PROGRESS TOWARDS A PRIMARY DNA MARKER MAP IN CATTLE.

M. GEORGES, A. MISHRA, L. SARGEANT, M. STEELE, X. ZHAO.
GENMARK Inc., Salt Lake City, Utah, US.

SUMMARY

More than 130 highly polymorphic bovine DNA markers have been identified which can be classified into three groups: 1. VNTR-type markers (Variable Number of Tandem Repeats); 2. Multisite haplotypes; 3. (TG)n repeats or microsatellites. The latter markers are detected using the Polymerase Chain Reaction, offering new perspectives for automated genetic typing of large samples. These markers are mapped by linkage analysis in bovine pedigrees generated by Multiple Ovulation and Embryo Transfer. One of these pedigrees segregates for Weaver, a disease shown to be linked to a Quantitative Trait Locus affecting milk yield.

INTRODUCTION

One of the major opportunities offered by molecular biology to animal breeding, is the possibility to unravel the genetic component of quantitative traits at the molecular level. This should not only lead to a better understanding of the physiology underlying those traits, but should as well allow us to increase genetic response by affecting both time and accuracy of selection (Soller and Beckmann, 1982; Smith and Simpson, 1986).
One strategy to identify the genes behind complex traits is through so-called "reverse genetics" (Orkin, 1986). In a first stage, this implies the determination of the map location of the corresponding genes, which can be obtained by linkage studies. Such studies, however, require a tool: a genetic map covering as much as possible of the genome of interest, if not the whole.
The purpose of this paper is to summarize our progress in the development of such a DNA marker map in cattle.

A. CHARACTERIZATION OF HIGHLY POLYMORPHIC BOVINE DNA MARKERS:

Assuming a total bovine map length of ± 30 Morgans, by analogy with the human, and a desired maximum distance of 20cM between adjacent markers, approximayely 150 of them would be required to obtain such a map. However, at least twice as many markers will be needed to generate a reasonable map, and this essentially for two reasons: 1. most of the time we have no a priori information on the location of the characterized markers. Hence, some chromosomal regions will be overrepresented in our map, others underrepresented. This problem is expected to become critical in the later stages of the development of the map. At that point, however, it should be

possible to exploit comparative data (Womack and Moll, 1986) to look for markers whose location is known in other species in order to fill the remaining gaps in our bovine map. 2. An individual will only be informative for the markers for which he is heterozygous; parts of his genome won't thus be explorable, because he will be homozygous for the corresponding markers. To compensate for this, one will have to identify more markers, the number required being inversely proportional to their heterozygosity. Hence the importance of highly informative systems.
Within one year of work, we have isolated over 130 highly polymorphic DNA markers, having a mean heterozygosity ≥ 50%, and which should already cover a significant part of the bovine genome. These markers can be classified into 3 categories:

α) VNTR-type markers (Variable Number of Tandem Repeats):
Hypervariable minisatellites have proven an invaluable source of genetic markers in the human, characterized by an exceptionally high heterozygosity (Nakamura et al.,1987). Exploiting the cross-hybridization observed between families of hypervariable minisatellites from different species, we have screened purpose built genomic libraries with human, rodent and the bacteriophage M13 minisatellite, in order to isolate cattle-specific VNTRs. The libraries used were obtained by ligating size selected restriction fragments (≥2Kb) generated by either of two four-cutters: HaeIII or MboI, into plasmid vectors. The rationale behind this approach is three-fold: 1. the size fractions used in the ligation are rich in minisatellite sequences, reducing the number of recombinants to screen and allowing us to use plasmid vectors; 2. as suggested by Jeffreys et al. (1985), large minisatellites may be more mutation - prone and, hence, more variable; 3. only one restriction enzyme is used in the Southern blot hybridization experiments to detect the genetic polymorphism - the same as the one used to construct the libraries -, reducing the screening costs.
Approximately 50 polymorphic VNTRs have been isolated using this strategy. The mean heterozygosity for all these markers was estimated at 51% in the Holstein breed.
Besides being a valuable source of DNA markers for linkage studies, the best of these markers could be used for identification and paternity diagnosis. Based on allelic distributions in the Holstein breed, the following parameters were estimated by Monte-Carlo simulation: 1. Identification: Matching Probability for 2 Randomly selected individuals (MPR), Matching Probability for 2 full-Sibs (MPS); 2. Paternity diagnosis: Exclusion Power when putative father is unrelated to real father (EPR), Exclusion Power when putative father and real father are full-Sibs (EPS), Exclusion Power when only one parent is available (EPSP). The estimated values are: MPR: 3.3E-8 and 4.0E-13; MPS: 0.00372 and 0.000035; EPR: 0.998524 and 0.999966; EPS: 0.884005 and 0.975671; EPSP: 0985092 and 0.998330, when combining our 5 or our 10 best probes respectively.
70% of our bovine probes revealed locus-specific patterns in sheep as well, with 85% of them being polymorphic. One of the bovine

clones was mapped by in situ hybridization to the proterminal region of chromosome 21 in cattle and of the homologous chromosome 18 in sheep (R. Fries, personal communication). Hence, minisatellite sequences exhibit conservation of both DNA sequence and map location within the Bovidae family. In consequence, a significant part of the data generated in cattle will be readily usable in other Bovidae, including sheep.

β) Multisite haplotypes:

Analysis of the map location of human VNTRs (Royle et al., 1988) and results from linkage studies performed with bovine multilocus DNA fingerprints (Georges et al., 1990) points towards non-random distribution of minisatellite sequences, which seems to be organised in clusters especially at proterminal regions. As a consequence, constructing complete genetic maps relying exclusively on VNTR-type markers is unlikely and we had to look for alternative sources of highly polymorphic markers.

Our first approach was to use random cosmids as probes in Southern blot hybridization to screen for RFLPs with 15 restriction enzymes, in order to generate multisite haplotypes (White and Lalouel, 1988). Approximately 75 such cosmids have now been screened. Surprisingly, 92.75% of the cosmids revealed one or more polymorphic events. The mean heterozygosity for these multisite haplotypes was estimated at ≥ 48% Using the RFLPs due to point mutations, we estimated the nucleotide diversity, π, or average heterozygosity per nucleotide site at 0.0007. This relatively low value compared to the human (human $\pi \pm 0.002$), confirms our previous results in the Belgian Blue Cattle breed (Georges et al., 1987; Hilbert et al.,1989). It is, however, compensated by a surprisingly large number of insertion/deletions. Indeed, out of 193 RFLPs identified, 55 (28.5%) were attributed to insertion/deletions because they were detected by more than one enzyme. As a matter of fact 64% of the cosmids detected such an event. It is tempting to speculate that at least part of these insertion/deletions may be attributed to mobile elements in the bovine genome.

γ) Microsatellites:

A potential source of highly informative, well dispersed DNA markers are the (TG)n repeats or microsatellites as identified in the genome of man and other organisms (Weber et al., 1989; Litt et al., 1989; Tautz et al., 1989). Screening a bovine cosmid library with a (TG)9 oligonucleotide demonstrated that at least 75% of the clones contained one or more such repeats. A library of MboI fragments between 250 and 500 bp was screened with the same oligonucleotide probe. One out of 50 clones cross-hybridized, allowing us to estimate the number of (TG)n repeats in the bovine genome at ≥ 150,000.

Approximately 75 of these have now been isolated and sequenced. At present, we have been able to successfully amplify 15 of these in vitro, using the Polymerase Chain Reaction. When analysing the product on denaturing polyacrylamide gels, all of them exhibited

substantial genetic polymorphism. The mean heterozygosity over all systems has been estimated at 65% Hence, it makes no doubt that microsatellites will be an invaluable source of highly polymorphic DNA markers for cattle too.
The possibility to use the Polymerase Chain Reaction to detect polymorphism opens entirely new perspectives for the efficient genetic typing of large samples. Several pairs of our (TG)n repeats can be amplified in duplex and results reported by others demonstrate that the multiplex amplification of a larger number of systems is achievable. Moreover, the possibility to tag fluorescent dyes to the primers used in the amplification, allows the automated detection of several systems simultaneously.
Moreover, the widespread distribution of (TG)n repeats offers new possibilities when using the "candidate gene" approach or exploiting comparative data. In both instances, one looks for polymorphism at specific loci, whether studying genes considered as candidates for a major role in the determinism of the trait of interest, or genes whose map location is known in other species. The corresponding loci have a high chance to harbor (TG)n or other dinucleotide repeats, susceptible to be a rich source of genetic polymorphism indispensable for further study and whose isolation is fairly straitforward.

B. EXPLOITING "DISEASE-TAGGED QTLs" TO SIMULTANEOUSLY CONSTRUCT LINKAGE MAPS AND MAP QTLs:

The map location of our DNA markers is determined by a combination of linkage studies and other mapping approaches, such as <u>in situ</u> hybridisation and the use of somatic cell hybrid panels, in collaboration with other laboratories (R. Fries - ETH-Zentrum, Zurich; J. Womack - Texas A.& M.).
To perform the linkage studies, we are using bovine pedigrees generated by MOET (Multiple Ovulation and Embryo Transfer) in order to maximize the number of informative meiosises with a minimum number of individuals.
One of our pedigrees segregates for the recessive Weaver condition. Besides helping us to build the bovine map, this pedigree may provide us with a marker for the Weaver condition, severely threatening the Brown Swiss breed. Potentially more important, however, is the demonstration by Hoeschele and Meinert (1989) of the presence of a QTL with a major effect on milk yield in the same chromosomal region. By going through the relatively easy exercice of mapping a single gene disorder, we hope to be able to identify a major QTL for milk yield. The existance of other "disease-tagged QTLs" is presently explored.

CONCLUSIONS:

We demonstrated in this work, that a large number of DNA markers can be identified in a reasonable time-span and that, hence,

complete genetic maps should soon become available for domestic species, especially cattle. We demonstrate as well that recent developments in marker technology, exploiting PCR technology in particular, will allow to type large samples fast and at lower cost. We propose to use "disease-tagged QTLs" as a short cut to map genes affecting complex traits.

ACKNOWLEDGEMENTS

We are grateful to Tom Holm, Professors R. White, J.M. Lalouel, R. Gesteland, Dr. J. Massey for their continuous support and critical interest.

REFERENCES

GEORGES, M.; LEQUARRE, A.S.; HANSET, R.; VASSART, G. (1987). Animal Genetics **18**:41-50.

GEORGES, M.; LATHROP, M.; HILBERT, P.; MARCOTTE, A.; SCHWERS, A.; SWILLENS, S.; VASSART, G.; HANSET, R. (1990). Genomics **6**: in press.

HILBERT, P.; MARCOTTE, A.; SCHWERS, A.; HANSET, R.; VASSART, G.; GEORGES, M. (1989). Animal Genetics **20**:383-394.

HOESCHELE, I.; MEINERT, T.R. (1989). Association of genetic defects with yield and type traits: a major production gene is linked to Weaver. Journal of Dairy Science, submitted.

JEFFREYS, A.J.; WILSON, V.; THEIN, S.L. (1985). Nature **314**:67-73.
LITT, M.; LUTY, J.A. (1989). American Journal of Human Genetics **44**:397-401.

NAKAMURA, Y.; LEPPERT, M.; O'CONNELL, P.; WOLFF, R.; HOLM, T.; CULVER, M.; MARTIN, C.; FUJIMOTO, E.; HOFF, M.; KUMLIN, E.; WHITE, R. (1987). Science **235**:1616-1622.

ORKIN, S.H. (1986). Cell **47**:845-850.

ROYLE, N.J.; CLARKSON, R.E.; WONG, Z.; JEFFREYS, A.J. (1988). Genomics **3**:352-360.

SMITH, C. & SIMPSON, S.P. (1986). J. Anim. Breedg. Genet. **103**:205-217.

SOLLER, M. & BECKMAN, J.S. (1982). 2nd World Congress on Genetics Applied to Livestock Production. Madrid **6**:396-404.

TAUTZ, D. (1989). Nucleic Acids Research **17**(16):6463-6471.

WEBER, J.L.; MAY, P.E. (1989). American Journal of Human Genetics 44:388-396.

WHITE, R.; LALOUEL, J.M. (1988). Annual Reviews of Genetics 22: 259-279.

WOMACK, J.E.; MOLL, Y.D. (1986). Journal of Hererdity 77:2-7.