

Pokročilé metody hodnocení sekvencí DNA a multilokusových dat

3. Analýza sekvenačních dat III

- Fylogeneze na „mělké“ taxonomické úrovni, speciace, hranice druhů
- Gene trees vs. species tree
- ILS (ancestrální polymorfismus) a koalescenční teorie
- Hybridizace, polyploidizace
- haplotypové sítě a detekce ancestrálního polymorfismu [**TCS**]
- detekce hybridů/polyploidů a jejich rodičů [**Splittree**]
- Fylogeneze (=species tree) při incomplete lineage sorting [***BEAST**]
- detekce druhových hranic pomocí GMYC [**BEAST, R, Genie**]
- **praktická část – fylogenetická analýza multilokusových sekvenačních dat**

Vhodné a nepoužívanější markery

- Sekvenační data
 - cpDNA, rDNA (ITS, 18S...), nDNA (low-copy)
 - Nejlépe kombinace více lokusů z různých částí genomu
- Fingerprinting data (AFLP, SSR) – viz další přednášky



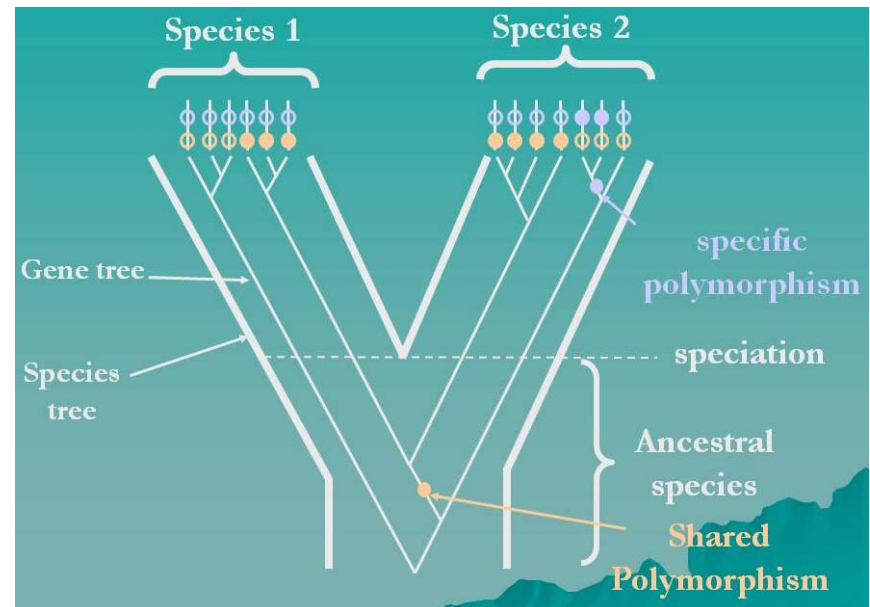
obvykle ale...

- malé rozdíly mezi sekvencemi
- nízká podpora kládů
- často v rámci „reproductive units“ – možnost introgrese
→ inkongruence mezi stromy na základě jednotlivých lokusů
- nutno odlišovat „gene trees“ a „species trees“

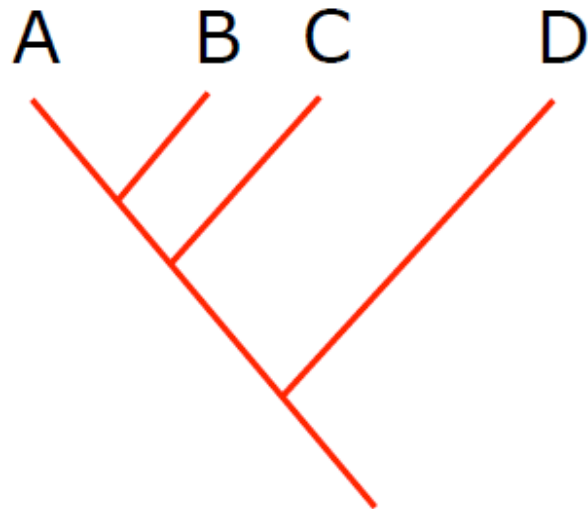
Speciace – hranice druhu

- **Speciace** – někdy rychlý, ale někdy i velmi pomalý proces (i 100-1000 mil. let)
 - Než dojde k úplné speciaci - hranice mezi druhy rozmyté a je těžké určit zda se jedná o druhy, poddruhy či variety.
- Během speciace **může docházet ke genovému toku mezi vznikajícími druhy**
 - Míra genového toku – liší se často mezi jednotlivými lokusy → **různá míra genetické diferenciace podél genomu.**
 - Ke genovému toku může docházet i po „speciaci“ (hybridizace)
- **Genový tok mezi druhy a neúplné sortování linií** → genové genealogie nemusí vždy odpovídat fylogenezi.
 - Pro rekonstrukci fylogeneze je proto potřeba použít více lokusů (gene trees).

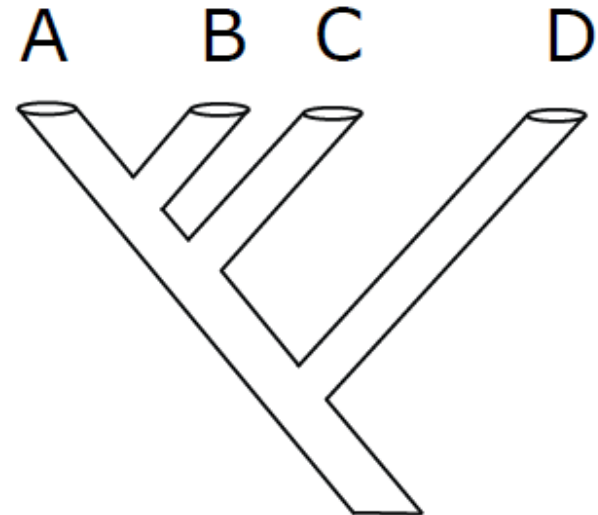
→ Koalescenční modely umožňují odhadnout dobu divergence mezi druhy či míru mezidruhového genového toku.



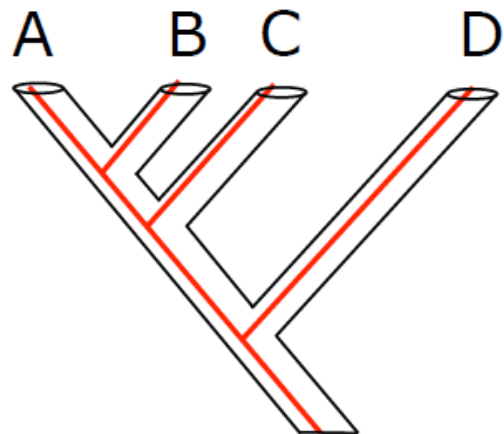
Gene trees vs. Species trees



Gene tree



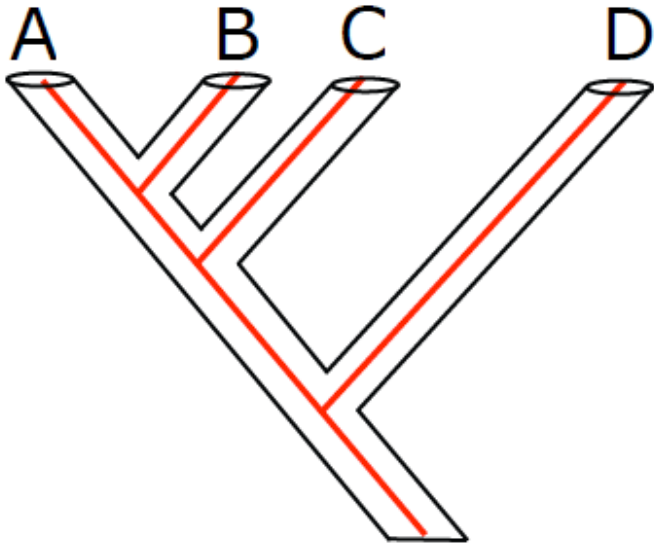
Species tree



Gene tree = Species tree

Gene trees vs. Species trees

- Při analýze jediného lokusu DNA předpokládáme, že tento lokus je dobrým odhadem skutečných vztahů mezi druhy – tj. species tree



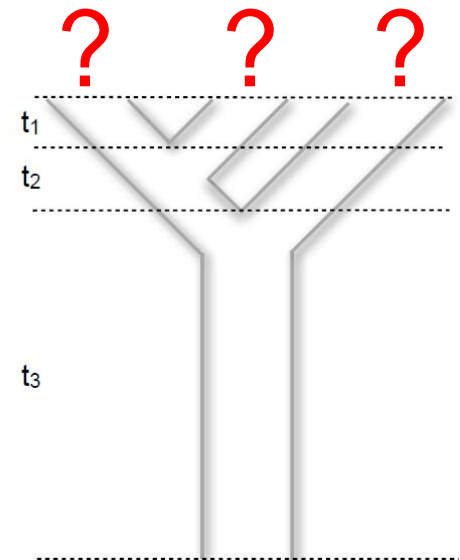
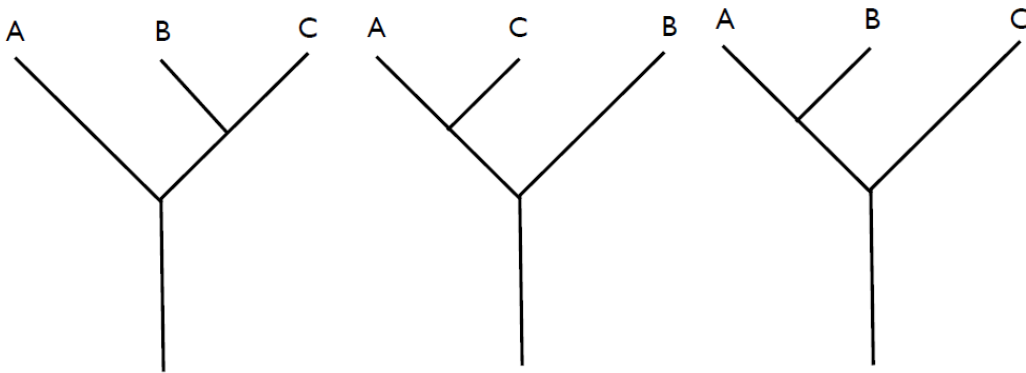
Například analyzujeme:

- **ITS, ETS, 18S...**, ale ty nejsou nezávislé
- několik lokusů **cpDNA**, ale ty se dědí společně
- často je předpoklad shody gene tree a species tree zavádějící!
- ne vždy Gene tree = species tree, spíše naopak

→ Pro odvození species tree potřebujeme více nezávislých gene trees...

Gene trees vs. Species trees

- máme několik fylogenezí na základě různých markerů („gene trees“) a
 - topologie některých z nich **si odporují**
 - topologie některých z nich **neodpovídají předchozí představě o tom, co je druh** (např. na základě morfologie, geografie apod)
- Který gene tree můžeme považovat za správný odhad species tree?
- Jak můžeme informace z jednotlivých lokusů poskládat dohromady?



Jak kombinovat data z jednotlivých lokusů

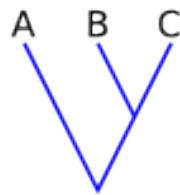
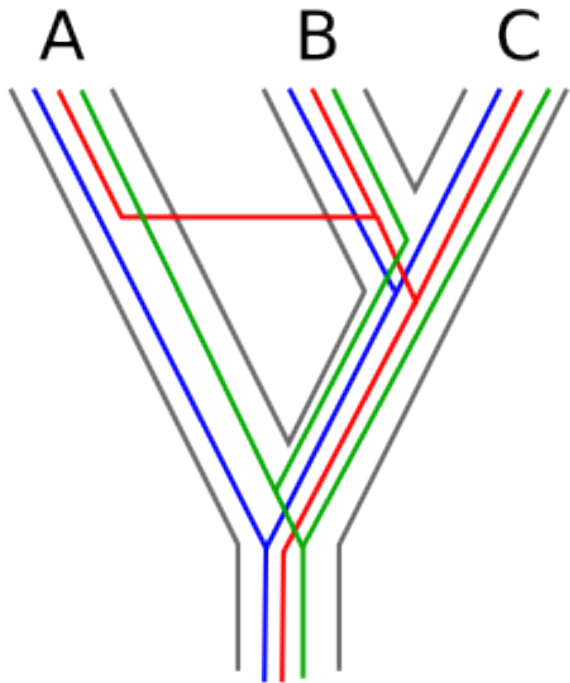
- **Konkatenace** (spojení alignmentů za sebe)
 - Předpokládá, že geny sdílejí společnou evoluční minulost - ??
 - Použitelné pokud umožníme, aby pro každý gen platily jiné parametry substitučního modelu.
 - ALE... Simulační studie ukazují, že **konkatenace vede častěji ke „špatnému“ species tree, i když je velmi dobře podpořen**
 - Kubatko & Degnan, 2007
- **Konsensuální metody na základě Maximální parsimonie**
 - MDC – minimize deep coalescence – na základě gene tree se hledá species tree, který vyžaduje nejmenší množství projevů „deep coalescence“ - nebere do úvahy pravděpodobnost těchto jevů

Jak kombinovat data z jednotlivých lokusů

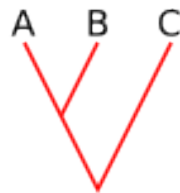
- **Multispecies coalescence na základě ML a Bayesovské analýzy**
- V současnosti nejpoužívanější, ale mají striktní předpoklady...
 - Odvození species tree na základě gene trees za předpokladu, že veškeré rozdíly mezi gene tree a species tree jsou způsobeny rozdílným průběhem koalescence
 - **Nepřipouští** - rekombinaci nebo vazbu mezi lokusy, žádný genový tok po speciaci...(tj. žádná hybridizace)
- **STEM** – ML analýza
 - vyžaduje gene trees, velikosti populací, mutační rychlost lokusů (Kubatko et al, 2009)
- ***BEAST** – Bayesovská analýza - odhaduje species tree, časy divergence gene trees, efektivní velikost populace
 - jako vstup alignemty na základě více lokusů
 - potřeba více jedinců z populace/druhu

Proč topologie gene tree neodpovídají species tree?

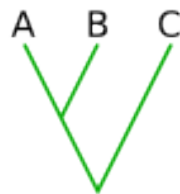
- Ancestrální polymorfismus, tj. neúplné vysortování linií pro některý z analyzovaných lokusů (tzv. **incomplete lineage sorting**)
- Genový tok mezi druhy (**hybridizace**, introgrese)
- **Genová duplikace a následná ztráta** genu (i na diploidní úrovni)
- Polyploidizace



← Gene tree = species tree



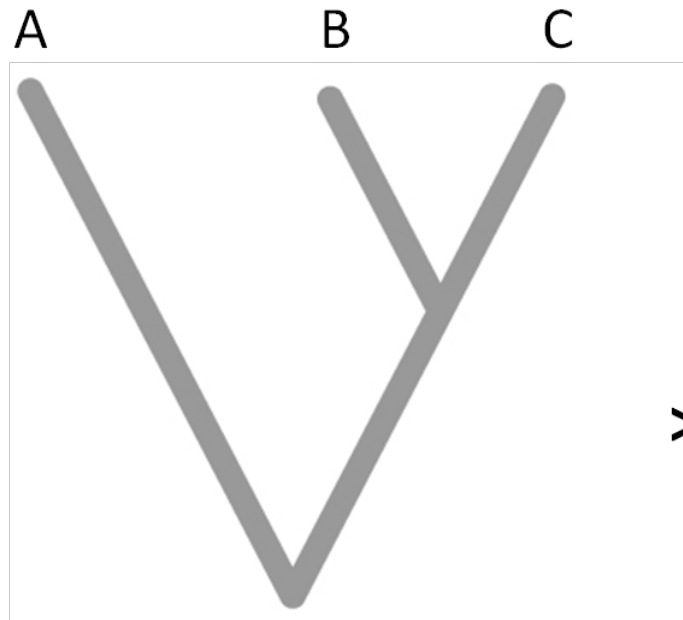
← Gene flow/hybridizace



← Deep coalescence

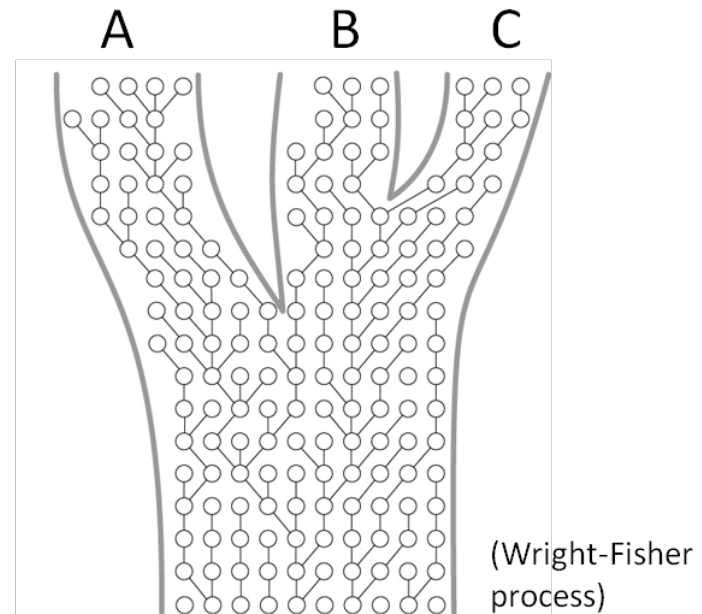
ILS – incomplete lineage sorting

- O procesech na úrovni blízce příbuzných druhů je vhodné přemýšlet jako o procesech na populační úrovni
 - Využívají se teorie populační genetiky, např. koalescenční teorie
 - Druhy jsou reprezentovány množstvím vývojových linií, které v průběhu evoluce vznikají (např. mutacemi, rekombinacemi) zanikají, nebo přežívají



Species phylogeny

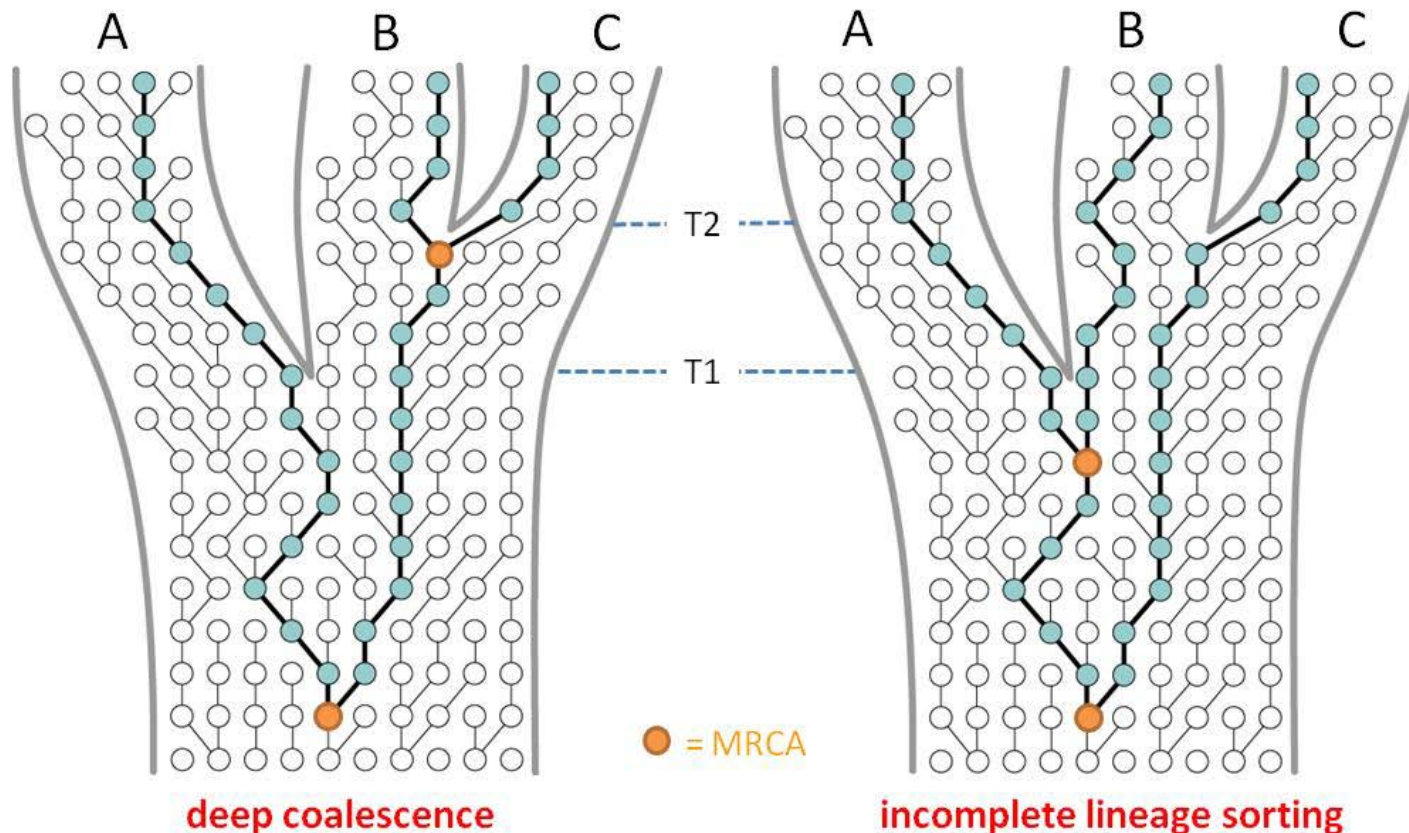
>>



Population genetics:
coalescence process

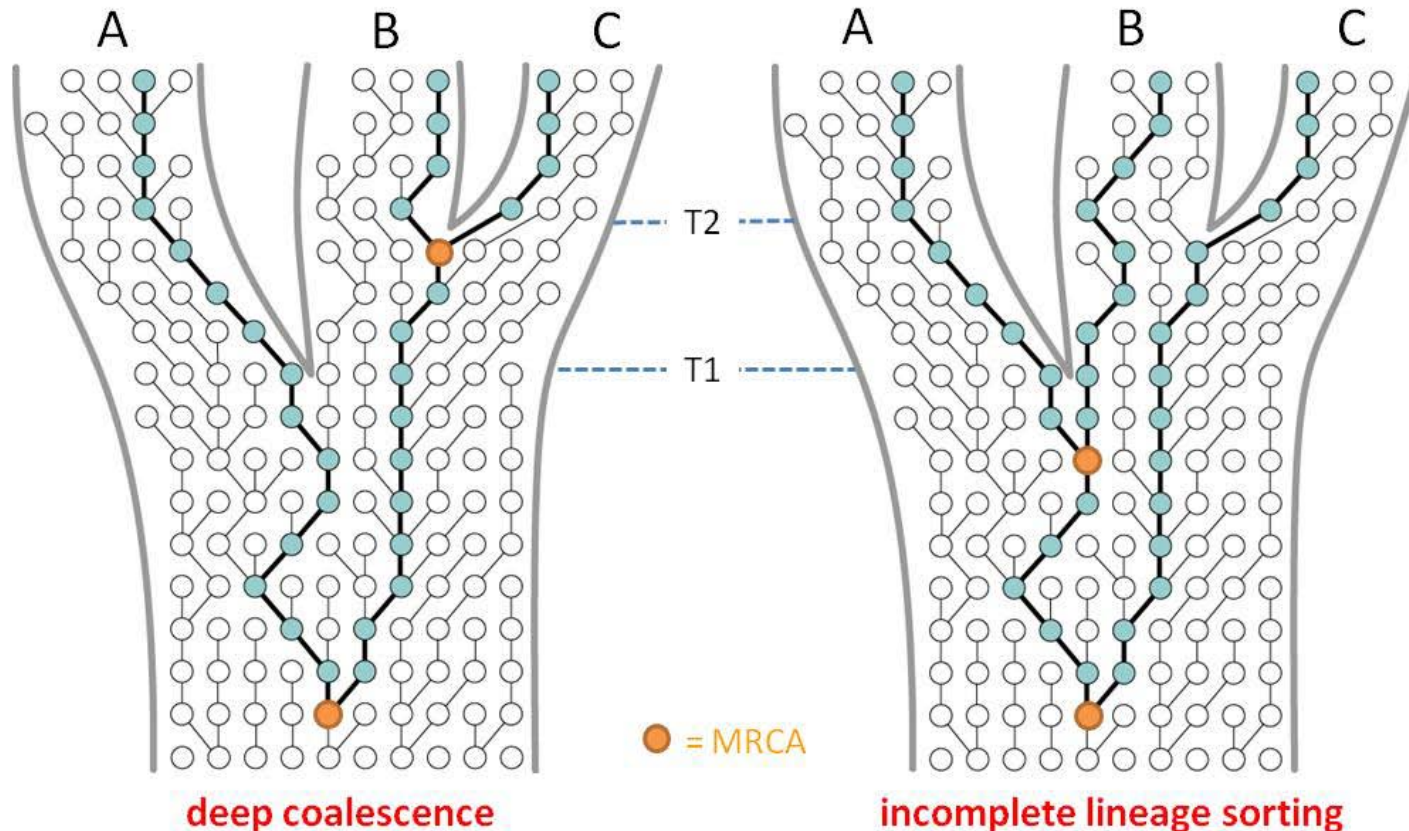
ILS – incomplete lineage sorting

- Na mělké taxonomické úrovni se může stát, že analyzujeme
 - Jedince z populace **B**, který nese alelu podobnou jedinci z populace **A**, i když **B** je sesterskou populací **C** – tzv. **ancestrální polymorfismus**
 - Pravděpodobnost, že najdeme takovou příbuznost se odhaduje pomocí **koalescenčního modelu** (závisí na velikosti populace N)



ILS – incomplete lineage sorting

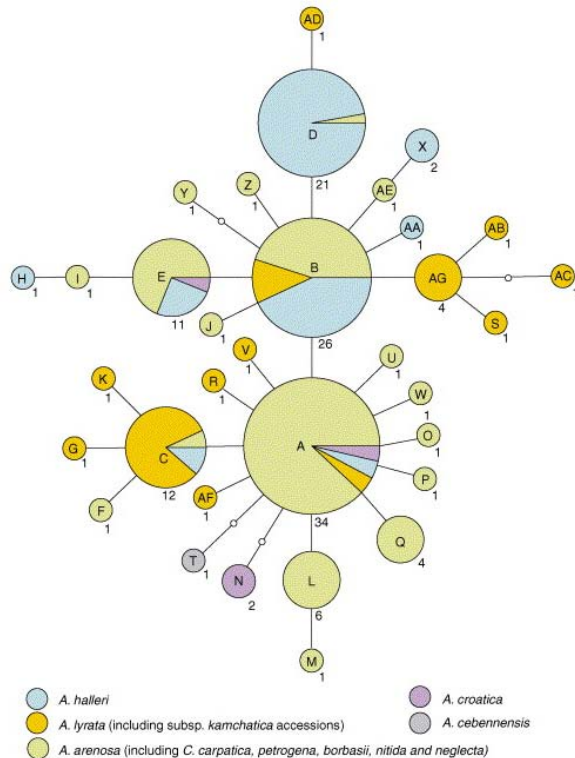
- Koalescenční model nám ukáže, zda by za určitých podmínek měly být linie daného lokusu už vysortované mezi druhy nebo ne...
- dvě náhodné kopie genu budou sdílet MRCA průměrně za $2N$ generací
- za $4N$ generací by měly všechny linie splynout (kompletní lineage sorting)



ILS \approx ancestrální polymorfismus v praxi

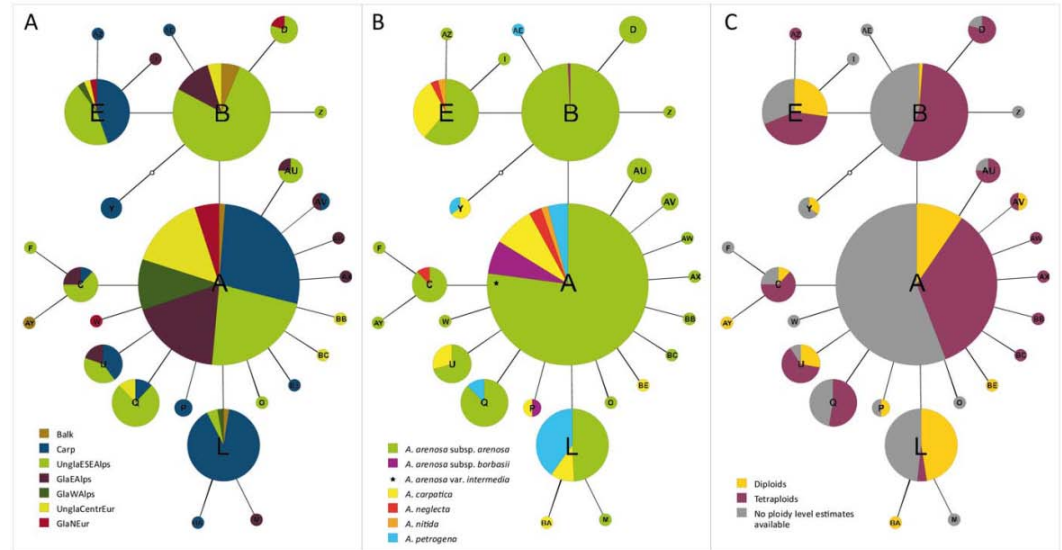
- např. cpDNA pattern, které nereflektuje hranice druhu ani geografii

→ Může být ancestrální polymorfismus, ale i důsledek genového toku/hybridizace!



TRENDS in Plant Science

Clauss and Koch, 2006; vnitrorodová úroveň – spíše hybridizace?



Chloroplast DNA trnL/F suprahaplotype networks of the *Arabidopsis arenosa* species complex (Schmickl et al., 2012) – vztahy v rámci polyploidního komplexu – ILS vs. gene flow....

Rekonstrukce species tree při ILS

- Pokud jsme si jisti, že nedochází k hybridizaci – můžeme odvodit species tree na základě gene trees pomocí **koalescenčního modelu**...

→ veškerá inkongruence je vysvětlena jako důsledek ILS

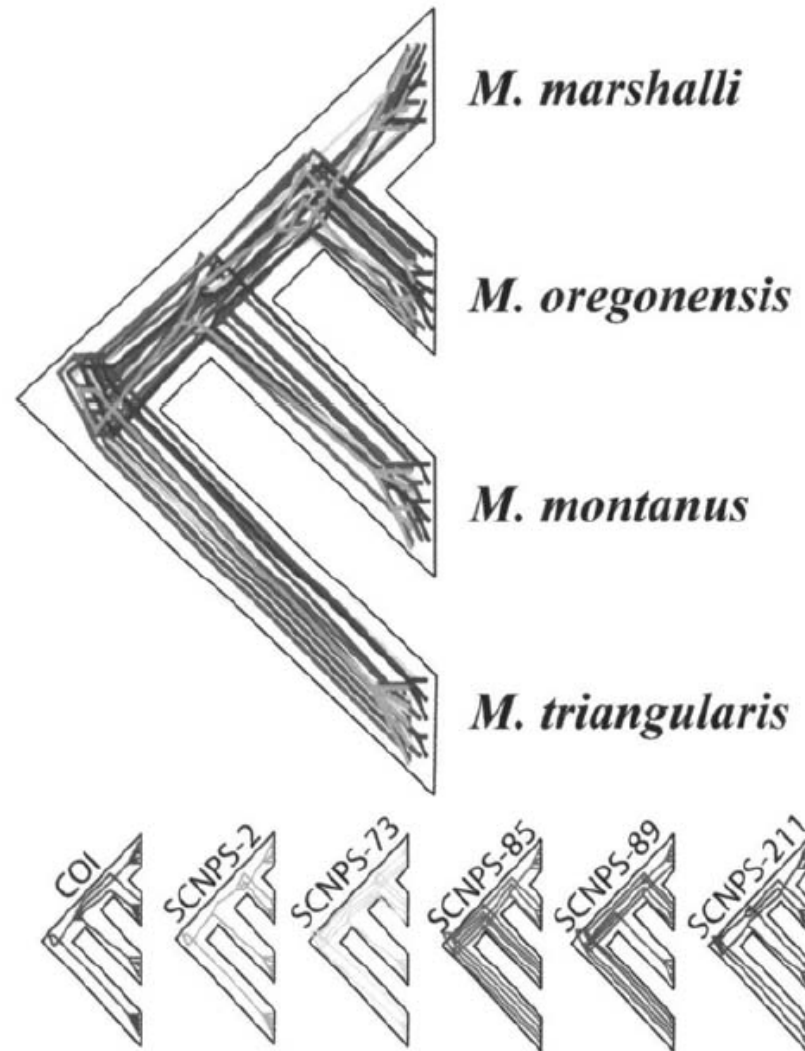


FIGURE 3. Estimate of the species phylogeny (ESP) with the highest likelihood score ($-\ln L = -10,487.1938$) calculated with the coalescent-based approach from the probabilities of the gene trees (see Fig. 2), with each gene tree shown within the species tree.

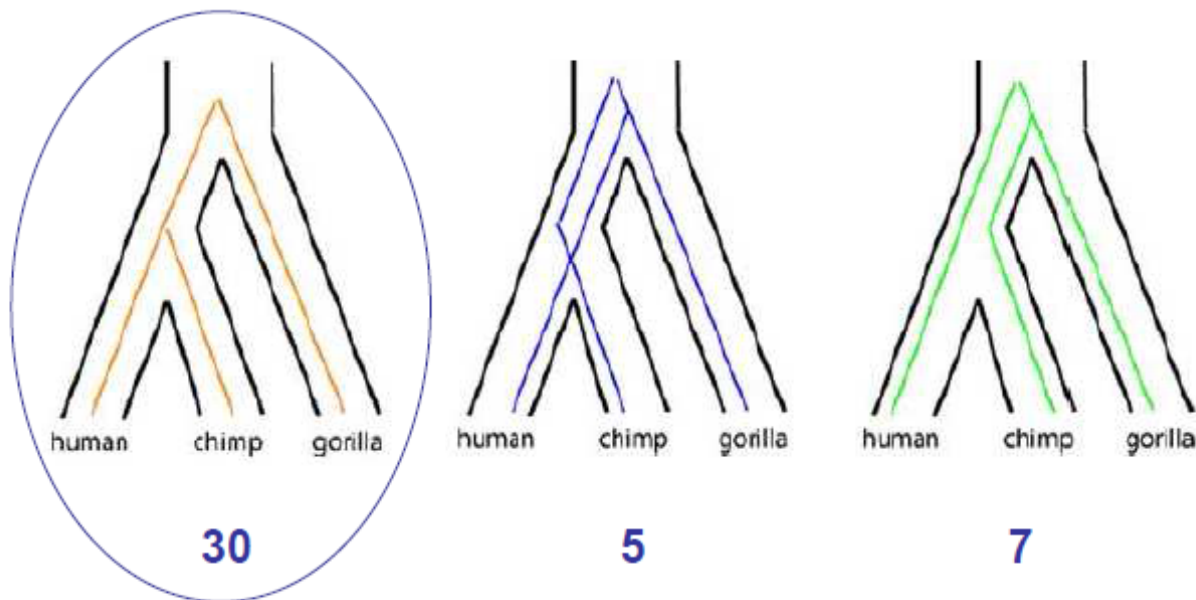
Programy

- BEAST/*BEAST
- BEST
- STEM

Většinové pravidlo (?)

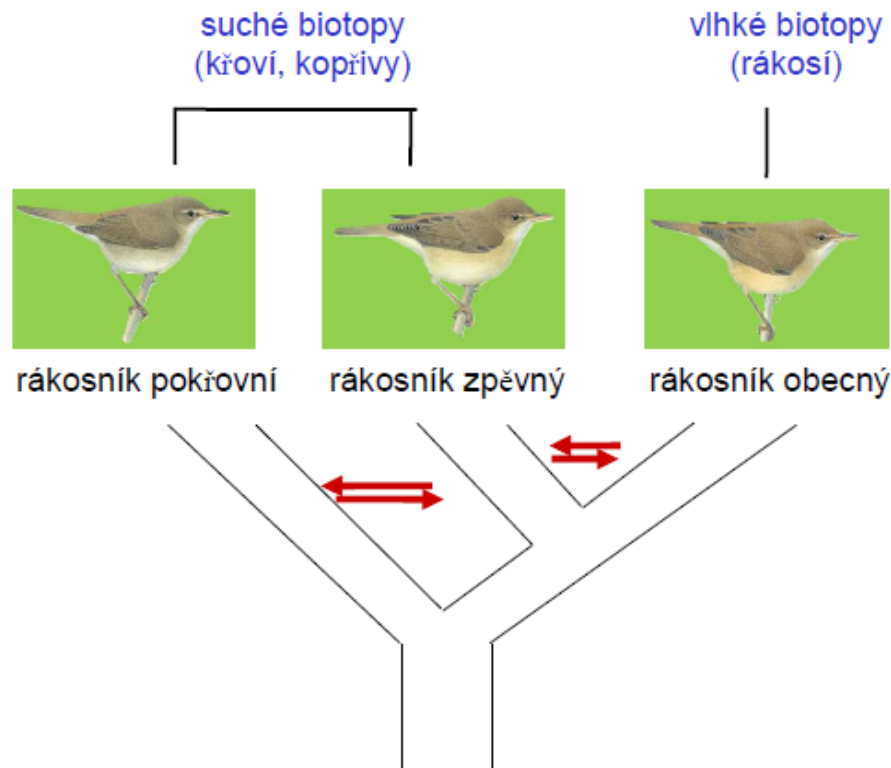
U blízce příbuzných druhů, které stále nesou ancestrální polymorfismus, tj. u kterých stále nedošlo k úplnému sortování linií, je třeba pro rekonstrukci fylogeneze použít víc genů.

Správná je ta fylogeneze, která je podpořena většinou lokusů.



Genový tok a hybridizace

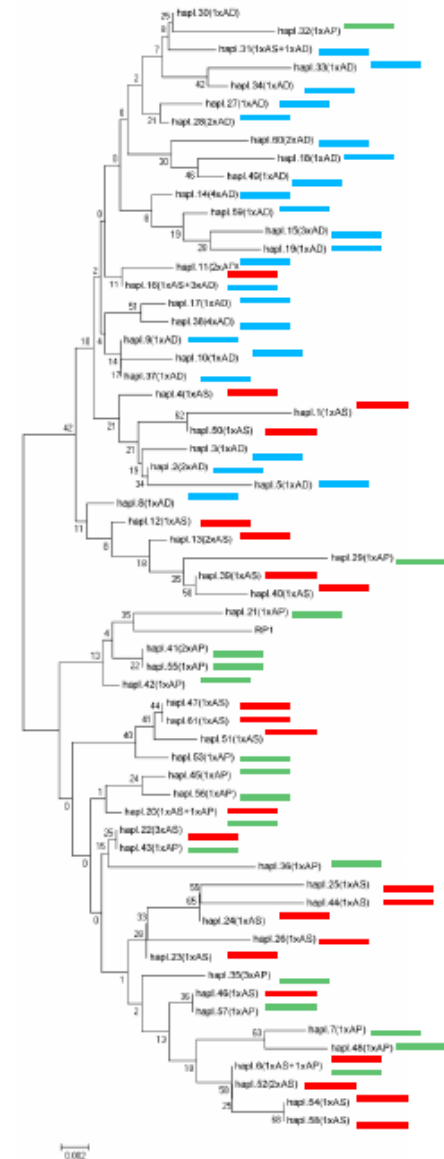
- Co když není mezi druhy ještě vytvořena reprodukční bariéra a dochází ke genovému toku?
→ další problém pro rekonstrukci fylogeneze u blízce příbuzných druhů.
- Genový tok nemusí probíhat pouze mezi sesterskými druhy (hybridizace)
- I když ke genovému toku nedochází dnes, mohlo k němu docházet v minulosti.



ILS vs. hybridizace

- Odlišení vlivu ILS a hybridizace je zapeklitý problém!
- Multispecies coalescent přístup genový tok ani hybridizaci nepřipouští
- Pro omezený počet druhů/populací lze vliv genového toku na speciaci a sortování linií modelovat...

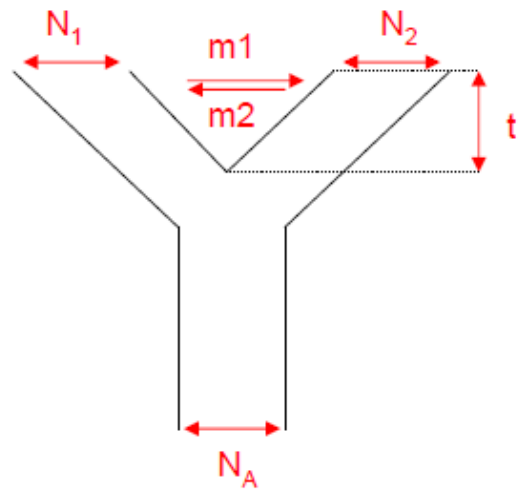
r. obecný —
r. zpěvný —
r. pokřovní —



Jak odlišit vliv hybridizace od ILS

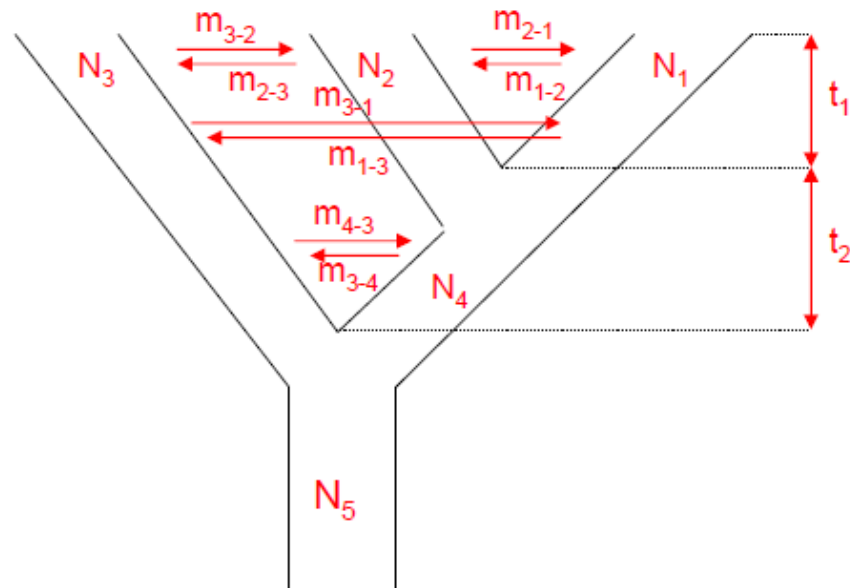
- Koalescenční modely a odhad míry genového toku mezi druhy

Model Izolace s Migrací (IM)
pro dvě populace



6 parametrů

Model Izolace s Migrací (IM)
pro tři populace



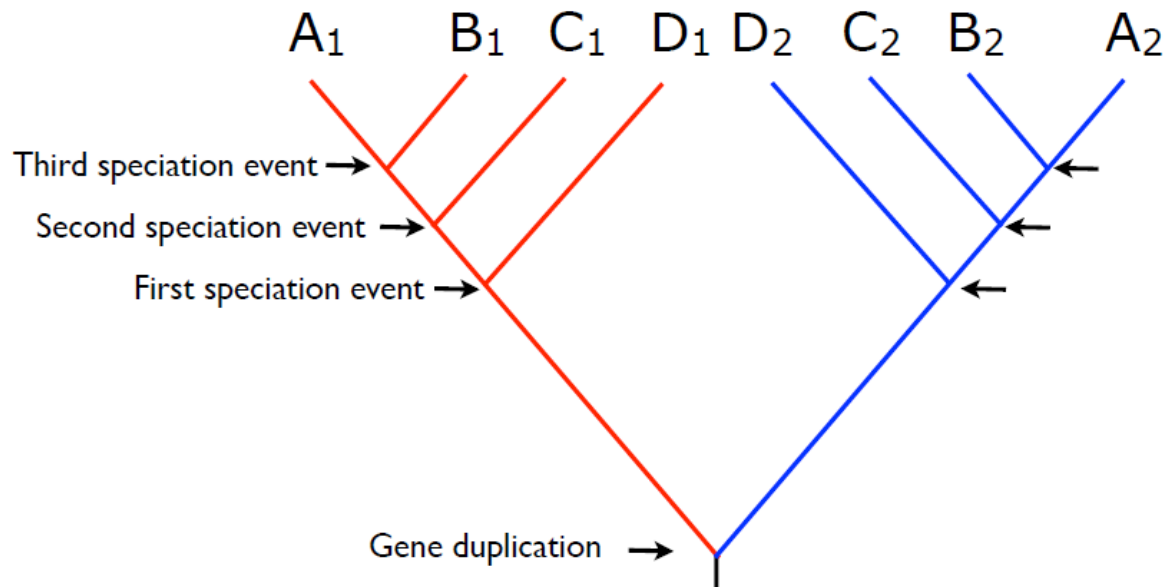
15 parametrů

Hey J, Nielsen R (2004). *Genetics* 167(2):747-60.

Hey J (2010). *Mol.Biol.Evol.* 27(4):905-920.

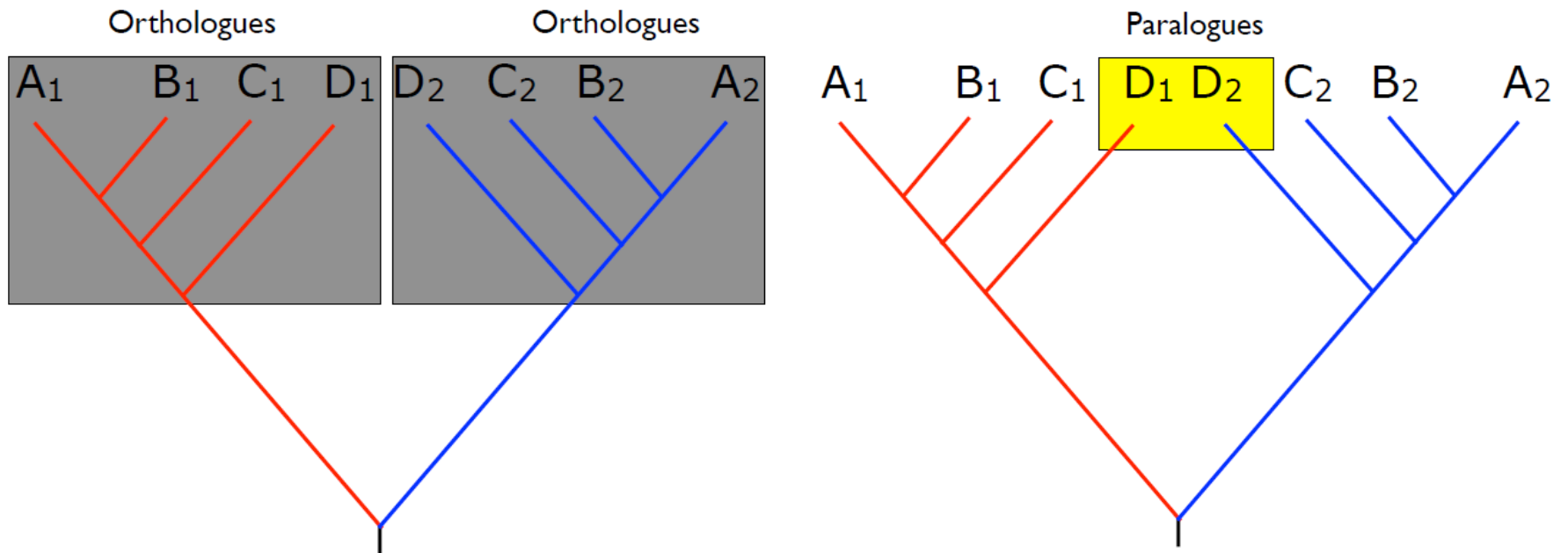
Genová duplikace a ztráta

- K duplikacím genů dochází poměrně často (ca 0.01/gen/milion let)
- Nové kopie genů mohou např.
 - získat novou funkci (neofunkcionalizace)
 - pseudogenizovat (postupná degradace genu)
 - nebo mohou být úplně ztraceny...
- Po duplikaci se každá kopie vyvíjí nezávisle, představuje jiný gene tree
- Zásada – pracovat vždy s orthology



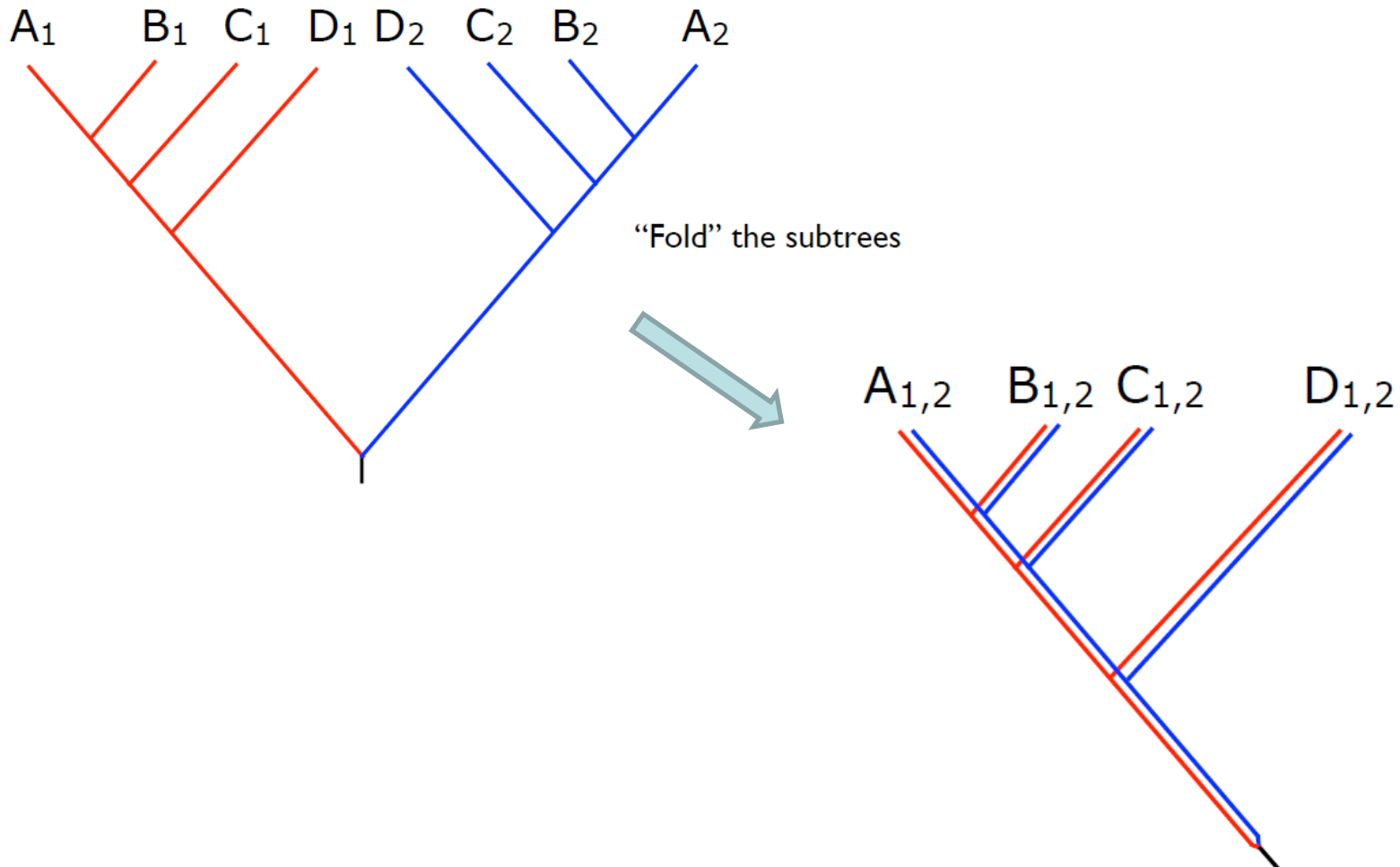
Genová duplikace a ztráta

Ortology vs. paralogy



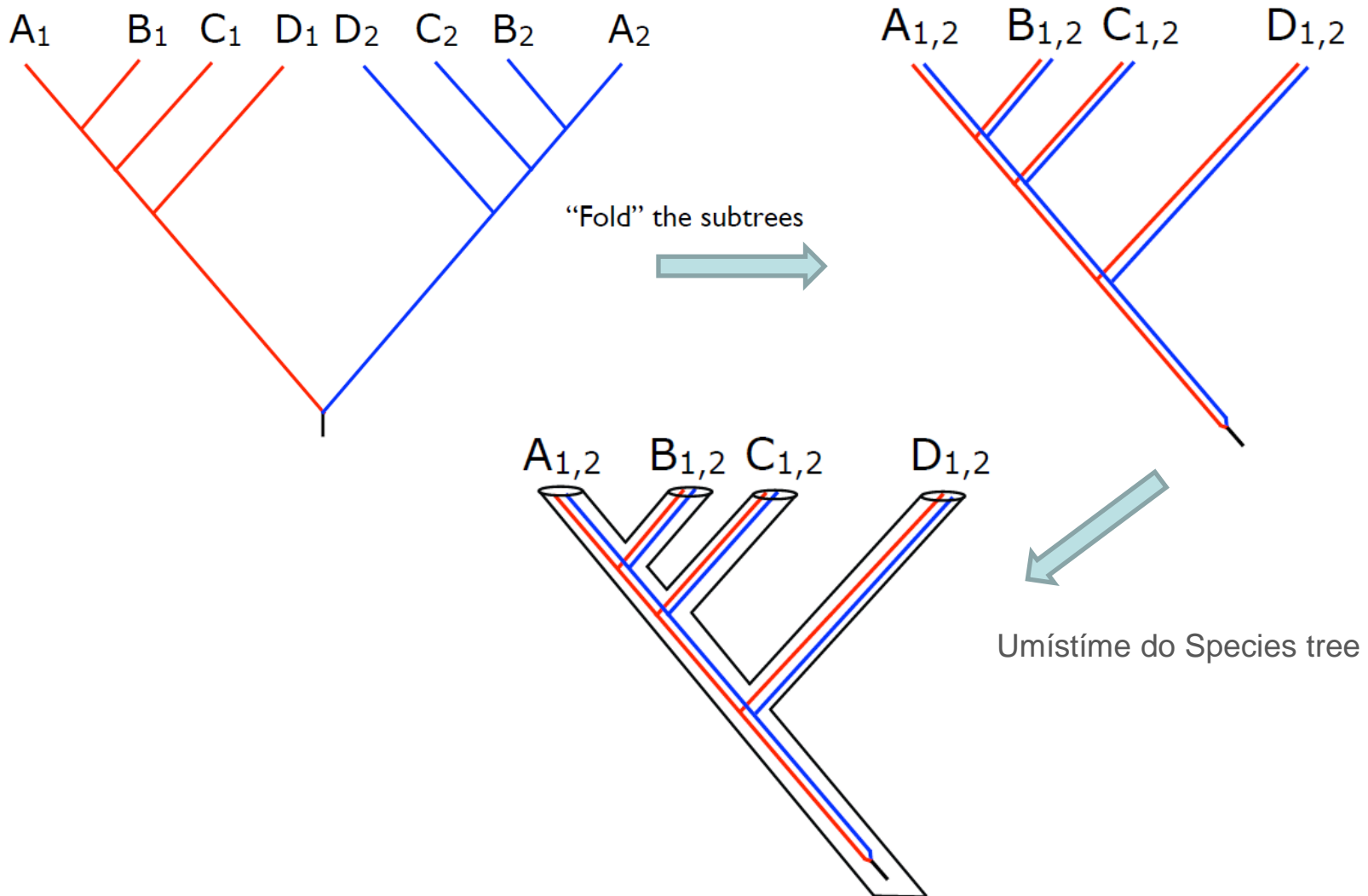
Genová duplikace a ztráta

Napasování duplikovaného gene tree do species tree



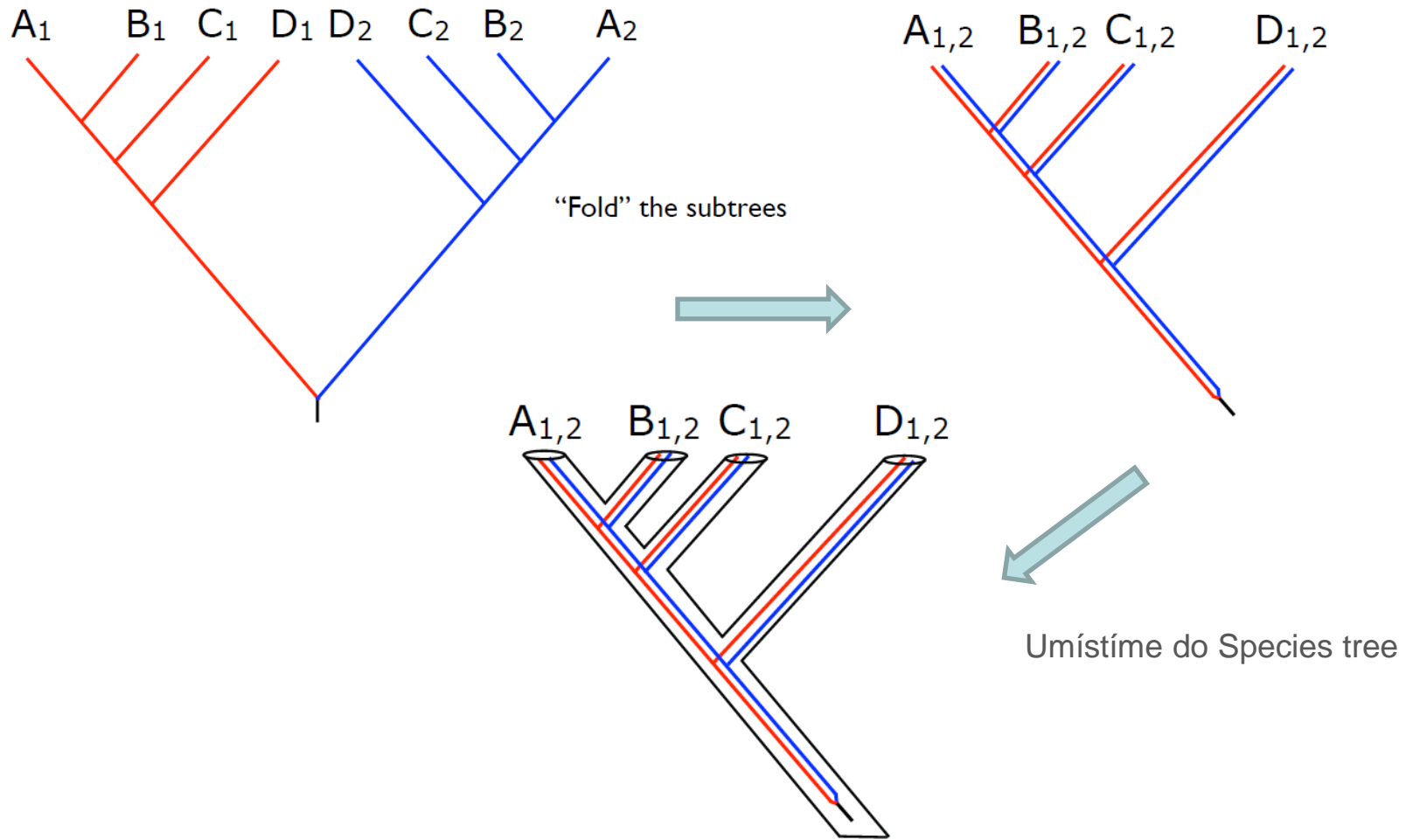
Genová duplikace a ztráta

Napasování duplikovaného gene tree do species tree



Genová duplikace a ztráta

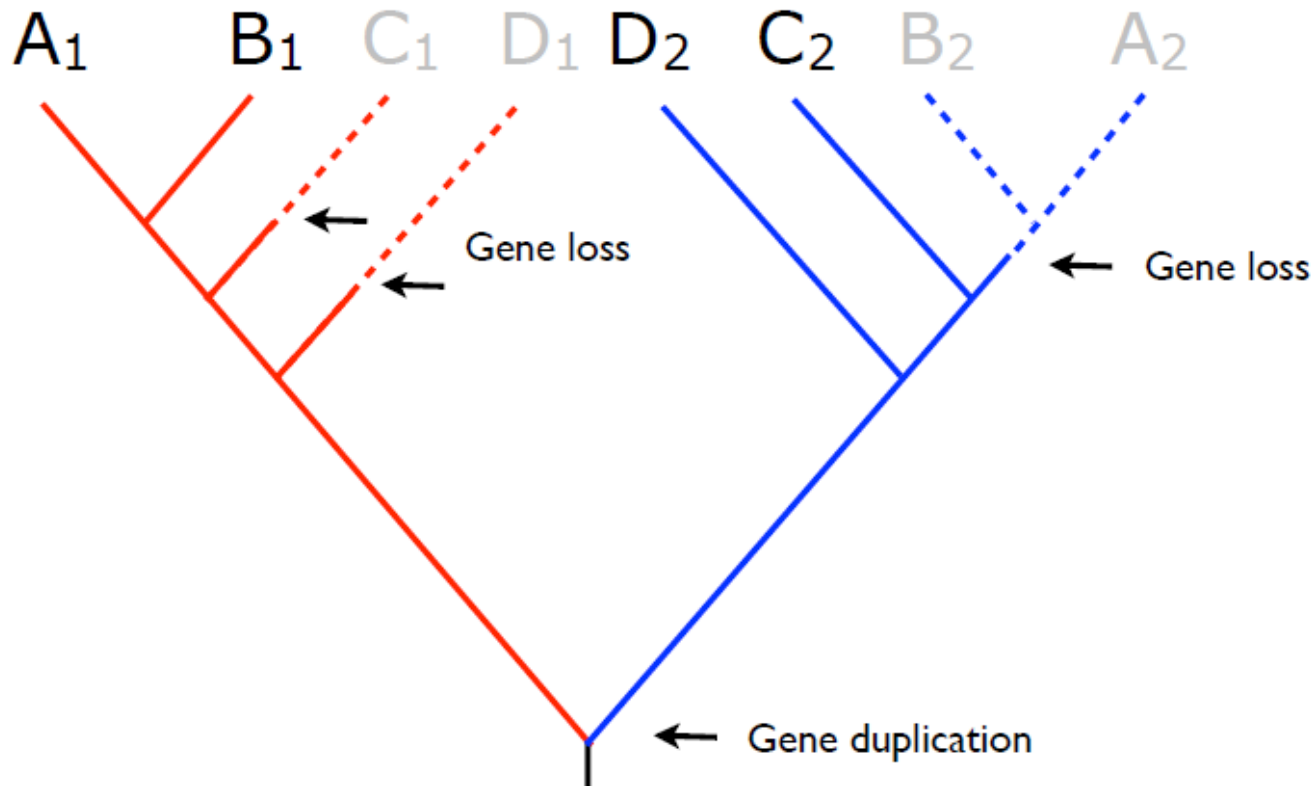
Napasování duplikovaného gene tree do species tree



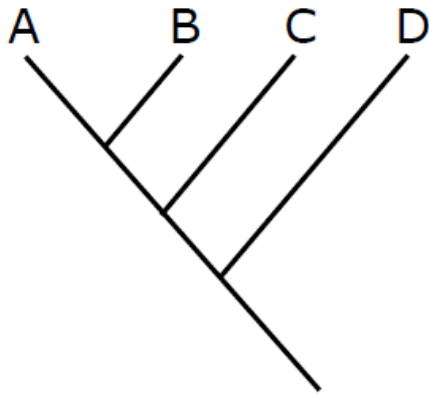
...pokud máme všechny kopie genu z každého jedince, není problém

Genová duplikace a ztráta

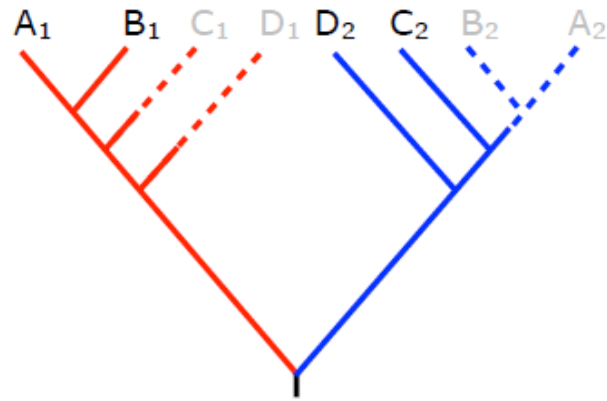
Problém nastává, pokud jsme nenašli všechny kopie, nebo některé prostě zmizely v průběhu evoluce...



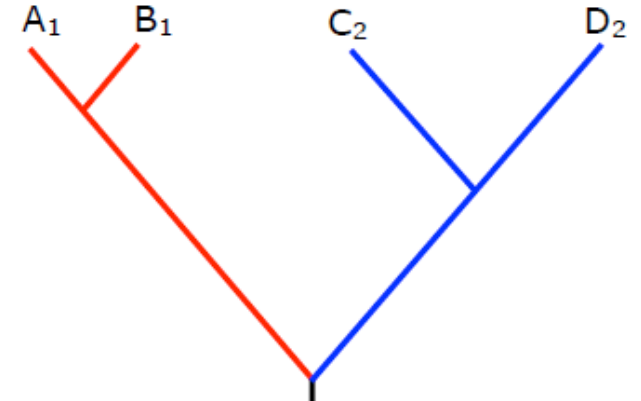
Genová duplikace a ztráta



Species tree



Náš gene tree v evoluci



Náš gene tree, jak ho vidíme

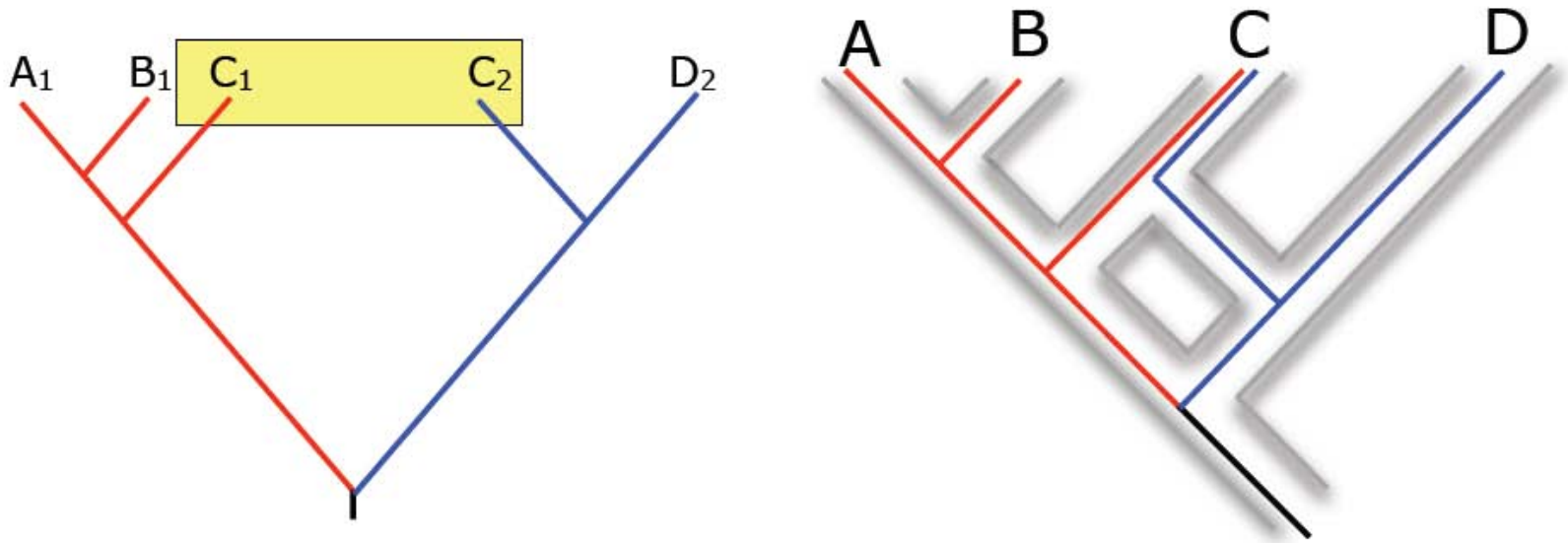
Gene tree \neq Species tree

... a pokud neznáme species tree, ani nevíme, že máme problém

A co když máme navíc polyploidy?

Vše se komplikuje (ne)úměrně se stupněm ploidie ☺

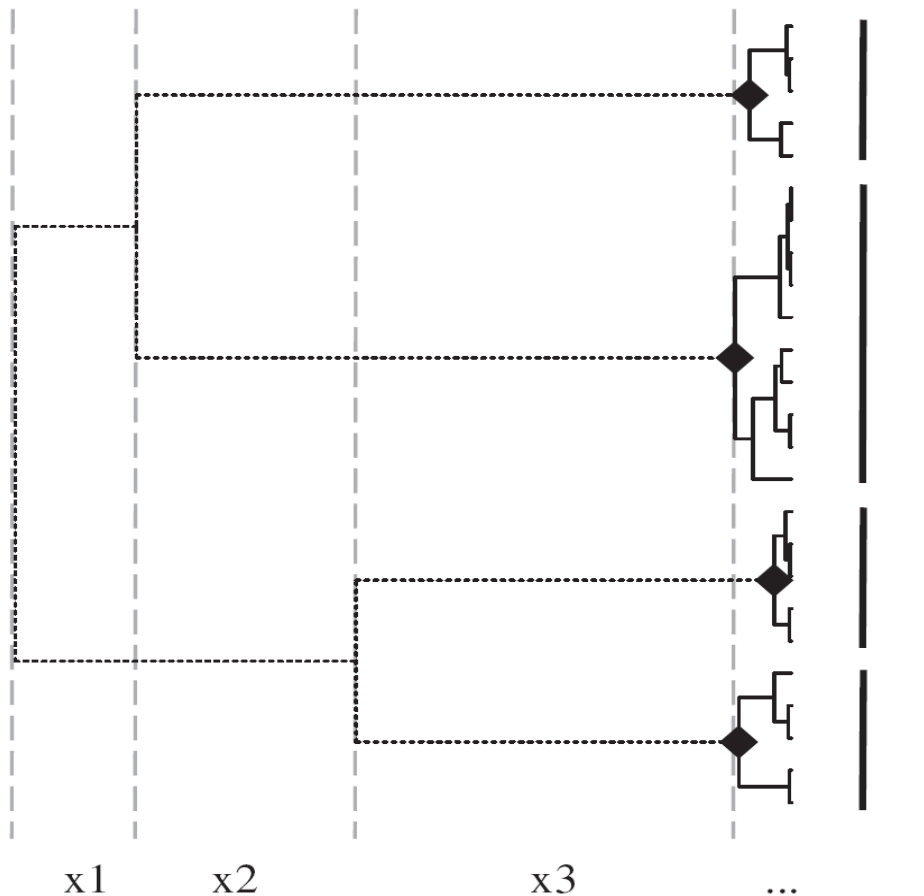
Kromě homologů, orthologů a paralogů se objevují ještě homoeology



Obecně platí – pracovat s jen orthology a sjednocovat homoeology

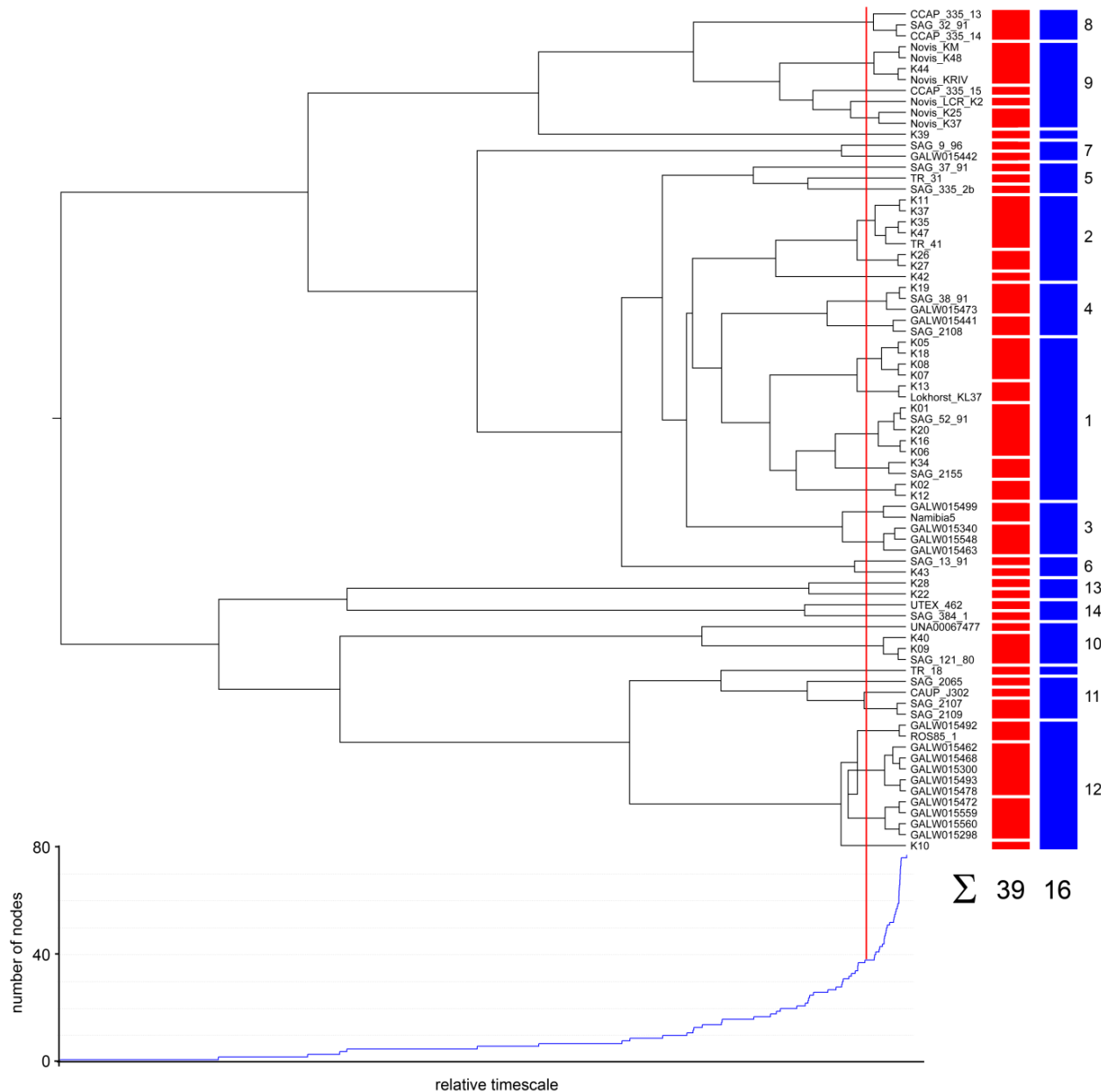
General mixed Yule-coalescent model - GMYC

GMYC: detekce druhových hranic podle rozdílu v rychlosti větvení stromů na úrovni druhů a populací



Black diamonds indicate the ancestral nodes of the biological species. Branches in dashed and solid lines represent speciation and coalescent branches, respectively. The bar next to the tips represents the species delimitation defined by the ancestral nodes shown on the tree. The GMYC approach analyses first the timing of branching events (x_1 , x_2 , x_3 periods between the dashed vertical lines). Then, the likelihood of observed waiting times on the tree is calculated, assuming a mixed model of coalescence within species and diversification among species.

GMYC – příkladový dataset



- Použití GMYC modelu ke stanovení druhových hranic u komplexu kryptických druhů rodu *Klebsormidium*

Programy

- BEAST
- FigTree

Nástroje pro analýzu populací a blízce příbuzných druhů

- haplotypové sítě, detekce ancestrálního polymorfismu [**TCS**]
- síť NNet a detekce rodičovství hybridů a polyploidů [**Splitstree**]
- rekonstrukce species tree při incomplete lineage sorting [***BEAST**]
- vizualizace gene trees ve species trees [**PrimeTV**]
- detekce druhových hranic pomocí GMYC [**BEAST, R, Genie**]

Analýza haplotypů pomocí statistické parsimonie

- pravděpodobnost parsimonie na základě párových rozdílů sekvencí
- analýza na diploidní i polyploidní úrovni (detekce rodičovských linií)
- vstupní data – alignmenty sekvencí v Nexus nebo Phylip formátu
 - cpDNA
 - nrDNA a nDNA (někdy potřeba získat klonováním)

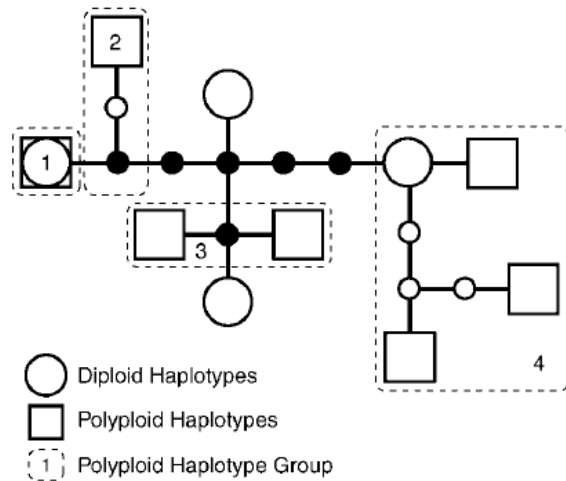


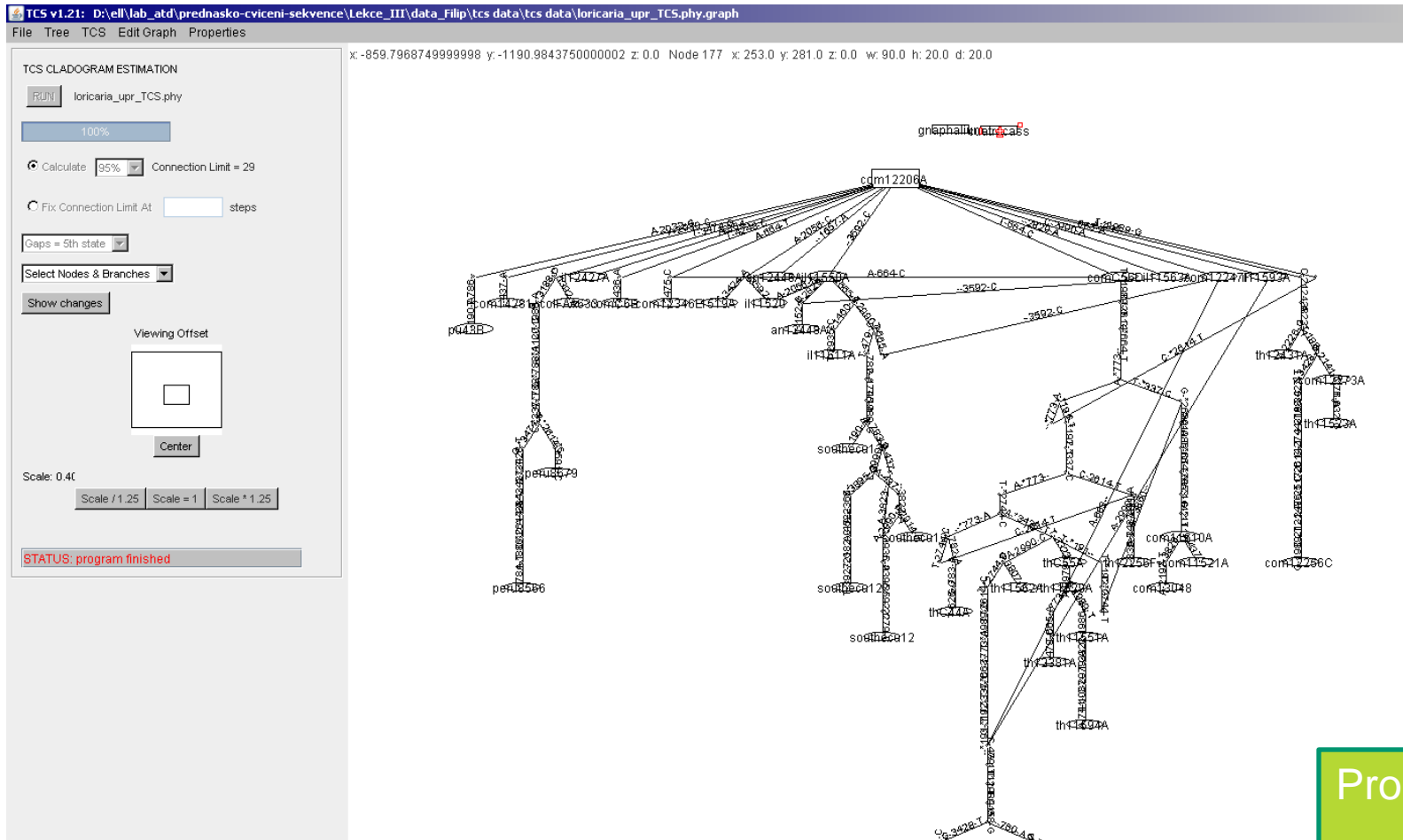
Fig. 3. Illustration of a network used to evaluate the minimum number of haplotypes contributed by diploids to polyploids when estimating the number of independent origins of a polyploid species. Large open circles (diploids) and squares (polyploids) represent sampled haplotypes. Small filled circles represent unsampled interior haplotypes inferred to have occurred in diploids, whereas small open circles represent interior unsampled haplotypes that cannot be inferred to have occurred in the diploids. Polyploid haplotype groups (dashed lines) are comprised of polyploid haplotypes that have a most recent common diploid or expected diploid haplotype ancestor.

Programy

- TCS

Analýza haplotypů pomocí statistické parsimonie

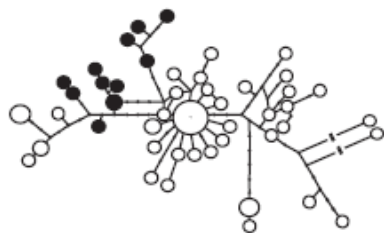
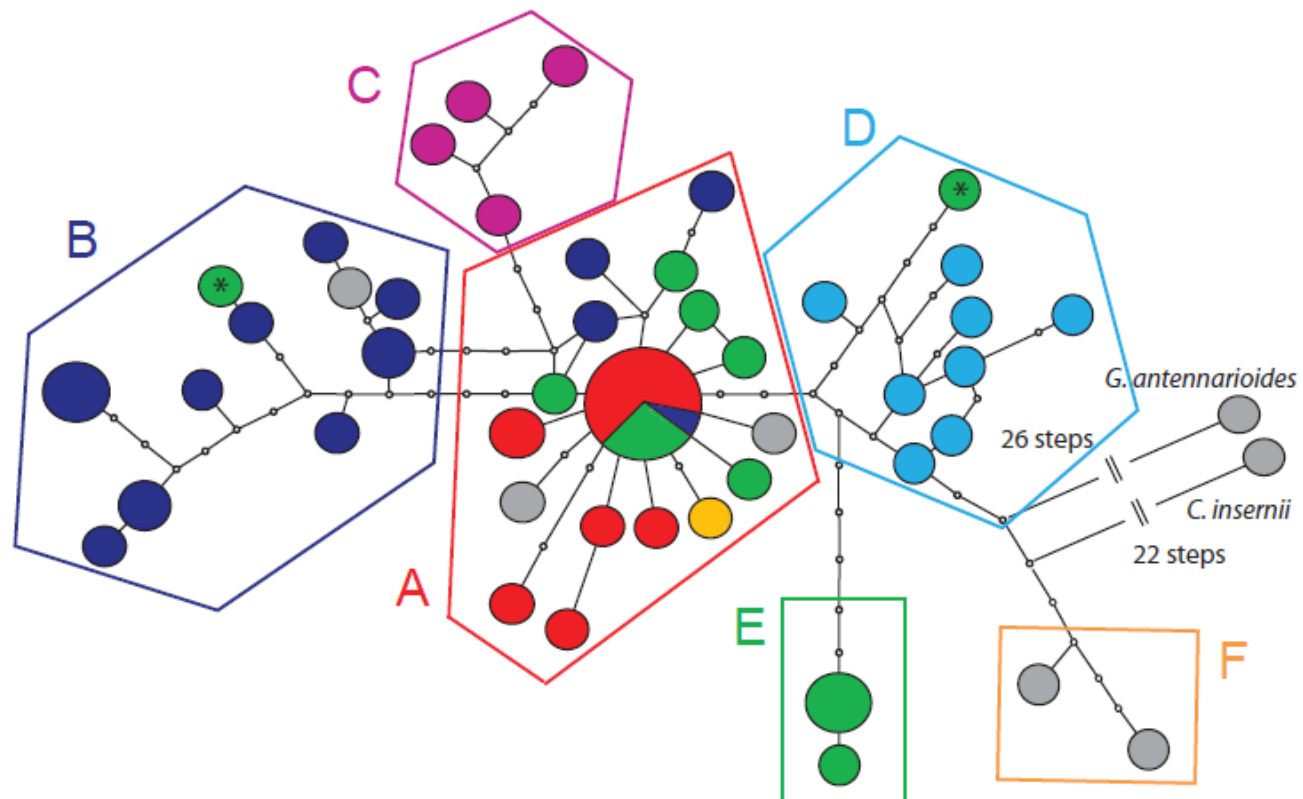
- analýza dvou druhů rodu *Loricaria* (cpDNA) – primární výstup analýzy
- soubor „loricaria_upr_TCS.phy“



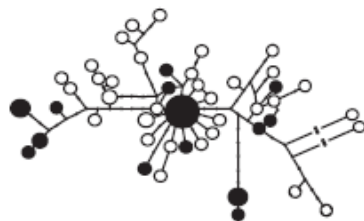
Program

- TCS

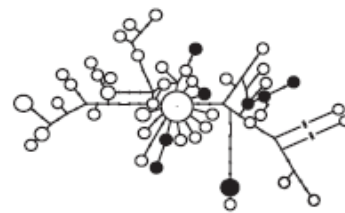
....po úpravě v Illustratoru.....



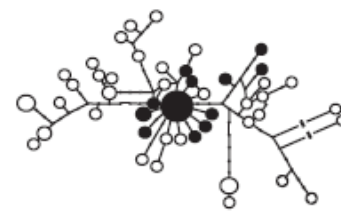
S Ecuador



C Ecuador

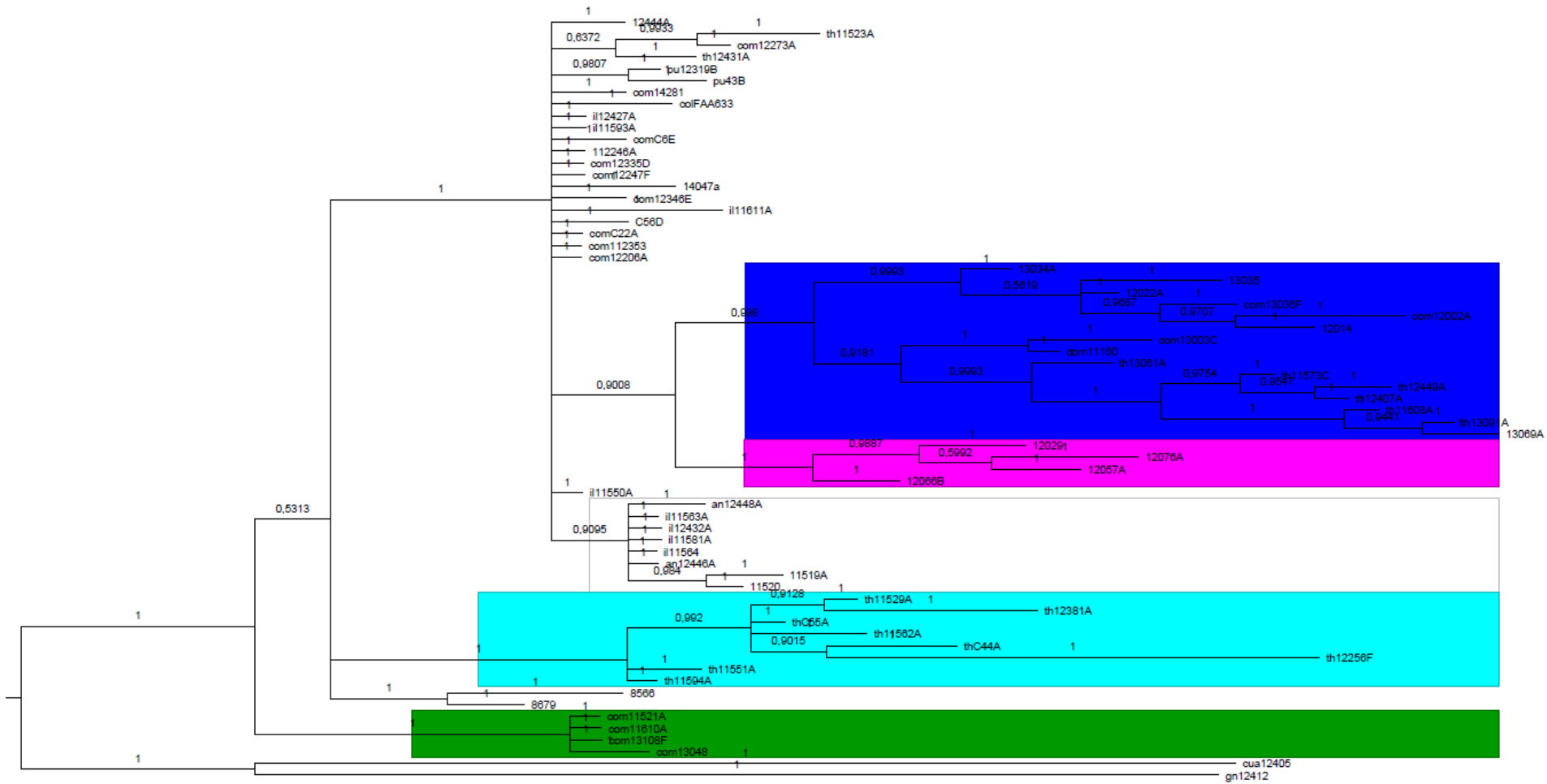


N Ecuador+S Col.



C+N Colombia

Stejná data analyzována jinou metodou – fylogenetický bayesovský strom



Detekce vztahů hybridů a polyploidů k jejich rodičům

- Princip – spojování informací z více konfliktních stromů, založeno na genetické vzdálenosti (obdobně jako NJ)
- Konflikt je vizualizován jako „box“ – hybridy umístěni ve vrcholech boxů
- Analýza na diploidní i polyploidní úrovni (detekce rodičovských linií)
- Vstupní data – alignment jednoho lokusu, či konkatenovaná data (Nexus)



Hybridizace na diploidní úrovni

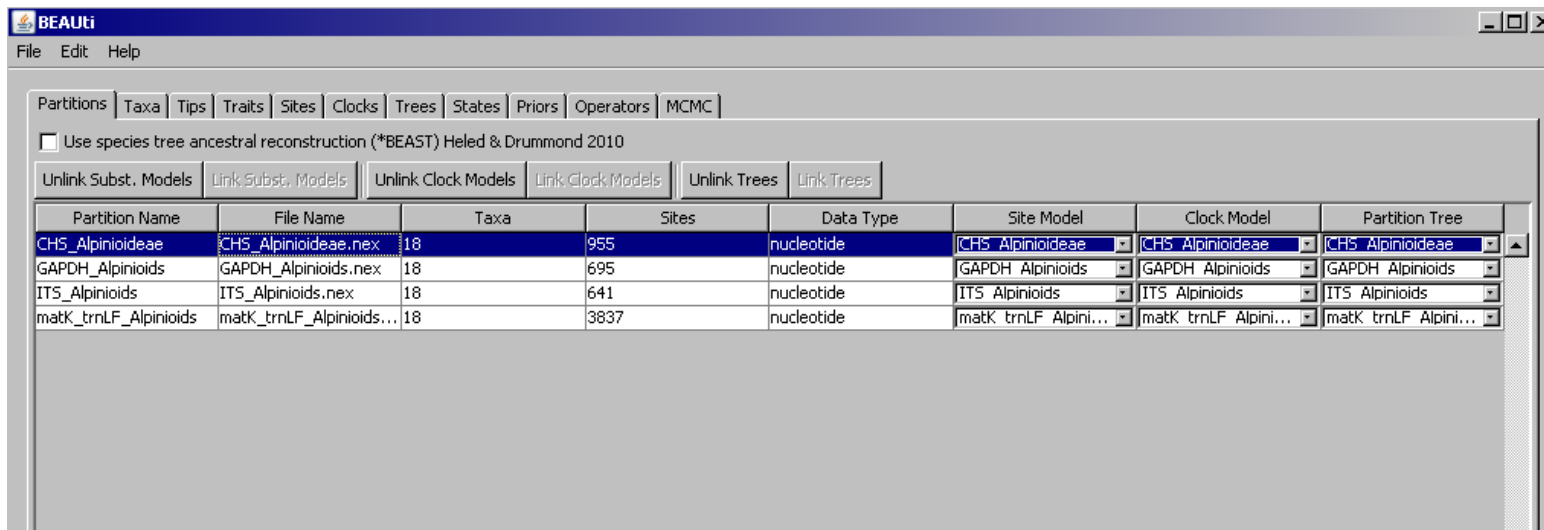
Hybridizace na polyploidní úrovni

Programy

- Splitstree

Rekonstrukce fylogeneze při incomplete lineage sorting

- Konstrukce species tree na základě gene trees a koalescenčních modelů
- Zatím prakticky aplikovatelné jen na diploidní úrovni a bez vlivu hybridizace
- Inkongruence stromů je vysvětlena vlivem ILS
- Nastavení analýzy v pomocném programu **Beauti**
- Vstupní data – alignmenty z několika lokusů (*.nex) se stejným samplingem
- Výsledkem – jediný strom – „species tree“

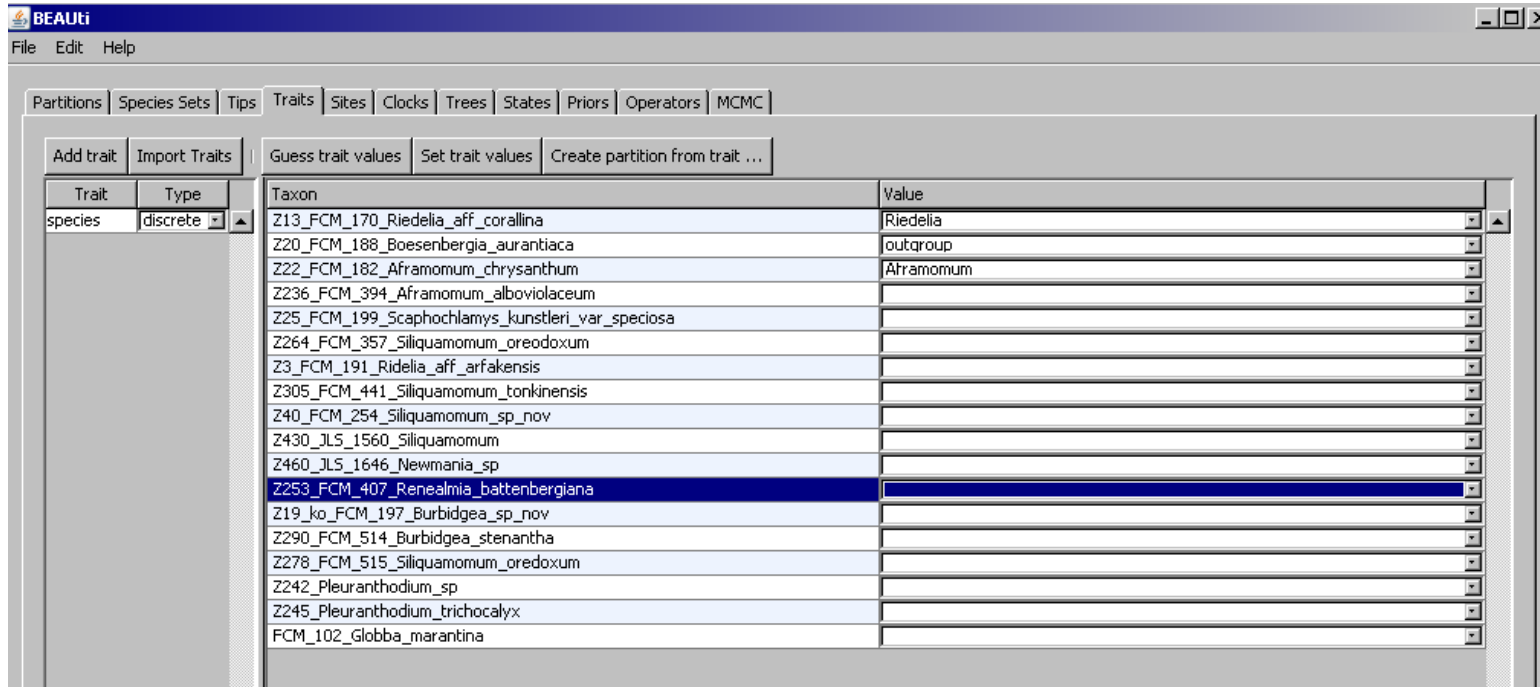


Programy

- BEAST/*BEAST
 - Beauti

Rekonstrukce species tree pomocí *BEAST

- Přiřazení jedinců k požadovanému druhu
- Nastavení evolučního modelu pro každý gene tree
- Nastavení koalescenčního modelu pro species tree



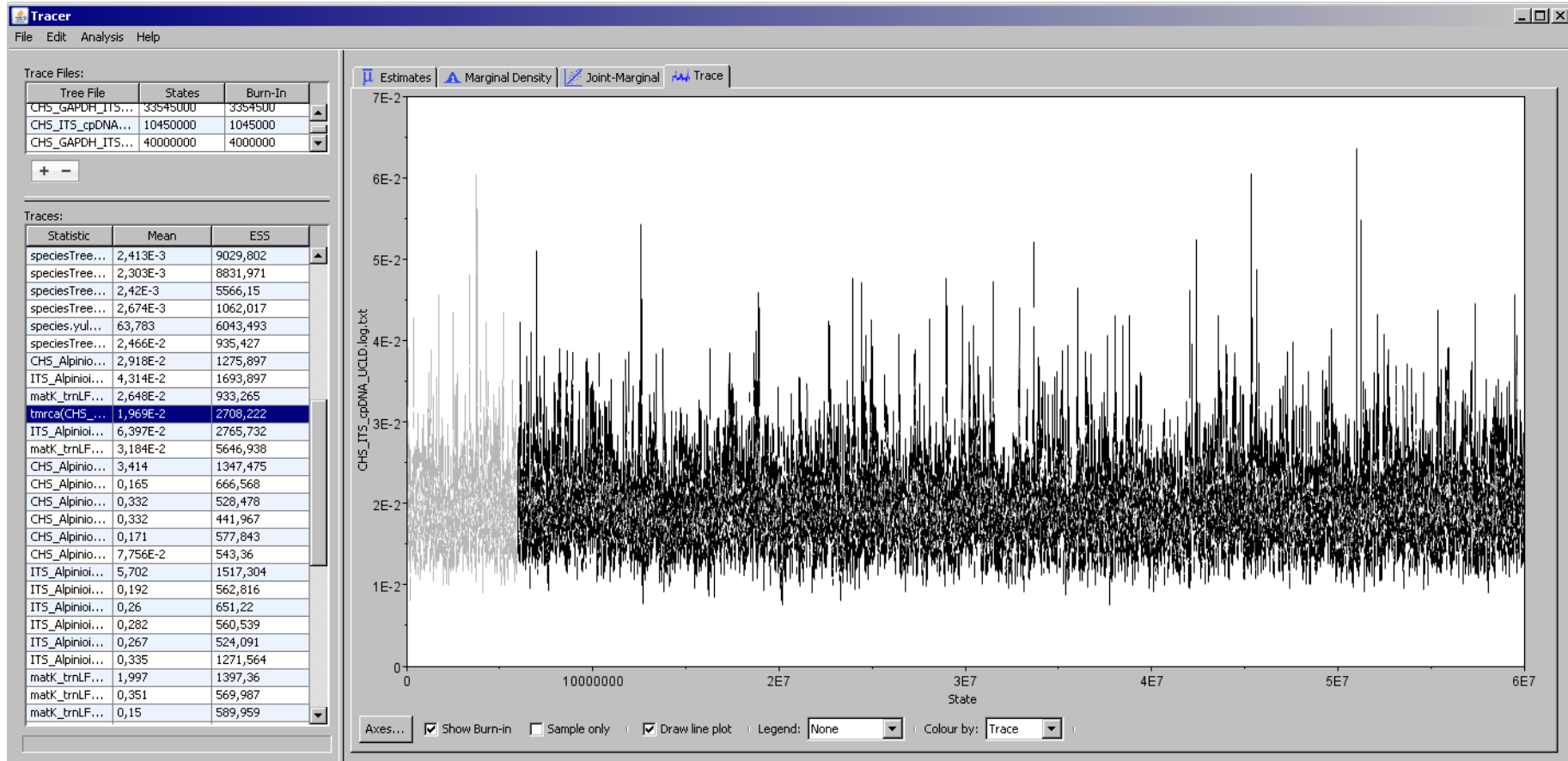
- Spuštění analýzy v programu BEAST

Programy

- BEAST/*BEAST
 - Beauti
 - Beast

Rekonstrukce species tree pomocí *BEAST

- Kontrola průběhu analýzy pomocí programu Tracer
- ESS hodnoty – měly by být > 200 (zobrazeny černě)

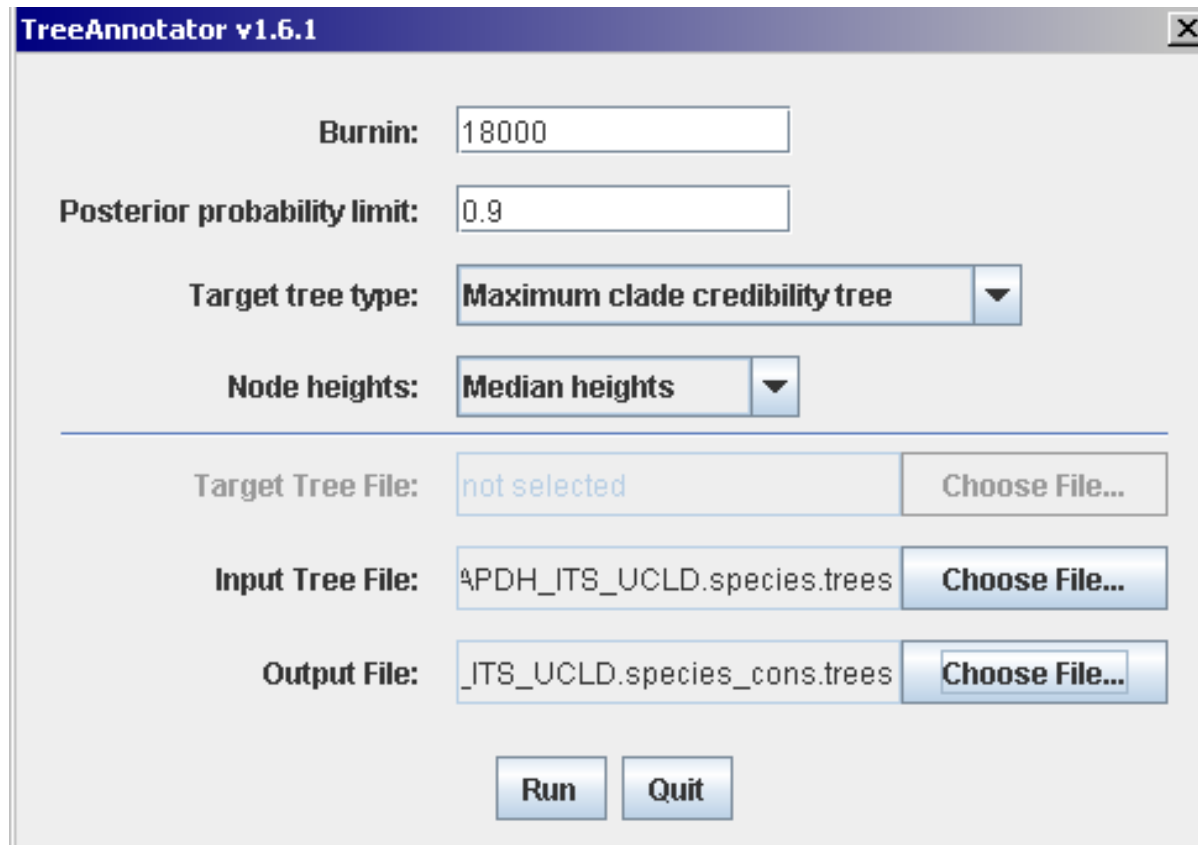


Programy

- Tracer

Rekonstrukce species tree pomocí *BEAST

- Tvorba konsenzuálního stromu ze stromů získaných během analýzy



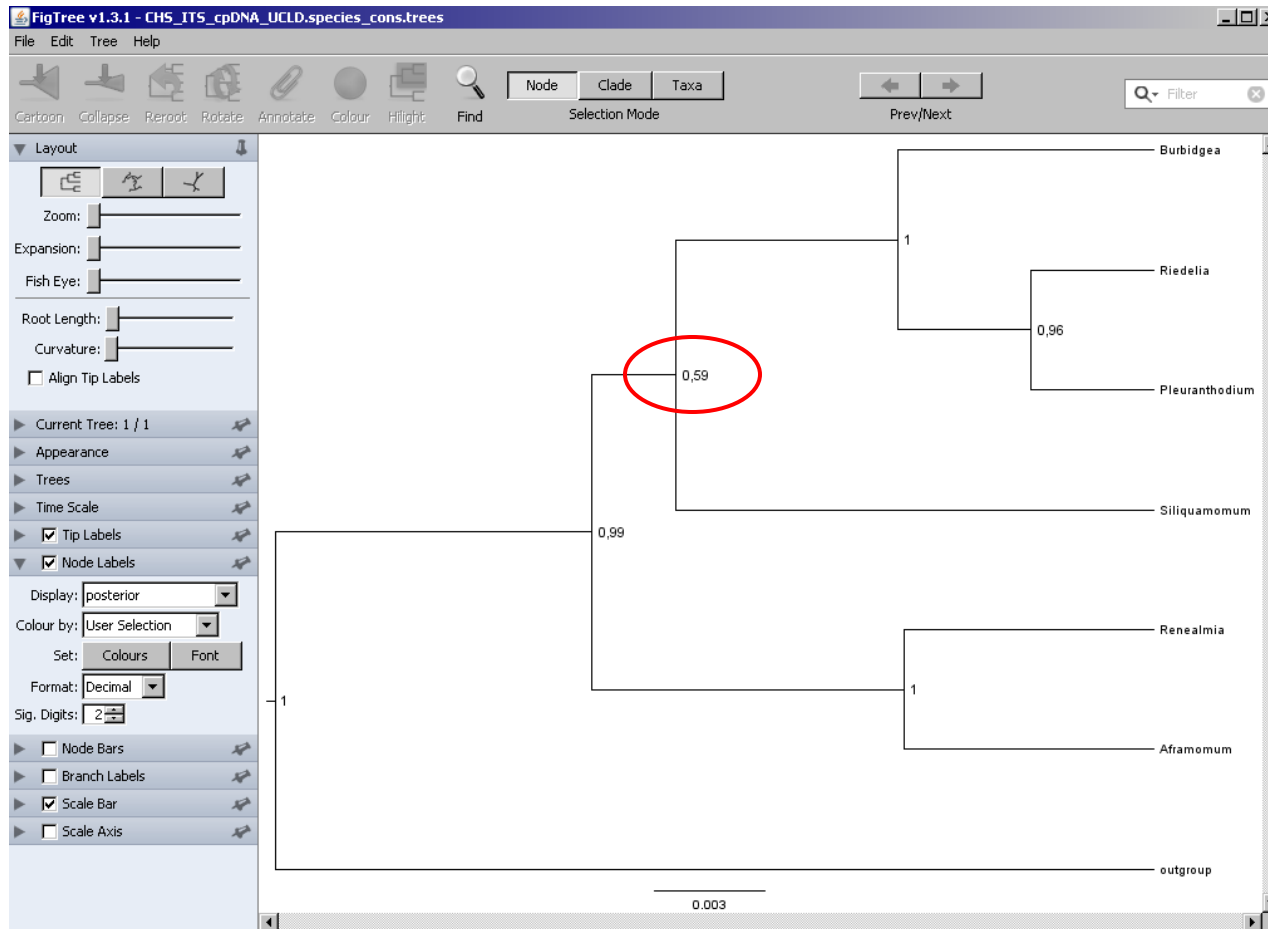
The screenshot shows the TreeAnnotator v1.6.1 window. It has a title bar with the text 'TreeAnnotator v1.6.1' and a close button. The main area contains several input fields and buttons. At the top, there is a 'Burnin:' field with the value '18000'. Below it is a 'Posterior probability limit:' field with the value '0.9'. Then there is a 'Target tree type:' dropdown menu set to 'Maximum clade credibility tree'. Below that is a 'Node heights:' dropdown menu set to 'Median heights'. A horizontal line separates these settings from the file selection section. In this section, there are three rows: 'Target Tree File:' with a text box containing 'not selected' and a 'Choose File...' button; 'Input Tree File:' with a text box containing 'APDH_ITS_UCLD.species.trees' and a 'Choose File...' button; and 'Output File:' with a text box containing '_ITS_UCLD.species_cons.trees' and a 'Choose File...' button. At the bottom of the window are two buttons: 'Run' and 'Quit'.

Programy

- TreeAnnotator

Rekonstrukce species tree pomocí *BEAST

- Zobrazení species tree v programu FigTree



- Interpretace nízkých podpor větví?

→ Inkongruence mezi gene trees a species tree není asi způsobena jen ILS

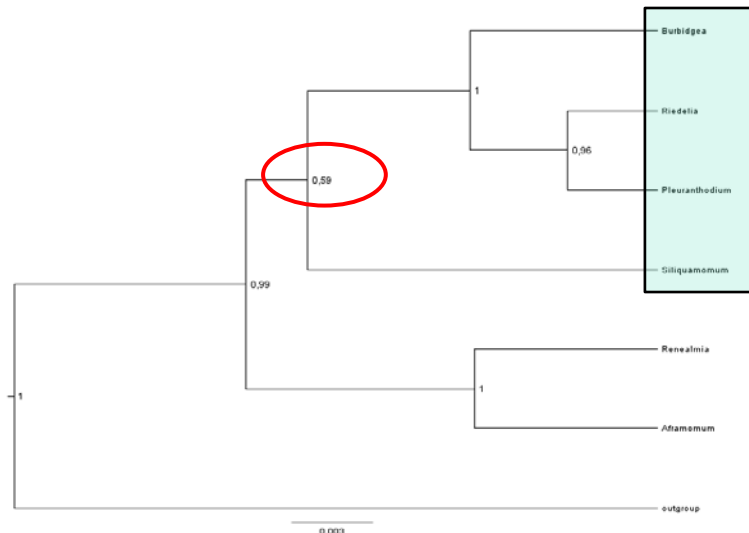
- Hybrididi v datasetu?

Programy

- FigTree

Rekonstrukce species tree pomocí *BEAST

- Přidání dalšího markeru neznamená vždy lepší rozlišení stromu ☺
- Možná jen podporuje hypotézu o hybridním původu rodu *Silicquomomum*
- Přestože ILS zde asi nehraje roli, může zde hrát roli duplikace/ztráta genu
→ další testy jsou nezbytné....

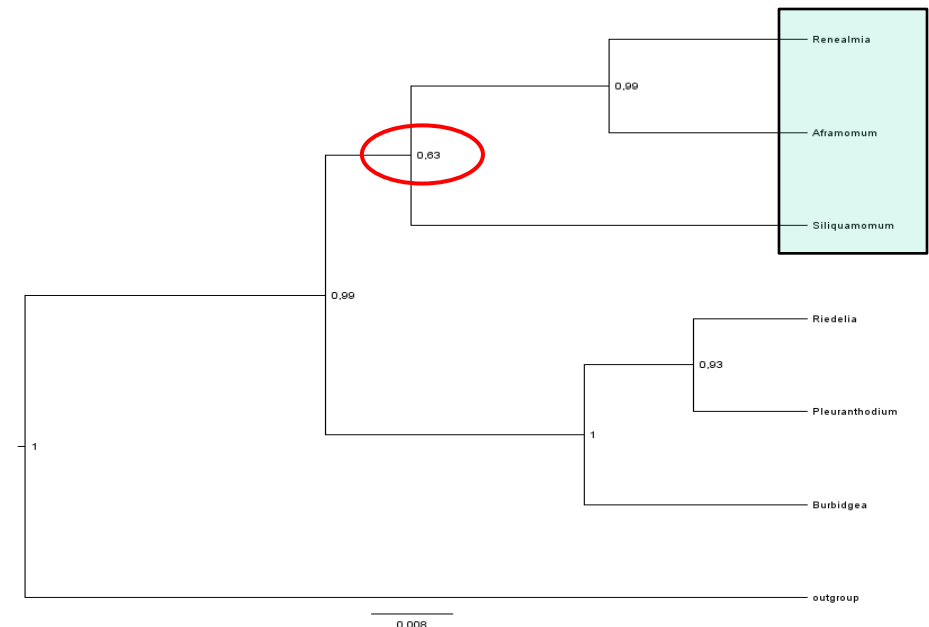


4 úseky:

- CHS (nDNA)
- ITS (nrDNA)
- matK, trnLF (cpDNA)

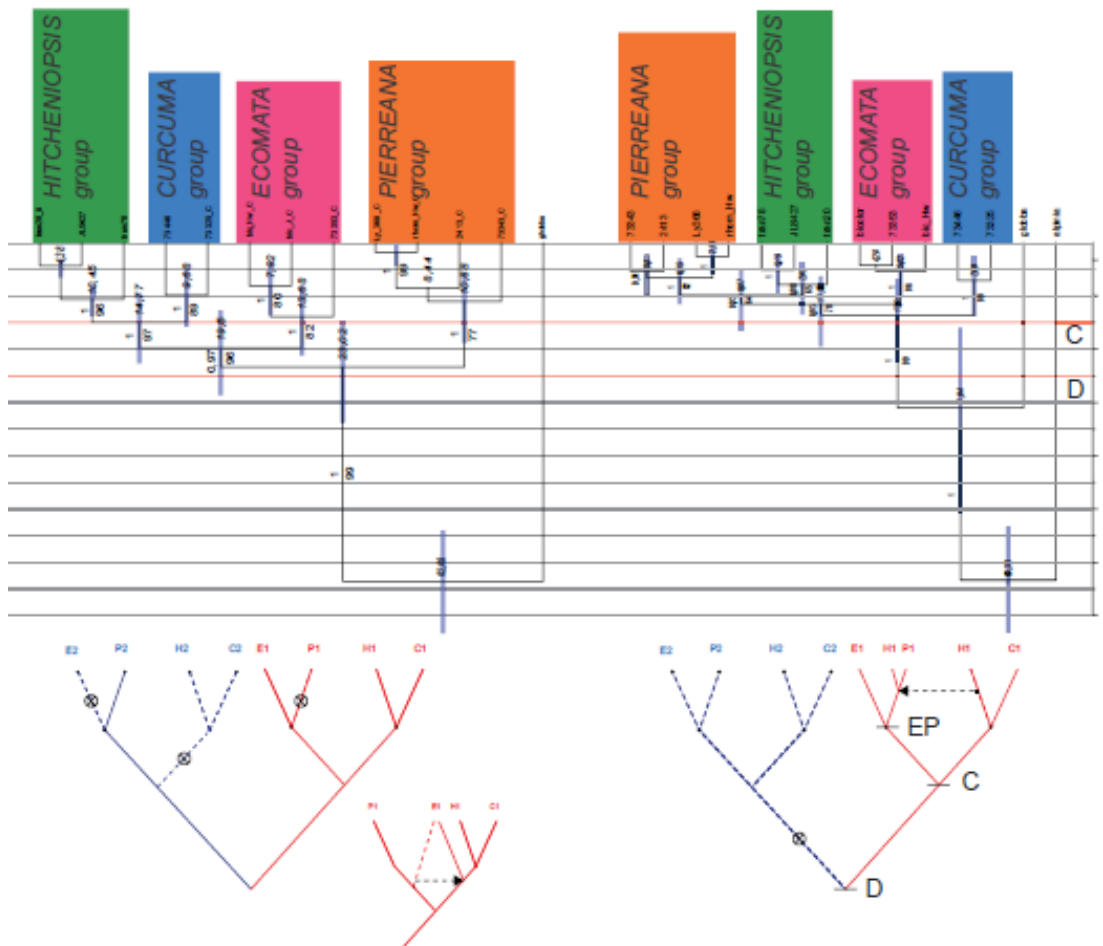
3 úseky:

- CHS, GAPDH (nDNA)
- ITS (nrDNA)



Odhalení genové duplikace v rámci gene trees

- (Relativní) datování pomocí BEAST
- Konstrukce species tree (*BEAST)
- Porovnání datovaných gene trees s topologií species tree
- Interpretace rozdílů

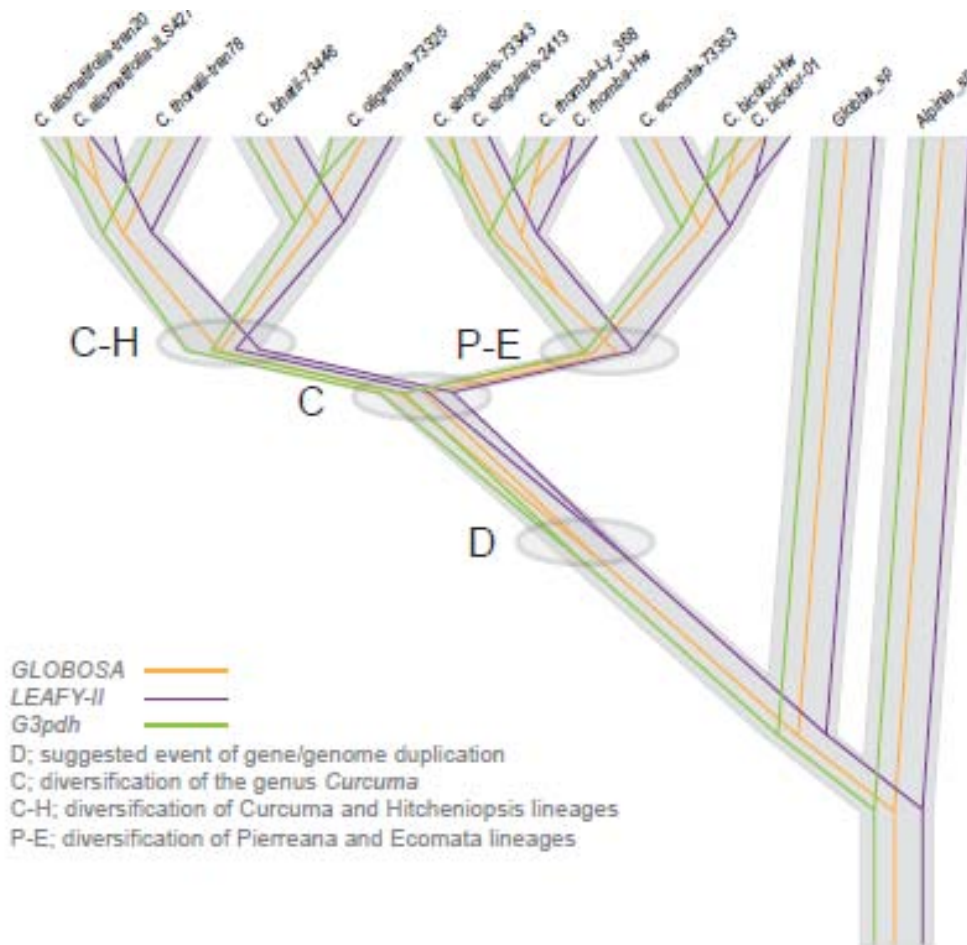


Programy

- BEAST
- *BEAST

Vizualizace gene trees ve species tree

- Např. Prime TV (<http://prime.sbc.su.se/cgi-bin/primetv.cgi>)
- Vstupní soubory – topologie species trees a gene trees (bez délek větví!!)
- ignoruje možnost laterálního přenosu (hybridizace)



Programy

- PrimeTV

Vizualizace gene trees ve species tree

- Např. Prime TV (<http://prime.sbc.su.se/cgi-bin/primetv.cgi>)
- Vstupní soubory – topologie gene trees a species tree (newick formát s délkami větví)
- Jedince v gene trees je třeba přejmenovat tak, aby pasovaly se jmény ve species tree
→ „Automatic reconciliation“ – navrhne nejparsimoničtější řešení

Tree input

Relate trees

Adjust drawing

Species tree

Branch lengths are interpreted as time information. Unit lengths are used if no branch lengths are present in the tree

```
7398312384):0.024738047721685877,  
(thorelii:0.018371510770020485,alismaticifolia:0.018  
371510770020485):0.02277677434997776):0.008229888  
796457814):0.06614466748513596,Globba_sp:0.1155228  
4140159204):0.014554536551069544,Alpinia_sp:0.1300  
7737795266158);
```

Gene tree

Branch lengths are ignored in this tree.

```
058315475805):6.128311666483107):2.065294844779906  
,  
(73446_bhatii:7.092450823234351,73325_oligantha:7.  
092450823234351):6.5212615195762424):17.4240770645  
18876,01_Globba_sp:31.03778940732947):32.292150334  
56823,01_Alpinia_sp:63.3299397418977);
```

Obnovit

Alternatives:

1.	<input type="button" value="Automatic reconciliation"/>	<input type="button" value="Sample data"/>	Produces the most parsimonious reconciliation.
2.	<input type="button" value="Use reconciliation in tree"/>		If your gene tree contains a reconciliation in PrIME format.
3.	<input type="button" value="Manual reconciliation"/>		Enter a reconciliation in tabular format.

*For bug reports, feature requests, wild praise, et.c., please contact [Lars Arvestad](#)
[PrIME TV homepage](#)*

Program

- PrimeTV

Vizualizace gene trees ve species tree

- Kontrola mapování jmen u gene trees ke species tree
→ „Reconcile and view“

Tree input

Relate trees

Adjust drawing

New trees

Gene-to-species mapping

bicolor
| ecomata
| | rhomba
| | | singularis
| | | | oligantha
| | | | | bhatii
| | | | | | thorelii
| | | | | | | alismatifolia
| | | | | | | | Globba_sp
| | | | | | | | | Alpinia_sp

01_alismatifolia	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
01_thorelii	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
02_alismatifolia	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
73343_singularis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2413_singularis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ly368_rhomba	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
02_rhomba	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
01_bicolor	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
73353_ecomata	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
02_bicolor	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
73446_bhatii	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
73325_oligantha	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
01_Globba_sp	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
01_Alpinia_sp	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

Reconcile and view

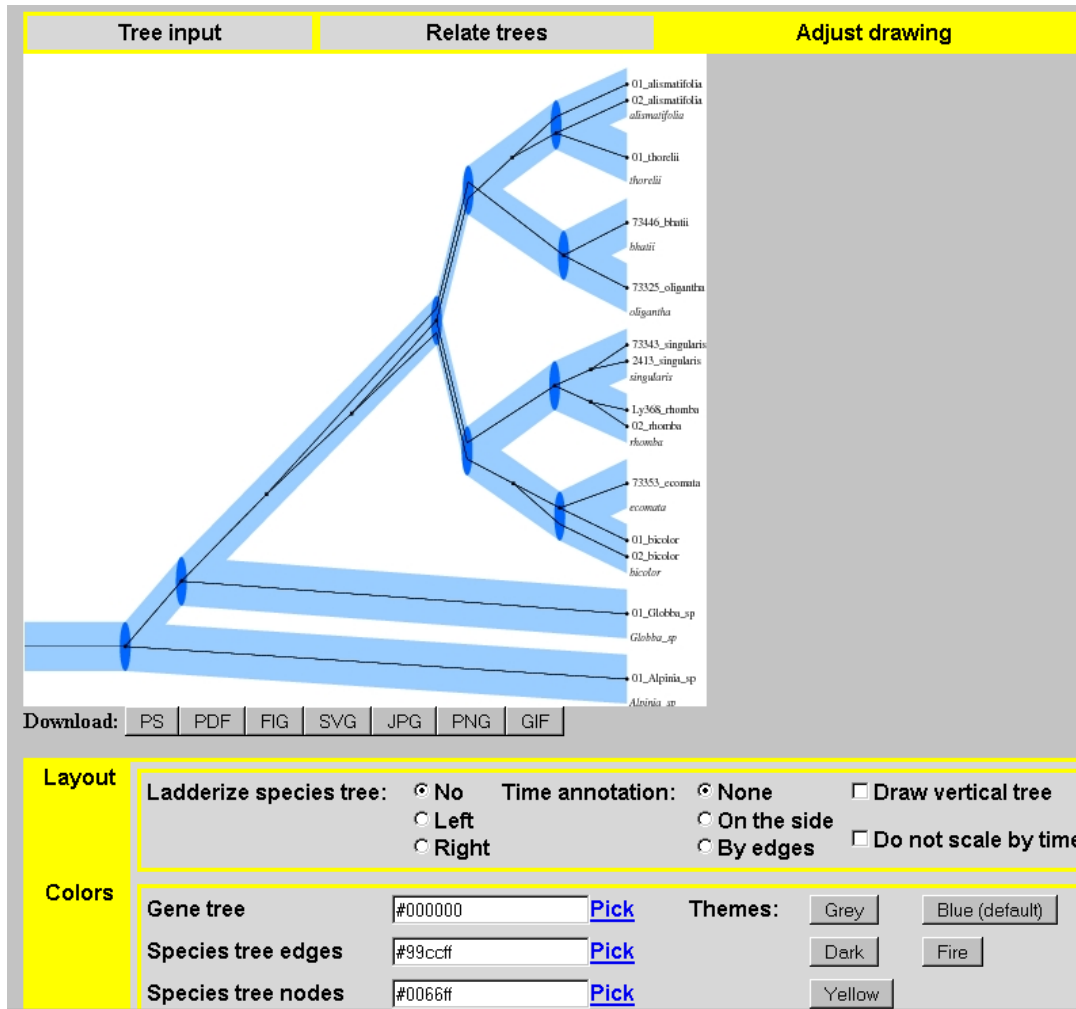
Show text-version of this

Program

- PrimeTV

Vizualizace gene trees ve species tree

- Výstup programu – lze stáhnout v různých formátech a upravit v grafickém programu (např. *.ps jako vstup pro Illustrator)
- Na Species tree se mapuje vždy jen jeden gene tree



Program

- PrimeTV

Praktické cvičení – konstrukce haplotypové sítě

- Na základě jednoho alignmentu vytvořte haplotypovou síť pomocí programu **TCS**
 - Pokud nemáte vlastní data, použijte soubor „**loricaria_upr_TCS.phy**“
 - Spusťte analýzu v programu TCS podle návodu a zkuste nastavit různé limity pro propojení jedinců
 - Proveďte stejnou analýzu s datasetem „**GAPDH_Alpinoids.nex**“
 - Jak a proč se výstupy analýzy liší?
- Na základě stejných alignmentů vytvořte síť pomocí **NeighbourNet**
 - Formát phylip je nejdříve nutné převést do formátu nexus
 - Porovnejte výsledky analýz a míru neshod v jednotlivých datasetech
 - Je možné, že v některém z datasetů je přítomen hybridní jedinec?

Praktické cvičení – konstrukce species tree

- Vytvořte species tree na základě několika vstupních souborů reprezentujících gene trees
 - Pokud nemáte vlastní data použijte soubory „GAPDH_Alpinoids.nex“, „CHS_Alpinioideae.nex“, „ITS_Alpinoids.nex“ a „matK_trnLF_Alpinoids.nex“
 - Nastavte analýzu *BEAST v programu Beauti a spusťte v BEAST
 - Průběh analýzy si prohlédněte v Tracer
 - V programu TreeAnnotator si zkonstruuje konsenzuální species tree z připraveného souboru „CHS_GAPDH_ITS_UCLD.species.trees“ nebo „CHS_GAPDH_ITS_cpDNA_strict_clock.species.trees“
 - Postupujte dle návodu, případně nahlédněte i do originálního manuálu
- Neboť analýza potrvá dlouho, vytvořte ze vstupních souborů konkatenovaný soubor a spusťte na něm analýzu pomocí Maximální parsimonie, ML nebo Bayesovskou analýzy
 - Pro přípravu konkatenovaného datasetu můžete použít konverze do fasta souboru a spojit jednotlivé soubory pomocí Fabox (viz Lekce 1)
 - Z konkatenovaného fasta souboru vytvořte opět nexus

Praktické cvičení – vizualizace gene trees ve species tree

- Namapujte topologii jednoho z gene trees na topologii odhadnutého species tree pomocí webového rozhraní PrimeTV
 - <http://prime.sbc.su.se/cgi-bin/primetv.cgi>
 - Použijte připravené soubory ve složce „primeTV/funguje“
 - Postupujte dle návodu
 - Zhodnoťte, zda se topologie liší, případně jak, a jak byste odlišnosti interpretovali
- Pokuste se připravit správný formát gene trees a species trees pro obdobnou vizualizaci pro stromy připravené ve složce „primeTV“
 - „CHS_coalesc_newick.tre“ a
 - „CHS_GAPDH_ITS_cpDNA_species_tree_newick.tre“
- Bude nutné správně přejmenovat OTU v gene tree....