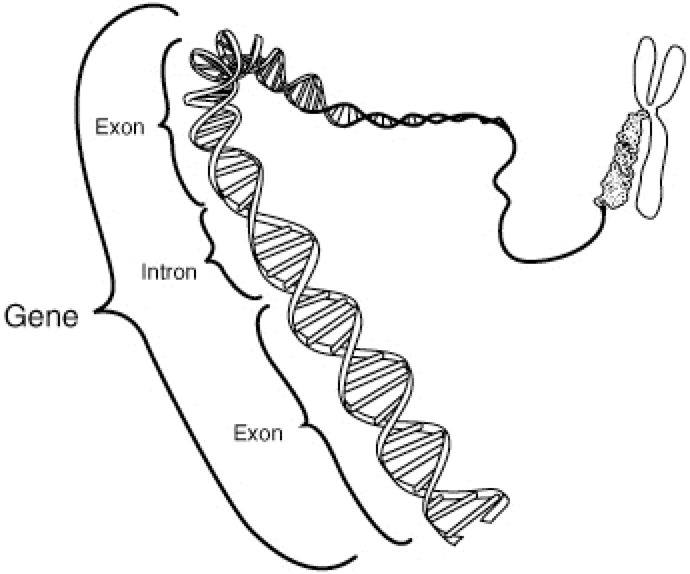
Population structure and paternity studies

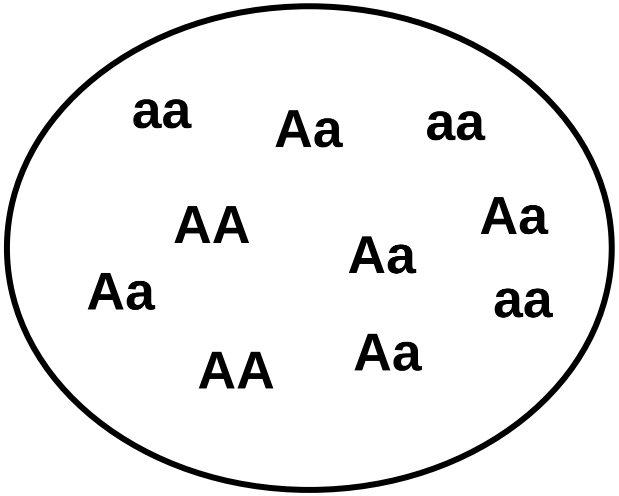
Terms



- Gene
- Locus (loci)
- Allele – alternative forms of a gene
- Genome – single copy of all of the genes
- Diploid - 2 genomes
  (exceptions Chr Y, cpDNA a mtDNA)
- Alleles at a single locus – genotype (AA, Aa, aa)
- Same alleles – homozygote (AA aa)
- Different alleles – heterozygote (Aa)
- Haplotype – array of alleles of linked genes, DNA sequence

Heterozygosity

- Proportion of heterozygous loci
- More loci in a sigle individual,
  single lokus in more individuals
- Whole genome



- **Allele frequency**
- A 9/20 = 0.45 (*p*)
- a 11/20 = 0.55 (*q*)

- **Genotype frequency**
- AA 2/10 = 0.2
- aa 3/10 = 0.3
- Aa 5/10 = 0.5
  → **observed** heterozygosity (H_O) = 0.5

Parentage **Overview** of the **Problem**

- Both parents unknown
- One known, second unknown
- One or two parents?
  Hermaphrodites
- **→ alleles only from mother**
  → heterozygote deficit (< HW)
- Selfing or clonality (parthenogenesis)
  clones → fixed heterozygosity (no segregation)

Chen caerulescens

Limacidae
outcrossing

Epiactis prolifera
selfing

Parent Aa, gamets A a
selfing → progeny AA, Aa, aa
clonality → progeny Aa

Deroceras laeve
clonality

Molecular techniques

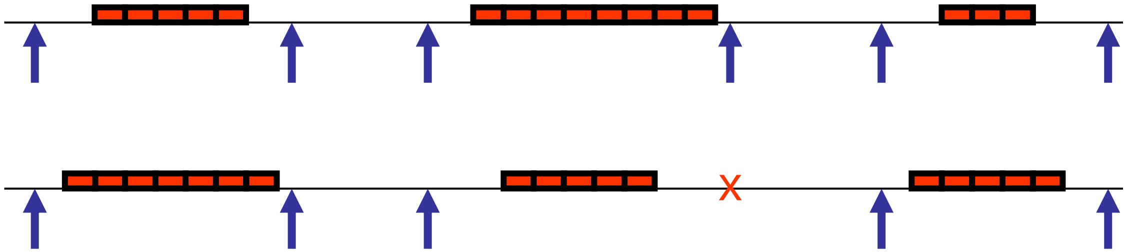
- **DNA fingerprinting**

- **Microsatellites**

- **SNPs**

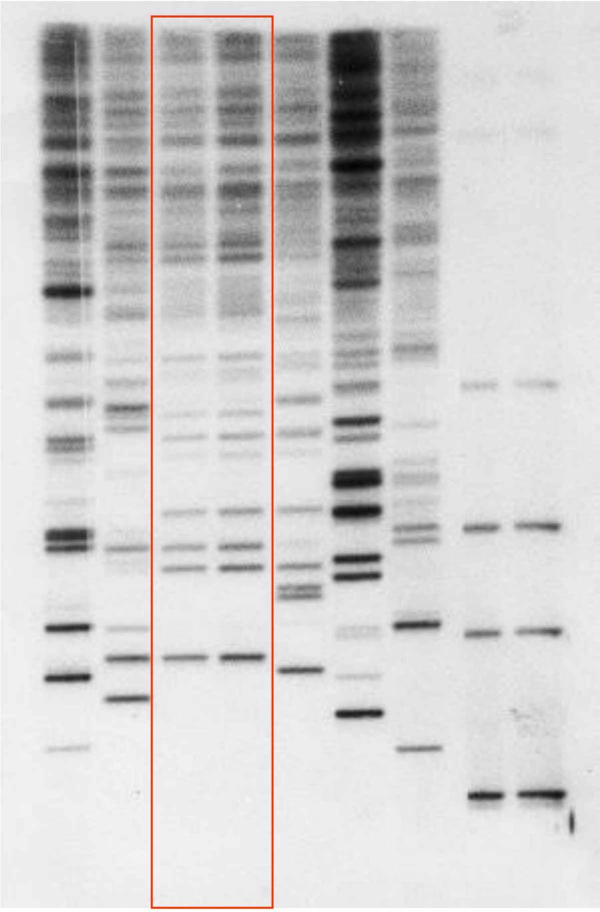
Sir
Alec Jeffreys

DNA fingerprinting



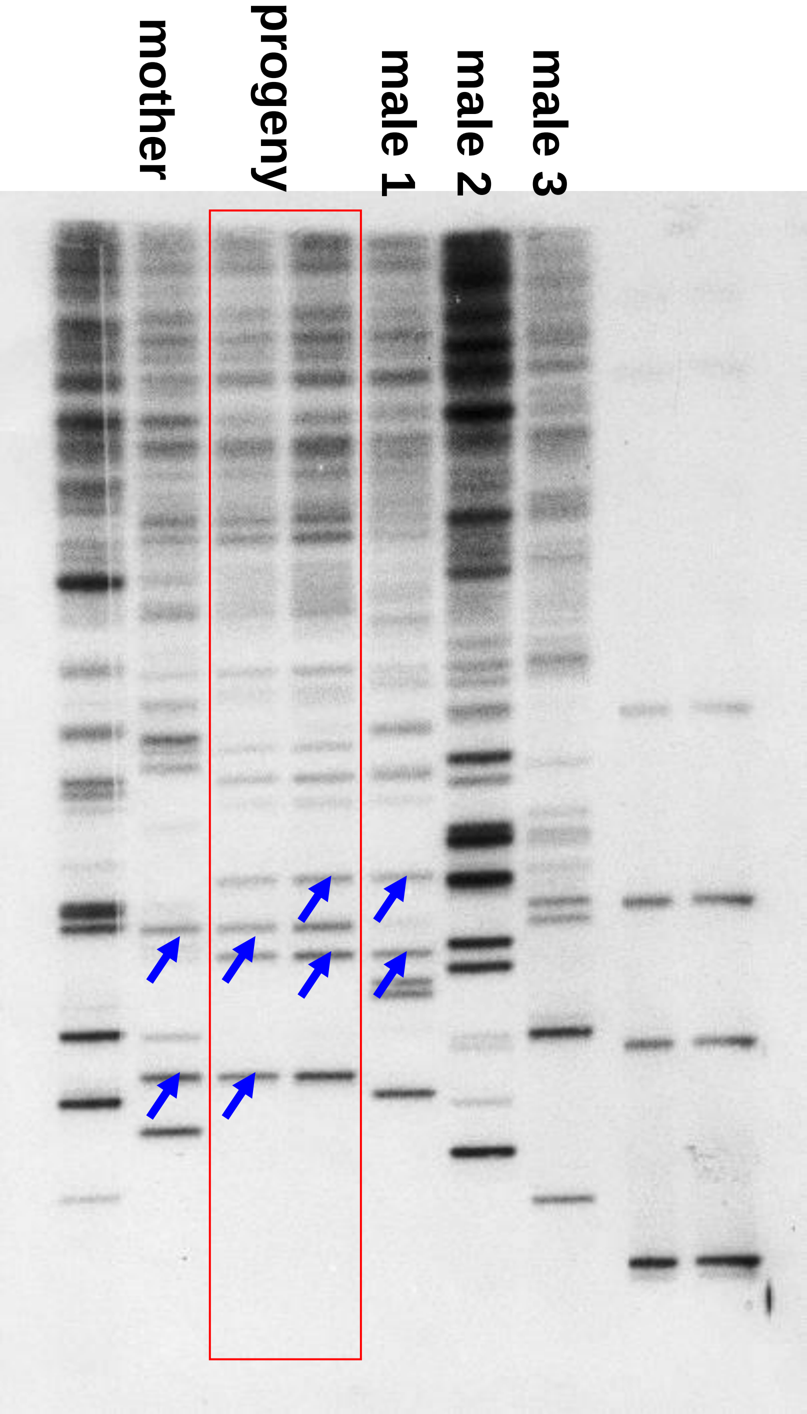
- Jeffreys et al. 1985
  - Restriction enzymes
  - Electrophoresis
  - Hybridization (minisatellite as a probe)

- High-quality DNA
- All samples on a single gel



DNA fingerprinting

→ percentage of shared bands

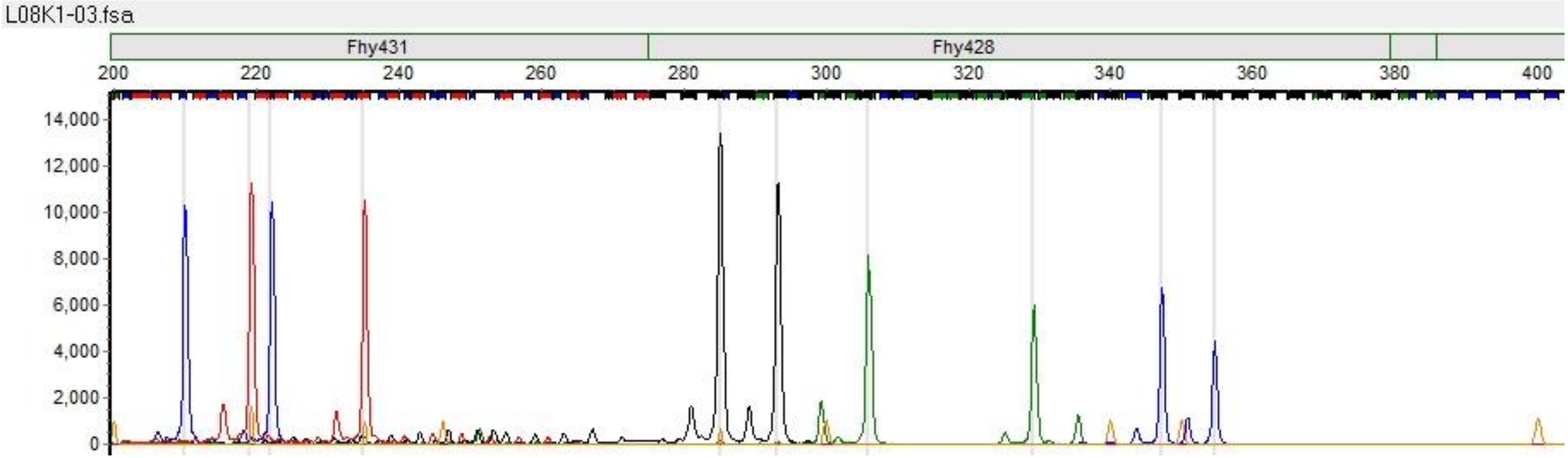


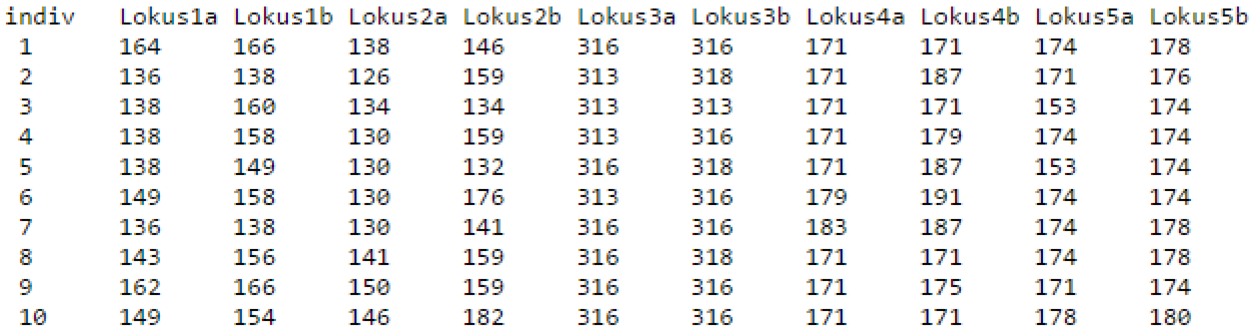
Microsatellites
Fragment analyses

generates a size estimate for DNA fragments relative to a size standard of DNA fragments with known lengths (the size standard).

CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT

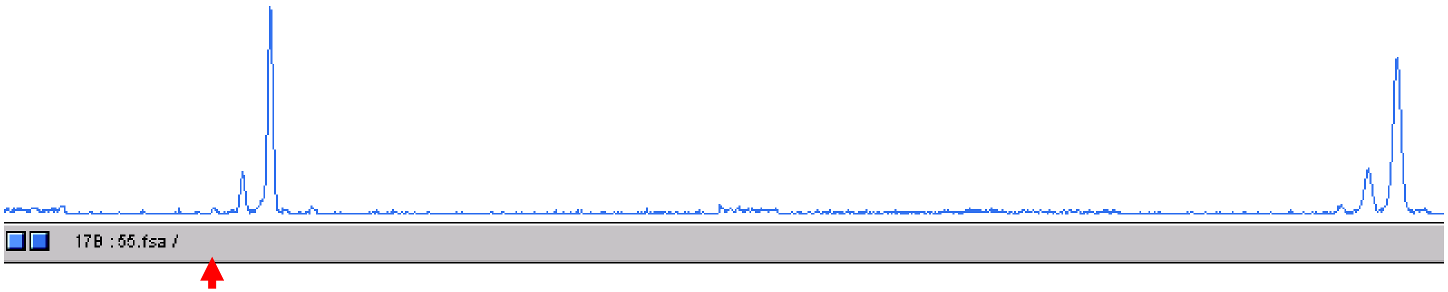
CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT





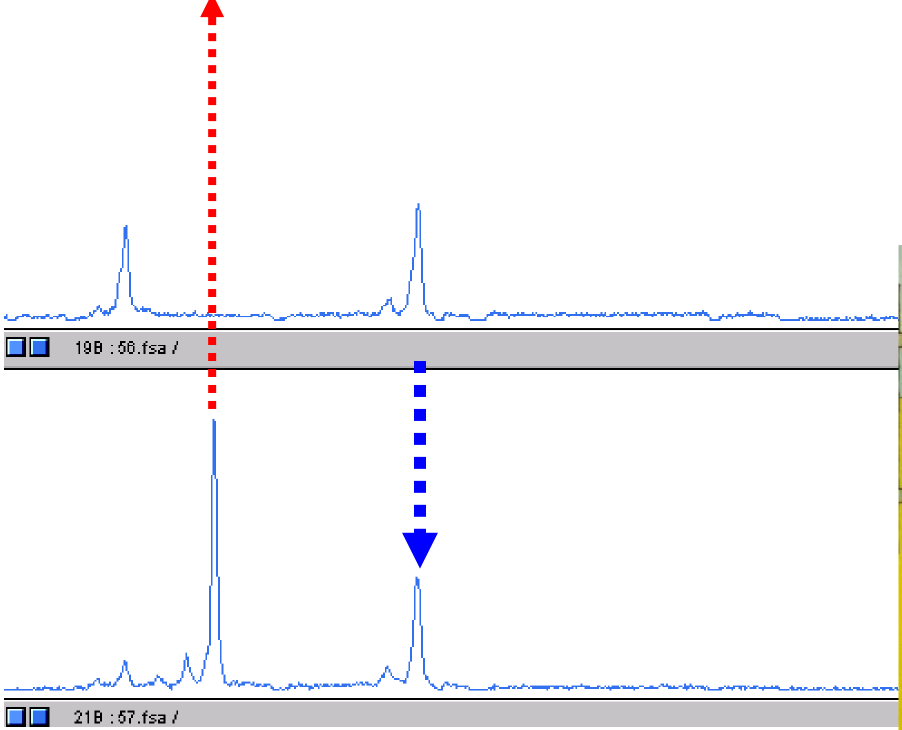
indiv	Lokus1a	Lokus1b	Lokus2a	Lokus2b	Lokus3a	Lokus3b	Lokus4a	Lokus4b	Lokus5a	Lokus5b
1	164	166	138	146	316	316	171	171	174	178
2	136	138	126	159	313	318	171	187	171	176
3	138	160	134	134	313	313	171	171	153	174
4	138	158	130	159	313	316	171	179	174	174
5	138	149	130	132	316	318	171	187	153	174
6	149	158	130	176	313	316	179	191	174	174
7	136	138	130	141	316	316	183	187	174	178
8	143	156	141	159	316	318	171	171	174	178
9	162	166	150	159	316	316	171	175	171	174
10	149	154	146	182	316	316	171	171	178	180

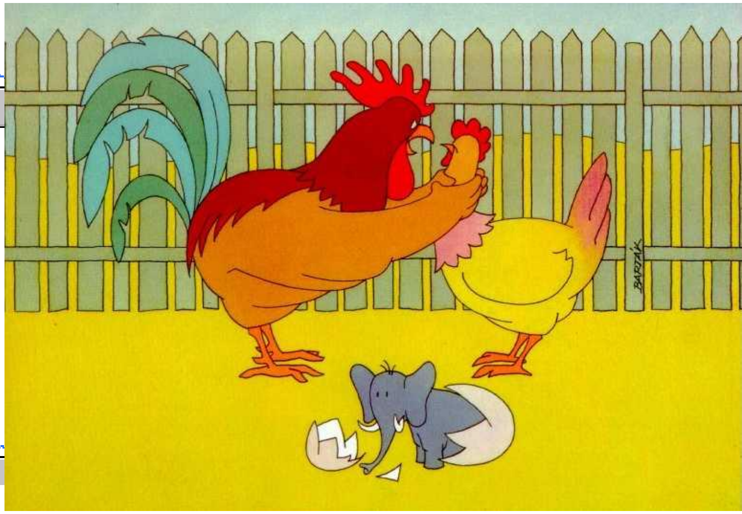
Male
(father?)



female

young





The horned passalus
Odontotaenius disjunctus
Dillard 2017

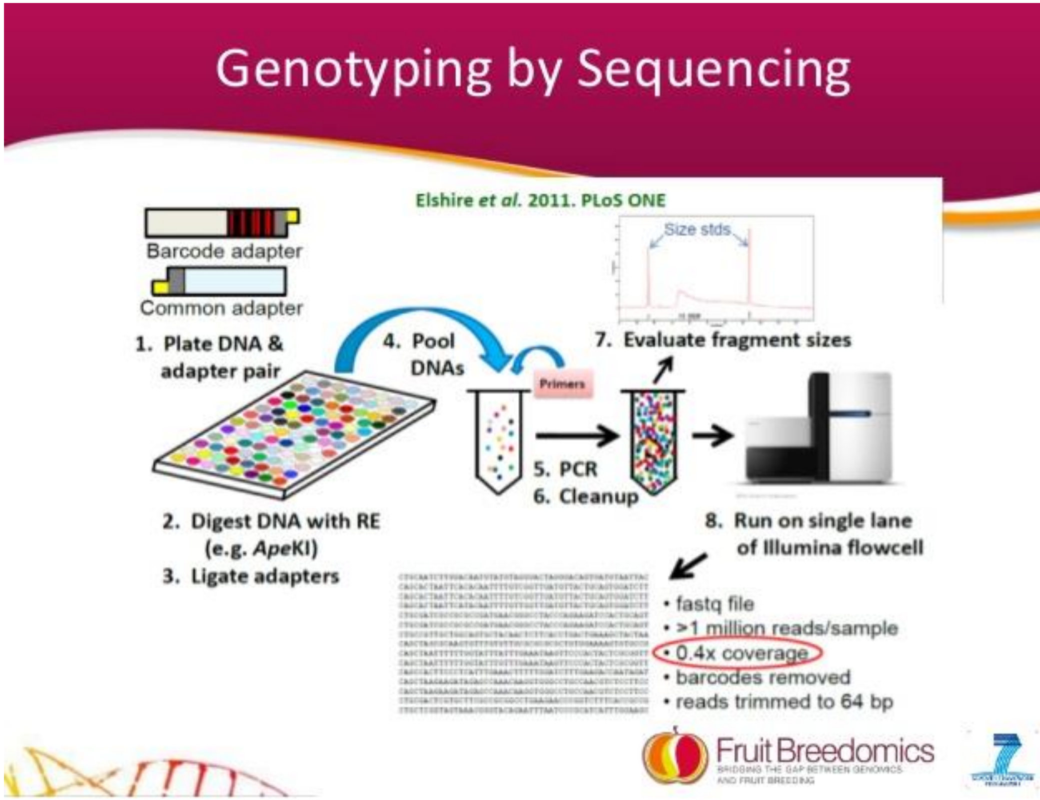
- Socially monogamous
- biparental care

- Genotyping-by-sequencing

- High levels of extra-pair paternity

- 54.8% of offspring sired by extra-pair males

- 70% of nests containing extra-pair young



Genotyping by Sequencing

Elshire *et al.* 2011. PLoS ONE

1. Plate DNA & adapter pair
2. Digest DNA with RE (e.g. *Ape*KI)
3. Ligate adapters
4. Pool DNAs
5. PCR
6. Cleanup
7. Evaluate fragment sizes
8. Run on single lane of Illumina flowcell

- fastq file
- >1 million reads/sample
- 0.4x coverage
- barcodes removed
- reads trimmed to 64 bp

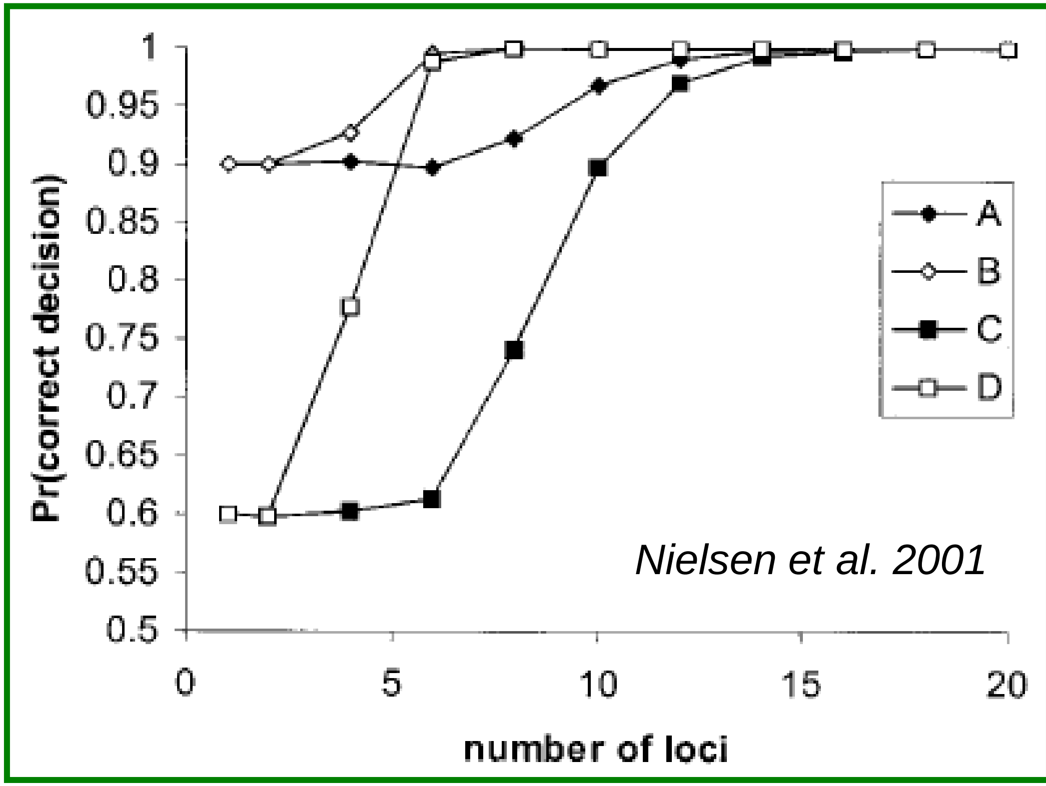
Fruit Breedomics
BRIDGING THE GAP BETWEEN GENOMICS AND FRUIT BREEDING

- Computer programs → automated assignment

Cervus
Colony
SOLOMON (R, Bayes)
Newpat
Probmax
Kinship
Famoz
Pasos
Papa
Parente
Patri
a others, see Jones et al. 2010, Flanagan and Jones 2019

How many loci?

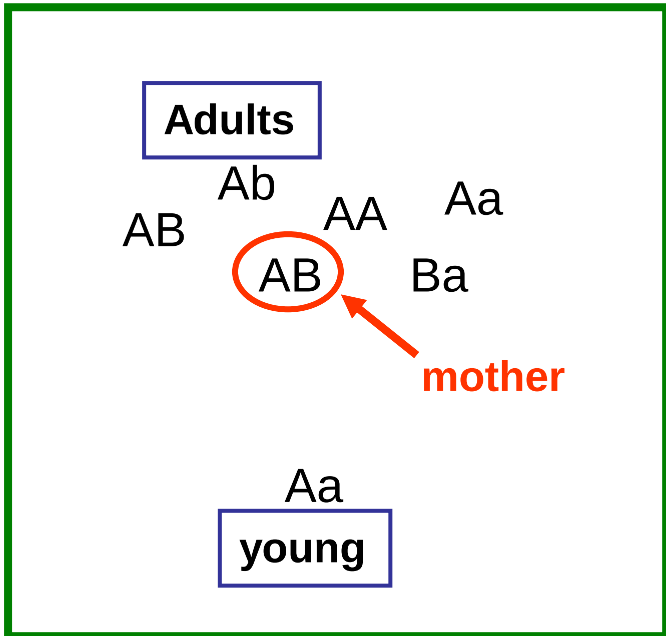
- Usually 6 or more
- Pilot study!
- **Exclusion probability, LD**



Nielsen et al. 2001

Exclusion probability

- measure of efficiency in paternity testing
- the probability of excluding an individual chosen by random
- non-exclusion probability (1 – Exclusion probability)
- Depends on loci
- 0.01 and less is fine (non-exclusion)
- Program **Cervus**
- First and second parent



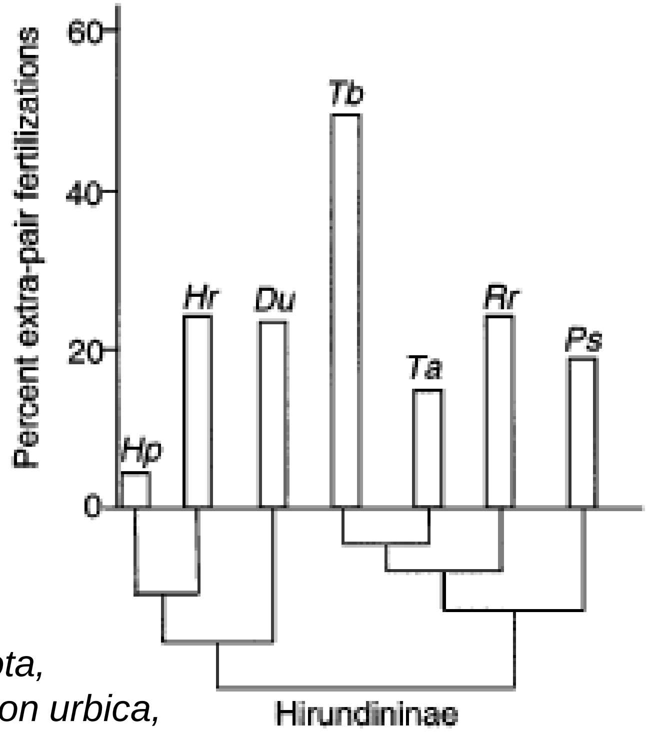
- Exclusion probability – rough estimate

- → simulations
  (including errors)

Birds

- Monogamous?
- EPC in 75 % bird species

Emberiza schoeniclus
55% EPY



Hirundo pyrrhonota,
H. rustica, Delichon urbica,
Tachycineta bicolor, T. albilinea,
Riparia riparia, Progne subis

Malurus cyaneus
72% EPY

Sperm storage

Painted turtle
Chrysemys picta

- *Chrysemys picta*
- msats
- → young with the same father in 3 years
- repeated mating with the same male is not probable
- → long sperm storage

Parental reconstruction

genotypes of a female and her litter

	locus 1		locus 2	
female	100	150	300	350
1. embryo	100	115	300	320
2. embryo	150	120	350	310
3. embryo	120	100	350	365

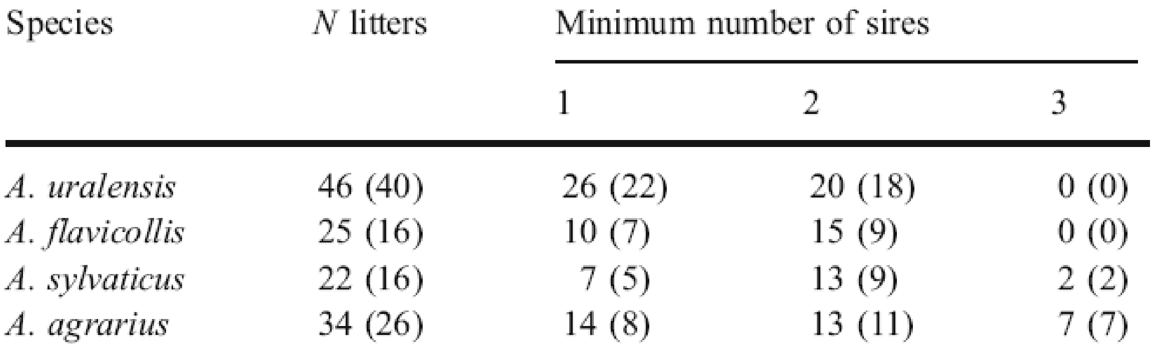
Maternal alleles

	locus 1		locus 2	
female	100	150	300	350
1. embryo	100	115	300	320
2. embryo	150	120	350	310
3. embryo	120	100	350	365

More than two paternal alleles
→ more fathers?

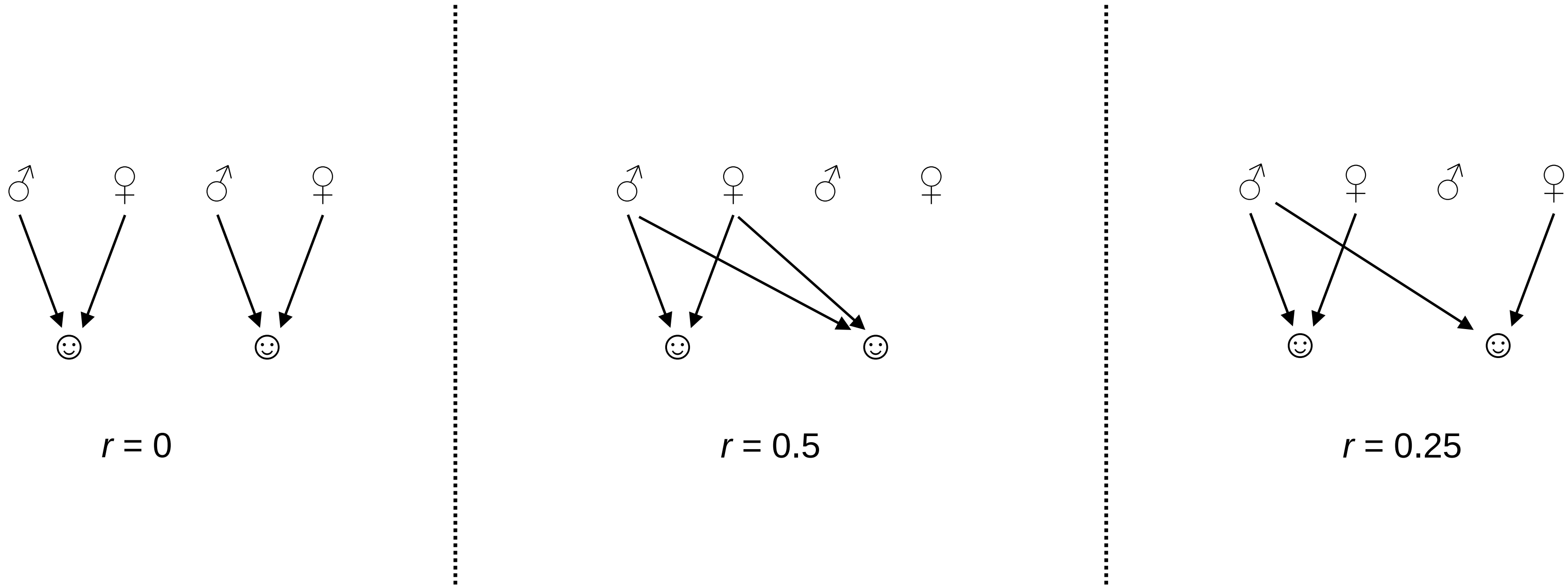
Apodemus agrarius a sylvaticus

(*Bryja et al. 2008*)

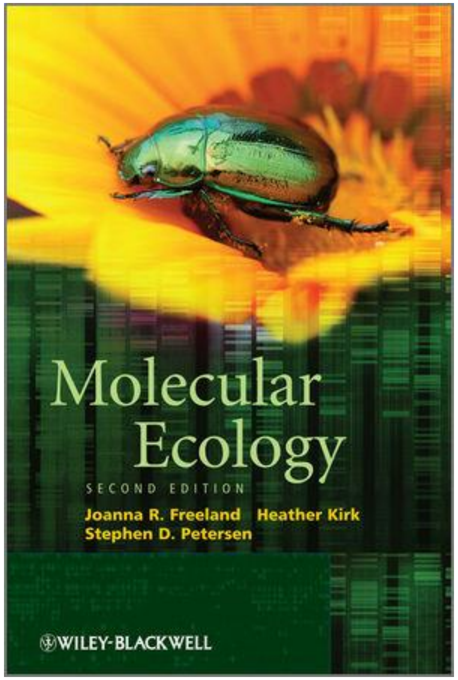


Species	*N* litters	Minimum number of sires		
		1	2	3
A. uralensis	46 (40)	26 (22)	20 (18)	0 (0)
A. flavicollis	25 (16)	10 (7)	15 (9)	0 (0)
A. sylvaticus	22 (16)	7 (5)	13 (9)	2 (2)
A. agrarius	34 (26)	14 (8)	13 (11)	7 (7)

Relatedness coefficient *r*



From Pedigrees
Estimates from genotypes



7

Behavioural Ecology

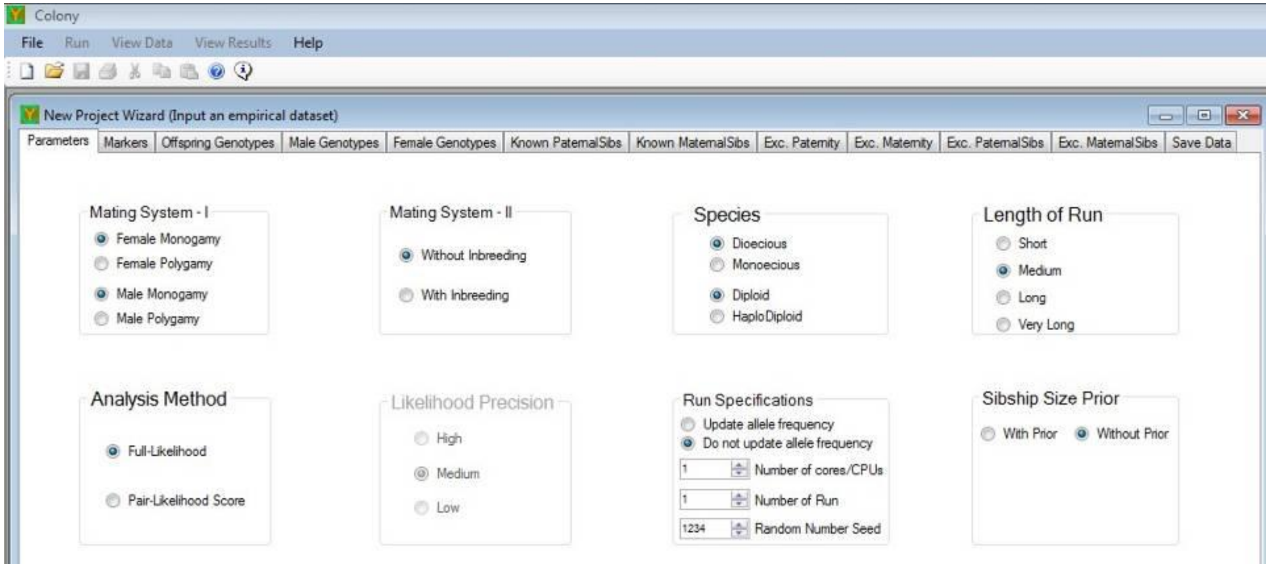
We already know how to estimate population allele frequencies, and the frequency of an allele in a diploid individual must be either 1.0 (homozygote), 0.5 (heterozygote) or 0 (allele absent). Based on this information, the relatedness of one individual to another can be calculated from allele frequency data as:

$$\frac{\sum(p_y - p)}{\sum(p_x - p)} \qquad (7.2)$$

in which for each allele, p is the frequency within the population, p_x is the frequency within the focal individual, and p_y is the frequency within the individual whose relationship to the focal individual we wish to know. Only those alleles that are found in the focal individual (x) are included in the equation (Queller and Goodnight, 1989).

Alternative approach

- Colony program
- Maximum likelihood method to assign individuals in a sample into full-sib families
- Stronger than „classical“ paternity approaches
- Larger litters



Mating System - I
● Female Monogamy
○ Female Polygamy
● Male Monogamy
○ Male Polygamy

Mating System - II
● Without Inbreeding
○ With Inbreeding

Species
● Dioecious
○ Monoecious
● Diploid
○ HaploDiploid

Length of Run
○ Short
● Medium
○ Long
○ Very Long

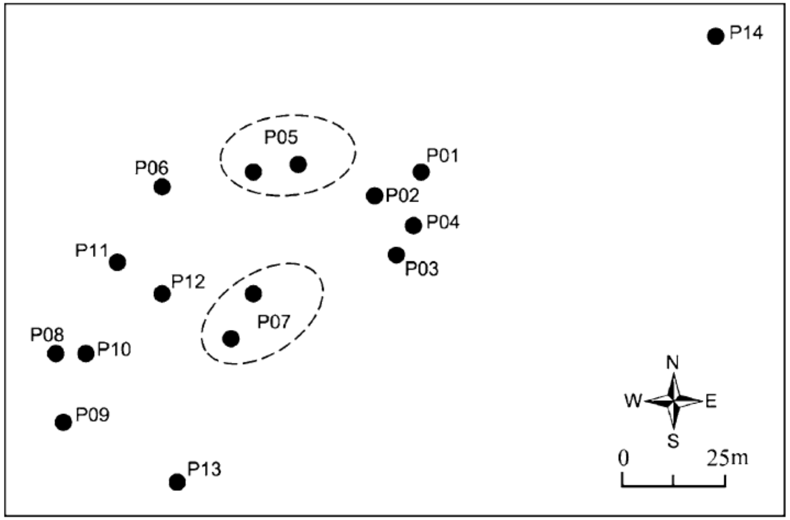
Analysis Method
● Full-Likelihood
○ Pair-Likelihood Score

Likelihood Precision
○ High
● Medium
○ Low

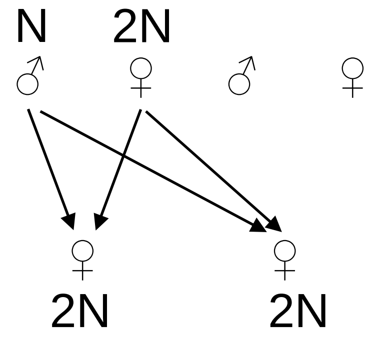
Run Specifications
○ Update allele frequency
● Do not update allele frequency
1 Number of cores/CPUs
1 Number of Run
1234 Random Number Seed

Sibship Size Prior
○ With Prior ● Without Prior

Qian et al. 2012 *Myrmecia pilosula*



- 319 workers (17 – 36 for a nest)
- 9 msats
- 1 - 4 queens per colony
- Queens mate with 1 – 9 males



***r* = 0.75**

Ducks

(Kreisinger et al. 2010)

- Non-invasive
- Feathers from nest
  → Genotype of mother
- Egg membrane
  → Genotypes of ducklings
- Nest parasitism
- Nests of parasitic females
- Renesting
- Reconstruction of father genotypes
- EPP

Altruism:

a behaviour that is costly to the actor and beneficial to the recipient.

Reciprocity

Kin selection

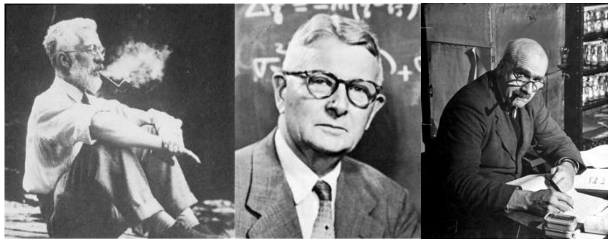
…

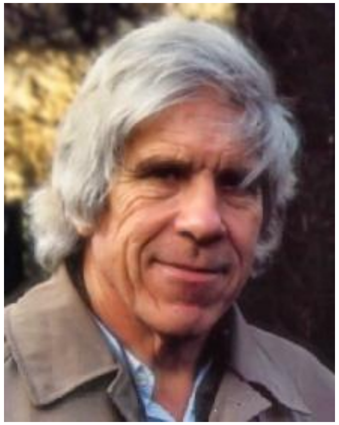
Photo: Badgers grooming (by Sue North) www.uksafari.com

Desmodus rotundus

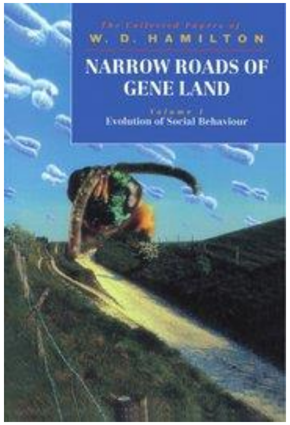
New (modern) synthesis:

Evolution = changes of allele frequencies over time

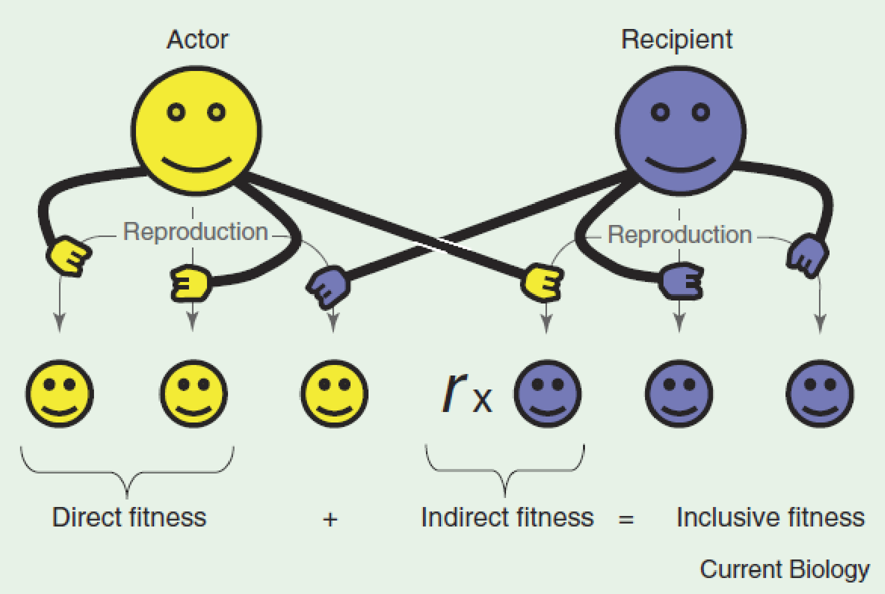




William Donald Hamilton



The Collected Papers of
W. D. HAMILTON
NARROW ROADS OF GENE LAND
Volume 1
Evolution of Social Behaviour



Current Biology

Inclusive fitness of an allele

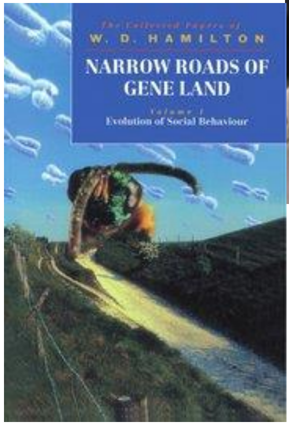
Inclusive fitness = Direct fitness + Indirect fitness
Direct fitness – fitness of the individual bearing the allele
Indirect fitness – fitness of other individuals carrying copies of the same allele

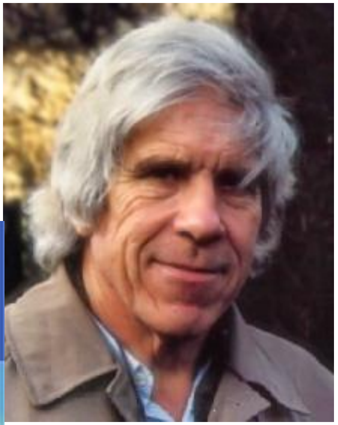
altruistic cooperation between relatives

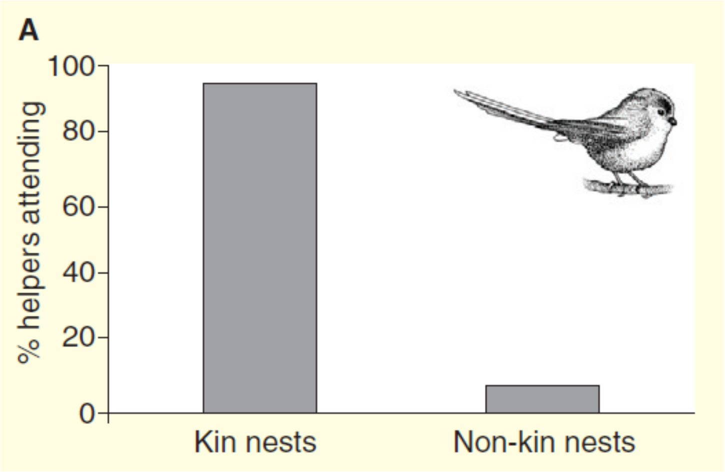
- **Hamilton’s rule**
- a behaviour or trait will be favoured by selection (altruism spreads), when

$rb - c > 0$

- c - the fitness cost to the actor
- b - the fitness benefit to the recipient
- r - their genetic relatedness
- r - fraction of genes that are identical by descent







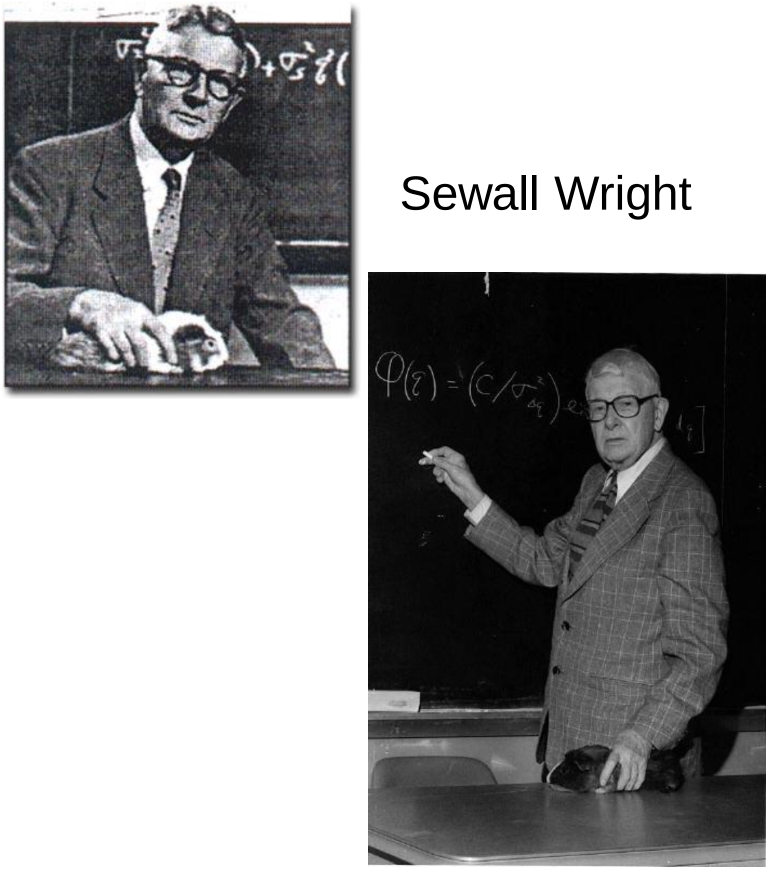
The long-tailed tit (*Aegithalos caudatus*)
Mlynařík dlouhoocasý
Russel & Hatchwell 2001

Ideal population

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

Ideal population = null model

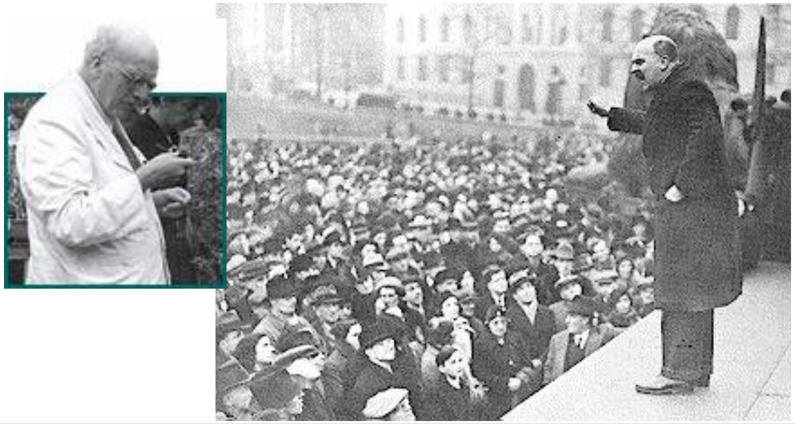
- Founders of the New synthesis
- Genetics, evolution, mathematics



Sewall Wright

Ronald
Fisher

John Burbon Sanderson Haldane



Ideal population
Fischer Wright model

- **Random mating** – depends on the trait
  - Sex chromosomes versus blood system
- **Nonoverlaping generations**

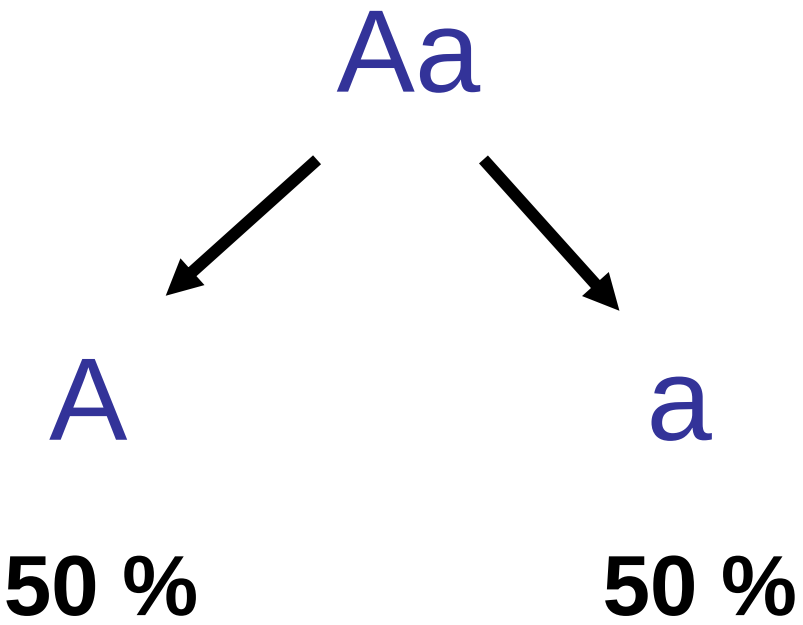
Assumptions

- *Random mating*
- *Nonoverlaping generations*
- **Mendel’s laws**
- Diploid (haploid)
- Sexual reproduction
- Two alleles (di-allelic model)
- Allele frequencies identical in males and females
- 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

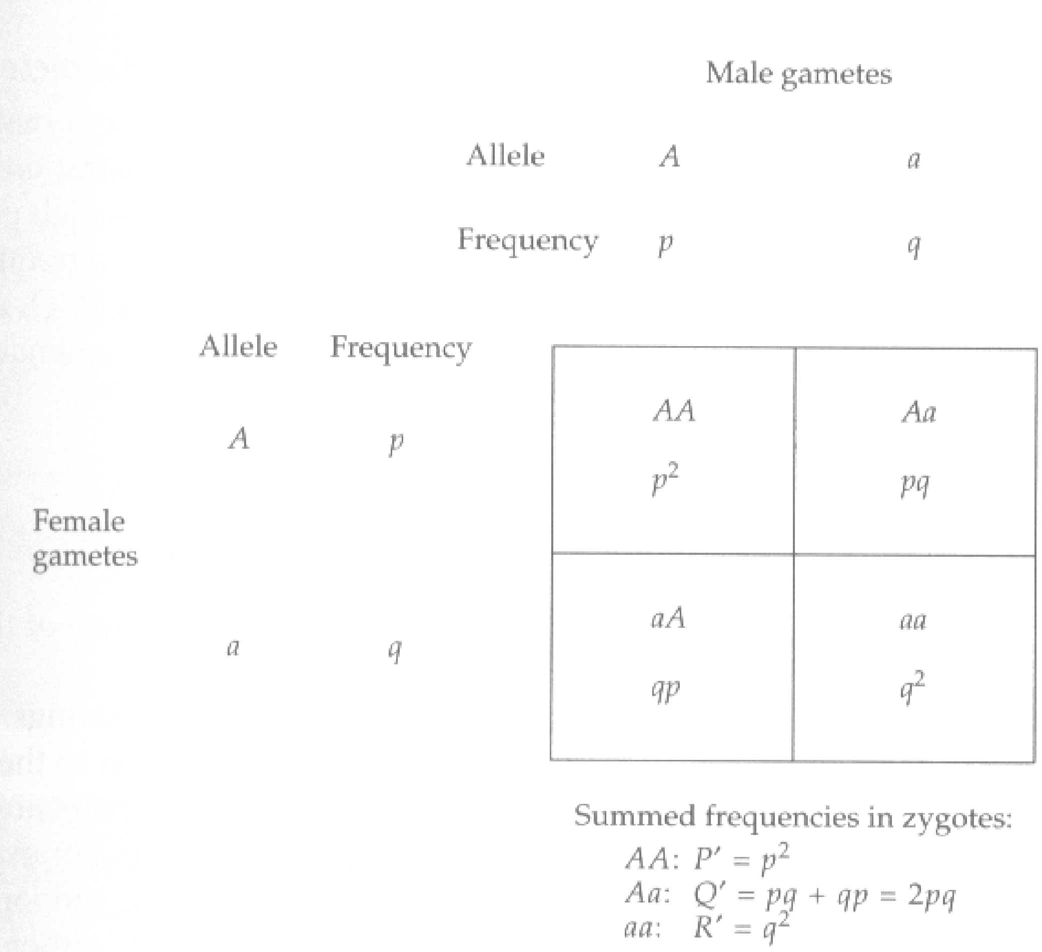
A

a

50 %

50 %

Hardy-Weinberg principle (equilibrium)



Male gametes

Female gametes

Allele	Frequency	A p	a q
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: $R' = q^2$

Joint probability

Genotypes assembled in each generation *de novo*

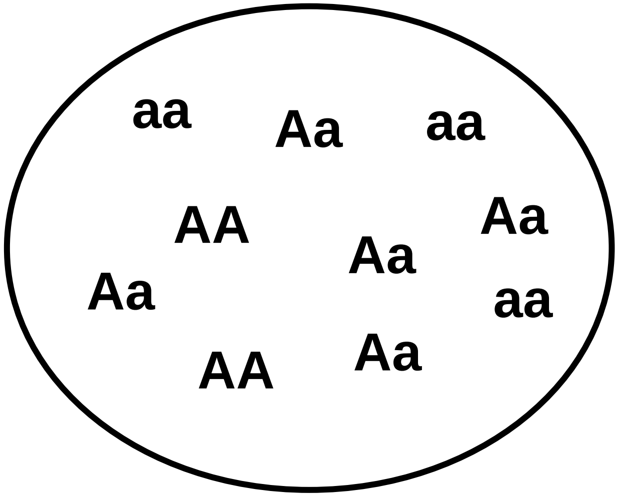
Relation of allele and genotype frequencies

AA: p^2 *Aa*: $2pq$ *aa*: q^2

Expected heterozygosity

$H_E = 2pq$

Established in one generation



- **Allele frequency**
- A 9/20 = 0.45 (p)
- a 11/20 = 0.55 (q)

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

- **Expected heterozygosity**
- $2pq$
- 2 * 0.45 * 0.55 = 0.495

Deviations from H-W

observed ≠ expected heterozygosity

- → **scoring errors**
  - Null alleles
  - Allele dropout
  - Duplications AA aa
- → **violations of assumptions**
  - **Decrease**
    - Selection against heterozygotes **underdominance**
    - Inbreeding
    - Positive-assortative mating
    - Population structure
  - **Increase**
    - Selection favoring heterozygotes **overdominance**
    - Outbreeding
    - Negative-assortative mating
    - Gene flow

Dispersal limitation → population structure

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

Demes – local interbreeding units

(Subpopulations, local populations)

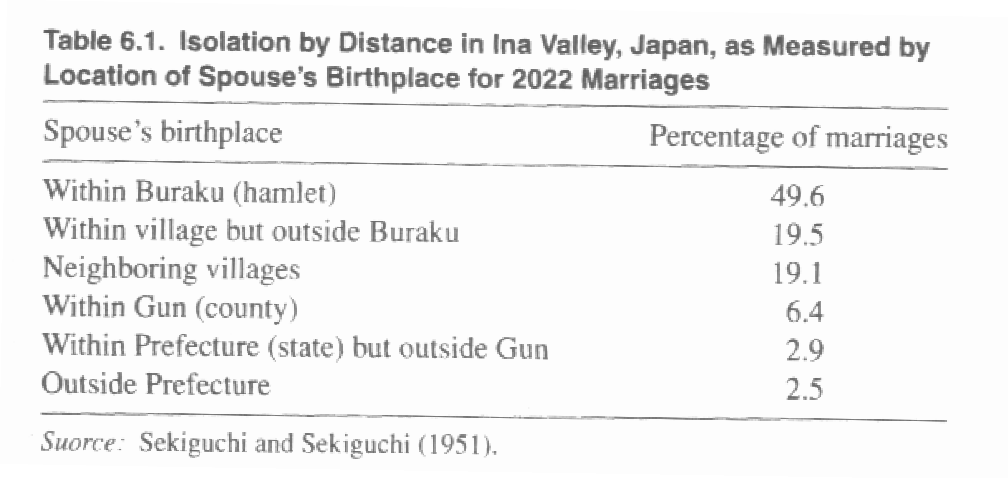
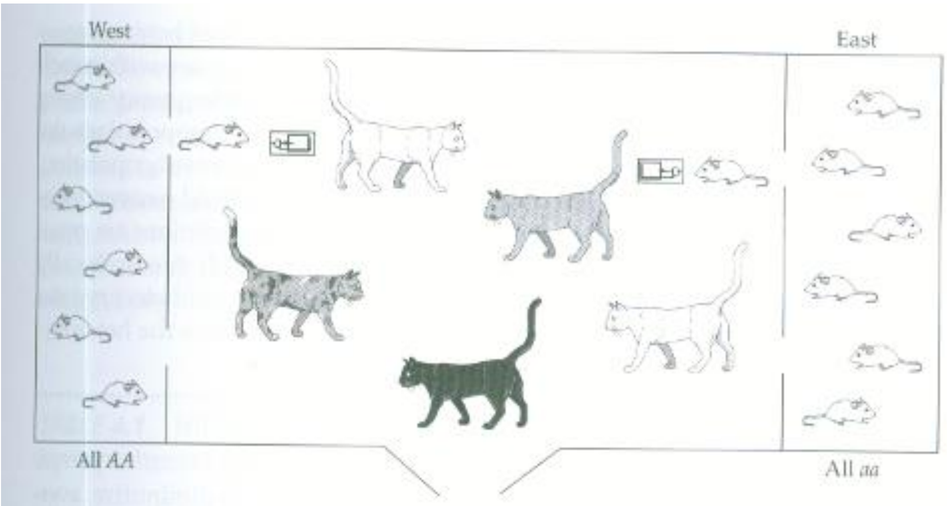
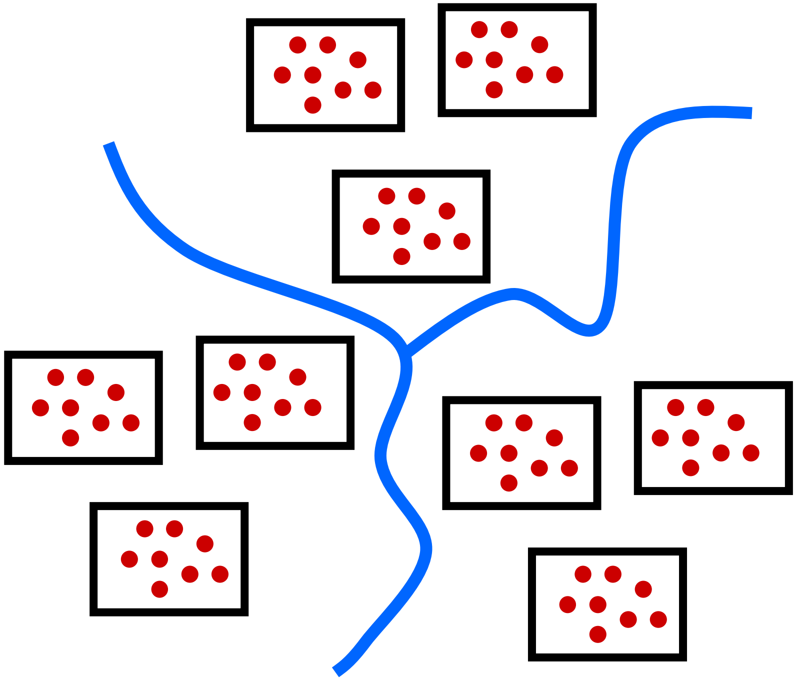


Table 6.1. Isolation by Distance in Ina Valley, Japan, as Measured by Location of Spouse's Birthplace for 2022 Marriages

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

Suorce: Sekiguchi and Sekiguchi (1951).

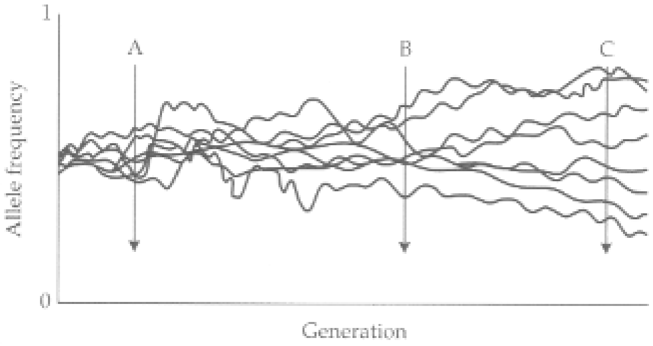


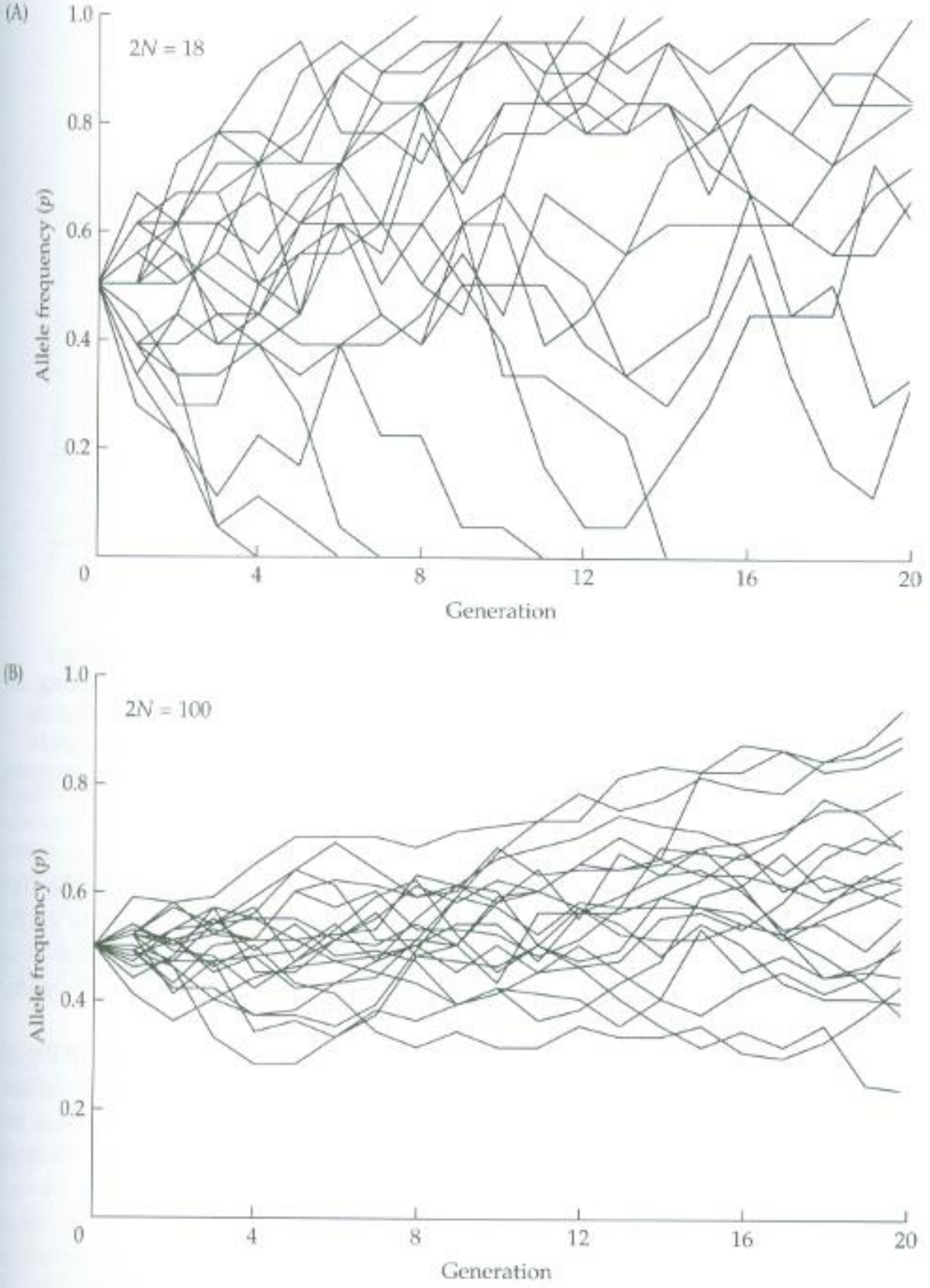


Differentiation of populations

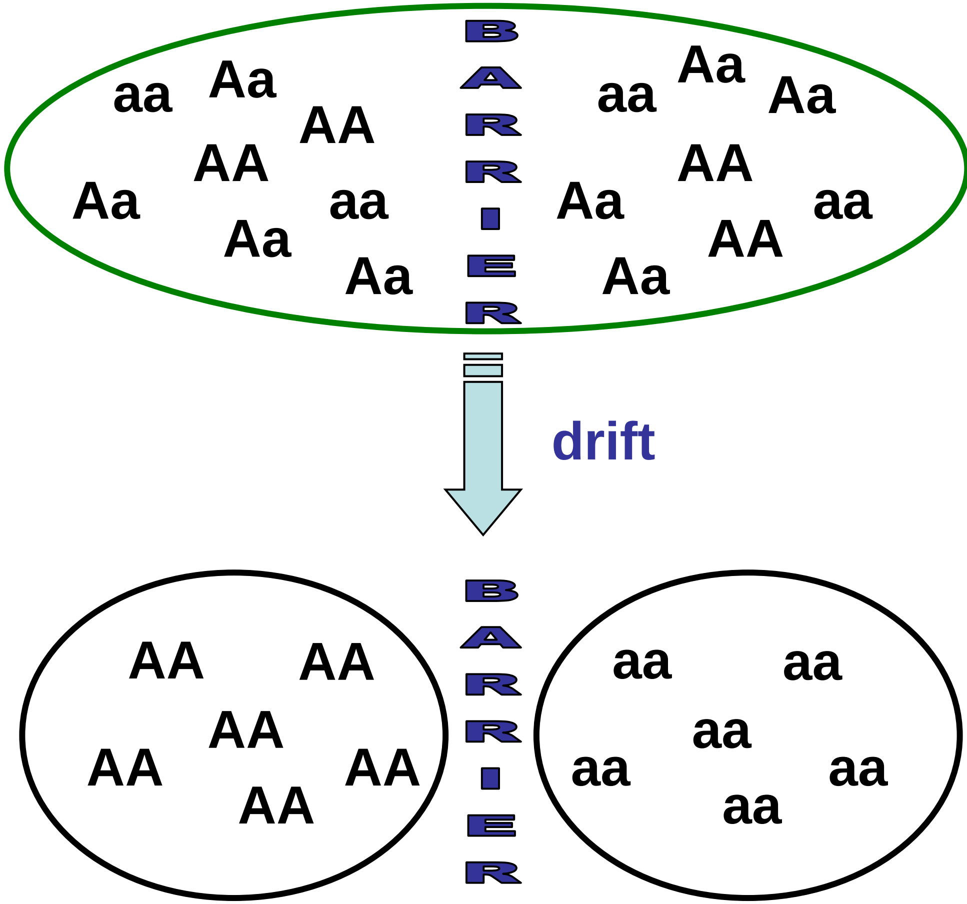
- Drift
- Selection
- Mutations

drift


1
A
B
C
Allele frequency
0
Generation

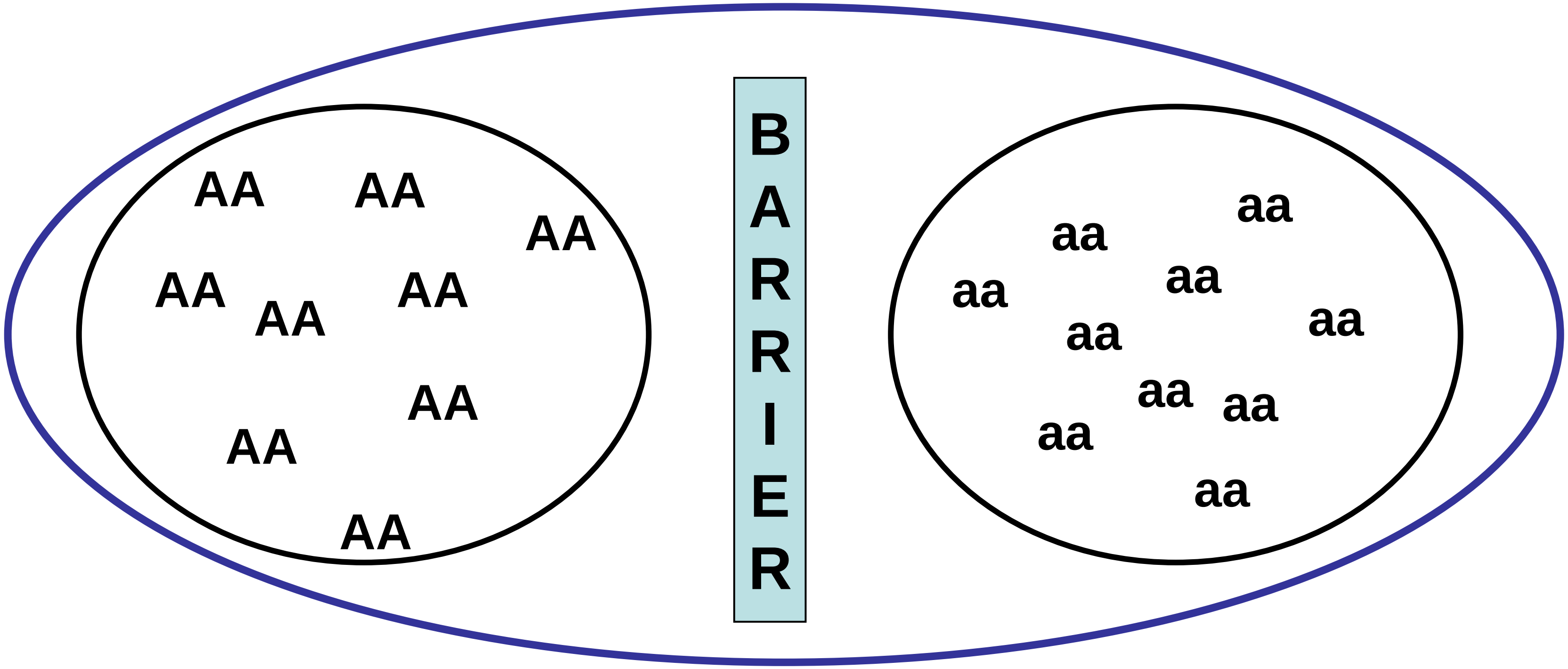

(A)
1.0
2N = 18
0.8
0.6
0.4
0.2
0
Allele frequency (p)
4
8
12
16
20
Generation
(B)
1.0
2N = 100
0.8
0.6
0.4
0.2
0
Allele frequency (p)
4
8
12
16
20
Generation

- **Drift**
  → extreme example
  (fixation of alternative alleles)



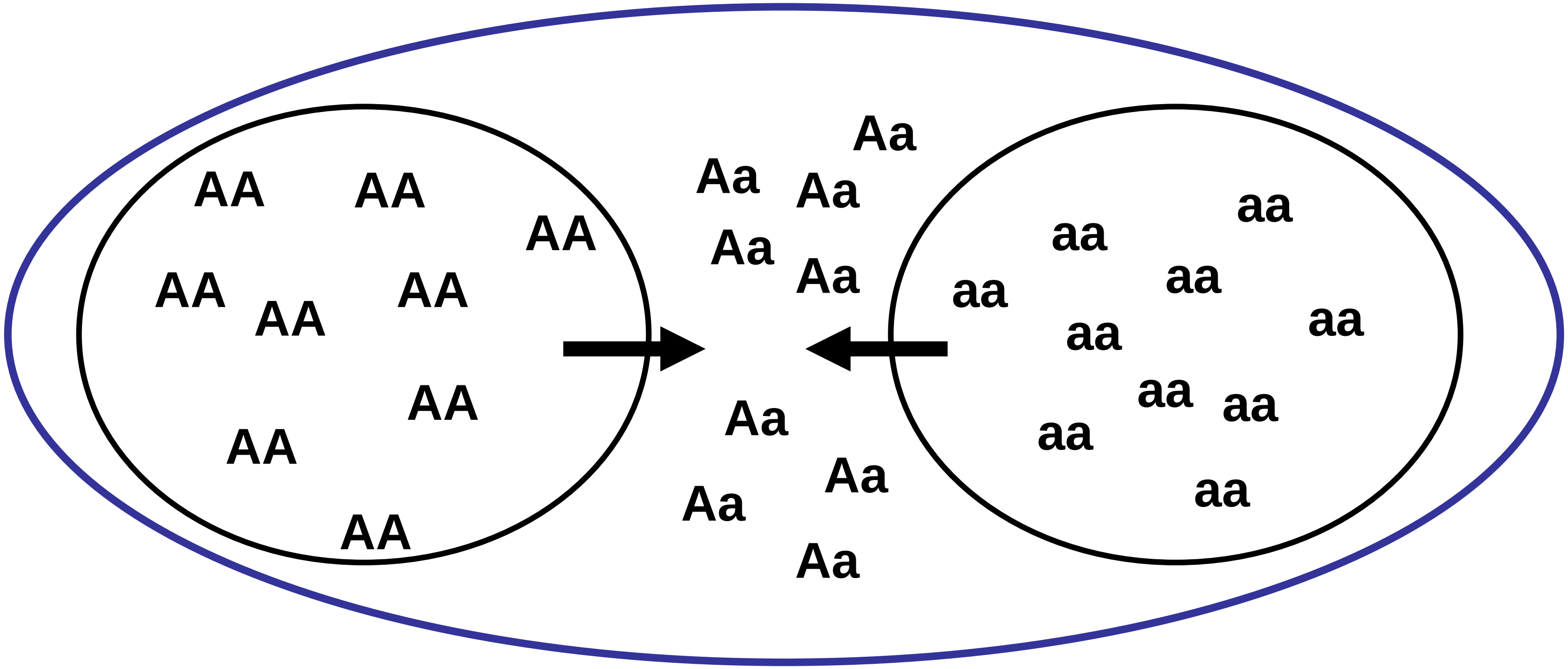
heterozygosity

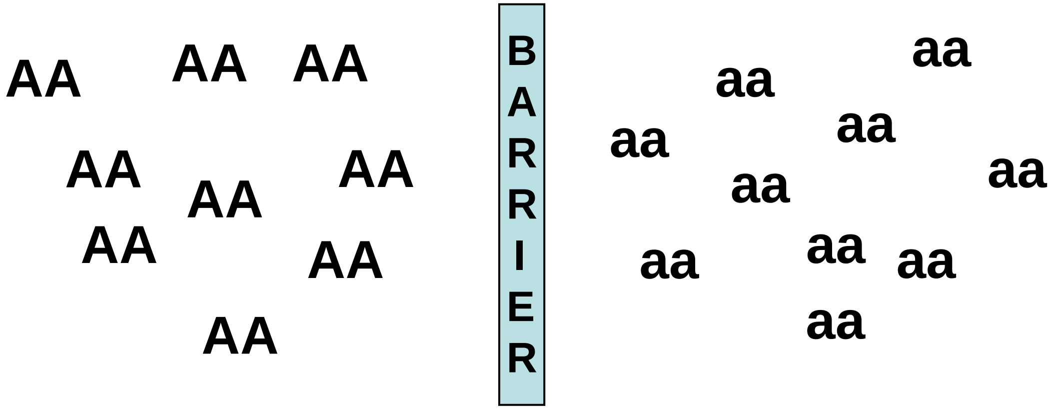
- Extreme example – fixed alleles
- Subpopulations in HW,
- total population – deficit of heterozygotes



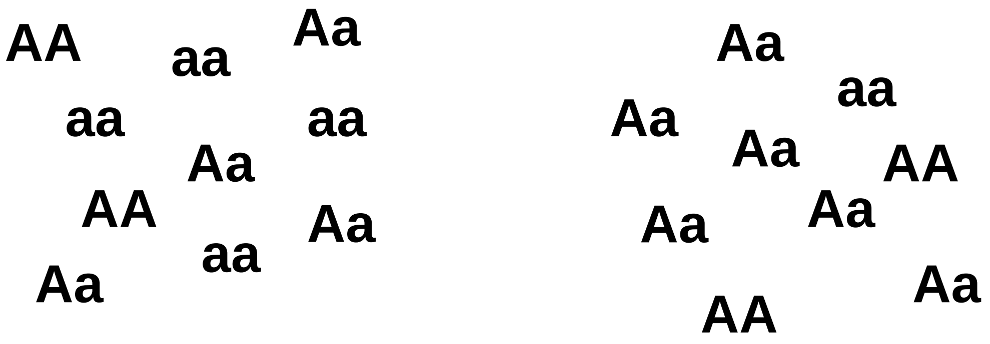
Wahlund effect (principle) = isolate breaking

- Increase of heterozygosity





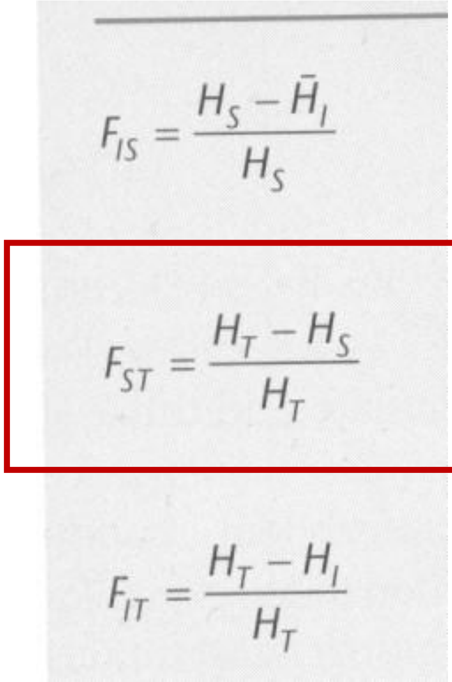
Isolated subpopulations (small gene flow)
Deviations from HW in total pop
Smaller dev in subpops



Mixed subpops (high gene flow)
Small dev from HW in total pop
Similar dev in subpops

F-statistics

- 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} **Inbreeding in subpops**

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

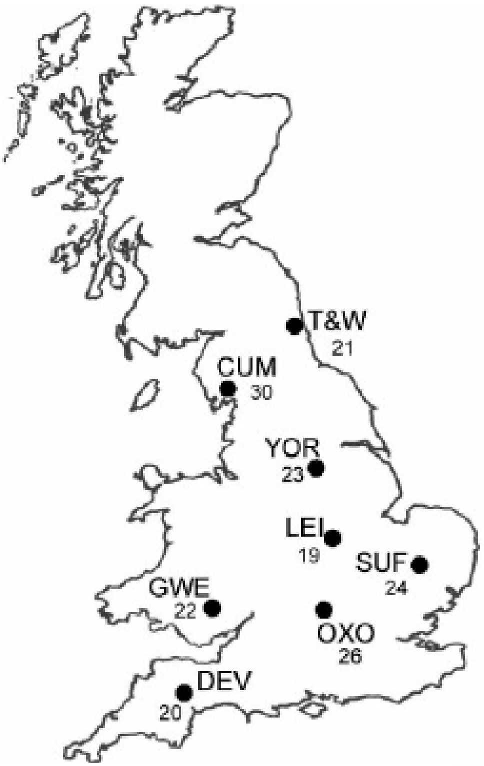
F_{ST} **Population structure**

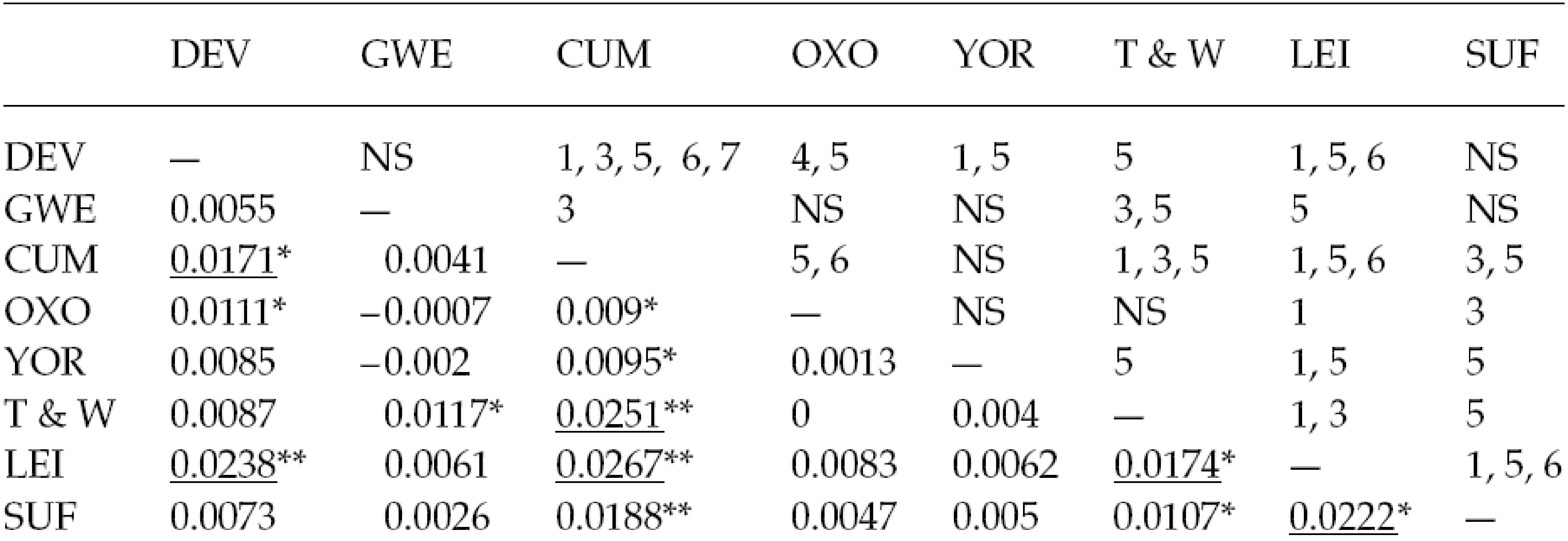
$$F_{IT} = \frac{H_T - H_I}{H_T}$$

F_{IT} **Total population**

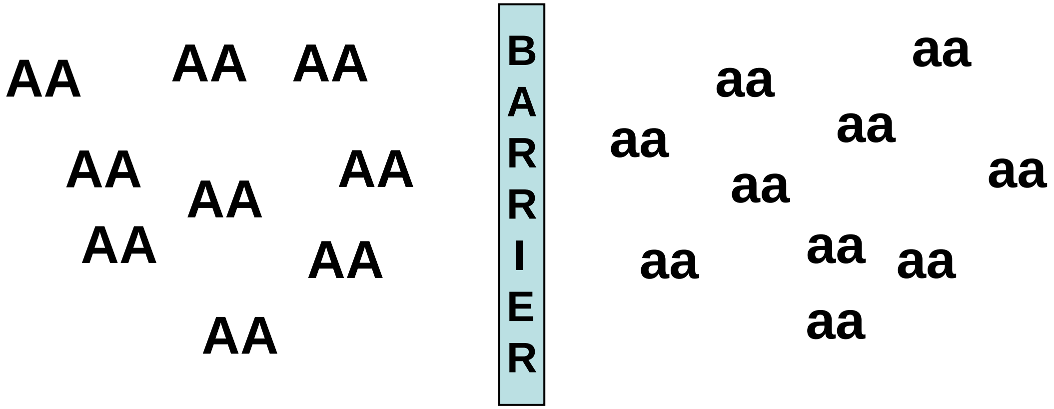
Emberiza citrinella

Lee et al. 2001



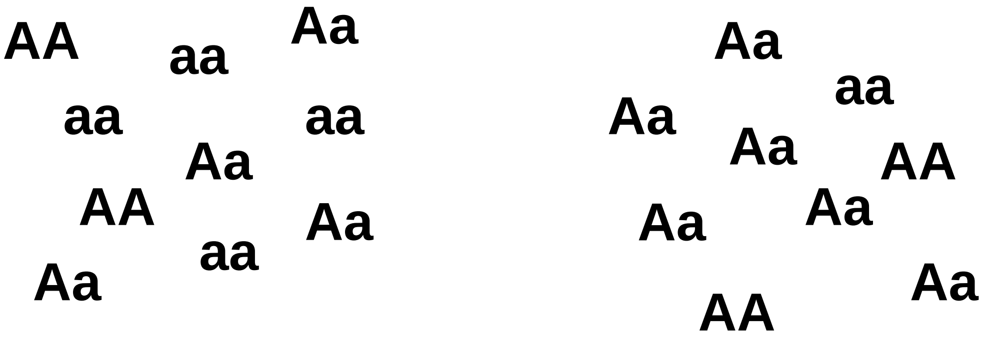


	DEV	GWE	CUM	OXO	YOR	T & W	LEI	SUF
DEV	—	NS	1, 3, 5, 6, 7	4, 5	1, 5	5	1, 5, 6	NS
GWE	0.0055	—	3	NS	NS	3, 5	5	NS
CUM	0.0171*	0.0041	—	5, 6	NS	1, 3, 5	1, 5, 6	3, 5
OXO	0.0111*	–0.0007	0.009*	—	NS	NS	1	3
YOR	0.0085	–0.002	0.0095*	0.0013	—	5	1, 5	5
T & W	0.0087	0.0117*	0.0251**	0	0.004	—	1, 3	5
LEI	0.0238**	0.0061	0.0267**	0.0083	0.0062	0.0174*	—	1, 5, 6
SUF	0.0073	0.0026	0.0188**	0.0047	0.005	0.0107*	0.0222*	—



Alternative:
Variance
instead of
deviations from HW

Isolated subpopulations (small gene flow)
Some variance in total pop
Smaller variance in subpops



Mixed subpops (high gene flow)
Similar variance in subpops and total pop

Myotis bechsteinii

Kerth et al. 2002

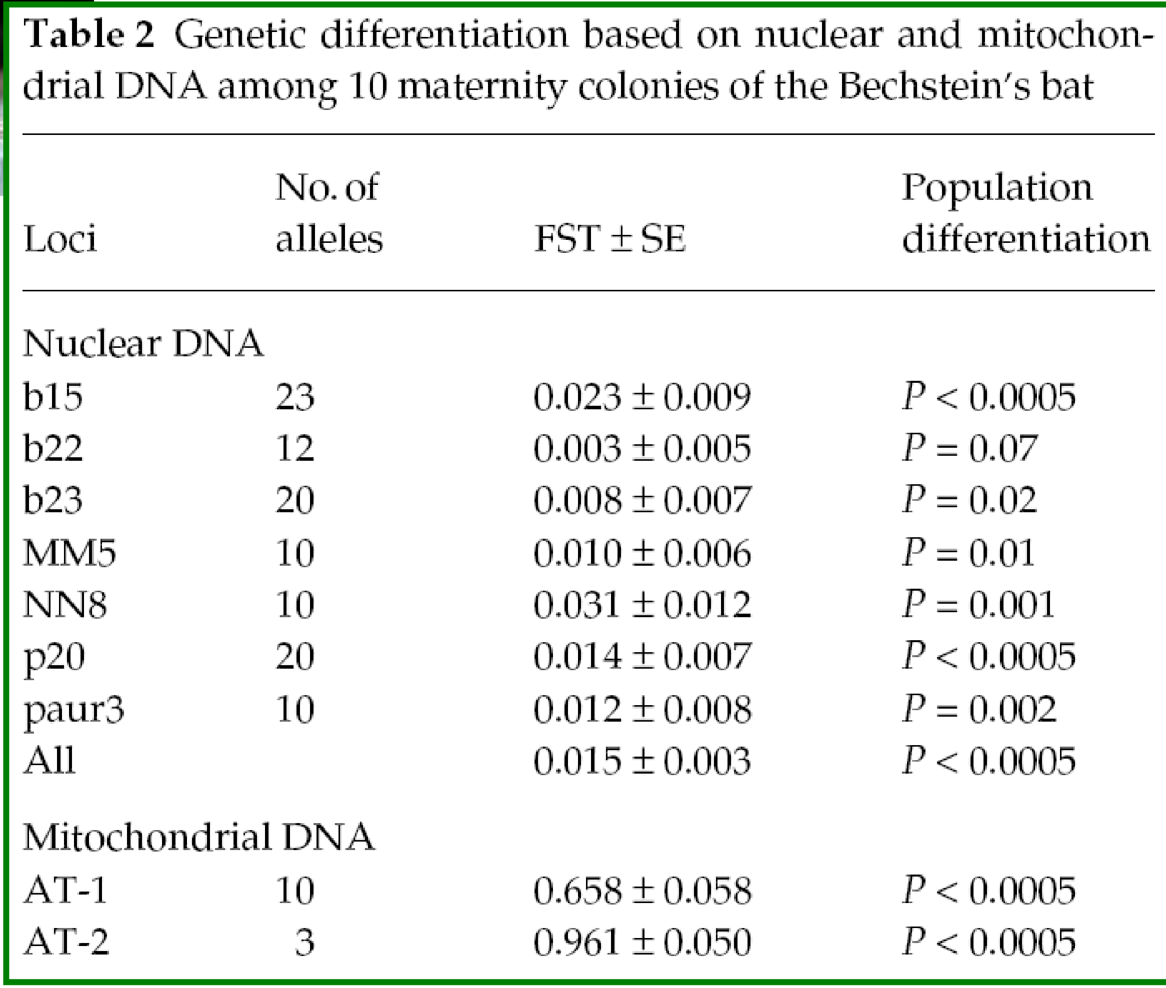
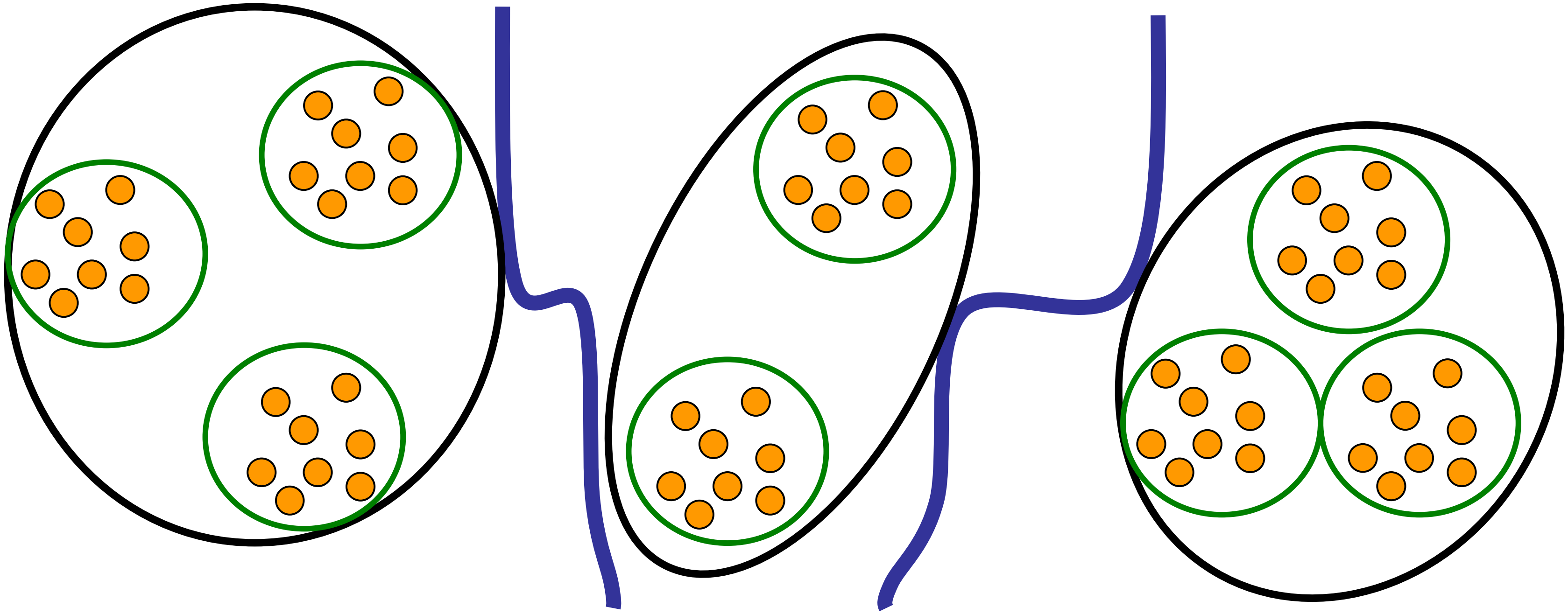


Table 2 Genetic differentiation based on nuclear and mitochondrial DNA among 10 maternity colonies of the Bechstein's bat

Loci	No. of alleles	FST ± SE	Population differentiation
Nuclear DNA			
b15	23	0.023 ± 0.009	$P < 0.0005$
b22	12	0.003 ± 0.005	$P = 0.07$
b23	20	0.008 ± 0.007	$P = 0.02$
MM5	10	0.010 ± 0.006	$P = 0.01$
NN8	10	0.031 ± 0.012	$P = 0.001$
p20	20	0.014 ± 0.007	$P < 0.0005$
paur3	10	0.012 ± 0.008	$P = 0.002$
All		0.015 ± 0.003	$P < 0.0005$
Mitochondrial DNA			
AT-1	10	0.658 ± 0.058	$P < 0.0005$
AT-2	3	0.961 ± 0.050	$P < 0.0005$

Hierarchical AMOVA



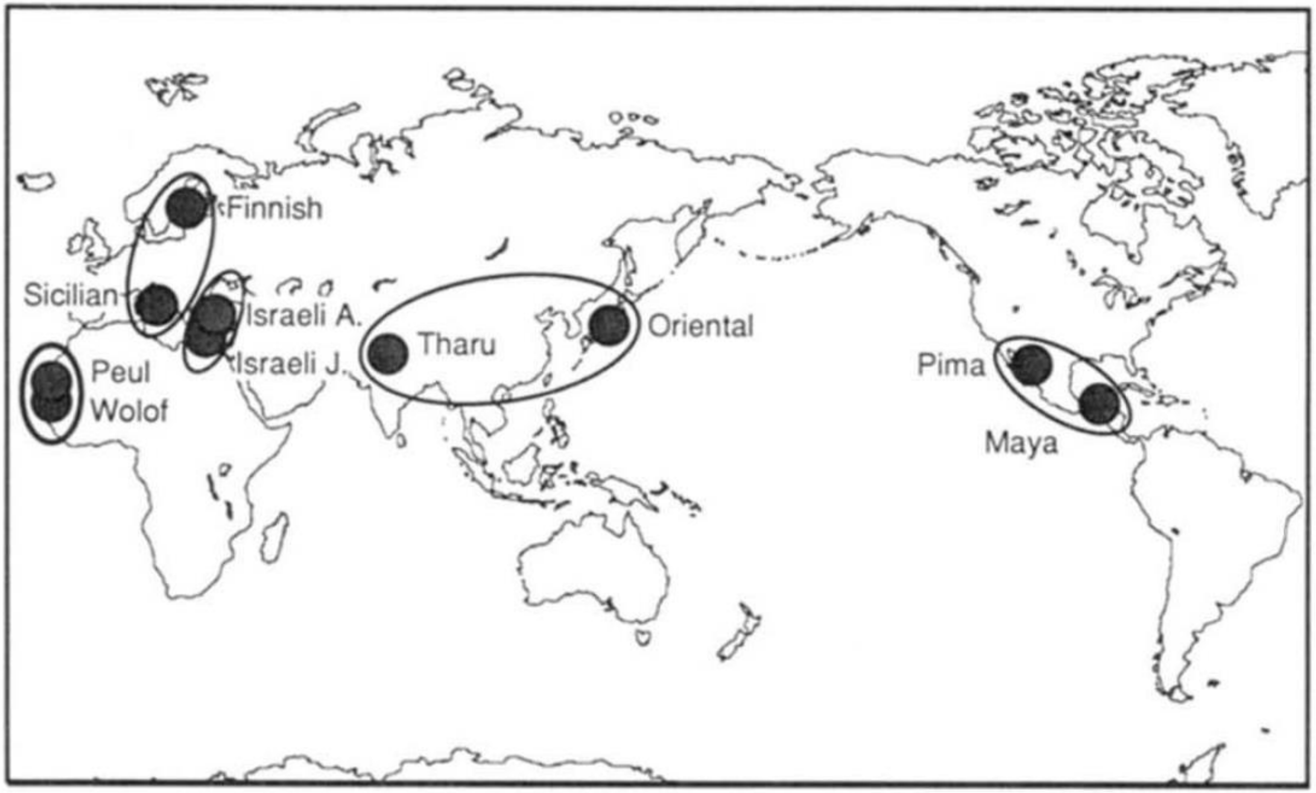
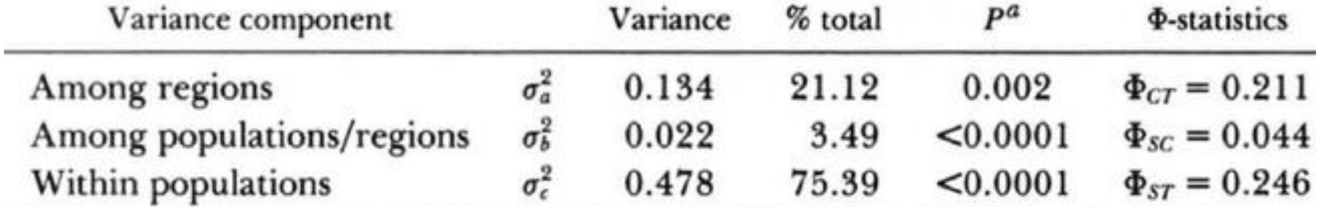


FIGURE 2.—Geographic location of the population samples.

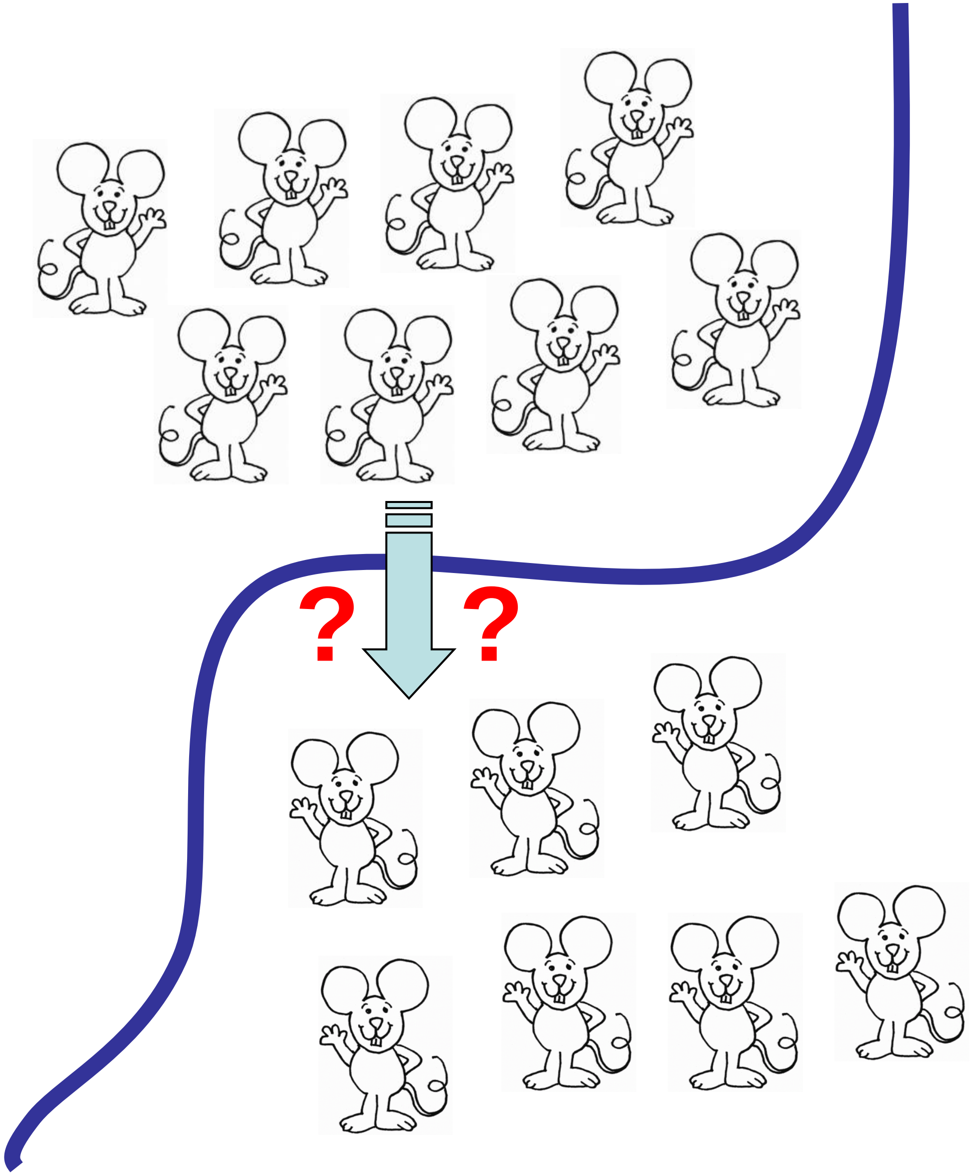


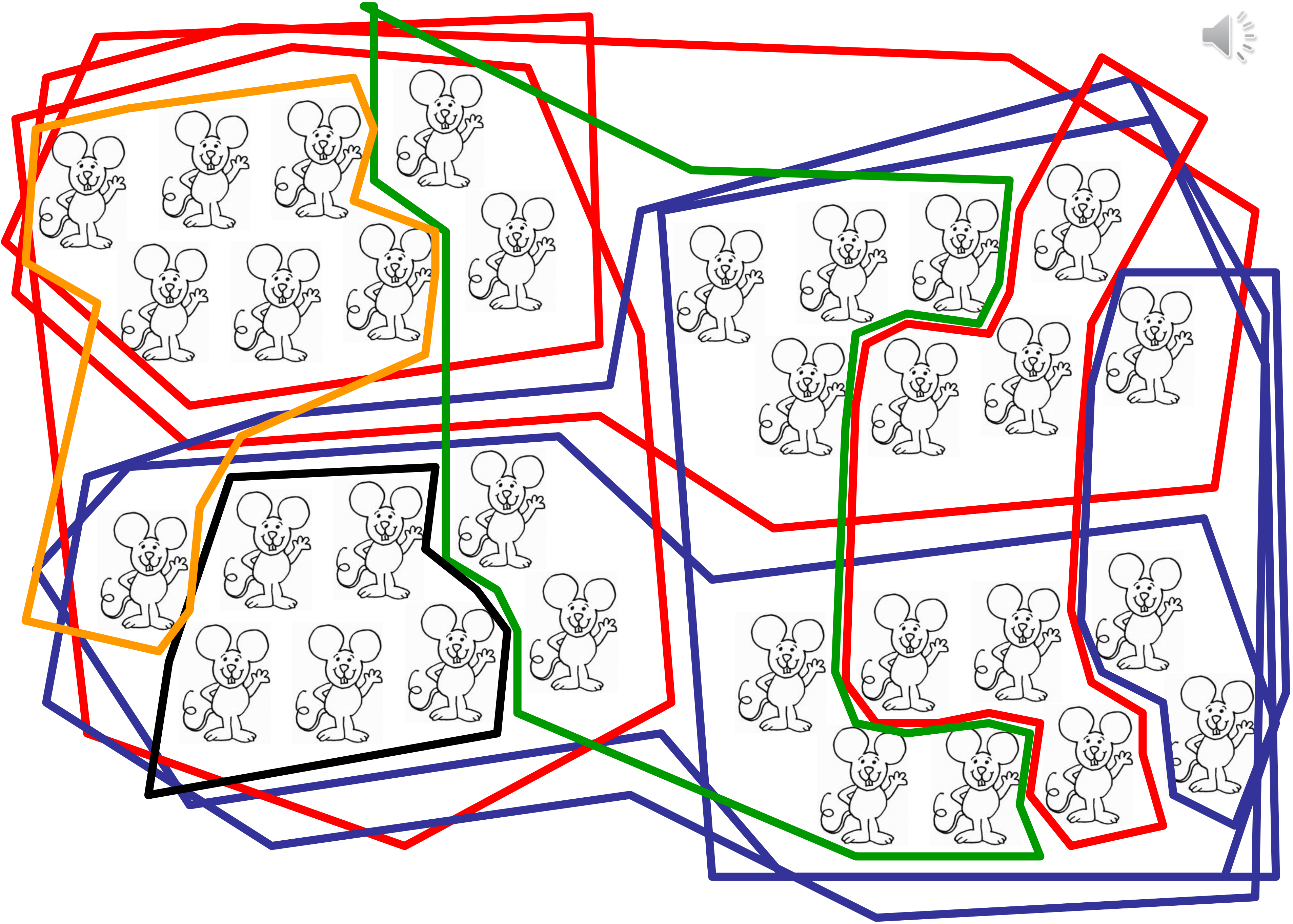
Variance component		Variance	% total	P^a	Φ-statistics
Among regions	σ_a^2	0.134	21.12	0.002	$\Phi_{CT} = 0.211$
Among populations/regions	σ_b^2	0.022	3.49	<0.0001	$\Phi_{SC} = 0.044$
Within populations	σ_c^2	0.478	75.39	<0.0001	$\Phi_{ST} = 0.246$

Excoffier et al. 1992

Traditional approaches

- pre-defined groups
- Allele frequencies
- Subgroups in HW





- Cryptic population structure
- Unknown number of clusters (K)
- Individual genotypes and coordinates
- Model-based clustering
  (Bayesian clustering –Structure)

Group 1		AA	bb	cc	dd
		AA	Bb	cc	DD
		AA	bb	cc	dd
Group 2		aa	BB	CC	DD
		aa	Bb	CC	dd

Group 1

AA bb cc dd

Group 2

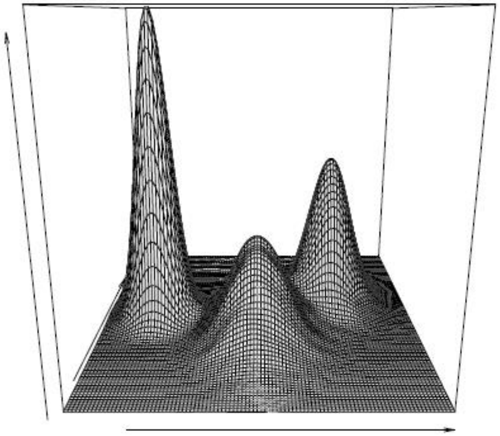
AA Bb cc DD

AA bb cc dd

aa BB CC DD

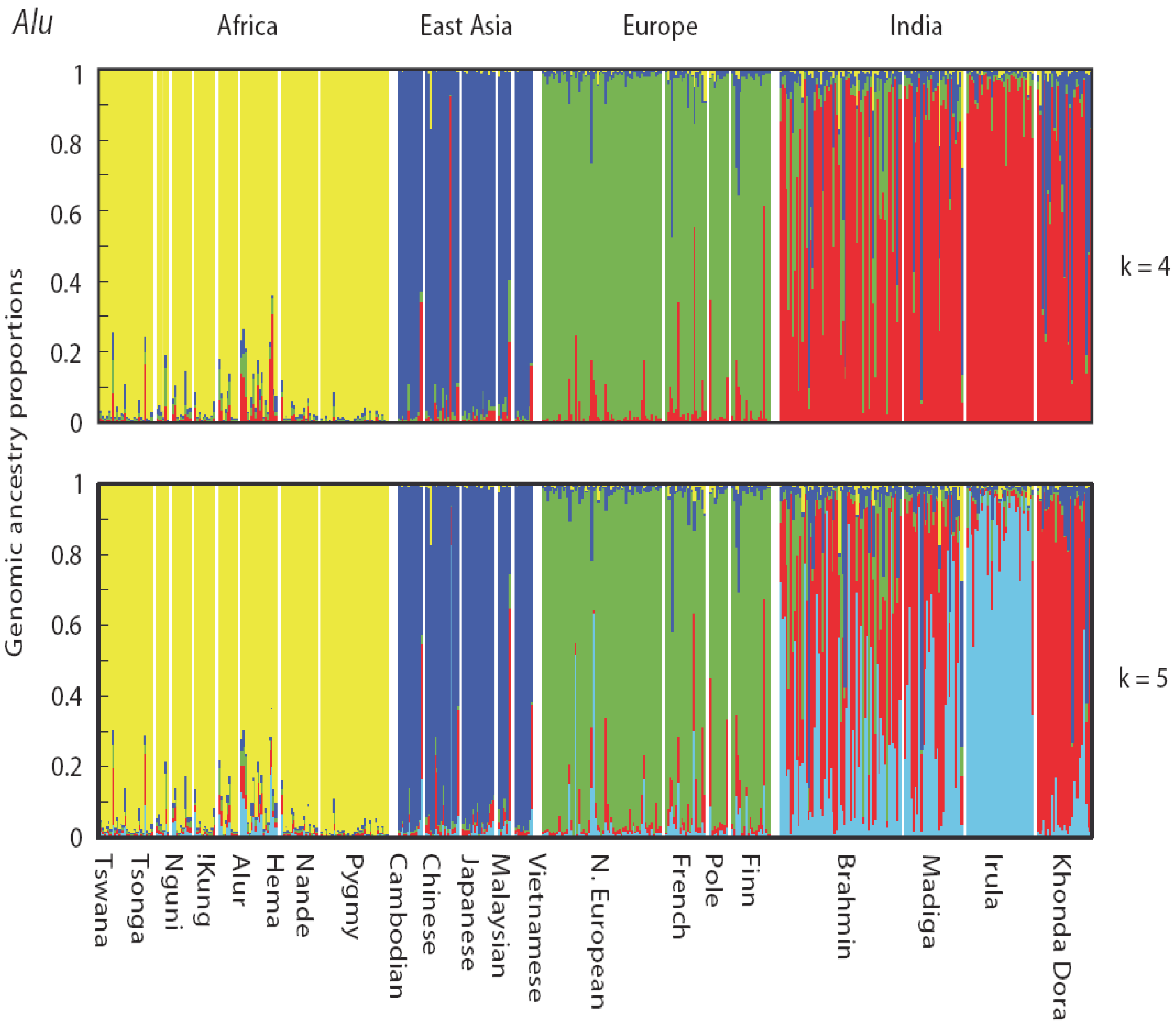
aa Bb CC dd

The number of partitions as a function of n.



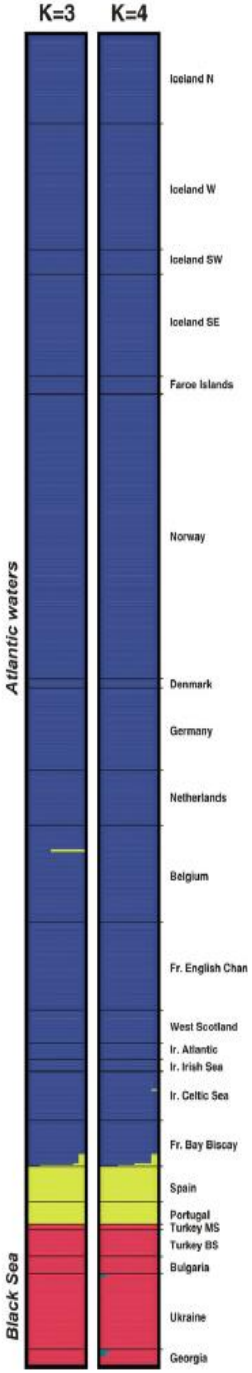
n		#S	n		#S
1		1	11		678570
2		2	12		4213597
3		5	13		27644437
4		15	14		190899322
5		52	15		1382958545
6		203	16		10480142147
7		877	17		82864869804
8		4140	18		682076806159
9		21147	19		5832742205057
10		11597	20		51724158235372

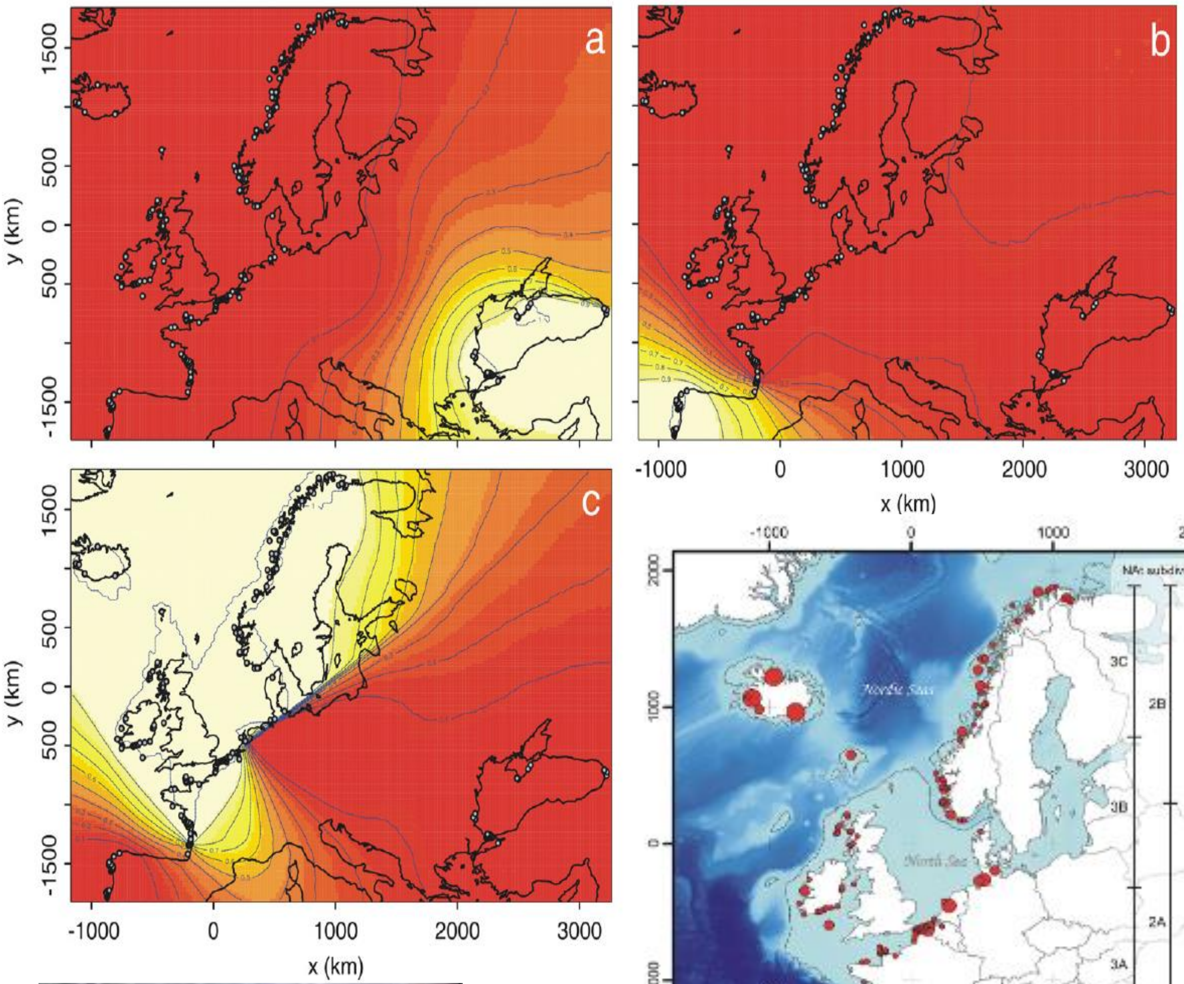
Human – Alu, Witherspoon et al. 2006

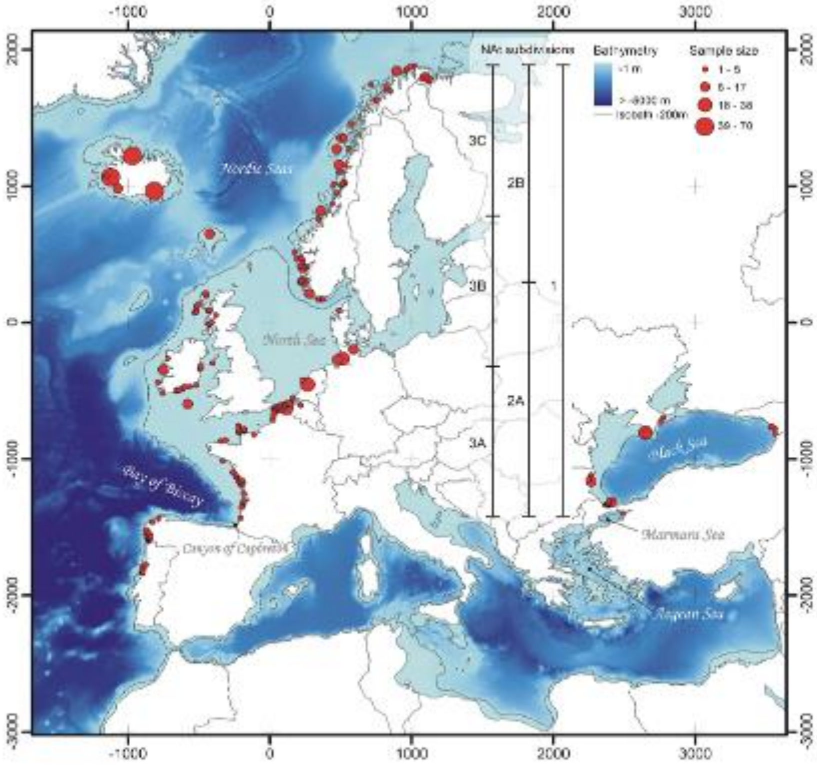


Spatial population (=landscape) genetics

Fontaine et al. 2007







harbor porpoise *Phocoena phocoena*