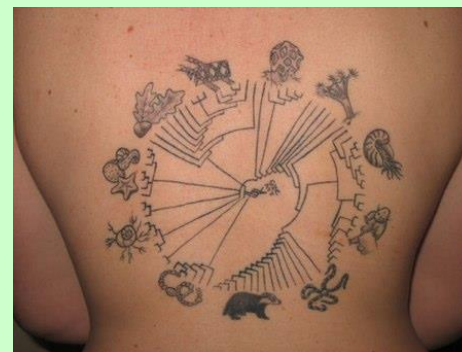
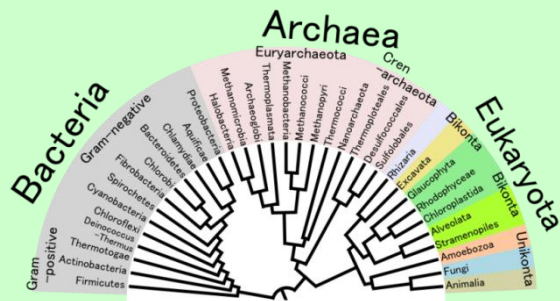
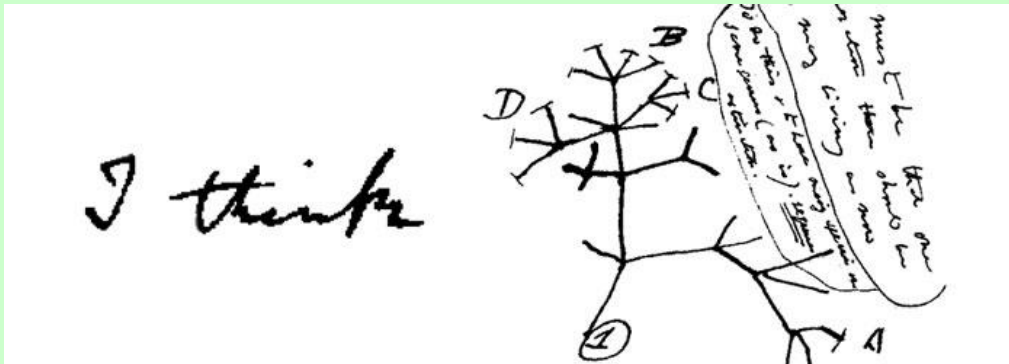




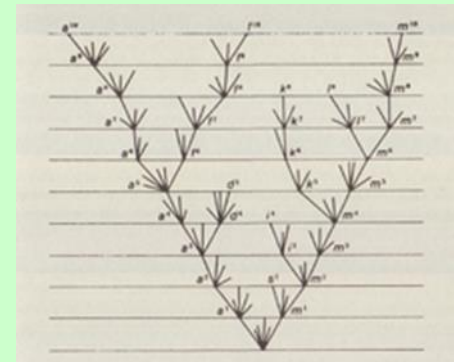
Iva Buriánková
Katedra ekologie
PřF UP

Kde „vyrostl“ první fylogenetický strom?

- Charles Darwin (1809 – 1882) – na základě fosilního záznamu



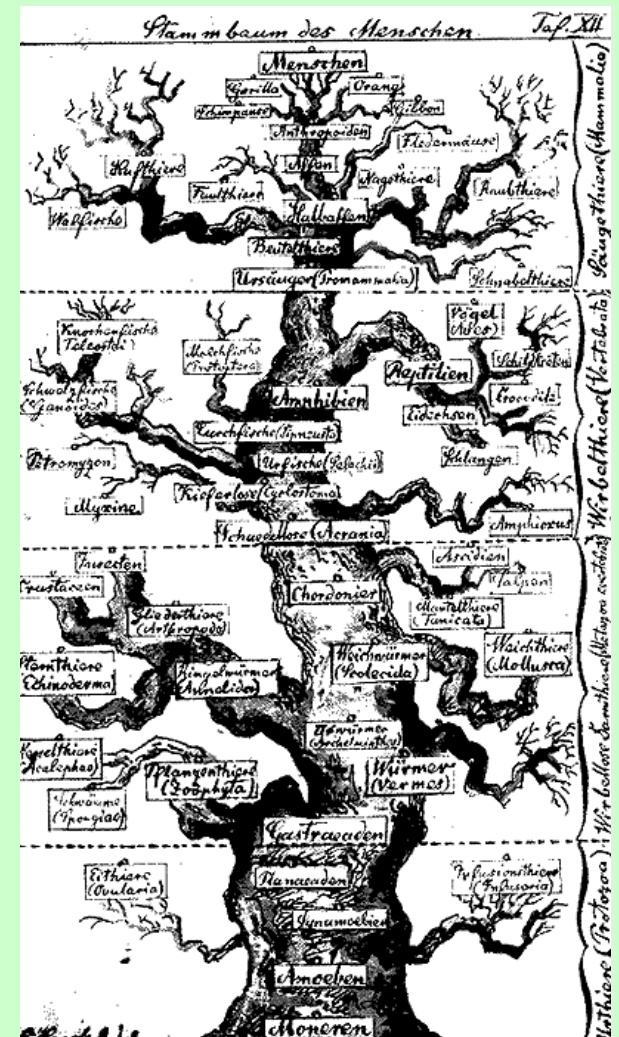
Darwinův deník – 1837, <http://www.icr.org/article/darwins-evolutionary-tree-annihilated/>



- Schéma vysvětlující základní princip biologické evoluce z Darwinovy knihy O původu druhů...
- Evoluce (přesněji biologická evoluce) je dlouhodobý a samovolný proces, v jehož průběhu se rozvíjí a diverzifikuje pozemský život...

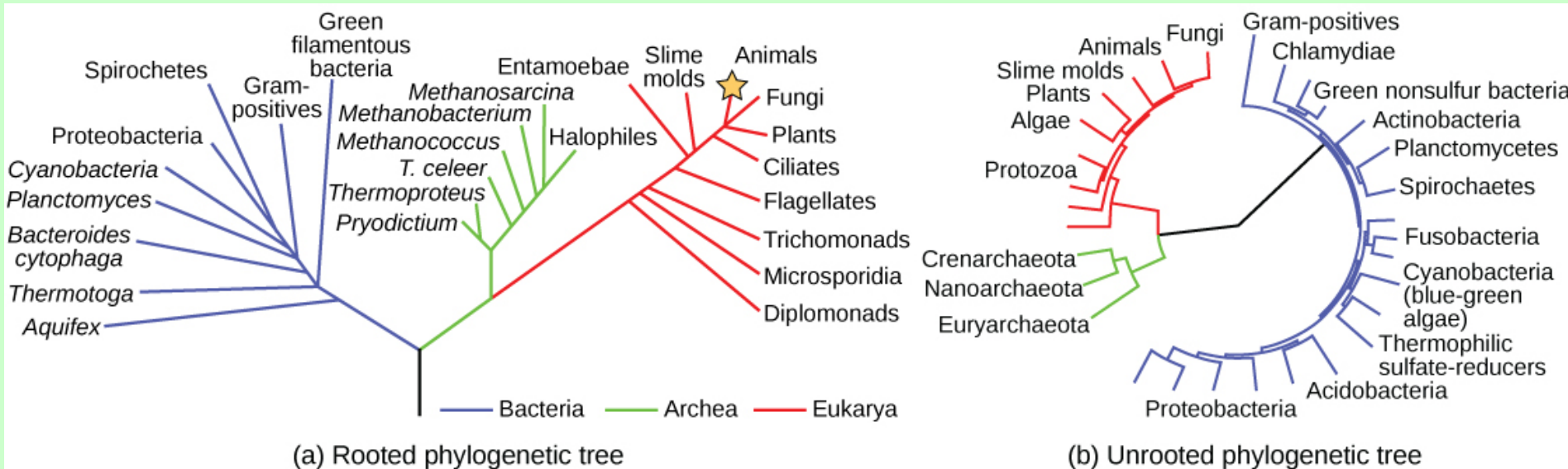
Fylogenetický strom

- grafické znázornění příbuzenských vztahů mezi různými taxonomickými jednotkami, o nichž lze předpokládat, že mají společného předka
- Příbuzenské vztahy se zde posuzují na základě morfologické či genetické podobnosti
- Místo taxonomických jednotek mohou v některých stromech vystupovat přímo jednotlivé biologické druhy nebo i jednotlivé geny
- V závislosti na typu fylogenetického stromu může délka hrany udávat dobu vývoje nebo míru podobnosti mezi příslušnými taxonomickými jednotkami
- Při tvorbě fylogenetických stromů se vychází z údajů o podobnosti mezi jednotlivými taxonomickými jednotkami
- Vychází ze sekvencí bází v genomech jednotlivých biologických druhů, případně lze použít i informace o příslušných aminokyselinových a proteinových produktech
- Na základě těchto dat je možné určit genetické vzdálenosti mezi jednotlivými dvojicemi taxonomických jednotek

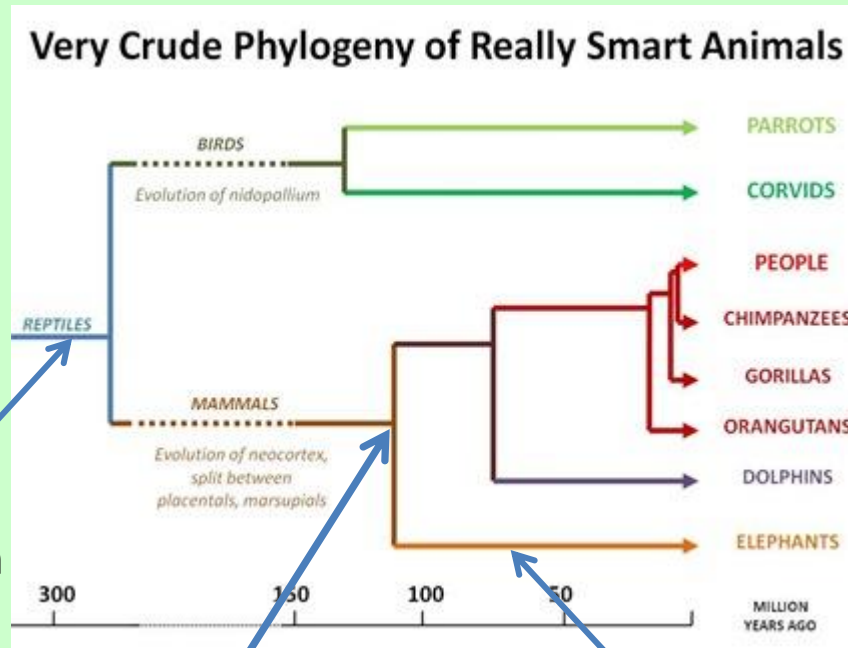


- **Nezakořeněný strom** - znázorňuje vztahy mezi taxonomickými jednotkami, aniž by specifikoval jejich společného předka...
- **Zakořeněný fylogenetický strom** je strom, u kterého byl jeden z vnitřních vrcholů označen jako *kořen* - reprezentuje společného předchůdce všech taxonomických jednotek znázorněných stromem...

Ze zakořeněného stromu je možné kdykoliv zkonstruovat nezakořeněný strom pouhým zrušením označení kořene, opačný postup je možný pouze s dodatečnými informacemi o průběhu evoluce...



Popis fylogenetického stromu



Clade – klad

Root – kořen

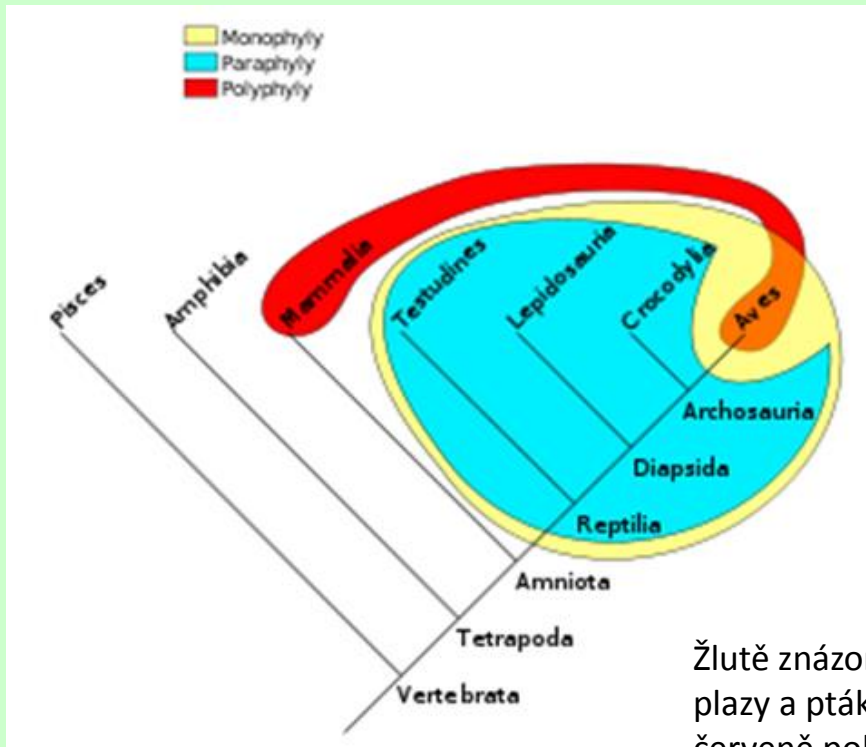
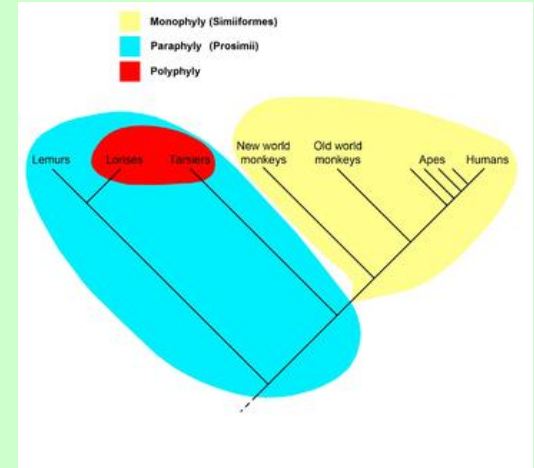
Node – uzel

Branch – větev

http://watchingtheworldwakeup.blogspot.cz/2009_11_01_archive.html

Vztahy kladů

- taxony monofyletické s jediným společným předkem a zahrnující všechny jeho potomky, základním požadavkem na přirozené biologické taxony
- polyfyletické bez blízkého společného předka
- parafyletické, nezahrnující všechny potomky společného předka



Žlutě znázorněna monofyletická skupina, zahrnující plazy a ptáky, tyrkysově parafyletická třída plazů a červeně polyfyletická skupina teplokrevných, zahrnující mimo jiné i ptáky.

Fundamentální podstata fylogenetiky

Bodové mutace – tranzice ,změna purinu na purin nebo pyrimidinu na pyrimidin (C → T, T → C, A → G, G → A)

- transverze, změna purinu na pyrimidin nebo naopak (A → T, T → A, C → A, A → C, G → T, T → G, G → C, C → G)
- delece (ztráta nukleotidu)
- inserce (vlození nukleotidu)
- ...

Teorie neutrální evoluce – Kimura (1968)

- evoluce většiny genů je zapříčiněna fixací neutrálních mutací (Kimura, 1983). Adaptivní mutace, přestože se vyskytují s velice malou četností, se mohou na evoluci genových sekvencí rovněž výraznou měrou podílet, protože jsou fixovány nejen působením náhodného genetického driftu, ale i přímé pozitivní selekce na úrovni proteinů. (Endo *et al.*, 1996)

Předpoklad homologie genů

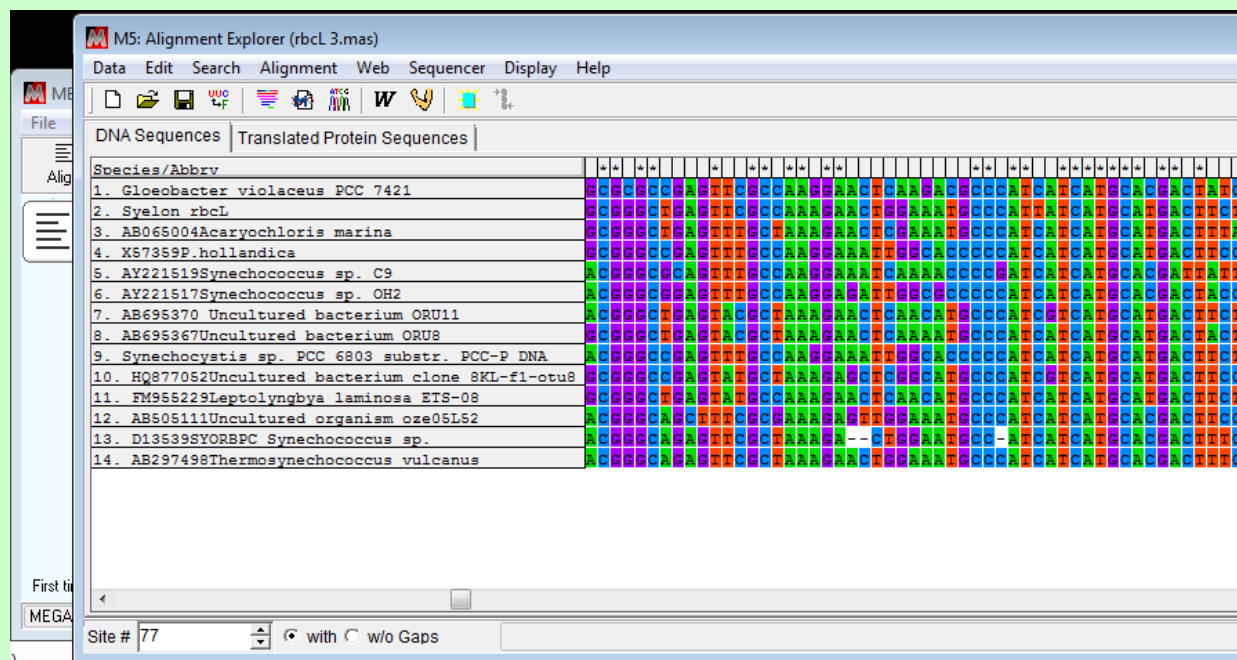
- Historická homologie– ortologie kdy se stejný znak u několika druhů odvozuje ze společného předka
- x biologická hom. paralogie, kdy existují dva homologické znaky na jednom jedinci: příkladem jsou hmyzí tykadlo a hmyzí noha, které se vyvinuly ze stejného základu
- homoplasie – konvergentní evoluce
- podobnosti v sekvencích či tvarech nebo v jejich životní historii, což ale *nelze* vysvětlit z jejich rodokmenu - podobnosti v sekvenci či struktuře, kterou naprosto nelze označit za důsledek původu ze společného předka
- podobnosti, které jdou napříč vývojovými stromy života



Multiple Sequence alignment- uspořádání do sloupců

- Přesný výpočet vzdálenosti mezi dvojicemi tax.jednotek vyžaduje nejprve vhodné zarovnání porovnávaných DNA sekvencí – tzv. aligning
- výpočetně velmi obtížnou úlohu v praxi se proto používá celá řada heuristických metod, které jsou schopny nalézt alespoň suboptimální řešení v přijatelném čase. ..
- U zarovnaných sekvencí je možné určit vzdálenost například na základě procenta odlišných bází mezi sekvencemi. ..
- Sofistikovanější metody se pokoušejí odhadnout počet mutací, které jsou zapotřebí pro přechod od jedné sekvence k druhé

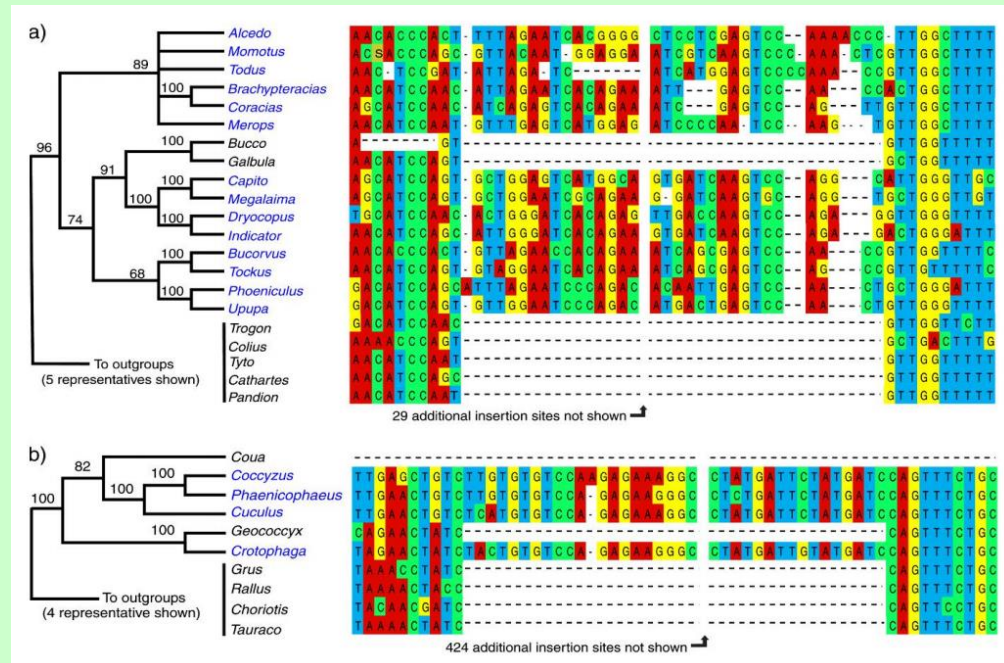
Pojem heuristický algoritmus se obvykle používá pro algoritmy, které neposkytují (matematické) záruky kvality řešení, případně kdy nevíme, zda heuristika uspěje.



Multiple Sequence alignment algorithmy

- **Progressive alignment construction – ClustalW**
 - Fylogenetický strom pro přesnější vyhledání podobných sekvencí
- Iterative method – Muscle
 - Podobné, ale je možné se vrátit zpět a alignment score vylepšit
- HMM – Hidden Markov Chain
 - Pravděpodobnostní metoda, přiřazování mezerám a kombinacím bází pravděpodobnosti

...



Metody konstrukce fylogenetických stromů

- Tyto metody vycházejí z matice distancí, která udává vzájemné vzdálenosti mezi všemi dvojicemi taxonomických jednotek, pro které konstruuujeme fylogenetický strom
- Jako vzdálenost se v tomto případě používá genetickou vzdálenost...

UPGMA

Neighbor joining

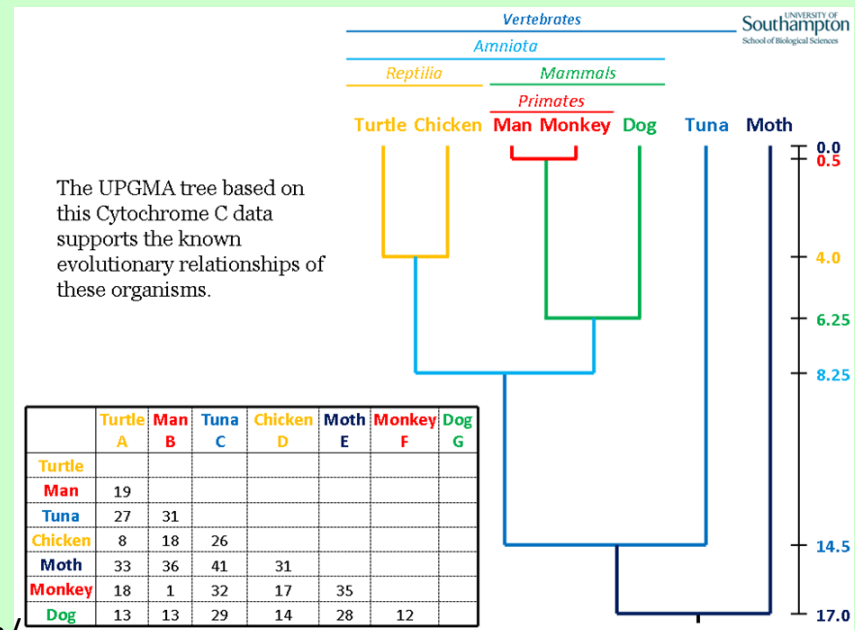
Maximum parsimony

Maximum likelihood

Molekulární hodiny

UPGMA (Unweighted Pair Group Method with Arithmetic mean)

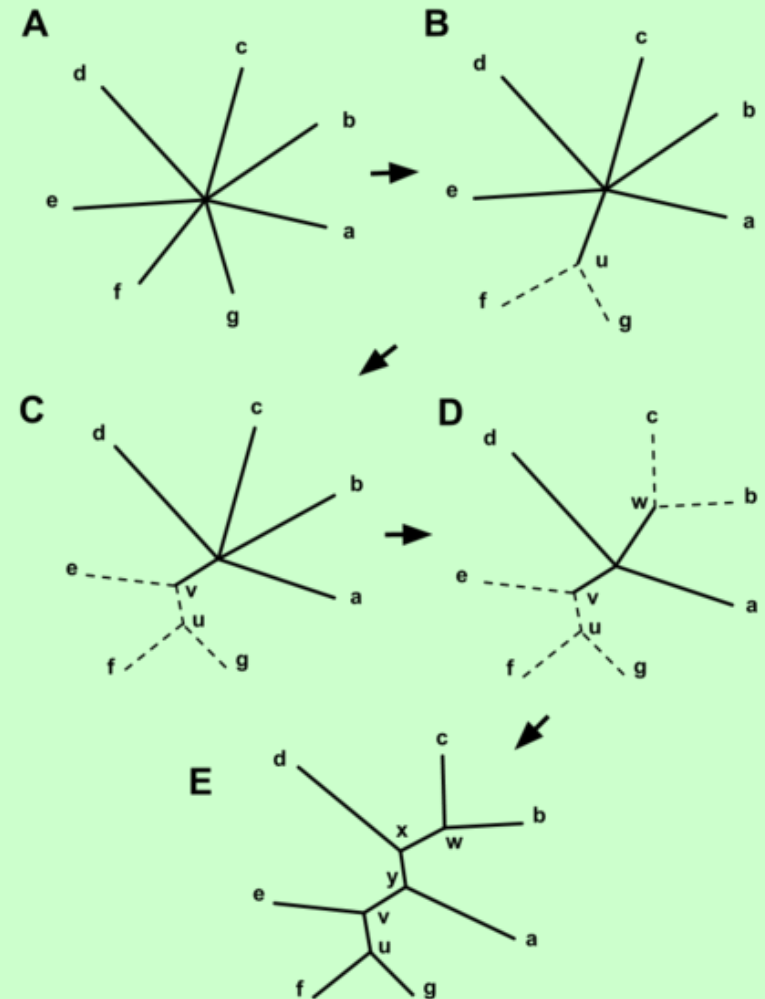
- UPGMA, zjednodušeně Shlukovací analýza, je nejjednodušší algoritmickou metodou konstrukce fylogenetického stromu. Postup je následující:
- Nalézt v distanční matici nejmenší hodnotu (odpovídá dvojici taxonomických jednotek, které mají k sobě nejbližší)
- Příslušné taxonomické jednotky sloučit do jedné skupiny a spočítat vzdálenost této nové skupiny ke všem ostatním taxonomickým jednotkám
- Vzdálenost taxonomické jednotky T k této nové skupině S se spočítá jako aritmetický průměr vzdáleností mezi jednotkou T a všemi prvky skupiny S
- Skupinu S lze dále považovat za hypotetickou taxonomickou jednotku
- Pokud máme k dispozici více než jednu taxonomickou jednotku, opakovat postup od 1. kroku
- Znázorníme-li graficky postup shlukování v průběhu popsaného algoritmu, získáme požadovaný fylogenetický strom
- Hypotetická taxonomická jednotka, která vznikla jako poslední, je jeho kořenem



Neighbor joining

Neighbor-joining

- Na začátku se vytvoří jeden hvězdicový strom, kde je jeden vnitřní vrchol, a všechny řešené taxonomické jednotky jsou reprezentovány pomocí listů
- Tento strom se postupně rozkládá shlukováním nejbližších taxonomických jednotek tak, aby se v každém kroku co možná nejvíce zmenšila celková délka stromu.
- Délka větví má informativní charakter
- „star like“ iniciální strom - nody se sbíhají na základě podobnosti

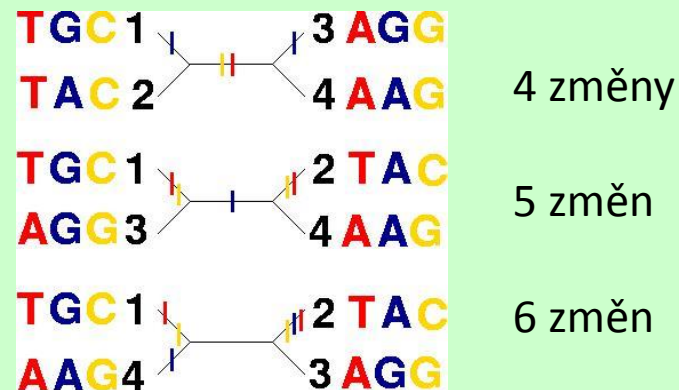


Maximum parsimony

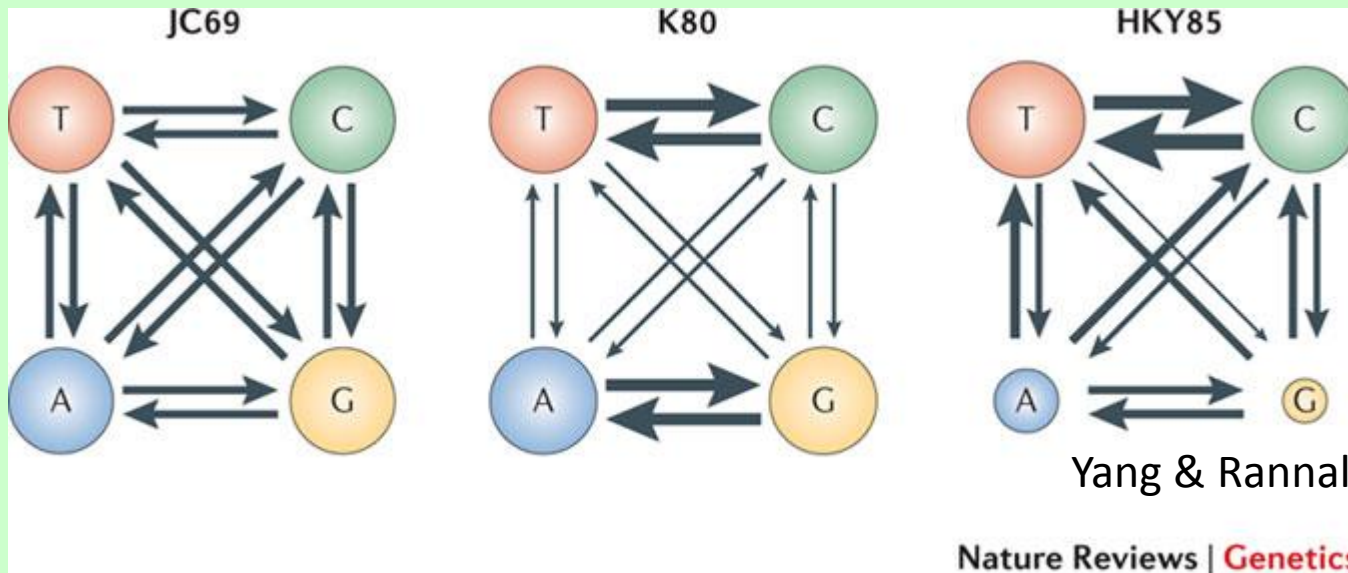
- Nejmenší počet změn nejlépe vysvětluje data
 - (Ockhamova břitva, 14. století) *Pluralitas non est ponenda sine necessitate*. tj. *Množství (tj. důvodů, příčin) se nemá dokládat, není-li to nezbytné.*
 - *Pokud pro nějaký jev existuje vícero vysvětlení, je lépe upřednostňovat to nejméně komplikované..*
 - *Pokud nějaká část teorie není pro dosažení výsledků nezbytná, do teorie nepatří.*
- Vytvoření stromů a vypočítání skóre – jak moc je strom parsimonní, vybráno maximálně parsimonní strom – nejlepší skóre
- Analyzovány jsou jednotlivé změny v alignmentu NE distanční matrix

Maximální parsimonie

Metoda maximální parsimonie se snaží nalézt takový fylogenetický strom, který vyžaduje co nejmenší množství evolučních událostí, ke kterým by muselo dojít, pokud by tento strom odpovídal průběhu evoluce. V některých případech se při hodnocení stromů přiřazuje jednotlivým evolučním událostem různá váha, například je-li známo, že některé nukleotidy či aminokyseliny mutují snáze či hůře než ostatní.



Substituční modely



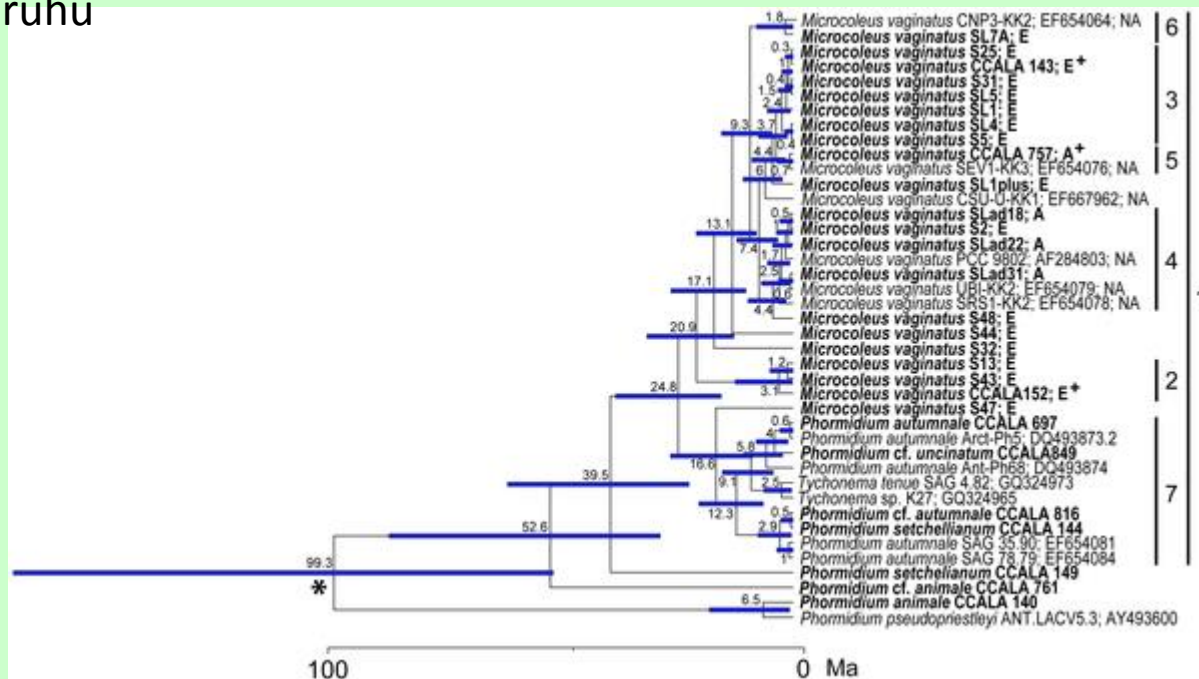
Síla šipky označuje významnost substituce

Maximum likelihood

- vychází ze statistických metod a aposteriorní pravděpodobnosti
 - Snaží se odhadnout, jaká je pravděpodobnost, že platí statistická hypotéza představovaná konkrétním fylogenetickým stromem pro data, která máme k dispozici.
 - Metoda vyžaduje substituční model, na základě kterého určíme pravděpodobnost jednotlivých evolučních změn (mutací).
 - Strom, který pro vysvětlení dostupných fylogenetických dat potřebuje těchto změn více, bude mít menší věrohodnost než strom, který si vystačí s menším počtem změn
 - Mimo toho si všímáme i délek jednotlivých větví.
-
- Každé místo v alignmentu – určitá pravděpodobnost, podle modelu
 - Totální pravděpodobnost je produktem místních
 - Výsledný strom je takový, který má největší pravděpodobnost k danému substitučnímu modelu
 - V průběhu analýzy převrácení větví stromu a hledání maximum likelihood – swapovací algoritmus

Molekulární hodiny

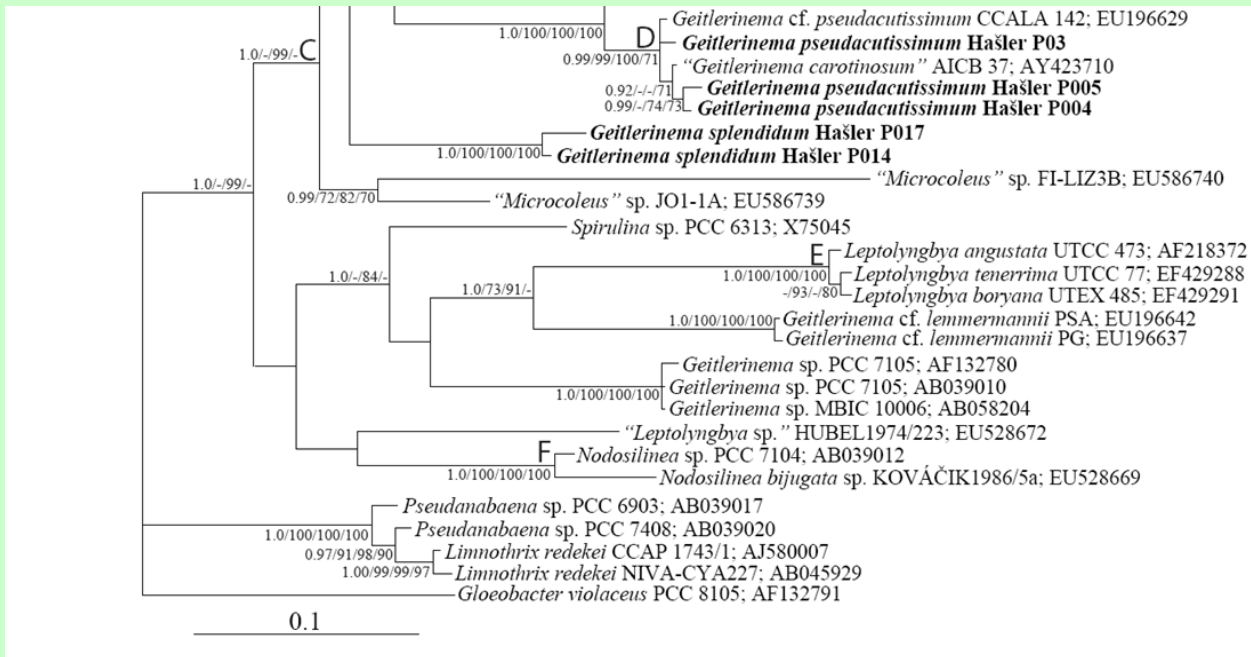
- Molecular clock hypothesis – Zuckerkandl, Pauling (1962) a další
- Mutace v DNA probíhají konstantně v čase
- Předpoklad neutrální evoluce – Kimura (1968)
- Počet mutací/časová jednotka, konstrukce fylogenetického stromu, odhad divergence druhů

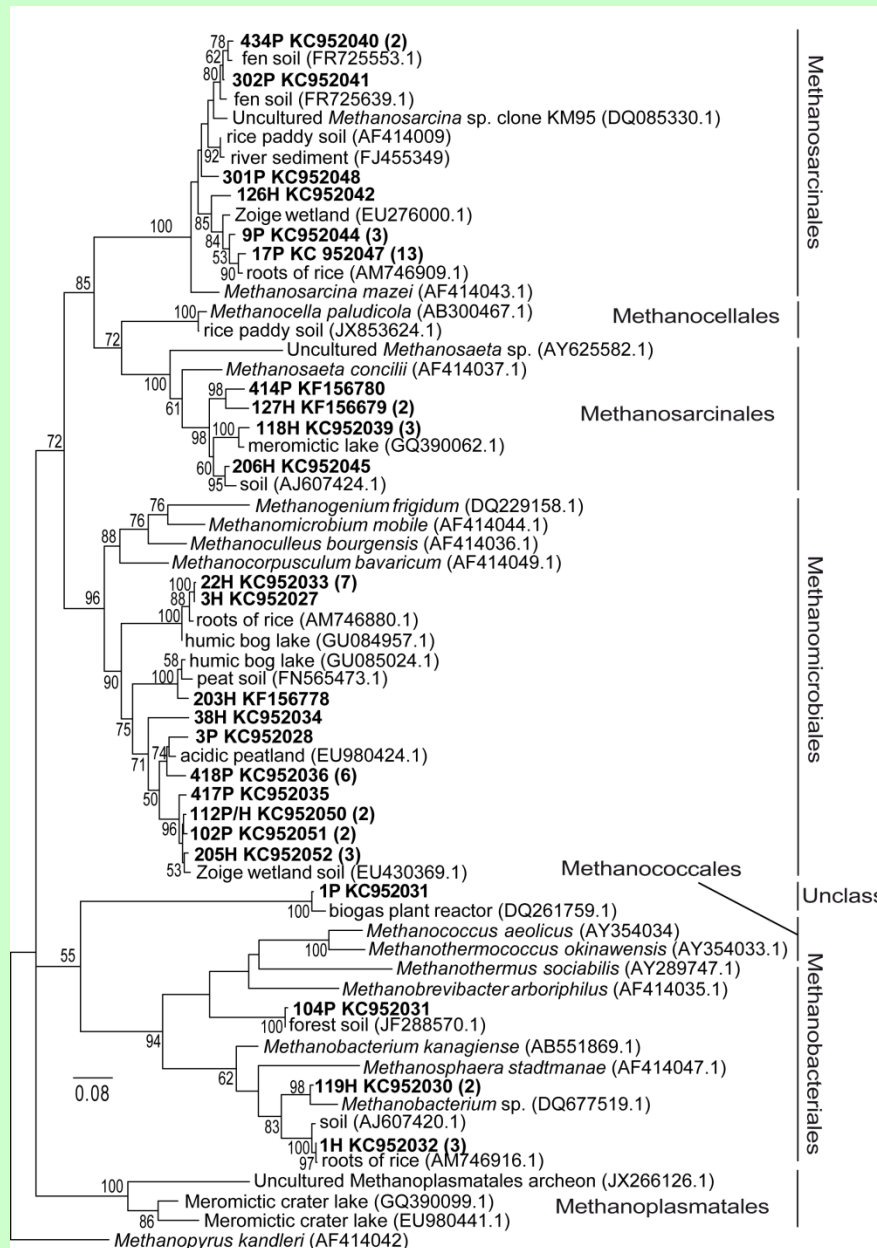


Dvořák et al. 2012, 16S rRNA

Statistické testy

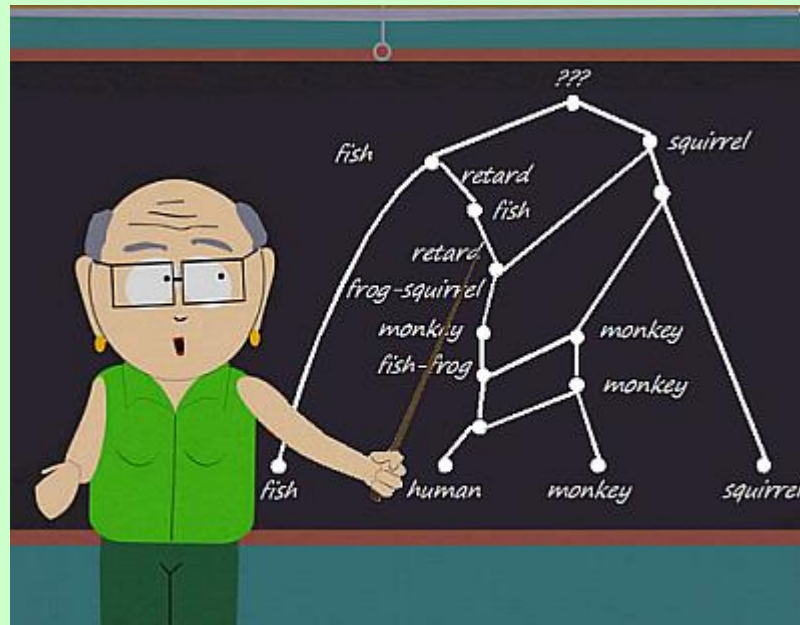
- Bootstrapping
 - Jednotlivé sloupce alignmentu jsou „resamplovány“ (počet opakování, obvykle více než 100) náhodně – některé mohou být vynechány
 - Konstrukce stromů
 - Zastoupení nodů při jednotlivých opakování – bootstrapp (0-100) významné – nad 50 (70)
 - Nejběžnější
- Jackknifing -probíhá analýza postupně jednotlivých částí datasetu bez vynechání



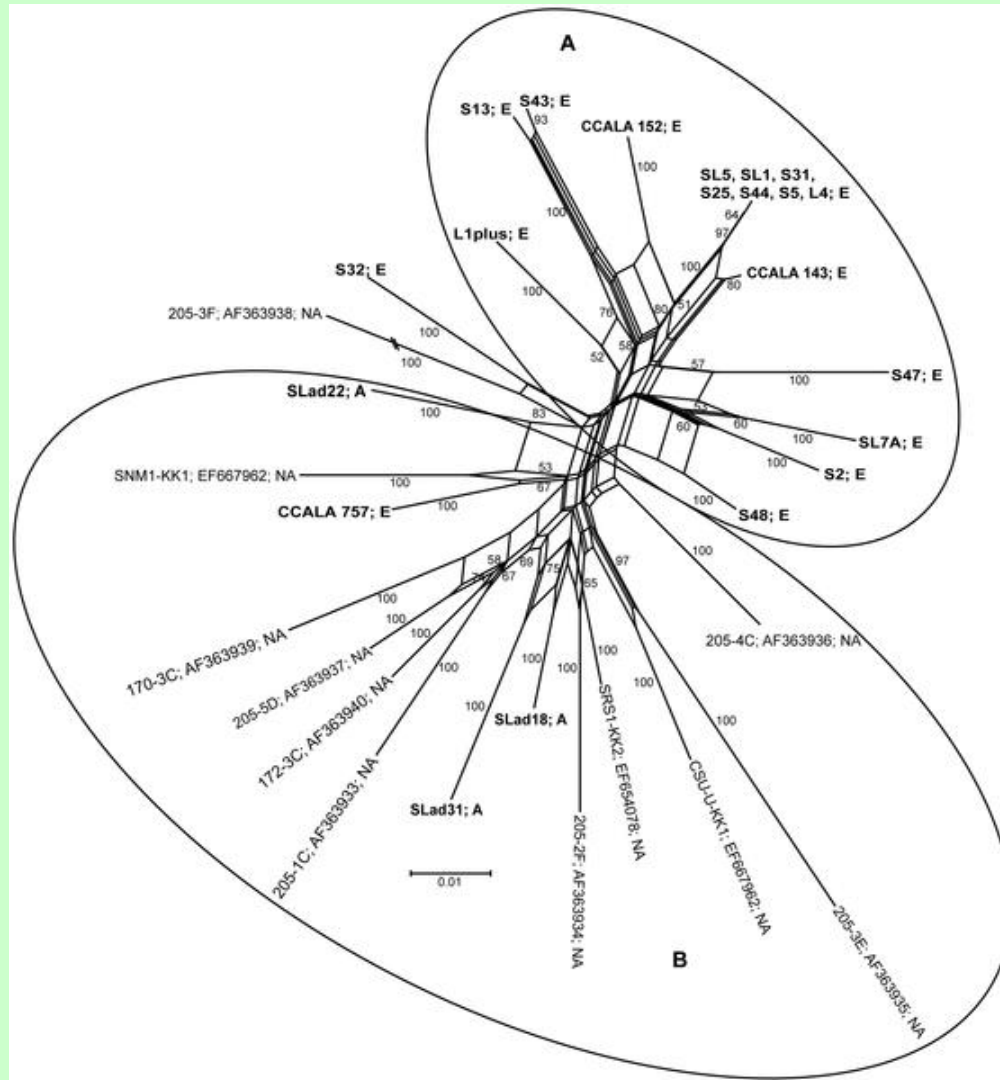


Fylogenetická síť

- Pro rekonstrukci evoluce bez bifurkace, kterou předpokládá fylogenetický strom
- Rekombinace, hybridizace, transfer genů
- Hlavně – viry, bakterie, některé rostliny

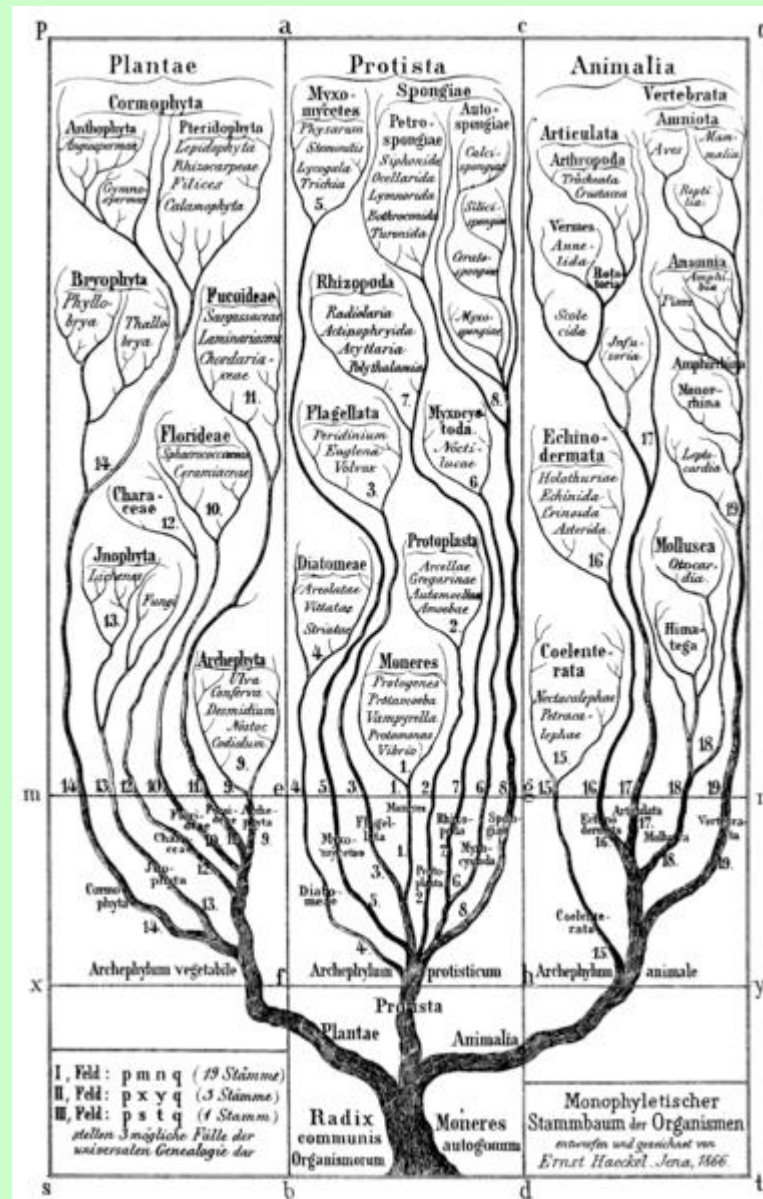


Fylogenetická síť

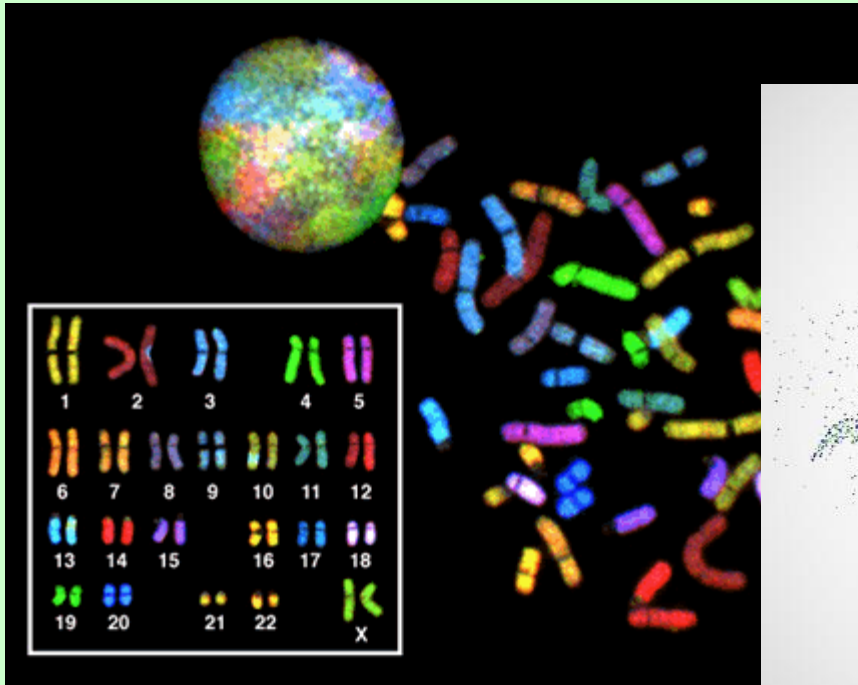


Dvořák et al. (2012)
16S-23S ITS

DĚKUJI ZA POZORNOST



Lidský genom



Lidský genom

- souhrn veškeré genetické informace zapsané v DNA uvnitř lidských buněk
- Homo sapiens sapiens - DNA uvnitř buněčného jádra a dále ovšem i v mitochondriích (tzv. mitochondriální DNA)
- 1990 spuštěn Human Genome Project, kladoucí si za cíl přečíst (osekvenovat) celou jadernou DNA člověka; tento projekt byl úspěšně završen v 2003...
- Do současnosti však již bylo osekvenováno několik genomů různých lidí, včetně biochemika Jamese Watsona.....
- Moderní metody sekvenace ovšem cenu dramaticky snížily a v současnosti je známa sekvence více než tisíce lidských genomů
- dříve byl počet lidských genů odhadován na 100 000, v současné době jsou kvalifikované odhady pětkrát nižší

- Lidská jaderná DNA se skládá v haploidním stavu z asi 3,1–3,2 miliardy párů bází
- kdyby se seřadila jednotlivá vlákna za sebe, byla by vzniklá řada dlouhá asi jeden metr
- jaderná DNA obsahuje 20–25 tisíc genů, což je mimochodem počet genů srovnatelný například s hlísticí *Caenorhabditis elegans*
- počet nefunkčních genů (pseudogenů) představuje však dalších asi 20 000
- záhadné geny kódující různé druhy malých RNA – např. bylo odhaleno 800 genů pro microRNA
- Lidé (*Homo*) se od evoluční linie šimpanzů (*Pan*), svých nejbližších příbuzných, oddělili před 5 až 7 miliony lety- hlavní změnou je splynutí dvou chromozomů, redukce z 48 na 46 -vznikl lidský chromozom 2
- genom šimpanze a člověka -udává se, že je genetická informace z 95–99% zcela identická
- Za některé rozdíly mezi genomem člověka a šimpanze jsou zodpovědné převážně genové duplikace a v menší míře i substituce v rámci jednoho nukleotidu (SNP).

- Soukromá společnost Celera použila k rychlému čtení lidského genomu velmi odvážný postup, při kterém je celá lidská DNA „rozsekána“ na malé kousky
- po přečtení velmi výkonnými počítači se speciálními programy kompletovány do souvislé řady písmen genetického kódu
- Mezinárodní konsorcium vyšlo z pomalejší ale důkladnější metody čtení, při níž se v lidské DNA nejprve určí pevně „orientační body“ a podle nich se pak jednotlivé přečtené úseky dávají dohromady
- První pohled do „vyspraveného“ lidského genomu odhalil 1183 genů, které vznikly relativně nedávno tím, že si staré geny našich předků vytvořily kopii a ty postupem času začaly sloužit novému účelu
- Podobnému „dublování“ podlehl v evolučně nedávných dobách asi 5% lidské DNA
- Genetici narazili také na 37 „čerstvých hrobů“ – tedy míst, kde naši předci měli funkční geny, ale ty utrpěly těžká poškození a v naší dědičné informaci už nefungují

- Když čtení lidské dědičné informace startovalo, věřili vědci v existenci až 100 tisíc lidských genů. Většinou očekávali 50 tisíc genů (sázky)
- Z „hrubého nárysu“ ale vyplývalo, že člověka dělá člověkem něco kolem 35 tisíc genů
- Poslední rozbor „vyspravené“ verze - pouhých 20 až 25 tisíc genů, kukuřice může mít kolem 50 tisíc genů
- Tajemství lidské výjimečnosti zřejmě tkví v organizaci a souhře našich genů, ale zatím zůstávají z drtivé většiny nepoznány
- dnes sice známe některé geny, které našim předkům zcela jistě dopomohly k významným evolučním úspěchům, ale ani všechny dohromady by zjevně nestačily k proměně šimpanze v člověka. ...
- současné době jsou „slušně rozečtení“ myš, potkan, šimpanz, makak rhesus, skot, pes, vačice a ptakopysk....rostliny, bakterie....
- V dohledné době - orangutan, gibbon, jihoamerická opička kosman a drobný madagaskarský lemur maki trpasličí. ...
- Ve frontě na „přečtení“ své DNA rejsek, ježek, kočka, slon, pásovec, madagaskarský hmyzožravec bodlín či morče, jež možná nahradí některý z četných netopýrů. ...

DĚKUJI ZA POZORNOST

