Chapter 3: Genetic Linkage

Models and methods for data at two loci

3.1 LINKAGE AND RECOMBINATION 3.1.1 MEIOSIS INDICATORS

- For one locus we have seen (2.4.1) that *segregation* of genes at a locus is fully specified by *meiosis indicators:*
  S_i = 0 if copied DNA is from parent's maternal genome
  . = 1 if copied DNA is from parent's paternal genome
  where i = 1,...,m indexes the meioses.
- S_i are independent with P(S_i = 0) = P(S_i = 1) = 1/2.
- ibd at a locus is a function of the S_i at that locus.

- For multiple loci:
  . S_ij = 0 if copied DNA in meiosis i at locus j is parent's maternal DNA
  . = 1 if copied DNA in meiosis i at locus j is parent's paternal DNA
- Here i=1,...,m indexes the meioses of the pedigree, and j = 1,..., L indexes the genetic loci.
- The marginal distribution of each S_ij is as before:
  . P(S_ij = 0) = P(S_ij = 1) = 1/2.
- For different meioses i, the S_ij are independent.

3.1.2 Recombination frequencies

- Chromosomes are inherited in chunks. In meiosis, the chromosomes of a pair duplicate, align and exchange material.
- Offspring chromosome consists of segments of the two parental chromosomes (length approx 10^8 bp).
- There is dependence in DNA inherited at nearby locations: dependence is stronger for closer locations.
- The pairwise distribution of (S_ij, S_ij') is determined by the *recombination frequency*, which is a measure of the independence in inheritance between the two loci. (Larger recombination frequency: closer to independence.)
- For two given loci (j and j') the recombination frequency ρ between them is
  . ρ = P(S_ij ≠ S_ij') for each i, 0 ≤ ρ ≤ 1/2.
- For loci that are close together on a chromosome, ρ is close to 0.
- For independently segregating loci, ρ = 1/2.
- Note there is recombination if the genes at the two loci derive from different grandparents.
- In practice, recombination frequencies vary among meioses, a major factor in this variation being the sex of the parent. Computationally, this can be incorporated.

3.1.3 Haplotypes and linkage (2 loci)

- Two diallelic loci, one with codominant alleles A1 and A2, and the other with codominant alleles B1 and B2.
- There are four haplotypes A1 B1, A1 B2, A2 B1 and A2 B2.
- Suppose haplotype frequencies are q1, q2, q3, q4, respectively.
- There are 10 two-locus genotypes, but only 9 phenotypes.
- Genotypes A1 B1 / A2 B2 and A1 B2 / A2 B1 both have double-heterozygote phenotype A1 A2 at locus A, B1 B2 at locus B.
- Recall section 1.6:
  Notation A1 B1 / A2 B2 denotes that alleles A1 and B1 are on a single haplotype (phase), and A2 and B2 are on the other.

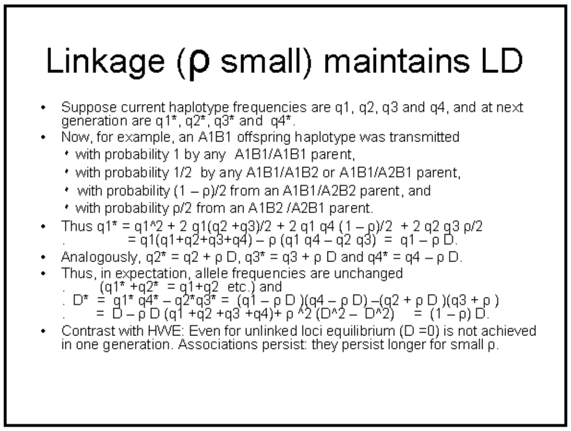
	B1B1	B1B2	B2B2
A1A1	A1B1/ A1B1	A1B1/ A1B2	A1B2/ A1B2
A1A2	A1B1/ A2B1	A1B1/ A2B2 . or A1B2/ A2B1	A1B2/ A2B2
A2A2	A2B1/ A2B1	A2B1/ A2B2	A2B2/ A2B2

Frequencies and Segregation of Haplotypes

- Assuming HWE for haplotypes, haplotype frequencies can be estimated from phenotype frequencies via the EM algorithm (see section 1.6):
  P(A1 B1 / A2 B2 | A1 A2, B1 B2) = 2 q1 q4 /(2 q1 q4 + 2 q2 q3).
- Given a set of current haplotype frequency estimates qi, i=1,... ,4 and the phenotypic counts, the conditional expected genotypic counts are easily obtained.
- New haplotype frequency estimates then are the estimated proportions of each haplotype using the expected phased genotype counts.
- Homozygous individuals (both loci): for example an A1A1, B2B2 individual segregates only A1B2 haplotypes, regardless of recombination ρ.
- Homozygote/Heterozygote: for example, an A1A1, B1B2 individual segregates on A1B1 or A1B2 each with probability 1/2 regardless of ρ.
- Only the double heterozygote A1A2, B1B2 is *informative for linkage;* the segregation probabilities depend on ρ. That is, this individual passes each of the four haplotypes A1B1, A1B2, A2B1 and A2B2, with probabilities (1– ρ)/2, ρ/2, ρ/2 and (1 – ρ)/2 if his genotype is A1B1/A2B2, and probabilities ρ/2, (1–ρ)/2, (1 – ρ)/2, and ρ/2 if his genotype is A1B2/A2B1.

3.1.4 Allelic Association

- A measure of allelic association between the two loci is
  . D = P(A1B1) – P(A1) P(B1) = q1 – (q1 +q2)(q1+q3)
  . = (q1q4 – q2q3) since q4 = 1 – (q1+q2 +q3).
- This measure is known as the coefficient of *linkage disequilibrium (LD)*.
- Allelic associations between loci arise from population structure, admixture and history, or from selection.
- Example of mixture/subdivision: Suppose one subpopulation has high A1 and B1 allele frequencies, and the other high A2 and B2. Even with no LD within subpopulations, in the overall population haplotypes A1B1 and A2B2 will have higher frequencies than expected if the overall D=0.
- Example of original mutation on some genetic background: Suppose a population has A1 and A2 at locus A, but only B1 at nearby locus B. New mutation B2 arises, say on A1 haplotype, and by random drift increases in frequency. Until recombination with an A2B1 haplotype occurs, there can be no A2B2 haplotypes. This may be a long time if ρ is small. In terms of frequencies, the association of A1 with B2 and A2 with B1 will be maintained much longer.
- Associations are, however, maintained by tight linkage (ρ ≈ 0).



Linkage (ρ small) maintains LD

- Suppose current haplotype frequencies are q1, q2, q3 and q4, and at next generation are q1*, q2*, q3* and q4*.
- Now, for example, an A1B1 offspring haplotype was transmitted
  - with probability 1 by any A1B1/A1B1 parent,
  - with probability 1/2 by any A1B1/A1B2 or A1B1/A2B1 parent,
  - with probability $(1 - \rho)/2$ from an A1B1/A2B2 parent, and
  - with probability $\rho/2$ from an A1B2 /A2B1 parent.
- Thus $q1^* = q1^2 + 2\,q1(q2 + q3)/2 + 2\,q1\,q4\,(1 - \rho)/2 + 2\,q2\,q3\,\rho/2$
  . $= q1(q1+q2+q3+q4) - \rho\,(q1\,q4 - q2\,q3) = q1 - \rho\,D.$
- Analogously, $q2^* = q2 + \rho\,D$, $q3^* = q3 + \rho\,D$ and $q4^* = q4 - \rho\,D$.
- Thus, in expectation, allele frequencies are unchanged
  . $(q1^* + q2^* = q1 + q2$ etc.) and
  . $D^* = q1^*\,q4^* - q2^*q3^* = (q1 - \rho\,D)(q4 - \rho\,D) - (q2 + \rho\,D)(q3 + \rho)$
  . $= D - \rho\,D\,(q1 + q2 + q3 + q4) + \rho^2\,(D^2 - D^2) = (1 - \rho)\,D.$
- Contrast with HWE: Even for unlinked loci equilibrium (D =0) is not achieved in one generation. Associations persist: they persist longer for small ρ.