

Proměnlivost a evoluce rostlin

Petr Smýkal

Katedra botaniky, PřF UPOL

2012/13

Metody detekce a analýzy variability a evoluce rostlin

- 1. Morfometrie**
- 2. Biochemické a proteinové markery**
- 3. DNA markery**
- 4. Markery pro fylogenetické analýzy**

Počátky biometrie

Michel Adanson (1727-1806)

Familles naturelles des plantes , 1763

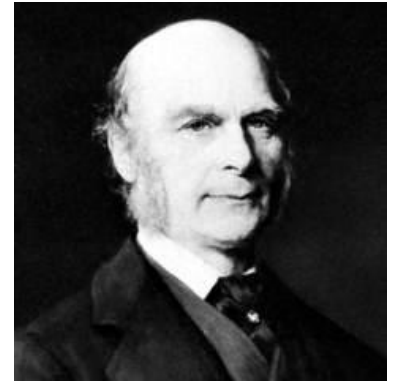
"In 1774 Adanson submitted to the consideration of the French Academy of Sciences an immense work, extending to all known beings and substances. It consisted of **27 large volumes of manuscript**, employed in displaying the general relations of all these matters, and their distribution; 150 volumes more, occupied with the alphabetical arrangement of **40,000 species**; a vocabulary, containing 200,000 words, with their explanations; and a number of detached memoirs, 40,000 figures and 30,000 specimens of the three kingdoms of nature. The committee to which the inspection of this enormous mass was entrusted strongly recommended Adanson to separate and publish all that was peculiarly his own, leaving out what was merely compilation. He obstinately rejected this advice; and the huge work, at which he continued to labour, **was never published.**„



Potřeboval metodu pro taxonomickou klasifikaci – vytvořil **taxon by character matrix s 65 odlišnými znaky.**

Položil tak základy metod numerické taxonomie.

Francis Galton (1822-1911)



anglický vědec, činný ve velmi mnoha různých oborech: psychologii a antropologii, statistice, geografii a dalších.

V matematice rozpracoval metody statistického zpracování výsledků pozorování (především metodu vypočítání korelací mezi proměnnými); zavedl **korelační koeficient**

Galton je zakladatelem eugeniky. Galton byl bratrancem Charlese Darwina.

Statisticky studoval lidskou proměnlivost a dědičnost, v kontextu se sociálními otázkami

Hereditary genius (1869) první příklad antropometrie. Dědičnost inteligence – eugenické ideje.

Snaha u důkazu gemulí (volně cirkulujících v organismech a vstupujících do pohlavních buněk) pomocí krve králíků - neúspěšné

Typy znaků

Kvalitativní vs. Kvantitativní znaky, ne-numerické – tvary, formy vs. numerické

Diskrétní (nespojité) vs. Spojité (kontinuální)

1. **Diskrétní** = celá čísla např. počet prašníků
 - a. **Binární** = 2 stavy (0, 1) např. přítomno/absence listů
 - b. **Vícestavové** = více než 2 stavy (0, 1, 2). Květy červené, purpurové, načervenalé. (mohou se re-kódovat do série binárních)
2. **Kontinuální** (nekonečné série čísel mezi stavy) např. měření výšky (1.2 cm, 1.238 cm).

Karl Pearson

(1857 – 1936)

Pearsonova práce se uplatnila v rozvoji matematické statistiky pro oblast biologie, epidemiologie, antropometrie, medicíny a sociálních dějin.

Korelační koeficient - parametrický statistický test (předpokládající normální rozdělení) zjišťující, jak těsný je vztah proměnných a jaký má směr (kladný nebo záporný)

Chí vzdálenost - je založena na korelaci mezi proměnnými, které mohou být identifikovány a analyzovány pomocí vzorů. Je to užitečný způsob určení podobnosti neznámého vzorku s jedním známým. Liší se od Euklidovské vzdálenosti v tom, že bere v úvahu korelace souboru dat, kdy je měřítko neměnná, tj. není závislé na rozsahu měření.

P hodnota - je u testu, kde má tato definice smysl pravděpodobnost, s jakou testovací statistika nabývá hodnot horších (více svědčící o testované hypotéze), než je pozorovaná hodnota statistiky.

Chí kvadrát rozdělení - Toto rozdělení je odvozeno ze součtu nezávislých náhodných veličin s normovaným normálním rozdělením.



počet bliznových laloků u *Papaver rhoeas*

Průměr, odchylka, variační koeficient

Table 3.1. Calculation of mean, variance, standard deviation and coefficient of variation in number of stigmatic bands in capsules of the Poppy, *Papaver rhoeas* (Data from Pearson, 1900)

Number of bands x	Frequency, f	fx	Difference from mean $x - \bar{x}$	Square of difference $(x - \bar{x})^2$	$f \times$ square of difference $f(x - \bar{x})^2$
5	1	5	- 4.8	23.04	23.04
6	12	72	- 3.8	14.44	173.28
7	91	637	- 2.8	7.84	713.44
8	295	2360	- 1.8	3.24	955.80
9	550	4950	- 0.8	0.64	352.00
10	619	6190	+ 0.2	0.04	24.76
11	418	4598	+ 1.2	1.44	601.92
12	195	2340	+ 2.2	4.84	943.80
13	54	702	+ 3.2	10.24	552.96
14	25	350	+ 4.2	17.64	441.00
15	5	75	+ 5.2	27.04	135.20
16	3	48	+ 6.2	38.44	115.32
	2268	22 327			5032.52

$$\text{Mean} = \frac{22\,327}{2268} = 9.8$$

$$s^2 = \frac{\sum f(x - \bar{x})^2}{n - 1} = \frac{\sum d^2}{n - 1} = \frac{5032.52}{2267} = 2.2 \quad s = 1.49$$

$$\text{Coefficient of variation} = \frac{s}{\bar{x}} \times 100 = \frac{1.49}{9.8} \times 100 = 15.2\%$$

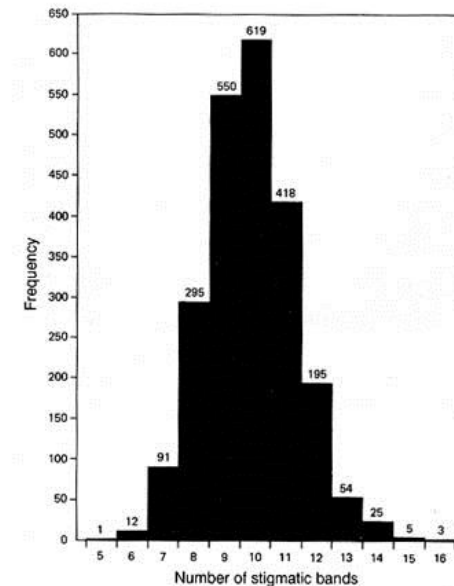


Fig. 3.3. Histogram of Pearson's data (Table 3.1) for the variation in the number of stigmatic bands in a sample of capsules of *Papaver rhoeas*. Such histograms were often used in early biometrical studies. Campbell (1967), discussing the use of histograms, suggests that they should be used only in cases of continuous variation. For examples of meristic or other discontinuous variation, frequency diagrams are preferable; in these, the frequency of each class is indicated by a vertical line on the graph.

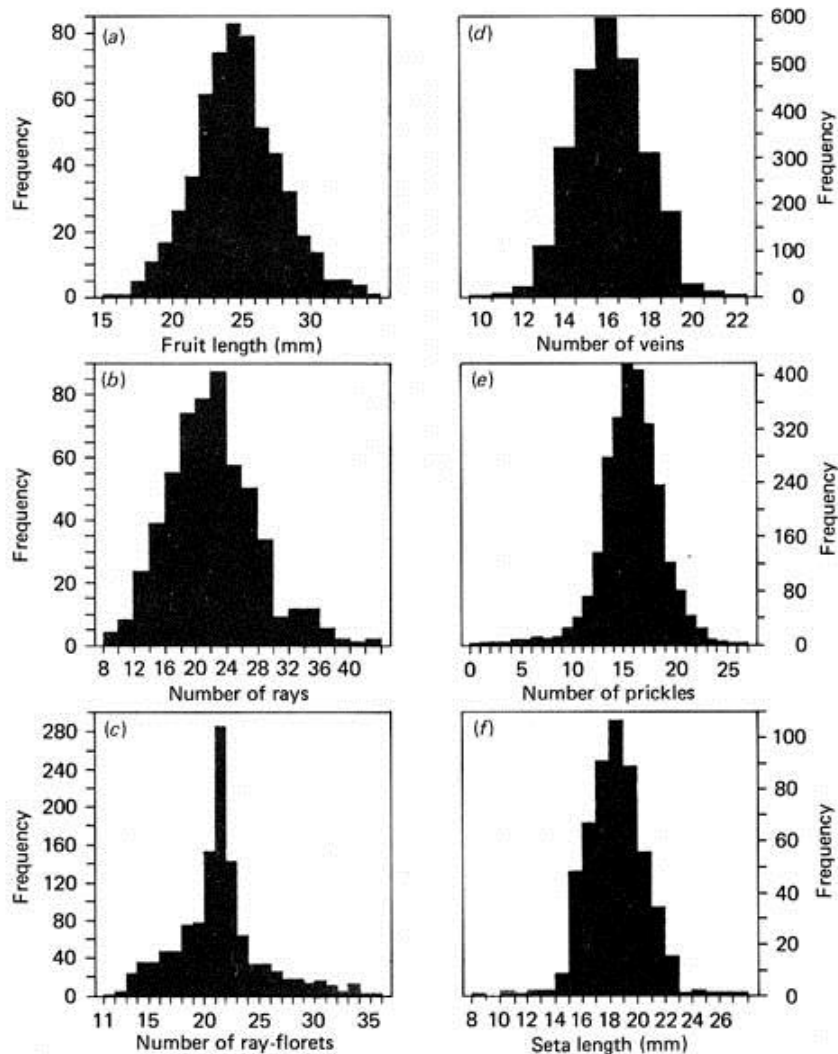


Fig. 3.4. Early botanical results showing approximately normal distributions. (a) Fruit length in the Evening Primrose, *Oenothera glazioviana* (*O. lamarckiana*), for 568 plants collected in October 1893 (De Vries, 1894). (b) Number of primary umbel rays in *Anethum graveolens* for 552 plants collected in July 1893 (De Vries, 1894). (c) Number of ray-florets in Ox-Eye Daisy, *Leucanthemum vulgare* (*Chrysanthemum leucanthemum*), collected from 1133 heads in Keswick, England, July 1895 (Pearson & Yule, 1902). (d) Number of main-branch veins in leaves of Beech (*Fagus sylvatica*) from 2600 leaves (Pearson, 1900). (e) Number of prickles on 2600 leaves of Holly (*Ilex aquifolium*) from trees in Somerset, England (Pearson *et al.*, 1901). (f) Seta length in the moss, *Bryum cirrhatum*, for 522 plants (Amann, 1896).

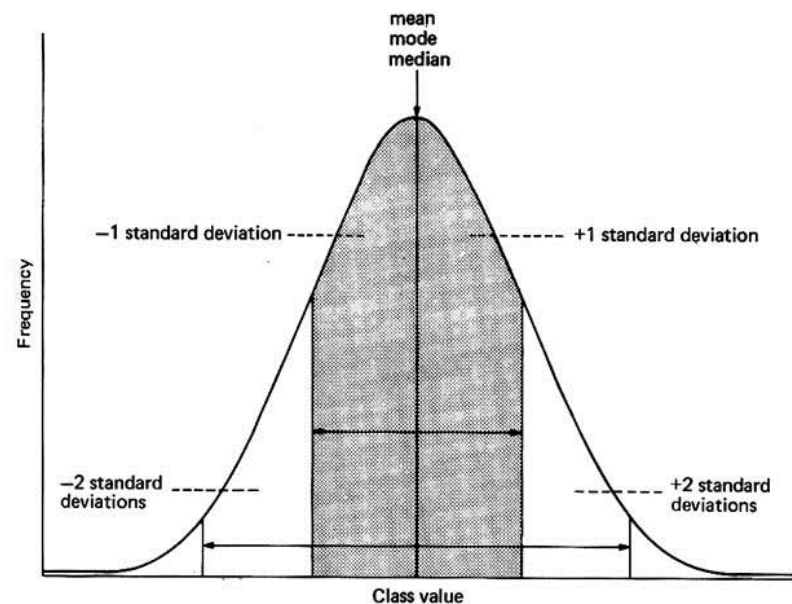


Fig. 3.5. A normal distribution, showing proportions of the distribution that are included between ± 1 standard deviation and ± 2 standard deviations with reference to the mean. (From Srb & Owen, 1958.)

Table 3.5. *Variation in the number of ray-florets in Leucanthemum vulgare* (*Chrysanthemum leucanthemum*)

Number	Ludwig (1895) ^a	5 July 1901 Tower (1902) ^b	30 July 1901 Tower (1902) ^b	Total Tower (1902) ^b
7	2			
8	9			
9	13			
10	36			
11	65			
12	148		1	1
13	427		8	8
14	383		3	3
15	455		6	6
16	479	1	8	9
17	525	0	9	9
18	625	0	8	8
19	856	2	12	14
20	1568	8	19	27
21	3650	17	26	43
22	1790	23	11	34
23	1147	22	10	32
24	812	21	10	31
25	602	22	8	30
26	614	19	5	24
27	375	16	4	20
28	377	14	6	20
29	294	12	4	16
30	196	10	2	12
31	183	16	4	20
32	187	18	2	20
33	307	29	1	30
34	346	20	1	21
35	186	6	0	6
36	64	6	0	6
37	28	0	0	0
38	16	0	0	0
39	16	2	0	2
40	14			
41	0			
42	3			
43	2			
<i>Total</i>	16 800	284	168	452

^aLudwig (1895): Plants from Greiz, Plauen, Altenberg and Leipzig, 1890–5

^bTower (1902): Plants from Yellow Springs, Ohio, USA (two collections and total from same locality)



Morfometrie

odkazuje na kvantitativní analýzu formy, což je pojem, který zahrnuje velikost a tvar. Morfometrické analýzy jsou běžně prováděny na organismech, a jsou užitečné při analýze jejich fosilního záznamu, posouzení vlivu mutací na tvar, vývojové změny, kovariance mezi ekologickými faktory a tvarem, jakož i pro odhad kvantitativně-genetických parametry tvaru.

Morfometrie je možné použít ke kvantifikaci znaků evolučního významu, a tím, že detekuje změny ve tvaru, odvodit něco z jejich ontogeneze, funkce nebo evolučních vztahů.

Hlavním cílem morfometrie je statisticky testovat hypotézy o faktorech, které ovlivňují tvar.

Rozlišují se tři obecné přístupy :

- **tradiční morfometrie**
- **landmark-based morfometrie**
- **outline-based morfometrie**

Morfologické znaky



Stipules-character of anthocyan spot

Flower-wings colour

Flower-vexillum colour

Leaflet-margin shape on the second real leaf

Seed-funiculus stability

Leaflet-margin shape at the first flowering node

Seed-colour at full ripeness

Seed-cotyledons colour

Leaf-type

Seed-hilum colour

Leaflet-colour

Leaflet-shape (in the first flowering node)

Leaflet-apex shape

Seed-testa colour

Seed-surface

Kvantitativní

Plant-seeds number

Plant-pods number

Stem-length to first productive node

Stem-length

Thousand seeds weight

Plant-seeds weight

Stem-length of internode under the first productive node

Stem-number of sterile nodes

Kvalitativní

Morfologické (Kategorické) data podlehající variabilitě vlivem prostředí

ECN	NAZEV	Shluk	Lodyha tvar	Lodyha délka	Lodyha délka do 1. prod. nodu	Lodyha déka pod 1. prod. nodem	Lodyha počet sterilních nodů	Lodyha typ vitví	Lodyha vitvení na bázi	Lodyha olistiní	List typ	Lístek tvar v 1 kvít. nodu	Lístek tvar okraje u 2 prav. listu
	Var. koeficient		0,0	12,5	12,9	20,2	11,3	75,5	50,3	82,7	67,2	23,1	43,0
L0100762	Adept	1	1	4	5	6	5	6	1	7	5	2	1
L0100777	Alan	1	1	4	6	6	6	4	3	7	5	5	1
L0100530	Bohatyr	1	1	4	4	5	5	4	3	6	5	3	1
L0100732	Janus	1	1	4	5	6	5	8	1	6	5	3	2
L0100736	Komet	1	1	4	5	6	5	9	3	6	5	3	2
L0100763	Merkur	1	1	4	5	5	5	9	2	7	5	3	1
L0100766	Pegas	1	1	4	6	5	5	5	2	6	5	2	1
L0100765	Primus	1	1	4	5	6	5	9	2	7	5	2	1
L0100688	Romeo	1	1	4	6	6	5	7	1	6	5	2	1
L0100761	Sonet	1	1	4	5	6	5	6	2	6	5	4	1
L0100978	Baryton	2	1	4	5	6	4	1	1	1	1	0	0
L0100982	Hardy	2	1	4	5	6	4	1	1	1	1	0	0

ECN	NAZEV	Opakování	Lodyha tvar	Lodyha délka	Lodyha délka do 1. prod. nodu	Lodyha déka pod 1. prod. nodem	Lodyha počet sterilních nodů	Lodyha typ vitví	Lodyha vitvení na bázi	Lodyha olistiní	List typ	Lístek tvar v 1 kvít. nodu	Lístek tvar okraje u 2 prav. listu	Lístek tvar okraje v 1 kvít. nodu	Lístek tvar vrcholu	Lístek barva
L0100762	Adept	A	1	4	5	6	5	9	1	7	5	2	1	1	2	7
L0100762	Adept	B	1	4	5	6	5	9	0	7	5	2	1	1	2	7
L0100762	Adept	C	1	4	5	7	5	5	1	5	5	3	1	1	2	3

Tradiční morfometrie

Tradiční morfometrie **analyzuje délky, šířky, hmotnosti, úhly a poměry oblasti. Obecně platí, že tradiční morfometrické údaje jsou měření velikosti.**

Nevýhodou použití mnoha měření velikosti je to, že většina z nich je vysoce korelovaná, jako výsledek, existuje několik nezávislých proměnných navzdory mnoha měření.

Tradiční morfometrika, začala aplikovat multivarietní statistiku na množinu proměnných, získaných měřeními objektu. Obvykle se zpracovávaly lineární vzdálenosti, ale užíval se i součet, podíl (procento, koeficient), úhel. Takto se statisticky porovnávaly odchylky a jejich tendence, vzhledem k nějakému vzoru. Tyto metody přinášely řadu problémů. Měřené vzdálenosti nebyly zcela jasně definované, tedy se měřila například maximální délka objektu, což není úplně ideální, maximální délky mohou dosáhnout různé body, které si fylogeneticky nemusí vůbec odpovídat. Měření může dát pro jiné tvary stejné hodnoty měření, neboť místo měření je relativní (např. tvar elipsy a kapkovitý tvar mohou mít stejnou maximální šířku i délku a přesto jde o tvar na první pohled zcela odlišný). Navíc nelze tvar odvodit zpětně, výsledkem je množina čísel, jež odpovídají vzdálenostem, obvykle uspořádané do matice, ze které je nemožné vykreslit původní tvar. **Předmět zkoumání – tvar objektu – je tedy měřením ztracen.**

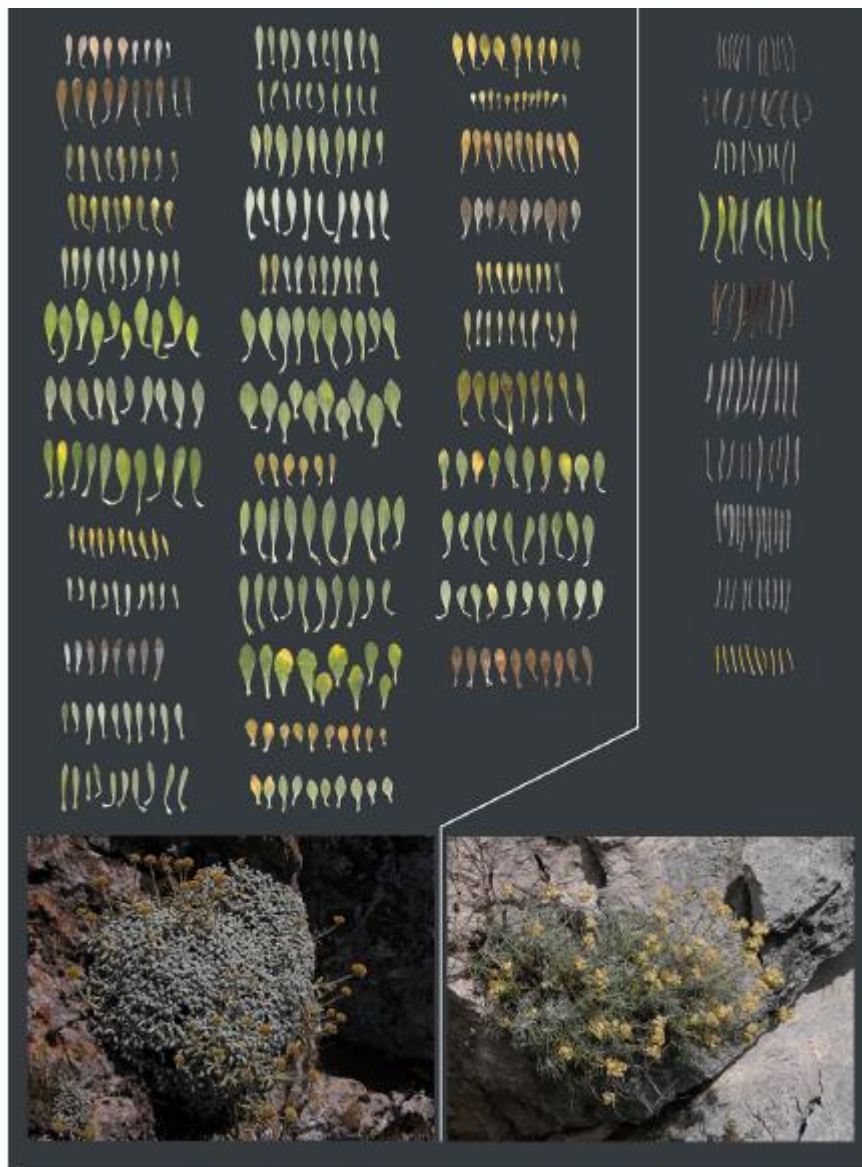


Figure 1. Patterns of morphological variation of *Helichrysum crassifolium* and *Helichrysum pendulum* leaves from the Balearic Islands. Ten leaves from each sampled population are shown.

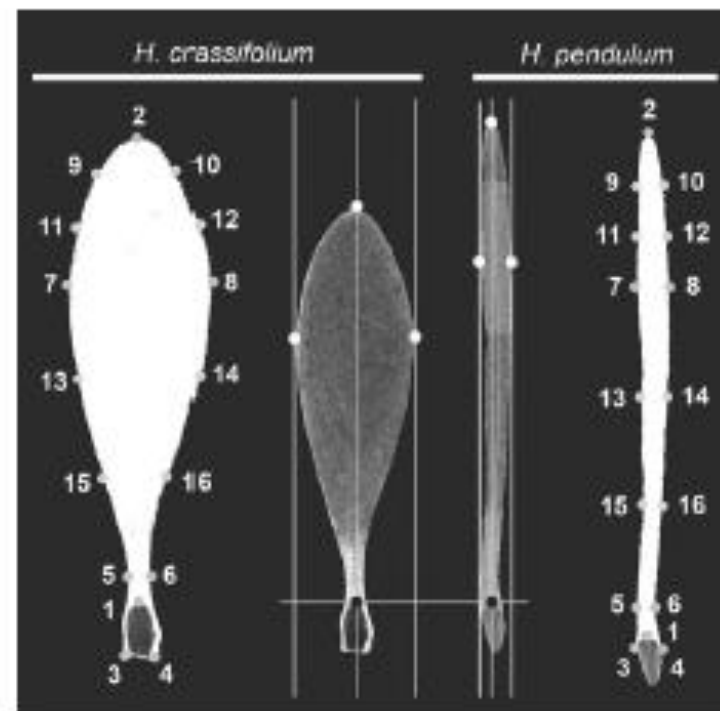
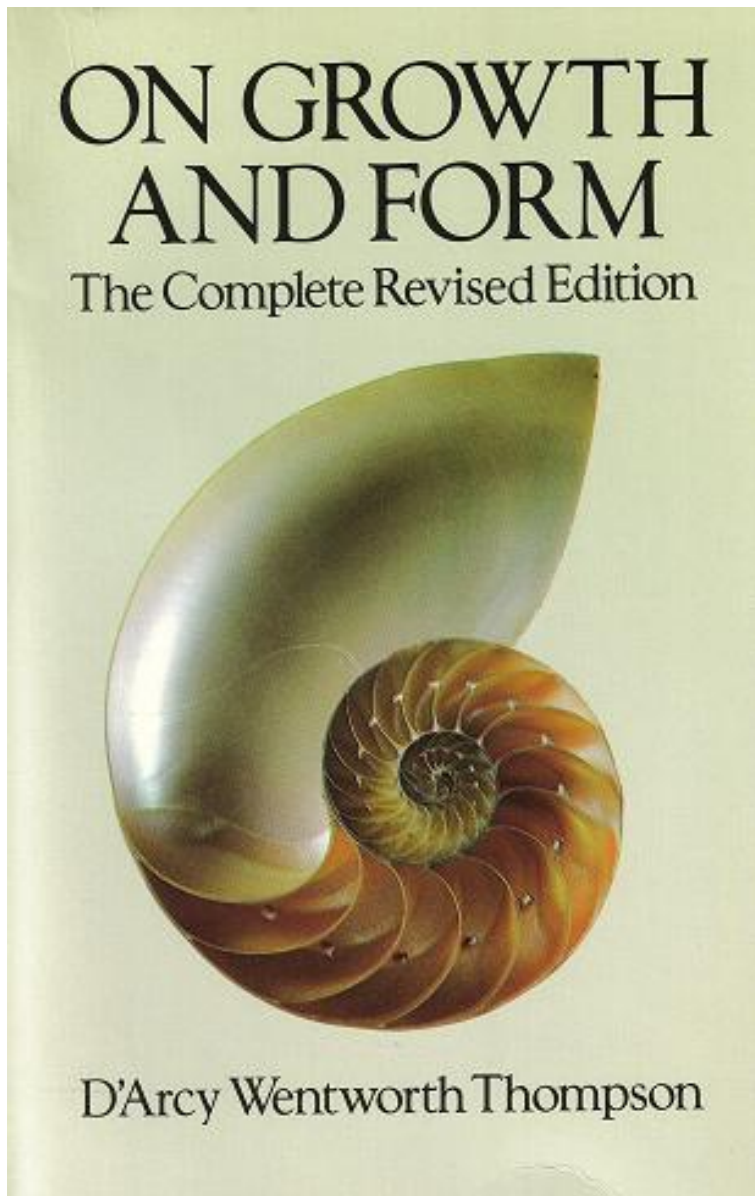


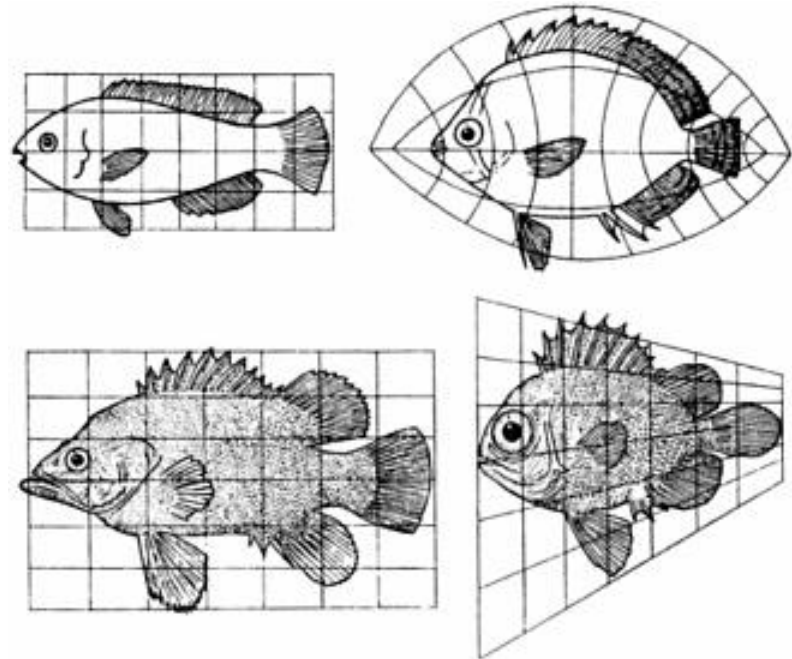
Figure 3. Landmark configuration recorded for each leaf. Representative *H. crassifolium* and *H. pendulum* leaves are shown, with the locations of the 16 landmarks on their digitalized shapes.

D'Arcy Wentworth Thompson (1860 - 1948)



Vydání 1917

Základy biologické matematiky



Transformační mřížky

základní konceptuální východiska transformace tvaru jako objektu biometrického zkoumání poprvé explicitně formuloval právě D'Arcy Thompson.

Jeho filozofie vycházela z platonského přesvědčení, že proměnlivost forem v přírodě je určitým způsobem uspořádána a ovlivňována více či méně jednoduchou a poznatelnou soustavou usměrňujících sil a že složité tvary organismů lze převést na jednoduchý systém pravoúhlých souřadnic. Tvar těla příbuzných druhů je tak možno vzájemně odvodit ze změn (deformací) karteziánské sítě, proložené libovolně zvolenými body či částmi objektu.

Transformační analýza, v zásadě omezená pouze na grafické výstupy, však postrádá jakoukoli kvantifikaci. Navíc Thompson a jeho následovníci kreslili sítě subjektivně, takže myšlené body sloužící jako vodítko při deformaci sítě většinou postrádaly biologické opodstatnění.

A právě tyto skutečnosti byly až donedávna stínem, který – přes snahy několika generací matematicky orientovaných biologů přitahovaných jejím elegantním grafickým stylem – tato metoda nedokázala překročit.

Přístupy založené na landmarcích

Landmark je specifický bod na objektu (např. špička zubu), vybraný dle určitého pravidla. Landmarky jsou biologicky definované (a matematicky významné, právě je pokládáme za základ našich měření). Důležitá je možnost je lokalizovat opakovaně. Pozice těchto landmarků budeme porovnávat na jednotlivých objektech měření. Lze tedy říci, že landmarky jsou významné body, které se nalézají na všech měřených objektech a jež můžeme snadno nalézt.

Většina landmarkových metod přiřadí landmarkům souřadnice, jež poté různými způsoby vyhodnocuje. Přiřazení souřadnic s sebou ale nese jisté problémy – objekty získají pozici a orientaci, již je nutno upravit, mohou být zrcadlově převrácené, a zároveň je zde stále problém s velikostí, takže musely být vyvinuty metody na potlačení těchto deformací.

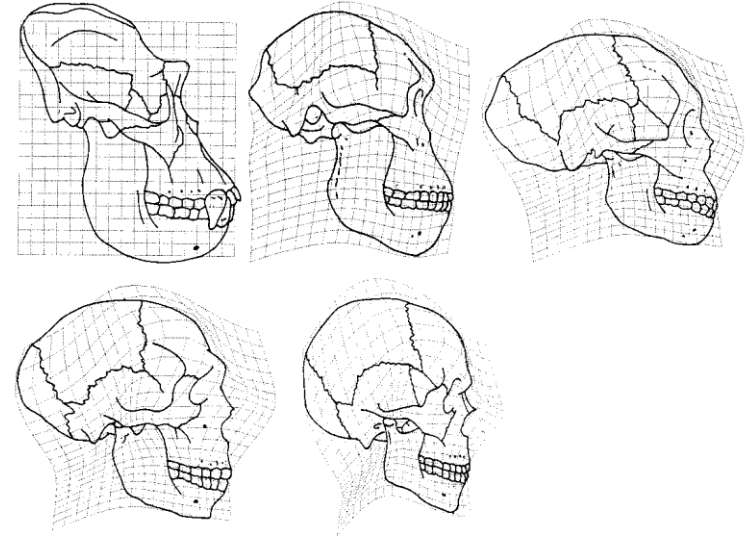
Prokrustova analýza

je rovněž založena na zkoumání tvaru pomocí význačných bodů. Kořeny této metody sahají do přelomu 60. a 70. let, využívat se však začala až v posledním desetiletí. Nazvána byla po bájném obrovi řecké mytologie, který přepadal pocestné, přičemž kratší natahoval na délku svého lože a delší naopak useknutím nohou zkracoval. Toto přiléhavé pojmenování vyjadřuje samu podstatu superpoziční techniky, jejímž cílem je s využitím rotace, posunu a celkové změny velikosti dosáhnout takové konfigurace význačných bodů dvou a více objektů, aby byla optimalizována určitá míra shody, v tomto případě suma druhých mocnin vzdáleností mezi příslušnými homologickými body těchto objektů (podobně jako např. u obyčejné lineární regrese). Proto se této technice někdy říká Least Squares, zkráceně LS, při více než dvou objektech potom Generalized Least Squares, GLS .

Avšak i tato metoda má některá omezení a nedostatky, z nichž nejzávažnější se týkají nemožnosti detekce složitějších tvarových odlišností a lokálních deformací a s tím spojené ztráty podstatné části informace. Nicméně díky své jednoduchosti je Prokrustova analýza i její varianty (např. GRF) obzvláště výhodná jako prostředek předběžné analýzy dat a při testování jednodušších nulových hypotéz.

Geometrická morfometrie

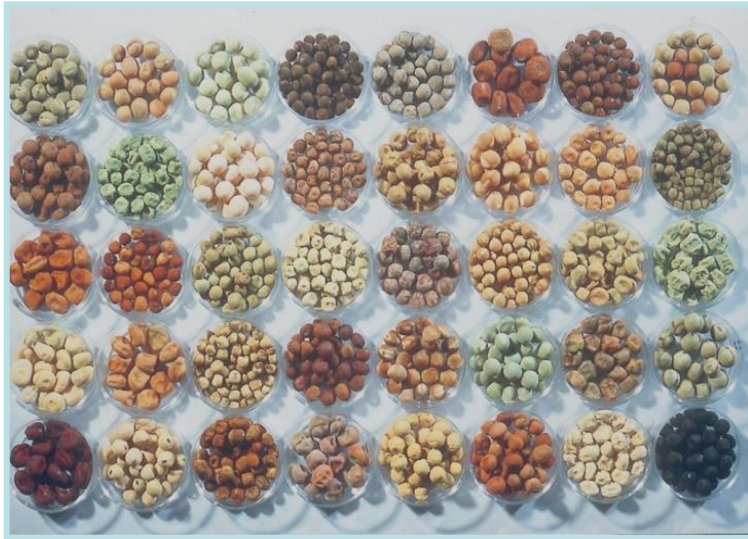
Geometrická morfometrika (nebo také morfometrie) se zajímá studiem tvaru biologických objektů. Aplikuje statistické metody na porovnávání tvaru. Účelem je zjistit, v jakém místě se tvar vyznačuje největší variabilitou, což vede posléze k úvahám dalších vědních disciplín, které zkoumají, proč tomu tak je, nebo jak by se zjištěné výsledky daly využít.



Miloš Macholán, Vesmír 78, 1999/1

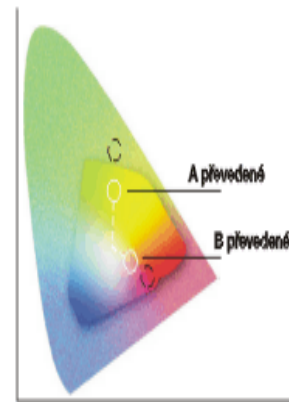
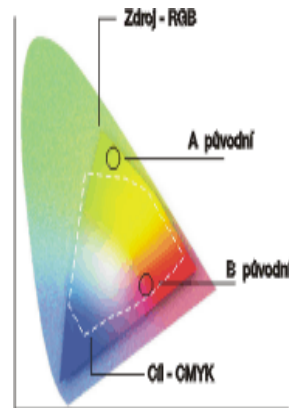
Rozdíly mezi nimi jsou pak ukázány ve formě lokálních deformací původně pravoúhlé sítě. Oproti Thompsonovu přístupu však metoda ohebných pásků postrádá jakoukoli subjektivitu, neboť poloha každého bodu je v prostoru jednoznačně určena. Síť zde není vlastním objektem analýzy jako u Thompsona, ale naopak prostředkem zobrazení, komunikačním médiem zprostředkujícím výsledky biologické analýzy.

Popis fenotypového projevu



Semeno-barva v plné zralosti

1. světležlutá
2. žlutorůžová
3. vosková/dvoubarevná
4. žlutozelená
5. šedozelená
6. tmavozelená
7. světlehnědá
8. hnědá
9. černá



Číselné zhodnocení
průměrných pozic souřadnic
L,a,b v CIE-LAB barevném
prostoru (3D)

Colour deviation:

$\Delta E_{a,b}$

$$\Delta E_{a,b} = (\Delta L^2 + \Delta a^2 + \Delta b^2)^{1/2}$$

Linum: velikost a tvar semen; parametry průměrné hodnoty

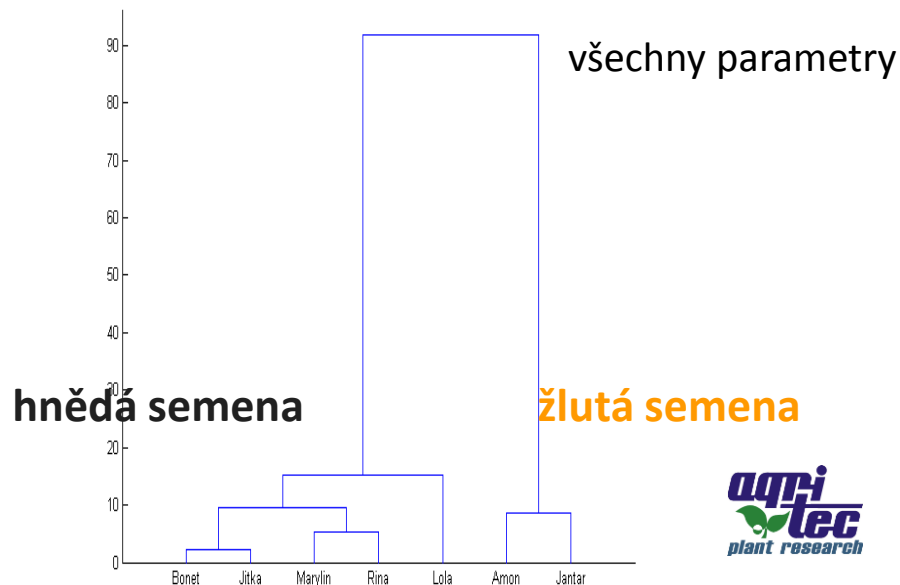
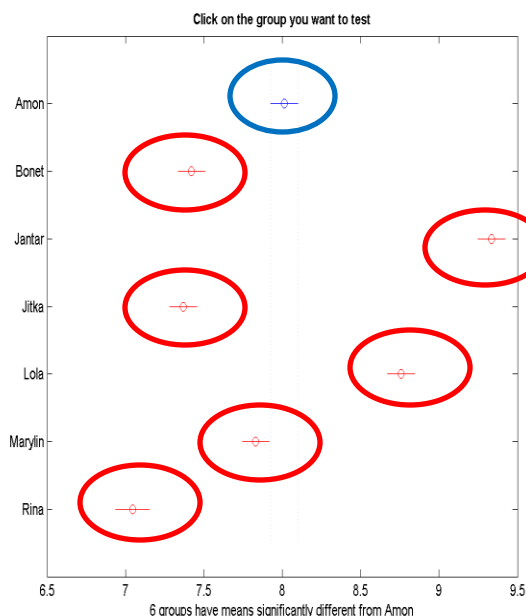
Kalibrace Area – 0,08 µm/px

Flax	Area	EqDiameter	Perimeter	MaxFeret	MinFeret	Circularity	Elongation
Amon	8,01 ^a	3,192	11,190	4,585	2,338	0,804	1,962
Bonet	7,42 ^{ab}	3,056	12,033	4,814	2,266	0,755	2,065
Jantar	9,34 ^a	3,445	12,082	4,953	2,518	0,802	1,969
Jitka	7,37 ^{ab}	3,061	10,584	4,164	2,398	0,828	1,738
Lola	8,76 ^a	3,336	11,851	4,731	2,522	0,785	1,878
Marylin	7,83 ^{ab}	3,155	11,367	4,471	2,393	0,771	1,873
Rina	7,04 ^b	2,992	10,269	4,125	2,298	0,839	1,798

Statistické rozdíly
(ANOVA, $\alpha < 0.05$)

Vyhodnocení dat v programu MATLAB R2009a ... shluková analýza

Area
statisticky
odlišné
od odrůdy
AMON



Example of fruit shape and fruit color

Phenotyping as an objective and archived resource: Tomato Analyzer
(Brewer et al., *Plant Physiology*, 2006, **141**, 15-25; Darrigues et al., 2008 *J. Amer. Soc. Hort. Sci.* 133:579-586)

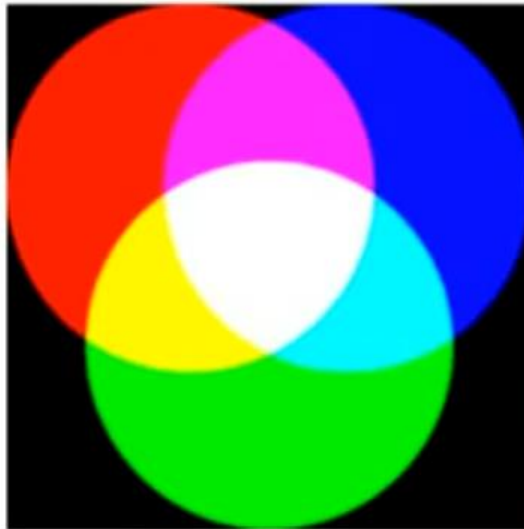


In many cases tools already exist to collect high quality objective phenotypic data

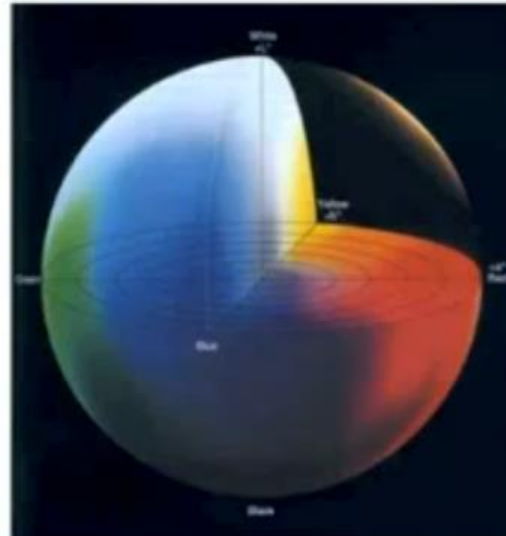
Example of fruit color

Where we need to go: from descriptor to objective data.

1 = GREEN, 2 = YELLOW (GRIN)



R, G, B



CIELAB

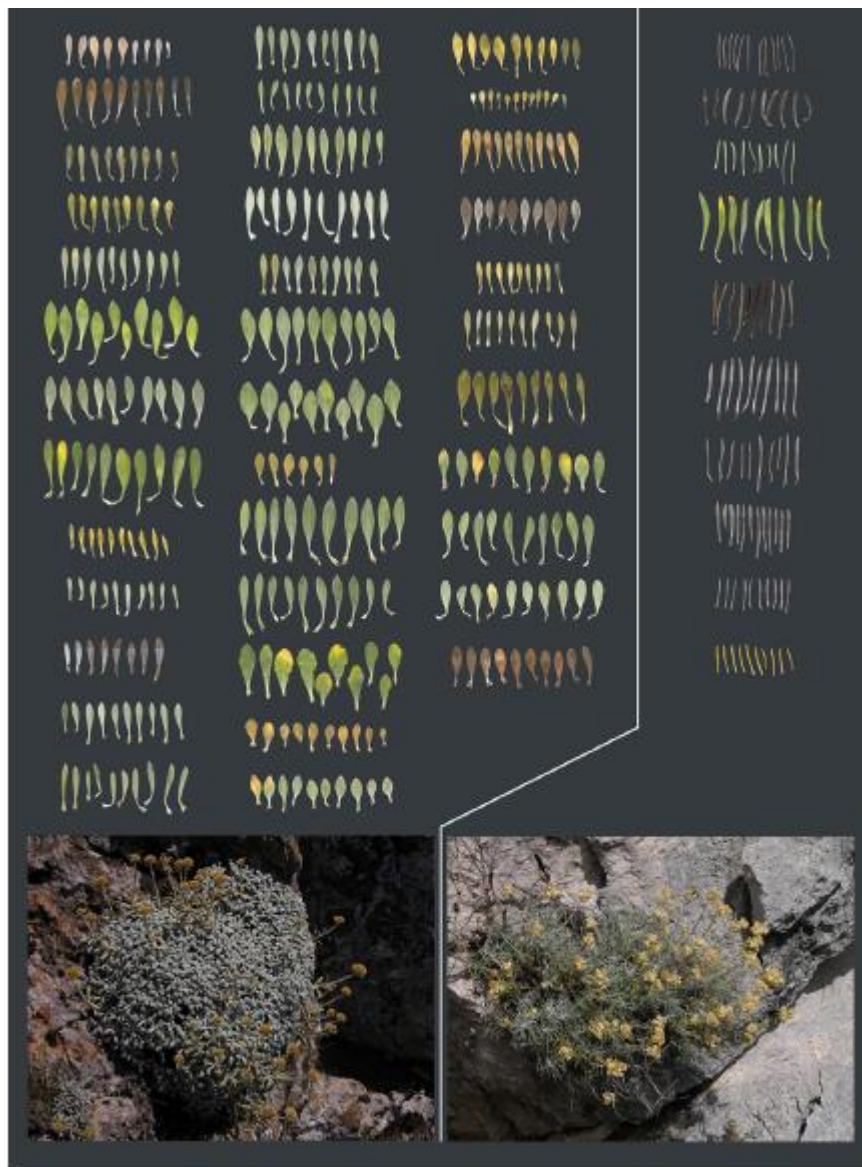


Figure 1. Patterns of morphological variation of *Helichrysum crassifolium* and *Helichrysum pendulum* leaves from the Balearic Islands. Ten leaves from each sampled population are shown.

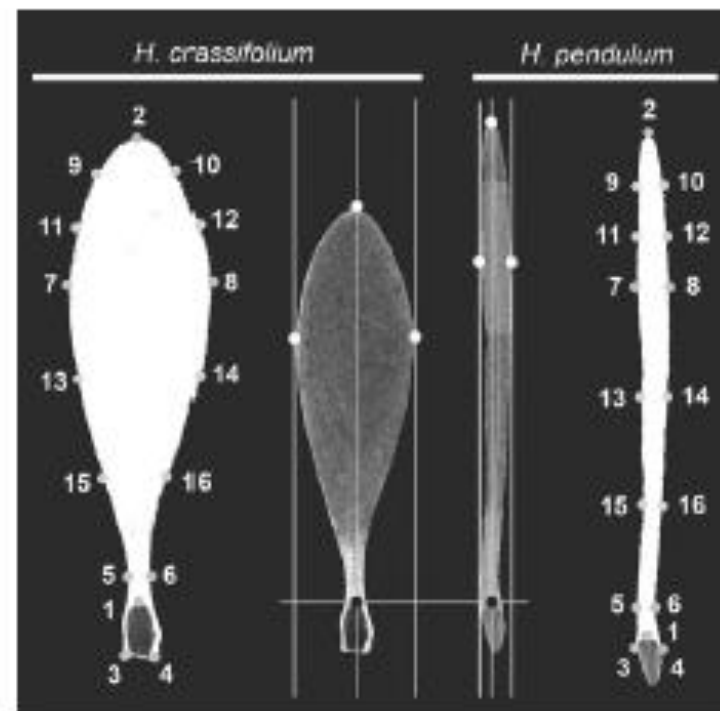


Figure 3. Landmark configuration recorded for each leaf. Representative *H. crassifolium* and *H. pendulum* leaves are shown, with the locations of the 16 landmarks on their digitalized shapes.

Table 2. Linear measurements (leaf length, leaf width, and ratio, L/W) and associated one-way ANOVA *F*-values

Variable	Species	Range	Mean	SD	SE	<i>F</i>
Length (mm)	CRA	18.6–83.0	44.98	10.74	0.59	2.67
	PEN	21.0–97.8	47.26	13.86	1.51	
Width (mm)	CRA	4.2–24.2	11.17	3.28	0.18	483.35*
	PEN	1.1–8.4	3.06	1.59	0.17	
L/W	CRA	1.8–9.4	4.21	1.09	0.06	1183.23*
	PEN	6.4–43.8	17.93	7.00	0.76	

Sample sizes for *Helichrysum crassifolium* (CRA) and *Helichrysum pendulum* (PEN) are $n = 333$ and 84 , respectively.
* $P < 0.0001$.

© 2012 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2012, **106**, 498–513

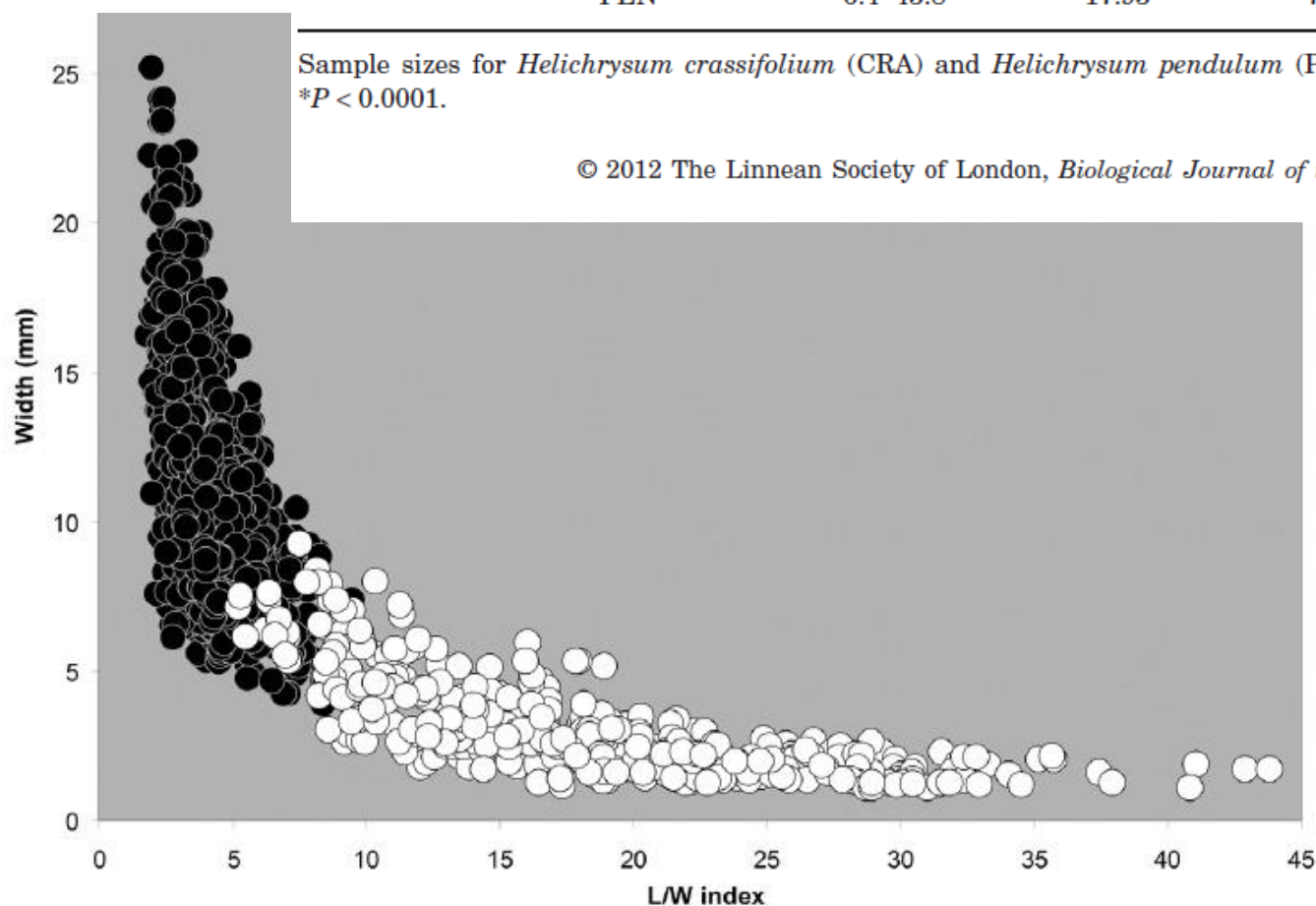


Figure 4. Scatter plot of the leaf length/width ratio and width values for *Helichrysum crassifolium* (black circles) and *Helichrysum pendulum* (white circles).

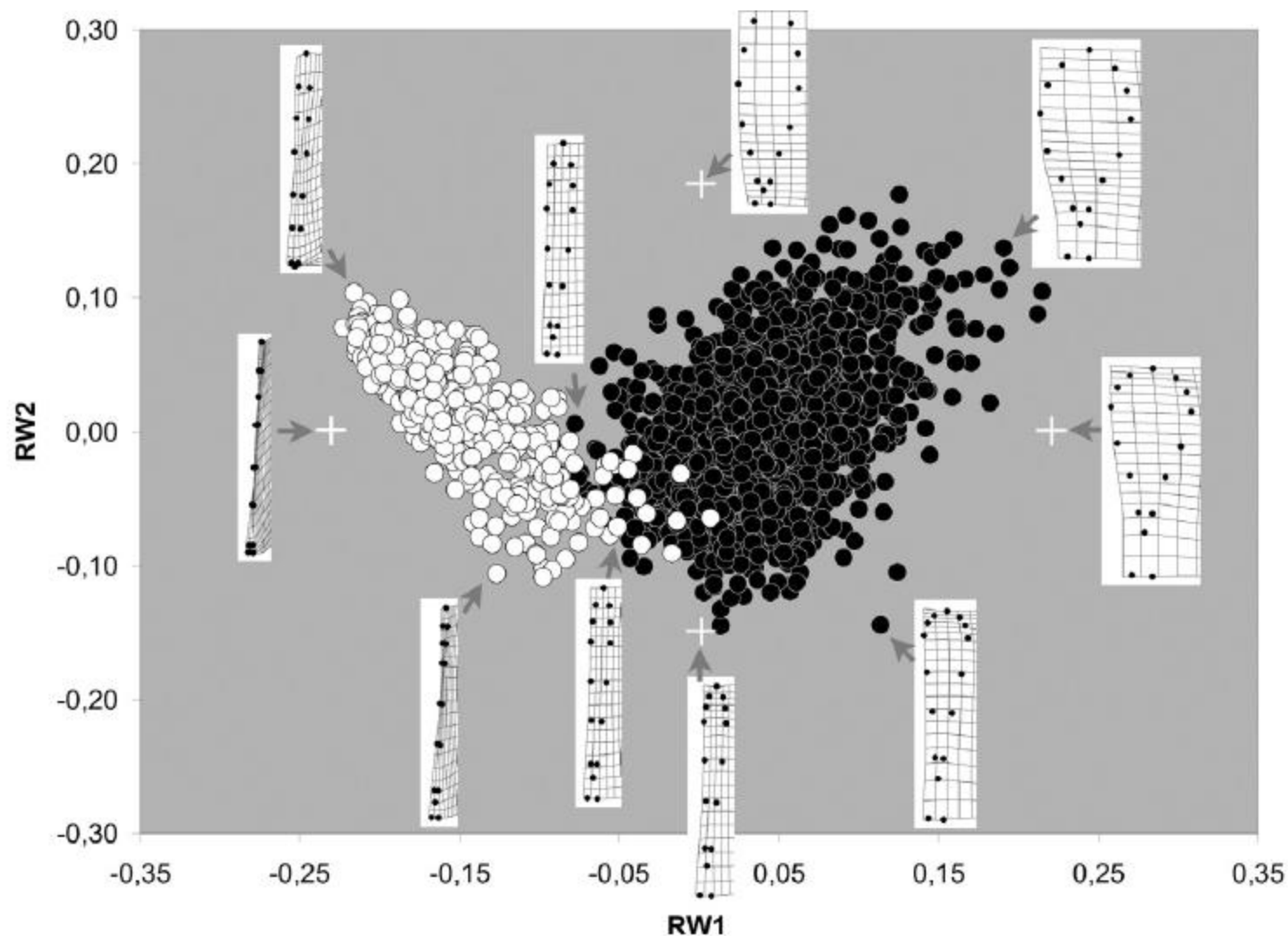


Figure 5. Scatter plot of the two first relative warps obtained with tpsRELW on 16 leaf landmarks from 2085 samples (black circles, *Helichrysum crassifolium*; white circles, *Helichrysum pendulum*). Thin-plate splines analysis (TPS) deformation grids for some extreme specimens have been situated around the plot (denoted with arrows). Virtual ranges of the two axes are marked with a white cross. The corresponding deformation grids are also depicted.

Leaf Morphology, Taxonomy and Geometric Morphometrics: A Simplified Protocol for Beginners

Vincenzo Viscosi^{1*}, Andrea Cardini^{2,3,4}

1 Museo Erbario del Molise, Dipartimento di Scienze e Tecnologie per l'Ambiente e il Territorio, Università del Molise, Contrada Fonte Lappone, Pesche, Italy, **2** Dipartimento di Biologia, Università di Modena e Reggio Emilia, Modena, Italy, **3** Hull York Medical School, The University of Hull, Hull, United Kingdom, **4** Centre for Forensic Science, The University of Western Australia, Crawley, Western Australia, Australia

Abstract

Taxonomy relies greatly on morphology to discriminate groups. Computerized geometric morphometric methods for quantitative shape analysis measure, test and visualize differences in form in a highly effective, reproducible, accurate and statistically powerful way. Plant leaves are commonly used in taxonomic analyses and are particularly suitable to landmark based geometric morphometrics. However, botanists do not yet seem to have taken advantage of this set of methods in their studies as much as zoologists have done. Using free software and an example dataset from two geographical populations of sessile oak leaves, we describe in detailed but simple terms how to: a) compute size and shape variables using Procrustes methods; b) test measurement error and the main levels of variation (population and trees) using a hierarchical design; c) estimate the accuracy of group discrimination; d) repeat this estimate after controlling for the effect of size differences on shape (i.e., allometry). Measurement error was completely negligible; individual variation in leaf morphology was large and differences between trees were generally bigger than within trees; differences between the two geographic populations were small in both size and shape; despite a weak allometric trend, controlling for the effect of size on shape slightly increased discrimination accuracy. Procrustes based methods for the analysis of landmarks were highly efficient in measuring the hierarchical structure of differences in leaves and in revealing very small-scale variation. In taxonomy and many other fields of botany and biology, the application of geometric morphometrics contributes to increase scientific rigour in the description of important aspects of the phenotypic dimension of biodiversity. Easy to follow but detailed step by step example studies can promote a more extensive use of these numerical methods, as they provide an introduction to the discipline which, for many biologists, is less intimidating than the often inaccessible specialist literature.

Heritabilita a vzájemná korelace znaků

Matrix of eigenvalues and vectors of principal components for 15 qualitative characters of field and fodder pea assessed in morphological trials

	Principal components (PC)					
	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalues						
Variance	4.429	2.783	1.838	1.180	0.867	0.790
% Total contribution	29.53	18.55	12.26	7.87	5.79	5.27
% Accumulated	29.53	48.09	60.35	68.22	74.00	79.27
Eigenvectors						
Stipules-character of anthocyan spot	0,917	-0,123	-0,256	-0,054	-0,095	0,010
Flower-wings colour	0,898	-0,094	-0,280	-0,037	-0,166	-0,023
Flower-vexillum colour	0,876	-0,078	-0,253	-0,102	-0,187	-0,087
Leaflet-margin shape on the second realleaf	0,084	0,694	0,282	-0,531	-0,080	0,069
Seed-funiculus stability	0,269	-0,171	0,677	-0,090	0,518	0,151
Leaflet-margin shape at the first flowering node	0,080	0,735	0,294	-0,414	-0,146	0,116
Seed-colour at full ripeness	0,805	-0,086	0,344	0,195	-0,140	-0,055
Seed-cotyledons colour	0,078	0,000	0,736	0,430	-0,255	-0,119
Leaf-type	0,235	0,730	-0,105	0,268	0,210	-0,216
Seed-hilum colour	0,575	-0,165	0,062	-0,048	0,491	-0,393
Leaflet-colour	0,101	0,730	0,024	0,058	-0,036	-0,443
Leaflet-shape (in the first flowering node)	0,234	0,460	-0,384	0,246	0,305	0,406
Leaflet-appex shape	0,165	0,615	-0,122	0,458	0,024	0,227
Seed-testa colour	0,734	-0,096	0,046	-0,284	0,114	0,181
Seed-surface	0,509	-0,048	0,406	0,230	-0,148	0,277

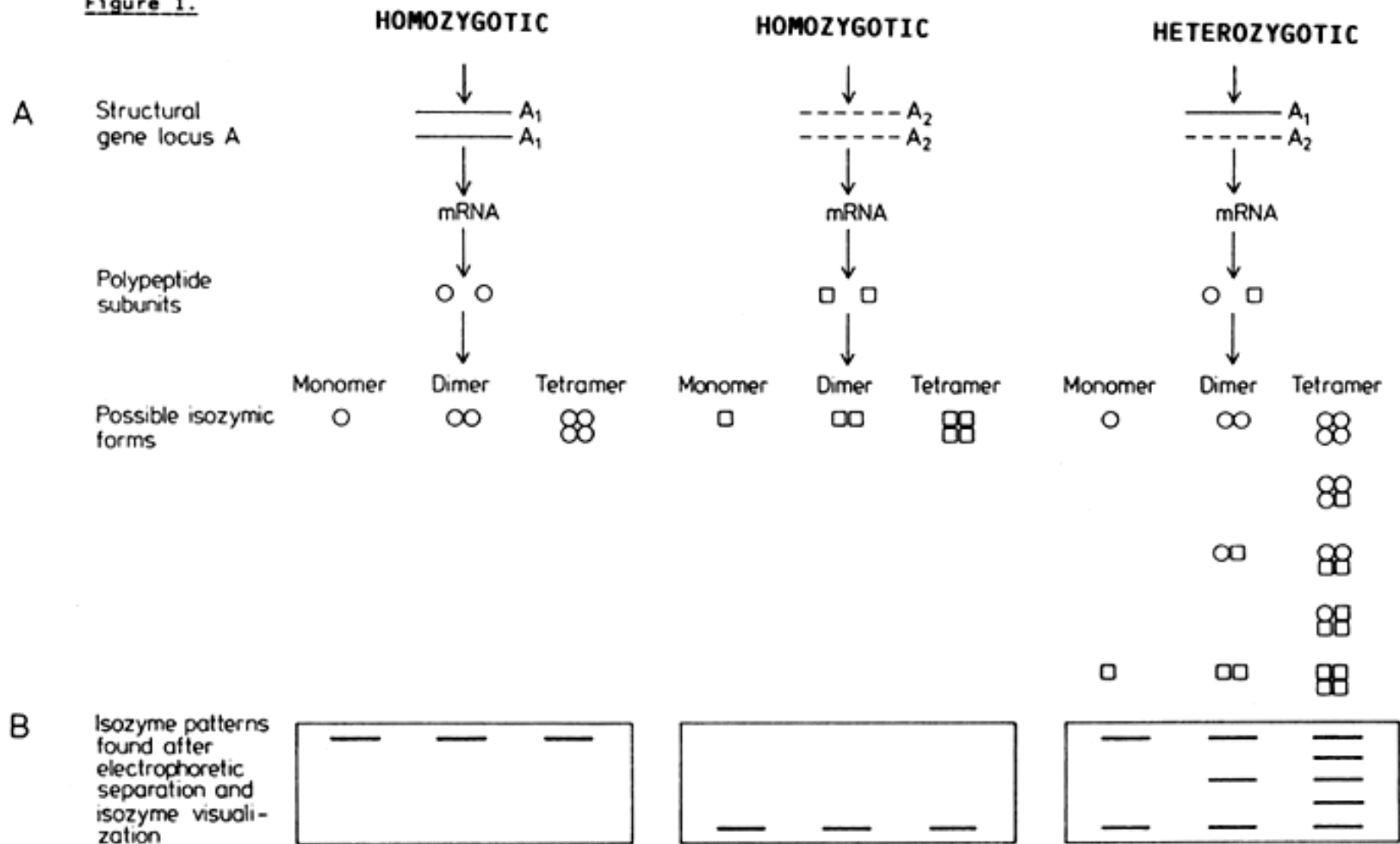
Biochemické markery

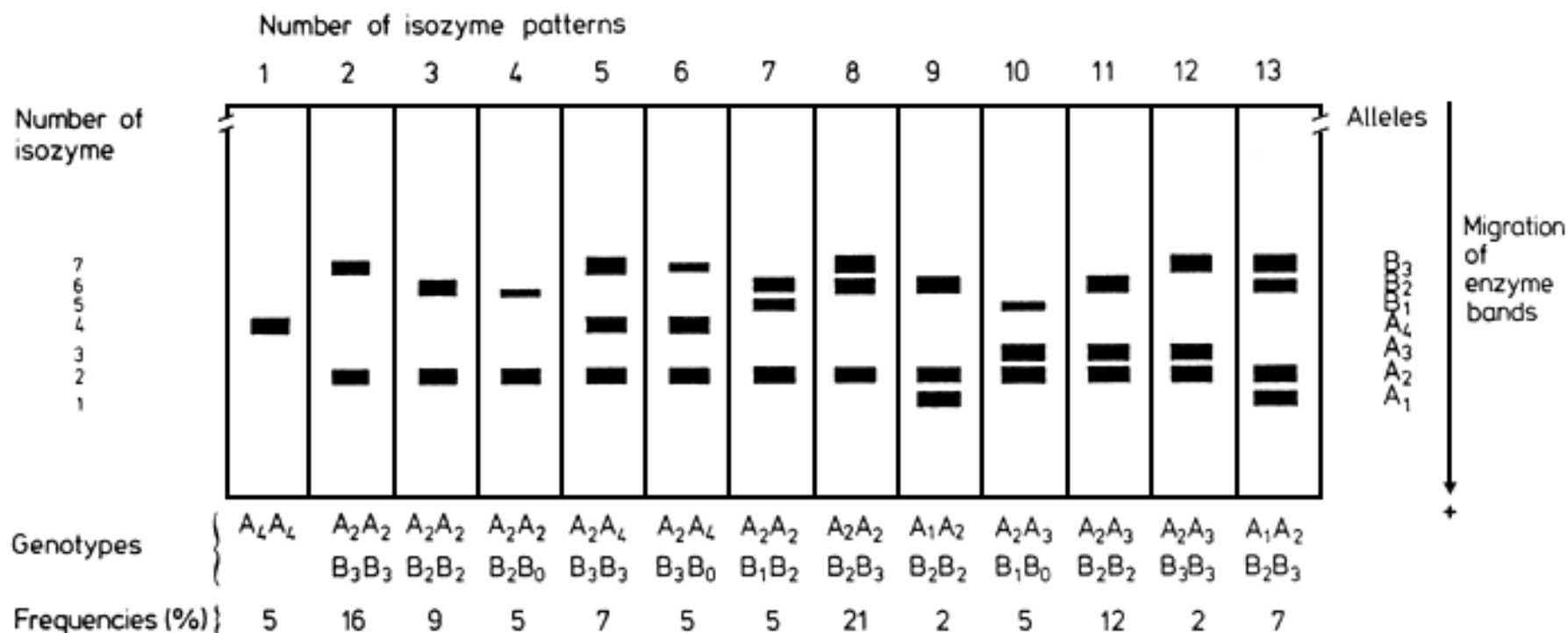
Isozymy (isoenzymy) – enzymy odlišující se v aminokyselinové sekvenci ale katalyzující stejnou chemickou reakci

poprvé popsány RL Hunter a Clement Markert (1957), kteří je definuje jako různé varianty téhož enzymu, které mají stejnou funkci ve stejném jedinci.

- enzymové varianty, které jsou výsledkem různých genů, a tím představují různé lokusy (isoenzymy, isozymy)
- enzymy, které jsou výsledkem různých alel téhož genu (allozymy)

Figure 1.





Ukázka dostupných isozymových detekčních systémů

Aspartate aminotransferase	Aat-A 3 4 2 2	Aat-B 3 4 1 3	Aat-C 1 4 3		Lagercrantz 1988 Lundquist 1979 Poulsen et al. 1983 Muona et al. 1987
Acid phosphatase	Acp-A 1 7	Acp-B 4			Bergmann 1978 Lunquist 1977
Diaphorase	Dia-A 2	Dia-B 1			Lagercrantz 1988
Esterase	Est-A 3	Est-B 3	Est-C 2		Bergmann 1973
Glucose-6-phosphate dehydrogenase	Gpdh-A 2 3				Scholz & Bergmann 1984 Muona et al. 1987
Glutamate dehydrogenase	Gdh-A 2 2	Gdh-A 2	Gdh-B		Scholz & Bergmann 1984 Lundquist 1979 Lagercrantz 1988
Glucose-6-phosphate dehydrogenase	Gpi-A 5	Gpi-A 1	Gpi-B		Lagercrantz 1988
Isocitrate dehydrogenase	Idh-A 2 3	Idh-A 2 1	Idh-B		Scholz & Bergmann 1984 Lagercrantz 1988
Leucine aminopeptidase	Lap-A 1	Lap-A 4	Lap-B		Bergmann 1973
Malate dehydrogenase	Mdh-A 1	Mdh-A 2	Mdh-B 2 4	Mdh-C 2	Mdh-D Poulsen et al. 1983 Scholz & Bergmann 1984 Lagercrantz 1988
Phosphoglucose isomerase	Pgi-A ?	Pgi-A 2 4	Pgi-B	?	Muona et al. 1987
6-Phospho-gluconate dehydrogenase	Pgd-A 2 4	Pgd-A 4 1	Pgd-B 3 1	Pgd-C	Poulsen et al. 1983 Lagercrantz 1988
Phosphogluco-mutase	Pgm-A 2 2	Pgm-A 2 2	Pgm-B		Poulsen et al. 1983 Scholz & Bergmann 1984
Superoxide dismutase	Sod-A 1	Sod-A 1	Sod-B		Lagercrantz 1988

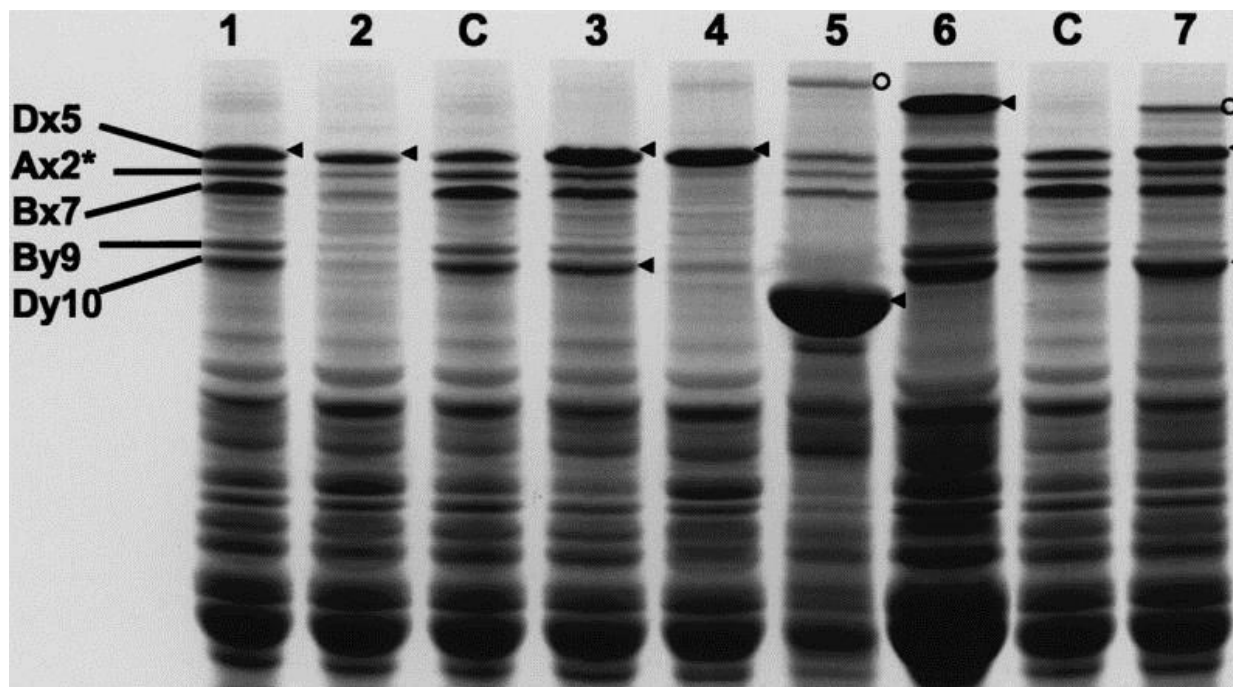
Výhody isozymových markerů:

- Univerzálnost
- Jednoduchost stanovení
- Vysoký polymorfismus
- Ko-dominance (možnost detekce heterozygotů)

Nevýhody isozymových markerů:

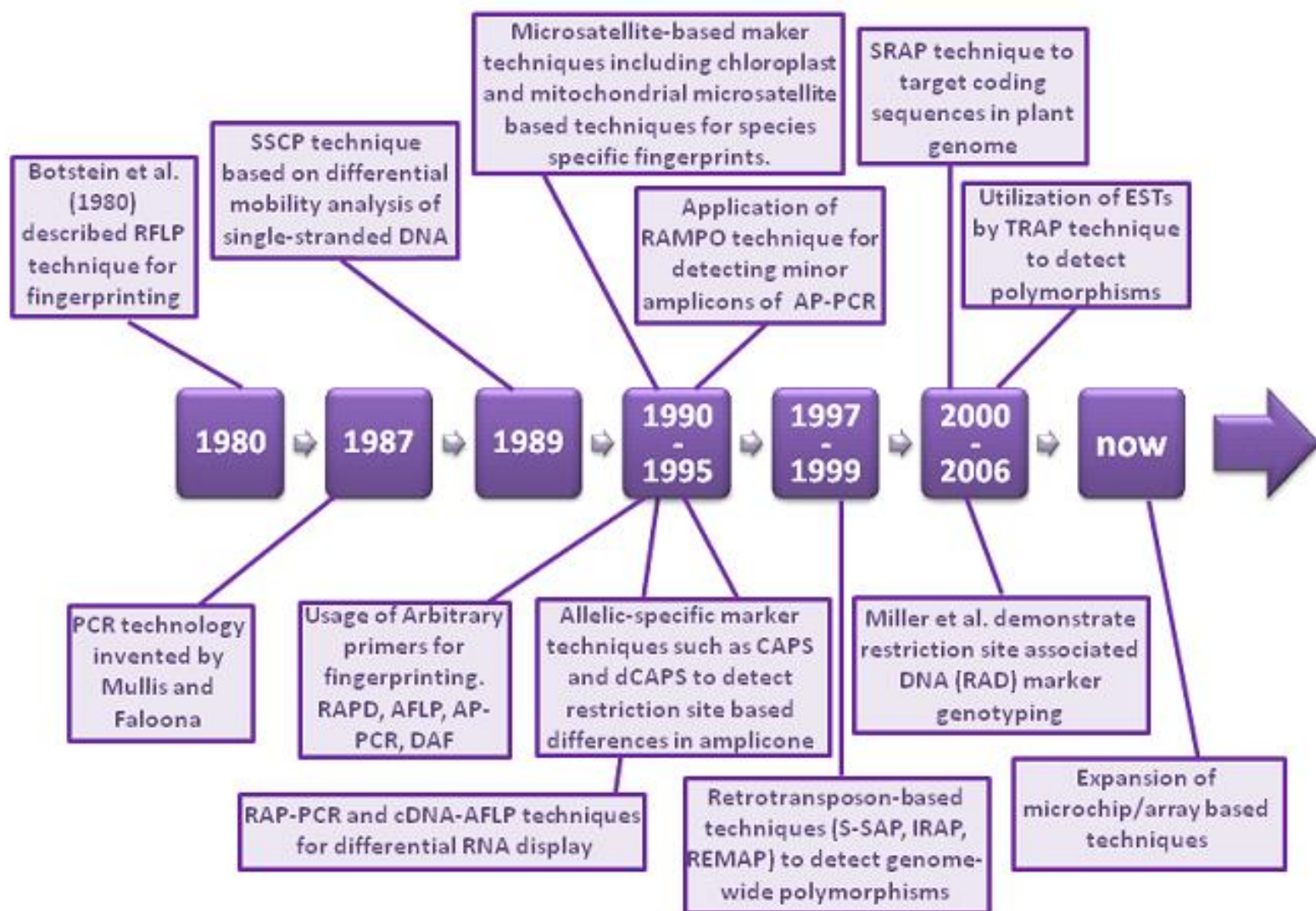
- Nutnost práce s čerstvým materiálem
- Závislost na ontogenetickém stáří a typu pletiva
- Vliv prostředí – problém porovnání mezi lokalitami, lety
- Limitovaný polymorfismus
- Dostupnost přibližně 40 isozymových systémů

Proteiny zásobních orgánů – semen/hlíz



SDS-PAGE (denaturační podmínky elektroforetické separace)

Podobné výhody i nevýhody jako isozymy, menší polymorfismus, ale i menší závislost na prostředí



Genetický marker je známá sekvence DNA, která může být jednoduše identifikována.

Lze jej popsat jako jistý druh variace, způsobený mutací či pozměněním původní sekvence, který se sleduje na předem daném místě.

Genetickým markerem může být krátká sekvence DNA, například úsek DNA se změnou jednoho páru bází (SNP marker), případně delší repetitivní úsek, zvaný mikrosatelit.

Některé běžné typy genetických markerů jsou

- **RFLP (neboli Restriction Fragment Length Polymorphism)**
- **AFLP (neboli Amplified Fragment Length Polymorphism)**
- **RAPD (neboli Random Amplification of Polymorphic DNA)**
- **VNTR (neboli Variable Number of Tandem Repeat)**
- **SNP (neboli Single Nucleotide Polymorphism)**
- **STR (neboli Short Tandem Repeat)**

RAPD

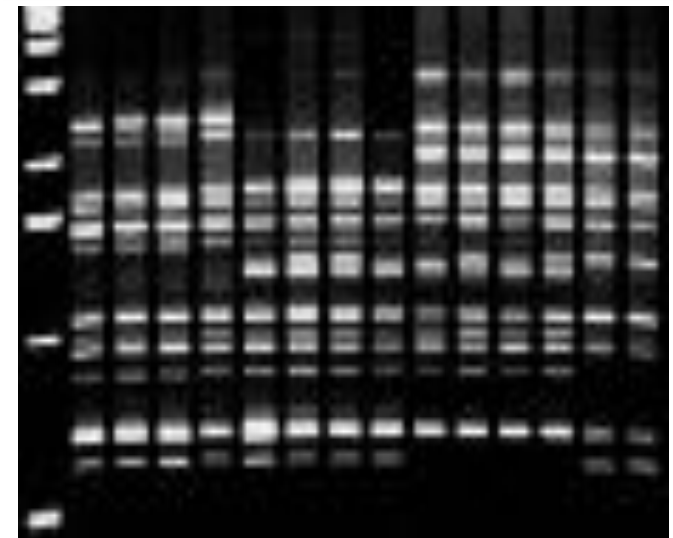
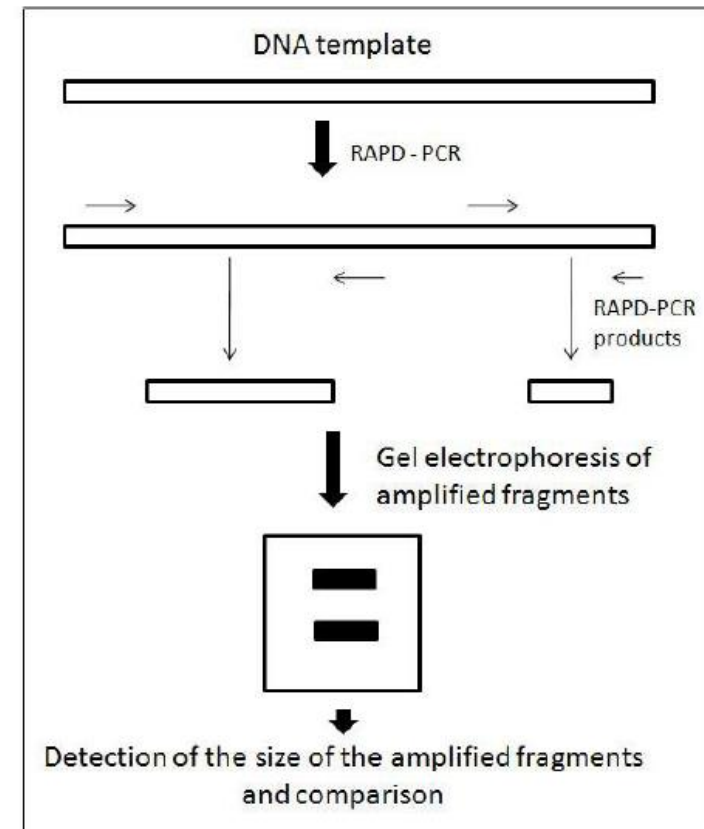
(Random Amplification of Polymorphic DNA)

Není zapotřebí žádné znalosti sekvence DNA pro cílené geny, není to nutné, protože primery se váží náhodně.

RAPD markery jsou decamery (10 nukleotidů délky) vedoucí k PCR amplifikace náhodných úseků genomové DNA, které jsou schopny rozlišit mezi geneticky odlišných jedinců, i když ne nutně reprodukovatelným způsobem

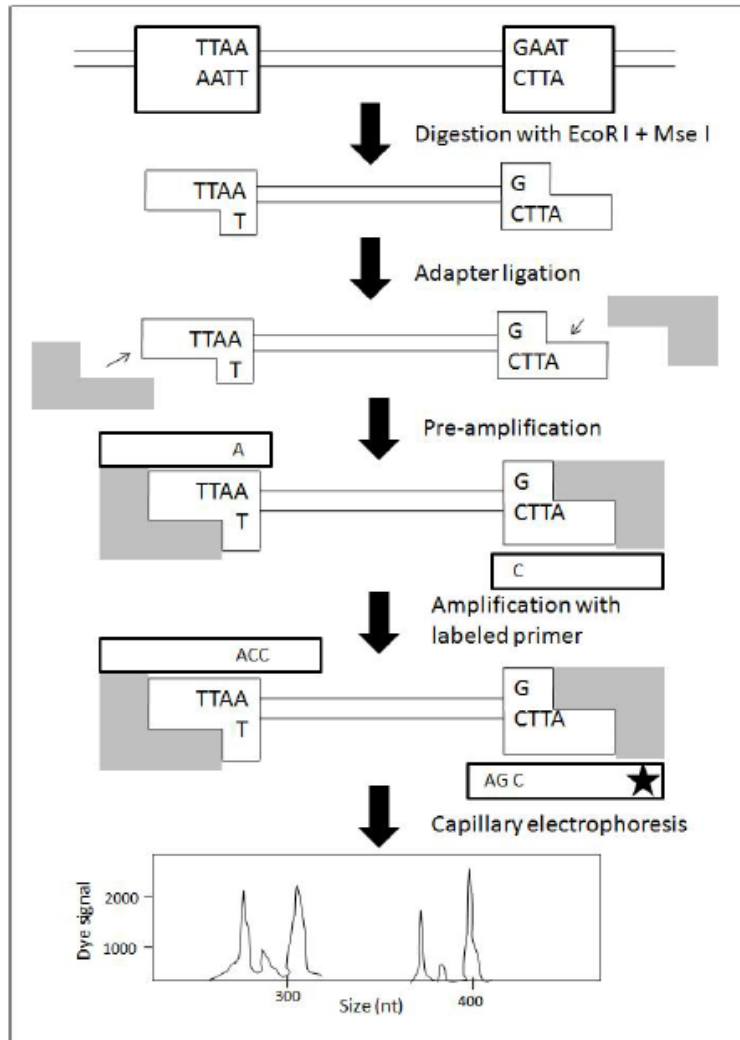
Nevýhody:

Dominantní marker, nízká reprodukovatelnost, homoplasie fragmentů



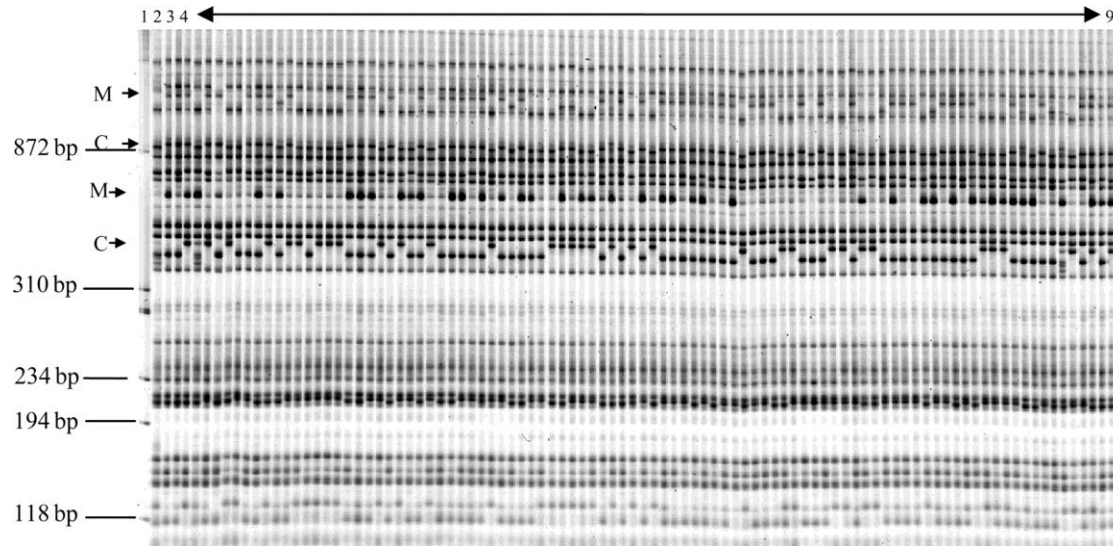
AFLP

(Amplified Fragment Length Polymorphism)



AFLP

(Amplified Fragment Length Polymorphism)



Výhody:

Univerzálnost

Není nutné znát sekvenci DNA

Vysoký polymorfismus

Fingerprinting – hodně dat/experiment

Nevýhody:

Technická a metodická náročnost

Riziko homoplasie

Mikrosatelitní markery

Typy- klasifikace (dle sekvence):

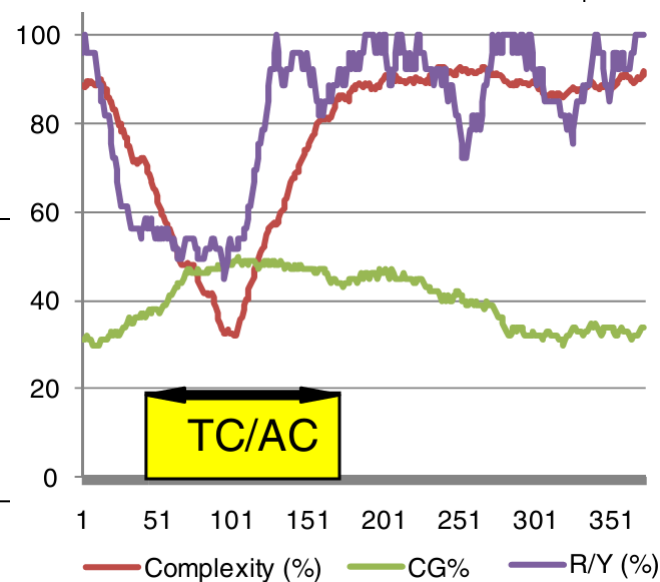
- **Perfect / simple perfect** - složené z jednoho motivu $(CA)_n$
DI/TRI/TETRA – CA/CCA/CCAA
- **Simple imperfect repeats** - přerušovaný motiv $(CA)_n N (CA)_n$
- **Compound repeats** - složené ze 2 a více motivů $(CA)_n (GA)_n$
- **Interrupted compound repeats** - přerušované, složené s mutací
 $(CCA)_n NN(GCA)_{n+1}$

Mikrosatelitní markery

V genech - v kódujících oblastech (exonech) často Trinukleotidové

- v intronech , UTR oblastech, promotorech

Mezi geny - intergenerické



Mechanismus vzniku

- **Single-strand slippage** (svezení) DNA polymerasy během replikace
- **Unequal crossing over** a genová konverze během DNA rekombinace
- Interakce mezi replikační svezením a rekombinací

Mikrosatelitní markery - detekce

1) PCR amplifikace

----- Forward primer--->>>>>>

[illegible]

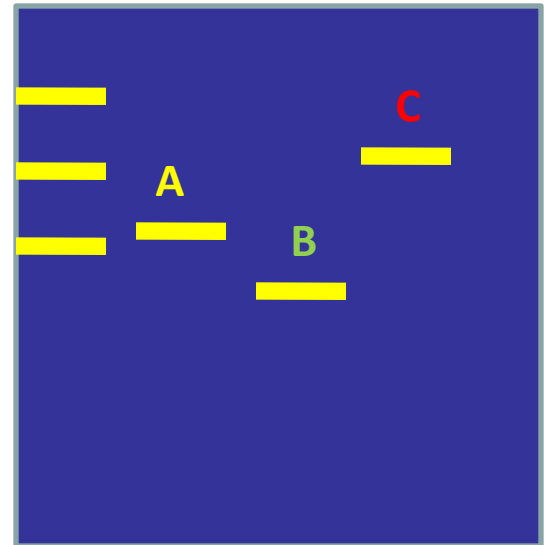
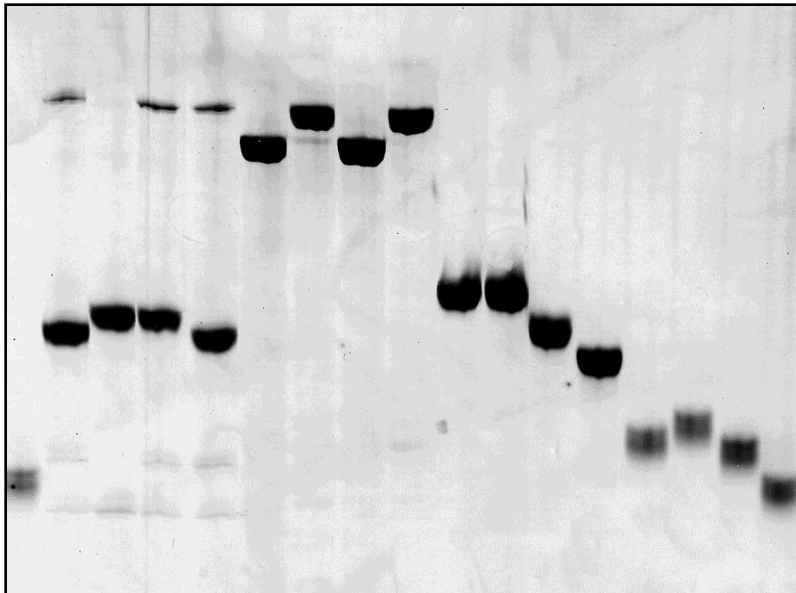
<<<< ---Reverse primer -----

Nutnost znát sekvenci přilehlou k repetici – tj. druhová specificita což je nejnákladnější pro vývoj nových SSR markerů pro daný druh a omezuje to následně i transferibilitu

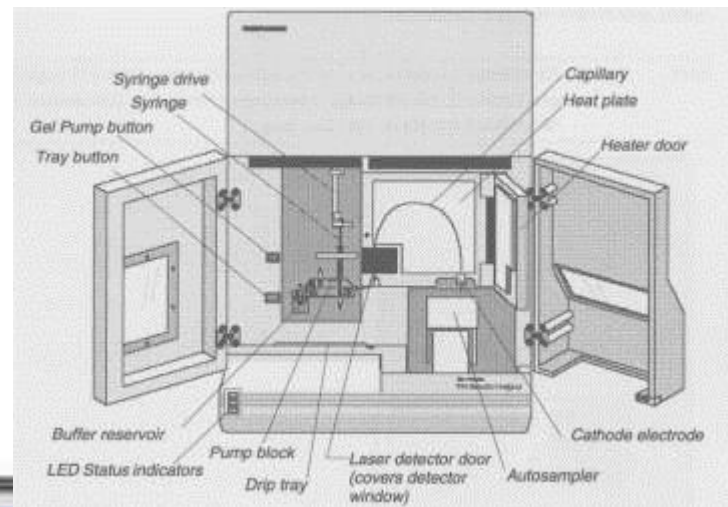
Mikrosatelitní markery - detekce

- 1) Elektroforetická separace fragmentů (odlišení cca 8-10 bp)
2-3% NuSieve agarózový gel, 10-12% PAGE

----- CACACACACACA ----- A
----- CACACA ----- B
----- CACACACACACACA ---- C

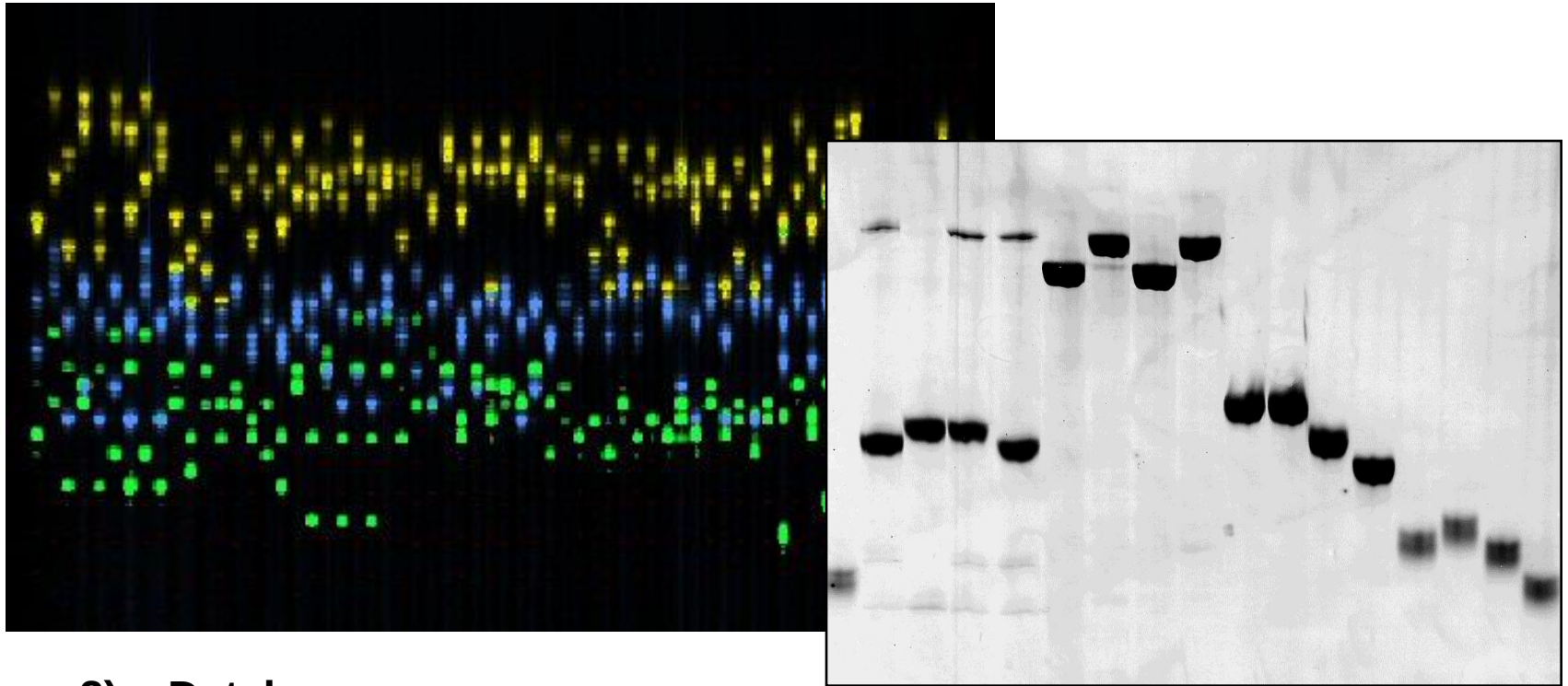


Mikrosatelitní markery automatická fragmentační analýza



Mikrosatelitní markery - detekce

- 1) Elektroforetická separace fragmentů (ideálně 1- 2 bp)
sekvenační denaturační Urea-akrylamidový gel (6% PAGE)

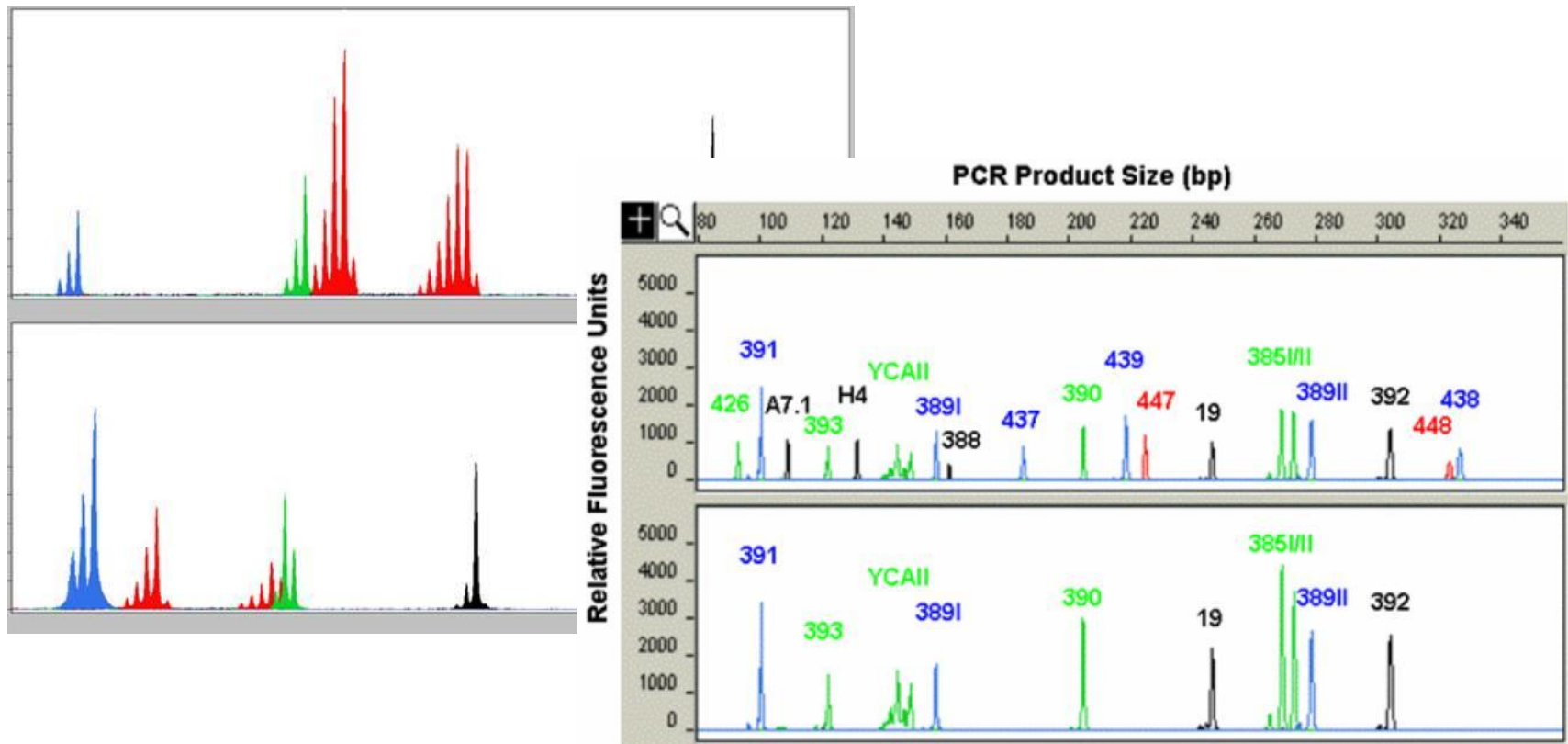


- 2) Detekce
 - barvení gelu (EtBr/UV nebo Ag/ VIS)
 - fluorescenční značení primerů - automatizace

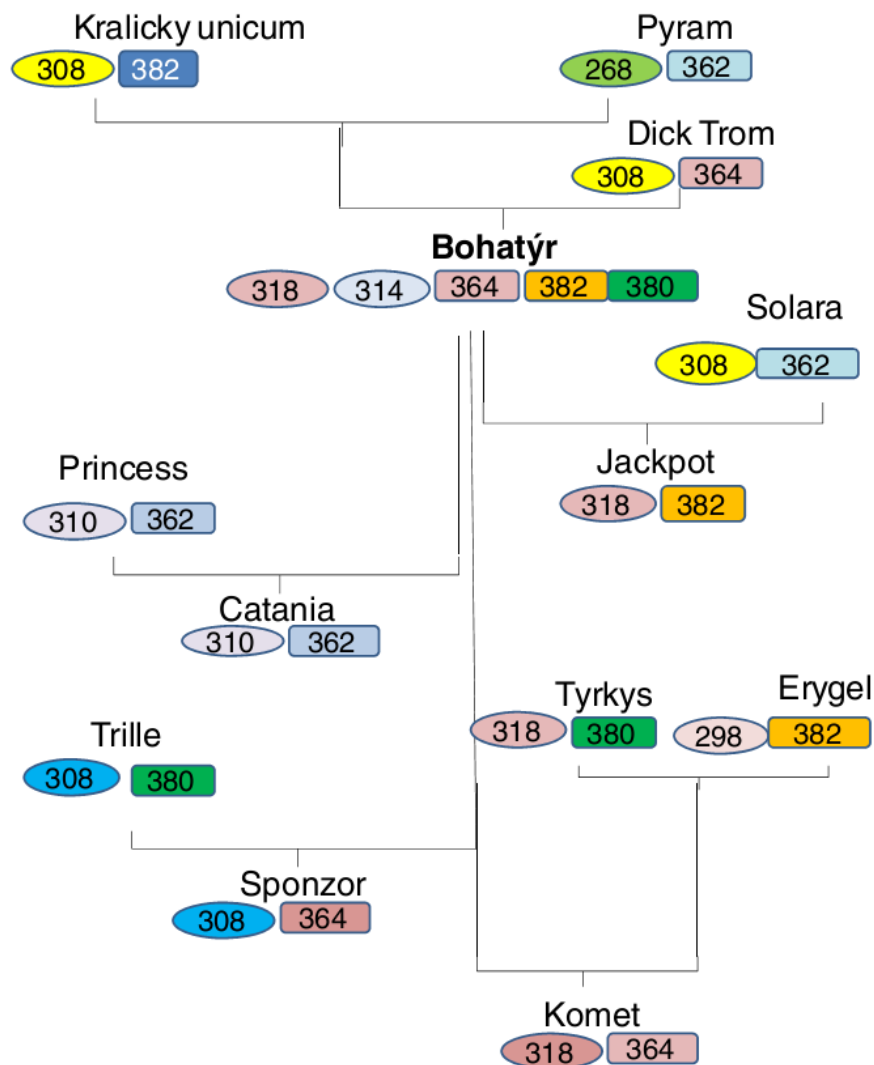
Mikrosatelitní markery - odečet velikosti

1) Elektroforetická separace fragmentů

Fluorescenčně značené fragmenty - odečet laserem, přesný standard



Mikrosatelity a stabilita



C

A9 locus		Pedigree set							Flanking region						
Accession-genotype	TC	AC	GCGC	AC	AT	AC	Repeat length	Total lenght	133	150	154	177	204	251	
Kralicky-unicum	13	19	1	5	1	6	90	382	A	A	C	C	C	A	
Pyram	16	11	0	0	1	8	70	362	A	A	C	C	C	A	
Dick_Trom	17	11	0	0	1	8	72	364	A	A	C	C	C	A	
Solara	16	12	0	0	1	7	70	362*	G	C	C	C	C	A	
Jackpot	13	18	1	1	1	6	92	382	A	A	C	C	C	A	
Princess	16	11	0	0	1	8	70	362	A	A	C	C	C	A	
Catania	17	10	0	0	1	7	70	362	A	A	C	C	C	A	
Trille	13	17	1	5	1	5	88	380	A	A	C	C	C	G	
Sponzor	17	11	0	0	1	8	72	364	G	C	A	T	T	A	
Erygel	13	19	1	5	1	6	90	382	A	A	C	C	C	A	
Tyrkys	13	17	1	5	1	5	88	380	A	A	C	C	C	G	
Komet	16	11	0	0	1	5	72	364	A	A	C	C	C	A	
Rondo	17	11	0	0	1	8	72	364	A	A	C	C	C	A	
Saxa	13	18	1	5	1	5	90	382	A	A	C	C	C	A	
Unica	13	18	1	5	1	6	90	380	A	A	C	C	C	A	

Fig. 1 Summary of pedigree related accessions with indicated sizes of AD270 (circles) and A9 loci (squares)

Mikrosatelitní markery

Výhody:

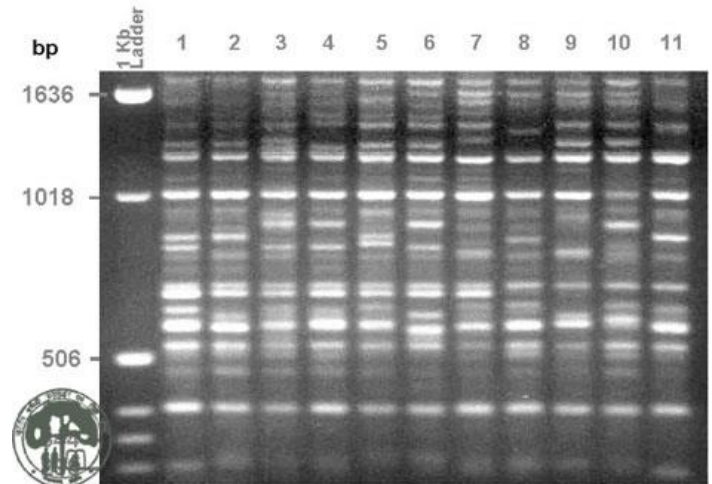
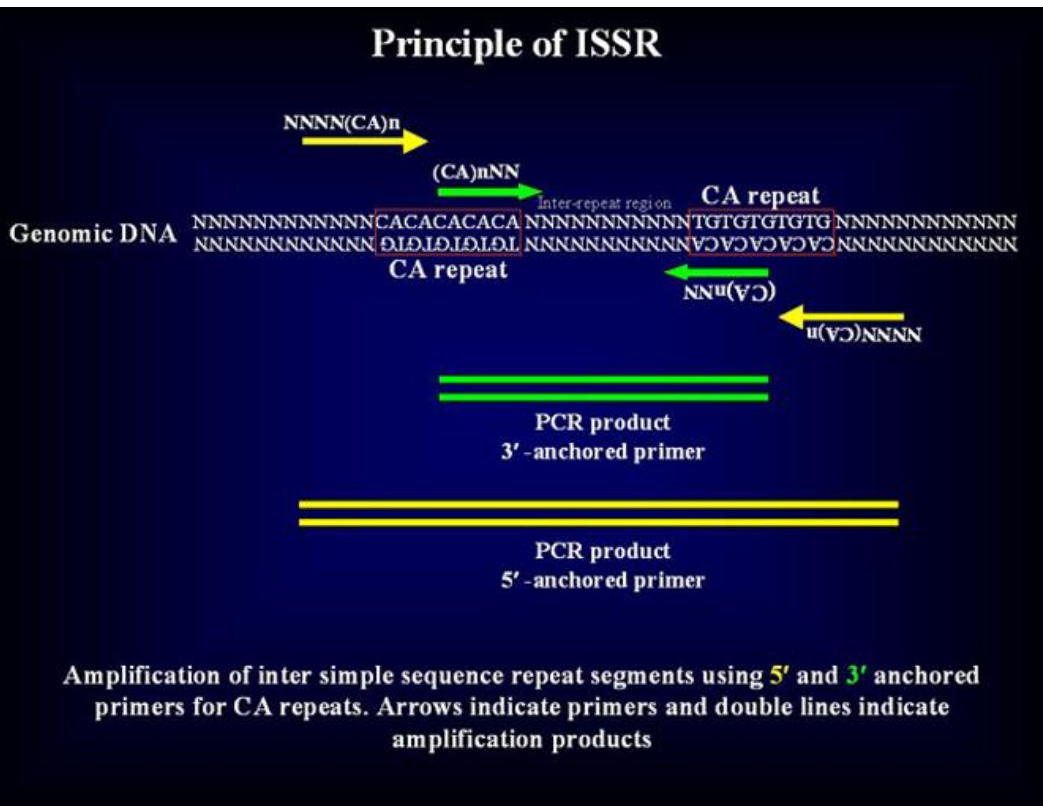
- * vysoký polymorfismus – multi-alelismus
- * ko-dominance (možnost detekce heterozygotního stavu, směsi)
- * jednoduchost aplikace metodiky
- * Mapovatelnost – znalost pozice (lokusu)

Nevýhody:

- * **homoplasie**, vysoká mutační rychlost (10^{-3} – 10^{-4} na lokus a generaci)
- * přenos mezi druhy není zcela možný/ aplikovatelný
- * pro *de novo* vyvinutí - potřeba sekvenování, obohacené knihovny
- * přesnost odečtu alel

C3_G1	...(CT) ₂₃	ATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT - - - CTCTCT...	296
C3_G10	...(CT) ₂₃	ATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTCTCTCT...	300

Inter Simple Sequence Repeats (ISSR)



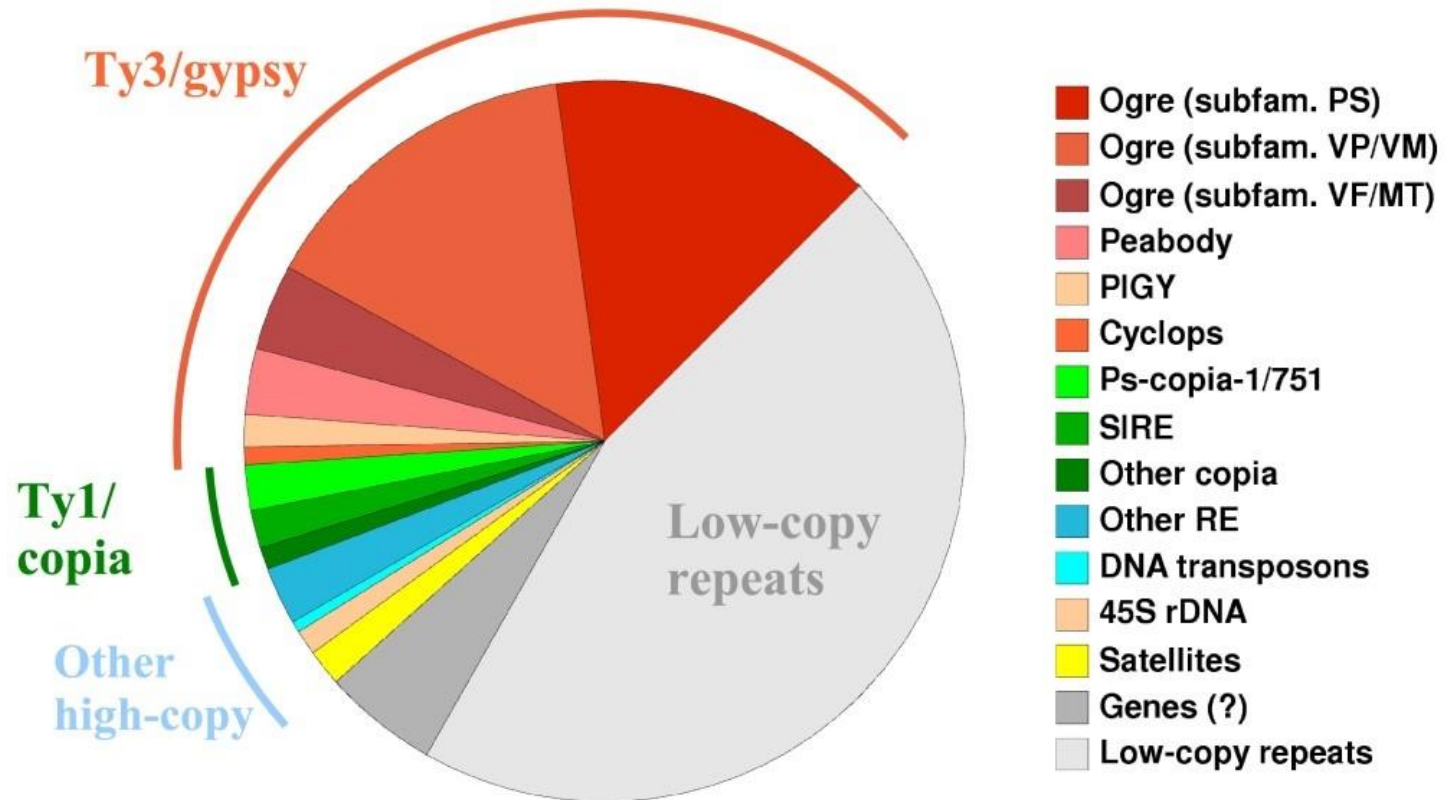
Výhody:

- * vysoký polymorfismus
- * není třeba spec. primerů
- * jednoduchost aplikace metodiky

Nevýhody:

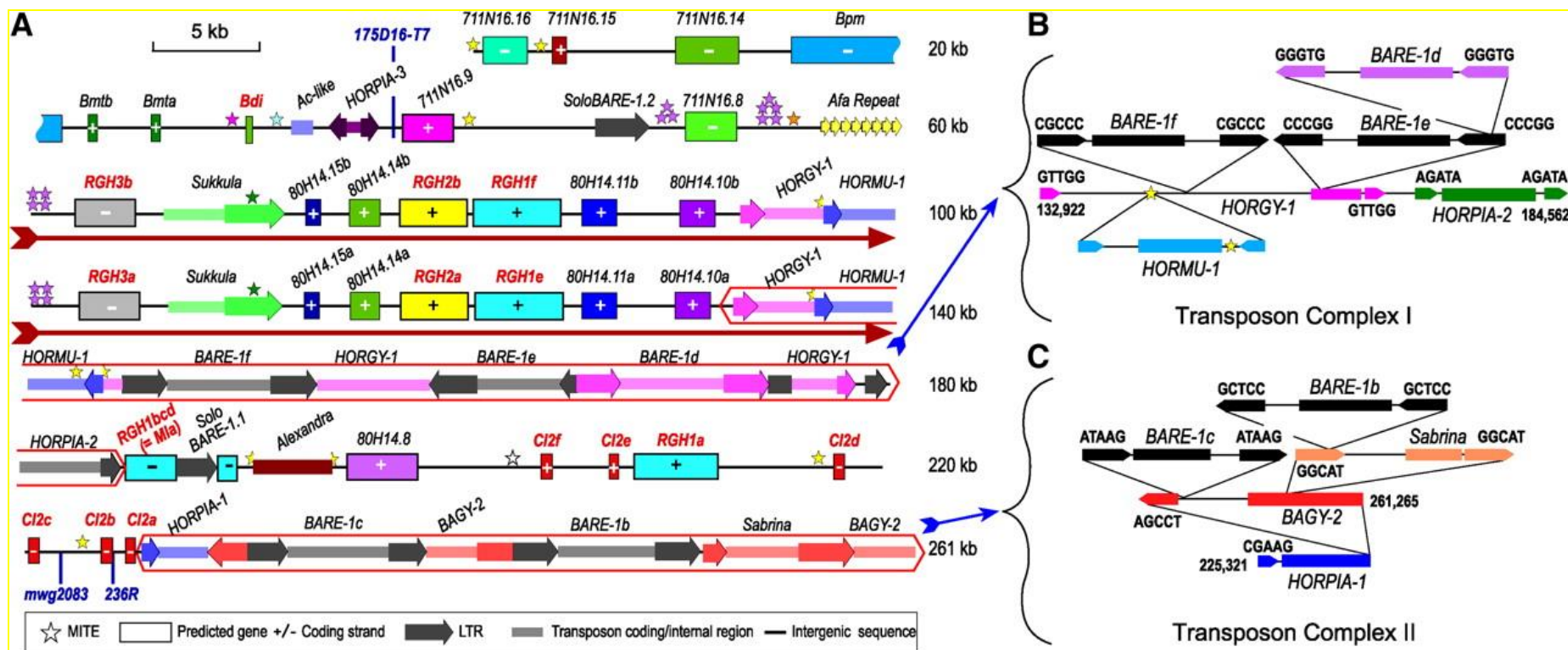
- * malá reprodukovatelnost
- * dominantní marker

Složení eukaryotického genomu (hrách) množství repetitivních sekvencí, méně genů



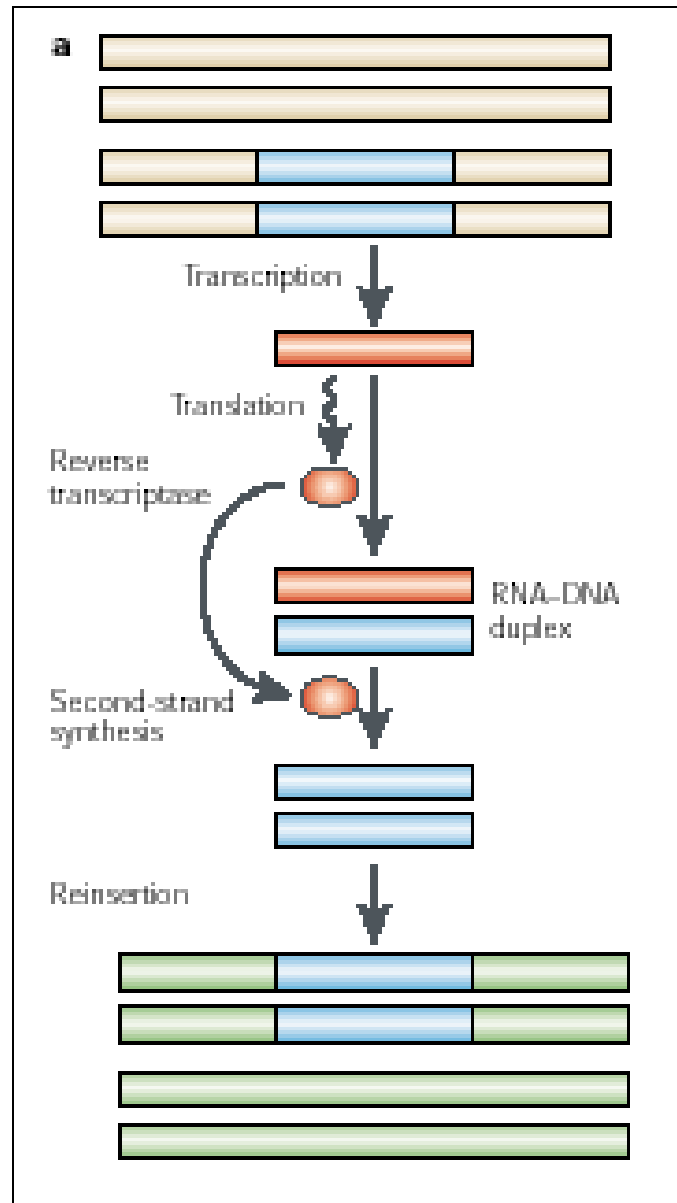
60 - 80 % repetitivní DNA, 20 – 40% geny + regulační sekvence

Repetitivní sekvence jsou často ve shlucích

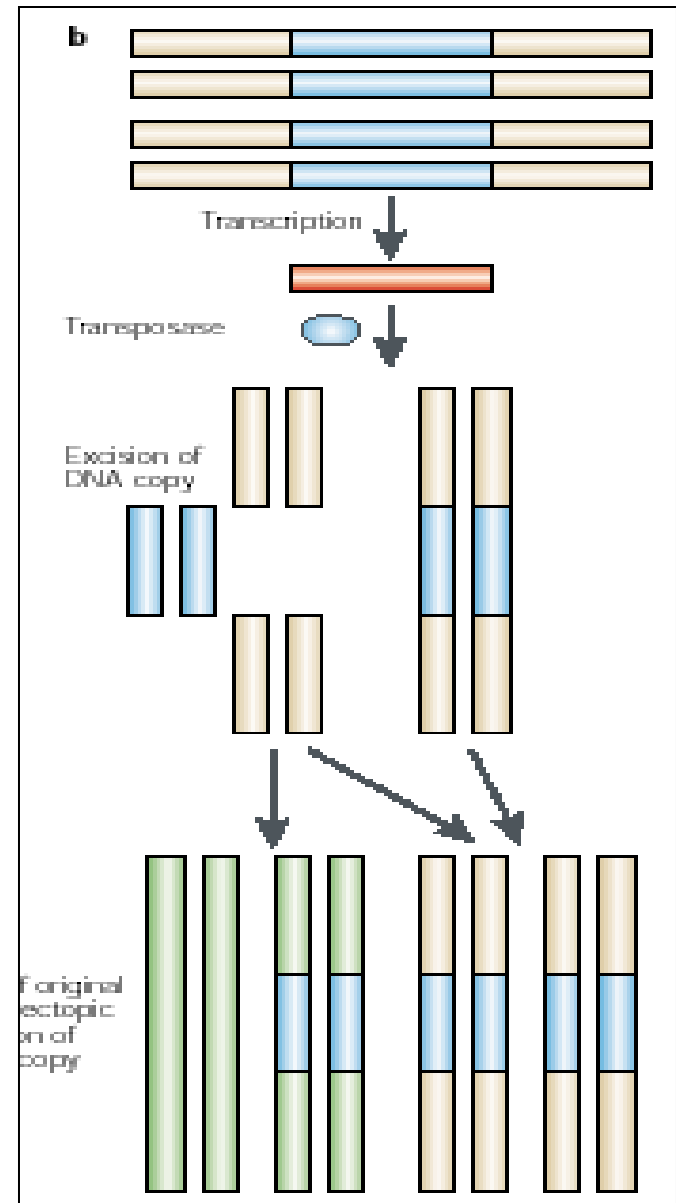


Mla lokus rezistence ječmene k padlí - Weis et al. Plant Cell 2002

Retrotransposony



Transposony



Rozdíl mezi transposony a retrotransposony

Transposony

šíření mechanismem

cut and paste

inzerce v daném místě

nestabilní/mobilní

četnost v genomu

X 100

využití jako markeru

nevhodné

Retrotransposony

copy and paste

stabilní

x 1000/100 000

vhodné / fylogeneticky informativní

LTR retrotransposons

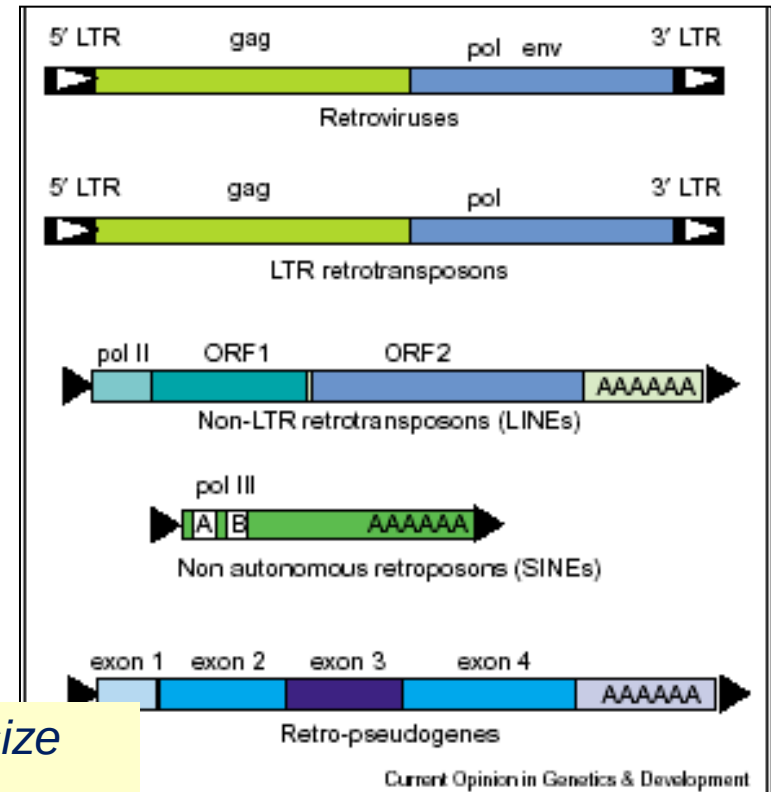
Ty1-copia type

Ty3-gypsy type

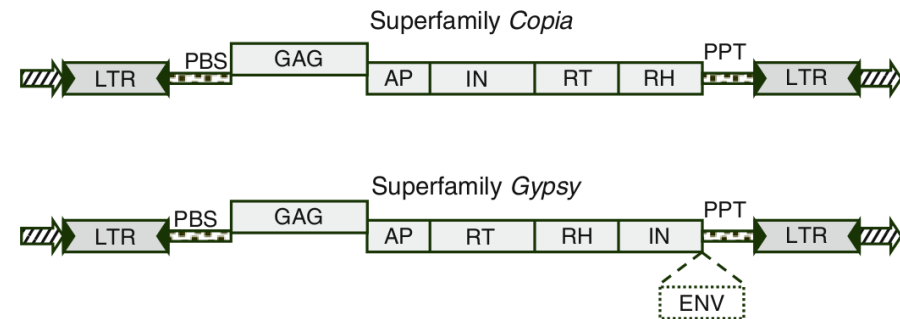
non-LTR retrotransposons

autonomous *LINE*

non-autonomous *SINE*

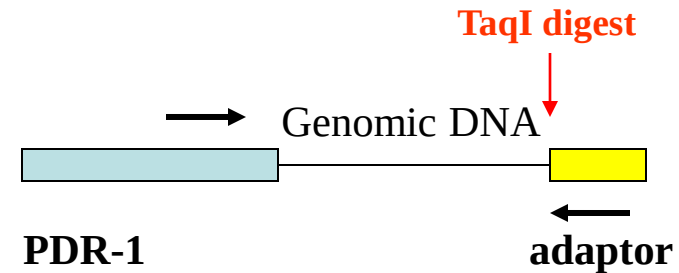
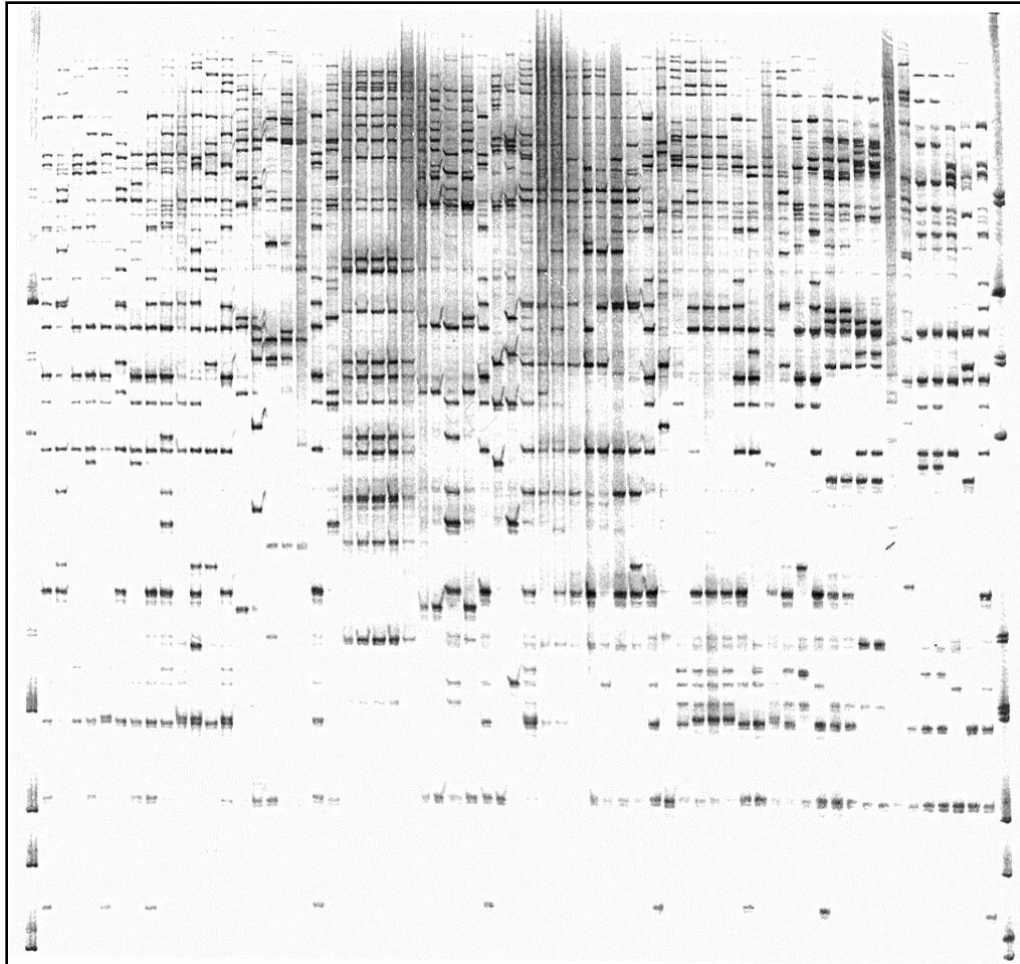


<i>E. coli</i>	4.403 genes	4.6 Mb size
<i>Yeast</i>	5.538	13
<i>Fruitfly</i>	13.350	120
<i>human</i>	30.000	3.500
<i>pea</i>	30.000	4.000
<i>Arabidopsis</i>	25.000	100
<i>Medicago</i>	30.000	550
<i>Lotus</i>	30.000	400
<i>maize</i>	25.000	4.500
<i>wheat</i>	25.000	17.00



Aplikace retrotransposonů jako markerů

Příklad SSAP analýzy s PDR-1 elementem



- genomic DNA restriction
- adaptor ligation
- PCR amplification
- gel separation – analysis

Nevýhody:

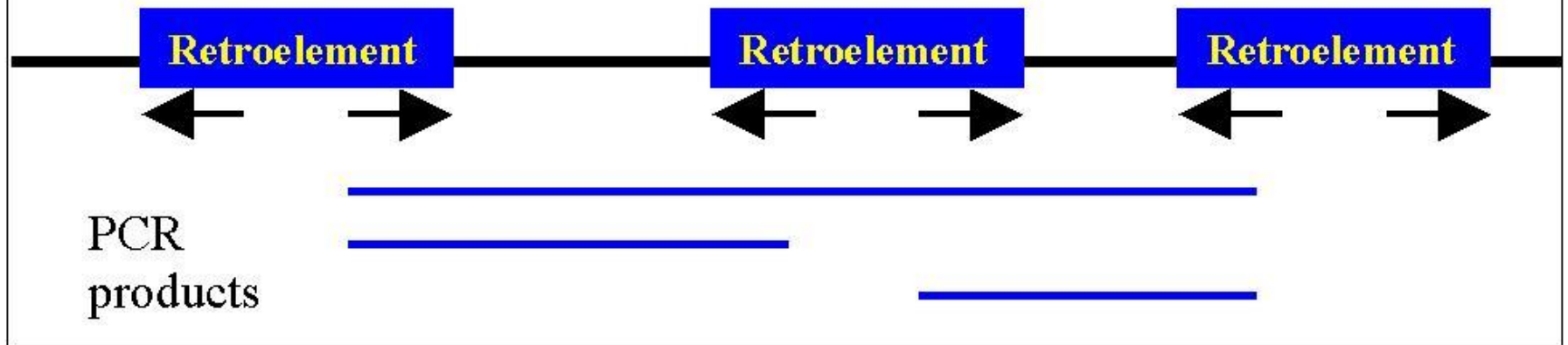
jako AFLP – technicky náročné
na přesnost provedení
mnoho kroků – možnost chyby

Inter-Retrotransposone Amplified Polymorphism (IRAP)

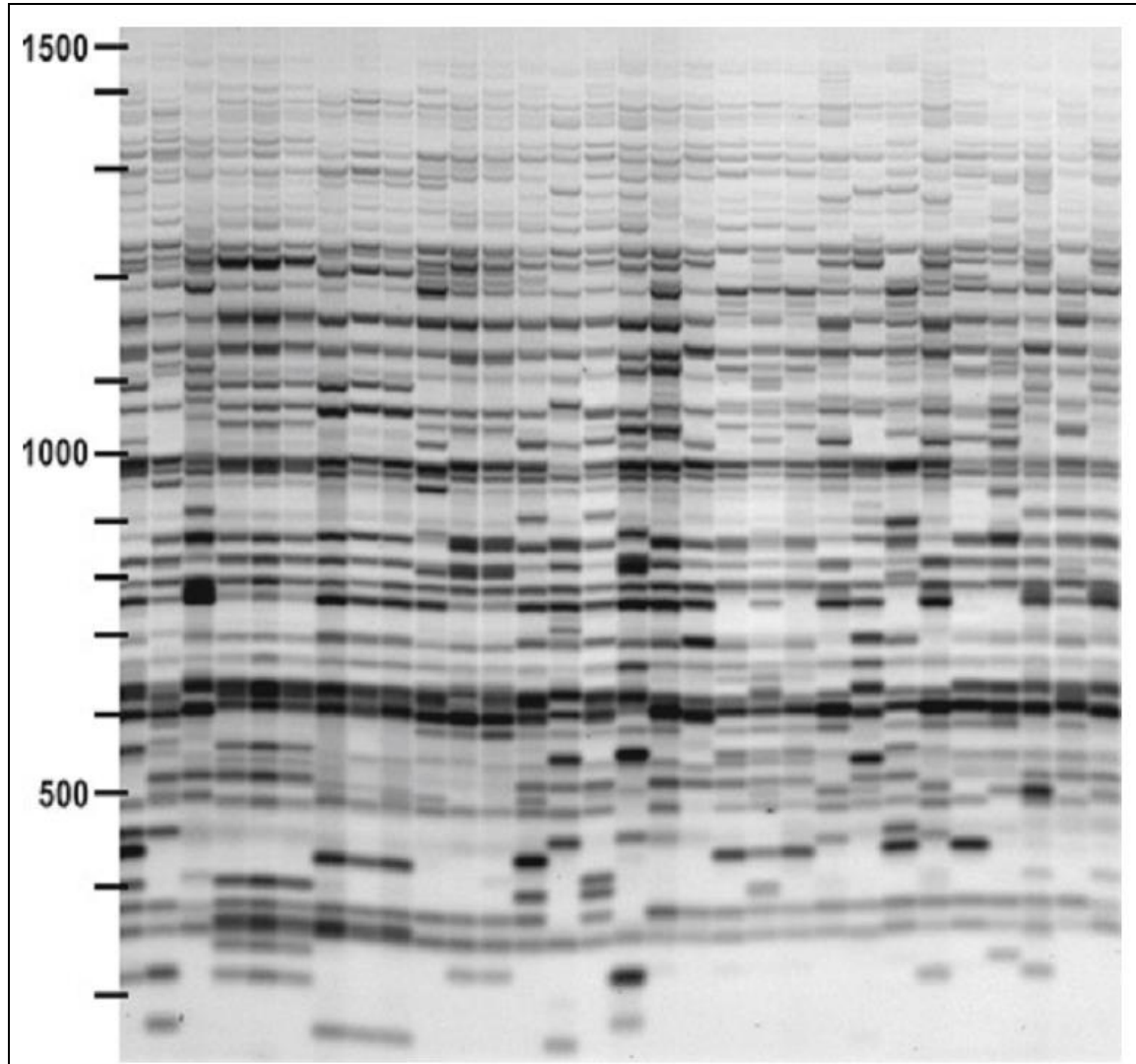
Long Terminal Repeat



Principle of IRAP method



Inter-Retrotransposone Amplified Polymorphism (IRAP)



Vliv výběru retrotransposonu na IRAP

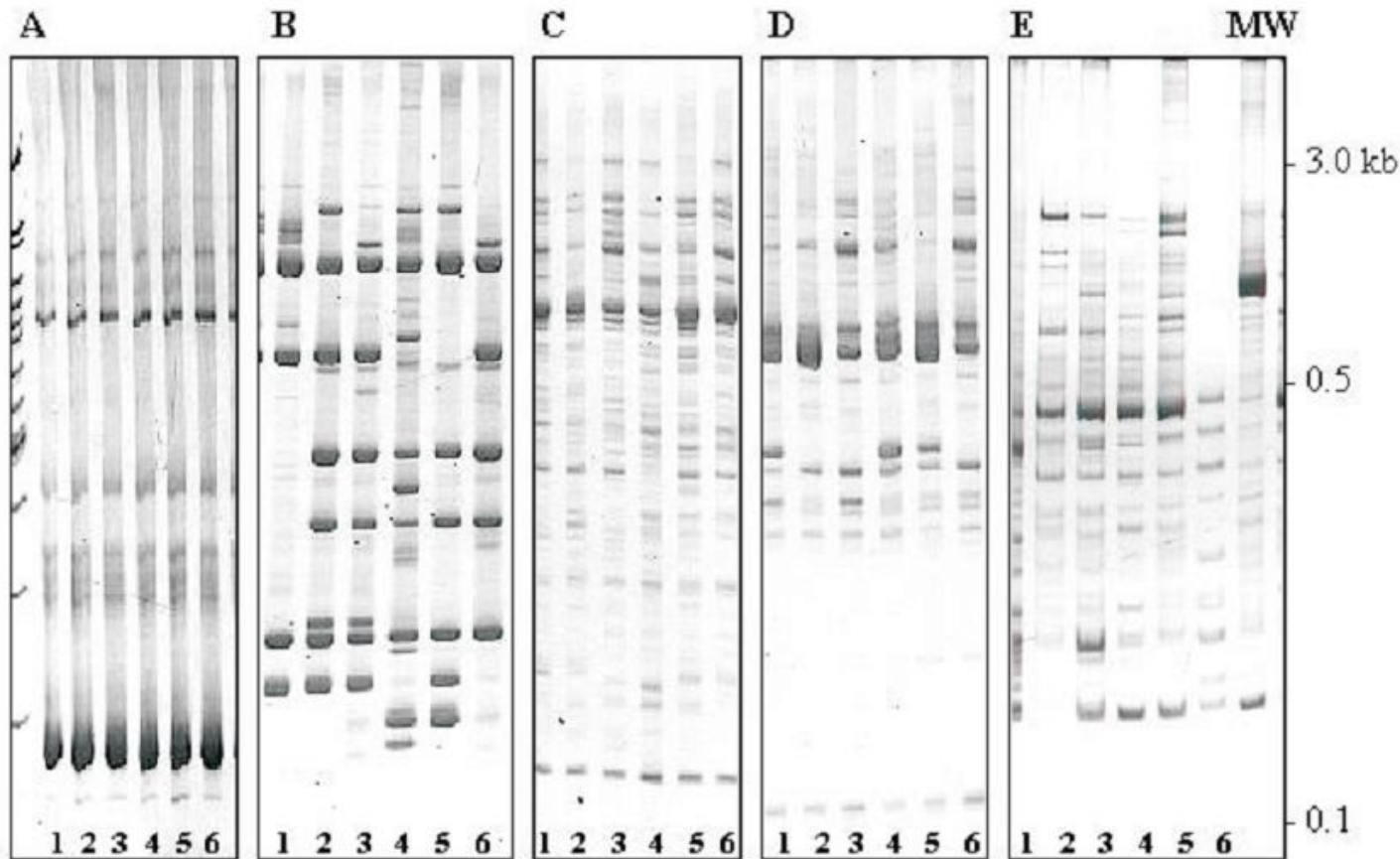


Figure 1. IRAP analysis with various pea *Ty3-gypsy* type retrotransposons—*Pis-TRA* (A), *Cyclop* (B), *MITE* (C), *Pigy* (D) and *Ogre* (E) – performed with F and R primers. Genomic DNA (50 ng) from cvs. Alan (1), Bohatýr (2), Merkur (3), Menhir (4), Komet (5), or Adept (6) was used per PCR reaction. Negative image of EtBr stained and UV visualized on 20-cm 8% PAGE gel is shown. MW = 100-bp DNA molecular weight marker (Fermentas).

Vliv metody izolace DNA a Taq polymerázy na IRAP fingerprinting

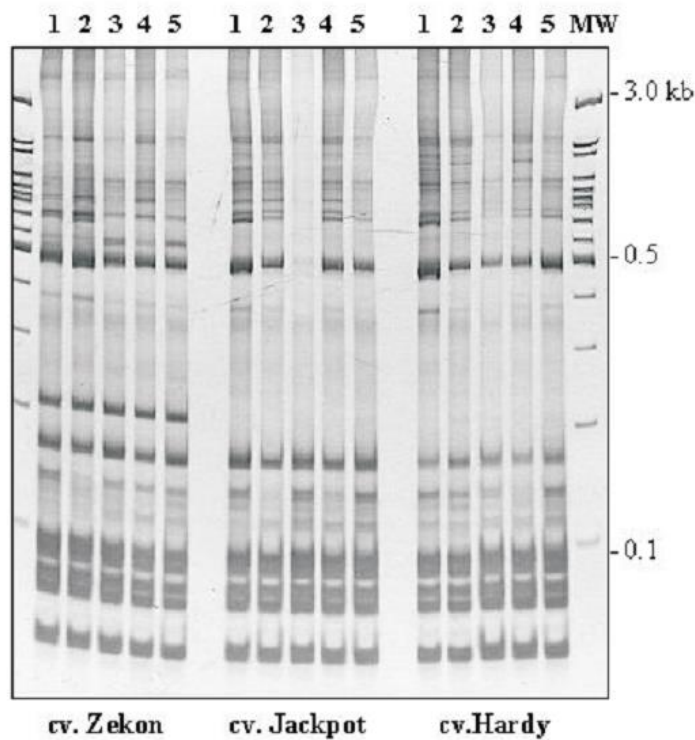


Figure 3. Effect of DNA isolation method – Qiagen DNeasy kit (1), Invitex (2), salt-acidic (3) and CTAB protocol (4, 5) – on performance of IRAP *Ogre* F and R primers. In each case, 50 ng of DNA was taken in 2 μ L for the PCR reaction. MW = 100-bp DNA molecular weight marker (Fermentas).

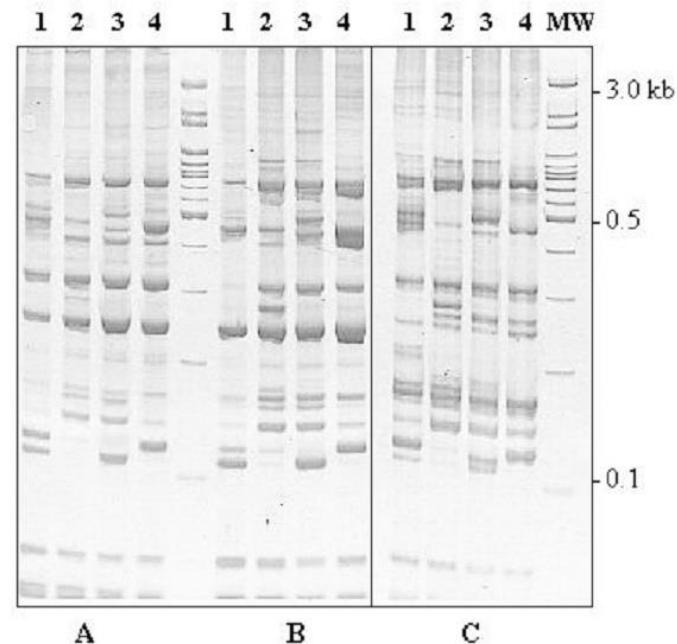


Figure 4. Effect of *Taq* polymerase source – TAKARA polymerase (A), BioTools (B) and Dynazine II, Finzymes (C) – on performance of IRAP *Cyclop* F and R primers. In each case, 50 ng of DNA and 2.5 mM of Mg ions were used for the PCR reaction with cvs. Merkur (1), Garde (2), Tempra (3) or Kamelot (4). MW = 100-bp DNA molecular weight marker (Fermentas).

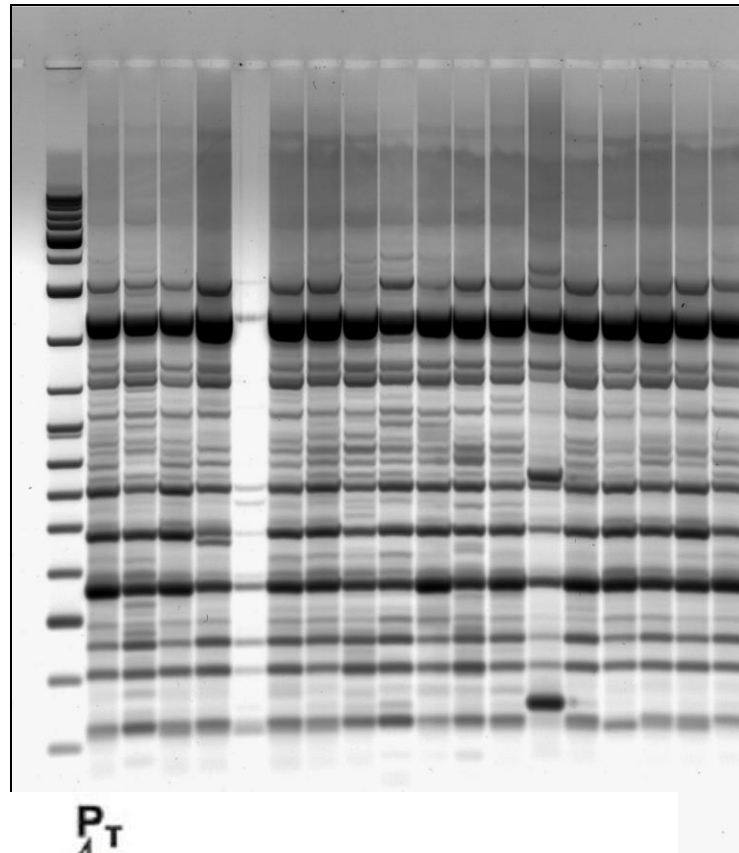
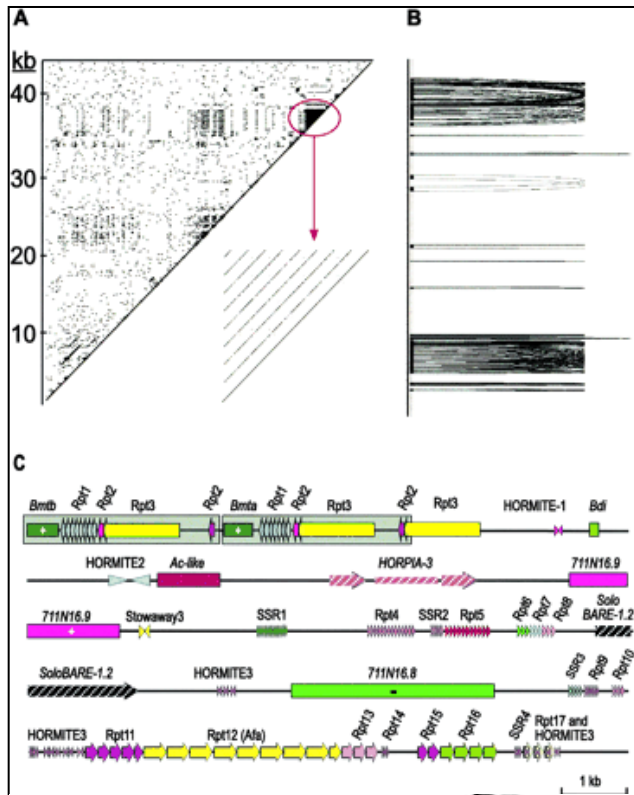
Inter-Retrotransposone Amplified Polymorphism (IRAP)

RETROTRANSPONOVÉ MARKERY										
Nr.	Odrůda	A	B	C	D	F	G	I	J	M
2	Alan	0	1	1	0	1	1	0	0	1
4	Bohatýr	0	1	1	0	1	1	0	0	1
5	Canis	0	1	1	0	1	1	0	0	1
14	Jackpot	0	1	1	0	1	1	0	0	1
8	Garde	1	1	1	0	0	0	1	1	1
10	Grana	1	1	1	0	0	0	1	1	1
12	Harnas	1	1	1	0	0	0	1	1	1
19	Madonna	1	1	1	0	0	0	1	1	1
16	Kamelot	0	1	0	0	1	1	0	0	1
13	Herold	0	1	0	0	1	1	0	0	1
18	Menhir	0	1	0	0	1	1	0	0	1
15	Janus	0	1	1	1	1	1	0	0	1
27	Romeo	0	1	1	1	1	1	0	0	1
34	Zekon	0	1	1	0	1	1	0	0	1
11	Hardy	0	1	1	0	1	1	0	0	1
25	Primus	0	1	1	0	1	1	0	0	1
23	Merkur	0	1	1	0	1	1	0	0	1
29	Sonet	0	1	1	1	1	0	0	0	1
30	Sponsor	0	1	1	1	1	0	0	0	1
33	Tyrkys	1	1	1	1	1	0	0	0	1
35	Olivín	1	1	1	1	1	0	0	0	1
32	Terno	0	1	1	0	1	0	1	0	1
7	Catania	0	1	1	0	1	1	1	0	0
3	Baryton	0	1	1	0	1	1	1	0	1
6	Carrera	0	1	0	1	1	1	0	0	1
9	Gotik	0	1	1	0	0	0	1	0	1
21	Komet	1	1	1	0	1	1	0	1	0
24	Power	1	1	1	0	1	0	0	1	1
28	Smaragd	1	1	1	1	1	0	0	1	1
31	Tempra	1	1	1	1	1	0	1	1	0
17	Lantra	1	1	1	0	1	0	1	1	0
1	Adept	1	1	1	0	1	1	1	1	0

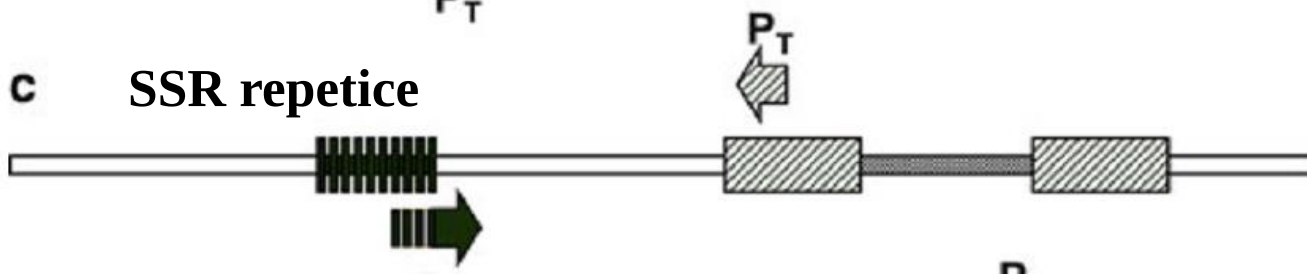
Odrůdy hrachu

Cyclop a Ogre LTR
fingerprinting

REMAP - Retrotransposon x Mikrosatelitní polymorfismus

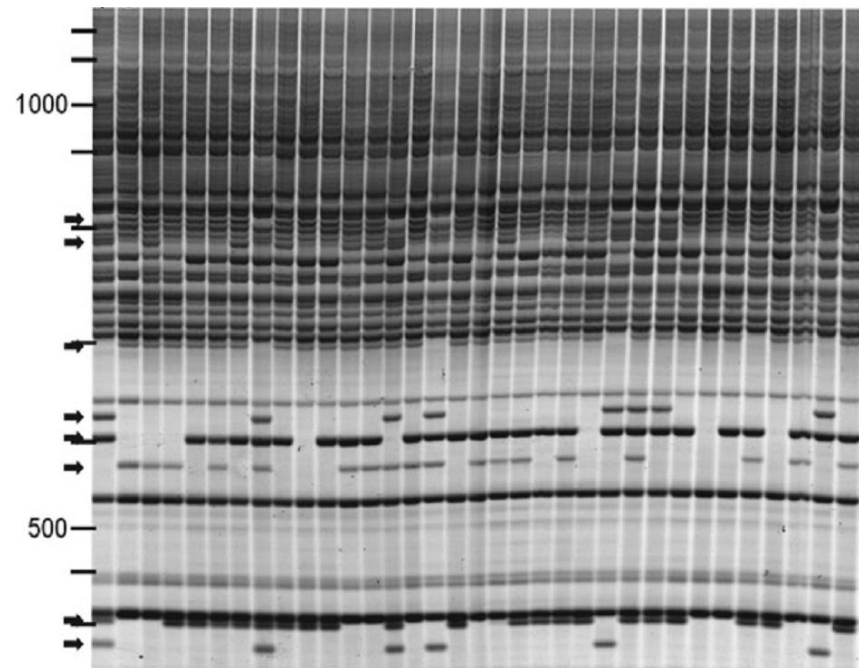
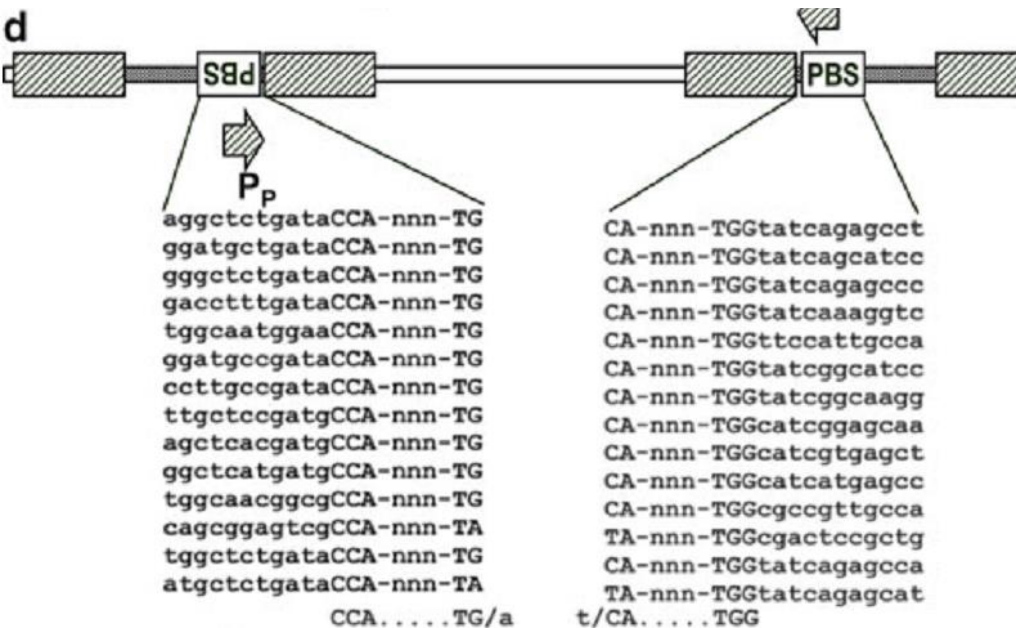


C SSR repete



iPBS: a universal method for DNA fingerprinting and retrotransposon isolation

Ruslan Kalendar · Kristiina Antonius · Petr Smýkal · Alan H. Schulman



Superfamily *Copia*



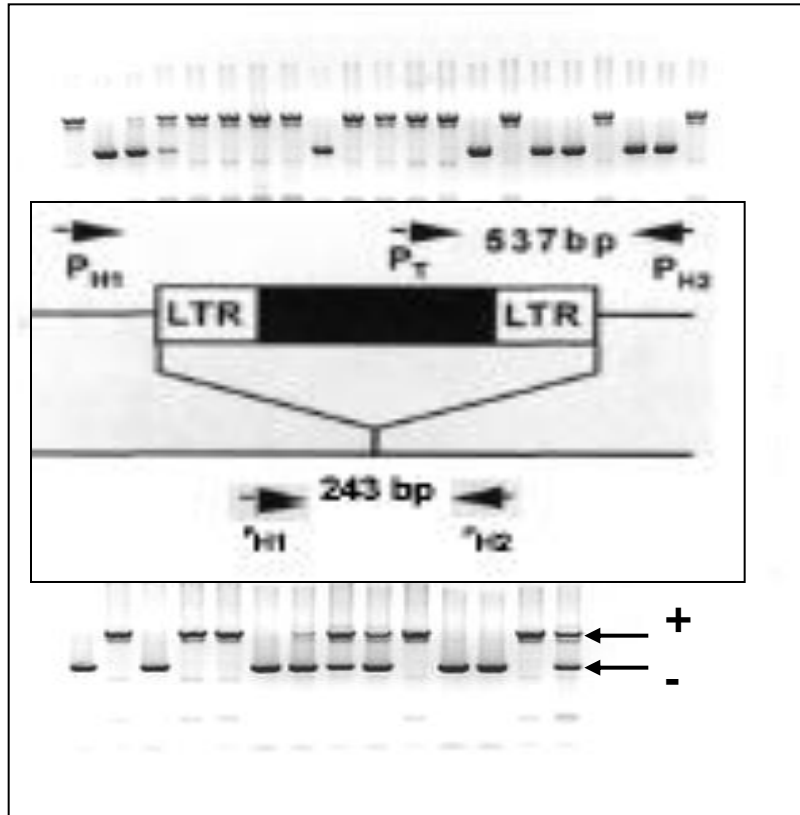
Výhody použití IRAP / REMAP / iPBS metod

- jednoduchost
- není třeba žádných předchozích kroků, jen izolované DNA - PCR
- univerzálnost (iPBS)
- vhodné pro studium vnitrodruhového polymorfismu/diverzity

Nevýhody:

- menší reprodukovatelnost - jako u jiných fingerprintových metod
- potřeba použití stejné Taq polymerázy
- odečitelnost délkového polymorfismu fragmentů
- problematické srovnání mezi laboratořemi
- homoplasie fragmentů mezi odlišnými druhy

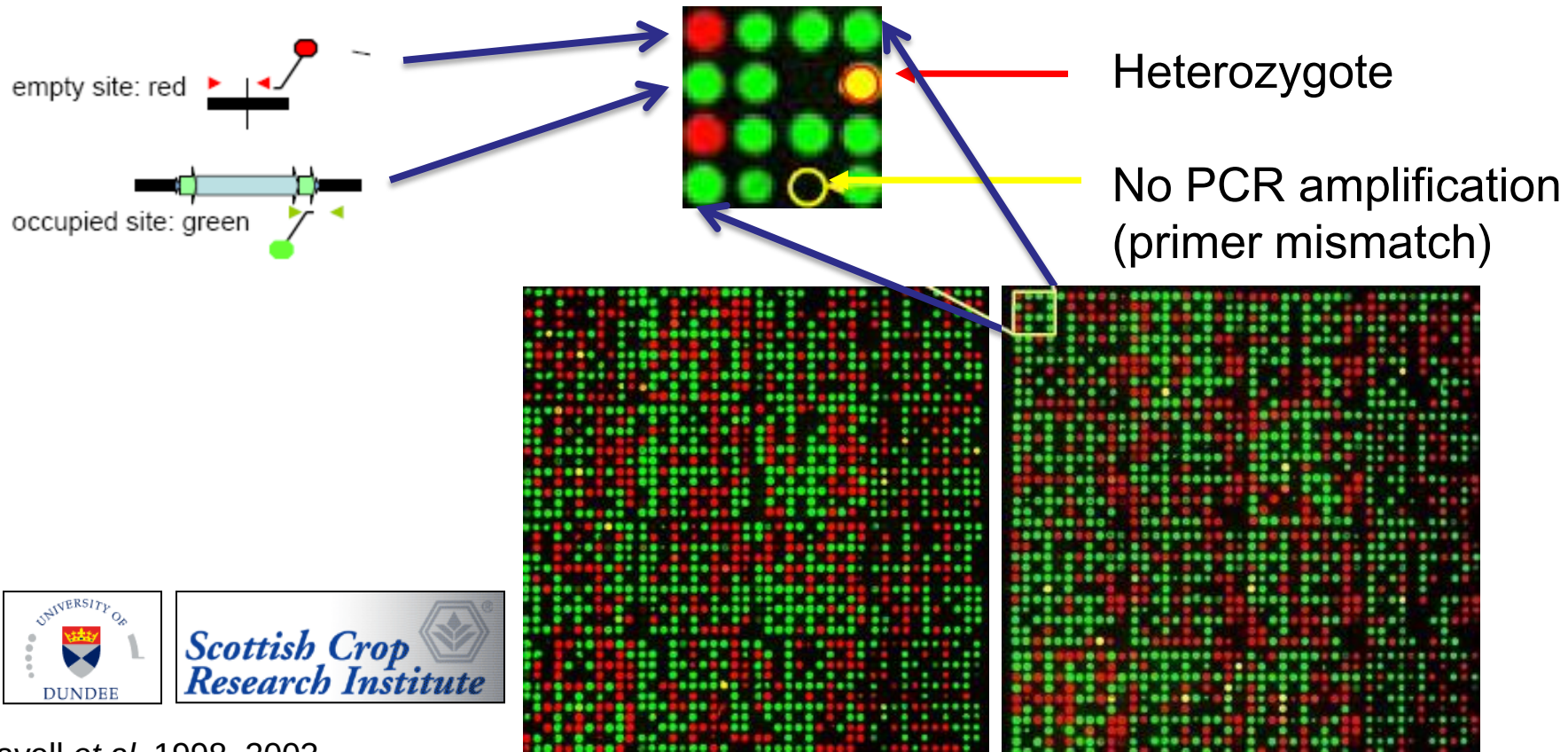
Retrotransposon-based insertion polymorphism (RBIP)



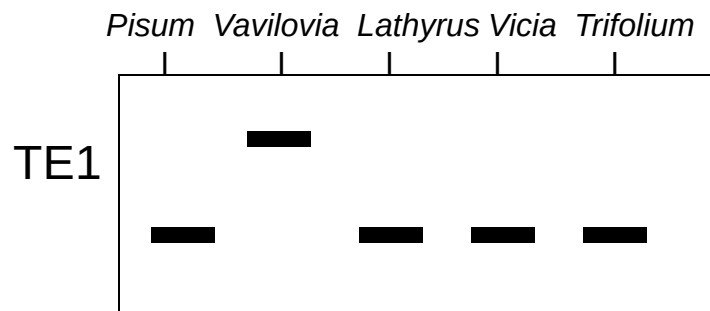
Nr.	Name	Year	RBIP-1	RBIP-2	RBIP-3	RBIP-4	RBIP-5	RBIP-6	RBIP-7	RBIP-8	RBIP-9	RBIP-10	RBIP-11
00581	HS 30-18 nsl.	1977	+	0	+/+	+	+	+	+	-	0	+	-
00582	LU-DIK ZL.	1979	-	+	+	+	+	+	0	+	0	-	-
00583	LU-DIK ZEL	1979	0	+	+	-	-	+	+	-	0	-	-
00584	Tolar	1980	+	+	+	+	+	+	+	-	0	+	-
00585	CH-719	1979	+	-	+	+	+	-	-	+	+	+	-
00586	Odeon	1980	-	0	+	+	+	+	+	-	0	+	-
00603	Tyrkys	1983	-	0	+	+	+	+	-	+	+	+	-
00604	HM 1926 nsl.	1983	-	0	+	-	-	+	0	-	0	+	+/+
00605	HC 52 nsl.	1983	+	+	+	+	+	-	+	-	0	+	-
00606	LU FK nsl.	1983	+	-	+	-	-	+	-	+	+	+	-
00607	HS 16502 nsl.	1983	-	0	+	-	-	-	-	+	0	+	-
00624	Senator	1971	+	+	+	+	+	+	-	+	0	+	-
00642	Junak	1985	+	+	+	-	-	+	-	+	0	-	-
00644	Emerald	1984	-	+	+	-	-	+	-	+	0	-	-

High throughput analysis

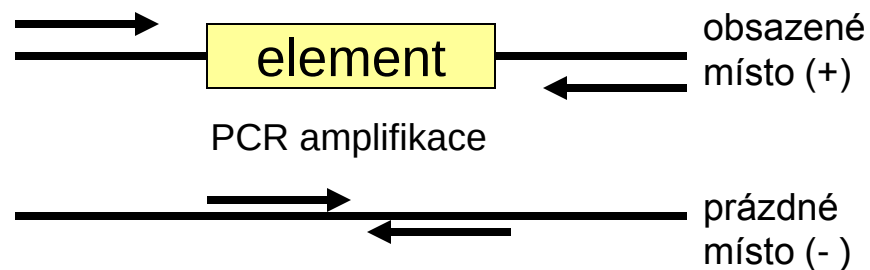
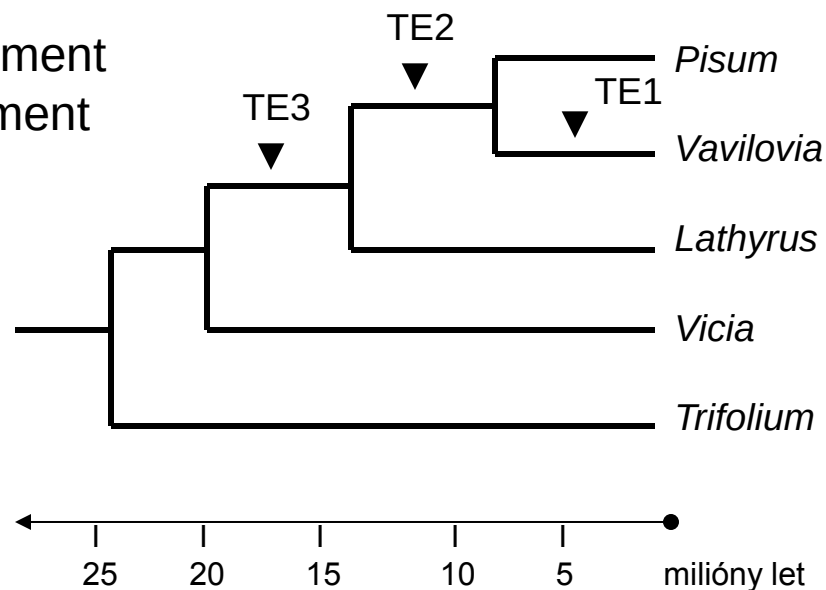
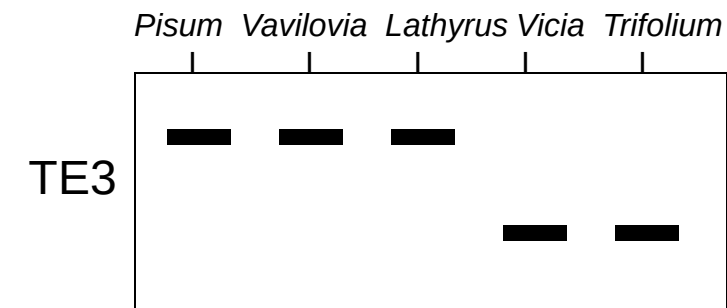
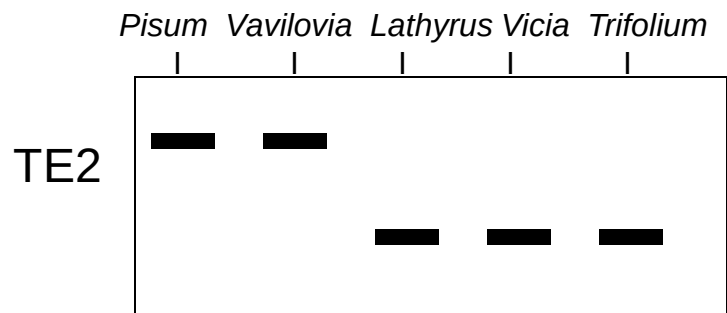
Retrotransposon-based insertion polymorphism (RBIP)



Flavell *et al.* 1998, 2003
Jing *et al.* 2005

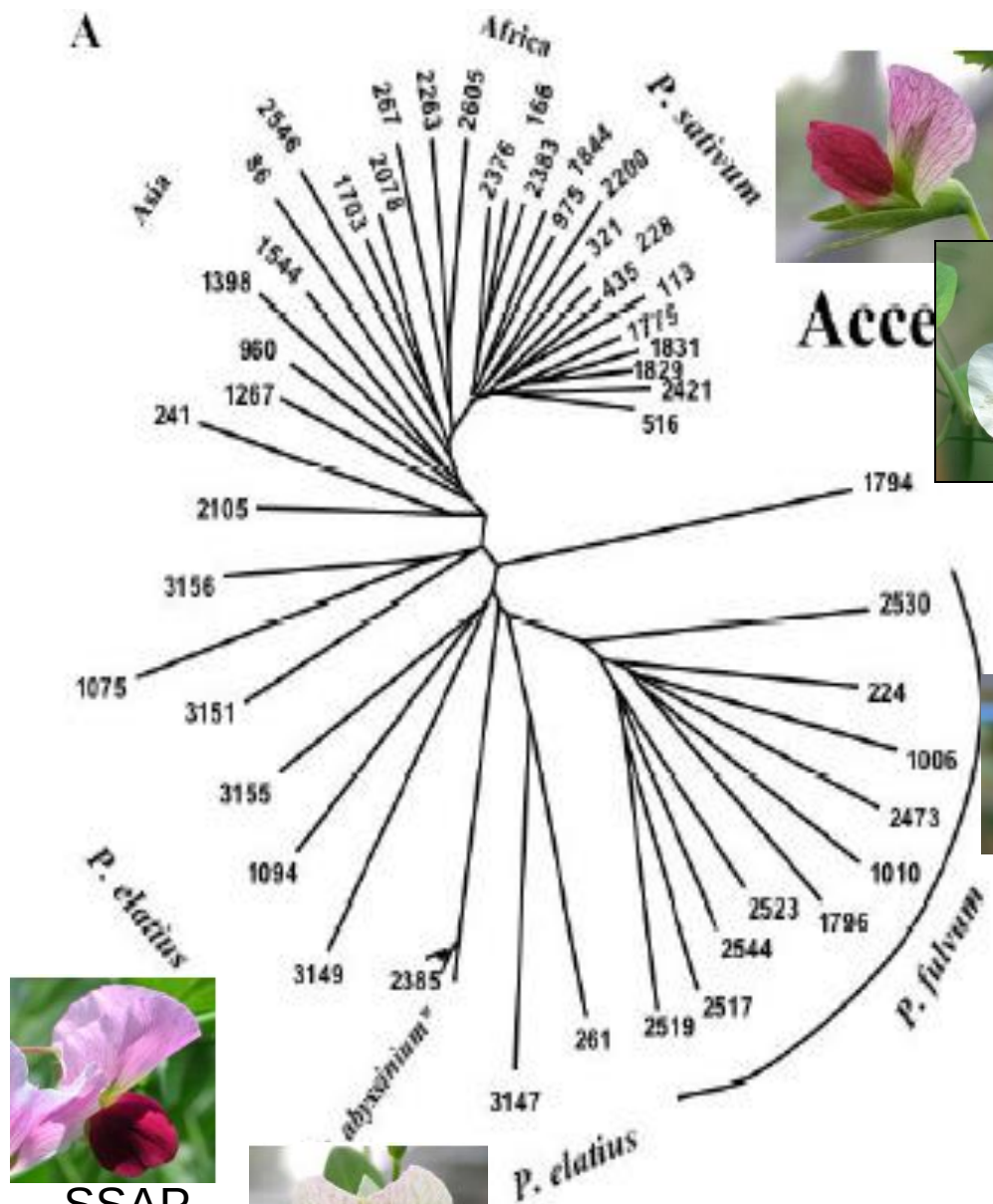


+ element
- element

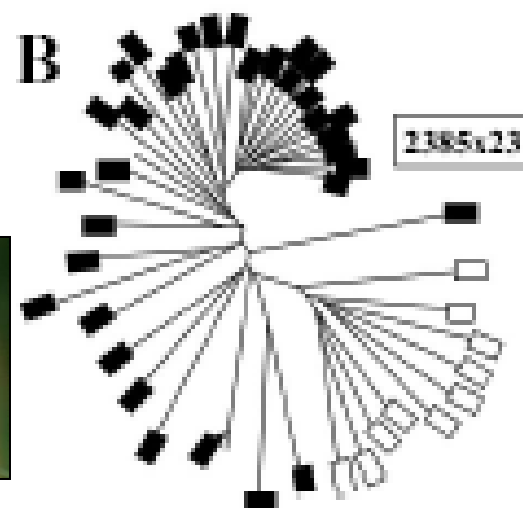


Inzerce retrotransposonu jako fylogenetická značka

A

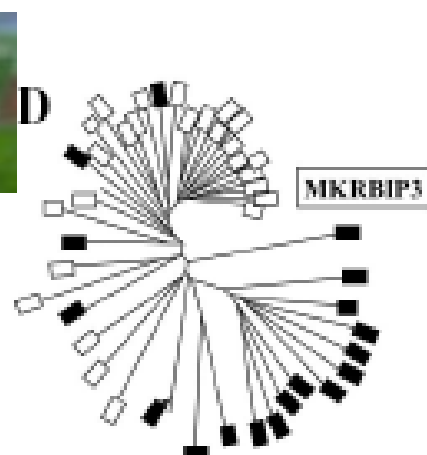


B



Hrách

D



SSAP



Retrotransposonové markery

Výhody:

- * Abundantní složka eukaryotických genomů
- * Dominantní (IRAP/SSAP) / ko-dominantní markery (RBIP)
- * Rychlá aplikace na nové druhy

Nevýhody:

- Dominantní markery (IRAP/SSAP) nepřesné pro fylogenezi –
- vhodné pro studium vnitrodruhové diverzity
- Ko-dominantní (RBIP) – přesné, ale náročné na vývoj

Genově specifické markery

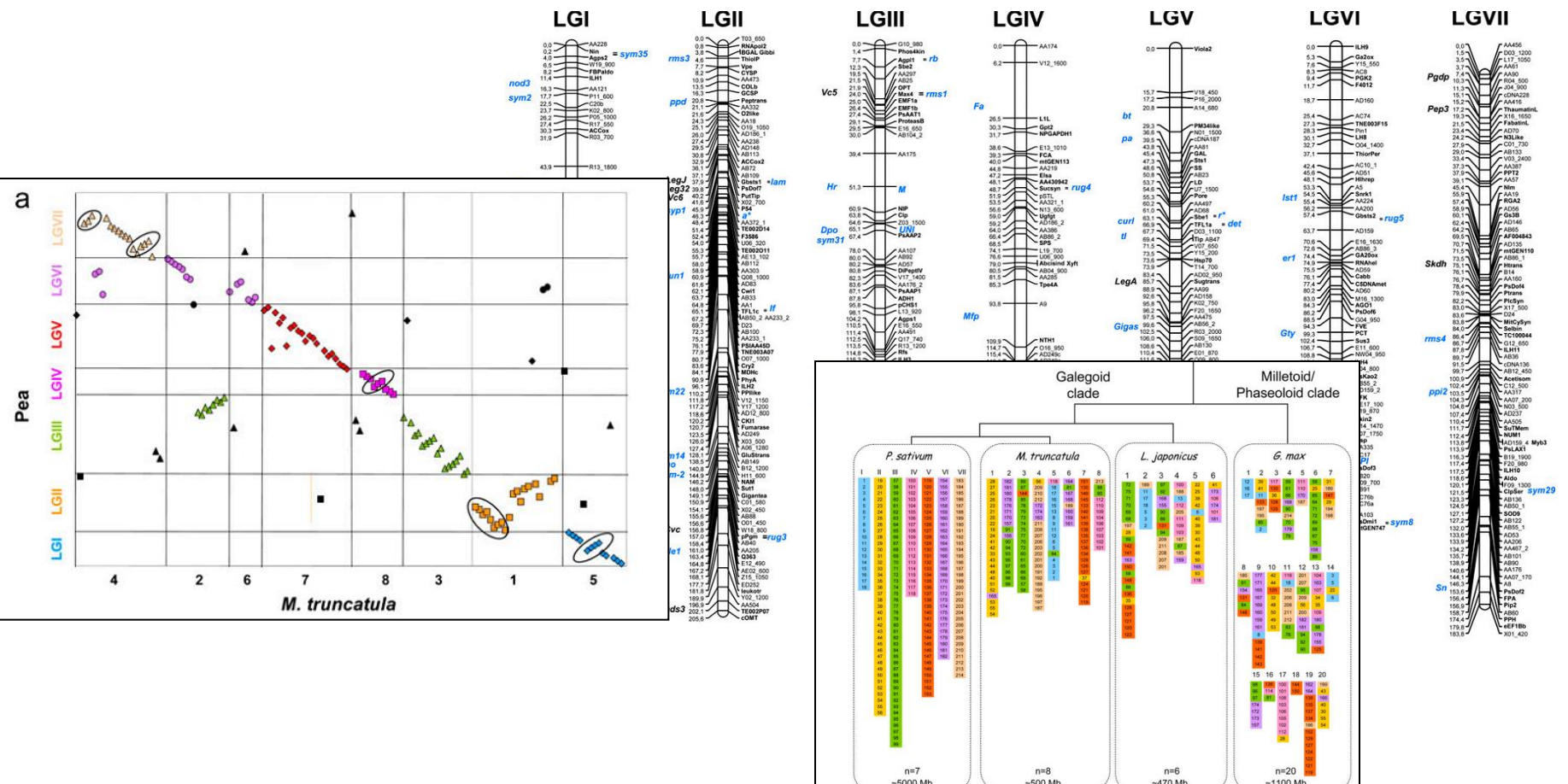
Translational Genomics in Legumes Allowed Placing *In Silico* 5460 Unigenes on the Pea Functional Map and Identified Candidate Genes in *Pisum sativum* L.

Amandine Bordat,^{*1,2} Vincent Savoie,^{*1} Marie Nicolas,^{*} Jérôme Salse,[†] Aurélie Chauveau,^{*3}

Michael Bourgeois,^{*4} Jean Potier,^{*} Hervé Houtin,^{*} Céline Rond,^{*} Florent Murat,[†] Pascal Marget,^{*}

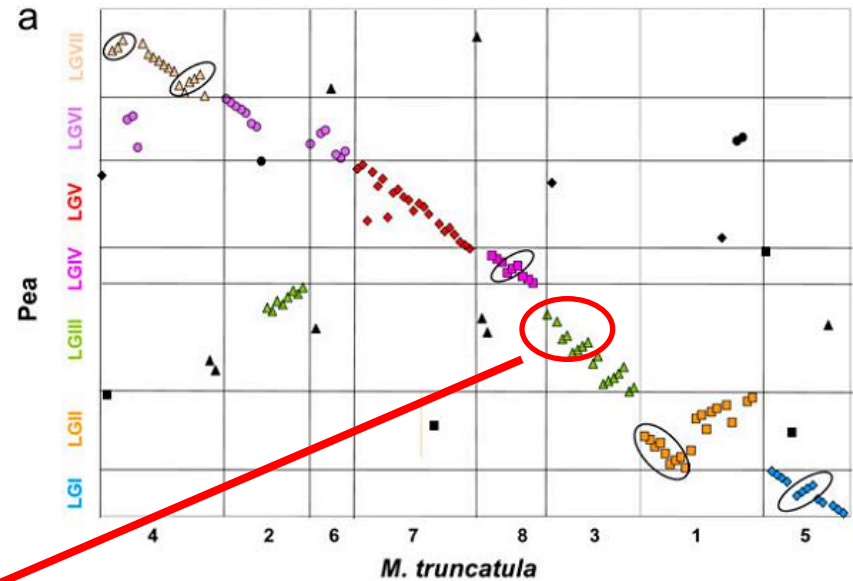
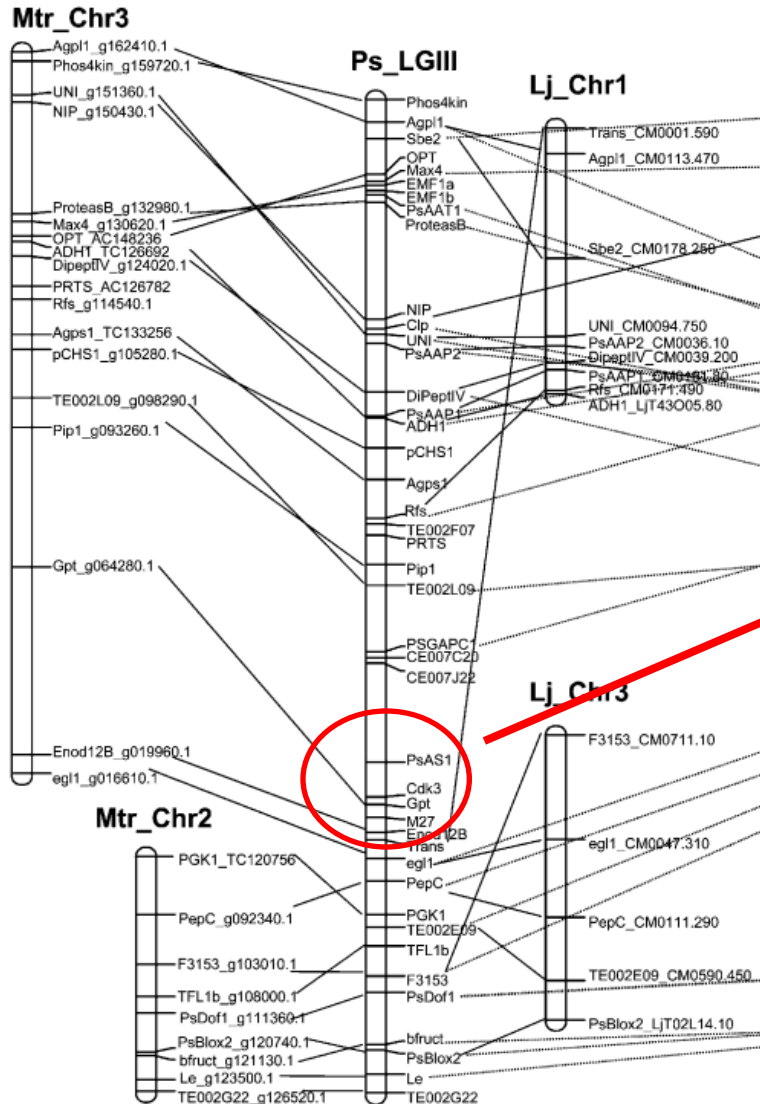
Grégoire Aubert,^{*} and Judith Burstin^{*5}

^{*}INRA, UMRLEG, 21065 Dijon Cedex, France [†]INRA, UMR1095, 63100 Clermont-Ferrand, France



Využití syntenie s modelovými organismy

A

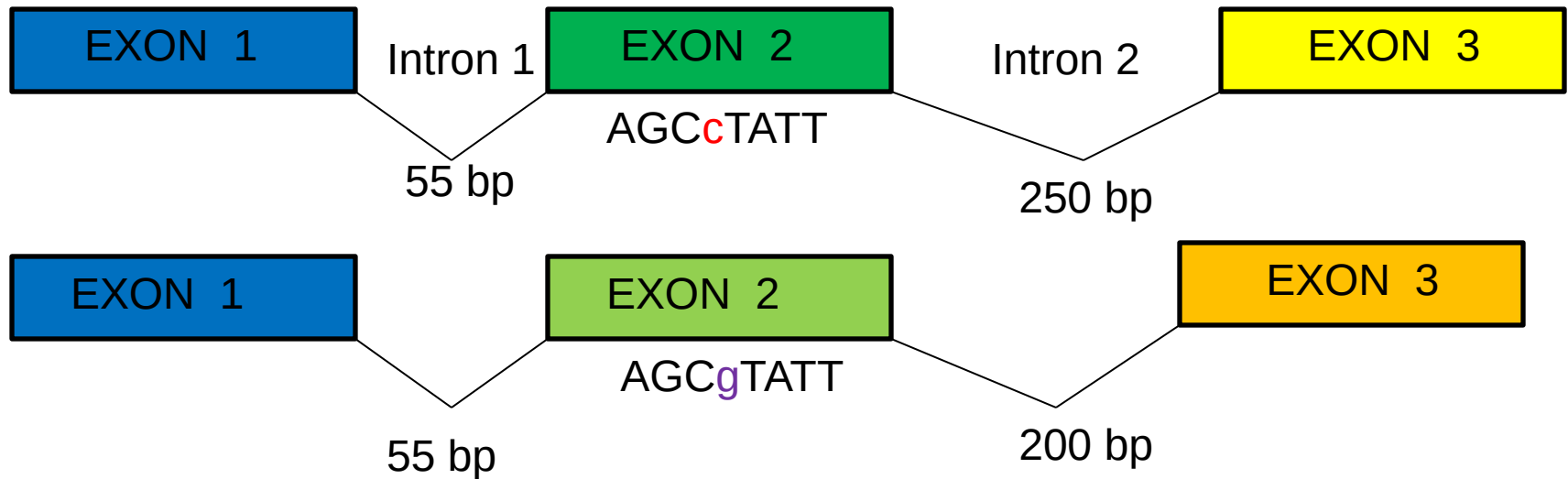


Hrách

Versus

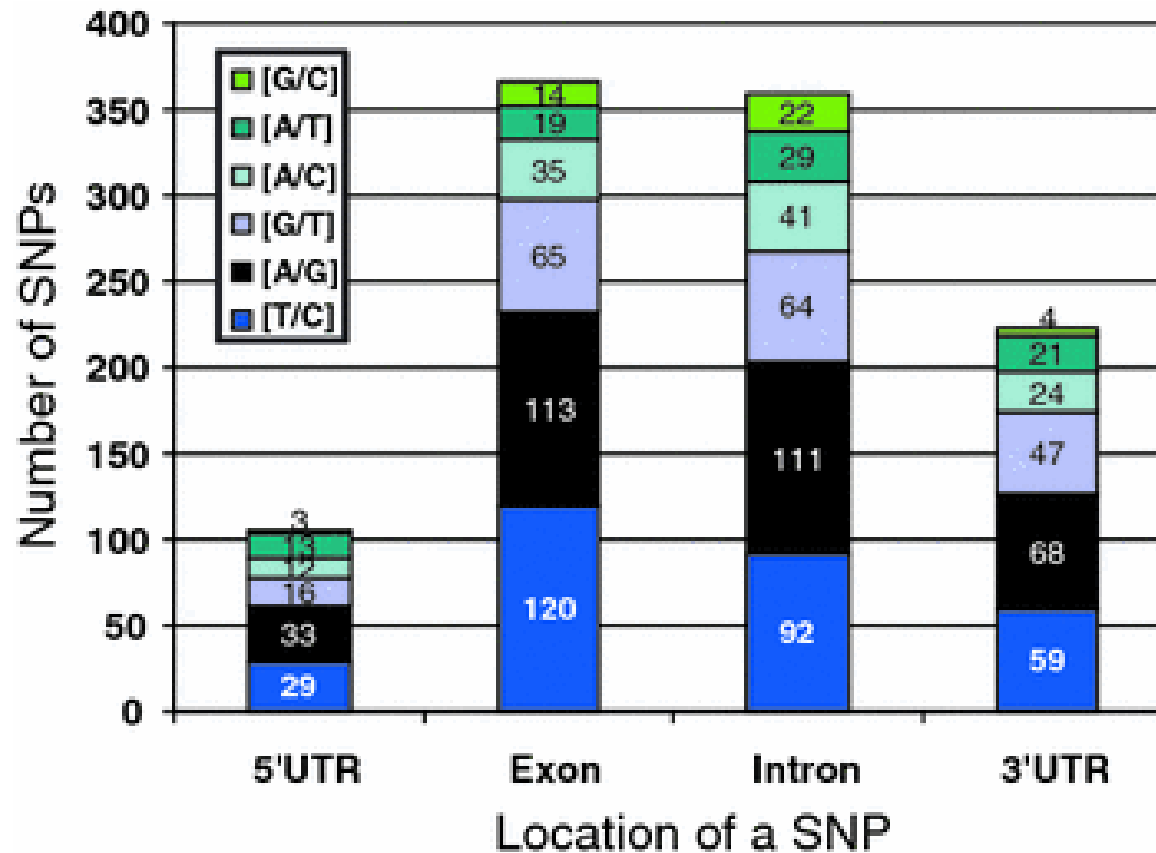
Medicago truncatula

Single nucleotide polymorphism

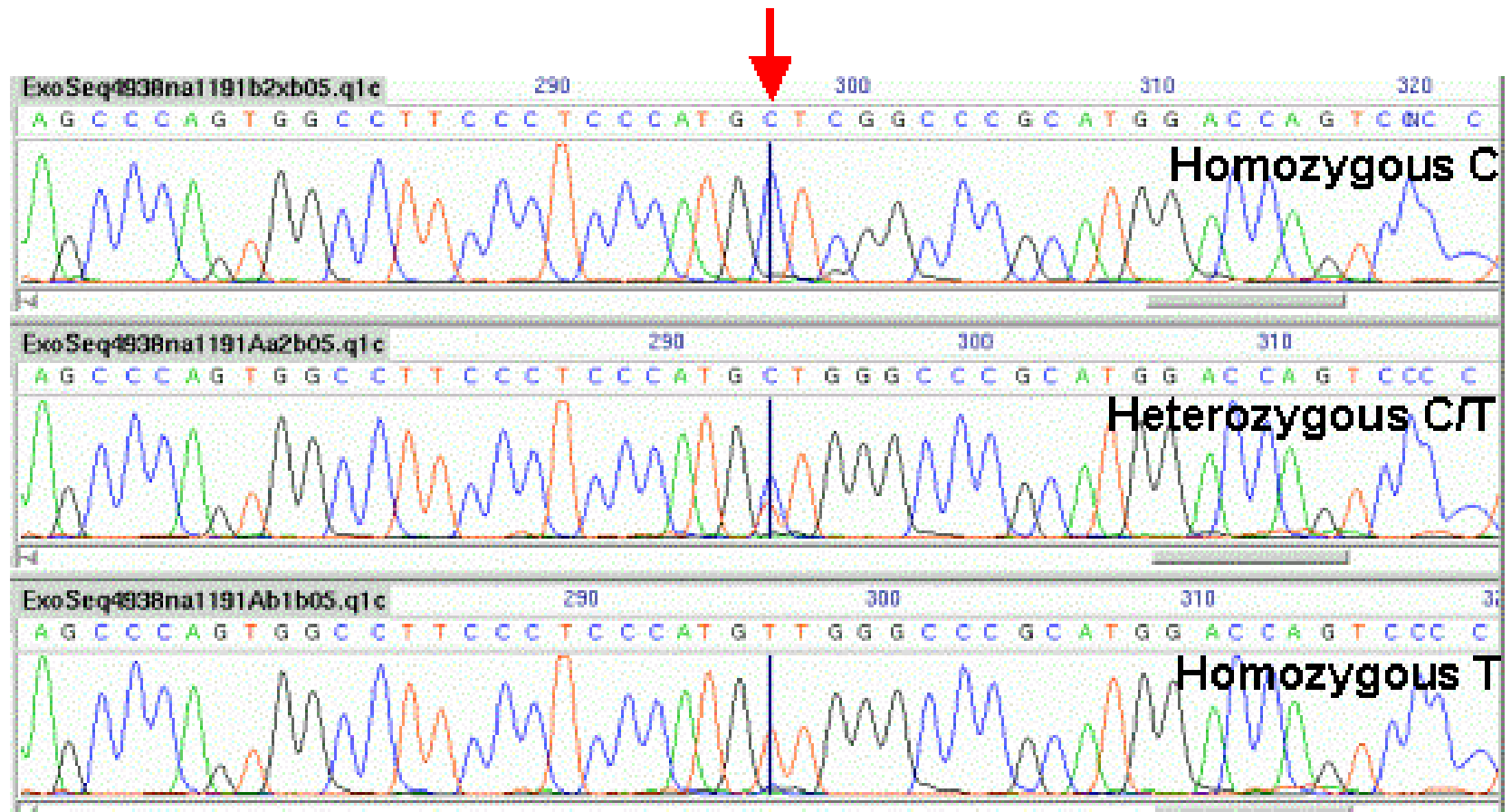


indel – inzertion / deletion

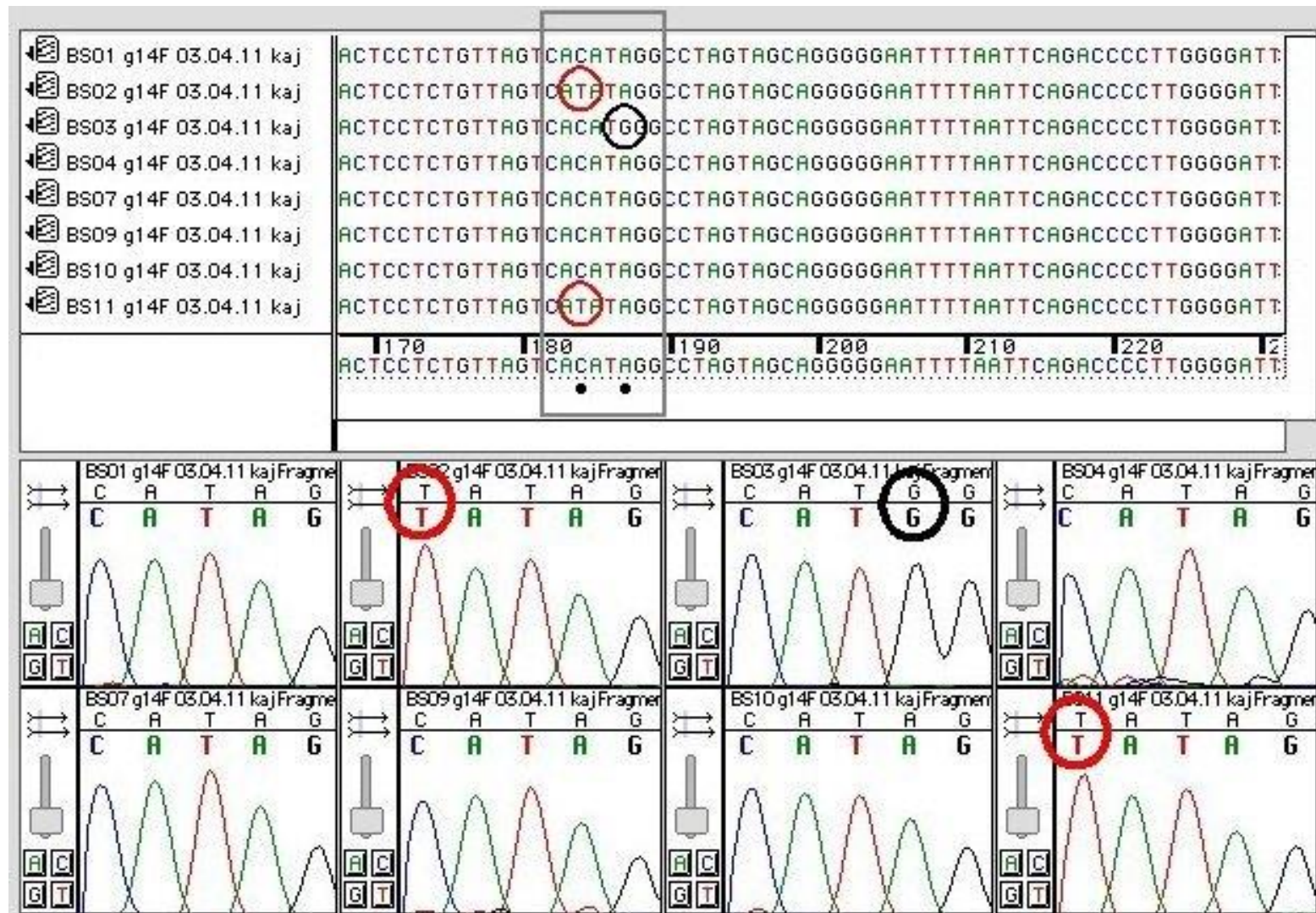
Distribuce a typy SNP(ů)



Single Nucleotide Polymorphism (SNP)

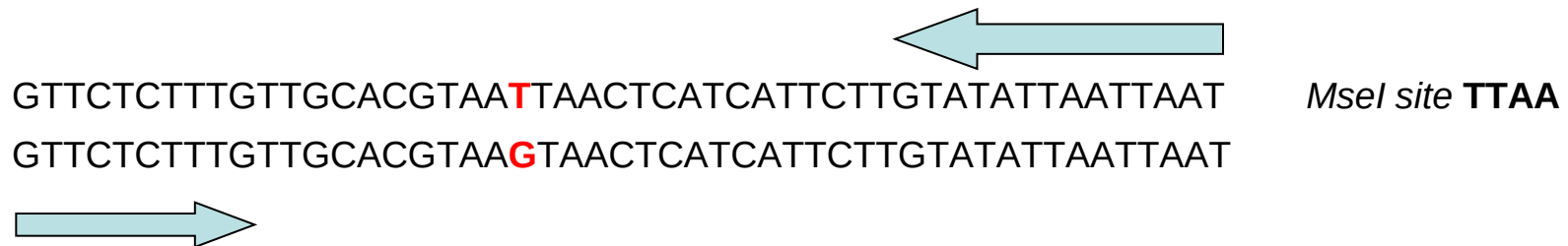


Gen po genu - sekvenováním

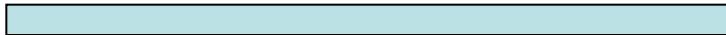


Pomalé, nákladné - vhodné pro menší počet vzorků

Cleavage Amplified Polymorphic Sequence (CAPS-PCR)



1. PCR

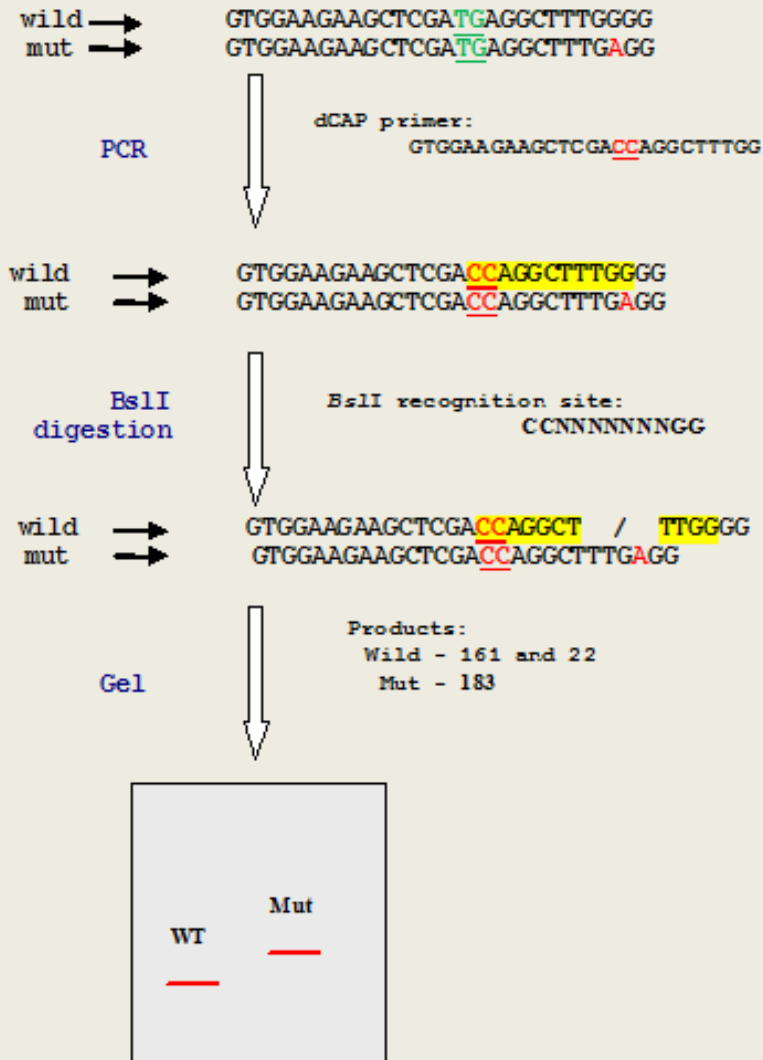


2. štěpení



Pokud existuje vhodné restrikční místo v cílovém fragmentu

The *Bs*II dCAPS assay: creating the RE recognition site to discern wild and mutant sequences



Derived Cleaved Amplified Polymorphic Sequences (dCAPS)

Pokud NEexistuje vhodné restrikční místo v cílovém fragmentu

Nutno jej vytvořit nově v PCR primeru

Následně je rozdíl jen o délku primeru (16-25 bp) – potřeba dobrého rozlišení

Genově specifické markery - CAPS/dCAPS

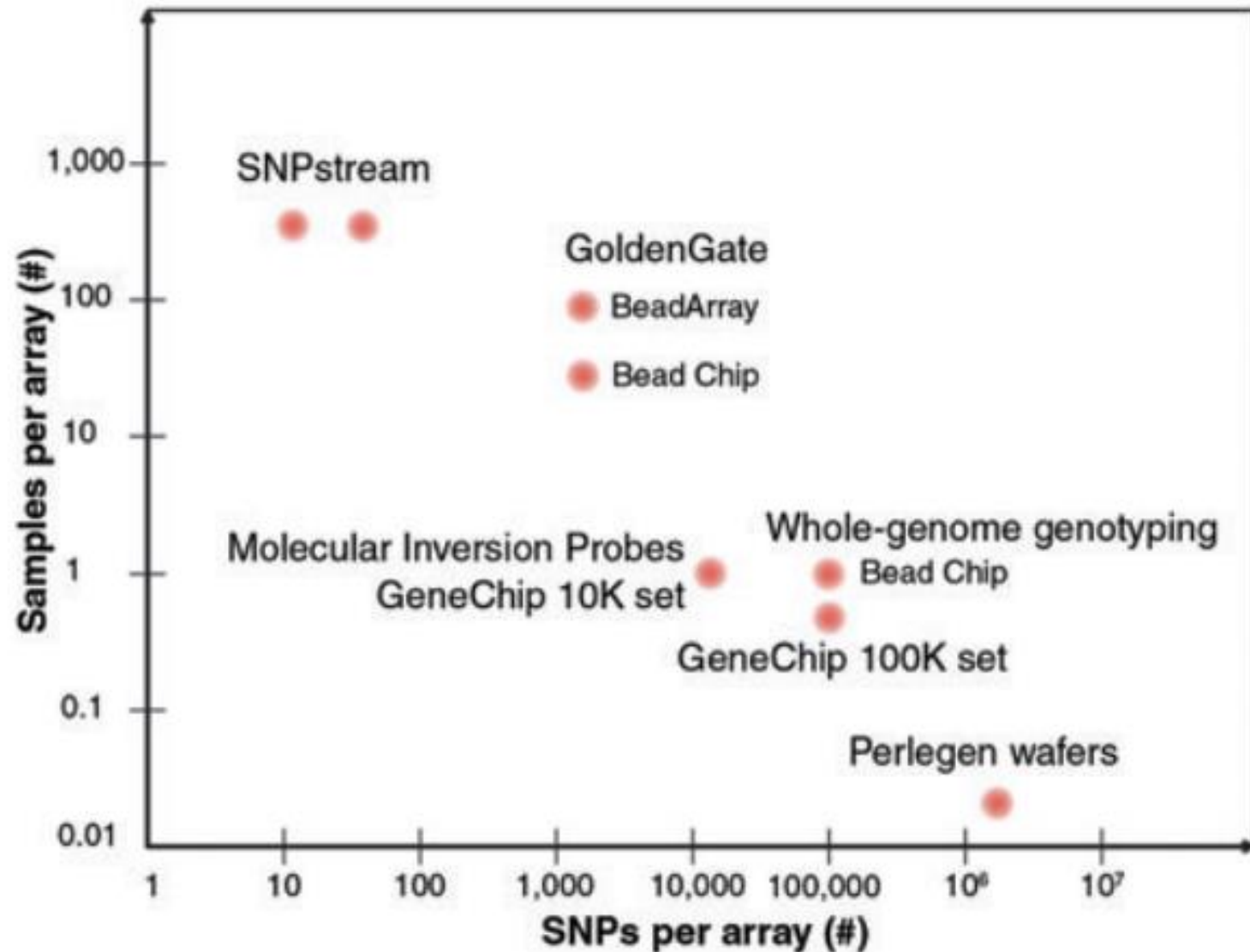
Výhody:

- ko-dominantní
- locus-specifické
- jednoduše hodnotitelné a interpretovatelné
- lehce adaptovatelné mezi laboratořemi
- Nejlépe pro mapování mezi 2 genotypy/rodiči

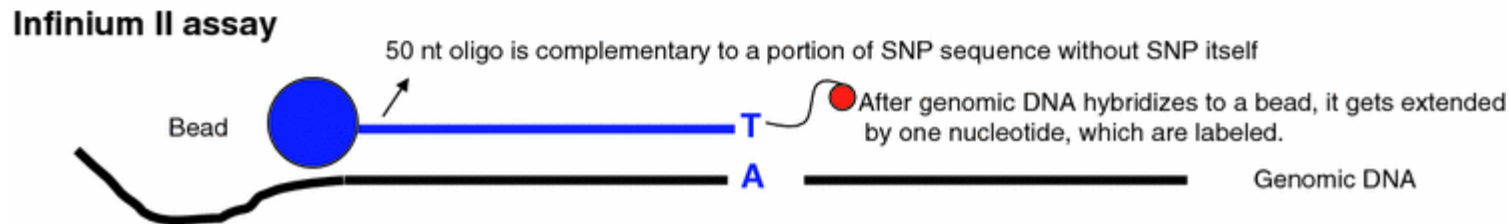
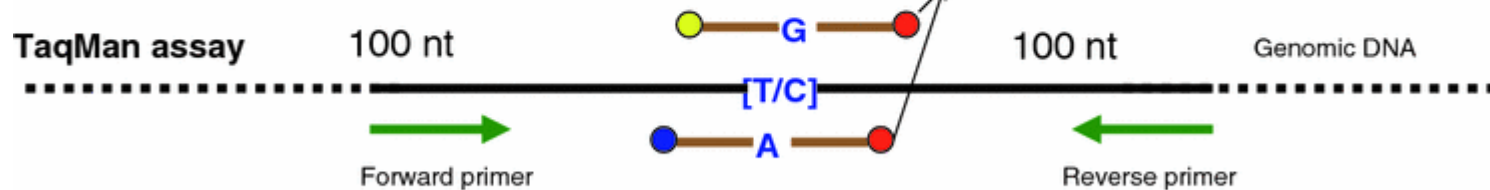
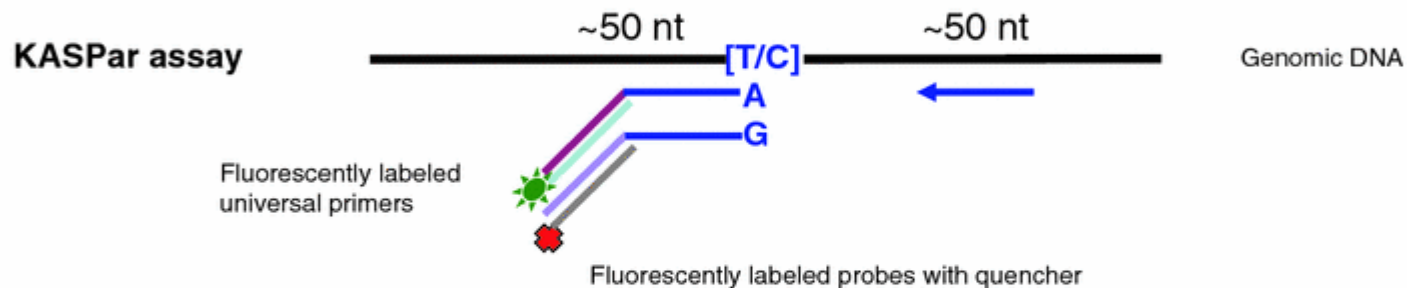
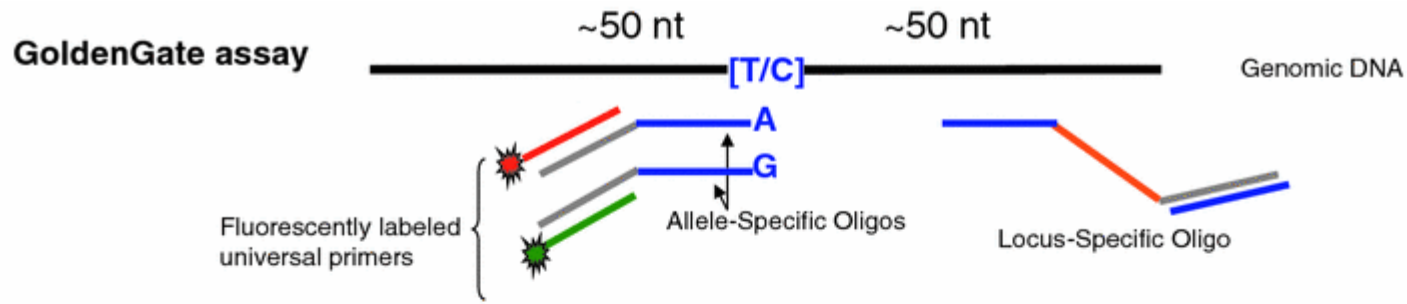
Nevýhody:

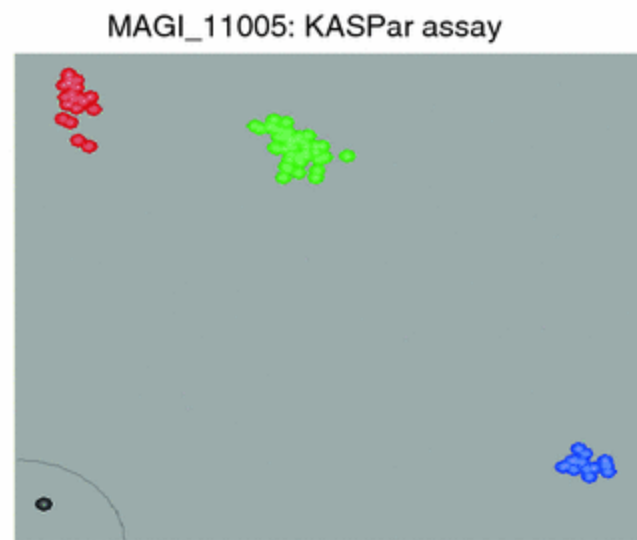
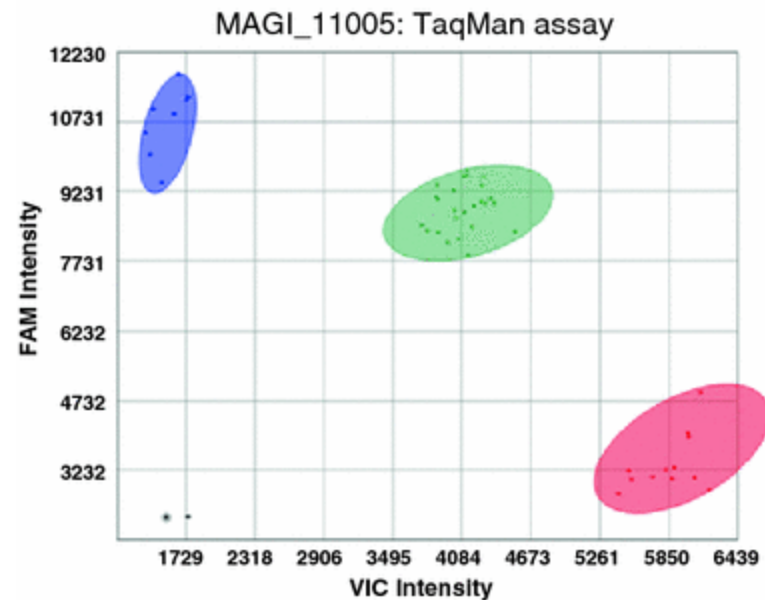
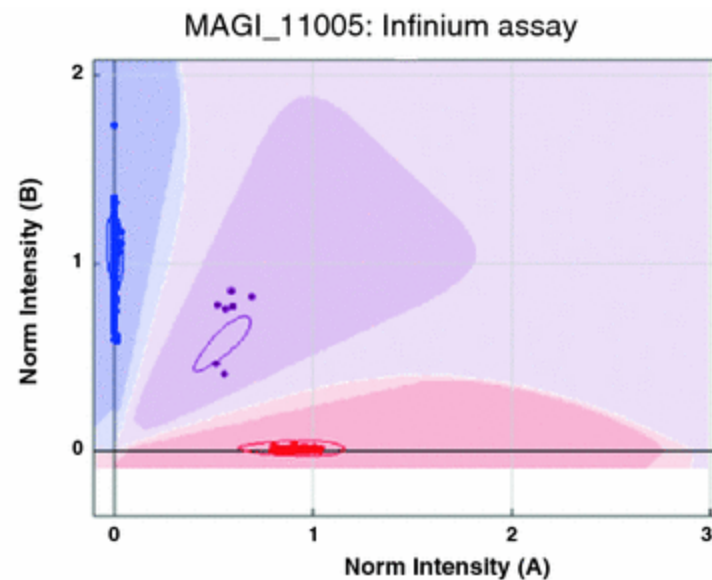
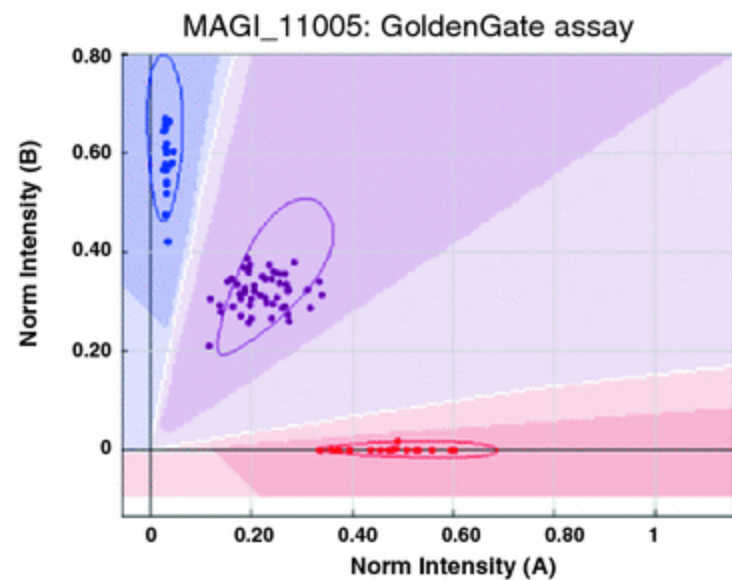
- nutnost sekvenování
- často malý polymorfismus mezi příbuznými genotypy
- malá produktivita - gen po genu

Kolik SNPs detekujeme versus kolik vzorků najednou

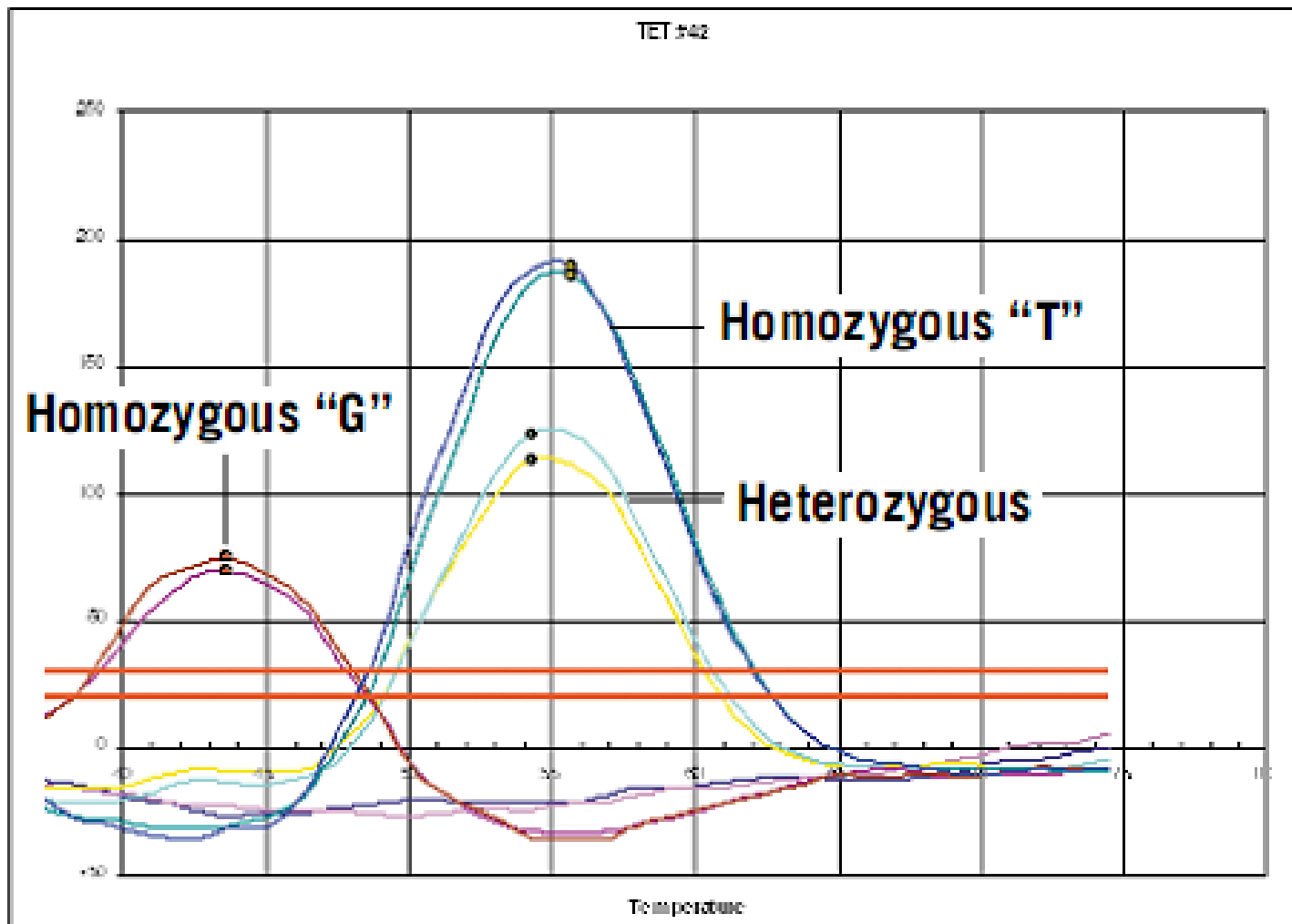


Specificita detekce dané během PCR reakce

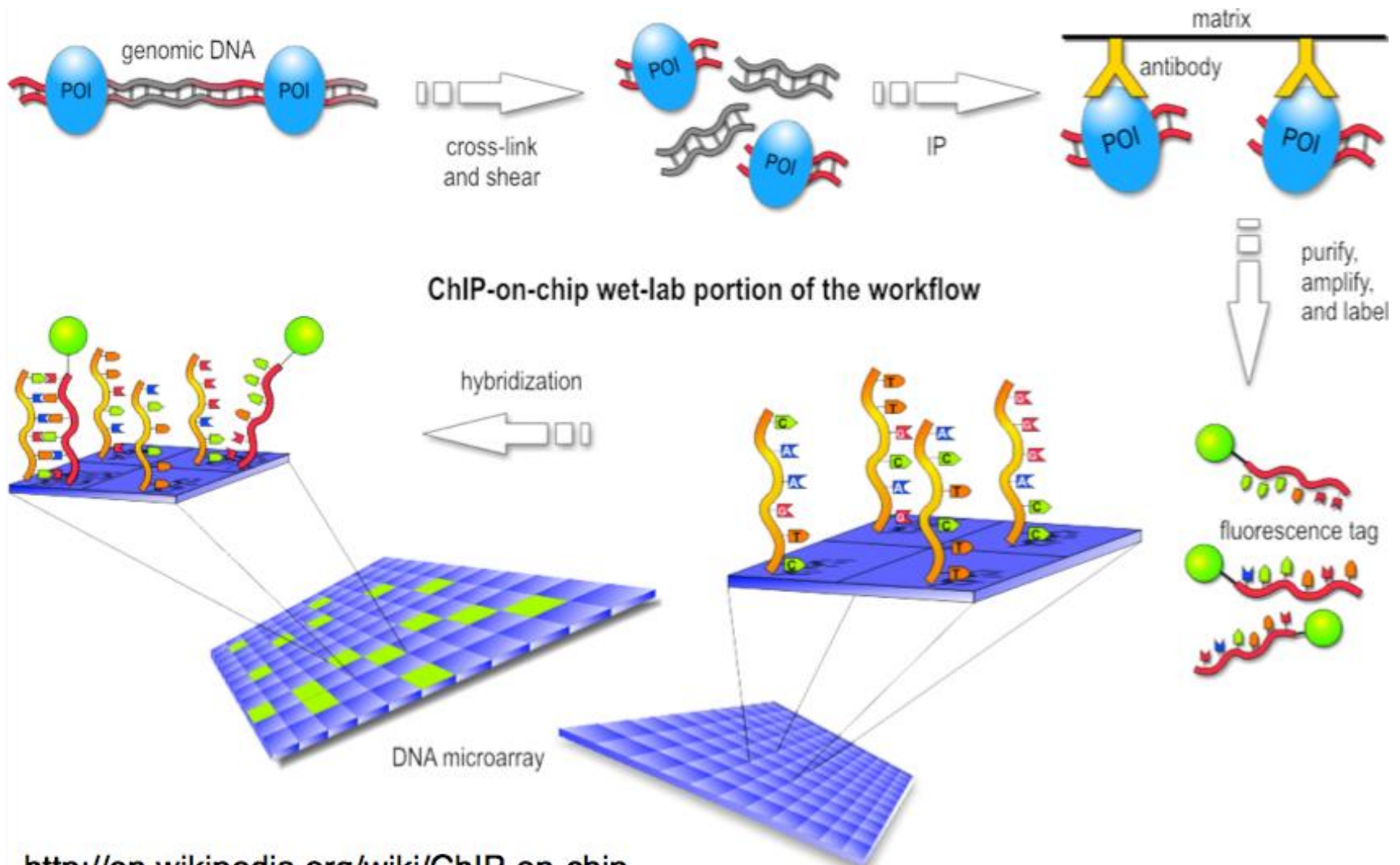




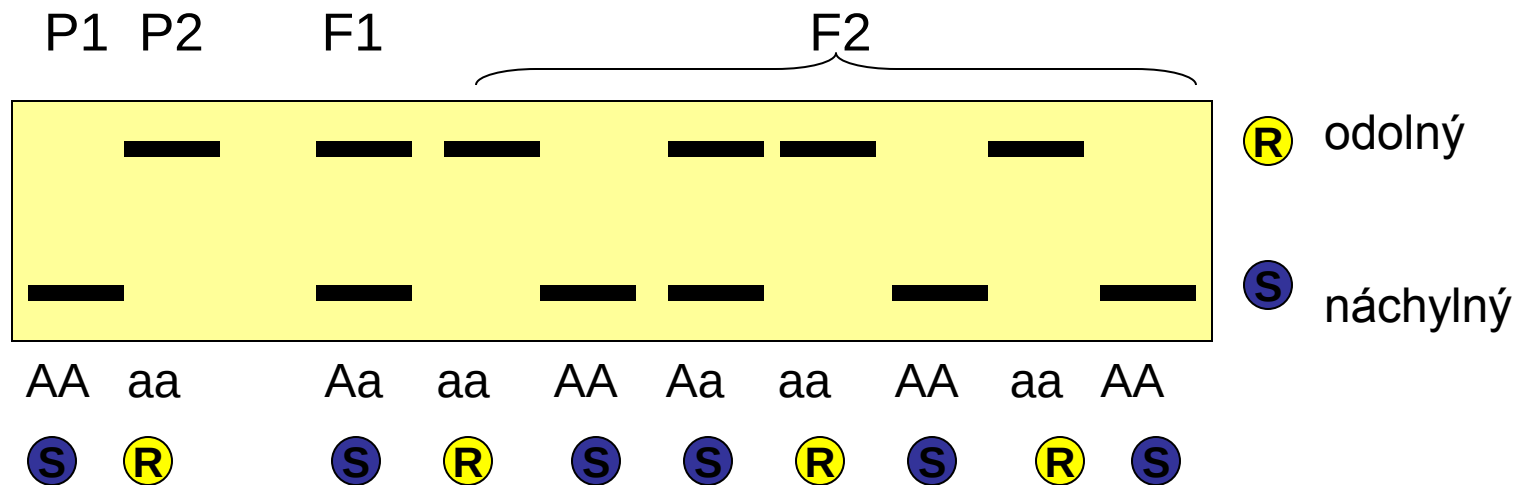
Detekce založená na rozdílné T_m teplotě



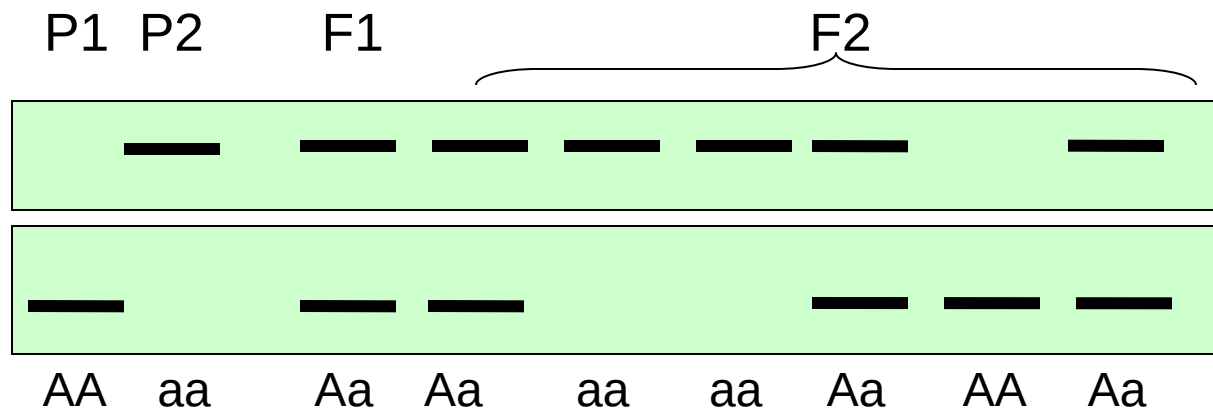
Hybridizací na chipech



Kodominantní marker – odlišení i heterozygota



Dominantní marker – nutnost 2 analýz



Mathematical evaluation and data scoring

Used methods :

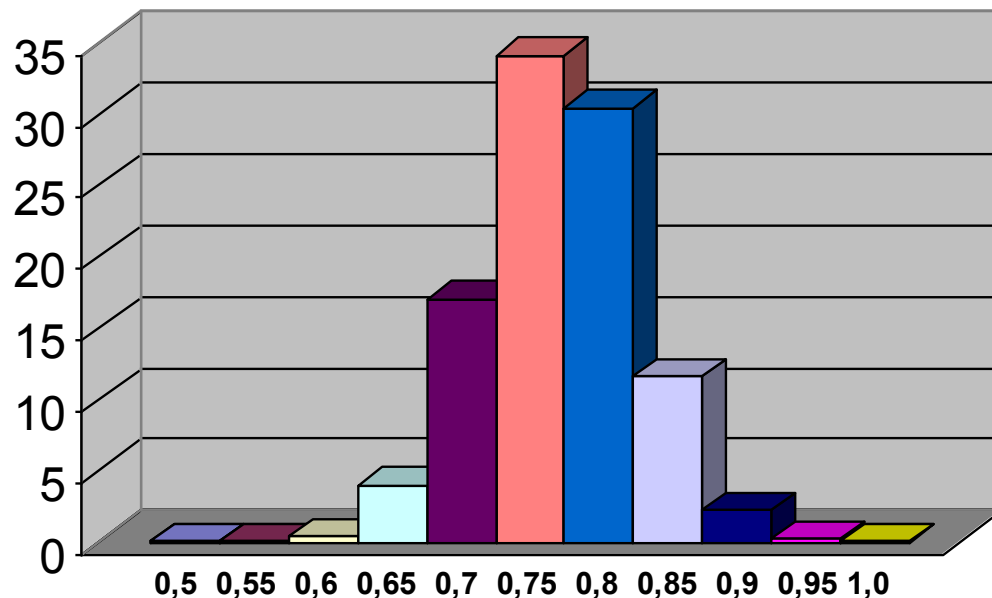
- **genetic distances calculations - coefficients of dis-/similarity**
 - **ordination, clustering (PCA, PCO, MDS)**
 - **dendrograms (cluster analysis - UPGMA, Ward's method)**
 - **allele frequencies, genetic distances between groups**
 - **model-based (Bayesian inference)**
-

Data evaluation and scoring

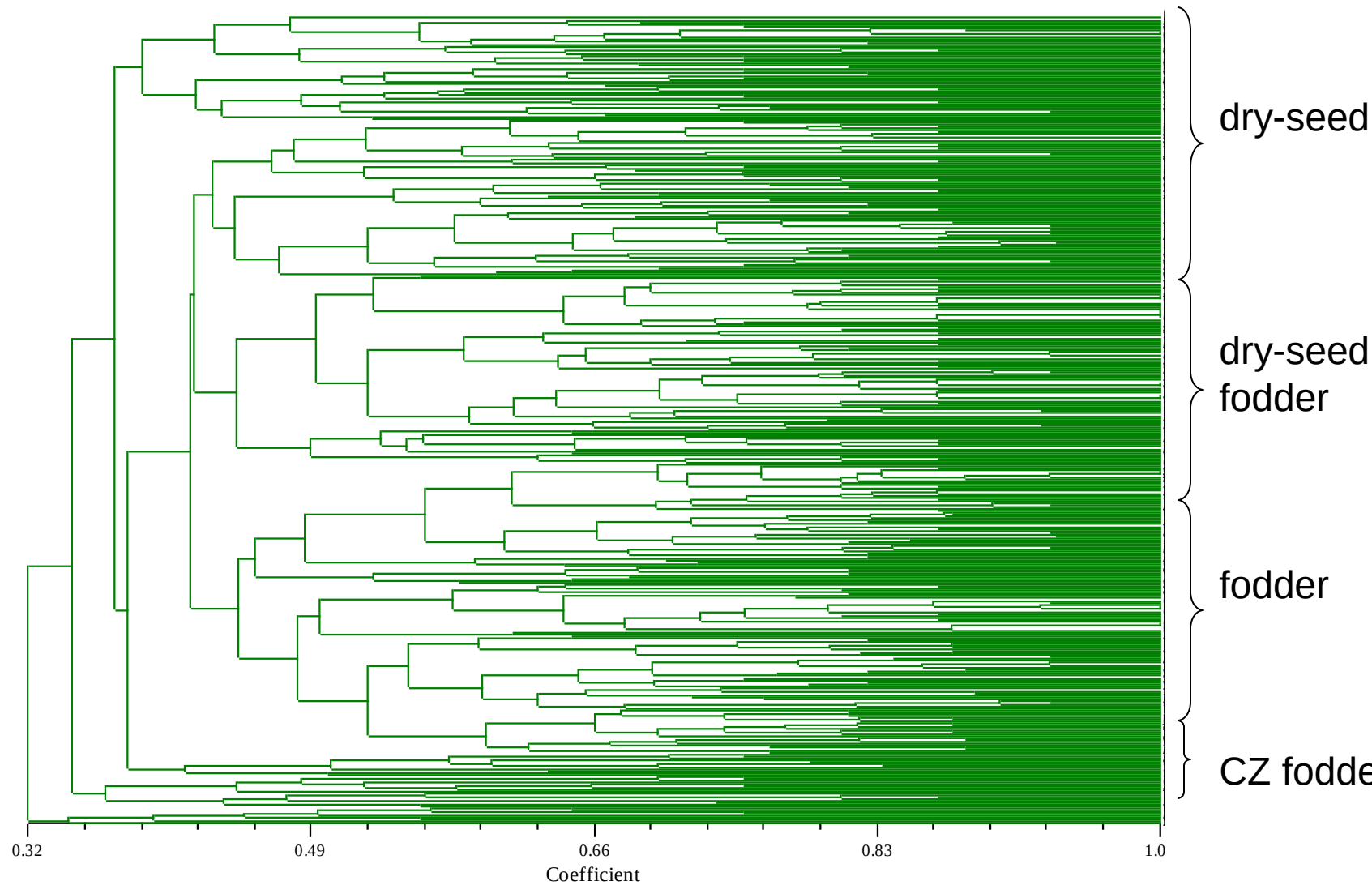
Distance-based approaches

Distribution of Nei genetic distances in investigated set of pea germplasm

genetic distance distribution -
SSR+RBIP data



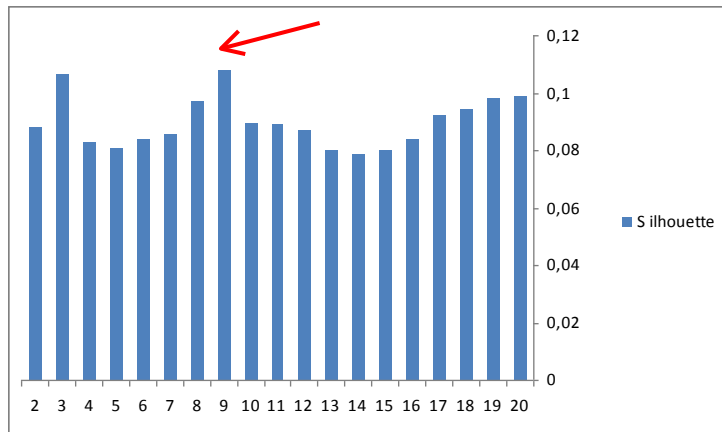
UPGMA analysis of 450 pea accessions



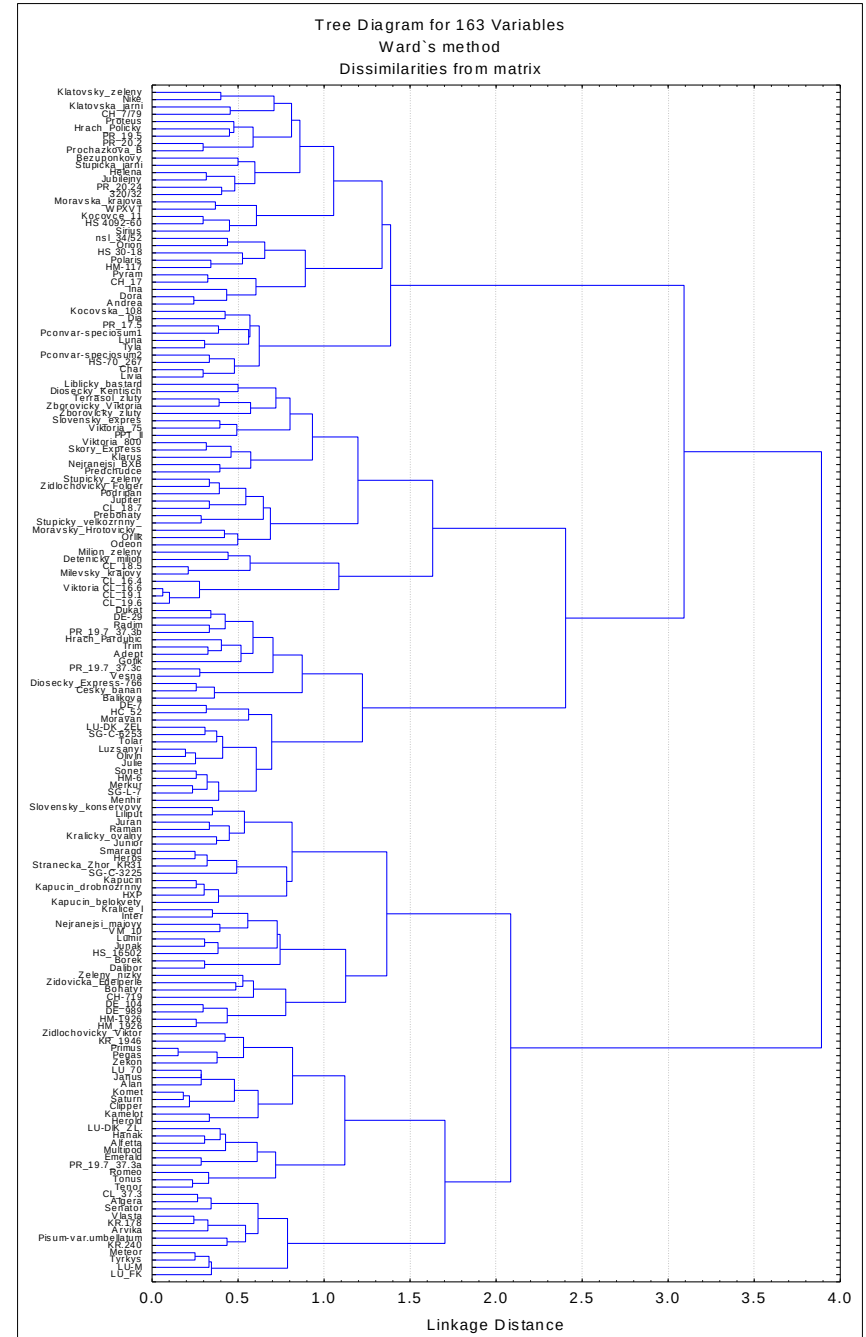
Ward's hierarchical cluster analysis

(minimalization of statistical variability within clusters
maximalization between - as ANOVA)

RBIP + SSR loci
dissimilarities matrix



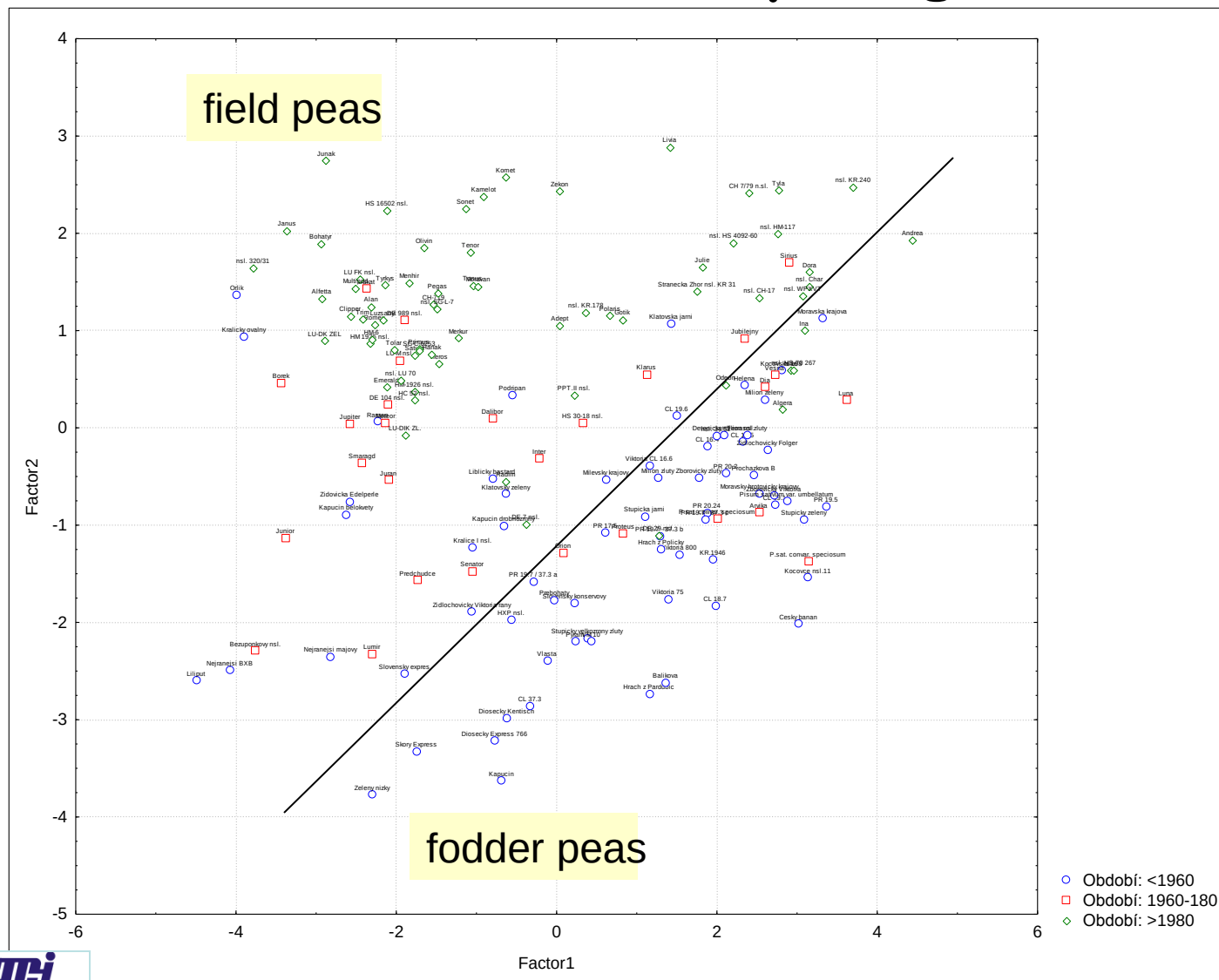
Silhouette method
for cluster number estimation



Multivariate statistics - ordination approaches

- to condense the differences into fewer characters and visualize them in multidimensional space
 - to identify the most discriminating characters
-

Principal component analysis (PCA) of morphological descriptors



Parametric data

PCA1: 33.2%

PCA2: 18.5%

finds components
that accounts for
as much original
variance as
possible

Bayesian and other models-based methods

Molecular data analysis by Bayesian inference STRUCTURE software

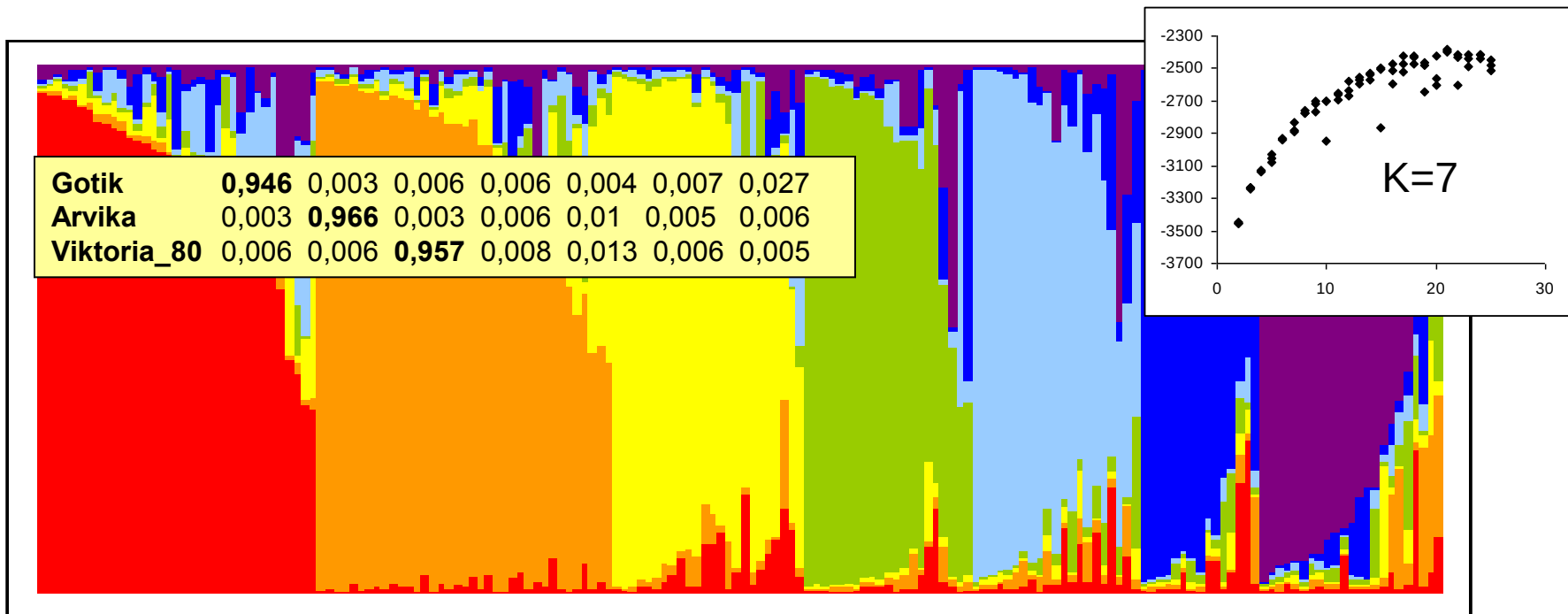


Table 2. Comparison of different characteristics of most frequently used molecular markers techniques.

Molecular Markers	RFLP	RAPD	AFLP	SSR	CAPS	SCAR	IRAP/ REMAP	RAMP	SSCP	SNP
Degree of polymorphism	M	M	M	M	L	M	M	M	L	H
Locus specificity	Y	N	N	N	Y	Y	Y	Y	Y	Y
Dominance (D)/Co-dom. (C)	C	D	D	C	C	C	D	D	C	C
Ease of Replication	H	L	H	M	H	H	H	M	M	H
Abundance	H	H	H	M	L	L	H	M	L	H
Sequence information required	Y	N	N	N	Y	Y	Y	N	Y	Y
Quantity of DNA required	H	L	M	L	L	L	L	L	L	L
Automation	N	Y	Y	Y	Y	Y	Y	Y	N	Y
Costs per assay	H	L	M	L/M	M	L	L	M	H	L
Technical requirement	H	L	M	L/M	H	M	H	H	H	M

Key: H = High; M= Medium; L = Low; Y = Yes; N= No

Markery pro fylogenetické studie

Požadavky:

- 1.) **univerzálnost** – aplikovatelnost napříč druhy, rody i čeleděmi
- 2.) **dostatečný polymorfismus**
- 3.) **liniovost** – možnost sledování linie (maternální dědičnost)

DNA barcoding

- Chloroplastové sekvence (*matK*, *rbcLA*, *trnSG*)
- Jaderně kódované (Internal transcribed spacer rDNA – ITS)

DNA barcoding a molecular operational taxonomic units (MOTU)

648 bp mtDNA *cytochrome oxidase 1* (COX1) lokus
pro **obratlovce a hmyz**

The Consortium for the Barcode of Life (CBOL)

Rychlejší evoluce, mutační rychlost než ntDNA

Univerzálnost sekvence (primerů)

Odhadnuto že veškerí živočichové by mohly být takto
popsány během 10 let za cca 300 mil USDA (ale
pravděpodobně méně)

Kontraverze:

Polymorfismus populací, bias pro samičí pohlaví, gene
flow a hybridizace, přenos mtDNA do ntDNA

Barcode of Life Datasystems

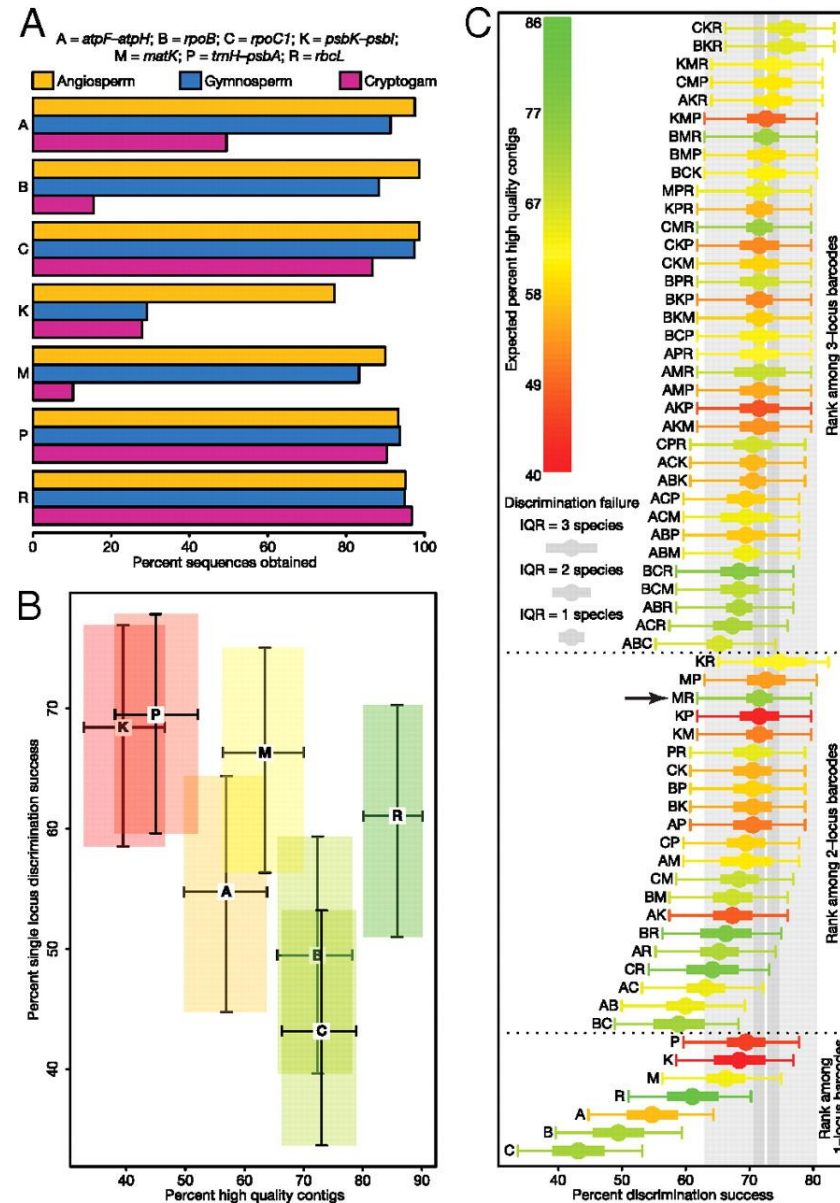
Specimen Records : 3,561,705 Specimens with Barcodes : 2,651,524

Species with Barcodes : 193,653

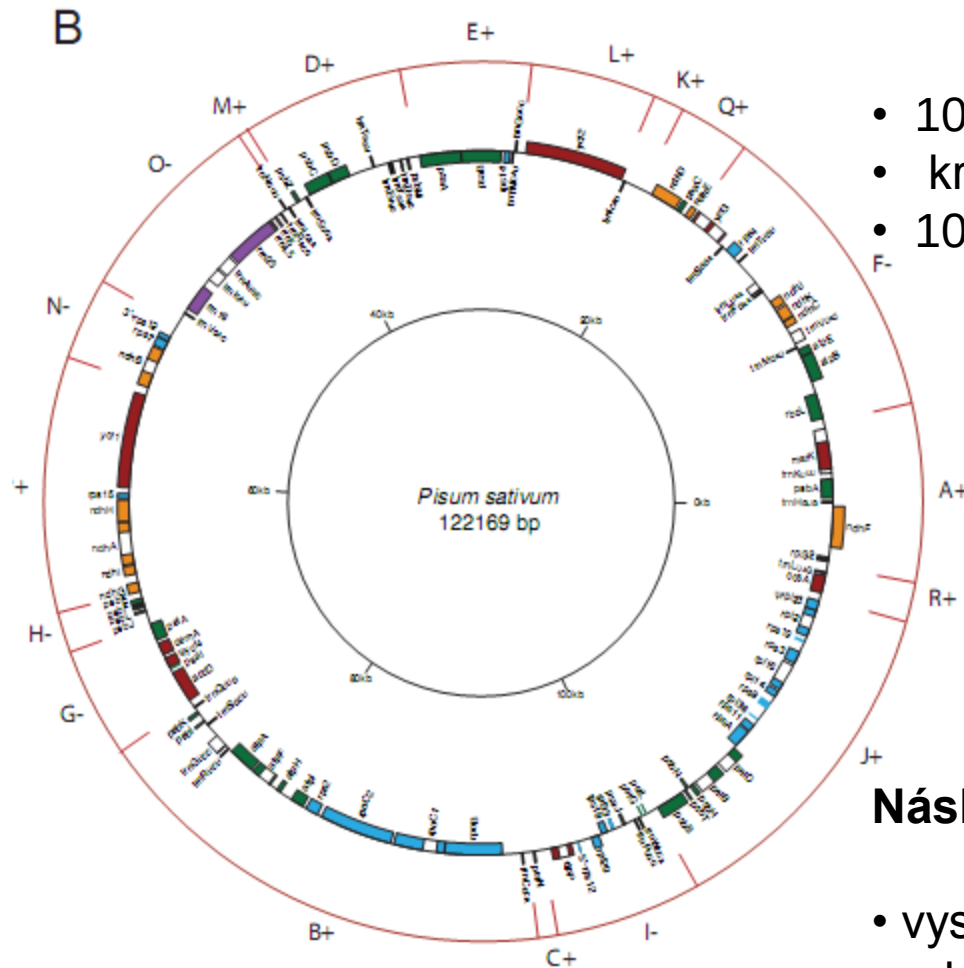
Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi
(PNAS 109 (16), Mar 27 2012)

The Chaperonin-60 Universal Target Is a Barcode for Bacteria That Enables De Novo Assembly of Metagenomic Sequence Data
(PLoS One 7(11):e49755, Nov 26 2012)

Plant DNA barcodes
rbcL+*matK*+*trnH-psbA*
ITS



Chloroplastová DNA



- 100 – 250 kb
- kruhová DNA
- 10 -100 kopií na buňku

Následně:

- vysoká stabilita DNA
- robustnost PCR detekce
- možnost detekce i z herbářových materiálů

Highly Variable Chloroplast Markers for Evaluating Plant Phylogeny at Low Taxonomic Levels and for DNA Barcoding

Wenpan Dong^{1,2}, Jing Liu^{1,3}, Jing Yu^{1,3}, Ling Wang², Shiliang Zhou^{1*}

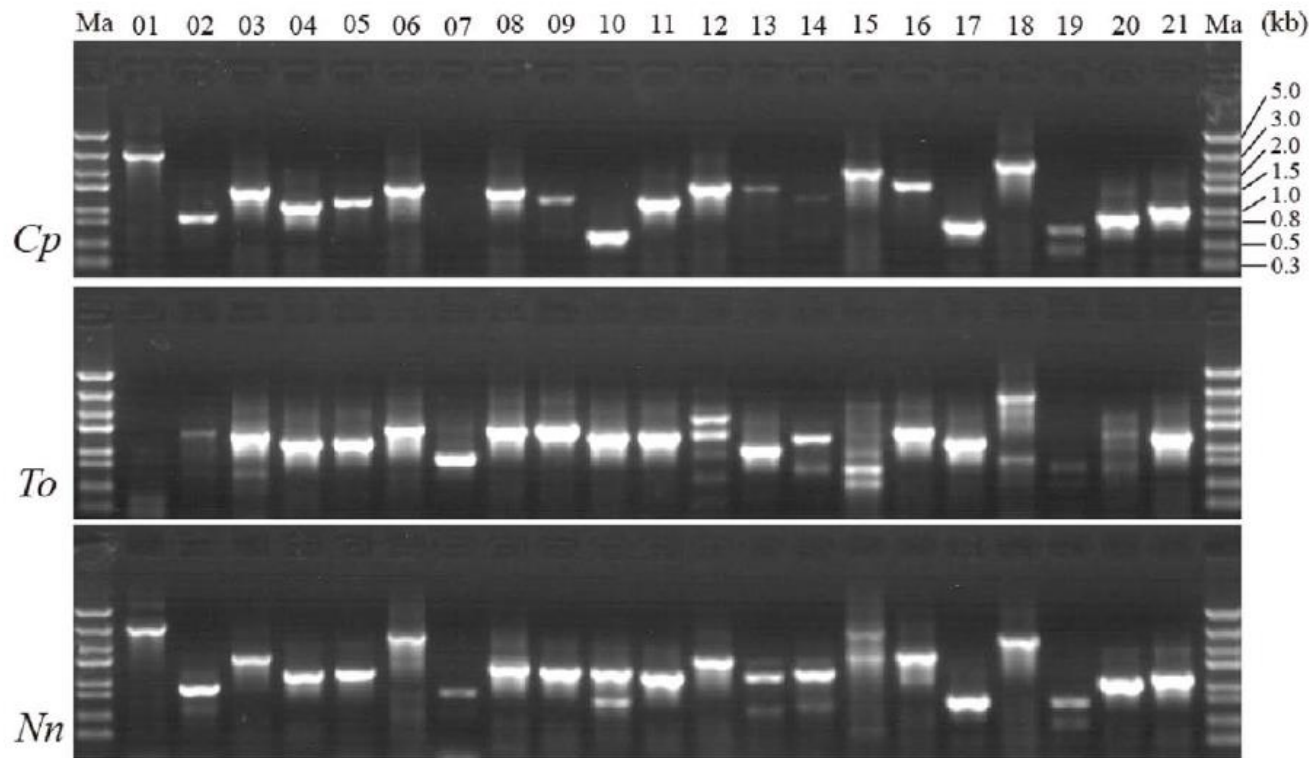
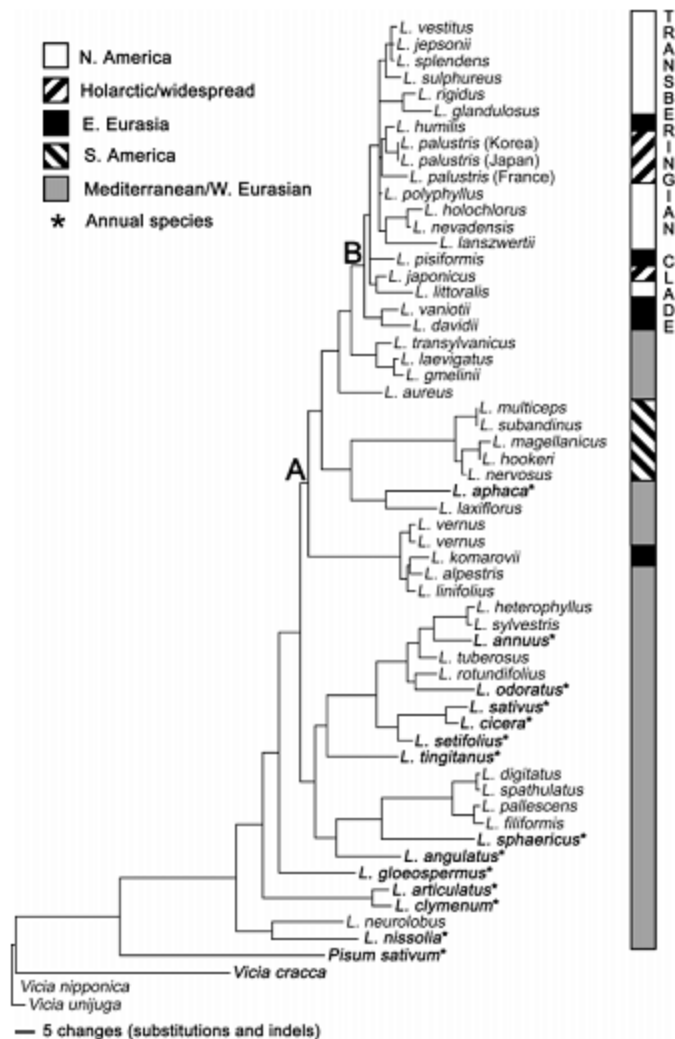


Figure 2. Gel profiles of fragments amplified from six species using 21 pairs of primers. Numbers shown at top are sequential order of loci as in Table 2 and Fig. 1. Numbers on right are size markers (kbp). Letters on left indicate species as follows: Cp: *Chimonanthus praecox* (L.) Link; To: *Typha orientalis* Presl.; Nn: *Nelumbo nucifera* Gaertn.; Ps: *Paeonia suffruticosa* Andrews; Pp: *Prunus persica* (L.) Batsch.; and Pb: *Panax bipinnatifidus* Seem.

doi:10.1371/journal.pone.0035071.g002

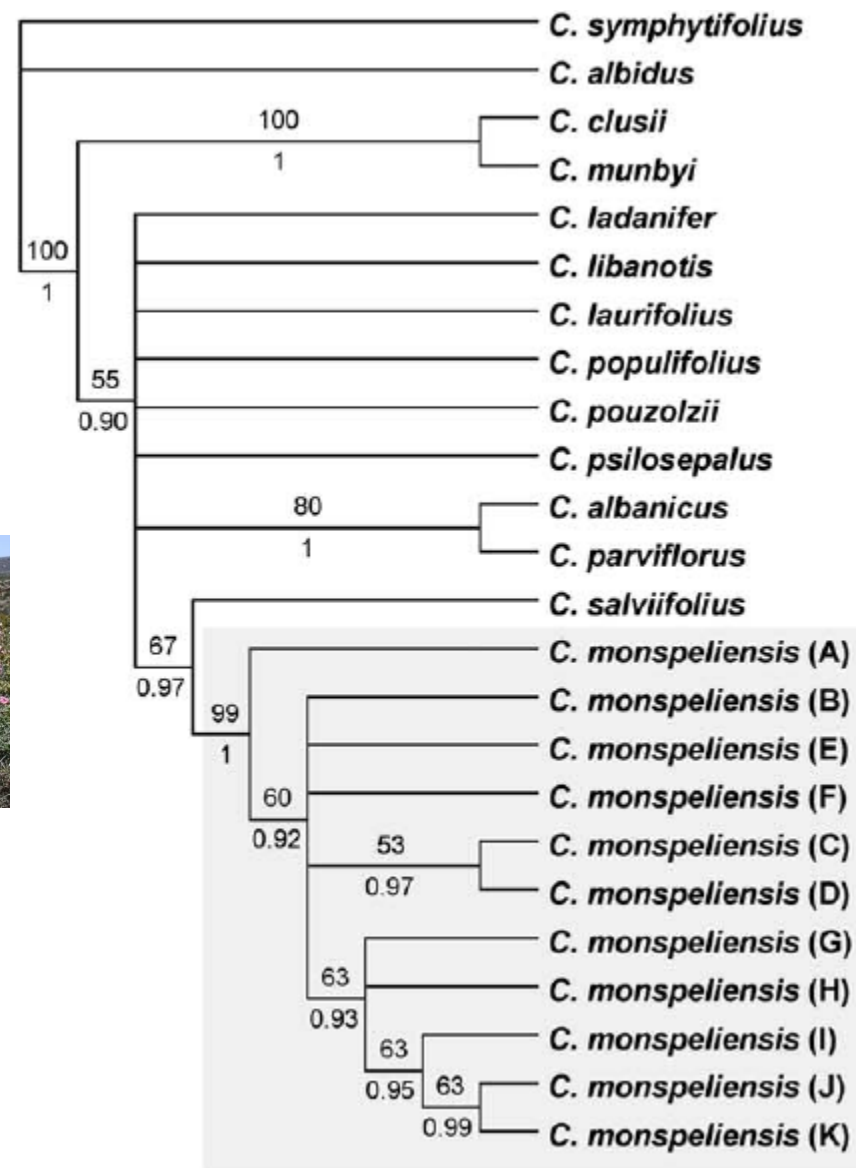
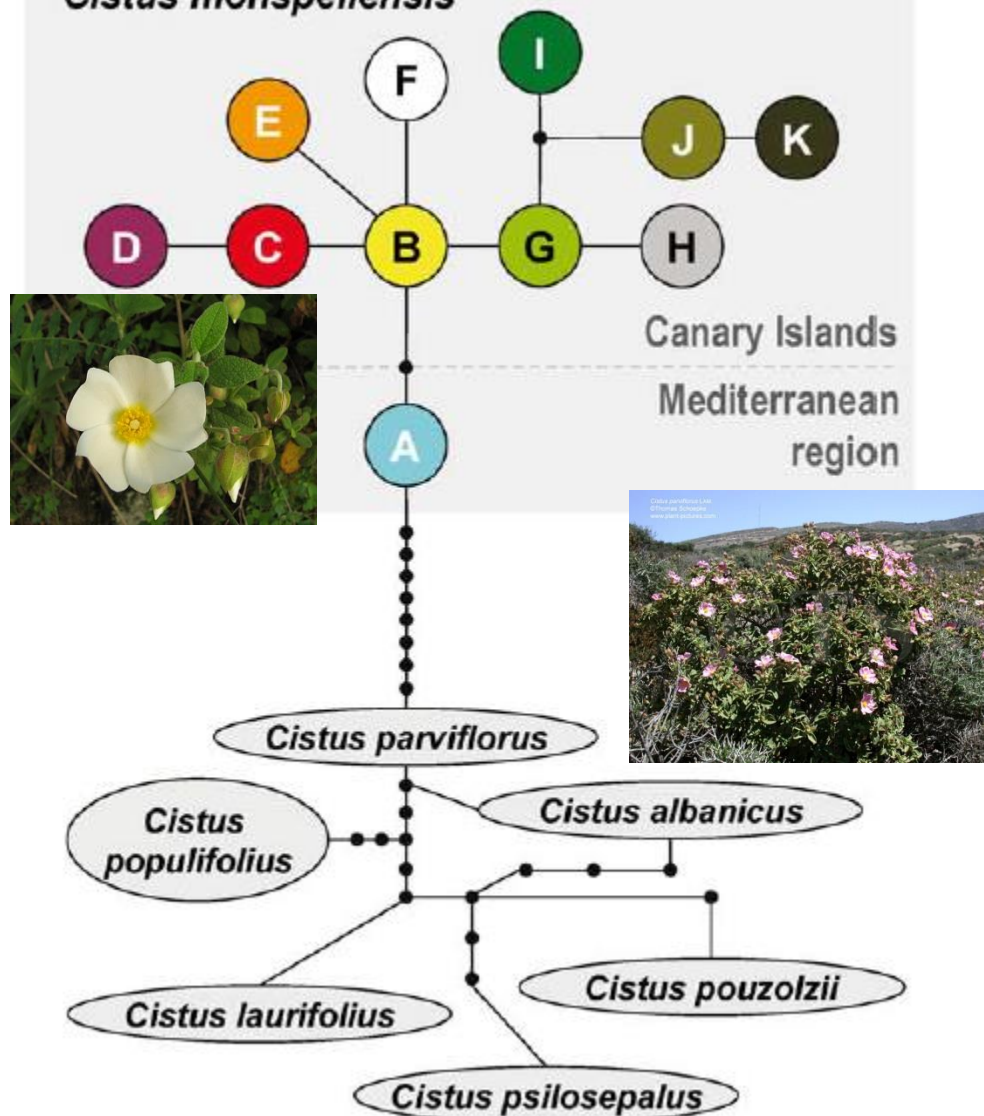
	40	50	250	320	470	620	630
p_abyssinicum_AGT102	CATCCATACCATCGC	ATTATAGGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
p_abyssinicum_1974	CATCCATACCATCGC	ATTATAGGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
p_abyssinicum_2385	CATCCATACCATCGC	ATTATAGGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
p_humile_241	CATCCCTACCATCGC	ATTATAAGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
p_sativum_atfc6927	CATCCCTACCATCGC	ATTATAAGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
p_humile_1794	CATCCCTACCATCGC	ATTATAAGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
p_asiaticum-matk	CATCCCTACCATCGC	ATTATAAGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
p_transcaasicum-	CATCCCTACCATCGC	ATTATAAGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
p_elatius_3151	CATCCCTACCATCGC	ATTATAAGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
p_elatius_3156	CATCCCTACCATCGC	ATTATAAGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
p_elatius_1703	CATCCCTACCATCGC	ATTATAAGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
p_elatius_3155	CATCCATACCATCGC	ATTATAGGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
p_elatius_3147	CATCCATACCATCGC	ATTATAGGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
p_elatius_3149	CATCCATACCATCGC	ATTATAGGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
p_fulvum_1006	CATCCATACCATCGC	ATTATAGGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGA-----ATA		
p_fulvum_2517	CATCCATACCATCGC	ATTATAGGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGA-----ATA		
p_fulvum_2473	CATCCATACCATCGC	ATTATAGGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGA-----ATA		
p_fulvum_2523	CATCCATACCATCGC	ATTATAGGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGA-----ATA		
p_fulvum_1010	CATCCATACCATCGC	ATTATAGGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGA-----ATA		
matk-10925	CATCCCTACCATCGC	ATTATAAGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
20092x20	CATCCATACCATCGC	ATTATAGGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
vir1851	CATCCCTACCATCGC	ATTATAAGGAAAC	TAAA-AAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
26127x22	CATCCATACCATCGC	ATTATAGGGAAAC	TAAATAAAGG	TTTTT-CCAA	ATATAGATATAGAATA		
343972	CATCCATACCATCGC	ATTATAGGGAAAC	TAAA-AAAGG	TTTTTTCCAA	ATATAGATATAGAATA		
ji3557	CATCCATACCATCGC	ATTATAGGGAAAC	TAAA-AAAGG	TTTTTTCCAA	ATATAGATATAGAATA		
ji3558	CATCCATACCATCGC	ATTATAGGGAAAC	TAAA-AAAGG	TTTTTTCCAA	ATATAGATATAGAATA		
26109x21	CATCCATACCATCGC	ATTATAGGGAAAC	TAAATAAAGG	TTTTTT-CCAA	ATATAGATATAGAATA		

matK



Fylogram odvozený na základě analýzy
cpDNA (*matK*, *trnSG*)

Cistus monspeliensis



Genetic diversity and origin of timopheevi wheat inferred by chloroplast DNA fingerprinting

Naoki Mori^{*1)}, Yasutaka Kondo¹⁾, Takashige Ishii²⁾, Taihachi Kawahara³⁾, Jan Valkoun⁴⁾ and Chiharu Nakamura¹⁾

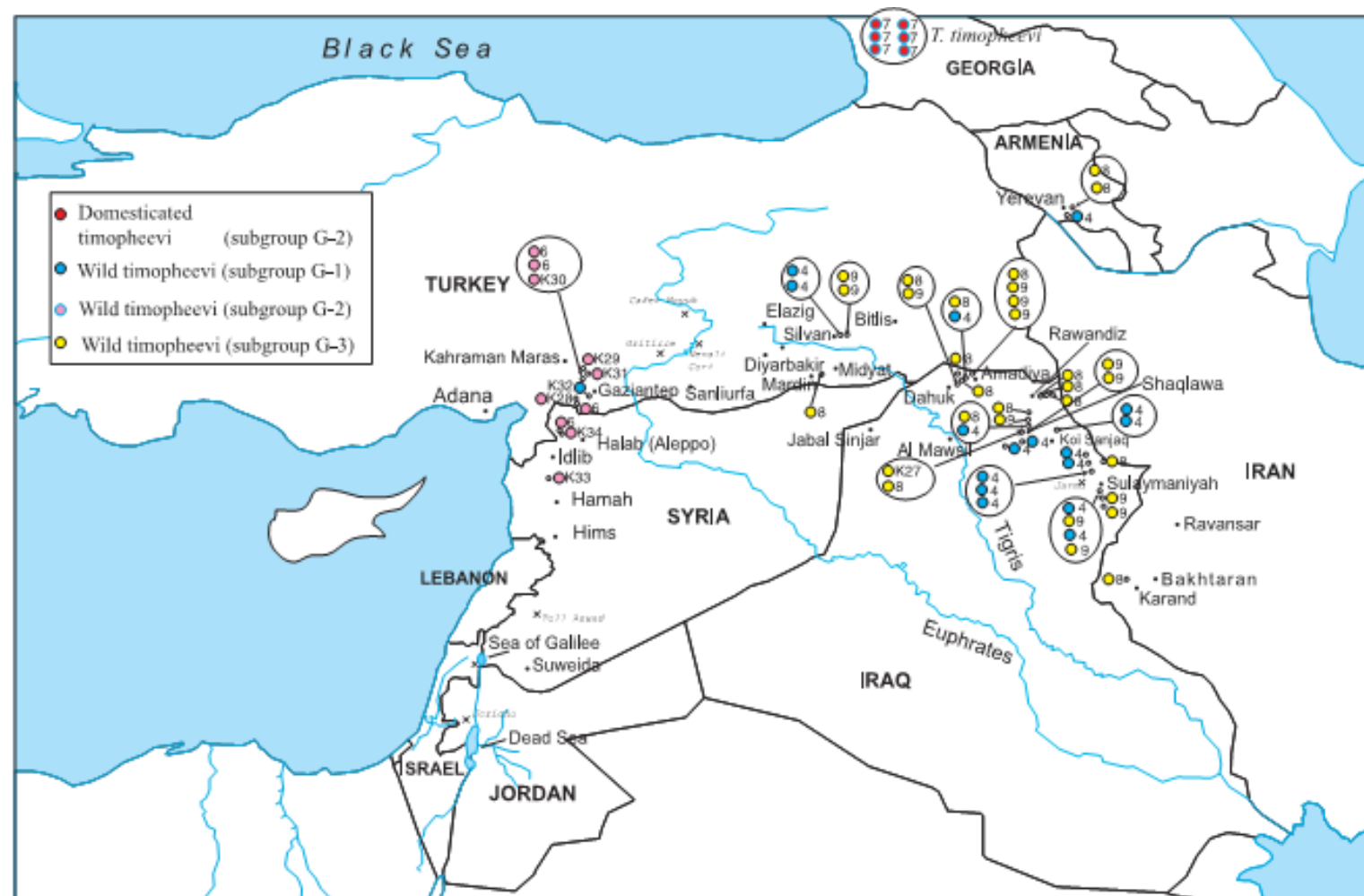


Fig.3. Geographical distribution of plastotypes found in 57 accessions of wild timopheevi wheat and six accessions of domesticated timopheevi

Vnitřní transkribovaný mezerník (ITS) mezi rDNA geny

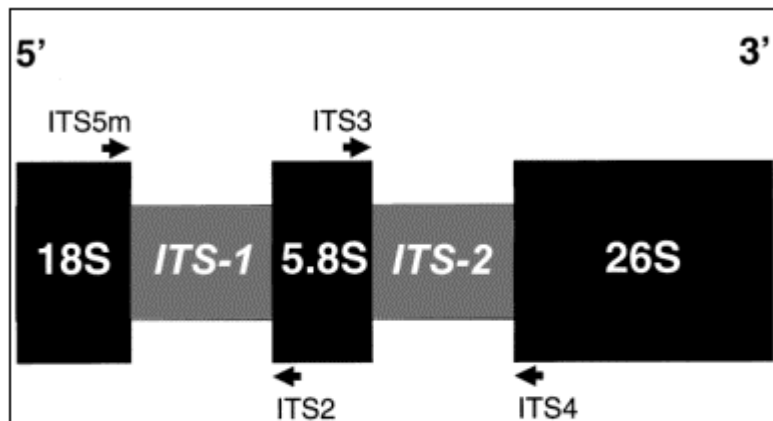
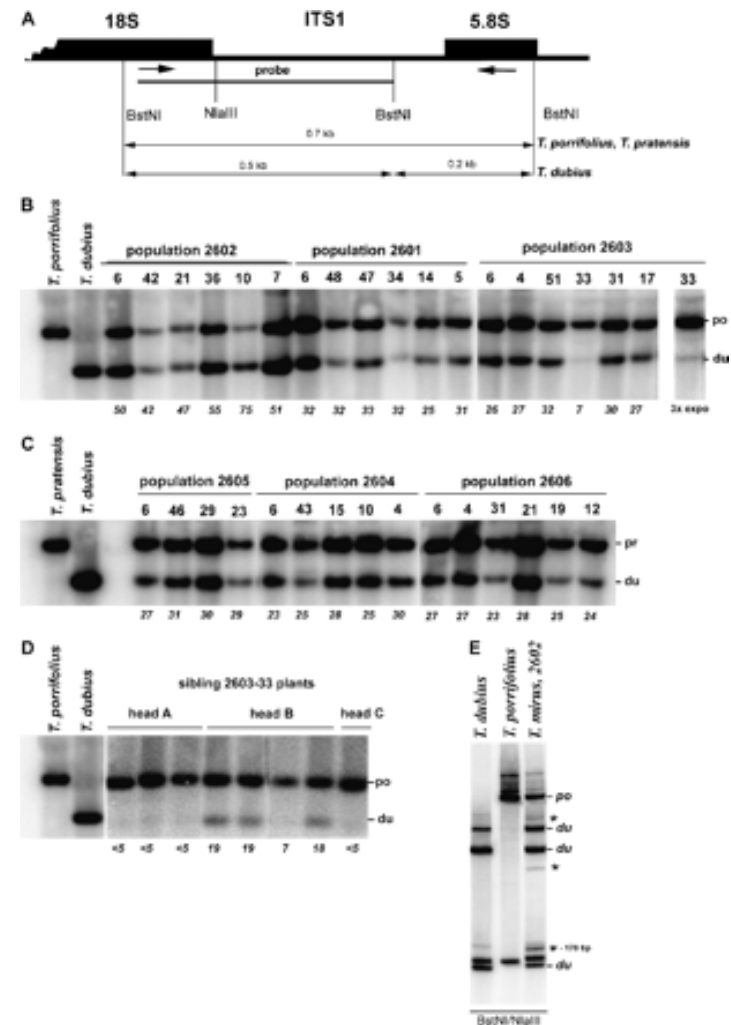


Fig. 1. The three coding and two internal transcribed spacer regions of the nuclear ribosomal DNA repeat unit of a typical angiosperm (not drawn to scale). Arrows indicate approximate locations of the four primers used for PCR amplification.

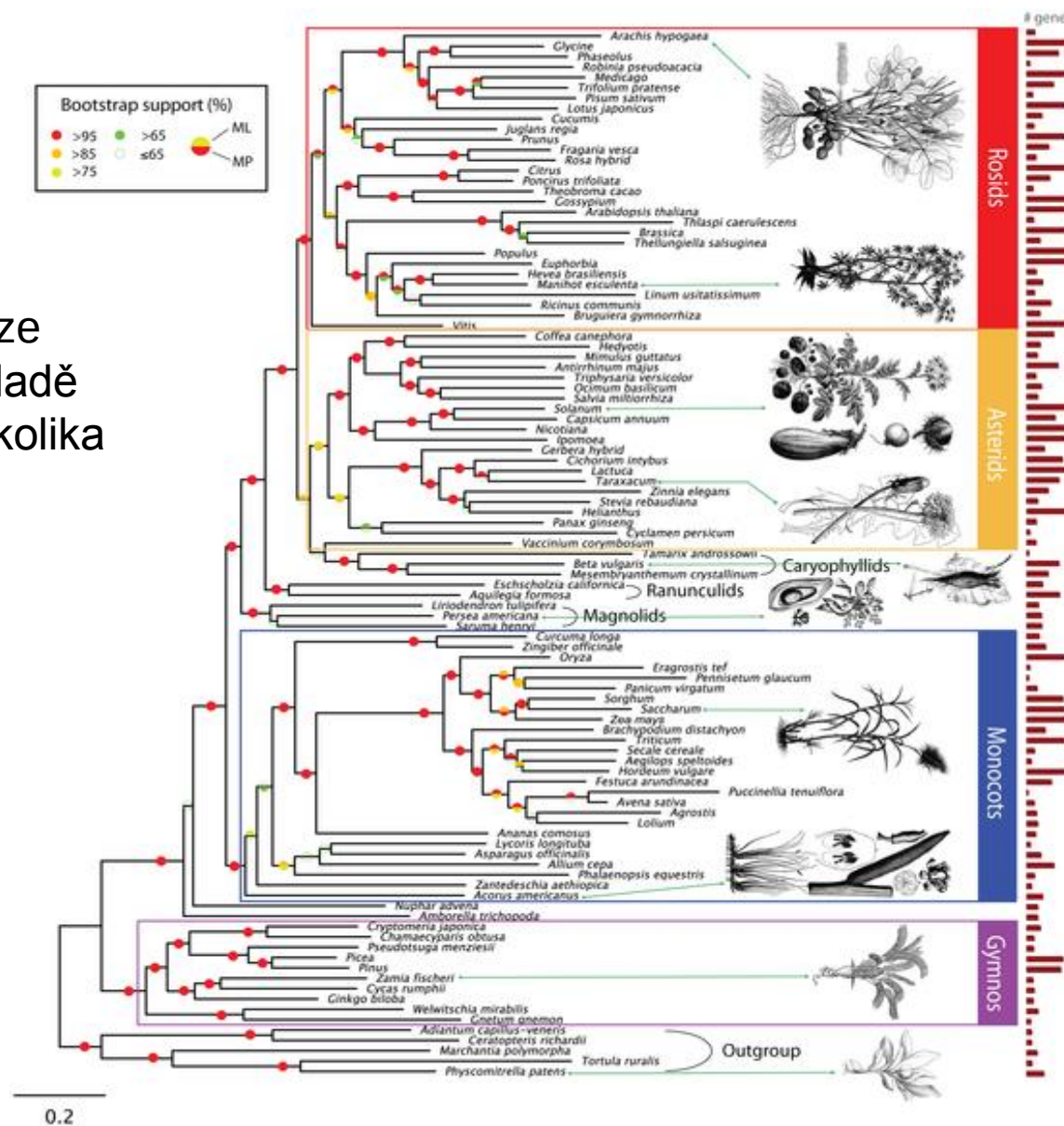
Výhody :

- Univerzálnost
- Možnost detekce hybridizace
- Dostatečná variabilita i uvnitř druhu

Uznáný v rámci barcoding projektu



Fylogenezi nelze odvodit na základě analýzy jen několika málo genů.



Lee EK, Cibrian-Jaramillo A, Kolokotronis S-O, Katari MS, et al. (2011) A Functional Phylogenomic View of the Seed Plants. PLoS Genet 7(12): e1002411. doi:10.1371/journal.pgen.1002411
<http://www.plosgenetics.org/article/info:doi/10.1371/journal.pgen.1002411>

Genome-Scale Phylogenetics: Inferring the Plant Tree of Life from 18,896 Gene Trees

J. GORDON BURLEIGH^{1,2,*}, MUKUL S. BANSAL^{3,4}, OLIVER EULENSTEIN³, STEFANIE HARTMANN^{5,6},
ANDRÉ WEHE³, AND TODD J. VISION^{2,5}

