2. Analýza sekvenačnch dat – II

- genetické distance **[MEGA]**
- konstrukce fylogenetických stromů **[MEGA, PAUP, Garli, MrBayes]**
- testy topologických hypotéz **[PAUP]**
- testy inkongruence v datasetu **[PAUP]**
- vizualizace stromů **[Treeview, Figtree, Dendroscope]**

- **praktická část – fylogenetická analýza sekvenačních dat**

Typy fylogenetických analýz

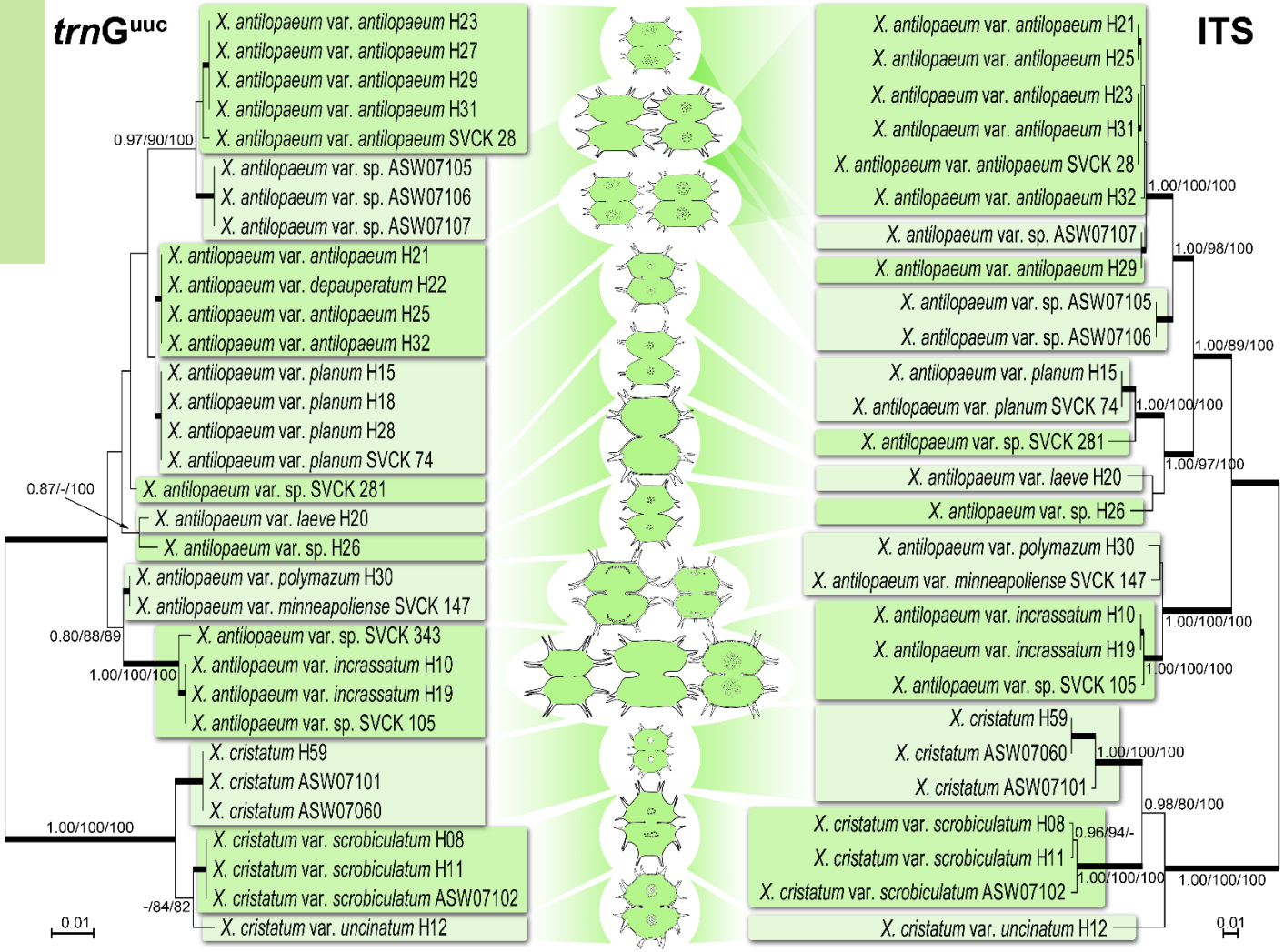
Distanční metody:
- Neighbor-Joining
- Minimum Evolution
- UPGMA, ...

Maximum Likelihood

Bayesian Inference

Maximum Parsimony

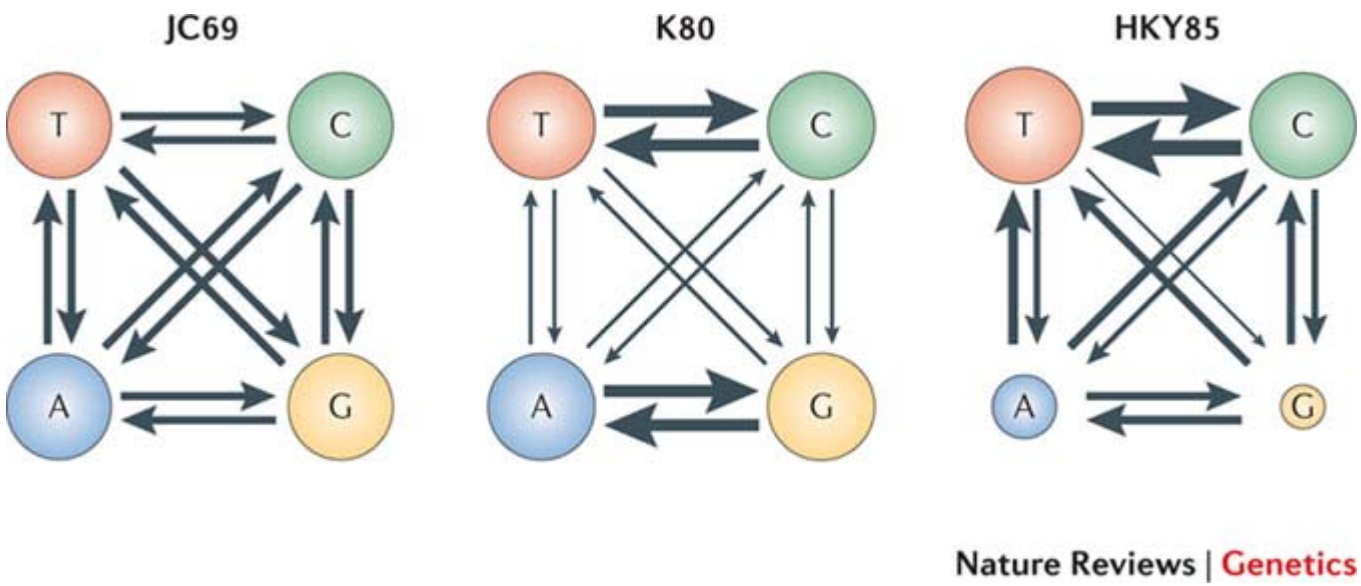
Detailní přehled metod rekonstrukce evoluce viz:
http://ibot.sav.sk/usr/Karol/teaching_mat.html



Genetické distance, substituční modely

- pro výpočet fylogenetických analýz je nutné stanovit genetické (evoluční) distance mezi sekvencemi
- v případě molekulárních hodin je distance přímo úměrná času

- ***p-distance***: prostý rozdíl sekvencí – výrazné podhodnocení reálných distancí (saturace)
- ***substituční modely*** – odhady jednotlivých substitučních rychlostí pomocí Markovových modelů + frekvence výskytu nukleotidů **= Q matice**



Substituční modely

- **Jukes-Cantor (JC69)**: během evoluce mají všechny nukleotidy stejnou pravděpodobnost substitucí i stejnou frekvenci výskytu bází (**nst=1**)

	A	T	C	G
A	-	α	α	α
T	α	-	α	α
C	α	α	-	α
G	α	α	α	-

- **Felsenstein (F81)**: nukleotidy mají stejnou pravděpodobnost substitucí ale jinou frekvenci výskytu bází (**nst=1**)

	A	T	C	G
A	-	$\alpha \Pi_T$	$\alpha \Pi_C$	$\alpha \Pi_G$
T	$\alpha \Pi_A$	-	$\alpha \Pi_C$	$\alpha \Pi_G$
C	$\alpha \Pi_A$	$\alpha \Pi_T$	-	$\alpha \Pi_G$
G	$\alpha \Pi_A$	$\alpha \Pi_T$	$\alpha \Pi_C$	-

- **Kimura (K80)**: jiné substituční rychlosti pro transice a transverze, shodné frekvence bází (**nst=2**)

	A	T	C	G
A	-	β	β	α
T	β	-	α	β
C	β	α	-	β
G	α	β	β	-

- **Hasegawa-Kishino-Yano (HKY)**: jiné substituční rychlosti pro transice a transverze, různé frekvence bází (**nst=2**)

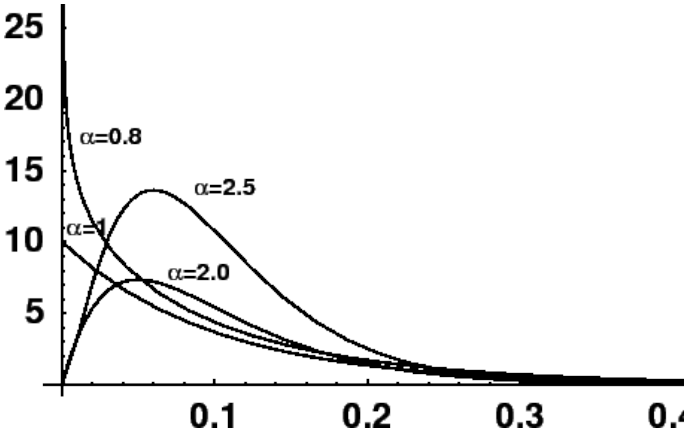
	A	T	C	G
A	-	$\beta \Pi_T$	$\beta \Pi_C$	$\alpha \Pi_G$
T	$\beta \Pi_A$	-	$\beta \Pi_C$	$\beta \Pi_G$
C	$\beta \Pi_A$	$\beta \Pi_T$	-	$\beta \Pi_G$
G	$\beta \Pi_A$	$\beta \Pi_T$	$\beta \Pi_C$	-

- **General time reverdible (GTR)**: pravděpodobnosti substitucí a frekvence bází jsou specifikovány pro každou možnost (**nst=6**)

	A	T	C	G
A	-	$\alpha \Pi_T$	$\beta \Pi_C$	$\gamma \Pi_G$
T	$\alpha \Pi_A$	-	$\delta \Pi_C$	$\varepsilon \Pi_G$
C	$\beta \Pi_A$	$\delta \Pi_T$	-	$\zeta \Pi_G$
G	$\alpha \Pi_A$	$\varepsilon \Pi_T$	$\zeta \Pi_C$	-

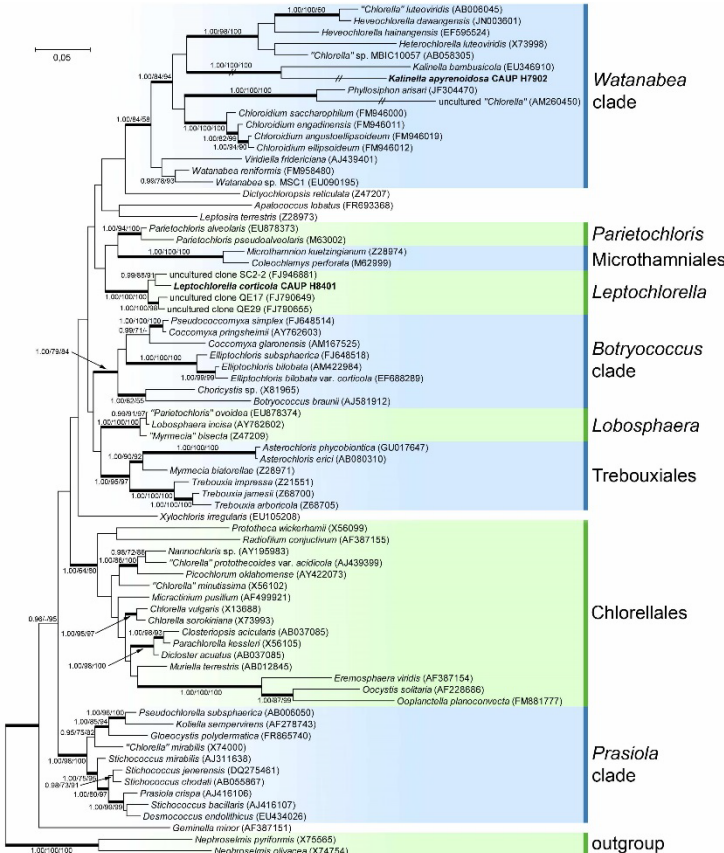
Substituční modely

- **Gamma distribuce (Γ)**: modeluje variabilitu v míře nukleotidových substitucí na různých pozicích alignmentu. Většinou se model zjednodušuje do 4 α kategorií



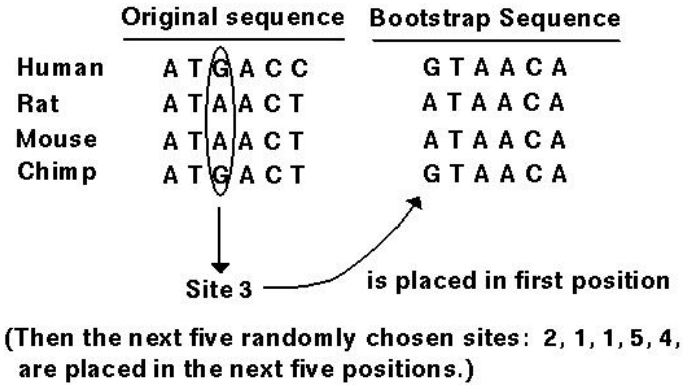
- **Proporce nevariabilních míst (I)**: existence velkého množství nevariabilních pozic negativně ovlivňuje odhad genetických distancí. Aplikace **I modelu** je např. důležitá při současné přítomnosti krátkých a dlouhých větví

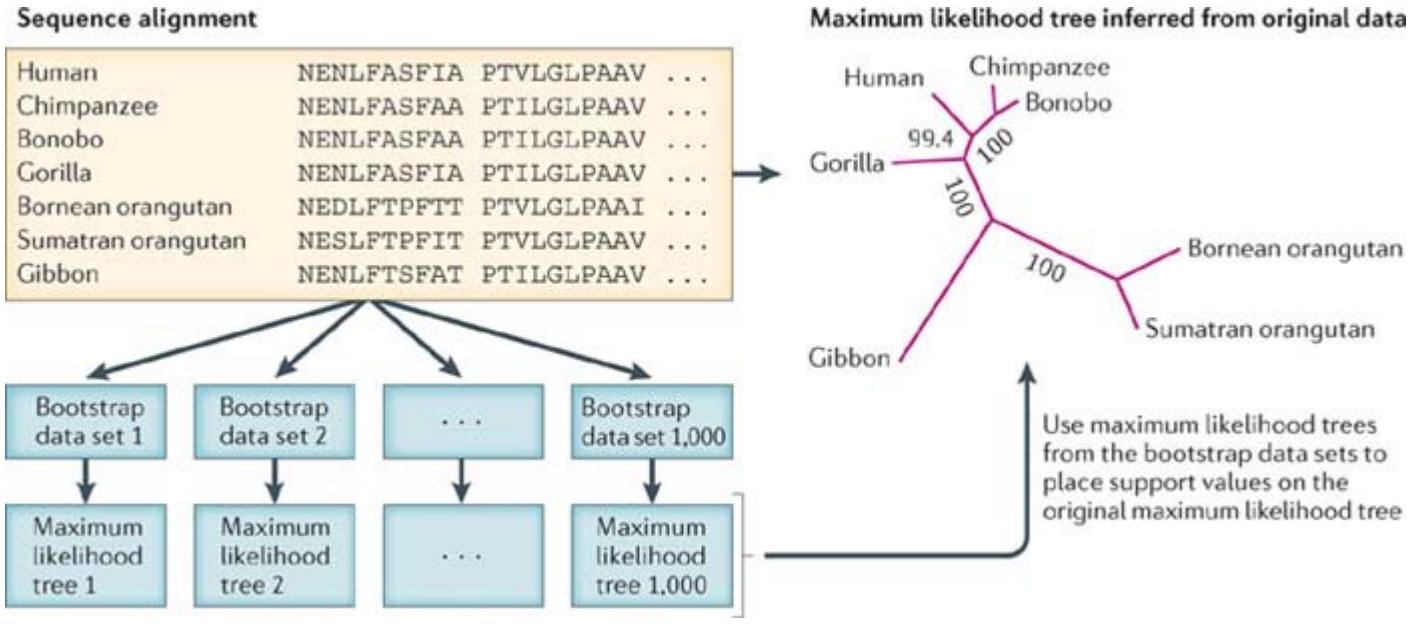
- **Kovarion (cov)**: modeluje variabilitu v míře nukleotidových substitucí v závislosti na fylogenetické pozici dané sekvence



Test topologie - bootstrapping

- výpočet stromů na základě nově generovaných alignmentů
- konstrukce majority-rule konsenzuálního stromu
- hodnoty bootstrapů je nutné zobrazit na topologii stromu zkonstruovaného na základě originálního alignmentu





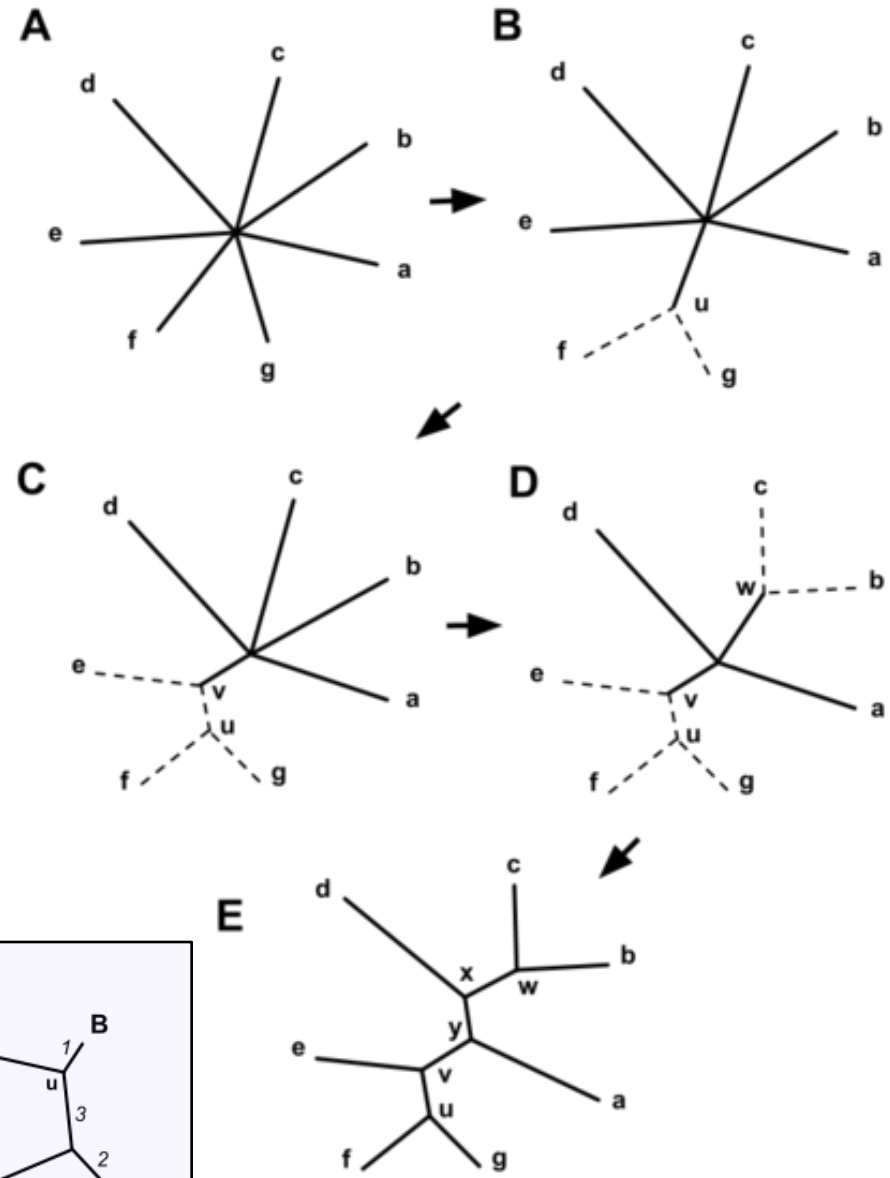
Fylogenetické analýzy na základě distančních matic

- stromy se počítají na základě distancí, spočtených pro každou dvojici sekvencí
- aplikují se substituční modely
- vhodné pro rychlou analýzu velkého množství sekvencí (v řádu několika stovek až tisíc)

- **Minimum Evolution (ME)**: hledá se strom o nejmenším součtu délek všech větví

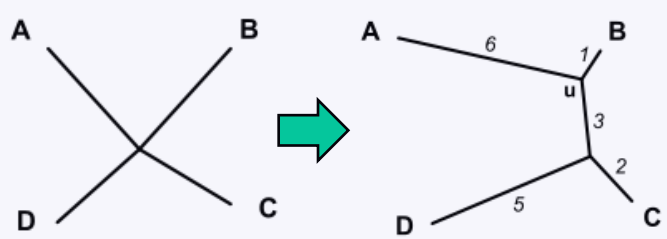
- **Neighbor Joining (NJ)**: heuristický algoritmus na rychlé nalezení ME stromu (začíná se u hvězdicovitého stromu)

- **BioNJ**: lepší přesnost topologie u vzdáleně příbuzných sekvencí



	A	B	C	D
A	0	7	11	14
B	7	0	6	9
C	11	6	0	7
D	14	9	7	0

	u	C	D
u	0	5	8
C	5	0	7
D	8	7	0



Neighbor-Joining – příkladový dataset

Fylogeneze rodu *Micrasterias*

1) NJ v programu MEGA
   - Vytvoření alignmentu ve formátu MEGA
   - Spuštění NJ analýzy
   - Jednoduchá úprava a export stromu

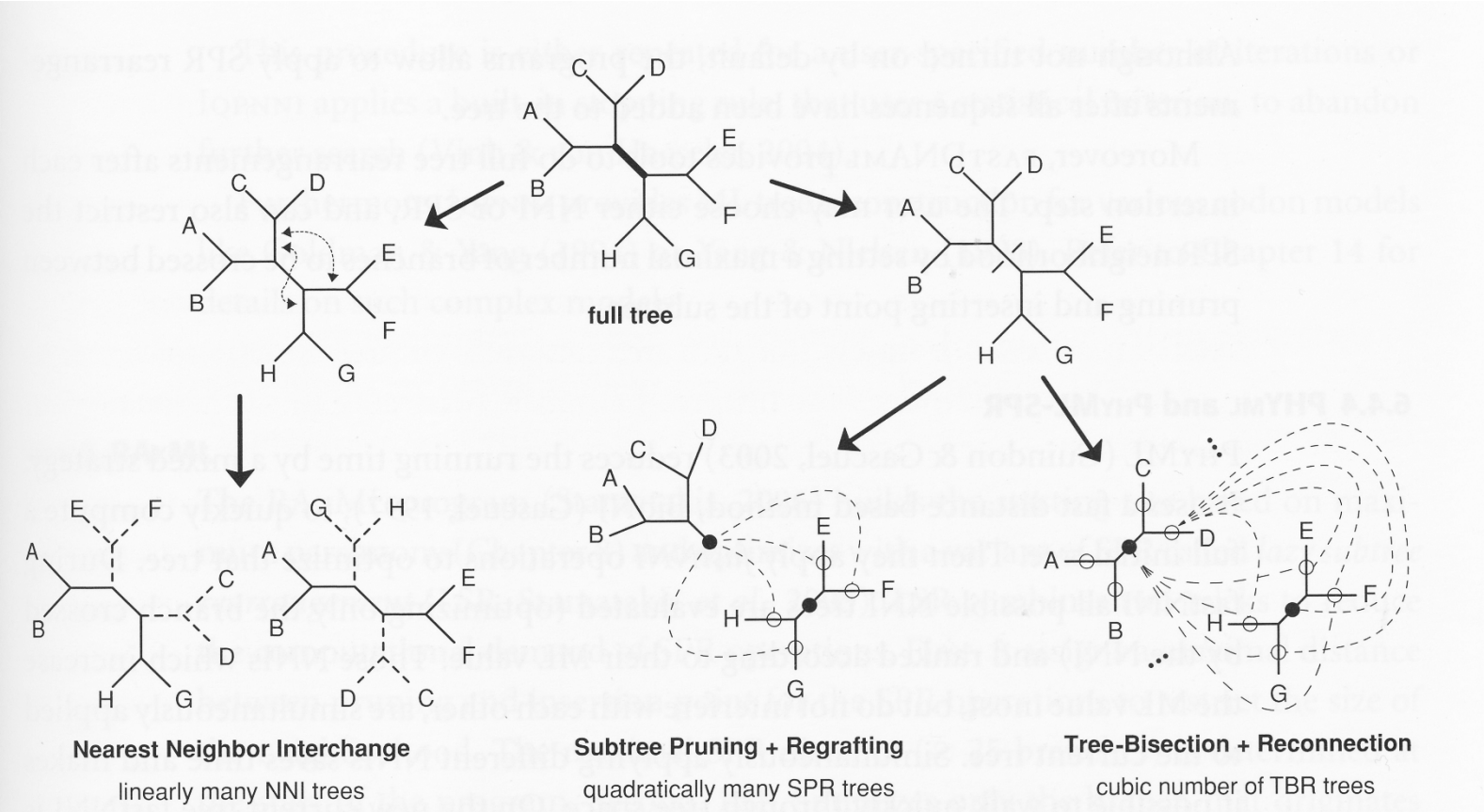
2) BioNJ v programu PAUP
   - Sekvence ve formátu Nexus
   - Spuštění NJ analýzy pomocí PAUP příkazového bloku

Programy

- MEGA
- PAUP

Maximum Likelihood (ML)

- hledání nejpravděpodobnějšího stromu odrážejícího evoluci sekvencí
- posuzování pravděpodobností nekonečně velkého množství stromů (různé topologie, délky větví, parametrů substitučních modelů, ...)
- heuristické metody pro hledání struktury stromu:
  - Nearest Neighbor Interchange (NNI)
  - Subtree Pruning + Regrafting (SPR)
  - Tree-Bisection + Reconnection (TBR)



Maximum Likelihood – příkladový dataset

Fylogeneze rodu *Micrasterias*

- 3 partitions (18S rDNA, psaA, coxIII)
- alignment ve formátu nexus
- nastavení a spuštění analýzy
- vytvoření konsenzuálního stromu a vypočtení hodnot bootstrapu

Programy

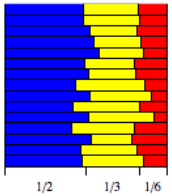
- Garli
- PAUP

Bayesova analýza (BI)

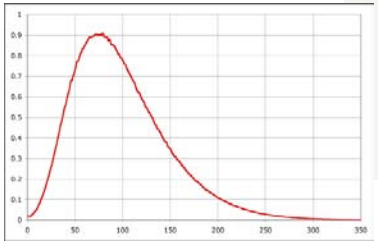
- **tradiční otázka**: pokud je v košíku stejně modrých a červených kuliček, jaká je pravděpodobnost, že si vytáhnu 3 modré a 3 červené kuličky?

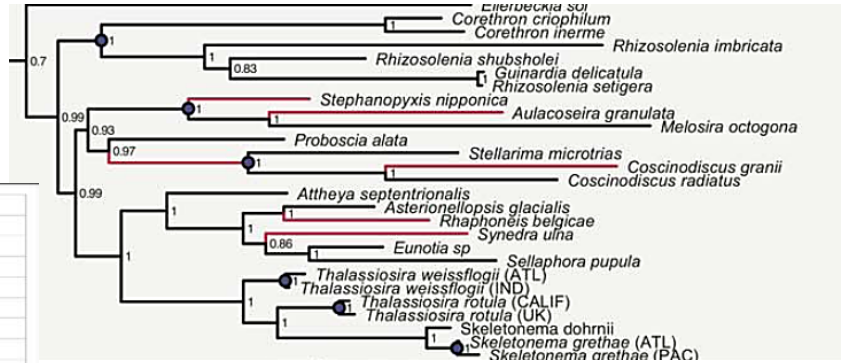
- **Bayesovská otázka**: pokud si vytáhnu 3 modré a 3 červené kuličky, jaká je pravděpodobnost, že je v košíku stejně modrých a červených kuliček?
  ➡ *podmíněná pravděpodobnost* (**posterior**)

- existuje nekonečné množství předpokladů (víc modrých, víc červených,...), které ale mohou mít různou pravděpodobnost ➡ tzv. **priors**
  - **uniform priors** – stejná pravděpodobnost, žádné předpoklady (*topologie*)
  - **exponencial priors** – např. *délky větví* (likelihood je negativní exponenciální funkcí)
  - **dirichlet priors** – pravděpodobnosti oscilují okolo dané hodnoty (*frekvence bází, substituční modely, I, ...)*
  - **lognormal priors** – např. *kalibrace stromu fosilními daty*



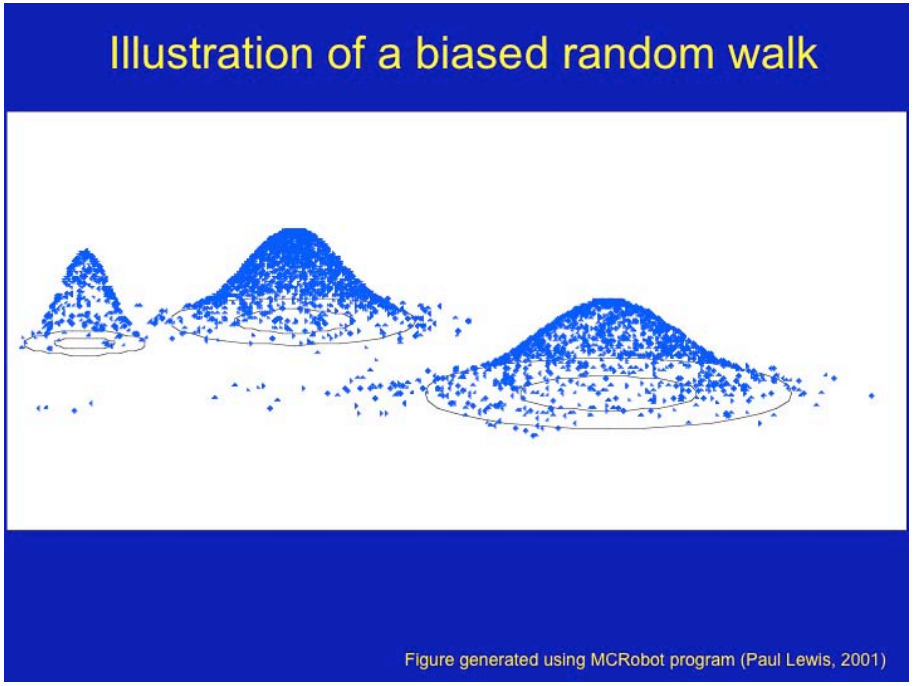
- **priors** se během analýzy mění na základě analyzovaných dat (alignment sekvencí), pomocí stochastických modelů
  ➡ získáme **posteriorní** pravděpodobnosti

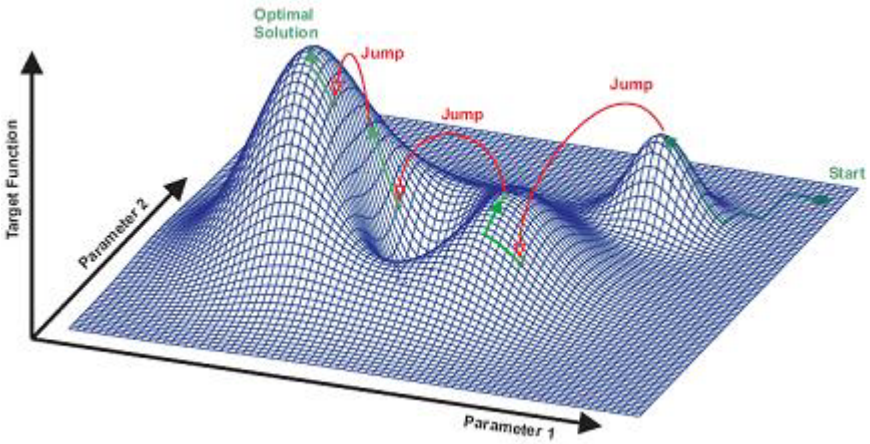


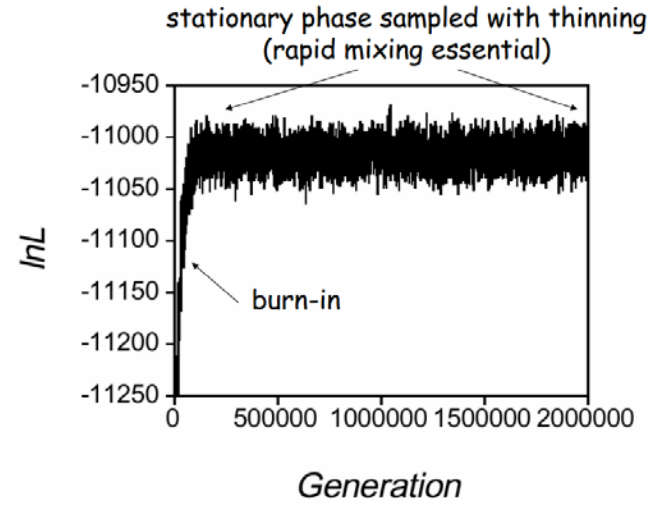


Bayesova analýza (BI)

- **Markov chain Monte Carlo (MCMC) sampling**
  - výpočet posteriorních pravděpodobností pomocí náhodných změn prior parametrů a ty buď zamítnout či přijmout na základě jejich pravděpodobností
  - Metropolic coupling MCMC = $(MC)^3$ – 1 studený a 3 horké řetězce







- **burn-in**: odstranění iniciální fáze MCMC
- **výsledná topologie**: konsenzuální strom posteriorních topologií

Bayesovská analýza – příkladový dataset

Fylogeneze rodu *Micrasterias*

- 3 partitions (18S rDNA, psaA, coxIII)
- alignment ve formátu nexus
- nastavení a spuštění analýzy
- vyhodnocení analýzy

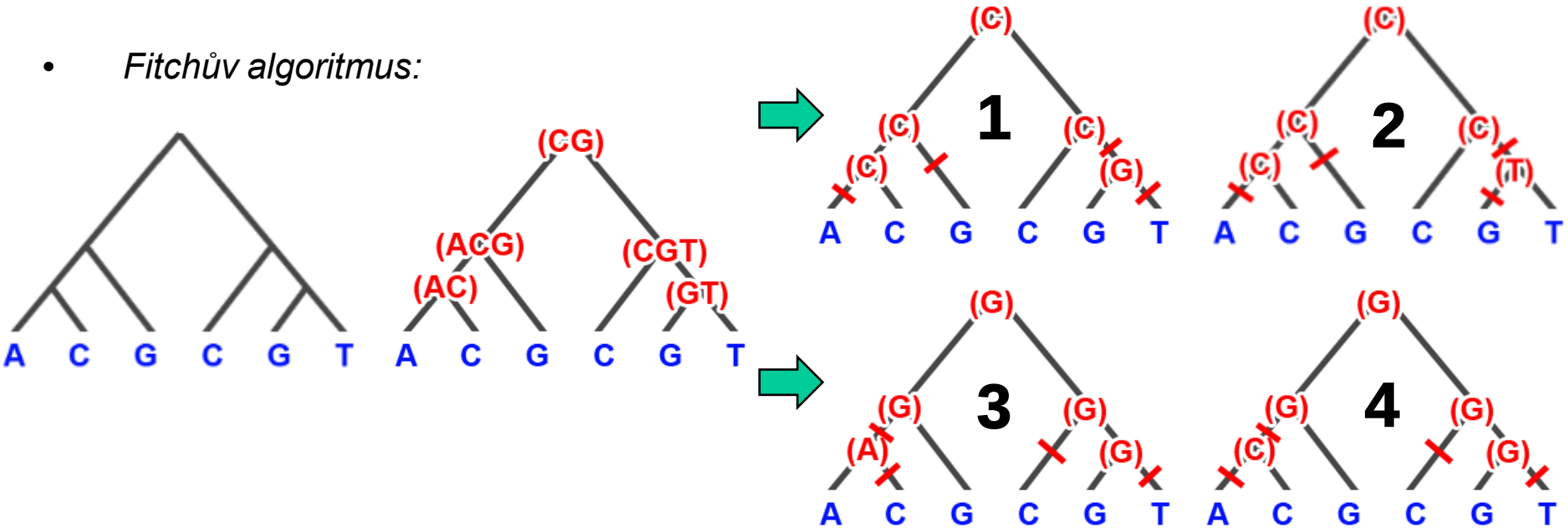
Programy

- MrBayes

Maximální parsimonie (MP)

- hledání stromu s co nejmenším počtem evolučních kroků
- heuristické prohledávání stromů stejné jako u ML: **NNI**, **SPR**, **TBR**

- *Fitchův algoritmus:*



- **vážená parsimonie** (wMP):
  - MP výpočet parsimonního skóre: **1** pro substituci, **0** pro žádnou změnu
  - wMP výpočet: každý typ substituce je vážen např. pomocí frekvence jeho výskytu (*rescaled consistency index* - méně častým mutacím je dáno vyšší skóre)

Bayesovská analýza – příkladový dataset

Fylogeneze rodu *Micrasterias*

- alignment ve formátu nexus
- nastavení a spuštění analýzy
- MP, wMP, MP bootstrapping

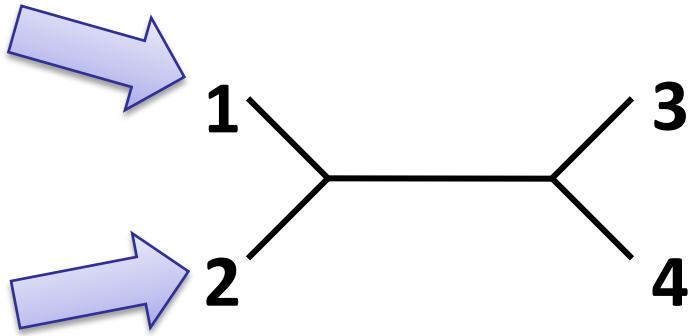
Programy

- MrBayes

Testy (in)kongruence topologií stromů

- *PROČ chceme už jednou vypočtené stromy testovat? Existuje nejistota...*
- Jak silně podporují naše data (tj. např. alignment) příbuzenské vztahy na stromu, který jsme získali?
- Je náš strom skutečně lepší než nějaký jiný?
- Je vůbec vhodné vysvětlovat příbuzenské vztahy mezi mými OTU pomocí stromu?
- *Každá data poskytnou strom, ale mohou obsahovat i zavádějící signál (saturace, nedostatek informaci v datech, artefakty apod...)*

1 ACCGAATGA
2 ACCGAGCAG
3 GTTAGGCAG
4 GTTAGATGA



Co a jak můžeme testovat?

- Testy **vycházející z původních dat** (tj. alignmentu)
  - pro testování využívají jak původní znaky, tak topologie stromů
  - výpočet rozdílů optimálních skóre (S) pro alternativní hypotézy

$$\delta \propto \mathbf{S}_{\mathrm{H}_0} - \mathbf{S}_{\mathrm{H}_1}.$$

  - **„Sitewise“ nebo také „Paired-sites“ – využívá likelihood**
    - Porovnání H_1 a H_2 – tj. dvou protichůdných hypotéz (topologií)
    - Rozdíly v testech – dle způsobu výpočtu rozložení statistiky δ.
    - KH test (Kishino-Hasegawa), SH test (Shimodaira-Hasegawa), AU test

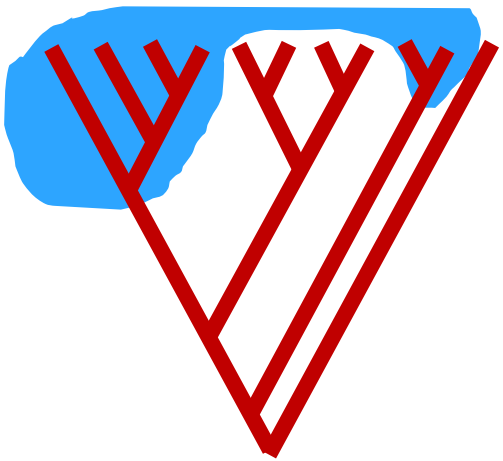
$$\delta_{\mathrm{likelihood}} = \log \mathrm{likelihood}_{\mathrm{H}_1} - \log \mathrm{likelihood}_{\mathrm{H}_2}.$$

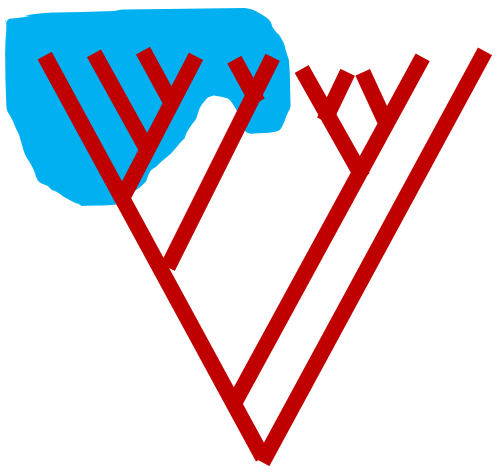
  - **ILD test – založen na parsimonii**
    - Incongruence lenght test, též „Partitions homogeneity test“
    - H_0 – strom na základě kompletního datasetu!

$$\delta_{\mathrm{ILD}} = \mathrm{Length}_{\mathrm{C}} - \sum_{i=1}^{n} \mathrm{Length}_i.$$

- Testy **porovnávajícící topologii stromů**
  - možnost porovnávat i jinak neporovnatelné datasety, např. mol./morfol
  - Tanglegramy, + různé indexy na základě vzdáleností – málo spolehlivé

Testy topologických hypotéz





L1 **L0**

δ= ln L1-lnL0

Je L1 signifikantně vyšší než L0? Potřebujeme znát rozložení δ….

- Testovat můžeme např. topologii stromu získanou pomocí molekulárních dat, vůči topologii získané z dat morfologických
- Porovnávat **stromy s constrains** (např. preferujeme některé vztahy apriori) **vůči stromům bez apriorních preferencí**

Testy topologických hypotéz

- Obecně se nejprve určí rozdíl věrohodnosti mezi dvěma topologiemi.

$$\delta = \ln L1 - \ln L0$$

- Rozložení δ statistiky lze určit různě -> největší rozdíly mezi testy.

- **KH test (Kishino-Hasegawa test)** - počítá rozložení δ analyticky (Možné vypočítat pomocí PAUP, Treepuzzle, příp. PHYLIP).

- **SH test (Shimodaira-Hasegawa)** – obdoba KH, je jednostranný a tedy silnější.

- **AU TEST** (approximately unbiased test)
  - V současnosti nejuznávanějším testem
  - obsažený v programu Consel.
  - Tento test používá „resampling“ metody (podobné bootstrapingu).
  - Pro testované topologie se vypočtou „likelihoody“ pro jednotlivé pozice alignmentu. (návod viz http://web.natur.cuni.cz/~vlada/moltax/blok6.htm)

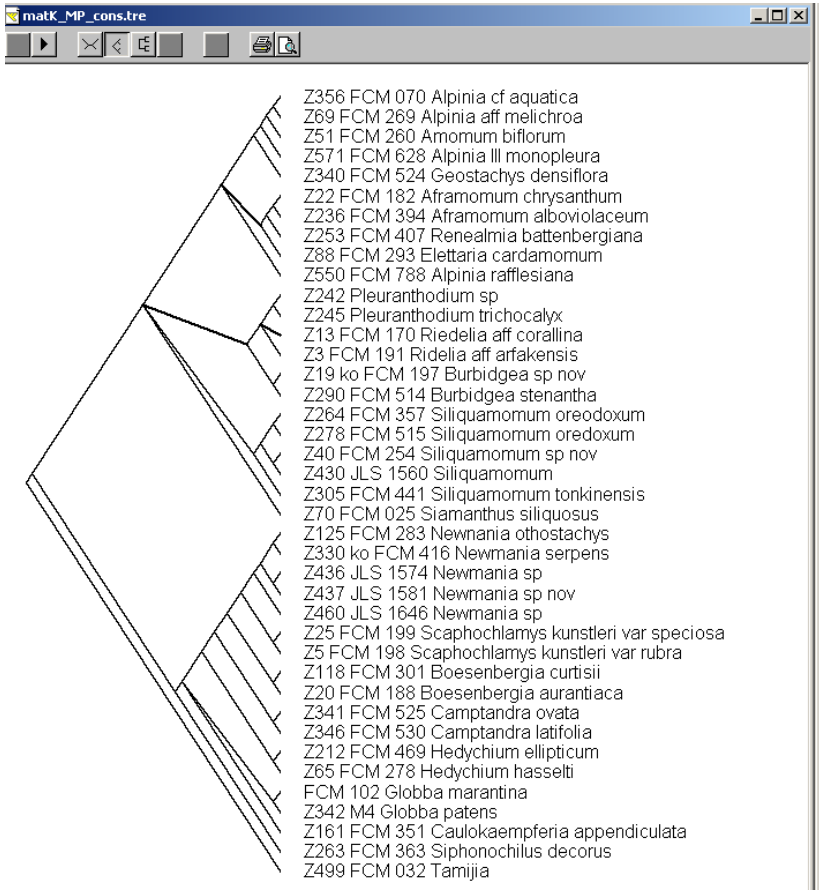
Literatura

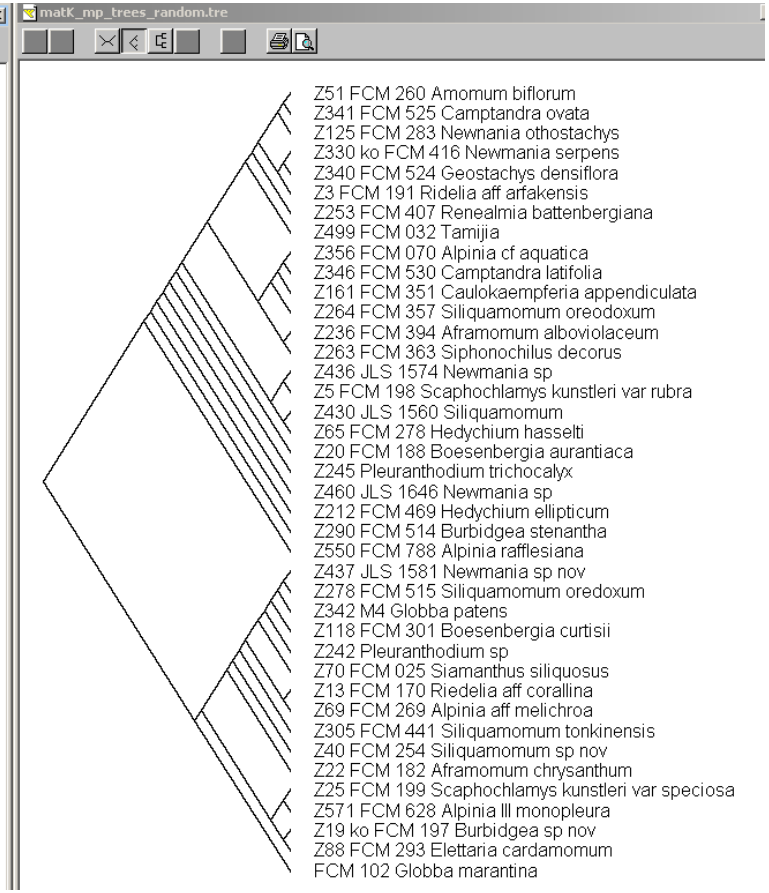
- Planet, 2005
- Goldman et al., 2000

Programy

- PAUP
- TreePuzzle
- Consel

Testy topologických hypotéz - příklad






```
Kishino-Hasegawa test:
  KH test using normal approximation, two-tailed test

                                        KH-test
    Tree          -ln L     Diff -ln L       P
-----------------------------------------------
       1     8908.26955         (best)
       2    12009.93169     3101.66215   0.000*
  * P < 0.05

Processing of file "D:\programy\PAUP\matK_commands_KH.nex" completed.
```

Programy

- PAUP
- TreePuzzle
- Consel

Testy homogenity datasetu – ILD test

- *Když chceme věďet zda:*

1) Dopomůže spojení datasetů ke zvýšení přesnosti fylogenetické analýzy?
2) Prodělaly jednotlivé partition různé evoluční procesy, nebo se vyvíjely jinak rychle?
3) Prodělaly jednotlivé partition různou evoluční historii (hybridizace, genová duplikace...)?

- *Udává míru kongruence mezi znaky v rámci datasetu – na základě parsimonie*

- Porovnává situaci, kdy všechny partition musí odpovídat jednomu stromu (H0), a situaci, kdy každá partition může mít vlastní strom (H1)

$$\delta_{\mathrm{ILD}} = \mathrm{Length}_{\mathrm{C}} - \sum_{i=1}^{n} \mathrm{Length}_i .$$

- Silná neshoda mezi partitions –> **LenghtC** - daleko větší počet kroků ve stromu než je součet délek stromů z jednotlivých partition

- Problém s H0 – může být nereálná
- + spoustu dalších nevýhod, ale může být dobrý startovací test, když tušíme problém v datech

ILD Test – příkladový dataset

- Test pracuje na datasetu, který je rozdělen na dvě či více partition

```
Begin sets;
  charset CHS_exon = 1-959;
  charset matK = 960-3872;
  charpartition genes = CHS_exon:CHS_exon, matK:matK;
End;

Begin PAUP;
hompart partition=genes nreps=500 / start=stepwise addseq=random nreps=10 savereps=no randomize=addseq rstatus=no
hold=1 swap=tbr multrees=yes;
log stop;
end ;
```

- P hodnota < 0.05 → **inkongruence mezi partitions**

```
500 partition-homogeneity test replicates completed
Note: Effectiveness of search may have been diminished due to tree-buffer
      overflow.
Time used = 10:49:10.3

Results of partition-homogeneity test:

       Sum of       Number of
 tree lengths       replicates
 ----------------------------
        1226*            1
        1228             4
        1229             5
        1230            13
        1231            28
        1232            47
        1233            60
        1234            85
        1235            64
        1236            87
        1237            71
        1238            29
        1239             5
        1240             1
            * = sum of lengths for original partition

P value = 1 - (499/500) = 0.002000
```

Programy

- PAUP

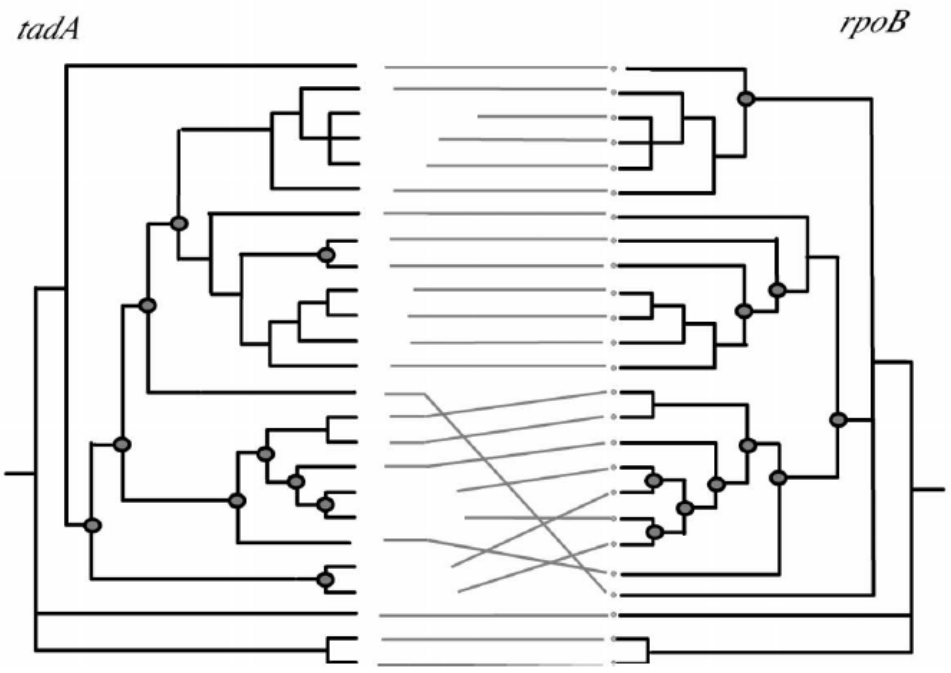
Testy porovnávající pouze topologie

- Míra kongruence stromů – odvozena pouze z pattern větvení stromů
- Někdy jso brány v úvahu i délky větví

- *Nebere v úvahu původní data (nebo např. podpory jednotlivých větví!)*

- **Mohou ale sloužit jako praktický pomocník!**

- Když chceme *VIĎET* rozsah inkongruence – zobrazeni pomocí **tanglegramů**



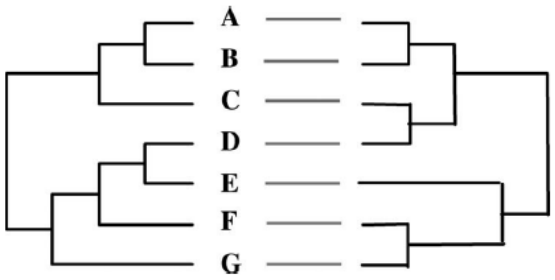
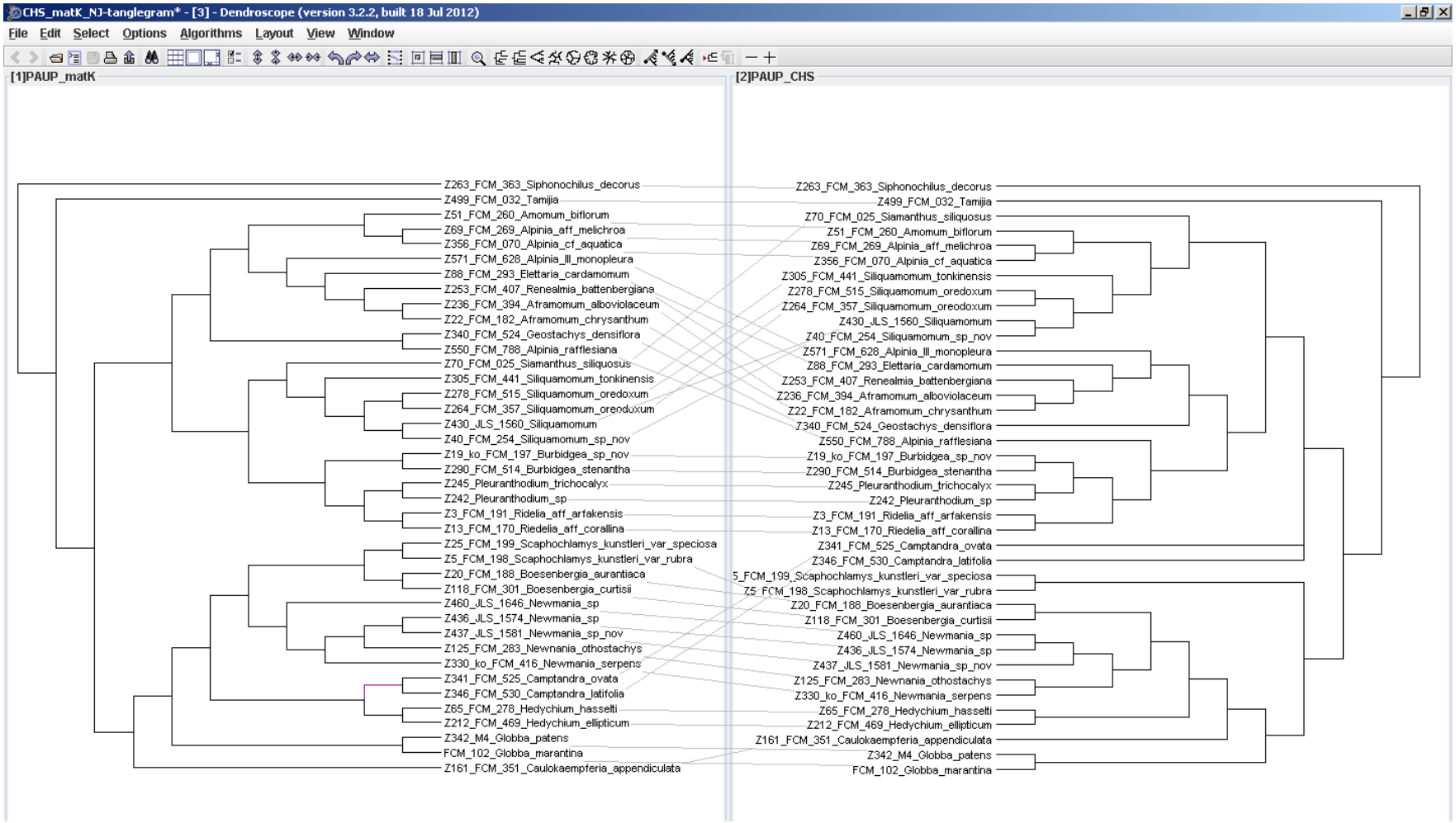


Fig. 2. Problems with tanglegrams. Shown is a tanglegram between two incongruent tree topologies. Note that the lines connecting corresponding terminals do not cross.

Vizualizace inkongruence stromů pomocí tanglegramů



- V programu Dendroscope otevřeme nexus soubor obsahující porovnávané stromy
- Algorithms → Tenglegrams

Programy

- Dendroscope

Vizualizace stromů

- Různé formáty stromů (spíše než na metodě záleží na použitém programu)
- **Newick (Phylip)**

```
(,,(,));                                  no nodes are named
(A,B,(C,D));                              leaf nodes are named
(A:0.1,B:0.2,(C:0.3,D:0.4):0.5);          distances and leaf names
```

```
((((((((Z264_FCM_357_Siliquamomum_oreodoxum:1.0,Z278_FCM_515_Siliquamomum_oredoxum:1.0):1.0,(Z40_FCM_254_Siliqua
_sp_nov:1.0,Z430_JLS_1560_Siliquamomum:1.0):1.0,Z305_FCM_441_Siliquamomum_tonkinensis:2.0):1.0,((Z356_FCM_070_Alp
cf_aquatica:1.0,Z69_FCM_269_Alpinia_aff_melichroa:1.0):1.0,Z51_FCM_260_Amomum_biflorum:2.0):1.0):2.0,((Z19_ko_FCM
Burbidgea_sp_nov:1.0,Z290_FCM_514_Burbidgea_stenantha:1.0):1.0,(Z242_Pleuranthodium_sp:1.0,Z245_Pleuranthodium_tr
alyx:1.0):1.0,Z13_FCM_170_Riedelia_aff_corallina:2.0,Z3_FCM_191_Ridelia_aff_arfakensis:2.0):3.0,((((Z22_FCM_182_A
mum_chrysanthum:1.0,Z236_FCM_394_Aframomum_alboviolaceum:1.0):1.0,Z253_FCM_407_Renealmia_battenbergiana:2.0):1.0,
CM_293_Elettaria_cardamomum:3.0):1.0,Z571_FCM_628_Alpinia_III_monopleura:4.0):1.0,Z340_FCM_524_Geostachys_densifl
.0,Z550_FCM_788_Alpinia_rafflesiana:5.0):1.0,Z70_FCM_025_Siamanthus_siliquosus:6.0):1.0,(Z341_FCM_525_Camptandra_
:1.0,Z346_FCM_530_Camptandra_latifolia:1.0):6.0):1.0,(((Z125_FCM_283_Newnania_othostachys:1.0,Z330_ko_FCM_416_New
_serpens:1.0,Z436_JLS_1574_Newmania_sp:1.0,Z437_JLS_1581_Newmania_sp_nov:1.0,Z460_JLS_1646_Newmania_sp:1.0):1.0,(
FCM_301_Boesenbergia_curtisii:1.0,Z20_FCM_188_Boesenbergia_aurantiaca:1.0):1.0):1.0,((Z212_FCM_469_Hedychium_elli
m:1.0,Z65_FCM_278_Hedychium_hasselti:1.0):1.0,Z161_FCM_351_Caulokaempferia_appendiculata:2.0):1.0):5.0,(Z25_FCM_1
aphochlamys_kunstleri_var_speciosa:1.0,Z5_FCM_198_Scaphochlamys_kunstleri_var_rubra:1.0):7.0,(FCM_102_Globba_mara
:1.0,Z342_M4_Globba_patens:1.0):7.0):1.0,Z499_FCM_032_Tamijia:9.0):1.0,Z263_FCM_363_Siphonochilus_decorus:10.0);(
(((Z264_FCM_357_Siliquamomum_oreodoxum:1.0,Z278_FCM_515_Siliquamomum_oredoxum:1.0):1.0,(Z40_FCM_254_Siliquamomum_
v:1.0,Z430_JLS_1560_Siliquamomum:1.0):1.0,Z305_FCM_441_Siliquamomum_tonkinensis:2.0):1.0,((Z356_FCM_070_Alpinia_c
atica:1.0,Z69_FCM_269_Alpinia_aff_melichroa:1.0):1.0,Z51_FCM_260_Amomum_biflorum:2.0):1.0):3.0,((Z19_ko_FCM_197_E
gea_sp_nov:1.0,Z290_FCM_514_Burbidgea_stenantha:1.0):1.0,(Z242_Pleuranthodium_sp:1.0,Z245_Pleuranthodium_trichoca
.0):1.0,Z13_FCM_170_Riedelia_aff_corallina:2.0,Z3_FCM_191_Ridelia_aff_arfakensis:2.0):4.0,(((((Z22_FCM_182_Aframc
hrysanthum:1.0,Z236_FCM_394_Aframomum_alboviolaceum:1.0):1.0,Z253_FCM_407_Renealmia_battenbergiana:2.0):1.0,Z88_F
3_Elettaria_cardamomum:3.0):1.0,Z571_FCM_628_Alpinia_III_monopleura:4.0):1.0,Z340_FCM_524_Geostachys_densiflora:5
.0,Z550_FCM_788_Alpinia_rafflesiana:6.0):1.0,Z70_FCM_025_Siamanthus_siliquosus:7.0):1.0,(Z341_FCM_525_Camptandra_
:1.0,Z346_FCM_530_Camptandra_latifolia:1.0):7.0):1.0,((((Z125_FCM_283_Newnania_othostachys:1.0,Z330_ko_FCM_416_Ne
a_serpens:1.0,Z436_JLS_1574_Newmania_sp:1.0,Z437_JLS_1581_Newmania_sp_nov:1.0,Z460_JLS_1646_Newmania_sp:1.0):1.0,
_FCM_301_Boesenbergia_curtisii:1.0,Z20_FCM_188_Boesenbergia_aurantiaca:1.0):1.0):1.0,((Z212_FCM_469_Hedychium_ell
um:1.0,Z65_FCM_278_Hedychium_hasselti:1.0):1.0,Z161_FCM_351_Caulokaempferia_appendiculata:2.0):1.0):1.0,(Z25_FCM_
caphochlamys_kunstleri_var_speciosa:1.0,Z5_FCM_198_Scaphochlamys_kunstleri_var_rubra:1.0):3.0):5.0,(FCM_102_Globb
antina:1.0,Z342_M4_Globba_patens:1.0):8.0):1.0,Z499_FCM_032_Tamijia:10.0):1.0,Z263_FCM_363_Siphonochilus_decorus:
;
```

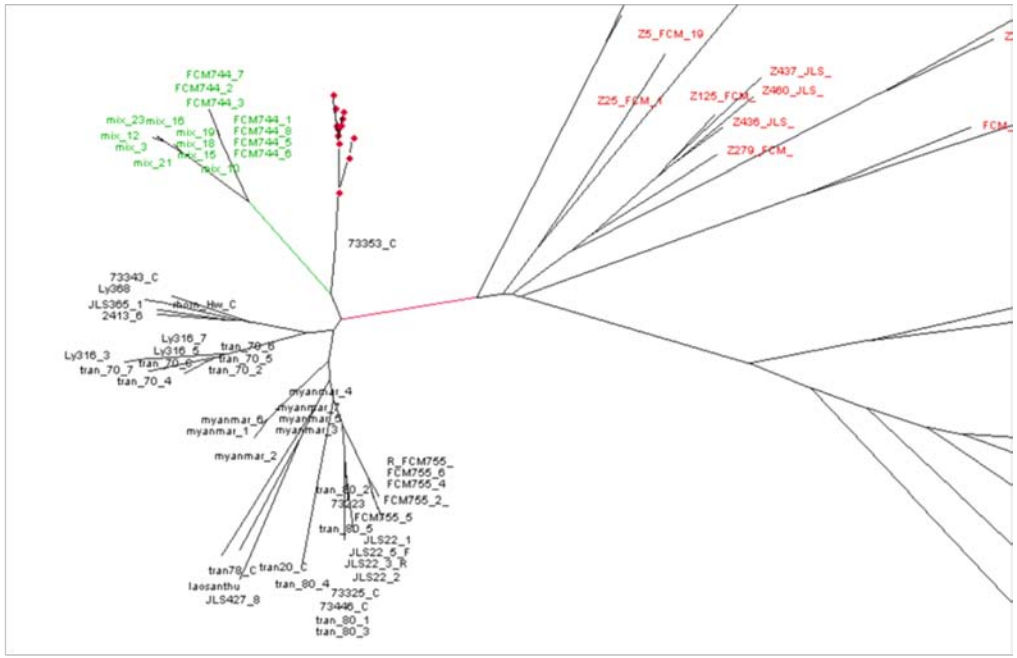
- **Nexus**

```
#NEXUS

BEGIN TREES;
        TRANSLATE
                1       Z118_FCM_301_Boesenbergia_curtisii,
                2       Z125_FCM_283_Newnania_othostachys,
                3       Z330_ko_FCM_416_Newmania_serpens,
                4       Z436_JLS_1574_Newmania_sp,
                5       Z437_JLS_1581_Newmania_sp_nov,
                6       Z460_JLS_1646_Newmania_sp,
                7       Z13_FCM_170_Riedelia_aff_corallina,

                33      Z499_FCM_032_Tamijia,
                34      Z263_FCM_363_Siphonochilus_decorus,
                35      FCM_102_Globba_marantina,
                36      Z342_M4_Globba_patens,
                37      Z161_FCM_351_Caulokaempferia_appendiculata,
                38      Z212_FCM_469_Hedychium_ellipticum,
                39      Z65_FCM_278_Hedychium_hasselti,
                40      Z20_FCM_188_Boesenbergia_aurantiaca
                ;
        TREE * Strict=
(34,(33,(((28,(26,27,(7,8,(9,10),(11,12)),(17,(16,(15,(13,14)))),((20,(18,19),(21,22)),(23,(24,25))))),(29,30)),(31,32
),(35,36),((37,(38,39)),((2,3,4,5,6),(1,40))))));
        TREE MajRule=
(34,(33,(((28,(27,(7,8,(9,10),(11,12)),26,(17,(16,(15,(13,14))))),((20,(18,19),(21,22)),(23,(24,25))))),(29,30)),(35,
36),((31,32),((37,(38,39)),((2,3,4,5,6),(1,40))))));
ENDBLOCK;
```

Vizualizace stromů

- Výsledné stromy z analýz NJ, MP, ML nebo Bayesovské analýzy (resp. z programů pro tyto analýzy používaných) – nejčastěji ve formátu NEXUS
- Programy pro práci se stromy – zobrazují i **doplňující informace**
  - délky větví
  - Boostrap hodnoty
  - PP hodnoty

- **Editace stromů**
  - Barvení větví
  - Rotace větví
  - Zakořenění stromů



Programy

- TreeView
- FigTree
- Splitstree

Praktické cvičení – konstrukce stromů

- Na základě jednoho alignmentu vytvořte fylogenetické stromy pomocí dvou (nebo i více) metod pro rekonstrukci fylogeneze (např. MP, ML, MrBayes)

  - vstupní soubor - alignment ve formátu nexus (vlastní data nebo příkladové datasety - **Micrasterias.nex, CHS_Exon.nex**, nebo **matK.nex**)

  - Pro analýzu MP použijte např. PAUP
  - Pro analýzu ML použijte např. Garli
  - Pro Bayesovskou analýzu použijte např. MrBayes

  - Spuštění analýz konzultujte s návodem...

- Po skončení analýz porovnejte výsledné stromy
  - identifikujte rozdíly v topologiích (použijte např. Dendroscope)
  - Porovnejte získané podpory pro jednotlivé clady

- Jak budete interpretovat hodnoty PP < 0.95 získané z analýzy pomocí MrBayes?
- Jak budete interpretovat hodnoty bootstrapu kolem 70% ?

Praktické cvičení – porovnání topologií z různých metod

- Otestujte topologie dvou stromů získaných např. z analýzy MP a MrBayes pomocí některého z testů využívajících ML (např. KH test, tj. Kishino-Hasegawa test :)

  - Vložte porovnávané stromy (výsledné stromy z MP analýzy [např. strict consensus tree] a Bayesovské analýzy) do jednoho souboru. Zachovejte formát nexus.
  - Upravte nexus soubor s alignmentem pro výpočet KH testu v programu PAUP přidáním bloku příkazů (viz návod) a uložte do stejné složky, kde máte uložen
  - Vložte soubor s porovnávanými daty a alignment s příkazy do složky, kde máte PAUP executable soubor (tj. „win-paup4b10.exe“)
  - Spusťte PAUP a načtěte alignment s příkazy

- O čem vypovídá výsledná p-hodnota?

Pokud nemáte svá data, nebo nedoběhly všechny potřebné analýzy, použijte jako vstupní soubory “**matK_commands_KH.nex** “ a soubor se stromy “**matk.tre**”

Praktické cvičení – inkongruence v rámci datasetu

- Otestujte alignment obsahující data z různých markerů/lokusů pomocí ILD testu a zjistěte, zda všechny markery navrhují stejnou fylogenezi, či navrhují–li signifikantně odlišné topologii stromu
  - vstupní soubor - konkatenovaný alignment rozdělený na dvě a více partitions v nexus formátu (vlastní data, příp. **Micrasterias.nex** nebo **CHS_matK_concatenated.nex**)
  - Ke vstupnímu souboru **připojte příkazy pro ILD** test a spusťte v programu PAUP (**viz návod**)

- Jaký postup zvolíte, pokud vyjde test nesignifikantně (tj. P>0.005)

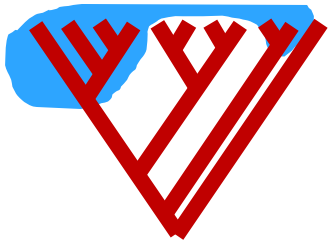
- Jaký postup zvolíte, pokud vyjde test signifikantně (P<0.005)?

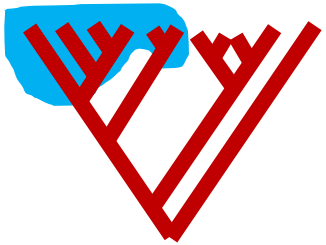
- Vypočtěte stromy na základě jednotlivých partition a topologie porovnejte pomocí tanglegramů, nebo jen pouhým okem
  - Pro výpočet stromů zvolte libovolnou metodu (např. MP v PAUP)

- Odstraňte z konkatenovaného datasetu jedince, kteří zjevně generují inkongruenci strumů a opakujte ILD test na redukovaném datasetu

Testy topologických hypotéz

AU test





	$L_1L_2L_3L_4L_5L_6$
A	catcga
B	ccgggt
C	gcggga

Vypočteme „site likelihoods“ $L_1, L_2, L_3, L_4, L_5, L_6$ $L_1, L_2, L_3, L_4, L_5, L_6$

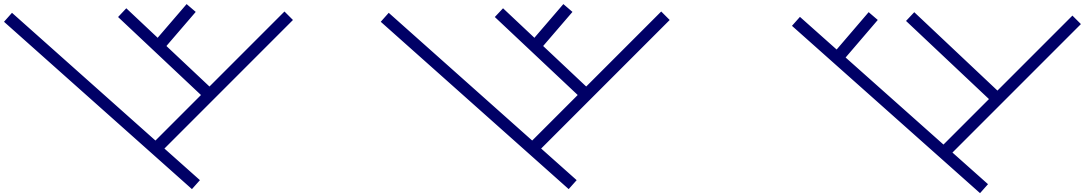
Provedeme permutaci „site likelihoods“ a vypočteme celkový Likelihood $L1= L_1*L_2* L_2* L_3*L_4* L_2$ $L0= L_1*L_1* L_6* L_3*L_4* L_5$

Spočítáme δ $\delta= \ln L1-\ln L0$

Opakujeme mnohokrát **Procento případů, kdy δ ≤ 0 je hodnota p s jakou můžeme H0 zavrhnout**

Partitions homogeneity test (Incongruence lenght difference - ILD)

	gen A	gen B	genC
taxon 1	A A G G C C T T	A A G G C C T T	A A G G C C T T
taxon 2	A G G G C C T T	A G G G C C T T	A G G G C C T T
taxon 3	A G T T C C A T	A G T T C C A T	A G T T C C A T
taxon 4	A G T T C A G T	A G T T C A G T	A G T T C A G T



porovnání rozdílů mezi topologiemi

	gen A	gen B	genC
taxon 1	A A G G C C T T	A A G G C C T T	A A G G C C T T
taxon 2	A G G G C C T T	A G G G C C T T	A G G G C C T T
taxon 3	A G T T C C A T	A G T T C C A T	A G T T C C A T
taxon 4	A G T T C A G T	A G T T C A G T	A G T T C A G T