

Hodnocení multilokusových molekulárních dat II

Tomáš Fér & Filip Kolář
Katedra Botaniky PřF UK, Praha

Multilokusová data

- **dominantní** – nejsme schopni odlišit heterozygoty od dominantních homozygotů
- **binární** – bialelická povaha lokusu (fragmentu)
 - přítomnost (dominantní alela/heterozygot)
 - nepřítomnost (recesivní alela)
 - tj. skórování 0-1
- **anonymní** – nevíme z jaké části genomu pocházejí
- **multilokusová** – často zároveň analyzujeme stovky lokusů, tj. analýza pokrývá „celý genom“
- RAPD, AFLP, ISSR...
- **kodominantní** – odlišíme homo- a heterozygoty, tj. detekujeme všechny alely
- **alelická** – známe frekvence alel v lokusech, v populacích...
- **anonymní** – nevíme z jaké části genomu pocházejí
- **multilokusová** – nejčastěji analýza malého množství lokusů (5-20)
- mikrosatelity (SSRs), isozymy

Možnosti analýzy

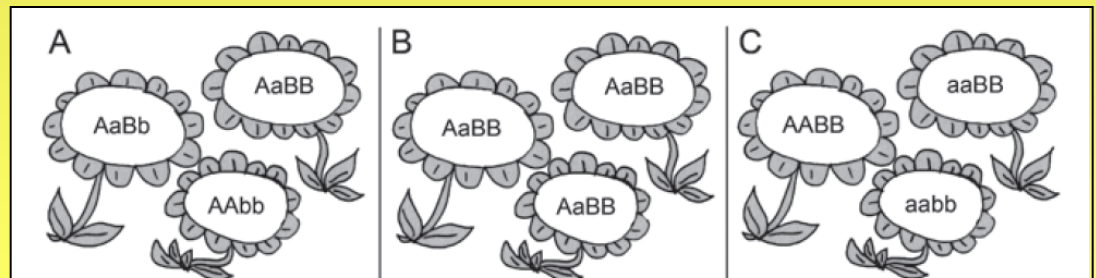
- vztahy mezi jedinci – základní orientace ve struktuře
 - distanční stromy (NJ, UPGMA), sítě
 - mnohorozměrné analýzy (PCoA)
 - Bayesovské clusterování
- populačně-genetické parametry
 - diverzita (% polymorfních fragmentů, index diverzity)
 - divergence (% unikátních fragmentů, DW-index)
 - F-statistika, R-statistika (mikrosatelity)
- testování a zjišťování prostorové struktury
 - AMOVA
 - F-statistika (odhady diferenciace), Mantelovy testy, prostorová autokorelace
- testování specifických hypotéz
 - podobnost a evoluční vztahy identifikovaných skupin
 - hybridizace
 - původ polyploidů
 - ...

Možnosti analýzy

- vztahy mezi jedinci – základní orientace ve struktuře
 - distanční stromy (NJ, UPGMA), sítě
 - mnohorozměrné analýzy (PCoA)
 - Bayesovské clusterování
- populačně-genetické parametry
 - diverzita (% polymorfních fragmentů, index diverzity)
 - divergence (% unikátních fragmentů, DW-index)
 - F-statistika, R-statistika (mikrosatelity)
- testování a zjišťování prostorové struktury
 - AMOVA
 - F-statistika (odhady diferenciace), Mantelovy testy, prostorová autokorelace
- testování specifických hypotéz
 - podobnost a evoluční vztahy identifikovaných skupin
 - hybridizace
 - původ polyploidů
 - ...

Populační data

- hierarchická struktura
 - frekvence alel v populacích, ve skupinách populací
- genetická diverzita populací
- proporce vzácných alel
- Hardy-Weinbergova rovnováha (equilibrium; HWE) a odchylky od něj
 - selekce
 - migrace
 - genetický drift
 - reprodukční systém, inbreeding, ...

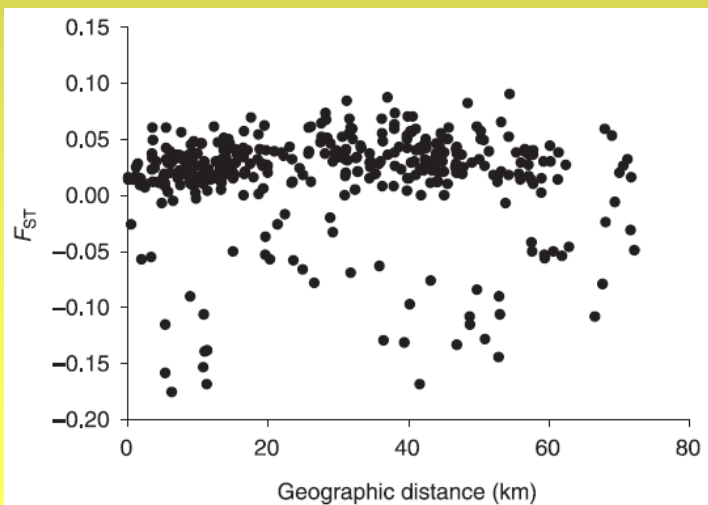
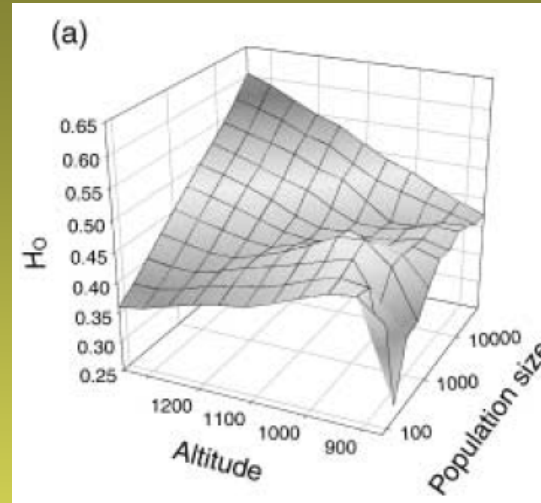
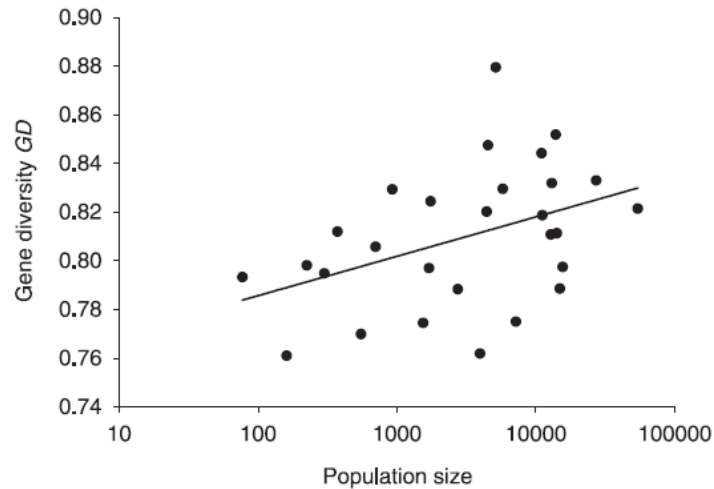


Obr. 3.26: Příklad rozložení variability u dvou genů (s alelami A, a a alelami B, b) v populacích rozmnožujících se různými strategiemi. A – pohlavní rozmnožování, B – nepohlavní rozmnožování, C – samosprašení.

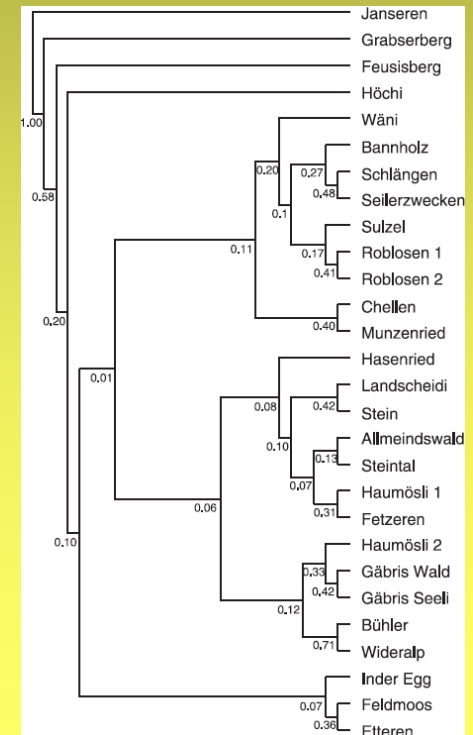
Populační diverzita – SSRs

- počet alel, allelic richness, privátní alely
- H_E , H_O , F_{IS} , gene diversity
- odchylky od HWE, deficience heterozygotů
- vztah k velikosti populace, geografické poloze apod.
- rozmnožovací systém – inbreeding, outcrossing rate
- identifikace bottlenecku (méně alel, žádné privátní alely...)

Lychnis flos-cuculi ve Švýcarsku



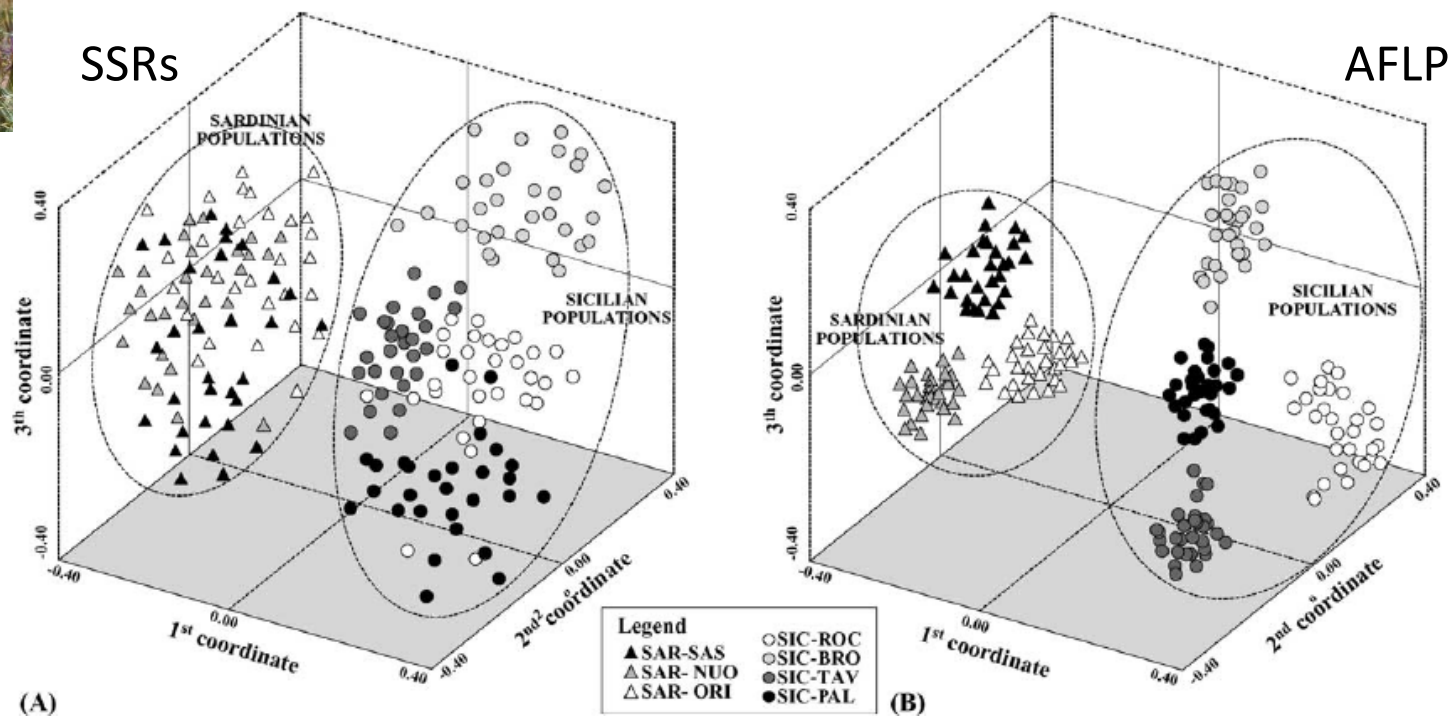
Source of variation	Variance component			
	d.f.	Absolute	%	P
Among regions	2	0.0147	0.68	< 0.001
Among populations within regions	25	0.0321	1.49	< 0.001
Within populations	906	2.0987	97.83	< 0.001
Total	933	2.1455	100.00	



Galeuchet et al. 2005

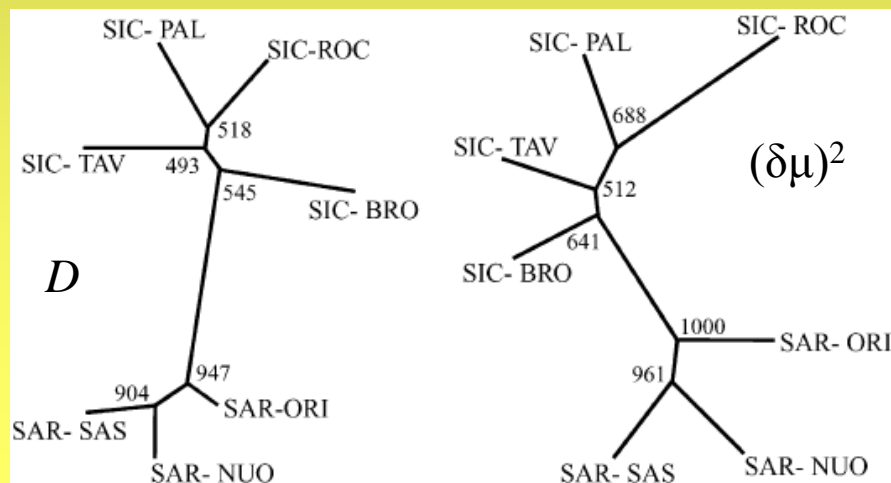


Cynara cardunculus – 5 lokusů



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*

* $P < 0.0001$.



Populační diverzita a divergence - AFLP

- **genetická diverzita** (jak moc je populace diverzifikovaná, rozdíly mezi jedinci v rámci populace)
 - % polymorfních fragmentů
 - Shannonův index diverzity / average gene diversity (Nei)
 - reflektuje reprodukční systém, recentní procesy (genový tok, velikost populace...)
 - software – POPGENE, Arlequin, AFLPdat...
- **divergence – vzácnost (*rarity*)** (jak moc se populace odlišuje od ostatních, jak moc je unikátní)
 - počet (nebo %) vzácných fragmentů (arbitrární definice vzácnosti)
 - počet (nebo %) unikátních (privátních) fragmentů
 - DW-index (*frequency down-weighted marker values*)
 - reflektuje historické procesy (dlouhodobá izolovanost...)
 - software – AFLPdat...

Shannonův index diverzity

$$H_{Sh} = - \sum_{i=1}^k p_i \ln p_i$$

p_i – frekvence i -tého fragmentu



Saponaria pumila, AFLP

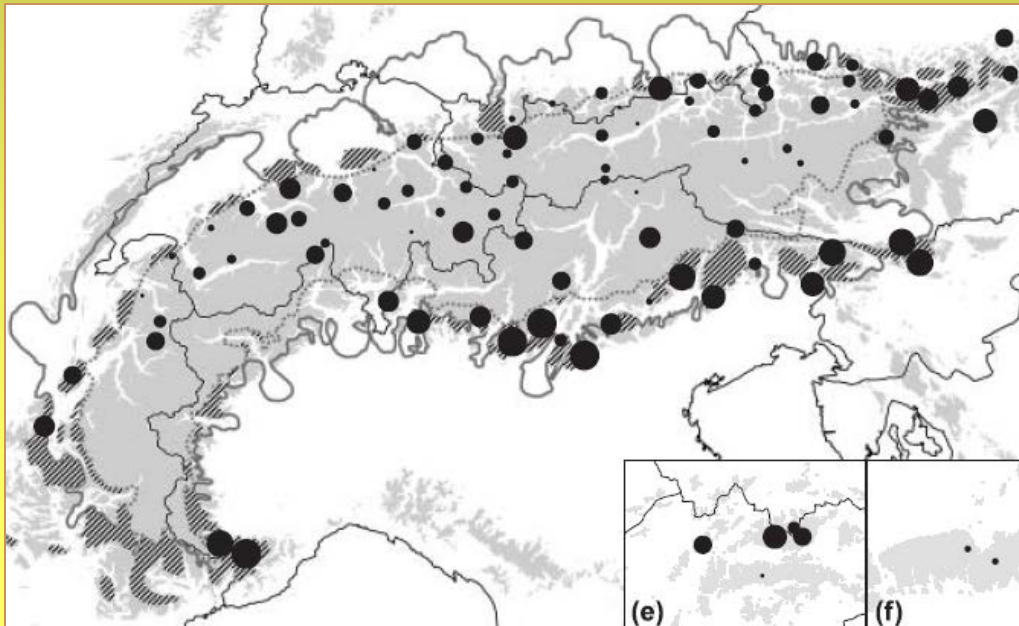
Velikost H_{Sh} ve vztahu k
maximálnímu zalednění
na území Alp

(Tribsch et al. 2002)

Divergence – DW-index

frequency down-weighted marker values

- počet výskytů AFLP markeru v populaci/počet výskytů markeru v celém datovém souboru → součet hodnot pro všechny markery
- vyšší hodnoty v dlouhodobě izolovaných populacích (akumulace díky mutacím)
- nižší hodnoty v nově vzniklých populacích (recentní šíření)

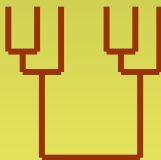





Ranunculus alpestris, AFLP

distribuce DW-indexu ve
vztahu k maximálnímu
zalednění na území Alp

(Paun et al. 2008)

Diverzita, divergence vs. populační historie

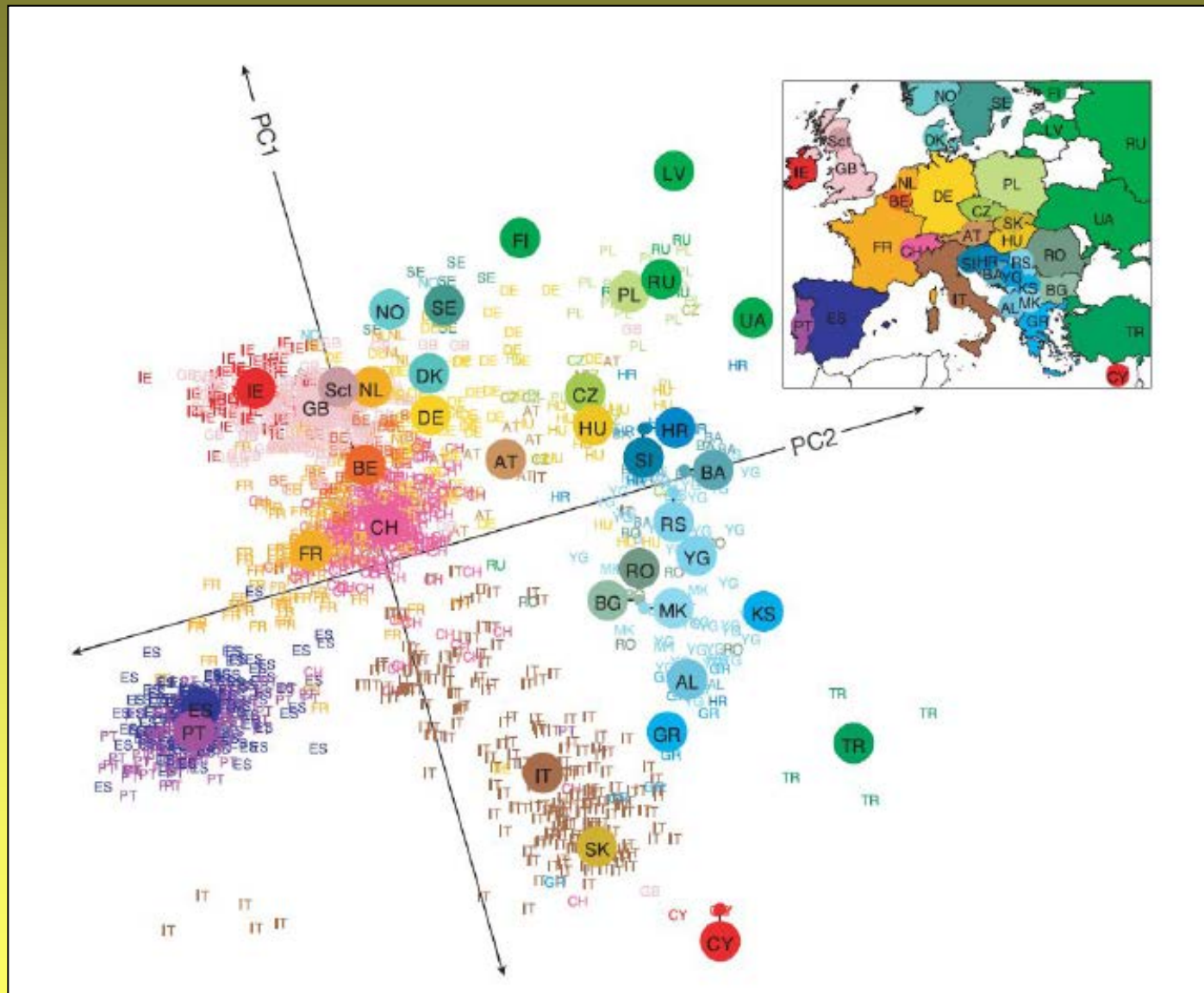
	vysoká genetická diverzita	nízká genetická diverzita
vysoká divergence (např. privátní & vzácné markery, vysoké DW)	omezený genový tok, přežívání <i>in situ</i> ve velkém refugiu 	bottleneck díky malým, izolovaným refugiálním populacím, přežívání <i>in situ</i> v refugiu 
nízká divergence	vysoký genový tok, imigrace 	bottleneck díky dálkovému šíření, imigrace 

podle Tribsch et al.

Možnosti analýzy

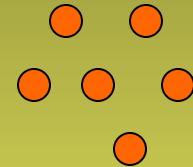
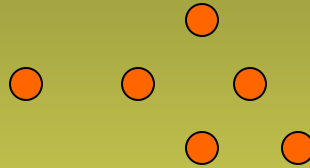
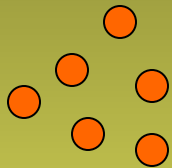
- vztahy mezi jedinci – základní orientace ve struktuře
 - distanční stromy (NJ, UPGMA), sítě
 - mnohorozměrné analýzy (PCoA)
 - Bayesovské clusterování
- populačně-genetické parametry
 - diverzita (% polymorfních fragmentů, index diverzity)
 - divergence (% unikátních fragmentů, DW-index)
 - F-statistika, R-statistika (mikrosatelity)
- testování a zjišťování prostorové struktury
 - AMOVA
 - F-statistika (odhady diferenciace), Mantelovy testy, prostorová autokorelace
- testování specifických hypotéz
 - podobnost a evoluční vztahy identifikovaných skupin
 - hybridizace
 - původ polyploidů
 - ...

Prostor a genetika



PCO z 500 000 SNPů na lidech,
Novembre et al, 2008, Nature

Testování a zjišťování prostorové struktury



- F-statistika, odhady diferenciace
- analýza molekulární variance (AMOVA)
- Mantelovy testy (běžný, regionální...)
- prostorová autokorelace

Populační diferenciace (F statistiky)

- mikrosatelity – alelické vyhodnocení
 - F-statistika (F_{ST} , F_{IS}), odchylky od H-W rovnováhy (test)...
 - 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} , Rho (Rousset 1996)
 - vzdálenosti
 - delta mu – $(\delta\mu)^2$, D_{dm} (Goldstein et al. 1995)
 - D_{sw} – stepwise weighted genetic distance
- software
- MICROSAT (Minch 1996), RSTcalc
 - MSA – Microsatellite Analyser (Dieringer & Schlötterer 2003)
- AFLP – dominantní marker, pouze nepřímé odhady F-statistiky
 - např. Bayesovský model k odhadu F_{ST} (ϑ – theta) a F_{IS} (f – inbreeding coefficient)
 - odhad F_{ST} srovnatelný s AMOVA, odhad F_{IS} velmi hrubý s širokým intervalem spolehlivosti
- software
- HICKORY (Holsinger et al. 2002)

Analýza molekulární variance (AMOVA)

- testování populačně genetické struktury
- rozdělení celkové variance na (ko)varianční komponenty skupiny, populace a jedince
- kolik procent celkové variability je
 - mezi jedinci v rámci populace
 - mezi populacemi
 - mezi skupinami populací ...
- výpočet F-statistik (F_{ST} analog)
- software
 - Arlequin 3.5 (Schneider, Roessli & Excoffier 2000)
 - FAMD (Schlüter & Harris 2006)

AMOVA (Arlequin 3.5)



Distance method: Pairwise difference

AMOVA design and results :

Reference: Weir, B.S. and Cockerham, C.C. 1984.
Excoffier, L., Smouse, P., and Quattro, J. 1992.
Weir, B. S., 1996.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	1	147.449	1.22895 Va	7.28
Among populations within groups	64	2288.706	7.00306 Vb	41.51
Within populations	192	1658.710	8.63911 Vc	51.21
Total	257	4094.864	16.87112	

Fixation Indices

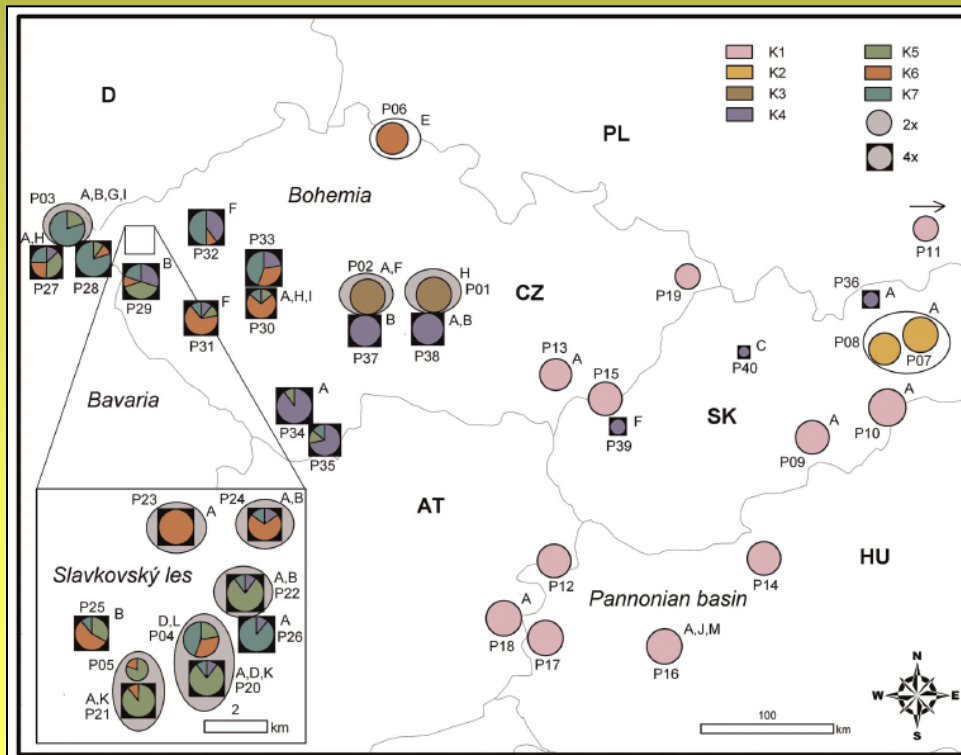
FSC : 0.44770

FST : 0.48793

FCT : 0.07284

AMOVA

- strukturování genetické diverzity např.
 - 1) celková hierarchická struktura v datech (celek/druh/pop. v rámci druhu)
 - 2) mezipop. diferenciace v subsetu populací např. v různých oblastech/v rámci cytotypu/...
 - 3) mezi druhy/regiony/cytotypy

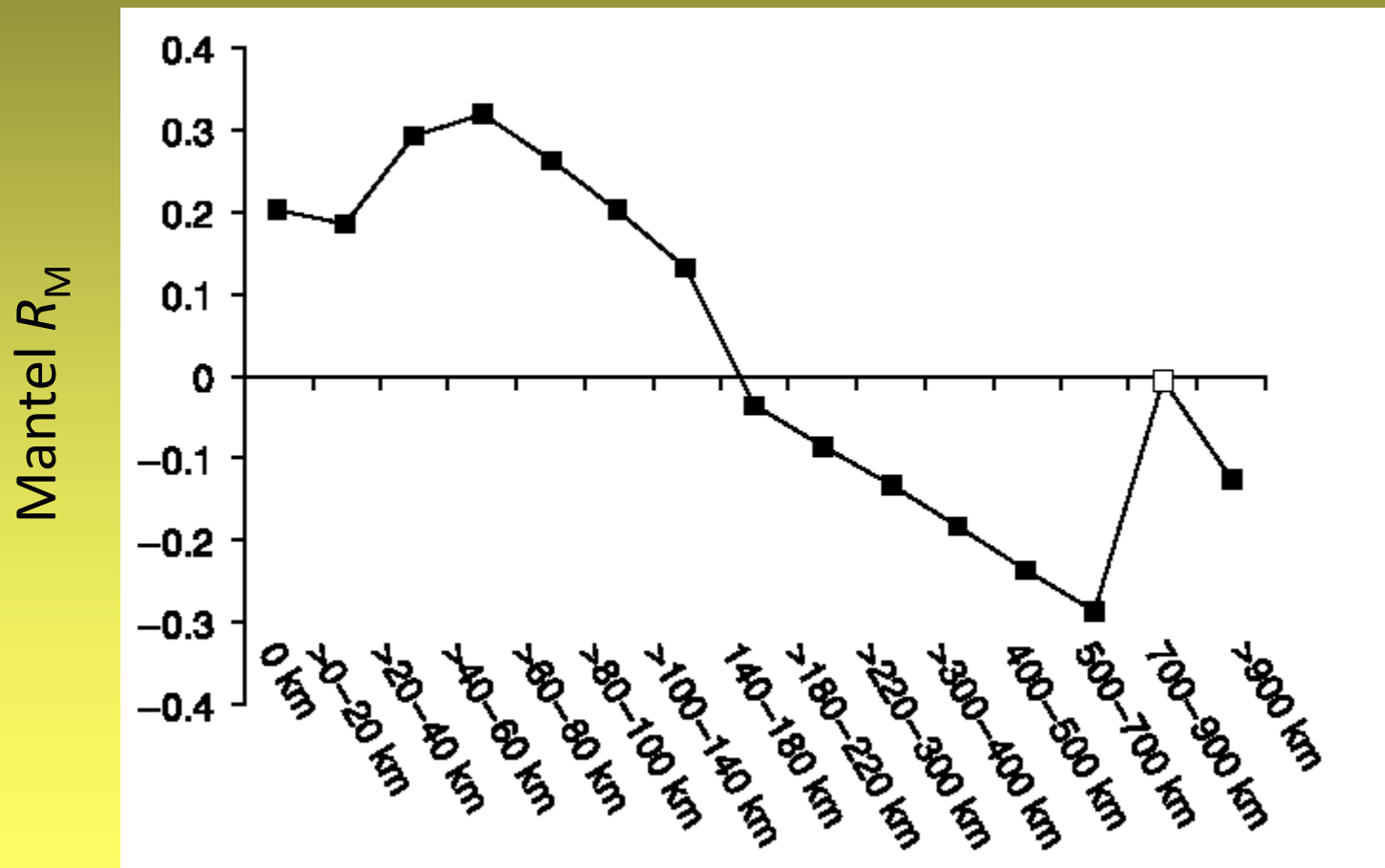


	d.f.	% of variation	Fst ^a
A. Complete dataset			
Among all populations	38	37.1	0.371
Within populations	296	62.9	
Species grouping			
Among species*	2	4.0	0.396
Among populations within species	34	35.6	
Within populations	282	60.4	
Ploidy level grouping			
Among all 2x vs. 4x	1	18.9	0.429
Among populations within groups	37	24.0	
Within populations	296	57.1	
Genome size grouping			
Among relict 2x vs. non-relict 2x vs. 4x	2	27.5	0.434
Among populations within groups	36	15.9	
Within populations	296	56.6	
Among populations of relict 2x	7	30.5	0.305
Within populations	62	69.5	
Among populations of non-relict 2x	10	14.3	0.143
Within populations	84	85.7	
Among populations of 4x	21	24.8	0.248
Within populations	157	75.2	
B. Only Slavkovský les area			
Among all populations in Slavkovský les	8	22.9	0.229
Within populations	73	77.1	
Among 2x vs. 4x in Slavkovský les	1	0.7	0.233
Among populations within groups	7	22.6	
Within populations	73	76.7	

Mantelův test

- klasický – korelace dvou matic vzdáleností (genetické vs. geografické)
 - porovnání dvou matic \rightarrow Mantel R_M
 - permutační test \rightarrow signifikance korelace
- korelace matice genetických vzdáleností a matice tzv. distančních tříd
 - série Mantelových testů \rightarrow Mantelův korelogram
- software – zt, Arlequin, R

Mantel correlogram



Phyteuma globulariifolium v Alpách (Schönswetter *et al.* 2002)

Prostorová autokorelace

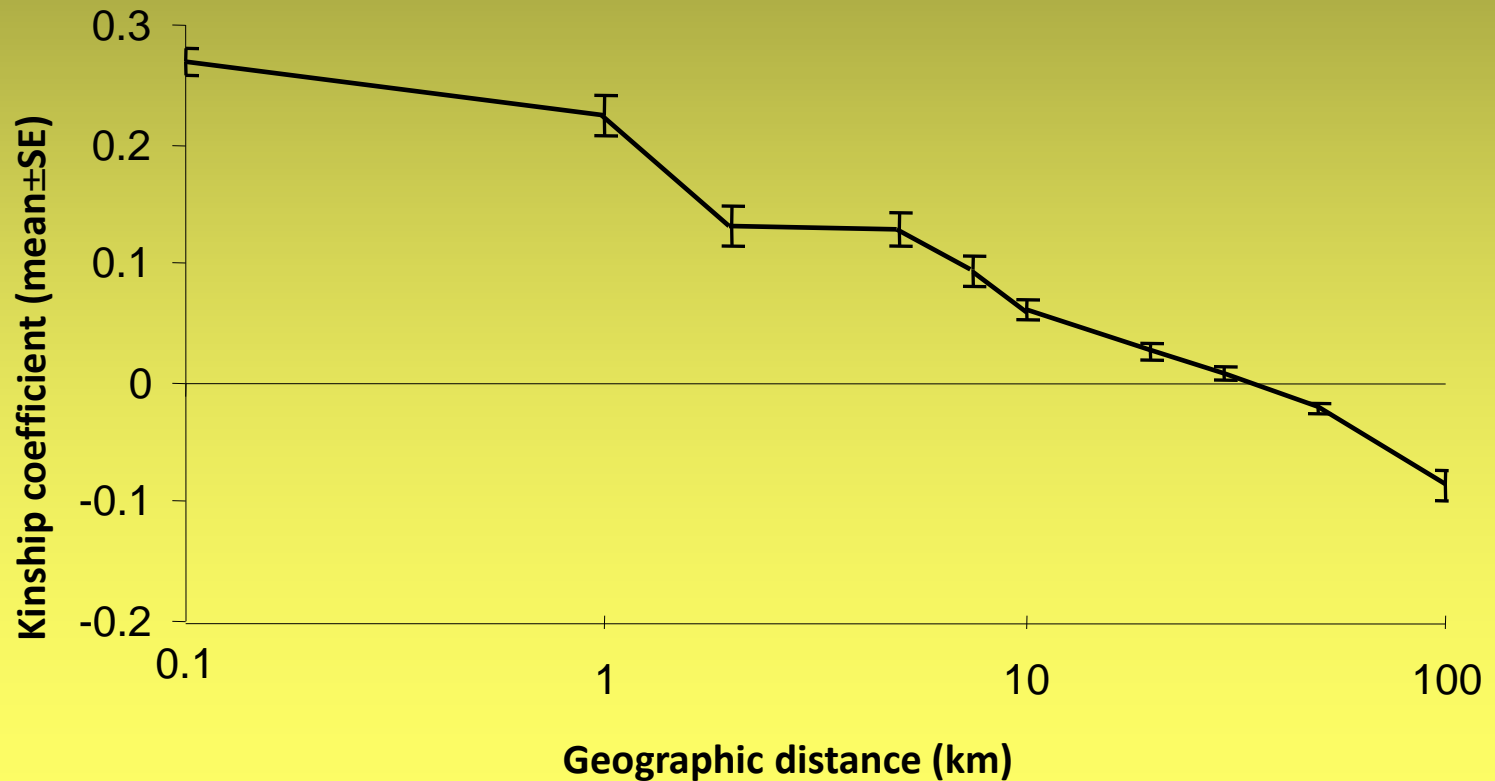
SPAGeDi 1.3

Spatial Pattern Analysis of Genetic Diversity

<http://ebe.ulb.ac.be/ebe/Software.html>

- výpočet *kinship koeficientů* mezi páry jedinců, tj. pravděpodobnost, že dva geny náhodně vytažené z každého jedince jsou identické
- kinship koeficient pro dominantní data (Hardy 2003)
- analýza vztahu mezi těmito hodnotami a geografickými vzdálenostmi
- permutace – testování prostorové struktury

Změna korelačního koeficientu se vzdáleností



Bayesovské analýzy pop. struktury

- Využití geografické polohy jako PRIOR

- likelihood u všech programů stejný (odvozený od Hardy-Weinberg equilibrium, HWE) – každý cluster v HWE + různé frekvence alel mezi clustery)

- zde ale navíc **prior distribution** – něco, co dopředu volíme; apriorní informace „injikovaná“ do modelu

- umožňuje efektivnější MCMC hledání v parametrovém prostoru

- nejlépe na malé datasety (ne mnoho lokusů) s rovnoměrně sebranými jedinci

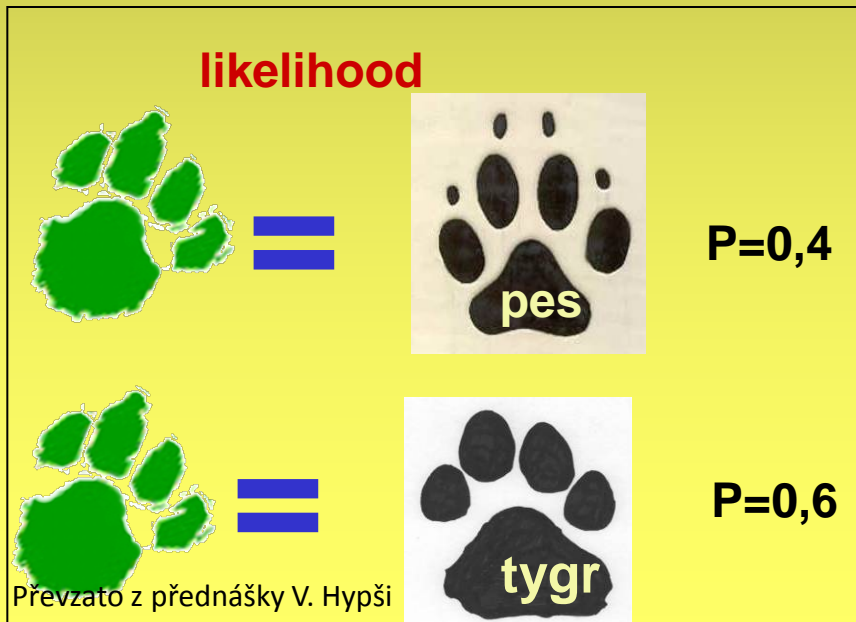
Programy

➤ Structure

➤ BAPS

➤ Geneland

likelihood



P=0,4

P=0,6

tygr

pes

Převzato z přednášky V. Hypši

prior probability

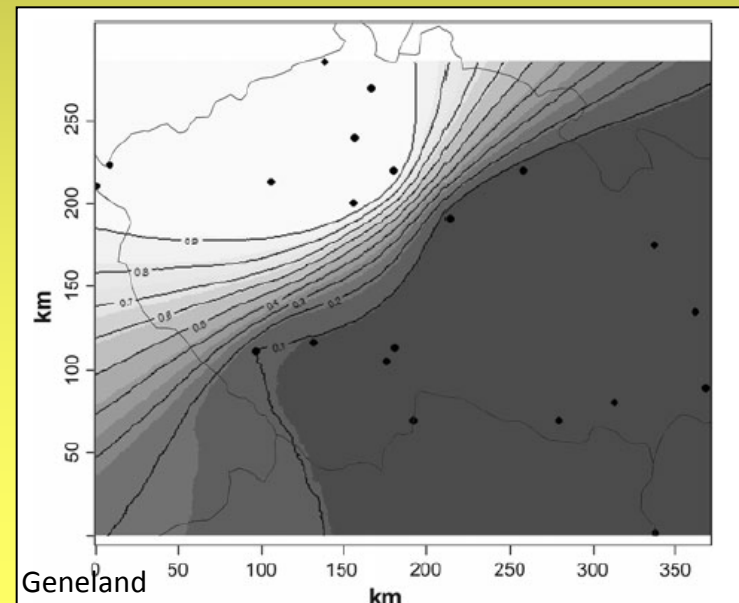
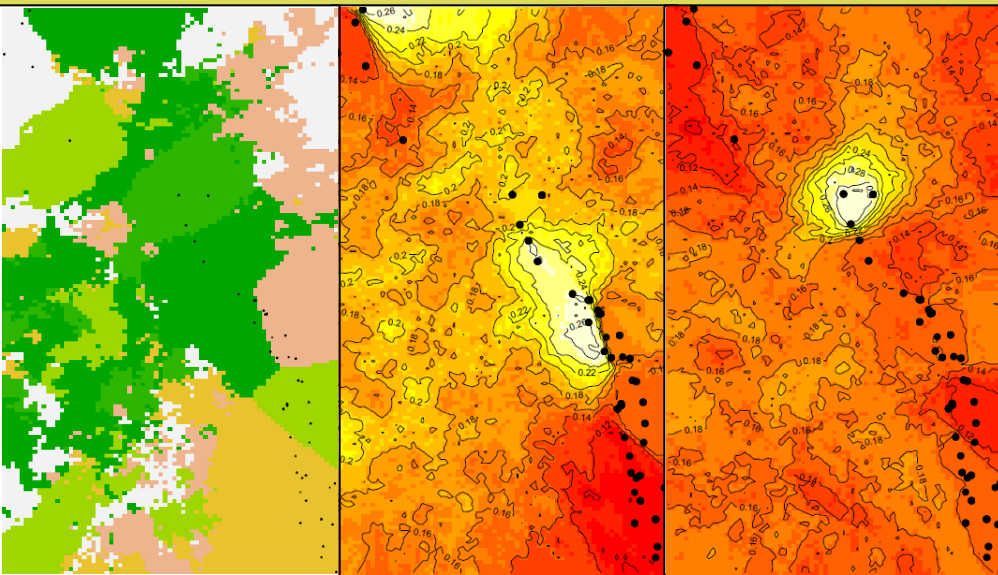
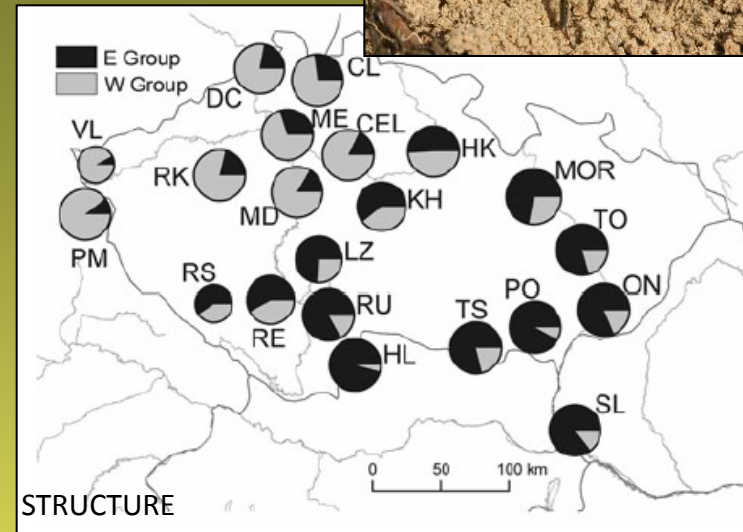
P=0,95

P=0,05



Geneland


- mapa posterior probability příslušnosti k jedné ze skupin
- MCMC + výstupy v R
- možná i non-spatial analýza, admixture model pro hybridní zóny
- dobré na prostorově spojitá a geneticky málo diferencovaná data





Geneland


Geneland - Graphical Interface Menu Tools Help


-Inference-

Data Files 


Run 


Postprocess 

Admixture & Hybrid zone 

Output 

-Simulation-

Panmictic data 

Output 

Maximum rate of poisson process: 100

Uncertainty on coordinate: 0

pop min pop init pop max

Number of populations: 1 10

Maximum number of nodes: 300


Number of iterations: 100000

Thinning: 100


Allele frequency model: Correlated

Spatial model: TRUE

Null allele model: FALSE



Multiple independent runs: No Yes 1

Run 

Coordinate: File loaded

Genotype: Dominant file loaded

Population graph network

- stanovení signifikantní genetické kovariance mezi populacemi (testována každá dvojice po odstranění vlivu ostatních pop.)
- spoje testovány bootstrapováním
- geografické vzdálenosti spoju: vikariance vs. LDD
- **POPGRAPHS**
 - Dyer and Nason (2004)
 - dyerlab.bio.vcu.edu/software/

testy spoju mezi regiony (Escobar García et al. 2012 Mol Ecol)

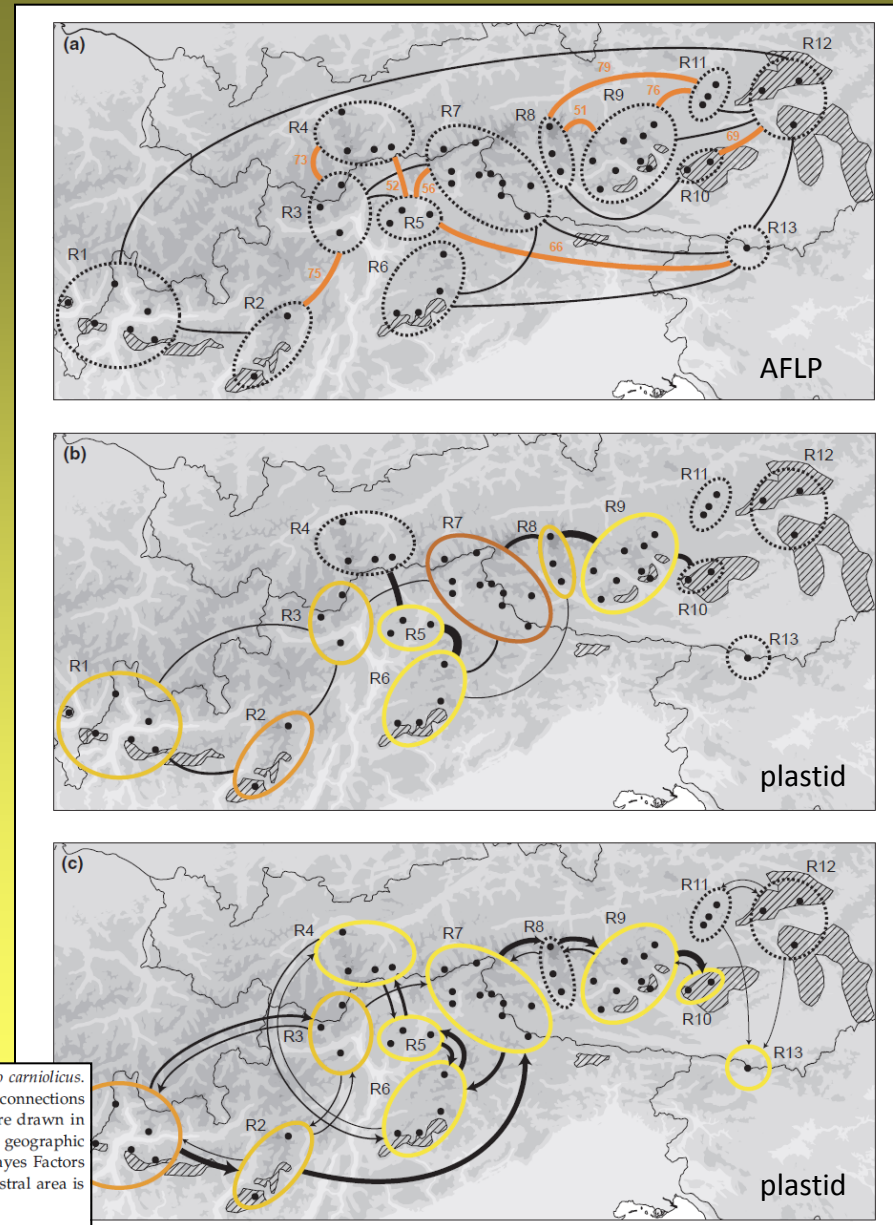


Fig. 3 Range connectivity among discrete geographic regions (their designations R1–R13 are indicated) in diploid *Senecio carnifolius*. (a) Population Graphs network illustrating the genetic covariance structure among regions based on AFLP fingerprints; connections between regions indicated in orange are stable edges (with a bootstrap support >50% given at the edges), other edges are drawn in black. (b–c) Range connectivity inferred from plastid sequence data using (b) reversible and (c) nonreversible models of geographic diffusion between discrete geographic regions, the thickness of the connections being proportional to their support by Bayes Factors (see text for details); decreasing posterior probability (shown in four classes from 0.05 to 0.19) of a region to be the ancestral area is indicated by increasingly lighter colours. Hatched areas in (a) to (c) as in Fig. 1.

Praktické cvičení

1. AMOVA

- AMOVA
 - SSRs data – z programu MSA vygenerovat soubor ve formátu pro Arlequin (případně ho vytvořit konverzí pomocí PGDSpider), v něm spočítat AMOVA pro Fst-like a Rst-like a porovnat
 - AFLP data – FAMD (dataset: Galium_data.xls) - nutné si nadefinovat skupiny (BI, CE, N) v Group manager
 - AFLPdata 2. možnost: Arlequin (dataset: Gal2x.arp nutné ručně dopravit – nadefinovat skupiny BI, CE – viz vzor Gal2x_2groups.arp

2. Geneland

- analyzujte diploidní Galium AFLP data v programu Geneland (R-package)
 - dataset = Galium_data_geneland.xls – z něj si vytvořte požadované vstupní txt soubory: soubor s 0/1 daty a soubor s koordinátami
- zvolte K 1-5, 100 000 opakování, *correlated allele frequency* model

3. Populační charakteristiky

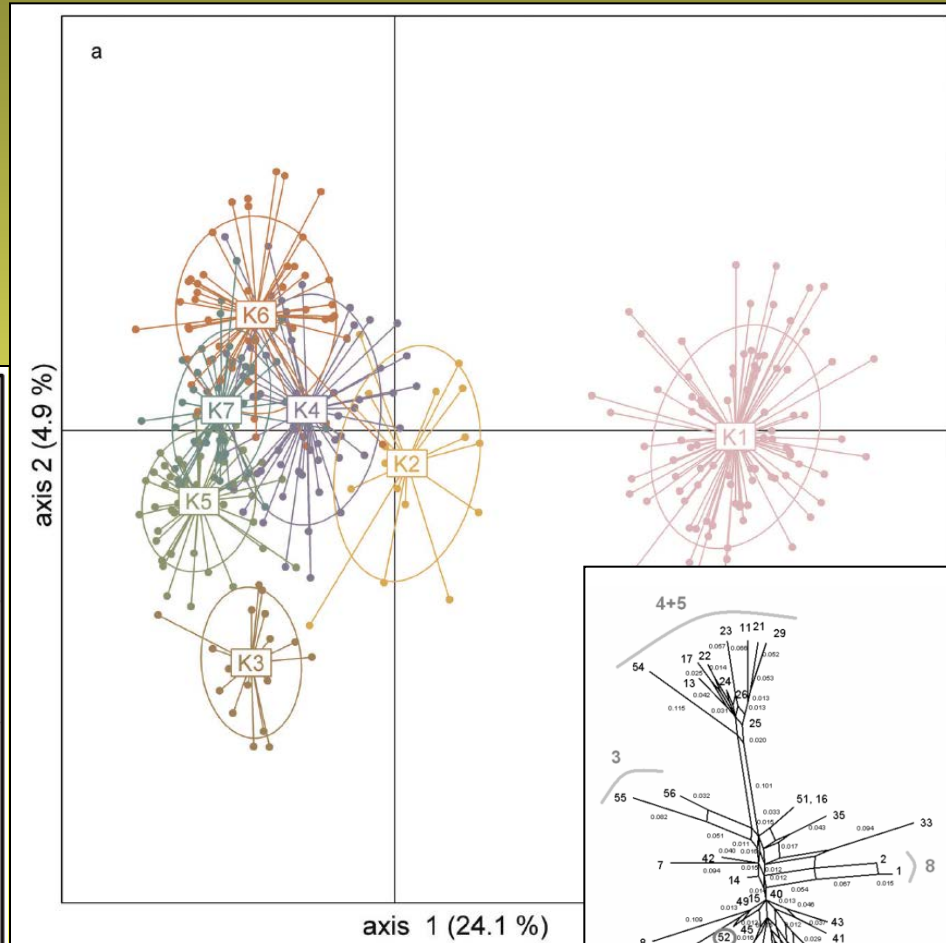
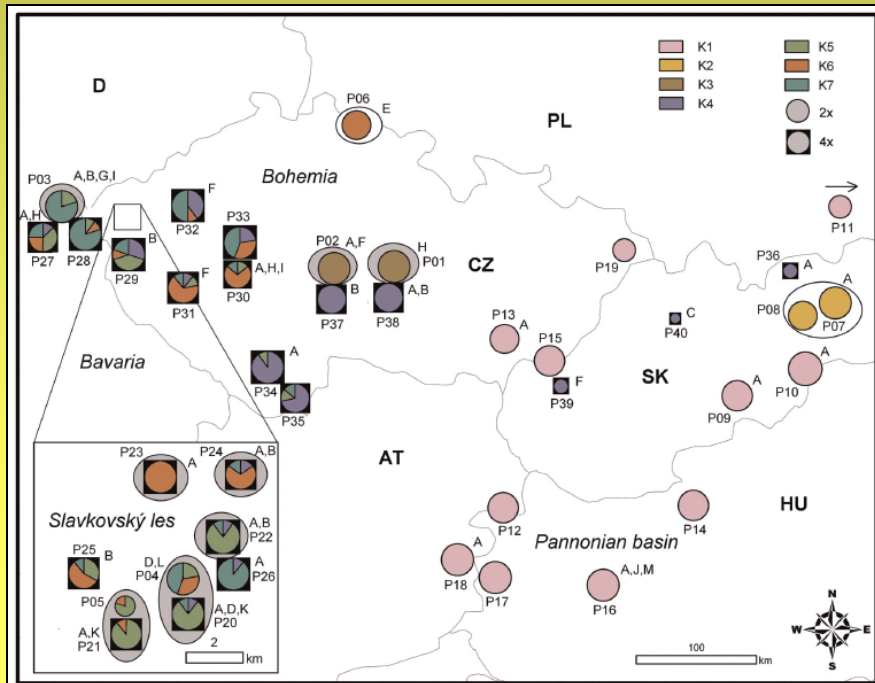
- v programu MSA nebo FSTAT (pro dataset Typha) spočítejte vnitropopulační diverzity, koeficient inbreedingu a zjistěte, jestli je signifikantní
- populace = druhy

Možnosti analýzy

- vztahy mezi jedinci – základní orientace ve struktuře
 - distanční stromy (NJ, UPGMA), sítě
 - mnohorozměrné analýzy (PCoA)
 - Bayesovské clusterování
- populačně-genetické parametry
 - diverzita (% polymorfních fragmentů, index diverzity)
 - divergence (% unikátních fragmentů, DW-index)
 - F-statistika, R-statistika (mikrosatelity)
- testování a zjišťování prostorové struktury
 - AMOVA
 - F-statistika (odhady diferenciace), Mantelovy testy, prostorová autokorelace
- testování specifických hypotéz
 - podobnost a evoluční vztahy identifikovaných skupin
 - hybridizace
 - původ polyploidů
 - ...

Vztahy mezi skupinami

- které detekované skupiny si jsou podobnější?
 - Arlequin – porovnání párových F_{st} skupin mezi populacemi/clustery
 - BAPS : K-L divergence matrix
 - PCO
 -



Kolář et al. 2012 PLoS ONE

Vztahy mezi skupinami

- Assignment testy

- předem definované skupiny („zdrojové populace), zařazují neznámé jedince + testování pomocí simulací
- rozhodování podle frekvence proužků v polymorfních lokusech
- např. program AFLPOP (Duchesne and Bernatchez 2002)

Table 3 Proportion of assignments of individuals of *Vaccinium uliginosum* to geographical groups inferred by STRUCTURE analyses of AFLP data using AFLPOP (Duchesne & Bernatchez 2002). Assignment of a plant to its group of origin was not allowed. Stringency level for assignment was set to log-likelihood difference > 2. D, average population diversity within groups (14 populations represented by single individuals were excluded). Abbreviations of main group names: Beringian/northern Canadian, Ber/N Can; east Canadian/Greenlandic, E Can/Grl; north European, N Eur; south European, S Eur

Assigned to Assigned from	Siberian	Ber/N Can (Ber E)	Ber/N Can (Arctic)	Ber/N Can (Ber W)	E Can/ Grl (South)	E Can/ Grl (East)	E Can/ Grl (Central)	N Eur (West)	N Eur (South)	N Eur (East)	S Eur (Caucasus)	S Eur (East)	S Eur (West)	Not assigned	No. of plants	D ± SD
Siberian	—	0.02	0.02	—	—	0.34	—	—	—	0.04	—	—	—	0.58	126	0.154 ± 0.051
Ber/N Can (Ber E)	—	—	0.77	0.05	—	—	—	—	—	—	—	—	—	0.18	39	0.155 ± 0.063
Ber/N Can (Arctic)	0.03	0.10	—	0.07	—	0.01	0.04	—	—	—	—	—	—	0.76	113	0.143 ± 0.032
Ber/N Can (Ber W)	0.12	0.03	0.16	—	0.01	—	—	—	—	—	—	—	—	0.67	73	0.154 ± 0.021
E Can/Grl (South)	0.02	—	0.03	0.02	—	—	0.52	—	—	—	—	—	—	0.42	60	0.160 ± 0.015
E Can/Grl (East)	0.31	—	—	—	0.09	—	0.04	—	—	—	—	—	—	0.56	75	0.167 ± 0.040
E Can/Grl (Central)	—	—	0.05	—	0.29	0.03	—	—	—	—	—	—	—	0.63	76	0.150 ± 0.021
N Eur (West)	—	—	—	—	—	—	—	—	0.01	0.71	—	—	—	0.29	168	0.140 ± 0.018
N Eur (South)	—	—	—	—	—	—	—	0.12	—	—	—	—	—	0.88	43	0.100 ± 0.013
N Eur (East)	—	0.01	—	—	—	—	—	0.74	—	—	—	—	—	0.25	89	0.166 ± 0.021
S Eur (Caucasus)	—	—	0.18	—	—	—	—	—	0.27	—	—	—	—	0.55	11	0.055*
S Eur (East)	—	—	—	—	—	—	—	—	—	0.13	—	—	0.08	0.79	48	0.100 ± 0.034
S Eur (West)	—	—	—	—	—	—	—	0.18	—	—	—	0.21	—	0.61	28	0.101 ± 0.016

Hybridizace

<i>T. latifolia</i>	176	176	278	278	176	190	269	269	179	179	93	93	278	278
<i>T. angustifolia</i>	210	210	286	286	196	196	287	287	193	193	101	101	280	280
<i>T. x glauca</i>	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

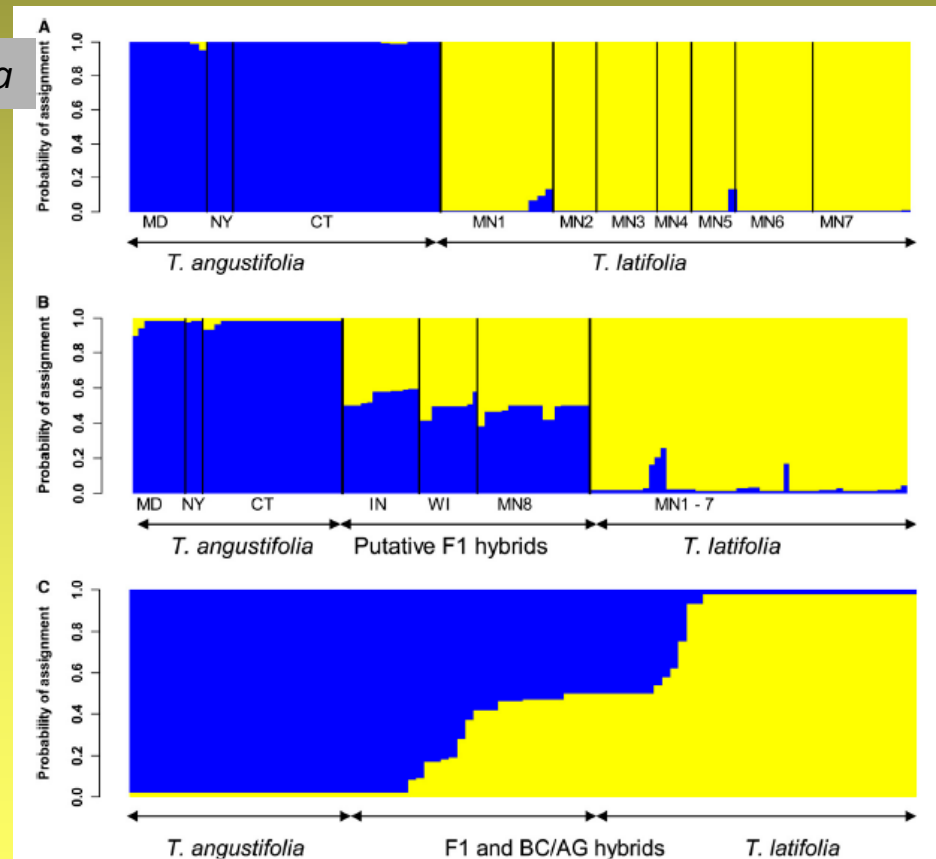
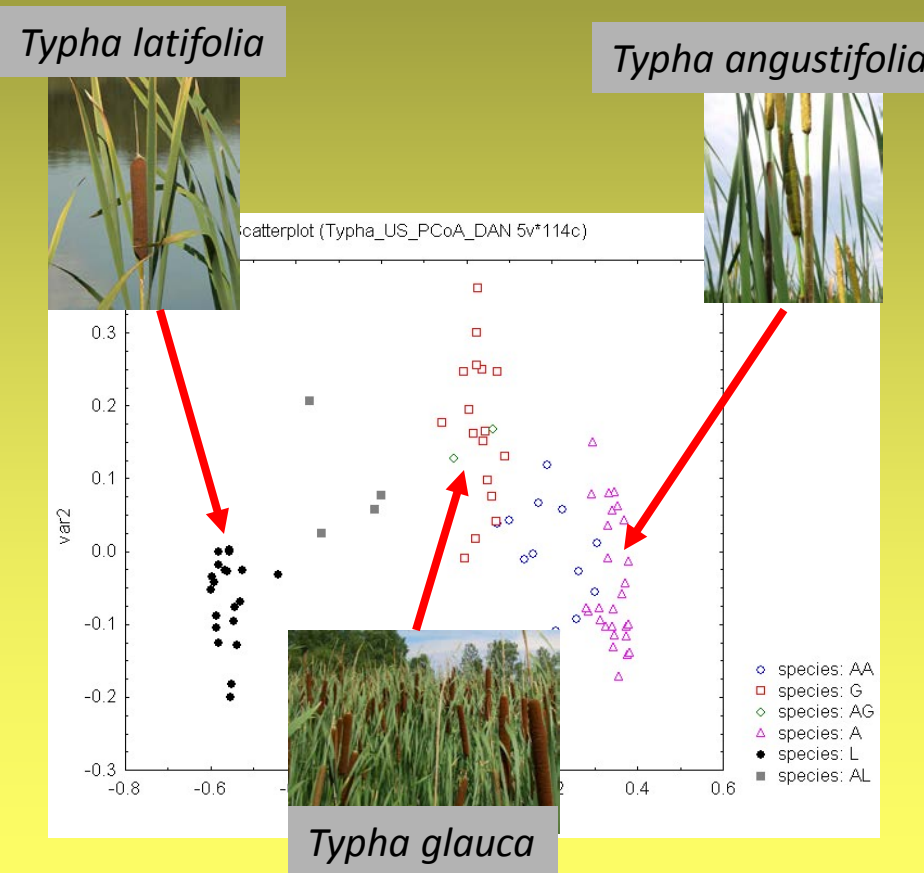
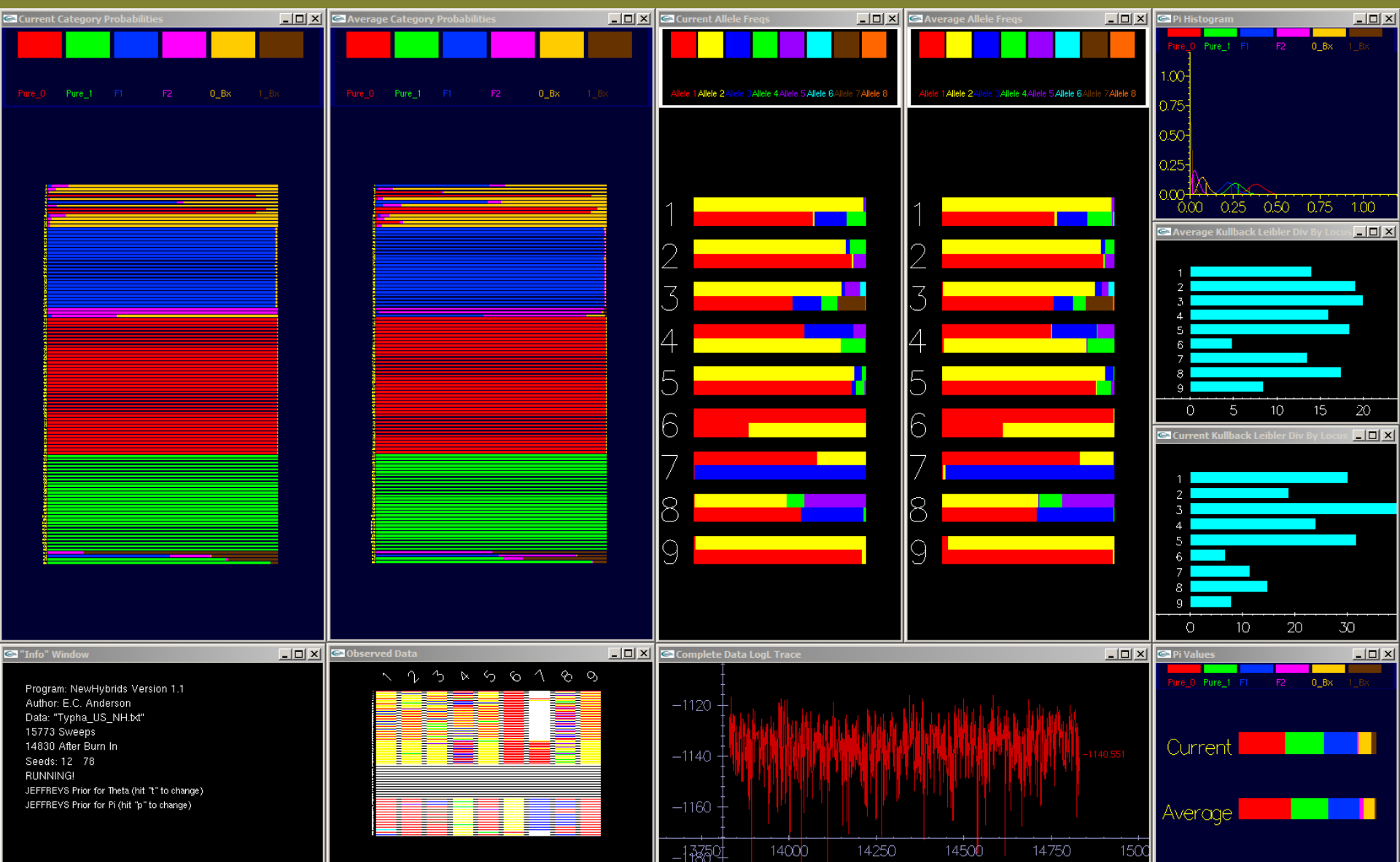


Fig. 1. Barplots of STRUCTURE analyses for (A) RAPD-identified *T. angustifolia* and *T. latifolia* from Collection Group 1; (B) RAPD-identified *T. angustifolia*, *T. latifolia*, and putative F₁ hybrids from additional Collection Group 1 sites in Indiana, Wisconsin, and Minnesota; and (C) mixed populations 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).

Hybridizace – software NewHybrids

- identifikace hybridů mezi dvěma druhy
- založen na Bayesovském modelovém přístupu (MCMC)
- počítá pravděpodobnost příslušnosti jedince do jedé z předem definovaných hybridních tříd
 - druh 1
 - druh 2
 - F1
 - F2
 - Bx1 (zpětný kříženec s druhem 1)
 - Bx2 (zpětný kříženec s druhem 2)
- nemusí být specifikováno, co jsou čisté rodičovské druhy

Hybridizace – software NewHybrids



Polyploidi – AFLP

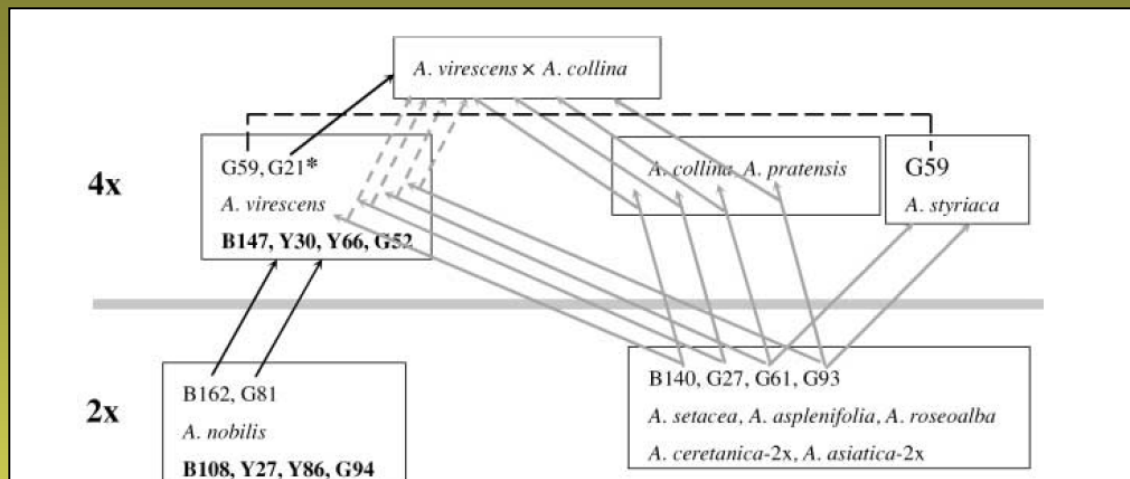
- **specifika polyploidních AFLP dat**

- dominantní => nepoznám (ani mne nezajímá) počet alel, tj. není problém s *allele dosage*
- často větší množství proužků u polyploidů
- umožňují meziploidní srovnání, tj. současnou analýzu di- a polyploidů

0	0	0	0	0	0	0		
074A	074	1	0	1	1	0	0	0
074A	074	1	0	1	1	0	0	0
074B	074	1	1	1	1	0	1	0
074B	074	1	1	1	1	0	1	0
...								

Allopolyploidi a AFLP

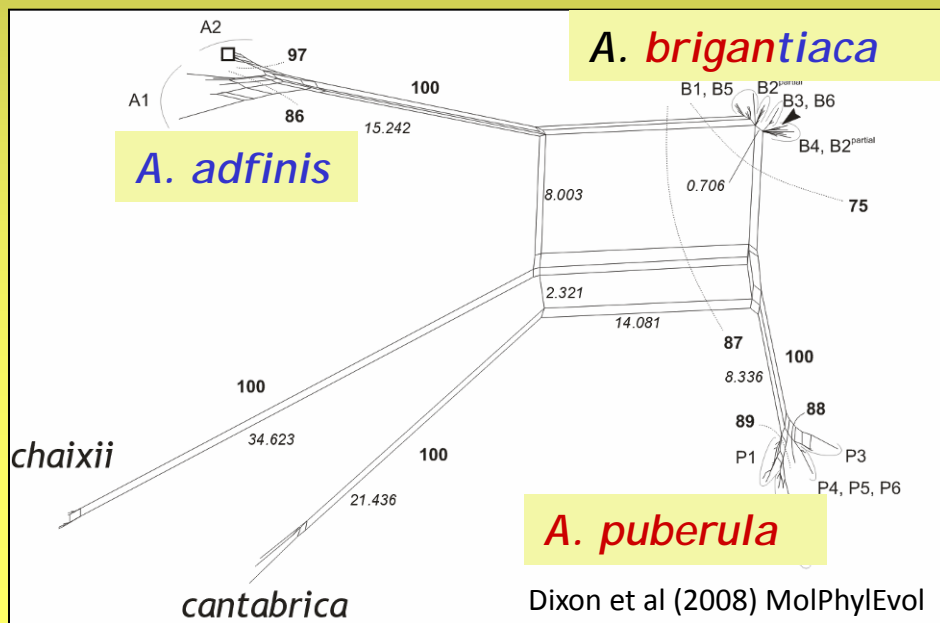
- druhově specifické AFLP proužky



Achillea (Guo et al. 2005 New Phytol)

➤ program Excel ...

- distanční síť



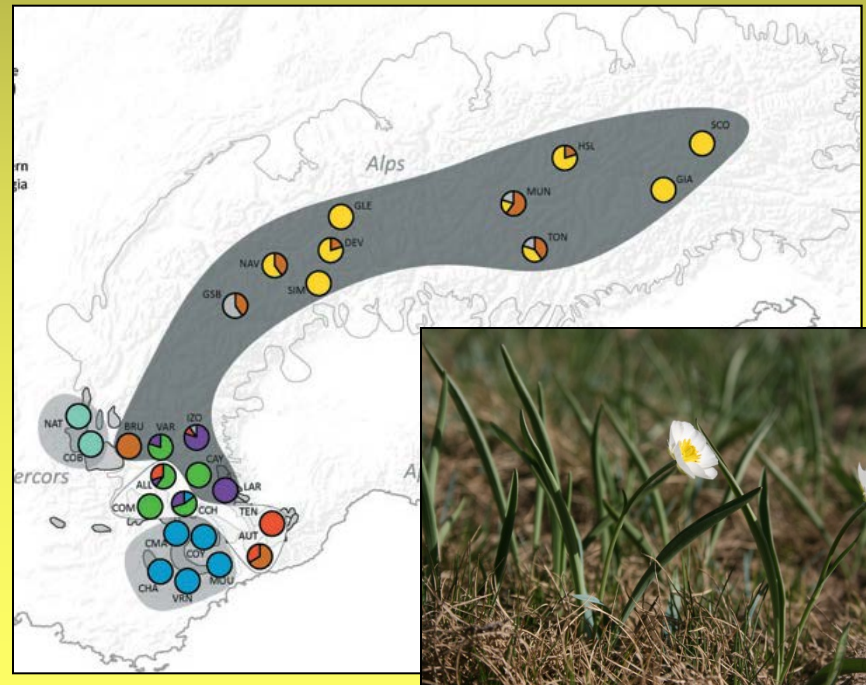
➤ program Splits Tree

Polyploidi – K-means clustering

- **Non-hierarchical K-means clustering**
 - neparametrická metoda, nezaložena na modelu ani předpokladech HW rovnováhy
 - prostá snaha nalézt při daném K takové rozdělení do K skupin, které maximalizuje mezi-skupinovou varianci („inter-cluster inertia“)
 - R-skript pro opakované spuštění - details viz Arrigo et al. 2010 New Phytol

K-means	Structure								NA
	S1	S2	S3	S4	S5	S6	S7	S8	
R1	28								
R2		10							
R3			43						2
R4			1	12					14
R5				2	22	5			7
R6					4	18			
R7							37*		2
R8				21	14				
R9								1	3
R10								4	
R11								3	
R12								6	

Ranunculus kuepferi (Burnier et al (2009) Mol Ecol



➤ **R skript**

Polyploidi – Structure

- **Analýza smíšených di-polyploidních datasetů**
 - statistické modely si dovedou poradit s nejistotou ...
 - viz i diploidní AFLP data – recessive alleles model

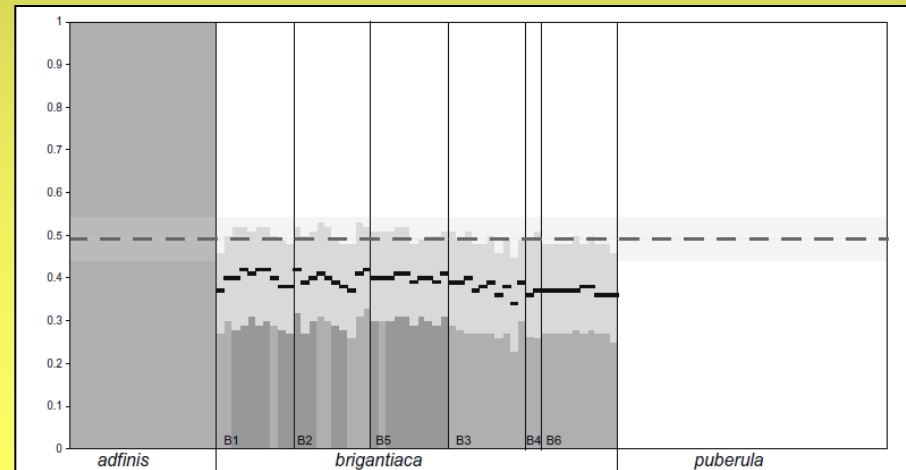
0	0	0	0	0	0	0		
074A	074	1	0	1	1	0	0	0
074A	074	1	0	1	1	0	0	0
074B	074	1	1	1	1	0	1	0
074B	074	1	1	1	1	0	1	0
...								

- **kódování polyploidních dat**
 - Stock et al. 2010 Evolution
 - níže 1 diploid a jeden tetraploid v di-tetraploidním datasetu:

0	0	0	0	0	0	0		
074A	074	1	0	1	1	0	0	0
074A	074	1	0	1	1	0	0	0
074A	074	-9	-9	-9	-9	-9	-9	-9
074A	074	-9	-9	-9	-9	-9	-9	-9
102B	102	1	1	1	1	0	1	0
102B	102	1	1	1	1	0	1	0
102B	102	1	1	1	1	0	1	0
102B	102	1	1	1	1	0	1	0
...								

Polyploidi – BAPS

- nutno mít sebrány všechny potenciální diploidní předky
- diploid = training dataset, polyploidi = neznámí
- **BAPS – trained clustering**
 - mixture analýza, při omezení K na počet předem definovaných diploidních skupin přiřadí polyploidy k nejbližší skupině
 - zařazované jedince možné předem rozčlenit do skupin (\sim prior) – budou přiřazováni jako celek
- **BAPS-admixture based on pre-defined clustering**
 - admixture obdoba

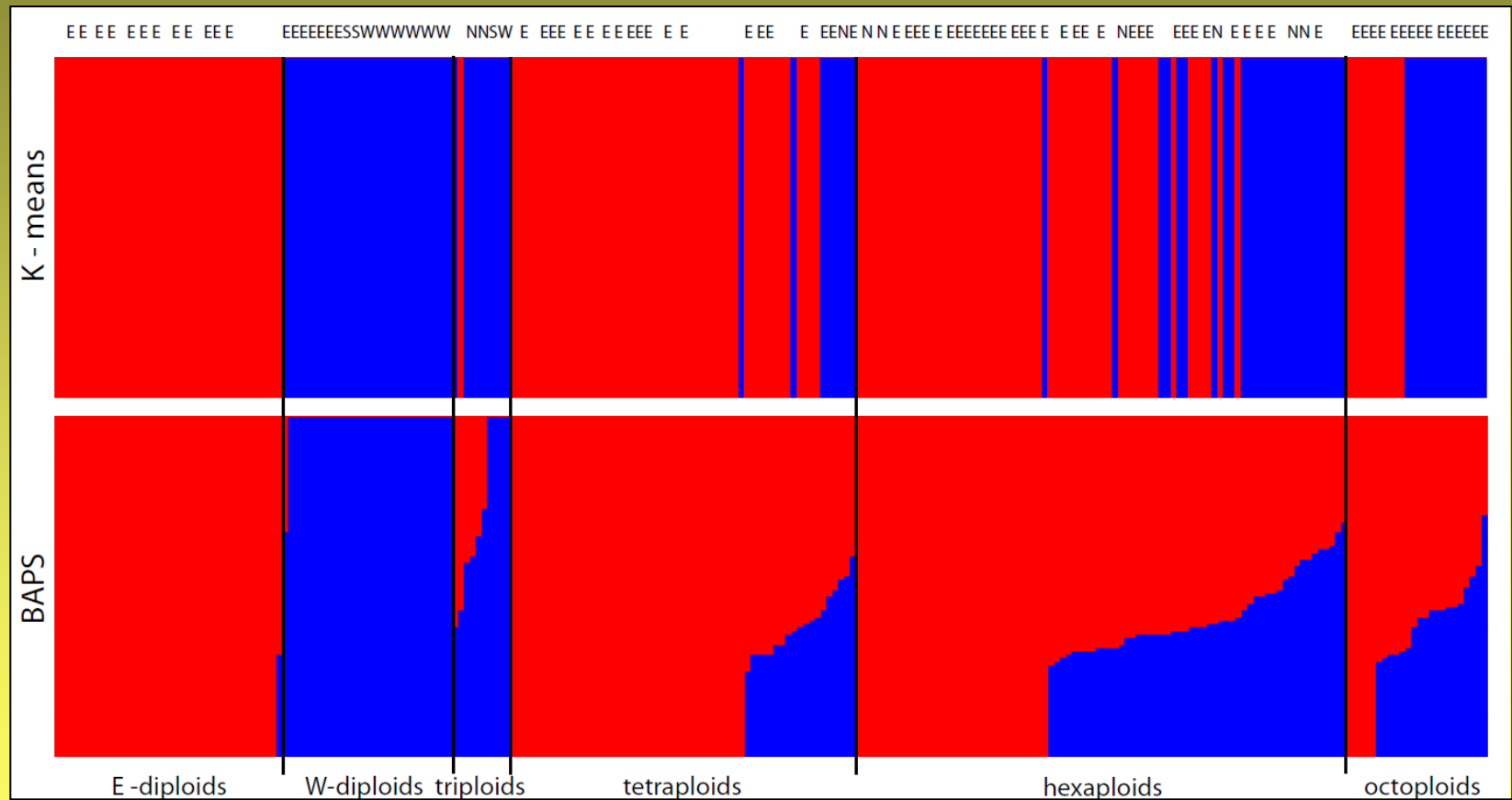


Dixon et al (208) Mol Phyl Evol

Fig. 4. Admixture proportions of *A. brigantia* individuals (mixed bars) with *A. adfinis* (full grey bars) and *A. puberula* (full white bars) pre-defined as parental genomes. In *A. brigantia*, the observed admixture coefficients (black lines) and their confidence intervals (light grey bar segments) as well as the maternal parentage (dark grey *A. adfinis*, medium grey *A. puberula*) are indicated. The range of admixture coefficients from artificial additive profiles (see text for details) and their overall mean are indicated by a horizontal light grey bar and the dashed line, respectively.

Porovnání

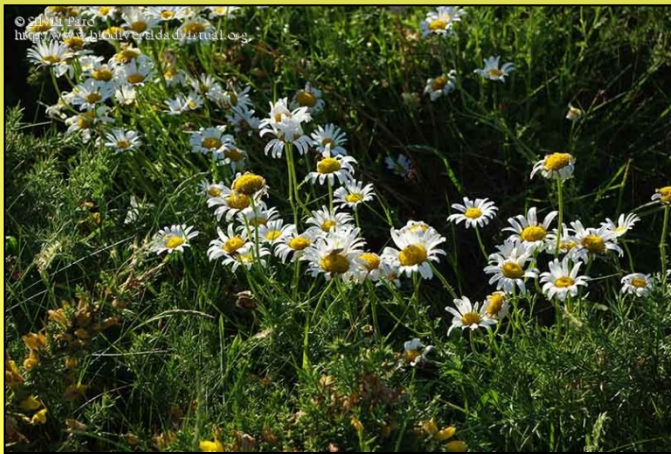
- **BAPS-admixture based on pre-defined clustering + K means clustering**



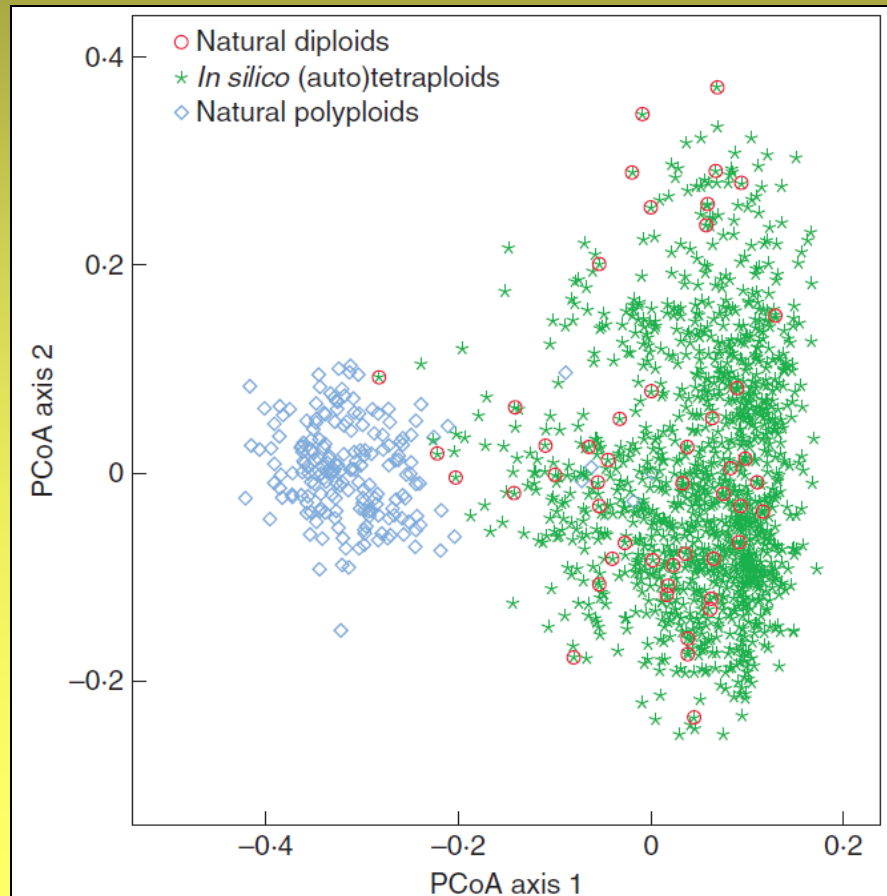
Polyploidi – *in silico* profily

- AFLP data

- snaha o korekci vlivu většího množství proužků u polyploidů
- vytvoření „hybridních“ AFLP profilů kombinací z reálných diploidů (0/0 v rodičovském AFLP profilu = 0 v „hybridním“ profilu; 1/0=1, 0/1=1 a 1/1=1)
- vynesení obou skupin např. v PCoA

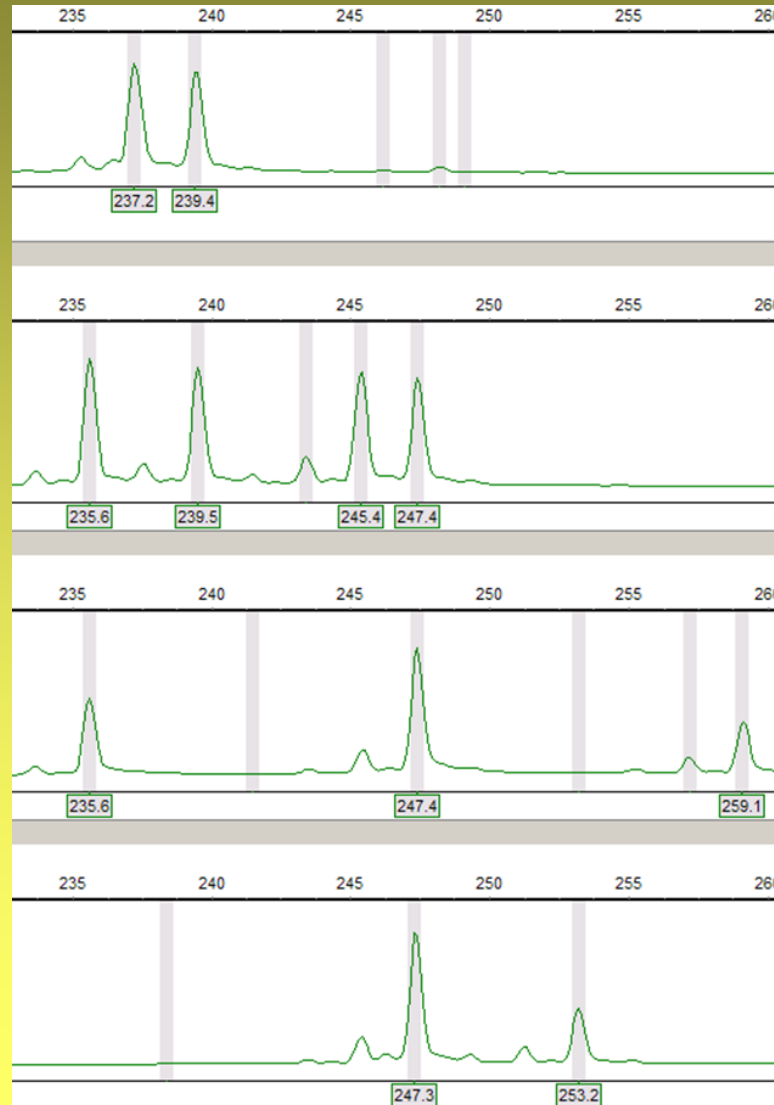


Leucanthemum pluriflorum, Greiner et al (2013) Ann Bot



Polyploidi – kodominantní data

Tetraploidní data (*Betula*)



Jak hodnotit tetraploidy

- jako dominantní data – přítomnost/nepřítomnost alel
 - nelze použít přístupy založené na frekvencích alel
- kodominantně – problém s *allele dosage* (vidíme alely, tj. fenotyp, 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 plochy peaku (MAC-PR; Esselink et al. 2004)
- autopolyploidi/allopolyploidi – polysomická/disomická dědičnost
- nulové alely?
- software umožňující analýzu různých ploidních úrovní – SPAGeDi, TETRASAT, BAPS, STRUCTURE, GenoDive, PolySat...

Odhad alelických frekvencí

- nutné pro výpočet H_e a fixačních indexů
- *allele dosage uncertainty* – inkorporování této nejistoty do výpočtu odhadu
 - každá alela má stejnou frekvenci být vícekrát (SPAGeDi - polysomic inheritance) - nadhodnocení frekvence vzácných alel
 - alela s vyšší frekvencí má větší šanci být vícekrát (GenoDive - polysomic inheritance) - kruhový argument
 - ignorování genotypů s chybějícími alelami, k výpočtu použity pouze úplné genotypy (StAMPP, atetra, TETRASAT) - odchýlené (*biased*) frekvence alel - problém u málo variabilních lokusů - mnoho neúplných genotypů a odhad založený na malém počtu dat
 - použití fenotypů místo genotypů k výpočtu - ML (PolySat, GenoDive)
- chybí detailní simulace těchto přístupů

Populační struktura

- STRUCTURE - autopolyploidi, společná analýza více ploidních úrovní
- problém - odchylky od základního předpokladu, tj. HWE uvnitř clusterů?
 - mnoho polyploidů má tendence k inbreedingu nebo asex. rozmnožování - tj. může způsobit divné clusterování, případně nedetekování skutečné struktury

Populační diferenciace

- F_{ST} - jednoduše rozšiřitelné na autopolyploidy s disomickou dědičností
- počítáno na základě frekvence alel, tj. je téměř nemožné počítat F-statistiky u polyploidů
- pokud lze spolehlivě odhadnout frekvence alel (vzácné případy) - stále je problém s polysomickou dědičností
 - GenoDive a adegenet umí testovat odchylky od HWE u polyploidů
 - pokud je špatný předpoklad polysomické dědičnosti (lokus je ve skutečnosti disomický) - nadhodnocení vnitropopulační heterozygosity a podhodnocení divergence mezi populacemi - G_{ST}), jediné R_{ST} je nezávislé na ploidní úrovni a typu segregace

Distanční analýzy

- SMC - jednoduše aplikovatelný na polyploidy
- Jaccard, Dice - také aplikovatelné na polyploidy a směsné populace
- tyto koeficienty nepočítají s genotypy, ale fenotypy, tj. neberou v úvahu počty alel - odchylky při srovnávání různých ploidii
- speciální koeficienty pro porovnání různých ploidii
 - Tomiuk & Loeschke - proporce různých typů genotypů/fenotypů
 - band sharing measurement (Tomiuk et al. 2009) - sdílení běžných alel
 - Bruvo distance - sdílená presence/absence, délková variabilita alel (GenoDive, PolySat) - nadhodnocuje genetické vzdálenost mezi ploidii
- tj. díky problému s allele dosage je lepší použít distance, které nevyžadují úplné genotypy

Praktikum 2

- **NewHybrids**
 - MSA soubor (*Typha*) převed'te do formátu pro NH
 - spus'tte program, pohrajte si s ním
 - jaké jsou hodnoty jednotlivých parametrů?
- **K-means clustering**
 - na datasetu *Galium* v R
 - srovnejte s výsledky z BAPS z minule

Práce s programy pro hodnocení binárních dat



FSTAT

<http://www2.unil.ch/popgen/softwares/fstat.htm>

- počty a frekvence alel na vzorek a lokus
- Nei's gene diversity
- F_{IS}
- Nei's F statistics
- Weir & Cockerham (1984) odhady
 - F_{IT} , F_{ST} and F_{IS}
- test HW ekvilibria (signifikance F_{IS})
 - uvnitř populací, globálně



FSTAT

<http://www2.unil.ch/popgen/softwares/fstat.htm>

- Utilities – File Conversion – Genepop->Fstat
- otevřít Genepop.gen vygenerovaný MSA

5 4 4 2 (*počet vzorků-populací, počet lokusů, nejvyšší číslo alely, 2=alela má 2 znaky*)

NLGA1

NLGA3

NLGA5

NLGA6

1	303	303	202	202
1	103	202	101	202
2	303	202	204	202
2	304	202	203	202
2	304	202	204	202
2	304	202	202	202
2	404	202	202	102
3	203	202	303	102
3	204	202	203	202
3	303	101	303	202
4	303	202	202	202
4	303	202	202	202
4	303	202	202	202
4	303	202	202	0
5	303	404	102	202



FSTAT

<http://www2.unil.ch/popgen/softwares/fstat.htm>

- File – Open
- Gene diversities and F-Statistics
 - Per locus and sample
 - Global statistics
- Testing
- Run – výsledky jsou zapsány do souboru

The screenshot shows the 'Fstat for Windows' application window. The menu bar includes 'File', 'Options', 'Utilities', 'Biased dispersal', 'Mantelise it!', and 'Help'. The main window is divided into several sections:

- Gene diversities and F-Statistics:**
 - Per locus and sample statistics:** Includes checkboxes for 'Allele frequencies' (unchecked), 'Genotypic frequencies' (unchecked), 'Number of alleles' (checked), 'Allelic richness' (unchecked), 'Gene diversity' (checked), and 'Fis' (checked).
 - Global statistics:** Includes checkboxes for 'Nei's Fstatistics' (checked), 'Weir and Cockerham Fstatistics' (checked), 'Rho_st' (unchecked), and 'Fst per pair of samples' (unchecked).
- Testing:**
 - Global Tests:** Includes checkboxes for 'Hardy Weinberg within samples' (checked) and 'Hardy Weinberg overall samples' (checked), each with a dropdown menu set to '100'.
 - Population differentiation:** Includes a radio button for 'NO test' (selected) and two radio buttons for 'Test NOT assuming HW within samples' and 'Test assuming HW within samples', each with a dropdown menu set to '0'.
- Genotypic disequilibrium:** Includes radio buttons for 'Overall or for each sample' (selected), 'NO tests', 'Tests between all pairs of loci', and 'Tests between all pairs of loci in each sample'.
- Nominal level for multiple tests:** Includes radio buttons for '5/100' (selected), '1/100', and '1/1000'.
- Save Tables:** Includes a checkbox (unchecked).

At the bottom right, there is a large button labeled 'Run'.

Arlequin 3.5

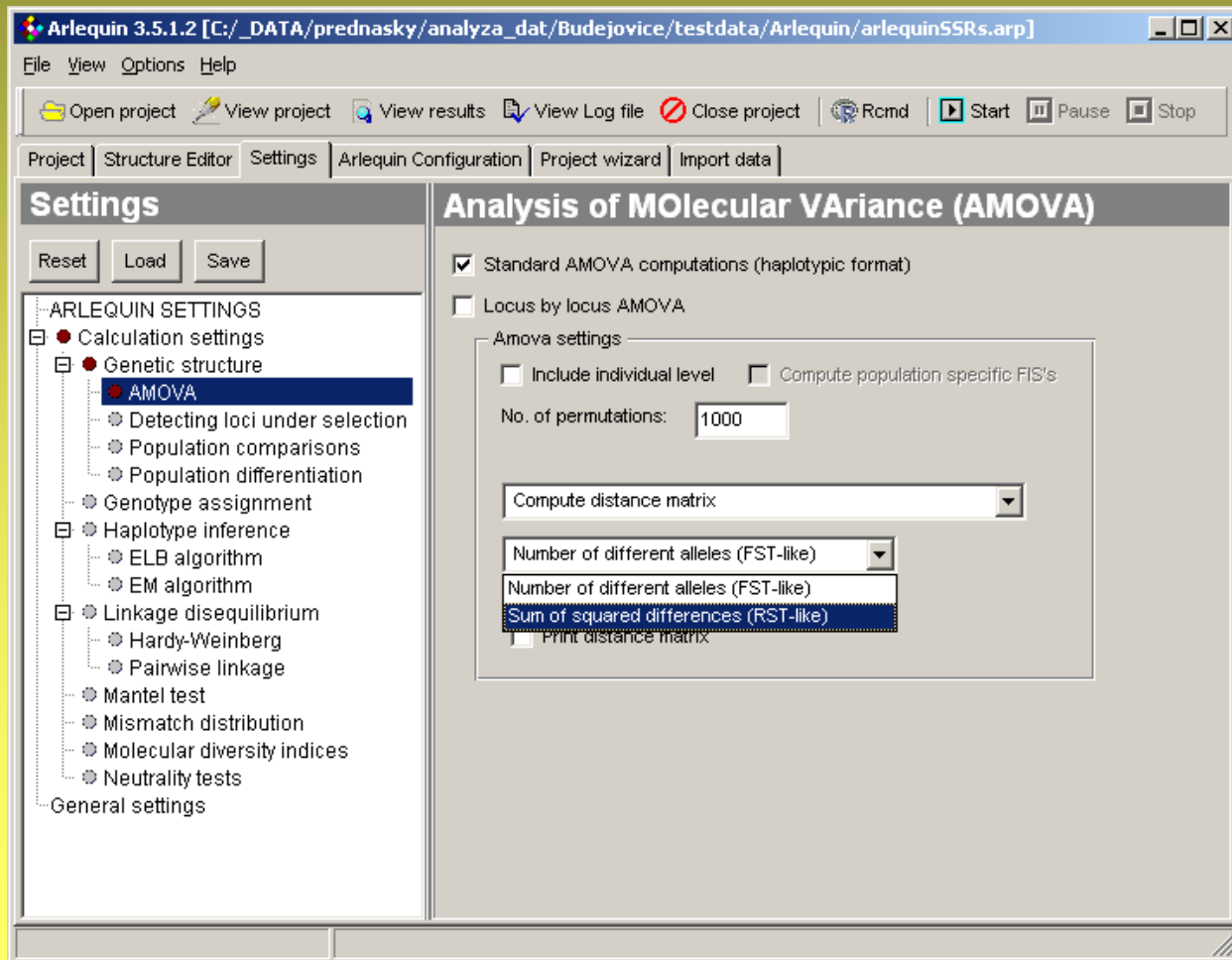
<http://cmpg.unibe.ch/software/arlequin35>

- *.arp soubor vygeneroval MSA
- AMOVA
 - F_{ST} nebo R_{ST} -like analýzy
- srovnání populací – párové F_{ST}
 - Slatkin's linearized [$F_{ST}/(1-F_{ST})$]
 - Reynold's linearized [$-\ln(1-F_{ST})$]
 - $(\delta\mu)^2$
- HW-ekvilibrium (exaktní test)
- Start
- Rcmd (vyžaduje nainstalované R) – doplní grafy do výstupu

Arlequin 3.5

<http://cmpg.unibe.ch/software/arlequin35>

AMOVA – F_{ST} nebo R_{ST} -like analýzy



FAMD

Fingerprint Analysis with Missing Data

<http://www.famd.me.uk/famd.html>

- různé indexy podobnosti (Jaccard, Dice, SMC, Euclidean...)
- stromy (NJ, UPGMA) + bootstrap (prohlížení v TreeView)
- PCoA (+ grafické 3D zobrazení)
- AMOVA
- Shannonův index (+ testy signifikantních rozdílů)
- statistika proužků (polymorfní, fixované, privátní)
- zacházení s chybějícími daty

FAMD

Fingerprint Analysis with Missing Data

<http://www.famd.me.uk/famd.html>

```
vz1 0 1 1 0 1 1 1 1 1
vz2 0 1 1 0 1 1 1 0 1
vz3 0 1 0 1 1 0 0 0 1
vz4 1 1 0 1 1 0 0 0 1
vz5 0 0 1 0 1 1 1 1 0
vz6 0 0 1 0 1 1 1 0 0
vz7 1 0 0 0 1 0 0 0 0
vz8 1 0 0 0 1 0 0 0 0
```

*

[Groups]

```
AllData/skup1= vz1, vz2, vz3, vz4;
```

```
AllData/skup2= vz5, vz6, vz7, vz8;
```

```
AllData/skup1/pop1= vz1, vz2;
```

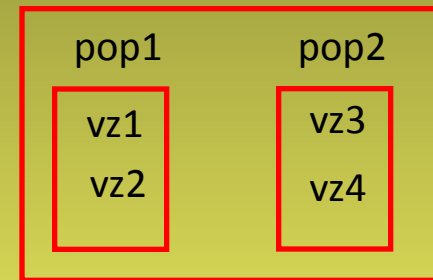
```
AllData/skup1/pop2= vz3, vz4;
```

```
AllData/skup2/pop3= vz5, vz6;
```

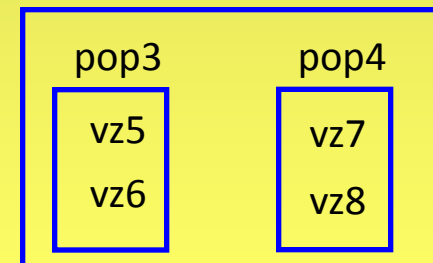
```
AllData/skup2/pop4= vz7, vz8;
```

*

skup1



skup2

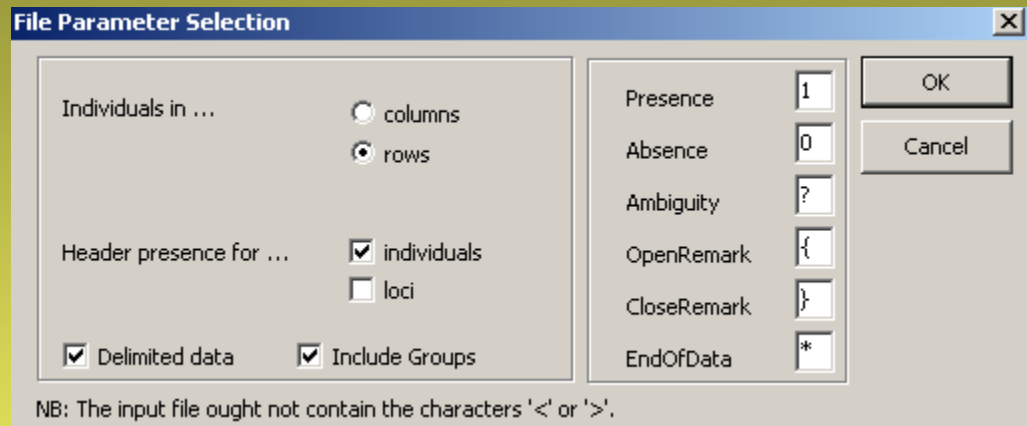


FAMD

Fingerprint Analysis with Missing Data

<http://www.famd.me.uk/famd.html>

- File – Load



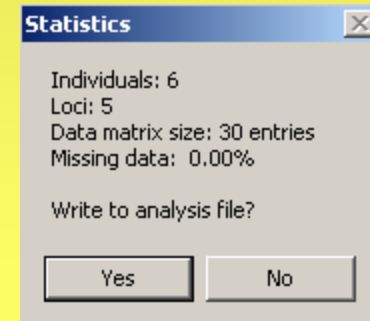
The 'File Parameter Selection' dialog box contains the following settings:

- Individuals in ...: ☐ columns, ☒ rows
- Header presence for ...: ☒ individuals, ☐ loci
- ☒ Delimited data, ☒ Include Groups
- Presence: 1
- Absence: 0
- Ambiguity: ?
- OpenRemark: {
- CloseRemark: }
- EndOfData: *

Buttons: OK, Cancel

NB: The input file ought not contain the characters '<' or '>'.

- DataMatrix – Matrix Statistics



The 'Statistics' dialog box displays the following information:

- Individuals: 6
- Loci: 5
- Data matrix size: 30 entries
- Missing data: 0.00%

Write to analysis file?

Buttons: Yes, No

FAMD

Fingerprint Analysis with Missing Data

<http://www.famd.me.uk/famd.html>

- Options – (Dis)Similarity Coefficients

The screenshot shows the 'Options and Settings' dialog box with the '(Dis)Similarity coefficients' tab selected. The dialog has a title bar with standard window controls. Below the title bar are several tabs: 'Population distance', 'I/O options', '[Unused]', 'Project', 'Bootstrapping & Replicates', 'Weights', 'Distance transformation', 'Trees', 'Consensus trees', 'R-support', 'AMOVA', 'PCoA', 'Allele frequencies', 'Missing data replacement', 'Shannon scaling', and '(Dis)Similarity coefficients'. The '(Dis)Similarity coefficients' tab is active, showing a list of radio buttons for different coefficients: 'Jaccard' (selected), 'Dice; Sørensen', 'SMC (Simple Matching Coefficient)', 'Nei-Li* with r=' (with a text box containing '6.000'), 'Euclidean*', and 'Squared Euclidean*'. Below this list is a note '*distance rather than similarity'. There are two checkboxes: 'Write similarity matrix to analysis file (where available)' (unchecked) and 'Write distance matrix to analysis file' (checked). At the bottom, there is a label 'JC: Calculate average coefficient from how many random draws (r):' followed by a text box containing '100'. On the right side of the dialog are 'OK' and 'Cancel' buttons.

Options and Settings

Population distance | I/O options | [Unused] | Project | Bootstrapping & Replicates | Weights
Distance transformation | Trees | Consensus trees | R-support | AMOVA | PCoA | Allele frequencies
Missing data replacement | Shannon scaling | (Dis)Similarity coefficients

(Dis)Similarity coefficient

☒ Jaccard
☐ Dice; Sørensen
☐ SMC (Simple Matching Coefficient)
☐ Nei-Li* with r= 6.000
☐ Euclidean*
☐ Squared Euclidean*

*distance rather than similarity

☐ Write similarity matrix to analysis file (where available)
☒ Write distance matrix to analysis file

JC: Calculate average coefficient from how many random draws (r): 100

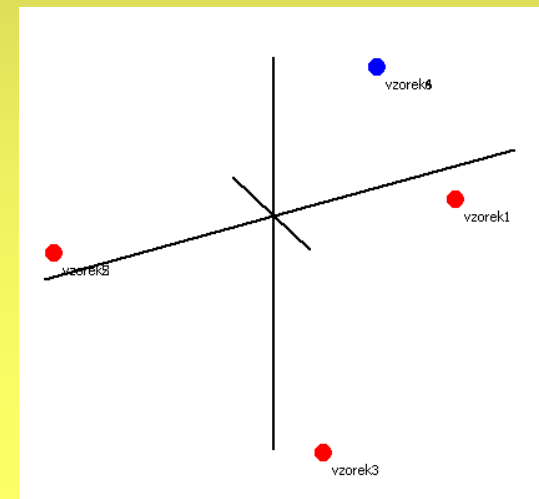
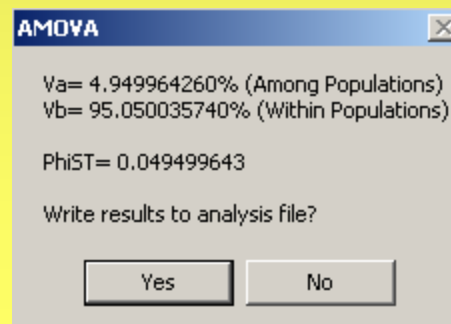
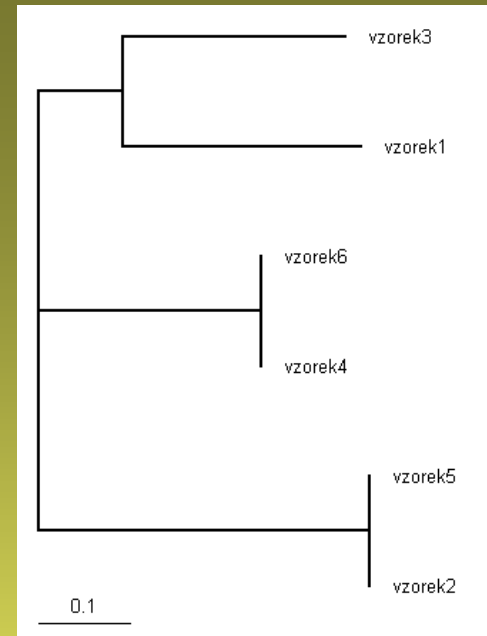
OK
Cancel

- Analysis – Standard Similarity

FAMD

Fingerprint Analysis with Missing Data
<http://www.famd.me.uk/famd.html>

- Trees – Neighbour Joining
- View – Tree File (vyžaduje TreeView)
- Trees – Principal Coordinate Analysis
- Analysis – AMOVA (s Euclidean dist.)



FAMD

Fingerprint Analysis with Missing Data

<http://www.famd.me.uk/famd.html>

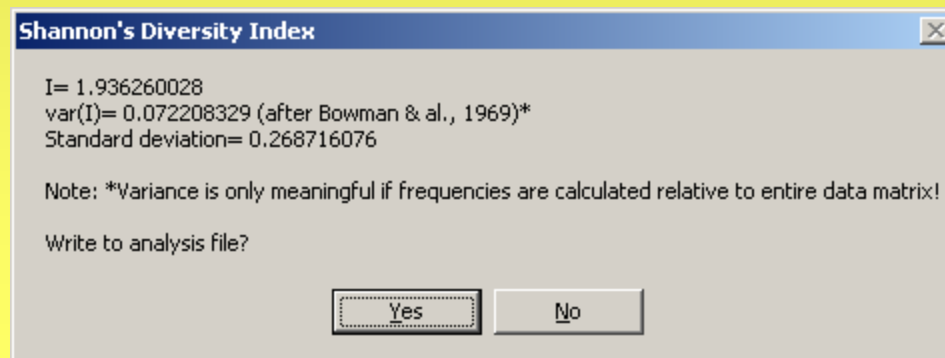
- DataMatrix – Count Bands... – Polymorphic bands
(Fixed Bands/Private Bands/Fixed Private Bands)

Number of polymorphic bands found in group AllData: 4

Number of polymorphic bands found in group AllData/306: 4

Number of polymorphic bands found in group AllData/307: 3

- Analysis – Shannon's index

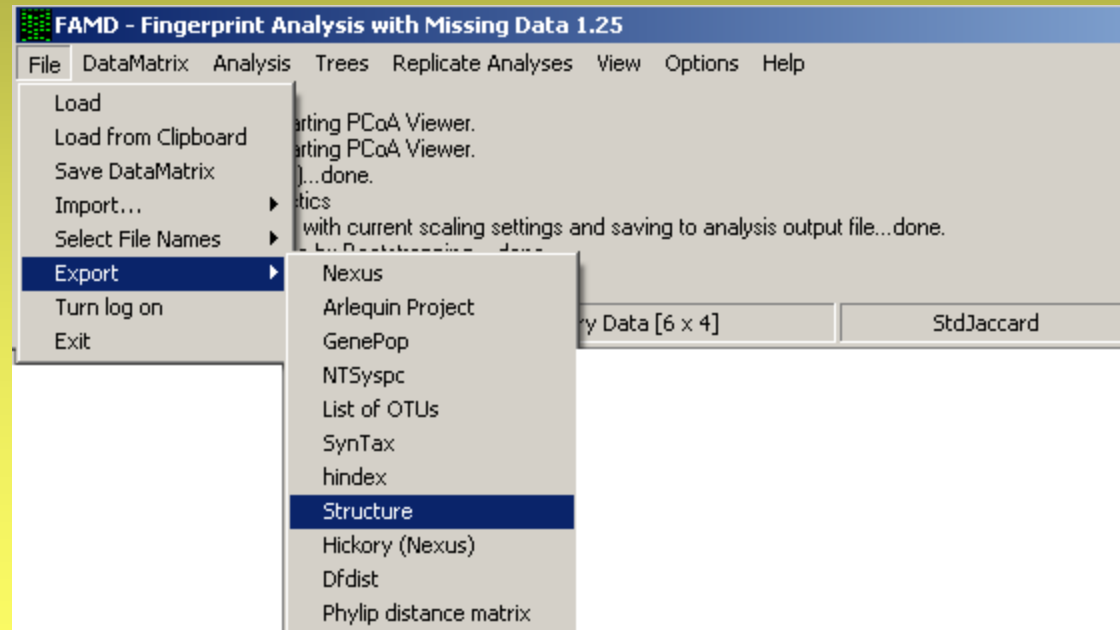


FAMD

Fingerprint Analysis with Missing Data

<http://www.famd.me.uk/famd.html>

- File – Export



AFLPdat

<http://www.nhm.uio.no/english/research/ncb/aflpdat>

- spustit R
- File – Source R code... (vybrat AFLPdat.R)
- File – Change dir... (vybrat adresář s daty)

number	pop	g68	g75	g76	g78	g79	g83	g84
01-1	pop1	1	1	1	0	0	1	1
01-2	pop1	1	1	1	1	0	1	1
01-3	pop1	1	1	1	1	0	1	1
01-4	pop1	1	1	1	0	0	1	1
01-5	pop1	1	1	1	1	1	1	1
04-1	pop2	1	1	1	0	0	1	1
04-2	pop2	1	1	1	0	0	1	1
04-3	pop2	1	1	1	1	0	1	1
04-4	pop2	1	1	1	0	0	1	1
04-5	pop2	1	1	1	0	0	1	1

AFLPdat

<http://www.nhm.uio.no/english/research/ncb/aflpdat>

- Diversity ("AFLPdat.txt") (Nei's gene diversity)

sample	n	proportion of variable markers	gene diversity
pop1	5	0.242718446601942	0.112621359223301
pop2	5	0.199029126213592	0.0941747572815534
pop3	5	0.135922330097087	0.0689320388349514
pop4	4	0.199029126213592	0.104368932038835

diversities.txt

- Diversity.boot ("AFLPdat.txt ", 1000)

sample	observed diversity	95% conf.int.(lower bound)	upper bound
pop1	0.112621359223301	0.0864077669902912	0.139805825242718
pop2	0.0941747572815534	0.0689320388349514	0.118446601941748
pop3	0.0689320388349514	0.0466019417475728	0.0932038834951456
pop4	0.104368932038835	0.0760517799352751	0.132686084142395

div-boot.txt

AFLPdat

<http://www.nhm.uio.no/english/research/ncb/aflpdat>

- Rarity ("AFLPdat.txt")

(DW index)

sample	n	rarity 1	rarity 2
pop1	5	5.37479533538357	26.25
pop2	5	4.39386146533205	22.0833333333333
pop3	5	4.73310202869026	21.25
pop4	4	5.37280146324264	24.4166666666667

rarity-pops.txt

AFLPdat

<http://www.nhm.uio.no/english/research/ncb/aflpdat>

- **Clones.list ("AFLPdat.txt", x)**

(seznam klonů lišících se o x proužků)

List of possible clones

04-104-4 4

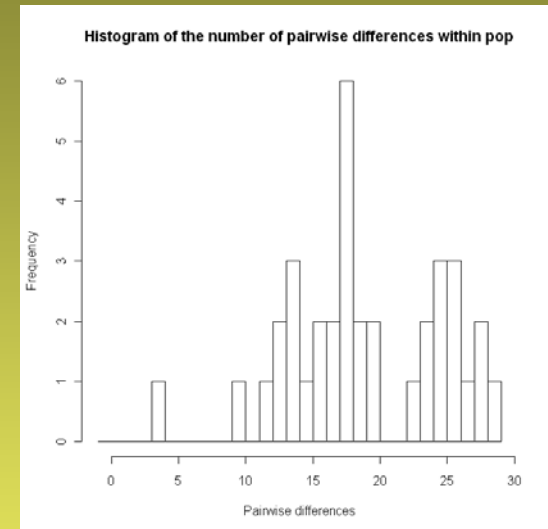
08-208-5 10

- **Clones ("AFLPdat.txt", x)**

(počet různých klonů v populacích, Dg – genotype diversity, effective number of genotypes)

$Dg = n/(n-1) * [1 - \sum (\text{genotype frequencies}^2)]$

Effective nb = $1 / \sum (\text{genotype frequencies}^2)$



sample	n	nb of genotypes	genotype diversity	eff. nb of genotypes	gene diversity
pop1	5	5	1	5	0.112621359
pop2	5	4	0.9	3.57142857142857	0.101941747
pop3	5	4	0.9	3.57142857142857	0.068770226
pop4	4	4	1	4	0.104368932

AFLPdat – exporty

<http://www.nhm.uio.no/english/research/ncb/aflpdat>

- Arlequin ("AFLPdat.txt")
- Structure ("AFLPdat.txt")
- Baps ("AFLPdat.txt")
- Popgene ("AFLPdat.txt")
- Hickory ("AFLPdat.txt")
- Nexus ("AFLPdat.txt")

Arlequin 3.5

<http://cmpg.unibe.ch/software/arlequin35>



```
[Profile]
Title="sparganium"
NbSamples=66
DataType=RFLP
GenotypicData=0
LocusSeparator=NONE
CompDistMatrix=1

[Data]
  [[HaplotypeDefinition]]
    HaplListName="List of AFLP phenotypes"
    HaplList= {
1-S300-1
  1011101010011111111011100111101110110011111111101111011110110
  11111111111011111110111010101010011011011
2-S300-2
  10111010100111111110111001111011100100011111111101111011110110
  11111111111011111110111010101011011011011010
3-S301-1
  10011001100101111111011110111110101110001111111111111110011111
  01100110101011101110111110101001001011010
4-S301-2
  10011001100101111111011110111110101110001111111111111110011111
  01100110101011101110111110101011001011010
5-S301-3
  10011001100101111111011110111110101110011111111111111110011111
  01100110101011101110111110100001001011010
```

Arlequin 3.5

<http://cmpg.unibe.ch/software/arlequin35>



```
[[Samples]]

SampleName="S300"
SampleSize=2
SampleData= {
1-S300-1  1
2-S300-2  1
}
SampleName="S301"
SampleSize=6
SampleData= {
3-S301-1  1
4-S301-2  1
5-S301-3  1
6-S301-4  1
7-S301-5  1
8-S301-6  1
}
SampleName="S302"
SampleSize=5
SampleData= {
9-S302-1  1
10-S302-2 1
11-S302-3 1
12-S302-4 1
13-S302-5 1
}
```

```
[[Structure]]
StructureName="All
Populations"
NbGroups=2

#Mrlina
Group={
"S300"
"S301"
"S302"
"S303"
"S304"
"S305"
"S306"
"S307"
}

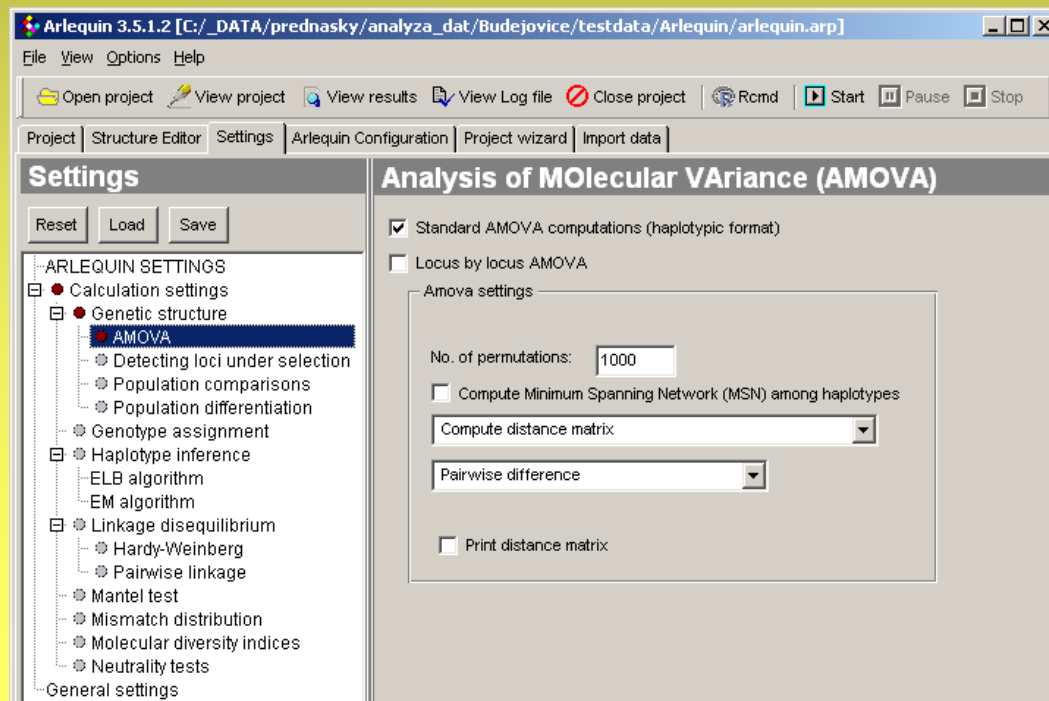
#Cidlina
Group={
"S310"
"S311"
"S312"
"S313"
"S314"
```



Arlequin 3.5

<http://cmpg.unibe.ch/software/arlequin35>

- File – Open Project (otevřít *.arp soubor)
- karta Settings: Calculation settings-Genetic structure-AMOVA-Standard AMOVA computations



Arlequin 3.5

<http://cmpg.unibe.ch/software/arlequin35>



Distance method: Pairwise difference

AMOVA design and results :

Reference: Weir, B.S. and Cockerham, C.C. 1984.
Excoffier, L., Smouse, P., and Quattro, J. 1992.
Weir, B. S., 1996.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	1	147.449	1.22895 Va	7.28
Among populations within groups	64	2288.706	7.00306 Vb	41.51
Within populations	192	1658.710	8.63911 Vc	51.21
Total	257	4094.864	16.87112	

Fixation Indices

FSC : 0.44770
FST : 0.48793
FCT : 0.07284

zt (Mantelův test)

<http://www.psb.ugent.be/~erbon/mantel/>

matice genetických vzdáleností

6					
0.03					
0.64	0.61				
0.28	0.25	0.36			
0.3	0.27	0.34	0.02		
0.36	0.33	0.28	0.08	0.06	
0.18	0.15	0.46	0.1	0.12	0.18

matice geografických vzdáleností

6					
0.07					
7.37	7.36				
7.64	7.63	0.3			
7.66	7.65	0.3	0.08		
7.68	7.67	0.34	0.04	0.1	
7.81	7.79	0.71	0.47	0.53	0.43

File A: gene.txt

File B: geo.txt

Size of matrices: 6 x 6

Number of iterations: 10000

Options: simple

Randomizing...

r = 0.489822

p = 0.000100 (one-tailed)

SPAGeDi 1.3

Spatial Pattern Analysis of Genetic Diversity

<http://ebe.ulb.ac.be/ebe/Software.html>

- diferenciace mezi populacemi (F_{ST} , G_{ST} ad.), globální míry
- diferenciace mezi jedinci (kinship koeficient pro dominantní markery, Hardy 2003)
- variance, jackknife, permutace
- práce z příkazové řádky
- vstupní soubor ve stejné složce jako SPAGeDi-1.3.exe

SPAGeDi 1.3

Spatial Pattern Analysis of Genetic Diversity

<http://ebe.ulb.ac.be/ebe/Software.html>

populací
jedinců

koordináty

lokusů

AFLP

ploidie

horní limity tříd

počet distančních tříd

```
//genetická data
ind pop 1 2 3 4 5 6 7 8
1 31 1 0 0 1 0 1 0
2 31 1 0 0 1 0 1 0
3 31 1 0 0 1 0 1 0
4 31 1 0 0 1 0 1 0
5 31 1 0 1 1 0 0 0
6 32 1 0 0 0 0 1 0
7 32 1 0 0 1 0 1 0
8 32 1 0 0 1 0 0 0
9 33 1 0 0 1 1 0 1
END

//matice prostorových vzdáleností
M9 1 2 3 4 5 6 7 8 9
1 0 0 0 0 0 182 182 182 778
2 0 0 0 0 0 182 182 182 778
3 0 0 0 0 0 182 182 182 778
4 0 0 0 0 0 182 182 182 778
5 0 0 0 0 0 182 182 182 778
6 182 182 182 182 182 0 0 0 927
7 182 182 182 182 182 0 0 0 927
8 182 182 182 182 182 0 0 0 927
9 778 778 778 778 778 927 927 927 0
END
```


SPAGeDi 1.3

Spatial Pattern Analysis of Genetic Diversity

<http://ebe.ulb.ac.be/ebe/Software.html>

- dvojklik na SPAGeDi-1.3.exe
- napsat název vstupního a výstupního souboru (včetně koncovky!)
- program sumarizuje vstupní data, stisknout ENTER
- 1 pro *kinship coefficient*
- vložit inbreeding koeficient (např. 0.25)
- 134 (matice, permutace, jackknife)
- ENTER (matice je v souboru), ENTER
- vložit počet permutací
- 123 (frekvence pro kategorie, všechny statistiky, matice)
- 3 (multilokusově matice a sloupce)
- výstupní soubor otevřít nejlépe v Excelu

SPAGeDi 1.3

<http://ebe.ulb.ac.be/ebe/Software.html>

VALUES OF THE COMPUTED STATISTICS

Genetic analyses at INDIVIDUAL level (156 individuals)

Dist classes	1	2	3	4	5	6
Max distance	0.1000	1000.0000	2000.0000	5000.0000	10000.0000	15000.0000
Number of pairs	306	38	113	213	456	498
Mean distance	0.0000	676.0000	1228.2124	3760.4883	7292.0658	12377.5884

Pairwise KINSHIP coefficients for a DOMINANT marker, assuming inbreeding coef=0.0000 (Hardy, 2003)

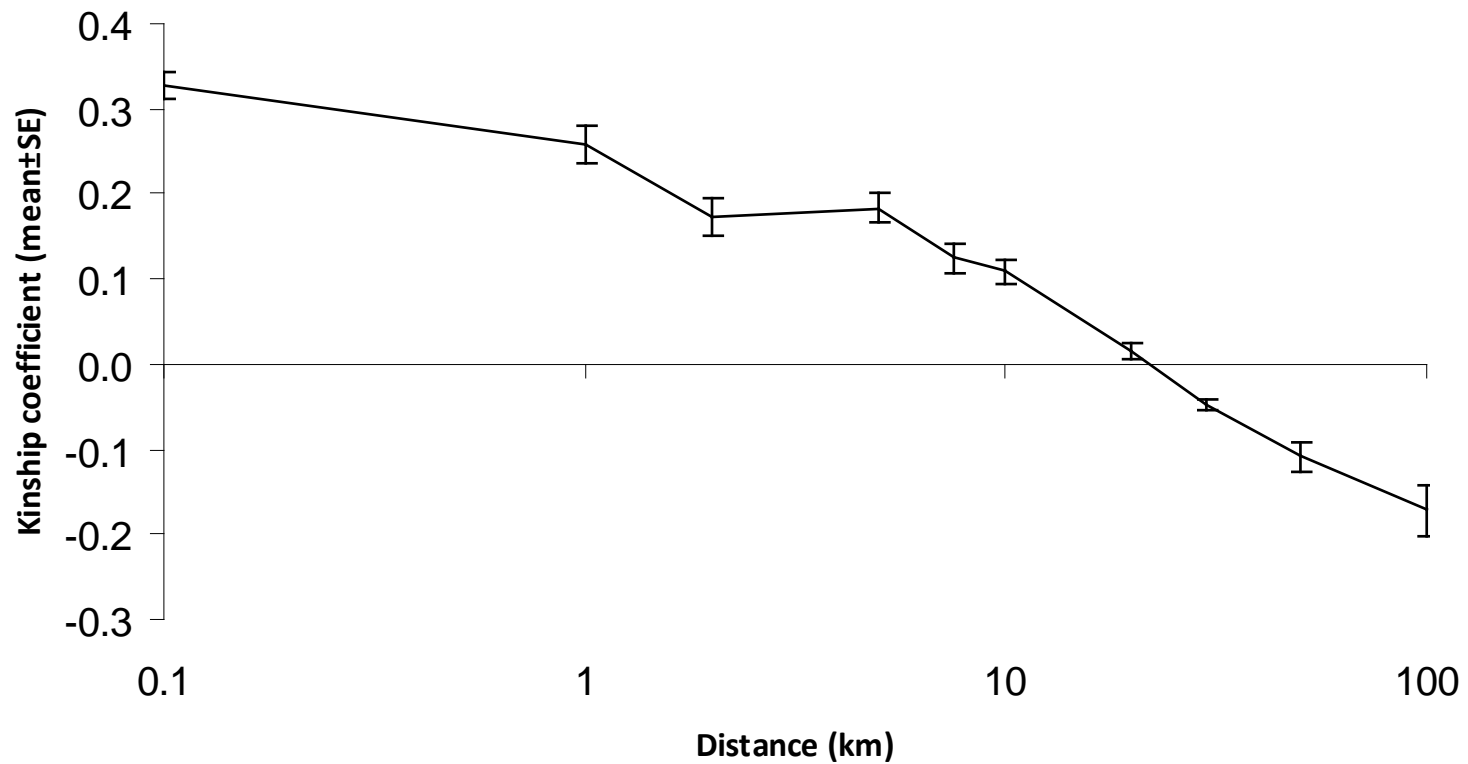
Locus	1	2	3	4	5	6
ALL LOCI	0.2699	0.1536	0.1556	0.1158	0.0635	0.0331
1	0.3257	0.0528	0.1953	0.0242	0.0167	-0.056
2	0.2362	0.0153	0.0561	0.0279	0.0421	0.006
3	0.3127	0.3375	0.3247	0.1141	0.0963	-0.06
4	0.2766	0.2369	0.1309	0.0632	-0.0767	-0.0871

Jackknifed estimators (over loci)

Mean	0.3389	0.2670	0.1751	0.1858	0.1288	0.1121
SE	0.0149	0.0222	0.0238	0.0180	0.0190	0.0154

SPAGeDi 1.3

<http://ebe.ulb.ac.be/ebe/Software.html>



Další software a odkazy ...

- Phylogeny Programs

<http://evolution.genetics.washington.edu/phylip/software.html>

- Genetic Data Analysis Software

<http://courses.washington.edu/fish543/Software.htm>

- Web Sites with Software for Population Genetic Analyses

<http://www.discat.unipi.it/BiolMar/utilities/progs.htm>

- DNA laboratoř Katedry botaniky PřF UK

<http://botany.natur.cuni.cz/dna>

Literatura - SSR

- 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. & Schlötterer Ch. (1999): *Microsatellites. Evolution and Applications*. Oxford University Press
- 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
- 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.