SELECTIVE GENOTYPING FOR DETERMINATION OF A MAJOR GENE ASSOCIATED WITH CRANIAL CRUCIATE LIGAMENT DISEASE IN THE NEWFOUNDLAND DOG

P.E. Macrossan[1], B.P. Kinghorn[1,2], V.L. Wilke[3] & M.F. Rothschild[3]
[1]Department of Animal Science, University of New England, Armidale
[2]Sygen Chair of Genetic Information Systems
[3]Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University

SUMMARY
Cranial cruciate ligament disease (CCLD) is the leading cause of lameness in the dog. The objective of this study was to determine the most informative subset of animals on which to carry out selective genotyping for the identification of any major gene affecting CCLD. Two groups of animals were chosen for genotyping from candidate groups of animals having high probabilities (calculated from segregation analysis) for carrying zero and two copies, respectively, of the putative causative allele. A stochastic search algorithm was employed to search the solution space for the 'best' two groups of individuals for genotyping based on two alternating fitness functions. The first fitness function was designed to minimize the genetic relationship amongst dogs within groups for each group individually. The second fitness function was designed to maximize the genetic relationship amongst dogs between groups. This approach, and the ramifications of its use, are discussed.
Keywords: Cranial cruciate ligament disease (CCLD), selective genotyping, segregation analysis.

INTRODUCTION
Cranial cruciate ligament disease (CCLD) is the leading cause of lameness in the dog (Johnson *et al.* 1994). Particular breeds of dogs are predisposed to CCLD, leading to the hypothesis that the disease has a heritable component. Experience has shown that significant morphological defects having all-or-none expression, such as CCLD, are often associated with single loci of major effect (termed Quantitative Trait Loci or QTL) (Falconer and Mackay 1996). Since whole genome scanning is an expensive procedure, the objective of this study was to determine the most informative subset of animals on which to carry out genotyping to identify the major gene responsible for CCLD.

Selective genotyping or DNA pooling has been used to choose individuals for genotyping on the basis of phenotype, or some other criteria of merit (Lander and Botstein 1989; Darvasi and Soller 1992). It can be used to reduce costs in QTL detection, by ensuring strong contrast in merit among individuals with genetic marker information. Selective genotyping can be seen as a special case of the approach to group genotyping formulated in the simulation study of Macrossan and Kinghorn (2005). In the work presented here, pre-selection on probabilities of carrying opposing disease genotypes was first carried out. Final selection was made in a manner that:

1. Reduced the relationships among animals within each selected group, to increase the effective number of dogs within each group, and

2. Increased the relationship between animals in different groups, to reduce the confounding effect of polygenes in the between-group contrast.

These genetic relationships were measured using CON, defined as the average of the numerator relationship matrix (NRM) of an animal to all live animals in the population, and grpCON, defined as the average of the numerator relationship of each animal in the selected group with all other animals in the group. Two groups of individuals were selected for genotyping from those available. The two groups were chosen from 'candidate pools' of animals having high probabilities for carrying zero and two copies, respectively, of the putative causative allele. A stochastic (randomized) search algorithm was employed to search the solution space for the 'best' two groups of individuals for genotyping based on two alternating fitness functions, both employing CON.

MATERIALS AND METHODS

Medical records, systemically evaluated for evidence of CCLD, were collected from the Iowa State University College of Veterinary Medicine for all Newfoundland dogs presented for care from 1995 to 2002. Additionally, a large-scale recruitment exercise was performed, with the resultant pedigree containing 282 animals of which 205 were available for DNA testing. Foundation animals were included to give a total of 747 animals in the pedigree. Segregation analysis on incidence of CCLD was carried out via the method of Kerr and Kinghorn (1996). A recessive mode of inheritance, a population frequency of 0.65 and a penetrance of 59% were inferred for the allele (a) of interest, using the technique described by Kinghorn (2003).

The project budget allowed for a total of 90 animals to be genotyped from the 205 animals available. Two 'candidate pools' of animals were assigned from which to choose a final 45 animals per group. Darvasi and Soller (1992) recommended genotyping animals located within the extreme 25% of the population distribution for the trait of interest. A stochastic search algorithm was employed to explore the solution space for the 'best' two groups of animals for genotyping, Grpaa and GrpAA, based on two alternating fitness functions. Grpaa and GrpAA were composed of animals from the extremes of the probability distributions for being homozygous recessive (aa) and homozygous dominant (AA), respectively. The first fitness function was designed to minimize the relationship amongst dogs within groups for each group individually (i.e. Fitness (1) = $\downarrow$ grpCON). The second fitness function was designed to maximize the relationship amongst dogs between groups in order to maximize the connectivity of the genotyped dogs to the rest of the pedigree, thereby potentially maximizing the information gained from selective genotyping groups (i.e. Fitness (2) = $\uparrow$ pairwiseCON). PairwiseCON is defined as the average across n times m NRM elements where n and m are the numbers of individuals in Grpaa and GrpAA respectively (thus, $n = m = 45$). A stochastic hill-climbing algorithm was used as the mechanism for the search , based on comparisons amongst a number of search algorithms used in a similar scenario (Macrossan, 2005).

RESULTS AND DISCUSSION

The search algorithm was run over 25 trials with randomly chosen initial 'titleholder' groups. In each of the 25 trials, individual grpCON for both Grpaa and GrpAA, and pairwiseCON for the groups, converged at approximately 240 iterations to the same values, viz. 0.0807, 0.0856 and 1.23×10^{-4}.

Since the pedigree contained animals both available and unavailable for genotyping, the final two groups chosen had different population percentages and probability cut-off points. The first group (CandidateGrpAA) contained 71 animals in the top 25% of the total population for the probability of

being homozygous dominant (p(AA) > 0.18). The second group (CandidateGrpaa) contained 54 animals in the top 31% of the total population for the probability of being homozygous recessive (p(aa) > 0.47). The individual evolution of grpCON for groups Grpaa and GrpAA, over 250 iterations of the search algorithm is shown in Figure 1 (top) whilst the concurrent evolution of pairwiseCON for the two groups over the same 250 iterations of the hill-climbing algorithm is shown in Figure 1 (bottom).

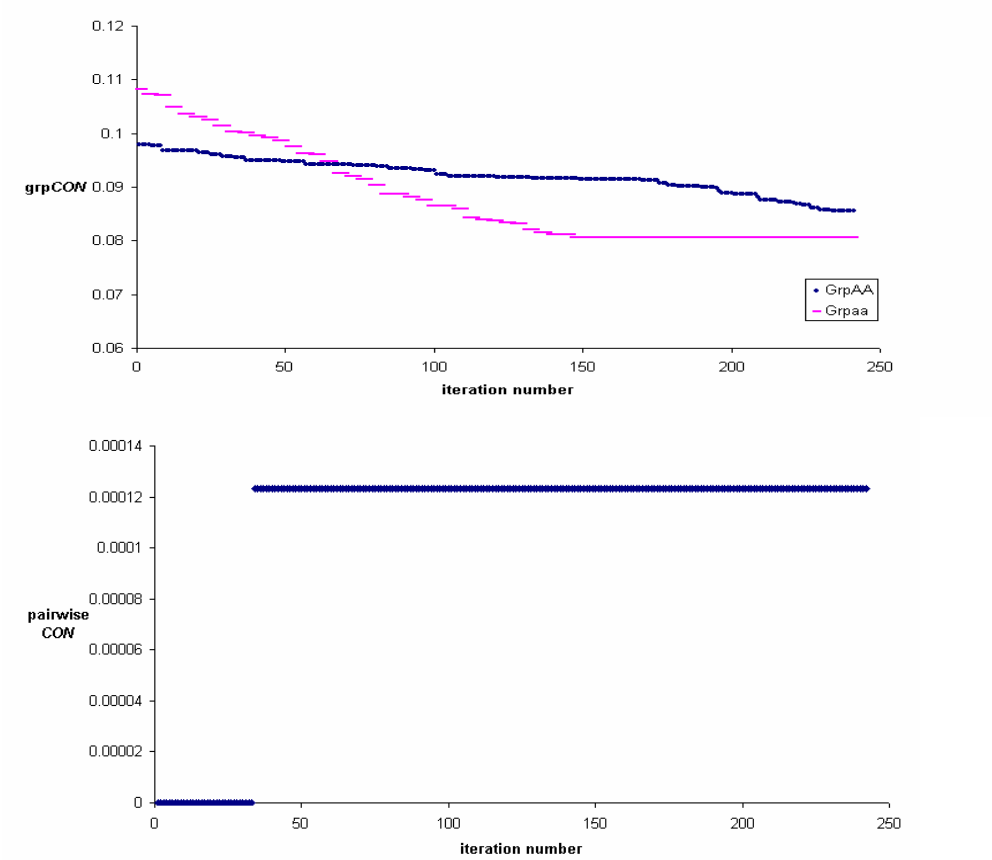

Figure 1 (top). The individual evolution of grpCON for Grpaa and GrpAA over 250 iterations of the search algorithm, and (bottom) the concurrent evolution of pairwiseCON for Grpaa and GrpAA over the same 250 iterations.

Inspection of Figure 1 (top) shows convergence for GrpAA being more gradual than that for Grpaa, which can be explained by the larger amount of choice available for the former group. The numbers of animals in the two candidate groups differed, with 71 candidates for GrpAA and 54 candidates for Grpaa. The 71 candidates for GrpAA represented the top 25% of the total population for the

probability of being homozygous recessive, in line with the recommendations made by Darvasi and Soller. The situation differed for candidate Grpaa for a number of reasons. Firstly, 44 animals available for DNA testing were CCLD positive, and were obvious choices for genotyping, having p(aa) equal or close to one. Secondly, a number of animals with a high p(aa) were unavailable for genotyping (i.e. deceased). Thirdly, a number of animals available for genotyping had the same high p(aa). To allow the algorithm sufficient flexibility in its choice of animals for both groups, flexibility was also necessary in the choice of animals for Grpaa. The combination of an individual's availability for genotyping, a high probability of carrying two copies of the allele, and the number of individuals having the same probability of the latter, led to 31% of the population, equal to 54 animals, being admitted into the candidate Grpaa.

It was earlier noted that CandidateGrpAA contained 71 animals in the top 25% of the total population for the probability of being homozygous dominant (p(AA) > 0.18). The cut-off point for this probability is significantly lower than that for the homozygous recessive candidate group (p(aa) > 0.47). This is a result of the phenotype for the latter being distinguishable from other genotypes in 59% of cases (equal to the penetrance of the phenotype), whilst the phenotype for both the homozygote dominant (AA) and the heterozygote (Aa) are undistinguishable from each other, leading to increased uncertainty in the assignment of probabilities for these two genotypes.

Inspection of Figure 1 (bottom) shows the lack of flexibility in the choice of animals in both groups with respect to pairwiseCON. The phenomenon of a single jump from zero to the final value of 1.23×10^{-4} occurred at some stage in each of the 25 trials and must be seen as a peculiarity of the particular pedigree involved, combined with the nature of the possibly conflicting fitness functions.

CONCLUSIONS

Selective genotyping was employed to choose animals for QTL investigations into CCLD, given pedigree restrictions and budgetary constraints. From two candidate pools of animals having a high probability of being either homozygous dominant or homozygous recessive, two sub-groups of animals were chosen for selective genotyping by means of a stochastic search algorithm with a two-step fitness function designed to minimize the relationship amongst dogs within groups whilst concurrently maximizing the relationship amongst dogs between groups. Since the total population contained animals both available and unavailable for genotyping, the two groups contained different percentages of the population and probability cut-off points. The nature of the possibly conflicting fitness functions led to a lack of flexibility in the choice of animals in both groups.

REFERENCES

Darvasi, A. and Soller, M. (1992) *Theoret. Appl. Genet*. **85**: 353.
Falconer,D.S and Mackay,T.F. (1996) "Introduction to Quantitative Genetics" 4th ed. Longman Press.
Johnson, J.A., Austin, C. and Breur, G.J. (1994) *Vet. Comp. Orthop. Traumotol*. **7**: 56.
Kinghorn, B. P. (2003), *Proc. Assoc. Advmt. Anim. Breed. Genet*. **15**: 103.
Kerr, R. J. and B. P. Kinghorn (1996). *J. Anim. Breed. Genet.* **113**: 457.
Lander, E. S. and Botstein, D. (1989) *Genetics* **121**: 185.
Macrossan, P. E. (2005) PhD Thesis, University of NewEngland.
Macrossan, P. E. and Kinghorn, B. P. (2005) *J. Anim. Breed. Genet.* Accepted.