MCB 142 Discussion

Hana Lee

November 14, 2008

1 Rationale

The linkage mapping experiments we've done so far benefited from the advantages of using model organisms, where researchers can do controlled crosses between individuals of known genetic background and obtain progeny in sufficient numbers to calculate recombination frequency with high statistical significance.

In humans, we no longer have those advantages. However, we can still utilize the variation present in the human population. We look at *phenotypic* variation (usually diseases although there are other human phenotypes of interest) and associate it with *genotypic* variation (by genotyping the individuals being studied at markers throughout the genome). Phenotypic information is readily obtained through medical histories and clinical observations. The lectures are focusing on how we determine the genotype at polymorphic markers and determine which markers are linked to our phenotype of interest, which will in turn help us figure out what gene causes that phenotype.

This sort of analysis can be useful for model organisms as well for studying:

- Complex traits: phenotypes that do not show simple Mendelian inheritance
- Multigenic traits: phenotypes that are regulated by multiple genes whose interactions are too complex to be dissected by epistasis analysis
- Natural variation: phenotypes present in the natural population rather than artificial mutations generated in the laboratory

2 Terminology

- haplotype, haplotype phase, ancestral haplotype
- polymorphism versus mutation
- marker, locus, gene
- physical map versus genetic map
- genetic heterogeneity, compound heterozygotes

- recurrent mutations

3 Techniques

How do the following techniques work and what are they used for?

- sequencing: Sanger sequencing, shotgun sequencing
- restriction fragment length polymorphism (RFLP), simple sequence repeat (SSR), single nucleotide polymorphism (SNP)
- polymerase chain reaction (PCR)
- fluorescent in situ hybridization
- positional cloning
- pedigree analysis
- linkage disequilibrium (LD) mapping
- genome-wide association mapping

4 Some Key Concepts

Your mapping resolution depends on two factors: the number of informative meioses in your sample set and the number of markers. Depending on the rate of recombination in the species, after a certain threshold, increasing the number of markers gives you diminishing returns because there aren't enough informative meioses in your sample set to pick out rare recombination events between two very closely linked loci.

When you notice that a disease phenotype is linked to a particular marker polymorphism, that polymorphism is usually not the actual sequence change that causes the phenotype. The resolution of your linkage mapping is usually not high enough that you can narrow down the location to a single gene. Hence, researchers usually look for candidate genes in the region to identify the causative mutation.

We previously used chi-square tests to calculate the statistical significance of linkage observed between two loci. You can also use LOD score, which compares the odds of your observed numbers of parentals (p) to recombinants (r) to the odds expected if the two loci were unlinked.

$$LOD = \frac{\left(\frac{p}{p+r}\right)^p \left(\frac{r}{p+r}\right)^r}{\left(\frac{1}{2}\right)^{p+r}} \tag{1}$$