

Využití molekulárních markerů v systematice a populační biologii rostlin

7. Mikrosatelity

Typy mikrosatelitů

- jednoduché (*simple*)

...CACACACACACACACACACA...

- složené (*compound*)

...CACACACACATGTGTGTGTGTG...

- přerušované (*interrupted*)

...CACACA~~TT~~CACACA~~TT~~CACA...

Repetitivní sekvence

- dinukleotidy
 - nejběžnější u rostlin AT opakování
 - cca každých 30-50 kb
 - počet jednotek je do 30
- trinukleotidy
 - vyskytují se i v exonech
 - neporušují čtecí rámeček
 - GTG – subtelomerická lokalizace na chromozomu
- tetranukleotidy
 - téměř výhradně GATA/GACA
 - výskyt poblíž centromer
 - často složené nebo přerušované

Charakteristiky mikrosatelitů

- *single locus* – vysoce specifické
- častý výskyt v genomu
- rozmístěny po celém genomu
- vysoký polymorfismus – mnoho alel
- kodominantní dědičnost

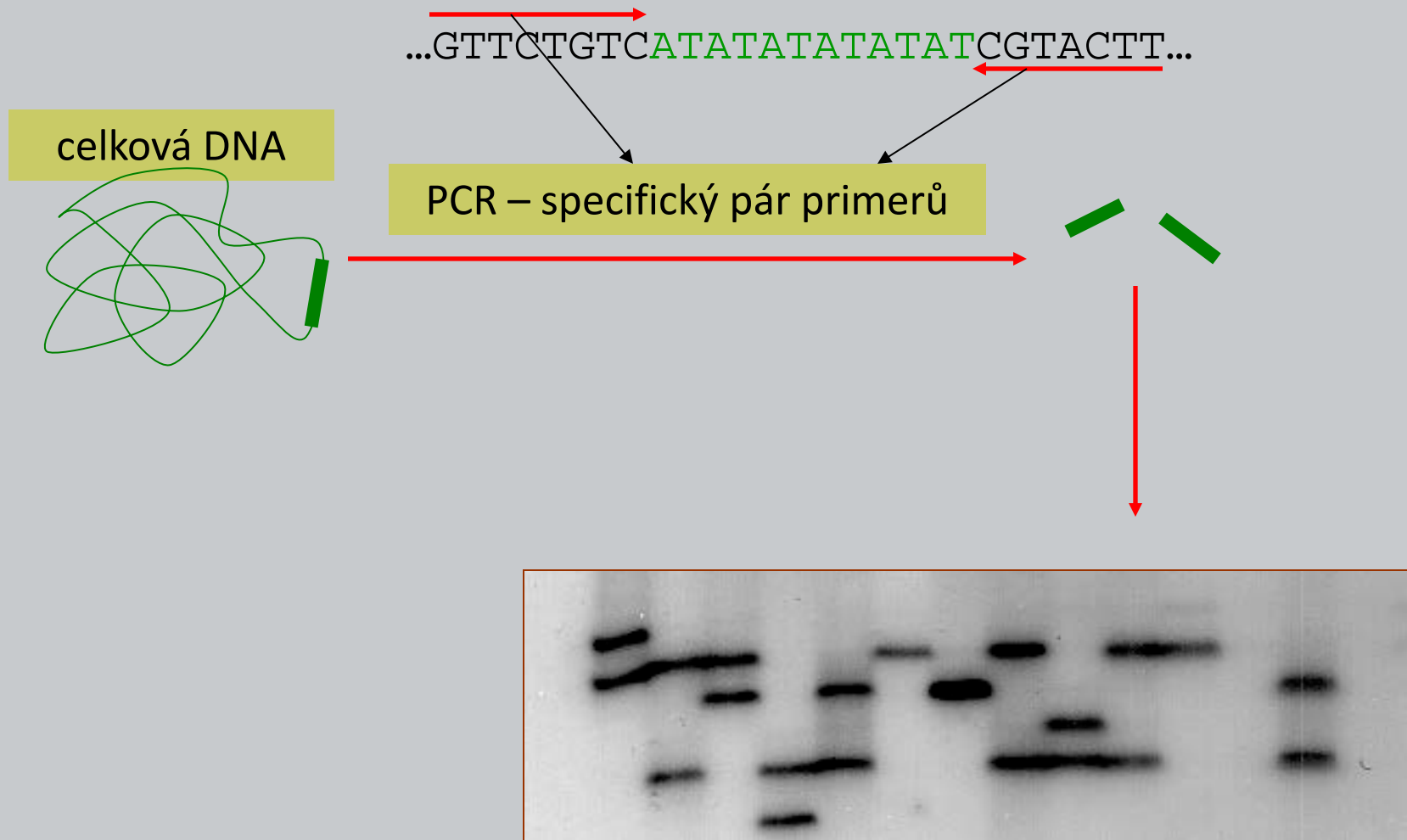
- ALE – musíme znát primery (tj. sekvence *flanking regions* – okolních úseků)

...GTTCTGTC  ATATATATATATATATATAT  CGTACTTA...

Rozmístění v genomu

- výskyt v celém genomu (ALE – odráží variabilitu pouze studovaných lokusů, tj. omezeného množství)
- spíše v nekodujících oblastech
- jaderné mikrosatelity
 - druhově specifické
- chloroplastové mikrosatelity
 - téměř výhradně opakování jediné báze – např. (T)₁₂
 - okolí (*flanking regions*) – méně variabilní – možnost designu konsenzuálních primerů

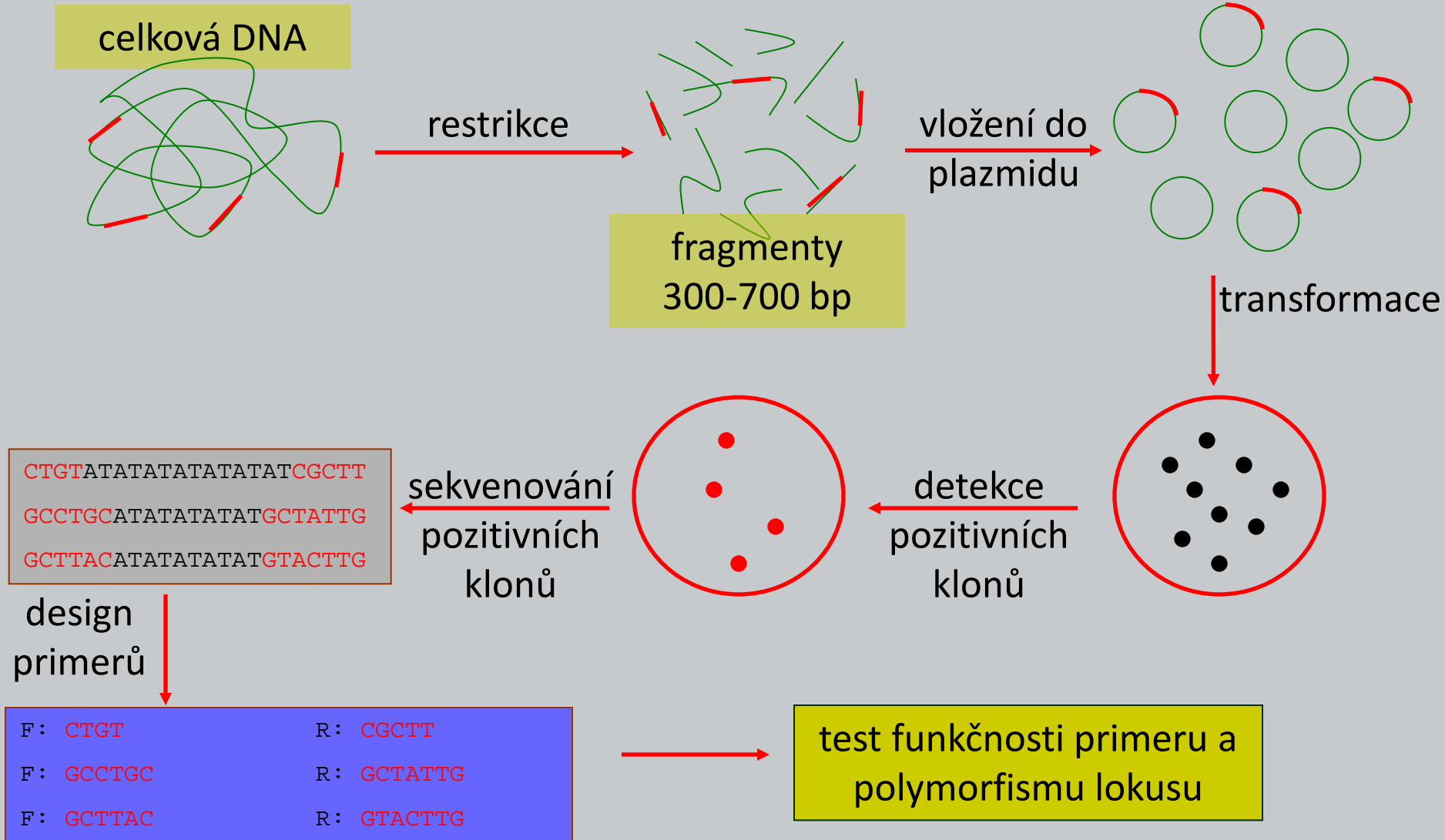
Zjištění polymorfismu



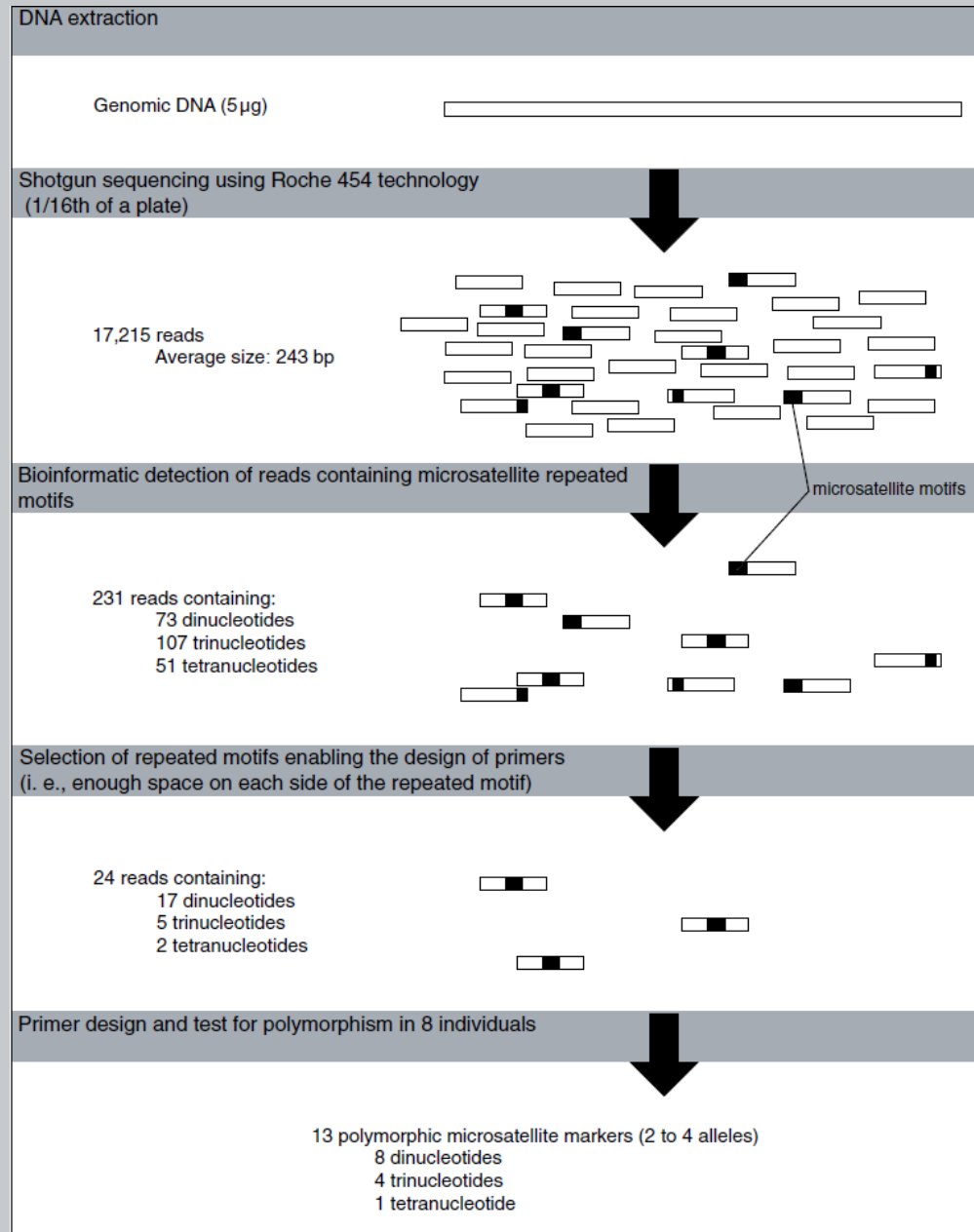
Mikrosatelitové primery

- specifické pro daný lokus – pouze jednou v genomu
- specifické pro konkrétní druh
- pro studovaný druh existují (publikované)
(např. v *Molecular Ecology Resources*, dříve *ME Notes*)
(viz také databáze na <http://tomato.bio.trinity.edu>)
- rešerše – <http://botany.natur.cuni.cz/dna> (Primery/Mikrosatelitové primery)
- prohledání genové banky – pro modelové organismy
- otestovat z příbuzných organismů (téhož rodu), tzv. *cross-amplifikace* – většinou buď nefunguje nebo problém s nulovými alelami
- nutno si vytvořit (izolovat)
 - klasické klonování
 - pyrosekvenování (shotgun sequencing) + hledání *reads* s mikrosatelity

Izolace primerů

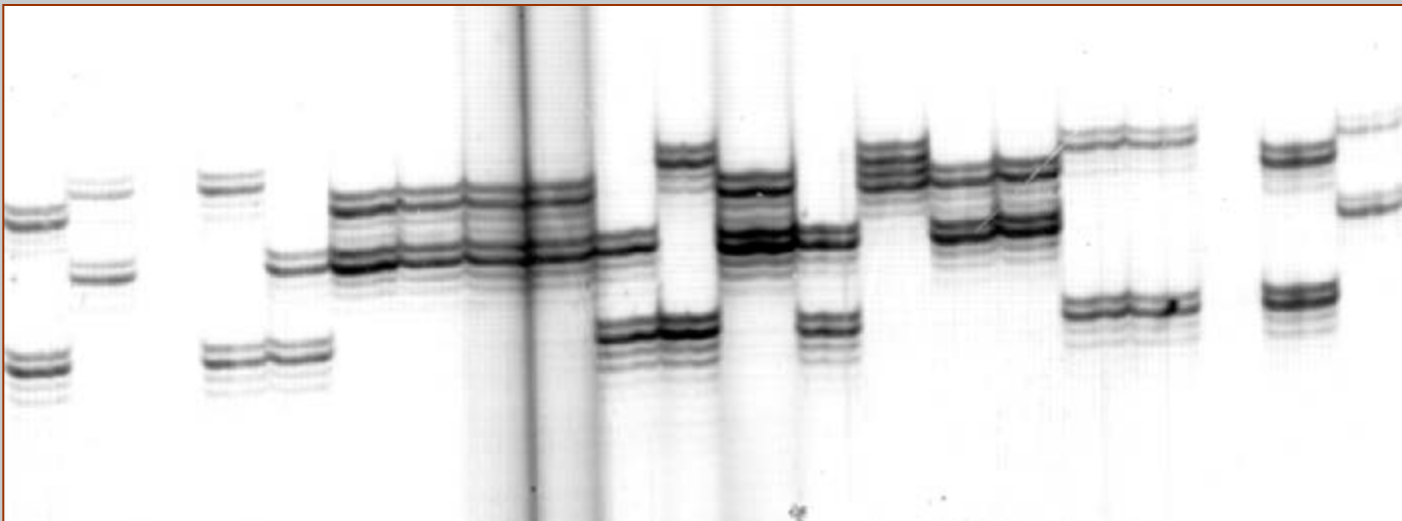


Izolace primerů – pyrosekvenování



Interpretace gelů

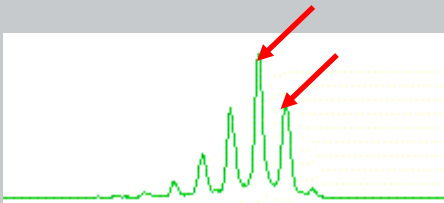
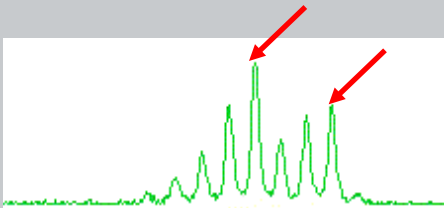
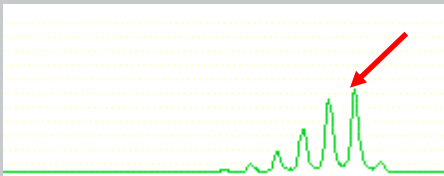
- „*stutter bands*“ – přidavné proužky okolo správně dlouhého (nejintenzivnějšího) – *in vitro DNA slippage*
- „*terminal transferase activity*“ – tendence *Taq* polymerázy přidávat A na 3'-konec



Interpretace gelů II.

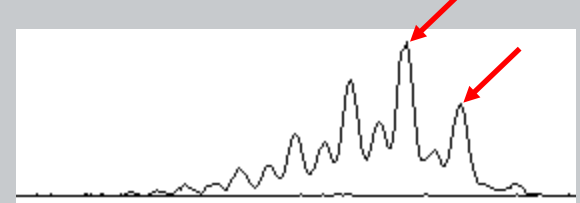
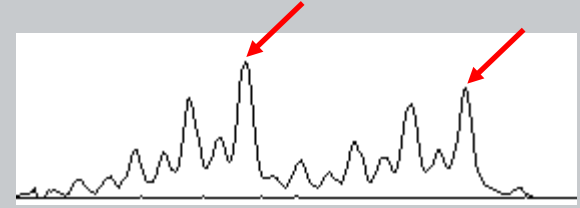
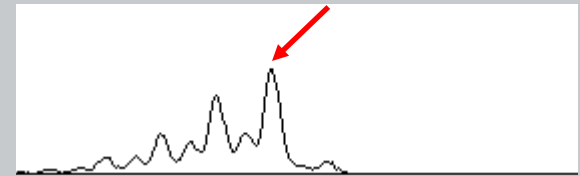
stutter bands

- produkty o 2, 4, 6 atd. bp kratší
- nejvyšší *peak* nejdelší – skutečná alela



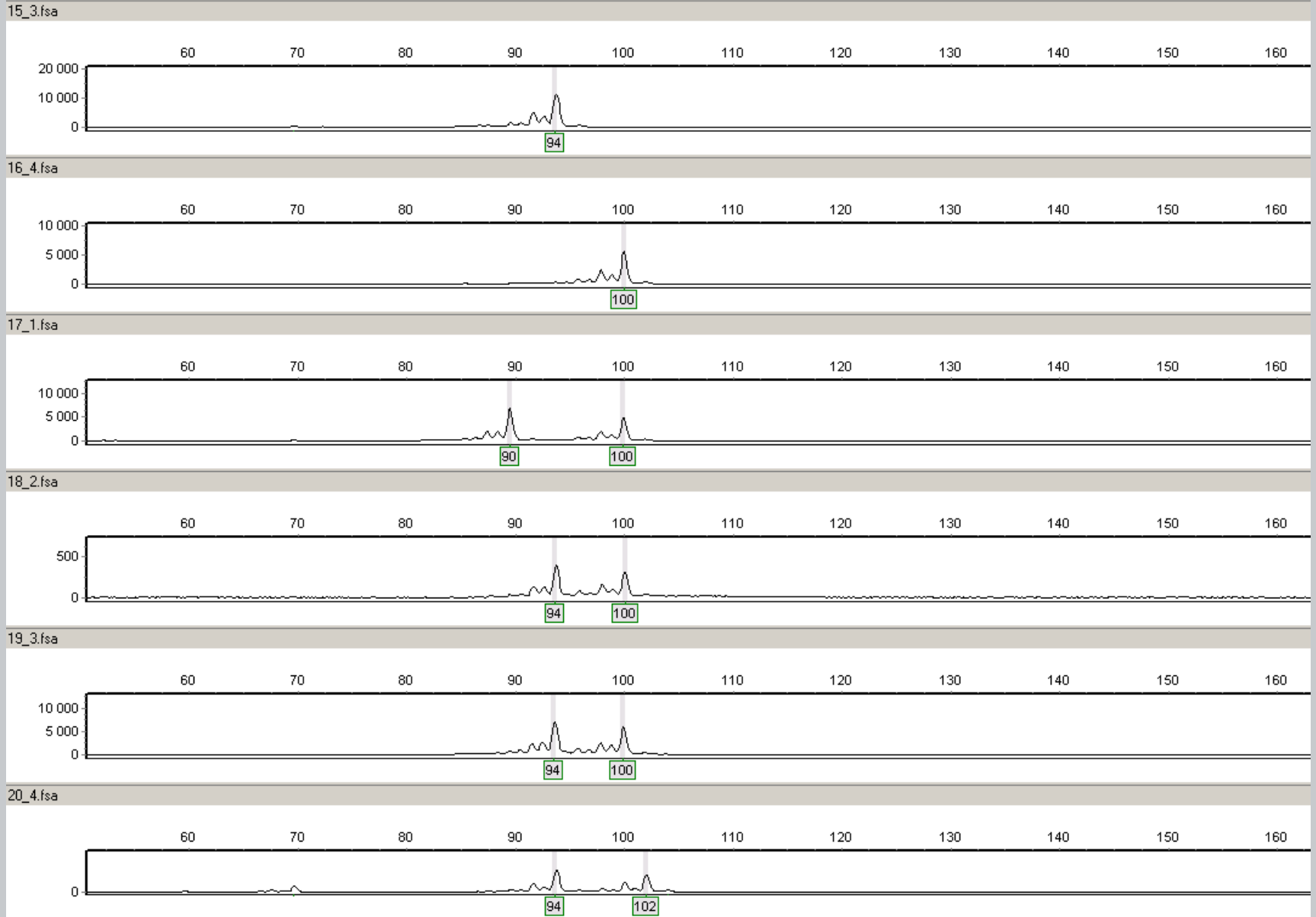
stutter bands a **-A** produkty

- *stutter bands* o 2, 4, 6 atd. bp kratší
- ke každému proužku i -A produkt



↙ označují skutečné alely

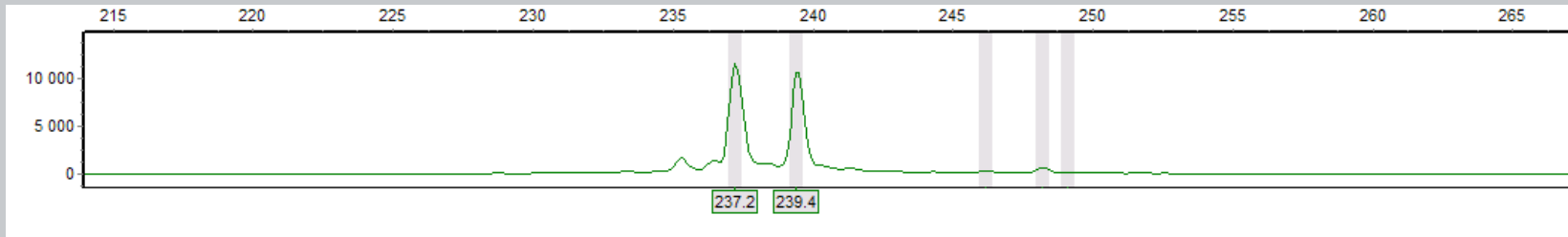
Automatická analýza (GeneMarker)



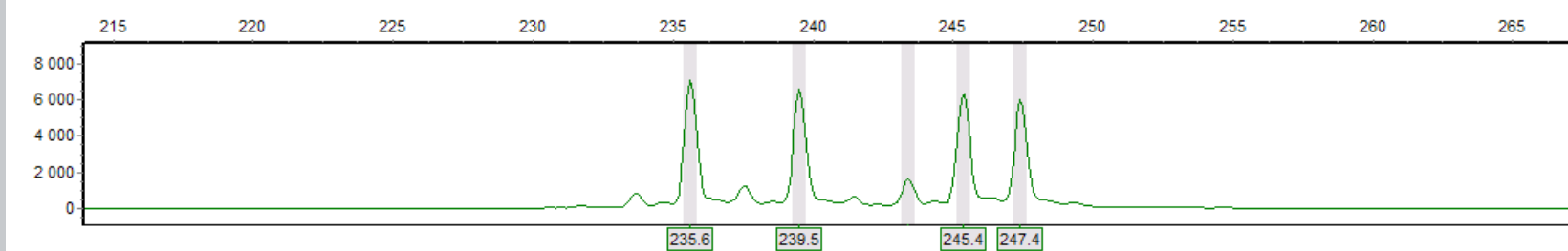
Jak hodnotit tetraploidy

- jako dominantní data – přítomnost/nepřítomnost alel
- kodominantně (vidíme alely, ale jaký je genotyp?)
 - tři alely – jedna z nich je dvakrát, ale která? (tj. hodnotíme jako 3 alely + chybějící)
 - dvě alely – každá dvakrát nebo jedna třikrát? (tj. hodnotíme jako 2 + 2 chybějící)
 - problém – velké množství chybějících dat
 - alternativa – počet alel určen z výšky/plochy peaku (?)
- autopolyploidi/allopolyploidi
- software umožňující analýzu různých ploidních úrovní – SPAGeDi, TETRASAT, POLYSAT (R), BAPS, STRUCTURE...

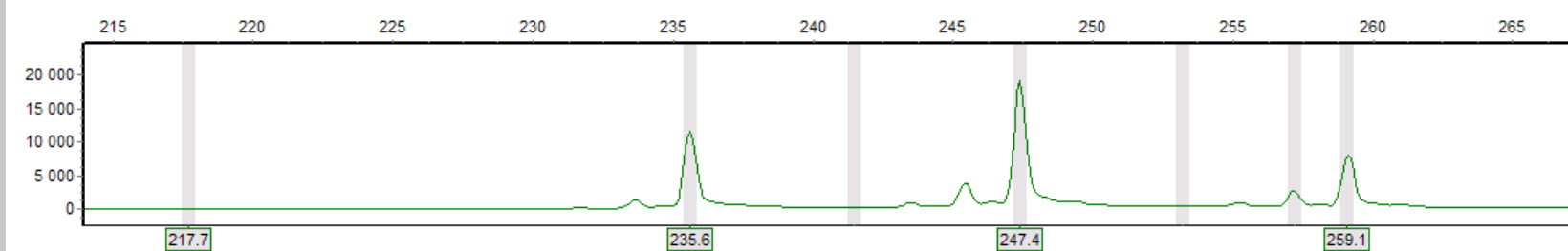
Tetraploidní data (*Betula*)



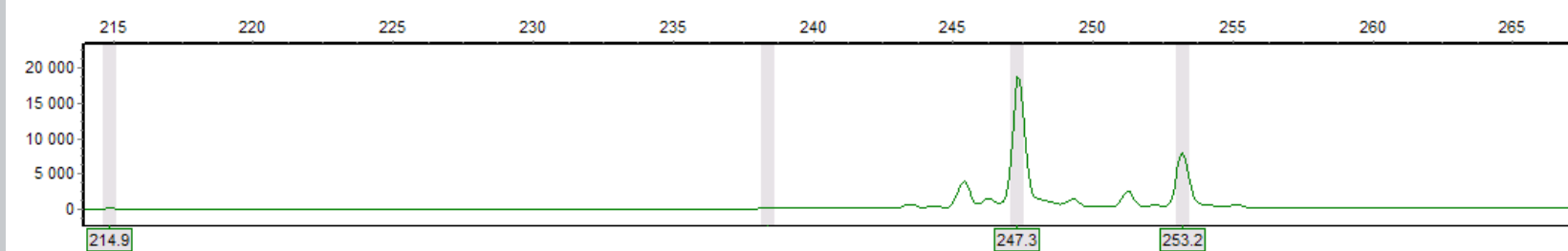
7_B02_2011-08-20-11-11-59_FA_POP7_Inj3sec_20min.fsa



8_B03_2011-08-20-11-11-59_FA_POP7_Inj3sec_20min.fsa



9_B04_2011-08-20-11-11-59_FA_POP7_Inj3sec_20min.fsa



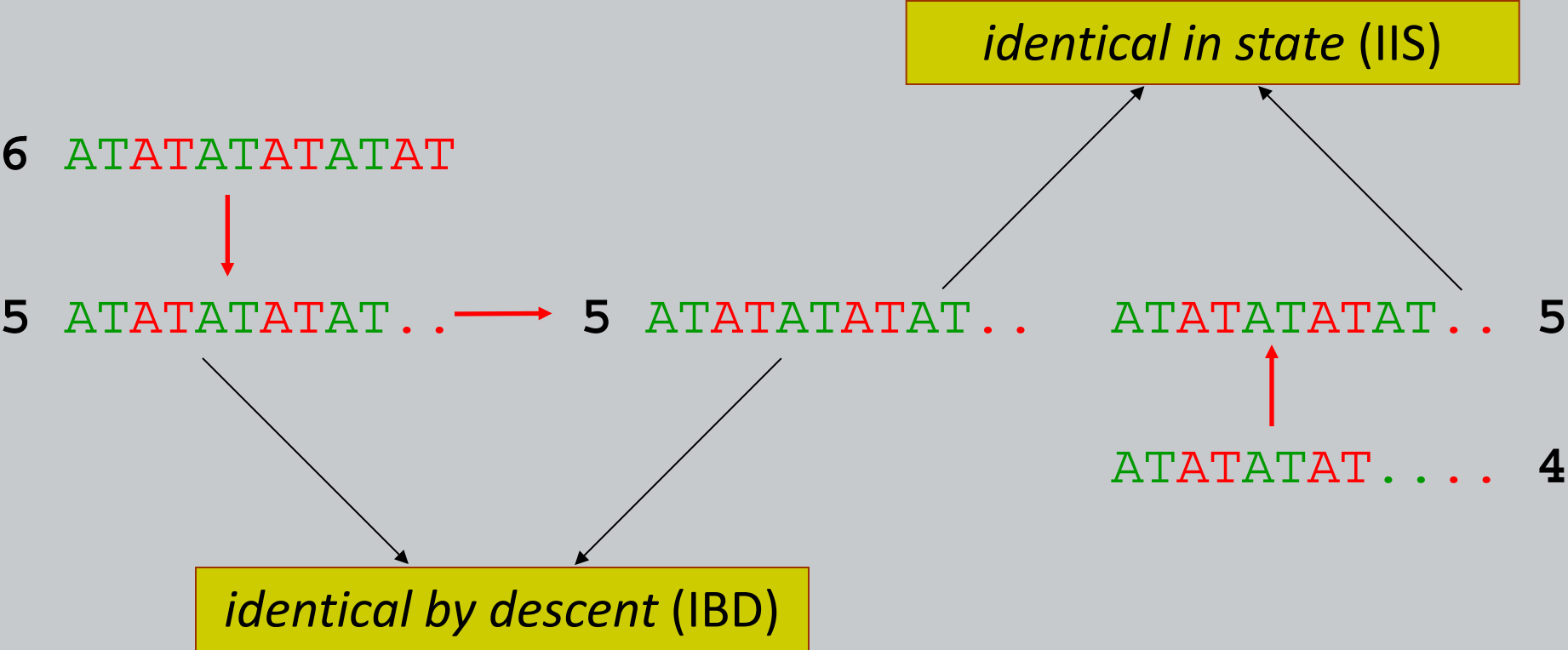
Vznik polymorfismu

- *DNA „slippage“*
 - „sklouznutí“ DNA polymerázy při replikaci
 - prodloužení nebo zkrácení o jednotku
- *„unequal crossing over“*
 - rozsáhlejší změny
- mutační rychlost je vysoká – 10^{-3} - 10^{-5}

Mutace mikrosatelitů

- o podstatě mutací mikrosatelitů se ví málo
- rychlost se odhaduje na 10^{-3} - 10^{-5}
 - liší se u 2, 3 a 4 bp opakování
 - podle typu mikrosatelitu
 - různá u různých druhů ...
- mutační rychlost – balance mezi vznikem mutací a jejich opravou
- nejčastěji – ztráta nebo získání jedné jednotky

Homologie alel



Mutační modely

- *infinite alleles model (IAM) – Kimura & Crow 1964*
 - mutace vytváří novou alelu danou rychlostí u
 - nepovoluje homoplasie
 - identické alely jsou IBD
- *stepwise mutation model (SMM) – Kimura & Ohta 1978*
 - nová alela vzniká přidáním nebo ztrátou jedné jednotky
 - stejná pravděpodobnost ztráty i přidání ($u/2$)
 - generuje homoplasie (alely nejsou IBD, pouze IIS)
 - alely o podobné délce jsou více příbuzné
- *two-phase model (TPM) – DiRienzo et al. 1994*
 - modifikace o 1 jednotku s pravděpodobností p
 - modifikace o více než 1 jednotku s pravděpodobností $1-p$

Nulové alely

- ztráta PCR produktu mutací v *priming site*
- tj. podhodnocení heterozygotů – někteří se jeví jako homozygoti
- identifikace pouze studiem potomků – alela není zděděna
- frekvence se zvyšuje použitím heterologních primerů (z příbuzných druhů)
- frekvenci možno odhadnout na základě odchylky od H-W rovnováhy (např. program Cervus)

Vyhodnocování dat

- kodominantní marker – alelické vyhodnocení (jako allozomy)
 - heterozygosita (pozorovaná, očekávaná)
 - F-statistika (F_{ST} , F_{IS})...
 - vzdálenosti (mezi populacemi, jedinci)
 - proporce sdílených alel (D_{ps})
 - Nei's chord distance (D_a)
 - Nei's standard distance (D)
- specifické koeficienty pro mikrosatelity
 - R_{ST} – analog F_{ST} (Slatkin 1995)
 - obsahuje logiku SMM (*stepwise mutation model* – založen na varianci v délce alel)
 - odhady – ρ_{ST} (Rousset 1996)
 - vzdálenosti
 - delta mu – $(dm)^2$, D_{dm} (Goldstein et al. 1995)
 - D_{sw} – *stepwise weighted genetic distance*
 - ...
- software
 - MICROSAT (Minch 1996)
 - MSA – Microsatellite Analyser (Dieringer & Schlötterer 2003)
 - RSTcalc (Goodman 1997)

Využití mikrosatelitů

- analýza rodičovství (*parentage analysis*)
 - určení rodičů semen (semenáčků) v populacích
 - outcrossing rate
- identifikace klonů
- populačně-genetické studie
 - inbreeding, odchylky od H-W rovnováhy
 - genový tok, migrace
 - historie populací, efektivní velikost populací ...
- systematika
 - problematické využití – homologie alel?
 - pouze do úrovně blízce příbuzných druhů
 - nutno použít mnoho lokusů
 - spíše cpDNA SSRs – díky vlastnostem cpDNA...
- hybridizace
 - odlišení F1 hybridů a zpětných hybridů (dynamika hybridizace, introgrese)

Analýza rodičovství (*parentage analysis*)

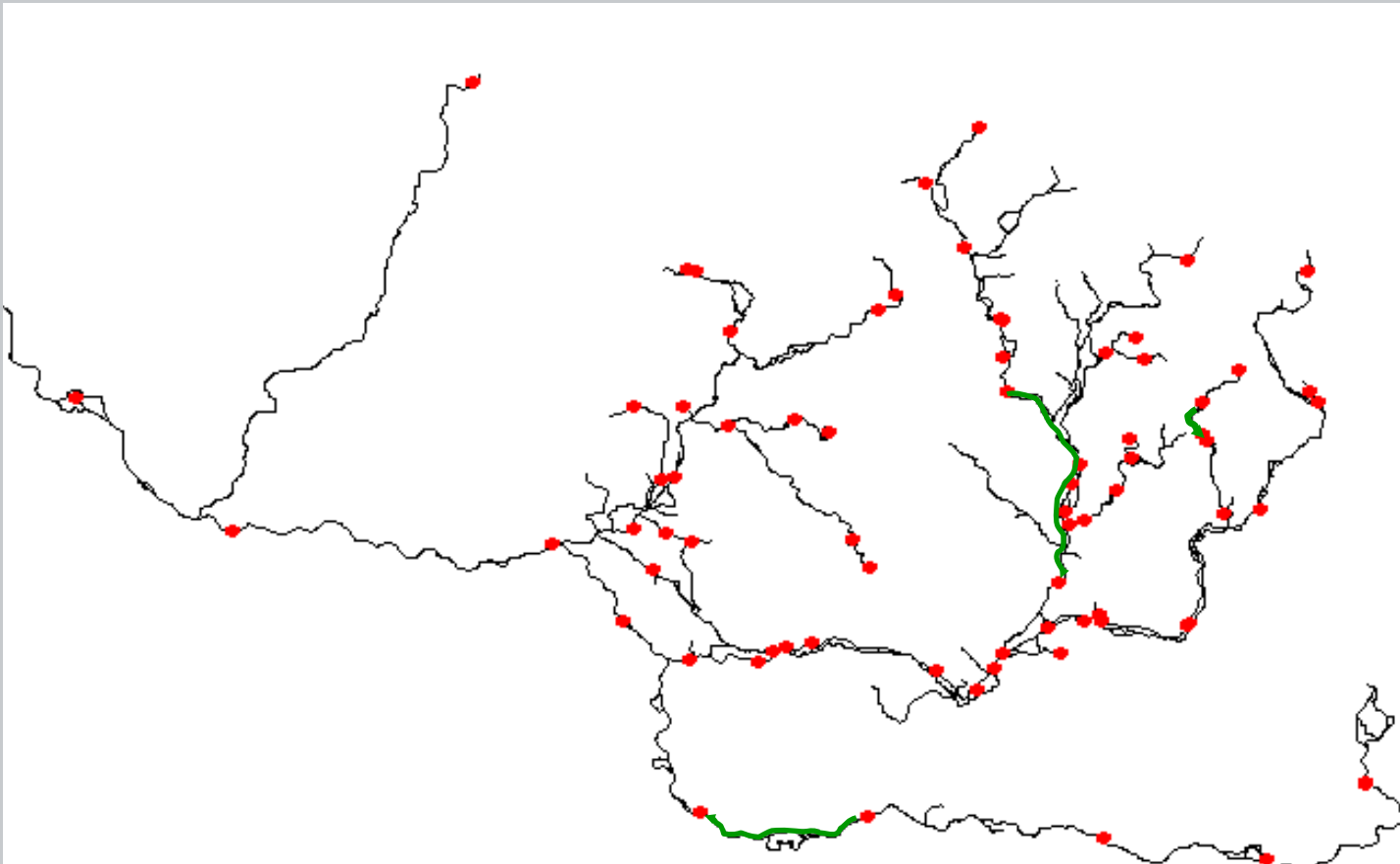
- přímé určení vzdálenosti a četnosti šíření
 - semen – vzdálenost mezi semenem a rodičem
 - pylu – vzdálenost mezi rodičovskými páry
- úspěch jednotlivých genotypů v populaci
 - podíl „jedinců-otců“ na opylení
- outcrossing rate
 - % semen vzniklých allogamicky
- předpoklady
 - genotypy všech potenciálních rodičů (rel. nízký počet)
 - variabilní marker – mikrosatelity, AFLP

Metody *parentage analysis*

- vylučovací metodou – *exclusion analysis*
 - nekompatibilita mezi genotypy rodičů a předpokládaného potomka → zamítnutí hypotézy
 - tj. zamítnutí všech až na jednoho, resp. dva
 - problémy – chyby ve skórování gelu, nulové alely, mutace
- *categorical allocation*
 - výpočet LOD score (*logarithm of the likelihood ratio*)
 - rodiče mají nejvyšší LOD score
 - výhoda – méně citlivé k chybám a mutacím
- software – např. CERVUS (Marshall et al. 1998)

Identifikace klonů

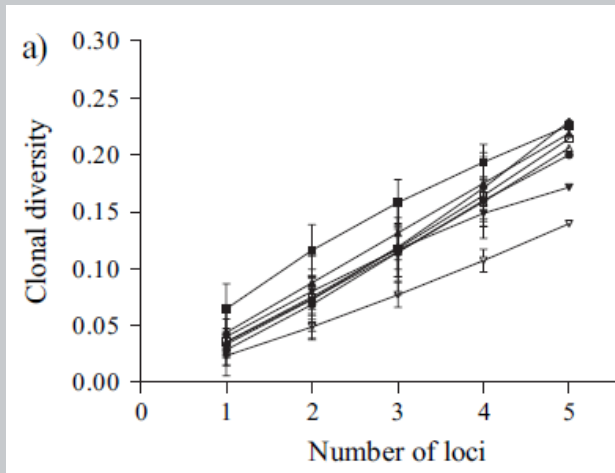
- klon = stejný multilokusový genotyp (tj. stejné alely ve všech lokusech)



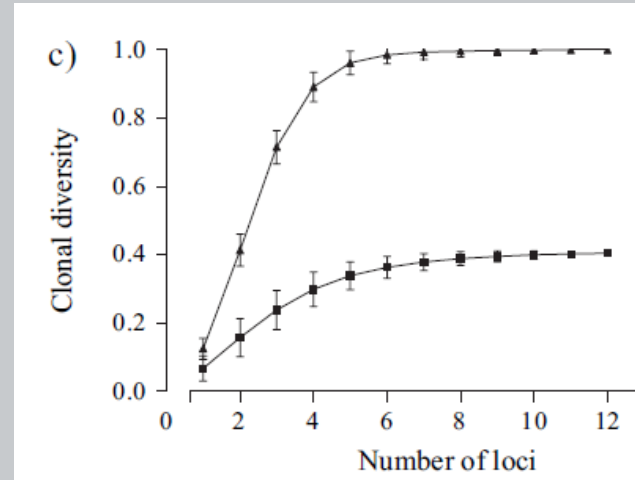
Phragmites australis v povodí Labe (Fér & Hroudová 2009)

Identifikace klonů

- pozor na rozlišovací schopnost markerů
- síla markeru (*marker power*)



nedostatečná variabilita



dostatečná variabilita

- MLG (*multilocus genotype*)
 - pokud nalezen více než jedenkrát – spočítat tzv. P_{sex} , tj. pravděpodobnost, že tento MLG mohl vzniknout náhodou při jiné generativní události – software GenClone, MLGSIM

Arnaud-Haond et al. (2005): Assessing genetic diversity in clonal organisms: Low diversity or low resolution? Combining power and cost efficiency in selecting markers. *Journal of Heredity* 96:434-440

Arnaud-Haond et al. (2007): Standardizing methods to address clonality in population studies. *Molecular Ecology* 16: 5115–5139

Genový tok – nepřímé určení

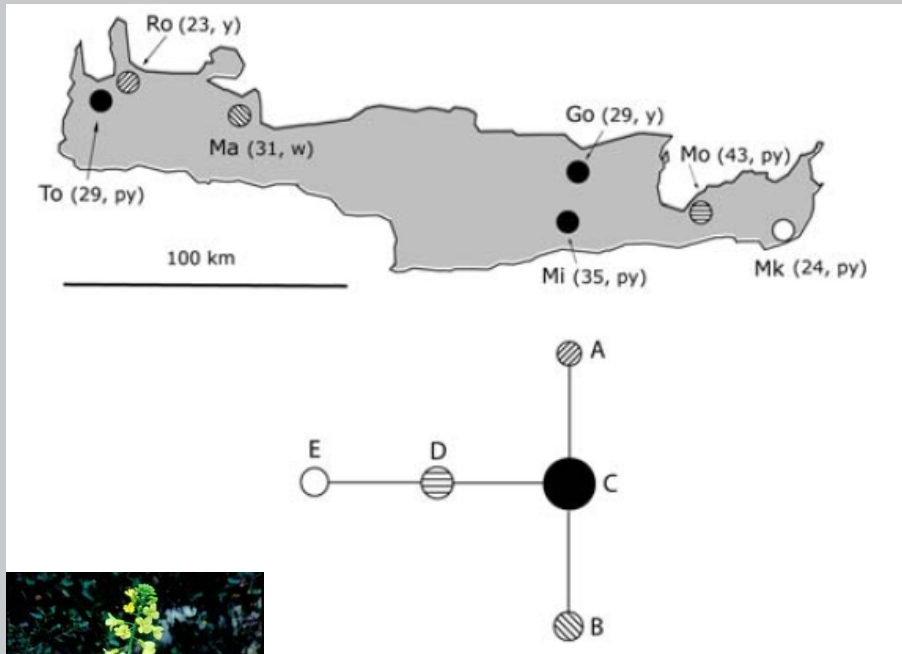


Table 2 Estimates of the total (A_{tot}) and average within-population (A_{pop}) number of alleles, total (H_T) and within-population (H_S) gene diversity, and population differentiation as F_{ST} at ten nuclear and four chloroplast microsatellite loci in *Brassica cretica*

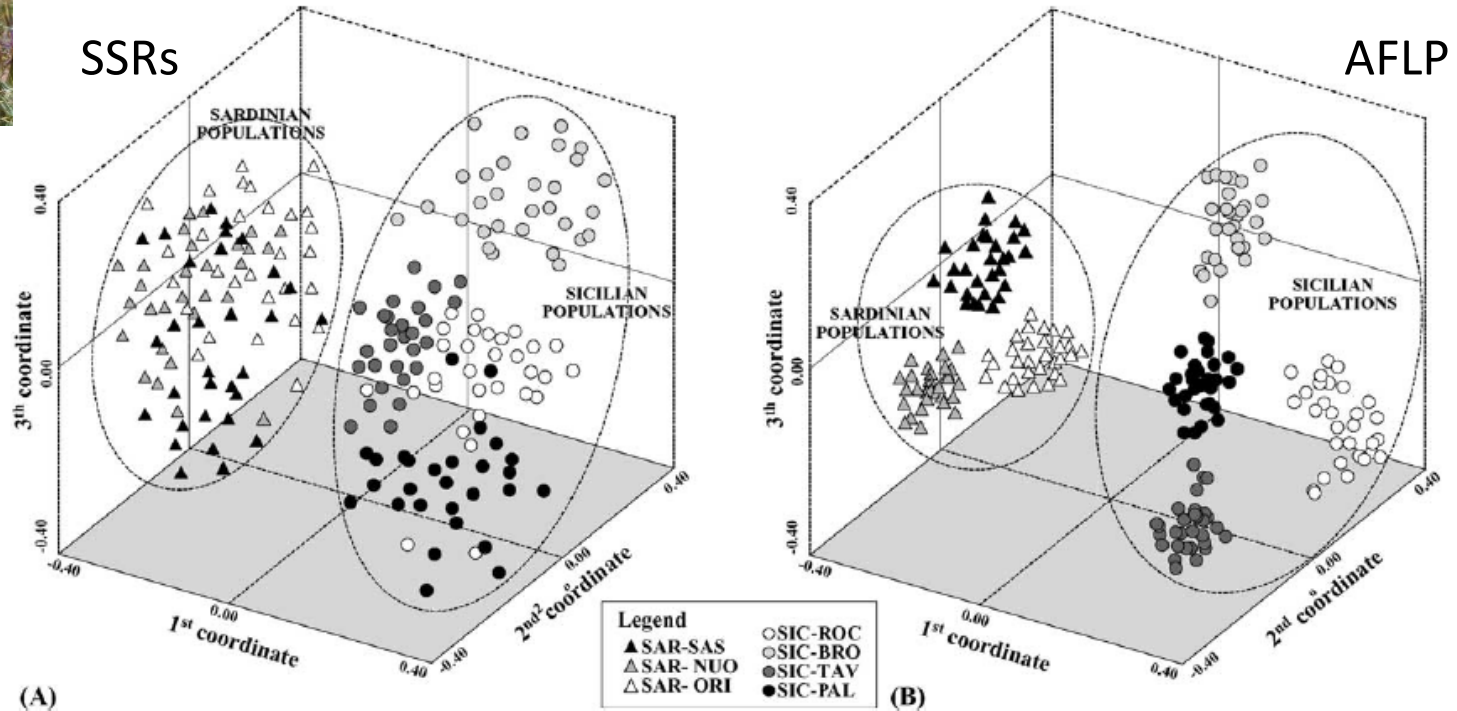
Locus	A_{tot}	A_{pop}	H_T	H_S	F_{ST}
Nuclear					
Ol10-F11	2	1.3	0.169	0.074	0.601
Ni4-B10	7	3.0	0.802	0.400	0.549
Ol9-A06	1	1.0	0.000	0.000	—*
Na12-A07	4	2.1	0.519	0.369	0.320
sORA26	2	1.4	0.156	0.121	0.237
BN12A	6	2.1	0.721	0.271	0.634
Na10-F06	5	1.7	0.647	0.160	0.766
Ol12-F02	20	4.0	0.851	0.436	0.556
nga111	5	1.7	0.752	0.176	0.763
MB4	2	1.0	0.408	0.000	1.000
Chloroplast					
ATCP28673	2	1.0	0.245	0.000	1.000
ATCP70189	1	1.0	0.000	0.000	—*
ccmp6	3	1.0	0.449	0.000	1.000
ccmp10	2	1.0	0.408	0.000	1.000

*Monomorphic locus, F_{ST} is not defined.

Edh K. et al. (2007): Nuclear and chloroplast microsatellites reveal extreme population differentiation and limited gene flow in the Aegean endemic *Brassica cretica* (Brassicaceae). Mol. Ecol. 16, 4972-4983.

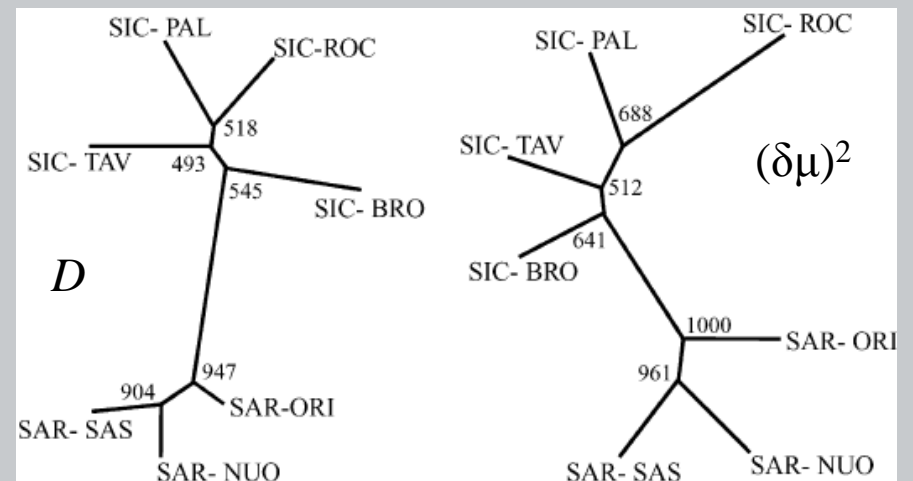


Cynara cardunculus – 5 lokusů



Locus	F_{IS}	F_{IT}	F_{ST}	R_{ST}
CDAT-01	-0.119	-0.016	0.093	0.078
CLIB-02	-0.089	0.139	0.201	0.178
CMAL-06	-0.076	0.068	0.133	0.182
CMAL-24	-0.036	0.136	0.166	0.210
CMAL-108	-0.014	0.071	0.083	0.185
Overall loci	-0.064	0.086*	0.141*	0.168*

* $P < 0.0001$.



Určení fylogeneze

Hordeum
cpDNA mikrosatelity

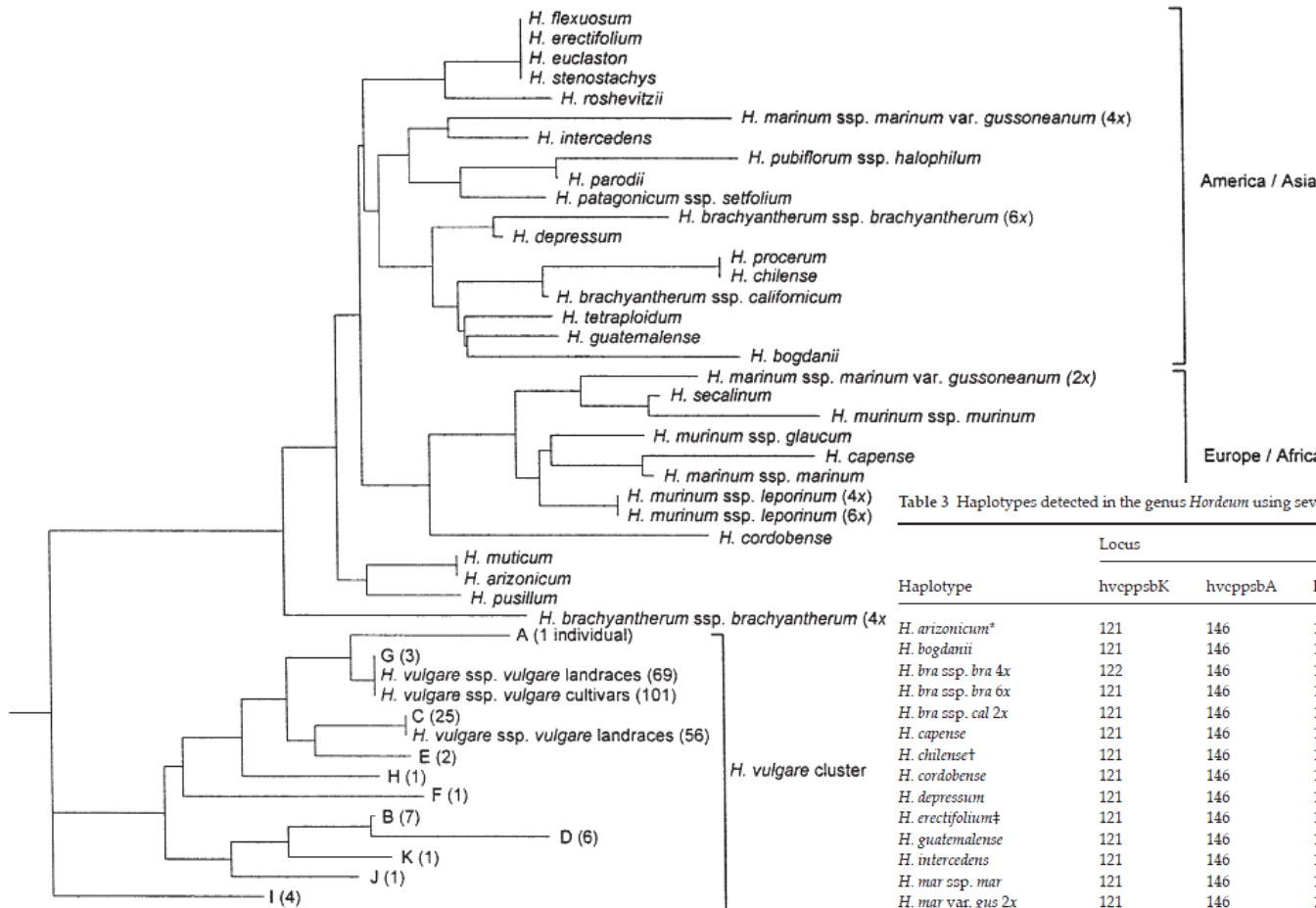


Table 3 Haplotypes detected in the genus *Hordeum* using seven cpSSRs

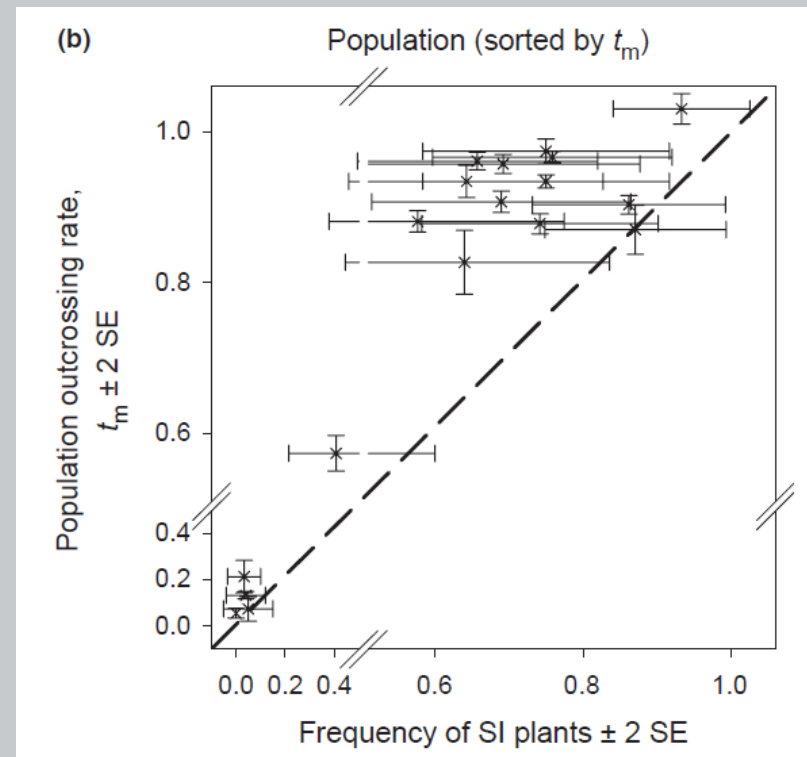
Haplotype	Locus						
	hvcppsK	hvcppsA	hvcprpA	hvcprps12	hvcptrn51	hvcptrn52	hvcptrnLF
<i>H. arizonicum</i> *	121	146	122	148	128	115	101
<i>H. bogdanii</i>	121	146	122	148	135	102	99
<i>H. bra ssp. bra 4x</i>	122	146	122	148	128	112	101
<i>H. bra ssp. bra 6x</i>	121	146	116	148	128	116	99
<i>H. bra ssp. cal 2x</i>	121	146	122	148	128	114	99
<i>H. capense</i>	121	146	118	148	135	113	101
<i>H. chilense</i> †	121	146	122	152	128	114	99
<i>H. cordobense</i>	121	146	126	148	128	114	103
<i>H. depressum</i>	121	146	122	148	128	116	99
<i>H. erectifolium</i> ‡	121	146	122	148	128	114	100
<i>H. guatemalense</i>	121	146	122	148	128	113	99
<i>H. intercedens</i>	121	146	122	148	128	116	101
<i>H. mar ssp. mar</i>	121	146	118	148	128	113	101
<i>H. mar var. gus 2x</i>	121	146	118	148	128	112	102
<i>H. mar var. gus 4x</i>	120	146	122	148	128	116	97
<i>H. mur ssp. mur</i>	120	146	118	148	128	112	101
<i>H. mur ssp. gla</i>	121	146	118	148	128	115	101
<i>H. mur ssp. lep 4x</i> §	121	146	118	148	128	114	101
<i>H. parodii</i>	121	146	122	148	128	115	102
<i>H. pat ssp. set</i>	121	146	122	148	128	116	102
<i>H. procerum</i>	121	146	122	152	128	114	99
<i>H. pub ssp. hal</i>	121	145	122	148	128	115	102
<i>H. pusillum</i>	121	146	122	148	128	113	101
<i>H. roshevitzii</i>	121	146	122	148	128	103	100
<i>H. secalinum</i>	121	146	118	148	128	112	101
<i>H. stenostachys</i>	121	146	122	148	128	114	100
<i>H. tetraploidum</i>	121	146	122	148	128	115	99

Provan J. et al. (1999): Polymorphic chloroplast simple sequence repeats for systematic and population studies in the genus *Hordeum*. *Molecular Ecology* 8, 505-511.

Self-(in)compatibility

- procento semen vzniklých autogamicky, tj. při *parentage analysis* byl stejný jedinec přiřazen jako první i druhý rodič
- *outcrossing rate*

Willi Y. & Määtänen K. (2010): Evolutionary dynamics of mating system shifts in *Arabidopsis lyrata*.
Journal of Evolutionary Biology 23: 2123–2131.



Hybridizace

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. x glauca	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

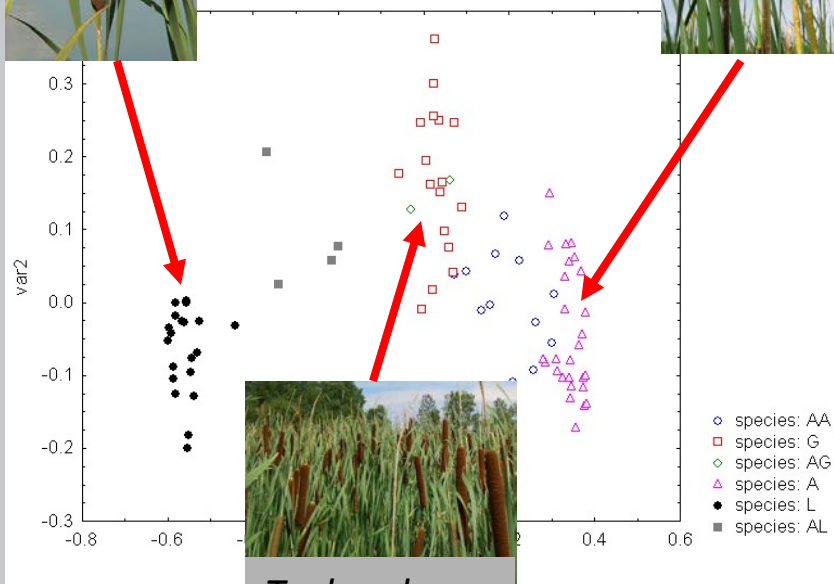
Typha latifolia



Typha angustifolia



scatterplot (Typha_US_PCoA_DAN 5v*114c)



Typha glauca

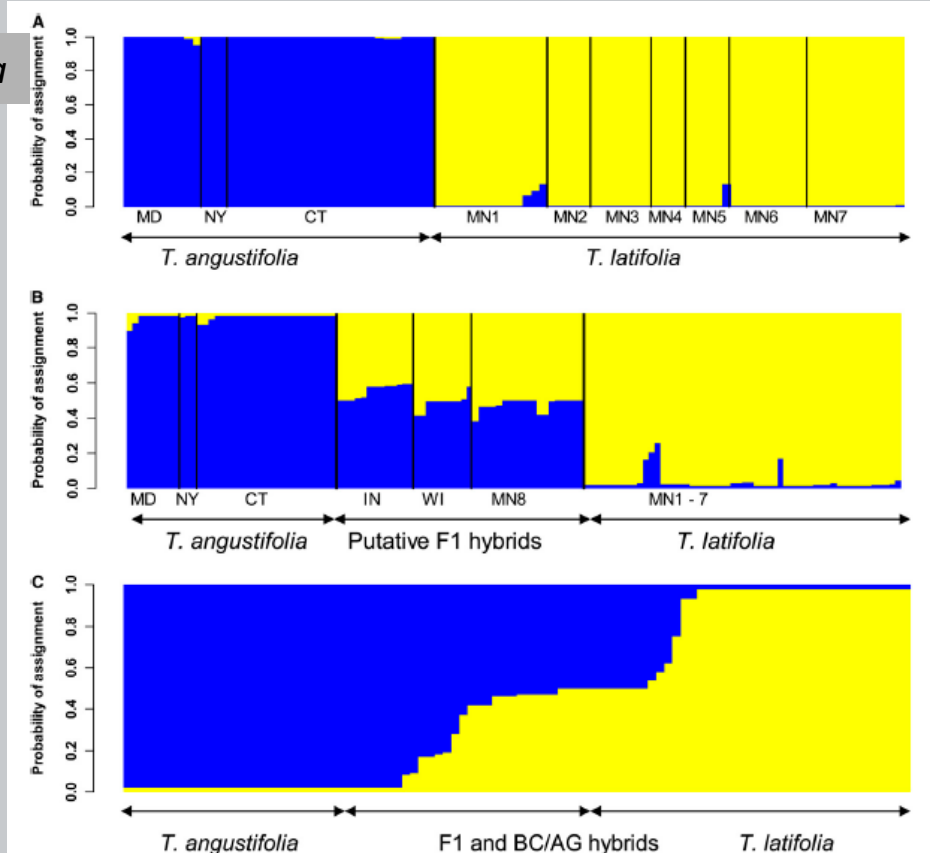


Fig. 1. Barplots of STRUCTURE analyses for (A) RAPD-identified *T. angustifolia* and *T. latifolia* from Collection Group 1; (B) RAPD-identified *T. angustifolia*, *T. latifolia*, and putative F₁ hybrids from additional Collection Group 1 sites in Indiana, Wisconsin, and Minnesota; and (C) mixed populations in Michigan (Collection Group 2) that included advanced-generation/backcrossed individuals (AG/BC). Each bar represents a single individual. Locations sampled are listed in Tables 1 and 2. Six SSR loci were used in these STRUCTURE analyses (TA 21 was not included).

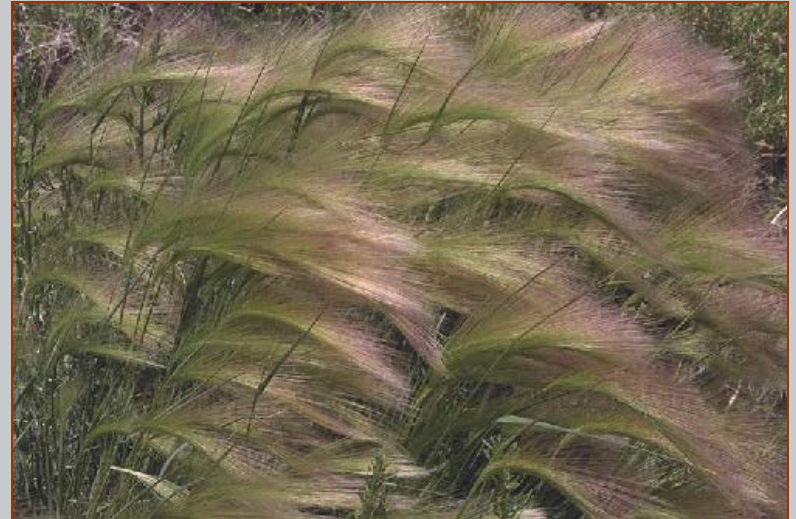
Populační studie

Kameyama Y. et al. (2001): Patterns and levels of gene flow in *Rhododendron metternichii* var. *hondoense* revealed by microsatellite analysis. *Molecular Ecology* 10:205-216



Systematická studie

Provan J. et al. (1999): Polymorphic chloroplast simple sequence repeat primers for systematic and population studies in the genus *Hordeum*. *Molecular Ecology* 8:505-511



Literatura

- Jarne P. & Lagoda P.J.L. (1996): *Microsatellites, from molecules to populations and back*. Trends in Ecology & Evolution 11(10):424-429
- Goldstein D.B. & Schlotterer Ch. (1999): *Microsatellites. Evolution and Applications*. Oxford University Press
- Provan J. et al. (2001): *Chloroplast microsatellites: new tools for studies in plant ecology and evolution*. Trends in Ecology & Evolution 16(3):142-147
- Lulkart G. & England P.R. (1999): *Statistical analysis of microsatellite DNA data*. Trends in Ecology & Evolution 14(7):253-256
- Balloux F. & Lugon-Moulin N. (2002): *The estimation of population differentiation with microsatellite markers*. Molecular Ecology 11:155-165
- Robinson J.P. & Harris S.A. (1999): *Amplified Fragment Length Polymorphisms and Microsatellites: A phylogenetic perspective*. In: Gillet E.M.[ed.]: Which DNA Marker for Which Purpose?
<http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm>
- Jones A.G. & Ardren W.R. (2003): *Methods of parentage analysis in natural populations*. Molecular Ecology 12:2511-2523
- Selkoe K.A. & Toonen R.J. (2006): *Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers*. Ecology Letters 9: 615-629.
- Guichoux E. et al. (2011): *Current trends in microsatellite genotyping*. Molecular Ecology Resources 11: 591-611.