Restriction Fragment Length Polymorphisms in Dairy Cattle Genetic Improvement

Y. Kashi[1], M. Soller[1], E. Hallerman & J.S. Beckmann[2]

[1]Department of Genetics, The Hebrew University of Jerusalem, 91904 Jerusalem Israel

[2]Department of Plant Genetics and Breeding, Agricultural Research Organization, The Volcani Center, 50250 Bet Dagan, Israel.

SUMMARY

Restriction fragment length polymorphisms are a new class of genetic marker uncovered by means of recombinant DNA technologies, that promise to be exceedingly numerous in agricultural livestock species. These markers open broad prospects for the mapping and breeding manipulation of quantitative trait loci (QTL). Mapping a large number of probes at a species level, could be most readily accomplished by utilizing wide crosses between divergent races or breeds Simulation studies show that it would take about 125 random polymorphic markers to provide 80% coverage of the cattle genome, at a maximum spacing such that all points are within 20 cM of a marker. Marker-QTL linkage studies, based on data from sire daughter progeny groups, could map 50% of the QTL segregating in a population, once five sires have been evaluated, and virtually all segregating QTL once 20 sires have been evaluated. This would involve about 1000 daughters each evaluated for five traits and 100 markers per sire. Selection programs based on marker-evaluated sires could contribute an additional 25-50% overall genetic progress. Costs would be about $1000,000 per marker-evaluated sire, and discounted returns about 1.5-3.0 x costs.

INTRODUCTION

Restriction fragment length polymorphisms (RFLPs) are a new class of genetic marker uncovered by means of recombinant DNA methodologies. RFLPs are detected by hybridization patterns of cloned and labelled DNA sequences ("probes") with fragments of DNA formed on digestion of genomic DNA by restriction endonucleases and separated according to length by gel electrophoresis. In all cases where a species has been studied intensively, RFLPs have been found to be exceedingly numerous, and there is every reason to expect them to be equally numerous in populations of agricultural livestock species as well. Indeed, initial studies have demonstrated the presence of RFLPs in the Israeli-Friesian dairy cattle population (Beckmann et al., 1986). Possible applications of RFLPs in animal genetic improvement have been described by Soller and Beckmann (1982, 1983, 1985), Beckmann and Soller (1983) and Smith and Simpson (1985). Of particular interest are the prospects for the mapping and manipulation of quantitative trait loci by means of linked RFLPs. Recently, Jeffreys et al. (1975) uncovered a class of probe that detects multiple highly variable regions in genomic DNA in man. Should such regions prove to be generally present in agricultural species, the costs of RFLP analyses would be greatly reduced, bringing these prospects closer to realization.

The general procedure for detecting linkage between a polymorphic genetic marker and a QTL in a segregating population has been described by Soller and Genizi (1978). In the present study we consider the number of RFLPs required for comprehensive mapping of QTL in dairy cattle, the power of marker-QTL linkage analyses, and the potential contribution and costs of marker-assisted selection of young sires in a dairy cattle improvement program.

RESULTS

The number of RFLPs required for comprehensive mapping of QTL in dairy cattle

The power of an experiment aimed at detecting linkage between marker and QTL decreases in proportion to $(1-2r)^2$, where r is the proportion of recombination between marker and QTL (Soller et al., 1976). Consequently, in most cases marker-QTL linkage will not be detected when $r>0.2$. Thus, for purposes of mapping QTL within a particular population, it is necessary to locate a set of RFLPs, polymorphic in that population, that are spaced in such a manner that all points of the genome to be mapped are located within no more than 20cM of a polymorphic marker.

An efficient way to achieve this would be to first map a very large battery of probes at a species level. These would then be tested, according to chromosomal location, for their ability to detect polymorphisms within any particular population of interest, until an appropriately spaced subset, giving adequate chromosome coverage in the population studied, was identified. If none of the probes mapped to a particular region at the species level proved to be polymorphic in the particular population studied, the mapped probes could still be used as a base to develope additional probes in that region, through chromosomal walking (Rohme et al., 1984) or hopping. (Collins and Weissman, 1984). These could be tested without further mapping, for their ability to detect polymorphism in the region in question.

A large number of probes could most readily be mapped to the chromosomal level, by means of a suitable panel of somatic cell hybrid lines. This approach has been successfully used to map a number of loci in cattle by Womack (1984). Mapping a large number of probes to chromosomal regions might most readily be achieved by utilizing the F-2 or backcross of a wide cross between different races of breeds. In such a cross most probes should be polymorphic. If large DNA samples were taken from a single set of offspring out of such a cross, these samples could be exchanged between laboratories and used virtually indefinitely as international standards to map additional probes as they became available. This would be analogous to the use of recombinant inbred lines for mapping purposes in mice (Taylor, 1978). Newly available probe would first be tested for polymorphism by comparing the hybridizing fragments in the parental breeds of the wide cross. As each new polymorphic probes was identified its banding pattern across the F-2 individuals or backcross would be compared to that of other probes, and it would be automatically mapped with respect to all previously mapped probes. Two or three such wide crosses should allow virtually any probe to be mapped within the species.

Wide crosses could be carried out specifically in order to provide an F-2 or backcross population for mapping purposes. Alternatively, since crossing of breeds and races is often carried out in the cattle world for various breeding purposes, information as to the existence of any such populations could be brought to the attention of appropriate laboratories, and DNA samples obtained from 100-200 individuals as required for mapping purposes. In this way within a very few years, a few hundred mapped probes might be available for dairy cattle as a species.

If a large battery of mapped probes is not available, genome coverage for a particular population will require that a large number of probes be tested for polymorphisms and sufficient RFLPs within that population identified to provide adequate genome coverage. The proportion of the genome covered to the required density for QTL mapping, as a function of the total number of RFLPs uncovered in the population can be calculated on the assumption that RFLPs will be distributed in a random or quasi-random manner among the chromosomal genome.

It unexpectedly turned outthat the degree of chromosome coverage for given total number of markers per genome was little affected by chromosome length, in the range 50-150 cM. Table 1 shows expected degree of chromosome coverage as a function of number of RFLPs per 1000 cM. Taking a total genome size of 2500 cM would seem to be a reasonable estimate for cattle. On this basis, 80% coverage of the genome by randomly distributed RFLPs would require about 125 RFLPs all told.

Table 1. Expected degree of chromosome coverage, at a maximum distance of 20cM between any point on the genome and a polymorphic marker, according to number of polymorphic markers per 1000 cM (assumed chromosome size equal 100 cM).

Degree of chromosome coverage	Number of polymorphic markers per 1000 cM
0.40	14.3
0.50	19.0
0.60	25.7
0.70	33.3
0.80	45.2
0.90	65.7
0.95	86.2

Experimental size for determining marker-QTL linkage relationships.

An experiment aimed at identifying marker-QTL relationships within a segregating population will involve the above number of markers, each tested against a series of five to ten or more quantitative traits, i.e., a matrix of 500 to 1000 comparisons. Thus, even a small Type 1 error will result in a large number of significant false effects, e.g., if 500 comparisons are carried out, a Type I error of 0.05, will provide 25 falsely identified marker-QTL linkages. On the other hand, if Type I error is set very small, say, 0.01 or 0.001, then Type II errors will become large and a large proportion of any useful marker-QTL linkages present in the population may be missed in the analysis.

Table 2 shows the power of an RFLP-QTL linkage analysis by number, k, of sires assuming number of daughters per size marker-genotype class, N=500, proportion of recombination between RFLP locus and QTL, r=0.10, and gene effect, d=0.15σ. The results show that for this many daughters, once as few as five sires have been evaluated, close to 60% of the QTL segregating in the population will be identified, with a Type I error of 0.05; and close to 45% of the segregating QTL will be identified, even if Type I error is set at 0.01. When 20 sires have been evaluated, virtually all of the segregating QTL will have been correctly identified, while very few false identifications will remain.

The total number of daughters and of sires that have to be screened to provide N=500 daughters per marker-allele progeny group, will depend greatly on whether probes detecting multiple variable regions, of the sort described by Jeffreys et al. (1985) prove to be generally available in dairy cattle. If sufficient probes of this nature are obtained to provide total genome coverage, then due to the rarity of individual alleles, each sire can be expected to heterozygous at the great majority of marker loci and it will also be possible

Table 2. Power of marker-QTL linkage analysis as a function of number of sires evaluated, and Type I error, assuming number of daughters per sire marker-allele progeny group: N=500, gene effect: d=0.15 , and proportion of recombination: r=0.10.

Number of sires evaluated	Type I error	
	0.05	0.01
5	0.59	0.44
10	0.70	0.54
20	0.91	0.85

to assign virtually all daughters unequivocally to one or the other sire marker allele classes. In this case, the total number of sires and daughters tested for an effective k and N will tend to 2kN. That is, for k=5 and N=500, a total of 5,000 daughters would have to evaluated for RFLP markers and quantitative traits of interest. The same would hold true if probes detecting individual hypervariable regions are available in sufficient numbers to cover the genome.

If hypervariable polymorphisms are not available, total numer of daughters tested for effective K=5 and N=500 might range up to 20,000. This leads to the conclusion, that for effective genome coverage and comprehensive mapping of QTL in the genome, a sina qua non will be the availability of a battery of probes detecting highly polymorphic regions.

Marker-assisted selection of young sires

In marker-assisted selection of young sires, alternative homologous chromosomal regions of a proven sire, differentially marked by alternative alleles at a marker locus, are evaluated by comparing the mean quantitative trait value of daughter progeny groups that received the alternative sire-marker alleles. Young sires would then be chosen on the basis of the summed breeding value of the alternative markers that they received from their marker-evaluated sire. It is anticipated that such a program would be instituted only after preliminary mapping of QTL in the population so that information was available as to those markers that had QTL for specific traits associated with them. As shown above, there will be a proportion of falsely identified effects, in the sense that some of the marker-QTL linkage identifications will be the result of Type I error. The proportion of falsely identified marker-QTL linkages will effect the additional breeding value (B.V.) of the young sires chosen in this manner; assuming that breeding value of the dams will be unaffected by selection of young sires, since it will be possible to raise large numbers of offspring from highly selected dams by means of induced multiple ovulation and embryo transfer.

Since the superior marked chromosomal region obtained from the parental proven sire will be transmitted to only half of the daughters of the young sires, additional value of the daughters of these young sires will be half of the additional breeding value of the young sires themselves.

Table 3 shows estimated superiority of the daughters of the marker-selected young sires, as a function number of sires tested and number of segregating

-QTL accessible to the markers and assuming d=0.15, r=0.10, N=500, and 100 markers tested against five quantitative traits (e.g., milk yield, fat%, udder conformation, legs and mastitis susceptibility). Daughter superiority is given in milk yield equivalents, assuming the phenotypic standard deviation for all quantitative traits monitored equal in value to 1000 additional liters of milk (or $100) per year. Expected daughter superiority increases with increased number of segregating QTL in the portion of the genome covered by polymorphic markers, but is almost unaffected by the magnitude of Type I error: apparently the increase in false marker-QTL identifications with greater Type I error is balanced by the increase in true marker-QTL identifications in its effect on daughter superiority. With 30 segregating QTL in the marker-covered portion of the genome, daughter superiority ranges from the equivalent of 80 liters per year at the beginning of the program, when marker-QTL linkage information is based on data from five evaluated sires only, to 115 kg per year, as information accumulates and marker-QTL linkage information is based on data from 20 evaluated sires. With 50 segregating QTL in the marker-covered portion of the genome the corresponding values are 110 and 150 kg per year respectively. As information accumulates on the location of QTL of interest, it will be possible to search for additional polymorphic markers in the near vicinity of the QTL, or to include the QTL in marker brackets. In this case the proportion of recombination between marker and QTL will decrease, and expected gains will be somewhat higher. In the best case, where r is reduced to zero, and information is available on 20 sires, gains in the daughters of the marker-selected sires will equal 160 to 210 liters per year, depending on the number of QTL in the marker-covered portion of the genome. Assuming an average generation interval of six years, the increase in annual genetic gains per year would be equivalent to 15-35 liters per year. This would represent a 25-50% increase in current rates of genetic progress. Useful, but not world-shaking! Major increases in rates of genetic progress would come only as information on marker-QTL linkages is used to select directly for the QTL (Soller and Beckmann, 1982; 1983; Smith and Simpson, 1985).

Costs and benefits

Given the availability of probes detecting multiple highly variable regions, screening 1000 daughters for 50-100 markers would involve two technician years of work plus about $40,000 of materials, i.e., about $100,000 to obtain a comprehensive marker-evaluation of a single proven sire. Assuming that the marker-selected sons of this proven sire, returned to service and retained after progeny testing, sire 10,000 daughters all told, the expexted total return on this investment (after discounting) would range from $120,000 at the beginning of the program, to $200,000 when information has accumulated on 20 sires, to a potential maximum of $300,000 when r is reduced to zero. Thus, the potential return would suffice to cover costs with a bit left over. The program would therefore be worthwhile economically, at least initially, only if the information were obtained for an outstanding proven sire, whose sons could be expected to produce in excess of 10,000 daughters. For such a sire, the added market value of his marker-selected sons might also prove a significant source of income. Marker-QTL linkage information might also enable direct or quasi-direct selection for QTL via haplotyping or other approaches. The value of such selection has been considered by Smith (1967), Soller (1978), Soller and Beckmann (1982, 1983) and Smith and Simpson (1985), but still awaits definitive modelling and cost/benefit analysis. It is possible that marker-QTL information and marker-assisted selection may find useful application in the context of a nucleus breeding herd scheme based on multiple ovulation and embryo transfer as described by Nicholas and Smith (1983).

Table 3. Superiority of daughters of marker-selected young sires in milk equivalents (kg/year) as a function of number of sires evaluated, Type I error, and number of QTL segregating in the portion of the genome covered by polymorphic markers; assuming 100 polymorphic markers and five quantitative traits tested for marker-QTL linkage other assumptions as a in Table 2.

Number of sires	Number of QTL in marker-covered genome			
	30		50	
	Type I error		Type I error	
	0.05	0.01	0.05	0.01
5	80	80	110	105
10	85	90	120	120
20	110	115	140	150
Maximum		160		210

ACKNOWLEDGEMENT

This research was supported by a grant from the U.S. -Israel Binational Agricultural Research and Development Fund (BARD).

REFERENCES

Beckmann, J.S. and M. Soller. (1983). Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. Theor. Appl. Genet. 67: 35-43.

Beckmann, J.S., Y. Kashi, E. Hallerman, A. Naveh and M. Soller. (1986). Restriction fragment length polymorphisms in Israeli-Friesian dairy cattle. Animal Genetics (in press).

Collins, R.S. and S.M. Weissman. (1984). Directional cloning of DNA fragments at a large distance from an initial probe: A circularization method. Proc. Nat. Acad. Sci. (U.S.) 81: 6812-6816.

Nicholas, F.W. and C. Smith. (1983). Animal Production 36: 341-353.

Rohme, D., H. Fox, B. Herrmann, A. Frischauf, J. Edstrom, P. Mains, L.M. Silver, H. Lehrach. (1984). Molecular clones of the mouse t complex derived from microdissected metaphase chromosomes. Cell 36: 783-788.

Smith, C. (1967). Improvement of metric traits through specific genetic loci. Animal Production 9: 349-358.

Smith, C. and S.P. Simpson. (1985). The use of genetic polymorphisms in livestock improvement. European Association of Animal Production, Greece, 1985.

Soller, M. (1978). The use of loci associated with quantitative traits in dairy cattle improvement. Animal Production 27: 133-139.

Soller, M. and J.S. Beckmann. (1982). Restriction fragment length polymorphisms and genetic improvement. Second World Congress on Genetics Applied to Livestock Production, Madrid, 6: 396-404.

Soller, M. and J.S. Beckmann. (1983). Genetic polymorphisms in varietal identification and genetic improvement. Theor. Appl. Genet. 67: 25-33.

Soller, M. and J.S. Beckmann. (1985). Restriction fragment length polymorphisms and animal genetic improvement. Reviews in Rural Sciences 6: 25-33.
Soller, M., A. Genizi and T. Brody. (1976). On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor. Appl. Genet. 47: 35-39.
Soller, M. and A. Genizi. (1978). The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. Biometrics 34: 47-55.
Taylor, B.A. (1978). Recombinant inbred strains: use in gene mapping. In: Origins of Inbred Mice, ed. H.C. Morse III, Academic Press, N.Y. pp.: 423-438.
Womack, J.E. (1984). Gene map of the cow. Genetic Maps 3: 402.