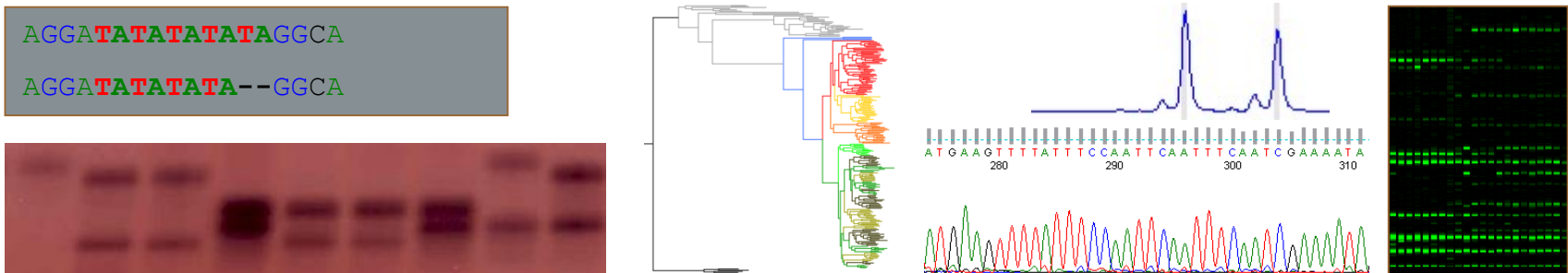


# Uplatnění molekulárních přístupů ve studiu biodiverzity, v taxonomii a v rekonstrukci evoluce



1. přehled metod, použití při studiu hybridizace a  
fylogeneze

Tomáš Fér  
Katedra botaniky PŘF UK

# Uložení genetické informace

## jádro

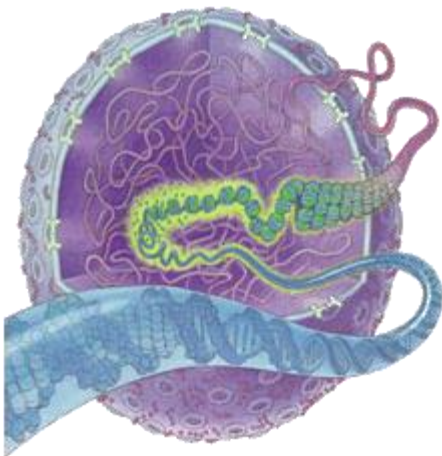
- různá ploidie
- rekombince
- biparentální přenos

## plastidy

- 1 kruhová molekula
- bez rekombince
- uniparentální přenos

## mitochondrie

- strukturně komplikovaná
- strukturní přestavby
- uniparentální přenos



# Charakteristika genomů

	nDNA zvířata	rostliny	cpDNA	mtDNA zvířata	rostliny
<b>dědičnost</b>	biparentální	biparentální	krytosemené maternální, konifery paternální	maternální	maternální
<b>struktura</b>	lineární	lineární	cirkulární	cirkulární	cirkulární, komplexní
<b>velikost (kb)</b>	$4.9 \times 10^4 - 7.0 \times 10^8$	$5.0 \times 10^4 - 3.0 \times 10^8$	71 – 214	15 – 20	200 – 2400
<b>substituční rychlost</b>	$3.5 \times 10^{-9}$	$4.1 - 5.7 \times 10^{-9}$	$0.86 - 1.20 \times 10^{-9}$	$56 \times 10^{-9}$	$0.36 - 0.50 \times 10^{-9}$
<b>substituční rychlost relativně vůči rostlinné mtDNA</b>	8.1	11.4	2.4	130.2	1.0
<b>cizorodé sekvence</b>	běžné	běžné	vzácné	vzácné	běžné
<b>strukturní mutace</b>	běžné	běžné	vzácné	vzácné	běžné
<b>rekombinace</b>	ano	ano	intramolekulární	ne	inter- a intramolekulární

# Základní molekulární metody využívané v systematice

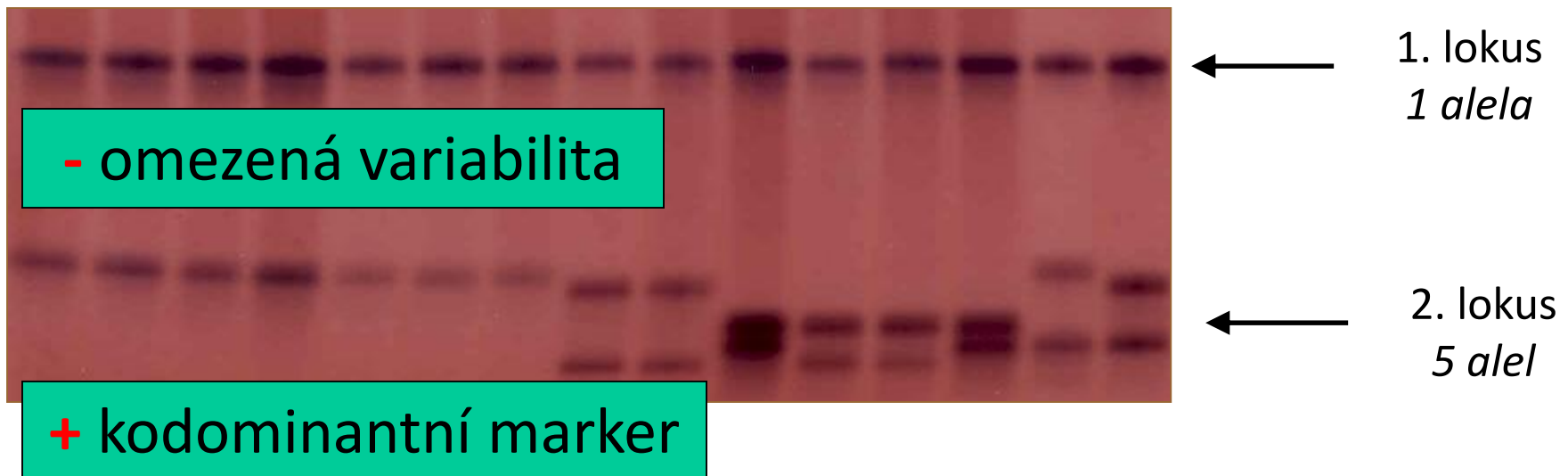
- izozymy
- RFLP, PCR-RFLP
- dominantní markery – RAPD, AFLP, ISSR
- mikrosatelity
- sekvence
  - klasické (Sangerovo, dideoxy) sekvenování
  - next-generation sequencing (NGS) – RNAseq, resequencing, Hyb-Seq, RADSeq...

# Markery – důležité rozdíly

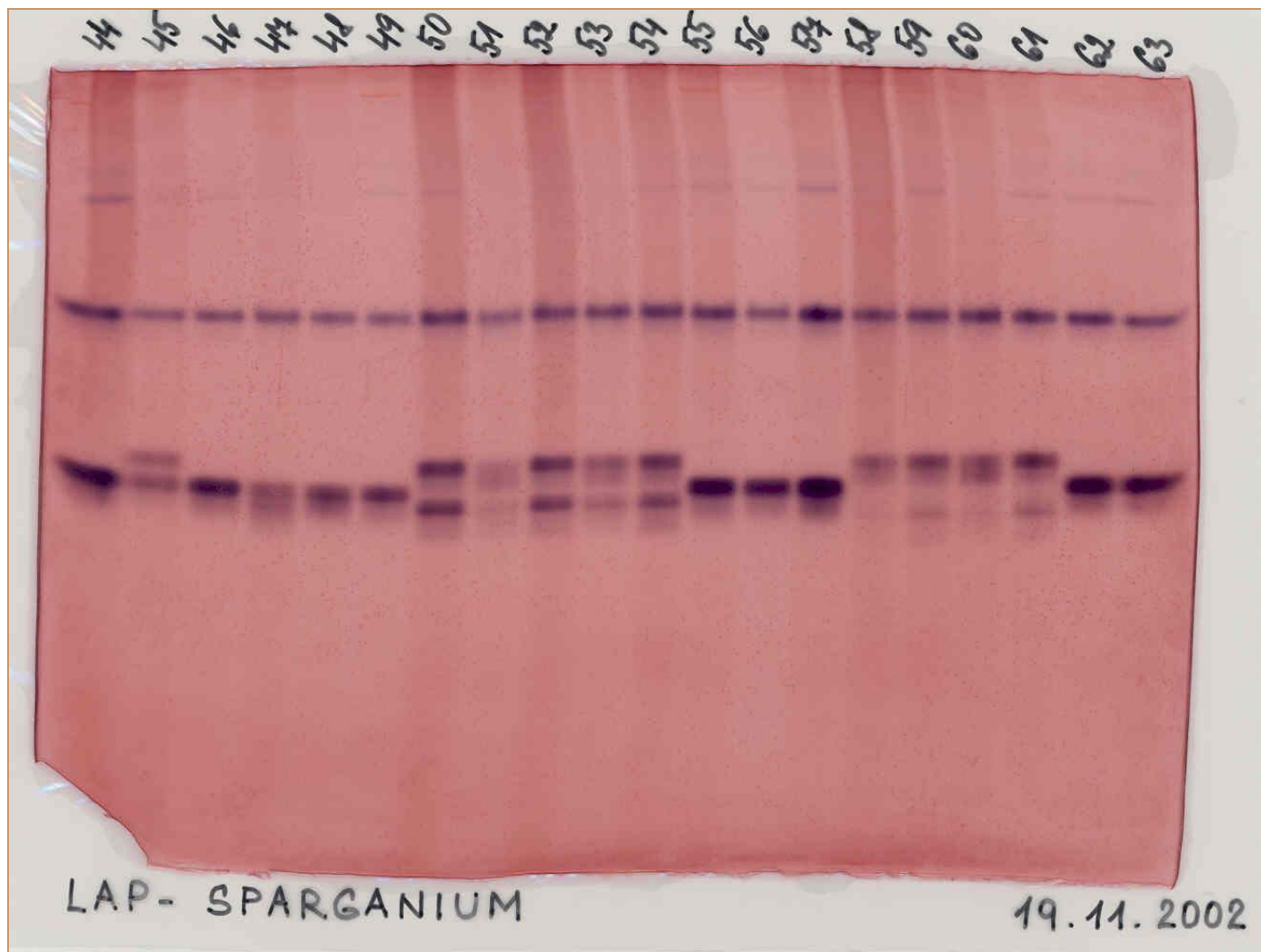
	variabilita	dědičnost	výhody	nevýhody	
<b>isozymy</b>	xx	kodominantní	levné, univerzální	čerstvý materiál, omezená variabilita, selekce?	
<b>RAPD</b>	xxx	dominantní	levné, mnoho proužků	omezená opakovatelnost	
<b>AFLP</b>	xxxx	dominantní	vysoký polymorfismus, reproducibilita	komplikované	
<b>ISSR</b>	xxx	dominantní	jednoduché, polymorfní		
<b>SSR</b>	xxxx	kodominantní	vysoce variabilní	druhově specifické	
<b>sekvence</b>	<b>cpDNA</b>	x	haploidní	možnost srovnání, nerekombinované	relativně drahé, často nízká variabilita
	<b>nDNA</b>	x - xxxx	kodominantní	rekombinované, mnoho analytických metod	zjistíme <i>gene tree</i>

# Isoenzymy (isozymy, allozymy)

- proteiny katalyzující základní biochemické reakce
- extrakce z jakýchkoliv živých pletiv – nejlépe listů
- na základě různě velkého elektrického náboje jsou jednotlivé „typy“ (= alely) rozděleny elektroforézou
- vizualizace pomocí „barevných reakcí“



# Příklad isoenzymového gelu



# RFLP

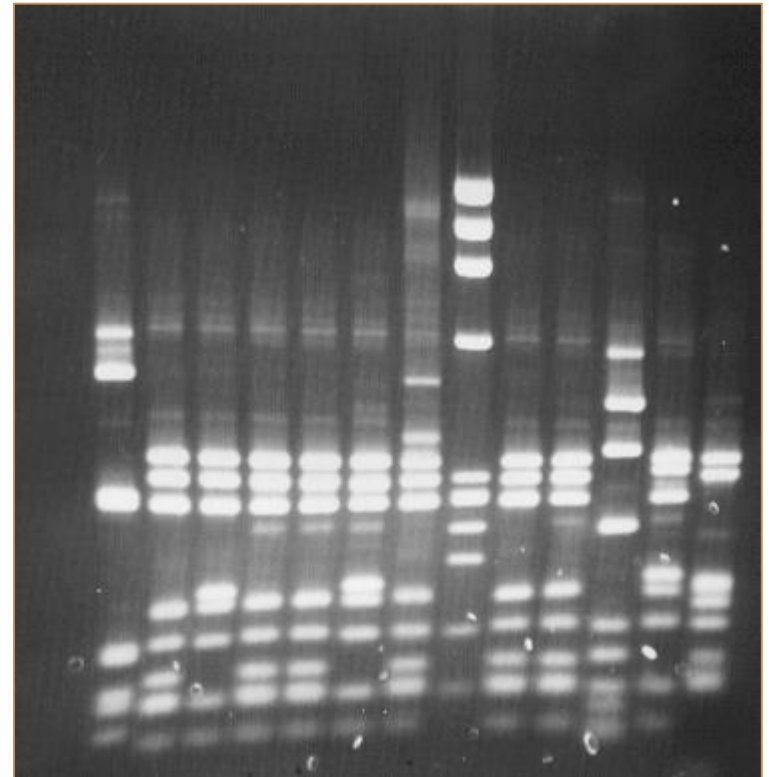
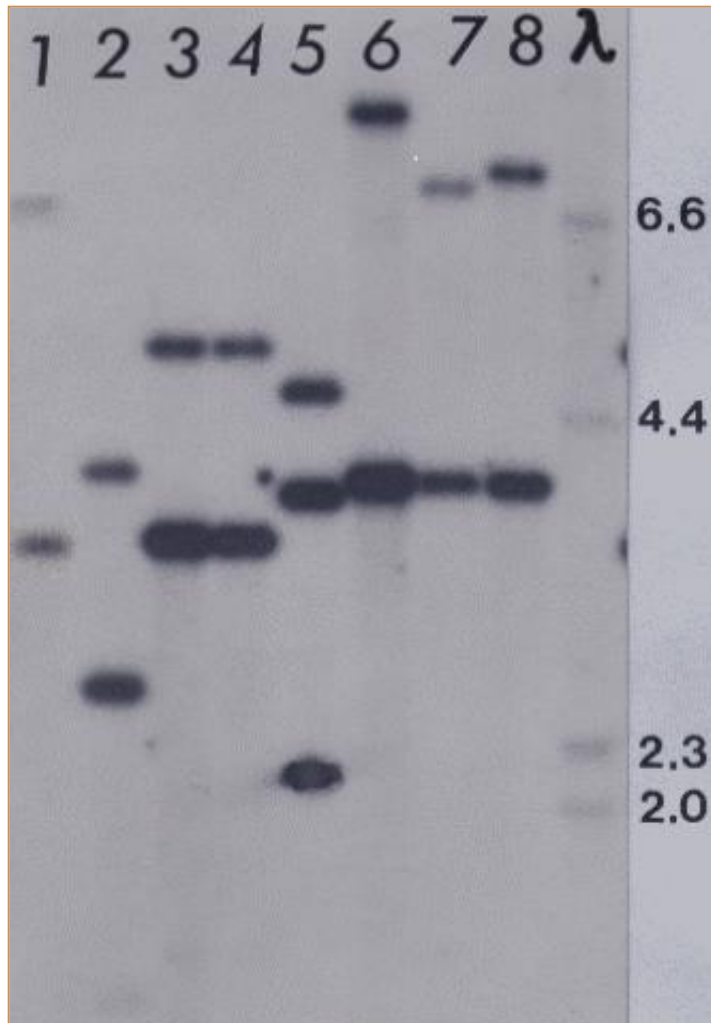
## Restriction Fragment Length Polymorphism

- DNA je specificky štěpena na fragmenty restričními enzymem
- elektroforéza – délkové rozdělení
- velké množství fragmentů – část je vizualizována hybridizací se značenou sondou (*Southern blotting*) – například vizualizace pouze chloroplastové DNA
- variabilita – inserce/delece nebo mutace v restričním místě

- + vysoce reprodukovatelné pattern
- potřeba značených sond



# Příklad RFLP gelů

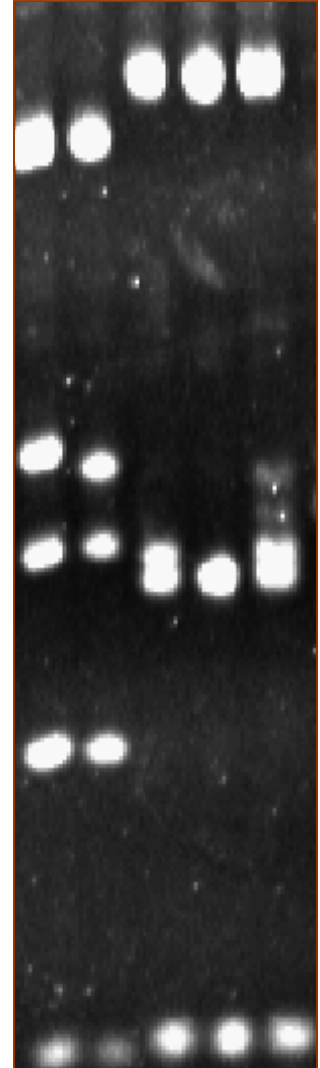


<http://www.ufpe.br/biolmol/Tec-mol-biol/RFLP-real.JPG>

# PCR-RFLP

- PCR s dvojicí specifických primerů je namnožen konkrétní úsek DNA
- využití konsenzuálních primerů (např. cpDNA)
  - použitelné takřka u všech druhů
- štěpení namnoženého úseku různými restriktčními enzymy
- elektroforéza fragmentů

- + univerzální metoda
- + kodominatní marker v případě nDNA
- nižší variabilita při použití cpDNA

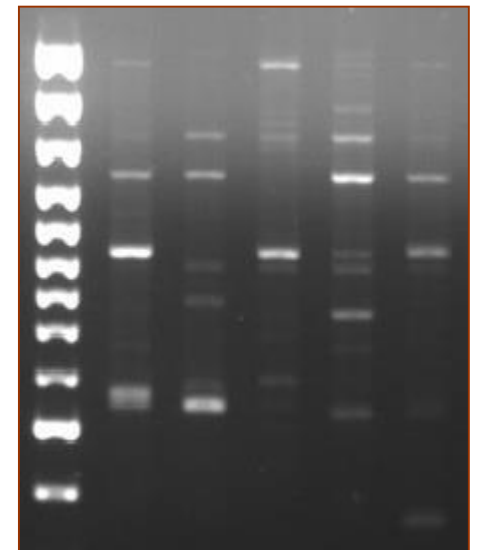


# RAPD

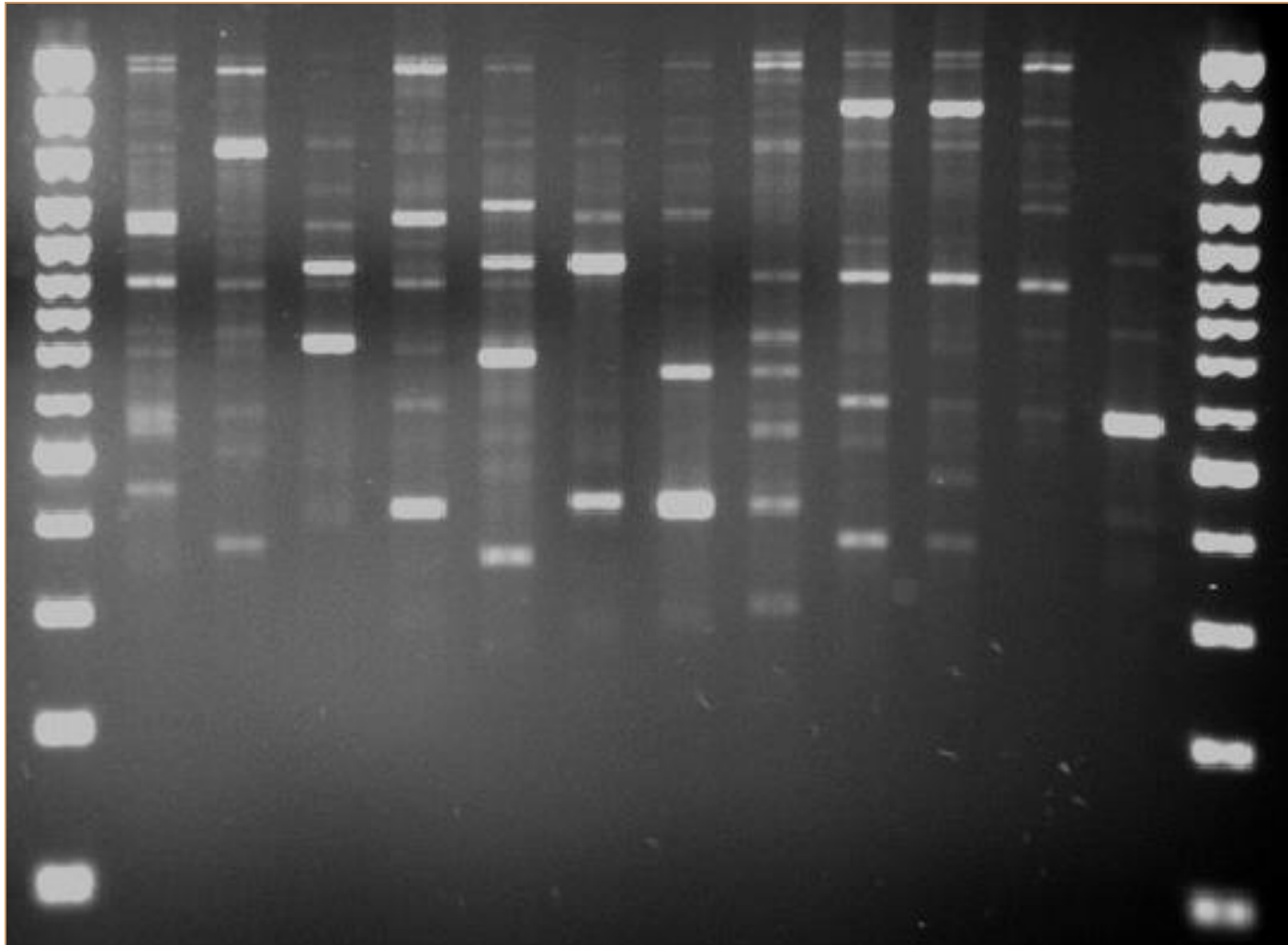
## Random **A**mplified **P**olymorphic **D**N

- generování fragmentů pomocí PCR s použitím arbitrárního primeru (dekanukleotid)
- elektroforetické rozdělení fragmentů dle délky
- polymorfismus je dán
  - mutací v místě nasedání primerů (*priming site*)
  - insercí/delecí v amplifikované části DNA

- + jednoduchá metoda
- dominantní marker
- výsledky obtížně reprodukovatelné



# Příklad RAPD gelu



# AFLP

## Amplified Fragment Length Polymorphism

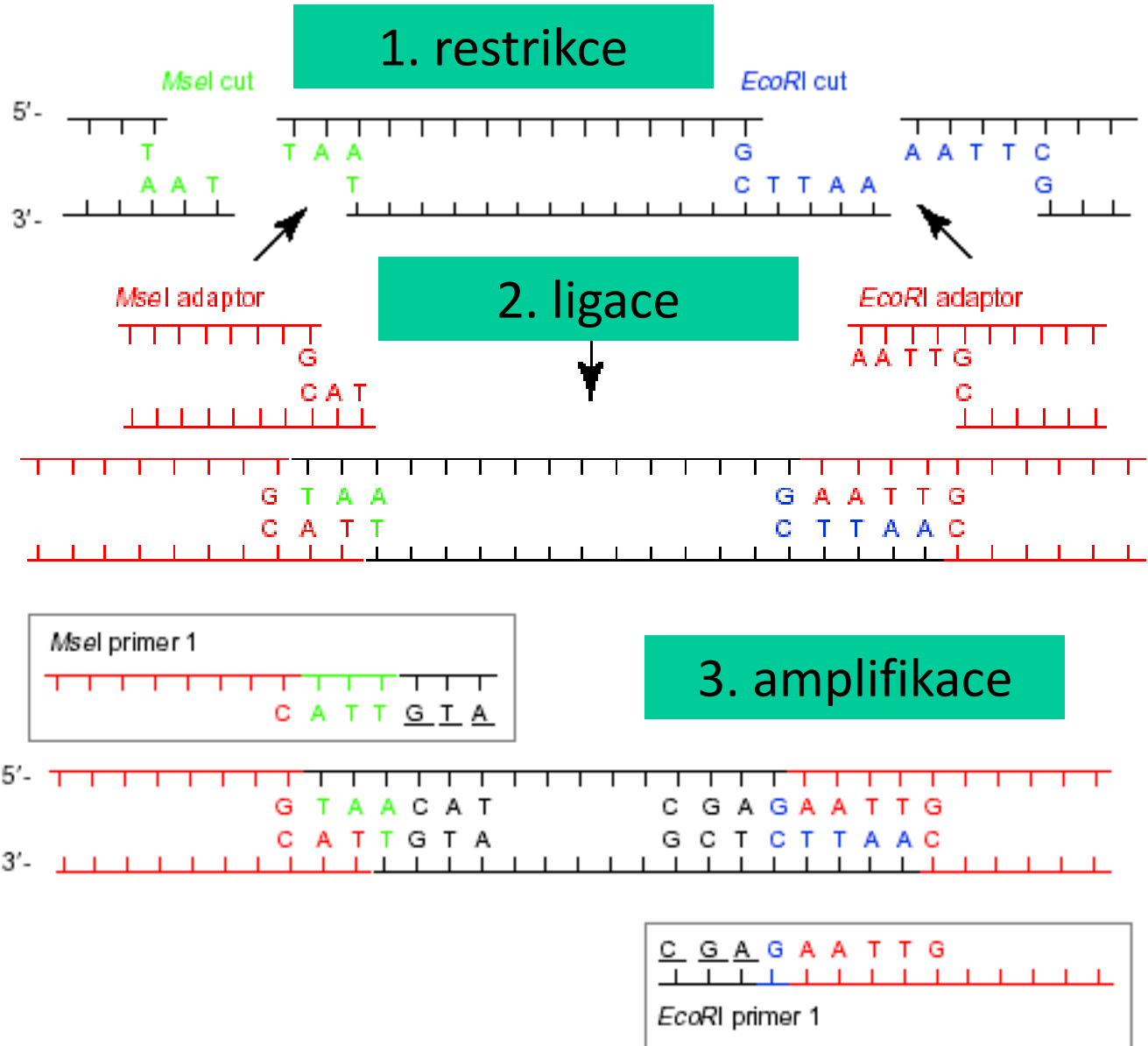
- kombinace RFLP a následné PCR určitých fragmentů
- založená na restrikci DNA dvěma enzymy
- selektivní namnožení jen některých proužků
- fluorescenční vizualizace proužků na gelu

- + velmi polymorfní, odráží variabilitu „celého“ genomu
- + vysoká reprodukovatelnost, spolehlivost
- dominantní marker – neodlišení homo- a heterozygotů

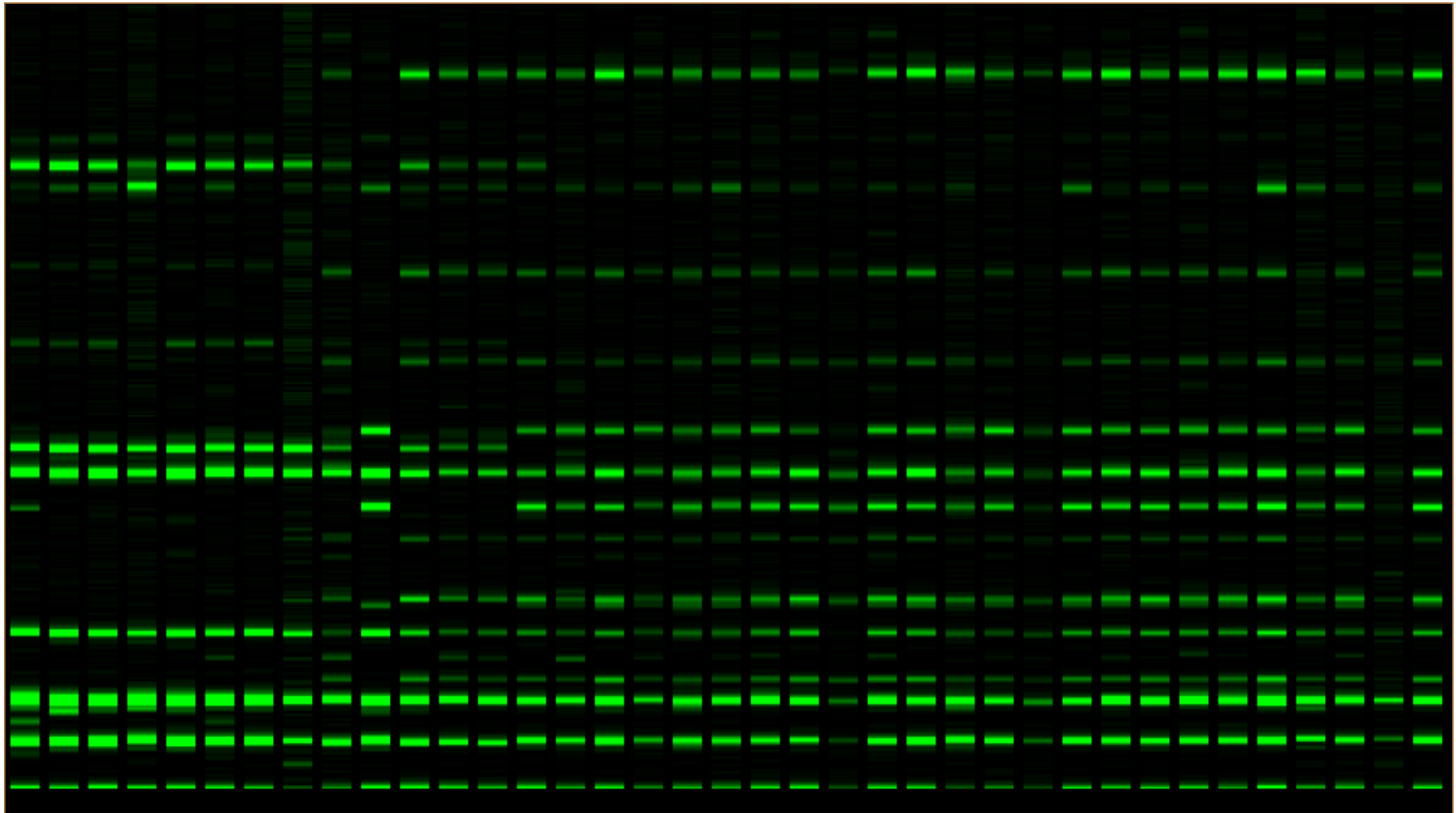
# AFLP — Amplified Fragment Length Polymorphism



celková DNA



# Příklad AFLP gelu



# Mikrosatelity

SSRs – simple sequence repeats

- opakující se několikanukleotidové sekvence

AGGATATATATATAGGCA 1

AGGATATATATA--GGCA 2

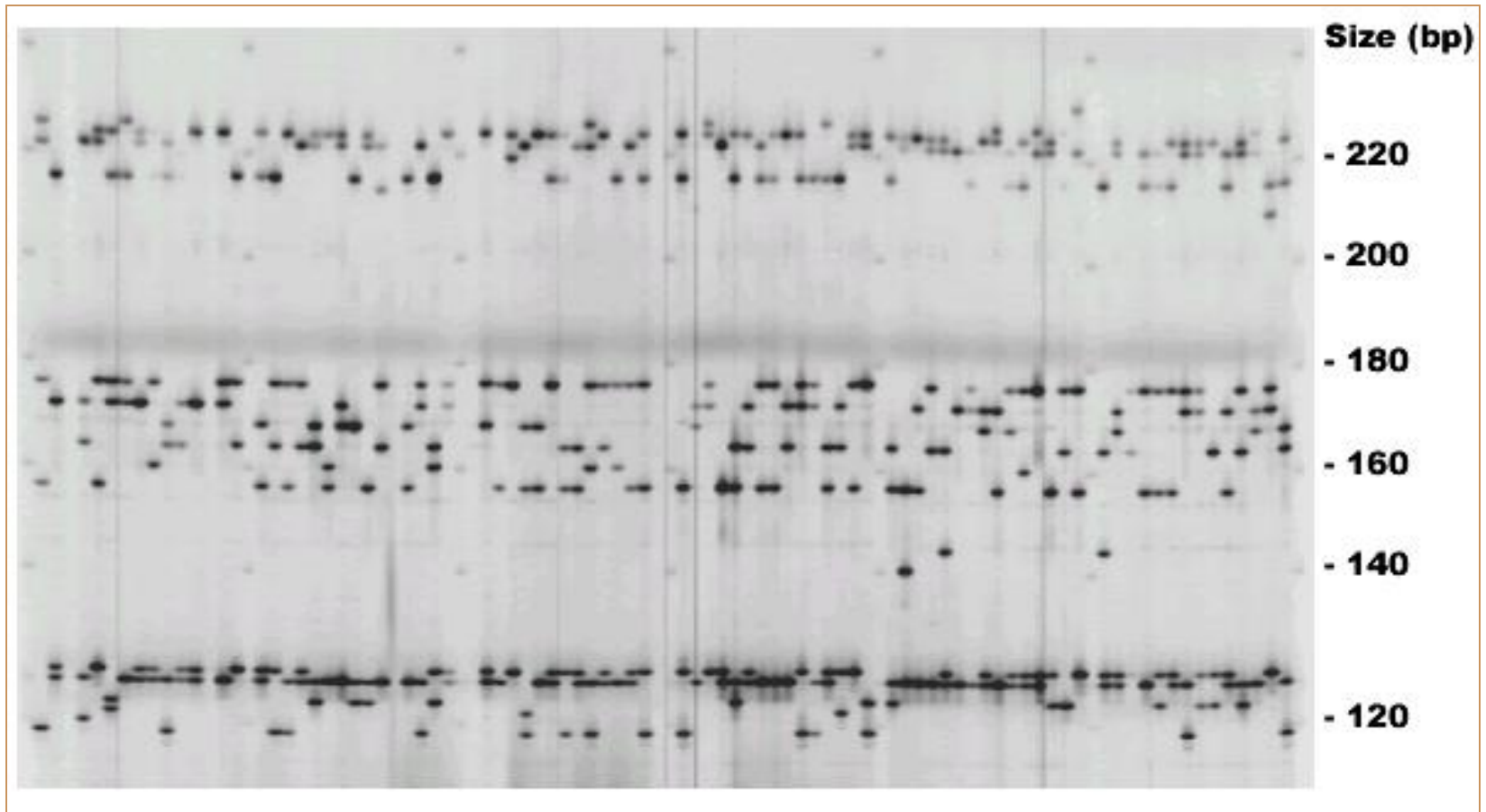


- alely – liší se počtem opakování

- + kodominantní marker, velmi variabilní
- + určitá představa o vztazích mezi alelami
- nutnost vytvořit primery pro studovaný druh

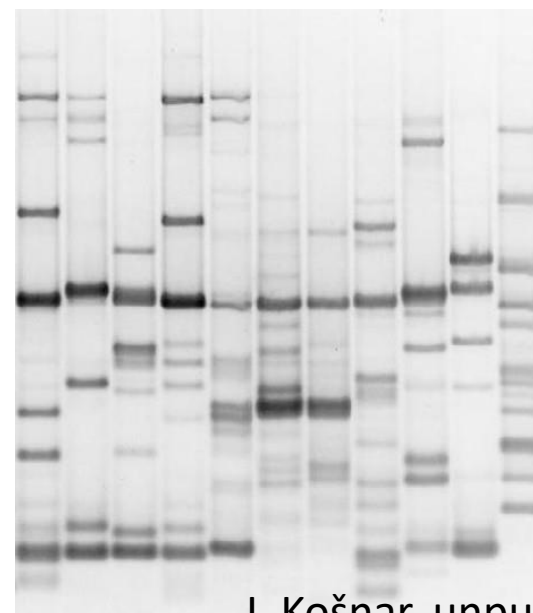


# Příklad analýzy mikrosatelitů



# ISSRs – Inter Simple Sequence Repeats

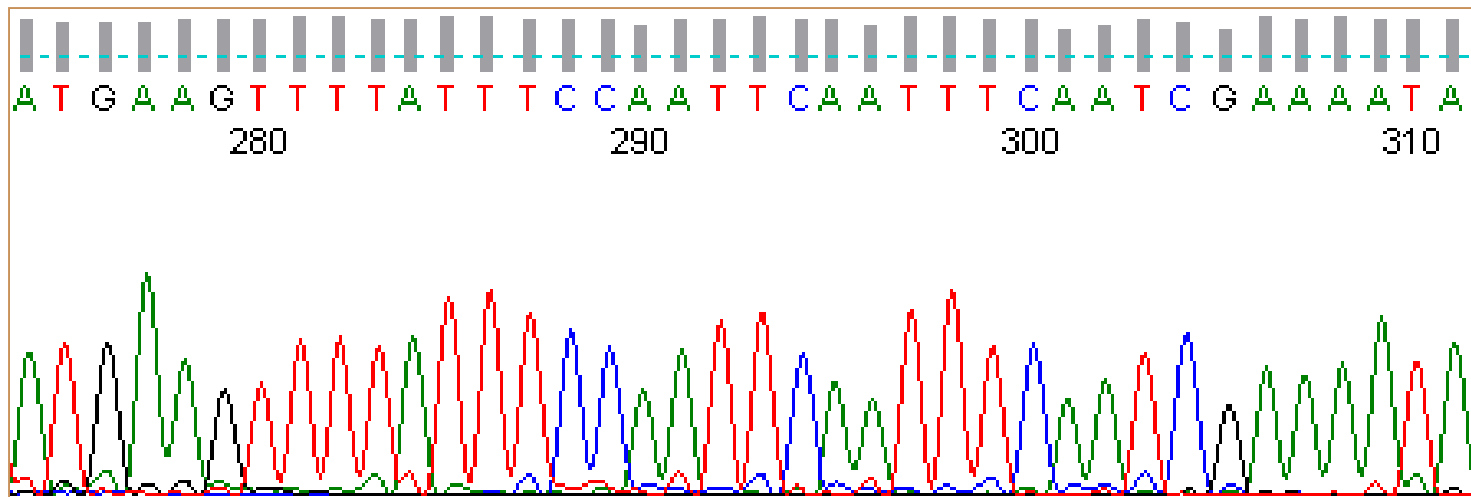
- primer – mikrosatelitová sekvence
- amplifikuje se úsek mezi mikrosatelitovými lokusy
- polymorfismus
  - změna počtu jednotek mikrosatelitu
  - indely mezi mikrosatelitovými motivy
- velmi variabilní marker
- dominantní marker
  - nutná optimalizace
  - robustnější než RAPD



J. Košnar, unpubl.

# DNA sekvenování

- determination of the sequence of nucleotides in DNA chain
- Sanger sequencing – use of automated sequencers – *fluorescence base labelling*
- specific primers for PCR amplification of the target region



# Next generation sequencing (NGS)

- **masivní** (mnoho sekvencí, až stovky milionů v jednom běhu)
- **paralelní** (současné sekvenování)
- **několik platforem**
  - **pyrosekvenování (Roche/454)** – dlouhé úseky (500-700 bp), chybovost v poly-úsecích, nízký počet sekvencí (GS FLX, GS Junior)
  - **Illumina (Solexa)** – krátké úseky (150-300 bází), high-throughput (GA II, MiniSeq, MiSeq, NextSeq, HiSeq, NovaSeq)
  - **PacBio** – single-molecule, vyšší chybovost, velmi dlouhé úseky (RS II, Sequel)
  - **Ion Torrent, Oxford nanopore (minION), ABI SOLiD, ...**

# NGS – princip a přístupy

- základní obecný postup
  - příprava knihovny – náhodné štěpení
  - prostorová separace fragmentů, imobilizace
  - klonální amplifikace/jednomolekulové sekvenování
  - sekvenování, záznam dat, analýza
- k čemu NGS?
  - sekvenování genomu de-novo
    - cílené obohacení genomu (*targeted enrichment*) – Hyb-Seq
  - re-sekvenování genomu
  - sekvenování transkriptomu (RNA-Seq)
  - ampliconové sekvenování
  - (environmentální) metasekvenování
  - RADseq (Restriction-site-associated DNA sequencing)
    - SNP – *single nucleotide polymorphism*

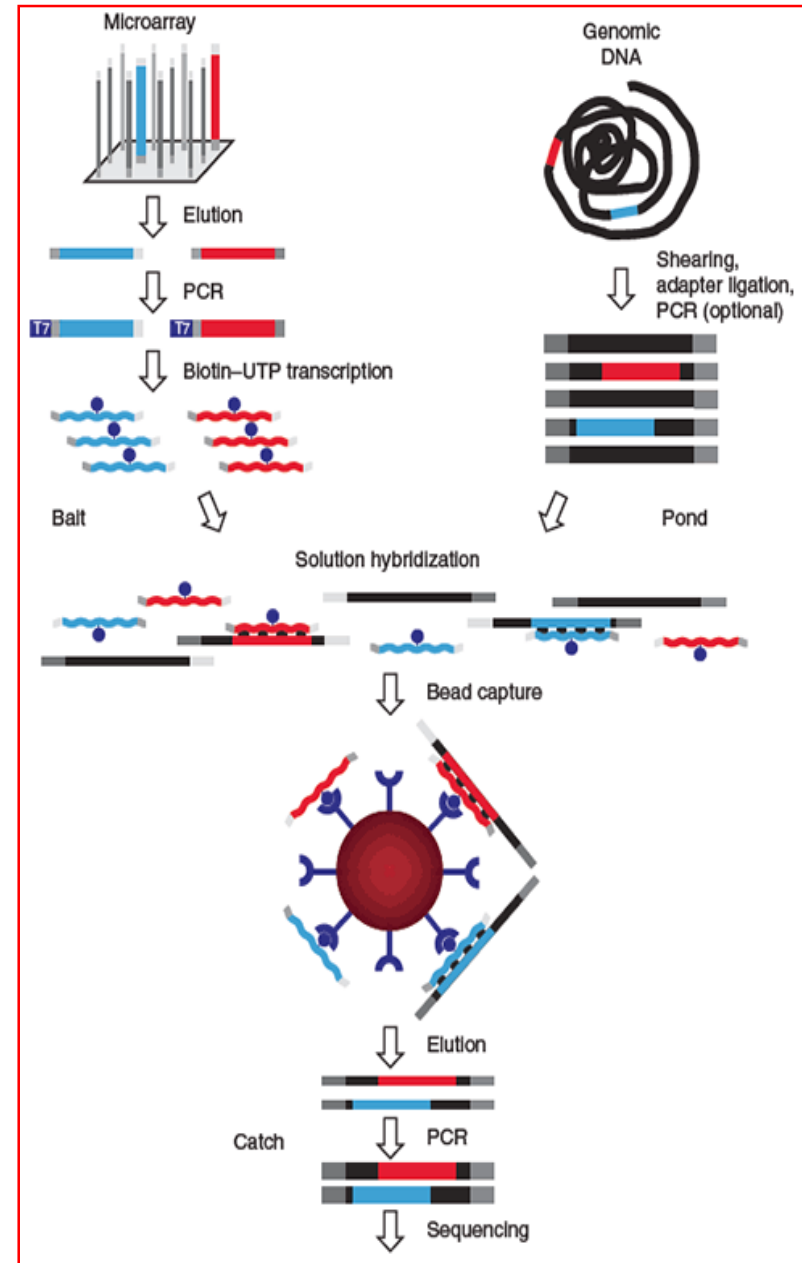
# Hyb-Seq & genome skimming

- ‘baits’ (krátké RNA fragmenty)
- hybridizace v roztoku
- obohacení knihovny o cílové sekvence
- tisíce exonů, stovky genů
- celý chloroplast, celý rDNA cistron

Weitemier et al. (2014) Appl Plant Sci. 2: apps.1400042

Cronn et al. (2012) Amer. J. Bot 99: 291-311

Straub et al. (2012) Amer. J. Bot 99: 349-364

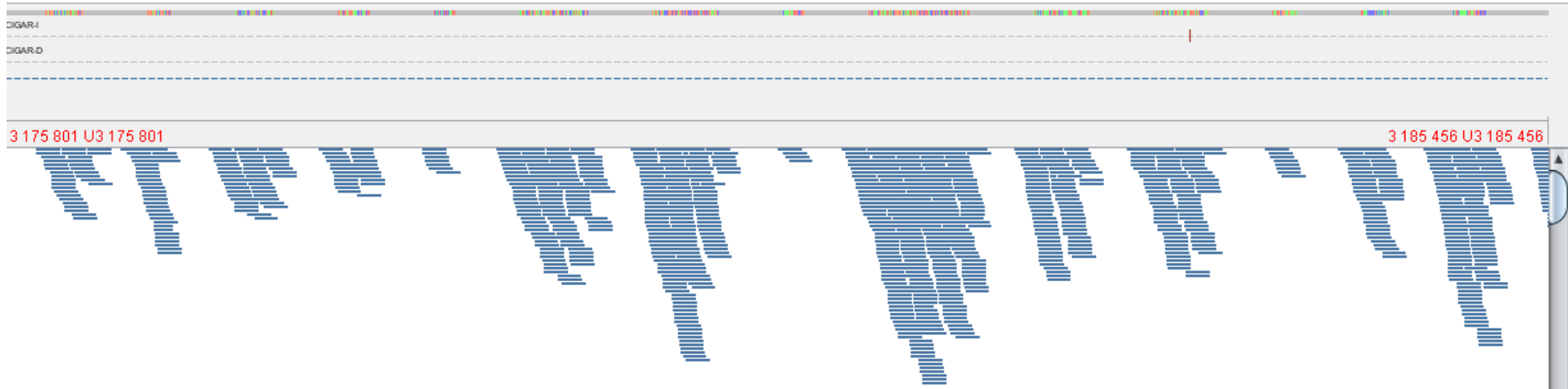


Microarray

# Hyb-Seq – reads mapped to reference

1 to 3 418 800 (3,4 Mbp)

3 175 801 to 3 185 457 (9,7 Kb)



3 175 801 U3 175 801

3 185 456 U3 185 456



# Využití markerů pro různé okruhy otázek

	Allo- zomy	Fragment-based			Sekvenování				NGS		
		RAPD	AFLP	SSR	nDNA	cpDNA	mtDNA (rostliny)	mtDNA (zvířata)	Hyb-Seq	RADseq	resekvenování
Genetická diverzita	++	++	++	++	+++	++	+	++	+++	+++	+++
Diferenciace pooplací	+++	++	++	++	+++	++	++	+++	++?	+++	+++
Genový tok	++	(+)	(+)	+++	+++	++	(+)	++	?	+++	+++
Polyploidizace	+++	-	(+)	+	++	++	-	-	+++	++	+++
Hybridizace	++	++	++	+	++	++	+	+	+++	+++	+++
Fylogeneze	(+)	-	++	(+)	+++	+++	(+)	+++	+++	++	+++
Genotypování jedinců	(+)	+++	+++	+++	+++	-	-	-	?		+++
Fylogeografie	(+)	-	++	-	(+)	+++	(+)	+++	(+)	+++	+++
Selekce	(+)	(+)	(+)	+	++	-	-	-	++	++	+++
Diverzifikace	?	?	(+)	-	++	++	?	++	+++	+++	+++

+++ velmi vhodné (+)

++ dobře použitelné -

+ OK ?

bylo použito

nevhodné

nejisté nebo nepoužito

první část podle Lowe et al. 2004



# Hybridizace

- velmi častý fenomén
- významný proces při vzniku nových druhů
- mezidruhov<sup>á</sup>, ale i mezirodov<sup>á</sup>
- jaderná DNA
  - identifikace informace obou rodičovských genomů v genomu hybrida
  - intenzita a pattern “promíchání genomů”
- chloroplastová DNA – který z rodičů je mateřský?

# Metody používané ke zjištění hybridizace

- PCR-RFLP ITS
- sekvenování ITS (+ klonování)
- srovnání nDNA a cpDNA fylogeneze
- AFLP – sdílené fragmenty, pravděpodobnostní modelování
- mikrosatelity – odlišení F1 a pokročilých hybridů (F2, B1...)
- RADseq, re-sequencing – whole genome SNPs

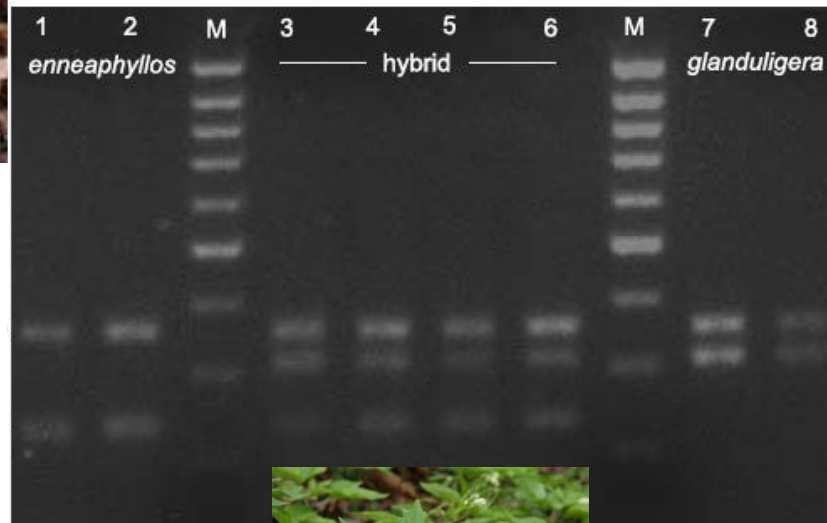
# Hybridizace – PCR-RFLP



*Dentaria enneaphylos*



*Dentaria glandulosa*



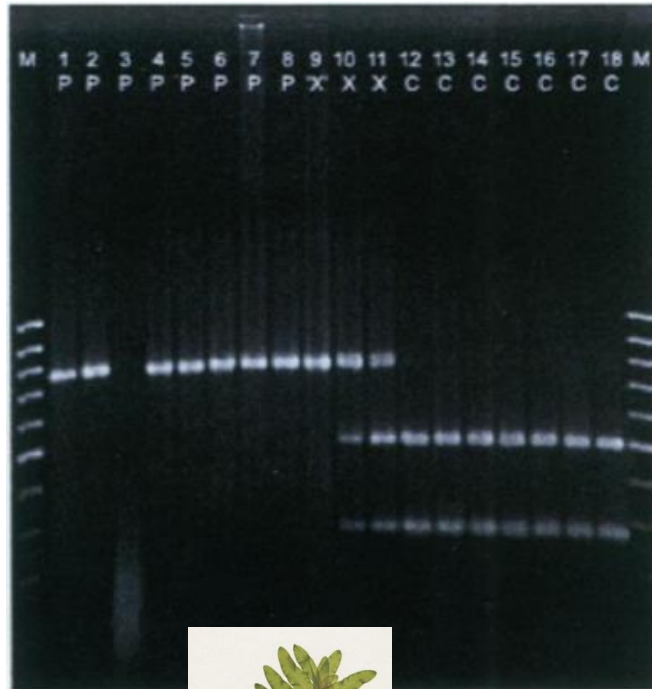
*Dentaria x paxiana*



# Hybridize – PCR-RFLP



*Potamogeton perfoliatus*



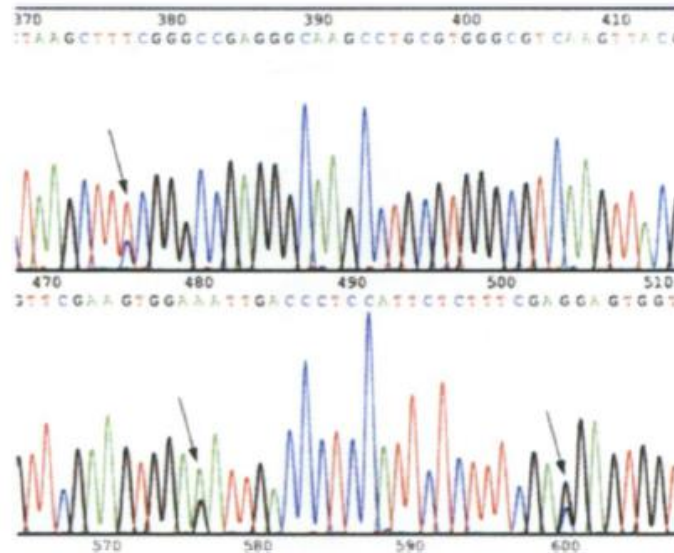
*Potamogeton crispus*



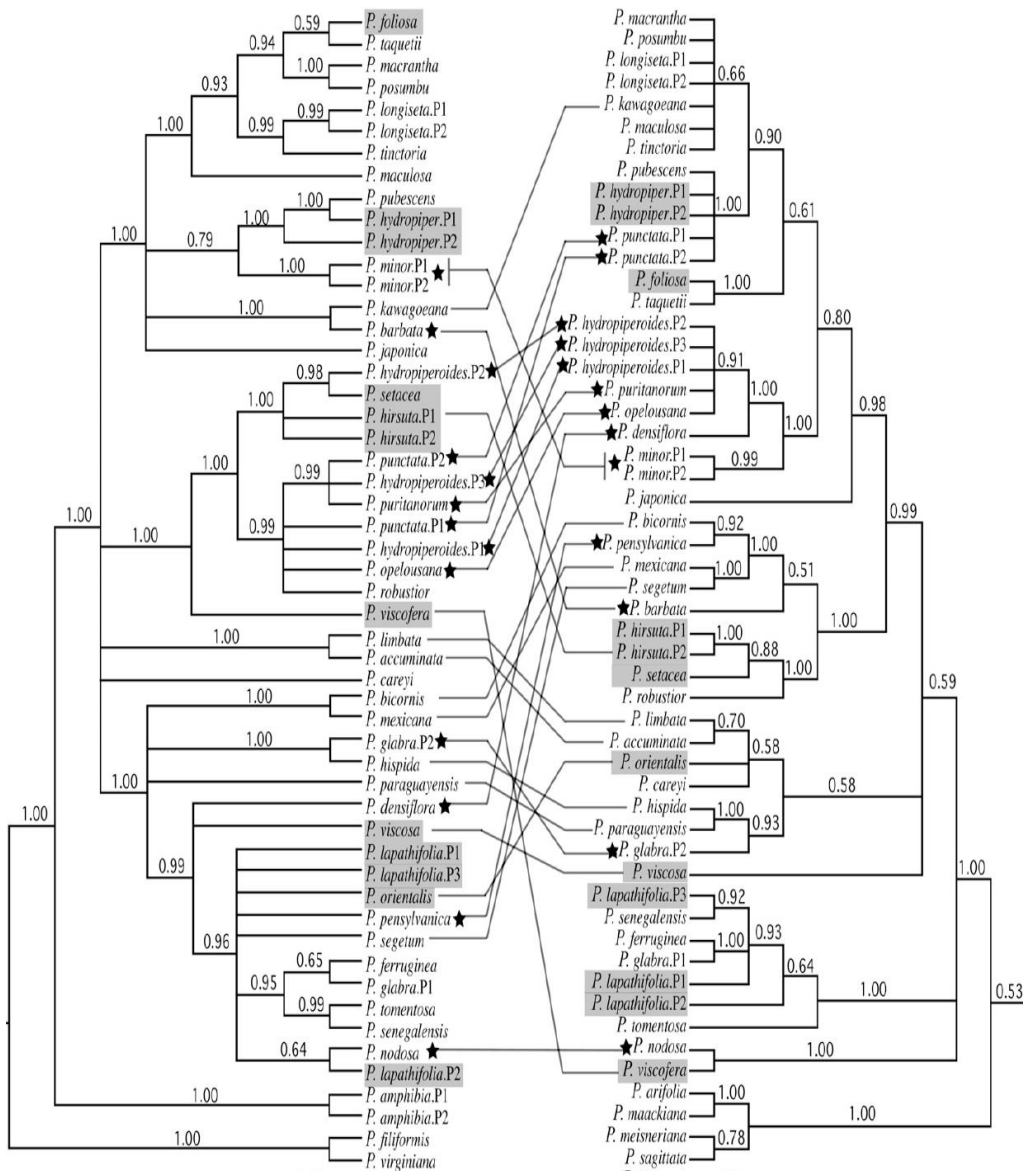
*Potamogeton x cooperi*

# Hybridizace – sekvenování ITS

- „*double peaky*“ v přímých sekvencích



# Srovnání nDNA a cpDNA fylogeneze



rozpor mezi cpDNA a nDNA

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

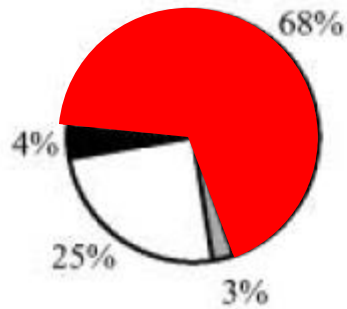




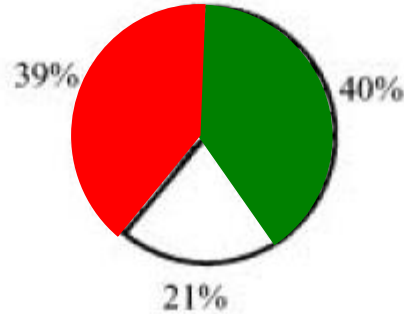
# Hybridizace – AFLP



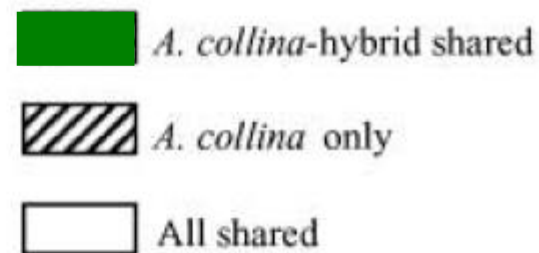
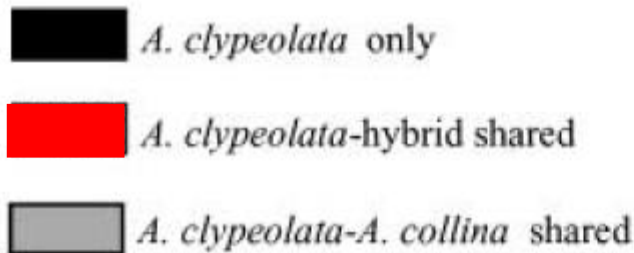
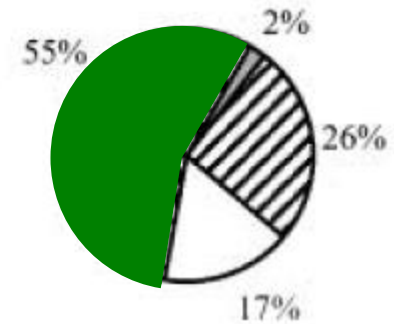
*A. clypeolata* \_40



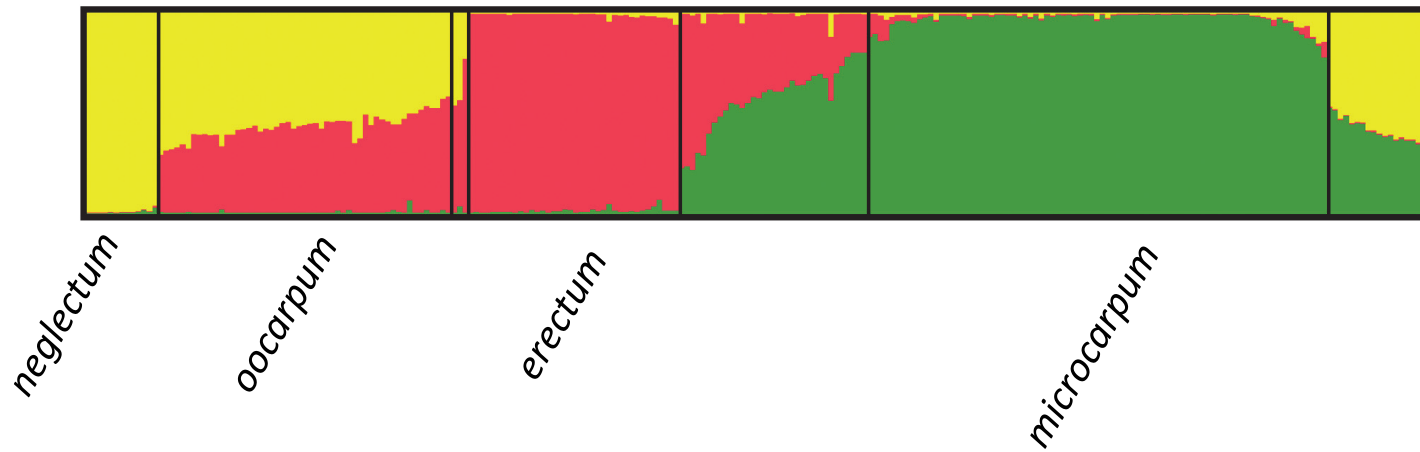
*A. clypeolata* × *A. collina* \_42



*A. collina* \_41



# Hybridizace – AFLP



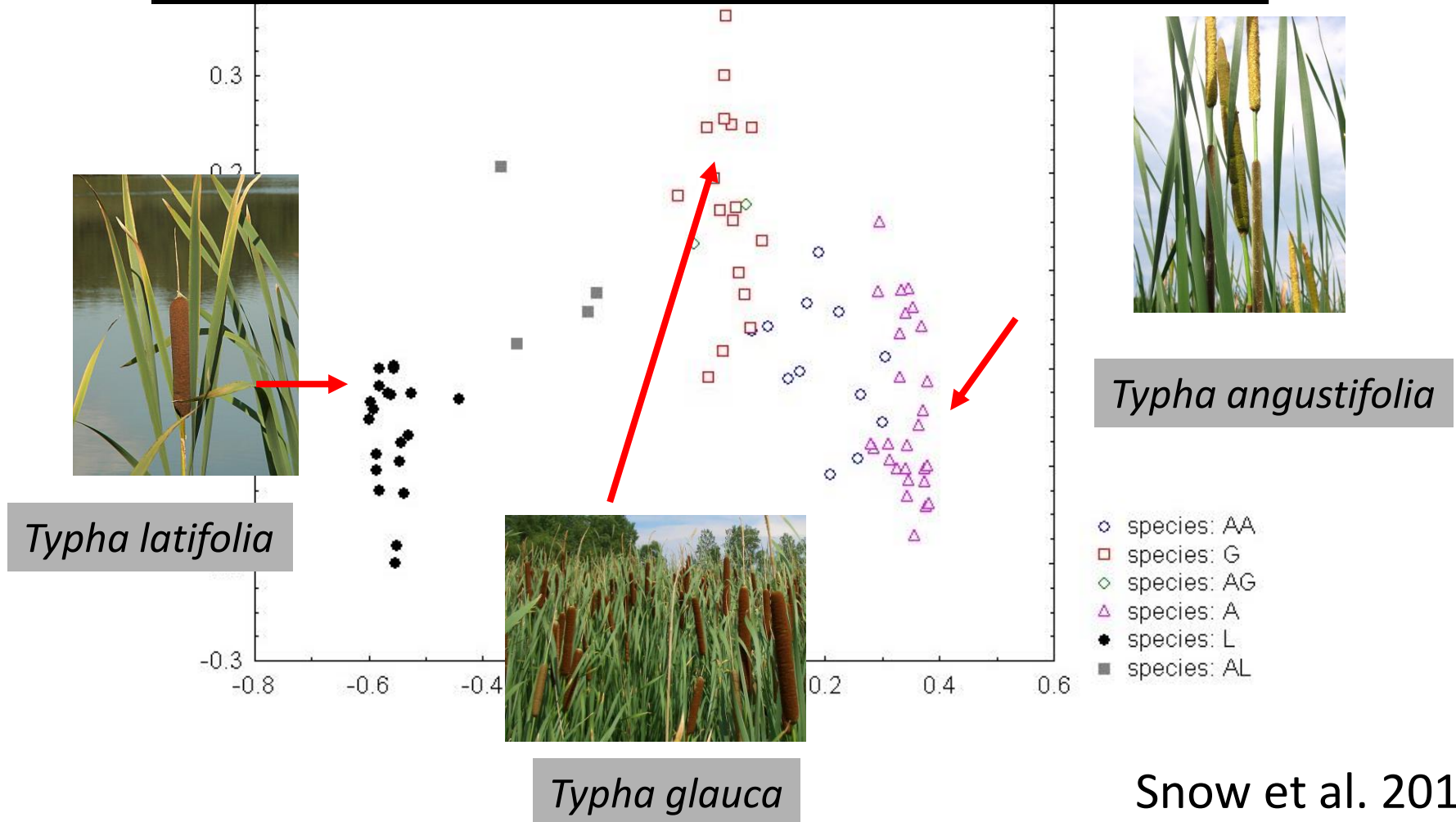
*Sparganium erectum*



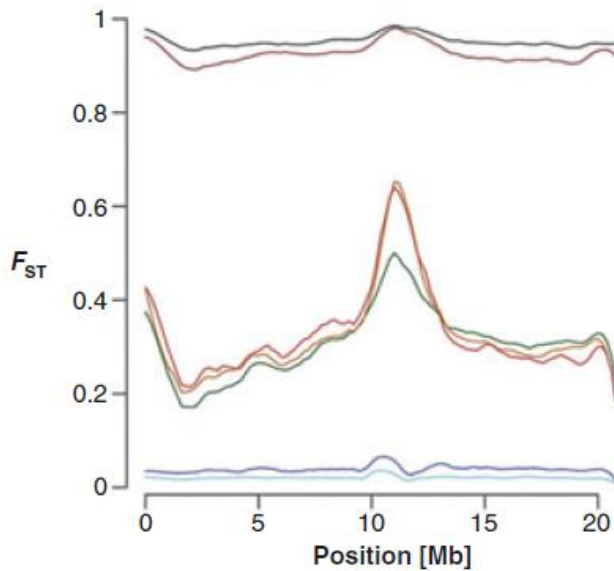


# Hybridizace – mikrosatelity

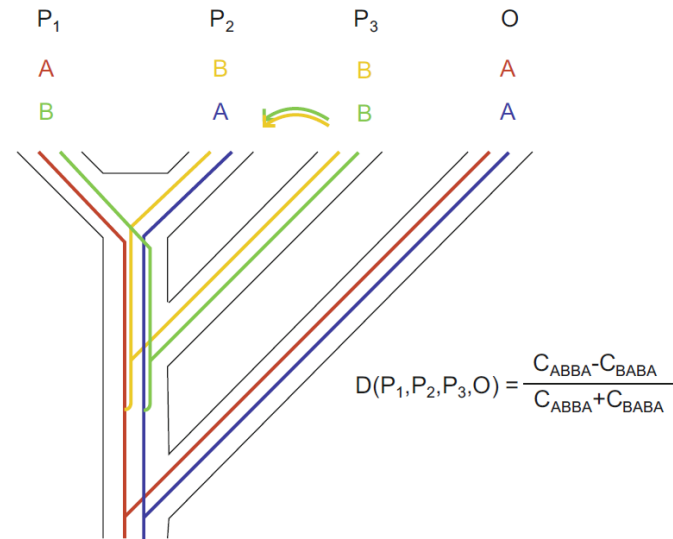
T. latifolia	176	176	278	278	176	190	269	269	179	179	93	93	278	278
T. angustifolia	210	210	286	286	196	196	287	287	193	193	101	101	280	280
T. xglauca	180	210	278	286	190	196	269	287	179	193	93	101	278	280
advanced hybrid	176	210	278	286	190	196	287	287	179	193	93	101	278	280



# Hybridizace – genomické studie

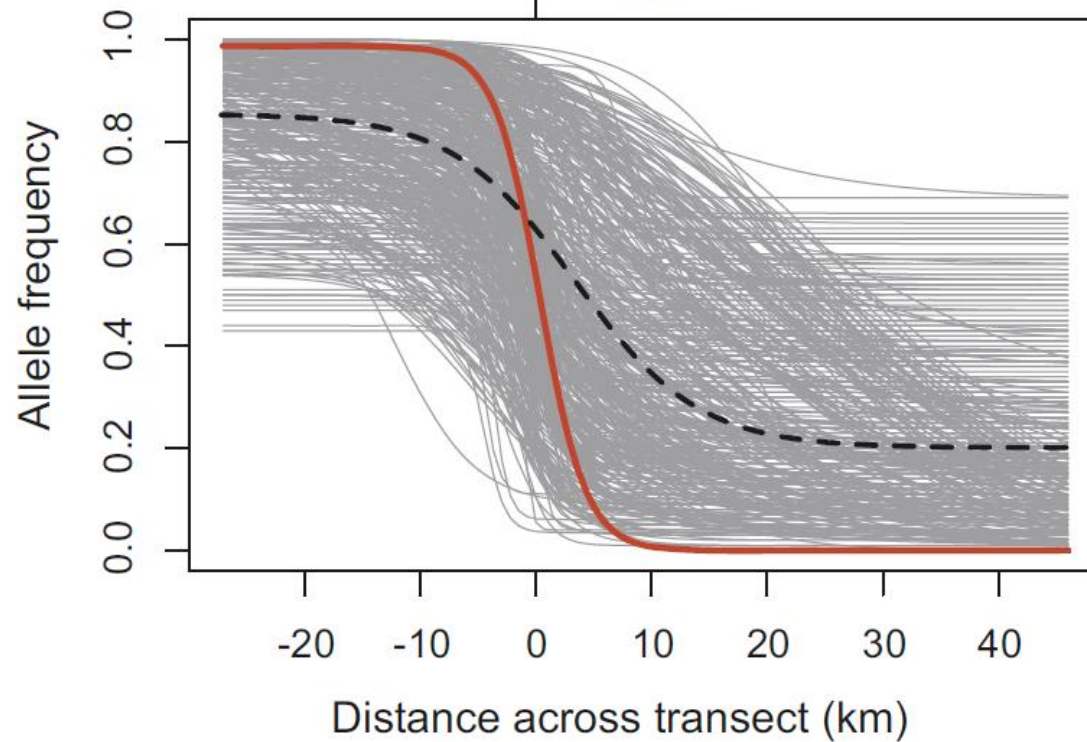


differentiation islands



ABBA-BABA D-statistics

# Hybridizace – genomické studie



*Mimulus aurantiacus*  
hybridní zóna, 426 lokusů

Stankowski et al. 2017

# Fylogenetické studie

- fylogeneze krytosemenných
- fylogenetické studie čeledí
- studie rodů a blízce příbuzných druhů
- druhové komplexy, kryptická diverzita
  
- ancestrální, odvozené skupiny
- monofylie, polyfylie, parafylie
  
- vztah fylogeneze a morfologie/ekologie/velikosti genomu/biologie ad.
- barcoding

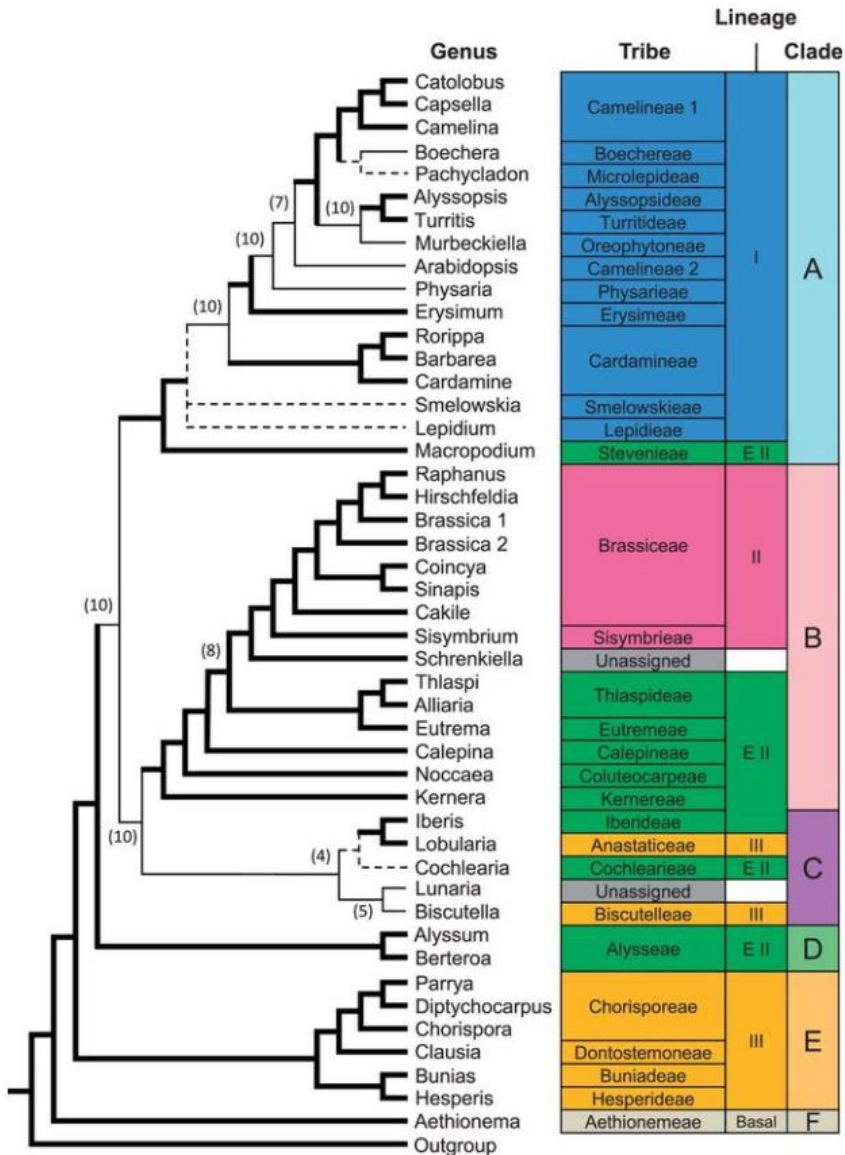
# Metody používané k rekonstrukci fylogeneze

- sekvence
  - ITS, low-copy, cpDNA, mtDNA
  - kodující × nekodující
- AFLP/RAPD/ISSR
- cpDNA mikrosatelity
- izozymy?
- RADseq (SNPs)
  
- stromy
  - neighbour-joining (NJ)
  - maximum parsimony + bootstrap support (MP+BS)
  - maximum likelihood (ML)
  - Bayesian inference + posterior probability (BI+PP)
- síť (např. neighbour-net)





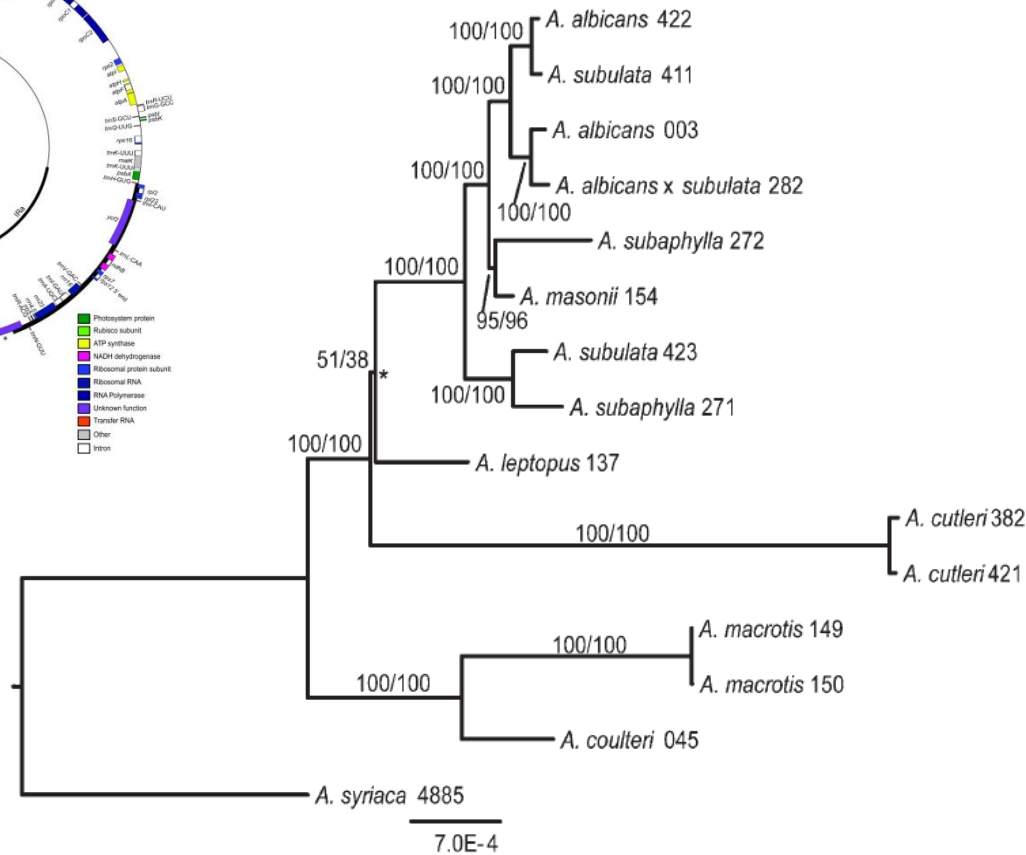
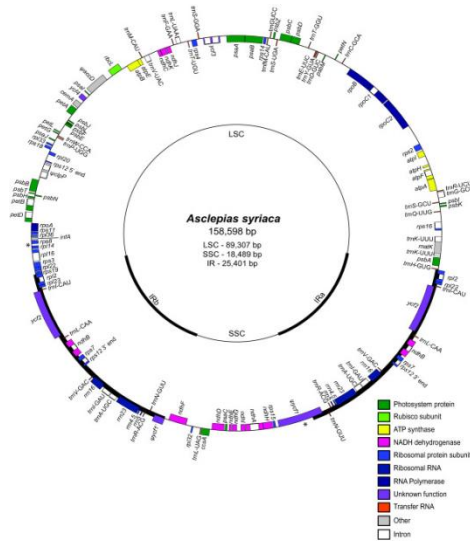
# Fylogeneze čeledi



*Brassicaceae*

113 jaderných genů  
transkriptom

# Sekvenování celých chloroplastů

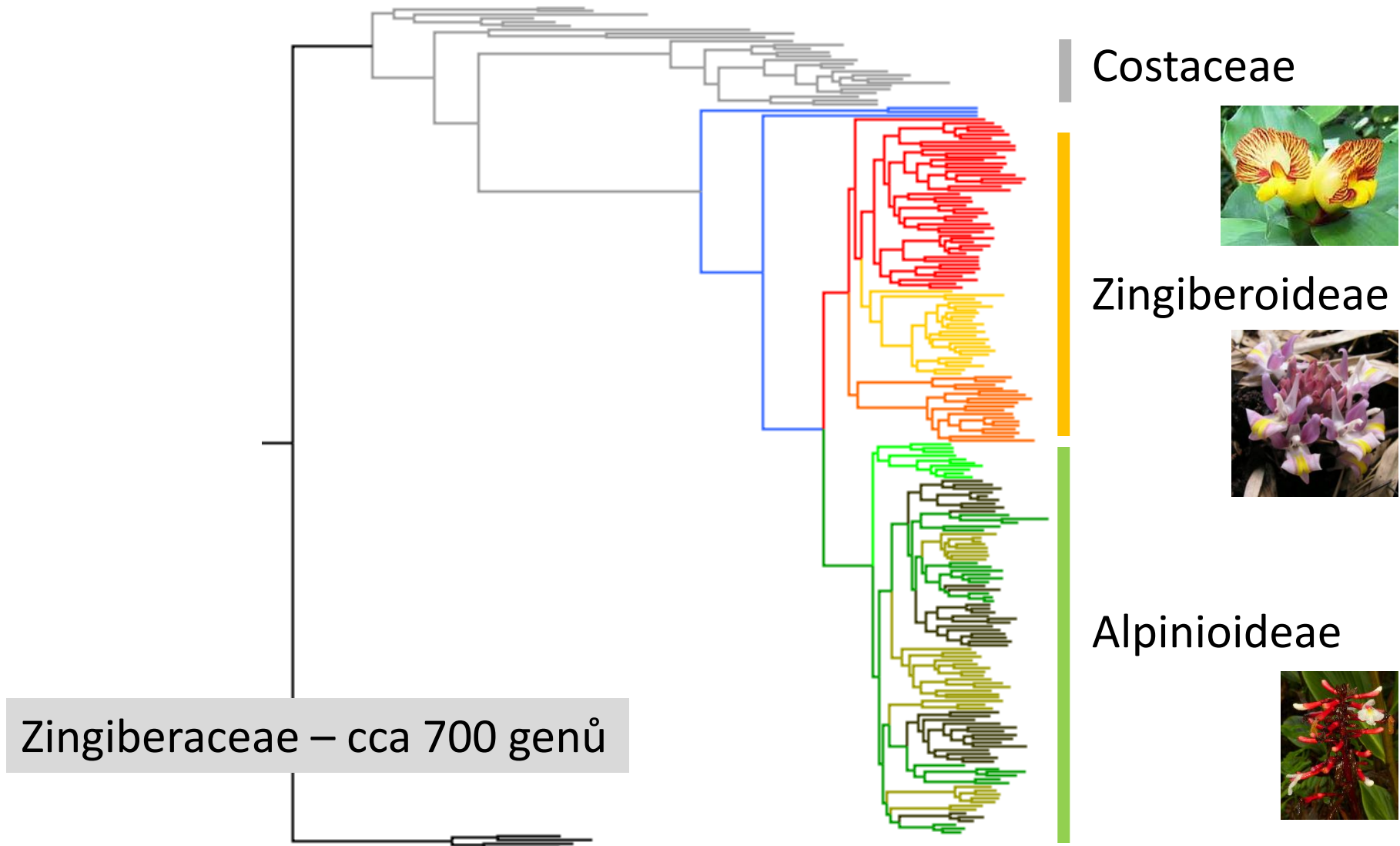


*Asclepias*

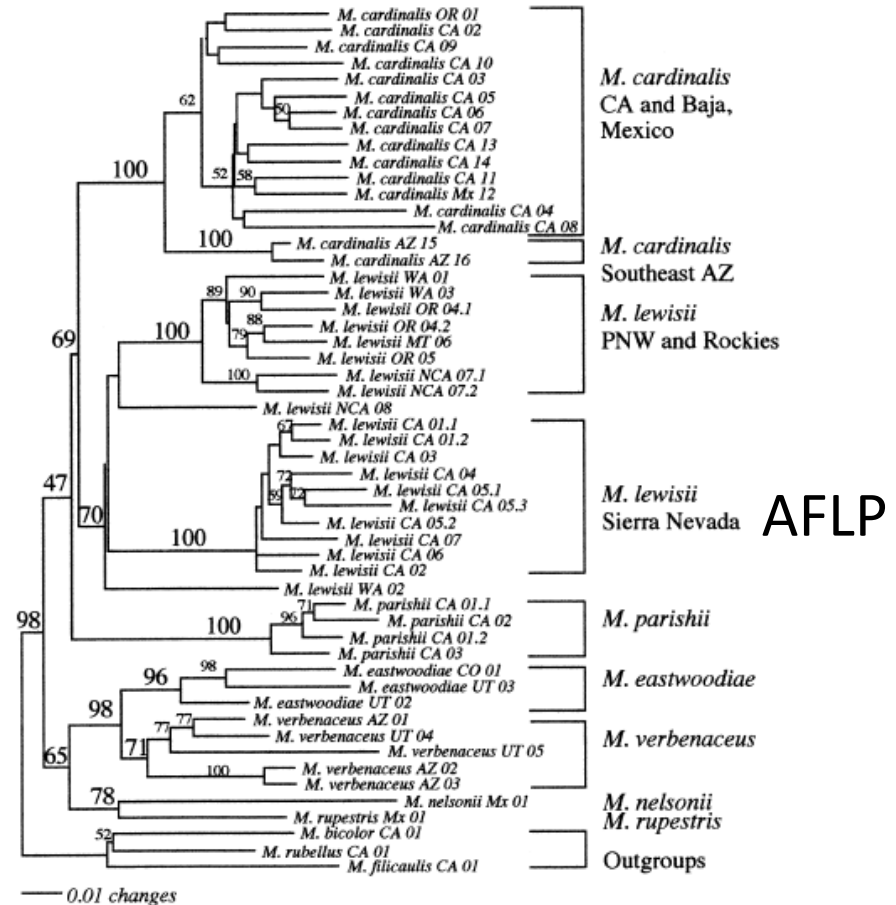
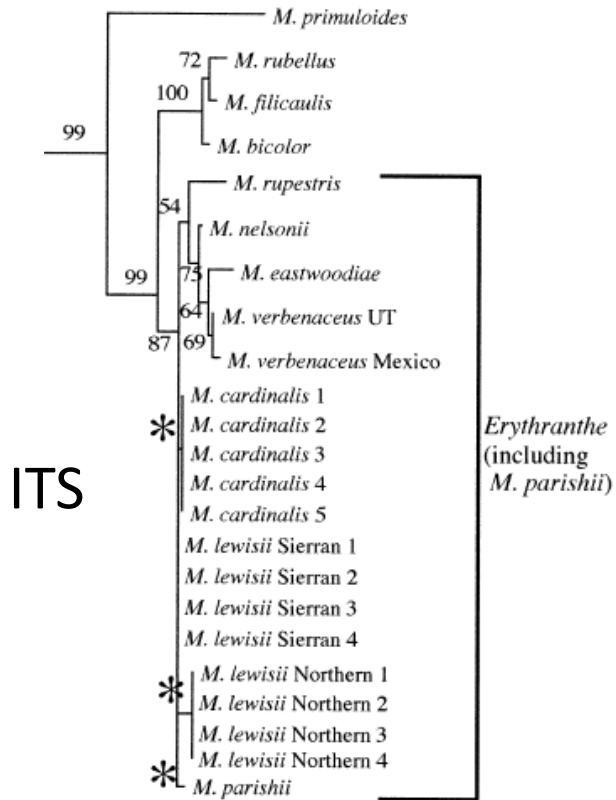
Straub et al. (2012): *Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics*. American Journal of Botany 99: 349–364.



# Hyb-Seq – sekvenování stovek genů



# Fylogeneze ITS vs. AFLP

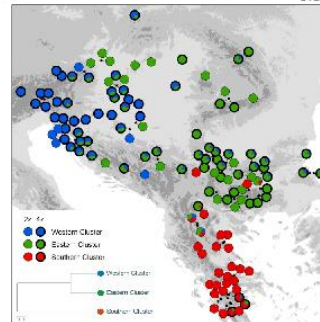
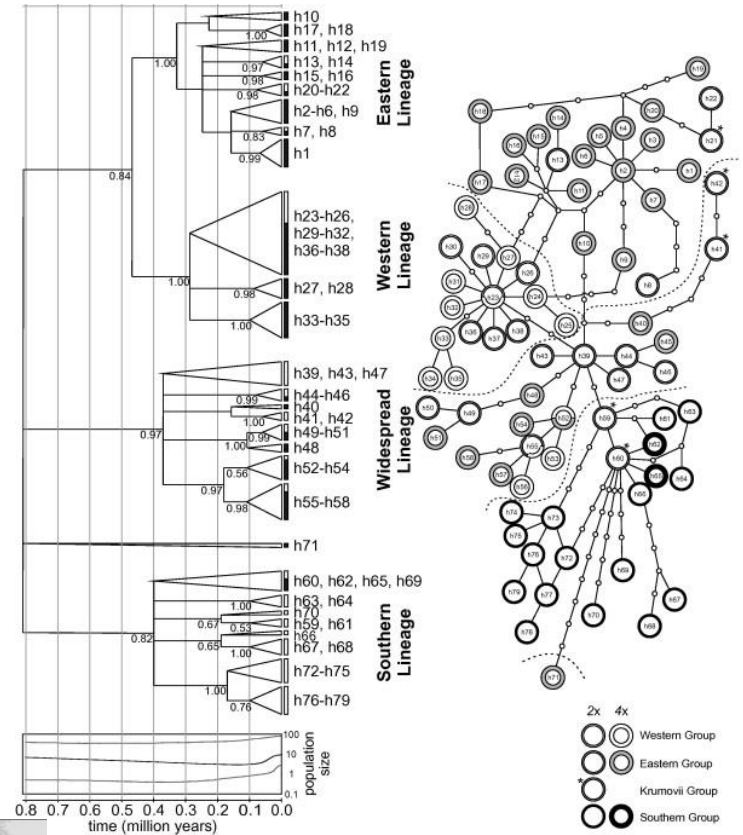
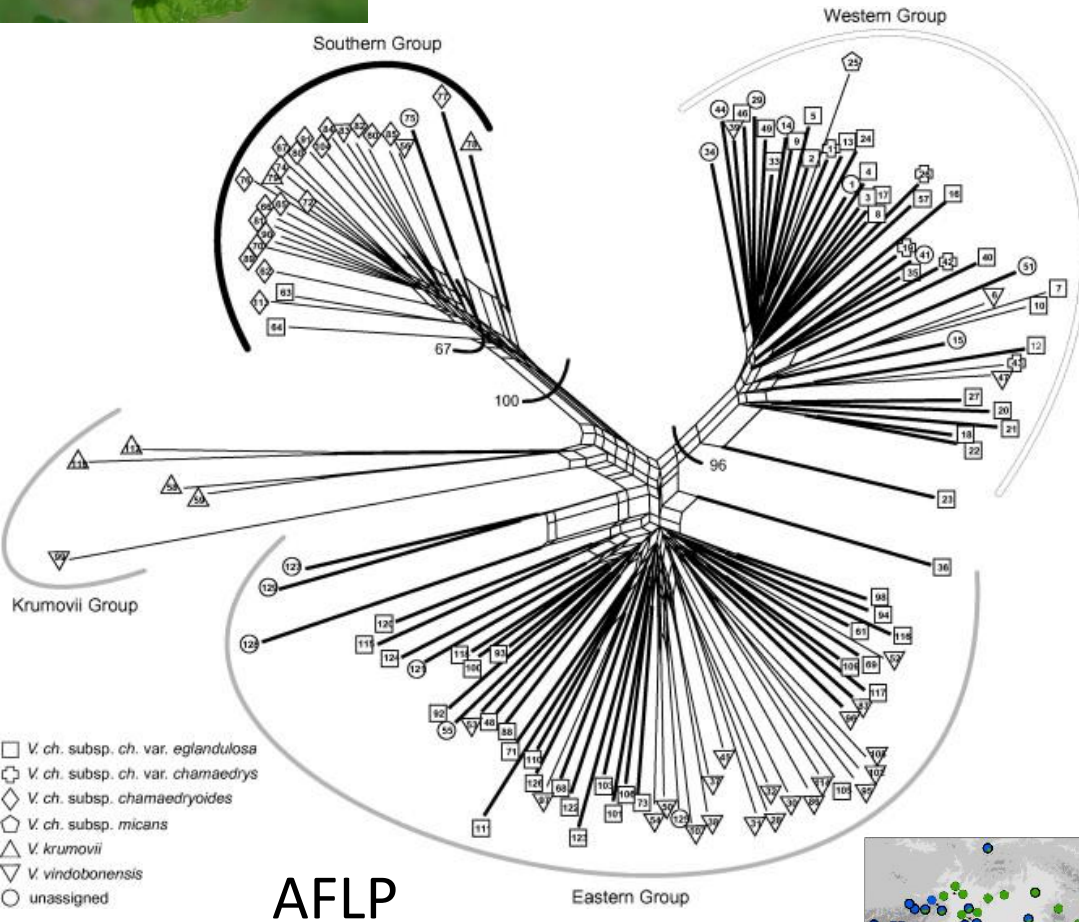


druhový komplex – *Mimulus* sect. *Erythranthe*

Beardsley et al. 2003



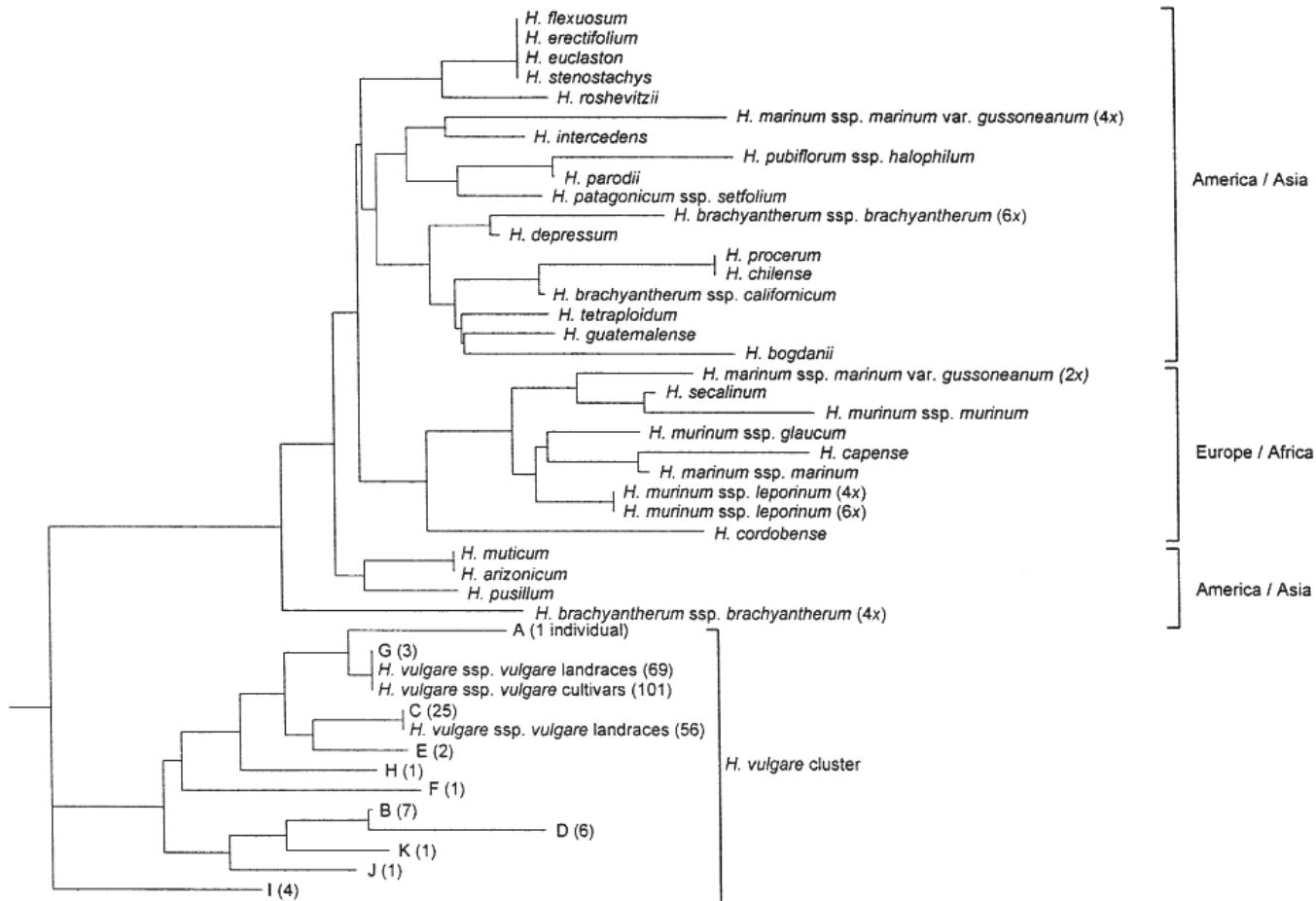
# Fylogeneze – sít'



komplex *Veronica chamaedrys*

Bardy et al. 2010

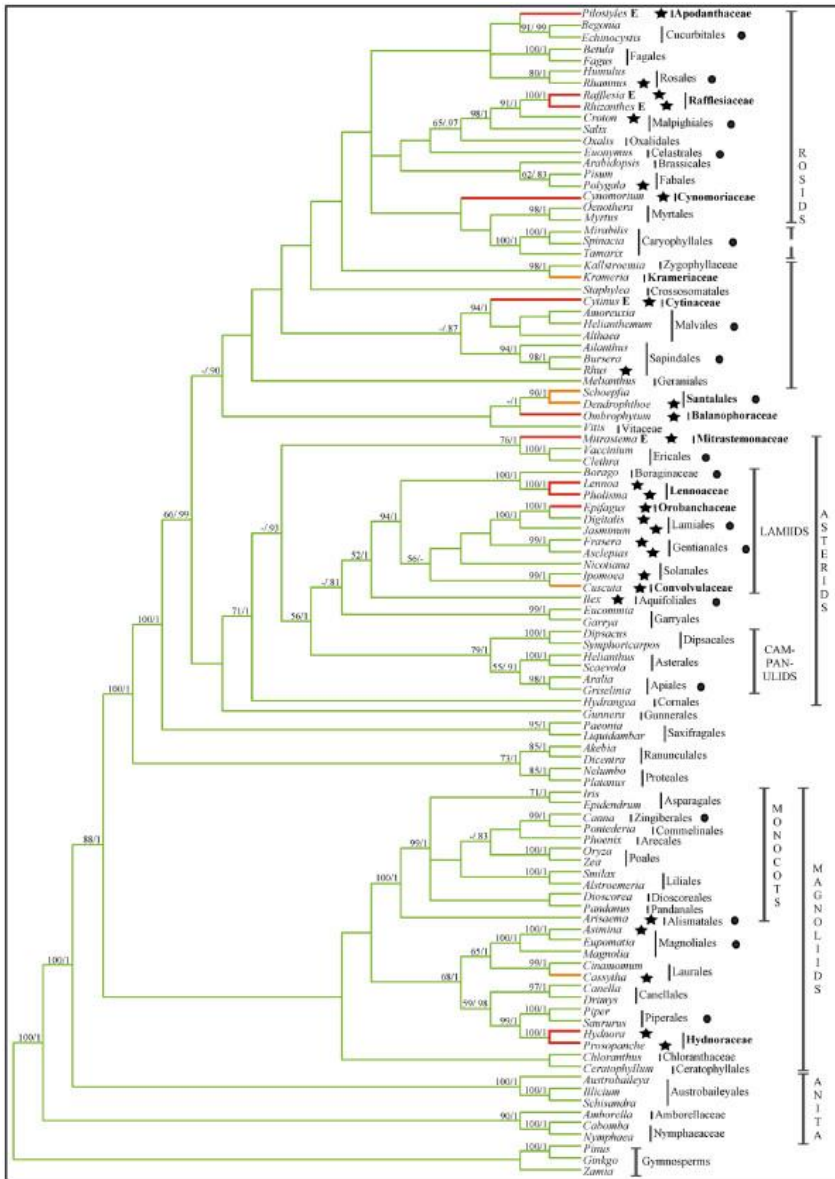
# Fylogeneze – cpDNA mikrosatelity



*Hordeum*



# Fylogeneze – mtDNA



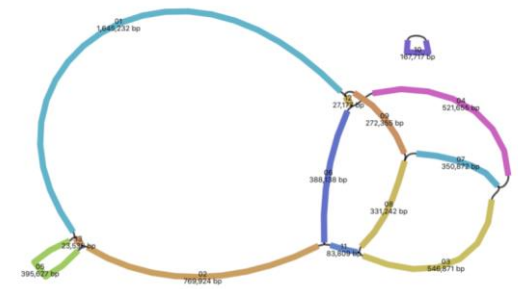
Barkman et al. 2007

3 mtDNA geny

- *atpl*
- *cox1*
- *matR*

krytosemenné

11 vzniků parazitismu

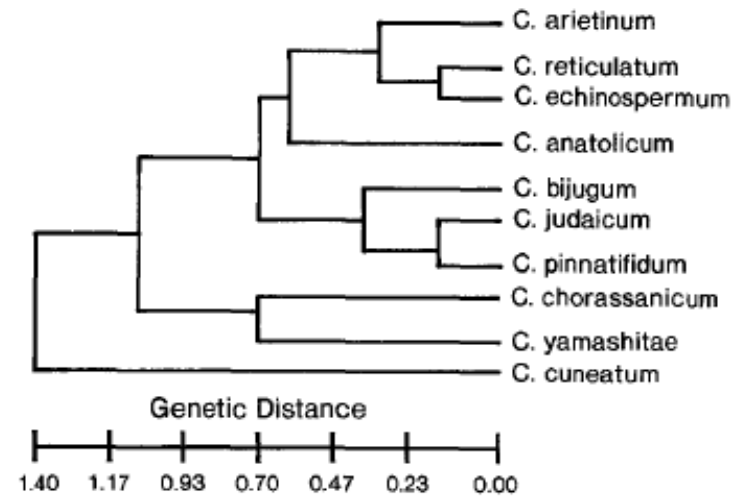
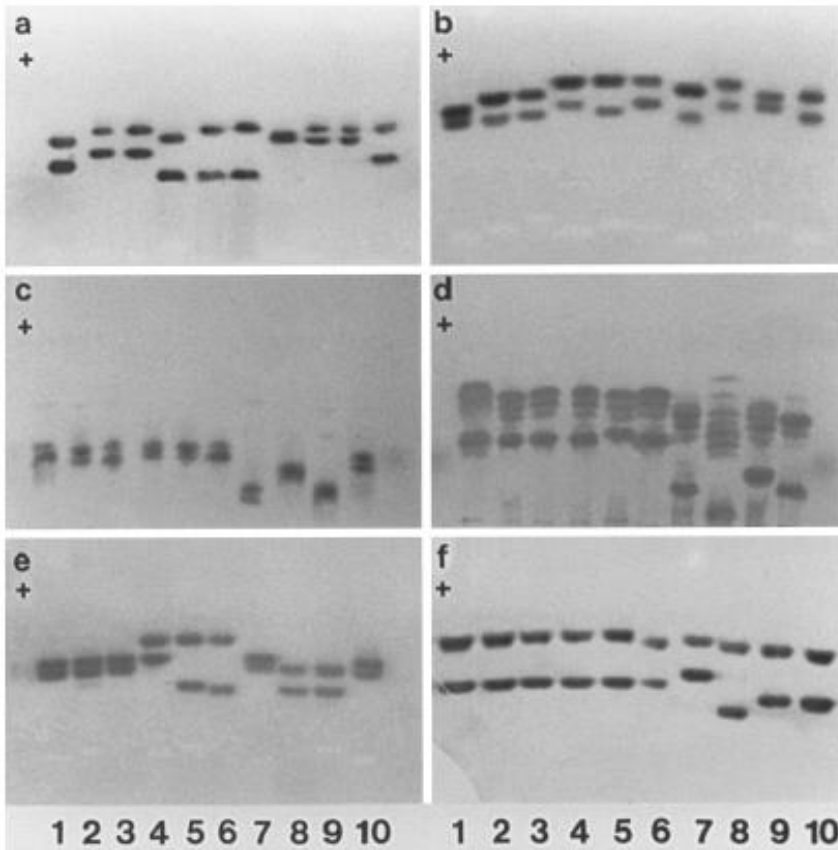


*Picea sitchensis*

5.52 Mbp, 13 segmentů

Jackman et al. 2019

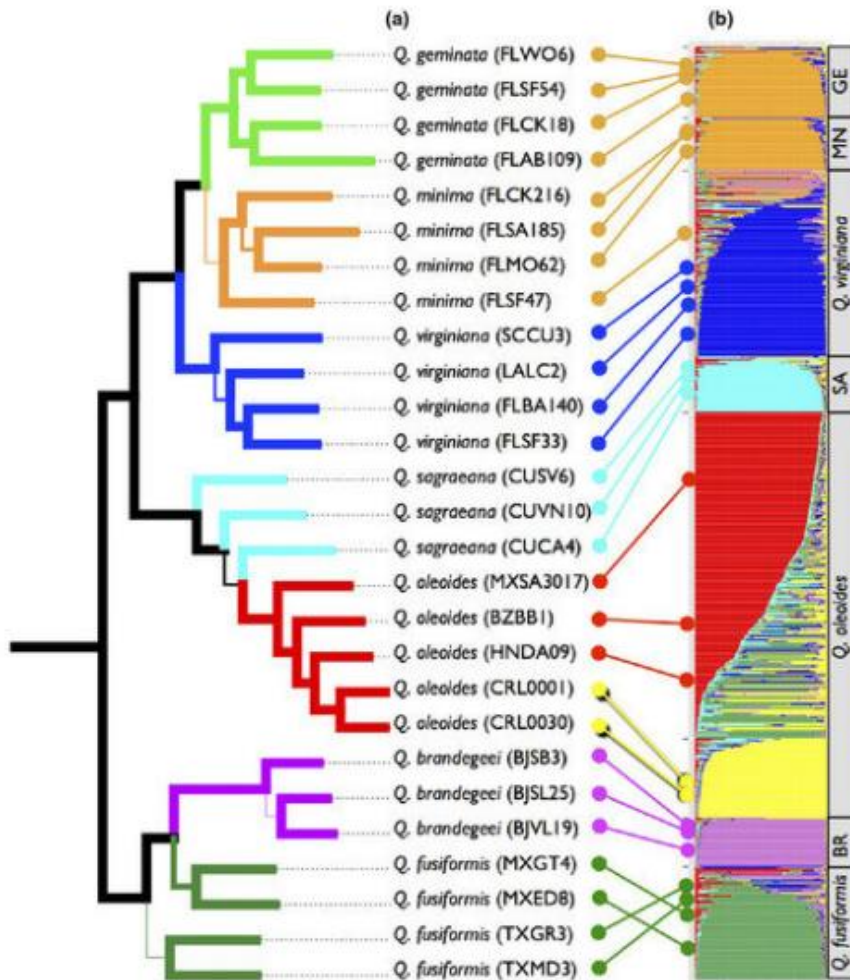
# Fylogeneze – isozymy



*Cicer arietinum* (cizrna)

Kazan & Muehlbauer 1991

# Fylogeneze – RADseq



ML fylogeneze vs. SSRs

*Quercus* subsect. *Virentes*

Cavender-Bares et al. 2015



# Barcoding

- unikátní databázovaná informace o organizmu určená k jeho jednoznačné identifikaci
- univerzální, dostatečně krátký, dobře sekvenovatelný, dobré rozlišení mezi druhy
- živočichové – cytochrome c oxidase I (mtDNA)
- rostliny – *rbcL*, *matK*, (*trnH-psbA*)
  
- CBOL – Consortium for the Barcode of Life
- [http://www.barcoding.si.edu/plant\\_working\\_group.html](http://www.barcoding.si.edu/plant_working_group.html)



# Polyploidizace, fylogeografie...

- příště... (27.11. 2023)