Parasite Resistance: A genome scan approach to finding markers and genes

A.M. CRAWFORD, J.C. McEWAN,[1] K.G. DODDS,[1] S.A. BISSET,[2] P.A. MACDONALD,
K.J. KNOWLER,[3] G.J. GREER,[1] R.S. GREEN,[2] R.P. CUTHBERTSON, C.S. WRIGHT,[1] A. VLASSOFF,[2]
D.R. SQUIRE,[2] C.J. WEST,[2] K.A. PATERSON AND S.H. PHUA

AgResearch, Molecular Biology Unit, Centre for Gene Research and Department of Biochemistry,
University of Otago, Box 56, Dunedin, New Zealand.

INTRODUCTION

A systematic approach for finding regions of the genome that are important in the control of productive traits has only recently become available. The advent of microsatellite markers for sheep and other domestic animals (Crawford *et al.* 1991) and their subsequent placement on linkage maps (Crawford *et al.* 1995, Kappes et al 1997) has provided a set of highly informative markers densely spread across the genome. With these reagents we now have the ability to follow the inheritance of each region of every chromosome in any cross. Providing we have pedigrees in which genes which have an important influence on the trait of interest are segregating, we can systematically identify the regions of the genome involved. This process has become known as the genomic scan.

Resistance to nematode parasites in sheep is associated with the development of an immune response. Once they have developed an immunity to parasites healthy adult sheep have a low nematode parasite burden. Animals resistant to parasites therefore, are those capable of rapidly developing an appropriate immune response, whereas susceptible animals are slow at responding appropriately to parasite challenge. The control of the type, rapidity and size of the immune response is very complex. There are so many possible candidate genes that could conceivably affect this process that a genomic scan is a far more efficient approach to finding genes affecting parasite resistance than taking genes involved in the immune response and checking them individually. This paper describes the pedigrees and markers used for a genomic scan for parasite resistance. Results of the scan of sheep chromosome 1 are then used to illustrate the type of data obtained, its interpretation, limitations and further uses.

METHODS

Selection lines: The Wallaceville divergent faecal worm egg count (FEC) selection lines of Romneys commenced in 1979 (Baker *et al.* 1991). Selection is currently based on FEC levels after natural challenge, twice between weaning (at approximately 3 months old) and 6 months old, in both ram and ewe lambs. At present, the high line has about an 8 fold higher FEC than the low line during the autumn. On a logarithmic scale this translates to a 2.6 σ_p divergence between lines. We would expect the lines to differ in allele frequencies at a quantitative trait locus (QTL) for host resistance that was segregating in the base population. The selection lines were too small to generate adequate backcross pedigrees in one year, so initially an outcross was used to search for regions of the genome segregating with resistance or susceptibility to parasites. Backcrosses are currently being generated.

Outcross Pedigrees: Reciprocal crosses between the Wallaceville divergent selection lines were made to generate rams which had a high chance of being heterozygous at QTL for parasite resistance. These rams were then each mated (via artificail insemination) with between 100 and 300 unselected Coopworth ewes to generate 5 half-sib pedigrees (Table 1). The progeny of these animals should exhibit considerable variation in their resistance to parasites depending on whether they received "resistance" or "susceptibility" alleles from their sire.

TABLE 1: Half-sib outcross sheep families used to detect QTL for parasite resistance.

Family	Sire	Number of progeny
1	92/066	225
2	92/153	175
3	92/154	348
4	92/155	111
5	93/124	101

Phenotype measurement: Two natural field challenges of infective nematode larvae (Baker *et al.* 1991) were given to all outcross lambs. Faecal egg counts were determined using the McMaster method with 3 separate samples taken over 5 days at the end of each challenge and results averaged. Mean strongyle and *Nematodirus* infection levels (epg) of 1091 (FEC1) and 70 respectively were achieved for the first challenge and 1462 (FEC2) and 36 respectively for the second challenge. Live weight measurements were taken at birth, weaning, and after each challenge. *T colubriformis* L3 antibodies were measured by ELISA (Douch *et al.* 1994) from serum samples taken at approximately monthly intervals being at the beginning middle and end of each challenge. Dagginess was scored at weaning, and after each challenge. At slaughter, which

[1]AgResearch, Invermay Agricultural Centre, Private Bag 50034, Mosgiel, New Zealand.
[2]AgResearch, Wallaceville Animal Research Centre, Box 40063, Upper Hutt, New Zealand.
[3]AgResearch, Woodlands, Private Bag 90121, Invercargill, New Zealand.

took place approximately a week after the end of the second challenge, the abomasum and the first 10m of small intestine were collected and adult parasites in the lumen were identified to genus and counted.

Genotyping and analysis: All of the markers used in this analysis were microsatellites, which are small regions of the genome that contain dinucleotide repeats. These are amplified using the polymerase chain reaction (PCR) and variation in the length of the amplified fragment is used to distinguish the different alleles at each microsatellite locus (Weber and May 1989).

Whenever a sire was heterozygous at any of these loci, his 22 most susceptible and 22 most resistant progeny were genotyped to determine, where possible, which of his alleles they had received. Five techniques of selecting the extreme progeny of each sire for genotyping were examined including mixed model repeatability analysis. Little difference was observed between the various methods. The final method chosen used the log-transformed mean of the 3 faecal egg counts ($\log_e[(x1 + x2 + x3 + 50)/3]$) recorded for the two challenges for both the strongyle and *Nematodirus* counts, with the latter also subsequently scaled to equalise variance between contemporary groups (McEwan *et al.* 1997). Animals were then segregated into groups based on the weighting of the first principal component derived from a principal components analysis of the residuals (from a least squares model fitting the fixed effects of contemporary group, sex, sire, birth and rearing rank and birth date) of the above four traits.

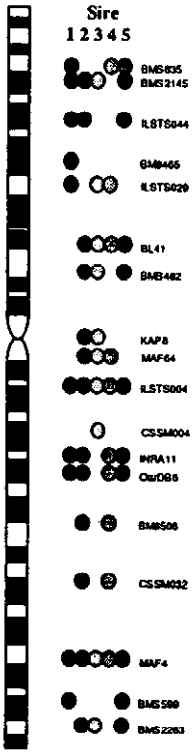
We used single marker regression methods to test for allele within sire effects for initial analyses (Soller, M. 1990). This has the advantages of speed, flexibility of model fitted and the ability to analyse each marker locus on a incremental basis as they are genotyped. Subsequently, regions of interest were analysed in more detail with ANIMAP (Georges *et al.* 1995, Nielsen *et al.* 1995). This program uses the map and half-sib family data from multiple markers simultaneously to test for and estimate the size and position of segregating QTL.

RESULTS

Details of all chromosome 1 loci screened are shown in Figure 1. The loci are shown at their approximate position on the chromosome based on the linkage map (Crawford *et al.* 1995; DeGotari *et al.* in preparation). All sires were tested to determine which were heterozygous at each of the loci. The five vertical lines of filled circles to the right of the chromosome ideogram indicate which sires were heterozygous for which loci, and therefore the locus and family combinations for which resistant and susceptible progeny were genotyped.

It can be noted from Figure 1 that coverage of the genome is different for each of the families. In particular, family 1 has a 76 cM region between ILSTS029 and ILSTS004 and family 3 a 100 cM region between CSSM4 and MAF4 where segregation analysis could not be undertaken as all loci screened between these markers were homozygous.

FIGURE 1: Sheep chromosome 1 showing the location of all the markers used in the genomic scan. The five vertical lines of filled circles to the right of the chromosome ideogram indicate which sires were heterozygous for which markers, and hence had the 22 most resistant and 22 most susceptible progeny genotyped for that marker.



The results from the single marker regression methods for markers on sheep chromosome 1 are shown in Figure 2. Although many other phenotypic measurements were taken, the four major measures of parasite load; abomasal *Trichostrongylus* spp. ($\log_e[X+100]$ transformed), small intestinal *Trichostrongylus* spp. ($\log_e[X+100]$), FEC1 ($\log_e[X+50]$) and FEC2 ($\log_e[X+50]$) are presented. The highest F statistic obtained (13.4) was with the marker MAF64 in sire family 3 with the trait FEC1 ($\log_e[X+50]$). The five measurements of serum antibodies to L3 *T. colubriformis* and the specific and total serum IgE measurements levels showed no evidence of segregation with any of the chromosome 1 markers. The highest F statistic for any of the marker/ family/ phenotype measurements was 6.2

To further test this result a multipoint analysis (Figure 3) was undertaken using ANIMAP. Because it is computationally intensive a maximum of 10 markers can be analysed at any one time. The result shown in Figure 3 is for the 10 markers spanning the region around MAF64 from ILSTS029 to CSSM32. The multipoint analysis is also suggestive of segregation occurring around MAF64 in sire family 3.

DISCUSSION

Genomic scan experiments, with outbred animals such as sheep and when measuring highly variable

FIGURE 2: F scores for comparing least squares means of progeny receiving different sire alleles for a) Abomasal *Trichostongylus* spp. (Log_e[X+100] transformed), b) Small Intestinal *Trichostrongylus* spp. (Log_e[X+100]), c) FEC1 (Log_e[X+50]) and d) FEC2 (Log_e[X+50]) for sires 920066 (+), 920153 (x), 920154 (*), 920155 (#), 930124 (o). For p=0.05 (multiple genomic test) the critical value is between F=12.1, and F=14.0, depending on the family.

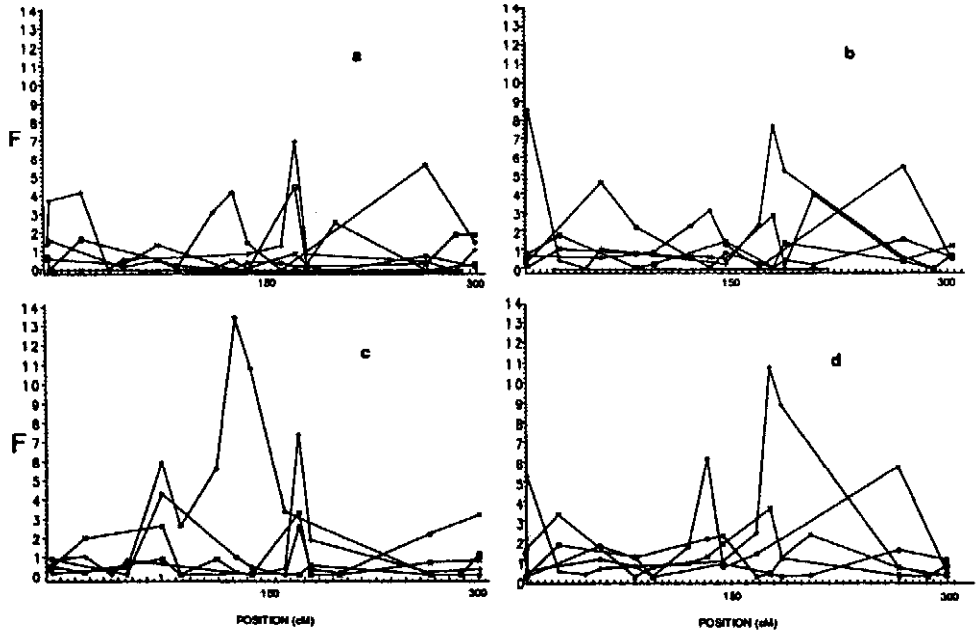
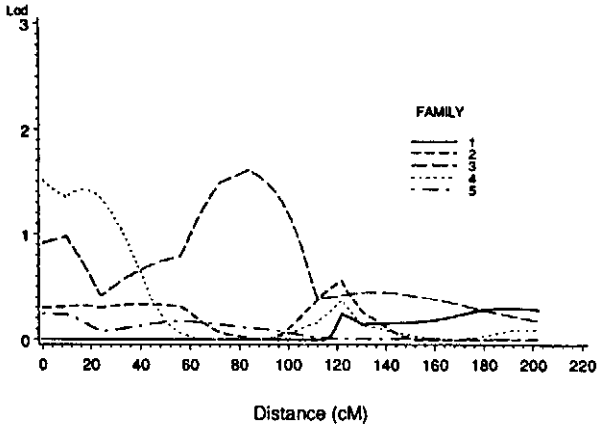


FIGURE 3: Multi-point linkage analysis of sheep chromosome 1 for the trait FEC1 (Log_e[X+50]). This distance is that from ILSTS029 towards CSSM32.



phenotypic traits such as faecal egg counts, are of necessity very large scale. They require hundreds if not thousands, of animals each of which must be accurately measured for the trait of interest. In the case of parasite resistance this work also involves a great deal of expertise in both parasitology and immunology. In addition the animals must be genotyped at approximately 200 loci to give reasonable coverage of the sheep genome. Even if only the extreme animals are genotyped from each pedigree it is common for most of these experiments to require a minimum of 20,000 to 30,000 genotypes. The numbers required are a trade off between the cost of the experiment and the size of the QTL which can be detected. A lower bound can be set, as smaller QTL are likely to be worth less economically. In the current experiment the design criterion was that there would be a 50% chance of detecting a QTL within any half-sib family if it had an allele substitution effect larger than 0.6 of a phenotypic standard deviation, and a greater than 95% chance if the effect was greater than 1.0 phenotypic standard deviation. Given the resources required, such work should only be undertaken when there is reasonable evidence that segregation is occurring for the trait of interest. The animals used in the current experiment satisfied this criterion, because of their selection history, the divergence observed after selection, the results of a segregation analysis on these flocks (McEwan *et al.* 1997) and the fact that a preliminary DNA screening exercise using RAPD's had provided some evidence of large allele frequency differences between the selection lines (Pulford *et al.*, 1995).

It has been suggested that markers spaced at 30 cM intervals will give good coverage for any genome scan (Darvasi and Soller 1994). For a marker to be informative the sire must be heterozygous. If the sire is found to be homozygous for all markers in a particular region then good overall coverage cannot be achieved. The preferred solution is to increase marker density on the linkage map. This allows an increased choice of markers and increases the chances that at least one marker will be informative for all regions of the genome. With the current set of markers available for genomic scans (approximately 600 microsatellites) we are achieving between 60 and 70 % coverage of the sheep genome in any particular family scan. It has been estimated that we would require 1000 more markers to achieve coverage greater than 90% with a high degree of certainty. However, while availability of additional markers would be useful, the current experiment was designed to account for both this factor and the possibility that the alleles were not fixed in the parental lines by using several sires so that if QTL were present segregation would be detected in at least one family.

There is considerable debate amongst geneticists as to what constitutes significant linkage of a marker locus to a trait (Lander and Schork 1994). With the large number of marker genotype/trait phenotype comparisons being made in a genomic scan the chance that type I errors (false positives) will occur is very high if inappropriate significance levels are used. The results presented here demonstrate two important aspects of this problem. An F statistic of 13.4 for MAF64 in Sire family 3 from the single marker analysis has a single test significance of P=0.0004. However simulations using the genotypes generated by the *current work* suggest that the true significance is P<0.05 when comparisons are done on a genome wide basis, because of the large number of tests involved (Wright *et al.* 1996). A second aspect is that single point analyses do not make full use of the information available and are therefore less sensitive than more sophisticated methods. ANIMAP, a multipoint analysis method, utilises additional information about the inheritance of the chromosomal region by using several markers in a combined analysis and can also use inferred dam information. In the current example the ANIMAP results produced a non-significant result (LOD score 1.6 the real P value is well over 0.1) for this region of the genome in sire family 3. Because none of the other four families show evidence of linkage in this region it would be difficult to justify further research on this area of the genome on the basis of this study.

The choice of an appropriate threshold is very important because of the likely costs incurred if incorrect decisions are made. The benefit of the genomic scan is that, unlike the candidate gene approach, it can systematically exclude regions of the genome from containing QTL alleles above a certain experimentally selected size and population frequency for the population examined.

The search for genes responsible for complex multigenic traits is likely to have different outcomes depending on the trait that is being examined. Some traits may be due to a small number of genes with large effects in which case it is likely that a genomic scan will identify the regions containing them. If however the trait is due to the additive effect of a large number of genes each with a small effect the genomic scan is unlikely to identify any region linked to the trait. Segregation analysis using programmes such as FINDGENE (McEwan *et al.* 1997) attempts to identify those pedigrees where genes of large effect are present. Given the high cost of genomic scans it would be prudent to undertake some form of segregation analysis in the pedigrees prior to the commencement of marker analysis or even generation of resource families. In the current work efficient segregation methods were not available at the commencement of the trial. In retrospect such an analysis would have suggested that the mode of inheritance could be recessive and therefore backcross pedigrees would be more efficient even given the delays involved in generating sufficient progeny.

In this report we presented an example of a 200 cM region with little evidence that segregation for the host resistance for parasite resistance is occurring, because it clearly illustrates the issues underlying the design, analysis and interpretation of the experimental results obtained. The majority of the rest of the approximately 3000 cM length sheep genome examined but not presented produced similar results. In several cases apparent segregation was detected, but additional work will be required before we can confidently report that QTL have been identified.

After QTL are detected further work will usually be required before the results can be used by industry or even published. Initially, the QTL region may not be well localised spanning a range of 20-40 cM with moderate confidence (similar to the region corresponding to the "bump" in Figure 3). To more closely localise the region of interest a three pronged approach needs to be used. These factors include: genotyping all animals (rather than just the extremes) in the regions of interest, the development of further polymorphic markers specific to the locality in an attempt to make additional pedigrees fully informative, and testing the validity of the results in independently generated populations to both verify and more closely define the likely QTL location. If this work is successful, marker assisted selection (MAS) may become a viable addition to existing genetic selection tools, where closely bracketing marker loci are available and sufficiently polymorphic, without the need for identifying the actual gene involved (Meuwissen and Goddard 1996).

The economic benefits depend on the value of the trait, the inheritance pattern of the QTL identified, the nature of the bracketing loci available and the ease, cost, proportion of the animals needing to be measured and/or genotyped, and age when phenotypic measurement normally occurs. Selection for host resistance to internal parasites is a good candidate for this selection technique, because existing methods involve exposing the animal to disease with consequent production losses and stock management difficulties. In addition faecal samples are labour intensive to collect, highly variable and expensive to analyse. However, prior to industry use a formal economic evaluation would need to be undertaken for each QTL identified.

CONCLUSIONS

We have outlined how we are currently conducting a genome scan in the search for QTL for host resistance to internal parasites. Great care has to be taken in the design, analysis and interpretation of these experiments because of their size and cost. Before attempting to undertake such an experiment, it is desirable to have evidence that QTL may be segregating in the population being examined. Typical results were presented for a region showing little real evidence of segregation to demonstrate the methodology currently used to phenotype, genotype and analyse the information collected. While the number of available polymorphic markers has increased dramatically in recent years it is still currently a limiting factor in these studies. The primary intention of this work is to utilise any QTL identified for MAS and we have alluded to some of the factors involved in determining whether their use will be financially viable. Disease resistance traits in general would appear to be ideal candidates for this technology.