The neutral theory of molecular evolution

The neutral theory of molecular evolution

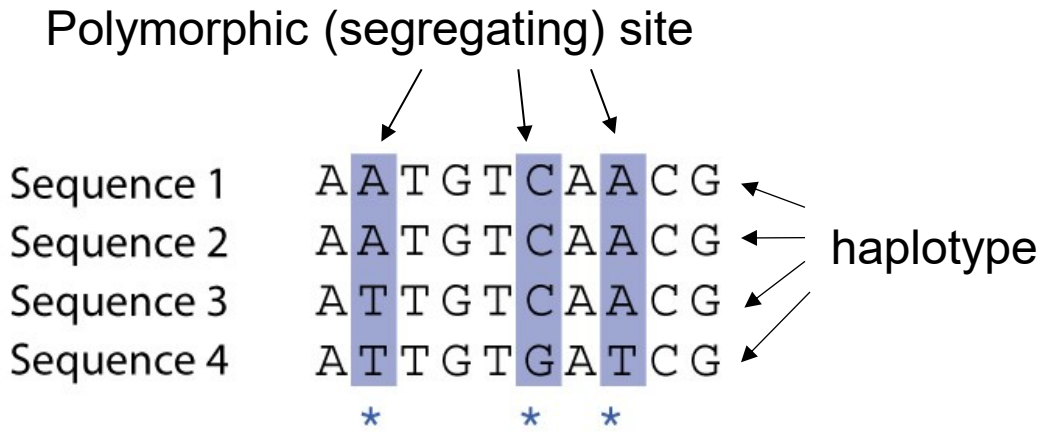
- Describes the rate of molecular evolution and levels of genetic variation if mutations are neutral and their frequency is affected only by genetic drift.

Mottoo Kimura

Genetic diversity - genetic polymorfism (θ)

For neutral sequence:

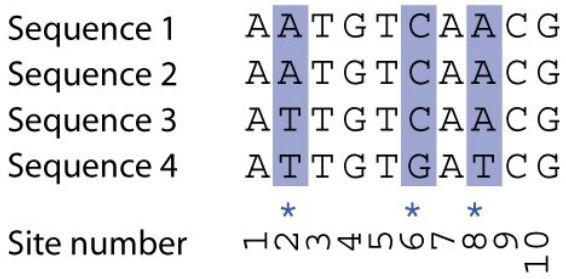
$$\theta = 4N_e\mu$$



- Levels of genetic diversity in population increases with increasing population size and mutation rate

Empirical estimates of genetic diversity

Nucleotide diversity (average heterozygosity) (π)



Nucleotide diversity (π):

1 A A T G T C A A C G
2 A A T G T C A A C G $d_{12} = 0$

1 A A T G T C A A C G
3 A T T G T C A A C G $d_{13} = 1$

2 A A T G T C A A C G
3 A T T G T C A A C G $d_{23} = 1$

1 A A T G T C A A C G
4 A T T G T G A T C G $d_{14} = 3$

2 A A T G T C A A C G
4 A T T G T G A T C G $d_{24} = 3$

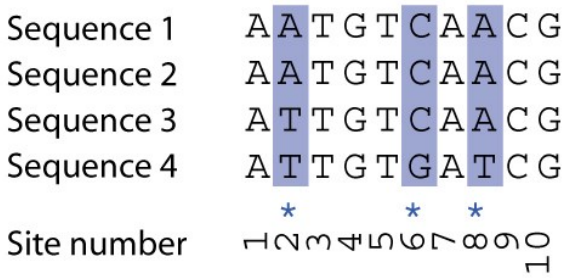
3 A T T G T C A A C G
4 A T T G T G A T C G $d_{34} = 2$

$$\Sigma d_{ij} = 0 + 1 + 3 + 1 + 3 + 2 = 10$$

Number of pairs of sequences compared = $[n(n-1)]/2 = [4(3)]/2 = 6$
$\hat{\pi}$ = 10 differences/6 pairs = 1.67 average pairwise differences
$\hat{\pi}$ = 1.67 avg. differences/10 sites = 0.167 pairwise differences per site

π = 0.01 on average one polymorphic site per 100 bp

Proportion of polymorphic sites (θ_W)



Segregating sites (S and p_S):

Sites 2, 6, and 8 have variable base pairs among the four sequences (columns marked with *). These are segretating sites. Therefore, for these sequences $S = 3$ segregating sites and $p_S = 3/10 = 0.3$ segregating sites per nucleotide site examined.

$$\theta = S / n / H_{k-1}$$

$$H_{k-1} = 1 + \frac{1}{2} + \frac{1}{3}$$

$$\theta = 3 / 10 / 1{,}83 = 0{,}164$$

S … number of segregating sites

n … number of nucleotides in the sequence

k … number of sequences

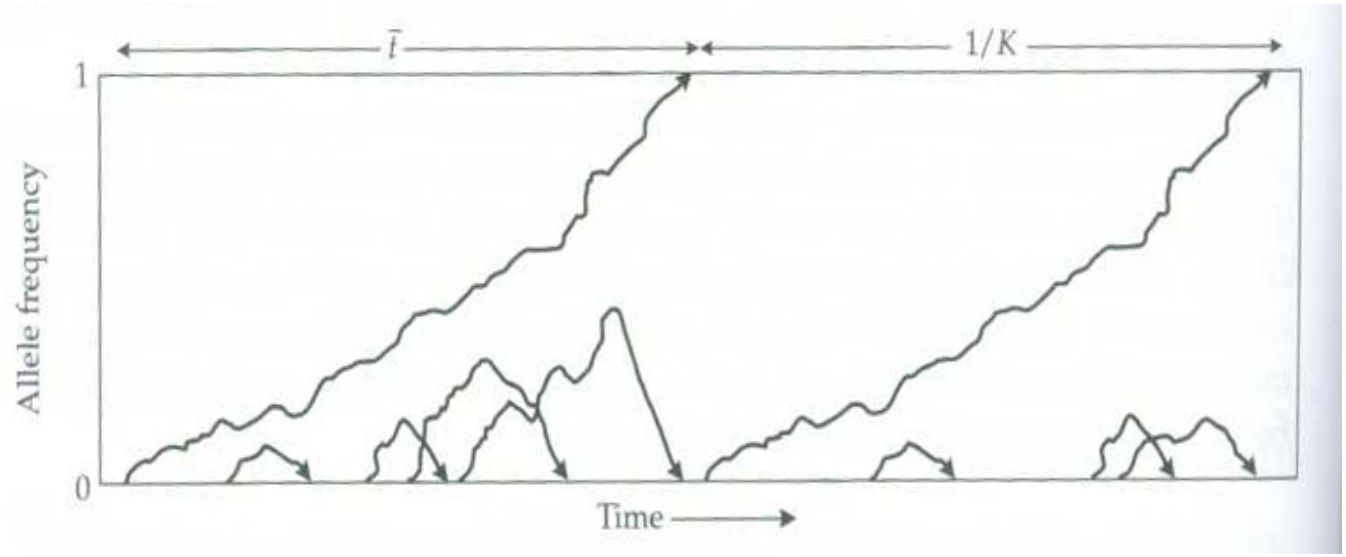
H_{k-1}… harmonic number

Program: DnaSP

Mutation vs. substitution rate

Substitution rate

Rate of fixation of new mutations.



Mutation rate

Rate of origin of new mutations.

Substitution rate

= number of new mutations x probability of their fixation

$$k = 2N_e\mu \frac{1}{2N_e} = \mu$$

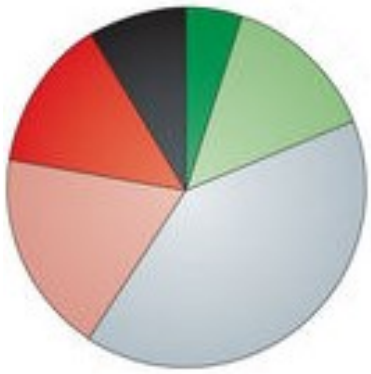
Probability of their fixation

Number of new mutations

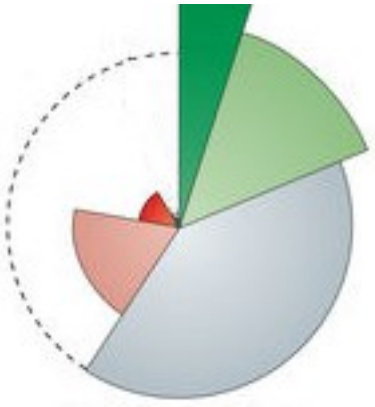
N_e … effective population size

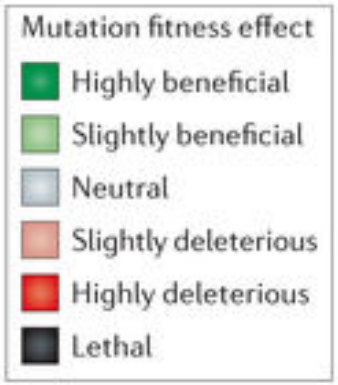
μ … mutation rate

Mutation rate



Substitution rate

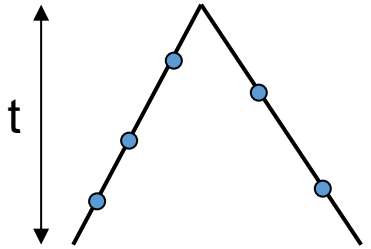




Molecular clocks

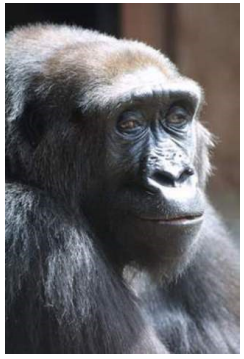
- Substitution rate of neutral mutations is given only by mutation rate.
- If mutation rate is constant over time, subtitution rate is also constant.

Genetic divergence



= substitution rate x time

$D = 2\mu t$



- Number of substitutions between species.
- For neutral substitutions increases lineary with time

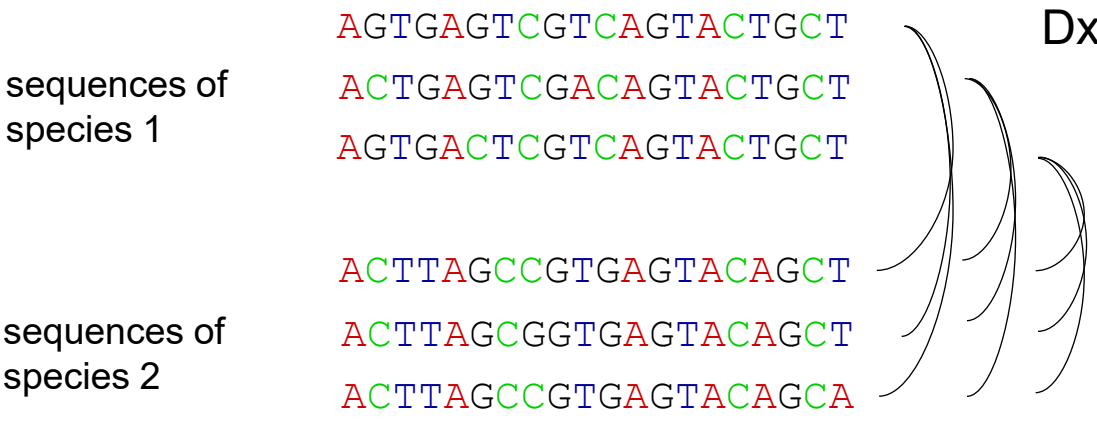
Estimation of genetic divergence

sequence 1 AGTGAGTCGTCAGTACTGCT

sequence 2 ACTTAGCCGTGAGTACAGCT

$$D = \frac{5}{20} = 0.25\ (25\%)$$

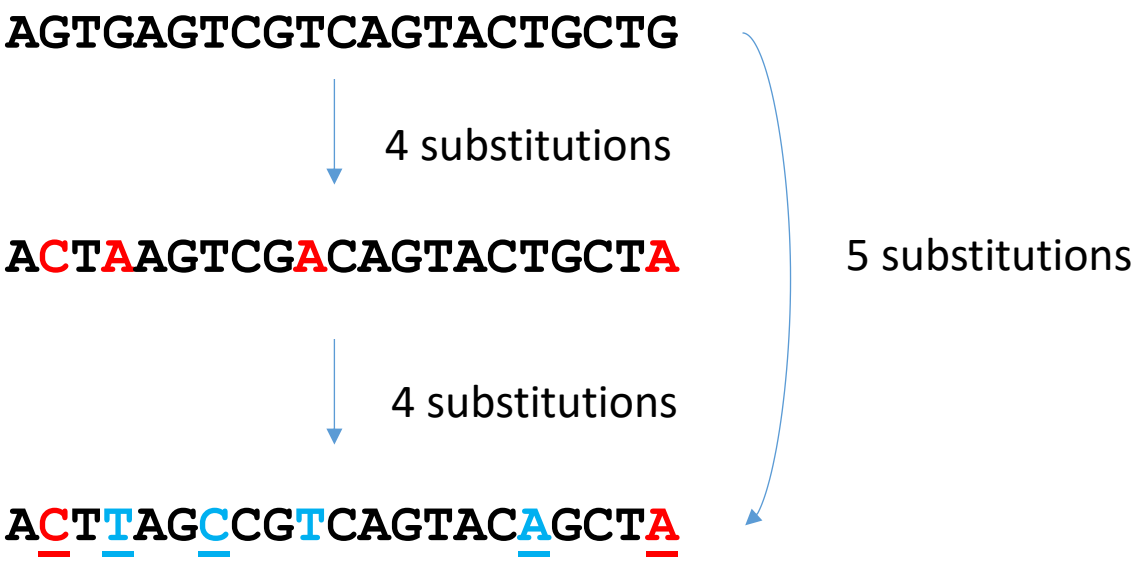
Proportion of nucleotides differing between two sequences (p-distance).



Dxy, avergage pairwise divergence

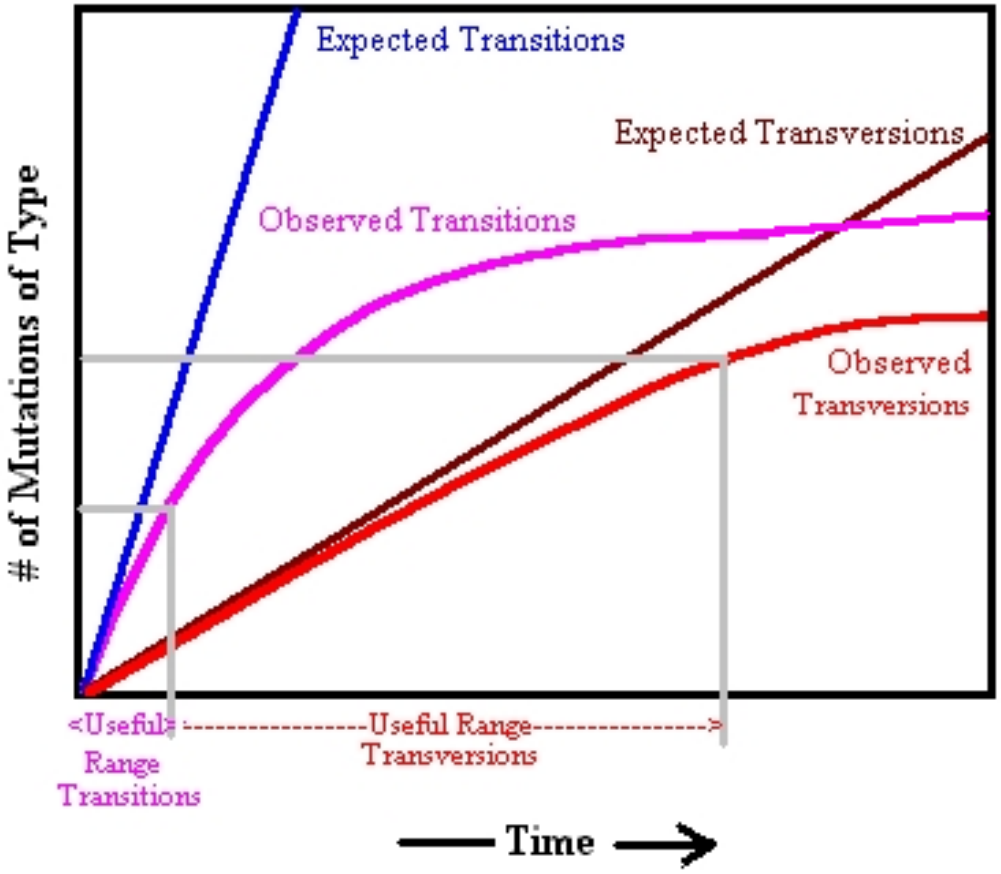
Genetic saturation

- The result of multiple substitutions at the same site in a sequence.



Genetic saturation

- Observed number of substitutions is underestimated at higher divergences.



Nucleotide substitution models

Correction for multiple substitutions and estimates of the true genetic divergences.

Jukes & Cantor (1 parameter) model
-all substitutions the same probability (α) and all nucleotides the same frequency

Kimura (2 parameter) model
-probability of transitions higher than probability of transversions

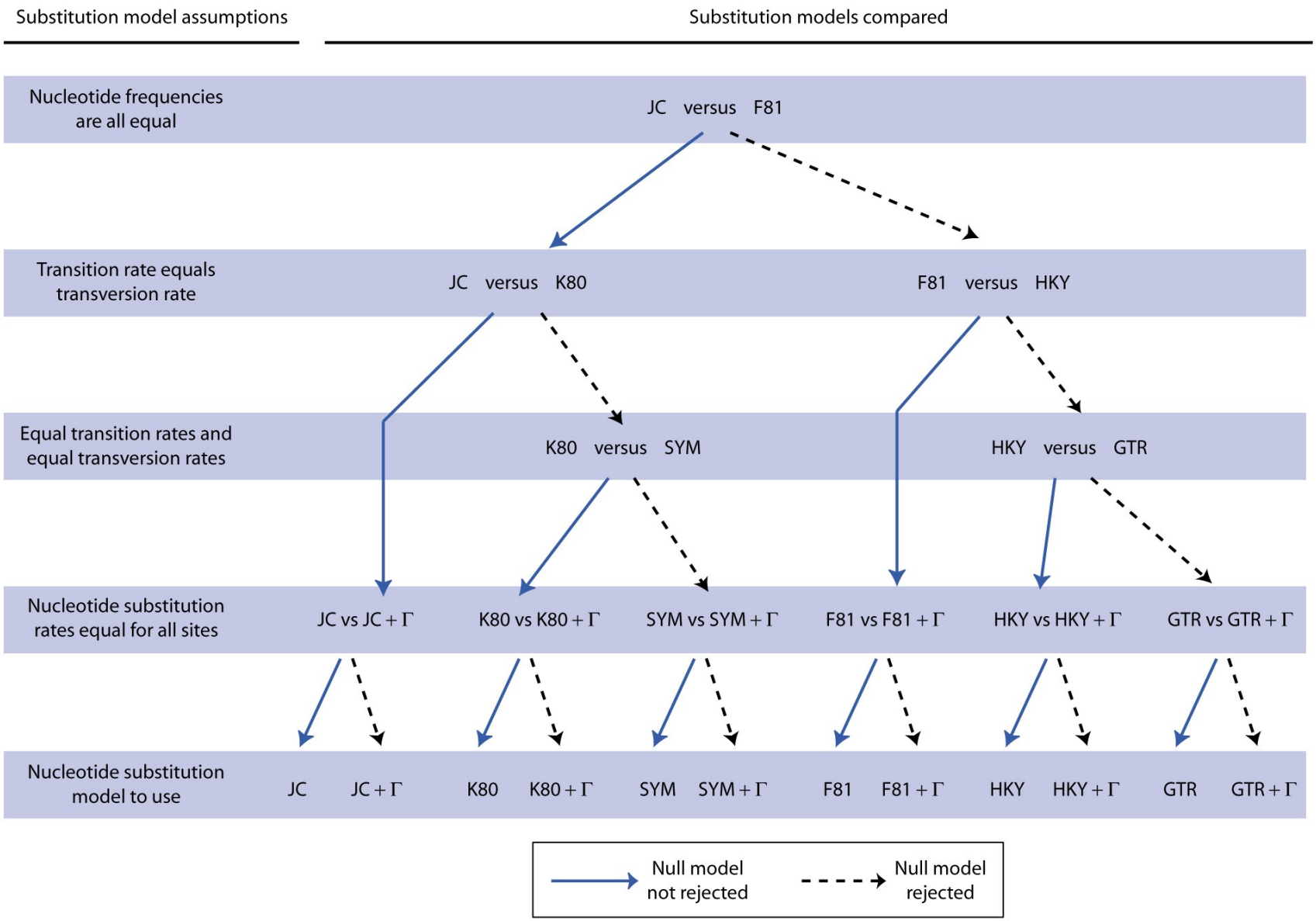
Felsenstein 81 (4 parameter) model
-Different frequencies of nucleotides

HKY-Hasegawa,Kishino,Yano (5 parameter) model
-Different probabilities of transitions and transversions and different nucleotide frequencies

HKY + Γ
-Gama parametr can model different substitution rate at different sites

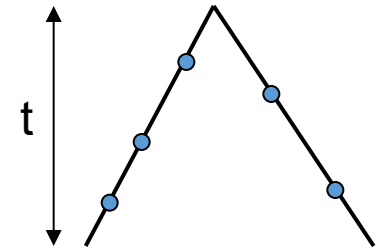
Program ModelTest

Nucleotide substitution models

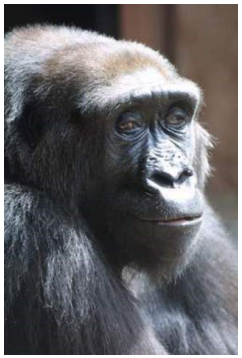


Molecular clocks
dating the divergence time

- Divergence time (t) can be calculated if we know the genetic distance (D) and substitution rate (μ)

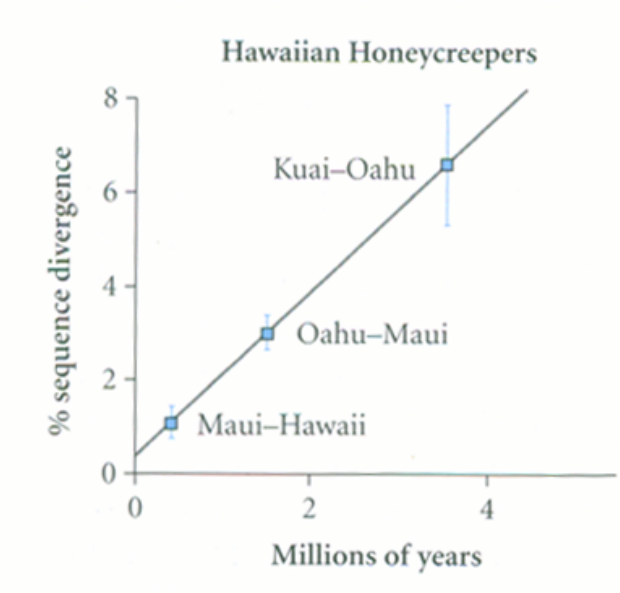


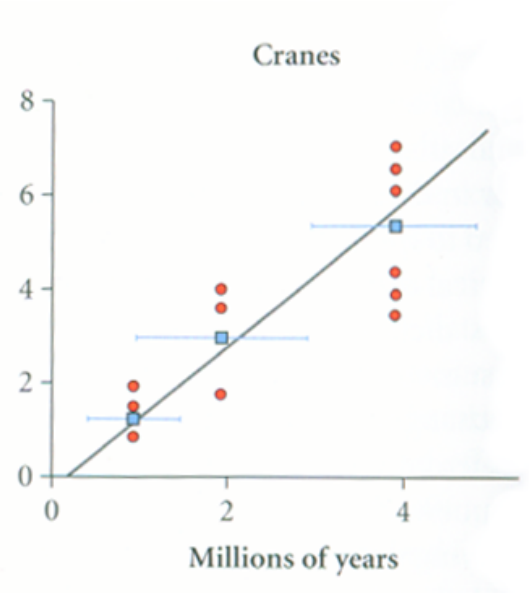
$$D = 2\mu t$$



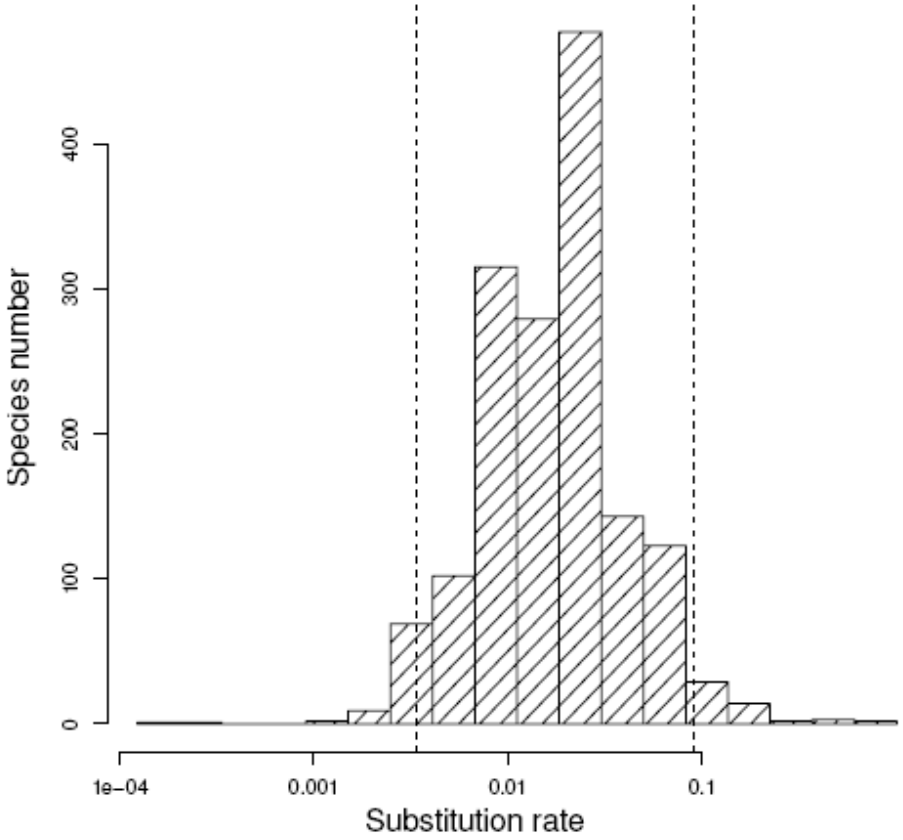
Molecular clocks
dating the divergence time

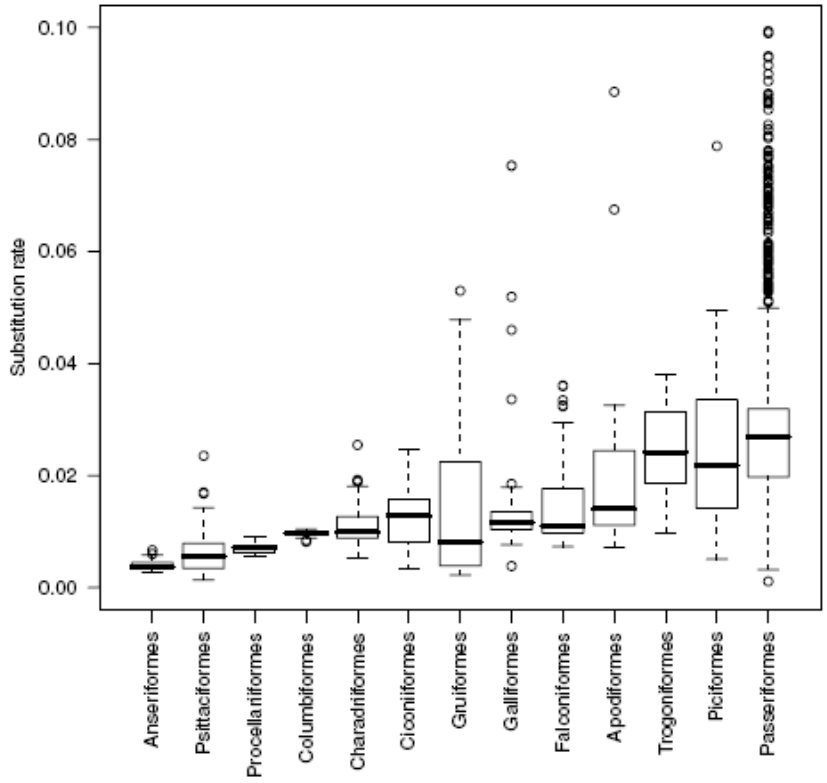
- Estimation of substitution rate (**calibration of molecular clocks**).
- Fossil records, known geological events
- cytochromu b sequence (mt DNA), substitution rate in mammals and birds cca 0,01. **2% divergence ~ 1 mil let.**





Substitution rate varies among species



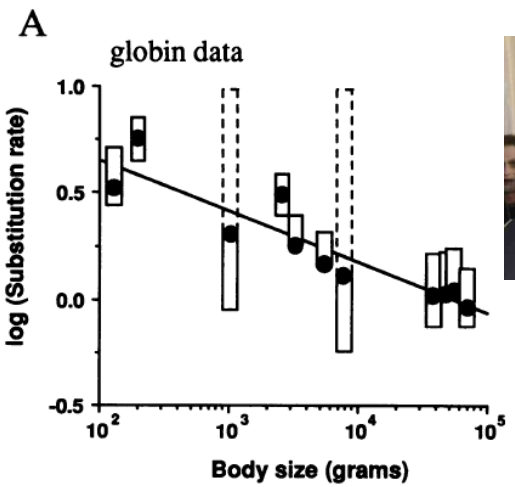


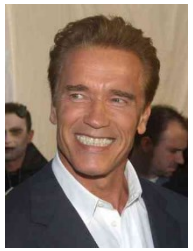
Substitution rate for cytochrome b in birds

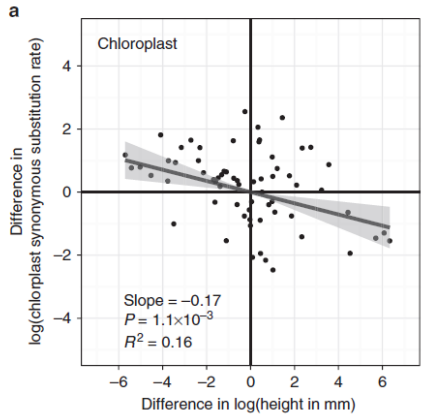
Causes of variation in substitution rate

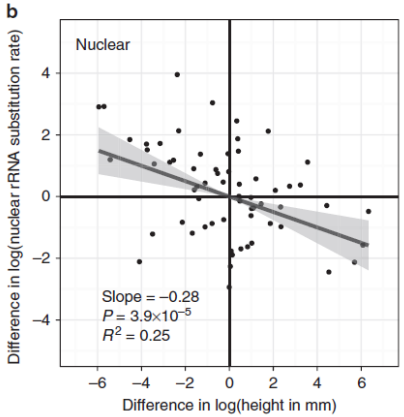
Body size

Organisms with smaller body size have higher substitution rate.



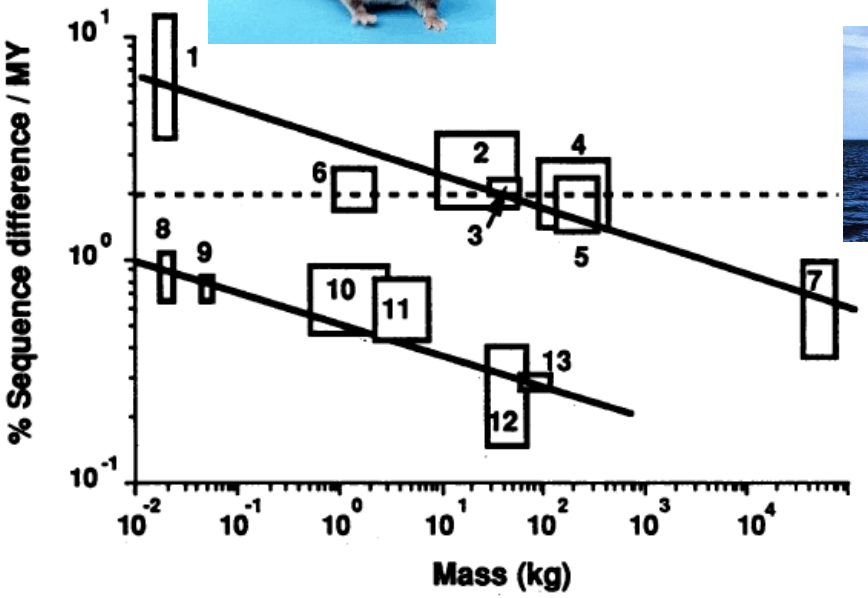






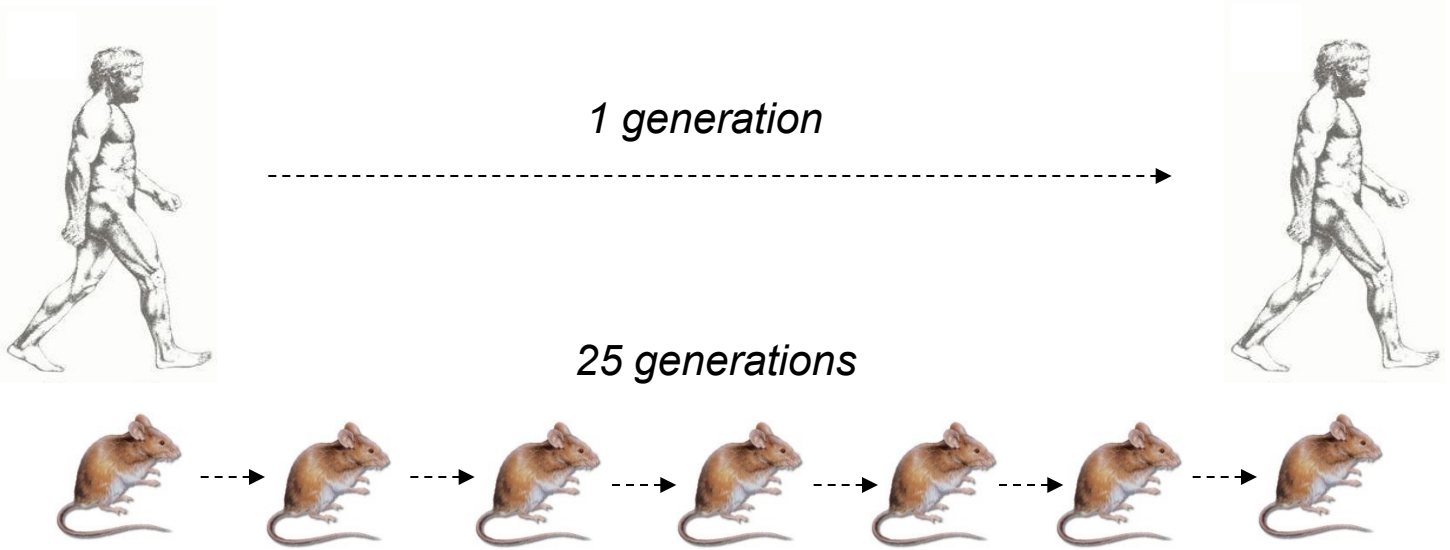
Metabolic rate

- Species with higher metabolic rate (i.e. smaller body size) have higher substitution rate.
- Homoiothermic vertebrates have higher substitution rate than poikilotermic vertebrates.



Generation time

- Species with shorter generation time (usually smaller body size) have higher substitution rate (per year). More cell divisions in the germline per year -> higher mutation rate per year.
- Species with longer generation time, however, have higher substitution rate per generation. They produce gametes longer time -> higher mutation rate per generation.

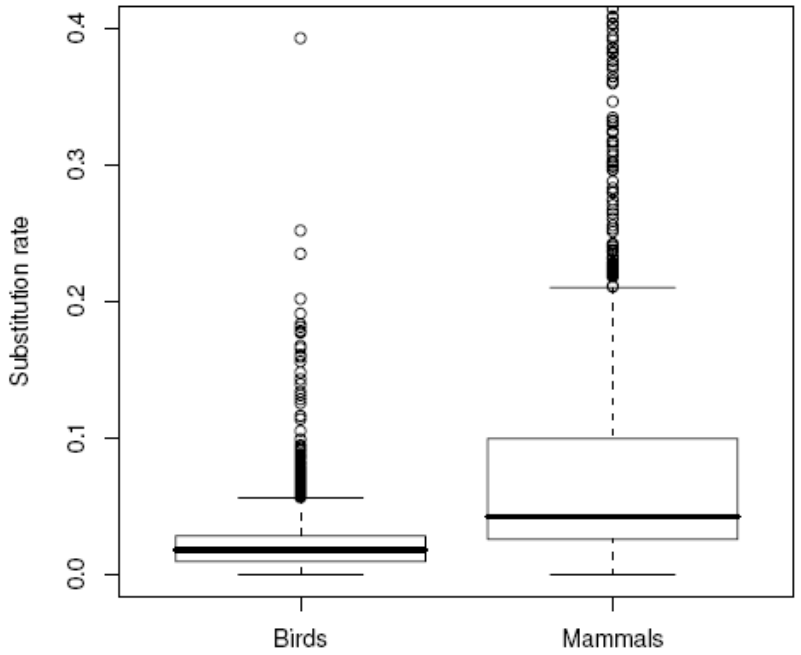


Longevity

- Long-lived species have more efficient DNA repair mechanisms -> lower mutation rate.

Gerbillus nanus
0.741 substitution/site/Myr

Balaenoptera borealis
0.007 substitution/site/Myr



Temperature

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION
SSE
SOCIETY for the STUDY of EVOLUTION

BRIEF COMMUNICATION | Full Access

Temperature predicts the rate of molecular evolution in Australian Eugongylinae skinks

Jeremias Ivan, Craig Moritz, Sally Potter, Jason Bragg, Rust Turakulov, Xia Hua,

First published: 05 September 2021 | https://doi.org/10.1111/evo.14342

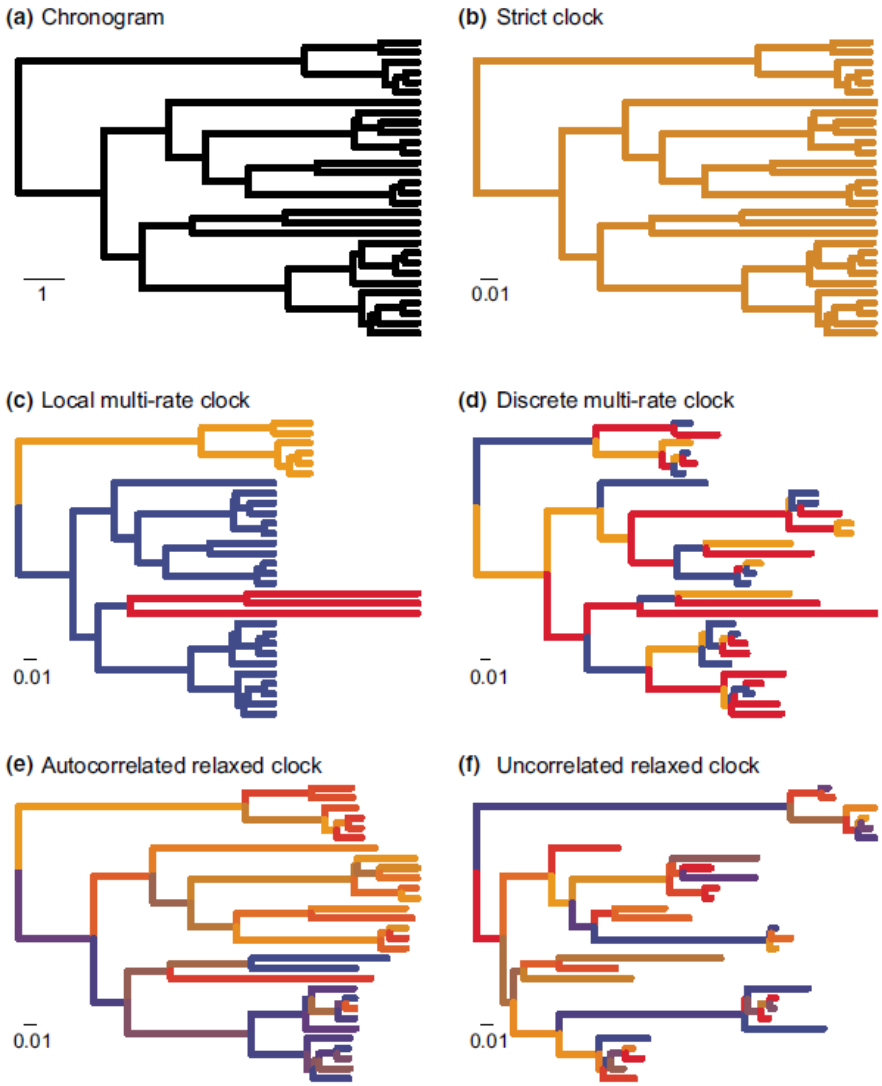
Molecular clocks

Strict clocks

- Substitution rate is homogenous across the phylogeny.

Relaxed clocks

- Variability in substitution rate among lineages.



Ho and Duchene, 2014

The nearly neutral theory of evolution

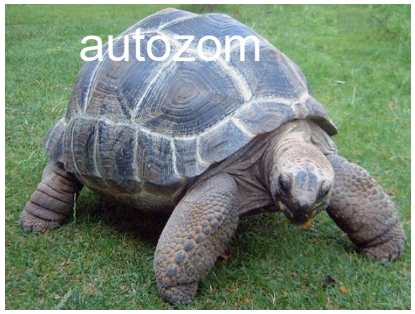
The nearly neutral mutations: $|2s| < \frac{1}{2N_e}$

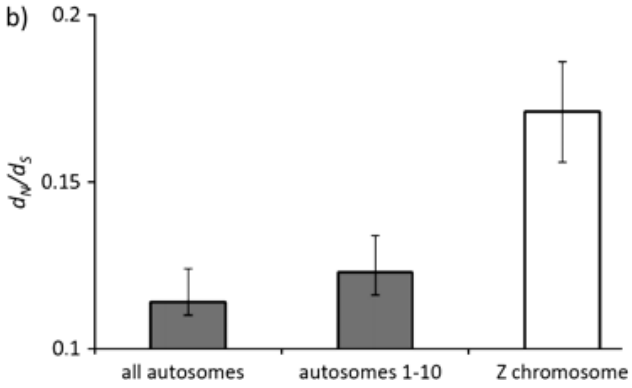
In small populations they behave as neutral (genetic drift), while in large populations as beneficial or disadvantageous (selection).

Tomoko Ohta

- Given that most mutations are disadvantageous, more mutations will be fixed in small populations than in large populations.
- Species with small populations will have higher substitution rate compared to species with large populations.

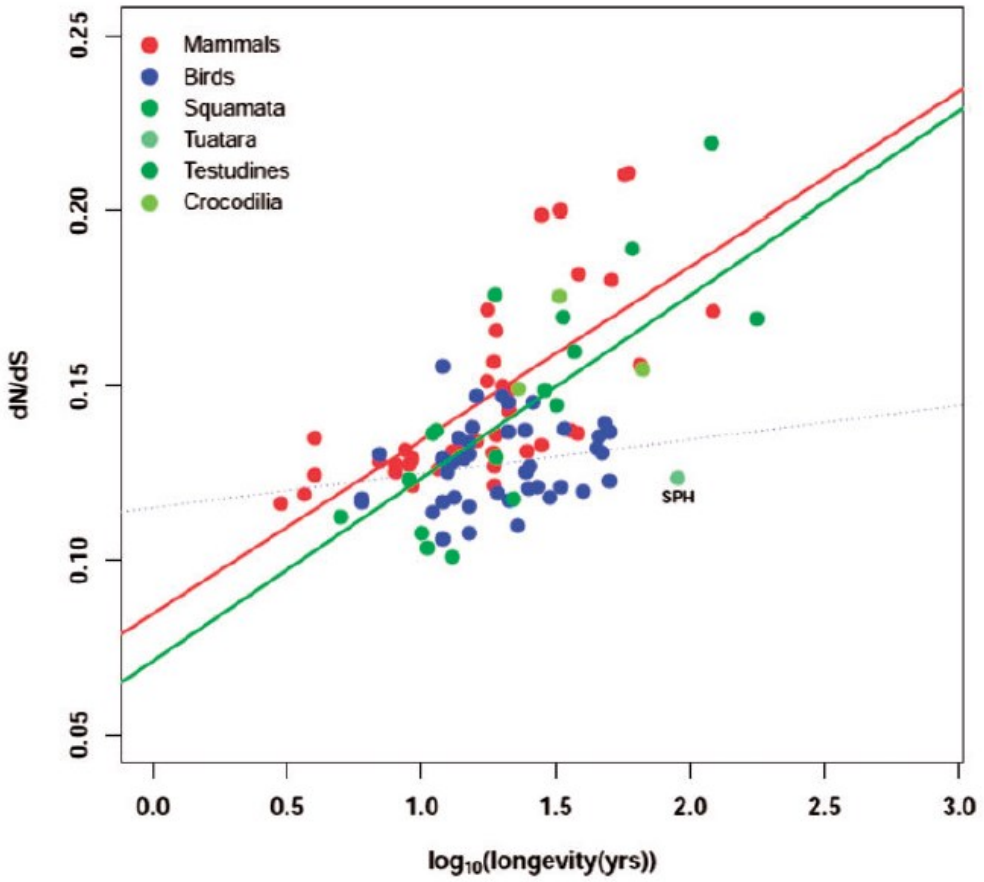
Faster X evolution





Mank et al. 2010

Species with smaller populations (higher longevity) show relatively more non-synonymous substitutions (dN) compared to synonymous substitutions (dS).

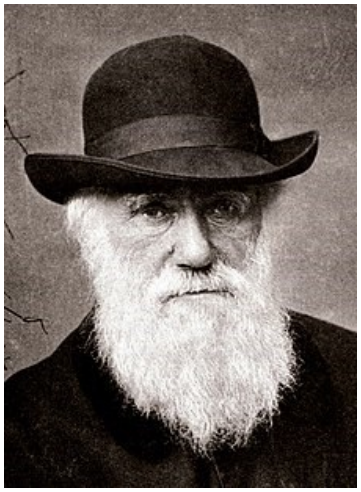


Figuet et al. 2016

Selection

Natural selection

- Main driver of evolution
- Origin of adaptations.

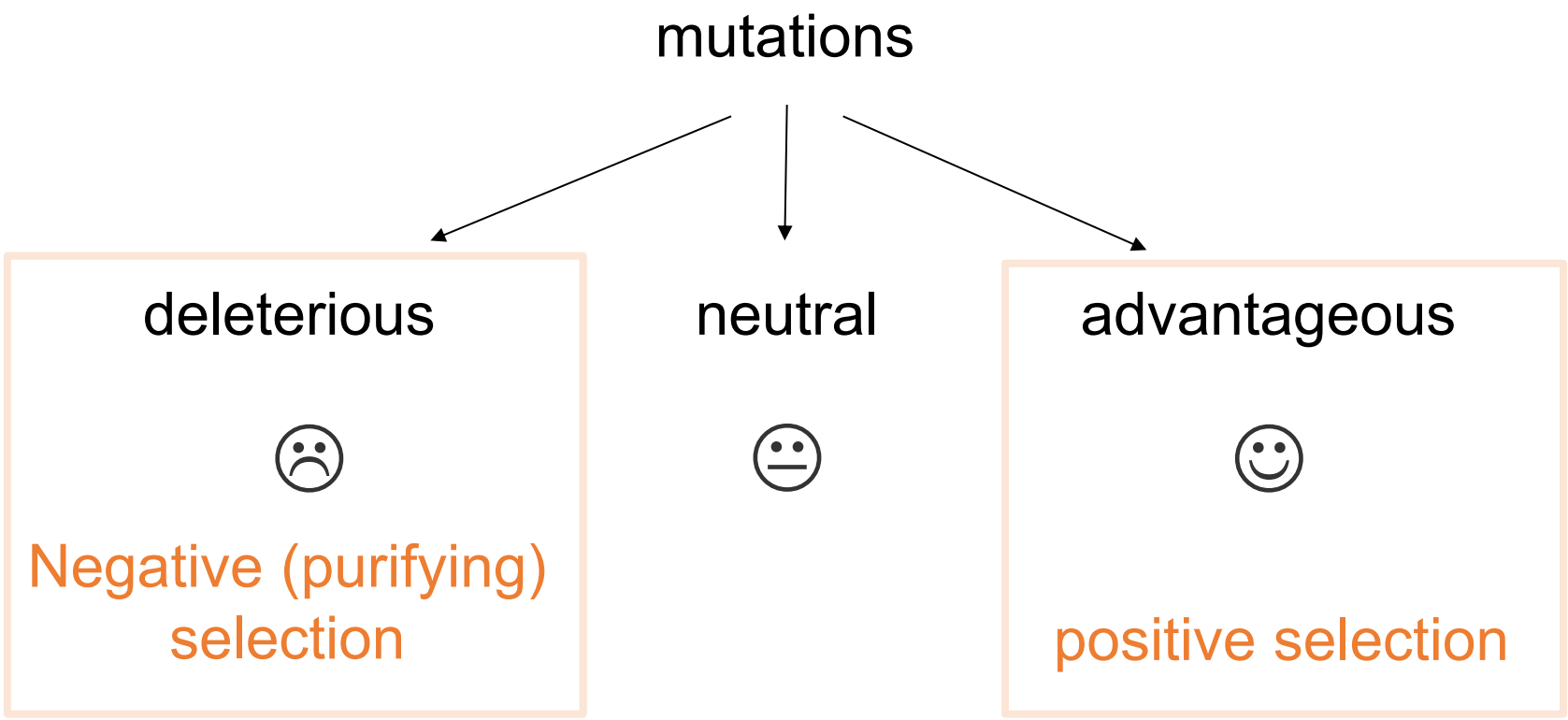


Charles Darwin

Biological fitness

- The ability of an individual to transfer its genes to the next generation (i.e. survive and reproduce).

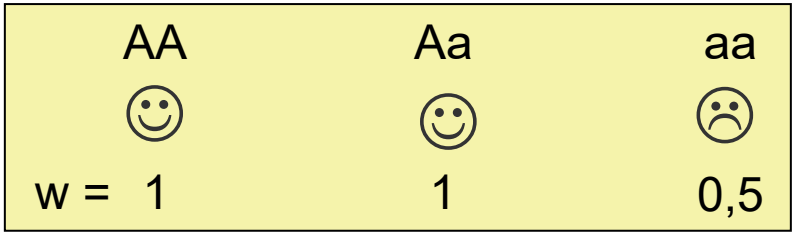
Selection


mutations
deleterious
☹
Negative (purifying) selection
neutral
😐
advantageous
☺
positive selection

Population genetic models of selection

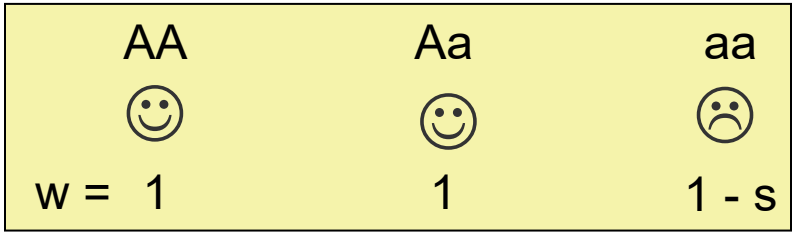
Relative fitness (w)

- Relative differences in fitness between genotypes.
  Maximum w = 1.
  Minimum w = 0.



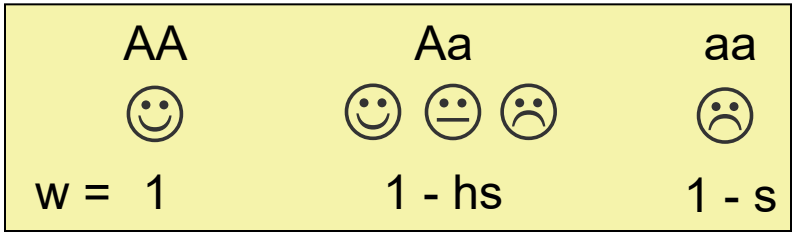
Selection coefficient (s)

- Increase or decrese in fitness in particular genotype(s).



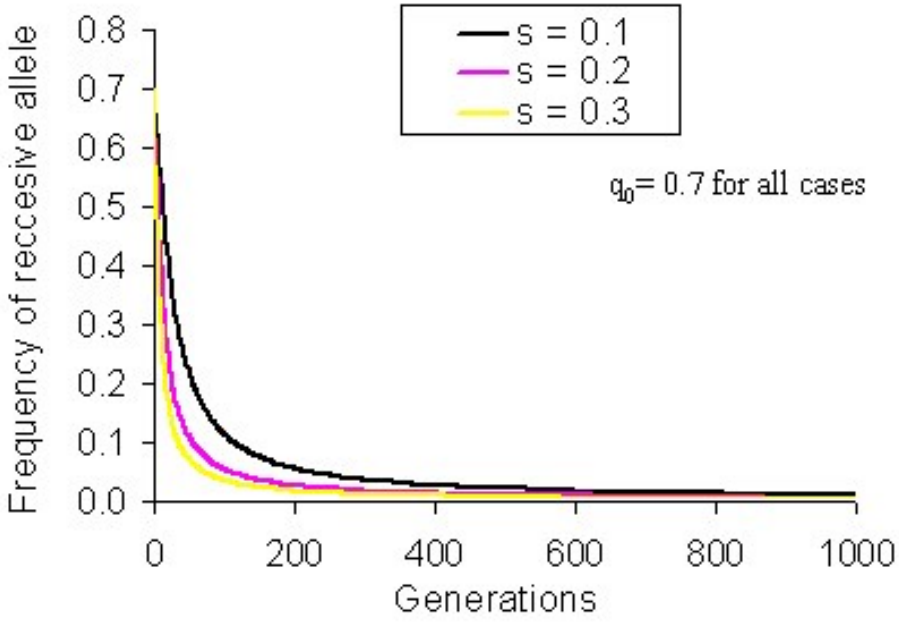
Coefficient of dominance (h)

- Level of dominance between alleles
  h = 0 či 1 complete dominance
  0 < h < 1 incomplete dominance



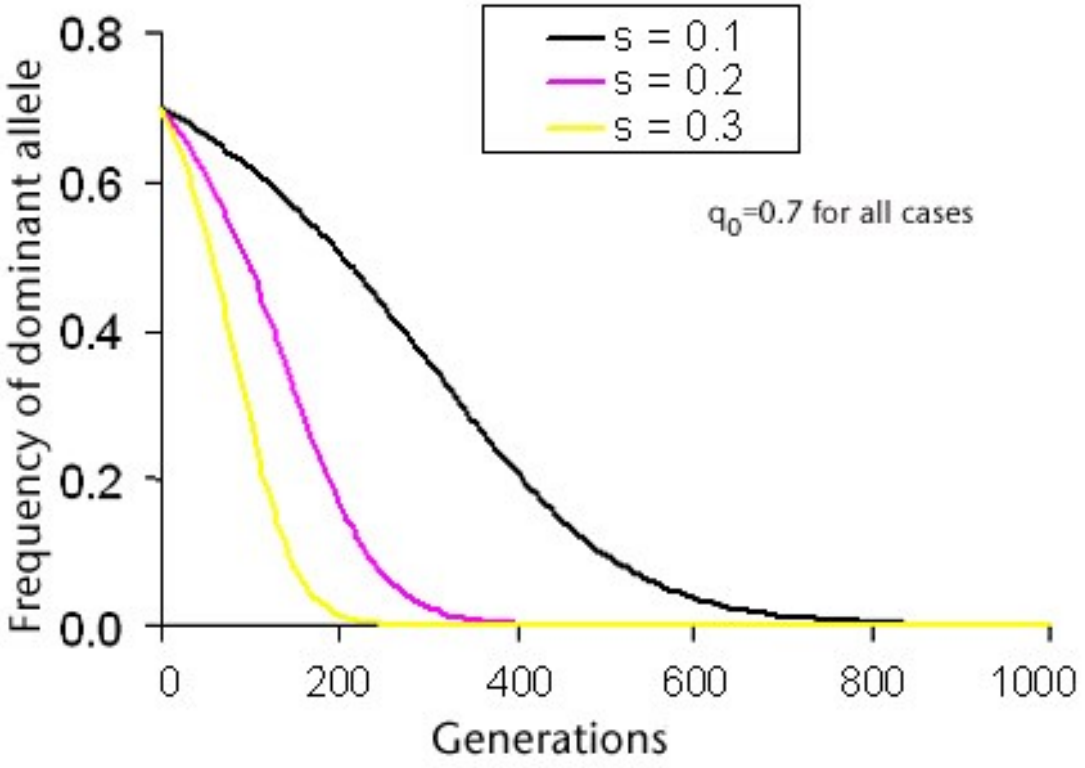
Negative selection against recessive mutations

- Reduces fitness of the deleterious mutation, but the mutation is not eliminated completely from the population if it is recessive (hidden in heterozygotes).
- Human disseases are often caused by recessive mutations (např. cystic fibrosis, phenylketonuria etc.)

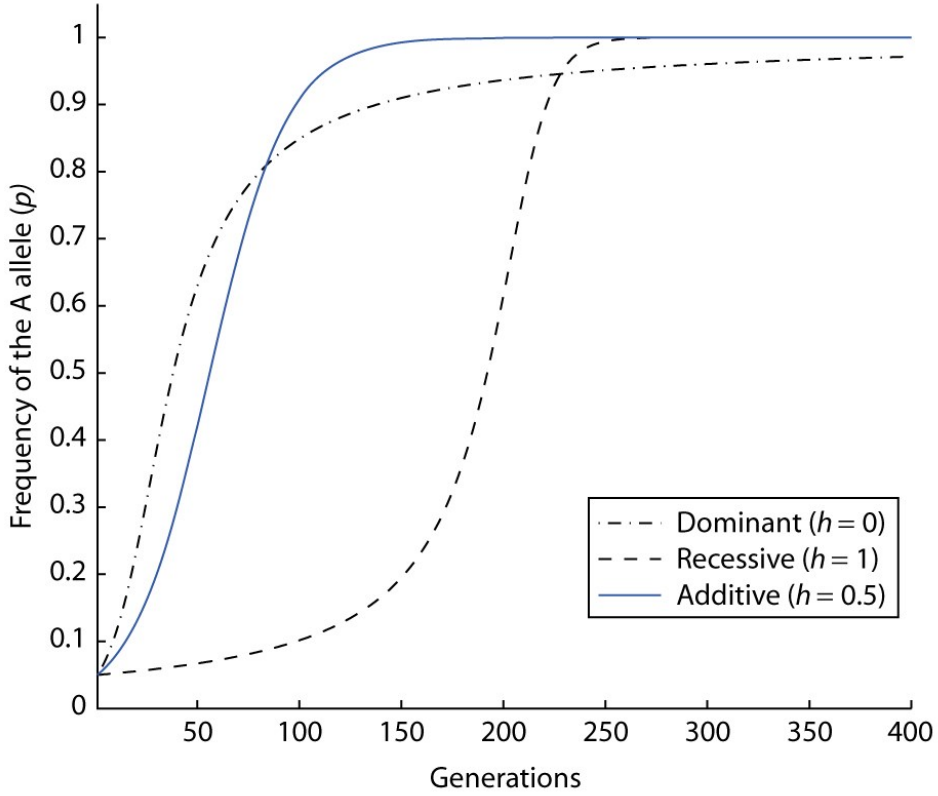


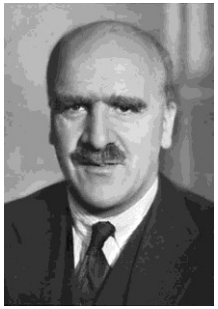
Negative selection against dominant mutations

- Leads to complete elimination of deleterious mutations from the population.
- Diseases caused by dominant mutations usually appear at post-reproductive age (e.g. Huntington disease).



Positive selection

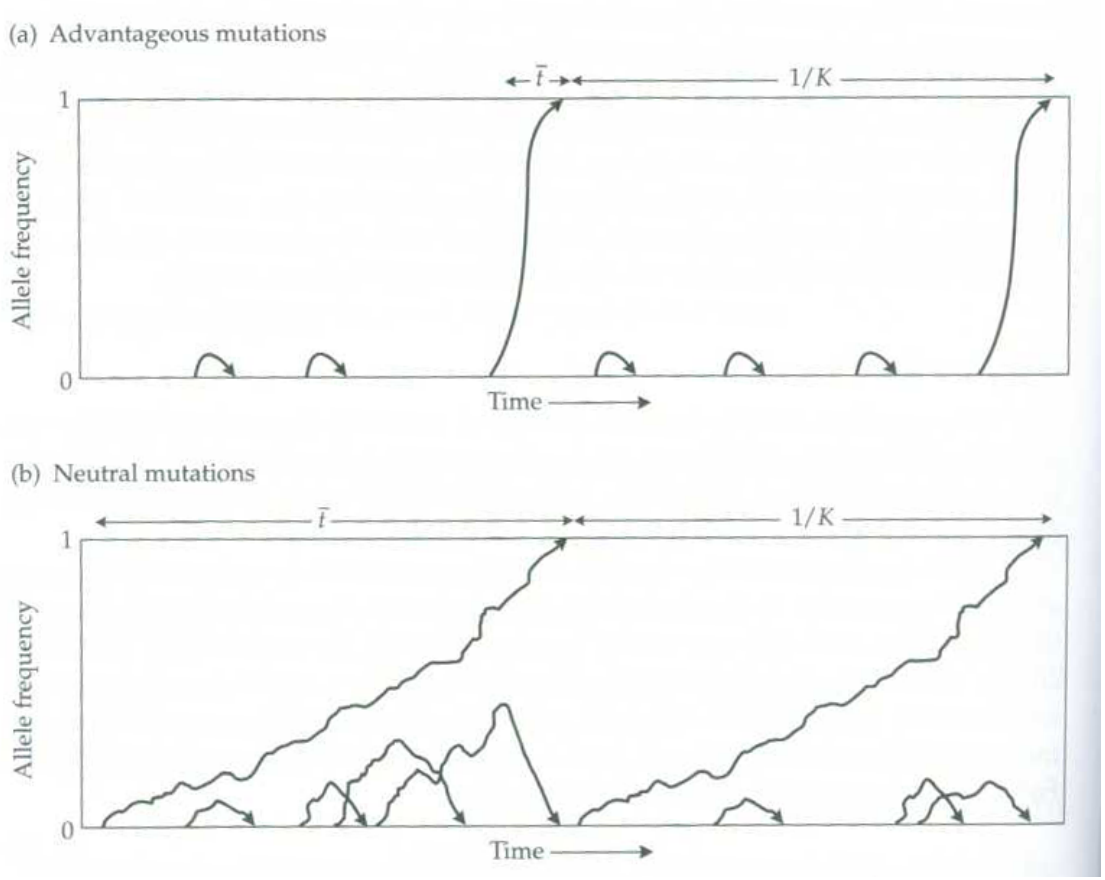




Haldane's sieve.

Dominant advantageous alleles are more likely to fix in the population than recessive alleles.

Time to fixation of beneficial mutation



Advantageous mutation

$t = 2\ln(2N_e)/s$ generations

Neutral mutation

$t = 4N_e$ generations

What is the time to fixation of neutral mutation in human population? Generation time 25 years. N_e 10 000.

$t = 4N_e$ generations

t = 4 . 10.000 generations

t = 40.000 . 25 = 1.000.000 years

What is the time to fixation of beneficial mutation (s = 5%) in human population?

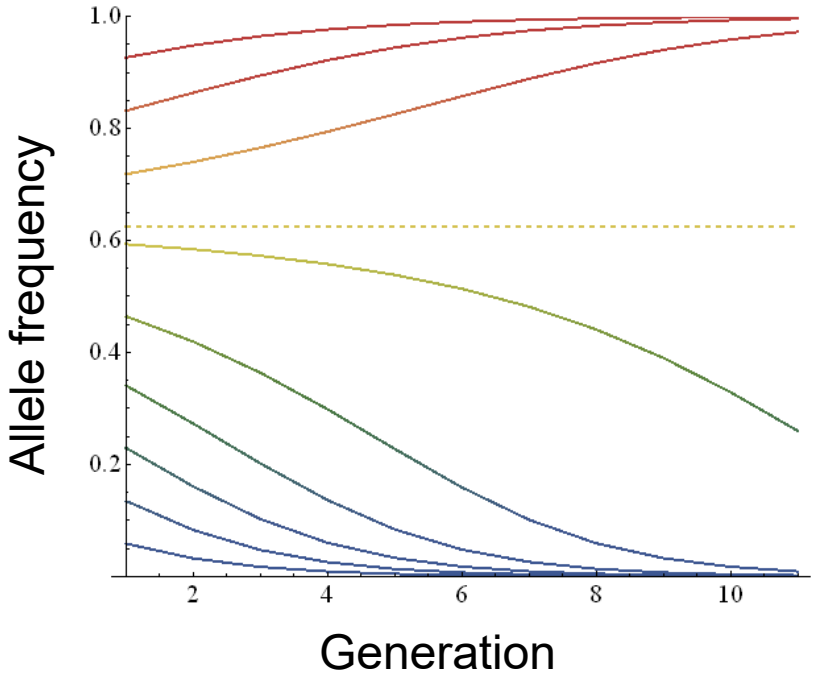
$t = 2\ln(2N_e)/s$ generations

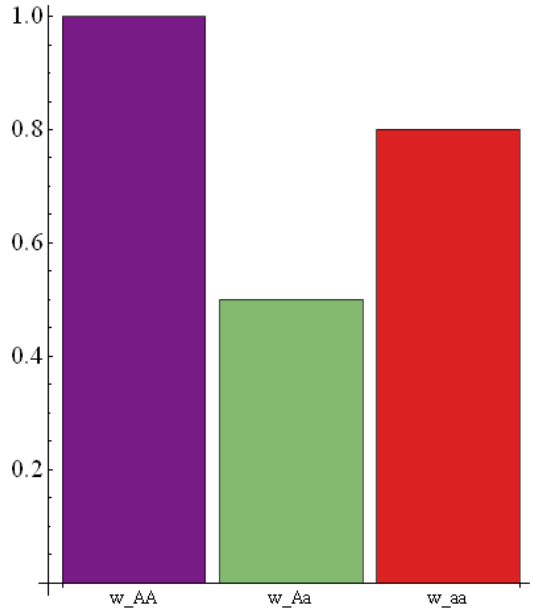
t = (2 . 9,9) / 0.05 = 396 generations

t = 396 . 25 = 9.900 years

Selection against heterozygotes (underdominance)

- Leads to fixation of one or the other allele depending on their frequency in the population and fitness of homozygote genotypes.





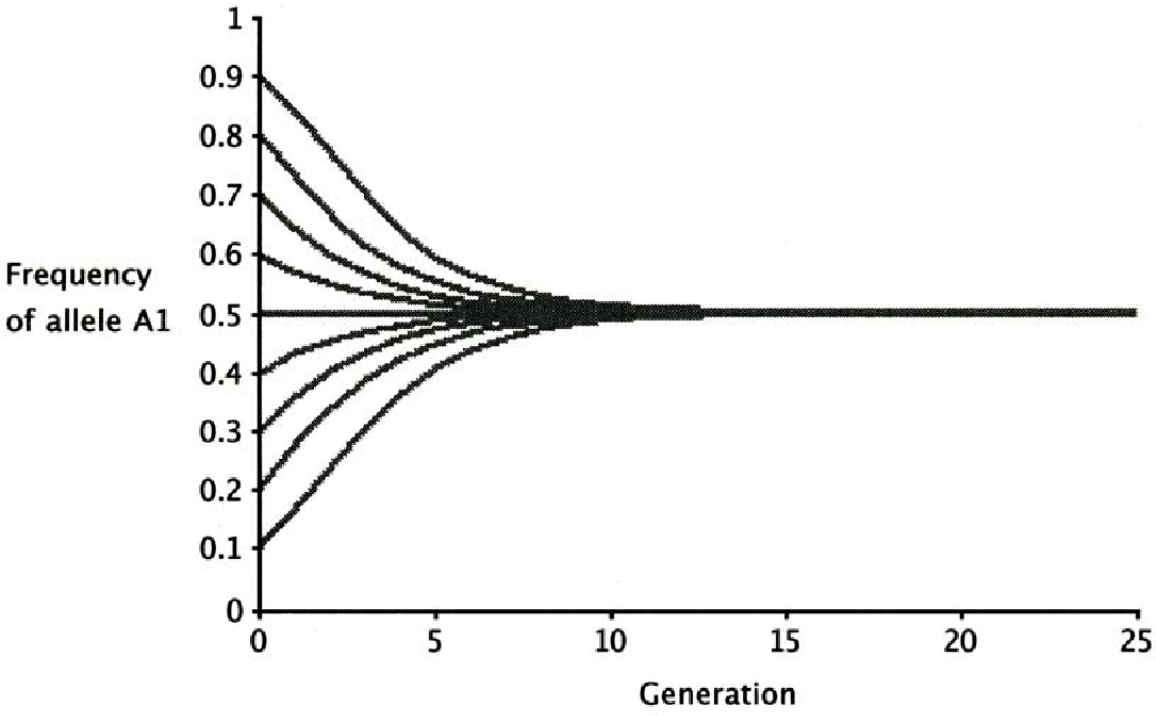
Selection against heterozygotes (underdominance)

Pseudacraea eurytus

Bateson mimicry

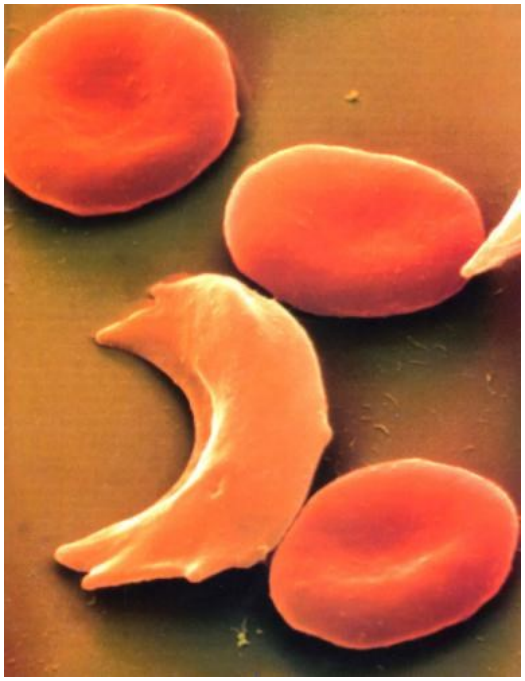
Selection in favor of heterozygotes (overdominance)

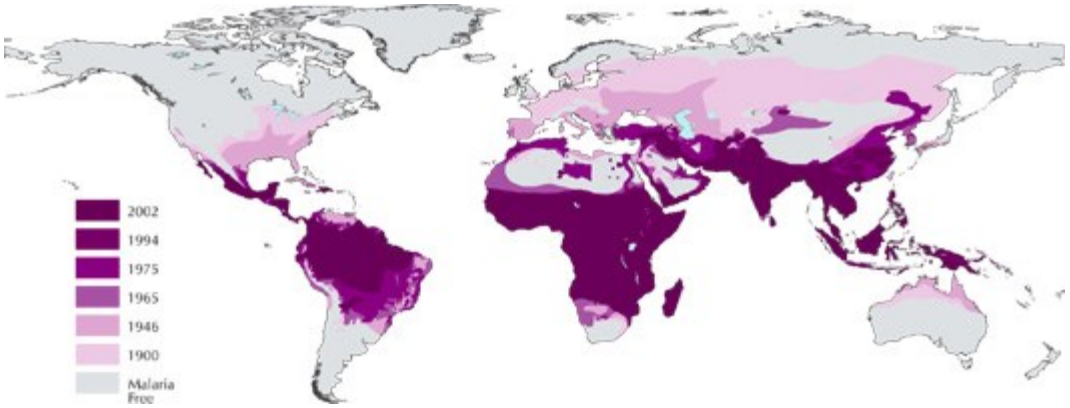
- Long-term maintenance of polymorphism in the population.
- If the fitness of homozygotes (AA and aa) is the same, frequency of the two alleles will be also the same (highest frequency of heterozygotes).



Sickle cell anemia and malaria.

- Caused by recessive mutation in the β-globin gene. Recessive homozygotes suffer from anemia, high mortality. Heterozygotes do not have sympoms and are resistant against malaria.

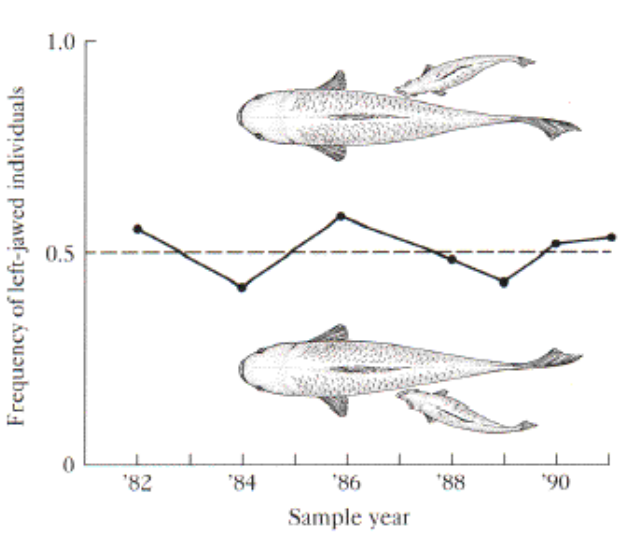




Distribuce malárie 1900 - 2002

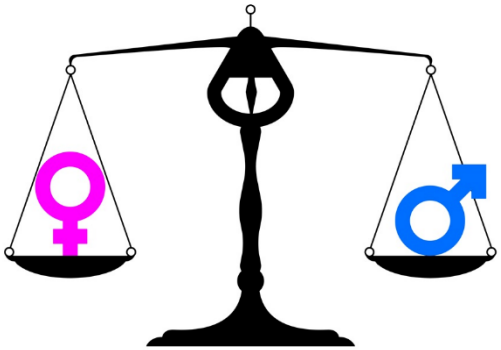
Balancing selection

- Selection in favor of heterozygotes
- **Frequency dependent selection**
- Cyclical selection



Cichlids (*Perissodus microlepsis*)

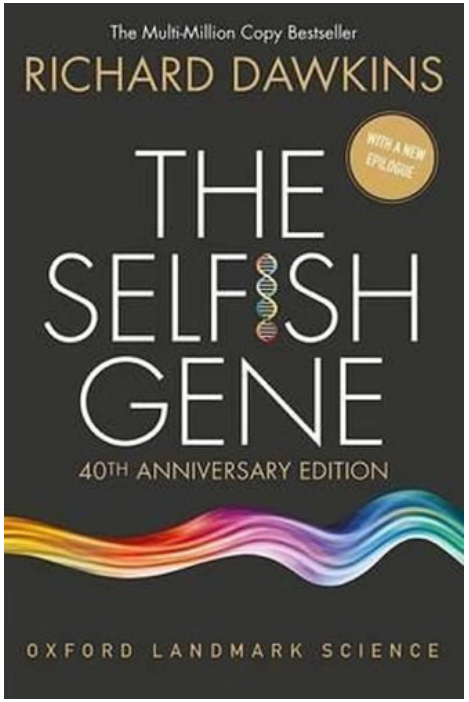
Red crossbill (*Loxia curvirostra*)

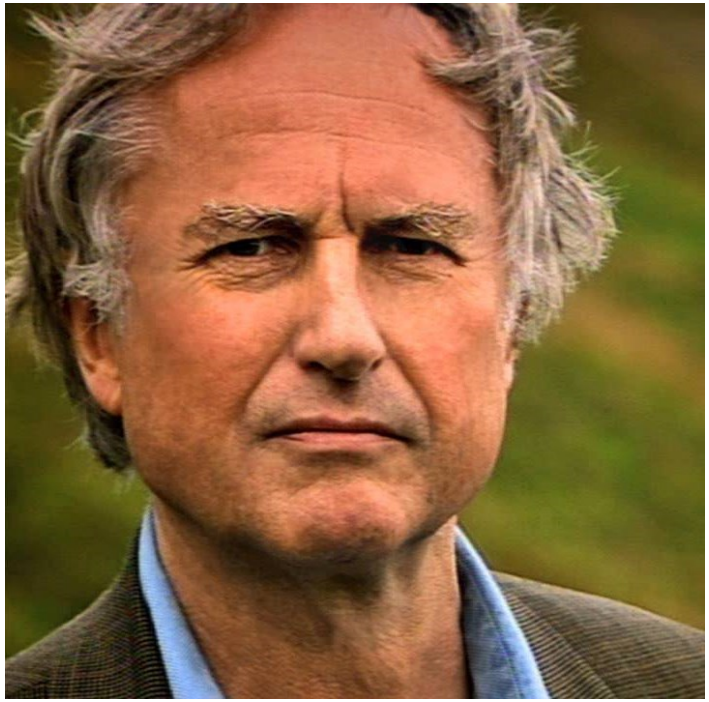


Sex ratio 1:1

The selfish gene theory (neodarwinism)

- Genocentric view on evolution
- Competition occurrs among alleles of individual genes rather than among individuals of the same species.





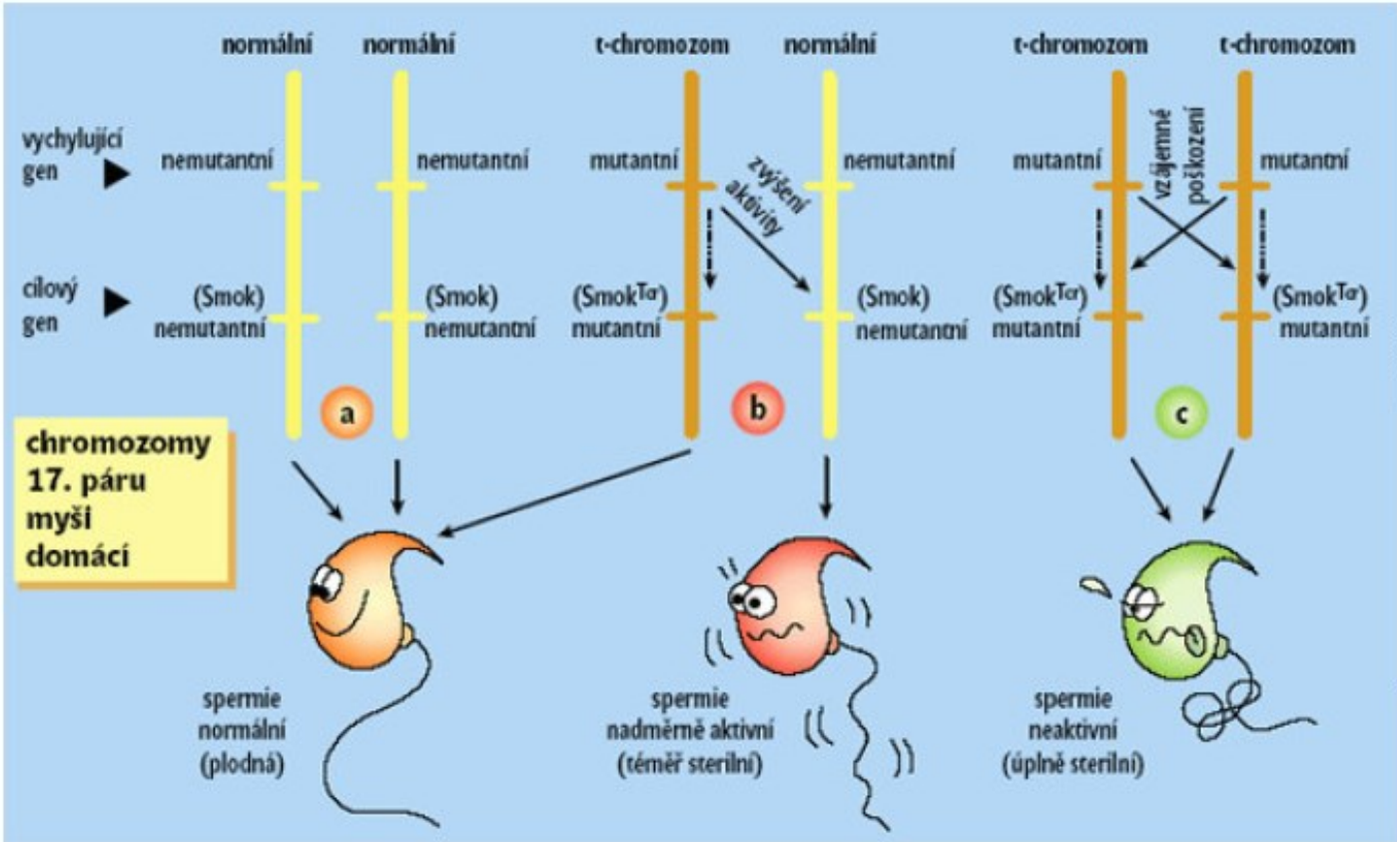
Drive

Rapid spread of gene variants independent of selection.

- **Meiotic drive**
- **Postmeiotic drive**
- **Zygotic drive**
- **Molecular drive** (e.g. through gene conversion)
- **Mutation/reparation drive**

t-haplotype

- Inversion on chromosome 17



Genetic draft
(genetic hitchhiking, linked selection)

- Change in the frequency of an allele because of linkage with beneficial or detrimental allele.