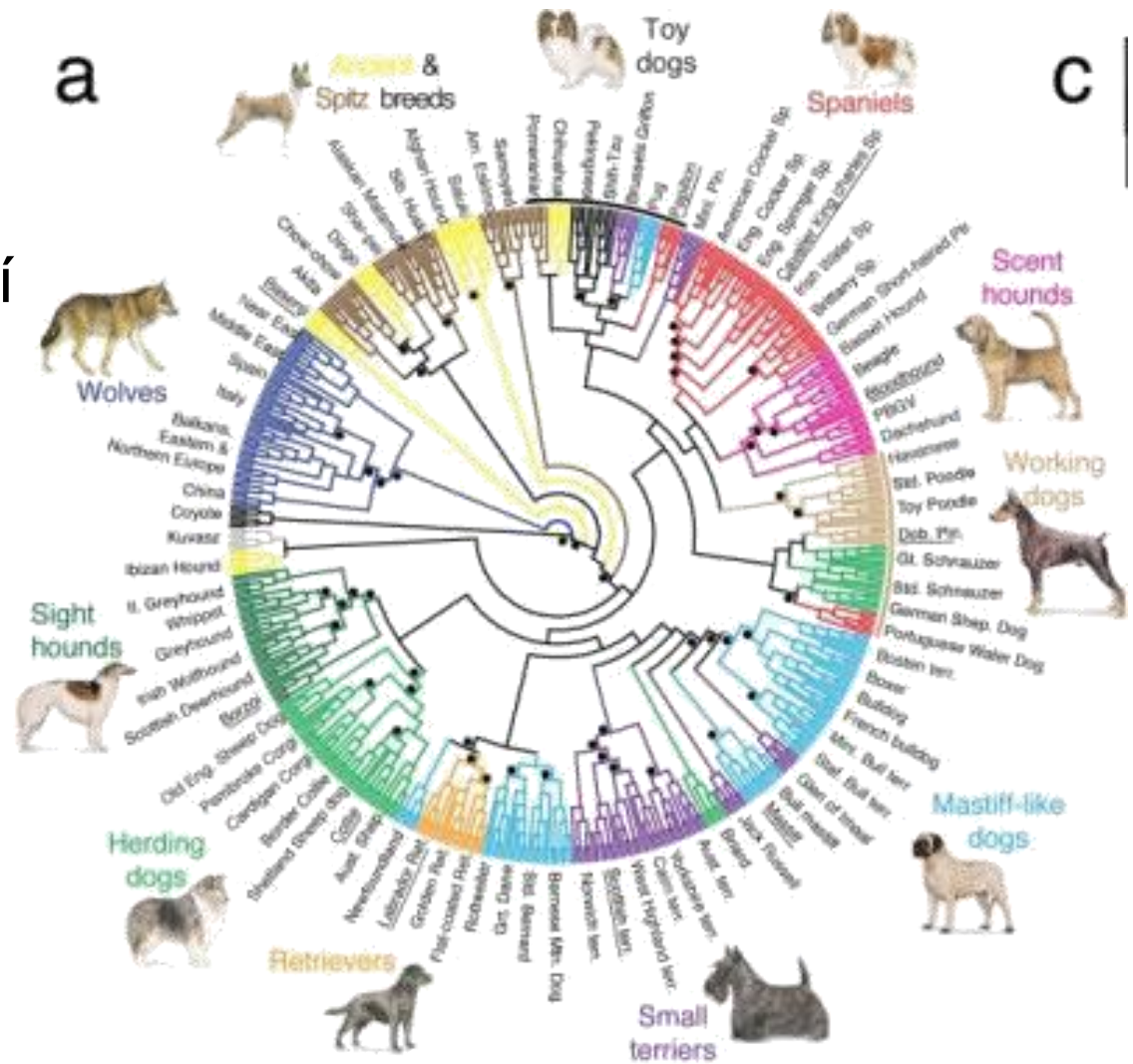


VII. Fylogenetická analýza

RNDr. Karel Berka, Ph.D.

Univerzita Palackého v Olomouci

- **Fylogeneze**
 - Vznik a vývoj jednotlivých linií organismů

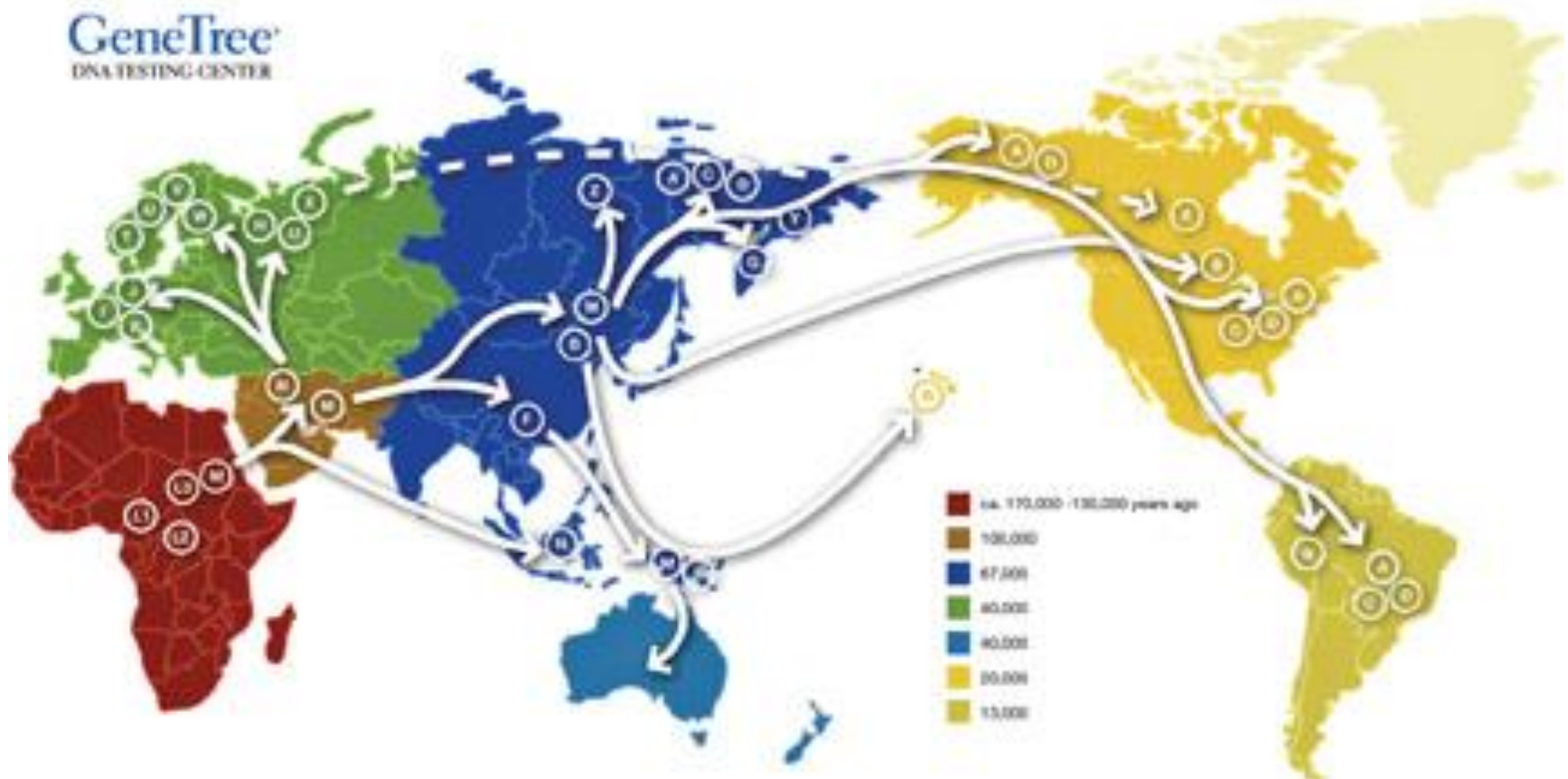


Vývoj člověka



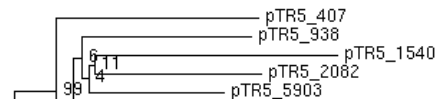
GeneTree
DNA TESTING CENTER

mtDNA Haplogroup Migration Patterns



phylogenetic analysis

pTR5.phb



Odhaduje evoluční souvislosti mezi daty

Výchozí předpoklady:

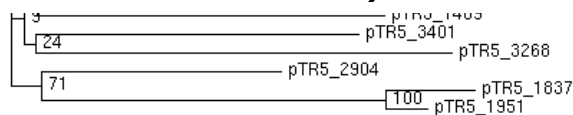
společný předek

kumulace jednotlivých změn

změny jsou náhodné

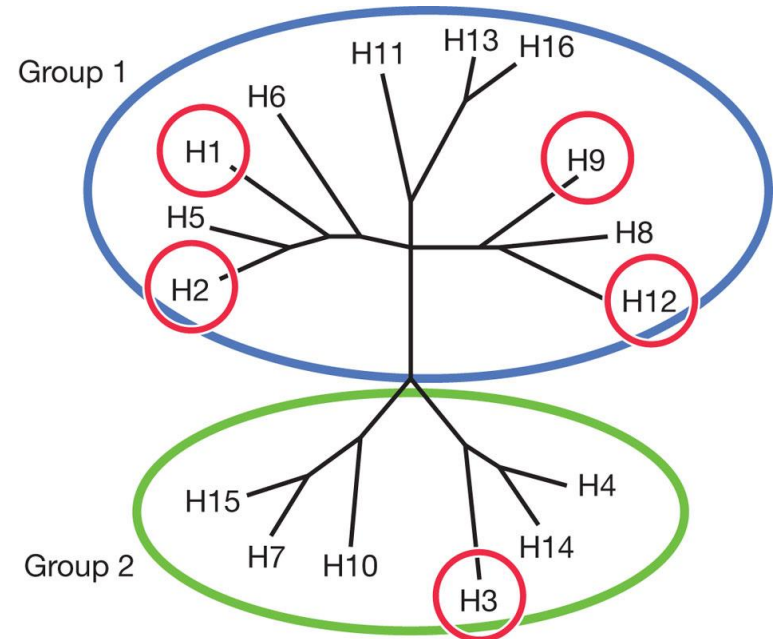
přibližně stejná evoluční rychlost

(molekulární čas)



A k čemu je to dobré ... ?

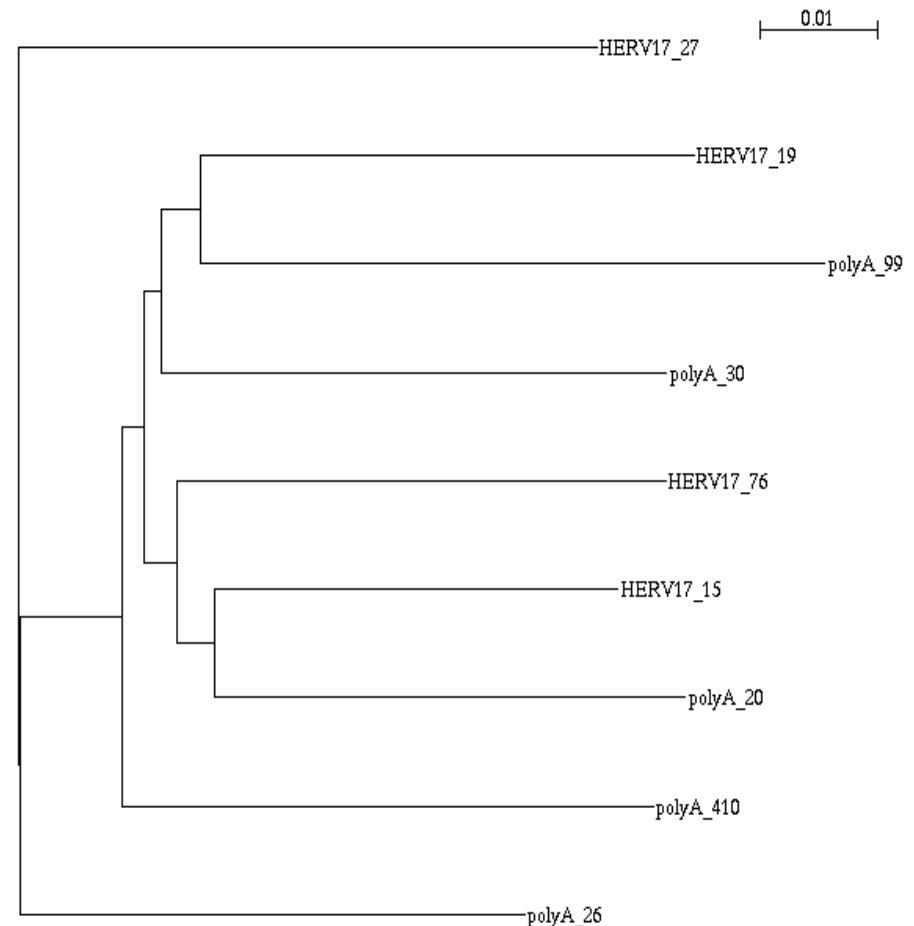
- Odhalení biodiverzity
- Klasifikace
- Testování evolučních hypotéz
- Biogeografie
- Genetické inženýrství

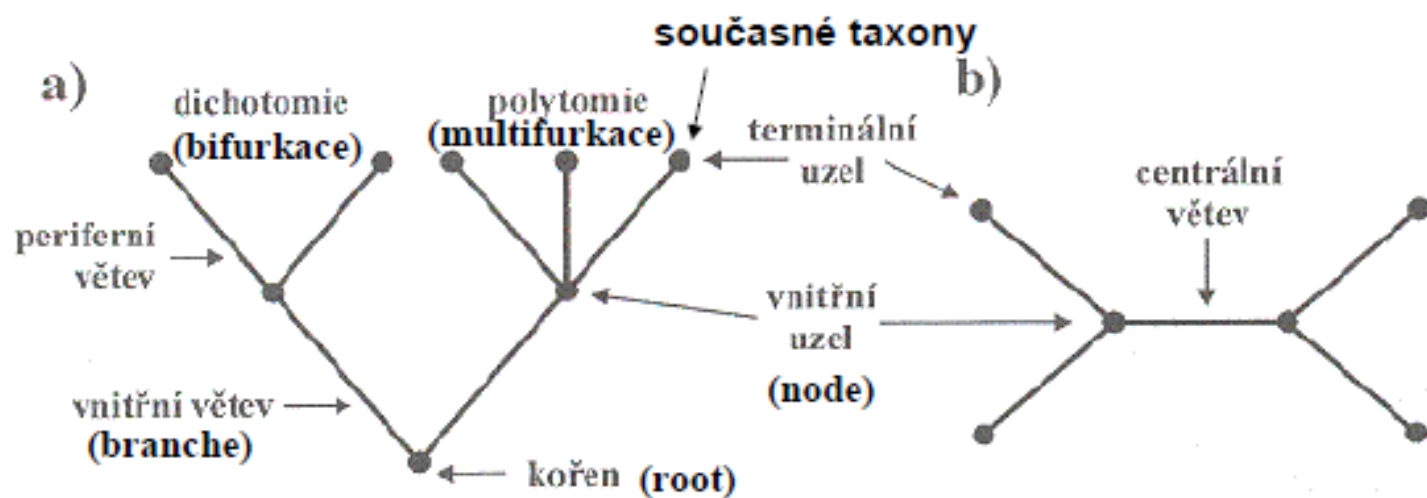


terminology

nody (uzly)

- vnitřní
- vnější
- **větve**
- **topologie stromu**
- **bifurkační strom**
- **aditivní strom**
- **ultrametrický strom**
- **kořen (root) stromu**
- **pravdivý (korektní)**
- **odvozený**





zakořeněný strom (rooted tree)

nezakořeněný strom (unrooted tree)

c)



hvězdicový



částečně vyřešený



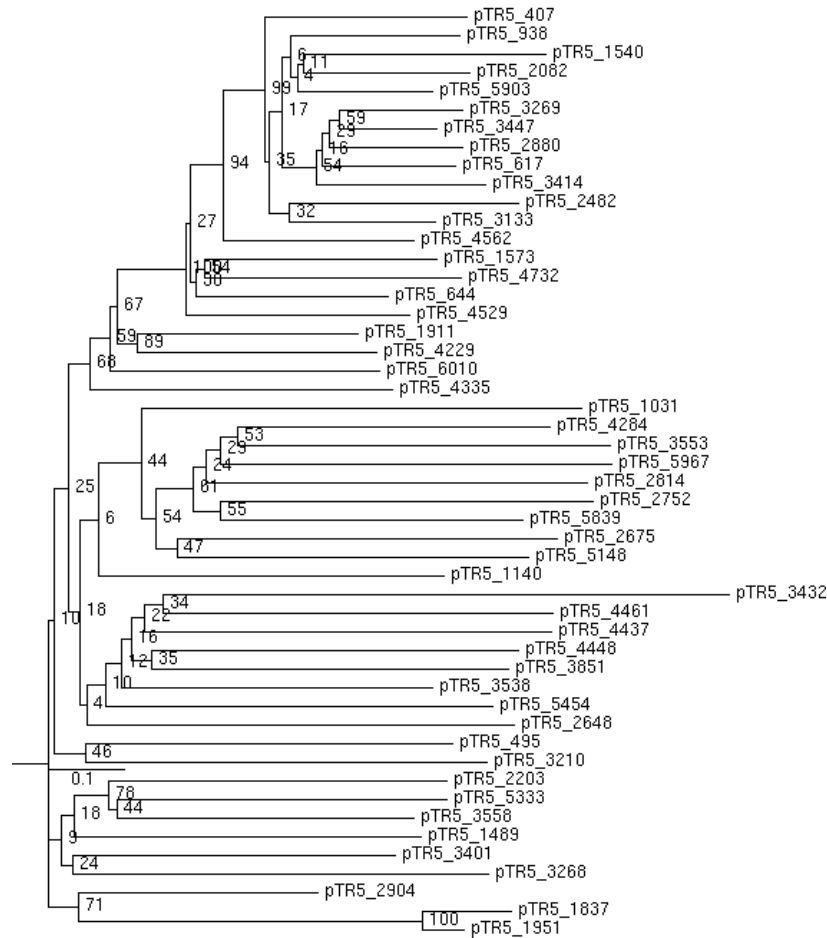
plně vyřešený

example

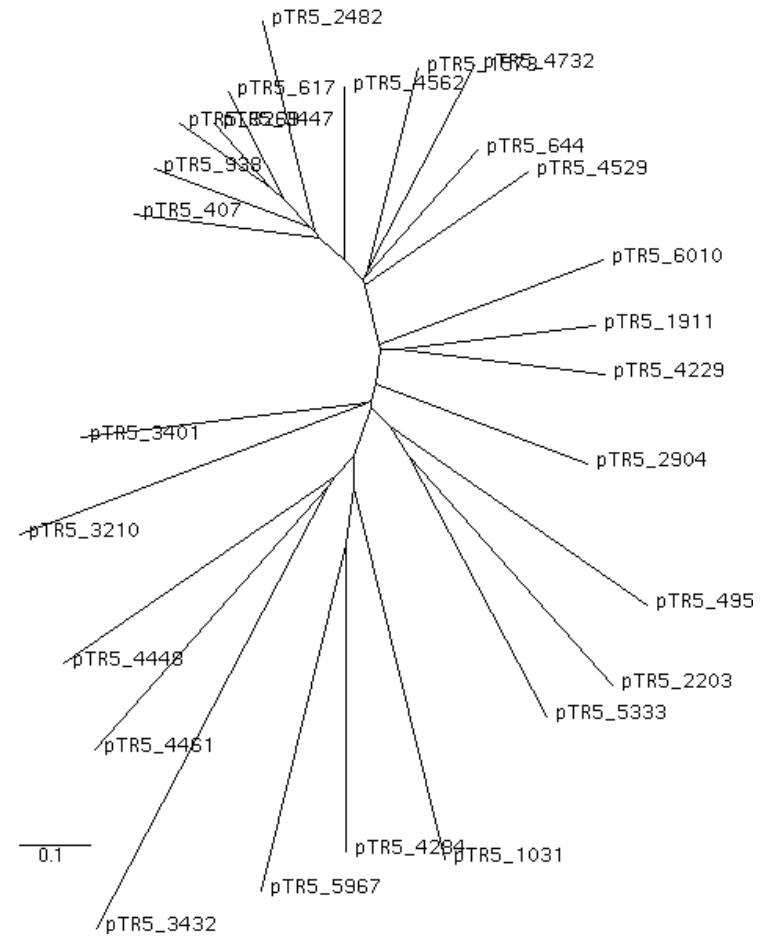
```
( ( ( (
    polyA_26:0.042779,
    HERV17_27:0.049179
  ):0.008643,
  polyA_410:0.045034
):0.001912,
( (
    polyA_20:0.039953,
    HERV17_15:0.034230
  ):0.003074,
  HERV17_76:0.041414
):0.002812
):0.001440,
polyA_30:0.042838,
(
  polyA_99:0.052972,
  HERV17_19:0.041888
):0.003257
)
```


example

pTR5.phb



pTR5.phb2



Jak na to ...

- Alignment
- (Výpočet distancí)
- Fylogenetický strom
- Spolehlivost
- Vizualizace

methods

algorithmical methods:

fast

giving one result (**jeden strom**), but not everytime the best one (local optimum)

= používají sled specifických kroků

optimisation methods:

slower, but can found global maximum

gives often range of the best results

= mají kritérium optimálnosti

methods

Requirement for input data:

Alignment only of homologous parts.
Skip gaps.

(trees based on other data:
restriction analysis
unique insertions or deletions)

algorhitmical (distant) methods

Metoda:

shluková analýza

Input:

matice vzdáleností

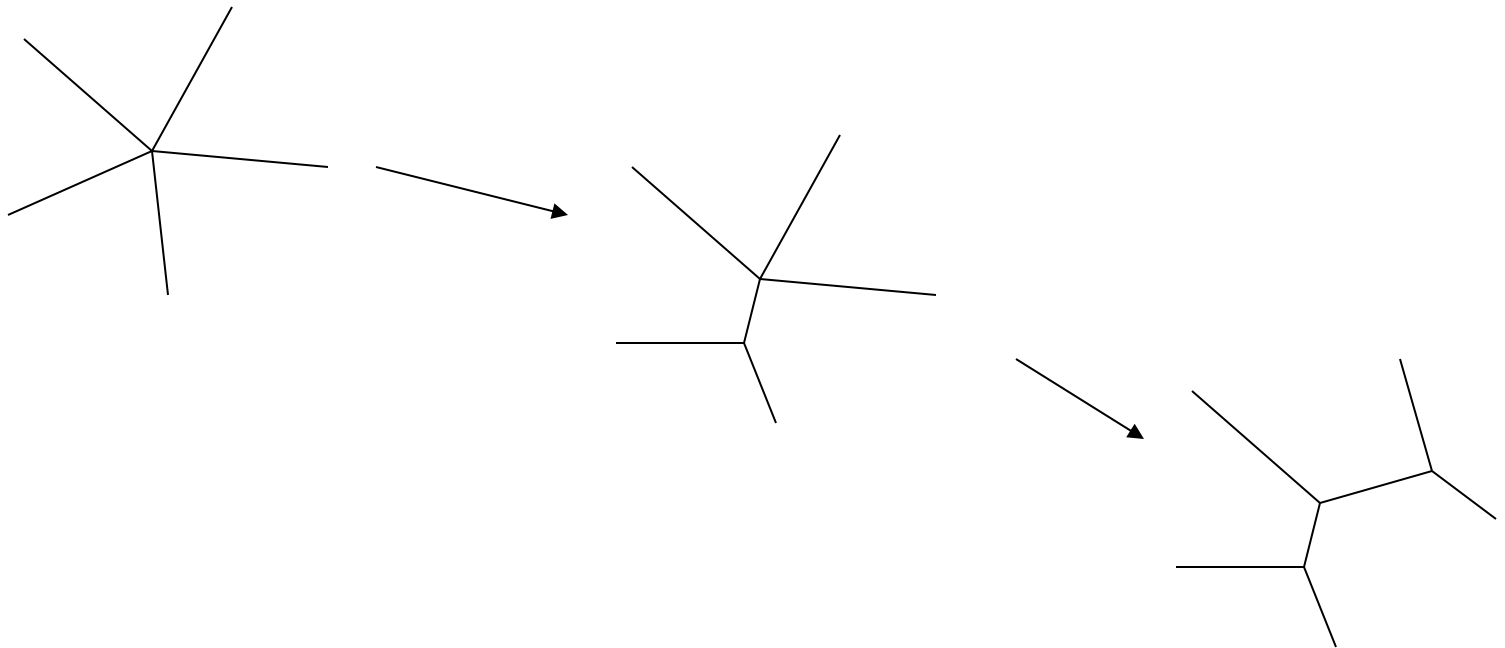
UGPMA (Unweighted pair group method with
arithmetic averages)

WGPMA

Neighbour-joining

neighbour-joining

Star decomposition method



substitutional models

DNA:

Single parametric: Jukes-Cantor

Two parametric: Kimura

Transition: purin - purin

Transversion: pyrimidin - purin

For proteins:

Substitution matrix (BLOSUM etc.)

Výpočet p distancí

p = podíl rozdílných nukleotidů mezi 2 sekvencemi

p = počet rozdílných bází v sekvenci / počet všech nukleotidů

$$p = n_d/n$$

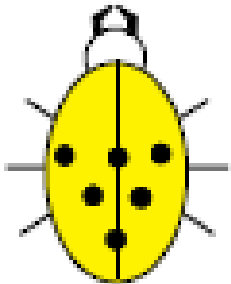
GATCATTAAATGCGATAT
GACCGTTATTGCCATAT

$$p = 4/17 = 0,23$$

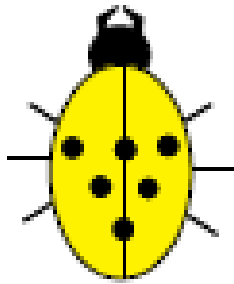
matrix of distances

```

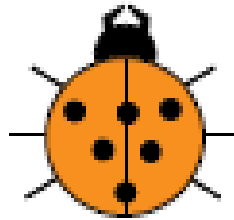
      9
polyA_26
polyA_30      0.1102
polyA_20      0.1144      0.1027
polyA_99      0.1326      0.1100      0.1237
polyA_410     0.1089      0.1009      0.1067      0.1150
HERV17_27     0.1070      0.1263      0.1285      0.1504      0.1198
HERV17_76     0.0960      0.1024      0.0953      0.1221      0.1036      0.1188
HERV17_19     0.1045      0.0994      0.1019      0.1097      0.1059      0.1304      0.0975
HERV17_15     0.0980      0.0975      0.0841      0.1170      0.0977      0.1127      0.0860      0.0927
```



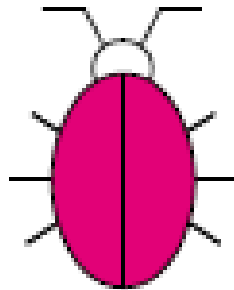
A



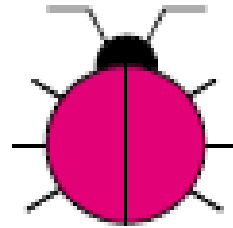
B



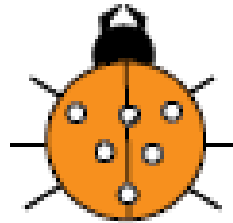
C



D



E



F

optimisation methods

Method: search for optimal tree

Input: multiple alignment

parsimony

maximální věrohodnost - maximum likelihood - ML

pairwise distant methods

parsimony

Parsimony is the use of the simplest or most frugal route of [explanation](#) available
- preference for the least complex explanation for an observation.

A: TATGTTTC

B: TATTTTC

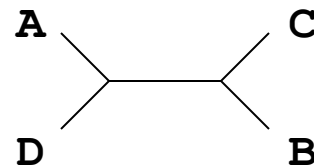
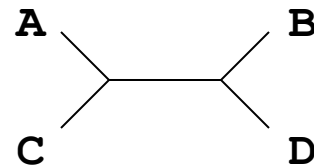
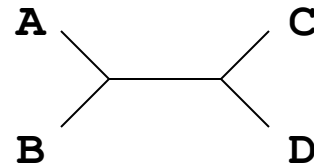
C: TACGTAC

D: GACTTAA

vybíráme strom s
minimální délkou

- tj. nejmenším počtem
evolučních kroků

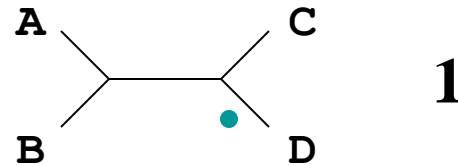
3 možné modely



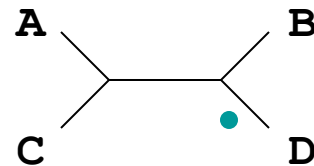
parsimony - step 1

A: **T**ATGTTTC
B: **T**ATTTTC
C: **T**ACGTAC
D: **G**ACTTAA

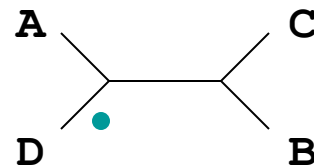
bod ukazuje, kde se to dělí



1



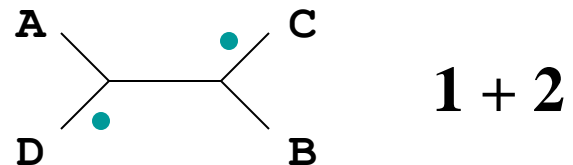
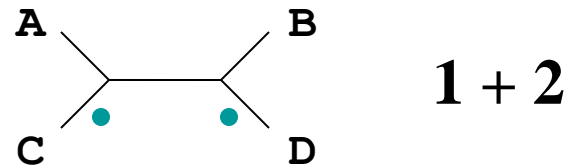
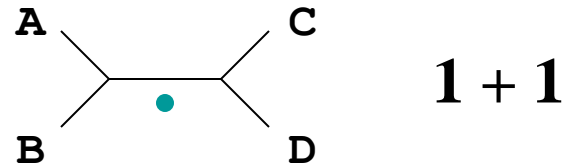
1



1

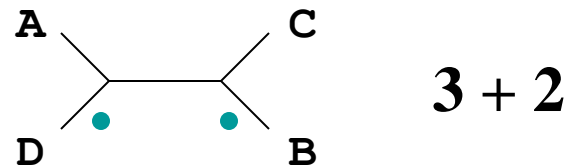
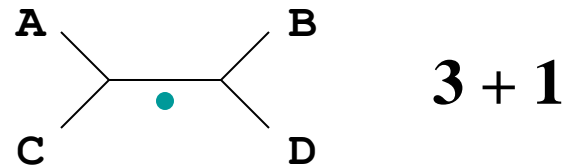
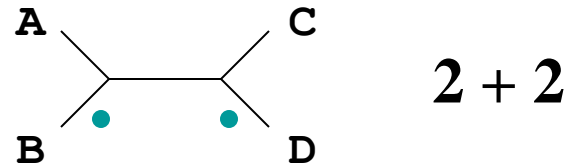
parsimony - step 2

A: TATGTTTC
B: TATTTTC
C: TACGTAC
D: GACTTAA



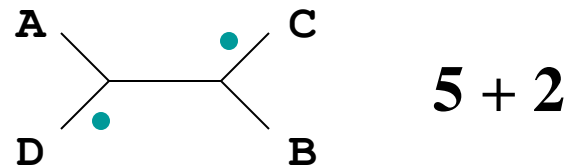
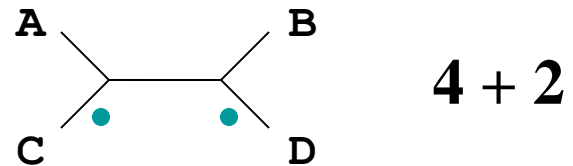
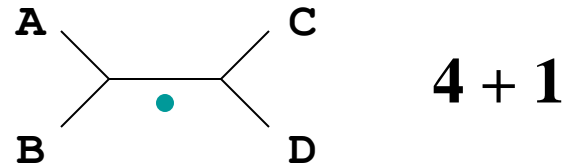
parsimony - step 3

A: TAT**G**TTC
B: TAT**T**TTC
C: TAC**G**TAC
D: GAC**T**TAA



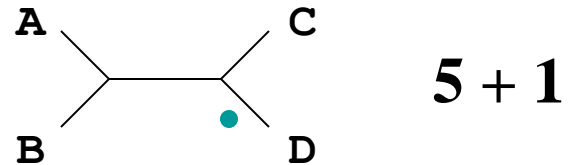
parsimony - step 4

A: TATG**T**C
B: TAT**T**T**C**
C: TACG**T**AC
D: GACT**T**AA

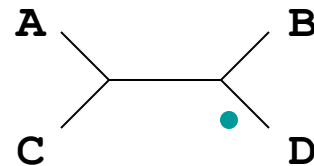


parsimony - step 5

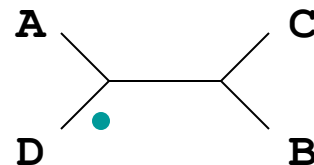
A: TATGTT**C**
B: TATTTT**C**
C: TACGTAC**C**
D: GACTTAA**A**



5 + 1



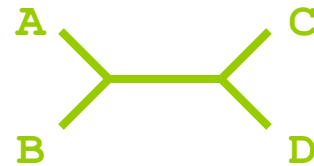
6 + 1



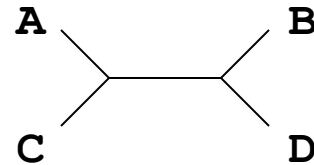
7 + 1

parsimony - result

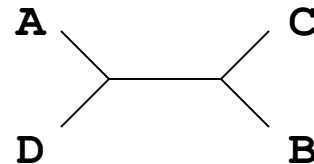
A: TATGTTTC
B: TATTTTC
C: TACGTAC
D: GACTTAA



6



7



8

optimisation methods

Parsimony does not count length of branches and probabilities of individual changes.

Maximum likelihood choose the trees, where less probable events are on longer branches.

differencies

DISTANCE, PARSIMONY, AND MAXIMUM LIKELIHOOD

Distance matrix methods simply count the number of differences between two sequences. This number is referred to as the evolutionary distance, and its exact size depends on the evolutionary model used.

The principle of **maximum parsimony** searches for a tree that requires the smallest number of changes to explain the differences observed among the taxa under study.

A maximum-likelihood approach to phylogenetic inference evaluates the probability that the chosen evolutionary model has generated the observed data.

topology testing

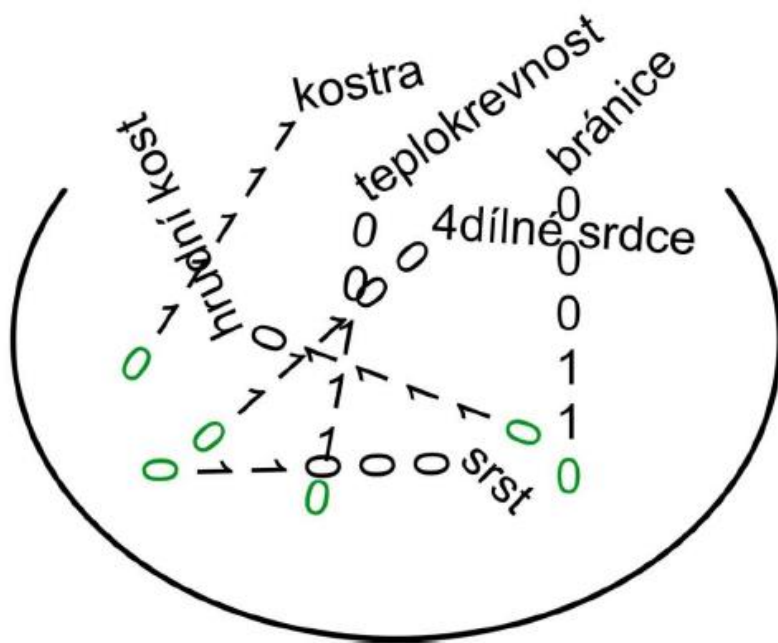
Bootstrap: selection without repeat

Jack Knife: selection without repeat, but shorter sequences or lower number.

	kostra	teplokrevnost	hrudní kost	čtyřdílné srdce	bránice	srst
ryba	1	0	0	0	0	0
žába	1	0	1	0	0	0
pták	1	1	1	1	0	0
králík	1	1	1	1	1	1
člověk	1	1	1	1	1	1
<i>pavouk</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>

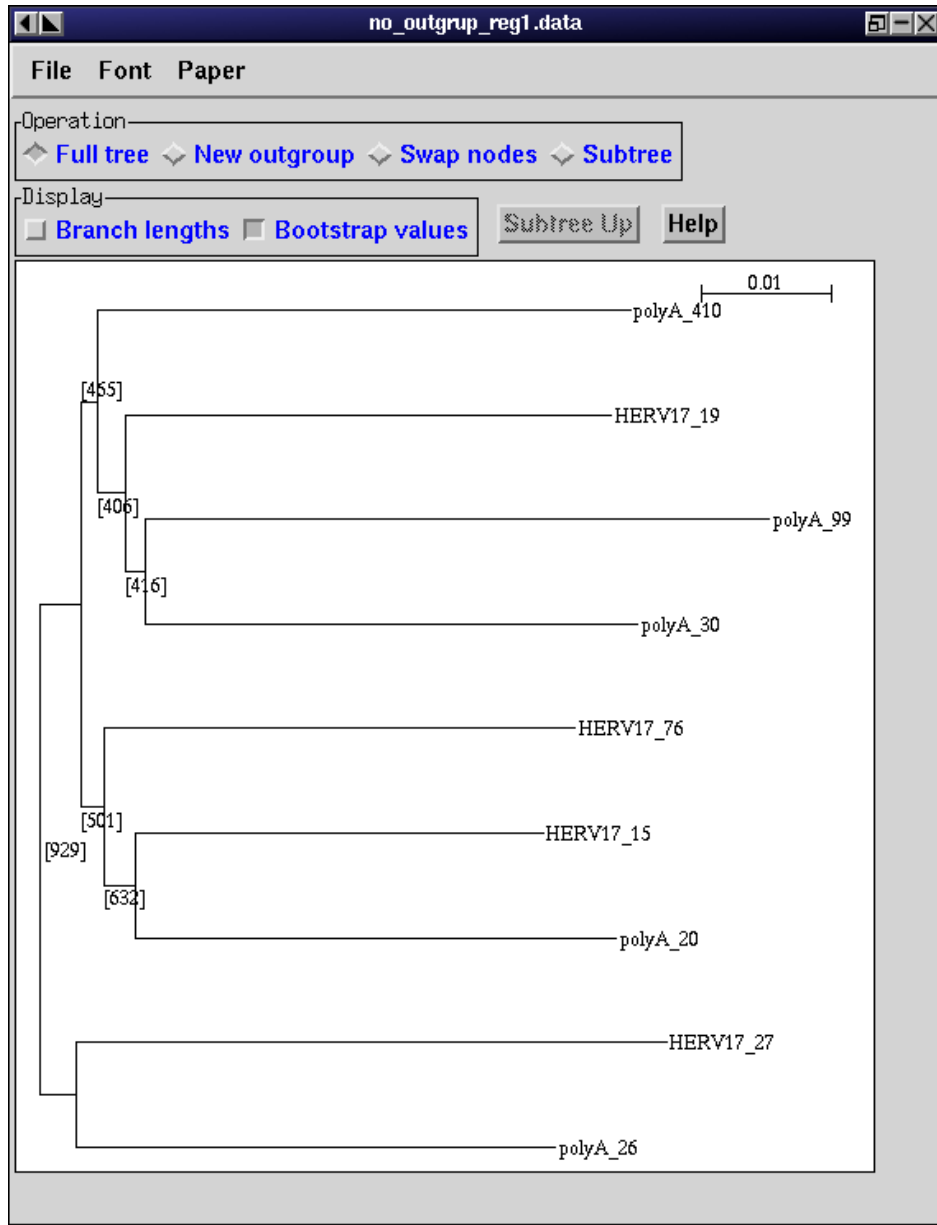
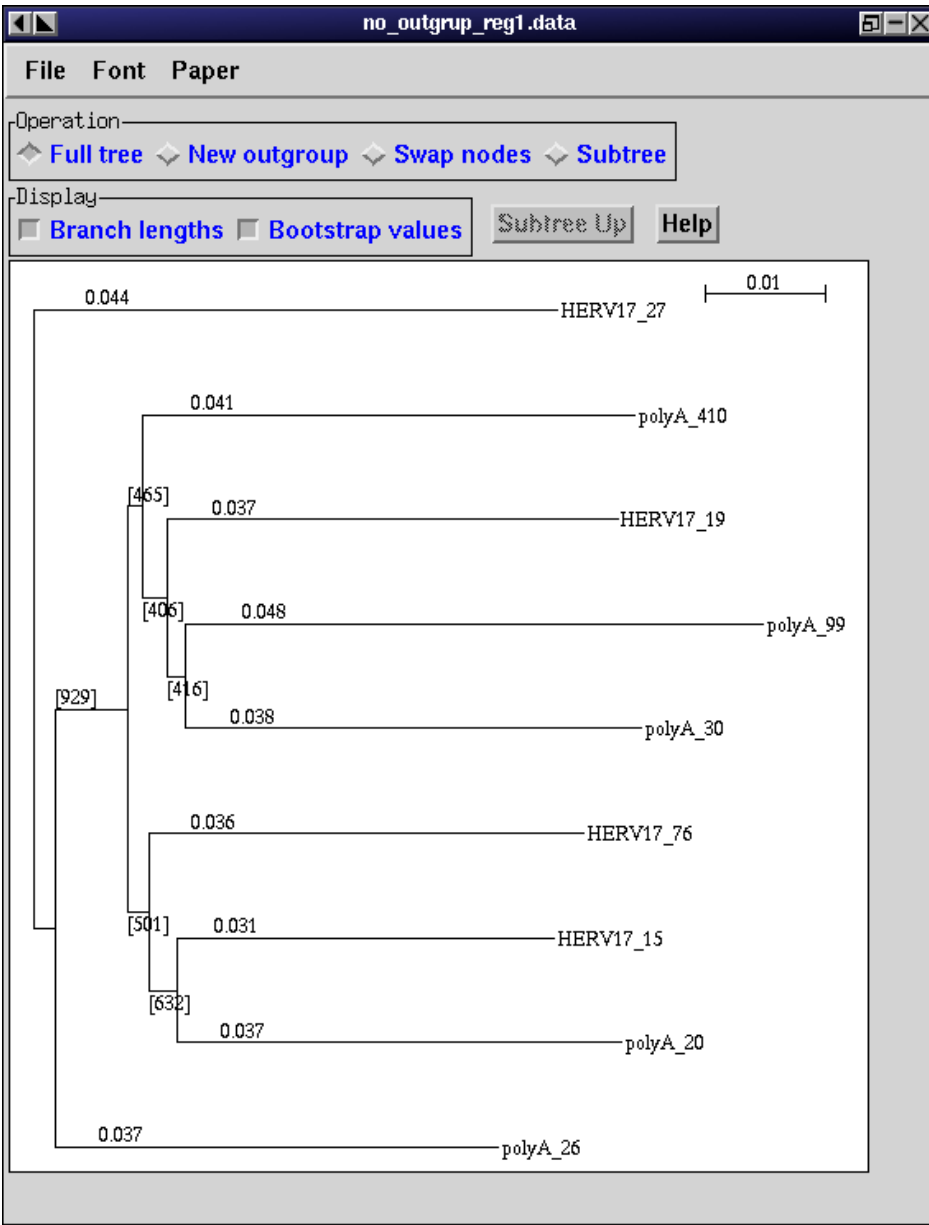
	kostra	teplokrevnost	hrudní kost	čtyřdílné srdce	bránice	srst
ryba	1	0	0	0	0	0
žába	1	0	1	0	0	0
pták	1	1	1	1	0	0
králík	1	1	1	1	1	1
člověk	1	1	1	1	1	1
<i>pavouk</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>

✂ ✂ ✂ ✂ ✂

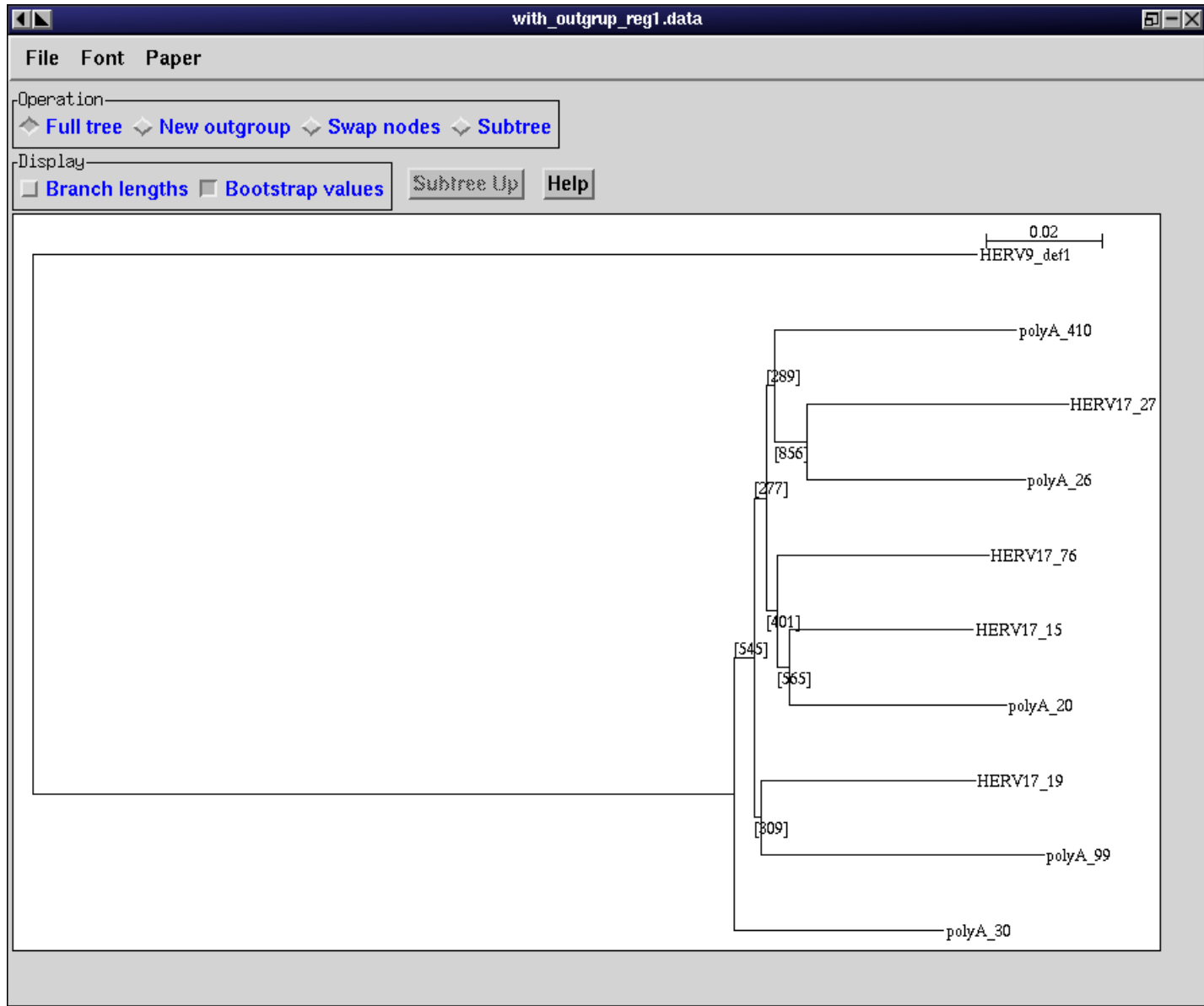


➔ následuje "losování"
a výroba tzv. **pseudomatic**,
které mají stejný počet
druhů i znaků;
znaky se mohou opakovat

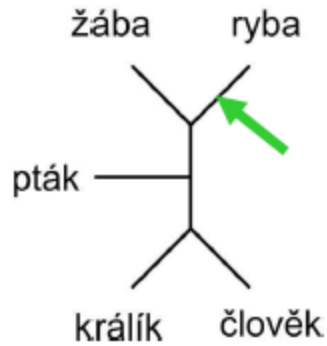
root of the tree



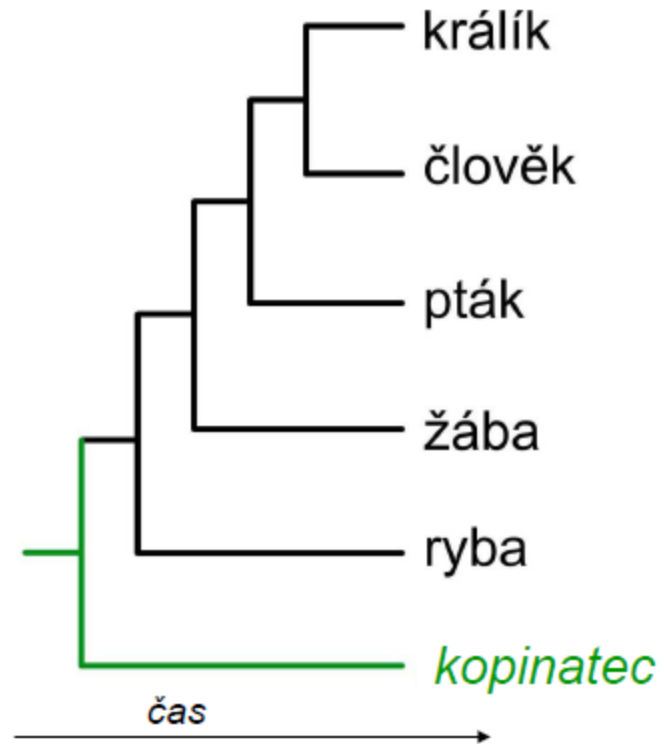
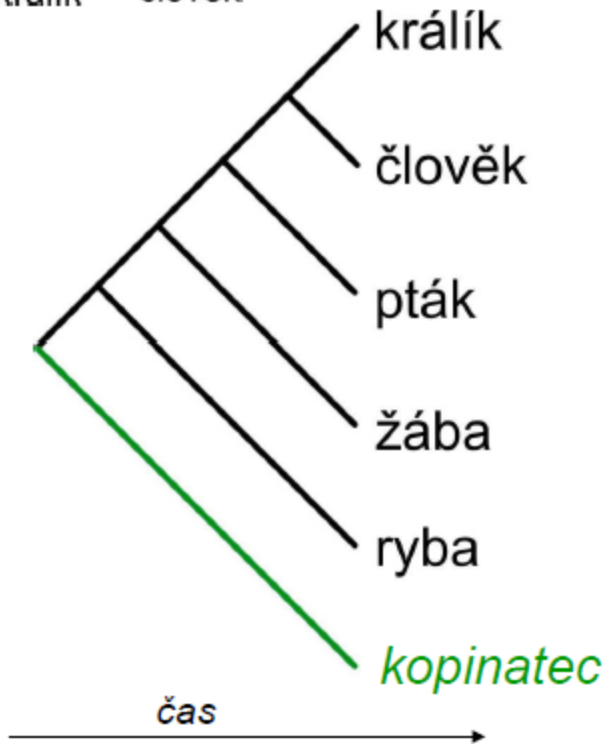
root of the tree



nezakořeněný strom:



kořen stromu pak umístíme do poloviny větve spojující **outgroup** se zbytkem stromu



programs

[http://geta.life.uiuc.edu/~nikos/LINKS/
biocomputing_servers.html](http://geta.life.uiuc.edu/~nikos/LINKS/biocomputing_servers.html)

[http://bioweb.pasteur.fr/seqanal/
phylogeny/phylip-uk.html](http://bioweb.pasteur.fr/seqanal/phylogeny/phylip-uk.html)

[http://evolution.genetics.washington.edu/
phylip/software.html](http://evolution.genetics.washington.edu/phylip/software.html)

- http://web.natur.cuni.cz/zoologie/biodiversity/prednasky/GenetickeMetodyVZoologii/Prednasky_2012/ZpracovaniSekvencnichDatFylogeneze_2012.pdf

Cvičení

- http://wiki.bio.dtu.dk/teaching/images/a/a0/L18_CDS.fasta
 - DNA ribozomálních proteinů
- Clustal W2
 - Multiple alignment
 - Phylogenetic tree