

5. Kvantitativní kladistika

- Hennigova metoda konstrukce kladogramu
- kódování znaku pro kladistickou analýzu
- algoritmy konstrukce fylogenetických stromů na základě kritéria parsimonie
- optimalizační kritéria
- popisné charakteristiky a porovnávání stromů
- konsenzuální stromy
- míry spolehlivosti výsledků kladistické analýzy
- kladistický software

Postup při fylogenetické analýze

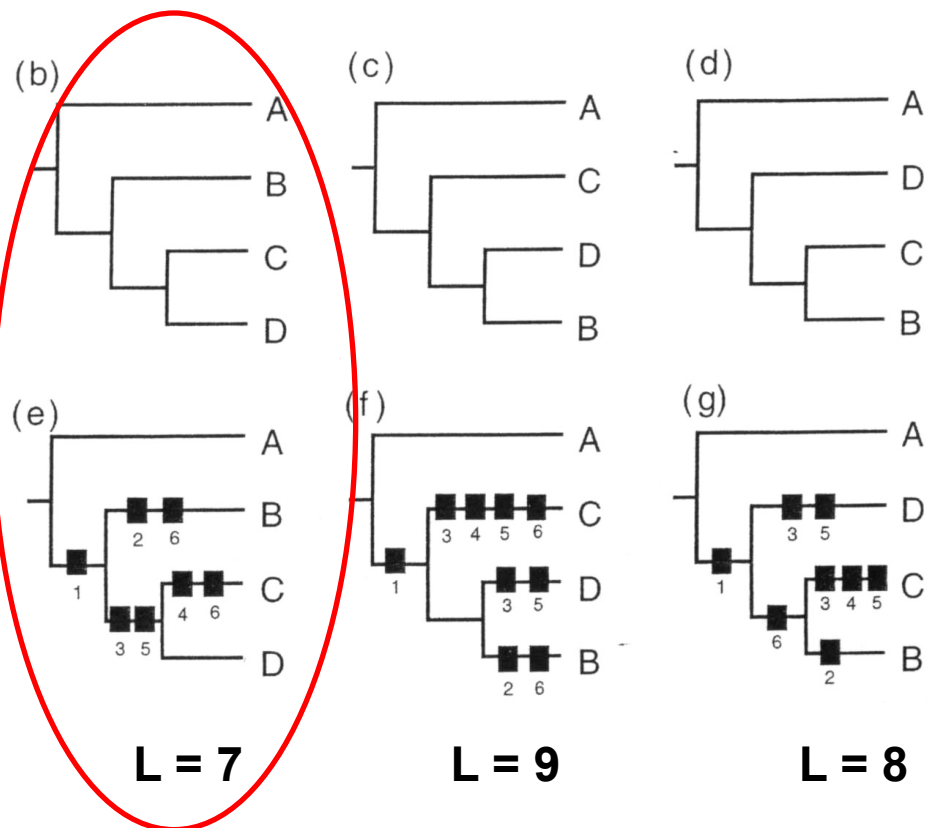
- 1) výběr taxonů
- 2) výběr znaků (posouzení homologie)
- 3) kódování znaků do matice
- 4) polarizace znaků (mimoskupinové srovnávání, event. ontogenetická pravidla)
- 5) vlastní analýza – konstrukce kladogramu

Parsimonie v kladistice: hledáme strom, na kterém rozložení znaků předpokládá minimum evolučních změn (*minimum evolution*), tj. takový, který má nejmenší délku (= *Hennigův pomocný princip*: nikdy dopředu nepředpokládejme homoplázií)

(a)

TAXA	CHARACTERS					
	1	2	3	4	5	6
A	□	□	□	□	□	□
B	■	■	□	□	□	■
C	■	□	■	■	■	■
D	■	□	■	□	■	□

délka stromu (*tree length*) – celkový součet všech změn ve všech znacích (kroků) na dané topologii stromu, tj. nutných k vysvětlení dané hypotézy o příbuznosti



Hennigova metoda

- každý znak při stavbě kladogramu posuzujeme zvlášť
- polaritu znaku (apomorfie/plesiomorfie) určujeme dopředu
- taxony seskupujeme na základě synapomorfii
- rychlá a jednoduchá, ale jen pro malé soubory s malým množstvím homoplázií

	Homalozoa (Outgroup)	Echinoid	Asteroid	Crinoid	Holothuroid	Ophiuroid
Number of ambulacral grooves	3	5	5	5	5	5
Suckers	Not on podia	On podia	On podia	Absent	On podia	On podia
Closed ambulacral grooves	No	Yes	Yes	No	Yes	Yes
Aboral surface reduced	No	Yes	No	No	Yes	Yes
Ambulacral grooves extended from oral to aboral	No	Yes	No	No	Yes	No

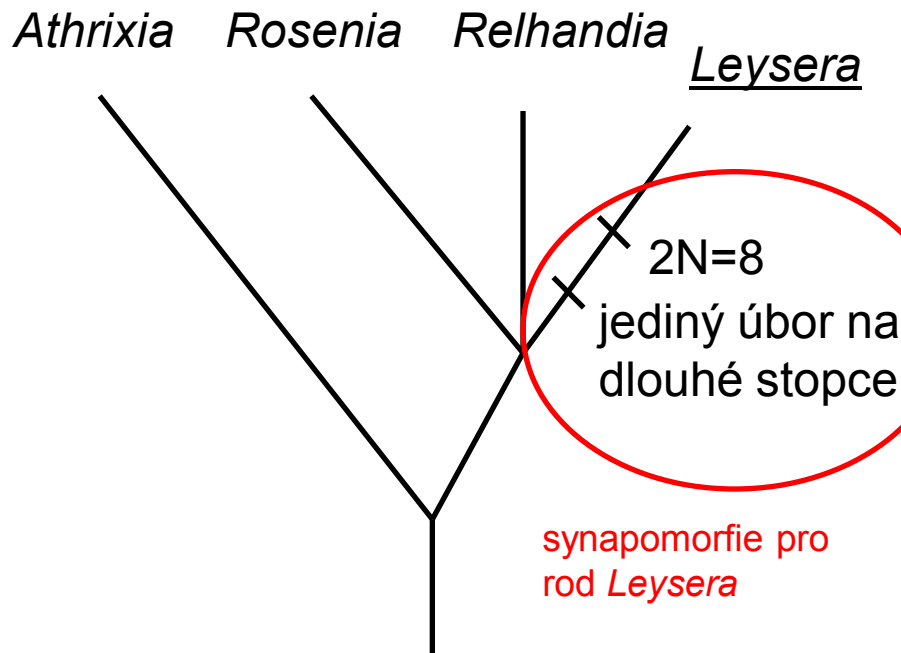
Příklad: rod *Leysera* (Bremer 1978)

- hvězdnicovité rostliny (4 druhy v JAR a Středomoří)

Taxon	Lůžko	Trubkovité květy	Chmýr	Povrch nažky	Šupiny na chmýru	Životní cyklus
<i>L. longipes</i>	hladké	se žlázkami	osinatý	hladký	šídlovité	trvalka
<i>L. leyseroides</i>	drsne	ochlupené	opeřený	drsne	široké, ploché	jednoletka
<i>L. tenella</i>	drsne	ochlupené	opeřený	drsne	široké, ploché	jednoletka
<i>L. gnaphaloides</i>	drsne	ochlupené	opeřený	drsne	šídlovité	trvalka

Příklad: rod *Leysera* (Bremer 1978)

Polarizace znaků pomocí
mimoskupinového srovnávání
(*outgroup comparison*):



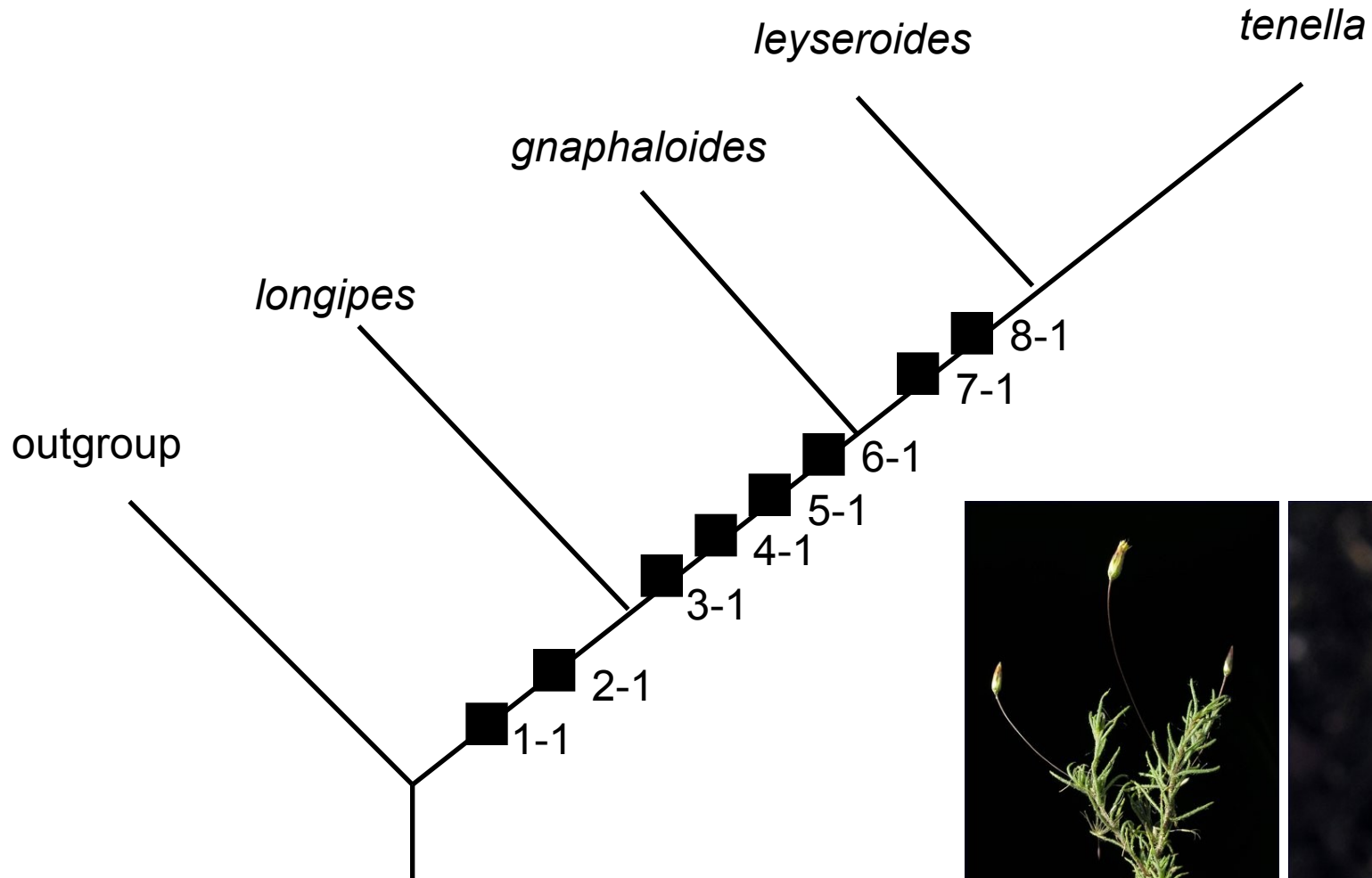
podíváme se, jaké stavy znaků se
vyskytují u sesterských rodů *Leysera*: ty
budeme považovat za plesiomorfní



plesiomorfní znaky:

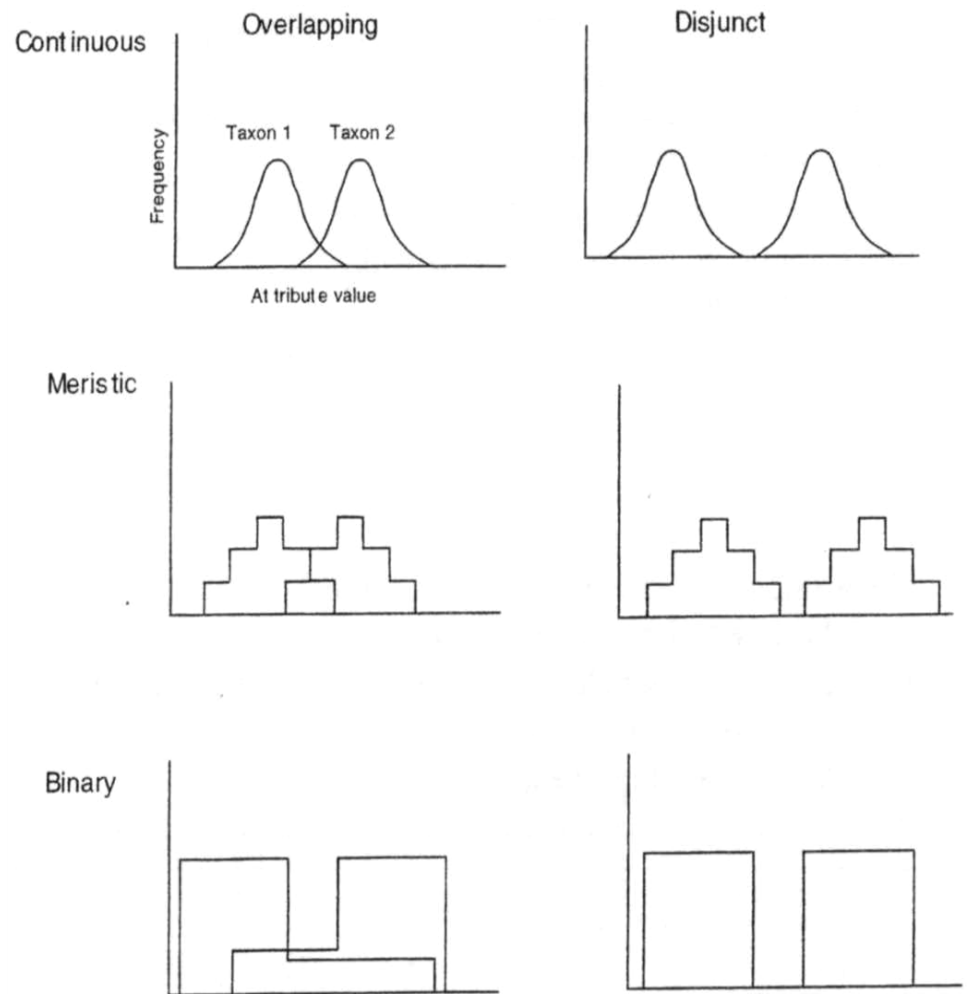
- lůžko hladké
- trubkovité květy se žlázkami
- chmýr osinatý
- povrch nažky hladký
- šupiny na chmýru šídlovité
- životní cyklus vytrvalý

Výsledný kladogram



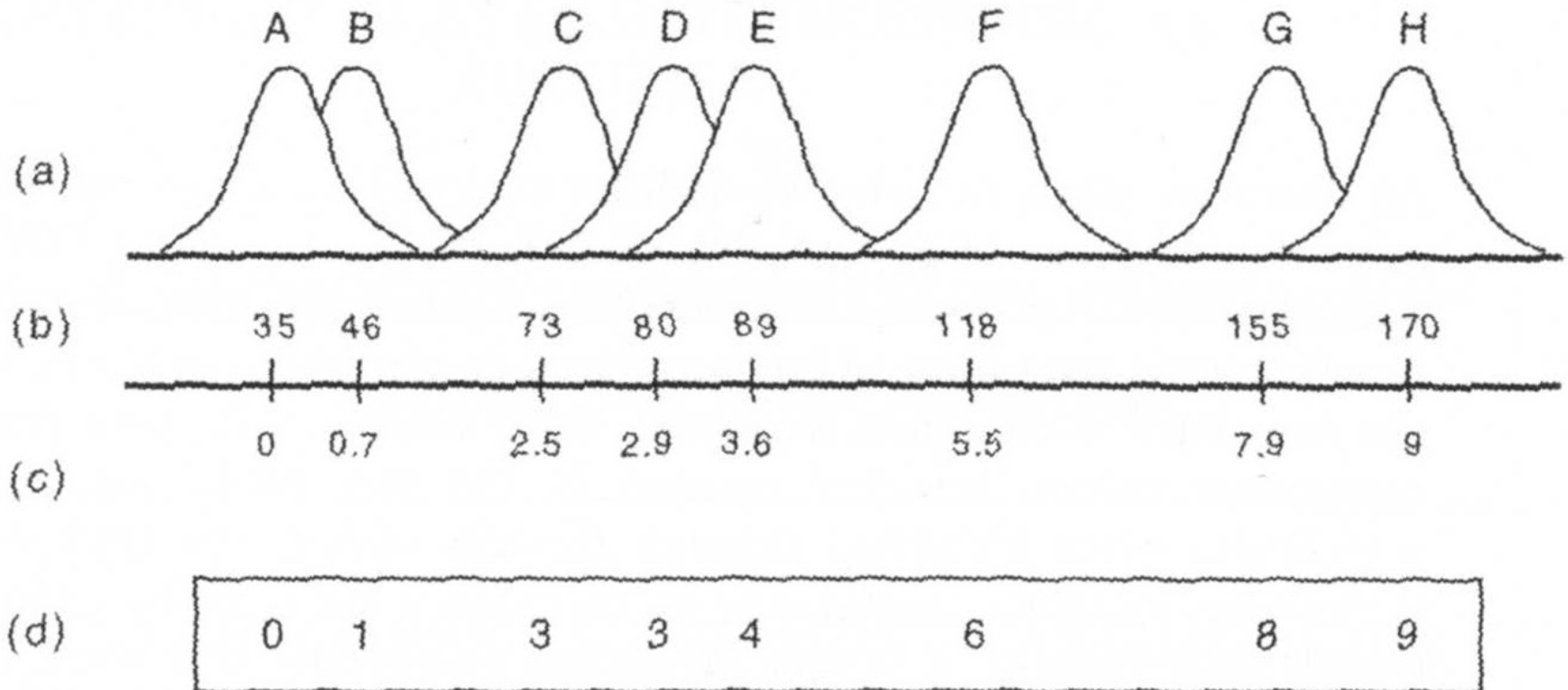
Kódování znaků

- převedení informace do formy vhodné pro kladistickou analýzu
- celá čísla (0, 1, 2, 3,9)
- musíme od sebe odlišit jednotlivé stavy
- znaky, které se nepřekrývají, jsou vhodnější
- v praxi jsou nejčastěji využívány binární znaky (0,1)



Kódování spojitých znaků

- různé metody, např. „gap-weighting“ (Thiele 1993)
- v praxi relativně málo používané



Kódování diskrétních znaků

vícestavové
kódování:
jeden znak se
závislými
stavy

vícestavové
kódování:
dva nezávislé
znaky (barva,
tvar)

hierarchické
kódování:
tři znaky
(přítomnost,
barva, tvar)

binární kódování:
pět nezávislých znaků
(přítomnost/
nepřítomnost)



0

0 0

0 ? ?

0 0 0 0 0

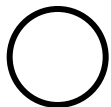


1

1 1

1 0 0

1 1 0 1 0



2

1 2

1 0 1

1 1 0 0 1

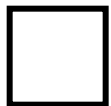


3

2 1

1 1 0

1 0 1 1 0



4

2 2

1 1 1

1 0 1 0 1

Chybějící data (*missing values*)

- a) nemáme údaje k dispozici (neznámá vývojová stádia, pohlaví, fosílie, poškozené kusy)
 - b) znaky nelze kódovat, protože příslušné struktury u organismů přirozeně chybí
- kódujeme “?”, resp. ”-”
 - problémy: naroste počet řešení, klesá rozlišení kladogramů, někdy dokonce zcela chybné výsledky

- počet všech možných rozlišených stromů pro t taxonů

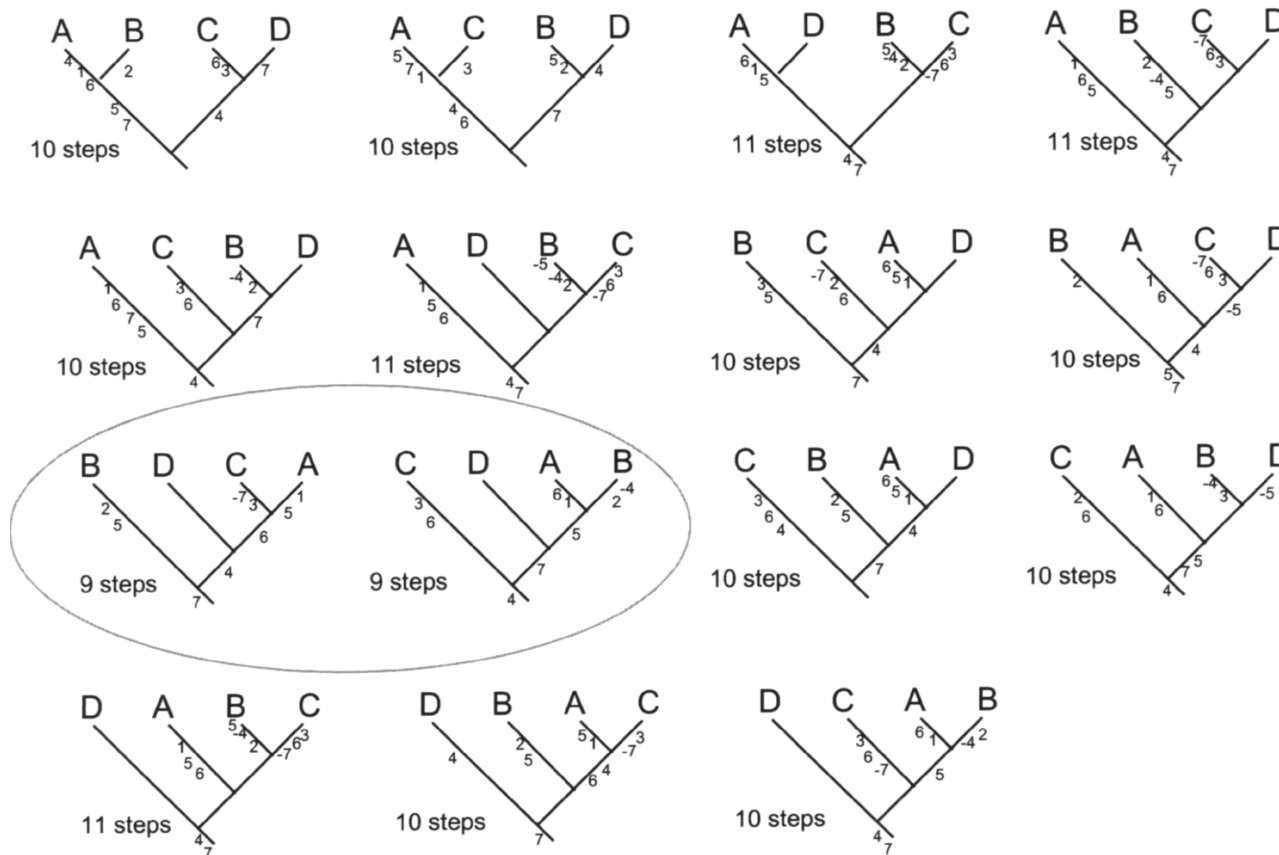
$$n = (2t - 5)! / [2^{t-3} (t - 3)!]$$

Species	Number of trees
1	1
2	1
3	3
4	15
5	105
6	945
7	10,395
8	135,135
9	2,027,025
10	34,459,425
11	654,729,075
12	13,749,310,575
13	316,234,143,225
14	7,905,853,580,625
15	213,458,046,676,875
16	6,190,283,353,629,375
17	191,898,783,962,510,625
18	6,332,659,870,762,850,625
19	221,643,095,476,699,771,875
20	8,200,794,532,637,891,559,375
30	4.9518×10^{38}
40	1.00985×10^{57}
50	2.75292×10^{76}

Počítačové algoritmy hledání nejúspornějších kladogramů

Přesné metody (*exact methods*):

- vyčerpávající hledání (*exhaustive search, implicit enumeration*) – jen pro malé datové soubory (do 10 taxonů – více než 2 miliony stromů)

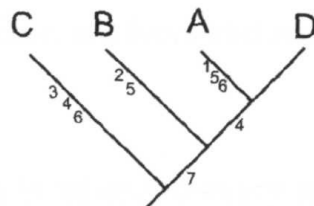


Přesné metody (*exact methods*):

- *branch-and-bound search* – jako vstupní minimum se vezme např. náhodný strom; pokud při hledání délka stromu přeskočí limit, celá skupina stromů se zavrhne a zkouší se jiná (realizovatelné do 25 taxonů)

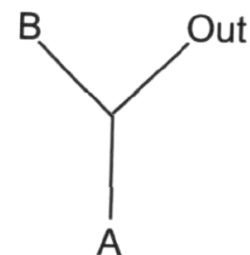
1)

	1	2	3	4	5	6	7
A	1	0	0	1	1	1	1
B	0	1	0	0	1	0	1
C	0	0	1	1	0	1	0
D	0	0	0	1	0	0	1
Out	0	0	0	0	0	0	0

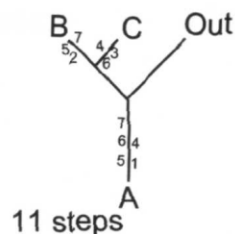


(Length=10)

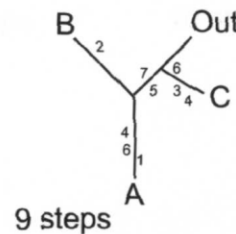
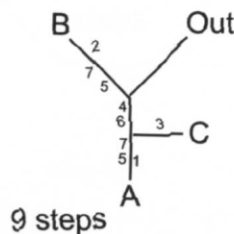
2)



3)



Upper bound reached

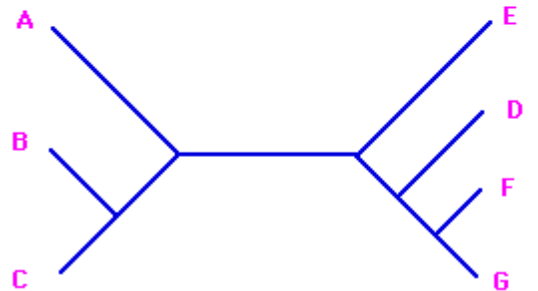
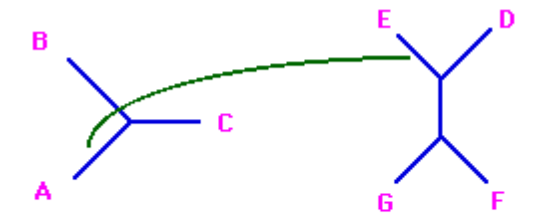
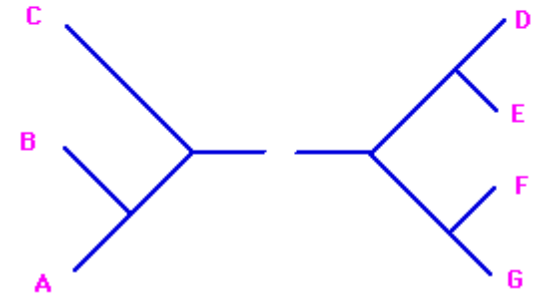
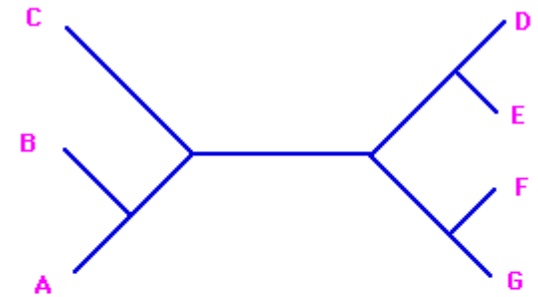
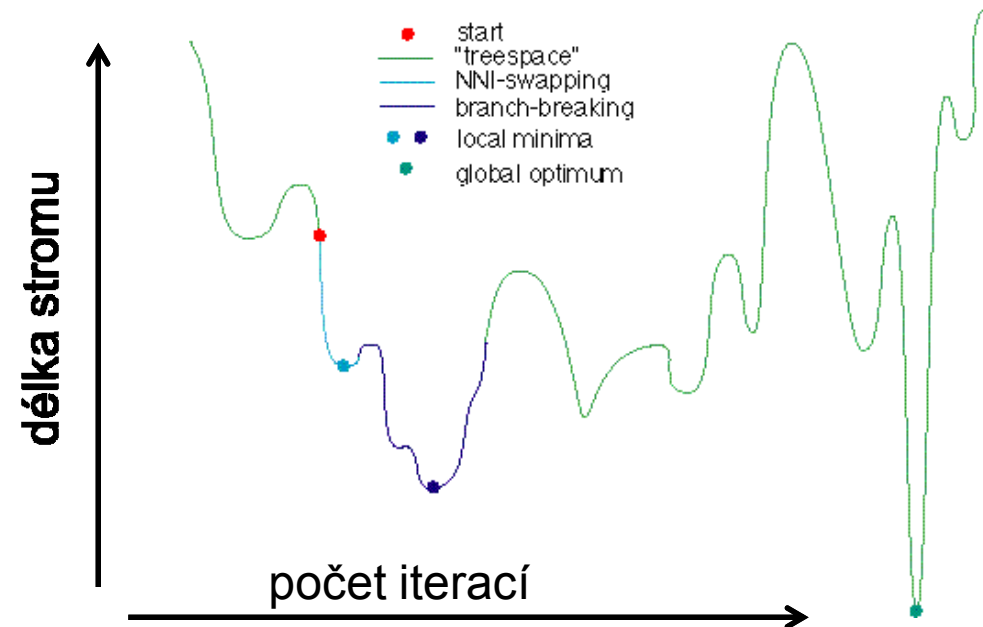


heuristické metody (*heuristics*)

- nepřesné (možnost lokálního optima), ale jediné možné při velkém počtu stromů

výměna větví (*branch swapping*): ca. do 100 taxonů):

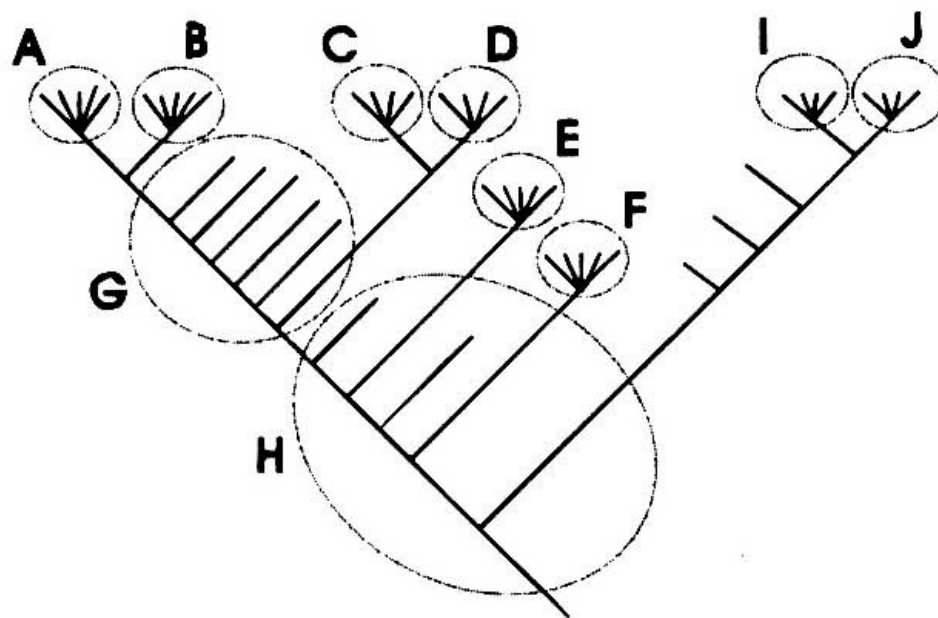
- *nearest neighbour interchange*
- *subtree pruning and regrafting*
- *tree bisection+reconnection (TBR)*



heuristické metody (*heuristics*)

analýza rozsáhlých maticí (nad 100 taxonů) pomocí parsimonie:

- problém tzv. ostrovů (*islands*) a *composite optima* (Goloboff 1999, Nixon 1999)
- *parsimony ratchet*
- *tree-fusing* (TF)
- *sectorial searches* (SS)
- *tree-drifting* (DFT)
- kombinace různých algoritmů

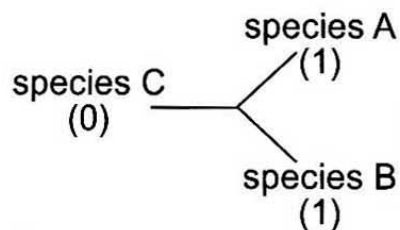


Polarizace znaků pomocí outgroup comparison

- po skončení hledání nejúspornějšího stromu vybereme jeden taxon z matice jako outgroup, podle něj jsou potom polarizovány všechny znaky na stromu

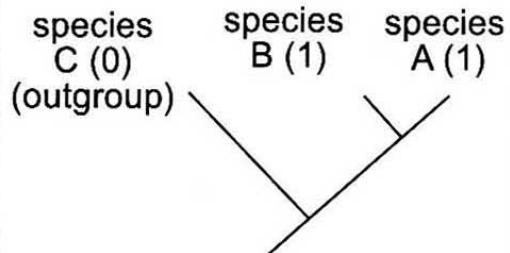
Cladistic determination of character polarity:

1) character distribution in a topology:



(0) and (1) = characters

2) selection of an outgroup:



3) identification of character state:

outgroup character state = plesiomorphy

conclusion:
0 = plesiomorphy
1 = apomorphy

Hennigian determination of character polarity:

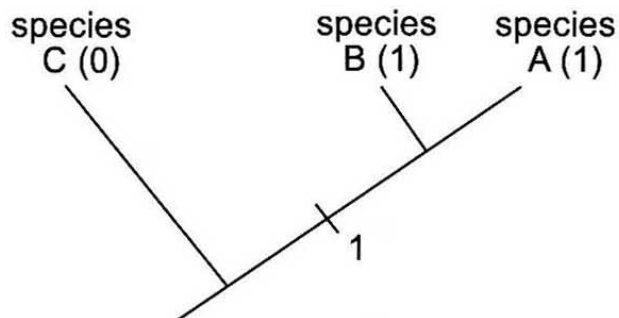
1) character distribution:

species	character
species A	1
species B	1
species C	0

2) character analysis:

1 = apomorphy
0 = plesiomorphy

3) reconstruction of a rooted tree:



Kritéria optimality pro vícestavové znaky

- nutné upřesnění podmínek parsimonie
- optimalizace (*optimization, character mapping*) = určení umístění znaků na kladogramu, tj. pořadí jejich změn, na základě nějakého kritéria
- Wagnerova (= Farrisova) optimalizace: aditivní, seřazené znaky (transformační série), přičítají se kroky do délky stromu (např. $1 \rightarrow 3 = 2$ kroky)

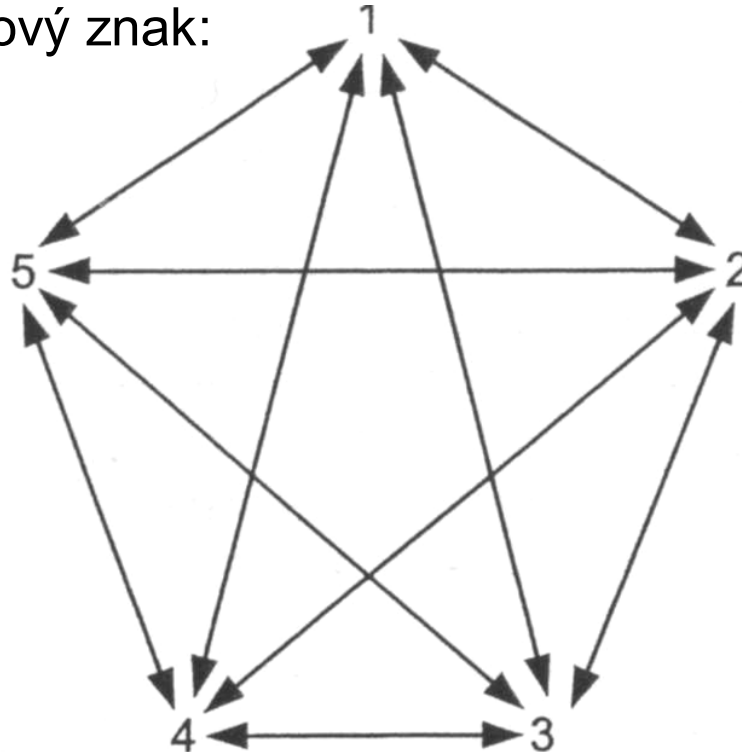
seřazený vícestavový znak:



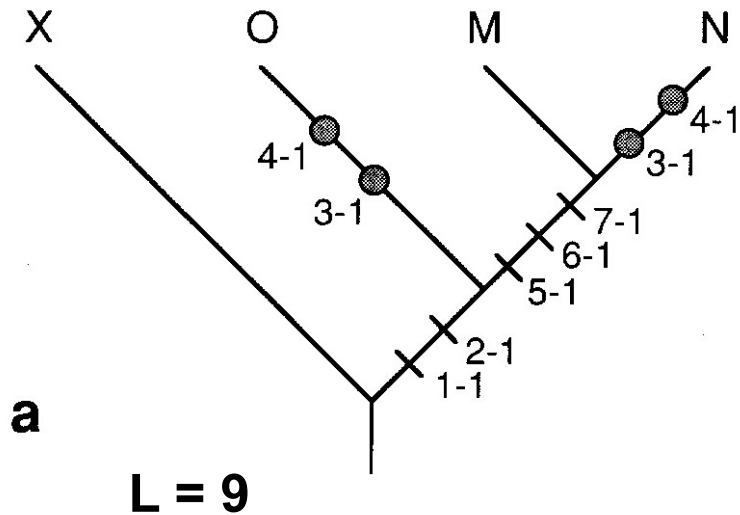
Kritéria optimality pro vícestavové znaky

- Fitchova optimalizace: neaditivní, neseřazené znaky (transformační série), přičítá se vždy jen jeden krok k celkové délce stromu (např. $1 \rightarrow 3 = 1$ krok)

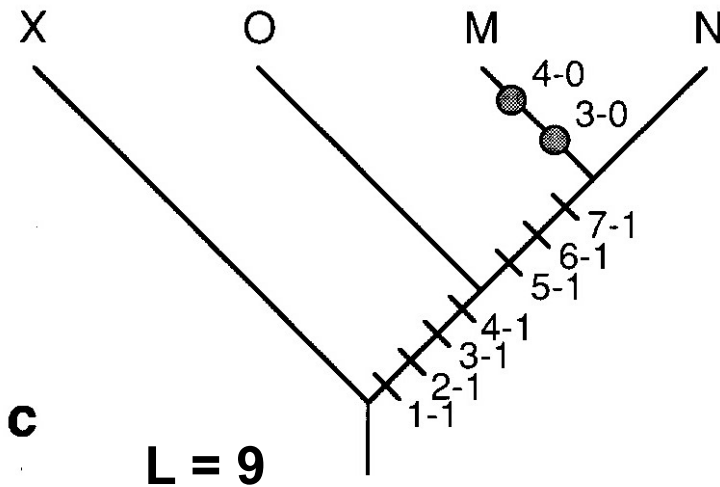
neseřazený vícestavový znak:



Optimalizační kritéria pro homoplázie



DELTRAN – *slow optimization*
„zpomaluje“ (*delays*) evoluční
změnu, posouvá ji co nejvíc po
stromě nahoru = upřednostňuje
paralelismy před reverzemi



ACCTRAN – *fast optimization*
„zrychluje“ (*accelerates*)
evoluční změnu, posouvá ji co
nejvíc po stromě dolů =
upřednostňuje reverze před
paralelismy

Popisné charakteristiky stromů

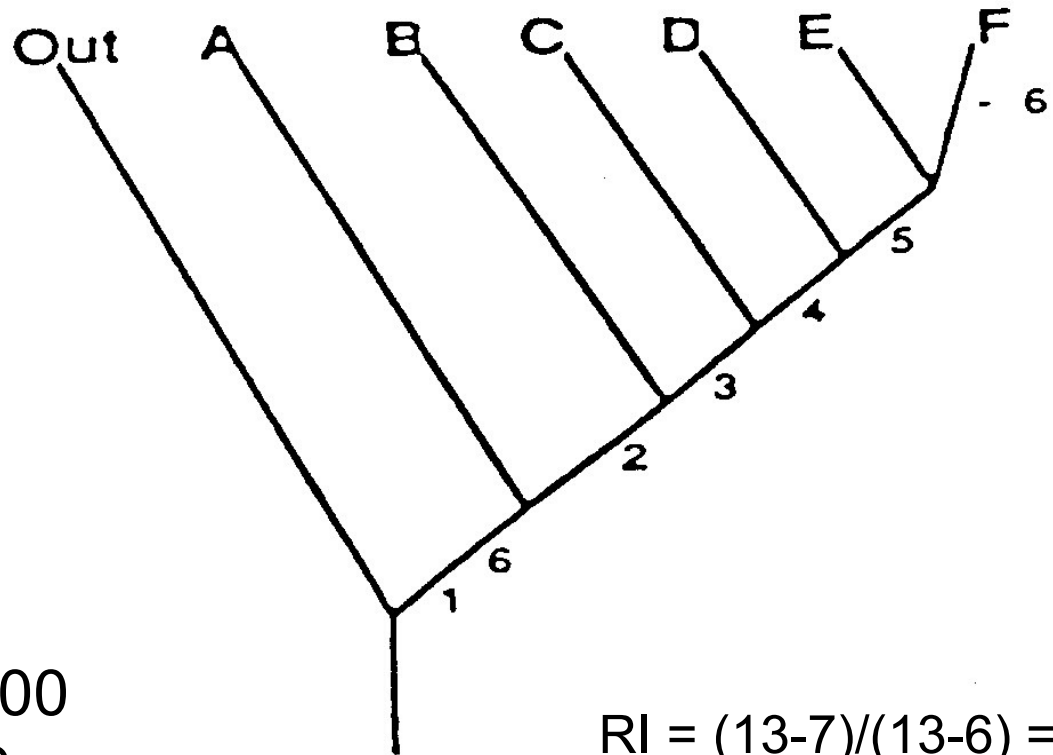
- délka stromu (*tree length*), L
- index konzistence (*consistency index*), CI
relativní míra homoplázie pro daný kladogram („*goodness of fit*“)

$$CI = \frac{\text{minimální možný (předpokládaný) počet kroků}}{\text{skutečný počet kroků (délka stromu)}}$$

	1	2	3	4	5	6
Outgroup	0	0	0	0	0	0
Taxa A	1	0	0	0	0	1
Taxa B	1	1	0	0	0	1
Taxa C	1	1	1	0	0	1
Taxa D	1	1	1	1	0	1
Taxa E	1	1	1	1	1	1
Taxa F	1	1	1	1	1	0

Character State Changes on Tree:

Character	State Changes
1	0 --> 1
2	0 --> 1
3	0 --> 1
4	0 --> 1
5	0 --> 1
6	0 --> 1 --> 0



$$L = 7$$

$$CI = 6/7 = 0.86$$

$$ci(1-5) = 1/1 = 1.00$$

$$ci(6) = 1/2 = 0.50$$

$$RI = (13-7)/(13-6) = 0.86$$

- index retence (*retention index*), RI
relativní míra synapomorfie pro daný kladogram

$$RI = \frac{\text{maximální možný počet kroků} - \text{skutečný počet kroků na stromu}}{\text{maximální možný počet kroků} - \text{minimální možný počet kroků}}$$

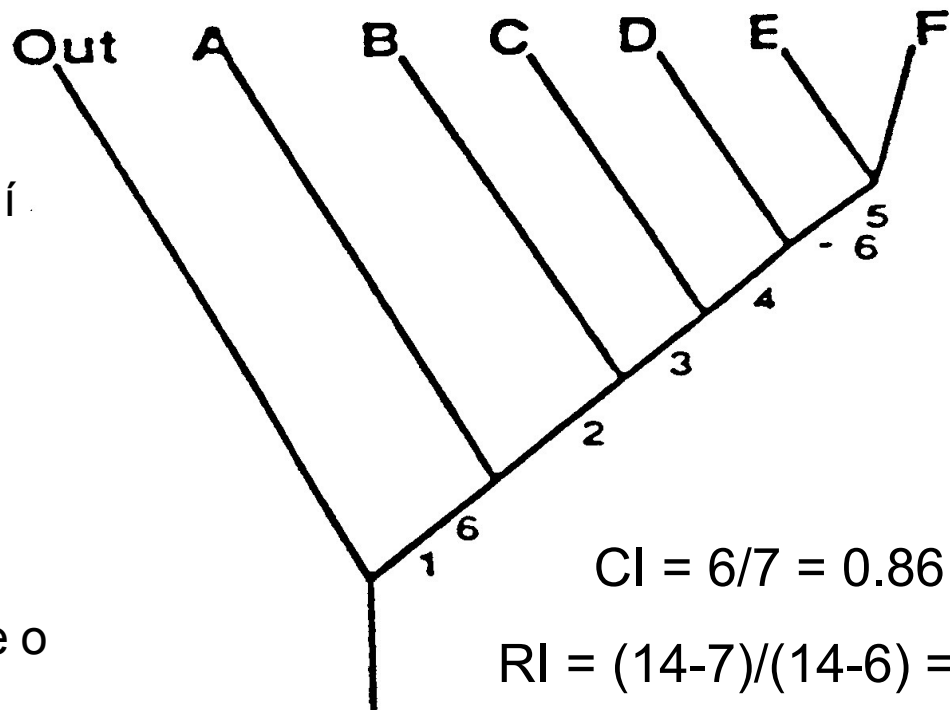
$$ri = 1.00$$

znak je zcela v souladu s kladogramem, plně informativní

čím více se ri blíží k 0, tím méně informace o příbuznosti taxonů

$$ri = 0.00$$

autapomorfie, žádná informace o příbuznosti

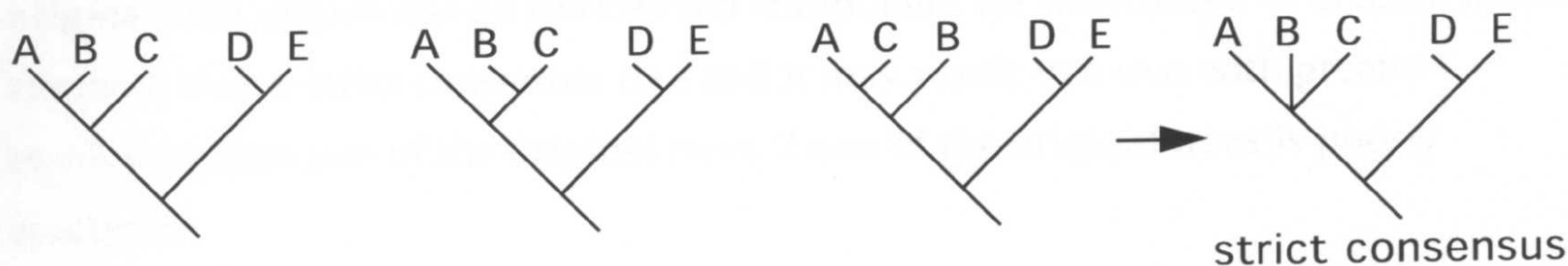


$$CI = 6/7 = 0.86$$

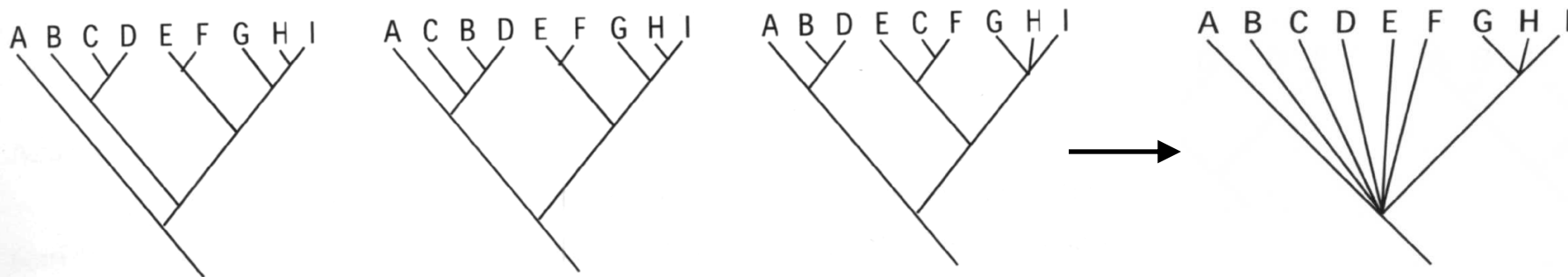
$$RI = (14-7)/(14-6) = 0.88$$

Konsenzuální stromy

- při výskytu homoplázie má kladistická analýza často více stejně úsporných řešení
- přísný konsensus (*strict consensus*)

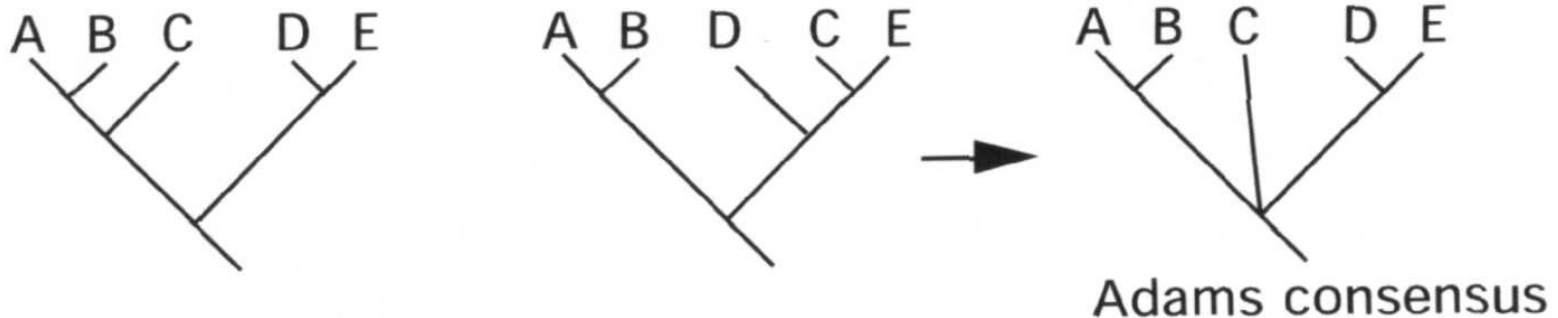


- výhoda: kombinuje jen shodnou informaci ze všech kladogramů
- nevýhoda: často malé rozlišení



Konsenzuální stromy

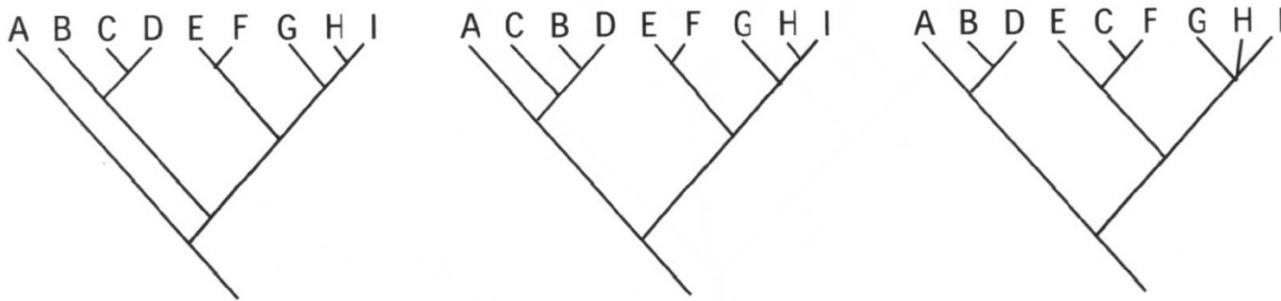
- Adamsův konsensus



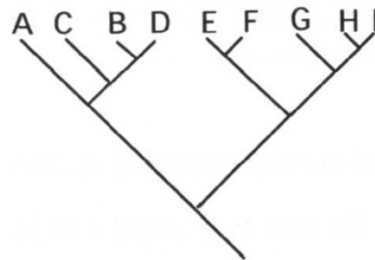
- výhoda: odstraňuje rozpory vyvolané jediným taxonem
- nevýhoda: může obsahovat skupiny, které nemají opodstatnění v žádném původním kladogramu

Konsensuální stromy

- většinový konsensus (*majority consensus*)



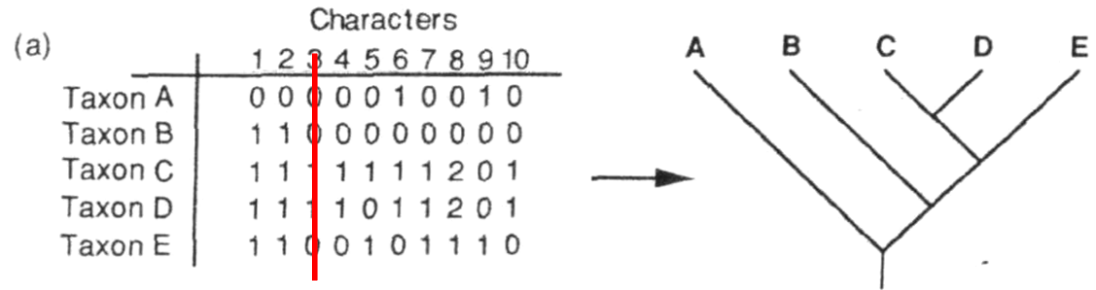
The majority rules consensus tree is:



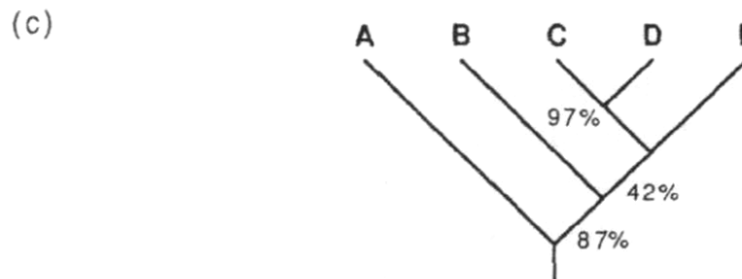
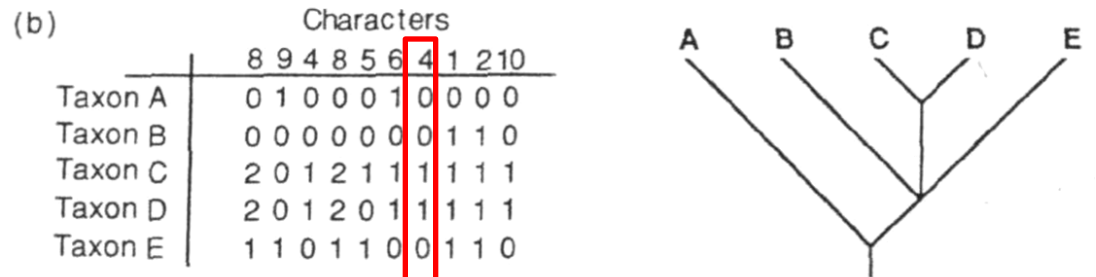
- zobrazí jen skupiny vyskytující se ve více než X % kladogramů
- např. zobrazení výsledků bootstrapu a jackknife
- nevýhoda: můžeme ztratit důležitou menšinovou informaci

Míry spolehlivosti výsledků klad. analýzy

- bootstrap – opakovaný náhodný výběr a nahrazení jednoho znaku zdvojením jiného
- jackknife – opakovaný náhodný výběr a odstranění jednoho či více znaků/taxonů bez náhrady

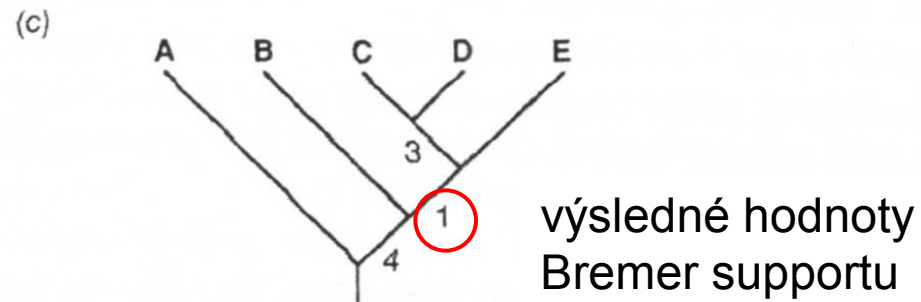
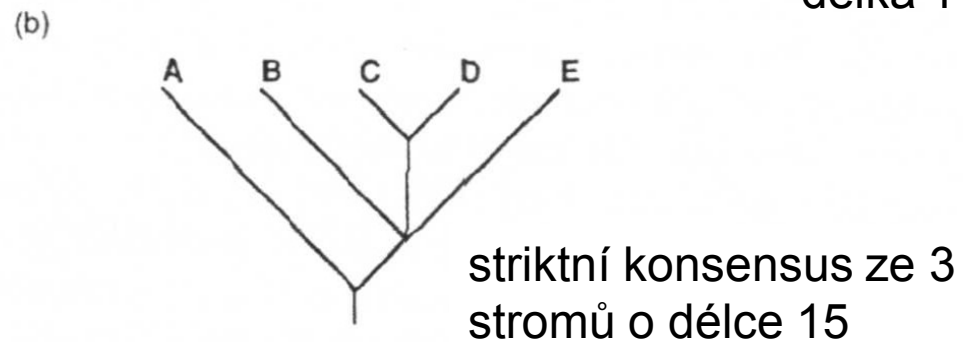
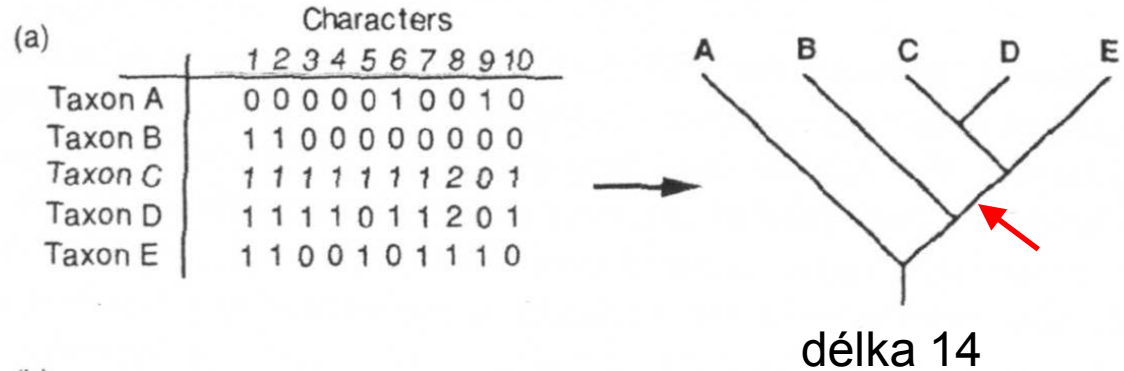


bootstrap

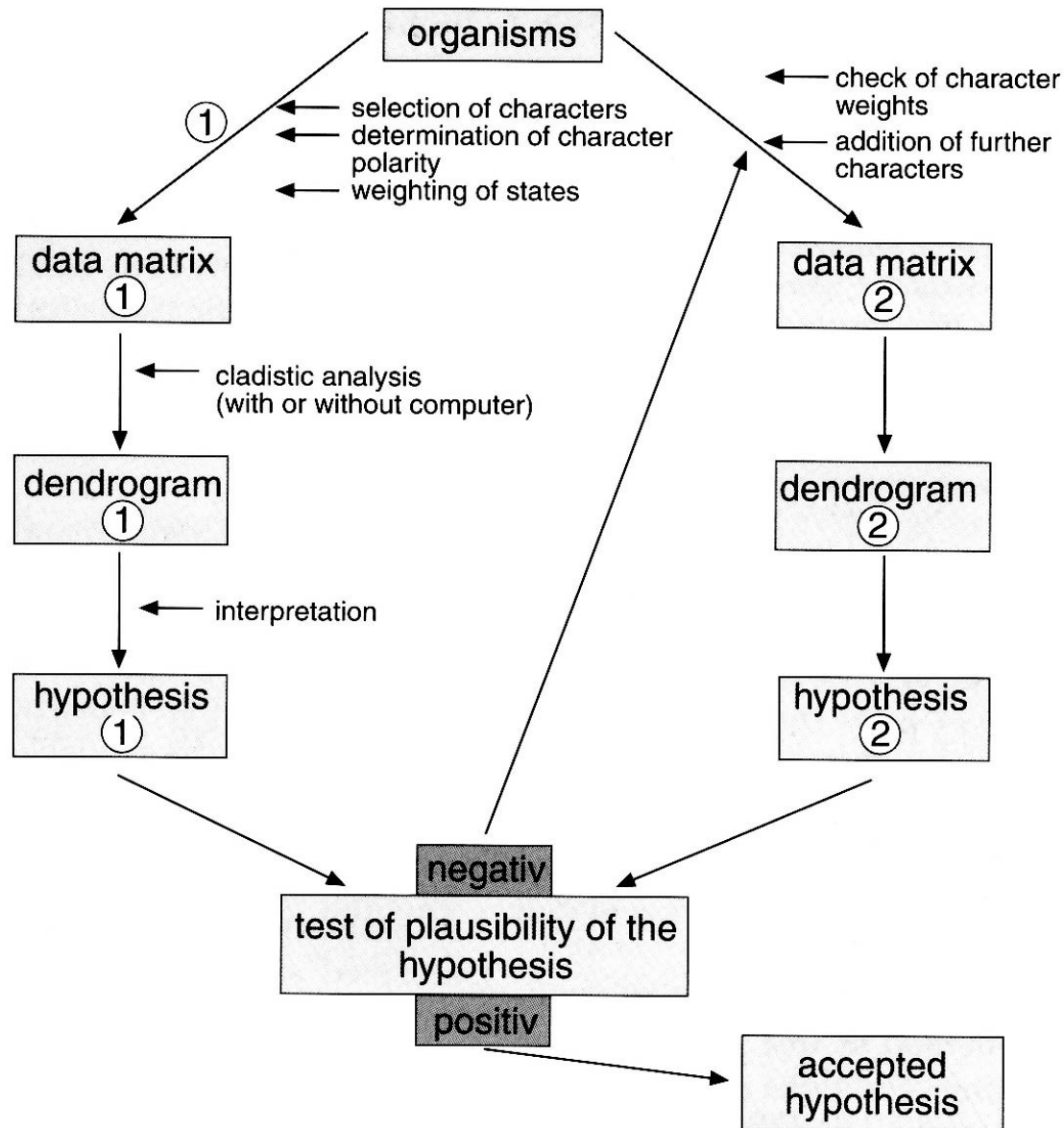


Míry spolehlivosti výsledků klad. analýzy

- Bremer support - počet kroků nutných k tomu, aby se daný klád ztratil ze striktního konsensu suboptimálních kladogramů



Postup v kladistické analýze: *checking & rechecking*



Kladistický software

- kompletní přehled fylogenetických programů:
<http://evolution.genetics.washington.edu/phylip/software.html>
- **NONA 2.0** (Goloboff 1993) + **Winclada 1.00.08** (Nixon 2002)
- **TNT** (Goloboff, Nixon & Farris 2008)
- **PAUP* 4.0 beta** (Swofford 2002) + **MacClade 4.07** (Maddison & Madison 2002)
- **Hennig86** (Farris 1989) + **TreeGardener** (Ramos 1996)

Literatura:

- Kitching, I. J., Forey, P. L., Humphries, C. J. & Williams, D. M. (1998): *Cladistics. The theory and practice of parsimony analysis*. 2. vyd. Oxford University Press, Oxford - New York - Tokyo, 228 s. ISBN 0-19-850138-2.
- Lipscomb D. (1998): *Basics of cladistic analysis*. George Washington University, Washington D.C., 75 s. (volně dostupné na www.gwu.edu/~clade/faculty/lipscomb/Cladistics.pdf)
- Schuh, R. T. (2000): *Biological systematics. Principles and applications*. Cornell University Press, Ithaca - London, 236 s.
- Skelton P., Smith A. & Monks N. (2002): *Cladistics. A practical primer on CD-ROM*. Cambridge University Press, Cambridge, 80 s.
- Wiley E., Siegel-Causey D., Brooks D.R. & Funk V.A. (1991): *The Compleat Cladist*. The University of Kansas Museum of Natural History Special publication No. 19. Lawrence, Kansas, 158 s.
- Williams D. M. & Ebach M. C. (2008): *Foundations of Systematics and Biogeography*. Springer, New York, 309 pp.
- Wägele, J.-W. (2005): *Foundations of phylogenetic systematics*. Verlag Dr. Friedrich Pfeil, München, 365 s.

Cladistics



... it's the real thing.