Finding QTL without markers: experience with FINDGENE

J.C. McEWAN, C.A. MORRIS[1], S.A. BISSET[2], R.S. GREEN[2] AND R.J. KERR[3]

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

ABSTRACT

FINDGENE is a mixed inheritance model computer program which can analyse extensive animal pedigrees. It partitions the observed variance into fixed effects, polygenic inheritance, a two allele major locus effect and a residual environmental variance. It does not require genotype information. Data from three New Zealand sheep experiments where faecal nematode egg count (FEC) information had been collected were examined by this method. They included a long term selection experiment for FEC in Romneys at Wallaceville, another in Perendales at Ruakura, and a randomly mated Coopworth flock at Woodlands Research Station. All FEC traits were initially transformed using the function $\log_e(x+100)$. The final model included the effect of sex/year contemporary group, birthday and polygenic inheritance in addition to a 2 allele Mendelian gene effect. Results from all three data sets suggested a dominant autosomal gene for host susceptibility was affecting strongyle FEC values in the summer/autumn period with estimates of 0.37, 0.69 and 0.72 log eggs/g for the single allele effect from Romney, Perendale and Coopworth flocks respectively. Corresponding values for dominance were 0.47, 0.44 and 0.88 log eggs/g with susceptible allele frequency 0.14, 0.55 and 0.69. When considered together, these results suggest that at least one QTL is affecting the development of host resistance to internal parasites in young sheep. The identification and characterisation of this QTL may result in useful novel methods of internal parasite control.

Keywords: QTL; FINDGENE; parasite; nematode; sheep; host resistance.

INTRODUCTION

A large amount of research is currently underway to identify and locate polymorphic genes whose alleles have a large effect on productive traits in humans and livestock. Such genes are commonly called quantitative trait loci (QTL). Progress in sheep has been enhanced by the availability of mapped highly polymorphic DNA markers (Crawford *et al.*, 1995) which makes comprehensive linkage studies possible. While linkage studies are highly effective in locating and describing loci if a QTL is segregating, they are not an efficient method to screen large populations. Rather it is preferable to use a type of analysis, commonly referred to as "segregation analysis", which does not involve genotyping. This study reports on a computer program 'FINDGENE', which implements a form of segregation analysis especially designed for use in large animal populations. FINDGENE was used to characterise putative QTL for host resistance to internal parasites and screen individuals on their genotype status for the locus

METHODS

The outline of the procedures used in the computer program has been described in Kinghorn *et al.* (1993) with modifications described by Kerr *et al.* (1995). Briefly, the FINDGENE mixed inheritance model definition is:

$$Y = X\beta^* + ZQ^* g^* + Zu^* + e^*$$

where Y is a vector of observations, X is a design matrix relating observations to fixed effects, ß is a vector of fixed effects, Z is a design matrix relating observations to animals, Q is a design matrix of n x 2, with n being the number of animals containing the probabilities of carrying one or two copies (q_1 and q_2) of the putative QTL allele, g is a vector of QTL genotype effects, u is a vector of polygenic breeding values and e is a vector of residual errors. All terms labelled with an asterisk have to be estimated from the data. The method uses a 2 step iterative procedure. Initially Q^* is estimated from the individual's own phenotype, relatives information and initial parameters provided, using a form of iterative peeling described in Kerr and Kinghorn (1996). In the second step (ß*, g*, u* and e* are estimated using best linear unbiased prediction (BLUP). Because BLUP is essentially regression, all regression variables X, Z, and Q* are assumed independent to Y. Because Q* is not independent (Y is used to calculate Q* in the previous step) estimates of ß*, g*, u^* and e^* will be biased. To overcome this problem a correction is applied in each BLUP iteration such that all best linear unbiased estimates (BLUEs) and BLUPs are close to unbiased. The two steps are repeated until appropriate convergence criteria are satisfied.

The criterion r =variance(QTL)/[variance(QTL) + variance(residual error)] was used to infer the significance of the results. Suitable significance thresholds were derived from simulation and are only approximate. Values greater than 0.2 were considered to be evidence of segregation.

Three data sets were examined in the current work. The first was the Wallaceville Romney strongyle faecal

[1] AgResearch, Ruakura Agricultural Centre, Private Bag 3123, Hamilton, New Zealand.
[2] AgResearch, Wallaceville Animal Research Centre, PO Box 40063, Upper Hutt, New Zealand.
[3] Department of Animal Science, University of New England, Australia, New Zealand.

nematode egg count (FEC) selection flocks (Baker *et al.*, 1991). Briefly these flocks, which commenced in 1979 from a common genetic base, were selected as lambs for consistently high or low FEC following an extended period of exposure to a natural field challenge. In the initial years lambs received no anthelmintic treatment during the test period, but subsequently the procedure involved the lambs being drenched at weaning and then monitored until the flock mean FEC rose to around 1500 eggs per gram of faeces (epg), when all individuals were sampled and drenched. This process was repeated twice more before late autumn/early winter. Sampling occurred at approximately 4 (FEC1), 6 (FEC2) and 7 (FEC3) months of age. Parasite infections consisted mainly of mixed *Ostertagia* and *Trichostrongylus* spp. infections although *Haemonchus contortus* was occasionally also present. Data from the 1979 to 1992 years were used, comprising a pedigree of 1897 individuals from 65 sires and 563 dams.

The second data set was the Ruakura Perendale FEC selection lines (Watson *et al.*, 1992). This differed from the Romney flock, in that only two challenges were used post weaning and an artificial infection of *H. contortus* was used during the first challenge. Data from the 1984 to 1994 years were used in the analysis, comprising a pedigree of 1152 individuals, with 66 sires and 360 dams.

Information from the Coopworth progeny test flock based at Woodlands Research Station in Southland comprised the final data set. This flock used a variety of homebred and commercial rams and was not subjected to intensive selection during the trial period. FEC from 2 natural field challenges were recorded approximately 8 and 16 weeks after weaning at 10 weeks of age. The challenges consisted of predominantly *Ostertagia* spp. for the first challenge, although *Nematodirus* spp. (NEM) was also identified and recorded separately at this time, and mainly *Trichostrongylus* spp. in the second challenge. Circulating host antibodies to the secretory/excretory proteins of L3 *T. colubriformis* larvae (ELFC2) were also measured at the end of the second challenge (Douch *et al.*, 1996). Data collected from the 1990 to 1994 years were used, consisting of a pedigree of 5844 individuals with 113 sires and 1406 dams.

The FEC for all flocks and NEM egg counts from the Coopworths were transformed using the function $\log_e$(measurement + count unit) prior to analysis. The count unit was 100 epg for the Perendale and Romney flocks and 50 epg for the Coopworth flock. Fixed effects included contemporary group (year and sex) and birthday as a covariate. The transformations used and fixed effects fitted in the final FINDGENE model were based on preliminary analyses using general linear models. Tests of the residuals from these preliminary analyses suggested that they were either, not significantly different from being normally distributed, or any departures were minor. The starting parameters for FINDGENE were, heritability of 0.1, initial frequency for a non wild type allele at the putative locus of 0.5 and effects of one and two copies of this allele of 0.5 and 1.0 phenotypic standard deviations, respectively. Generally, the robustness of the convergence estimates was checked by varying the initial parameters used. In cases where convergence did not occur the starting parameters were modified to investigate whether convergence would occur over a more restricted range. Selection line effects were not fitted as a separate fixed effect, because this factor was implicitly corrected via the animal relationships included as part of the model.

RESULTS

The results from the Wallaceville data set are tabulated in Table 1. Decreasing numbers of FEC records were available between the first and last FEC challenge. FINDGENE converged for all FEC traits although for FEC1 either the estimate of the polygenic heritability had to be reduced to 0.05, or the allele frequency fixed at 0.08 for convergence to occur. Only the FEC2 trait exceeded the approximate threshold criterion and then only by a small margin. The estimates obtained from this challenge suggest an allele with slight over-dominance exists for FEC2 susceptibility (S) which was present at moderate frequency in the population (p=0.14), or alternatively a recessive allele is present for resistance (r). The primary component of the infection at this challenge time would be *Trichostrongylus* spp. Estimates obtained from the other challenge times suggest a similar mode of inheritance. Back transformed onto the original scale, Sr and SS groups would average 242 and 218 percent the FEC2 levels of rr individuals respectively.

TABLE 1: Estimates of the size ($\log_e$ eggs/g), mode of inheritance and frequency of host parasite resistance QTLs in the Wallaceville Romney FEC selection lines.[1]

	FEC1	FEC2	FEC3
pedigree (animals)	1897	1897	1897
observations (n)	1557	1523	1324
infection level (nil alleles)	6.76	7.03	6.95
a	+0.24	+0.37	+0.21
d	+0.47	+0.47	+0.36
p	0.08	0.14	0.23
r	0.11	0.21	0.11

[1]abbreviations: a= additive allele effect; d =dominance effect, p= frequency of the allele in foundation animals, r =variance(QTL)/[variance(QTL)+variance(residual error)]. FEC1, FEC2 and FEC3 are transformed (loge[x+100]) strongyle faecal egg counts at the first, second and third challenges respectively.

The results from the Ruakura Perendale data set are presented in Table 2. FINDGENE converged for both FEC traits with FEC1 exceeding the threshold. The estimates obtained from this challenge suggest a dominant allele exists for FEC susceptibility (S) which is present at a higher frequency than Wallaceville. The primary component of the infection would be *Haemonchus*. Estimates obtained from the second natural field challenge suggest a similar mode of inheritance. Back transformed onto the original scale, Sr and SS groups would average 373 and 489 percent the FEC1 levels of rr individuals respectively.

TABLE 2: Estimates of the size($\log_e$ eggs/g), mode of inheritance and frequency of host parasite resistance QTLs in the Ruakura Perendale FEC selection lines.[1]

	FEC1	FEC2
pedigree (animals)	1152	1093
observations (n)	943	907
infection level (nil alleles)	6.09	6.41
a	+0.69	+0.37
d	+0.44	+0.27
p	0.55	0.40
r	0.29	0.14

[1]abbreviations see Table 1.

The results from the Woodlands Coopworth progeny test analysis are presented in Table 3. FINDGENE converged for FEC at both challenges and for *Nematodirus* spp. egg counts at the first challenge, but not for host antibody levels at the end of the second challenge. Both the FEC2 and NEM1 trait exceeded the approximate threshold criterion, with the NEM1 (r = 0.70) markedly higher than any other trait. The estimates obtained suggest a slightly over-dominant allele exists for FEC2 susceptibility (S) which was present at the highest frequency of any of the three data sets examined. In contrast, results for *Nematodirus* spp. egg counts suggest a completely recessive allele for susceptibility (s) to infection was present. Back transformed onto the original scale FEC2, Sr and SS animals would average 584 and 493 percent the FEC levels of rr individuals respectively. Corresponding values for NEM1 sR and ss relative to RR animals are 69 and 493 percent.

DISCUSSION

These results provide encouraging evidence that marker linkage studies investigating inheritance of host resistance to internal parasites in sheep will detect QTL. They are of particular relevance to the AgResearch experiment currently in progress using large half-sib pedigrees of F1 rams from the Wallaceville Romney FEC selection lines (Crawford *et al.*, 1997). Similar studies with FINDGENE have already suggested that a QTL exists for resistance to ticks, an external parasite in cattle (Kerr *et al.*, 1994) and human studies using approximate maximum likelihood techniques have indicated the existence of QTL for resistance to infection from both malaria (Abel *et al.*, 1992) and *Schistosoma mansoni* (Abel *et al.*, 1991). This evidence suggests that QTL for disease resistance traits may be relatively common in vertebrates. Both the size and pattern of inheritance of the putative FEC QTL were remarkably similar in all three data sets examined and this provides additional support that FEC QTL exist. The magnitude of the QTL effects detected suggest that they could be of significant benefit to the New Zealand industry, if closely linked DNA markers can be isolated. The results also demonstrate the ability of mixed inheritance models to identify traits and individuals within populations of interest for subsequent intensive study using linkage techniques.

TABLE 3: Estimates of the size, mode of inheritance and frequency of host parasite resistance QTLs ($\log_e$ eggs/g and O.D. for ELFC2) in the Woodlands Coopworth progeny test flock.[1]

	FEC1	FEC2	NEM1	ELFC2
pedigree (animals)	5844	5844	5844	5844
observations (n)	4054	3983	4054	3611
infection level (nil alleles)	7.00	5.68	4.69	
a	0.00	+0.72	+0.57	N/C
d	-0.78	+0.89	-0.75	N/C
p	0.9	0.69	0.60	N/C
r	0.18	0.36	0.70	N/C

[1]abbreviations as per Table 1 except that FEC1, FEC2 and NEM1 were transformed using $\log_e[x+50]$ prior to analysis. NEM1 is the *Nematodirus* epg faeces at the end of the first challenge and ELFC2 the host antibody levels to *T. colubriformis* at the end of the second challenge. N/C means no estimates were available because the analysis did not converge.

However while exciting, the present results should be considered preliminary until confirmed by linkage analysis. This is because the FINDGENE methodology makes a number of assumptions in an effort to make the analytical problem more tractable. The assumptions include: a simple autosomal 2 allele major locus coupled with polygenic inheritance and a fixed known heritability for the polygenic inheritance. Clearly, the first assumption may in fact be false. A much more likely pattern would be a number of QTLs, with a variable number of alleles and frequencies, with individual alleles having differing effects on the quantitative trait of interest. Only a small number of loci could have alleles with effects of the magnitude described in the current results.

The second assumption is that the heritability of the polygenic variance is known. This is obviously false although an upper bound for the heritability can be estimated. However, analyses on simulated data sets show the technique is quite robust to incorrect estimates, provided the given estimate is low to medium (0.1 - .25). For FEC the real polygenic heritability would be 0.3 or less. Alternative techniques are available, but most suffer from several problems. The first is mixed model (i.e. polygenic and Mendelian) maximum likelihood techniques (Lalouel *et al.*, 1983). However, derivation of approximate maximum likelihood equations is often difficult and the procedure cannot be easily generalised to handle complex pedigree structures. A second more general technique is the recent utilisation of Markov Chain Monte Carlo methods to this problem (Janss *et al.*, 1995). This technique does not require a known prior polygenic heritability estimate and also estimates the standard error of the allele effects and population frequency. While this method offers considerable potential, the computational overhead is 100-1000 fold larger. This means that it will be used for more detailed analyses, but is unlikely to be used for routine screening in the near future.

Obviously, the current results are preliminary and much further work remains to be done. It is not known if