**Chapter 13**

# Fine mapping and IBD mapping

*Julius van der Werf, Department of Animal Science*
*University of New England*



## Mapping resolution using linkage analysis

QTL mapping based on linkage analysis, as discussed in Chapters 6-12 can be applied for genome wide scans based on many markers. The evidence for the existence of a QTL is based on genetic linkage between the QTL and the marker. In most designs, individuals from 2 or 3 generations were marker genotyped, and the association between marker and QTL depends on their recombination frequencies. To find a marker reasonably close to a QTL we need a low recombination frequency. The mapping principle is based on finding those markers that have lowest recombination frequencies with the QTL (therefore having highest effect).

The problem with linkage analysis is that the mapping resolution is poor: around 20 cM. The problem is not that there are not enough markers, as marker densities can be around 0.5 to 3 cM with the current maps, and this density is still increasing (Georges and Anderson, 1996). The reason for the low resolution is that there are not enough observed meiosis in most experiments in order to distinguish between few and very few recombination events.

Maximal mapping resolution in a given pedigree corresponds to the interval between the two nearest flanking cross-overs (Boehnke, 1994, ref in Georges, 1999)
Darvasi and Soller (1997) presented an equation to determine the resolution of 'high density' mapping.

$$CI95 = 3000/kN\alpha^2$$

where CI is the 95% confidence interval for the map position (in cM)
k is the number of informative parents per individual
(k = 1 for HS and BC, k = 2 for F2)
N is the total effective mapping population size
$\alpha$ = the allele substitution effect at the QTL.

For example, in a half sib design (k = 1) for a QTL effect of 0.5 ($\alpha$ = 0.5 within HS family Standard Deviations), and 500 animals genotyped (N = 500), the CI95 would be equal to 24 cM. To achieve a CI95 of 2 cM would require 6000 animals genotyped. Usually, it is hard to obtain such large half sib families (except in dairy cattle). However, map positions with such precision are needed as 1 cM is equivalent to around 1 million base pairs or 30 genes, which is still very large if comparative mapping or positional cloning is considered.

Methods other than linkage analysis have been considered. For example, when using advanced interline crosses (F3, F4 etc) the number of crossovers is increased and gives more resolution to map QTL. However, advanced inter-crosses will take several generations to be established, which is not practical. Other methods to be discussed hereafter are 1) population wide linkage disequilibrium and 2) identity by descent (IBD) mapping

## Population wide linkage disequilibrium

Recall from Chapter 5 the definition of linkage disequilibrium:

Given bi-alleleic loci A and B with alleles A,a and B,b, the loci are said to be in linkage disequilibrium if the chance of finding a B is dependent on the alleles in A.
Linkage equilibrium if the

freq(AB) = freq (Ab) = freq (aB) = freq (ab).

And the amount of disequilibrium is measured as

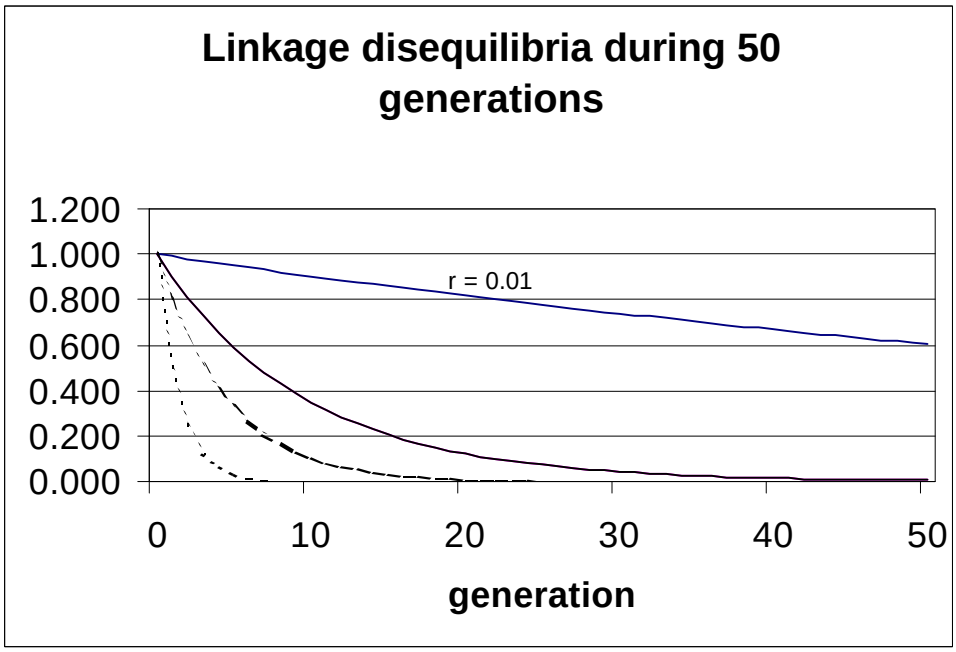
D = freq(AB).freq(ab) – freq(Ab).freq(aB).

A more general definition is that the probability of an allele at locus A is unconditional to the allele at locus B.

Now suppose the polymorphism at the QTL is due to a mutation from Q into q, at some point back in time. This mutation was transmitted from an individual to its offspring. The Q alleles was accompanied by marker alleles M, one of the alleles in the population at marker locus M. If the M and Q alleles are so close that they rarely recombine, than the mutant Q-alleles will mostly be accompanied by the marker alleles M. In linkage analysis, the probability of recombination between Q and M is based on one meiosis (between two generations), and linkage phase can be different in different families. However, most animals in a population might relate back to a common founder animal (or a common ancestor). In that case, for all animals within a population, there may be a LD between Q and M. In other words, if we find M in any arbitrary animal in the population, it might have an increased probability of carrying a Q allele. The linkage phase does not vary across generations and across families.

Let the mutant alleles Q be linked to the marker alleles M. In the offspring after the first mutant the frequency of recombinants (qM and Qm) is equal to r (where r = recombination rate). In the next generation we have $r.(1-r)+(1-r).r = 2r-2r^2$ recombinants, etc.

The next figure shows the linkage disequilibrium (D-values) between two loci depending on recombination fraction and generations since the mutation took place. It shows that with small recombination fractions, it takes many generations before, at population level, the linkage disequilibrium disappears.




Linkage disequilibrium has been used in human genetics. When a certain marker allele is close to a disease allele (i.e. when the affected individuals all carry the same allele at a certain marker locus), than that would be a sign that the marker would be linked to the disease gene. The wider the linkage disequilibrium (i.e. if the LD is valid over more genetically distant individuals, the closer the marker can be expected to be to the gene.

Some comments on the use of linkage disequilibrium:

Linkage disequilibrium is not only caused by physical linkage, but also by drift, selection, crossing (hybrydization) or by the large influence of founder individuals. As the effects of crossing and selection are more common in livestock, it may be expected that in livestock data LD exists between more distant loci (Haley, 1999).

Linkage disequilibrium would be a useful method if there were only two alleles at the QTL (one mutation), However, if there are more alleles at the QTL it will be more difficult to detect LD. IN the human literature there is evidence that many diseases indeed are caused by multiple alleles, although other have argued that for quantitative

traits it is more likely that most variation is due to only a few alleles (Haley, 1999). However, with quantitative traits, with part of the variation due to environmental effect, QTL alleles can not be so easily identified or inferred upon, making LS studies possibly not very powerful.

## Identity by descent mapping

AS LD mapping, IBD mapping is based on observing linkage disequilibrium between QTL and close markers loci. Hence, we determine proximity based on the number of cases where marker alleles and QTL alleles have not recombined. This can be contrasted with linkage analysis where we need recombinations to estimate map distance.

In IBD mapping we use pedigree information in order to look at historical crossovers (back pedigree). Within a pedigree, genes are in linkage disequilibrium, but the region where this applies to becomes smaller when IBD probabilities become smaller. In a particular genomic region, e.g. around the QTL, we may expect that sequences are be shared by individuals from the same founders. Because of crossovers, this region will become smaller if more generations have passed since the common ancestor (e.g. founder). Hence, the deeper the pedigree, the finer we can map markers near QTL's.

The probability of no recombination at a genetic distance X from a gene is $(1-X)^N$, where N is the number of meiosis events. Suppose we have two progeny from the same great grand parents. The value for N would be 6. The probability of no recombination would be 0.53 for a region of 10cM. With larger N, the size of the conserved regions would be smaller. For example, for N = 100 the probability of no recombination would be 5% for a region not greater than 3 cM. The number of 100 meiosis can be achieved in a very deep pedigree (lots of generations) but also in one half-sib family of 100 progeny. However, in a general pedigree, a combination of these is possible, making such values for N not unrealistic.

The idea behind IBD mapping is to find a region that is conserved in individuals with a common ancestor. Hence, a sequence of marker alleles at different loci remains intact. The method has been used in human genetics, where allele sequences (haplotypes) were related to individuals that were related and had all received a disease gene. The method tries to find a region that is Identical by Descent (IBD) among the affected individuals, since such a region may carry the disease gene. The IBD region is detected by closely linked marker loci that carry identical alleles at this region in the diseased individuals.

There are few references that have discussed IBD mapping for quantitative traits. Meuwissen and Goddard (2000) proposed a regression of phenotype on marker haplotypes. The haplotype effects were taken as random effects and covariances between haplotypes were derived, depending on the position of the QTL in the

haplotype. The expected covariances between the haplotype effects are proportional to the probability that the QTL position is identical by descent (IBD) given the marker haplotype information. The most likely position was found using Restricted Maximum Likelihood.

A problem with IBD methods is that it is quite complicated to determined IBD probabilities, particularly in a general pedigree where many marker genotypes may be missing or non-informative. Meuwissen and Goddard used a gene drop procedure, which might become problematic for large pedigrees. Henshall et al (1999) have proposed an efficient algorithm (GEIC: Genotype Elimination through Inheritance State Constraint) that might be useful for such problems.

## References

Darvasi, A. and M. Soller. 1997. A simple method to calculate resolving power and confidence interval of QTL map locations. Behavor. Genetics. 27:125-132.

Georges, M. and L. Andersson. 1996. Livestock genomics comes of age. Genome Research 6: 907-921.

Georges, M. 1999, From phenotype to genotype: towards positional cloning of quantitative trait loci. in livestock. In Dekkers et al. (Eds). Proc. of the Conf. “From Lush to Genomics. Visions for animal breeding and genetics. Iowa. State University

Haley, C.S. 1999. Advances in QTL mapping. In Dekkers et al. (Eds). Proc. of the Conf. “From Lush to Genomics: Visions for animal breeding and genetics. Iowa. State University.

Henshall, J.M., B. Tier and R.J. Kerr. 1999. Inferring genotype probabilities for untyped individuals in complex pedigrees. Proc. of the Assoc. of Advmt. Anim. Breed. genet. Vol. 13:329-332.

Meuwissen, T.H.E. and M. E. Goddard , 2000. Fine Mapping of Quantitative Trait Loci Using Linkage Disequilibria With Closely Linked Marker Loci . Genetics 155: 421-430.