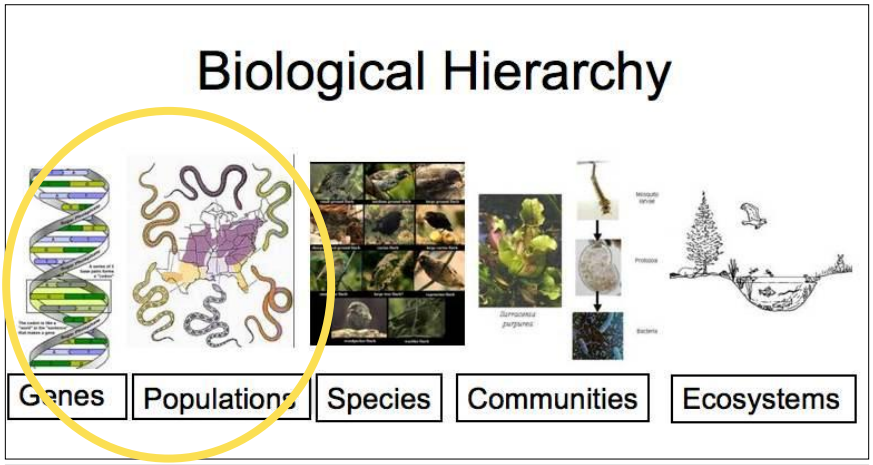
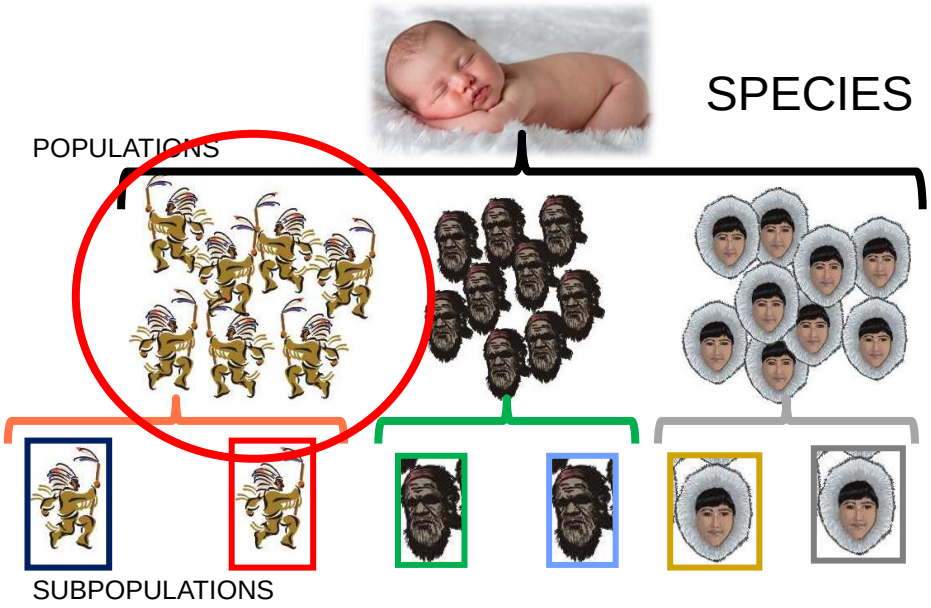
POPULATION GENETICS



I. GENETIC DIVERSITY

17 March 2016



I. GENETIC DIVERSITY – ANALYSIS OF SINGLE POPULATIONS

POPULATION
and problems of definition

- a population is a group of interbreeding indiviuals that exist together in time and space

- to develop the basic concepts of population genetics, we initially consider the **ideal population** = large, random-mating

ALLELE FREQUENCY

- proportion of an allele in comparison to all the others alleles of the same locus (gene) in a population sample

- basic characteristics for genetic diversity (variation) of a population

- population genetics studies genetic diversity and processes that have created it and influence it – i.e. the dynamics of distribution and frequency of alleles (genotypes → phenotypes), i.e. processes shaping **evolution**:

  increase of gen. diversity: mutation and migration
  decrease of gen. diversity: genetic drift (and natural selection)

MUTATIONS

increase genetic diversity
responsible for variation/heterogeneity in populations – essential to **evolution**

1. substitutions (transitions, transversions)

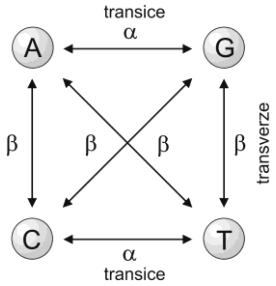
non-coding regions

synonymous GTC → GTA, Val → Val } = silent substitutions

nonsynonymous

missense GTC → TTC, Val → Phe

nonsense AAG → TAG, Lys → *ochre* (stop)



2.
insertion
deletion
ACGGT → ACAGGT
ACGGT → AGGT
} = indels
→ frameshift mutations

Mutation rate – rate at which number of various types of mutations occur in a given position over time

OBSERVATION

Callimorpha dominula

přástevník hluchavkový

Scarlet tiger moth

OBSERVATION

Callimorpha dominula

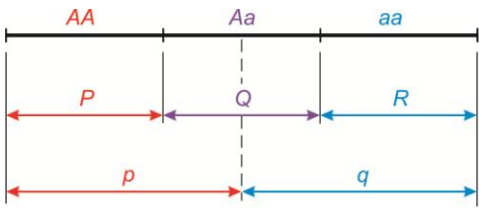
přástevník hluchavkový

Scarlet tiger mo

Table 3.1. Data from a collection of 1612 scarlet tiger moths.

Phenotype	No. of individuals
White spotting	1469
Intermediate	138
Little spotting	5

Genotype and allele frequency



Relative numbers = frequencies: genotype f.: P (G_{AA}), Q (G_{Aa}), R (G_{aa})

allele (gene) f.: p (A), q (a)

$P + Q + R = 1$
$p + q = 1$

Genotype	A_1A_1	A_1A_2	A_2A_2	Total
Number	n_1	n_2	n_3	N
Frequency	$P=n_1/N$	$Q=n_2/N$	$R=n_3/N$	
	$p = (2n_1 + n_2)/2N$		$q = (n_2 + 2n_3)/2N$	

Hardy-Weinberg Equilibrium (HWE)

Ex. Single locus with 2 alleles

Allele	Allele frequency
A	p
a	q

$p + q = 1$
p, q - Allele frequencies known from our samples

Genotype	Expected genotype frequency
AA	p^2
Aa	2pq
aa	q^2

= Hardy-Weinberg equilibrium

- Observed genotype frequencies (H_o) are known from our samples
- deviation of H_o from HWE conditions ⇒ for example χ^2 test

Expected heterozygosity, (H_e) under HWE
$H_e=1-(p^2+q^2)$ ..... for 1 locus with the allele frequencies *p* and *q*

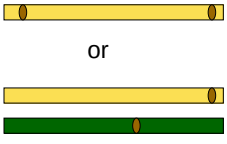
Assumptions for ideal population in HWE

- random-mating
- negligible effect of mutations and migration („closed populations“)
- infinitely large population (negligible effect of random fluctuations in allele frequencies in time – genetic drift) – **in HWE population the allele frequencies are stable = do not change between generations**
- Mendelian inheritance of the analysed loci
- neutral loci – not under selection
- diploid, sexually reproducing organisms with discrete generations

- loci are independent from each other – test for „linkage disequilibrium“

2 loci physically close to each other
(decreased probability of recombination
- **linkage disequilibrium**)

vs.



or

2 loci physically distant
(probability of recombination not influenced
- **linkage equilibrium**)

LINKAGE DISEQUILIBRIUM (LD)

loci in LINKAGE EQUILIBRIUM – segregate independently of each other during meiosis

the most common reason for non-random association among loci (LD) is the **proximity of two loci on a chromosome** (others e.g. small pop. size – gen. drift, immigration, overlapping generations, admixture, etc.)

haplotype diversity – p(AB) ≠ p(A) x p(B)

in presence of LD:

we have **fewer** independent loci for our genetic analysis than anticipated

neutral loci (alleles) linked to selected ones will appear non-neutral

presence of LD **needs to be tested** when analysing data from multiple loci

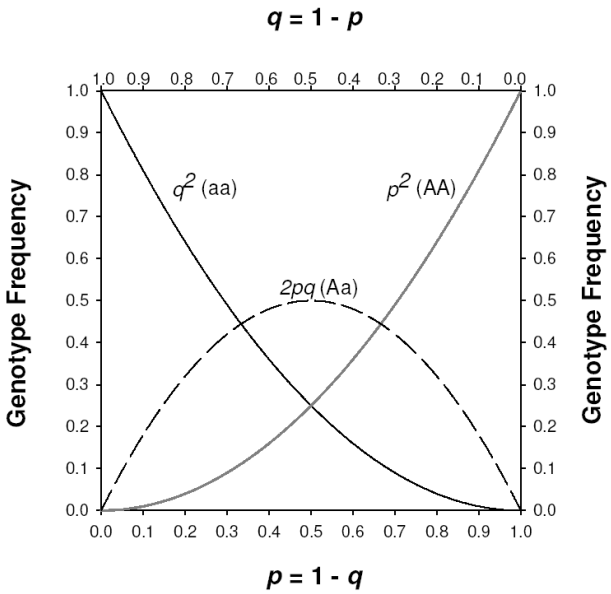


Figure 3.4 The combinations of homozygote and heterozygote frequencies that can be found in populations that are in HWE. Note that the frequency of heterozygotes is at its maximum when $p = q = 0.5$. When the allele frequencies are between 1/3 and 2/3, the genotype with the highest frequency will be the heterozygote.

Example of genetic diversity estimation in a sample of 4 individuals (on 4 loci)

Individual	Locus 1	Locus 2	Locus 3	Locus 4	Average
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**

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

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

Proportion of polymorphic loci (polymorphism) = 0,75

Is our population in HWE?

Callimorpha dominula

přástevník hluchavkový

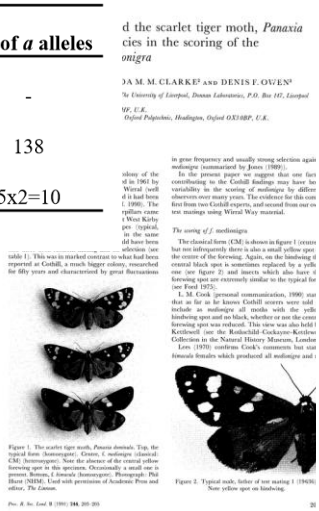
Scarlet tiger moth

Is our population in HWE?

Table 3.1. Data from a collection of 1612 scarlet tiger moths.

Phenotype	No. of individuals	Assumed genotype	No. of *A* alleles	No. of *a* alleles
White spotting	1469	*AA*	1469x2=2938	-
Intermediate	138	*Aa*	138	138
Little spotting	5	*aa*	-	5x2=10

www.shutterstock.com · 60840859



d the scarlet tiger moth, *Panaxia*
cies in the scoring of the
onigra

DA M. M. CLARKE[1] AND DENIS F. OWEN[2]

Deviation from HWE

- HWE test – e.g. Genepop software („exact probability tests“) – any case of **significant deviations from HWE** indicates that some of HWE assumptions were not fulfilled → detailed inspection required:
- heterozygote excess
  - negative **assortative mating** (i.e. intentional mating of distinct individuals)
  - used loci are advantageous in heterozygote situation (= balancing **selection** favouring heterozygotes, e.g. MHC genes)
  - **mutation**
  - **migration**
- heterozygote deficit
  - **inbreeding** (all loci are equally affected), assortative mating
  - genetic **structure** in populations
  - **null alleles** (only some loci affected by heterozygote deficit)

Quantifying genetic diversity

Polymorfism (proportion of polymorphic loci) - P

- **polymorphic locus** = with at least two alleles with having frequency of more numerous allele being **less or equal 0.95** (or 0.99)
- e.g. a population sample with four polymorphic loci out of five → P = 0.8

Number of alleles - N_a

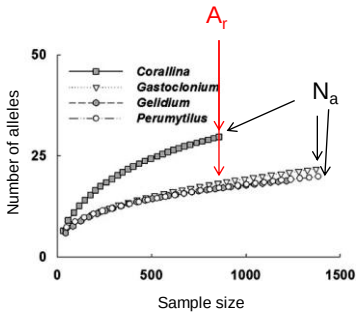
- number of alleles per locus (mean over loci)

Allelic richness - A_r

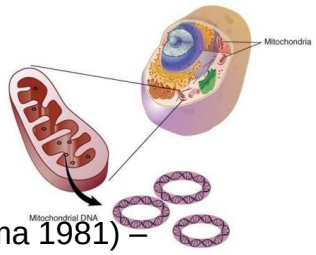
- number of alleles corrected for sample size (rarefaction method e.g. in FSTAT software)

Observed heterozygosity - H_o

- observed frequency of heterozygote genotypes (mean over loci)



HAPLOID DIVERSITY



- genetic diversity for haploid data

HAPLOTYPE DIVERSITY (h; Nei et Tajima 1981) – frequency of different haplotypes

$$H = \frac{N}{N-1}(1 - \sum_i x_i^2)$$

x_i –haplotype frequency of each haplotype in the sample
N – sample size

NUCLEOTIDE DIVERSITY (π; Nei 1987)

– quantifies the mean nucleotide divergence between sequences

– probability that two randomly chosen homologous nucleotides will be identical

$$\pi = \sum_{ij} x_i x_j \pi_{ij}$$

x_i and x_j – respective frequencies of the *i*th and *j*th sequences
π_{ij} – number of nucleotide differences per nucleotide site between the *i*th and *j*th sequences

WHAT INFLUENCES GENETIC DIVERSITY?

- influenced by a multitude of factors
- varies considerably between populations

MOST IMPORTANT DETERMINANTS OF GENETIC DIVERSITY:

- genetic drift
- population bottlenecks
- natural selection
- methods of reproduction

GENETIC DRIFT

population not infinitely large → population not in HWE → increase of influence of CHANCE → allele freqencies vary between generations

in absence of selection, each allele goes to:

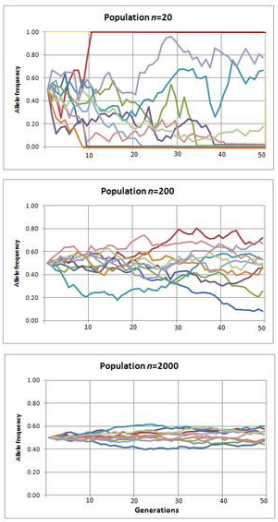
1. fixation
2. extinction

more quickly in smaller populations

→ **DECREASE of genetic diversity**

genetic drift – process causing a population´s allele frequencies to change from one generation to the next as a result of **CHANCE**

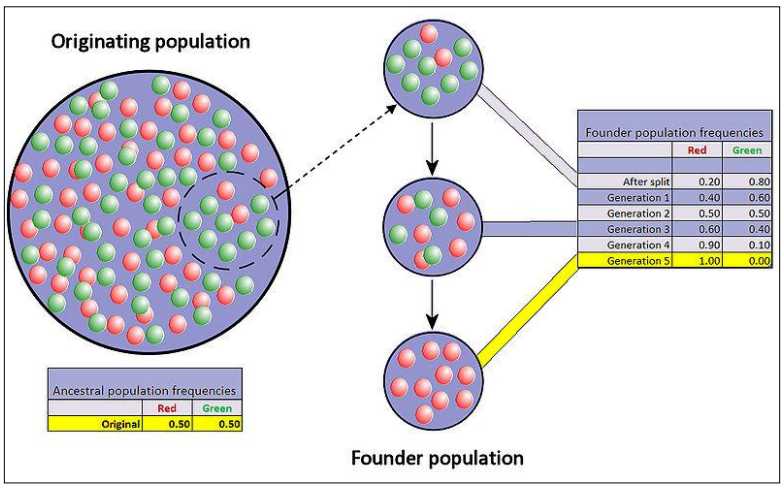
GENETIC DRIFT



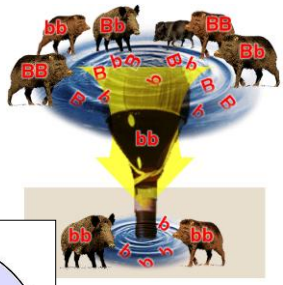
very profound effect of genetic drift in small populations – **founder effect, bottleneck**

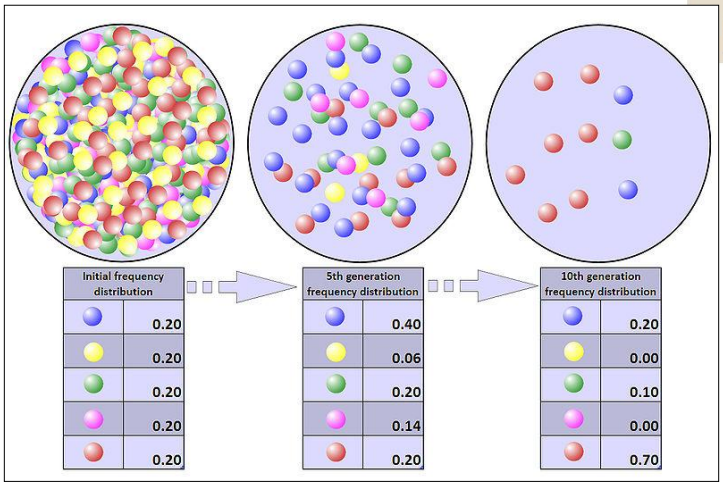
inextricable link between genetic drift and population size – **the effective population size**

Founder effect

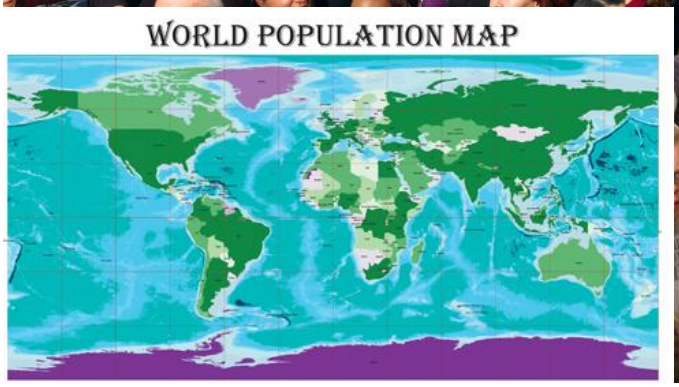


Bottleneck





N_e – effective population size



N_e – effective population size

vs. N_c – census population size (may be estimated from N_e – see Luikart *et al.* 2010 *Conserv Genet*)

all else being equal, LARGE pops are MORE LIKELY to survive than small pops

N_e – reflects the rate at which genetic diversity will be lost following genetic drift (this rate is inversely proportional to a population´s N_e)

single-sample estimators of Ne – level of LD due to drift

double sample estimators of Ne – temporal changes in allele frequencies due to genetic drift

OVERVIEW

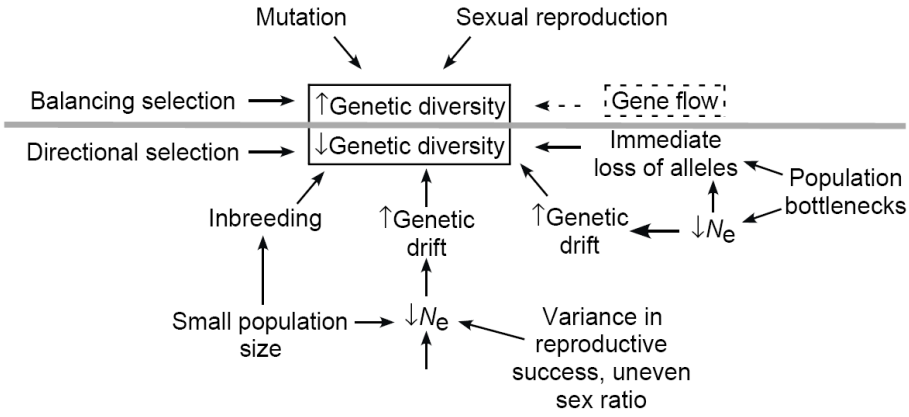


Figure 3.16 An overview of some of the main factors that influence levels of genetic diversity within populations.

Freeland *et al.* 2011