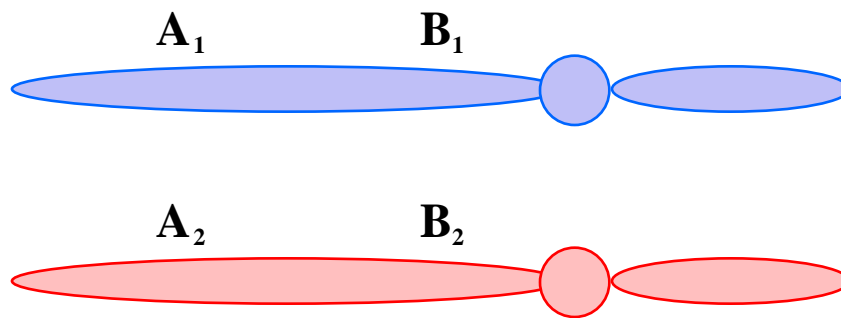


# Organization of Genetic Information

## II) Multi-locus

### A) Linkage Equilibrium

- 1) **Hardy-Weinberg at two loci – during meiosis different chromosomes are randomly segregated into different gametes so there should be no interchromosome association of alleles. What about random association of alleles at different loci on the same chromosome.**



Locus A — 2 alleles  $p_1$  and  $p_2$

$$p_1 + p_2 = 1.0$$

Locus B — 2 alleles  $q_1$  and  $q_2$

$$q_1 + q_2 = 1.0$$

- a) **Types of gametes produced if random association of alleles at different loci**

		Locus A	
		A <sub>1</sub>	A <sub>2</sub>
Locus B	B <sub>1</sub>	$p_1q_1$	$p_2q_1$
	B <sub>2</sub>	$p_1q_2$	$p_2q_2$

**In linkage equilibrium:**

$$A_1B_1 = p_1q_1 = P_{11}$$

$$A_1B_2 = p_1q_2 = P_{12}$$

$$A_2B_1 = p_2q_1 = P_{21}$$

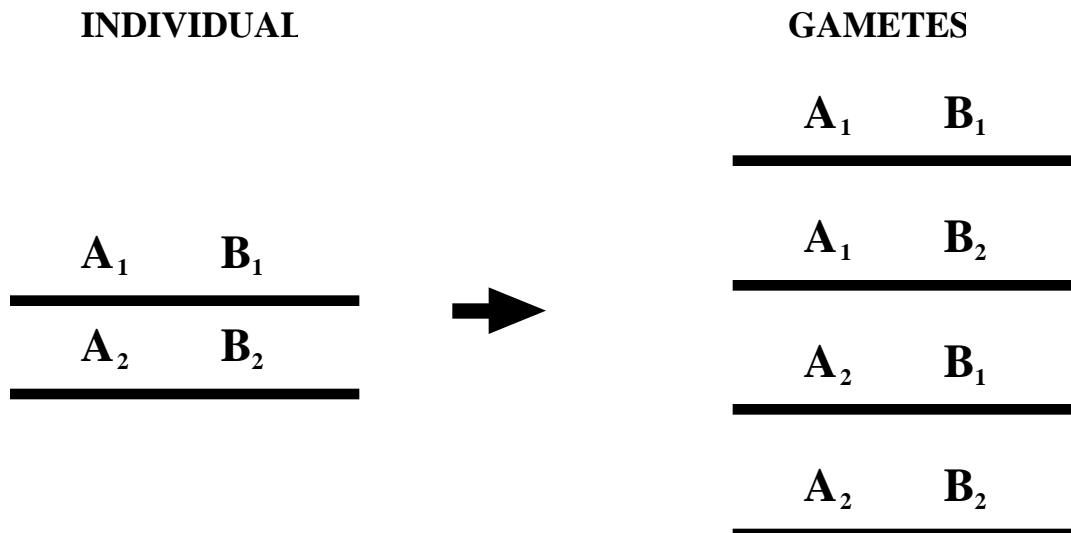
$$A_2B_2 = p_2q_2 = P_{22}$$

**AND**

$$P_{11} + P_{12} + P_{21} + P_{22} = 1.0 \quad \text{ALWAYS}$$

## **B) Linkage Disequilibria**

### **1) Basic Idea**



**r = fraction of the gametes that are recombinant**

**1 — r = fraction of the gametes that are not recombinant**

$$F(A_1B_1) + F(A_2B_2) = 1 - r$$

$$F(A_1B_1) = F(A_2B_2) = \frac{1-r}{2}$$

Similarly

$$F(A_1B_2) + F(A_2B_1) = r$$

$$F(A_1B_2) = F(A_2B_1) = \frac{r}{2}$$

## 2) Frequency of chromosomes in Gamete Pool

### OH 4.1

a) Assume complete linkage (i.e.,  $r = 0.0$ )

$$\begin{aligned} P_{11} &= P_{11}^2 + P_{11}P_{12} + P_{11}P_{21} + P_{11}P_{22} \\ &= (p_1q_1)^2 + p_1q_1p_1q_2 + p_1q_1p_2q_1 + p_1q_1p_2q_2 \\ &= p_1^2q_1^2 + p_1q_1p_2q_1 + p_1^2q_1q_2 + p_1q_1p_2q_2 \\ &= p_1q_1^2(p_1 + p_2) + p_1q_1q_2(p_1 + p_2) \\ &= p_1q_1^2 + p_1q_1q_2 \\ &= p_1q_1(q_1 + q_2) \end{aligned}$$

$P_{11} = p_1q_1$  (i.e., if there is no recombination then the frequency of each chromosome type in the gamete pool is equal to its frequency in the parental population.)

E.G., assume  $r = 0.0$  and the population consists of only  $A_1B_1$  and  $A_2B_2$  chromosomes

### OH 4.1A

**b) With recombination (maximum  $0.0 \leq r \leq 0.5$ )**

$$P_{11} = P_{11}(1 - r) + p_1q_1r$$

**Similarly:**

$$P_{12} = P_{12}(1 - r) + p_1q_2r$$

$$P_{21} = P_{21}(1 - r) + p_2q_1r$$

$$P_{22} = P_{22}(1 - r) + p_2q_2r$$

**i) Formulation of Disequilibria parameter:**

**E.G., Subtract  $p_1q_1$  from both sides and simplify:**

$$* \quad P_{11} - p_1q_1 = P_{11}(1 - r) + p_1q_1r - p_1q_1$$

$$P_{11} - p_1q_1 = (1 - r)(P_{11} - p_1q_1)$$

**Define the disequilibrium variable “D” as (deviation from expected)**

$$D = (P_{11} - p_1q_1) \quad (D_{ij} = (1 - r)(P_{ij} - p_iq_j) \text{ note } D \leq 0.0 \geq D)$$

**the \* equation then becomes:**

$$D' = (1 - r)D$$

**OR**

$$D_1 = (1 - r)D_{N-0}$$

**AND**

$$D_2 = (1 - r)(1 - r)D_0$$

**AND SO ON, SO:**

$$D_t = (1 - r)^t D_0 \quad \text{since } 0.0 \leq r \leq 0.5 \text{ as } t \Rightarrow \infty \quad (1 - r)_t \Rightarrow 0.0$$

**How fast disequilibrium approaches zero depends on the magnitude of r. (OH 4.2).**

c)  $D_{\max}$  Calculations – the maximum value that D can have is related to the frequency of the frequency of the alleles.

$D = (P_{11} - p_1q_1)$  can be rewritten as

$$P_{11} = p_1q_1 + D$$

$$P_{12} = p_1q_2 - D$$

$$P_{21} = p_2q_1 - D$$

$$P_{22} = p_2q_2 + D$$

It can also be shown that:

$$D = P_{11}P_{22} - P_{12}P_{21}$$

Since  $P_{11}$ ,  $P_{12}$ ,  $P_{21}$ , and  $P_{22}$  are all non-negative for any value of  $p_1$ ,  $p_2$ ,  $q_1$ , and  $q_2$ :

$$D_{\min} = \text{the larger of } -p_1q_2 \text{ or } -p_2q_1$$

$$D_{\max} = \text{the smaller of } p_1q_2 \text{ or } p_2q_1$$

For comparative studies we define a new parameter:

$$D' = \frac{D}{D_{\max}} \text{ where } -1.0 \leq D' \leq 1.0 \text{ and } D' \text{ corresponds to the}$$

percentage of disequilibrium relative to the maximum possible given the allele frequencies.

### C) Factors Producing Linkage Disequilibrium

- 1) **Mutation – not a significant factor in natural populations since mutation rate ( $\mu$ ) is usually very small (i.e.,  $10^{-7}$ )**

Assume two loci one fixed for  $A_1$  the other has two alleles  $B_1$  and  $B_2$  at frequency  $q_1$  and  $q_2$

$A_1 \Rightarrow A_2$  at a rate of  $\mu$

When  $A_1 \Rightarrow A_2$  the initial frequency of  $A_2$  is  $\frac{1}{2N}$  where  $N$  is the number of individuals in the population.

#### OH 4.2.1

$$\begin{aligned} D &= P_{11}P_{22} - P_{12}P_{21} \\ &= \left(q_1 - \frac{1}{2N}\right) 0.0 - q_2\left(\frac{1}{2N}\right) \\ &= -\frac{q_2}{2N} \end{aligned}$$

$D$  is quite small, however, given the frequencies  $D' = -1.0$  or  $D' = 1.0$  (depending on the phase).

- 2) **Inbreeding** – when populations inbreed it can have a similar effect to reduced recombination (i.e., produce gametic disequilibrium). With inbreeding, overall heterozygosity is reduced. The reduction in the frequency of the double heterozygotes reduces the changes for recombination to produce new chromosomes. Inbreeding slows the rate of decay of inbreeding. Genome wide.

**Remember**

$$D_t = (1 - r)D_{t-1}$$

**With inbreeding and recombination:**

$$D_{t+1} = \frac{1}{2} \frac{1 + \lambda + S}{2} + \frac{1 + \lambda + S}{2}^2 - 2S\lambda^{\frac{1}{2}} D_t$$

Where  $\lambda = 1 - 2r$  and  $S =$  selfing rate ( $\lambda$  is used only to make recombination max and min the same as selfing (i.e.,  $0.0 \leq S \leq 1.0$ ))

if  $S = 0.0$

$$\begin{aligned} D_{t+1} &= \frac{1}{2}(1 + \lambda)D_t \\ &= \frac{1}{2}(1 + 1 - 2r)D_t \\ &= \frac{1}{2}(2 - 2r)D_t \\ &= (1 - r)D_t \end{aligned}$$

**OH 4.2.2**

- 3) **Drift and Gametic Disequilibrium – stochastic change in the allele frequency due to sampling also can produce gametic disequilibrium (i.e., correlations of alleles at two loci).**

**Similar to a single locus with 4 alleles with drift there can be a loss of a gamete type. Disequilibrium between two loci with two alleles each is similar in that a chromosome type can be lost. Can use the square of the correlation ( $r^2$ ) among alleles (correlations can be either + or –) to show the effect of population size on disequilibrium.**

$$E(r^2) = \frac{(1 - c)^2 + c^2}{2N_e c(2 - c)} + \frac{1}{N}$$

**where:**

**$r^2$  is the correlation coefficient,  $c$  is the recombination rate,  $N$  is the population size, and  $N_e$  is the genetically effective population size**

**As  $c$  becomes small (i.e., little recombination) and  $N$  becomes large (i.e., large population size)**

$$E(r^2) \approx \frac{1}{1 + 4N_e c}$$

**When  $N_e c$  is large, the correlation between alleles at different loci approaches 0.0 (i.e., in equilibrium), and when  $N_e c$  is small it approaches 1.0 (i.e., in disequilibrium)**

**OH 4.2.2A**

when  $c = 0.5$  (i.e., unlinked)

$$E(r^2) \approx \frac{1}{3N_e} + \frac{1}{N_e}$$

**AND**

$$N_E = \frac{1}{3 [E(R^2)]^2 - \frac{1}{N}}$$

- 4) **Selection – differential survival of chromosome types can also produce disequilibrium.**
  - a) **Hitchhiking of neutral alleles – alleles that are not themselves under selection respond to selection at linked loci.**
    - i) **magnitude of effect of selection is dependent on degree of linkage, selfing rate, and type and amount of initial (i.e., before selection) gametic disequilibrium. If not in disequilibrium (i.e., correlation between alleles at different loci is 0.0) then hitchhiking cannot happen.**
    - ii) **similar to drift, selective sweeps can cause linked neutral loci to be in disequilibrium**

**b) Multilocus selection – selection can operate at two loci at once.**

**i) Epistasis – the interaction of alleles at nonhomologous loci to affect a single trait.**

➤ **multiplicative – effects of single loci are multiplied**

**e.g. Survivorship of  $A_1A_1$  genotype = 90%**

**Survivorship of  $B_1B_1$  genotype = 80%**

**Survivorship of  $A_1A_1B_1B_1$  genotype =  $0.90 * 0.80 = 0.72$**

➤ **Additive - effects of single loci are added.**

**e.g., Genotype  $A_1A_1$  produces 2 seeds above average**

**Genotype  $B_1B_1$  produces 3 seeds above average**

**Genotype  $A_1A_1B_1B_1$  produces 5 seeds above average**

**Consider a haploid model of two loci under selection and epistatic**

**Fitnesses:  $A_1B_1 = W_1$**

**$A_1B_2 = W_2$**

**$A_2B_1 = W_3$**

**$A_2B_2 = W_4$**

**The degree of epistasis is (E)**

**$E = W_1 - W_2 - W_3 + W_4$  when Additive and**

**$E = \ln W_1 - \ln W_2 - \ln W_3 + \ln W_4$**

**$= \ln \frac{W_1 W_4}{W_2 W_3}$  When multiplicative**

**∴ if  $E > 0.0$  some genotypes will be more common than expected based on the single locus frequency.**

ii) **Genetic Background** – epistatic effects can come from a variety of sources and many loci.

5) **Admixture (migration or hybridization)** – the mixing of populations that have been isolated or of good biological species will cause linkage disequilibrium.

➤ **allele frequencies in the separate populations must be different.**

➤ **allele frequency change must be large for any significant amount of disequilibrium to be generated.**

○ **Assume no gametic disequilibrium in parental populations.**

$$D = m_x m_y (p_{1 \cdot x} - p_{1 \cdot y}) \times (q_{1 \cdot x} - q_{1 \cdot y})$$

**Where:**  $m_x$  and  $m_y$  are the proportion in the combined pool of individuals from population x and y, respectively and  $m_x = (1 - m_y)$ .

$p_{1 \cdot x}$  and  $p_{1 \cdot y}$  are the frequency of the  $A_1$  allele in populations x and y respectively

$$p_{1 \cdot i} = p_{11 \cdot i} + p_{12 \cdot i}$$

$q_{1 \cdot x}$  and  $q_{1 \cdot y}$  are the frequency of the  $B_1$  allele in populations x and y respectively

$$q_{1 \cdot i} = q_{11 \cdot i} + q_{21 \cdot i}$$

**OH 4.3**  
**OH 4.3.5**  
**Movie Admix.mov**

**With populations that themselves are in gametic linkage disequilibrium:**

$$D = m_x D_x + m_y D_y + m_x m_y (p_{1 \cdot x} - p_{1 \cdot y}) \times (q_{1 \cdot x} - q_{1 \cdot y})$$

**It can be shown that:**

**Rate of decay of gametic linkage disequilibrium is:**

$$D_{t+1} = (1 - r)D_t$$

$$D_{t+1} = (1 - 2m)^2 D_t \quad \text{whichever is larger.}$$

**D) Cytonuclear Disequilibrium – correlation between nuclear allele frequencies and mitochondrial or chloroplast DNA haplotypes.**

- **assume 2 mtDNA alleles (haplotypes) and 1 nuclear locus with two alleles**
- **most useful in situations of hybridization and introgression**

**OH 4.4**

**OH 4.5**

## Review of Factors Producing Gametic Linkage Disequilibrium

### 1) Mutation:

$$D = - \frac{q_2}{2N}$$

$$- 1.0 \leq D' \leq 1.0$$

- a. recombination (r) reduced D'
- b. Disease gene markers

### 2) Inbreeding

$$D_{t+1} = \frac{1}{2} \frac{1 + \lambda + S}{2} + \frac{1 + \lambda + S}{2}^2 - 2S\lambda^{\frac{1}{2}} D_t$$

- a. reduces number of double heterozygotes necessary for recombination to reduce disequilibrium

### 3) Drift

$$E(r^2) = \frac{(1 - c)^2 + c^2}{2N_e c(2 - c)} + \frac{1}{N}$$

### 4) Selection and Epistasis

$$E = \ln W_1 - \ln W_2 - \ln W_3 + \ln W_4$$

OR

$$= \ln \frac{W_1 W_4}{W_2 W_3}$$

### 5) Admixture and Migration

$$D = m_x D_x + m_y D_y + m_x m_y (p_{1 \cdot x} - p_{1 \cdot y}) \times (q_{1 \cdot x} - q_{1 \cdot y})$$

$$p_{1 \cdot i} = p_{11 \cdot i} + p_{12 \cdot i}$$

$$q_{1 \cdot i} = q_{11 \cdot i} + q_{21 \cdot i}$$

and Rate of decay:

$$D_{t+1} = (1 - r)D_t$$

$$D_{t+1} = (1 - 2m)^2 D_t \quad \text{whichever is larger.}$$