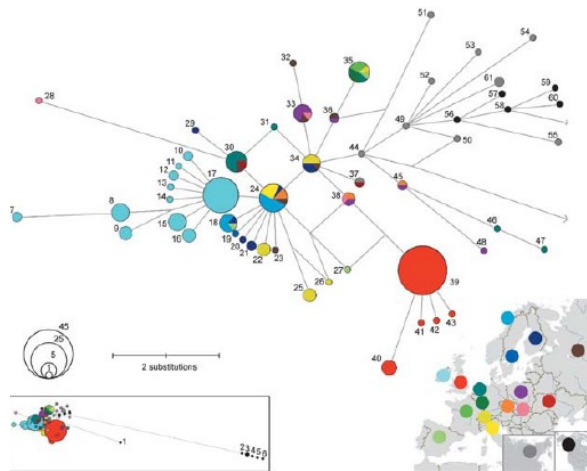
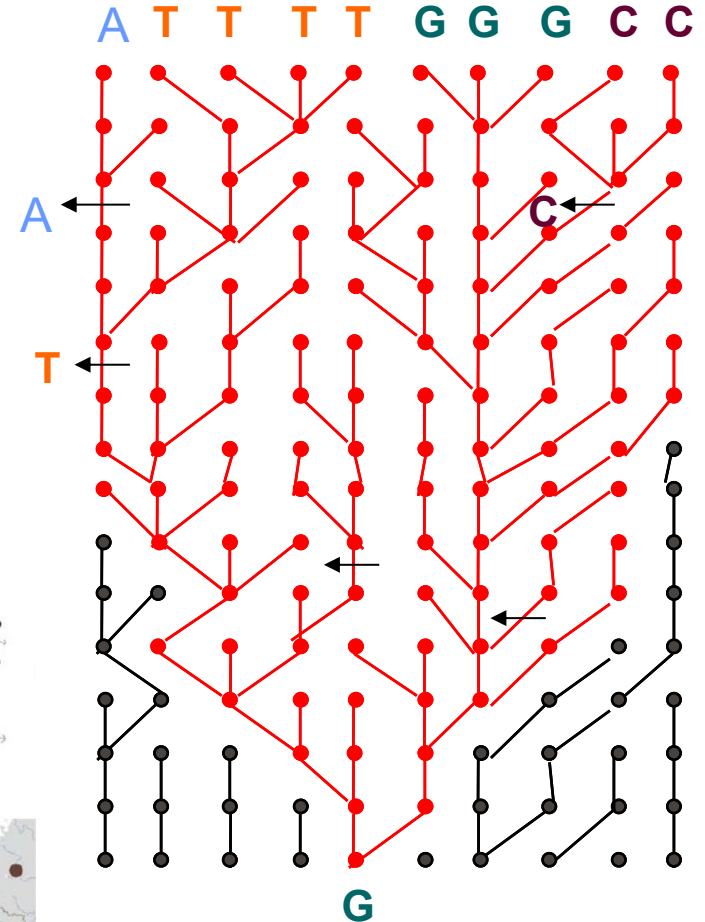
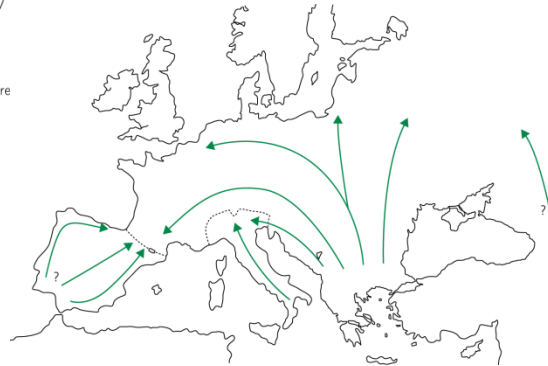
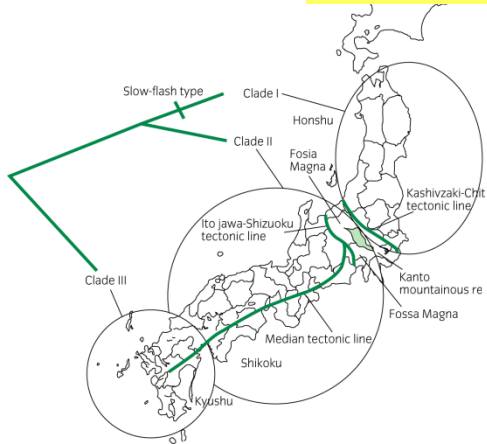


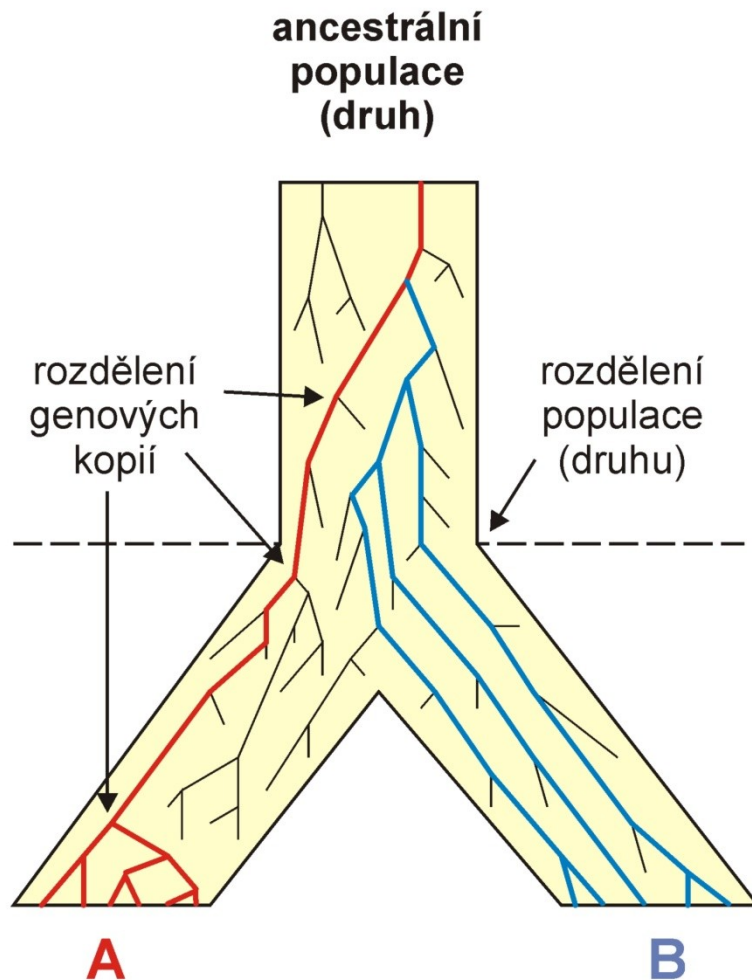
# KOALESCENCE A FYLOGEOGRAFIE





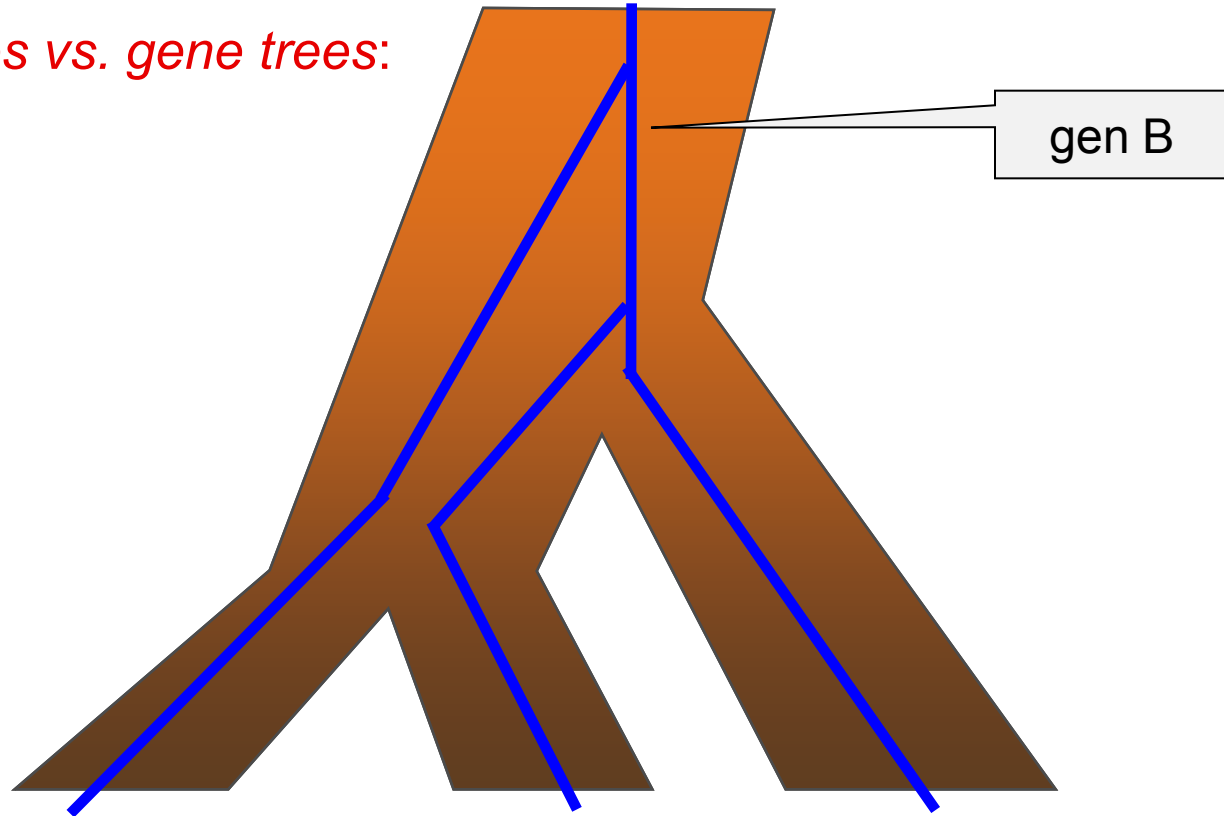
# KOALESCENCE

Osud jednotlivých kopií genů v populaci → **genové stromy**

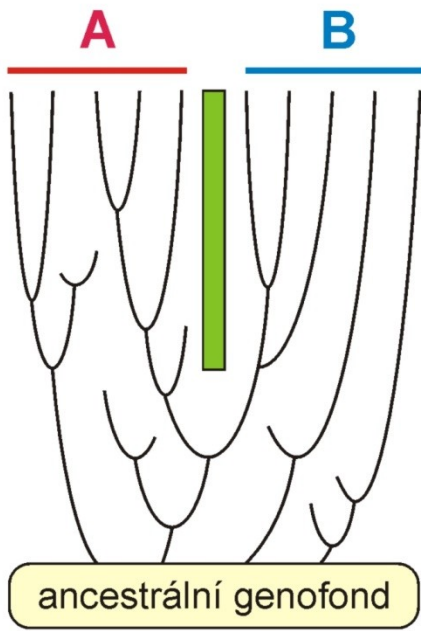




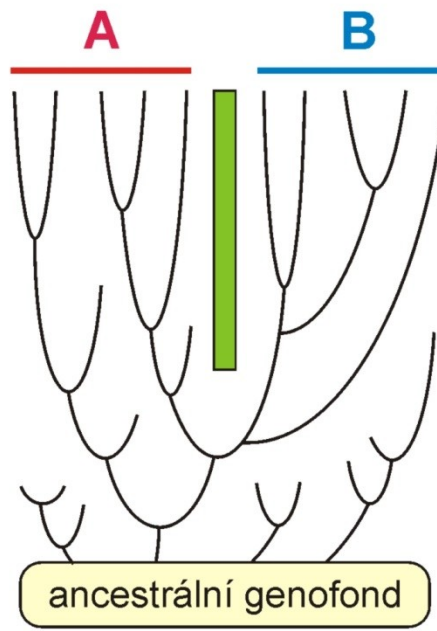
*Species trees vs. gene trees:*



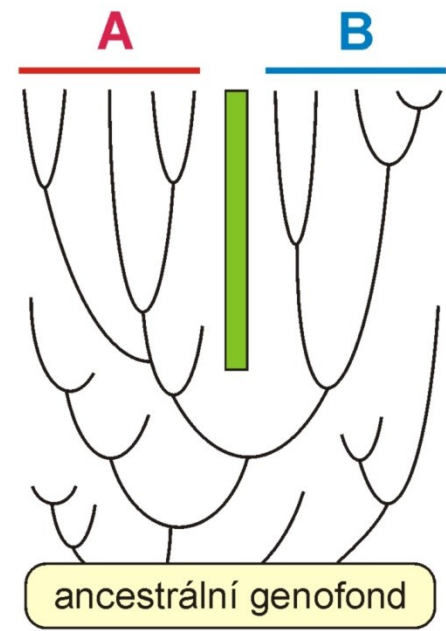
## Fylogenetické vztahy 2 dceřiných populací (např. mtDNA):



polyfylye

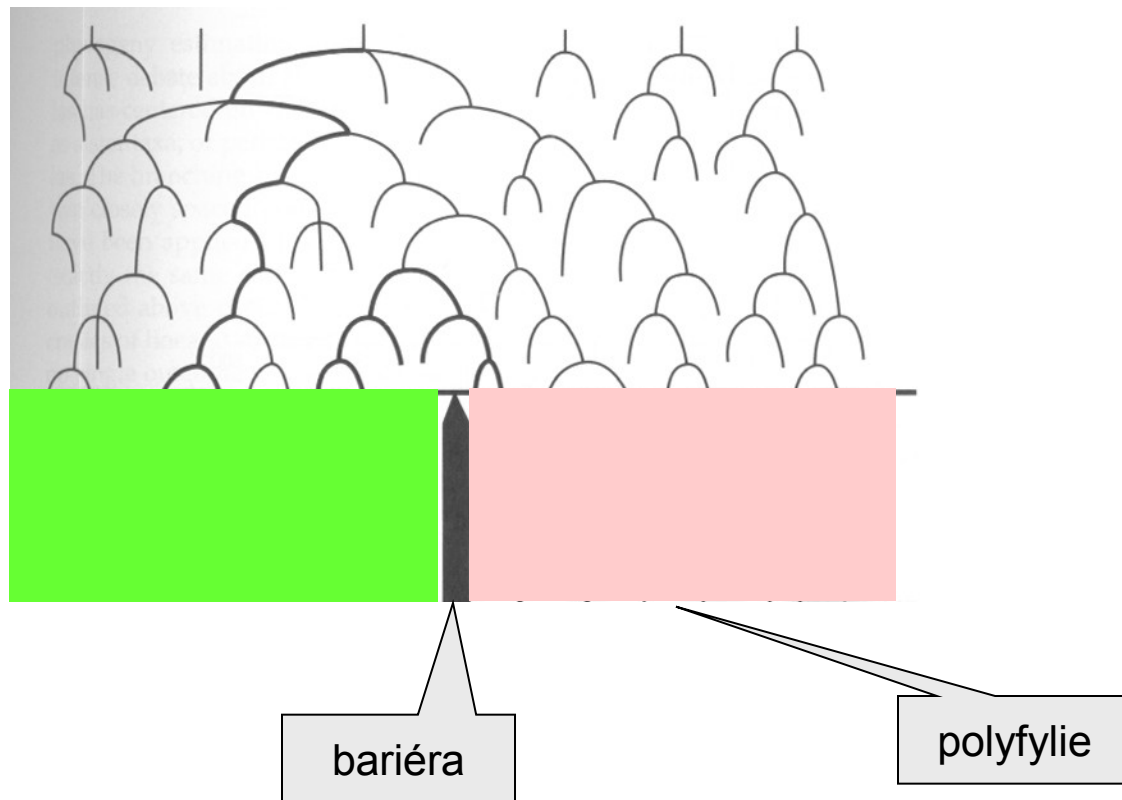


parafylye

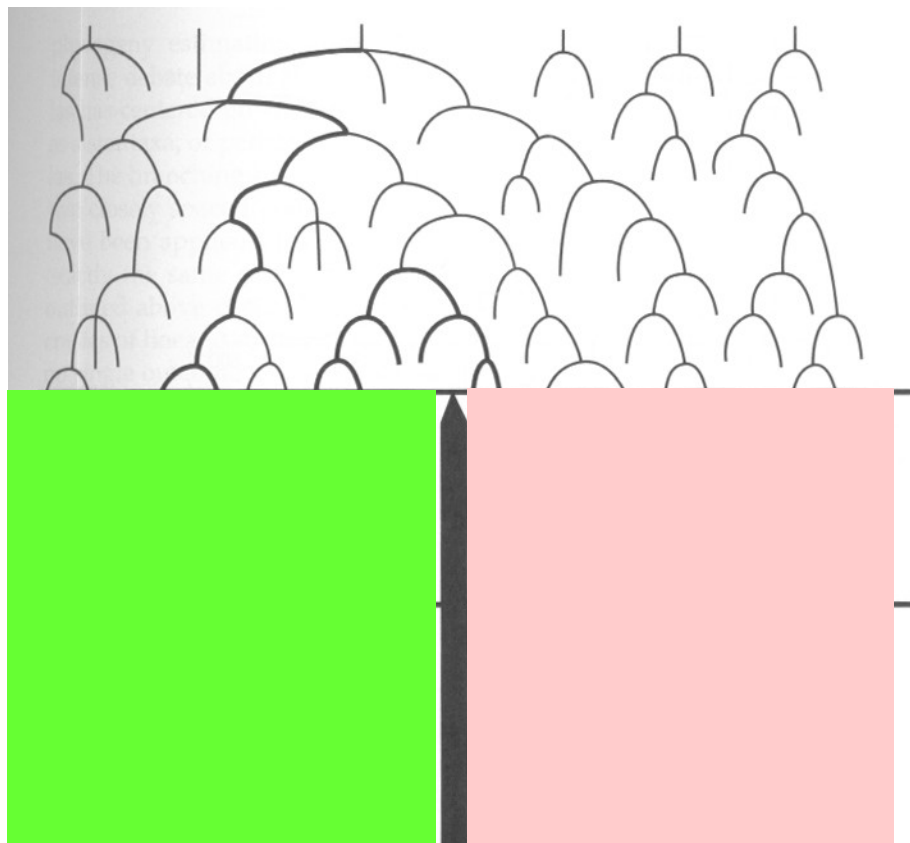


reciproční monofylye

# Ancestrální polymorfismus a sortování linií



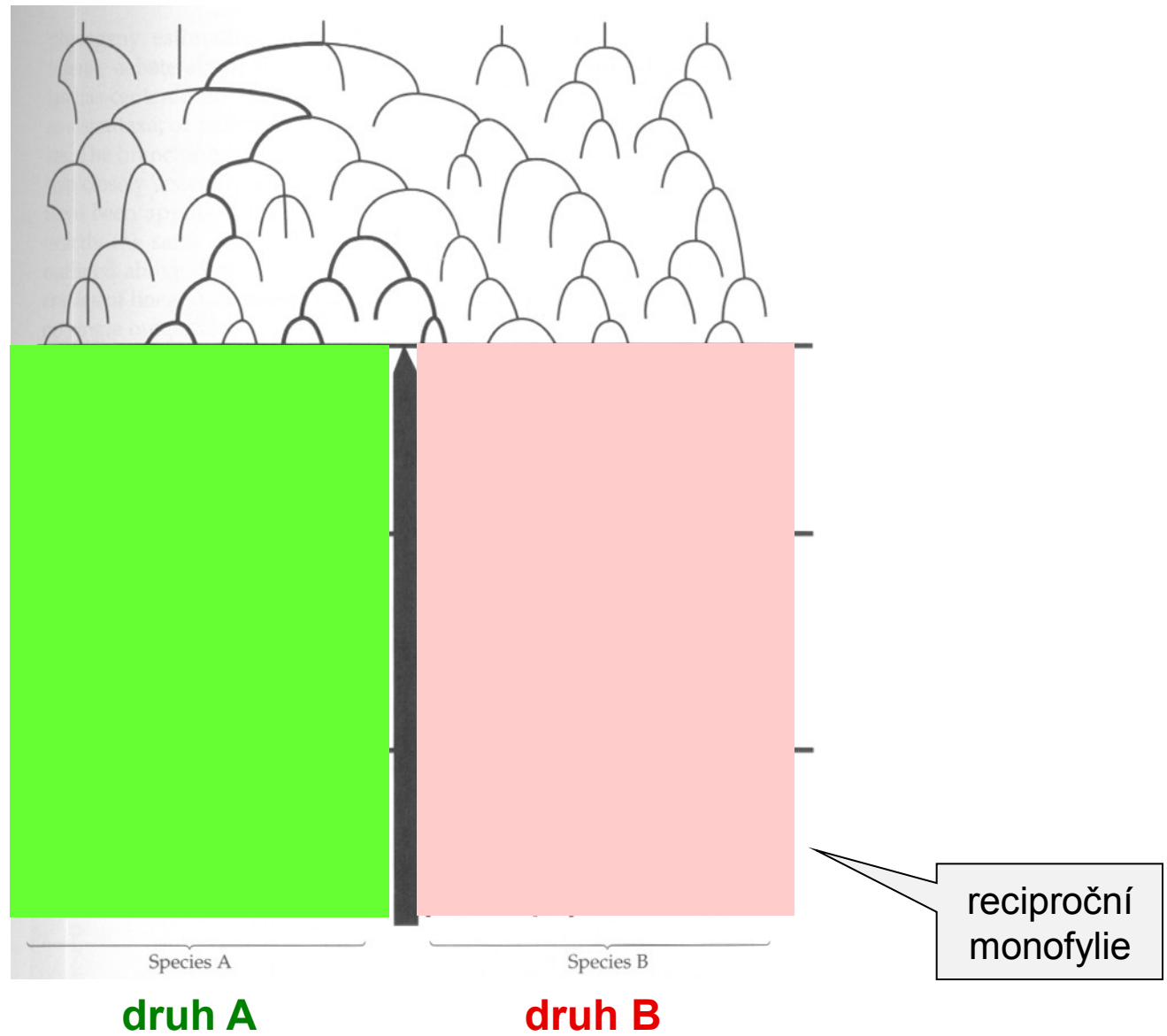
# Ancestrální polymorfismus a sortování linií



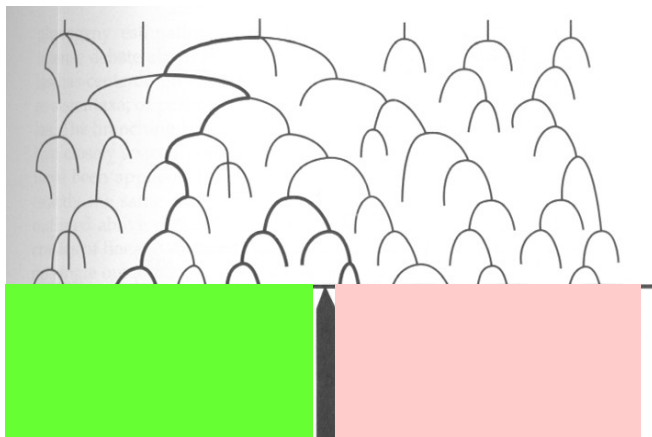
parafyletická  
fáze



# Ancestrální polymorfismus a sortování linií



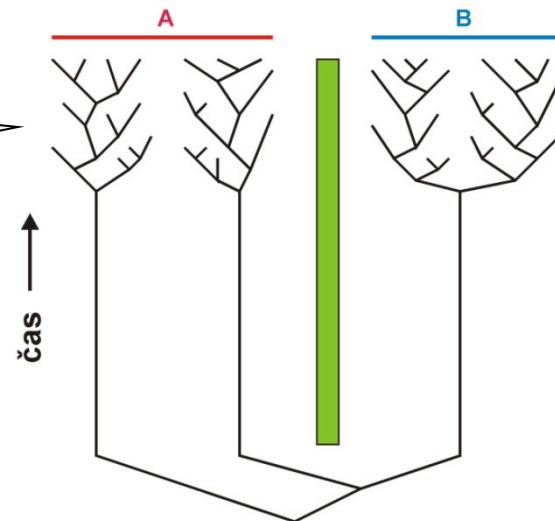
# Ancestrální polymorfismus a sortování linií



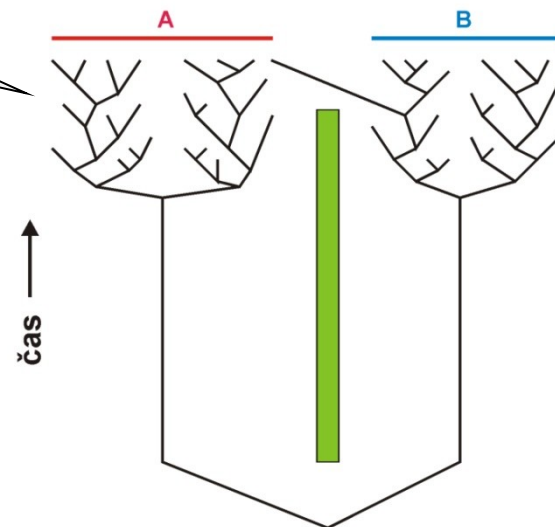
druh A

druh B

nekompletní  
sortování linií



recentní  
tok genů

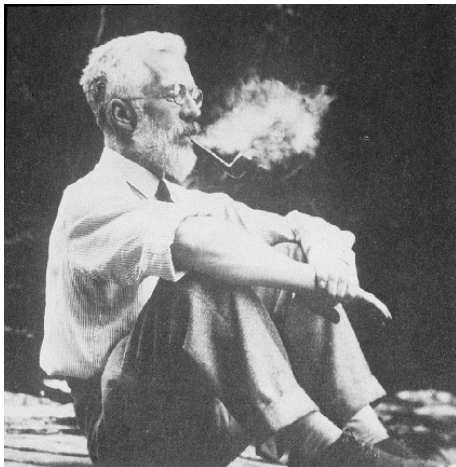


Problém: „incomplete lineage sorting“ je většinou obtížné odlišit od důsledků toku genů

# Wrightův-Fisherův model:



Sewall Wright



Ronald A. Fisher

W-F populace:

diploidní, hermafrodit

velikost omezená, žádné fluktuace  $N$

náhodné oplození

kompletní izolace (žádný tok genů)

diskrétní generace

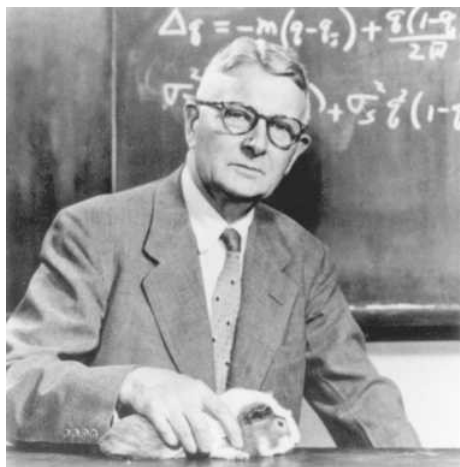
žádná věková struktura

žádná selekce

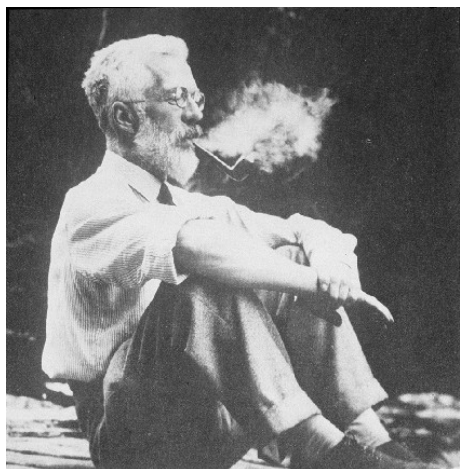
rozptyl výběru gamet do další generace

→ Poissonovo rozdělení

# Sortování linií ve W-F modelu:



Sewall Wright



Ronald A. Fisher

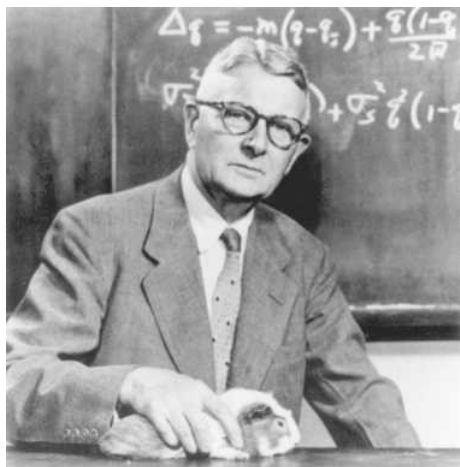
generace 1



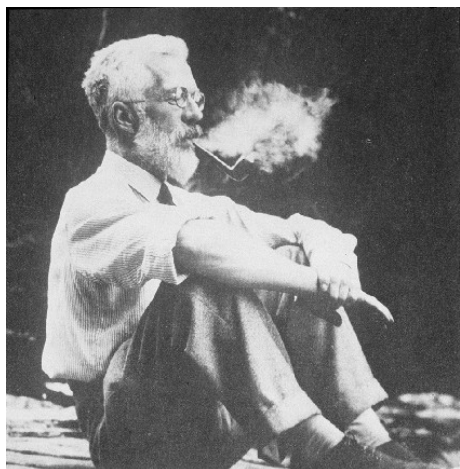
čas



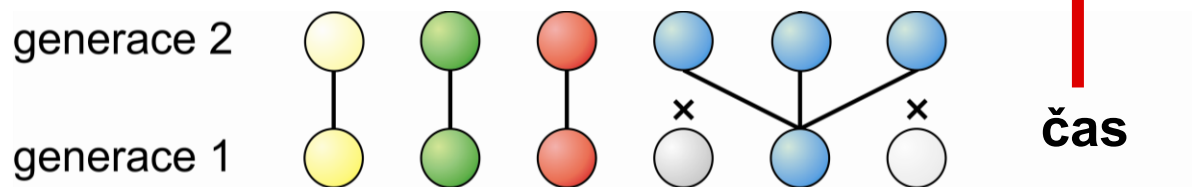
# Sortování linií ve W-F modelu:



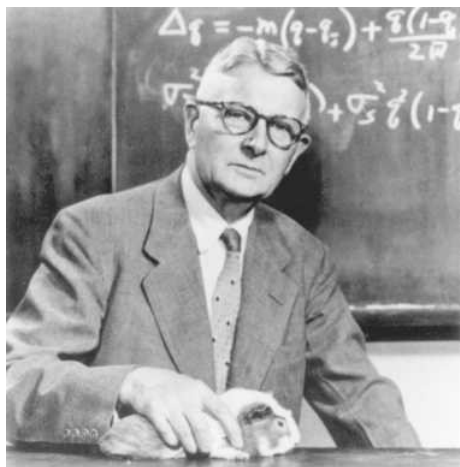
Sewall Wright



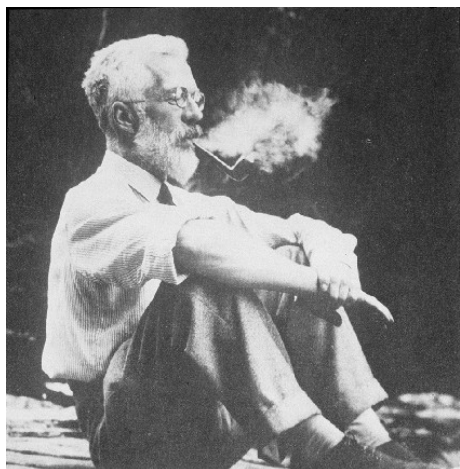
Ronald A. Fisher



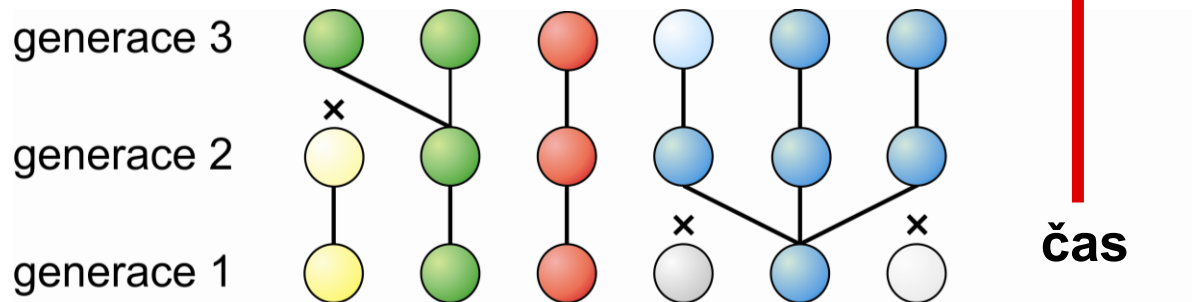
# Sortování linií ve W-F modelu:



Sewall Wright



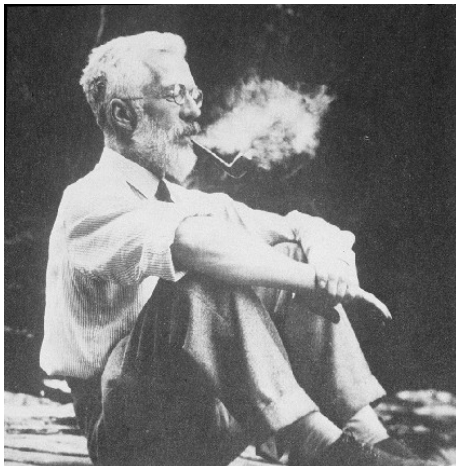
Ronald A. Fisher



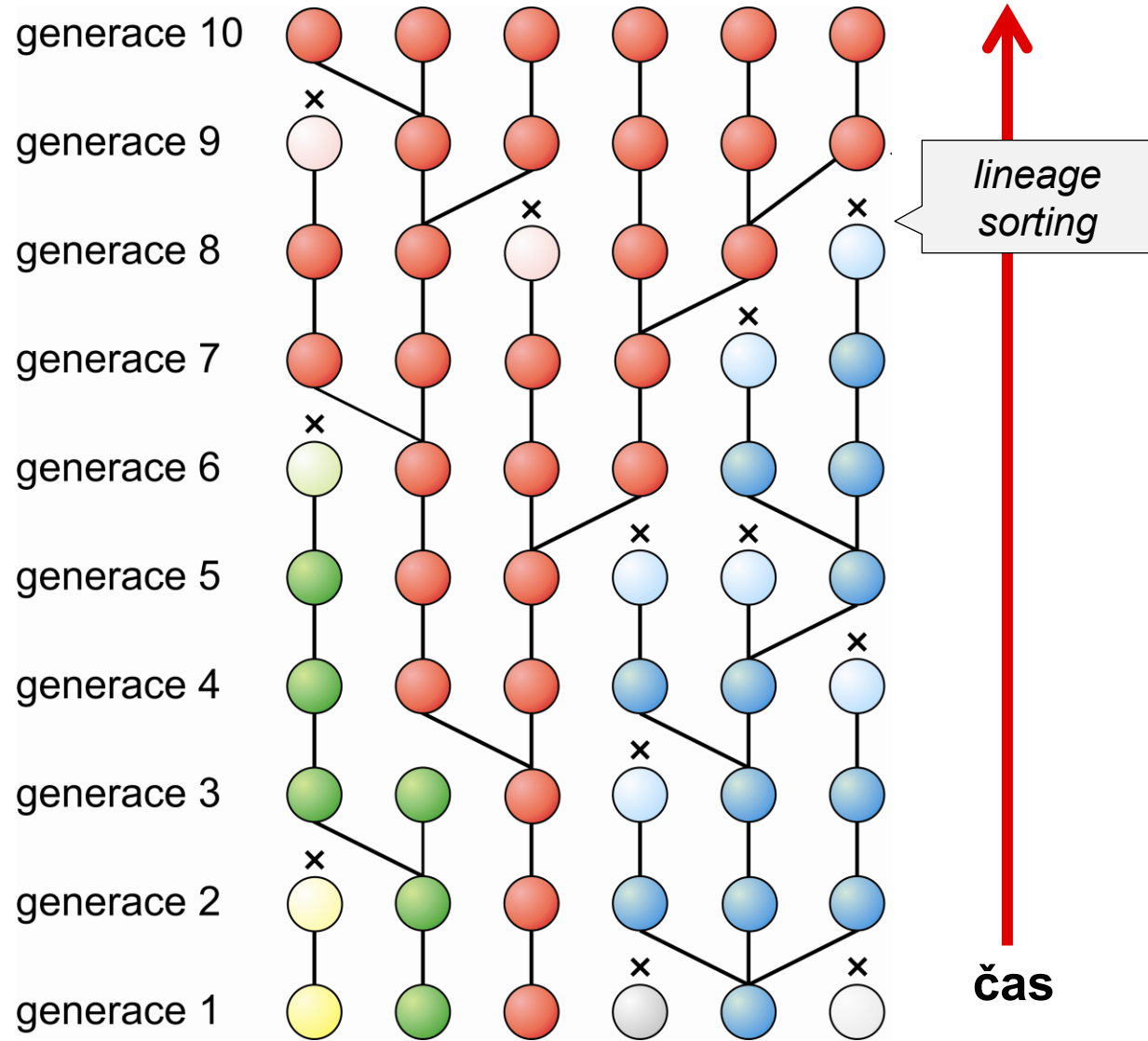
# Sortování linií ve W-F modelu:

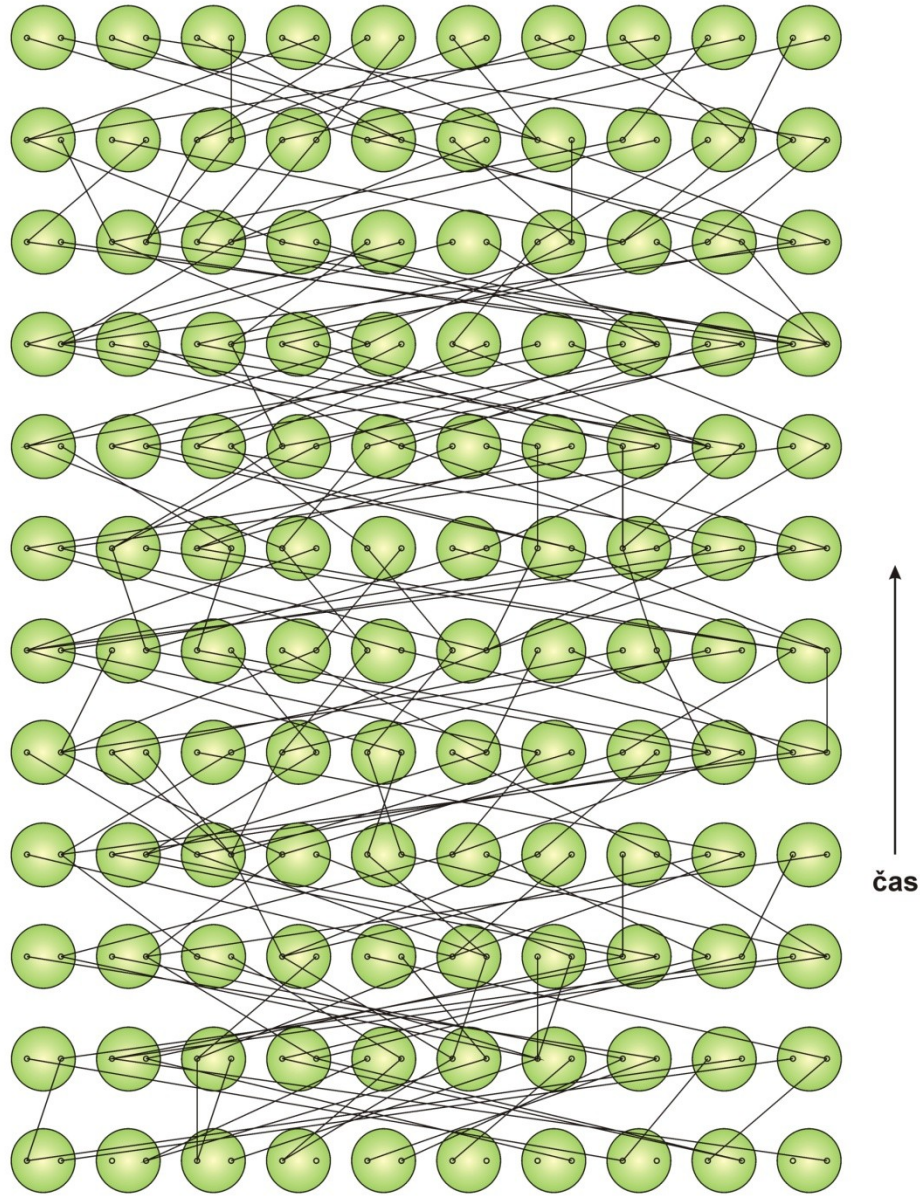


Sewall Wright

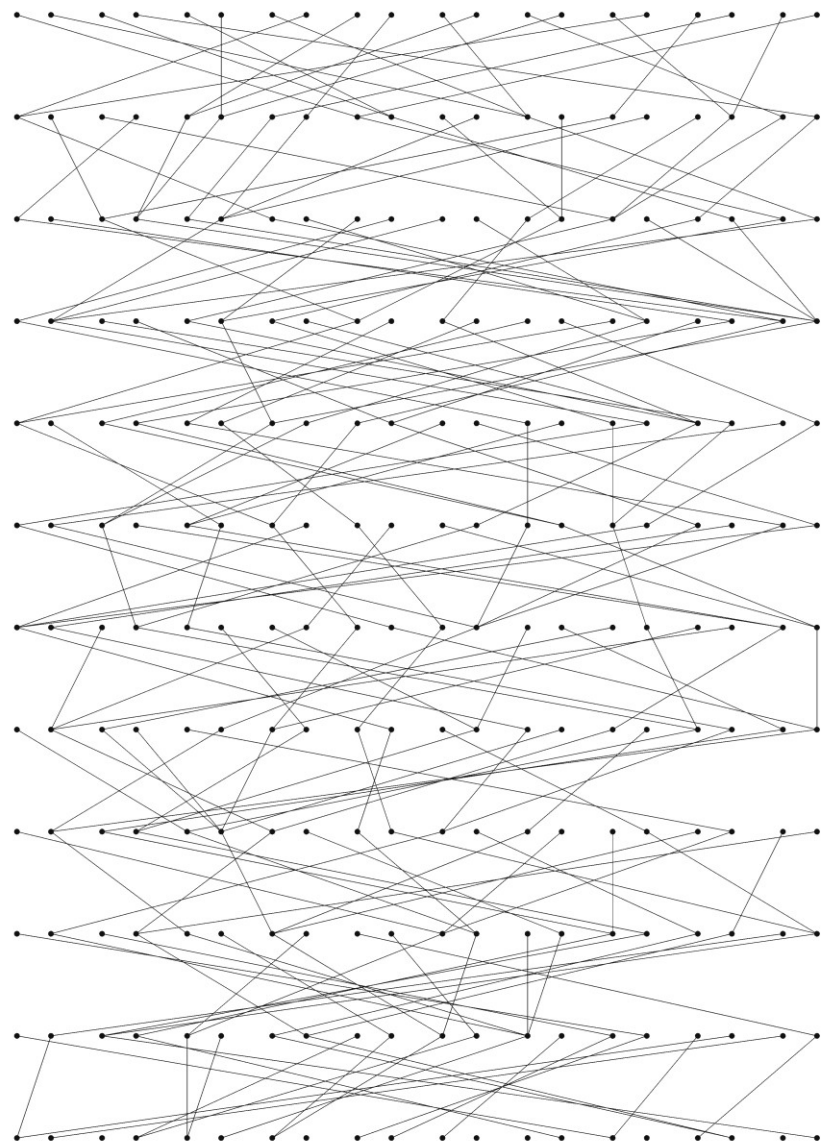


Ronald A. Fisher

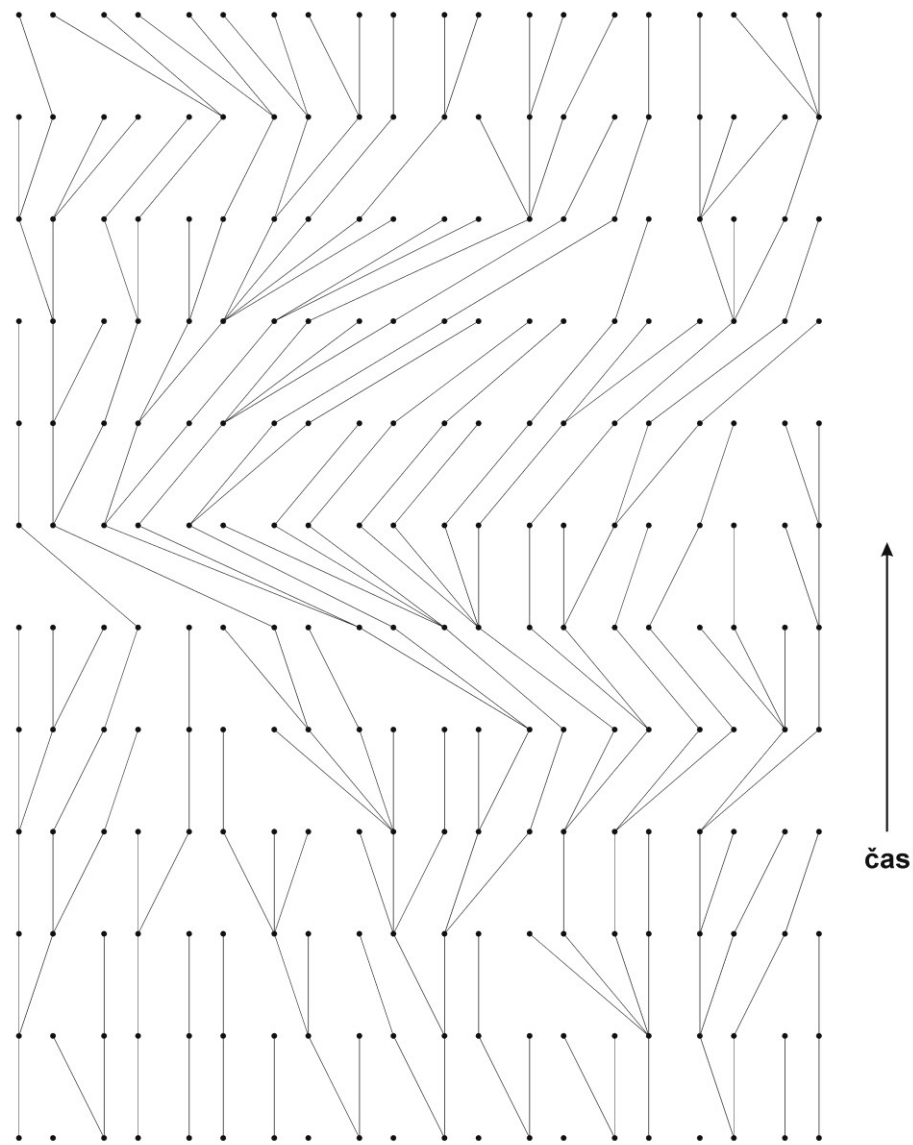




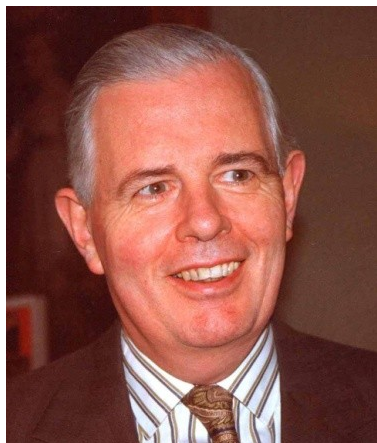




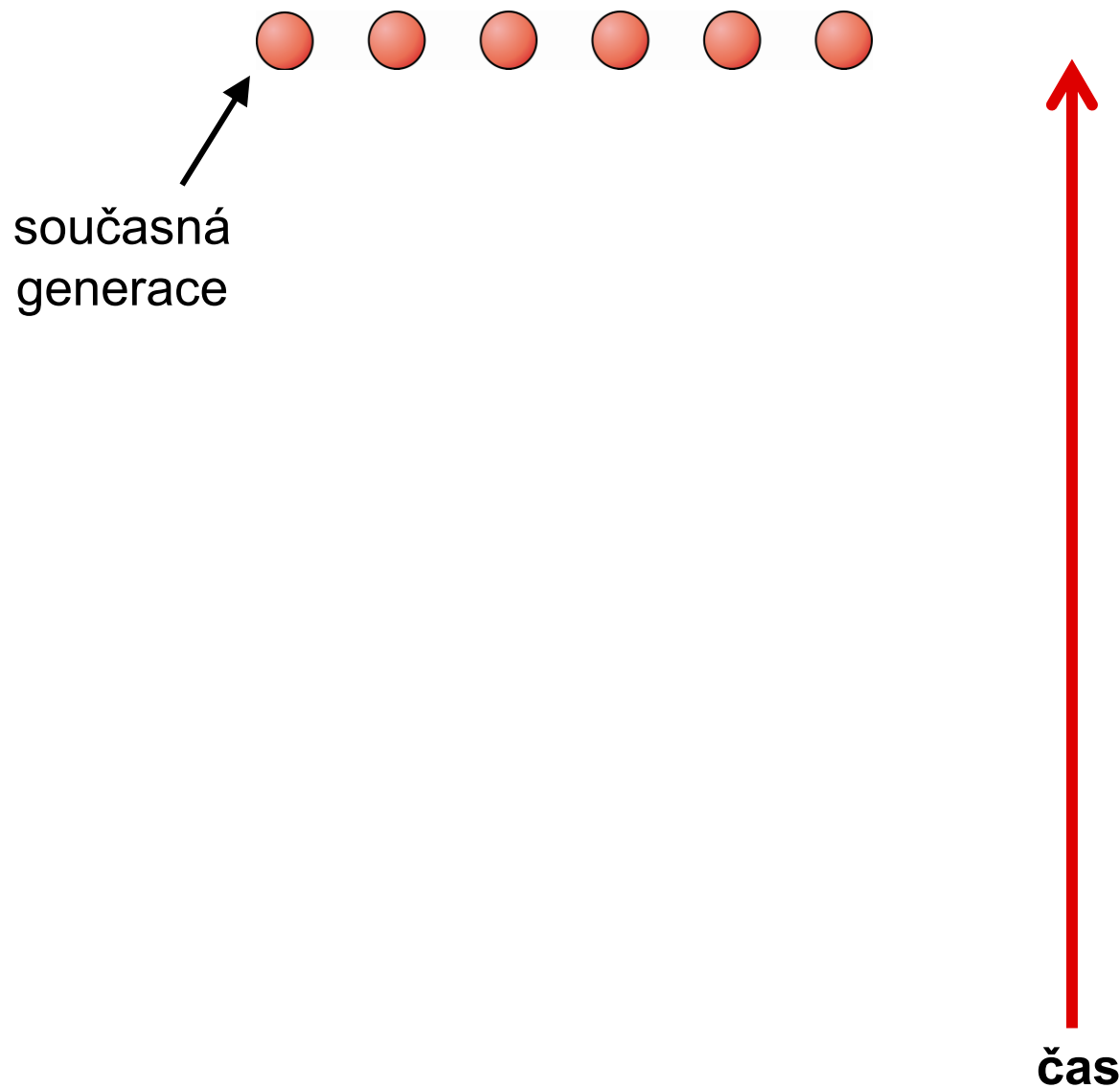
↑  
čas



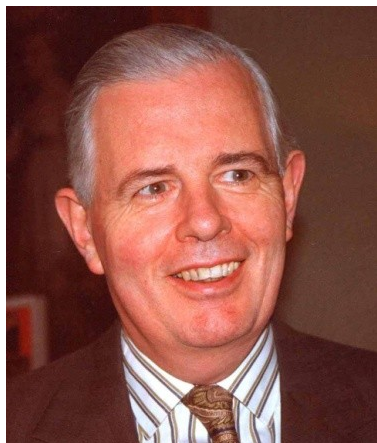
# Koalescence:



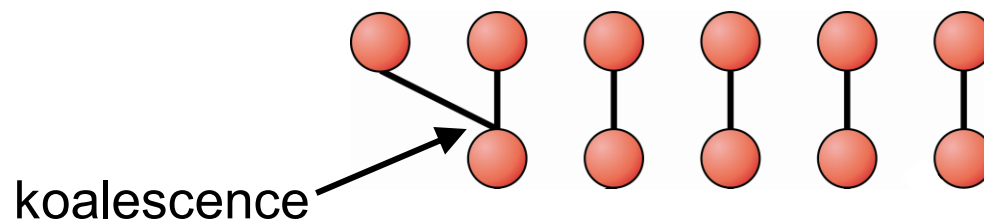
John F.C. Kingman



# Koalescence:

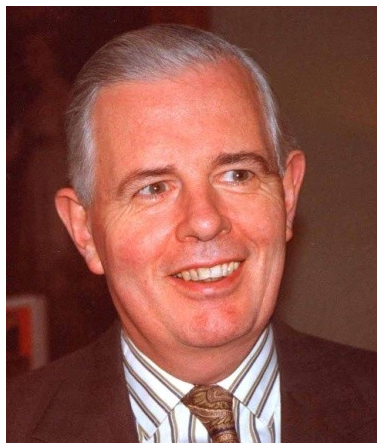


John F.C. Kingman

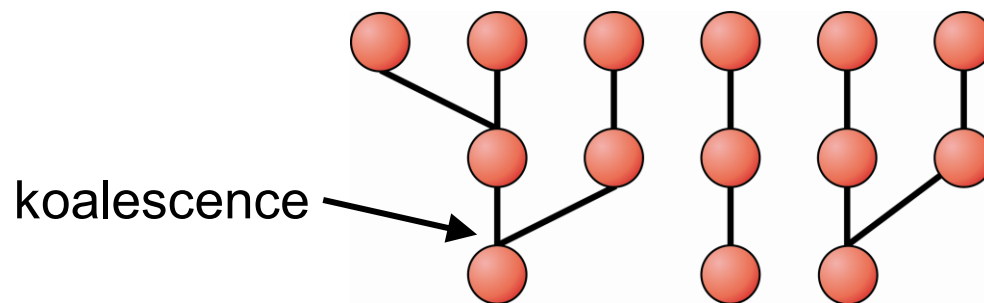


čas

# Koalescence:

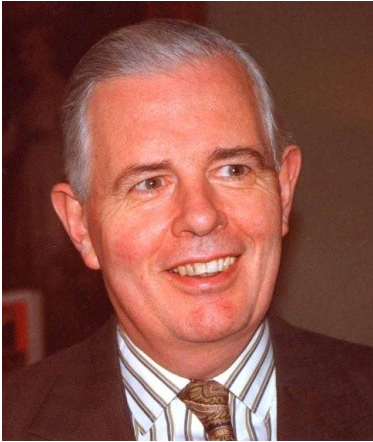


John F.C. Kingman

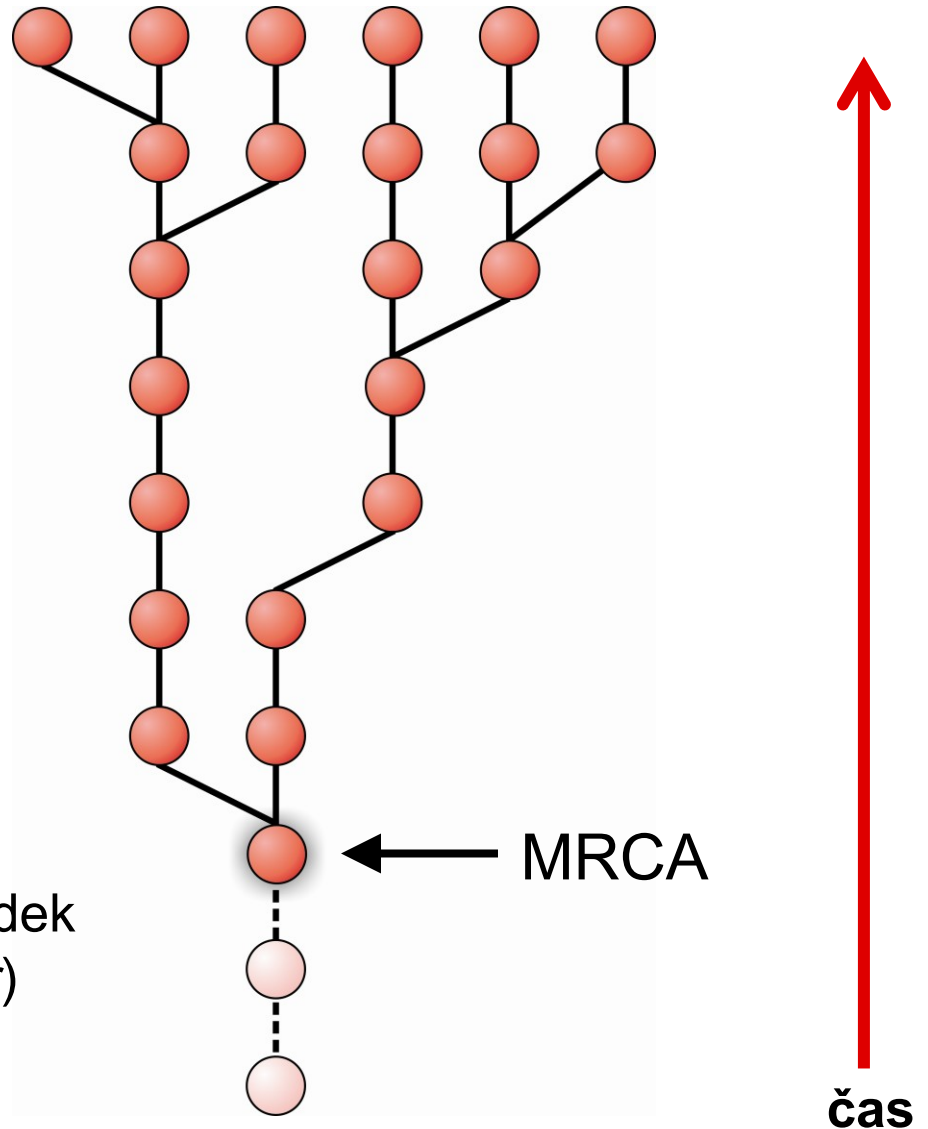


čas

# Koalescence:

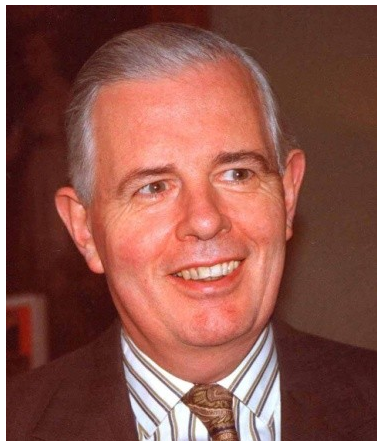


John F.C. Kingman



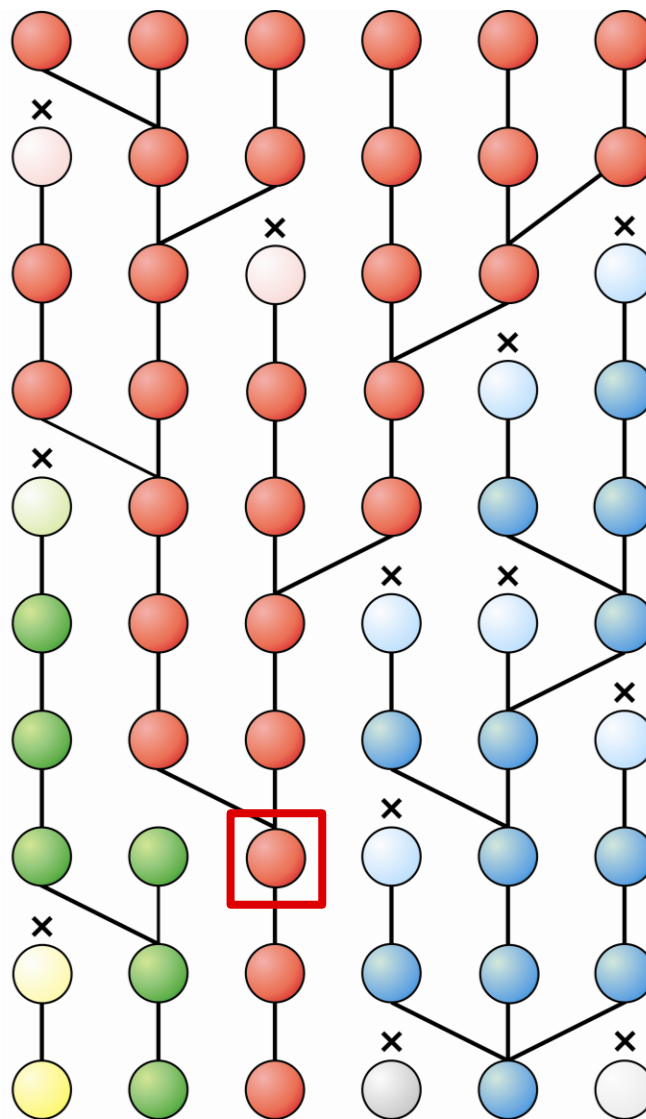
MRCA = nejblížší společný předek  
(*most recent common ancestor*)

# Koalescence:



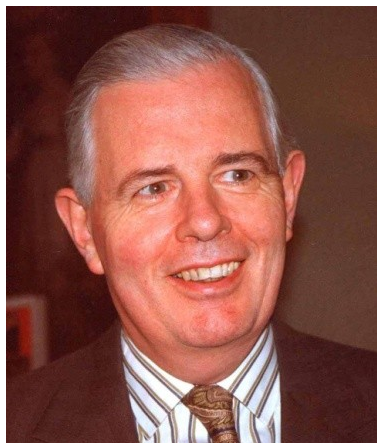
John F.C. Kingman

nevíme, kolik kopií bylo v generaci MRCA

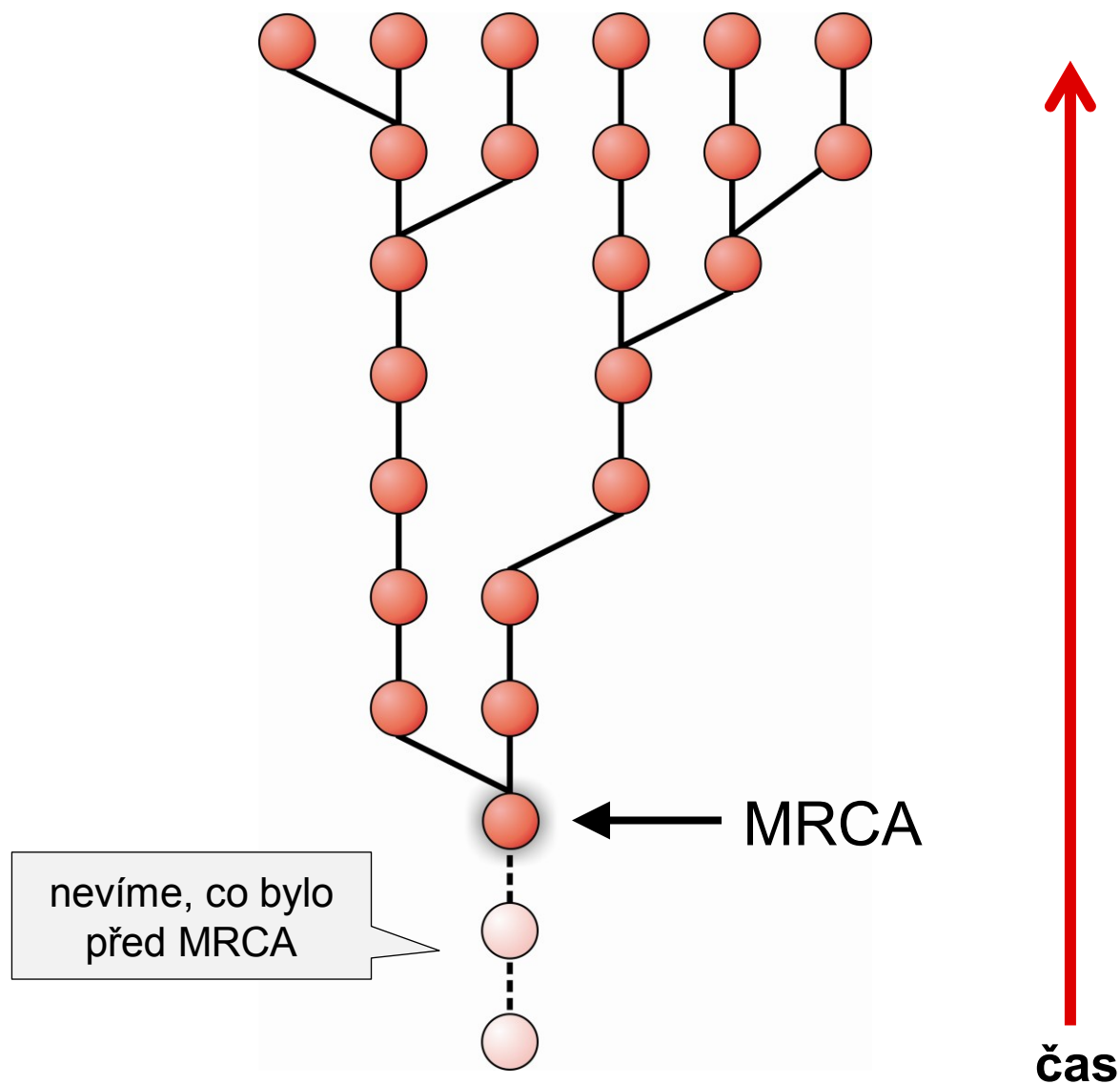


čas

# Koalescence:



John F.C. Kingman

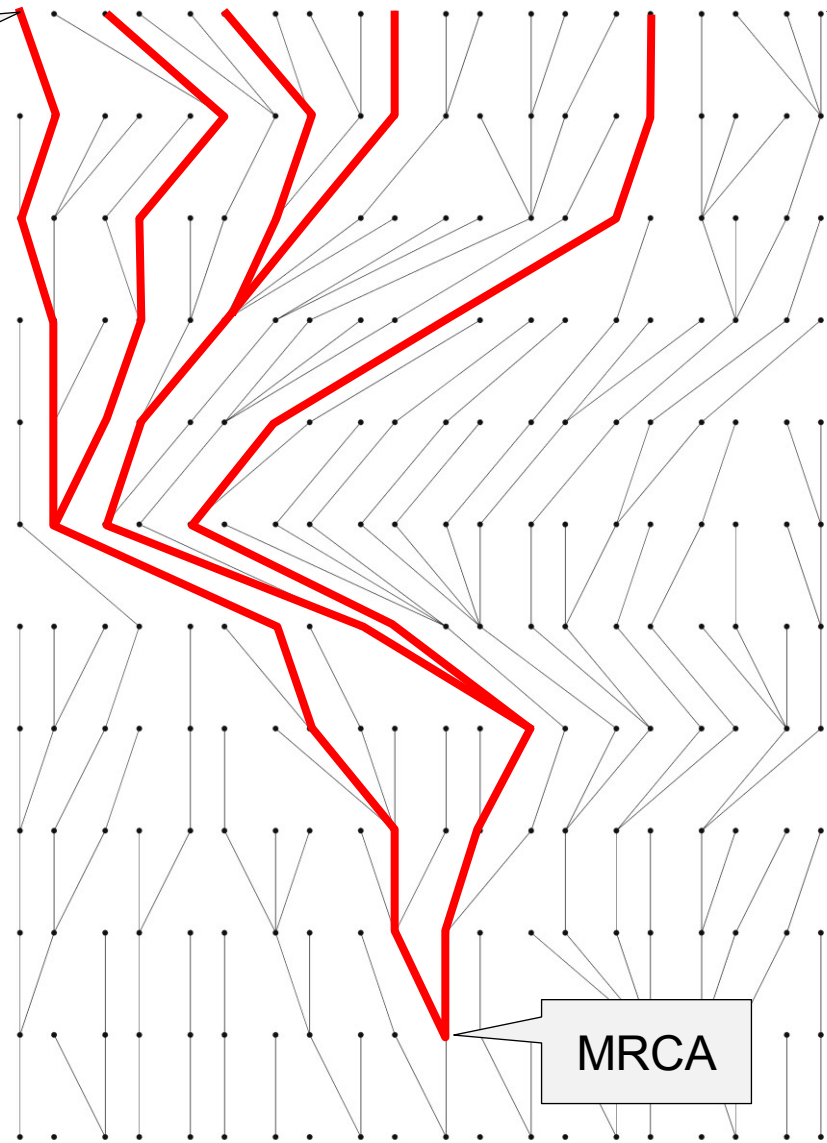




$n = 5$  kopií  
ve vzorku

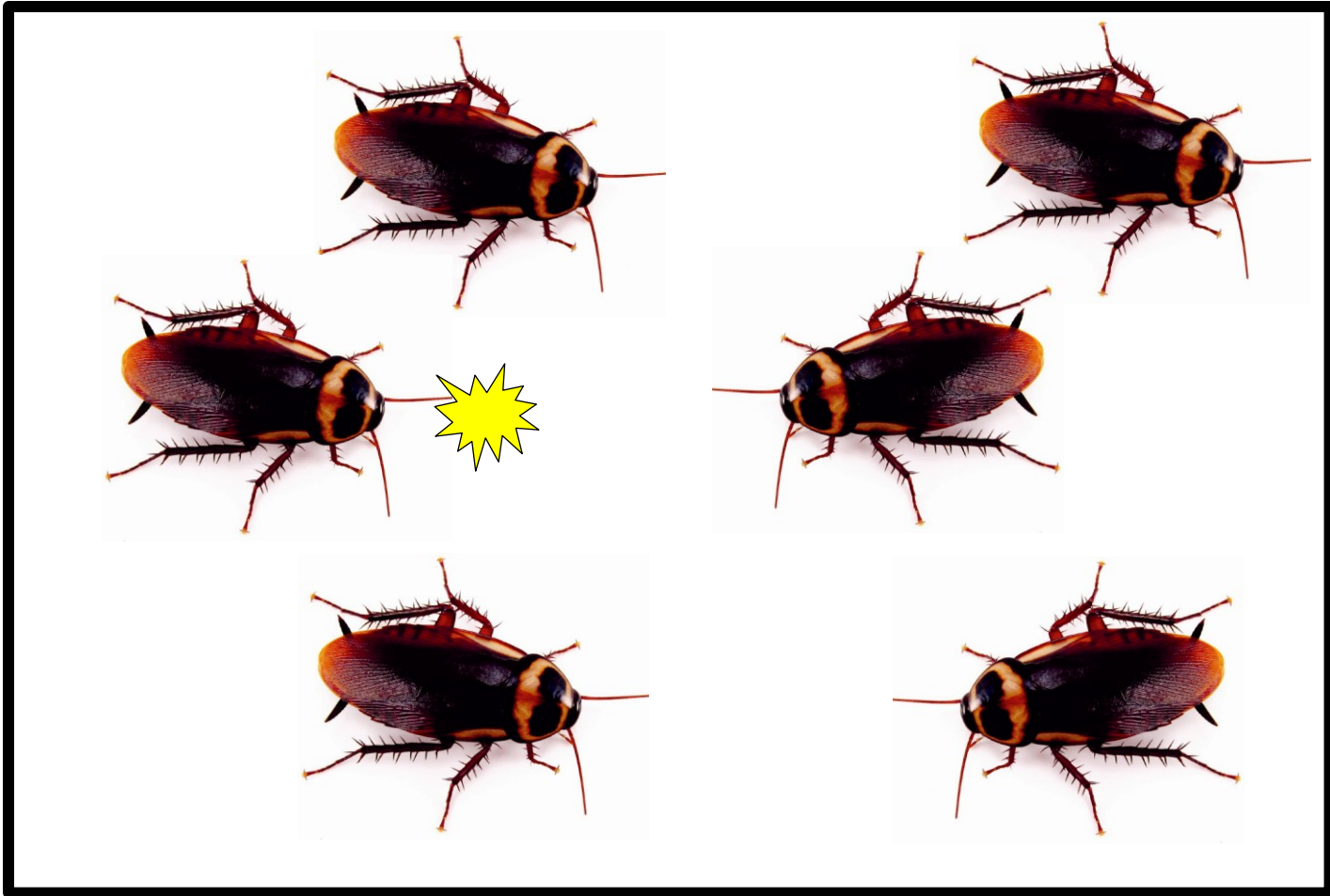
většinou  
 $n \ll N$

$N = 20$  kopií  
v populaci

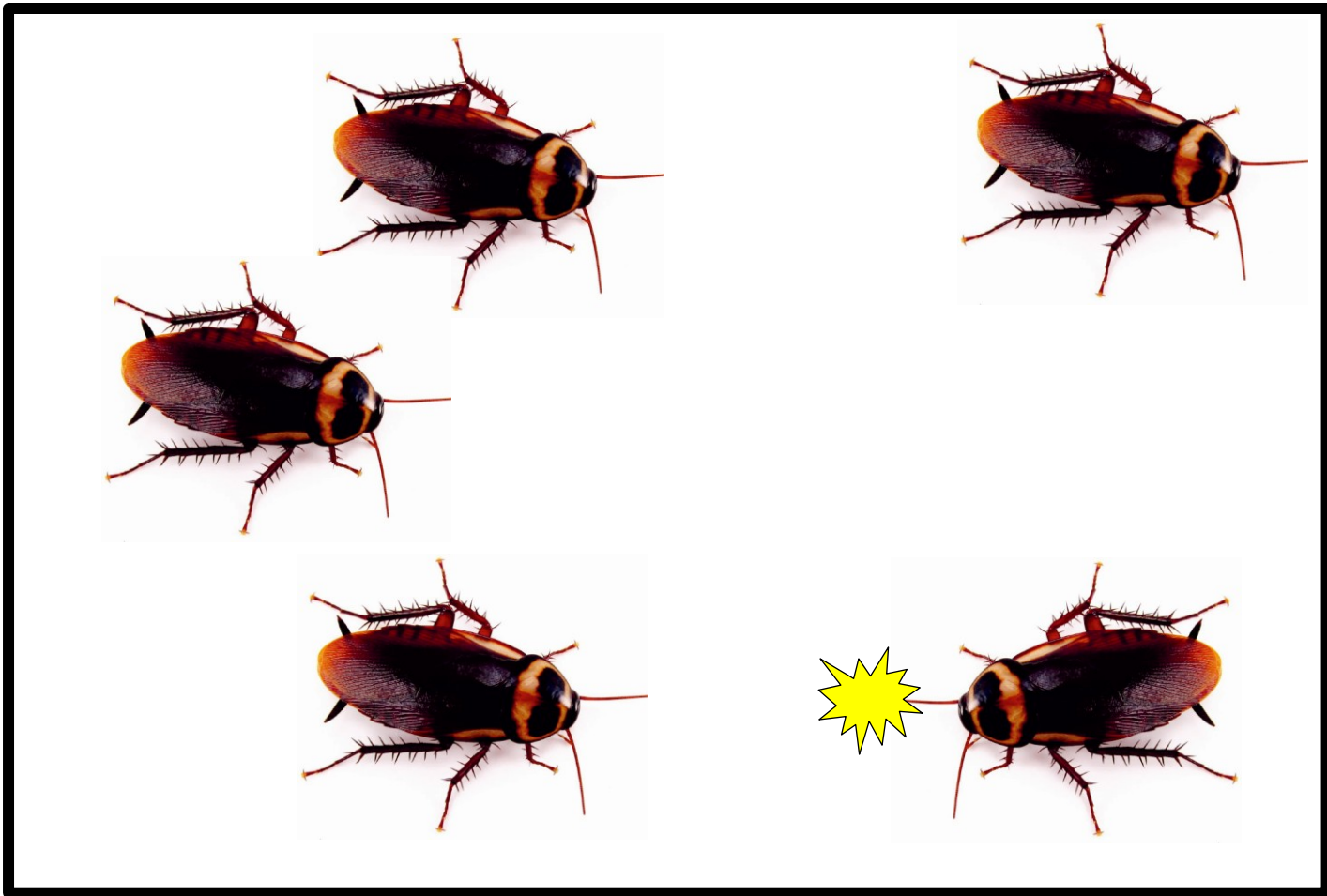


MRCA

čas



Pravděpodobnost setkání 2 švábů je  $n(n - 1)/4N$ , kde  
 $n$  = počet švábů v krabici,  $N$  = počet „míst“ v krabici



při koalescenci se počet švábů (kopií) sníží o 1 ...

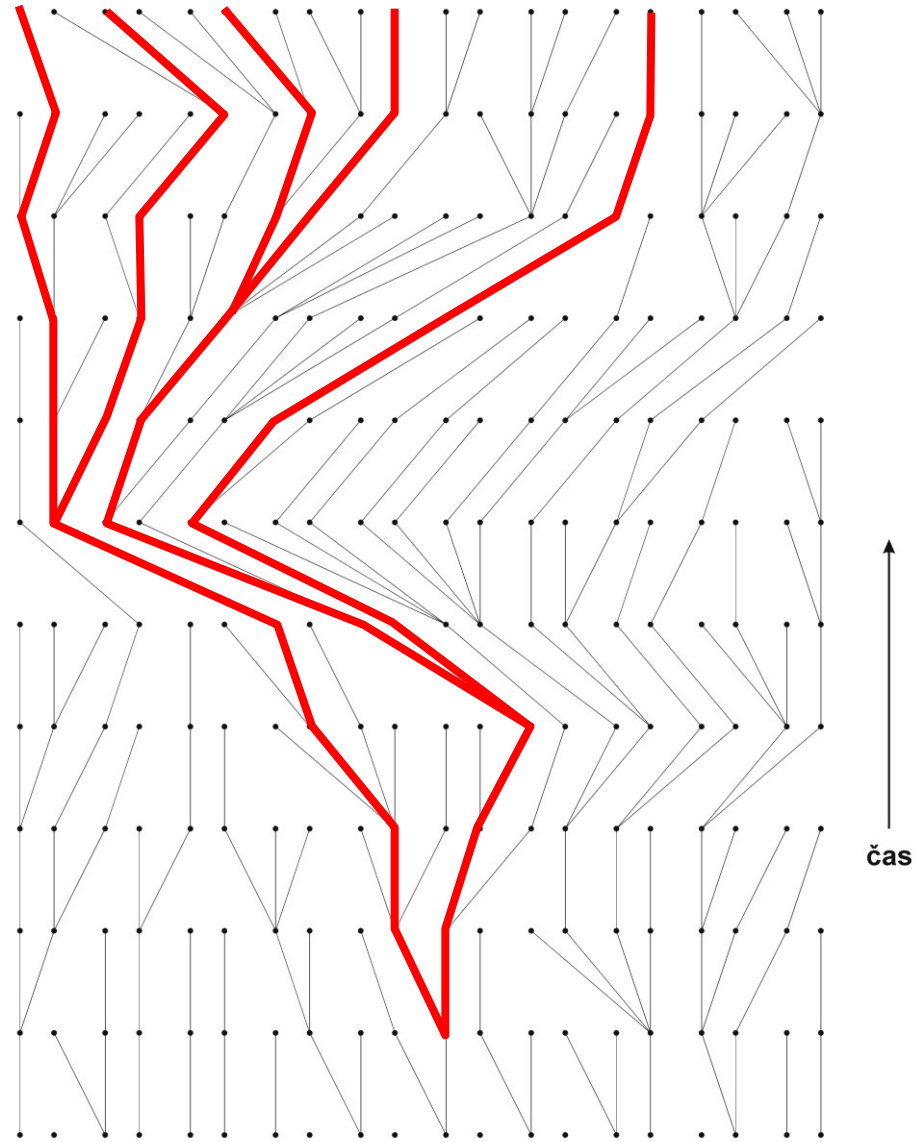
s tím, jak klesá počet švábů ( $n$ ), roste čas k dalšímu kontaktu (koalescenci)

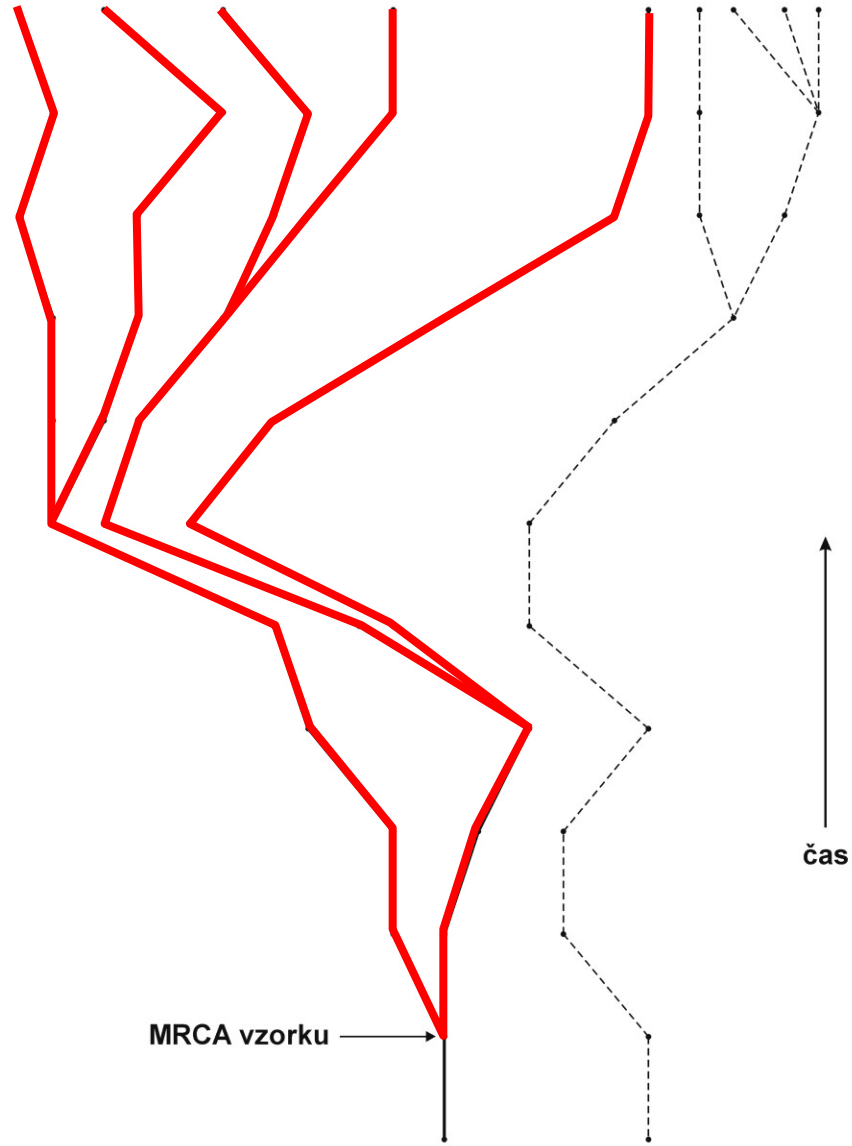


při koalescenci se počet švábů (kopií) sníží o 1 ...



... až nakonec zůstane jen 1 kopie





## Kingmanova koalescence:

čím menší počet zbývajících kopií, tím se **proces koalescence zpomaluje** (pro velká  $n \sim 4N$ , pro 2 kopie  $\sim 2N$ )

splynutí posledních  $k$  kopií zabere  $(1 - 1/n)/(1 - 1/k)$

⇒ **prvních 90% kopií splyne během 9% celkového času, zbývajících 91% času se čeká na splynutí posledních 10 kopií!**

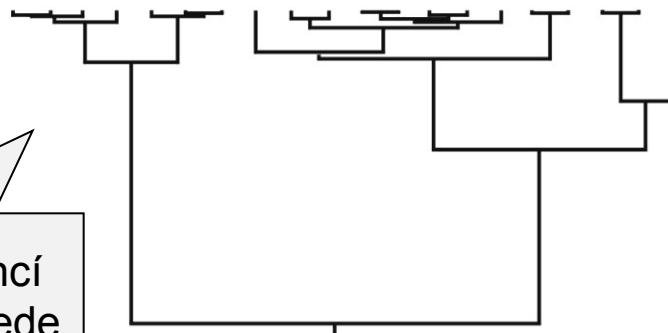
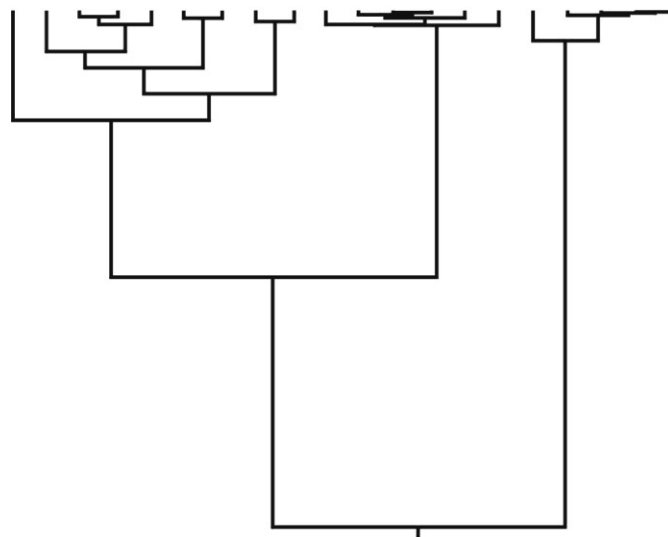
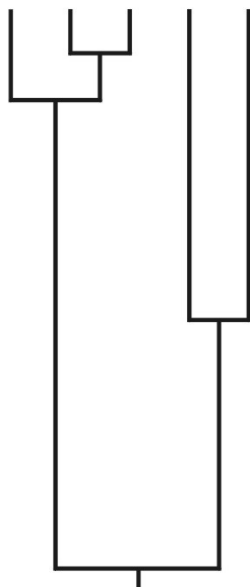
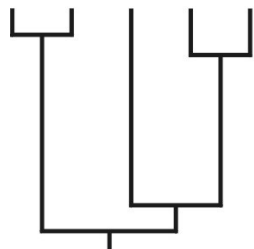
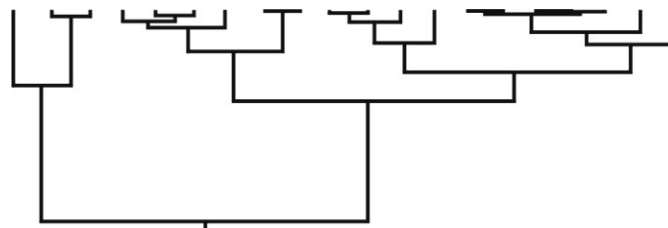
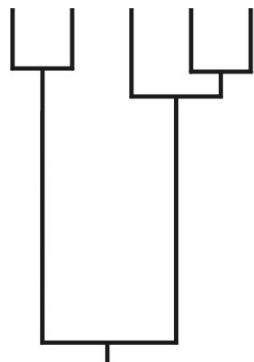
jestliže 100 linií, pravděpodobnost, že 101. linie přidá hlubší kořen je pouze 0,02% ⇒ **přidání další genové kopie pravděpodobně nepovede k hlubší (starší) koalescenci**



rozdělení času mezi koalescencemi je přibližně exponenciální:



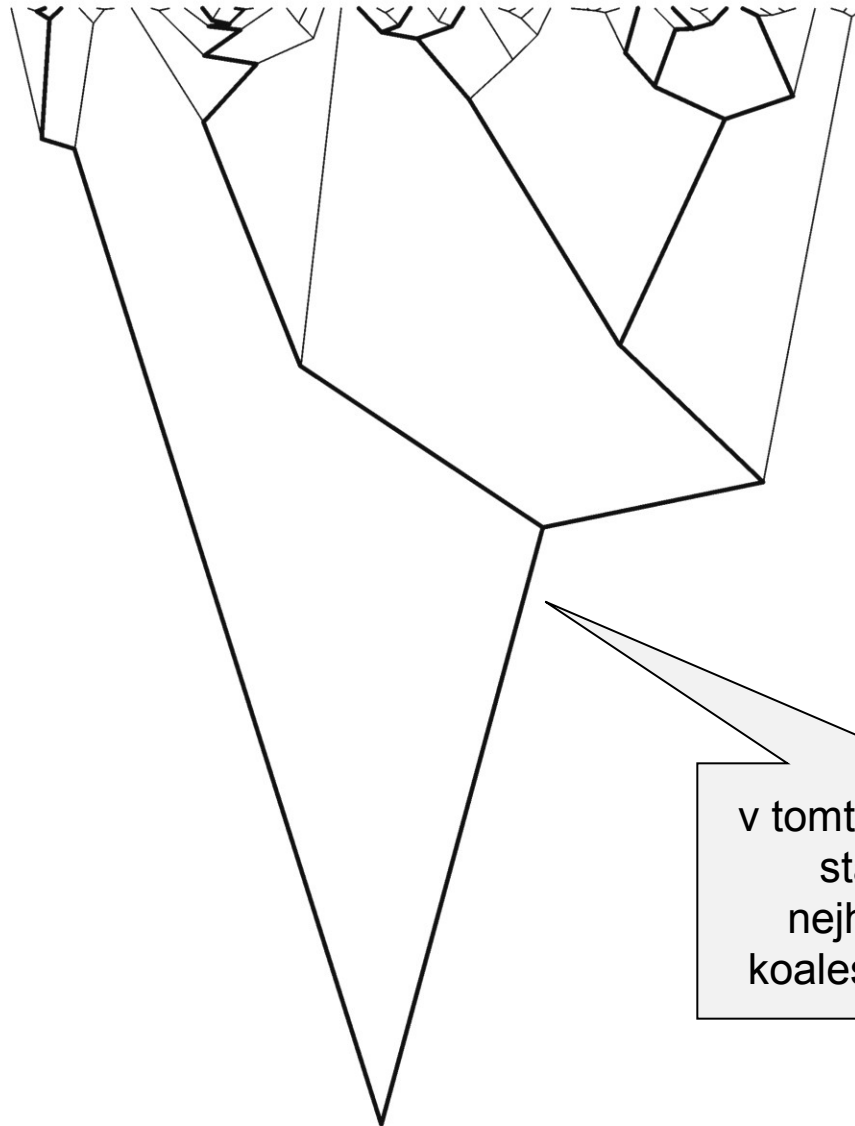
\*) viz počet švábů v krabici



s klesajícím počtem  
volných kopií se proces  
zpomaluje ...

přidání dalších sekvencí  
pravděpodobně nepovede  
k hlubší koalescenci ...

50 genových kopií, 10 náhodně vybraných:



v tomto případě 10 kopií  
stačí k nalezení  
nejhlubšího kořene  
koalescenčního stromu

Pokud nás zajímají „staré“ koalescence, nepotřebujeme velké vzorky

např. pouhé 2 kopie poskytují v průměru 50 % koalescenčního času pro celou populaci!

Naopak pokud nás zajímá čas do první koalescence z  $n$  na  $n - 1$ , odhad  $N_e/[n/(n - 1)]$  je citlivý vůči  $n$

např. rozptyl průměrné doby první a poslední koalescence pro 10 genů je  $0,0444N_e$  až  $3,60N_e$ ; zvýšením  $n$  na 100 genů, rozmezí bude  $0,0004N_e - 3,96N_e$

zvýšením  $n$  10× se  
rozdíl zvýší 100× ...

... pro poslední  
koalescenci prakticky  
žádný rozdíl

**Z toho plyne, že pro odhady starých evolučních genových událostí stačí malé vzorky, pro odhady recentních událostí jsou velké vzorky nezbytné**

Koalescence je ovlivněna různými faktory, např.:

mutací

rekombinací

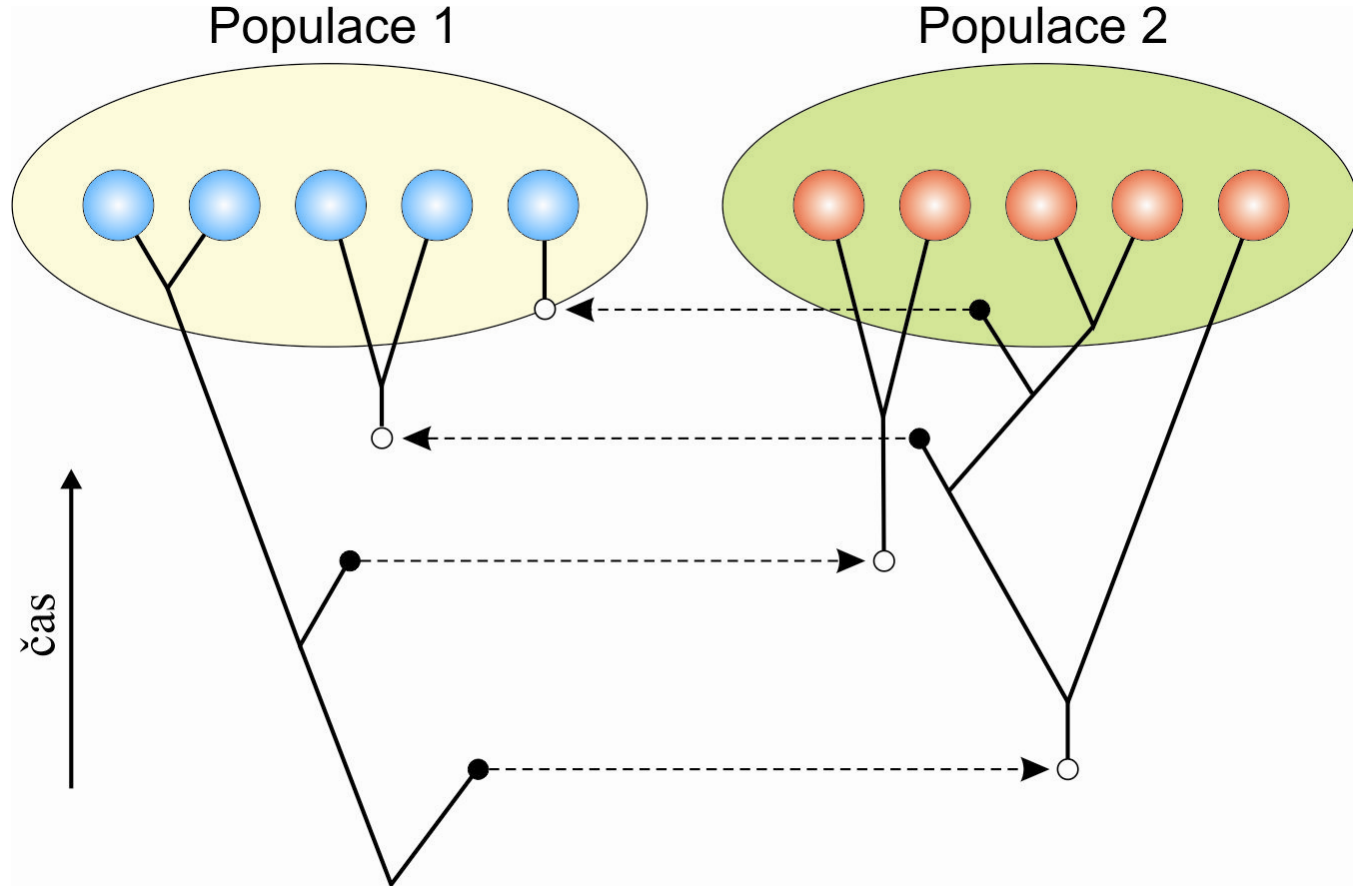
selekcí

změnami velikosti populace

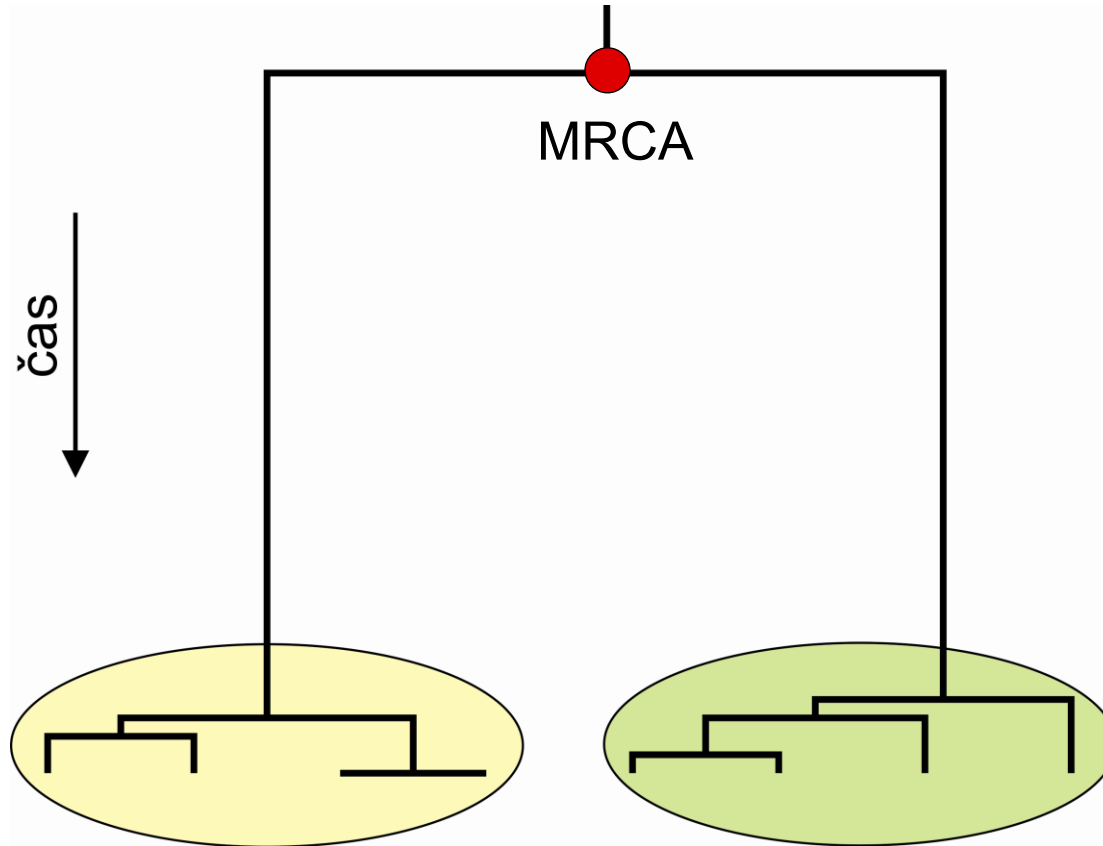
⇒ koalescenční teorii lze použít k odhadu těchto parametrů

Koalescence je ovlivněna různými faktory, např.:

migrací



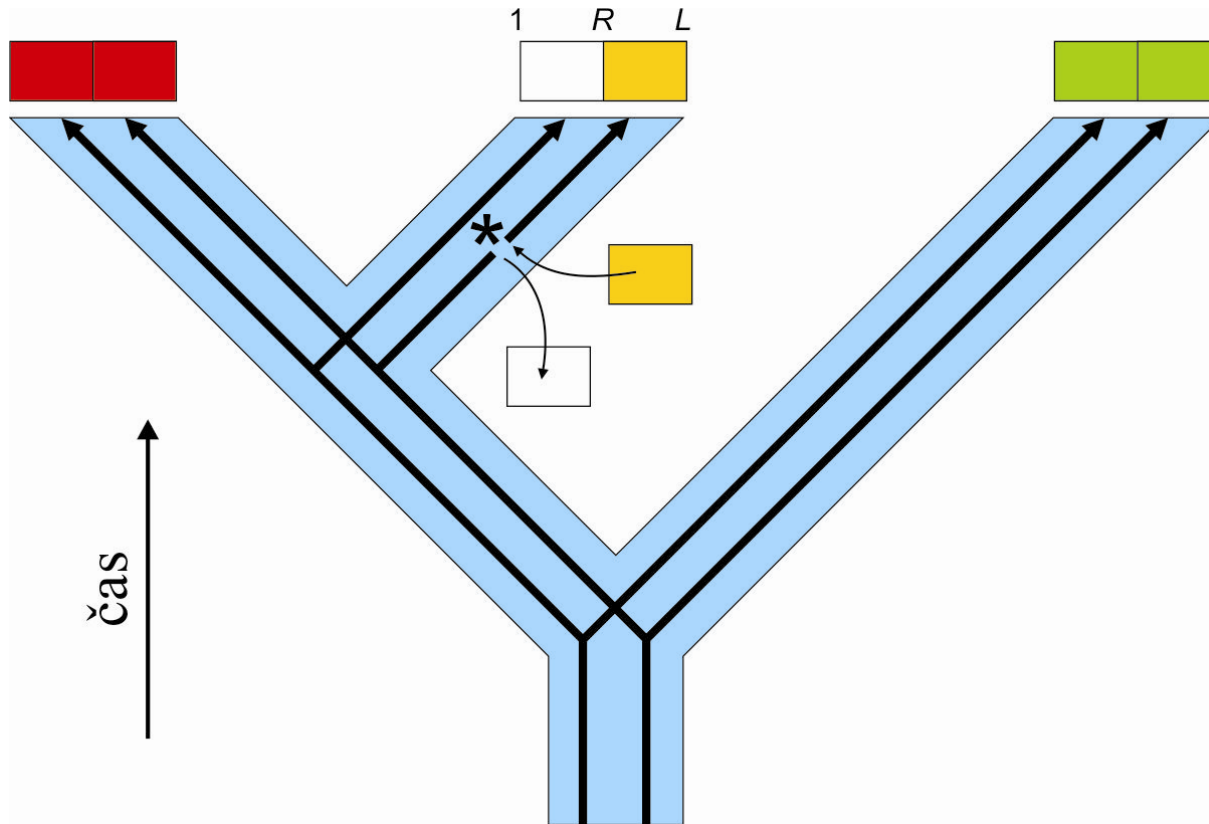
Slabá migrace vede k většině koalescencí uvnitř lokálních populací,.....



.... k prodloužení času k MRCA a zvýšení jeho rozptylu.

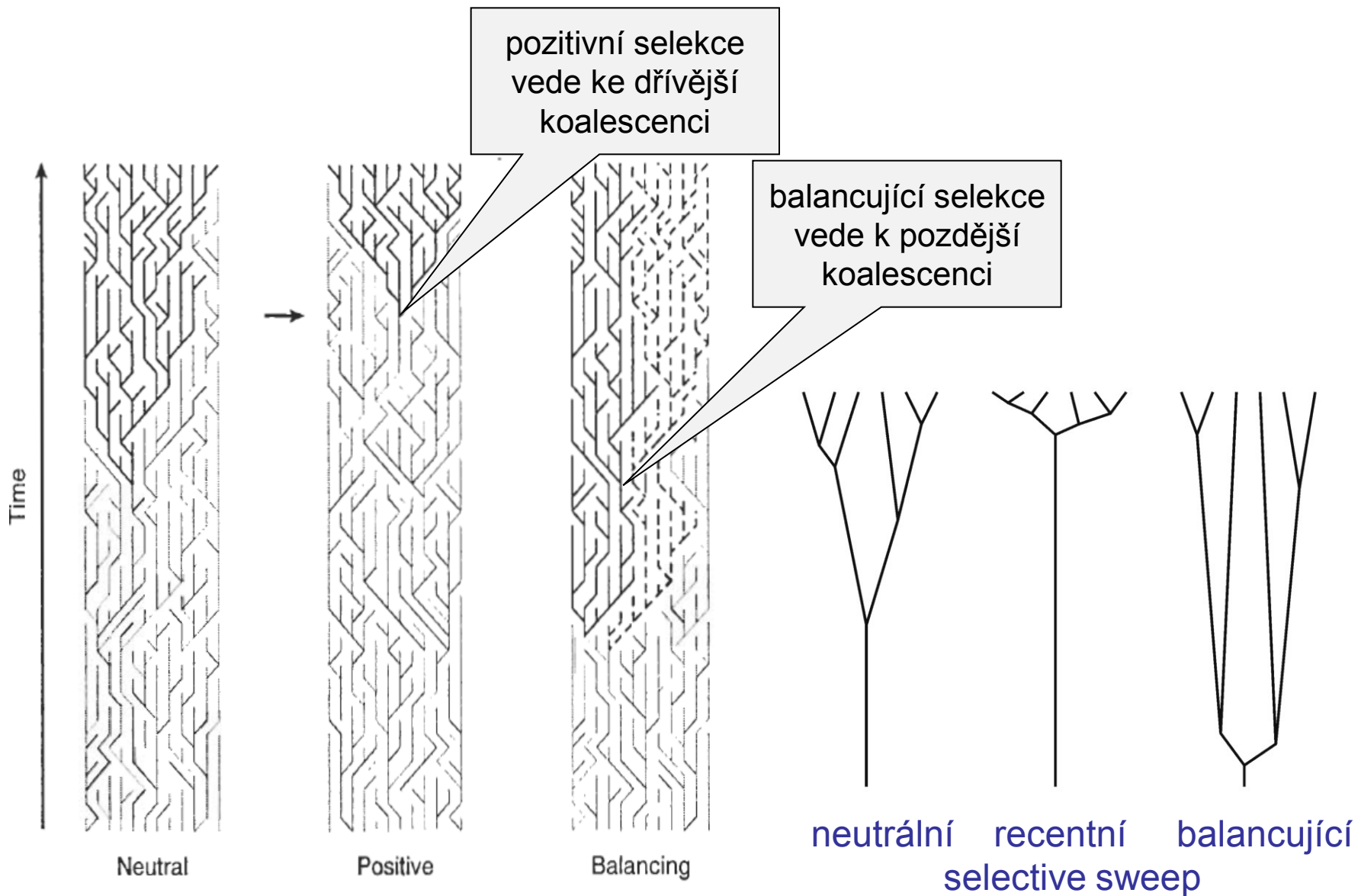
Koalescence je ovlivněna různými faktory, např.:

rekombinací

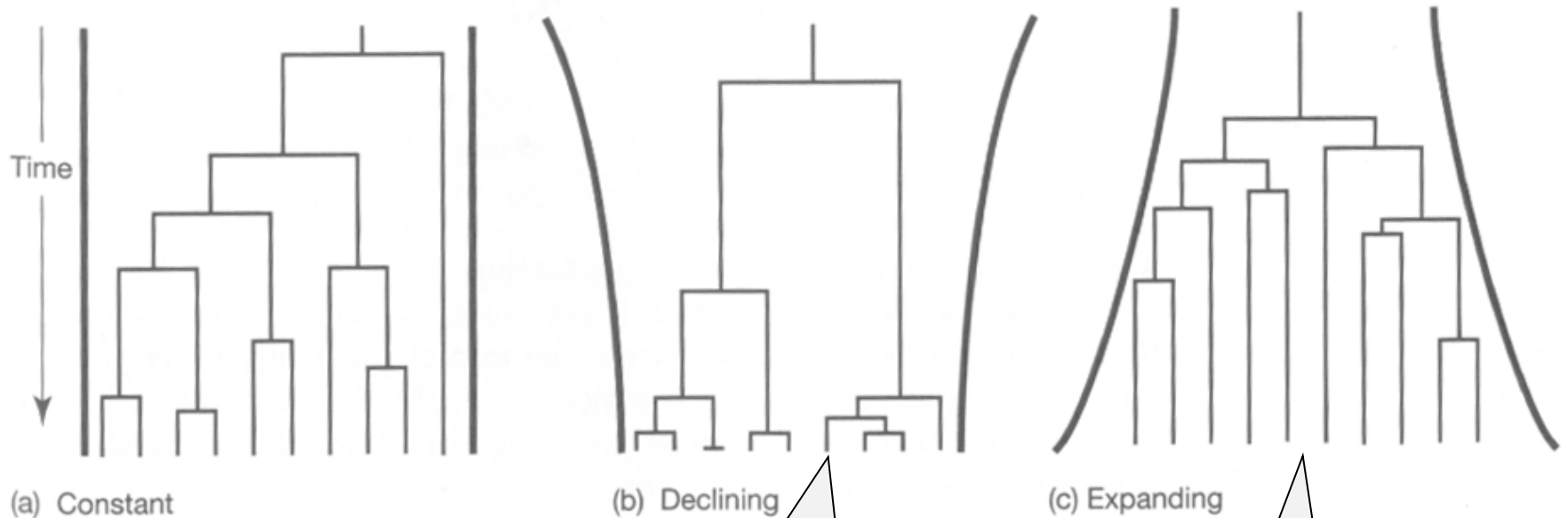




# Vliv selekce na tvar koalescenčního stromu



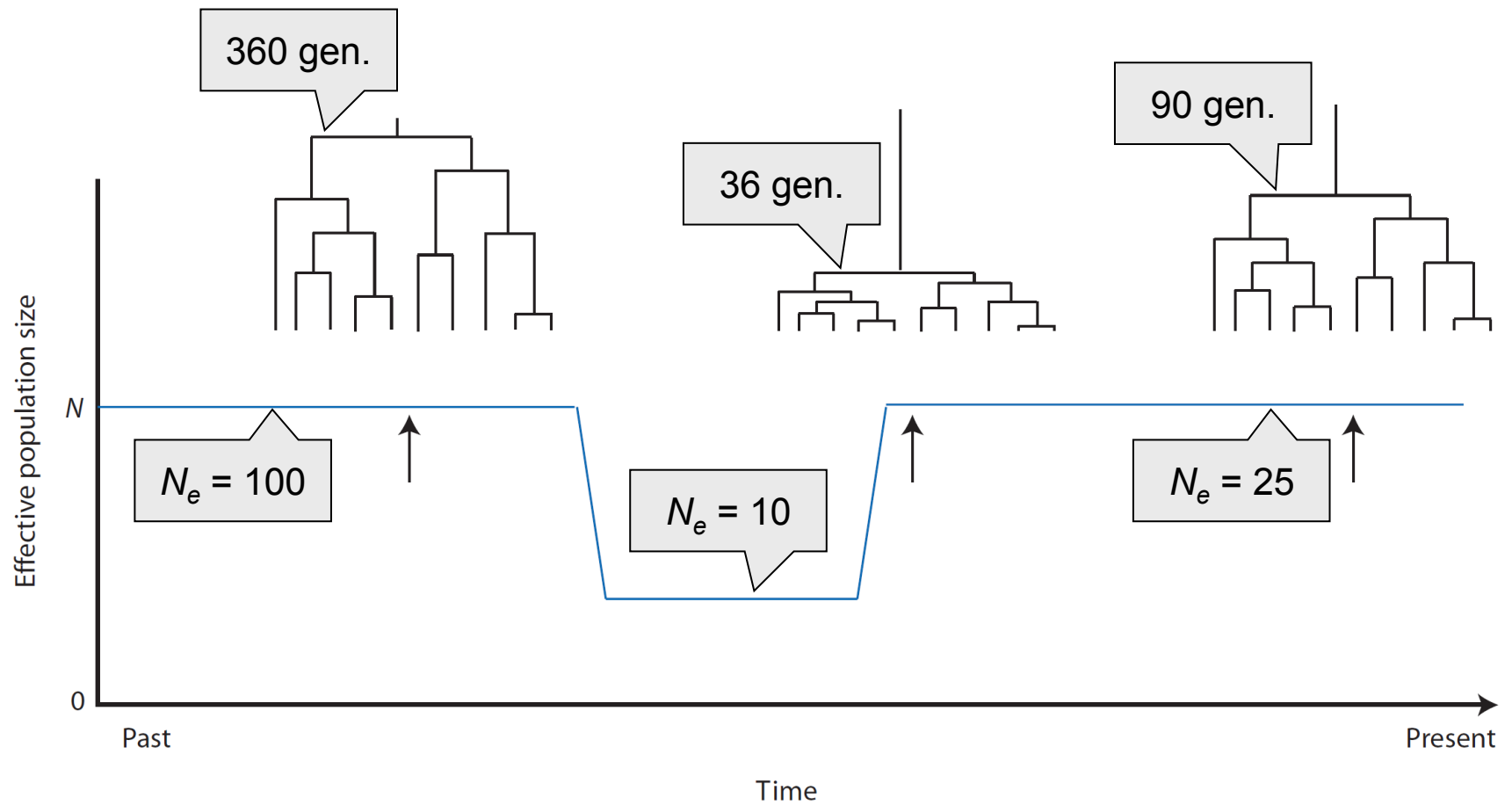
# Vliv změn velikosti populace na tvar koalescenčního stromu



zmenšující se populace:  
koalescence se postupně  
zrychlují

rostoucí populace:  
koalescence se  
postupně zpomalují

$n = 10$



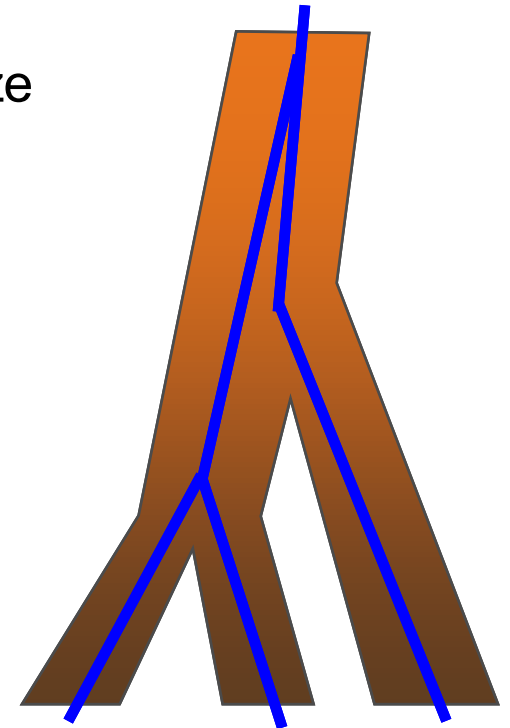
## Genové vs. druhové stromy ještě jednou:

dlouhé intervaly mezi speciálními událostmi → genové a druhové stromy stejné

krátké intervaly mezi speciálními událostmi → genové stromy mohou být jiné než druhové (hemiplazie)

protože odhadujeme divergenci mezi sekvencemi a ne mezi druhy, jsou naše odhady nutně nadhodnocené

nesrovnalosti mezi genovými a druhovými stromy lze minimalizovat použitím markerů s nízkou  $N_e$ , např. mtDNA nebo chromozom Y



# FYLOGEOGRAFIE

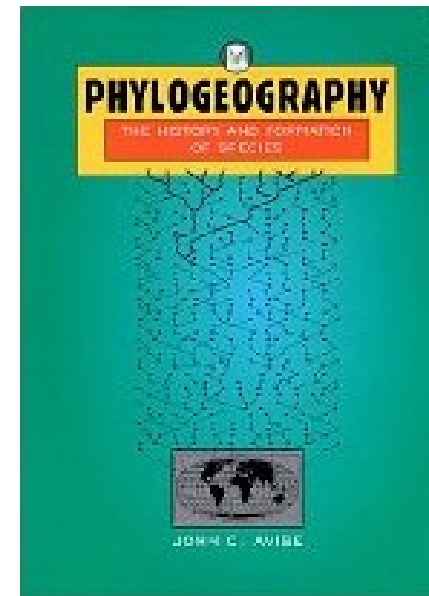
studuje principy a procesy ovlivňující  
geografické rozložení genealogických linií

svým způsobem propojuje mikroevoluční  
procesy (populační genetika)  
s makroevolucí (fylogeneze)

většinou vnitrodruhové studie nebo  
blízce příbuzné druhy



John C. Avise





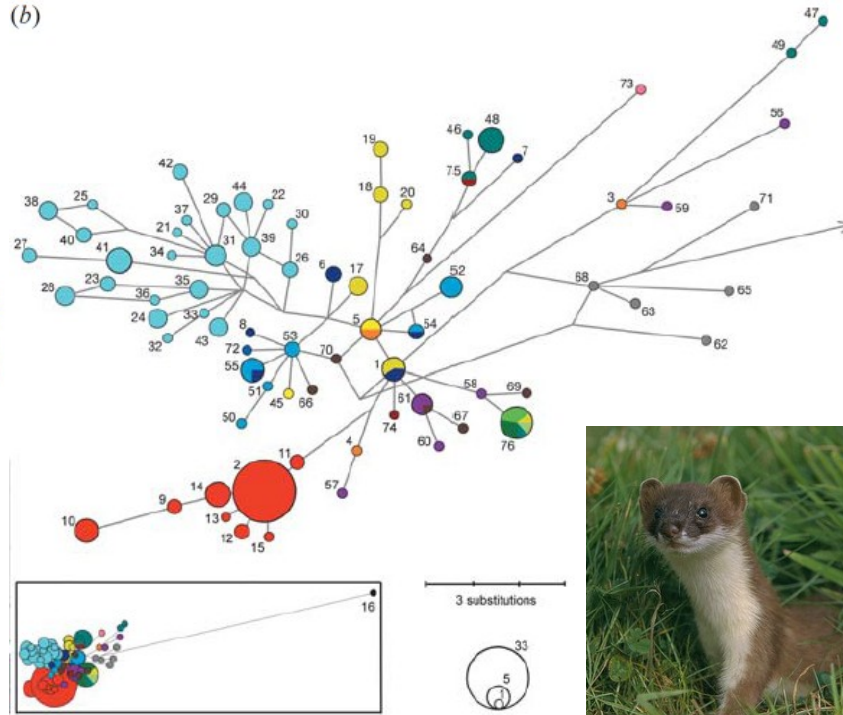
*Mus macedonicus*

Asia

Minimum Spanning Tree (MST)  
 Mimum Spanning Network (MSN)  
 Median-joining network

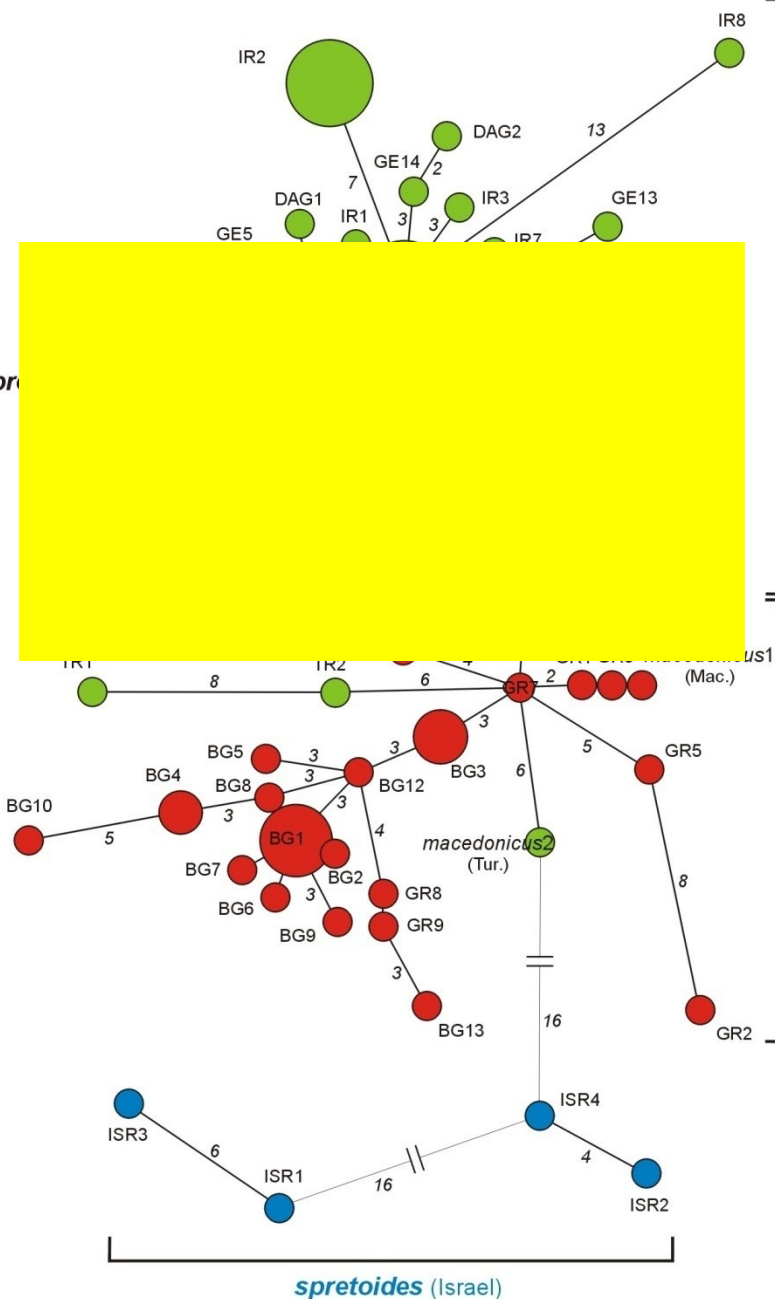
(b)

Europe



*Mustela erminea*

*M. spr*



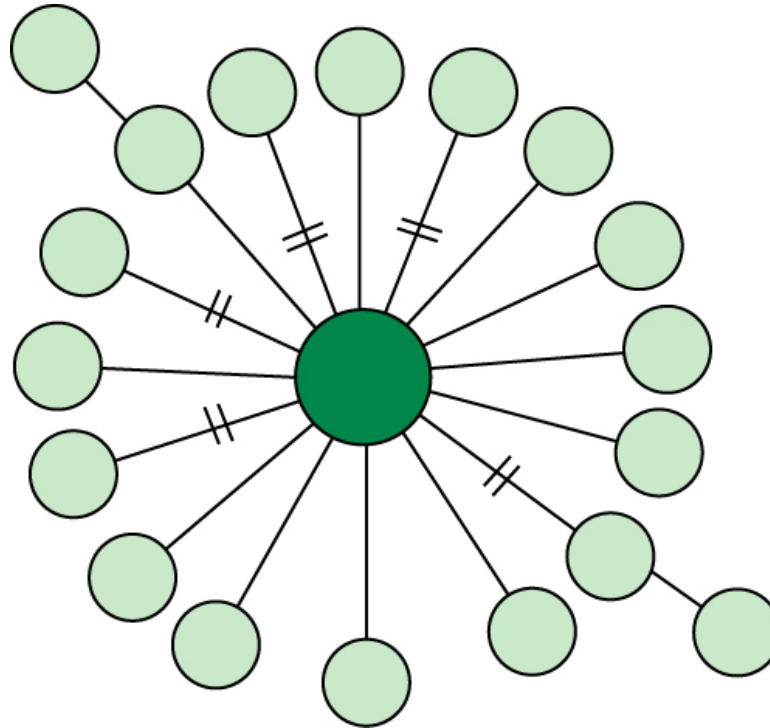
*spretoides* (Israel)

Recentní expanze:

rychlé rozšíření jednoho haplotypu

akumulace malého počtu mutací

hvězdicová struktura



# Změny velikosti populace

Tajimův test (*Tajima's D*)

rozdělení párových rozdílů (*mismatch distribution*)

koalescence, ML nebo BA, MCMC

*Bayesian Skyline Plot*



# 1. Tajimův test (*Tajima's D*)

založený na porovnání haplotypové diverzity a nukleotidové diverzity

primárně jde o test selektivní neutrality, ale může indikovat i růst populace nebo bottleneck

silně záporné hodnoty indikují populační expanzi – převládá „mladý“ polymorfismus, kdy se vytvořily nové haplotypy, ale nukleotidová diverzita je stále nízká

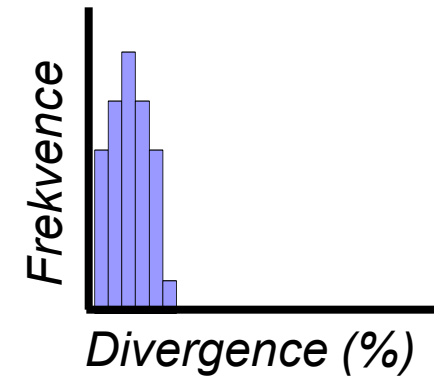
programy Arlequin, DnaSP

podobně Fu's test a další

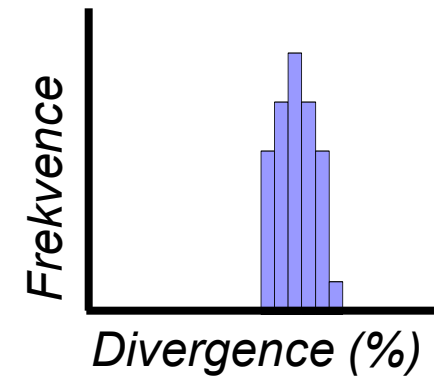
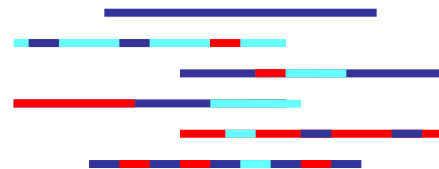
## 2. Rozložení párových neshod (*mismatch distribution*)

párová srovnání všech sekvencí → histogram

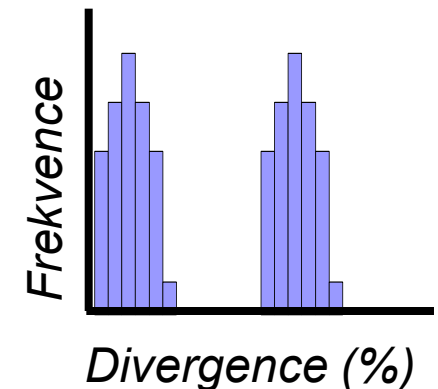
Sekvence navzájem velmi podobné



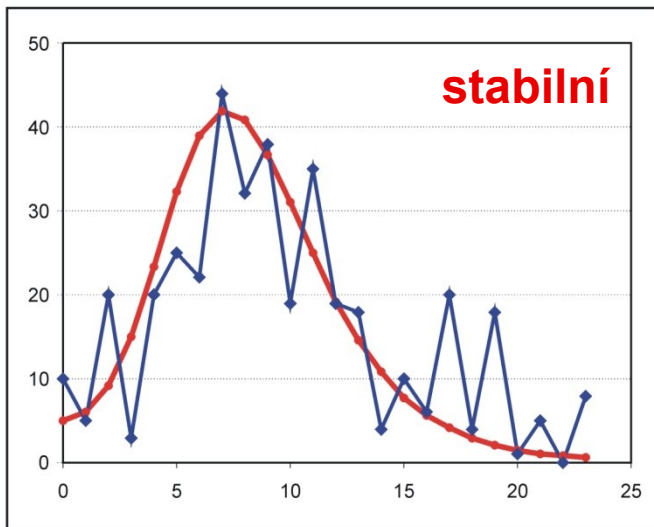
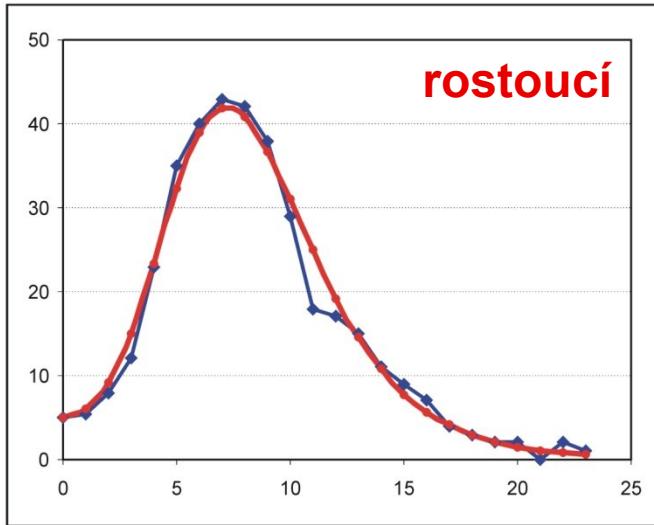
Sekvence navzájem velmi odlišné



Směs podobných a odlišných sekvencí



frekvence



párové rozdíly

test shody skutečného rozdělení  
s teoretickou predikcí:

*Harpending's raggedness index*  
(Harpending 1994)

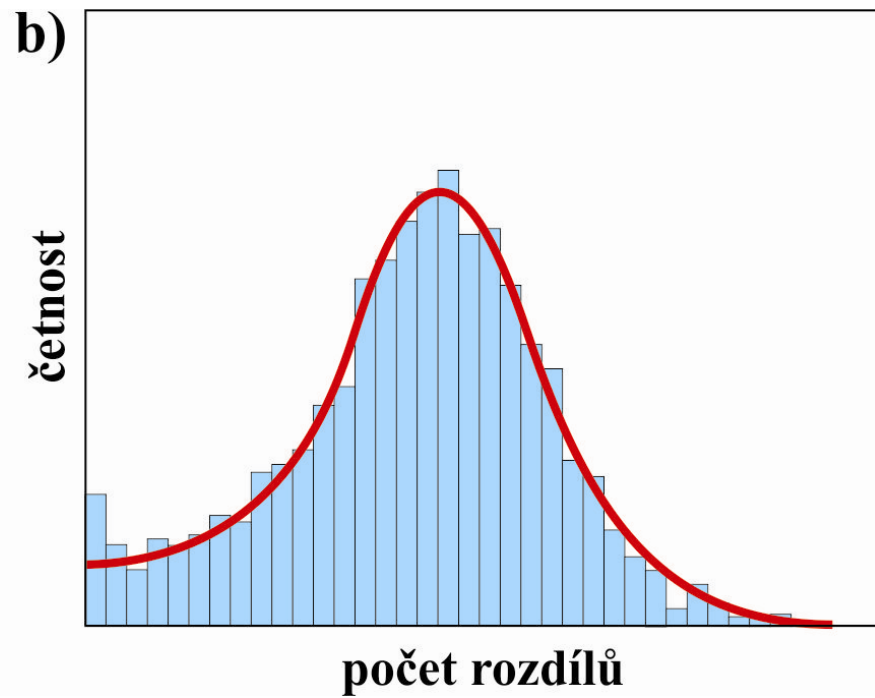
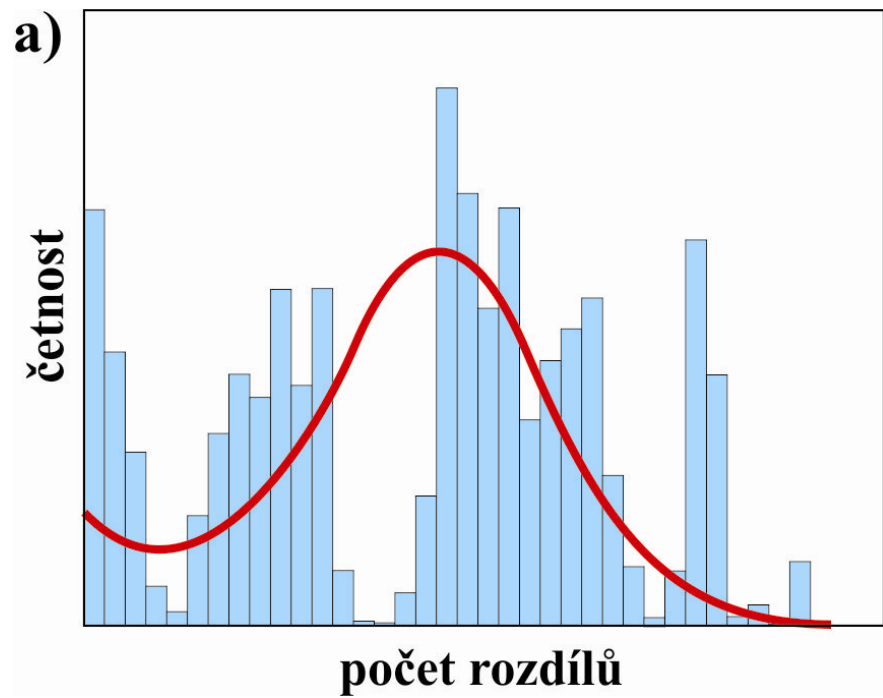
*sum of squared deviations*

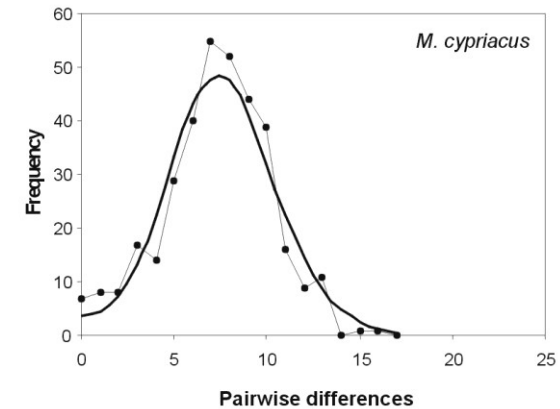
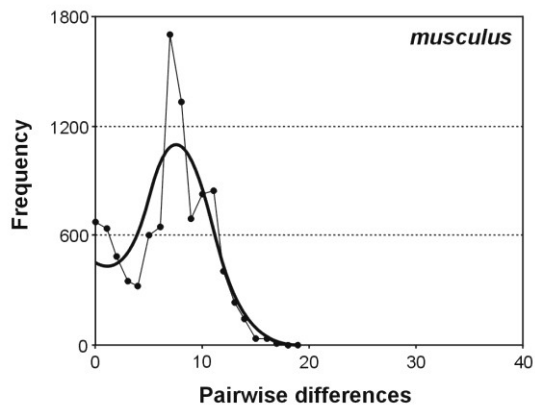
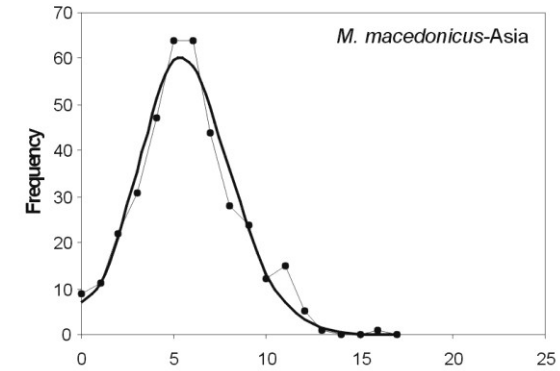
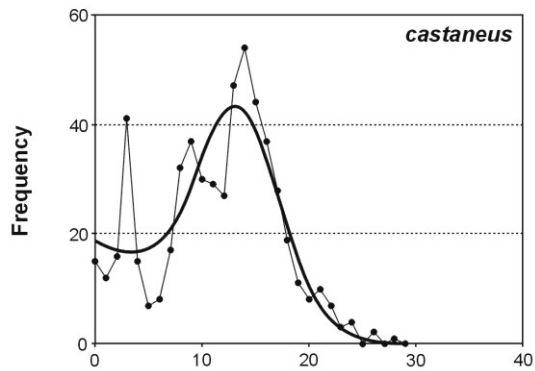
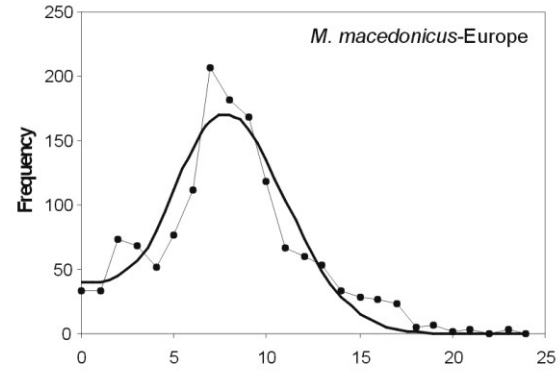
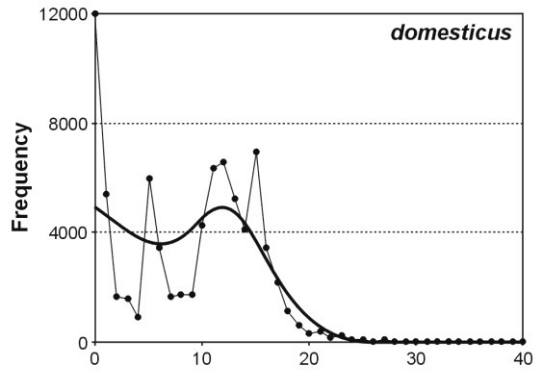
doba expanze/bottlenecku:

$$\tau = 1/2u,$$

kde  $u$  je mutační frekvence pro  
celou sekvenci

můžeme odhadnout i velikost  
populace před a po expanzi





### 3. ML a bayesiánské odhady

MCMC

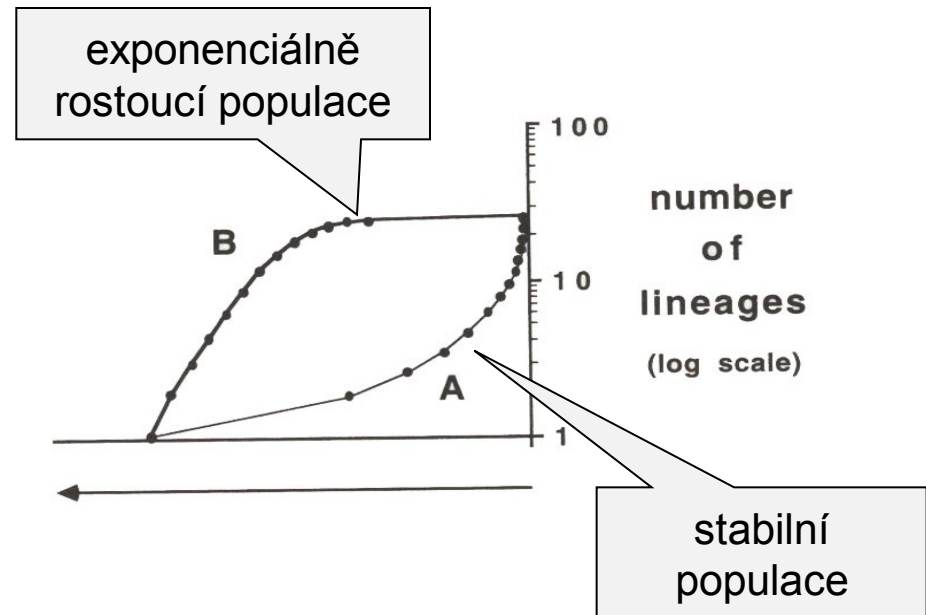
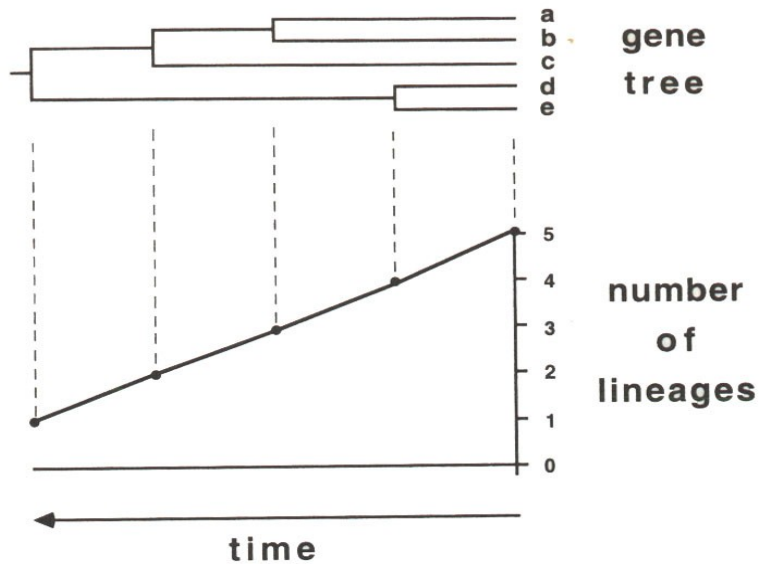
srovnání modelu stabilní velikosti a modelu exponenciálního  
růstu/redukce populace pomocí LRT s 1 stupněm volnosti

program Fluctuate:

parametr růstu  $g$   
ML i BA přístup



## 4. Bayesovský panoramatický graf (*Bayesian Skyline Plot – BSP*)

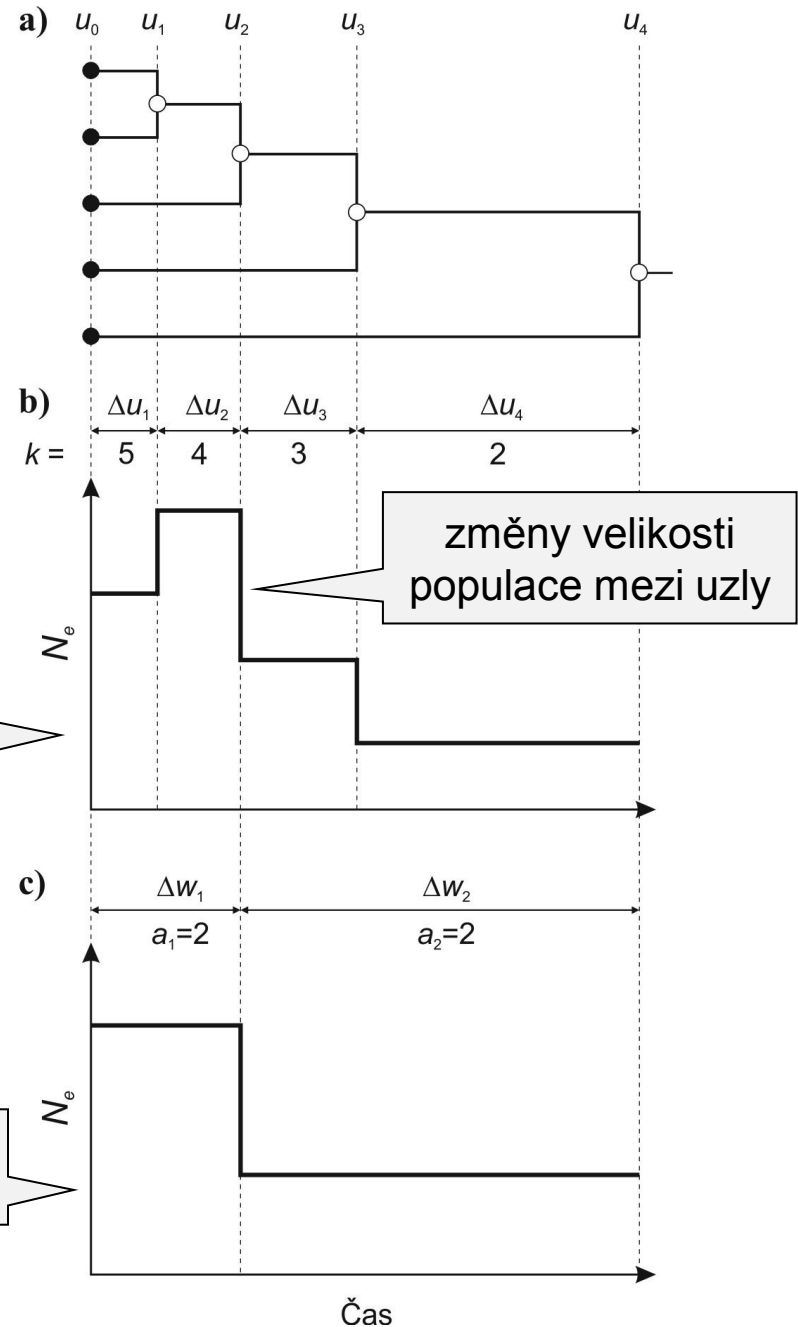


# Bayesovský panoramatický graf

rozložení genealogických linií v čase

BSP vychází z tohoto postupu

programy BEAST a Tracer

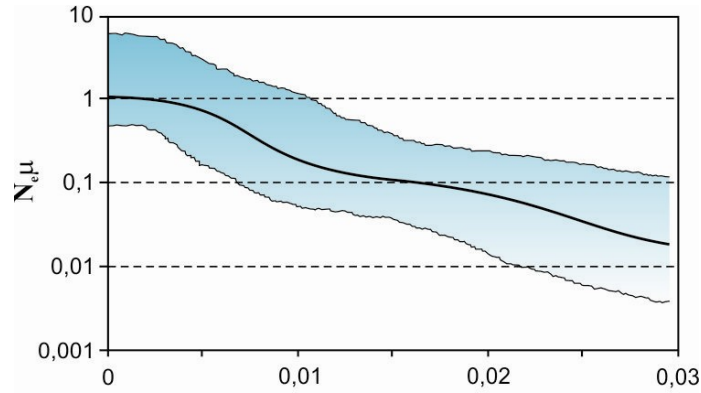


klasický

generalizovaný

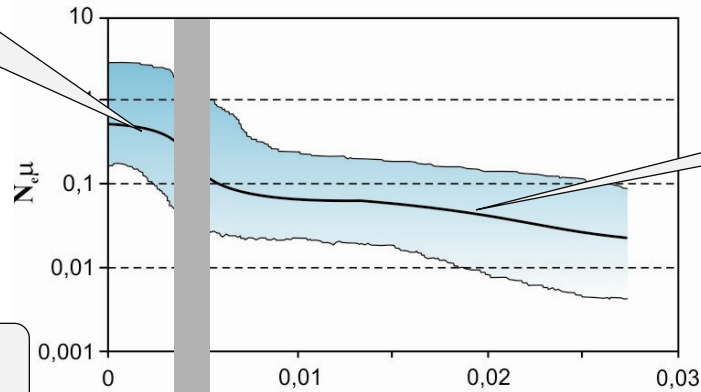


# Myší kolonizace Evropy



*domesticus*

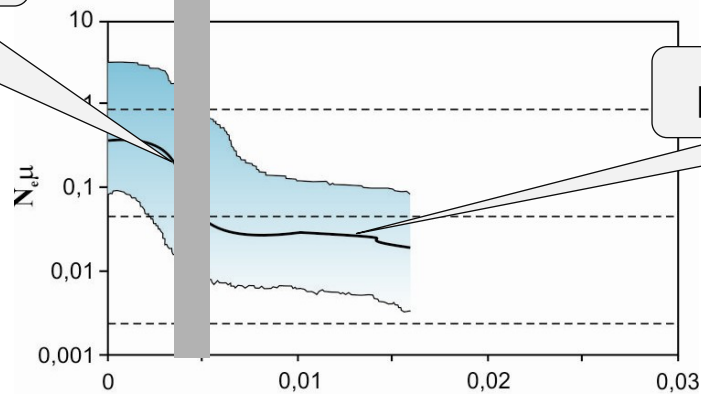
expanze do Evropy



původ mimo Evropu

*domesticus* - Evropa

expanze do Evropy

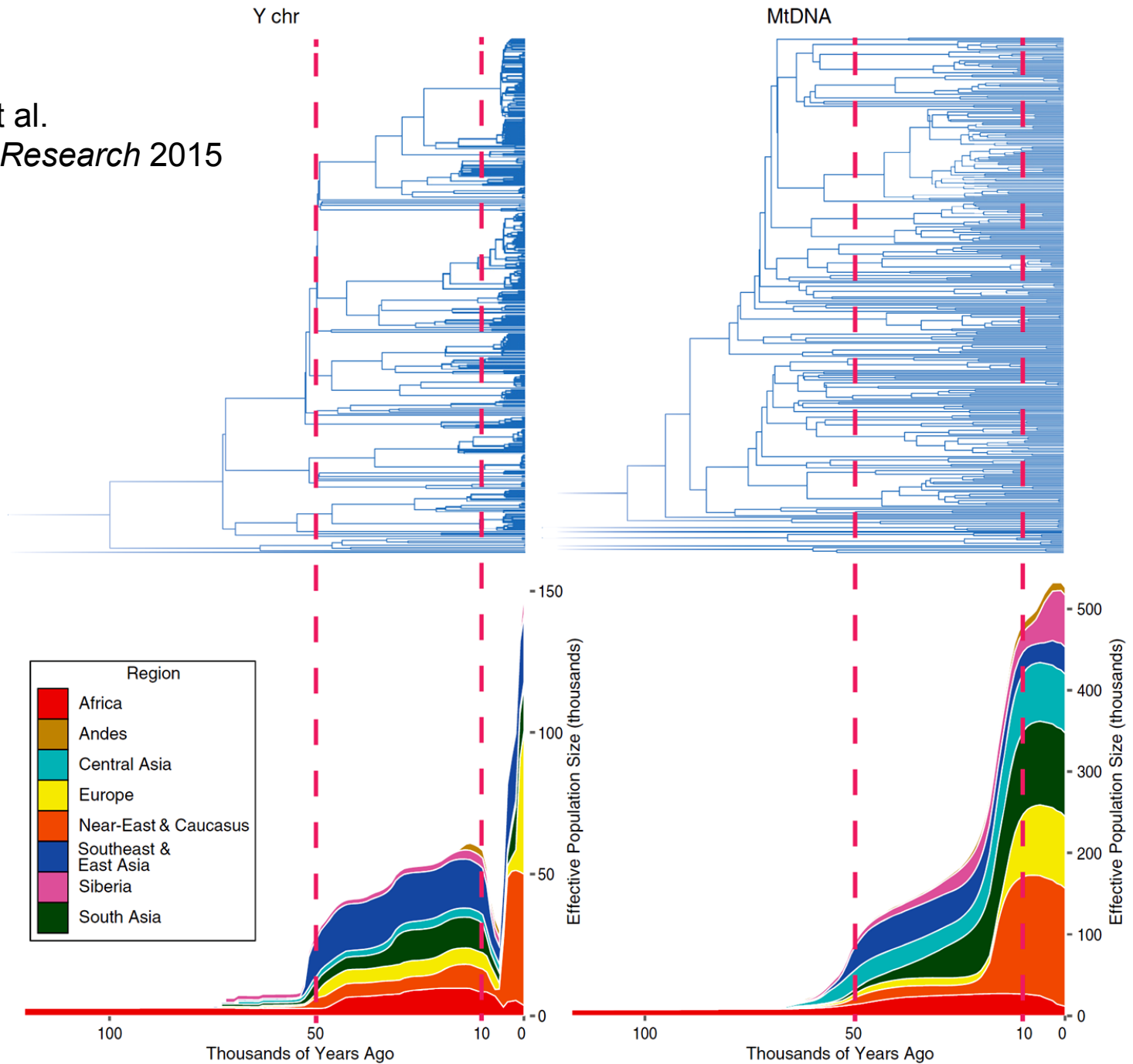


původ mimo Evropu

*musculus* - Evropa

Čas (mil. let)

Karmin et al.  
*Genome Research* 2015



# Možné výsledky fylogeografických studií

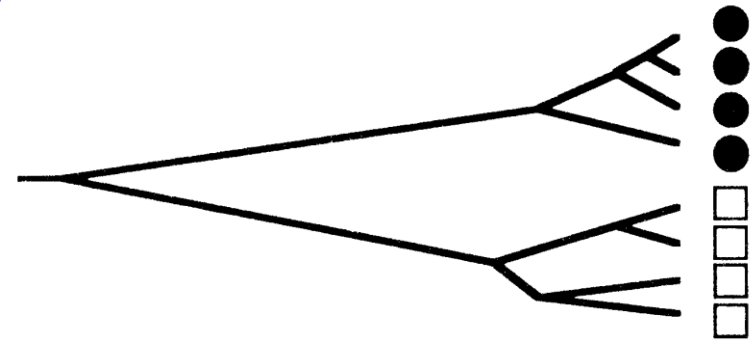
(Avice 2000)

## Kategorie I:

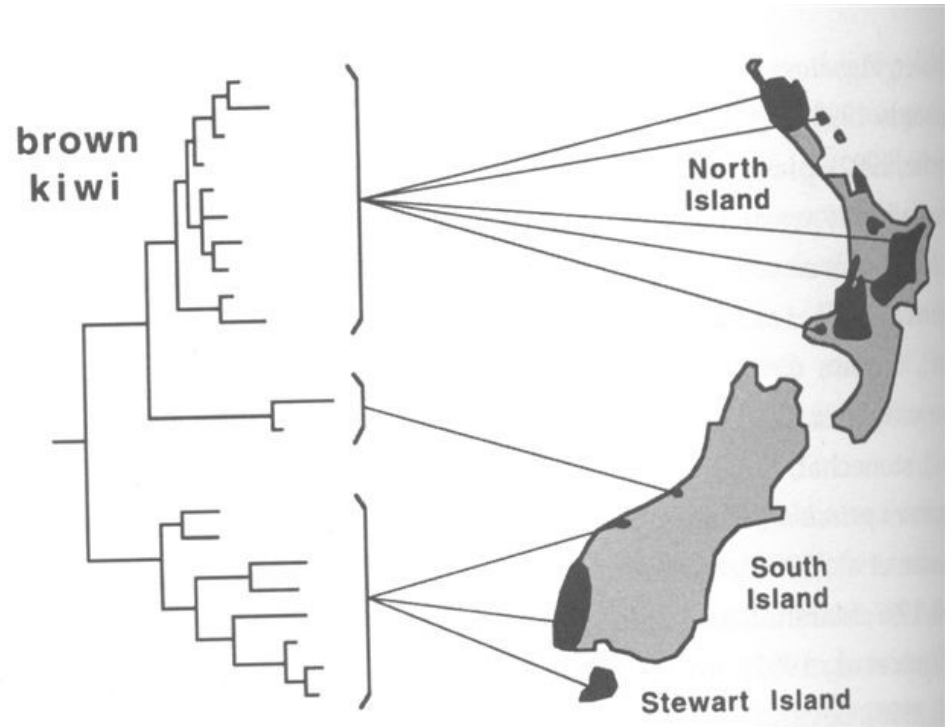
zřetelně oddělené alopatické linie

bariéry toku genů nebo nízká disperze

rozdíly díky sortování linií nebo akumulaci nových mutací

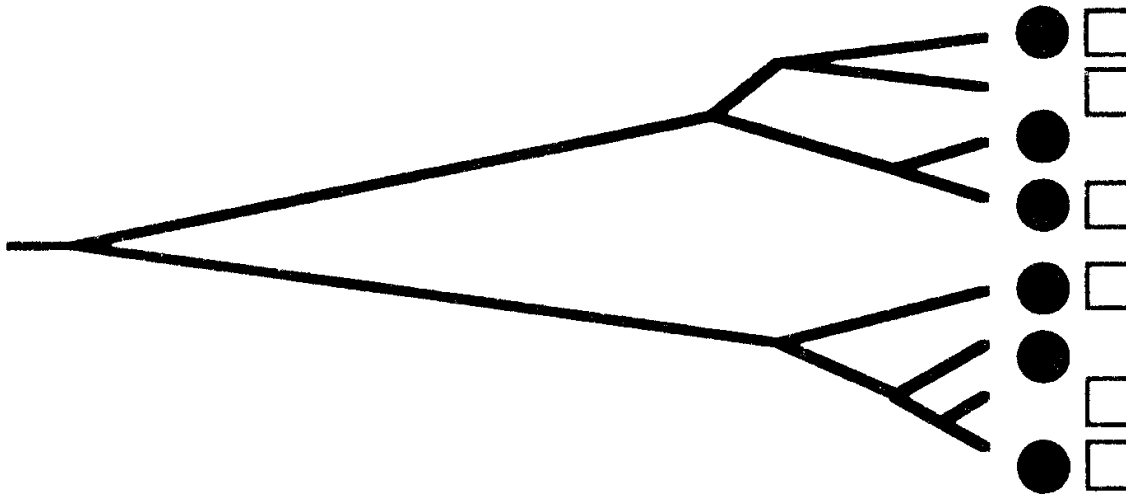


*Apteryx australis*



## Kategorie II:

sympatrické, ale jasně hluboce oddělené linie  $\Rightarrow$  sekundární kontakt  
dříve oddělených populací



## Kategorie III:

alopatrické, málo oddělené linie

blízce příbuzné ale geograficky lokalizované haplotypy

populace v nedávném historickém kontaktu

ale: tok genů dostatečně nízký

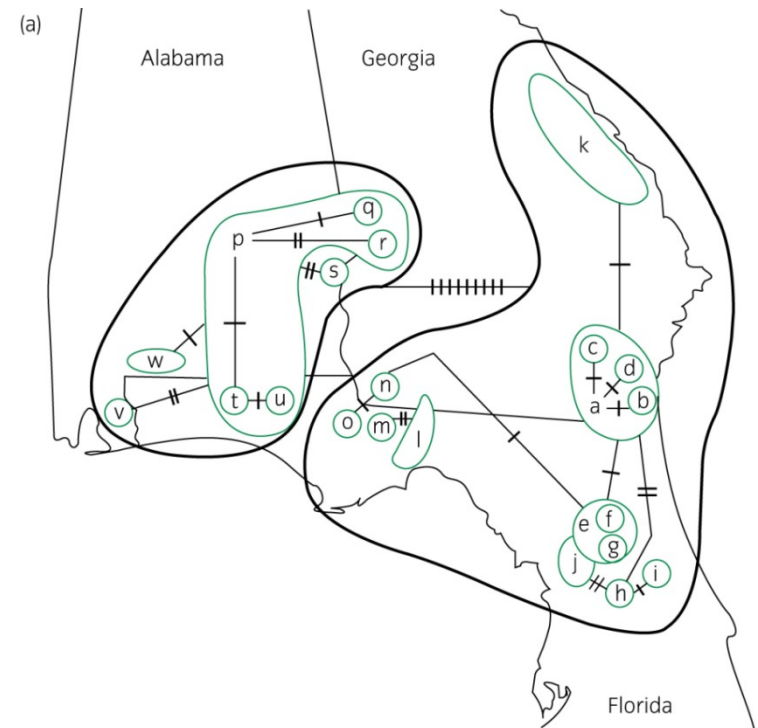
→ drift a „lineage sorting“ → divergence populací

často:

na větší škále Kategorie I

na menší škále Kategorie III

př.: *Geomys pinetis*



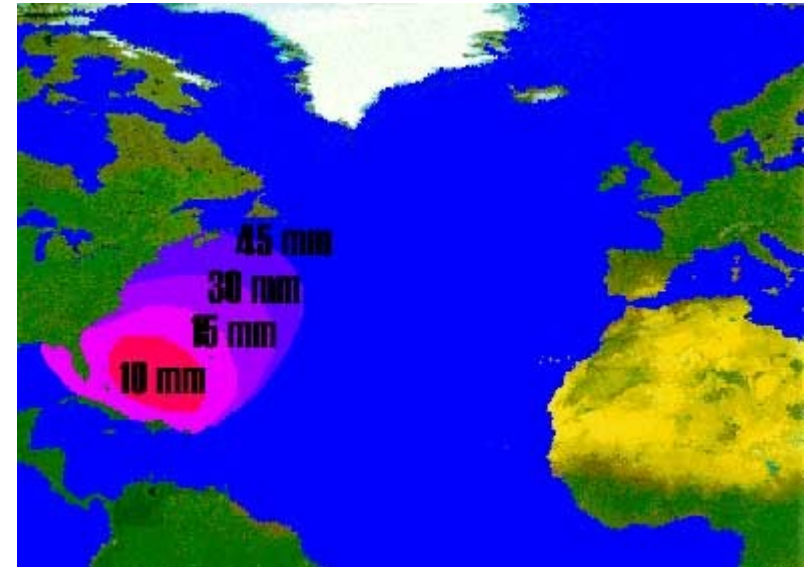
## Kategorie IV:

sympatrické, málo oddělené linie

silný tok genů

absence geografických bariér nebo

recentní expanze



*Anguilla rostrata*

Náhodná disperze larev

Panmiktické agregace  
během tření

## Kategorie V:

kombinace III a IV

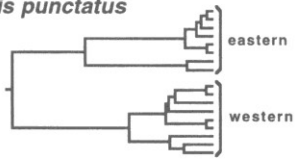
málo oddělené linie

některé linie značně rozšířeny (zřejmě ancestrální), jiné (nové)  
geograficky omezeny

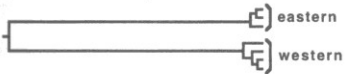
jako vhodné znaky je nutné vybrat privátní haplotypy

# Genealogické konkordance Ryby JV USA

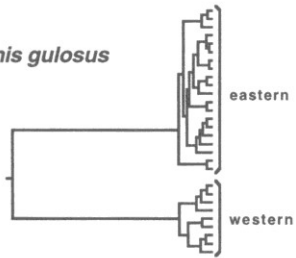
*Lepomis punctatus*



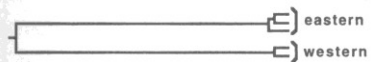
*Lepomis microlophus*



*Lepomis gulosus*



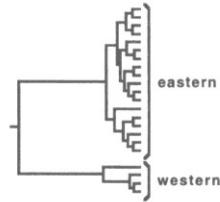
*Lepomis macrochirus*



8.0 6.0 4.0 2.0 0.0

sequence divergence (%)

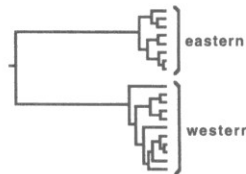
*Gambusia affinis/ G. holbrooki*



*Amia calva*



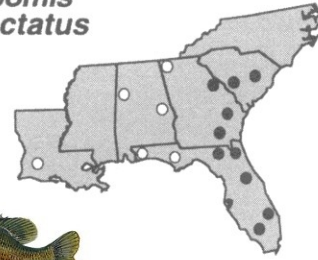
*Micropterus salmoides*



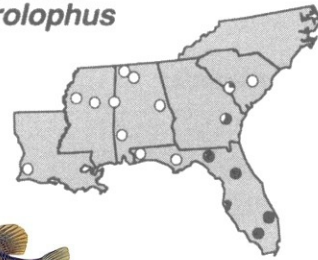
8.0 6.0 4.0 2.0 0.0

sequence divergence (%)

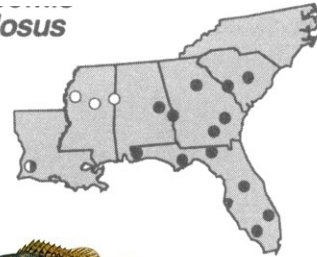
*Lepomis punctatus*



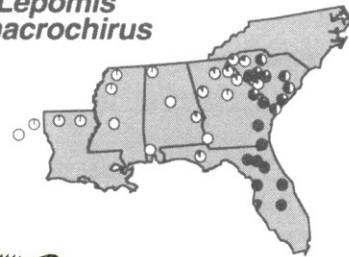
*Lepomis microlophus*



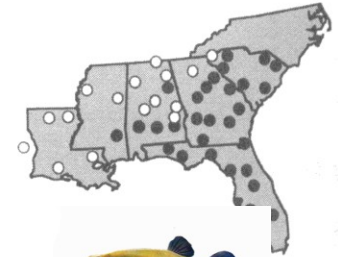
*gulosus*



*Lepomis macrochirus*



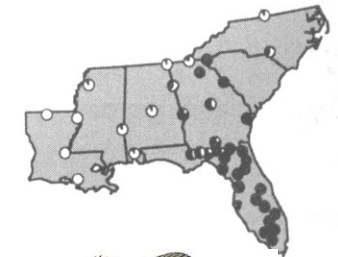
*Gambusia affinis, G. holbrooki*



*Amia calva*



*Micropterus salmoides*



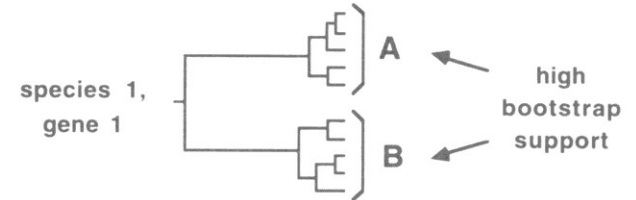


# Genealogické konkordance (shody na různých úrovních)

Různé části sekvence genu



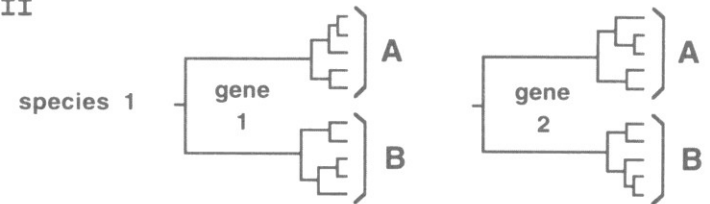
Aspect I



Více sekvencí (genů) od stejného druhu



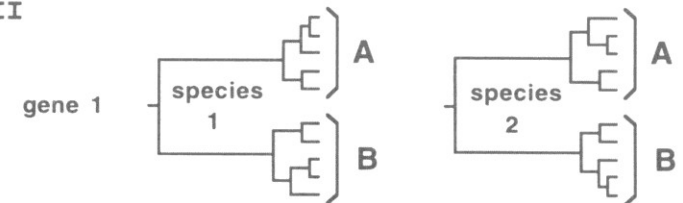
Aspect II



Více druhů ve stejné oblasti



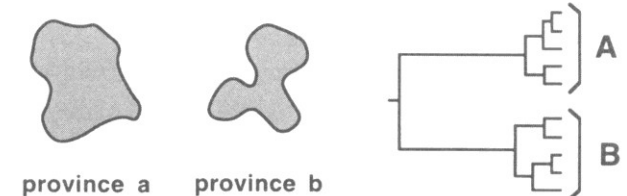
Aspect III



Podpora biogeografických oblastí  
(více druhů, více oblastí)



Aspect IV



# Genetické důsledky ledových dob

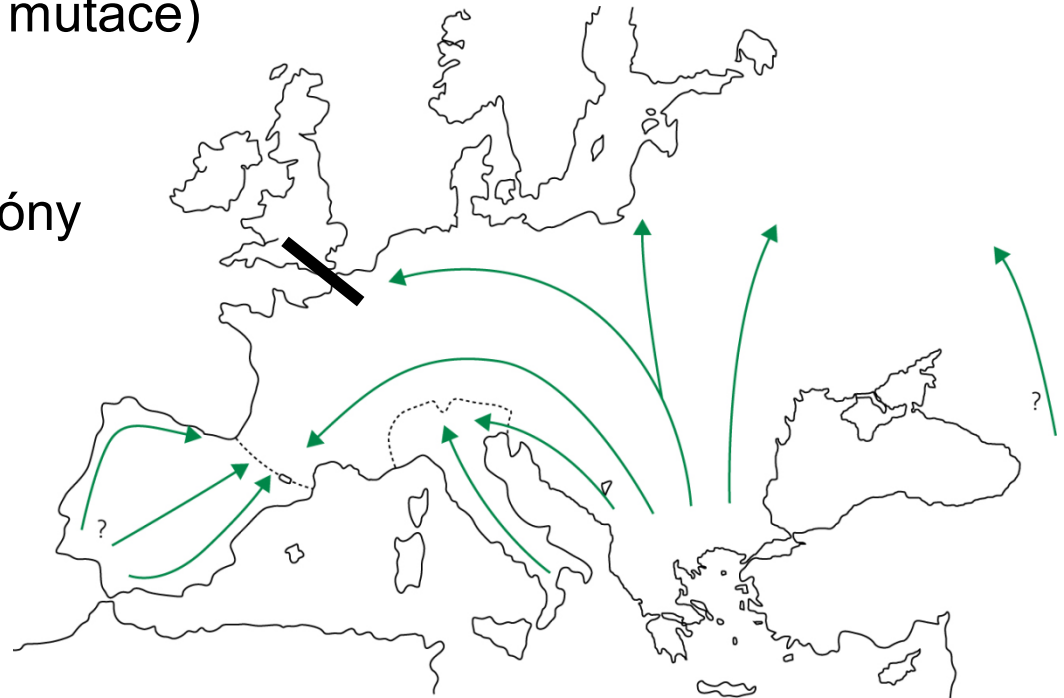
Refugia (iberské, apeninské, balkánské)

V refugiích malé populace relativně dlouhou dobu

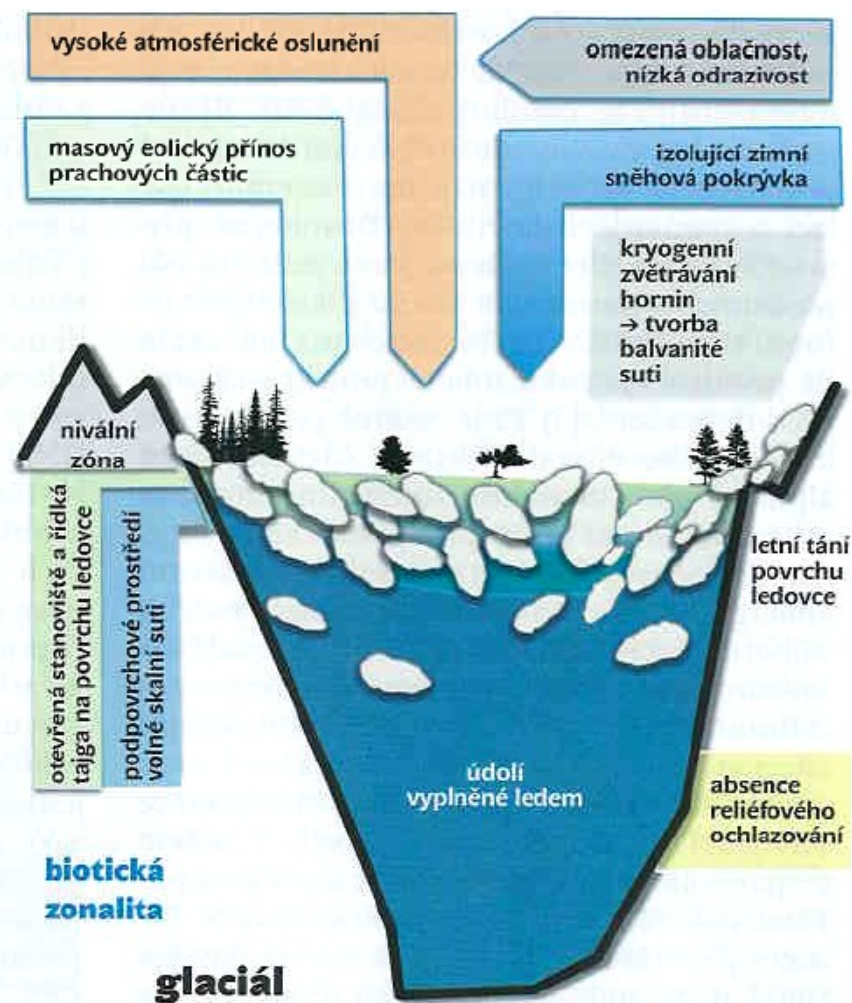
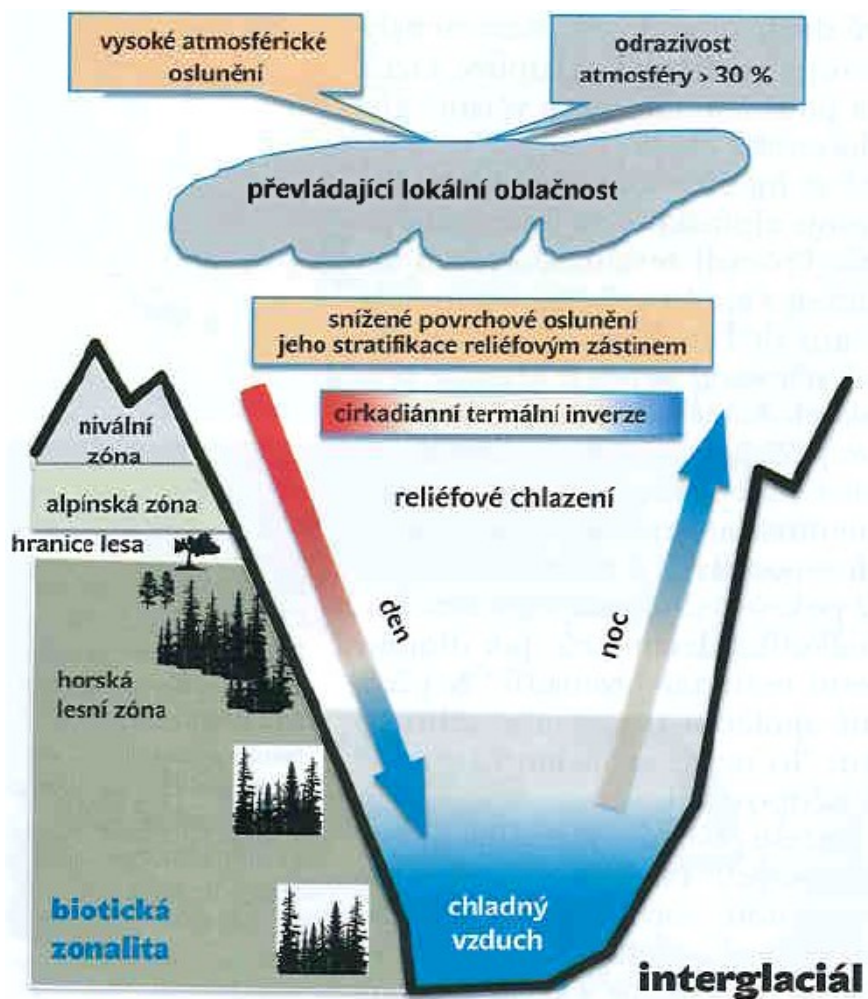
*Linneage sorting* (případně mutace)

Následné šíření →  
vnitrodruhové hybridní zóny

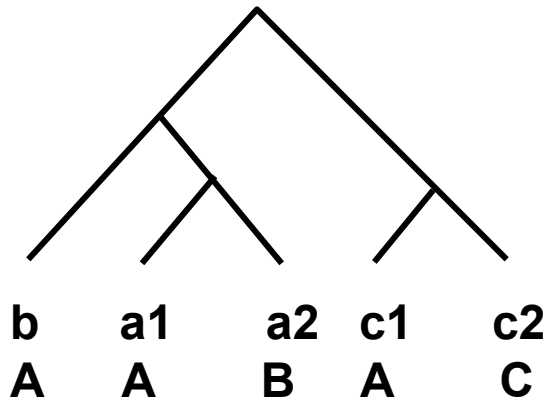
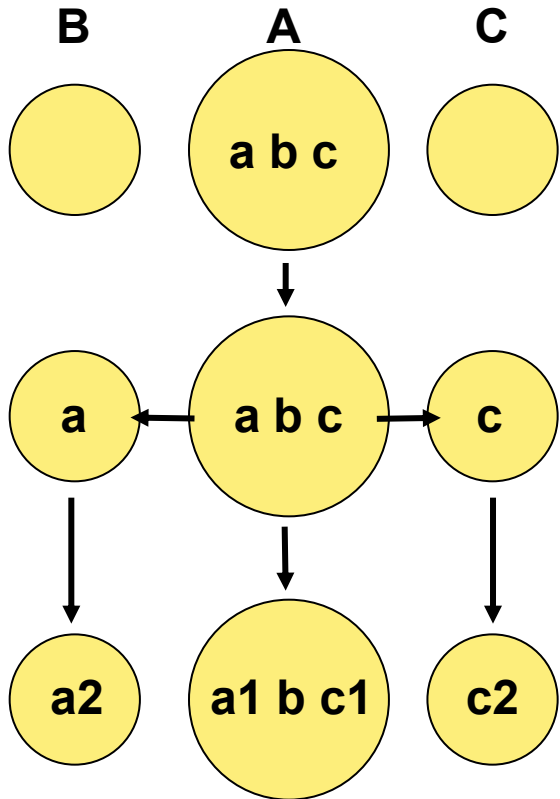
Ale některé druhy měly  
i severská refugia!



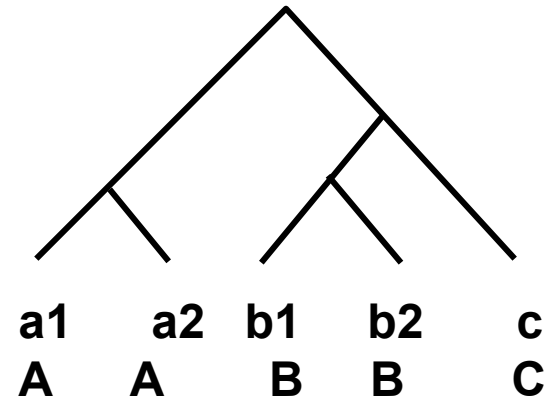
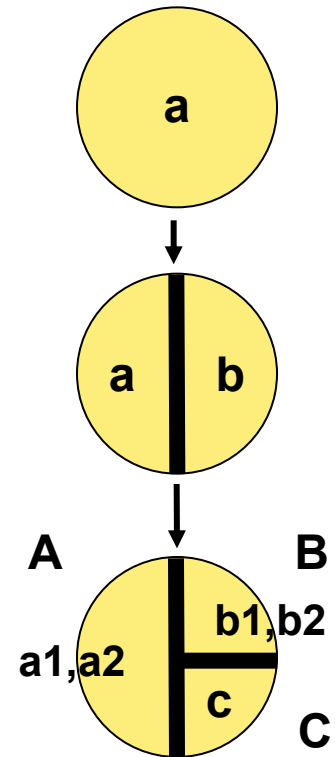
*Chorthippus parallelus*



## disperze



## vikariance



# Vztah mezi genetickou strukturou populace, pohlavně-specifickou disperzí a režimy toku genů (Avice 2000)

samičí disperze a tok genů

nízké  $\longrightarrow$  vysoké

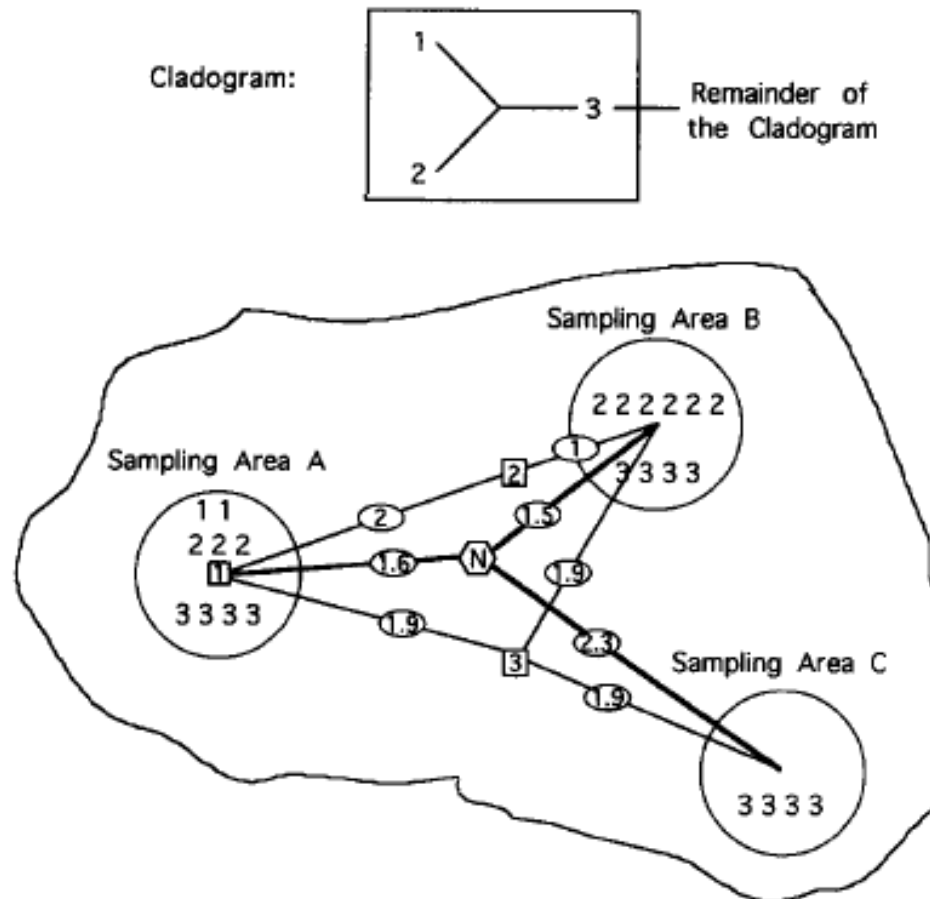
<p style="writing-mode: vertical-rl; transform: rotate(180deg);">samičí disperze a tok genů</p> <p style="text-align: center;">nízké <math>\longleftarrow</math></p> <p style="text-align: center;">vysoké <math>\longleftarrow</math></p>	<p>geografická struktura v:</p> <p>mtDNA ANO</p> <p>autozomy ano</p> <p>chr. Y ano</p>	<p>geografická struktura v:</p> <p>mtDNA NE</p> <p>autozomy ano</p> <p>chr. Y ***</p>
	<p>geografická struktura v:</p> <p>mtDNA (u samic) ANO</p> <p>autozomy ne</p> <p>chr. Y ne</p>	<p>geografická struktura v:</p> <p>mtDNA NE</p> <p>autozomy ne</p> <p>chr. Y ne</p>

# Nested Clade Analysis (NCA)

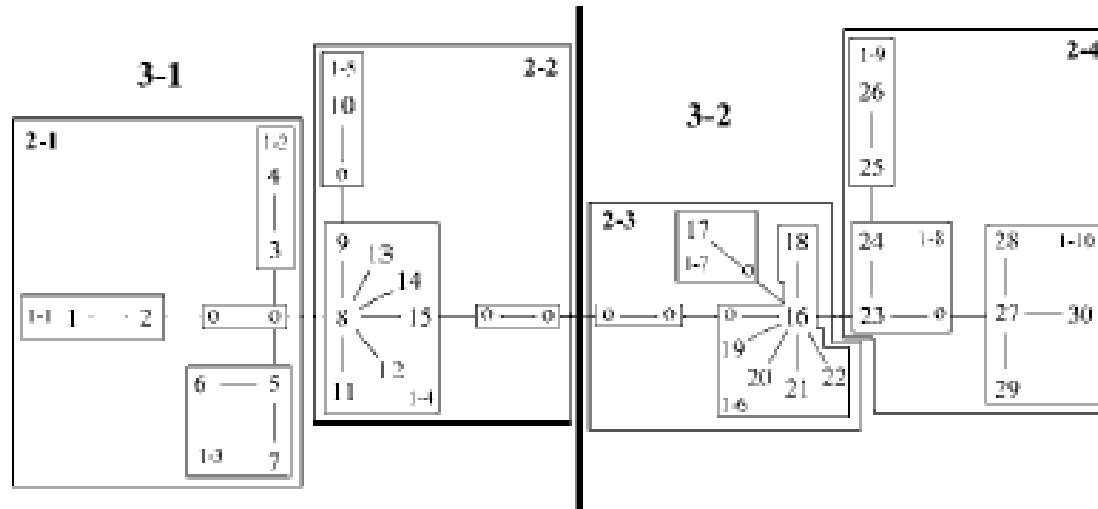
Alan R. Templeton

clade distance ( $D_c$ ): měří, jak daleko se klad rozšířil

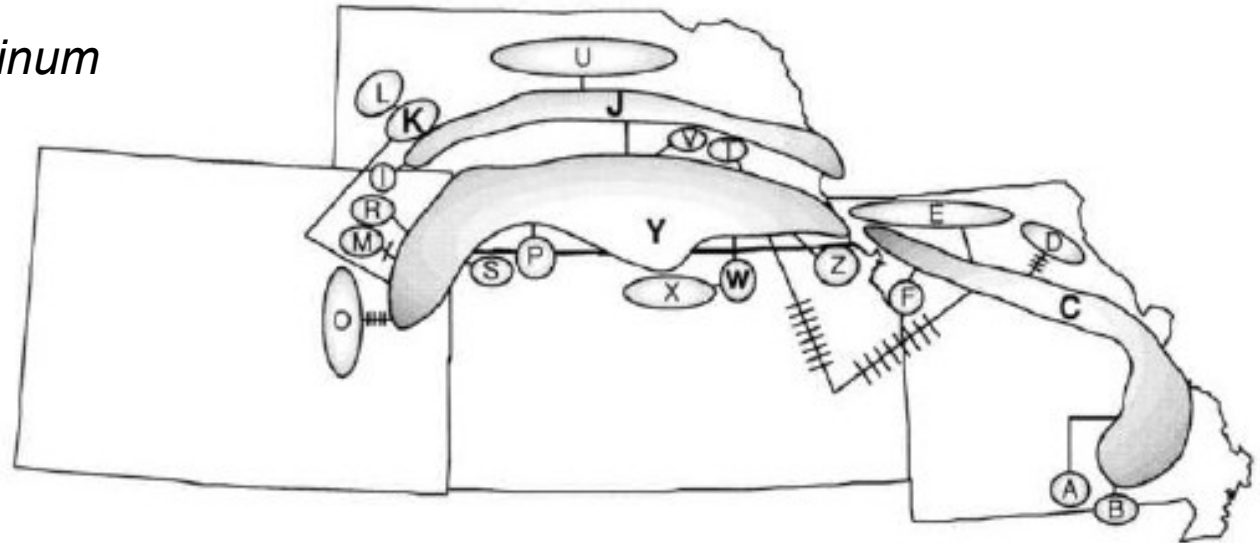
nested clade dist. ( $D_n$ ): měří, do jaké míry změnil klad svoji polohu ve vztahu ke kladu, ze kterého vzniknul



# Nested Clade Analysis (NCA)



*Ambystoma tigrinum*



# programy GeoDis, TCS

## Expected patterns under the different models of population structure and historical events

---

### Pattern 1. Restricted gene flow

- Significantly small  $D_c$ s, primarily for tip clades. Some interior clades with significantly large  $D_c$ s.
- $\bar{D}_c(I) - \bar{D}_c(T)$  significantly large, where  $\bar{D}_c(I)$  is the average clade distance of interior clades within the nested category, and  $\bar{D}_c(T)$  is the average clade distance of tip clades.
- Average  $D_c$ s should increase (and occasionally level off) with increasing clade level in a nested series of clades. If the distances level-off, the null hypothesis of no geographical association should no longer be rejected even though rejected at lower clade levels.
- The above patterns also hold for the  $D_n$ s unless some gene flow is due to long-distance dispersal events, then significant reversals of the above pattern can occur with the  $D_n$ s.

### Pattern 2. Range expansion

- Significantly large  $D_c$ s and  $D_n$ s for tip clades, and sometimes significantly small for interior clades under contiguous range expansion, but some tip clades should show significantly small  $D_c$ s under long-distance colonization.
- $\bar{D}_i(I) - \bar{D}_i(T)$  significantly small for  $i = c, n$  for contiguous range expansion and for  $i = n$  for long-distance colonization.
- The above patterns are not recurrent in the cladogram or are geographically congruent.

### Pattern 3. Allopatric fragmentation

- Significantly small  $D_c$ s, primarily at the higher clade levels. The  $D_n$ s at this clade level may suddenly increase rapidly while the  $D_c$ s remain restricted, depending upon the geographical configuration of the isolates.
  - The pattern of distances described in (a) should represent a break or a reversal of the distance pattern established by the lower level nested clades.
  - Clades showing pattern (a and b) should tend to be connected to the remainder of the cladogram by a larger-than-average number of mutational steps.
  - The above patterns are not recurrent in the cladogram or are geographically congruent.
-



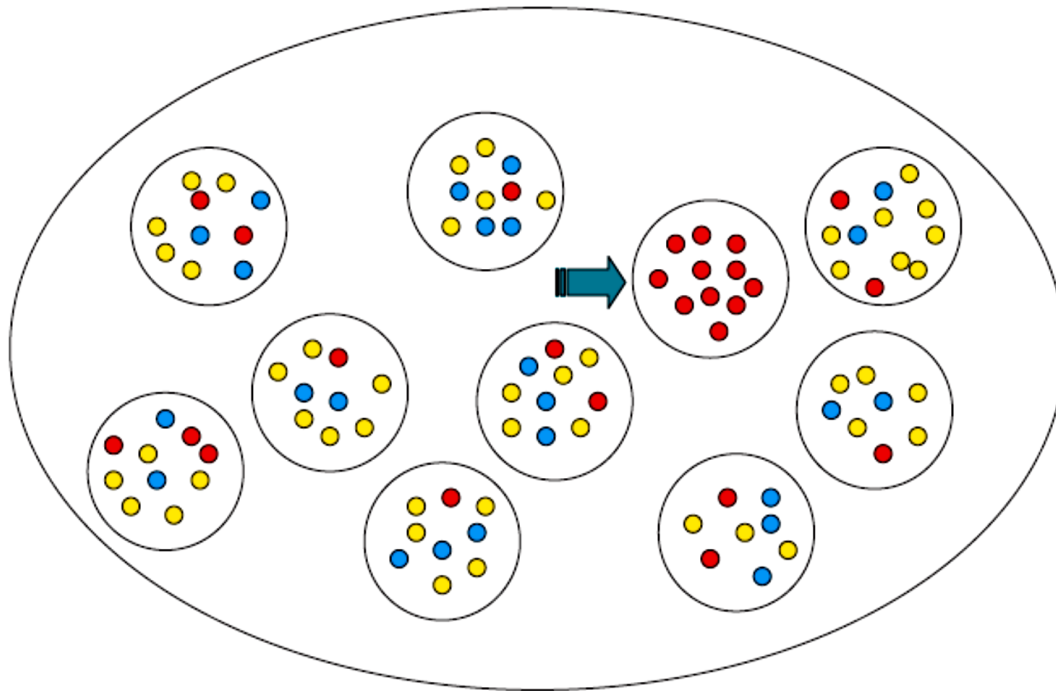
# Nested Clade Analysis (NCA)

Problémy:

absence evolučního modelu

simulace: vysoké procento „*false positives*“ (>75%; Petit 2008)

lokální procesy → zdánlivé expanze...



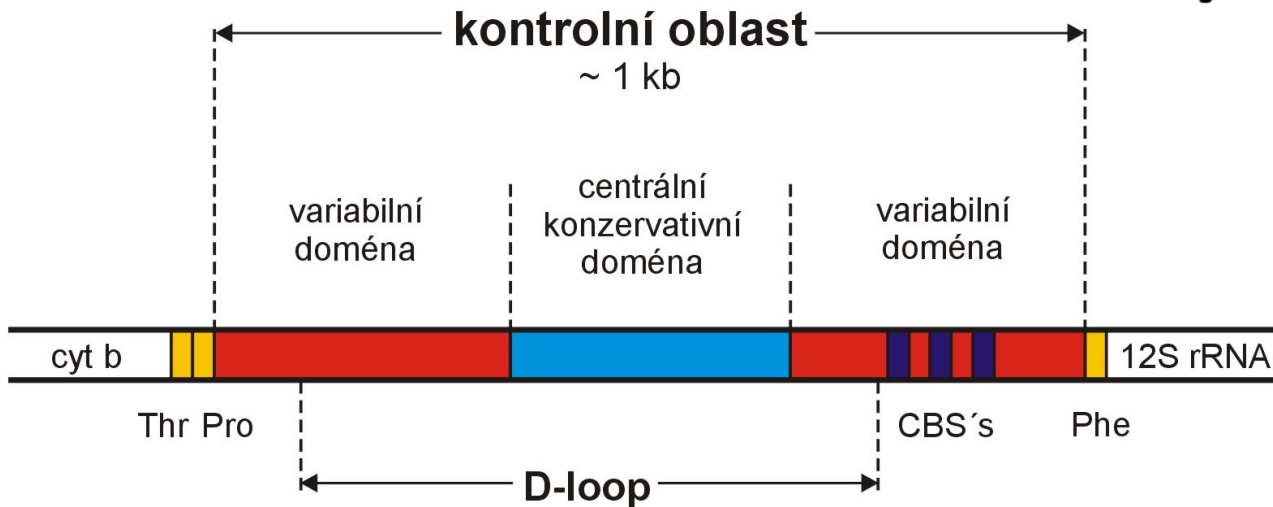
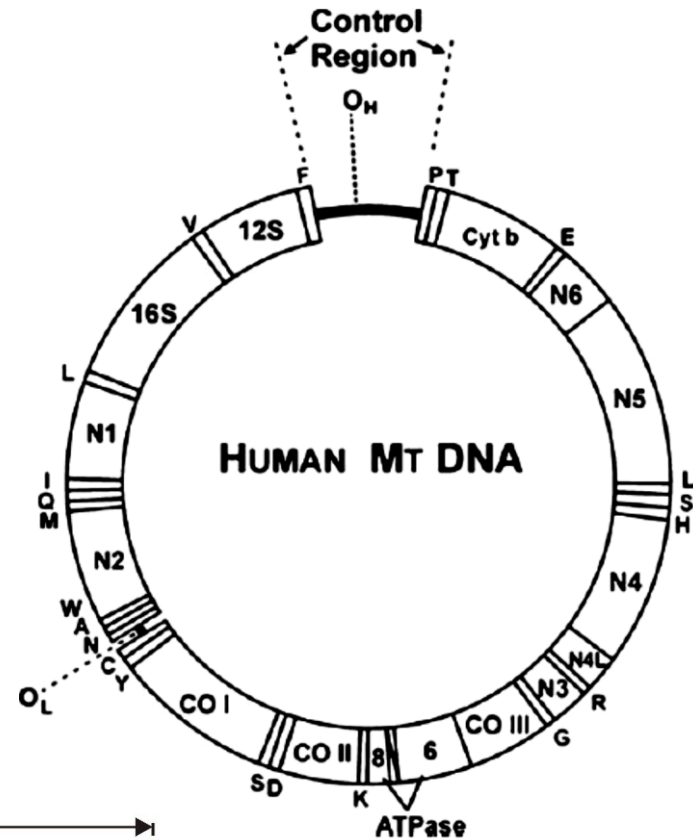
markery:

sekvence mtDNA

sekvence Y

mikrosatelity

SNP



## Proč je mtDNA výhodná?

- ? Malá (15-20 kb), kruhová molekula
- ? Bez intronů
- ? Minimum nekódujících oblastí
- ? Dědí se jen po jednom rodiči (po matce)
- ? Nerekombinuje
- ? V buňce jen jeden typ v mnoha kopiích
- ? Neutralita (různé varianty v populaci mají stejnou fitness)

... a proč ty otazníky?

# Problémy pro populační genetiku:

Neutralita

Mezidruhový přenos

Nukleární pseudogeny

Biparentální dědičnost

Rekombinace

## Neutralita?

vliv variant mtDNA na fitness (experimentální důkazy):

myš (*Mus*)

octomilka (*Drosophila*)

člověk

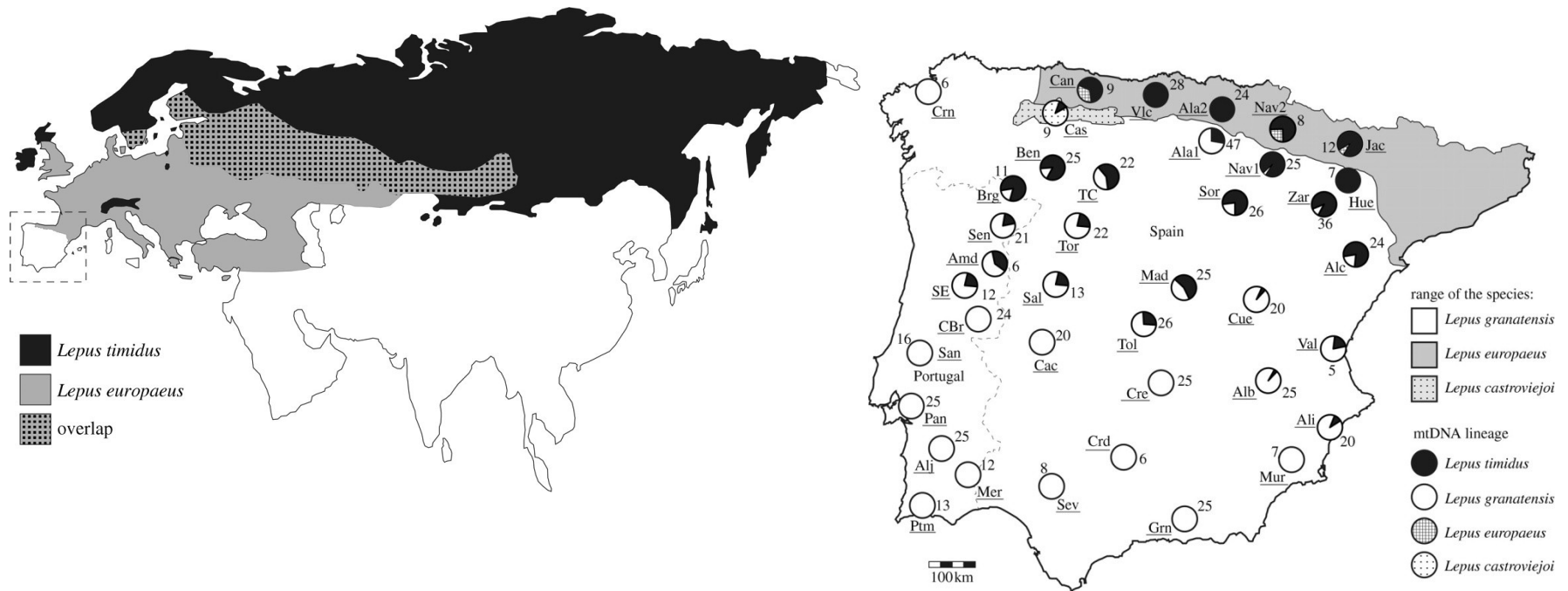
# Mezidruhový přenos:

zajáci ve Španělsku:

existence mtDNA *Lepus timidus* u druhů *L. granatensis*, *L. castroviejoi* a *L. europaeus*

*L. timidus* však vymizel na konci posledního glaciálu několikanásobný přenos různých mtDNA linií

= „*mtDNA capture*“



# Jaderné pseudogeny mitochondriálního původu = NUMT (Nuclear Mitochondrial DNA):

kopie segmentů mtDNA integrované do jaderné DNA

ztráta funkce

molekulární fosilie

podobnost s původní sekvencí → riziko amplifikace namísto mtDNA  
⇒ problém!!

výskyt různý u různých skupin i u různých druhů téže skupiny  
např. u 7 druhů kočkovitých šelem: numt > 12,5 kb  
člověk: 27 numt v linii po oddělení od společného předka  
se šimpanzem

# Jak na numt?

ultracentrifugace (většinou nutný čerstvý vzorek, nebo alespoň hluboce zmražený)

použití tkání s velkým množstvím mitochondrií (např. svaly)

long-range PCR

RT-PCR

elektronická PCR (u druhů se známým genomem)



## Rekombinace mtDNA:

nutné podmínky:

biparentální dědičnost – fúze mitochondrií

existence proteinového aparátu pro rekombinaci: existuje i u člověka

biparentální dědičnost:

navzdory mýtům, mitochondrie otce obvykle přeneseny do zygoty – tam jsou označeny a následně zlikvidovány (u savců značení provádí jaderné geny otce)

→ u někt. druhů „*paternal leakage*“: *Mus*, *Drosophila*, *Parus*, *Homo*

# Rekombinace mtDNA:

biparentální dědičnost:

Gyllensten et al., 1991: Paternal inheritance of mitochondrial DNA in mice. *Nature* 352: 255–257.

F1 hybridi *Mus spretus* × C57BL

frekvence otcovské mtDNA ve vztahu k mateřským  $\approx 10^{-4}$

**Maternal Inheritance of Mouse mtDNA in Interspecific Hybrids: Segregation of the Leaked Paternal mtDNA Followed by the Prevention of Subsequent Paternal Leakage**

Hiroshi Shitara,<sup>\*,†</sup> Jun-Ichi Hayashi,<sup>\*</sup> Sumiyo Takahama,<sup>†</sup> Hideki Kaneda<sup>†</sup> and Hiromichi Yonekawa<sup>†</sup>

Shitara et al., 1998: *Genetics* 148: 851–857.

F1 hybridi *Mus spretus* × C57BL

„prosakování“ otcovské mtDNA ne ve všech tkáních  
pouze u F1, ne v dalších generacích (u zpětných kříženců) → druhově  
specifická exkluze

## Zbývá zodpovědět:

Frekvence biparentální dědičnosti

Frekvence rekombinací

Jen somatické tkáně?

Dědičnost rekombinantních molekul

Rekombinace s nukleárními pseudogeny?

Mechanismus biparentální dědičnosti a rekombinace

Výjimečné události u živočichů?