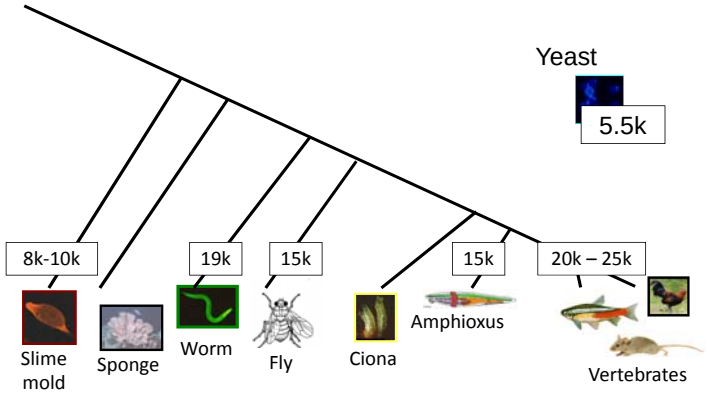
Where do new genes come from?

03-327/727 Lecture 4

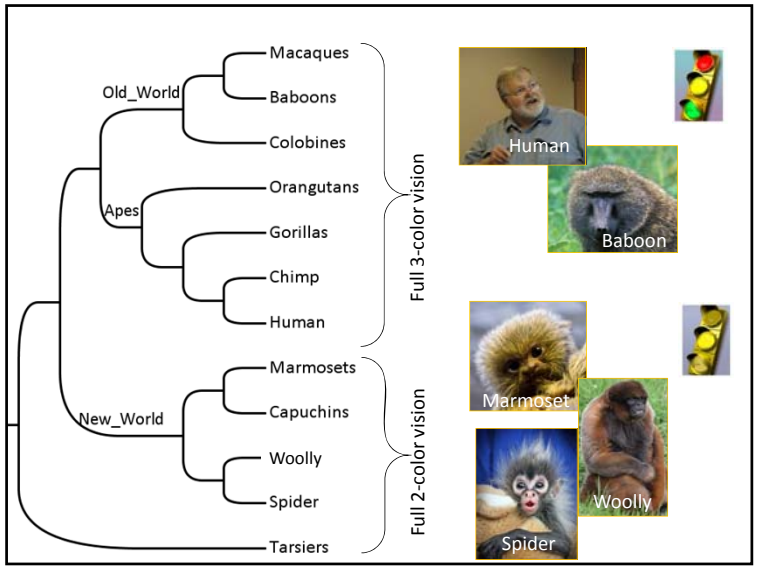
The number of genes in selected genomes

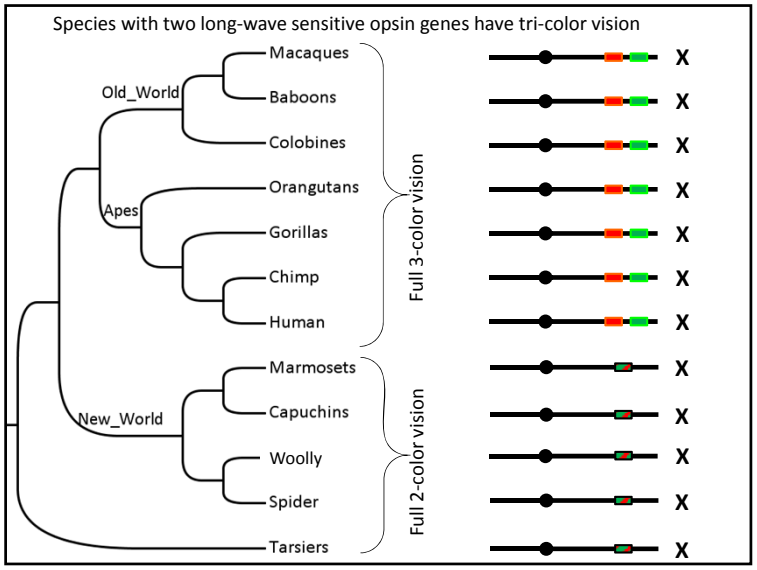


Where do new genes come from?

- Gene duplication
  - Mechanisms of duplication
  - Fate of duplicated genes
- *de novo* gene formation
- Formation of mosaic genes by duplication and shuffling of pieces of genes
- Horizontal gene transfer

Example 1: Color vision

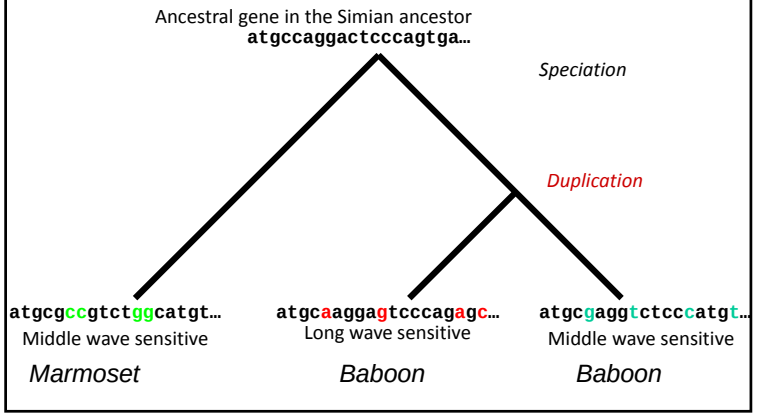




Where do new genes come from?

- Gene duplication
  - Mechanisms of duplication
  - Fate of duplicated genes
- *de novo* gene formation
- Formation of mosaic genes by duplication and shuffling of pieces of genes
- Horizontal gene transfer

New genes arise through duplication and modification of existing genes



Gene duplication
Gene families evolve on a range of scales

- Single genes
- Chromosomal segments a few genes)
- ~~Partial chromosomes~~
- ~~Entire chromosomes~~
- Whole genome duplication

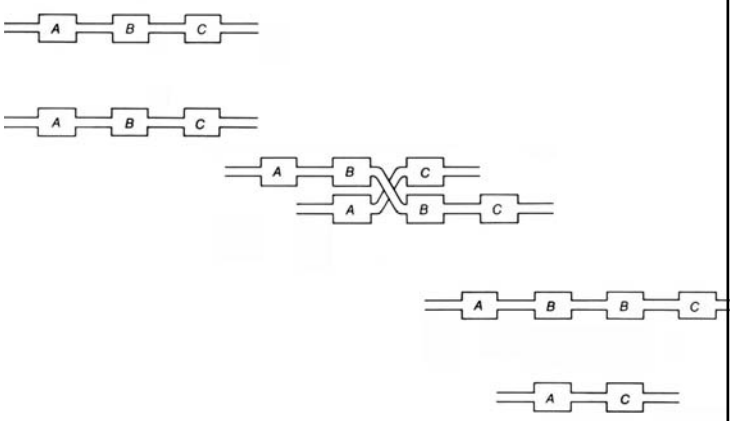
Gene duplication
Gene families evolve on a range of scales

- Single genes
- Chromosomal segments (a few genes)

Tandem duplication

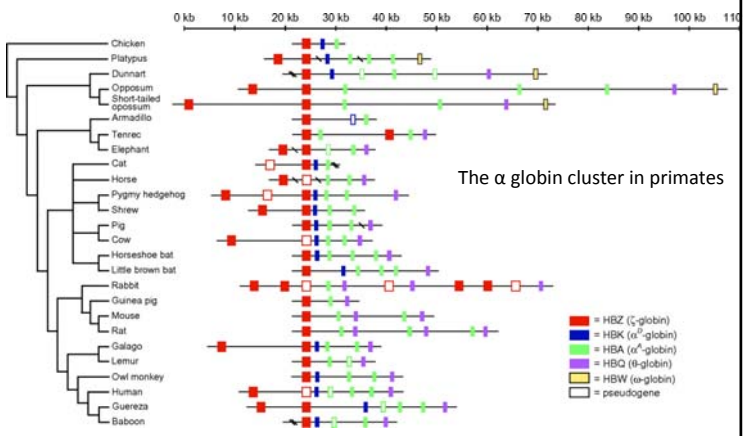
- ~~Partial chromosomes~~
- ~~Entire chromosomes~~
- Whole genome duplication

Tandem duplication and gene loss via unequal crossing over



- Type of duplication: Tandem duplication
- Scale: One or a small number of genes
- Mechanism: Unequal crossing over due to non-allelic homologous recombination
- Genomic pattern:
  - A B C C C C D E F
  - P Q R S T R S T R S T U V

Examples of tandem duplications: The α globin



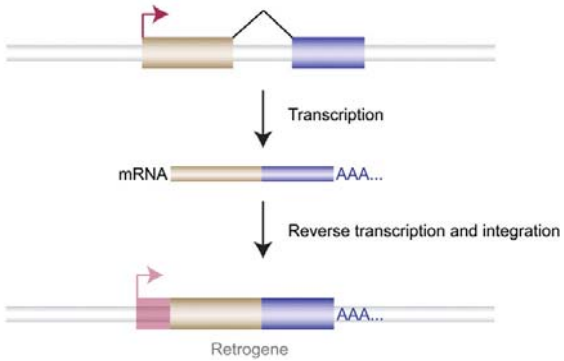
Gene duplication
Gene families evolve on a range of scales

- Single genes
- Chromosomal segments (a few genes)

Retrotransposition

- ~~Partial chromosomes~~
- ~~Entire chromosomes~~
- Whole genome duplication

"Retrogenes" arise through reverse transcription by an RNA transposon

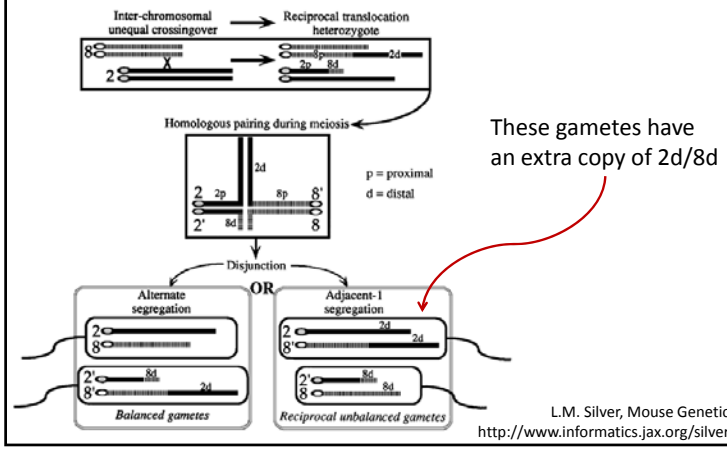


- Type of duplication: Retrotransposition
- Scale: One or a small number of genes
- Mechanism: Reverse transcription by a reverse transcriptase enzyme encoded by an RNA transposon and subsequent reintegration in the genome.
- Genomic pattern (in eukaryotes): A gene that lacks introns and has similar sequence to a gene that possesses introns.

Gene duplication
Gene families evolve on a range of scales

- Single genes
- Chromosomal segments (a few genes)
- Partial chromosomes
- Entire chromosomes
- Whole genome duplication

Chromosome segregation from a reciprocal translocation heterozygote



L.M. Silver, Mouse Genetics
http://www.informatics.jax.org/silver/

- Type of duplication: Block duplication
- Scale: Chromosomal fragment
- Mechanism: Reciprocal translocation followed by meiosis.
- Genomic pattern:

A B C D E F

P Q R S T U V C D E F

Gene duplication
Gene families evolve on a range of scales

- Single genes
- Chromosomal segments (a few genes)
- ~~Partial chromosomes~~
- ~~Entire chromosomes~~
- Whole genome duplication

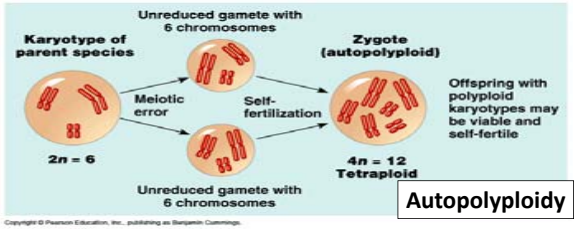
Complete and partial chromosomal duplications are almost always deleterious

nichd.nih.gov

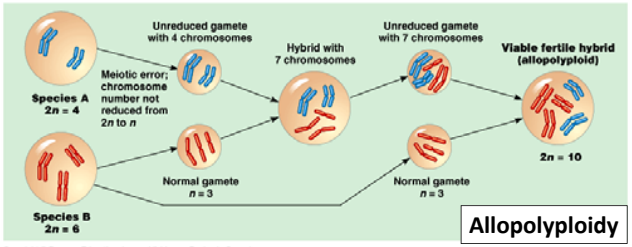
e.g. Down's syndrome

Gene duplication
Gene families evolve on a range of scales

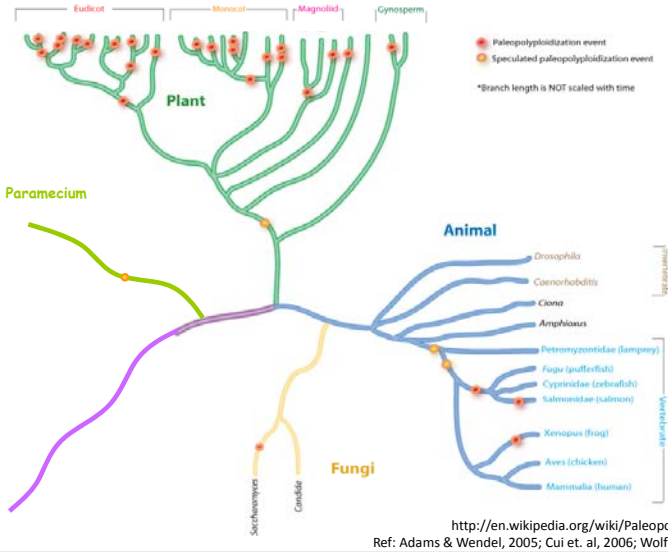
- Single genes
- Chromosomal segments (a few genes)
- ~~Partial chromosomes~~
- ~~Entire chromosomes~~
- Whole genome duplication



Autopolyploidy

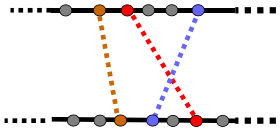


Allopolyploidy



http://en.wikipedia.org/wiki/Paleopolyploidy
Ref: Adams & Wendel, 2005; Cui et. al, 2006; Wolfe, 2001.

- Type of duplication: Whole genome duplication
- Scale: Whole gene duplication
- Mechanism:
  - Autopolyploidy: errors in meiosis
  - Allopolyploidy: hybridization
- Genomic pattern:



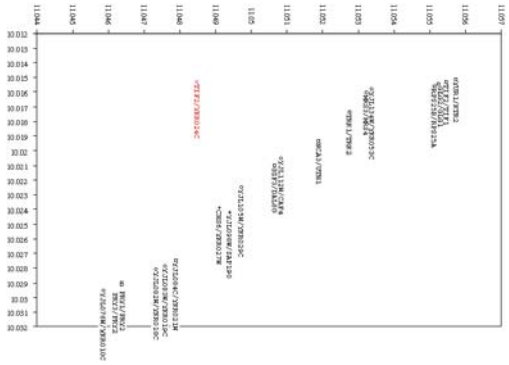
Distinct chromosomal regions with similar gene content.

Comparison matrix for yeast

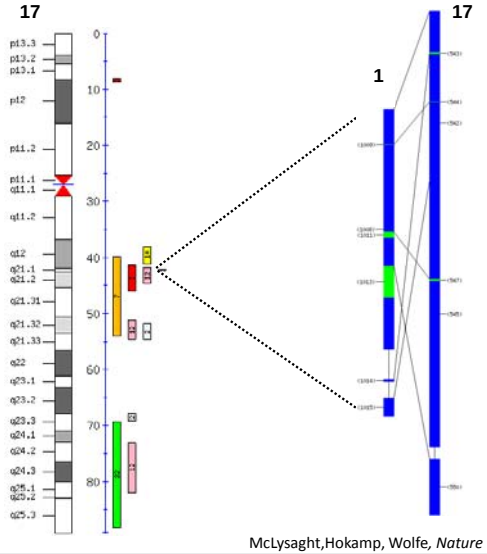
Wolfe and Shields, Nature, 1997

Chromosome 11

Chromosome 10



Human gene clusters

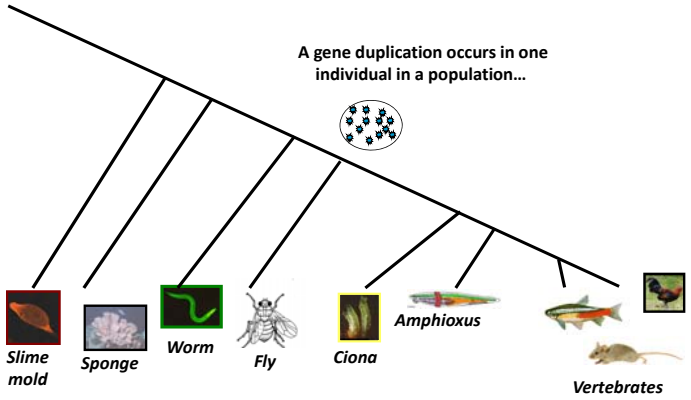


McLysaght,Hokamp, Wolfe, *Nature Genetics*, 2002

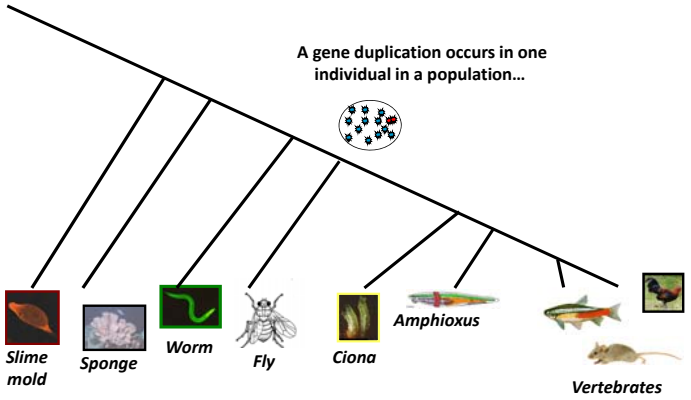
Where do new genes come from?

- Gene duplication
  - Mechanisms of duplication
  - Fate of duplicated genes
- Horizontal gene transfer
- Formation of mosaic genes by duplication and shuffling of pieces of genes
- *de novo* gene formation

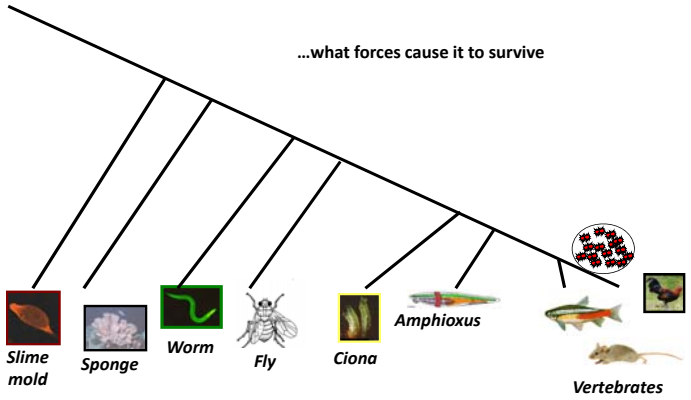
Animal evolution and genome sizes



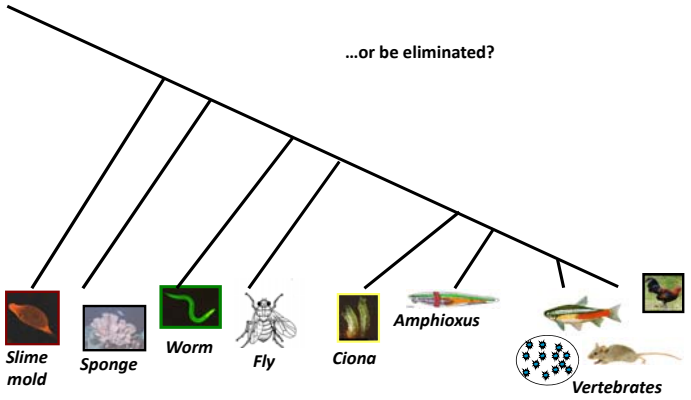
Animal evolution and genome sizes



Animal evolution and genome sizes

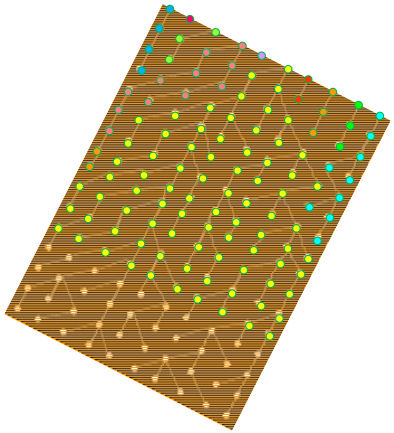


Animal evolution and genome sizes



Fixation of an allele in a population

An allele reaches fixation when all individuals in the population have that allele.



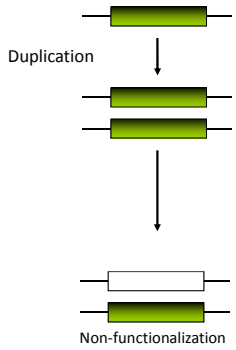
Genetic drift: In the absence of selection, alleles will either be eliminated or reach fixation due random fluctuations in allele frequencies.

Where do new genes come from?

- Gene duplication
  - Mechanisms of duplication
  - Fate of duplicated genes
- *de novo* gene formation
- Formation of mosaic genes by duplication and shuffling of pieces of genes
- Horizontal gene transfer

Fates of duplicate genes

Non-functionalization: One copy sustains deleterious mutations, loses its function, and becomes a "pseudogene"



Pseudogenes

A segment of DNA that looks like a gene

e.g., possesses

- Start codon, stop codon
- Upstream regulatory sequences
- Sequence composition typical of coding sequences
- Sequence similarity to known functional genes.

Pseudogenes

A segment of DNA that looks like a gene, but <u>does not encode a functional gene product</u>

e.g., Start and stop codon are not in frame

- Sequence composition *not* typical of coding sequences
- Sequence similarity to known functional genes, but contains premature stop codons

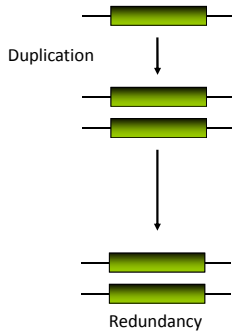
Fates of duplicated genes

Non-functionalization: One copy sustains deleterious mutations, loses its function, and becomes a "pseudogene"

Redundancy: Both copies are retained and continue to perform the same function.

- The most common explanation for redundancy is *dosage:*
- Multiple copies of a gene are beneficial because the cells demand for the gene product is very high.
- The canonical example: ribosomal genes.

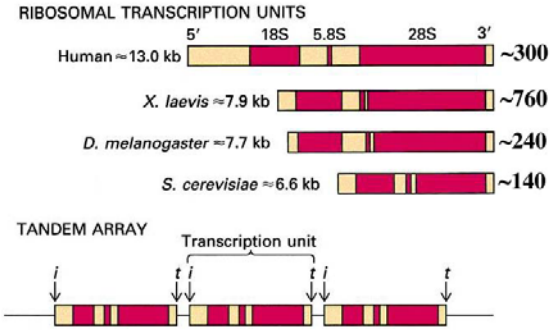
Fates of duplicate genes



Both copies continue to perform the same function

Typically, both copies are retained provide more ribosomal proteins, supporting rapid, high volume protein synthesis.

"dosage"



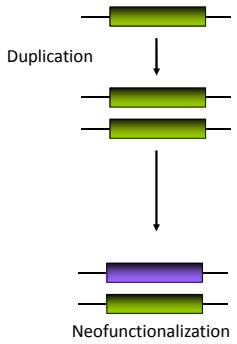
Fates of duplicated genes

Non-functionalization: One copy sustains deleterious mutations, loses its function, and becomes a "pseudogene"

Redundancy: Both copies are retained and continue to perform the same function

Neofunctionalization: Both copies are retained and one (or both) takes on a new function

Fates of duplicate genes



One (or both) copies evolve to perform a new or modified function that is beneficial

Fates of duplicated genes

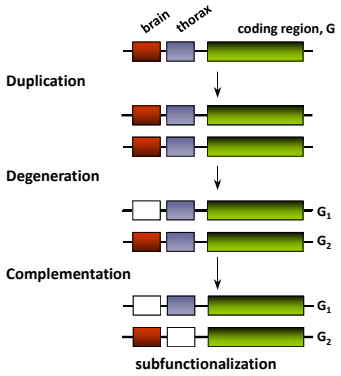
Non-functionalization: One copy sustains deleterious mutations, loses its function, and becomes a "pseudogene"

Redundancy: Both copies are retained and continue to perform the same function

Neofunctionalization: Both copies are retained and one (or both) takes on a new function

Subfunctionalization: Both copies are retained and the functions of the original gene are partitioned between them

Fates of duplicate genes
The subfunctionalization model

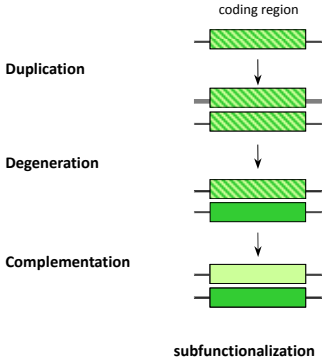


If the function of G in the brain contributes to fitness, then G2 is protected by selection as soon as the expression of G1 in the brain is lost.

Similarly, G2 is protected by selection as soon as the expression of G1 in the thorax is lost.

The loss of a transcription factor binding site requires just a few mutations and can happen more easily than mutation to a new biochemical function.

Fates of duplicate genes
The subfunctionalization model



Subfunctionalization can also involve partitioning of biochemical functions of a multifunctional protein.

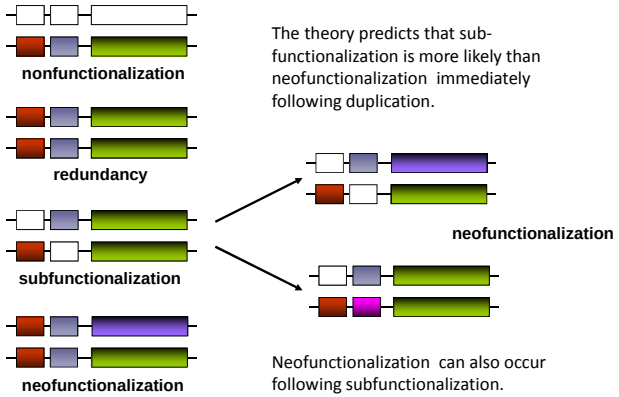
Rationale for the subfunctionalization model

Survival of a new duplicate gene depends on a race between

- fixation in the population (i.e., the new duplicate has spread to all members of the population)
- Deletion or loss of function mutations (i.e., the new duplicate becomes a pseudogene).

If the new gene is not protected by selection, then it is likely to lose the race. However, it can be **preserved** through a beneficial mutation or acquisition of an essential function. Acquisition of an essential function by loss of a transcription factor binding site, can occur much more quickly than a series of mutations leading to a new biochemical function.

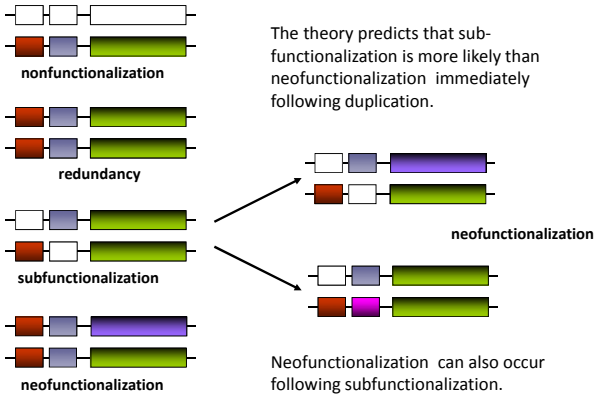
Fates of duplicate genes



The theory predicts that sub-functionalization is more likely than neofunctionalization immediately following duplication.

Neofunctionalization can also occur following subfunctionalization.

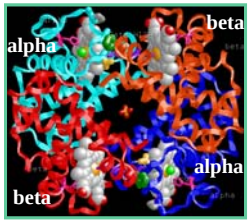
Fates of duplicate genes



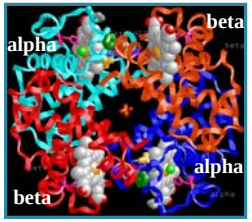
The theory predicts that sub-functionalization is more likely than neofunctionalization immediately following duplication.

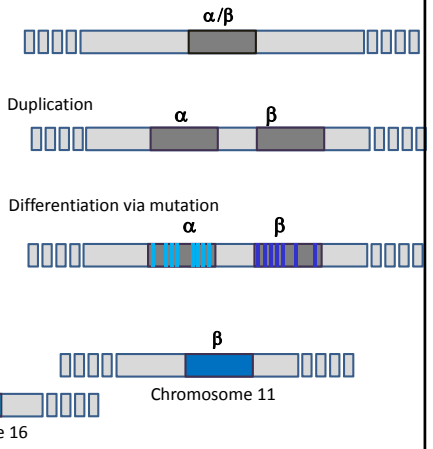
Neofunctionalization can also occur following subfunctionalization.

Example 2: Hemoglobin

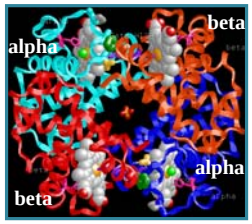


Vertebrate α and β globins arose via duplication of an ancestral globin gene in a vertebrate ancestor



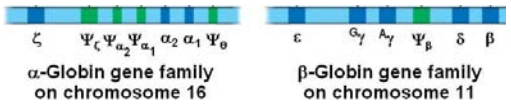


Additional duplication gave rise to the human globin family

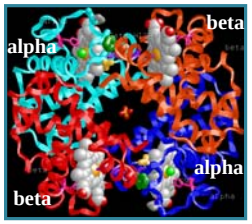


Some copies evolved new functions, others became *pseudogenes*

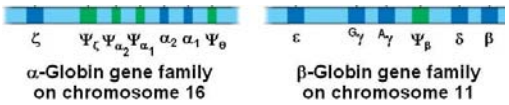
Ψ (pseudogenes):
Sequences that have lost function, but still look enough like genes to be recognized as former genes.

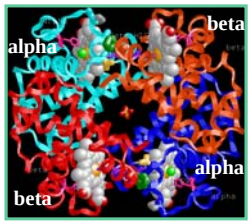


Additional duplication gave rise to the human globin family

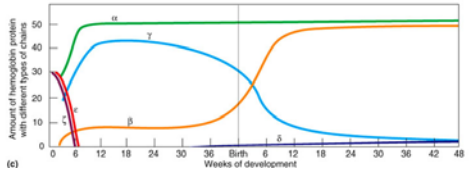


What is the benefit of having several copies of the globins?





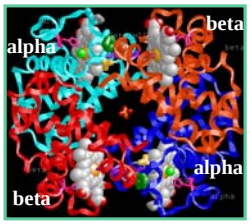
Functional roles of the globins



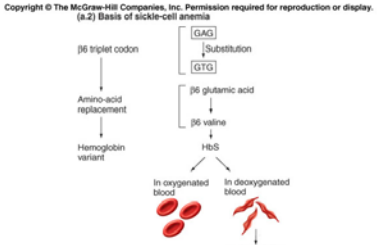
ζε αγ αβ, αδ

0 wks 8 wks birth

Embryo Fetus Adult

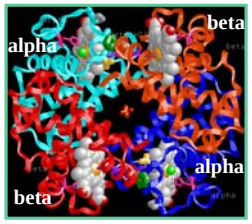


Sickle cell disease

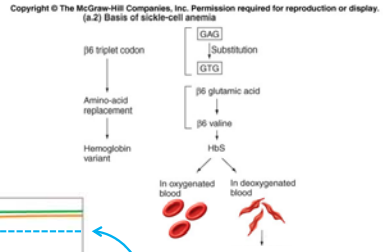


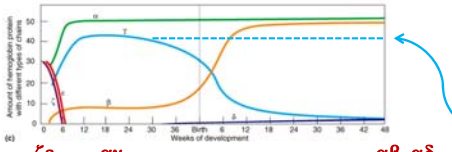
α-Globin gene family on chromosome 16

β-Globin gene family on chromosome 11



But fetal β globin (γ) can partially compensate for the sickle cell mutation in adult fetal β globin!





ζε αγ αβ, αδ

0 wks 8 wks birth

Embryo Fetus Adult

In some individuals, fetal globin continues to be expressed into adulthood