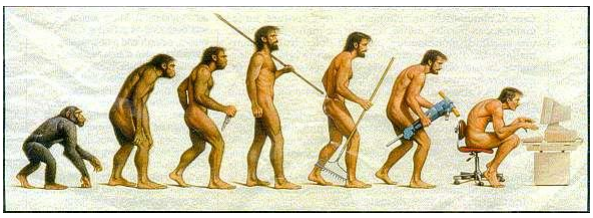
Fylogenetická evoluční analýza

Fylogeneze = vývoj druhu (vývoj nových druhů) procesem evoluce.

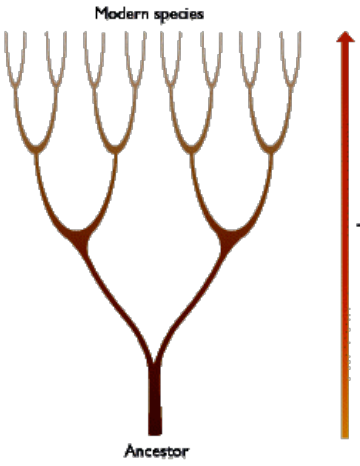
Fylogenetika = věda zkoumající fylogenezi, příbuzenské vztahy a vývoj organismů.



Evoluce bioinformatika

Fylogeneze

Fylogeneze nezahrnuje pouze podobnosti a rozdíly mezi organismy (taxonomie)…



…ale také jejich evoluční vztahy.

Fylogenetická data

- **Fylogenetická data jsou získávána zkoumáním charakteristických znaků studovaných organismů.**

  Prvotně používány MORFOLOGICKÉ znaky.

  Problém – fosilní pozůstatky většinou NEKVALITNÍ, neposkytují žádané informace nebo se VŮBEC nedochovají.



Molekulární fylogenetická data

- **Jediný experiment může poskytnout informace o mnoha znacích.**

```
AAGACGGCACCGACAACGACTACAACGACGCCGTCGTGGTGATCAACTGGCCGCTCGGCT
AGGATGGTACCGACATGGACTACAACGACTCCATCGTCATCCTGAACTGGCCGCTGGGCT
GGGACGGCAACGGC-TGGAC--CAAGGGCGCCTACACCGCCACGAACTGA----------
ACGACGTGCCCGGAACCTATGGCAATAACTCCGGC-TCGTTCAGTGTCAATATTGGAAAG
```

Každá nukleotidová pozice v sekvenci může být považována za jeden ZNAK, který se vyskytuje ve ČTYŘECH rozdílných STAVECH.

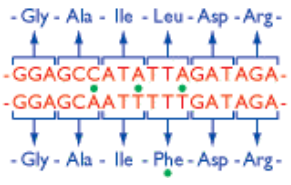
- **Jednotlivé stavy jsou jednoznačné a nezaměnitelné (A x C x G x T).** Na rozdíl od morfologických znaků (tvar), u nichž existuje mnoho přechodových forem.

- **Molekulární data se dají snadno převést do „číselné“ formy.** Vhodné pro matematické a statistické analýzy.

Proteinové sekvence x DNA sekvence

- **Pro fylogenetickou analýzu využívány PŘEVÁŽNĚ DNA sekvence.**

DNA poskytuje mnohem více fylogenetických informací než protein.



Tiché mutace

Variabilita uspořádání genomu (kódující x nekódují oblasti)

PCR, automatické sekvencování

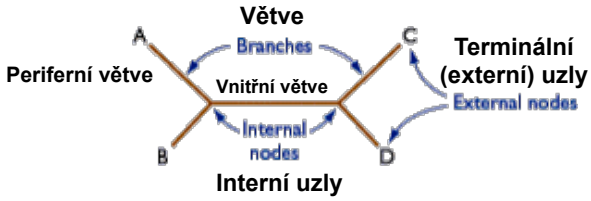
Fylogenetický strom

- **Cíl fylogenetické analýzy** - **fylogenetický strom popisující evoluční vztahy mezi studovanými organismy.**

Současné taxony (geny) = terminální (externí) uzly, vrcholy

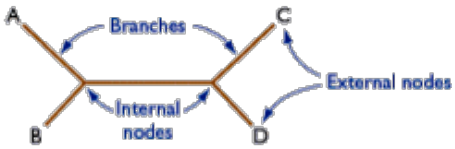
Interní uzly = rozdělení společného „předka“

Délky větví = uměrné velikosti změny v průběhu evoluce

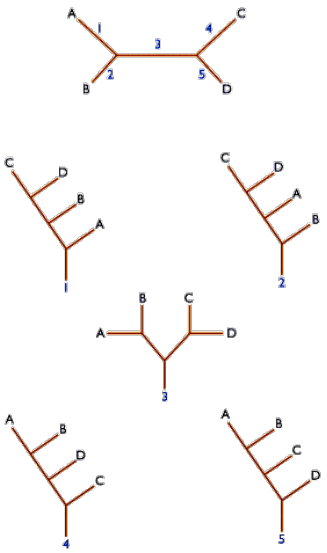


Fylogenetický strom (strom)

Fylogenetický strom

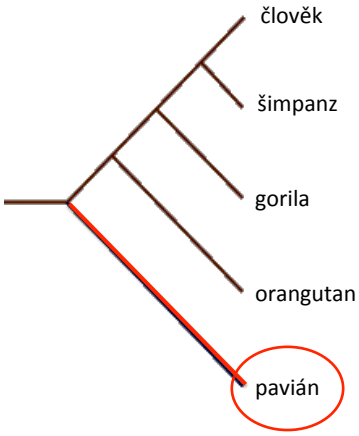


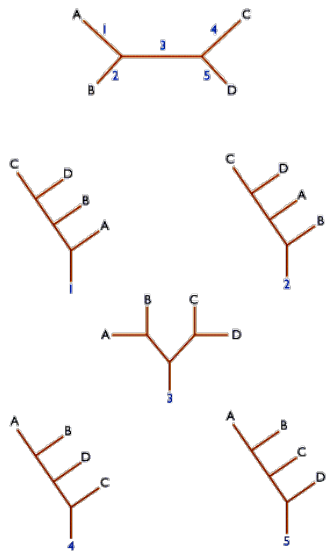
Fylogenetický strom BEZ KOŘENE (unrooted). Není známý nejstarší společný předek (bod). Vypovídá pouze o příbuzenských vztazích mezi geny, ne o „cestě“ kterou se evoluce ubírala.



Fylogenetický strom S KOŘENEM (rooted). Nutný alespoň jeden gen, který je méně příbuzný s A,B,C,D, než jsou tyto geny mezi sebou navzájem = „outgroup“

Fylogenetický strom





Fylogenetický strom S KOŘENEM (rooted). Nutný alespoň jeden gen, který je méně příbuzný s A,B,C,D, než jsou tyto geny mezi sebou navzájem = „outgroup“

„Genový“ strom x „druhový strom“

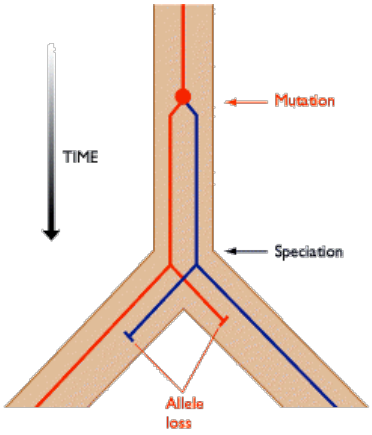
- **Genový strom – odvozen ze srovnání ortologních genů. Předpokládá se, že bude přesnější než strom získaný pomocí morfologických dat.**
- **Genový strom ≠ druhový strom.**

  Genový strom – vnitřní uzly představují rozdělení původního GENU (mutace).

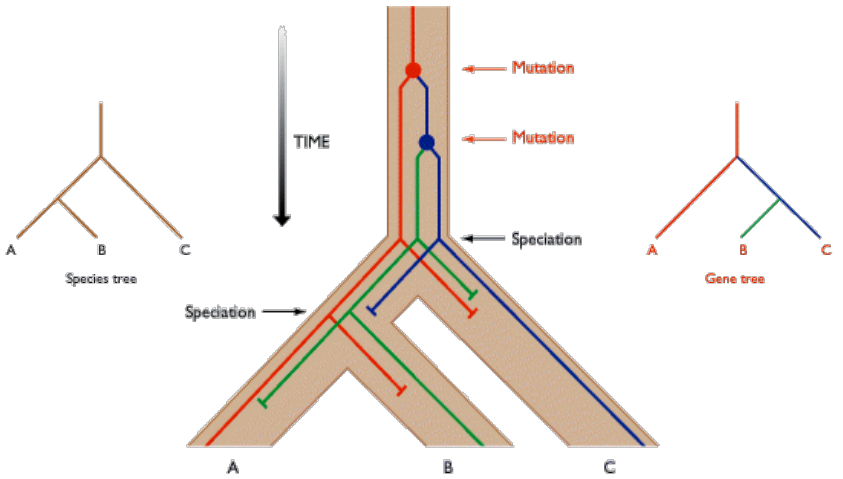
  Druhový strom – vnitřní uzly představují rozdělení populace původního DRUHU do dvou skupin (geografická izolace).

„Genový“ strom x „druhový strom“

- **Mutace a vznik nového druhu se s největší pravděpodobností neodehrají současně.**
- **Mutace předchází separaci – v populaci se nacházejí obě alely genu. Po rozdělení populací může dojít ke ztrátě jedné alely.**

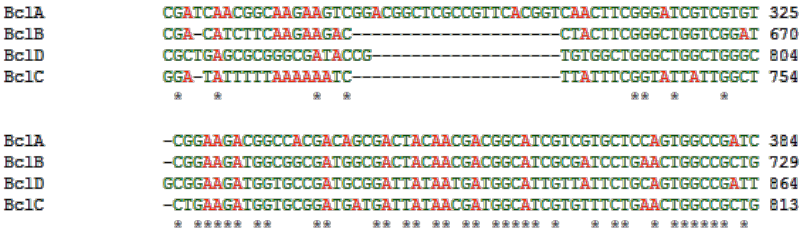


„Genový“ strom x „druhový strom“



Tvorba evolučních stromů

- **„Alignment“ sekvencí – nezbytný pro vytvoření stromu.** **Vyhodnocení rozdílů mezi jednotlivými nukletidovými sekvencemi, většinou „multiple alignment“.**



Jak převést „multiple alignment“ na strom?

- **Neexistuje „nejlepší metoda“.** **Několik metod je používáno souběžně, žádnou nelze označit za lepší než ostatní.**

Jak převést „multiple alignment“ na strom?

- **Distanční matice.**

  Slouží k určení délky větví.

Multiple alignment

```
1 AGGCCAAGCCATAGCTGTCC
2 AGGCAAAGACATACCTGACC
3 AGGCCAAGACATAGCTGTCC
4 AGGCAAAGACATACCTGTCC
```

4/20

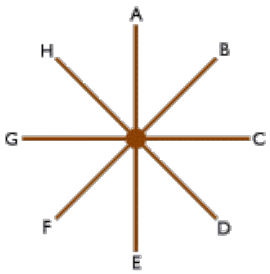
Distance matrix

	1	2	3	4
1	–	0.20	0.05	0.15
2		–	0.15	0.05
3			–	0.10
4				–

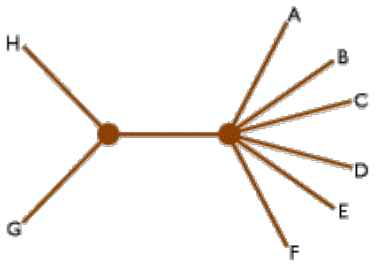
Jak převést „multiple alignment“ na strom?

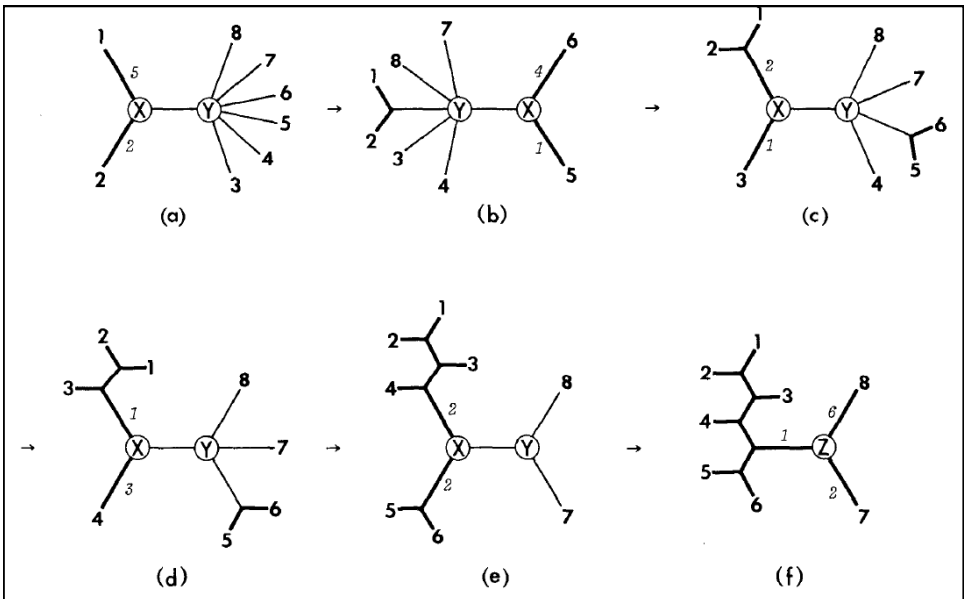
- **Neighbor-joining method– „spojování sousedních objektů“ (Saitou a Nei 1987) . Využívá distanční matici.**

(A) The starting point for the neighbor-joining method



(B) Removal of two sequences from the star





The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees[1]

Naruya Saitou[2] *and Masatoshi Nei*

Jak převést „multiple alignment“ na strom?

- **Neighbor-joining method – „spojování sousedních objektů“ (Saitou a Nei 1987) . Využívá distanční matrici.**
  - \+ Jednoduché = rychlé
  - \+ Vhodné pro velké soubory dat
  - \+ Vhodné pro prvotní analýzu
  - \- Informace z alignmentu velmi zredukována
  - \- Poskytuje pouze jeden výsledný strom (unrooted)

Jak převést „multiple alignment“ na strom?

- **Unweighted Pair Group Method with Arithmetic Mean**
- – Využívá distanční matrici.

Ultrametrická metoda, očekává, že všechny terminální konce jsou stejně vzdálené od počátku (molekulární hodiny) - všecny linie se vyvíjejí stejnou rychlostí...

Výsledkem je “rooted” tree

Jak převést „multiple alignment“ na strom?

preciznější metody

- **Metody maximální úspornosti** – maximum parsimony method. Předpokládá (správně???), že evoluce jde nejkratší možnou cestou, tj. správný fylogenetický strom je ten, který požaduje minimum nukleotidových změn, aby bylo dosaženo daného rozdílů mezi sekvencemi.

+ Preciznější
- Větší nároky na manipulaci s daty
- Čím vice sekvencí, tím více topologií stromů je nutné vyzkoušet
- 5 sekvencí = 15 stromů, 10 sekvencí = 2 027 025 stromů

Jak převést „multiple alignment“ na strom?

- **Parsimonie**: Fitchova parsimonie
  Wagnerova parsimonie (reverzibilita změn)
  Dollova parsimonie („novinka“ může zaniknout)
  Caminova-Sokalova parsimonie (změny ireverzibilní)
  Vážená parsimonie
  Generalizovaná parsimonie
- **Metoda maximální pravděpodobnosti**
- **Metoda minimální evoluce**

Jak převést „multiple alignment“ na strom?

preciznější metody

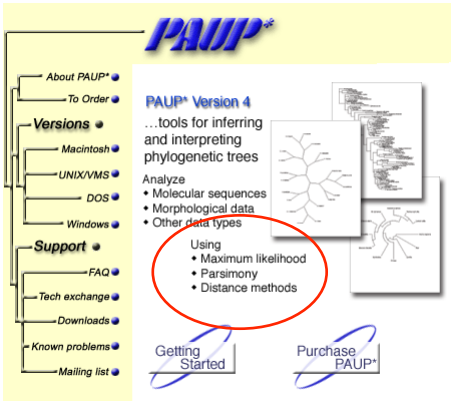
- **Metoda maximální pravděpodobnosti (maximum likehood)**

  statistická metoda – vyhodnocuje pravděpodobnost pro jednotlivé modely- (více mutací v interních větvích snižují pravděpodobnost navrhovaného modelu - podobná maximum parsimony method (např. umožňuje odlišné rychlosti evoluce)
- **Bayesian inference**

  založena na Monte Carlo metodě

Software pro fylogenetickou analýzu

- **BioNJ** (Neighbor-joining method)
- **PAUP** - Phylogenetic Analysis Using Parsimony



http://paup.csit.fsu.edu/index.html

Software pro fylogenetickou analýzu

PHYLIP

PHYLIP (the *PHYL*ogeny *I*nference *P*ackage) is a package of programs for inferring phylogenies (evolutionary trees). It is available free over the Internet, and written to work on as many different kinds of computer systems as possible. The source code is distributed (in C), and executables are also distributed. In particular, already-compiled executables are available for Windows (95/98/NT/2000/me/xp/Vista), Mac OS X, Mac OS 8 and 9, and Linux systems. Complete documentation is available on documentation files that come with the package.

- **PHYLIP** – ***PHYL*ogeny *I*nference *P*ackage**

Methods that are available in the package include parsimony, distance matrix, and likelihood methods

http://evolution.genetics.washington.edu/phylip.html

Software pro fylogenetickou analýzu

Phylogenetic Analysis by Maximum Likelihood (PAML)

Introduction

PAML is a package of programs for phylogenetic analyses of DNA or protein sequences using maximum likelihood. It is maintained and distributed for academic use free of charge by Ziheng Yang. ANSI C source codes are distributed for UNIX/Linux/Mac OSX, and executables are provided for MS Windows. PAML is not good for tree making. It may be used to estimate parameters and test hypotheses to study the evolutionary process, when you have reconstructed trees using other programs such as PAUP*, PHYLIP, MOLPHY, PhyML, RaxML, etc.

http://abacus.gene.ucl.ac.uk/software/paml.html

MacClade

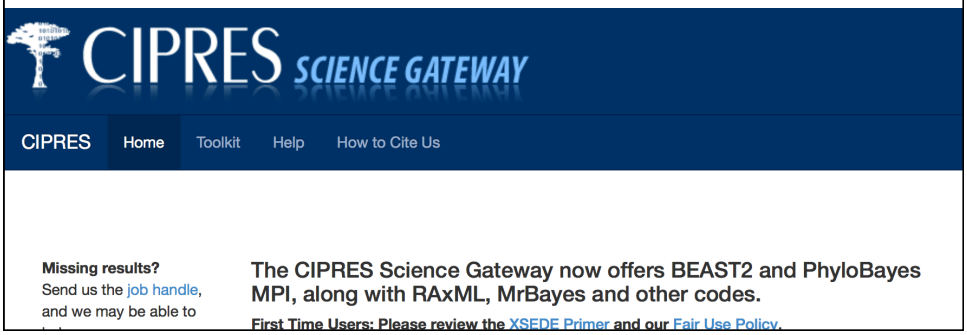
http://macclade.org/index.html

Software pro fylogenetickou analýzu

Portal pro ML a BI

- Maximum likelihood tree (PhyML, RAxML)
- Bayesian tree (Mr.Bayes, BEAST).

https://www.phylo.org/portal2/login!input.action



Postranslational modifications

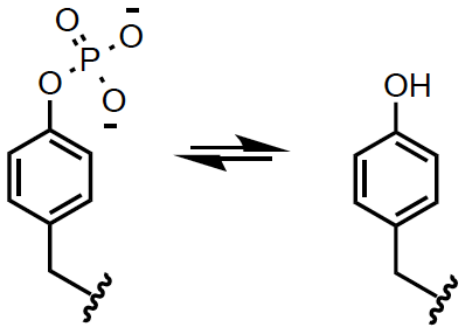
Phosphorylation

-Ser, Thr, Tyr

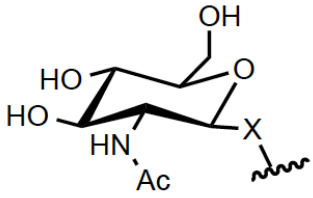
-Control protein activty and structure, as well as protein-protein and protein/nucleic acid interactions

-Kinases phosphorylate, phosphatases dephosphorylate

-Kinases are major drug targets



Glycosylation
-Ser, Thr, Asn
-regulated by glycosyl transferases
-Control protein structure,stability, and trafficking. Regulate protein activity.

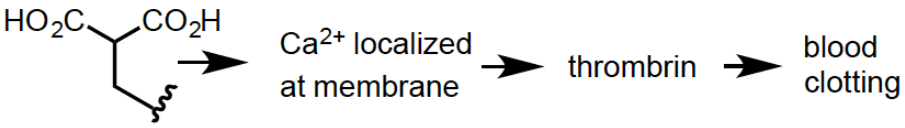


***O*-glykosylace**

***N*-glykosylace**

Acetylation
-N-terminus, Lysine side chains
-Affects chromatin structure and gene expression

Carboxylation
-most common is γ-carboxy-glutamate
-Vitamin K, CO_2, O_2 dependent.
ex. Prothrombrin



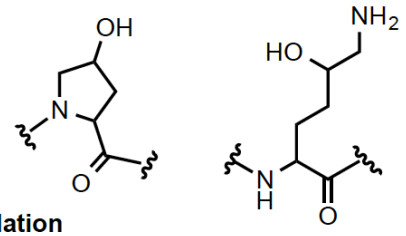
Vitamin K dependent process; Warfarin inhibits turnover of Vitamin K by epoxide reductase and prevents clotting

Hydroxylation

-Pro, Lys

-Proline hydroxylation is important in transcriptional control and protein structure.

-Hydroxylation and subsequent crosslinking of lysine residues in collagen cause conformational restriction and stabilize the coil-coil structure.



Thiol oxidation

-caused by reactive oxygen species

-unclear whether this has natural regulatory activity

Sulfatace

Prenylace

Myristoylace

SUMOylace

....

--- více Vás navede domácí úkol---