II. Population Genetics

Linkage Disequilibrium

Lecture 11

Instructor: Su-In Lee

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

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

- 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

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)

When does linkage equilibrium hold?

Equilibrium or Disequilibrium?

- We will present simple argument for why linkage equilibrium holds for most loci
- Balance of factors
  - Genetic drift (a function of population size)
  - Random mating
  - Distance between markers
  - …

Why Equilibrium is Reached…

- Eventually, random mating and recombination should ensure that mutations spread from original haplotype to all haplotypes in the population…
- Simple argument:
  - Assume fixed allele frequencies over time

Generation t, Initial Configuration

- Assume arbitrary values for the allele frequencies and disequilibrium coefficient

	B	b	
A	$p_A p_B + D_{AB}$	$p_A p_b - D_{AB}$	p_A
a	$p_a p_B - D_{AB}$	$p_a p_b + D_{AB}$	p_a
	p_B	p_b	

Generation t+1, Without Recombination

- Haplotype Frequencies Are Function of Allele Frequencies
- Outcome has probability r

	B	b	
A	$p_A p_B$	$p_A p_b$	p_A
a	$p_A p_b$	$p_a p_b$	p_a
	p_B	p_b	

Generation t+1, Overall

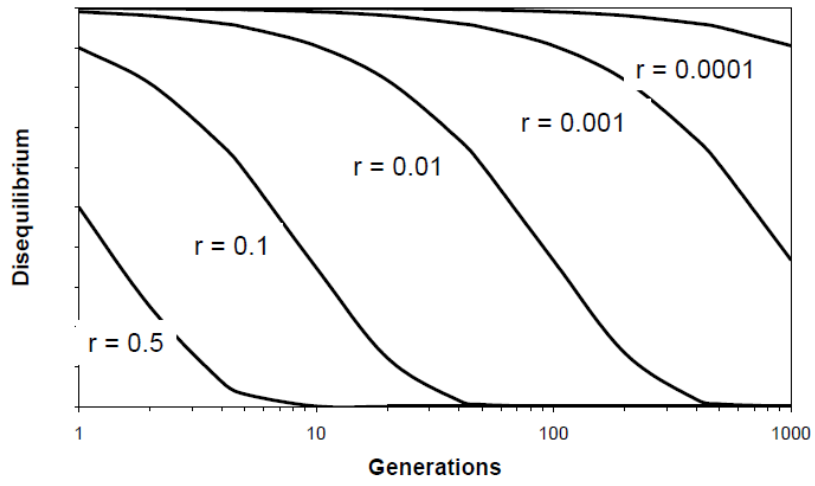
- Disequilibrium Decreases…

	B	b	
A	$p_A p_B + (1-\theta) D_{AB}$	$p_A p_b - (1-\theta) D_{AB}$	p_A
a	$p_A p_b - (1-\theta) D_{AB}$	$p_a p_b + (1-\theta) D_{AB}$	p_a
	p_B	p_b	

Recombination Rate (r)

- Probability of an odd number of crossovers between two loci
- Proportion of time alleles from two different grand-parents occur in the same gamete
- Increases with physical (base-pair) distance, but rate of increase varies across genome

Decay of D with Time



Predictions

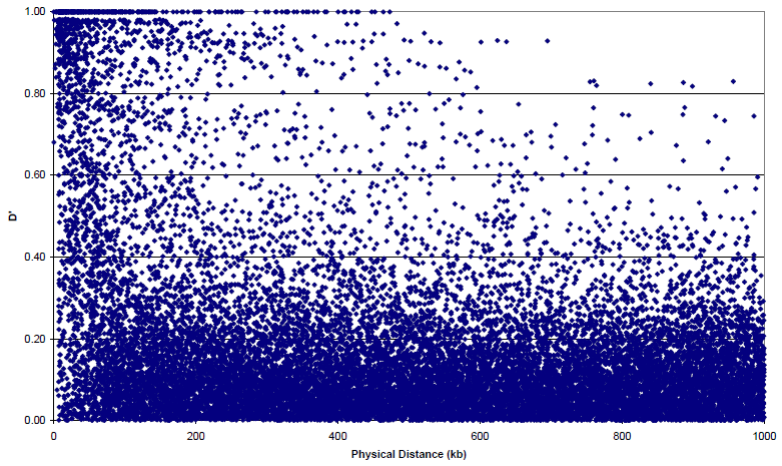
- Disequilibrium will decay each generation
  - In a large population
- After t generations…
  - $D^t_{AB} = (1-\Theta)^t D^0_{AB}$
- A better model should allow for changes in allele frequencies over time…

Linkage Equilibrium

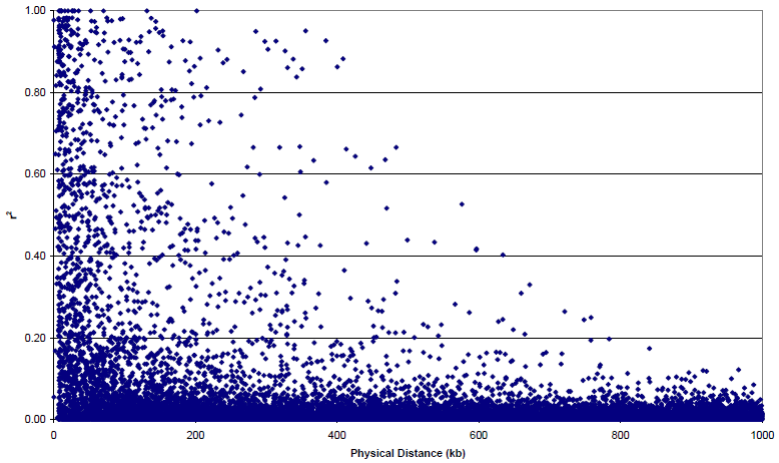
- In a large random mating population haplotype frequencies converge to a simple function of allele frequencies

Some Examples of Linkage Disequilibrium Data

Raw |D'| data from human Chr22

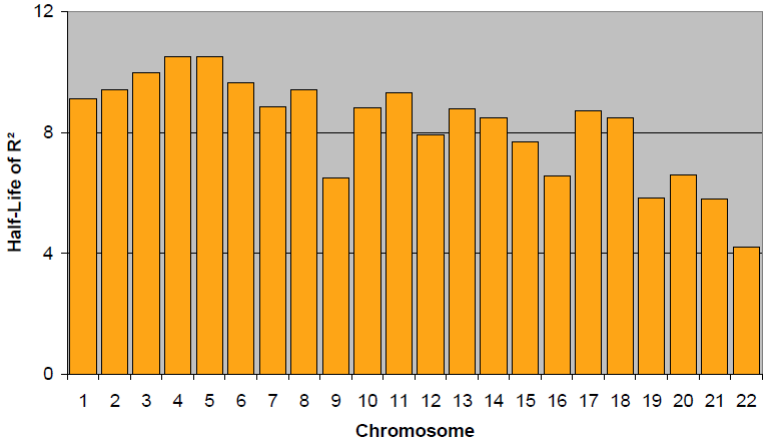


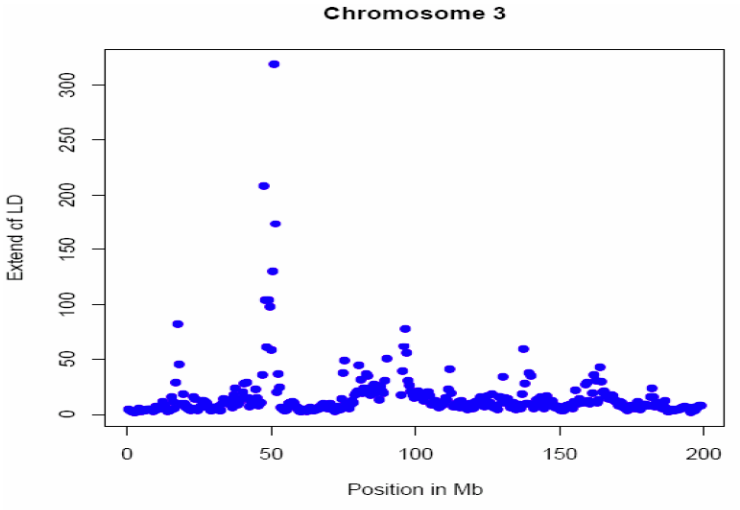
Raw Δ^2 data from human Chr22



Extent of Linkage Disequilibrium

- LD extends further in larger chromosomes, which have lower recombination rates





But within each chromosome, there is still huge variability!

Acknowledgement

- This lecture note is based on material by Prof Goncalo Abecasis (Univ of Michigan).
- Chapter 2 of *Principles of Population Genetics* by Daniel L. Hartl & Andrew Clark