Počítačová část

1. seznámení s on-line databázemi, nástroji a softwarem (databáze, vyhledání sekvencí, základní manipulace se sekvencemi, navržení primerů)

Pavel Munclinger, Petr Synek

2. fylogenetická analýza

Zuzana Starostová, Darina Koubínová, Zuzana Musilová

začátky

chci studovat populační strukturu, příbuzenské vztahy uvnitř nějaké skupiny organismů, …

- podívat se na NCBI a do databází článků (WOS, PubMed) co se o daném ogranismu/ skupině ví
- jaké geny byly použity v předchozích studiích?
- stačí použité geny na rozřešení fylogenetických vztahů na námi zvolené taxonomické úrovni?

Vyhledávání sekvencí v genové bance (GenBank)

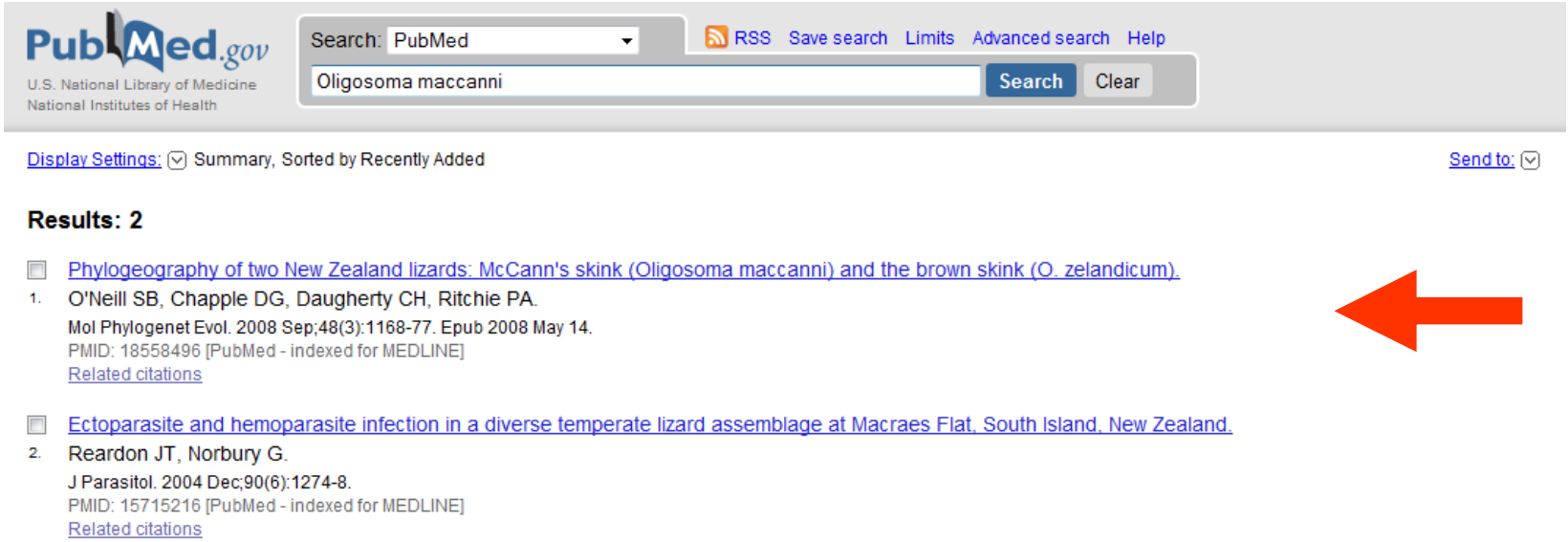
- jsou známy nějaké sekvence leguánů rodu *Cyclura*?

  Chci dělat fylogeografii tohoto druhu – jaké geny použili jiní badatelé? Stažení sekvencí se kterými pracovali přímo v určitém článku snadno a rychle …

přes GenBank na NCBI

Jak postupovat?
V NCBI http://www.ncbi.nlm.nih.gov/

Cyclura



Mol Phylogenet Evol. 2000 Nov;17(2):269-79.

Phylogeography of the Caribbean rock iguana (Cyclura): implications for conservation and insights on the biogeographic history of the West Indies.

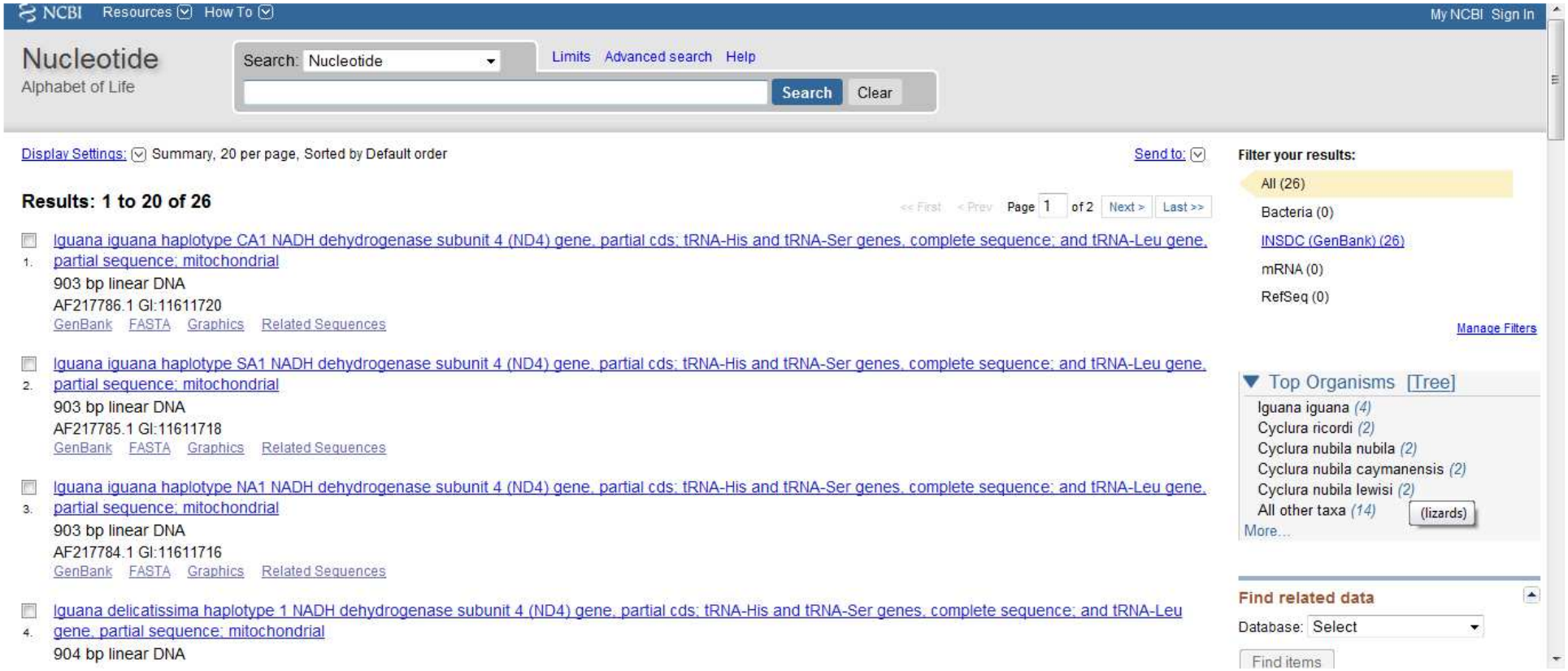
Malone CL, Wheeler T, Taylor JF, Davis SK.

Program in Genetics, Texas A&M University, MS 2471, College Station, Texas 77843, USA. clm0668@pop.tamu.edu

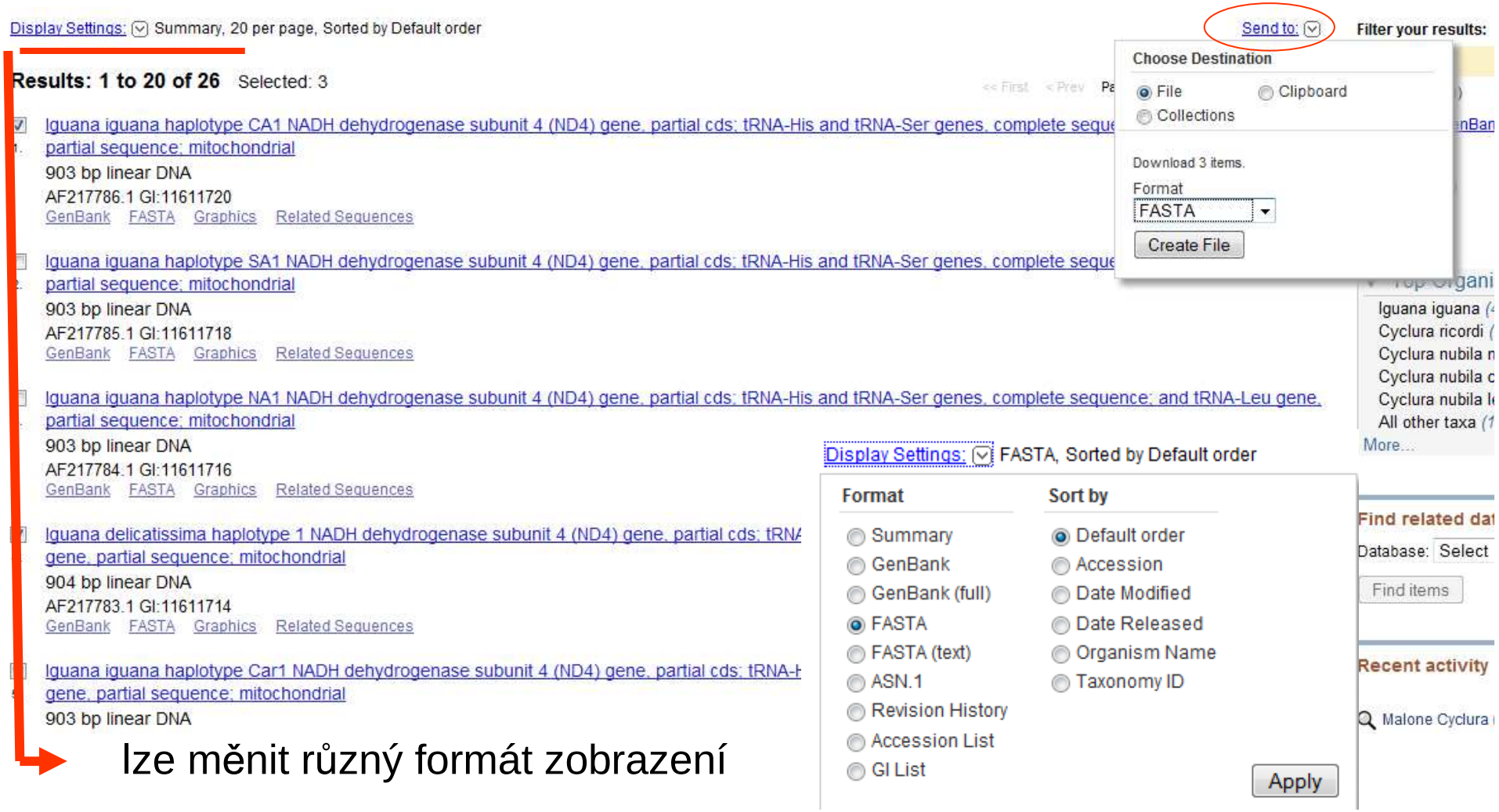
Abstract

The Caribbean rock iguana, Cyclura, has had an unstable intrageneric taxonomy and an unclear phylogenetic position within the family Iguanidae. We use mtDNA sequence data to address these issues and explore the phylogeographic history of the genus. ND4 to leucine tRNA sequence data were collected from multiple individuals of each of the eight species of Cyclura (including 15 of 16 subspecies) and from four localities of Iguana iguana (representative of this species' broad geographic range). This data set was combined with sequence data from Sites et al. (1996, Mol. Biol. Evol. 13, 1087-1105) and analyzed under maximum-parsimony and maximum-likelihood optimization criteria. The ND4 region provided good resolution for the majority of nodes, as indicated by high bootstrap support. In agreement with several recent molecular studies, Cyclura is recovered as monophyletic and is not closely related to any other genus, whereas Iguana is strongly supported as the sister taxon to Sauromalus. This result is statistically more likely than other published hypotheses of Iguanid relationships. Cyclura shows a southeast to northwest speciation sequence in the Caribbean, with the most ancient lineage on the Puerto Rican Bank. The amount of interspecific sequence divergence within Cyclura (maximum 11.4%) is very high in comparison to data from other iguanid taxa at this locus, suggesting that this group either has been in the Caribbean for a very long time or has gone through a very rapid rate of evolution at this locus. Using dates from other published studies, we calculate a molecular clock that suggests that Cyclura colonized the Caribbean between 15 and 35 mya. Several questions regarding subspecific taxonomy are raised in the analysis and await further investigation using a more rapidly evolving marker such as nuclear microsatellites.

PMID: 11083940 [PubMed - indexed for MEDLINE]

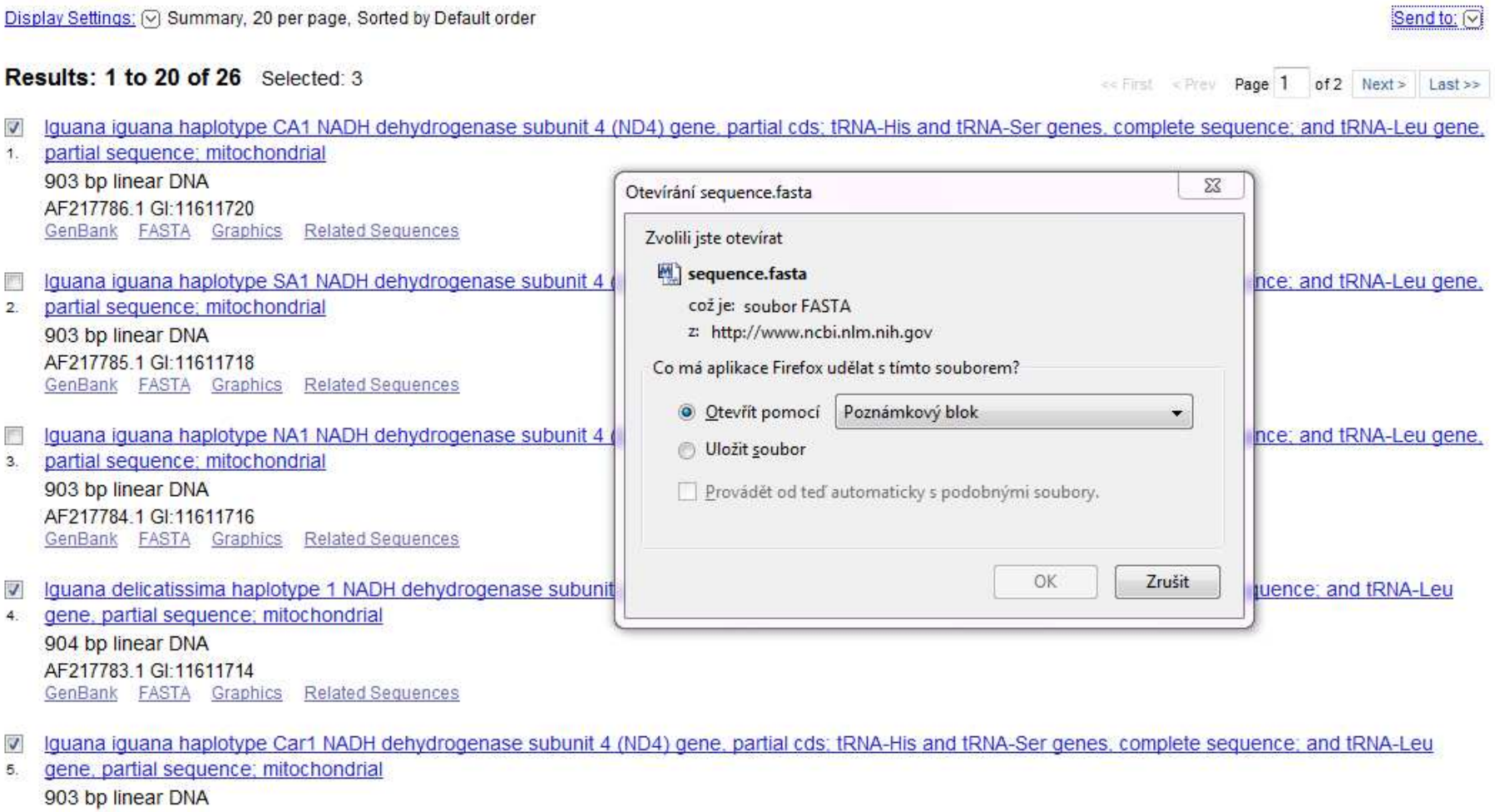


prohlédutí sekvencí, zatrhnutí těch, které chceme (sekvence leguánů rodu *Cyclura* – vždy jednu od druhu + jednu sekvenci *Iguana iguana*)

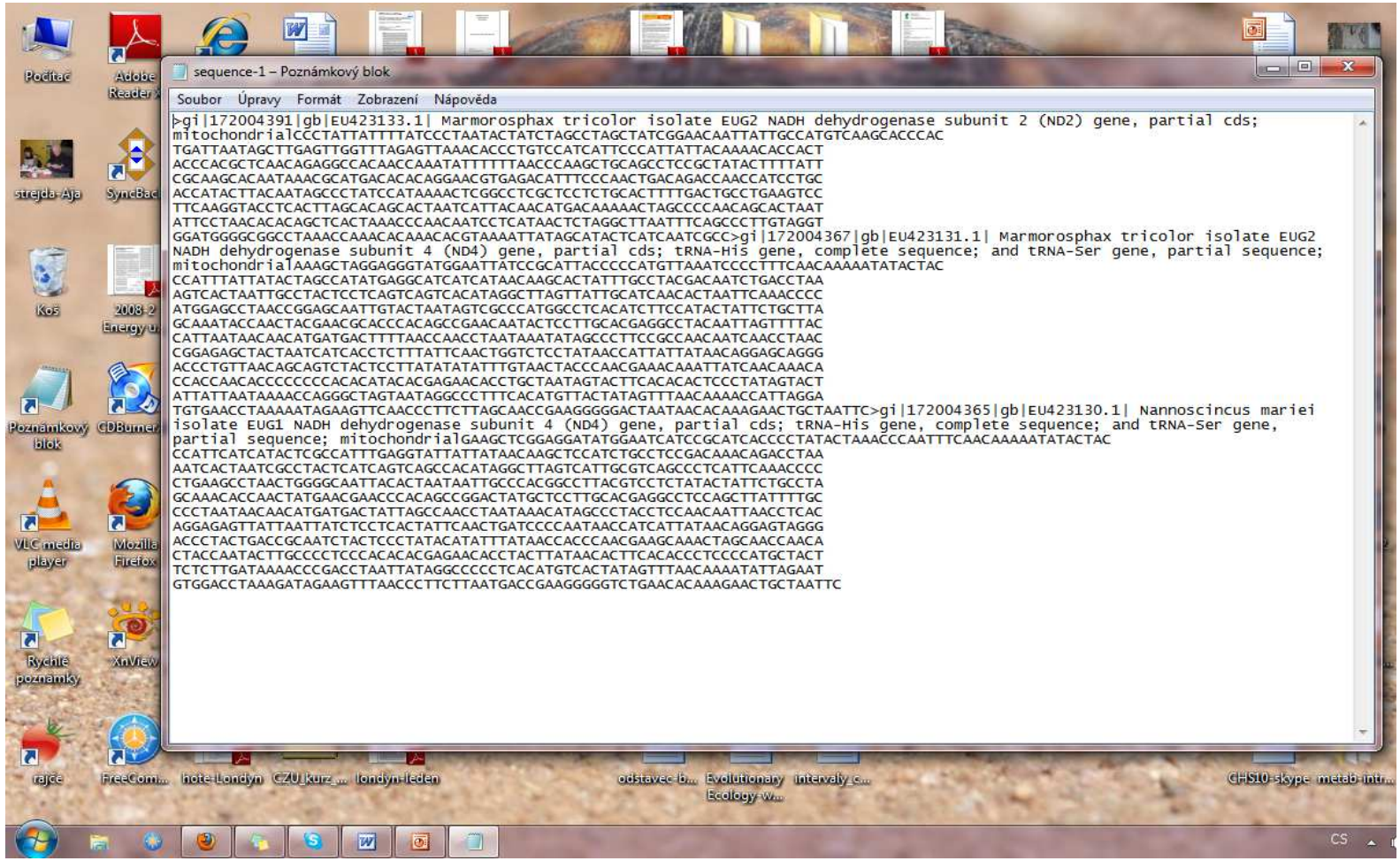


lze měnit různý formát zobrazení

Po vybrání sekvencí zvolíme nahoře nebo dole na stránce Send to zaškrtneme, že chceme sekvence uložit do souboru (File) a jako preferovaný formát sekvencí (FASTA)



Uložit/otevřít vybrané sekvence ve formátu FASTA v programu Poznámkový blok (lze i rovnou v BioEditu)



soubor se sekvencemi lze upravit (editovat názvy dle požadavku programů, atd.) – upravte název na jednoslovný

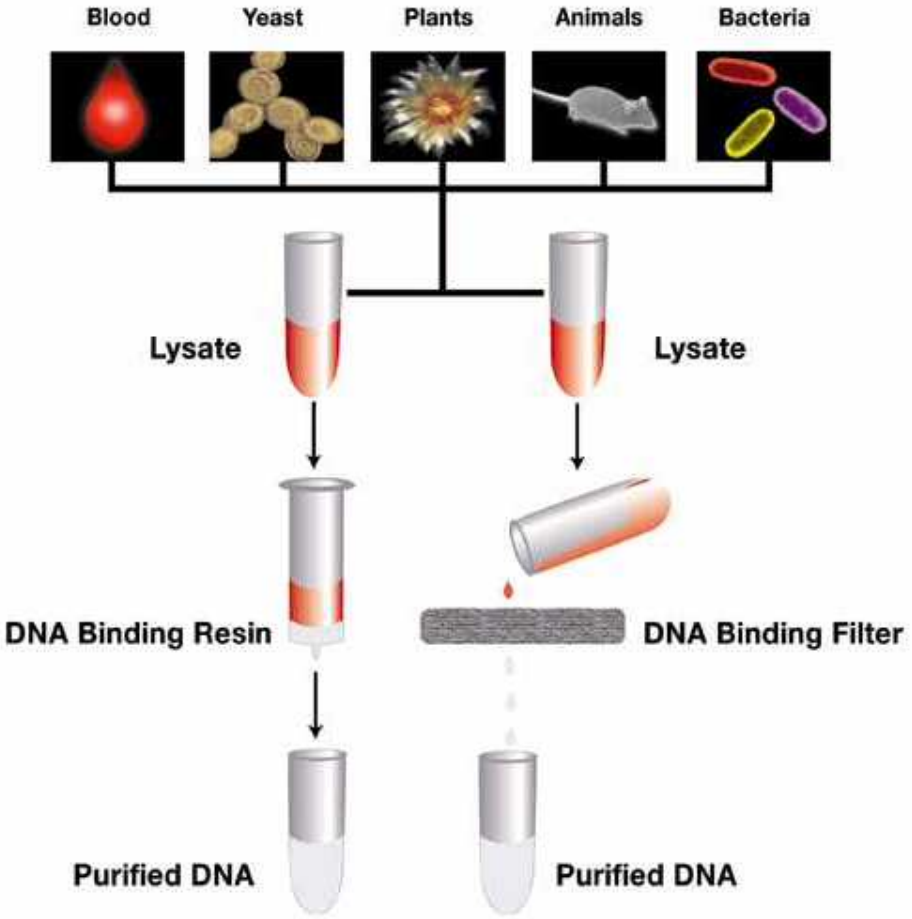
>iguana

CTACCTAAATGGCTAGCC

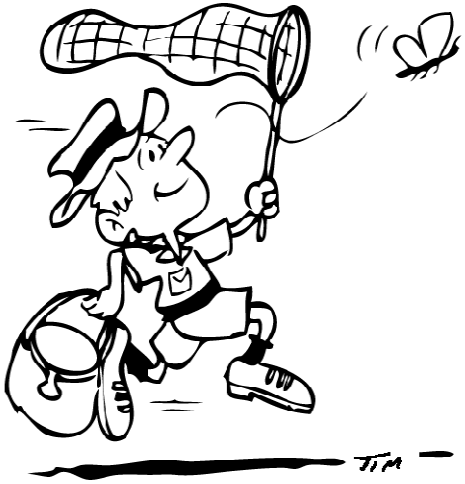
uložit do adresaře gen-metody-odpoledne/sekvence-praktika jako leguani.fas

pro naše analýzy můžeme skombinovat sekvence stažené z databází (GeneBank) se sekvencemi získanými z vzorků druhů, které chceme studovat

+



Zisk vlastních sekvenačních dat:

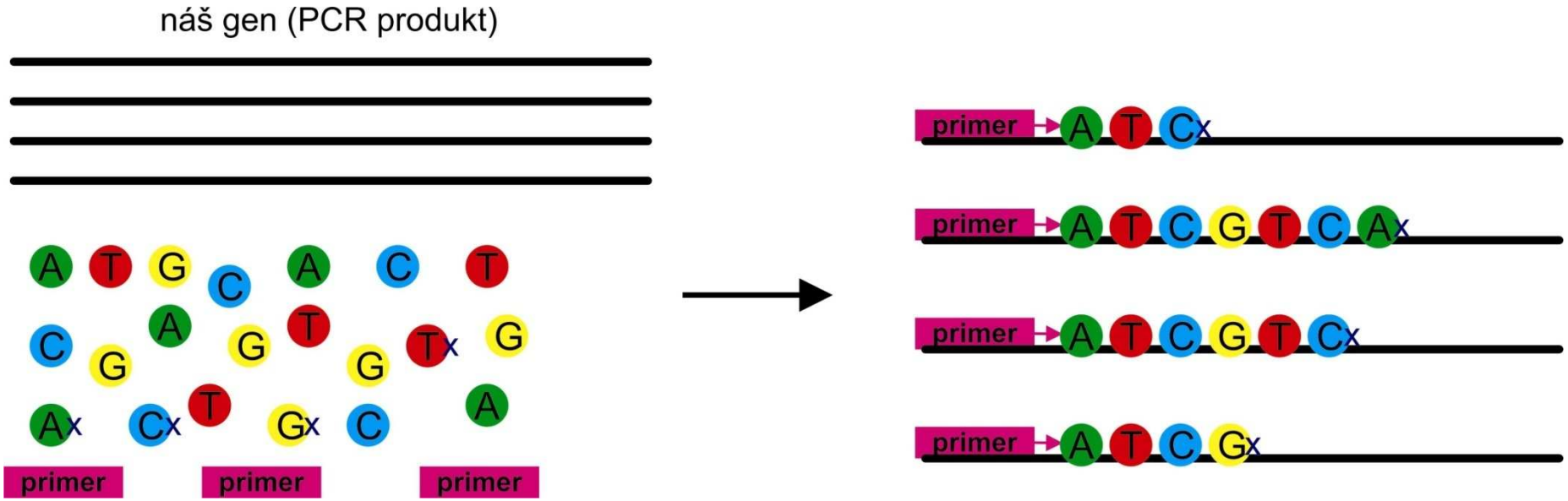


sběr vzorků, laboratorní zpracování (izolace DNA, PCR)

sekvenační reakce:

Sangerova metoda; Macrogen

speciální forma PCR s jedním primerem a fluorescenčně značenými dideoxynukleotidy (ddNTP, ddATP, ddGTP, ddCTP)



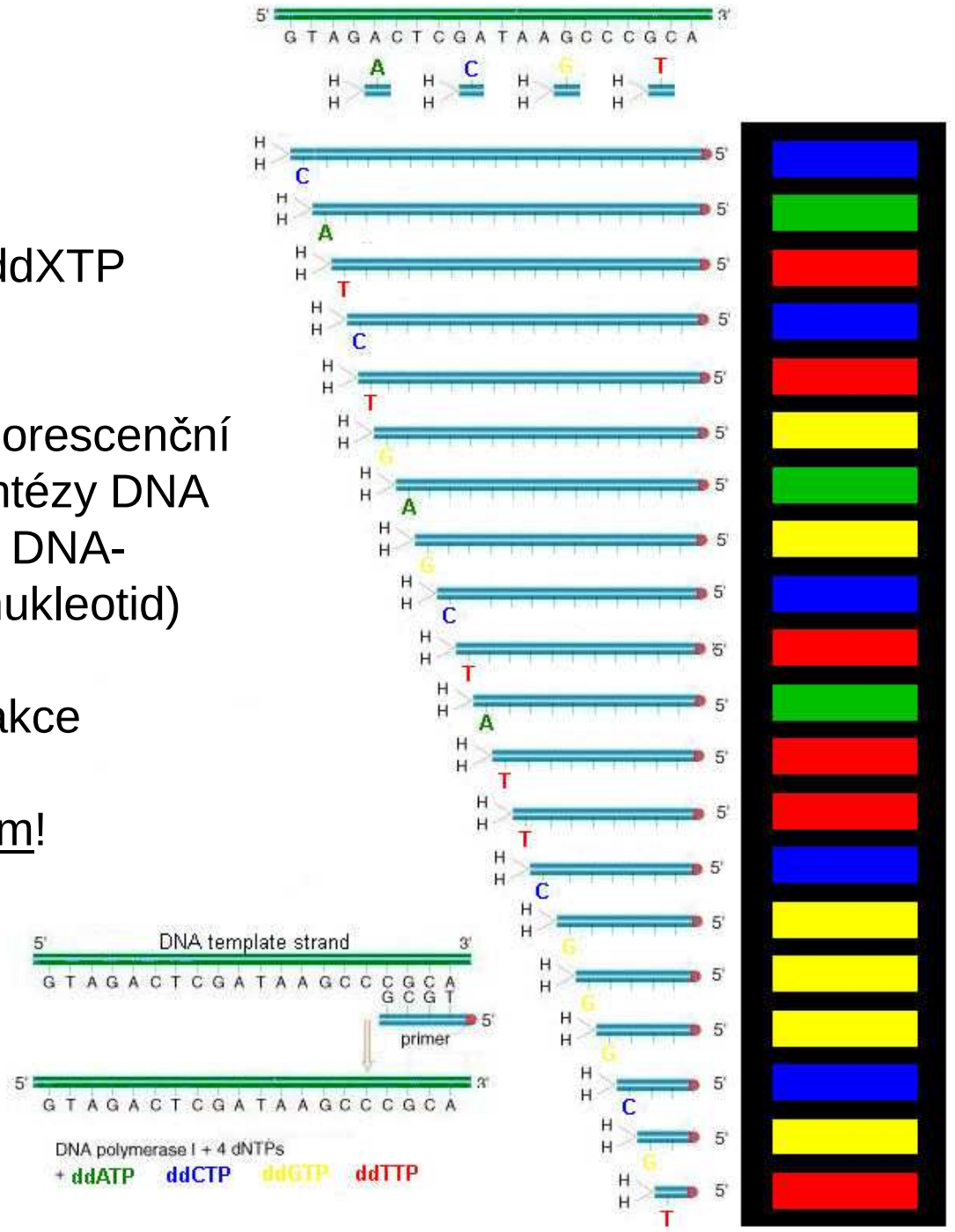
sekvenační reakce:

optimálně zvolený poměr dXTP a ddXTP zaručí produkty o všech délkách

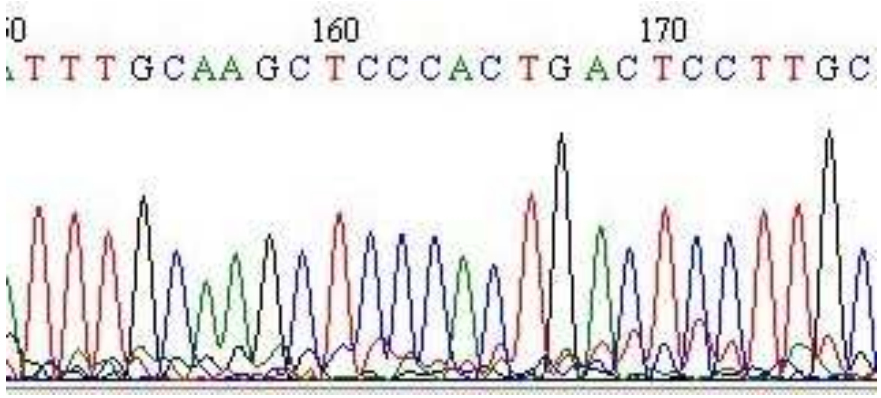
dideoxynukleotidy jsou značeny fluorescenční Barvou a slouží jako terminátor syntézy DNA (nemá na 3'-konci OH skupinu, tak DNA-polymeráza nemůže připojit další nukleotid)

přečištění produktu sekvenační reakce

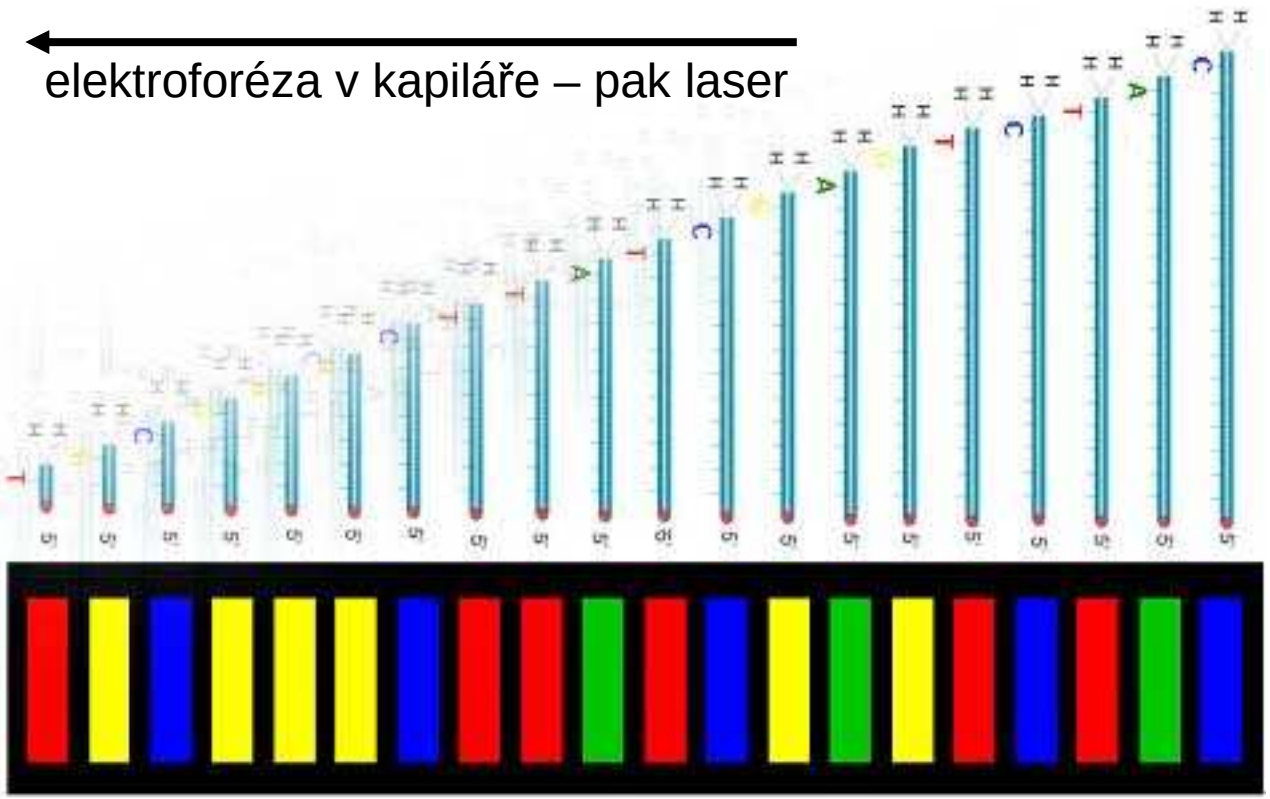
⇒ analýza laserovým sekvenátorem!



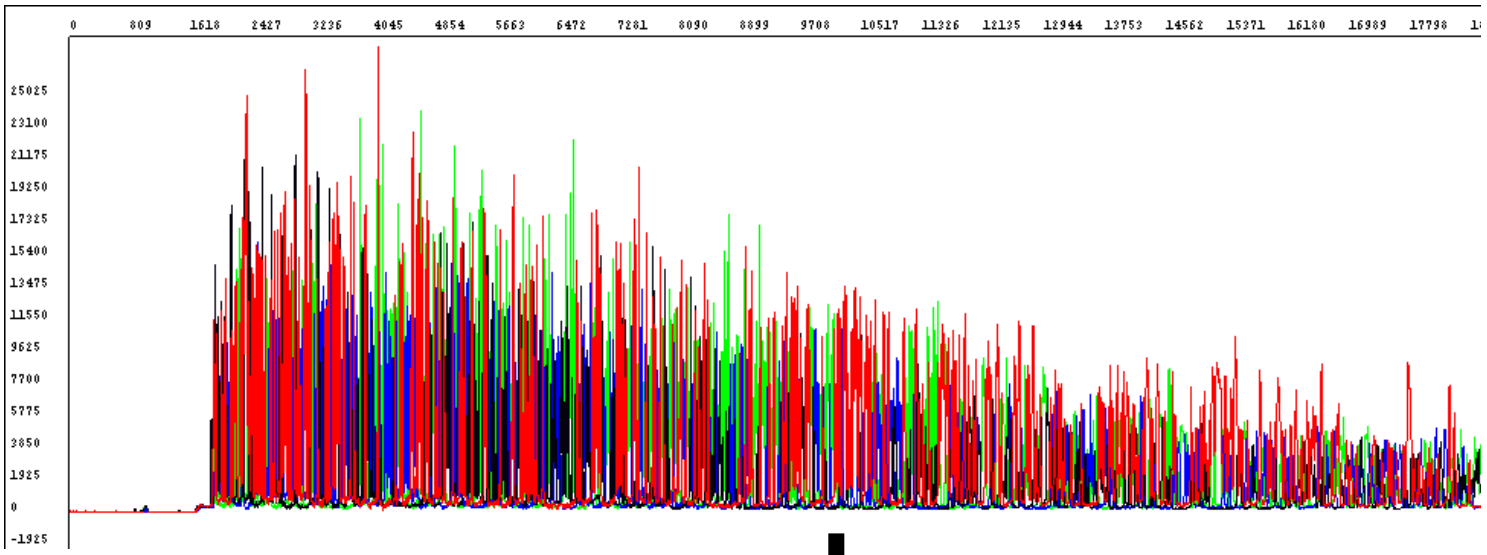
analýza na sekvenátoru:

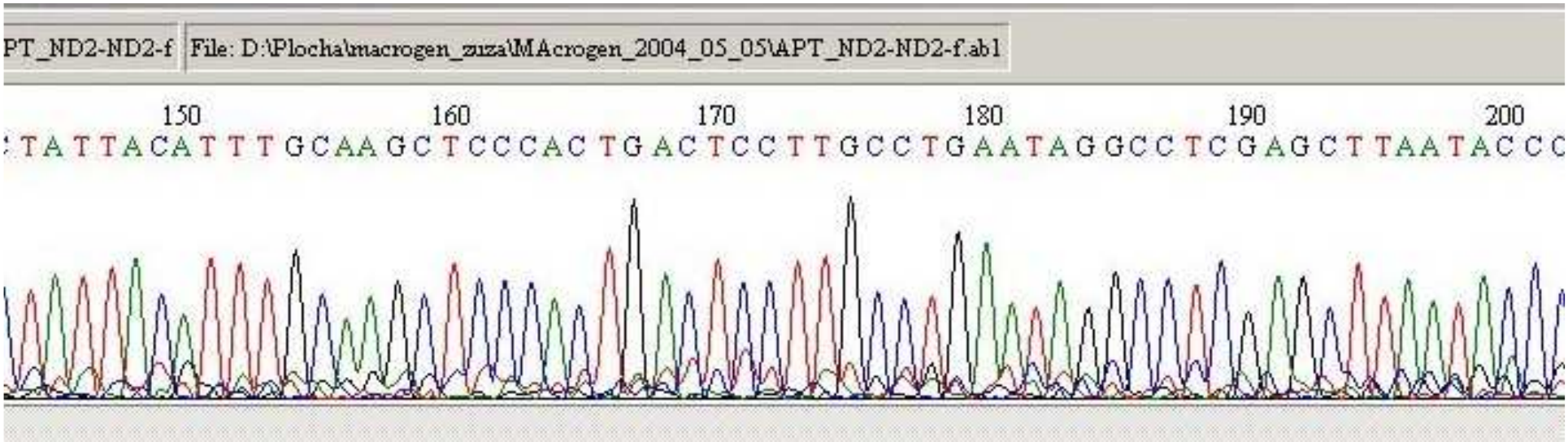


elektroforéza v kapiláře – pak laser

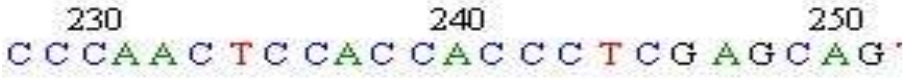


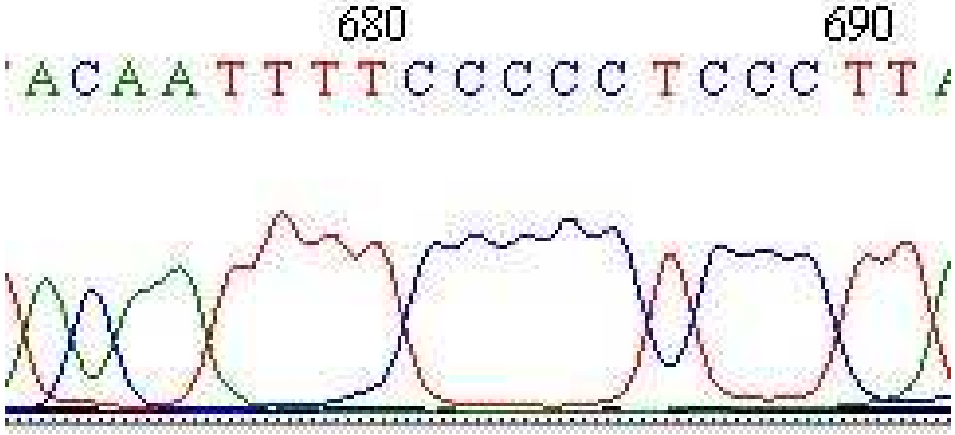
Chromatogram:

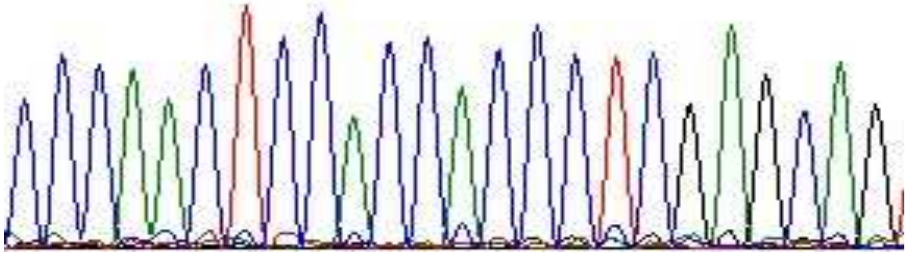




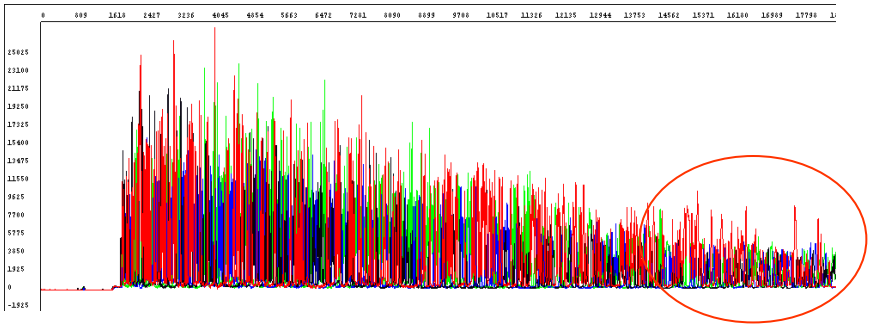
Co je špatně?



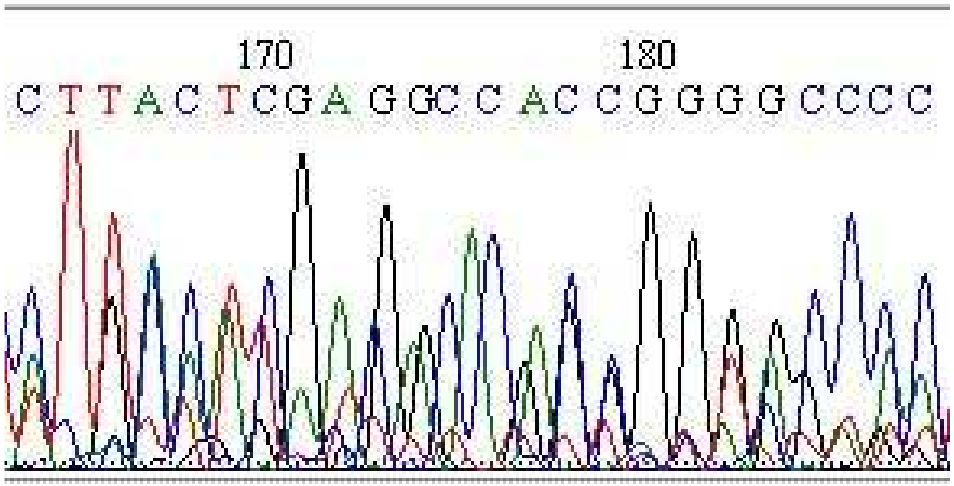




konec sekvence, zhoršuje se čtení laseru



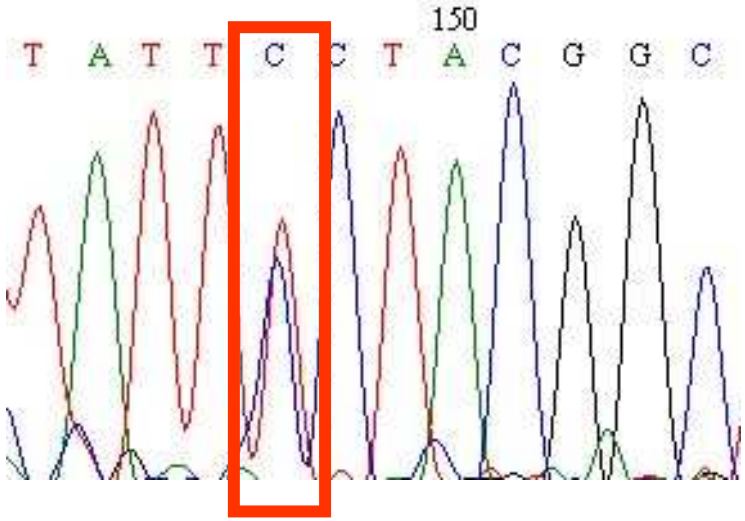
špatná sekvence s velkým množstvím šumu:



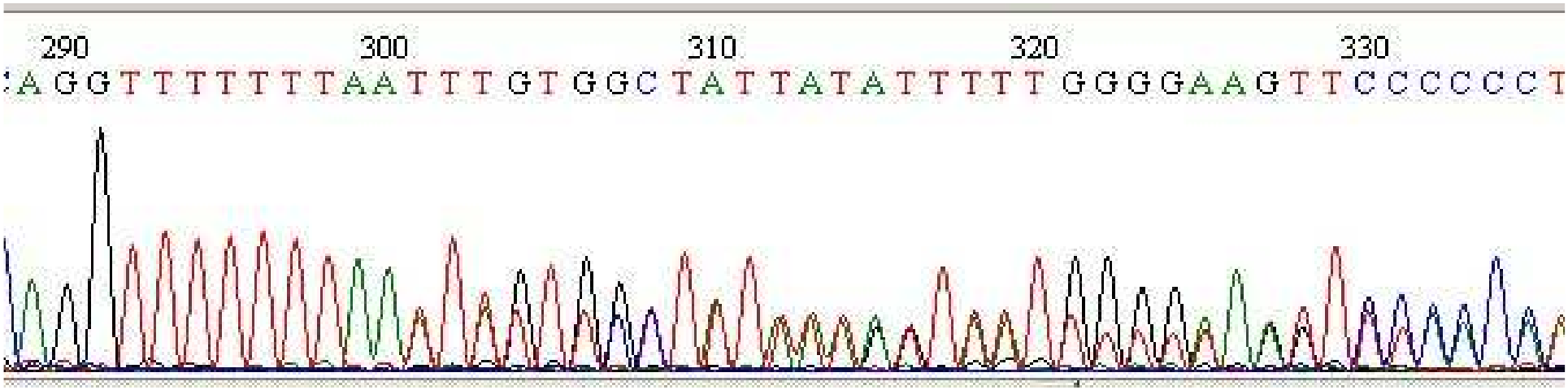
Co je špatně?

pouze u jaderných genů:

heterozygot na jedné pozici:



heterozygot inserce/delece:



délka přečtení sekvence:

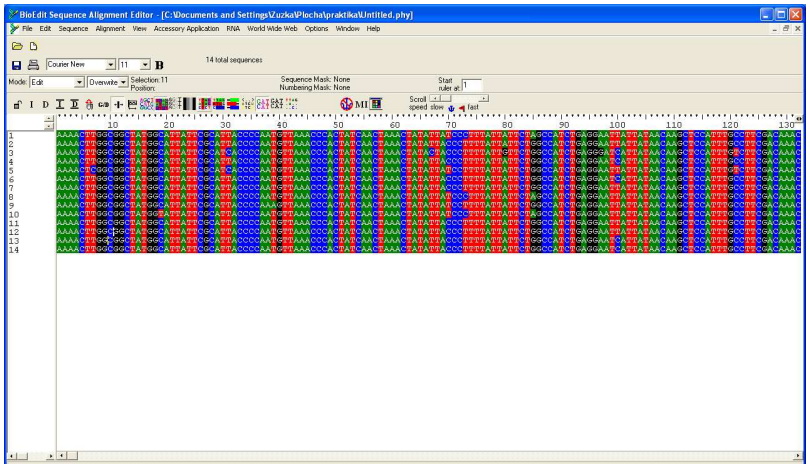
v současnosti lze z kvalitní DNA spolehlivě osekvenovat cca 700 – 800 párů bazí, delší geny nutno sekvenovat z obou stran:

BioEdit

Copyright © 1997-2007
Tom Hall
Ibis Biosciences
Carlsbad, CA 92008

Biological sequence alignment editor for Win95/98/NT/2K/XP

http://www.mbio.ncsu.edu/BioEdit/bioedit.html

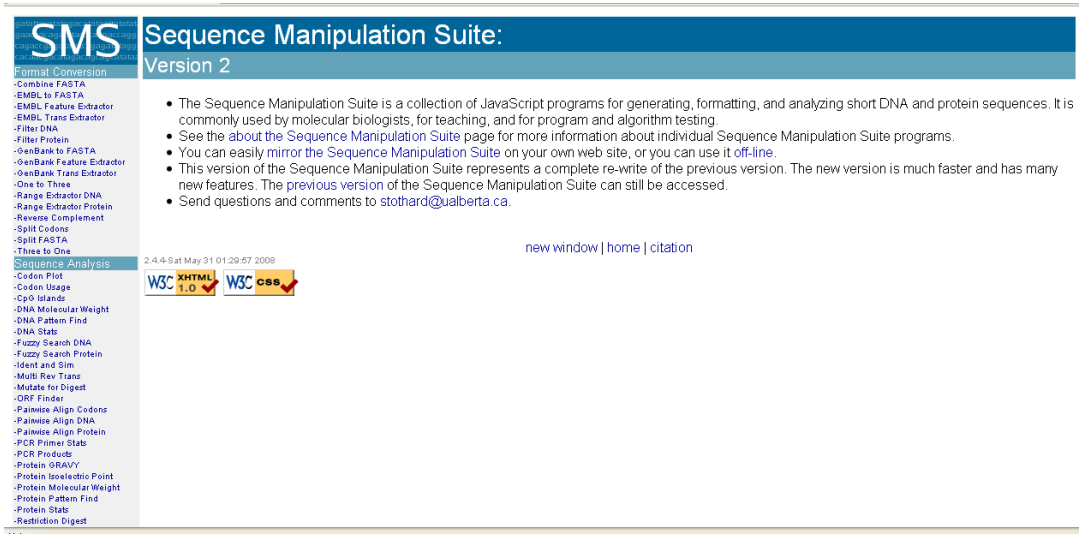


BioEdit - editor sekvencí

- barevné rozlišení
- možnost čtení a editace ABI souborů ze sekvenátoru
- implementováno mnoho dalších funkcí - alignment - CluslalW
- umožňuje slučovat alignmenty, přidávat referenční sekvence, …
- obsahuje část fylogenetického balíku Phylip - možnost jednoduchých fylogenetických analýz – např. neighbour-joining
- BLAST

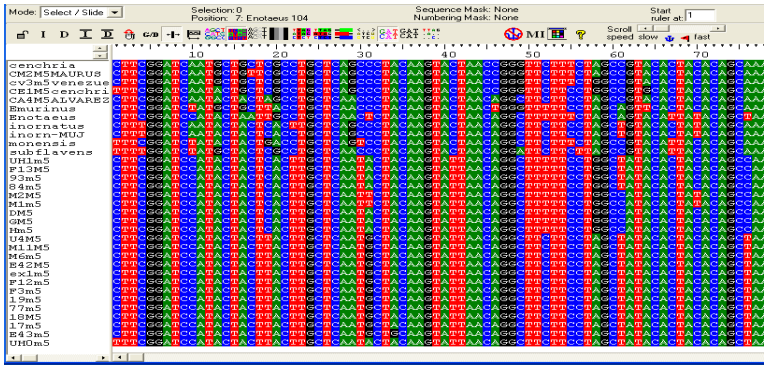
Práce se sekvencemi on-line

SMS- sequence manipulation suite
http://www.bioinformatics.org/sms2/
- různá manipulace se sekvencemi, převod formátů, reverse complement, náhodná sekvence, …



FaBox- umožňuje práci s datasety, převod formátů, rozpojování a spojování datasetů
http://www.birc.au.dk/~biopv/php/fabox/

Kde a čím se sekvence liší?



alignment ➔ start pro fylogenetické analýzy

jedná se o ustanovení poziční homologie jednotlivých bází v sekvenci (jednotlivých znaků vstupujících do analýzy)

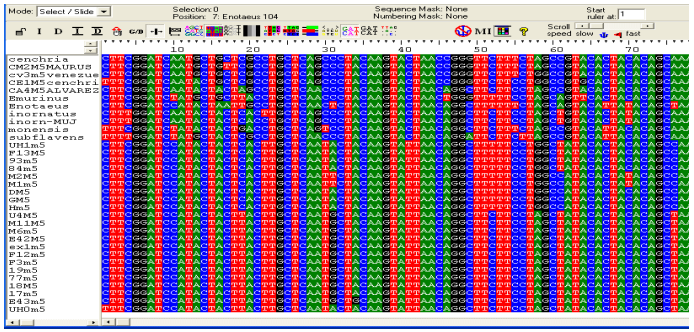
vždy se musí pracovat s homologickými znaky

➔ existují různé programy a editory na tvorbu a úpravu alignmentů - ruční BioEdit, Macaw

- automatické – pracují s různými algoritmy

Clustal X, PileUp, Multalin, Mafft - mnoho z těchto programů jsou online

Kde a čím se sekvence liší?



Alignment - pairwise alignment (dvě sekvence)

- multiple alignment (víc sekvencí)

```
AATGCCCTAAA
AATGCGGCTAAA
AACGCGCTAAA
ATGCTAA
```

```
AATGCC-CTAAA
AATGCGGCTAAA
AACGCG-CTAAA
-ATG---CTAA-
```

Alignment rady a programy:

Clustal W (Clustal X = Clustal W s rozhraním pro Windows) – www.clustal.org - jeden z nejpoužívanějších programů, součást BioEditu

- při prvním pokusu o alignment- ponechat defaultní parametry
- zvyšovat penaltu za otevření mezery a snižovat penaltu za prodlužování gapu
- geny kódující proteiny je dobré alignovat na základě sekvence aminokyselin (při překladu z DNA na aminokyseliny pozor na čtecí rámec a různé genetické kódy (mitochondrie, obratlovci, bezobratlí, kvasinky, nálevníci,…) - GenBank v informacích o sekvenci (codon start=?, translation table=?).
- lze zohlednit sekundární strukturu (geny pro 12S a 16S RNA) – databáze alignmentů udělaných podle sekundární struktury, lze vyhledávat, stahovat, přialignovávat naše sekvence (http://www.arb-silva.de/)

Další metody jak zacházet se sekvencemi složitými na alignování:

Culling - do konečného alignmentu zařazena pouze báze, pro které byla ve všech alignmentech (s různými parametry) určena stejná homologní pozice (Gatesy et al. 1993).

Elision - Více alignmentů - různé penalizace za otevření mezery - zřetězeny v jeden (Wheeler et al. 1995)

Úloha: práce ze sekvencí

Ze sekvenační laboratoře nám poslali sekvenci našeho vzorku a my musíme:

- prohlédnout si sekvence, spojit a případně upravit:
- otevřít v programu BioEdit soubory **vzorekLEU** a **vzorekND4** ze složky „Sekvence “
- otočit reversní primer – **vzorekLEU** (*Sequence – Nucleic Acid – Reverse Complement*)
- vkopírovat otočenou reversní sekvenci (vzorekLEU) k forwardu (vzorekND4) do jednoho souboru
- alignovat pomocí párového alignmentu – *Sequence – Pairwise Alignment – allow ends to slide* – ponechá možnost dlouhých konců
- chromatogram (**vzorekLEU**) sekvence zobrazit převedený: *View – Reverse Complement*
- srovnat pod sebe alignment a chromatogramy obou sekvencí sekvence
- přepnout okno s alignmentem do módu *Edit*
- překontrolovat sekvence, editovat, ořezat konce
- všechny úpravy dělat v jednom řádku tzn. vkopírovat část vzorekLEU do sekvence vzorekND4
- přepnutí na mód *Select/slide* a odstranění všech řádků kromě vzorekND4 (označit sekvenci a dát CTRL+X)
- uložení a máme složenou sekvenci – uložit **vzorek.fas**

porovnání pairwise alignmentu s chromatogramy obou sekvencí

Základy fylogenetických analýz

dva typy dat

znaková data (maximální úspornost (parsimony, MP); maximální pravděpodobnost (likelihood, ML), Bayesiánská analýza, BA)

distanční data (Neighbour-joining, UPGMA)

Přístup metod k výpočtu fylogenetického stromu je dvojí: používají k výpočtu algoritmus (sled specifických kroků) nebo nějaké kritérium optimálnosti

Algoritmus - najde jen jeden strom postupným přidáváním sekvencí (analýza UPGMA, Neighbour-joining = distanční metody).

Prohledávání stromového prostoru – heuristické hledání, Marcov chain Monte Carlo (MCMC) - a hodnocení stromů podle různých kritérií optimálnosti (nejmenší počet kroků, největší likelihood)

Metody výpočtu fylogenetických stromů:

distanční metody (UPGMA, NJ, minimal evolution)

---

znakové metody

- maximální úspornost = max. parsimony
- maximální pravděpodobnost = max. likelihood
- Bayesovská analýza

distanční metody výpočtu stromů:

distanční metody: př.Neighbour-joining (NJ) - používají genetické distance (= vzdálenosti (v %) znaků (=sekvencí)), používá korekce distancí, aby umožnila odhadnout počet nezjištěných změn
výpočet vzdáleností (rozdílů) každé sekvence od každé – vznikne matice vzdáleností
p distance = nekorigovaný rozdíl v sekvenci dvou vzorků

p= počet rozdílných bází v sekvenci/ počet všech nukleotidů

jaké korigované distance použít? Jukes-Cantor, Kimurova K2P distance, …
dle Kumar et al. 1993: $d_{JC} < 0.05$, použít d_{JC}
$0.05 < d_{JC} < 0.30$, použít d_{JC} pokud není poměr transice:transverse příliš vysoký (potom d_{K2P} nebo $d_{K2P+\Gamma}$)
$0.30 < d_{JC} < 1.00$ a rozdílná rychlost změn na různých místech sekvence – gamma distance; záleží i na délce sekvence – víc znaků snese složitější model

maximální parsimonie:

princip metody je vybrat strom s minimální celkovou délkou (nejmenší počtem evolučních kroků – např. nejméně substitucí nukleotidů)

problém: všechny stromy lze dělat pouze při nízkém počtu sekvencí, počet možných stromů roste podle vzorce:

$$\frac{(2n-3)!}{2^{n-2}(n-2)}$$

3 taxony: (6-3)!/2(1) = 6/2 = **3 stromy**

6 taxonů: (12-3)!/2^4(4) = **5670 stromů**

9 taxonů: (18-3)!/2^7(7) = **1 459 458 000 stromů**

12 taxonů: (24-3)!/2^{10}(10) = **498 934 949 821 000 stromů**

exponenciální nárůst počtu potenciálních stromů

parsimonie

není tedy možné prohlédnout všechny stromy, spočítat pro ně počty evolučních změn a vybrat ten nejlepší

⇒ **heuristické hledání stromů**

tj. - vytvoří se náhodný strom,
- spočítají se evoluční změny,
- náhodně se v něm přehodí dvě větve,
- spočítají se evoluční změny
- dál postupuje jen ten, který měl méně změn – opakuje se přehazování
- na konci řady je strom s nejméně změnami

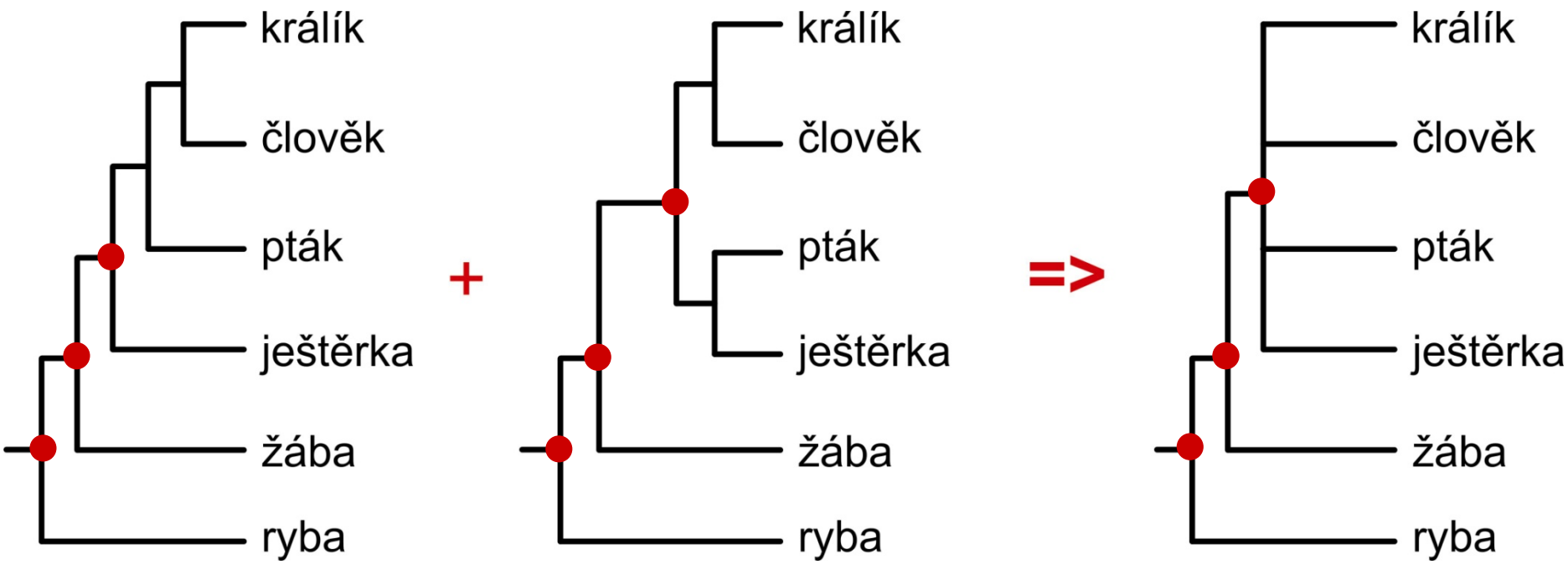
výběr náhodného stromu se opakuje několikrát, porovnání výsledných stromů

→ metodou se často nalezne několik stejně parsimonních stromů
- pro získání jednoho stromu – nutno udělat konsenzuální strom

konsensuální strom:

je-li výsledkem MP analýzy několik stromů, je nutné provést tzv. **konsensus**

- **strict consensus** (ve výsledném stromu jen bifurkace nalezené ve všech stromech)
- **majority-rule consensus** (zůstanou bifurkace, které ve většině stromů)



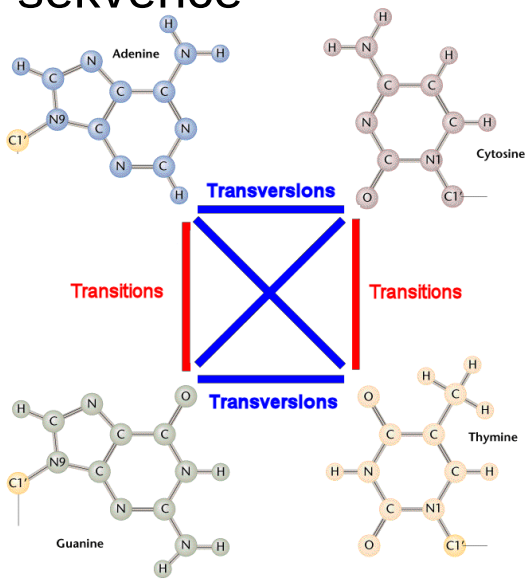
Maximální pravděpodobnost = maximum likelihood

- metoda posuzuje hypotézy o evoluční historii zkoumaných taxonů z hlediska pravděpodobnosti, že jsou v souladu se získanými daty. Vyšší pravděpodobnost stromu je preferována nad nižší, nutno zadat model evoluce sekvencí
- vychází z modelu evolučního procesu, který vede ke změně jedné sekvence v druhou (substituční model)
- často se model navrhuje pomocí dalších programů (Modeltest, jModeltest) na základě vstupních dat - obecně modely uvažují:

• frekvenci jednotlivých bází
• pravděpodobnost změny jednoho nukleotidu v druhý (transice x transverze)
• heterogenita substitučních rychlostí v různých částech sekvence

příklady modelů: Jukes-Cantor,
Kimurův dvouparametrový model,
general time-reversible model (GTR)

Výsledkem ML je jeden strom



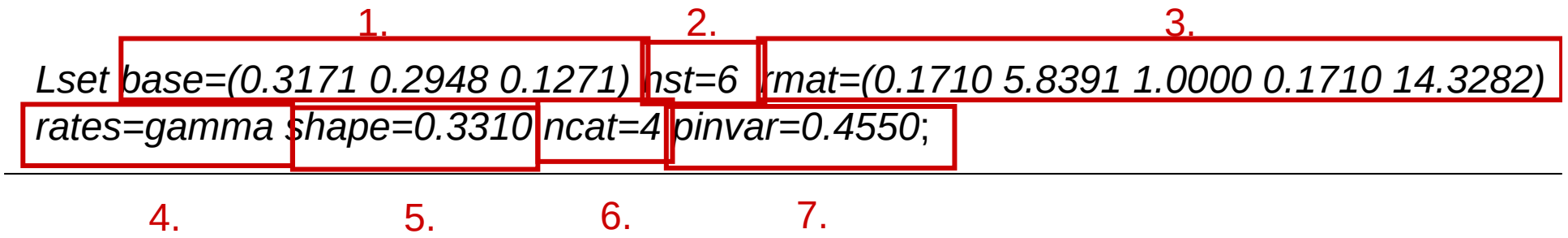
výběr nejvhodnějšího modelu

program jModeltest (Modeltest)

Návod na spuštění: http://fyloshop.webnode.cz/news/navod-na-beh-modeltestu

příklad modelu:

Lset base=(0.3171 0.2948 0.1271) nst=6 rmat=(0.1710 5.8391 1.0000 0.1710 14.3282) rates=gamma shape=0.3310 ncat=4 pinvar=0.4550;

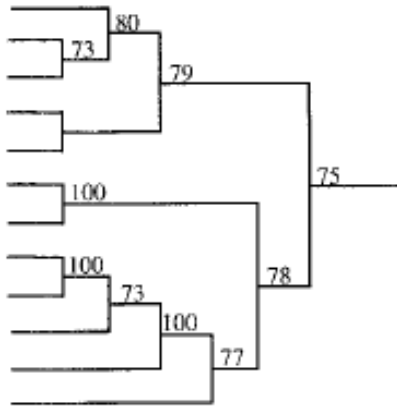


1. - poměry jednotlivých bází (čtvrtá se dopočítává do 1)
2. - počet substitučních typů (1 = všechny záměny stejně pravděpodobné, 6 = každá jinak)
3. - matice míry každého typů záměny (poslední chybí, je to vždy 1 a ostatní hodnoty jsou relativní k ní)
4. - typ rozložení pravděpodobností záměn na jednotlivých pozicích (equal = shodné pro vše, gamma = různé, mající gamma rozložení
5. - sklon; parametr gamma distribution
6. - kategorie gamma distribution
7. - poměr nevariabilních míst

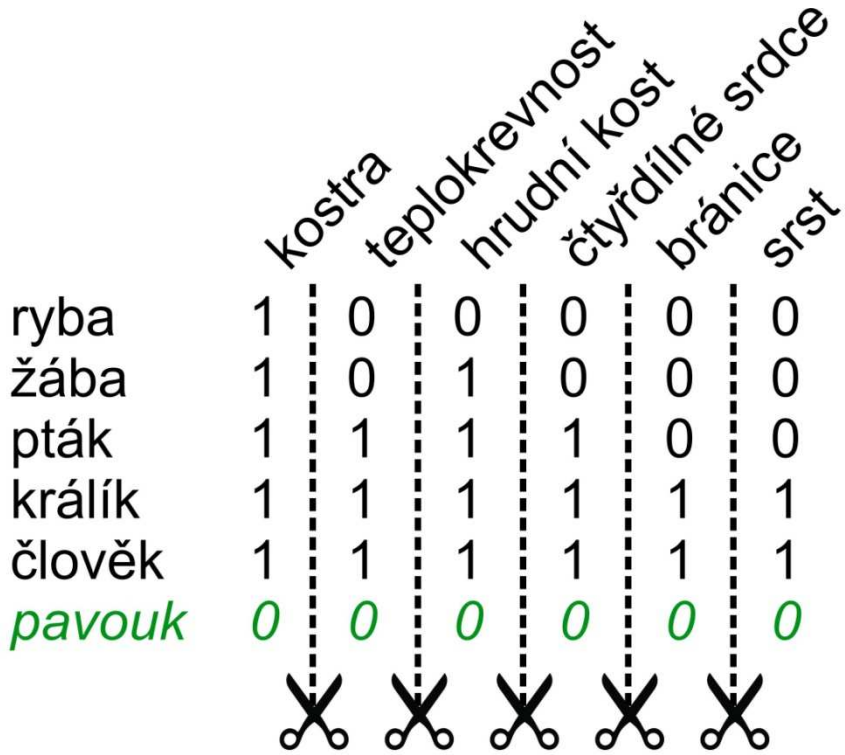
podpora zjištěných topologií (stanovení spolehlivosti)
-nejčastěji využití neparametrických technik opakovaného výběru
bootstrapping, jackknifing

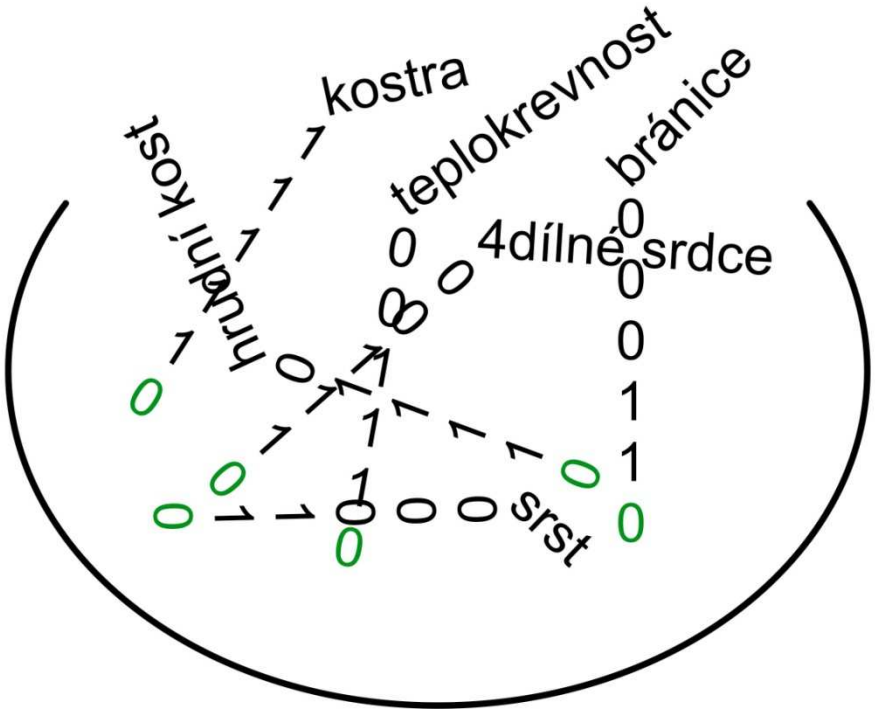
➡ náhodné vybírání znaků z původního datasetu (nukleotidů)
-ve fylogenetice vlastně vytvoříme mnoho datasetů (založených na stejných datech jako originální dataset, ale v jiném zastoupení), z každého datasetu vypočteme strom, ze všech takto vzniklých stromů vytvoříme konsenzuální strom a procentuální zastoupení jednotlivých větvení (topologie taxonu) ukazuje na míru spolehlivosti

větve s bootstrapem < 50% - nelze topologii věřit – může se jednat o náhodu, nad 75% u MP, ML – uspokojivě spolehlivé, 95-100% velmi dobré



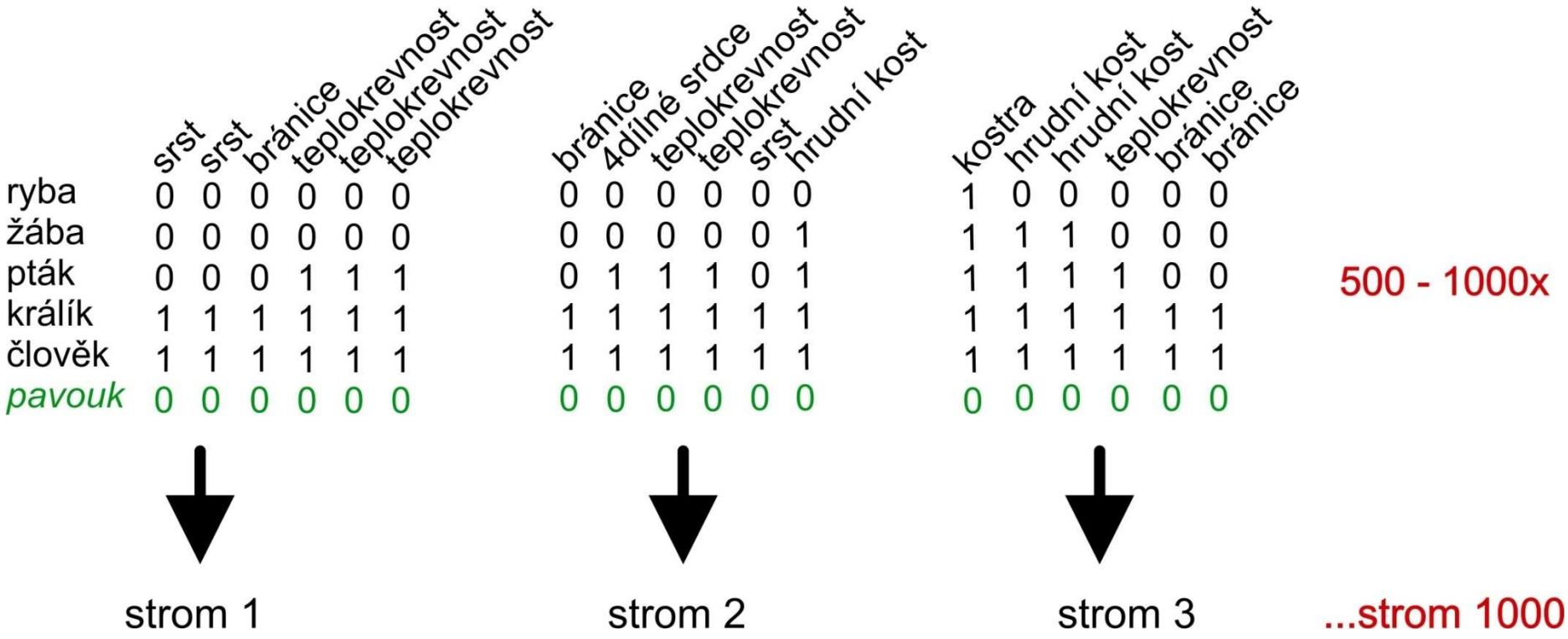
	kostra	teplokrevnost	hrudní kost	čtyřdílné srdce	bránice	srst
ryba	1	0	0	0	0	0
žába	1	0	1	0	0	0
pták	1	1	1	1	0	0
králík	1	1	1	1	1	1
člověk	1	1	1	1	1	1
pavouk	*0*	*0*	*0*	*0*	*0*	*0*





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

Bootstrap - pseudomatice



Bayesovská analýza

znaková metoda, založená na modelu evoluce sekvencí a používající při výpočtu posteriorní pravděpodobnosti (probability) = pravděpodobnosti vypočtené na základě nějakých předpokladů (priors)

Thomas Bayes (18. století) vymyslel statistickou metodu a tzv. Bayesův teorém

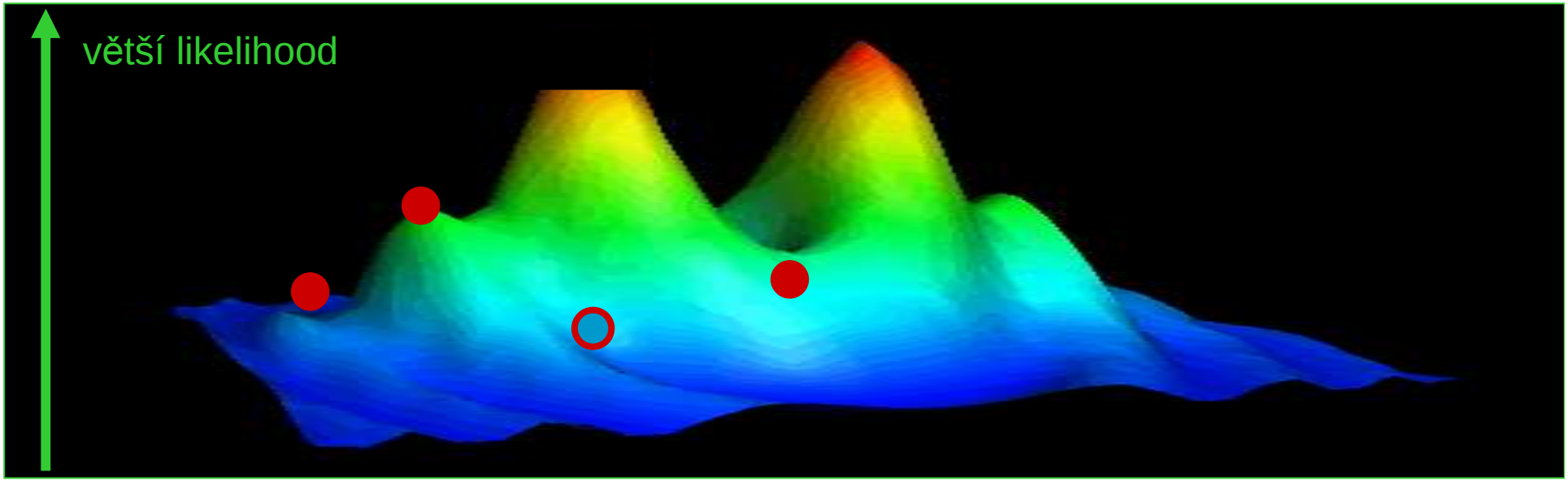
je to mírně modifikovaná forma likelihoodu

velmi zjednodušeně:
Likelihood = pravděpodobnost stromu z našich dat
BT = pravděpodobnost našich dat při určitém stromu...

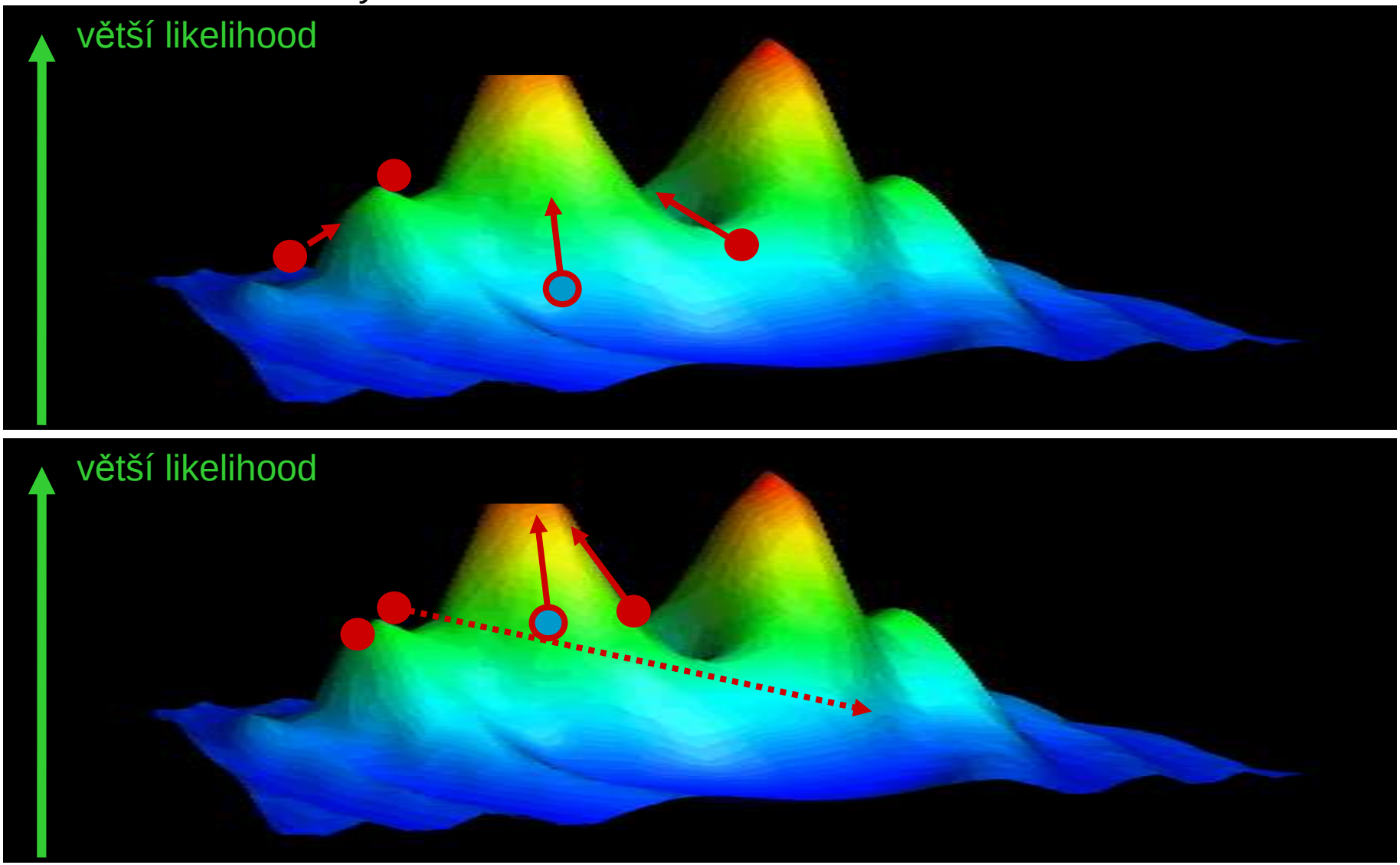
(také hledá nejlepší strom)

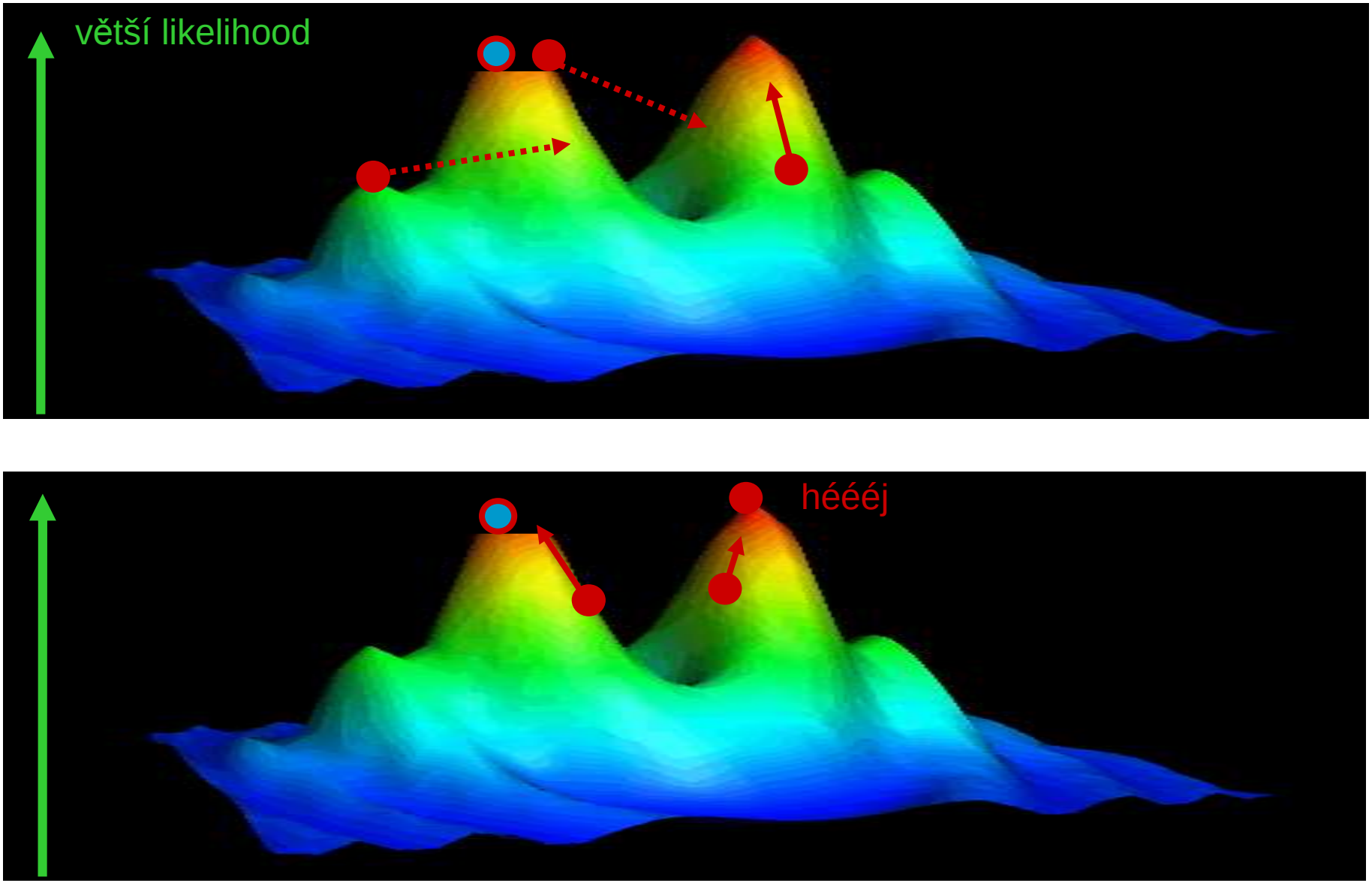
Bayesovská analýza

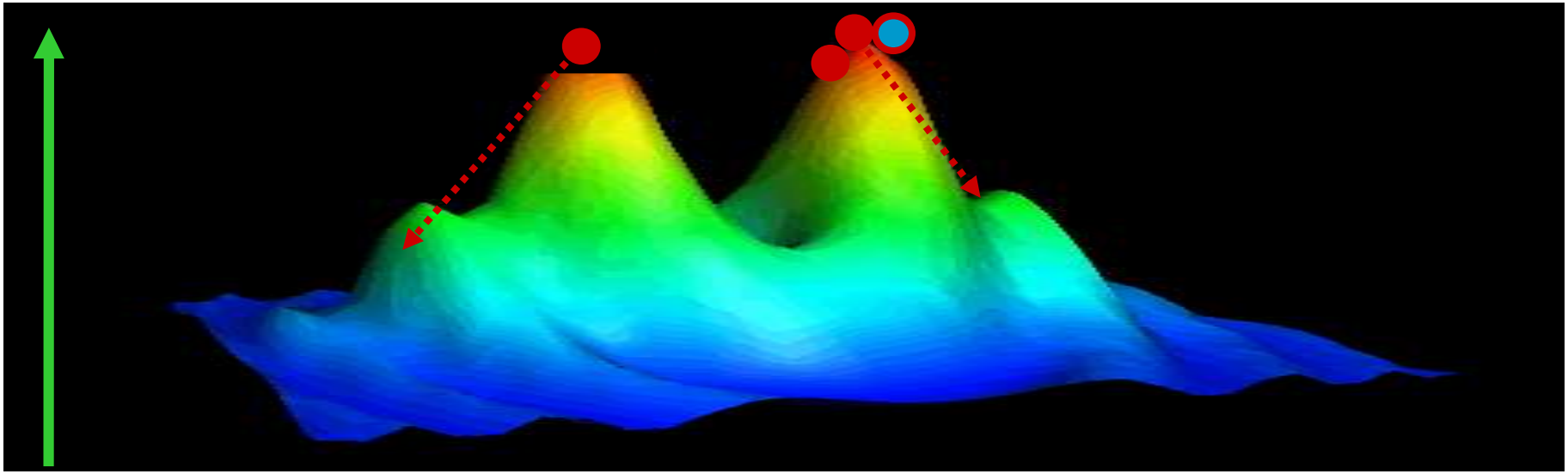
průběh Bayesovského hledání
krok 1: máme 4 řetězce, které se vydaly hledat do krajiny nejlepší strom...
hledání je opět heuristické, tedy zkusím strom, spočítám jeho L, zkusím další, posunu se pouze, je-li nový strom lepší...



1 řetězec je studený – tzn. konzervativní, posune se pouze nahoru – tedy pokud je další strom lepší
3 řetězce jsou teplé – tzn. mohou se vrátit i dolů + skáčou náhodně na jiná místa
teplé řetězce volají studeného, pokud najdou lepší strom = vyšší vrcholek, než na které se usídlil studený




větší likelihood
héééj



při dostatečném počtu generací (tj. hledacích kroků) najde studený řetězec nejvyšší vrchol v krajině, tj. strom s nejlepším Likelihoodem

Posteriorní pravděpodobnosti

PP je parametr Bayesovské analýzy – slouží namísto Bootstrapů

výstupem programu **MrBayes** bude soubor, který zaznamenává nalezené stromy studeného řetězce v průběhu hledání.

Celkem jsou to třeba až statisíce stromů.

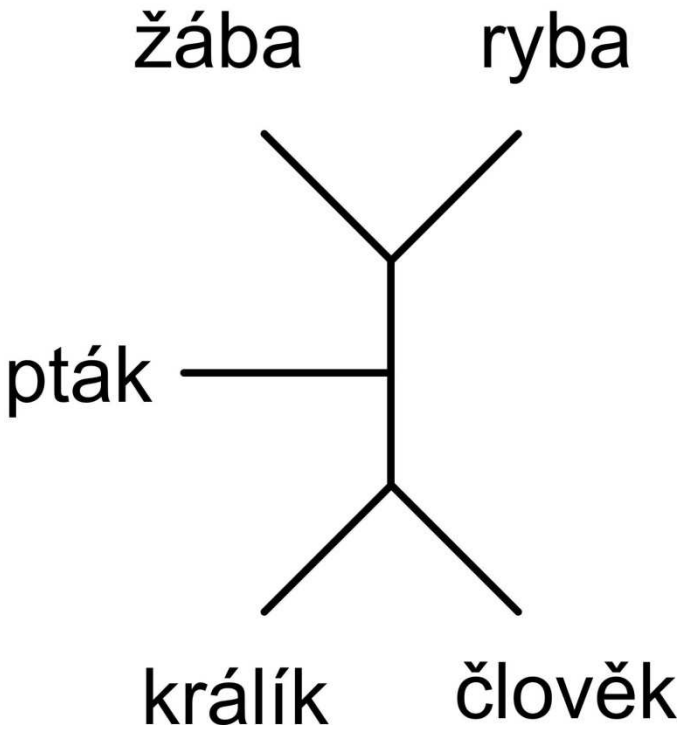
Stromy na začátku hledání nebudou dobré – musíme je odstranit z další analýzy

PP: říká nám v kolika procentech zaznamenaných stromů studeným řetězcem se vyskytuje daný uzel
posteriorní probability- pod 95 až 90 (resp. 0.95 – 0.9) topologie nejistá

Zakořeněné vs. nezakořeněné fylogenetické stromy:

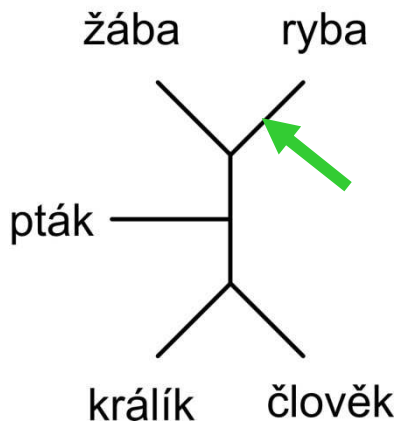
Outgroup ukáže, kde je kořen našeho stromu. Outgroup by měl být fylogeneticky co možná nejbližší studované skupině

nezakořeněný strom:



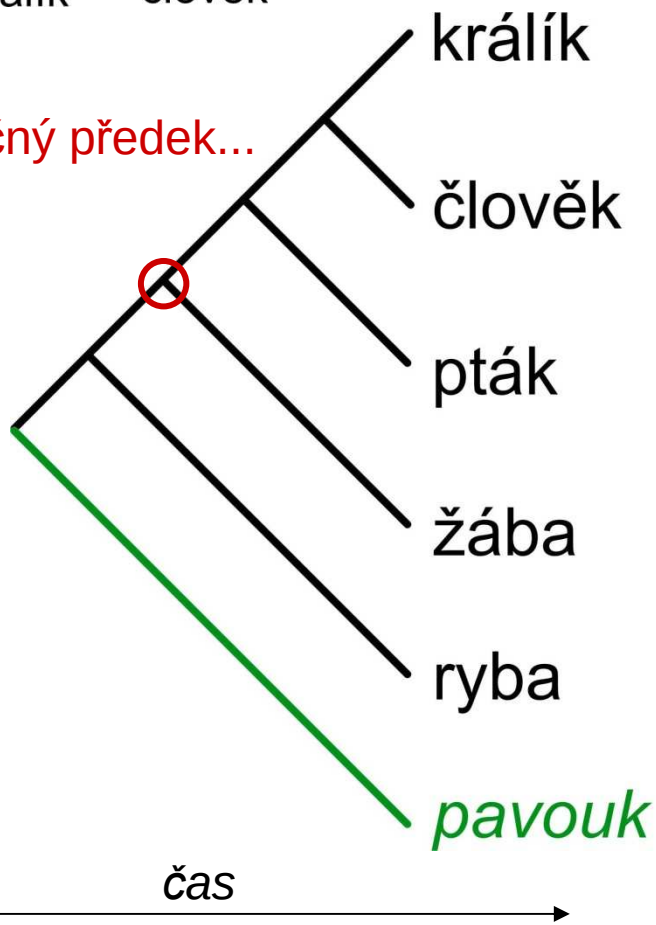
přidat druh, který nepatří do skupiny, tzv. *outgroup*, tj. zde druh, který není obratlovec...

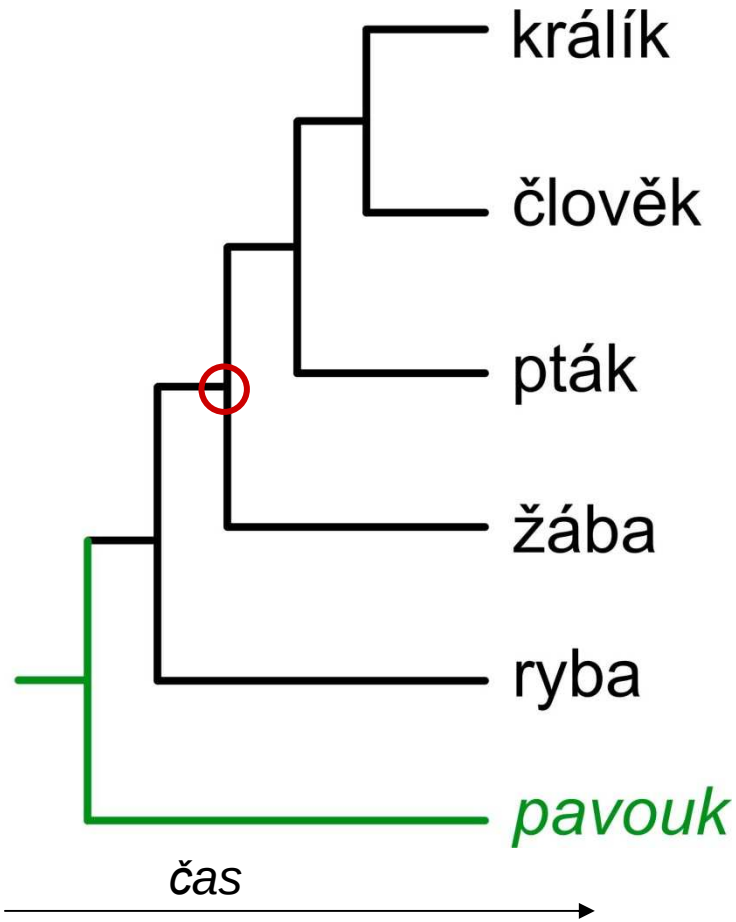
nezakořeněný strom:



zakořeněné stromy = **kladogramy:**

společný předek...





Take -Home Message!

celkem je několik metod rekonstrukce fylogenetických stromů
fylogenetický strom jako výstup jakékoliv metody je jen hypotéza - musíme zjistit spolehlivost topologií
výsledek vždy závisí na kvalitě vstupních dat a na dobrém alignmentu (určení homologických znaků, které pak porovnáváme)

Software: MP: PAUP*, TNT, Phylip, …
ML: PAUP*, PHYML, GARLI, RAxML, Phylip
BA: MrBayes
NJ: PAUP*, Phylip, MEGA, …

+ různé internetové aplikace

úloha:

příbuzenské vztahy leguánů – zejm. rod *Cyclura*

• stáhneme sekvence z GenBank – pro tento případ si stáhneme sekvence použité Malone et al. 2000 - **Phylogeography of the Caribbean rock iguana (*Cyclura*): implications for conservation and insights on the biogeographic history of the West Indies.**

• uložit sekvence leguánů rodu *Cyclura* – vždy jednu od druhu + jednu sekvenci *Iguana iguana*)

• stáhnout sekvence použité autory
do souboru –FASTA formát – otevřít v BioEdit - pozměnit jména aby byla zpracovatelná programy a srozumitelná (8 znaků) (možná líp se to dělá tak, že sekvence uložím do poznámkového bloku a tam přejmenuji a pak uložím a otevřu v BioEditu)

• přidat k datasetu z článku naši složenou a překontrolovanou sekvenci
• udělat alignment - Accesory Application - ClustalW, oříznout konce, exportovat do formátu nexus
• vytvořit fylogenetický strom metodou NJ a Maximální parsimonie v PAUPu

Postup práce se sekvencemi, příprava alignmentu:

- otevřít dataset ze sekvencemi v programu BioEdit
- přikopírovat pomocí Ctrl+C a Ctrl+V sekvenci vzorekND4 uložit soubor „leguani-cely.fas“
- provést alignment volbou: *Accesory application – ClustalW Multiple Alignment*
- ořezat sekvence tak, aby začínaly všechny stejně na první místo přesunout zvíře, které bude jako outgroup (*Iguana*)
- exportovat do formátu Nexus: file - *EXPORT – SEQUENCE ALIGNMENT*, volba Nexus, jmeno souboru včetně přípony („leguani.nex“)

Bayesiánská analýza

program MrBayes – ovládá se příkazovým řádkem

důležité příkazy:
execute leguani.nex (načtení souboru)
lset nst=6 rates=invgamma (zadání modelu – zjednodušeně dle modeltestu)
mcmcp ngen=500000 samplefreq=100; (nastavení parametrů, počet generací, jak často ukládat stromy)
mcmc (spuštění)

po doběhnutí překontrolovat jaká je hodnota „average standard deviation of split frequencies“ - měla by být 0,01
prohlédnout -lnL v každé zaznamenané generaci – dosáhli jsme plató?
(programy Excel – udělat graf a vynést –lnL v každé generaci, Tracer, AWTY)
-jestliže hodnota vyšší nebo není dostatečně dlouho plató, tak přidat generace

po analýze:
sump burnin=1250
sumt burnin=1250 (odstranění prvních 25% získaných stromů) – udělá „.con“ soubor, který se otevře v programu na prohlížení stromů
(nebo lze napsat: relburnin=yes burninfrac=0.25)

seznámení s programem PAUP*:

základní příkazy v PAUPu:
Pro Windows se ovládá pomocí příkazů psaných do příkazové řádky
Vstupem alignment ve formátu nexus
pro výpočet stromů obecně:

set criterion= (nastavení metody – tj. distance/parsimony/likelihood
showtree (ukáže výsledný strom v okně programu)
savetree file= (uloží výsledný strom jako soubor)

výpočet distancí:

showdist (výpočet distancí mezi sekvencemi; ukázaní v okně programu)
savedist file= (uloží spočítané distance jako soubor)

otevřít soubor leguani.nex v PAUPu

metoda NJ v programu PAUP

Nj strom

set criterion=distance (přepnem do distančního modu PAUPu –pozor!)
dset distance= jc (nastavime model jc – mužeme i jiny)
nj; (příkaz pro zahájení algoritmu metody nj)
savetrees file=leguaninj.tre brlens=yes; (uložíme nj strom s délkou větví)

Bootstrap

bootstrap nrep=1000 search=nj;
savetrees file=leguaninjboot.tre from=1 to=1000 savebootp=nodelabels;
(uloží strom do souboru leguaninjboot.tre s hodnotami bootstrapu uvedených na nodech)

Nápověda v PAUPu: dáme příkaz a ? př.: dset ? nebo pouze ?
pak zobrazí seznam možných příkazů, podrobný manuál

metoda MP v programu PAUP

příkazy pro MP analýzu:

set criterion = parsimony

hsearch addseq=random nrep=10 swap=TBR;

„number of replication“ = počet opakování hledání náhodného stromu

heuristické hledání

„addition of sequences“ = postupné přidávání sekvencí, volba „random“, proto aby nemohl být výsledek ovlivněn pořadím druhů v alignmentu

describe all/plot=p;

savetrees file=leguanimp.tre brlens=yes;

pokud víc stromů:

contree all/majrule=yes treefile=leguanmajrulepars.tre

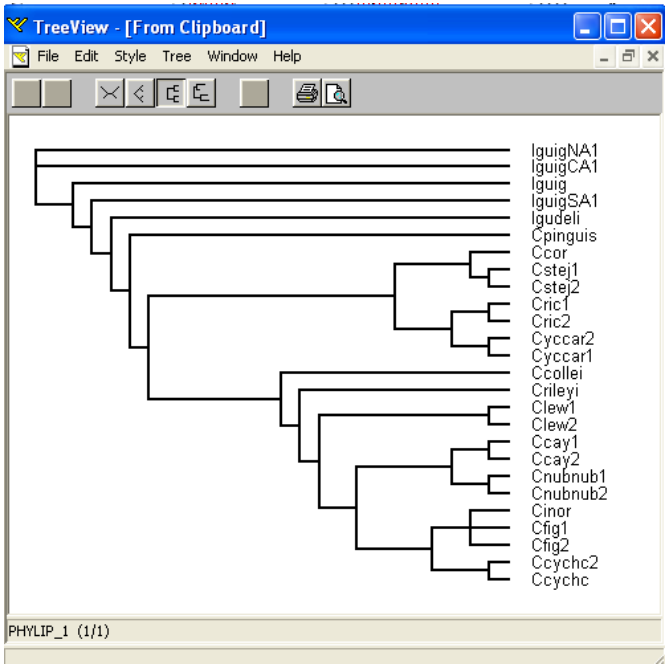
Bootstrap:

bootstrap search=heuristic nrep=1000; (parametry hledání nastavené stejně jako při hledání stromu)

savetrees file=leguanimpboot.tre from=0 to=10000 savebootp=nodelabels;

Strom v závorkové konvenci – lze vložit např. do TreeView a zobrazit v grafické podobě

(IguigNA1_:0.00221,(Iguig:0.01733,(((Cpinguis:0.05228,((((Cstej1:0.00012,
Cstej2:0.00098):0.00354,Ccor:0.00543):0.03863,((Cric1:0.00184,
Cric2:0.00184):0.01853,(Cyccar2:0.00636,Cyccar1:0.00498):0.01702):0.03298):0.01722,
(Ccollei:0.04462,(Crileyi:0.01300,((Clew1:0.00221,Clew2:0.00221):0.00885,
(((Ccay1:0.00442,Ccay2:0.00111):0.00111,(Cnubnub1:0.00111,Cnubnub2:0.00000):0.00332):0.
00885,((Cinor:0.00000,Cfig1:0.00111,Cfig2:0.00332):0.00221,(Ccychc2:0.00000,
Ccychc:0.00000):0.00221):0.01051):0.00442):0.01023):0.02222):0.01447):0.02213):0.06215,Igu
deli:0.05379):0.02871,IguigSA1_:0.02231):0.01069):0.02471,IguigCA1_:0.00553);

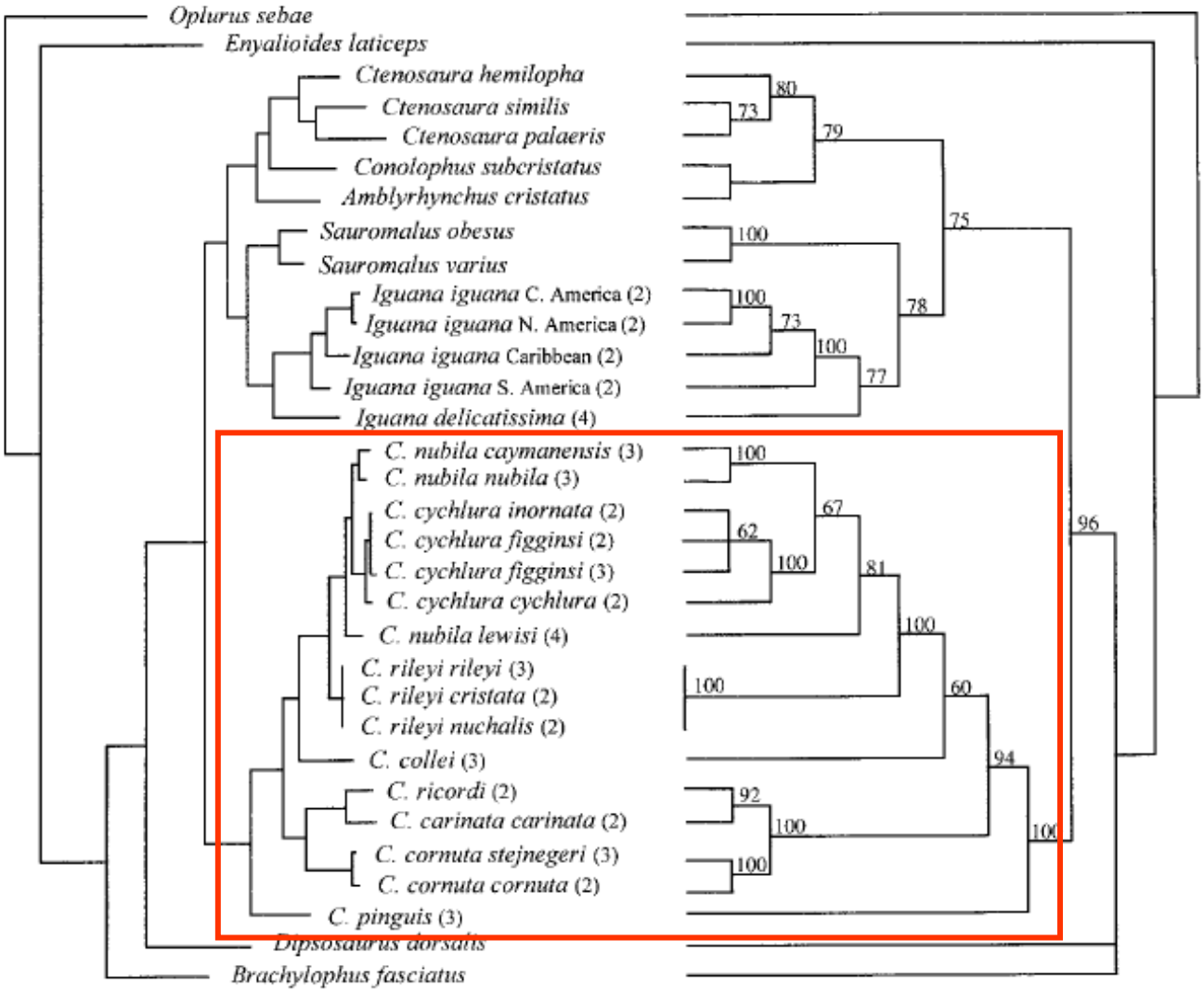


stromy lze prohlížet v různých programech TreeView, FigTree, Dendroscope

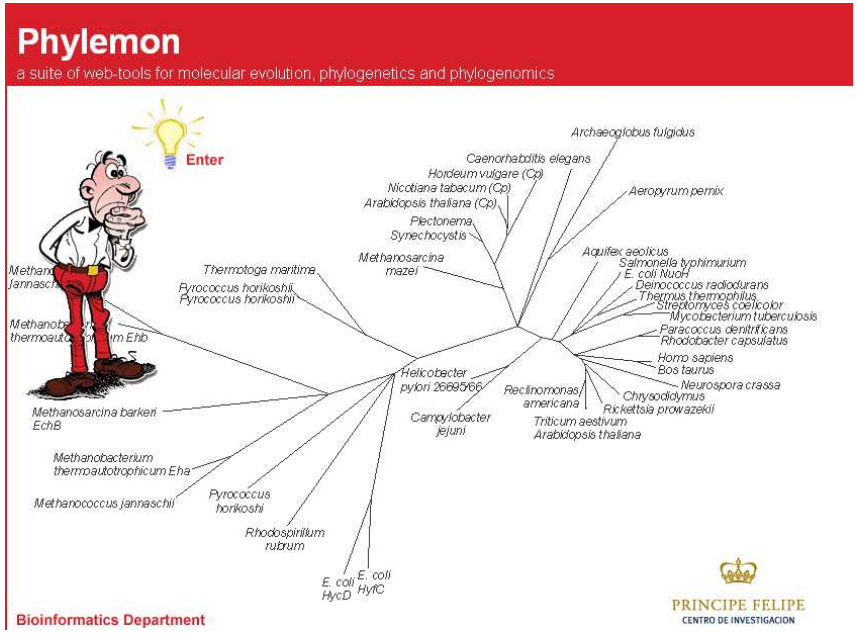
Molecular Phylogenetics and Evolution
Vol. 17, No. 2, November, pp. 269–279, 2000
doi:10.1006/mpev.2000.0836, available online at http://www.idealibrary.com on IDEAL®

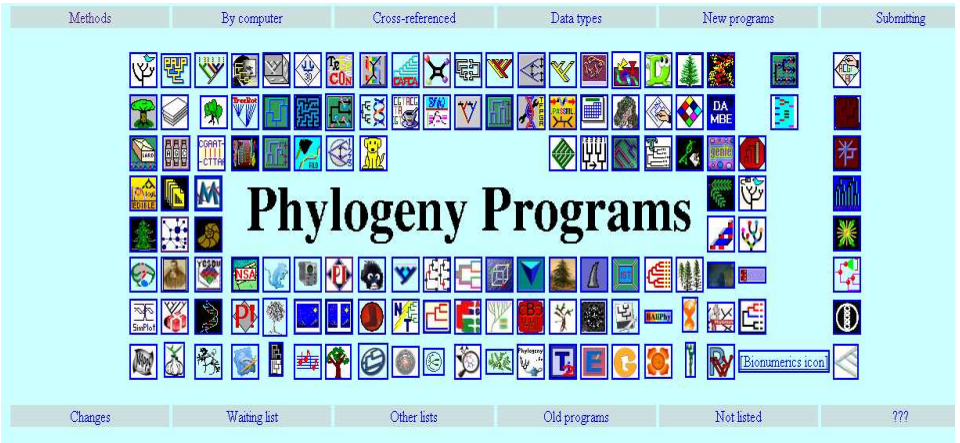
Phylogeography of the Caribbean Rock Iguana (*Cyclura*): Implications for Conservation and Insights on the Biogeographic History of the West Indies[1]

Catherine L. Malone,[2] Tana Wheeler, Jeremy F. Taylor, and Scott K. Davis



Kde hledat na internetu?
stránky J. Felsensteina:
http://evolution.genetics.washington.edu/phylip/software.html
-téměř úplný přehled programů pro fylogenetické analýzy, zobrazování stromů, testy, a vše další potřebné
portál Phylemon: http://phylemon.bioinfo.cipf.es/cgi-bin/home.cgi
-některé analýzy on-line





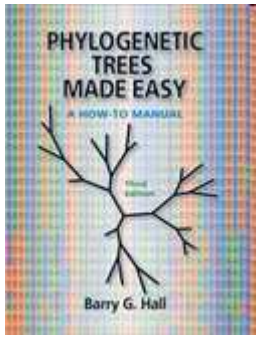
Kde se dozvědět více?

- **Kurz Computational Genomics**
  (Marc VanRanst, Bioinformatics bookmarks http://www.kuleuven.ac.be/rega/mvr/bioinformatics.htm)
- **Introduction to Bioinformatics**
  (F. Cvrčková, http://kfrserver.natur.cuni.cz/studium/prednasky/bioinfo/index.html)
- **Molekulární ekologie**
  (Pavel Munclinger, letní semestr, populační genetika, analýza paternity)
- **Evoluční genetika**
  (Pavel Munclinger a Radka Storchová, zimní semestr)
- **Molekulární taxonomie + cvičení**
  (http://web.natur.cuni.cz/~vlada/moltax/)
- **Fylogenetický workshop**
  (http://www.fyloshop.webnode.cz/)

Kde najdu adresy stránek z tohoto praktika?

(http://www.natur.cuni.cz/~muncling a na stránce Laboratoře pro výzkum biodiverzity/přednášky – http://web.natur.cuni.cz/zoologie/biodiversity/index.php?page=prednasky)

Další čtení:



Phylogenetic Trees Made Easy: A How-To Manual, Edition
Barry G. Hall, Emeritus, University of Rochester

Genetické metody v zoologii
J. Zima, M. Macholán, P. Munclinger, J. Piálek
Karolinum 2004

Úvod do praktické bioinformatiky
F. Cvrčková
Academia