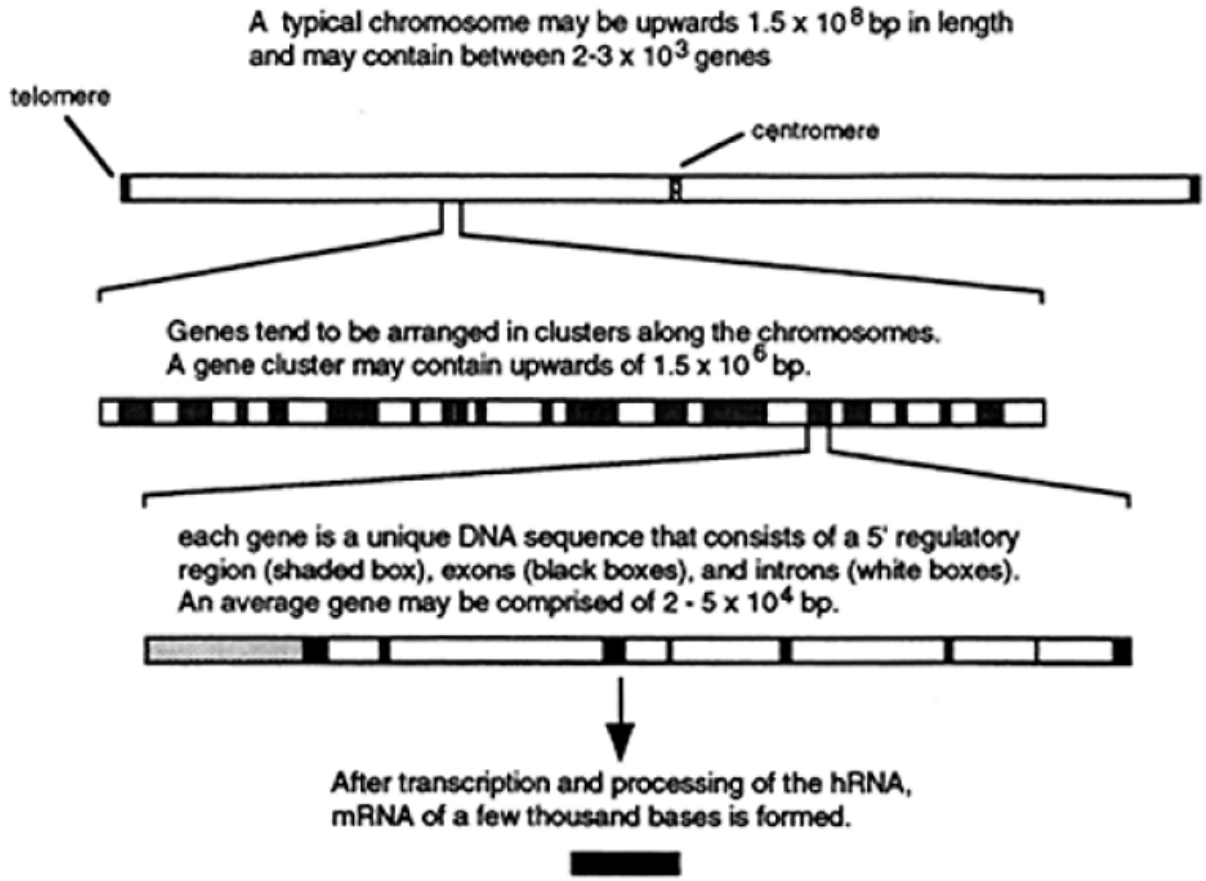
# Types of DNA sequences

**On the basis of copies of nucleotide sequences, the chromosomal DNA is of three types, viz.**

(1) Unique or single-copy DNA sequences (1 - 10 copies/genome)
(2) Palindromic DNA
(3) Repetitive DNA

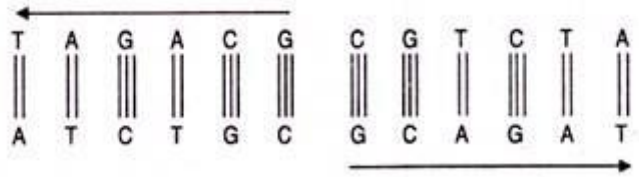
**Unique Sequences**

Greater than 50% of the eukaryotic genome consists of DNA that is unique in sequence. The human genome encodes for about 100,000 proteins. The average coding portions of a gene (the exons) consist of about 2,000 base pairs of DNA that is unique in sequence. This number represents less than 7% of the total DNA comprising the human genome and less than 14% of that DNA is unique. Most of the coding sequences are interrupted by from 1 to 50 noncoding sequences or introns. The total length of the introns that interrupt a gene generally far exceeds the total length of the exons. The spatial distribution of genes, exons, introns and regulatory sequences along each chromosome is shown below.



**Palindromic DNA:**

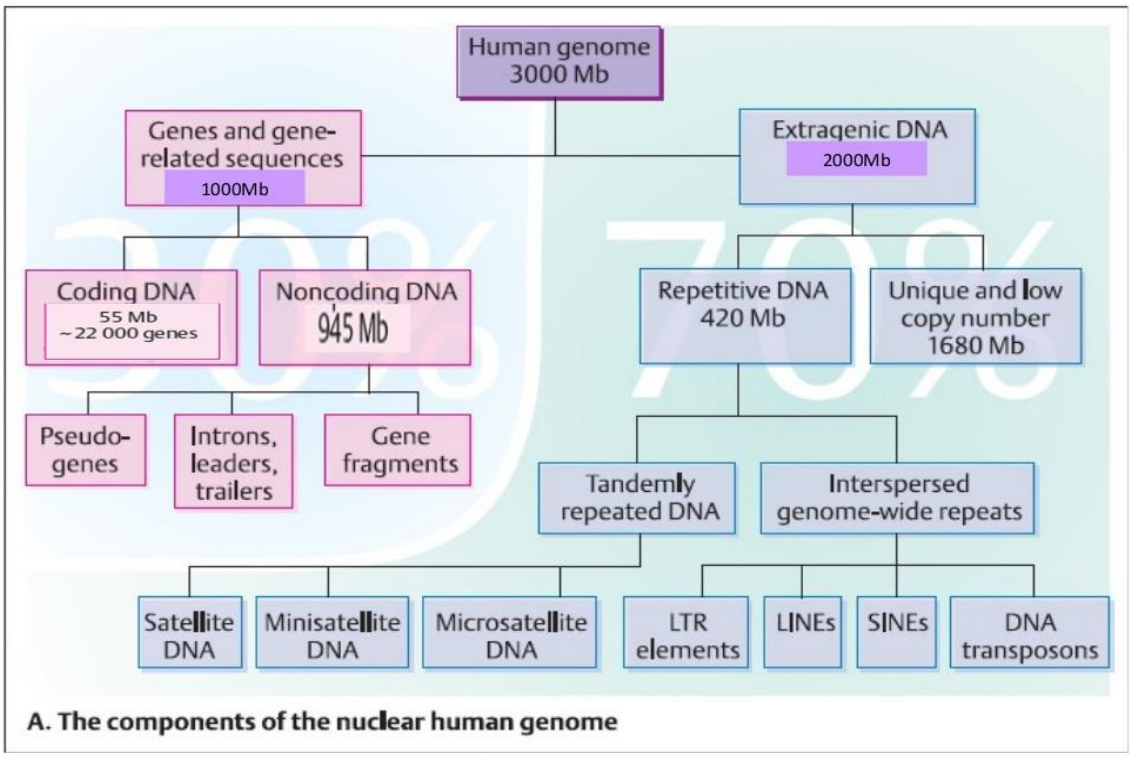
A palindrome is a sentence that reads the same in forward and backward direction. As for example: 'AND MADAM DNA'. Palindromic sequence of nucleotides is sometimes found in DNA. DNA duplex possesses areas where sequence of nucleotides is the same but opposite in the two strands. These sequences are recognised by restriction endonucleases and are used in genetic engineering. Such a DNA sequence is called palindromic DNA. This term was first used by Thomas et. al. (1973).



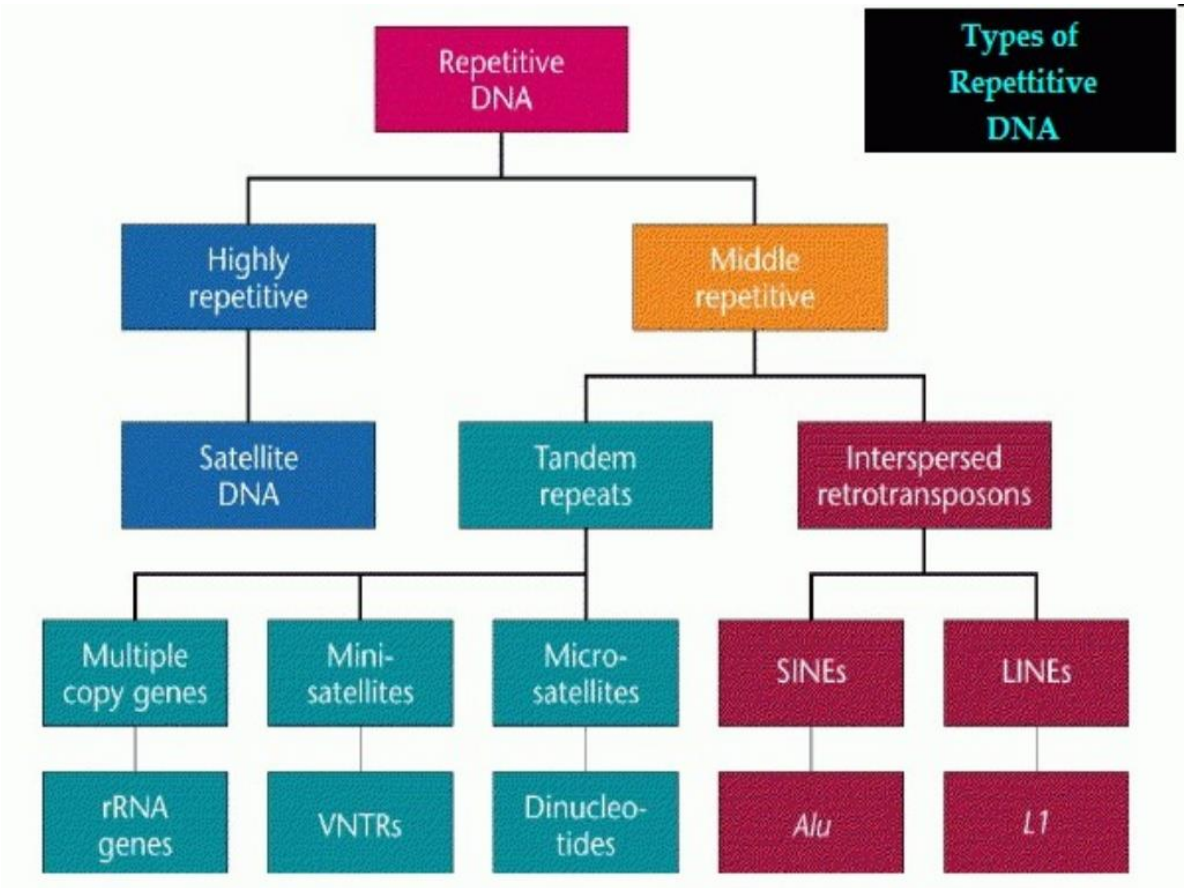

**Repetitive DNA:**

Most of the non-coding DNA having multiple repeats of the same nucleotides or base sequences is known as repetitive DNA. The function of repetitive DNA is not really known but approximately 30% of the human genome consists of repetitive DNA.

Repetitive DNA is found almost in all eukaryotes (except yeast) but almost negligible (0.3%) in prokaryotes. Such DNA is found near the centromere in the chromosome. The percentage of repetitive DNA varies from species to species. In many cases, the repeating sequences have different base composition from the remaining DNA. In such case, separation of repetitive DNA is very easy by way of ultra centrifugation.




A. The components of the nuclear human genome




There are multiple classes of repetitive DNA, two of these classes include: highly repetitive and moderately repetitive DNA.

***Highly Repetitive DNA Sequences*** or satellite DNA consists of several different sets (>$10^5$ copies/genome) of short repeated polynucleotides, generally the repeats range from 5 to 500 base pairs in length and exist in tandem arrays. Highly repetitive DNA comprises about 10-15% of the total genomic DNA, is present in over a million copies and is transcriptionally inactive. Some of the highly repetitive DNA is clustered in structural regions of chromosomes particularly in the cetromeric and telomeric regions.

***Moderately Repetitive DNA Sequences*** contains a large variety of repeated sequences (100 - $10^5$ copies/genome) ranging from a few hundred to tens of thousands of base pairs with different characteristics. Moderately repetitive DNA can be clustered at specific chromosomal locations or distributed throughout the genome. One type of moderately repetitive human DNA sequence is the rRNA precursor gene. Each rRNA precursor gene is contained in a DNA segment of about 43,000 base pairs. The actual transcript is 13,400 bases which is processed into the mature 28S, 18S and 5.8S rRNA's. This means that at least 30,000 base pairs are not transcribed and apparently serve as spacer DNA. About 280 copies of the rRNA precursor gene are distributed in clusters on five chromosomes and account for about 0.4% of the genomic DNA. Most types of moderately repetitive DNA are short about 300 base pairs in length, are interspersed with unique sequences, are often transcribed but do not code for gene product.

There may be many thousands of these copies present tandemly repeated (follow one another directly) and called as ***Satellite DNA***. When multiple copies found scattered in the genome is called as dispersed repetitive DNA.

On the basis of the length of repeating sequence satellite DNA is of two types i.e. mini-satellites (variable number tandem repeats, VNTRs) and microsatellites (simple tandem repeats, STRs).

Human satellite DNA is comprised of very large arrays of tandemly repeated DNA with the repeat unit being a simple or moderately complex sequence (100 Kb to several Mb). Repeted DNA of this type is not transcribed. Satellite DNA accounts for the bulk of the heterochromatic regions of the genome, being notably found in the vicinity of the centromeres.

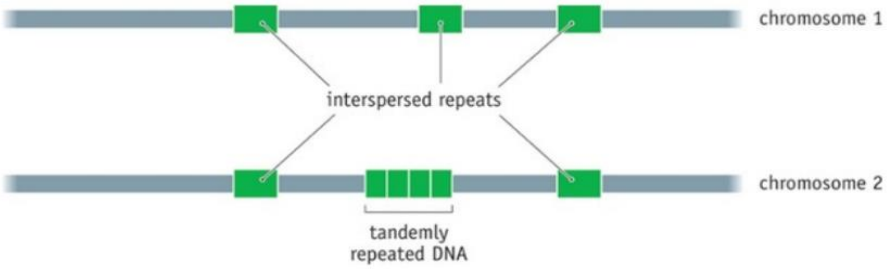
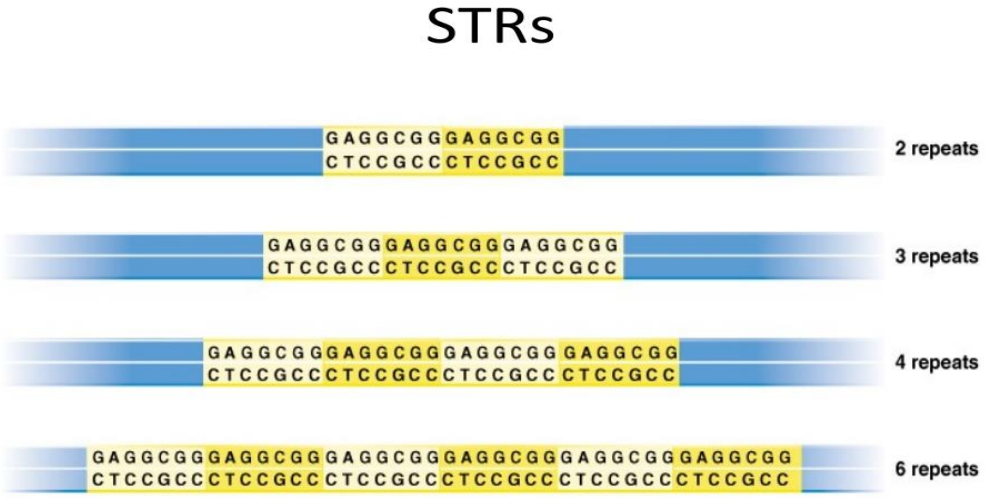



**Figure 19.9** The two types of repetitive DNA: interspersed repeats and…

***Mini-satellite DNA*** comprises a collection of moderately sized arrays of tandemly repeated DNA sequences which are dispersed over considerable portions of the nuclear genome. Like Satellite DNA sequences they are not normally transcribed. In humans, 90% of Mini-satellites are found at the sub-telomeric region of chromosomes. The telomere sequence itself is a tandem repeat: TTAGGGTTAGGGTTAGGG. The variation in size (array length) of these regions between indivisuals in humans was originally the basis for DNA fingerprinting.

A ***Variable Number Tandem Repeat (VNTR)*** is a location in a genome where a short nucleotide sequence is organised as a tandem repeat. These can be found on many chromosomes, and often show variations in length between individuals. Not only that, the number of repeats varies widely in the populations, although the repeat number is usually well preserved during transmission. Therefore each variant acts as an inherited allele, allowing them to be used for personal or parental identification. The site with each different repeat number can be treated as a highly polymorphic site with multiple allele and is known as VNTR site. Their analysis is useful in genetics and biology research, forensics, and DNA fingerprinting, DNA profiling.

***Micro-satellite DNA*** comprises of repeating sequences of 1-6 base pairs of DNA and can be repeated 10 to 100 times. In human the most common is the (CA)n sequence where n varies from 5-50 or more. The Trinucleotide and Tetranucleotide tandem repeats are comparatively rare. The lengths of particular micro-satellite sequences tend to be highly variable among individuals. These differences make up molecular 'alleles'. It is also known as Short Tandem Repeat (STR), Simple Sequence Length Polymorphism (SSLP) and Simple Sequence Repeat (SSR).




Although micro-satellite DNA has generally been identified in intragenetic DNA or within the introns of genes, a few example have been recorded within the coding sequences of genes.

Two types of dispersed repeated sequences are known as SINEs and LINEs. All eukaryotic organisms have SINEs and LINEs with a wide variation in their relative proportions. Humans and frogs, for example, have mostly SINEs whereas Drosophila and birds have mostly LINEs. SINEs and LINEs represent a significant proportion of the entire moderately repeated DNA in the genome.

***SINEs [Short Interspersed Nuclear Element(s)]*** are 150-300 base pair (bp) repeated elements that are found in the "short interspersion pattern" – typically possess an 8-20 bp inverted repeat (characteristic of "insertion" sequences) called 'target-site duplications'

In primates and rodents, alternatively, there are only a few prominent families, with on family usually outnumbering all other families; in humans, for example, the *Alu*I family consists of up to 500,000 copies of the *Alu*I repeat (300 bp) and comprises up to 5% of the human genome.

SINE sequences are transcribed but are not translated -- in humans, *Alu*I sequences are found in 20% of (pre-) mRNA but are removed during mRNA processing.

***LINEs [Long Interspersed Nuclear Elements]*** are an interesting and heterogeneous class of sequences comprised in part of transposons (mobile genetic elements, jumping genes, nomadic sequences, etc.) and retrotransposons (mobile elements depending on reverse transcription). These elements that are 3,000 - 5,000 bp in length that are dispersed (interspersed) throughout genomes (hence LINE).

LINEs comprise about 21% of the human genome and consist of repetitive sequences up to 6500bp long that are adenine rich at their 3' ends. They are clearly mobile (able to "move" from location to location within a genome) and inducible. The latter accounts for the phenomenon of "hybrid dysgenesis".

Repetitive DNA is the part of the DNA which has repetitin of the base pairs no matter how much and in what way like palindrome repeats or mirror repeats or trinucleotide repeats or simple repeats...

Satellite DNA is the part of the repetitive DNA which is very specific in nature; DNA that contains many tandem (not inverted) repeats of a short basic repeating unit. Satellite DNA is located at very specific spots in the genome (on chromosomes 1, 9, 16 and the Y chromosome, the tiny short arms of chromosomes 13-15 and 21 and 22, and near the centromeres of chromosomes). Satellite DNA is highly repetitive DNA sequences found in heterochromatin, composed of simple sequences (very short) repeated in tandem many times to form large blocks of sequence. Additionally, following the accumulation of mutations, these blocks of repeats have been repeated in tandem themselves. The degree of repetition is on the order of 1000 to 10 million at each locus. Loci are few, usually one or two per chromosome. They were called satellites since in density gradients, they often sediment as distinct, satellite bands separate from the bulk of genomic DNA owing to a distinct base composition.