Genetics and Molecular Biology

Topic 1: Bacterial Genetics

Bacteria

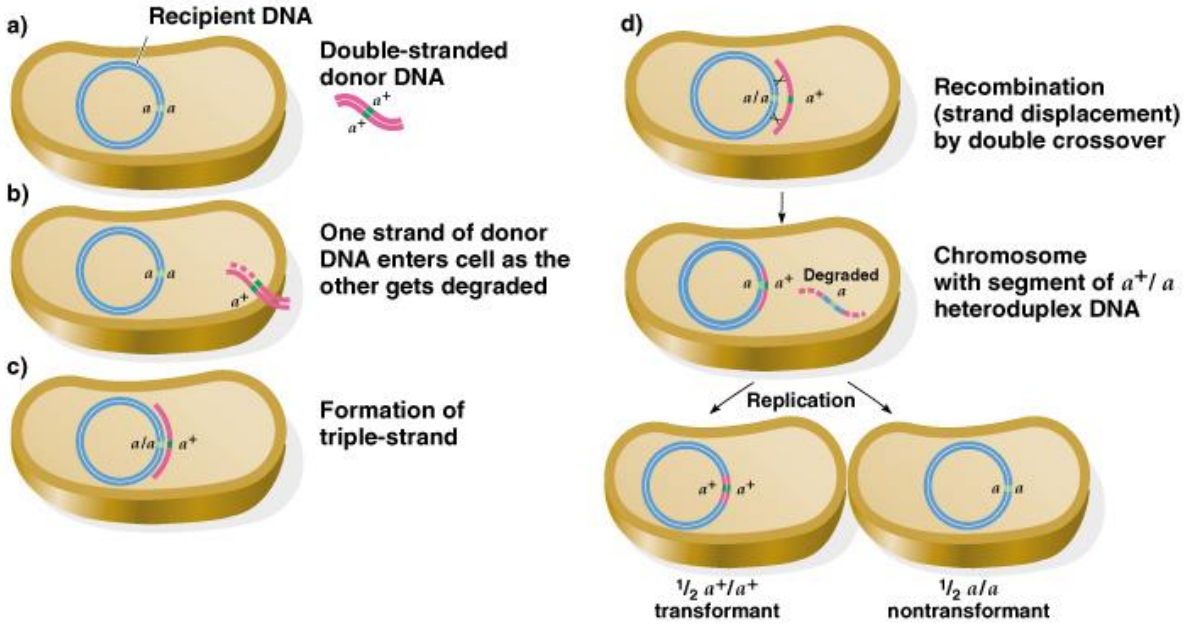
- Can be grown in liquid media or on agar plates
- **Prototrophic** bacteria: can grow on minimal media containing only carbon source, water, and inorganic salts
- **Auxotrophic** bacteria: mutants which need media supplements, eg adenine, threonine, biotin
- Antibiotic resistant mutants: able to grow in the presence of antibiotics like streptomycin, tetracycline, ampicillin
- Carbon source mutants: cannot utilise a particular carbon source, e.g. Lac^- cannot utilise lactose
- Nutritional mutants: unable to synthesise essential nutrient such as nucleic or amino acids

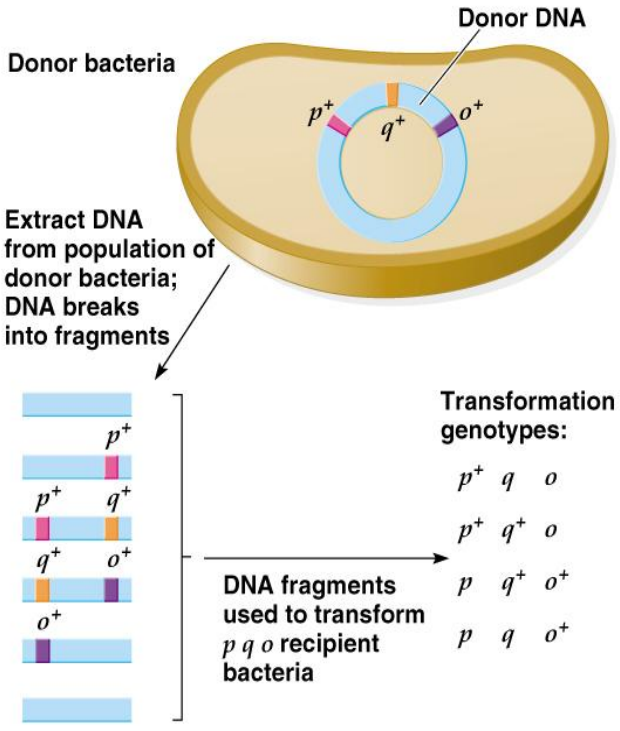
Conventions in Bacterial Genetics

- Phenotypes: written with 3 letters, the first one capitalised, followed by a + or -, or superscript ^{s} or ^{r} (for susceptible and resistant)
- Genotypes: writtein in lowercase with italicised letters
- Example: a leucine auxotroph is Leu^- and results from a *leu*$^-$ mutation

Transformation

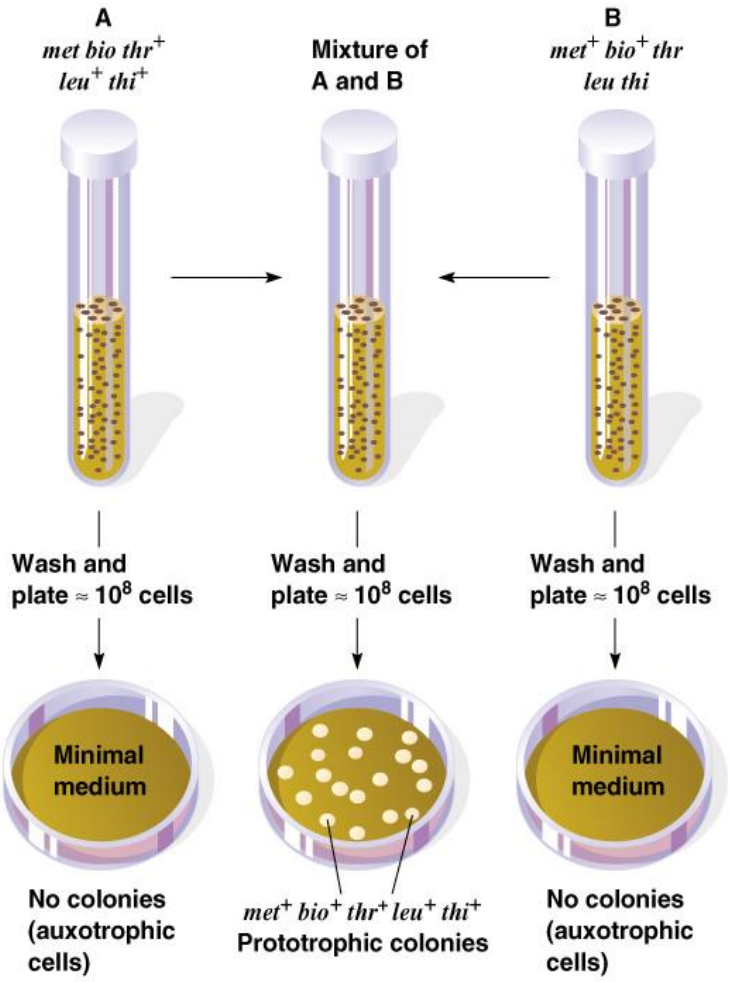
- **Transformation** is the genetic alteration of a cell resulting from the direct uptake and incorporation of exogenous genetic material (exogenous DNA) from its surroundings and taken up through the cell membrane(s)
- Recipients acquire DNA from the medium, not needing direct contact
- Transformation was first demonstrated in 1928 by British bacteriologist Frederick Griffith. Griffith discovered that a strain of *Streptococcus pneumoniae* could be made virulent after being exposed to heat-killed virulent strains
- **Cotransformation** is the simultaneous transformation of two or more genes. Only genes in the same chromosomal vicinity can be transformed; the closer together the genes lie, the more frequently they will be cotransformed
- If markers are very far apart get frequency equivalent to $1/10^6$, while if so near as to co-transform get frequency nearer to $1/10^3$
- Cotransformation frequencies can also be used to determine gene order



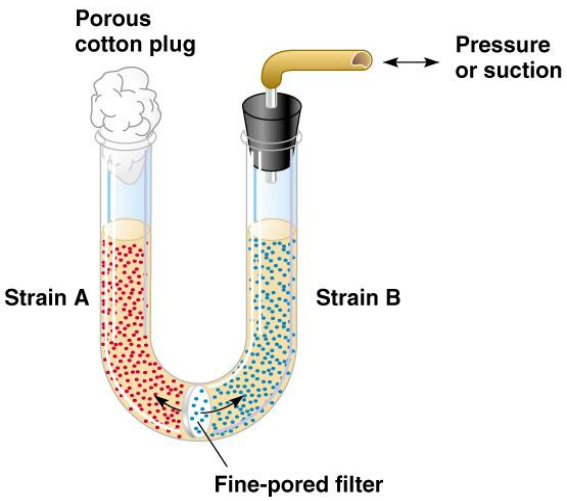


Discovery of Conjugation

- In a milestone experiment in 1946, **Joshua Lederberg and Edward Tatum** discovered bacterial conjugation
- They had 2 strains of E.coli with different nutritional requirements
- Strain A, met^- bio^- thr^+ leu^+ thi^+
- Strain B, met^+ bio^+ thr^- leu^- thi^-
- They plated only strain A, only strain B, and a mix of both strains on separate plates
- The mixed stains A&B were the only one to show colonies, so they concluded that there had been some form of genetic recombination between the strains



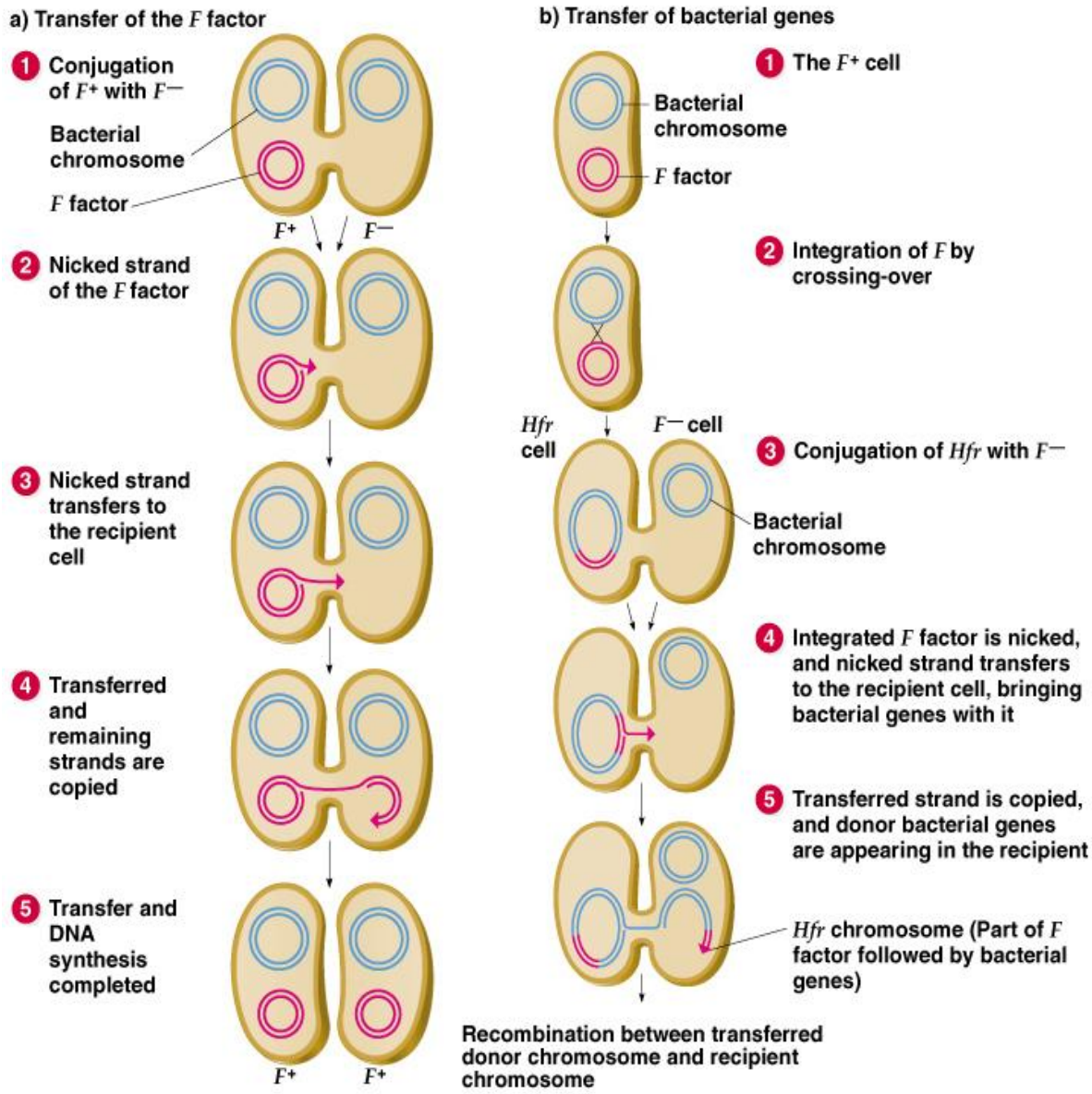
- In a separate experiment, Bernard Davis used a T-tube to show that physical contact was necessary for conjugation to occur



Conjugation

- In 1953 **William Hayes** showed that DNA transfer was uni-directional, one way only
- He proposed that the transfer of genetic material was mediated by something called the sex factor F. Cells carrying F are F^+ , those without are F^-

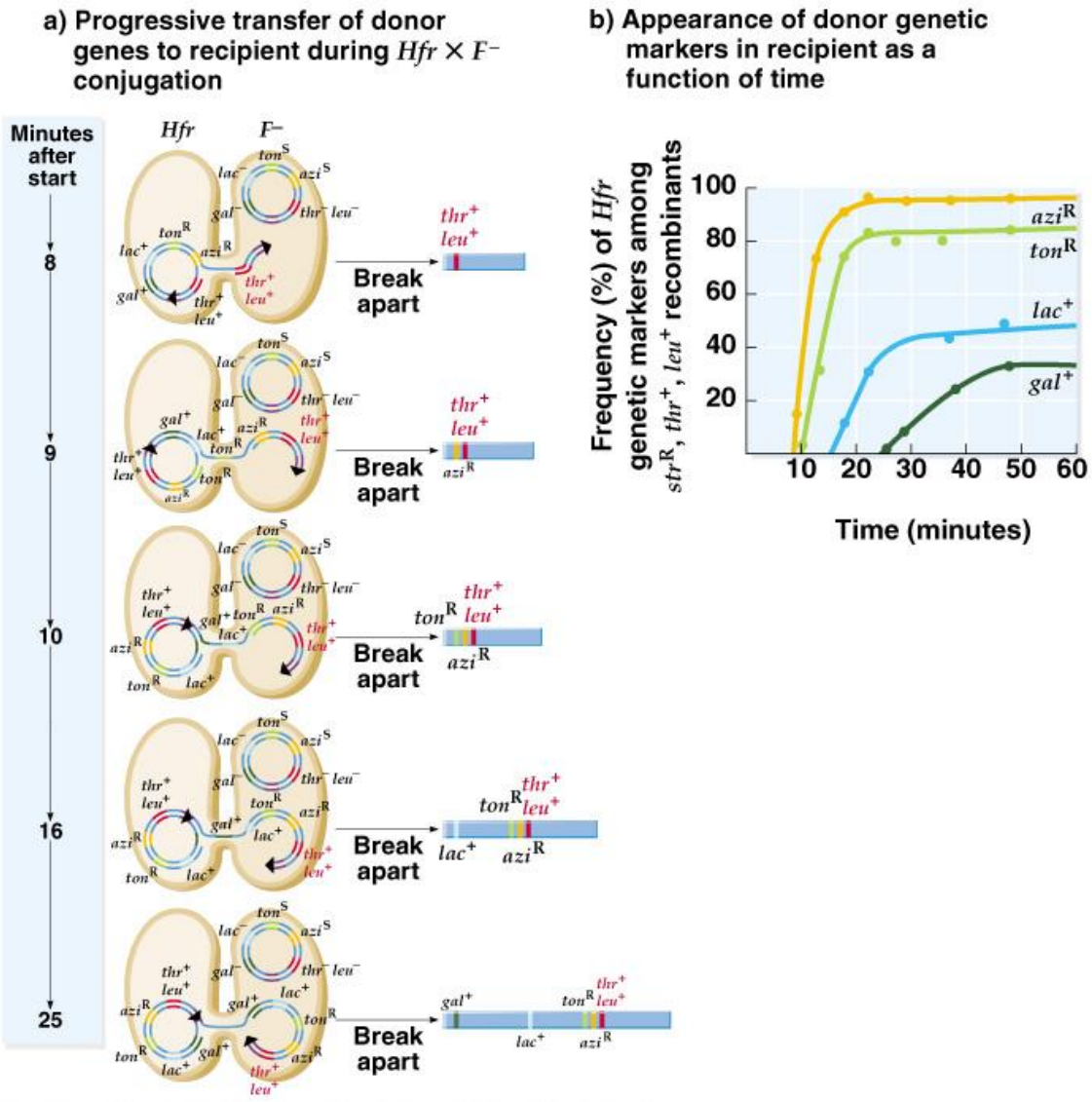
- F is a small plasmid containing about 100 genes coding for the ability to produce pili (pilus) needed for conjugation

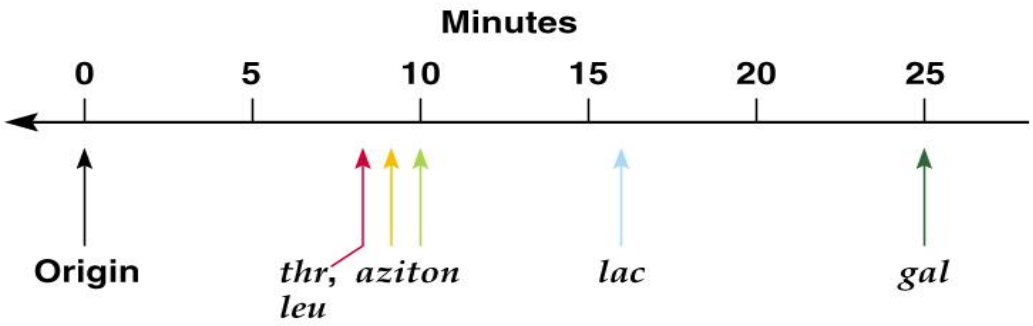


Interrupted-mating Experiments

- Occasionally F leaves the cytoplasm and incorporates into the host chromosome
- When it transfers it occasionally transfers chromosomal markers, while allow partial or complete genome mapping
- These cells are called Hfr, or high-frequency recombination cells (also called an Hfr strain)
- Unlike a normal F^+ cell, Hfr strains will, upon conjugation with a F^- cell, attempt to transfer their entire DNA through the mating bridge
- A structure as fragile as a mating bridge will, however, likely break, and so the transfer is rarely complete. Thus, the F^- cell uses only part of the genomic DNA of the Hfr cell for recombination
- The amount of chromosomal DNA that is transferred depends on how long the two conjugating bacteria remain in contact. In common laboratory strains of *E. coli* the transfer of the entire bacterial chromosome takes about 100 minutes

- The genome can thus be mapped, with the distance between neighbouring genes measured in the number of minutes between when they are transferred to the recipient





a) Orders of gene transfer

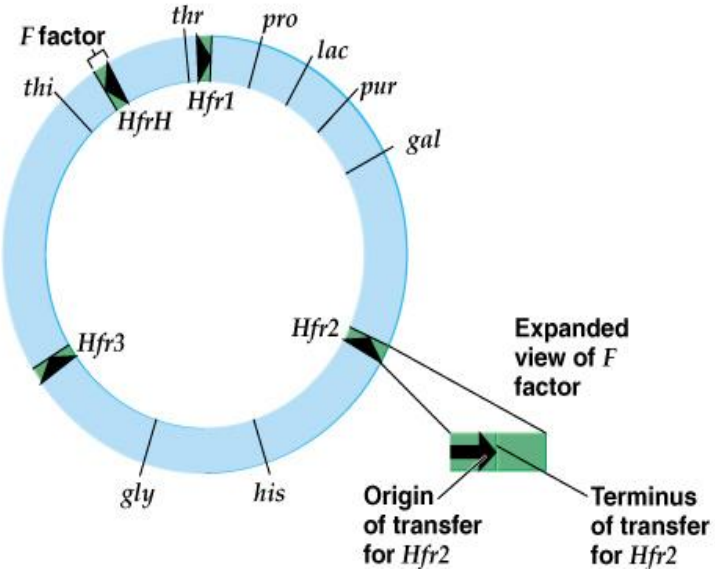
Hfr **strains:**

H origin—thr–pro–lac–pur–gal
1 origin—thr–thi–gly–his
2 origin—his–gly–thi–thr–pro–lac
3 origin—gly–his–gal–pur–lac–pro

b) Alignment of gene transfer for the *Hfr* strains

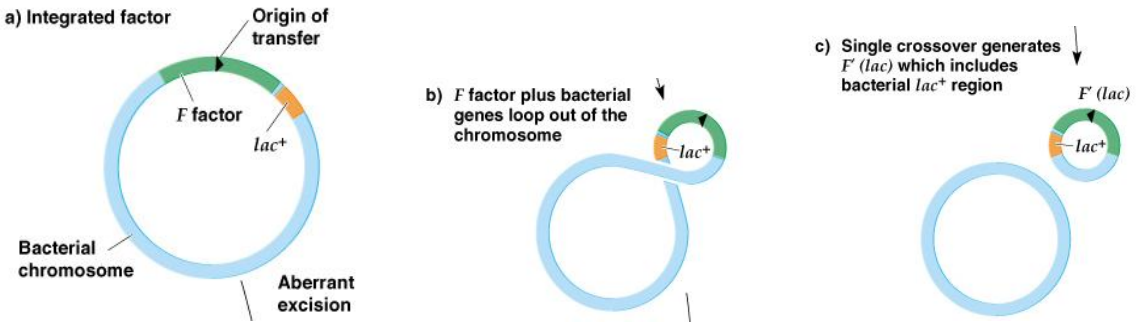
H thr–pro–lac–pur–gal
1 his–gly–thi–thr
2 his–gly–thi–thr–pro–lac
3 pro–lac–pur–gal–his–gly

c) Circular *E. coli* chromosome map derived from *Hfr* gene transfer data



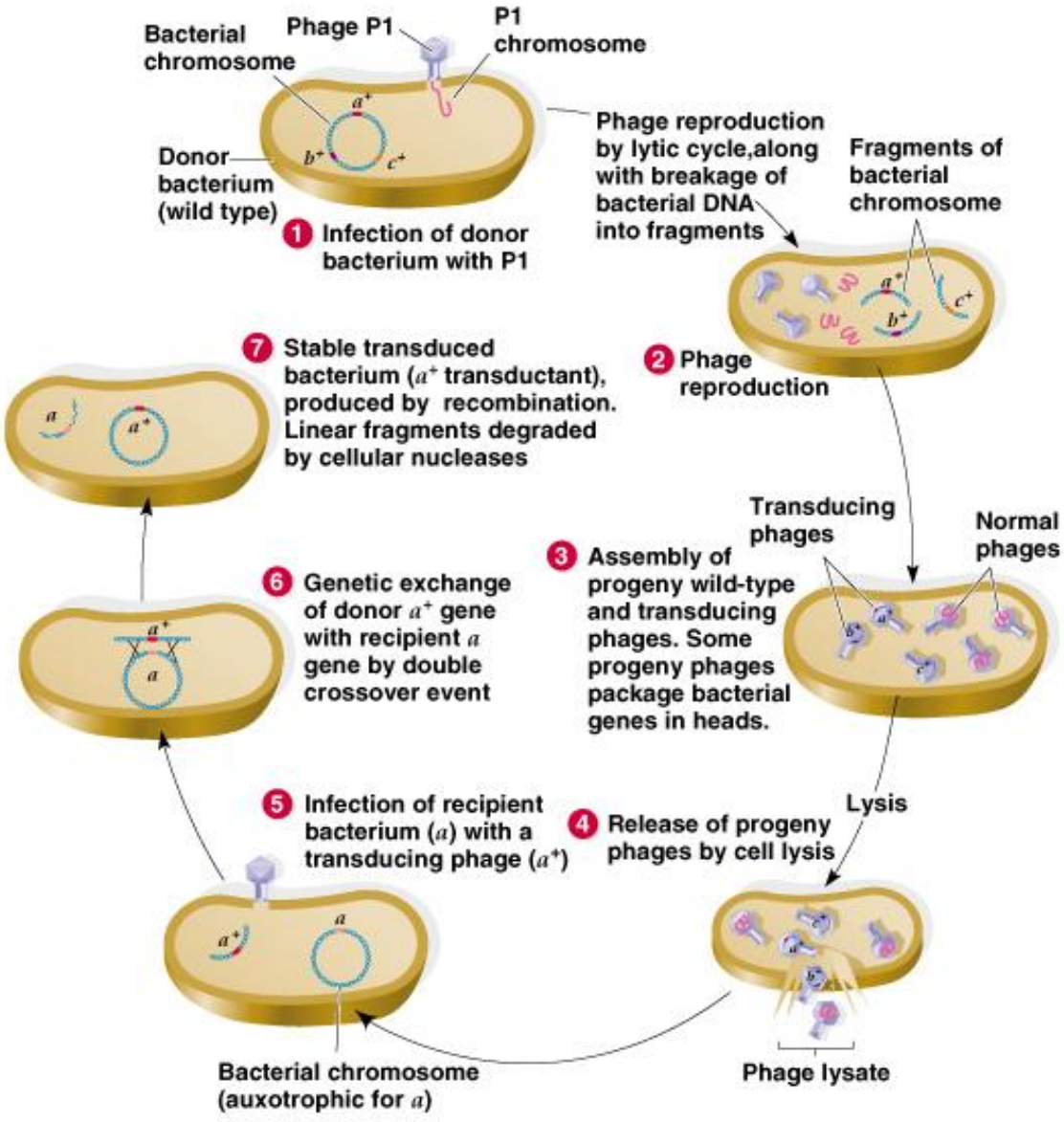
The F' Factor

- Occasionally F is excised from Hfr by genetic exchange
- This excision is not always precise, and can create a plasmid with part of the chromosome contained in it
- This can lead to a recipient which is partial diploid or **merodiploid**
- **F' (F-prime) bacteria** are formed by incorrect excision from the chromosome, resulting in F plasmid carrying bacterial sequences that are next to where the F episome has been inserted



Transduction

- Transduction is the process by which DNA is transferred from one bacteria to another by a virus
- Transduction does not require physical contact between the cell donating the DNA and the cell receiving the DNA (which occurs in conjugation)
- Bacterial DNA is transferred from one bacteria to another by a phage particle
- Generalised transducing phage: process by which any bacterial gene may be transferred to another bacterium via a bacteriophage, and typically carries only bacterial DNA and no viral DNA. In essence, this is the packaging of bacterial DNA into a viral envelope
- Specialised transducing phage: process by which a *restricted* set of bacterial genes is transferred to another bacterium. The genes that get transferred (donor genes) depend on where the phage genome is located on the chromosome. Specialized transduction occurs when the prophage excises imprecisely from the chromosome so that bacterial genes lying adjacent to the prophage are included in the excised DNA



Cotransduction Frequency

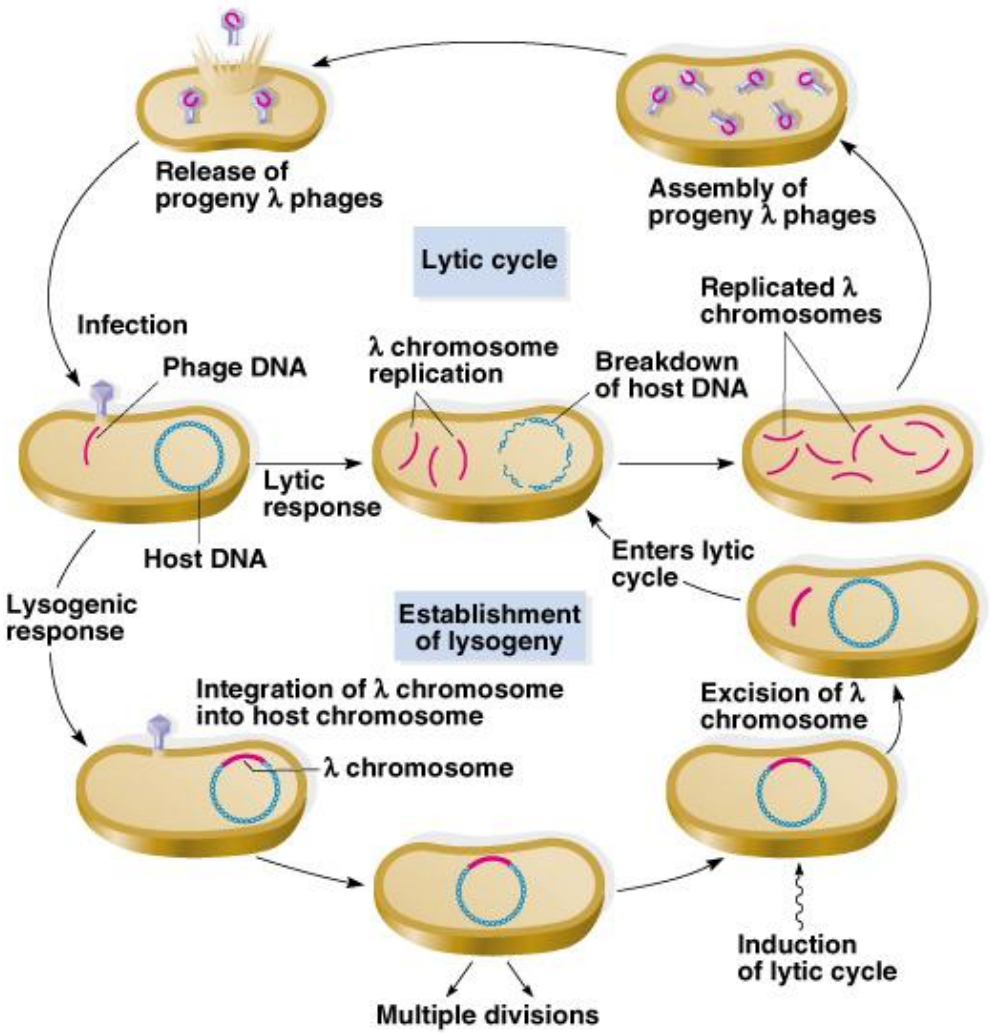
- Cotransduction occurs when two genes are transduced together
- Frequency depends on how close the markers are in the donor genome, and is calculated as the number of recipient cells which display both traits, divided by the total number of recipient cells displaying one of the traits
- Frequency of cotransduction falls off very quickly as distance between markers increases

Lytic Life Cycle

- Viruses of the lytic cycle are called virulent viruses. The lytic cycle is a six-stage cycle
- In the first stage, called "penetration," the virus injects its own nucleic acids into a host cell. Then the viral acids form a circle in the center of the cell
- The cell then mistakenly copies the viral acids instead of its own nucleic acids. Then the viral DNA organize themselves as viruses inside the cell
- When the number of viruses inside becomes too much for the cell to hold, the membrane splits and the viruses are free to infect other cells

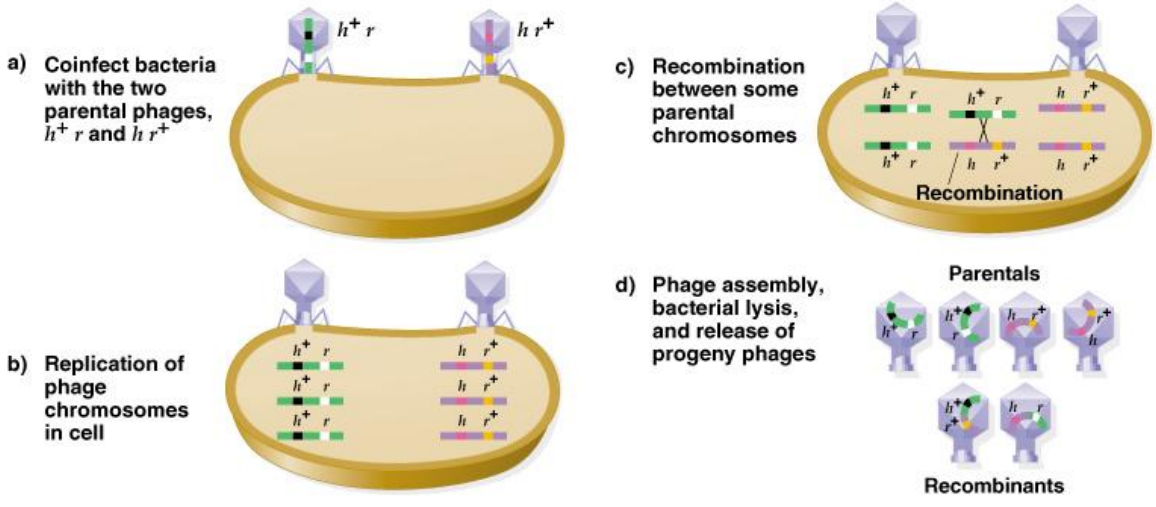
Lysogenic Life Cycle

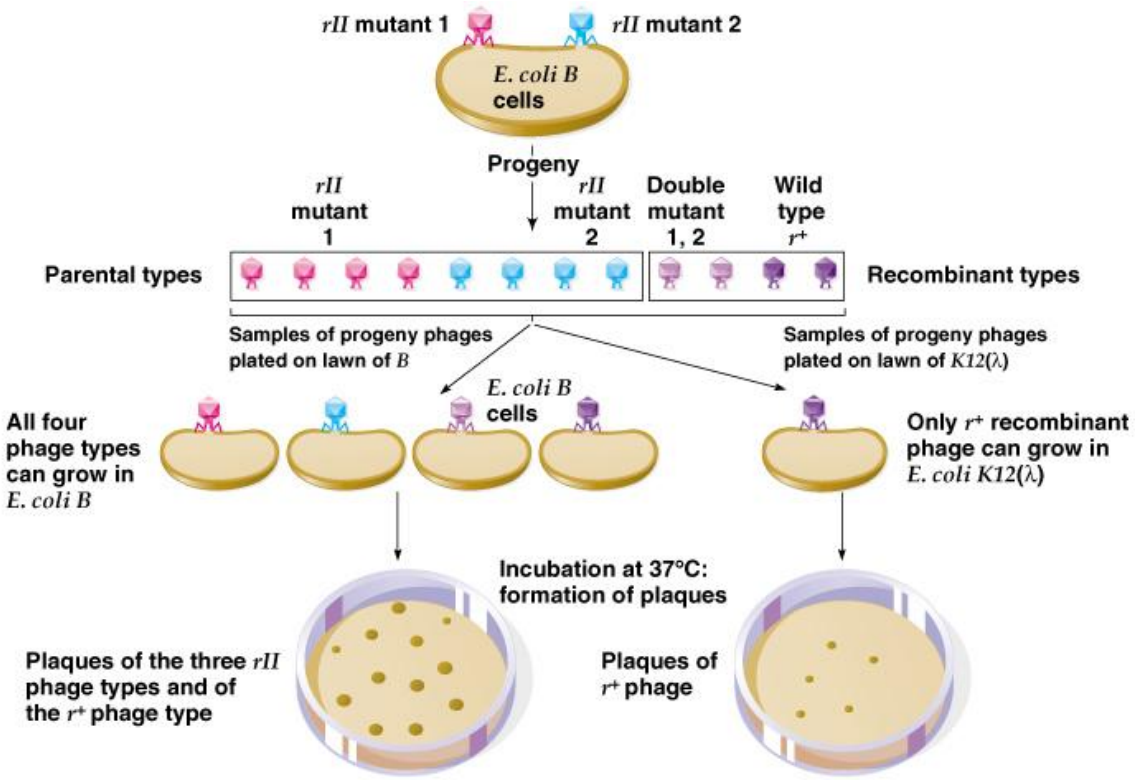
- **Lysogeny**, or the **lysogenic cycle**, is one of two methods of viral reproduction (the lytic cycle is the other)
- Phages that replicate only via the lytic cycle are known as virulent phages while phages that replicate using both lytic and lysogenic cycles are known as temperate phages
- In the lysogenic cycle, the phage DNA first integrates into the bacterial chromosome to produce the prophage
- When the bacterium reproduces, the prophage is also copied and is present in each of the daughter cells
- The daughter cells can continue to replicate with the prophage present or the prophage can exit the bacterial chromosome to initiate the lytic cycle
- A **prophage** is a bacteriophage (often shortened to "phage") genome inserted and integrated into the circular bacterial DNA or existing as an extrachromosomal plasmid
- The host is termed a lysogen when a prophage is present



Mapping Bacteriophage Genes

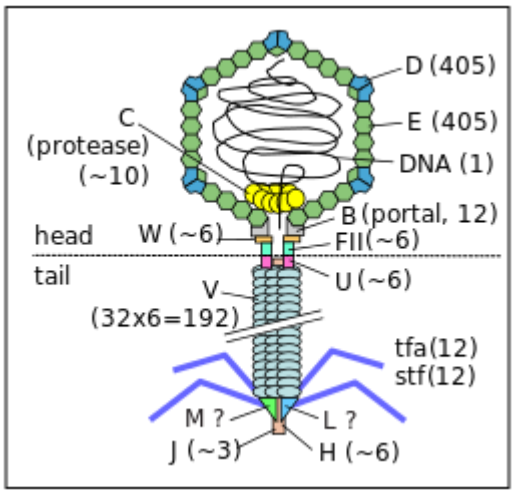
- The distance between bacteriophage genes can be determined by the recombination frequency of those two genes when they coinfect a bacteria

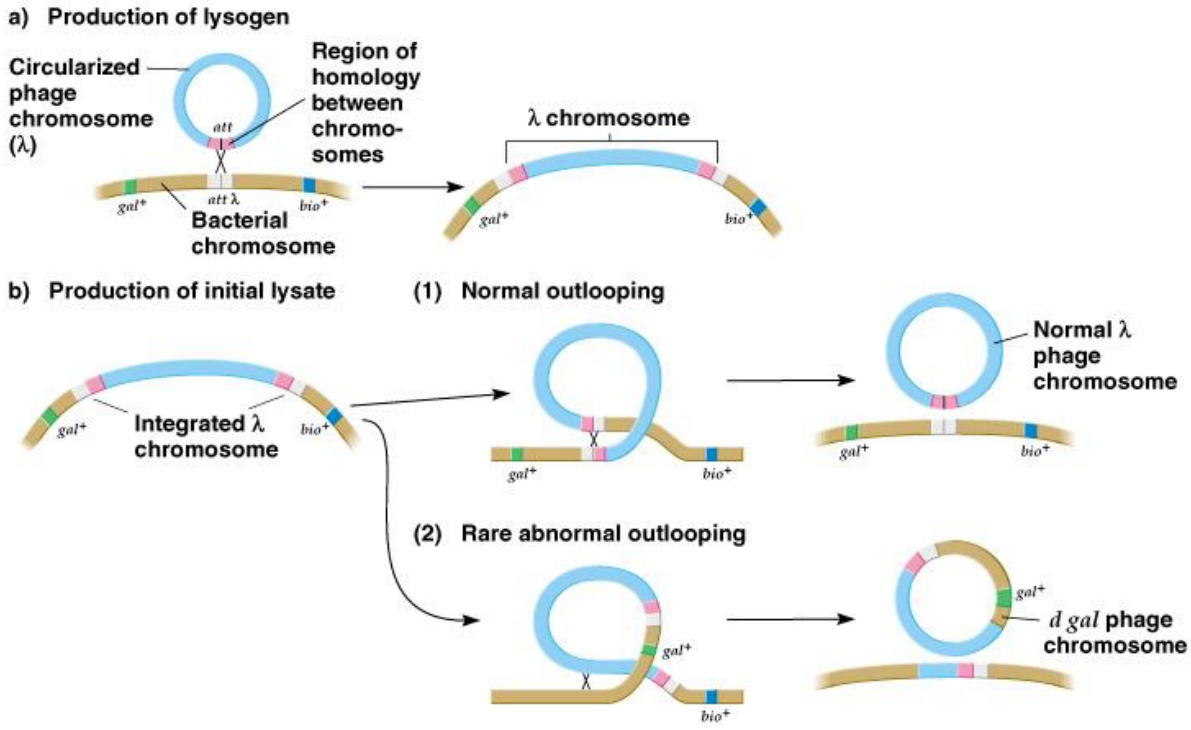




Lambda Phage

- **Enterobacteria phage λ** (**lambda phage**) is a bacterial virus, or bacteriophage, that infects the bacterial species *Escherichia coli*
- This virus has a temperate lifecycle that allows it to either reside within the genome of its host through lysogeny or enter into a lytic phase
- The phage particle consists of a head (also known as a capsid), a tail, and tail fibers
- The genome contains 48,490 base pairs of double-stranded, linear DNA, with 12-base single-strand segments at both 5' ends. These two single-stranded segments are the "sticky ends" of what is called the *cos* site. The *cos* site circularizes the DNA in the host cytoplasm. In its circular form, the phage genome, therefore, is 48,502 base pairs in length
- Prophage induction: Induces lytic phase, initiated by damage to bacterial DNA
- Lambda phage has also been of major importance in the study of specialized transduction
- Usually the prophage is extracted 'clean' out of the host genome, however in 1 cell per 10^6 or 10^7 cells excision is imprecise
- This form of specialized transduction can only transduce certain bacterial genes close to the site of insertion of the phage DNA (which is fixed); in the case of lamda, the bio or gal genes can be transduced in this way



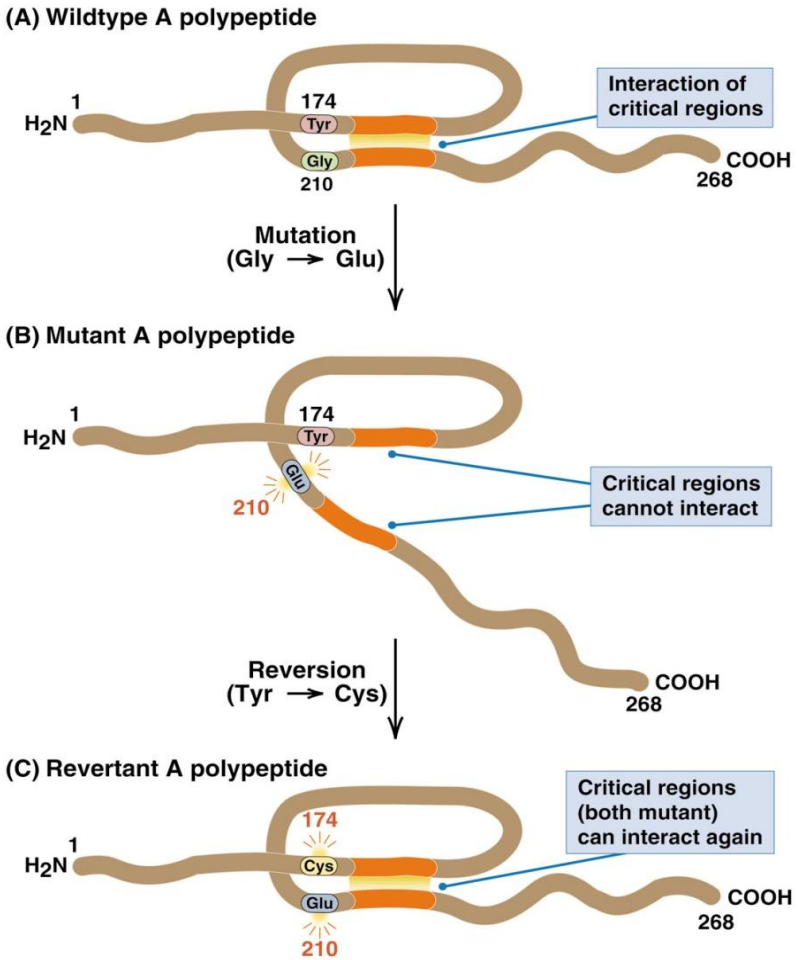


Topic 2: Mutations and DNA Repair

Types of Mutation by Function

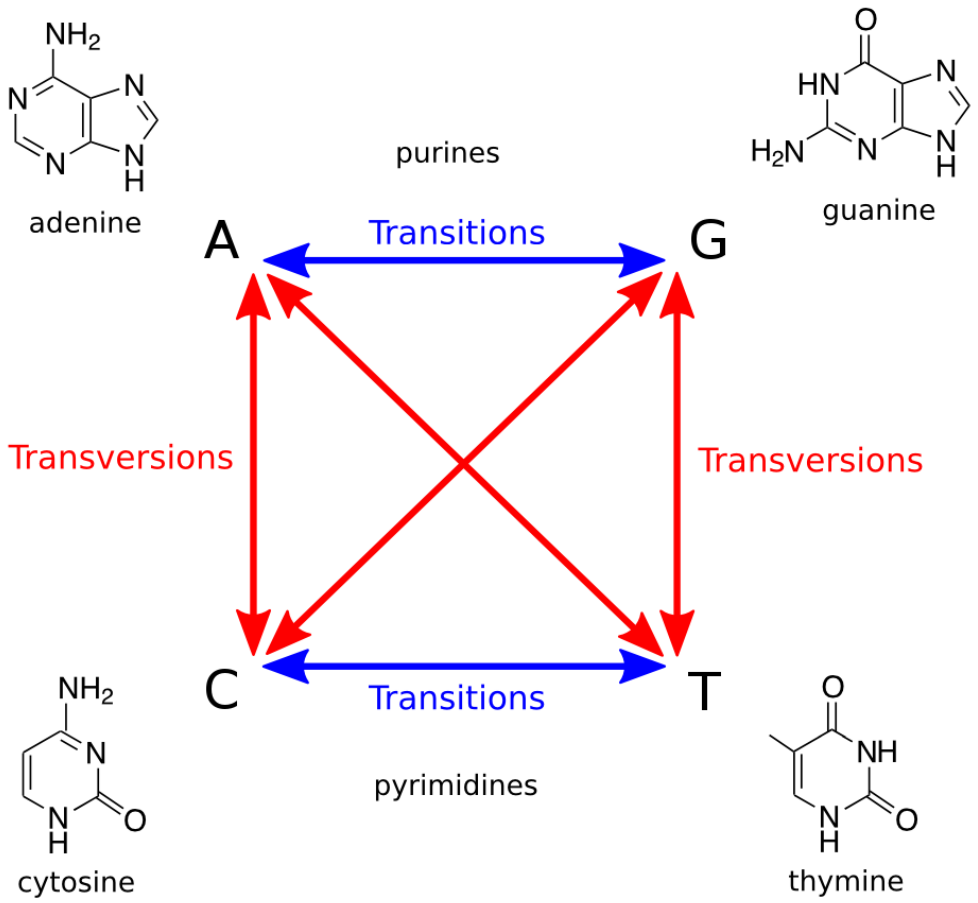
- **Amorphic** describes a mutation that causes complete loss of gene function, Also called **null** or knockout mutants. Usually recessive - heterozygotes are often wild type
- **Hypomorphic** describes a mutation that causes a partial loss of gene function. lower protein production than normal levels or amino acid replacement which impairs protein function
- **Hypermorphic** mutation causes an increase in normal gene function. Hypermorphic alleles are gain of function alleles. A hypermorph can result from an increase in gene dose (a gene duplication), from increased mRNA or protein expression, or constitutive protein activityA
- **Back mutation** or **reversion** is a point mutation that restores the original sequence and hence the original phenotype. Can be true (same amino acid), or partial (different)

- **Gain-of-function mutations** change the gene product such that it gains a new and abnormal function. These mutations usually have dominant phenotypes a **second** mutation at another site which compensates for the effect of the first mutation; these are more common
- **Ectopic expression** is the expression of a gene in an abnormal place in an organism. This can be caused by a disease, or it can be artificially produced as a way to help determine what the function of that gene is.
- **Antimorph**s are dominant mutations that act in opposition to normal gene activity. Antimorphs are also called **dominant negative** mutations. An antimorphic mutation might affect the function of a protein that acts as a dimer so that a dimer consisting of one normal and one mutated protein is no longer functional (e.g. sickle cell anemia)
- **Forward mutation**: a mutation which changes a wild type allele into a new allele
- **Mutator gene**: a gene in which a mutation results in a much higher-than-normal mutation frequency for all other genes
- **Suppressor mutation:** a second mutation that alleviates or reverts the phenotypic effects of an already existing mutation, but at a different site to the original (hence it is a partial reversion)
- **Suppressor gene:** genes which, when mutated, can cause partial reversion of some mutation in another gene. Often suppressor genes encode for tRNAs, changing the way the genetic code is read, and thereby offsetting the impact of (say) a missense or nonsense mutation
- **SOS response**: a global response to DNA damage in which the cell cycle is arrested and DNA repair and mutagenesis are induced. Repressed by **LexA gene/enzyme**



Types of Mutation by Structural Effect

- **Point mutations**, often caused by chemicals or malfunction of DNA replication, exchange a single nucleotide for another
- **Transition:** a purine replaces a purine or pyrimidine replaces a pyrimidine (4 possibilities)
- **Transversion:** purine replaced with pyrimidine or pyrimidine replaced with purine (8 possibilities)
- **Silent mutations**: new triplet encodes the same amino acid due to the degeneracy of the DNA code e.g. AGG → CGG, both encode Arg
- **Missense mutations**: different amino acid encoded by the new sequence changes have little or no effect - neutral mutation. Not all missense mutations lead to appreciable protein changes. An amino acid may be replaced by an amino acid of very similar chemical properties, in which case, the protein may still function normally; this is termed a neutral, "quiet", "silent" or conservative mutation
- **Neutral mutations**: new triplet encodes a different amino acid new amino acid similar or functionally equivalent**,** eg Lys → Arg
- **Nonsense mutation** is a point mutation in a sequence of DNA that results in a premature stop codon, or a *nonsense codon* in the transcribed mRNA, and in a truncated, incomplete, and usually nonfunctional protein product
- **Depurination:** The removal, by hydrolysis, of a purine base from DNA
- **Depyramidation:** A process in which the chemical bond linking a pyrimidine (cytosine or thymine) to a deoxyribose sugar breaks spontaneously

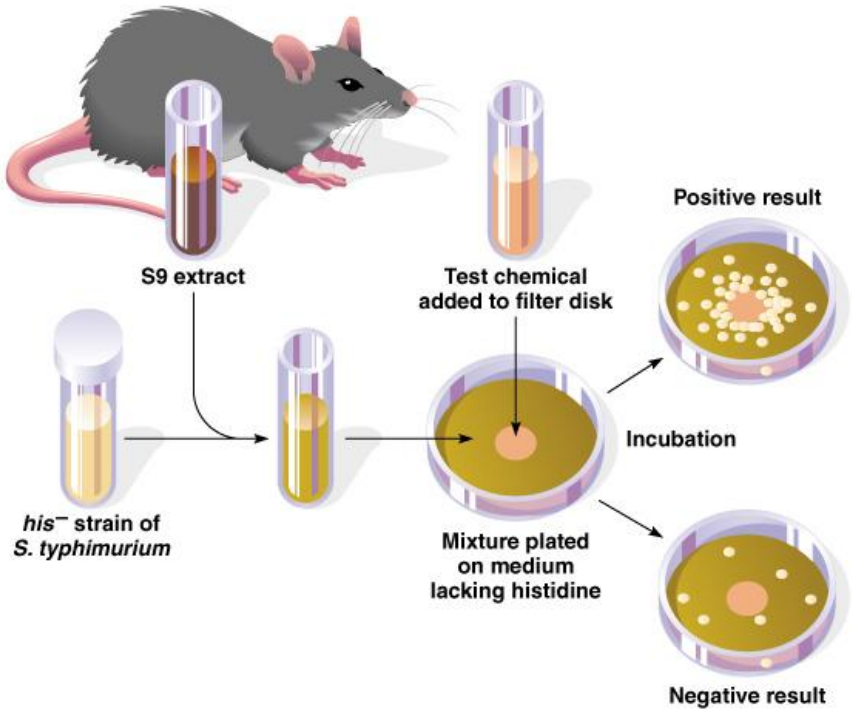


Examples

- Sickle cell anemia: missense transversion mutation of AT->TA in beta-globin chain of hemoglobin. Mutation causes defective haemoglobin with long, needle-shaped red blood cells
- Bacterial blight of rice: resistance to bacterial blight is conferred by *xa5* gene, a missense mutation from GTC (Val) to GAC (Glu)

Ames test

- Simple, inexpensive test for **screening mutagens**
- A positive test indicates that the chemical is mutagenic and therefore may act as a carcinogen, because cancer is often linked to mutation
- The test serves as a quick and convenient assay to estimate the carcinogenic potential of a compound because standard carcinogen assays on mice and rats are time-consuming and expensive. However, false-positives and false-negatives are known



Mutation rates

- Mutation rate: The number of mutations occurring in some unit of time
- Mutation frequency: the number of occurances of a particular kind of mutation, expressed as the proportion of cells or individuals in apopulation, such as the number of mutations per 100,000 organisms or per 1 million gametes
- Vary between organisms: Drosophila yellow body 1 in 10^{-4}/gamete/generation, *E. coli* strep resistance 10^{-9} cell/generation
- Mutation rates vary **between genes** within an individual organism, e.g. *E. coli* genes can vary in mutation rate between 10^{-5} cell/generation - 10^{-9}cell per generation

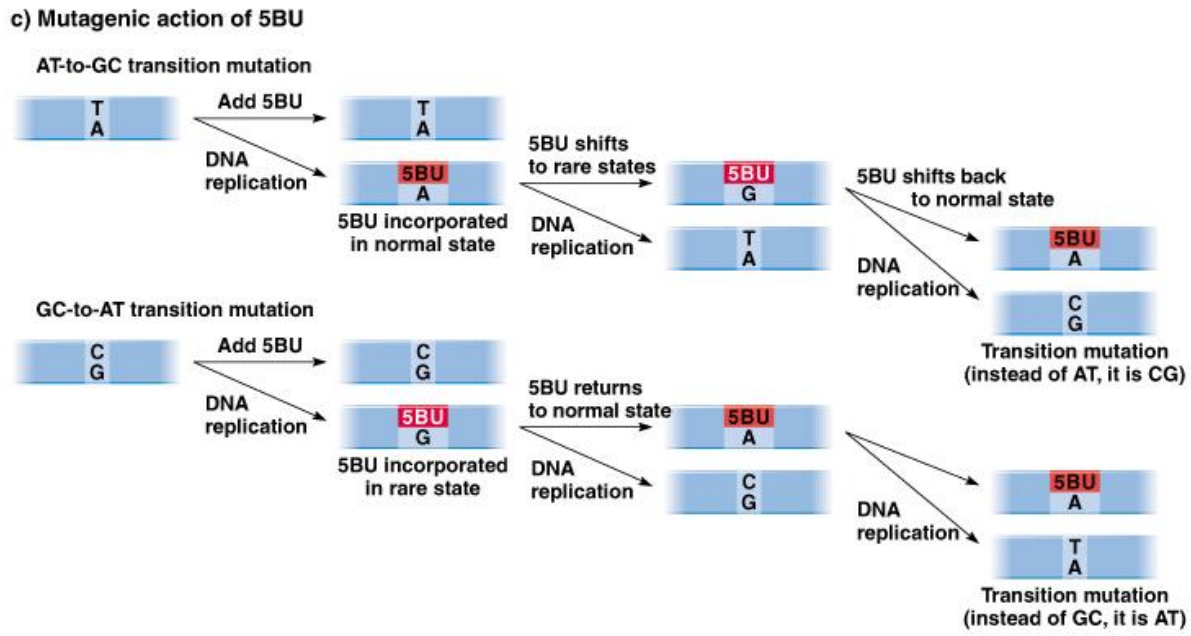
Causes of Mutation

- Tautomerism: A base is changed by the repositioning of a hydrogen atom, altering the hydrogen bonding pattern of that base, resulting in incorrect base pairing during replication.

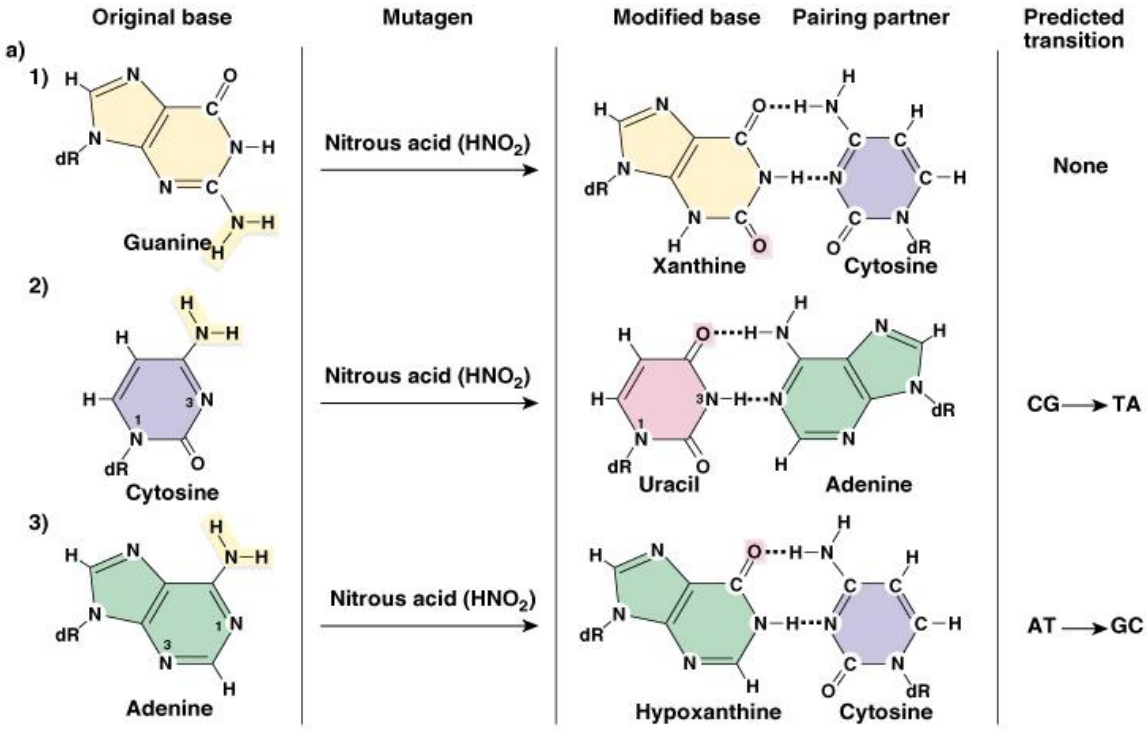
- Depurination: Loss of a purine base (A or G) to form an apurinic site (AP site).
- Deamination: Hydrolysis changes a normal base to an atypical base containing a keto group in place of the original amine group. Examples include C → U and A → HX (hypoxanthine), which can be corrected by DNA repair mechanisms; and 5MeC (5-methylcytosine) → T, which is less likely to be detected as a mutation because thymine is a normal DNA base.
- Slipped strand mispairing: Denaturation of the new strand from the template during replication, followed by renaturation in a different spot ("slipping"). This can lead to insertions or deletions.

Mutagens

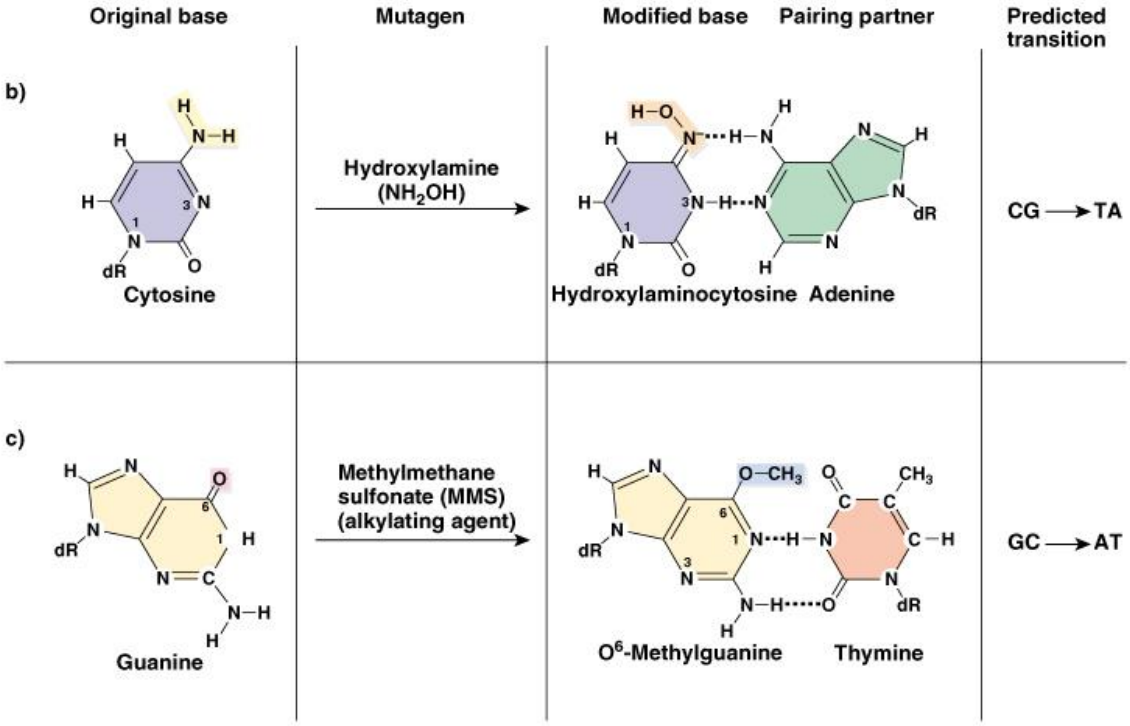
- A **mutagen** is a physical or chemical agent that changes the genetic material of an organism and thus increases the frequency of mutations above the natural background level
- **Base analogues**: Chemically similar to bases incorporate during replication. One of the more common base analogs is 5-bromouracil (5BU), which is an analog of thymine. When a nucleotide containing **5-bromouracil** is incorporated into the DNA, it is most likely to pair with adenine; however, it can spontaneously shift into another isomer which pairs with a different nucleobase, guanine. If this happens during DNA replication, a guanine will be inserted as the opposite base analog, and in the next DNA replication, that guanine will pair with a cytosine. This results in a change in one base pair of DNA, a transition mutation



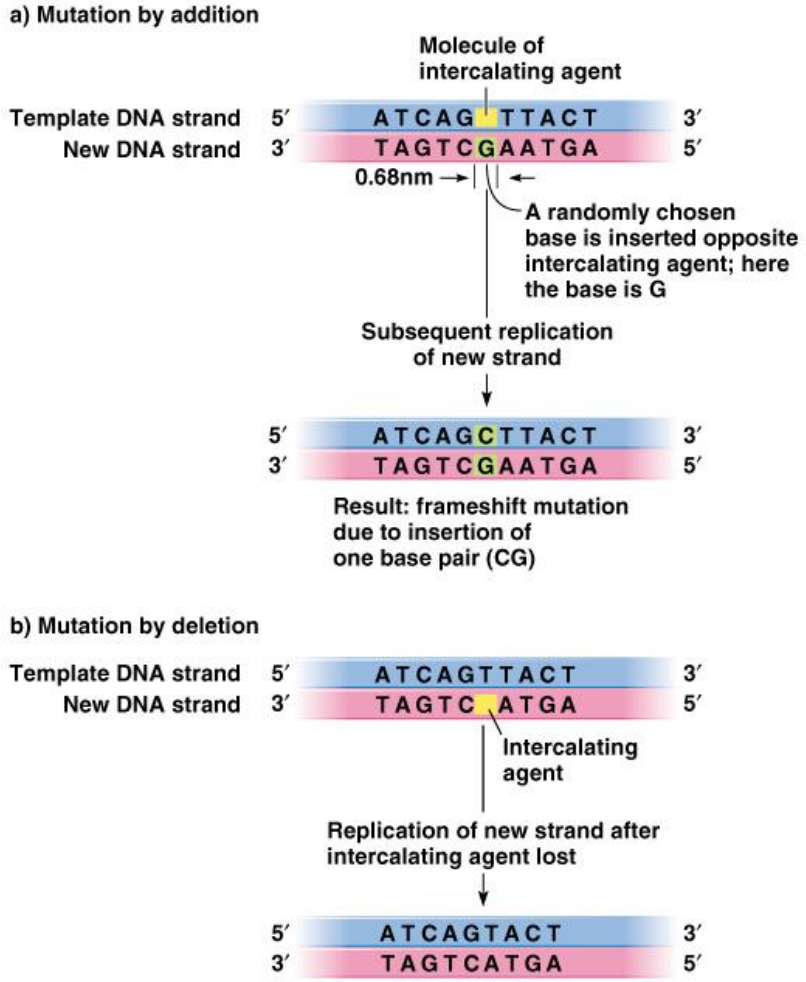
- **Nitrous acid**: a deaminating agents that changes adenine to hypoxanthine, which pairs with cytosine leading to A-T → G-C transitions, and also changes cytosine to uracil, which pairs with adenine, leading to G-C → A-T transitions



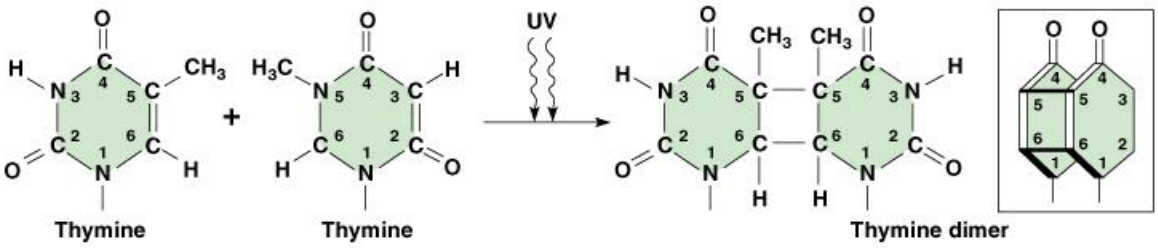
- **Alkylating agents**: such as ethylnitrosourea and EMS, which transfer methyl or ethyl groups to bases or the backbone phosphate groups, yielding A-T → G-C and G-C → A-T transitions



- **Intercalating agents**: such as ethidium bromide, **acridine orange** and proflavine, molecules that may insert between bases in DNA, causing frameshift mutation during replication



- **UV light**: creates pyrimidine dimers, particularly thymidine, distorting the helix

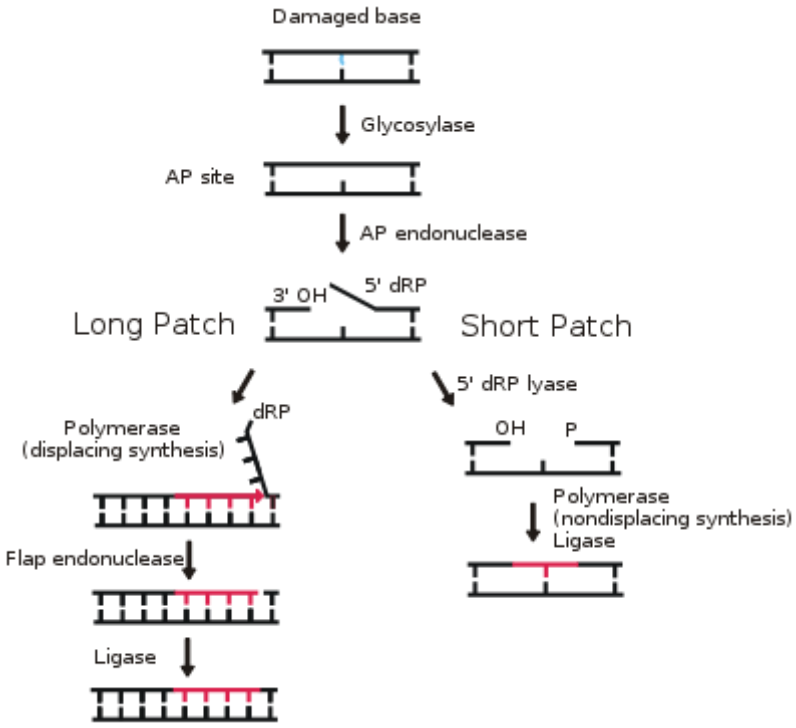


- **Ionizing radiation**: X rays, gamma rays, create free radicals which react and damage DNA, may cause DNA breakage and other damages like alteration in bases
- **Mutation "hot spots"**: some DNA sequences are more likely to mutate, including long stretches of single nucleotide and tandem repeats

Base Excision Repair

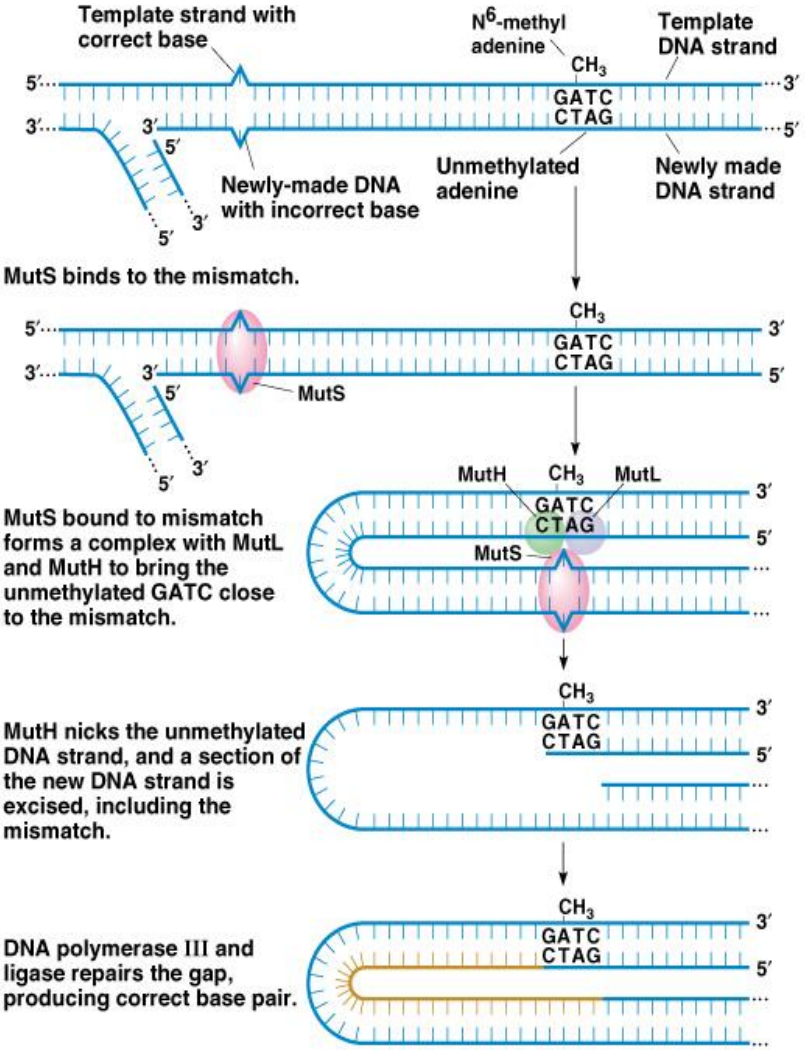
- A cellular mechanism that repairs damaged DNA throughout the cell cycle. It is responsible primarily for removing small, non-helix-distorting base lesions. BER is initiated by DNA glycosylases, which recognize and remove specific damaged or inappropriate bases, forming AP sites. These are then cleaved by an AP endonuclease. The resulting single-strand break is then

processed by either short-patch (where a single nucleotide is replaced) or long-patch BER (where 2-10 nucleotides are replaced)



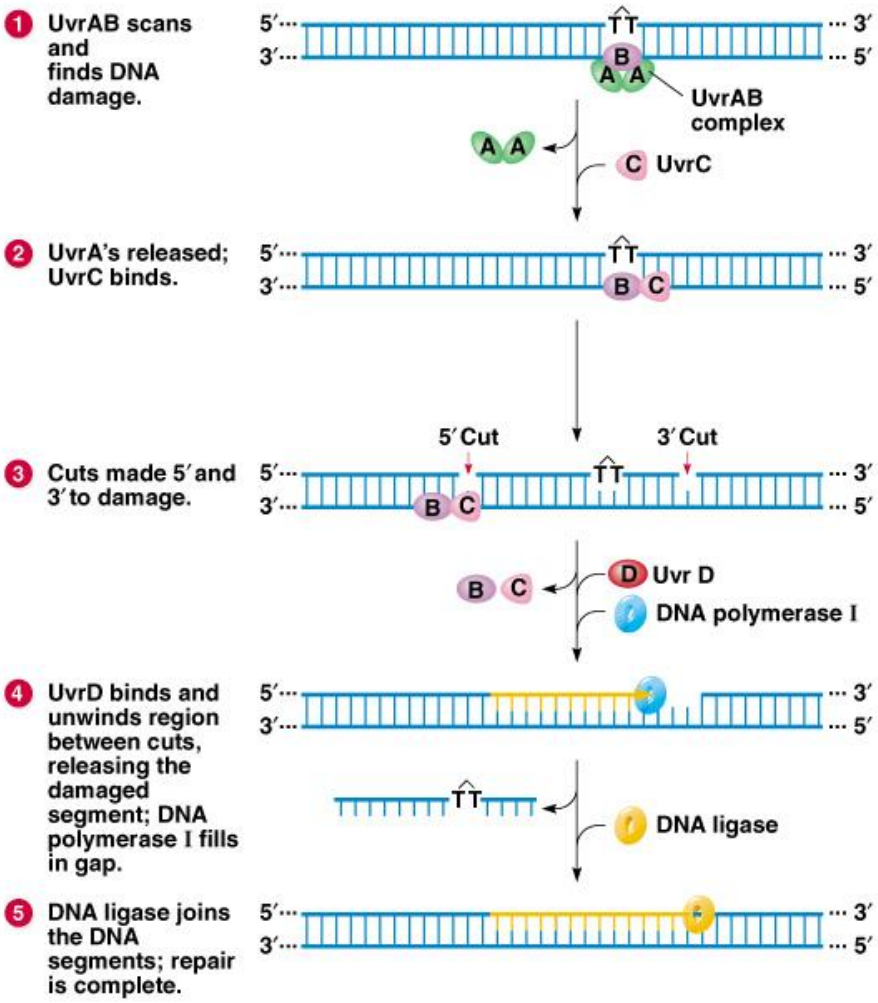
DNA Mismatch Repair

- Mismatch repair is strand-specific. During DNA synthesis the newly synthesised (daughter) strand will commonly include errors. In order to begin repair, the mismatch repair machinery distinguishes the newly synthesised strand from the template (parental). In gram-negative bacteria, transient hemimethylation distinguishes the strands (the parental is methylated and daughter is not). However, in other prokaryotes and eukaryotes, the exact mechanism is not clear
- Three of these proteins are essential in detecting the mismatch and directing repair machinery to it: MutS, MutH and MutL (MutS is a homologue of HexA and MutL of HexB)
- MutS forms a dimer ($MutS_2$) that recognises the mismatched base on the daughter strand and binds the mutated DNA. MutH binds at hemimethylated sites along the daughter DNA, but its action is latent, being activated only upon contact by a MutL dimer ($MutL_2$), which binds the MutS-DNA complex and acts as a mediator between $MutS_2$ and MutH, activating the latter. The DNA is looped out to search for the nearest d(GATC) methylation site to the mismatch, which could be up to 1 kb away
- Upon activation by the MutS-DNA complex, MutH nicks the daughter strand near the hemimethylated site and recruits DNA Helicase II to separate the two strands with a specific 3' to 5' polarity. The entire MutSHL complex then slides along the DNA in the direction of the mismatch, liberating the strand to be excised as it goes
- The single-strand gap created by the exonuclease can then be repaired by DNA Polymerase III (assisted by single-strand-binding protein), which uses the other strand as a template, and finally sealed by DNA ligase



Nucleotide Excision Repair

- particularly important excision mechanism that removes DNA damage induced by ultraviolet light (UV). UV DNA damage results in bulky DNA adducts - these adducts are mostly thymine dimers and 6,4-photoproducts
- The process of nucleotide excision repair is controlled in *Escherichia coli* by the UvrABC endonuclease enzyme complex, which consists of four Uvr proteins: UvrA, UvrB, UvrC, and DNA helicase II (sometimes also known as UvrD in this complex)
- First, a UvrA-UvrB complex scans the DNA, with the UvrA subunit recognizing distortions in the helix, caused for example by pyrimidine dimers. When the complex recognizes such a distortion, the UvrA subunit leaves and an UvrC protein comes in and binds to the UvrB monomer and, hence, forms a new UvrBC dimer
- UvrB cleaves a phosphodiester bond 4 nucleotides downstream of the DNA damage, and the UvrC cleaves a phosphodiester bond 8 nucleotides upstream of the. DNA helicase II (sometimes called UvrD) then comes in and removes the excised segment by actively breaking the hydrogen bonds between the complementary bases. The resultant gap is then filled in using DNA polymerase I and DNA ligase
- The basic excision process is very similar in higher cells, but these cells usually involve many more proteins – *E.coli* is a simple example



Xeroderma pigmentosa

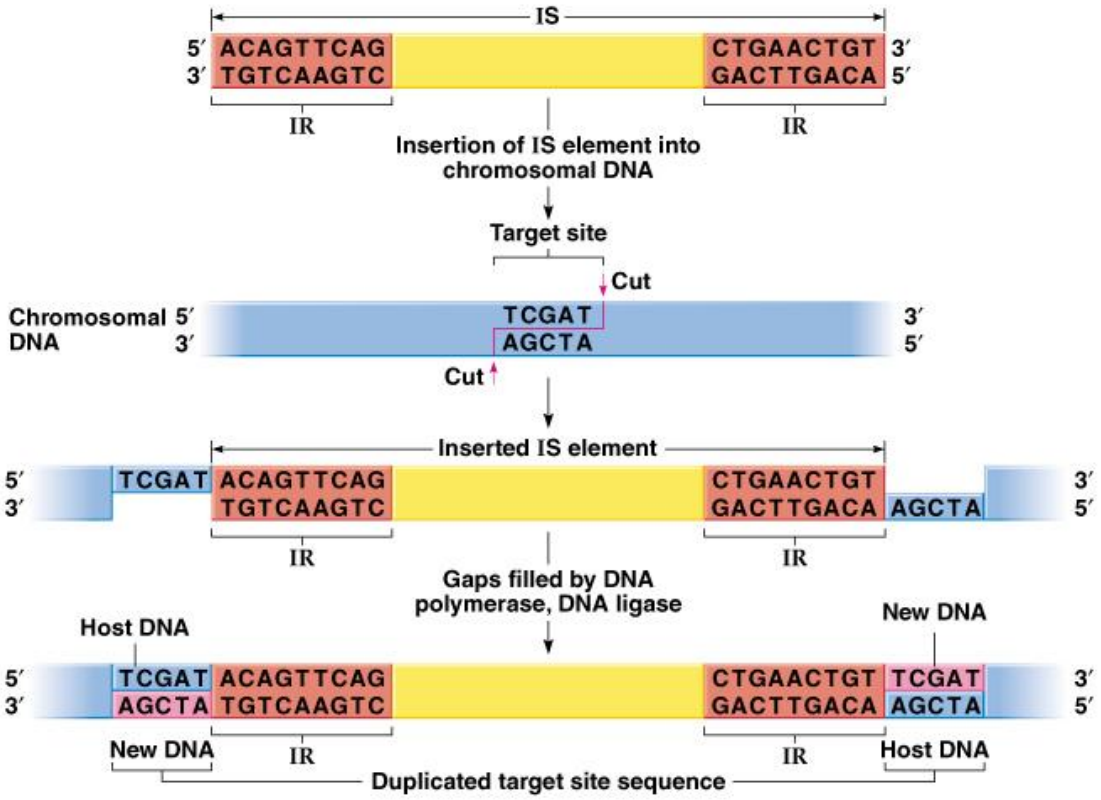
- An autosomal recessive genetic disorder of DNA repair in which the ability to repair damage caused by ultraviolet (UV) light is deficient
- Multiple basal cell carcinomas (basaliomas) and other skin malignancies frequently occur at a young age in those with XP

Dynamic Mutations

- A dynamic mutation is an unstable heritable element where the probability of expression of a mutant phenotype is a function of the number of copies of the mutation
- Frequently caused by mutations in trinucleotide repeats e.g. fragile X syndrome, which affects 1/2500 children
- Trinucleotide repeat, GGC is found in normal individuals who have 6- 54 tandem repeats, average about 30, but affected individuals have 230-2300
- Before the *FMR1* gene was discovered, analysis of pedigrees showed the presence of male carriers who were asymptomatic, with their grandchildren affected by the condition at a higher rate than their siblings suggesting that genetic anticipation was occurring
- Loss of function of *FMR1* (Fragile mental retardation 1 gene) caused by replication slippage

Transposable Elements

- A **transposable element** (**TE** or **transposon**) is a DNA sequence that can change its position within the genome, sometimes creating or reversing mutations and altering the cell's genome size
- They are generally considered non-coding DNA, although it has been shown that TEs are important in genome function and evolution
- TEs are assigned to one of two classes according to their mechanism of transposition, which can be described as either *copy and paste* (class I TEs) or *cut and paste* (class II TEs)



Topic 3: Prokaryotic Gene Regulation

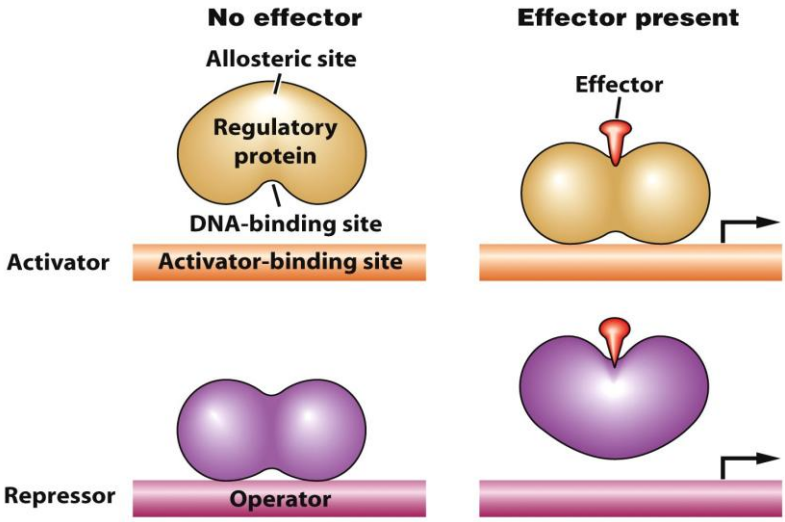
Gene regulation in prokaryotes

- Most genes are not used all the time - how is expression regulated?
- **Anabolic** enzyme synthesis: production of complex molecules (e.g. amino acid synthesis), usually controlled by repression caused by the end product
- **Catabolic** enzyme synthesis: breakdown of complex molecules (e.g. lactose degradation), usually to obtain energy, induced by presence of substrate
- Enzyme synthesis is regulated by small molecules called effectors
- **Effectors** effect allosteric enzymes which control enzymes synthesis

Initiation of Transcription

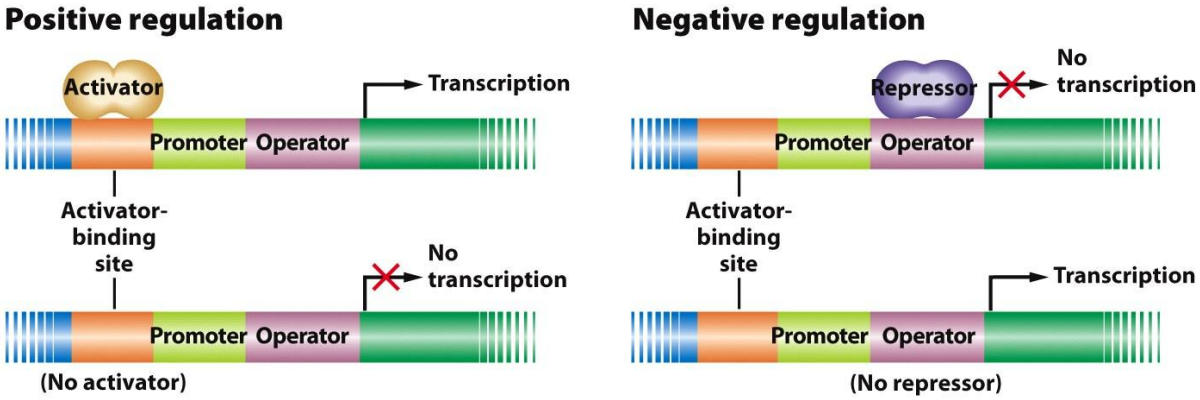
- A **sigma factor** (**σ factor**) is a protein needed only for initiation of RNA synthesis. It is a bacterial transcription initiation factor that enables specific binding of RNA polymerase to gene promoters

- Once σ binds to promoter, the poly III and σ factor complex initiate DNA unwinding
- Each bacteria contains a variety of σ factors, with slight differences in promoter recognition and preference. Alternate σ factors adjust the specificity of DNA sequence for binding to promoters
- Note that in bacteria there is always a basal level of transcription, so 'off' really means very little transcription



Types of Regulation

- **Positive regulation**: default state is transcription off, and binding of a regulatory protein required to turn transcription on
- **Negative regulation**: default state is transcription on unless transcription is turned off by a repressor binding upstream from the start site

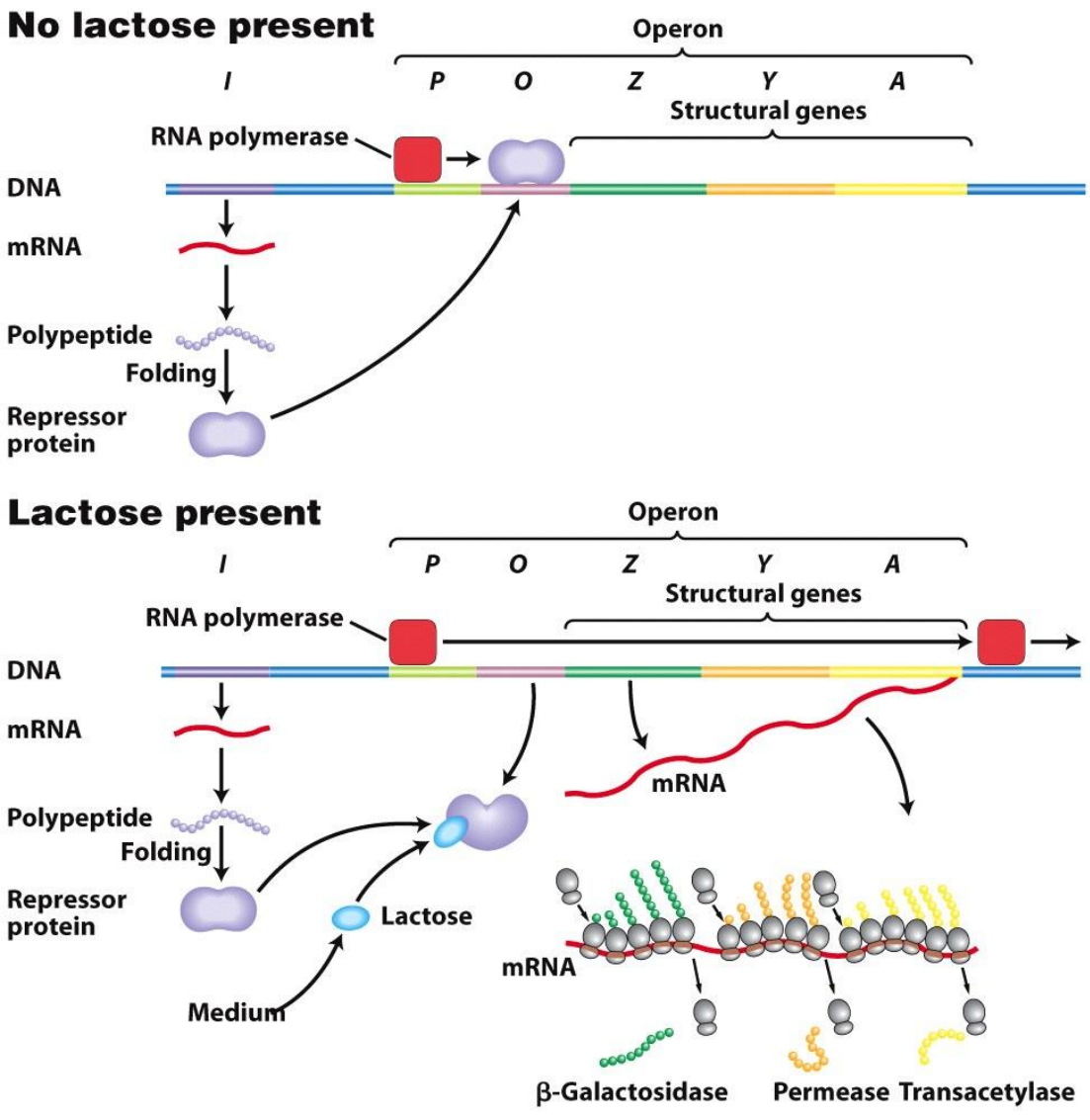


Lactose metabolism in *E. Coli*

- Lactose is broken down into two sugars by an enzyme called β-galactosidase
- Entry of lactose into the cell requires the operation of another protein, lactose permease
- Transcription of these lactose utilisation genes is by default turned off. If lactose is added to the media, both galactosidase and permease induced similtaneously
- *lacZ* encodes the β-galactosidase
- *lacY* encodes the permease
- lacA encodes β-galactoside transacetylase, function unkown

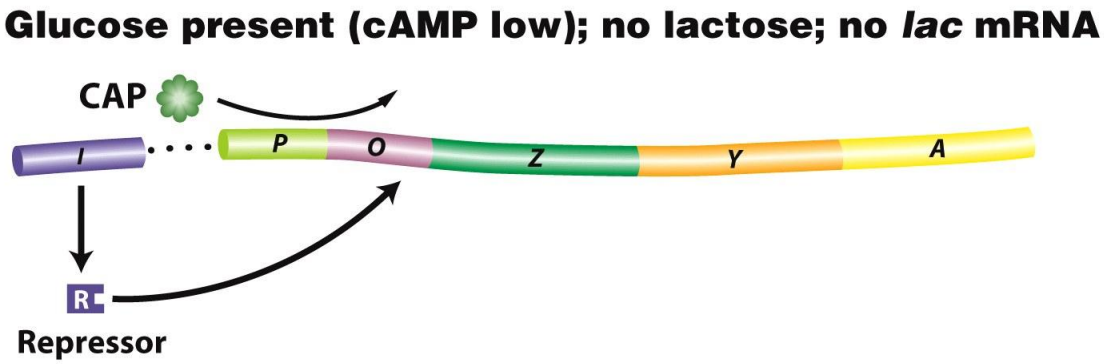
Lac Operon - Negative Regulation

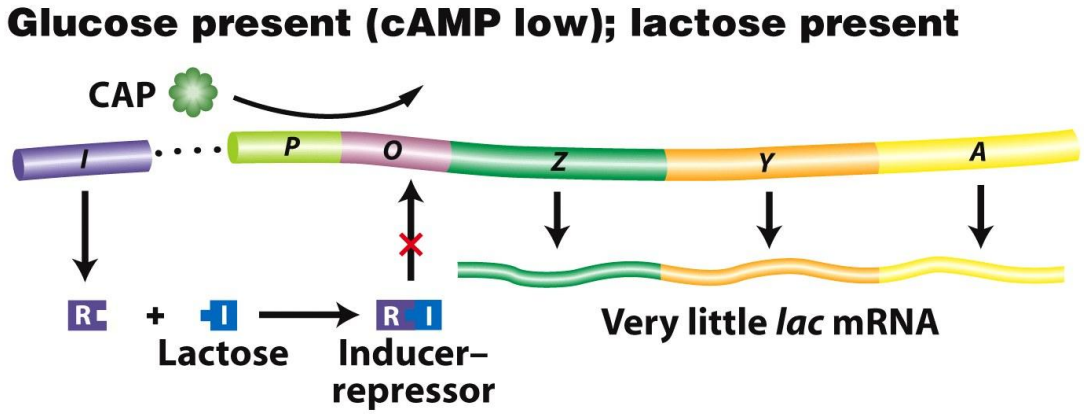
- An **operon** is a functioning unit of genomic DNA containing a cluster of genes under the control of a single promoter
- An operon is made up of 3 basic DNA components:
    - Promoter: a nucleotide sequence that enables a gene to be transcribed. The promoter is recognized by RNA polymerase, which then initiates transcription
    - Operator: a segment of DNA that a repressor binds to. It is classically defined in the lac operon as a segment between the promoter and the genes of the operon. In the case of a repressor, the repressor protein physically obstructs the RNA polymerase from transcribing the genes
    - Structural genes: the genes that are co-regulated by the operon
- The lac operon (a transcriptional unit encoding the structural genes) is controlled by both positive and negative regulation
- *lacI* gene is located outside the promoter (it has its own promoter), and so is outside the lac operon. It encodes for the repressor protein that blocks lac operon transcription
- Lactose serves as an inducer molecule, which binds to the repressor protein, dramatically reducing its binding affinity for the operator, and hence removing it from the DNA

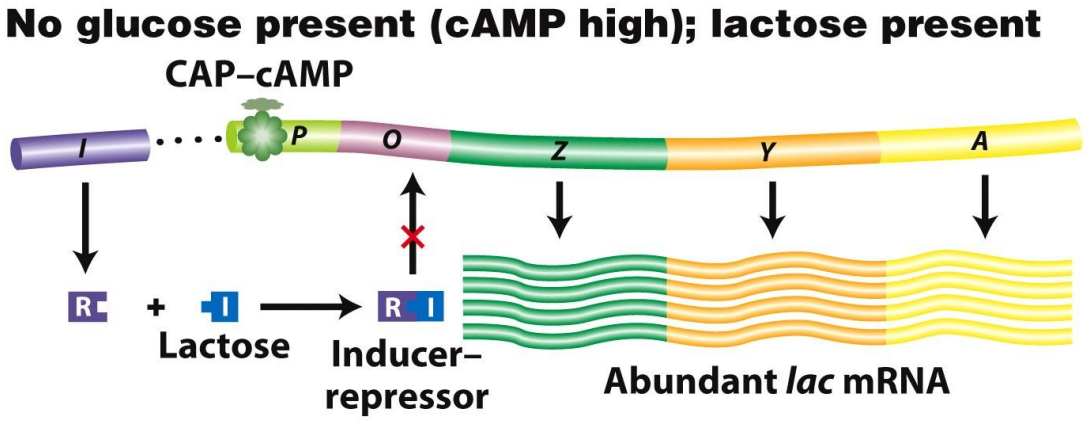


Lac Operon - Positive Regulation

- Glucose is a preferred carbon source, so if both glucose and lactose are present in medium no *lac* operon transcription should occur. This inhibition of *lac* operon is mediated by cyclic adenosine monophosphate (cAMP)
- Cyclic AMP is a molecule whose concentration regulated by glucose metabolism. Glucose high - cAMP low, and glucose low - cAMP high
- A protein called CAP binds with cAMP to form a complex which binds to the CAP site on the promoter, and recruits the RNA polymerase to begin transcription
- If glucose is low, then cAMP is high, and so the CAP site complex is 'ready to go'

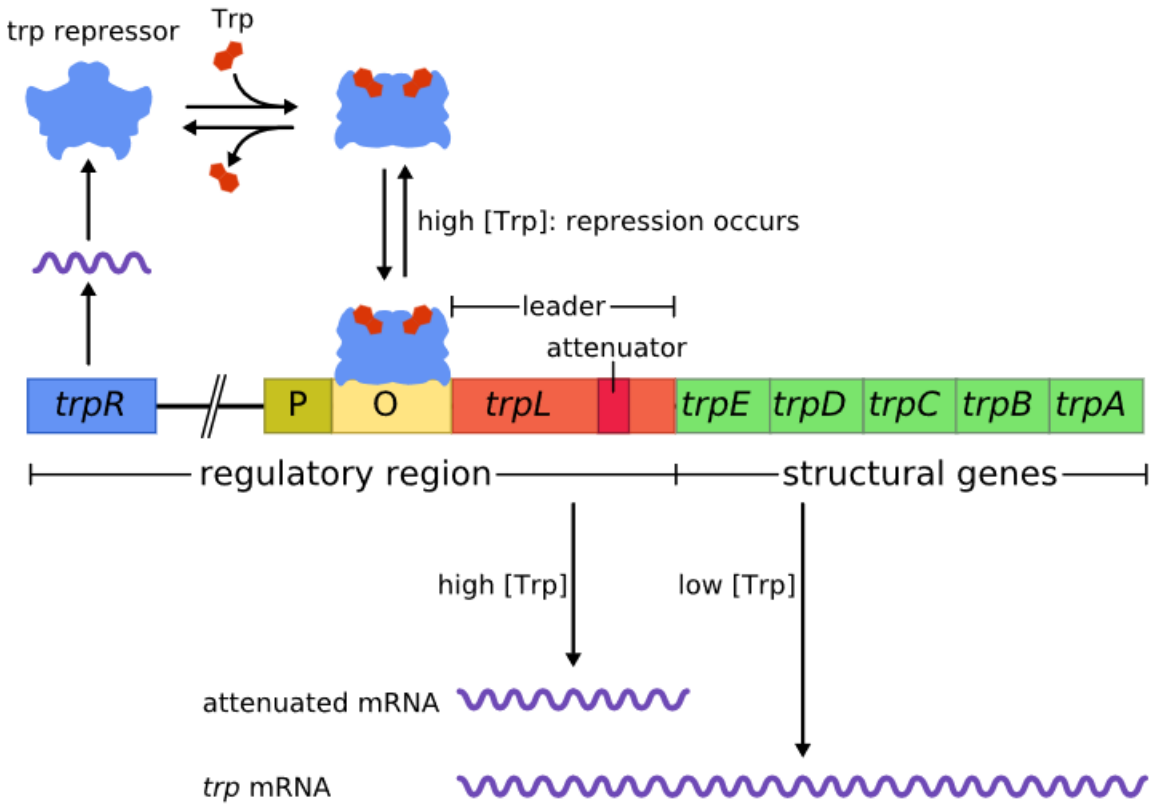






Tryptophan Operon

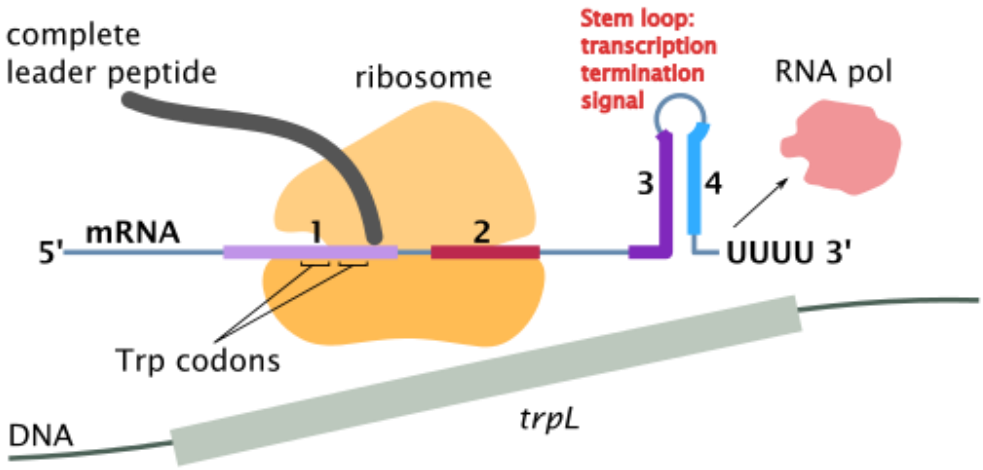
- The tryptophan operon codes for five enzymes, with the genes in same order as they function in biosynthetic pathway
- It is an example of a repressible operon, as its default state is on, but can be turned off by the amino acid tryptophan is added
- The operon operates by a negative repressible feedback mechanism. The repressor for the trp operon is produced upstream by the trpR gene, which is constitutively expressed at a low level. Synthesized TrpR monomers associate into tetramers
- When tryptophan is present, these tryptophan repressor tetramers bind to tryptophan, causing a change in the repressor conformation, allowing the repressor to bind to the operator. This prevents RNA polymerase from binding to and transcribing the operon, so tryptophan is not produced from its precursor
- When tryptophan is not present, the repressor is in its inactive conformation and cannot bind the operator region, so transcription is not inhibited by the repressor



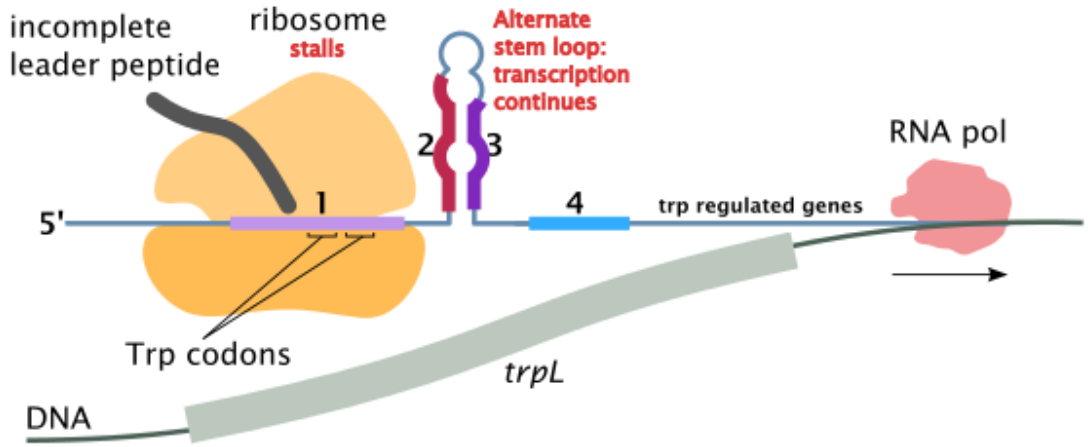
- Attenuation is a second mechanism of negative feedback in the *trp* operon. The trpR repressor decreases gene expression by altering the initiation of transcription, while attenuation does so by altering the process of transcription that's already in progress
- At the beginning of the transcribed genes of the *trp* operon is a sequence of at least 130 nucleotides termed the leader transcript (trpL)
- Part of the leader transcript codes for a short polypeptide of 14 amino acids, termed the leader peptide. This peptide contains two adjacent tryptophan residues, which is unusual, since tryptophan is a fairly uncommon amino acid

- If the ribosome attempts to translate this peptide while tryptophan levels in the cell are low, it will stall at either of the two trp codons. While it is stalled, the ribosome physically shields sequence 1 of the transcript, preventing the formation of the 1-2 secondary structure
- Sequence 2 is then free to hybridize with sequence 3 to form the 2-3 structure, which then prevents the formation of the 3-4 termination hairpin, which is why the 2-3 structure is called an anti-termination hairpin, and RNA polymerase is free to continue transcribing
- If tryptophan levels in the cell are high, the ribosome will translate the entire leader peptide without interruption and will only stall during translation termination at the stop codon. At this point the ribosome physically shields both sequences 1 and 2. Sequences 3 and 4 are thus free to form the 3-4 structure which terminates transcription.

High level of tryptophan



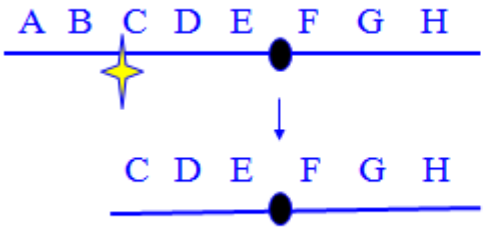
Low level of tryptophan

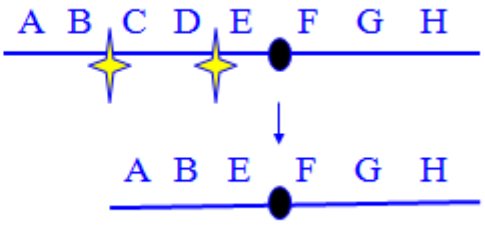


Topic 4: Changes in Chromosome Structure and Number

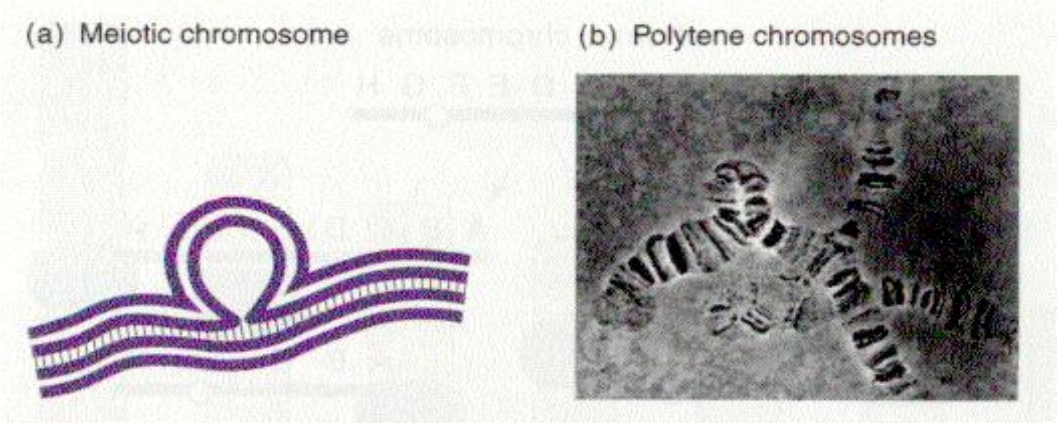
Deletions

- Deletion is the loss of genetic material. Any number of nucleotides can be deleted, from a single base to an entire piece of chromosome
- Changes in chromosome structure usually involve chromosome breakage and broken chromosome ends are highly reactive -tend to join with other broken chromosome ends
- Deletions must involve 2 chromosome breaks induced by ionising radiation. Can be terminal deletion or interstitial deletion
- **Intragenic deletion**: small deletions within a gene inactivate the gene. Can be distinguished from single base mutations because they are non-reversible
- **Multigenic deletion**: remove several thousand genes therefore more severe phenotypes. Homozygous nearly always lethal, and often heterozygotes also not viable. Multigenic deletions uncover recessive alleles and deleterious mutations normally masked by homologous chromosomes





- **Pseudodominance:** unmasking of recessive alleles on other homologue following deletion
- **Cri du chat syndrome**: a rare genetic disorder due to a missing part (deletion) of chromosome 5. Its name is a French term (*cat-cry* or *call of the cat*) referring to the characteristic cat-like cry of affected children
- **Prader–Willi syndrome:** a rare genetic disorder in which seven genes (or some subset thereof) on chromosome 15 (q 11–13) are deleted on the paternal chromosome
- Chromosome deletions often arise in cancer cells, especially solid tumors
- Polyploid plant more tolerant to deletions than diploid

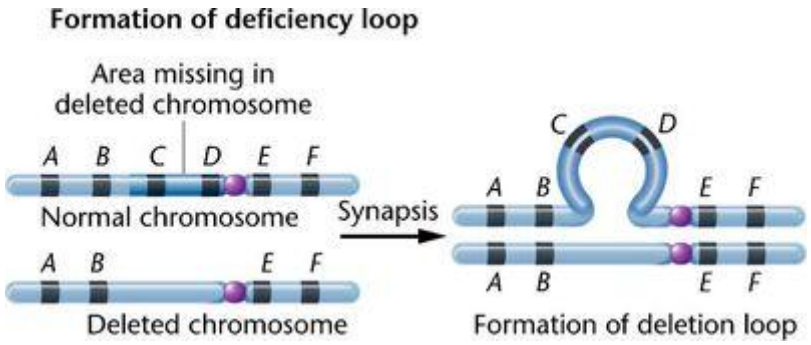


Polytene Chromosomes

- Special chromosomes which replicate without separating into distinct chromosomes
- Found in certain tissues, such as salivary glands in the larval stages of insects

Deletion Loops

- These are an identifiable cytological feature in deletion heterozygotes, and can also be seen in insect polytenes, eg notch wing in Drosophila
- The result of a deletion in one chromosome, and so during recombination the other has nothing to pair with, so forms a loop
- Heterozygotes recombination frequencies between genes flanking the deletion are lower than controls, as the region contains an unpaired chromosome which cannot cross over



Duplications

- In diploids duplication usually present together with standard chromosomes, leaving a set of three copies of the genes in duplication also called gene amplification
- Duplicated heterozygotes show interesting pairing at meiosis
- Tandem duplication: ABCD**BCD**
- Reverse duplicatiopn: ABC**CB**D

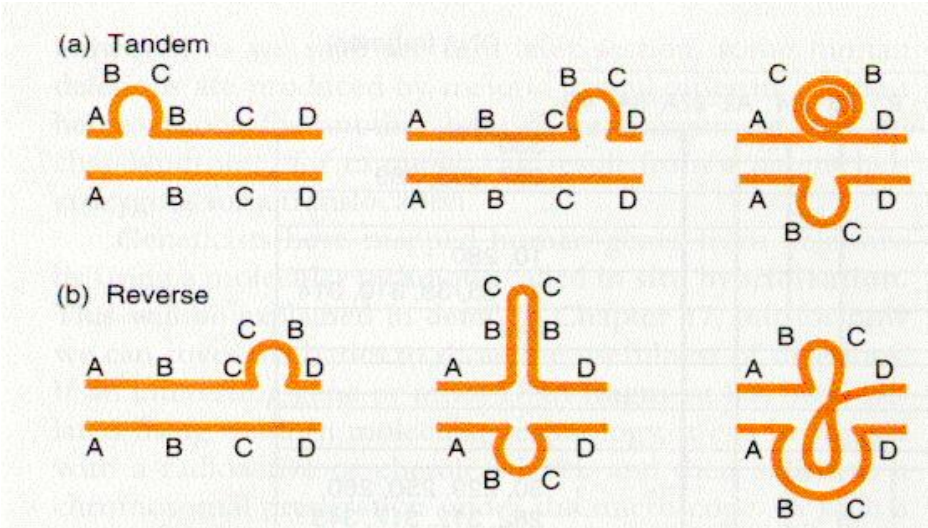
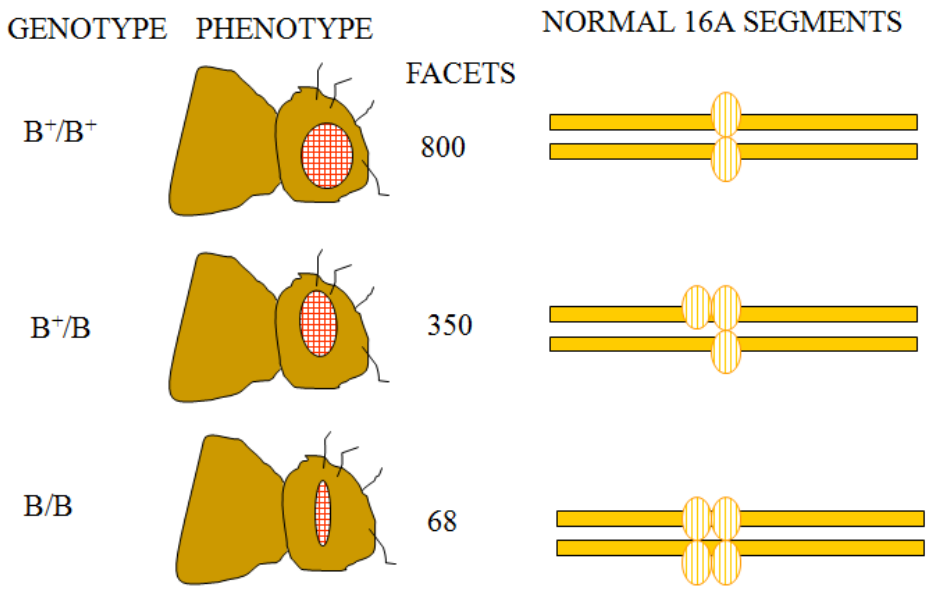
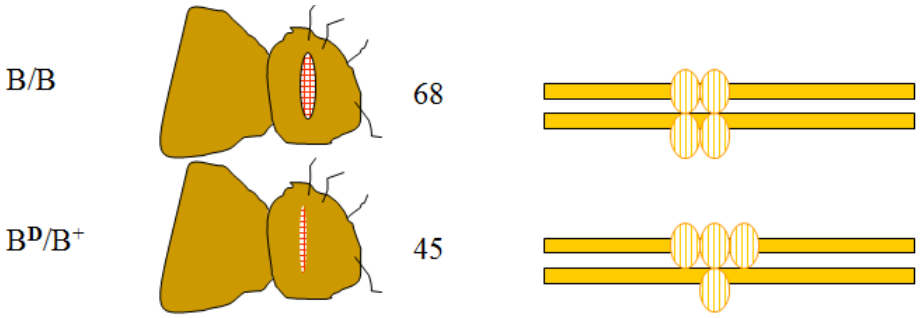


Figure 8-13 Possible pairing configurations in heterozygotes of a standard chromosome and a side-by-side duplication. Duplicated segments may be (a) in tandem or (b) in reverse order. A particular duplicated segment may assume different configurations in different meioses.

- Bar mutants in Drosophila have slit like eyes with reduced number of facets owing to tandem duplication of 16A region, probably arose from unequal crossing over

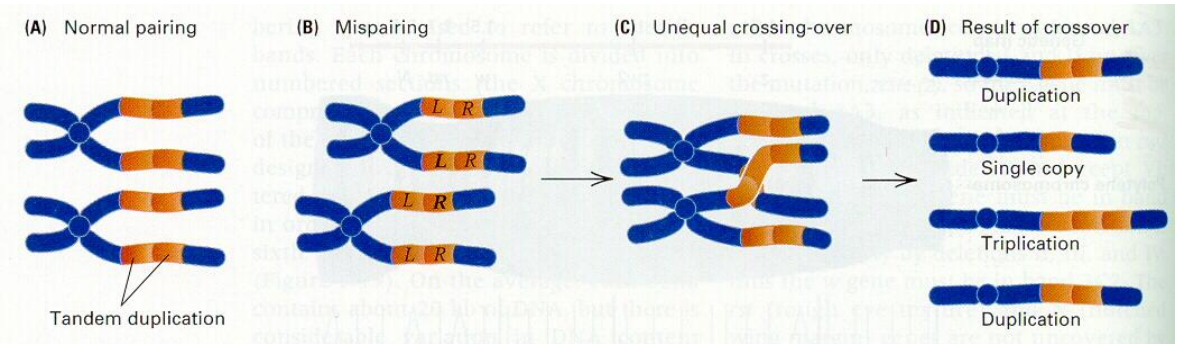


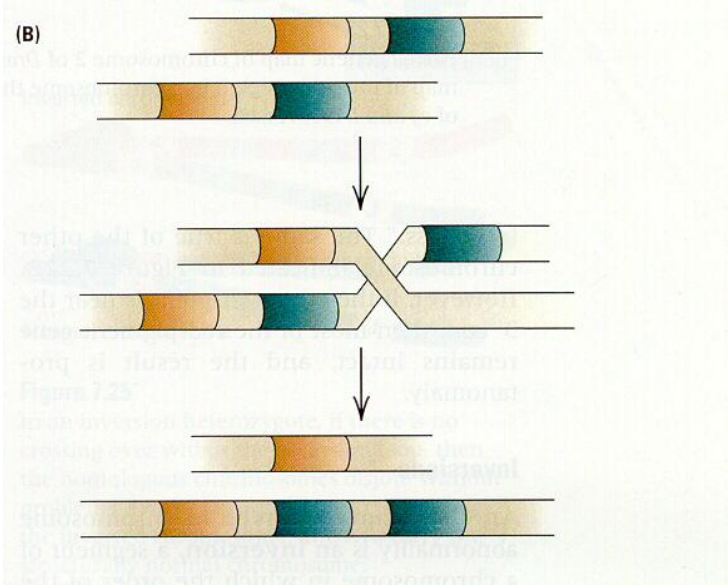
- Double bar mutants result from asymmetric crossing over and have three copies of Bar



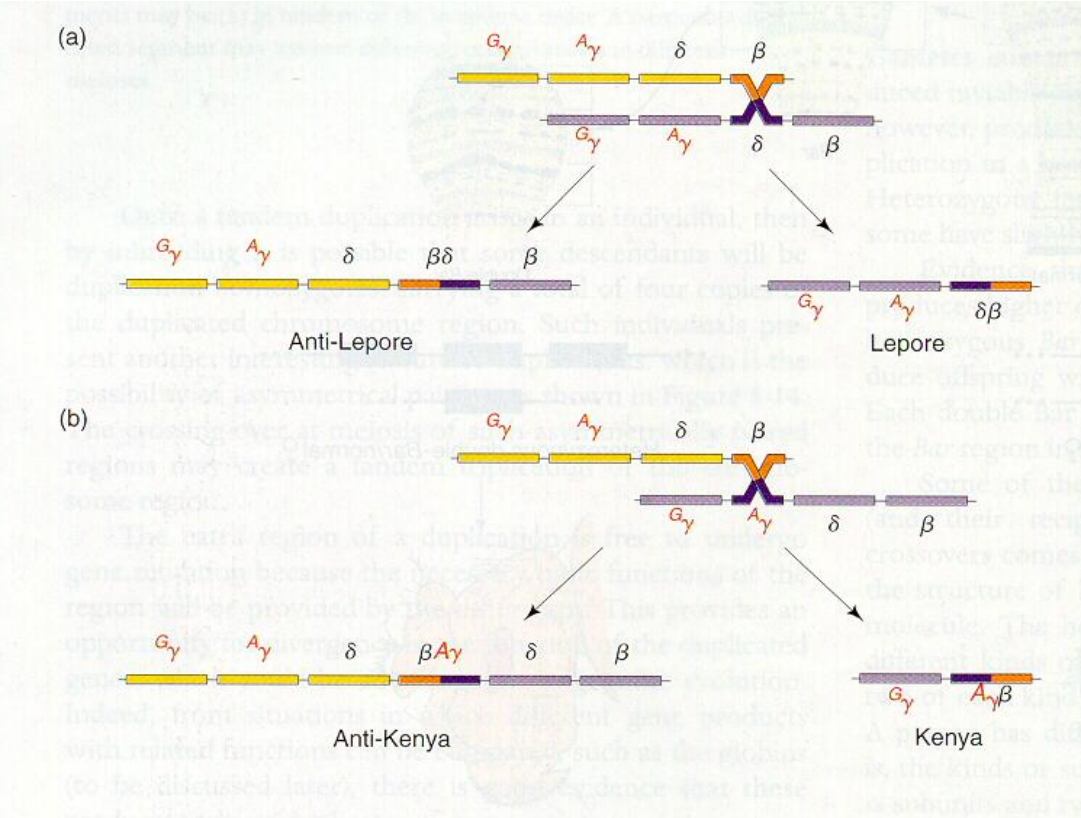
Unequal Crossing Over

- Unequal crossing over is a type of gene duplication or deletion event that deletes a sequence in one strand and replaces it with a duplication from its sister chromatid in mitosis or from its homologous chromosome during meiosis
- It is a type of chromosomal crossover between homologous sequences that are not paired precisely. It exchanges sequences of different links between chromosomes
- Along with gene conversion, it is believed to be the main driver for the generation of gene duplications and is a source of mutation in the genome





- Thought to be important in the evolution of multigene families such as hemoglobin



Red Green Colour Blindness

- Basis of red green colour blindness is unequal crossing over
- Red and green pigment genes Xq28, separated by less than 10 cM. They arose by duplication, and are 96% identical at the amino acid level
- Arrangement of pigment genes in normal men indicates normal X chromosome contains 2 or 3 green pigment genes, which arise by unequal crossing over

Inversions

- Sometimes piece in break rotates 180 degrees and reinserts - no overall change in genetic material
- Usually no major abnormalities or problems, however if break is within an essential gene can get lethal gene mutation
- During meiosis, one chromosome twists around at the ends to pair with the other and forms and inversion loop
- Recombinant class which normally generates recombinants generates lethals owing to gene imbalance or lack of genetic material

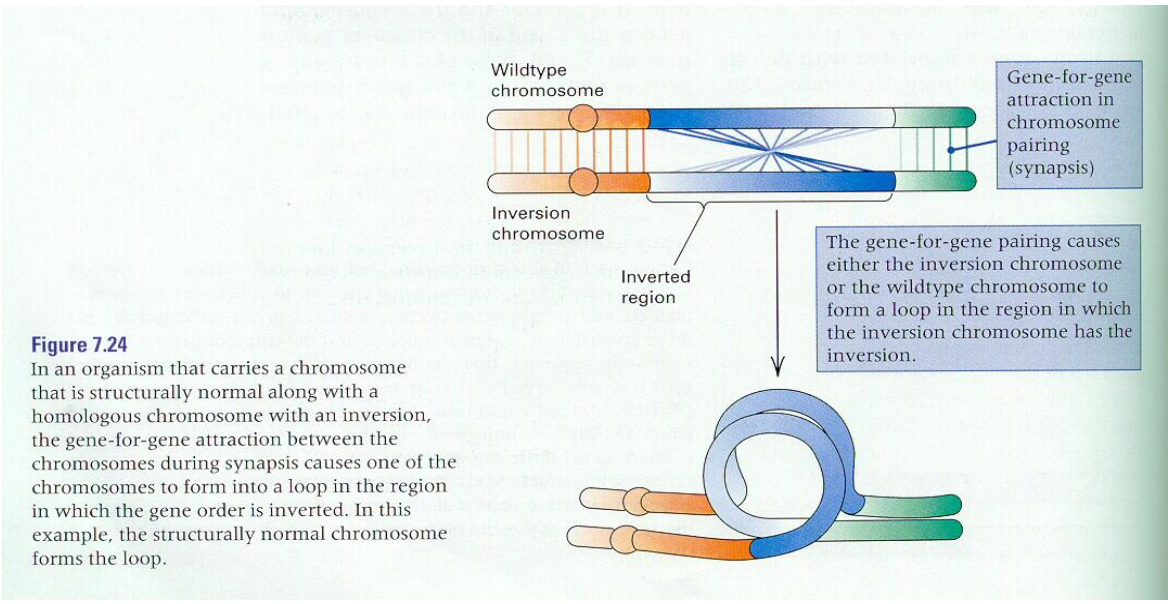
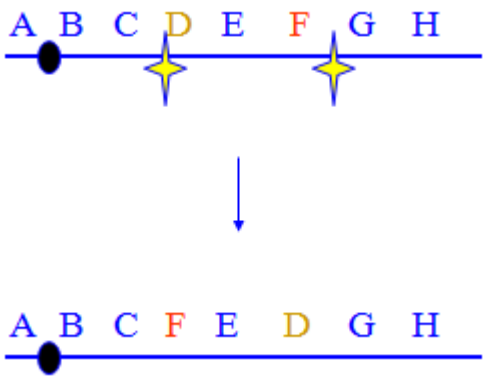


Figure 7.24
In an organism that carries a chromosome that is structurally normal along with a homologous chromosome with an inversion, the gene-for-gene attraction between the chromosomes during synapsis causes one of the chromosomes to form into a loop in the region in which the gene order is inverted. In this example, the structurally normal chromosome forms the loop.

- **Paracentric inversion**: centromere is outside the inversion



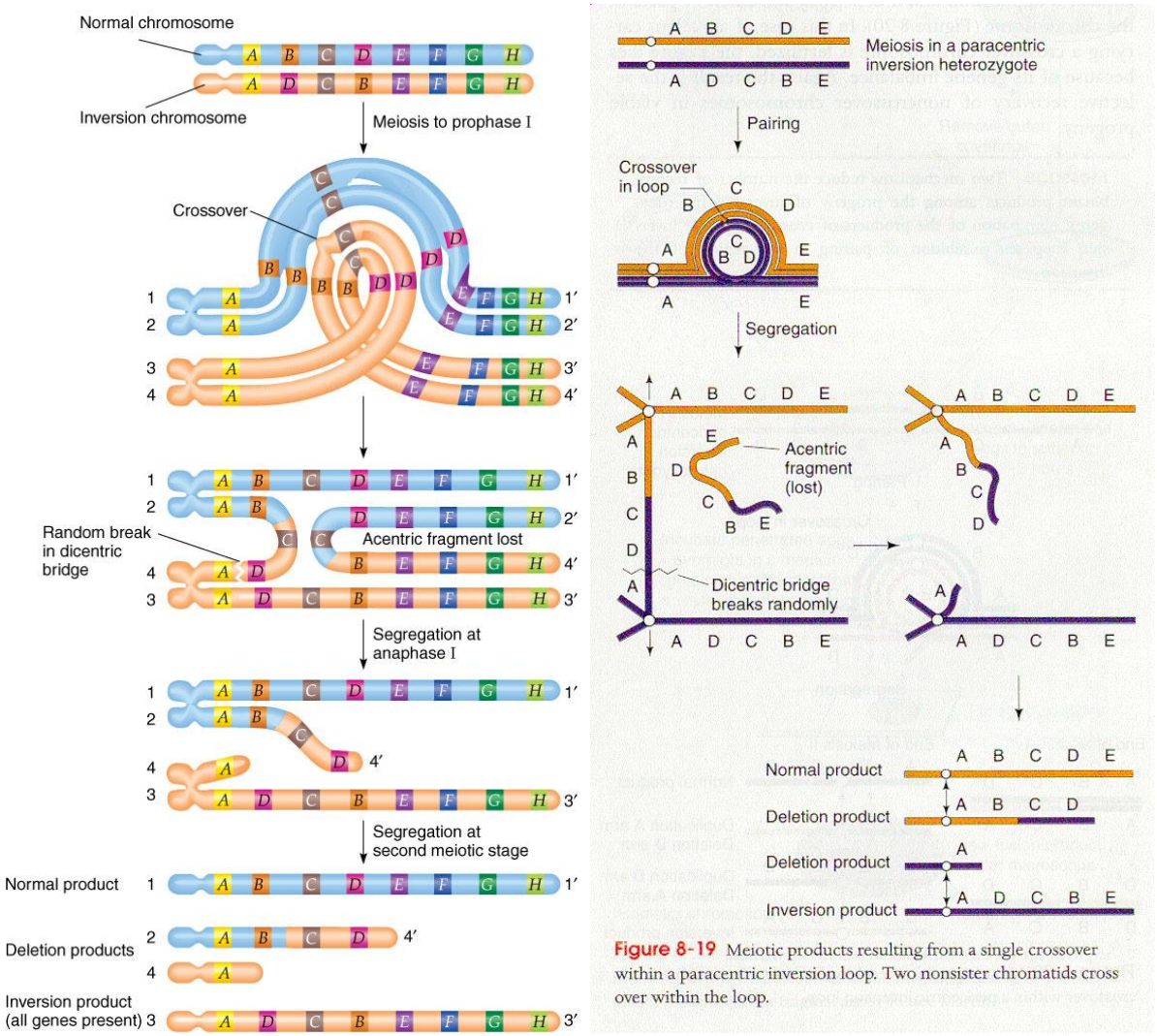
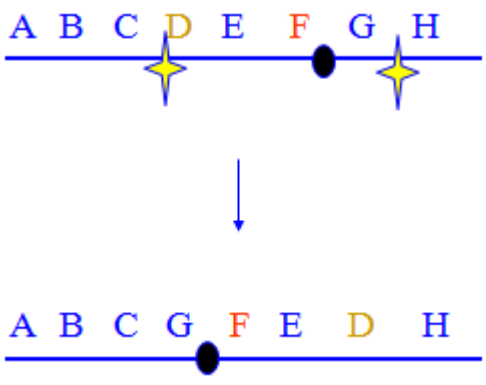
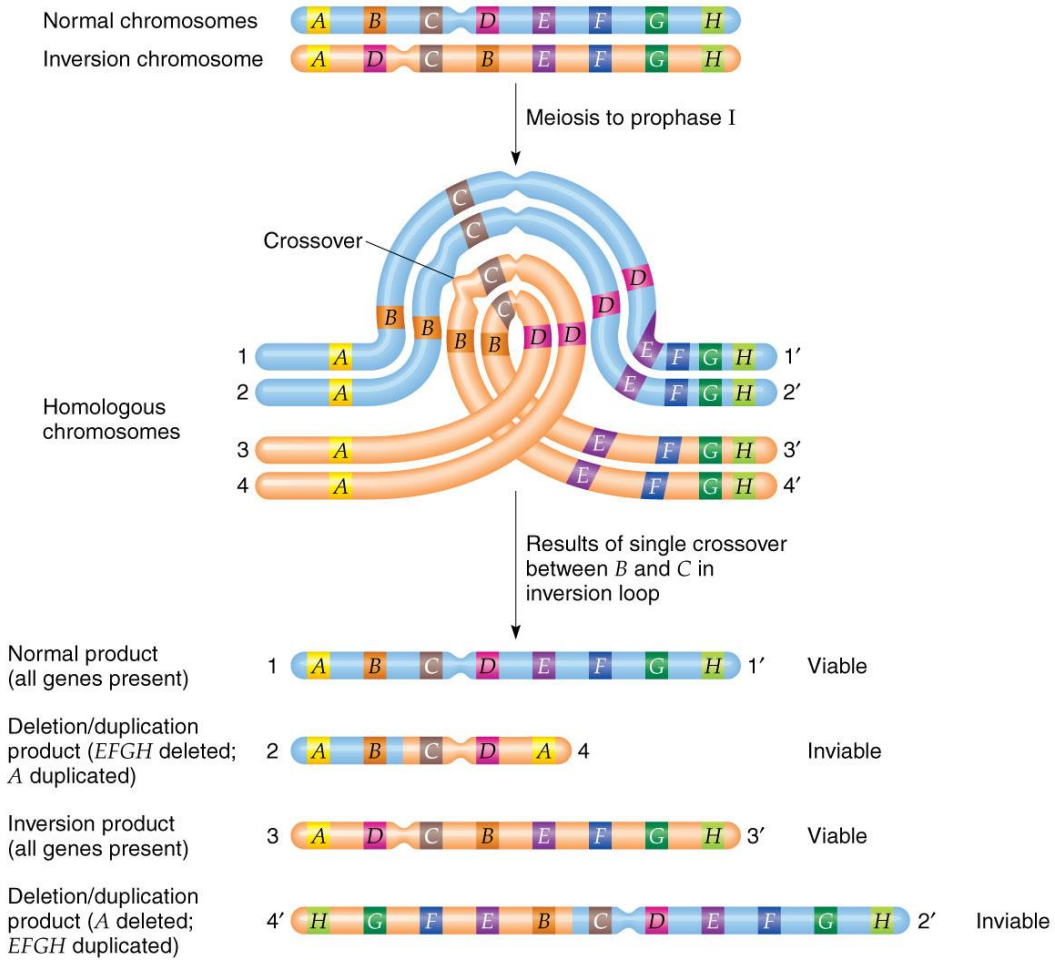


Figure 8-19 Meiotic products resulting from a single crossover within a paracentric inversion loop. Two nonsister chromatids cross over within the loop.

- **Pericentric inversion**: centromere is inside the inversion





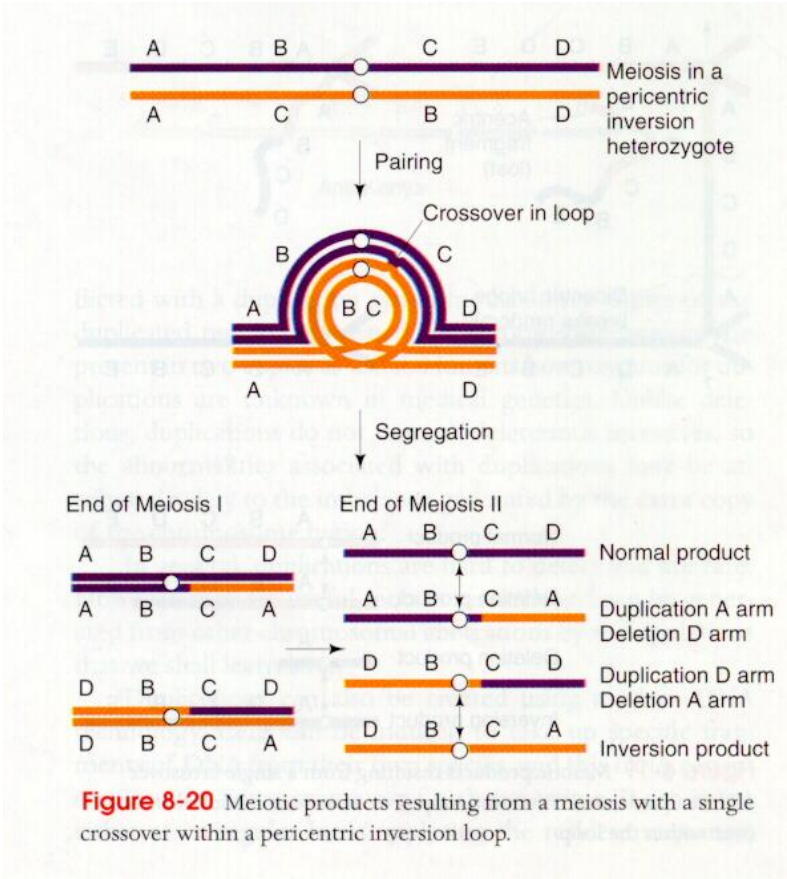


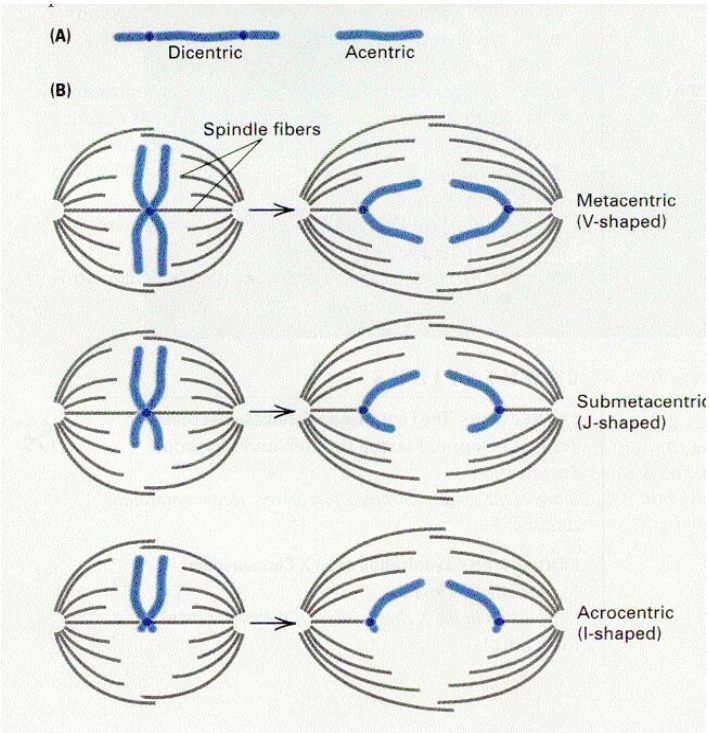
Figure 8-20 Meiotic products resulting from a meiosis with a single crossover within a pericentric inversion loop.

Translocations

- Chromosome mutations in which a change in position of chromosome segments occurs
- Gene sequence moves to a different genomic position
- **Reciprocal translocations**: interchange of parts of 2 non-homologous chromosomes
- **Nonreciprocal translocation**: transfer of genes from one chromosome to another, nonhomologous chromosome
- This means heterozygotes for the reciprocal translocation, but homozygotes for translocation
- Robertsonian translocations: non-reciprocal translocation in which the centromeric regions of two non-homologous chromosomes form a single centromere
- **Position effect** is the effect on the expression of a gene when its location in a chromosome is changed, often by translocation. This has been well described in *Drosophila* with respect to eye color and is known as position effect variegation (PEV)
- **Philadelphia chromosome:** the result of a reciprocal translocation between chromosome 9 and chromosome 22

Centromeres

- Centromeres add genetic stability, and stable transmission relies on single centromere
- In eukaryotes virtually all chromosomes have single centromere and are rod shaped
- **Dicentric** chromosomes: two centromeres, are genetically unstable and unpredictable
- **Acentric chromosomes:** no centromere, lost during cell division
- **Metacentric:** centromere about middle, arms equal, V shape at anaphase
- **Submetacentric:** centromere off centre, J shape at anaphase
- **Acrocentric:** centromere close to end, I shape at anaphase
- **Telocentric:** centromere at end



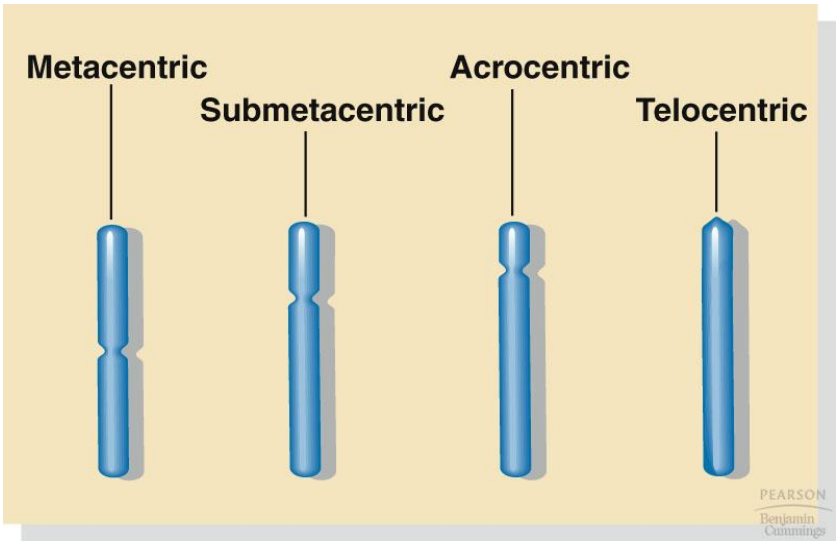


Table 8-1 Human Chromosomes

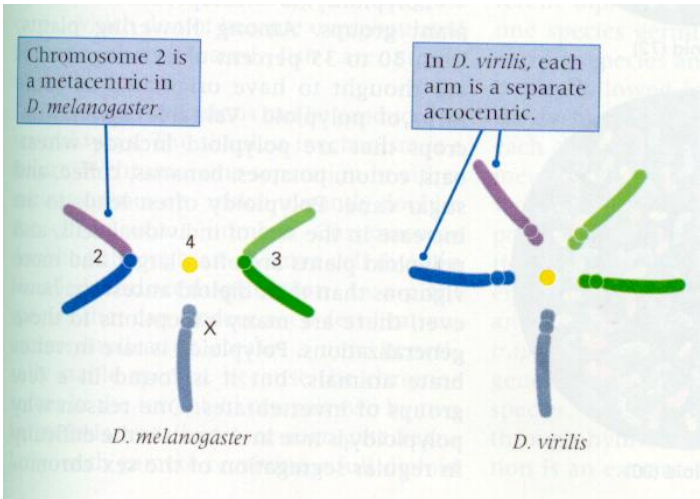
Group	Number	Diagrammatic representation	Relative length*	Centromeric index†
Large chromosomes				
A	1		8.4	48 (M)
	2		8.0	39
	3		6.8	47 (M)
B	4		6.3	29
	5		6.1	29
Medium chromosomes				
C	6		5.9	39
	7		5.4	39
	8		4.9	34
	9		4.8	35
	10		4.6	34
	11		4.6	40
	12		4.7	30
D	13		3.7	17 (A)
	14		3.6	19 (A)
	15		3.5	20 (A)
Small chromosomes				
E	16		3.4	41
	17		3.3	34
	18		2.9	31
F	19		2.7	47 (M)
	20		2.6	45 (M)
G	21		1.9	31
	22		2.0	30
Sex chromosomes				
	X		5.1 (group C)	40
	Y		2.2 (group G)	27 (A)

* Percentage of the total combined length of a haploid set of 22 autosomes.

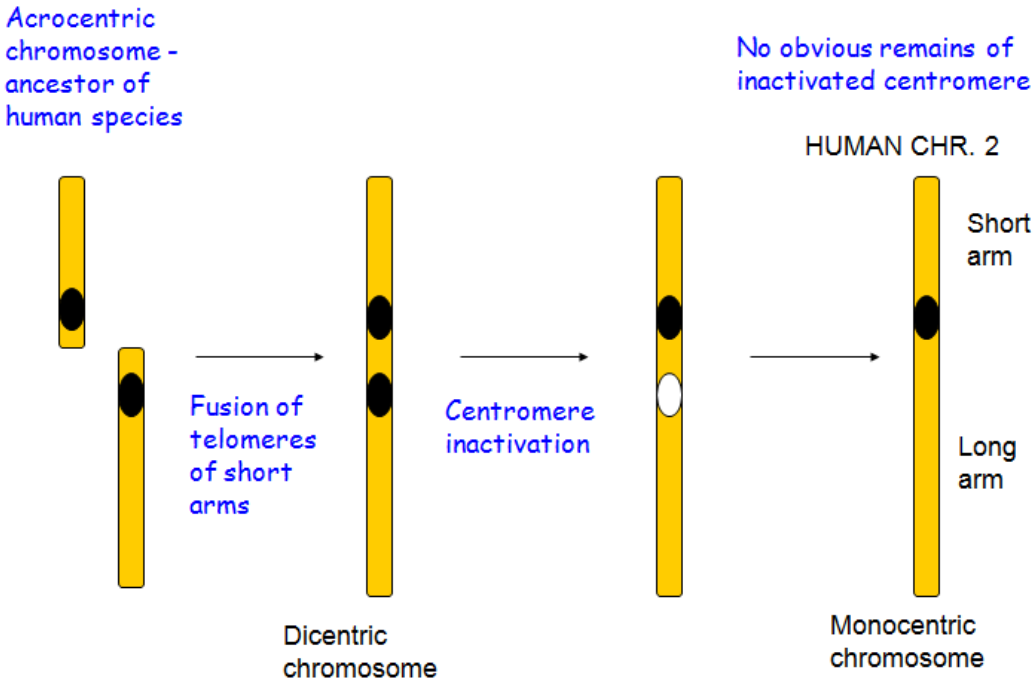
† Percentage of a chromosome's length spanned by its short arm. The four most metacentric chromosomes are indicated by an (M); the four most acrocentric by an (A).

Conservation of Chromosome Arms

- In evolution number of chromosome arms is conserved without conservation of chromosome number
- *D. melanogaster*: 2 large metacentric autosomes
- Other *Drosophila* species: 4 acrocentric autosomes - these correspond arm for arm to large metacentrics in *D. melanogaster*



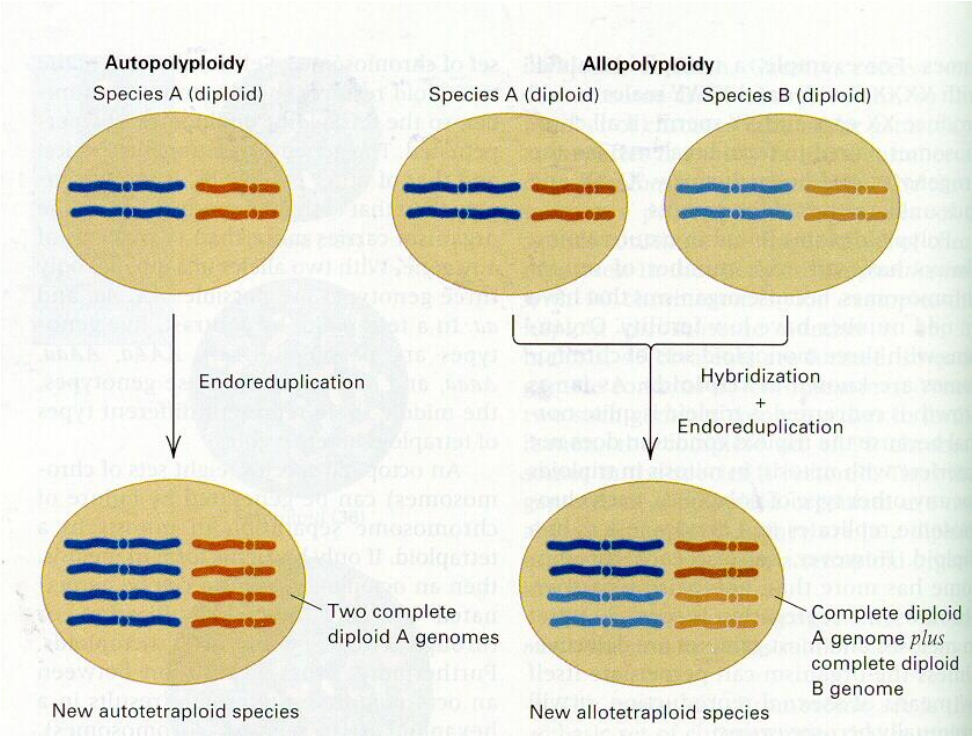
- Humans (2n = 46) have 22 autosomes, while Chimpanzee (2n = 48) have 23 autosomes
- Human metacentric chromosome two formed via fusion of ancestor chromosomes at the telomeres



Ploidy

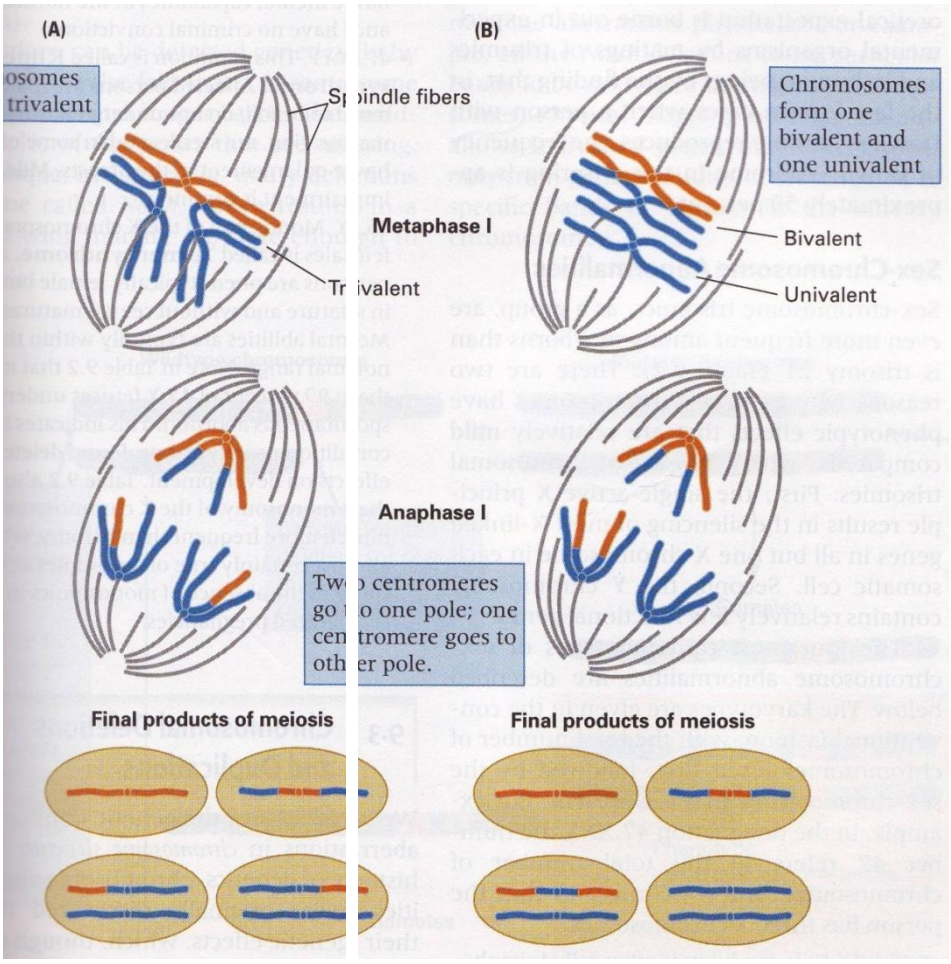
- **Monoploid** number (n): the basic set of chromosomes multiplied in polyploids e.g. A B C
- **Haploid** number: the number of chromosomes in a gamete, only the same as the monoploid number in diploid organisms (e.g. tetraploid has 4 monoploid sets so haploid is diploid)
- Monoploid (1n): fairly rare, found in bees and ants
- **Polyploidy** (3n, 4n, etc): common among flowering plants, often leads to increase in cell size cells can be larger and more vigorous. Most often have even number of sets of chromosomes
- **Triploids** (3n): 3 monoploid sets of chromosomes, e.g. AAA BBB CCC. Growth often normal, however chromosome segregation is upset in meiosis, with gametes defective (sterile). Agricultural applications – seedless varieties of bananas, watermelons

- **Tetraploids** (4n): chromosomes do not separate in meiosis or mitosis. Possibility of differing tetraploid heterozygotes, e.g. AAAA AAAa AAaa Aaaa aaaa
- **Autopolyploidy**: all chromosomes in polyploid from single ancestral species, e.g. *Chrysanthemum* with a diploid number of 18
- **Allopolyploidy**: complete sets of chromosomes from two or more different ancestral species (closely related), e.g. wheat
- A chromosome in a diploid organism is **hemizygous** when only one copy is present

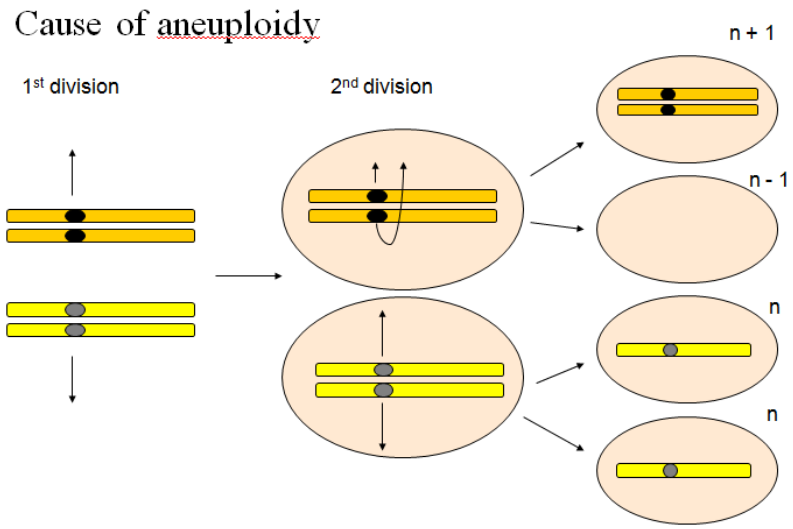


Polysomy

- **Polysomy** is a condition in which an organism has at least one more chromosome than normal, i.e., there may be three or more copies of the chromosome rather than the expected two copies
- Usually **less vigorous** than diploids and usually have **abnormal phenotypes**
- Perhaps surprisingly, the presence of an extra (or missing) single chromosome has a more severe effect than presence of a complete extra set of chromosomes, owing to the imbalance in gene proportions



- **Trisomics** (2n+1): an otherwise diploid organism with an **extra copy** of an individual chromosome; segregation of chromosomes in meiosis is upset
- **Monosomics** (2n-1): a **chromosome is missing, even** more harmful effects than an extra copy of the same chromosome, indeed is often lethal. Allows expression of all recessive alleles, and leads to **non-disjunction** in meiosis
- **Nullisomics** (2n-2): is a genetic condition involving the lack of one of the normal chromosomal pairs for a species. Humans with this condition will not survive
- **Euploidy**: organisms with multiple complete sets of the basic chromosome set, e.g. plants and animals (diploids) and fungi (haploid)
- **Aneuploidy**: changes in parts of chromosome sets leading to unbalanced chromosome complements, e.g. trisomics. Often caused by n**ondisjunction**, the failure of homologous chromosomes or sister chromatids to separate properly during cell division
- **Polysomy** is a condition in which an organism has at least **one more** chromosome than normal, i.e., there may be three or more copies of the chromosome rather than the expected two copies. Aneuploids possess chromosome numbers that are not exact multiples of the haploid number and polysomy is a type of aneuploidy



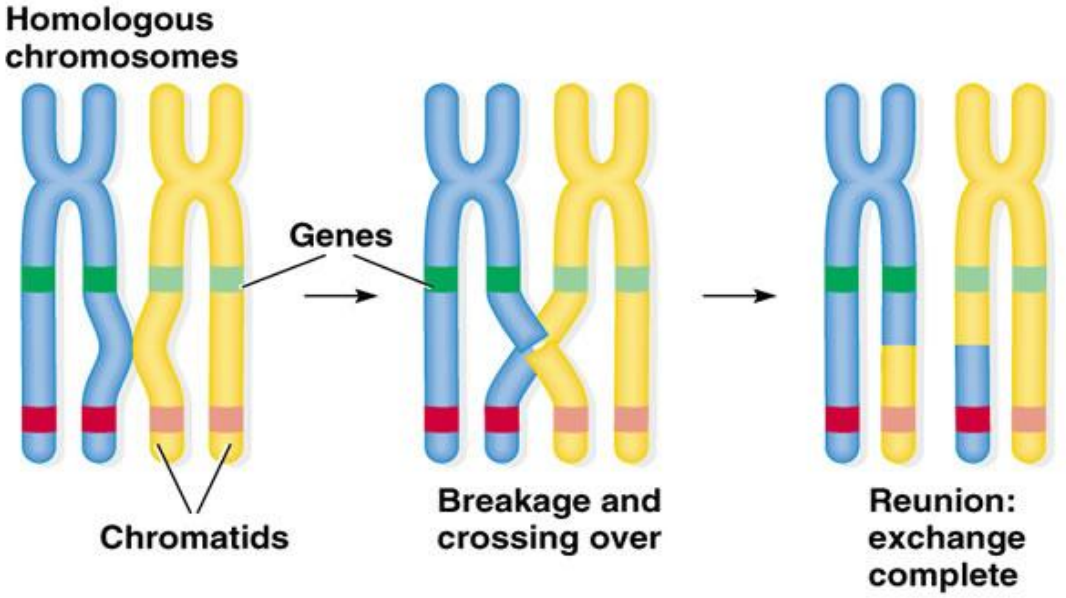
Human Aneuploidy

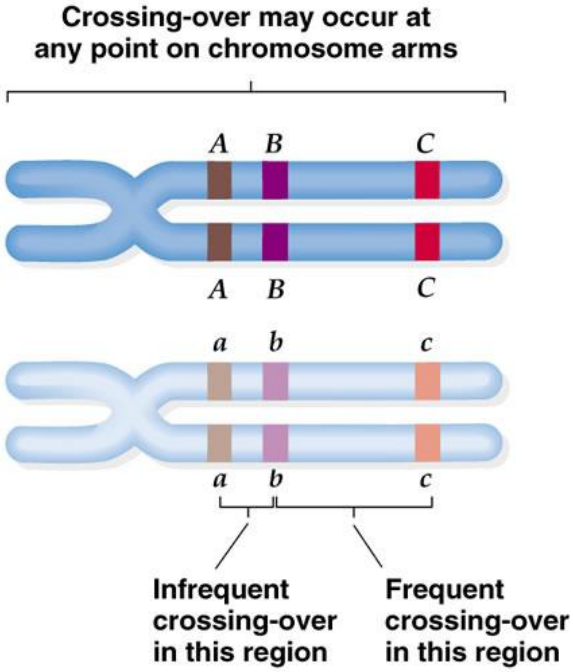
- **Turners syndrome (XO):** 1 X chromosome leading to sterile females with short stature and a characteristic web of skin between neck and shoulder, 1 in 5000 female births
- **Klinefelters syndrome (XXY):** 1 in 1000 live male births with lanky build and mental retardation; sterile
- **Double Y (XYY**): 1 in 1000 live male births, usually fertile with meiosis of the XY type, the extra Y not transmitted
- **Trisomy 21**: the most common viable human trisomy, in 0.15% of live births, caused by non-disjunction of chromosome 21. A major cause of Downs syndrome
- Monosomics for all human autosomes die *in utero*
- **Fragile X syndrome:** associated with the expansion of the CGG trinucleotide repeat affecting the *Fragile X mental retardation 1* (*FMR1*) gene on the X chromosome

Genetic Mapping

- Alfred Sturtevant in 1913 was the first to propose the idea of using recombination to construct a genetic map
- He argued that the recombination frequencies between the genes is additive
- **Genetic linkage** is the tendency of alleles that are located close together on a chromosome to be inherited together during meiosis
- Genes whose loci are nearer to each other are less likely to be separated onto different chromatids during chromosomal crossover, and are therefore said to be genetically *linked*. In other words, the nearer two genes are on a chromosome, the lower is the chance of a swap occurring between them, and the more likely they are to be inherited together
- By working out the number of recombinants it is possible to obtain a measure for the distance between the genes. This distance is expressed in terms of a **genetic map unit (m.u.)**, or a **centimorgan** and is defined as the distance between genes for which one product of meiosis in 100 is recombinant
- But this equivalence is only a good approximation for small percentages; the largest percentage of recombinants cannot exceed 50%, which would be the situation where the two genes are at the extreme opposite ends of the same chromosomes

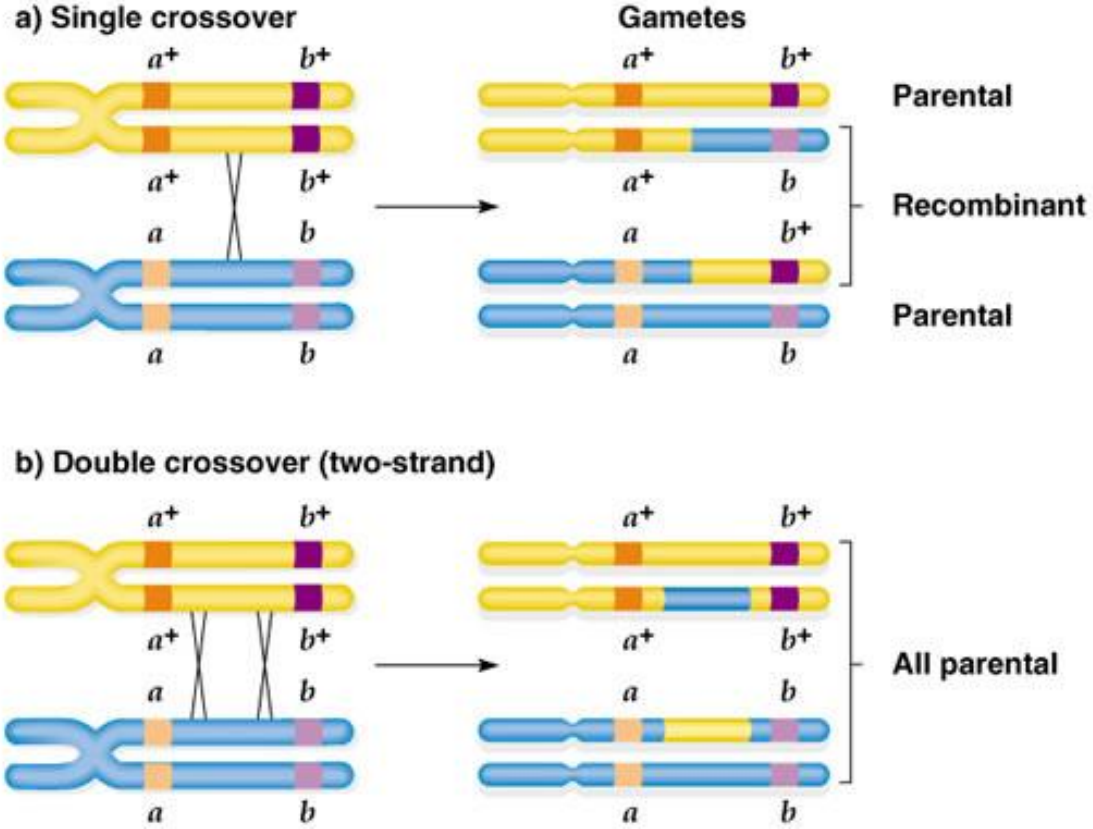
- A linkage map is a genetic map of a species or experimental population that shows the position of its known genes or genetic markers relative to each other in terms of recombination frequency, rather than a specific physical distance along each chromosome
- Note that the crossing over frequency is twice that as the recombination frequency, as the recombination frequency counts chromatids as the denominator, only two of which are recombinant out of four, whereas the crossing over frequency counts chromosomes as the denominator, two of two are involved in a cross over event

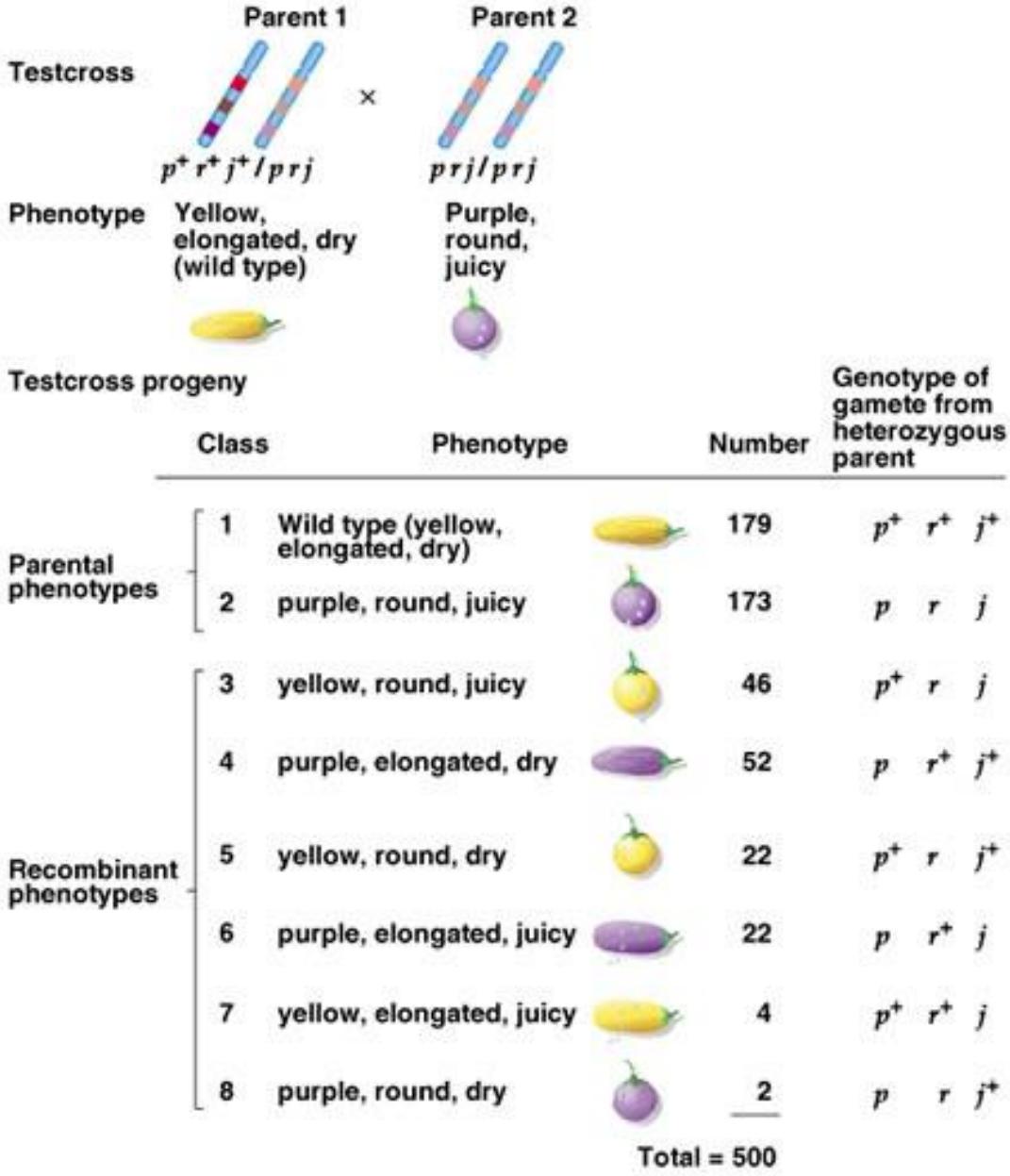




Three-Point Test Cross

- The test cross is usually performed with a double heterozygote recessive
- The most common phenotypes will be the parentals, and the least common will always be those resulting from double cross-over events
- Reasoning backwards we can determine the gene order, and then the distances between the two extreme alleles and the centre can be found by dividing the total number of recombinants of each end allele (single and double) by the total number of gametes



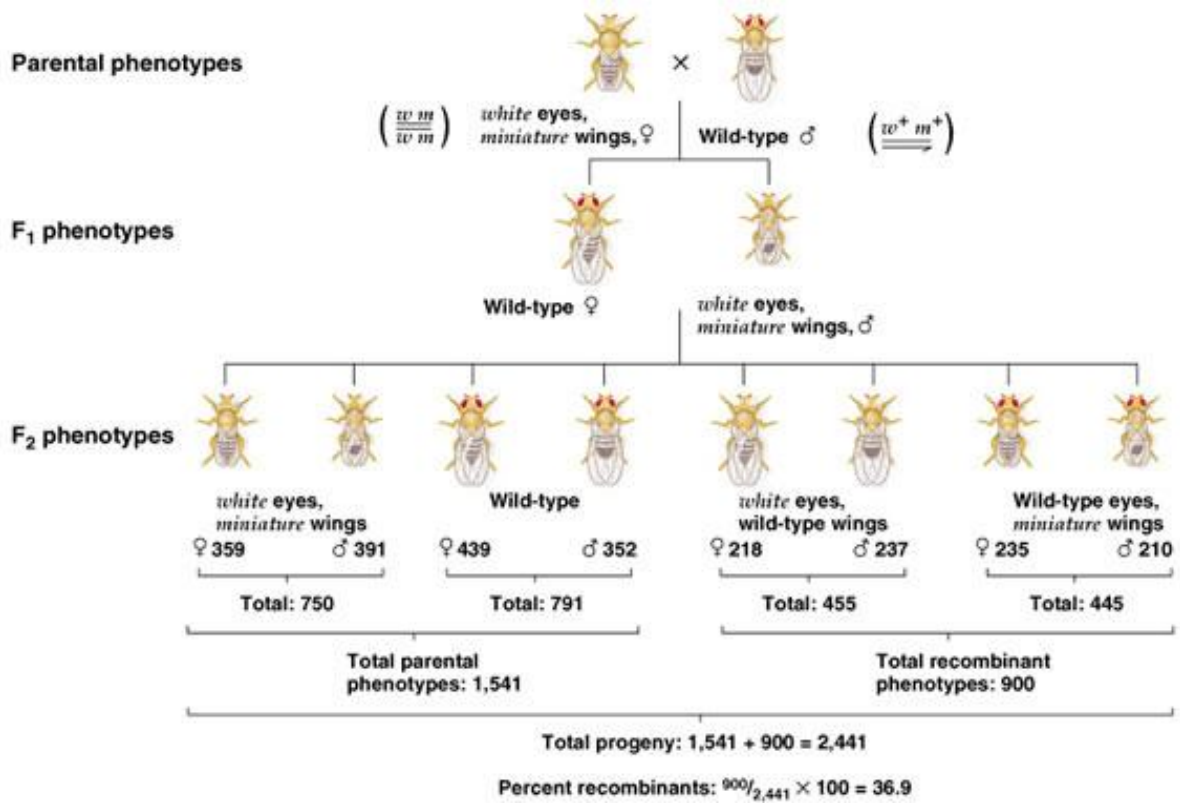


- For example, in this instance 46+52+4+2/500 = 0.208, which is the distance in cm from p to j

Topic 5: Sex Determination and Sex Linkage

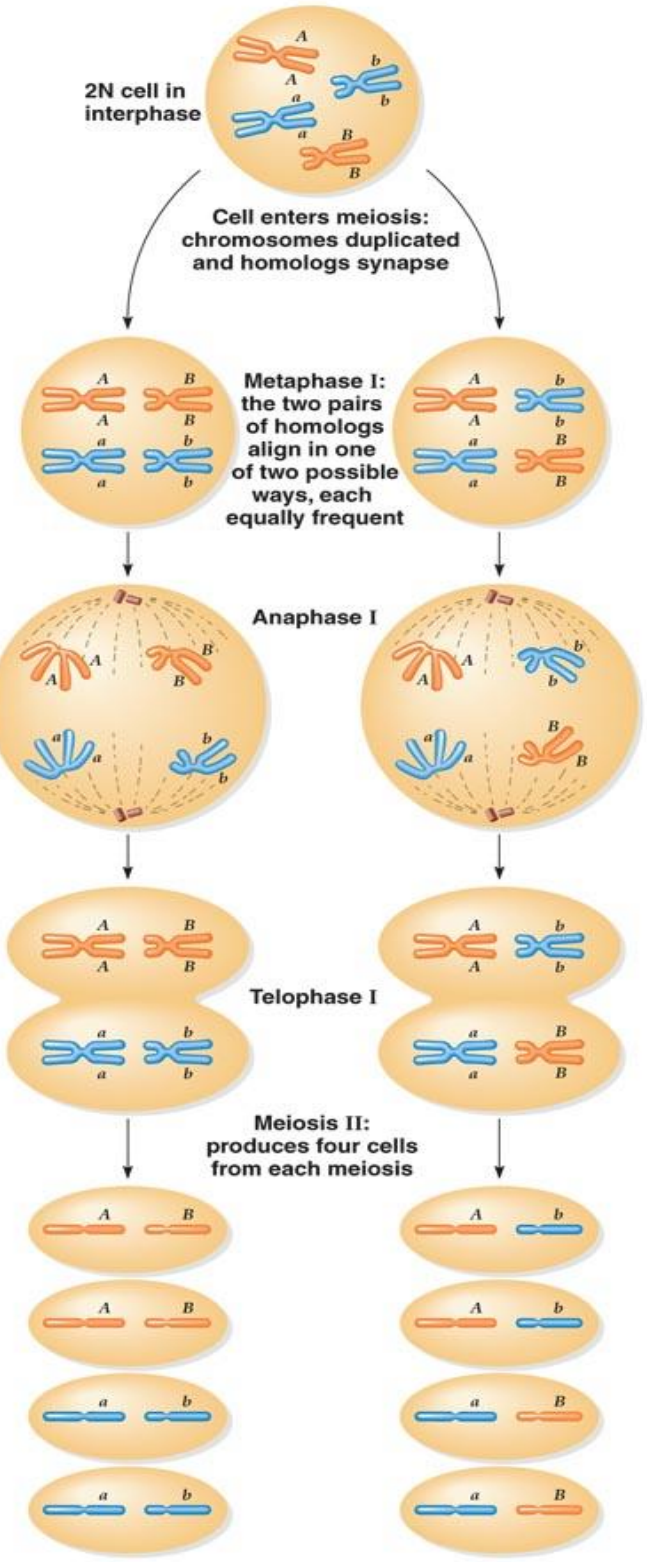
Morgan's Studies

- Studying fruit flies, Morgan found that the white eyed phenotype was somehow connected to sex as all the white eyed flies were male



Law of Independent Assortment

- States the traits assort independently in gametes
- Molecular mechanism shown by chromosome assortment possibilities in meiosis



Sex Determination Systems

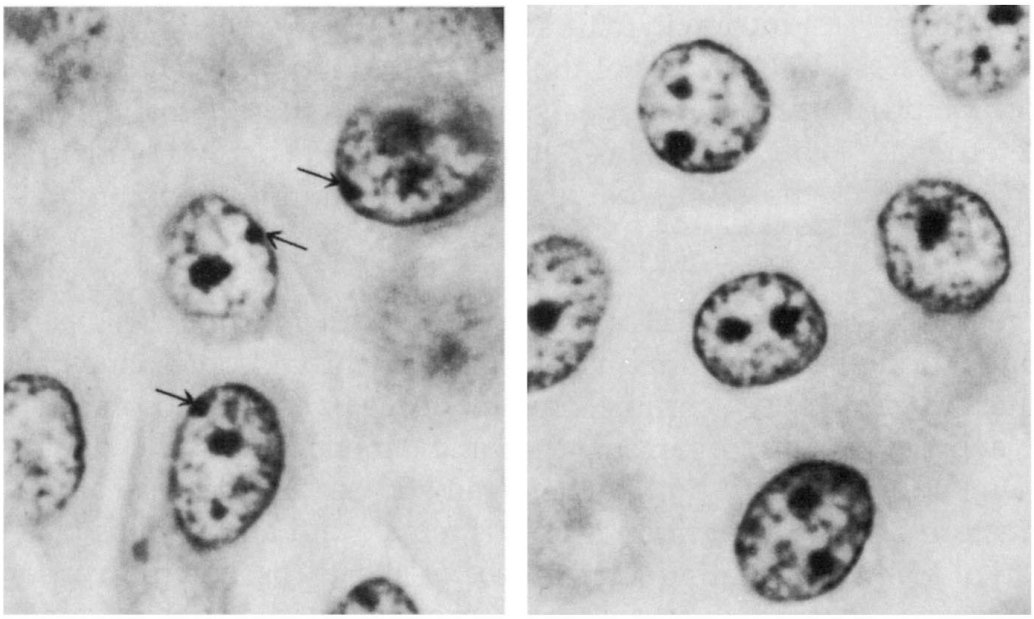
- Environmental sex determination: any sex-determination system in which sex is established by a nongenetic cue (e.g. nutrient availability, temperature in crocodiles and turtles) experienced within a discrete period after conception
- Sex determination by one gene: in many dioecious plants sex is determined by 2 alleles at one locus; typically one sex is heterozygous and the other is homozygous

- **XX/XY sex determination**: In the system, females have two of the same kind of sex chromosome (XX), while males have two distinct sex chromosomes (XY). The XY sex chromosomes are different in shape and size from each other, unlike the autosomes, and are termed allosomes. In this case the male is heterogametic, meaning that they produce 2 different types of gametes, while the female is homogametic, meaning that all eggs carry the same sex chromosomes. Males are said to be hemizygous for the sex chromosomes
- **XX/X0 sex determination**: In this variant of the XY system, females have two copies of the sex chromosome (XX) but males have only one (X0). The 0 denotes the absence of a second sex chromosome. Generally in this method, the sex is determined by amount of genes expressed across the two chromosomes. This system is observed in a number of insects, including the grasshoppers and crickets
- **ZW/ZZ sex determination**: found in birds, some reptiles, and some insects and other organisms. The ZW sex-determination system is reversed compared to the XY system: females have two different kinds of chromosomes (ZW), and males have two of the same chromosomes (ZZ)
- **Haplodiploidy**: found in insects belonging to Hymenoptera, such as ants and bees. Unfertilized eggs develop into haploid individuals, which are the males. Diploid individuals are generally female but may be sterile males. Males cannot have sons or fathers
- **X chromosome-autosome balance**: the ratio between the numbers of X chromosomes and number of sets of autosomes is the primary determinant of sex, found in fruit flies. XX, XXX, and XXY are female (if they have the usual number of autosomes), while XY and X are male (though only XY are fertile)
- **SRY** is the gene on the Y chromosome which triggers the formation of testes, and hence produces testosterone, making it phenotypically male. If the SRY gene is transferred to the X chromosome during the production of sperm, an XX male can result.

Dosage Compensation

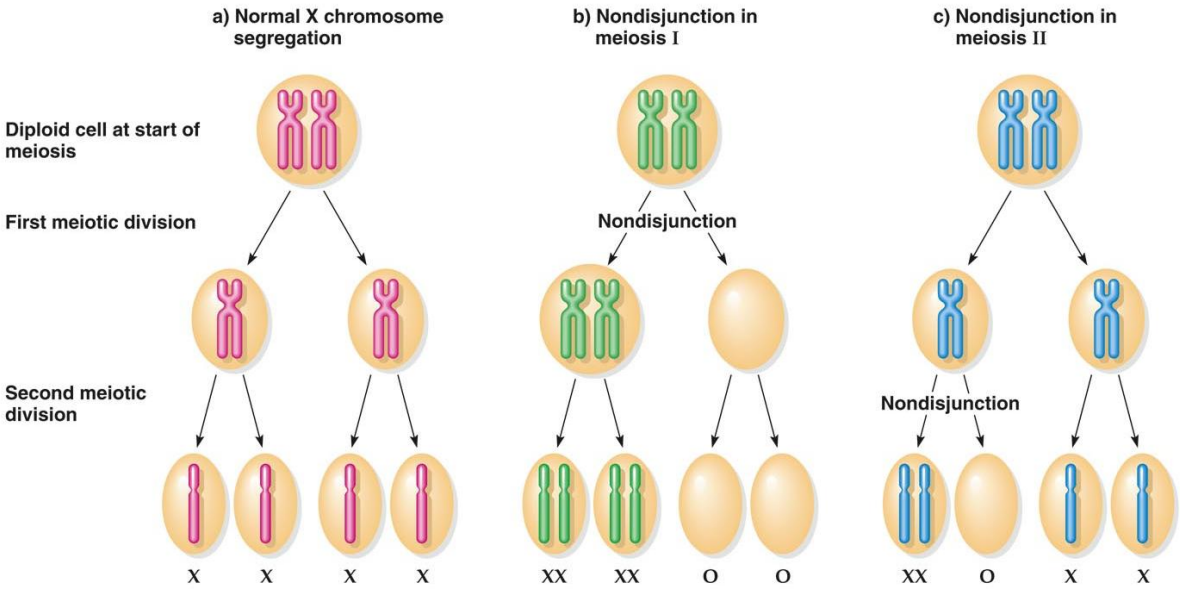
- X chromosome is much longer than Y
- Thus, in the XX-XY, the ZZ-ZW and the XX-XO systems for sex determination it would seem that the chromosomes are out of balance
- Given there are a lot of genes on the X chromosome, does that mean there will be twice as much gene product in the females than in males?
- Enzymes encoded by X-linked genes are normally present in equal amounts in male and female cells, so there must exist some mechanism to compensate for the lack of homologous chromosomes in males
- X chromosome inactivation: in 1949 Barr bodies were discovered, which are inactivated X chromosomes found as dark staining bodies seen in the nucleus of every mammalian species, but only in females
- X inactivation equalises the number of active copies of X-linked genes in males and females
- Lyon hypothesis: the X chromosome which is inactivated is random, occurring at the blastocyst stage
- Once an X chromosome is inactivated in a cell, all progeny from that cell will have that X chromosome inactivated, so a normal female is mosaic for X-linked genes
- This can be seen in heterozygous X-linked diseases, e.g. mutation affecting sweat glands, where patches of the skin have active sweat glands and patches where cells are normal. Also observed in the patches of orange and black skin on Calico cats

- In marsupials and some other animals, the X chromosome which is inactivated is always the one contributed by the father



Sex Chromosome Abnormalities

- Caused by nondisjunction during meiosis, which is the failure of chromosomes to move to opposite poles during anaphase
- Nondisjunction can occur during meiosis I or meiosis II



Genetic Disorders

- XO: Turner syndrome, short sexually immature sterile females
- XXX: Triplo-X, caused by nondisjunction in meiosis I or meiosis II of mother, found in 1/1000 - 1/2500 female births, have on average 11 IQ points deficit
- XXXX+: nondisjunction in meiosis of both, mother and father, which is very rare, less than 1/10,000 live births. Normal sexual development but severely retarded

- XYY: XYY syndrome, caused by nondisjunction of Y chromosome during meiosis, occurs in 1/1000 males. Tend to be tall with slightly impaired mental function, usually fertile
- XXY/XXXY: Klinefelter syndrome, occurs in 1/1000 males, who tend to be tall and have mild mental impairment. Abnormal sexual development, infertile, aneuploid

Sex-Linked Disorders

- **X-linked recessive disorders**: affected males always get abnormal allele from their mother, among sons of carrier mothers about half are affected and half are unaffected. Much more common in males than females, e.g. hemophilia, muscular dystophy, colour blindness
- **X-linked dominant disorders**: affected males always get the mutant allele from their mother, while heterozygous females pass it to approx half their sons and half daughters. Relatively few disorders are of this sort, enamel hypoplasia being one example
- **Y-linked disorders**: passed from father to all his sons. Several have been suggested but genetic evidence is poor e.g. hairy ears trait
- **Sex-Influenced traits**: sex of an individual influences the expression of a phenotype, e.g. pattern baldness. Although females show baldness, the phenotype is much more prevalent in males

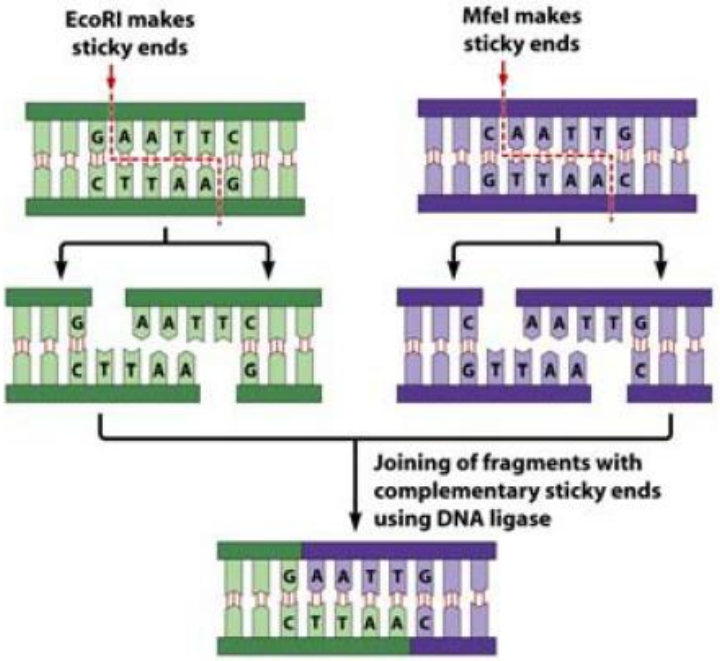
Topic 6: Gene Technologies

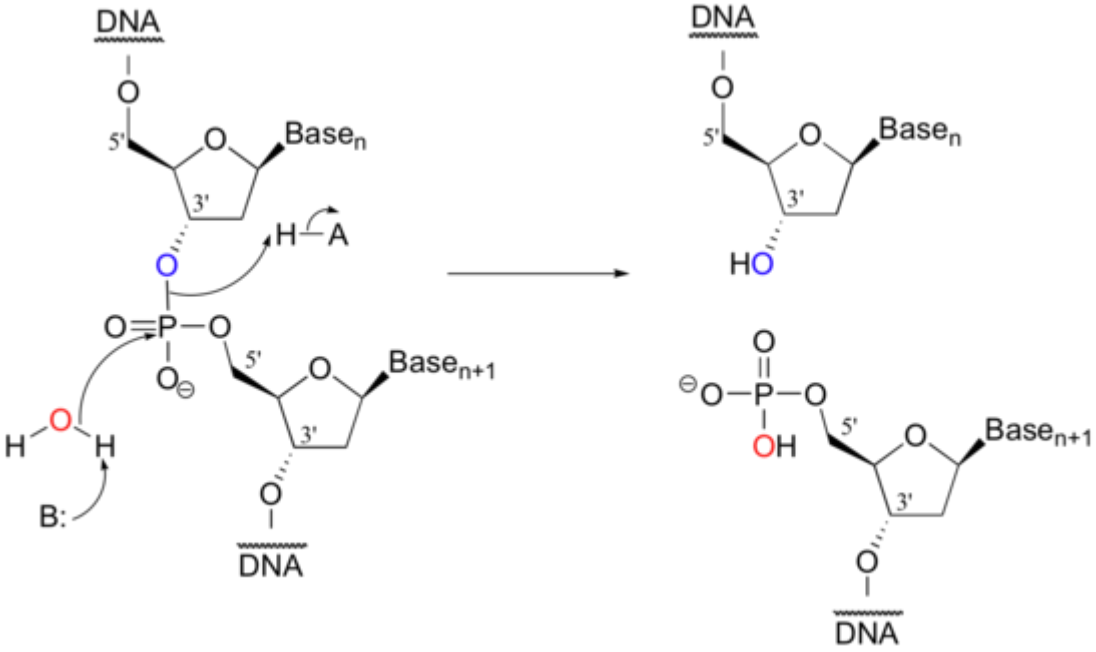
Restriction Endonucleases

- An enzyme that cuts DNA at or near specific recognition nucleotide sequences known as restriction sites
- *Eco*RI digestion produces "sticky" ends

- Whereas SmaI restriction enzyme cleavage produces "blunt" ends

- Bacteria use restriction endonucleases to degrade viral DNA, distinguishing it from their own DNA by its absence of methyl groups; foreign DNA is non-methylated and hence digested
- Type I enzymes cleave at sites remote from recognition site; require both ATP and S-adenosyl-L-methionine to function
- Type II enzymes cleave within or at short specific distances from recognition site
- Type III enzymes cleave at sites a short distance from recognition site; require ATP but do not hydrolyse it
- Type IV enzymes target modified DNA, e.g. methylated, hydroxymethylated and glucosyl-hydroxymethylated DNA
- These enzymes function by hydrolysing specific phosphodiester bonds in the backbone
- **Isoschizomers** are pairs of restriction enzymes that recognise the same recognition sequence, though some cut at different places
- It is also possible for two enzymes to recognise different sequences but have compatible end





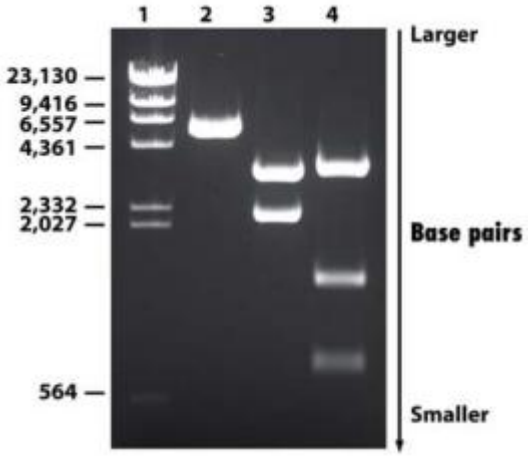
Polylinkers

- **pBluescript** (pBS) or **pBluescript II** is a commercially available phagemid containing several useful sequences for use in cloning with bacteriophag
- The multiple cloning site sequence is located within a LacZ controlled gene designed to provide a blue coloration when expressed in bacteria. If the **β-galactosidase** gene is disrupted by insertion of a DNA sequence, the bacteria exhibit a white coloration in Blue white screening, distinguishing successful recombination from those phagemids which were not altered

Gel Electrophoresis

- A technique whereby biomolecules are separated by applying an electric field to move the charged molecules through an agarose matrix, and the biomolecules are separated by size in the agarose gel matrix

- The separated DNA may be viewed with stain, most commonly under UV light, and the DNA fragments can be extracted from the gel with relative ease

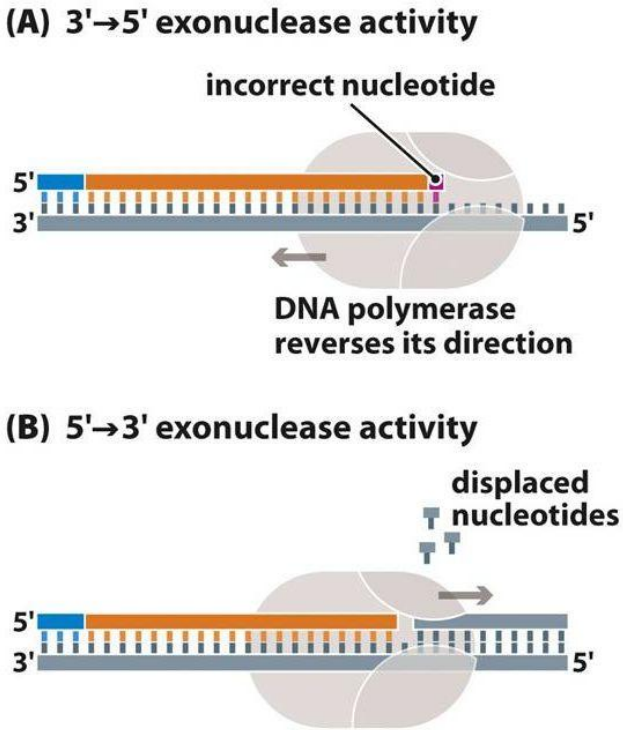


Restriction Maps

- A **restriction map** is a map of known restriction sites within a sequence of DNA
- The experimental procedure first requires an aliquot of purified plasmid DNA (see appendix) for each digest to be run. Digestion is then performed with each enzyme(s) chosen. The resulting samples are subsequently run on an electrophoresis gel, typically on agarose gel
- Comparing the sizes of fragments produced when different enzymes are used, a map of the location of the restriction sites can be made

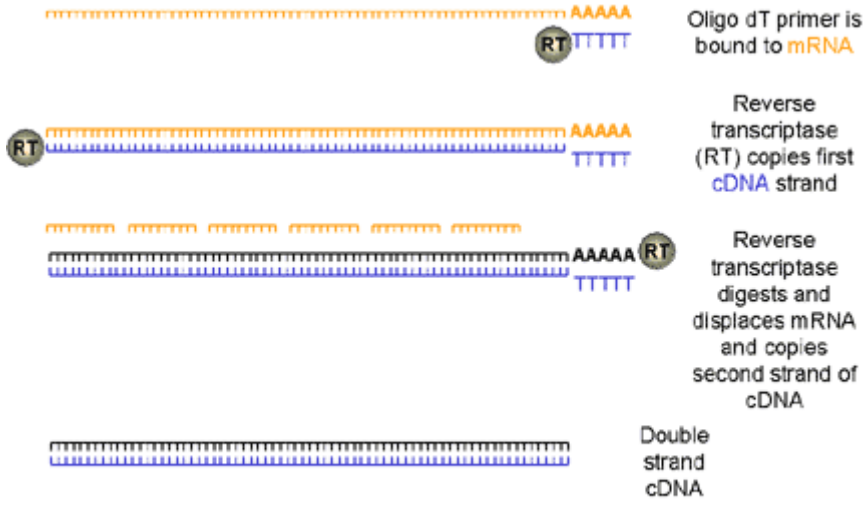
DNA Polymerase Exonuclease

- Exonucleases are enzymes that work by cleaving nucleotides one at a time from the end (exo) of a polynucleotide chain



Reverse Transcriptase

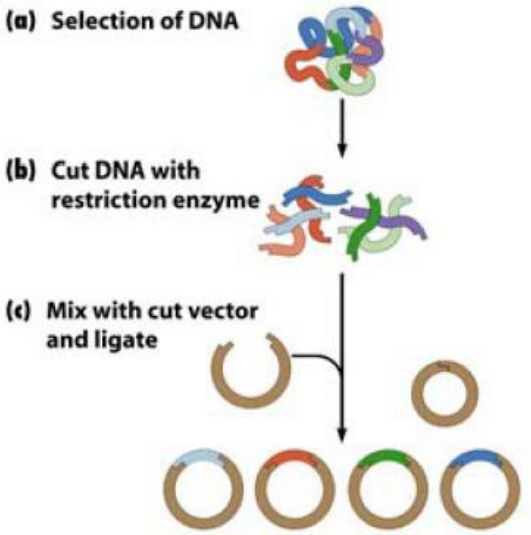
- An enzyme used to generate complementary DNA (cDNA) from an RNA template, a process termed *reverse transcription*. It is mainly associated with retroviruses

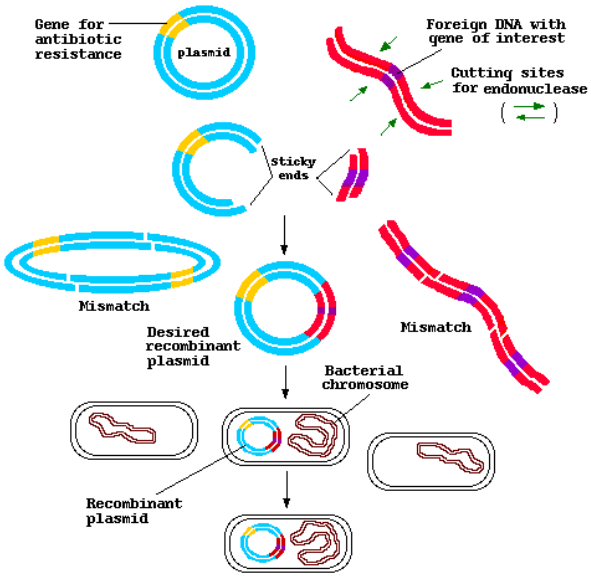


Conversion of mRNA to cDNA by Reverse Transcription

Plasmid Vectors

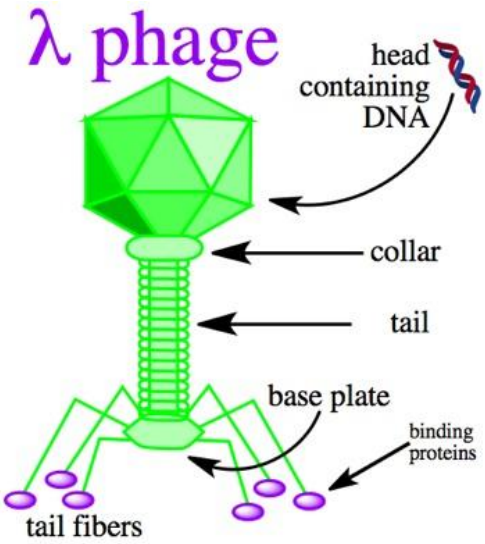
- DNA to be cloned goes into a vector, which acts as a vehicle to insert the DNA into the host cell, usually a bacterium
- Plasmids have their independent own origin of replication and so can multiply independent of the bacterial chromosome
- Good plasmid vectors will be small, present in multiple copies in the cell, include the genes of interest, unique restriction enzyme sites, and a selectable marker, such as antibiotic resistance or auxotrophy for some nutrient
- A **multiple cloning site** (**MCS**), also called a **polylinker**, is a short segment of DNA which contains many (up to 20) restriction sites - a standard feature of engineered plasmids





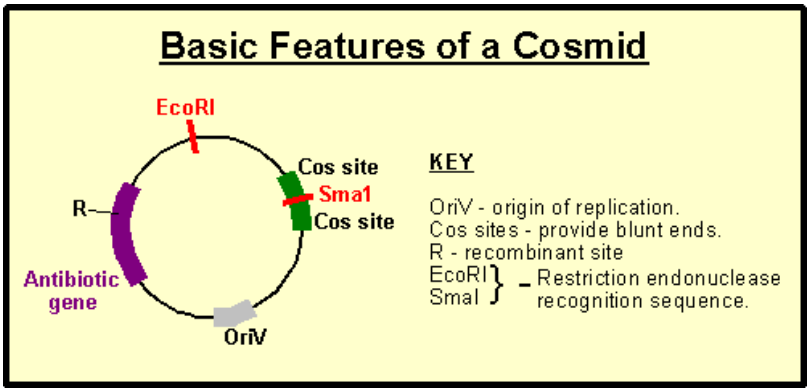
Phage Vectors

- The bacteriophages used for cloning are the phage λ and M13 phage
- There is an upper limit on the amount of DNA that can be packed into a phage (a maximum of 53 kb), therefore to allow foreign DNA to be inserted into phage DNA, phage cloning vectors need to have some non-essential genes deleted
- Linear molecule with single stranded complementary cohesive ends; genome circularises and transcribes as a circle
- A major advantage of lambda is that DNA can be packaged into phages in vitro, and transferred with an efficiency approaching 10%, compared to bacterial transformation less than 1/1000 plasmids become transformed



Cosmid Vectors

- A cosmid is a type of hybrid plasmid that contains a Lambda phage *cos sequence*
- They can replicate as plasmids if they have a suitable origin of replication
- Unlike plasmids, they can also be packaged in phage capsids, which allows the foreign genes to be transferred into or between cells by transduction
- This is made possible by the *cohesive ends*, also known as *cos* sites. In this way, they are similar to using the lambda phage as a vector, but only that *all* the lambda genes have been deleted with the exception of the *cos* sequence
- Cosmid phages are infectious, adsorbing to bacterial cells and injecting their DNA, which circularises due to the cos site. The molecule then replicates as a plasmid



Artificial Chromosomes

- Yeast artificial chromosomes are genetically engineered linear chromosomes derived from the DNA of the yeast, *Saccharomyces cerevisiae*, which is then ligated into a bacterial plasmid
- The minimal requirements of a linear eukaryote chromosome are an origin of replication (ars), telomeres, a centromere
- To make a TAC, combine ars from yeast, a centromere from yeast, and telomeres from Tetrahmena, generated a linear molecule that behaved in yeast like a chromosome
- Bacterial artificial chromosomes (BACs) are circular DNA molecules, usually about 7kb in length, that are capable of holding inserts up to 300kb in size. BAC vectors contain a replicon derived from E. coli F factor, which ensures they are maintained at one copy per cell
- Includes λ cosN and P1 loxP sites, as well as HindIII and BamH1 sites and T7 and SP6 promoter sequences

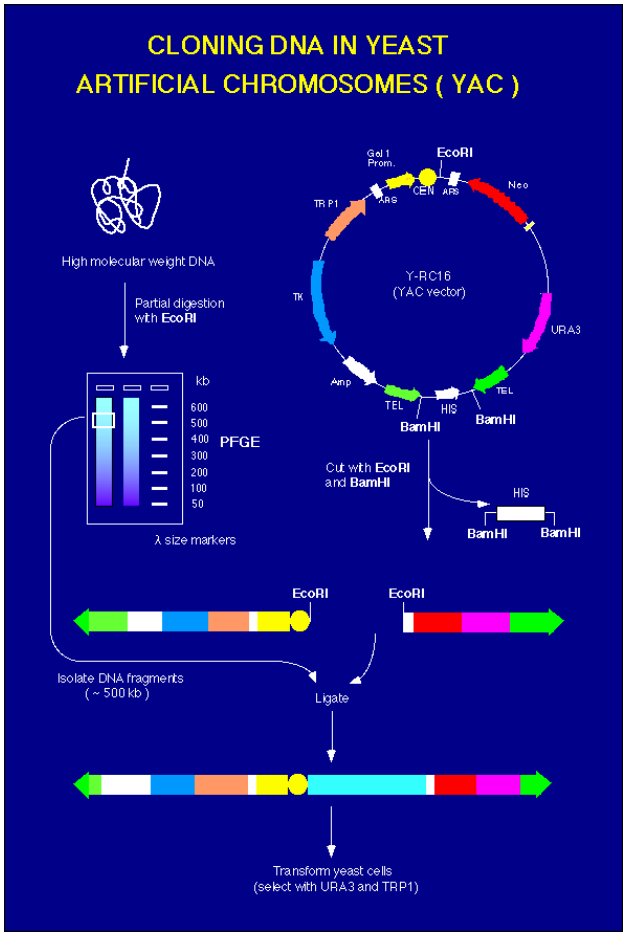
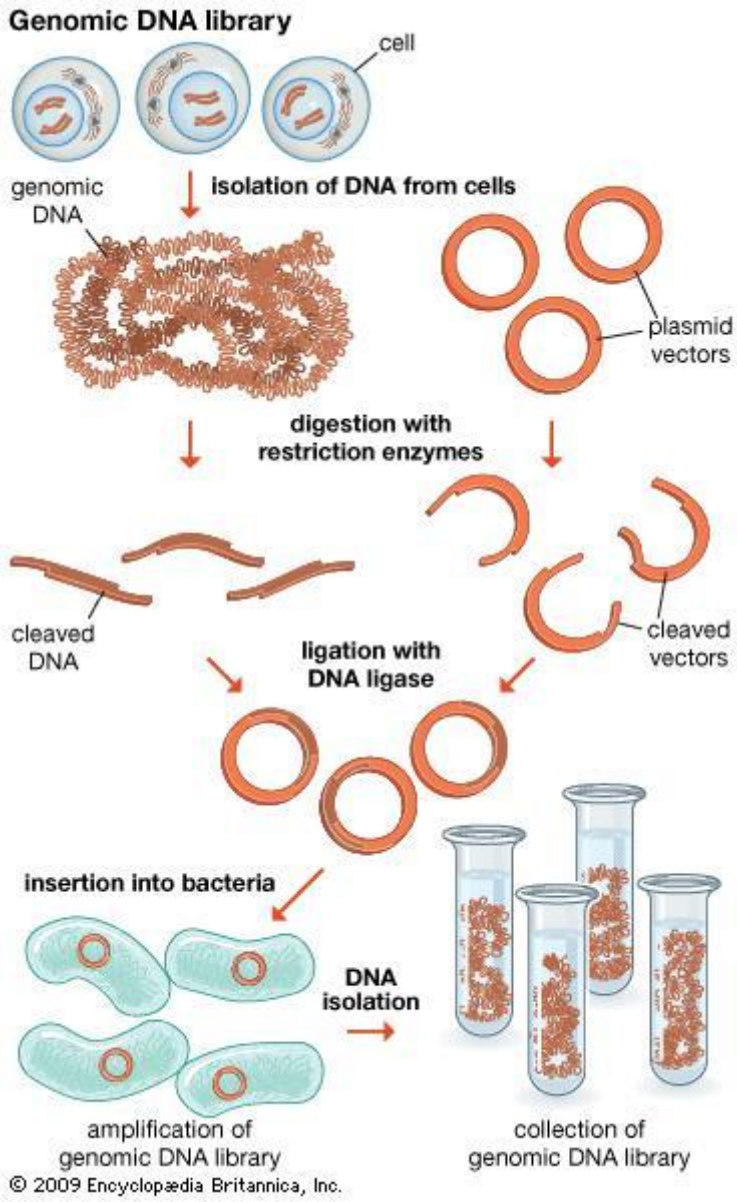


TABLE 4-2. Cloning capacity of commonly used vectors

Vector	Insert size range (kb)
Plasmid	<10
Phage	<23
Cosmid	30–46
P1 artificial chromosome (PAC)	130–150
Bacterial artificial chromosome (BAC)	<300
Yeast artificial chromosome (YAC)	200–2000

Genomic Libraries

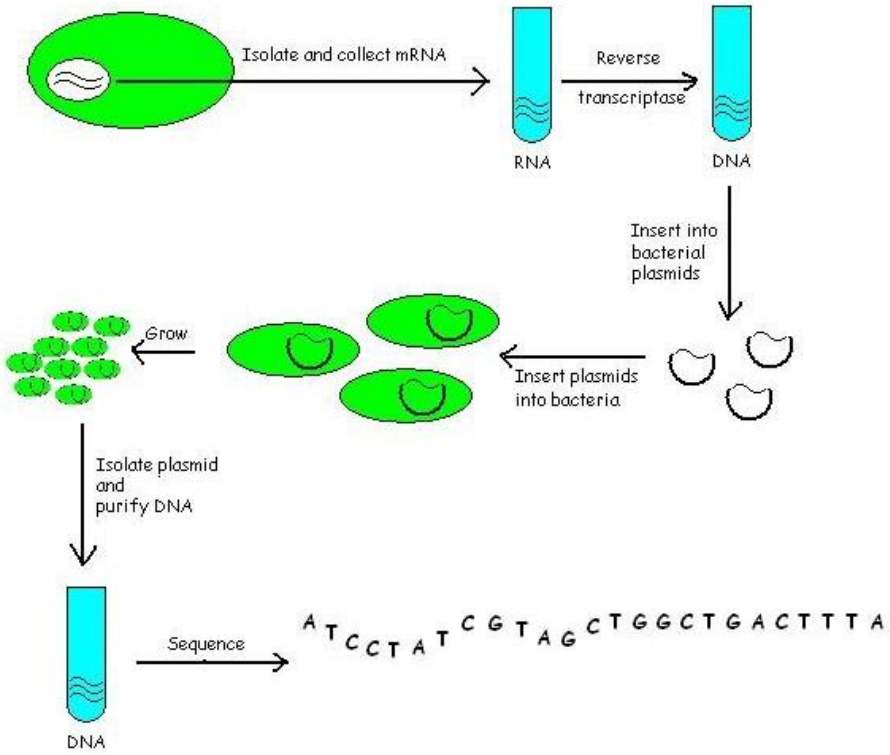
- A genomic library is a collection of the total genomic DNA from a single organism. The DNA is stored in a population of identical vectors, each containing a different insert of DNA
- In order to construct a genomic library, the organism's DNA is extracted from cells and then digested with a restriction enzyme to cut the DNA into fragments of a specific size
- The fragments are then inserted into the vector using the enzyme, DNA ligase. Next, the vector DNA can be taken up by a host organism- commonly a population of Escherichia coli or yeast- with each cell containing only one vector molecule

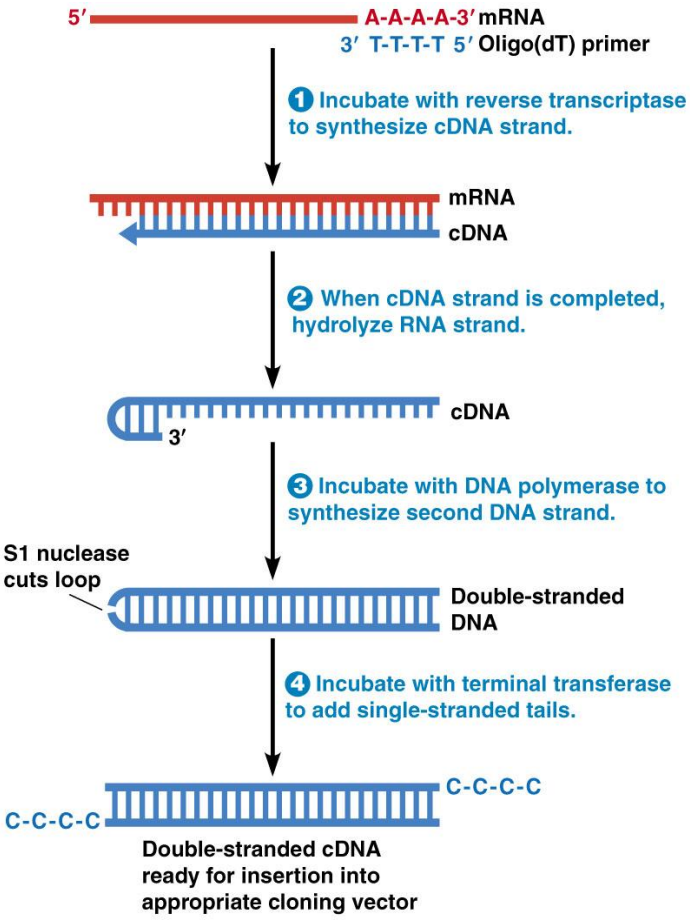


cDNA Libraries

- a combination of cloned cDNA (complementary DNA) fragments inserted into a collection of host cells, which together constitute some portion of the transcriptome of the organism
- cDNA is produced from fully transcribed mRNA found in the nucleus and therefore contains only the expressed genes of an organism

- While information in cDNA libraries is a powerful and useful tool since gene products are easily identified, the libraries lack information about enhancers, introns, and other regulatory elements found in a genomic DNA library





Genomic vs cDNA Library Construction

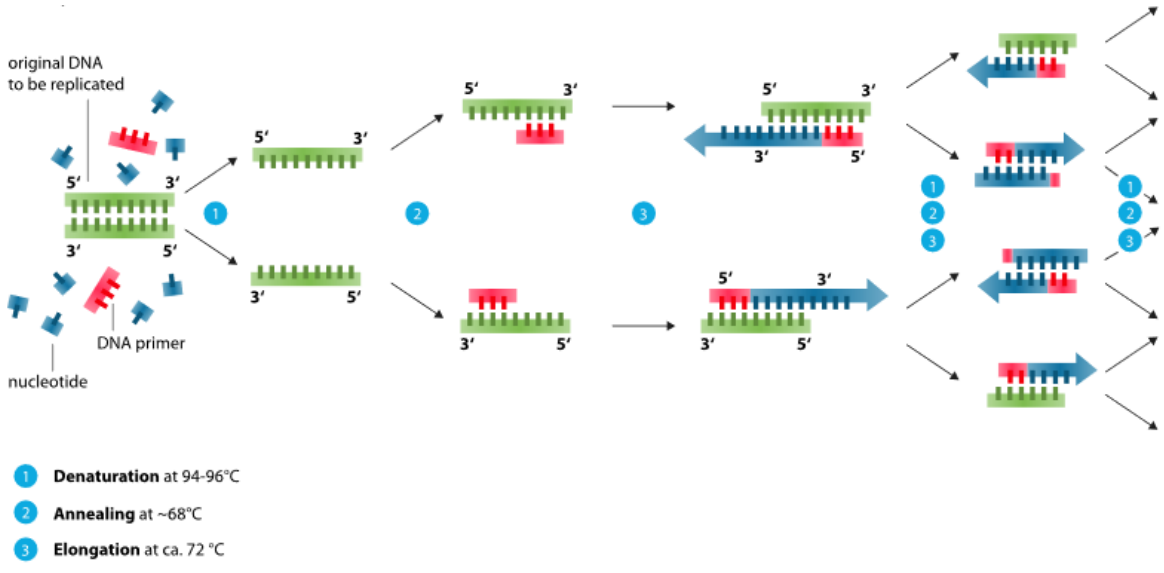
- Advantages of genomic: includes entire genome incl. non-coding regions and genes which are not expressed (pseudogenes, retroelements, introns), contains upstream elements which are necessary for tissue specific and regulated gene expression
- Disadvantages of genomic: includes all the 'junk' DNA, means a higher number of recombinant clones, no expression of genes possible in bacteria or yeast since splicing signal not recognised in those hosts.
- cDNA cloning: library contains only clones from genes which are transcribed, easy screening possible in particular when mRNA is abundant

Isolating the Desired Gene

- In order to isolate clones that contain regions of interest from a library, the library must first be screened. One method of screening is hybridization
- Each transformed host cell of a library will contain only one vector with one insert of DNA. The whole library can be plated onto a filter over media. The filter and colonies are prepared for hybridization and then labeled with a probe
- The target DNA- insert of interest- can be identified by detection such as autoradiography because of the hybridization with the probe as seen below

Polymerase Chain Reaction

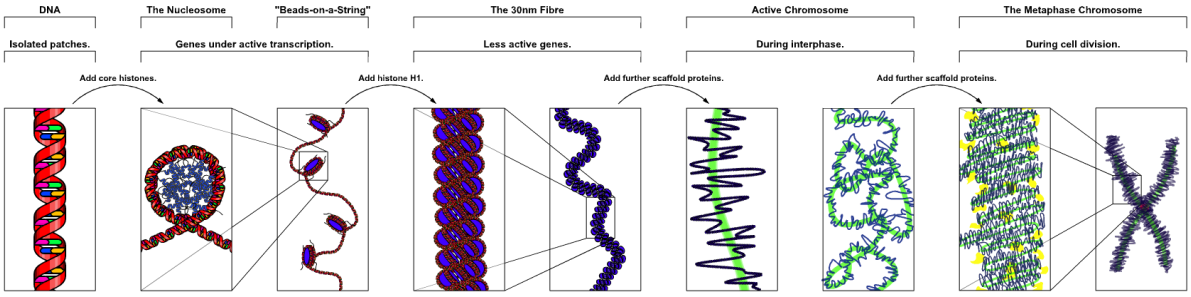
- The **polymerase chain reaction** (**PCR**) is a technology in molecular biology used to amplify a single copy or a few copies of a piece of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence.
- *Taq* polymerase is typically used as it is thermostable, as well as DNA ligase to join fragments
- The method relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA
- Primers (short DNA fragments) containing sequences complementary to the target region along with a DNA polymerase, after which the method is named, are key components to enable selective and repeated amplification
- As PCR progresses, the DNA generated is itself used as a template for replication, setting in motion a chain reaction in which the DNA template is exponentially amplified

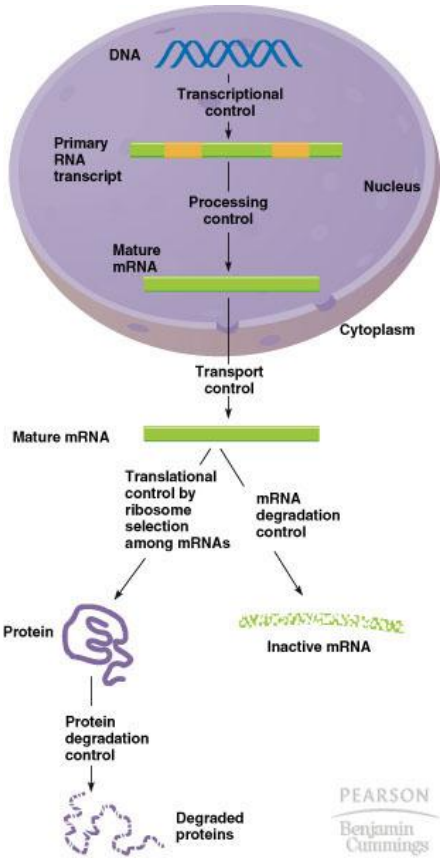


Topic 7: Eukaryotic Gene Regulation

The Eukaryote Genome

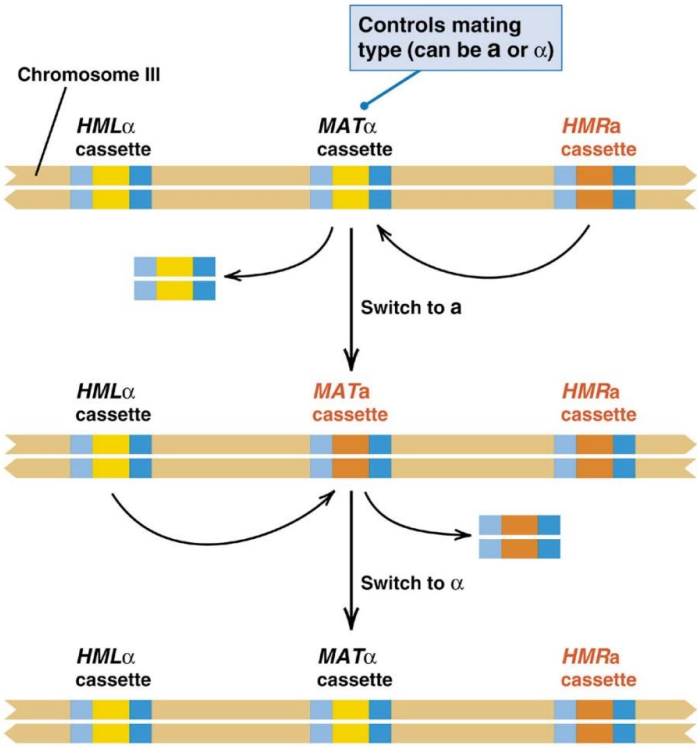
- Genome of eukaryotes tend to be 3-4 orders of magnitude larger than prokaryotes
- In most eukaryotes the coding regions are only a small proportion of the genome - bulk of the genome is repeat DNA, and at any given time about 1% of the genome is expressed
- In the eukaryote the genome is compartmentalised in the nucleus, while protein synthesis occurs in the cytoplasm, so transcription and translation are separated
- Most nuclear RNA never appears in the cytoplasm as functional mRNA
- The eukaryotic genome is tightly coiled and wrapped about histones and into fibre structures; only relatively unwound regions are actively transcribed
- Unlike prokaryotes, eukaryote genes tend to produce monocistronic mRNA
- mRNA stability also tends to be greater in eukaryotes
- In general no proteolytic enzymes in prokaryotes, they just dilute proteins in successive cell divisions. Eukaryotes cannot dilute proteins, but destroy them when not needed

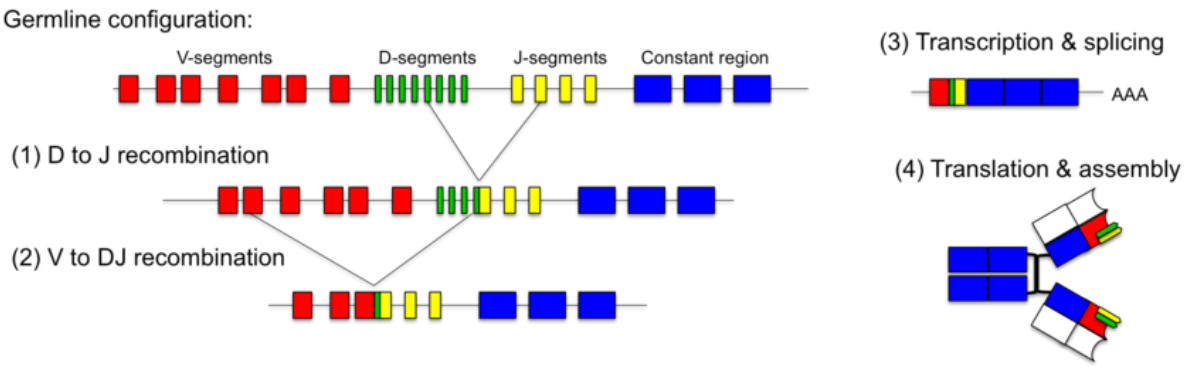




Genomic Control of Gene Regulation

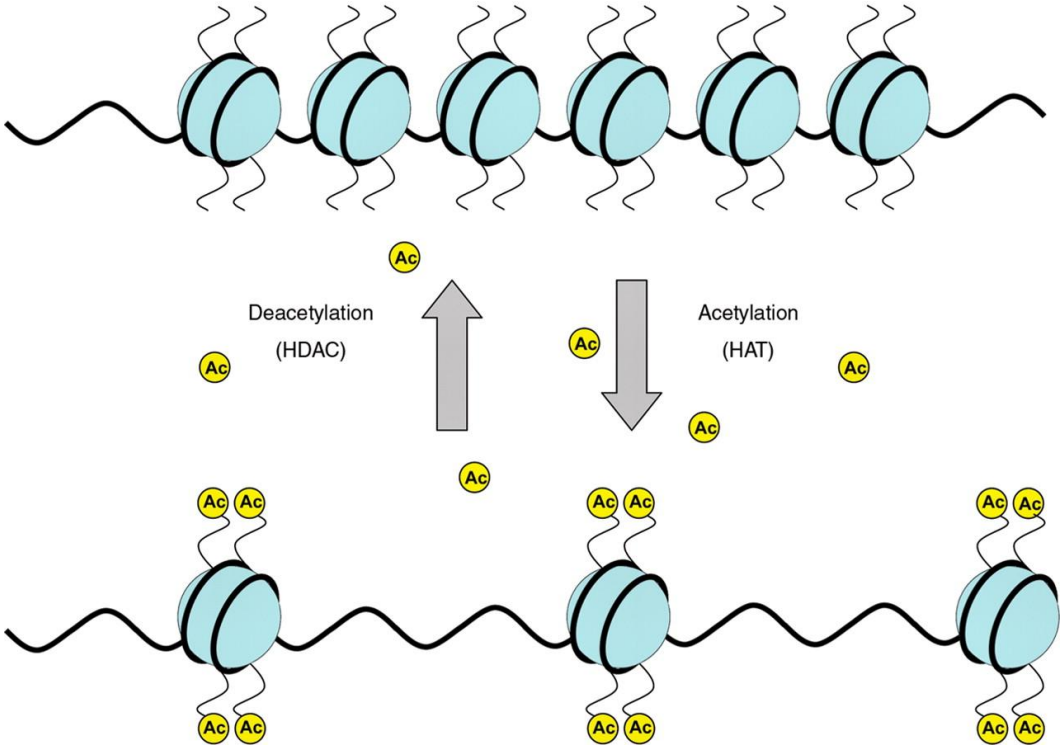
- A relatively rare type of control, in which there is selective loss or amplification of a specific segment of the genome
- Gene amplification: ribosomal RNA genes in *Xenopus laevis* during oogenesis amplified 4000 fold, such that the mature oocyte contains 2 x 10^6 copies, found as extra chromosomal circles
- Genome rearrangement: found in S. cervisiae yeast which has two mating types alpha and a. all haploids carry both alleles, a and α The mating type depends on which allele is at a specific site in genome called the MAT locus. Cells switch mating type by moving the appropriate allele to this locus
- **VDJ recombination**, less commonly known as **somatic recombination or** as antigen receptor gene rearrangement, is the unique mechanism of genetic recombination that occurs only in developing lymphocytes during the early stages of T and B cell maturation
- It occurs in the primary lymphoid organs (bone marrow for B cells and thymus for T cells) and in a nearly random fashion rearranges variable (V), joining (J), and in some cases, diversity (D) gene segments in the genome itself

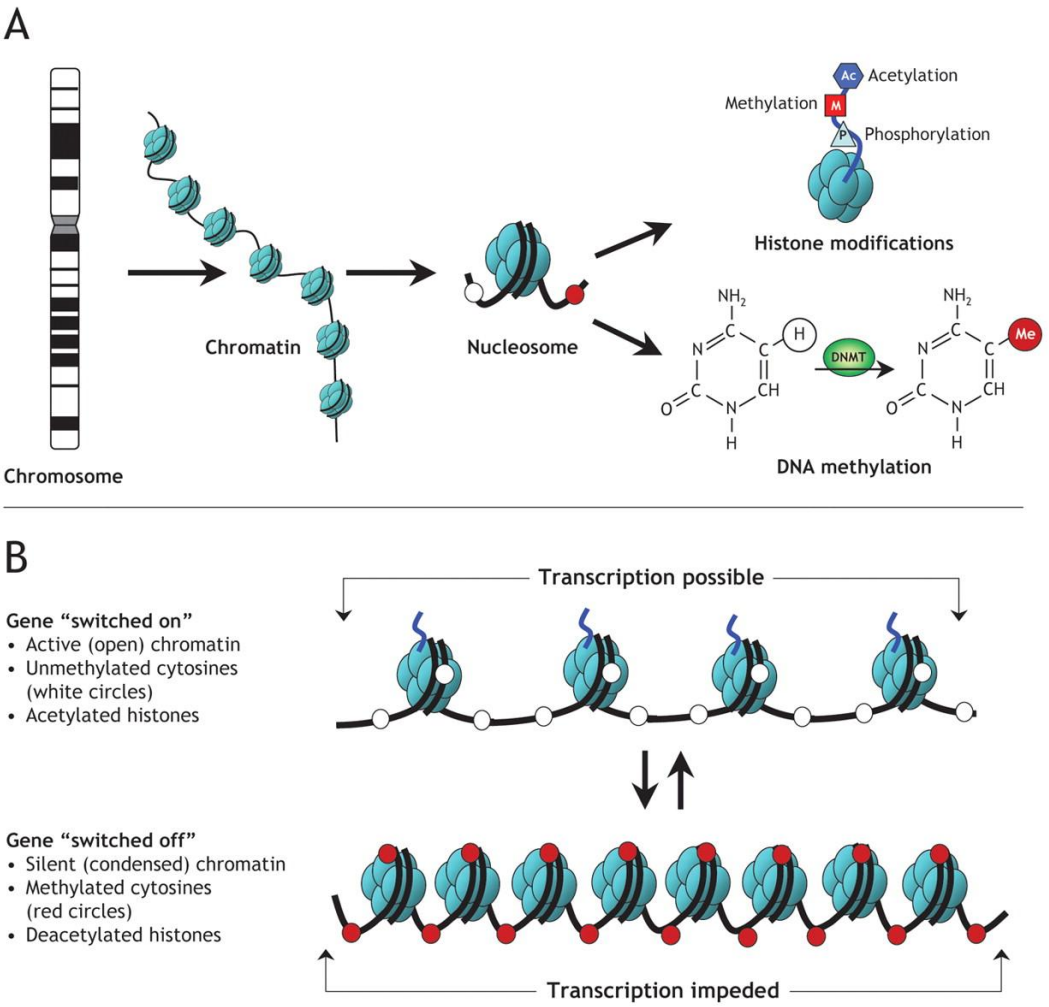




Chromosome Decondensation

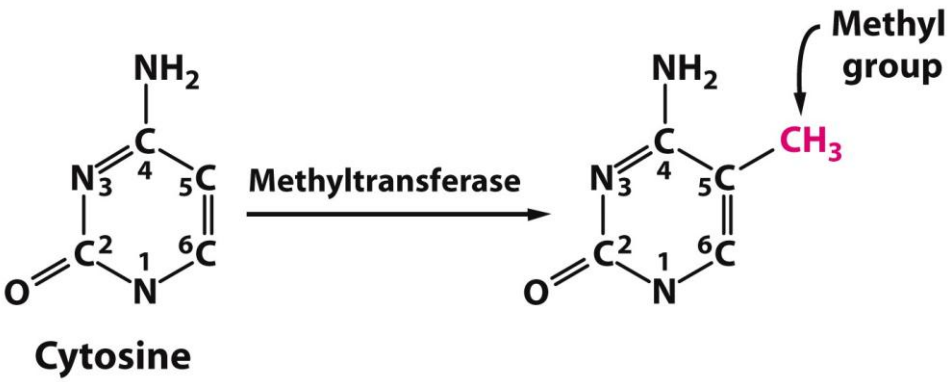
- For gene transcription to occur, the promoter region must decondense so that the DNA polymerase can bind to it
- Histone acetylation and deacetylation are the processes by which the lysine residues within the N-terminal tail protruding from the histone core of the nucleosome are acetylated and deacetylated as part of gene regulation
- Acetylation removes the positive charge on the histones, thereby decreasing the interaction of the N termini of histones with the negatively charged phosphate groups of DNA. As a consequence, the condensed chromatin is transformed into a more relaxed structure that is associated with greater levels of gene transcription
- **Chromosome puffs** are diffused uncoiled regions of the polytene chromosome that are sites of RNA transcription





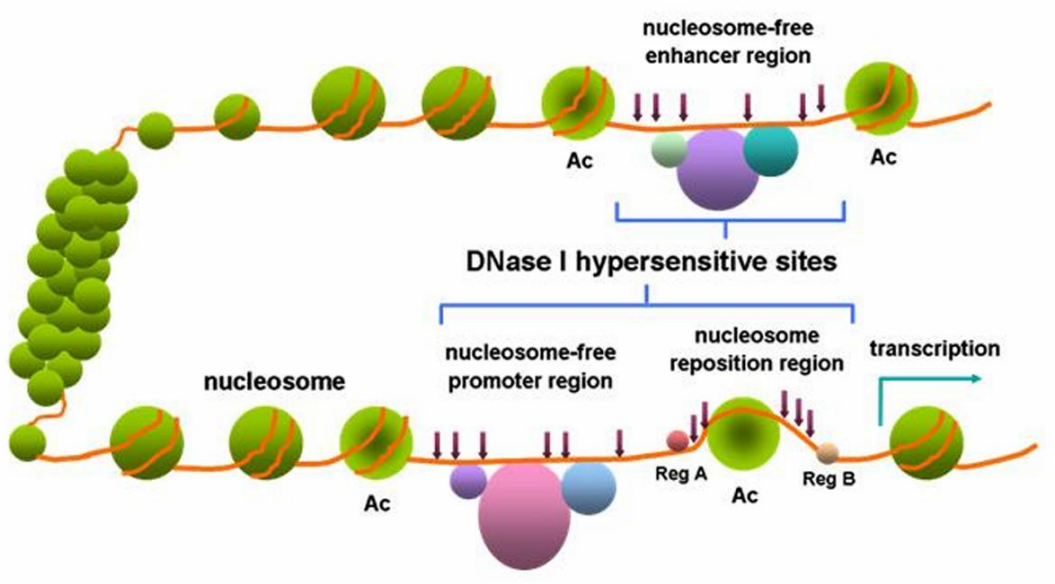
DNA Methylation

- DNA methylation refers to the selective addition of methyl groups to selected cytosine residues
- DNA of **inactive genes** tends to be more **heavily methylated** than DNA of active genes
- Example: sperm cells are transcriptionally completely inactive, as almost all CG doublets are fully methylated, while those not methylated are active immediately after fertilisation (eg histones). Also in the Barr bodies of females, the inactive X chromosome has significantly higher methylation than does active X



DNase Hypersensitive Regions

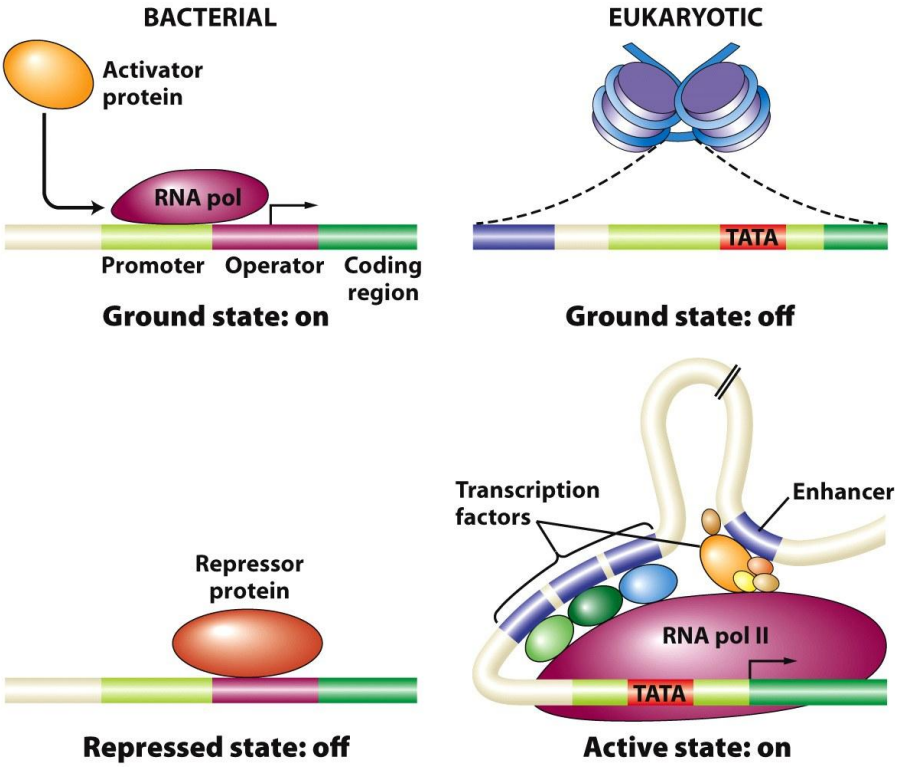
- **DNase I hypersensitive sites (DHSs)** are regions of chromatin that are sensitive to cleavage by the DNase I enzyme
- In these specific regions of the genome, chromatin has lost its condensed structure, exposing the DNA and making it accessible
- These accessible chromatin zones are functionally related to transcriptional activity, since this remodeled state is necessary for the binding of proteins such as transcription factors

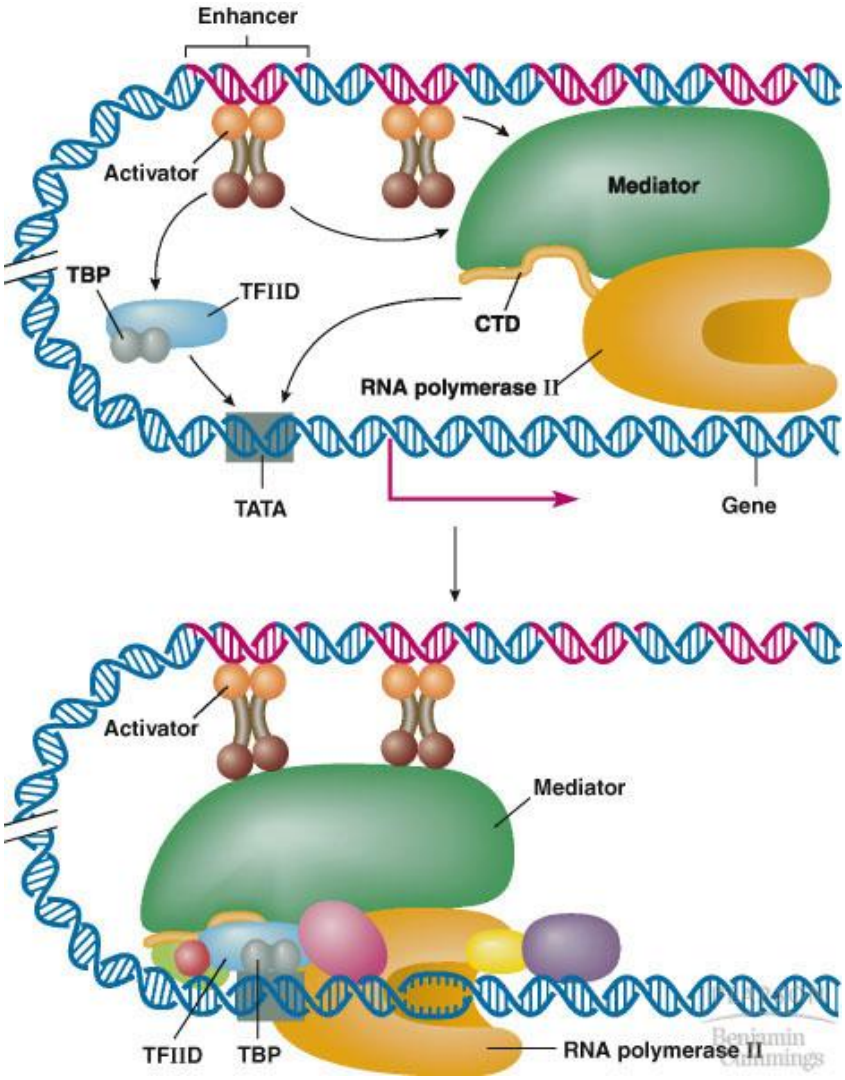


Transcriptional Regulation

- Control of transcriptional initiation is the usual way transcription is regulated in eukaryotes
- Unlike bacterial genes which are mostly 'on' by default, most eukaryotic genes are 'off' by default, and their synthesis needs to be triggered by special molecules
- **Transcription factors** are proteins that bind to specific DNA sequences in order to regulate the expression of a given gene
- **General transcription factors** (GTFs) are a set of factors in eukaryotes that are required for all transcription events, and their presence alone (along with an accessible promoter and DNA polymerase) results in a certain basal level of transcription
- A **promoter** is a region of DNA that initiates transcription of a particular gene. Promoters are located near the transcription start sites of genes, on the same strand and upstream on the DNA. Promoters can be about 100–1000 base pairs long. Promoter elements include the core promoter region (where the RNA polymerase binds), and sites for transcription factors
- **Activators** (also called trans-activators) are DNA-binding proteins (often dimers, but not always) which bind to DNA regions called **enhancers,** which can be located upstream or downstream of the gene, near or far (or even in introns). They activate transcription by interacting with the core transcription machinery
- A **coactivator** is a protein that increases gene expression by binding to an activator (or transcription factor) which contains a DNA binding domain. The coactivator is unable to bind DNA by itself

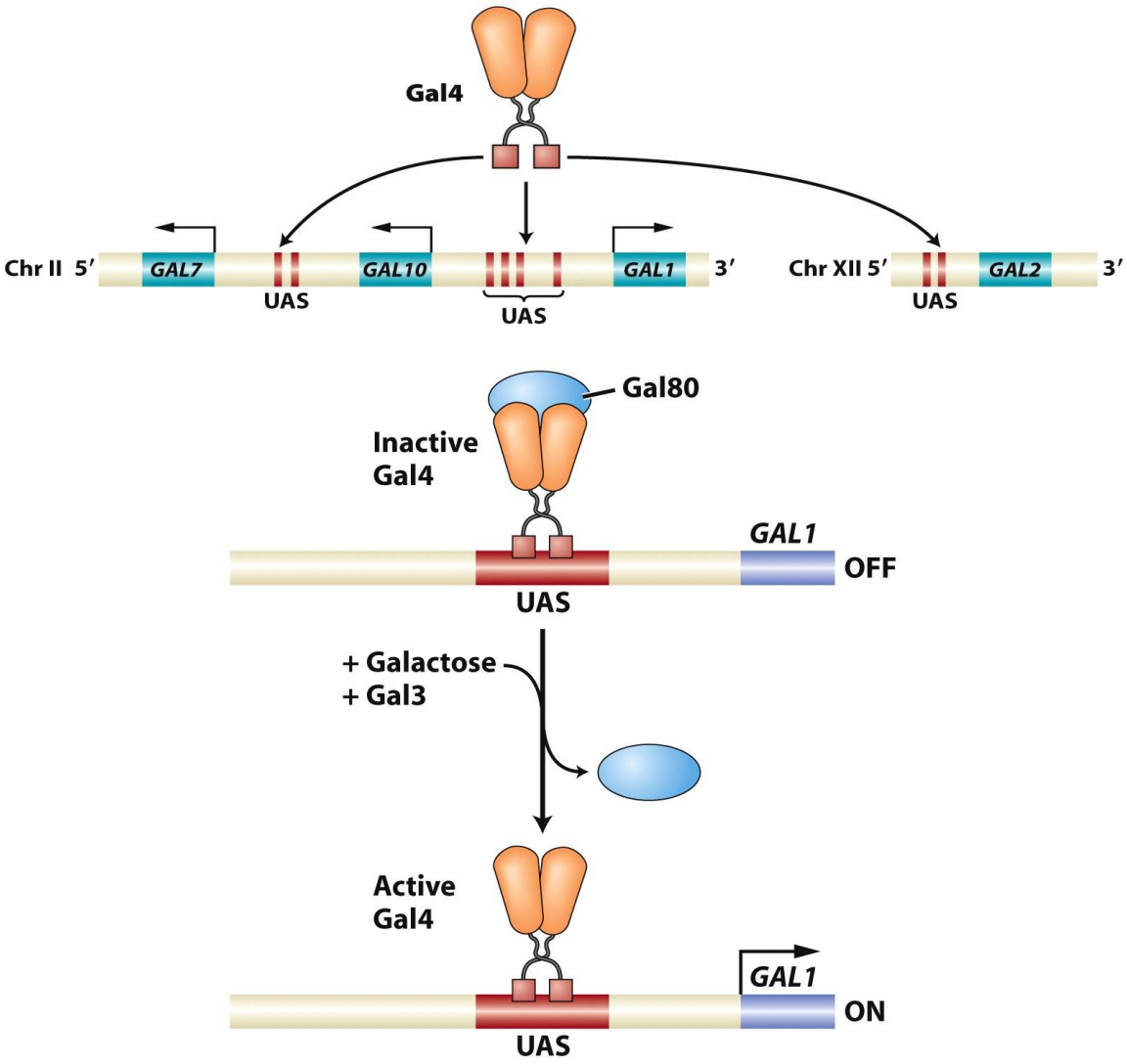
- An **inducer** is a molecule that regulates gene expression by binding to repressors or activators. Usually inducers will deactivate repressors or allow activators to work. The lactose operon is one example of an inducible system, with lactose being the inducer
- An **effector molecule** is usually a small molecule that selectively binds to a protein and regulates its biological activity. In this manner, effector molecules act as ligands that can increase or decrease enzyme activity, gene expression, or cell signalling.
- **Repressors** predominantly recruit co-repressor complexes leading to transcriptional repression by chromatin condensation of enhancer region
- Bacterial repressors generally work by binding to the promoter so that the DNA polymerase cannot bind there. This is not seen in eukaryotes. Instead, repressors can act by blocking access to the activator's binding site, or through chromatin remodelling
- **Silencers** are regions of DNA sequences that, when bound by particular transcription factors, can silence expression of the gene





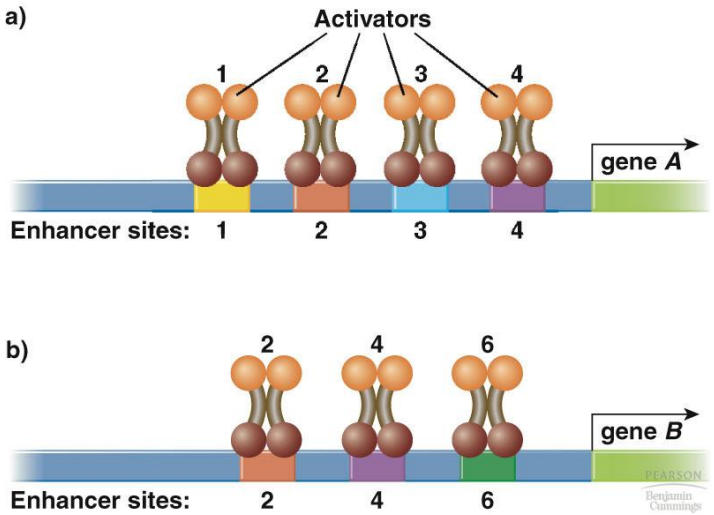
GAL4/UAS system

- The system has two parts: the GAL4 gene, encoding the yeast transcription activator protein GAL4, and the UAS (Upstream Activation Sequence), an enhancer to which GAL4 specifically binds to activate gene transcription
- Three genes (GAL1, GAL7 and GAL 10) encode for enzymes needed to utilise galactose as a carbon source. They are not transcribed in the absence of galactose
- The three genes are located near each other, but are **not an operon**
- The genes reside near regulatory sequences called UAS for **'upstream activator sequence'**, which is like an enhancer sequence but only function upstream of genes, so technically are not enhancers proper
- Galactose acts as an inducer molecule which binds to the repressore Gal80p, which in turn enables the activation of the Gal4p activator, and so the genes to be transcribed
- The genes will also not be active if glucose (a preferred carbon source) is present, thus this is an instance of **catabolite repression**
- Though not an operon, this system controls multiple related enzymes at once, and so is an example of coordinate induction



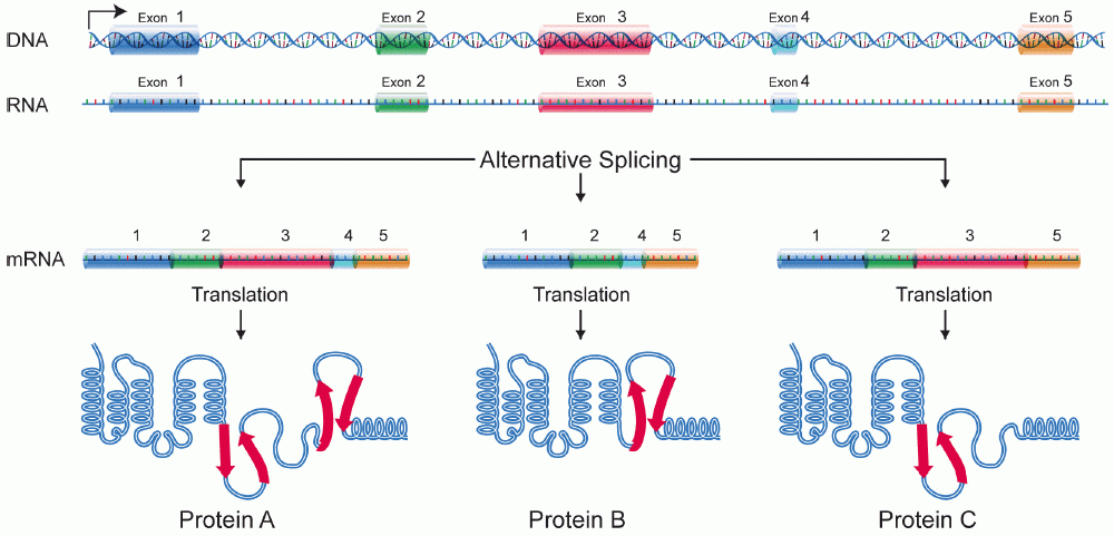
Combinatorial Gene Regulation

- The term combinatorial control refers to the way that groups of regulatory proteins work together to determine the expression of a single gene
- The gene expressed may not be determined by a single activator protein, but by the particular combination of activators and suppressors that are present

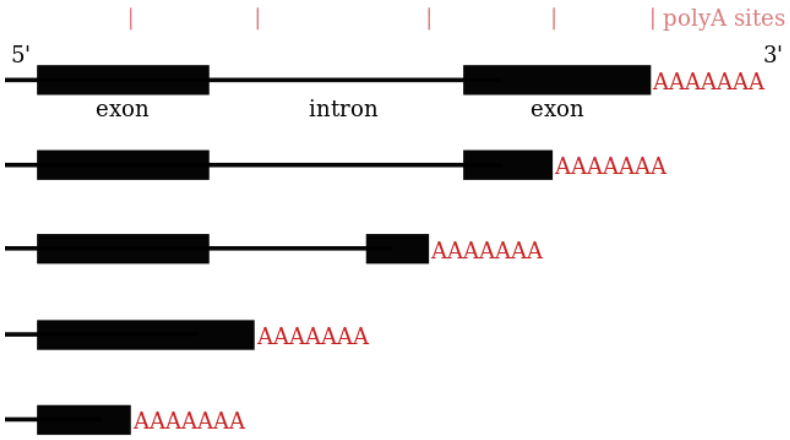


Post-Transcriptional Regulation

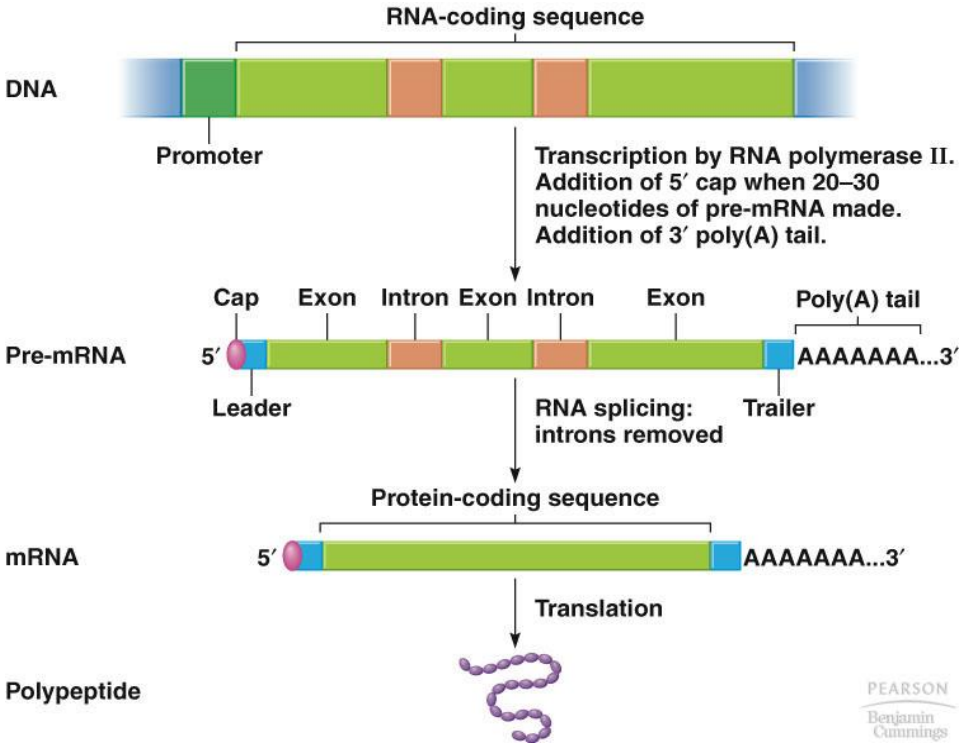
- After being produced, the stability and distribution of the different transcripts is regulated (post-transcriptional regulation) by means of RNA binding protein (RBP) that control the various steps and rates of the transcripts
- **Capping** changes the five prime end of the mRNA to a three prime end by 5'-5' linkage, which protects the mRNA from 5' exonuclease, which degrades foreign RNA. The cap also helps in ribosomal binding, and allows transport out of the nucleus
- **Splicing** removes the introns, noncoding regions that are transcribed into RNA, in order to make the mRNA able to create proteins. Cells do this by spliceosomes binding on either side of an intron, looping the intron into a circle and then cleaving it off. Alternative splicing can lead to alternate protein products



- **Addition of poly(A) tail** otherwise known as polyadenylation. That is, a stretch of RNA that is made solely of adenine bases is added to the 3' end, and acts as a buffer to the 3' exonuclease in order to increase the half life of mRNA. Many protein-coding genes have more than one polyadenylation site, so a gene can code for several mRNAs that differ in their 3' end. Since alternative polyadenylation changes the length of the 3' untranslated region, it can change which binding sites for microRNAs the 3' untranslated region contains. Longer poly(A) tails are also known to lead to more stable mRNA transcripts



- **mRNA Stability** can be manipulated in order to control its half-life. Stable mRNA can have a half life of up to a day or more which allows for the production of more protein product; unstable mRNA is used in regulation that must occur quickly. Various chemicals can be applied to do this; for example, prolactin prolongs the half-life of casein mRNA



Translational control

- Important for the rapid short term adjustments in protein synthesis
- Example: iron responsive elements are mRNA segments contained within the mRNA sequences that code for transferrin receptors and for ferritin, which is an iron storage molecule. Normally, IRE binding proteins bind to these segments, inhibiting the translation of the ferritin molecules. However, when sufficient iron is present in the cell, it binds to the IRE binding proteins, causing them to change shape and disassociate from the mRNA, thus freeing it up for expression
- Post translation control: modulation of protein function can be permanent (proteolysis, glycosylation) or reversible (phosphorylation)
- **Ubiquitin** is a small regulatory protein that has been found in almost all tissues of eukaryotic organisms. The addition of ubiquitin can affect proteins in many ways: It can signal for their degradation via the proteasome, alter their cellular location, affect their activity, and promote or prevent protein interactions
- The ***N*-end rule** is a rule related to ubiquitination. The rule states that the *N*-terminal amino acid of a protein determines its half-life (likelihood of being degraded). The rule applies to both eukaryotic and prokaryotic organisms, but with different strength

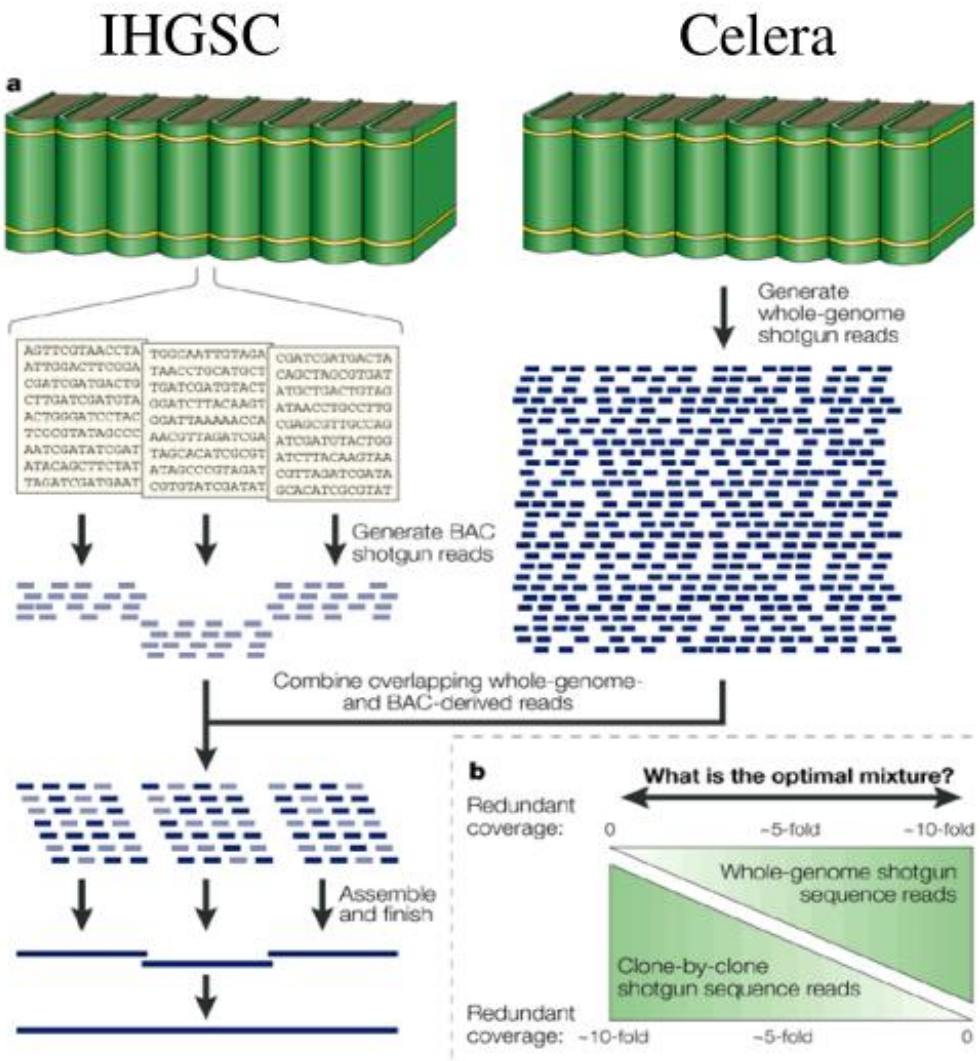
Topic 8: Genomics

What is Genomics?

- "The development and application of new mapping, sequencing, and computational procedures for the analysis of the entire genomes of organisms"
- Structural genomics: determine the full genome sequence and the structure of the proteins
- Functional genomics: describe the transcription, translation, and function of proteins and genes, and how they interact. Strong focus on data-driven analysis of genome
- Comparative genomics: detailed comparison of genomes of different organisms, including gene order, number, genomic sequence, regulatory sequences, and other structural landmarks

Methods of Genome Sequencing

- Genetic mapping:
- **Shotgun sequencing**, also known as **shotgun cloning**, is a method used for sequencing long DNA strands
- a high-molecular-weight DNA strand is sheared into random fragments, size-selected (usually 2, 10, 50, and 150 kb), and cloned into an appropriate vector. The clones are then sequenced from both ends using the chain termination method yielding two short sequences



SNP Typing

- A **Single Nucleotide Polymorphism** is a DNA sequence variation occurring commonly within a population (e.g. 1%) in which a single nucleotide in the genome (or other shared sequence) differs between members of a biological species
- SNP genotyping is the measurement of genetic variations of single nucleotide polymorphisms (SNPs) between members of a species. It is a form of genotyping, which is the measurement of more general genetic variation
- Restriction fragment length polymorphism, or RFLP, refers to a difference between samples of homologous DNA molecules from differing locations of restriction enzyme sites
- The basic technique for detecting RFLPs involves fragmenting a sample of DNA by a restriction enzyme, which can recognize and cut DNA wherever a specific short sequence occurs, in a process known as a restriction digest
- The resulting DNA fragments are then separated by length through a process known as agarose gel electrophoresis, and transferred to a membrane via the Southern blot procedure. Hybridization of the membrane to a labeled DNA probe then determines the length of the fragments which are complementary to the probe
- An RFLP occurs when the length of a detected fragment varies between individuals. Each fragment length is considered an allele, and can be used in genetic analysis.

Goals of the Human Genome Project

- Identify the 30,000 genes (approx) in human DNA
- Determine the 3.2 billion bp that make up our genome
- Improve tools for data analysis
- Transfer technologies to the private sector
- Address the ethical, legal and social issues from this work

Results of the Human Genome Project

- Prior to the sequencing we thought there would be about 100,000 genes in the genome, but now it is thought there are only about 25,000
- Less than 2% of the genome codes for protein
- We still don't know what a lot of the genes do, or what the non-coding regions are for, and how genes expression is regulated in detailed
- The IHGSC cost approx $3 billion dollars and took 13 years

Yeast Knockout Project

- Yeast deletion project is a project to create data for a near-complete collection of gene-deletion mutants of the yeast Saccharomyces cerevisiae
- Each strain carries a precise deletion of one of the genes in the genome, and has been functionally characterized. The fitness contribution of each gene was quantitatively assessed under six different growth conditions
- Of all the knocked out genes, 4,200 of 6,200 proved to be non-essential (non-lethal to knockout)
- Half show no significant phenotype change. Probably depends on how the analysis is done

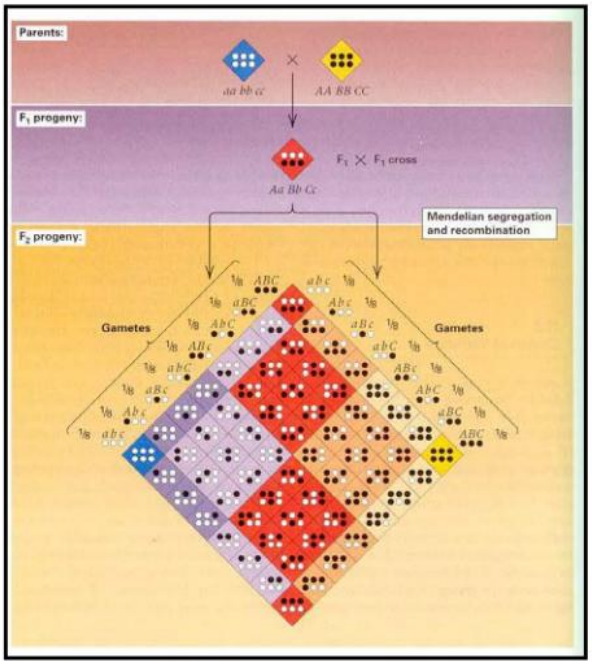
Topic 9: Quantitative Genetics

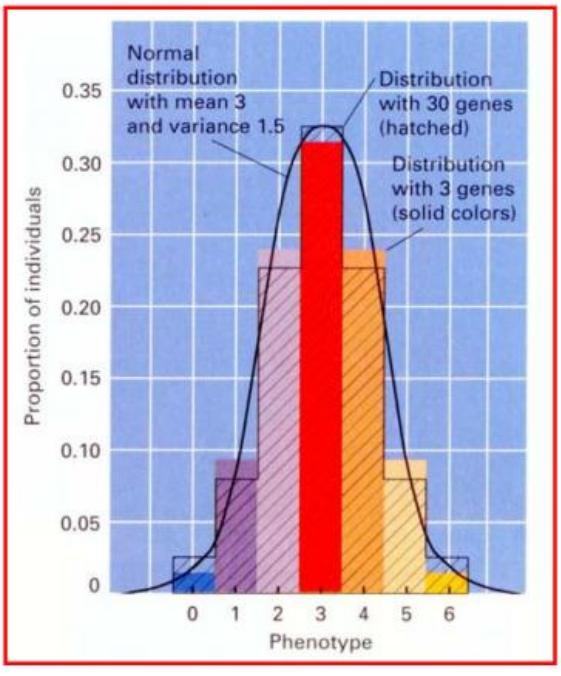
Quantitative Traits

- A quantitative trait is a measurable phenotype that depends on the cumulative actions of many genes and the environment. These traits can vary among individuals, over a range, to produce a continuous distribution of phenotypes. Examples include height, weight and blood pressure
- Originally studied by Francis Dalton 1889
- Three different types of quantitative traits:
    - Continuous (e.g. weight)
    - Meristic (sometimes called "metric"), variation established by counting, e.g. chaeta number in Drosophila, number of seeds per spike in wheat)
    - Threshold (only few phenotypic classes, but controlled by multiple genes e.g. adult-onset diabetes, schizophrenia)

Distributions of QT

- When a trait is controlled by many genes and its expression is affected by the environment, it is not possible to follow the segregation of each separate polygene e.g. the F2
- Instead, we look for similarities/differences among relatives and families using various statistics like mean, variance, and covariance
- For many quantitative traits, the distribution of phenotypes often approximate that of a normal distribution
- Consider the phenotypic ratios for a case where a trait is determined by a mere three genes





Causes of Variation in QT

- Variation observed in phenotype caused by genotype and environment
- The effects of the genotype and environment on the mean of the phenotype may be estimated if inbred [plant] or genetically identical populations available

- In many cases, esp. human genetics, genetically identical populations do not exist. However, it is still possible to gauge the effects of the environment and genotype on the variance of the trait
- The phenotypic variance for a particular trait may be expressed as:

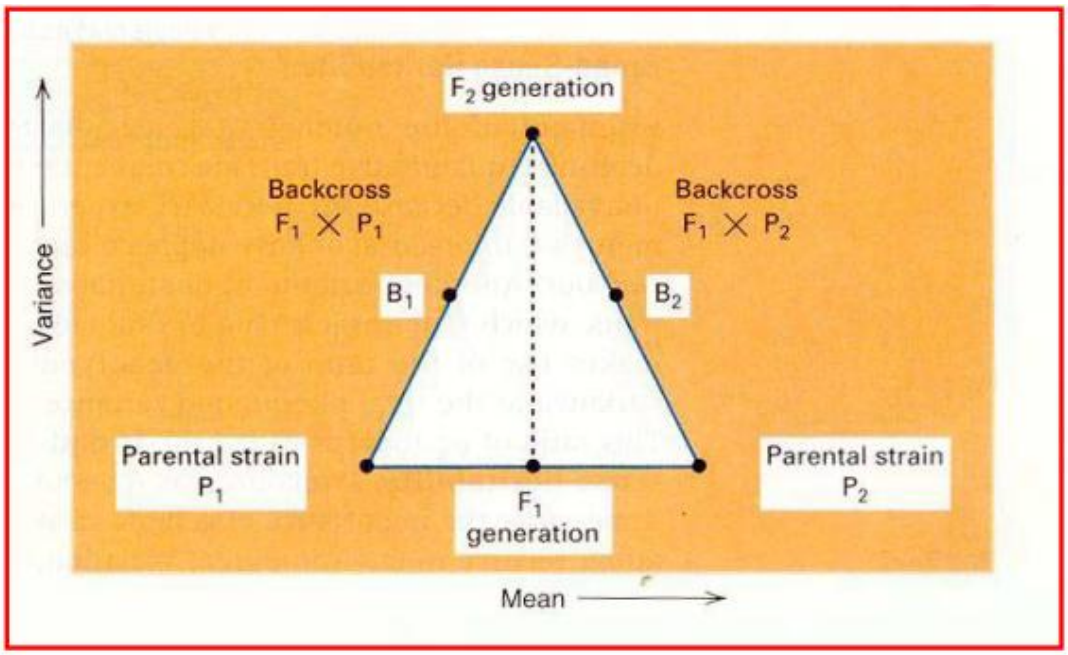
$$V_P = V_G + V_E + V_{GE}$$

- The genotypic variance (VG) is due to variation in phenotype due to differences in
- genotype between individuals
- The environmental variance (VE) is due to variation in phenotype due to the exposure of individuals to differing environmental conditions
- The genotype x environment variance (V GE) is due to the environmental effects on the phenotype differing according to genotype

Analysis of QT

- How to separate out the different sources of variation? Need a genetically uniform population where $V_G = 0$ and $V_{GE} = 0$ to measure $V_P = V_E$
- Example: Eye diameter in cave-dwelling fish. If F1 and F2 progeny from homozygous parents are available, and no G x E Interaction is present, then the genotypic and environmental variance may be estimated:

$$F_1: V_P = V_E = 0.057$$
$$F_2: V_P = V_G + V_E = 0.563$$
$$V_G = 0.563 - 0.057 = 0.506$$



Number of Genes Affecting a QT

- If: the alleles of each polygene are additive, each polygene contributes equally to the trait [acting additively], the polygenes are unlinked, and the original parental strains are homozygous for alternative alleles for each polygene
- Then the number of polygenes (n) contributing to the trait is given by:

$$n = \frac{D^2}{8V_G}$$

- The number of genes controlling a QT affects the extent to which artificial selection may be used to improve the qualities of a given population. The larger the number of genes, the greater the eventual improvement
- Most breeding populations too small to express the full range of phenotypes of a QT, which is given by 4^n

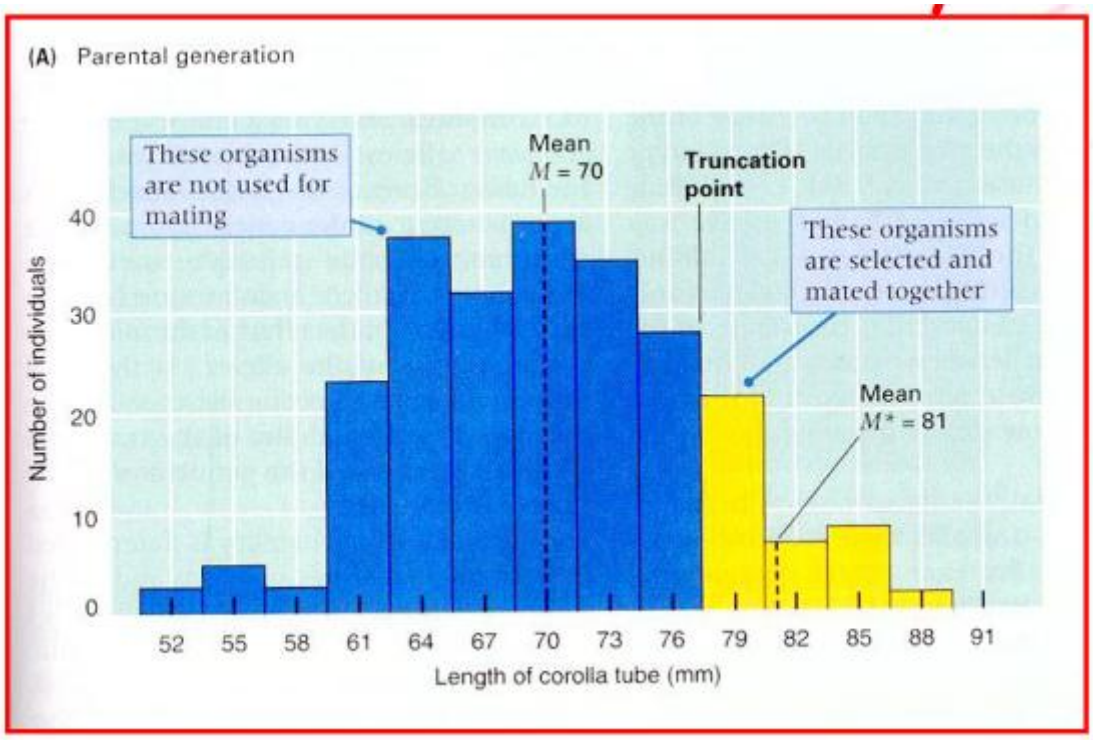
Heritability

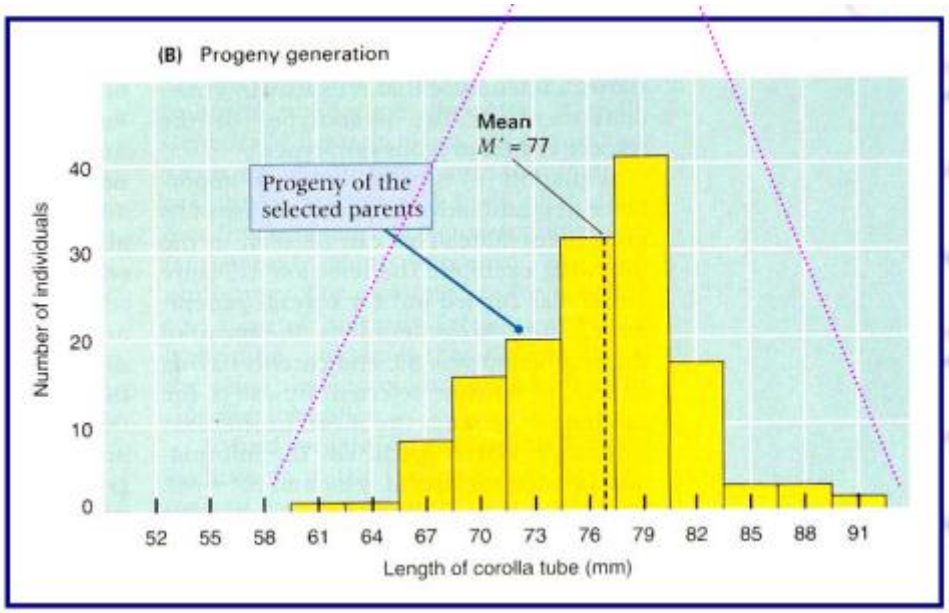
- Broad-sense heritability: H^2 is the broad-sense heritability. This reflects all the genetic contributions to a population's phenotypic variance

$$H^2 = \frac{V_G}{V_P} = \frac{V_G}{V_G + V_E}$$

- Narrow-sense heritability: h^2 is the degree to which a trait is passed from parent to offspring expressed as the ratio of the additive genetic variance to the total phenotypic variance. M is the population mean, M^* is the mean of chosen individuals, and M' is the mean of the next generation (from chosen individuals).

$$h^2 = \frac{M' - M}{M^* - M}$$





Artificial Selection

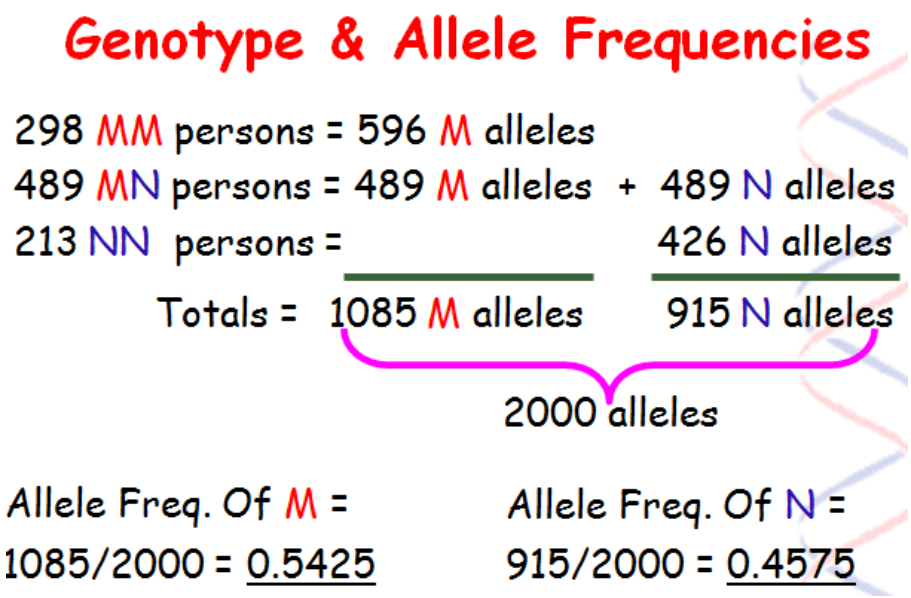
- Involves choosing a certain number of individuals from a current population to become parents for the next generation
- In the next generation, segregation and recombination produces some individuals which are better than the parents, but many also tend to be genetically inferior
- Narrow-sense heritability is often used for the process as artificial selection

What is Population Genetics?

- The study of the nature and causes of genetic variation in natural populations
- What is the origin of new genetic variation at the gene and chromosome level?
- What are the causes of changes in the frequency with which these genes/alleles occur in populations?
- Population: A group of organisms of the same species living within a given area.
- Gene Pool: The complete set of genetic information contained within a population.

Allele Frequencies

- Allele frequency is the proportion of a particular allele (variant of a gene) among all allele copies being considered
- For example, if the frequency of an allele is 20% in a given population, then among population members, one in five chromosomes will carry that allele. Four out of five will be occupied by other variant(s) of the gene
- When the allele frequency in a population is 1.0, the allele is said to be fixed
- If the allele frequency is 0, it is said to be lost



The Hardy-Weinberg Principle

- Assumptions:
  - generations are non overlapping
  - mating is random
  - population size is infinitely large
  - allele frequencies are equal in the sexes
  - there is no migration, mutation or selection
- Random mating of individuals is equivalent to the random union of gametes. As a result of random union, the genotype frequencies of the progeny may be calculated as in the Punnett square below

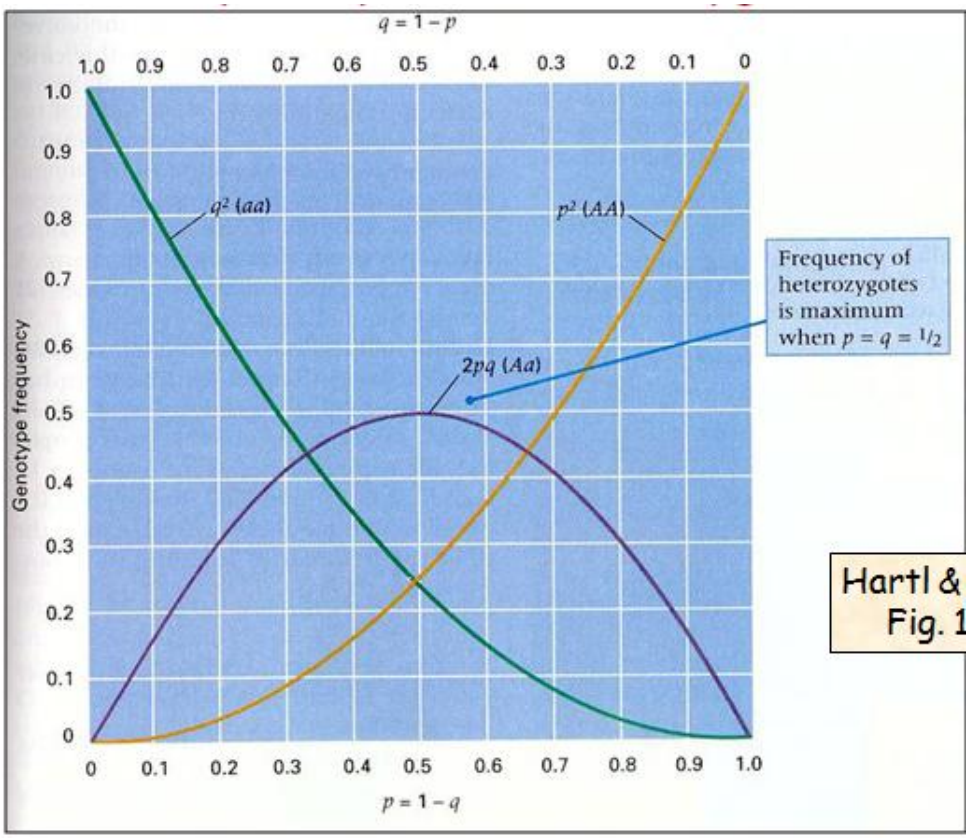
	p (M)	q (N)
p (M)	p^2 (MM)	pq (MN)
q (N)	pq (MN)	q^2 (NN)

- The Hardy-Weinberg Principle implies that allele frequencies remain constant from generation to generation, i.e. an equilibrium is reached
- Migration, mutation, selective mating, genetic drift in small populations, non-random mating, and selective pressures will all upset the HW equilibrium

Frequency of Heterozygotes

- For a rare allele, the frequency of heterozygotes far exceeds the frequency of the rare homozygote
- The ratio of heterozygotes to homozygotes is given by:

$$\frac{2pq}{q^2} = \frac{2p}{q}$$



Inbreeding

- At any diallelic locus, the proportion of heterozygotes (e.g. *Aa*) in the population after ***n*** successive generations of self-fertilisation is $\left(\frac{1}{2}\right)^n$
- In a randomly mating population, the frequency of heterozygotes for a single di-allelic locus is 2pq
- An inbred population has **fewer** heterozygotes, so let: H be the frequency of heterozygous genotypes in an **inbred** population, and ***F*** be the inbreeding coefficient (a measure of the degree of inbreeding), then:

$$F = \frac{2pq - H}{2pq}$$

- If ***H*** = ***2pq*** (no difference in freq. to random mating, i.e**. no inbreeding**) then from the equation below, ***F*** = 0
- If ***H*** = 0 (no heterozygous individuals left due to complete inbreeding), ***F*** = 1

Heterozygote Superiority

- When the heterozygote possesses greater fitness than both homozygotes, the condition is called **Overdominance** or **Heterozygote Superiority**
- Because selection favours the heterozygote, both alleles of a di-allelic gene is retained in the population
- Example: sickle cell trait

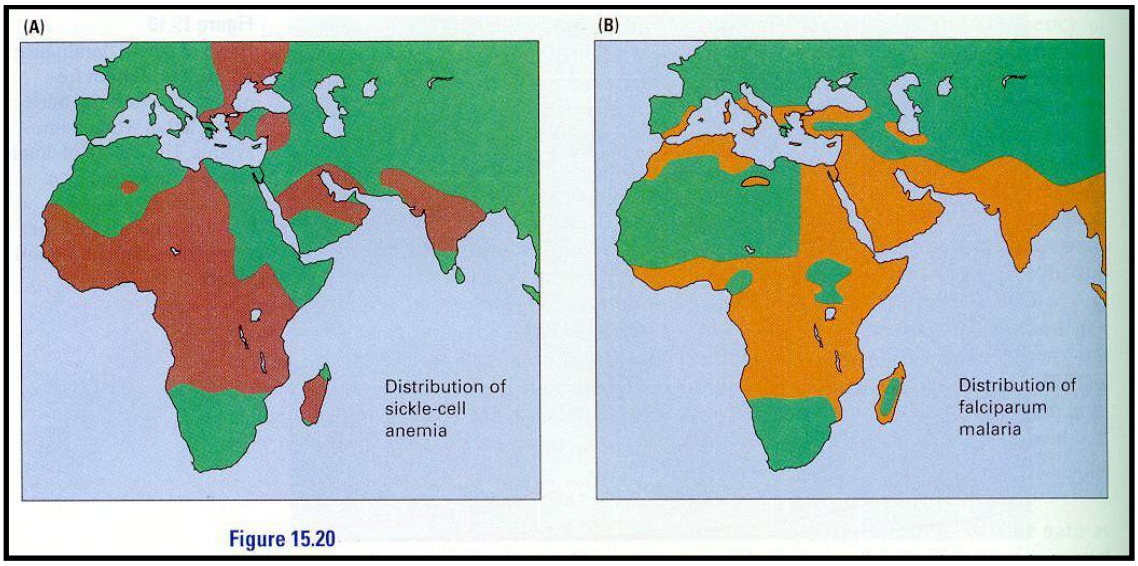


Figure 15.20