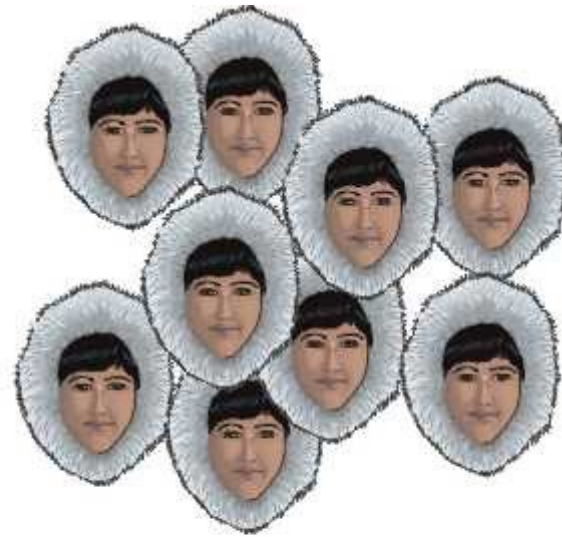


# Analýza populační variability a struktury



# Hierarchická populační struktura

Druh → populace → subpopulace (demy)

## Předpoklady studia populačně-genetické struktury

- lokusy používané pro analýzu populační struktury jsou neutrální vůči selekci
- množství adaptivní variability koreluje s neutrální variabilitou
- popsáno mnoho výjimek (rozdíly v subpopulacích - neutrální vs. adaptivní)

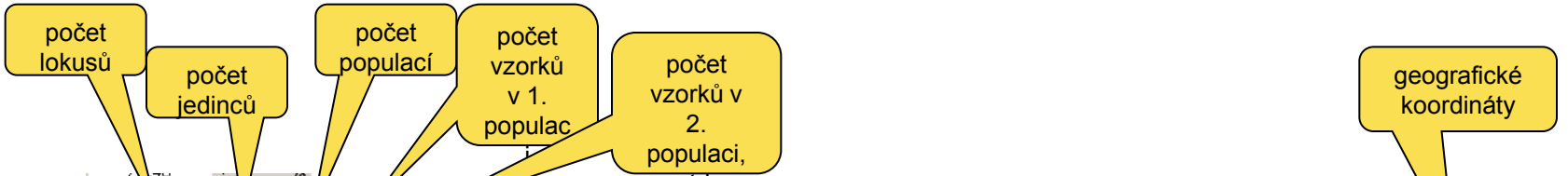
# Hierarchická populační struktura

Druh → populace → subpopulace (demy)

## Hlavní otázky populačně-genetické analýzy:

- Jak je variabilní daná populace a jaká je její efektivní velikost (neutrální genetická teorie:  $H_e = 4N_e\mu / [4N_e\mu + 1]$ )
- Nachází se daná populace v období demografické expanze nebo poklesu?
- Existuje mezi dvěma subpopulacemi bariéra toku genů a jak je silná? Jaká je prostorová genetická struktura?
- Vyskytují se v populacích imigranti nebo jejich potomci (hybridi)?

# Diploidní single-locus markery (mikrosatelity)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB		
1	11	233	10	30	22	25	22	18	23	12	27	25	29																	
2	Allele frequency																													
3	ID	Pop	Ppip01	Ppip02	Ppip04	Ppip06	EF1	EF4	EF6	Paur05	NN18	NnP217	NnP219	X	Y															
4	TYN1	TYN	174	176	128	128	213	215	120	132	166	175	236	247	174	188	249	251	287	291	205	209	140	170			48.76500	17.00750		
5	TYN2	TYN	167	167	132	132	207	207	130	136	169	175	230	245	182	188	247	249	287	287	205	218	166	166			48.76500	17.00750		
6	TYN3	TYN	176	178	128	128	207	207	130	130	169	169	0	184	190	190	249	249	283	293	209	213	166	170			48.76500	17.00750		
7	TYN4	TYN	180	180	130	132	207	207	118	124	171	173	220	243	184	190	247	249	283	285	209	214	140	166			48.76500	17.00750		
8	TYN5	TYN	174	174	128	128	207	207	128	138	179	181	241	241	182	188	247	253	0	0	205	209	158	162			48.76500	17.00750		
9	TYN6	TYN	180	182	128	130	209	209	136	136	173	175	241	243	184	186	253	253	295	325	209	213	162	170			48.76500	17.00750		
10	TYN7	TYN	174	180	128	130	215	215	130	130	173	173	241	241	190	190	249	249	285	325	209	213	158	162			48.76500	17.00750		
11	TYN8	TYN	174	175	128	134	207	221	128	138	171	177	220	240	178	186	249	249	287	293	213	213	162	166			48.76500	17.00750		
12	TYN9	TYN	156	178	126	130	209	209	128	149	177	177	220	230	184	188	249	255	287	287	217	218	158	166			48.76500	17.00750		
13	TYN10	TYN	167	178	128	130	209	209	132	134	171	175	243	245	184	186	245	247	285	287	205	218	158	174			48.76500	17.00750		
14	TYN11	TYN	170	174	130	130	217	217	130	138	173	175	0	0	182	190	251	257	289	291	213	213	166	170			48.76500	17.00750		
15	TYN12	TYN	174	176	130	132	0	0	134	134	175	177	220	220	186	186	249	251	293	297	209	209	158	162			48.76500	17.00750		
16	TYN13	TYN	166																		3	162	170			48.76500	17.00750			
17	TYN14	TYN	162																		5	166	166			48.76500	17.00750			
18	TYN15	TYN	156																		3	166	174			48.76500	17.00750			
19	TYN16	TYN	168	178	128	128	215	217	132	134	173	175	243	243	178	178	247	249	289	289	205	209	162	170			48.76500	17.00750		
20	TYN17	TYN	178	178	126	130	207	211	132	132	171	175	220	238	176	180	249	255	289	289	178	209	162	166			48.76500	17.00750		
21	TYN18	TYN	161	174	130	130	213	213	132	132	173	177	245	245	178	178	243	249	289	289	205	209	162	182			48.76500	17.00750		
22	TYN19	TYN	174	180	128	132	213	213	134	134	169	169	220	220	174	182	249	249	285	285	209	213	131	131			48.76500	17.00750		
23	TYN20	TYN	176	176	130	130	0	0	138	151	171	171	247	247	186	188	249	255	287	287	209	209	131	166			48.76500	17.00750		
24	TYN21	TYN	178	178	128	134	213	213	132	134	171	173	245	245	180	188	247	249	285	287	209	209	144	154			48.76500	17.00750		
25	TYN22	TYN	156	175	128	132	217	217	124	132	177	177	220	220	180	182	251	253	287	297	209	217	170	170			48.76500	17.00750		
26	TYN23	TYN	168	176	128	128	213	213	120	153	173	173	220	236	178	180	245	253	285	287	209	213	154	162			48.76500	17.00750		
27	TYN24	TYN	176	176	130	130	219	219	132	149	175	175	216	245	182	184	255	255	283	297	205	209	131	174			48.76500	17.00750		
28	TYN25	TYN	177	179	128	130	209	213	126	126	156	173	241	241	188	188	249	255	289	297	198	213	158	166			48.76500	17.00750		
29	TYN26	TYN	177	179	126	130	201	201	118	124	171	179	218	245	174	184	249	255	295	295	205	205	158	166			48.76500	17.00750		
30	TYN27	TYN	176	176	126	128	207	207	120	153	169	173	220	220	182	188	249	255	291	291	205	209	136	162			48.76500	17.00750		
31	TYN28	TYN	168	172	128	132	219	219	130	130	169	175	249	249	186	188	245	253	285	289	209	213	132	170			48.76500	17.00750		
32	TYN29	TYN	170	180	128	130	219	219	130	130	171	171	243	243	184	184	249	255	285	289	209	213	166	178			48.76500	17.00750		
33	TYN30	TYN	170	177	124	130	215	215	138	138	171	177	232	236	184	186	249	253	285	291	209	209	162	174			48.76500	17.00750		
34	NOV1	NOV	172	176	128	134	209	219	120	120	169	175	253	253	182	184	251	255	289	295	178	209	162	166			48.83320	16.50610		
35	NOV2	NOV	178	178	130	130	209	209	128	132	173	173	238	243	182	184	249	251	285	287	205	209	136	170			48.83320	16.50610		
36	NOV3	NOV	163	165	126	126	207	213	122	124	173	175	220	245	178	186	247	247	285	285	205	213	133	179			48.83320	16.50610		
37	NOV4	NOV	167	178	128	134	205	205	118	124	173	177	220	220	186	188	253	255	287	289	205	213	162	166			48.83320	16.50610		
38	NOV5	NOV	176	178	128	130	201	201	130	130	173	177	243	245	174	184	247	249	287	293	201	209	162	170			48.83320	16.50610		
39	NOV6	NOV	176	178	126	130	209	215	138	134	166	173	0	0	184	184	245	245	287	287	205	213	162	166			48.83320	16.50610		

pop1  
pop2

genotypy, tj. velikosti fragmentů v populaci

# Velké množství populačně-genetických programů

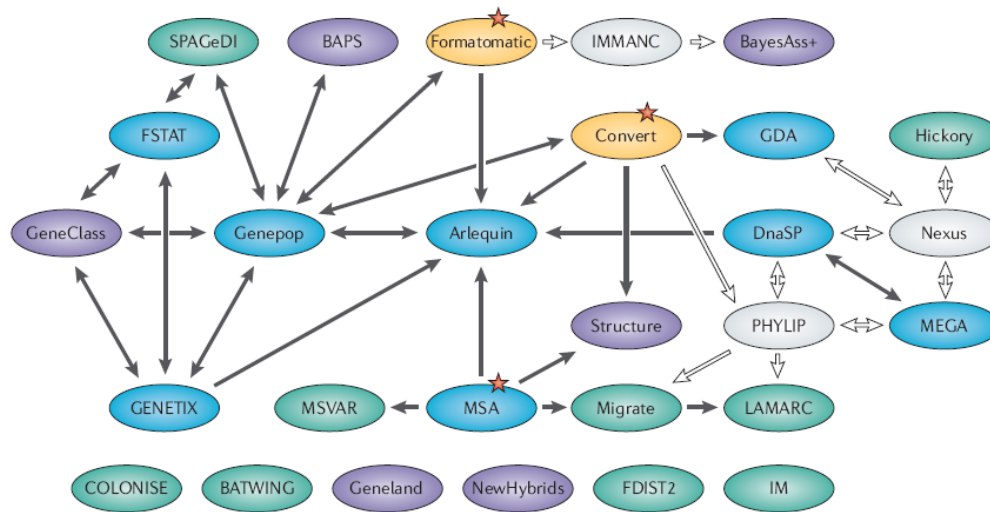
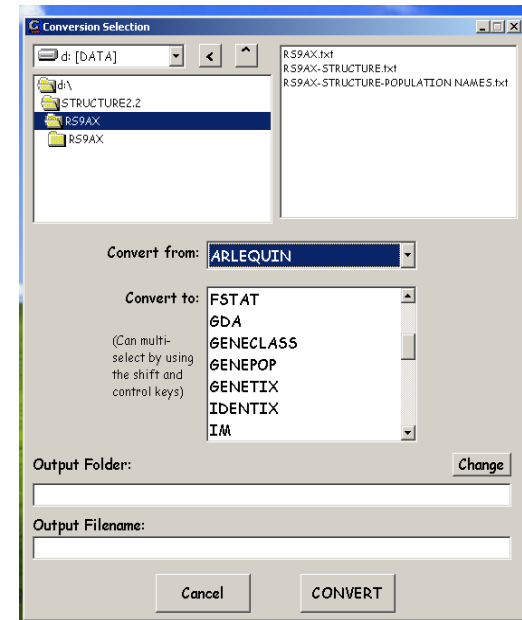
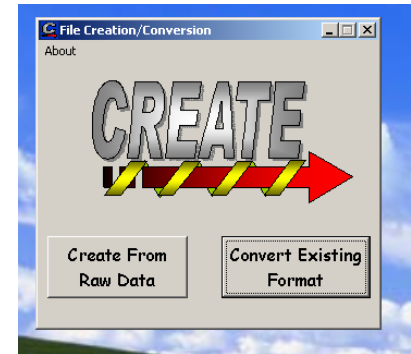


Figure 1 | **Flow chart of possible data exchange between different population genetics programs.** Although many programs have their own input-file specification, data files can still be exchanged between most programs (black arrows), avoiding tedious reformatting processes. The red stars are recommended starting points to format an initial data set. Blue ellipses represent multi-purpose packages, whereas individual-centred programs are shown in violet. The two conversion programs are shown in yellow. Specialized programs are shown in green, and light grey ellipses represent programs that are not reviewed here, but the data formats of which are used by other programs allowing indirect data exchange (white arrows). The data files associated with the programs listed on the bottom row cannot be exchanged directly with the other programs.



## Computer programs for population genetics data analysis: a survival guide

Laurent Excoffier and Gerald Heckel

# Účel populačně-genetické analýzy

Table 5 | List of computer programs suited for a given analysis and genetic marker

	Multi-allelic markers*	STR	Dominant markers (AFLP)	SNP	DNA sequences
Descriptive statistics			SPAGeDi		Arlequin, DnaSP, MEGA
Linkage disequilibrium					
Analysis of population subdivision			Hickory		Arlequin, DnaSP, MEGA
Detection of new immigrants: known populations					
Detection of new immigrants: inferred populations					
Demographic expansion or decline					Arlequin, DnaSP, IM, LAMARC, Migrate
Population size					IM, LAMARC, Migrate
Divergence time					DnaSP, IM, LAMARC, Migrate
Migration rates					DnaSP, IM, LAMARC, Migrate
Neutrality tests					Arlequin, DnaSP, MEGA
Spatially explicit analyses					

\*By multi-allelic markers, we mean loci for which no specific mutation model is assumed. Computations are based on allele frequencies only, as well as nucleotide frequencies. Some software packages are assumed to be independent of the marker type.

ed. In the latter case, computations are based on a sequence, STR and SNP allele frequencies, and to detect new

frekvence alel

frekvence alel + mutační model

# 1. Vnitropopulační variabilita (popis získaných dat)

## Polymorfismus

- podíl polymorfních lokusů (znaků) – 95 % nebo 99 % (např. 0,8 = 4 z pěti zkoumaných mikrosatelitů mají v populaci alespoň 2 alely, z nichž ta vzácnější dosahuje frekvence alespoň 1% nebo 5%)

## Počet alel (number of alleles)

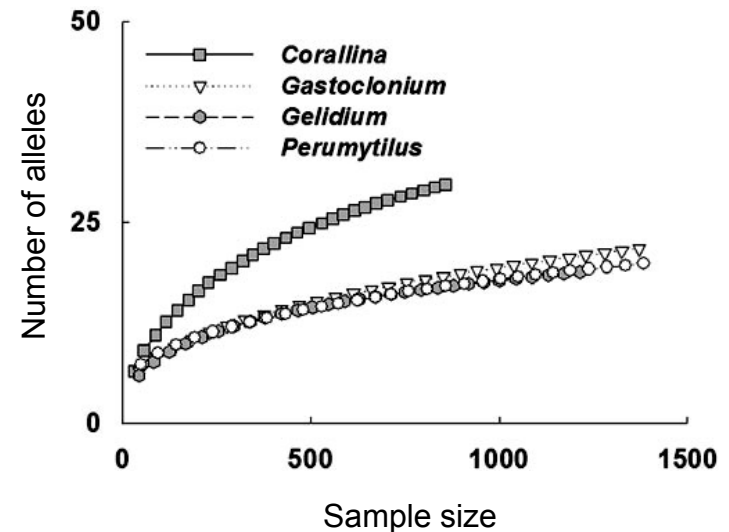
- počet alel na lokus

## Alelická bohatost (allelic richness)

- počet alel na lokus vztažený k velikosti vzorku (metodou „rarefaction“)

## Pozorovaná heterozygotnost (observed heterozygosity)

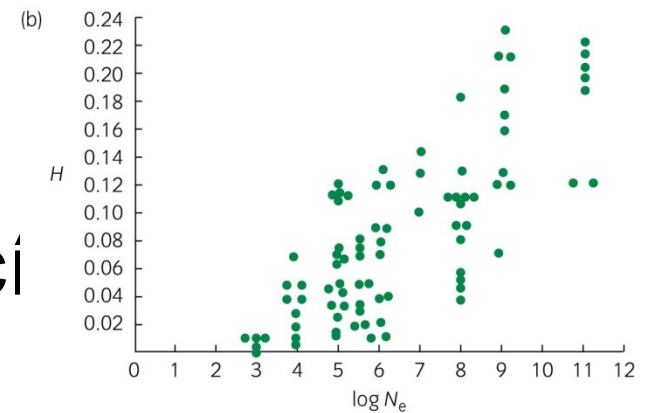
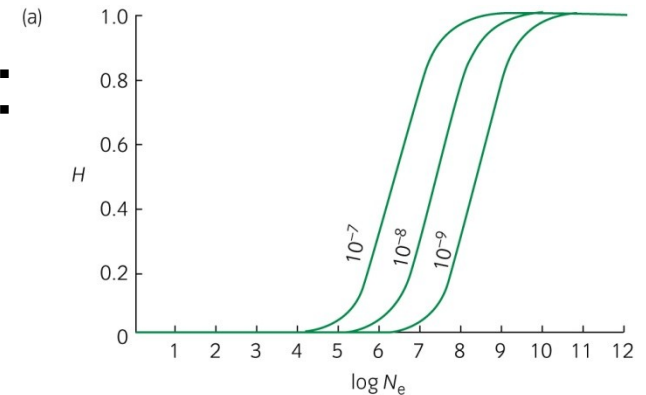
- průměrná četnost heterozygotů v jednotlivých lokusech





# Použití údajů o genetické variabilitě

- neutrální genetická teorie:  
 $H_e = 4N_e\mu / [4N_e\mu + 1]$
- mutation-drift equilibrium
- srovnání různých populací  
a jejich  $N_e$  ( $H_e$ , AR atd.)





# $N \sim H_e$ ... nemusí to být pravda

- vliv historického vývoje populací („bottlenecks“)
- Northern elephant seals *Mirounga angustirostris*
  - 120 000 jedinců – 50 allozymových lokusů – žádná variabilita



rypouš severní



# Indian rhinoceros

- 29 allozyme loci – výrazně vyšší variabilita než byla očekávána
- ještě nedávno až 100 000 nosorožců v Nepálu a Indii
- 1960 – 60-80 jedinců v Nepálu
- 1980 – nárůst populace na několik set – max. 2 alely na lokus
- méně intenzivní bottleneck ( $H_0$  odpovídá historické  $N_e$ )



# Hardy-Weinbergova rovnováha (HWE)

Př. Jeden lokus se 2 alelami

Alela	Četnost alely
A	p
a	q

$$p + q = 1$$

p, q - zjistíme analýzou svých vzorků

Genotyp	Očekávaná četnost genotypu
AA	$p^2$
Aa	$2pq$
aa	$q^2$

= Hardy-Weinbergova rovnováha

- četnosti genotypů zjistíme analýzou svých vzorků
- odchylky od očekávaných četností  $\Rightarrow$  např.  $\chi^2$  test

Očekávaná heterozygotnost (expected heterozygosity,  $H_e$ ) při HWE

$H_e = 1 - (p^2 + q^2)$  ..... pro 1 lokus se 2 alelami s četností p a q

# Předpoklady HWE

- random mating (panmixia)
- negligible effects of mutation and migration („closed populations“)
- infinitely large population size
- Mendelian inheritance
- no selection
  
- check for linkage disequilibrium (vazebná nerovnováha)



2 lokusy ve fyzické blízkosti  
(snížená pravděpodobnost rekombinace  
linkage disequilibrium)

vs.



nebo



2 lokusy fyzicky vzdálené  
(pravděpodobnost rekombinace není ovlivněna  
linkage equilibrium)

# Odchylky od HW rovnováhy

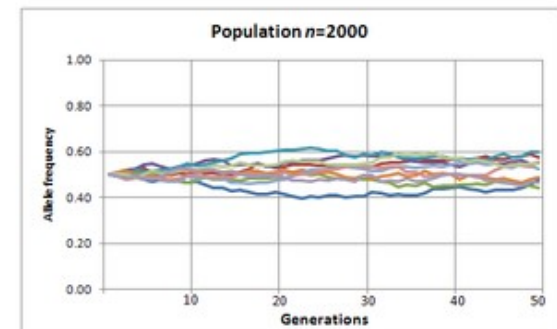
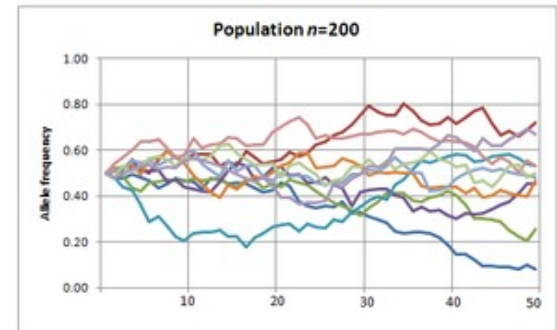
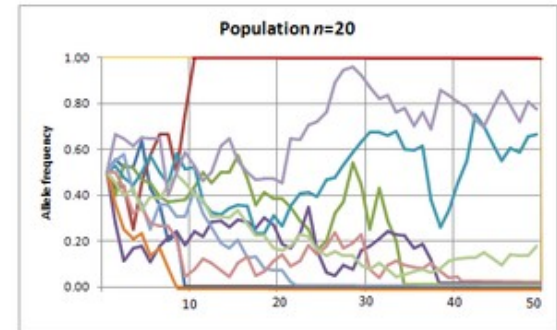
- **Test HWE** – nejlépe Genepop („exact probability tests“) – pokud jsou odchylky, tak některý předpoklad HWE nebyl splněn
- **nadbytek heterozygotů** = negativní asortativní páření (tj. cílené rozmnožování nepodobných jedinců) – použité lokusy mohou být výhodné v heterozygotním stavu (např. geny MHC)
- **nedostatek heterozygotů**
  - inbreeding (postihuje všechny lokusy stejně)
  - nulové alely (jen na některých lokusech bude deficit heterozygotů)



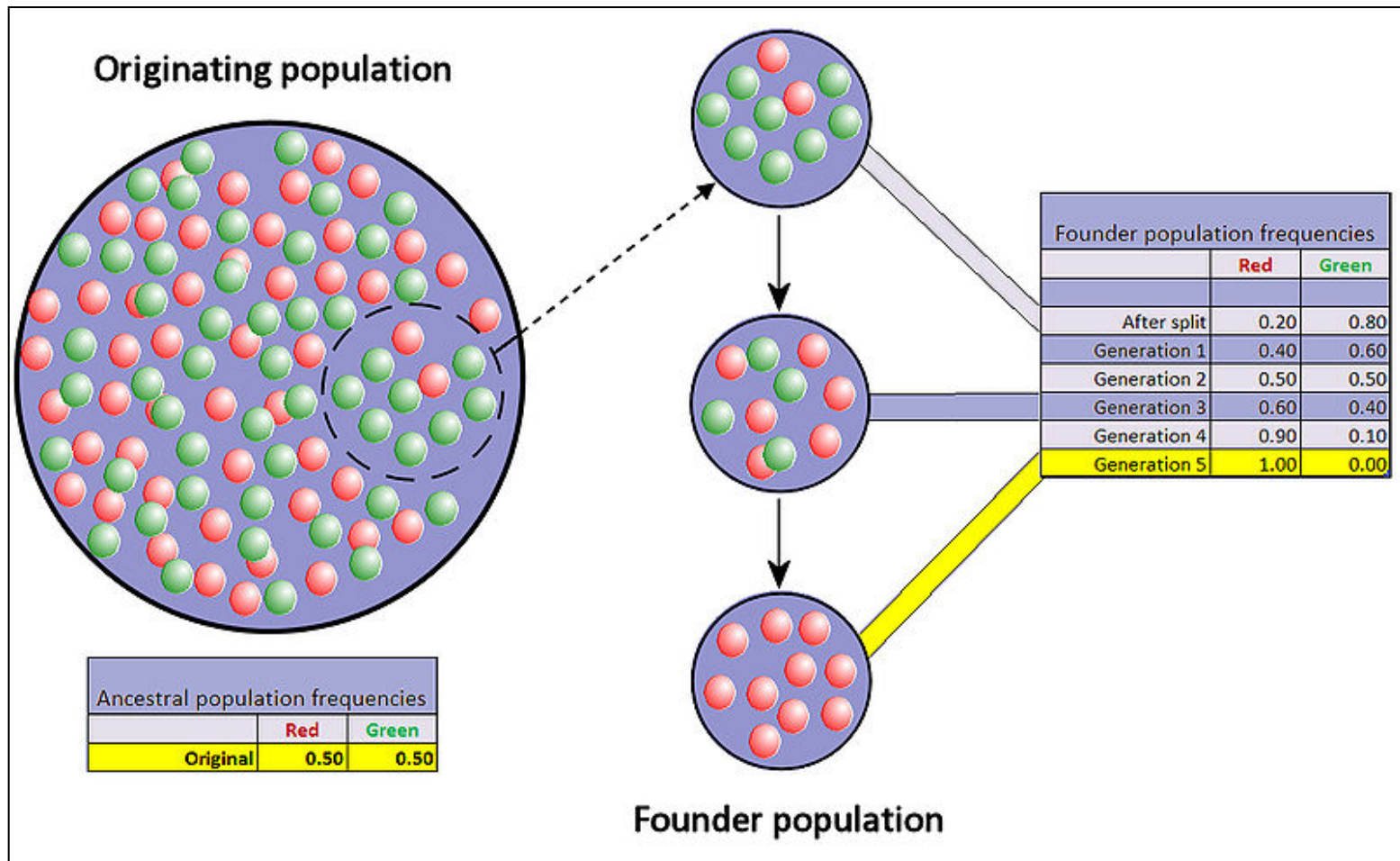
# Genetický drift



- Náhodné změny ve frekvencích alel
- Intenzita driftu závisí na velikosti populace (viz ochranářská genetika)
- Specifické případy – founder effect, bottleneck



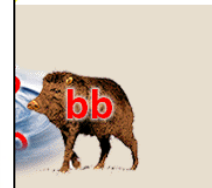
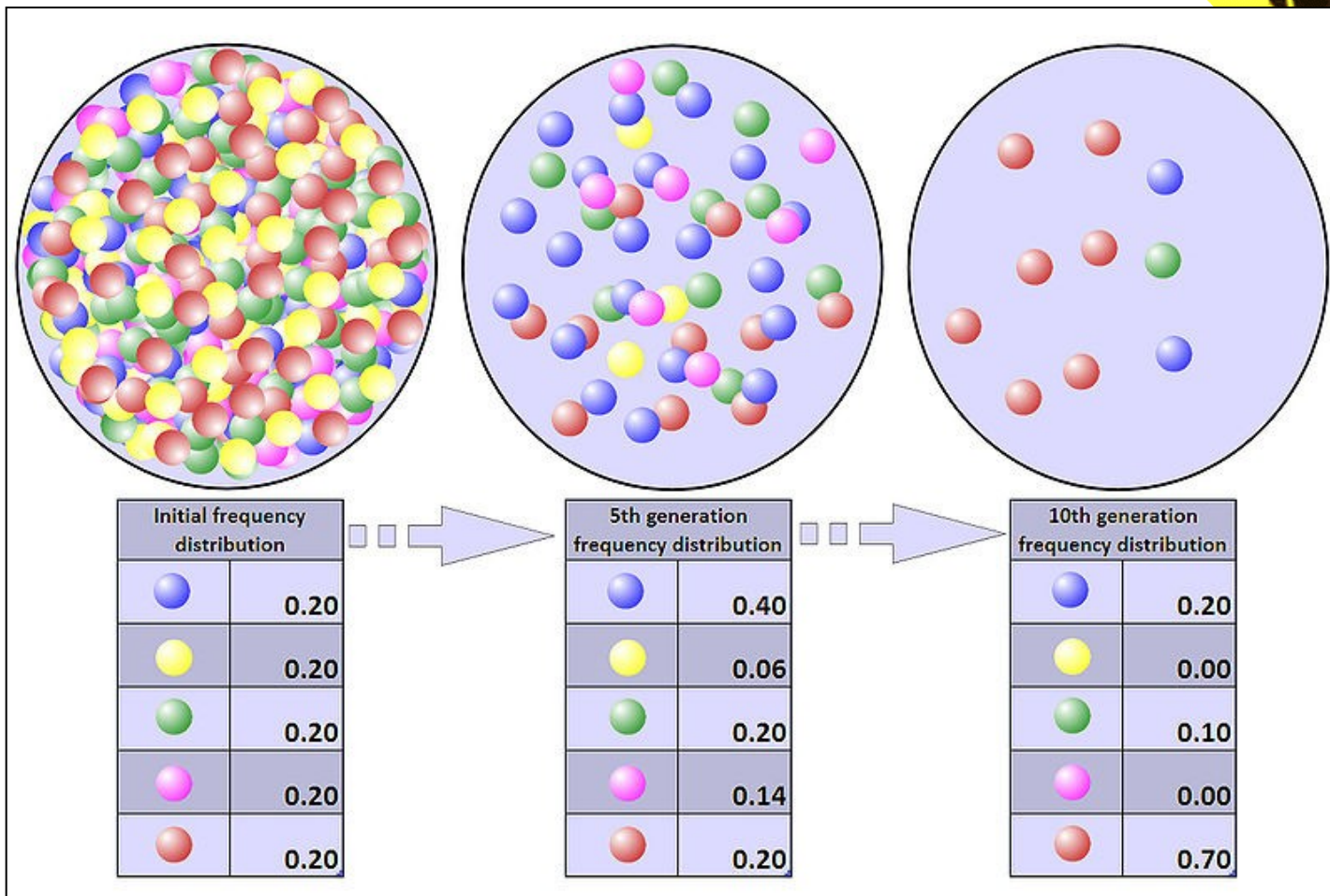
# Founder effect („jev zakladatele“)



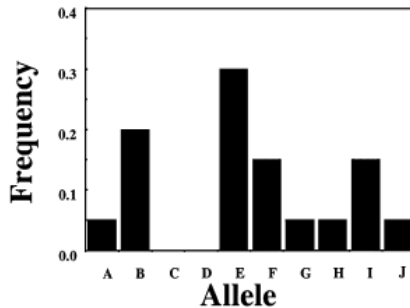




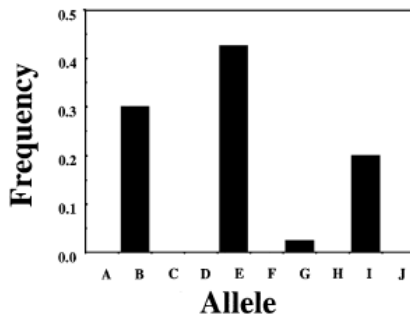
# Bottleneck



# Detekce bottlenecku



Reduction in population size



50% reduction in no. of alleles  
20% reduction in size range

Při bottlenecku vymizí nejdříve vzácné alely, rychleji než se sníží heterozygotnost nebo rozsah alel

Nutno znát (předpokládat) mutační model – pak se nasimuluje „mutation-drift“ rovnováha a srovnává se se skutečným stavem

**Program BOTTLENECK:** počet alel vs. heterozygotnost

**Program M Ratio:** počet alel vs. rozsah alel

Předpoklad testů: Populace v HW rovnováze

# Bottleneck

- Při bottlenecku
  - redukce počtu alel
  - ovlivnění heterozygotnosti není tak rychlé
  - více heterozygotů než by vyplývalo z populačního modelu (IAM, TPM, SMM)
- Nutno definovat mutační model, předpokladem je HW rovnováha, testuje se mutation-drift equilibrium
- Program BOTTLENECK

- Záleží na zvoleném modelu
- Ale SMM asi neplatí stoprocentně

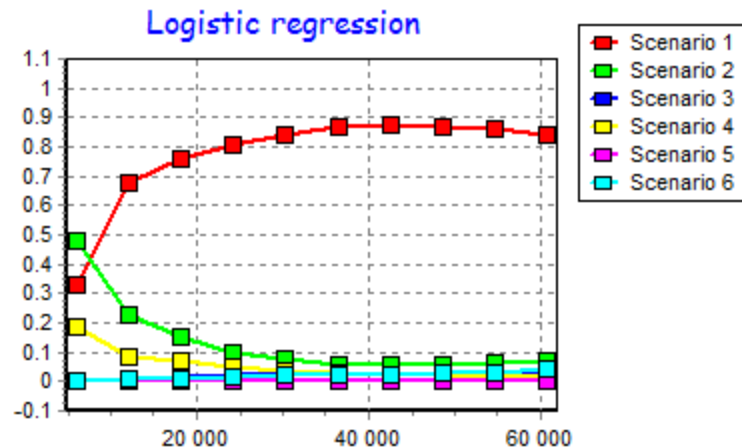
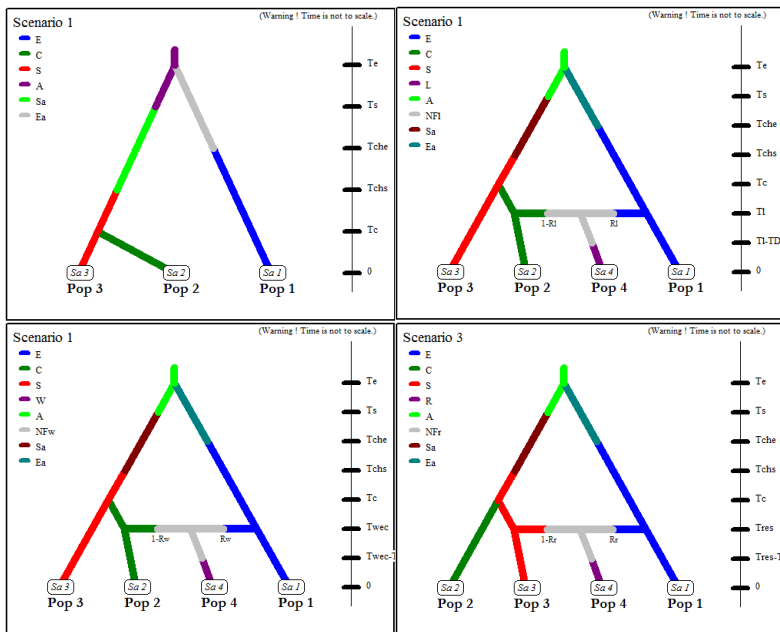


Table 4 Departures from mutation-drift equilibrium in yellowhammer samples

Sampling sites	IAM	TPM	SMM
DEV	0.0039**	0.0078**	0.3438
GWE	0.0039**	0.0391*	0.7656
CUM	0.0039**	0.0391*	0.8125
OXO	0.0039**	0.0117*	0.3438
YOR	0.0039**	0.0117*	0.7656
T & W	0.0078**	0.1875	0.8125
LEI	0.0195*	0.2891	0.6563
SUF	0.0039**	0.0195*	0.6563

# Detekce bottlenecku nebo expanze – Bayesiánský přístup

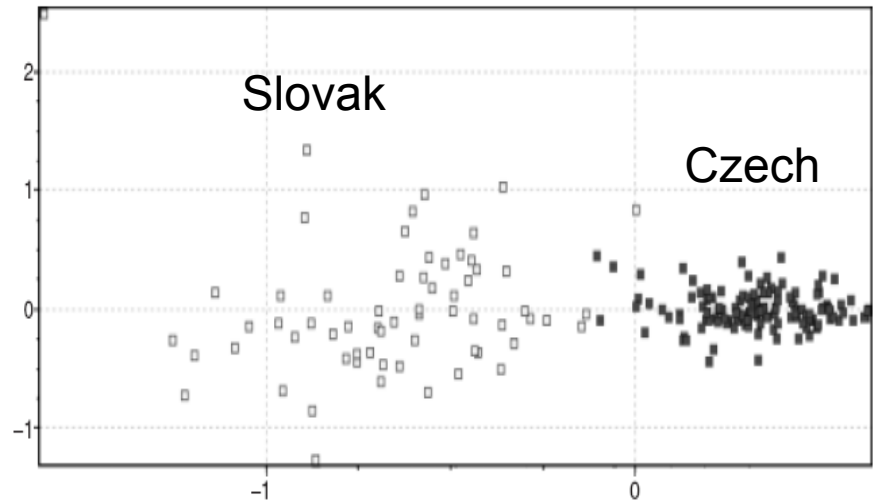
- **Komplexní Bayesiánský přístup** (založen na koalescenční teorii)
  - Detekce bottlenecku i expanze
  - Vhodné při dlouhodobějších změnách
  - Markov chain Monte Carlo simulations
  - Beaumont M. – např. program MSVAR nebo DIY ABC



Testování alternativních scénářů vývoje populací v programu DIY ABC

# Otters in central Europe

- strong decline of population numbers in last century
- fragmentation of distribution area



**Table 3** Simulated data from five replicate runs of the MSVAR programme

Population	Model	$n_0$	$n_1$	$\mu$	Xa (generations)	Xa (years)
CZE	EXP	<b>66.1</b> (1.7; 245 470.9)	<b>257.0</b> (12.6; 7079.5)	<b><math>1.6 \times 10^{-3}</math></b> ( $5.6 \times 10^{-5}$ ; $4.2 \times 10^{-2}$ )	<b>5.9</b> (0.2; 549.5)	<b>17.7</b>
	LIN	<b>28.2</b> (0.4; 93 325.4)	<b>281.8</b> (13.5; 7413.1)	<b><math>1.3 \times 10^{-3}</math></b> ( $4.3 \times 10^{-5}$ ; $3.7 \times 10^{-2}$ )	<b>7.2</b> (0.2; 660.7)	<b>21.6</b>
SVK	EXP	<b>147.9</b> (2.6; 323 593.6)	<b>588.8</b> (26.9; 14 791.1)	<b><math>1.1 \times 10^{-3}</math></b> ( $4.4 \times 10^{-5}$ ; $2.3 \times 10^{-2}$ )	<b>4.3</b> (0.1; 162.2)	<b>12.9</b>
	LIN	<b>123.0</b> (0.5; 363 078.1)	<b>575.4</b> (29.5; 16 218.1)	<b><math>1.0 \times 10^{-3}</math></b> ( $4.1 \times 10^{-5}$ ; $2.1 \times 10^{-2}$ )	<b>4.7</b> (0.1; 173.8)	<b>14.1</b>

$n_0$ , estimated mode (and limits) of current effective population size;  $n_1$ , estimated mode (and limits) of past effective population size;  $\mu$ , estimated mode (and limits) of mutation rate; Xa, estimated mode (and limits) of time between current and past population size in generations and years; CZE, Czech population; SVK, Slovak population; EXP, exponential model; LIN, linear model.



## 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

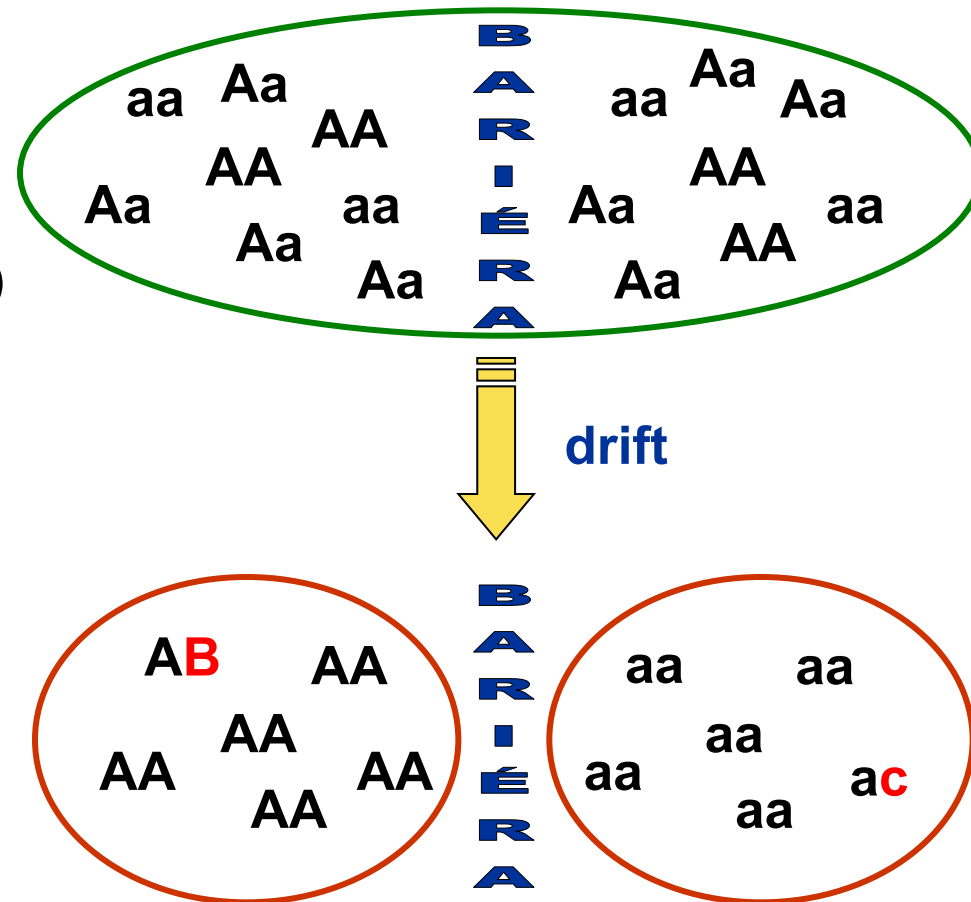
## drift, mutace

- **Drift**

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

- **Mutace**

mohou zvýšit diferenciaci (nebo ne – homoplázií)

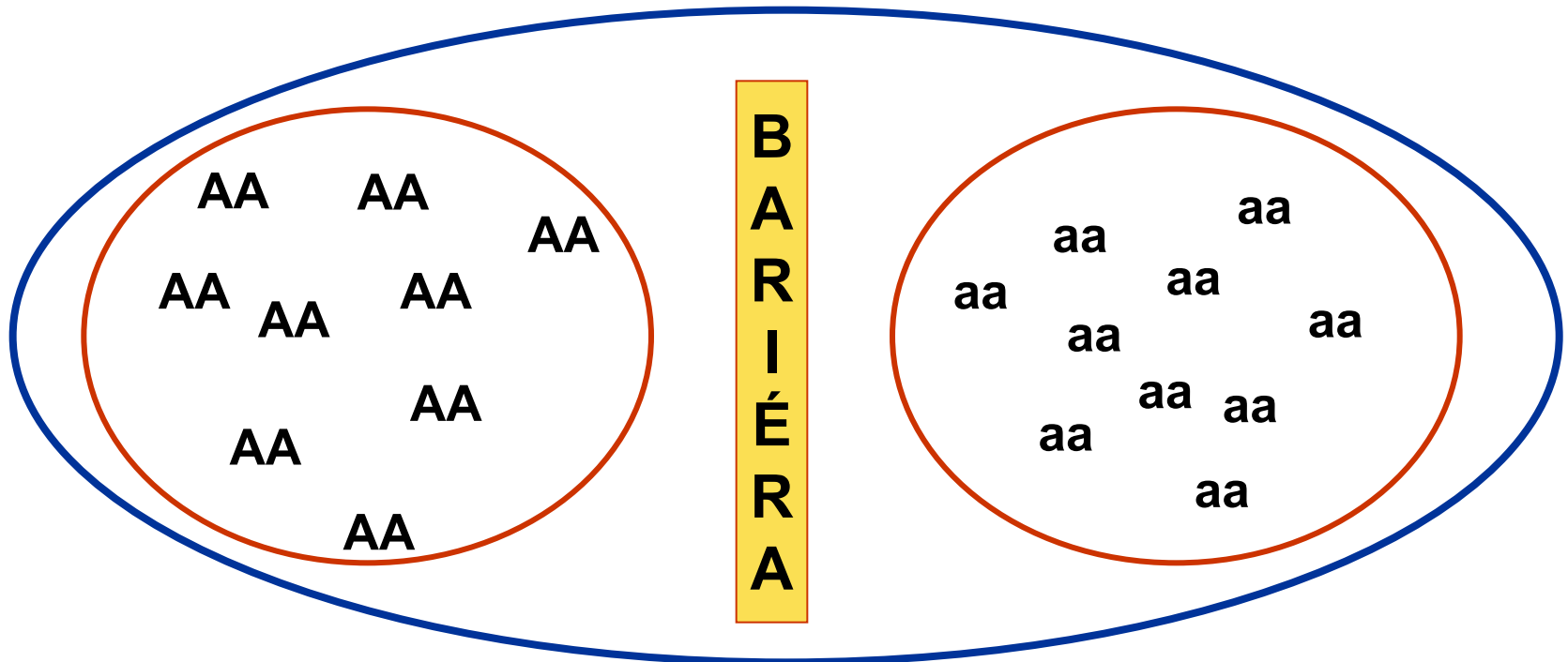


### Migrace (genový tok)

- působí proti diferenciaci subpopulací

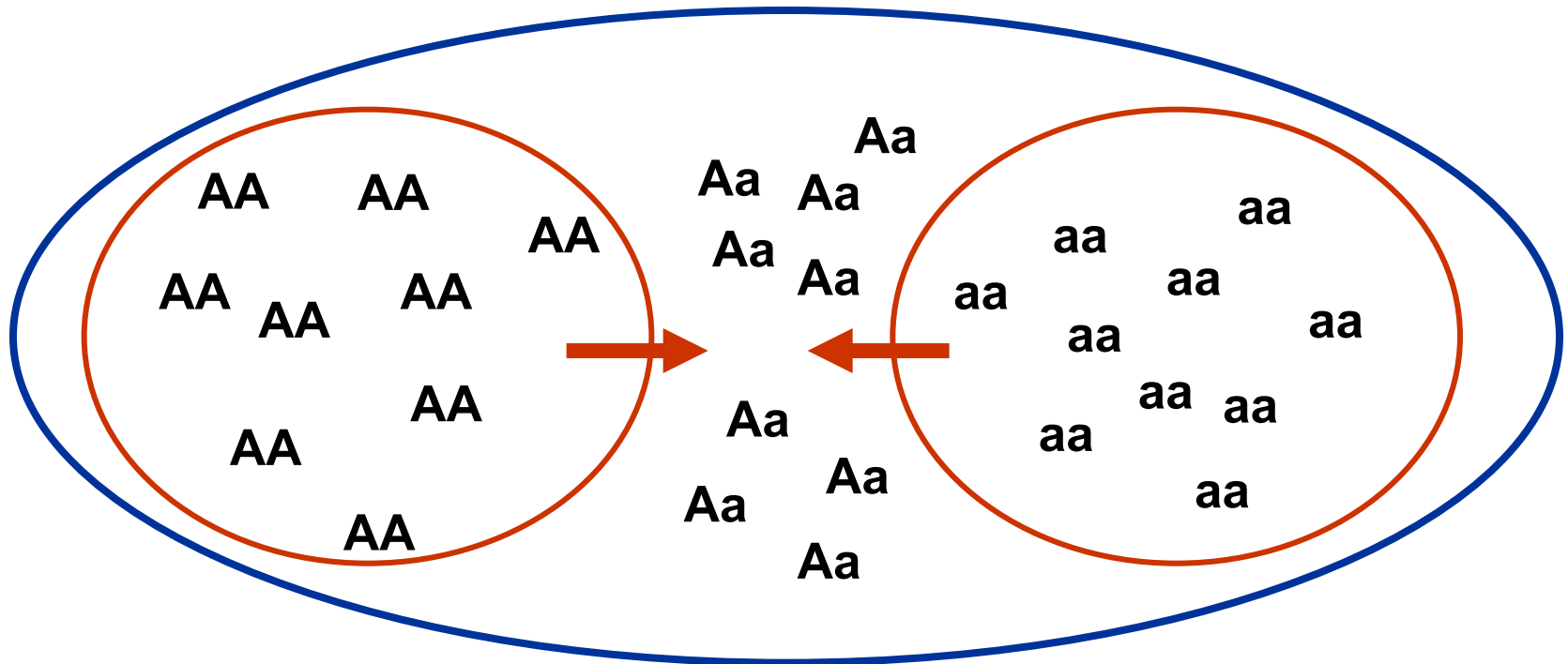
# Vliv populační struktury na heterozygotnost

- 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 Bunnarsjöarna (severní Švédsko) – brown trout
- 2 alely na alozymovém markeru

	100/100	100/null	null/null	Total	p	2pq
Přítok	50	0	0	50	1.000	0.000
Odtok	1	13	36	50	0.150	0.255
Celé jezero (expected)	51 (33.1)	13 (48.9)	36 (18.1)	100	0.575	0.489



# Jak zjišťovat strukturovanost populací?



Factorial correspondence analysis  
(Genetix)



# F-statistiky



**Masatoshi Nei**  
\*1931

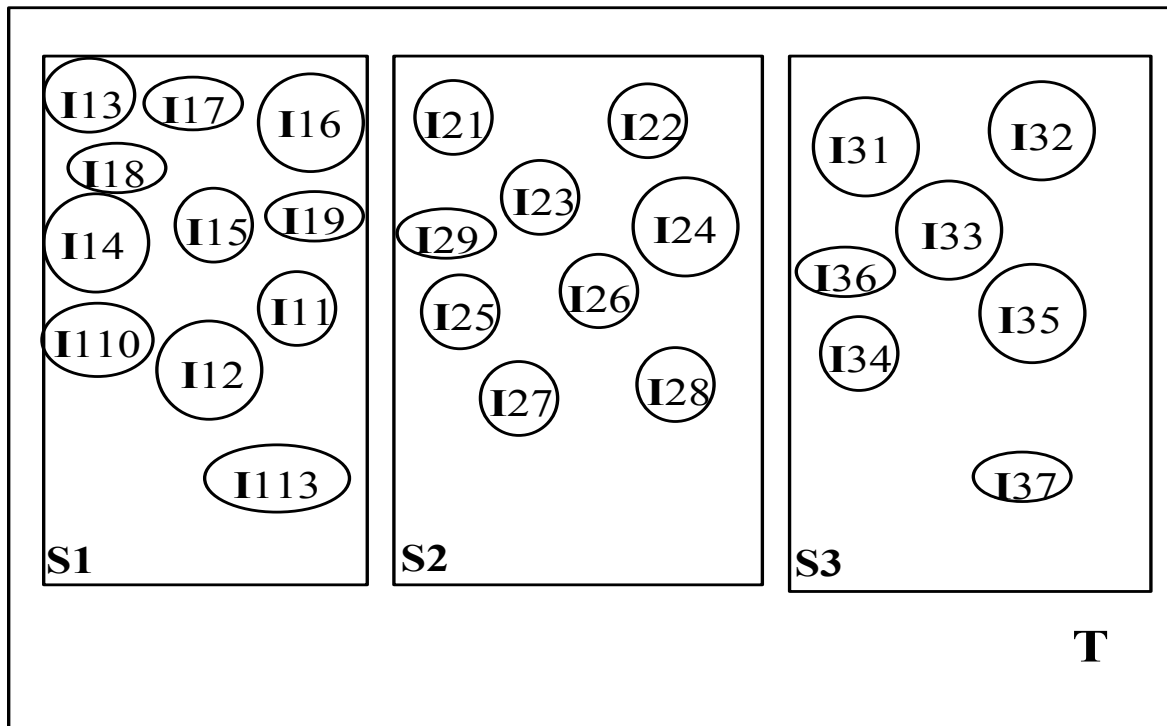


**Sewall Wright**  
1889 - 1988

- Wright, Nei  $F_{IS}$ ,  $F_{ST}$ ,  $F_{IT}$
- Popisují heterozygotnost (odchylky 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, Aa, aa – 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

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

$$H_S = 1 - \sum_{i=1}^j p_{i,x}^2 \quad p_{i,x}^2 = \text{frekvence } i\text{-té alely v subpopulaci } x \quad \overline{H_S} = \sum_{x=1}^j s / k \quad \text{průměrná oček. heterozygotnost v populaci}$$

$$H_T = 2 p_0 q_0$$

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

# F statistiky

$$F_{IS} = \frac{\overline{H}_s - I_I}{\overline{H}_s}$$

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

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

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 odlišnosti od nuly – nejčastěji permutace (FSTAT, Genetix, Genepop etc.)

# 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 <sub>0(j)</sub>	Pozn.
	AA	AB	BB	p <sub>1(j)</sub>	AA	AB	BB	p <sub>2(j)</sub>		
A	10	20	10	0.5	5	10	5	0.5	0.5	H.-W. rovnováha
B	16	8	16	0.5	4	4	12	0.3	0.4	deficit heterozygotů
C	12	28	0	0.65	6	12	2	0.6	0.625	přebytek heterozygotů
D	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 <sub>1(j)</sub>	H <sub>2(j)</sub>	H <sub>I(j)</sub>	H <sub>S(j)</sub>	H <sub>T(j)</sub>	F <sub>IS(j)</sub>	F <sub>ST(j)</sub>	F <sub>IT(j)</sub>
A	0.5	0.5	0.5	0.5	0.5	0.0	0.0	0.0
B	0.2	0.2	0.2	0.46	0.48	0.565	0.042	0.583
C	0.7	0.6	0.65	0.4675	0.46875	-0.39	0.0027	-0.387
D	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

# F-statistika

- $F_{IS}$  snížení heterozygotnosti v lokální subpopulaci  
vysoké hodnoty – inbreeding
- $F_{IT}$  souhrnná hodnota, heterozygotnost v celé populaci
- $F_{ST}$  míra „rozdělenosti“ = snížení toku mezi subpopulacemi (tj. existence bariéry – Wahlundův princip)  
Vliv driftu – fixuje odlišné alely v subpopulacích

# Výpočet $F_{ST}$ - příklad

- Jezero Bunnarsjöarna (severní Švédsko) – brown trout
- 2 alely na alozymovém markeru

	100/100	100/null	null/null	Total	p	2pq
Přítok	50	0	0	50	1.000	0.000
Odtok	1	13	36	50	0.150	0.255
Celé jezero (expected)	51 (33.1)	13 (48.9)	36 (18.1)	100	0.575	0.489

$$F_{ST} = \frac{H_T - \bar{H}_s}{H_T} = \frac{0.489 - 0.128}{0.489} = 0.728$$

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



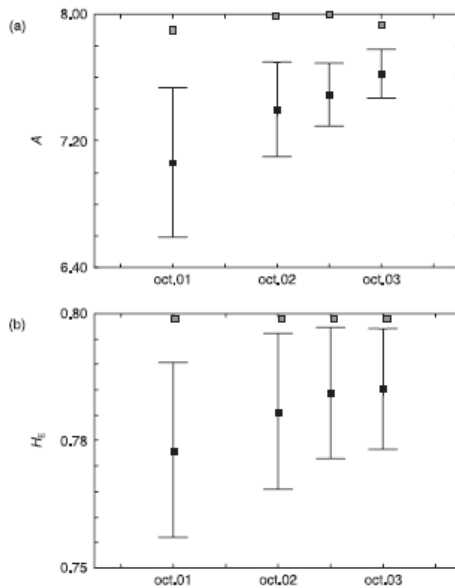


# Populační cykly hrabošovitých hlodavců (Berthier et al. 2006)

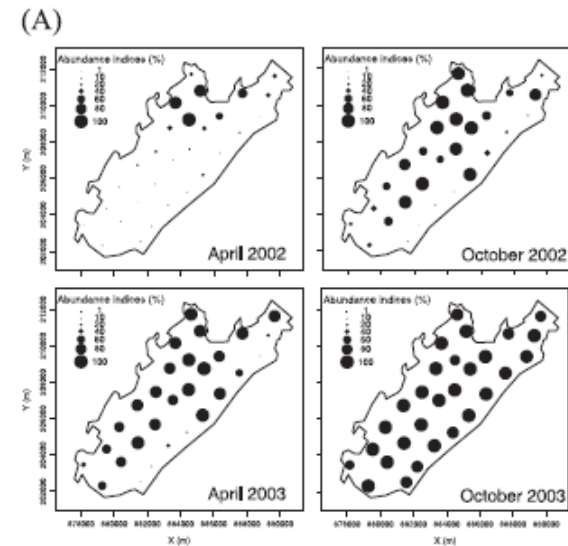


*Arvicola terrestris*

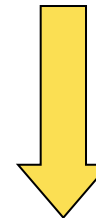
- Výrazné cyklické změny v početnosti populací – střídají se období „izolovanosti“ s „kontinuálním rozšířením“



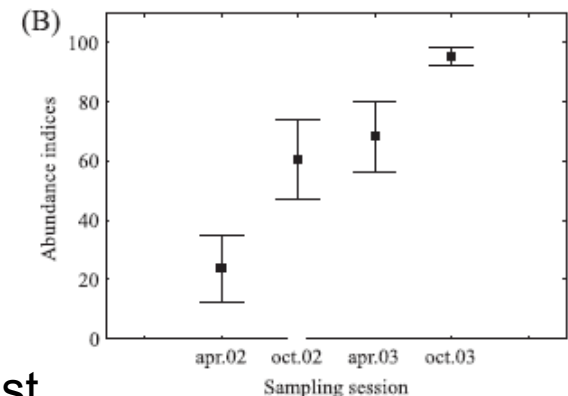
- Nízká hustota – extrémní vliv driftu
- S nárůstem populační hustoty se zvyšuje intenzita toku genů mezi subpopulacemi a narůstá genetická variabilita subpopulací



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  $F_{ST}$





# Hodnoty $F_{ST}$

- 0 – 0.05 malá diferenciace (zanedbatelná)
- 0.05 – 0.15 střední
- 0.15 – 0.25 velká
- > 0.25 velmi velká

**Absolutní hodnoty závisí na heterozygotnosti !!!**

**Nutno testovat odlišnost od nuly !!!**

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

# $G_{ST}$ (Nei 1973)

- Analogie  $F_{ST}$
- Haploidní (haplodiploidní) organismy, sekvence mtDNA
- Počítá s „gene diversity“ místo heterozygotnosti
- 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	$0.010 \pm 0.006$	$P = 0.01$
NN8	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	$0.961 \pm 0.050$	$P < 0.0005$

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).

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

# $R_{ST}$

- 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 rozdílů randomizačními testy (Hardy et al. 2003, program SPAGeDi 1.1)

# Lagopus lagopus scoticus

Piertney et al. 1998 a 2000

- Mikrosatelity a mtDNA
- $R_{ST}$ , AMOVA
- NcDNA strukturovanost
- mtDNA není strukturovanost
- **Filopatrie samců, disperze samic, průkazný vztah  $R_{ST}$  a vzdálenosti**

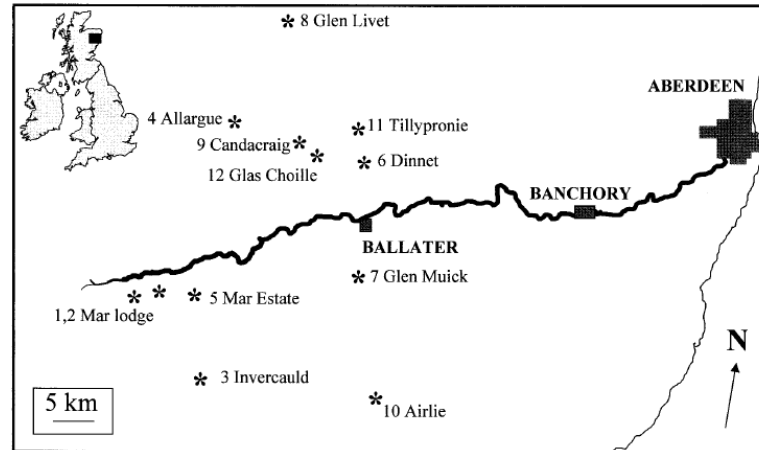


Table 3 Unbiased, multilocus estimates of Slatkin's (1995)  $R_{ST}$  between all population pairs. Significant values are underlined

	Mar Lodge 1	Mar Lodge 2	Invercauld 1	Invercauld 2	Allargue	Mar Estate	Dinnet	Glen Muick	Glen Livet	Tillypronie	Candacraig 1	Candacraig 2	Airlie
Mar Lodge 1													
Mar Lodge 2	0.039												
Invercauld 1	<u>0.227</u>	0.060											
Invercauld 2	-0.020	-0.011	<u>0.117</u>										
Allargue	0.048	<u>0.174</u>	<u>0.380</u>	0.075									
Mar Estate	<u>0.152</u>	<u>0.285</u>	<u>0.477</u>	<u>0.177</u>	0.009								
Dinnet	0.017	<u>0.137</u>	<u>0.352</u>	0.040	-0.016	<u>0.049</u>							
Glen Muick	-0.032	0.009	<u>0.153</u>	-0.033	0.024	<u>0.110</u>	-0.002						
Glen Livet	<u>0.181</u>	0.309	<u>0.491</u>	<u>0.204</u>	0.028	-0.012	<u>0.074</u>	<u>0.137</u>					
Tillypronie	0.328	<u>0.448</u>	<u>0.619</u>	<u>0.334</u>	<u>0.117</u>	0.026	<u>0.195</u>	<u>0.252</u>	-0.004				
Candacraig 1	0.332	<u>0.458</u>	<u>0.641</u>	<u>0.333</u>	<u>0.100</u>	0.009	<u>0.185</u>	<u>0.243</u>	-0.017	-0.039			
Candacraig 2	0.236	<u>0.374</u>	<u>0.565</u>	<u>0.254</u>	0.040	-0.019	<u>0.103</u>	<u>0.173</u>	-0.030	-0.025	-0.045		
Airlie	<u>0.249</u>	<u>0.384</u>	<u>0.574</u>	<u>0.263</u>	0.051	-0.008	<u>0.166</u>	<u>0.182</u>	-0.020	-0.012	-0.028	-0.039	
Glas Choille	-0.001	<u>0.040</u>	<u>0.208</u>	-0.013	<u>0.044</u>	<u>0.136</u>	0.017	-0.025	<u>0.163</u>	<u>0.239</u>	0.291	<u>0.209</u>	<u>0.220</u>

# AMOVA

*Excoffier et al. 1992*

Arlequin ver. 2.000

A software for  
population genetics  
data analysis

Authors:

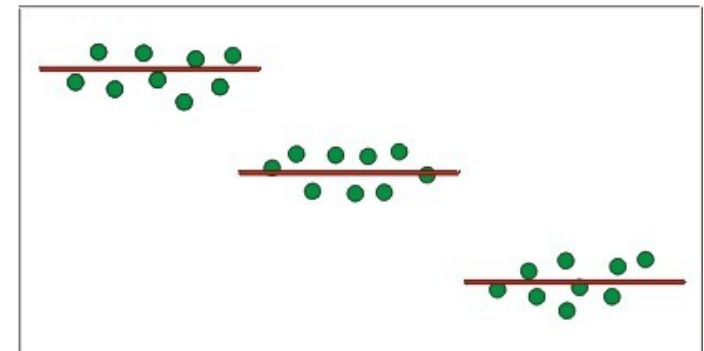
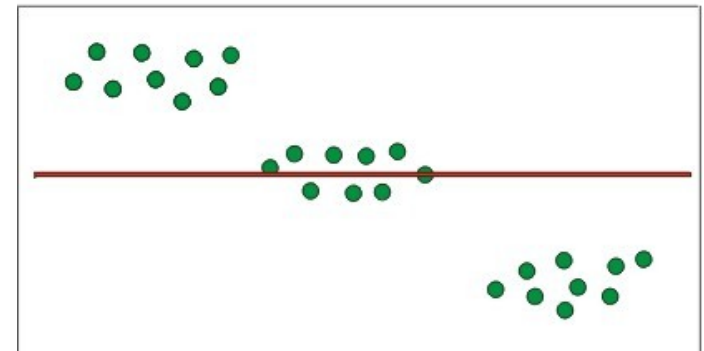
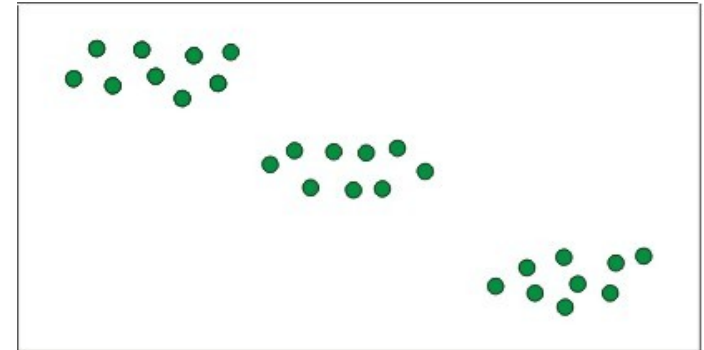
Stefan Schneider  
David Roessli  
Laurent Excoffier

Contact Arlequin:

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

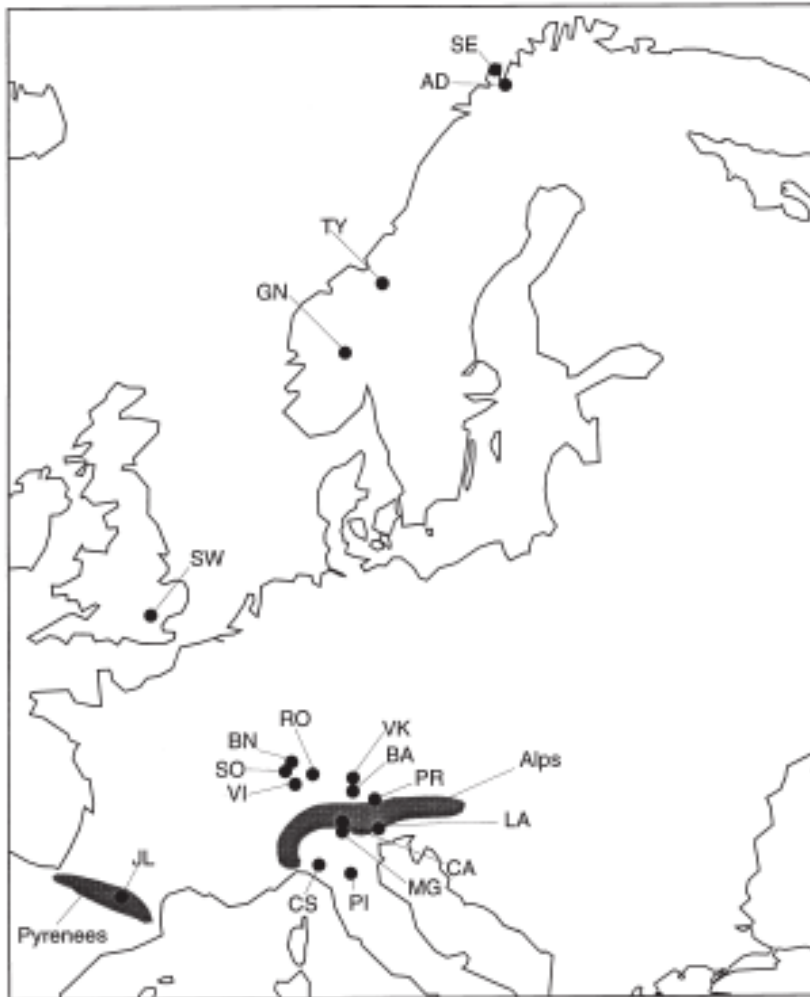


- **A**nalýza **M**olekulární **V**ariance
- Analogie  $R_{ST}$  a spojení s ANOVA
- Analýza variance alelických frekvencí  
(již dříve Cockerham & Weir  
1987, 1993)
- Započítává se ale rozdílnost (mutace)  
alel
- $\Phi_{ST}$ , Program ARLEQUIN (umí i  $F_{ST}$ )
- Data:  
sekvence  
mikrosatelity (jen při platnosti SMM  
*stepwise mutation model*)



# *Bombus pascuorum*

Widmer & Schmid-Hempel 1999



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

†Sum of squared deviations.

\*  $P < 0.001$ .

**Mikrosatelite, AMOVA  
Nejvíc vysvětlují Alpy**





# Neighbour-joining

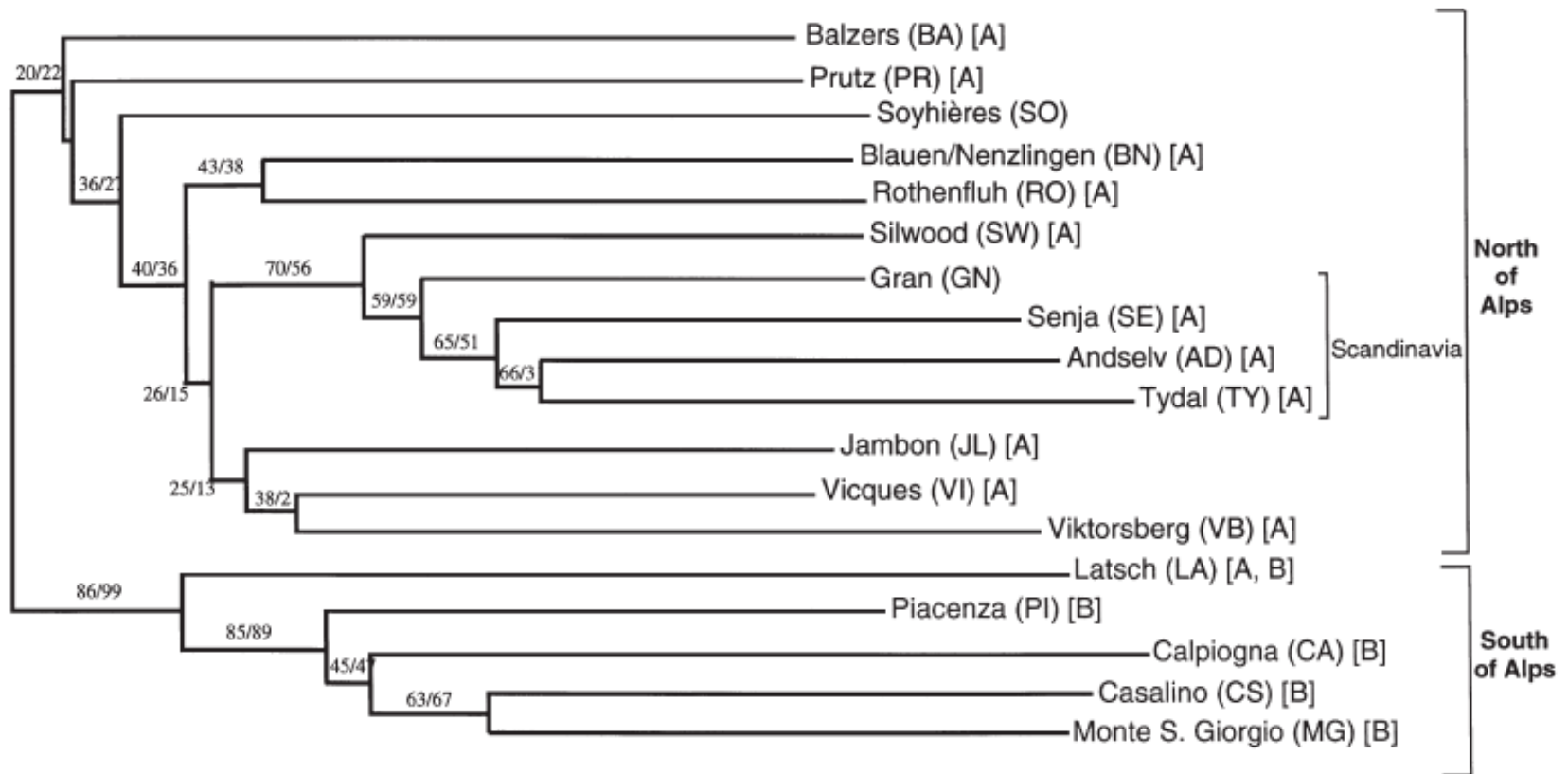


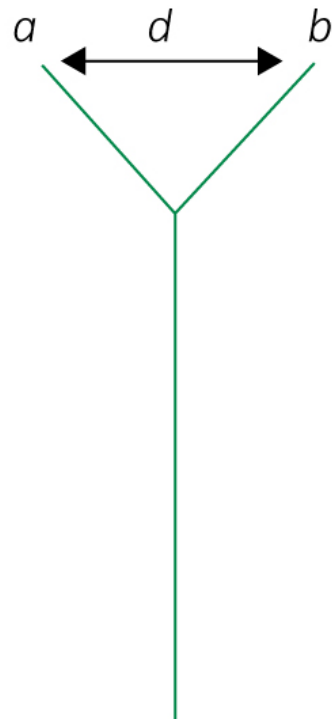
Fig. 2 Neighbour-joining (NJ) dendrogram based on the chord distance of microsatellite data (Cavalli-Sforza & Edwards 1967) connecting 18 populations of *Bombus pascuorum*. Bootstrap values (percentages) were computed over 2000 replications by resampling individuals within populations (left values) and by resampling loci (right values). Letters in parentheses denote mtDNA haplotypes.



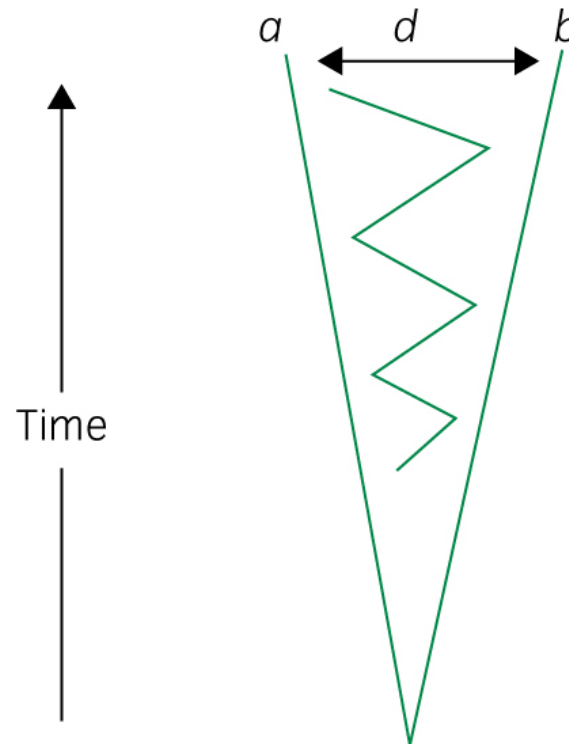
# AMOVA a F statistika

popis výsledku nikoliv příčin → možná alternativní vysvětlení

Recent separation,  
no gene flow

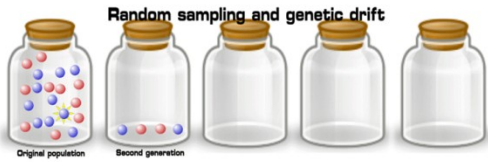


Old separation, but  
continuous (low)  
gene flow

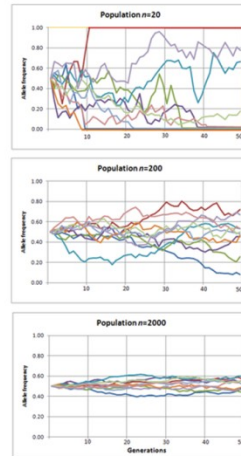


Time

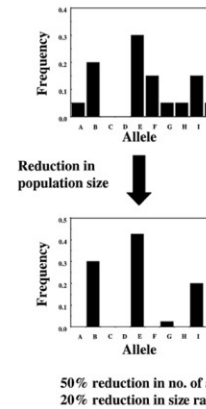
# Genetický drift



- Náhodné změny ve frekvencích alel
- Intenzita driftu závisí na velikosti populace (viz ochránářská genetika)
- Specifické případy – founder effect, bottleneck



# Detekce bottlenecku



Při bottlenecku vymizí nejdříve vzácné alely, rychleji než se sníží heterozygotnost nebo rozsah alel

Nutno znát (předpokládat) mutační model – pak se nasimuluje „mutation-drift“ rovnováha a srovnává se se skutečným stavem

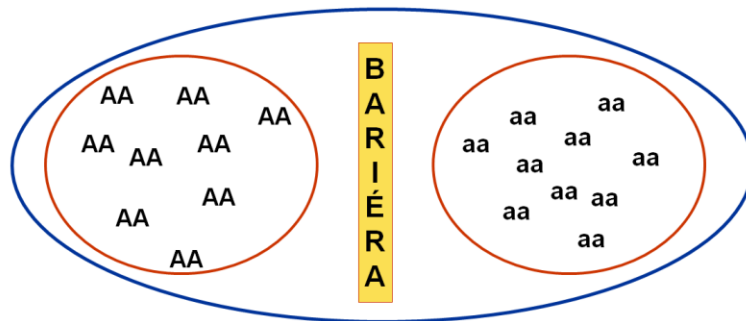
**Program BOTTLENECK:** počet alel vs. heterozygotnost

**Program M Ratio:** počet alel vs. rozsah alel

Předpoklad testů: Populace v HW rovnováze

# Vliv populační struktury na heterozygotnost

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



# F statistiky

$$F_{IS} = \frac{\bar{H}_S - H_I}{\bar{H}_S} \quad \text{Snížení heterozygotnosti jedince kvůli nenáhodnému páření v subpopulaci (~ HWE)}$$

$$F_{ST} = \frac{H_T - \bar{H}_S}{H_T} \quad \text{Vliv rozdělení populace na subpopulace}$$

$$F_{IT} = \frac{H_T - H_I}{H_T} \quad \text{Celkový koeficient inbreedingu } F_{IT} \text{ - měří redukcí heterozygotnosti jedince ve vztahu k celkové populaci}$$

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

Weir & Cockerham (1984)  $f, \theta, F$   
Korekce na velikost vzorku a počet subpopulací

# Tok genů – metody

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

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

- **Privátní alely** (Slatkin)  
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** 
$$F_{ST} = \frac{1}{1 + 4N_e m}$$

## Předpoklady použití $N_e m$ :

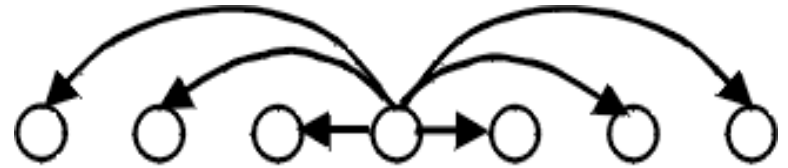
- „island model“ (= infinite number of populations, absence of selection, the same size of all populations etc.)
- migration-drift equilibrium (= no range expansions, habitat fragmentation or population bottlenecks)

## 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  $N_e m$

# Modely toku genů (alel)

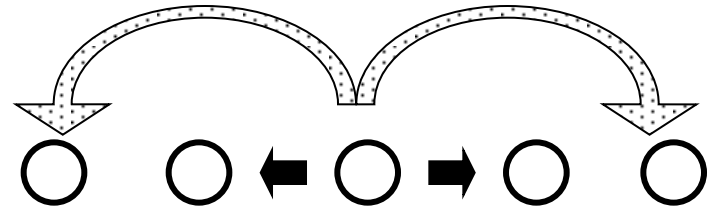
- **Island model** (Wright)  
Stejně velké subpopulace  
Symetrický tok genů  
Stejná pravděpodobnost výměny  
mezi jakýmkoliv subpopulacemi



- **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

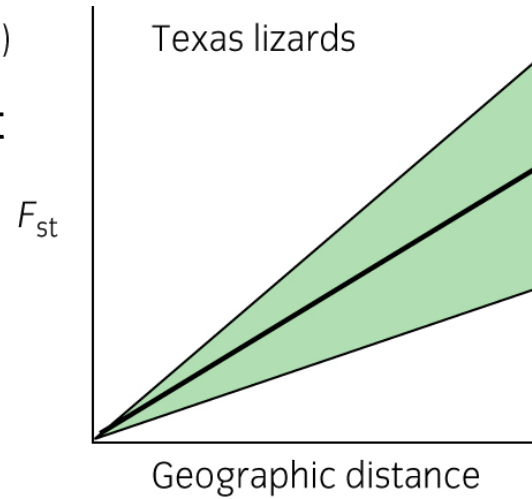
*Crotaphytus collaris*

Hutchinson & Templeton 1999

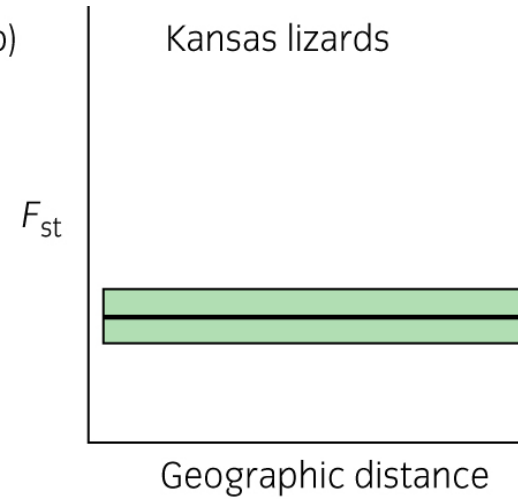


desítky tisíc let  
nejsou bariéry  
**rovnováha**  
mezi driftem  
a migrací

(a)



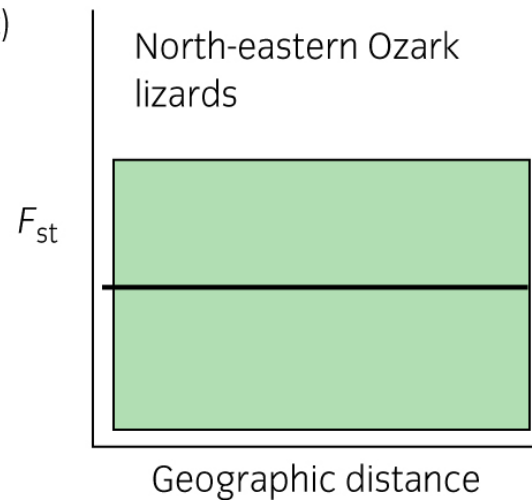
(b)



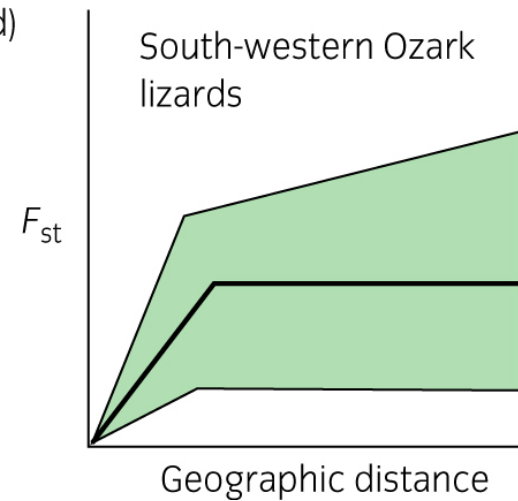
postglaciálně  
nejsou bariéry  
**vliv migrace**

postglaciálně  
fragmentace  
**vliv driftu**

(c)

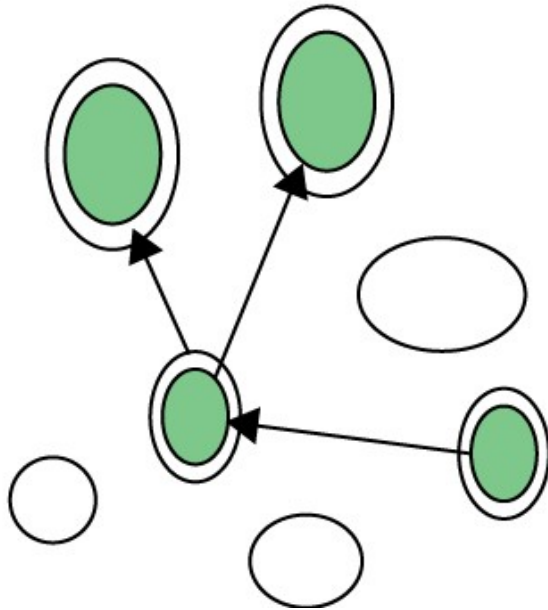


(d)



postglaciálně  
vzrůstající  
fragmentace  
**vliv driftu**  
na velkých  
měřítkách  
**rovnováha**  
na malých  
měřítkách

# 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

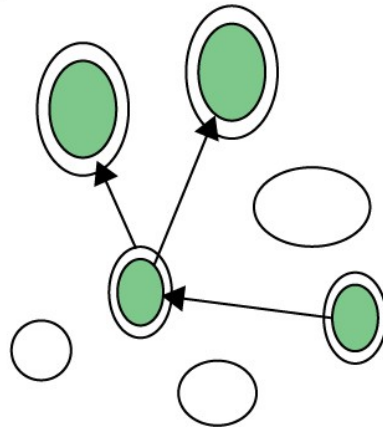


Variance component	D.f.	% of total variance	F-statistic	P
Mezi přítoky	1	0,1	0,0014	0,167
Mezi lokalitami	6	0,9	0,0085	<b>0,001</b>
Mezi roky uvnitř lokality	7	2,5	0,0255	<b>&lt; 0,0001</b>
Uvnitř vzorku	1352	96,6	0,0337	<b>&lt; 0,0001</b>

# Který typ metapopulační struktury?

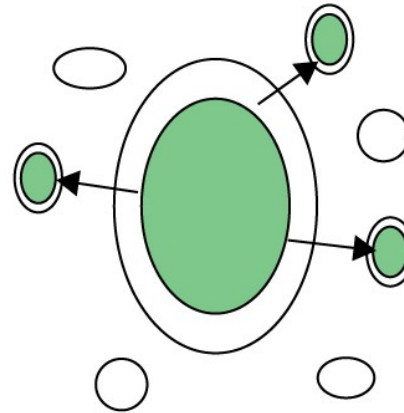
„classical“

(a)



„mainland-island“

(b)

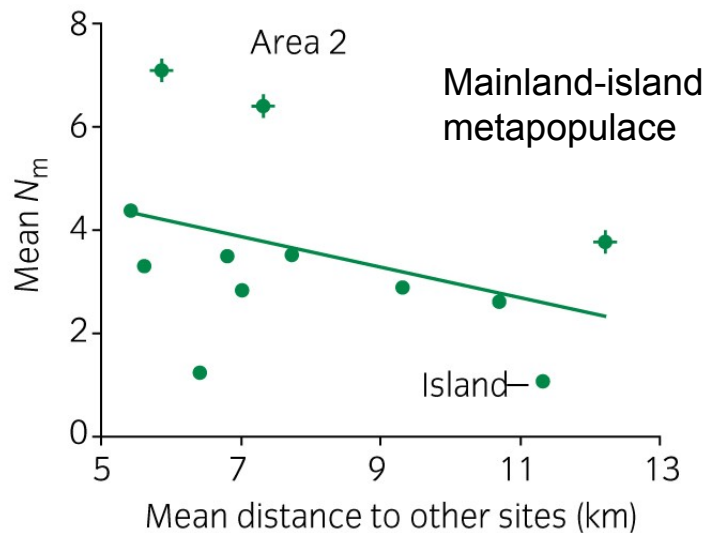
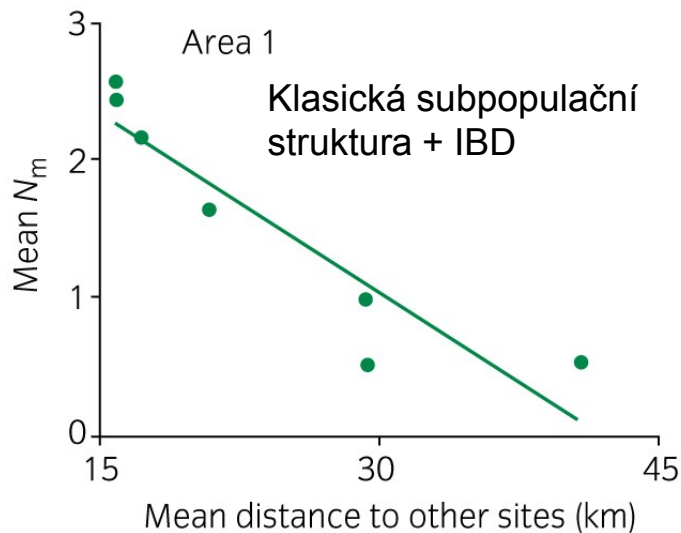


„source and sinks“

- 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 isolation-by-distance** in genetic differentiation)

# 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“)



$$F_{ST} = 0.06-0.22 \longrightarrow F_{ST} = \frac{1}{1 + 4N_m}$$

# Populační struktura - shrnutí

	Connected populations (gene flow)	Isolated populations (no gene flow)
Ne	<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"/>



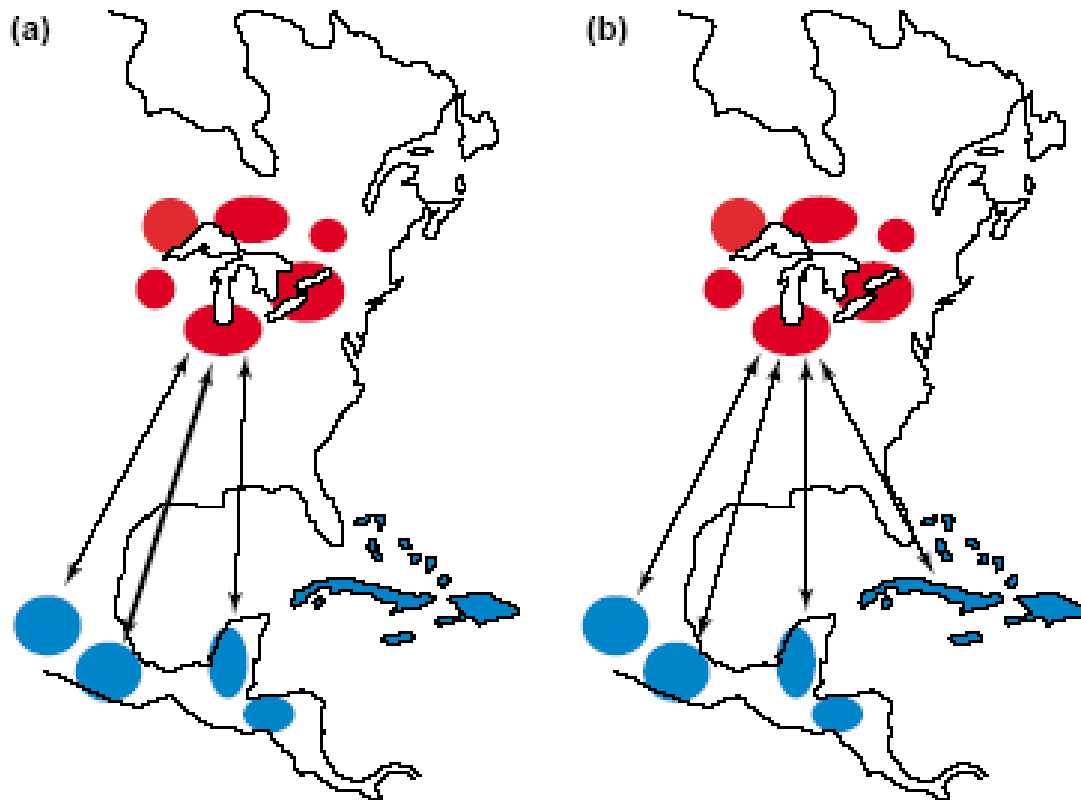
1. vnitropopulační variabilita – deskriptivní statistiky
2. popis populačně-genetické struktury, bariéry toku genů

## 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**)

# 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

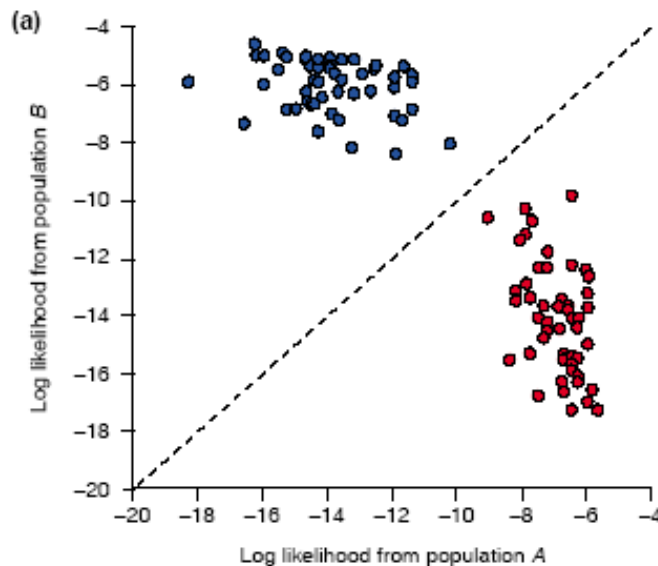
(3) tissue samples from individuals at different parts of their annual cycle

# C. Genetic approaches

- « very few birds have bands, but all have genotypes »
- genetic data on population structure
- problems: (1) weak genetic differentiation among populations (widespread dispersal), (2) lack of differentiation in northern temperate zone – recent postglaciation expansion
- solution: (1) more genetic markers, (2) study of avian parasites DNA

# Population assignment tests

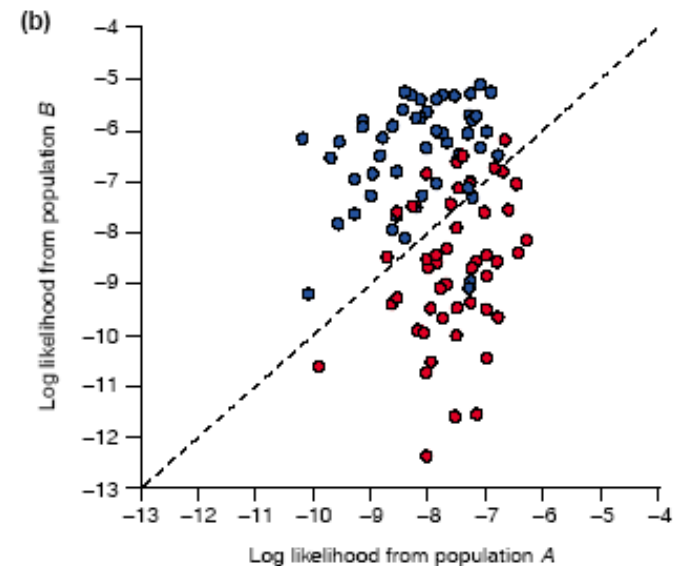
- program **GeneClass** (Piry et al. 2004)
- calculates the probability that an individual's genotype might exist in a particular population
- can combine data from multiple genetic marker types



5 microsatellite loci

$$F_{st} = 0.14$$

99.9% assigned correctly



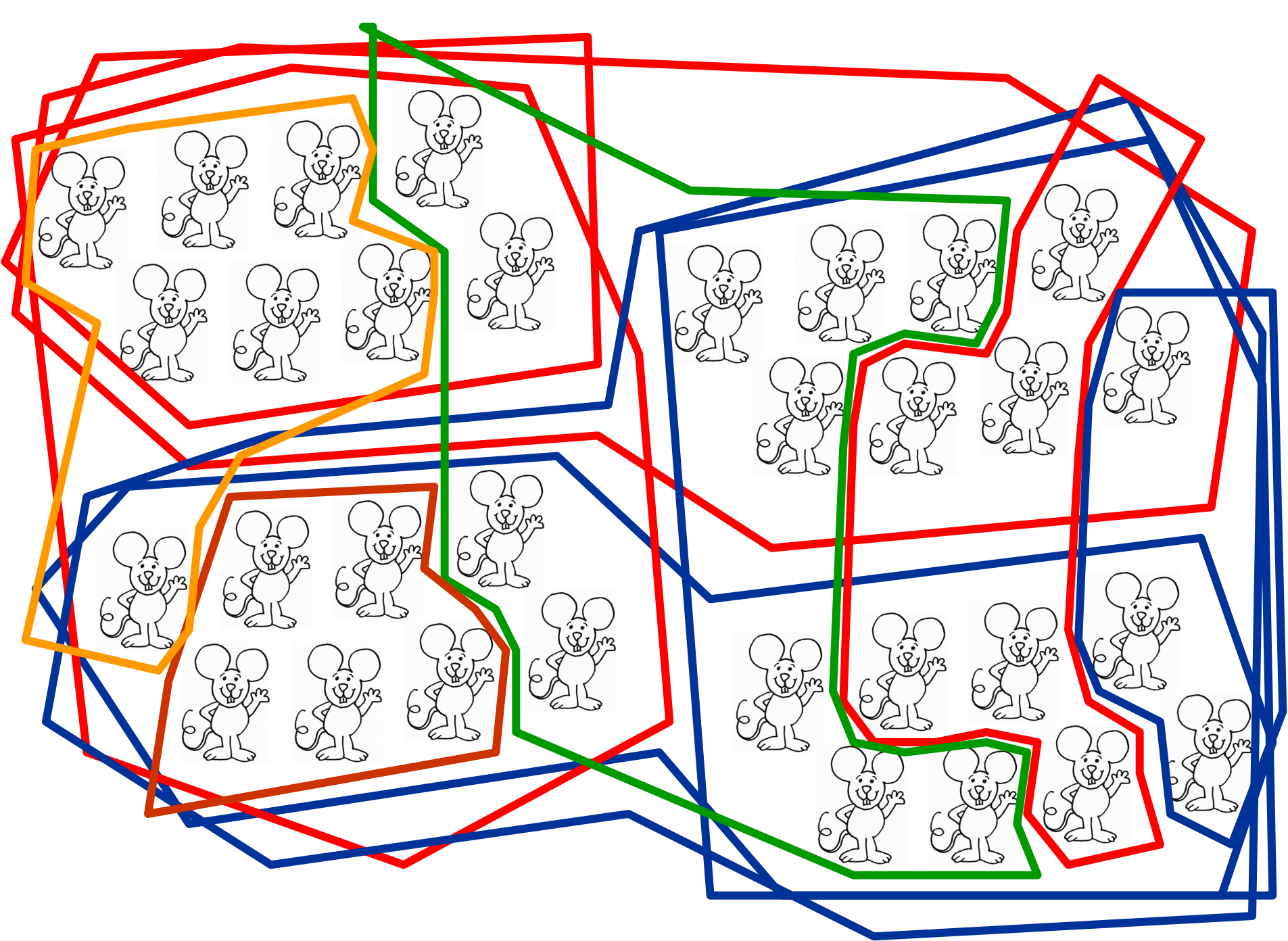
5 microsatellite loci

$$F_{st} = 0.04$$

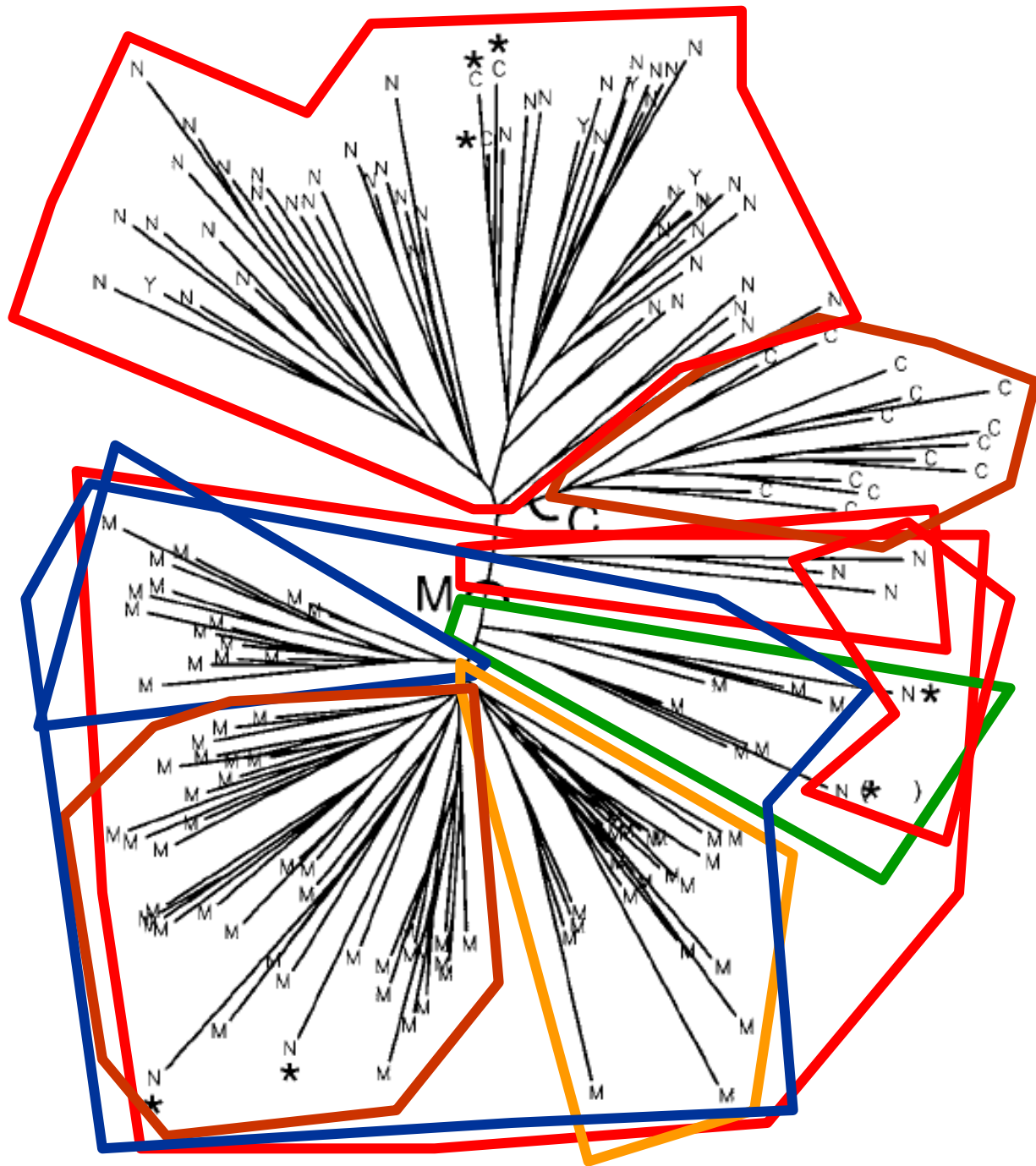
90.2% assigned correctly

# 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, *Alu*), AFLP







INVITED REVIEW

# Statistical inferences in phylogeography

RASMUS NIELSEN\*† and MARK A. BEAUMONT‡

*\*Departments of Integrative Biology and Statistics, University of California, Berkeley, 4096 VLSB, Berkeley, CA 94720, USA,*

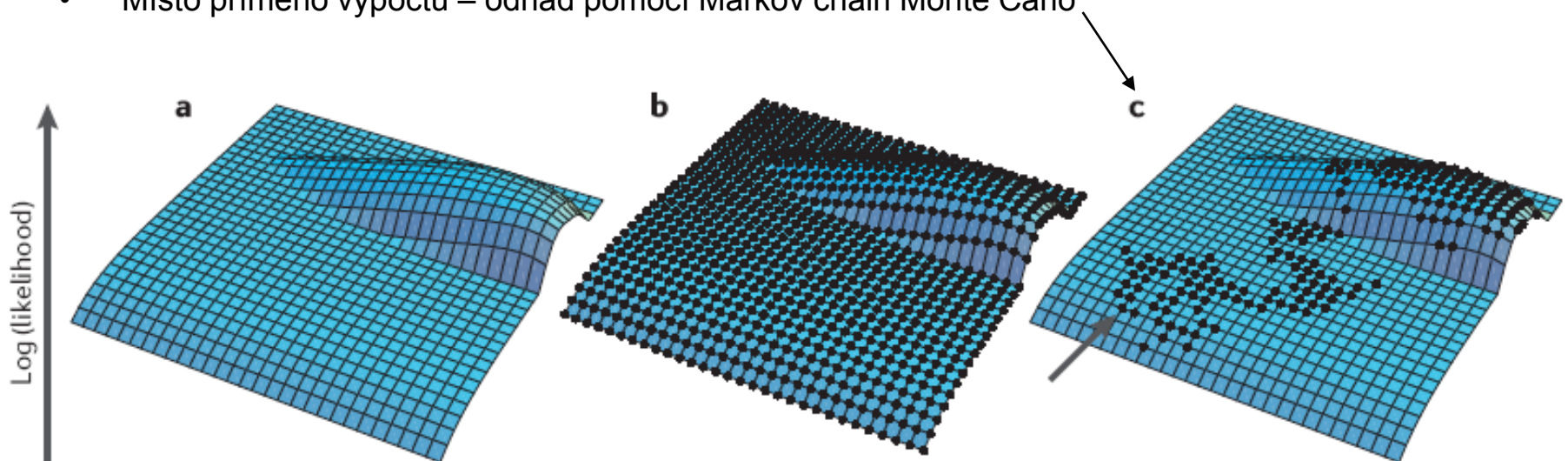
*†Department of Biology, University of Copenhagen, Universitetsparken 15, 2100 Kbh O, Denmark, ‡School of Biological Sciences, University of Reading, PO Box 68, Whiteknights, Reading RG6 6BX, UK*

- Therein lies the fundamental difference between phylogeographic and theoretical population genetic thinking: phylogeographic studies traditionally assume that ancestral history can be directly deduced from estimated gene trees, whereas population genetic theory asserts that **gene trees are random outcomes of stochastic population-level processes.**

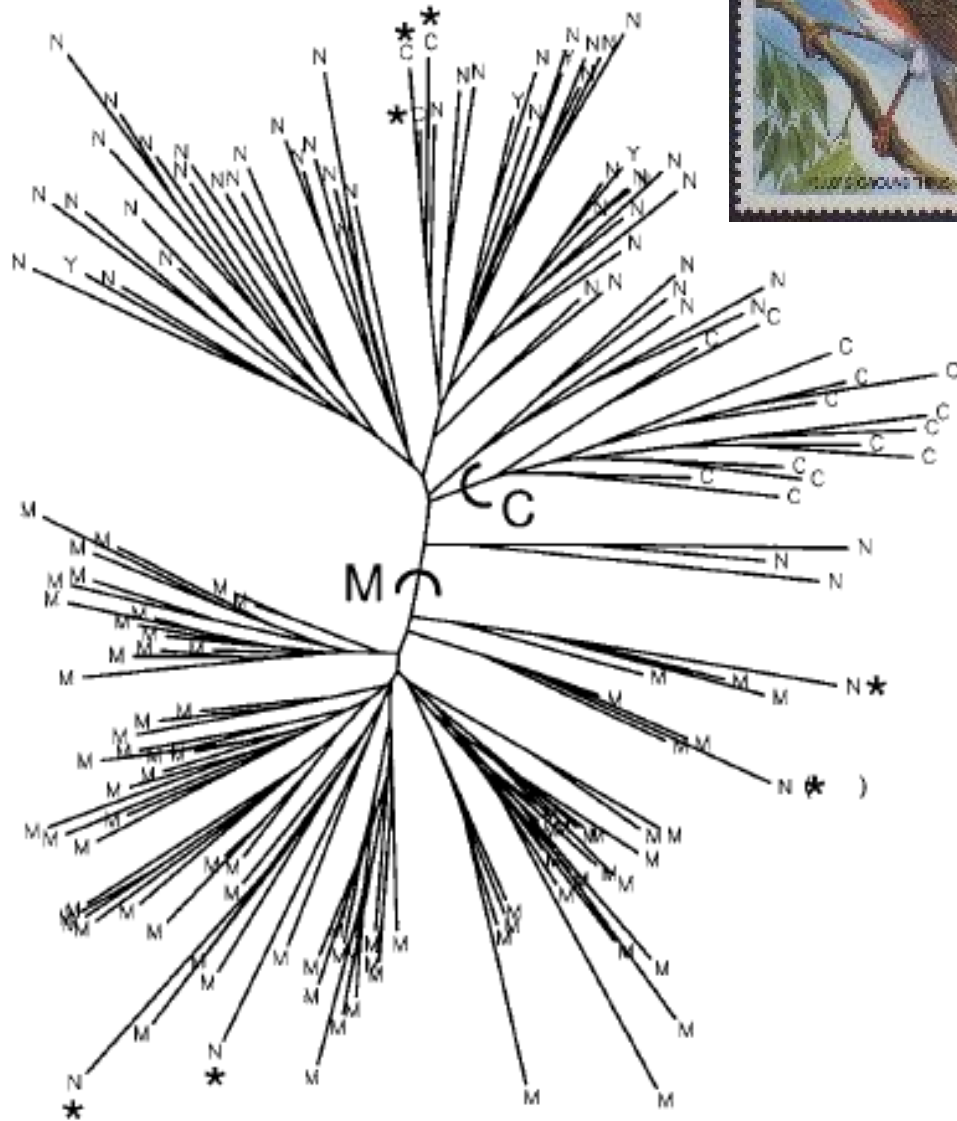
# Bayesian clustering approach

## 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)
- Možno předem zahrnout geografickou polohu individuí
- Model se snaží vyložit HW nebo vazebnou nerovnováhu zavedením populační struktury
- Místo přímého výpočtu – odhad pomocí Markov chain Monte Carlo



# *Turdus helleri*

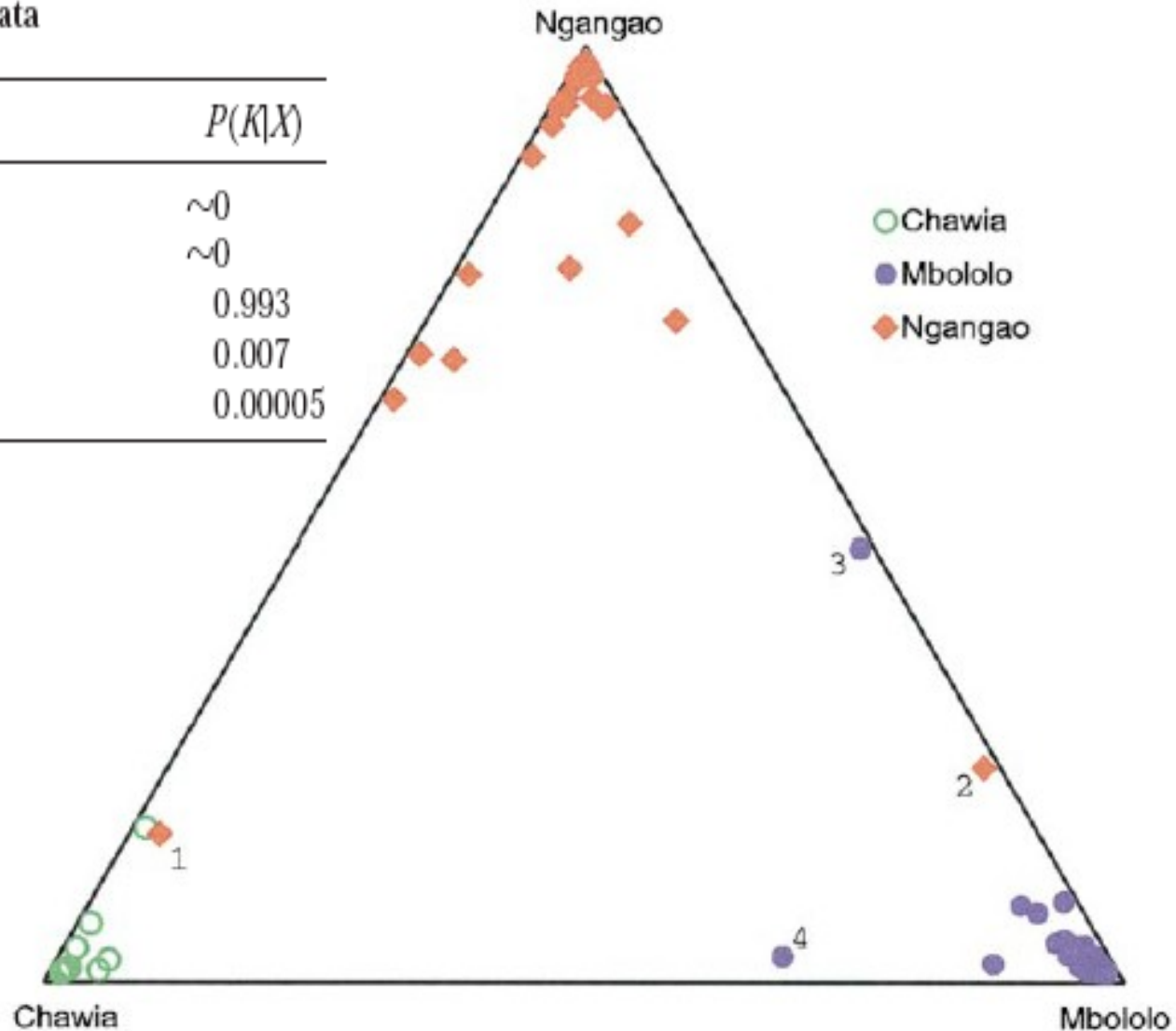


- Fragmenty deštného pralesa
- Lokality Chawia, Ngangao, Mbololo, Yale (Kenya)
- 7 mikrosatelitových lokusů
- Neighbour-joining
- \* špatně zařazení jedinci

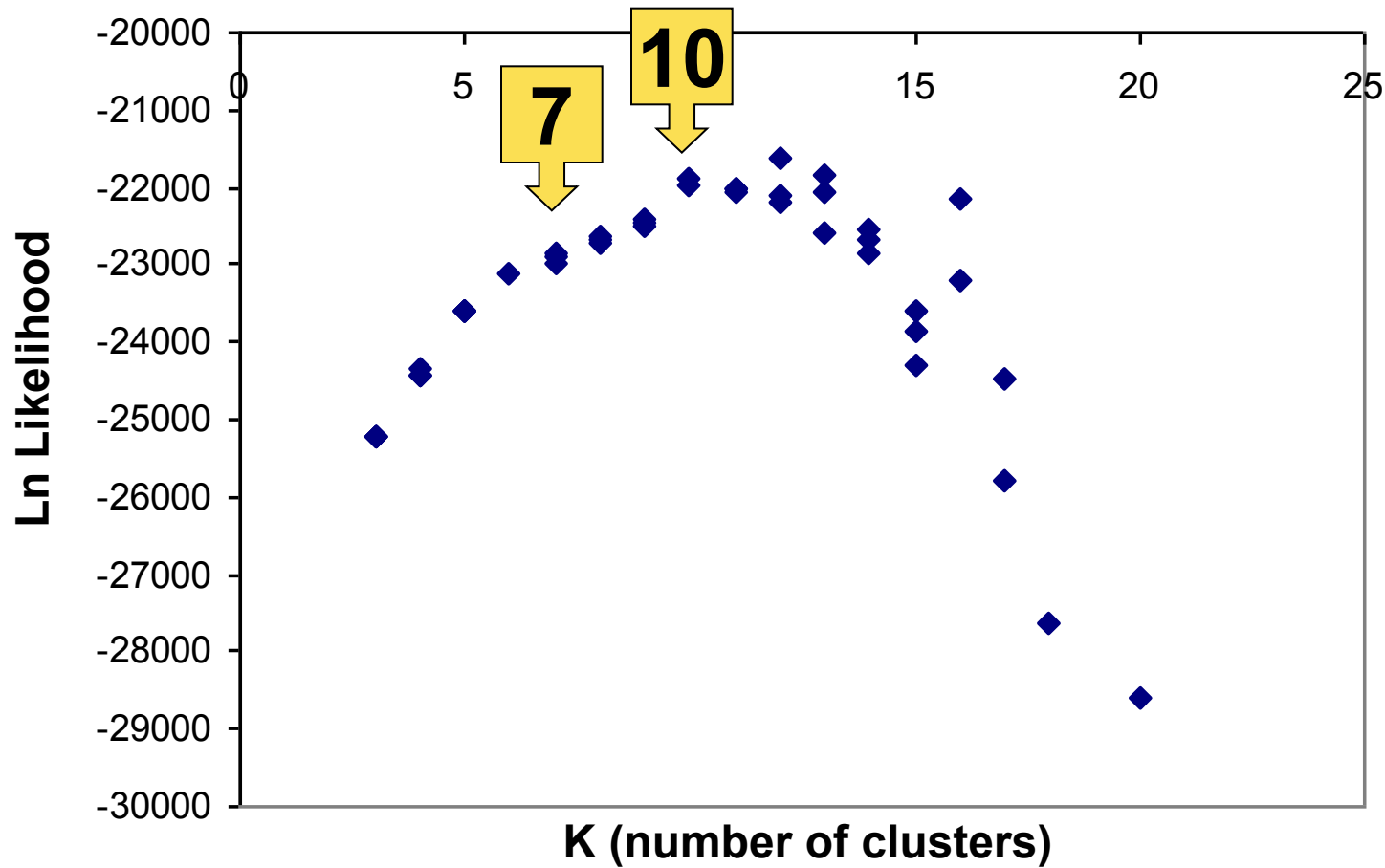
# Program STRUCTURE - Bayesiánský přístup

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



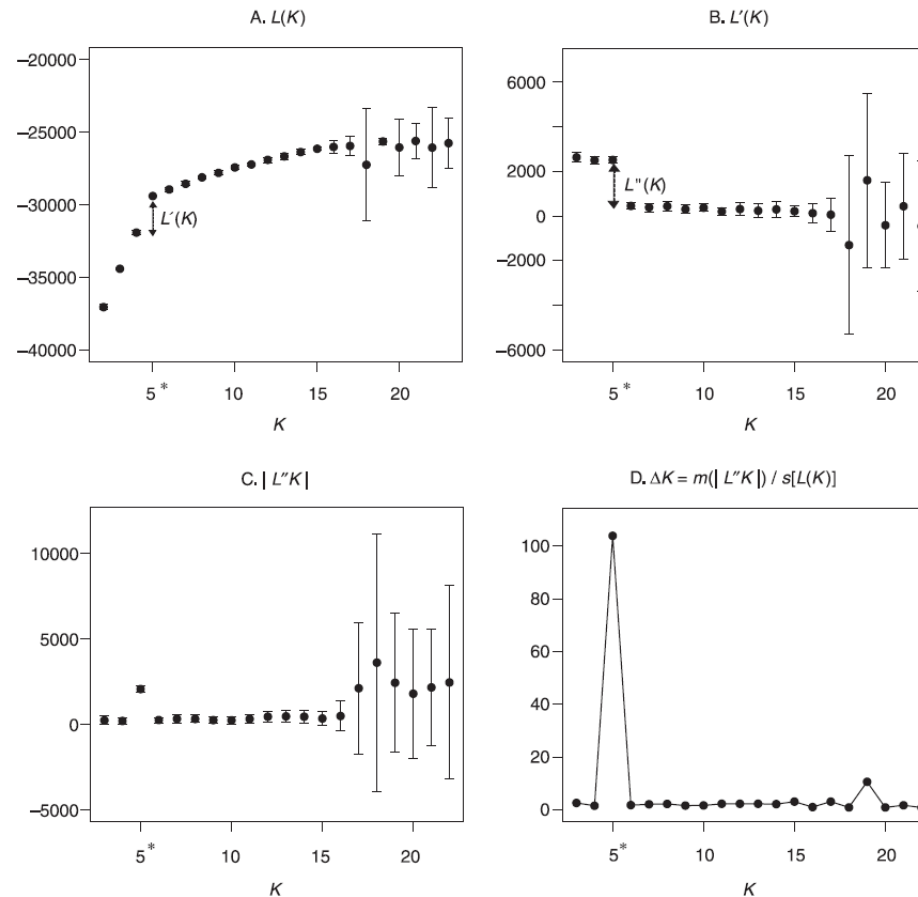
# Stanovení počtu „přirozených“ subpopulací



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

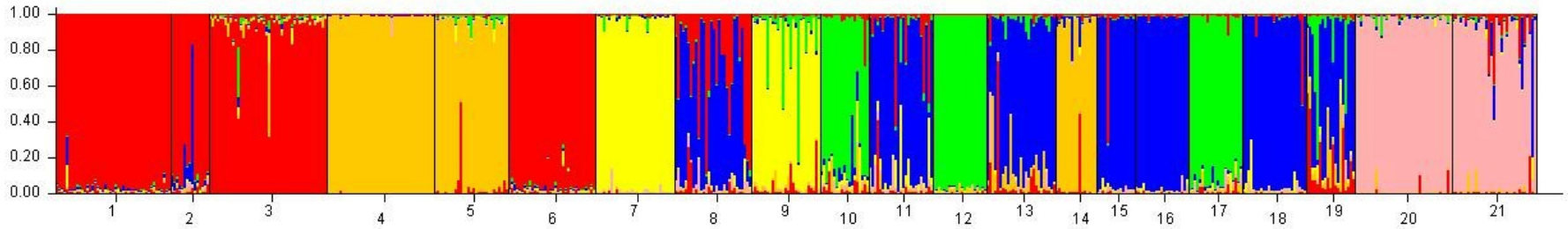
G. EVANNO, S. REGNAUT and J. GOUDET

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

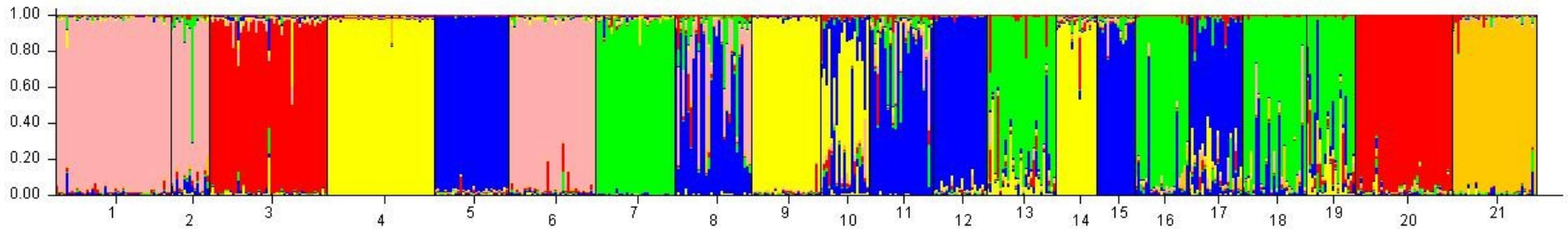




# K = 7



# K = 10

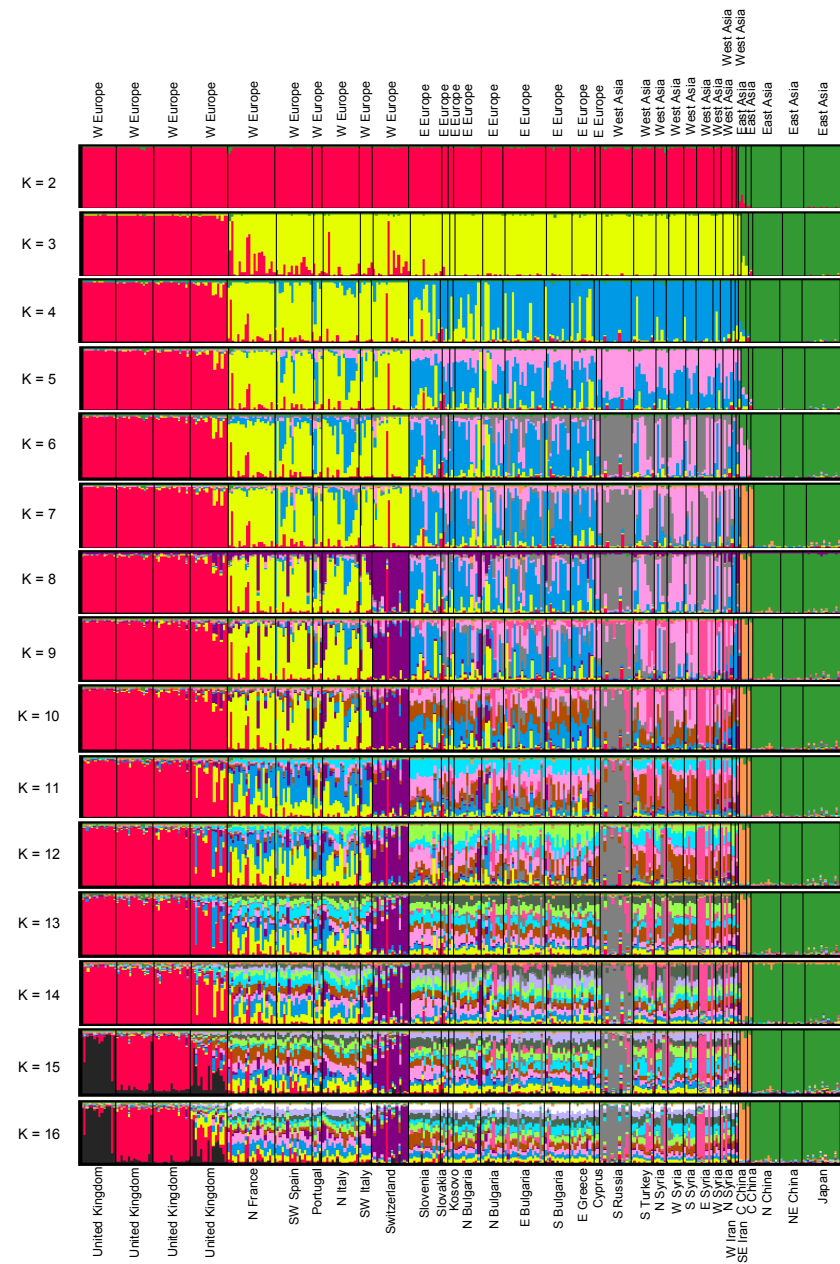


Proporce genomu každého jedince náležející určitému „clusteru“

# Alternativní vizualizace výsledků ze STRUCTURE

„forced clustering“

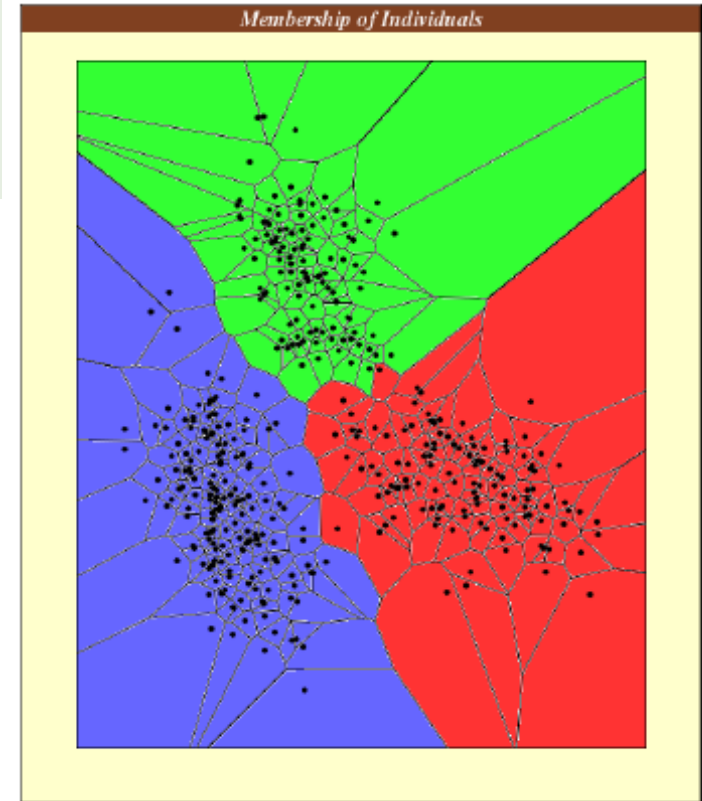
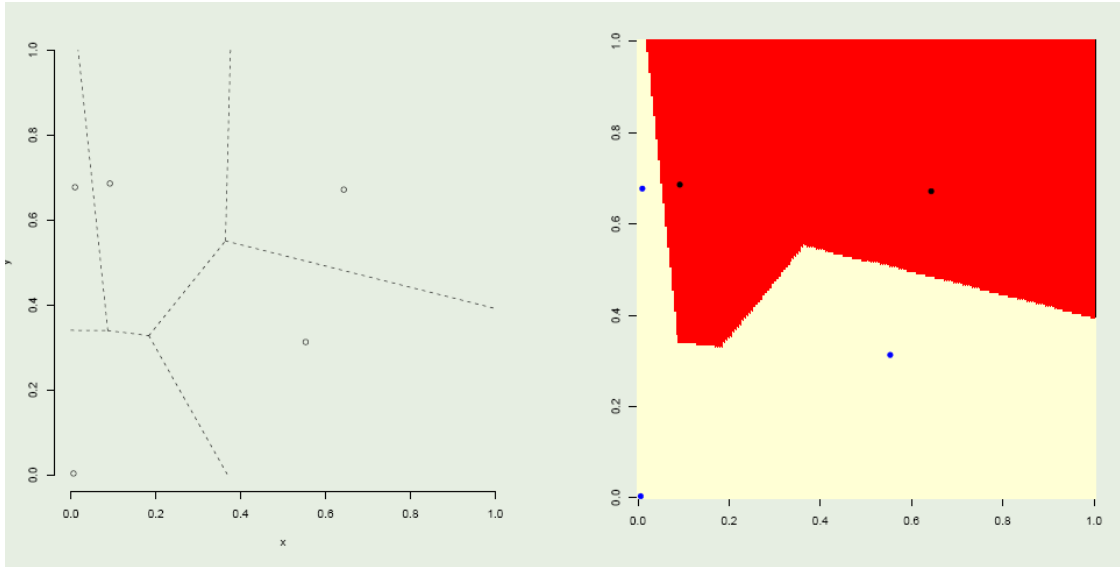
Zobrazení hierarchické struktury mezi populacemi



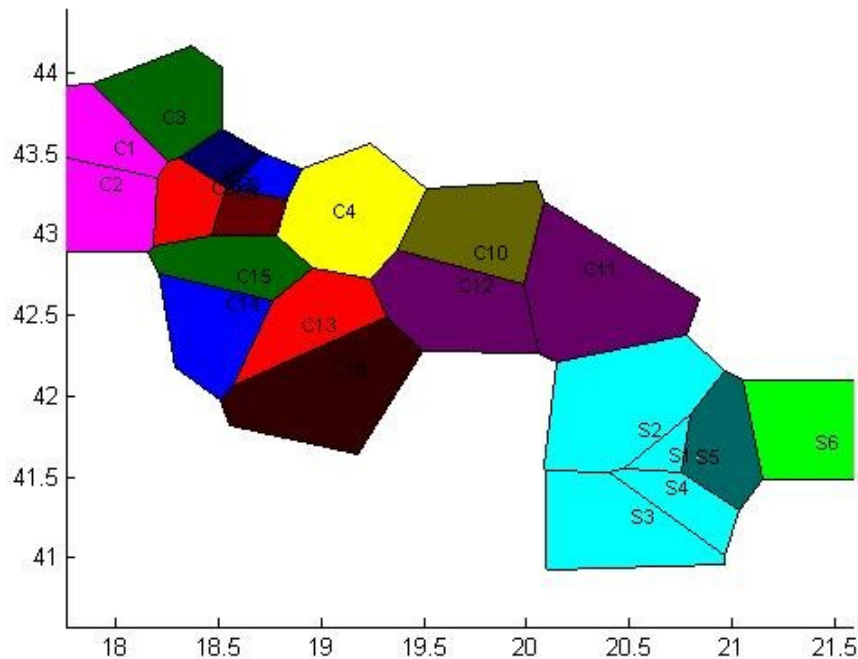
## 4. Spatially explicit analyses = spatial genetics = landscape genetics

- 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)

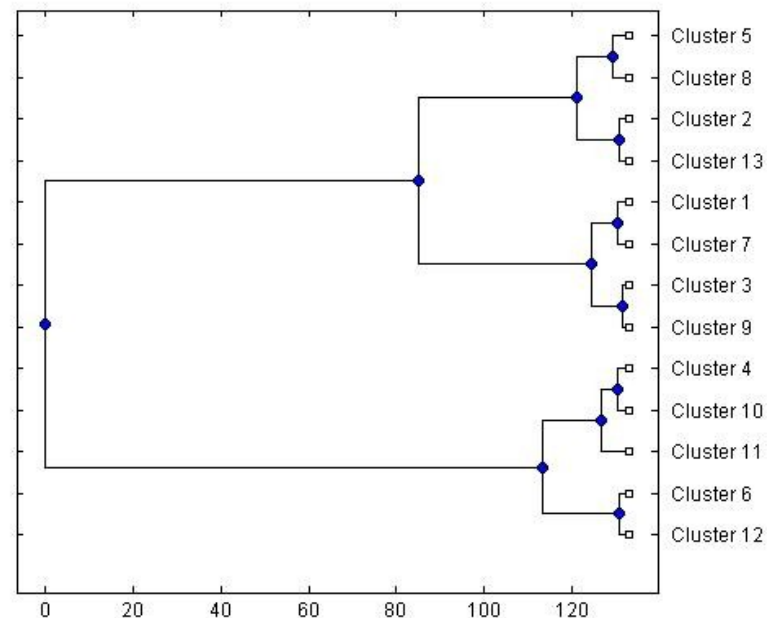
# Spatial models – tessellation, Voronoi polygons



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)

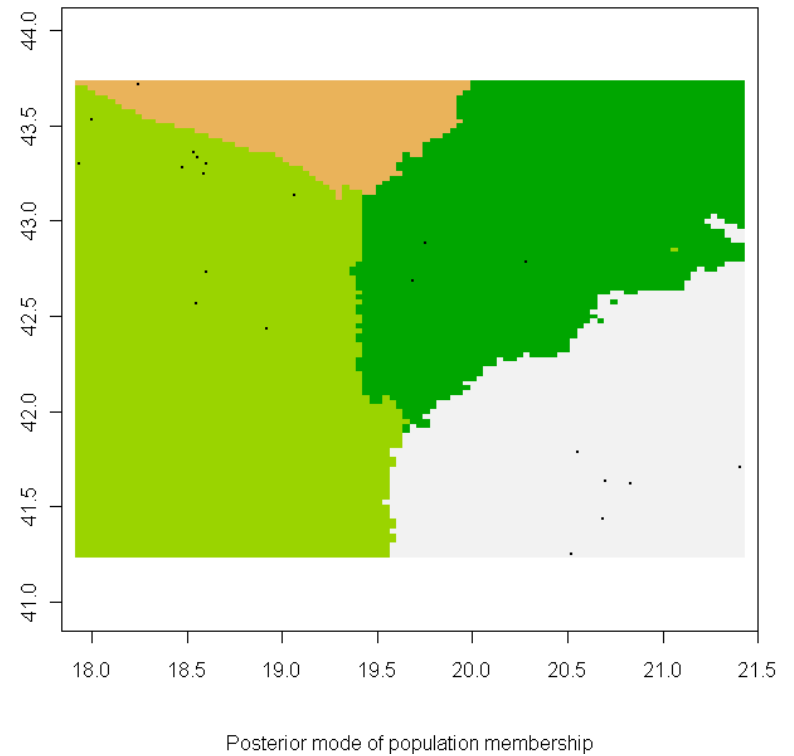
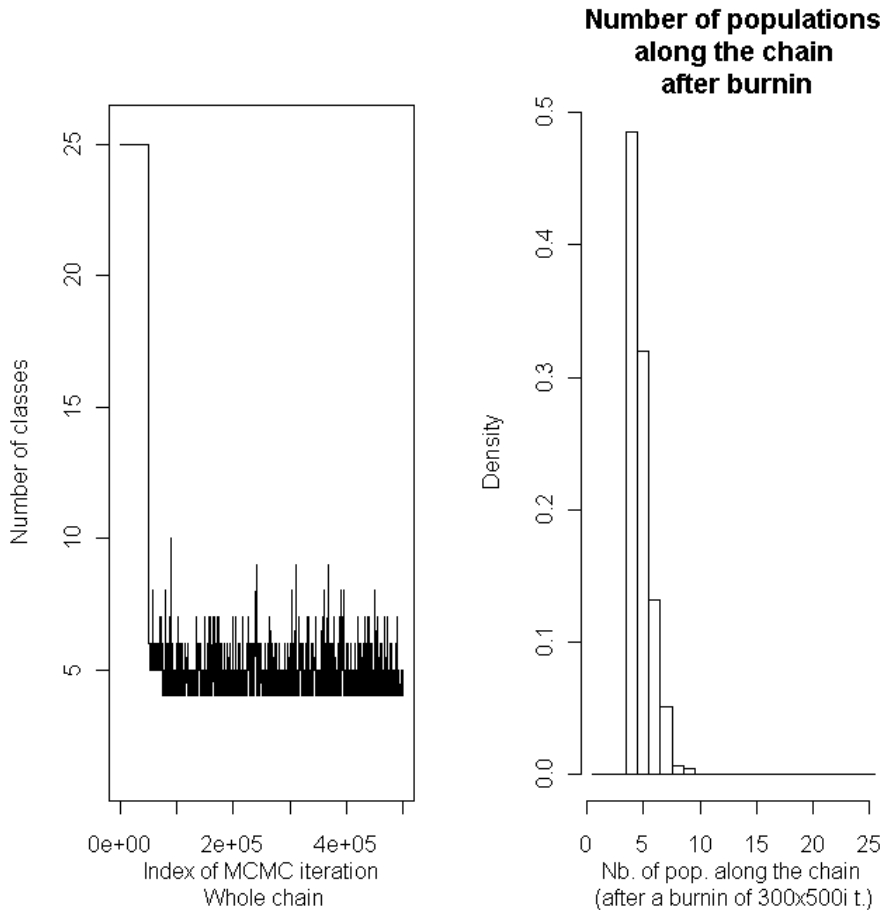


- Best Partition:
- Cluster 1: {C9, C13}
  - Cluster 2: {S6}
  - Cluster 3: {C8, C14}
  - Cluster 4: {C4}
  - Cluster 5: {C1, C2}
  - Cluster 6: {S1, S2, S3, S4}
  - Cluster 7: {C6}
  - Cluster 8: {C3, C15}
  - Cluster 9: {C5, C7}
  - Cluster 10: {C10}
  - Cluster 11: {C11, C12}
  - Cluster 12: {S5}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



Pr. program BAPS

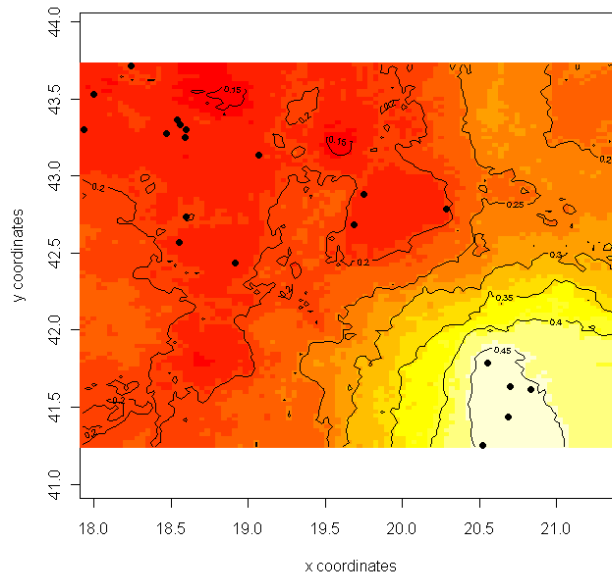
# Př.: Geneland



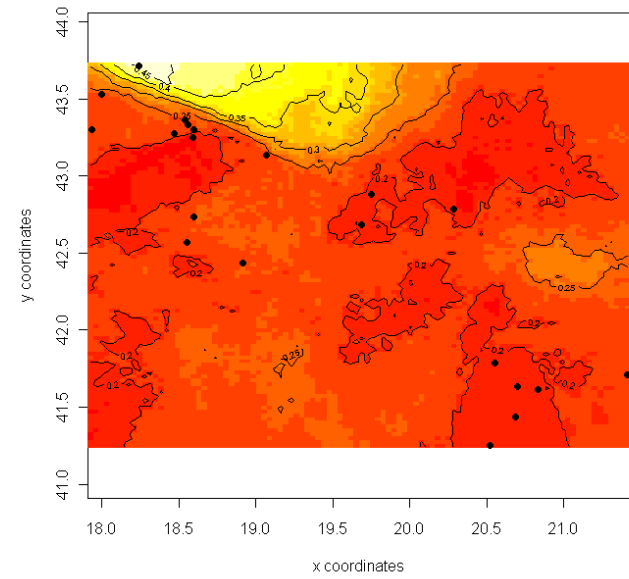
R platform

Posterior probability maps

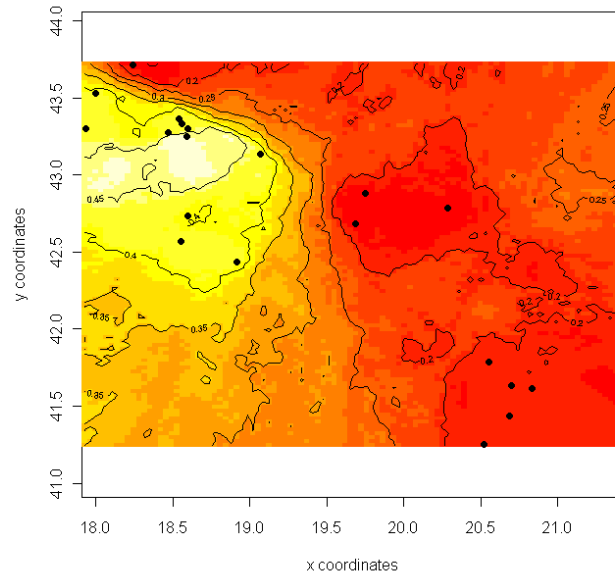
Map of posterior probability to belong to class 4



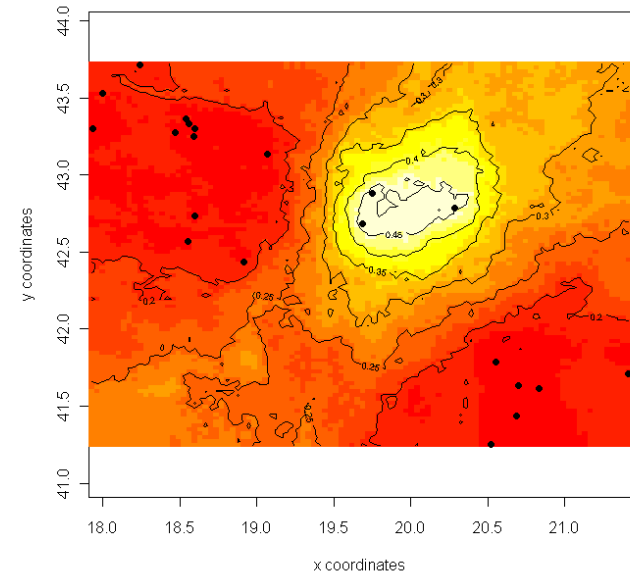
Map of posterior probability to belong to class 3



Map of posterior probability to belong to class 2



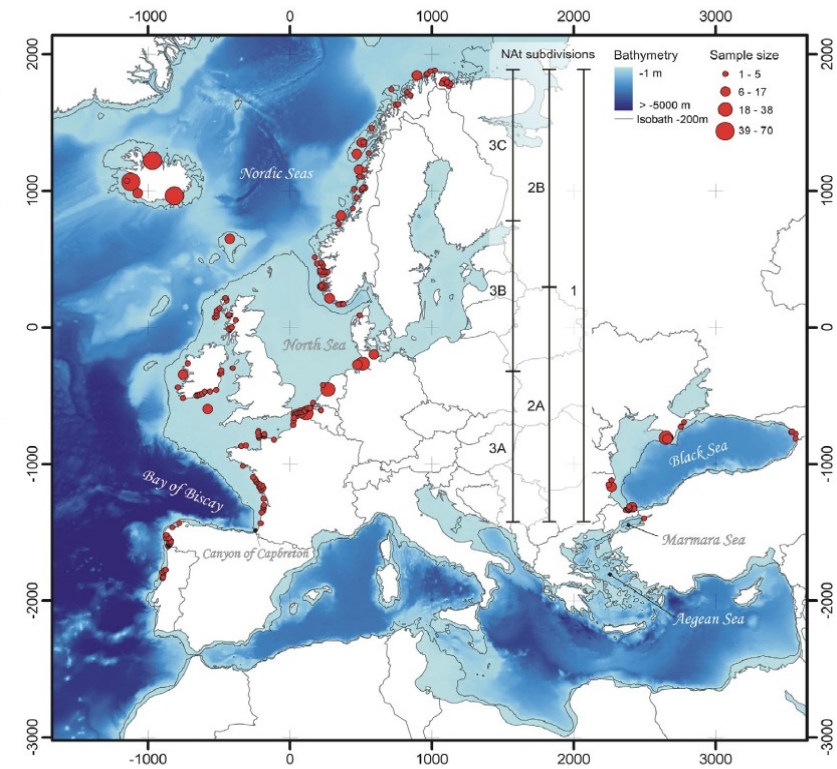
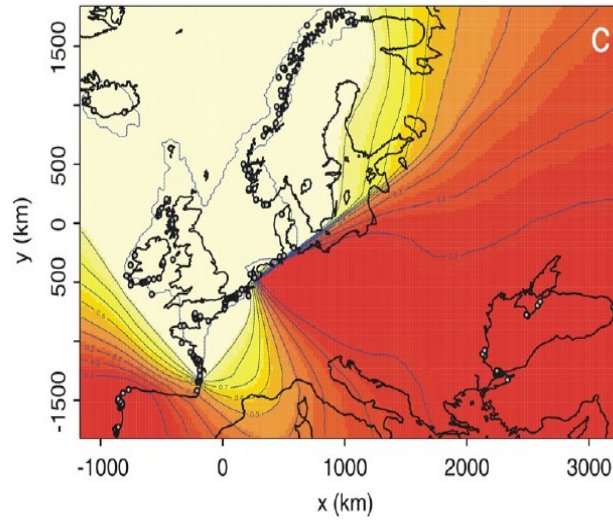
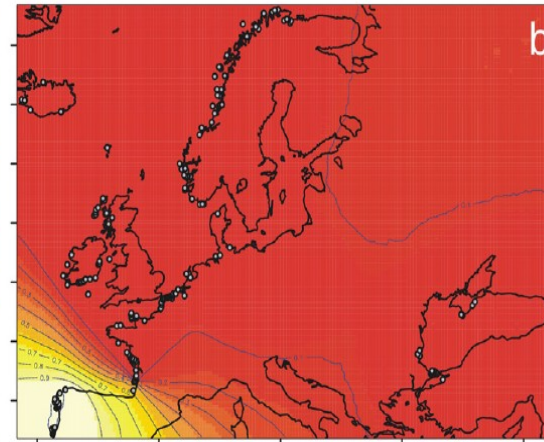
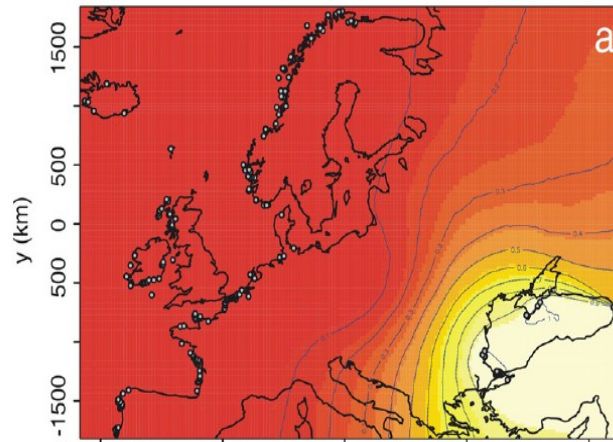
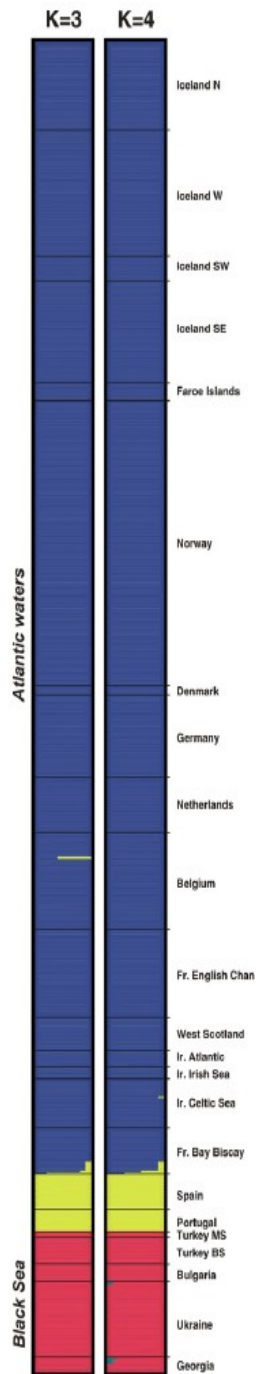
Map of posterior probability to belong to class 1





# Spatial population genetics

Fontaine et al. 2007



*Phocoena phocoena*