

6. Kombinace dat různého charakteru a interpretace nezávislých datových souborů na základě „známé“ fylogeneze (zpravidla molekulární)

- možnosti a úskalí kombinace více standardních molekulárních markerů
- případové studie
- interpretace inkongruentních výsledků (plastidová DNA+ITS, plastidová DNA+AFLP/mikrosatelity)
- analýzy významných vlastností (např. morfologie, cytologie) ve fylogenetickém kontextu [R-package ape, BayesTraits, Mesquite]
- praktická část – analýza významných znaků ve fylogenetickém kontextu

analýzy významných vlastností ve fylogenetickém kontextu

- „známe-li“ molekulární evoluční historii skupiny, tj. fylogenezi, můžeme studovat také
 - evoluci různých významných vlastností/znaků dané skupiny, např.
 - **morfologických** (velikost, tvar, přítomnost/nepřítomnost orgánu)
 - **cytologických** (velikost genomu, počet chromosomů,...)
 - apod...
- Důležitou podmínkou je mít (zcela) kompatibilní dataset pro molekulární fylogenezi a pozorovaný soubor znaků!

	A	B	C	D	E	F	G	H	I	
	sekvenční data	matK sequence	FCM no	2C-DNA amount	1C-DNA amount (pg)	chrom. No	ploidy level	1Cx (2C/ploidy level)	presece of epi glands	Binární data
1	species									
2	CURCUMA									
3	Curcuma caesia	caataaattcg	FCM 879	2,56397	1,281985	63	9	0,040698	1	
4	Curcuma sp. "elata-latifolia"	caataaattcg	FCM 889	2,589383	1,294691	63	9	0,041101	1	
5	Curcuma longa	caataaattcg	FCM 904	2,37882	1,18941	63	9	0,037759	1	
6	Curcuma amada	caataaattcg	FCM 807	1,704733	0,852367	63	9	0,027059	0	
7	Curcuma longa	caataaattcg	FCM 905	2,341608	1,170804	63	9	0,037168	1	
8	Curcuma sp. "elata-latifolia"	GGTAG-AGTC	FCM 890	2,669252	1,334626	42	6	0,063554	0	
9	Curcuma sp. "fucata"	caataaattcg	FCM 922	2,545818	1,272909	42	6	0,060615	0	

S jakými typy dat pracujeme

(1) kvalitativní (qualitative):

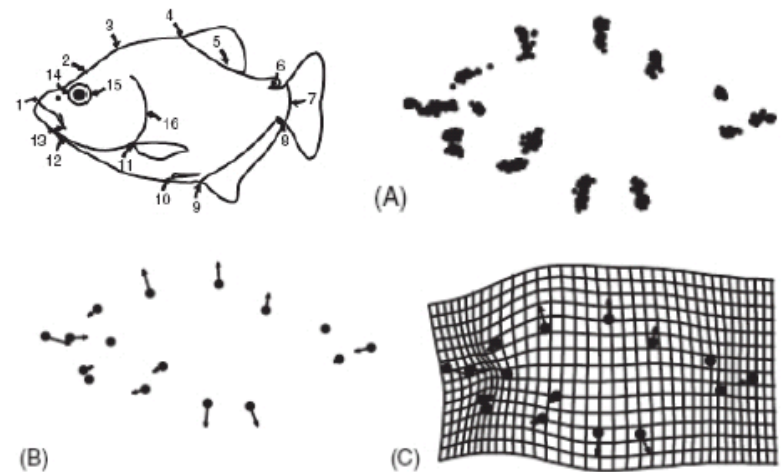
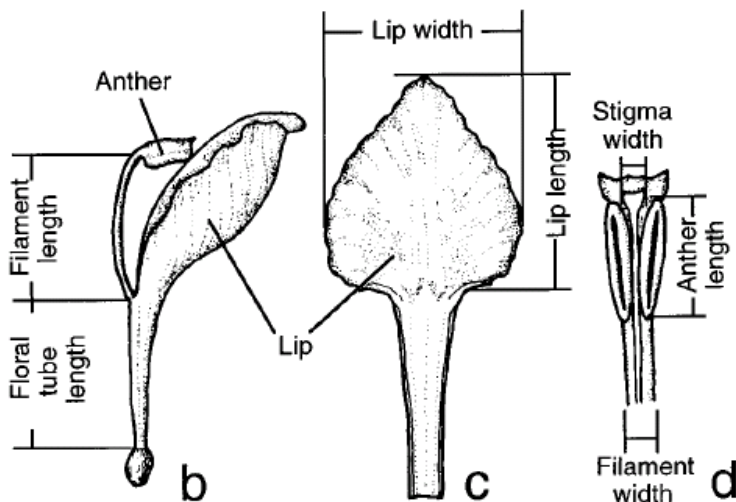
- binární (přítomnost, nepřítomnost, 0,1)
- víceřadová (kategorická - omezený počet kategorií) – např. počet chromosomů, počet tyčinek, apod.

(2) kvantitativní (quantitative)

- nespojitá, diskrétní (meristic) – např. počet listů, buněk, apod.
- spojitá (continuous) – např. délka listu, velikost genomu, apod.

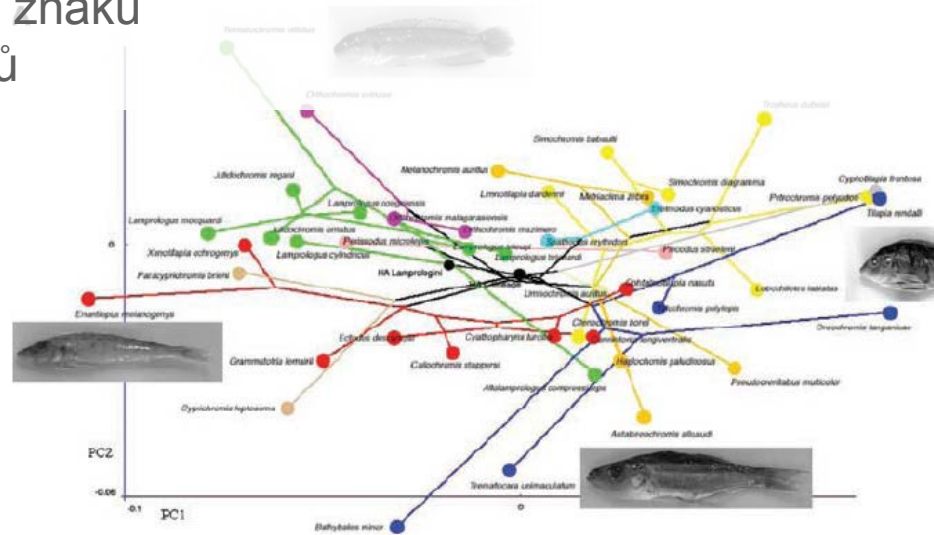
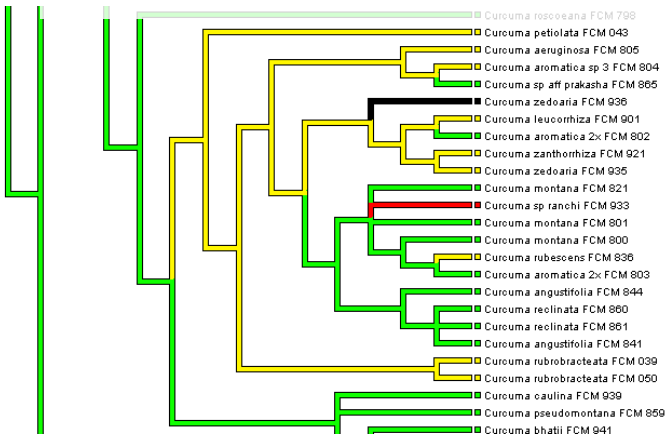
(3) „vícerozměrná data“

- dvě a více spojitých proměnných charakterizuje jeden objekt
- např. geometricko-morfometrická landmarková data charakterizující tvar objektu

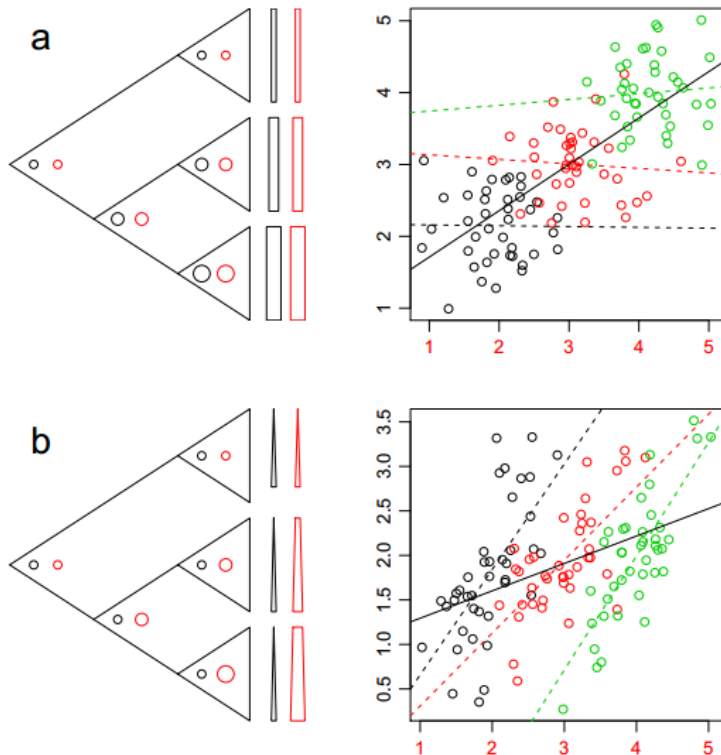


Co se můžeme dozvědět o charakteru a evoluci znaků

- Vyvíjí se morfologické, cytologické a jiné znaky v závislosti na molekulární evoluci, nebo je jejich vývoj závislý spíše na jiných faktorech (ekologie, abiotické faktory apod.)?
 - tj. testujeme přítomnost fylogenetického signálu v datech – **λ (lambda)**
- Jakým způsobem a jak rychle se vyvíjí znaky v porovnání s molekulární fylogenezí?
 - mění se znaky spíše skokově nebo naopak pozvolna? – **κ (kappa)**
 - Odpovídá evoluce znaků spíše adaptivní radiaci nebo recentní diversifikaci? – **δ (delta)**
- Jaké znaky a vlastnosti mohli mít předkové námi studované skupiny?
 - Rekonstrukce ancestrálních stavů znaků
 - Rekonstrukce ancestrálních tvarů



Je závislost fylogeneticky korektní? ...fylogenetická autokorelace...



- silná závislost mezi dvěma znaky daná relativně málo nezávislými srovnáními
- mizí po fylogenetické korekci
- silná závislost mezi znaky maskována fylogenetickými rozdíly mezi skupinami

- phylogenetic independent contrast (PIC) – Felsenstein (1985)
- phylogenetic generalized least square (PGLS)

PGLS (phylogenetic generalized least square)

- lineární regrese: $y = \beta X + \varepsilon$
 - y – vysvětlovaná proměnná (*response variable*)
 - β – odhadované koeficienty
 - X – vysvětlující proměnná (*explanatory variable*)
 - ε – rezidua
- v případě příbuzných taxonů nejsou y a X nezávislé
- řešení – zahrnutí kovarianční matice (V), která definuje evoluční vzdálenost mezi taxony

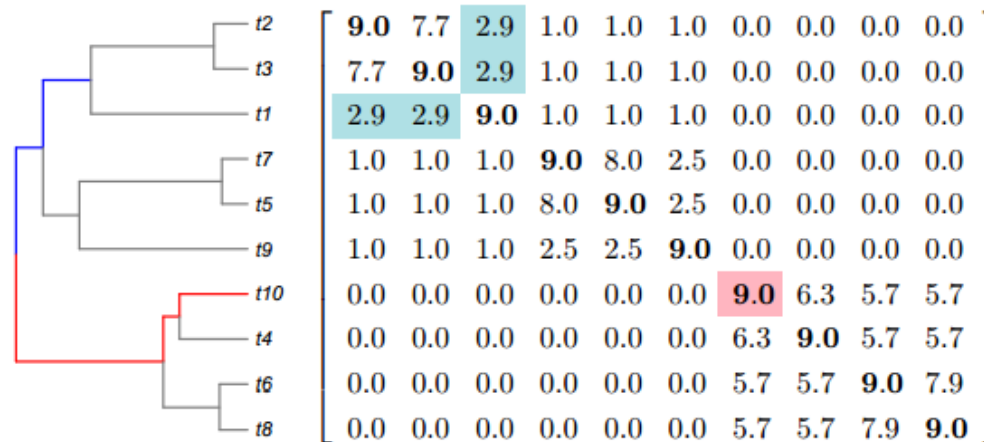


Figure 2: A phylogenetic tree of 10 taxa and the variance covariance matrix (V) of that phylogeny. The diagonal of the matrix (bold values) shows the path length from each tip to the root (example in red). Off diagonal values show the shared path length for a given pair of tips (example in blue).

- předpokládá se *Brownian motion* vývoje znaků, tj. znaky se vyvíjejí proporcionálně vzhledem k délce větví
- ne vždy vyhovuje – tři různé typy transformací pro lepší fitování modelu

PGLS – scaling parameters

- λ – *phylogenetic association*
 - interní délky větví násobené konstantou
 - $\lambda = 0$: interní větve nulové, tj. variabilita je nezávislá na fylogenezi
 - $\lambda = 1$: žádné škálování, variabilita kopíruje fylogenezi
 - $\lambda > 1$: větší kovariance než předpokládá *Brownian model*
- δ – *tempo of evolution*
 - umocnění všech hodnot na ' δ ' – transformace sumy sdílených délek mezi jedinci, tj. škáluje '*root to tip*' vzdálenost
 - detekce zrychlení nebo zpomalení evoluce znaku
 - $\delta < 1$: časné změny důležité (adaptivní radiace), pomalé změny mezi blízce příbuznými
 - $\delta = 1$: graduální evoluce znaku, tj. proporcionálně délce větví
 - $\delta > 1$: pozdější změny důležité, tj. zrychlování změn v průběhu evoluce (druhově specifická adaptace)
- κ – *mode of evolution*
 - umocnění všech délek větví na ' κ ', tj. škáluje délky všech větví
 - $\kappa < 1$: délky větví se stávají stejné (zkracuje delší větve více než kratší), tj. variabilita akumulována pouze pokud se skupiny liší
 - $\kappa = 0$: evoluce znaku nezávislá na délce větví – punktualismus
 - $\kappa = 1$: gradualismus

PGLS – scaling parameters

Three scaling parameters and their interpretation when applied to trait evolution on a phylogeny

Parameter	Action	0	<1	1	>1
λ (lambda)	Assess contribution of phylogeny	star phylogeny (species independent)	phylogenetic history has minimal effect	default phylogeny	not defined
κ (kappa)	Scale branch lengths in tree	punctuational evolution	stasis in longer branches	default gradualism	longer branches more change
δ (delta)	Scale total path (root to tip) in tree	not defined	temporally early change important (adaptive radiation)	default gradualism	temporally later change (species-specific adaptation)

PGLS – scaling parameters

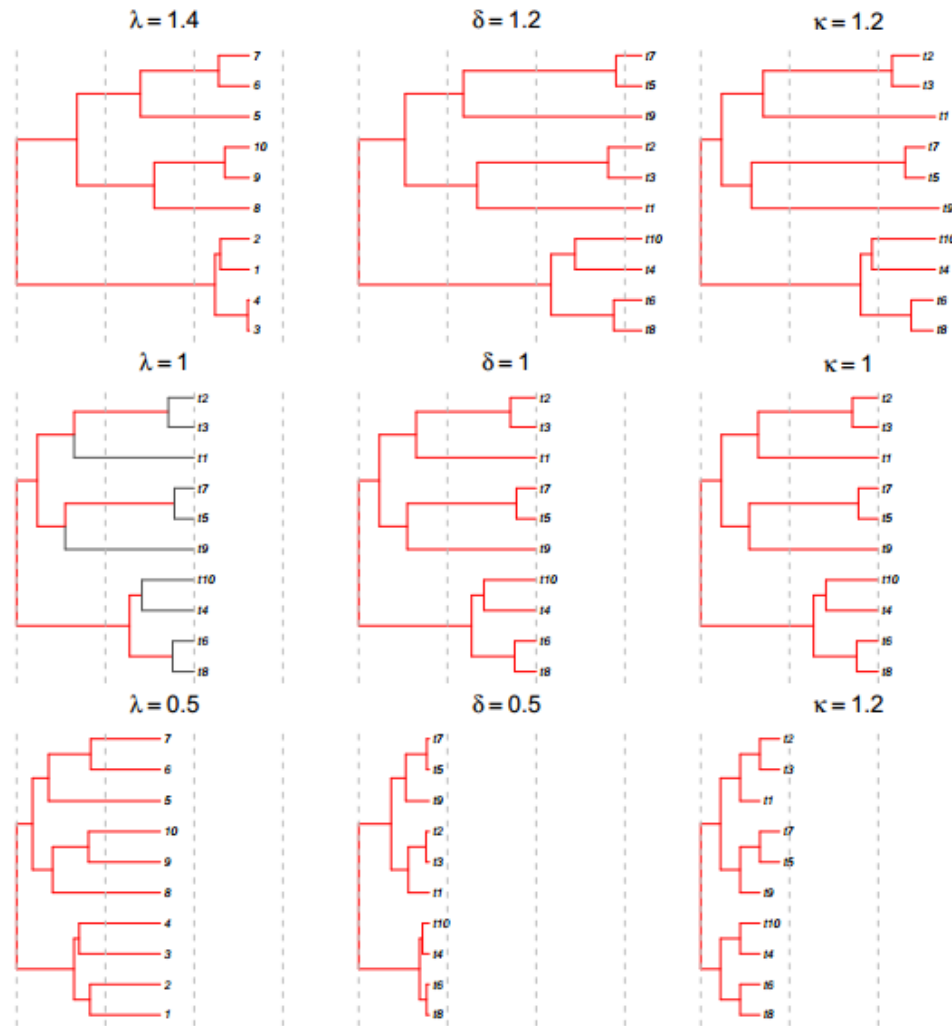
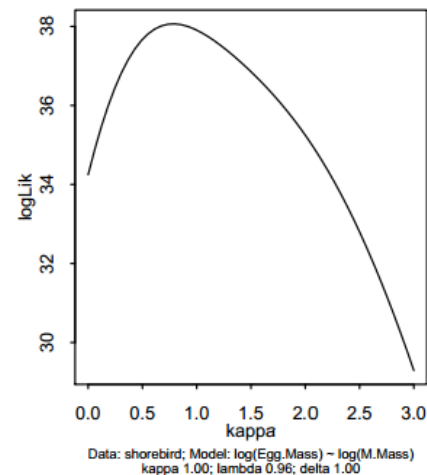
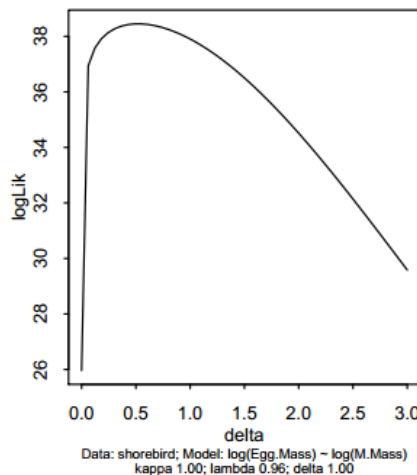
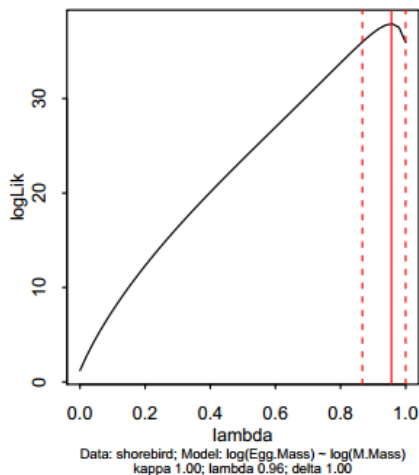


Figure 3: Examples of λ , δ and κ branch length transformations. The branches affected by a given transformation are shown in red.

PGLS

- fitování evoluce znaku nebo korelace dvou znaků s ohledem na fylogenezi
- odhad parametrů λ , δ a κ pro lepší fit modelu
 - ML – caper (funkce pgls) – hodnota + konfidenční interval
 - Bayesian – BayesTraits
- parametry můžou být odhadovány najednou – interpretace komplexního modelu?

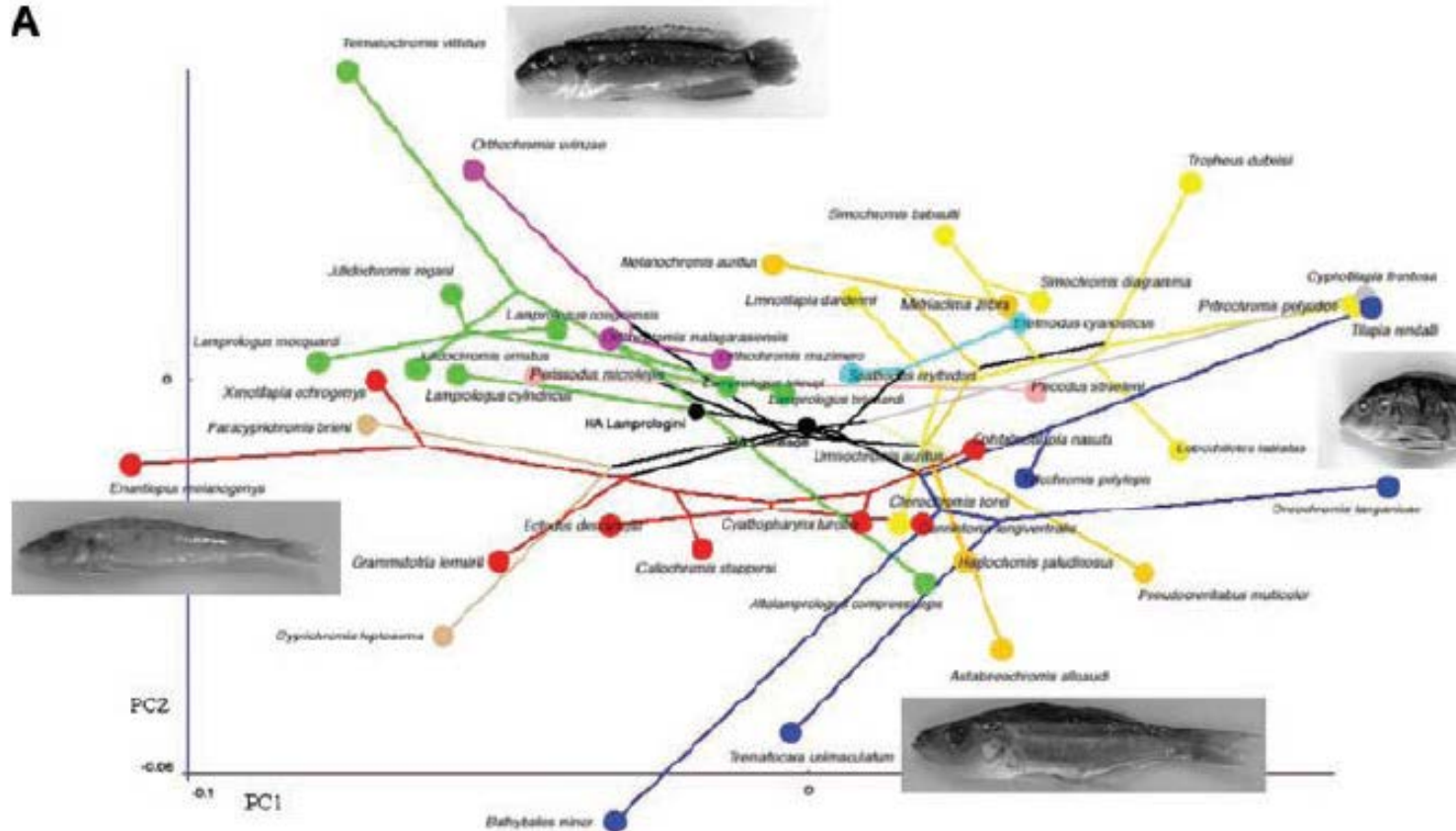


Jaké metody lze využít

- Maximální parsimonie
 - zejména odhady ancestrálních stavů znaků (diskrétní i spojitá data) a tvarů – tzv. square change parsimony
 - Programy: Mesquite, tpsTree
 - Výhoda – možnost vizualizace (Mesquite), ale R už dobíhá...
- Maximum likelihood
 - Odhady ancestrálních stavů znaků (diskrétní i spojitá data)
 - Program: R – knihovna Geiger
 - Odhady parametrů charakterizujících evoluci znaku – λ , κ , δ
 - Programy: R – knihovny Geiger, Caper; BayesTraits
 - Výhoda – rychlost
- Bayesovská analýza
 - Odhady ancestrálního stavu v kořeni stromu (diskrétní i spojitá data)
 - Program: BayesTraits
 - Odhady parametrů charakterizujících evoluci znaku – λ , κ , δ
 - Program: BayesTraits
 - Výhoda – může pracovat na základě více různých fylogenezí

Fylogenetický signál – lambda λ

A



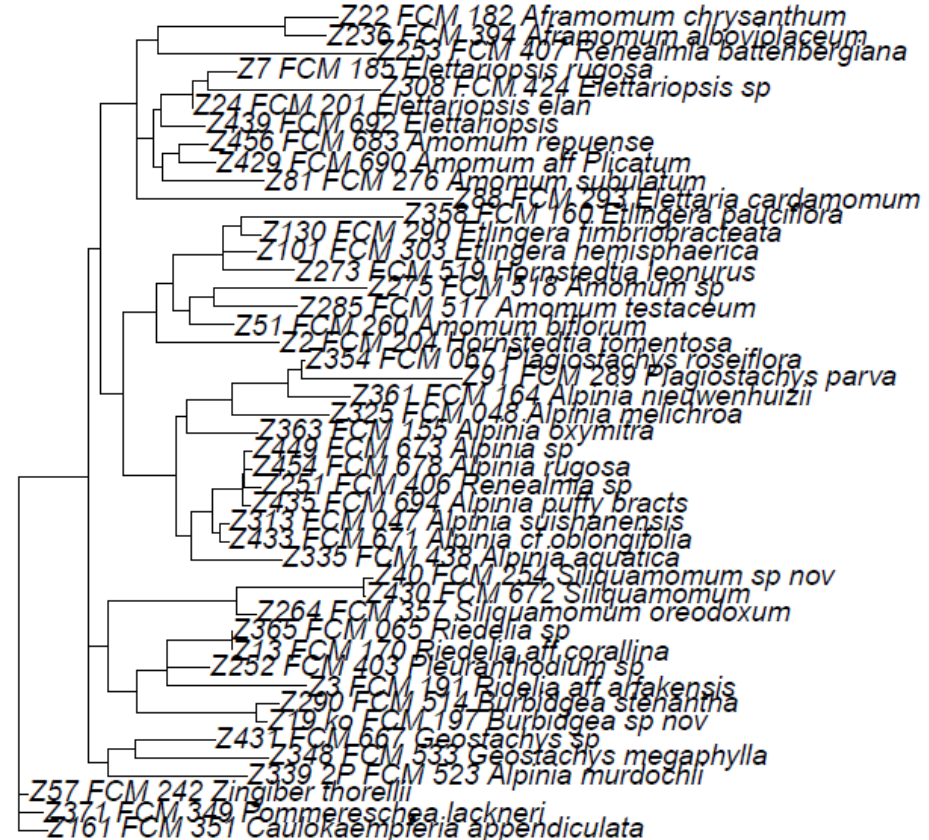
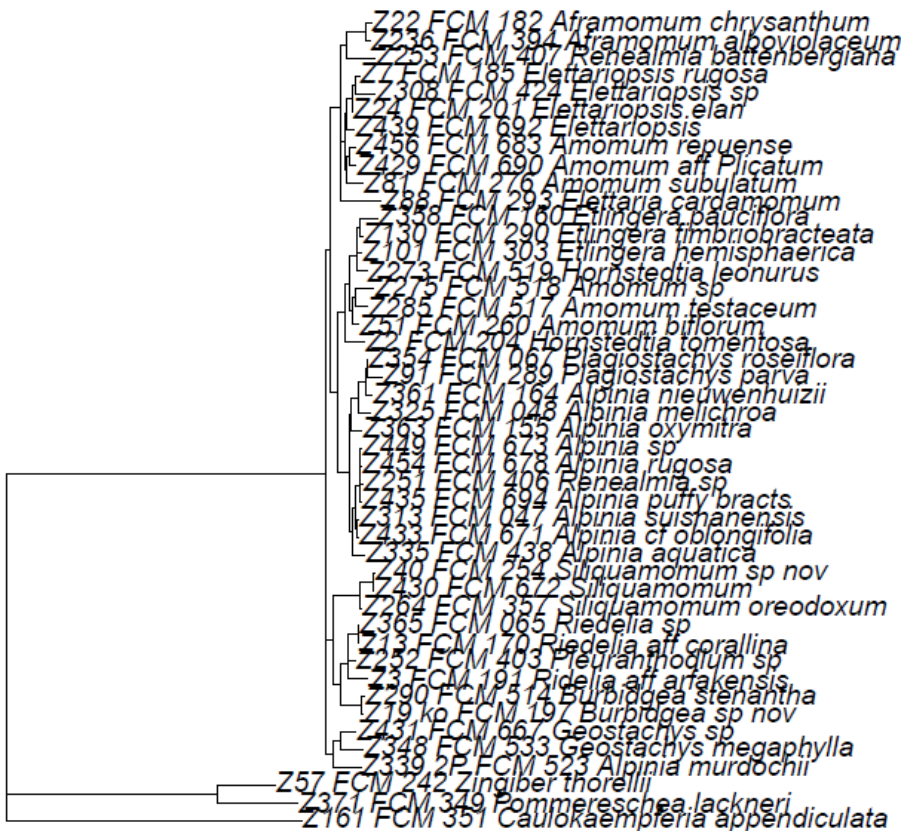
**GEOMETRIC MORPHOMETRIC ANALYSES
PROVIDE EVIDENCE FOR THE ADAPTIVE
CHARACTER OF THE TANGANYIKAN
CICHLID FISH RADIATIONS**

Clabaut et al. (2006) - studie vztahů mezi ekologickou a morfologickou diversitou a fylogenezi

Tempo evolve znaku – delta δ

• D = 0.2

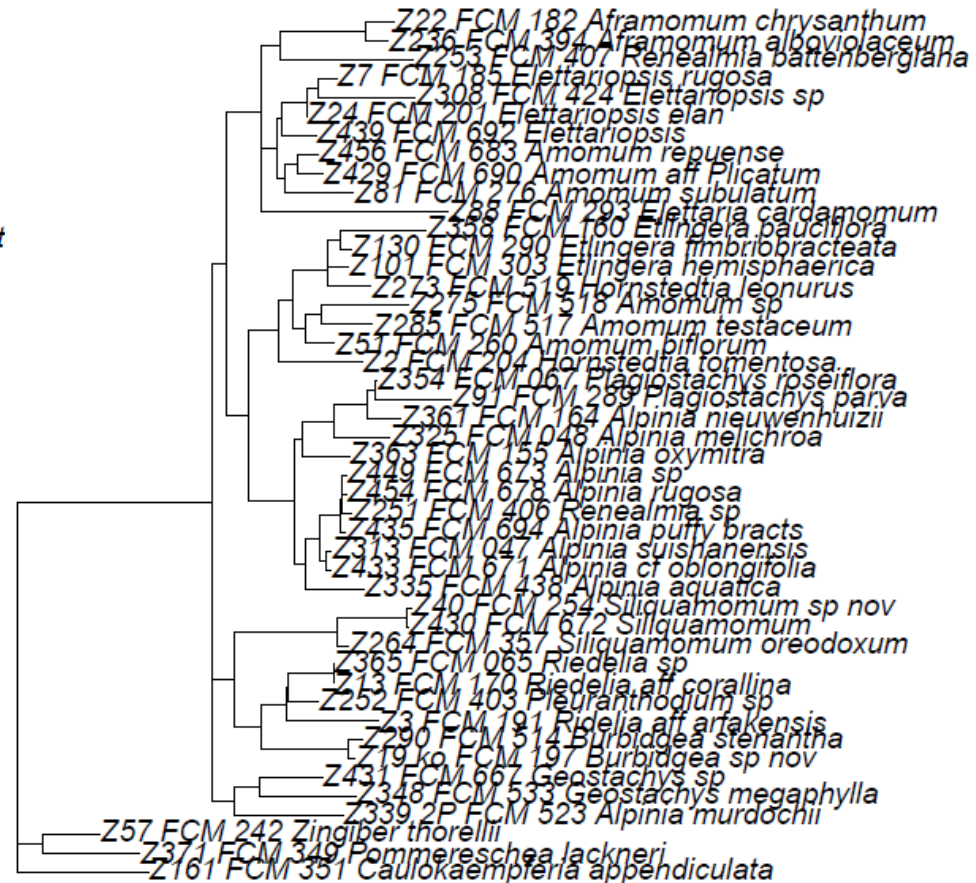
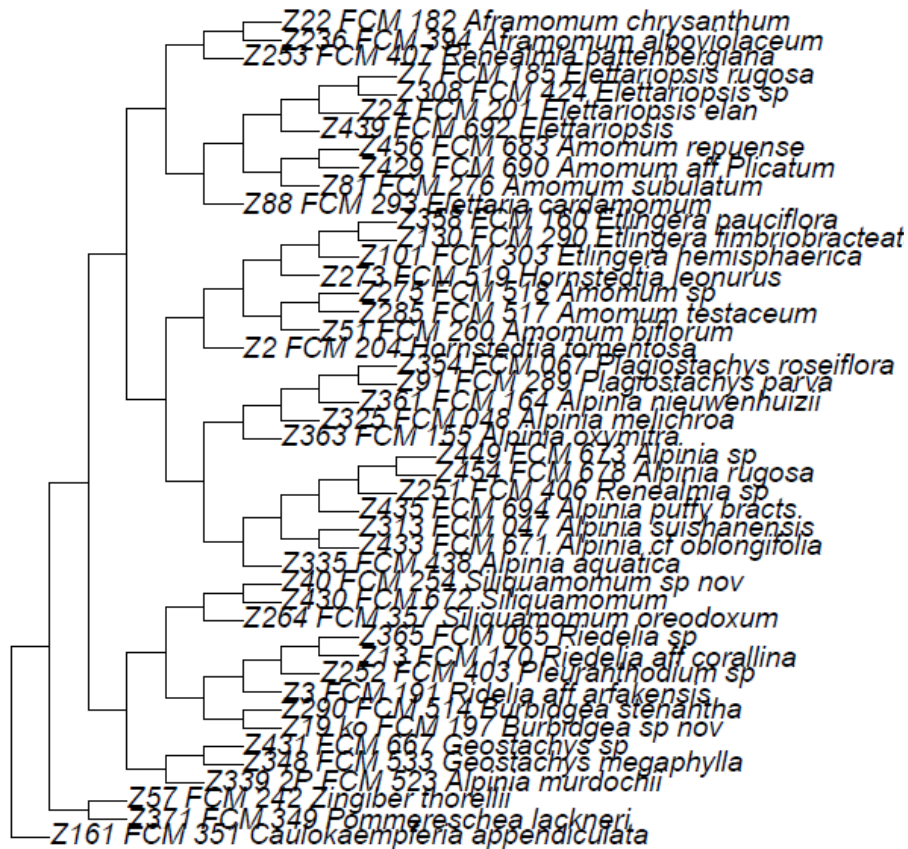
• D = 2.3



Charakter evoluce znaku – kappa κ

- $K=0$
- evoluce znaku nezávislá na délce větví \rightarrow skokové změny znaků

- $K=1$
- čím delší větev, tím výraznější změna znaku



Odhad parametrů lambda, kappa a delta pomocí BayesTraits

- Alternativní způsob odhadu parametrů oproti skriptům v R
- Výhodou, může pracovat i na základě více vstupních fylogenezí (zahrnutí nejistoty)
- Nevýhoda – nutno separátně počítat **Likelihood ratio test** (např. v R)
- Možnost odhadů na základě
 - ML
 - MCMC (bayes) – jiný program bayesovské odhady neumí

```
cd d:\ell\zazvory_2011\clanek-GS\listopad_2012\analzy\ALPINOIDEAE\bayesTraits\pok_N3-funkcni/m...
Rand Seed      1386862517
Please Select the model of evolution to use.
1) MultiState.
2) Discrete: Independent
3) Discrete: Depend
4) Continuous: Random Walk <Model A>
5) Continuous: Directional <Model B>
4
Please Select the analysis method to use.
1) Maximum Likelihood.
2) MCMC
2
Options:
Model: Continuous Random Walk
Tree File Name: ITS_matk_rooted.trees
Data File Name: Alpinoideae_log2C.txt
Log File Name: Alpinoideae_log2C.txt.log.txt
Summary: False
Analysis Type: MCMC
Sample Period: 100
Iterations: 5050000
Burn in: 50000
Rate Dev: 2.000000
No of Rates: 1
Test for trait correlation: True
Kappa: Not in use
Delta: Not in use
Lambda: Not in use
Restrictions:
  alpha-1: None
Prior Information:
  Prior Categories: 100
  alpha-1: uniform -100.00 100.00
Tree Information
  Trees: 1
  Taxa: 95
  Sites: 1
> lambda
> info
Options:
Model: Continuous Random Walk
Tree File Name: ITS_matk_rooted.trees
Data File Name: Alpinoideae_log2C.txt
Log File Name: Alpinoideae_log2C.txt.log.txt
Summary: False
Analysis Type: MCMC
Sample Period: 100
Iterations: 5050000
Burn in: 50000
Rate Dev: 2.000000
No of Rates: 1
Test for trait correlation: True
Kappa: Not in use
Delta: Not in use
Lambda: Estimate
Restrictions:
  alpha-1: None
Prior Information:
  Prior Categories: 100
  alpha-1: uniform -100.00 100.00
Tree Information
  Trees: 1
  Taxa: 95
  Sites: 1
>
```

rekonstrukce ancestrálních stavů znaků - Mesquite

- Kategorická, diskrétní i spojitá data – maximum parsimony
- maximum likelihood – jen kategorická data
- Potřebná data:
 - Soubor s fylogenetickým stromem
 - nexus formát - blok „TRANSLATE“, délky větví (nulové převede na jednotkové)
 - možno jeden i více stromů v souboru
 - Stromy na základě MP, ML i bayesovské
 - Tabulka pozorovaných dat (např. morfologických, cytologických, apod)
 - Nejlépe uspořádaná podle „TRANSLATE“ bloku ve stromech
 - Desetinná čísla s desetinnou tečkou

Data:

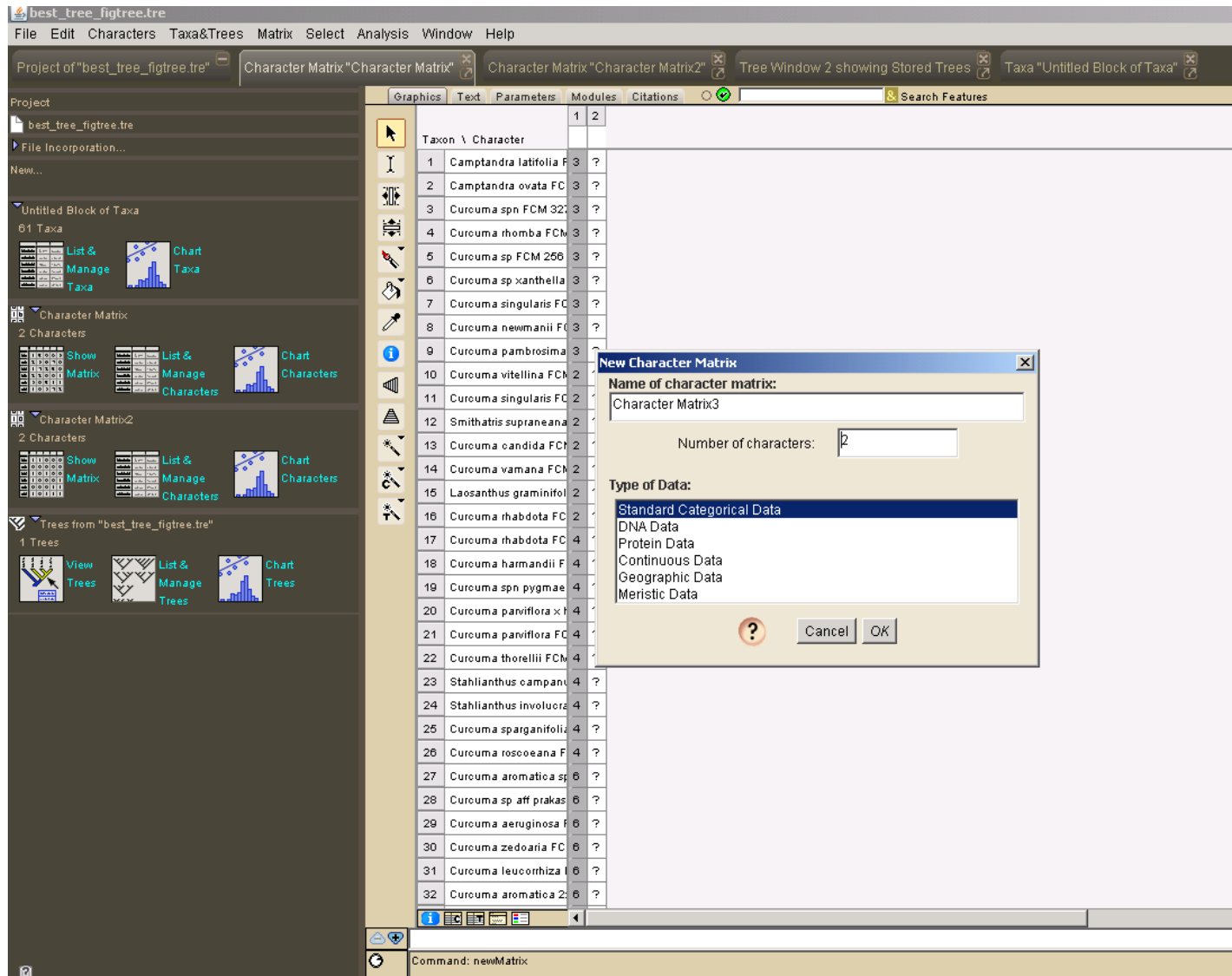
- ML Strom: best_tree_figtree.tre
- Data: Curcuma_1Cx.txt

Programy

- Mesquite

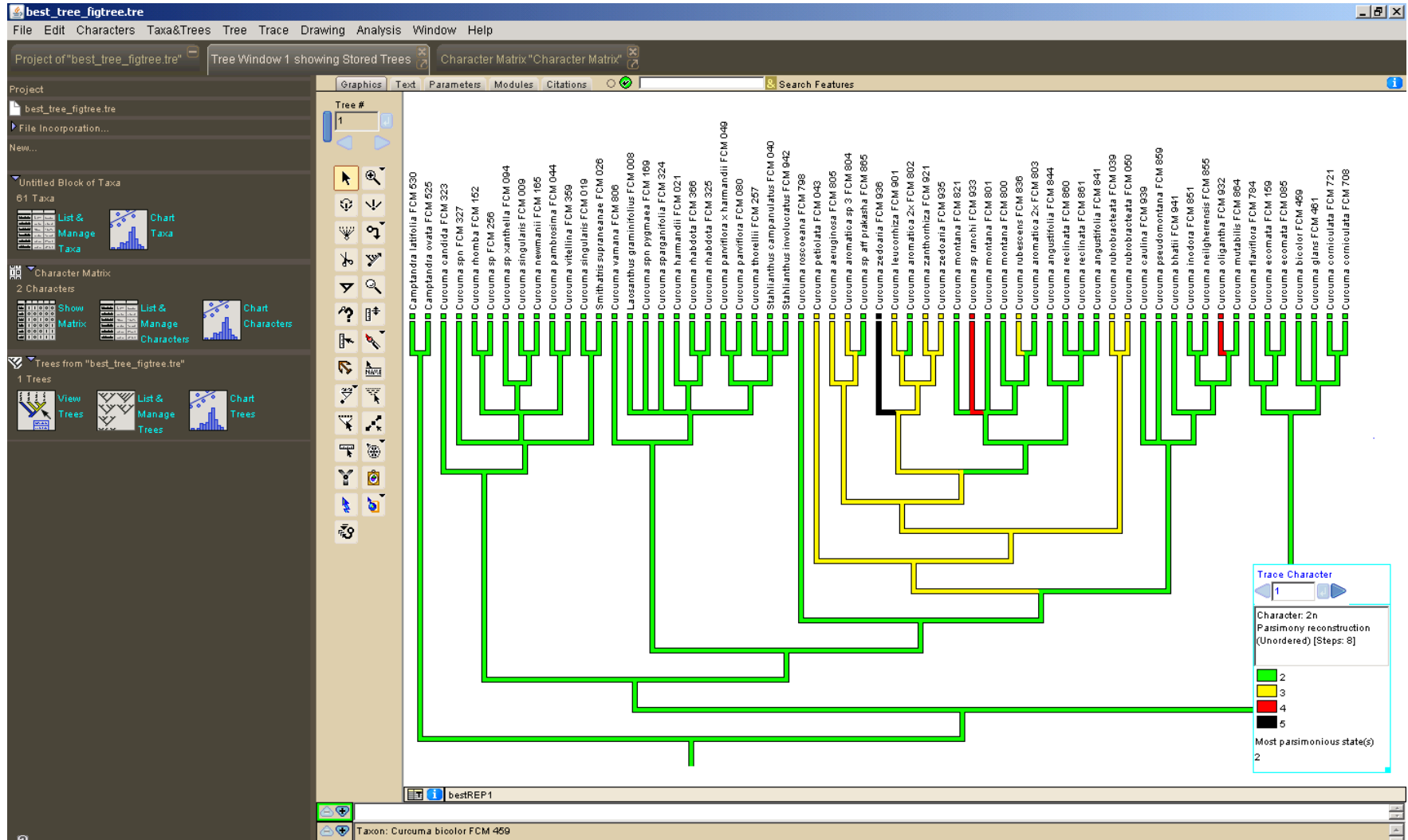
Import dat do programu Mesquite

- Otevření souboru s fylogenezí (nexus formát)
- Tvorba prázdné matice pro import hodnot znaků (kategorická, spojitá data)



Rekonstrukce ancestrálních stavů znaku

→ Analysis → Trace Character History



rekonstrukce ancestrálních stavů znaků – R - Geiger

- Kategorická i spojitá data – maximum likelihood
- Potřebná data:
 - Soubor s fylogenetickým stromem
 - nexus formát - blok „TRANSLATE“, nenulové délky větví
 - možno jen jeden strom (na základě MP, ML i bayesovské)
 - Tabulka pozorovaných dat (např. morfologických, cytologických, apod)
 - Nejlépe uspořádaná podle „TRANSLATE“ bloku ve stromech
 - Desetinná čísla s desetinnou tečkou,
 - Nejlépe *.csv soubor

Data:

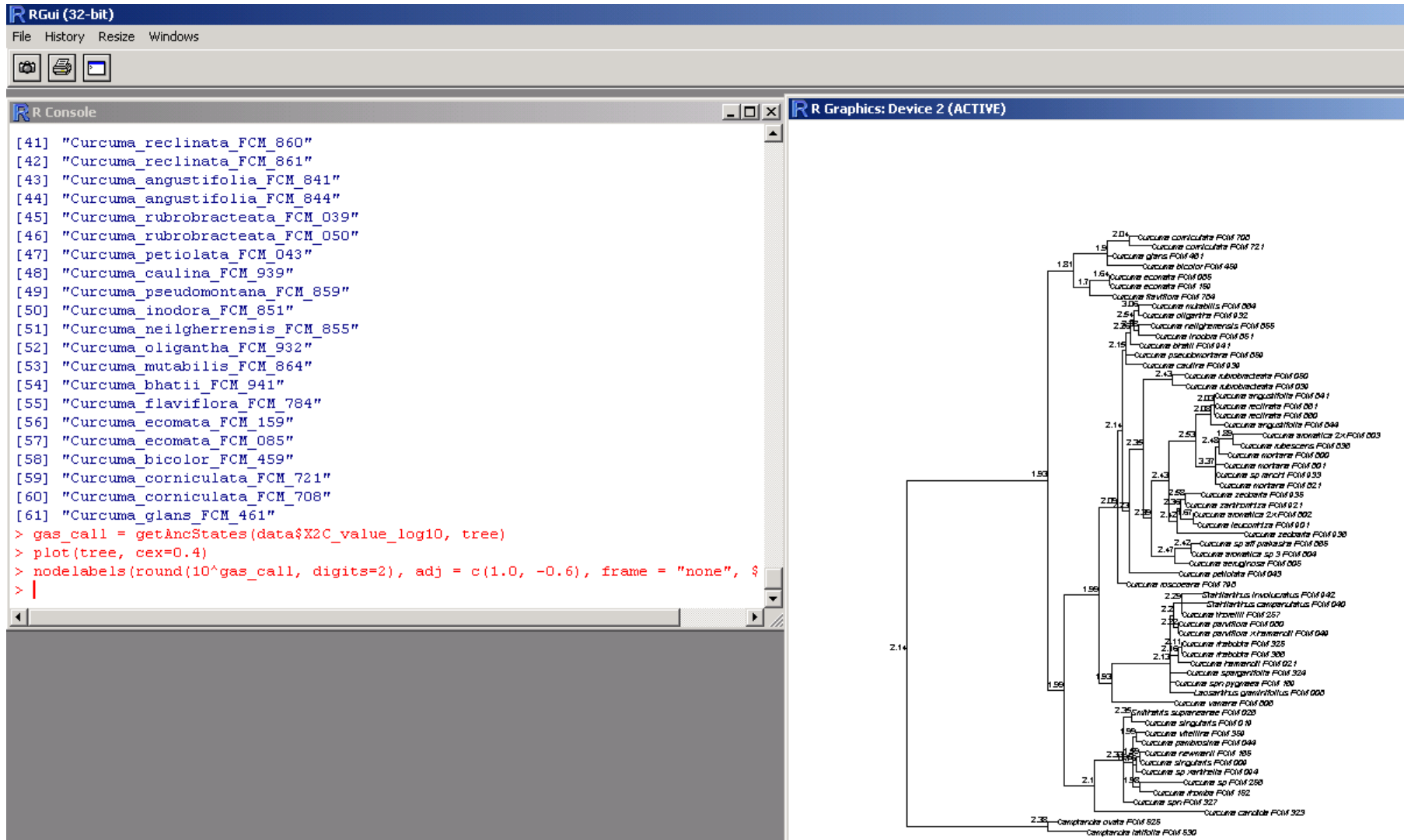
- ML Strom: ITS_ML_curcuma.txt
- Data: Curcuma_GS.csv

Programy

- R, knihovna Geiger

rekonstrukce ancestrálních stavů znaků – R - Geiger

- Ancestrální hodnoty namapované přímo k odpovídajícím uzlům stromu



Rekonstrukce ancestrálních tvarů na základě geometricko-morfometrických dat

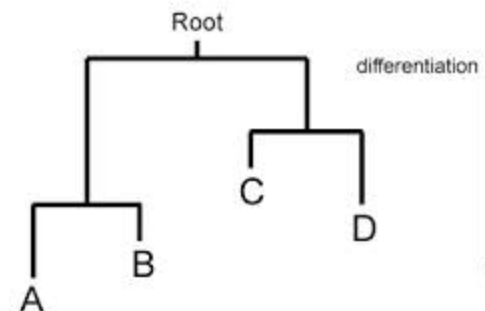
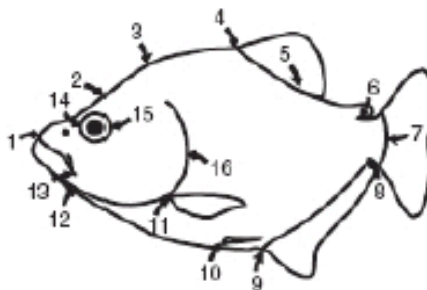
- Tvar – popsán pomocí několika bodů v 2D prostor – tzv. **landmarků**
- každý landmark definován dvěma hodnotami (souřadnice na ose x a y)
- Data popisující tvar – vícerozměrná – $X \text{ landmarků} \times 2 \text{ osy}$

→ Analýza landmarkových dat – specifický aparát – např. analýza TPS

- Potřebná data:
 - Soubor s fylogenetickým stromem
 - nexus formát - blok „TRANSLATE“, nenulové délky větví
 - možno jen jeden strom (na základě MP, ML i bayesovské analýzy)
 - soubor s landmarkovými daty charakterizujícími objekt/tvar
 - tzv. *TPS formát – každý jedinec popsán souborem landmarků
 - jedinci musí být v souboru uspořádáni jako v souboru stromu

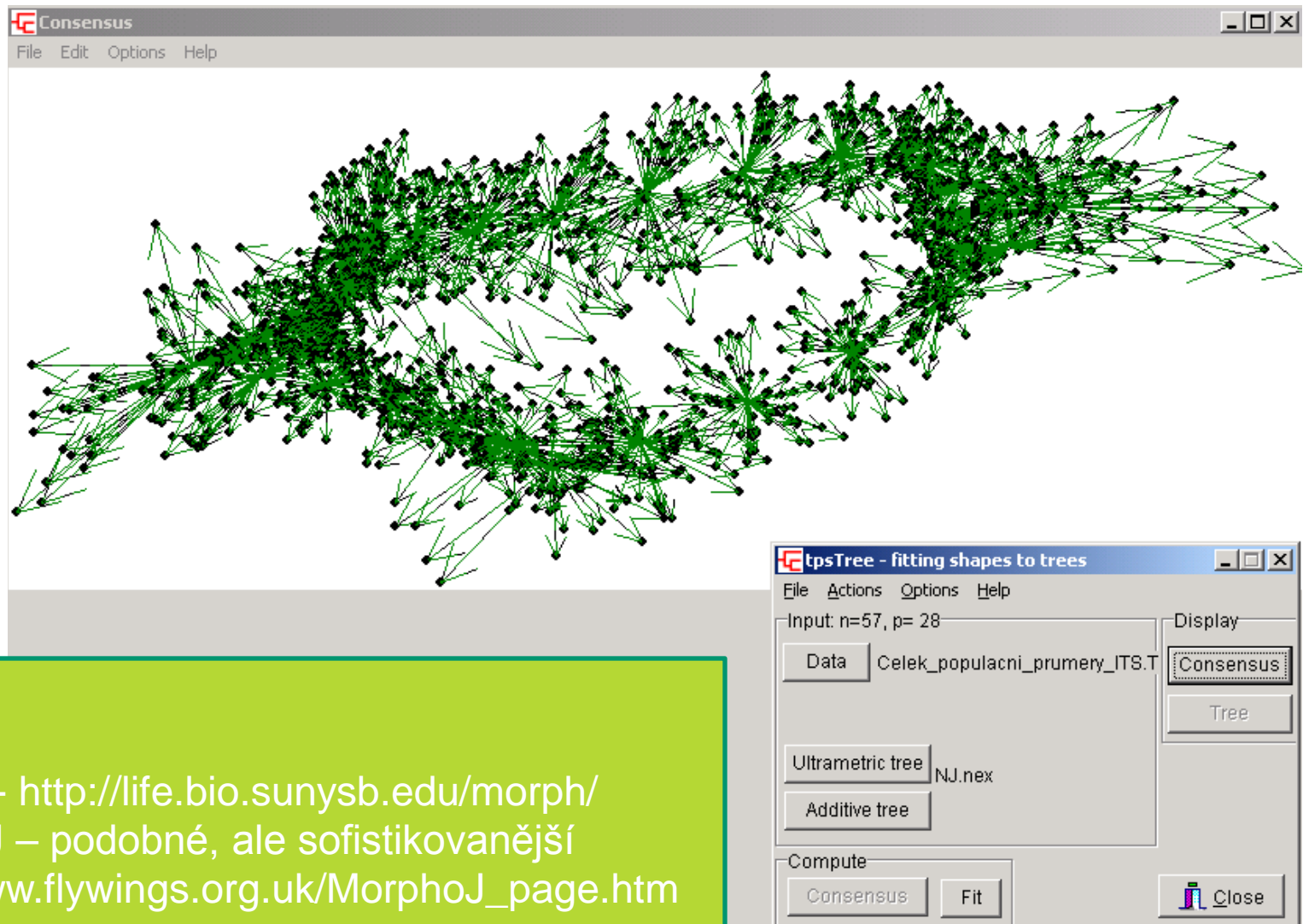
LM=3

353.00000	878.00000
501.00000	978.00000
480.00000	1038.00000



Rekonstrukce ancestrálních tvarů na základě geometricko-morfometrických dat

- TpsTree – program z balíku tps pro analýzu metodou Thin-plate spline

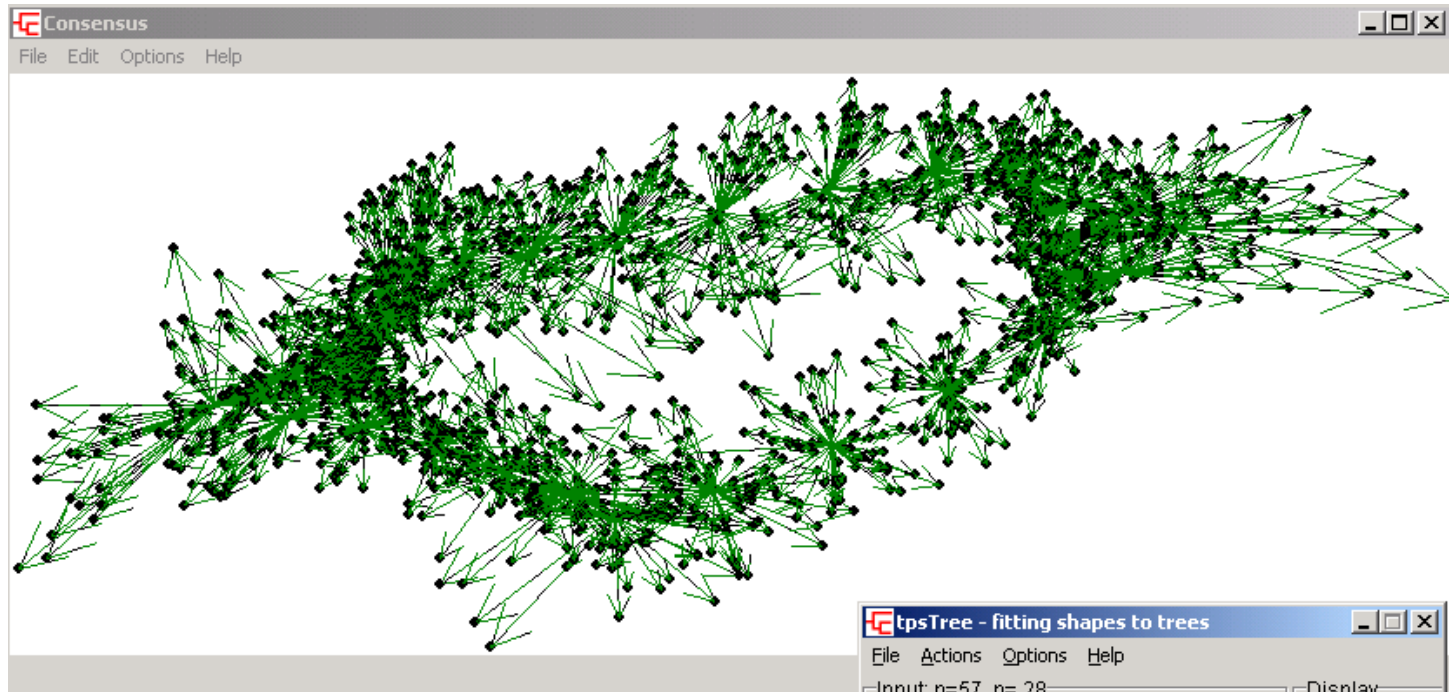


Programy

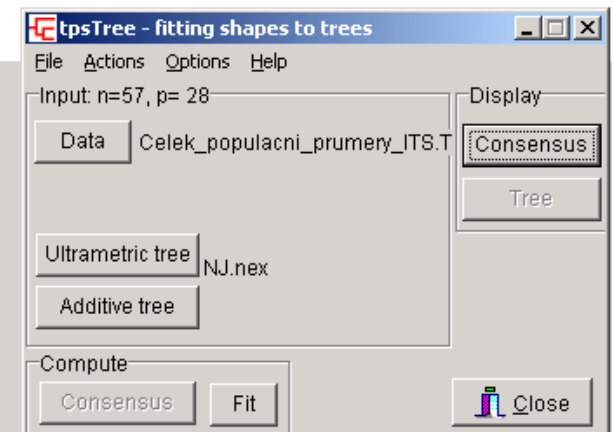
- tpsTree - <http://life.bio.sunysb.edu/morph/>
- MorphoJ – podobné, ale sofistikovanější
http://www.flywings.org.uk/MorphoJ_page.htm

Rekonstrukce ancestrálních tvarů na základě geometricko-morfometrických dat

- TpsTree – program z balíku tps pro analýzu metodou Thin-plate spline
- **Superimpozice** jednotlivých objektů a výpočet **konsensuálního tvaru**

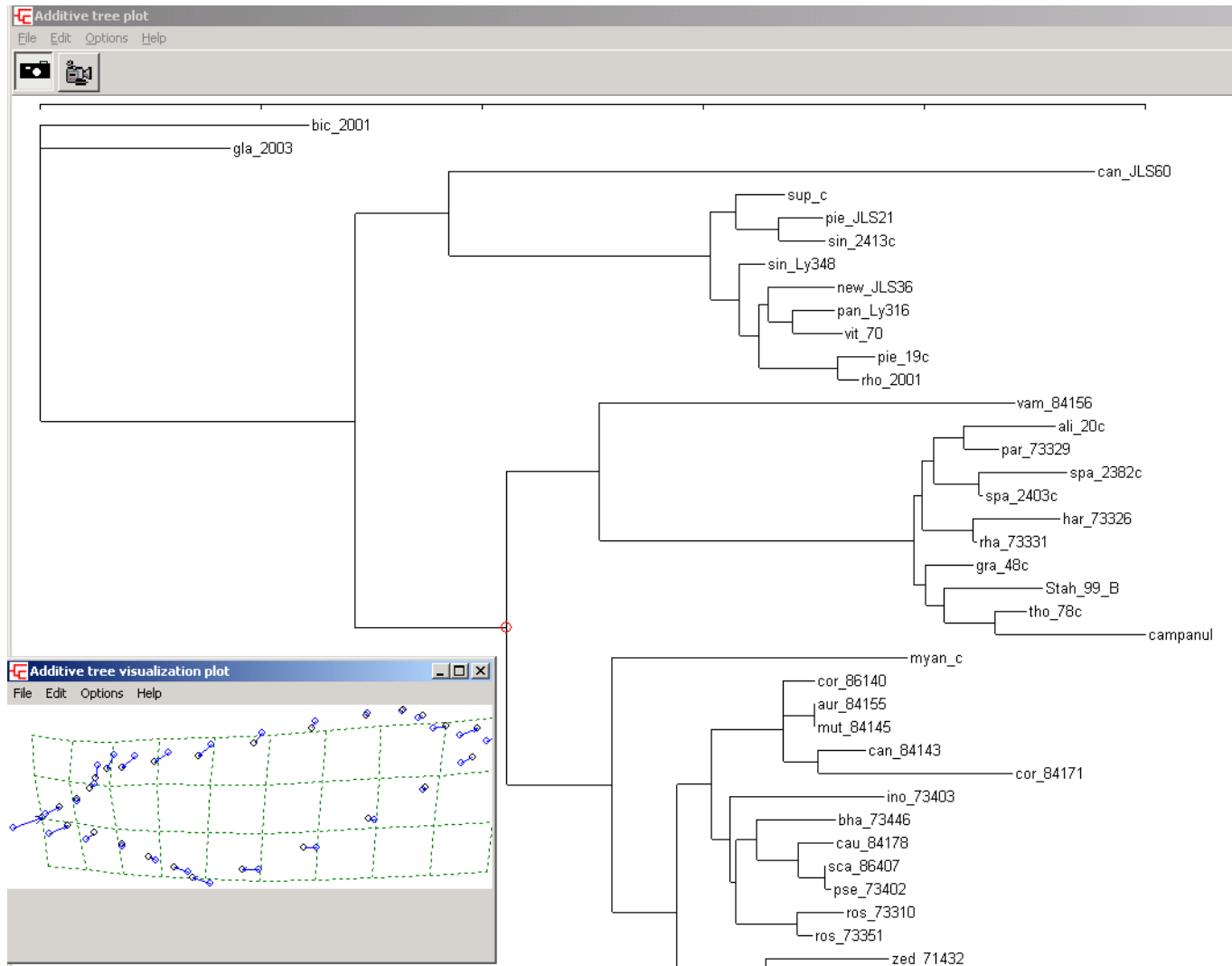


Programy



Rekonstrukce ancestrálních tvarů na základě geometricko-morfometrických dat

- Výpočet nejparsimoničtějších tvarů v každém místě fylogeneze
- square change parsimony



Praktické cvičení – testování fylogenetického signálu

- Vyberte si jeden soubor dat obsahujících fylogenetický strom a k němu pasující data nemolekulární povahy (např. cytologická, morfologická)
 - Uvědomte si, s jakými daty pracujete (kategorická, spojitá apod.)
 - Vyberte si jeden z nástrojů pro odhad parametru **lambda** λ
 - Odhadněte parametr lambda a otestujte, zda je signifikantně větší než 0 a signifikantně menší než 1
 - K jakému výsledku jste došli a co výsledek znamená?

Data

- Alpinoideae_log2C.txt , ITS_matK_ML_rooted.nex (složka „DATA_EZ/R_caper_geiger_k_d“)
- v_nisbet.txt, tree.tre (složka „DATA_TF“)

Programy

- R – Geiger (spojitá i diskrétní data), ML odhad
- BayesTraits (spojitá i diskrétní data), ML a Bayesovské odhady

Praktické cvičení – rekonstrukce ancestrálních znaků

- Vyberte si jeden soubor dat obsahujících fylogenetický strom a k němu pasující data nemolekulární povahy (např. cytologická, morfologická)
 - Uvědomte si, s jakými daty pracujete (kategorická, spojitá apod.)
 - Vyberte si jeden z nástrojů pro rekonstrukci ancestrálních znaků
 - Proveďte analýzu dle návodu k dané metodě/programu
 - Vyberte alternativní nástroj, který vypočítá ancestrální hodnoty jiným způsobem, a proveďte analýzu i touto jinou metodou
 - Porovnejte výsledky
 - Je možné, že pro provedení alternativní analýzy bude nutné upravit vztupní soubory....

Data

- Pro Mesquite - best_tree_figtree.tre, Curcuma_2C.txt
- Pro R – Geiger - Curcuma_GS.csv, ITS_ML_curcuma.txt

Programy

- Mesquite (parsimonické rekonstrukce, spojitá, disktrétní i kategorická data)
- R – Geiger (ML odhady – spojitá i disktrétní data)

Literatura

- Pagel M. (1992): *A method for the analysis of comparative data*. Journal of Theoretical Biology, 156 (4):431-442
- Pagel M (1997): *Inferring evolutionary processes from phylogenies*. Zoologica Scripta 26, 331-348
- Freckleton R.P., Harvey P.H. & Pagel M. (2002): *Phylogenetic analysis and comparative data: A test and review of evidence*. American Naturalist, 160:712-726
- Nunn C. L. (2011): *The Comparative Approach in Evolutionary Anthropology and Biology*. The University of Chicago Press. 380 pp.