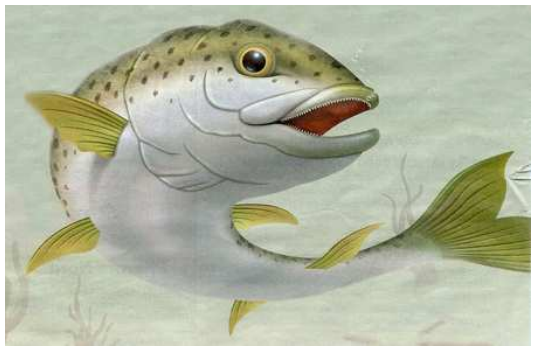
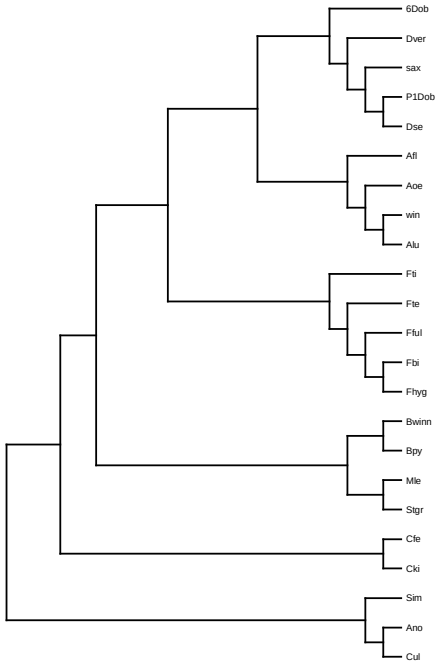
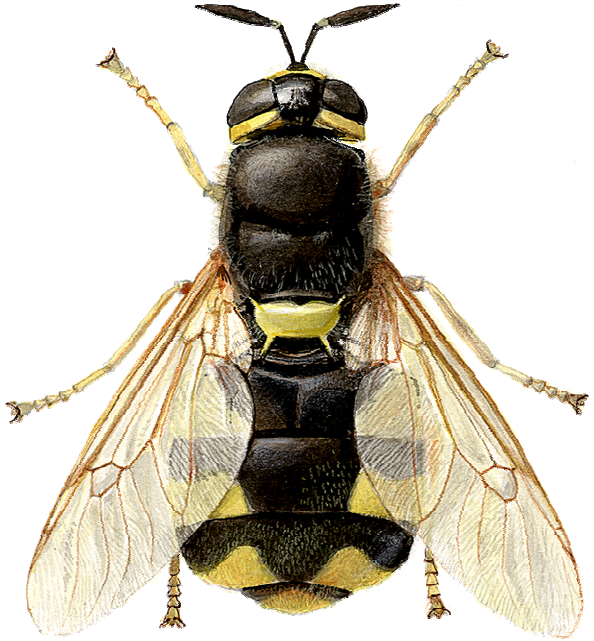
Molekulární taxonomie a fylogenetika
II
(rekonstrukce fylogenetických vztahů)







Andrea Tóthová

Základné pojmy

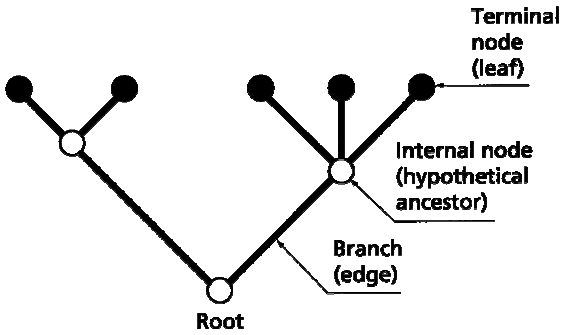
Fylogenetický strom – s kořenem („rooted“)
– bez kořene („unrooted“) – nejstarší společný bod není naznačen, nedefinuje evoluční cestu

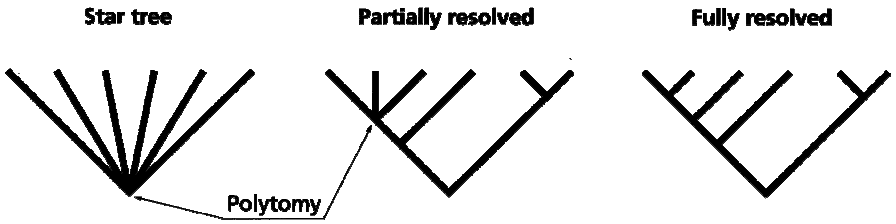
(Kladogram – info jen o pořadí větvení taxonů, délka větví není relevantní, taxony vždy na vrcholu větví bez ohledu na extinkci)

Uzly – terminální (externí)
– vnitřní (interní)

Větve – periférní (končí terminálním uzlem)
- vnitřní (spojují 2 vnitřné uzly)
- centrální (spojují 4 periférní větvy)

Bifurkace (dichotomie), multifurkace (polytomie) – plně vyřešený
- částečně vyřešený
- hvězdicovitý





homologie – obě sekvence mají stav znaku přímo od ancestora
homoplázie – podobnost znaku mezi sekvencemi se vyskytuje nezávisle

Variabilní místa

→ singletons – pouze jeden taxon nese mutaci
→ sites phylogenetically informative – alespoň dva taxony nesou mutaci

sites phylogenetically uninformative – invariable sites + singletons

Consistency index – hodnota počtu homoplázií charakteru v dosaženém kladogramu
CI=m/s (minimum/observed)

Retention index – hodnota počtu podobností charakteru
RI=(g-s)/(g-m)
g the maximum number of the substitutions

Rescaled index = R*C

Monofyletická/parafyletická/polyfyletická skupina

Zpracování sekvenačních dat

Manuální korekce sekvencí - Sequencher software (Gene Codes Corp.)
„Zalinení“ sekvencí – NCBI BLAST, ED program v MUST package (Philippe, 1993), Clustal W v BioEdit 5.0.9 (Hall, 1999)
Fylogenetická analýza – MEGA 6 program (Kumar et al., 2011)
PAUP* 4b10 (Swofford, 2001)

```
A T G C G T C G T T
| | | | | | | | | |
A T G C G T C G T
```

```
A T G - - - C G T C G T T
| | |       | | | |
A T G C G T C G T
```

```
A T G C G T C G T T
| |   | | | |   |
A T C C G T C A T
```

MEGA v6 – úprava sekvencí (alignment)

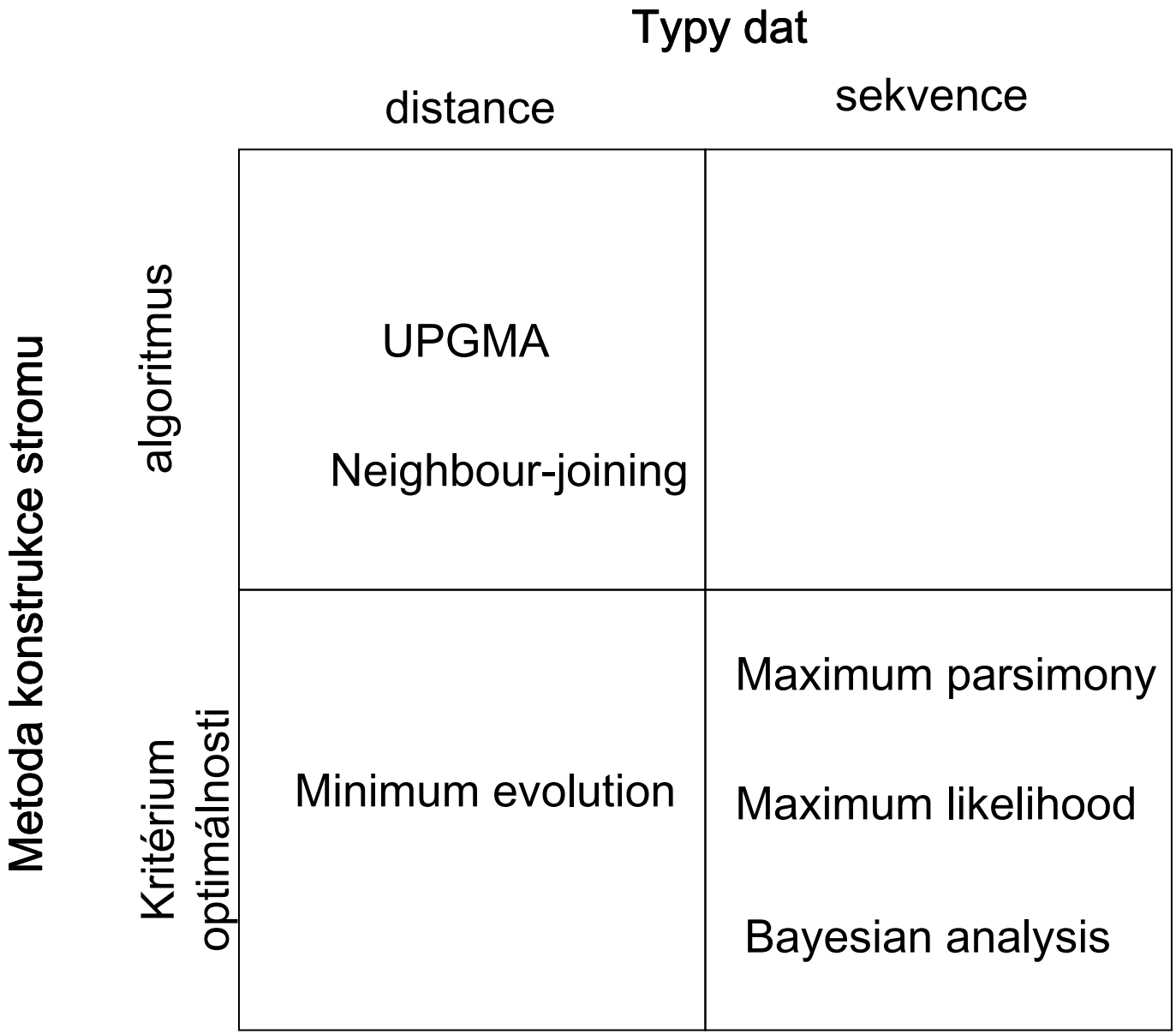
Alignment

Zarovnání sekvencí - dochází k maximalizování reziduí – vytvoření mezer („gaps“) – inzercie a delece vyskytujíce se v sekvenci, kterými se odlišuje od „common ancestor“

Gap Opening Penalty (0-100), Gap Extension Penalty (0-100) – snižování – povolí vložení více „gaps“ – míň nesrovnalostí v alignmentu – falešné shody neukazující homologie, a naopak

Alignment – není věc absolutní, ale najlepší alternativa, kterou algoritmus v počítači vybere

ClustalW, Clustal X, BioEdit…


Typy dat
distance
sekvence
Metoda konstrukce stromu
algoritmus
Kritérium optimálnosti
UPGMA
Neighbour-joining
Minimum evolution
Maximum parsimony
Maximum likelihood
Bayesian analysis

Fylogenetické analýzy

Fylogenetický strom – hypotéza, která vznikla co nejlepším odhadem na základě omezeného zdroje informací

Metody FA – dva přístupy

1. Algoritmus – jde přímo k výsledku, co je jediný strom (odpadá srovnání vzájemně si konkurujících stromů) – metody shlukové analýzy (UPGMA), Neighbour-joining (NJ) – obě využívají data vzdáleností (distance)

2. Kritérium optimálnosti – **dva kroky** – definování kritéria, podle kterého je hodnocen každý strom určitým skóre, které se použije k následnému srovnání všech stromů
   - použití specifického algoritmu pro výpočet funkce (kritérium optimálnosti) a pro získání stromu s nejlepší hodnotou této funkce

Jaká by měla vybraná metoda být?

Výkonnost – „tempus fugit“ nebo „time is money“
pomoc – heuristické metody hledání v případe vyššího počtu taxonů či znaků

Síla – kolik dat musíme shromáždit, aby byly výsledky správné

Konzistence – s přidáváním dalších znaků spějeme k správnému výsledku

Robustnost – do jaké míry vedou drobné odchýlky od vstupných předpokladů k nesprávným závěrům

Falzifikovatelnost – určení nevhodnosti modelu na základě odchýlky od předpokladu

IDEÁLNÍ METODA NEEXISTUJE…

Metoda maximální parsimonie – úspornosti (MP)

Jedna z nejpoužívanějších metod - rychlá, jednoduchá, preferuje jednoduší hypotézy před složitějšími (široká filozofická platnost), tzn. vybere možnost (strom) s minimálním počtem evolučních kroků nutných k vysvětlení vstupních dat
Ne všechny znaky jsou použitelné, parsimony - informative

Metoda maximální parsimonie – úspornosti (MP)

Fitchova a **Wagnerova parsimonie** – nejjednoduší, nezatížené žádnými (F) nebo minimálními (W) omezeními vůči možným typům změn (X → Y, Y → X)

Camin-Sokalova parsimonie – jen X → Y, evoluce je ireverzibilná, neumožňuje ztrátu získaného znaku, moc se nepoužíva

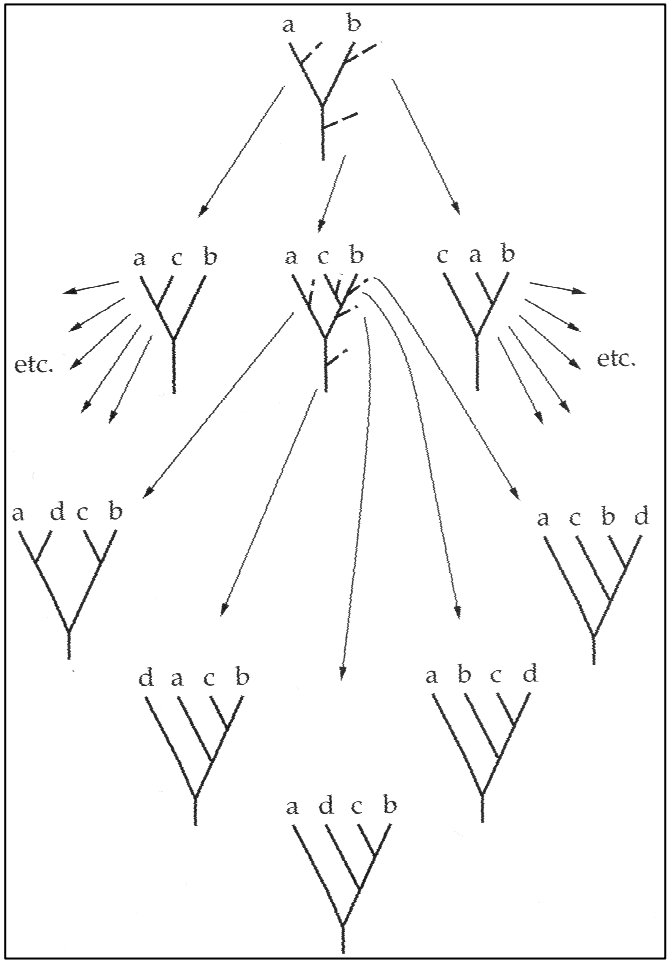
Dollova parsimonie – takisto předpokládá symetričnost změn, ale povoluje znaku vzniknout jen jednou, paralelní a konvergentní získání znaku není povolené

Vážená parsimonie – ne všechny znaky jsou stejně informativní, je subjektivní

Generalizovaná parsimonie – zobecnění uvedených typů, přirazení „costs“ všem možným typům změn

Maximum parsimony

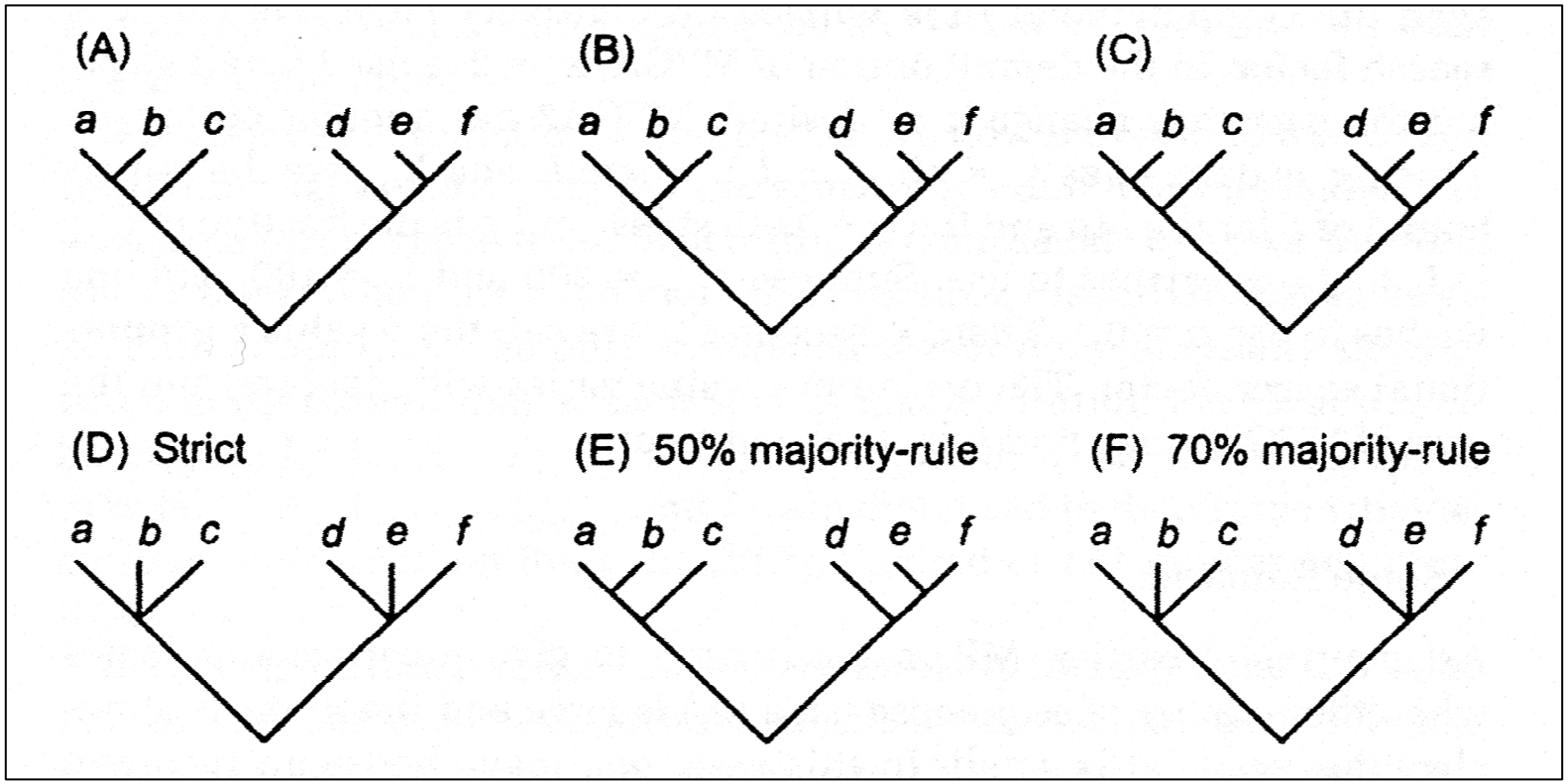
Počet stromů



Species	Number of trees
1	1
2	1
3	3
4	15
5	105
6	945
7	10,395
8	135,135
9	2,027,025
10	34,459,425
11	654,729,075
12	13,749,310,575
13	316,234,143,225
14	7,905,853,580,625
15	213,458,046,676,875
16	6,190,283,353,629,375
17	191,898,783,962,510,625
18	6,332,659,870,762,850,625
19	221,643,095,476,699,771,875
20	8,200,794,532,637,891,559,375
30	4.9518×10^{38}
40	1.00985×10^{57}
50	2.75292×10^{76}

Maximum parsimony

Konsenzuálne stromy



Výhody a nevýhody parsimonie

+

Dobrá pochopitelnost, jednoduchost, rychlost, nízký počet předpokladů (jakákoliv evoluční změna je vzácná, takže MP strom lze považovat za nejlepší odhad skutečné evoluce)

-

Nekonzistentnost, přitažlivost dlouhých větví (LBA)

Metoda maximální pravděpodobnosti (Maximum likelihood, ML)

- posuzují se jednotlivé hypotézy o evoluční historii zkoumaných taxonů z hlediska pravděpodobnosti, že jsou v souladu se získanými daty, výsledek – maximálně pravděpodobný odhad

Tři součásti - vstupné data
evoluční model
fylogenetický strom s topologií i délkou větví

Výhody a nevýhody maximální pravděpodobnosti

\+

Nízká náchylnost k chybě, robustnost vůči odchylkám

\-

Vysoká výpočetná náročnost

Bayesian inference

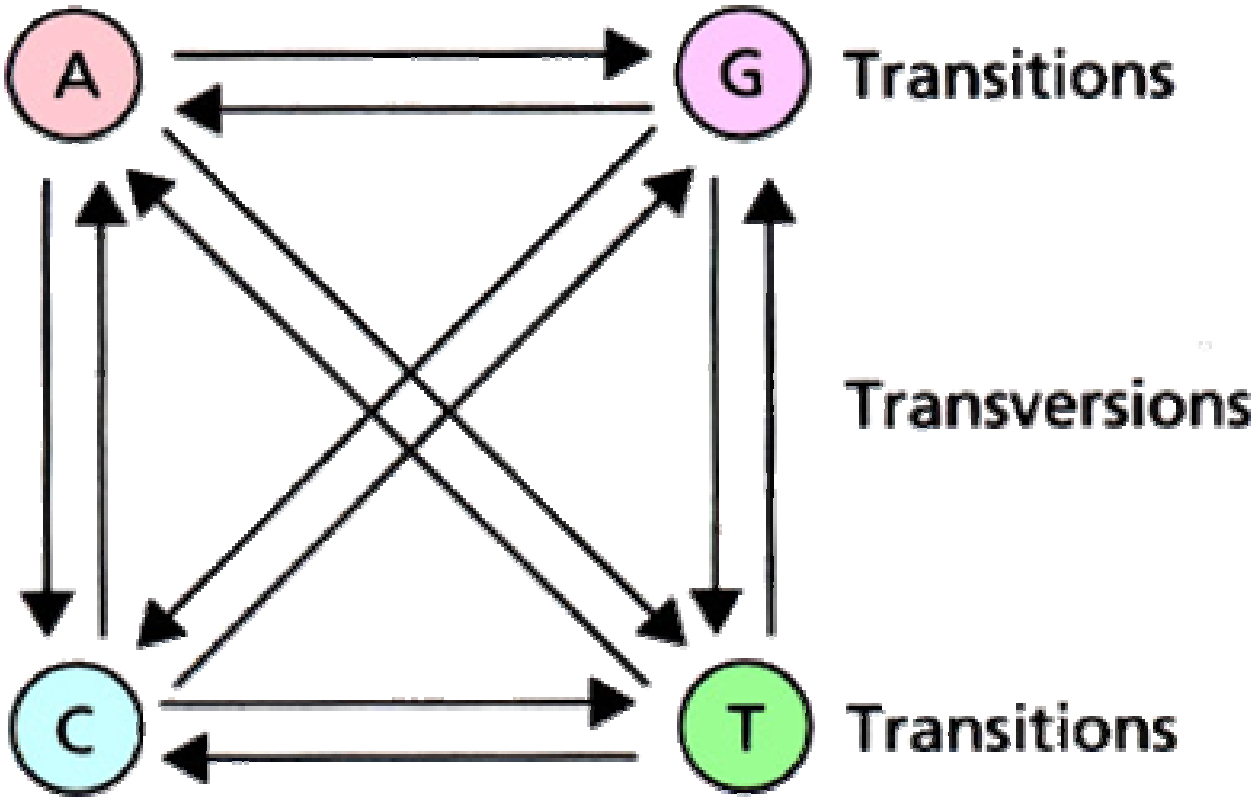
Výpočet pravděpodobnosti na základě specifikovaného modelu a na základě toho, co jsme o charakteru dat zjistili

Základ – strom s danou topologií a délkami větví, model nukleotidových substitucí a rozložení substitučních frekvencí mezi jednotlivými nukleotidy

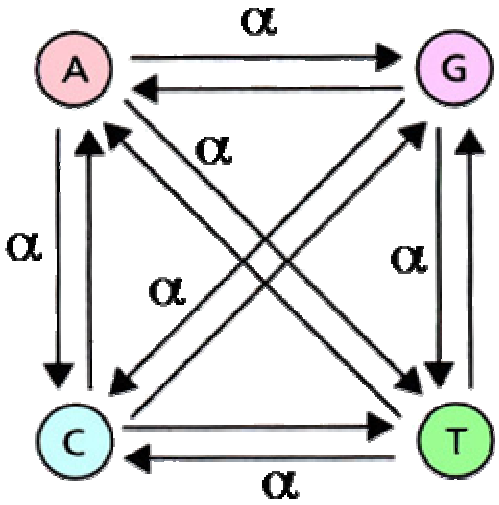
Princip přístupu jako u ML

VÝHODY – menší časová náročnost, strom zohledňující fylogenetický signál v datasetu, možnost použít i pro smíšený dataset

Modely evoluce sekvencí (substitučné modely)



Jukes - Cantor model (JC)



Rate = 3a

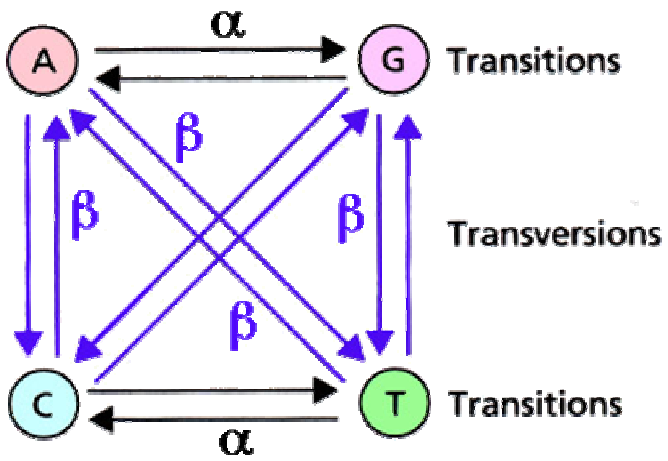
$$K = -\left(\frac{3}{4}\right)\ell n\left(1 - \frac{4p}{3}\right)$$

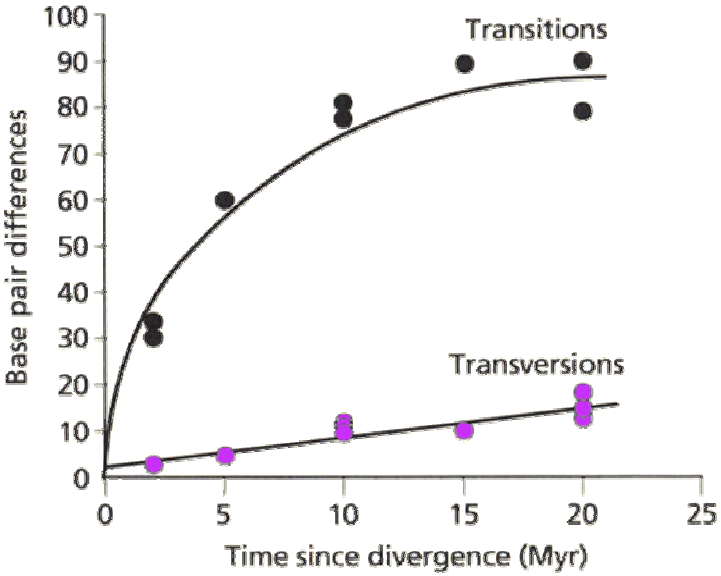
Frekvence jednotlivých bazí jsou totožné a pravděpodobnosti změny jednoho nukleotidu v kterýkoli jiný jsou stejné

$$\mathbf{P}_t = \begin{bmatrix} . & \alpha & \alpha & \alpha \\ \alpha & . & \alpha & \alpha \\ \alpha & \alpha & . & \alpha \\ \alpha & \alpha & \alpha & . \end{bmatrix}, \qquad \mathbf{f} = [\tfrac{1}{4}\ \tfrac{1}{4}\ \tfrac{1}{4}\ \tfrac{1}{4}]$$

Nejjednodušší,
nejméně realistický

Kimura´s 2-parameter model (K2P)





$$\mathbf{P}_t = \begin{bmatrix} . & \beta & \alpha & \beta \\ \beta & . & \beta & \alpha \\ \alpha & \beta & . & \beta \\ \beta & \alpha & \beta & . \end{bmatrix}, \qquad \mathbf{f} = [\tfrac{1}{4}\ \tfrac{1}{4}\ \tfrac{1}{4}\ \tfrac{1}{4}].$$

Rate = a+2β

Frekvence substitucí na 1 nukleotidové místo je 1 tranzice a 2 transverze

Pokud alfa = beta ⇒ K2P=JC

Felsenstein 1981's model (F81)

Některé typy substitucí můžou být častější než jiné proto, že jsou v zkoumaných sekvencích početnejší

Tento model uvažuje nestejné frekvence pro všechny 4 nukleotidy

Jukes-Cantor je speciální případ tohoto modelu, kdy mají všechny nukleotidy stejnou frekvenci

$$\mathbf{P}_t = \begin{bmatrix} . & \pi_C\alpha & \pi_G\alpha & \pi_T\alpha \\ \pi_A\alpha & . & \pi_G\alpha & \pi_T\alpha \\ \pi_A\alpha & \pi_C\alpha & . & \pi_T\alpha \\ \pi_A\alpha & \pi_C\alpha & \pi_G\alpha & . \end{bmatrix}, \qquad \mathbf{f} = [\pi_A\ \pi_C\ \pi_G\ \pi_T]$$

Pí je průměrná frekvence baze *i* v porovnávaných sekvencích

Pokud pA=pC=pG=pT , pak F81 = JC

Hasegawa, Kishino, Yano 1985 model (HKY85)

Spojuje vlastnosti obou předchozích modelů (K2P a F81)

Bere v potaz nestejné zastoupení jednotlivých bazí a rozdílnou frekvenci tranzicí a transverzí

$$\mathbf{P}_t = \begin{bmatrix} . & \pi_C\beta & \pi_G\alpha & \pi_T\beta \\ \pi_A\beta & . & \pi_G\beta & \pi_T\alpha \\ \pi_A\alpha & \pi_C\beta & . & \pi_T\beta \\ \pi_A\beta & \pi_C\alpha & \pi_G\beta & . \end{bmatrix}, \qquad \mathbf{f} = [\pi_A\ \pi_C\ \pi_G\ \pi_T]$$

General time-reversible model (GTR)

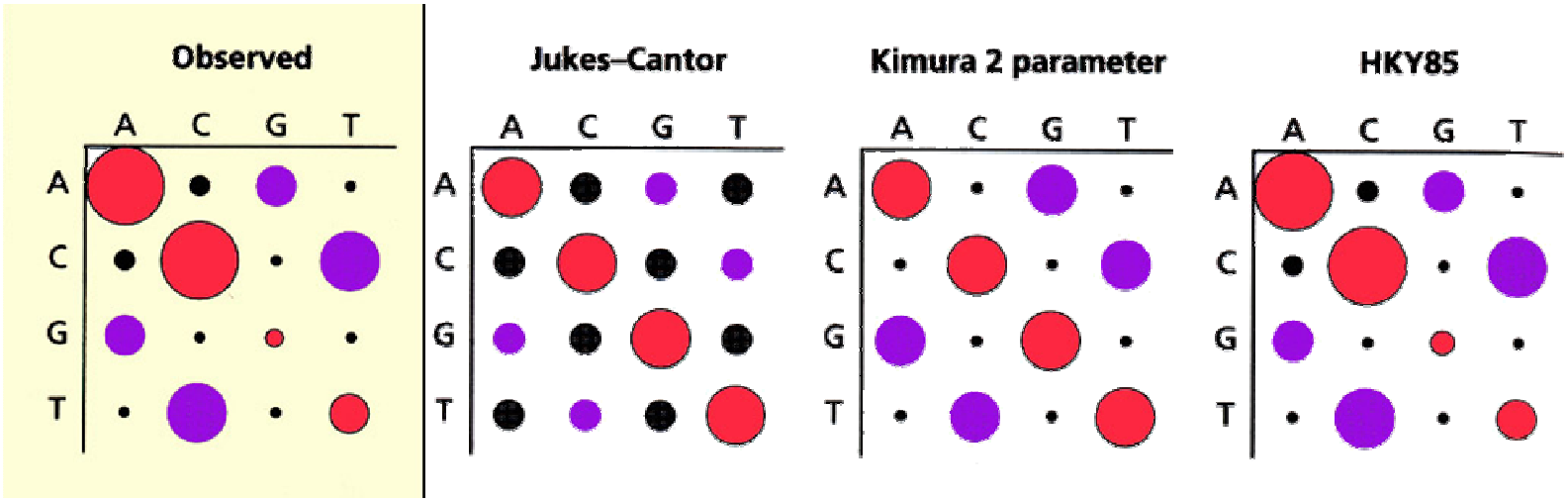
Nejobecnější model, všech 6 typů substitucí má rozdílnou frekvenci

$$\mathbf{P}_t = \begin{bmatrix} . & \pi_C a & \pi_G b & \pi_T c \\ \pi_A a & . & \pi_G d & \pi_T e \\ \pi_A b & \pi_C d & . & \pi_T f \\ \pi_A c & \pi_C e & \pi_G f & . \end{bmatrix}, \qquad \mathbf{f} = [\pi_A \; \pi_C \; \pi_G \; \pi_T]$$

Omezením některého z parametrů se můžeme dopracovat ke kterémukoliv z předešlých modelů

Skutečné data: pozorované a očekávané změny

- Srovnání lidské a šimpanzí mtDNA (307/1333 bp je rozdílných)
- K2P odhaduje P=0.22, Q=0.011
- HKY85 odhaduje A=0.37, T=0.18, C=0.40, G=0.05
- Modely bohatší na parametry jsou blíže skutečnému stavu



Jak vybrat správný model…

Více parametrů, více realizmu, ale…

Více přidaných parametrů (naředení dat) – zvyšujeme nejistotu odhadu - zvýšení chyby výběru (sampling error)

Málo parametrů – nepřesné odhady

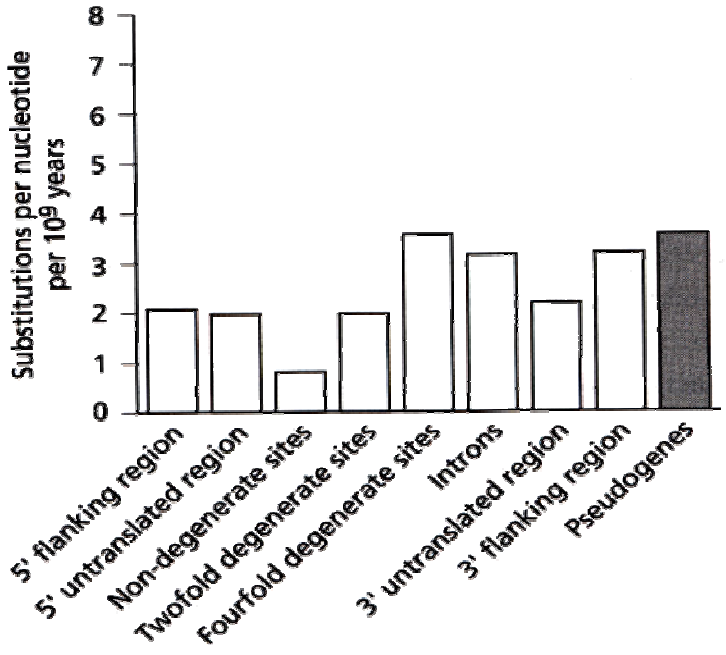
Hodně parametrů – nízká presnost

Heterogenita substitučních frekvencí v různých částech sekvence

Uvedené modely předpokládají, že každé nukleotidové místo se vyvíjí stejnou rychlostí

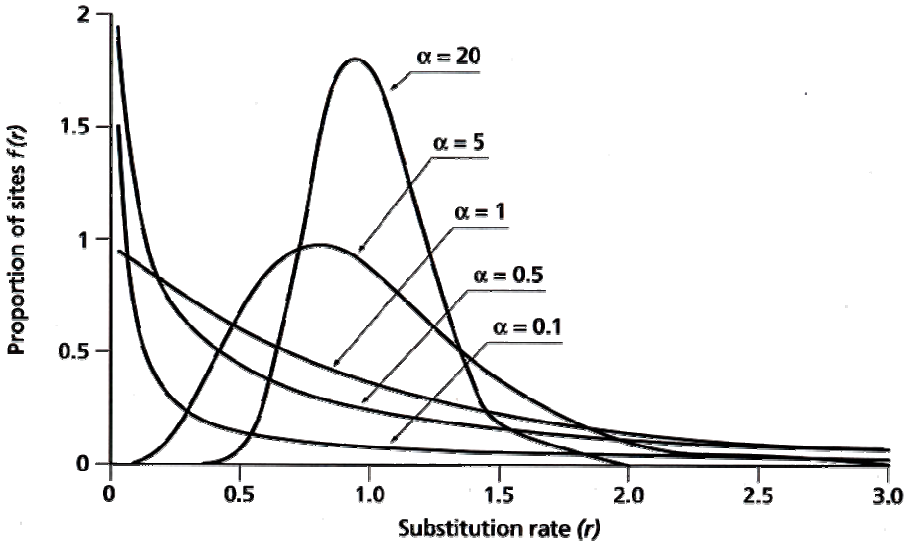
Heterogenitu subst. frekvencí je možné do ML analýzy začlenit:

Gamma distribuce – rozsah, tvar



Gamma distribuce

Umožňuje víc jako dvě kategorie



Type of sequences	α
Nuclear genes	
Albumin genes	1.05
Insulin genes	0.40
c-myc genes	0.47
Prolactin genes	1.37
16S-like rRNAs, stem region	0.29
16S-like rRNAs, loop region	0.58
ψη-globin pseudogenes	0.66
Viral genes	
Hepatitis B virus genomes	0.26
Mitochondrial genes	
12S rRNAs	0.16
Position 1 of four genes	0.18
Position 2 of four genes	0.08
Position 3 of four genes	1.58
D-loop region	0.17
Cytochrome *b*	0.44

Jak vybrat parametry modelu?

Alternativa 1: odhadnout přímo z dat při vyhodnocení každého stromu (časově náročné)

Alternativa 2: najít “vhodný strom” a odhadnout parametry na základě tohoto stromu

Alternativa 3: opakovat rekonstrukci ML stromů pomocí parametrů odhadovaných v kroku 2, kým se nic nebude měnit

Nejjednodušší způsob je použít Modeltest 3.8 (online) nebo MrModeltest

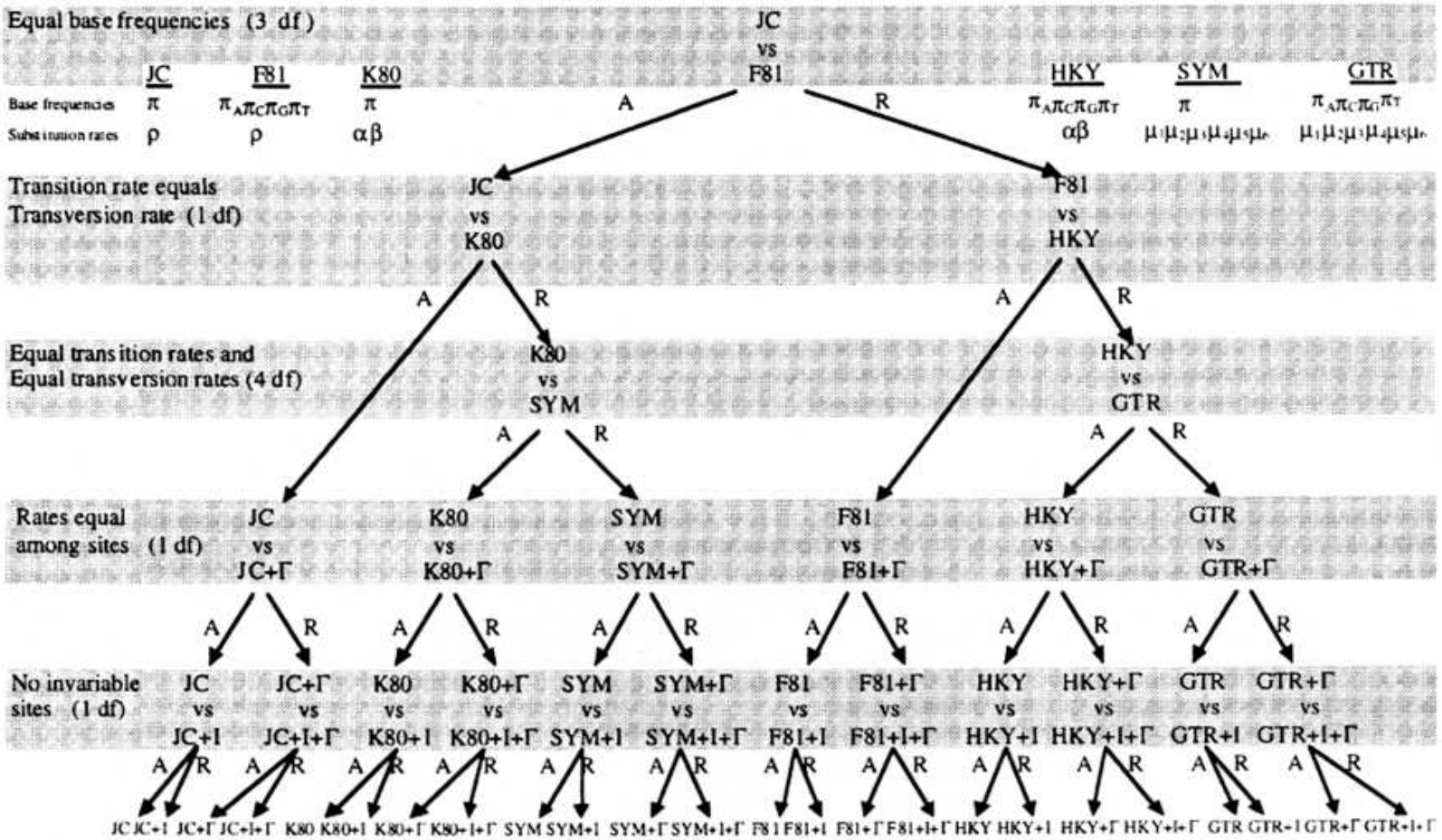


Fig. 1. Hierarchical hypothesis testing in MODELTEST. At each level the null hypothesis (upper model) is either accepted (A) or rejected (R). The models of DNA substitution are: JC (Jukes and Cantor, 1969), K80 (Kimura, 1980), SYM (Zharkikh, 1994), F81 (Felsenstein, 1981), HKY (Hasegawa *et al.*, 1985), and GTR (Rodríguez *et al.*, 1990). Γ: shape parameter of the gamma distribution; I: proportion of invariable sites. df: degrees of freedom. 1: equal base frequencies (0.25), π_A: frequency of adenine, π_C: frequency of cytosine, π_G: frequency of guanine, π_T: frequency of thymine. ρ: equal substitution rate, α: transition rate, β: transversion rate; μ_1: A⇒C rate, μ_2: A⇒G rate, μ_3: A⇒T rate, μ_4: C⇒G rate, μ_5: C⇒T rate, μ_6: G⇒T rate.

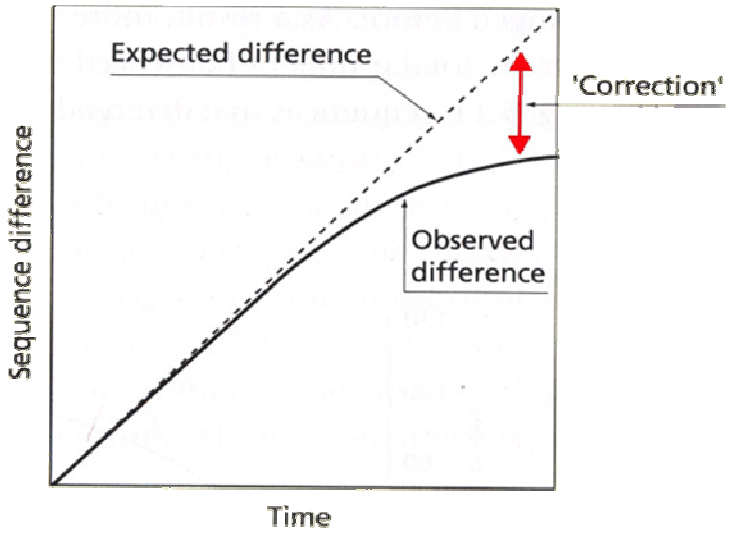
Distanční metody

Založené na podobnostech (vzdálenostech, rozdílech)

Poznání skutečné evoluční vzdálenosti mezi všemi členy studovaného souboru taxonů umožňuje velmi lehkou rekonstrukci evoluční historie těchto taxonů

Opakované změny jednoho znaku – korigované distance (jako u pravděpodobnosti)

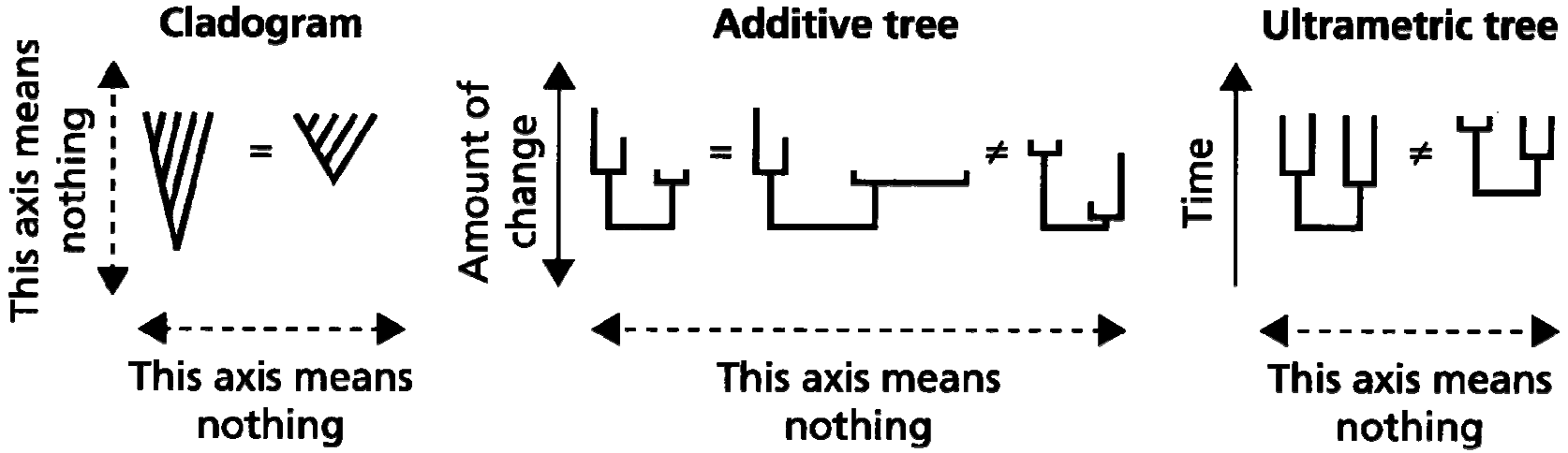
Nekorigovaná vzdálenost – p-distance
Korekce: JC, F81, K2P, F84, GTR



Aditivní a ultrametrické stromy

Aditivita – evoluční vzdálenost mezi kteroukoliv dvojicí taxonů je součtem délek všech větví, které je spojují

Ultrametrické distance – jsou také aditivní, všechny taxony se vyvíjí v souladu s předpoklady molekulárních hodin



Neighbour - joining (NJ) Saitou et Nei, 1987

Často používaná metoda k rekonstrukci stromů (barcoding)

Data nejsou ultrametrická, pracuje s uzlama, ne taxonama

Kombinuje rychlosť a jedinečnost výsledku – jediný strom

Nezohledňuje shodu mezi daty a výsledným stromem (je to víc klastrová metoda než optimality)

Dobrá heuristická metoda pro konstrukci „minimum evolution“ stromu

ME - Minimalizuje sumu délek větví, která je počítána z „pairwise distances“

Optimální stromy a jejich spolehlivost

Exaktní přístup – Exhaustive search – porovnávání všech možných stromů

Branch-and-Bound – „přidej větev, stanov limit“ - kritérium, které se snažíme minimalizovat, může byť jakékoli, a s přidáváním taxonů neklesá – nemusíme hodnotit všechny možné stromy, ale stanovíme horní hranici kritéria, kterou žádny strom nesmí překročit

Heuristický přístup – Stepwise addition – postupné přidávání taxonů
Star decomposition – hvězdicová dekompozice
Branch Swapping – výměna větví

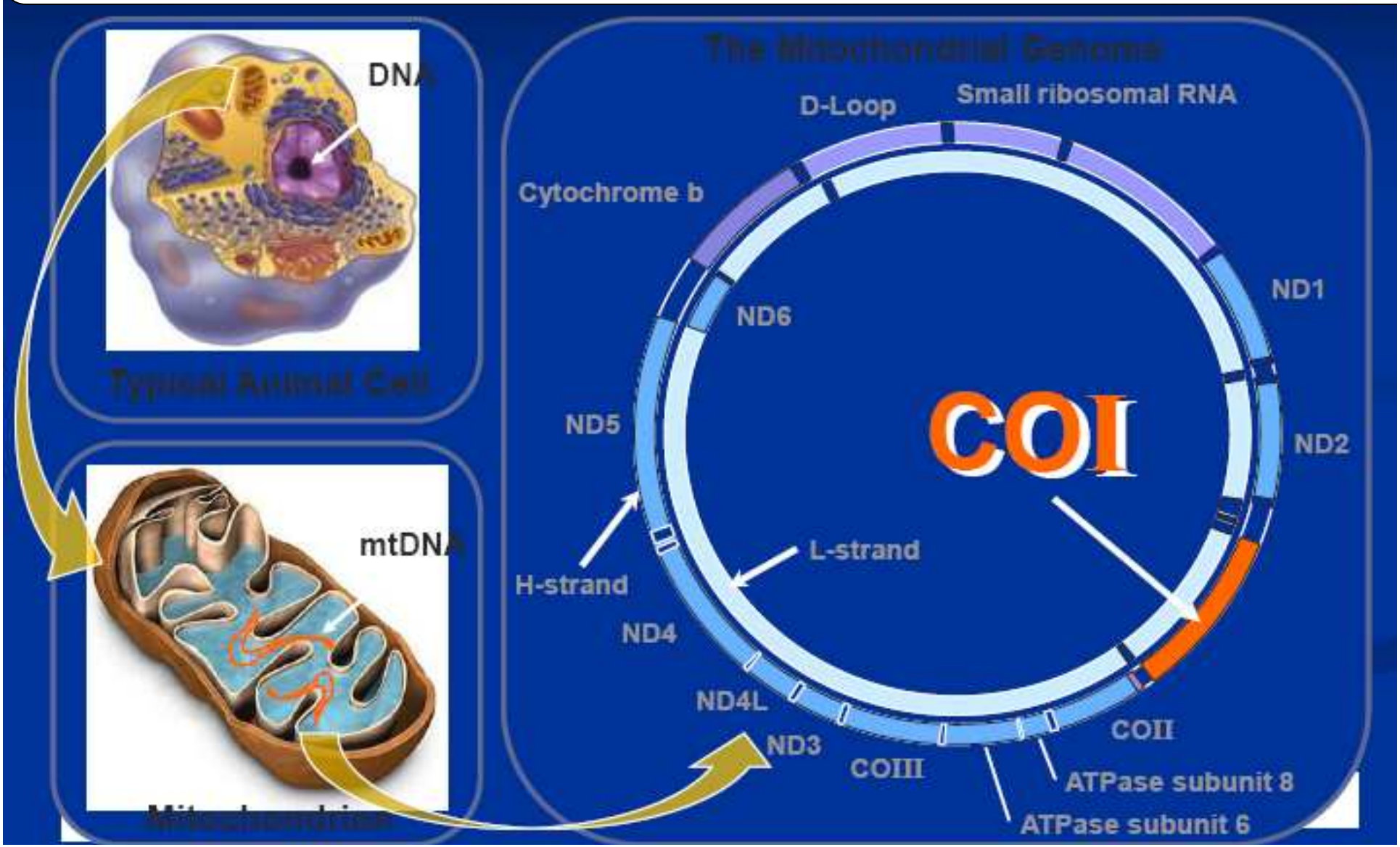
Test hierarchické struktury - randomizace

Test spolehlivosti jednotlivých větví – Bootstrapping
Jack-knife → konsenzuální strom

Systém čárového kódu

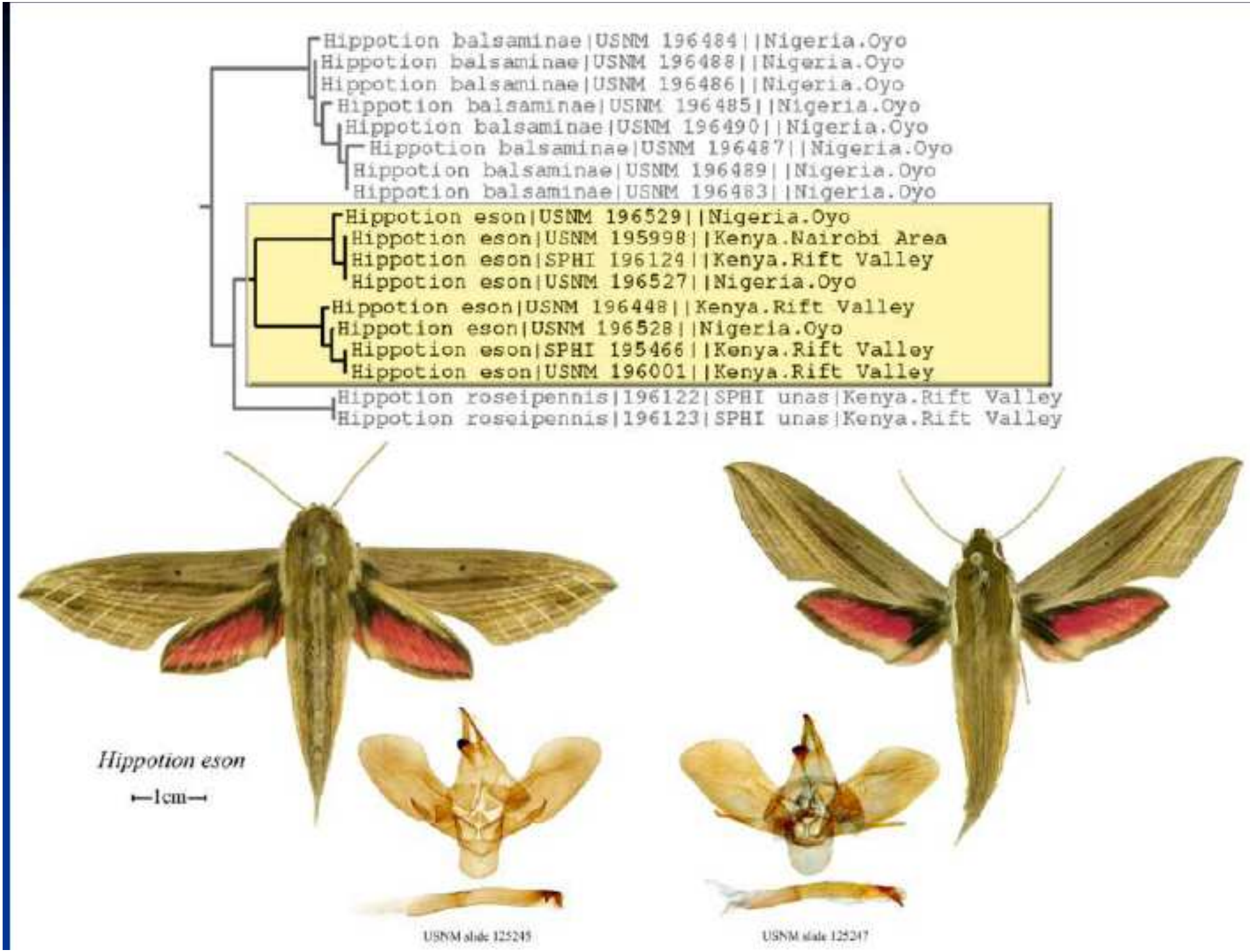
- DNA barcode je krátká genová sekvence standardizované části genomu použitá k druhové identifikaci

Interné ID pro všechny organizmy na Zem



Příklad: Afričtí motýli

Tvoří species komplexy
Jejich parazitoidi (Tachinidae) také
(Dittrich et al 2006)
Leguminivora ptychora na luštěninách je také species komplex
Ale někteří škůdci jsou široko rozšíření, např. *Spoladea recurvalis* a *Maruca vitrata*


Hippotion balsaminae|USNM 196484||Nigeria.Oyo
Hippotion balsaminae|USNM 196488||Nigeria.Oyo
Hippotion balsaminae|USNM 196486||Nigeria.Oyo
Hippotion balsaminae|USNM 196485||Nigeria.Oyo
Hippotion balsaminae|USNM 196490||Nigeria.Oyo
Hippotion balsaminae|USNM 196487||Nigeria.Oyo
Hippotion balsaminae|USNM 196489||Nigeria.Oyo
Hippotion balsaminae|USNM 196483||Nigeria.Oyo
Hippotion eson|USNM 196529||Nigeria.Oyo
Hippotion eson|USNM 195998||Kenya.Nairobi Area
Hippotion eson|SPHI 196124||Kenya.Rift Valley
Hippotion eson|USNM 196527||Nigeria.Oyo
Hippotion eson|USNM 196448||Kenya.Rift Valley
Hippotion eson|USNM 196528||Nigeria.Oyo
Hippotion eson|SPHI 195466||Kenya.Rift Valley
Hippotion eson|USNM 196001||Kenya.Rift Valley
Hippotion roseipennis|196122|SPHI unas|Kenya.Rift Valley
Hippotion roseipennis|196123|SPHI unas|Kenya.Rift Valley
Hippotion eson
←1cm→
USNM slide 125245
USNM slide 125247

Rychlé a efektivní

- Čeleď Sphingidae – vzorky 49 druhů za 6 měsíců (téměř kompletní lokální fauna)
- DNA barcoding rozlišil druhy jak v lokálním, tak v globálním měřítku
- Místní knihovny můžou být rychle srovnány a přispět ke globálním knihovnám

Jak Barcoding funguje

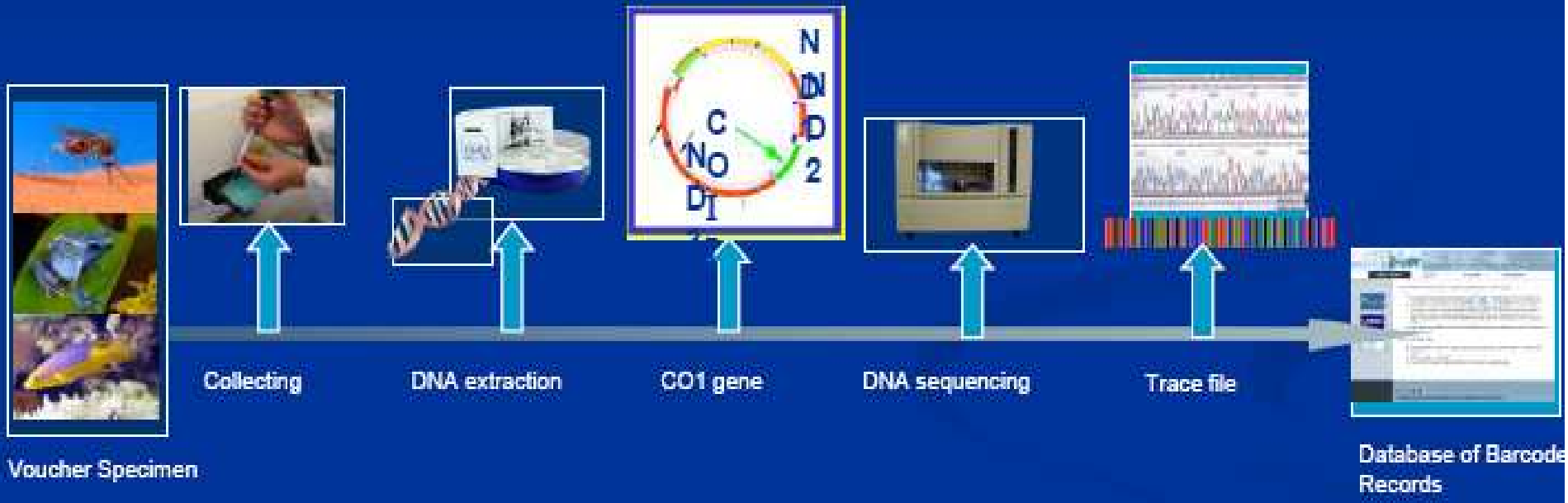
Tvorba referenční knihovny:

- Správně určený jedinec (vouchers)
- Vzorek tkáně
- DNA extrakce, PCR amplifikace
- DNA sekvenování
- Odeslání dat do GenBanku

Použití referenční knihovny :

- Neurčené druhy
- Tkáň, DNA, sekvenování
- Srovnání s referenčními sekvencemi

Jak se to vše děje od jedince přes sekvenci po druh?



Produkce dat v r. 2007

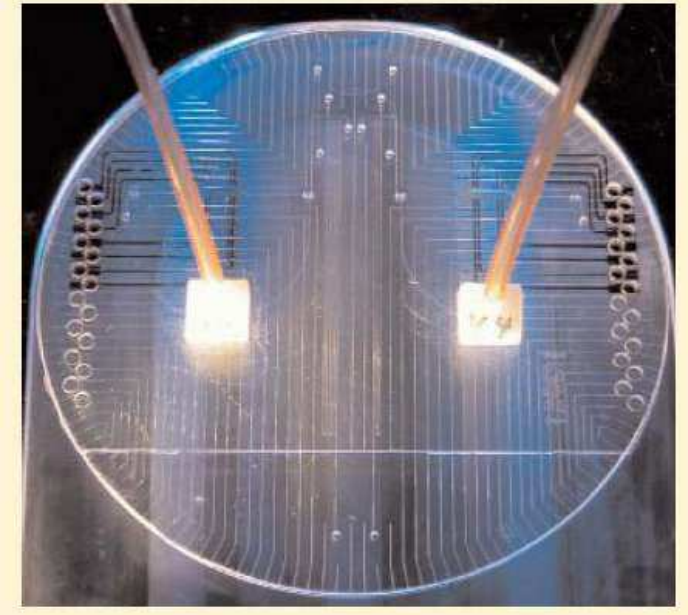
PCR amplifikační jednotka

ABI 3100 sekvenátor

Stovky vzorků denně, cena od několika centů po dolary

Produkce dat v r. 2008

Rychlejší a přenosnější systém – stovky vzorků za hodinu



Integrované DNA mikročipy

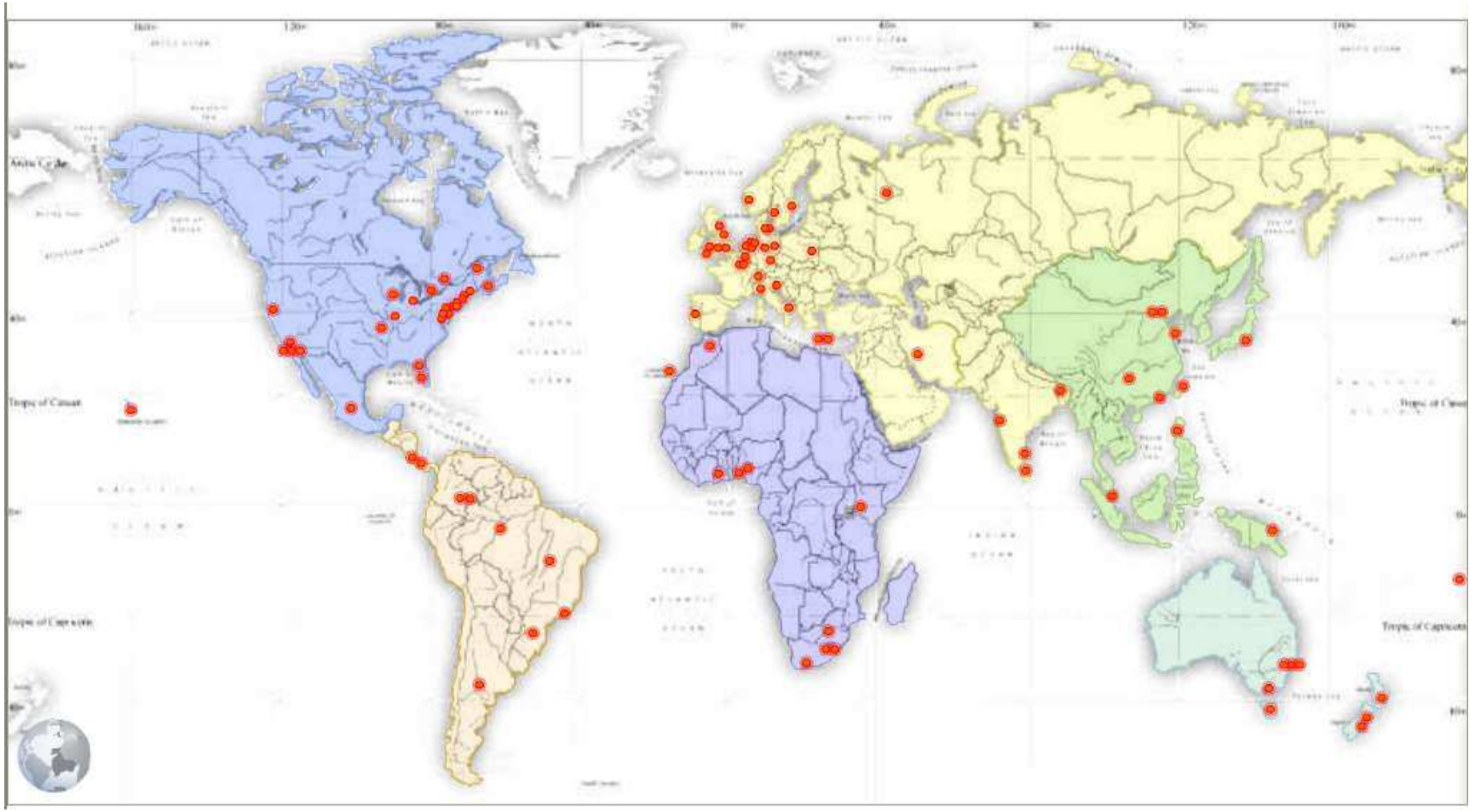
Stolní mikrofluidné systémy

Produkce dat v budoucnosti?

- Získání dat kdekoli, hned
- Cena několik haléřů
- Link do referenční databáze
- Taxonomická GPS
- Použitelné nespecialisty

CBOL – organizace členů od r. 2008

Více než 170 organizací z více než 50 zemí (z toho 54 organizací z 20 rozvojových zemí)

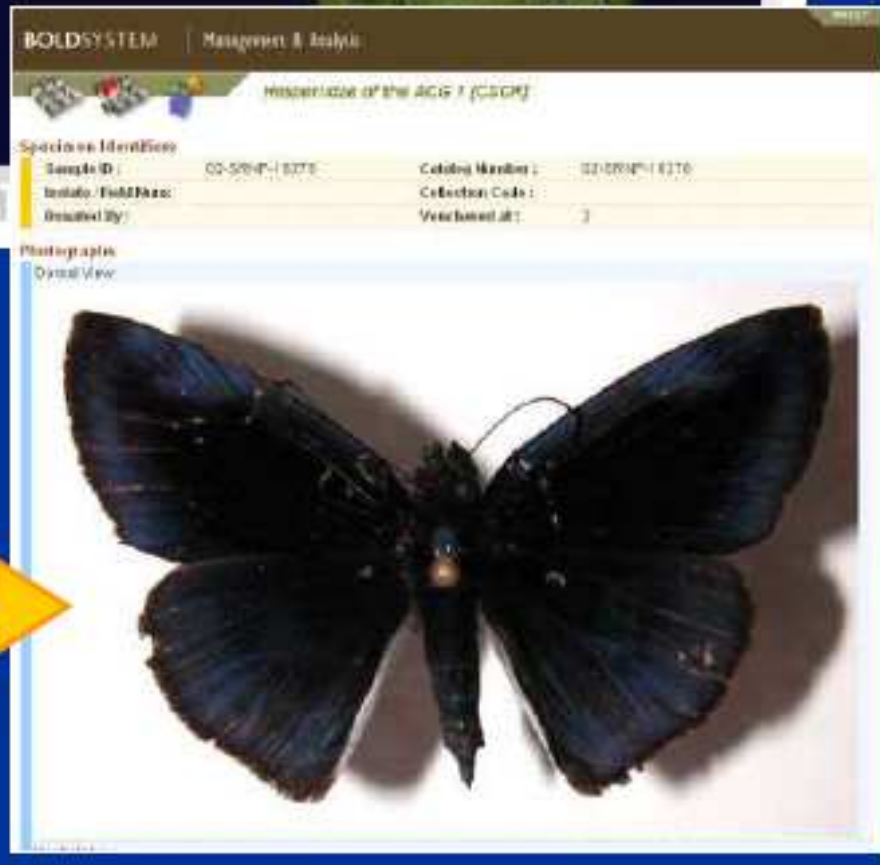


Mise CBOLu: uvést DNA Barcoding jako globální standard

1. Vyvinout a zvednout standardy komunity
2. Barcode projekty plnit databáze
3. Globalní participace a koordinace
4. Přijetí taxonomickou komunitou
5. Koordinace s jinými oblastmi vědy
6. Přijetí regulačními agenturami
7. Vyvíjení produktů soukromými společnostmi

Propojení GenBanku s vouchery

BOLD Specimen Webpage



BOLD Sequence Webpage

BOLDSYSTEM | Management & Analysis

Hesperiidae of the ACG 1 [CSCR]

Barcode Identifiers

Barcode ID :	CSCR010-04	Sample ID :	
Gene :	COX1	GenBank Accessi	
Last Updated :		Translation Matrix	

Sequencing Runs

Run Date	Run Site	Direction	Trace File

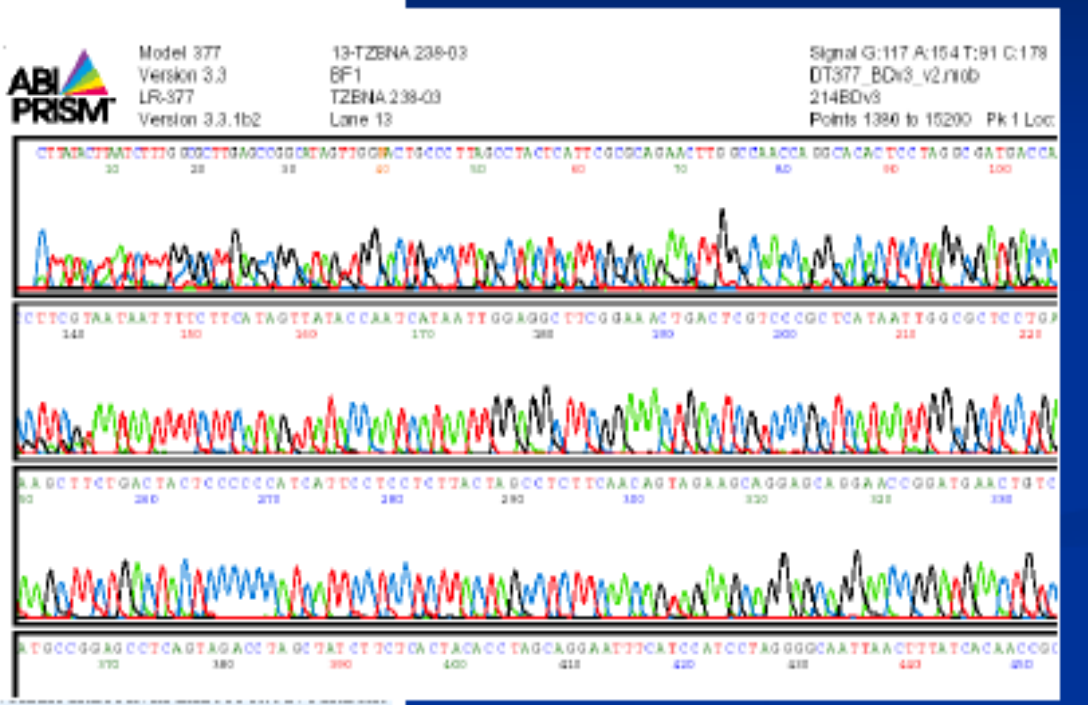
Nucleotide Sequence

Length :	617
Comp. A :	203
Comp. G :	85
Comp. C :	92
Comp. T :	237
Updated :	2005-09-09

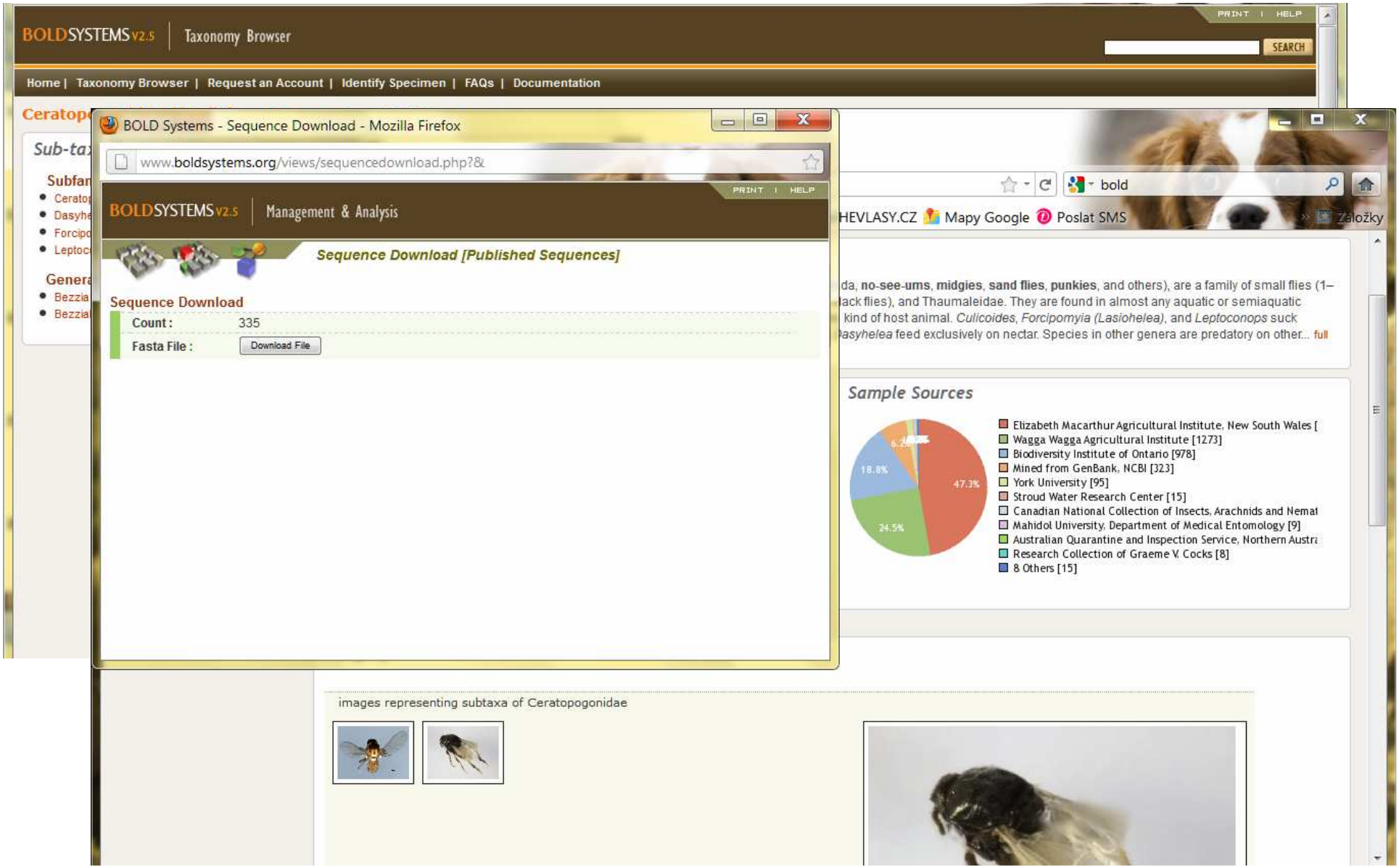
Amino Acid Sequence

Length : 220

Illustrative Barcode



Ceratopogonidae – 105 barcoding sekvencí



Globální projekty CBOL

- Fish Barcode of Life (FISH-BOL) - 30 000 mořských/sladkovodných druhů do r. 2010
- All Birds Barcoding Initiative (ABBI) - 10 000 druhů do r. 2010
- Tephritidae – 2 000 škůdců/prospěšných druhů do r. 2008
- Komáry - 3 300 druhů do r. 2008
- Ohrožené druhy
- Trees of the world

Staré a nové techniky

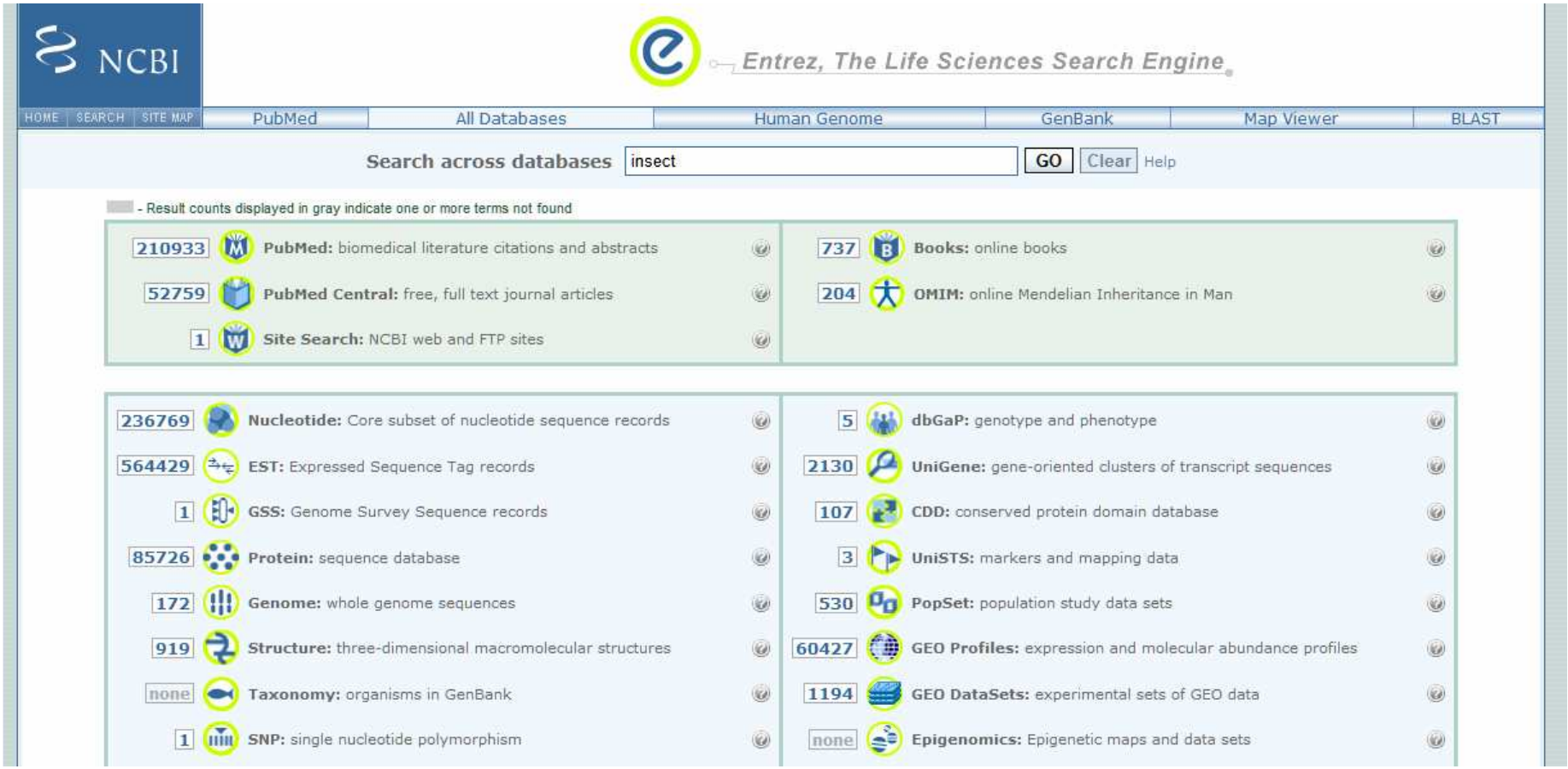
- CBOL staví na současných taxonomických poznatcích
- Sequence knihovny založeny na voucher jedincích, co dělá vědu opakovatelnou a testovatelnou
- Voucher jedince propojují historické, současné a budoucí výzkum

Příklad: CSIRO studie na bzučivkách mapující rezistence na insekticidy a zjišťování historie pomocí DNA z muzejních jedinců (PNAS 103: 8757)

GenBank

- http://www.ncbi.nlm.nih.gov/genbank/
- Několik databází – Nucleotide, Protein, PubMed, CoreNucleotide, Structure, Genome, etc.
- Věrohodnost sekvencí vyšší než v databázích CBOLu
- Součástí je BLAST - „multialign tool“

Po zadání hesla – Insect…



Musca domestica cytochrome oxidase subunit II (COII) gene, partial cds; mitochondrial

GenBank: DQ133110.1

FASTA Graphics PopSet

Go to:

```
LOCUS       DQ133110                663 bp    DNA     linear   INV 13-DEC-2005
DEFINITION  Musca domestica cytochrome oxidase subunit II (COII) gene, partial
            cds; mitochondrial.
ACCESSION   DQ133110
VERSION     DQ133110.1  GI:72398995
KEYWORDS    .
SOURCE      mitochondrion Musca domestica (house fly)
  ORGANISM  Musca domestica
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Muscoidea; Muscidae; Musca.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   Dittmar,K., Porter,M.L., Murray,S. and Whiting,M.F.
  TITLE     Molecular phylogenetic analysis of nycteribiid and streblid bat
            flies (Diptera: Brachycera, Calyptratae): implications for host
            associations and phylogeographic origins
  JOURNAL   Mol. Phylogenet. Evol. 38 (1), 155-170 (2006)
   PUBMED   16087354
REFERENCE   2  (bases 1 to 663)
  AUTHORS   Dittmar,K., Porter,M.L., Murray,S. and Whiting,M.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2005) Integrative Biology, Brigham Young
            University, 401 WIDB, Provo, UT 84602, USA
FEATURES             Location/Qualifiers
     source          1..663
                     /organism="Musca domestica"
                     /organelle="mitochondrion"
                     /mol_type="genomic DNA"
                     /specimen_voucher="Di180"
                     /db_xref="taxon:7370"
     gene            <1..>663
                     /gene="COII"
     CDS             <1..>663
                     /gene="COII"
                     /codon_start=1
                     /transl_table=5
                     /product="cytochrome oxidase subunit II"
                     /protein_id="AAZ72903.1"
                     /db_xref="GI:72398996"
                     /translation="MSTWANLGLQDSSSPLMEQLIFFHDHALMILVMITVLVGYLMFT
                     LFFNKYVNRYLLHGQTIEIIWTILPAIILLFIAFPSLRLLYLLDEINEPSVTLKAIGH
                     QWYWSYEYSDFNNVEFDSYMIPTNELPVDGFRLLDVDNRVVLPMNSQIRILVTAADVI
                     HSWTVPALGVKVDGTPGRLNQTNFLINRPGLFYGQCSEICGANHSFMPIVIESIPVNY
                     FIK"
ORIGIN
        1 atgtcaacat gagcaaattt aggtttacaa gatagttctt ctccattaat agaacaatta
       61 attttttttc atgatcatgc attaataatt ttagtaataa ttacagtatt agtcggatat
      121 ttaatgttta cattattttt taataaatat gttaatcgtt atttattaca tggacaaaca
      181 attgaaatta tttgaactat tttacctgca attattttat tattcattgc tttcccttct
      241 ttacgattat tatacttatt agatgaaatt aatgaaccat cagtaacttt aaaggctatt
      301 ggtcatcaat gatattgaag ttatgaatat tcagatttta ataatgttga atttgattct
```

Customize view

Analyze this sequence

Run BLAST
Pick Primers
Find in this Sequence

Related information

Related Sequences
PopSet
Protein
PubMed
Taxonomy

Recent activity

Turn Off Clear

Musca domestica cytochrome oxidase subunit II (COII) gene, partial cds; mitochon... Nucleotide

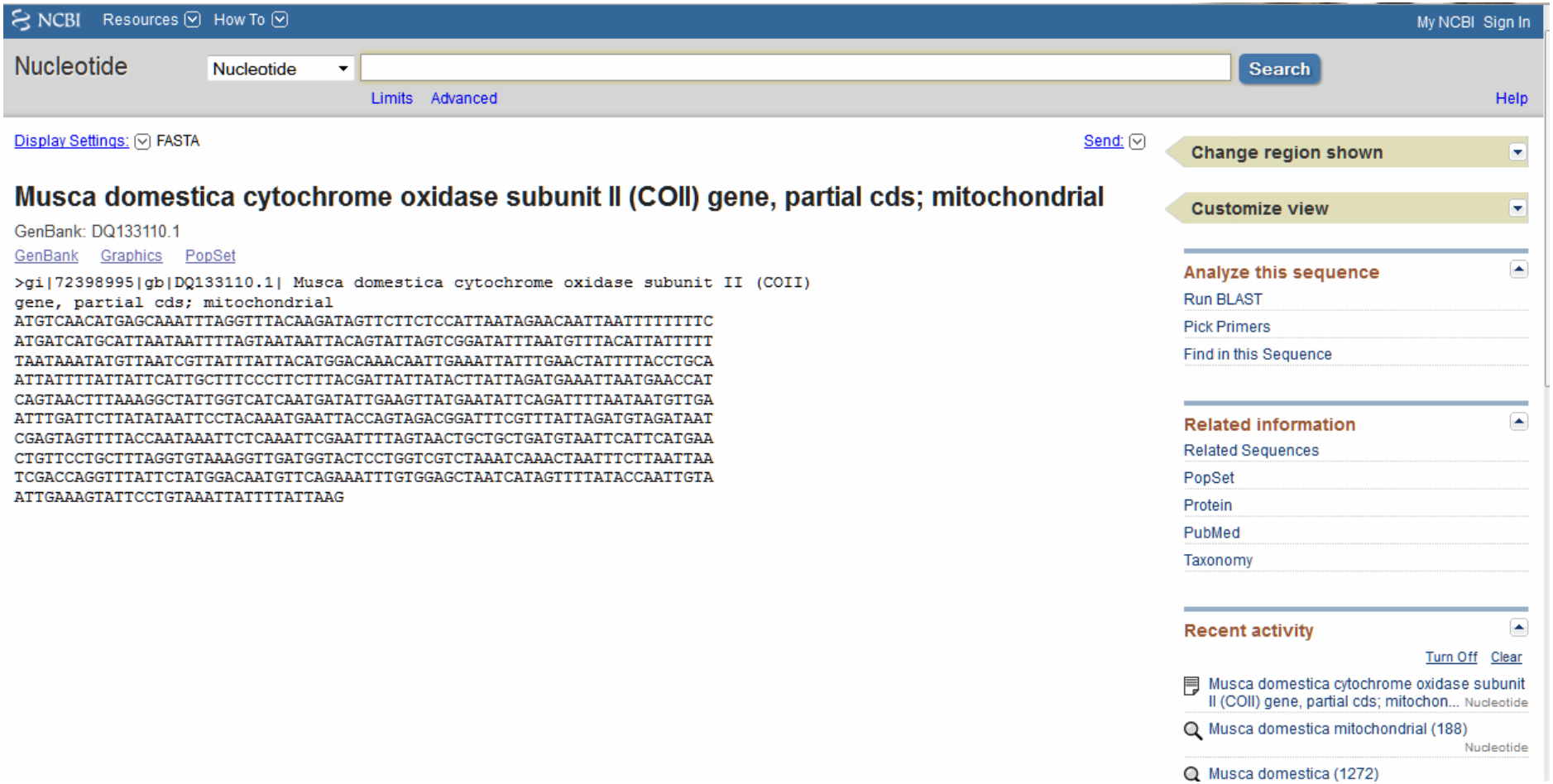
~~Musca domestica mitochondrial (188)~~ Nucleotide

Musca domestica (1272) Nucleotide

insect (236769) Nucleotide

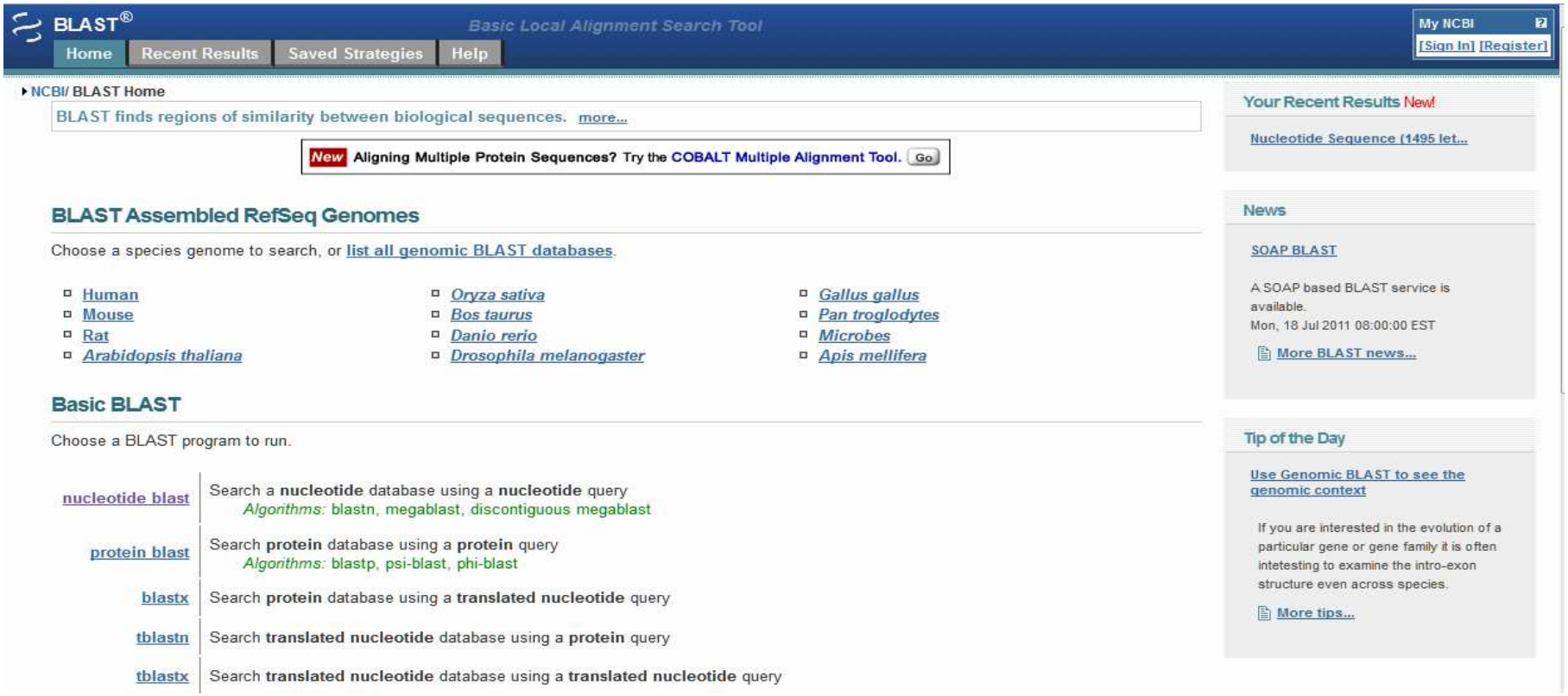
Nucleotide Sequence (1495 letters) BLAST

See more...


NCBI Resources How To
My NCBI Sign In
Nucleotide
Nucleotide
Search
Limits Advanced
Help
Display Settings: FASTA
Send:
Musca domestica cytochrome oxidase subunit II (COII) gene, partial cds; mitochondrial
GenBank: DQ133110.1
GenBank Graphics PopSet
>gi|72398995|gb|DQ133110.1| Musca domestica cytochrome oxidase subunit II (COII)
gene, partial cds; mitochondrial
ATGTCAACATGAGCAAATTTAGGTTTACAAGATAGTTCTTCTCCATTAATAGAACAATTAATTTTTTTTC
ATGATCATGCATTAATAATTTTAGTAATAATTACAGTATTAGTCGGATATTTAATGTTTACATTATTTTT
TAATAAATATGTTAATCGTTATTTATTACATGGACAAACAATTGAAATTATTTGAACTATTTTACCTGCA
ATTATTTTATTATTCATTGCTTTCCCTTCTTTACGATTATTATACTTATTAGATGAAATTAATGAACCAT
CAGTAACTTTAAAGGCTATTGGTCATCAATGATATTGAAGTTATGAATATTCAGATTTTAATAATGTTGA
ATTTGATTCTTATATAATTCCTACAAATGAATTACCAGTAGACGGATTTCGTTTATTAGATGTAGATAAT
CGAGTAGTTTTACCAATAAATTCTCAAATTCGAATTTTAGTAACTGCTGCTGATGTAATTCATTCATGAA
CTGTTCCTGCTTTAGGTGTAAAGGTTGATGGTACTCCTGGTCGTCTAAATCAAACTAATTTCTTAATTAA
TCGACCAGGTTTATTCTATGGACAATGTTCAGAAATTTGTGGAGCTAATCATAGTTTTATACCAATTGTA
ATTGAAAGTATTCCTGTAAATTATTTTATTAAG
Change region shown
Customize view
Analyze this sequence
Run BLAST
Pick Primers
Find in this Sequence
Related information
Related Sequences
PopSet
Protein
PubMed
Taxonomy
Recent activity
Turn Off Clear
Musca domestica cytochrome oxidase subunit II (COII) gene, partial cds; mitochon... Nucleotide
Musca domestica mitochondrial (188) Nucleotide
Musca domestica (1272)

BLAST

http://blast.ncbi.nlm.nih.gov/Blast.cgi



Other reports: Search Summary [Taxonomy reports] [Distance tree of results]

Graphic Summary

Distribution of 100 Blast Hits on the Query Sequence

Descriptions

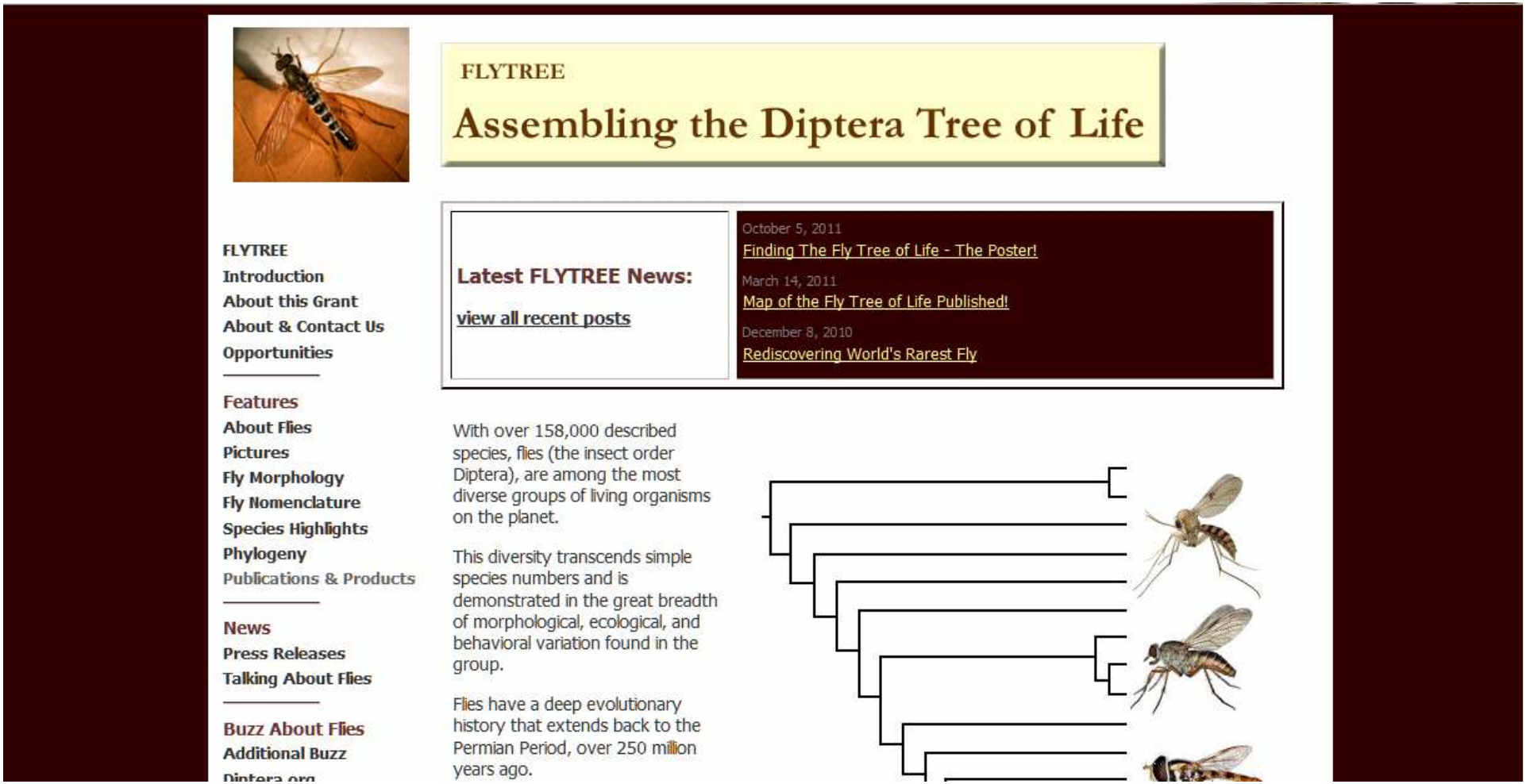
Legend for links to other resources: UniGene | GEO | Gene | Structure | Map Viewer | PubChem BioAssay

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
GQ465784.1	Musca domestica cytochrome oxidase subunit I (COI) gene, partial c	2761	2761	100%	0.0	100%	
AB479529.1	Musca domestica mitochondrial COI gene for cytochrome oxidase sub	2750	2750	100%	0.0	99%	
AB479528.1	Musca domestica mitochondrial COI gene for cytochrome oxidase sub	2750	2750	100%	0.0	99%	
AY526196.1	Musca domestica tRNA-Tyr gene, partial sequence; and cytochrome	2750	2750	100%	0.0	99%	
EU154477.1	Musca domestica mitochondrion, partial genome	2739	2739	100%	0.0	99%	
EU627696.1	Musca conducens cytochrome oxidase subunit I (COI) gene, partial c	2244	2244	99%	0.0	93%	
EU627694.1	Musca asiatica cytochrome oxidase subunit I (COI) gene, partial cds;	2222	2222	99%	0.0	93%	
EU627693.1	Musca sorbens cytochrome oxidase subunit I (COI) gene, partial cds;	2222	2222	99%	0.0	93%	
EU627700.1	Musca larvipara cytochrome oxidase subunit I (COI) gene, partial cds	2206	2206	99%	0.0	93%	
EU627698.1	Musca confiscata cytochrome oxidase subunit I (COI) gene, partial c	2178	2178	99%	0.0	92%	
AB479530.1	Musca crassirostris mitochondrial COI gene for cytochrome oxidase su	2167	2167	99%	0.0	92%	
AB479531.1	Musca crassirostris mitochondrial COI gene for cytochrome oxidase su	2161	2161	99%	0.0	92%	
EU627701.1	Musca convexifrons cytochrome oxidase subunit I (COI) gene, partial	2156	2156	99%	0.0	92%	
EU627702.1	Musca formosana cytochrome oxidase subunit I (COI) gene, partial c	2089	2089	99%	0.0	91%	
AB479533.1	Musca bezzii mitochondrial COI gene for cytochrome oxidase subunit	2061	2061	99%	0.0	91%	
EU627695.1	Musca crassirostris cytochrome oxidase subunit I (COI) gene, partial	2056	2056	99%	0.0	91%	
AB479532.1	Musca bezzii mitochondrial COI gene for cytochrome oxidase subunit	2056	2056	99%	0.0	91%	
EU815009.1	Musca domestica isolate JIA-A-1 cytochrome oxidase subunit I (COI)	2049	2049	74%	0.0	100%	
EU627699.1	Musca inferior cytochrome oxidase subunit I (COI) gene, partial cds;	2045	2045	99%	0.0	91%	
EU814999.1	Musca domestica isolate jia21 cytochrome oxidase subunit I (COI) ge	2045	2045	74%	0.0	100%	
EU814993.1	Musca domestica isolate jia14 cytochrome oxidase subunit I (COI) ge	2043	2043	73%	0.0	100%	
EU814992.1	Musca domestica isolate jia13 cytochrome oxidase subunit I (COI) ge	2039	2039	73%	0.0	100%	

FLY TREE

2004-2008, 30 mil. USD, 649 taxonů, desítky tisíc bp



Builders of the Dipteran Tree...

FLYTREE
Introduction
About this Grant
About & Contact Us
Opportunities

Features
About Flies
Pictures
Fly Morphology
Fly Nomenclature
Species Highlights
Phylogeny
Publications & Products

News
Press Releases
Talking About Flies

Buzz About Flies
Additional Buzz
Diptera.org
EDIT Diptera
Tree of Life

//www.cals.ncsu.edu/entomology/wiegmann/

FLYTREE

Builders of the Dipteran Tree...

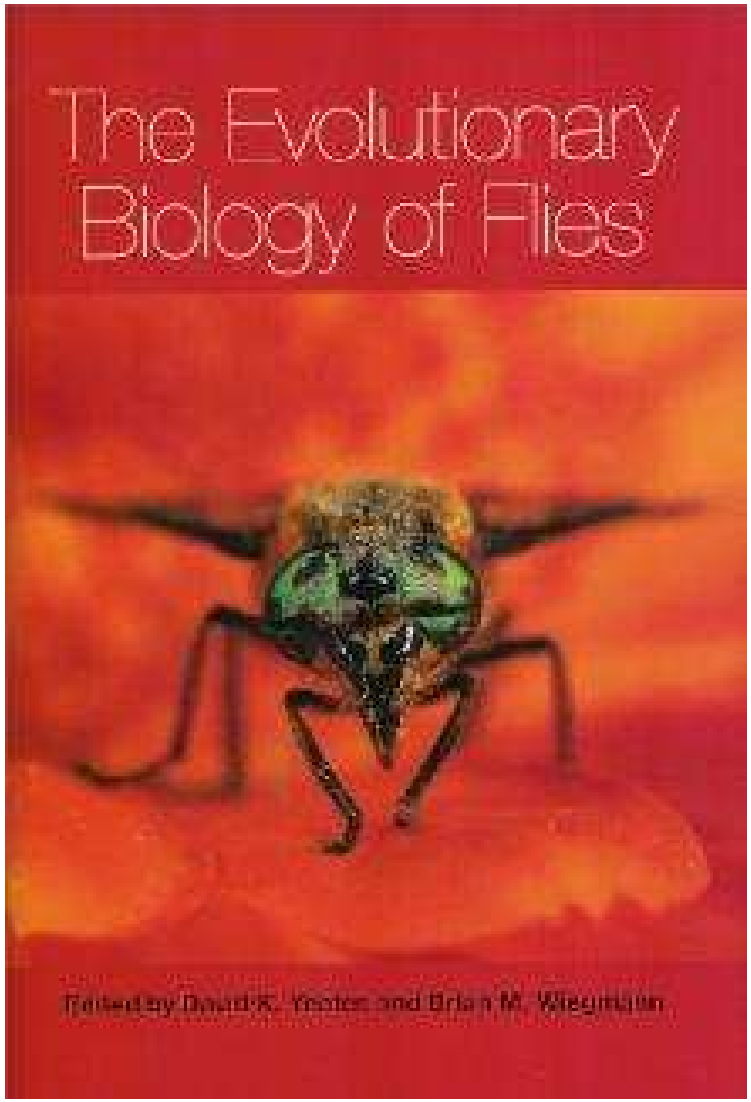
FLYTREE
Introduction
About this Grant
About & Contact Us
Opportunities

Features
About Flies
Pictures
Fly Morphology
Fly Nomenclature
Species Highlights
Phylogeny
Publications & Products

News
Press Releases
Talking About Flies

Buzz About Flies
Additional Buzz
Diptera.org
EDIT Diptera

OUTPUTS


The Evolutionary Biology of Flies


Diptera Diversity
STATUS, CHALLENGES AND TOOLS
Edited by Thomas Pape, Daniel Bickel and Rudolf Meier
BRILL

DNA Taxonomie – pomoc při řešení problému nebo vnášení chaosu?

První záznamy COI do GenBanku – 1996, od té doby ca. 1000 sekvencí dvoukřídlých

V současnosti – ca. 153 000 popsaných druhů dipter – méně než 1% je zařazeno do „Barcoding procesu“

COI nevhodný pro odlišení blízkých druhů

Stanovení hranic druhu – podobnost sekvencí (pairwise distances) - PROBLÉM

Fylogenetická rekonstrukce příbuzenských vztahů – možné řešení – multigenový přístup

Světové sbírky hmyzu – nemožnost použít materiál pro analýzy - PROBLÉM

Taxonomie založená výlučně/převážně na DNA analýze – zkreslený pohled

<u>Potřeba propojit s ostatními přístupy – INTEGRATIVNÍ TAXONOMIE</u>