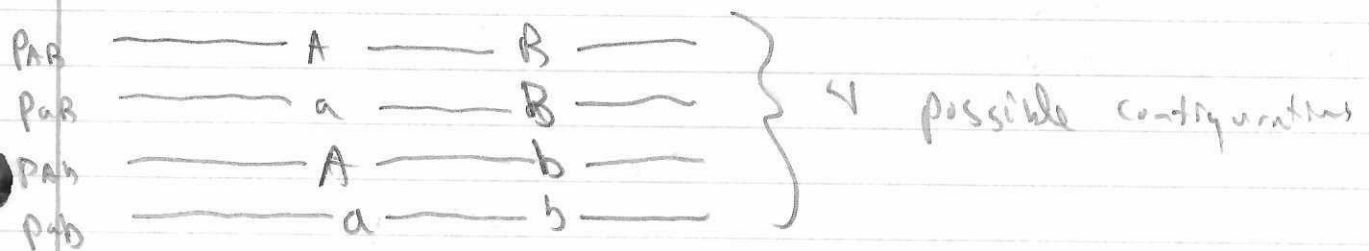


Linkage Disequilibrium

- LD = association between alleles at different loci
- Genetic LD - association between alleles on same chromosome (haplotypic)
- p_A = allele freq at locus 1
- p_B = allele freq at locus 2



- Alleles are independent if $p_{AB} = p_A p_B$

- LD coefficient:

$$D_{AB} = p_{AB} - \hat{p}_A \hat{p}_B \quad 0 \Rightarrow \text{Independent}$$

$$\hat{D}_{AB} = \hat{p}_{AB} - \hat{p}_A \hat{p}_B$$

Sample:

AB	$\hat{p}_{AB} = \frac{3}{7}$	$\hat{p}_A = \frac{4}{7}$	$\hat{p}_B = \frac{5}{7}$
AB			
Ab	$\hat{D}_{AB} = \frac{3}{7} - \frac{4}{7} \cdot \frac{5}{7}$		
aB			
aB			
ab			
AB			

Can get $E(\hat{D}_{AB})$, $var(\hat{D}_{AB})$

$$Z = \frac{\hat{D}_{AB} - E(\hat{D}_{AB})}{\sqrt{var(\hat{D}_{AB})}} \sim N(0,1)$$

↑ use to get χ^2 test

• or goodness of fit χ^2

Observed	n_{AB}	\hat{n}_{AB}	\hat{n}_{aB}	n_{ab}
Expected	$z \hat{p}_A \hat{p}_B$	$z \hat{p}_A \hat{p}_b$	$z \hat{p}_a \hat{p}_B$	$z \hat{p}_a \hat{p}_b$

3 parameters under alternative

$$p_{AB} + p_{aB} + p_{Ab} + p_{ab} = 1$$

2 parameters under null: p_A, p_B

⇒ χ^2 distribution

$$H_0: D_{AB} = 0$$

$$H_A: D_{AB} \neq 0$$

Standardized D_B equilibrium coefficient

D_{AB} is constrained by

$$\max(-p_A p_B, -(1-p_A)(1-p_B)) \leq D_{AB} \leq \min(p_A(1-p_B), (1-p_A)p_B)$$

Ex.) $p_A = 0.5, p_B = 0.01$ then

$$-0.005 \leq D_{AB} \leq 0.005$$

↑ very constrained, so hard to compare LD over the genome, since depends on p_A, p_B

Ex.) $p_A = p_B = 0.5$

$$-0.25 \leq D_{AB} \leq 0.25$$

$$D'_{AB} = \begin{cases} \frac{D_{AB}}{\min(p_A p_B, (1-p_A)(1-p_B))} & \text{if } D_{AB} < 0 \\ \frac{D_{AB}}{\min(p_A(1-p_A), (1-p_A)p_B)} & \text{if } D_{AB} > 0 \end{cases}$$

↑ free to range from -1 to 1

• Pearson correlation

$$r = \text{cor}(I(X_i = A), I(Y_i = B))$$

$$X_i = \begin{cases} 1 & \text{if game has } A \\ 0 & \text{otherwise} \end{cases}$$

$$Y_i = \begin{cases} 1 & \text{if game has } B \\ 0 & \text{otherwise} \end{cases}$$

$$= \frac{\text{Cov}(I(X_i = A), I(Y_i = B))}{\text{SD}(I(X_i = A)) \text{SD}(I(Y_i = B))}$$

$$= E[I(X_i = A, Y_i = B)] - E[I(X_i = A)] E[I(Y_i = B)]$$

$$= P_i(X_i = A, Y_i = B) - P_i(X_i = A) P_i(Y_i = B)$$

$$= P_{AB} - P_A P_B$$

$$= D_{AB}$$

$$\text{SD}(I(X_i = A)) = \frac{\sqrt{P_i(X_i = A)(1 - P_i(X_i = A))}}{\sqrt{P_A(1 - P_A)}}$$

$$r = \frac{D_{AB}}{\sqrt{P_A(1 - P_A) P_B(1 - P_B)}}$$

sometimes, phase is not known, so hard to estimate Genetic LD

Genotype LD

↙ Association b/t alleles at different loci and possibly different haplotypes.

↘ Association between dosages

G_{iA} = dosage at locus 1 $\in \{0, 1, 2\}$
 G_{iB} = dosage at locus 2 $\in \{0, 1, 2\}$

Ex.)

— A — B —	$G_A = 2$
— A — b —	$G_B = 1$

$\Delta_{AB} =$ "composite LD coefficient"
 $= \text{cov}(G_A, G_B)$

under HWE, $\frac{1}{2} \Delta_{AB} = D_{AB}$

What does HWE mean for haplotypes?
↳ multinomial distribution

under HWE: $(G_{AB}, G_{Ab}, G_{aB}, G_{ab}) \sim \text{Mult}(2, p_{AB}, p_{Ab}, p_{aB}, p_{ab})$

G_{AB} = # haplotypes an individual has w/ AB out of 2

Recall HWE at a single biallelic locus:

$G_A \sim \text{Bin}(2, p_A)$

↳ # A's an individual has

$$G_A = G_{AB} + G_{Ab}$$

$$G_B = G_{AB} + G_{aB}$$

$$\text{Cov}(G_A, G_B) = \text{Cov}(G_{AB} + G_{Ab}, G_{AB} + G_{aB})$$

$$= 2 D_{AB}$$

↑
math

• If HWE is not true, then

		G_B		
	0	1	2	
G_A	0	p_{00}	p_{01}	p_{02}
	1	p_{10}	p_{11}	p_{12}
	2	p_{20}	p_{21}	p_{22}

↑ 8 parameters, not 4 under HWE

• If HWE, then,

$$Pr(G_A=1, G_B=1) = Pr \left[\begin{array}{l} (G_{AB}=1, G_{Ab}=0, G_{aB}=0, G_{ab}=1) \\ \text{or} (G_{AB}=0, G_{Ab}=1, G_{aB}=1, G_{ab}=0) \end{array} \right]$$

$$= 4 p_{AB} p_{ab} + 4 p_{Ab} p_{aB}$$

etc...

• Δ is useful when

- ① HWE is not fulfilled (ex. experimental populations)
- ② Phasmy is not available

• $\rho = \text{Cor}(G_A, G_B)$

= r if HWE is fulfilled