Proc. Assoc. Advmt. Anim. Breed. Genet. ***17****: 471-478*

APPLICATION OF GENOMIC INFORMATION IN A DAIRY CATTLE BREEDING SCHEME

R.J. Spelman[1], M. Keehan[1], V. Obolonkin[1] and W. Coppieters[2]

[1]LIC, Private Bag 3016, Hamilton, New Zealand.
[2]University of Liege, Liege, Belgium.

SUMMARY

Marker assisted selection has been applied in dairy cattle breeding schemes with minor to moderate improvements in genetic gain. The cost effectiveness of it has been neutral at best in most cases. The completion of the sequencing of the bovine genome in 2006 has generated a large number of single nucleotide polymorphisms to become commercially available. In addition, the cost of genotyping SNPS is markedly reduced compared to microsatellite genotyping. This technology shift will enable the ability to use genomic information in breeding schemes in a far more efficient manner through genomic selection.

INTRODUCTION

DNA information has been used in dairy cattle breeding schemes since the early 1990's when parentage testing was transferred from blood protein polymorphisms to microsatellite markers. At the same time the detection of quantitative trait loci (QTL) for dairy cattle had just commenced (Georges *et al.* 1995). Since this date there has been a flood of QTL being reported for dairy cattle traits. In at least three instances, the QTL have been positionally cloned; chromosome 14 – DGAT1 (Grisart *et al.* 2002), chromosome 20 – GHR (Blott *et al.* 2003) and chromosome 6 (Ron *et al* 2006, Schnabel *et al* 2005). In the instance of chromosome 6 there is some debate over which gene and polymorphism underlies the QTL. With the small number of QTL that have been positionally cloned there has been a reliance of using linked markers rather than the functional polymorphism itself in dairy cattle breeding schemes.

Utilisation of QTL in dairy cattle breeding schemes via marker-assisted selection (MAS) has been thoroughly researched in a theoretical setting through simulation studies (eg. Meuwissen and Goddard 1996, Spelman and Garrick 1997). The increases in rates of genetic gain through MAS varied from 2-3% to an increase of 50%. This variation was mainly due to the underlying genetic model that was simulated and the proportion of genetic variance that was explained by the QTL.

The implementation of marker assisted selection has occurred in a number of dairy cattle breeding schemes (Spelman 2002, Boichard *et al.* 2006). Boichard *et al.* 2006 reported that the French dairy industries use of 43 markers had resulted in a improvement in genetic response by approximately 5-19%. They concluded that MAS was economically beneficial to the French breeding companies.

The sequencing of the bovine genome by the international consortium (Kappes *et al.* 2006) has generated a large number of single nucleotide polymorphisms that have been deposited in the public domain. In conjunction with companies such as Affymetrix and Illumina, large scale SNP genotyping at low cost can now be undertaken in bovine. Whereas the genotyping cost for a microsatellite was approximately NZ$2.50 per genotype two years ago, now a SNP genotype costs approximately one NZ cent when tens of thousands of SNP are typed in parallel. The technology shift to large scale SNP genotyping has and will have a major effect on the utilisation of markers in

dairy cattle breeding schemes.

This paper will outline the current use of markers in the LIC breeding scheme and describe the biological resources that have been generated to develop the next wave of markers and genomic tools that will be used in of LIC's breeding business.

PARENTAGE TESTING

DNA parentage testing of bulls entering the progeny test system has been undertaken for the last 15 years based on molecular markers. Prior to this parentage testing was undertaken on blood group systems. For the last 15 years, LIC has only progeny tested bulls that have passed (a animal that has one marker inconsistent with inheritance is accepted as passed) parentage to both sire and dam. In the past 10 years, this has been extended to testing the dam also against her sire. The parentage testing has identified an error rate of 4-6%.

In 1997 LIC, undertook DNA parentage (paternal only) testing of all progeny test daughters. The level of paternal misidentification was identified to be 12-15%. This was not unexpected, as the majority of cows calve over a 6 week period, and calving is managed by 1 person per 200 cows. The error rate is higher for the daughters than for the bulls, as the bulls are usually generated from specific matings where greater attention is given to that cow at calving.

In the first months of 2007 LIC has moved its parentage testing platform from a microsatellite based system to one using single nucleotide polymorphisms (SNPs). The major drive for the change was to reduce the cost of parentage testing and thus increase the level of uptake by dairy farmers.

Dairy farmers are utilising parentage on the whole herd. The farmer is sending in all of the sires used on in the herd and the majority of the dams (average herd size 300) and the retained progeny (approximately 20-25% of herd size). The expectation is that LIC will match all of the progeny to sire and dam. Fisher *et al.* (2007) calculated through stochastic simulation and with a genotyped herd that 40 SNPS that have a minor allele frequency of greater than 0.35 will have comparable statistical power to current microsatellite sets of 14 markers. It should be noted that 40 SNP markers will not have sufficient power of exclusion alone to establish sire and dam. However, in conjunction with mating records and calving dates the level of resolution can approach 100%.

SINGLE GENE TESTING

In the last 20 years many molecular tests have been available to dairy breeders for identification of single gene defects; BLAD, citrullinaemia, DUMPS and CVM. The strategy in the past with single gene tests has been to test all bulls entering progeny test and remove carriers. However, the use of electronic recording of matings has given LIC the opportunity to use carrier bulls but avoid carrier x carrier matings. This is achieved by calculating the probability that each cow is a carrier based on carrier status of her ancestors. Before the technician inseminates the cow he/she enters the cow and bull identification. If the bull is a carrier and the cow is above a certain probability, the technician is queried or advised to choose a non-carrier bull.

Ongoing screening of animals in the progeny test system periodically detects new genetic defects. During the 2002 calving season a skin condition in a number of newborn calves was reported from progeny-testing herds. In most cases the condition was so severe that the calves required euthanasia on humane grounds. Pedigree records showed that the calves were from the same sire and the inherited condition, epidermolysis bullosa (EB) was suspected. EB is a family of mechanobullous disorders characterized predominantly by skin fragility resulting in blistering and defects of the skin

and mucous membranes, among other tissues (Uitto and Pulkkinen 2001).

Of 302 confirmed pregnancies, 187 calves exhibited no symptoms, 17 had clinical signs of EB, three were stillborn (cause unknown), two aborted, and the status of 93 was unknown. The latter category was most likely made up of male calves sent for slaughter soon after birth.

Comparative mapping of 2 known mutations that cause EB in humans was undertaken to identify the putative positions on the bovine genome. Microsatellites were genotyped for both chromosomes, which identified a segregation pattern consistent with phenotype for one chromosome. The candidate gene on that chromosome was sequenced and a single point mutation was identified within the Keratin 5 gene (Ford *et al.* 2005). It was identified that one mutant allele was sufficient to cause disease. However, the sire was unaffected (although heterozygous) and the low proportion of affected offspring derived from this sire lead us to consider mosaicism. The ancestors of the sire did not have the mutation indicating that the mutation must have arisen spontaneously during the sire's development. The possibility of reduced penetrance was excluded through an analysis of 50 unaffected animals of the sire, none of which carried the mutant allele.

The continued identification of genetic defects such as that described above has been enhanced with the new SNP genotyping platforms. Genotyping approximately 10-20 progeny with a 10,000 or more SNP panel will identify SNPs that co-segregate with the phenotype if it is genetically controlled. Due to the greater density of markers compared to the previous microsatellite markers, the identification of smaller genomic regions and thus more targeted positional candidate gene identification will occur. This approach has been successfully undertaken for a genetic condition in Belgian Blue cattle (M. Georges personal communication).

MARKER ASSISTED SELECTION

As reported by Spelman (2002), LIC started MAS in 1998 using 6 QTL that had been identified from a grand-daughter design. The QTL, which affected milk production traits, were used in a within-family MAS setting where the phase of the QTL alleles were re-estimated for each of the sire families. The sires that were identified to be heterozygous for the QTL of interest were used to generate multiple full-sib sons. This was undertaken through a combination of MOET and IVP reproductive programmes. Full-sibs were deemed to be required as generating half sibs through artificial insemination would result in the selection advantage of the QTL information being negated through the loss of selection differential on the cow to dam pathway. The resulting male offspring were genotyped, with the sons that received the desired alleles selected to enter progeny testing.

The reproductive performance of the donor cows was poor and very few of the families had enough sons to allow within-family marker-assisted selection. After two years of poor reproductive performance the within-family MAS was abandoned. The results from these two years indicated that methods that used markers in a BLUP setting or markers that were in linkage disequilibrium would be required for more effective utilization of MAS.

In 2002 Grisart *et al.* reported the identification of a functional mutation in DGAT1, which was closely followed by Blott (2003) for the identification of a polymorphism in the GHR gene. Both of these genes have been used in the LIC breeding scheme for the last 4-5 years. For each of the two genes allelic effects have been estimated from over 3000 sires for the New Zealand dairy population (Spelman *et al.* 2002, 2003). All bull dams and bulls entering the progeny testing scheme are genotyped for the two genes. For each animal genotyped the breeding value of the applicable traits were adjusted by the previously estimated allelic effects, in a manner similar to that described by

Spelman and Garrick (1997). That is, where an animal has many lactations of information collected or progeny evaluated, a proportion of the allelic effects will already be represented in the breeding value. Thus the allelic effects need to be regressed by the reliability of the animals breeding value to ensure that the allelic effect is not double counted. Periodically the new genotypes and phenotypes were added to the initial experimental datasets to re-estimate the allelic effects.

The utility of the GHR gene has been reduced by the high allele frequency of the economically superior allele (Spelman *et al.* 2003). However, in some cases there were heterozygous sires that were being used to generate sons for entry into the progeny testing system. For these sires, the sons that received the favourable T allele from the sire were selected for the progeny testing scheme. In addition, all other young bulls were genotyped for GHR and their breeding values adjusted as previously outlined.

DGAT1 is at more intermediate frequency than GHR in the NZ dairy industry. Although DGAT1 has significant effects on fat, protein and milk yield, it's applicability is negated by the small overall impact on the breeding index applied in NZ (Spelman *et al.* 2002). For both the Jersey and HF breeds the overall effect is approximately 0.05 of a standard deviation of the breeding index.

To increase the pool of QTL available to the breeding scheme, LIC with ViaLactia BioSciences undertook a_n experiment involving the Holstein-Friesian and Jersey breeds (Spelman *et al.* 2004). Eight hundred F_2 animals were generated from 6 F_1 sires and approximately 800 high genetic merit F_1 cows. The F_2 cows born over 2 years commenced their first lactation in 2001 and 2002. The animals have been extensively measured over this period for phenotypes including milk production, milk characteristics, methane production, health traits (including a mastitis challenge), and fertility parameters. The animals (F_0 males, F_1 and F_2 animals) have been genotyped with approximately 300 microsatellite markers and the Affymetrix 10K SNP panel.

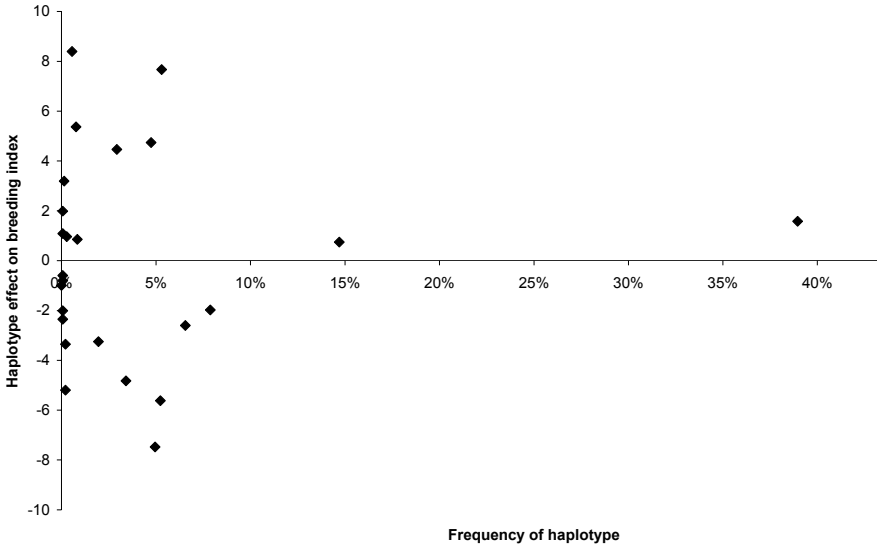

Figure 1: Haplotype frequency and haplotype effects for the *Enhancer* QTL in the Holstein-Friesian population.

The first QTL identified from this experiment is being utilised in the LIC breeding scheme. The QTL, termed Enhancer, has been fine mapped to where it is in linkage disequilibrium with a panel of markers. Haplotype effects have been estimated from the F2 experiment and then validated and re-estimated based on 4000 progeny tested sires. The Enhancer QTL affects all of the milk production traits in the breeding index as well as live weight. The haplotype effects have a range of 16 index points, which is equivalent to 0.3 of a standard deviation of the breeding index (Figure 1). The better haplotypes in terms of the effect of the breeding index are at frequencies less than 10% thus there is selection potential for this QTL in terms of effect and frequency.

In addition, 500 of the F_2 animals have been analysed on the Affymetrix gene expression microarray for both adipose and fat tissues. The microarray phenotypes are being used in a genetical genomics setting (Jansen and Nap 2001) with the objective of identifying regions of the genome that have an effect on the regulation of key pathways for fertility and milk production. Preliminary results to date have identified regions of the genome that have a large degree of trans control of RNA transcripts (Figure 2). The horizontal bands in Figure 2 identify genome regions where there is trans or distant control of a large number of transcripts (eg. chromosome 10). The diagonal line represents the cis or locally controlled transcripts.

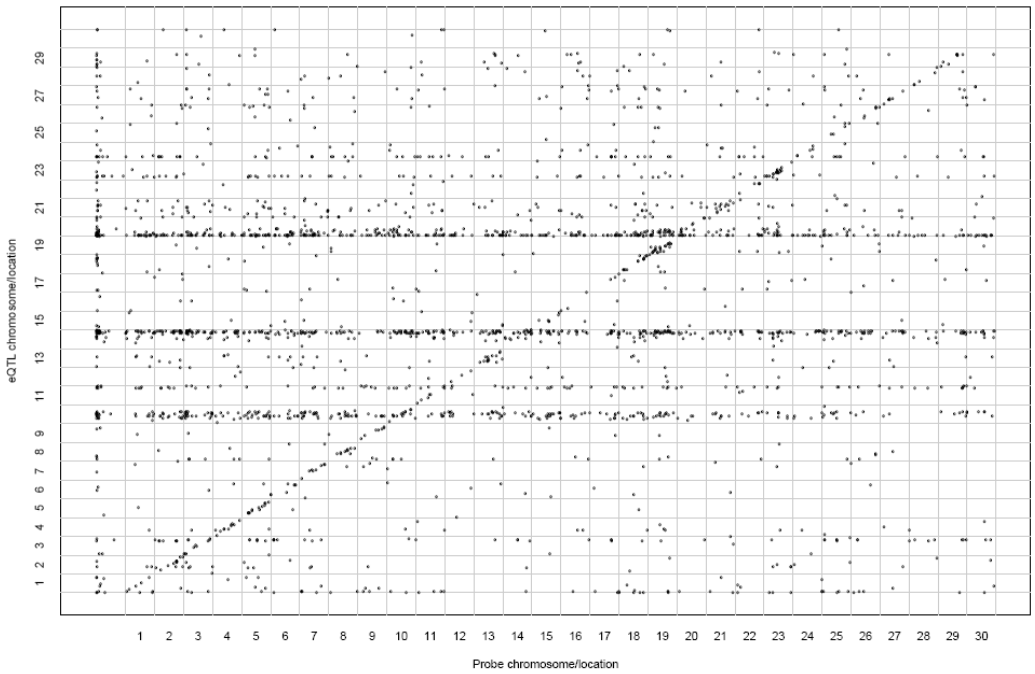

Figure 2: Results of mRNA expression profiling from fat tissue across the F_2 QTL mapping population. The x-axis is the physical chromosomal map position of the probe in the Affymetrix expression array. The y-axis is the chromosomal map position of the QTL where it is significant at the 0.001 threshold level in an additive QTL model for the probes on the expression array.

In total, there are 228 cis controlled transcripts from fat tissue that are significant at the 0.001 threshold level and 1036 trans controlled transcripts at the same threshold. The high number of trans controlled transcripts is inflated by the threshold level for the trans effects not accounting for the multiple testing of the transcript against many markers for linkage that cover the whole genome. In contrast, for the testing of cis effects the transcripts are only tested against the markers near the gene encoding the transcript. The significant transcripts are now being analysed in a network setting (eg. Bing and Hoeschele 2005) to identify the underlying regulatory pathways.

GENOMIC SELECTION

Meuwissen *et al.* (2001) first proposed the use of dense marker maps in a genomic selection (GS) setting. Through simulation they found that by estimating breeding values from genomic information (one microsatellite per cM), the accuracies of selection were 0.75-0.85. At the time of the paper being published, cost estimates to undertake the work on 2000 sires with 4000 microsatellites was approximately NZ$20 million. With the sequencing of the bovine genome and the commercial SNP panels being developed, the cost of this experiment is a now a factor of 10-20 less.

LIC progeny tests 300 bulls per year at an average cost of $40,000 per bull. Progeny testing in dairy cattle schemes is lengthy with the bulls receiving their progeny proof when they are 5 years of age. With genomic selection the bulls will have a genomic proof available a few days after birth. Schaeffer (2006) has presented a theoretical analysis of genomic selection based on the Canadian Holstein population. He proposed to firstly estimate the marker effects by genotyping 2500 sires previously progeny tested. Utilisation in the breeding scheme was to select firstly 2000 cows using conventional methods and then to select 1000 of these cows using genomic information. Twenty bulls were then selected from the 500 male progeny of these cows, using genomic information and these bulls were used on the general population from one year of age (for two to three years), without progeny testing, with the best few used to mate selected bull dams for the next year. This scheme is predicted to cost less than 10%, and result in more than double the genetic gain per year, compared to the current progeny testing scheme. The analysis of Schaeffer (2006) essentially assumes that GS is based on the same traits as are available with initial progeny test results. Additional gains could be made by incorporating information on traits such as longevity and fertility that would not normally be available until after selection of bulls for widespread usage. Currently the majority of information for both of these traits is collected after the sire has had one to two to years of widespread use and thus the majority of his direct genetic contribution is completed before accurate estimates are available. It also should be noted that there are some technical errors in the Schaeffer (2006) paper regarding selection intensities and generation intervals. Re-working the analysis results in a 75-80% increase in rate of genetic gain for GS (117% in the original paper) compared to the current progeny testing scheme, which is still a substantial improvement and is consistent with the general findings.

LIC is undertaking a genomic selection experiment with approximately 3000 bulls from DNA that has been stored since 1980. The animals will be genotyped on an Illumina panel of 60,000 SNPs. More recent estimates of the level of LD in dairy populations (Hayes *et al.* 2006, Spelman and Coppieters 2006) have identified that LD is shorter than previously thought (Farnir *et al.* 2000). To ensure the accuracies of selection reported by Meuwissen *et al.* 2001 at least 30,000 SNPs are required (Hayes *et al.* 2006). The LIC breeding scheme has three breeds represented in it; Holstein-Friesian, Jersey and crossbred bulls. Therefore a greater number of SNPs will be required to allow the transfer of allelic associations from one breed to the other. Goddard *et al.* (2006) investigated

conservation of LD between Angus and Holstein. They found SNPs that are within 10 kb of a QTL have a high probability (0.98) of being in strong LD with the QTL and this LD will be highly consistent between Angus and Holstein. This would require 300K SNPs with uniform coverage to be able to transfer information between the two breeds. The Jersey and Holstein-Friesian breeds have separated more recently than that of the Holstein and Angus and thus it would be expected that a lower number of SNPs would be required to transfer information from one to breed to the other. The crossbred population also will allow the opportunity to identify genomic regions where heterosis occurs.

Utilisation of genomic selection in the LIC breeding scheme will differ from that presented by Schaeffer (2006). Firstly sires will be used only sparingly at one year of age. The reasons for this are that the semen production of the bull limited at that age and also the danger of using a bull that may have a dominant defect, such as the one described previously. The commercial impact of using a sire with over 50,000 possible calvings that has a dominant defect is too greater risk. Therefore, in the first year the sire will have enough matings to ensure he can be screened for dominant defects and also generate enough daughters for a daughter proof when the sire is 5 years of age. This will accommodate those farmers who are not ready to accept genomic proofs. In the second year, and possibly for another 2 years, the bull will be used over the commercial population. Genomic selection is suited to the way LIC predominantly markets its bulls in a team. The genomic selected bulls will have a lower accuracy of selection compared to the current daughter-proven bulls but when sold in a slightly larger team, there will be no difference in the reliability between the daughter-proven team and genomic-proven team.

The power of genomic selection is to estimate the mendelian contribution the animal has received. Thus rather than focusing on the selection of bull dams using GS tools, it may be more profitable to focus on exploiting the mendelian variation in the bulls. Every year in NZ there are approximately 1.5 million bull calves born from artificial insemination matings. The majority of the bull calves are sold at 4 days of age into the beef market. Thus systems would have to be put in place to ensure potential sires are screened quickly with a reduced SNP panel with the important SNPs. The SNP information will be combined with the pedigree information to estimate a genomic breeding value for the animal. Once a sub-set of the bull population has been selected these bulls will then be screened with the more complete panel of SNPs to get a more accurate genomic evaluation and then enter the LIC bull team. Bulls that are progeny of cows that are not in the elite category will still potentially enter the LIC bull team due to them receiving a very favourable mendelian contribution. Thus the focus for sire analysts will move from selection of elite bull dams for the generation of sons to efficient screening procedures of young bull calves on the ground.

The rate of genetic gain presented by Schaeffer (2006) is greater than that that will be realised due to the use of sires in the commercial population as 2 year-olds thus increasing the generation interval for both sire paths by at least one year. However, there is more potential for increased selection intensity on the sire pathways if widespread screening is undertaken. Schaffer's (2006) prediction that a GS breeding scheme would only cost 10% of the current scheme is optimistic and also depends heavily on the amount of genotyping undertaken to screen animals. Unless the genomic information explains nearly all of the genetic variation there will still be a selection response benefit from including and combining pedigree information in the animal evaluation procedure. The SNP information may be further utilised to adjust the estimate of relatedness between animals from the expectations based on pedigree alone.

REFERENCES

Bing, N. and Hoeschele I. (2005) *Genetics* **170**: 533.

Blott S, Kim, JJ, Moisio, S, Schmidt-Kuntzel, A, Cornet, A, Berzi, P, Cambisano, N, Ford, C, Grisart, B, Johnson, D, Karim, L, Simon, P, Snell, R, Spelman, R, Wong, J, Vilkki, J, Georges, M, Farnir, F, Coppieters, W. (2003) *Genetics* **163**: 253.

Boichard, D., Fritz, S., Rossignol, M.N., Guillaume, F., Colleau, J.J., and Druet, T. (2006) *CD-ROM Eighth World Congr. Genet. Appl. Livest. Prod*. Comm No. 22-11.

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** (2): 220.

Fisher, P.J., Malthus, B., Walker, M.C., Corbett, G. and Spelman R.J.(2007) *Submitted Animal Genetics*

Ford, CA, Stanfield, AM, Spelman, R.J., Smits, B, Ankersmidt-Udy, AE, Cottier, K, Holloway, H, Walden, A, Al-Wahb, M, Bohm, E, Snell, R.G., and Sutherland G.T. (2005) *J Invest Dermatol*.**124** (6): 1170

Georges, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto, et al. (1995) *Genetics* **139**: 907.

Goddard, M.E., Hayes, B., McPartlan, H., and Chamberlain, A.J. (2006) *CD-ROM Eighth World Congr. Genet. Appl. Livest. Prod*. Comm No. 22-16.

Grisart, B., Coppieters, W., Farnir,F., Karim, L., Ford, C., Cambisano, N., Mni, M., Reid, S., Spelman, R., Georges, M. and Snell R. (2002) *Genome Res.* **12** *(2):* 222.

Hayes, B. J.; Chamberlain, A. J.; and Goddard, M. E. (2006) *CD-ROM Eighth World Congr. Genet. Appl. Livest. Prod*. Comm No. 30-06.

Kappes, S. M.; Green, R. D.; and Van Tassell, C. P. (2006) *CD-ROM Eighth World Congr. Genet. Appl. Livest. Prod*. Comm No. 22-01.

Meuwissen, T.H.E., and Goddard, M.E. (1996) *Genet. Sel. Evol*. **28***:* 161.

Meuwissen, T.H.E.; Hayes, B.J.; and Goddard, M.E. (2001) *Genetics* **157**: 1819.

Jansen R.C., and Nap, J.P. (2001) *Trends Genet.***17:**388.

Ron M, Cohen-Zinder M, Peter C, Weller J.I., and Erhardt G. (2006) *J Dairy Sci.* **89**: 4921.

Schaeffer, L.R. (2006): *J. of Animal Breeding and Genetics* **123***:* 218.

Schnabel R.D., Kim J.J., Ashwell M.S., Sonstegard T.S., Van Tassell C.P., Connor E.E., and Taylor J.F. (2005) *Proc Natl Acad Sci* **102***:* 6896.

Spelman, R.J., and Garrick, D.J. (1997) *Livest. Prod. Sci*. **47**: 139.

Spelman,. R.J., Ford, C.A. , McElhinney, P., Gregory, G.C., and Snell, R.G. (2002) *J. of Dairy Science* **85**: 3514

Spelman, R.J., (2002) *CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod*. Comm No. 22-02.

Spelman R.J., J.D. Hooper, Stanley G., Kayis, S.A., and Harcourt, S. (2004) *Proceedings of New Zealand Society of Animal Production* **64**:92.

Spelman R.J., J. Wong, C.A. Ford, R.G, Snell, (2003) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **15**: 94.

Spelman R.J. and Coppieters W. (2006) *CD-ROM Eighth World Congr. Genet. Appl. Livest. Prod*. Comm No. 22-21.

Uitto J, and Pulkkinen L: (2001) *Arch Dermatol* **137**:1458.