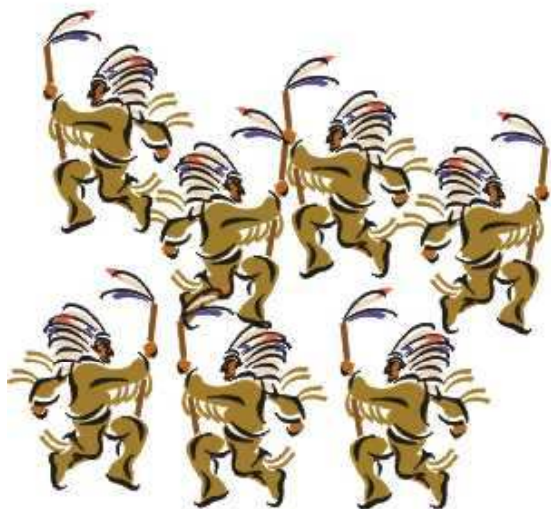


Analýza populační variability a struktury



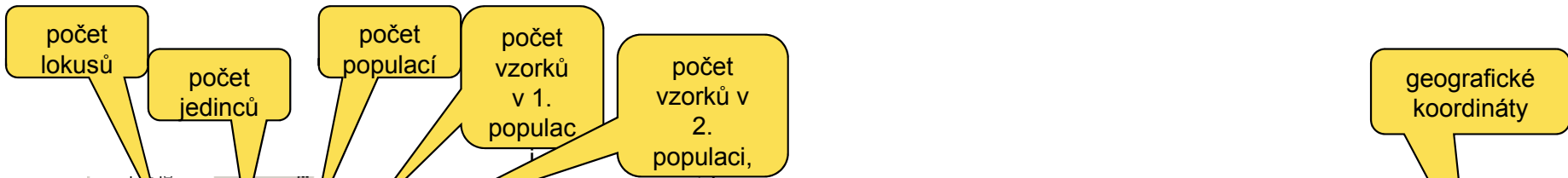
Hierarchická populační struktura

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	A
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	0	184	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	176	136	136	215	217	132	132	169	169	220	245	178	186	249	251	287	287	205	209	162	170			48.76500	17.00750
17	TYN14	TYN	162	176	136	136	215	217	132	132	169	169	220	245	178	186	249	251	287	287	205	209	162	170			48.76500	17.00750
18	TYN15	TYN	156	176	136	136	215	217	132	132	169	169	220	245	178	186	249	251	287	287	205	209	162	170			48.76500	17.00750
19	TYN16	TYN	168	176	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	128	134	166	173	0	0	184	184	245	245	287	287	205	213	162	166			48.83320	16.50610

genotypy, tj. velikosti fragmentů v populaci

pop1

pop2

Velké množství populačně-genetických programů

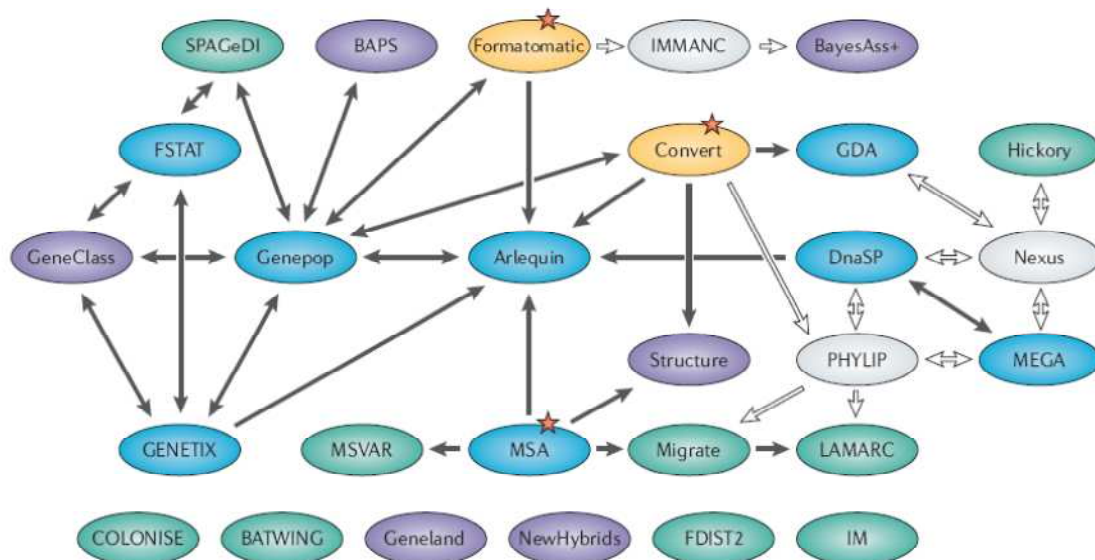
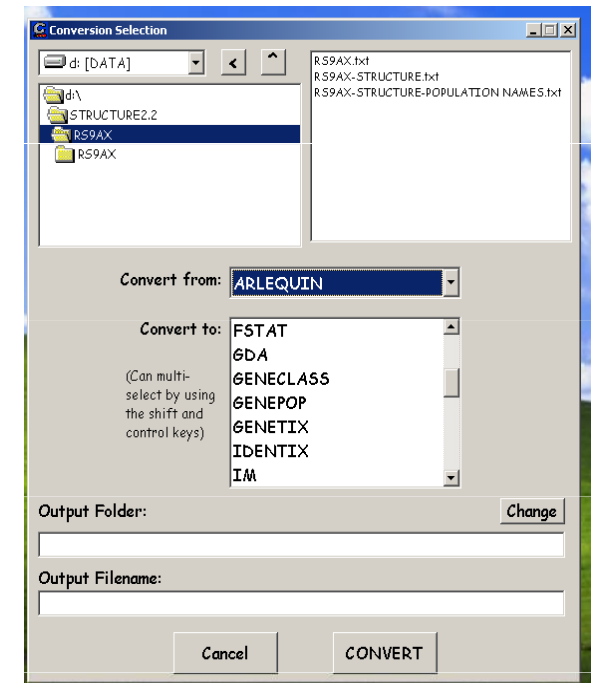
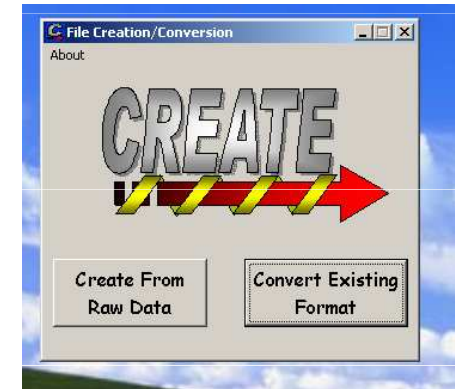


Figure 1 | **Flow chart of possible data exchange between different population genetics programs.** Although many programs have their own input-file specification, data files can still be exchanged between most programs (black arrows), avoiding tedious reformatting processes. The red stars are recommended 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	Arlequin, FSTAT, GDA, Genepop, GENETIX, MSA, SPAGeDi, Hickory		SPAGeDi		Arlequin, DnaSP, MEGA
Linkage disequilibrium	Arlequin, FSTAT, GDA, Genepop, GENETIX, Structure				
Analysis of population subdivision	Arlequin, FSTAT, GDA, Genepop, GENETIX, MSA, SPAGeDi, Hickory, Structure, BAPS, Geneland	Arlequin, FSTAT, GDA, Genepop, MSA, SPAGeDi	Hickory		Arlequin, DnaSP, MEGA
Detection of new immigrants: known populations	BayesAss+, GeneClass				
Detection of new immigrants: inferred populations	BAPS, NewHybrids, Structure, Geneland	BATWING, IM, LAMARC, MSVAR			
Demographic expansion or decline		BATWING, IM, LAMARC, Migrate, MSVAR		BATWING, LAMARC, Migrate	Arlequin, DnaSP, IM, LAMARC, Migrate
Population size	Migrate	BATWING, IM		BATWING, LAMARC, Migrate	IM, LAMARC, Migrate
Divergence time	Arlequin, FSTAT, GDA, Genepop, GENETIX	BATWING, IM, LAMARC, Migrate, MSVAR		BATWING, LAMARC, Migrate	DnaSP, IM, LAMARC, Migrate
Migration rates	Arlequin, FSTAT, Genepop, BayesAss+, COLONISE, Migrate	BATWING, IM, LAMARC, Migrate, MSVAR		BATWING, LAMARC, Migrate	DnaSP, IM, LAMARC, Migrate
Neutrality tests	Arlequin, FDIST2				Arlequin, DnaSP, MEGA
Spatially explicit analyses	SPAGeDi, Geneland, COLONISE				

*By multi-allelic markers, we mean loci for which no specific mutation model is assumed, or for which mutations can be neglected. In the latter case, computations are based on allele frequencies only. Otherwise, specific mutation models are assumed. For example, for STRs, the general packages to estimate descriptive statistics and linkage disequilibrium, and to detect new immigrants. AFLP, amplified fragment length polymorphism; STR, short tandem repeat.

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)

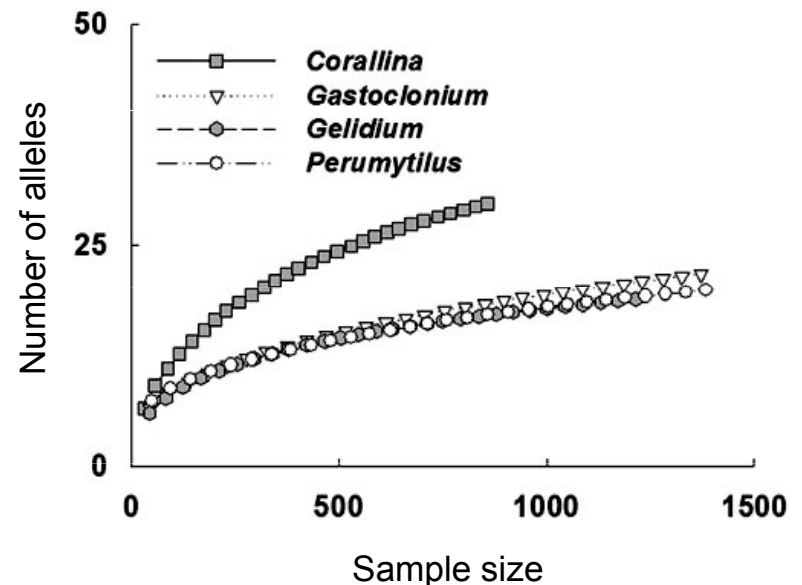
- počet alel na lokus

Alelická bohatost (allelic richness)

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

Pozorovaná heterozygotnost (observed heterozygosity)

- průměrná četnost heterozygotů v jednotlivých lokusech



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. χ^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 HWWE

- náhodné párování (panmixia)
 - zanedbatelný efekt mutací a migrací („closed populations“)
 - nekonečně velká populace
 - Mendelovská dědičnost použitých markerů
 - neutrální znaky – žádná selekce
-
- znaky nejsou ve vazbě – kontrola na „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)

Odchytky od HW rovnováhy

- **Test HWE** – nejlépe Genepop („exact probability tests“) – pokud jsou odchytky, 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)

Příklad – stanovení variability populace

Jedinec	Locus 1	Locus 2	Locus 3	Locus 4	Průměr
Ind 1	170/170	223/227	116/116	316/316	
Ind 2	170/172	223/225	112/112	316/316	
Ind 3	172/172	223/225	112/112	316/316	
Ind 4	170/172	223/227	112/112	316/316	
Počet alel	2	3	2	1	2
Ho	0,5	1,00	0	0	0,375
p	0,5	p = 0,5	0,75	1,00	
q	0,5	q = 0,25 r = 0,25	0,25	0	
He	0,5	0,625	0,375	0	0,375

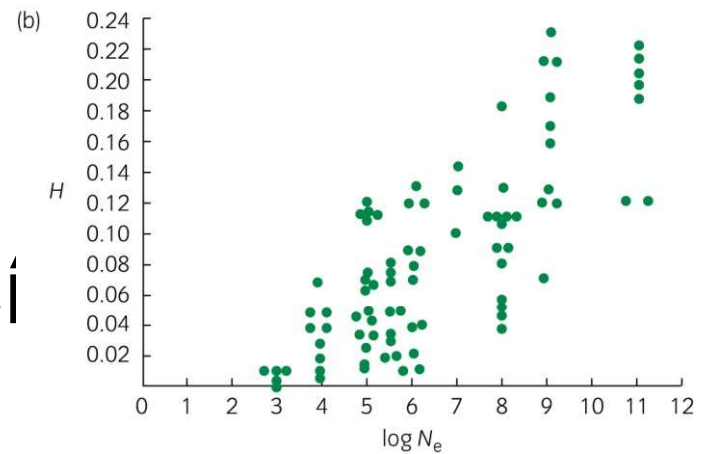
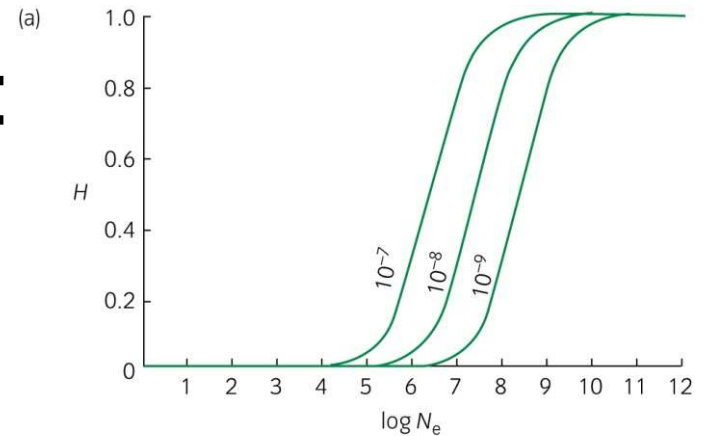
$$H_e = 1 - (p^2 + q^2)$$

Polymorfismus = 0,75

$$H_e = 1 - (p^2 + q^2 + r^2)$$

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

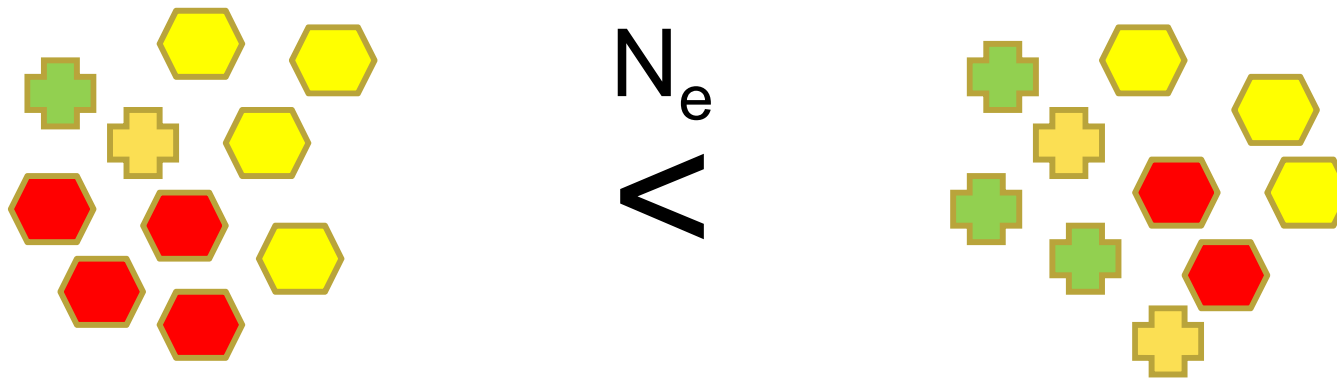


rypouš severní



Efektivní velikost populace (N_e)

- N_e = velikost ideální populace (náhodné páření, rovnoměrný poměr pohlaví), která ztrácí genetickou diverzitu stejnou rychlostí jako aktuální populace (vlivem náhodného driftu)
- N_e = velikost ideální panmiktické populace, kde všichni jedinci mají stejnou šanci stát se rodiči jakéhokoliv potomka v následující generaci
- ovlivněna genetickou a věkovou strukturou, poměrem pohlaví, intenzitou inbreedingu atd.

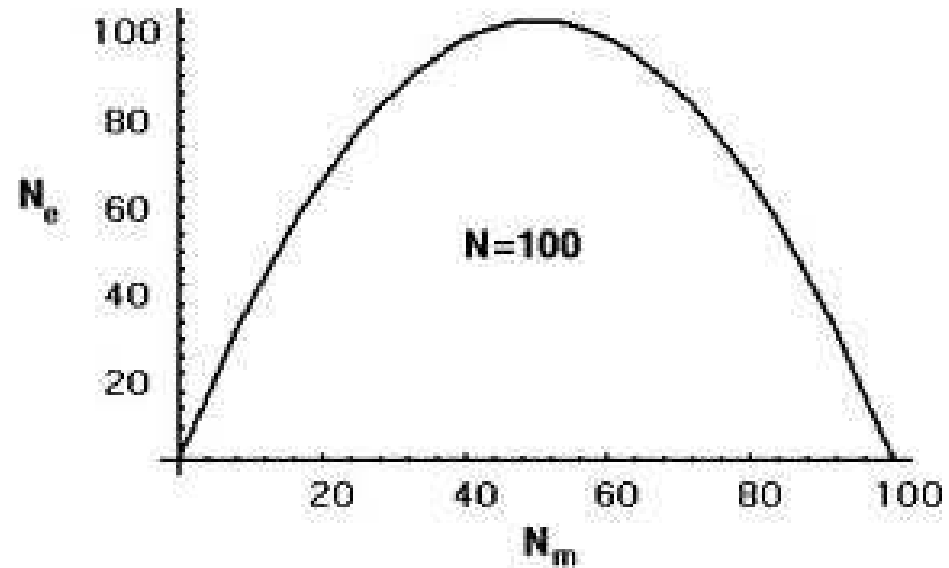


- vývoj genetické variability v malých populacích závisí na N_e více než na N

Efektivní velikost populace (N_e)

$$N_e = \frac{4N_f N_m}{N_f + N_m}$$

Effect of sex ratio



$$N_e = \frac{4N - 2}{V_k + 2}$$

Effect of variance in the number of progeny (V_k)
(for demographically stable population)

Efektivní velikost populace (N_e)

$$N_e = \frac{N_{ef}}{2}$$

mitochondrial DNA
(i.e. 4x lower than nuclear DNA)

$$N_e = \frac{N_{em}}{2}$$

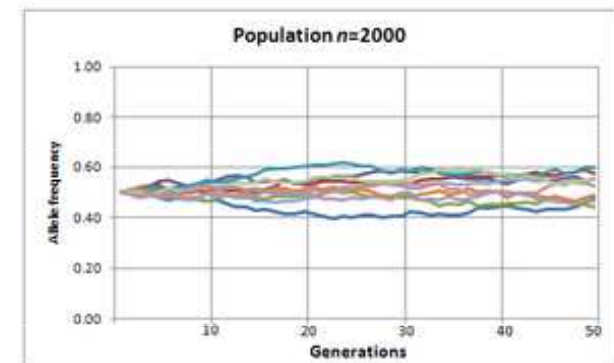
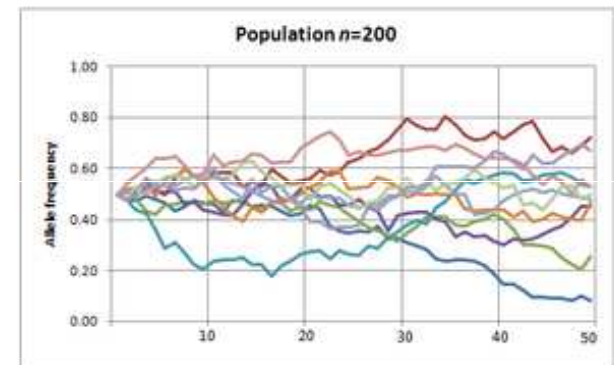
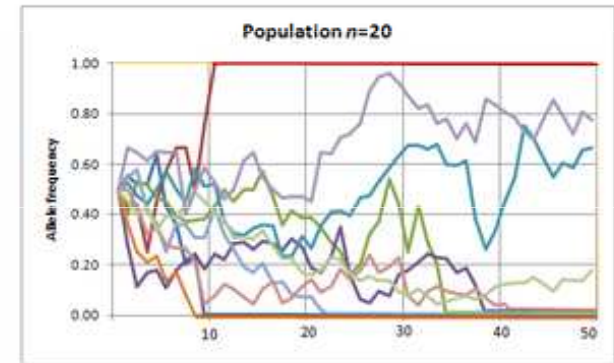
Y-chromosome

Genetický drift je 4x rychlejší na těchto znacích

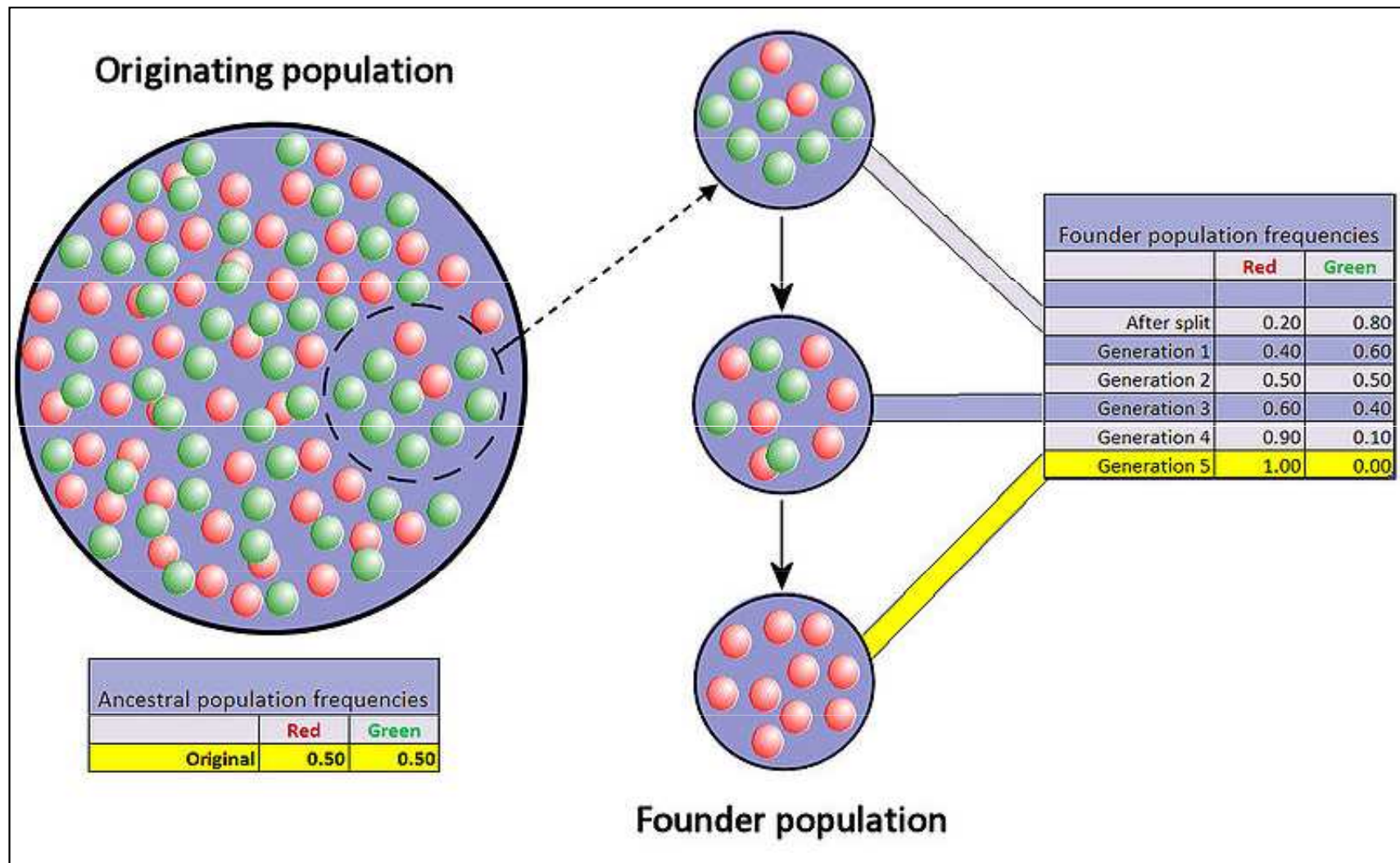
Genetický drift



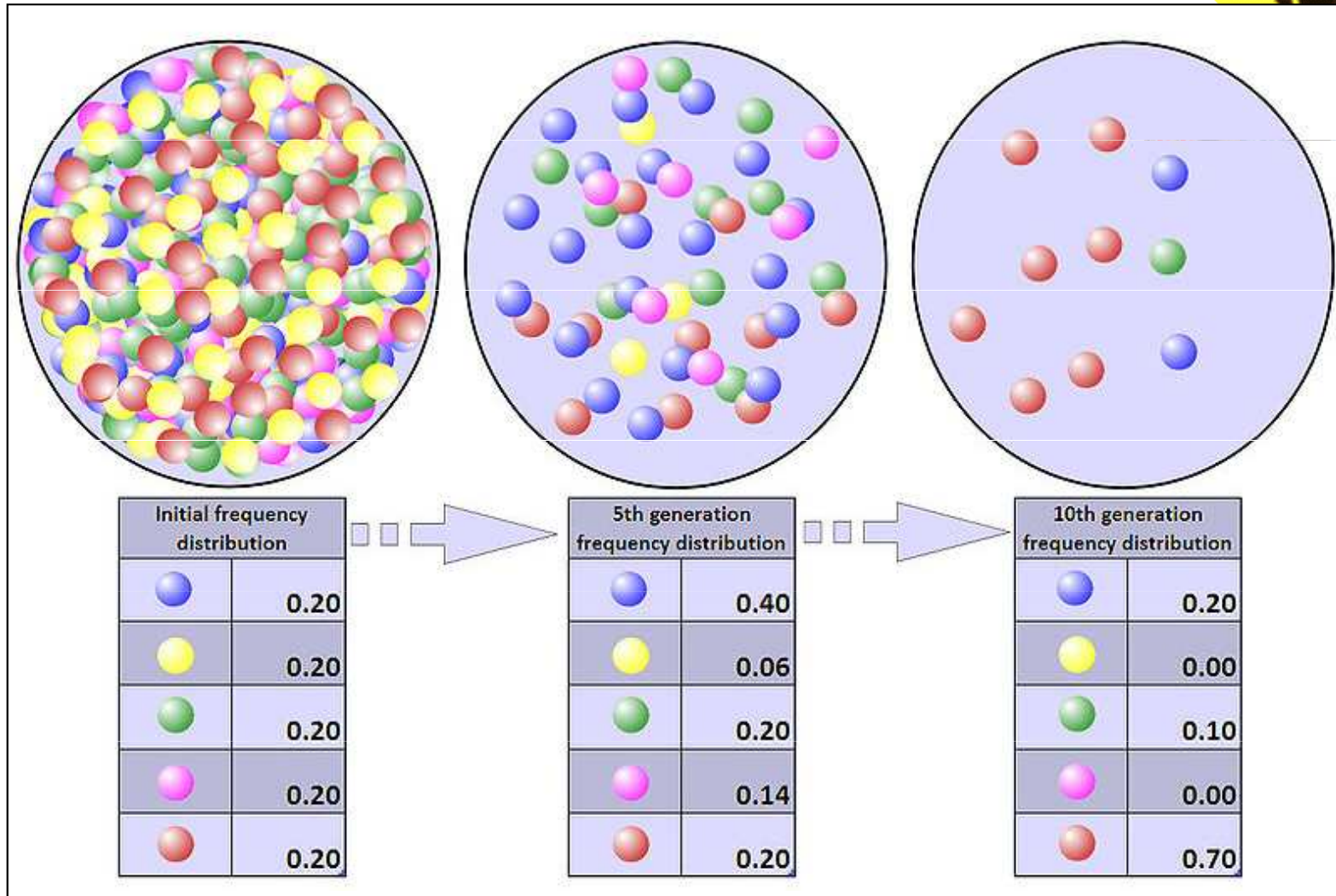
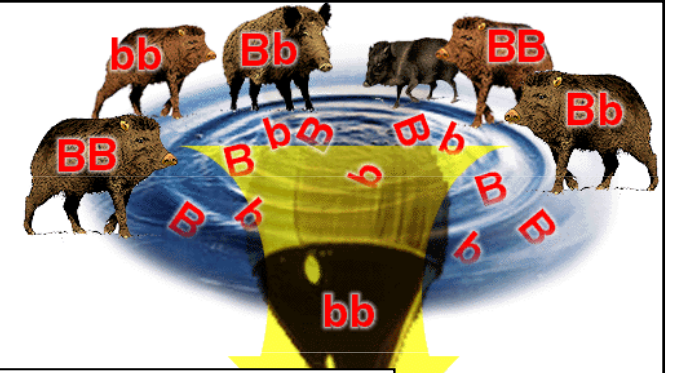
- Náhodné změny ve frekvencích alel
- Intenzita driftu závisí na velikosti populace (viz ochranářská genetika)
- Specifické případy – founder effect, bottleneck



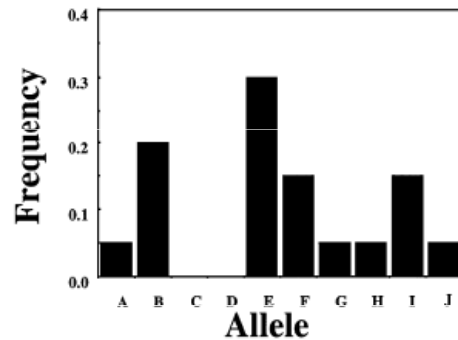
Founder effect („jev zakladatele“)



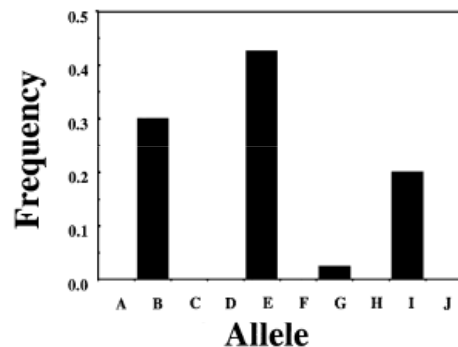
Bottleneck



Detekce bottlenecku (tj. intenzity genetického driftu)



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 rozsah alel nebo heterozygotnost

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

Program M Ratio: počet alel vs. rozsah alel

Program BOTTLENECK: počet alel vs. heterozygotnost (vzácné alely se ztrácejí rychleji než heterozygotnost)

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
- TPM je neivíce realistický (potvrzeno meta-analýzami velkých datasetů)

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

Otters in central Europe – program MSVAR

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

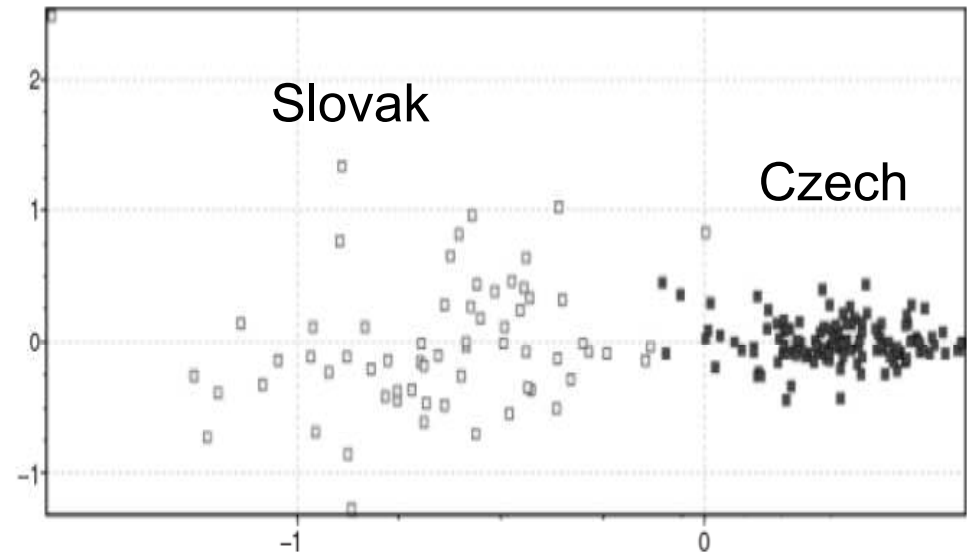


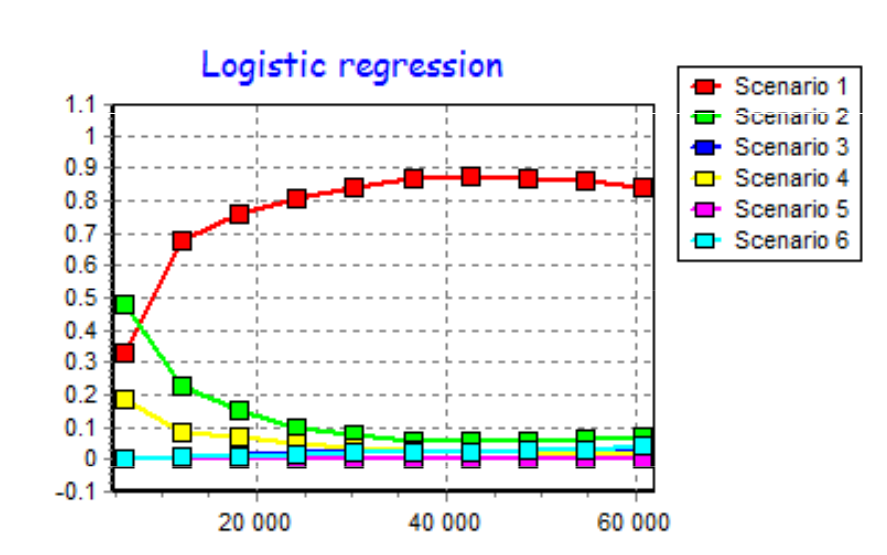
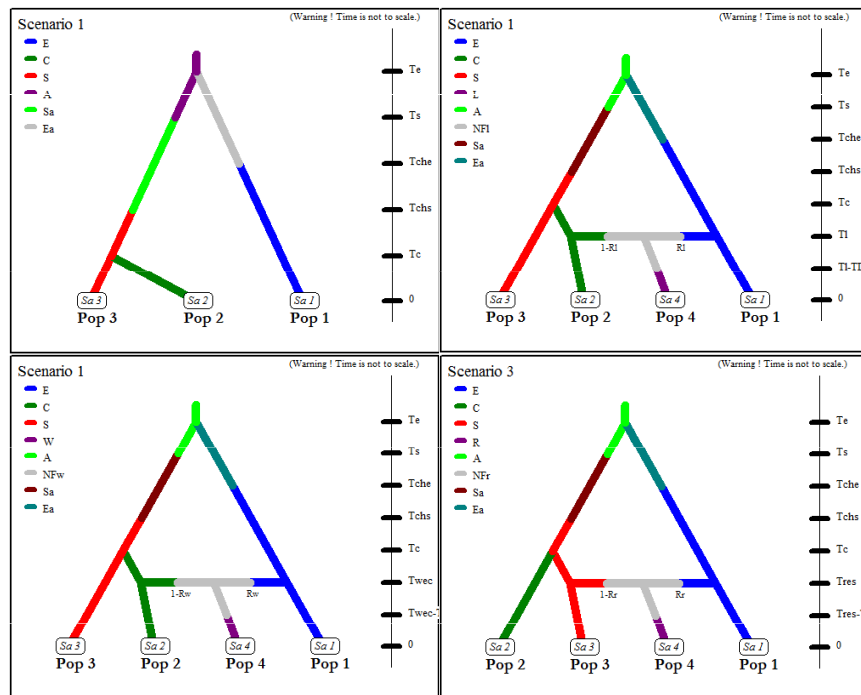
Table 3 Simulated data from five replicate runs of the MSVAR programme

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

n_0 , estimated mode (and limits) of current effective population size; n_1 , estimated mode (and limits) of past effective population size; μ , estimated mode (and limits) of mutation rate; X_a , 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.

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