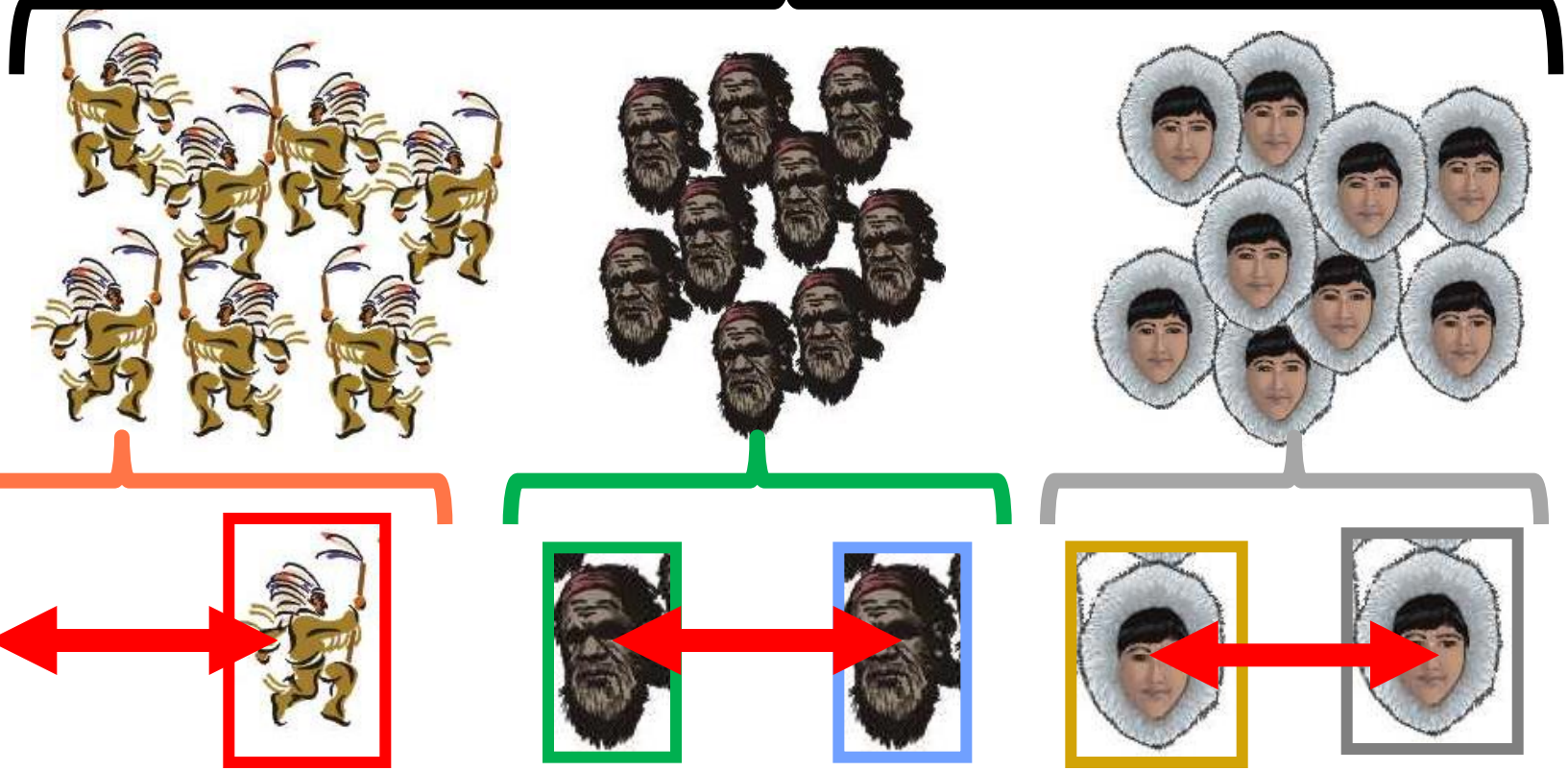
POPULATION GENETICS

SPECIES

POPULATION



SUBPOPULATIONS

II. GENETIC STRUCTURE OF POPULATION

Assumption for population structure analysis:

- **neutral loci** = no effect of natural selection included

- classical population genetics approach = populations are ***a priori*** *(thought to be)* known (e.g. we want to quantify level of genetic differentiation between two localities / ?populations)

- BUT populations are **not usually known** (e.g. due to no obvious spatial heterogeneity over the distribution range) – we want to **reveal any potential population differentiation/structure according to our genetic data -> non-*a priori* methods**

Genetic structure – any pattern in the genetic make-up of individuals within a population

AIMS:

- Detection of **any** genetic structure (subdivision) in a population (in my dataset)
- Are there any **differences** between „different“ (in space and time) populations?

- Quantification of such differences = **description of genetic structure in population (of genetic differentiation between (sub)populations)**

- What factors shape (have shaped) these differences? e.g. **population history**

- Is there any migration/connection between different populations? = detection and quantification of **gene flow**, what influences gene flow (e.g. **spatial heterogeneity**)

- What happens during migration/connection of populations? = **hybridisation**

Population genetic structure

neutral markers

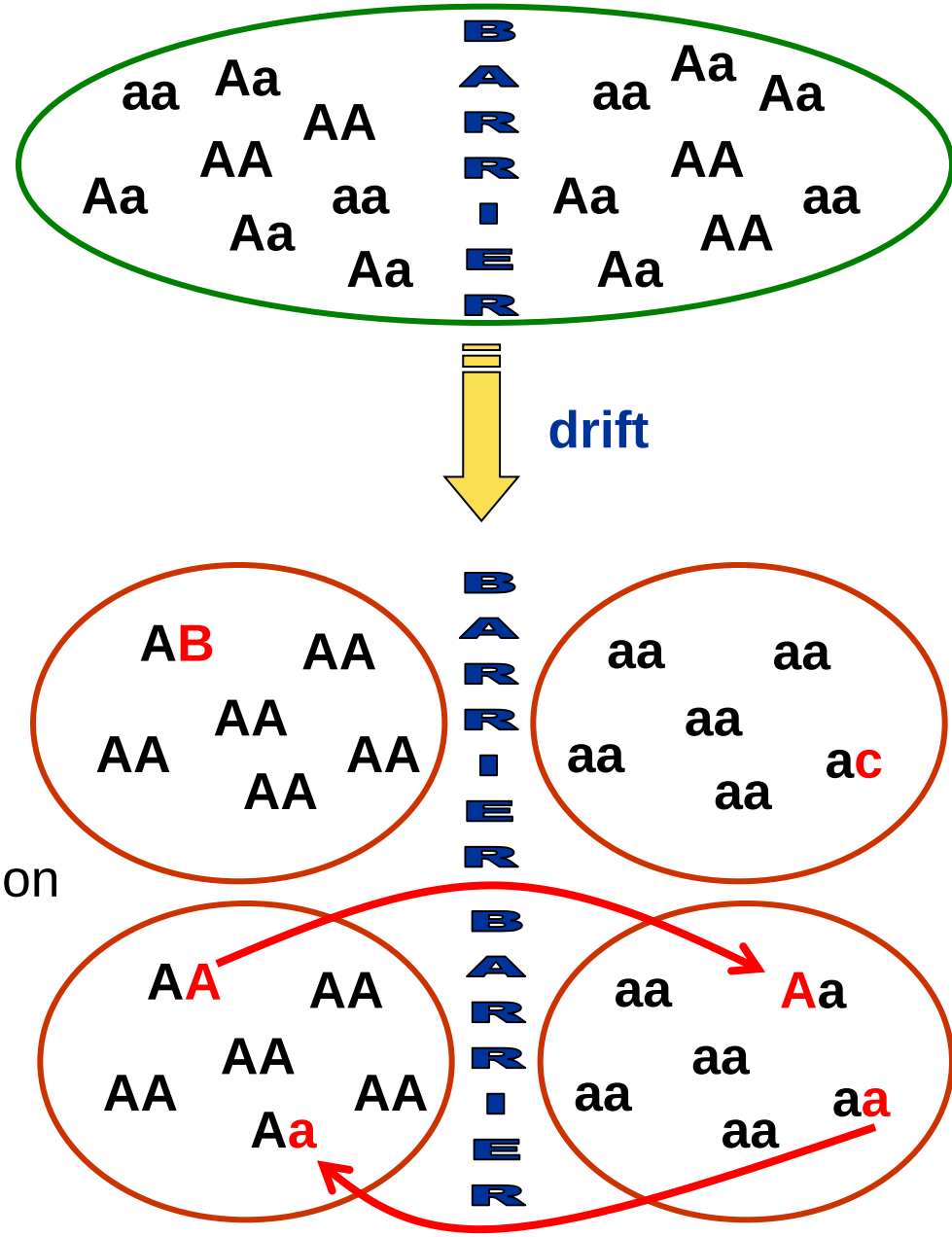
- **GENETIC DRIFT**
  - creates subpopulation differentiation

  (changes in allele frequencies – extremely up to fixation of distinct alleles)

- **MUTATION**
  may increase differentiation

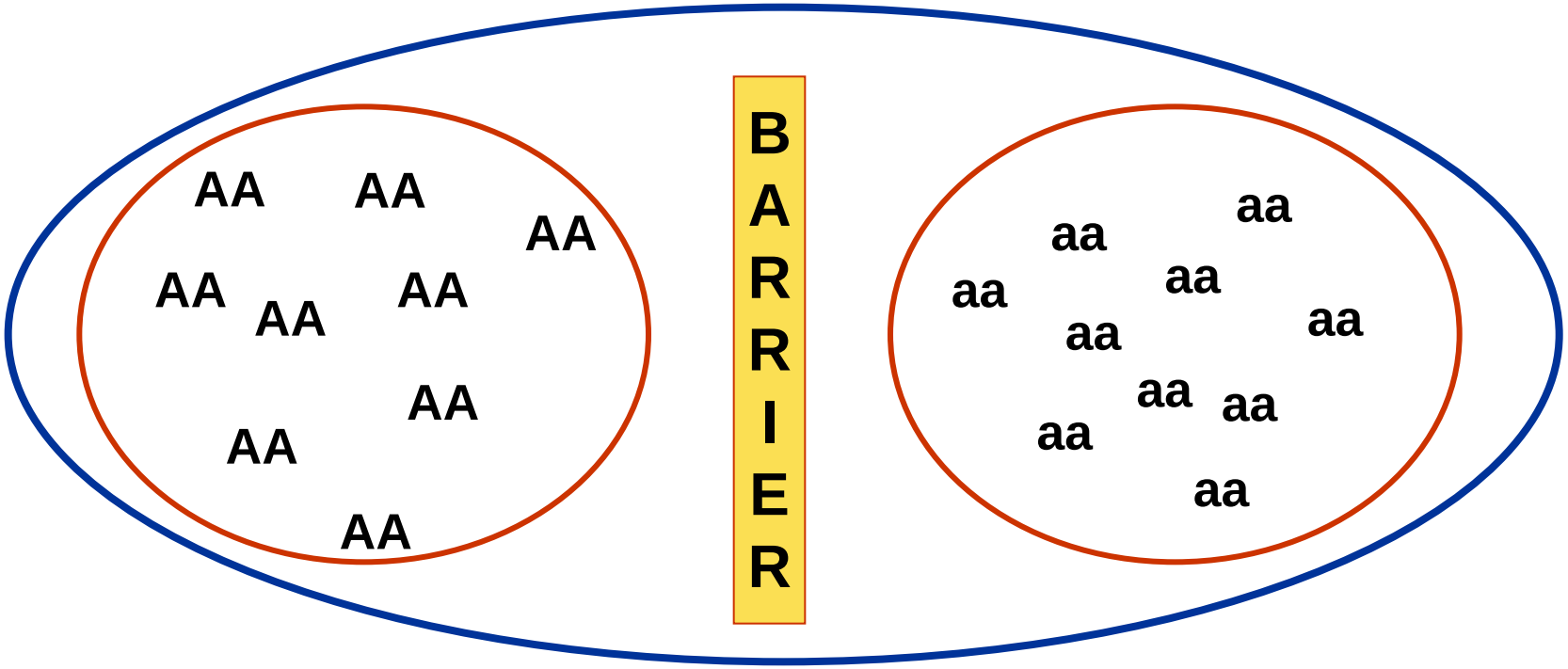
- **INBREEDING**
  increase of homozygotes proportion

MIGRATION (GENE FLOW)
- AGAINST subpopulation differentiation



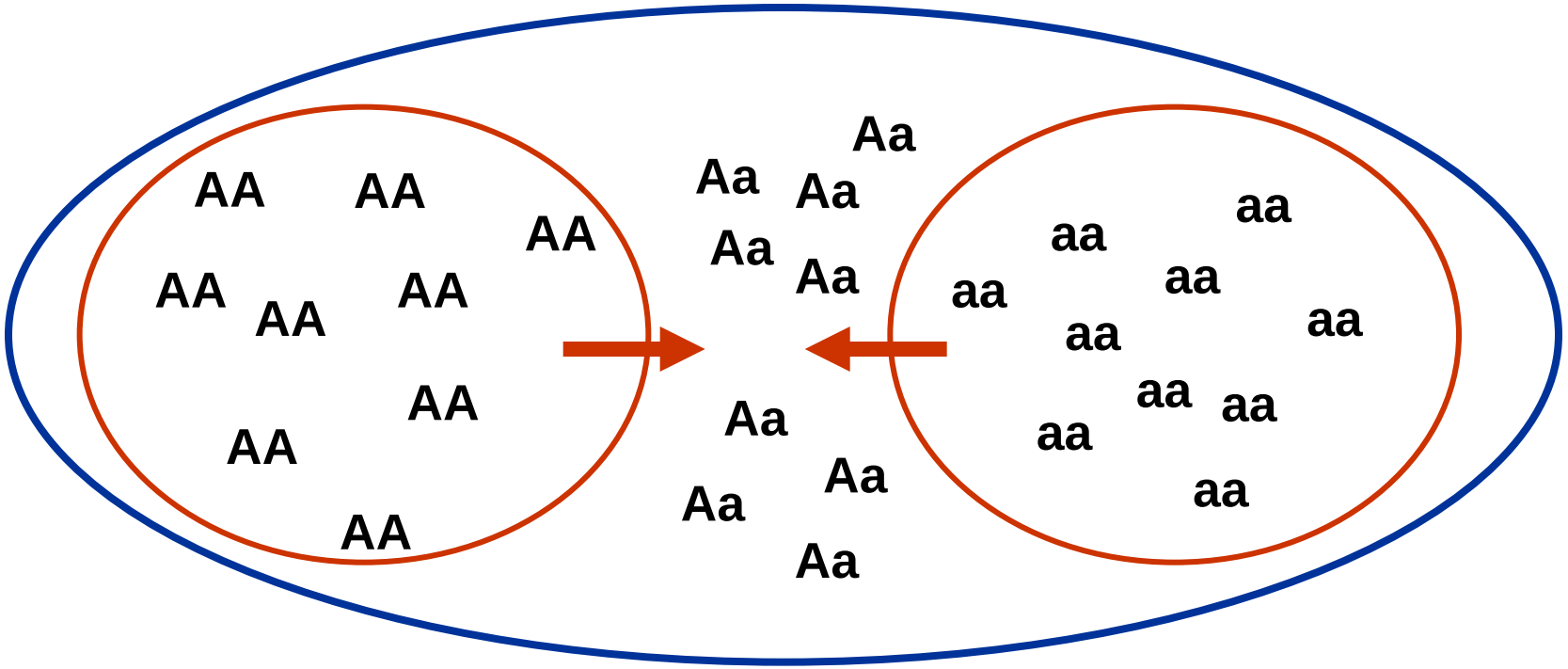
Effect of population structure on heterozygosity

- **Wahlund effect** – first documented by Swedish geneticist Sten Wahlund (1901-1976) in 1928
- both SUBPOPULATIONS are in HWE, but the pooled dataset (the whole POPULATION) shows **deficit of heterozygotes**
- Extreme ex.: two isolated subpopulations with fixed distinct alleles (more generally – subpopulations with different allele frequencies)



→ isolation breaking

Homozygosity reduction when subpopulations merge



Wahlund, S. (1928) Zusammensetzung von Population und Korrelationserscheinung vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas,* 11: 65–106

Wahlund effect – an example

- Bunnersjöarna lake (northern Sweden) – *Salmo trutta*
- one trait with 2 alleles

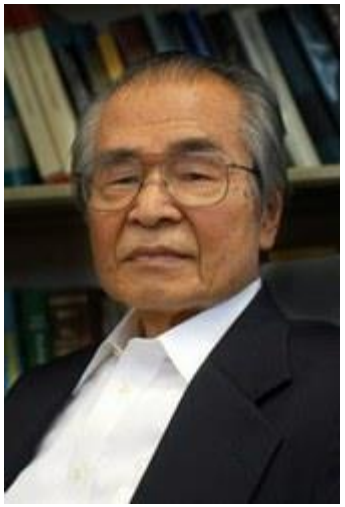
	170/170	170/172 (= Ho)	172/172	Total	p	2pq (=He)
Přítok	50	0 (**0**)	0	50	1.000	**0.000**
Odtok	1	13 (**0.26**)	36	50	0.150	**0.255**
Whole lake	51	13 **(0.13)**	36	100	0.575	**0.489**

DECREASE OF HETEROZYGOSITY DUE TO POPULATION SUBDIVISION

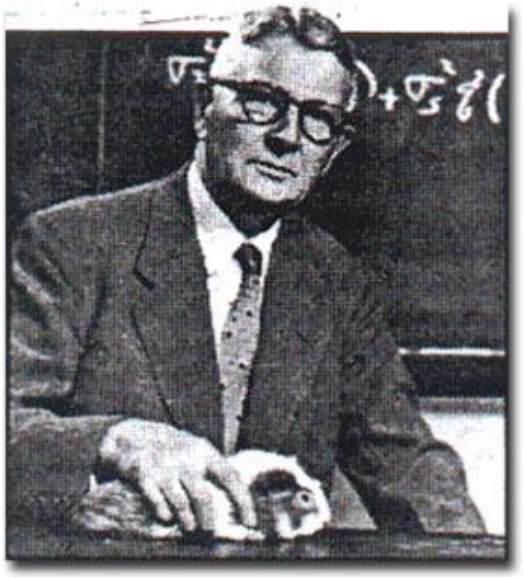
Ryman et al. 1979

Wright´s F-statistics

F_{IS}, F_{ST}, F_{IT}



Masatoshi Nei
***1931**



Sewall Wright
1889 - 1988

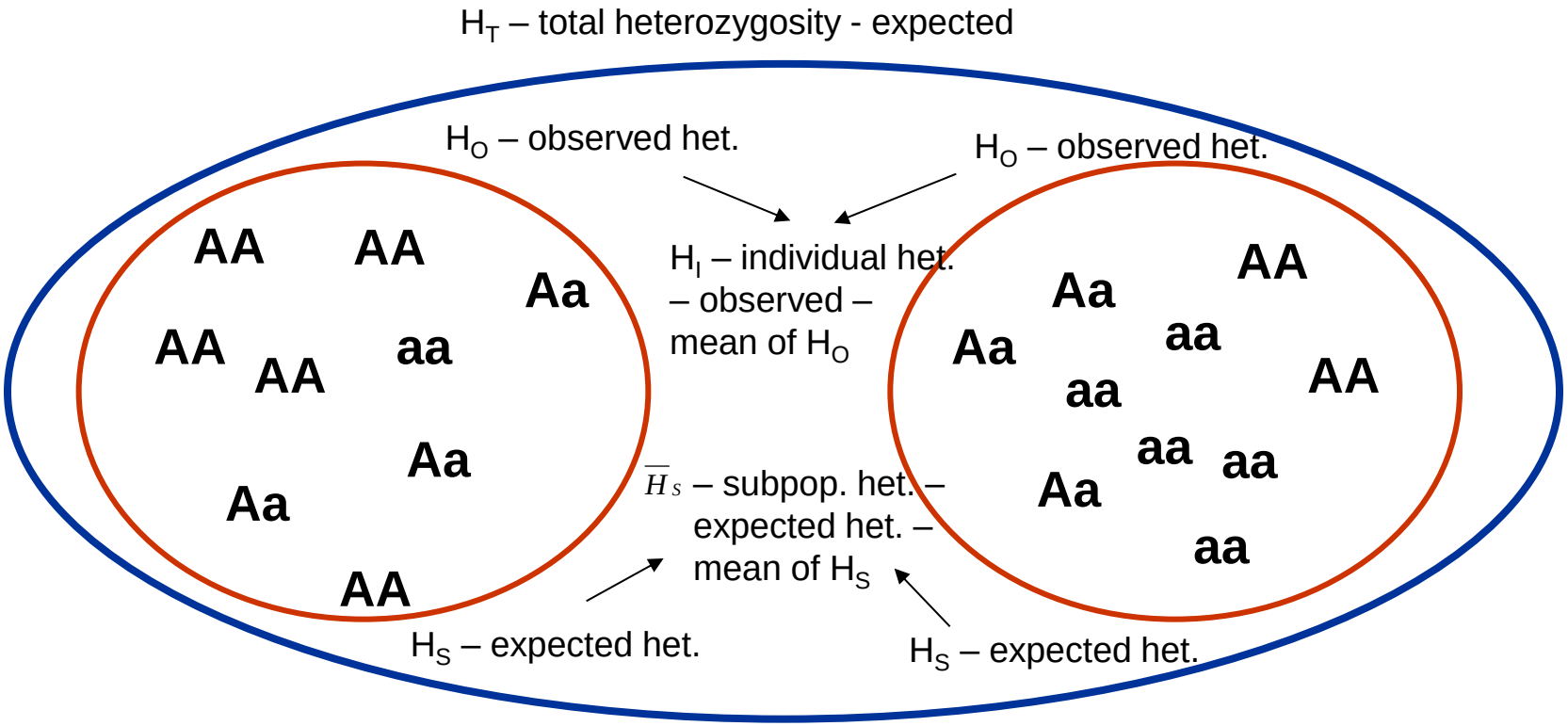
- Wright (1951), Nei (1987)
  - for two alleles at a single locus (Wright 1951)
  - more complicated for more alleles (Nei 1987)

- detecting and describing heterozygosity decrease
- describing heterozygosity (and its deviation from HWE) **at different levels**

TADY SKONČIL

F-statistics and heterozygosity

H_I – averaged **observed** heterozygosity of an individual in a subpopulation
H_S – **expected** heterozygosity of an individual in a subpopulation under HWE
H_T – **expected** heterozygosity of an individual over the total population under HWE



F-statistics and heterozygosity

H_I – averaged **observed** heterozygosity of an individual in a subpopulation
H_S – **expected** heterozygosity of an individual in a subpopulation under HWE
H_T – **expected** heterozygosity of an individual over the total population under HWE

$$H_I = \sum_{x=1}^{k} H_O / k$$

H_x = observed heterozygosity in subpopulation x

$$H_S = 1 - \sum_{i=1}^{j} p_{i,x}^2$$

$p_{i,x}^2$ = frequency of i-th allele in subpopulation x

$$\overline{H}_S = \sum_{x=1}^{k} H_S / k$$

averaged expected heterozygosity over subpopulations

$$H_T = 2 p_0 q_0$$

p_o = allele frequency in the total population

F-statistics

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

Heterozygosity decrease of an individual due to non-random mating in a subpopulation (vs. HWE)

Heterozygosity over all populations

Mean heterozygosity within subpopulations

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

Influence of division of the total population in subpopulations (i.e. heterozygosity decrease due to Wahlund effect)

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

Total coefficient of inbreeding F_{IT} - measures heterozygosity decrease of an individual in relation to the total population

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

Weir & Cockerham (1984) ***f*** **(~ F_{IS}),** ***θ*** **(~ F_{ST}),** ***F*** **(~ F_{IT})**

Correction for sample size and number of subpopulations

Computation of F-statistics

Computation of allele frequencies

Mean allele A frequency in the whole population

	Subpopulation 1 (N_1=40)				Subpopulation 2 (N_2=20)					
Locus	AA	AB	BB	$p_{1(j)}$	AA	AB	BB	$p_{2(j)}$	$p_{0(j)}$	Note
Loc I	10	20	10	0.5	5	10	5	0.5	0.5	**HWE**
Loc II	16	8	16	0.5	4	4	12	0.3	0.4	**heterozygote deficit**
Loc III	12	28	0	0.65	6	12	2	0.6	0.625	**heterozygote excess**
Loc IV	0	0	40	0.0	20	0	0	1.0	0.5	**alternatively fixed alleles**

Computation of heterozygosities

	Observed heterozygosity		Expected heterozygosity			Wright´s F-statistics		
Locus	$H_{1\,(j)}$	$H_{2\,(j)}$	$H_{I\,(j)}$	$H_{S\,(j)}$	$H_{T\,(j)}$	$F_{IS\,(j)}$	$F_{ST\,(j)}$	$F_{IT\,(j)}$
Loc I	0.5	0.5	0.5	0.5	0.5	0.0	0.0	0.0
Loc II	0.2	0.2	0.2	0.46	0.48	0.565	0.042	0.583
Loc III	0.7	0.6	0.65	0.4675	0.46875	-0.39	0.0027	-0.387
Loc IV	0.0	0.0	0.0	0.0	0.5	---	1.0	1.0
Mean						0.058	0.261	0.300

Mean values of F-statistics may hide distinct evolution history of different loci

F-statistics

- F_{IS} *decrease of heterozygosity in local subpopulation*
  high values –> inbreeding

- F_{IT} summary measure – limited use

- F_{ST} = **subdivision measure** = *limited gene flow between subpopulations* (i.e. existence of a barrier –> Wahlund effect)

  – originally developed for estimation of the amount of allelic fixation due to genetic drift (**fixation index**)

F_{ST} computation – an example

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

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

As a consequence of gene flow barrier:
Heterozygosity is about 72.8% lower than would be under HWE

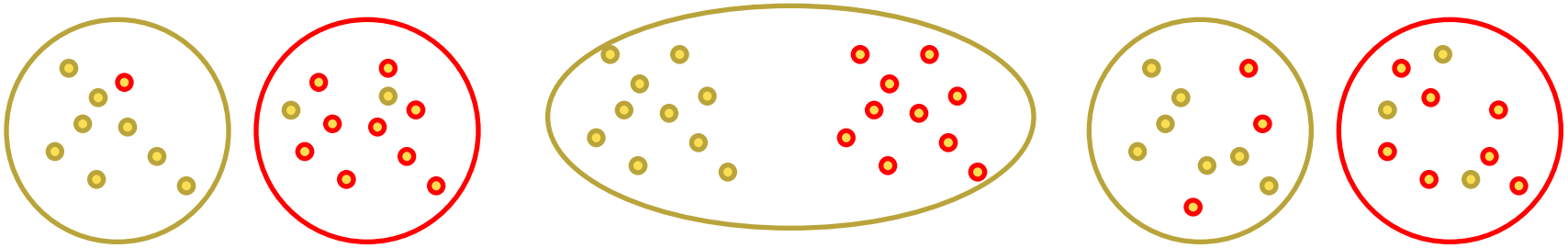
Ryman et al. 1979

Permutation test of F_{ST} significance

1. Real measured populations

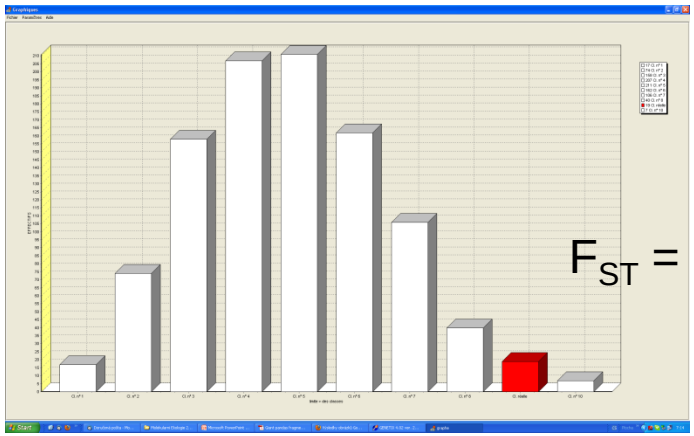
2. Merged into a single dataset

3. 1000 x randomly re-separated populations



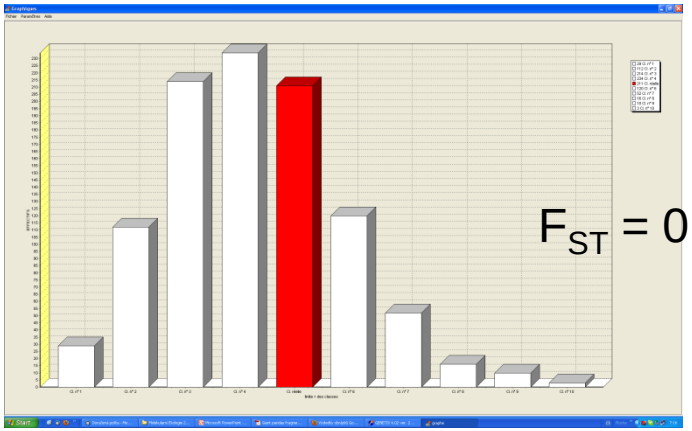
Real Fst

1000 x simulated Fst



F_{ST} = 0.072

0.8 % simulated values higher than real Fst
p = 0.008 (i.e. significant difference)



F_{ST} = 0.0013

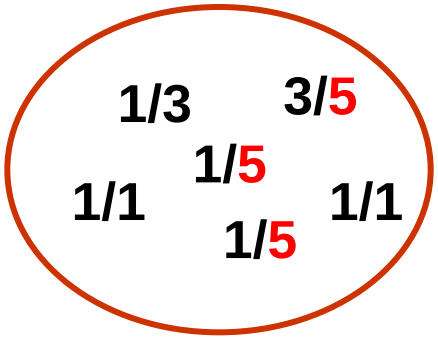
35.4 % simulated values higher than real Fst
p = 0.354 (e.g. non-significant difference)

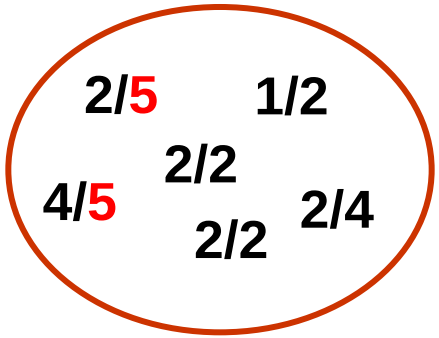
F_{ST} analysis – BE AWARE

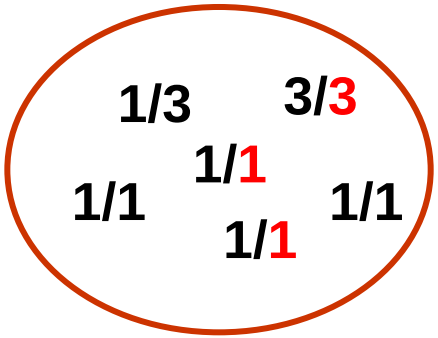
Global vs. pairwise indices

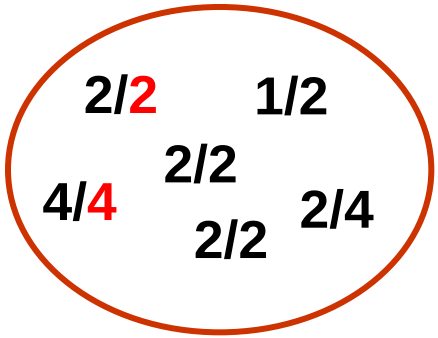
Absolute values depends on heterozygosity level of used loci!!!
(i.e. microsatellite-based F_{ST} cannot be compared to allozyme-based F_{ST})
Demands standardization: $F_{ST}' = F_{ST}/F_{STmax}$ (Hedrick 2005)
– e.g. GenAlEx

In case of null alleles presence: needs to be corrected!
(increase of homozygosity); FreeNA software









Giant Panda

- 192 feces samples→ 136 genotypes→ 53 unique genotypes
- separation by a river (ca 26 ky ago) and by roads (recently)
- even the roads are important barriers, even if less

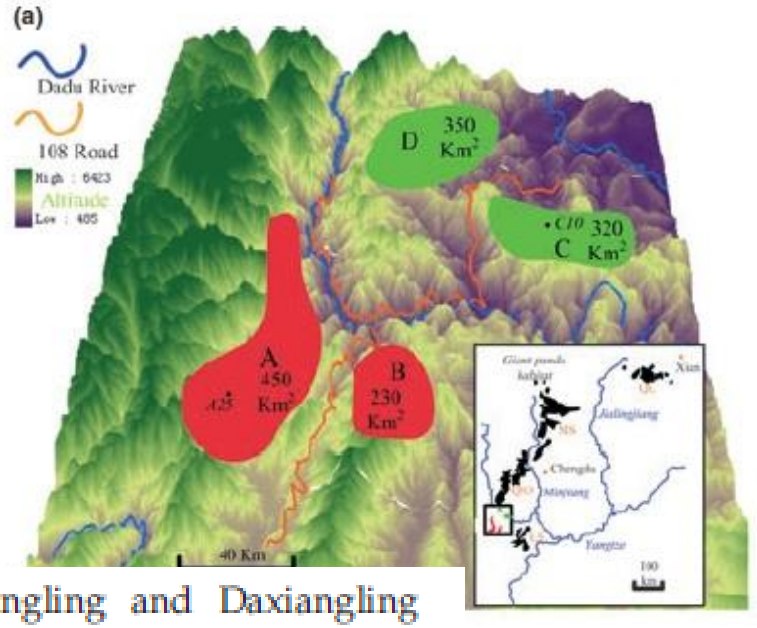


Table 3 Pairwise F_{ST} in the Xiaoxiangling and Daxiangling populations

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

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

(Zhu et al., 2011)

G_{ST} (Nei 1973)

- Analogy of F_{ST} for **haploid (haplodiploid) organisms, mtDNA sequences**

- Takes into account **haplotype (gene) diversity** instead of heterozygosity

- *Haplotype diversity* = probability that any two randomly chosen sequences in a population will be different

R_{ST}

- Analogy of F_{ST}
- Takes into account **the size of alleles** (number of repeats in microsatellite loci)
- Assumption of a known mutation model
  assumption of SMM (*stepwise mutation model*)
- Indicates traces of mutations
  - **$R_{ST} > F_{ST}$ higher effect of mutations**
  - **$R_{ST} = F_{ST}$ higher effect of genetic drift**
- Randomisation tests for R_{ST} significance (Hardy et al. 2003, program SPAGeDi 1.1)