

**Lecture 4. Linkage disequilibrium (LD). Inbreeding coefficient as probability.
Inbreeding coefficient as fixation index.**

1.7 Linkage disequilibrium (LD)

Two genes with two alleles each: A, a and B, b
actual gamete frequencies (left) and
linkage equilibrium frequencies (right)

	B	b	Tot		B	b	Tot
A	P_{11}	P_{12}	p_1	A	p_1q_1	p_1q_2	p_1
a	P_{21}	P_{22}	p_2	a	p_2q_1	p_2q_2	p_2
Tot	q_1	q_2	1	Tot	q_1	q_2	1

Measures of LD

$$P_{11} = p_1q_1 + D, P_{12} = p_1q_2 - D$$

$$P_{21} = p_2q_1 - D, P_{22} = p_2q_2 + D$$

Basic LD measure $D = P_{11}P_{22} - P_{12}P_{21} = \text{Cov}(1_A, 1_B)$
depends on allele frequencies difficult to interpret

Correlation coefficient $r = \frac{D}{\sqrt{p_1p_2q_1q_2}}, r^2 = \frac{X^2}{n}$

Normalized D

$$D' = \frac{D}{D_{\max}} \text{ if } D > 0, \text{ where } D_{\max} = \min(p_1q_2, p_2q_1)$$

$$D' = \frac{D}{D_{\min}} \text{ if } D < 0, \text{ where } D_{\min} = -\min(p_1q_1, p_2q_2)$$

Ex 15: MN and Ss blood groups

Two genes in chromosome 4: alleles (M, N) and (S, s)

British sample, 1000 ind, $n = 2000$ chromosomes

Observed gamete counts and frequencies

	S	s	Total		S	s	Total
M	474	611	1085	M	0.237	0.305	0.542
N	172	773	915	N	0.071	0.387	0.458
Tot	616	1384	2000	Tot	0.308	0.692	1

Linkage equilibrium frequencies and counts

	S	s	Total		S	s	Total
M	0.167	0.375	0.542	M	334.2	750.8	1085
N	0.141	0.317	0.458	N	281.8	633.2	915
Tot	0.308	0.692	1	Tot	616	1384	2000

Chi-square test of independence: $X^2 = 184.9, df = 1$

$\sqrt{184.9} = 13.6$, reject H_0 : linkage equilibrium

$$\hat{D} = 0.070, \hat{r} = 0.304, \hat{D}' = \frac{0.07}{0.141} = 0.5$$

Attainment of linkage equilibrium

Changing D over generations under H-W assumptions

Fig 3.9, p. 100: $D_0 \rightarrow D_1 \rightarrow D_2 \rightarrow \dots \rightarrow 0$

$$D_t = D_0(1 - \rho)^t, \text{ where } \rho = \text{recombination fraction}$$

Causes of LD

1. small ρ , chromosome inversion
2. small t , recent mutation
3. epistatic selection favoring some genotypes
4. effectively small ρ , excess of homozygotes

Ex 16: LD in plants

Two unlinked esterase genes in Barley

gametes	B_1D_1	B_1D_2	B_2D_1	B_2D_2
observed counts	1501	754	720	74
LE expected counts	1642.6	613.7	577.1	215.6

$$X^2 = 172.7, \text{ df} = 1, D = -0.046, D' = 0.66$$

significant LD due to 99% self-fertilization

Haldane's recombination model

Number of crossovers between two loci u Morgans apart

$X_u \in \text{Pois}(u)$ [definition of 1 Morgan: $E(X_1) = 1$]

$\rho = P(X_u \text{ is odd}) = \frac{1}{2}(1 - e^{-2u}), \rho \approx u$ for small u

$\rho \approx 0.5$ for large u , independent assortment

1.8 Inbreeding coefficient as probability

Two alleles are IBD if they are derived

from a single allele in an ancestral HWE population

For an individual genotype any locus is

either autozygous: two IBD alleles, probability $P(\text{IBD})$

or allozygous: non IBD alleles, probability $1 - P(\text{IBD})$

$$\text{Pedigree formula of inbreeding coefficient}$$
$$F = P(\text{IBD}), F \geq 0$$

$$Fp = P(\text{autozygosity}) \times P(\text{ancestral allele is } A)$$

$$(1 - F)p^2 = P(\text{allozygosity}) \times P(\text{ancestors are } A, A)$$

$$D = Fp + (1 - F)p^2 = p^2 + pqF$$

1.9 Inbreeding coefficient as fixation index

Metapopulation = K partially isolated HWE subpop-s

Diploid population sizes $N_i = w_i N$, $w_1 + \dots + w_K = 1$

genotype frequencies $D_i = p_i^2$, $H_i = 2p_i q_i$, $R_i = q_i^2$

Metapopulation averages

$$\bar{p} = \sum_{i=1}^K p_i w_i$$

$$D_S = \sum_{i=1}^K p_i^2 w_i = \overline{p^2}, H_S = 2\overline{pq}, R_S = \overline{q^2}$$

Observed variance of allele freqs across subpopulations

$$\sigma^2 = \overline{p^2} - (\bar{p})^2$$

Complete allele fixation case: if $p_i = 0$ or 1 , then

$$\sigma^2 = \bar{p} - (\bar{p})^2 = \bar{p}\bar{q}$$

Total population = hypothetical fused population
with random mating

Expected genotype frequencies for the total population

$$D_T = (\bar{p})^2, H_T = 2\bar{p}\bar{q}, R_T = (\bar{q})^2$$

Wahlund's principle

isolation breaking increases genetic variation

$$D_S - D_T = \sigma^2, R_S - R_T = \sigma^2, H_T - H_S = 2\sigma^2$$

Isolation contributes to allele fixation

$$\text{Fixation index } F_{ST} = 1 - \frac{H_S}{H_T} = \frac{\sigma^2}{\bar{p}\bar{q}}$$

Inbreeding effect of population structure

$$D_S = \bar{p}^2 + \bar{p}\bar{q}F_{ST}, R_S = \bar{q}^2 + \bar{p}\bar{q}F_{ST}$$

$$H_S = 2\bar{p}\bar{q}(1 - F_{ST})$$

Literature:

1. D.L.Hartl, A.G.Clarc. Principle of population genetics. Sinauer Associates, 2007.
2. R.Nielson, M. Statkin. An introduction to population genetics: theory and applications, Sinauer Associates. 2013.