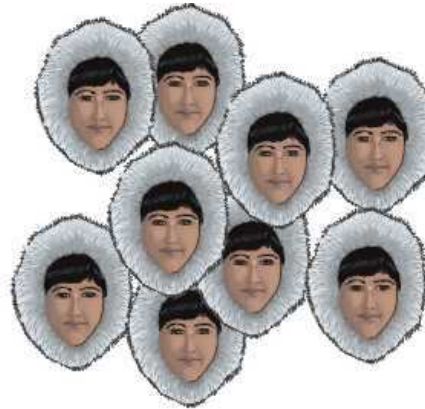


## 2. Analysis of population subdivision (mezipopulační variabilita)



# Hierarchická populační struktura

Druh → populace → subpopulace (demy)

## Předpoklady studia populačně-genetické struktury

- lokusy používané pro analýzu populační struktury jsou neutrální vůči selekci
- **klasický populačně-genetický přístup** = jednotlivé populace jsou předem známy (např. chceme zjistit úroveň genetických rozdílů mezi dvěma lokalitami = populacemi)

# Genetická struktura populací

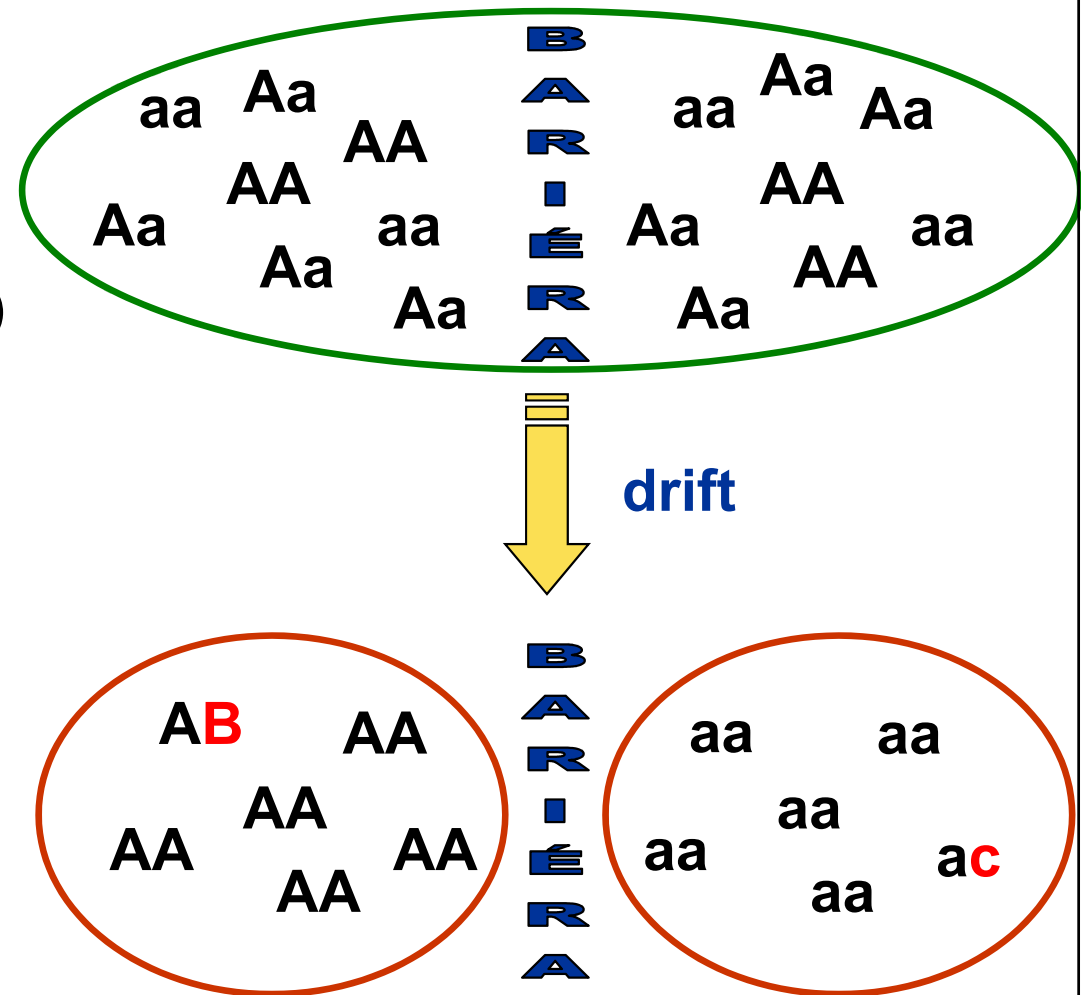
## drift, mutace

- Drift**

→ diferenciace subpopulací díky změnám frekvencí (až fixaci) alternativních alel

- Mutace**

mohou zvýšit diferenciaci (nebo ne – homoplázií)

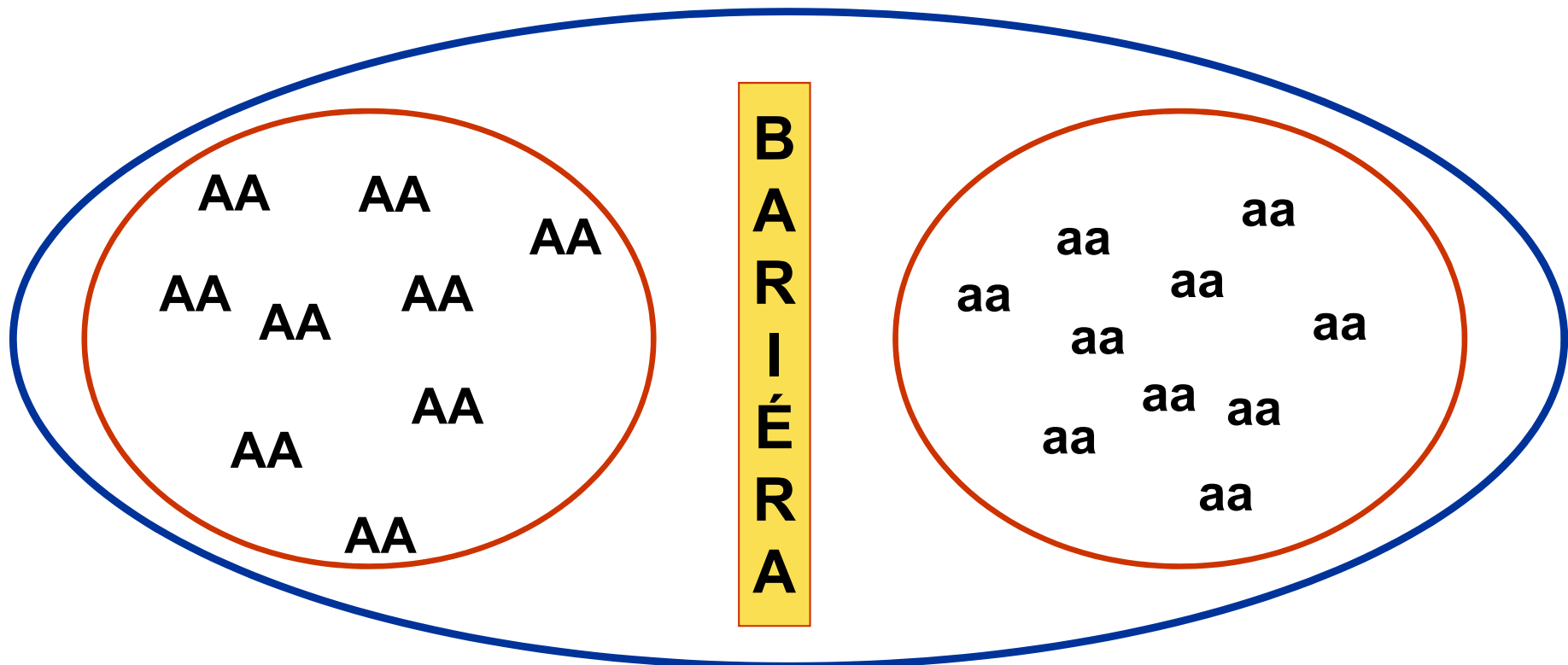


### Migrace (genový tok)

- působí proti diferenciaci subpopulací

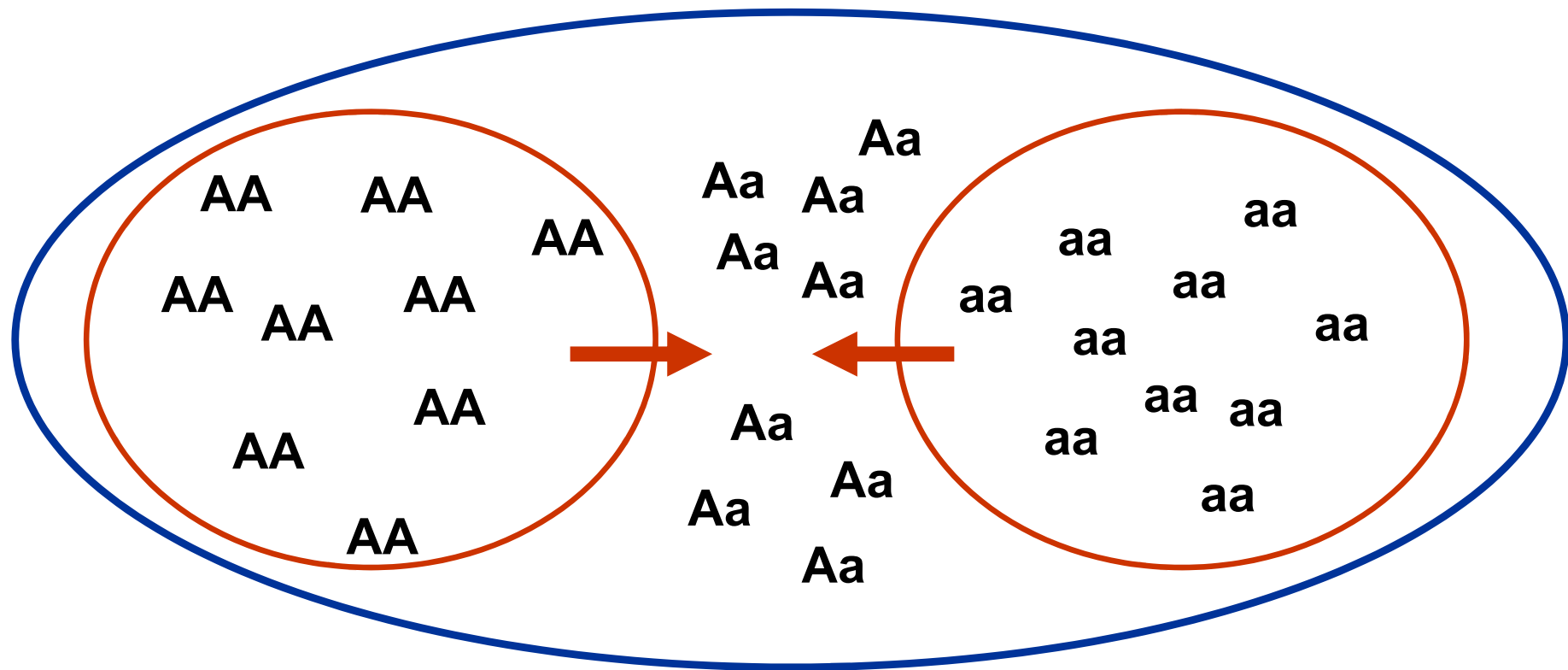
# Vliv populační struktury na heterozygotnost

- Wahlundův princip
- Dvě izolované subpopulace s fixovanými alelami
- Subpopulace v HW, celkově v populaci však nedostatek heterozygotů



# Wahlundův princip (isolate breaking)

- Pokles homozygotnosti při sloučení subpopulací



# Wahlundův princip - příklad

- Jezero Bunnersjöarna (severní Švédsko) – „brown trout“
- 2 alely na jednom znaku

	170/170	170/172 (= Ho)	172/172	Total	p	2pq (=He)
Přítok	50	0 (0)	0	50	1.000	0.000
Odtok	1	13 (0.26)	36	50	0.150	0.255
Celé jezero (expected)	51 (33.1)	13 (0.13) (48.9)	36 (18.1)	100	0.575	0.489

$$p^2 = 0.575^2$$

$$q^2 = 0.425^2$$



# Factorial correspondence analysis

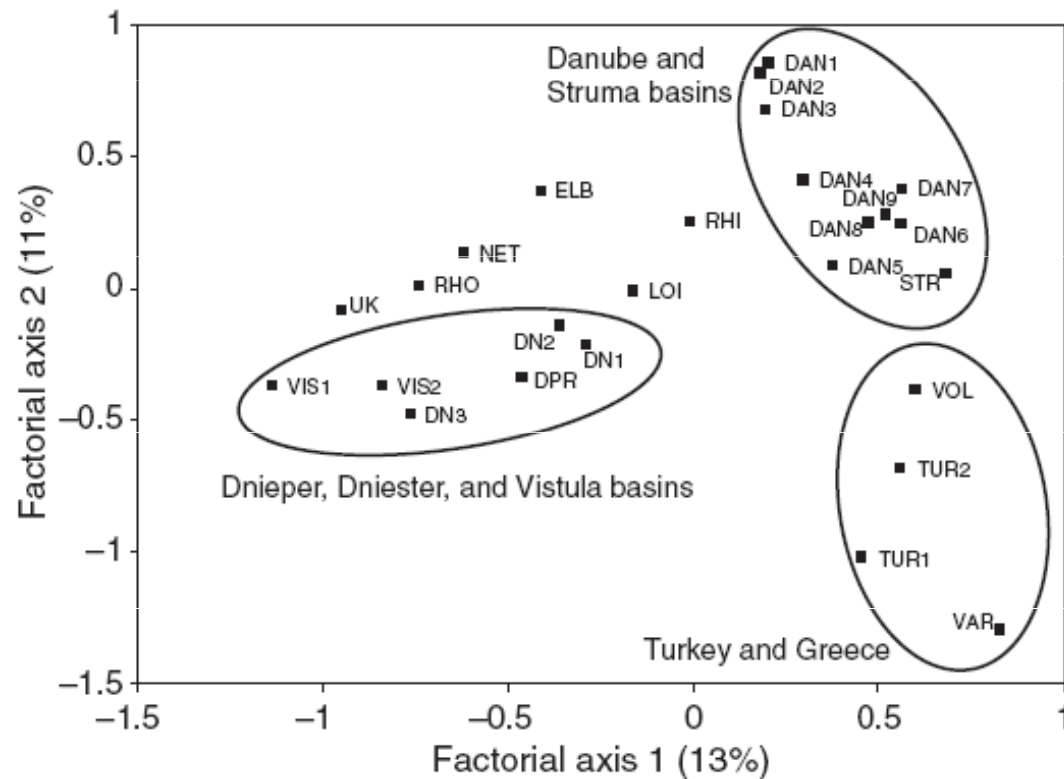


Fig. 2 A two-dimensional plot of the factorial correspondence analysis performed using GENETIX based on 12 microsatellite loci. Three geographical groups are bounded by grey lines.

- each locus as one variable, reduction of number of variables
- **Genetix** – orientační zjištění strukturovanosti populace
- individuals vs. populations

# F-statistiky



**Masatoshi Nei**  
\*1931



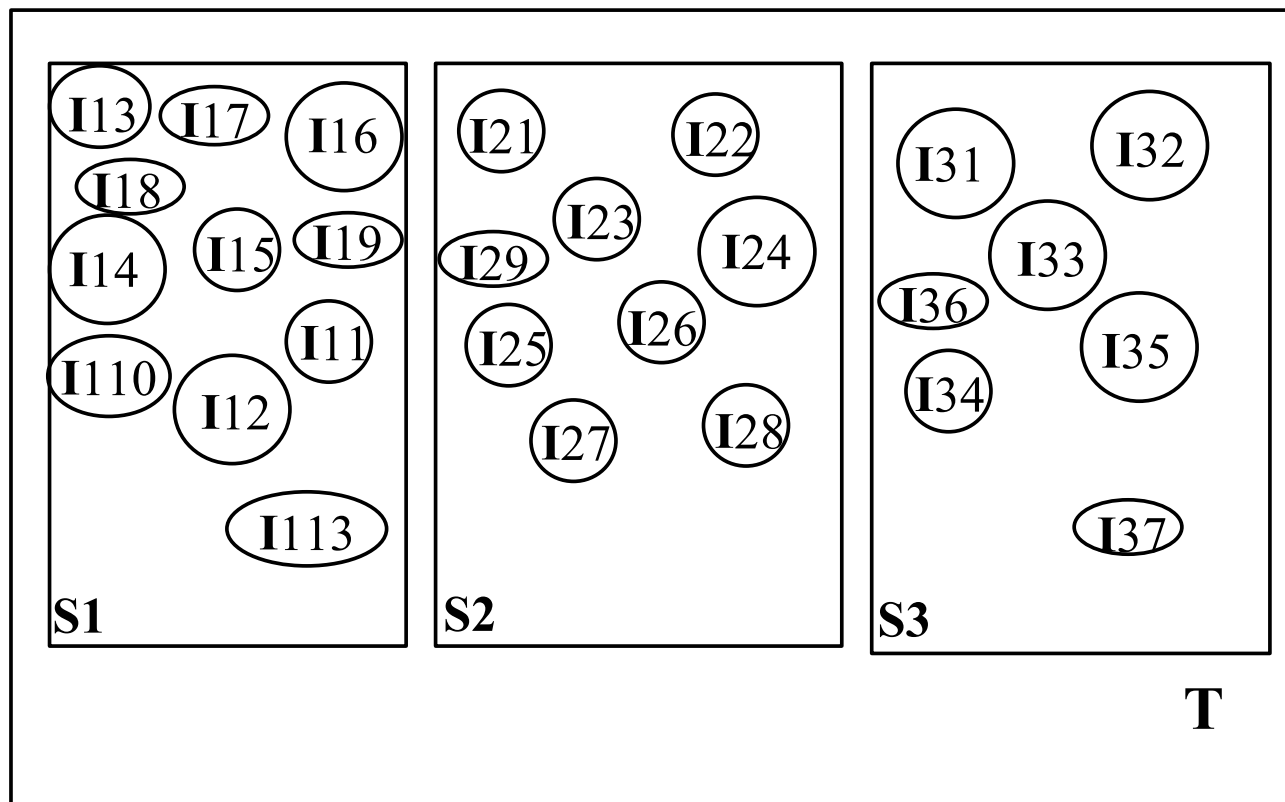
**Sewall Wright**  
1889 - 1988

- Wright, Nei  $F_{IS}$ ,  $F_{ST}$ ,  $F_{IT}$
- Popisují heterozygotnost (odchyly od HW) na různých měřítkách



# Odhad vlivu populační struktury na genetický make-up populace

Celková populace



- 3 úrovně (T, S, I)
- $x$  subpopulací ( $x = 1$  až  $k$ ; zde  $k = 3$ )
- každá subpopulace má  $N_x$  jedinců
- AA, AB, BB – odlišný symbol
- př. I1-13 = 13. jedinec z první subpopulace

# Koncept heterozygotnosti

$H_I$  – průměrná pozorovaná heterozygotnost jedince v subpopulaci

$H_S$  - očekávaná heterozygotnost jedince v subpopulaci za předpokladu náhodného páření

$H_T$  - očekávaná heterozygotnost jedince v celé populaci za předpokladu náhodného páření

$$H_I = \sum_{x=1}^k H_x / k \quad H_x = \text{pozorovaná heterozygotnost v subpopulaci } x$$

$$H_S = 1 - \sum_{i=1}^j p_{i,x}^2 \quad p_{i,x}^2 = \text{frekvence } i\text{-té alely v subpopulaci } x \quad \bar{H}_S = \sum_{x=1}^k H_S / k \quad \text{průměrná oček. heterozygotnost v populaci (tj. průměr ze všech subpopulací)}$$

$$H_T = 2p_0q_0 \quad p_0 = \text{frekvence alely v celé populaci}$$

- pouze pro dvě alely na jednom lokusu (Wright 1931)
- pro více alel je výpočet složitější (Nei 1987)

# F statistiky

$$F_{IS} = \frac{\overline{H}_S - H_I}{\overline{H}_S}$$

Snížení heterozygotnosti jedince kvůli nenáhodnému páření v subpopulaci (~ HWE)

$$F_{ST} = \frac{H_T - \overline{H}_S}{H_T}$$

Vliv rozdělení populace na subpopulace

$$F_{IT} = \frac{H_T - H_I}{H_T}$$

Celkový koeficient inbreedingu  $F_{IT}$  - měří redukci heterozygotnosti jedince ve vztahu k celkové populaci

$$(1-F_{IT}) = (1-F_{ST})(1-F_{IS})$$

Weir & Cockerham (1984)  $f, \theta, F$

Korekce na velikost vzorku a počet subpopulací

# Výpočet F statistik - příklad

Průměrná frekvence alely A v celé populaci

Lokus	Subpopulace 1 (N <sub>1</sub> =40)				Subpopulace 2 (N <sub>2</sub> =20)				$p_{0(i)}$	Pozn.
	AA	AB	BB	$p_{1(i)}$	AA	AB	BB	$p_{2(i)}$		
Loc I	10	20	10	0.5	5	10	5	0.5	0.5	H.-W. rovnováha
Loc II	16	8	16	0.5	4	4	12	0.3	0.4	deficit heterozygotů
Loc III	12	28	0	0.65	6	12	2	0.6	0.625	přebytek heterozygotů
Loc IV	0	0	40	0.0	20	0	0	1.0	0.5	alternativně fixované alely

Výpočet alelových frekvencí

Lokus	Pozorovaná heterozygotnost		Očekávaná heterozygotnost			Wrightova F-statistika		
	$H_{1(i)}$	$H_{2(i)}$	$H_{I(i)}$	$H_{S(i)}$	$H_{T(i)}$	$F_{IS(i)}$	$F_{ST(i)}$	$F_{IT(i)}$
Loc I	0.5	0.5	0.5	0.5	0.5	0.0	0.0	0.0
Loc II	0.2	0.2	0.2	0.46	0.48	0.565	0.042	0.583
Loc III	0.7	0.6	0.65	0.4675	0.46875	-0.39	0.0027	-0.387
Loc IV	0.0	0.0	0.0	0.0	0.5	---	1.0	1.0
Průměr						0.058	0.261	0.300

Průměrné hodnoty F statistik mohou maskovat odlišnou evoluční historii na různých lokusech

# Genetická struktura populací

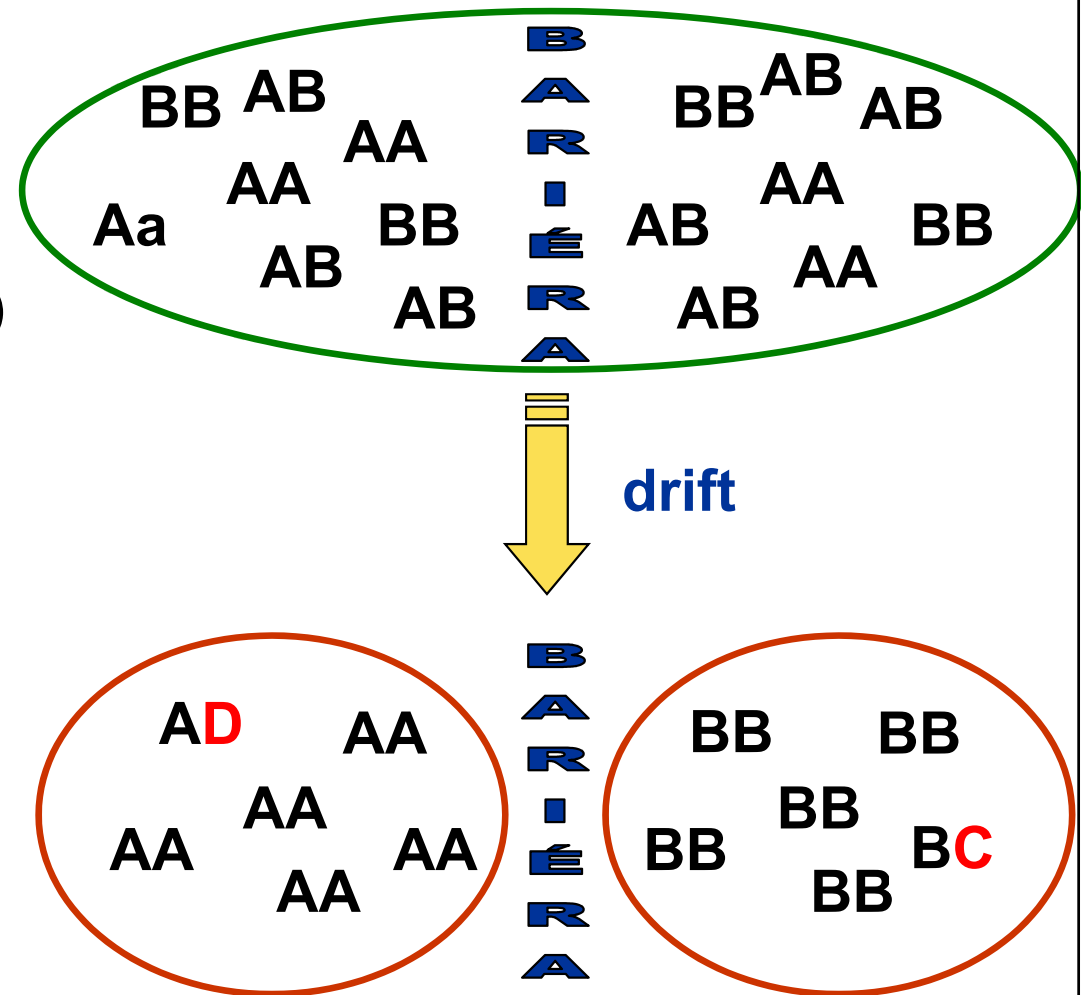
## drift, mutace

- Drift**

→ diferenciaci subpopulací díky změnám frekvencí (až fixaci) alternativních alel

- Mutace**

mohou zvýšit diferenciaci (nebo ne – homoplázií)



### Migrace (genový tok)

- působí proti diferenciaci subpopulací

# Wahlundův princip - příklad

- Jezero Bunnersjöarna (severní Švédsko) – „brown trout“
- 2 alely na jednom znaku

	A/A	A/B (= Ho)	B/B	Total	p	2pq (=He)
Přítok	50	0 (0)	0	50	1.000	0.000
Odtok	1	13 (0.26)	36	50	0.150	0.255
Celé jezero (expected)	51 (33.1)	13 (0.13) (48.9)	36 (18.1)	100	0.575	0.489

$$p^2 = 0.575^2$$

$$q^2 = 0.425^2$$



# F-statistiky

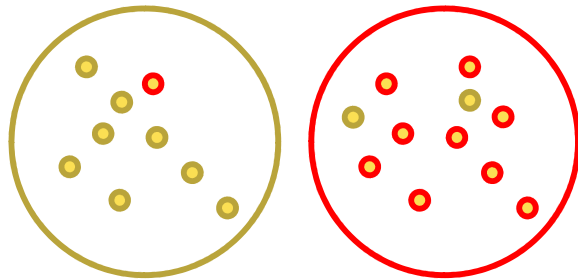
- $F_{IS}$  snížení heterozygotnosti v lokální subpopulaci  
vysoké hodnoty – inbreeding

- $F_{IT}$  souhrnná hodnota – příliš se nepoužívá

- $F_{ST}$  míra „rozdělenosti“ = snížení toku mezi subpopulacemi (tj. existence bariéry – Wahlundův princip)  
Vliv driftu – fixuje odlišné alely v subpopulacích

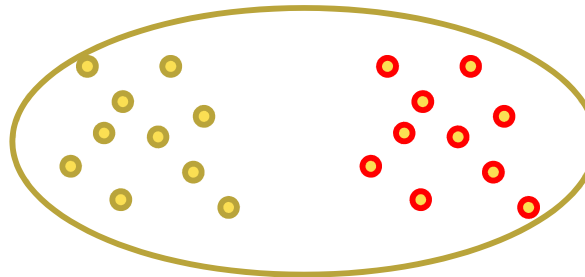
# Permutační test významnosti Fst

1. Skutečné populace

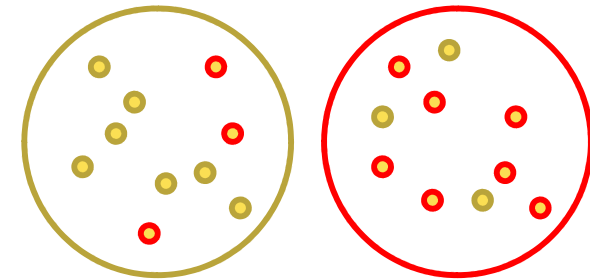


Reálné Fst

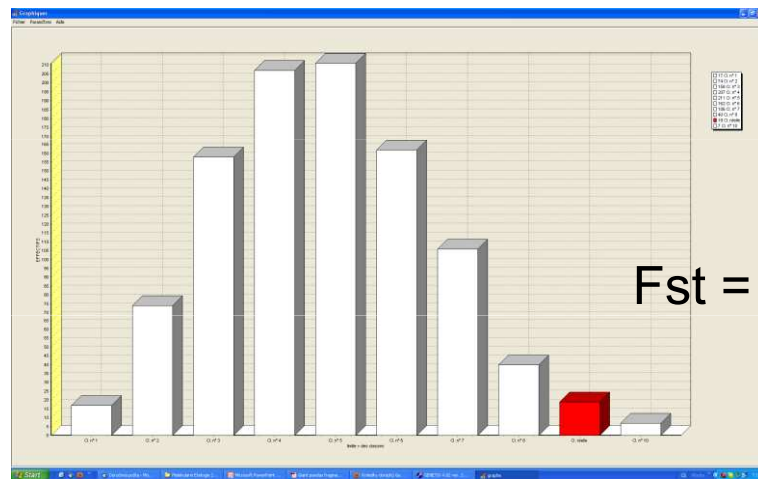
2. Sloučení jedinců



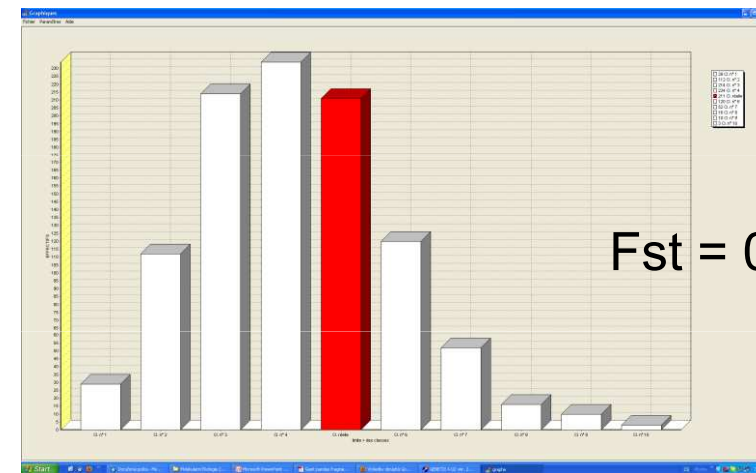
3. 1000 x náhodně vytvořené populace



1000 x nasimulované Fst za nepřítomnosti bariéry



0,80 % simulovaných hodnot větších než skutečné Fst  
 $p = 0,008$  (tj. významný rozdíl)



35,40 % simulovaných hodnot větších než skutečné Fst  
 $p = 0,354$  (tj. nevýznamný rozdíl)



# Výpočet $F_{ST}$ - příklad

	A/A	A/B (=H <sub>o</sub> )	B/B	Total	p	2pq (=H <sub>e</sub> )
Přítok	50	0 (0)	0	50	1.000	0.000
Odtok	1	13 (0.26)	36	50	0.150	0.255
Celé jezero	51	13 (0.13)	36	100	0.575	0.489
(expected)	(33.1)	(48.9)	(18.1)			

$$F_{ST} = \frac{H_T - \overline{H_s}}{H_T} = \frac{0.489 - 0.128}{0.489} = 0.728$$

V důsledku bariéry toku genů je heterozygotnost o 72.8% nižší než by byla v panmiktické populaci



# Zdroje chyb při analýze Fst

## Global vs. pairwise indices

Absolutní hodnoty závisí na heterozygotnosti !!!

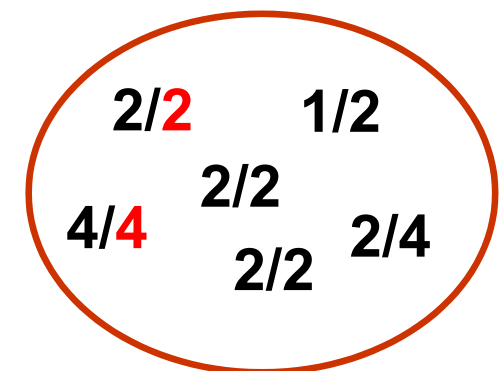
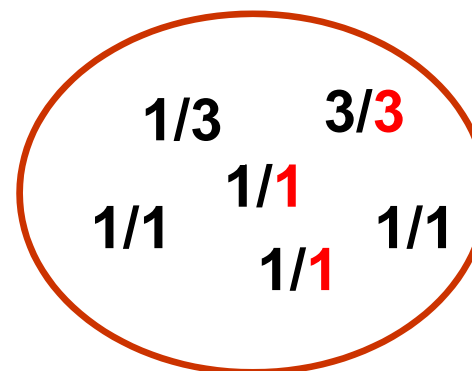
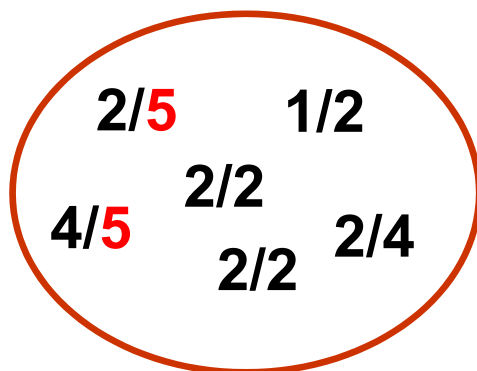
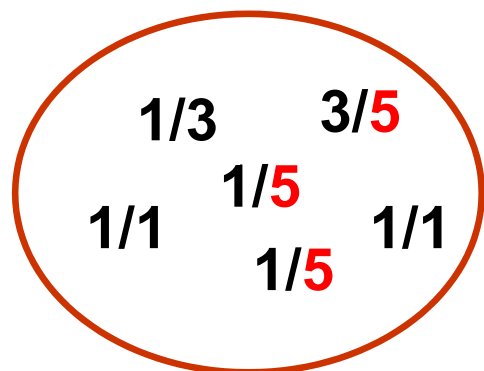
(např. nelze srovnávat alozomy s mikrosatelity)

Nutno standardizovat:  $F_{ST}' = F_{ST}/F_{STmax}$  (Hedrick 2005)

– např. GenAlex

Nutno korigovat na přítomnost nulových alel

(většinou uměle zvyšují Fst – zvýšení homozygotnosti) - FreeNA



# Panda velká



- 192 vzorků trusu → 136 genotypů → 53 unikátních genotypů
- rozdělení řekou (cca 26 tisíc let) a cestami (recentně)
- i cesty jsou významná bariéra, i když menší

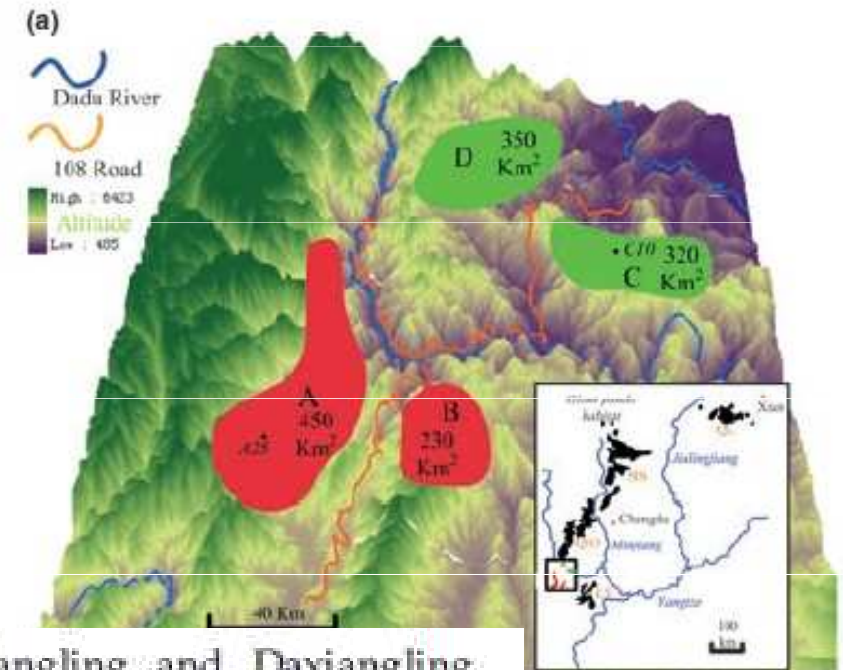


Table 3 Pairwise  $F_{ST}$  in the Xiaoxiangling and Daxiangling populations

Patch	A	B	C	D
A				
B	0.033*			
C	0.107*	0.062*		
D	0.107*	0.097*	0.037*	

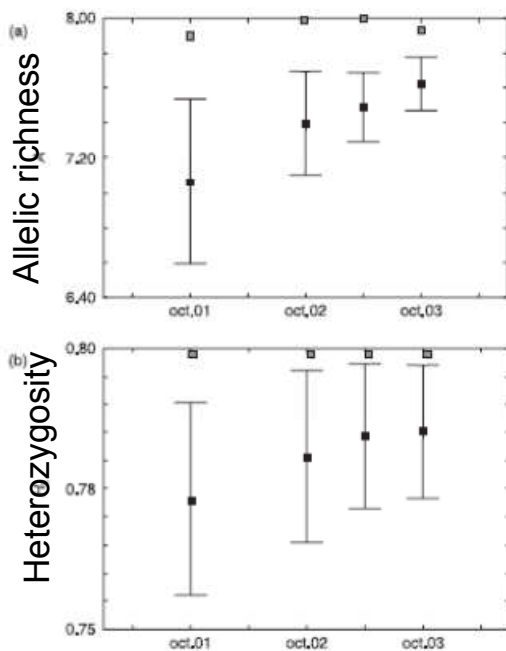
\*Significant level after Bonferroni correction ( $P < 0.01$ ).

# Populační cykly hrabošovitých hlodavců (Berthier et al. 2006)

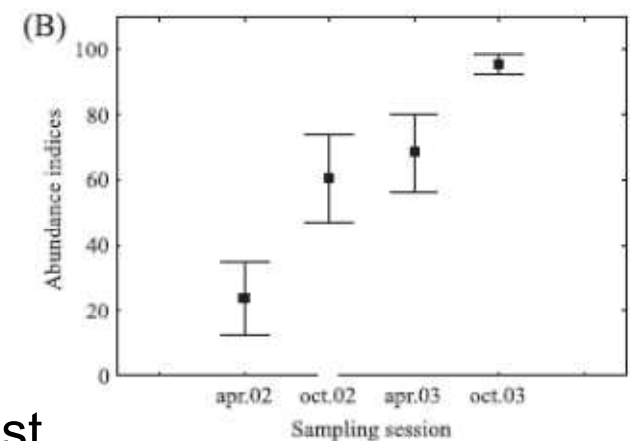
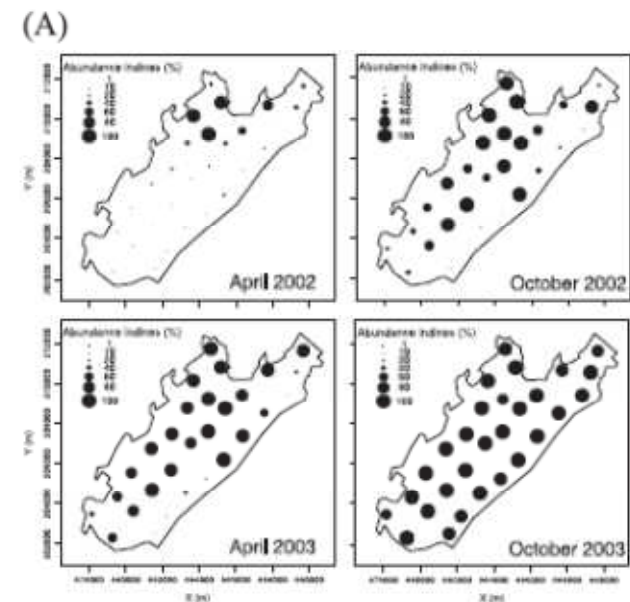


*Arvicola terrestris*

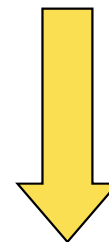
- Výrazné cyklické změny v početnosti populací – střídají se období „izolovanosti“ s „kontinuálním rozšířením“



- Nízká hustota – extrémní vliv driftu
- S nárůstem populační hustoty se zvyšuje intenzita toku genů mezi subpopulacemi a narůstá genetická variabilita subpopulací



Sampling session	Min. pairwise $F_{ST}$	Max. pairwise $F_{ST}$	Overall $F_{ST}$
October 2001	0.018	0.058	0.037
October 2002	0.010	0.047	0.026
April 2003	0.002	0.033	0.020
October 2003	0.007	0.035	0.017



Snižování  $F_{ST}$

# $G_{ST}$ (Nei 1973)

- Analogie  $F_{ST}$  - Haploidní (haplodiploidní) organismy, sekvence mtDNA
- Počítá s „haplotype (gene) diversity“ místo heterozygotnosti
- *Haplotype diversity* = pravděpodobnost, že dvě náhodně vybrané sekvence z populace budou odlišné
- Pracuje tedy jen s frekvencemi alel, ne s procentem heterozygotů



# *Myotis bechsteinii*

*Kerth et al. 2002*

**Table 2** Genetic differentiation based on nuclear and mitochondrial DNA among 10 maternity colonies of the Bechstein's bat

Loci	No. of alleles	$F_{ST} \pm SE$	Population differentiation
Nuclear DNA			
b15	23	$0.023 \pm 0.009$	$P < 0.0005$
b22	12	$0.003 \pm 0.005$	$P = 0.07$
b23	20	$0.008 \pm 0.007$	$P = 0.02$
MM5	10	$0.010 \pm 0.006$	$P = 0.01$
NNS	10	$0.031 \pm 0.012$	$P = 0.001$
p20	20	$0.014 \pm 0.007$	$P < 0.0005$
paur3	10	$0.012 \pm 0.008$	$P = 0.002$
All		$0.015 \pm 0.003$	$P < 0.0005$
Mitochondrial DNA			
AT-1	10	$0.658 \pm 0.058$	$P < 0.0005$
AT-2	3	$0.961 \pm 0.050$	$P < 0.0005$

- Letní kolonie samic (15-40 jedinců)
- Vzorky - jen ♀♀
- Nukleární i mitochondriální mikrosatelity
- ♀♀ filopatrie
- ♂♂ disperze

Wright's fixation index  $F_{ST}$  ( $\pm SE$ ) was calculated for seven nuclear and two mitochondrial microsatellite loci. Significant deviations from zero, indicating population differentiation, were assessed using permutation tests (see Materials and methods for details).

# $R_{ST}$

- Obdoba  $F_{ST}$
- Pracuje však s velikostí alel (počet repeatů u mikrosatelitů)
- Předpoklad známého mutačního modelu jen při platnosti SMM (*stepwise mutation model*)
- „Paměť“ mutací v minulosti
  - $R_{ST} > F_{ST}$  větší vliv mutací
  - $R_{ST} = F_{ST}$  větší vliv driftu
- Potvrzení významnosti rozdílu randomizačními testy (Hardy et al. 2003, program SPAGeDi 1.1)

# AMOVA

*Excoffier et al. 1992*

Arlequin ver. 2.000

A software for  
population genetics  
data analysis

Authors:

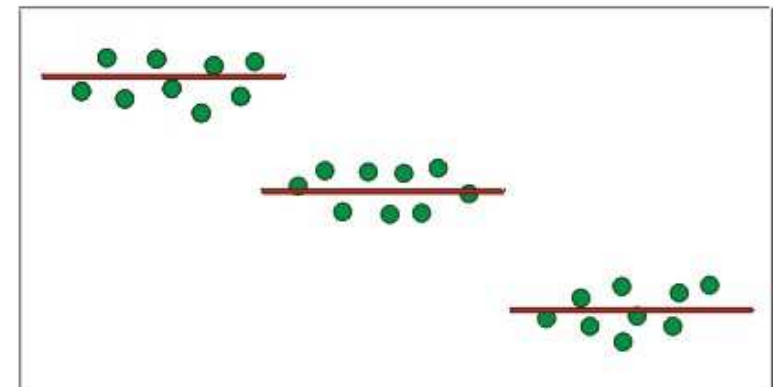
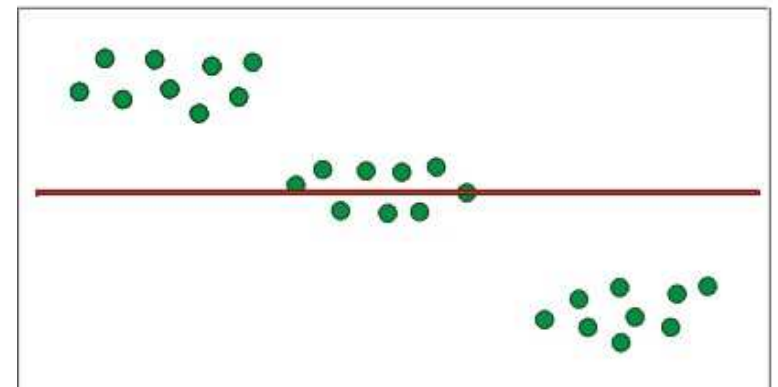
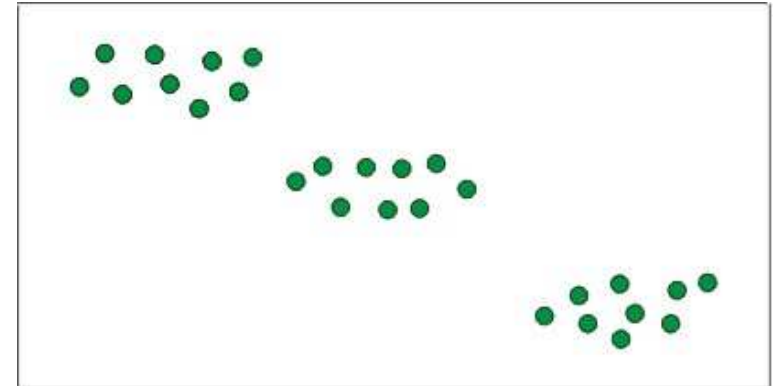
Stefan Schneider  
David Roessli  
Laurent Excoffier

Contact Arlequin:

Url: <http://anthropologie.unige.ch/arlequin/>  
Mail: [arlequin@sc2a.unige.ch](mailto:arlequin@sc2a.unige.ch)



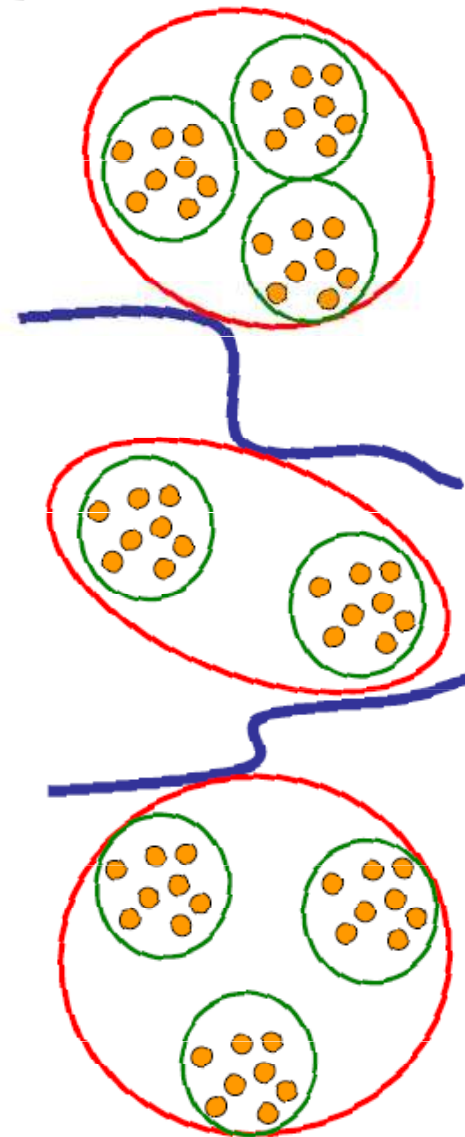
- **A**nalýza **M**olekulární **V**ariance
- Analýza variance alelických frekvencí  
(již dříve Cockerham & Weir  
1987, 1993)
- Započítává se ale rozdílnost (mutace)  
alel
- Program ARLEQUIN
- Data:  
sekvence  
mikrosatelity (jen při platnosti SMM  
*stepwise mutation model*)





# Hierarchická AMOVA

- Zjišťuje kolik variability vysvětlí:
- Rozdělení na velké skupiny populací
- Rozdělení na populace v rámci skupin
- Rozdíly mezi jedinci



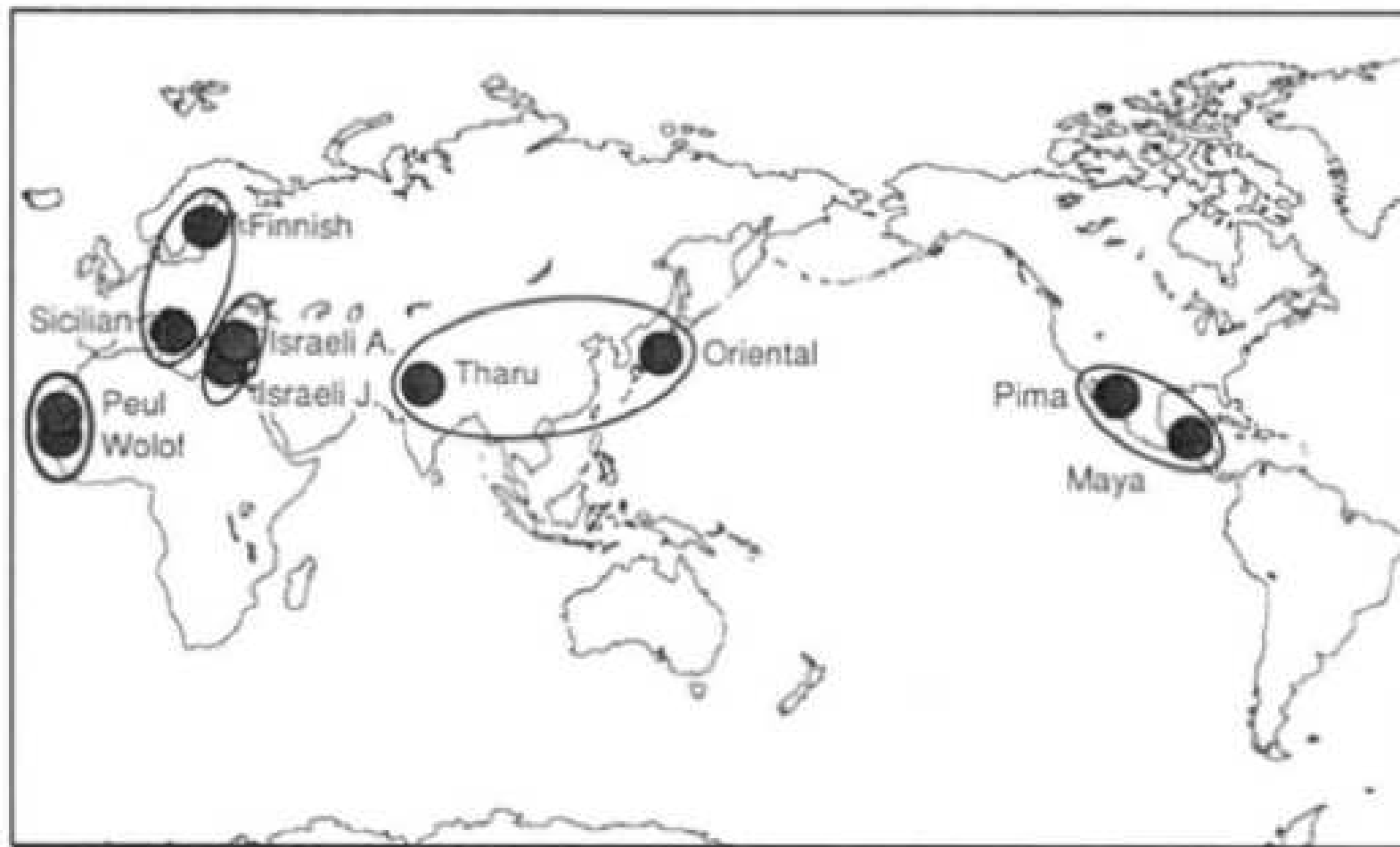
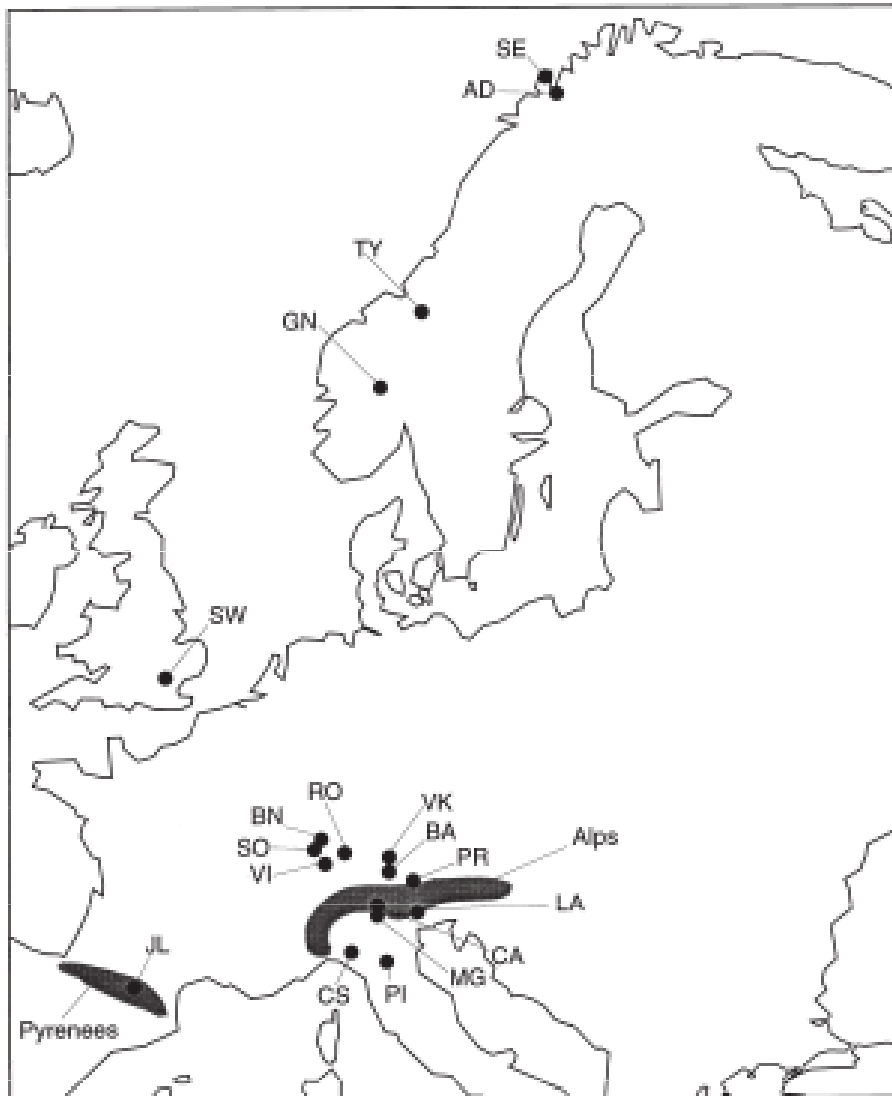


FIGURE 2.—Geographic location of the population samples.

Variance component		Variance	% total	$p^a$	$\Phi$ -statistics
Among regions	$\sigma_c^2$	0.134	21.12	0.002	$\Phi_{CT} = 0.211$
Among populations/regions	$\sigma_b^2$	0.022	3.49	<0.0001	$\Phi_{SC} = 0.044$
Within populations	$\sigma_i^2$	0.478	75.39	<0.0001	$\Phi_{ST} = 0.246$

# *Bombus pascuorum*

Widmer & Schmid-Hempel 1999



	$F/\Phi$	d.f.	SSD†	Variance component	% Total variance*
Among populations	$F$	17	77.71	0.07	4.51*
	$\Phi$	17	5198.20	5.02	8.74*
Among regions	$F$	4	56.15	0.08	5.16*
	$\Phi$	4	3464.94	4.58	7.49*
Among populations within regions	$F$	11	24.35	0.02	1.11*
	$\Phi$	11	1773.71	2.16	3.53*
Between north and south of Alps	$F$	1	38.57	0.11	7.12*
	$\Phi$	1	2622.89	7.25	11.74*
Among populations north and south of the Alps, respectively	$F$	16	39.14	0.02	1.46*
	$\Phi$	16	2575.31	2.18	3.53*

†Sum of squared deviations.

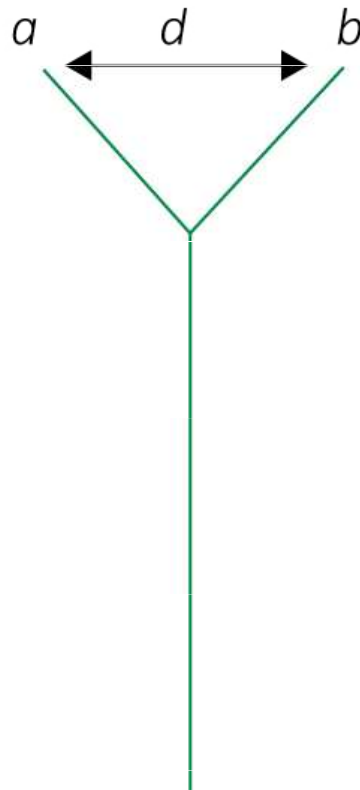
\*  $P < 0.001$ .

**Mikrosatelity, AMOVA  
Nejvíc vysvětlují Alpy**

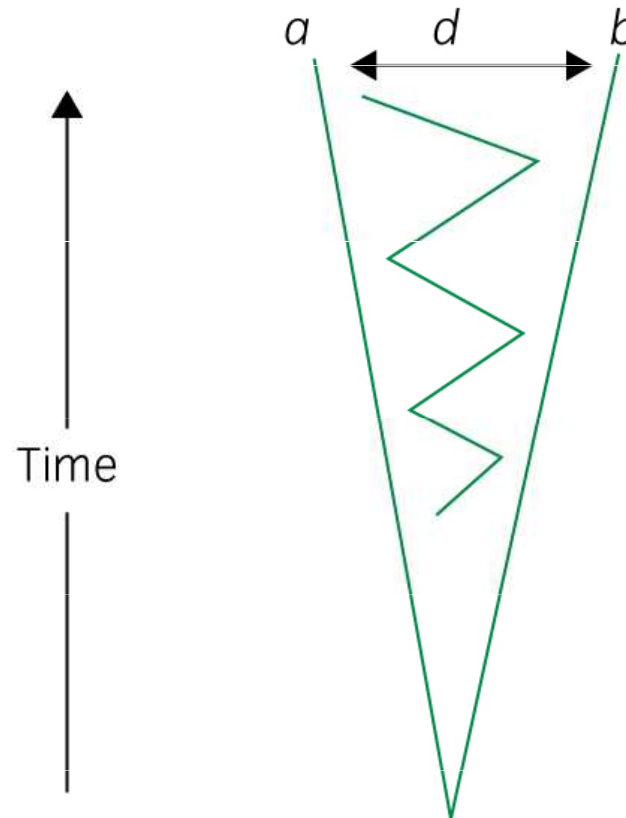
# AMOVA a F statistika

popis výsledku nikoliv příčin → možná alternativní vysvětlení  
(je možno je odlišit metodami založenými na koalescenci a fylogenezi alel)

Recent separation,  
no gene flow



Old separation, but  
continuous (low)  
gene flow



Time

# Tok genů – metody

1. **Přímé metody:** Capture-Mark-Recapture + parentage analýzy

## 2. Nepřímé metody stanovení toku genů – populačně-genetické modely

- **Island model** (Wright)

Stejně velké subpopulace  
Symetrický tok genů  
Stejná pravděpodobnost výměny  
mezi jakýmkoliv subpopulacemi



- **Stepping-stone model**  
(Kimura)

Výměna jen mezi sousedními  
subpopulacemi



# Tok genů – metody

1. **Přímé metody**: Capture-Mark-Recapture + parentage analýzy

2. **Nepřímé metody** – na základě distribuce genetické variability

$N_e m$  = počet migrantů / subpopulace / generace („island model“!)

Jde o hrubý odhad ve škále: málo, středně, hodně!

- **Privátní alely** (Slatkin 1985) – vhodné pro vysoce polymorfní znaky  
Alely vyskytující se jen v jedné subpopulaci

$p(1)$  frekvence privátních alel

$$\ln p(1) = -0,505 \ln(N_e m) - 2.44$$

- **F statistika** 
$$F_{ST} = \frac{1}{1 + 4N_e m}$$
 (jen pro  $F_{ST} > 0.05-0.10$ )

## Předpoklady použití $N_e m$ :

- „island model“ (= infinite number of populations, absence of selection, the same size of all populations etc.)
- migration-drift equilibrium (= no range expansions, habitat fragmentation or population bottlenecks)

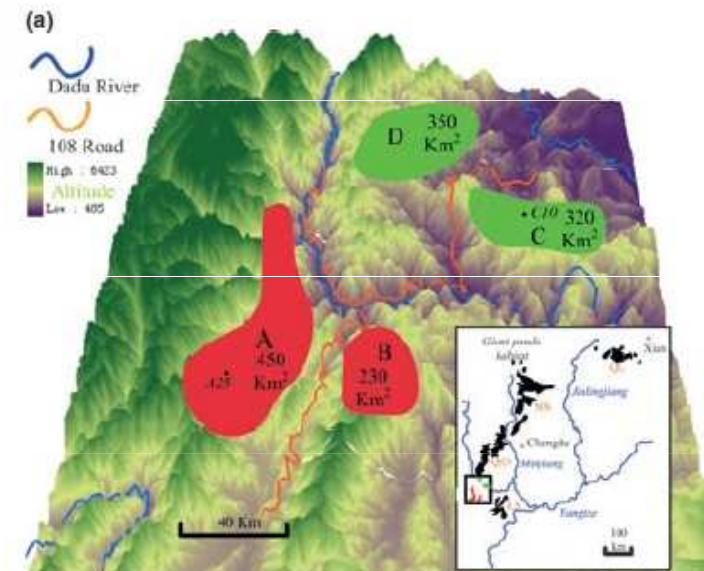
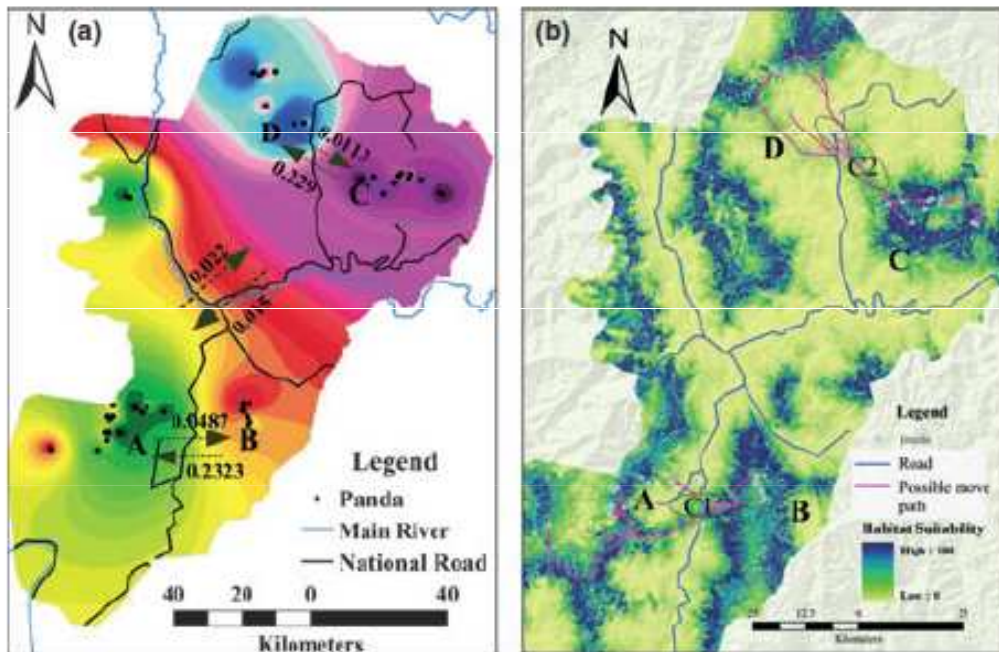
## Co ale rozhodně NE!

- Dvě hodně vzdálené populace
- $F_{ST} \rightarrow$  vždy bude nenulové  $N_e m \rightarrow$  v minulosti došlo k výměně jedinců
- I populace, které si nikdy nevyměnily migranty mohou mít nenulové  $N_e m$



# BayesASS

- Panda velká



- Bayesian estimates of gene flow
- doporučení pro ochrannářský management – stavba koridorů

# Modely toku genů (alel)

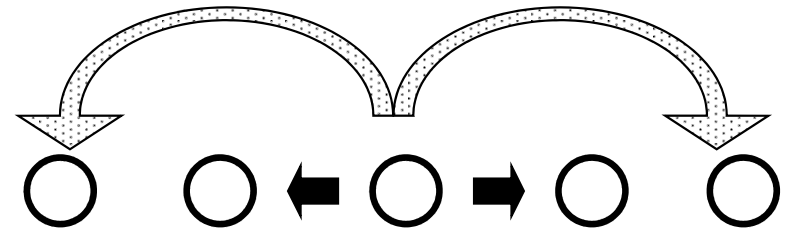
- **Island model** (Wright)  
Stejně velké subpopulace  
Symetrický tok genů  
Stejná pravděpodobnost výměny  
mezi jakýmkoliv subpopulacemi



- **Stepping-stone model** (Kimura)  
Výměna jen mezi sousedními  
subpopulacemi



- **Isolation by distance**  
Tok slábne se vzrůstající  
vzdáleností subpopulací



# Isolation by distance

- rozumné geografické měřítko  
(závisí na schopnosti disperze)
- musí být ustanovena rovnováha mezi migrací a driftem
- IBD (isolation-by-distance) nebude
  - u velmi recentně izolovaných populací
  - u zcela izolovaných populací
  - při značné migraci



# Isolation by distance

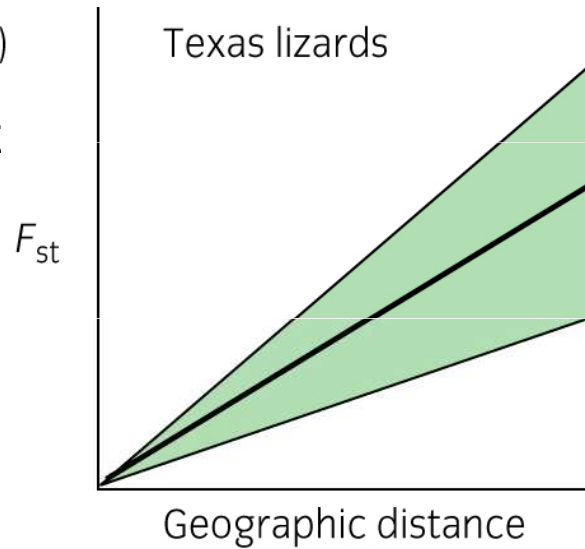
*Crotaphytus collaris*

Hutchinson & Templeton 1999



(a)

desítky tisíc let  
nejsou bariéry  
**rovnováha**  
mezi driftem  
a migrací



# Detekce IBD

- korelace mezi maticí genetických a geografických vzdáleností
- Mantelův test
- např. Genepop

# Příklad: IBD u netopýrů rodu *Pipistrellus*

## Mechanisms of radiation in a bat group from the genus *Pipistrellus* inferred by phylogeography, demography and population genetics

P. HULVA,\* A. FORNŮSKOVÁ,† A. CHUDÁRKOVÁ,\* A. EVIN,‡ B. ALLEGRINI,§ P. BENDA\*¶ and J. BRYJA†

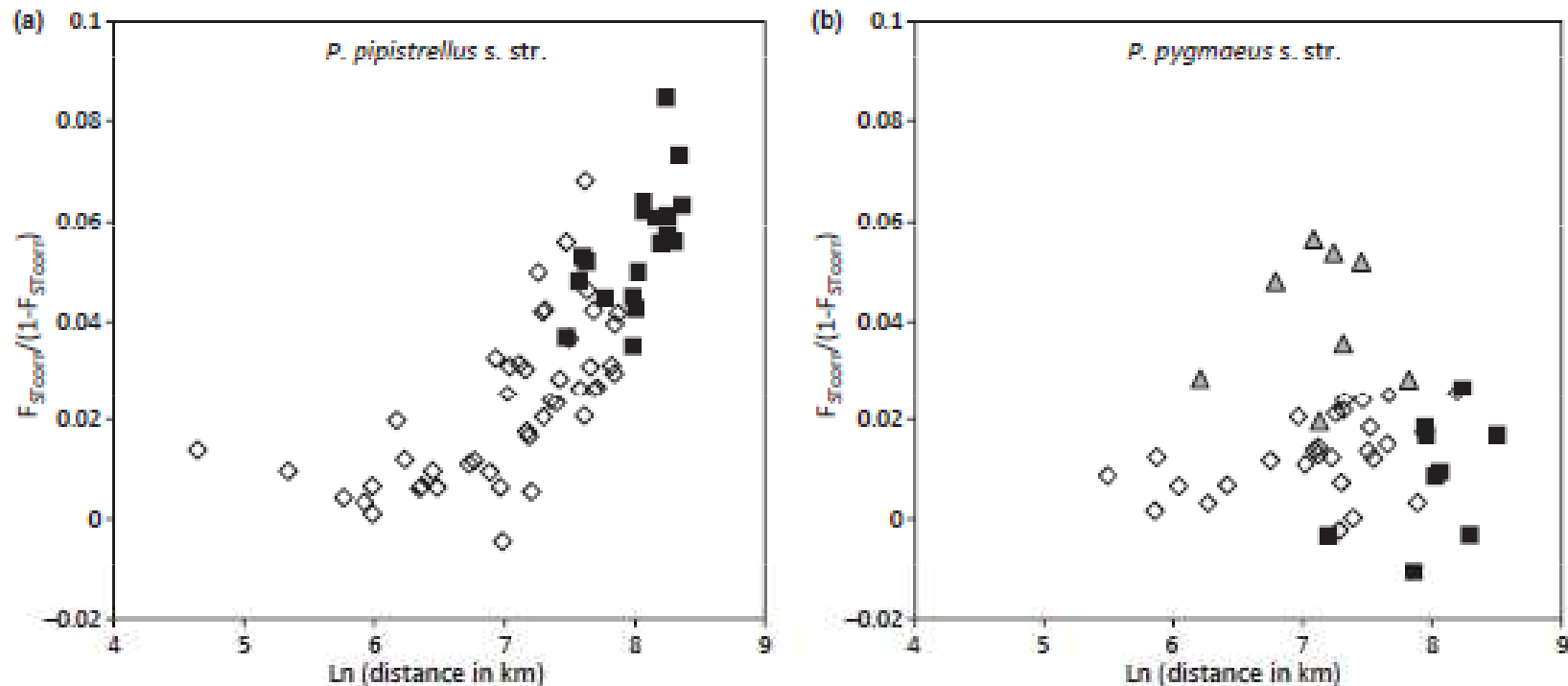
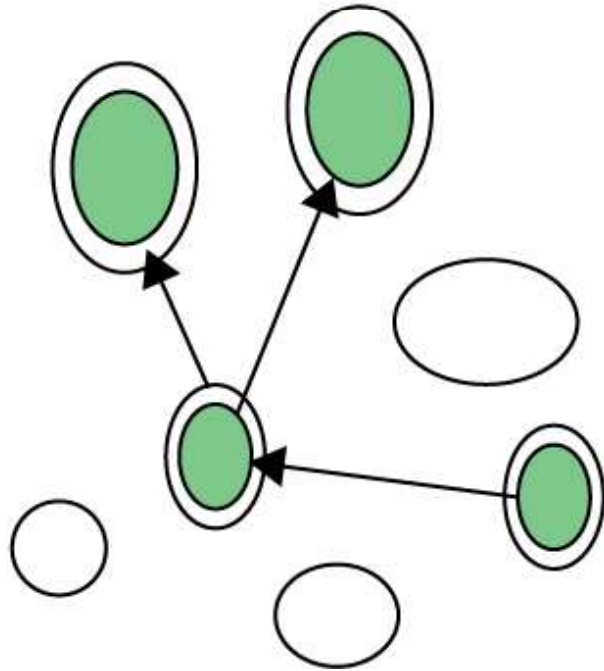


Fig. 5 Correlation between genetic  $F_{STcorr}/(1-F_{STcorr})$  and geographic ( $\ln$  scale) distances for *Pipistrellus pipistrellus* s.str. (a) and *Pipistrellus pygmaeus* s.str. (b). Black squares, population pairs including Iran (both species) and central Asia (only *P. pipistrellus* s.str.). Grey triangles, population pairs including Corsica.

# Genetika metapopulací



**Metapopulace:** časté lokální extinkce a kolonizace

- **Jednoduchá populační struktura vs. metapopulace** → populační genetika
- stabilní vs. nestabilní populačně-genetický pattern

# Atlantic salmon – Sainte Marguerite river (Quebec) (Garant et al. 2000)

- výrazná populační struktura mezi řekami
- v rámci jedné řeky – různé mikrohabitaty obývané (1) **demy** nebo (2) **metapopulacemi**?
- dvouleté sledování, mikrosatelity, hierarchická AMOVA
- lokálně adaptované demy, ale i metapopulace (např. po povodních je mikrohabitat obsazen náhodně a ne jedinci z původního demu)
- long-term (stable) vs. short-term (metapopulations) struktura



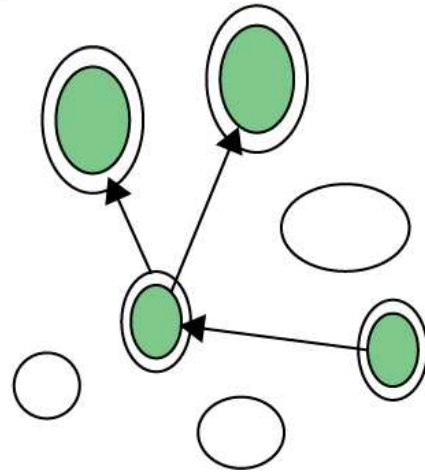
Variance component	D.f.	% of total variance	F-statistic	P
Mezi přítoky	1	0,1	0,0014	0,167
Mezi lokalitami	6	0,9	0,0085	<b>0,001</b>
Mezi roky uvnitř lokality	7	2,5	0,0255	<b>&lt; 0,0001</b>
Uvnitř vzorku	1352	96,6	0,0337	<b>&lt; 0,0001</b>



# Který typ metapopulační struktury?

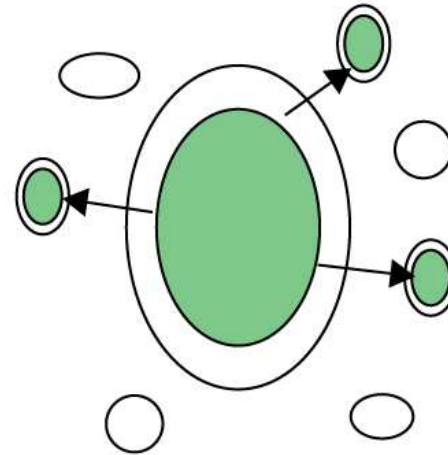
„classical“

(a)



„mainland-island“

(b)

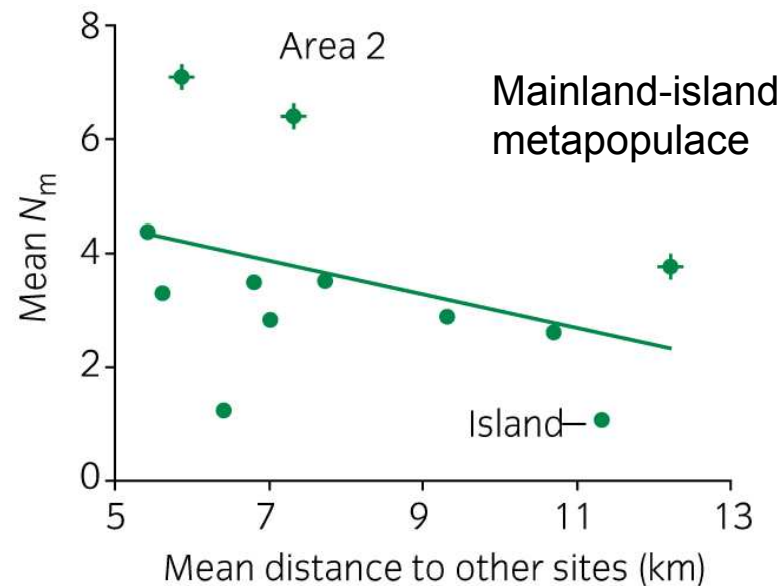
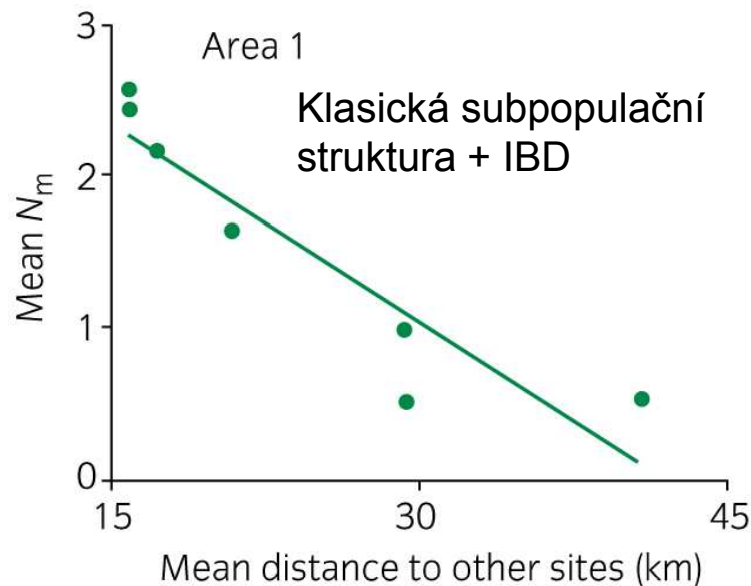


„source and sinks“

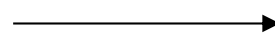
- ad a) *Cristatella mucedo* - Bryozoa (Freeland et al. 2000) – long-distance migration (birds), no IBD, 5 ms for clone identification
- black rats in Senegal – transport by traffic – no IBD (Konečný et al. 2012)

# ad b) Mainland-island model

- *Bufo calamita* (Rowe et al. 2000) – různé rybníky mají různý reprodukční úspěch
- klasické subpopulace (Area 1) i metapopulace (Area 2); křížek – vyšší tok genů než by se předpokládalo (jsou v nich obsaženy vysoce úspěšné rybníky = „mainland“)



$$F_{ST} = 0.06-0.22$$



$$F_{ST} = \frac{1}{1 + 4N_m}$$



# Populační struktura - shrnutí

	Connected populations (gene flow)	Isolated populations (no gene flow)
$N_e$	<input type="checkbox"/>	<input type="checkbox"/>
Genetic drift	<input type="checkbox"/>	<input type="checkbox"/>
Genetic diversity	<input type="checkbox"/>	<input type="checkbox"/>
Population differentiation	<input type="checkbox"/>	<input type="checkbox"/>

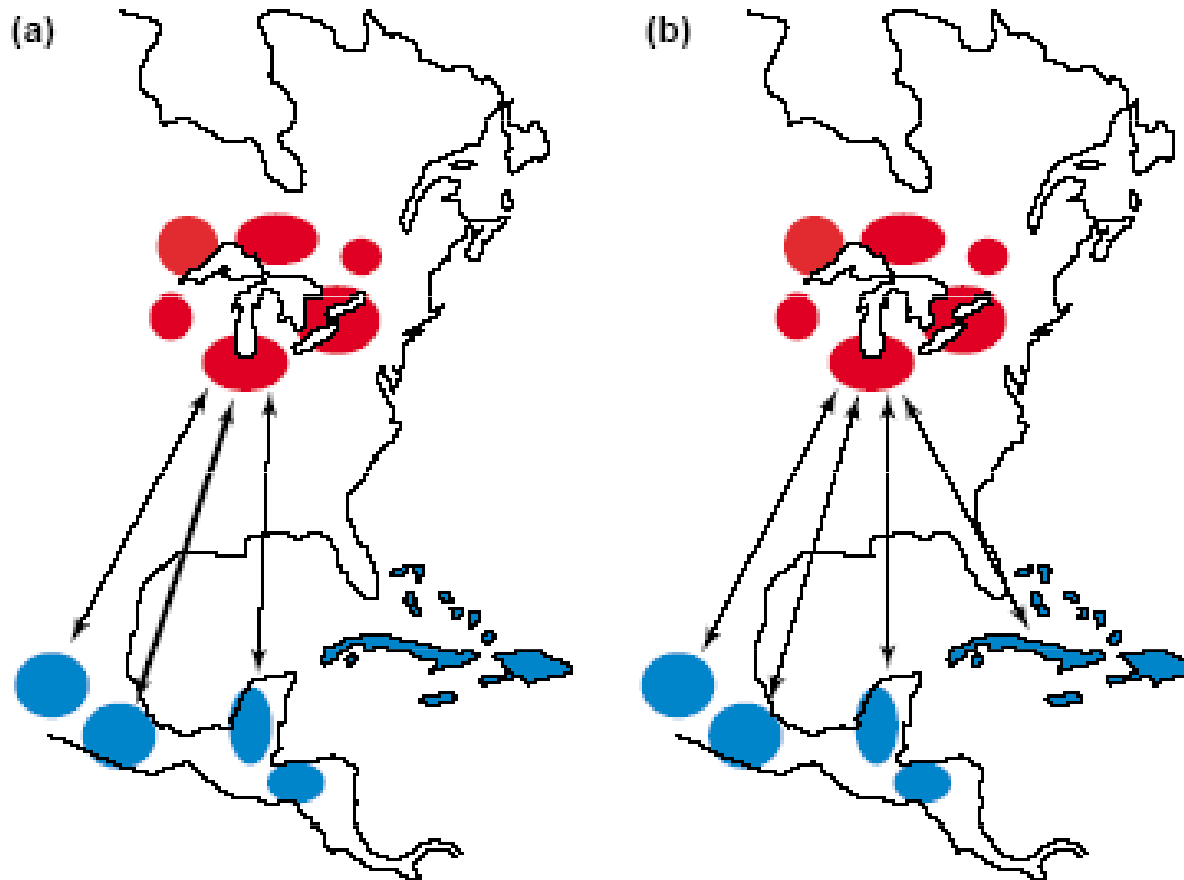
1. vnitropopulační variabilita – deskriptivní statistiky
2. popis populačně-genetické struktury, bariéry toku genů

## 3. „Population assignments“

### Klasické problémy populační genetiky

- Populace dány, jedinci předem zařazeni do populací, zajímají nás vlastnosti populací (F-statistiky) → ad (1) nebo (2)
- Populace sice definovány, ale chceme k nim přiřadit jedince neznámého původu
- Kryptická populační struktura = předem není dáno nic → chci zjistit klastry (tj. přirozené populace) a rozřadit individua do klastrů (**population assignments**)

# Př.: Unraveling migratory connectivity



(a) strong vs. (b) weak connectivity

## A. Direct methods

- morphological variation (geographical races)
- leg-bands or similar markers (ex. over one million *Ficedula hypoleuca* have been ringed in UK and Sweden – only six recaptured on wintering grounds in Africa)
- satellite telemetry – expensive, not useful for small animals

## B. Biogeochemical approaches

- ratios of stable isotopes of naturally occurring elements (C, H, N, Sr) vary across the landscape
  - determined by the relative frequency of C3 and C4 plants, climate, and bedrock
- (1) geographical structure of isotopic ratio distributions
  - (2) knowledge about where animals incorporate isotopes
  - (3) tissue samples from individuals at different parts of their annual cycle

# C. Genetic approaches

- « málo ptáků má kroužky, ale všichni mají genotypy »
- genetická data o populační struktuře
- problémy: (1) malá genetická diferenciace mezi populacemi (intenzivní disperze), (2) malá diferenciace v temperátní zóně – recentní postglaciální kolonizace
- Řešení: (1) použití více genetických znaků, (2) studium DNA parazitů (např. avian malaria) – mají rychlejší evoluci, jsou více diferencovaní

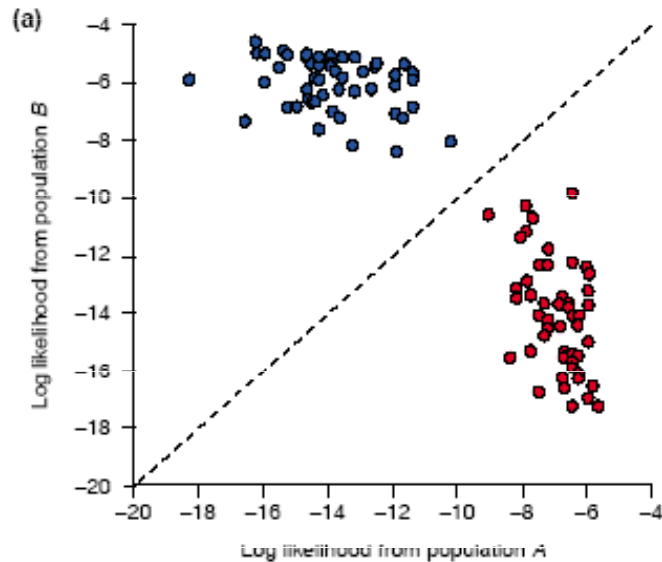
# Population assignment tests

- program **GeneClass** (Piry et al. 2004)
- počítá pravděpodobnost, že se určitý genotyp může vyskytovat v určité předem definované populaci – identifikace recentních migrantů nebo zvířat neznámého původu (pytláctví)
- může kombinovat data z různých genetických markerů

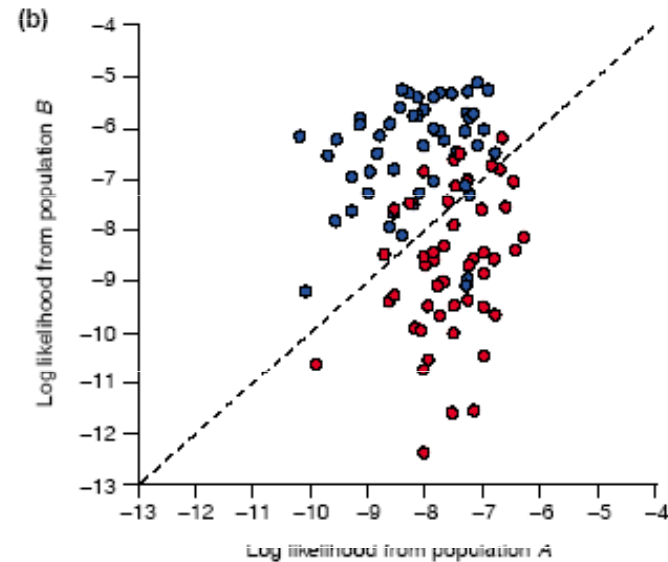




# Závisí na míře genetických rozdílů mezi populacemi



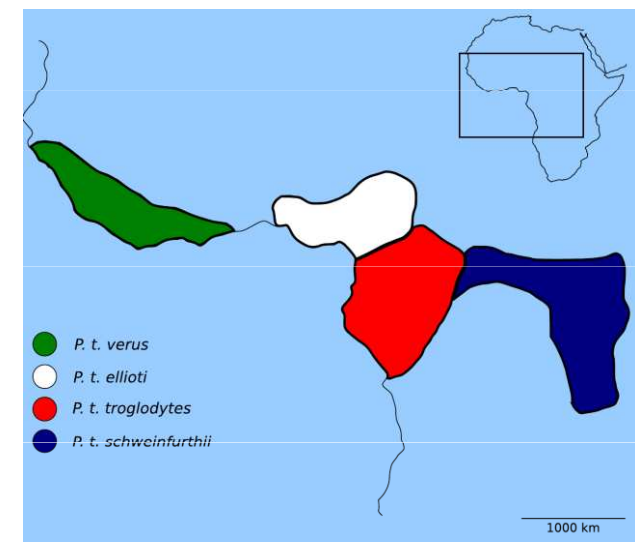
5 microsatellite loci  
 $F_{st} = 0.14$   
99.9% assigned correctly



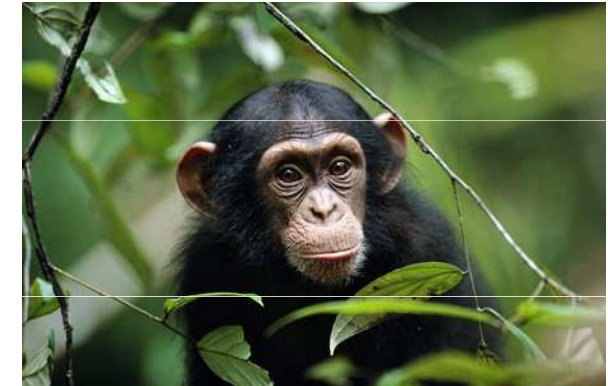
5 microsatellite loci  
 $F_{st} = 0.04$   
90.2% assigned correctly

# Poddruhová identifikace šimpanzů v českých ZOO

- šimpanzi v Zoo často nejasného původu
- genetická data z přírodních populací jsou dostupná (300 msats, Becquet et al. 2007)
- 30 nejvíce informativních mikrosatelitů – genotypizace všech šimpanzů v ČR
- GeneClass – přiřazení k jednotlivým poddruhům



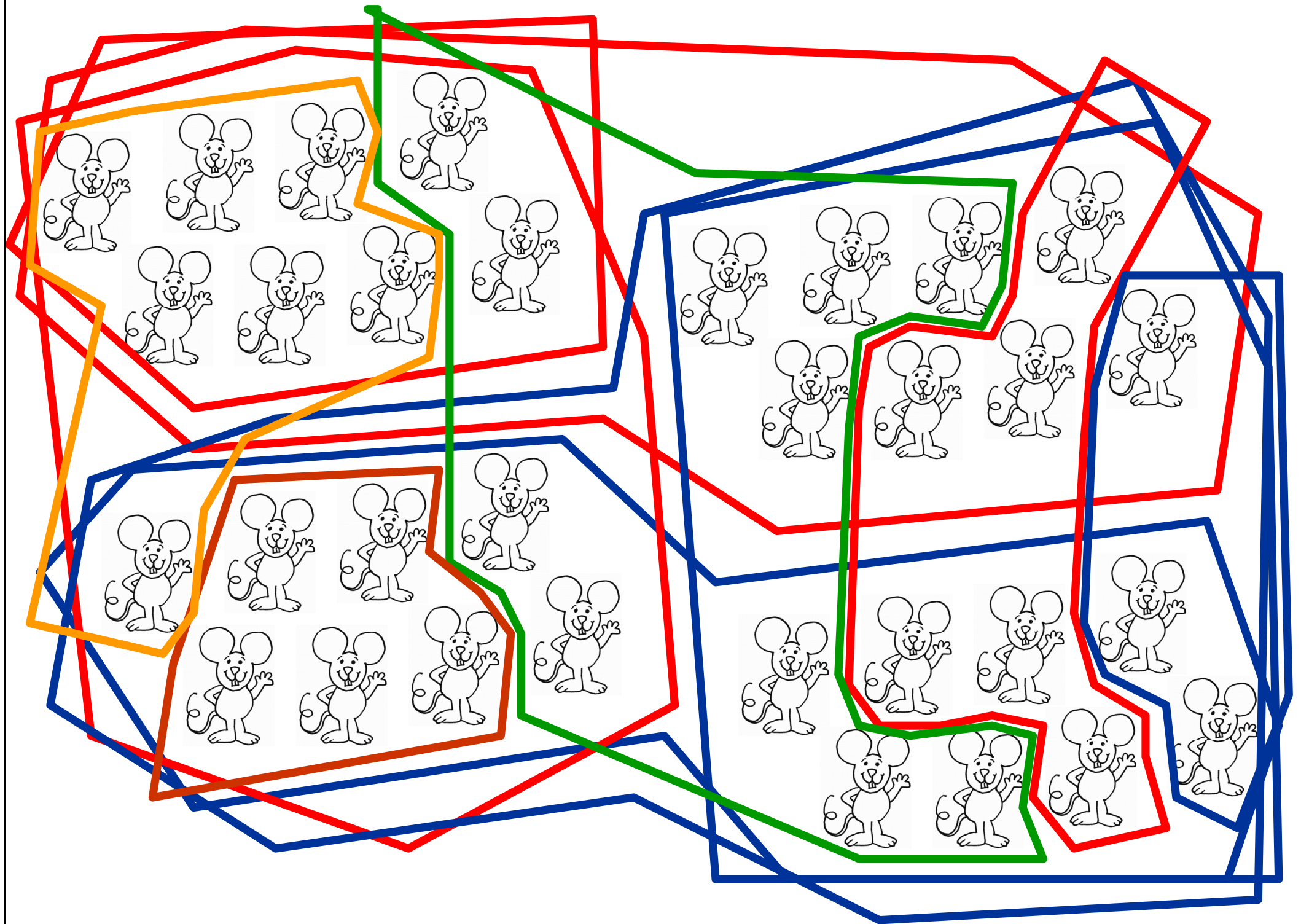
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		12	loci			27	loci			30	loci			
2		rank	score	rank	score	rank	score	rank	score	rank	score	rank	score	
3	Assigned samp	1	%	2	%	1	%	2	%	1	%	2	%	
9	77-pop5-60	Pop1	100	Pop2	0	Pop1	100	Pop2	0.001	Pop1	100	Pop2	0.004	
10	78-pop5-67	Pop1	100	Pop4	0.001	Pop2	80.65	Pop1	19.35	Pop2	99.76	Pop1	0.239	
11	Bamia	Pop2	57.13	Pop1	42.87	Pop1	100	Pop2	0	Pop1	100	Pop2	0	
12	Babeta	Pop1	95.21	Pop2	4.786	Pop1	98.17	Pop2	1.829	Pop1	64.26	Pop2	33.35	
13	Bambari	Pop2	94.66	Pop1	5.3	Pop1	84.77	Pop2	15.23	Pop2	83.34	Pop1	16.66	
14	Bonie	Pop1	100	Pop2	0	Pop1	100	Pop2	0	Pop1	100	Pop2	0	
15	Carl	Pop4	99.26	Pop2	0.645	Pop1	99.98	Pop3	0.019	Pop1	99.72	Pop3	0.268	
16	Cindy	Pop4	99.98	Pop1	0.022	Pop3	89.59	Pop4	8.614	Pop4	89.06	Pop3	10.19	
17	Dadula	Pop4	99.58	Pop1	0.415	Pop1	67.47	Pop4	32.53	Pop1	92.15	Pop4	7.854	
18	Dais	Pop1	92.04	Pop2	7.957	Pop1	100	Pop4	0	Pop1	100	Pop4	0	
19	Dingo	Pop1	100	Pop2	0.003	Pop1	98.98	Pop4	0.98	Pop1	99.84	Pop2	0.102	
20	Dorka	Pop1	99.34	Pop2	0.399	Pop2	99.48	Pop4	0.46	Pop2	99.67	Pop1	0.326	
21	Faben	Pop4	100	Pop2	0.001	Pop2	95.76	Pop1	4.236	Pop2	98.34	Pop4	0.874	
22	Gina	Pop2	99.26	Pop1	0.736	Pop2	71.24	Pop1	28.77	Pop1	52.48	Pop2	47.53	
23	Hope	Pop2	99.08	Pop1	0.918	Pop2	100	Pop1	0.001	Pop2	100	Pop1	0	
24	Ingridy	Pop3	56.69	Pop1	43.31	Pop3	99.52	Pop1	0.484	Pop3	99.93	Pop1	0.072	
25	Jakub	Pop1	99.99	Pop2	0.015	Pop1	100	Pop2	0	Pop1	100	Pop3	0	
26	Janis	Pop4	99.42	Pop3	0.499	Pop3	95.23	Pop2	4.756	Pop3	86.24	Pop4	9.103	
27	Jimmy	Pop4	99.42	Pop1	0.565	Pop1	100	Pop2	0	Pop1	100	Pop2	0	
28	Judy	Pop1	99.84	Pop4	0.158	Pop1	82.71	Pop3	17.29	Pop3	97.73	Pop1	2.273	

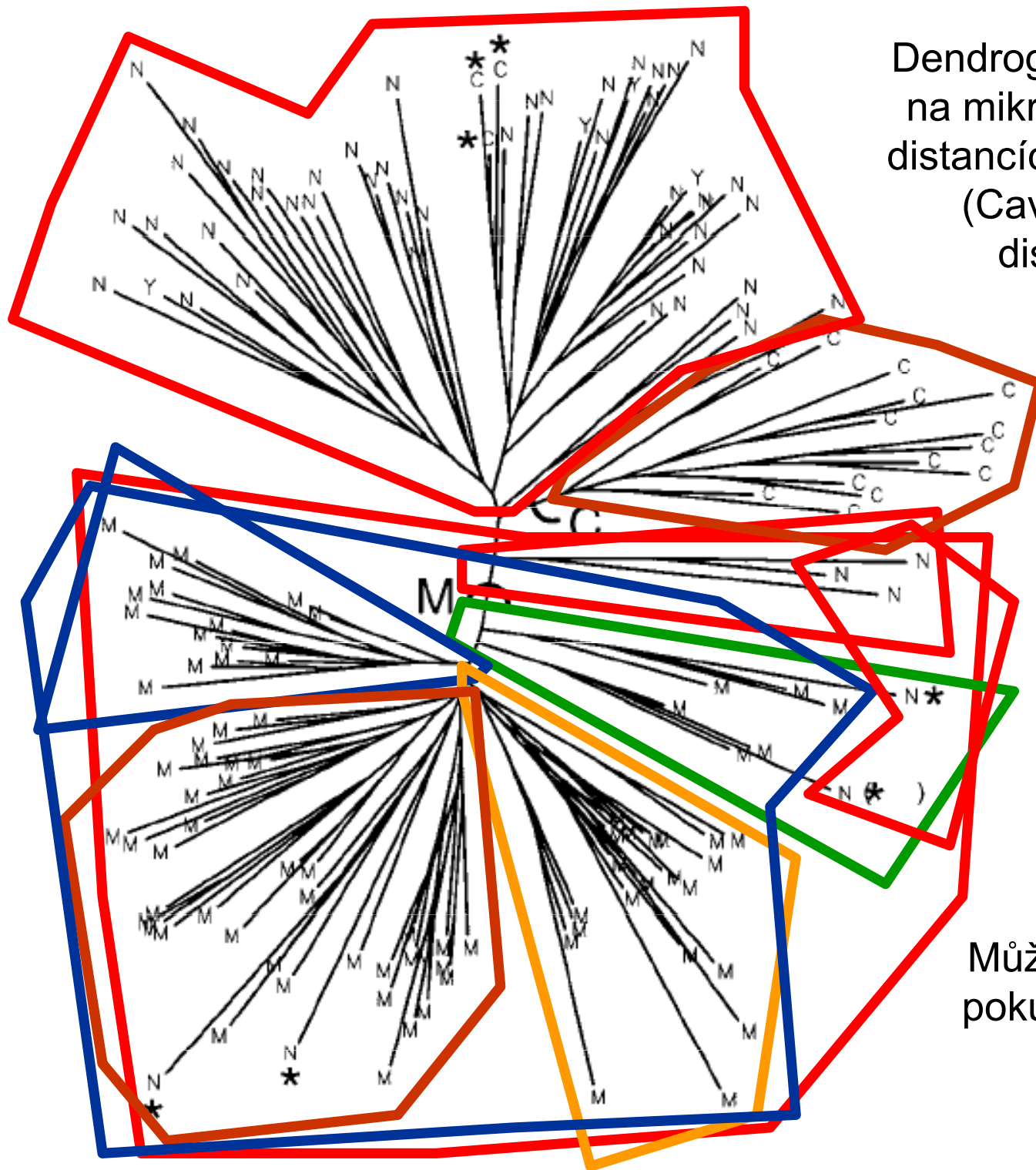


- někteří jedinci jsou geneticky jasně přiřaditelní k ESU („evolutionary significant units“ = poddruhy) – Liberec, Dvůr Králové
- ale i řada hybridů (hlavně Ostrava, Brno aj.)

# Individual-based assignments

- Kryptická populační struktura
- Neznámý počet skupin (klastřů)
- **Úroveň jedince**
- Vytvořit klastry a současně k nim přiřadit jedince
- K dispozici máme individuální genotypy (případně i souřadnice)
- Data: msat (jiné kodominantní lokusy, *SINE*), AFLP





Dendrogram založený  
na mikrosatelitových  
distancích mezi jedinci  
(Cavali-Sforza  
distances)

Může být „biased“  
pokud máme málo  
znaků

# Bayesian clustering approach

## STRUCTURE - *Pritchard et al. 2000*

- Neznámý počet populací charakterizovaných různými frekvencemi alel → počet populací a frekvence zjišťují
- Současně přiřazují individua do populací
- Lokusy, které nejsou ve vazbě, HW uvnitř subpopulací (např. mikrosatelity, SNPs)
- Model se snaží vyložit HW nebo vazebnou nerovnováhu zavedením populační struktury



**AA**

**bb**

**CC**

**dd**

Skupina 1



**AA**

**Bb**

**cc**

**DD**



**AA**

**bb**

**cc**

**dd**

Skupina 2



**aa**

**BB**

**CC**

**DD**



**aa**

**Bb**

**CC**

**dd**



# Reálně nemožné – příliš velký počet kombinací

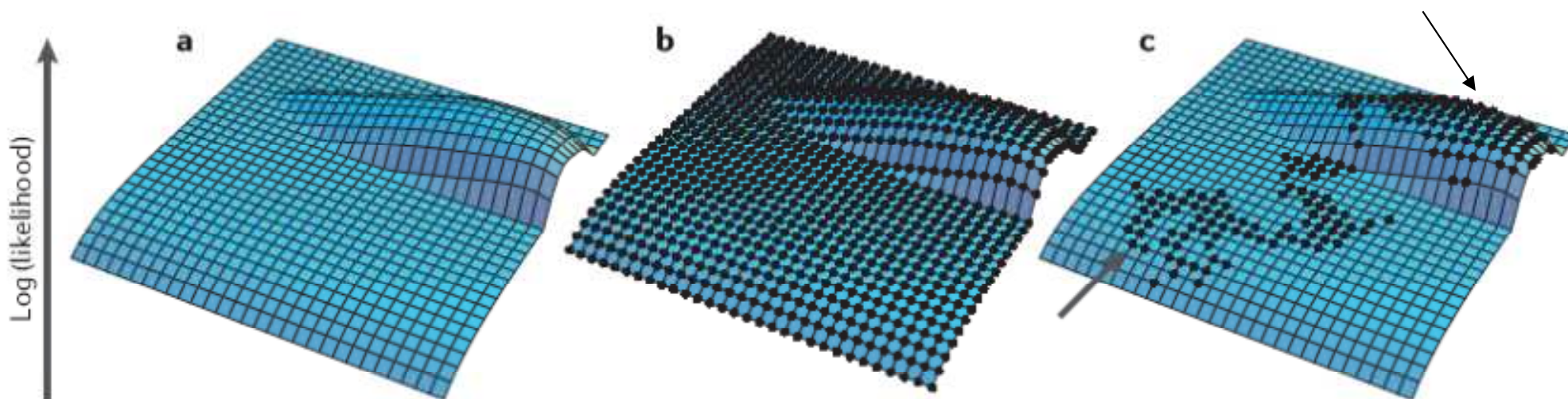
The number of partitions as a function of n.

n	#S	n	#S
1	1	11	678570
2	2	12	4213597
3	5	13	27644437
4	15	14	190899322
5	52	15	1382958545
6	203	16	10480142147
7	877	17	82864869804
8	4140	18	682076806159
9	21147	19	5832742205057
10	11597	20	51724158235372

# Bayesian clustering approach

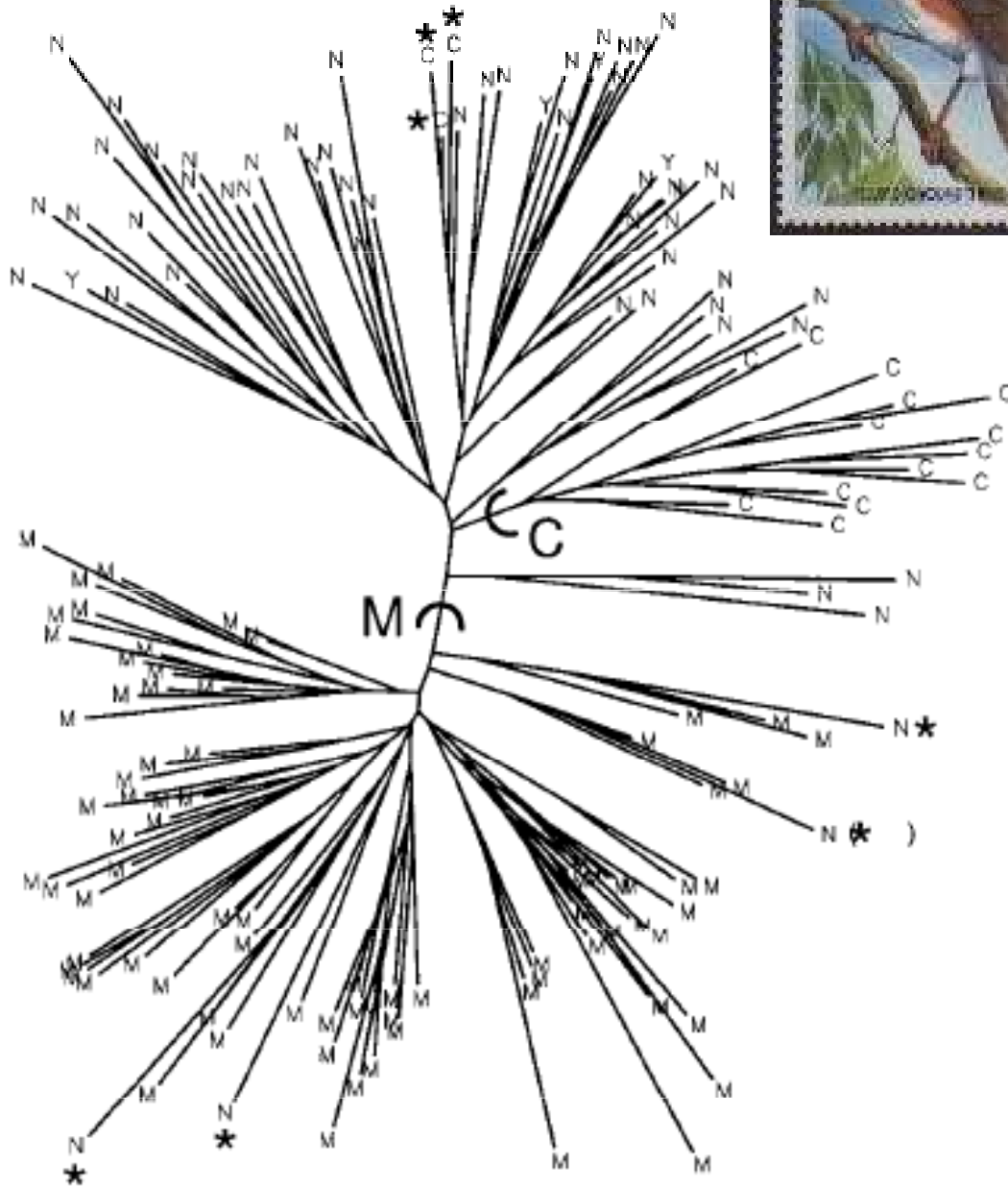
## STRUCTURE - *Pritchard et al. 2000*

- Neznámý počet populací charakterizovaných různými frekvencemi alel → počet populací a frekvence zjišťují
- Současně přiřazují individua do populací
- Lokusy, které nejsou ve vazbě, HW uvnitř subpopulací (např. mikrosatelity, SNPs)
- Model se snaží vyložit HW nebo vazebnou nerovnováhu zavedením populační struktury
- Místo přímého výpočtu – odhad pomocí Markov chain Monte Carlo



nutno opakovat několik běhů pro každý model

# *Turdus helleri*



- Fragmenty deštného pralesa
- Lokality Chawia, Ngangao, Mbololo, Yale (Kenya)
- 7 mikrosatelitových lokusů
- Neighbour-joining
- \* špatně zařazení jedinci

Shlukovací analýza na základě mikrosatelitových distancí

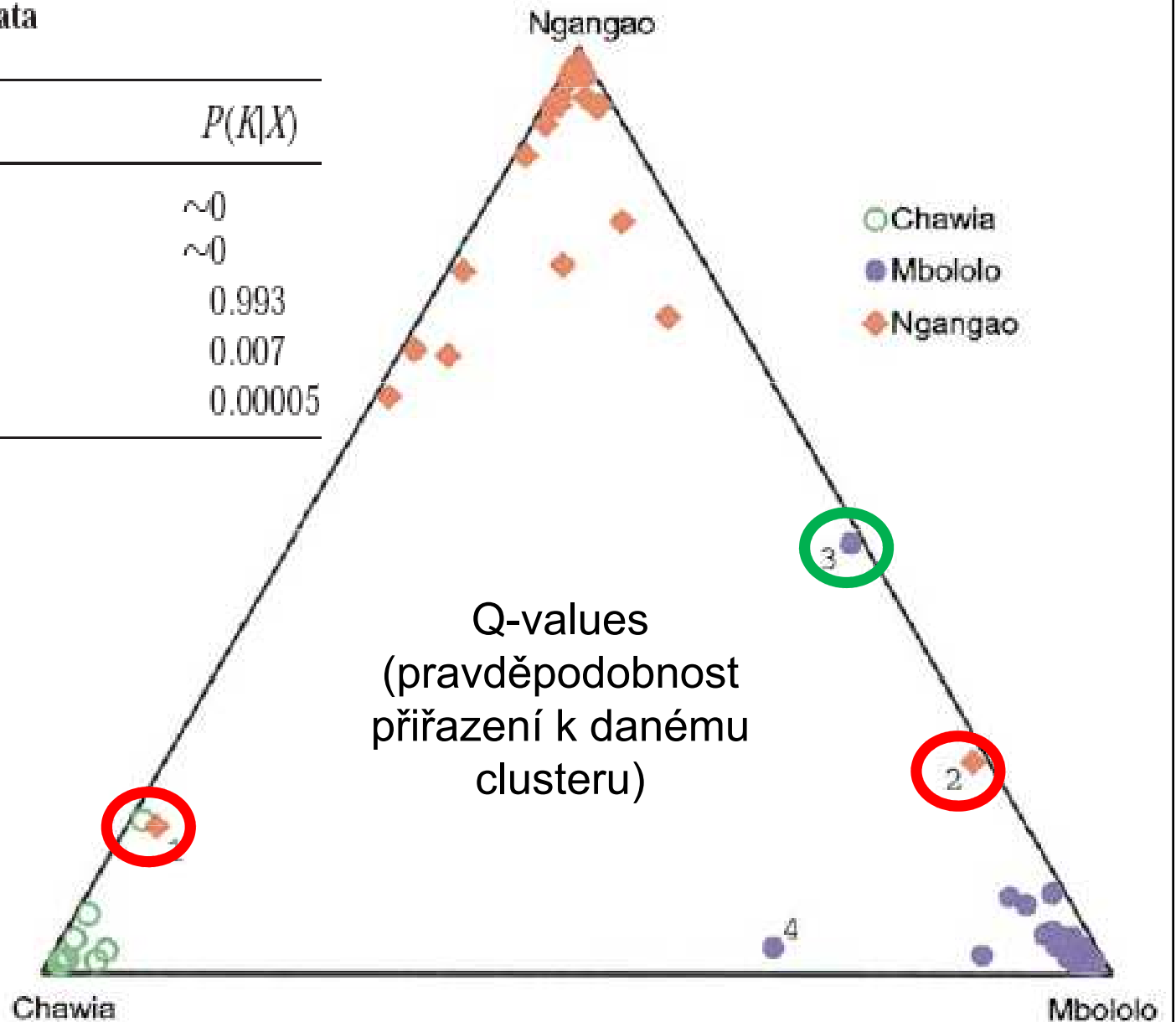
# Program STRUCTURE - Bayesiánský přístup

Inferring the value of  $K$ , the number of populations,  
for the *T. helleri* data

$K$	$\log P(X K)$	$P(K X)$
1	-3144	$\sim 0$
2	-2769	$\sim 0$
3	-2678	0.993
4	-2683	0.007
5	-2688	0.00005

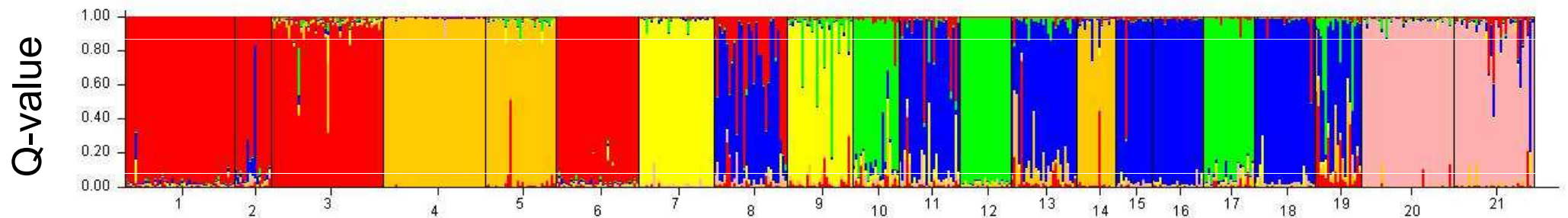
recentní migranti

hybrid?



„Admixture“ model – umožňuje zařazení jedince do více klastrů

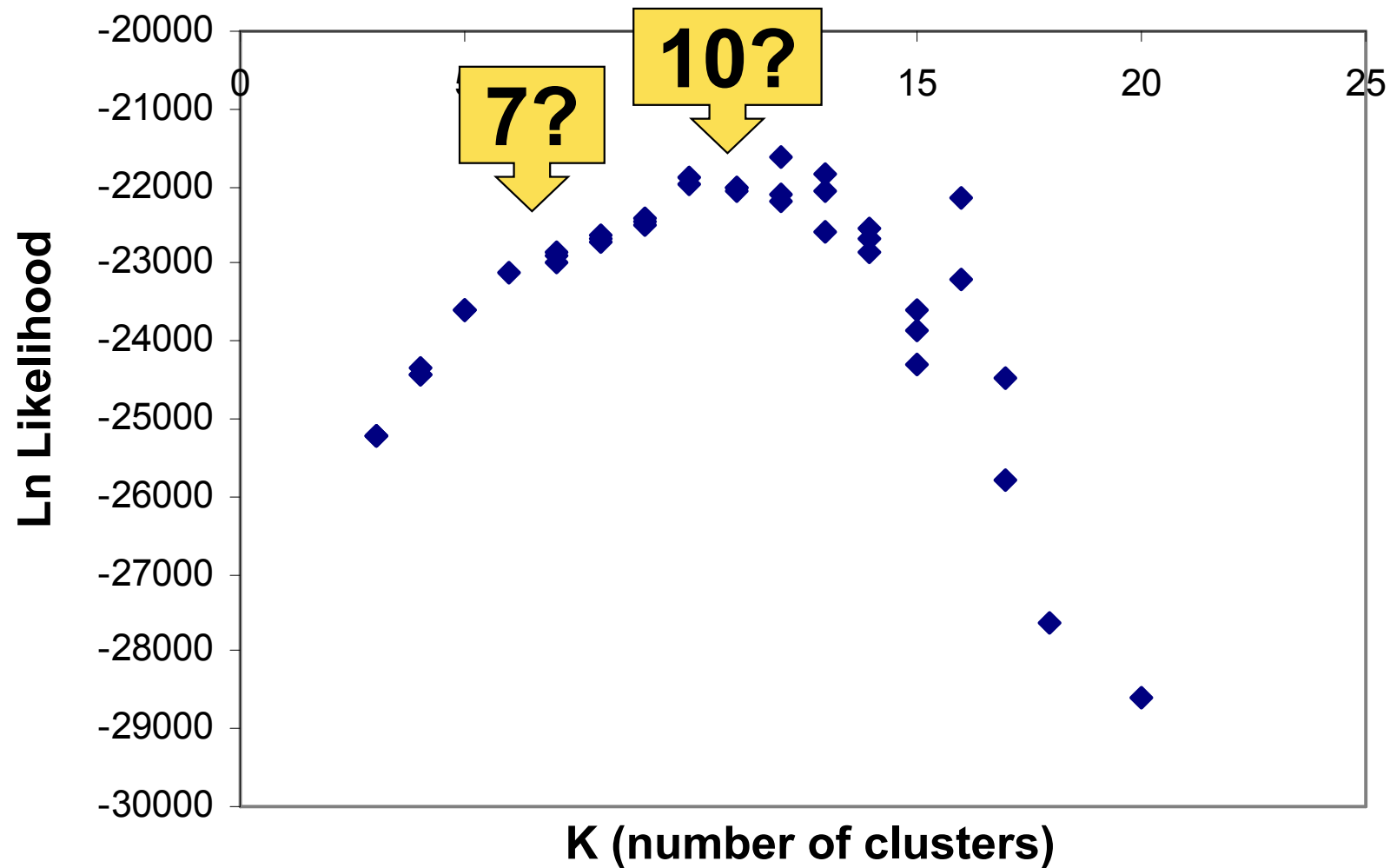
„Barplot“ pro  $K = 7$



„Proporce genomu“ každého jedince náležející určitému „clusteru“

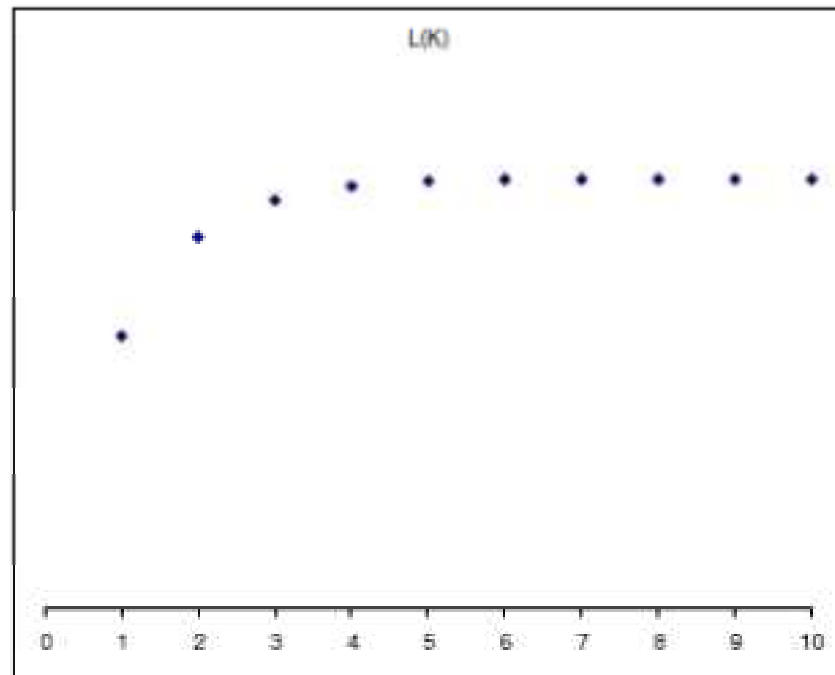
Existuje i jednodušší „non-admixture model“

# Jak poznáme nejlepší model

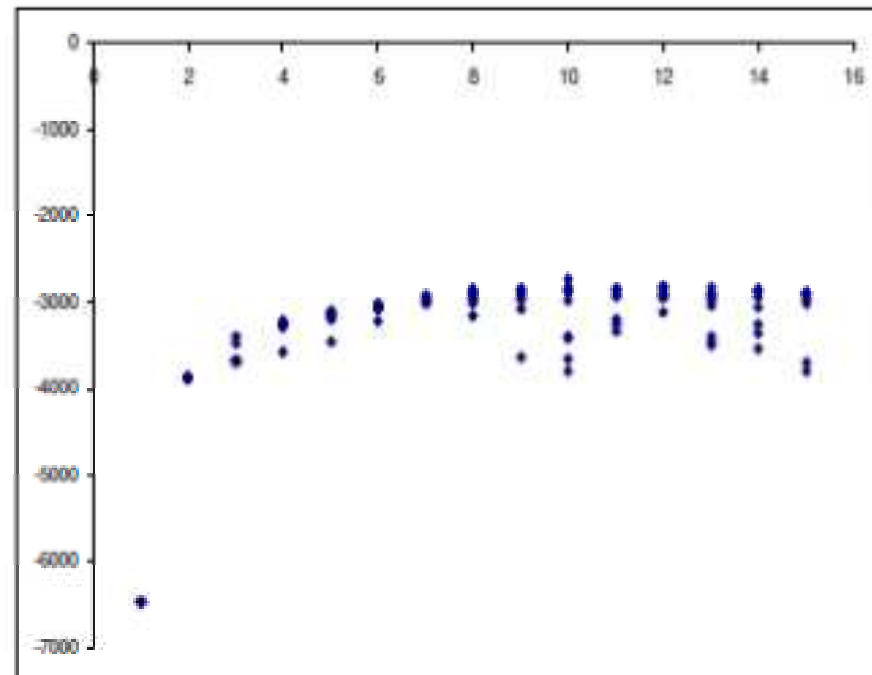


# Teorie vs. skutečnost

Teorie



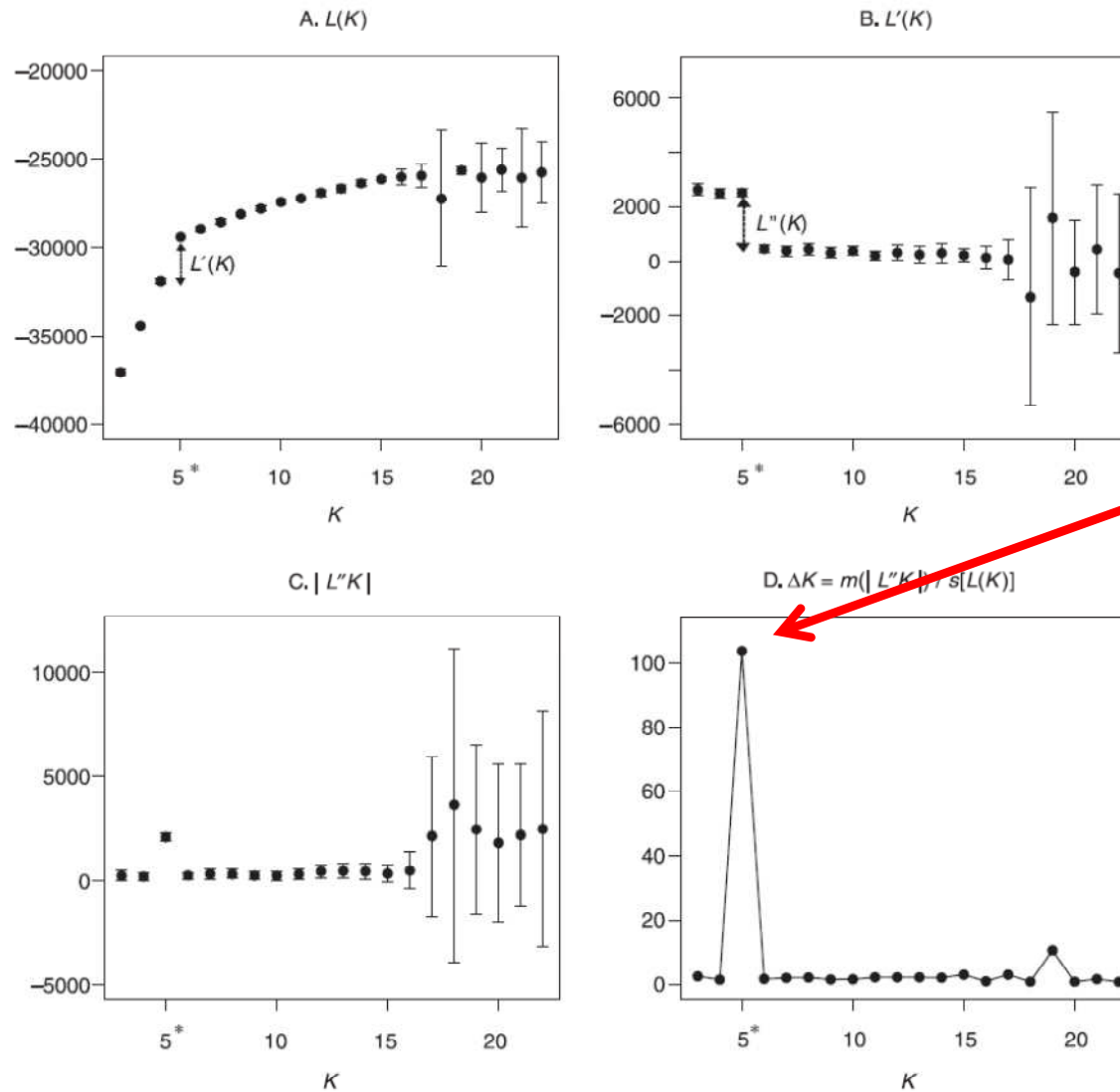
Skutečné výsledky



# Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study

G. EVANNO, S. REGNAUT and J. GOUDET

Department of Ecology and Evolution, Biology building, University of Lausanne, CH 1015 Lausanne, Switzerland

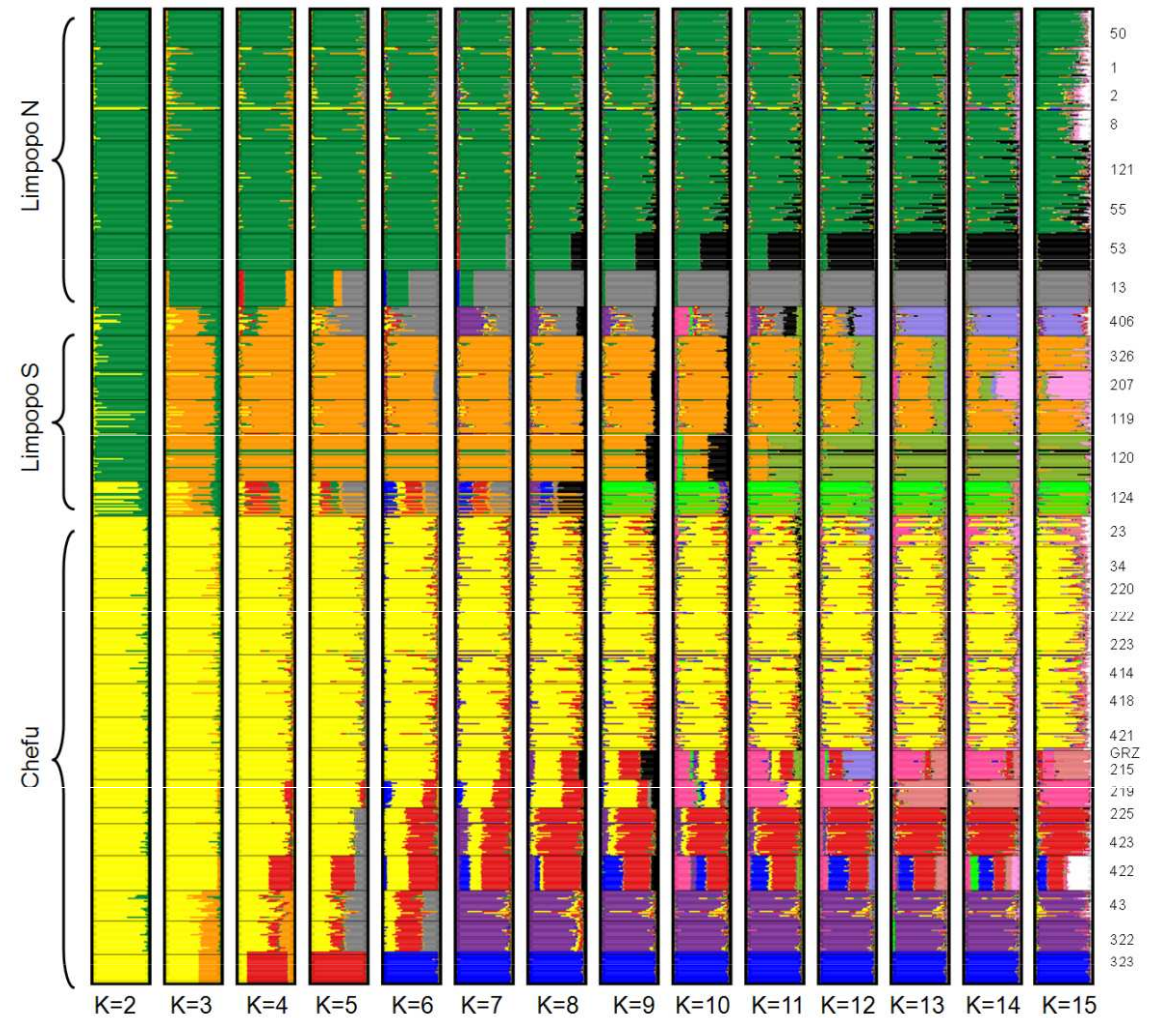




## Další úpravy výsledků ze STRUCTURE

- Structure Harvester – zpracování jednotlivých běhů (včetně odhadu  $K$  podle Evanno et al. 2005)
- CLUMPP – podobnost výsledku běhů pro stejná  $K$ ; permutace jednotlivých analýz pro stejná  $K$
- DISTRUCT – vizualizace, grafická úprava barplotů

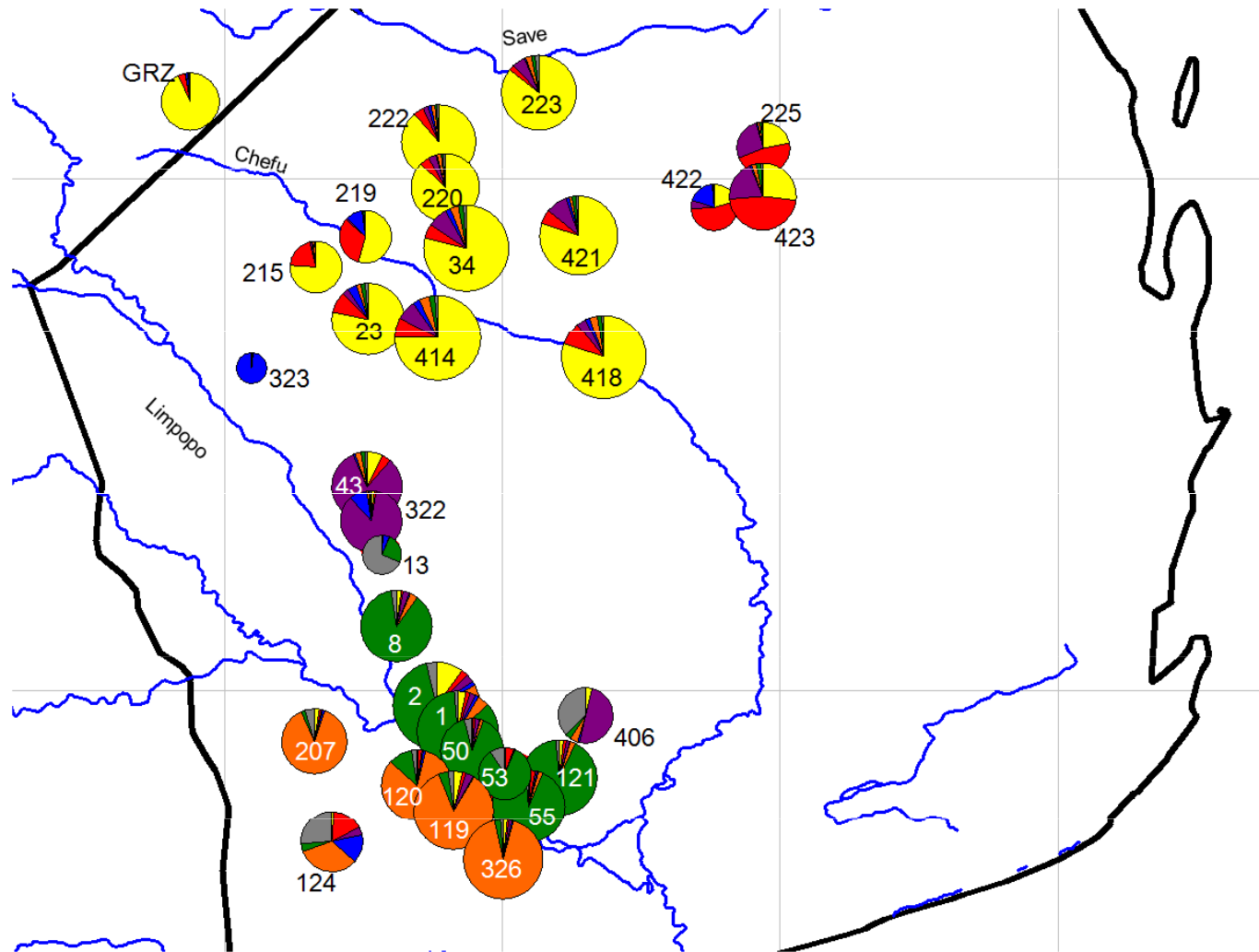
# Alternativní vizualizace výsledků ze STRUCTURE



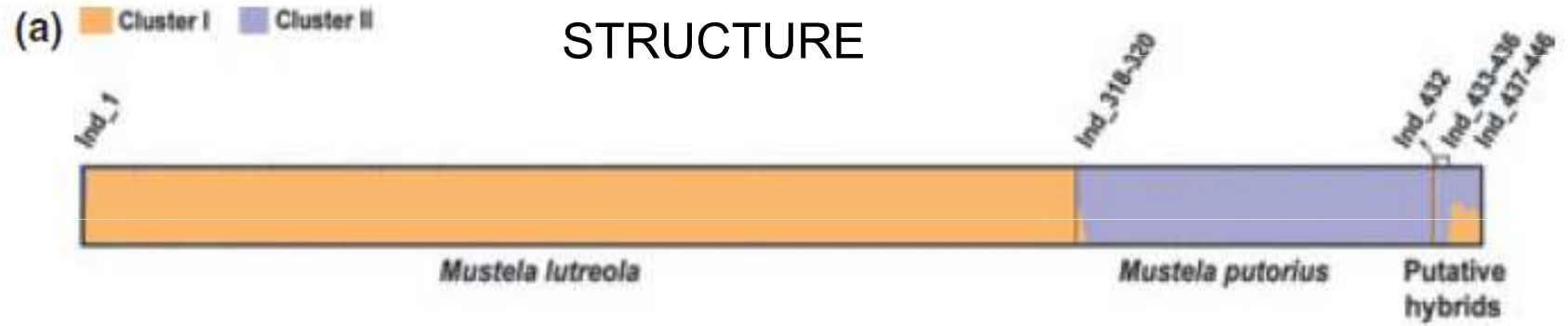
„forced clustering“

Zobrazení hierarchické struktury mezi populacemi

- Q-values pro celé populace (ne jedince)



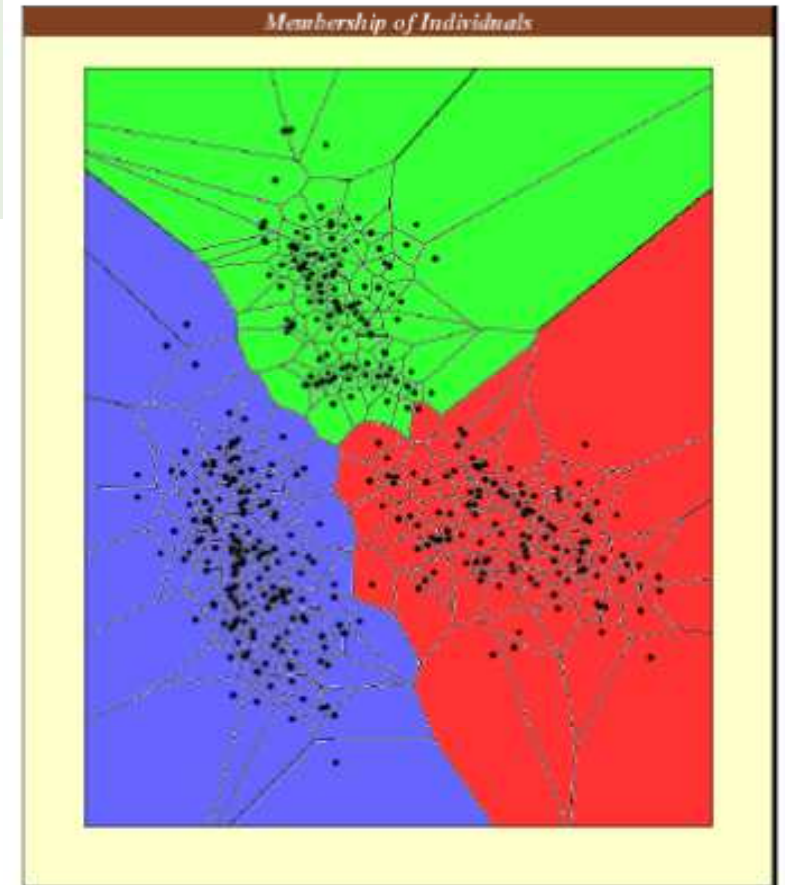
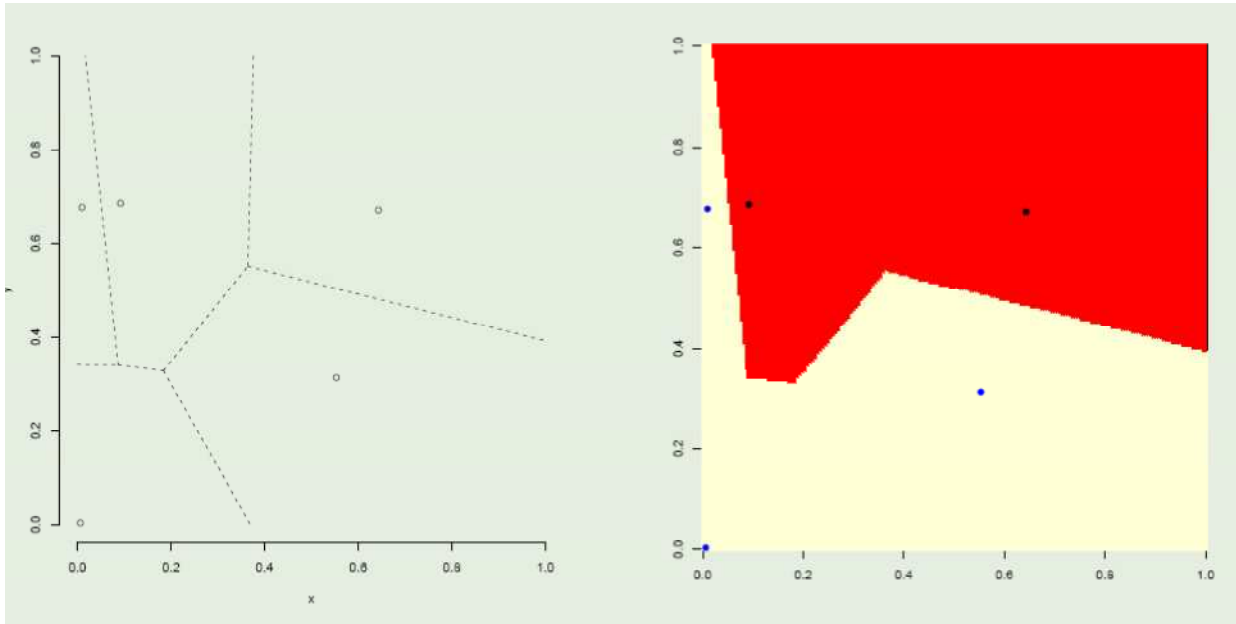
# NewHYBRIDS



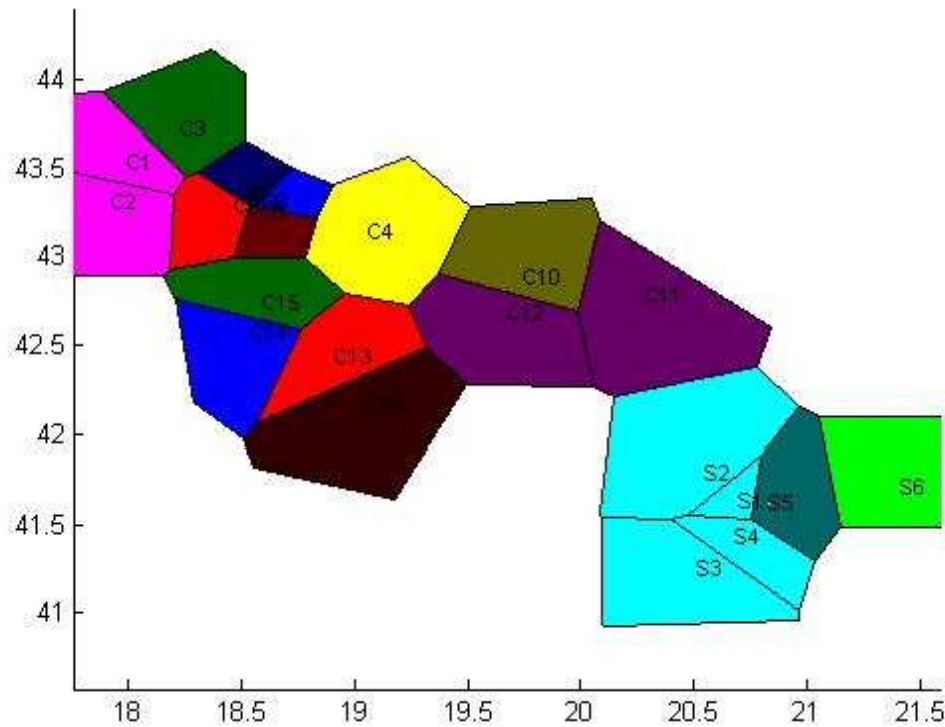
## 4. Spatially explicit analyses = spatial genetics = landscape genetics

- vychází z Bayesian clustering approach (typu STRUCTURE) – **individual-based models**
- do modelování genetické informace přidává i geografické koordináty
- např. programy BAPS, TESS, Geneland (automaticky stanovují nejlepší počet populací K)

# Spatial models – tessellation, Voronoi polygons

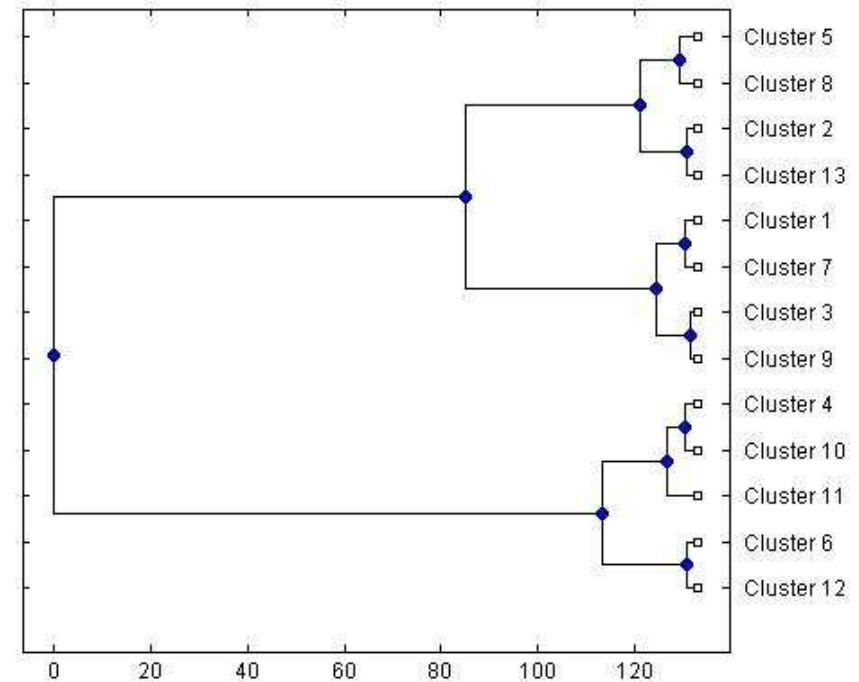


The example of very fragmented populations: the best model in BAPs for Central and Southern *Dinaromys* populations (spatial clustering of groups of individuals): K=13 (i.e. evidence of very high structuration)



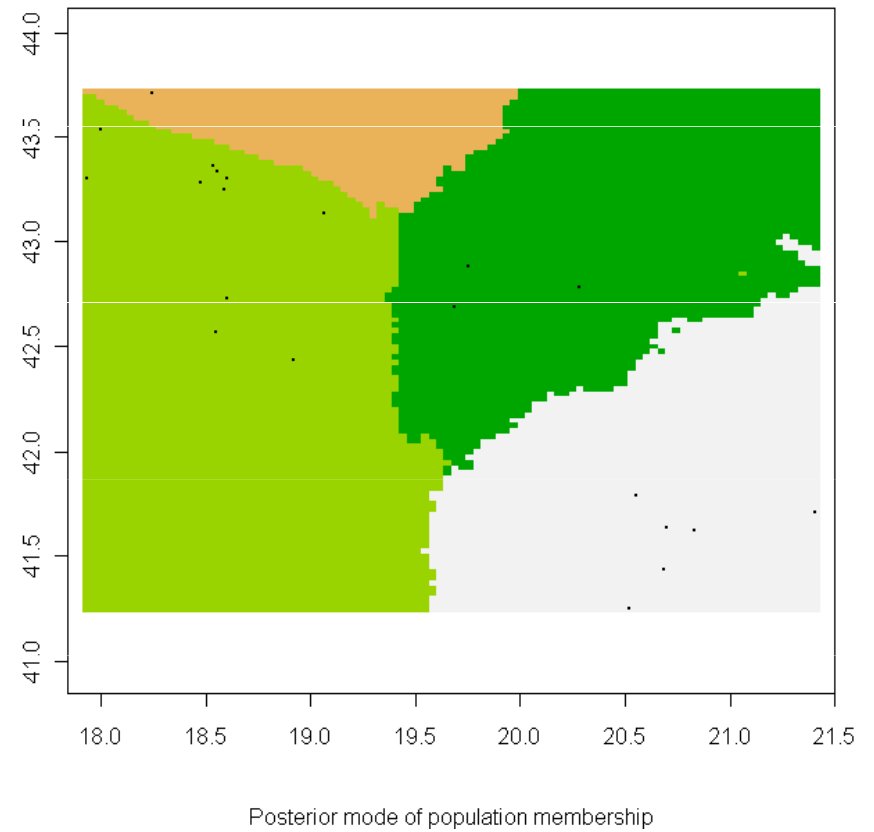
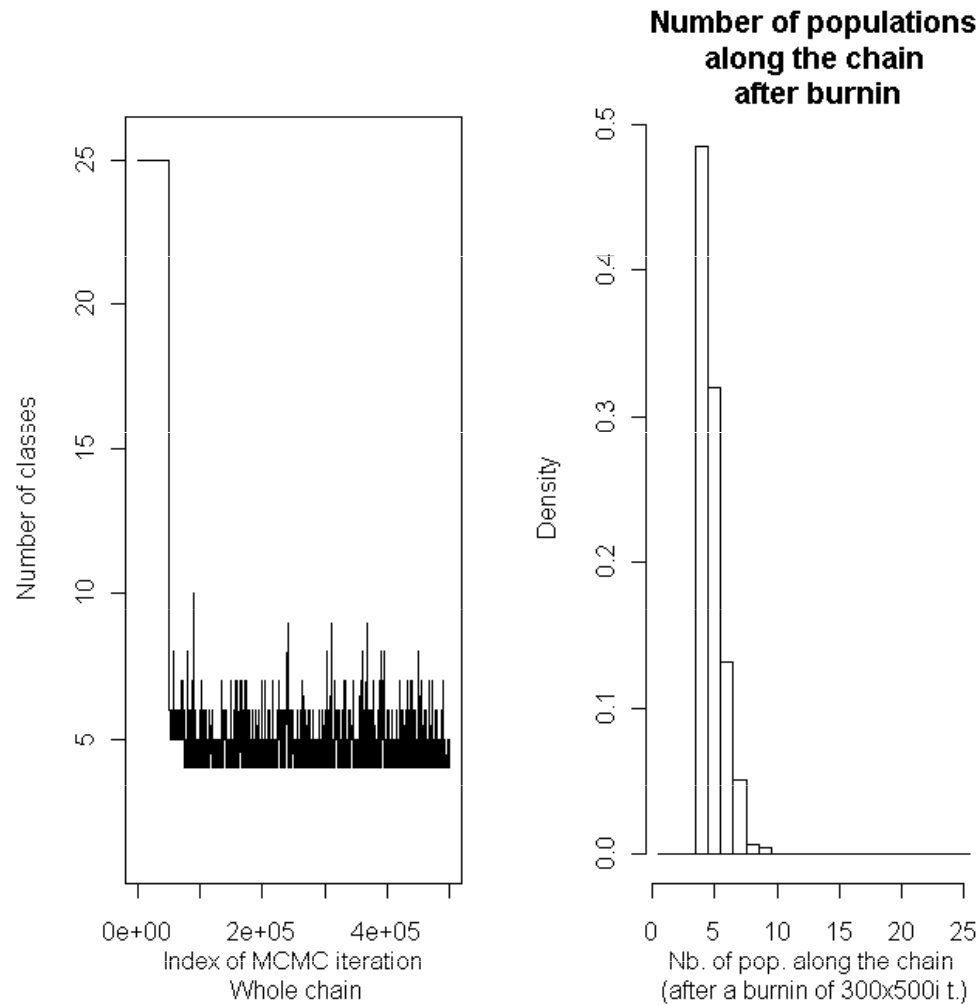
Best Partition:

- Cluster 1: {C9, C13}
- Cluster 2: {S6}
- Cluster 3: {C8, C14}
- Cluster 4: {C4}
- Cluster 5: {C1, C2}
- Cluster 6: {S1, S2, S3, S4}
- Cluster 7: {C6}
- Cluster 8: {C3, C15}
- Cluster 9: {C5, C7}
- Cluster 10: {C10}
- Cluster 11: {C11, C12}
- Cluster 12: {S5}
- Cluster 13: {C16}



Ř. program BAPS

# Př.: Geneland

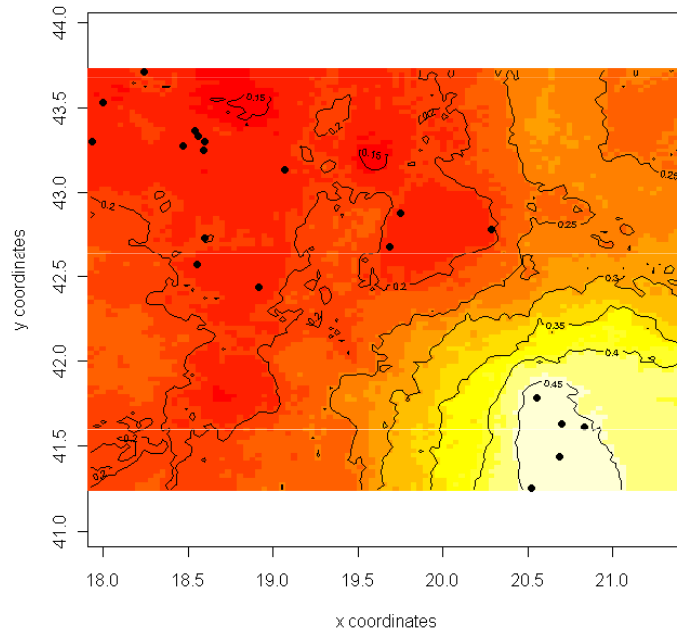




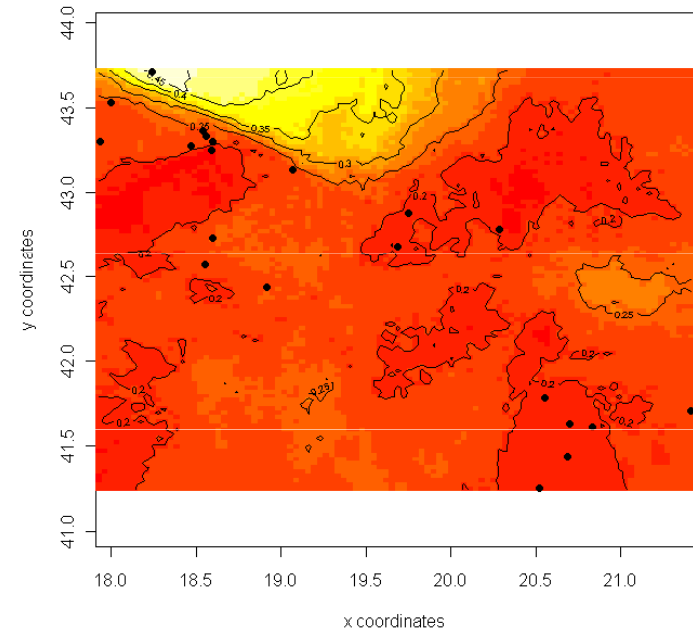
R platform

Posterior probability maps

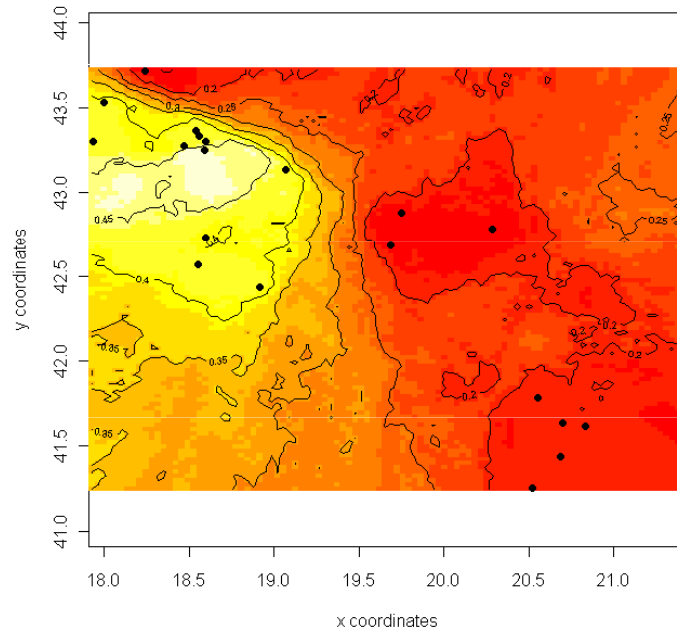
Map of posterior probability to belong to class 4



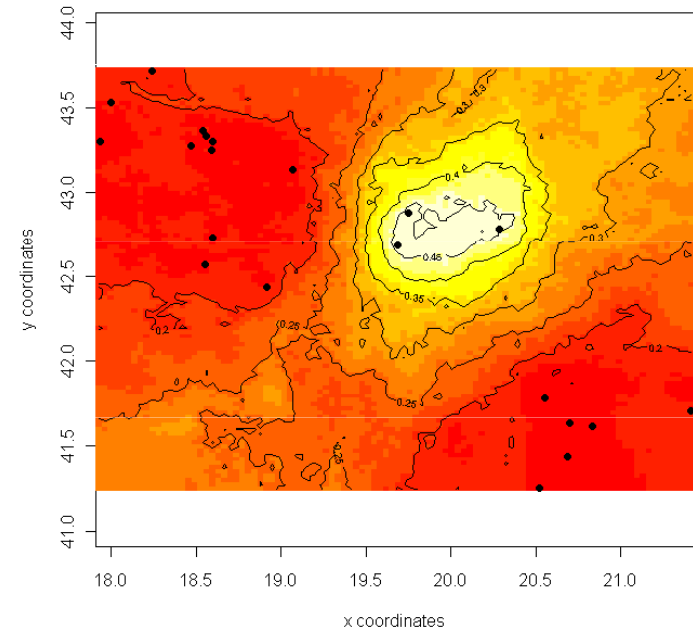
Map of posterior probability to belong to class 3



Map of posterior probability to belong to class 2

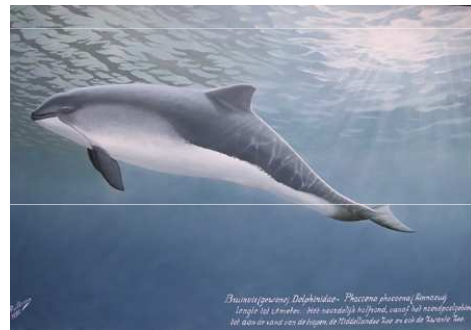
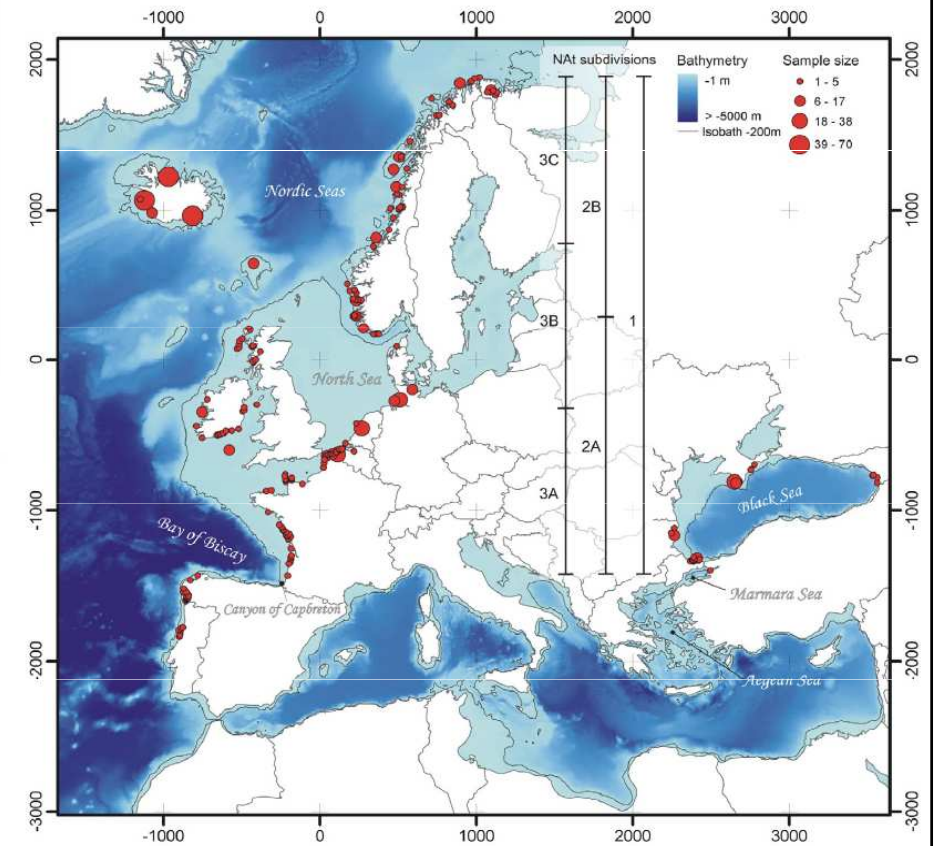
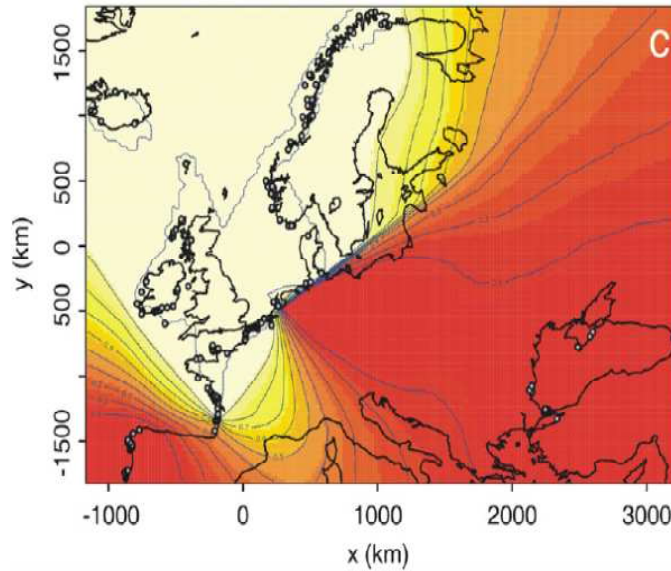
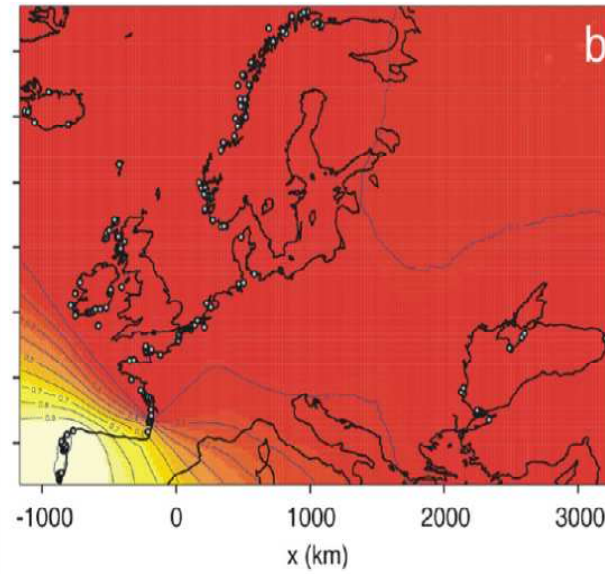
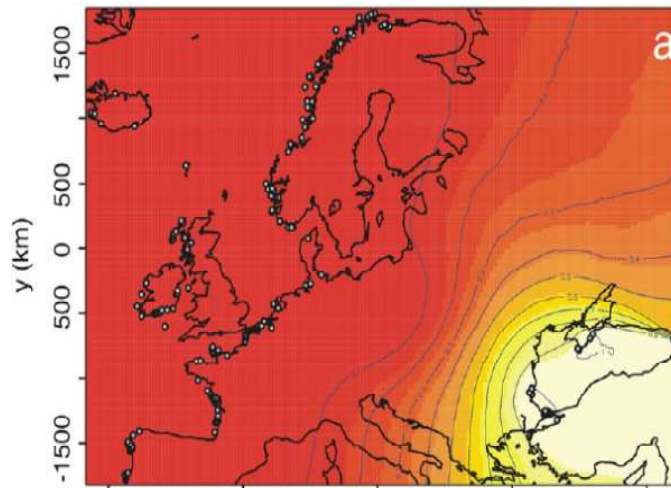
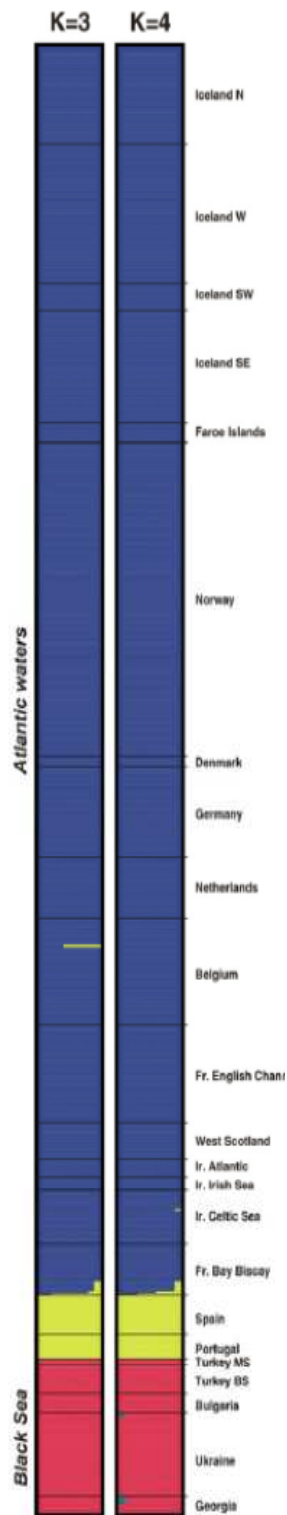


Map of posterior probability to belong to class 1



# Spatial population genetics

Fontaine et al. 2007

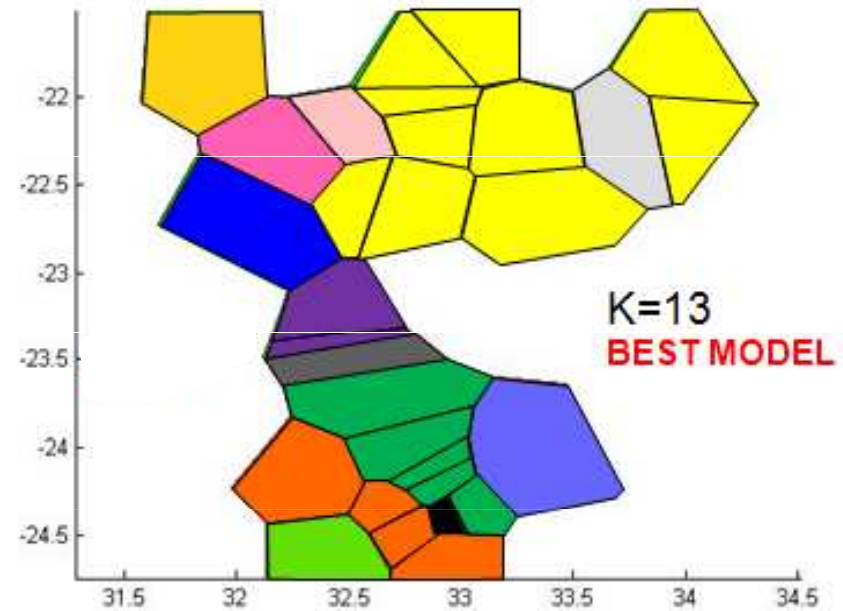
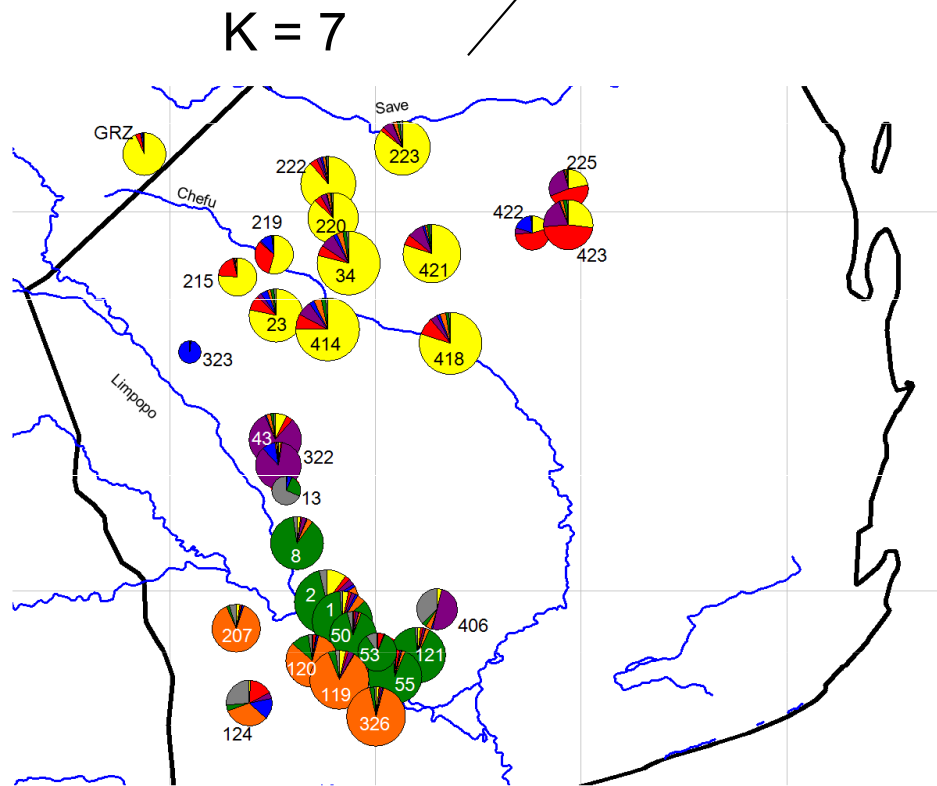


*Phocoena phocoena*

# Srovnání vlastností jednotlivých „individual-based assignment“ programů

	Structure	Partition	BAPS	Geneland
Estimate $K$	●	●	●	●
Spatial	●	●	●	●
Admixture	●	●	●	●
Inbreeding	●	●	●	●
Linked loci	●	●	●	●
Corr. freq.	●	●	●	●
Co-dom. markers	●	●	●	●
Null alleles	●	●	●	●

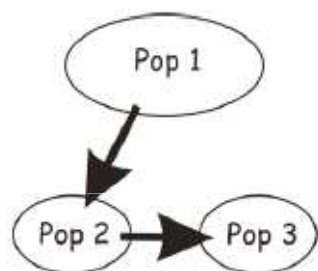
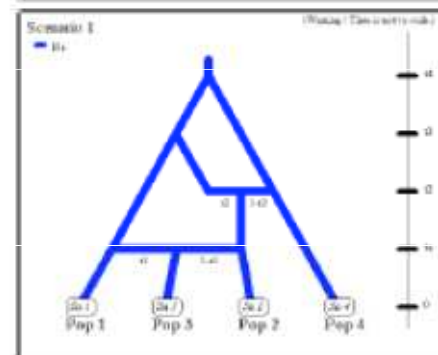
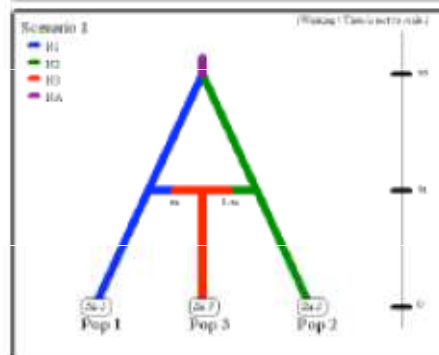
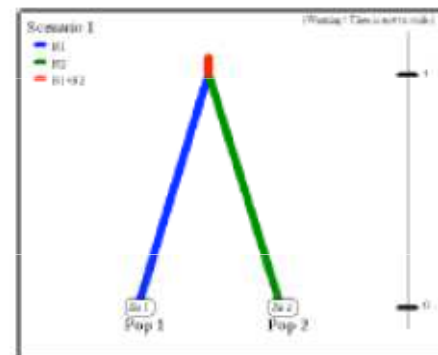
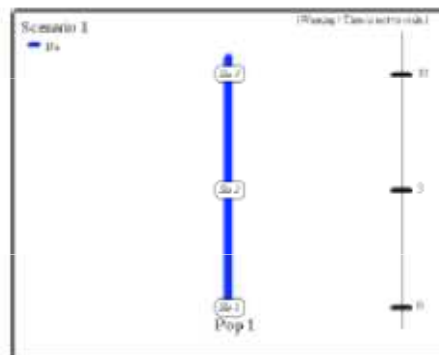
# STRUCTURE vs. BAPs



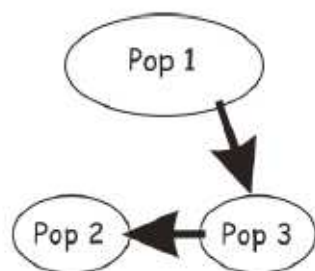
Robustní podpora genetické struktury

# Jak ta struktura vznikla?

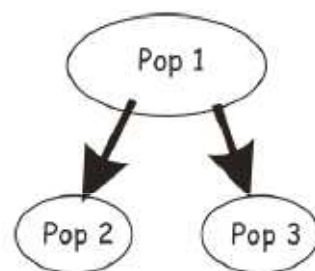
- ABC  
(Approximate Bayesian Computation)
- Program DIYABC
- Různé scénáře
- Mikrosatelitová data



Scenario 1



Scenario 2



Scenario 3