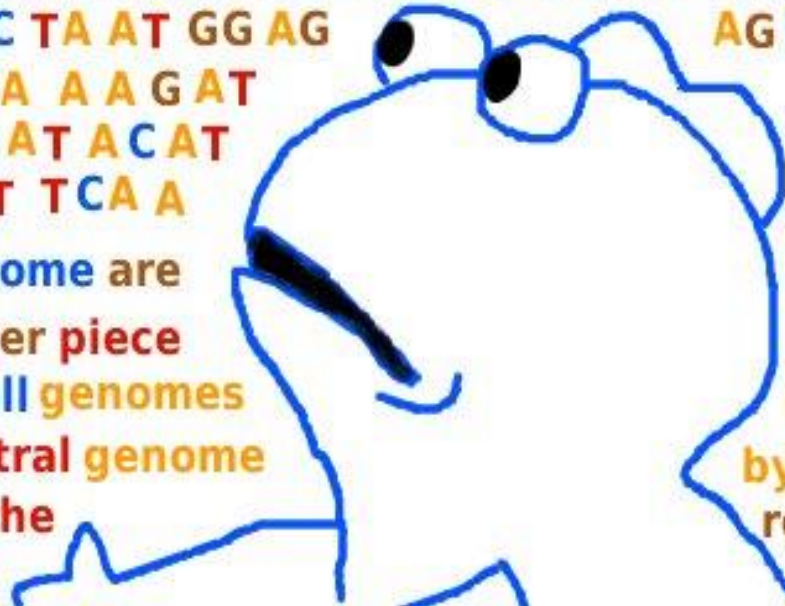# SYSTEMATICS & EVOLUTION

Molecular Evolution

GINCY C GEORGE
(Assistant Professor On Contract)

Molecular Evolution

ATGCAACTTACATGCCTTAGGATCGAGTCAAGATGTATCG
CTTAAAGATTGCCAGGACTGCAATACATTCAAACTGC
CATGCCTTAATTCAAGTTTAACAGGAAGATCCAGGATG
ATACATTCAATTAGGATCGAATGTATCGTAACCAGGATGC
AAGATATCCAGTGCAAACATGTTAGGACCCTTAATTCA
ATACATTCAAACTGCAAGATTGCCAAGTTAGGATCGAG
TTAGGATCGAATGTATCCTTAATTCAAGTAAGATTGCCAA
CCAAGGATCGAGTCAAGAATATCCAGCTGCAAGATTAATT
AAGAATATCTAATGGAG   AGCCTTATAAAG
  CCTTAATTTAAAGAT   AGATATCCAG
TTAGGATCGAATACAT   GCAAACATG
ATGCCTTAATTCAA   In principle all
parts of the genome are   descended from
one much smaller piece   of nucleic acid.
All genes, and all genomes   have originated
from this ancestral genome   by gene duplication
loss of parts of the   resulting genomes,
insertions, and   rearrangements

Molecular Evolution

- Molecular evolution is the area of evolutionary biology that studies evolutionary change at the level of the DNA sequence.

Molecular Evolution

- It includes the study of rates of sequence change, relative importance of adaptive and neutral changes, and changes in genome structure.

Molecular evolution examines DNA and proteins, addressing two types of questions:

How do DNA and proteins evolve?

How are genes and organisms evolutionarily related?

Study of how genes and proteins evolve and how are **organisms related based on their DNA sequence**

- Molecular evolution therefore is the determination and comparative study of DNA and deduced amino acid sequences.

- Sequences from different organisms or populations are matched or aligned

- Evolution at molecular level is observable at the base (nucleotide) level changes in the DNA and amino acid changes in proteins
- Both can be studied by examining the differences between species

- Both polymorphism and evolutionary changes between species can be explained by two processes ie;

- Natural selection and Neutral drift
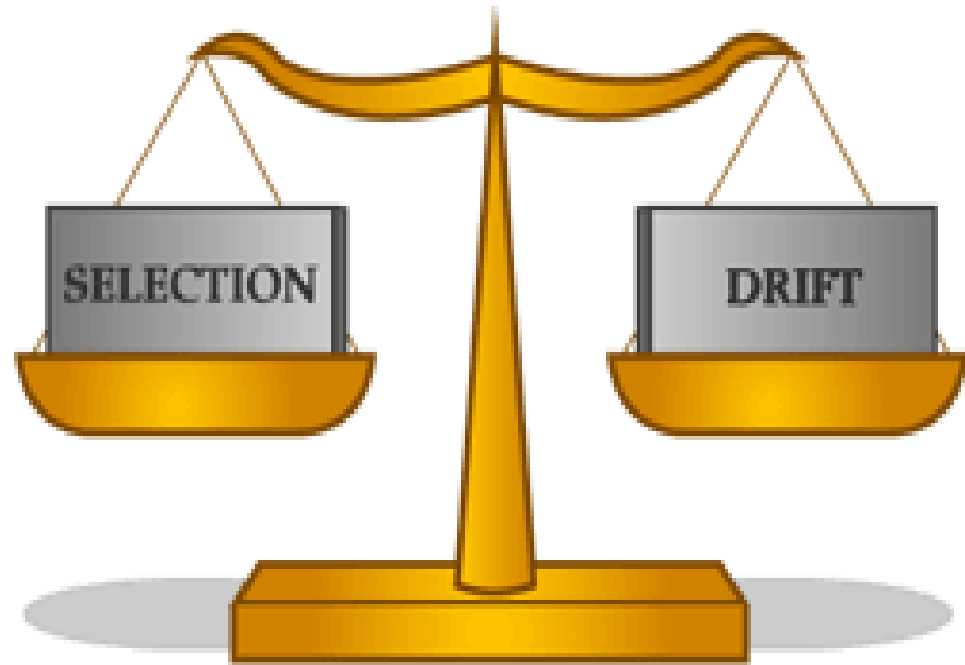
- The main factors that influence Natural selection and Neutral drift are population size and the selection coefficient of the different genotypes

- If the population is small and the selection coefficient low; genetic drift dominates,
- Whereas natural selection dominates if the population and selection coefficients are large

- **Evolution of Modern species has incorporated millions of molecular changes**

- In **1968** and **1969** Kimura and King and Jukes suggested that most evolutionary changes at the molecular level is driven by random drift rather than natural selection.

- Kimura called this view as the **neutral theory** of Molecular evolution

Neutral Theory of Molecular Evolution (MNT)

- Motoo Kimura (1968)

- 1. Vast majority of molecular differences are selectively neutral

- 2. Most evolutionary change is the result of genetic drift acting on neutral alleles

Neutral Theory of Molecular Evolution
(Motoo Kimura – 1968,69)

- The Neutral Theory states that evolution at the level of the DNA and proteins, but not at the level of adaptation or by natural selection
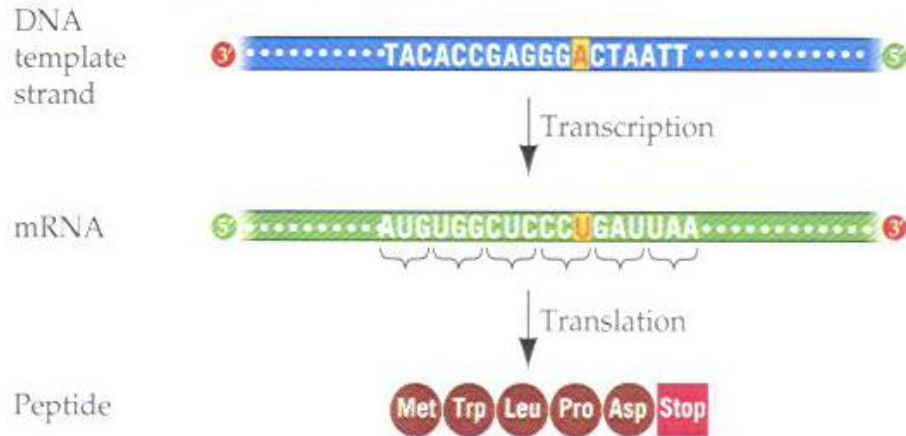
- Neutral theory of Motoo Kimura states that many mutations have such a small effect on the fitness of an organism that they can be considered as "neutral"

Many mutations in sequences of genes do not alter the proteins encoded by those genes

virtually always true for synonomous substitutions

Synonymous mutation

Mutation at position 12 in DNA: C→A

DNA template strand: 3' ...TACACCGAGGGACTAATT... 5'

↓ Transcription

mRNA: 5' ...AUGUGGCUCCCUGAUUAA... 3'

↓ Translation

Peptide: Met Trp Leu Pro Asp Stop

		Second base		
	U	C	A	G
U	UUU / UUC Phe; UUA / UUG Leu	UCU / UCC / UCA / UCG Ser	UAU / UAC Tyr; UAA Stop / UAG Stop	UGU / UGC Cys; UGA Stop; UGG Trp	U C A G
C	CUU / CUC / CUA / CUG Leu	CCU / CCC / CCA / CCG Pro	CAU / CAC His; CAA / CAG Gln	CGU / CGC / CGA / CGG Arg	U C A G
A	AUU / AUC / AUA Ile; AUG Met/Start	ACU / ACC / ACA / ACG Thr	AAU / AAC Asn; AAA / AAG Lys	AGU / AGC Ser; AGA / AGG Arg	U C A G
G	GUU / GUC / GUA / GUG Val	GCU / GCC / GCA / GCG Ala	GAU / GAC Asp; GAA / GAG Glu	GGU / GGC / GGA / GGG Gly	U C A G

First base — Third base

- If protein function is not altered by a mutation, the allelic variant that results from that mutation is unlikely to be influenced by natural selection...

- All populations have genetic variation that may or may not affect the fitness of the organism.

- The neutral theory of <u>molecular</u> evolution suggests that most of the genetic variation in populations is the result of **mutation and genetic drift** and not selection.

- This can be explained by the fact that many amino acids in a protein can be exchanged for other amino acids with similar biochemical properties, with negligible impact on the overall function or structure of protein

- At the level of DNA, many mutations in protein coding genes are synonymous because they do not result in any changes in the amino acid sequence of the protein.

- The fate of these mutation – neutral mutation is detrimental stochastically by a process known as <u>genetic drift</u>.
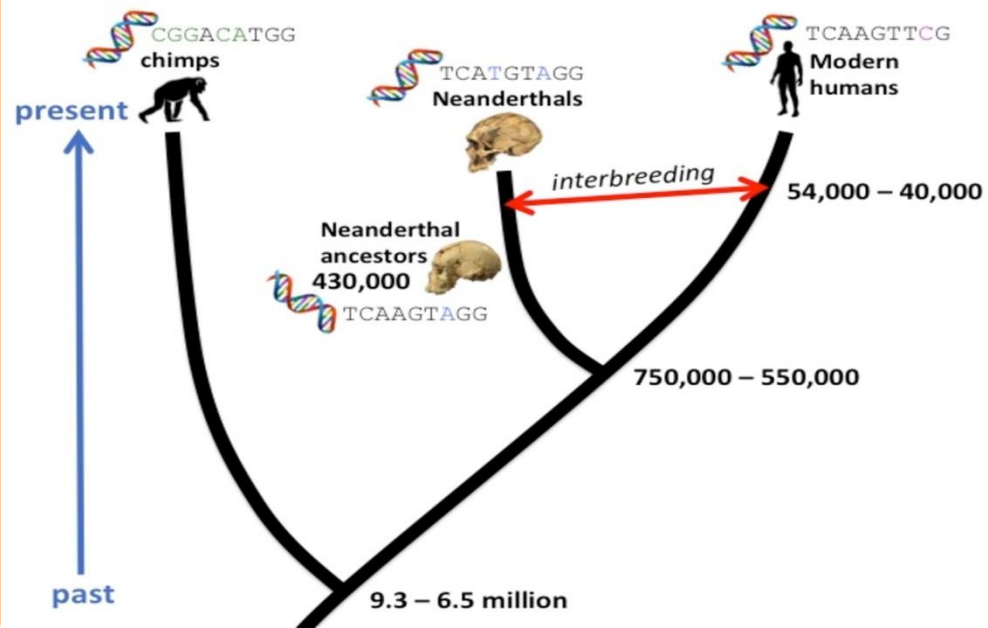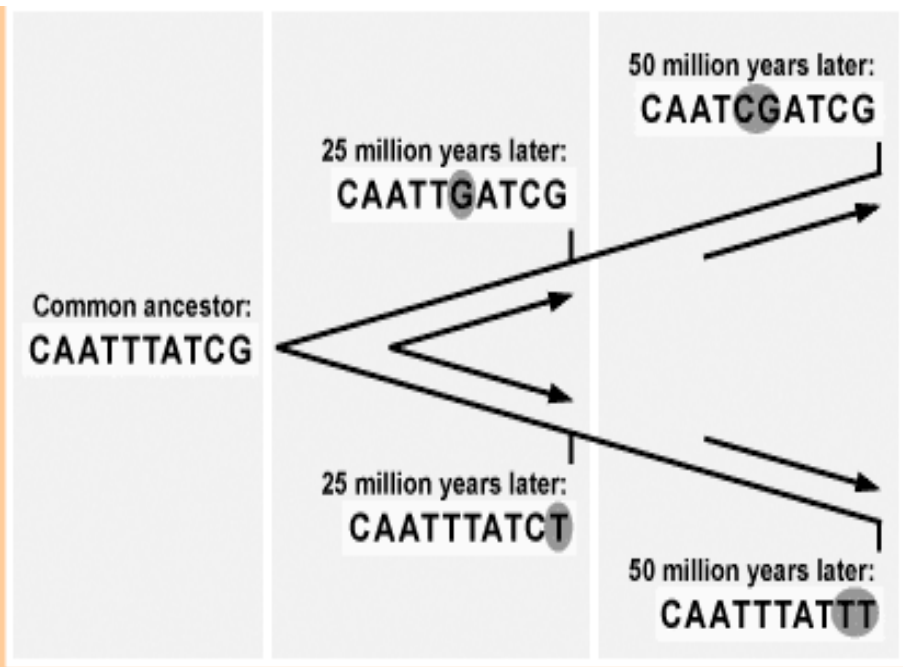
- One of the predictions of the neutral theory is that rates of molecular evolution are constant among lineages.

- This prediction refers significantly to the rate of genetic change per generation.

- As a consequence, the species with shorter generations tend to evolve more quickly per unit time.

- According to neutral theory of molecular evolution most mutations have no influence on fitness of the organism.

- Rate of Change is thus effected only by mutation rate, and so should be relatively constant within a species

Molecular Clock

- Molecular clock is an essential tool in many areas of evolutionary biology including systematics, molecular ecology and conservation genetics.

- Molecular clock is proposed in the year 1960s by Zuckerkandl and Pauling

- Molecular clock hypothesis states that, DNA and protein sequences evolve at a rate that is relatively constant over time and among different organism

Common ancestor:
CAATTTATCG

25 million years later:
CAATTGATCG

50 million years later:
CAATCGATCG

25 million years later:
CAATTTATCT

50 million years later:
CAATTTATTT

CGGACATGG
chimps

TCATGTAGG
Neanderthals

TCAAGTTCG
Modern humans

present

interbreeding

54,000 – 40,000

Neanderthal ancestors
430,000
TCAAGTAGG

750,000 – 550,000

past

9.3 – 6.5 million

- According to Molecular Clock hypothesis the amount of genetic differences between sequences is a function of time since separation and the rate of molecular changes is constant to predict times of divergence ie; it predicts a constant rate of molecular evolution among species

- It is also a method of genetic analysis that can be used to estimate evolutionary rates and time scales using data from DNA or protein.
- Traditionally such inferences were made from the fossil record, coupled with radiometric dating.

- Fossils can provide an estimate of when different lineages first appeared and when species diverged from each other
- Whereas, molecular clock can be used to estimate evolutionary rate from genetic data

- Molecular clock implies that the genetic difference between any two species is proportional to the time since they shared a common ancestor

➢ The molecular clock represents the only method for studying the evolutionary time scales of organisms that have failed to leave any trace in the fossil record, including many groups of invertebrates, bacteria and viruses

➢ It allows to estimate the timing of events that are too recent to be resolved by fossil evidence, such as divergence among conspecific populations

➢ It also used in phylogenetic analyses, which aims to reconstruct evolutionary trees that show the relationships among species of interest

Internal nodes in the tree represent evolutionary divergence events.

➢ The timings of these events can be estimated using molecular clocks

Disadvantages of molecular clocks

➢ The molecular clock does not run as smoothly as neutral theory predicts

➢ Irregularities result from natural selection in which some DNA changes are favored over others

➢ Estimates of evolutionary divergences older than the fossil record have a high degree of uncertainty.

Gene family

- Gene family is a set of several similar genes; formed by duplication of a single gene, and generally with similar biochemical functions
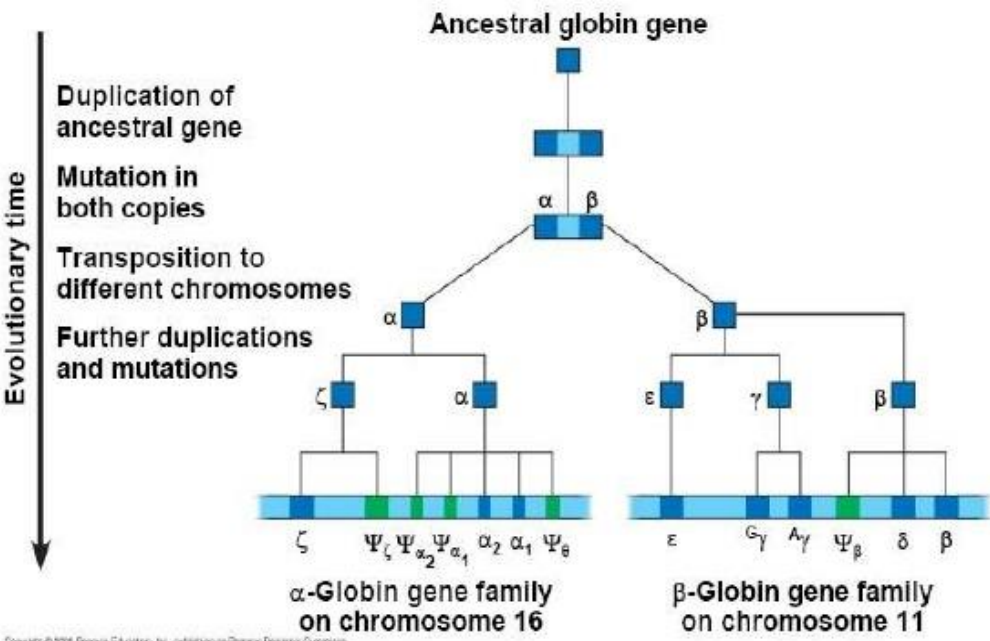
- It is set of homologous genes within one organism.
- The members of gene family may include expressed genes as well as non expressed sequences.
- Such non expressed sequences include promoters, operators, transposable genetic elements and pseudogenes

- A **promoter** is a region of DNA that initiates transcription of a particular gene.

- In genetics, an **operator** is a segment of DNA to which a transcription factor binds to regulate **gene** expression by repressing it. ... Repressors bind to **operators** to prevent transcription.

- A **transposable element** (**TE** or **transposon**) is a DNA sequence that can change its position within a genome, sometimes creating or reversing mutations and altering the cell's genetic identity and genome size.

- Pseudogenes are DNA sequences,, that have lost their protein-coding ability or are otherwise no longer expressed in the cell.

- The members of gene family may be either clustered together, dispersed on different chromosomes or present in a combination of both
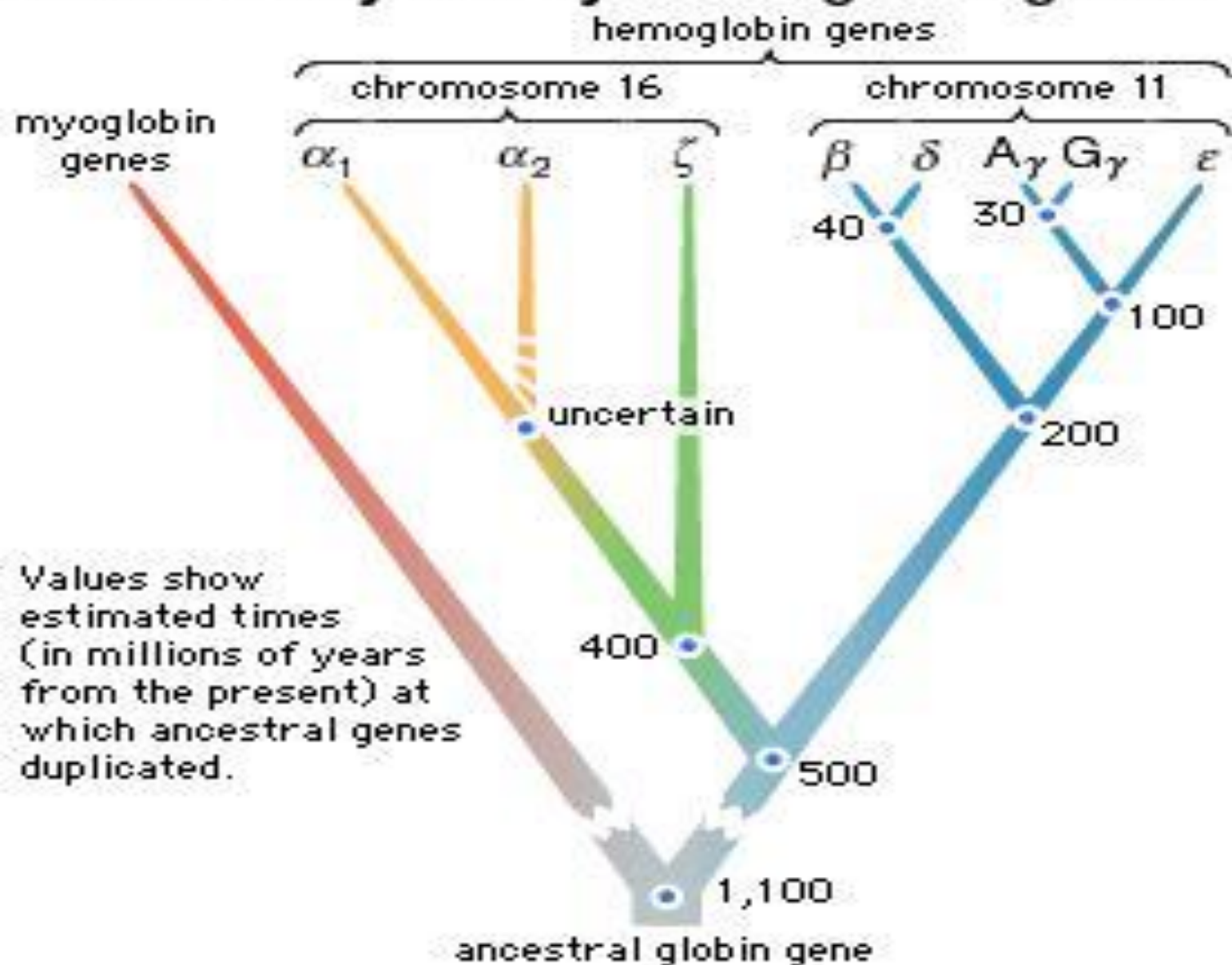
- Eg: Human Globin Genes. The ten **genes** are in two clusters on different chromosomes, called the α-globin and β-globin loci.

Human Globin Genes

Evolutionary time →

Ancestral globin gene

Duplication of ancestral gene

Mutation in both copies — α β

Transposition to different chromosomes — α β

Further duplications and mutations — ζ α | ε γ β

ζ Ψζ Ψα₂ Ψα₁ α₂ α₁ Ψθ | ε Gγ Aγ Ψβ δ β

α-Globin gene family on chromosome 16 | β-Globin gene family on chromosome 11

- Genes are categorized into families based on shared nucleotide or protein sequences
- The position of exons within the coding sequence can be used to infer common ancestry.

**Knowing the sequence of the protein encoded by a gene may provide the following;**

➢ similarities among protein sequences

➢ Similarities or differences among DNA sequences

➢ Knowledge of the protein secondary structures for further information about ancestry

- If the genes of a gene family encode proteins, the term **protein family** is often used in an analogous manner to gene family

- The expansion or contraction of gene families along a specific lineage can be due to chance or can be the result of natural selection

- Gene complexes are simply tightly linked groups of genes, often created through gene duplication (or segmental duplication)
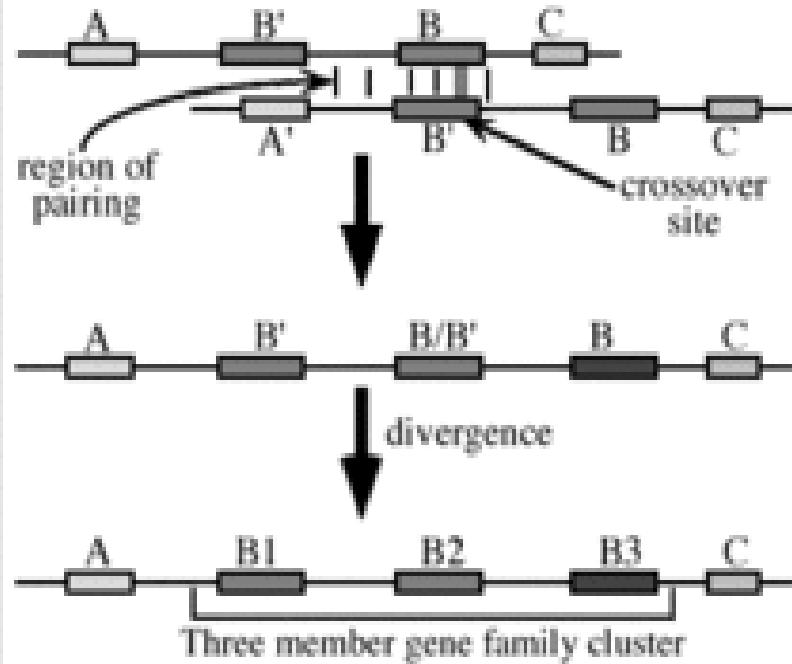
Unequal crossing over generates gene families.
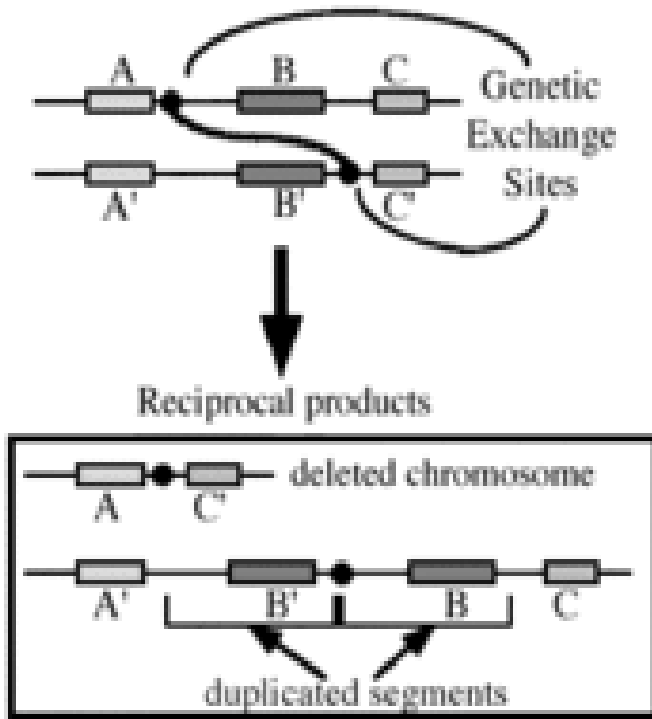
The left side illustrate an unequal crossing over event and the two products that are generated. One product is deleted and the other is duplicated for the same region.

The right side illustrates a second round of unequal crossing over that can occur in genome that is homozygous of the original duplicated chromosome. Over time the three copies of the B gene can diverge into three distinct functional units of a gene family

Molecular Drive

- Molecular drive is an evolutionary process, like natural selection and neutral drift, that changes the genetic composition of a population through generations

- Molecular drive is distinct from natural selection and neutral drift in that it emerges from the activities of a number of ubiquitous mechanisms of DNA turn over, such as gene conversion, unequal crossing over, slippage, transposition, retro transposition and so on.

- Molecular drive implies a known mechanism and a predictable one way movement of genetic information non-randomly into the population, causes a particular genes to pass preferentially to the next generation