HALEY-KNOTT REGRESSION MAPPING OF QUANTITATIVE TRAIT LOCI IN EXTENDED PEDIGREES

K. G. Dodds

AgResearch, Invermay Agricultural Centre, Private Bag 50034, Mosgiel, New Zealand

SUMMARY

The use of multiple generations of a founder's family to test or confirm the existence of a QTL is investigated. This is achieved via an extension of the Haley-Knott procedure to calculate the expected number of copies of each of the founder's alleles, at a particular genome position, in each descendant.

Keywords: Quantitative traits, regression, extended families

INTRODUCTION

Mapping quantitative trait loci (QTL) in livestock species usually proceeds by using the phenotypes of a single generation of one or more large families. QTL are detected by finding a phenotype difference between those individuals receiving each of a parent's alleles at a particular genome position. The data can be analysed using likelihood methods (Georges *et al.* 1995) or least squares methods (Haley and Knott 1992).

Generation of mapping populations large enough for adequate power is expensive and involves time delays. An alternative is to analyse extant general pedigrees, using segregation analysis or gamete relationship methods, incorporating information from markers. Such methods are under investigation (eg. Meuwissen and Goddard 1997), but are computationally demanding. Between these extremes is the situation where we trace the putative quantitative trait alleles of a particular "founder" individual to detect segregation of these alleles in its descendants. Unlike the segregation analysis methods, alleles from other individuals are essentially ignored. Data from grandprogeny of a sire were used to confirm the existence of a previously identified QTL (Arranz *et al.* 1998). Here we investigate the use of multiple generations of a family to test or confirm the existence of a QTL.

METHOD

The method of Knott *et al* (1996), which regresses phenotypes onto gamete probabilities for half-sib progeny of an individual, is extended to multiple generations. For simplicity we will focus on a single QTL position, although the method can be used for a genome scan by stepping this position along the genome. A founder of interest is chosen, and then for each descendant the expected number of copies of each of the founder's alleles is calculated from marker information.

Conditional expectations. We assume that the founder is genotyped with known marker phase, and that no individuals are inbred to the founder. Where there is reference to a parent or other ancestor it is taken to be the ancestor that is descended from the founder. The first step is to use marker information from all individuals in the pedigree to deduce which marker allele a parent passes to each of its progeny. Phase may be inferred from progeny, to assist in this deduction.

Let Q_i = the probability of individual i receiving the 'relevant' QTL allele from the parent,
x_{ij} = the number of copies of founder QTL allele j in individual i,
E_{ij}= the expectation of x_{ij}, and
$P(i)$ be the parent of i.

Probabilities and expectations are conditional on the marker information. For each individual, the closest informative markers flanking the QTL position are used. For progeny of the founder we take the 'relevant' allele to be that denoted as the first QTL allele. For subsequent generations the 'relevant' allele is the allele that the parent received from its parent in turn. We take the expectation of x_{ij}, rather than Prob(x_{ij} = 1) as used in the Haley-Knott procedure, to allow for an individual that is inbred to the founder, and therefore may have more than one copy of a QTL allele of the founder.

The Q_i for progeny of the founder can be calculated as in Knott *et al* (1996). Then $E_{i1}=Q_i$ and $E_{i2}=1-Q_i$. For a marker to be informative in subsequent generations, it must be possible to determine which allele the individual received from its parent, and which allele the parent received from its parent. An informative marker is treated as non-recombinant if an individual has received a grandparental allele, recombinant otherwise. Then $E_{ij}= Q_i E_{P(i)j}$, ie. Q_i is multiplied by the genotype probabilities of the parent to give the genotype probabilities for the current individual.

When a marker genotype for an individual is missing, that marker is treated as uninformative for that individual. If the ungenotyped individual has a genotyped progeny, the marker will be uninformative if it contains an allele in common with its most recent genotyped ancestor (since we cannot be certain that the allele came from this ancestor). If the progeny does not contain an allele in common with the genotyped ancestor, then ancestor's QTL alleles could only be passed on if there was recombination between that marker and the QTL, so the marker is scored as recombinant in the progeny.

Example of calculations. For simplicity an example with only two markers (flanking the putative QTL) and with only founder descendants genotyped is considered. The example pedigree is shown in Figure 1. Suppose we wish to test for a QTL placed between the first and second markers with recombination fractions of θ_1=0.1 and θ_2=0.2 respectively, so that the recombination fraction between markers is θ=0.26 (using Haldane's mapping function). We assume that the founder (Animal 1) has marker phase AA/BB. Genotype probabilities for the progeny of the founder are calculated as described in Knott *et al* (1996): Table 1 shows these calculations to determine the expectation for the founder's first QTL allele; the expectation for the second QTL allele is one minus this value. For these individuals the haplotype indicates whether the individual received the first (N) or second (R) marker allele from the founder. For the next generation the haplotype denotes whether the parent's allele came from the founder (N) or not (R). A 0 in the haplotype denotes non-informative cases. The Q_i for this generation are then calculated as for the previous generation, and the result is multiplied by the parent's E_{ij}s to give the individual's E_{ij}s (Figure 1).

Examples of uninformative markers are: Animal 8 - heterozygous for the same alleles as its parent; Animal 11 - parent is homozygous for the marker; Animal 10 - parent heterozygous for the same alleles as its grandparent. If Animal 3 had sufficiently many genotyped progeny to determine the

phase, the Marker 2 allele received from the sire could be determined, and this marker would then be informative.

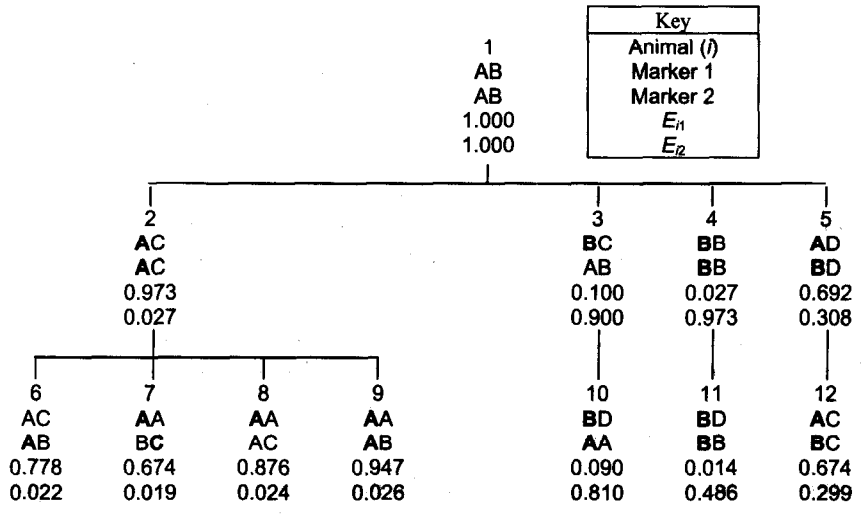

Figure 1. Pedigree for example calculations. Only links back to the founder (Animal 1) are shown. Alleles shown in bold denote which allele the parent passed to its progeny.

Table 1. Example calculations for pedigree in Figure 1

Haplotype	Q_i	Animals (i)
NN	$(1-\theta_1)(1-\theta_2)/(1-\theta)=0.973$	2,9,12
NR	$(1-\theta_1)\theta_2/\theta=0.692$	5,7
RR	$\theta_1\theta_2/(1-\theta)=0.027$	4
N0	$(1-\theta_1)=0.900$	8,10
R0	$\theta_1=0.100$	3
0N	$(1-\theta_2)=0.800$	6
00	0.500	11

Regression analysis. To perform the QTL analysis, the phenotypes are regressed on the two conditional expectations, to determine whether at least one of these has a non-zero slope. The slope estimates the effect of a substituting a randomly chosen allele by the corresponding founder allele.

DISCUSSION

Inbreeding. Individuals that are inbred with respect to the founder need to have E_{ij}s calculated for the lines of descent from both the sire and dam. These are then added together.

Conditional expectations. By definition, the founder's E_{ij} are both 1. For progeny of the founder $E_{i1}+E_{i2}=1$, so if this is the only generation included in the analysis, regression on each of these will be confounded. If no other generations have marker data, there is partial confounding, and it is advocated that either of the E_{ij} be used, but not both together. No E_{ij} will exceed that of a parent, unless the individual is inbred, in which case it will not exceed the sum of the respective E_{ij}s for the parents. The method traverses down the pedigree from the founder. In some cases there may be information from progeny, or from allele frequencies which could be of use in calculating the E_{ij}. Apart from possibly inferring phase this information is ignored, and is unlikely to warrant the extra computation.

Phase. For large half-sib families, Knott *et al* (1996) found that inferring phase had little effect on the results. In the current situation, there is the possibility of inferring phase based on very few progeny. The consequences of this have not been studied, and it is recommended to set some minimum criteria on using inferred phase.

Regression Analysis. The regression analysis could be performed by least squares or mixed model methods (sire model or animal model), the choice will depend on the data structure. If the founder has only a few progeny, and each of these has a large number of descendants, there could be partial confounding between the E_{ij} and the breeding values of the founder's progeny. It may be possible to reduce such an effect by using an animal model analysis with additional relatives outside the founder's family. However the results of this analysis would be influenced by the heritability used.

Conclusion. A method to perform QTL analysis within families spanning multiple generations has been presented. Unlike half-sib or similar designs, the founder does not need to be heterozygous for the QTL. However only segregation involving the founder's alleles may be detected. Further investigation is required to determine the consequences of inferring phase, of including many ungenotyped individuals across several generations, and to find the best ways to include non-QTL genetic effects and multiple founder families in the regression models. A measure of information content would also be useful to direct genotyping efforts and to compare alternative designs.

REFERENCES

Arranz J.-J., Coppieters W., Berzi P., Cambisano N., Grisart B., Karim L., Marcq F., Moreau L., Mezer C., Riquet J., Simon P., Vanmanshoven P., Wagenaar D. and Georges M. (1998) *Anim. Genet.* **29**:107

Georges M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A.T., Sargeant, L.S., Sorensen, A., Steele, M.R., Zhao, X.Y., Womack, J.E. and Hoeschele, I. (1995) *Genetics* **139**:907

Haley, C.S. and Knott, S.A. (1992) *Heredity* **69**:315

Knott, S.A., Elsen, J.M. and Haley, C.S. (1996) *Theor. Appl. Genet.* **93**:71

Meuwissen T.H.E. and Goddard M.E. (1997) *Genetics* **146**:409