ***p***, while the longer arm is designated ***q***. In terms of centromere location, chromosomes are classified as follows:

- **Metacentric**. The centromere of a metacentric chromosome is located near the center of the chromosome.
- **Submetacentric**. The centromere of a submetacentric chromosome is located slightly off center.
- **Acrocentric**. The centromere of an acrocentric chromosome is located significantly off center. In humans, there are five pairs of acrocentric chromosomes: 13, 14, 15, 21, and 22. These five pairs of chromosomes contain short *p* arms having multiple copies of the same **ribosomal RNA (rRNA)** genes. Having many copies of the ribosomal RNA genes ensures that the cell is able to produce enough ribosome components for translation (see Part 11). The number of rRNA gene copies varies among individuals, but the average is 100 copies per genome. Thus, the rRNA genes are considered moderately repetitive.
- **Telocentric**. The centromere of a telocentric chromosome is located near the end of the chromosome. The human genome does not contain telocentric chromosomes; however, the mouse genome contains telocentric chromosomes.

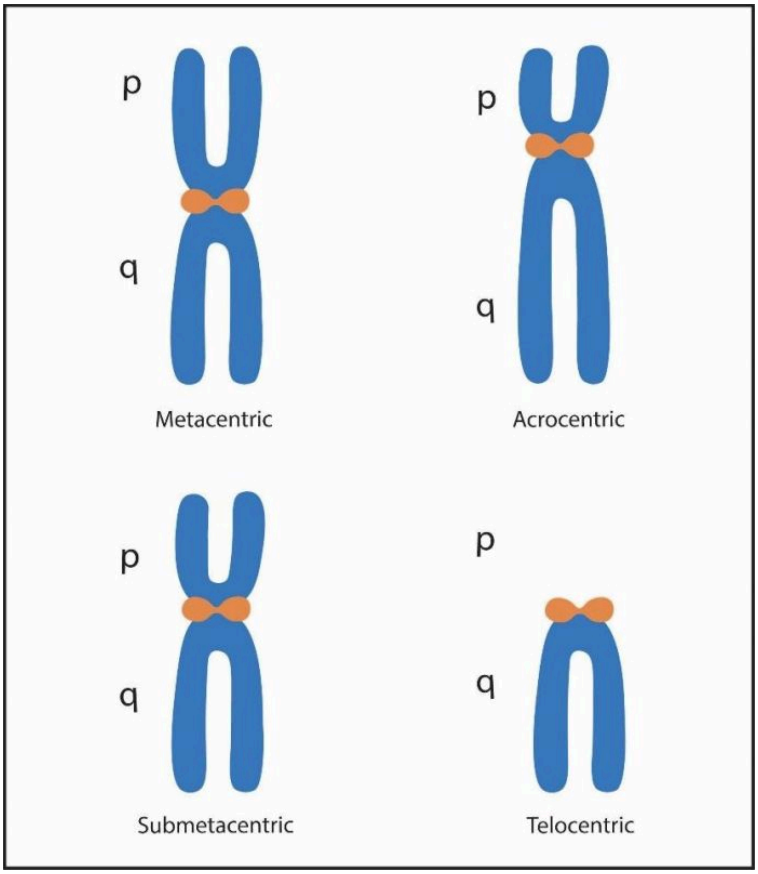


Figure 1.8 Centromere Location. *Note that each chromosome has gone through DNA replication, forming two sister chromatids per chromosome.--- Image created by SL*

Key Questions

- What is meant by the terms metacentric, submetacentric, acrocentric, and telocentric?
- What family of genes is found on the short arms of the five acrocentric human chromosomes?

Human Karyotype and Staining

A **karyotype** is an image of all of the chromosomes within a dividing cell, in which the homologous chromosomes (recall that one chromosome in a homologous pair is inherited from mom; the other chromosome is inherited from dad)

are arranged in pairs (see **Figure 1.9**). The chromosomes are aligned so that their *p* arms are above the centromere and the *q* arms are located below the centromere. Human **autosomes** (non-sex chromosomes) are numbered from the largest to the smallest chromosome, 1 to 22. The **sex chromosomes** are labeled X and Y.

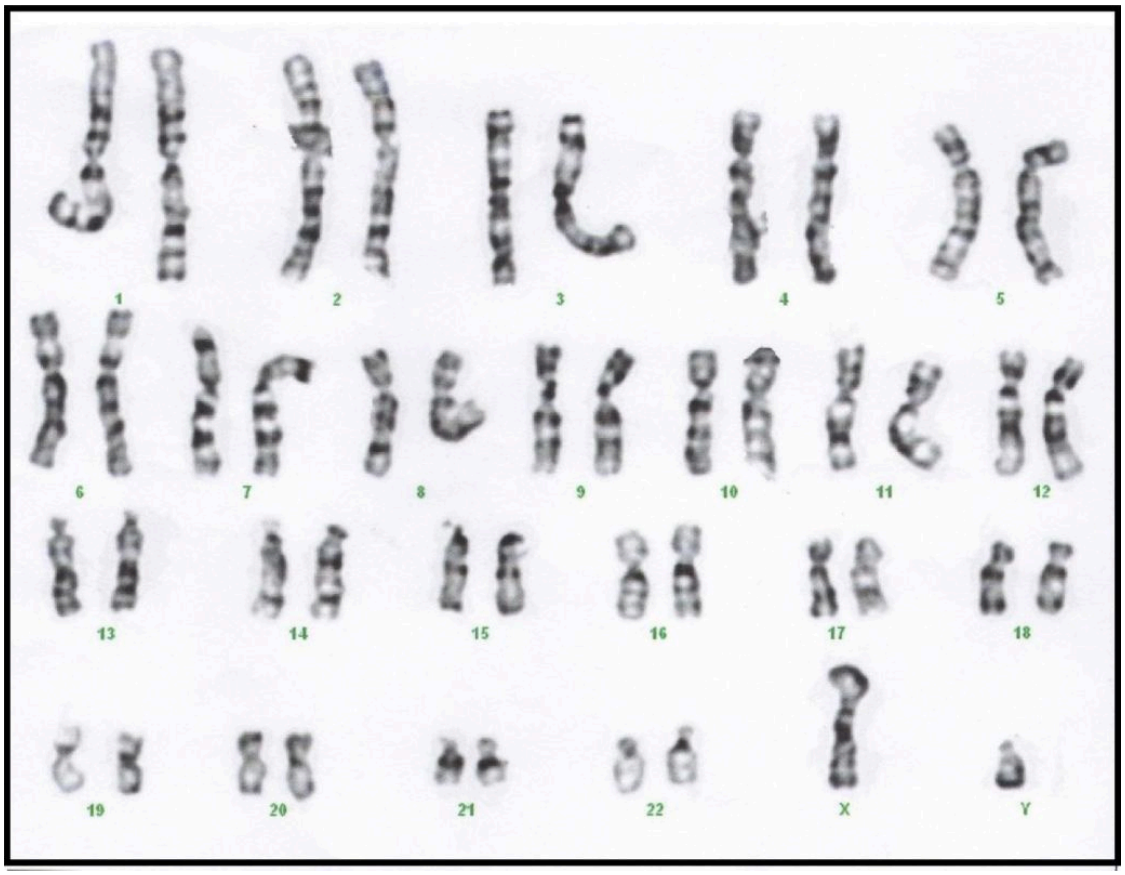


Figure 1.9 ***Karyotype of Human Male.*** *The chromosomes have been stained with the chemical dye Giemsa.* **---** *Karyotype by Can H. is licensed under CC BY 2.0*

Some chromosomes are similar in size and in centromere location. As a result, these chromosomes are difficult to distinguish from each other in the microscope, unless the chromosomes are stained with dyes to produce banding patterns that are unique to each chromosome. A common staining procedure involves the chemical dye **Giemsa**, which produces a unique pattern of light and dark bands (G banding) on each chromosome. Dark bands on the chromosomes represent areas of the DNA that are tightly compacted (heterochromatin); light bands represent areas of the DNA that are loosely compacted (euchromatin).

Key Questions

- What are three ways that scientists can distinguish chromosomes from each other?

Chromosome Nomenclature

A numbering system has been established to describe human chromosomes based on the size, centromere location, and banding pattern. This numbering system assists in determining where chromosome mutations (deletions, duplications, etc.) occur and helps to delineate the exact location of the abnormality. For example, band 22q12 refers to

chromosome 22, the long arm (q), region 1 (closest to the centromere), band 2. If a deletion removes a portion of chromosome 22, the exact location of that deletion can be identified based on this numbering system.

Review Questions

Fill in the blanks:

1. A(n) _____ is an enzyme that digests the ends of linear nucleic acid molecules.
2. A(n) _____ is an enzyme that cuts both linear and circular nucleic acid molecules.
3. Bacterial chromosomes are found in a region of the cytoplasm called the _____.
4. One distinction between prokaryotic and eukaryotic chromosomes is that bacterial chromosomes have _____ origin of replication while eukaryotic chromosomes have _____.
5. Eukaryotes contain highly condensed DNA that lacks genes, these regions called _____ are not generally transcribed and appear as _____ bands on a Giemsa-stained chromosome.
6. _____ are highly repetitive DNA sequences that compose up to 10% of the human genome.
7. The ends of a linear chromosome are called _____ and the portion where spindle proteins attach is called the _____.
8. Two types of genes that are moderately repetitive include _____ and _____ genes.
9. The shorter piece of a chromosome is called the _____ arm while the longer piece is called the _____ arm.
10. _____ genes are found on the *p* arms of chromosomes 13, 14, 15, 21, and 22.