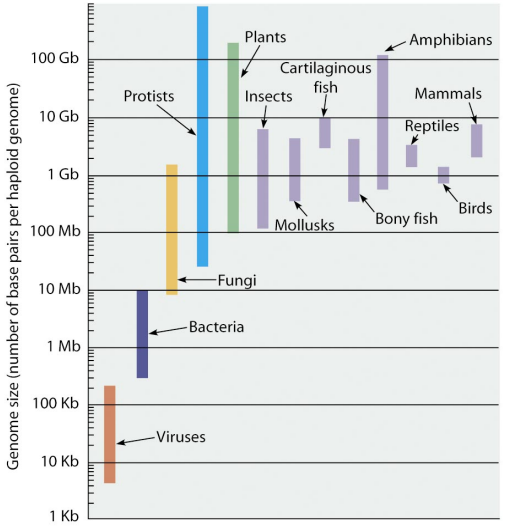
How does the human genome stack up?

Organism	Genome Size (bases)	Estimated Genes
Human (*Homo sapiens*)	3 billion	~20,500
Laboratory mouse (*M. musculus*)	2.6 billion	30,000
Mustard weed (*A. thaliana*)	100 million	25,000
Roundworm (*C. elegans*)	97 million	19,000
Fruit fly (*D. melanogaster*)	137 million	13,000
Yeast (*S. cerevisiae*)	12.1 million	6,000
Bacterium (*E. coli*)	4.6 million	3,200
Human Immunodeficiency Virus (HIV)	9700	9

Genomic Size



Eukaryotic genomes are generally larger.

Notice that **generally genome size increases with complexity of the group**, but there is **considerable variation**

(up to a thousandfold!) in some of the groups.

Genome Size

- There exists considerable variation amongst Eukaryotes!
- No link between genome size and phenotype of organisms.

  Ex: Lilly families *Fritallaria assyriaca* has 120 billion base pairs, 40 times the size of the human genome of 3.2 billion base pairs.

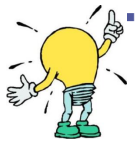
Number of Genes

- Prokaryotes have fewer genes than eukaryotes
  - BUT within eukaryotes the number of genes is often lower than expected from simply comparing the size of the genome and the number of known polypeptides.

Number of Genes

1990s -- thought humans had 100,000 genes
2000 -- 40,000 was considered a good estimate
2004 -- 30,000 seemed more accurate
2006 -- 25,000 is our best estimate

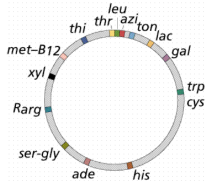
- **RNA SPLICING!!!**
  - 75% of multi-exon genes are spliced in two different ways!
- **POST-TRANSLATIONAL MODIFICATIONS!!!**
  - Cleavage of initial polypeptides can alter peptide final folding and function
  - Modification to polypeptides such as carbohydrate addition



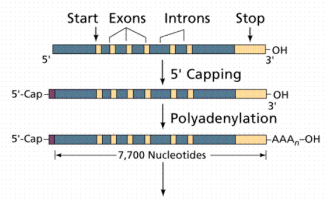
Organism	Number of genes in the genome
Myscoplasma genitalium	517
Saccharomyces cerevisiae	6,275
Arabidopsis thaliana	~ 20,000
Caenorhabditis elegans	19,099
Haemophilus influenzae	1,743
Drosophila melanogaster	13,601
Neisseria meningitdis	2,158
Homo sapiens	20,000–25,000

Gene Density & Noncoding DNA

- Eukaryotes have fewer genes per amount of bases (but have larger genomes) compared to Prokaryotes
  - Humans and mammals have the lowest gene density
- DNA composition of Prokaryotes like bacteria:
  - Genes for protein, tRNA, rRNA
  - Regulatory sequences (untranscribed) like promoters
  - No introns & very little noncoding DNA



- Eukaryotic DNA composition:
  - Noncoding sequences like introns
  - Complex regulatory sequences
  - Introns & other massive amounts of noncoding DNA

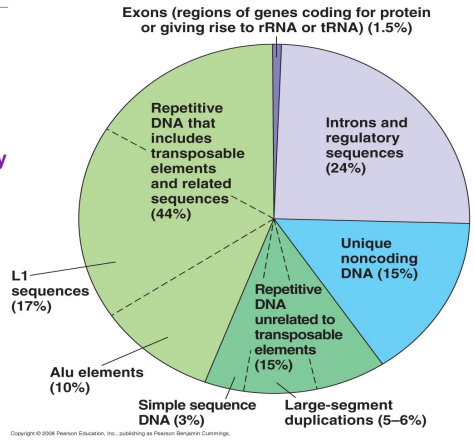


AP Biology

Multicellular Eukaryotic DNA Composition

Coding Regions = 1.5%
- tRNA, rRNA, miRNA, polypeptides

Noncoding Regions = 98.5%
- Not really 'junk' DNA!
- Function of all this DNA is still largely unknown
  - Gene related regulatory sequences and introns = 24%
- Some of this DNA are Pseudogenes
  - former genes that accumulated mutations and no longer function
- Most noncoding is repetitive DNA
  - sequences present in many copies in the genome
    - Short Tandem Repeats (STRs)
      - Side-by-side repeats of 2 to 5 bases
    - 44% are transposons

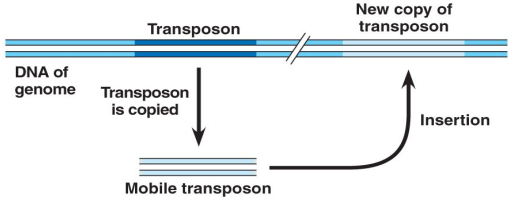


AP Biology

Transposable Elements

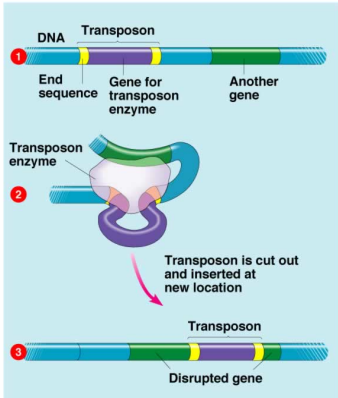
- Elements of DNA that move from one site to another by a type of DNA recombination process:

1. Transposons
   a. Cut-and-Paste mechanism of movement
      Moves from original site
   b. Copy-and-Paste mechanism of movement
      Leaves a copy behind and new copy added to new location



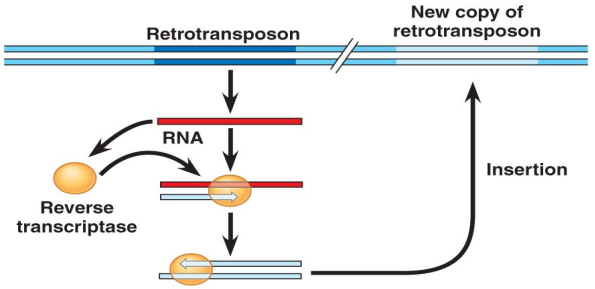
(a) Transposon movement ("copy-and-paste" mechanism)

The Yellow color is caused by transposons that moved adjacent to purple-pigment producing gene which does not get transcribed now



Transposable Elements

2. **Retrotransposons**
   - Moves by means of an RNA intermediate that is a transcript of the retrotransposon DNA
     - Reverse transcriptase encoded by the retrotransposon itself not from a viral source
   - Always leaves a copy of the retrotransposon DNA behind



(b) Retrotransposon movement

Genome Evolution

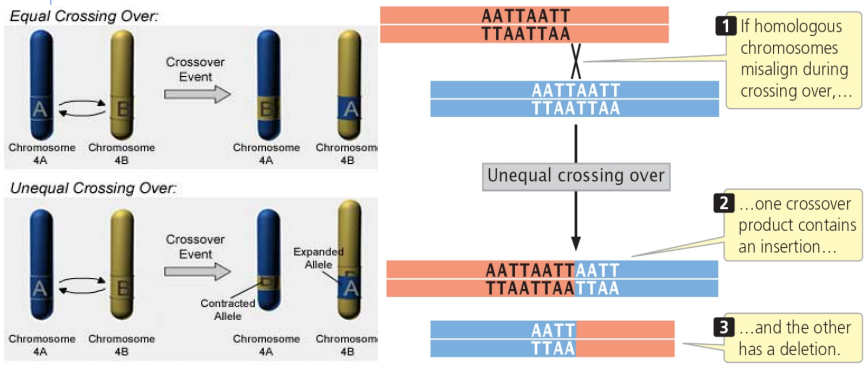
- For evolution, the genome must be altered in a way that alters phenotype in a beneficial manner.
  - Genome evolution involves
    - DNA duplications
    - DNA rearrangements
    - All other types of point mutations and sometimes even non-disjunction events that alter the ploidy number of a daughter cell
      - Polyploidy = result from errors in meiosis.

Polyploidy is common in plants. Ex: This day lilly is a tetraploid (4n)

- With gene or chromosomal DUPLICATIONS:
  - Organism still have one functional copy of genes needed so they can make any necessary RNA and protein correctly
  - Extra copies of genes can accumulate mutations easier because they do not stop the organisms form being able to produce the essential polypeptides or gene products to LIVE & REPRODUCE.

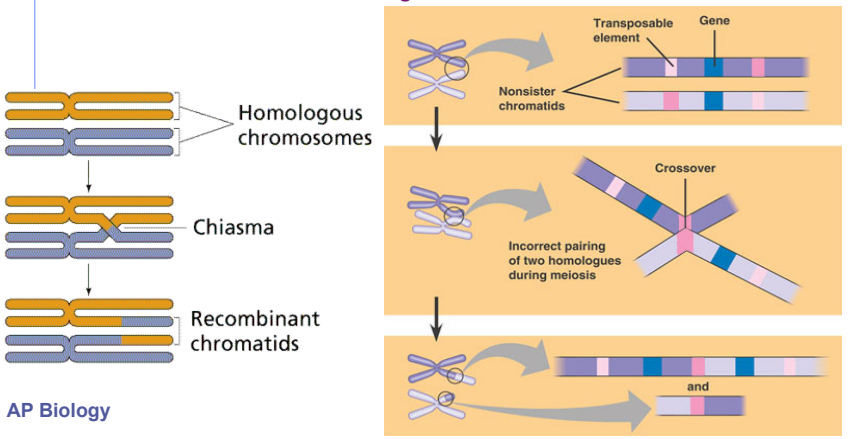
Genome Evolution

- LARGE-SCALE DUPLICATIONS
  - Duplication of gene-sized regions result from "unequal crossing over"
    - In crossing over, two non-sister chromatids line up locus by locus and exchange segments.
    - In unequal crossing over, chromosomes do not line up properly and exchange unequal amounts of DNA



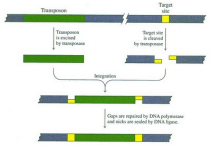
Genome Evolution

- Transposable elements can contribute to unequal crossing over
  - Crossing over may occur between similar transposable elements on two non-sister chromatids that are not located at the same locus but are located on the sides of a gene



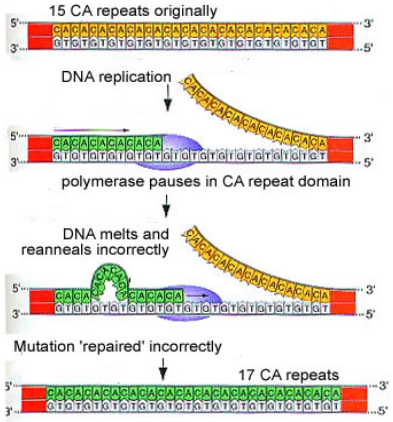
Transposable elements contribute to genome evolution

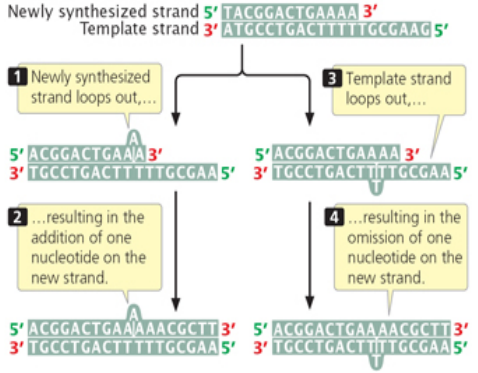
- Unequal crossing over can cause DNA duplication.
- Also because some transposons duplicate themselves, they create numerous homologous regions of DNA throughout the genome, leading to regions were non-homologous chromosomes could cross over
- Transposons could carry genes to new locations in genome.
- Transposons might carry an exon from one gene to another gene and thereby introduce a new region of DNA that will be coded into part of the protein (introducing a new domain into the polypeptide)
- Transposons can jump in the middle of protein-coding gene regulatory sequences, preventing normal transcription
- Transposons can insert themselves into regulatory sequences and lead to increased or decreased polypeptide production



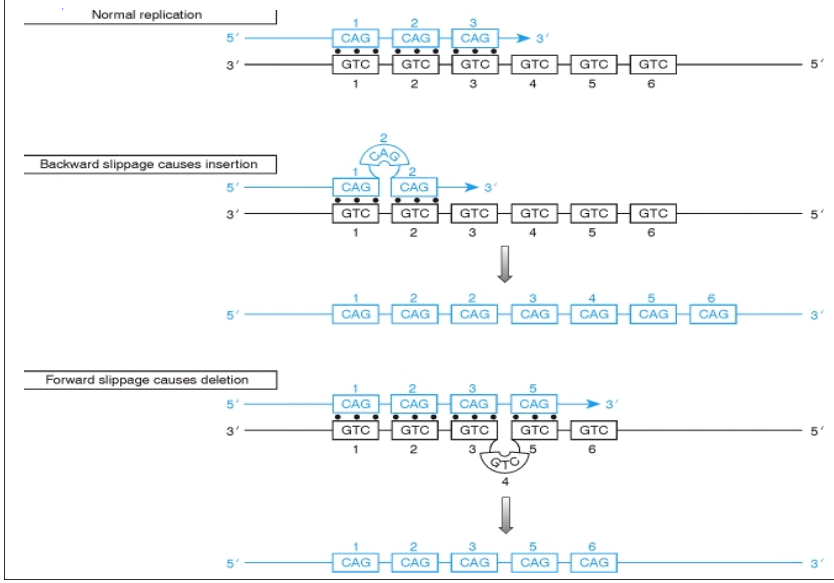
Genome Evolution

- Slippage during DNA replication can result in duplications
  - Template shifts with respect to new complimentary strand.
  - Part of template is either skipped or copied twice resulting in DNA being deleted or duplicated in the daughter strand





Slippage of DNA polymerase

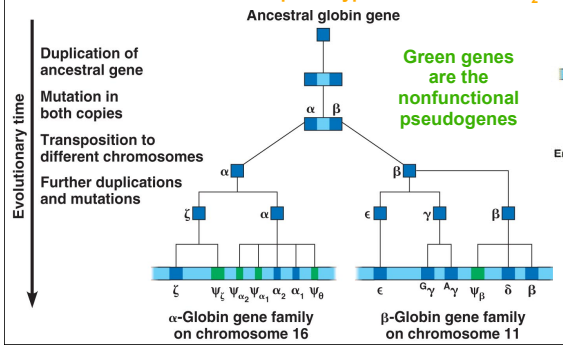


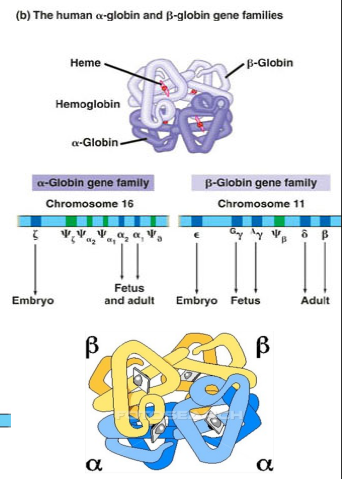
Duplicated Genes can change in sequence over time

- If genes are duplicated, the new copy can accumulate mutations and maybe alter phenotype of future cells

Ex: Hemoglobin's subunits are made by genes that belong to 2 globin multigene famlies = collections of two or more identical or very similar genes

- Different genes within each family are expressed at different points in development and produce polypeptides with different phenotypes & affinities for O_2

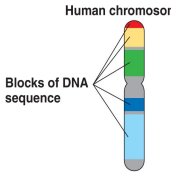


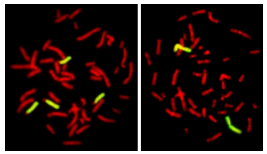


Genome Evolution

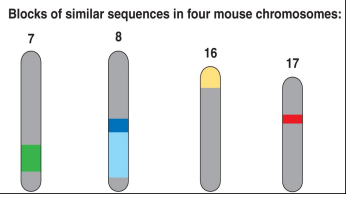
Rearrangement of chromosome structure

- Can result from mistakes during meiotic recombination if DNA broke and was joined back together incorrectly
  - Does not alter gene number BUT if these gametes fuse with normal gametes during reproduction the *offspring would be missing or have large scale duplications of DNA*.
    - Could be the beginning of species divergence and the evolution of two separate species from one ancestral species.





Chimp 48 Human 46 (fusion of two ancestral primate chromosomes)



Genome Evolution

- Exon Duplication & Exon Suffling contribute to genome evolution.
- Exons often code for 'domains' = distinct structural or functional regions of proteins.
  - One domain may be an enzyme active site, while another a receptor region.
  - Different exons code for the different domains of a protein.
- Exon Shuffling = the mixing or matching of different exons within a gene or between non-allelic genes
  - Can lead to significant genome rearrangement (evolutionarily significant) and new proteins with novel combinations of functions
    - exons are mixed & matched due to meiotic division errors via unequal crossing over at non-sister transposable element sites