EXTENT OF LINKAGE DISEQUILIBRIUM IN THE UK DAIRY CATTLE POPULATION

A. Tenesa[1], S. A. Knott[1], D. Ward[2], D. Smith[2], J. L. Williams[2] and P.M. Visscher[1]

[1]Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, Scotland, UK
[2]Roslin Institute, Roslin, Midlothian EH25 9PS, Scotland, UK

INTRODUCTION

Linkage disequilibrium (LD) mapping methods use linkage disequilibrium at the population level to map trait loci. LD mapping methods have higher power (Risch and Merikangas, 1996) and higher resolution than traditional linkage methods. The power of LD mapping methods depends on population parameters such as allele frequencies and level of LD. Although the extent and patterns of LD have been extensively studied in human populations, (Jeffreys *et al.*, 2001) farm animal populations have been rarely studied. Farnir *et al.* (2000) studied the extent of LD in the Dutch black-and-white dairy cattle population. In this paper, we assessed the extent of LD in two regions of the genome from a sample of 50 randomly selected dairy cattle bulls that were being progeny tested. They were assumed to produce a representative sample of the future extent of LD in the UK dairy cattle population.

MATERIAL AND METHODS

Data. The bulls were born between 1988-1995. Bulls were genotyped at six marker loci on chromosome two and at seven marker loci on chromosome six. Each bull pedigree was known up to three generations. Grand parents were assumed unrelated. Relationships between bulls are shown in Table 1. Genetic distances between markers were obtained from the map MARC97 (http://www.ri.bbsrc.ac.uk/cgi-bin/mapviewer?species=cattle).

Table 1. Additive genetic relationships (a) among bulls. Number of relationships (NR)

a	0	0.016	0.031	0.063	0.078	0.094	0.125	0.156	0.188	0.25	0.313	0.5
NR	837	15	70	135	2	6	105	8	7	31	3	6

Haplotype frequency estimation and Hardy-Weinberg equilibrium proportions. Maximum likelihood estimates of 78 two-marker loci haplotype frequencies were estimated by employing the expectation-maximization (EM) algorithm (Excoffier and Slatkin, 1995) as implemented in Gold (Abecasis and Cookson, 2000). Departures from Hardy-Weinberg equilibrium (HWE) proportions were tested using an exact test as described by Guo and Thompson (1992). This algorithm is implemented in Arlequin (http://lgb.unige.ch/arlequin/).

Level of linkage disequilibrium. Hedrick's normalised measure of disequilibrium (Hedrick, 1987) was obtained from the estimates of the two loci haplotype frequencies. It is defined as follows:

$$D' = \sum_{m=1}^{k} \sum_{n=1}^{l} m_m q_n |D'_{mn}|$$

where k and l are the number of alleles at locus M and Q respectively, m_m and q_n are the population allele frequencies of allele m at locus M and allele n at locus Q respectively. $|D'_{mn}|$ is the absolute value of Lewontin's normalised measure: $D'_{mn} = (D_{mn}/D^{max}_{mn}) = (h_{mn}-m_m q_n)/D^{max}_{mn}$ (Lewontin, 1964) where h_{mn} is the estimated population frequency of the haplotype M_mQ_n and D^{max}_{mn} is the maximum amount of disequilibrium possible between allele m at locus M and allele n at locus Q that equals: $min\{m_m q_n, (1-m_m)(1-q_n)\}; D_{mn} < 0\}$ or $min\{m_m(1-q_n), (1-m_m)q_n\}; D_{mn} > 0\}$.

In order to test the statistical significance of the allelic association, we compared the statistic $S = 2ln(L_{LD}/L_{LE})$ to a χ^2 distribution with $(k-1)*(l-1)$ degrees of freedom (Slatkin and Excoffier, 1996). Assuming random mating, L_{LD} is the likelihood computed using the haplotype frequencies found by the EM algorithm and L_{LE} is the likelihood under the assumption of linkage equilibrium. We performed a large number of tests (n = 78), therefore we applied a Bonferroni correction. The individual significance threshold after correction was 0.0007, corresponding to an overall significance level of 5%. Since a number of the 78 tests are likely to be correlated, our stringent threshold is likely to be conservative with respect to the type-I error rate.

RESULTS

Departures from HWE. Genetic positions of the markers, number of alleles at each locus and loci that showed significant departures from HWE are shown in Table 2.

Table 2. Genetic map and number of alleles at the marker locus (NA)[A,B]

Chromosome 2						
Marker	**TGLA226[A,B]**	**BMS829[B]**	**BMS2519[A,B]**	**BM2113**	**IDVGA37**	**IDVGA2[B]**
Genetic map	80 cM	91.5 cM	101.5 cM	106.2 cM	108.2 cM	117.8 cM
NA	5	5	5	6	3	5

Chromosome 6							
Marker	**RM28[B]**	**BM415[A,B]**	**CSN3[B]**	**BM1236**	**BMS511**	**AFR227[A,B]**	**BM8124[B]**
Genetic map	74.3 cM	76.3 cM	82.6 cM	83.9 cM	89.8 cM	90.4 cM	94.2 cM
NA	4	7	3	4	5	6	2

[A] Locus showed significant departure of HWE (P<0.001)
[B] Locus observed heterozygosity was smaller than the expected heterozygosity under HWE

Linkage disequilibrium between syntenic marker loci. Figure 1A plots the extent of disequilibrium (D') versus genetic distance measured in cM. The average D' was 44%. The most remarkable observation was that D' did not appear to vary as a function of the genetic distance. The level of association ($-\log_{10}(P)$) did show a clearer correlation with distance (Figure 1B) but still highly variable, especially for the smallest distances. All $P < 0.01$ correspond to genetic distances smaller than 10.3 cM. Only two pairs of markers were in significant linkage disequilibrium after accounting for multiple testing. These were BM1236-

BM8124 (P = 0.0007; distance = 10.3 cM) and BMS511-AFR227 (P = 0.0007; distance = 0.6 cM) on chromosome six. Although we observed a high average level of disequilibrium, only two pairs of loci showed a significant association. In order to test whether the mean level of disequilibrium observed was significant we calculated: the sum of the 36 X^2 statistics (X^2 = 646) and the sum of the 36 associated degrees of freedom (456 d.f.). This overall test for average level of LD across all pairs of syntenic loci was highly significant ($P << 10^{-7}$) indicating that the mean level of disequilibrium was significantly different from zero and that we lacked power when testing individual pairs.

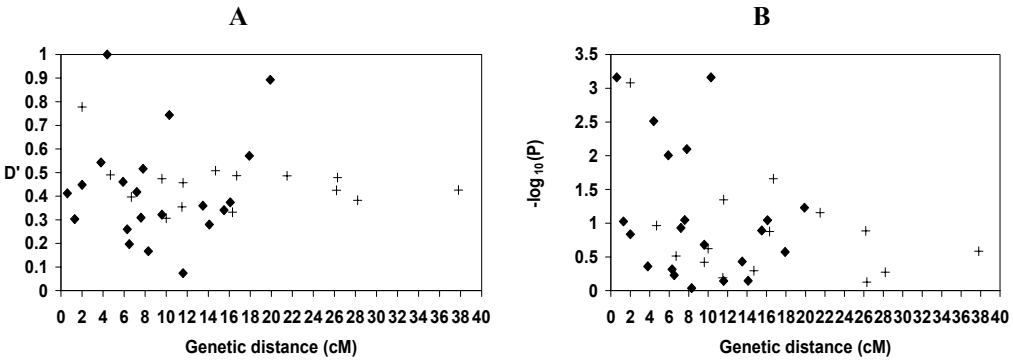

Figure 1. On the left, relationship between genetic distance (cM) and level of linkage disequilibrium (D'). On the right, relationship between level of significance ($-\log_{10}$ (P)) and genetic distance (cM) for syntenic loci pairs. Crosses and diamonds represent comparisons between pairs of loci on chromosome two and chromosome six respectively

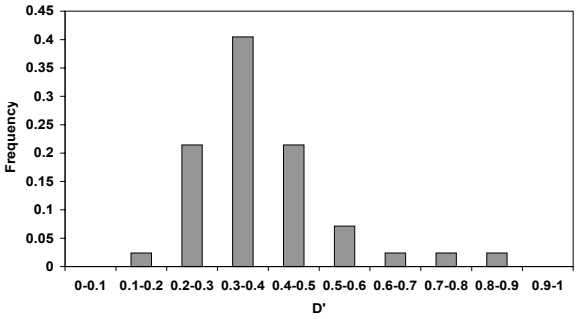

Figure 2. Distribution of D' values observed between pairs of non-syntenic loci

Linkage disequilibrium between non-syntenic marker loci. Figure 2 shows the distribution of D' values observed between pairs of non-syntenic loci. We estimated the mean level of LD between non-syntenic loci, measured as D', to be 39%. None of the loci pairs showed significant association between alleles. The sum of the 42 X^2 statistics obtained between non-

syntenic loci was 548 and the sum of the 42 associated degrees of freedom was 539. The overall level of association between pairs of non-syntenic loci was not significant (P = 0.39). Overall, average levels of LD were fairly similar between syntenic and non-syntenic loci. However, association could be statistically detected between syntenic loci but not between non-syntenic loci, even when the D' values were similar.

DISCUSSION

Our results show that LD mapping methods could be successfully applied to the future UK dairy cattle population with the available density of microsatellite markers. Significant linkage disequilibrium was found only for genetic distances smaller than about 10 cM, in addition significant association was never found between non-syntenic loci. This would have important implications for LD mapping. Firstly, the mapping resolution achievable with this level of disequilibrium would not be much finer than with traditional linkage methods. Secondly, if the lack of significant association found here between loci on chromosomes two and six were the same across the whole genome, then the number of false positives due to allelic associations between unlinked loci would be small when applying LD methods to map trait loci.

Given the observed level of LD, mapping methods based upon population-wide association might not provide much better resolution than traditional linkage methods, but could reduce the required sample sizes of the experiments.

We noted that the region of chromosome six where we detected the most significant LD has been reported to harbour QTLs influencing milk, fat and protein yield in the UK dairy population (Wiener *et al.*, 2000). This suggests that selection for milk production traits could have generated LD in this region, which was detectable even with the large amount of background LD observed.

REFERENCES

Abecasis, G.R. and Cookson, W.O.C. (2000) *Bioinformatics* **16:** 182-183.
Excoffier, L. and Slatkin, M. (1995) *Mol. Biol. Evol.* **12:** 921-927.
Farnir, F., Coppieters, W., Arranz, J.J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marcq, F., Moreau, L., Mni, M., Nezer, C., Simon, P., Vanmanshoven, P., Wagenaar, D. and Georges, M. (2000) *Genome Res.* **10:** 220-227.
Genstat 5 Committee (1993) Genstat 5 reference manual, Clarendon Press, Oxford.
Guo, S.W. and Thompson, E.A. (1992) *Biometrics* **48:** 361-372.
Hedrick, P. W. (1987) *Genetics* **117:** 331-341.
Jeffreys, A.J., Kauppi, L. and Neumann, R. (2001) *Nat. Genet.* **29:** 217-222.
Lewontin, R.C.(1964) *Genetics* **49:** 49-67.
Risch, N. and Merikangas, K. (1996) *Science* **273:** 1516-1517.
Slatkin, M. and Excoffier, L. (1996) *Heredity* **76:** 377-383.
Wiener, P., Maclean, I., Williams, J.L. and Woolliams, J.A. (2000) . *Anim. Genet.* **31:** 385-395.