Populační genetika

Radka Reifová

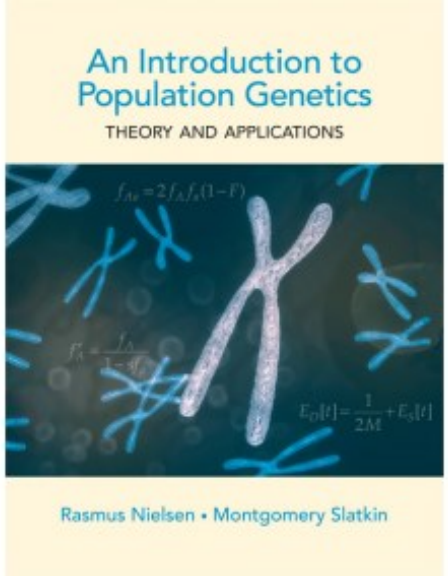
Prezentace ke stažení:

http://web.natur.cuni.cz/~radkas

v záložce Courses

Literatura

- **An Introduction to Population Genetics.**
  Rasmus Nielsen and Montgomery Slatkin. 2013. (v knihovně)
- Elements of Evolutionary Genetics (2010)
  Brian Charlesworth and Deborah Charlesworth (v knihovně).
- Population Genetics and Microevolutionary Theory (2006)
  Alan R. Templeton (v knihovně).
- Principles of Population Genetics (různá vydání)
  Daniel L. Hartl, Andrew G. Clark (v knihovně)
- Genetics of Populations (různá vydání)
  Philip W. Hedrick (v knihovně)
- Conservation and the genetics of populations.
  Fred W. Allendorf, Gordon H. Luikart, Sally N. Aitken. 2012
  http://site.ebrary.com/lib/cuni/docDetail.action?docID=10608188



Populační genetika

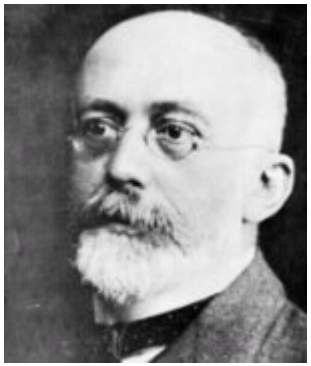
- Obor zabývající se genetickou variabilitou v populacích a procesy, které ji ovlivňují.
- Snaha vysvětlit evoluci pomocí mechanismů měnících frekvence alel v populaci.
- Metody umožňující vyčíst z genetické variability současné populace, co se dělo s populací v minulosti.

Frekvence alel a genotypů v populaci

Hardy-Weinbergův zákon

Godfrey Hardy

$$p^2 + 2pq + q^2 = 1$$



Wilhelm Weinberg

p …… frekvence alely A_1
q …… frekvence alely A_2

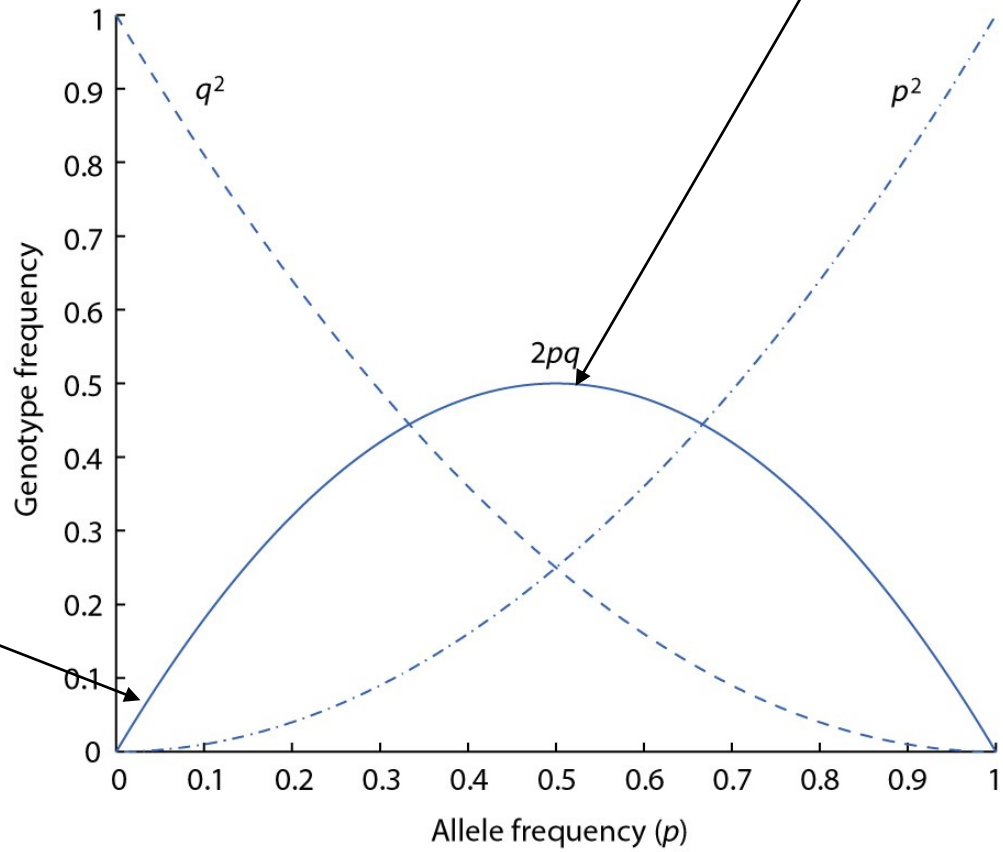
p^2 …… frekvence genotypu A_1A_1
q^2 …… frekvence genotypu A_2A_2
$2pq$ …… frekvence genotypu A_1A_2

Hardy-Weinbergův zákon

Nejvíce heterozygotů v populaci při stejné frekvenci obou alel.

Přetrvávání škodlivých recesivních mutací v populaci.

Vzácné alely skryty v heterozygotech.



Hardy-Weinbergův zákon pro více alel

2 alely

$$(p + q)^2 = 1$$

$$\boxed{p^2 + 2pq + q^2 = 1}$$

3 alely

$$(p + q + r)^2 = 1$$

$$\boxed{p^2 + q^2 + r^2 + 2pq + 2pr + 2qr = 1}$$

4 alely

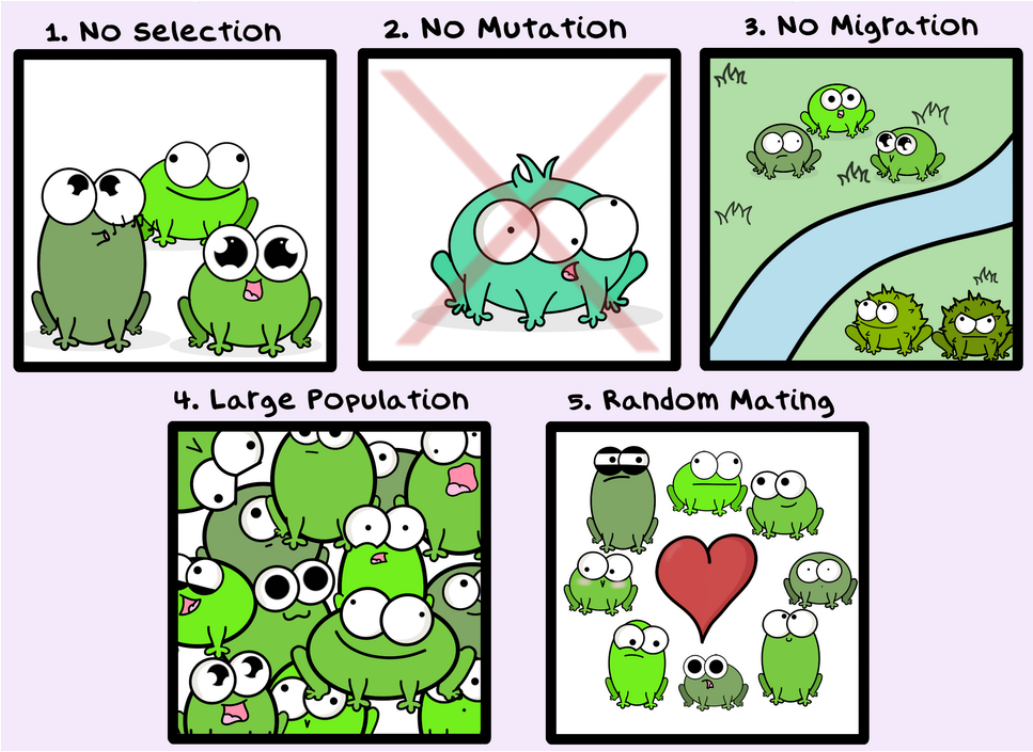
$$(p + q + r + s)^2 = 1$$

$$\boxed{p^2 + q^2 + r^2 + s^2 + 2pq + 2pr + 2ps + 2qr + 2qs + 2rs = 1}$$

atd.

Předpoklady Hardy-Weinbergovy rovnováhy

- Dostatečně velká populace, ve které se jedinci párují zcela náhodně bez ohledu na genotyp/fenotyp jedinců (**panmiktická populace**). Nehraje roli vzdálenost mezi jedinci. Neprobíhá selekce.
- Stačí jedna generace náhodného křížení, aby se ustavila HW rovnováha.
- To, zda je populace v HW rovnováze se může lišit pro různé geny v genomu.



Příklad:

Testování odchylky od Hardy-Weinbergovy rovnováhy

Pozorovali jsme níže uvedené počty genotypů v lokusu pro krevní skupinu MN.

Je populace v Hardy-Weinbergově rovnováze?

Genotypy	Pozorované počty
MM	125
MN	550
NN	310

Testování odchylky od Hardy-Weinbergovy rovnováhy

Genotypy	Pozorované počty
MM	125
MN	550
NN	325
celkem	1000

Odhad frekvence alel

$$M = \frac{2 \times 125 + 550}{2 \times 1000} = 0{,}40$$

$$N = 1 - M = 0{,}60$$

Odhad očekávaných počtů genotypů

MM $1000 \times (0{,}40)^2 = 160$

MN $1000 \times 2(0{,}40)(0{,}60) = 480$

NN $1000 \times (0{,}60)^2 = 360$

Výpočet rozdílů pozorovaných a očekávaných počtů genotypů

MM $125 - 160 = -35$

MN $550 - 480 = 70$

NN $325 - 360 = -35$

Chi-square statistika (X^2)

$$X^2 = \sum \frac{(\text{pozorované} - \text{očekávané})^2}{\text{očekávané}}$$

Počet stupnů volnosti (df)

počet porovnávaných tříd – počet odhadovaných parametrů – 1

Chi-square test

$$X^2 = \frac{(-35)^2}{160} + \frac{(70)^2}{480} + \frac{(-35)^2}{360} = 21{,}27$$

df = 3-1-1=1 **$p < 0.01$**

Populace není v HW rovnováze.

Mechanismy způsobující odchylku od HW rovnováhy

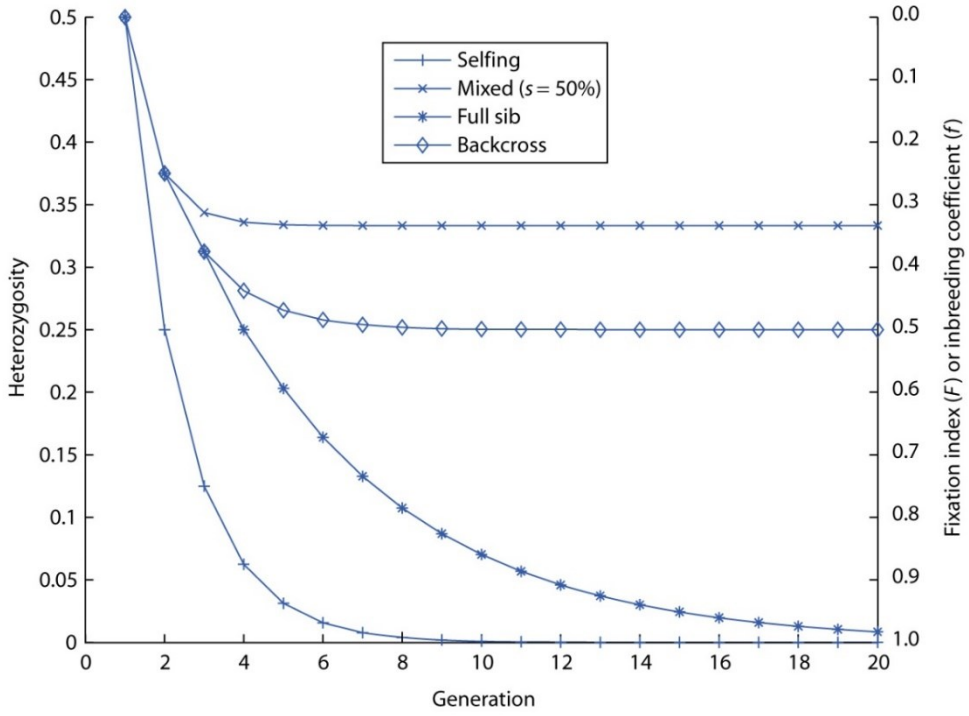
Nenáhodné (asortativní) párování

Pozitivní asortativní párování

- Jedinci s podobnými genotypy se vzájemně párují s větší pravděpodobností. Vede k přebytku homozygotů v populaci.
- Např. příbuzenské křížení (inbreeding)
- Vede k odhalení škodlivých recesivních mutací (inbrední deprese).

Inbreeding

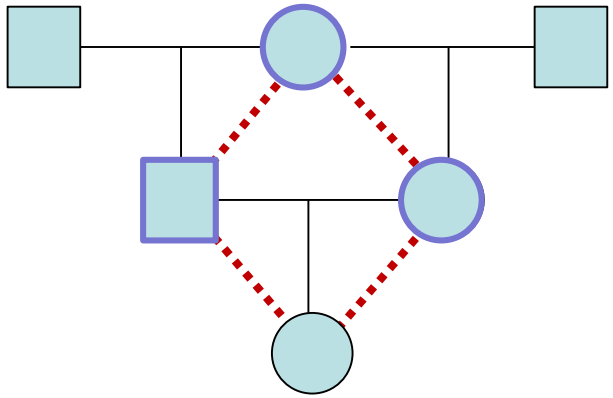
- V laboratoři se využívá při vytváření inbredních kmenů.
- V přírodě dochází k inbreedingu také v extrémně malých populacích.



Koeficient inbreedingu (F)

Pravděpodobnost, že jedinec zdědil obě alely určitého genu od jednoho předka.

$F = (1/2)^n$ n = počet jedinců v genealogii od daného jedince ke společnému předkovi



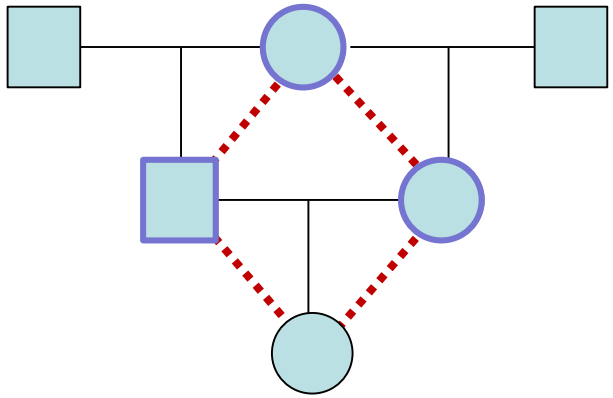
$F = (1/2)^3 = 1/8$

Koeficient inbreedingu (F)

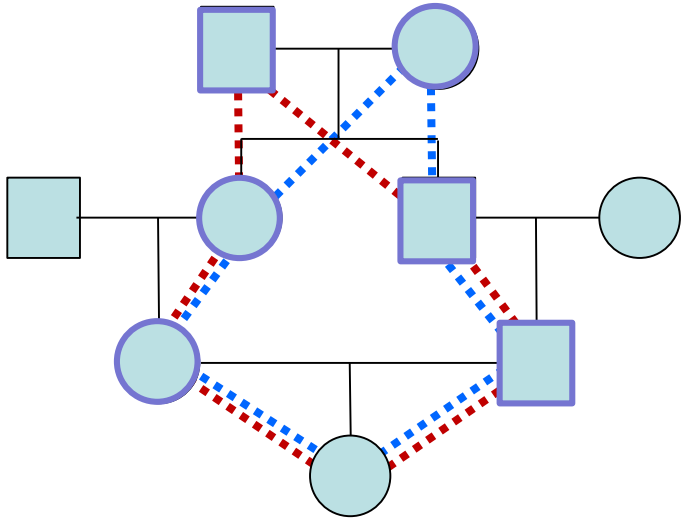
Pravděpodobnost, že jedinec zdědil obě alely téhož genu od jednoho předka.

$F = (1/2)^n$

n = počet jedinců v genealogii od daného jedince ke společnému předkovi



$F = (1/2)^3 = 1/8$



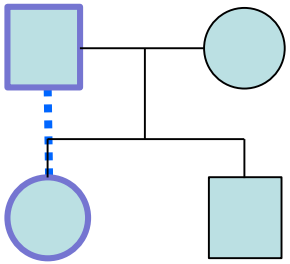
$F = (1/2)^5 + (1/2)^5$

$F = 1/32 + 1/32 = 1/16$

Koeficient příbuznosti (r)

Pravděpodobnost, že je pravděpodobnost, že dva příbuzní jedinci nesou stejnou alelu určitého genu zděděnou od společného předka.

$r = (1/2)^n$ n = počet kroků v genealogii spojující příbuzné jedince



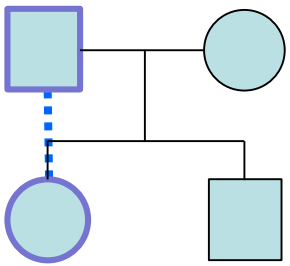
rodič - potomek

r = 1/2

Koeficient příbuznosti (r)

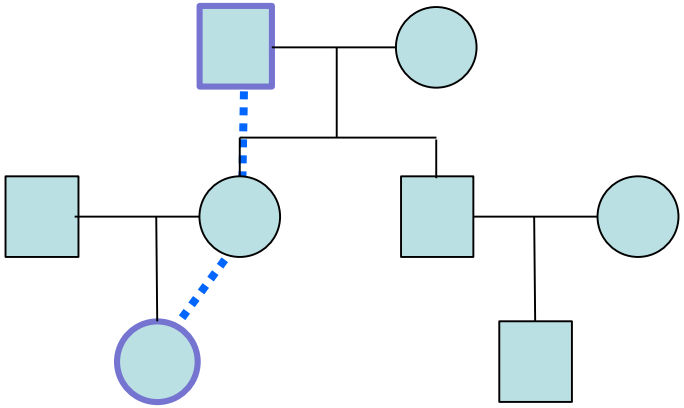
Pravděpodobnost, že je pravděpodobnost, že dva příbuzní jedinci nesou stejnou alelu určitého genu zděděnou od společného předka.

$r = (1/2)^n$ n = počet kroků v genealogii spojující příbuzné jedince



rodič - potomek

r = 1/2



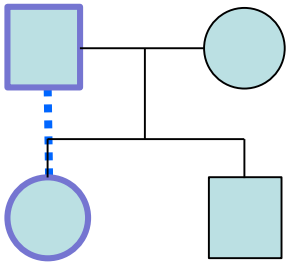
prarodič - vnouče

$r = (½)^2 = 1/4$

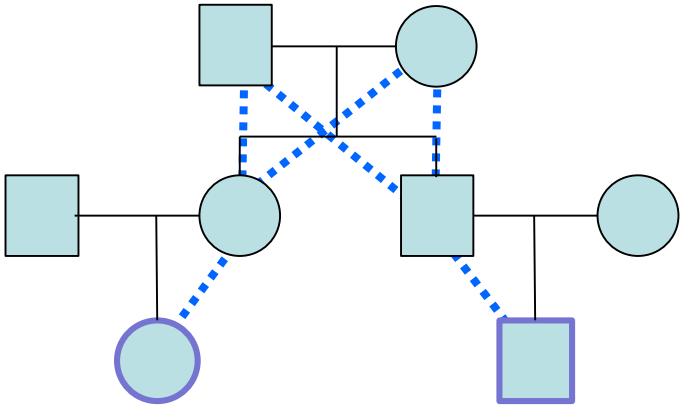
Koeficient příbuznosti (r)

Pravděpodobnost, že je pravděpodobnost, že dva příbuzní jedinci nesou stejnou alelu určitého genu zděděnou od společného předka.

$r = (1/2)^n$ n = počet kroků v genealogii spojující příbuzné jedince



otec - dcera
r = 1/2



bratranec - sestřenice

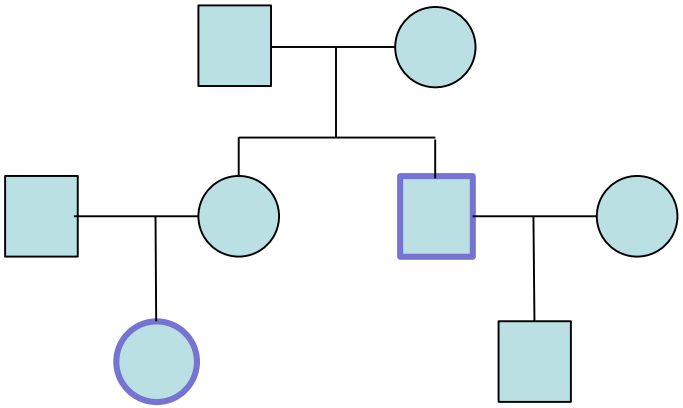
$$r = (½)^4 + (½)^4$$

$$r = (1/16) + (1/16) = 1/8$$

Domácí úkol

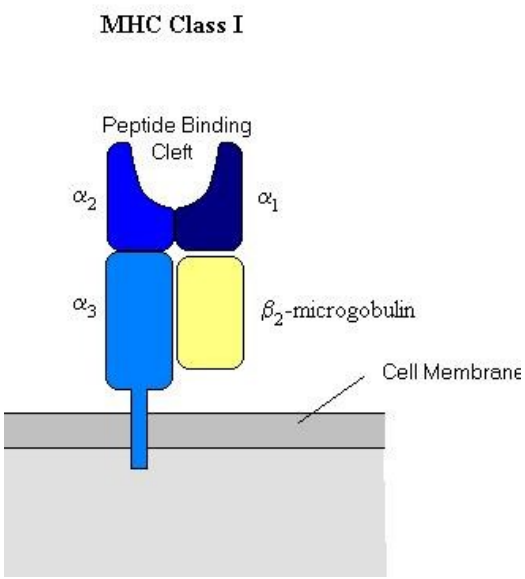
Jaký je koeficient příbuznosti mezi strýcem a jeho neteří?

Jaký by byl koeficient inbreedingu jejich případných společných dětí?



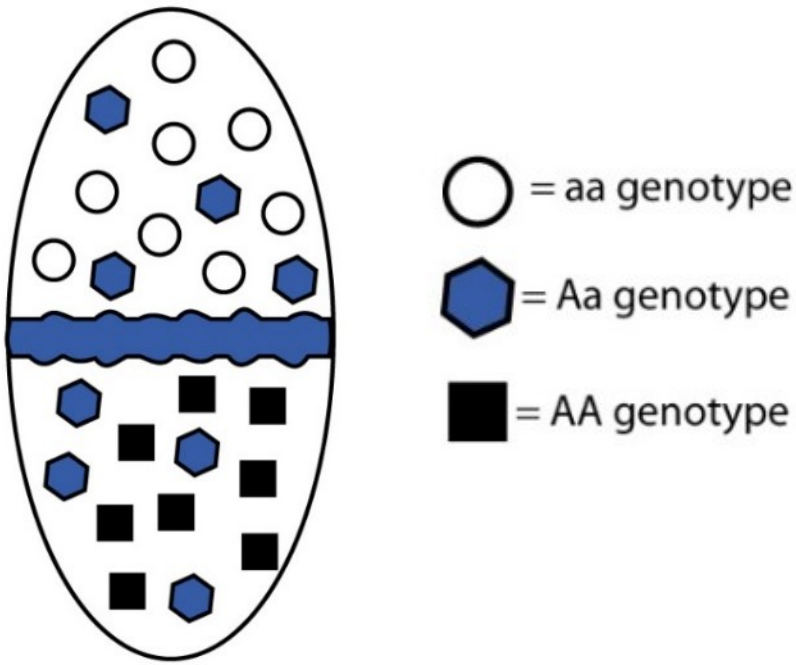
Negativní asortativní párování

- Jedinci s odlišnými genotypy se vzájemně párují s větší pravděpodobností.
- Vede k přebytku heterozygotů.
- V přírodě nastává např. pokud jsou zvýhodněni heterozygoti.
  - MHC I (*Major Histocompatibility Complex I*) geny



Geografická struktura populace

- Brání náhodnému párování mezi jedinci.
- Snížení celkové heterozygotnosti ve strukturované populaci = **Wahlundův princip.**



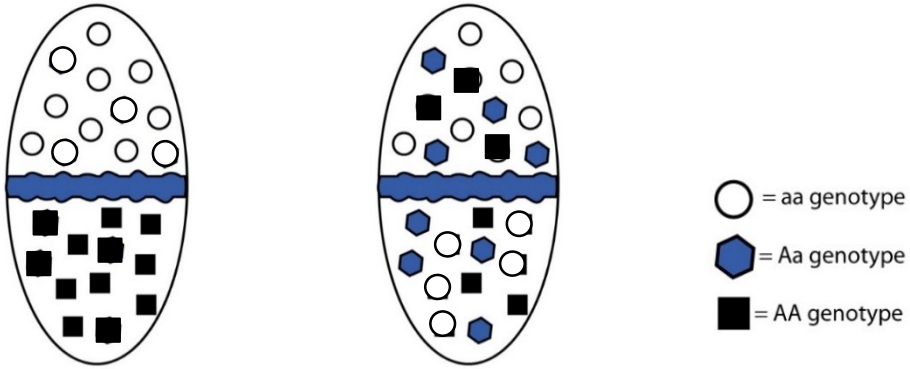
Odhad míry geografické strukturovanosti populací

F_{ST} statistika

Proporce o kterou je snížená heterozygotnost subpopulací ve srovnání s očekávanou heterozygotností za předpokladu náhodného křížení mezi všemi jedinci.

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

H_T očekávaná heterozygotnost pro celkovou populaci
H_S očekávaná heterozygotnost pro subpopulace



$F_{ST} = 1$ $F_{ST} = 0$

Příklad

Jaké je F_{ST} mezi dvěma populacemi?

	Počty genotypů			
	AA	**Aa**	**aa**	
subpopulace 1	25	50	25	100
subpopulace 2	49	42	9	100
celková populace	74	92	34	200

Frekvence alel	
p	**q**
0,5	0,5
0,7	0,3
0,6	0,4

$$H_T = 2pq = 2 \,.\, 0{,}6 \,.\, 0{,}4 = 0{,}48$$

$$H_S = p_1q_1 + p_2q_2 = 0{,}25 + 0{,}21 = 0{,}46$$

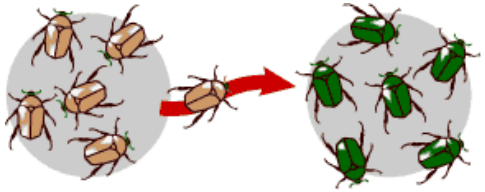
$$F_{ST} = 0{,}0416$$

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

Mechanismy měnící frekvence alel v populaci

mutace

drift

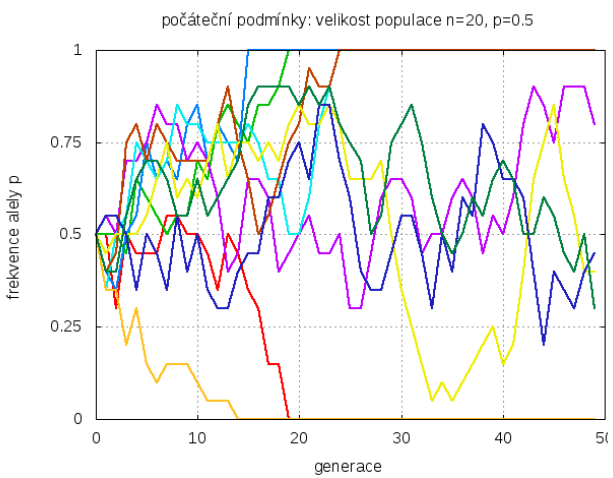


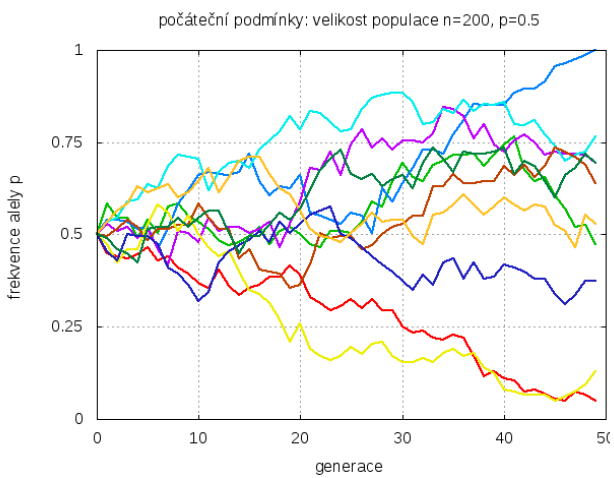
migrace

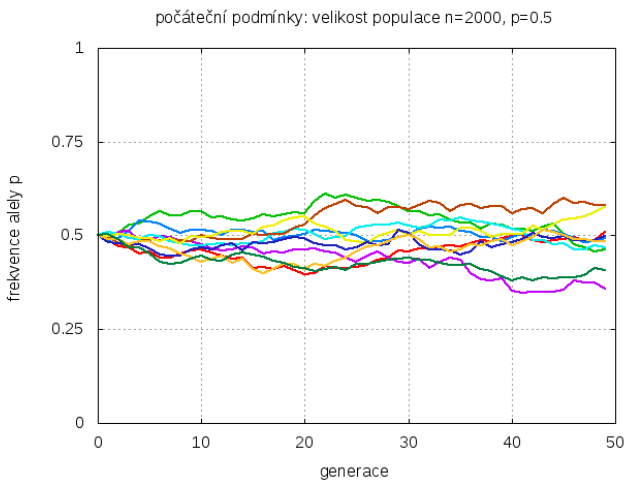
selekce

Genetický drift

- Náhodné změny ve frekvencích alel v omezeně velkých populacích.
- Působí silněji v malých populacích.
- Ochuzuje genetickou variabilitu populace.
- Směr genetického driftu je náhodný.
- Pravděpodobnost fixace alely působením genetického driftu odpovídá její frekvenci v populaci.

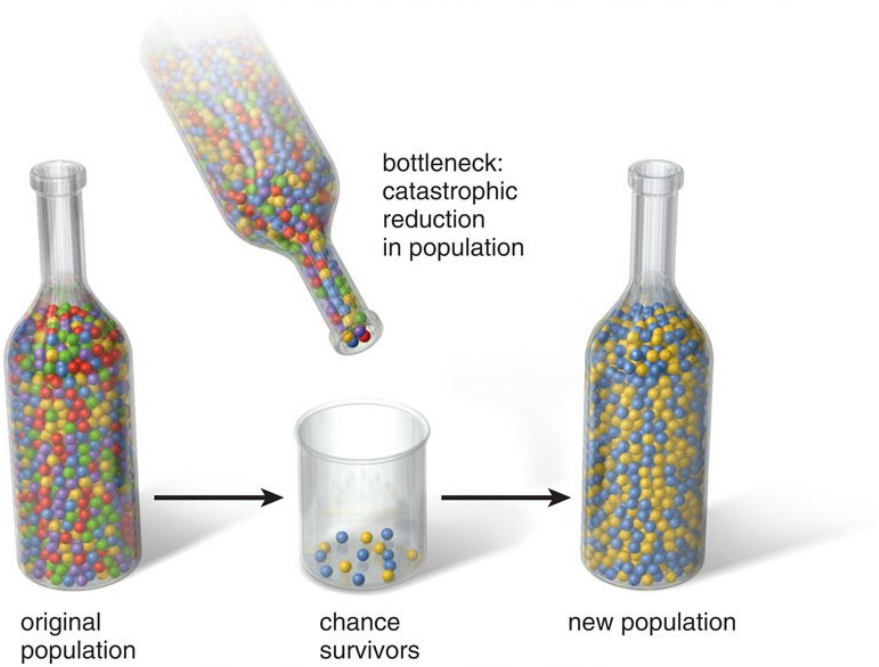




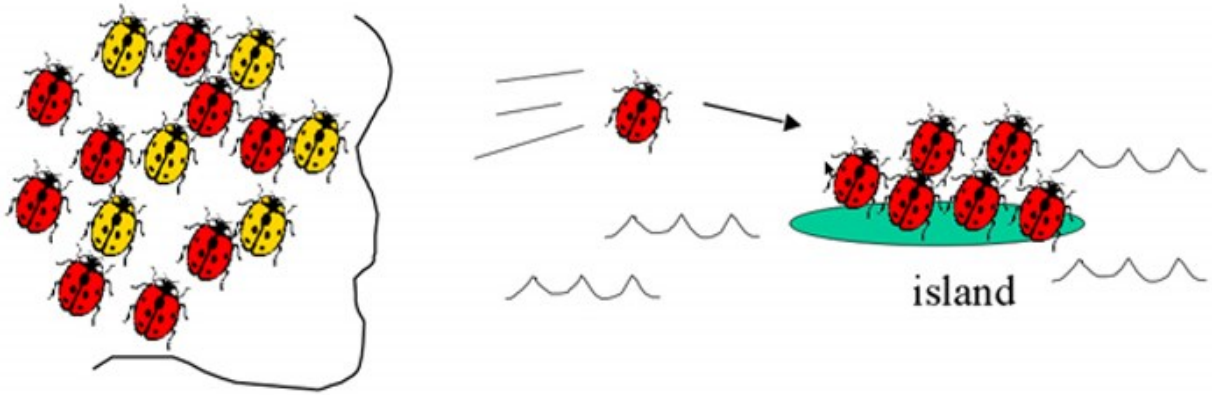


Genetický drift působí silně i při krátkodobém zmenšení velikosti populace

Efekt hrdla lahve



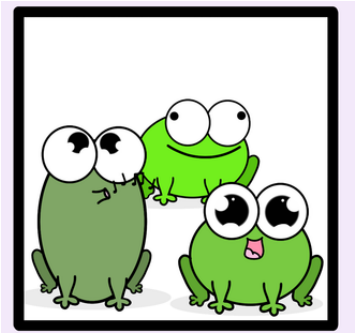
Efekt zakladatele



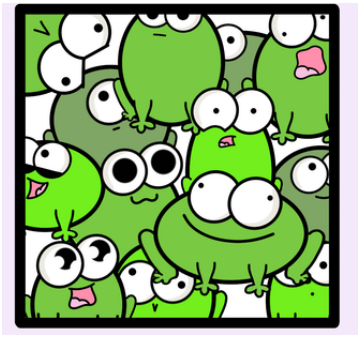
Efektivní velikost populace (N_e)

- Vyjadřuje velikost ideální panmiktické populace, ve které genetické procesy (např. genetický drift) probíhají stejnou rychlostí jako v dané reálné populaci.
- Daná pouze počtem rozmnožujících se jedinců v populaci.
- Odpovídá skutečnému počtu jedinců v populaci za předpokladu:
  - *konstantní velikosti populace*
  - *náhodného křížení mezi jedinci*
  - *rovnoměrného počtu potomků mezi jedinci (není selekce).*

Panmiktická populace



Reálná populace

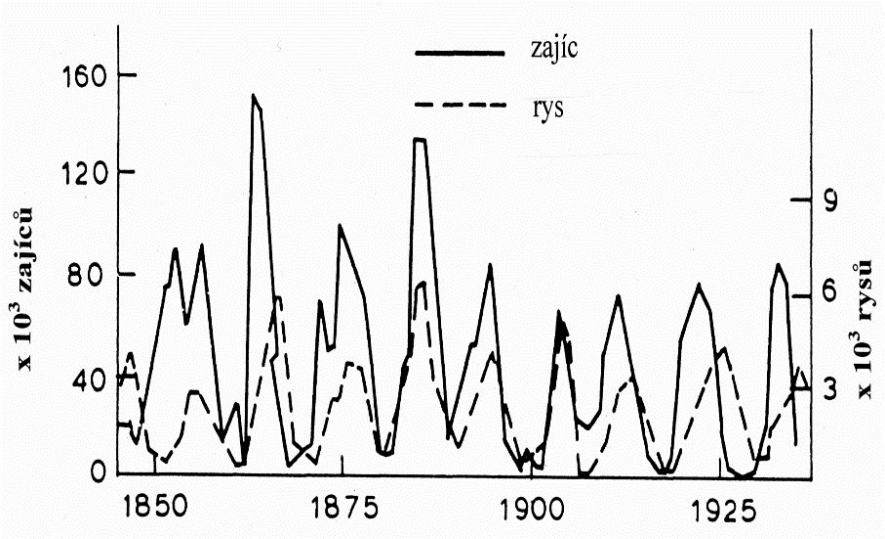


Síla genetického driftu stejná.

Faktory ovlivňující efektivní velikost populace

Kolísavá velikost populace, efekt hrdla lahve, efekt zakladatele

- N_e odráží nejen současnou velikost populace, ale také velikost populace v minulosti.
- U populací s proměnlivou velikostí se Ne blíží spíše nižším hodnotám celkového počtu jedinců v populaci.



Faktory ovlivňující efektivní velikost populace

Nenáhodné křížení a rozdílný počet rozmnožujících se samic a samců

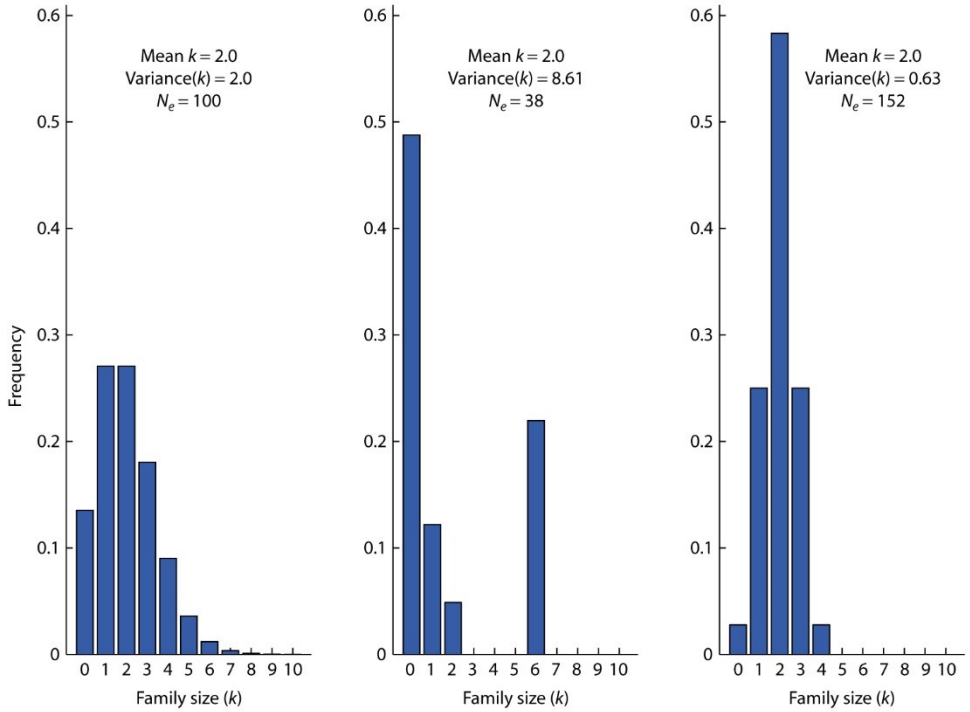
- Vede ke snížení efektivní velikosti populace.

$$N_e = \frac{4N_mN_f}{N_m + N_f}$$

Faktory ovlivňující efektivní velikost populace

Rozptyl v počtu potomků mezi jedinci (selekce)

- N_e se rovná skutečnému počtu jedinců pokud počty potomků jednotlivých jedinců mají Poisson rozložení. Tzn. rozptyl v počtu potomků je stejný jako průměrný počet potomků.
- Pokud je rozptyl větší, N_e je nižší než skutečný počet jedinců v populaci.



Odhady efektivní velikosti populace u různých druhů

~ 10 000

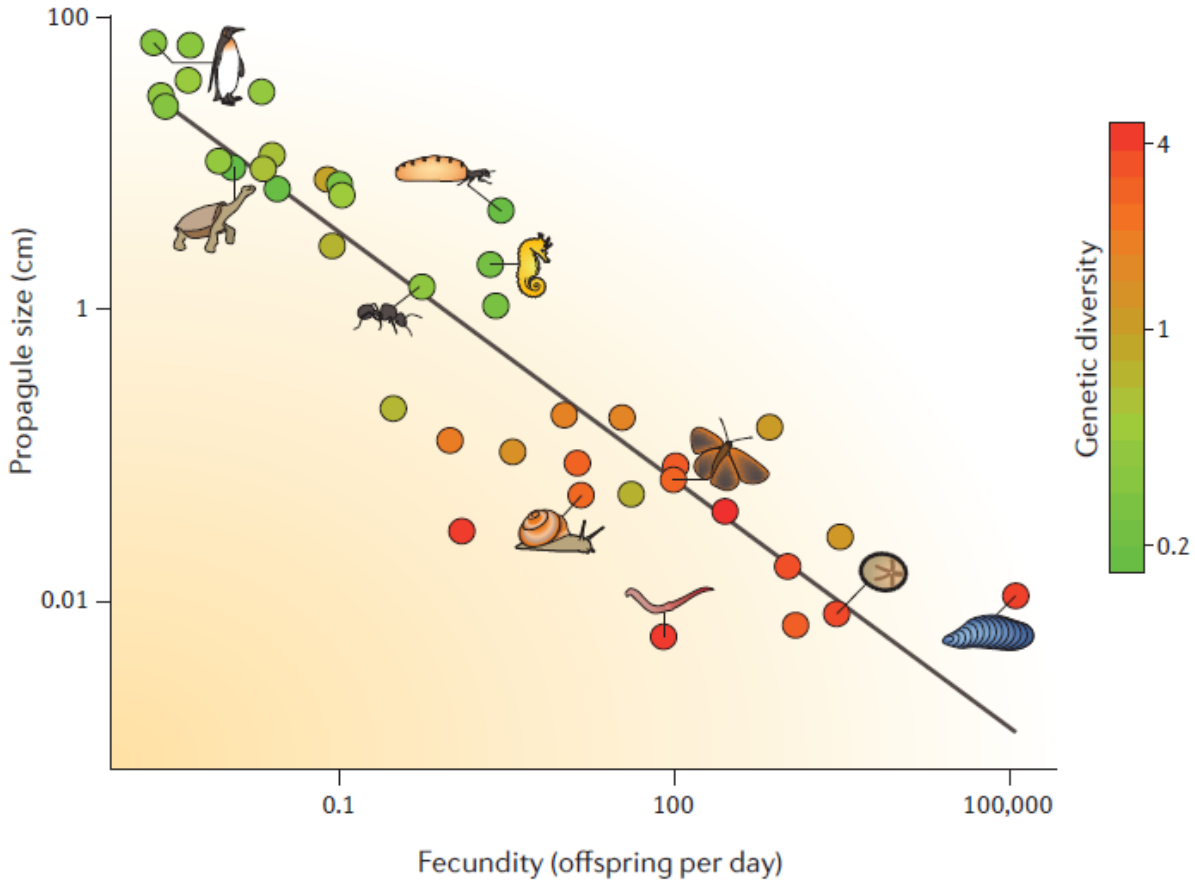
~ 30 000

~ 2 000 000

~ 1 000 000

~ 100 000

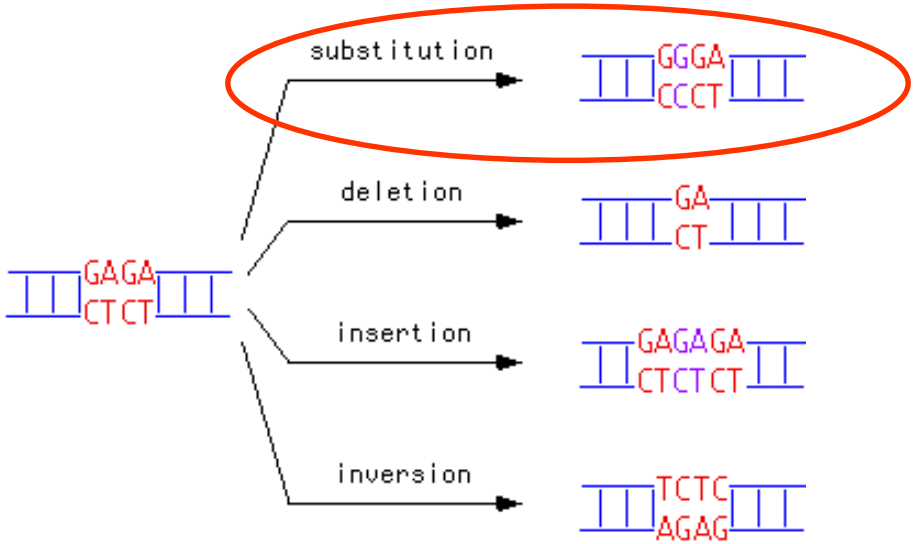
Efektivní velikost populace závisí na životních strategiích organismů



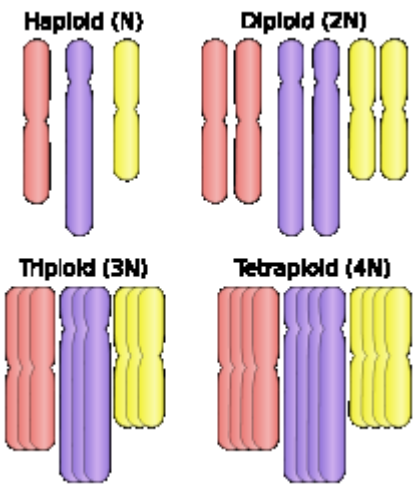
Ellegren and Galtier, 2016

Mutace jako zdroj genetické variability

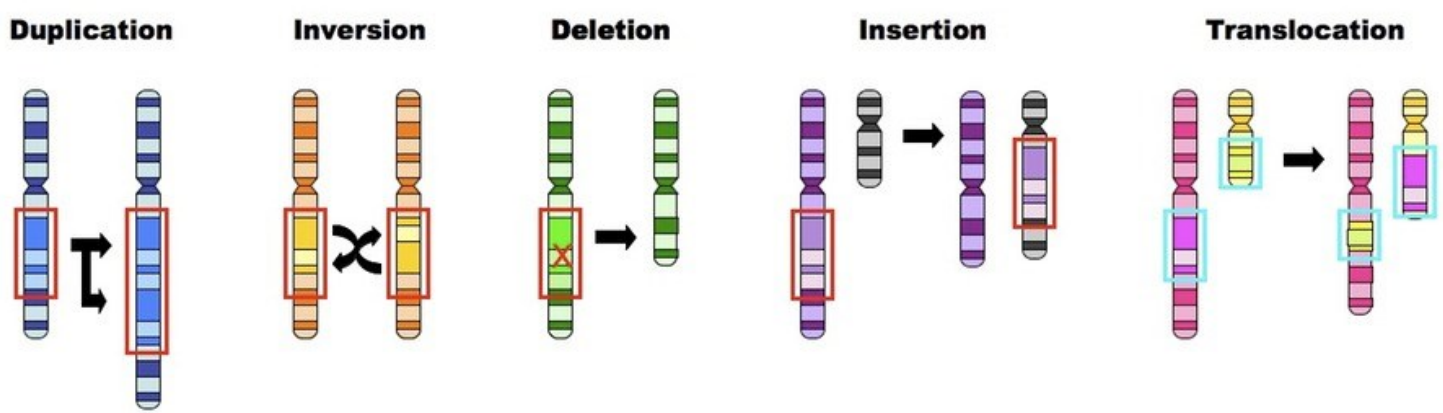
Bodové/genové mutace



Genomové mutace



Chromosomové mutace



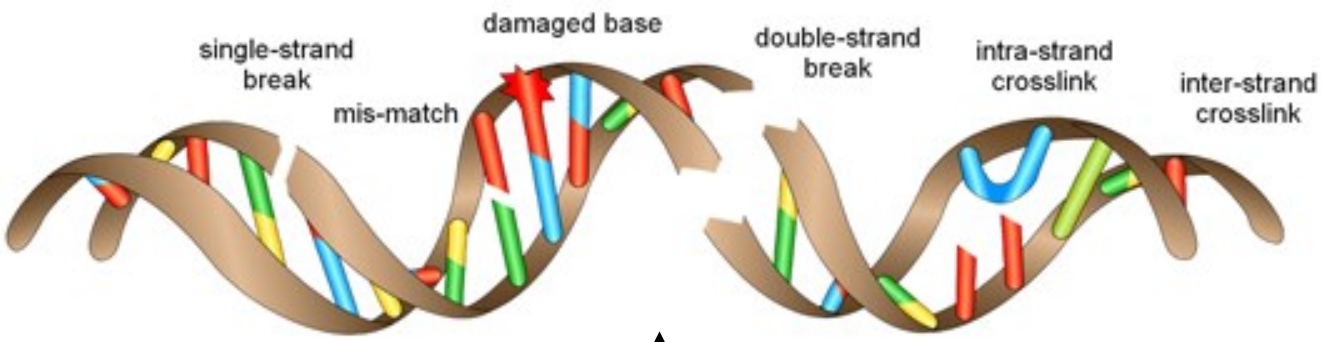
Mutace

Buněčný metabolismus

Záření

Chemické mutageny

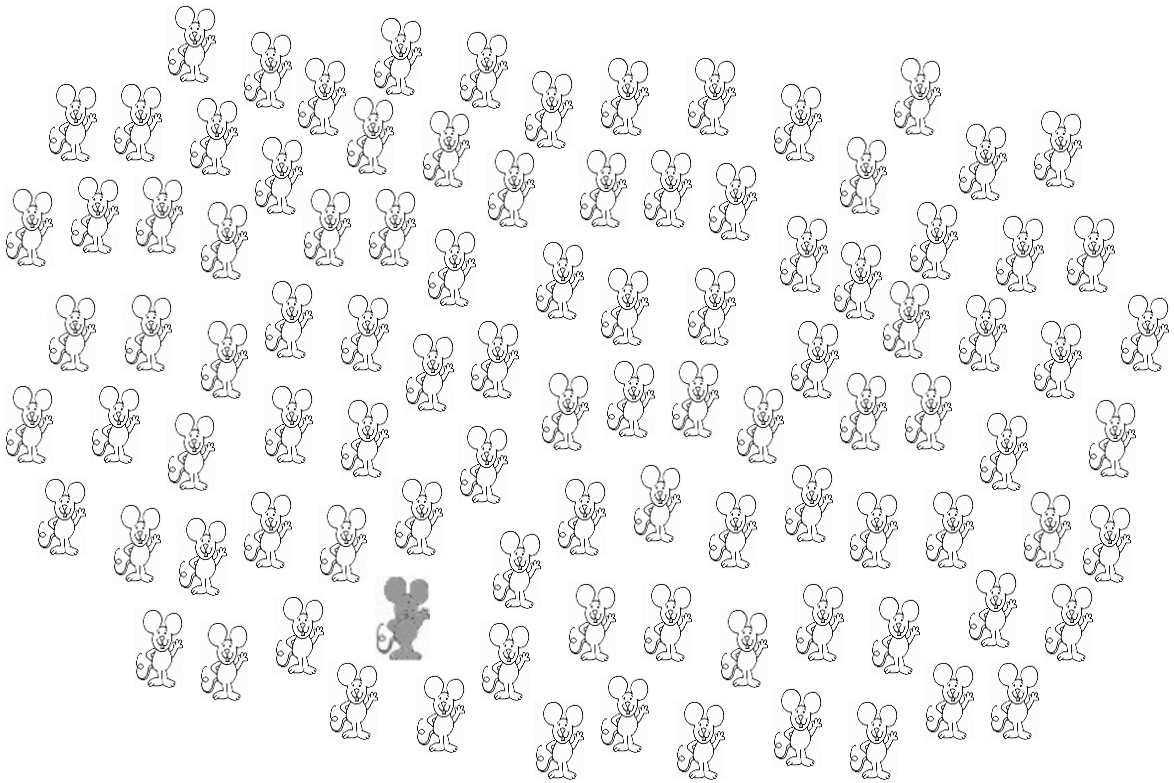
Chyby při replikaci



Reparační mechanismy DNA

Mutační rychlost (μ)

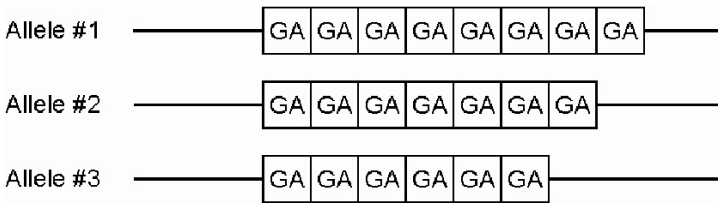
- frekvence nových mutací na generaci
- μ = 0.01 (1 mutace na 100 jedinců na 1 generaci)



- Mutační rychlost (pro nukleotidové substituce) u lidí je cca 1.10^{-8}.
  (necelých 100 nových nukleotidových substitucí v celém genomu na jedince)

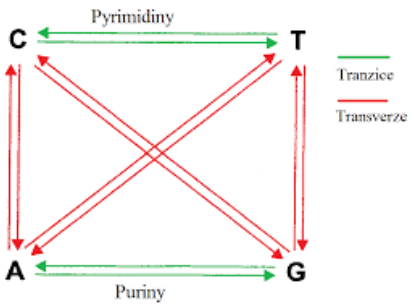
Mutační rychlost

mikrosatelit



Je odlišná pro různé typy mutací

- rozsáhlejší mutace obecně méně časté
- mutační „hotspots“: např. mikrosatelity.
- tranzice vznikají častěji než transverze



Je odlišná pro různé geny v rámci genomu

- Ovlivněna např. mírou transkripce dané sekvence, blízkostí k replikačnímu počátku, pozicí nukleozomu.
- Mutační rychlost u živočichů cca 10 x větší na mt DNA než v jaderné DNA

Liší se u různých organismů

- velikost genomu
- životní strategie (velikost těla, intenzita metabolismu, dlouhověkost, efektivní velikost populace…)

Mutační rychlost v germinální linii a somatických tkáních

Table 1. Mutation rates per nucleotide site ($\times 10^{-9}$) in different tissues[a]

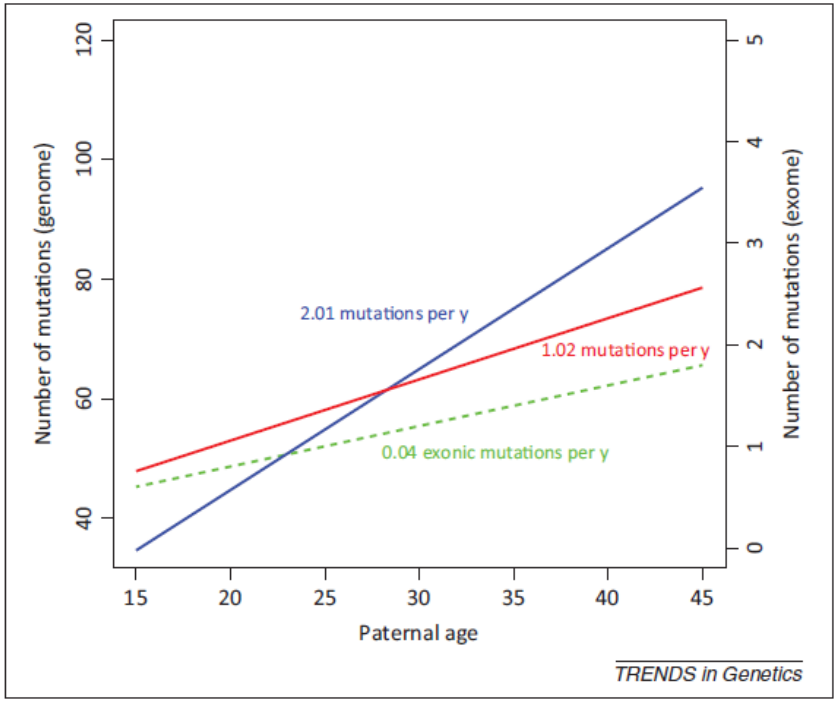
Species	Tissue	Cell divisions per generation[a]	Mutation rates[b]	
			Per generation	Per cell division
Homo sapiens	Germline	216	12.85	0.06
	Retina	55	54.45	0.99
	Intestinal epithelium	600	162.00	0.27
	Fibroblast (culture)			1.34
	Lymphocytes (culture)			1.47
Mus musculus	Male germline	39	38.00	0.97
	Brain		76.94	
	Colon		83.35	
	Epidermis		90.38	
	Intestine		117.69	
	Liver		237.88	
	Lung		166.83	
	Spleen		130.00	
Rattus norvegicus	Colon		178.38	
	Kidney		167.45	
	Liver		179.92	
	Lung		223.22	
	Mammary gland		57.70	
	Prostate		448.90	
	Spleen		101.62	
Drosophila melanogaster	Germline	36	4.65	0.13
	Whole body		380.92	
Caenorhabditis elegans	Germline	9	5.60	0.62
Arabidopsis thaliana	Germline	40	6.50	0.16
Saccharomyces cerevisiae		1	0.33	0.33
Escherichia coli		1	0.26	0.26

[a]References to data on numbers of germline cell divisions: human [Crow 2000]; *D. melanogaster* and mouse [57]; *C. elegans* [58]; and *A. thaliana* [59]. Numbers of cell divisions are unknown for the mouse and rat rates.

[b]Mammalian tissue-specific rates are given only for tissues in which at least two independent estimates have been acquired. All data on human mutation rates are taken from Lynch [36]. Data for somatic mutation rates in mouse and rat are derived from references contained within the supplementary material online. References to data on germline mutation rates are: *D. melanogaster* [5], *C. elegans* [4], *A. thaliana* [Ossowski *et al.*, 2009], *S. cerevisiae* [3], and *E. coli* [24].

Většinu mutací dědíme od otců

- Většina nukleotidových substitucí vzniká v samčí germinální linii (více replikací DNA) (**male-driven evolution**).
- Počet nově vzniklých nukleotidových substitucí vzrůstá s věkem otce (paternal age effect).

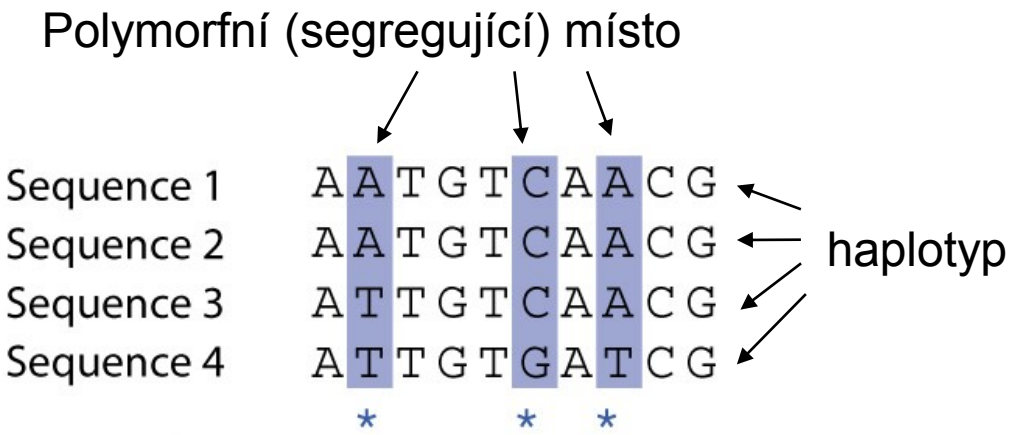


Campbell and Eicher, 2013

Genetická diverzita - genetický polymorfismus (θ)

Pro neutrální mutace platí:

$$\theta = 4N_e\mu$$



- Míra genetického polymorfismu v populaci závisí přímoúměrně na mutační rychlosti a velikosti populace.

Empirické odhady genetické diverzity v populaci

Nukleotidová diverzita (průměrná heterozygozita) (π)

Sequence 1	AATGTCAACG
Sequence 2	AATGTCAACG
Sequence 3	ATTGTCAACG
Sequence 4	ATTGTGATCG
	* * *
Site number	1 2 3 4 5 6 7 8 9 10

Nucleotide diversity (π):

1 AATGTCAACG
2 AATGTCAACG $d_{12} = 0$

1 AATGTCAACG
3 ATTGTCAACG $d_{13} = 1$

2 AATGTCAACG
3 ATTGTCAACG $d_{23} = 1$

1 AATGTCAACG
4 ATTGTGATCG $d_{14} = 3$

2 AATGTCAACG
4 ATTGTGATCG $d_{24} = 3$

3 ATTGTCAACG
4 ATTGTGATCG $d_{34} = 2$

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

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

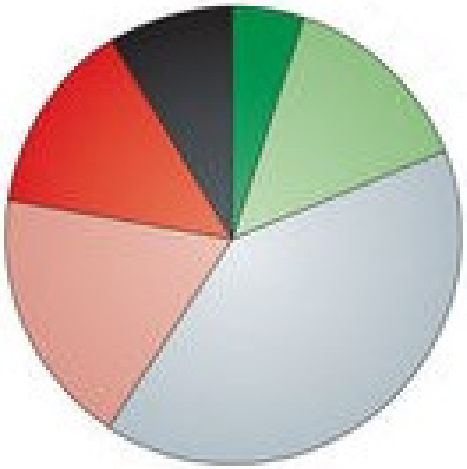
Teorie neutrální evoluce

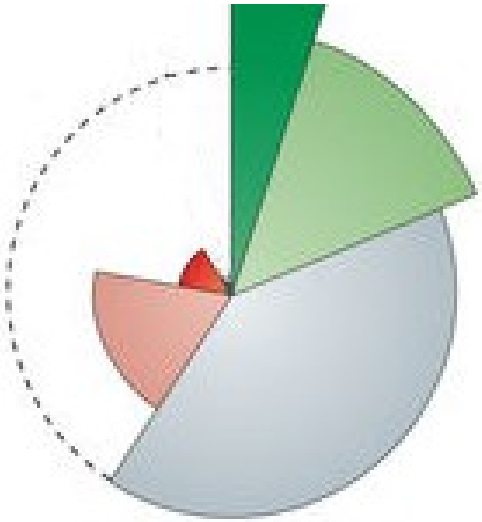
Konec 60. a začátek 70. let 20. stol.

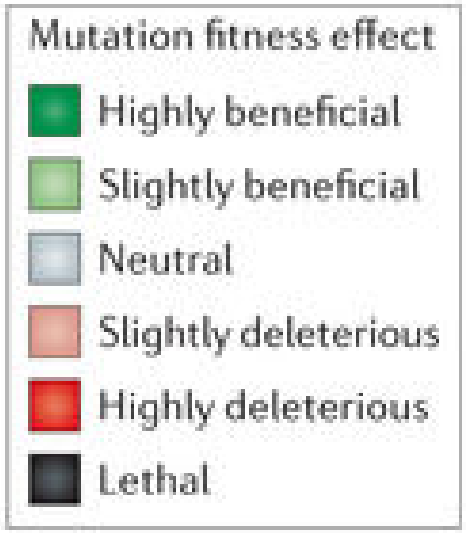
- Ukazuje jak bude vypadat genetická variabilita v populaci a jaká bude rychlost evoluce v případě, že mutace jsou neutrální a působí na ně pouze genetický drift.

Mottoo Kimura

Mutační vs. substituční rychlost







Mutační rychlost

Rychlost s jakou nové mutace vznikají.

Substituční rychlost

Rychlost s jakou se substituce fixují v populaci.

substituční rychlost

= počet nových mutací x pravděpodobnost jejich fixace

Pro neutrální mutace platí:

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

N_e … efektivní velikost populace
μ … mutační rychlost

pravděpodobnost fixace nové mutace

počet nových mutací v populaci

Substituční rychlost pro neutrální mutace je daná pouze mutační rychlostí. Nezávisí na efektivní velikosti populace. Za předpokladu neměnné mutační rychlosti se je substituční rychlost v čase také konstantní. **Molekulární hodiny**.

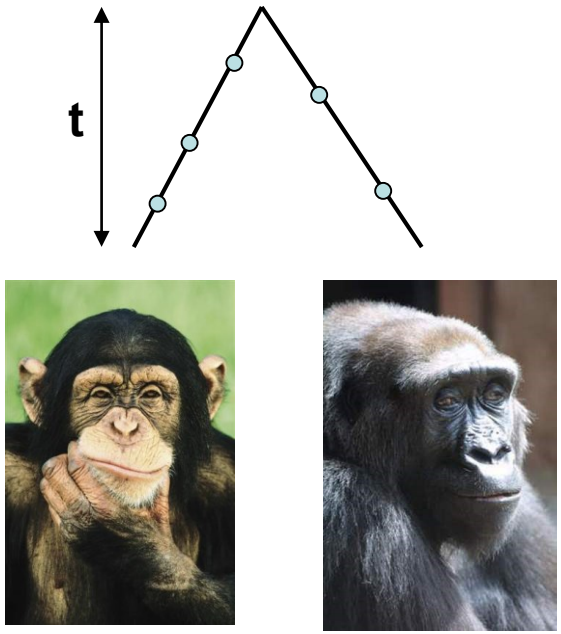
Molekulární hodiny a odhad doby divergence dvou druhů

Genetická divergence

= substituční rychlost x čas

$$D = 2\mu t$$

- Počet mutací, kterými se liší dva taxony.
- Pro neutrální sekvence vzrůstá genetická divergence lineárně s časem.



Odhad genetické divergence ze sekvenčních dat

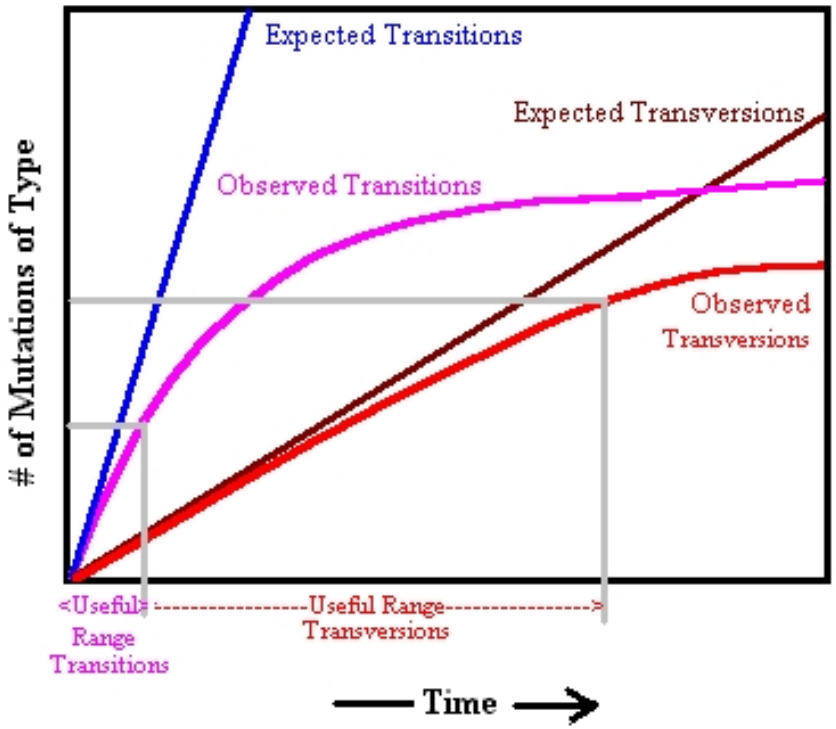
```
Druh 1 AGTGAGTCGTCAGACTGCTG
Druh 2 ACTTAGCCGTGAGACAGCTA
```

$$D = \frac{6}{20} = 0{,}3\ (30\%)$$

Proporce nukleotidových míst, kterými se liší dva taxony.

Saturace mutací

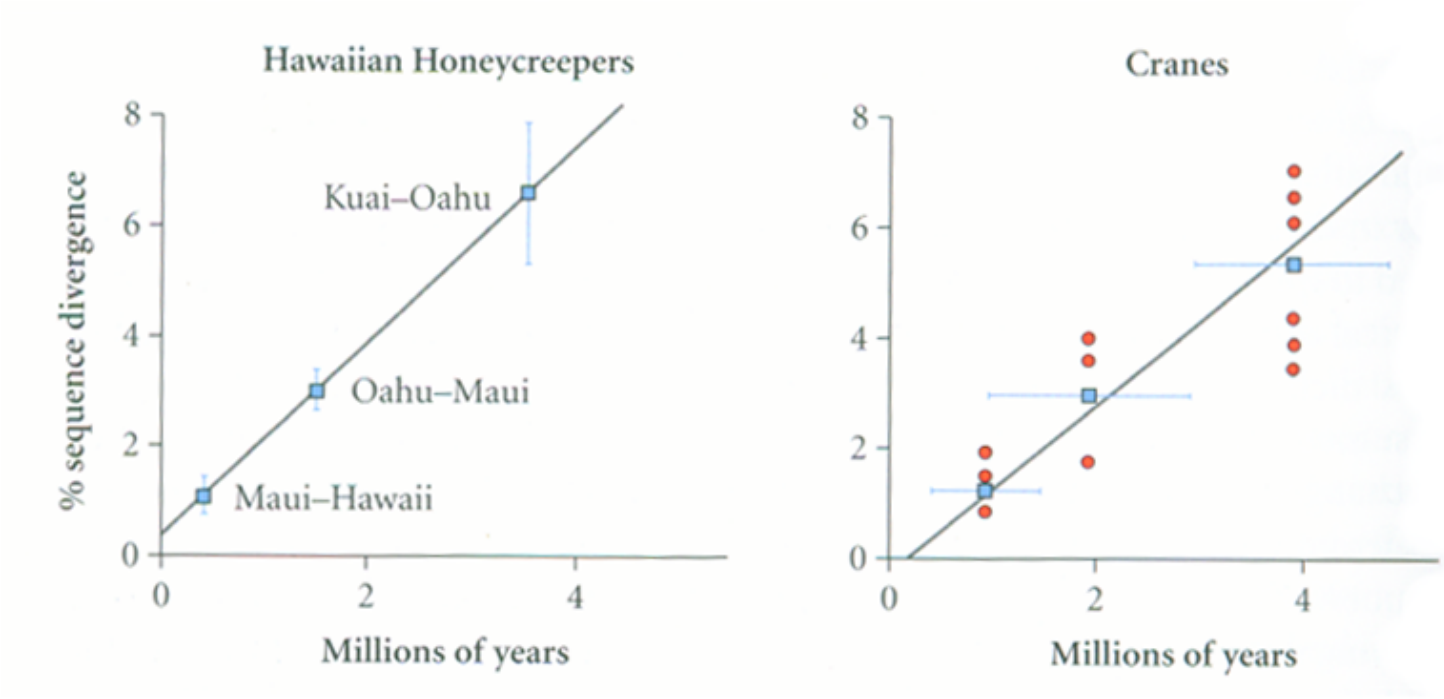
- Po určité době divergence přestane stoupat. Je to kvůli mnohočetným substitucím ve stejné pozici. Došlo k saturaci.
- Odhad divergence vzdálených sekvencí je podhodnocený. Nutno korigovat pomocí nukleotidových substitučních modelů.



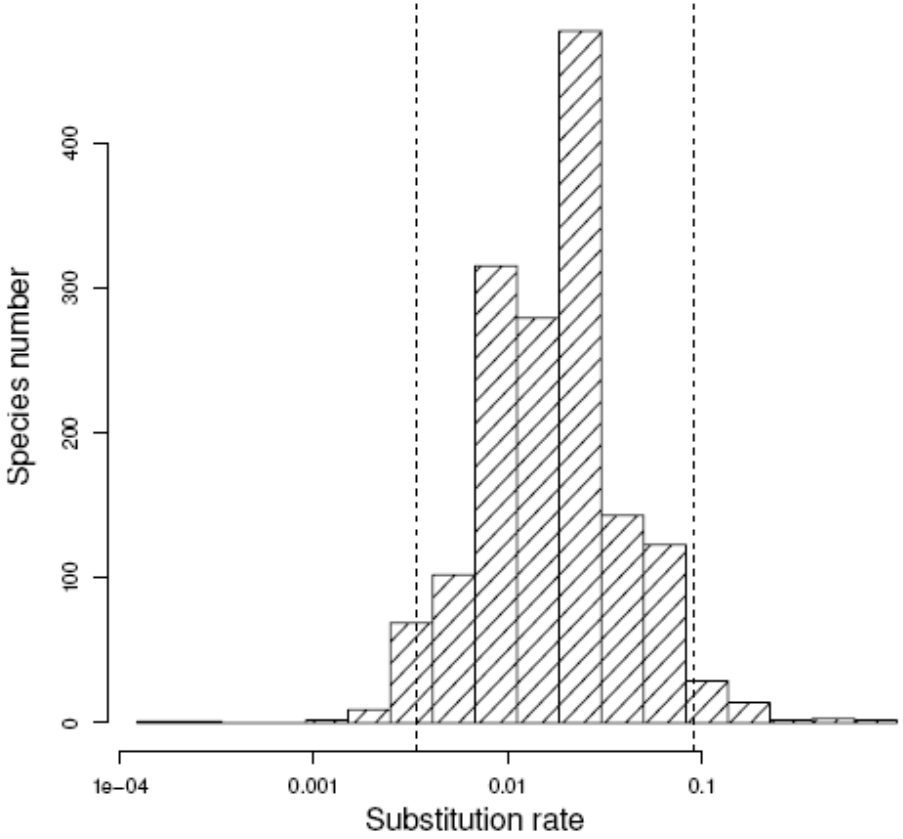
Odhad doby divergence dvou druhů z genetických dat

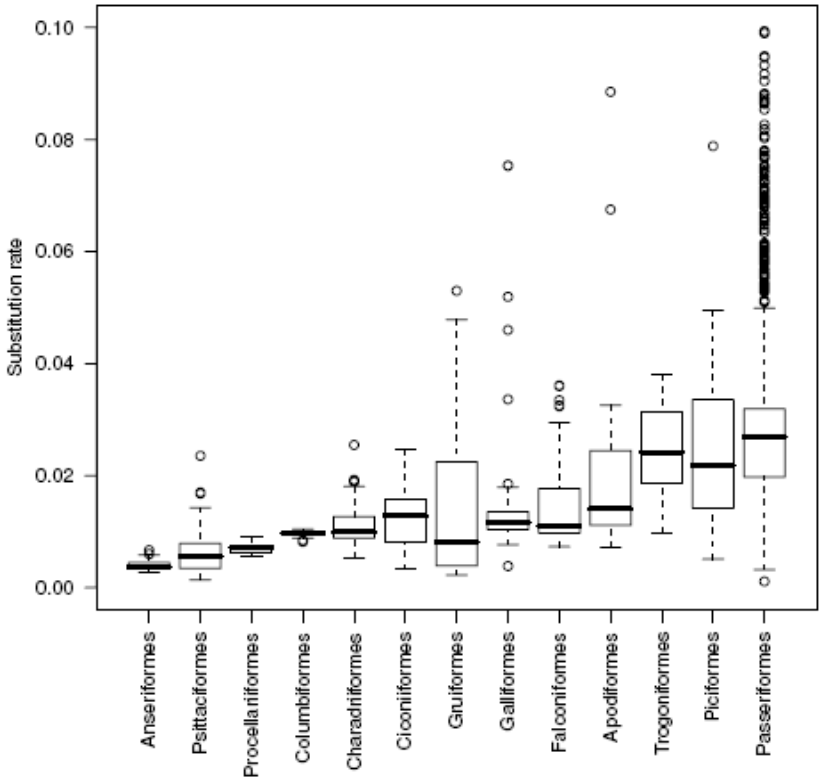
- **Kalibrace molekulárních hodin**: tj. odhad substituční rychlosti
- Provádí se na základě známé doby divergence některých taxonů na základě např. paleontologických dat.
- Substituční rychlost pro sekvenci genu cytochromu b (mt DNA) u savců a ptáků cca 0,01. Tzn. 2% divergence ~ 1 mil let.

$$D = 2\mu t$$



Substituční rychlost se může lišit u různých druhů





Substituční rychlosti pro cytochrom b u ptáků.

Pravidlo 2% divergence ~ 1 mil let, neplatí!

Molekulární hodiny

Striktní molekulární hodiny

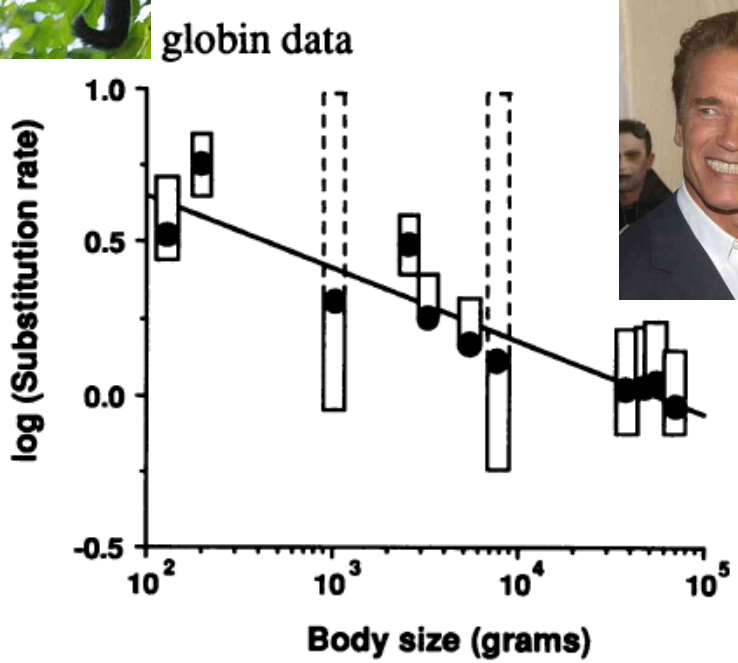
•Substituční rychlost homogenní napříč fylogenezí.

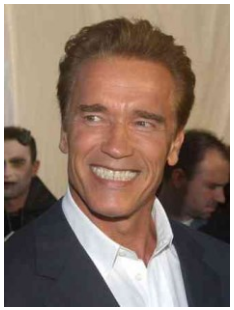
Relaxované molekulární hodiny

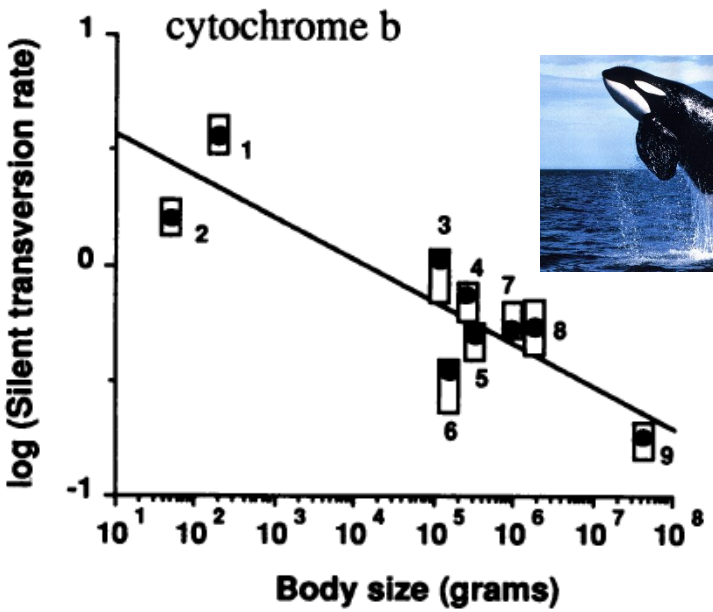
•Berou v potaz možnou variabilitu v substituční rychlosti mezi liniemi.

Proč se substituční rychlost liší u různých druhů?

Druhy s menší velikostí těla mají větší substituční rychlost.



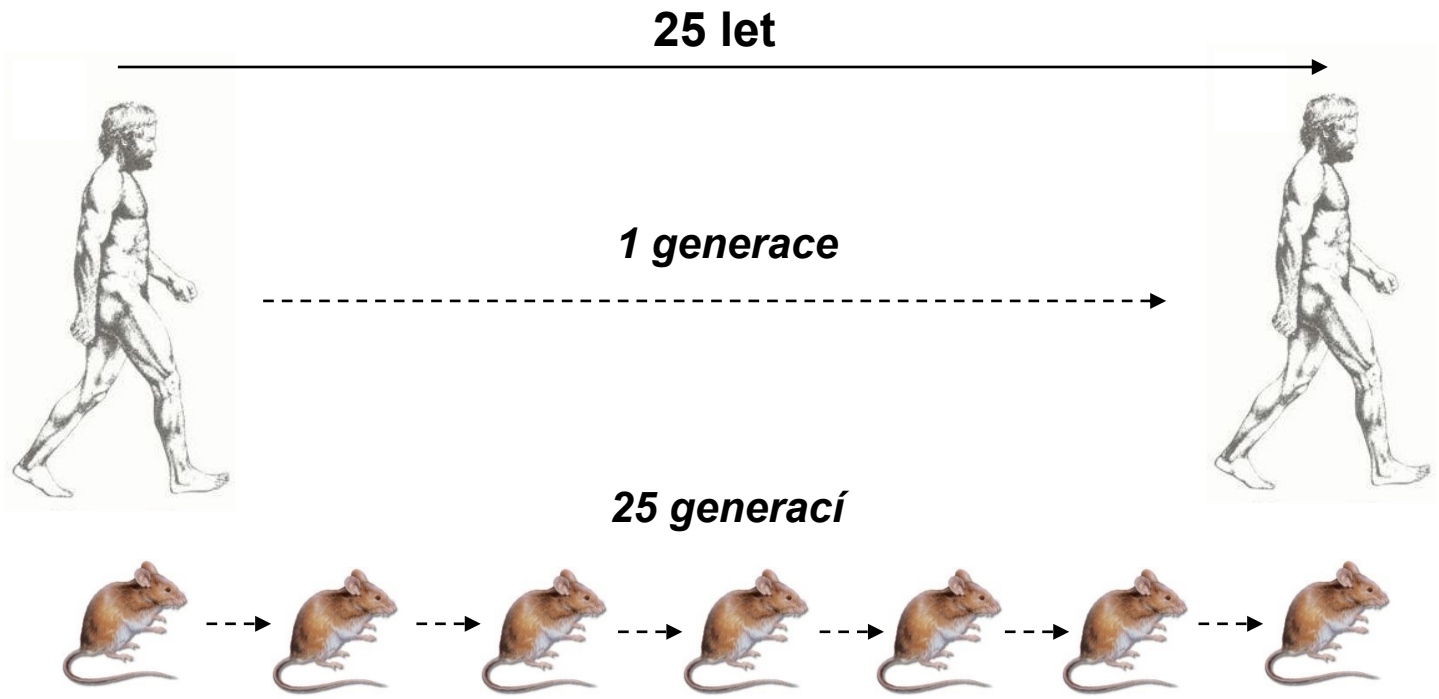




Proč je substituční rychlost vyšší u organismů s malou tělesnou velikostí?

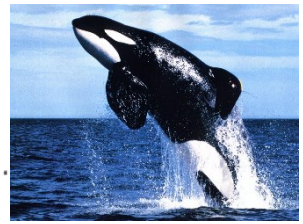
Hypotéza generační doby

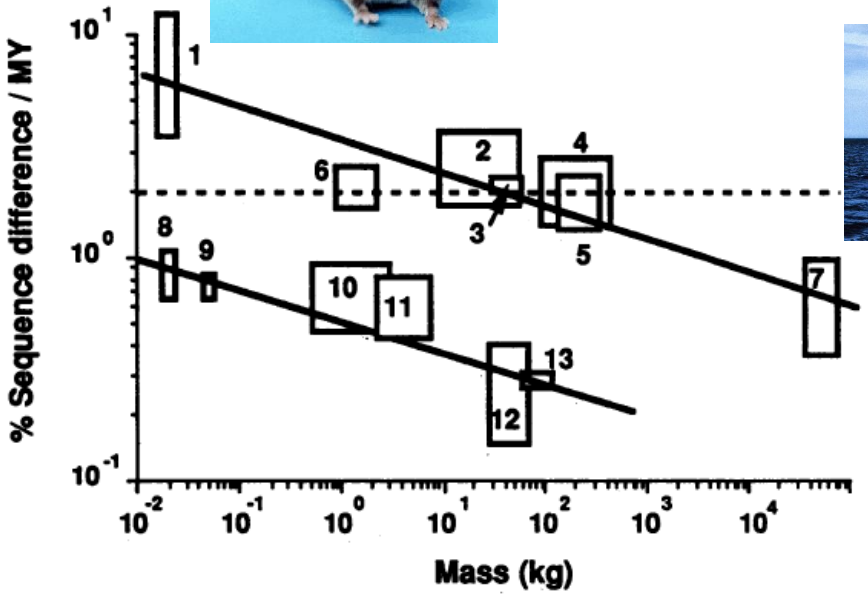
- Druhy s malou tělesnou velikostí mají obvykle kratší generační dobu. Za určitý časový interval se víckrát rozmnoží, v germinální linii proběhne celkově více buněčných dělení v přepočtu na jeden rok. Pokud většina mutací vzniká při replikaci, je u nich zvýšená mutační a tím pádem i substituční rychlost (vztažená na rok, ne však na generační dobu).



Metabolická hypotéza

- Mutační rychlost a tím pádem substituční rychlost je ovlivněná rychlostí metabolismu. Při vyšší rychlosti metabolismu, vyšší oxidativní stres. Druhy s malou velikostí mají rychlejší substituční rychlost.
- Vysvětlí proč teplokrevní mají vyšší substituční rychlost než studenokrevní.



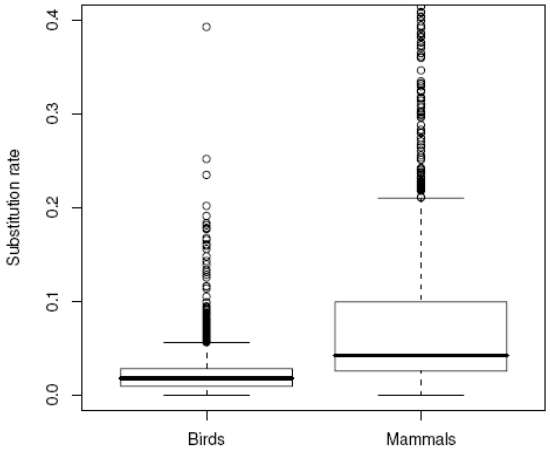


Hypotéza dlouhověkosti

- Dlouhověké druhy si vyvinuly mechanismy, jak se bránit mutační zátěži. Mají účinnější DNA reparační mechanismy. Mohou se tak dožít vysokého věku.

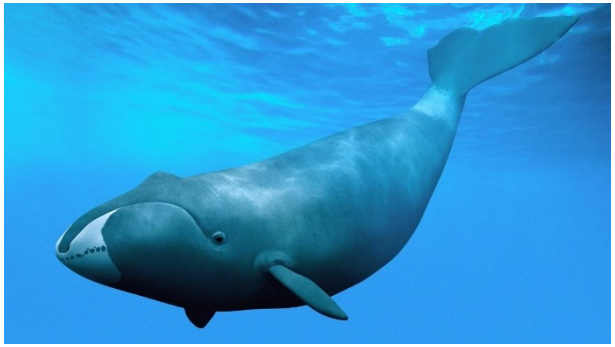
vysoký metabolismus

dlouhověcí



nízký metabolismus

krátkověcí



Velryba grónská

Teorie téměř neutrální evoluce

- neutrální mutace
  působí na ně drift
- škodlivé / výhodné mutace
  působí na ně zejména selekce
- mírně škodlivé / výhodné mutace
  působí na ně drift i selekce

V malých populacích se chovají spíše neutrálně,
ve velkých spíše jako výhodné či škodlivé.

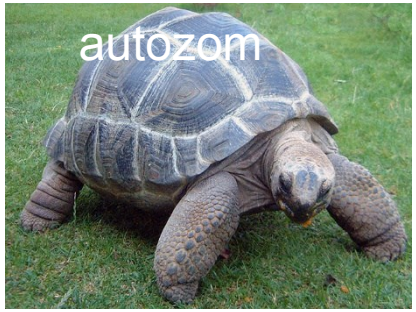
Tomoko Ohta

Téměř neutrální teorie evoluce předpovídá rychlejší substituční rychlost v malých populacích

- V malých populacích se fixuje více mutací než ve velkých (fixují se zde mírně škodlivé mutace).
- Druhy s malými populacemi mají rychlejší substituční rychlost než druhy s velkými populacemi. Zároveň se u nich fixuje více škodlivých mutací.

>

>

Rychlejší molekulární evoluce pohlavních chromosomů

- Platí i pro části genomu s nižší N_e (pohlavní chromozomy, mtDNA). Rychlejší evoluce na pohlavních chromosomech či mtDNA ve srovnání s autosomy.