Haplotype Frequencies

		Locus B		Totals
		B	b	
Locus A	A	p_{AB}	p_{Ab}	p_A
	a	p_{aB}	p_{ab}	p_a
Totals		p_B	p_b	1.0

Linkage Equilibrium
Expected for Distant Loci

$$p_{AB} = p_A p_B$$

$$p_{Ab} = p_A p_b = p_A(1 - p_B)$$

$$p_{aB} = p_a p_B = (1 - p_A)p_B$$

$$p_{ab} = p_a p_b = (1 - p_A)(1 - p_B)$$

Linkage Disequilibrium
Expected for Nearby Loci

$$p_{AB} \neq p_A p_B$$

$$p_{Ab} \neq p_A p_b = p_A(1 - p_B)$$

$$p_{aB} \neq p_a p_B = (1 - p_A)p_B$$

$$p_{ab} \neq p_a p_b = (1 - p_A)(1 - p_B)$$

Disequilibrium Coefficient D_{AB}

$$D_{AB} = p_{AB} - p_A p_B$$

$$p_{AB} = p_A p_B + D_{AB}$$

$$p_{Ab} = p_A p_b - D_{AB}$$

$$p_{aB} = p_a p_B - D_{AB}$$

$$p_{ab} = p_a p_b + D_{AB}$$

D_{AB} is hard to interpret

- Sign is arbitrary …
  - A common convention is to set A, B to be the common allele and a, b to be the rare allele

- Range depends on allele frequencies
  - Hard to compare between markers

What is the range of D_{AB}?

- What are the maximum and minimum possible values of D_{AB} when
  - $p_A = 0.3$ and $p_B = 0.3$
  - $p_A = 0.2$ and $p_B = 0.1$

- Can you derive a general formula for this range?

D' – A scaled version of D

$$D'_{AB} = \begin{cases} \dfrac{D_{AB}}{\min(p_A p_B, p_a p_b)} & D_{AB} < 0 \\[2ex] \dfrac{D_{AB}}{\min(p_A p_b, p_a p_B)} & D_{AB} > 0 \end{cases}$$

- Ranges between –1 and +1
  - More likely to take extreme values when allele frequencies are small
  - ±1 implies at least one of the observed haplotypes was not observed

More on D'

- Pluses:
  - D' = 1 or D' = -1 means no evidence for recombination between the markers
  - If allele frequencies are similar, high D' means the markers are good surrogates for each other
- Minuses:
  - D' estimates inflated in small samples
  - D' estimates inflated when one allele is rare

Δ^2 (also called r^2)

$$\Delta^2 = \frac{D_{AB}^2}{p_A(1-p_A)p_B(1-p_B)}$$

$$= \frac{\chi^2}{2n}$$

- Ranges between 0 and 1
  - 1 when the two markers provide identical information
  - 0 when they are in perfect equilibrium
- Expected value is $1/2n$

More on r^2

- $r^2 = 1$ implies the markers provide exactly the same information

- The measure preferred by population geneticists

- Measures loss in efficiency when marker A is replaced with marker B in an association study
  - With some simplifying assumptions (e.g. see Pritchard and Przeworski, 2001)