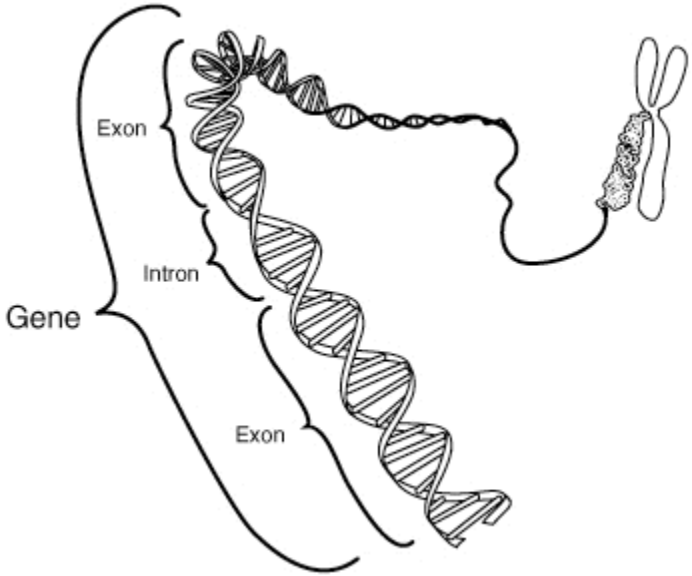
Úvod do populační genetiky

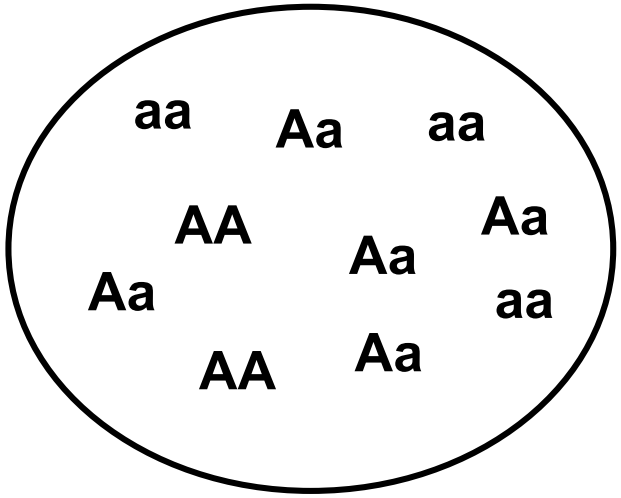
Termíny



- Gene
- Locus (loci)
- Allele – alternative forms of a gene
- Genome – single copy of all of the genes
- Diploid - 2 alely pro každý gen (výjimka Chr Y, cpDNA a mtDNA) – 2 genomy
- Soubor alel jednoho genu u jedince (na jednom lokusu) – genotyp (AA, Aa, aa)
- Stejné alely na lokusu – homozygote (AA aa)
- Různé alely – heterozygote (Aa)
- Haplotype – array of alleles of linked genes, DNA sequence

Heterozygosity

- Podíl heterozygotních lokusů
- Více lokusů jedince, jeden lokus u více jedinců
- Whole genome



- **Frekvence alel**
- A 9/20 = 0.45 (p)
- a 11/20 = 0.55 (q)

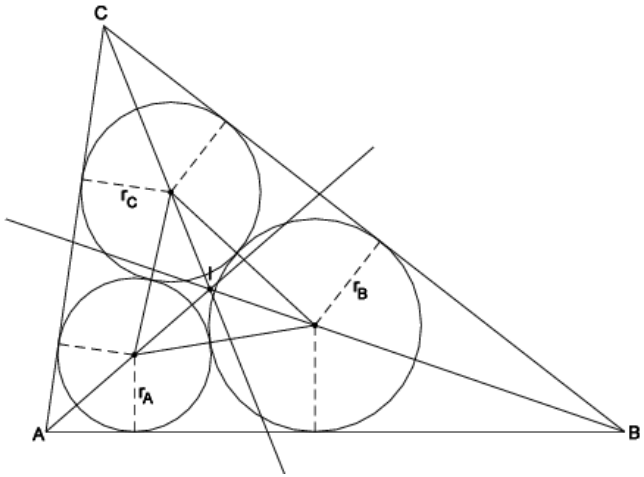
- **Frekvence genotypů**
- AA 2/10
- aa 3/10
- Aa 5/10
  → **observed** heterozygosity (H_O) = 0.5

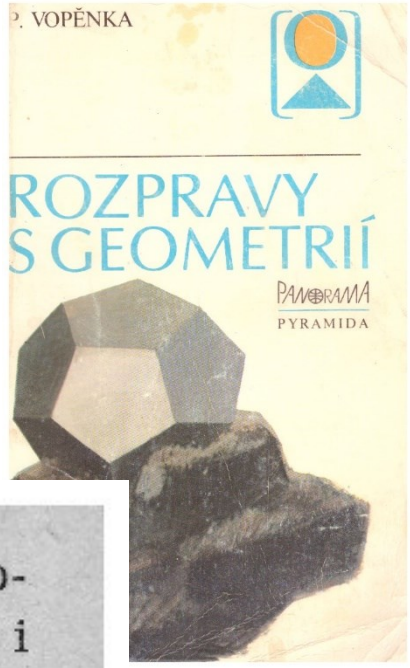
Ideální populace

Population - group of interbreeding individuals that exist together in time and space

Ideální populace = nulový model

Reálná populace neodpovídá ideální → něco se děje (selekce, demografické změny…)

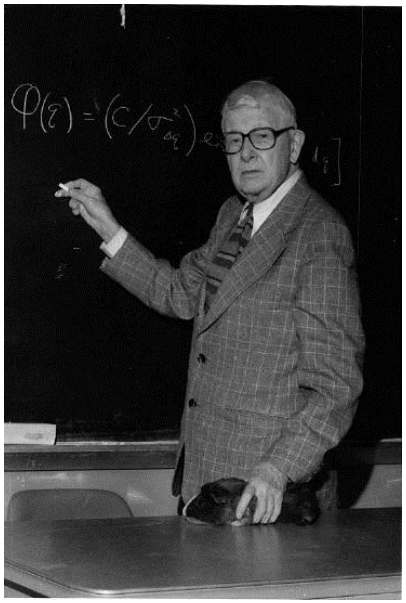




Geometr má před sebou list papíru pokreslený čarami rozmanitých tvarů, rovnými i křivými, vzájemně propletenými a protínajícími se v různých bodech. Jeho zrak spočinul na obrázku, jeho pohled však pronikl skrze obrázek, ven z reálného světa do světa geometrického. Tak například za rovnou čarou uviděl geometrickou úsečku, uviděl ji v její úplné čistotě a spolu s ní uviděl dokonalou přímost. Od okamžiku tohoto prohlédnutí je pro něj navždy úsečka úsečkou geometrickou, a ne čarou narýsovanou podle pravítka.

- Slavní zakladatelé moderní syntézy, třicátá léta
- Matematické modely spojující genetiku a evoluční teorii

Sewall Wright



Ronald Fisher

J. B. S. Haldane

Vlastnosti ideální populace Fischer Wright model

- **Random mating** – depends on the trait
  - **Mating = křížení**
  - Krevní skupiny versus pohlavní chromosomy
- **Nonoverlaping generations**
  - Často reálné
  - Tažní ptáci, rostliny

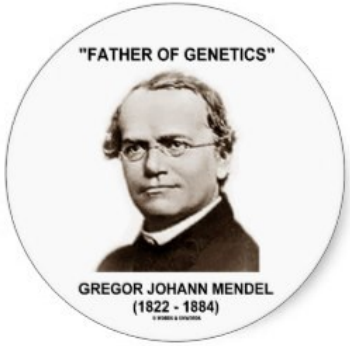
Předpoklady

- *Mendel's laws*
- *Random mating*
- *Nonoverlaping generations*
- Diploid (často i haploidní)
- Sexual reproduction
- Two alleles (di-allelic model)
- Allele frequencies identical in males and females
- *Někdy i infinite population size*

- ~~Migration~~, ~~population structure~~
- ~~Mutation~~
- ~~Selection~~

Mendel's first "law"

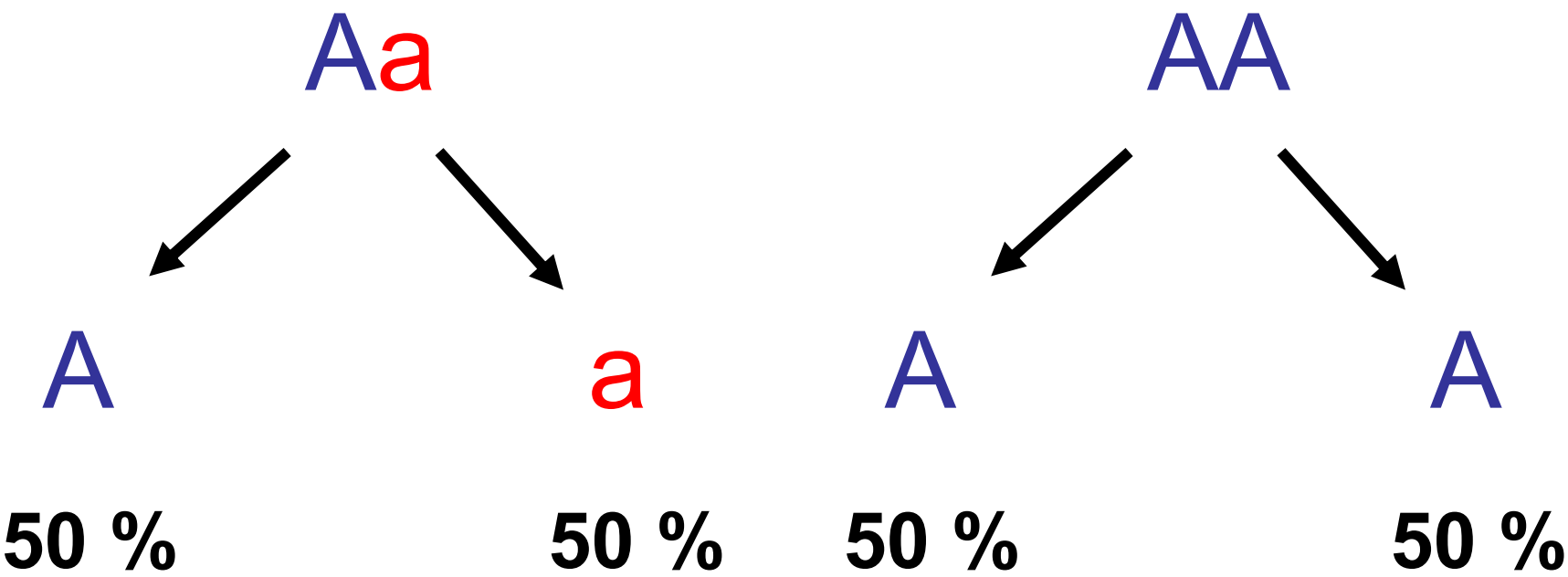
Predicts independent segregation of alleles at a single locus: two members of a gene pair (alleles) segregate separately into gametes so that halt of the gametes carry one allele and the other half carry the other allele.



Aa

A a

50 % **50 %**



Frekvence a 0.25 → a bude ve čtvrtině gamet

Hardy-Weinberg principle (equilibrium)

Male gametes

Allele	A	a
Frequency	p	q

	Allele	Frequency		
Female gametes	A	p	AA p^2	Aa pq
	a	q	aA qp	aa q^2

Summed frequencies in zygotes:

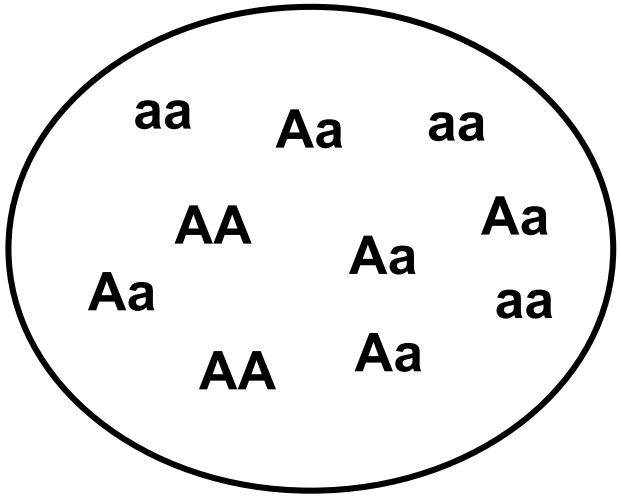
AA: $P' = p^2$
Aa: $Q' = pq + qp = 2pq$
aa: $R' = q^2$

- Genotypy rušeny a znovu sestavovány v každé generaci
- **Vztah mezi frekvencí alel a genotypů**
- AA: p^2 Aa: $2pq$ aa: q^2
- **Expected** heterozygosity $H_E = 2pq$
- **Během jedné generace**

Joint probability

Expected heterozygosity *2pq*

2 x 0.45 x 0.55 = 0.495

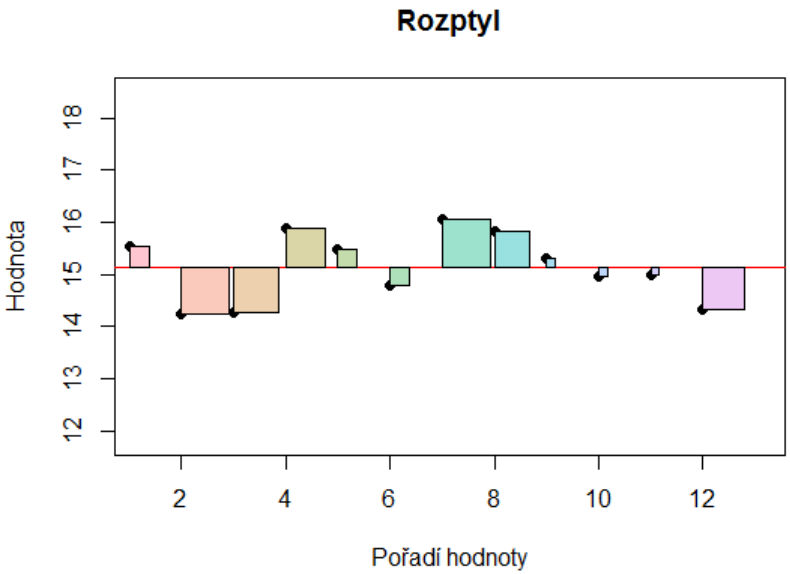


- **Frekvence alel**
- A 9/20 = 0.45 (*p*)
- a 11/20 = 0.55 (*q*)

- **Frekvence genotypů**
- AA 2/10
- aa 3/10
- Aa 5/10
  → **observed** heterozygosity (H_O) = 0.5

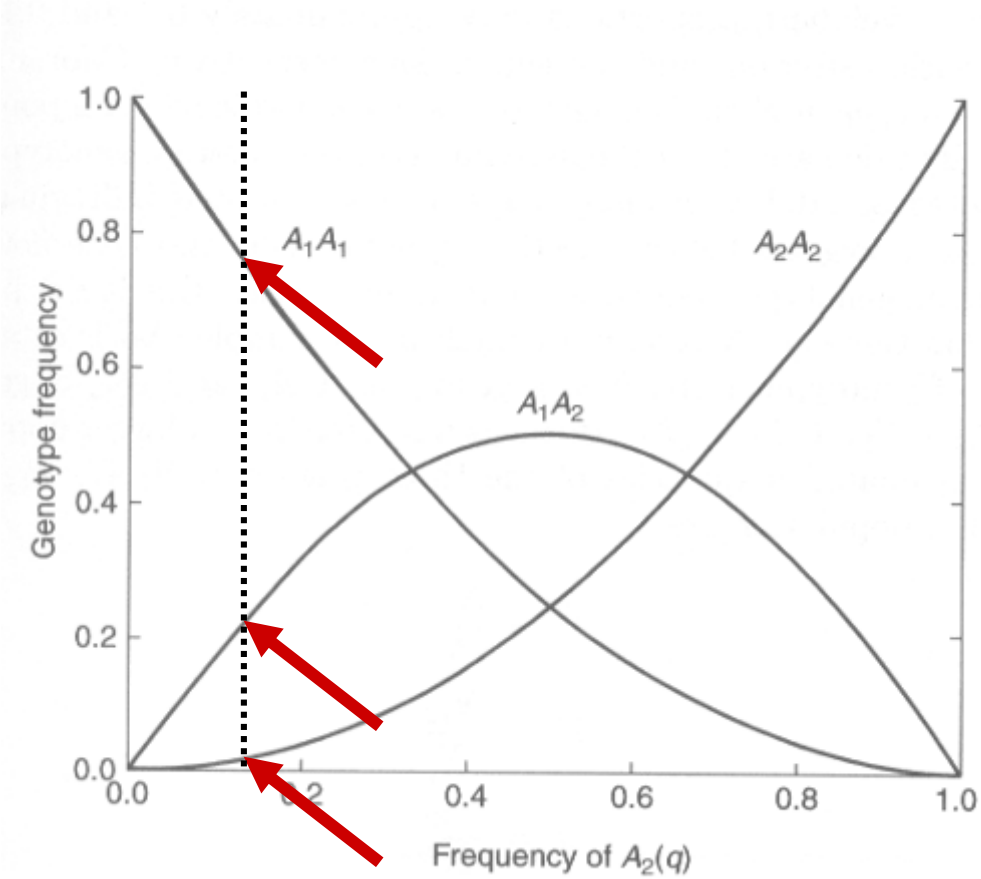
Zobecnění

- Více alel → **gene diversity** (= očekávaná heterozygotnost)
- Rozdílné frekvence alel mezi pohlavími

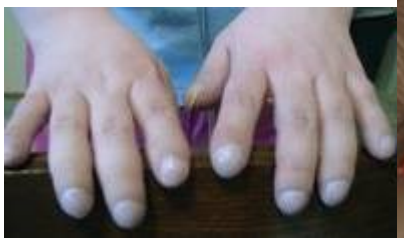


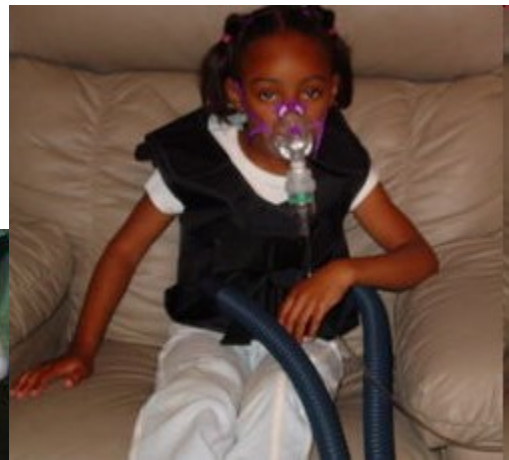
- Odpovídá rozptylu (variance)
  - Haplodiploidie, geny na pohlavních chromosomech

Rare recessive alleles



- Nízká frekvence alely
- → alela „schována“ v heterozygotech
- Cystic fibrosis
  - Homozygous recessive genotype 1 in 1700
  - Heterozygous carrier 1 in 24





Fixační index

popisuje odchylku od očekávání

$$F = \frac{H_e - H_o}{H_e}$$

THE GENETICAL STRUCTURE OF POPULATIONS

By SEWALL WRIGHT, *University of Chicago*
Galton Lecture at University College, London, 1950

the inbreeding coefficient F

The coefficient F measures the degree of approach toward fixation

H-W neplatí (neodpovídá počet heterozygotů)

- → **scoring errors**
  - Null alleles
  - Allele dropout
  - Duplications AA aa

- → **porušen některý z předpokladů**
  - **Decrease**
    - Selection against heterozygotes **underdominance**
    - Inbreeding
    - Positive-assortative mating
    - Population structure
  - **Increase**
    - Selection favoring heterozygotes **overdominance**
    - Outbreeding
    - Negative-assortative mating
    - Gene flow

HW Značně robustní, snese malá porušení předpokladů

HW = nulový model

Odchylky frekvencí → významně porušeny předpoklady

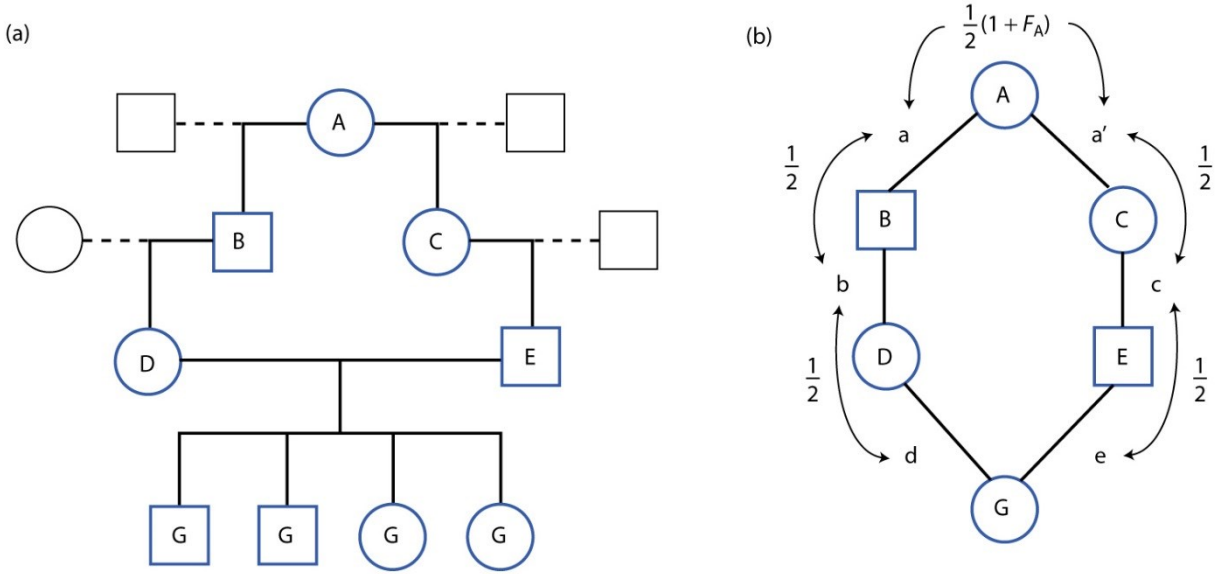
TABLE 2.4 The observed numbers and the numbers expected from Hardy–Weinberg proportions for the *MN* blood group locus in an English sample of 1000 blood donors (from Cleghorn, 1960).

Phenotype	*Genotype*	*Observed number*
M	*MM*	298
MN	*MN*	489
N	*NN*	213
Total		1000

Locus	n^a	k^b	Allele size range	Het (exp)[c]	Het (obs)[d]
LOX2	98	22	166–282	0.905	0.867
LOX7	93	70	130–357	0.983	0.968
LOX8	98	73	200–439	0.980	0.949

Inbreeding

- Inbreeding
  → větší šance, že se potkají stejné alely na lokusu



- Identical by state

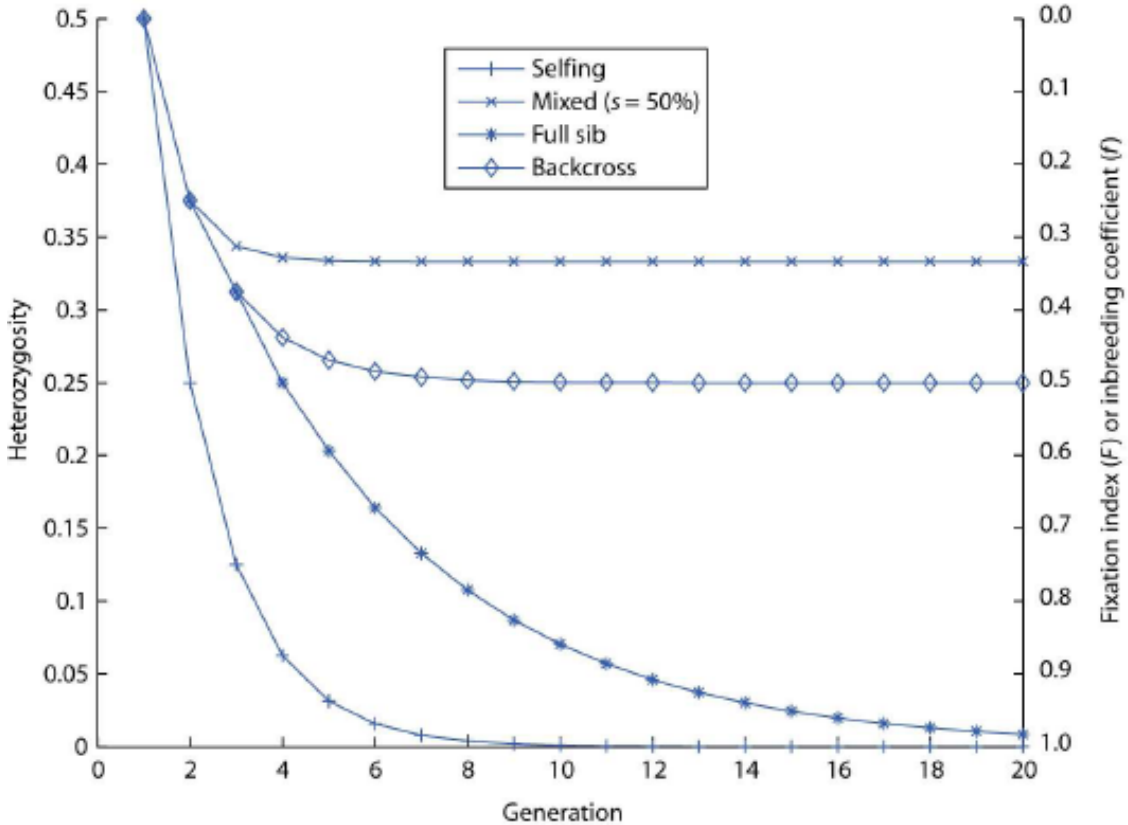
- Identical by descent

- Inbreeding coefficient *f*

  The probability that two alleles in a homozygous genotype are identical by descent.

- Pokud není heterozygotnost ovlivněna i něčím jiným: $f = F = (H_E - H_O)/H_E$

Inbreeding

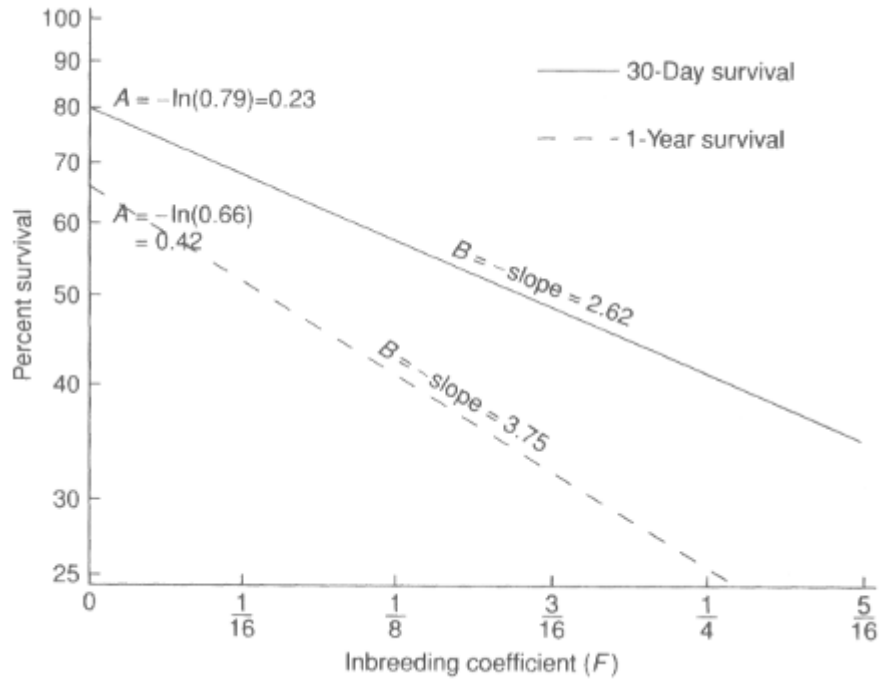


Inbreeding coefficient *f*

The probability that two alleles in a homozygous genotype are identical by descent.

Pokud není heterozygotnost ovlivněna i něčím jiným: $f = F = (H_E - H_O)/H_E$

Inbreeding



- **Inbreeding depression**

- **Dominance hypothesis**

  Expression of recessive alleles with deleterious effects

- **Ovedominance hypothesis**

  Heterozygotes have higher fitness
  Inbreeding → decrease in the frequency of heterozygotes

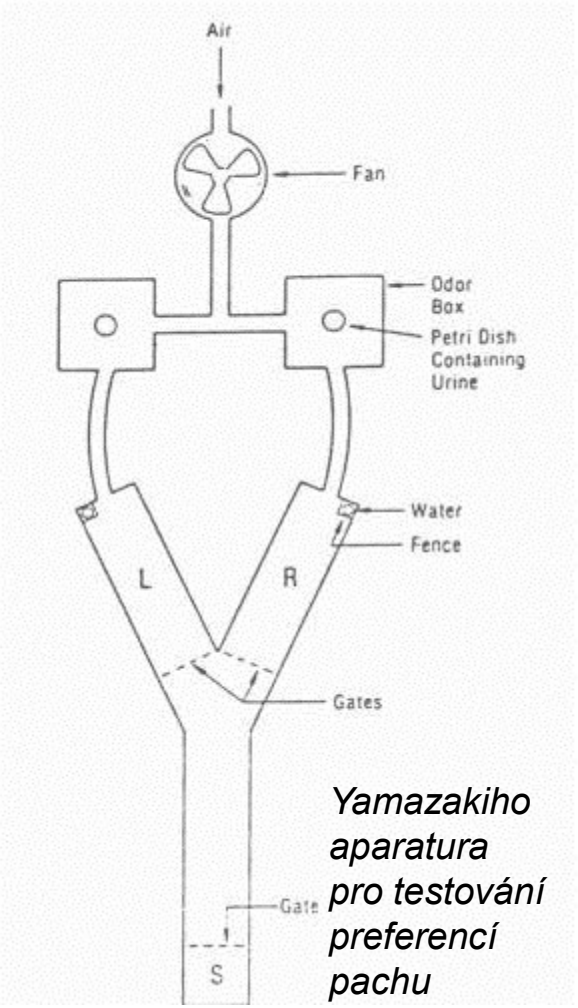
Gazella spekei

(Negative-)assortative mating

- Klasický příklad MHC u myší (možná i lidí)

MHC
a volba partnera

- Výhodná by měla být volba partnera s odlišným MHC (dissortative or negative-assortative preference)

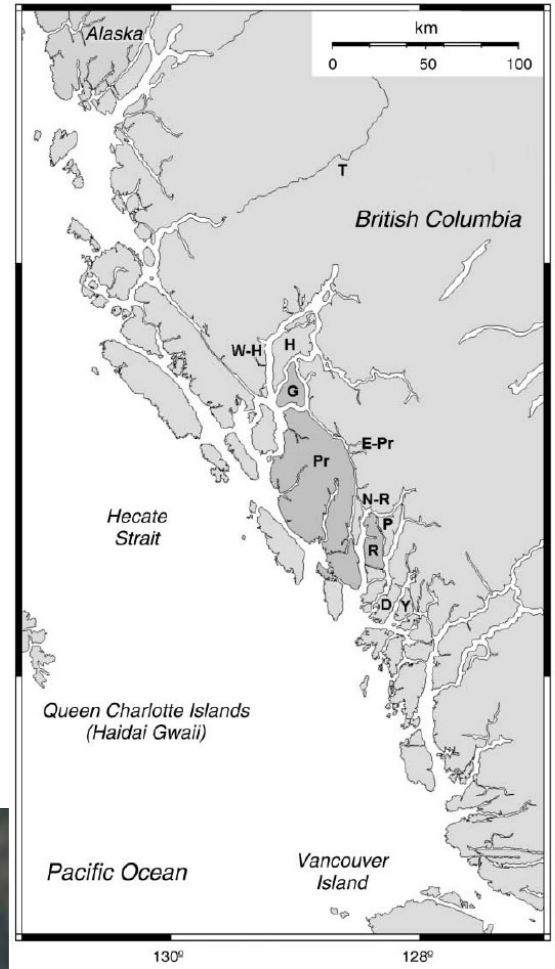


Yamazakiho aparatura pro testování preferencí pachu

Coat color genotypes, frequency of the recessive white allele, and expected number of heterozygotes, by locality.

	Mc1r genotype*			Frequency of G	Expected Number of heterozygotes
	AA	AG	GG		
Island localities					
Hawkesbury	24	1	0	0.02	0.98
Gribbell	7	6	10	0.56	11.33
Princess Royal	26	17	9	0.33	22.99
Roderick	9	1	2	0.21	3.98
Pooley	8	2	0	0.10	1.80
Yeo	9	1	0	0.05	0.95
Mainland localities					
West of Hawkesbury	5	1	0	0.08	0.88
East of Princess Royal	25	0	0	0.00	0.00
North of Roderick	10	1	1	0.13	2.71
Don Peninsula	22	2	0	0.04	1.84
Terrace	19	2	0	0.05	2.00

*GG is white, AA & AG are black.



Inbreeding versus assortative mating

- Ovlivnění konkrétního lokusu či nenáhodné skupiny lokusů
- Versus ovlivnění mnoha genů

Dispersal limitation → population structure

~~Infinite~~ population size, ~~random~~ mating

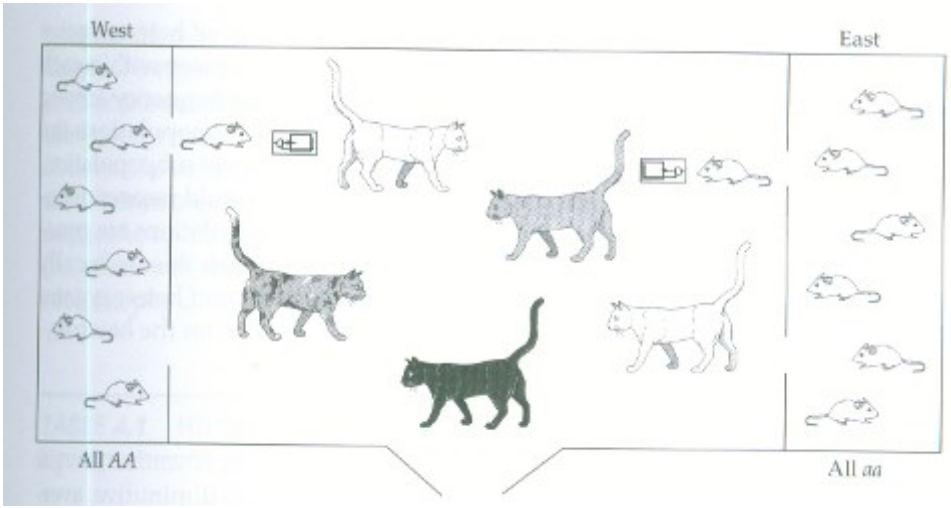
Demes – local interbreeding units

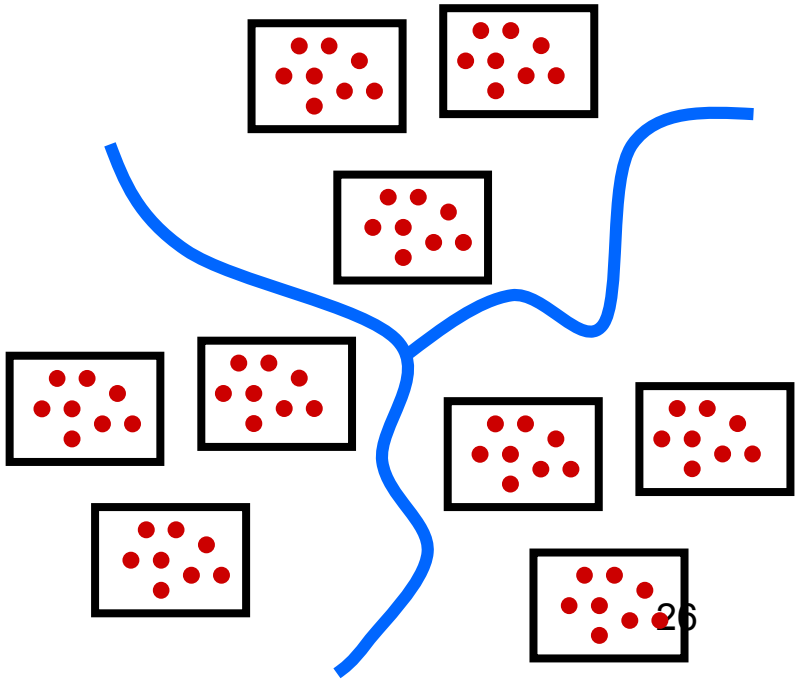
(Subpopulations, local populations)

Table 6.1. Isolation by Distance in Ina Valley, Japan, as Measured by Location of Spouse's Birthplace for 2022 Marriages

Spouse's birthplace	Percentage of marriages
Within Buraku (hamlet)	49.6
Within village but outside Buraku	19.5
Neighboring villages	19.1
Within Gun (county)	6.4
Within Prefecture (state) but outside Gun	2.9
Outside Prefecture	2.5

Source: Sekiguchi and Sekiguchi (1951).



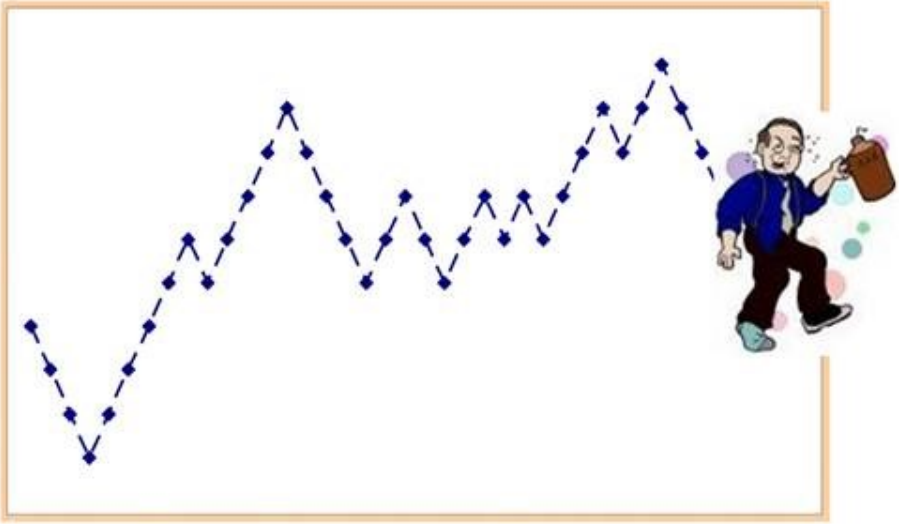


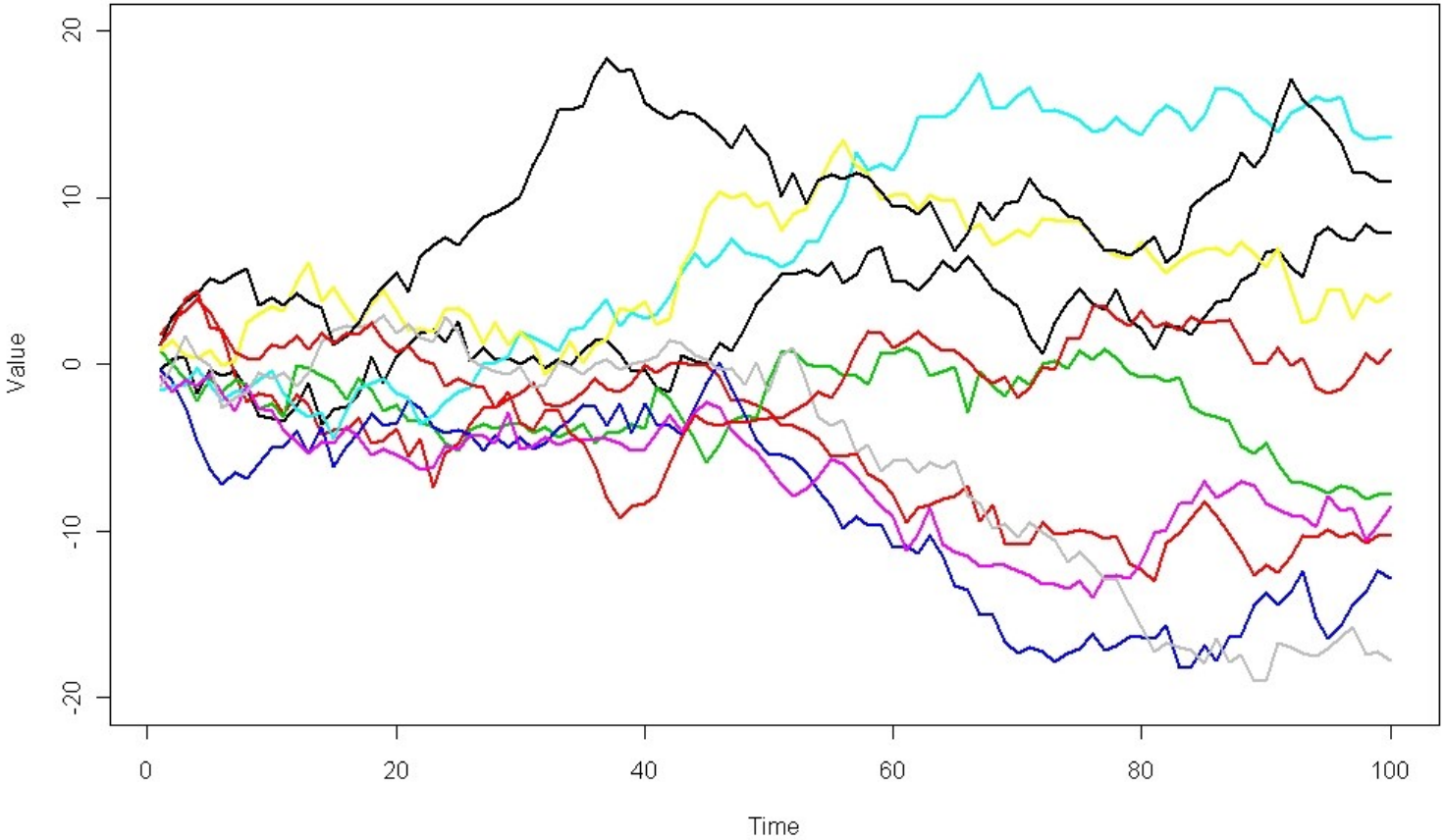
Změny frekvence alel - drift

- **Náhodný proces**
  - Omezená velikost populace
  - Některé alely se „vytáhnou“ častěji
  - Binomické rozdělení

- Jsou však i jiné příčiny změn frekvencí!
  - Mutace
  - Migrace
  - Selekce

Random walk

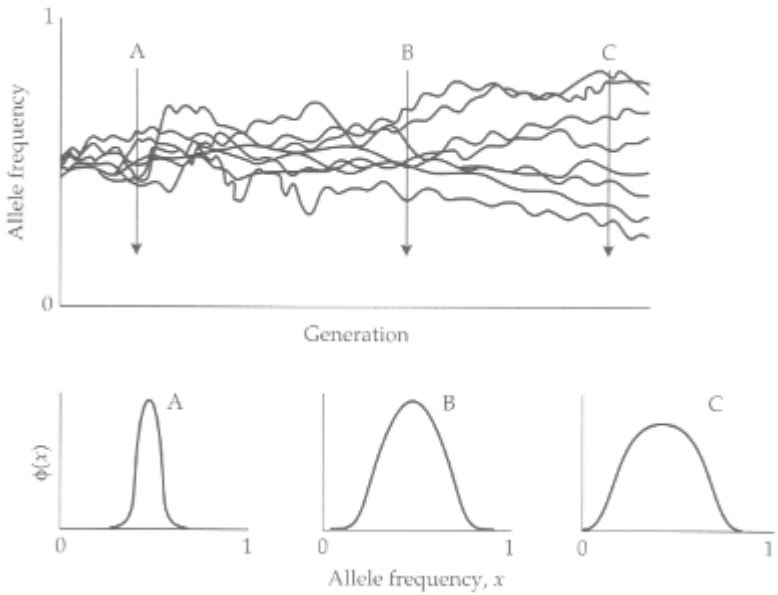


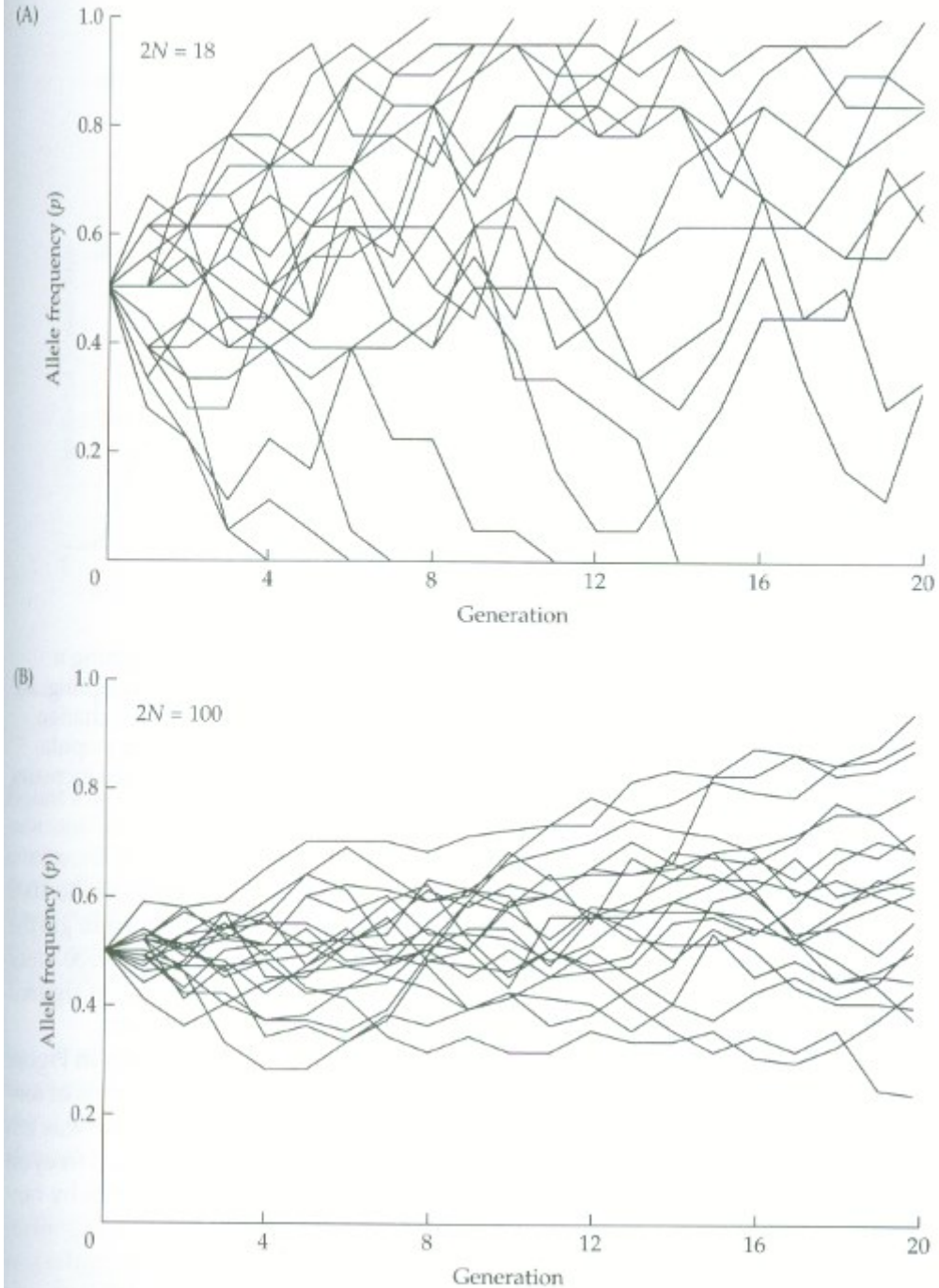


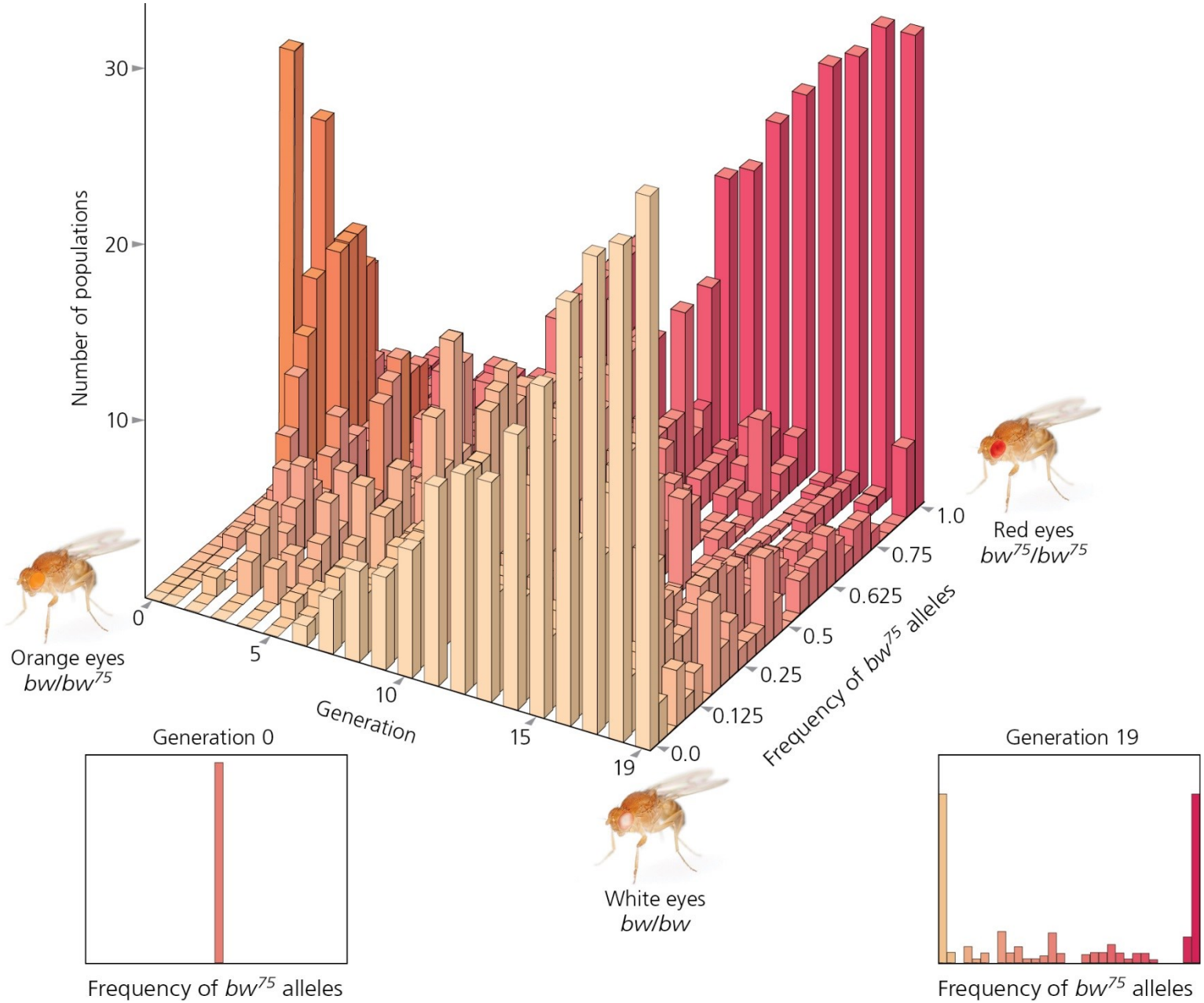
Simulace

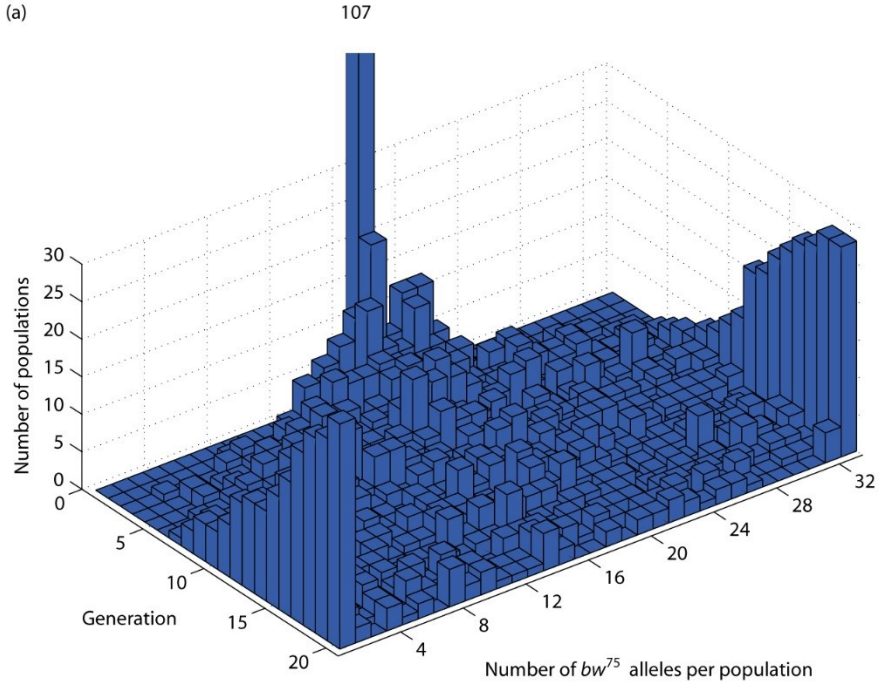
- Např PopG

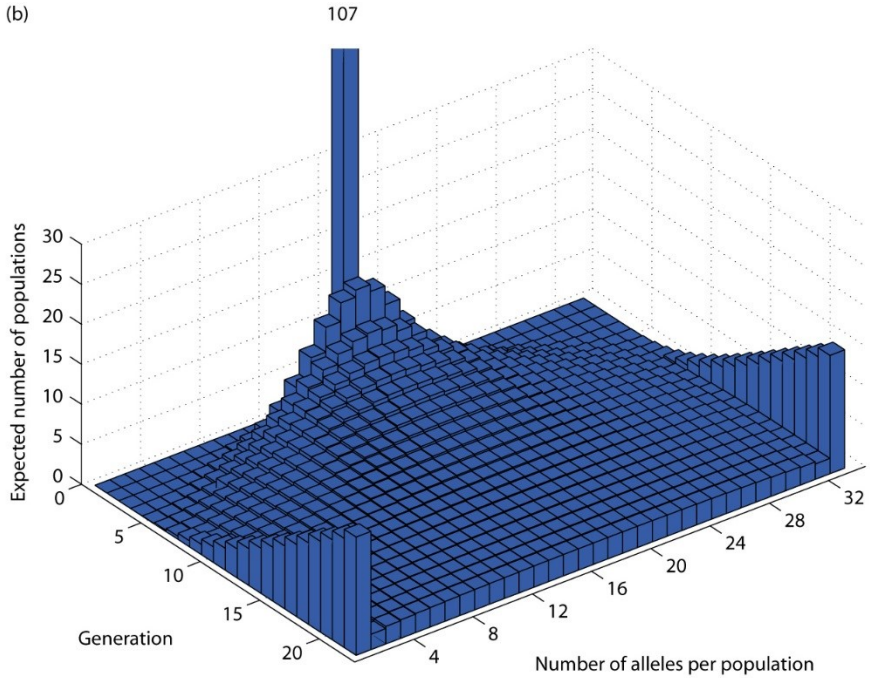
Random Genetic Drift






Number of populations
30
20
10
0
5
10
15
19
Generation
0.0
0.125
0.25
0.5
0.625
0.75
1.0
Frequency of bw75 alleles
Orange eyes
bw/bw75
White eyes
bw/bw
Red eyes
bw75/bw75
Generation 0
Frequency of bw75 alleles
Generation 19
Frequency of bw75 alleles

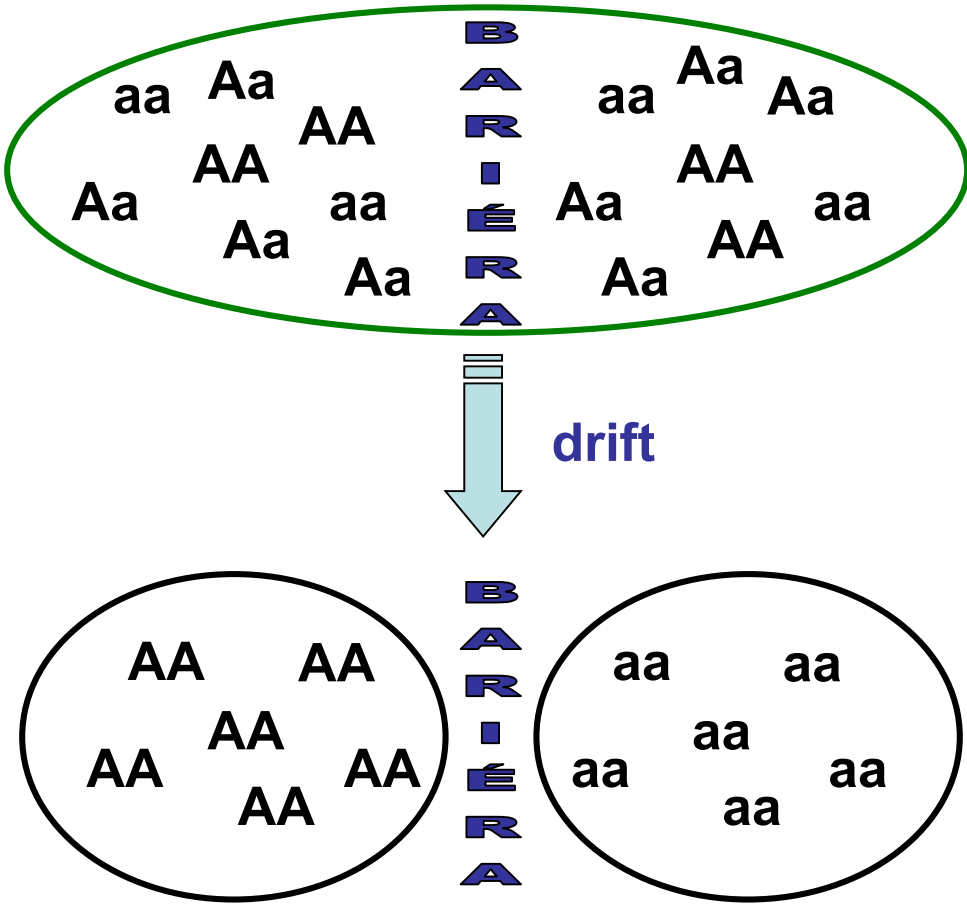

(a)
107
Number of populations
30
25
20
15
10
5
0
0
5
10
15
20
Generation
4
8
12
16
20
24
28
32
Number of bw^{75} alleles per population


(b)
107
Expected number of populations
30
25
20
15
10
5
0
0
5
10
15
20
Generation
4
8
12
16
20
24
28
32
Number of alleles per population

Drift

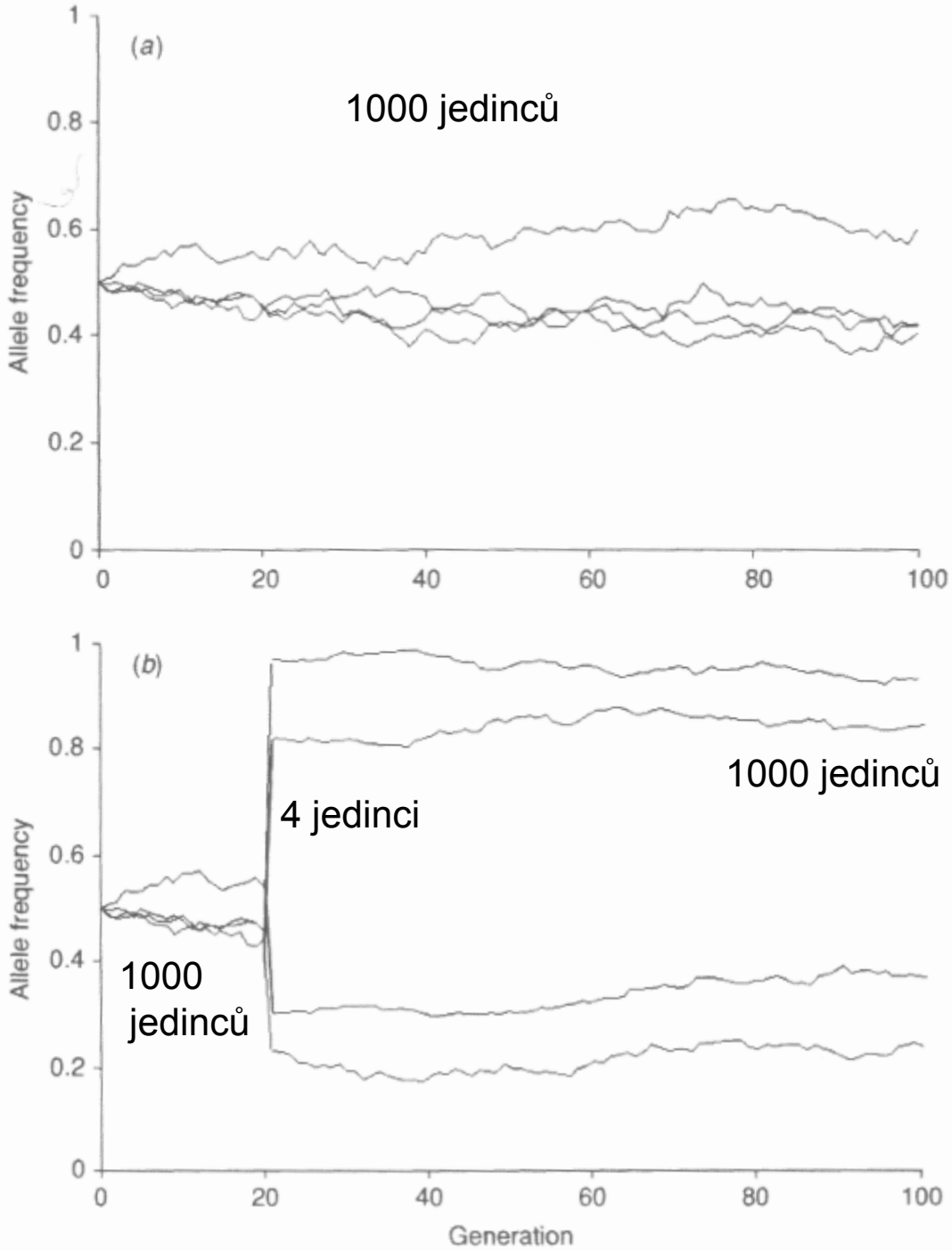
- Náhodné změny frekvence alely
- Stejná pravděpodobnost snížení i zvýšení frekvence
- Není žádná tendence se vracet k výchozím hodnotám
- Větší změny v malých populacích
- S časem rostou rozdíly mezi izolovanými populacemi
- S časem roste pravděpodobnost fixace (nebo vymizení) alely
- Vymizení nebo fixace dány počáteční frekvencí alely

- **Drift**
  → diferenciace subpopulací díky změně frekvence alel (až fixaci alternativních alel)



Drift

- Bottleneck
- Founder effect



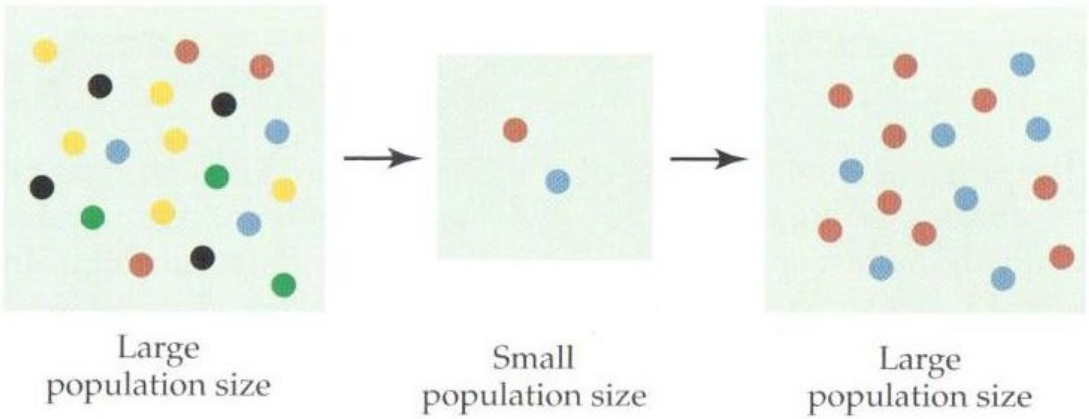


Figure 2.3 An illustration of the effect of a bottleneck in population size on genetic variability. The initial population has a high degree of variability, illustrated by the variety of colors of the balls in the box. The population goes through a bottleneck—a temporary decrease in size—and after the bottleneck, there are many fewer different alleles present.

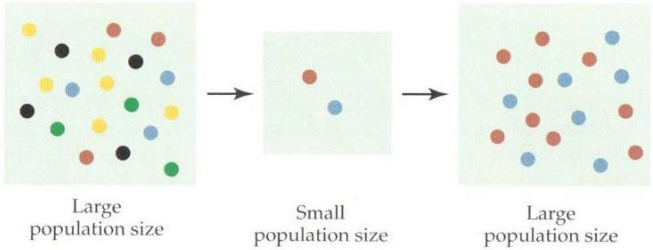
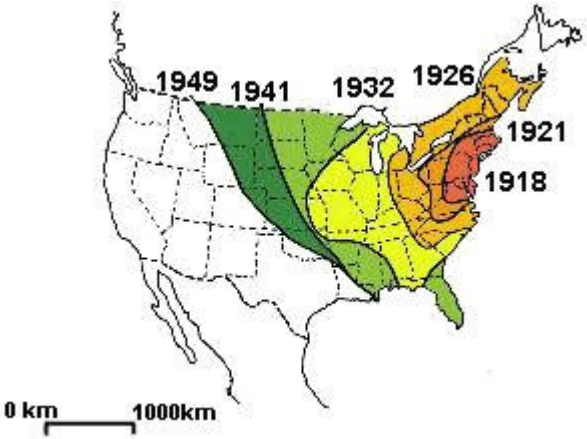


Figure 2.3 An illustration of the effect of a bottleneck in population size on genetic variability. The initial population has a high degree of variability, illustrated by the variety of colors of the balls in the box. The population goes through a bottleneck—a temporary decrease in size—and after the bottleneck, there are many fewer different alleles present.

Table 1 Allelic frequencies of uncommon alleles estimated from electrophoretic data for introduced European starlings (*Sturnus vulgaris*) in North America and New Zealand, and from their native range in Great Britain (data from New Zealand and Great Britain pooled over localities, from Ross, 1983; additional data from Great Britain from Evans, 1980)

Locus	Allele	Virginia	Vermont	Colorado	California	U.S.A. pooled	New Zealand ($N = 299$)	Great Britain ($N = 298$)
Est-1	*107*	0.0200	—	0.0500	0.0217	0.0219 (0.0405)[1]	—	0.0017
	117	0.0267	0.0526	0.0250	0.0435	0.0219 (0.0378)[1]	0.0217	0.0067
Est-2	*95*	—	—	—	—	—	0.1582	0.0684 (0.022)[2]
	105	0.0424	—	0.1111	0.0455	0.0408	0.1400	0.0516 (0.031)[2]
Idh-2	*144*	—	—	—	—	—	0.0318	0.0017
Ldh-1	*133*	—	—	—	—	—	0.0060	—
Ldh-2	*57*	—	—	—	—	—	0.0083	—
	76	—	—	—	—	—	—	0.0017
	181	—	—	—	—	—	—	0.0017
Mpi	*83*	—	—	—	—	—	—	0.0050
Pep-A	*89*	—	—	—	—	—	—	0.0017
	95	0.0833	0.0789	0.0455	0.0682	0.0745	0.1129	0.1347
	106	—	—	—	—	—	—	0.0067
Pep-C	*82*	—	—	—	—	—	—	0.0034
	89	—	—	—	—	—	—	0.0017
	107	0.1299	0.1316	0.1000	0.0682	0.1160	0.0468	0.0893
Pgdh	*143*	—	—	—	—	—	0.0600	0.0116 (0.016)[2]
Pgm-1	*150*	—	0.0263	0.0250	0	0.0072	0.0150	0.0084
	170	—	—	—	—	—	0.0084	0.0184
Pgm-2	*60*	0.0974	0.0789	0.2250	0.0227	0.1014	0.0100	0.0083
	150	—	—	—	—	—	—	0.0083
	210	—	—	—	—	—	—	0.0050



Špačci v severní Americe
Vysazeni 1890
Přežilo asi 15 párů
Expanze

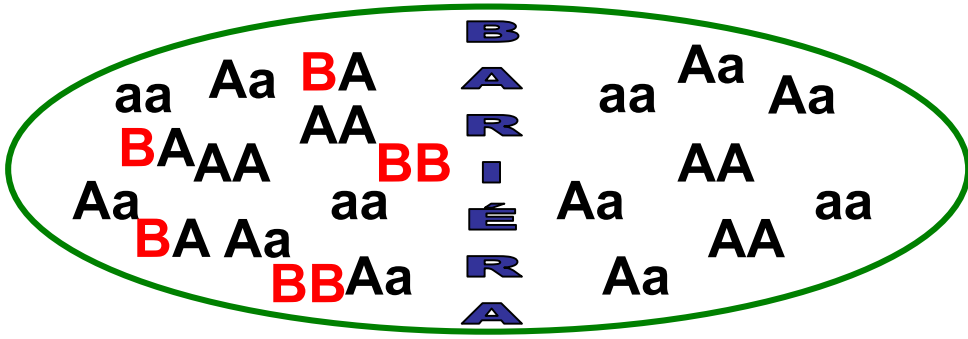
Cabe 1998

Locality		Karlovy Vary	Letňany	Hrádek	Trhovky	Kyjov	Vyškov	Bratislava
Locus	allele							
IGS-110b	137	0.880	0.3438	0.7917	0.0417	0.4286	0.5600	0.3519
	140	0.110	0.2656	-	-	0.3393	0.3400	0.2593
	143	-	-	-	-	-	-	0.2222
	146	-	0.3906	0.2083	0.9583	0.2321	0.1000	0.1667

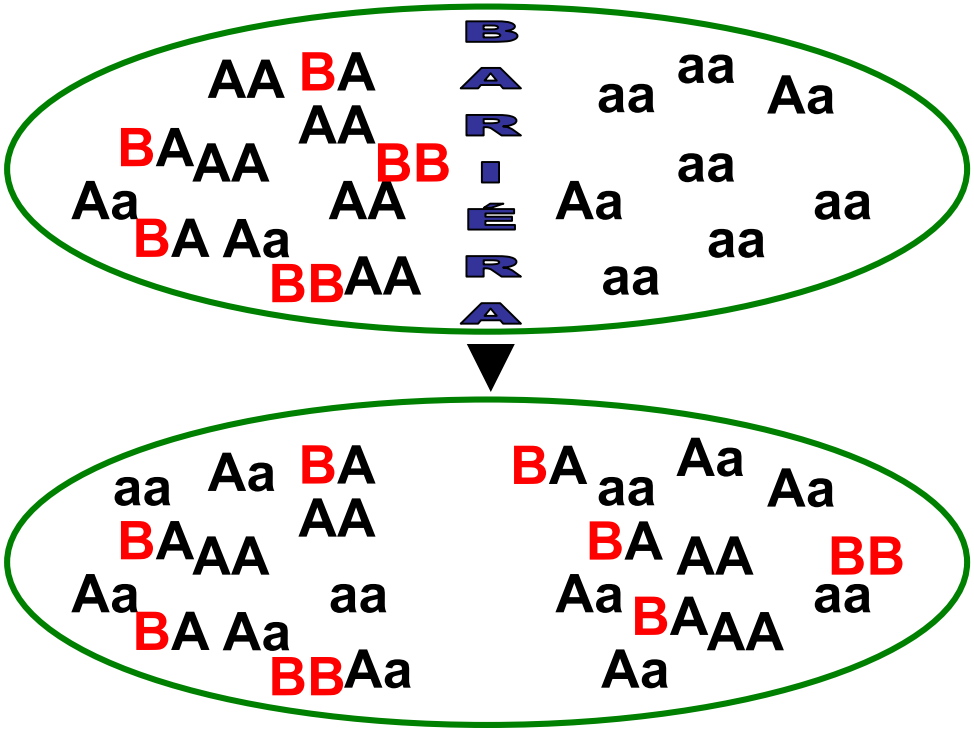
Hulová & Sedláček 2008

Mutace

- **Mutace**
  mohou zvýšit diferenciaci
  (odlišit subpopulace)



Migrace



- **Migrace**
  homogenizuje
  1 migrant na generaci – značný vliv!

Drift versus mutace

- Ztráta versus zvyšování variability v populaci

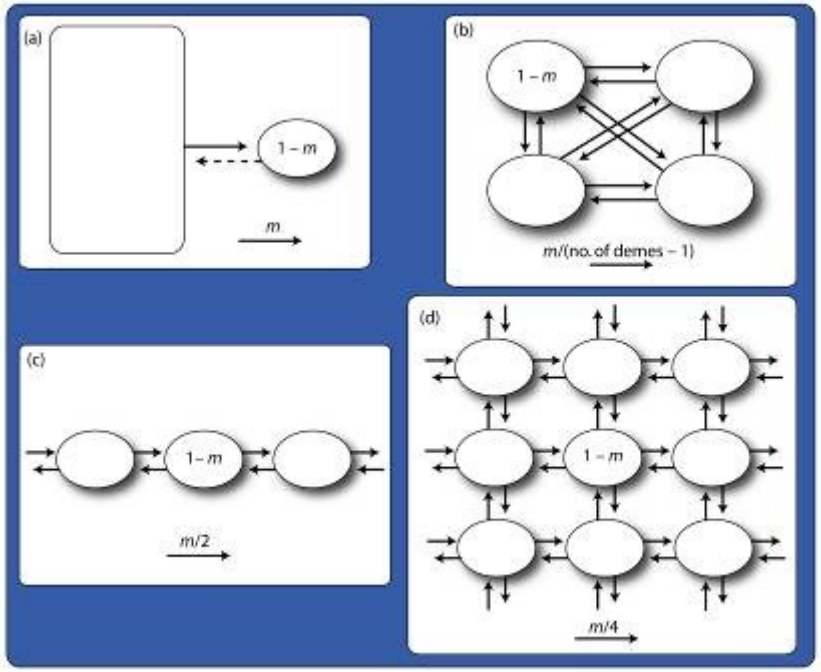
Drift a mutace versus migrace

- diferenciace versus homogenizace populací

Modely populační struktury (toku genů)

- Continent-island
- Island
- Stepping-stone

- Metapopulation
  extinction and recolonization



IBDL Isolation by dispersal limitation
→ IBD Isolation by distance

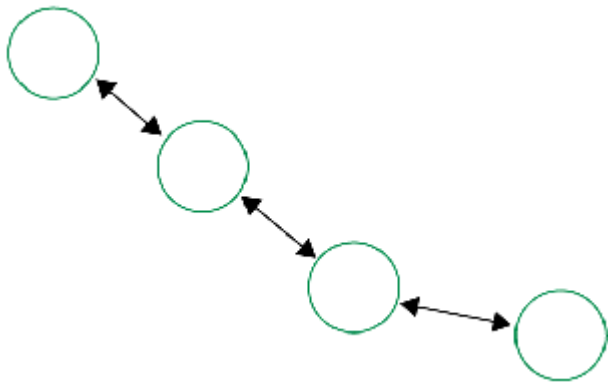
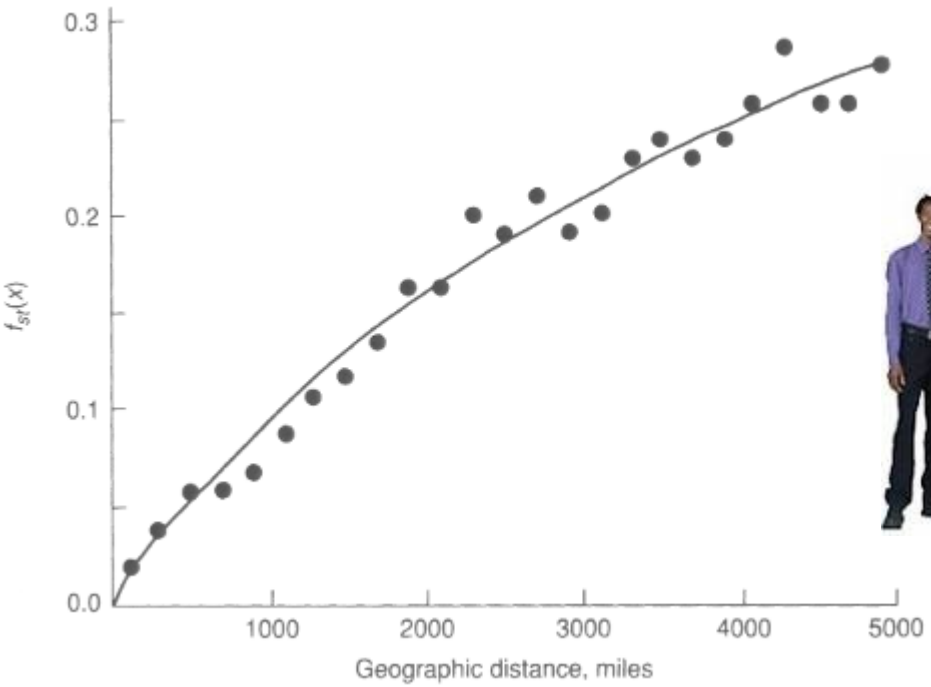


Table 6.1. Isolation by Distance in Ina Valley, Japan, as Measured by Location of Spouse's Birthplace for 2022 Marriages

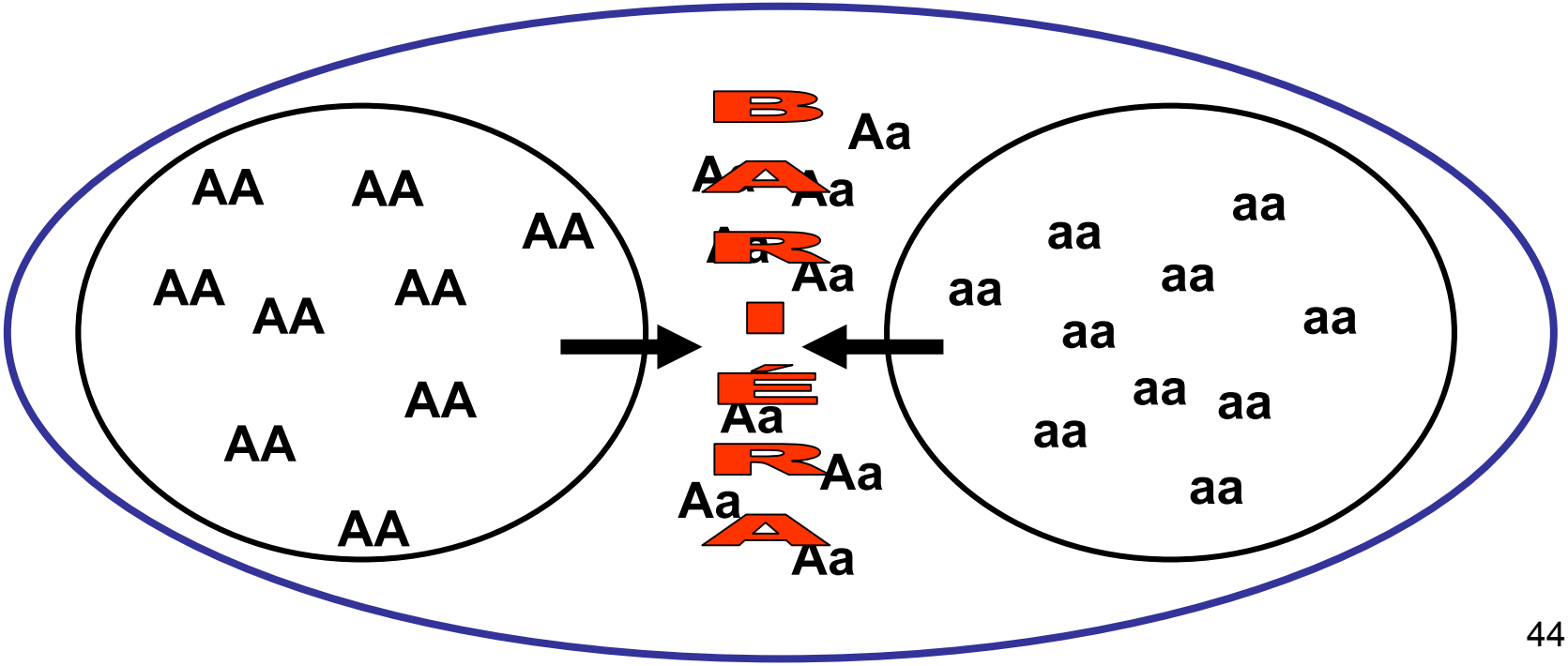
Spouse's birthplace	Percentage of marriages
Within Buraku (hamlet)	49.6
Within village but outside Buraku	19.5
Neighboring villages	19.1
Within Gun (county)	6.4
Within Prefecture (state) but outside Gun	2.9
Outside Prefecture	2.5

Suorce: Sekiguchi and Sekiguchi (1951).



Wahlundův princip

- Rozdělení na subpopulace – snížení celkové heterozygotnosti
- Isolate breaking – vzrůst heterozygotnosti při sloučení subpopulací
- F statistika F_{ST}, θ_{ST}



Fixační index

popisuje odchylku od očekávání

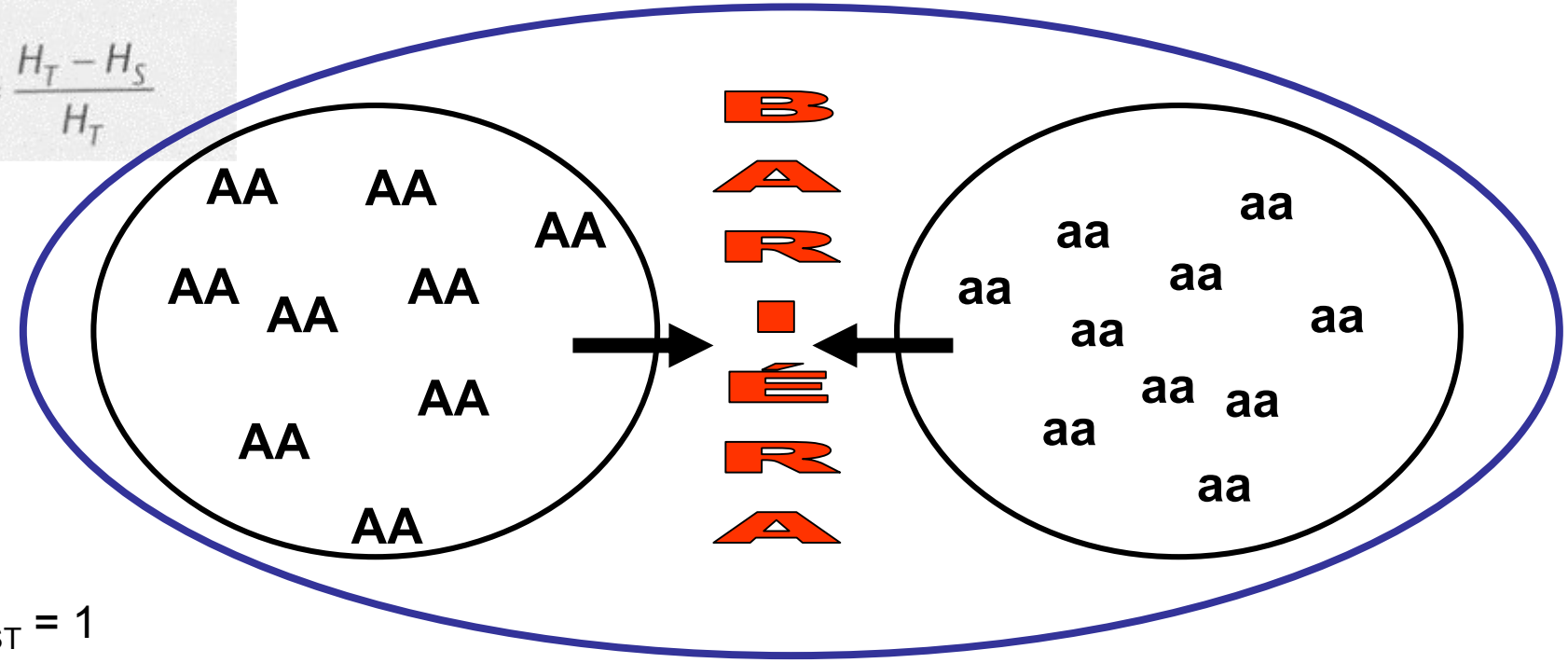
$$F = \frac{H_e - H_o}{H_e}$$

F-statistika

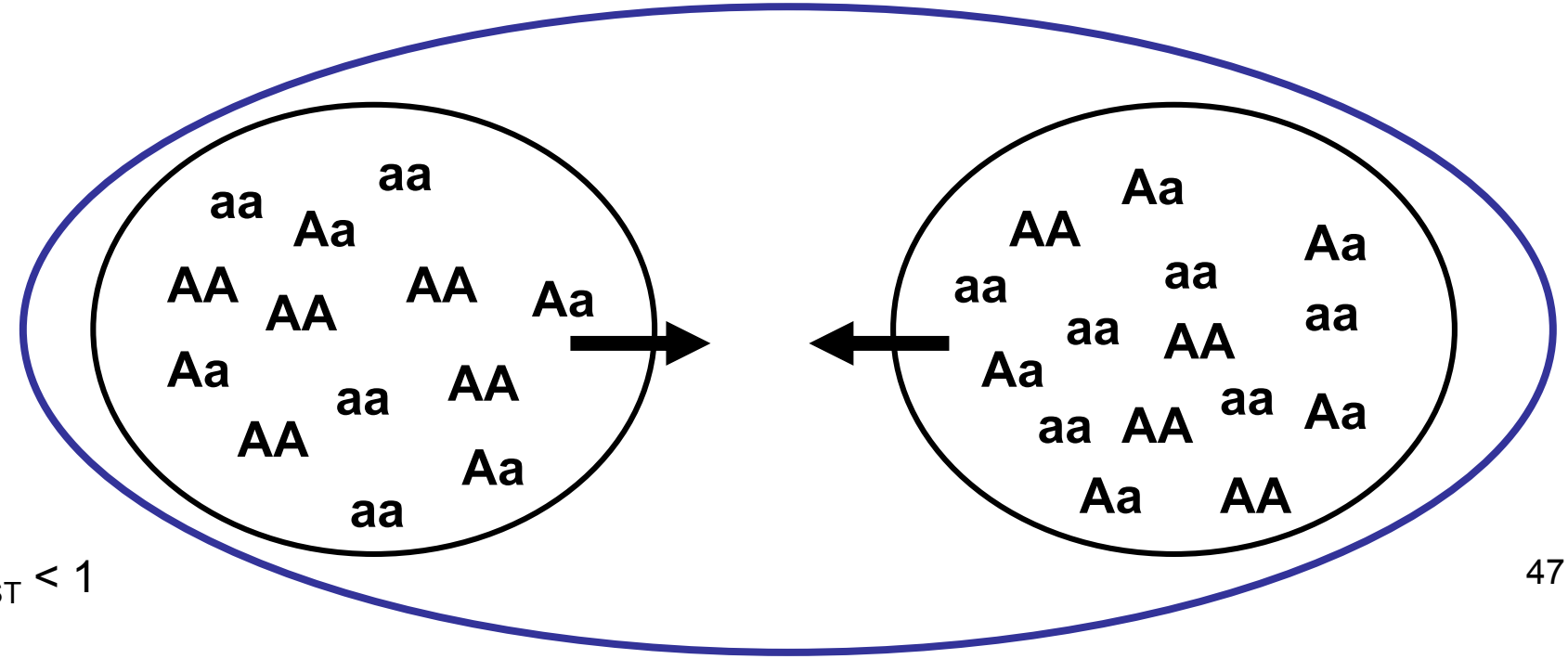
- H_I average observed heterozygosity within each population
- H_S average expected heterozygosity of subpopulations
- H_T expected heterozygosity of the total population

$F_{IS} = \frac{H_S - \bar{H}_I}{H_S}$	F_{IS}	**Snížení heterozygotnosti jedince kvůli nenáhodnému páření v subpopulaci**
$F_{ST} = \frac{H_T - H_S}{H_T}$	F_{ST}	**Vliv rozdělení populace na subpopulace (genetický drift)**
$F_{IT} = \frac{H_T - H_I}{H_T}$	F_{IT}	**Celkový koeficient inbreedingu F_{IT} - měří redukci heterozygotnosti jedince ve vztahu k celkové populaci**

$$F_{ST} = \frac{H_T - H_S}{H_T}$$



$F_{ST} = 1$



$F_{ST} < 1$

Obvykle nemáme genotypy všech jedinců

Frekvence odhadujeme (estimates)

→ výpočet je složitější

Census population size
≠ Effective population size

fluctuations

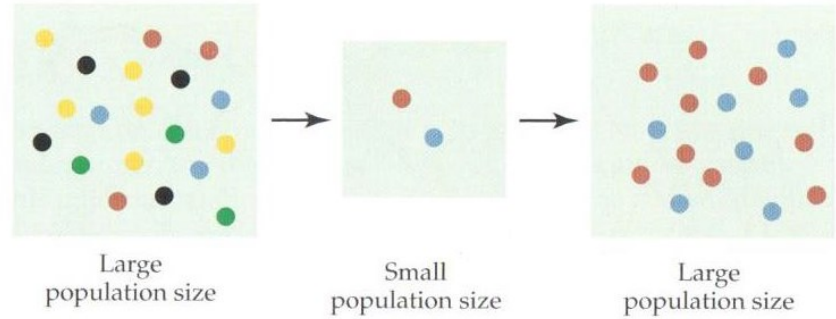
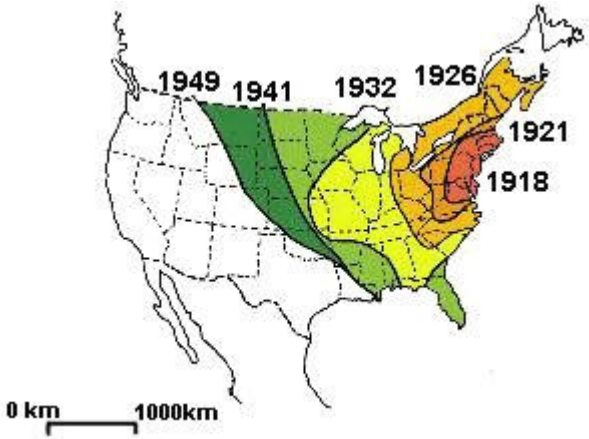
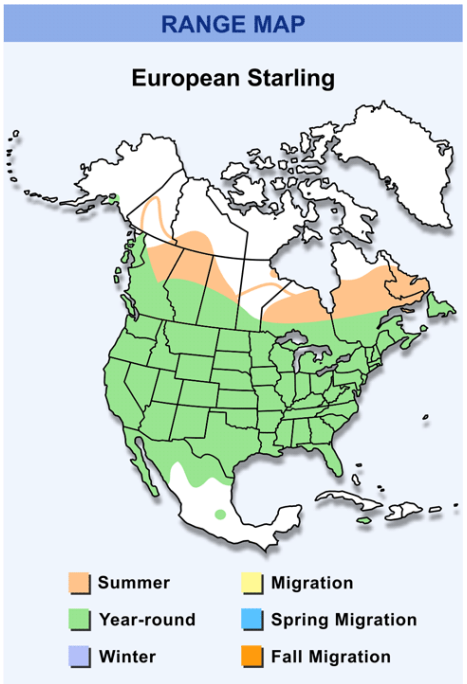


Figure 2.3 An illustration of the effect of a bottleneck in population size on genetic variability. The initial population has a high degree of variability, illustrated by the variety of colors of the balls in the box. The population goes through a bottleneck—a temporary decrease in size—and after the bottleneck, there are many fewer different alleles present.

Špačci v severní Americe
Vysazeni 1890
Přežilo asi 15 párů
Expanze





Census population size
≠ Effective population size

harem

Effective population size

- = number of individuals in an ideal population (in which every adult reproduces) in which the rate of genetic drift (measured by the rate of decline in heterozygosity) would be the same as it is in the actual population

- = the size of an ideal Wright-Fisher population that maintains as much genetic variation or experiences as much genetic drift as an actual population regardless of census size

- = the size of an idealized population that would show the same amount of genetic diversity as the population of interest

- **Census population size ≠ Effective population size**

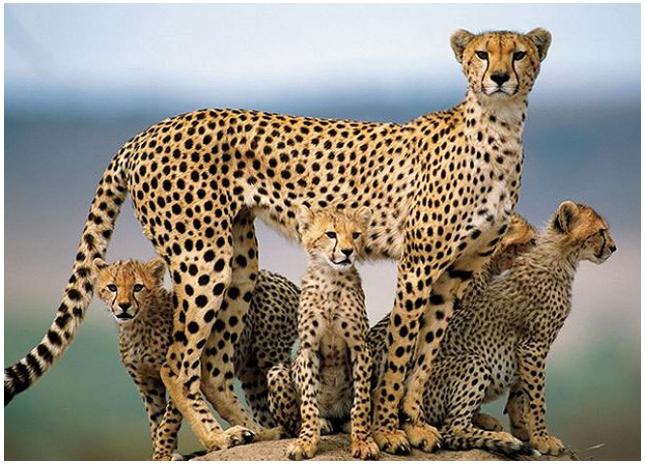
- Variation in the number of progeny
- Sex ration different from 1:1
- Natural selection
- Generation overlap
- Fluctuations in population size

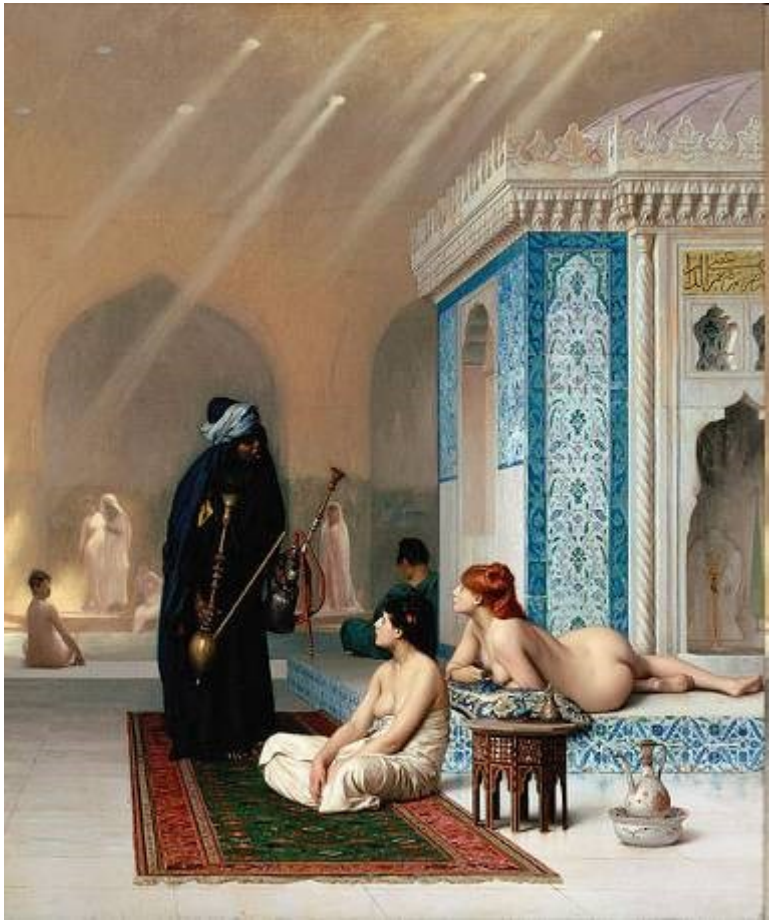
N_e/N averages about 0.10 to 0.14

N_e > 50 to minimize inbreeding depression

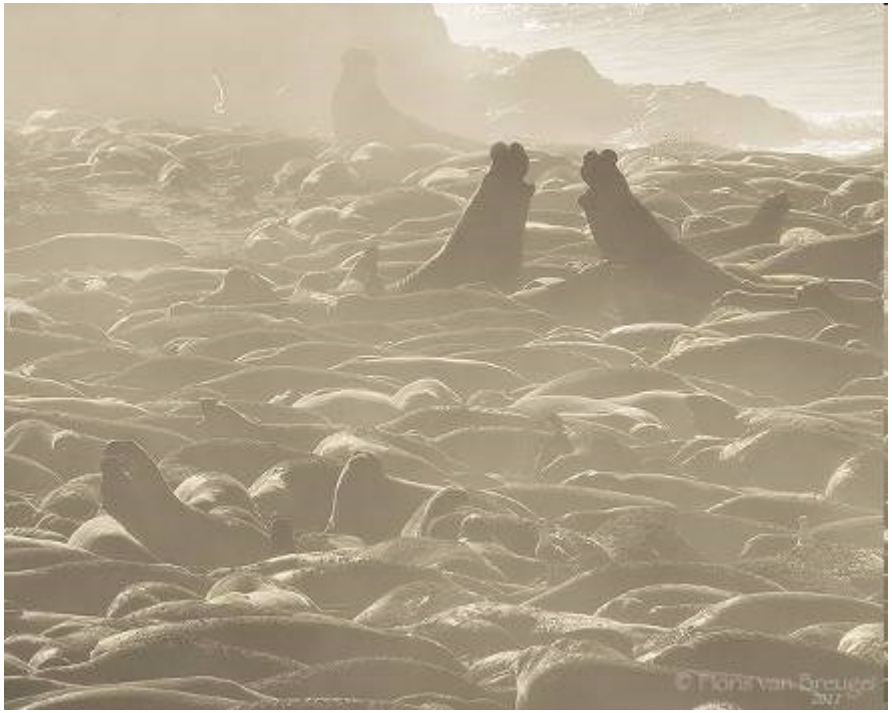
N_e > 500 to maintain sufficient evolutionary potential

Sice krásné ale ve skutečnosti totální hausnumera ☺





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



Příklad:

$N_f = 100$, $N_m = 100 \rightarrow N_e = 200$

$N_f = 100$, $N_m = 10 \rightarrow N_e = 20$

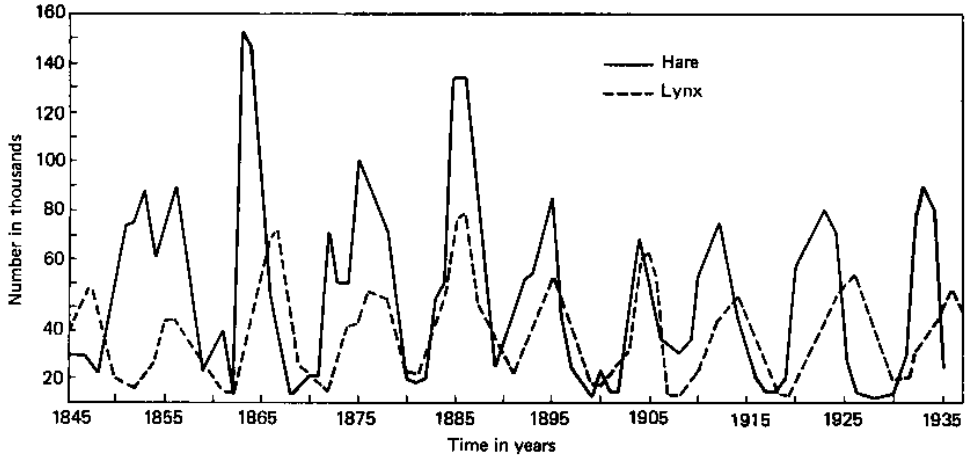


Figure 9-3. Changes in the abundance of the lynx and the snowshoe hare, as indicated by the number of pelts received by the Hudson's Bay Company. This is a classic case of cyclic oscillation in population density. (Redrawn from MacLulich 1937.)

- Fluctuating size
- $N_e \rightarrow$ harmonic mean

$$\bar{x} = n/(\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n})$$

- Dominated by the smallest terms

Arctocephalus australis
Lachtan jihoamerický
Rosa de Oliveira et al. 2006

The census population size 8223 individuals: 337 breeding males, 3215 breeding females

prior to the El Niño
24,481 individuals: 2903 reproductive males and 10,720 reproductive females.

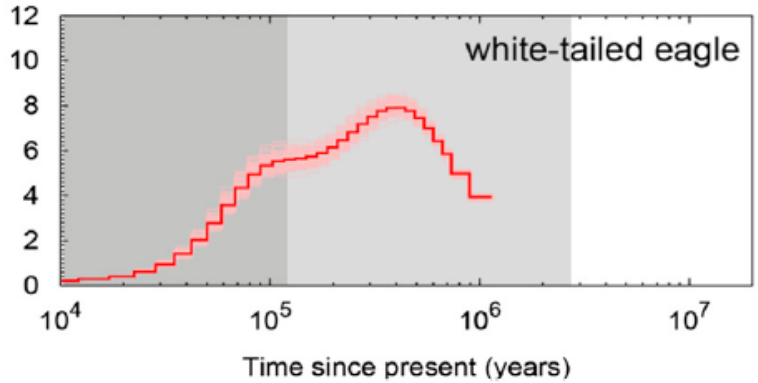
$$N_e = \frac{4N_{ef}N_{em}}{N_{ef} + N_{em}} \qquad N_e = \frac{t}{\sum 1/N_i}$$

→ *Ne* 2153

Výpočet Ne

- neutral genetic diversity estimates (i.e., $\theta = 4N\mu$; where μ is the mutation rate and N is the effective population size)
  - **'Lewontin's paradox'** — the discrepancy between the many orders of magnitude of variation in population size and the much narrower distribution of diversity levels.
- linkage disequilibrium (LD), heterozygote excess, molecular coancestry and sibship frequency
- Koalescenční přístupy

Ne



N_e je různé pro různé části genomu

- mtDNA
- Pohlavní chromosomy
  Ne of A, X and Y should scale 4:3:1
- Ale! Vliv variance v reprodukčním úspěchu
- When the variance in reproductive success among males is higher than among females, as is often the case with sexual selection, organisms with male heterogamety should show X:A > 0.75 and organisms with female heterogamety should show Z:A < 0.75

Huang & Rabovski 2015

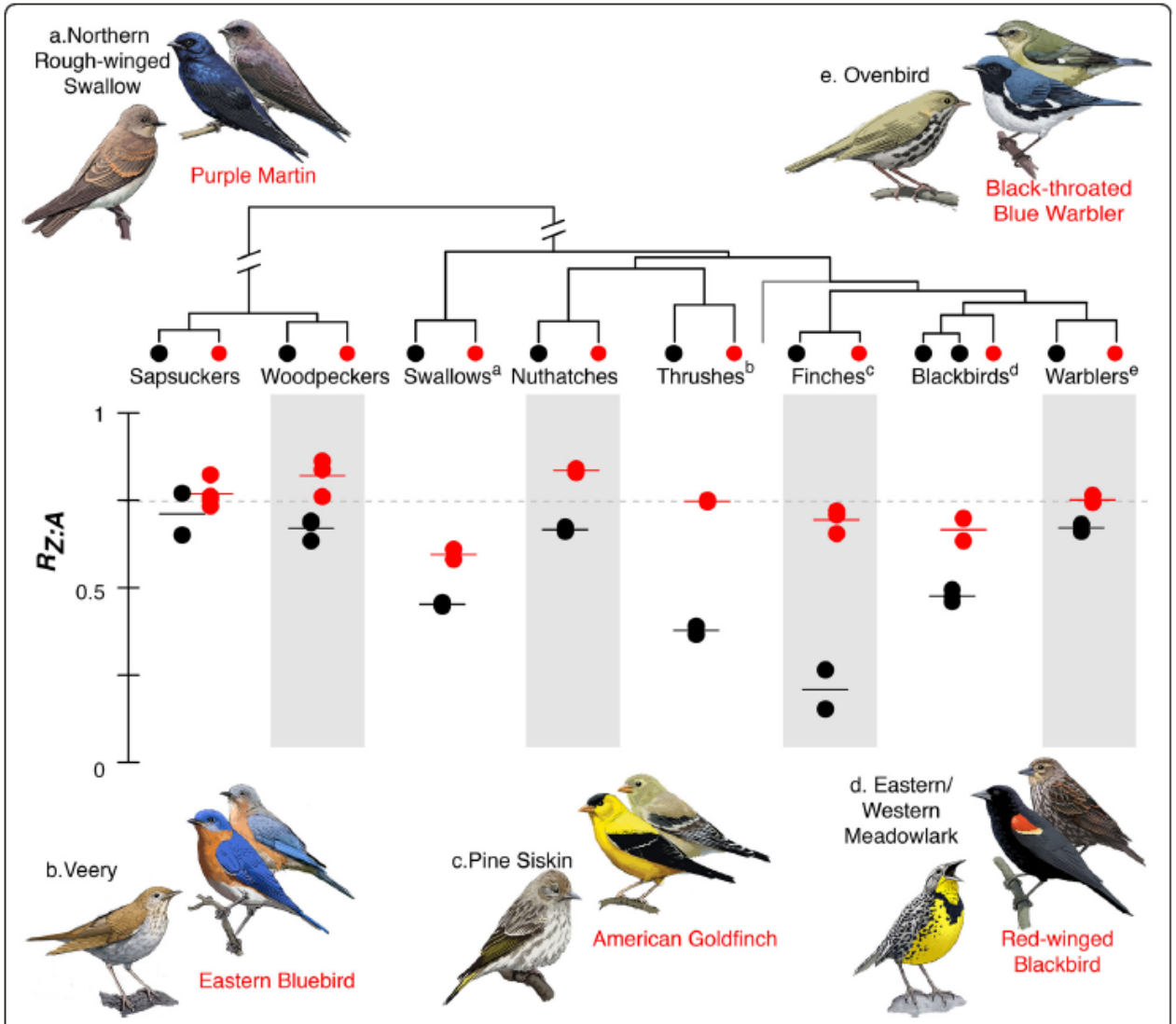


Fig. 1 Ratios of Z-associated and autosomal effective population sizes. Matched pairs of sexually dichromatic (red dots) and monochromatic (black dots) bird species are plotted together. Short horizontal lines represent species averages across individuals. The top panel shows the phylogenetic relationship among studied species (grey line represents the position of Zebra Finch). Five of the species pairs are illustrated in surrounding plates with upper right dichromatic species (males in front) and lower left monochromatic species. The three pairs not shown are: Williams' Sapsucker vs. Red-breasted Sapsucker, Red-bellied Woodpecker vs. Red-headed Woodpecker, and Red-breasted Nuthatch vs. Pygmy Nuthatch. Monochromatic species have reduced genetic variation on the Z chromosome compared to its matching dichromatic species (Wilcoxon test $p = 0.008$; mixed-effect linear regression $p < 0.001$; PGLS regression $p < 0.001$; see Additional file 6 for details). Grey dotted line represents the null expectation that Z-associated effective population size is ¾ of autosomal effective population size

člověk

- Odhady N_e zhruba desítky tisíc
- Důvody:
  - Bottleneck
  - Populační struktura