# BIO2133 – Test 1 – Textbook Notes

## 1.2 Genetics Progressed from Mendel to DNA in Less Than a Century

### 1.2.1 The Chromosome Theory of Inheritance

- **eukaryotes:** # of chromosomes in each cell = 2n (diploid)
  - ex**:** humans have a diploid # = 46
- **homologous chromosomes:** chromosome pairs in diploid cell
  - Identical in **size** and **location of centromere** and with **genes** for the same characteristics at corresponding **loci**.
- **Mitosis:** chromosomes copied and distributed so each daughter cell receives diploid set of chromosomes
- **Meiosis:** receive 1 chromosome from each chromosome pair

  ∴ get a haploid (n) #

  - essential so that offspring maintain constant # of chromosomes (Diploid number is regenerated with fusion of gametes)

- **Chromosomes VS genes – the same?**
  - Behavior of chromosomes during meiosis IS LIKE behavior of genes during gamete formation
    - genes & chromosomes both exist in pairs
    - Both pairs separate during gamete formation

  - Genes are however part of chromosomes (Sutton & Boveri)

- **chromosome theory of inheritance:** inherited traits are controlled by genes residing on chromosomes
  - these genes are transmitted through gametes, maintaining genetic continuity from generation to generation

### 1.2.2 Genetic Variation

- **Mutation:** any heritable change (changes in genomic sequence)
  - Source of all genetic variation
    - **alleles:** alternative forms of a gene
      - genotype: set of alleles for a given trait
      - phenotype: differences in set of alleles to produce observable changes
- **Mutant genes** = markers
  - Maps location of genes on chromosomes
  - Can result in phenotypic changes (depending on the type of mutation)

### 1.2.3 Chemical Nature of Genes: DNA or Proteins?

- Could be protein?
  - most abundant
  - different types
  - in nucleus and cytoplasm
- DNA proven as genetic material in bacteria (Avery, MacLeod, McCarty) but failed to convince; ∴ evidence DNA is genetic material: from viruses that infect and kill bacteria = **bacteriophage**
- **bacteriophage:** has protein coat surrounding DNA core
  - Showed: during infection, protein coat remains OUTSIDE bacteria cell, while viral DNA enters cell and directs synthesis of more phages.
  - Used a radioactive marker to verify this.

    ∴ DNA carries genetic info

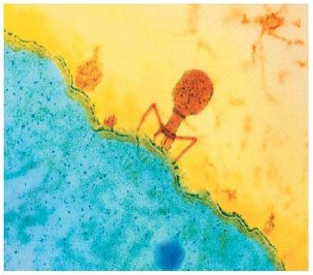


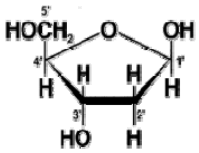
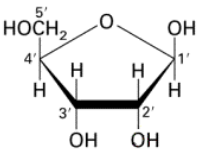
FIGURE 1–5 An electron micrograph showing T phage infecting a cell of the bacterium *E. coli*.

### 1.3 Structure of DNA and RNA

- **DNA:** strands that twist to form double helix
  - Each strand is a linear polymer
  - Strands run **antiparallel** to each other (1 runs 3'-5' the other 5'-3')
    - Made of subunits called **nucleotides** (linked together through 5'-3' phosphodiester bonds to form polynucleotide chains)
      - DNA→ 4 nucleotides
      - Each nucleotide has 1 of 4 nitrogenous bases (A G T C)
      - Nucleoside = Base + sugar
      - Nucleotide = Nucleoside + phosphate group

→ DNA dictates RNA synthesis, which dictates protein synthesis (Central dogma)

- RNA: structurally similar to DNA
  - Has different sugar (ribose, not deoxyribose)
  - Ribose has a 2'OH where deoxyribose has a 2'H (deoxy = removal of oxygen)
  - Uracil (U) instead of thymine (T)
  - Single stranded
  - Can form complementary structures with strands of either DNA or RNA

| | DNA | RNA |
|---|---|---|
| Predominant shape | Double helix | Single stranded |
| Sugar | Deoxyribose  | Ribose  |
| Bases | A=T, G≡C | A=U, G≡C |
| Location | Nucleus | Cytoplasm and nucleus |

**Section 1.3: Proteins and biological Function**

- Proteins are the end products of gene expression
- The diversity of protein and biological function is a result of the fact the proteins are made from a combination of **20 different amino acids**
- Ex. protein of 100 amino acids, where each position can be any of the 20 different types = $20^{100}$ possibilities
- Remember that amino acid tables are **degenerate** (1 amino acid can be made by more than 1 codon, usually the third letter differs)
- Enzymes make up largest category of proteins
- Function of enzymes: Biological catalysts to speed up biological processes by lowering activation energy to body temperature
- Examples of other proteins: haemoglobin (oxygen binding molecule in blood), insulin, histones, collagen

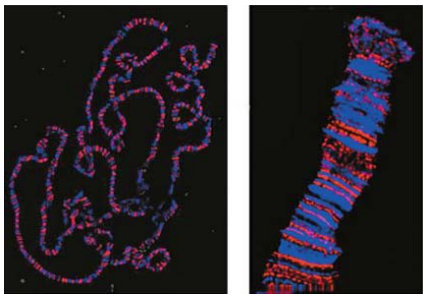
**Section 12.1: Viral and Bacterial Chromosomes are relatively simple DNA molecules**

- Viruses, eukaryotic cells, and bacteria can all package DNA into relatively small volumes
- Viruses and bacteria have much simpler chromosomes in comparison to eukaryotes
    - Usually consist of a single nucleic acid molecule
    - Prokaryote chromosomes lack associated proteins and contain less genetic information
- **Viruses:**
    - Contain one single or double stranded nucleic acid molecule (either DNA/RNA)
    - Can use RNA to store info
    - Can be circular or linear
    - Genetic information carried in viral heads is inert – packed extremely tightly
- **Bacterial chromosomes:**
    - Always double stranded DNA, compacted into an area called the **nucleoid**
    - They contain DNA binding proteins, which are called **HU and H1 proteins.**
        - These proteins are **positively charged** in order to bind to **negatively charged** phosphate groups in DNA
        - Structurally similar to histones, but do not compact DNA.
    - Genetic information in bacteria is not as tightly packed as that in viruses, which gives it the ability to replicate or transcribe – DNA is not inert.

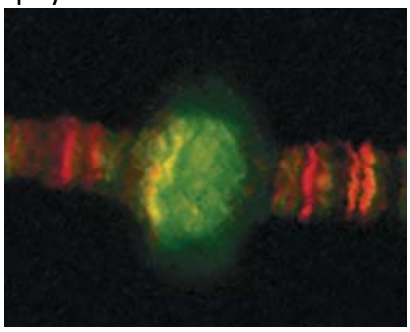
**Section 12.3: Specialized chromosomes reveal variation in the organization of DNA**

- There are two highlighted special cases of eukaryotic chromosomes:
    - Polytene chromosomes
    - Lampbrush chromosomes
- Both types of chromosomes are very large
- Only found in the cells of a few eukaryotic organisms
- Provided insight of many functions and arrangement of genetic information

- **Polytene Chromosomes**
  - Found in various tissues (salivary, midgut, rectal etc), larvae of some flies, protazoans and plants
  - Can be seen during interphase
  - 200-600um long
  - Linear series of bands and interbands under light microscope: banding pattern distinctive for each chromosome and individuals bands are referred to as **chromomeres**



  - **Represent paired homologs:** highly unusual do to the fact that they are present in somatic cells where chromosomal material is dispersed as chromatin and hemologs aren't paired (this is key).
  - **Large size and distinctive appearance** is a result of being composed of large number of identical DNA strands
  - <u>Replication occurs but without strand separation or cytoplasmic division</u>
  - Uncoiling events result in bulges known as **puffs: manifestation of high level of gene activity** (areas of high transcription)
    - Evidence by the high rate of radioactively labeled RNA precursors assayed by autoradiography.



- **Lampbrush Chromosome**
  - Discovered in 1892 by Walter Flemming in salamander ooyctes
    - 2nd time seen in shark oocytes by j. Ruckming in 1892
    - **At this point in time they are known to be a characteristic of vertebrate oocytes and the spermocytes of some insects**
    - **Meiotic chromosomes**
  - These chromosomes are easily isolated from oocytes in the diplotene stage of the prophase I of meiosis where they are active in directing metabolic activities of the developing cell
  - A special characteristic of this chromosome is that it **does not condense**, like most meiotic chromosomes do. They are between 500 to 800um long and do not revert to normal lengths of 15-20um until later on in meiosis
  - **Lampbrush chromosomes are interpreted as extended, uncoiled versions of normal meiotic chromosomes.**

- **Provide significant insight into morphology**
- The study of Lampbrush chromosomes proves the hypothesis that each meiotic chromosome is composed of a pair of sister chromatids; each chromosomal loop is composed of one DNA double helix, while the central axis is made up of two DNA helices.

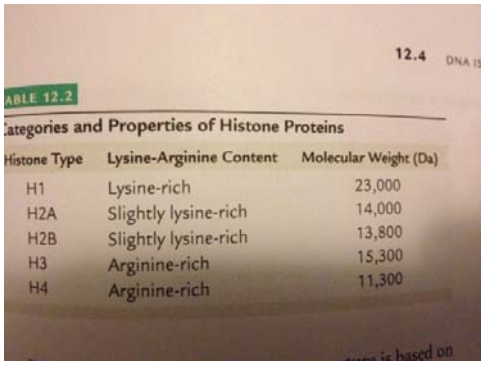
**12.4: DNA is organized into chromatin in eukaryotes.**

How DNA is organized?

- Seen as highly condensed structures during mitosis
  - After chromosome separation of cells, Interphase stage components of chromosomes uncoil and decondense into chromatin. (During interphase chromatin is scattered throughout the nucleus)
  - As cell cycle progresses they form chromosomes again.
  - Contraction of 10000 times for each chromatin fibre.
  - Organization of DNA in eukaryotes is more complex than in viruses and bacteria because greater amount of DNA per chromosome in eukaryotes and presence of large number of proteins which allow for greater amounts of compaction
    - Ex. DNA in E.coli chromosome is 1200um while a human chromosome length ranges from 19000-73000um.
    - Within a human cell nucleus, all 46 chromosomes have enough DNA to extend to 2 meter if uncoiled. This info along with the proteins is contained within the nucleus that measures about 5-10um in diameter.
  - Different cells activate different sets of genes

**Chromatin structure and nucleosomes**

- DNA and RNA strands of viruses and bacteria that are relatively devoid of proteins while eukaryotic chromatins have a substantial amount of protein that is associated with chromosomal DNA
  - The associated proteins can be positively charge histones or less charged non-histone proteins.
  - **Histones play a major structural role**:
    - They contain large amounts of positively charged amino acid, lysine and arginine, which make it possible to bind with negatively charged phosphate groups of nucleotides.



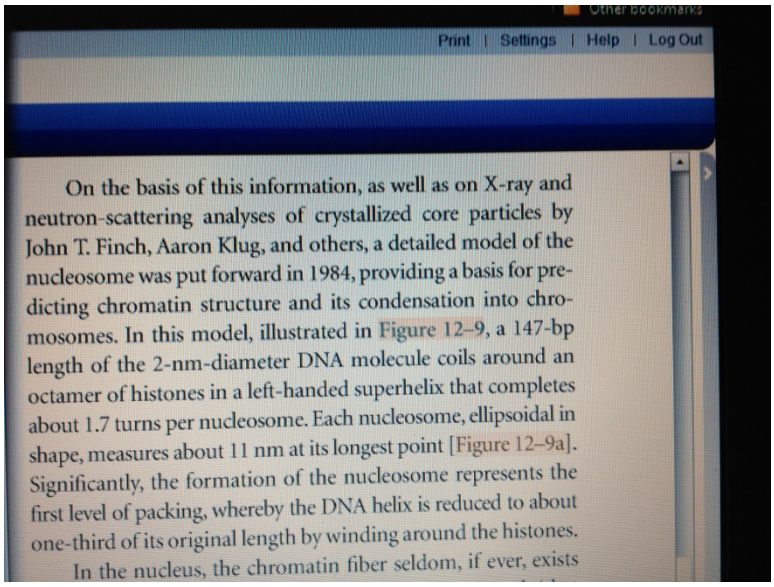
12.4

ABLE 12.2

ategories and Properties of Histone Proteins

| Histone Type | Lysine-Arginine Content | Molecular Weight (Da) |
|---|---|---|
| H1 | Lysine-rich | 23,000 |
| H2A | Slightly lysine-rich | 14,000 |
| H2B | Slightly lysine-rich | 13,800 |
| H3 | Arginine-rich | 15,300 |
| H4 | Arginine-rich | 11,300 |

- Chromatin fibres, composed of DNA and protein undergo coiling and folding to get condensed within the nucleus
- X-ray diffraction indicate that **histones play a role in chromatin structure:**
    - Chromatin produces spaced diffraction rings (suggests that repeating structural units occur along the chromatin axis. If the histones are removed the pattern of diffraction is disrupted.)
- 1984: detailed model of nucleosomes, that provided a basis for predicting chromatin structure




A 147-bp length of the 2-nm diameter **DNA molecule coils around an octamer of histones** in a left-handed superhelix which completes a 1.7 turns per nucleosome. The formation of nucleosome represents the first level of packing, where **DNA helix is reduced to about 1/3 of its original length by winding around histones.**
- DNA -> Nucleosomes -> Solenoid -> Chromatin

**Eukaryotic chromosomes**

- Consists of DNA complexed with histone and nonhistone proteins, called **chromatin**
- DNA wraps around a nucleosome (two molecules each of histones H2A, H2B, H3, H4, forms an octamer)
    - Linker DNA connects adjacent nucleosomes
    - Binding of histone H1 causes nucleosomes to package into a coiled 30nm diameter structure (solenoid)
        - Numerous nucleosomes coiled and stacked upon one another creating 2$^{nd}$ level of packing providing a 6 fold increase in compaction of DNA.
- Nonhistone proteins help control the expression of individual genes

- Further contraction occurs during transition to mitotic chromosome
    - 30nm structures fold into a series of looped domains and further condense the chromatin fibres into structure with 300nm diameter
    - Then they are compacted into the chromosome arms that constitute a chromatid.
- Process of the organization DNA into chromatin and chromatin into mitotic chromosomes can be seen with the follow example:
    - Human cell stores it genetic material within a nucleus of size 5-10um in diameter.
    - The haploid genome consists of over 3 billion base pairs of DNA distributed amongst 23 chromosomes
    - The diploid cell consists twice the amount of the haploid
    - 0.34nm per base pair amounts to a huge length (2m almost)
    - A ratio of DNA length to the length of the structure containing it→ 500 to 1

**Nucleosomes and Chromatin Fiber**

- Nucleosomes form a left handed helix

**Nucleosome Structure is Conserved**

Nucleosome parameters

- 2 each of 4 histones named H2A, H2B, H3 and H4 (the "octamer")
- DNA is wrapped around the outside
- Adding histone H1 brings nucleosomes together to form a 10nm fibre
- Sequences that are distant can now be closer together

Unwound portion is around 50 base pairs, if H1 is present, it will wrap around 2 times due to the change in structure

**The Solenoid, or 30 nm Fibre**

- 6-8 nucleosomes per turn
- Structure is not well defined
- Sequences that are more distant can now be closer together
- H1 binds to linker and brings nucleosomes together. The nucleosomes then bind together in a helical structure called a solenoid.
- It is universally conserved in all eukaryotes

**Chromatin**
Distributed between:

- Euchromatin (loosely packed region, genes are active in RNA transcription)
- Heterochromatin (densely packed masses, genes are inactive, makes up large part of centromere and telomere)
    - Euchromatin changes with time while heterochromatin stays the same

Folds and packs to form thick, rodlike chromosomes during nuclear division

**Chromatin remodelling**

- It's the process of the chromatin changing its structure to accommodate the protein DNA interactions of the proteins that directly interact with DNA. (During transcription for ex). To allow gene expression to occur, chromatin must expose the promoter regions
- Proteins bind to DNA; those proteins use certain types of amino acids to bind to DNA. Amino acids such as lys, arg, and his. Modify amino acids to make histones difficult to interact with each other. This would be a method to allow change in shape.
- **Mechanism for changing the structure; best that is known is the modification of amino acid by methylation or phosphorylation**. Just enough so DNA does not bind well.
- **These modifications affect the tails (amino termini)**
- Mapping sites are not the same on all cell types because transcription is different in all cell types.
- Modelling complexes must be formed only when you want the structure to be changed
- **Acetylation:** a well-practiced histone modification involves the action of enzyme histone acetyl tranferase (HAT).
- Acetylation is linked to gene activity → high levels of acetylation open up, remodel the chromatin fibre. (Increases in regions of active genes and decreases otherwise).
- The influence of these modifications on transcription is known as the "Histone Code"

**Histone Code**

- Post-translational modificaitons of histones introduce meaningful variations into chromatin and provide a regulatory platform for controlling and/or fine-tuning many important DNA-templated processes, including gene transcription, the repair of DNA damage and DNA replication
- Histone modifications, together with factors responsible for adding (writing), interpreting(reading) and removing (erasing) histone modifications, regulate specific and distinct functional outputs of our genomes, which constitute the basis of the 'histone code hypothesis'
- The misregulation of the histone code leads to deregulated gene expression and perturbation of cellular identity, and is therefore a major contributor to cancer initiaiton, progression and/or metastasis

**Chromatin Structure: summary**

- The structure of chromatin influences transcription, replicaiton, repair and recombination
- Chromatin remodeling can convert accessible portions of DNA to inaccessible
- Remodeling is influenced by alterations of the cell physiology
- Remodelling influences the expression of many genes, thus the 'transcriptome'
- The histone code hypothesis postulates that the sum of modifications to the "tails" of the histones mediates transcriptional changes
- Histone modification and histone variants can play a role in the regulation of gene expression.

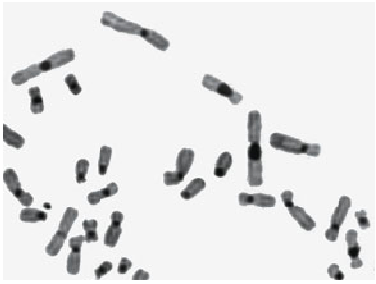
**Heterochromatin**

- Euchromatin and heterochromatin were coined to describe regions of chromosomes that are uncoiled and regions that are coiled and thus remain condensed, respectively.
- **Heterochromatin**
    - **Remain inactive** → lack genes or contain genes that are repressed.
    - They replicate later during the s-phase of the cell cycle than does euchromatin
    - Position effect→ position of a gene or group of genes relative to all other genetic material
    - When a certain heterochromatic area from 1 chromosome is translocated to a new area on the same or another chromosome that area may become inert if it lies adjacent to the translocated heterochromatin. Aka position effect.
    - Highly condensed
    - DNA is highly methylated
    - 15% of genome
    - Centromeric and telomeric regions
    - Constitutive or facultative
    - Highly enriched in DNA repeats
        - The only major chromosome that is highly condensed outside the metaphase is the barr body (females)
- **Euchromatin**
    - Dispersed
    - **Active**
    - Less methylated
    - Coding part of genome

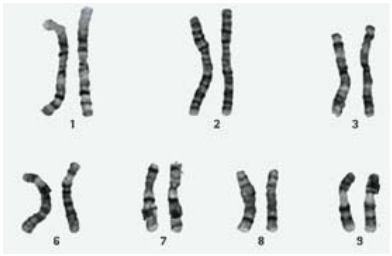
**12.5 - Chromosome-Banding Differentiates Regions Along The Mitotic Chromosome**

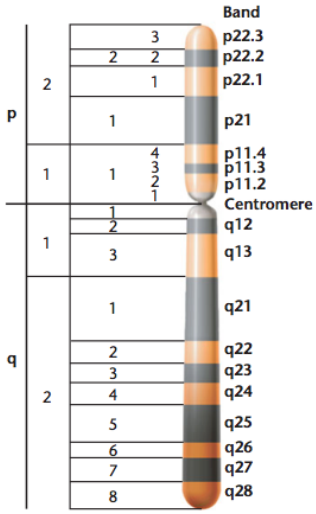
- Chromosome-Banding Techniques: Cytological procedures that allow for differential staining along the longitudinal axis of mitotic chromosomes, and the staining patterns look like the bands of polytene chromosomes.
- One of the first chromosome-banding techniques was using the Giemsa stain. It worked by detecting a specific area of the chromosome composed of heterochromatin using C-banding.

- C-Banding: Staining pattern where only the centromeric region of mitotic chromosomes take up the stain.



- The most useful banding technique produces a staining pattern differentially along the length of each chromosome, involving the digestion of the mitotic chromosomes with the proteolytic enzyme followed by Giemsa staining – these stains are called G-bands. These techniques show the heterogeneity and complexity of chromosomes along its length.





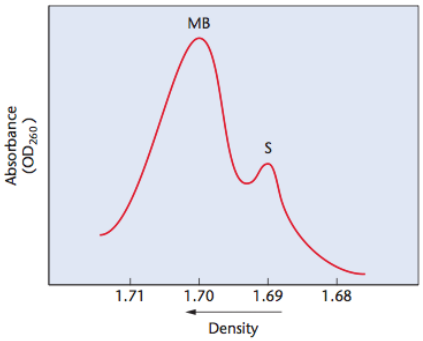

- A uniform nomenclature for human chromosome-banding patterns was established based on G-banding.
- These bands are important for cytogenetic analysis in humans.
- The nomenclature can be applied to each chromosome, allowing for identification of various regions along the p (short) and q (long) arms.
- The banding is unique to each chromosome, allowing a distinction to be made even between chromosomes that are identical in size and centromere placement. Even homologs can be distinguished from one another.

## 12.6 - Eukaryotic Genomes Demonstrate Complex Sequence Organization Characterized By Repetitive DNA

- Repetitive DNA: Much of the DNA sequences within eukaryotic chromosomes are repetitive in nature and various levels of repetition occur within the genome.
- Multiple-Copy Genes: Functional genes that are present in more than one copy.
- **The majority of repetitive sequences do not encode proteins.**
- There are 3 main categories of repetitive sequences: 1) Heterochromatin - centromeres and telomeres. 2) Tandem (adjacent) repeats of short and long DNA sequences. 3) Transposable sequences interspersed throughout the genome of eukaryotes.

**Satellite DNA**

- Nucleotide composition (ex. % of G=C vs. A=T) of specie's DNA is reflected in the DNA's density.
- A graph describes its composition as a single main peak (representing a single main band) of fairly uniform density.
- Usually there is one peak for the band of uniform density. Sometimes, there are other peaks, which differ in density, which are called satellite DNA**.**
- When analyzed, the DNA sequences fell under the categories of highly repetitive DNA, because it consisted of short sequences repeated a large number of times. They can be in specific areas, and is heterochromatic- flanking the centromere.




- Prokaryotes do not contain satellite DNA.
- Certain portions of DNA re-anneal faster than others. Rapid re-annealing is characteristic of multiple DNA fragments composed of identical nucleotide sequences, the basis of repetitive DNA.
- *In Situ* Molecular Hybridization: Molecular hybridization between an isolated fraction of DNA/RNA probes and the DNA contained in the chromosomes of a cytological preparation further concluded that: Satellite DNA, highly repetitive DNA, consists of short sequences repeated a large number of times clustered in heterochromatic areas (regions flanking centromeres).

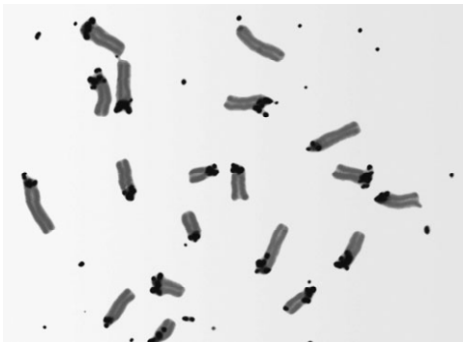
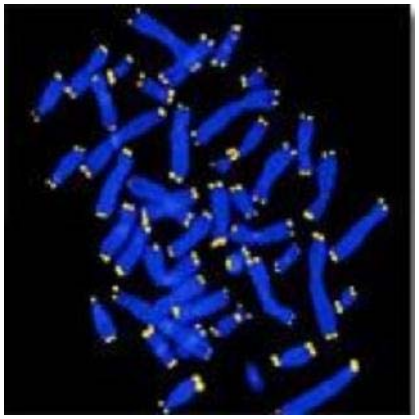


Figure 12-16: In situ hybridization between a radioactive probe representing mouse satellite DNA and mouse mitotic chromosomes. The grains in the autoradiograph are concentrated in the chromosome regions (the centromeres), revealing them to be the location of satellite DNA sequences.

**Centromeric DNA Sequences**

- Separation of homologs during meiosis & mitosis depends on centromeres (primary constrictions along eukaryotic chromosomes).
- CEN Region: Minimal region of the centromere that supports the function of chromosomal segregation, in which DNA binds a platform of proteins, including the kinetochore**,** which binds spindle fibres during cell division.
- Portions near the 3' end of CEN are more critical to centromere function because mutations in them disrupt centromere function. 3' is essential to the binding to the spindle fibre.
- Alphoid Family: one of the most recognized satellite DNA sequences in humans, found in the centromere regions. Alphoid sequences are common in humans.

**Telomeric DNA Sequences**

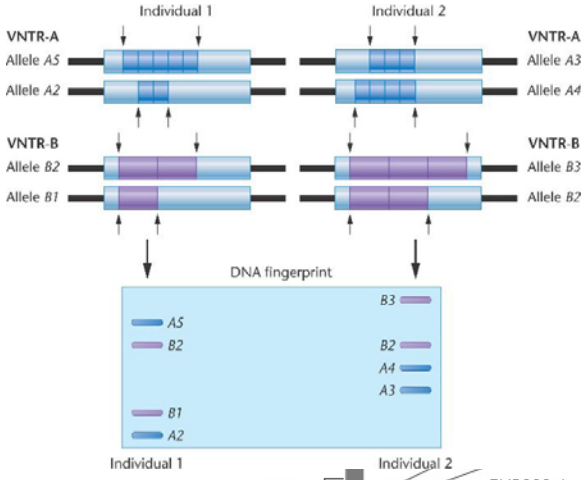


- Telomere: Structure that "caps" the ends of linear eukaryotic chromosomes, rendering chromosome ends inert in interactions with other chromosome ends and with enzymes that use DNA ends as substrates.
- Telomeric DNA sequences: Short tandem repeats. They add to the stability and integrity of the chromosome
- Help prevent degradation of DNA
- The sequences transcribed and RNA products, called TERRA (telomere repeat-containing RNA), facilitate methylation of the histone H3K9 and inhibit telomerase.

**Middle Repetitive Sequences: VNTRs & STRs**

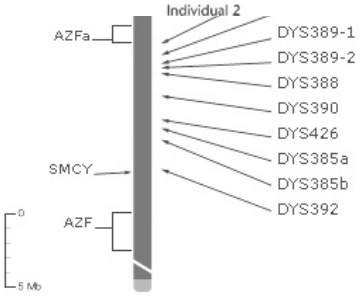
- Middle Repetitive DNA: Non-coding tandemly repeated or interspersed sequences.
- Variable Number Tandem Repeats (VNTRs): Repeating DNA sequences found within and between genes.
- Many are dispersed throughout the genome and often referred to as minisatellites.
- DNA Fingerprinting: The technique based on the variation in size (length) of the number of tandem copies of each specific sequence at each location between individual humans.

Microsatellites / Short Tandem Repeats (STRs): Another group of tandemly repeated sequences consisting of di-, tri-, tetra- and pentanucleotides that are dispersed throughout the genome and varying among individuals in the number of repeats present at any site.




- Example: $(CA)_n$, where *n* is the number of repeats between 5 and 50.
- These clusters serve as useful molecular markers for genome analysis.

**Repetitive Transposed Sequences: SINEs & LINEs**

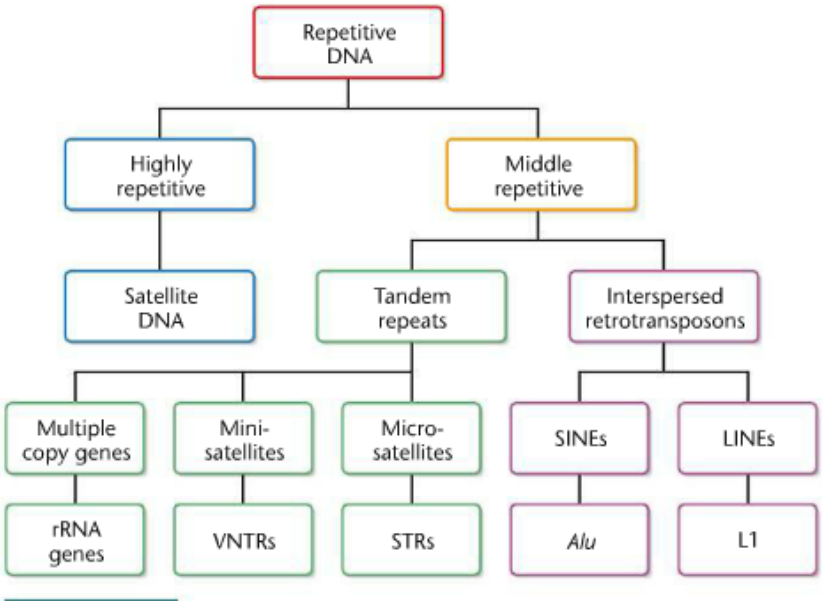



- Repetitive DNA sequences that are interspersed individually throughout the genome, rather than being tandemly repeated.
- Transposable Sequences: Ones that can move to different locations within the genome.
- Short Interspersed Elements (SINEs): Less than 500 base pairs long that are present 500,000 times or more in the human genome.
- Alu Family: Best characterized human SINE. It is 200-300 base pairs long and is dispersed uniformly throughout genome. Some members of the *Alu* family are transcribed into RNA.
- Long Interspersed Elements (LINEs): 6kb in length and present 850,000 times in the in human genome.

- L1 Family: The most prominent LINE in humans. The 5' end is highly variable. Its role within human genome is undefined.
- Retro-Transposons: Another name for LINEs because of the similarity of their transposition mechanism to that used by retroviruses.
- SINEs make up ~13% of the human genome, and LINEs make ~21%.

**Middle Repetitive Multiple-Copy Genes**

- Sometimes middle repetitive DNA includes functional genes present tandemly in multiple copies.
- Example: Many copies of the genes encoding ribosomal RNA exist.




**12.7 - The Vast Majority Of A Eukaryotic Genome Does Not Encode Functional Genes**

**What Proportion Of The Eukaryotic Genome Actually Encodes Functional Genes?**

- A genome is comprised of lots of repeating DNA.
- A lot of the DNA consists of single copy genes that are non-coding, like pseudogenes.
- Only a small amount of the genome actually codes for proteins but the proportion varies for each organism.
- Example: In humans, only 2% (20,000 functional genes) of the DNA sequence make up the genome.

**21.2 - DNA Sequence Analysis Relies on Bioinformatics Applications and Genome Databases**

**Bioinformatics**

- Uses computers combined with statistics to analyze data related to genomes (gene structure, sequence and expression) and proteins (structure, function).
- Before bioinformatics, data was gathered through cloning.
- GenBank is one of the important genomic databases used and its database is available to the public.
- Programs can be codon based.
- After a genome is sequenced, scientists need to map out the gene via annotation.

**Annotation**

- It is the identification of gene sequences within the genome along with their functions.
- It compares new genome sequence to known sequences through BLAST (Basic Local Alignment Search Tool).
- Only works to compare with known gene sequences, but not to identify new genes.
- Works easier for prokaryotes (no introns) than eukaryotes (with introns).
- It makes it harder when studying the genes that do not code for the proteins.
- There are clues with open reading frames (ORFs).

**Characteristics Used To Determine Gene Location**

- Splice Sites: Common nucleotide sequences within genomes that can be used to determine exon and intron boundaries.
- Open Reading Frames (ORFs): Area of genetic code that codes for amino acid sequence of polypeptide chain, characterized by:
- Promoter Sequences: TATA box, GC box or CAAT box.
- Initiation sequences: ATG, transcribed to AUG, the start codon in mRNA.
- Nucleotide Triplets: Exons and introns transcribed to mRNA that are read as codons (introns spliced out).
- Termination Sequences: Determine the end of a gene. TAA, TAG, or TGA transcribed to UAA, UAG or UGA, the stop codons in mRNA.
- Multiple open reading frames may exist on the same template strand.
- Codon Bias: Occurs in exons. One codon is favored over another (occurs more often) that codes for the same amino acid.
- Example: There are 4 different codons that code for valine (GUU, GUC, GUA and GUG), but GUG is used 38% of the time (codon bias in favor of GUG).

**Section 14.1 - Translation of mRNA Depends on Ribosomes and Transfer RNAs**

- Translation of mRNA is the process of polymerizing amino acids into polypeptide chains
- A tRNA molecule contains within its sequence three consecutive ribonucleotides (the anticodon) which complement, pair with, the mRNA codon.
    - Another region of tRNA is covalently bonded to its corresponding amino acid.
- tRNAs and mRNAs are bonded through hydrogen bonding so that the amino acids can be close enough together for a peptide bond to form in between the mRNA.
- Amino acids are polymerized into a polypeptide as an mRNA molecule proceeds continuously through the ribosome.

***Ribosomal structure***

- Both (large and small) subunits consist of one or more molecules of rRNA and ribosomal proteins
- When two subunits are associates in a single ribosome, sometimes it's called a monosome
- Differences between prokaryotic and eukaryotic ribosomes (monosomes) is prokaryotes have a 70s (50s and 30s) while eukaryotes have 80s (60s and 40s).

    - Subunits characterized on the basis of their sedimentation behaviour in sucrose gradients (the rate of migration when centrifuged). The rate is abbreviated as "S". Heavier molecule, larger S, essentially.
    - Sedimentation coefficients, "S", are not additive. Not how the 70S ribosome of prokaryotes is made up large (50S) and small (30S) subunits.
- Prokaryotes:
    - Large subunit: 23S + 5S + 31 ribosomal proteins
    - Small subunit: 16S + 21 ribosomal proteins
- Eukaryotes:
    - Large subunit: 28S + 5.8S + 5S + 46 ribosomal proteins
    - Small subunit: 18S + 33 ribosomal proteins
- The RNA components of the ribosome perform the catalytic functions in translation.
- The ribosomal proteins are thought to promote the binding of the various molecules involved in translation (fine-tuning the process)
- Many genomes contain genes that encode all three components. A single RNA molecule is produced (ex 30S), which is cleaved to produce the three components of a ribosome (ex 23S, 16S and 5S subunits). This way, equal quantities of all three are present as ribosomes are assembled.
    - Exception: in humans, the gene encoding 5S rRNA is located separately, on chromosome 1.

***tRNA structure***

- Small size and stable in the cell. Composed of only 75 to 90 nucleotides. Nearly identical structure in bacteria and eukaryotes. tRNAs are transcribed from DNA as larger precursors, which are cleaved afterwards.
- A number of nucleotides are unique to tRNA, containing a modification of on of the four N-bases. Modifications done after transcription.
- It is believed that the modified nucleotides enhance hydrogen bonding efficiency during translation
- The 3' end always ends with 2 cytosine's. The 5' end always ends with guanine
- The amino acid is covalently joined to tRNA at the 3' end.

***Charging tRNA***

- Charging occurs under the direction of enzymes called aminoacyl tRNA synthetases
- Because of the ability of the third member of the triplet codon to "wobble", the min. number of different tRNAs required is 31, but there is actually more than this number
- Charged tRNA's may participate directly in protein synthesis.
- Aminoacyl tRNA synthetases are highly specific enzymes because they only recognize one amino acid and the tRNAs corresponding to that amino acid, called iso-accepting tRNA's

**Section 21.3 - Functional Genomics Attempts to Identify Potential Functions of Genes and Other Elements in a Genome**

- ***Function genomics*** is the study of gene functions based on the resulting RNAs or possible proteins they encode and the functions of other components of the genome, such as gene-regulatory elements.
- One approach to assigning functions to genes is to use sequence similarity searches such as **BLAST**, which find alignments between the newly sequences genome and genes that have already been identified either in the same or different species
- If a genome sequence shows statistically significant similarity to the sequence of a gene whose function is known, it is likely that the genome sequence encodes a protein with a similar or related function
- Similarity searches are also used to identify homologous genes which are evolutionarily related.
    - Many open reading frames in the human genome were identified as protein-coding genes based on their alignment with related genes of known function in other species.
    - Ex. figure 21-9 compares portions of the human leptin gene with its homolog in mice. 85% identical.
- **Orthologs:** homologous genes in different species thought to have descended from a gene in a common ancestor.
- **Paralogs:** homologous genes in the same species. Paralogs have similar or identical functions. Arise from a gene-duplication event.
    - Ex: alpha- and beta-globin in humans.

***Predicting Function from Structural Analysis of Protein Domains and Motifs***

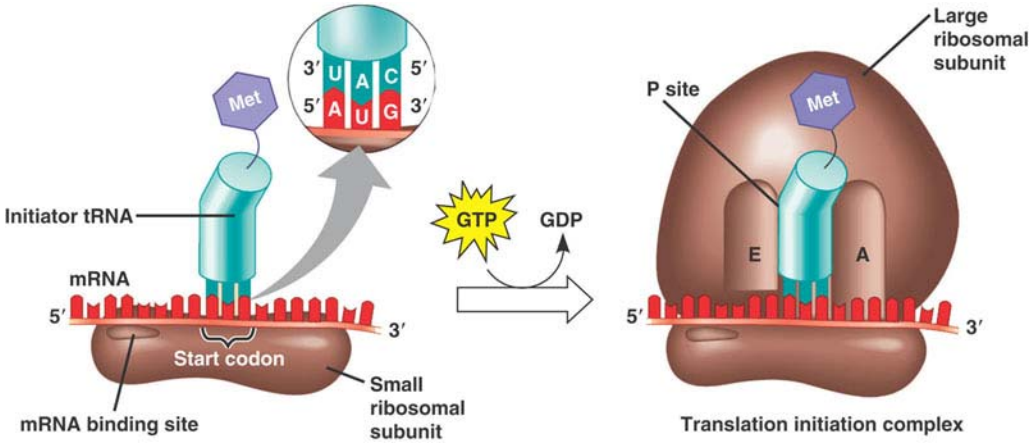
- Identification of protein **domains** such as ion channels and other structural aspects of a polypeptide that are encoded by a DNA sequence can be used to predict protein function
- **Motifs** (characteristic patterns), such as the helix-turn-helix, leucine zipper, zinc-finger motif on DNA-binding protein can be identified using bioinformatics software.
- Motif identification in a sequence is a common strategy for inferring the possible functions of a protein

***Investigators are using Genomics Techniques Such as Chromatin Immunoprecipitation to Investigate Aspects of Genome Function and Regulation***

- Interested in Chromatin immunoprecipitation (ChIP) and another related technique that couples ChIP to microarrays of gene chips, called ChIP-on-chip
- Used to map protein-DNA interactions
- Useful for identifying genes that are regulated by DNA-binding transcription factor

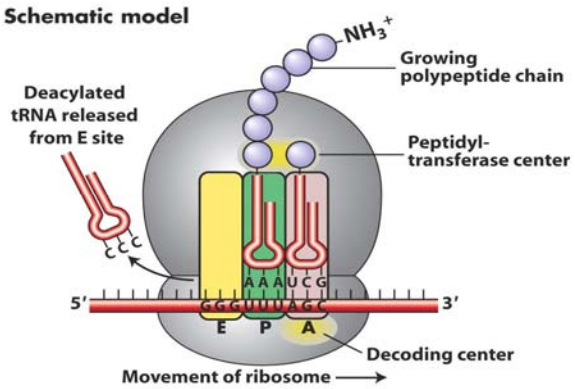
## 14.2 Translation of mRNA Can Be Divided into Three Steps

- Translation is a dynamic and continuous process.
- **INITIATION**



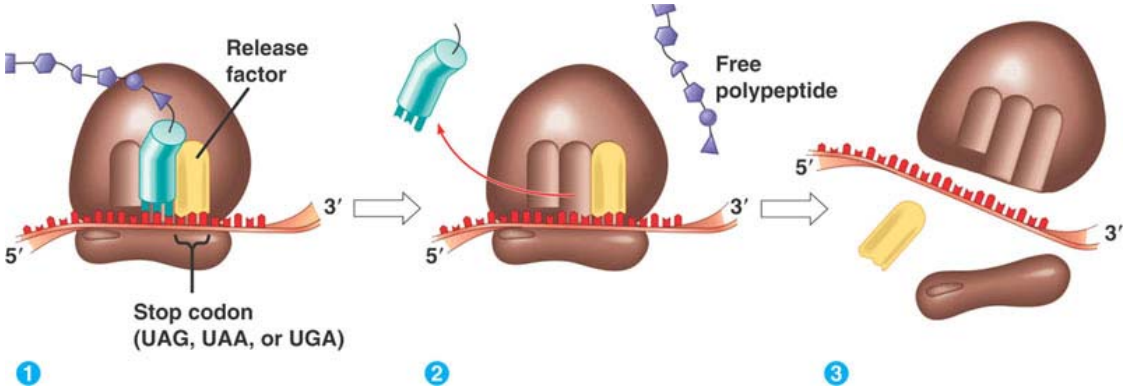

- Ribosomes serve as nonspecific workbenches for the translation process.
- They are divided into two subunits: large subunit and small subunit.
- In bacteria, initiation involves the small subunit, an mRNA molecule, a specific charged initiator tRNA, GTP, Mg2+, and three initiation factors (IFs) which enhance binding affinity of components.
- In prokaryotes, the initiation codon of 5'-(AUG)-3' calls for the modified amino acid N-formylmethionine (f-met).
- <u>Step one</u>
  - The small subunit binds to several initiation factors and the resulting complex binds to the mRNA at the initiation site (AUG).
    - In prokaryotes, this involves a sequence of 6 ribonucleotides (AGGAGG) that precedes AUG called the **Shine-Dalgarno sequence**; binds to small 16S unit of rRNA, facilitating initiation.
- <u>Step Two</u>
  - An initiation factor will enhance the binding of the charged formylmethionyl tRNA to the small subunit in response to the AUG triplet.
    - This step sets the reading frame.
- <u>Step three</u>
  - The aggregate represents the **initiation complex**, which binds with the large subunit
  - a molecule of GDP is hydrolyzed (for energy) and initiation factors are released.

- **ELONGATION**

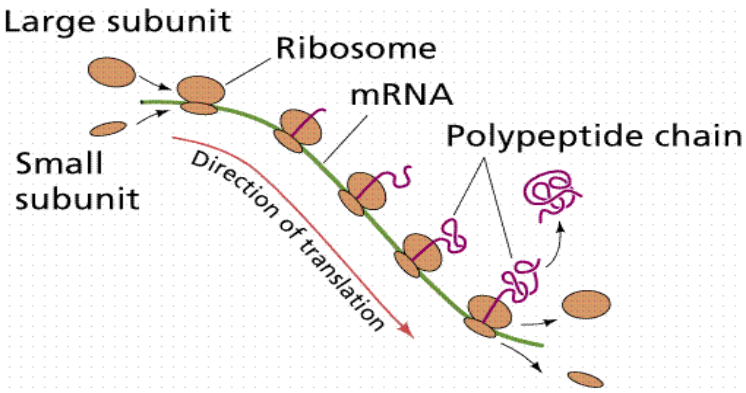



  - Once the subunits are assembled with mRNA, binding sites for the 2 charged tRNA molecules are formed; these are the **P (peptidyl) site** and the **A (aminoacyl) site**.
    - Charged initiator tRNA binds to the P site, provided that the AUG codon of mRNA is in the corresponding position of the small subunit.
  - Lengthening of polypeptide chain by 1 AA is called **elongation**.
  - <u>Step one</u>
    - The sequence of the 2nd triplet in mRNA dictates which charged tRNA will be positioned at the A site.
      - Once present, a reaction occurs within the large subunit catalyzing the formation of a peptide bond, linking 2 AAs together. Prior to this, the bond between the tRNA in the P site and its respective AA is broken by hydrolysis.
  - <u>Step two</u>
    - The newly formed dipeptide remains attached to the tRNA in the A site. Bond formation is catalyzed by **peptidyl transferase.**
      - Before elongation can repeat, the uncharged tRNA in the P site must be released. It moves briefly into the **E (exit) site.**
  - <u>Step three</u>
    - The entire mRNA-tRNA-AA2-AA1 complex shifts towards the P site by a distance of 3 nucleotides in an event called translocation.
      - Requires protein elongation factors (EFs)
  - <u>Step four</u>
    - As a result of translocation, the third codon of mRNA has moved into the A site and the site is now ready to accept the third charged tRNA.
  - <u>Step five</u>
    - Repetition of the sequence of elongation and translocation. The amino acid polypeptide chain keeps growing.
  - <u>Step six</u>
    - Once a polypeptide chain of about 30 AAs is assembled, it begins to emerge from the base of the large subunit.
    - The role of the small unit is to decode the codons in the mRNA, while the large unit is all about peptide bond synthesis.

- **TERMINATION**




  - Signaled by the presence of 1 of 3 triplet codons in the A site (UAG, UAA, or UGA). These are called **stop codons**, **termination codons**, or **nonsense codons.**
  - When stop codon is encountered, the finished polypeptide is still attached to the terminal tRNA at the P site, and the A site is empty.
  - <u>Step one</u>
    - The stop codons signal for action of a **GTP-dependent release factor**, which stimulates steps leading to release of the polypeptide chain from the terminal tRNA and from the translation complex.
  - <u>Step two</u>
    - The tRNA is released from the ribosome, which then dissociates into its subunits.
- **Polyribosomes**




  - As elongation proceeds, the initial portion of mRNA is free to bind with another small subunit to create a second initiation complex.
  - This process can be repeated several times creating a **polyribosome** (or **polysome**). Polypeptides are being translated in each ribosome.

**14.4: Translation is More Complex in Eukaryotes**

- The main difference between prokaryotes and eukaryotes: eukaryotic translation occurs on larger ribosomes whose rRNA and protein components are more complex.
- rRNA in both prokaryotic and eukaryotic cells share a *core sequence*, but in eukaryotes they are lengthened by the addition of *expansion sequences* (ES)
- In prokaryotes, transcription and translation are coupled (occur in the same place).
- In eukaryotes these processes are separated spatially (location) and temporally (time). Transcription occurs in the nucleus and translation in the cytoplasm.
    - More opportunities for regulation of gene expression in eukaryotes
- Differences in Initiation of Translation (focus on the mRNA)
    - 1. In eukaryotes, the 5' end of mRNA is capped with a **7-methylguanosine** residue (absent in prokaryotes), which is essential for initiation of translation.
    - 2. Many mRNAs contain a purine (A/G) three bases upstream from the AUG initiator codon, which is followed by a G (A/GNNAUGG). This sequence is called the **Kozak sequence**.
        - The sequence increases the efficiency of translation by interacting with initiator tRNA.
    - 3. Eukaryotic mRNAs require the posttranscriptional addition of a poly-A tail on their 3' end – a process called **poly-adenylation**.
        - mRNA with no tail are degraded in the cytoplasm. One exception: histone's mRNA is not poly-adenylated.
        - In eukaryotes, the AA f-met is not required as it is in prokaryotes, but the AUG triplet codon which encodes methionine is essential for formation of translation complex.
- Eukaryotic mRNAs are longer lived (exist for hours, compared to minutes→ more translation).
- They have similar proteins factors used during initiation, elongation and termination.
- Many eukaryotic factors are homologous to counterparts in prokaryotes.
- A great number of translation factors in eukaryotes are required during each step, and some are more complicated than in prokaryotes
- Many eukaryotic ribosomes are in association with the endoplasmic reticulum membranes, an association that facilitates secretion of newly synthesized proteins directly from the ribosome into the channels of the endoplasmic reticulum.
    - This is made possible by the tunnel found in the large subunit.
- Prokaryotic cytoplasm lacks the membranes that make up the endoplasmic reticulum.

- In prokaryotes, the polypeptide chain is released into the cytoplasm directly.
- Translation is a dynamic process as the subunits shift with relationships to one another as the mRNA proceeds through the ribosome.

## 14.10 – Posttranslational Modification Alters the Final Protein Product

- *Posttranslational modifications*: the modification of polypeptide chains after being synthesized – important for final protein product and function

**Table 1. 6 Main Posttranslational Modifications**

| Posttranslational modification | Description | Example |
|---|---|---|
| Removal or modification of N-terminus amino acid | - N-terminus: start of a polypeptide terminated by an amino acid with a free amine group<br>- also known as N-acetylation | - In eukaryotes methionine residue is often removed, N-terminal residue may be acetylated<br>- In bacteria the formyl group or entire formylmethionine residue is removed with enzymes. |
| Modification of individual amino acid residues. | - can affect function of the modified peptide chain<br>- might also allow for ionic bonding with other molecules | - Phosphates may be added to the hydroxyl groups of certain amino acids, allowing it to form ionic bonds with other molecules; this is the process of phosphorylation used to regulate many cellular activities. |
| Attachment of carbohydrate side chains, producing glycoproteins | - added covalently<br>- produce glycoproteins (category of cell-surface molecules) | -cell-surface molecules are those specifying the antigens in the ABO blood-type system |
| Trimming of polypeptide chains | - done through the use of enzymes | - after translation insulin gene is trimmed to 51-amino acid form into its active state |
| Removal of signal sequences | - Signal sequences are about 30 amino acids long and found at the N-terminal of a protein<br>- directs the protein to a specific location by determining its final destination in the cell (protein targeting)<br>- signal sequences are removed once the protein has reached its destination but before it functions | -proteins destined for secretion contain up to 15 hydrophobic amino acids preceded by positively charged amino acid at N-terminus |
| Polypeptide chains are often complexed with metals | - tertiary and quaternary levels of protein structure include/are dependent on metals to form properly | - Hemoglobin, has four iron atoms and four polypeptide chains |

**Protein Folding and Misfolding**

- The final 3-D structure of the protein is responsible for the protein's specific function
- Proteins are folded in 1 of 2 ways
    - A spontaneous process or chaperone mediated
    - Folding is a spontaneous process for most proteins (they achieve their thermodynamically stable conformation based on the chemical properties of the amino acid sequence)
    - For others, folding is dependent on **chaperones** (proteins that mediate the folding process by excluding the formation of incorrect patterns)
        - Chaperones bind to proteins but DO NOT become part of the final product
        - Chaperones in Drosophila are called heat-shock proteins
        - Chaperones are important in translation when it occurs on membrane-bound ribosomes
- Misfolding occurs sometimes even in the presence of chaperones
    - Misfolded proteins are a problem because they can be non-functional and accumulate and become detrimental to the cell
        - *Prion:* collection of misfolded protein
        - An example of a prion is mad cow disease (Creutzfeldt-Jakob disease in humans) is caused by prions in the brain. The normal protein is $PrP^{C}$ (folds into α helix), the misfolded version is $PrP^{Sc}$ (β-pleated sheet). When a $PrP^{C}$ protein touches a $PrP^{Sc}$ protein it changes to the misformed conformation.
- *Ubiquitin* is a small protein that will tag misfolded proteins as they exit the endoplasmic reticulum and end up in the cytoplasm.
    - Ubiquitin and a tagged misfolded protein is called a *polyubiquitin-protein complex*
    - The polyubiquitin-protein (tagged protein) complex moves to the *proteasome* where the misfolded protein is degraded by *proteases*

**15.1 – Gene Mutations are Classified in Various Ways**

- A mutation is the alternation in DNA sequence
- Mutations are the source for any variation in DNA
- Mutations can occur in protein coding regions (exons) or non-coding regions (introns)
- A mutation may or may not bring about a detectable change to the phenotype
- Mutations can occur in germ cells as well as somatic cells
    - *Germ cell mutations* are heritable, allowing for genetic diversity and evolution, as well as genetic diseases
        - Germ cell mutations may be transmitted to offspring as gametes and have the potential of being expressed in all cells of the offspring
        - X linked recessive mutations in gametes of a homogametic female can be expressed in homozygous male offspring, if he receives the affected X-chromosome

- *Somatic cell mutations* (any cell but germ cells) are not heritable, but they may cause altered cellular formation and can lead to altered functions.
    - A recessive autosomal mutation of a somatic cell in a diploid organism is not likely show up in the phenotype - the wild-type allele will mask it
    - Somatic mutations are most likely to be noticed when:
        I. They are dominant, X-linked in males
        II. The dominant/X-linked mutation happens early in development where the mutated cell will give rise to many more cells (in an adult the mutation can be masked by the other functioning cells)

- Spontaneous and Induced Mutations
    - *Spontaneous mutation*: Changes in the nucleotide sequence of genes that appear to have no known cause; they seem accidental or "random", may occur because of errors in the expression of the genes. Ex. A repetitive nucleotide sequence may be misread by an RNA-polymerase and will add the incorrect sequence of bases in the complementary RNA strand.
        - Often occur during enzymatic process of DNA replication
        - The spontaneous mutation rate is very low for all organisms, but varies a lot between different organisms
        - The spontaneous mutation rate can vary from gene to gene (even within the same species)
    - *Induced mutations*: Result from external factors, and can happen naturally (UV radiation from the sun) or artificially (causing mutation with x-rays)
- *Mutation rate* is the likelihood that a gene will become mutated after one generation/ the formation of one germ cell
- DNA sequences that are much more susceptible to mutations are called *mutation hot spots*

**Fluctuation Test: Are Mutations Random or Adaptive?**

- Basic concept: mutations occur randomly and are not directed by the environment in which the organism lives
- In other words, mutations arise spontaneously in the absence of selective pressure than being the consequence of selective pressure
- Luria-Delbruck Fluctuation Test is an experiment that showed that mutations do not occur as part of an adaptive mechanism but occur randomly

**Classification Based on Location of Mutation**

- Mutations can be classified based on cell type or on chromosomal locations on which they occur
- *Somatic mutations* occur on any cell except germ cells
- *Autosomal mutations* are mutations that happen within genes on the autosomes
- X-linked or Y-linked mutations happen in genes on the X or Y chromosome
- Inherited dominant autosomal mutations will be expressed phenotypically in the first generation

**Classification based on Type of Molecular Change**

- *Point mutation/Base substitution* is a change in one base pair to another in DNA
    - If the mutation causes a change to the codon so that a different amino acid is coded it is called a *missense mutation*
    - A *conservative missense mutation* is when the altered codon specifies for an amino acid with similar chemical properties to the original amino acid
    - A *non-conservative missense mutation* is when the mutation changes a codon to another amino acid and may result in the loss of function in the protein
    - If the mutation causes a stop codon, stopping translation early, it is called a *nonsense mutation*
    - If the mutation causes a change to the codon but still codes for the same amino acid (due to degeneracy of the genetic code) it is called a *silent mutation*
- *Frameshift mutations* are the insertion or deletion of one or more nucleotides in any part of the gene, which causes a change in the reading frame. The result of these mutations can be very severe since they can change a large portion of the amino acid chain. A frameshift mutation found at the beginning of a reading frame is much more hazardous than that found at the end, because it will shift the WHOLE reading frame over, as oppose to a couple of nucleotides if found at the end.



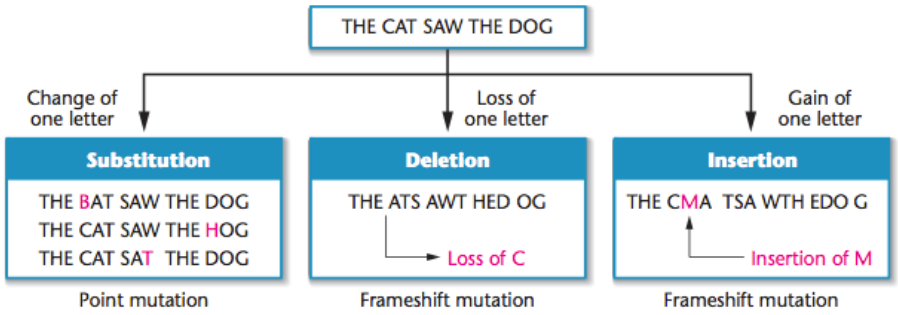



FIGURE 15–1 Analogy showing the effects of substitution, deletion, and insertion of one letter in a sentence composed of three-letter words to demonstrate point and frameshift mutations.

**15.2 Spontaneous Mutations Arise From Replication Errors and Base Modifications**

DNA REPLICATION ERRORS AND SLIPPAGE

- DNA polymerases often correct replication errors with **exonuclease** proof-reading
    - Errors in nucleotides can still persist. These errors lead to mutations.
- Point mutations: the main result of mispairing errors
- Tautomers*:* Different forms bases can take depending on the migration of a H atom within the base. Ex. Enol/Keto forms
    - This increases the chance of mispairing during replication
- Tautomerization: Keto/enol forms of thymine or guanine and amino/imino forms of adenine and cytosine. These shifts change the hydrogen bonding ability of the bases (T binds G and A binds C)

- Nucleotide insertions and deletions can also occur
    - Caused when:
        - one strand becomes displaced
        - DNA polymerase slips or accidentally repeats
    - Repeats leads to insertion of one or more nucleotides = unpaired loop
    - Can lead to frameshift mutations/amino acid insertions & deletions
- Replication slippage common in repeated sequences

## TAUTOMERIC SHIFTS

- Tautomeric shifts*:* different chemical forms, with only a proton shift as a difference
- Can lead to base-pair changes & mutations
    - The shifts change the bonding structure, meaning it can bond with bases it normally wouldn't (ex: A-C, G-T)

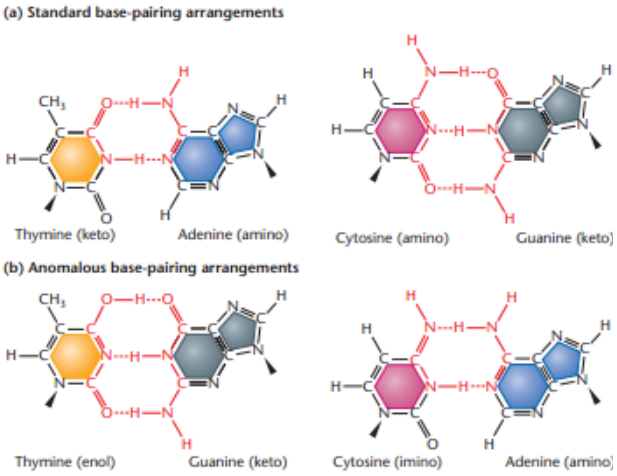
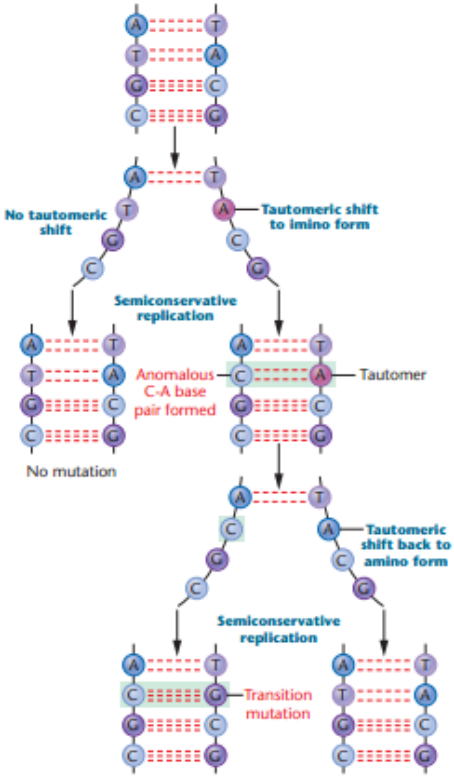



FIGURE 15–2 Standard base-pairing relationships (a) compared with examples of the anomalous base-pairing that occurs as a result of tautomeric shifts (b). The long triangle indicates the point at which the base bonds to the pentose sugar.




- Results in a point mutation as the two strands/mismatched pair would be separated

## DEPURINATION AND DEAMINATION

- Depurination*:* Loss of one of the nitrogenous bases in a DNA double-helix
    - Usually purine (guanine or adenine) is lost
    - Creates an apurinic site (a site where DNA has neither a purine nor a pyrimidine) Could lead to a random nucleotide being introduced via DNA polymerase
- Deamination*:* The amino group of a cytosine or adenine becomes a Keto group
    - Cytosine becomes uracil, adenine becomes hypoxanthine

**OXIDATIVE DAMAGE**

- Reactive oxidants can be generated via cellular metabolism as well as radiation
- Leads to modifications in DNA: base modification, base deletion, and single-stranded breaks
- Super hydroxides, hydroxyl oxides and hydrogen peroxide produce more than 100 types of modifications

**TRANSPOSONS**

- Transposons*:* (transposable genetic elements) DNA elements that can move within/between genomes if a lack of a complementary base sequence is present in the DNA
- Act as natural mutagens
- Can insert themselves into a gene, meaning they can change the reading frame or make an early stop codon
- Insertion in regulatory regions can change the expression of the gene
- Can also lead to: chromosomal damage (double-stranded breaks, inversions, and translocations)

**15.4 Single-Gene Mutations Cause A Wide Range Of Human Disease**

- Specific types of mutations contribute to human disorders.
- Most human genetic diseases are **POLYGENIC**
  - **Polygenic** : caused by variations in several genes
- Even a single base-pair change in one of the approx.20,000 human genes can lead to a serious inherited disorder.
  - These **MONOGENIC** diseases can be caused by many different types of single-gene mutations.

TABLE 15.3

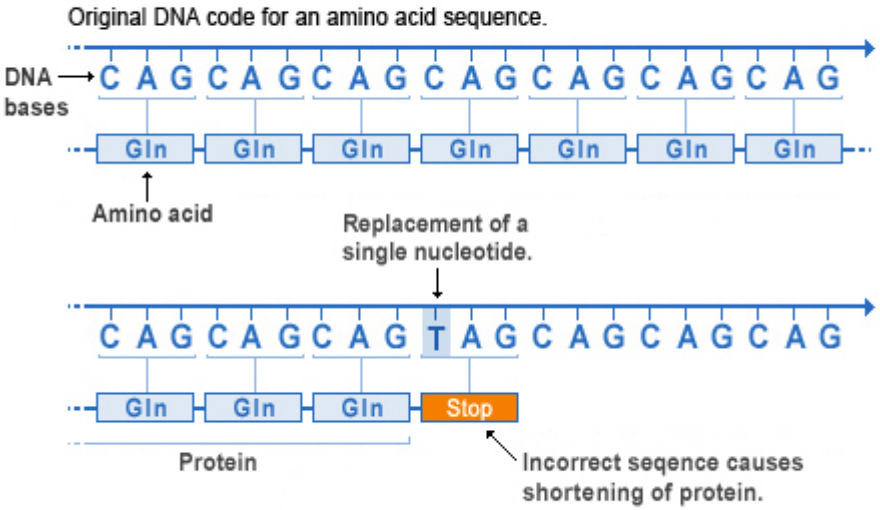
Examples of Human Disorders Caused by Single-Gene Mutations

| Type of DNA Mutation | Disorder | Molecular Change |
|---|---|---|
| Missense | Achondroplasia | Glycine to Arginine at position 380 of *FGFR2* gene |
| Nonsense | Marfan syndrome | Tyrosine to STOP codon at position 2113 of *fibrillin-1* gene |
| Insertion | Familial hypercholesterolemia | Various short insertions throughout the *LDLR* gene |
| Deletion | Cystic fibrosis | Three base-pair deletion of phenylalanine codon at position 508 of *CFTR* gene |
| Trinucleotide repeat expansions | Huntington disease | >40 repeats of (CAG) sequence in coding region of *Huntingtin* gene |

**TABLE 15.3**: lists some examples of the types of single-gene mutations that can lead to serious genetic diseases.

- Geneticists estimate that approx. 30% of mutations that cause human diseases are single base-pair changes that create **nonsense mutations.**
    - These mutations code for a prematurely terminated protein product (early stop codon), and also trigger rapid decay of the mRNA.

Nonsense mutation




U.S. National Library of Medicine

- **Missense mutations**
- Change in one base causing only one amino acid substitution

- **Frameshift mutations**
- Caused by insertion or deletion of a base(s)
- Changes the reading frame of the codons
- Usually affecting all bases following (unless a codon is added/deleted)
- Potentially drastic alterations in the protein sequence and create internal nonsense codons.

**Table 1** Types of Mutations

| Category | Type | Result |
|---|---|---|
| point mutation | **substitution**<br>AAG CCC GGC AAA<br>AAG ACC GGC AAA | **missense mutation**<br>only one amino acid substituted |
| | **deletion**<br>AAG CCC GGC AAA<br>AAC CCG GCA AA | **frameshift mutation**<br>can result in many different amino acids substituted or a stop codon read (nonsense mutation) |
| | **addition/insertion**<br>AAG CCC GGC AAA<br>AAG ACC GGG CAA A | |

- Other common disease-associated mutations affect the sequences of gene promoters, mRNA splicing signals, and other noncoding sequences that affect transcription, processing, and stability of mRNA or protein.
- **Recent study:** About 15 % of all point mutations that cause human genetic diseases result in abnormal mRNA splicing.
- 85% of these splicing mutations alter the sequence of 5' and 3' splice signals.
- 15% creates new splice sites within the gene.
- Splicing defects often result in degradation of the abnormal mRNA of creation or abnormal protein products.

## SINGLE BASE-PAIR MUTATIONS AND β -THALASSEMIA

- β thalasemmia (inherited blood disorder) provides an example of how one inherited disease can arise from a large number of possible mutations.
- **β -thalassemia**
- An **inherited autosomal recessive blood disorder** resulting from a reduction or absence of hemoglobin.
- The **most common single-gene disease in the world**, affecting people worldwide, but especially populations in Mediterranean, North African, Middle Eastern, Central Asian, and Southeast Asian countries.
- People with B-thalassemia have varying degrees of **anemia** (from severe to mild)

- **Symptoms:** weakness, delayed development, jaundice, enlarged organs, and often a need for frequent blood transfusions.
- Mutations in the β-globin gene (HBB gene) cause β-thalassemia.
- The HBB gene encodes the 146 amino acid B-globin polypeptide.
- 2 β-globin polypeptides associate with 2 A-globin polypeptides to form the adult hemoglobin tetramer.
- 250 diff. mutations in the HBB gene have been discovered that cause β-thalassemia
- But most cases worldwide are associated with about 20 of these mutations.
- Most mutations change a **single nucleotide** within or surrounding the HBB gene, or create **small insertions and deletions.**
- Each population affected by β-thalassemia has a unique mix of mutations.

- Ex: most prevalent mutation in a Sardanian population (a mutation that accounts for more than 95 % of cases) - is a single base-pair change at codon 39, creating a nonsense mutation and premature termination of the B-globin polypeptide.
- In contrast, a study of β- thalassemia mutations in a Yugoslavian population revealed 14 diff. mutations; with only 3 accounting for 75% of all cases (all in intron 1 splice signals).
- The types of mutations that cause B-thalassemia not only affect the B-globin amino acid sequence (missense, nonsense, and frame shift mutations, but also alter HBB transcription efficiency, mRNA splicing and stability, translation, and protein stability.

- This table provides a summary of the types of single gene mutations that cause β-thalassemia.
- **More than half** : single base pair changes. The remainder are short insertions, deletions, and duplications.

TABLE 15.4

**Types of Mutations in the *HBB* Gene that Cause *β*-thalassemia**

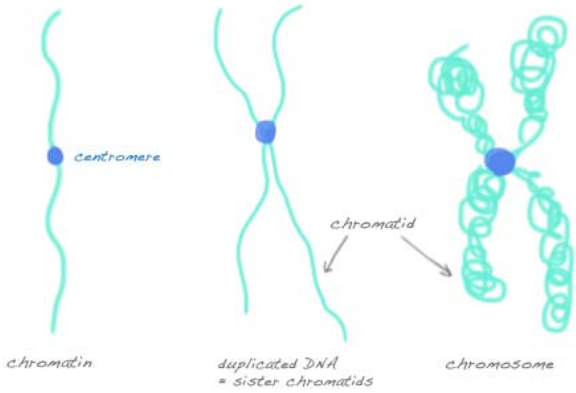
| Number of Mutations Known | Gene Region Affected | Description |
|---|---|---|
| 22 | 5′ upstream region | Single base-pair mutations occur between −101 and −25 upstream from transcription start site. For example, a T->A transition in the TATA sequence at −30 results in decreased gene transcription and severe disease. |
| 1 | mRNA CAP site | Single base-pair mutation (A->C transversion) at +1 position leads to decreased levels of mRNA. |
| 3 | 5′ untranslated region | Single base-pair mutations at +20, +22, and +33 cause decreases in transcription and translation and mild disease. |
| 7 | ATG translation initiation codon | Single base-pair mutations alter the mRNA AUG sequence, resulting in no translation and severe disease. |
| 36 | Exons 1, 2, and 3 coding regions | Single base-pair missense and nonsense mutations, and mutations that create abnormal mRNA splice sites. Disease severity varies from mild to extreme. |
| 38 | Introns 1 and 2 | Single base-pair transitions and transversions that reduce or abolish mRNA splicing, and create abnormal splice sites that affect mRNA stability. Most cause severe disease. |
| 6 | Polyadenylation site | Single base-pair changes in the AATAAA sequence reduce the efficiency of mRNA cleavage and polyadenylation, yielding long mRNAs or unstable mRNAs. Disease is mild. |
| > 100 | Throughout and surrounding the *HBB* gene | Short insertions, deletions, and duplications that alter coding sequences, create frameshift stop codons, and alter mRNA splicing. |

## Mutations Caused By Expandable DNA Repeats

- Some mutant genes contain expansion of **TRINUCLEOTIDE REPEAT SEQUENCES**
- **Trinucleotide repeat sequences**: specific short DNA sequences repeated many times.
- Normal individuals have fewer than 30 repetitions of these sequences
- Individuals with over 20 diff. human disorders appear to have abnormally large numbers of repeat sequences (often over 200) within and surrounding specific genes.
- Examples of diseases associated with these trinucleotide repeat expansions are:
  - Fragile X syndrome
  - Myotonic dystrophy
  - Hunnington disease

- When trinucleotide repeats such as $(CAG)_n$ occur within a coding region, they can be translated into long tracks of glumatine.
- These glumatine tracks may cause the protein to aggregate abnormally.

- When the repeats occur outside regions, but within the mRNA, it is thought that the mRNAS may act as “toxic” RNAs that bind to important regulatory proteins, sequestering them away from their normal functions in the cell.
- Another possible consequence of long trinucleotide repeats is that the regions of DNA containing the repeats may become abnormally methylated, leading to silencing of gene transcription.
- The mechanisms by which the repeated sequences expand from generation to generation are of great interest.
- It is thought that expansions may result from either errors during DNA replication or errors during DNA damage repair.




## 2.2 CHROMOSOMES EXIST IN HOMOLOGOUS PAIRS IN DIPLOID ORGANISMS

- **Chromosomes** are most easily visualized during mitosis.
  - Distinctive shapes and lengths are visible when observed carefully.
- Each chromosome contains a condensed region called a **centromere**.
  - The location of the centromere establishes the general appearance of each chromosome.

| Centromere location | Designation | Metaphase shape | Anaphase shape |
|---|---|---|---|
| Middle | Metacentric | Sister chromatids, Centromere | Migration |
| Between middle and end | Submetacentric | p arm, q arm | |
| Close to end | Acrocentric | | |
| At end | Telocentric | | |

**FIGURE 2–3** Centromere locations and the chromosome designations that are based on them. Note that the shape of the chromosome during anaphase is determined by the position of the centromere during metaphase.

- This figure shows chromosomes with centromere placements at different distances along their length.
- Arms of the chromosome extend from each side of the centromere .
- Chromosomes are classified as **metacentric, submetacentric, acrocentric, or telocentric** on the basis of the centromere location.

- Shorter arm **= p arm** ( p for petite )
- Longer arm = **q arm** (because q is the next letter after p in alphabet).

- **Somatic cells** derived from members of the same species have identical number of chromosomes.
- This represents the d**iploid number (2n).**
- All chromosomes except sex chromosomes exist in pairs and the members of each pair are called **homologous chromosomes.**
  - So: for each chromosome with a specific length and centromere placement, exists another with identical features.
- Exceptions
  - Many bacteria and viruses have one chromosome
  - Yeast, mold, and certain plants spend most of their life in haploid stage.
    - This means they contain only ONE member of each homologous pair of chromosomes during most of their lives.

**Figure 2-4**

- Shows the physical appearance of different pairs of homologous chromosomes
- This is a **karyotype**
  - Meaning the human mitotic chromosomes have been photographed, cut out of the print, and matched up, creating this display.
  - As noted, humans have a **2n number of 46 chromosomes**.
  - Exhibit a diversity of sizes and centromere placements
  - Each chromosome in the karyotype is a structure consisting **of 2 parallel sister chromatids** connected by a centromere.

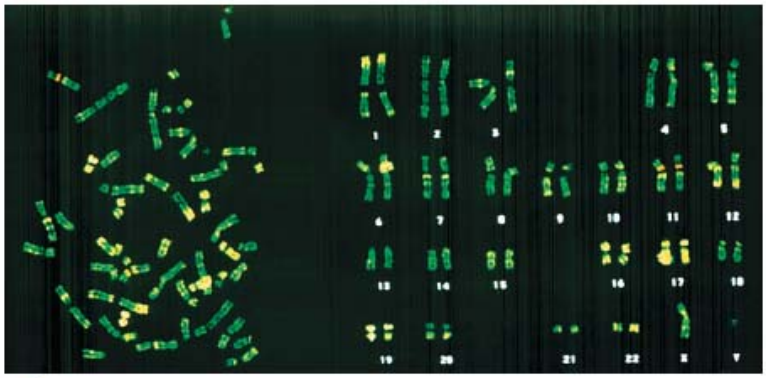


**FIGURE 2–4** A metaphase preparation of chromosomes derived from a dividing cell of a human male (left), and the karyotype derived from the metaphase preparation (right). All but the X and Y chromosomes are present in homologous pairs. Each chromosome is clearly a double structure consisting of a pair of sister chromatids joined by a common centromere.

- **The haploid number (n)** = half of the diploid number
- The genetic info in a haploid set of chromosomes constitutes the genome of the species.
  - This includes copies of all genes and large amount of noncoding DNA.

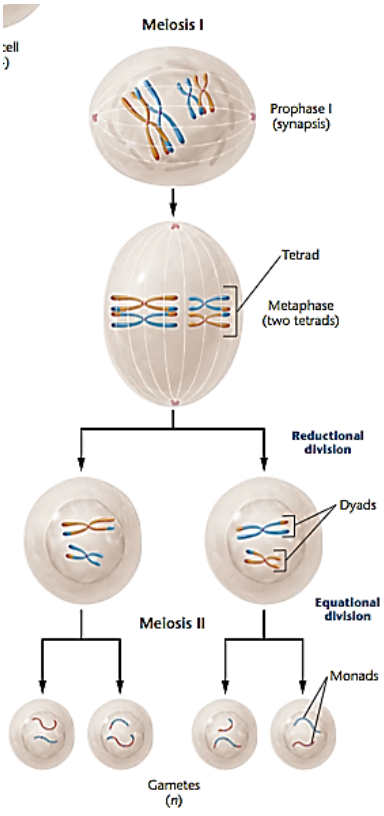
- **Homologous** chromosomes have important genetic similarities.
  - Contain identical gene sites along their lengths
  - Each site is called a locus (pl. loci)
  - So they are identical in the traits that they influence and in their genetic potential.

- **In sexually reproductive organisms :**
  - One member of each pair is derived from the maternal parent ( ovum)
  - And the other member is derived from the paternal parent ( sperm)

- So, each diploid organism contains 2 copies of each gene as a consequence of **biparental inheritance**- inheritance from 2 parents.
- The members of each **pair of genes** influencing the same characteristic or trait have to be identical
- But, many diff. alternative forms of the same gene, called **alleles** can exist.

- **Sex Determining chromosomes**
  - Not homologous in size, centromere placement, arm ratio or genetic content.
- Ex: In humans female carry 2 homologous X chromosomes, males usually carry one Y and one X (exception XXY).
  - These X and Y chromosomes are not strictly homologous
  - Y is smaller and lacks most of the gene loci contained on the X.
  - Still, they contain homologous regions and behave as homologs in meiosis so that gametes produced by males receive either one X or one Y chromosome.

**Section 2.4 & 2.5**




**2.4 Meiosis reduces the Chromosome Number from Diploid to Haploid in Germ Cells and Spores**

- Two divisions of meiosis (I and II) produce gametes with a haploid (n) set of chromosomes.
- Four haploid non-identical cells are produced (genetic variation)
- During fertilization, these gametes combine to restore a diploid (2n) cell (zygote).

**Meiosis I:**

-The first meiotic division (meiosis I) is a reductional division where each tetrad (homologous pair) separates into 2 dyads (separate homologs).

-Followed by another division, Meiosis II

**1) Prophase I** is divided into 5 substages:

i) **Leptonema stage:** Chromatin material condenses and chromosomes become visible under microscope. On each chromosome, chromomeres are present. A process called homology search also begins: each chromosome searches for its homolog (pairing of homologues).

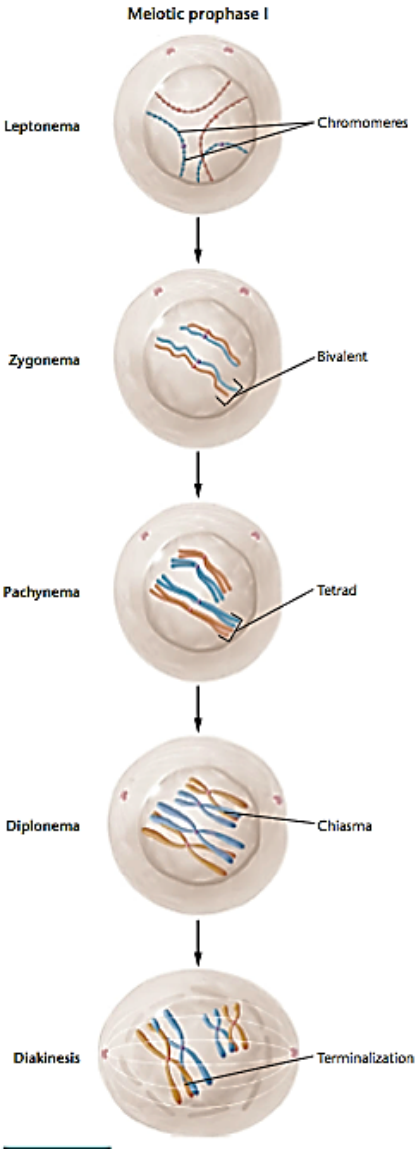
ii) **Zygonema stage:** Chromosomes continue to shorten and thicken. A process called rough pairing occurs: homologous chromosomes are paired and are aligned with one another. A synaptonemal complex (attachment) begins to form between the homologs. At the end of this stage, bivalents (paired homologs) are formed.

iii) **Pachynema stage**: Chromosomes continue to coil and shorten. The synaptonemal complex develops and synapsis begins: the aligned homologous chromosomes become closer together and this forms a tetrad.

iv) **Diplonema stage:** Each tetrad consists of 4 chromatids (2 pairs of sister chromatids) which are more visible in this stage. The sister chromatids slightly separate but are still connected at chiasma points of crossing over. Crossing over (exchange of genes between non-sister chromatids) generates genetic variability.

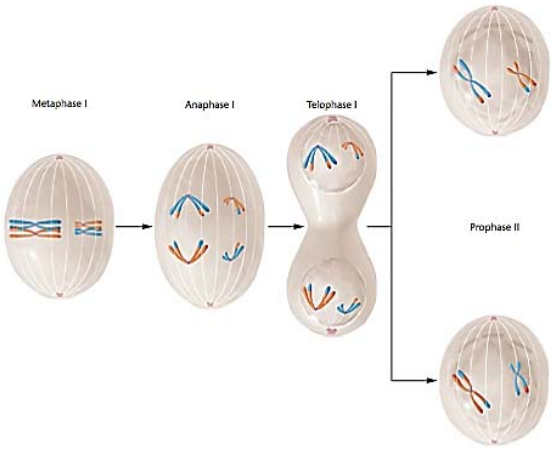
v) **Diakinesis stage:** Chromosomes are pulled further apart and the chiasmata move towards the ends of the tetrad. This is called terminalization. The nucleolus and nuclear envelope break down. Centromeres of each tetrad attaches to spindle fibers.

- At the end of prophase I, the centromeres of tetrads, are on the metaphase plate of the cell.




## Metaphase, Anaphase and Telophase I

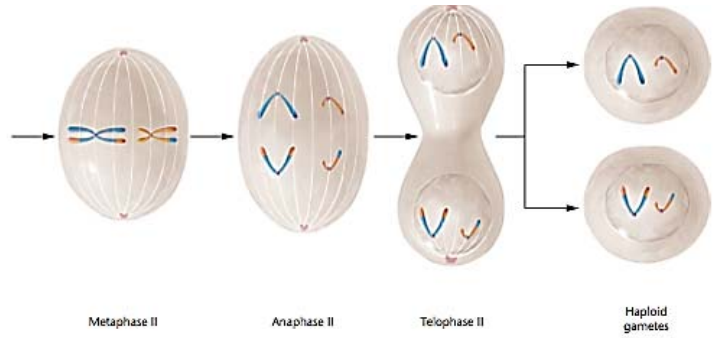
2) **Metaphase I:** Chromosome shortening and thickening is done. Alignment of each tetrad to the center (metaphase plate) is done by interacting with spindle fibers. Alignment of each tetrad is random (Law of Independent Assortment).
3) **Anaphase I:** Homologous pairs separated, forming haploid number of chromosomes at each pole. Nondisjunction occurs if separation (of homologs from one another) is not achieved in meiosis.
4) **Telophase I:** A nuclear membrane reforms around the dyads. The cell membrane gets "pinched" and the organelles and DNA are split. Following telophase is cytokinesis when the cell membrane is sealed off and the mother cell becomes two daughter cells.

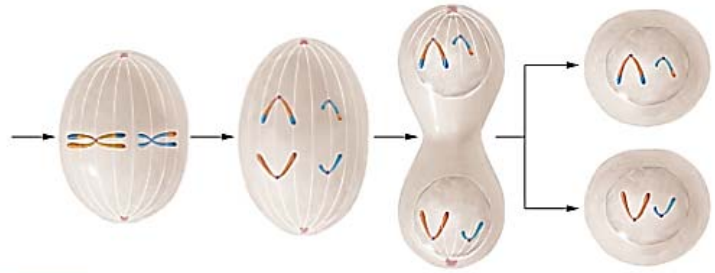


**Meiosis II:**

-The second division, meiosis II, is described as an equational division (# of centromeres remains equal).
-Similar to mitosis (but haploid)
-Each dyad splits into two monads.

1) **Prophase II:** each dyad consisting of one pair of sister chromatids is attached at the centromere.
2) **Metaphase II:** the centromeres connect to spindle fibers and are positioned at the equatorial plate.
3) **Anaphase II:** the sister chromatids of each dyad are pulled to opposite poles. Each chromosome is now a monad.
4) **Telophase II:** Once the chromosomes reach the end of the poles, they are called monads. Cytokinesis occurs and this results in 4 haploid cells in total.



**2.5 The Development of Gametes Varies in Spermatogenesis Compared to Oogenesis**

- Production of male gamete=spermatogenesis, production of female gamete=oogenesis.

**Spermatogenesis**

- Occurs in the testes and begins with the enlargement of a diploid germ cell called a spermatogonium (2n).
- This grows to become a primary spermatocyte (2n) that undergoes meiosis I.
- The products, secondary spermatocytes (n), undergo meiosis II.
- Each secondary spermatocyte produces two haploid (n) spermatids.
- Spermatids undergo developmental changes (spermiogenesis) and become spermatozoa (sperm).
- Each sperm cell is haploid and has equal amounts of cytoplasm.

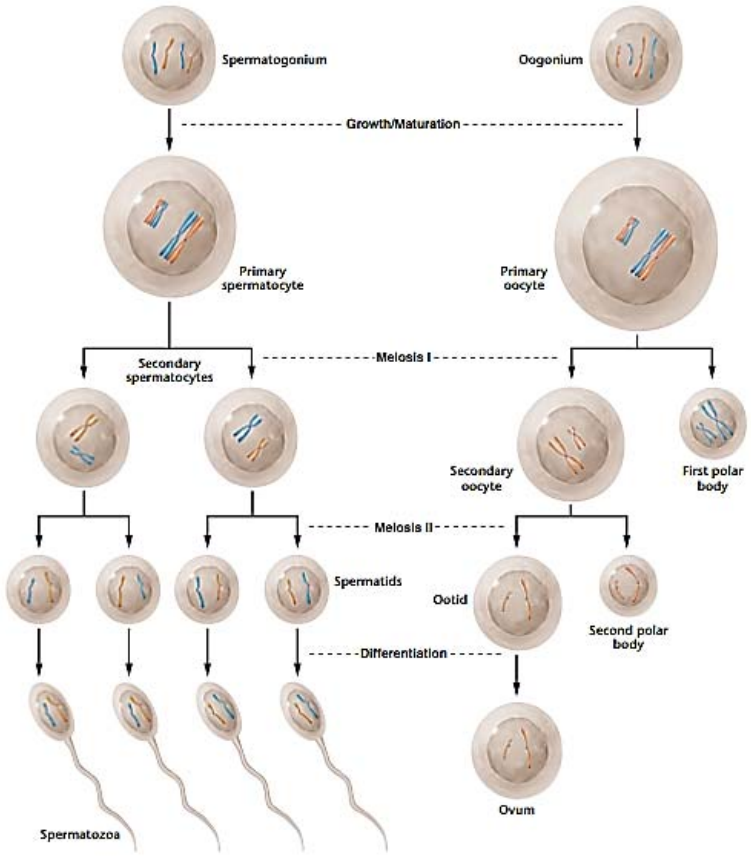
## Oogenesis

- In each division, almost all the cytoplasm of the primary oocyte (2n) , which is derived from the oogonium (2n germ cell), goes to **one** of the daughter cells (extra cytoplasm needed to nourish embryo).
- In telophase I, the dyads of one pole are pinched off to form the first polar body (n).
- The other daughter cell (not the polar body) is called the secondary oocyte (n).
- In meiosis II, an ootid (n) and a second polar body (n) is formed.
- The ootid differentiates into the mature ovum.
- The second division is only completed after fertilization (in humans).

### 2.6 Meiosis is Critical to Sexual Reproduction in All Diploid Organisms

- Meiosis is critical to successful sexual reproduction in all diploid organisms

- Causes diploid amount of genetic info to be reduced to haploid
- Main function in animals: Leads to formation of gametes
- Plants: Haploid spores are produced, than formed into haploid gametes
    - Also have alternation of generations life cycle – alternate between diploid sporophyte and the haploid gametophyte stages
    - Fungi: spend predominant amount of time in haploid vegetative state – to spread they undergo meiosis and then mitotic cell division




- Diploid organism store genetic info in homologous pairs of chromosomes

- One member of pair is from maternal parent and other from paternal

- Haploid cells can have paternal or maternal component
- Crossing over in the first meiotic prophase rearranges alleles from paternal and maternal components

- The components then separate and divide independently into gametes – source of genetic variability

- In fungi: exist mostly as haploid cells

- Arise by meiosis and reproduce by mitosis

- In multicellular plants: alternates between diploid sporophyte stage and haploid gametophyte stage (alternation of generations)

- Meiosis and fertilization act as "bridges" between sporophyte and gametophyte stages

**3.5 Mendel's Work Was Rediscovered in the Early Twentieth Century**

- Mendel started his work in 1856 (Brunn Society of Natural Science) but was not acknowledged for 35 years

- He used mathematical analysis of probability events – this was rarely done before
- His conclusions did not coincide with existing hypotheses about variation among organisms
    - At the time, evolutionary theorists believes in continuous variation

continuous variation: offspring were blend of parents' phenotypes

- Charles Darwin and A R Wallace defined the theory of continuous variation
- Mendel thought variation arose from discontinuous variation

discontinuous variation: dominance-recessive relationship between separate units

- He believed a $F_2$ offspring of a dihybrid cross could express traits combined from previous existing unit factors

Ex. pink flowers instead of simply red or white

- Mendel's generation did not realize he explained how variation was transferred to offspring
- Instead they interpreted his work as why certain phenotypes survived over others
- Obscured by Darwin's theory of natural selection

<u>The Chromosomal Theory of Inheritance</u>

- Flemming's discovery of chromosomes in nuclei in salamander cells helped for recognizing Mendel's work
    - Lead to the idea of units of inheritance (genetic material) in the nucleus

- Hugo de Vries, Karl Correns and Erich Tschermak also did research drawing conclusions similar to Mendel
- Walter Sutton and Theodor Boveri linked behaviour of chromosomes during meiosis to Mendelian principles of segregation and independent assortment

- Started the chromosomal theory of inheritance

Chromosomal theory of inheritance: genetic material in living organisms is contained in chromosomes – based on findings using Mendel's principles of segregation/independent assortment

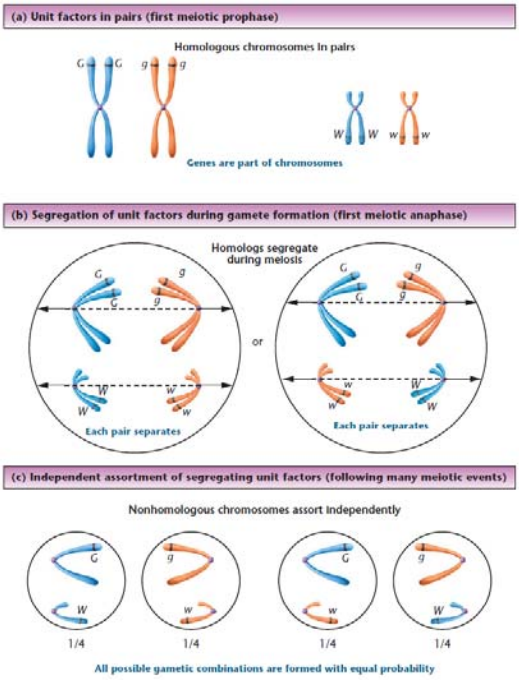
**Unit Factors, Genes and Homologous Chromosomes**

- Diploid number (2n): number of chromosomes in diploid organisms, characteristic for each species
- Somatic cells contain a certain number of chromosome (species dependent) – the 2n or diploid number
- Halved during meiosis (gamete formation) and re-established during fertilization (two gametes combine)
- Chromosomes are paired by morphological appearance and behaviour
- Each pair of homologous chromosomes is separated and contained in a gamete

  - Number of chromosomes in a gamete is haploid
  - One member is maternal and other is paternal
  - Fuse during fertilization – offspring has diploid number

- Independent behaviour of pairs of unit factors is due to their presence on separate pairs of chromosomes
- Unit factors: genes located on homologous pairs of chromosomes
- There are many more genes than chromosomes
- This causes each homolog to carry genetic info for more than one trait
- Chromosomes are composed of a large number of linearly ordered, information-containing genes
- Mendel's paired unit factors make up a pair of genes on a pair of homologous chromosomes
- locus (pl. loci): location on a chromosome where any gene occurs
- Different alleles of a gene contain slightly different genetic information that determines the same characteristic (G and g being green and yellow)

  - Most genes have more than two allelic forms




**Criteria For Classification of Chromosomes as Homologous**

1. both members of a homologous pair are the same size with identical centromere locations
2. During meiosis, homologous pairs form pairs and synapses
3. Homologs contain the identical linear order of gene

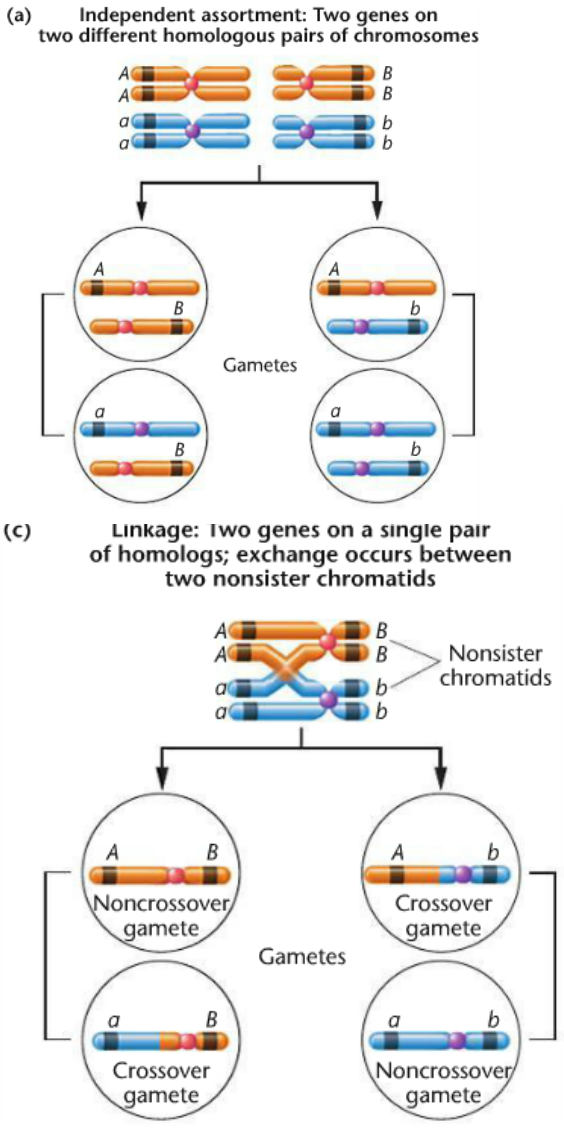
Overview:

- Chromosomes are visible in their distinctive shapes during mitosis and meiosis
- Both members of homologous pair are the same size and have the same centromere locations (except X and Y chromosomes
- Homologous chromosomes form pairs (synapse) during early meiosis
- Homologs contain identical linear order of gene loci

**3.6 – <u>Independent Assortment Leads to Extensive Genetic Variation:</u>**

- Genetic variation occurs because the homologous chromosomes of an individual are almost never genetically identical.
- Extensive genetic diversity is caused by the independent assortment of paternal and maternal chromosomes in the gametes.
- The number of possible gamete combinations can be calculated using the formula $2^n$, where n is the number of chromosomal pairs. Ex: in humans the possible gamete combinations = $2^{23}$ > 8 million.
- Since the gamete used from each parent can only be one of the 8 million, each offspring comes from the combination of one of the 8 million from each parent. Meaning that there is $(8x10^6)^2 = 64x10^{12}$ possibilities of what the offspring can turn out to be. That number is much larger than the world population. Thus, this is where our diverse genetic variation comes from.
- That is how independent assortment leads to a very large genetic variation in human beings.

**5.1 – <u>Genes Linked on the Same Chromosome Segregate Together:</u>**

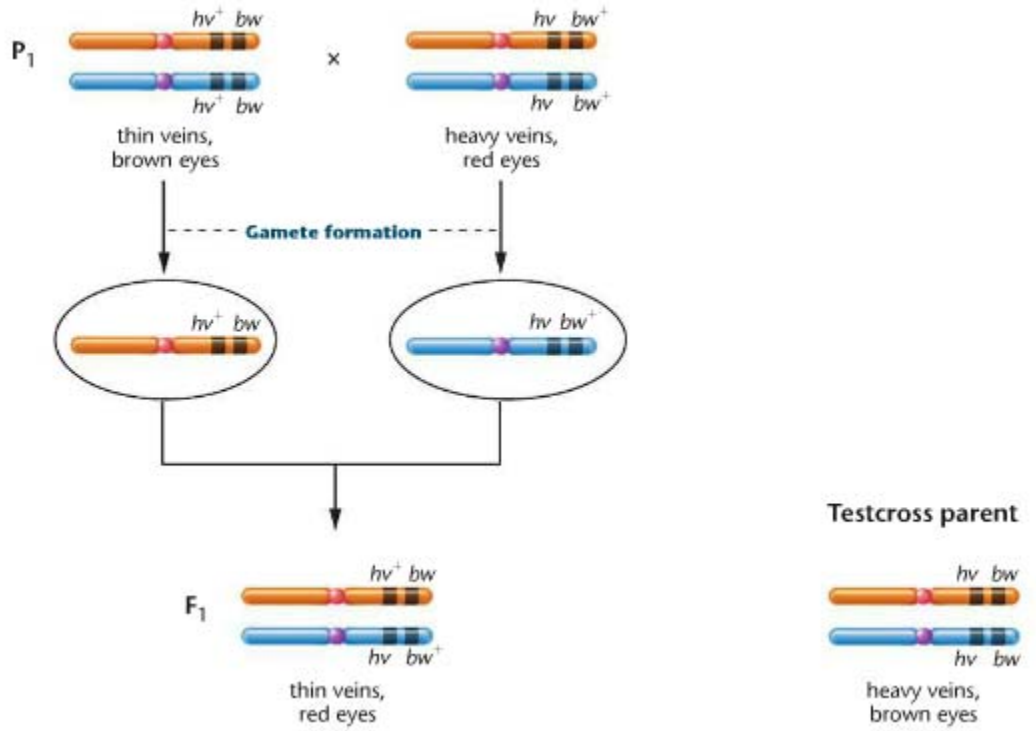


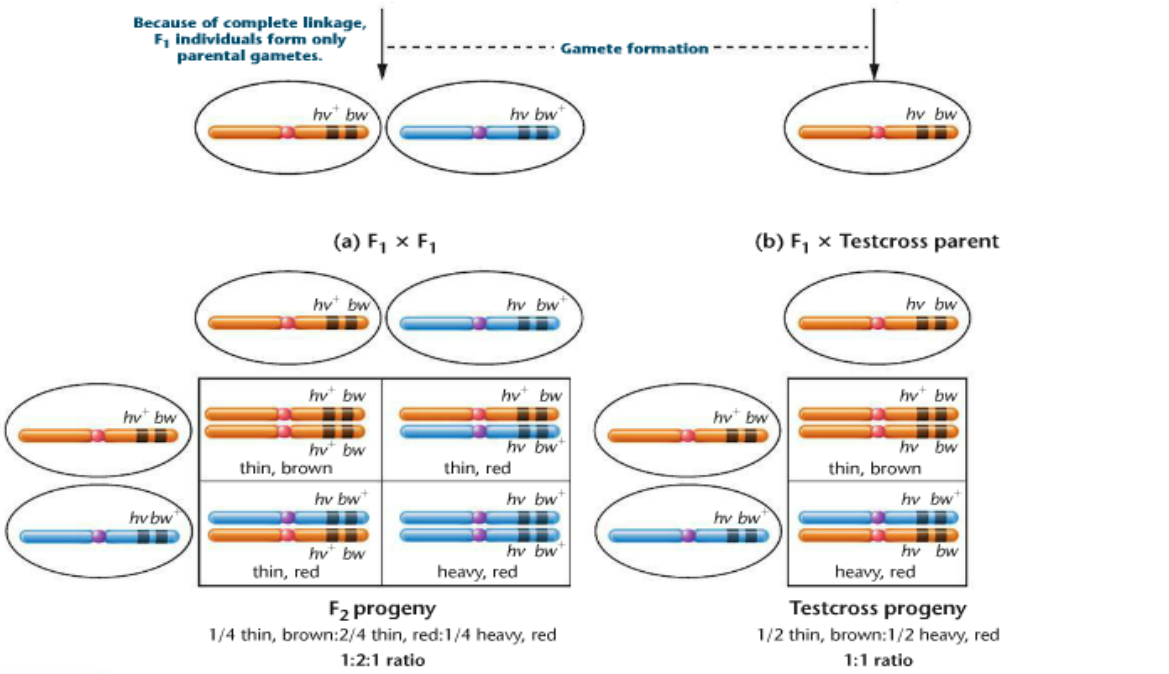

- Genetic studies revealed that certain genes segregate as if they were somehow joined or linked together if the genes are part of the same chromosome.
- Genes that are part of the same chromosome are said to be linked and to demonstrate **linkage** in genetic crosses. They are transmitted as a single unit.

- For Figure below: one band on one chromosome (a duplicated chromosome) equals one gene. Two bands equals two genes. Also, a homologous pair of chromosomes is designated by two different coloured chromosomes, but with the same letter, just one is capital and the other lowercase. Eg: AA and aa is a pair of homologous chromosomes.

- **(a) Independent assortment**
  - **We see independent assortment of 2 pairs of chromosomes, each containing one heterozygous gene pair. 4 distinct gametes are formed in equal proportions.**
- **(b) Linkage without crossing over**
  - **Only 2 genetically different kinds of gametes are formed if no crossing over occurs**
  - **Each gamete receives the allele of one homolog (this illustrates <u>complete linkage</u>, which produces only <u>parental</u> or <u>noncrossover gametes</u>)**
  - **The 2 parental gametes are formed in equal proportions**
- **(c) Linkage with crossing over**
  - **Involves 2 nonsister chromatids for crossing over, which forms 2 new allele combinations called <u>recombinant</u> or <u>crossover gametes</u>**
  - **As the distance between the two genes increases, the proportion of recombinant gametes increases and that of the parental gametes decreases.**
  - **Complete linkage or formation of many parental gametes occurs if genes are very close together that crossing over can't really be detected**
  - **when the loci of two linked genes are far apart, the number of recombinant gametes approaches, but does not exceed, 50 percent.**

**Figure 5-2:**

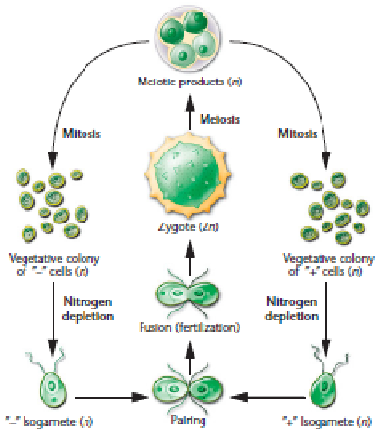





**Linkage Ratio:**

- A unique F2 generation ratio, called the linkage ratio, results from complete linkage of two genes.
- This theory is explained in figure 5-2. Where a fly of mutant brown eyes and wild thin veins is crossed with a fly of mutant heavy veins and wild red eyes.
- This can be represented by P1: where the ones above the line are located on the loci of one homolog, and the ones on the bottom are located on the homologous loci on the other homolog.
- Since these genes are located on an autosomal chromosome, the sex can be ignored.
- The F1 generation gets one chromosome from each parent giving us an F1 generation of phenotype thin, red but with heterozygous genes.
- Since this is a cross of complete linkage genes, the gametes of the F1 generation is all parental, and the F2 generation gives a ratio 1:2:1. Where ¼ of the F2 generation shows brown eyes thin veins, ½ shows wild phenotype, and ¼ shows red eyes and heavy veins. This is shown in 5-2 a).
- In 5-2 b) a test cross with the F1 flies was carried out to give us a 1:1 ratio of heavy, red and brown, thin. Had the genes been incompletely linked, we would have seen 4 phenotypes instead of two.
- Linkage groups can be identified by studying a large amount of mutant genes in any species. The number of linkage genes should correspond to the haploid number of chormosomes.
    - Linkage groups are genes located on the same chromosome with evidence and show evidence of being linked to one another.

**7.1 Life Cycles Depend on Sexual Differentiation**




- genes rather than chromosomes serve as underlying basis of sex determination
- Dimorphism: phenotypic differences between males and females
- Primary sexual differentiation: involves only gonads (where gametes are produced)
- Secondary sexual differentiation: involves overall appearance (reproductive and non reproductive organs)
- unisexual, dioecious, gonochoric : refer to an individual containing only male or only female reproductive organs.

bisexual/monoecious/hermaphroditic: containing both male and female reproductive organs (common in plants)

- Intersex: individuals of intermediate sexual condition, usually sterile

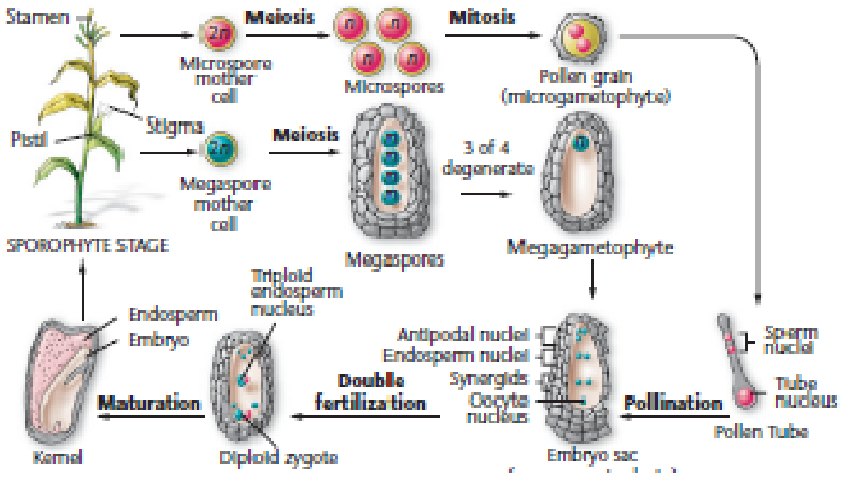
**Chlamydomonas**

- Green algae exhibiting infrequent periods of sexual reproduction
- Majority of life cycle is haploid reproducing in colonies by mitotic division
- In unfavorable conditions, two daughter cells may fuse creating a zygote
- Mating cells are isogametes and appear identical
- Only gametes from two different "mating types" can fuse zygotes ($mt^+$ with $mt^-$)
- Meiosis forms 4 zoospores/haploid cells (2 are mt+, 2 are mt-)
- Isogametes are morphologically similar, biochemically different
- This is a primitive mean of sex differentiation

**Zea mays (corn)**

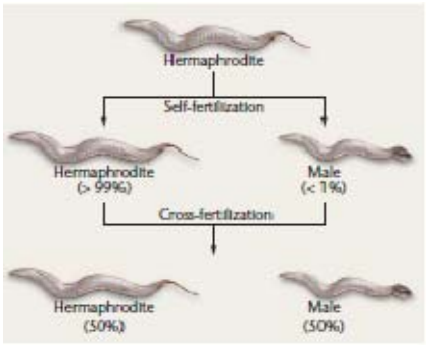
- Monoecious: Sporophyte phase and morphological structures predominate during the life cycle; both male and female structures present
- Stamens produce diploid microspores, which undergo meiosis and give 4 haploid microspores
- Haploid microspore becomes mature male micorgametophyte/pollen grain with two haploid sperm nuclei
- Megaspore (female diploid), produced in pistil of sporophyte and only 1 of 4 megaspores will survive after meiosis
- Three mitotic division in megaspore to give eight haploid nuclei in embryo sac
- Pollination: two sperm nuclei enter embryo sac→One sperm nucleus unites with haploid oocyte nucleus→Other sperm nucleus unites with two endosperm nuclei (double fertilization)
- Results in diploid zygote nucleus and triploid endosperm nucleus
- Kernel gives rise to new plant (sporophyte)→ Male and female gametes have same genetic constitution
- Mutant genes disrupt normal products and sex determination
- Mutant genes alter sexual development of florets

- Homozygous mutation " tassel seed" converts male tassels to female pistils
- Wild type induce cell death in pistils , making a normally monoeciuous plant to become female
- Recessive mutations change pistil making the plant only male




**Caenorhabditis elegans**

- a nematode worm , adults have 959 somatic cells which can be traced back to specific embryonic origins




- Two sexual phenotypes, males (only testes) and hermaphrodites ( two gonads produce both sperm and eggs→self fertilization)
- <1% are male , males mate with hermaphrodites producing half male and half hermaphrodites
- Lacks a Y chromosome, Hermaphrodites= XX, males= X
- Ratio of X chromosome # to autosome determine sex (Herm 2:2, Male (1:2)

**7.3 The Y Chromosome Determines Maleness in Humans**

**Klinefelter and Turner Syndromes**

Klinefelter syndrome (47, XXY)

- Tall, long arms and legs, large hands and feet
- Genitalia and internal ducts are male, testes are rudimentary and do not produce sperm
- Slight enlargement of breasts, round hips (slight feminine development)
- Less than average intelligence

Turner ( 45, X)

- Female external genitalia and internal ducts
- Rudimentary ovaries
- Short stature, cognitive impairment, underdeveloped breasts, skin folds on back of the neck, broad chest
- Mosaics – somatic cells display two different genetic cell lines (45, X)/(46, XX) or (45,X)/(46, XY)
- Most die in utero
- Nondisjuction causes both, this is the failure of the X chromosomes to segregate properly during meiosis
- Conclusion: since lack of Y results in female characteristics, then Y must make the male
- (45, Y) does not survive

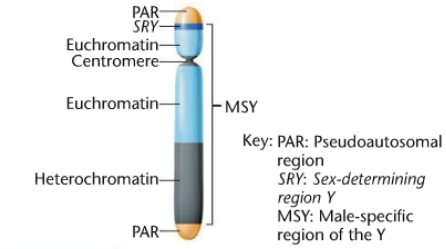


FIGURE 7–7 The regions of the human Y chromosome.

**47, XXX Syndrome**

- Female differentiation
- Often perfectly normal but, some have underdeveloped secondary sex characteristics, sterility, delay of language and motor skill development
- More X chromosomes disrupt the balance of normal development but do occur

**47, XYY Condition**

- Above average height males; generally normal
- Some below average intelligence

**Sexual Differentiation in Humans**

- By 5th week of gestation, hermaphroditic / sexually indifferent embryo
- 7 weeks, X or Y determines sex

**The Y Chromosome and Male Development**

- Y has far fewer genes than X (75 compared to 900+)
- On the end, they share a homologous region (pseudoautosomal regions- PARs) which can synapse and recombine during meiosis
- MSY is male-specific region of the Y, which does not synapse; has <u>some</u> homology with genes on X chromosomes
- MSY equal division between euchromatic and heterochromatic regions
- SRY is male sex-determining region Y (absent from X chromosome)
- SRY is activated in XY embryos (6-8 weeks), forming testes
- Rare cases where SRY can appear on XX, phenotypically showing as male or lack of SRY can appear in XY, phenotypically showing as female
- SRY activates TDF (testis-determining factor), which is a transcription factor
- Other autosomal genes are part of cascade of genetic expression started by SRY