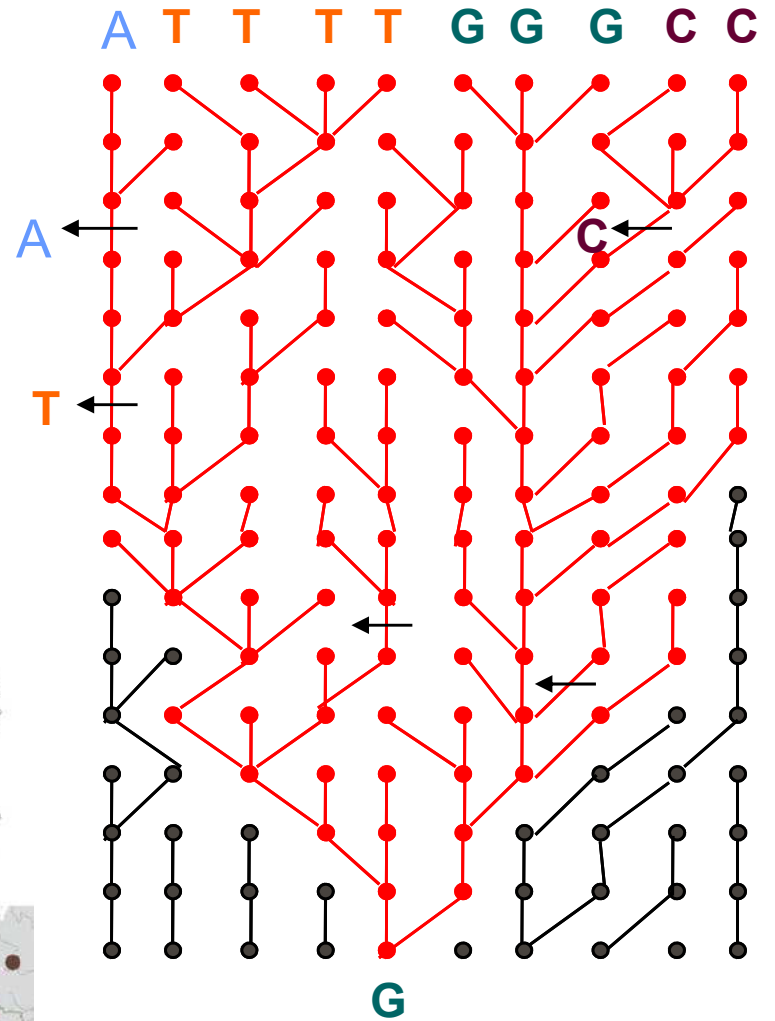
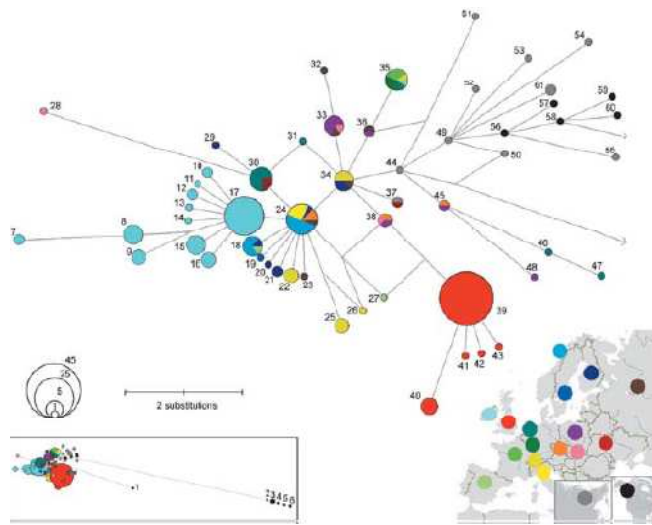
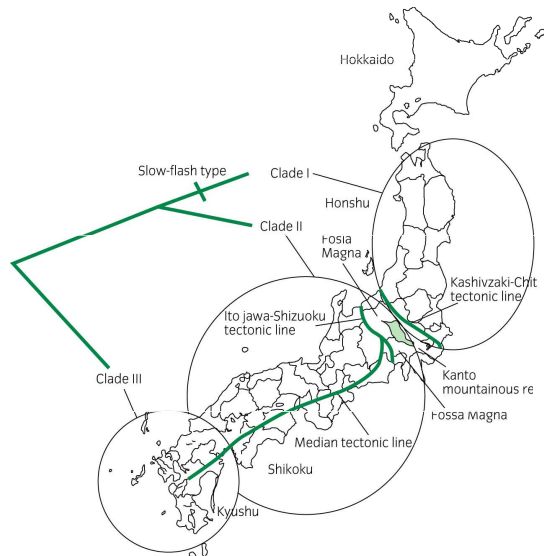
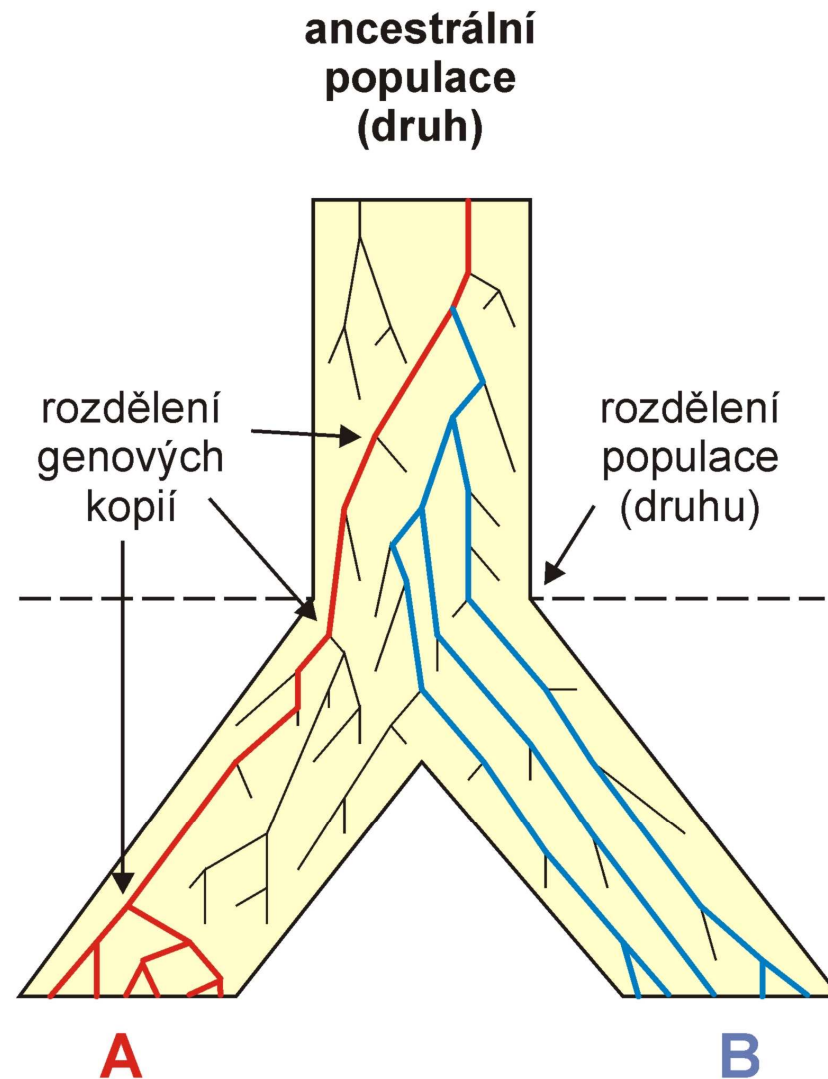


FYLOGEOGRAFIE A KOALESCENCE

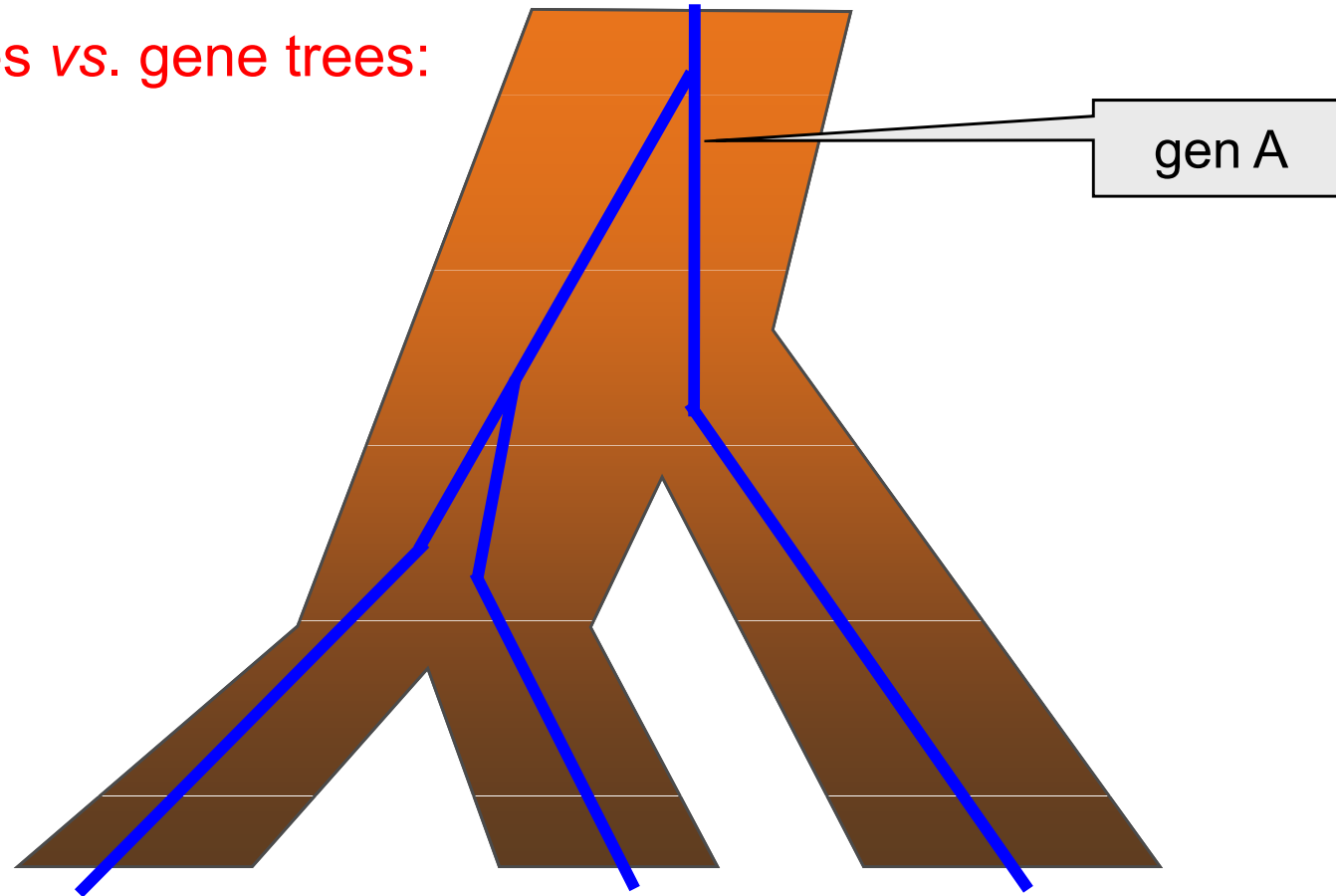


Koalescence

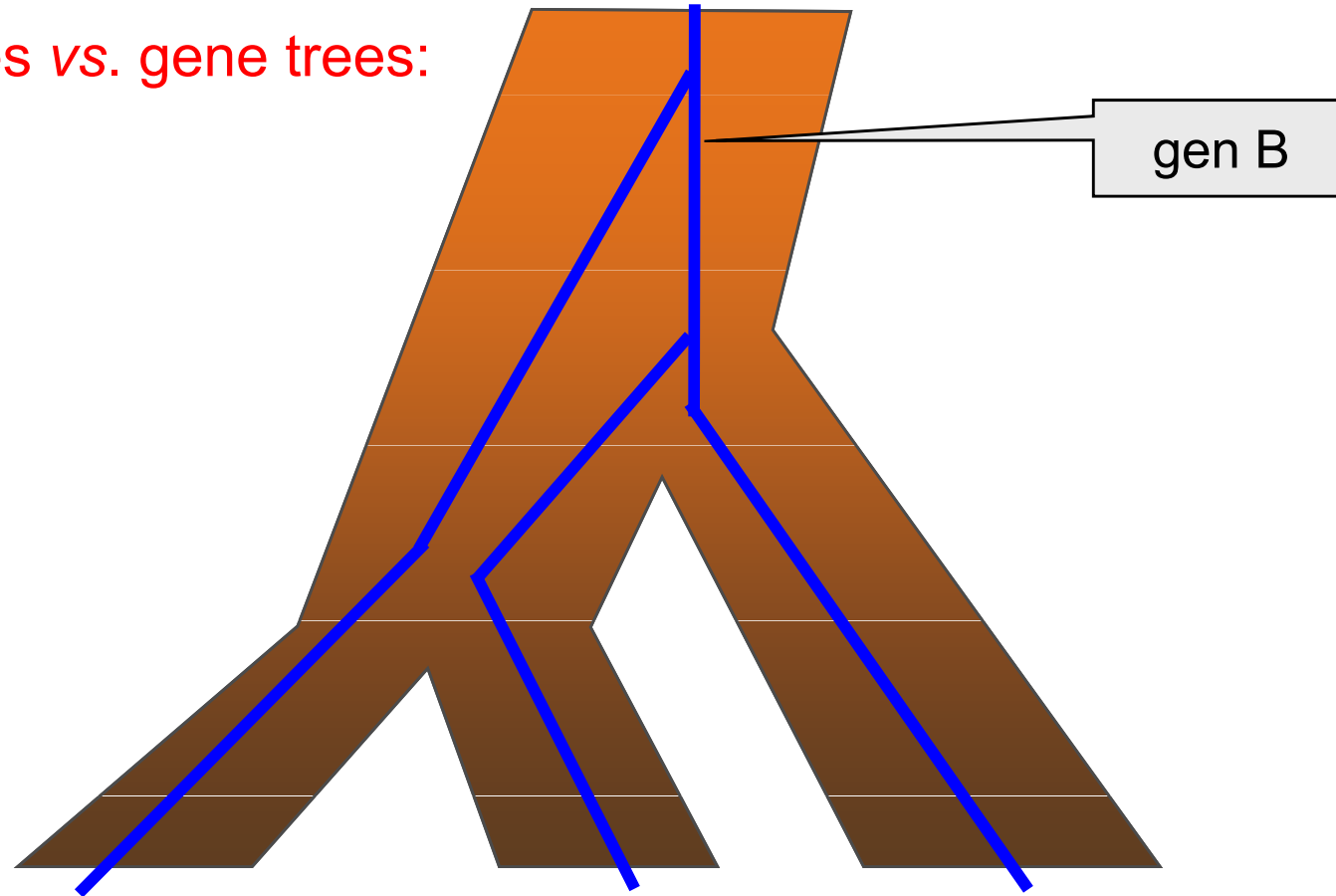
Osud jednotlivých kopií genů v populaci → **genové stromy**



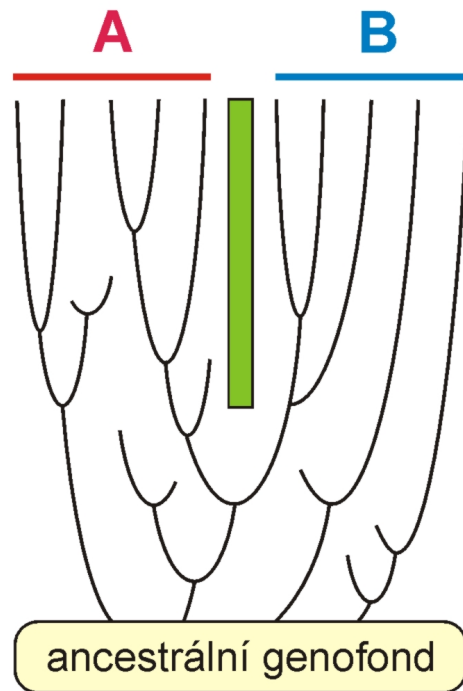
Species trees vs. gene trees:



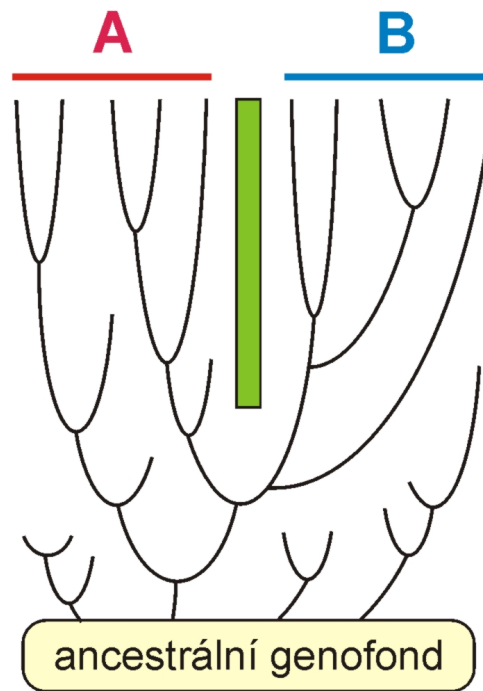
Species trees vs. gene trees:



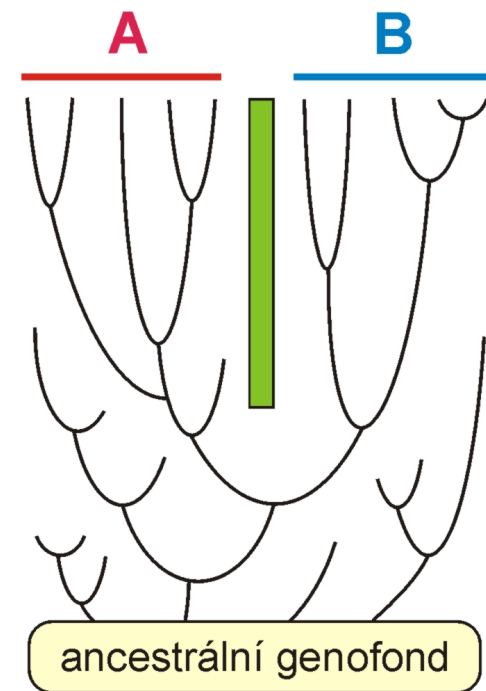
Fylogenetické vztahy 2 dceřinných populací mtDNA:



polyfylye

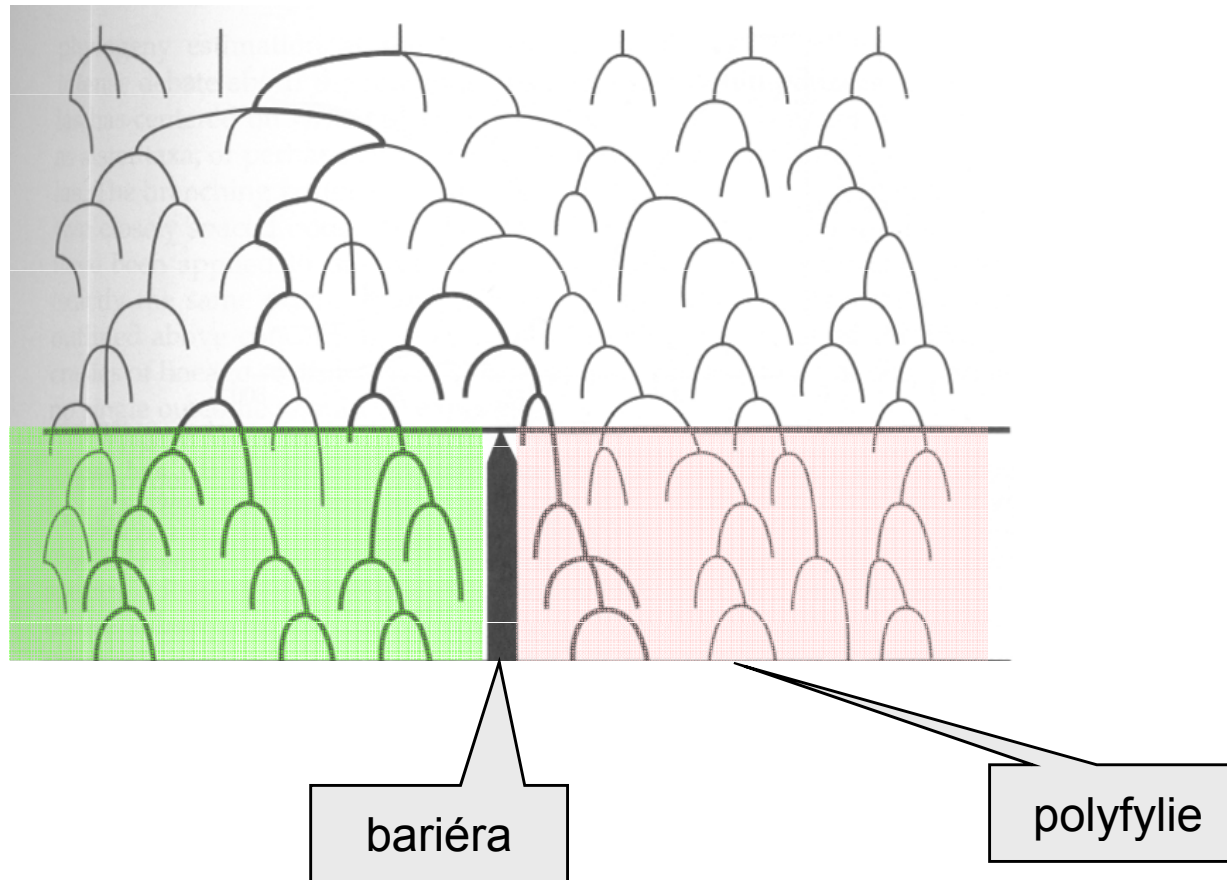


parafylye

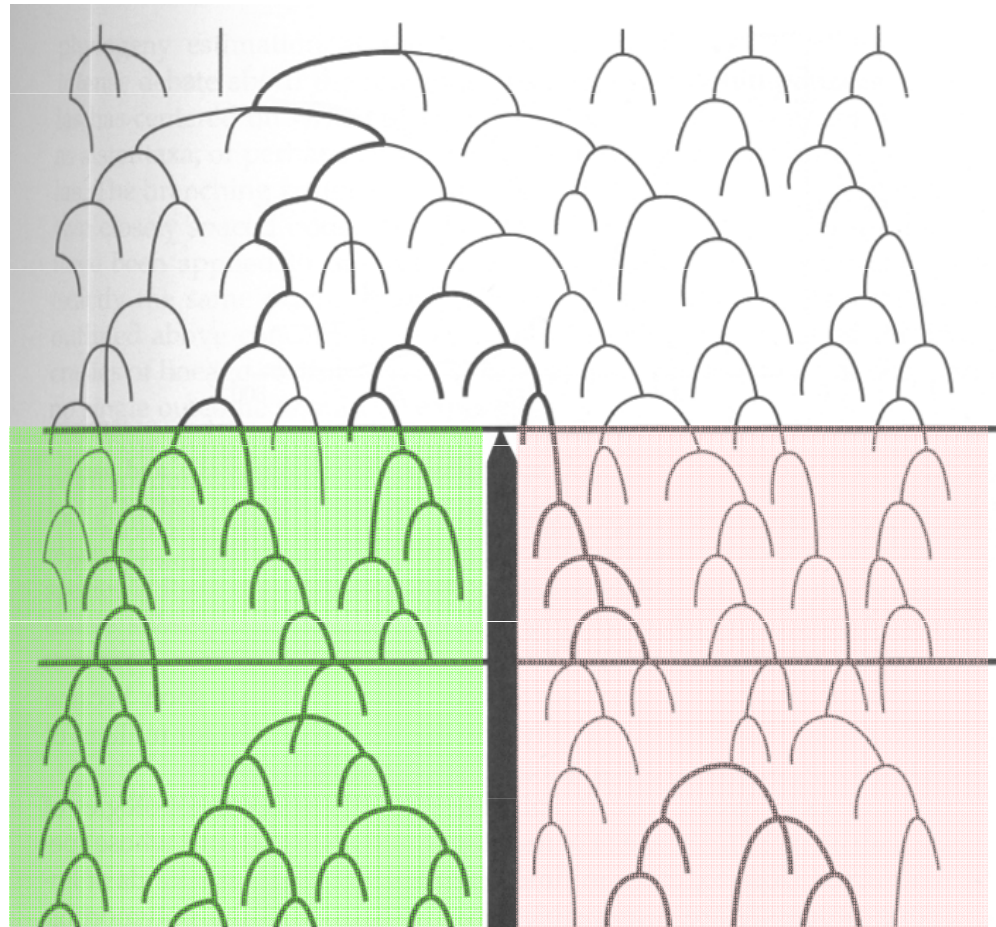


reciproční monofylye

Ancestrální polymorfismus a sortování linií

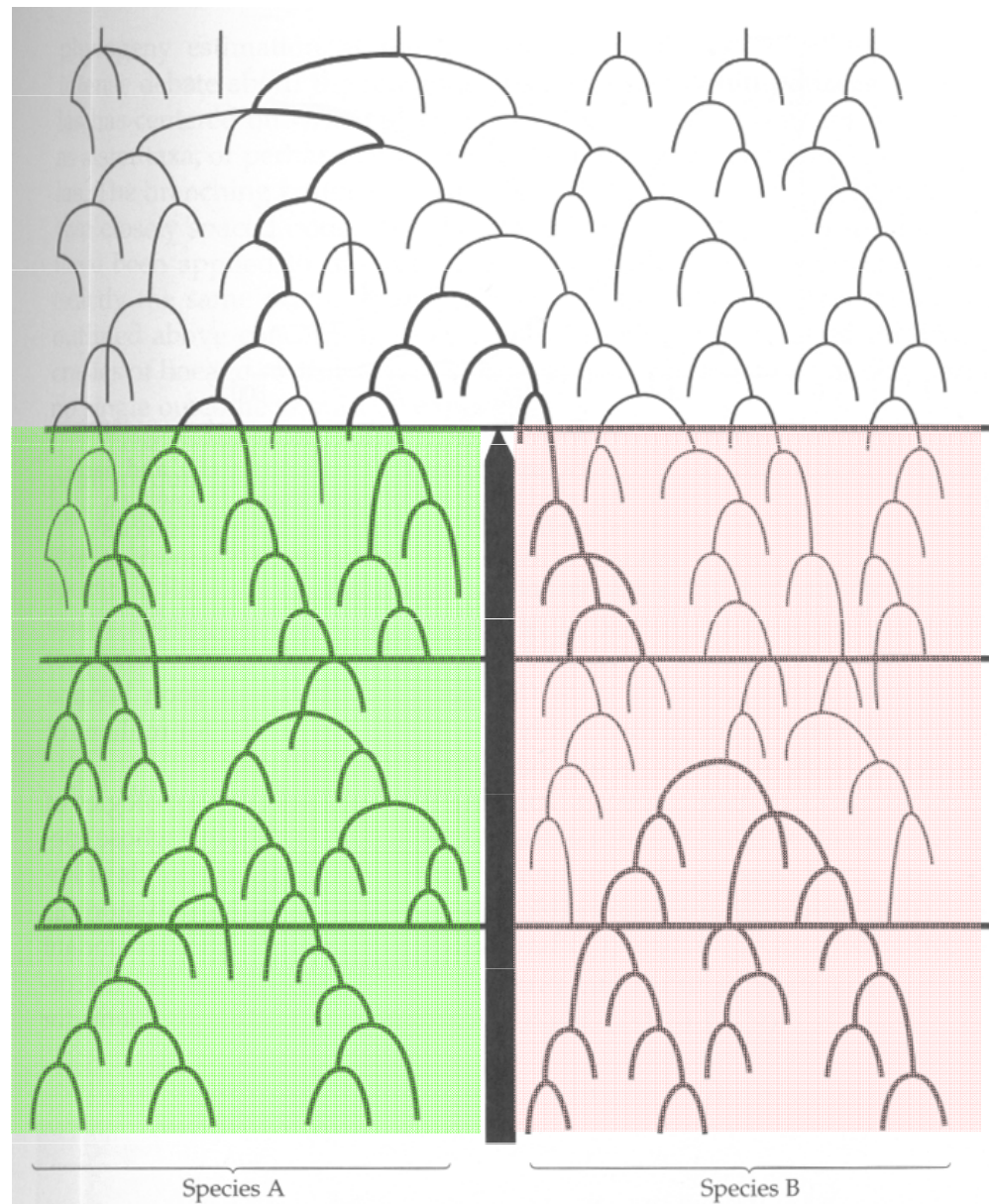


Ancestrální polymorfismus a sortování linií



parafyletická
fáze

Ancestrální polymorfismus a sortování linií

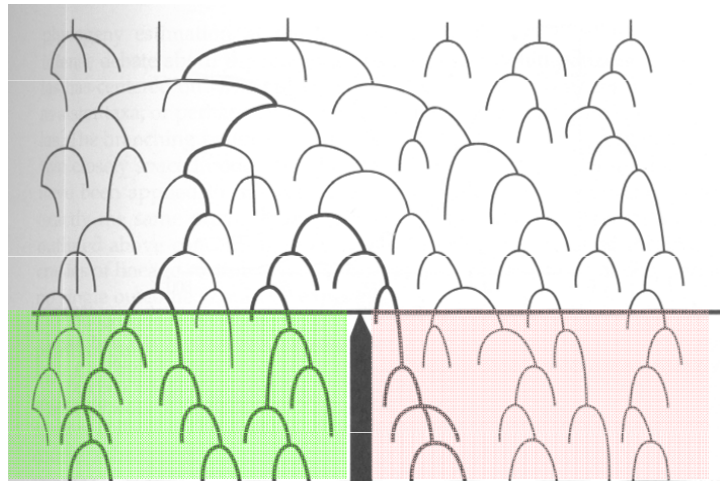


druh A

druh B

reciproční
monofylie

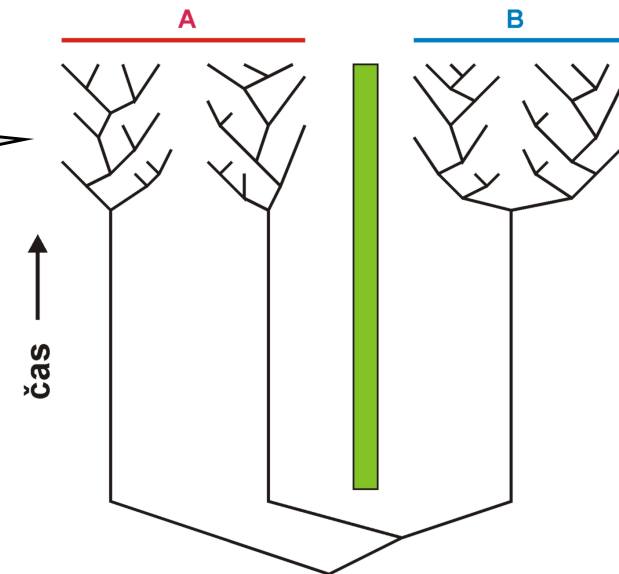
Ancestrální polymorfismus a sortování linií



druh A

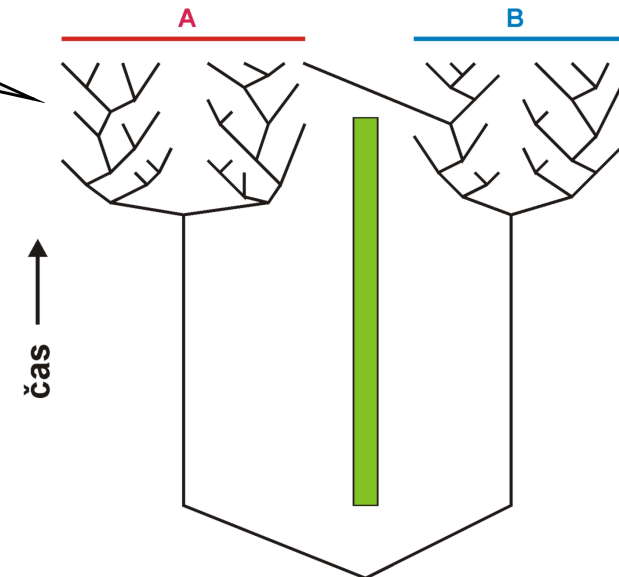
druh B

nekompletní
sortování linií



čas ↑

recentní
tok genů

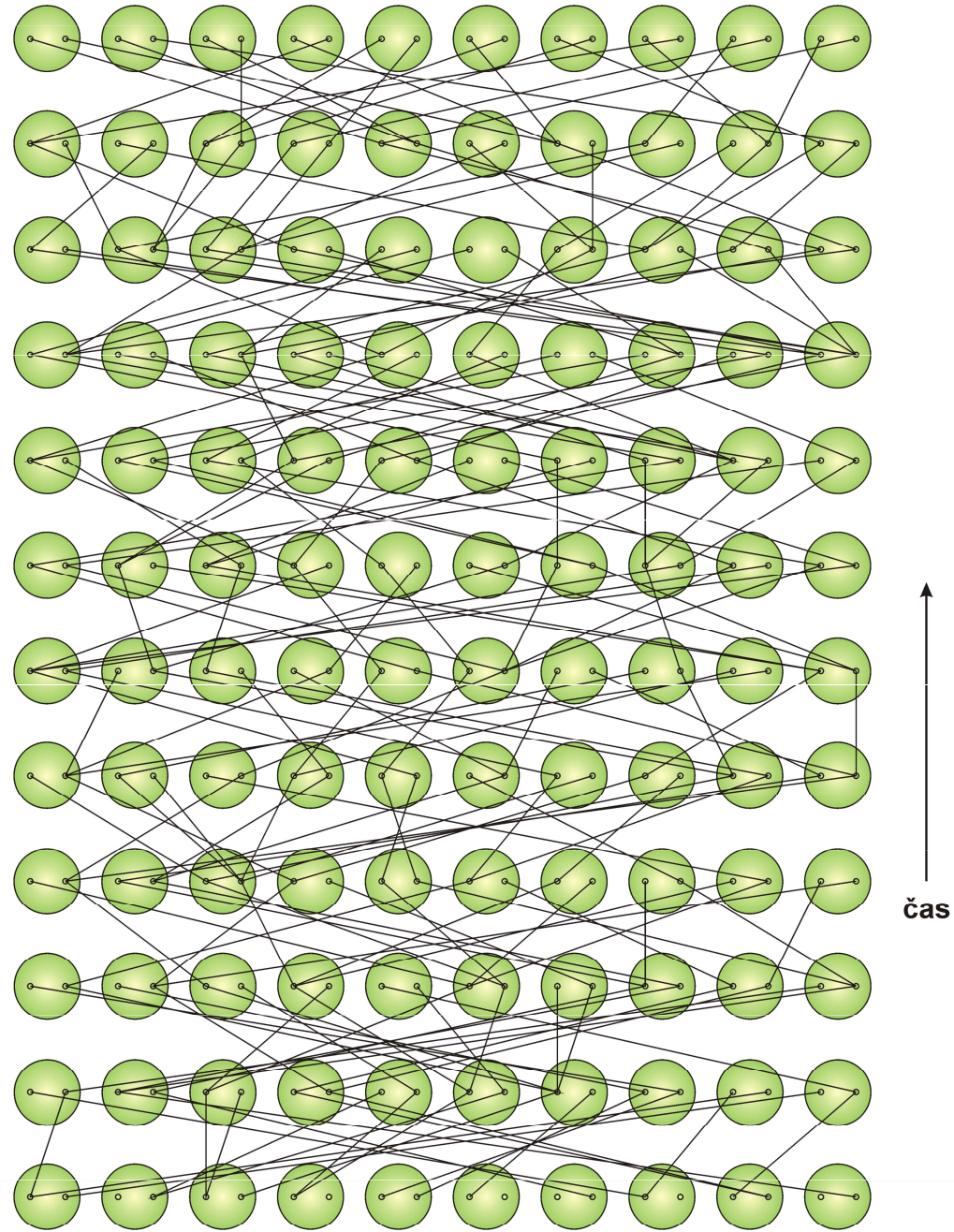


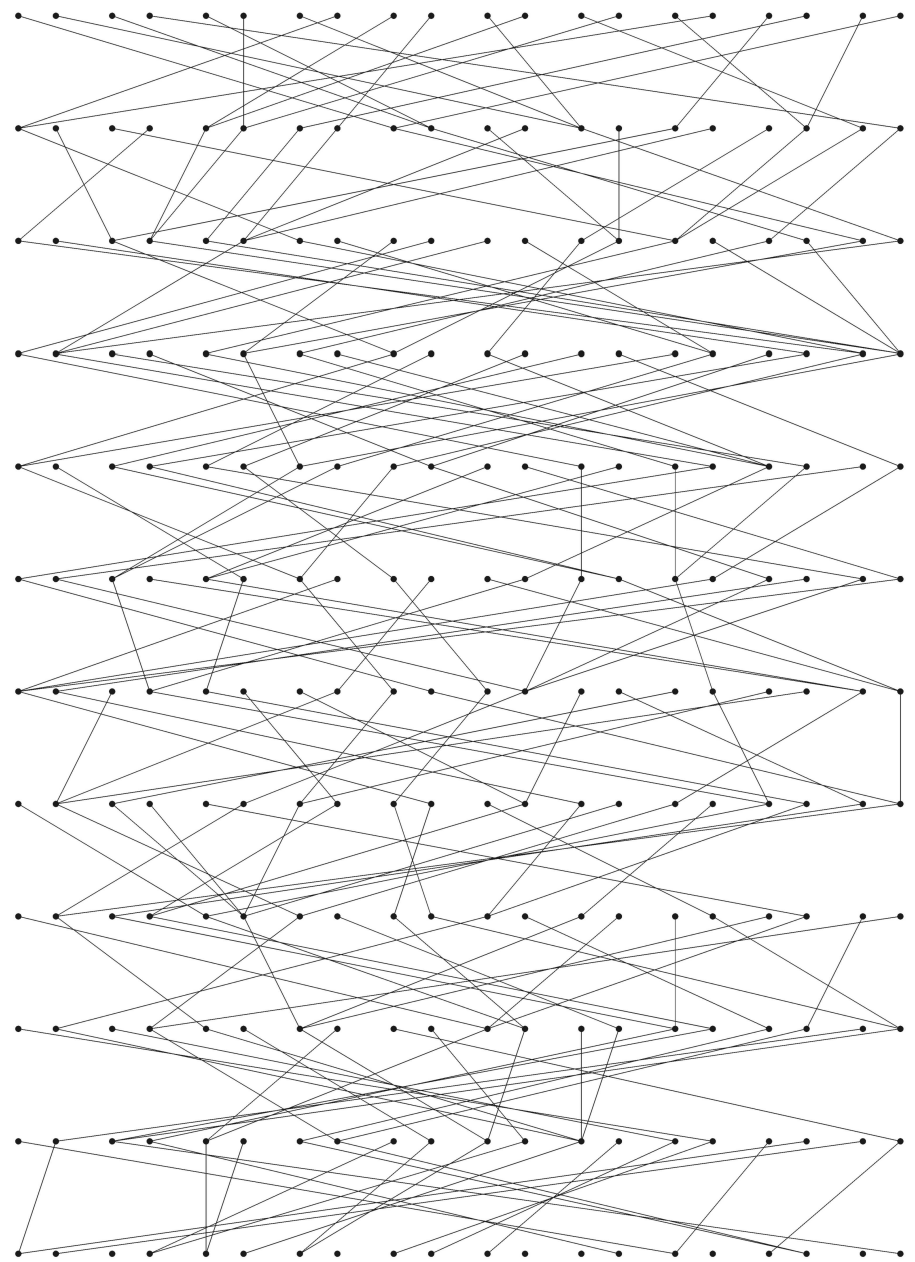
čas ↑

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

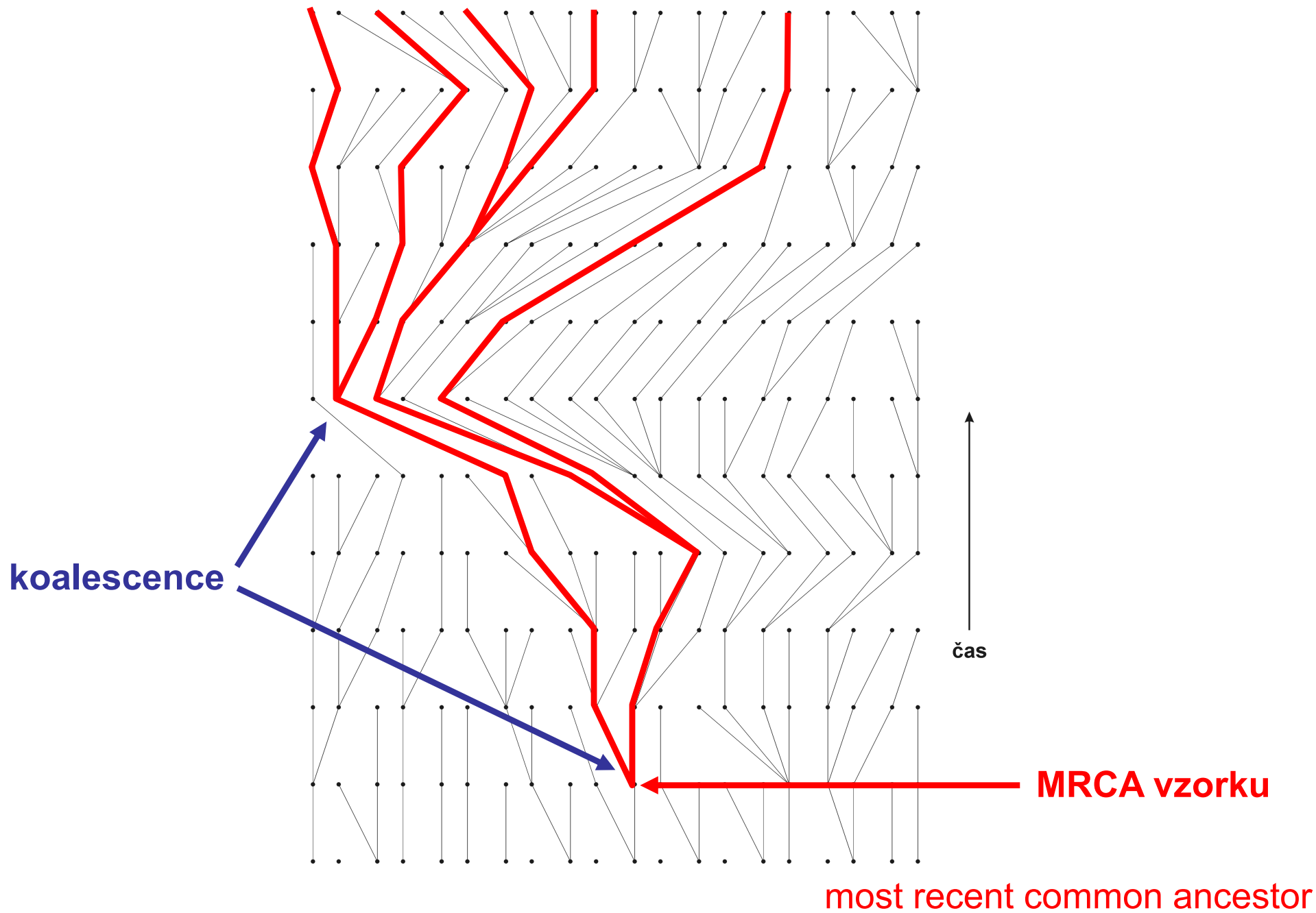
Wrightův-Fisherův model (drift):

- konstantní velikost populace o N jedincích
- náhodné křížení
- počet potomků jednotlivých jedinců náhodný (Poissonov rozdělení)
- Hardyho-Weinbergova rovnováha
- nepřekrývající se generace
- jedinci monoičtí (jednodomí)
- možné i samooplození
- gen selektivně neutrální
- populace není geograficky strukturovaná

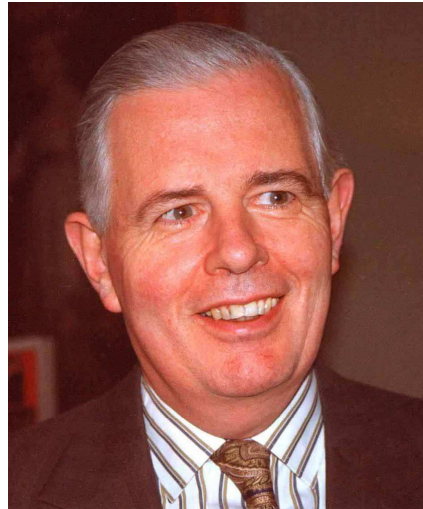




↑
čas



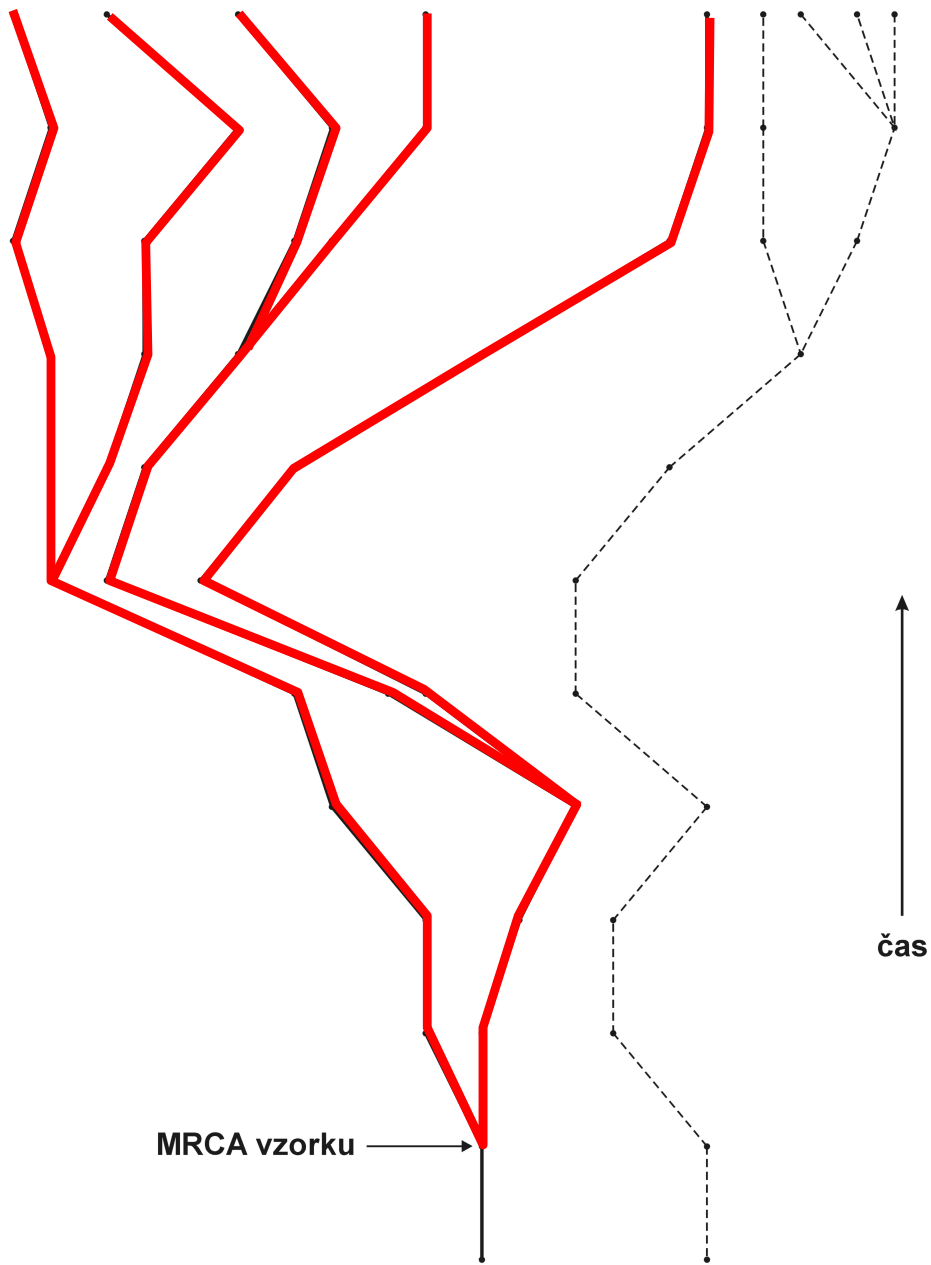
Analogie procesu koalescence: příklad švábů v krabici
pravděpodobnost srážky = $k(k - 1)/4N$ ($/2N$ v případě haploidie)



Sir John F.C. Kingman

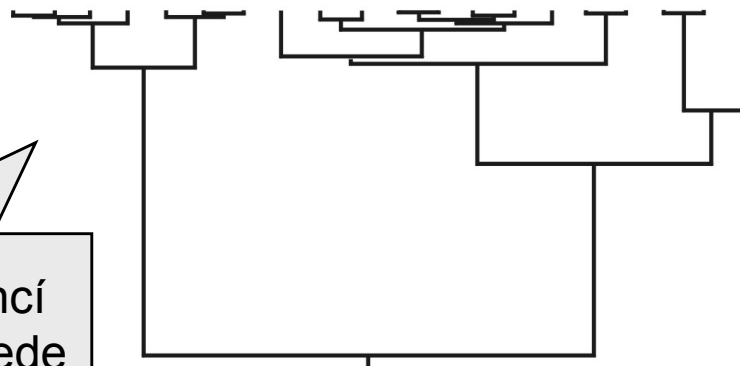
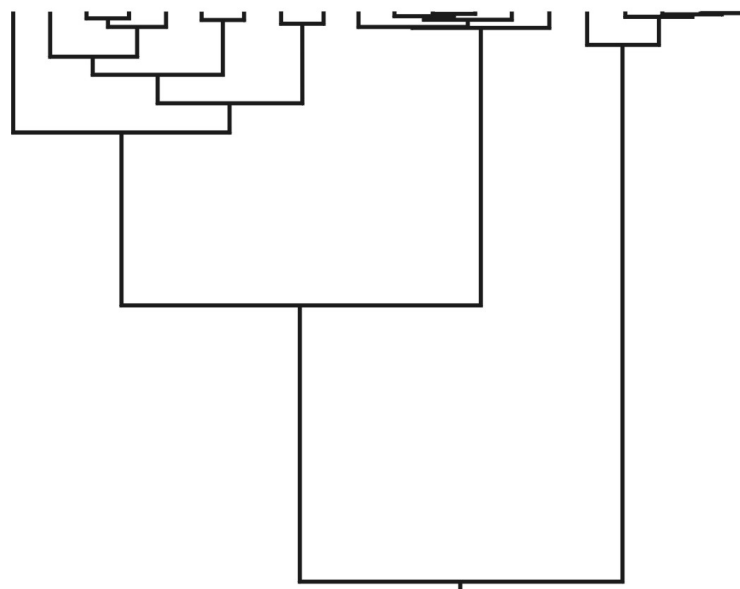
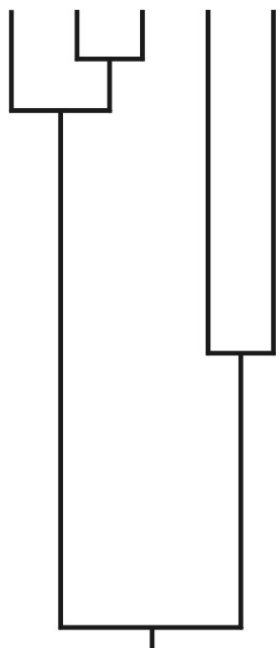
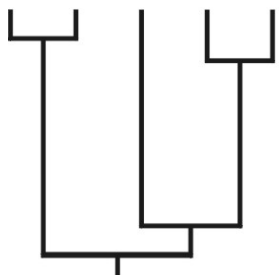
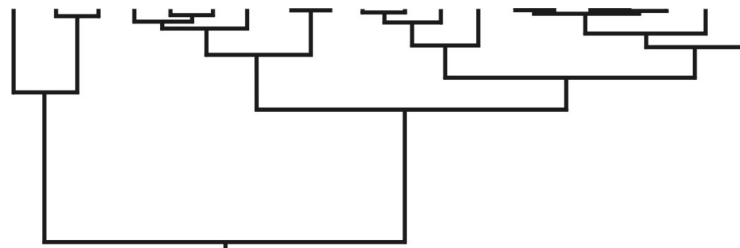
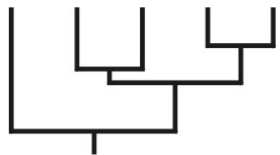
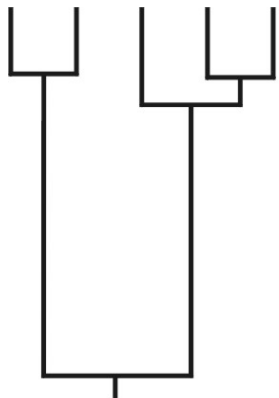
Kingmanova koalescence:

- vzorek k genových kopií z celkového počtu N (uvažujeme haploidní stav)
- pravděpodobnost, že 2 kopie pocházejí ze stejného předka v předchozí generaci je rovna $1/N$
- geometrické rozdělení – aproximace exponenciálním rozdělením
- zajímá nás pouze genealogie kopií, které zanechaly potomky do současnosti



Kingmanova koalescence:

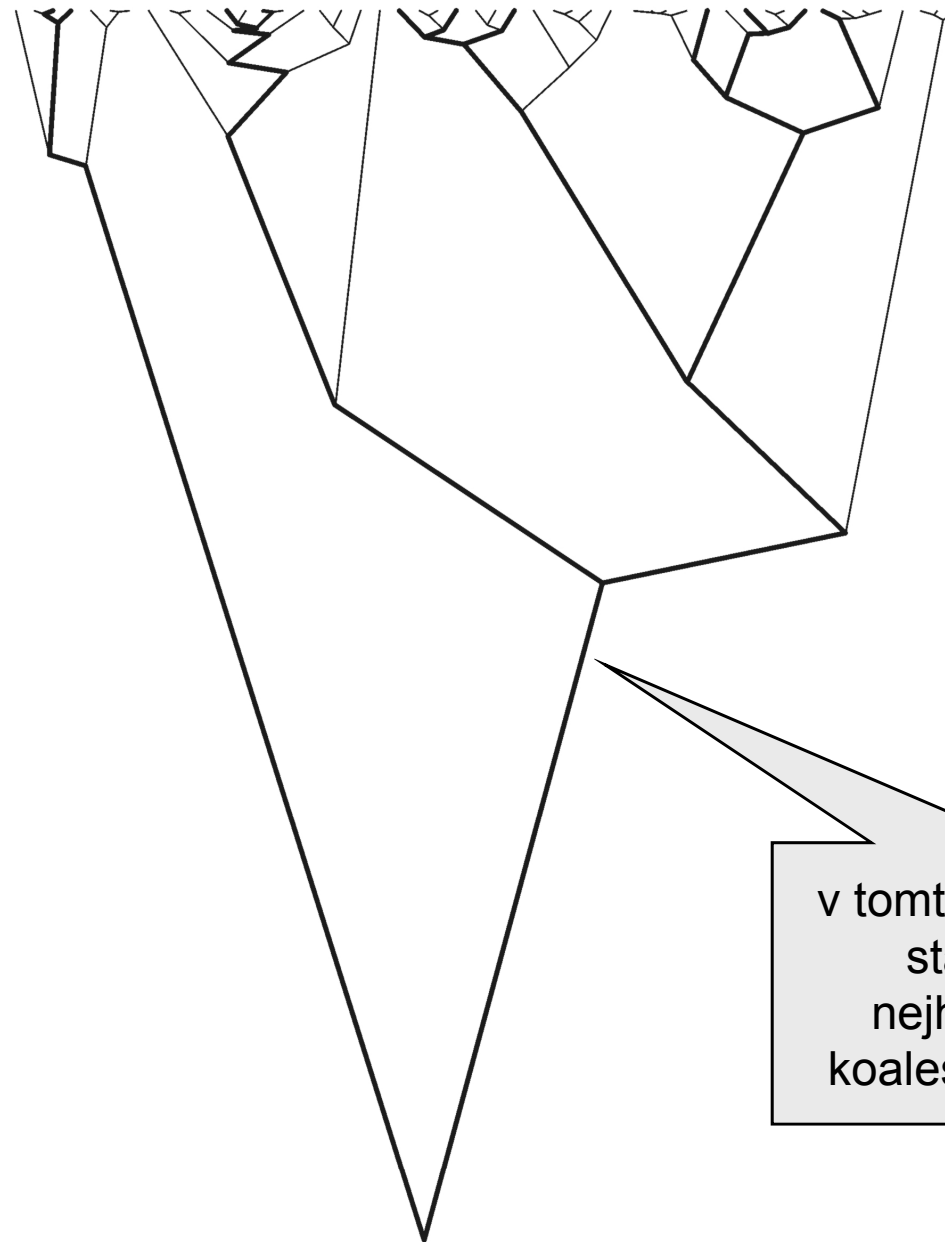
- čím menší počet zbývajících kopií, tím se **proces koalescence zpomaluje** (pro velká $k \sim 4N$, pro 2 kopie $\sim 2N$)
- splynutí posledních n kopií zabere $(1 - 1/n)/(1 - 1/k)$
 \Rightarrow **prvních 90% kopií splyne během 9% celkového času, zbývajících 91% času se čeká na splynutí posledních 10 kopií!**
- jestliže 100 linií, pravděpodobnost, že 101. linie přidá hlubší kořen je pouze 0,02% \Rightarrow **přidání další genové kopie pravděpodobně nepovede k hlubší (starší) koalescenci**



s klesajícím počtem
volných kopií se proces
zpomaluje ...

přidání dalších sekvencí
pravděpodobně nepovede
k hlubší koalescenci ...

50 genových kopií, 10 náhodně vybraných:



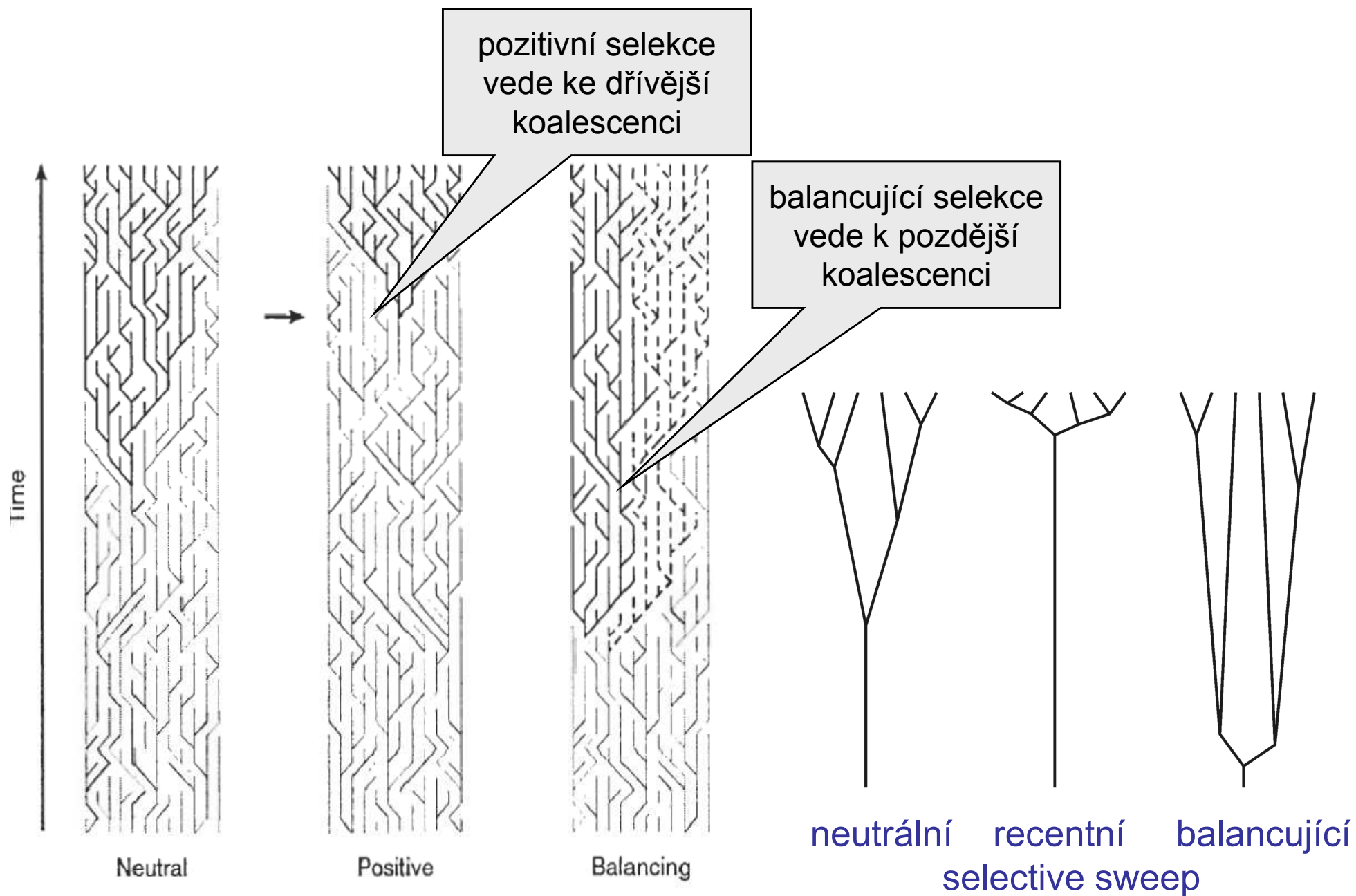
v tomto případě 10 kopií
stačí k nalezení
nejhlubšího kořene
koalescenčního stromu

Koalescence je ovlivněna různými faktory, např.:

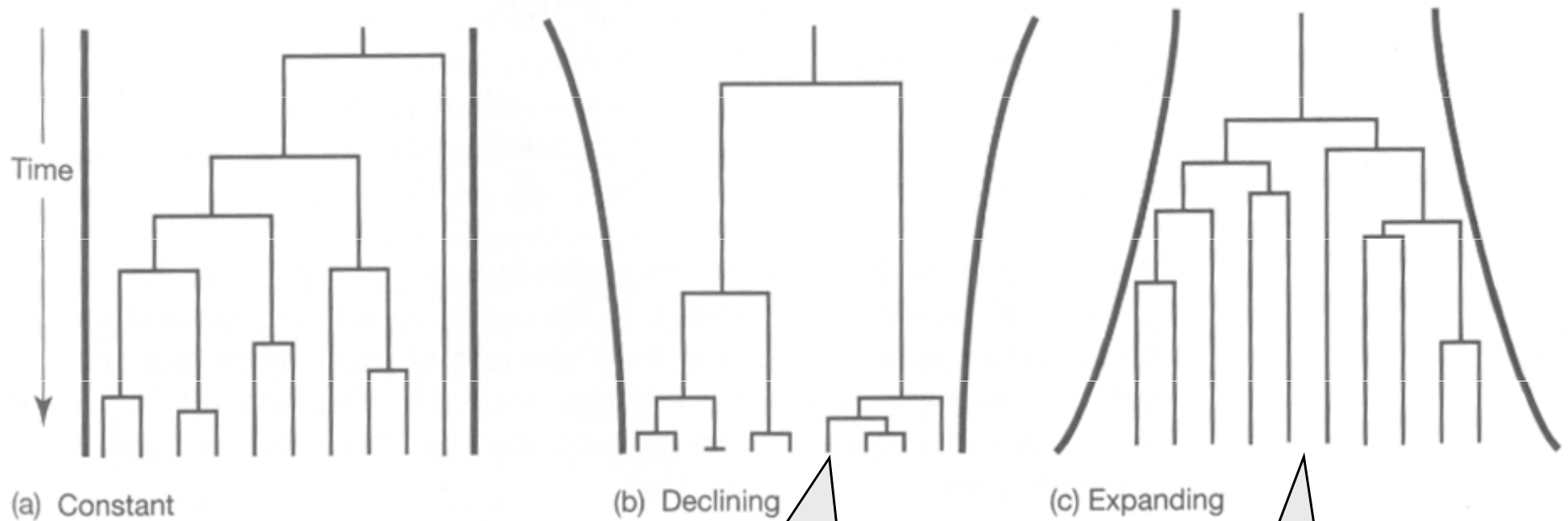
- mutací
- rekombinací
- selekcí
- změnami velikosti populace

⇒ koalescenční teorii lze použít k odhadu těchto parametrů

Vliv selekce na tvar koalescenčního stromu



Vliv změn velikosti populace na tvar koalescenčního stromu

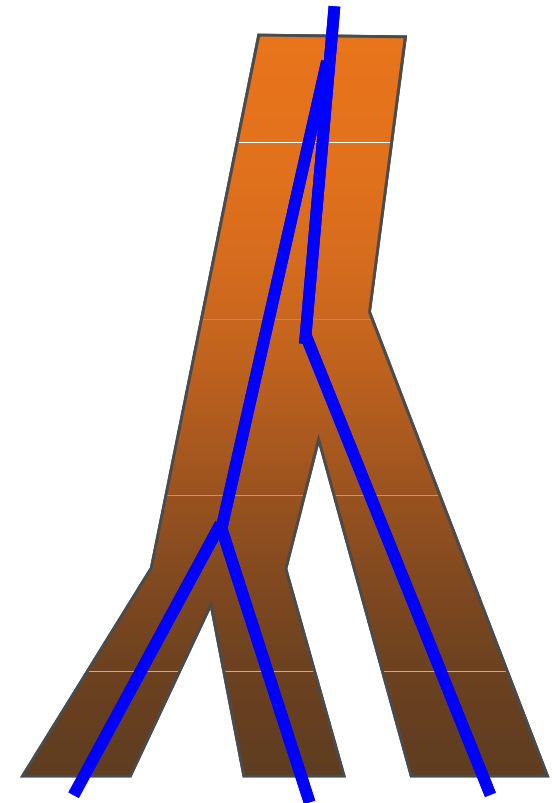


zmenšující se populace:
koalescence se postupně
zrychlují

rostoucí populace:
koalescence se
postupně zpomalují

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
- 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

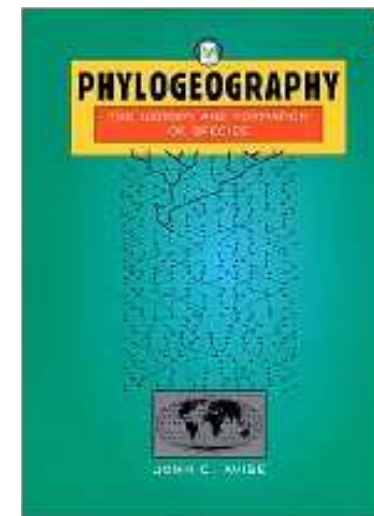


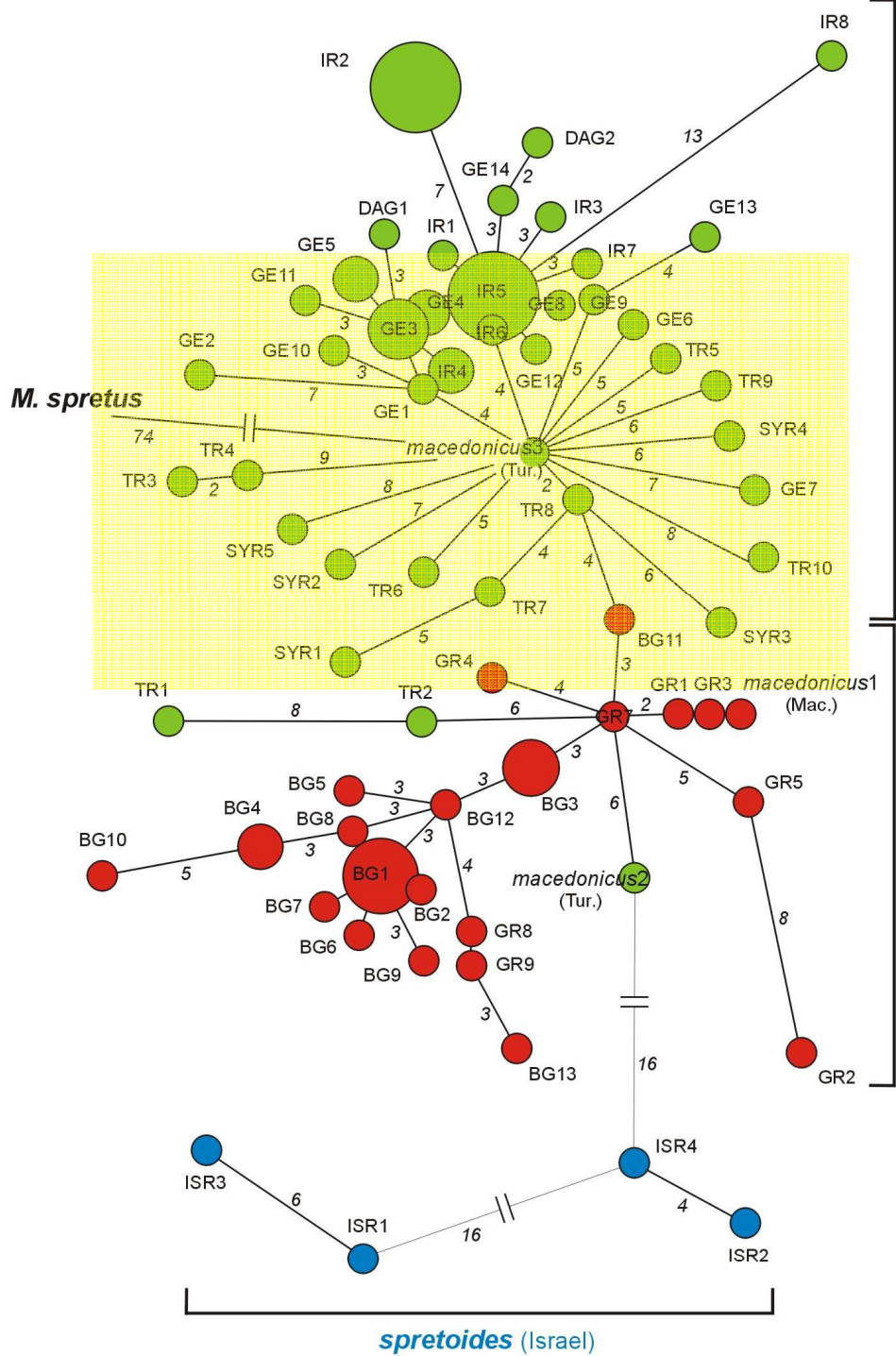
Fylogeografie

- 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

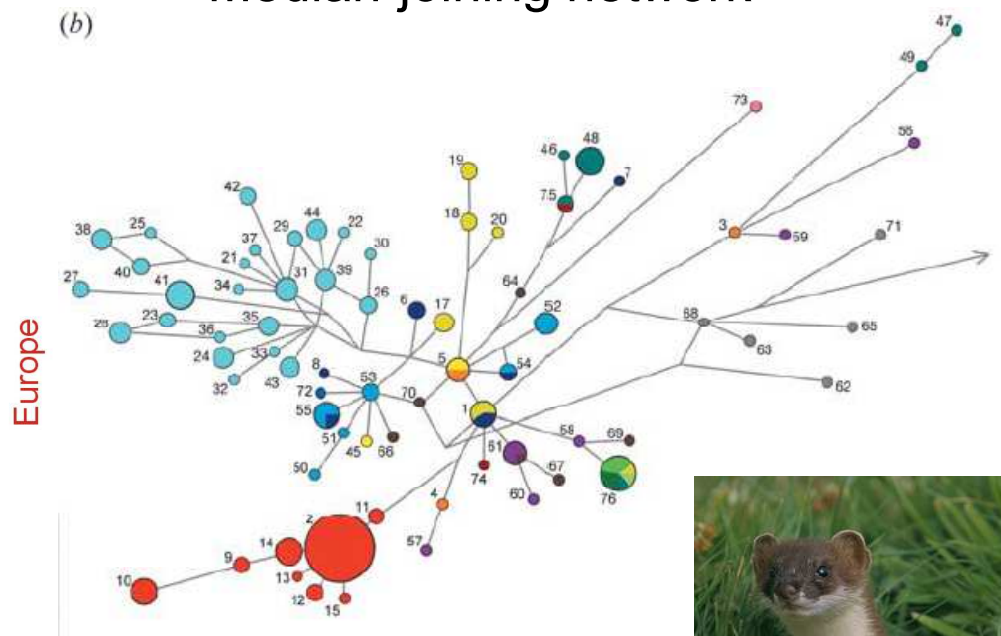




Mus macedonicus

Asia

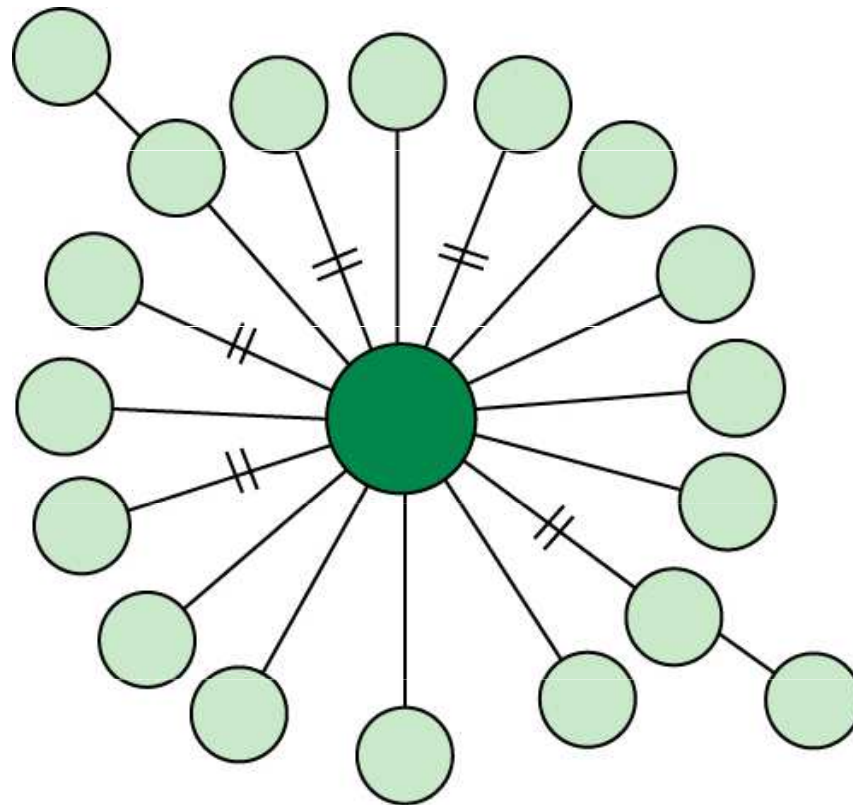
- Minimum Spanning Tree (MST)
- Mimum Spanning Network (MSN)
- Median-joining network



Mustela erminea

Recentní expanze:

- rychlé rozšíření jednoho haplotypu
- akumulace malého počtu mutací
- hvězdicová struktura



Změny velikosti populace

- Tajimův test (Tajima's D)
- mismatch distribution
- koalescence, ML nebo BA, MCMC
- Bayesian Skyline Plots

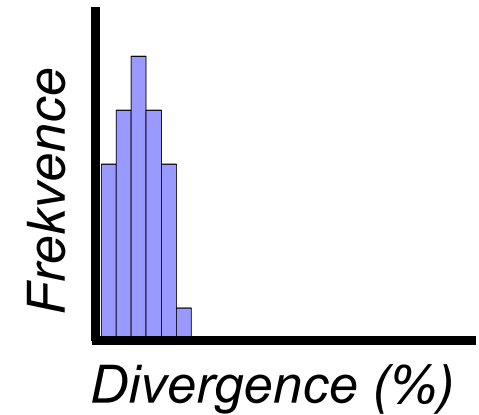
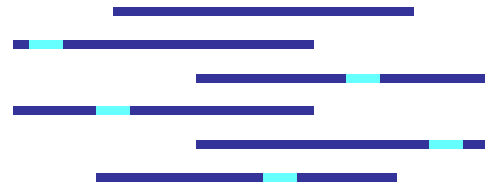
1. Tajimův test (Tajima's D)

- 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

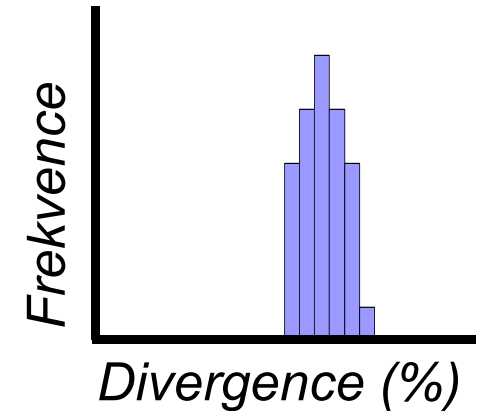
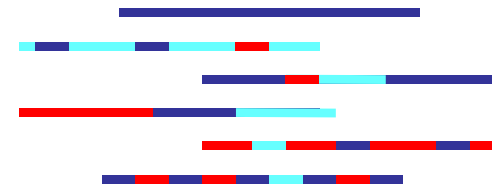
2. Mismatch distribution

- párová srovnání všech sekvencí → histogram

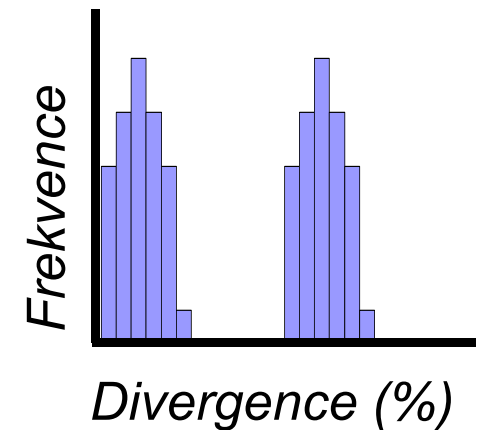
Sekvence navzájem velmi podobné



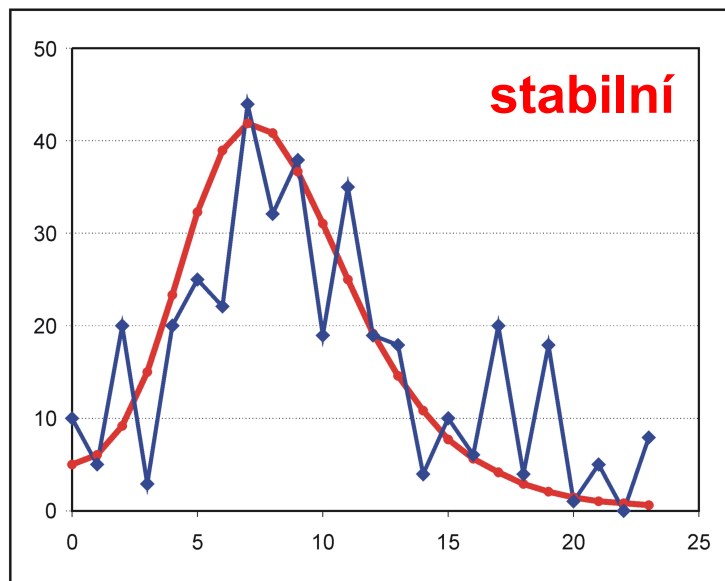
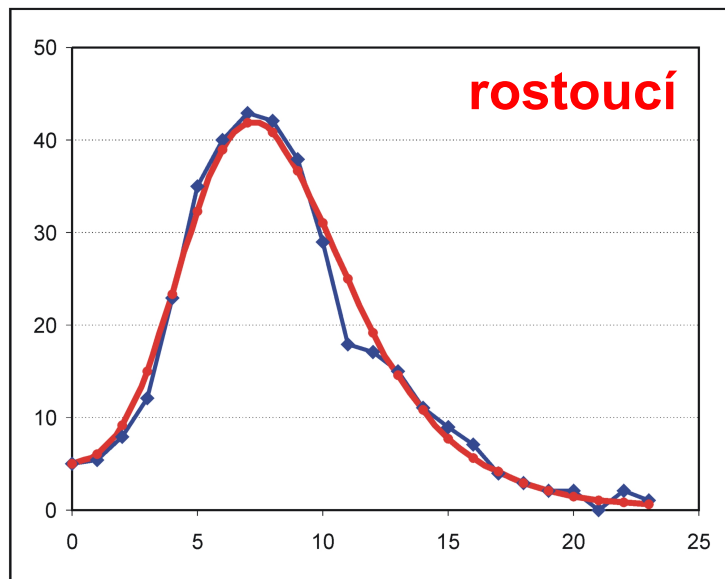
Sekvence navzájem velmi odlišné



Směs podobných a odlišných sekvencí

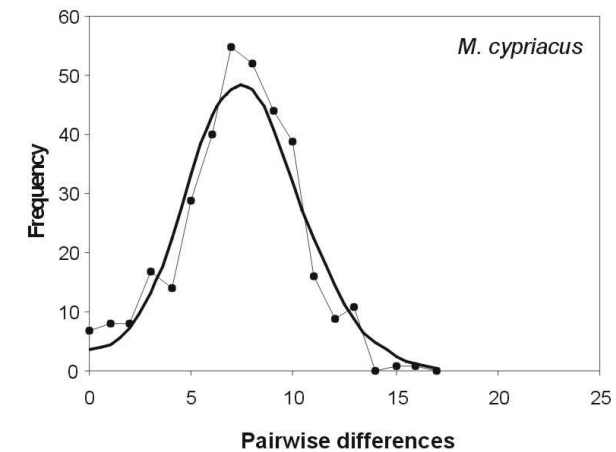
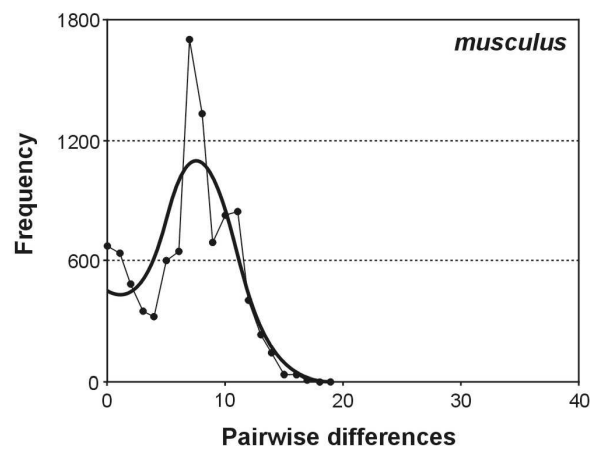
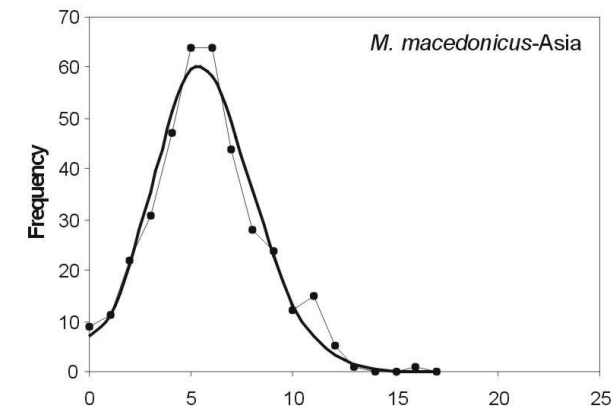
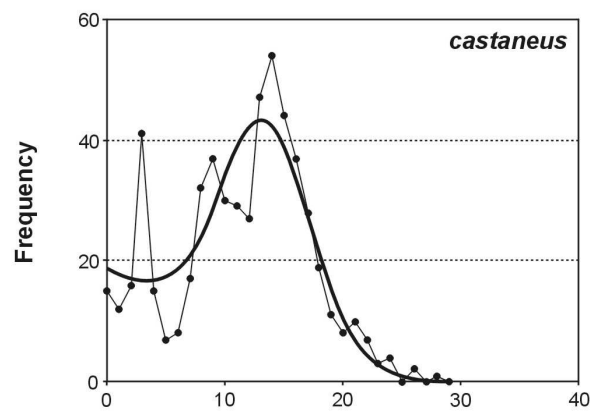
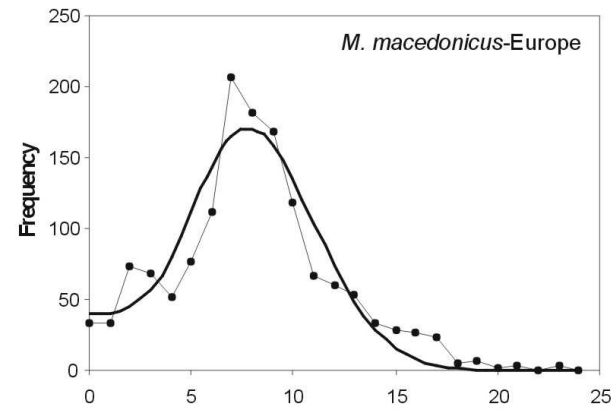
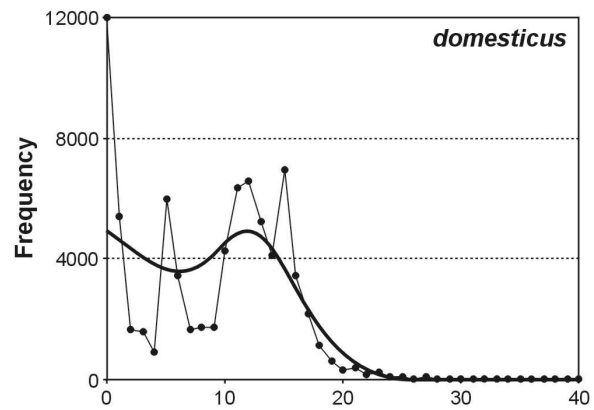


frekvence



párové rozdíly

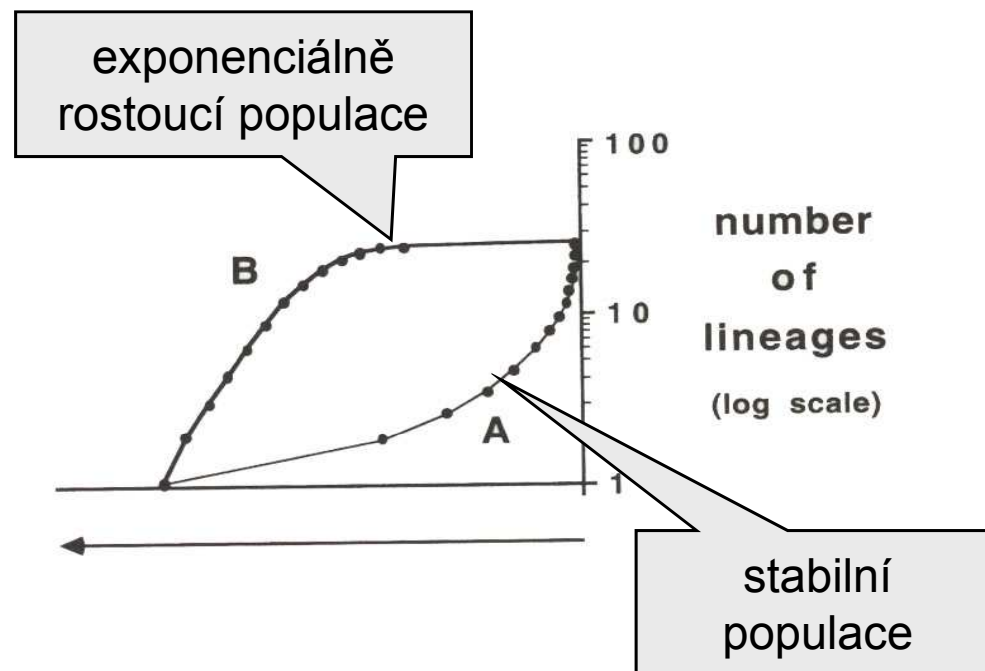
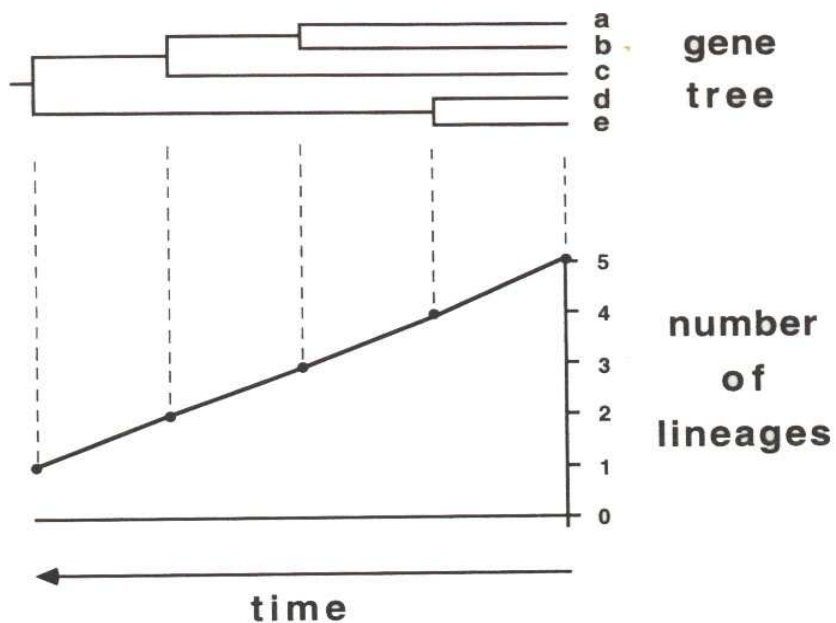
- test shody skutečného rozdělení s teoretickou predikcí:
- Harpending's raggedness index (Harpending 1994)
- sum of squared deviations
- doba expanze/bottlenecku:
 $\tau = 1/2u$,
kde u je mutační frekvence pro celou sekvenci
- můžeme odhadnout i velikost populace před a po expanzi



3. ML a bayesiánské odhady

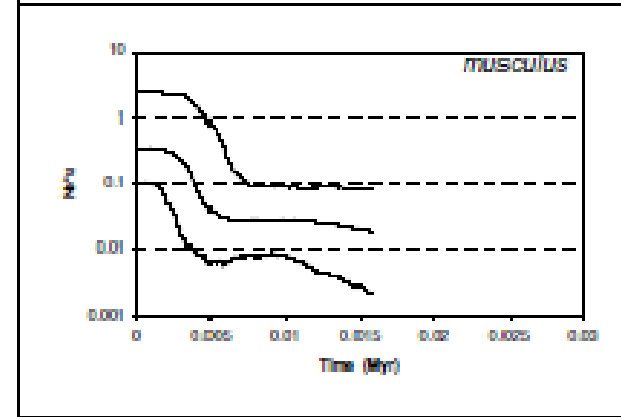
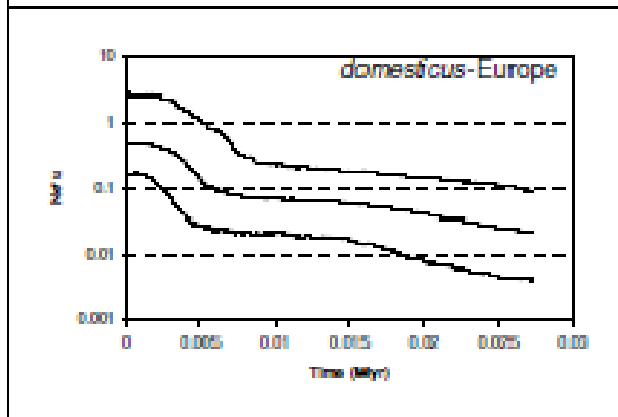
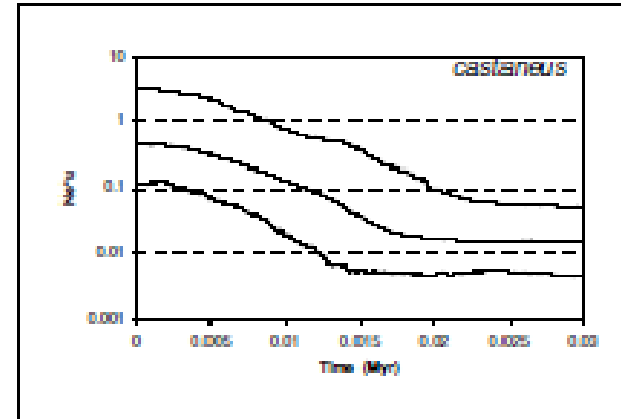
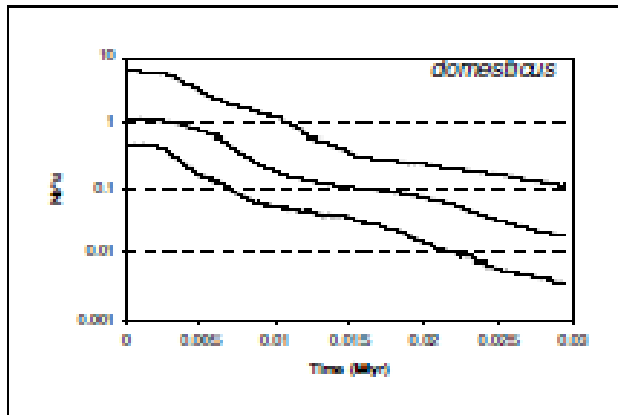
- MCMC
- srovnání modelu stabilní velikosti a modelu exponenciálního růstu/redukce populace pomocí LRT s 1 stupněm volnosti
- program Fluctuate:
- parametr růstu g
- ML i BA přístup

4. Bayesian Skyline Plot (BSP)



- rozložení genealogických linií v čase
- BSP vychází z tohoto postupu
- programy BEAST a Tracer

4. Bayesian Skyline Plot (BSP)

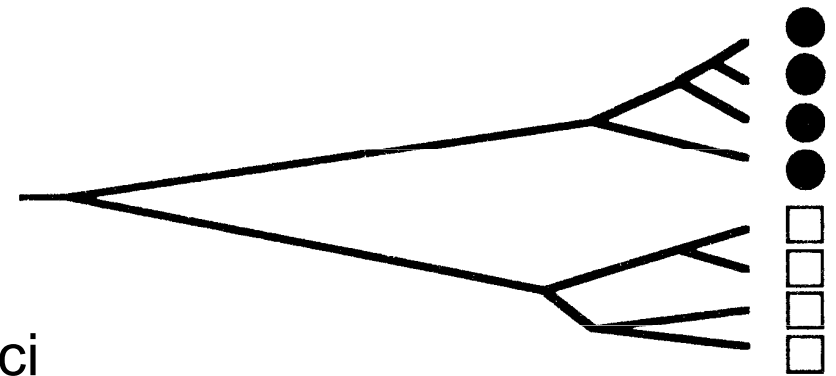


Možné výsledky fylogeografických studií

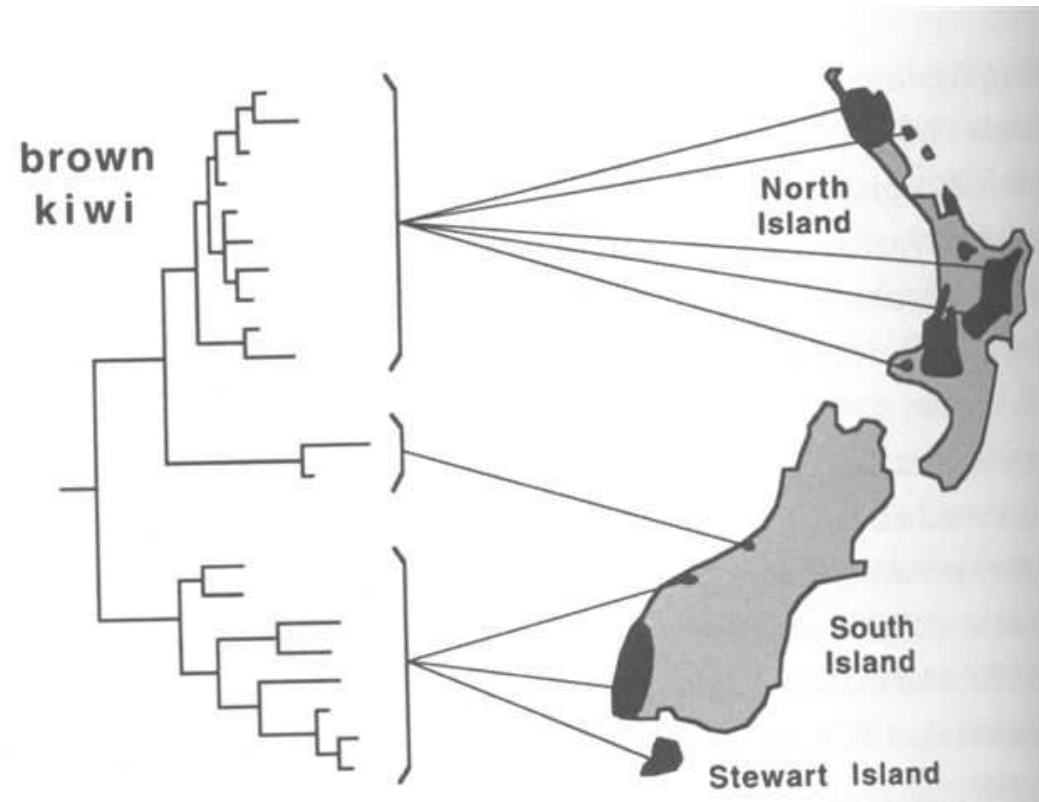
(Avice 2000)

Kategorie I:

- zřetelně oddělení alopatické linie
- bariéry toku genů nebo nízká disperze
- rozdíl díky sortování linií nebo akumulaci nových mutací

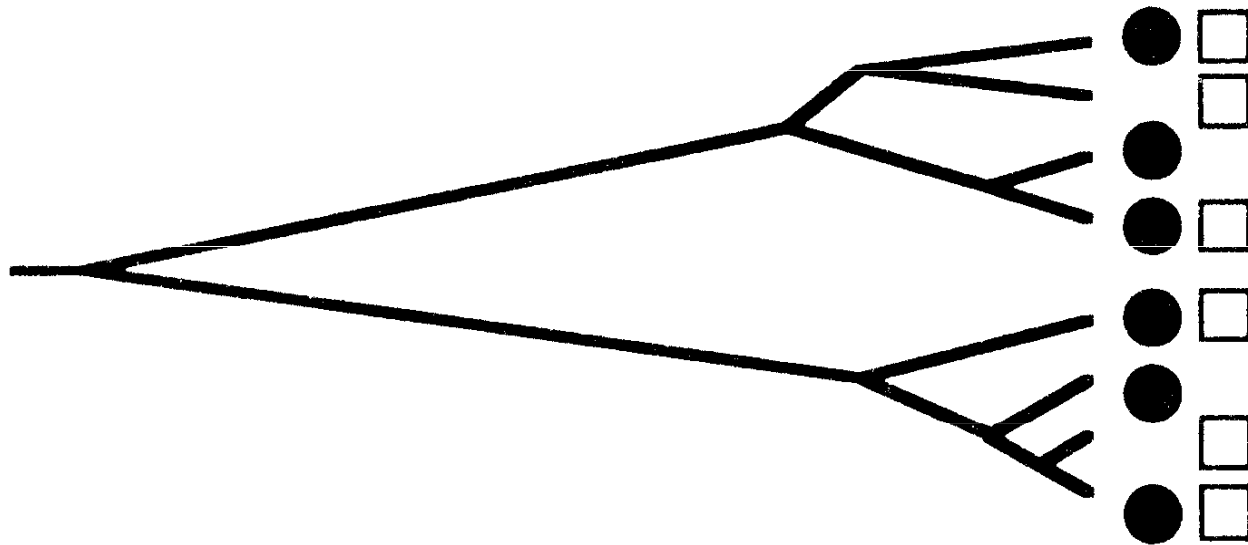


Apteryx australis



Kategorie II:

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

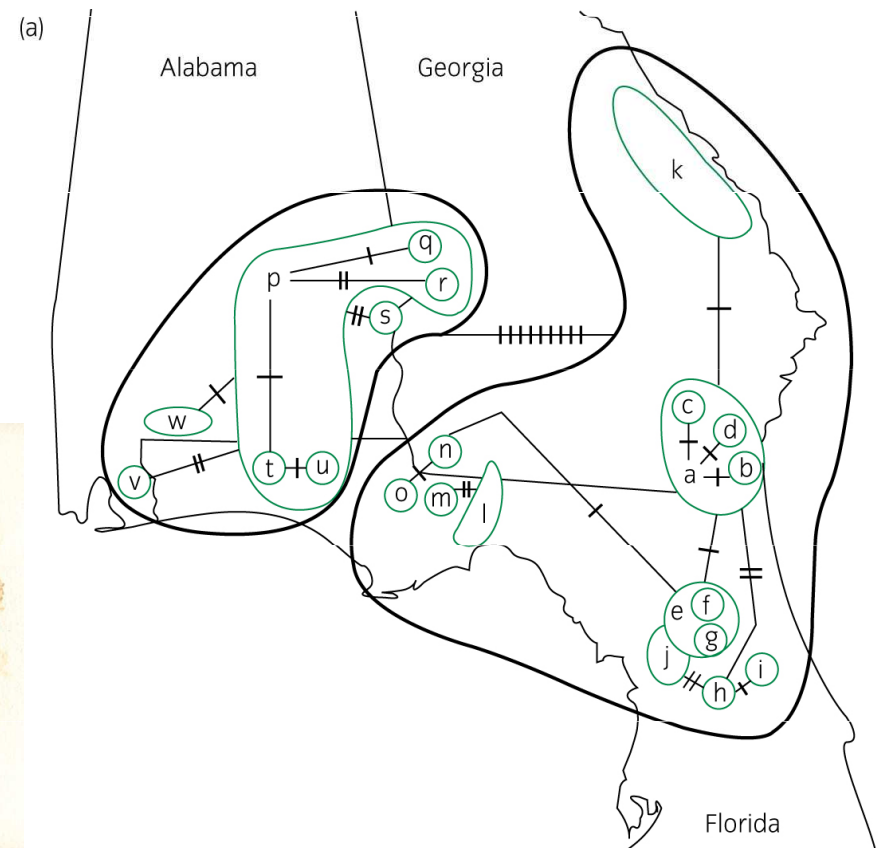


Kategorie III:

- 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

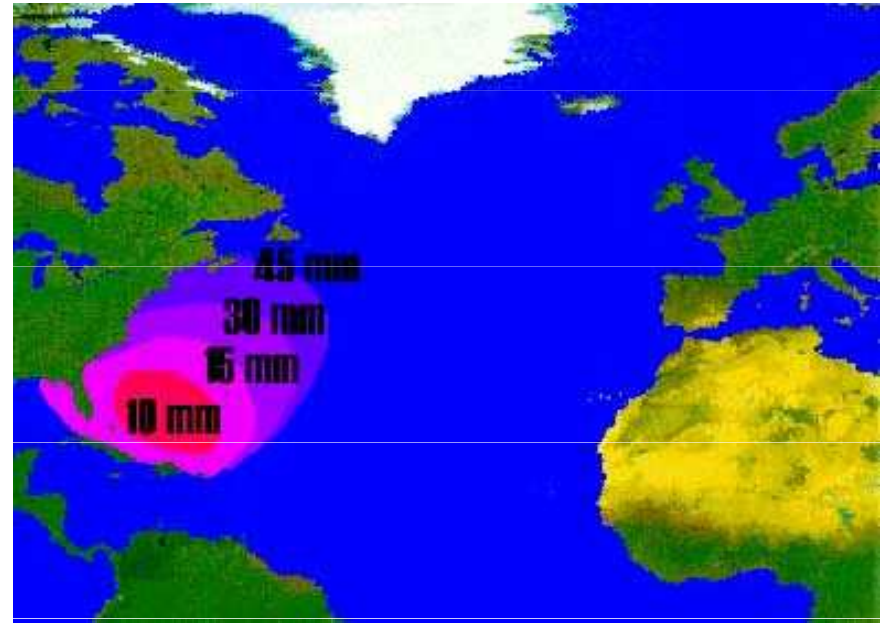
- často:
 - na větší škále Kategorie I
 - Na menší škále Kategorie III

- př.: *Geomys pinetis*



Kategorie IV:

- sympatrické, málo oddělené linie
- silný tok genů
- absence geografických bariér nebo
- recentní expanze



Anguilla rostrata

Náhodná disperze larev

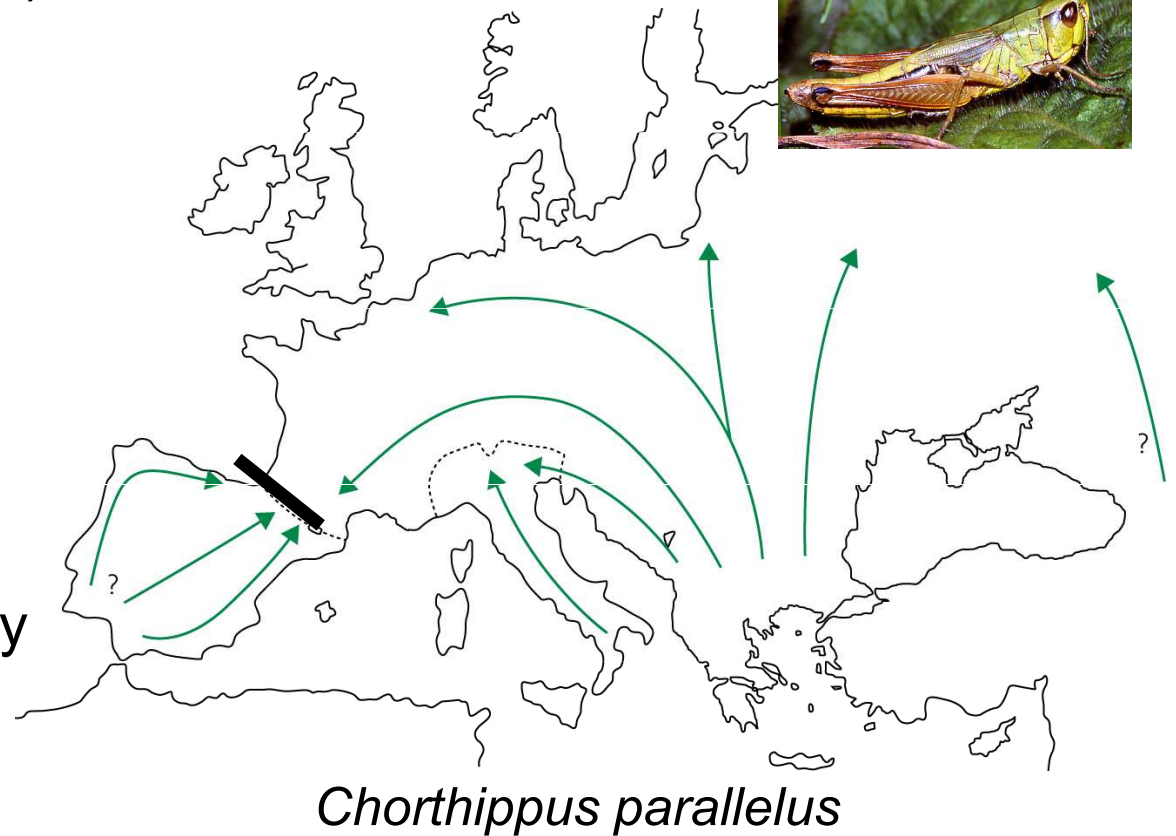
Panmiktické agregace
během tření

Kategorie V:

- kombinace III a IV
- kálo oddělené linie
- některé linie značně rozšířeny (zřejmě ancestrální), jiné (nové) geograficky omezeny
- jako vhodné znaky je nutné vybrat privátní haplotypy

Genetické důsledky ledových dob

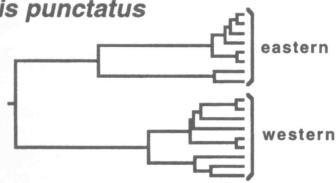
- Refugia (Iberské, Italské, Balkánské)
- V refugiích malé populace relativně dlouhou dobu
- Linneage sorting (případně mutace)
- Následné šíření → vnitrodruhové hybridní zóny
- Ale některé druhy měly i severská refugia!



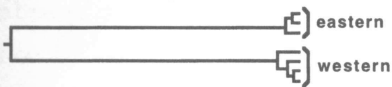
Genealogické koncordanace

Ryby JV USA

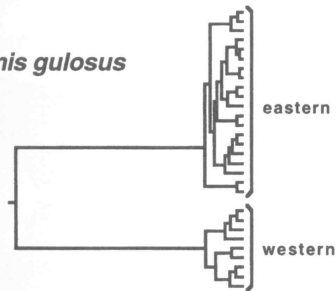
Lepomis punctatus



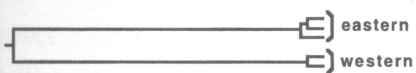
Lepomis microlophus



Lepomis gulosus

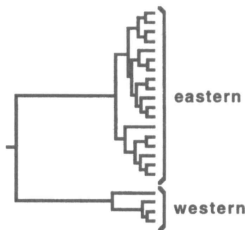


Lepomis macrochirus



8.0 6.0 4.0 2.0 0.0
sequence divergence (%)

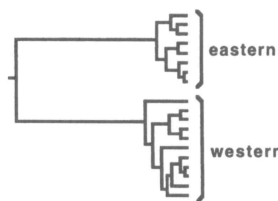
Gambusia affinis/ G. holbrooki



Amia calva

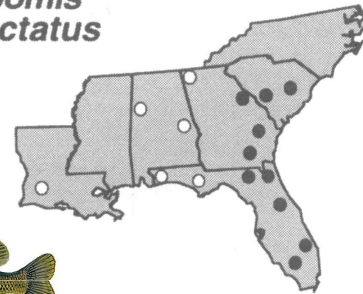


Micropterus salmoides

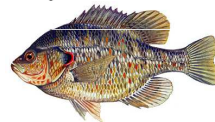
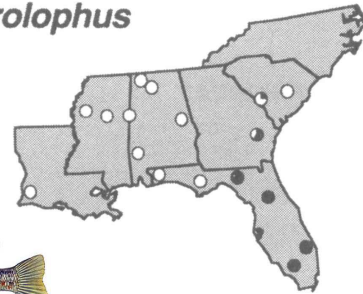


8.0 6.0 4.0 2.0 0.0
sequence divergence (%)

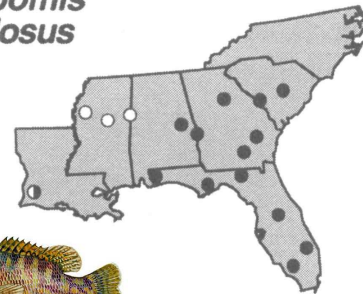
Lepomis punctatus



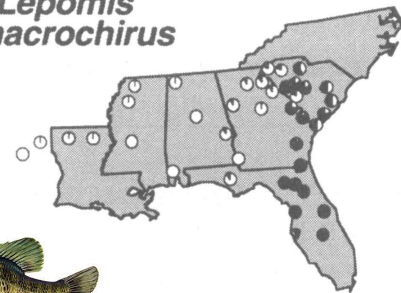
Lepomis microlophus



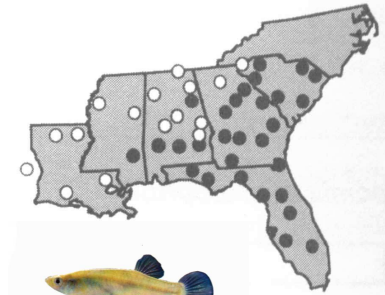
Lepomis gulosus



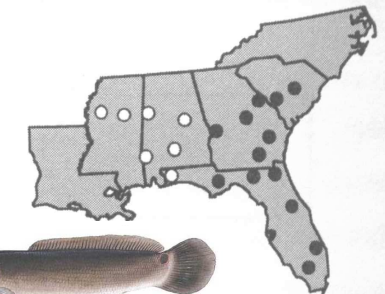
Lepomis macrochirus



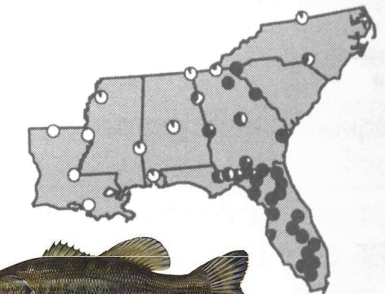
Gambusia affinis, G. holbrooki



Amia calva

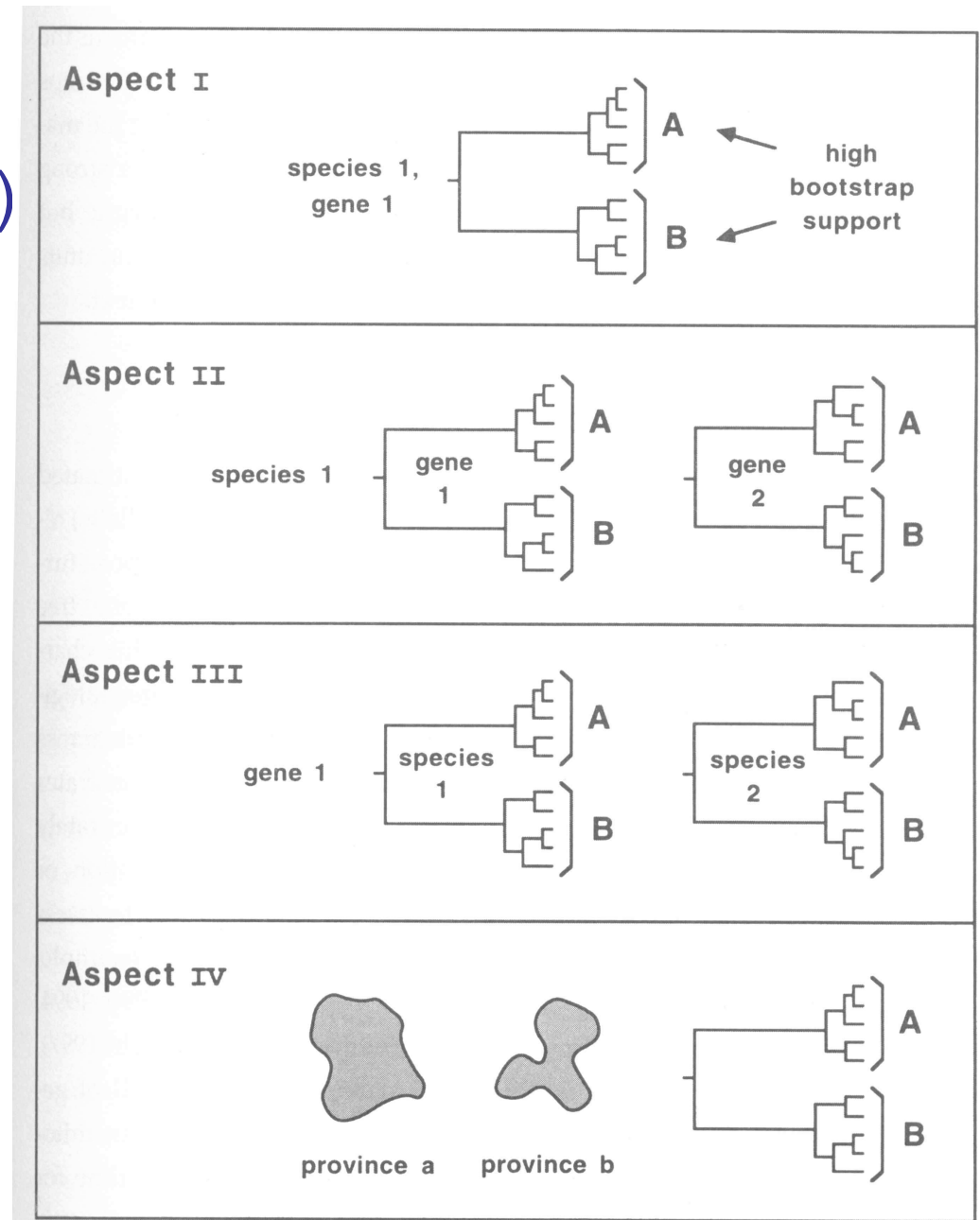


Micropterus salmoides

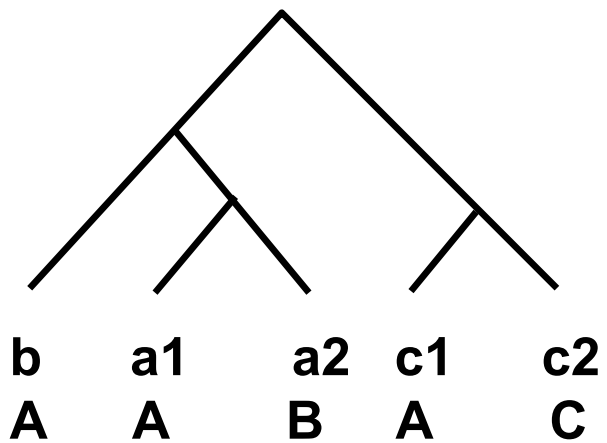
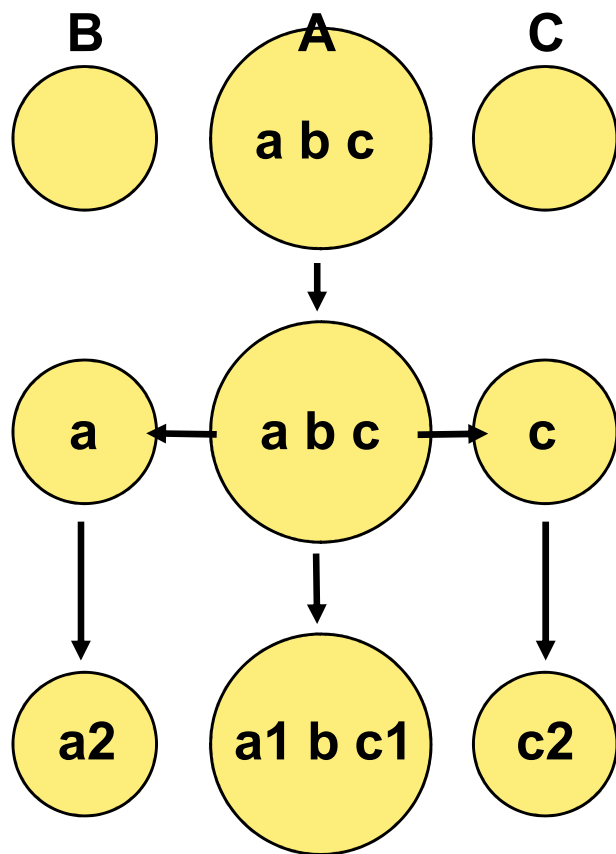


Genealogické koncordance (shody na různých úrovních)

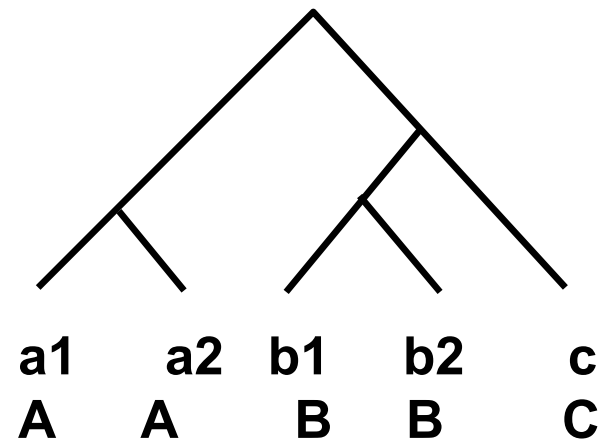
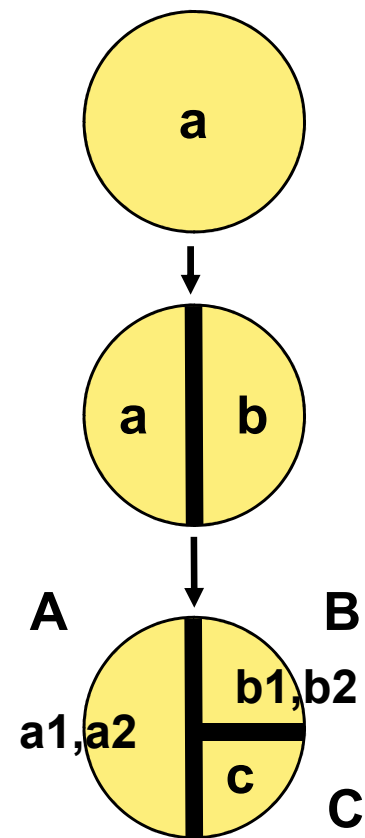
- Různé části sekvence genu
- Více sekvencí (genů)
od stejného druhu
- Více druhů ve stejné oblasti
- Podpora biogeografických
oblastí
(více druhů, více oblastí)



disperze



vikariance



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

samičí disperze a tok genů

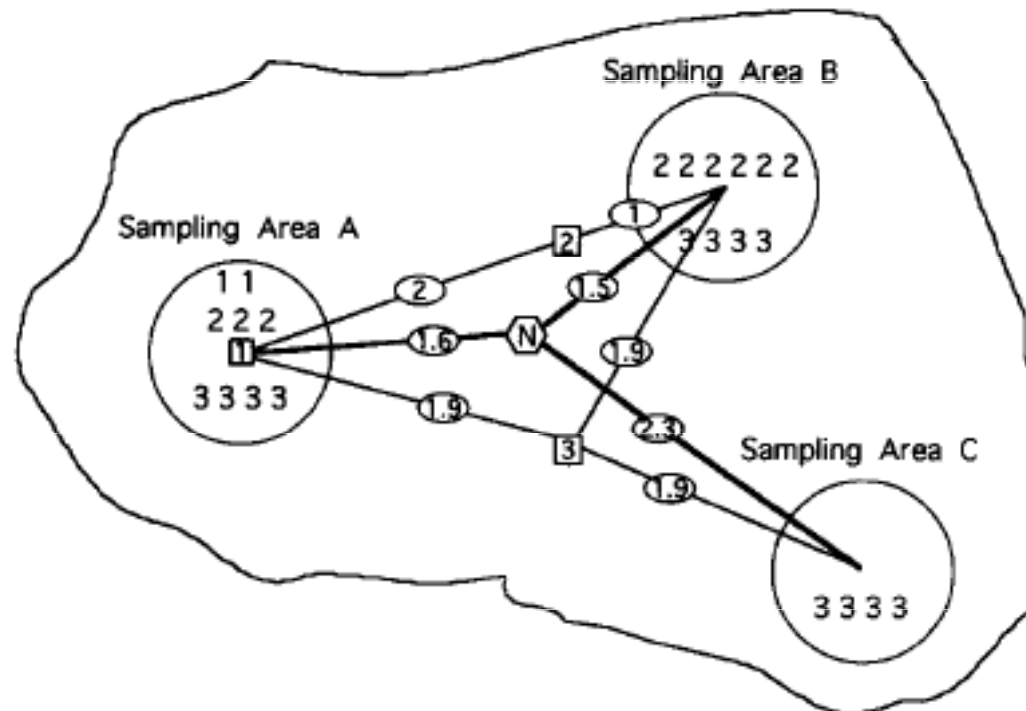
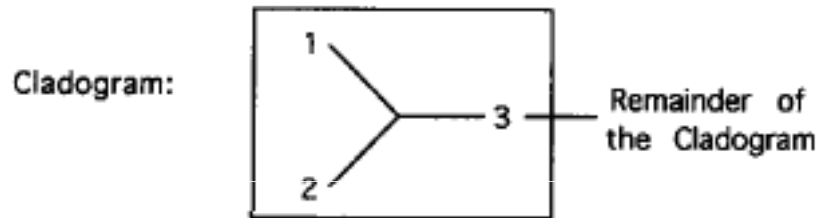
nízké -----> vysoké

samčí disperze a tok genů ↑ nízké -----> vysoké	geografická struktura v:		geografická struktura v:	
	mtDNA	ANO	mtDNA	NE
	autozomy	ano	autozomy	ano
	chr. Y	ano	chr. Y	***
	demografická autonomie	ANO	demografická autonomie	***
	geografická struktura v:		geografická struktura v:	
	mtDNA (u samic)	ANO	mtDNA	NE
	autozomy	ne	autozomy	ne
chr. Y	ne	chr. Y	ne	
demografická autonomie	ANO	demografická autonomie	NE	

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



- 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

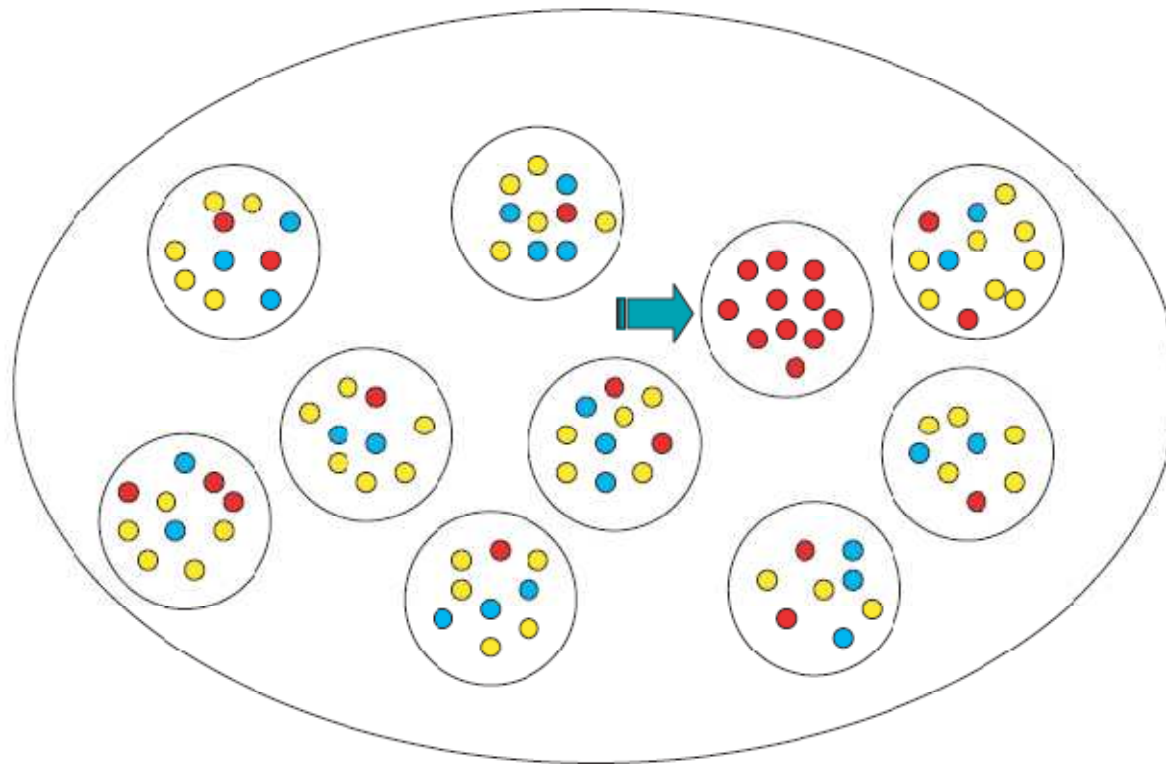
- Significantly small D_c s, primarily at the higher clade levels. The D_n s at this clade level may suddenly increase rapidly while the D_c s remain restricted, depending upon the geographical configuration of the isolates.
 - The pattern of distances described in (a) should represent a break or a reversal 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:

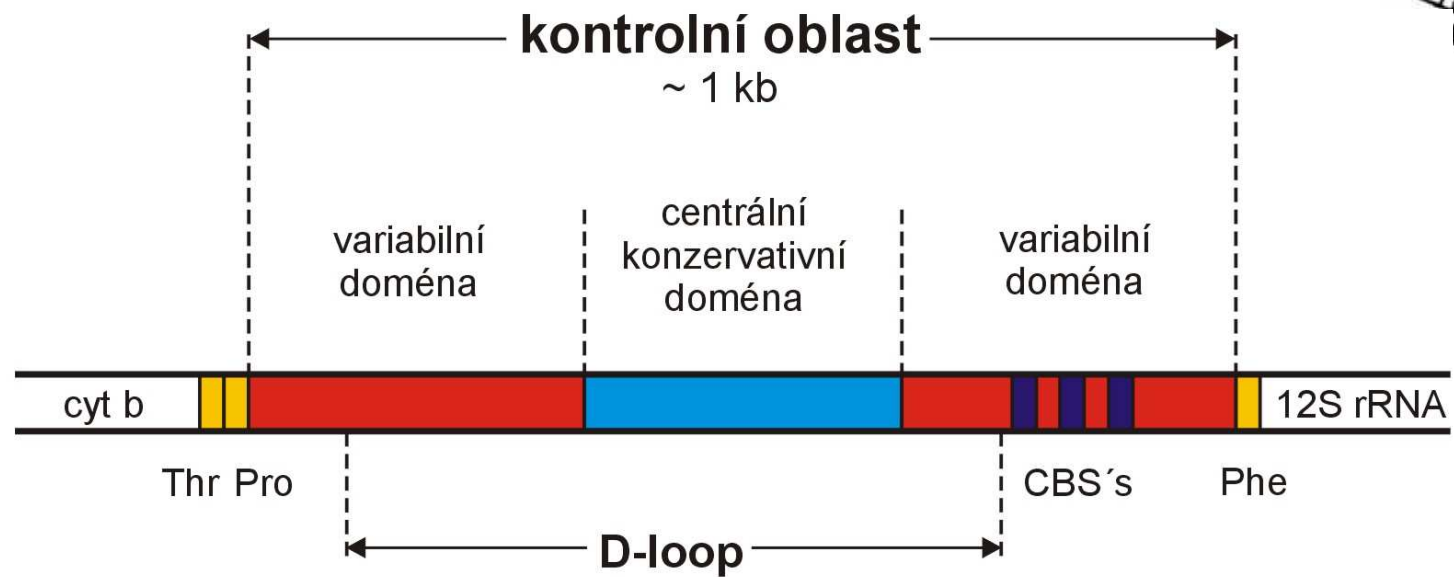
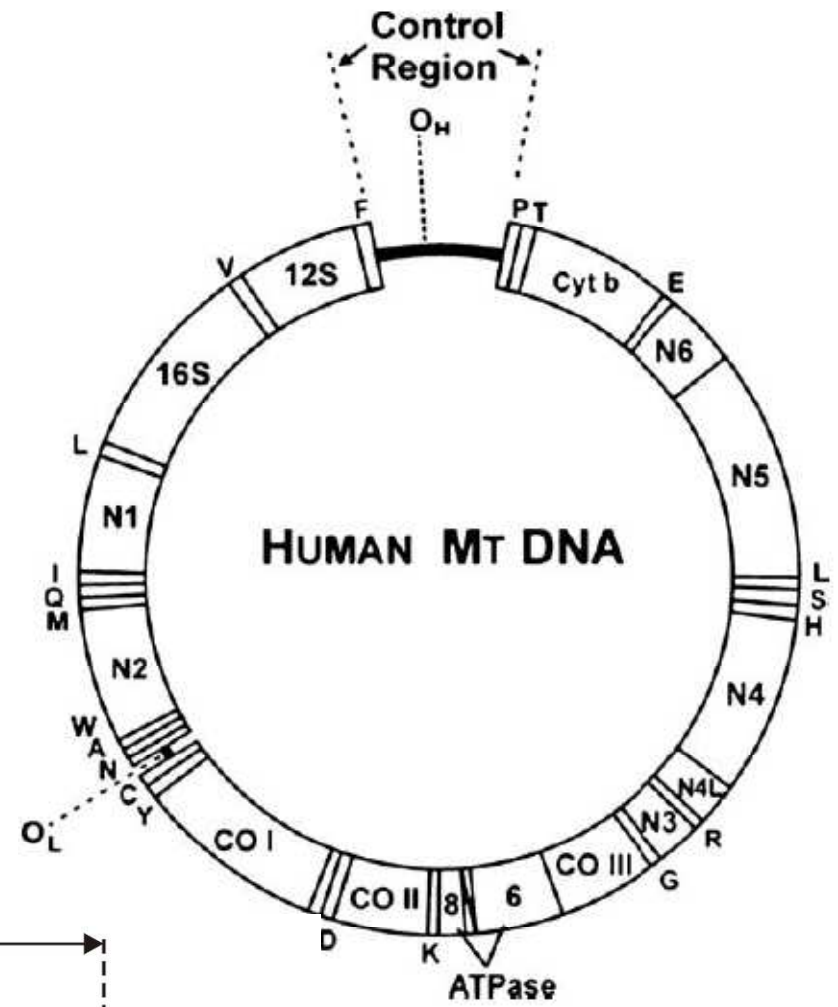
- absence evolučního modelu
- simulace: vysoké procento „false positives“ (>75%; Petit 2008)

Lokální procesy → zdánlivé expanze...



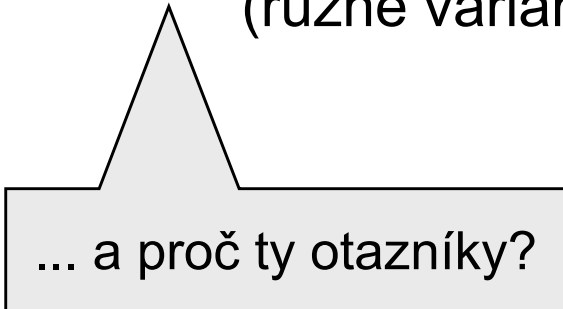
markery:

- mtDNA
- Y
- mikrosatelity



Proč je mtDNA výhodná?

- ? Malá (15-20 kb), kruhová molekula
- ? Bez intronů
- ? Minimum nekódujících oblastí
- ? Dědí se jen po jednom rodiči (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
- Mezigruhový přenos
- Nukleární pseudogeny
- Biparentální dědičnost
- Rekombinace

- Neutralita?

- vliv variant mtDNA na fitness (experimentální důkazy):

- myš (*Mus*)
- octomilka (*Drosophila*)
- člověk

- Mezidruhový přenos:

- zajíci ve Španělsku:

- existence mtDNA *Lepus timidus* u druhů *L. granatensis*, *L. castroviejo* a *L. europaeus*
- *L. timidus* však vymizel na konci posledního glaciálu
- několikanásobný přenos různých mtDNA linií

- **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 odhalit numt?
 - ultracentrifugace
 - 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*

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?