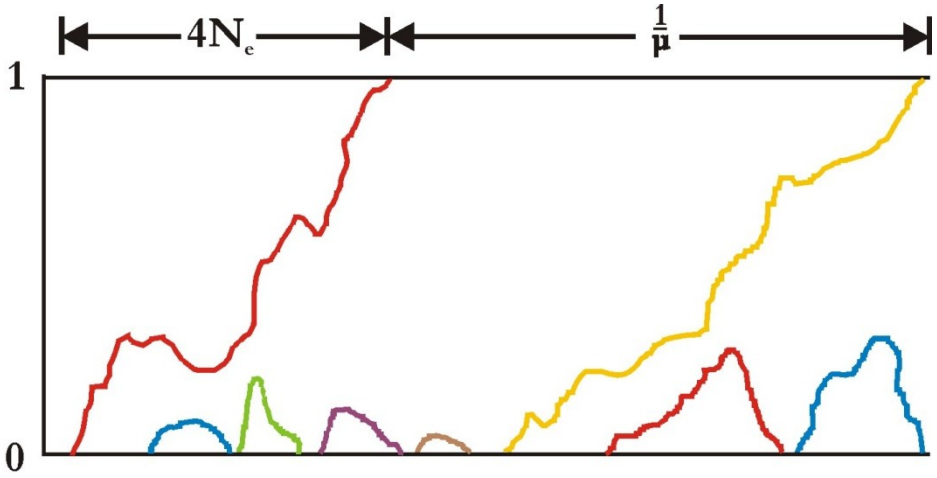
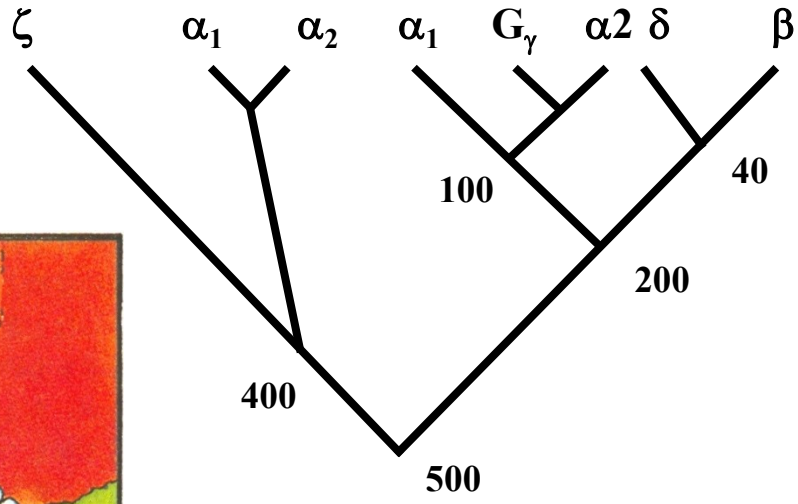
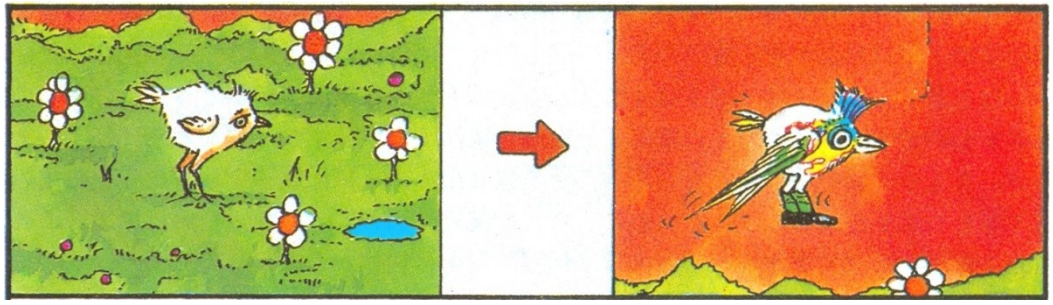
MOLECULAR EVOLUTION







Substitutional load and the cost of natural selection

substitution = replacement of one allele by other (tj. fixation of a new allele)

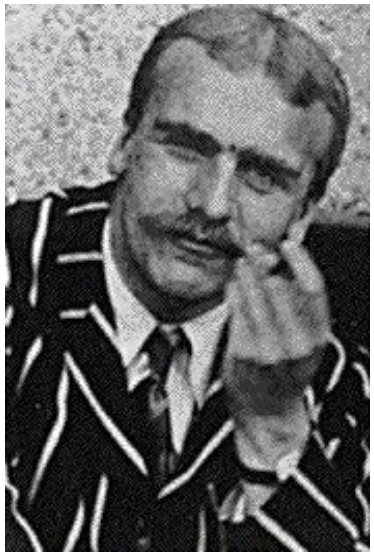
if an individual do not reproduce, we call it genetic death

J.B.S. Haldane (1957):

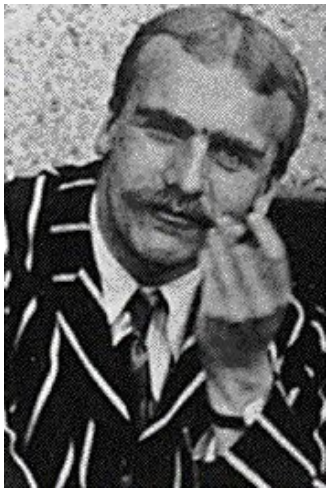
advantageous mutation $\rightarrow$ fixation and replacement of a deleterious allele

as long as the original allele exists in the population, mean fitness is lower than maximal fitness

J.B.S. Haldane



substitutional load*): $L = 1 - \bar{w}$; when $\bar{w} = w_{\max}$ $L = 0$



in general

$$L = \frac{w_{\max} - \bar{w}}{w_{\max}}$$

it measures to what extent an average member of the population is less fit than the best fit genotype

it expresses probability that an average individual dies before his/her reproduction

*) in general = genetic load; other loads: mutational: emergence of a deleterious allele; segregational: cost of homozygotes under overdominance

Cost of natural selection:

We can envisage replacement of one allele in a population by another as a „selective death“ of the original allele

The higher is the strength of selection, the more of the original (less advantageous) alleles are removed from the population (they „die“)

If the natural selection was too strong it could cause extinction of the whole population $\Rightarrow$ overproduction of the offspring necessary!

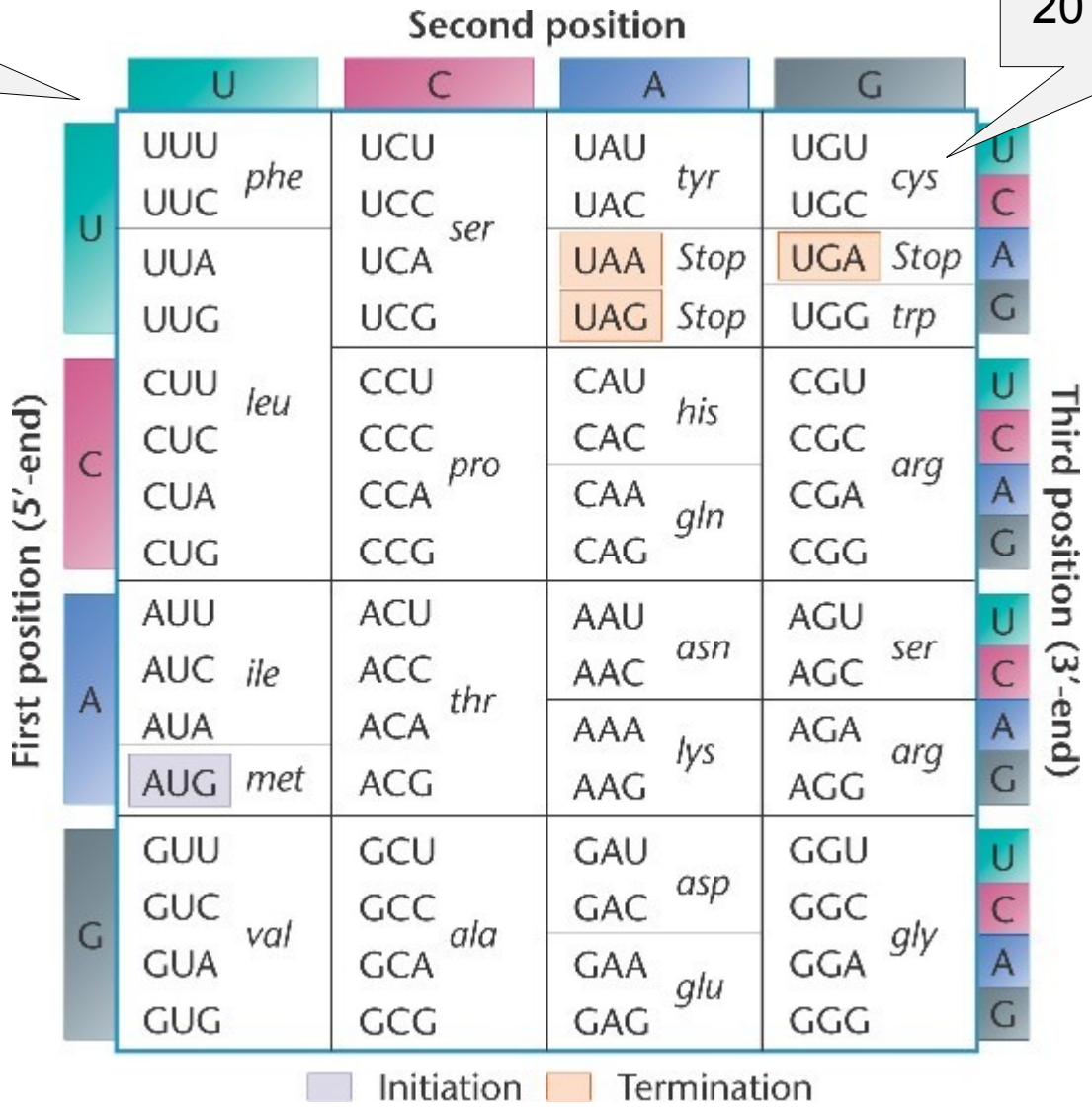
eg. if the ratio of non-surviving to surviving alleles is 0,1/0,9, each survivor have to produce by 1/9 more descendants,
but if the ratio is 0,999/0,001 $\rightarrow$ by ~1000 more descendants!

Haldane: upper limit of the cost of natural selection $\approx$ substitution of 1 gene per 300 generations

$\Rightarrow$ evolution cannot run too fast, the cost of selection would be too high

64 codons

20 amino acids



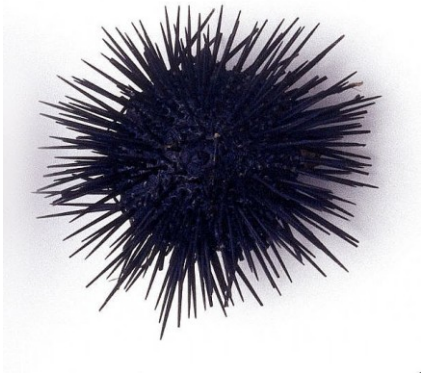
excess of synonymous nucleotide substitutions → esp. at the 3rd position

M. Kimura (1977): mRNA sequences of humans and rabbits →
of 53 nucleotide positions 6 differences, of them only 1 nonsynonymous

× theoretically only 24% of differences should be synonymous

likewise M. Grunstein (1976):

evolutionary rate of Histone H4 in 2 sea urchin species
84 bp mtDNA → 9 of 10 differences synonymous



NEUTRAL THEORY OF MOLECULAR EVOLUTION

Modern Synthesis: selection *vs*. drift debate

beginning of the 1960s → amino acid sequences in proteins

1966: protein elektrophoresis

Richard Lewontin and Jack Hubby - *Drosophila pseudoobscura*;

Harry Harris - humans

→ high level of polymorphism

Data gathered till the end of 1960s suggested that:

Rate of molecular evolution is too high

Genetic variation in natural populations is too high

... both would require too high cost of natural selection $\Rightarrow$
polymorphism cannot be maintained by selection

Rate of evolution at the molecular level is constant

Higher evolutionary rate in functionally less important parts of the molecule, in noncoding regions and pseudogenes

Why so high polymorphism in populations?

Motoo Kimura: because alleles are selectively neutral, it lasts many generations till a new mutation is fixed – meanwhile the population must be polymorphic = transient polymorphism

During the process of fixation often a new allele appears by mutation ⇒ in a sufficiently large population at each point in time there is a large amount of variation

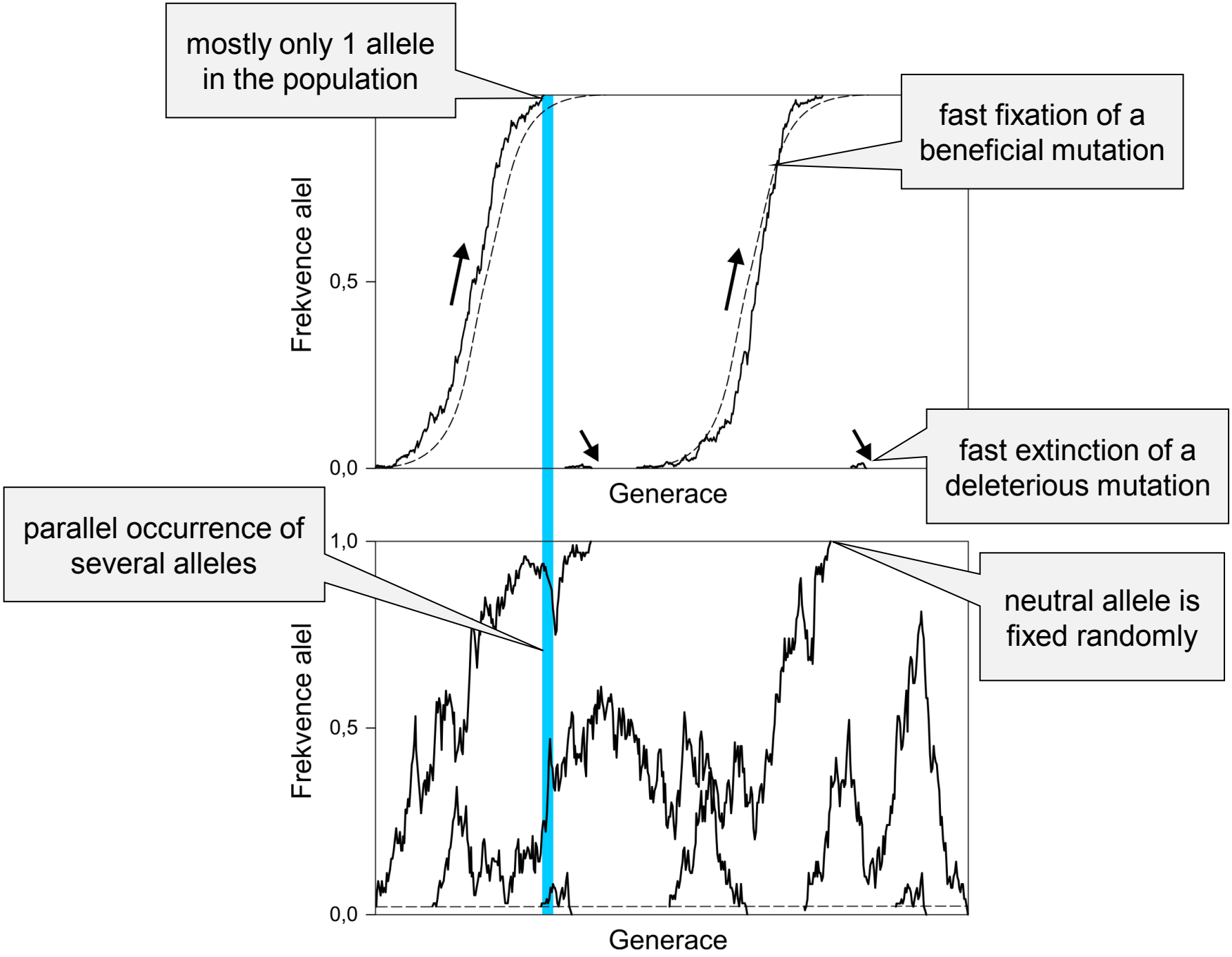
Population is in equilibrium between drift and mutation

M. Kimura

M. Kimura (1968)

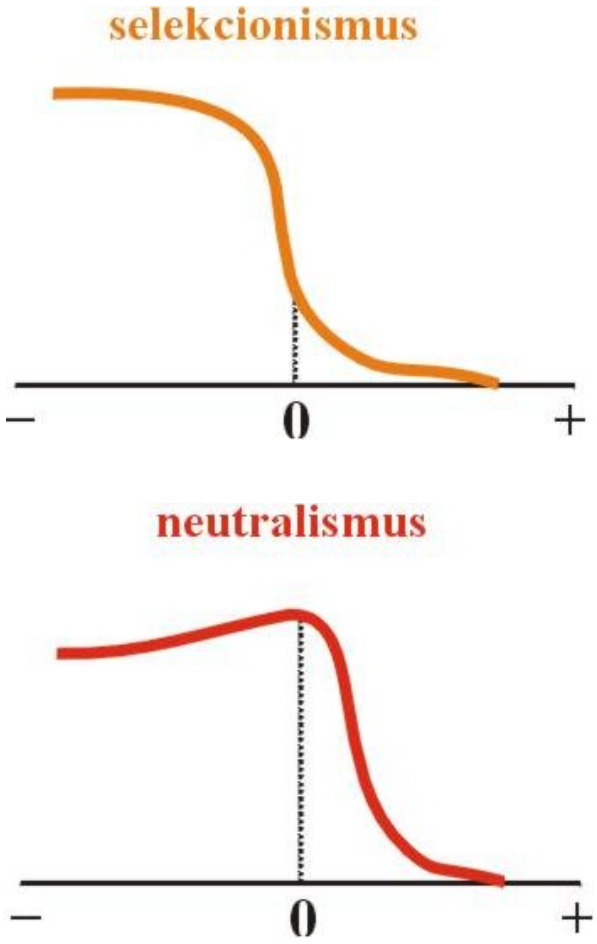
J.L. King & T.H. Jukes (1969)

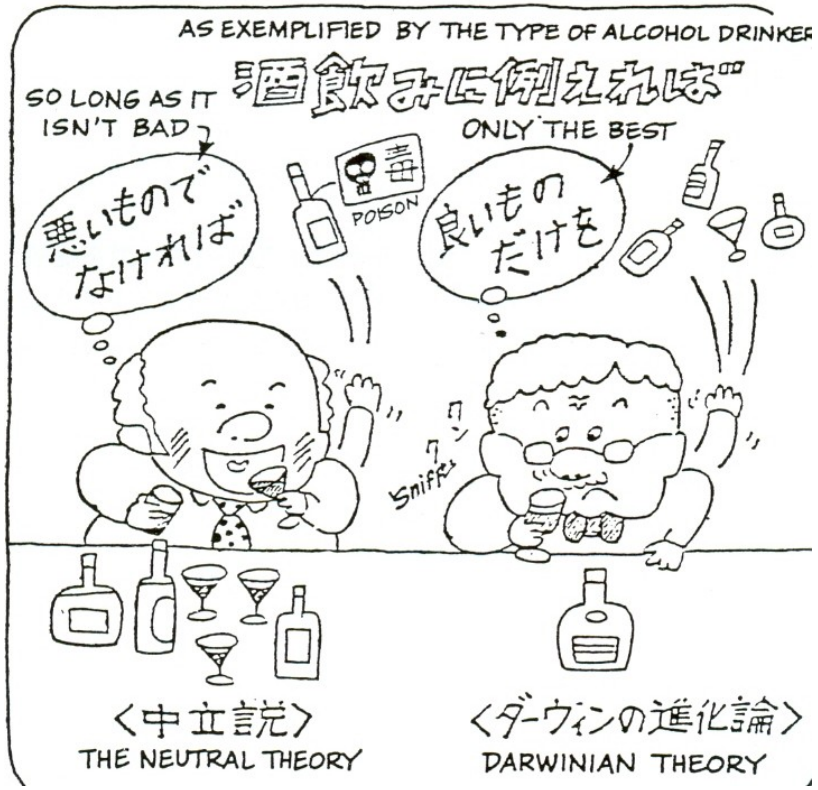
→ neutral theory of molecular evolution


mostly only 1 allele in the population
fast fixation of a beneficial mutation
fast extinction of a deleterious mutation
parallel occurrence of several alleles
neutral allele is fixed randomly
Frekvence alel
0,5
0,0
Generace
1,0
0,5
0,0
Frekvence alel
Generace

Basic principles of the neutral theory:

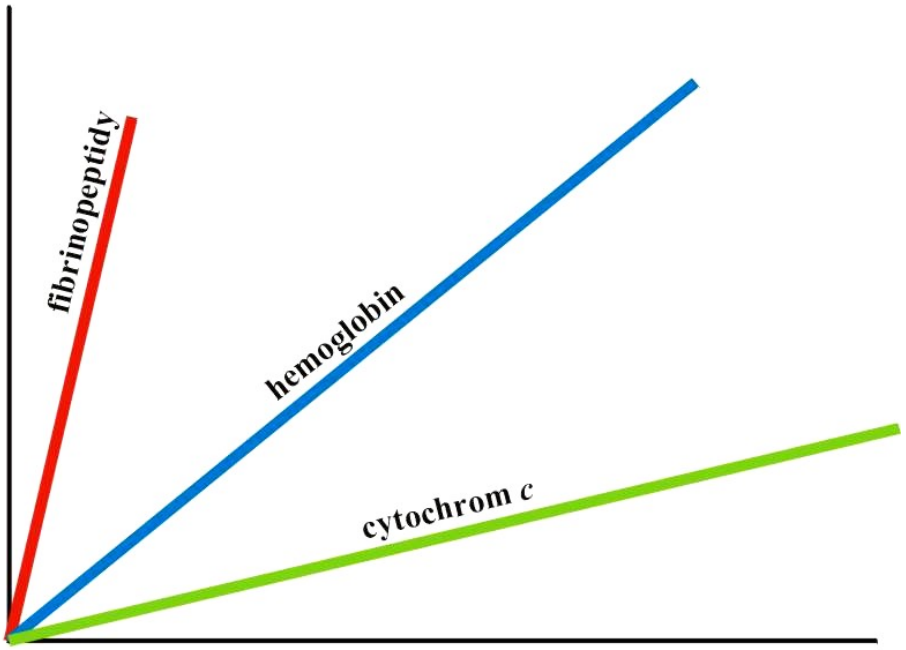
1. most allele substitutions in a population are neutral (⇒ drift)





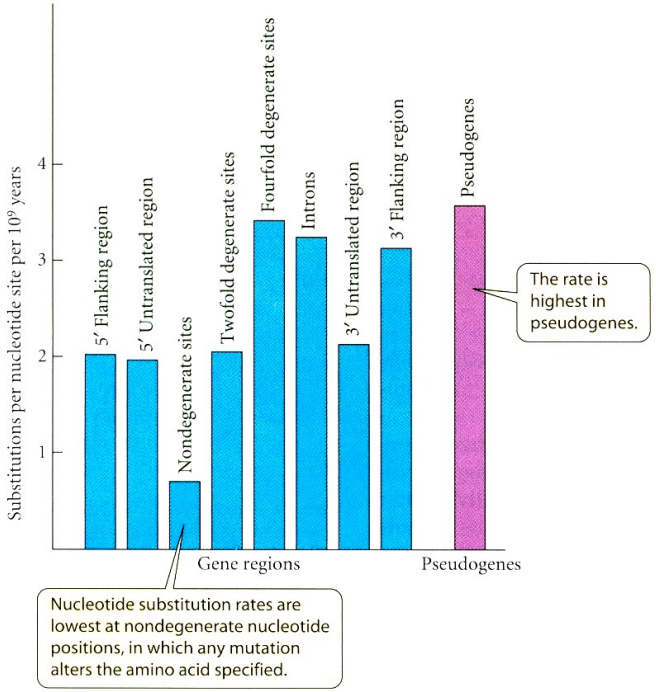
2. evolutionary rates in differently important proteins are different

No. AA substitutions per 100 molecules



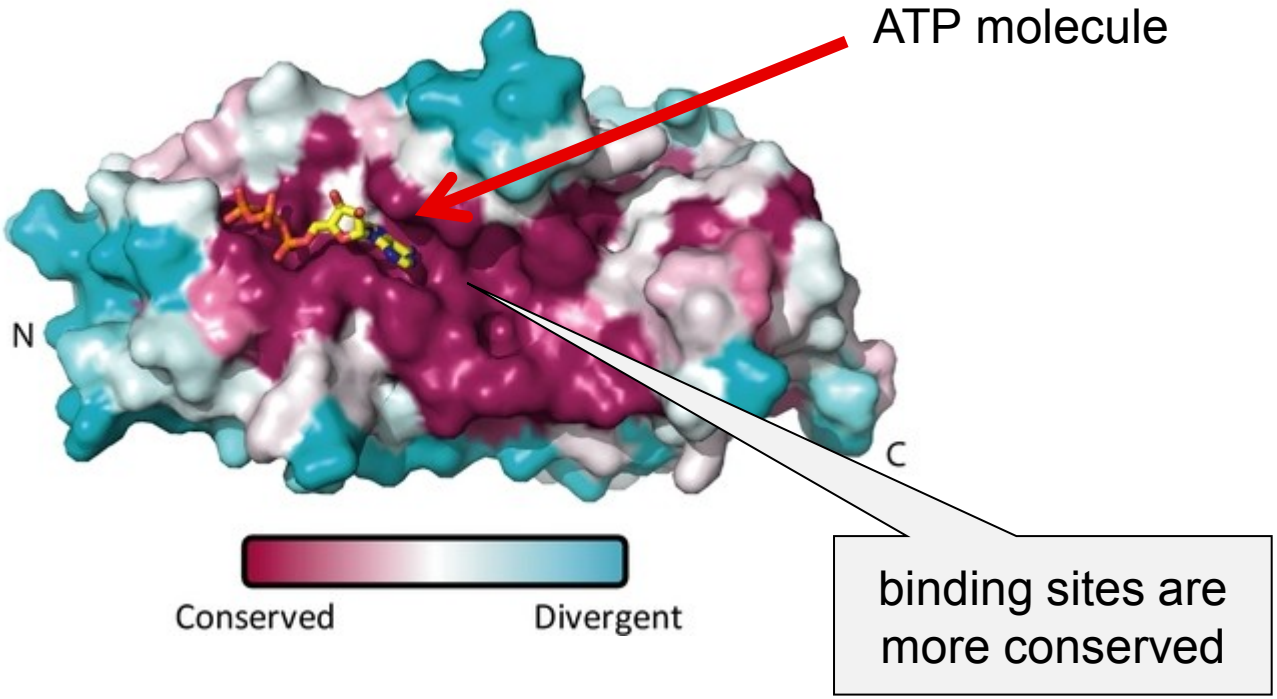
Time of divergence (millions years)

fibrinopeptides	8,3
pancreatic ribonuclease	2,1
lysozyme	2,0
alpha-globin	1,2
insulin	0,44
cytochrome *c*	0,3
histone H4	0,01



3. diverse evolutionary rate in different parts of proteins (binding sites × structural areas)

Eg.: transient receptor potential vanilloid (TRPV) channel protein:



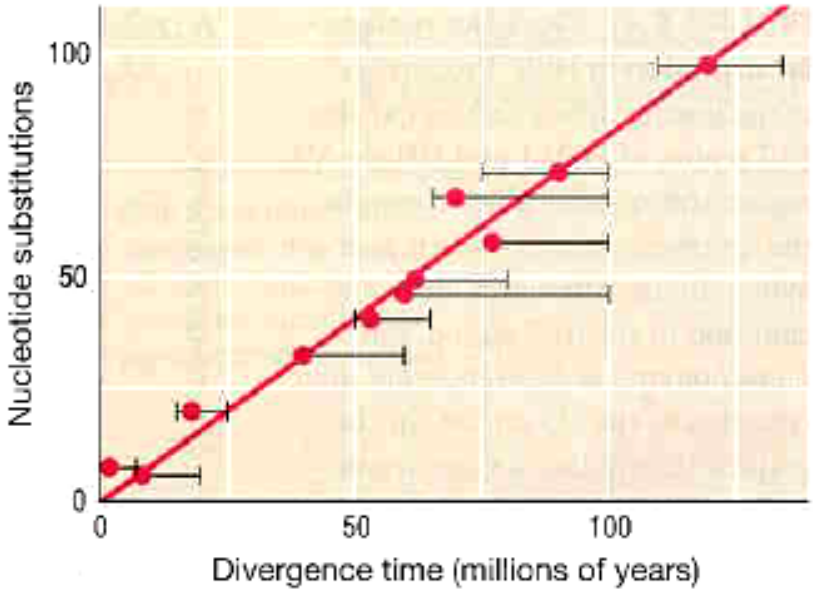
4. different evolutionary rates at individual codone positions

Table 4. Relative frequencies of different types of mutational substitutions in a random protein-coding sequence.

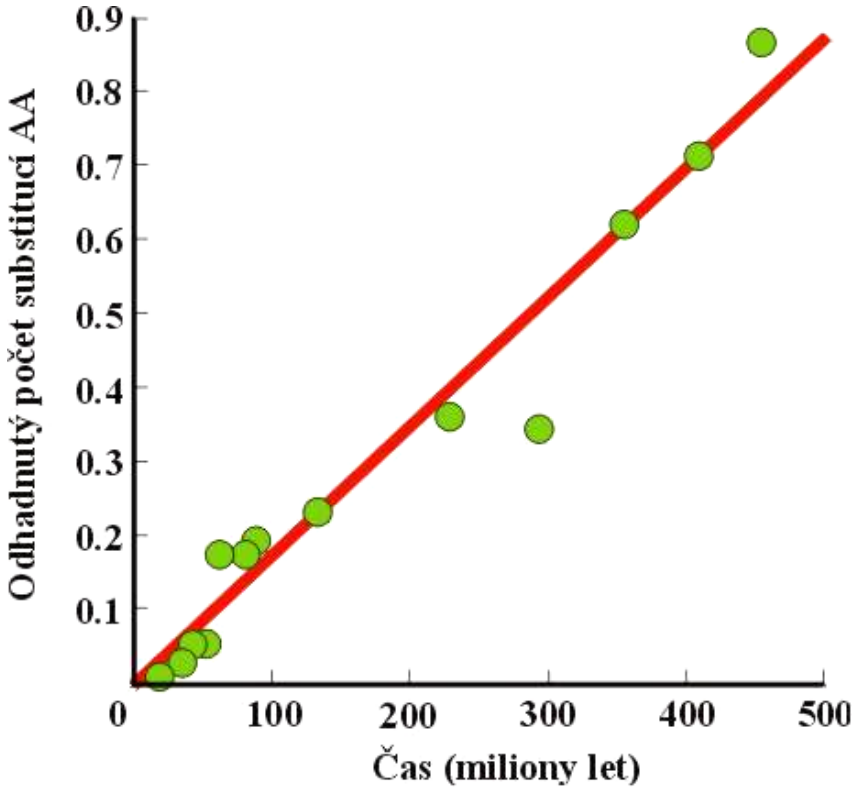
Substitution	Number	Percent
Total in all codons	549	100
Synonymous	134	25
Nonsynonymous	415	75
Missense	392	71
Nonsense	23	4
Total in first position	183	100
Missense	166	91
Nonsense	9	5
Total in second position	183	100
Missense	176	96
Nonsense	7	4
Total in third position	183	100
Missense	50	27
Nonsense	7	4

5. evolutionary rate of a given protein in various species roughly constant

Wilson (1977), mammals, 7 proteins:



Kimura (1983), vertebrates, α-globin:



mostly does not concern morphological, physiological, and behavioural trais

cannot explain adaptations

many deleterious mutations, however, these rapidly removed by selection

selection acts also at the molecular level but most mutations have only small effects on fitness $\Rightarrow$ important role of drift

Haldane s cost of selection was overestimated:

selection mostly soft, not hard

frequency-dependent selection rather than overdominance

selection does not affect individual loci independently (epistasis)

Theoretical principles of neutral theory:

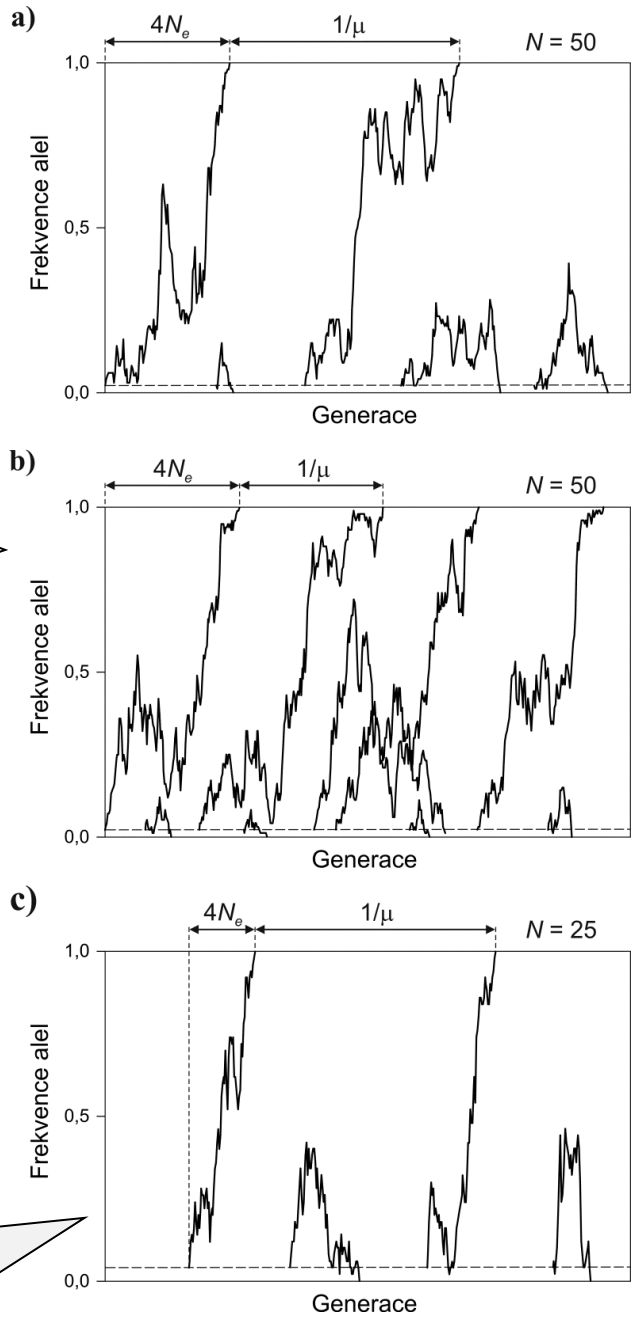
Mean time to fixation of a novel mutation
= $4N_e$

Mean interval between fixations
= $1/\mu$

moderately sized population: mutations more frequent

In small populations faster fixations but longer interval between them:

small population: mutations infrequent



Theoretical principles of neutral theory:

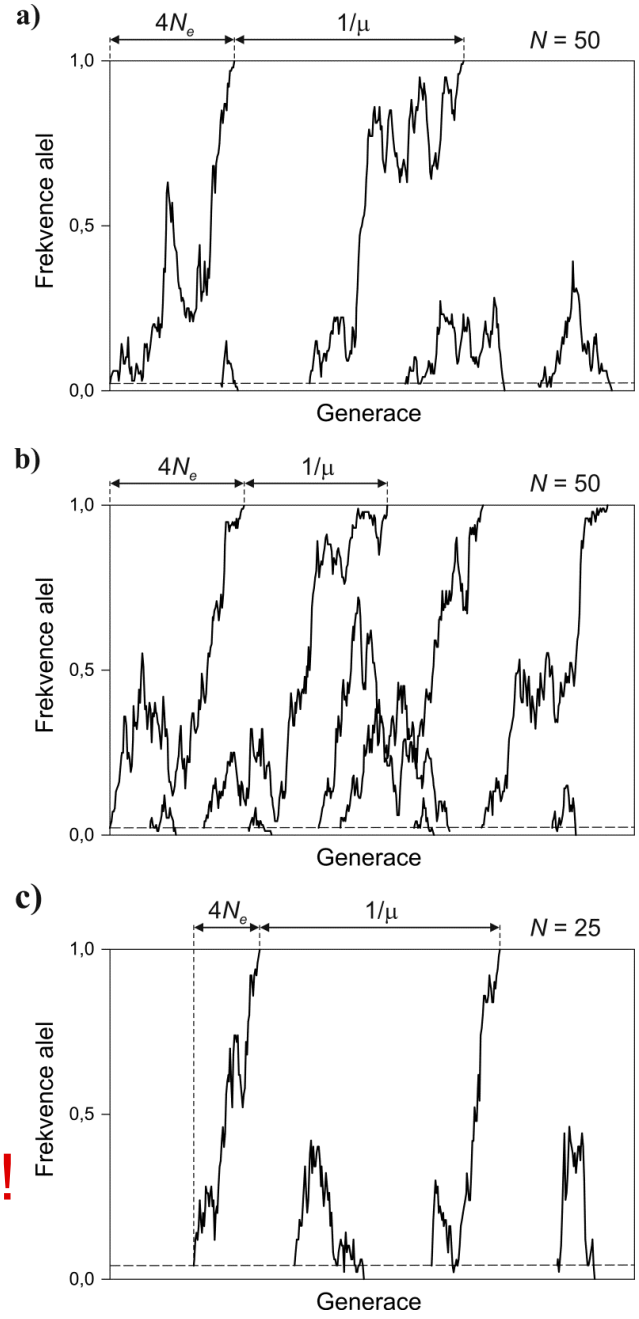
Mean time to fixation of a novel mutation
= $4N_e$

Mean interval between fixations
= $1/\mu$

Frequency of substitutions (replacements of one allele by another in populations):

$1/(2N_e) \times 2N_e\mu = \mu$

⇒ rate of neutral evolution is independent of N_e, depends only on frequency of neutral mutations μ !



Theoretical principles of neutral theory:

mean balanced heterozygosity:

$$\frac{\theta}{\theta+1}, \text{ where } \theta = 4N_e\mu$$

larger populations ⇒ higher heterozygosity

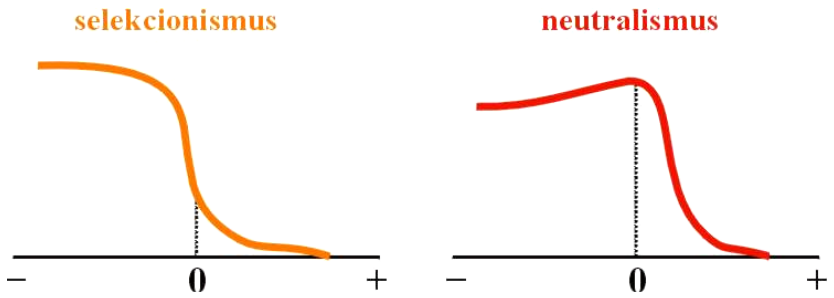
continual emergence of new mutations ⇒ increase of variation

× its erosion by drift

⇒ continual replacement of one allele by another

Equilibrium between mutation and drift ⇒ polymorphism (contrary to the mutation-selection eq.) is transient

Rate of neutral mutations:

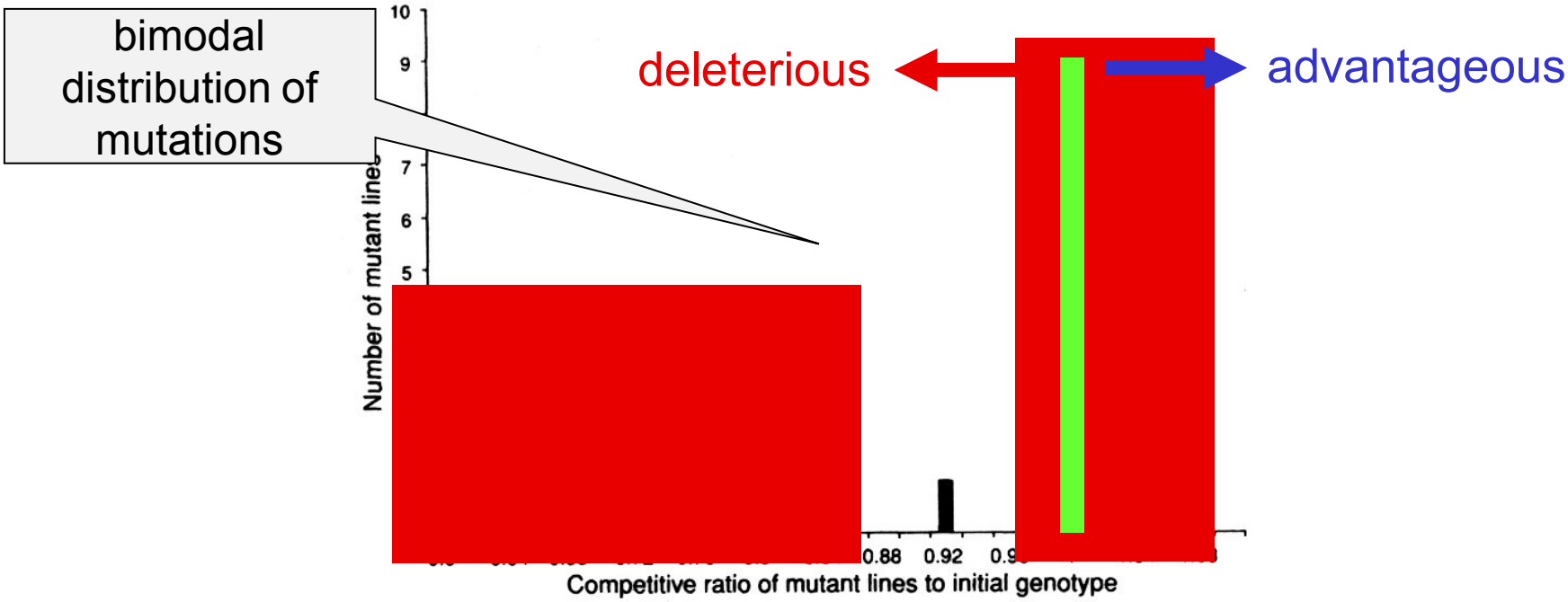


Zeyl & DeVisser (2001):

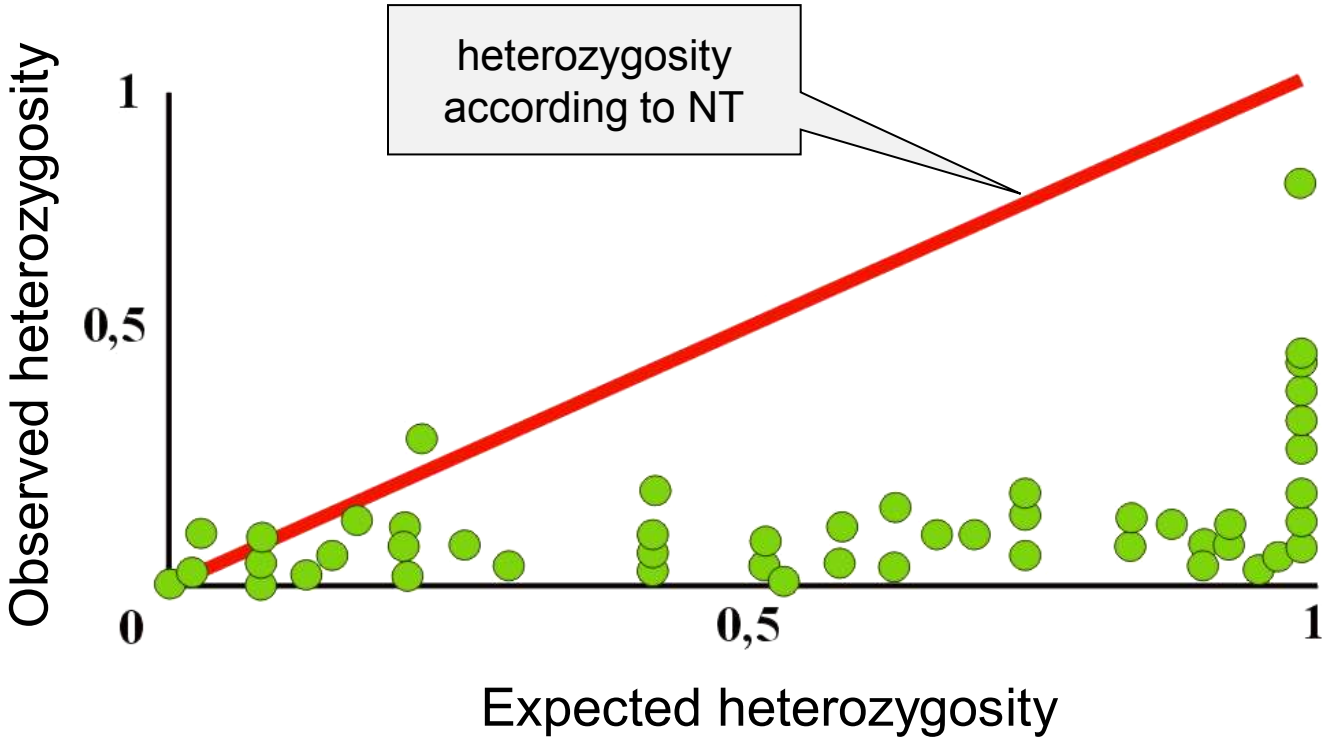
yeast *Saccharomyces cerevisiae*

50 replicated populations,
1 individual in each generation

the experiment does not detect extremely deleterious (lethal) mutations

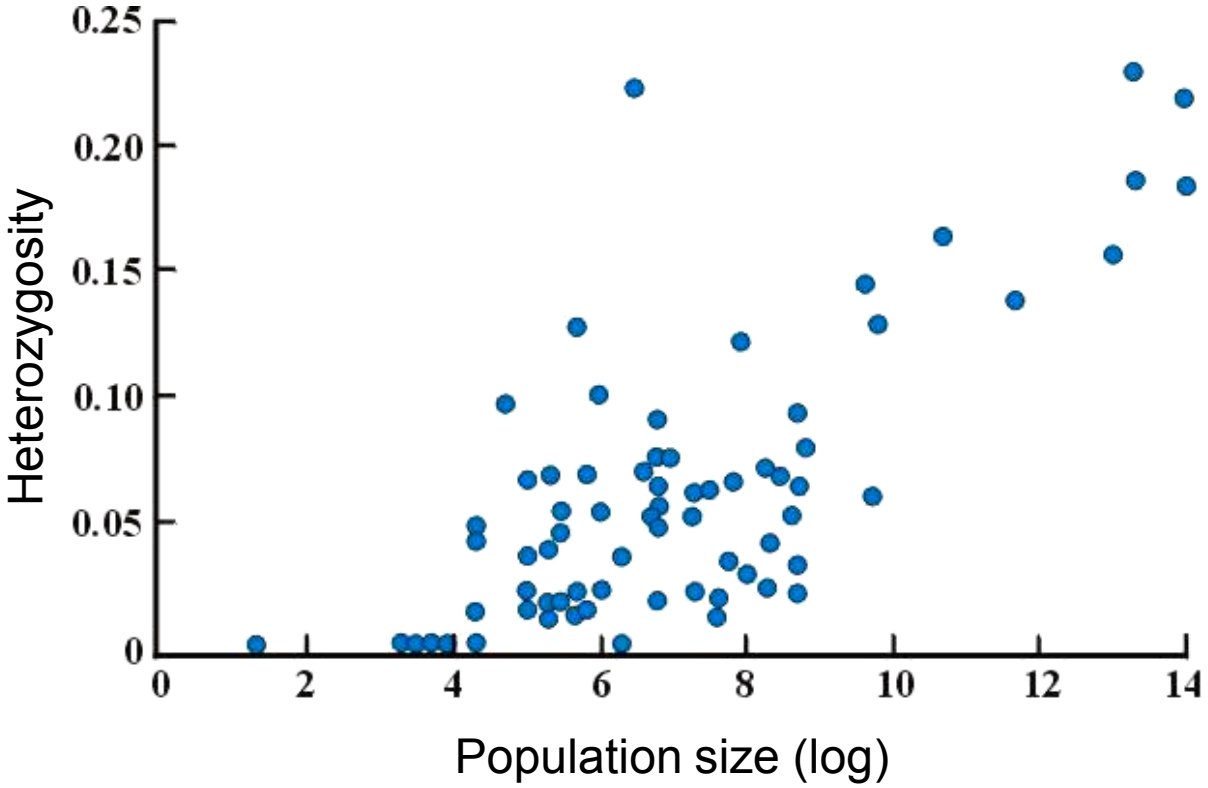


Test of neutral theory: range of heterozygosity



Observed <u>heterozygosity lower</u> than predicted by NT

Test of neutral theory: range of heterozygosity



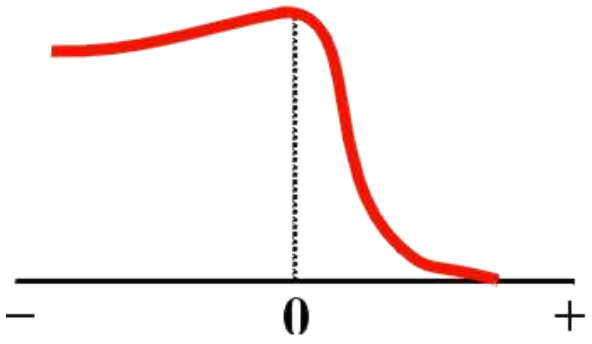
Given the enormous range of population sizes, the range of heterozygosities is too small

Tomoko Ohta tried to explain the deviations of the observed range of heterozygosity from predictions:

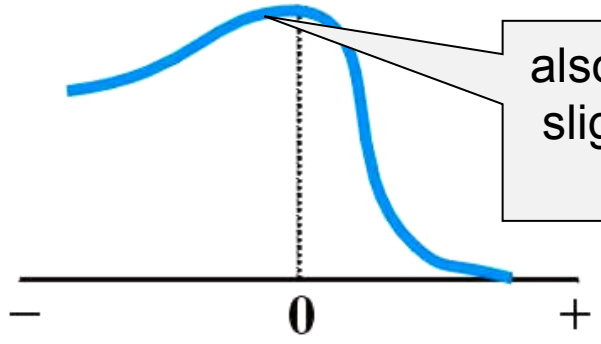
slightly deleterious mutations (SDM)

in small populations
alleles behave as
effectively neutral

neutralism



slightly deleterious mutations



also substitutions of slightly deleterious alleles

Fixation probability of neutral, beneficial, and deleterious mutation:

Eg.:
What is the fixation probability of a mutation in a population of N_e = 1000?

neutral mutation (s = 0): P = 0,05%

advantageous mutation (s = 0,01): P = 20%

advantageous mutation (s = 0,001): P = 2%

deleterious mutation (s = -0,001): P = 0,004%

as $s \rightarrow 0$ higher „neutrality“

We can conclude that

1) Not all advantageous mutations may be fixed in the population
2) Conversely, with low probability also deleterious mutations can be fixed

What is the fixation probability of a mutation in a population of N_e = 10 000?

neutral mutation (s = 0): P = 0,005%

advantageous mutation (s = 0,01): P = 20%

advantageous mutation (s = 0,001): P = 2%

deleterious mutation (s = -0,001): $P = 2.10^{-17}\%$

in a large population P of an advantegous allele is the same as in a small one but for a deleterious allele $P \rightarrow 0$

Conclusions:

1) In large populations selection plays a much more important role; conversely, with decreasing population size the role of drift is increasing

2) Harmfulness of a mutation is inversely proportional to population size: the more it approaches zero, the larger the population may be for allele fixation (drift exceeds negative selection) and vice versa: the stronger selection against an allele, the smaller the population must be to allow drift to play a substantial role

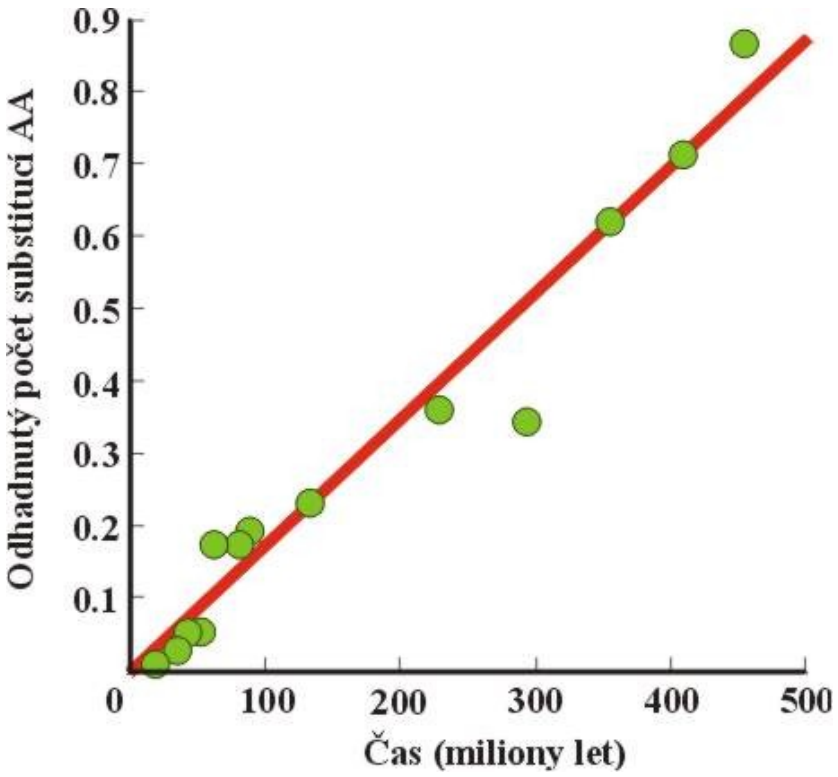
3) This means that in small populations slightly deleterious mutations behave as effectively neutral

MOLECULAR CLOCK

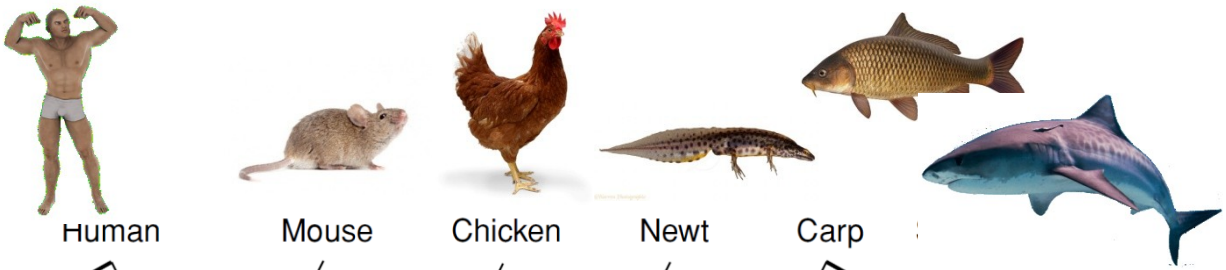
Zuckerkandl & Pauling (1962-65)

AA and/or nucleotide substitution rate is constant

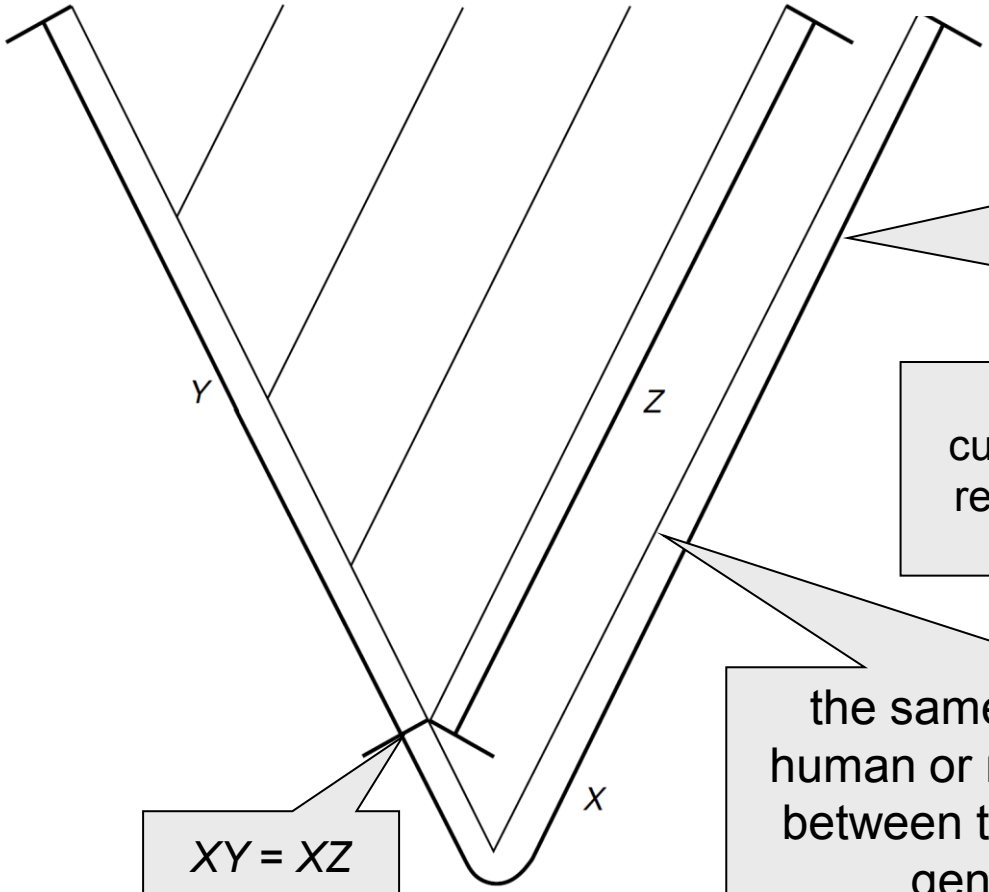
effect of generation time:
dependency on absolute or generation time?



AA sequences of the α-chain of hemoglobin of 6 vertebrate species:



Human Mouse Chicken Newt Carp



despite being morphologically more similar, the distance of the shark from humans is the same as from the carp

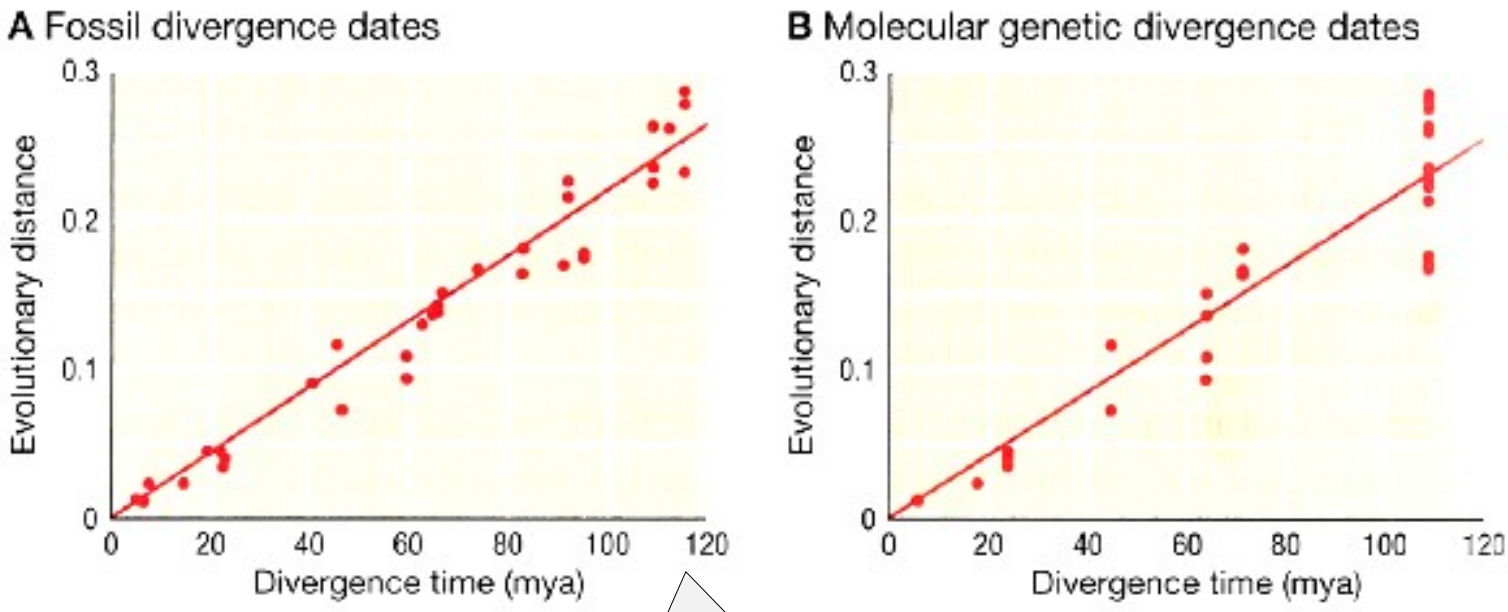
$\Rightarrow$ AA differences are cumulating constantly in time regardless of the phenotypic evolution

the same distance between the human or mouse from the shark as between the carp and shark $\Rightarrow$ no generation time effect

$XY = XZ$

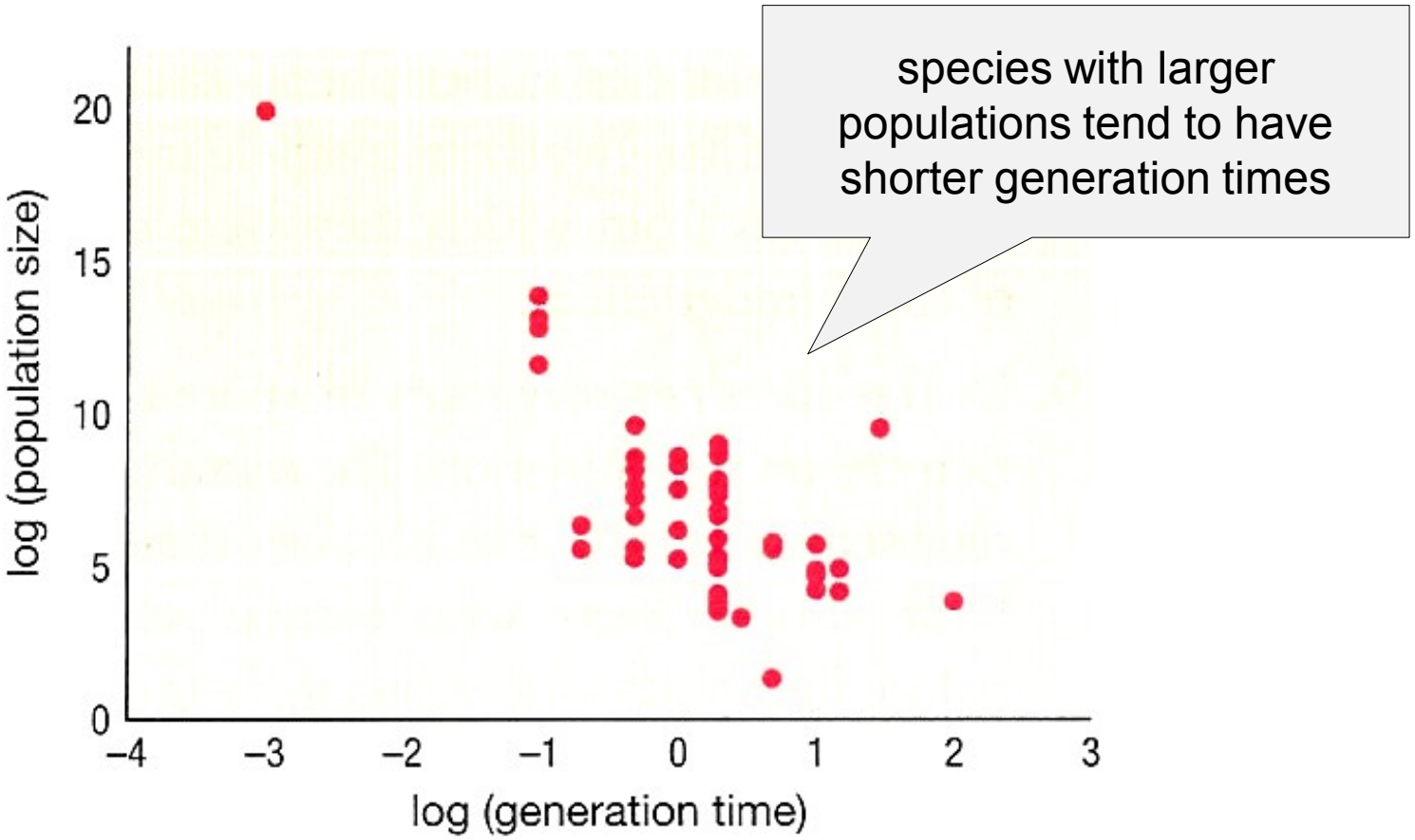
Generation or absolute time?

Accumulation of neutral substitutions in placental mammals:



both dating methods show almost constant rate independent of generation time

Population size and generation time:

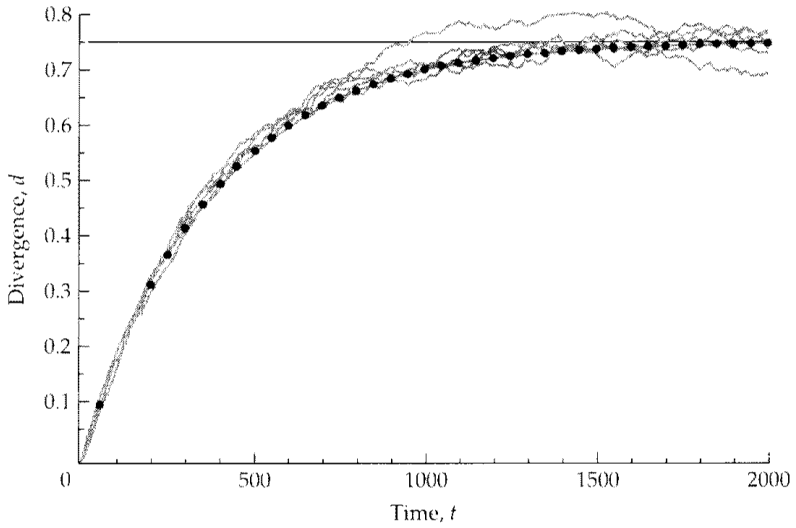


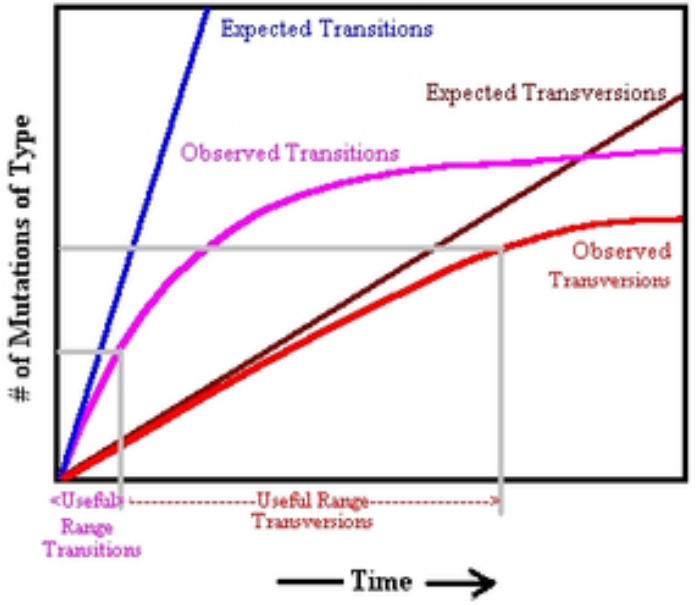
⇒ potential explanation of absolute time dependence: in smaller populations also slightly deleterious alleles are fixed

But molecular clock does not „tick“ with the same pace in different groups

eg. cetaceans < „artiodactyls“< primates < murine rodents
in primates Old World monkeys > „apes“ > humans

Problem of sequence saturation:



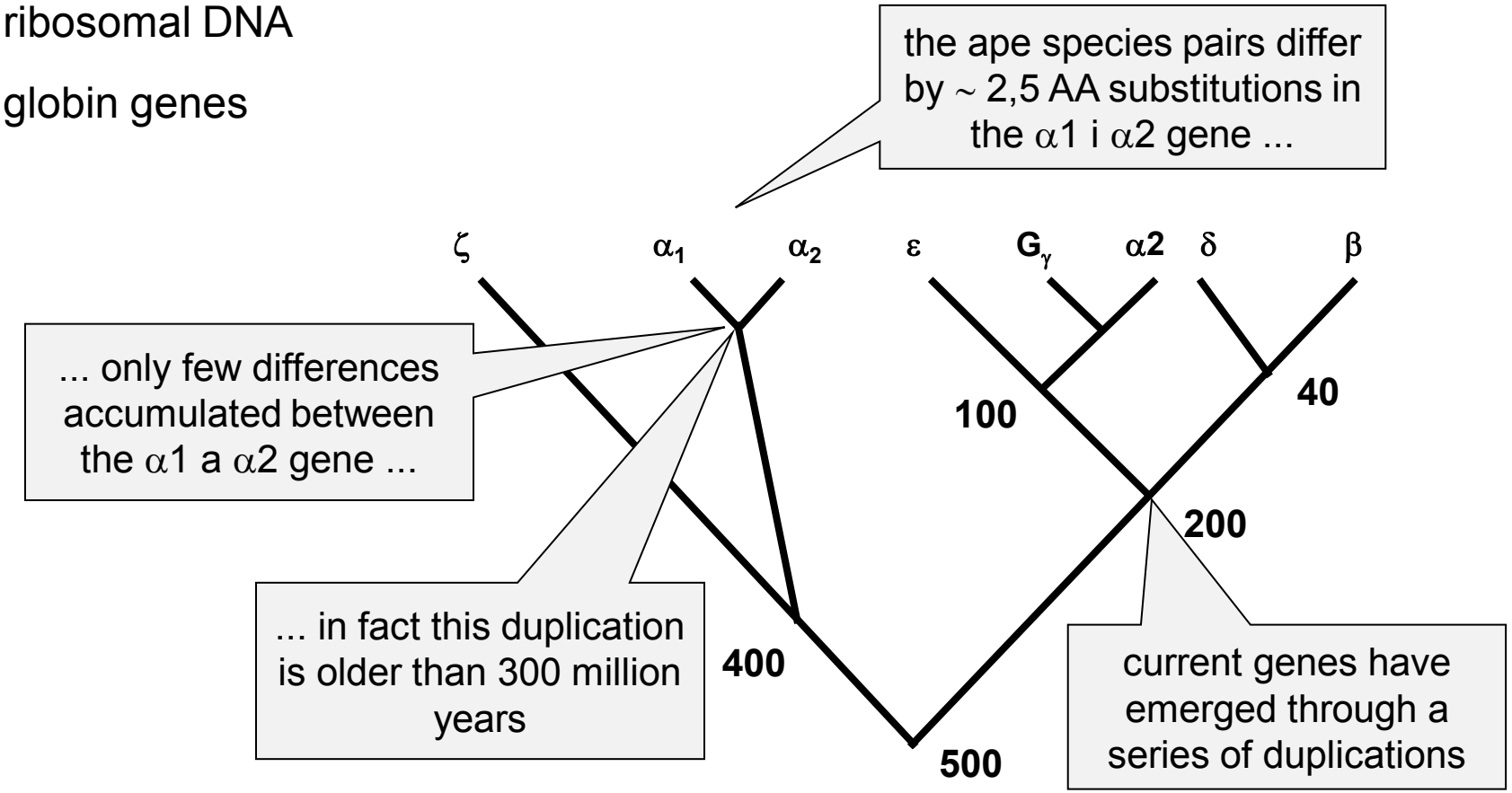


→ using an appropriate evolutionary model („straightening“ of the curve)
relaxed molecular clock method

CONCERTED EVOLUTION AND MOLECULAR DRIVE

ribosomal DNA

globin genes



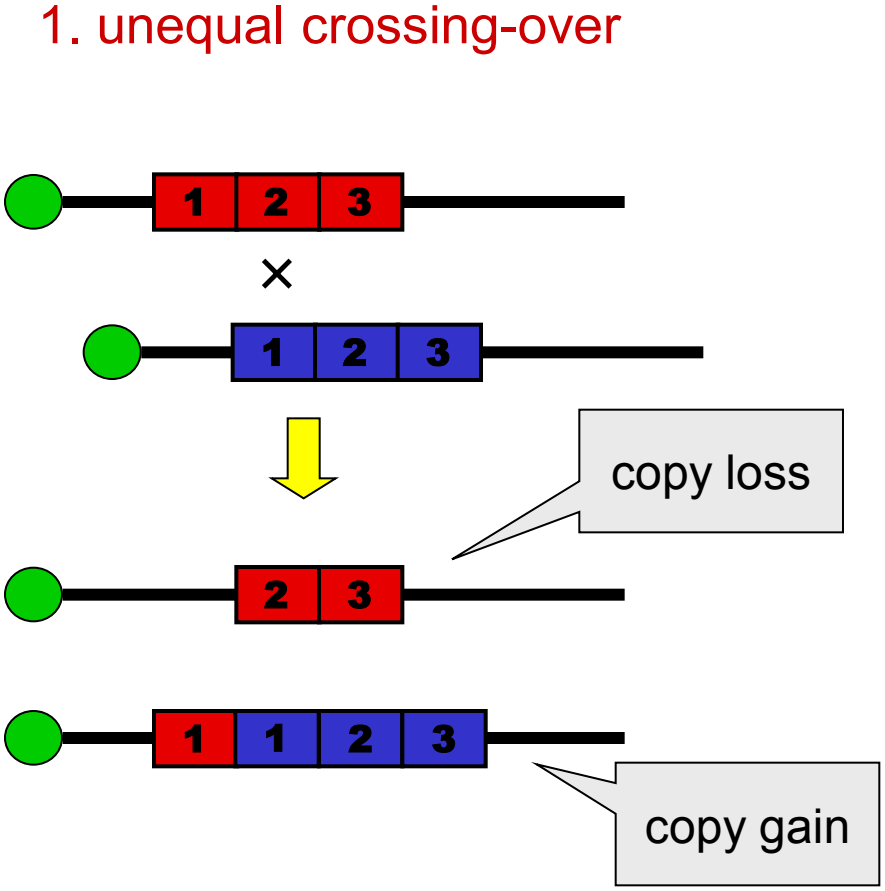
⇒molecular clock is invalid in this case, the genes do not evolve independently – the evolution is <u>concerted</u>

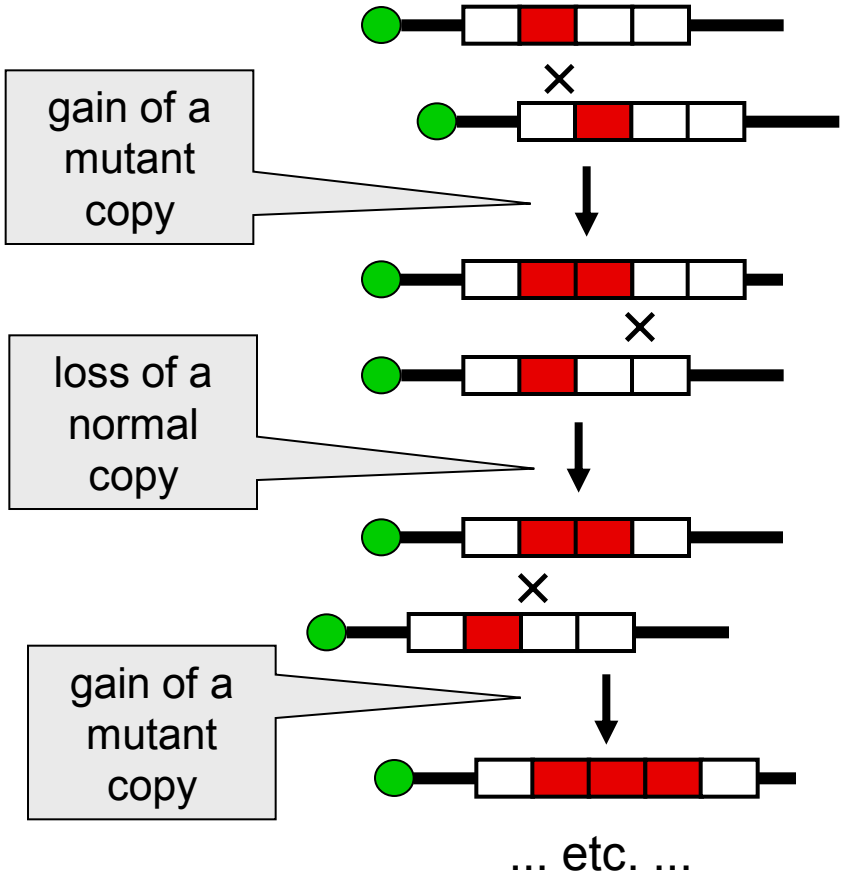
Gabriel Dover (1982): Molecular drive

mechanism different from selection and drift

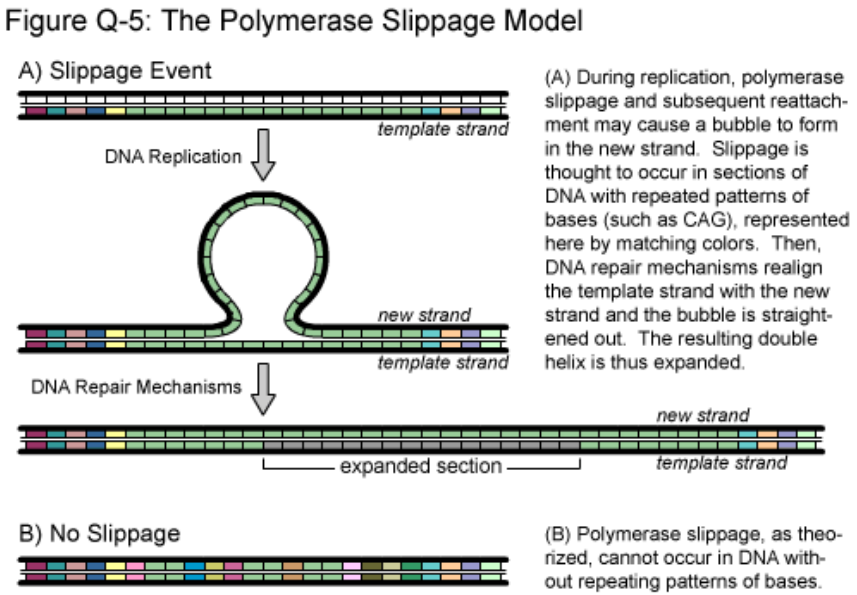
Mechanisms of concerted evolution:

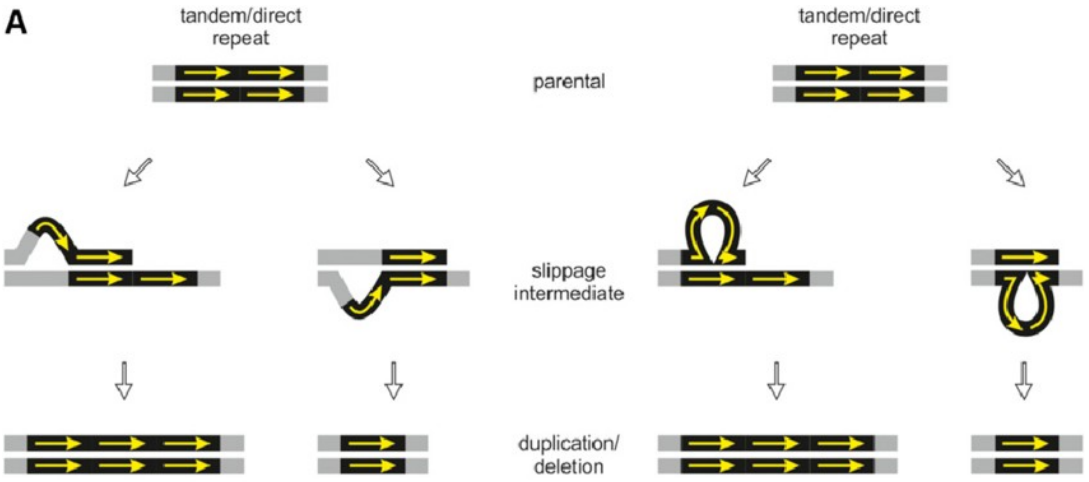
1. unequal crossing-over



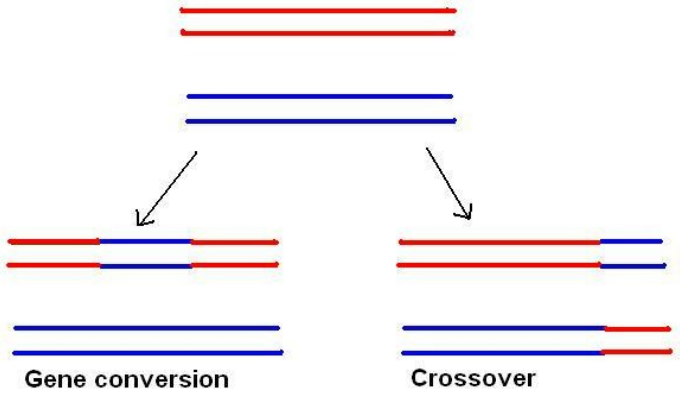


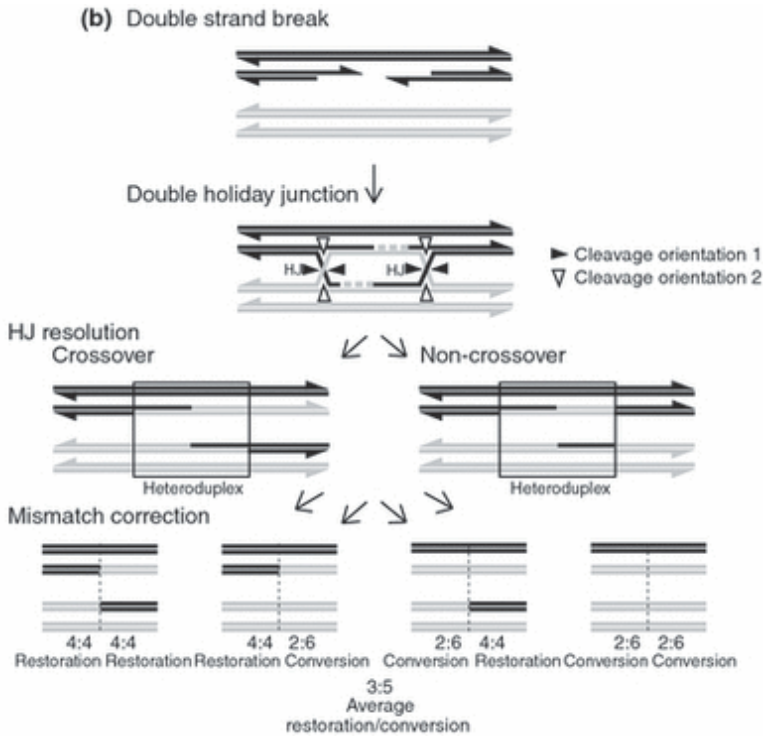
2. replication slippage





3. gene conversion





Conclusions:

a consequence of unequal crossing-over and slippage = <u>change of a copy number</u>

a consequence of unequal c-o and gene conversion = <u>sequence homogenization</u>