

Využití molekulárních markerů v systematice a populační biologii rostlin

8. Sekvenování DNA

Sekvenování DNA

- zjištění pořadí nukleotidů v řetězci DNA

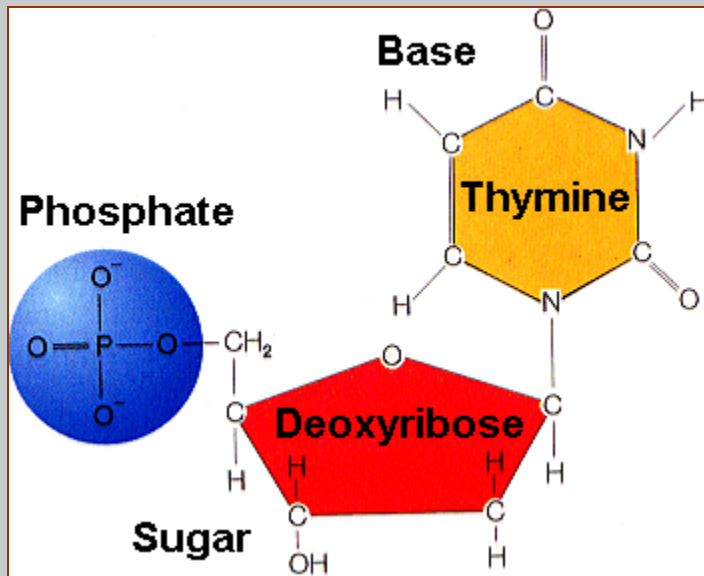
...ATATATAGGCAAGGAATCTCTATTATTAAATCATT...

- využití informace ke zjištění průběhu a rychlosti evoluce
- zjištění podobnosti a příbuznosti taxonů

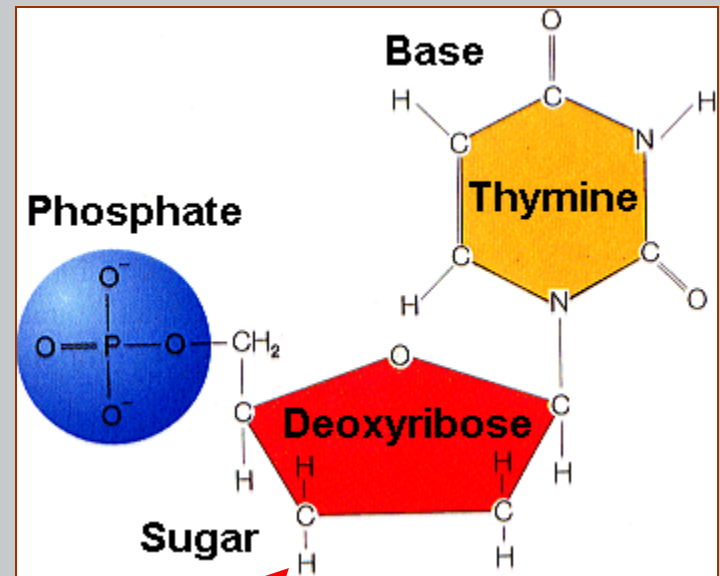
Princip sekvenování

- PCR s použitím dvojice primerů
 - namnožení studovaného úseku
- cyklické sekvenování (dideoxy, Sangerovo)
 - použití pouze jednoho primeru
 - kromě dNTP jsou ve směsi přítomny i ddNTP
 - produkce fragmentů lišících se přesně o 1 bázi
- elektroforetické oddělení fragmentů na gelu
 - automatický sekvenátor

2', 3'- dideoxy NTPs



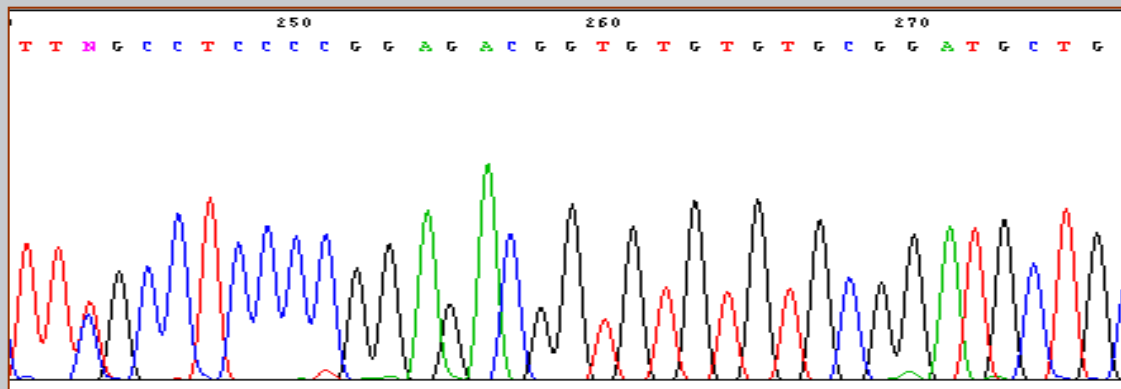
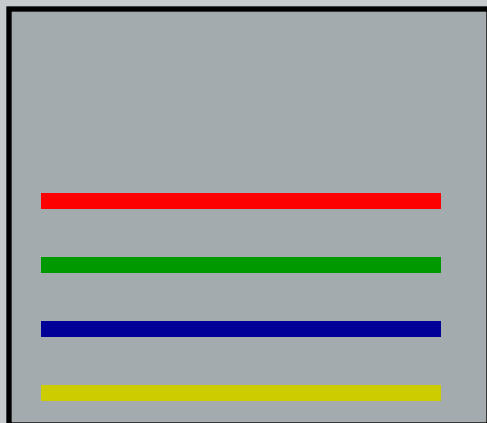
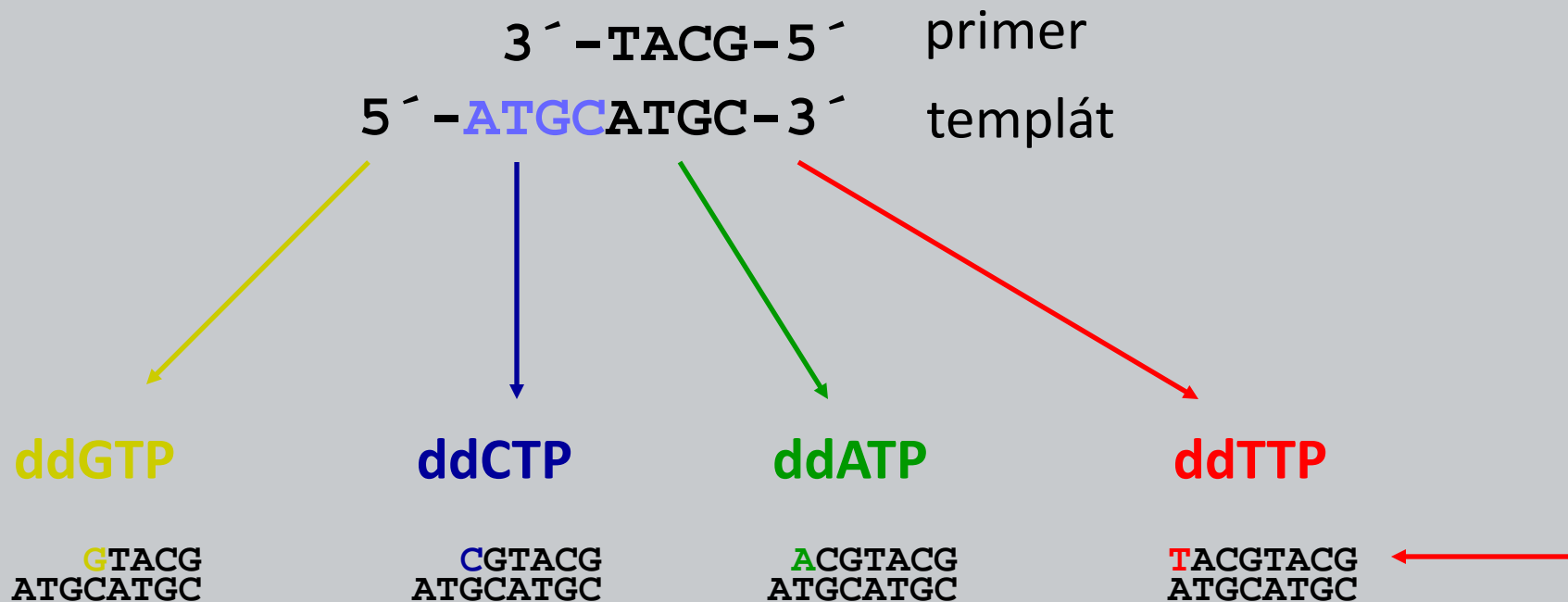
dTTP



ddTTP

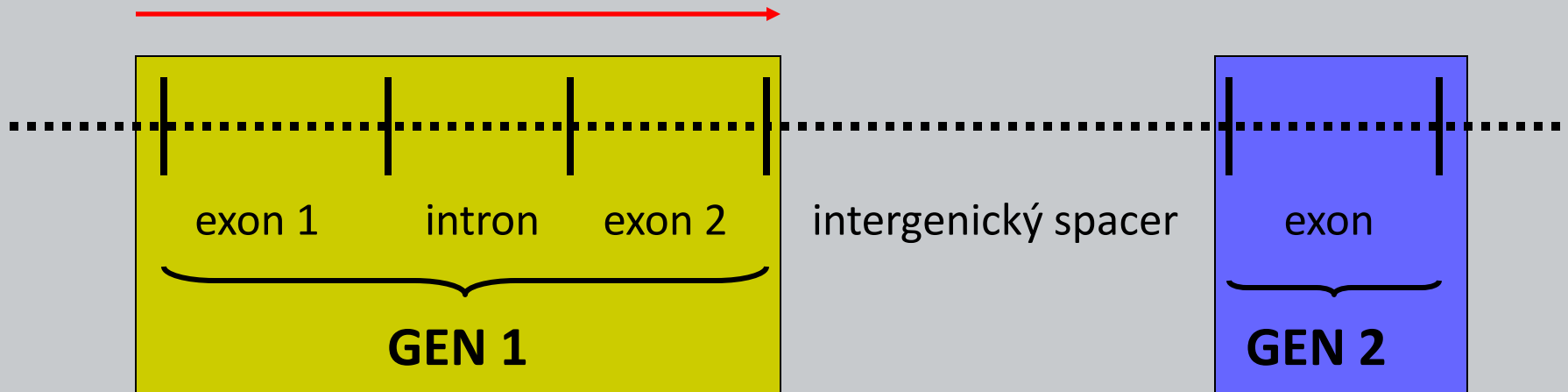
3' -CTGGACTGCA-5'
5' -GACCT

Automatické sekvenování



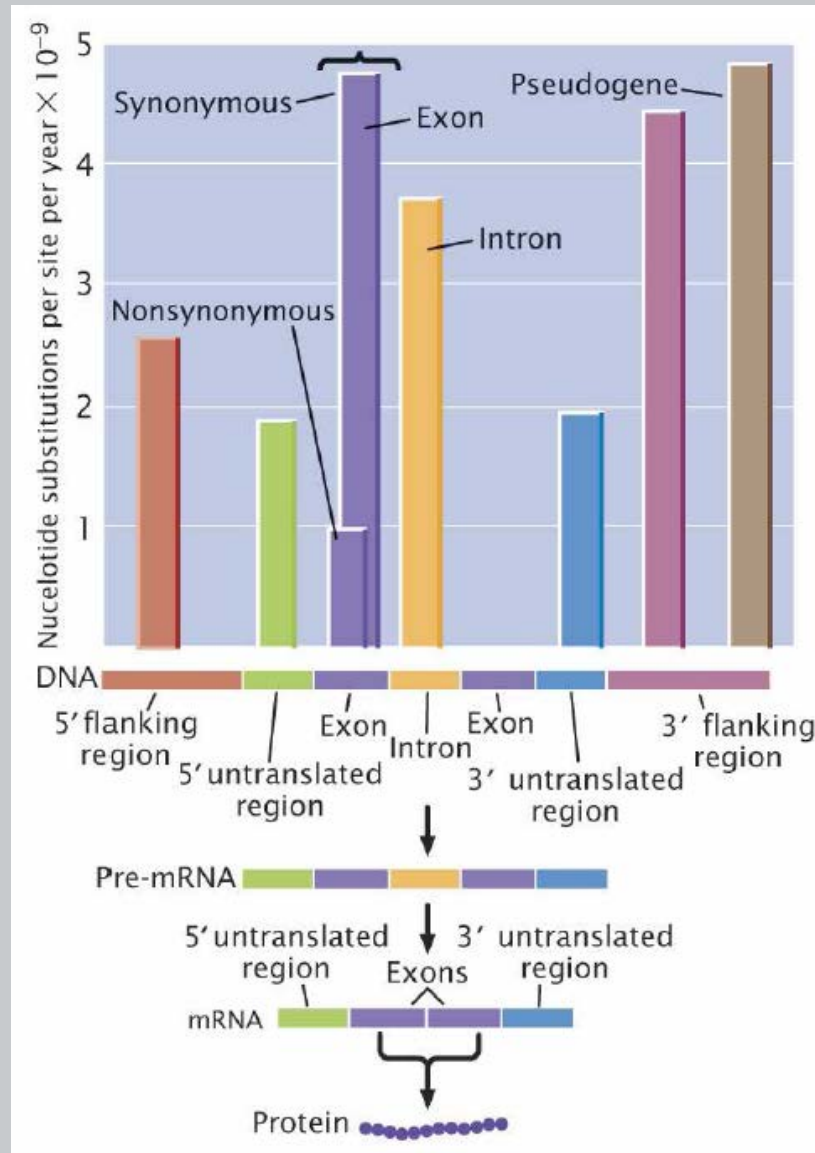
Struktura genomu

- genetická informace – pořadí nukleotidů (ACGT)

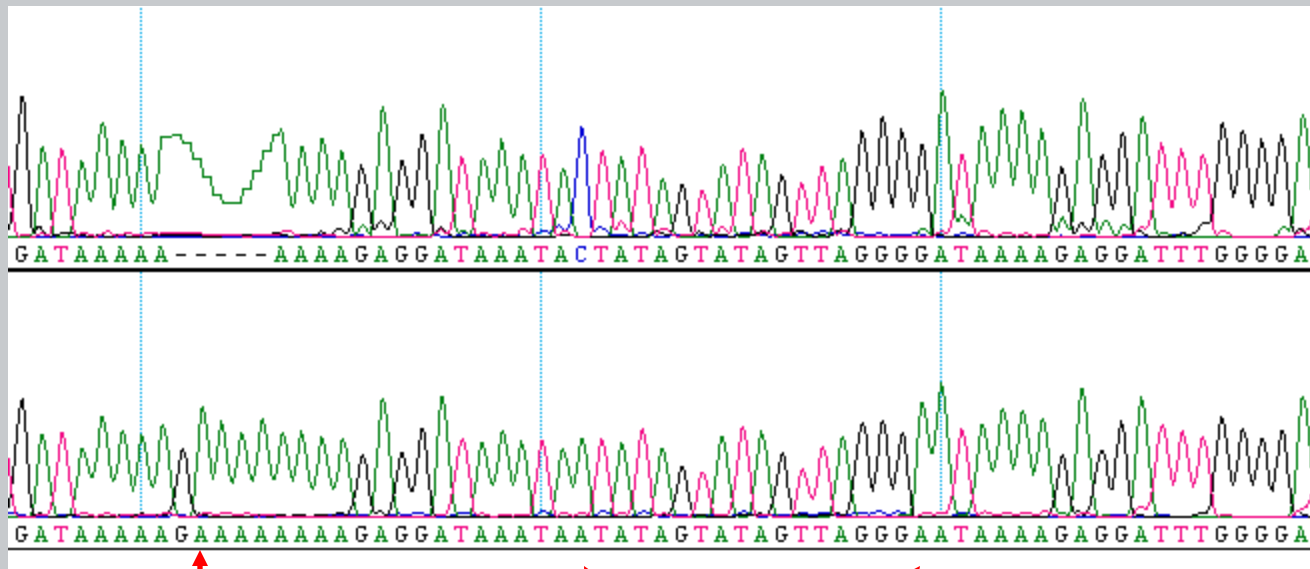


- kodující úseky – exony – *konzervativnější*
- nekodující úseky – introny, spacery – *variabilnější*
- jaderný, chloroplastový a mitochondriální genom

Rychlost evoluce sekvencí



Typy variability v sekvenci DNA



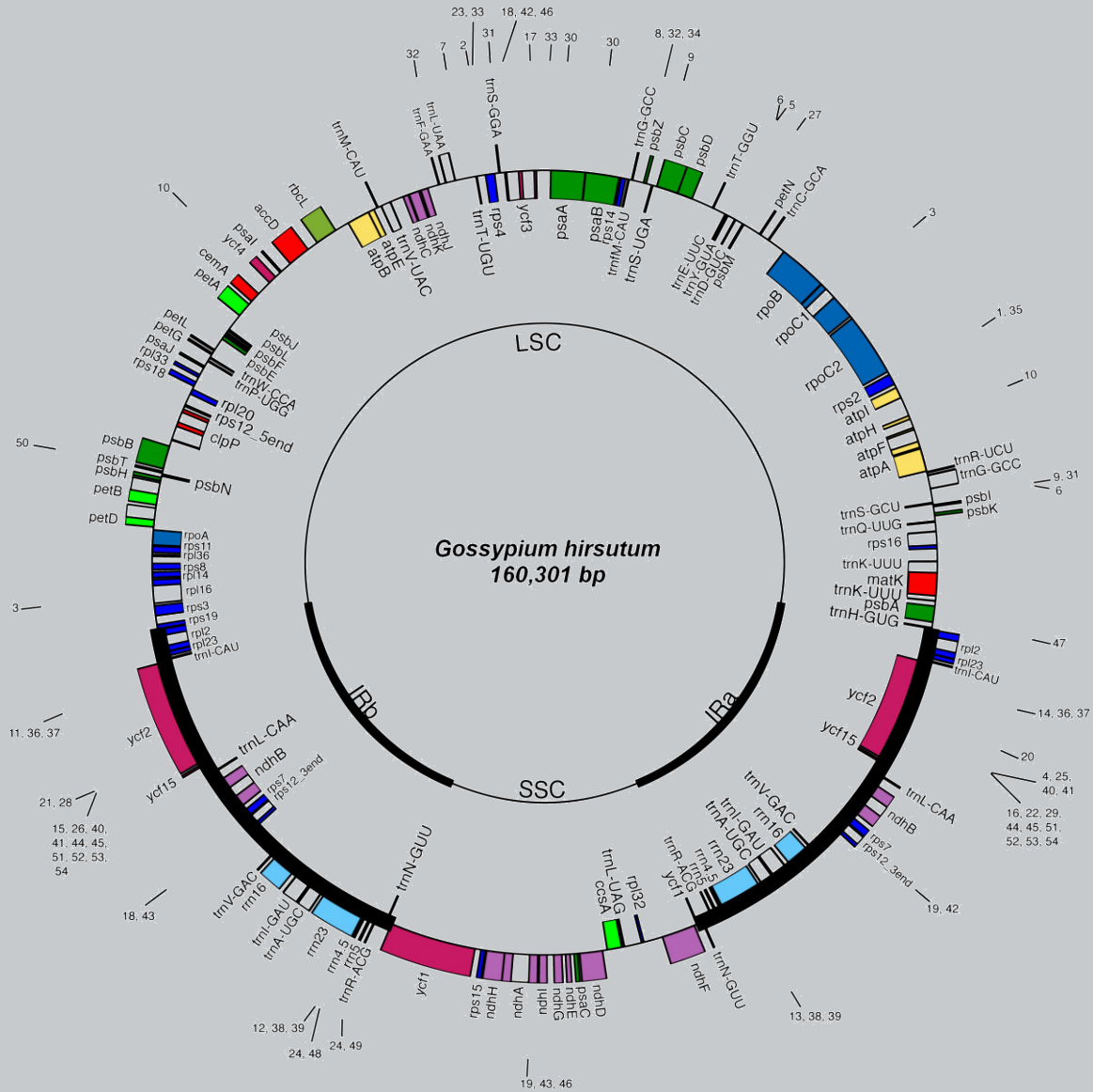
5bp indel

bodové mutace

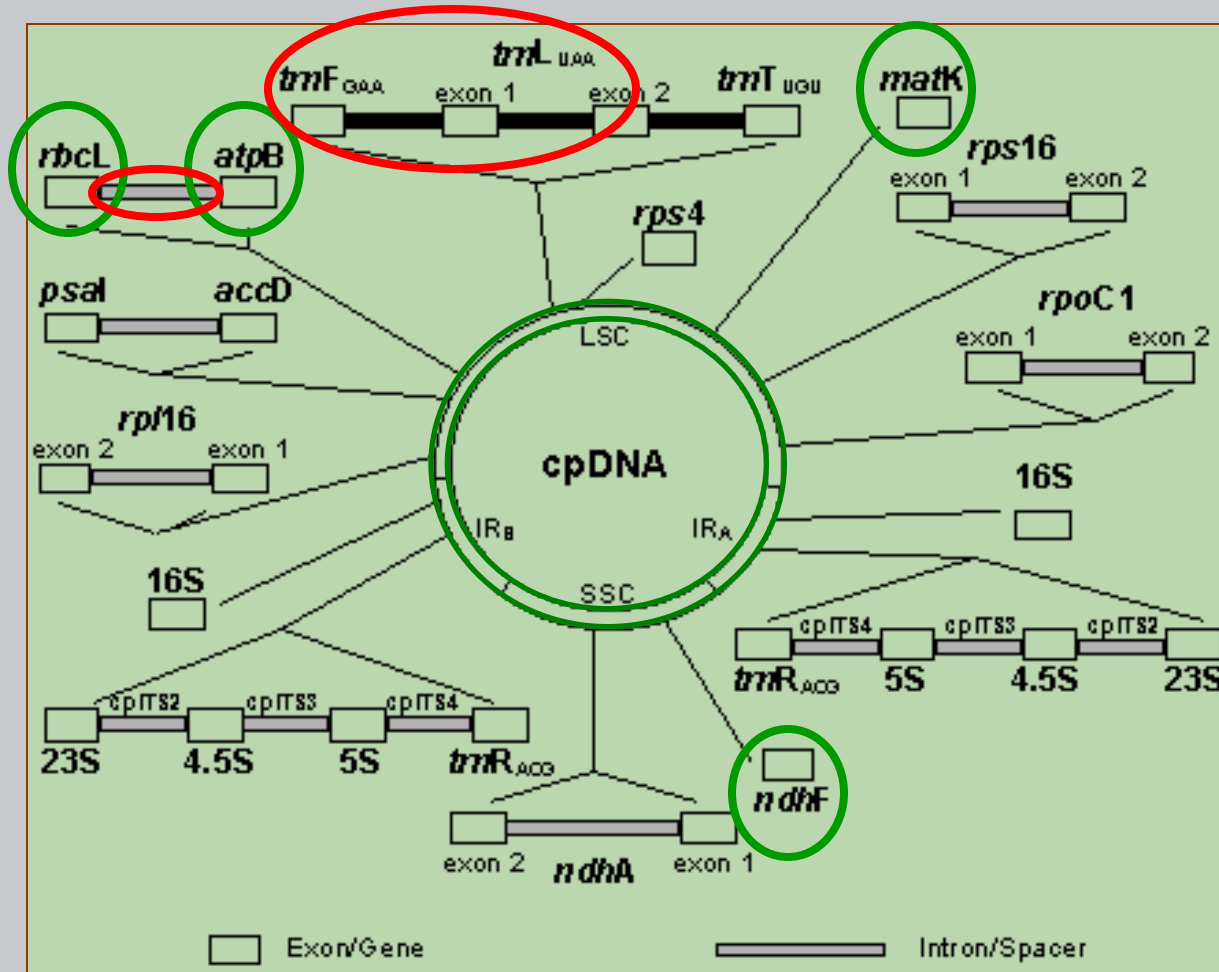
Chloroplastový genom

- mnoho genů je *single-copy* (pouze 1 kopie v celém genomu)
- konzervativní evoluce chloroplastového genomu
 - nevýhoda při studiu na intraspecifické a populační úrovni
 - mnoho konzervativních míst využitelných jako *priming sites* pro PCR amplifikaci úseků mezi nimi
- strukturní přestavby chloroplastového genomu
 - inverse – např. 30kb inverse odlišuje mechorosty od cévnatých rostlin
 - rozsáhlé delece
 - ztráta specifických genů a intronů
- *chloroplast capture*
 - přenos chloroplastu z jednoho druhu do druhého introgresí
 - může ovlivnit určení fylogeneze (pokud není rozpoznán)

Chloroplastový genom

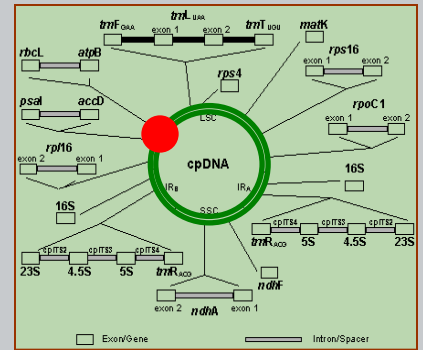


Často sekvenované cpDNA oblasti



+ mnoho dalších...

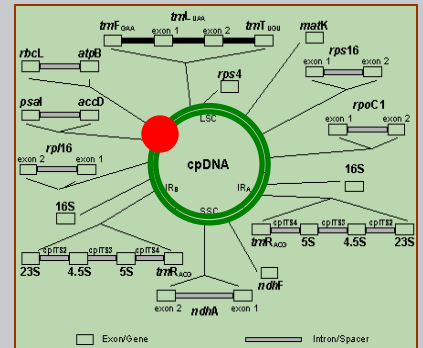
rbcl



- gen pro velkou podjednotku ribuloso-1,5-bisfosfát-karboxylasy/oxygenasy (RUBISCO)
- délka cca 1428, 1431 nebo 1434 bp – indely extrémně vzácné
- jeden z prvních sekvenovaných genů
- velmi konzervativní, systematika na úrovni čeledí a rodů, v některých skupinách i druhů

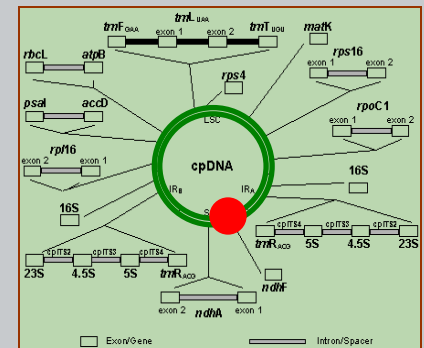
atpB

- gen kodující beta podjednotku ATP synthasy
- délka 1497 bp, indely nezjištěny
- využití podobně jako *rbcL*

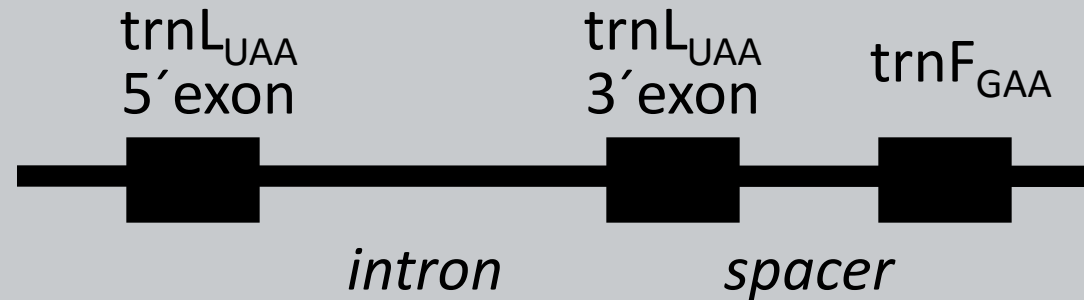
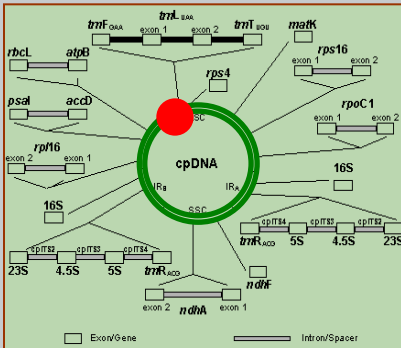


ndhF

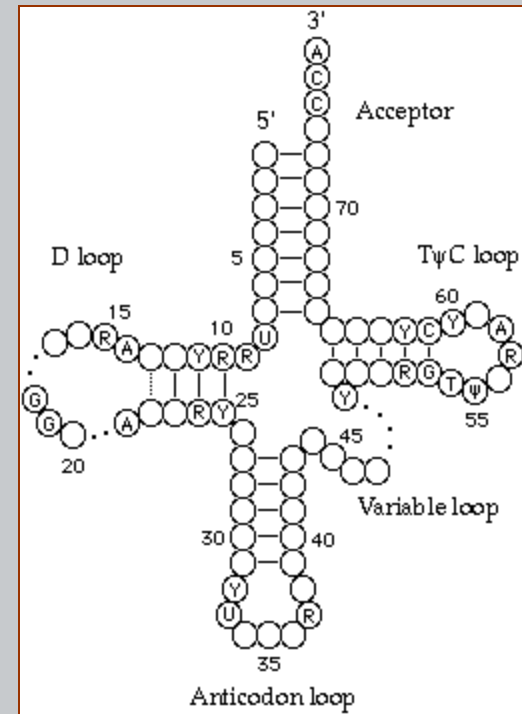
- koduje podjednotku chloroplastové NADH-dehydrogenasy
- délka 2233 bp (tabák)
- cca 2x více substitucí než *rbcL*
- využití na rodové úrovni



trnL intron a spacer mezi *trnL* a *trnF*

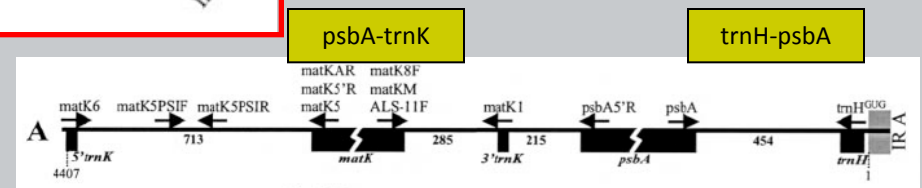
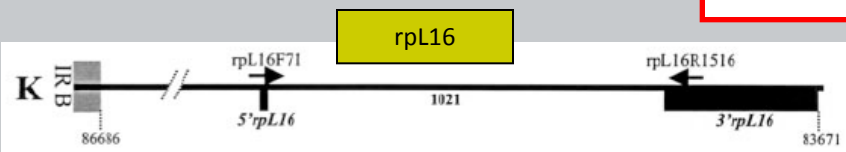
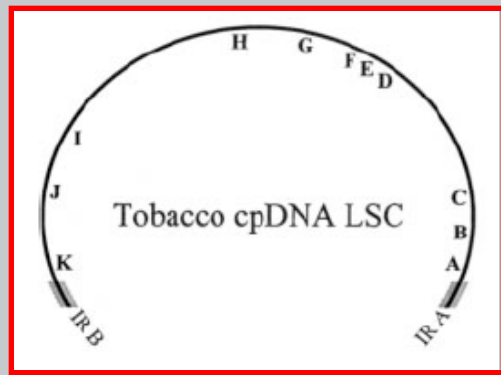
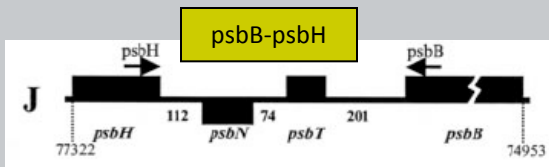
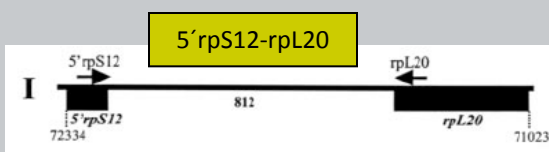
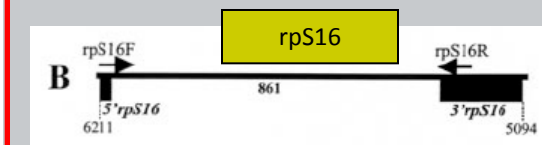
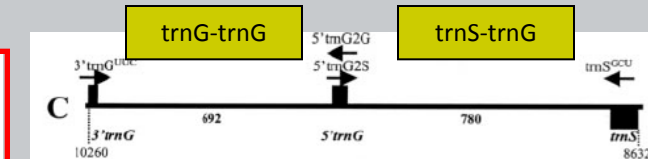
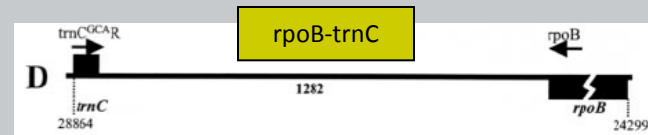
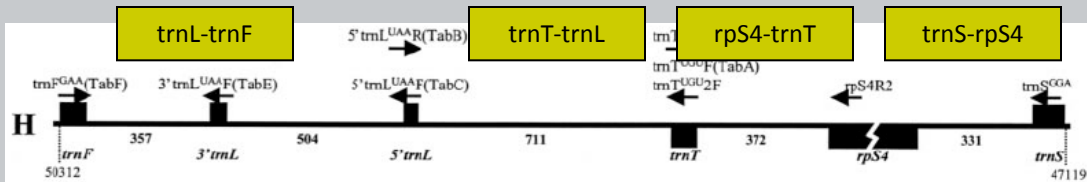
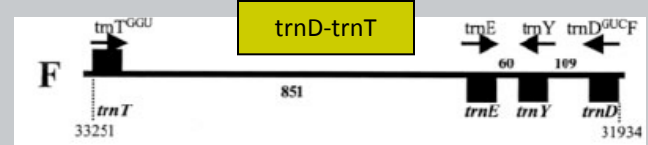
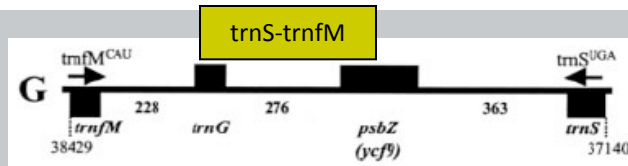
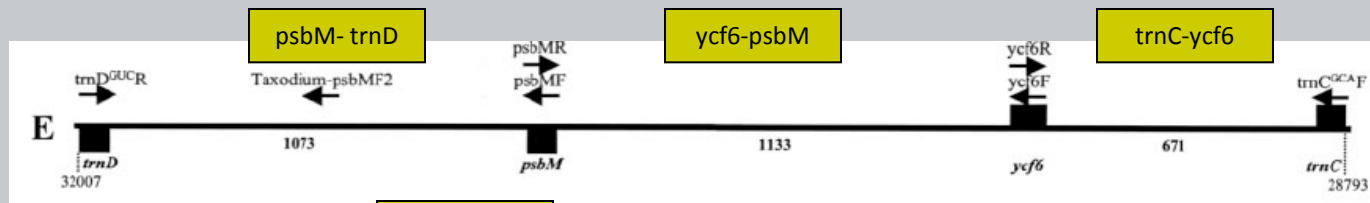


- tRNA geny – sekundární struktura
- akumulace insercí/delecí se stejnou rychlostí jako substituce nukleotidů
- problémy s alignmentem, zvláště u vzdálených organismů (už třeba na úrovni čeledi)
- vhodné pro systematiku na úrovni blízce příbuzných druhů



Variabilní nekodující oblasti cpDNA

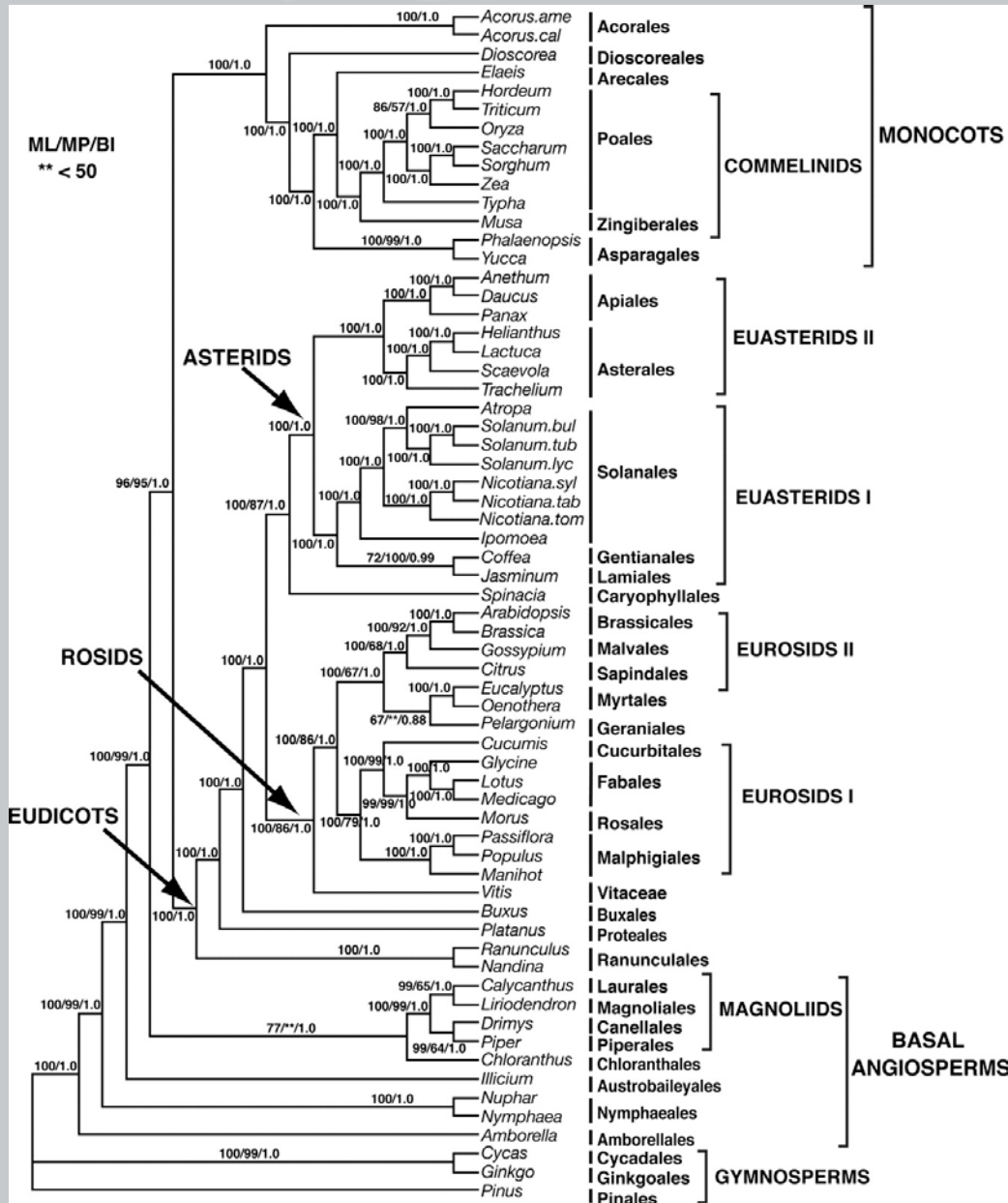
Shaw et al. (2005): The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* **92**: 142-166



Využití chloroplastových sekvencí

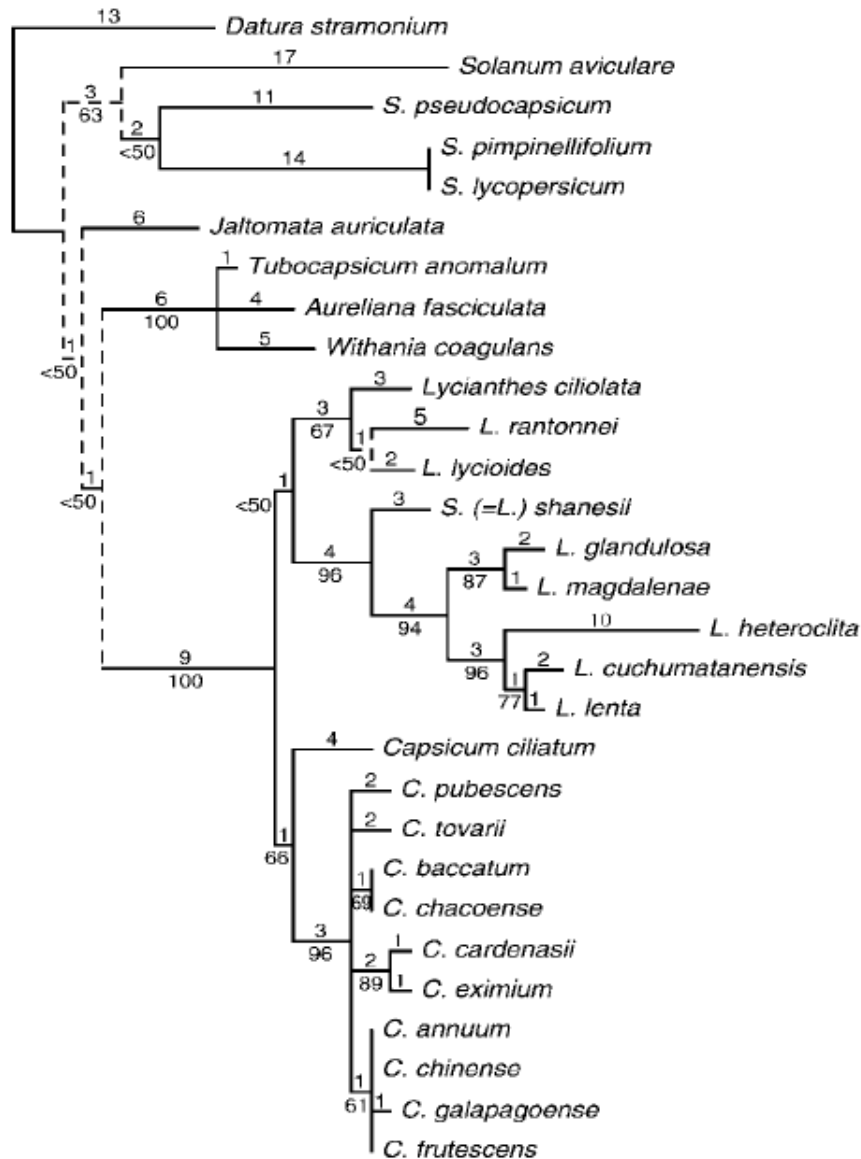
- fylogeneze krytosemenných rostlin
- mezidruhové vztahy v rámci rodu
- vnitrodruhová fylogeografie (definice hplotypů)
- hybridizace – zjištění mateřského taxonu (jedince) (maternální přenos chloroplastu)

Fylogeneze krytosemenných



Jansen et al. (2007)
81 plastidových genů
76 583 nukleotidů

Vztahy mezi druhy



Capsicum
atpB-rbcL spacer
 Walsh & Hoot (2001)

Fylogeografie

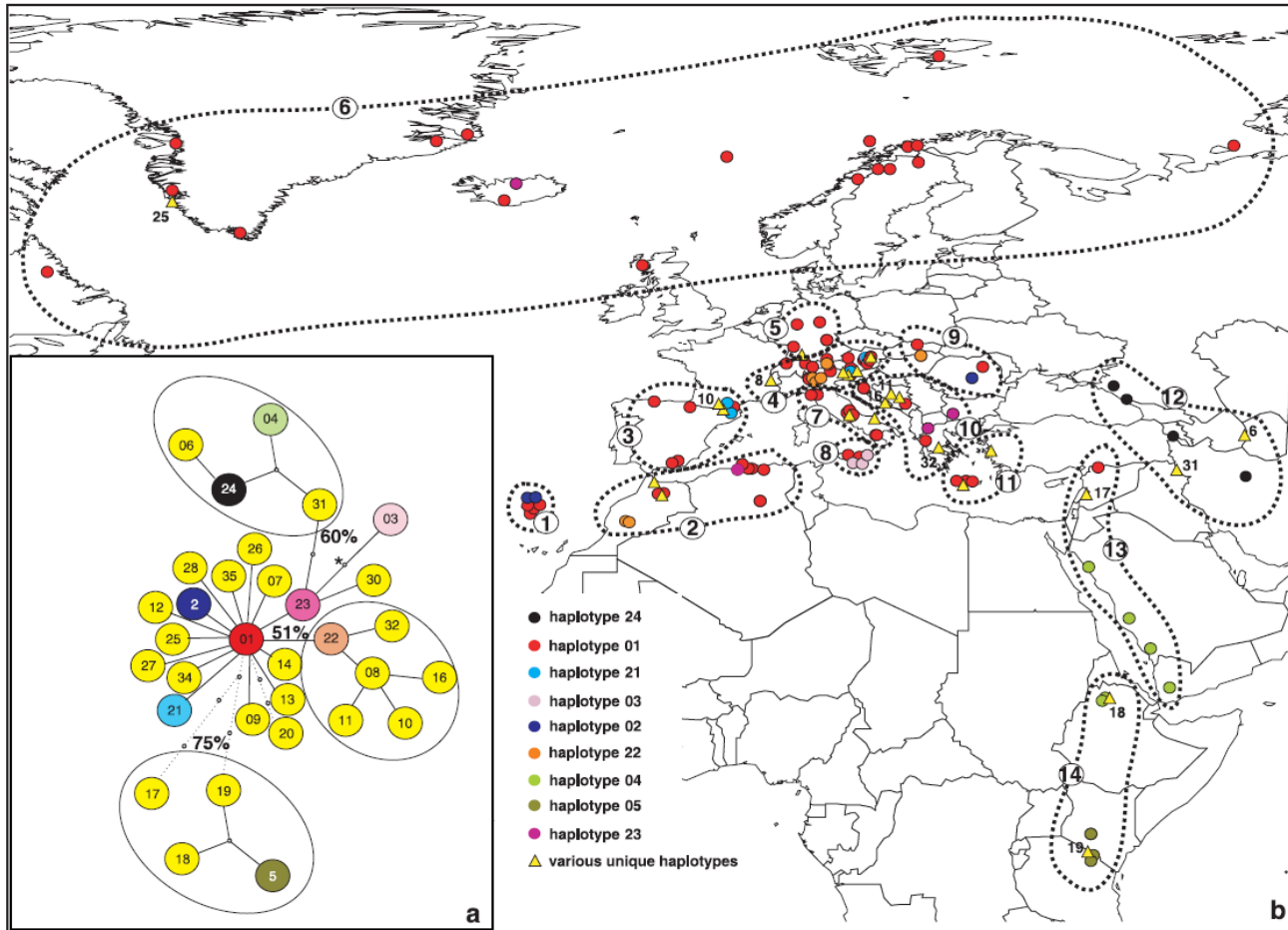


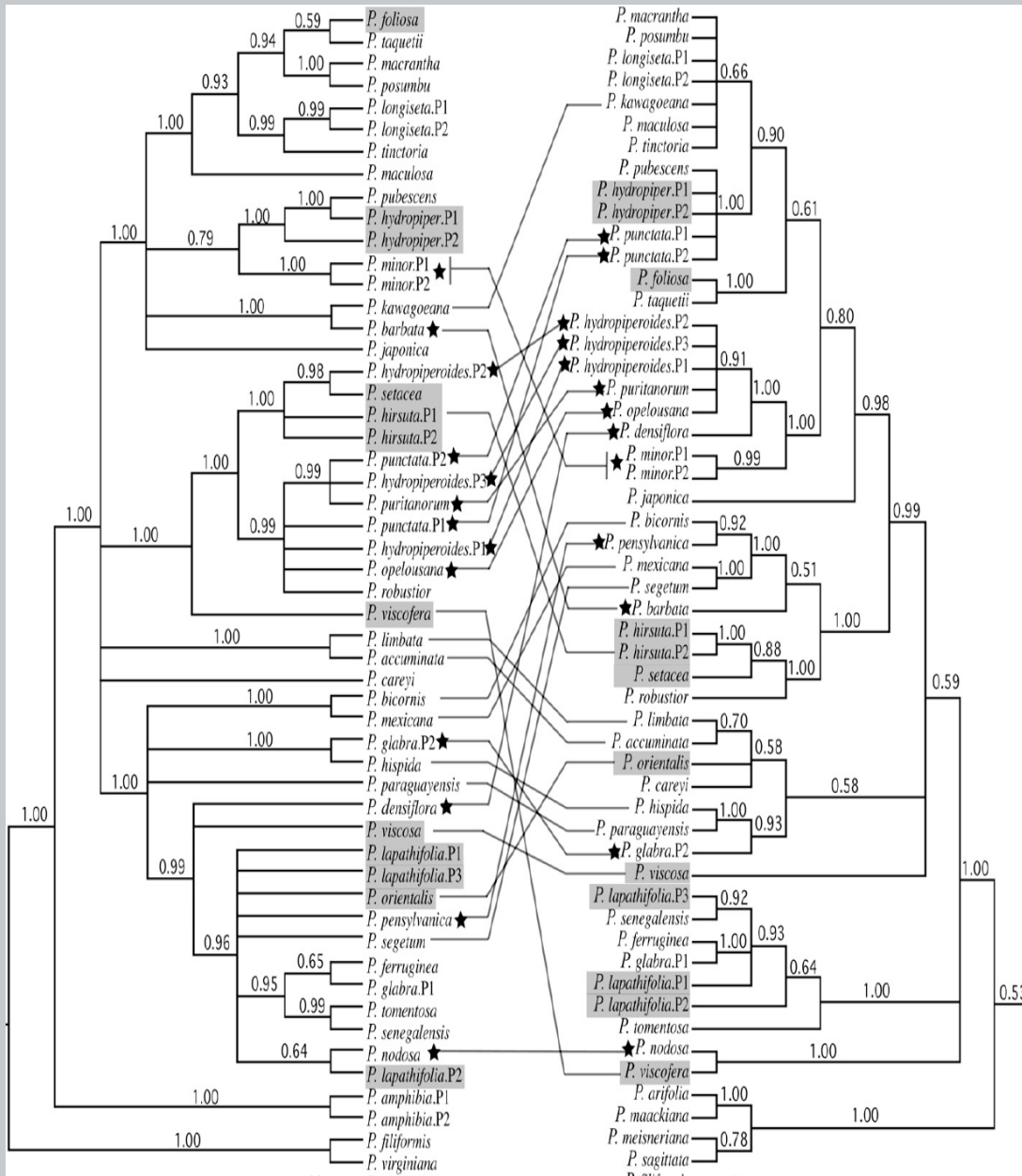
Fig. 1 (a) *Arabis alpina* cpDNA haplotype network as inferred from rcs (see Material and methods section for details). The position of the root as revealed from parsimony and ML analyses is indicated (*). Groups of haplotypes circled correspond to clades that have been also recognized by parsimony and ML analyses (bootstrap values redrawn from parsimony analysis). Colour code and haplotype designation follow Fig. 1 (b), with yellow symbols showing unique haplotypes. (b) Distribution of *A. alpina* cpDNA haplotypes. The distribution range of *A. alpina* is represented by the accessions. Regional subdivision of the accessions into 14 areas is indicated (refer to Table 1 and Table S1, Supplementary materials).

Arabis alpina

trnL-trnF

Koch et al. 2006

Mezidruhová hybridizace



rozpor mezi cpDNA a nDNA



Persicaria
matK, psbA-trnH, trnL-trnF
 vs. ITS
 Kim & Donoghue 2008

Analýza dat

- *multiple alignment*

S206	ATATATATATAGGCAAGGAATCTCTATTATTAAATCATTTAGAATCCATA
S207	ATATATATA--GGCAAGGAATCTCTATTATTAAATCATTTAGAATCCATA
S208	ATATATATA--GGCAAGGAATCTCTATTATTAAATCATTTAGAATCCATA
S209	ATATATATA--GGCAAGGAATCTCTATTATTAAATCATTTAGAATCCATA
S210	ATATATATA--GGCAAGGAATCTCTATTATTAAATCATTTAGAATCCATA
S0G3	ATATATATA--GGCAAGGAATCTCTATTATTAAATCATTTAGAATCCATA
TL	ATATATATATAGGCAAGGAATCTCTATTATTAAATCATTCATAATTCATA

- konstrukce fylogenetického stromu

- distanční metody
- maximální parsimonie (MP)
- maximální věrohodnost (*maximum likelihood* – ML)
- Bayesovský přístup

Maximální parsimonie (MP)

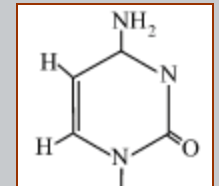
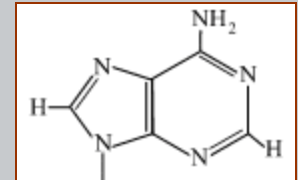
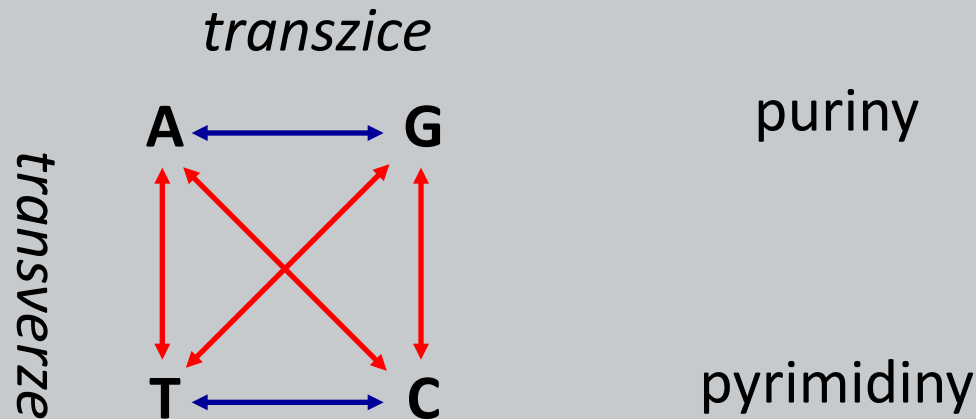
- kladistická metoda
- hledání nejjednoduššího stromu (*most parsimonious tree*)
- tj. stromu, při kterém je evoluce vysvětlena co nejmenším počtem substitucí
- software PAUP *
Phylogenetic Analysis Using Parsimony
(* and other methods)

Maximum likelihood (ML)

- najítí stromu s nejvyšší pravděpodobností (L)
- pravděpodobnost, že se při dané topologii stromu (za daného evolučního modelu) vyvinou pozorované sekvence
- software PAUP, GARLI, PAML, PhyML, RAxML...

Modely evoluce sekvence DNA

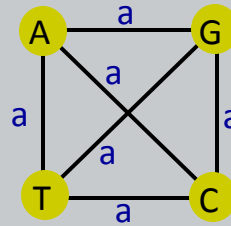
- modely změn v sekvenci



- parametry
 - frekvence bazí
 - typy substitucí (transzice, transverze)
 - heterogenita rychlosti substitucí
 - zastoupení invariantních míst

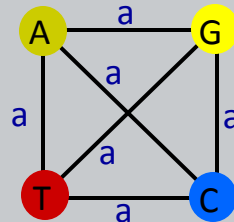
Substituční modely

Zvyšující se počet parametrů modelu



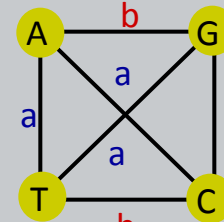
JC (Jukes-Cantor 1969)

- stejné rychlosti substituce
- stejné frekvence bazí



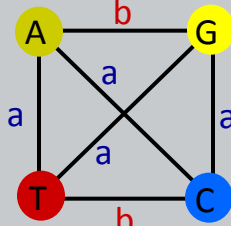
F81 (Felsenstein 1981)

- stejné rychlosti substituce
- nestejně frekvence bazí



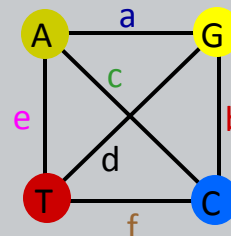
K2P (Kimura 2 parameter 1980)

- dvě různé rychlosti substituce
- stejné frekvence bazí



HKY (Hasegawa, Kishino & Yano 1985)

- dvě různé rychlosti substituce
- nestejně frekvence bazí



GTR (General time-reversible model)

(Tavaré et al. 1986)

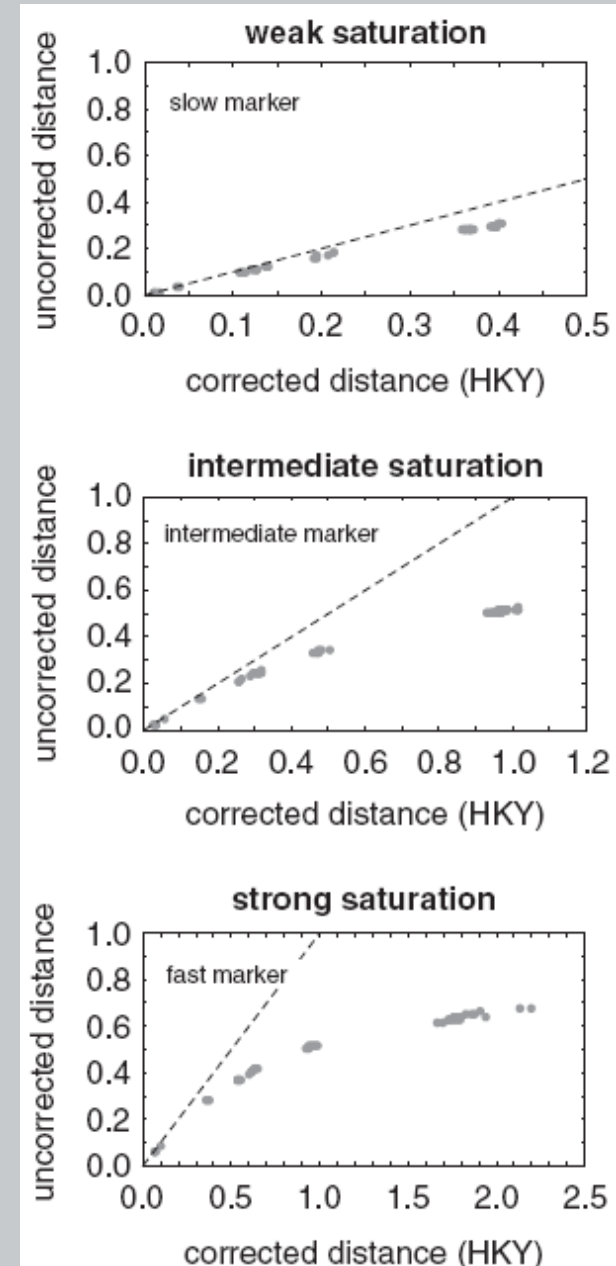
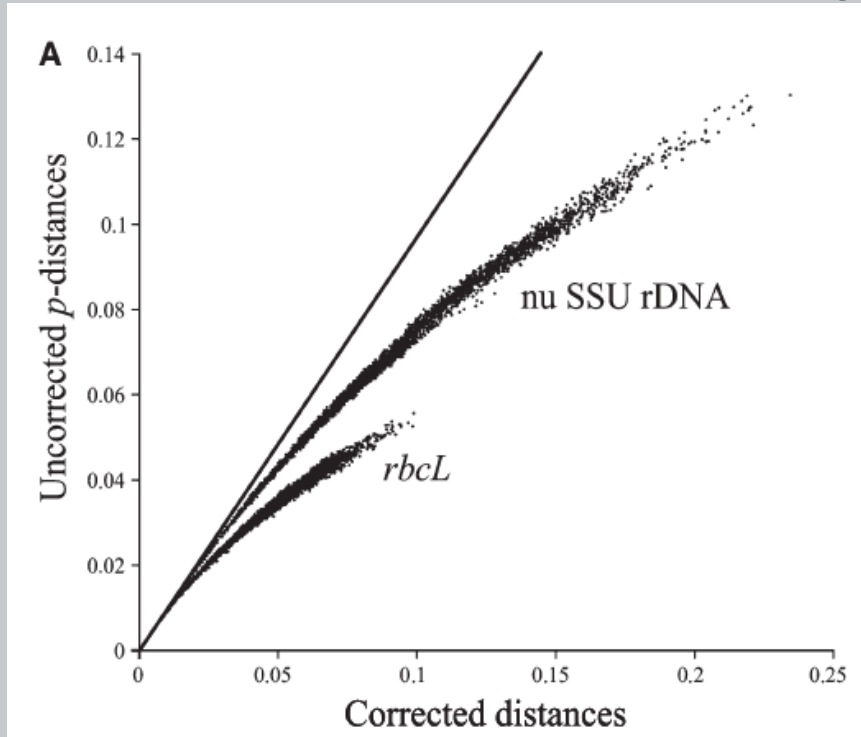
- 6 různých rychlostí substituce
- nestejně frekvence bazí

Jaký model vybrat?

- MODELTEST: A tool to select the best-fit model of nucleotide substitution (Posada et al.)
- testování různých modelů – výběr toho nejjednoduššího, který dostatečně vysvětluje data pomocí
 - hierarchical likelihood ratio tests (hLRTs)
 - Akaike information criterion (AIC)
- jModelTest2 (<https://code.google.com/p/jmodeltest2/>)

Saturace

- signál a šum v datech
- nekorigovaná vs. korigovaná vzdálenost
- skewness (g_1 -statistics), I_{SS}



Molekulární hodiny

- striktní (globální)

- *clocklike evolution*

- lokální

- *relaxed clocks*

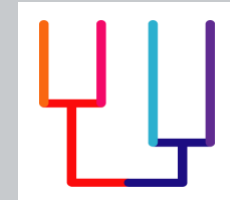
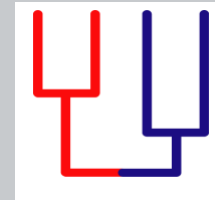
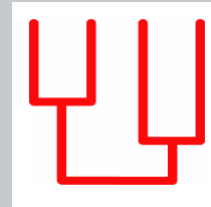
- autokorelované (blízce příbuzné taxony mají podobnou mutační rychlost)
- nekorelované (uncorrelated – lognormal, exponential)

- kalibrace

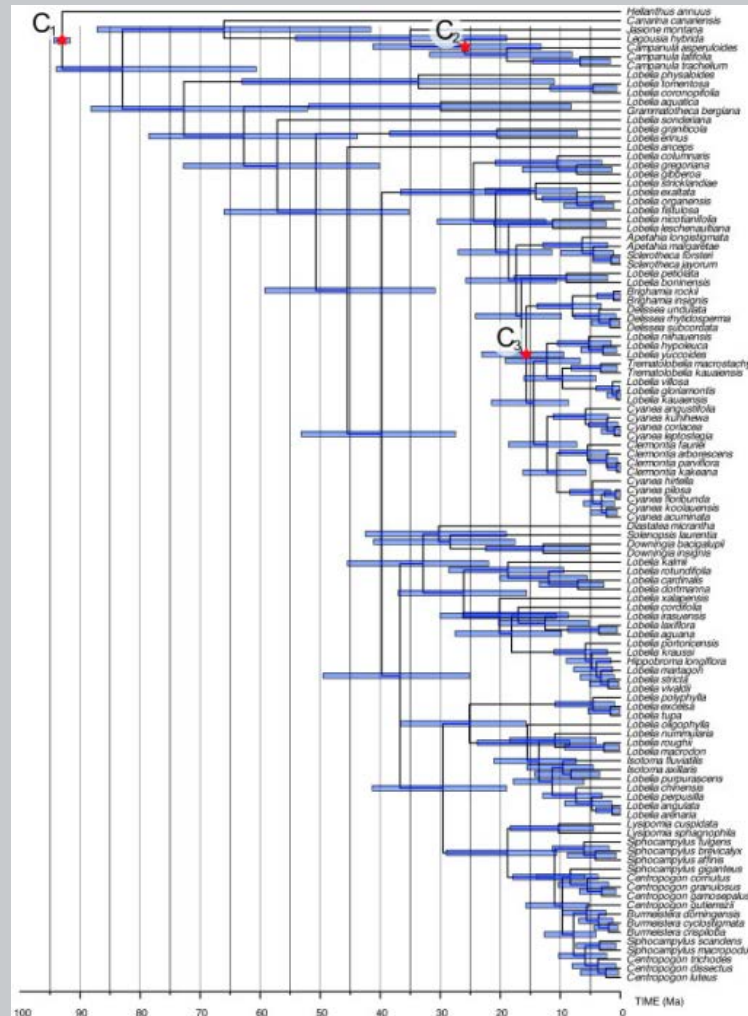
- substituční rychlost z jiné studie nebo obecná rychlost (např. pro cpDNA)
- fosilie
- biogeografie

- software

- BEAST (Bayesian), r8s (non-parametric rate smoothing, penalized likelihood), ...



Odhad doby divergence (BEAST)



Antonelli A. (2009): Have giant lobelias evolved several times independently? Life form shifts and historical biogeography of the cosmopolitan and highly diverse subfamily *Lobelioideae* (Campanulaceae). *BMC Biol.* 7:82

Genové banky – databáze sekvencí

- **GenBank**

National Centre for Biotechnology Information
(NCBI)

<http://www.ncbi.nlm.nih.gov/>

- **EMBL**

European Bioinformatics Institute (EBI)

<http://www.ebi.ac.uk/embl/>

Populační studie

Chiang T.Y. & Schaal B.A. (1999): Phylogeography of North American populations of the moss species *Hylocomium splendens* based on the nucleotide sequence of internal transcribed spacer 2 of nuclear ribosomal DNA. *Molecular Ecology* 8:1037-1042



Systematická studie

Martins L., Oberprieler C. & Hellwig F.H. (2003): A phylogenetic analysis of *Primulaceae* s.l. based on internal transcribed spacer (ITS) DNA sequence data. *Plant Systematics and Evolution* 237: 75-85



Literatura

Soltis D.E. & al. [eds.] (1998): *Molecular systematics of plants.II. DNA sequencing.*

Hollingsworth & al. [eds.] (1999): *Molecular systematics and plant evolution.*

Hall B.G. (2001): *Phylogenetic trees made easy.*

Felsenstein J. (2004): *Inferring phylogenies.*

Lemey P. & al. [eds.] (2009): *The phylogenetic handbook.* 2nd ed.

Wiley E.O. & Lieberman B.S. (2011): *Phylogenetics. Theory and Practice of Phylogenetic Systematics.* 2nd ed.

Wheeler W. C. (2012): *Systematics. A course of lectures.*

Drummond et al. (2009): *Relaxed phylogenetics and dating with confidence.* PLoS Biol 4(5): e88