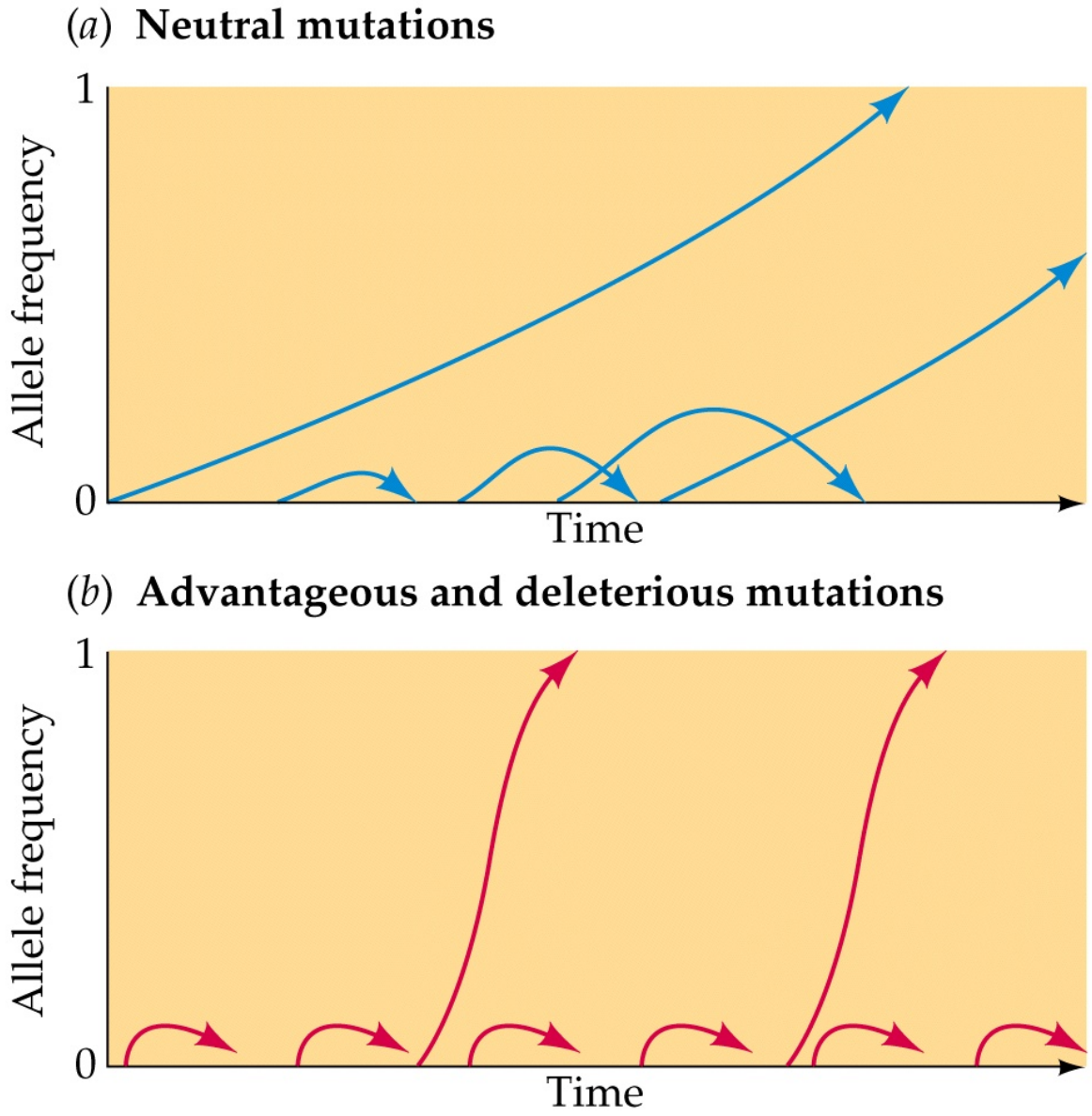
Molecular Evolution & the Origin of Variation

What Is Molecular Evolution?

- **Molecular evolution** differs from phenotypic evolution in that mutations and genetic drift are much more important determinants of molecular evolution.

- The goals of **molecular evolution** studies are to determine patterns of evolutionary change in organisms' molecules, determine the processes that caused the changes, and use those insights to solve other biological problems.

- **Neutral alleles** are fixed slowly, whereas **advantageous and disadvantageous alleles** are fixed rapidly.

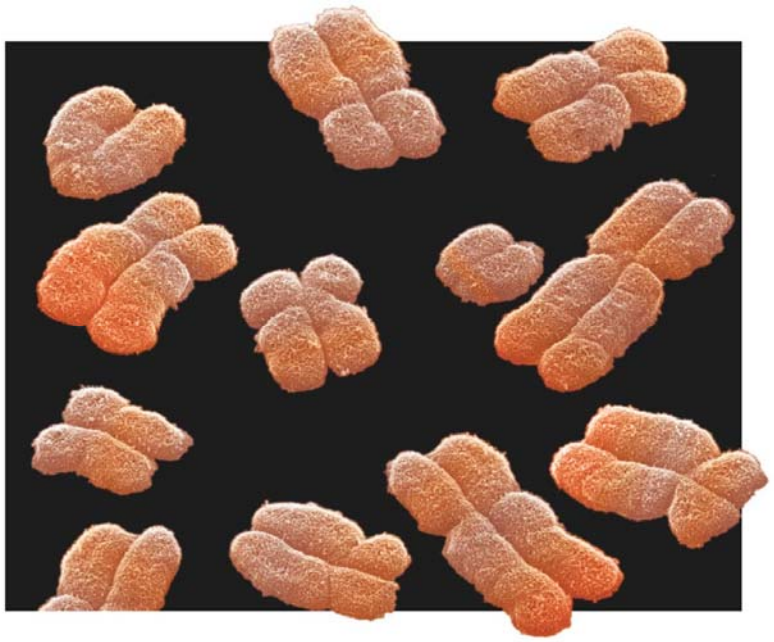

(a) Neutral mutations
1
Allele frequency
0
Time
(b) Advantageous and deleterious mutations
1
Allele frequency
0
Time

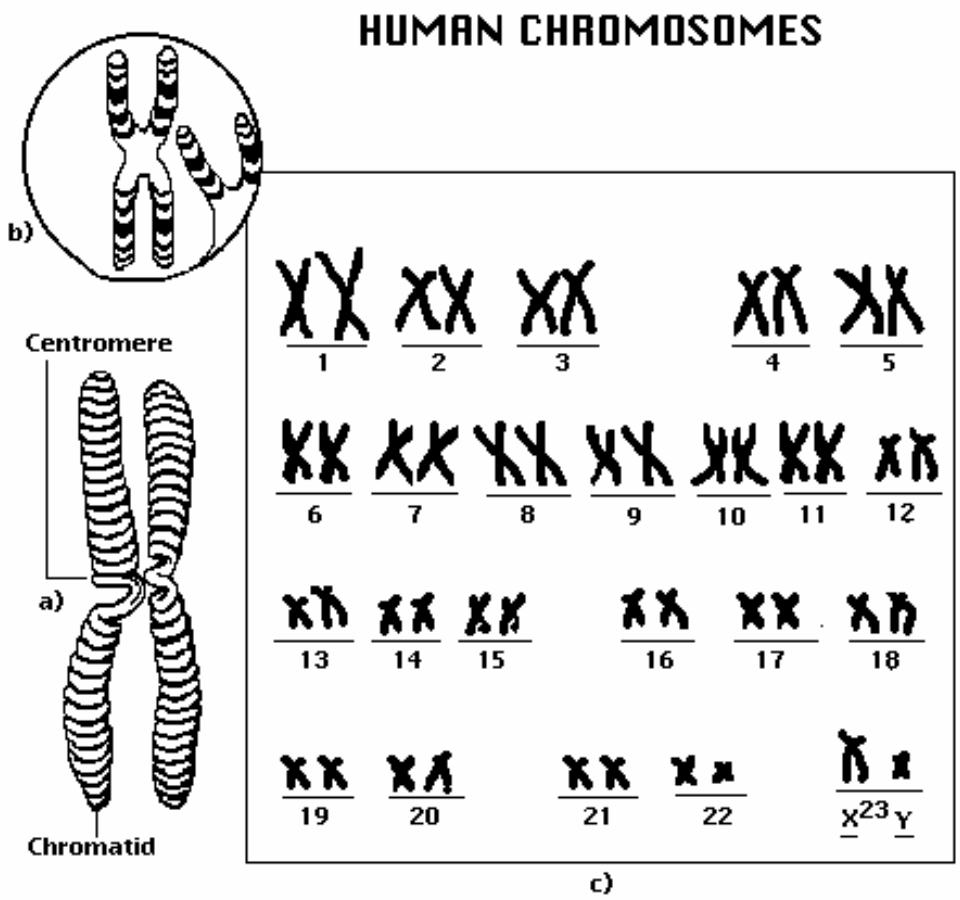
Mechanisms that Act on the Diversity of Genes and Alleles

- Mutation
- Drift (Dominant in Neutral theory)
- Selection (Dominant in Selectionist Theory)

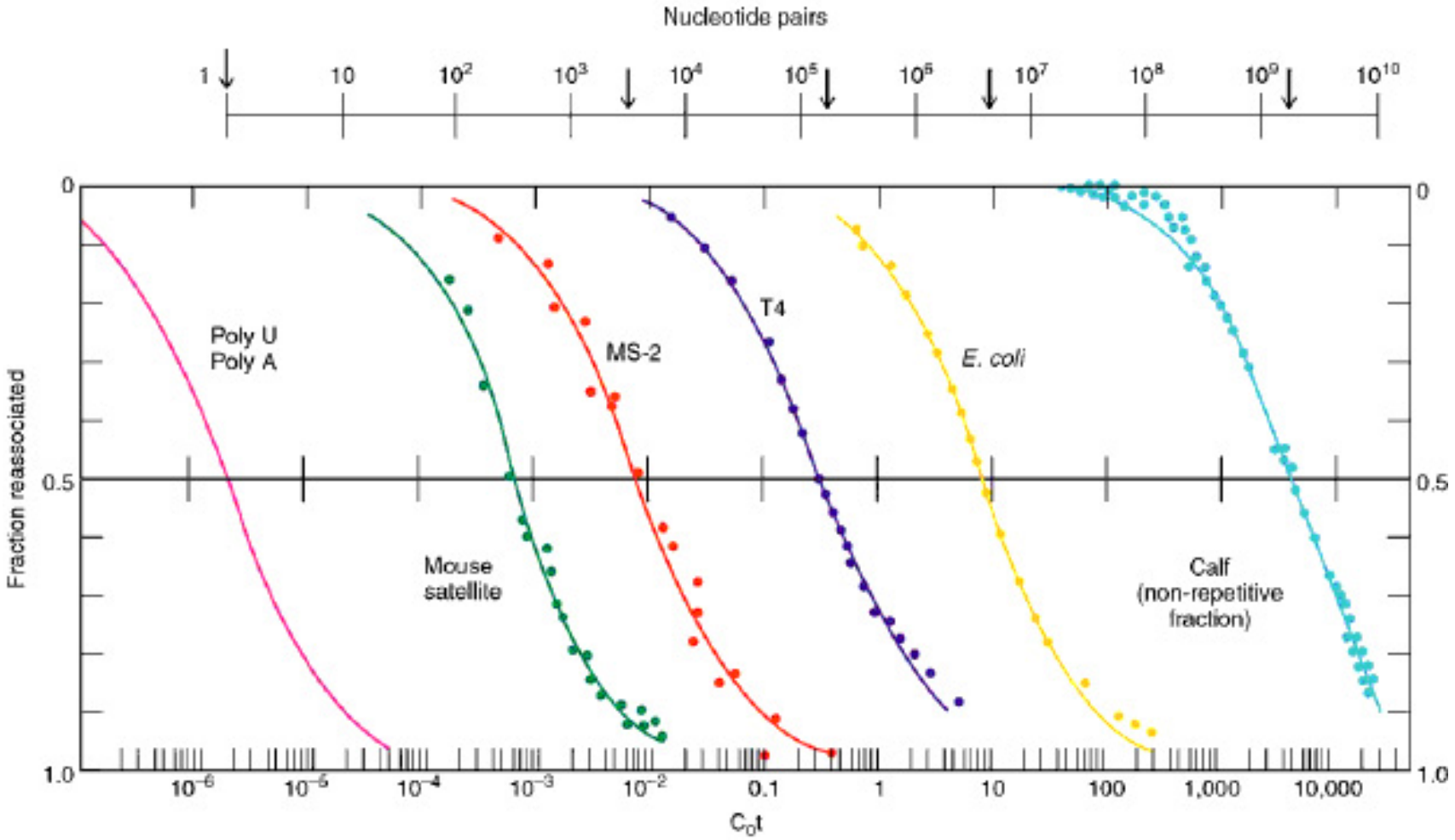
Genome Organization

- $C_o t$ curves – Three levels of structure in Eukaryotes.
- Size does not affect complexity of a Genome: "C-value Paradox."
- Zuckerkandl & Pauling – Clock-like thru time supporting Neutral Theory.

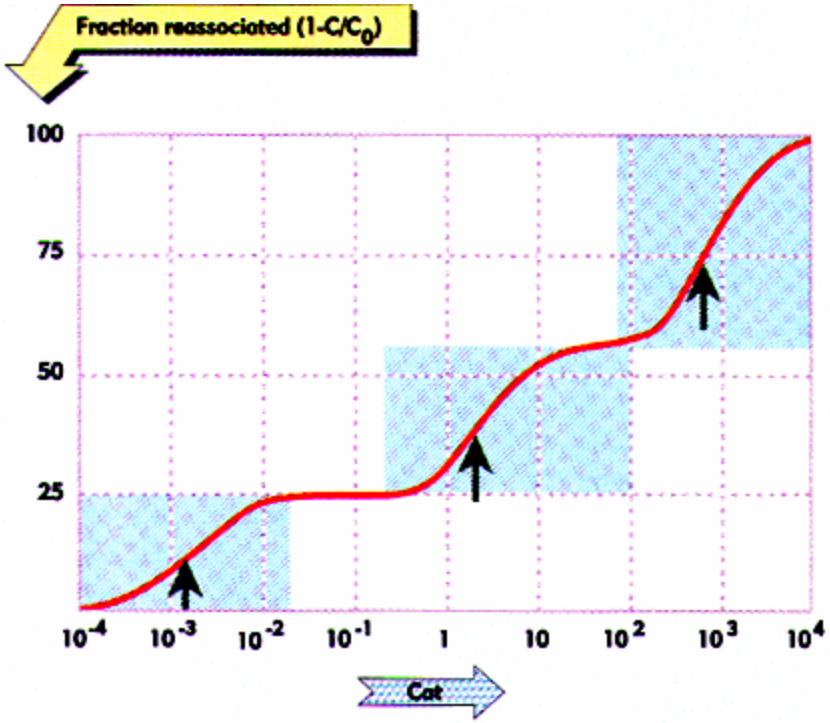




The sum of all the chromosome information is known as a **karyotype** with 22 pairs of **autosomes** and 1 pair of **sex chromosomes**.

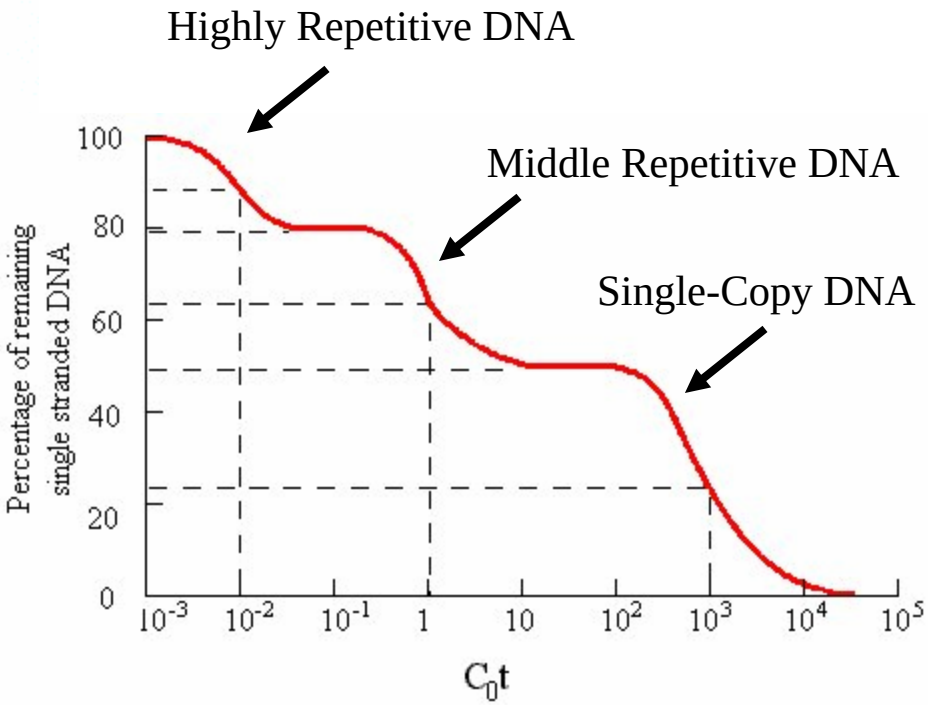


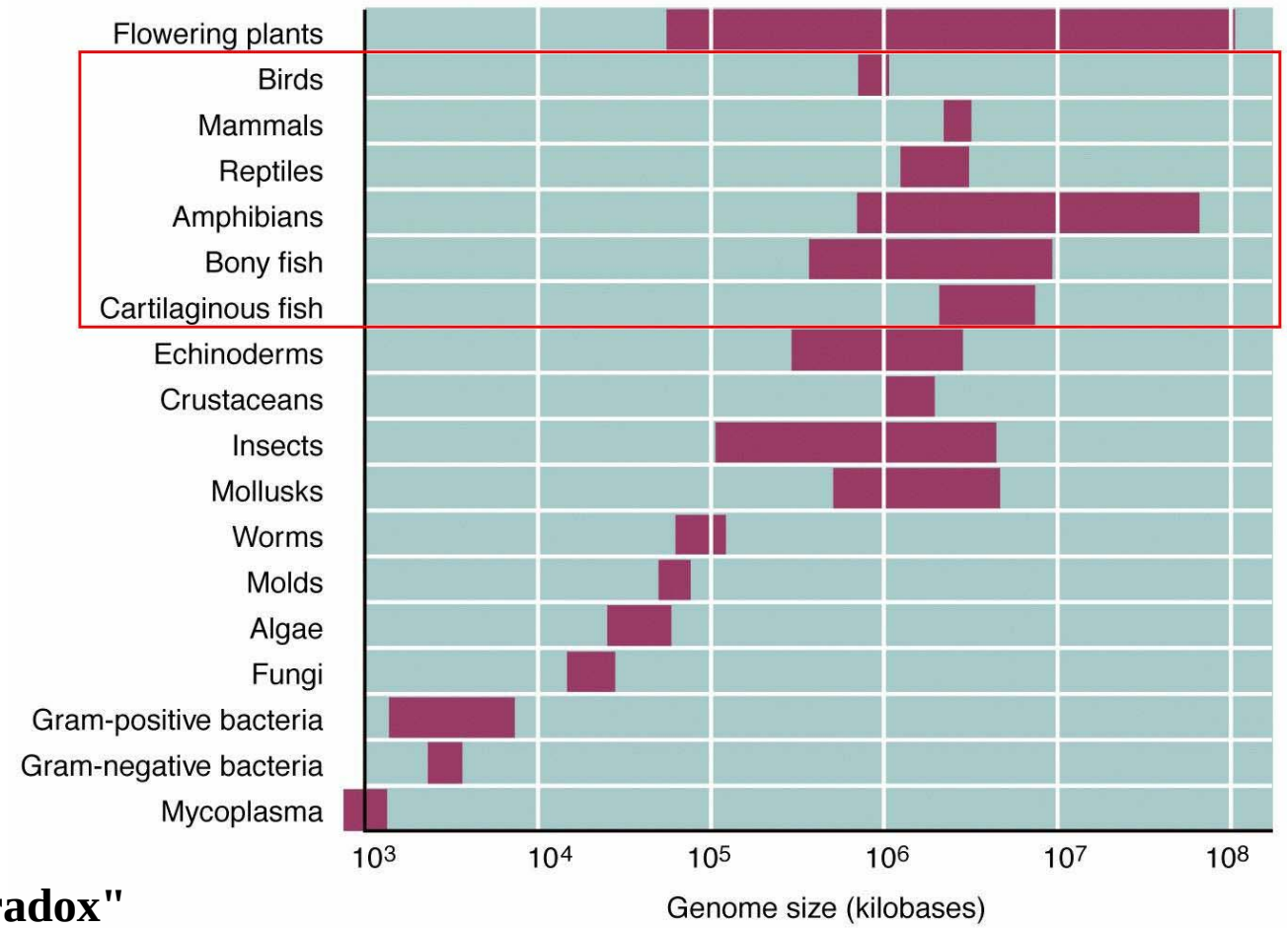
The of reassociation rate of dsDNA from various sources shows how the rate decreases as the complexity of the organism and its genome increases.



	Fast Component	Intermediate Component	Slow Component
Percent of genome	25	30	45
Cot ½	0.0013	1.9	630
Complexity, bp	340	6.0×10^5	3.0×10^8
Repetition frequency	500,000	350	1

If the same experiment is carried out using DNA purified from a complex eukaryote, such as human, then we do not see a simple sigmoidal curve. Instead we see a curve which is the sum of the reannealings of many different components.





The "C-Value Paradox"

This chart shows the range of **C-values** [genome sizes] for a variety of organisms. "Simple" prokaryotic organisms in general have less **DNA** per genome than do more complex, eukaryotic organisms, such as Plants and Animals, and vertebrate animals have more DNA than do invertebrates. The so-called **C-Value Paradox** refers to the observation that C-value does not uniformly increase with respect to perceived complexity of organisms, especially among "higher" vertebrate animals (red box). Note for examples that some Amphibians have more than 10-fold more **DNA** than do Mammals, including humans.

TABLE 8.4 C values from eukaryotic organisms ranked by genome size

Species	C value (Kb)
Saccharomyces cerevisiae (baker's yeast)	12,000
Neurospora crassa (fungus)	17,000
Navicula pelliculosa (pennate diatom)	35,000
Dysidea crawshagi (sponge)	54,000
Caenorhabditis elegans (nematode)	80,000
Chlorella ellipsoide (green alga)	80,000
Ascidia atra (sea squirt)	160,000
Drosophila melanogaster (fruitfly)	180,000
Paramecium aurelia (ciliate)	190,000
Oryza sativa (rice)	590,000
Strongylocentrotus purpuratus (sea urchin)	870,000
Scomber scombrus (mackerel)	950,000
Gallus domesticus (chicken)	1,200,000
Erysiphe cichoracearum (powdery mildew)	1,500,000
Cyprinus carpio (common carp)	1,700,000
Lampetra planeri (brook lamprey)	1,900,000
Boa constrictor (snake)	2,100,000
Parascaris equorum (roundworm)	2,500,000
Carcharias obscurus (sand-tiger shark)	2,700,000
Canis familiaris (dog)	2,900,000
Rattus norvegicus (rat)	2,900,000
Xenopus laevis (African clawed frog)	3,100,000
***Homo sapiens* (human)**	**3,600,000**
Nicotiana tabacum (tobacco plant)	3,800,000
Locusta migratoria (migratory locust)	6,600,000
Spirogyra setiformis (desmid alga)	7,000,000
Paramecium caudatum (ciliate)	8,600,000
Schistocerca gregaria (desert locust)	9,300,000
Allium cepa (onion)	15,000,000
Triturus cristatus (warty newt)	19,000,000
Thuja occidentalis (western giant cedar)	19,000,000
Coscinodiscus asteromphalus (centric diatom)	25,000,000
Lilium formosanum (lily)	36,000,000
Amphiuma means (two-toed salamander)	84,000,000
Pinus resinosa (Canadian red pine)	68,000,000
Lepidosiren paradoxa (South American lungfish)	120,000,000
Protopterus aethiopicus (marbled lungfish)	140,000,000
Ophioglossum petiolatum (adder's tongue fern)	160,000,000
Amoeba proteus (amoeba)	290,000,000
Amoeba dubia (amoeba)[a]	690,000,000

Data from Sparrow et al. (1972), Cavalier-Smith (1985), and many other sources.

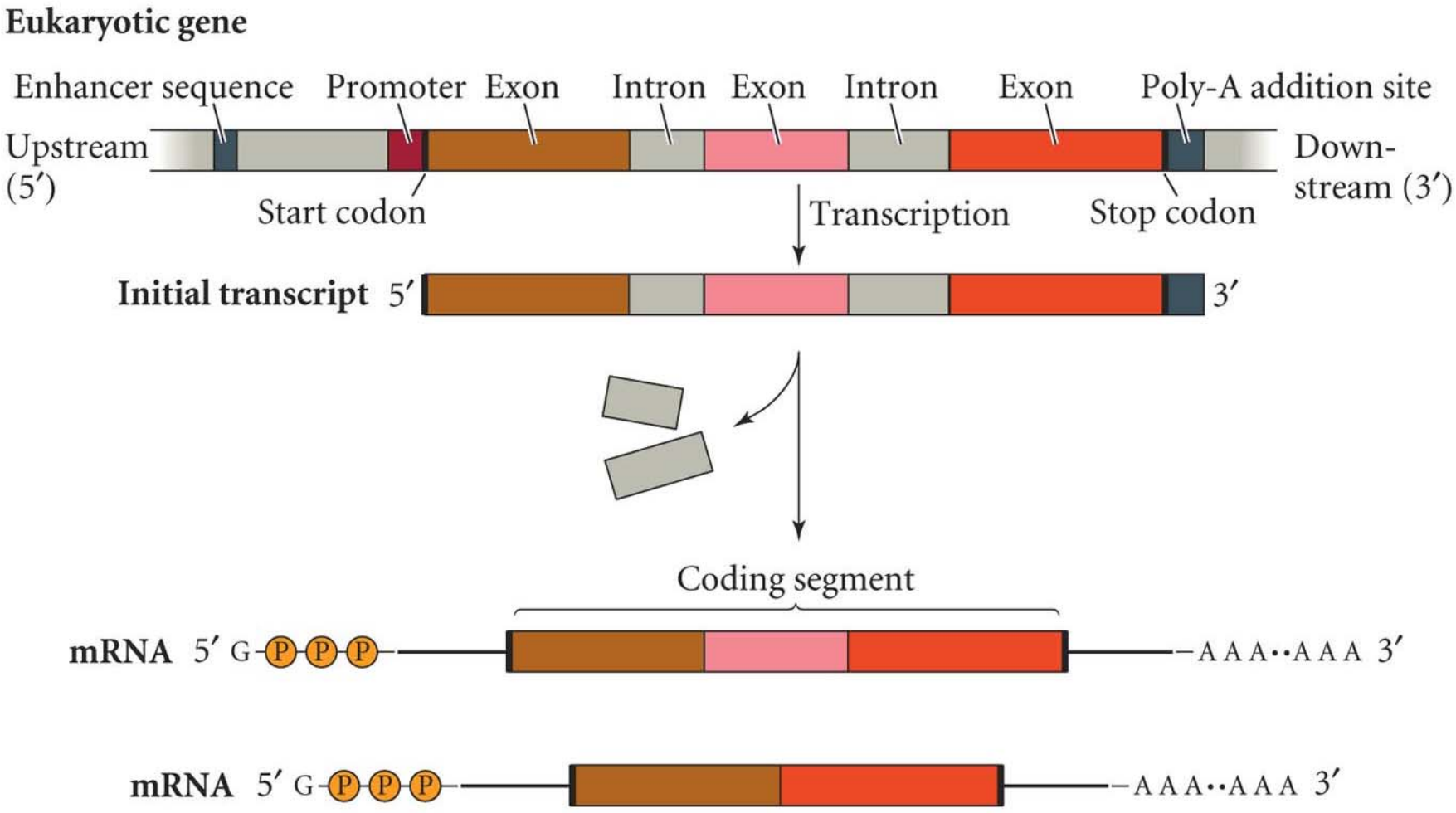
[a]The ploidy of the sarcodine amoeba *Chaos chaos* is not known, but it is highly probable that its C value is even higher than that of *Amoeba dubia* (Sparrow et al. 1972).

The "C-Value Paradox"

There is in fact **no** "paradox." Evolution does not proceed in a linear manner, nor is there a linear succession of organisms from "lower" to "higher."

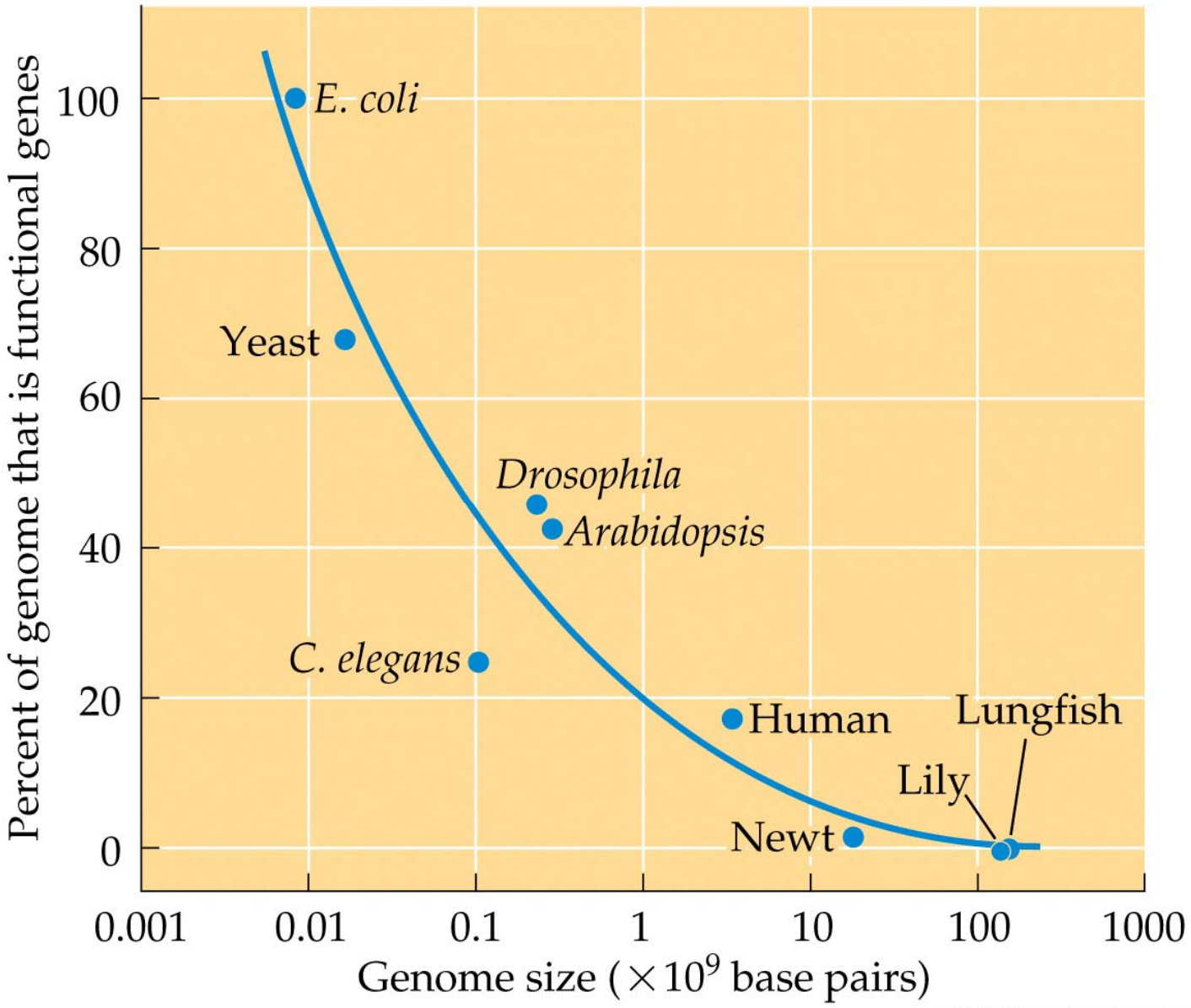
~200X

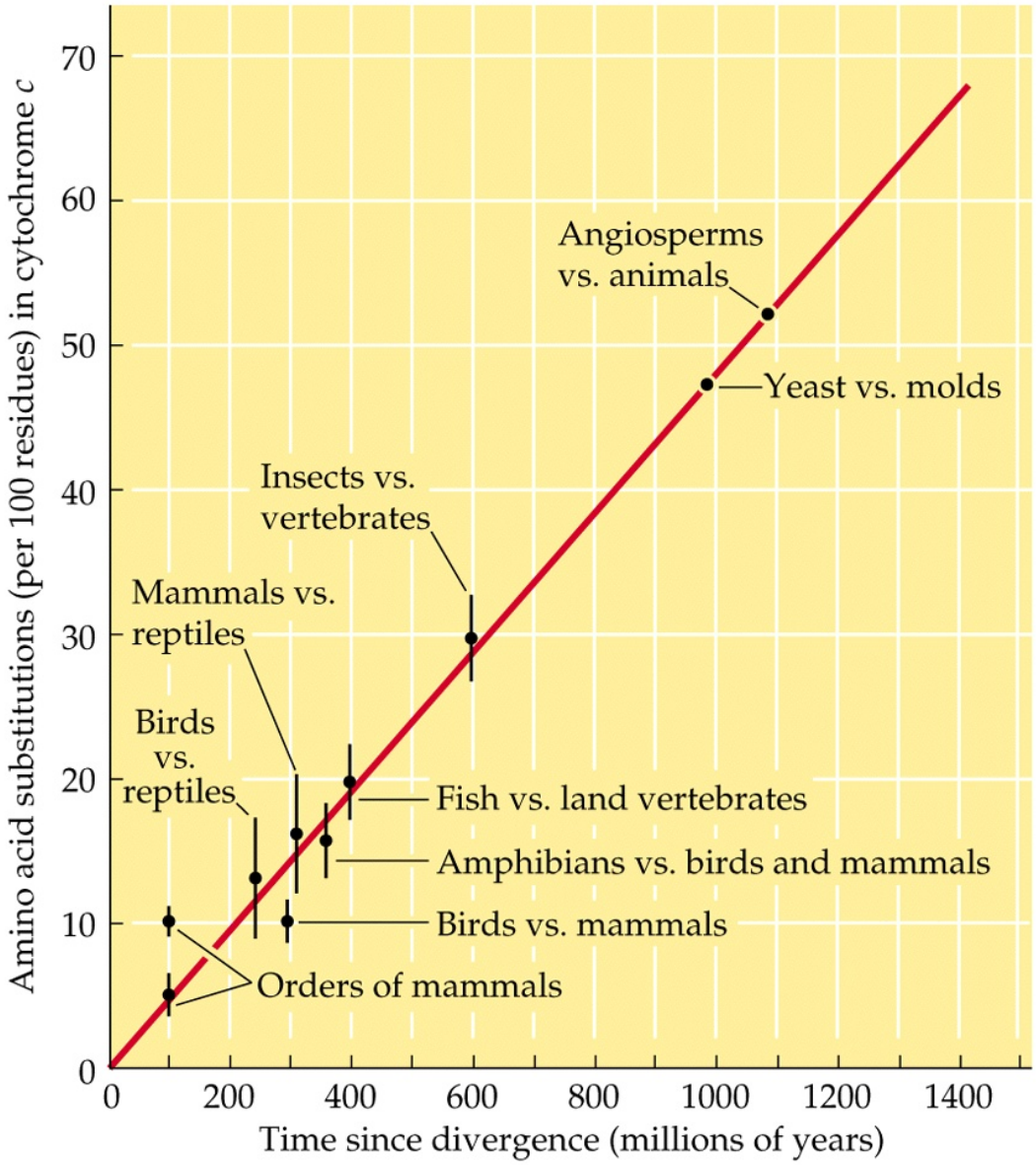
Diagram of a eukaryotic gene, its initial transcript, and the mature mRNA transcript



Typical Gene, of which we have only ~20K

Drake's Rule

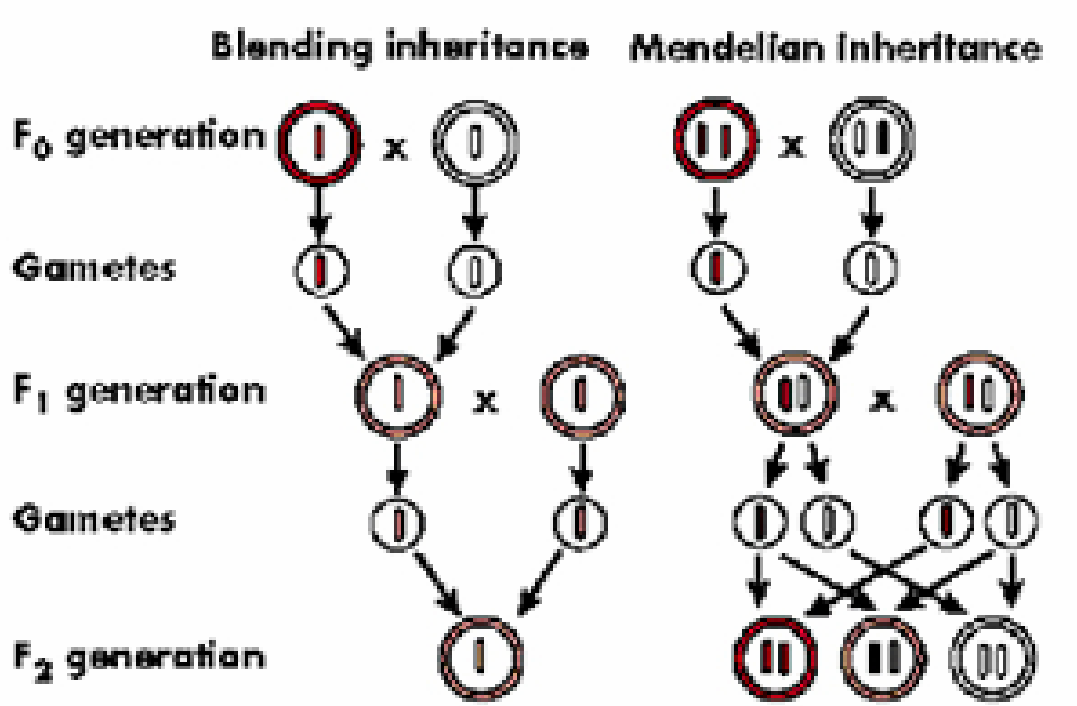

Percent of genome that is functional genes
100
80
60
40
20
0
E. coli
Yeast
Drosophila
Arabidopsis
C. elegans
Human
Lungfish
Lily
Newt
0.001
0.01
0.1
1
10
100
1000
Genome size (×10^9 base pairs)



Rates of amino acid substitutions in some molecules are relatively constant over evolutionary time.

Source of New Genes and Alleles

- Old view: Inheritance of acquired characters.
- New view: Mutation is ultimate source of all variation.
- Rem: Mutations in somatic vs. germ line cells.


Blending inheritance
Mendelian inheritance
F0 generation
x
x
Gametes
Gametes
F1 generation
x
x
F2 generation

Types of Genetic Change

- **Point mutations –** molecular scale (source of new alleles)
  - Base substitutions: transitions vs. transversions
  - Replacement (non-synonymous) vs. silent substitutions (synonymous)
  - Insertions and deletions may cause frameshift mutations
- **Chromosome Rearrangements –** macro-molecular scale (tighter linkage as heterozygotes cannot recombine)
- **Gene Duplications –** safety in numbers (unequal crossing over during meiosis)
- **Polyploidization –** change in chromosomal numbers (possible new species)

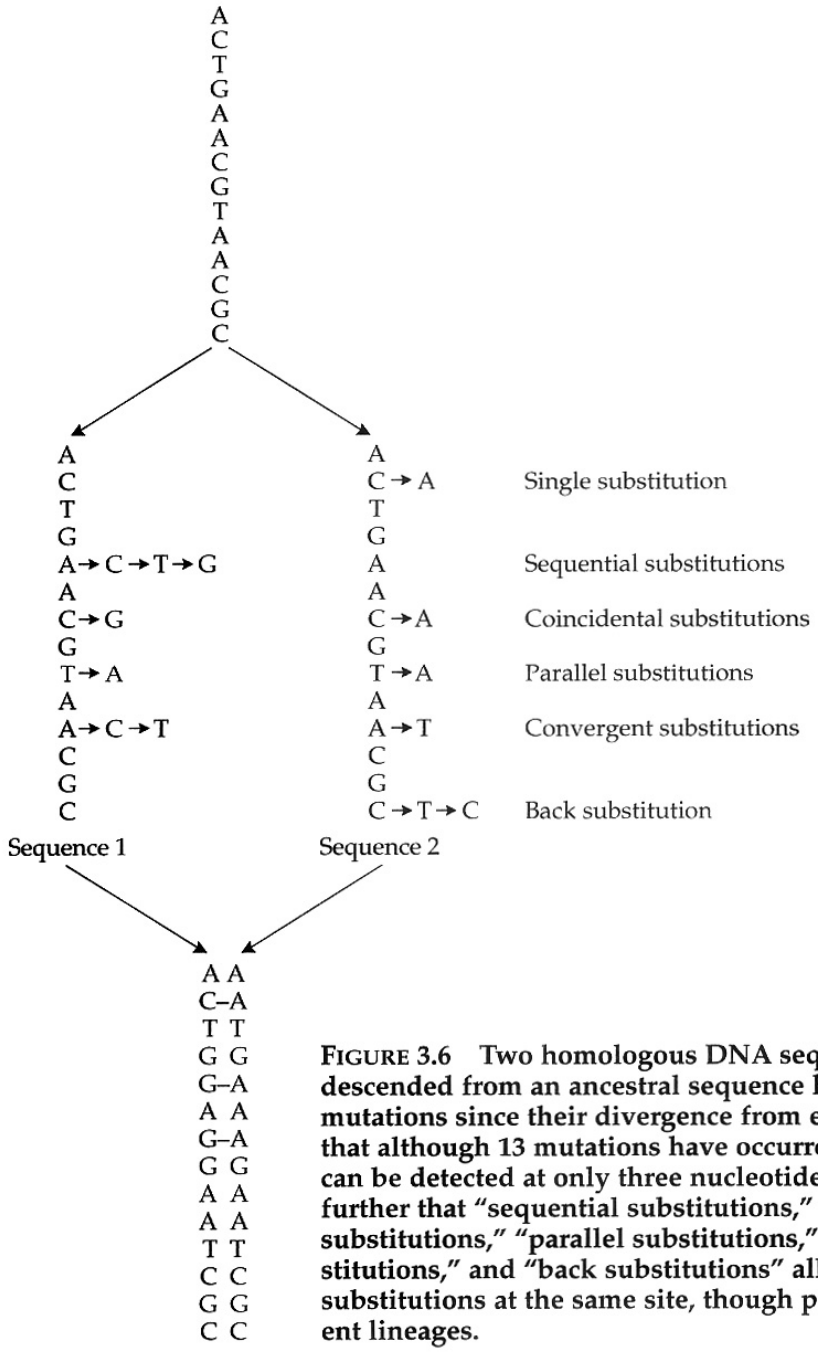


FIGURE 3.6 Two homologous DNA sequences that descended from an ancestral sequence have accumulated mutations since their divergence from each other. Note that although 13 mutations have occurred, differences can be detected at only three nucleotide sites (–). Note further that "sequential substitutions," "coincidental substitutions," "parallel substitutions," "convergent substitutions," and "back substitutions" all involve multiple substitutions at the same site, though perhaps in different lineages.

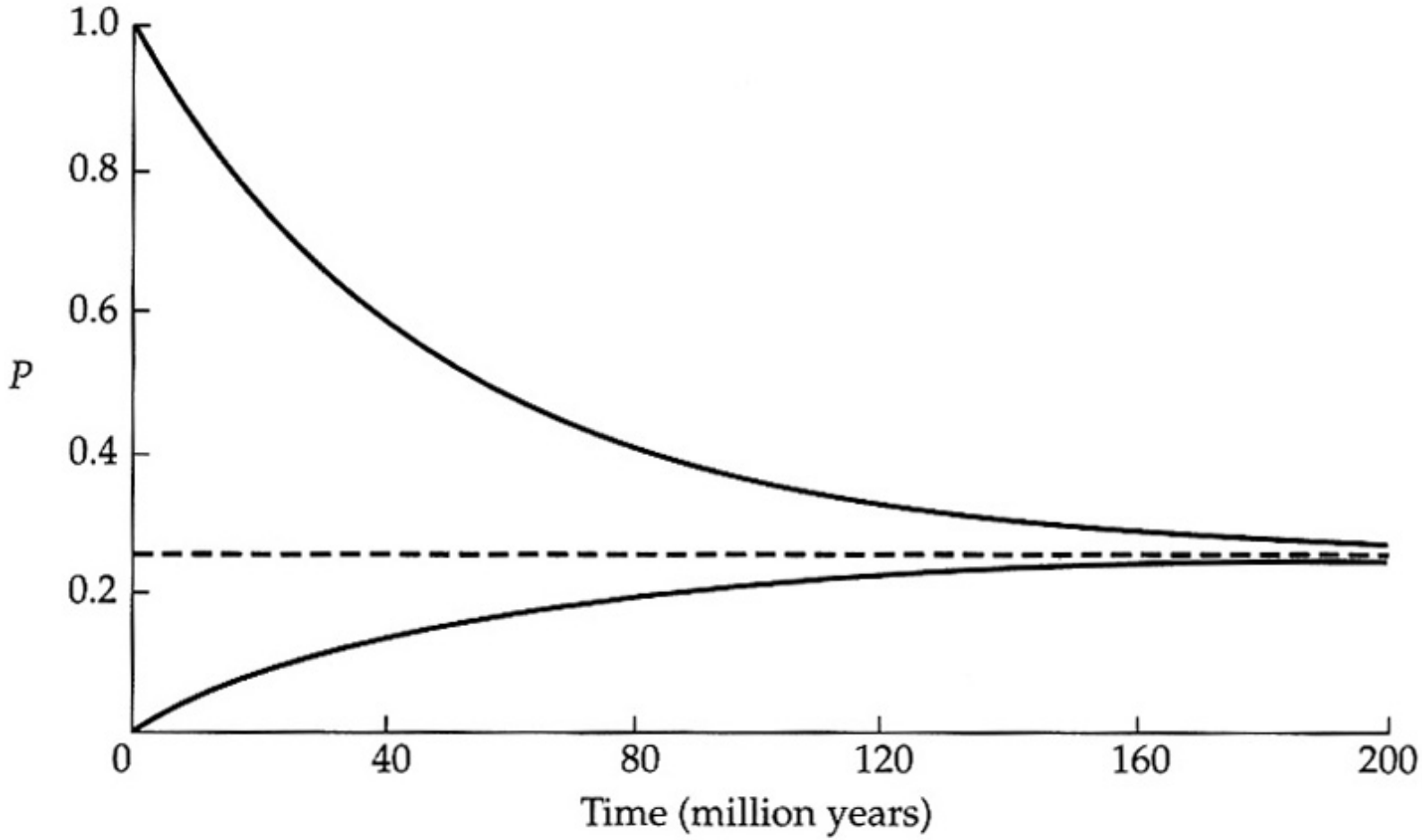


FIGURE 3.3 Temporal changes in the probability, P, of having a certain nucleotide at a position starting with either the same nucleotide (upper line) or with a different nucleotide (lower line). The dashed line denotes the equilibrium frequency ($P = 0.25$). $\alpha = 5 \times 10^{-9}$ substitutions per site per year.

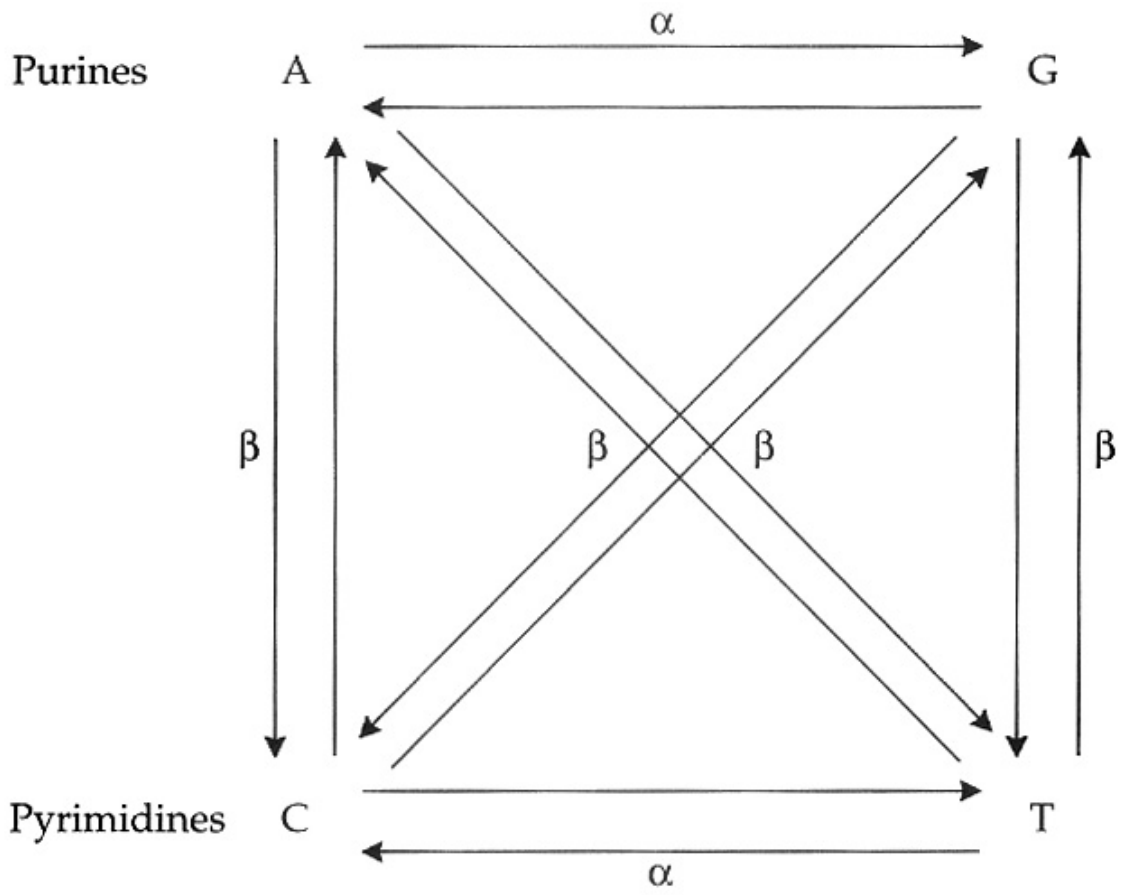
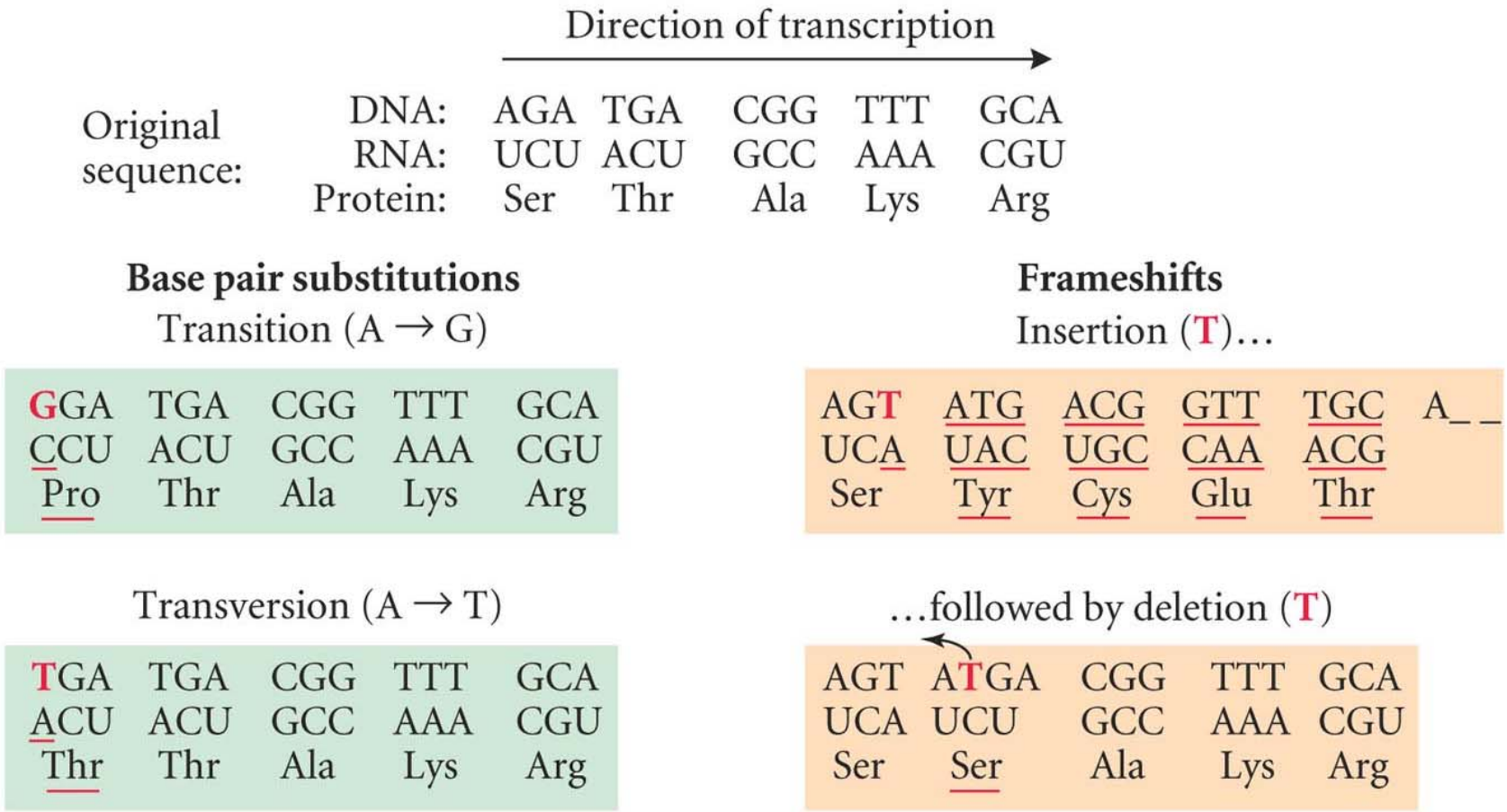


FIGURE 3.4 Two-parameter model of nucleotide substitution. The rate of transition (α) may not be equal to the rate of transversion (β).

Examples of **point mutations** and consequences for mRNA & amino acid sequences



Mutation Rates (rare for most part)

TABLE 8.3 *Estimates of spontaneous mutation rates per base pair and per genome*

Organism	Base pairs		Mutation rate			
	in haploid genome	in effective genome[a]	per base pair per replication	per replication per haploid genome	per replication per effective genome[a]	per sexual generation per effective genome[b]
T2, T4 phage	1.7×10^5	—	2.4×10^{-8}	0.0040	—	—
Escherichia coli	4.6×10^6	—	5.4×10^{-10}	0.0025	—	—
Saccharomyces cerevisiae (yeast)	1.2×10^7	—	2.2×10^{-10}	0.0027	—	—
Neurospora crassa (bread mold)	4.2×10^7	—	7.2×10^{-11}	0.0030	—	—
Caenorhabditis elegans	8.0×10^7	1.8×10^7	2.3×10^{-10}	0.018	0.004	0.036
Drosophila melanogaster	1.7×10^8	1.6×10^7	3.4×10^{-10}	0.058	0.005	0.14
Mouse	2.7×10^9	8.0×10^7	1.8×10^{-10}	0.49	0.014	0.9
Human	3.2×10^9	8.0×10^7	5.0×10^{-11}	0.16	0.004	1.6

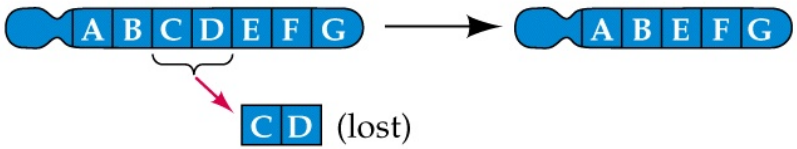
Source: After Drake et al. 1998.

[a] The effective genome is the number of base pairs in functional sequences that could potentially undergo mutations that reduce fitness.

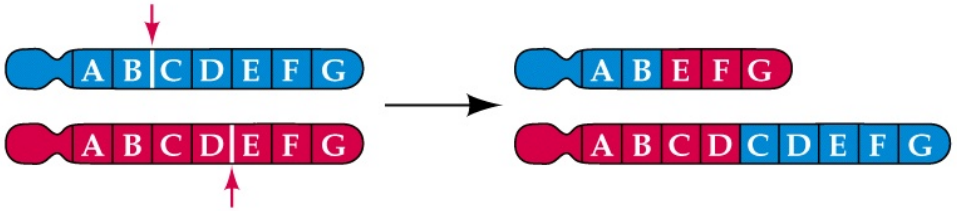
[b] Calculated for multicellular organisms in which multiple DNA replication events occur in development between zygote and gametogenesis.

Chromosome Rearrangements

(*a*) **Deletion**

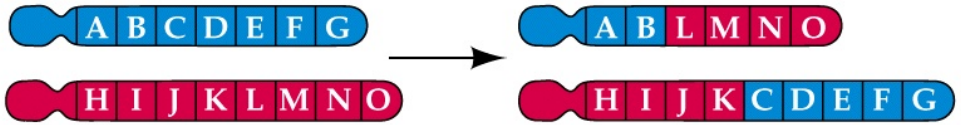


(*b*) **Duplication and deletion**

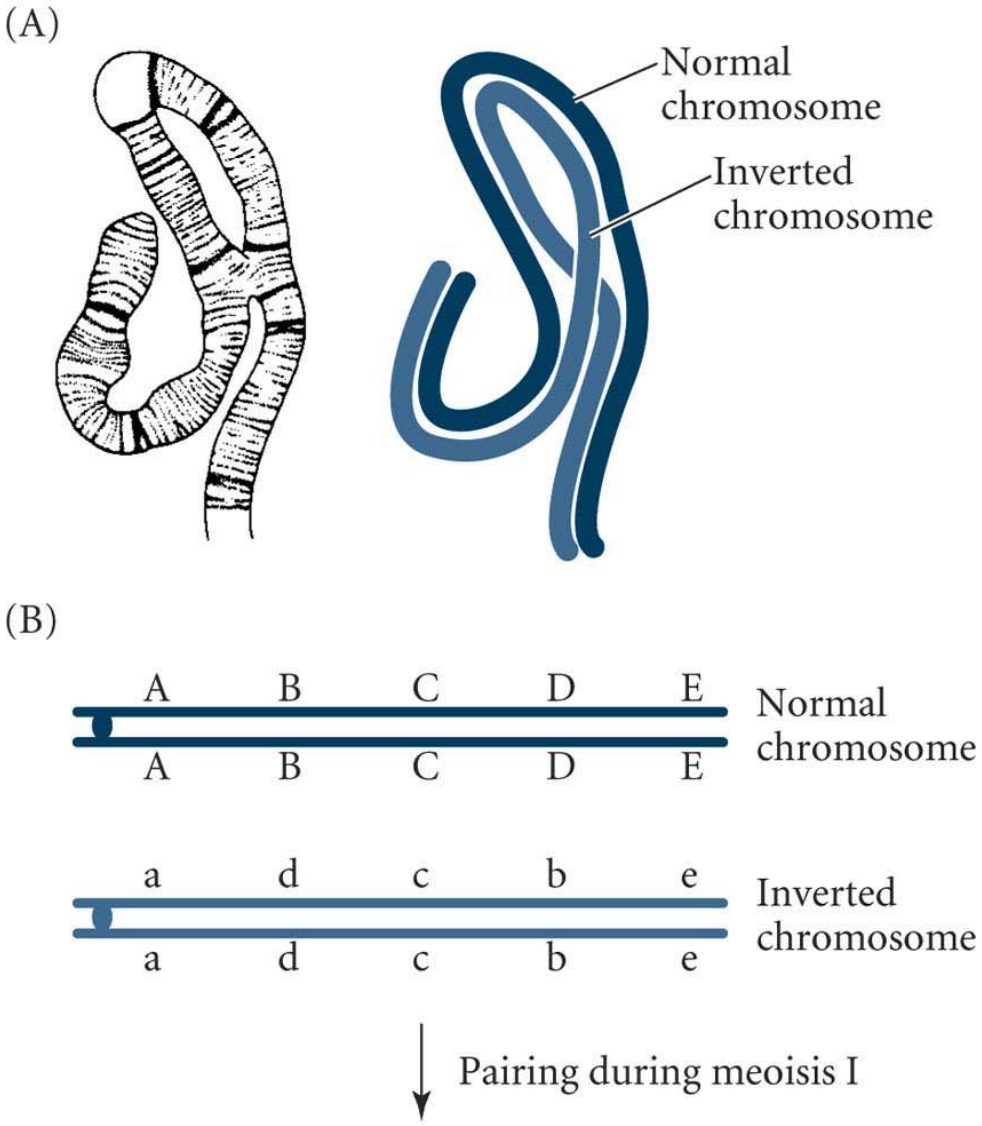


(*c*) **Inversion**

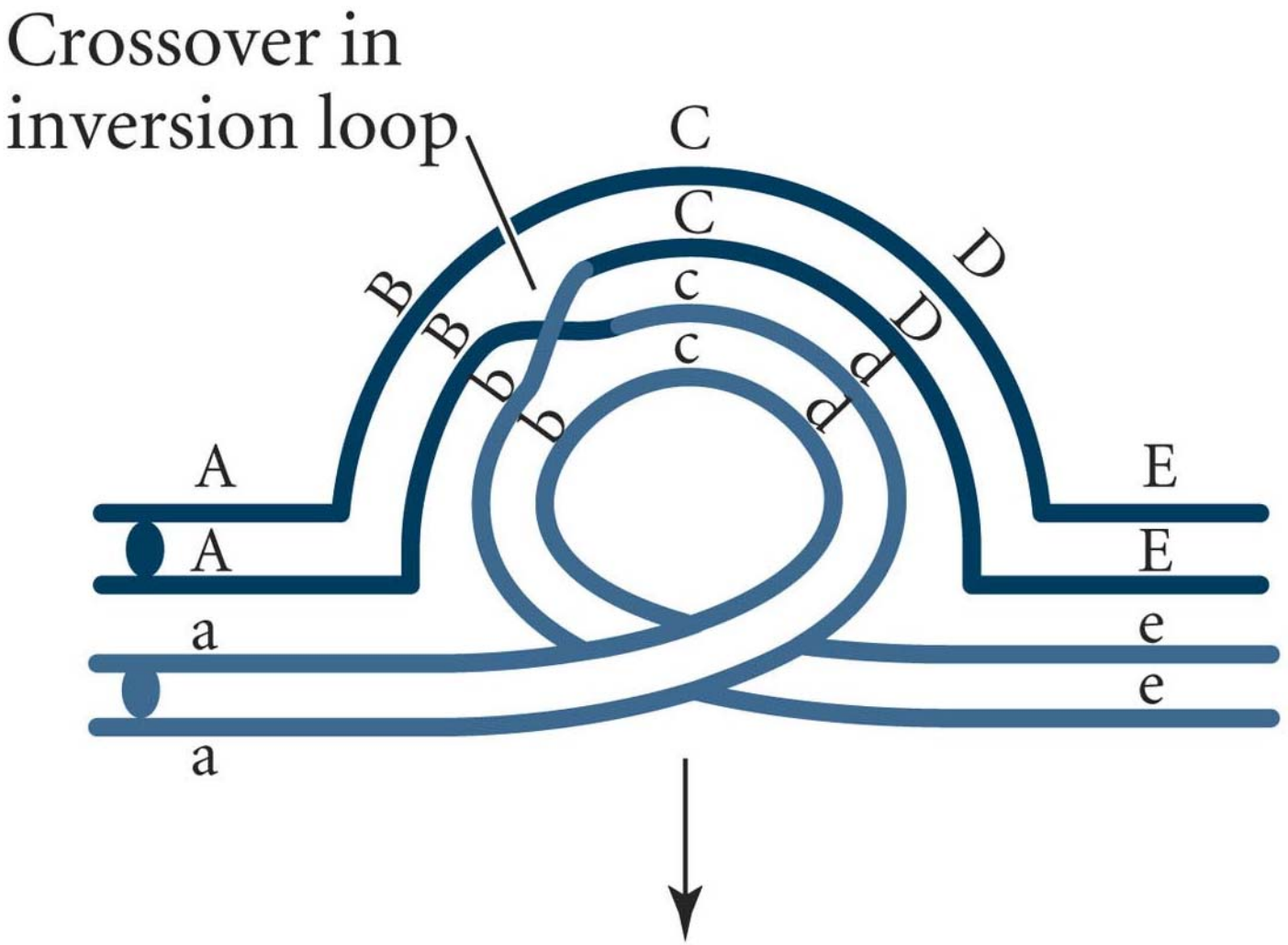
(*d*) **Reciprocal translocation**



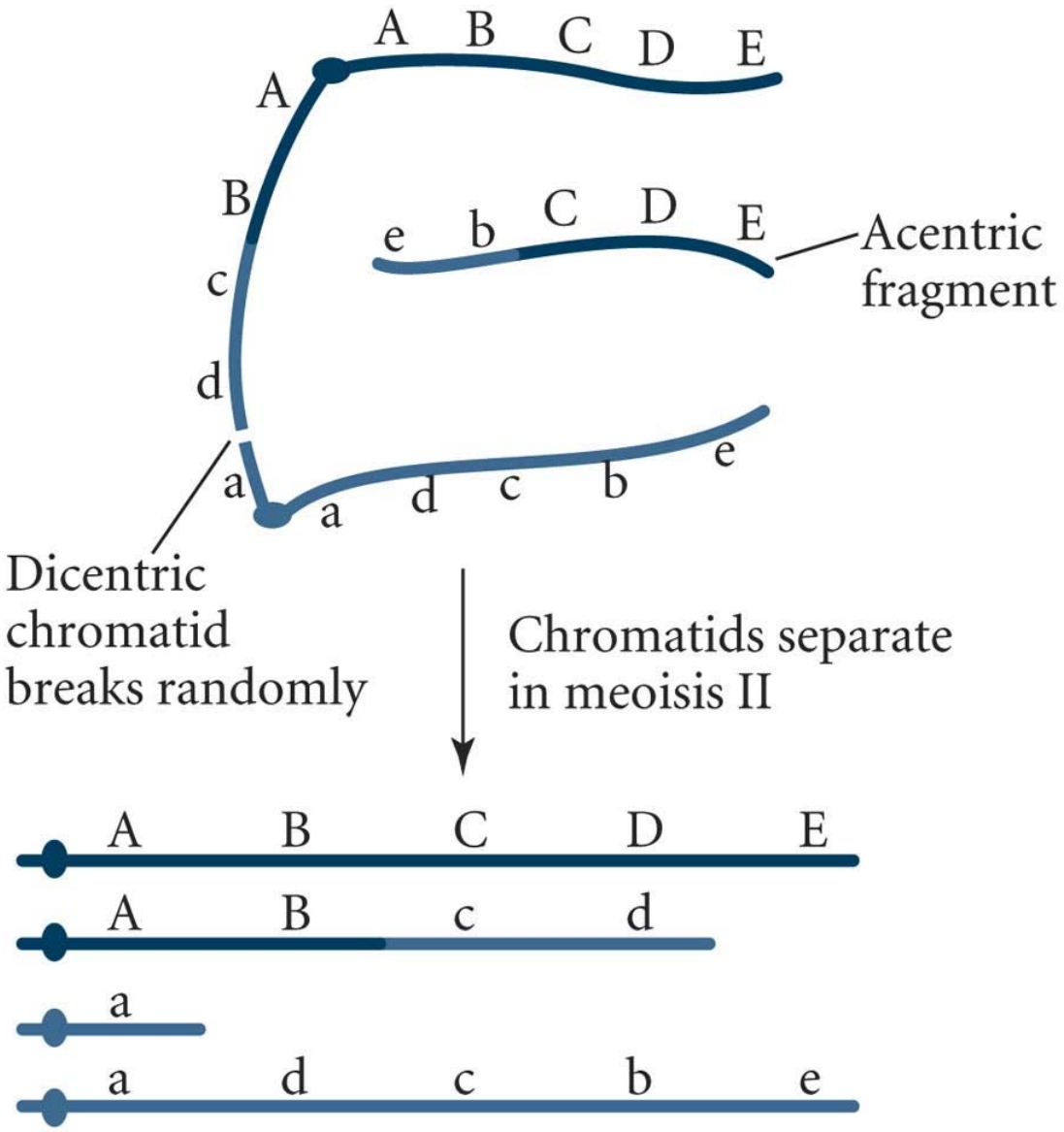
Chromosome Inversion



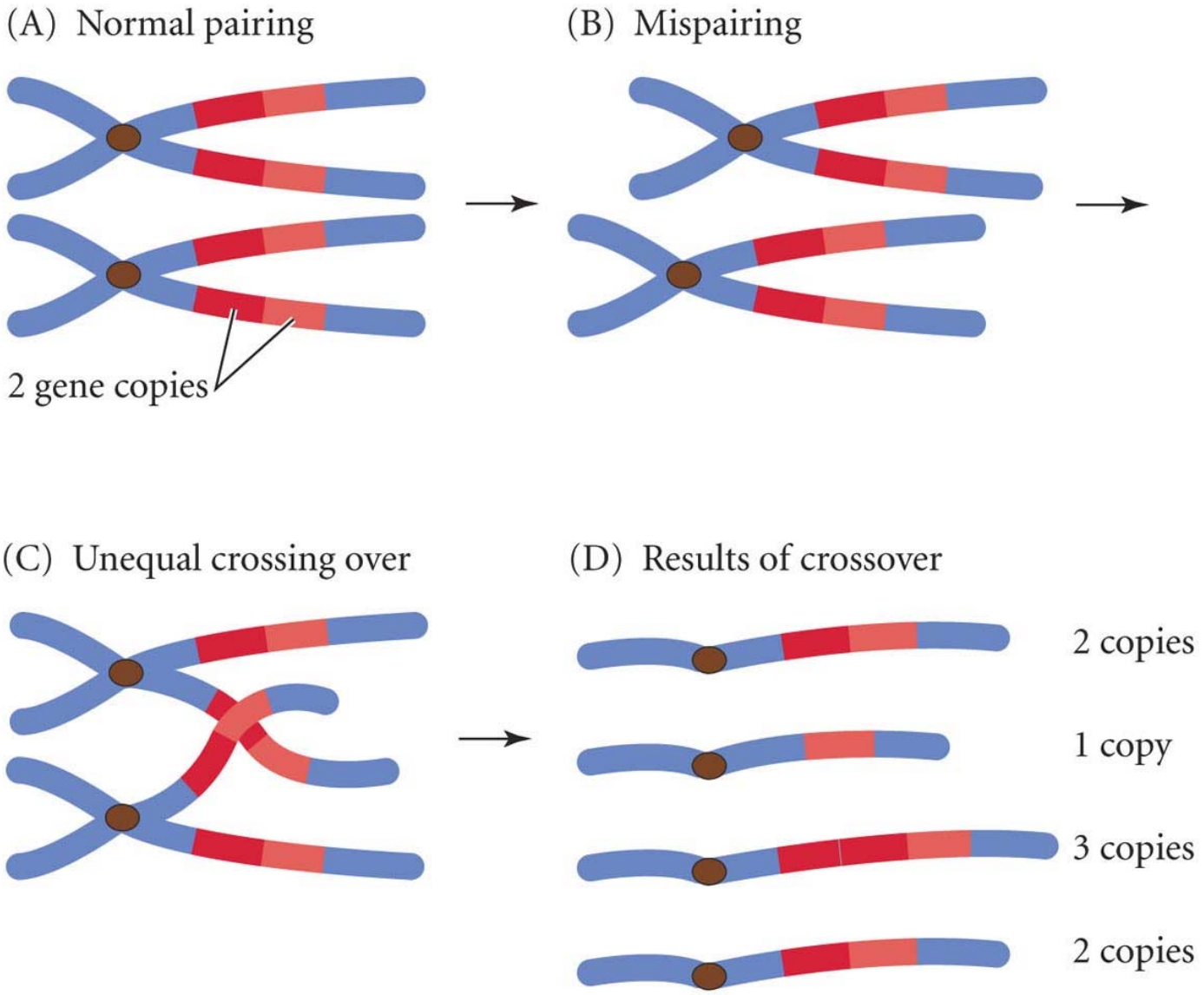
Chromosome Inversion



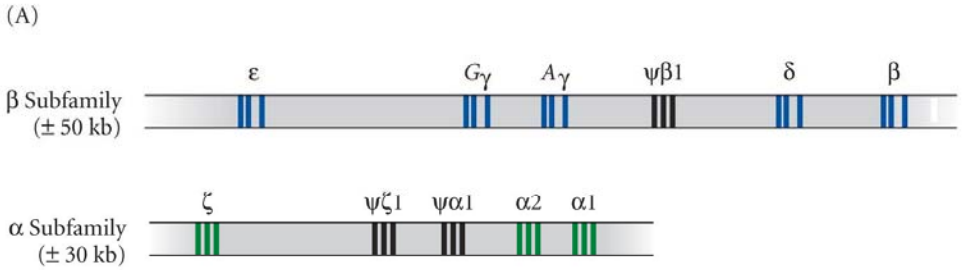
Chromosome Inversion

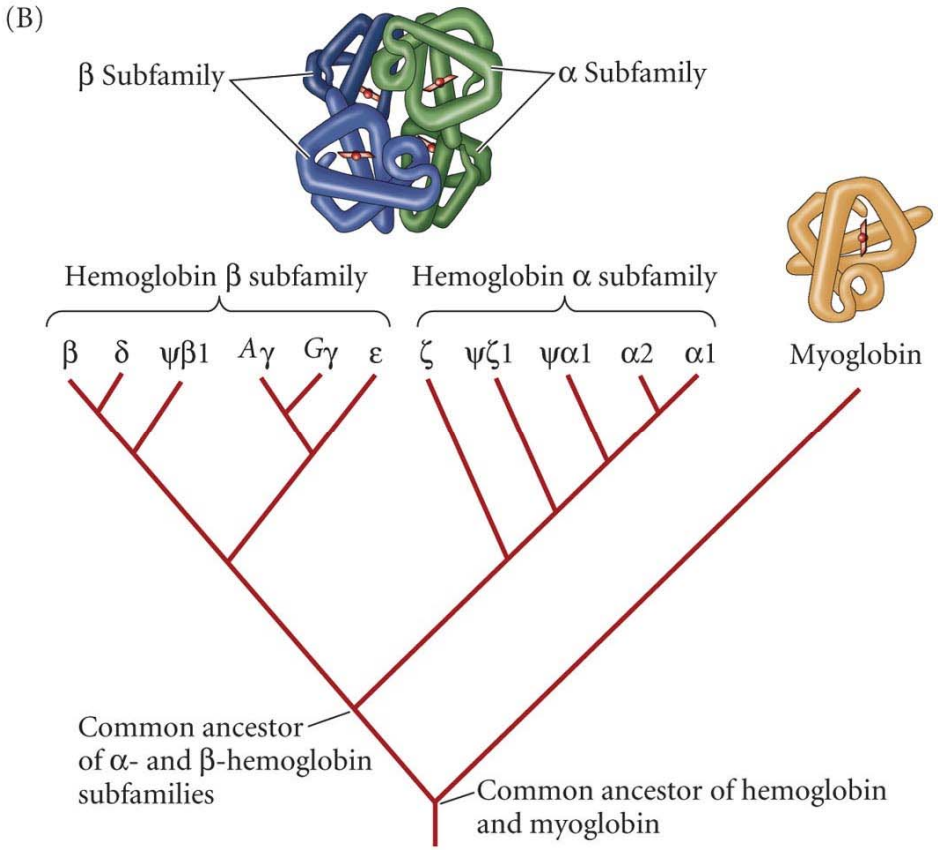


Gene Duplication: Unequal Crossing Over

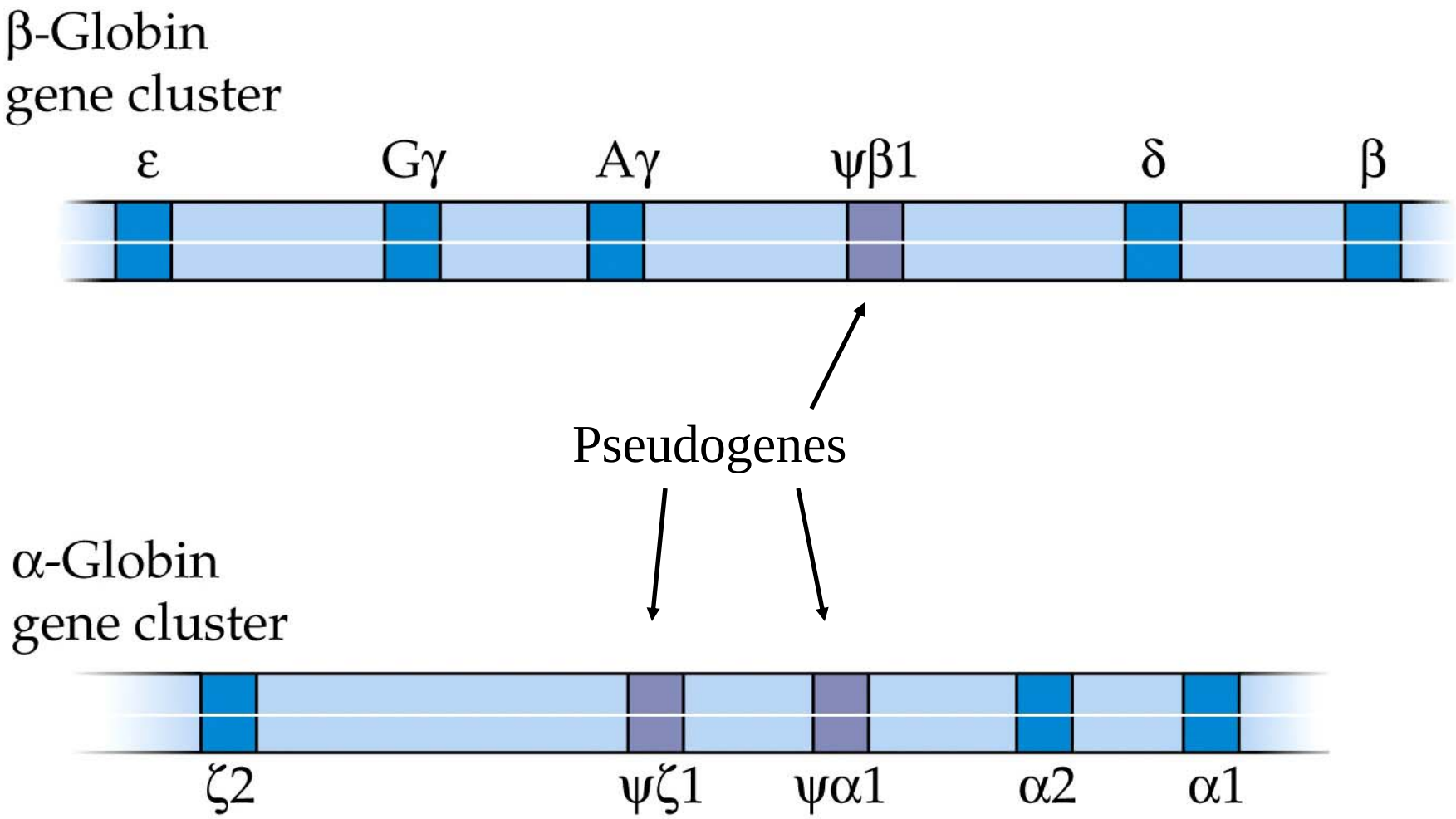


Genotypic variation - gene duplication

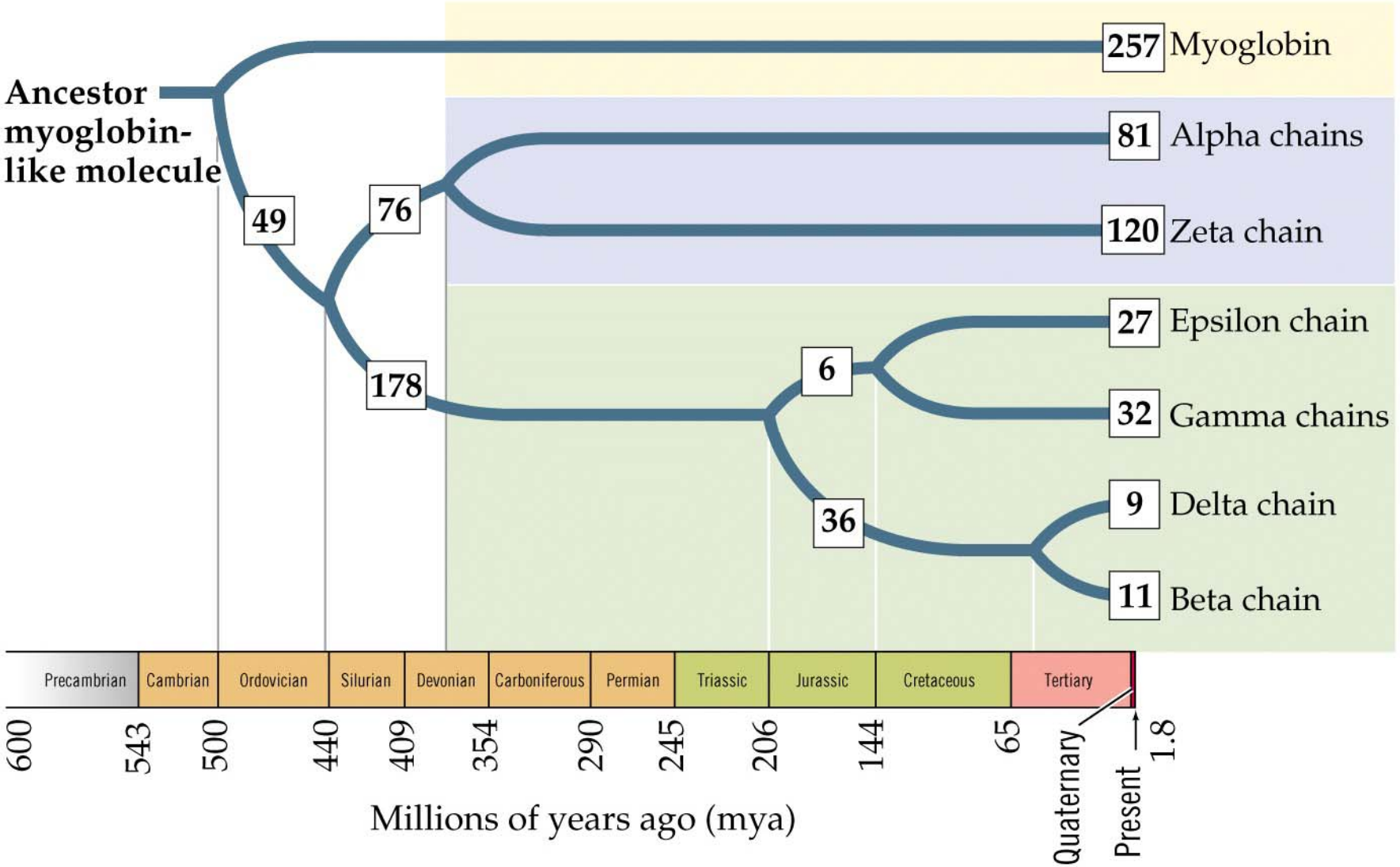




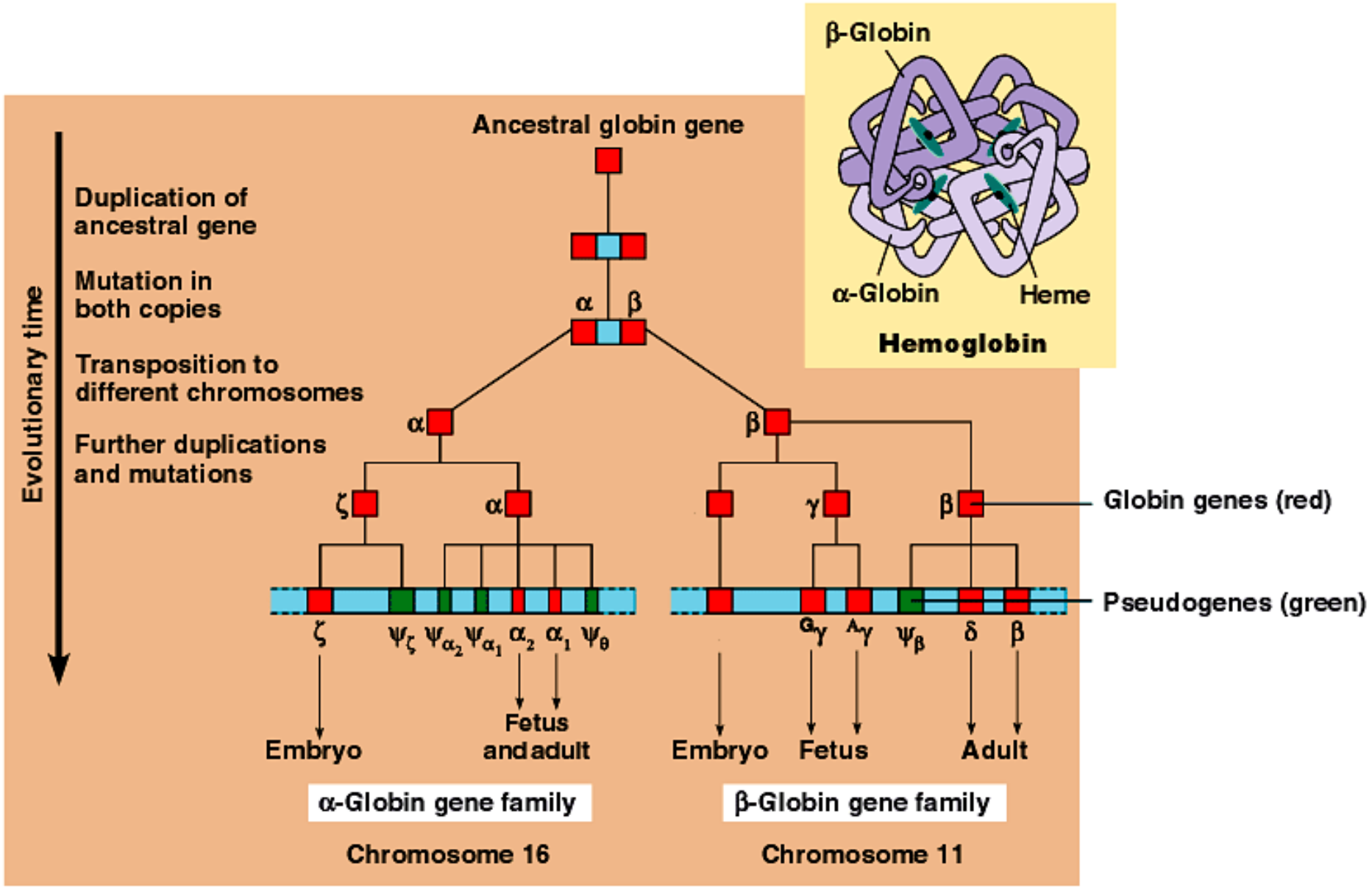
Gene Families



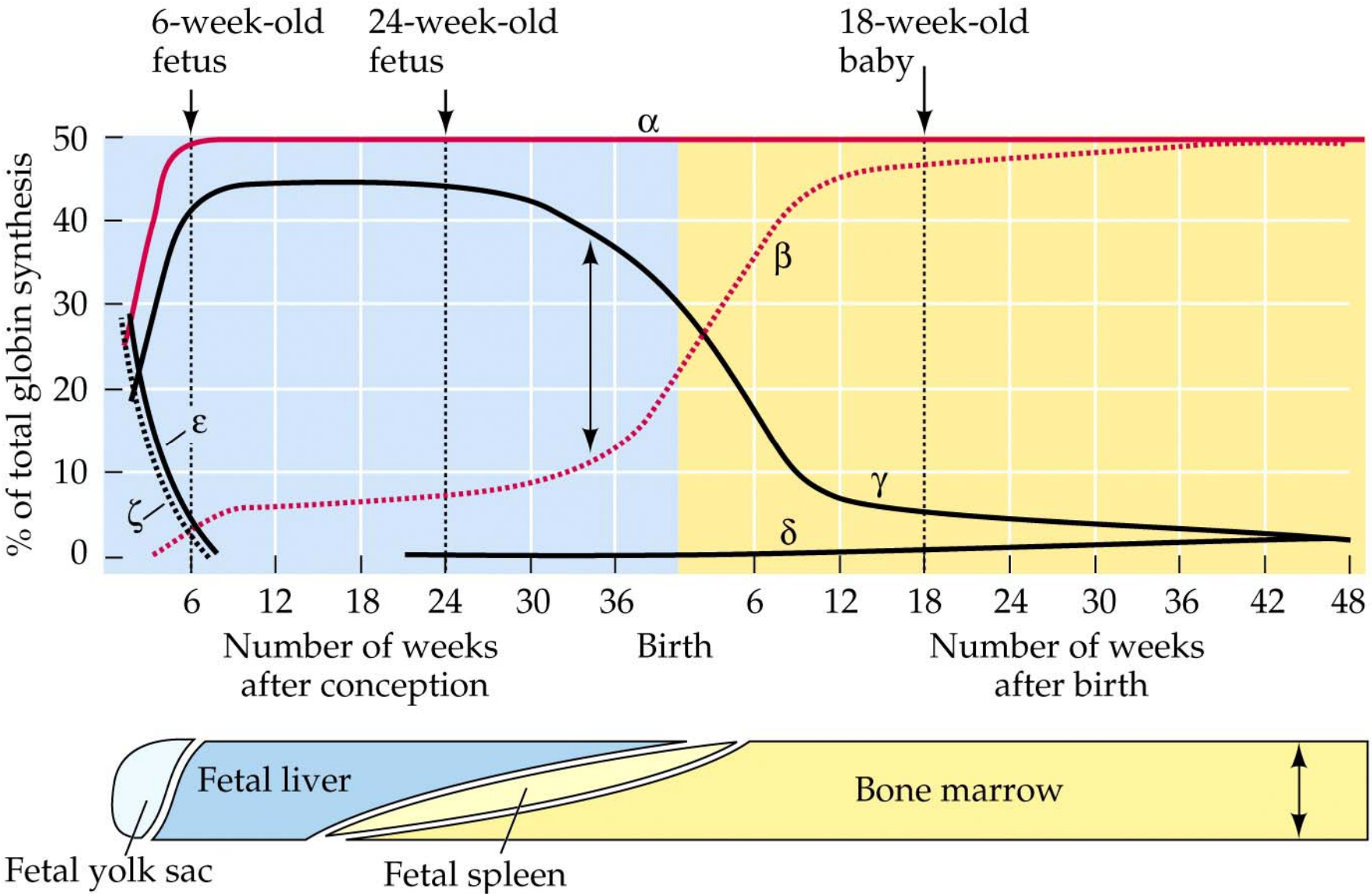
Evolution of α-globin and β-globin



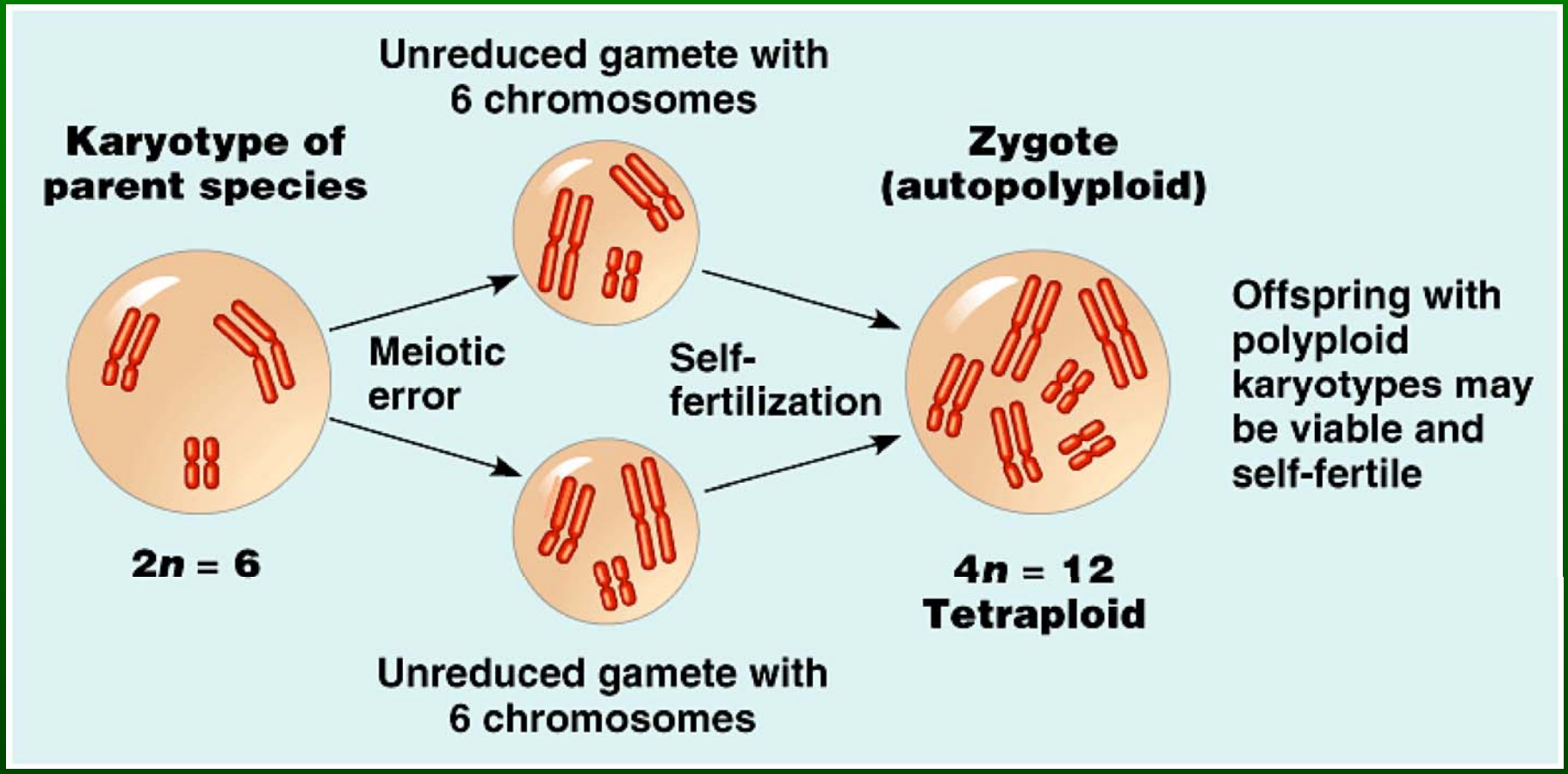
The Evolution of Human α-globin and β-globin Gene Families



Mechanisms: Duplication, Mutation, Transposition, etc.

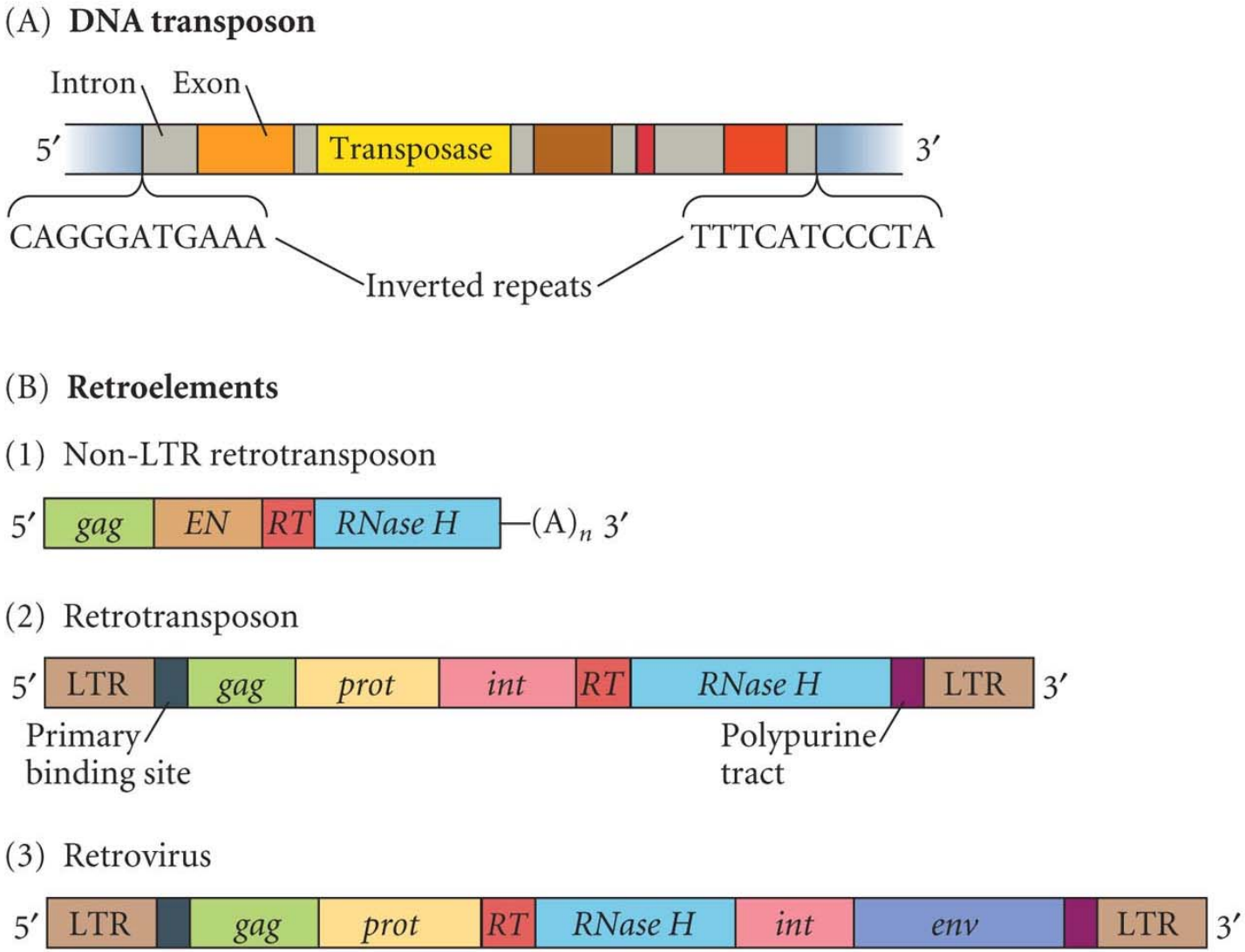

6-week-old fetus
24-week-old fetus
18-week-old baby
α
β
γ
δ
ε
ζ
% of total globin synthesis
50
40
30
20
10
0
6
12
18
24
30
36
6
12
18
24
30
36
42
48
Number of weeks after conception
Birth
Number of weeks after birth
Fetal liver
Bone marrow
Fetal yolk sac
Fetal spleen

Polyploidy


Unreduced gamete with
6 chromosomes
Karyotype of
parent species
Zygote
(autopolyploid)
Meiotic
error
Self-
fertilization
Offspring with
polyploid
karyotypes may
be viable and
self-fertile
2n = 6
4n = 12
Tetraploid
Unreduced gamete with
6 chromosomes

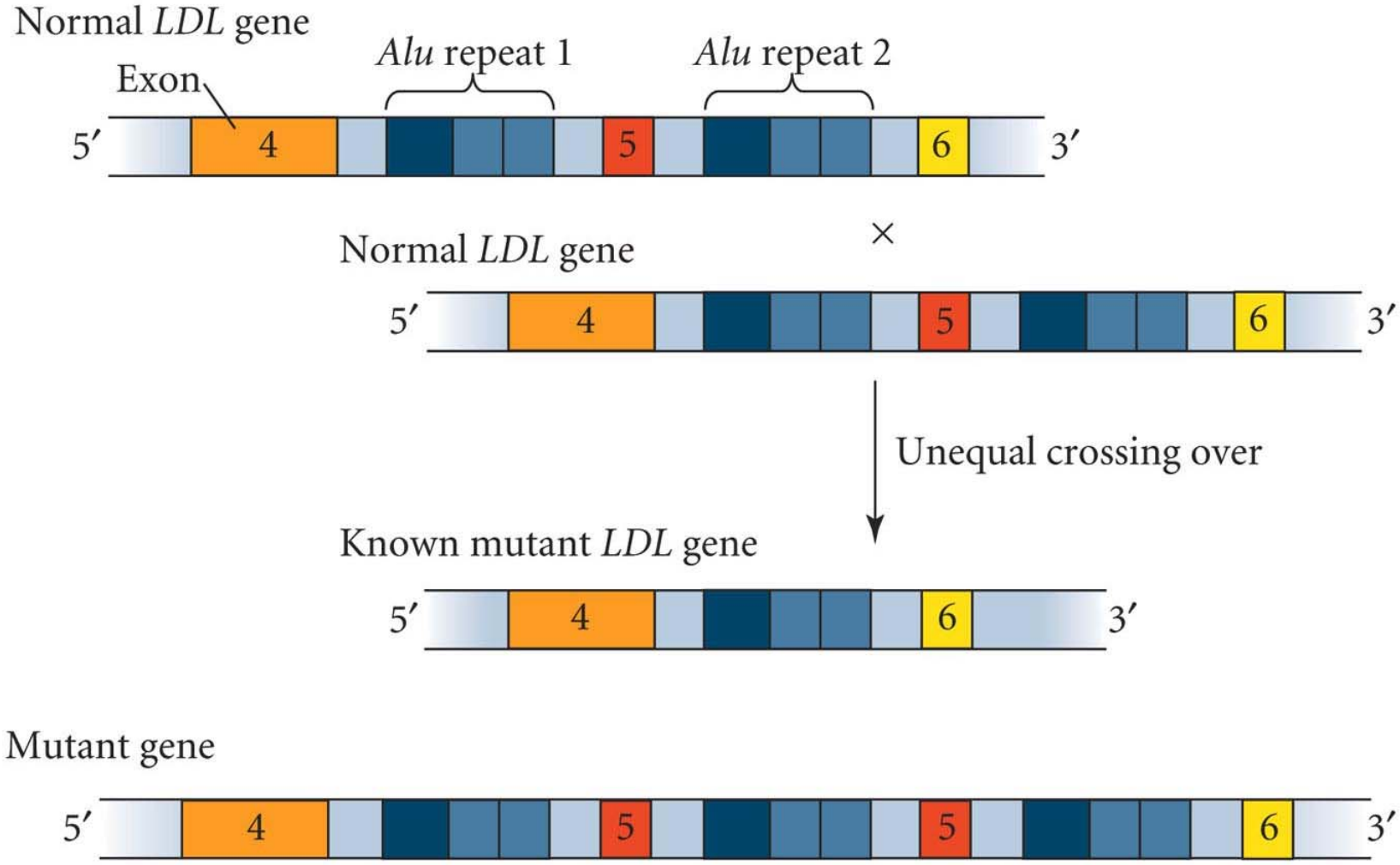
Transposition

Some different kinds of transposable elements



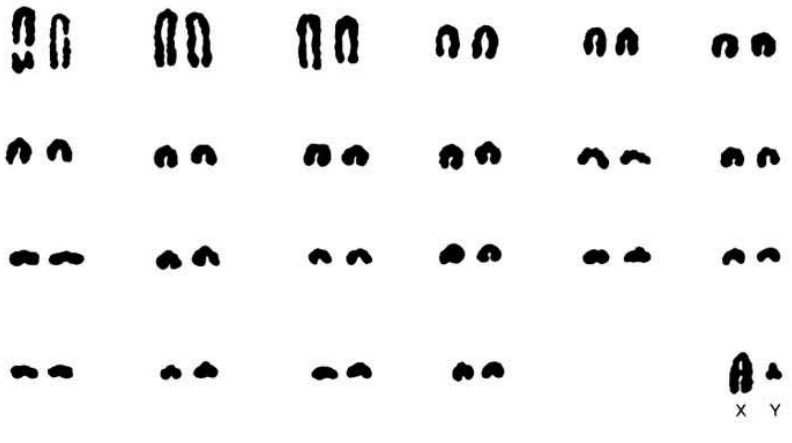
Transposition

A mutated low-density lipoprotein (*LDL*) gene in humans lacks exon 5



Translocation

Muntiacus reevesii ($2N = 46$)



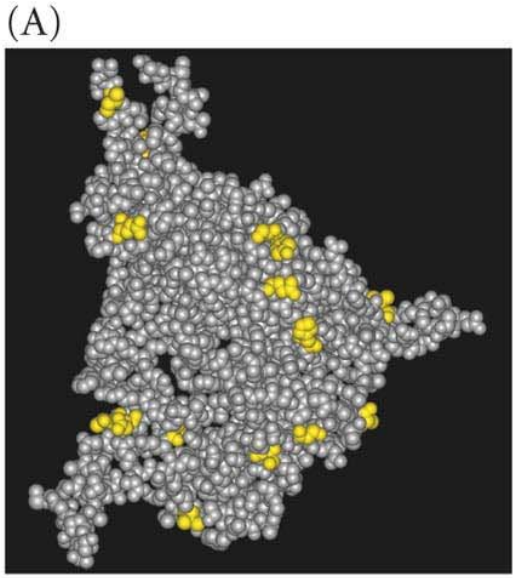
Muntiacus muntiacus ($2N = 8$)

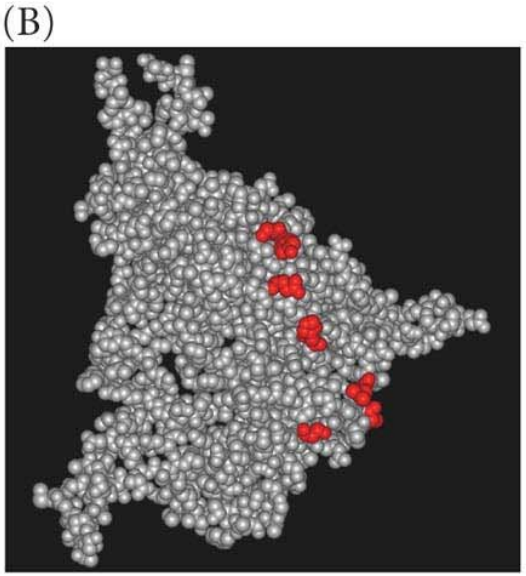
Barking Deer: Similar phenotype, dissimilar karyotype.

Structure and Function Considerations

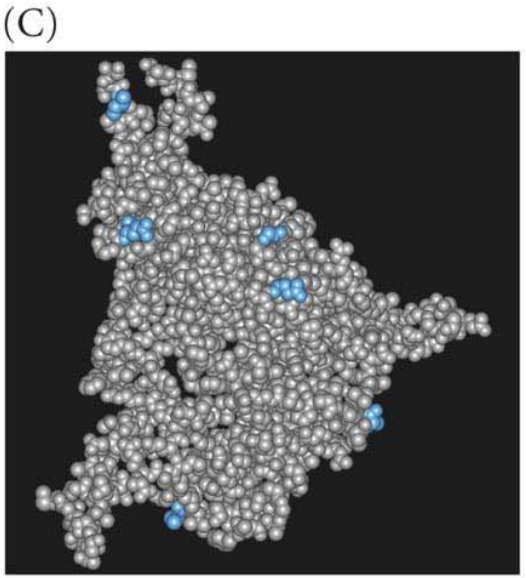
- Magnitude of genetic and phenotypic changes are not necessarily correlated, most have little effect on fitness.
- Repair mechanisms are not random, directed to specific exons.
- Point mutations at first and second position, usually replacement.
- Point mutations at third position, usually silent.
- Most populations harbor considerable allele diversity.

The surface of the major capsid protein (gpF) of phage strains ϕX174 and S13.





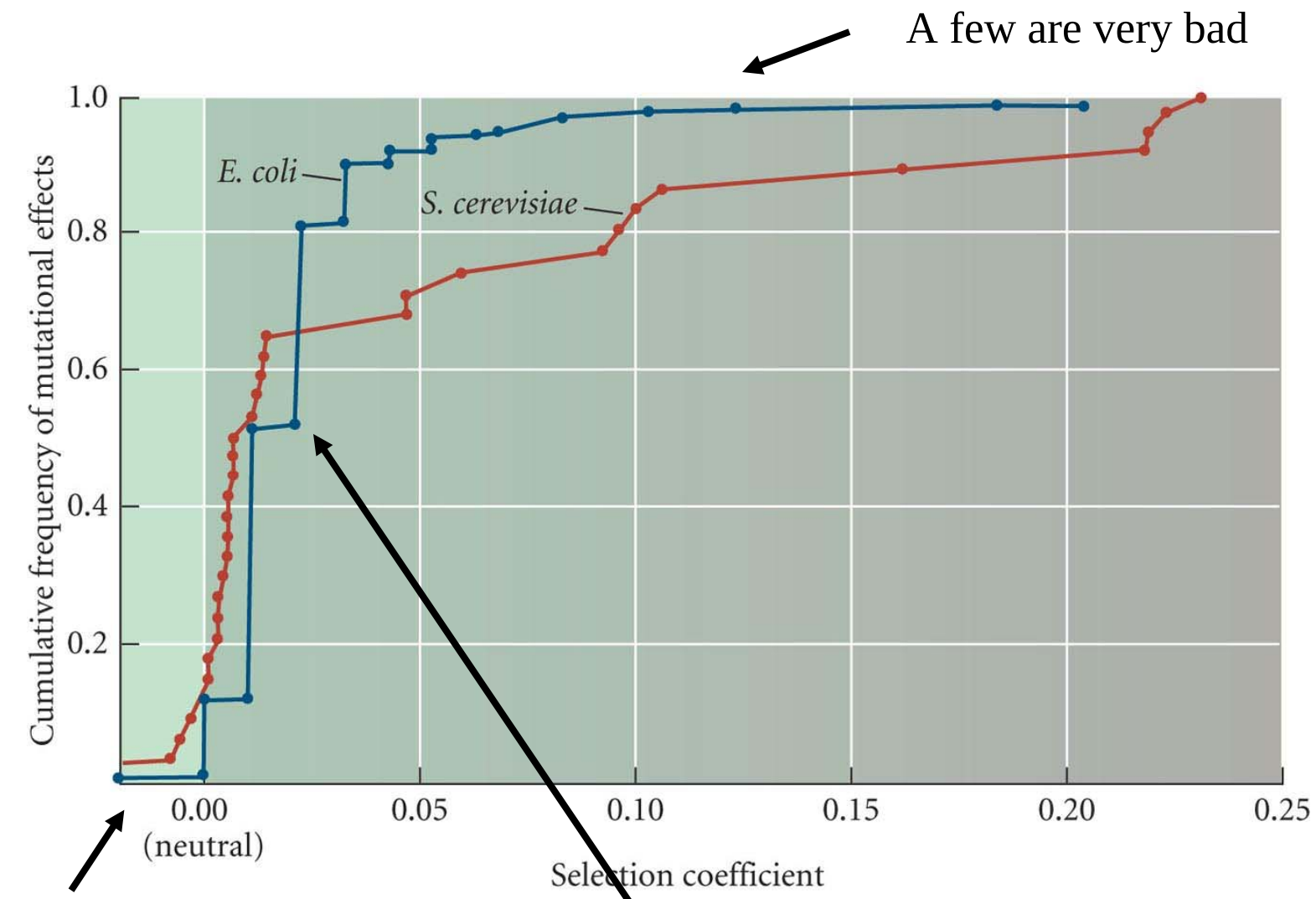
Affects Fitness



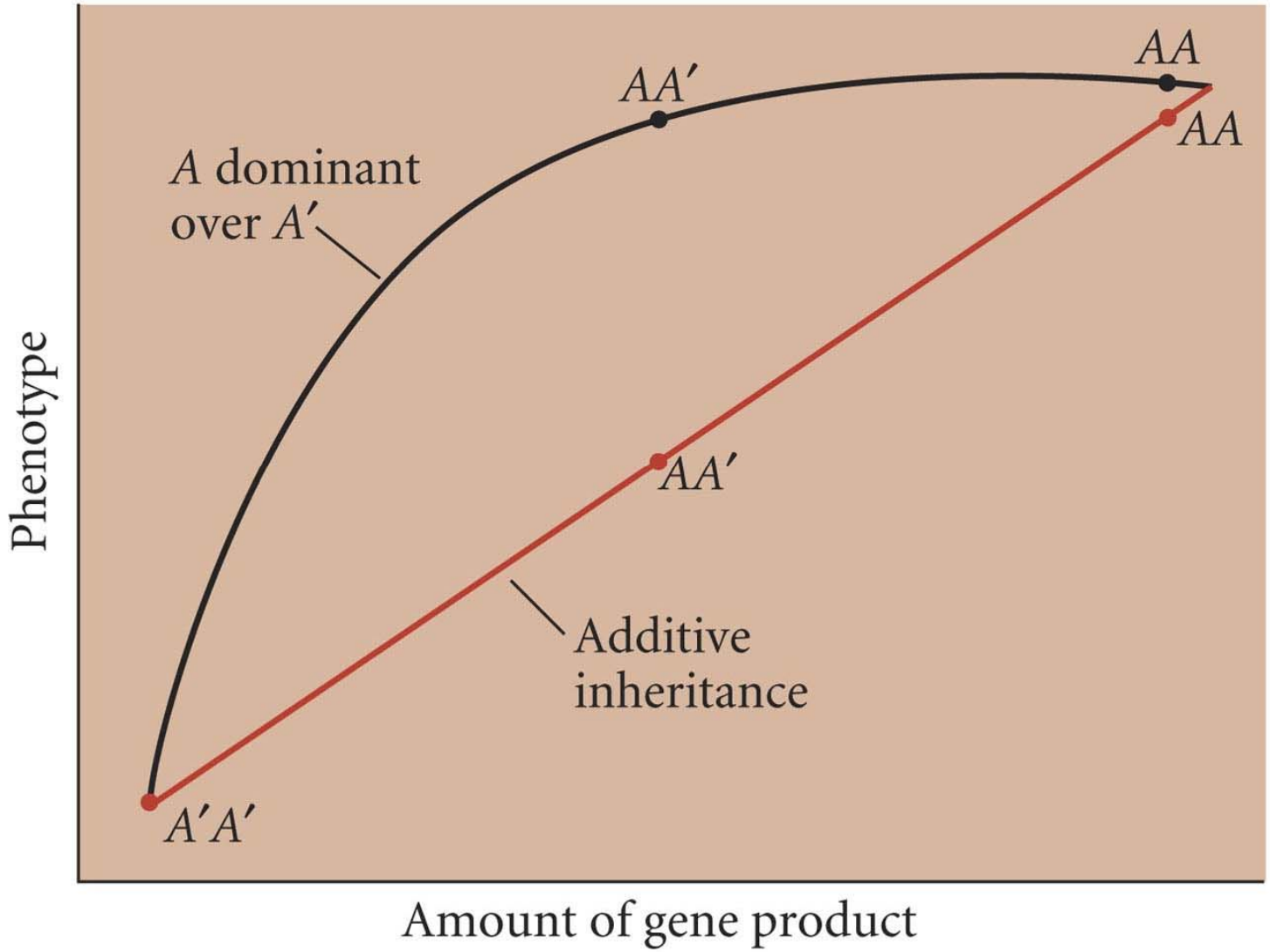
Difference

AA Replacements
wrt Wild type

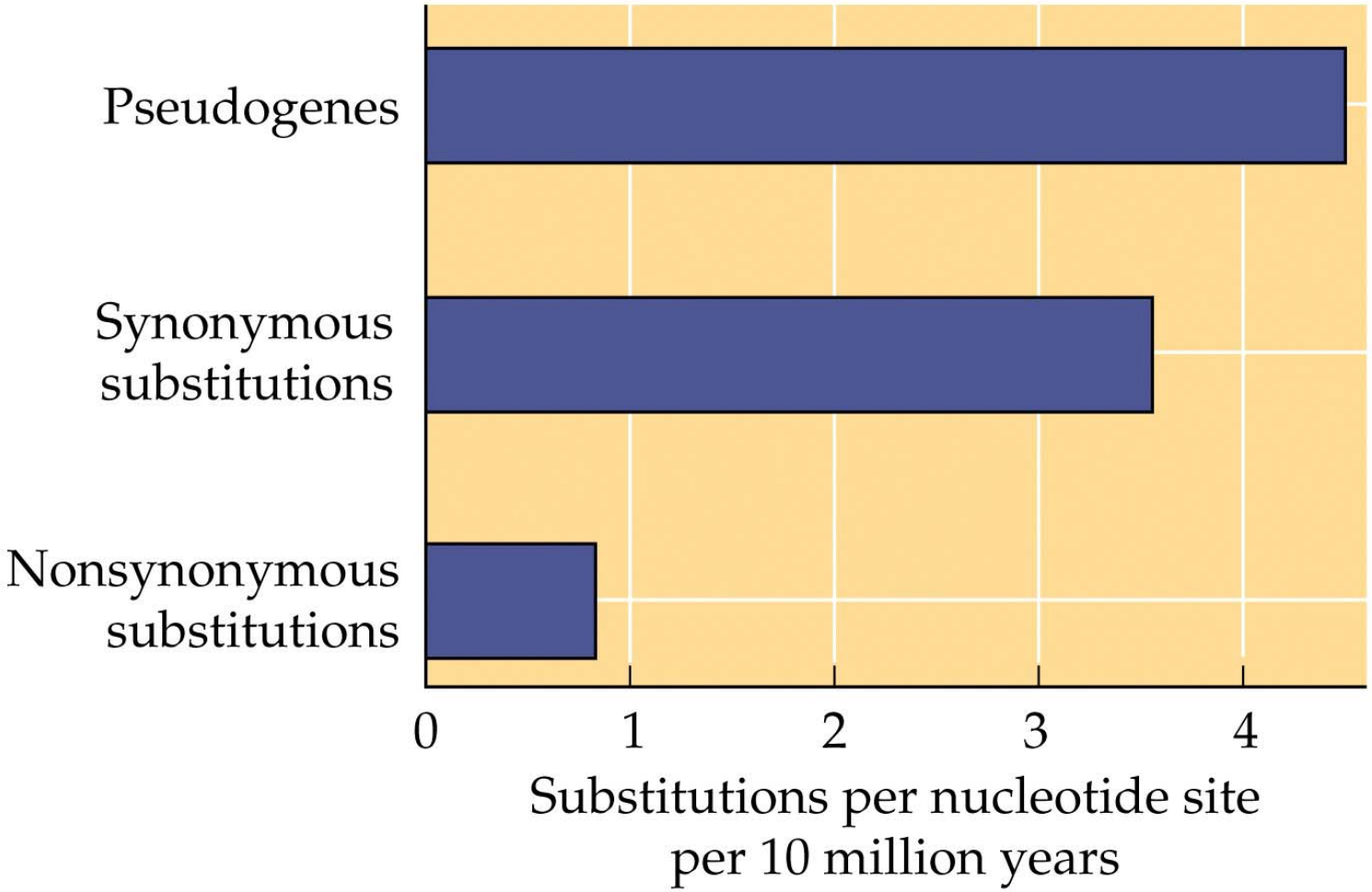
Most mutations have a weakly deleterious effect

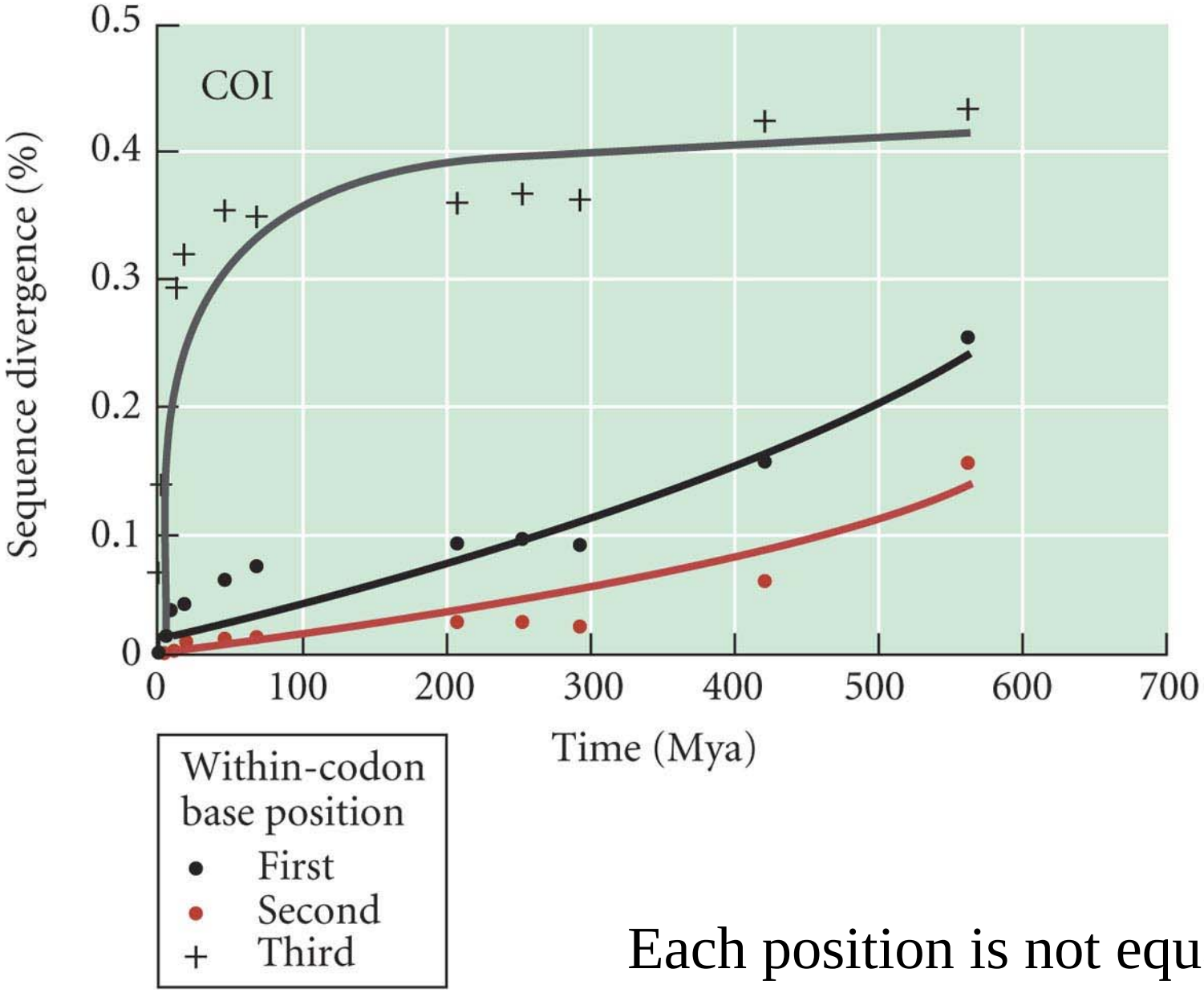


Dominance can influence their strength with nearly full expression



Changes evolve slowly in regions of functionally significant molecules, but more rapidly in regions where base substitutions do not affect molecule functioning.

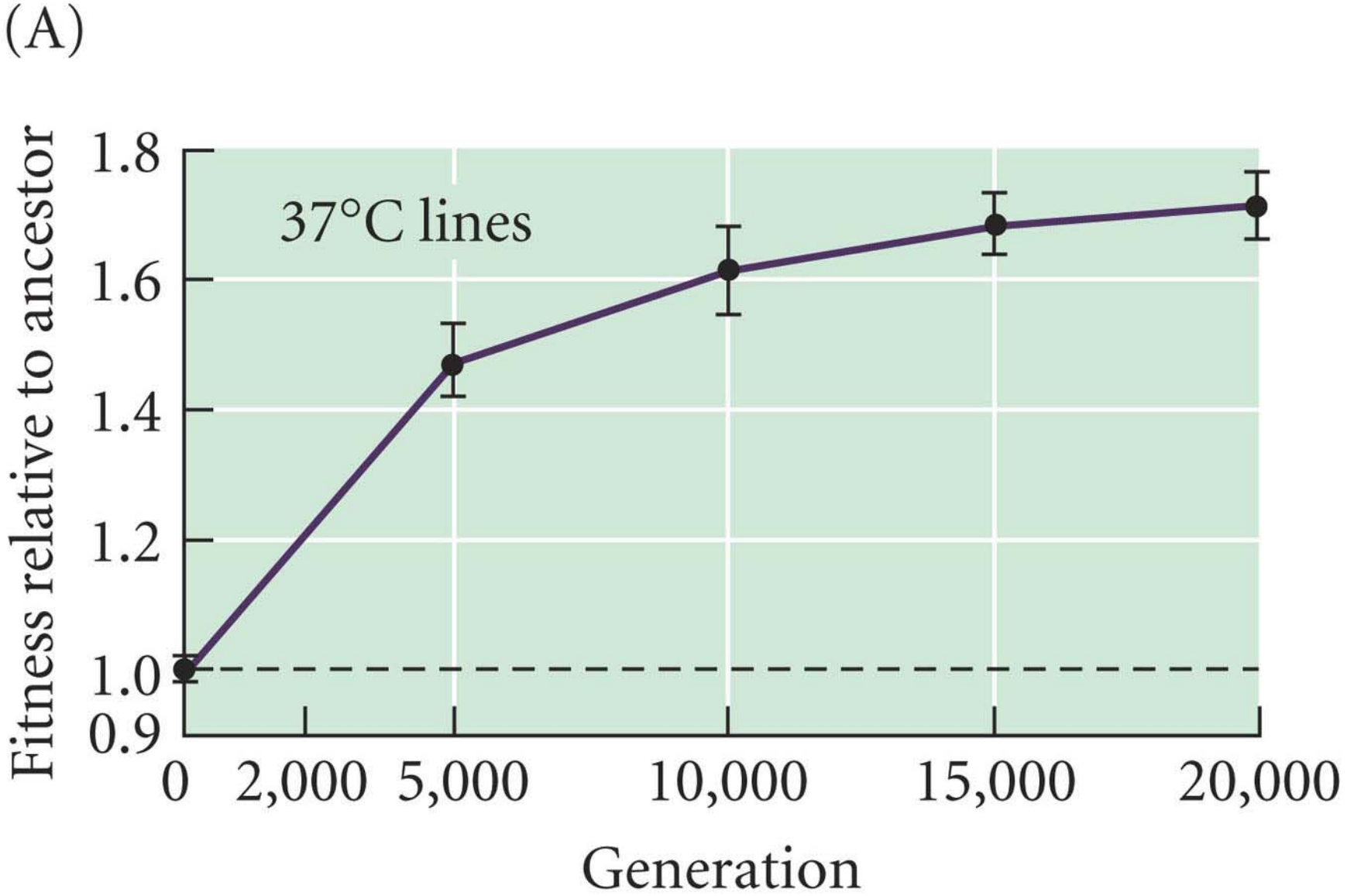




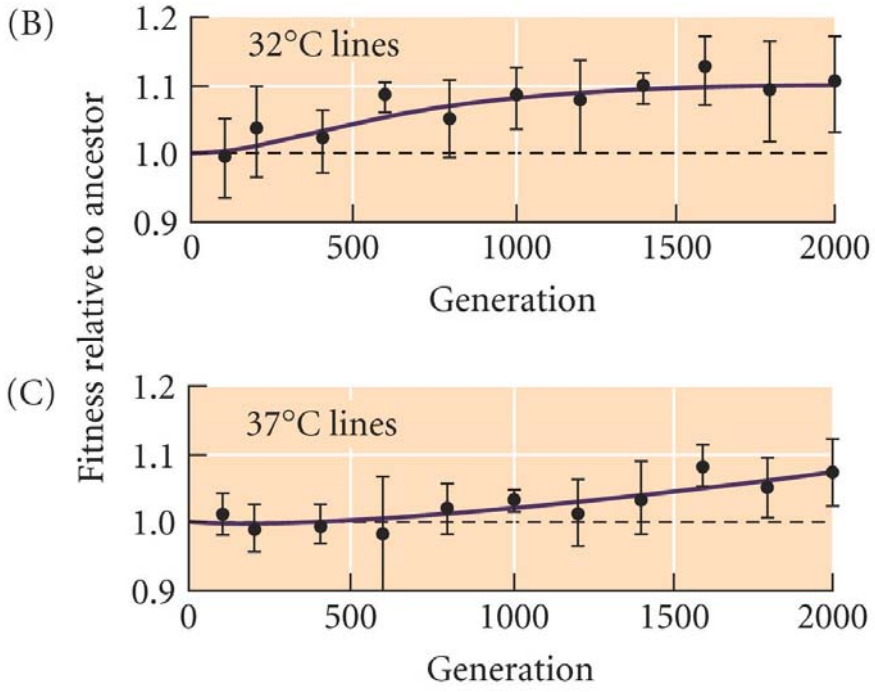
Each position is not equal!

Adaptation in experimental populations of *E. coli*

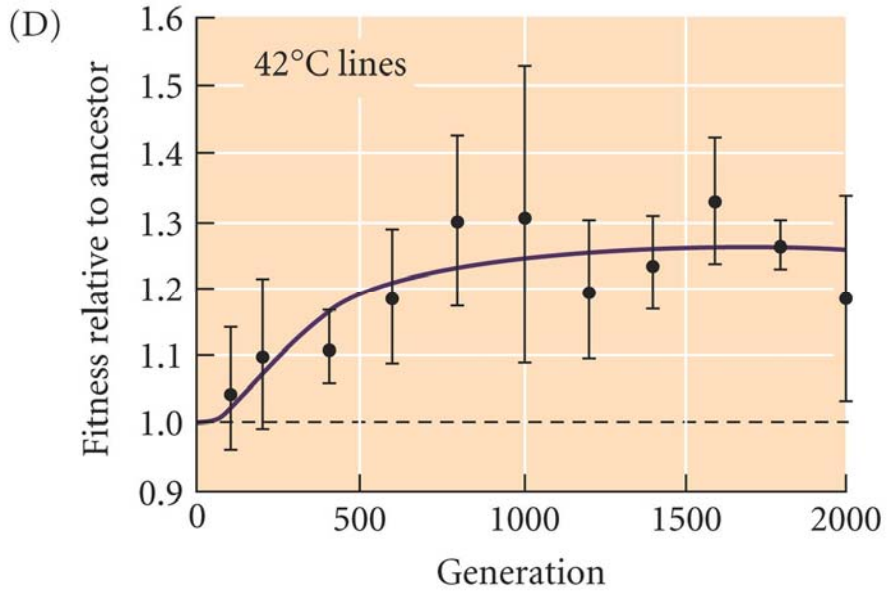
(Fitness is growth rate proportional to ancestor)



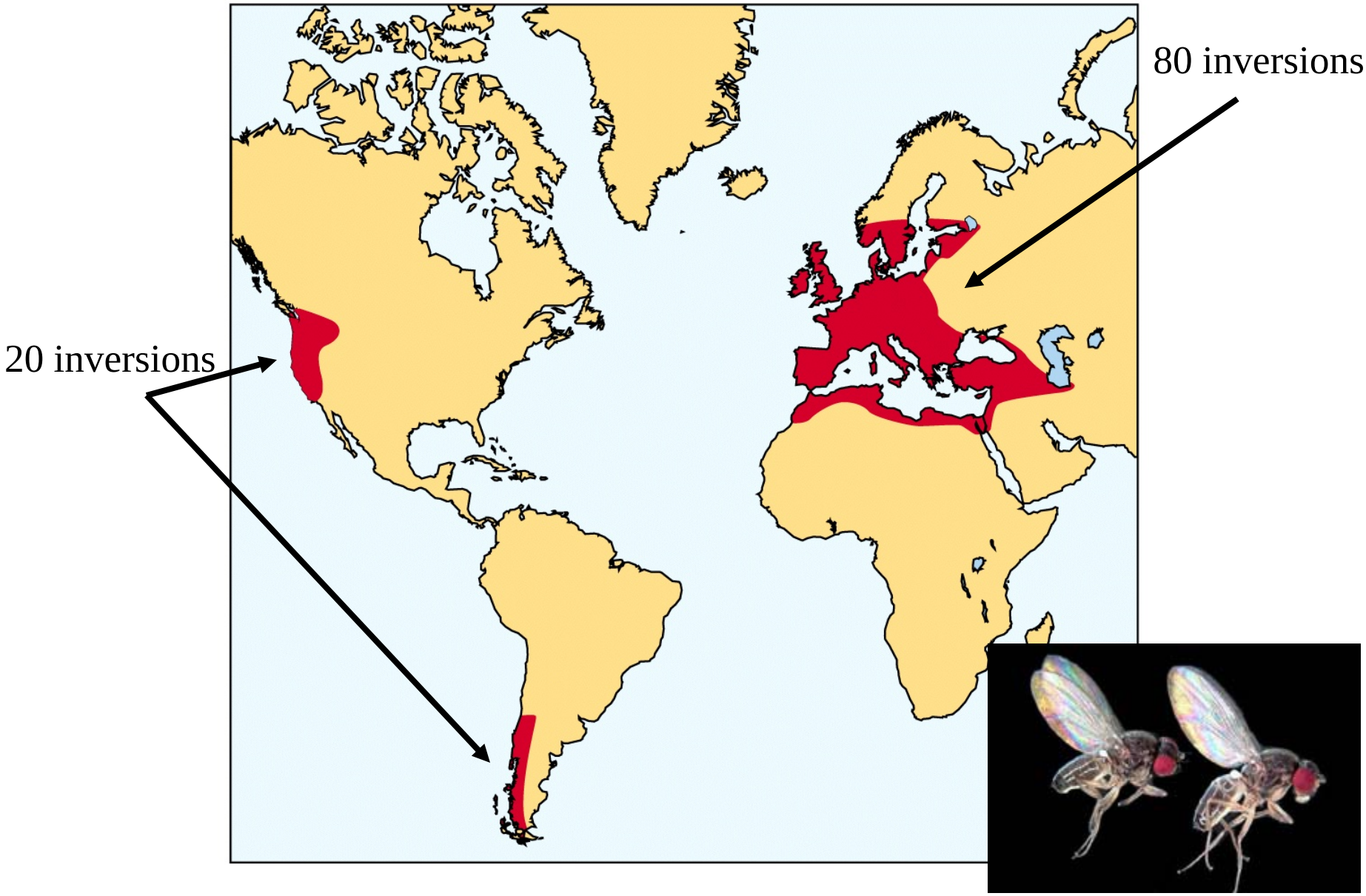
Adaptation in experimental populations of *E. coli*



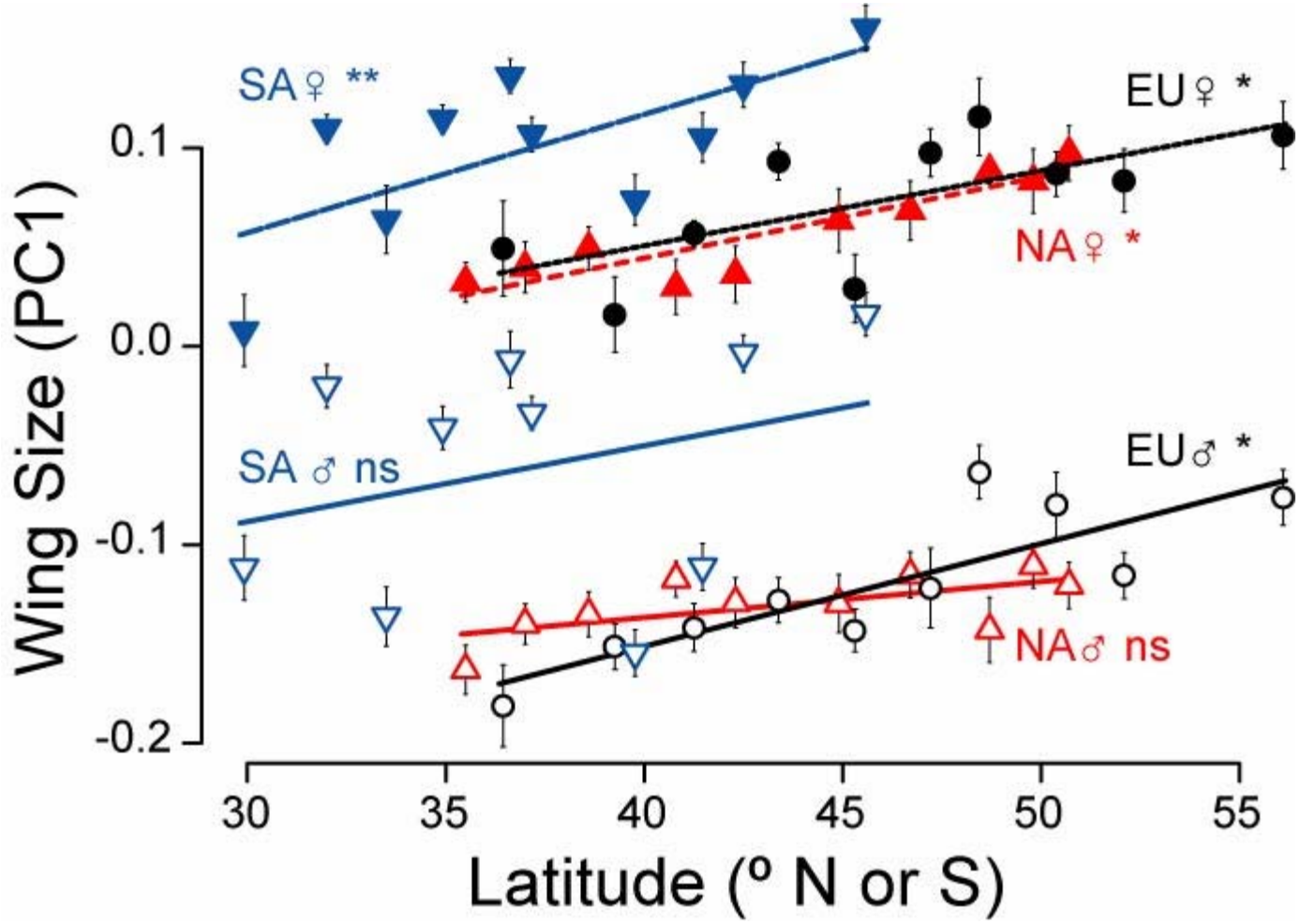
Initial populations lacked genetic diversity, increase in adaptation due to N.S. acting on new mutations.



Founder Effect in *Drosophila subobscura*



Clines = Gradients



NS: Larger sizes in colder wetter climates, greater number of inversions.