

Preface

Procedures for measuring and utilizing records of performance were developed and implemented in meetings of on the farm testing, central testing, national sire evaluation, and carcass evaluation committees held in connection with annual meetings of the Beef Improvement Federation until the early 1980's. As attendance and participation grew in the annual meetings, it became apparent that procedures and guidelines for genetic improvement in beef cattle could be developed and advanced more efficiently by holding workshops focused on specific needs for genetic evaluation. In the first workshop, procedures for computing expected progeny differences (EPDs) to compare animals across many herds were laid out, as an alternative to estimated breeding values (EBVs) which had been computed within herds. For the past two decades, progress in development of procedures and guidelines for genetic improvement of beef cattle have been marked by Genetic Prediction Workshops, a list of which follows:

- 1st Prediction of Genetic Values in Beef Cattle. Winrock International, Morrilton, AR, Dec.12-15, 1983.
- 2nd Prediction of Genetic Values in Beef Cattle II. Kansas City, Mar. 10-11, 1987.
- 3rd Genetic Prediction Workshop. Kansas City, Oct. 16-18, 1989.
- 4th Specification Beef Production. Kansas City, Jan. 21-23, 1994.
- 5th Beef Cattle Body Composition. Kansas City, Jan. 8-9, 1995.
- 6th Genetic Prediction Workshop for Ultrasound Technicians. Kansas City, Dec. 9-10, 1996.
- 7th Marker Assisted Selection, Multi-breed and International Evaluations. Kansas City, Dec. 3-4, 1999.

We are indebted to participants in the planning meeting held May 28, 2003 during the Beef Improvement Federation Annual Meeting in Lexington, KY. They represented the Genetic Prediction Committee and the Emerging Technologies Committee of the Beef Improvement Federation and included members of the technical committee for Regional Research Project NCR-199 on genetic improvement in beef cattle, all responsible for planning this workshop sponsored by the Beef Improvement Federation. Thanks are also extended to the staff at the Embassy Suites, to Debbie Brown (USMARC) who assisted with registration and other arrangements for the meeting, and to Jan Watts (USMARC) who assisted with typing and editing of the Proceedings. We are especially appreciative of the time, effort, and expense that each invited speaker has invested in this workshop.

Today, DNA based technology is emerging rapidly. Thousands of genetic markers have been mapped in the bovine genome, dozens of quantitative trait loci (QTL) have been discovered, DNA markers are used for parentage testing, and new genetic tests are emerging to genotype animals for specific alleles significantly associated with variation in traits of economic importance in beef production (e.g., marbling, tenderness). The challenge is to integrate these techniques with quantitative genetic procedures in genetic evaluation of beef cattle. The purpose of this workshop is to provide an update on the current state of the art and to focus on challenges regarding further development of molecular approaches to genetic improvement in beef cattle.

Larry V. Cundiff
Chairman, BIF Genetic Prediction Committee
U.S. Meat Animal Research Center, ARS, USDA
Clay Center, NE 68933

TABLE OF CONTENTS

Agenda	3
Molecular approaches to genetic improvement. R. Mark Thallman, USDA, ARS, U.S. Meat Animal Research Center	5
Examples of marker assisted selection in sheep and dairy improvement in New Zealand. Dorian Garrick, Colorado State University.....	16
Validation of genetic tests for QTL. Richard Quaas, Cornell University.	35
Industry Panel Jay Hetzel – Genetic Solutions	41
Jim Gibb – Frontier Beef Systems and Geneseek	46
Stewart Bauck and Leigh Marquess – Merial/Quantum	50
Practical implications of using DNA analyses for marker assisted selection. John Pollak, Cornell University.	53
DNA parentage testing. Mike MacNeil – Ft. Keogh Livestock & Range Research Lab	59
Bob Weaber – American Simmental Association and Cornell University	63
Report on carcass merit project. Mark Thallman, USDA, ARS, U.S. Meat Animal Research Center	70
Examples of using QTL in marker assisted selection. Jack Dekkers, Iowa State University.	91
Statistical issues in marker assisted selection. Rohan Fernando, Iowa State University.....	101
Panel discussion of statistical and computational challenges to development and application of marker assisted selection. Daniel Gianola – University of Wisconsin.....	109
Steve Kachman – University of Nebraska	121
Ignacy Misztal – University of Georgia	123

**BEEF IMPROVEMENT FEDERATION 8TH GENETIC PREDICTION WORKSHOP
“MOLECULAR APPROACHES TO GENETIC IMPROVEMENT”**

**Embassy Suites – KCI Airport, 7640 NW Tiffany Springs Parkway
Kansas City, Missouri
December 4-6, 2003**

December 4 (Thursday)

7:30 PM to 9:30 PM - NCR 199 Technical Committee Meeting

December 5 (Friday)

Chairman – Larry Cundiff, MARC, ARS, USDA

- 8:00 AM Molecular approaches to genetic improvement. Mark Thallman, MARC, ARS, USDA, Clay Center, NE
- 9:00 AM Integrating quantitative and molecular genetics- Daniel Pomp, University of Nebraska
- 9:45 AM Break

Chairman – Craig Huffhines, American Hereford Assoc.

- 10:15 AM Examples of marker assisted selection in sheep and dairy improvement in New Zealand. Dorian Garrick, Colorado State University
- 11:00 AM Validation of genetic tests for QTL. Richard Quaas, Cornell University
- 12:00 Lunch

Chairman – Robert Williams, American International Charolais Association

- 1:00 PM Sequencing the bovine genome - Steve Kappes, MARC, ARS, USDA, Clay Center, NE
- 1:30 PM Industry Panel on Currently Available Genetic Tests and Future Plans – Moderator, Ronnie Green , NPS, ARS, USDA
 - Jay Hetzel - Genetic Solutions
 - Jim Gibb –Frontier Beef Systems and Geneseek
 - Sue Denise - Metamorphics
 - Stewart Bauck and Leigh Marquess – Merial/Quantum
- 3:00 PM Break

3:30 PM Practical Implications of using DNA analyses for marker assisted selection. John Pollak, Cornell University

4:00 PM DNA parentage testing

Experimental setting. Mike MacNeil, Ft Keogh Livestock and Range Research Laboratory, ARS, USDA, Miles City, MT

Ranch setting. Bob Weaber . American Simmental Association and Cornell University

December 6, (Saturday)

Chairman – Ronnie Green, NPS, ARS, USDA

8:00 AM Report on carcass merit project – Mark Thallman

9:00 AM Examples of using QTL in marker assisted selection – Jack Dekkers, Iowa State University

9:45 AM Break

10:15 AM Statistical issues in marker assisted selection – Rohan Fernando, Iowa State University

11:00 AM Panel Discussion of Statistical and Computational Challenges to development and application of Marker Assisted Selection –

Moderator - Rob Tempelman, Michigan State University

Daniel Gianola – University of Wisconsin

Steve Kachman - University of Nebraska

Ignacy Misztal _ University of Georgia

Rohan Fernando – Iowa State University

12:00 Lunch

MOLECULAR APPROACHES TO GENETIC IMPROVEMENT

R. Mark Thallman
U.S. Meat Animal Research Center
Agricultural Research Service, U.S. Department of Agriculture
Clay Center, NE 68933

Benefits of Genetic Testing

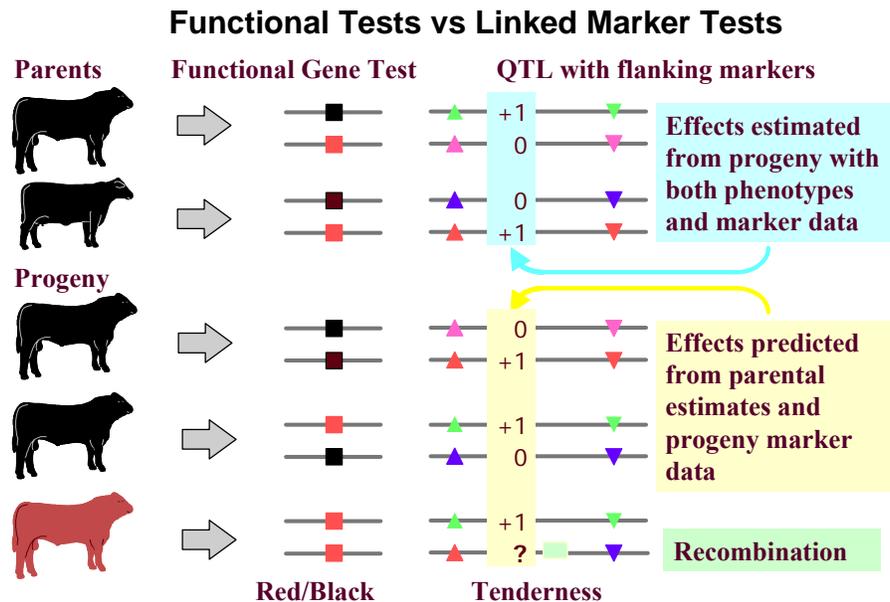
- Obtain evaluations earlier in the life cycle.
- Increase accuracy of selection, especially for traits that are expensive to measure, sex-limited, or measured postmortem.
- Capture more benefit from each phenotype that is measured.
- Increase opportunity to select for traits with antagonistic genetic relationships (e.g., birth weight and growth rate).

Realistic Expectations

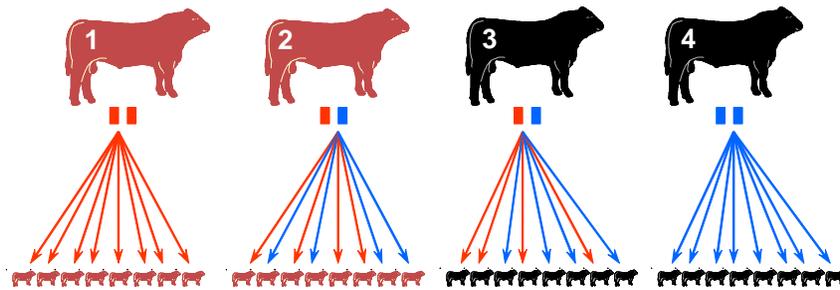
- DNA testing can increase the amount of information that each phenotype contributes.
- DNA testing can reduce the number of phenotypes needed, but DNA testing can not replace phenotypes.
- DNA testing will probably make cattle breeding more complicated, not easier.

Types of Genetic Tests that Could Be Available

- Linked Markers – highly polymorphic markers within a few cM of, and generally assumed to be in linkage equilibrium with, the functional polymorphism
- Functional Tests – polymorphisms that directly affect phenotypes of interest
- Association Tests – polymorphisms in high linkage disequilibrium (and generally physically close to) one or more functional polymorphisms that affect phenotypes



Linked Marker Tests



Advantages of Linked Marker Tests

- They are relatively inexpensive to develop
- We have considerable experience in using them, at least in QTL detection experiments
- They tend to be highly polymorphic

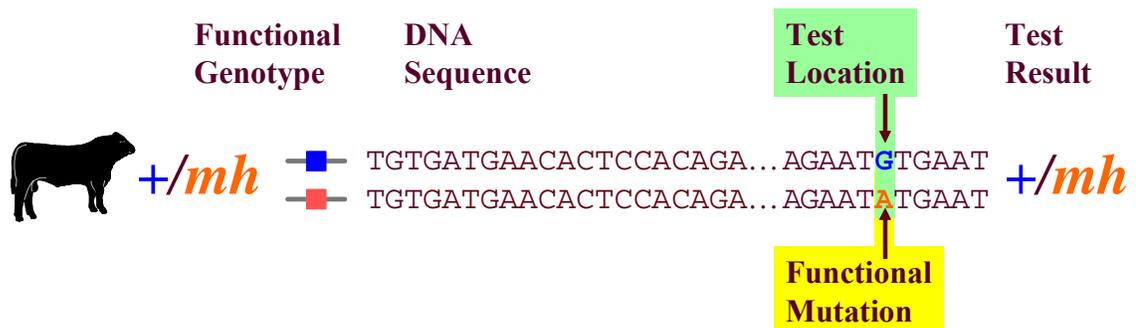
Challenges in Applying Linked Markers

- Must establish phase between markers and QTL within each family
- Must estimate QTL effects within family
- Therefore, marker data must be collected on a substantial number of individuals in each family
- The required statistical analysis is computationally demanding
- Breeders cannot easily interpret the test results themselves

Commercialization of Linked Markers Depends On:

- Testing technology that is inexpensive enough to be applied on a whole-herd basis
- Existence of the infrastructure to incorporate the information into NCE
- Not likely to be used widely

Functional Tests



Advantages of Functional Tests

- Test results are easier to interpret
- Does not require extensive testing of relatives
- *We don't have as much experience using them in livestock, so their disadvantages are not as apparent as those of linked markers.*

Challenges in Applying Functional Tests

- Very expensive to develop
- Some will not detect all of the functionally different alleles in the population.
- Different size of effect exists between breeds or production systems

Myostatin is an example in which a functional test is available.



Undetected Functional Alleles

	Functional Genotype	DNA Sequence	Test Location	Test Result
	$+/mh$	-  - TGTGATGAACACTCCACAGA... AGAATGTGAAT -  - TGTGATGAACACTCCACAGA... AGAATATGAAT		$+/mh$
	$+/mh$	-  - TGTGATGAACACTCCA CAGA... AGAATGTGAAT -  - TGTGA-----CAGA... AGAATGTGAAT		$+/+$

Undetected mutation causes same phenotype as animal above

- These undetected functional alleles are likely to be common and will only be detected if phenotypes continue to be collected and associated with DNA test results.

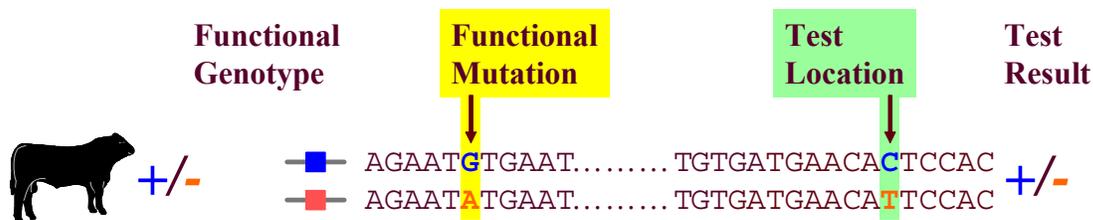
Consequences of Undetected Functional Alleles

- May result in decreased accuracy of the test, depending on the frequency of the undetected alleles
- The inaccuracies may be infrequent, but very large, and are likely to go undetected for a long time. When they are recognized, the consequences could be considerable
- Could cause underestimation of the effect and degree of dominance of the gene in populations with undetected alleles

Different Sizes of Effects Between Breeds will Present Challenges in Incorporating Test Data into NCE

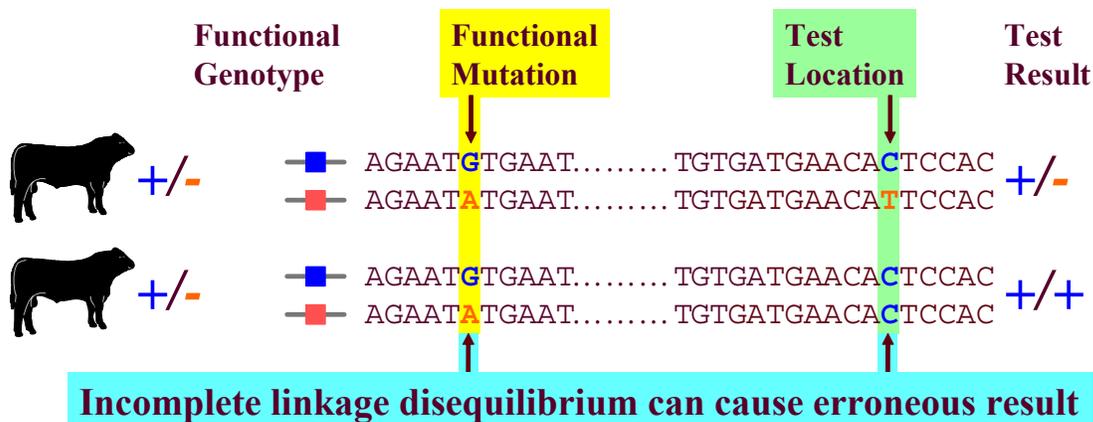
Genotype	+/+	-/-
Breed A	+5	-5
Breed B	+8	-8

Association Tests



- Association is due to linkage disequilibrium
- Share many of the advantages and disadvantages of functional tests
- May be very difficult to distinguish from functional tests
- Easier to develop because it is not necessary to prove that the polymorphism being tested causes the effect on phenotype

Challenge in Applying Association Tests



- When this situation is detected, more SNP should be added to the test

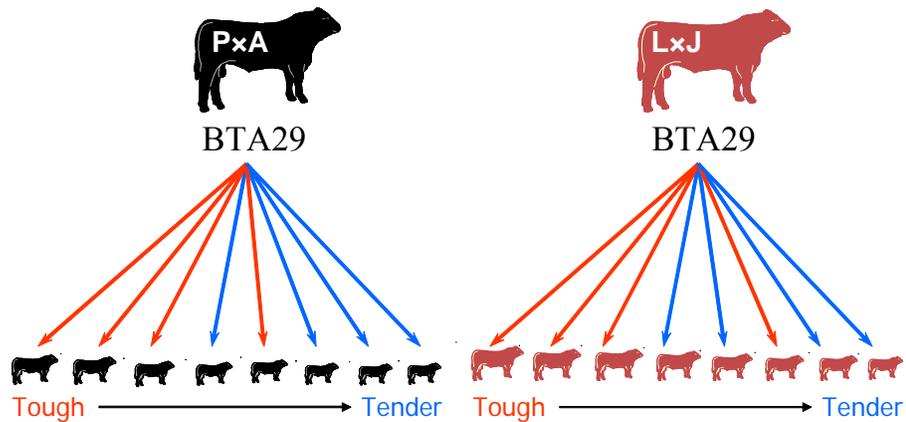
Which Type of DNA Test Will Be Most Widely Used in the Near Future?

- Association tests

Case Study: μ -Calpain

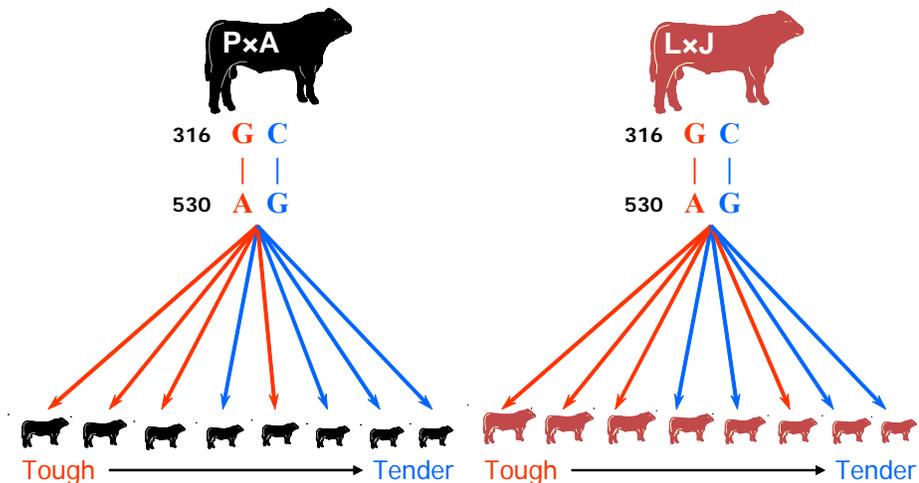
- QTL for Warner-Bratzler Shear Force on BTA29
- μ -calpain gene located under QTL peak
- μ -calpain is a proteolytic enzyme that plays a key role in postmortem tenderization of meat
- U.S. Meat Animal Research Center
- Tim Smith, Eduardo Casas, Roger Stone, Brent Page, Stephen White, Mohammad Koochmaraie
- μ -calpain was first discovered as a QTL on BTA29 in large paternal half sib families sired by a Piedmontese \times Angus F₁ bull at MARC and a Limousin \times Jersey F₁ bull in New Zealand.

BTA29 QTL



- The μ -calpain gene was sequenced in both of these F₁ bulls and they were found to have the same genotypes at the only two SNP that resulted in amino acid changes.

SNP Association



Possible Test Results Can be Viewed from Two Perspectives

		SNP 530			Haplotype				
		GG	AG	AA		C G	G G	G A	C A
SNP 316	CC	CC GG	CC AG	CC AA	C	CC	CG	CG	CC
	CG	CG GG	CG AG	CG AA	G	GC	GG	GG	GC
	GG	GG GG	GG AG	GG AA	G	GG	GG	GA	GA
					A	GC	GG	GG	GC
					A	AG	AG	AA	AA
					C	CC	CG	CG	CC

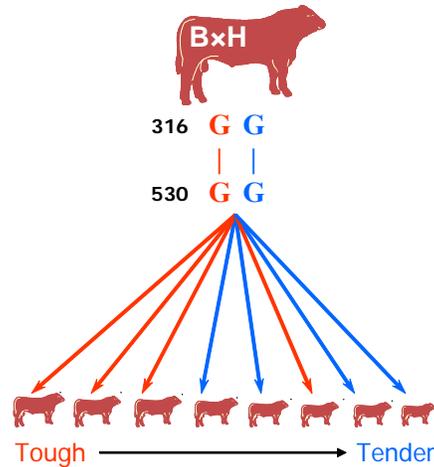
- Interpretation of the results becomes considerably more complicated with more than one SNP.

Two Possible Models for Observed Intermediate Haplotype Effect

Avg. Effect of Haplotype	Three Functional Alleles	Mixture of Two Functional Alleles
Tender	316 C 530 G	316 C 530 G
Intermediate	316 G 530 G	316 G G 530 G G
Tough	316 G 530 A	316 G 530 A
Too Rare to Estimate Well	316 C 530 A	316 C 530 A

A Challenge and an Opportunity

- A Brahman x Hereford F₁ bull was heterozygous for the QTL, but homozygous at the test loci.
- Red vs blue alleles in progeny are determined by flanking microsatellites.
- This situation suggests rather strongly that the G-G Haplotype is actually associated with a mixture of functional alleles.
- It was not the result we were looking for, but it provided an opportunity to make the test more powerful.



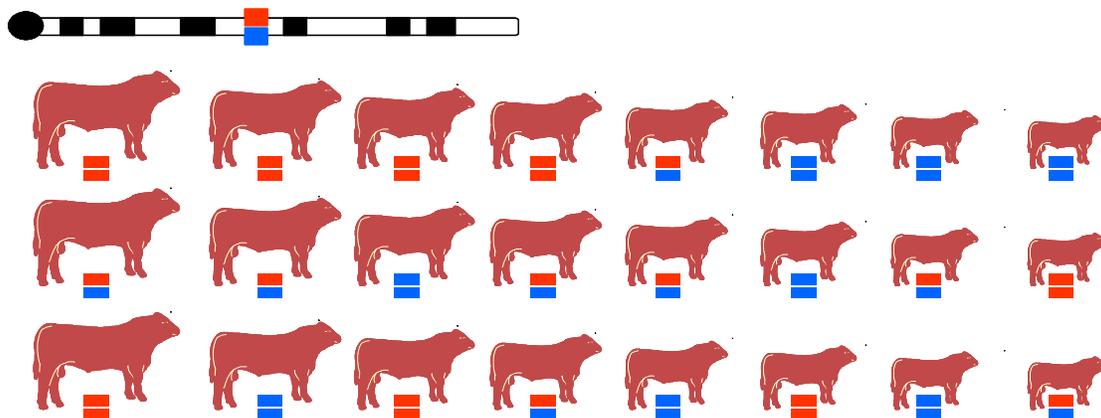
Why Did We Discover this Opportunity to Improve the Test?

- Because we looked
- Because we had an appropriate population in which to find it

We Are Learning a Great Deal From the μ -Calpain DNA Test

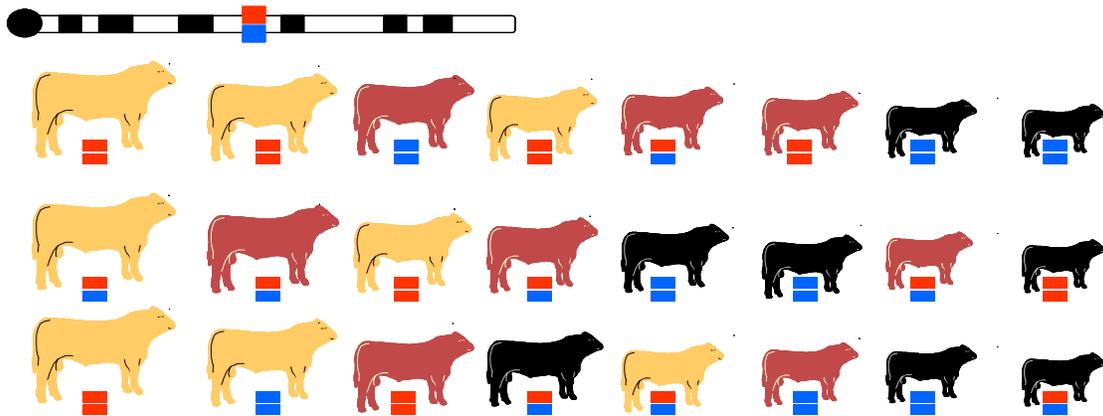
- Multiple SNP present some challenges.
- The test works very well as it is.
- I am confident that it will work even better when we add a few more SNP to it.

QTL Detection by Association (Linkage Disequilibrium)



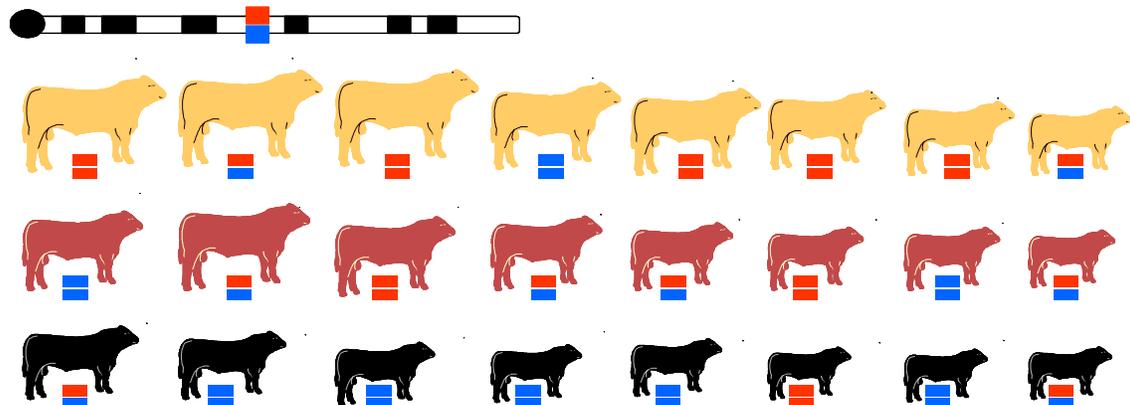
- In this figure, animals with red alleles tend to be larger than those with blue alleles. One possible explanation is that allele color is in linkage disequilibrium with a gene influencing size.
- This type of analysis usually assumes that animals are unrelated, so it is difficult to control for factors besides the QTL.

QTL Detection by Association (Linkage Disequilibrium)



- It is often desirable to use multiple breeds to increase the inference space in this type of project.
- In this figure, animals with red alleles certainly are larger than those with blue alleles.

QTL Detection by Association (Linkage Disequilibrium)



- However, when the cattle from the previous figure are sorted by breed, it is clear that there is no within-breed association between allele and size.
- The overall association between allele and size does not imply that the marker locus is in proximity to a gene that affects size.
- If breeds differ substantially for a phenotype, any marker for which the breeds differ in allele frequency will show an association.
- Therefore, it is critical to adjust for breed and pedigree in association studies.

Importance of Proper Design of Association Studies

- As illustrated in the previous set of diagrams, ignoring genetic effects (breed or pedigree) in association studies can easily lead to the conclusion that a polymorphism has a useful genetic effect, when in fact, it does not.
- A common tactic in association studies is to use populations in which pedigree (and perhaps even breed composition) is unknown. Then the lack of information is used as an excuse for ignoring these genetic effects in the analysis.
- The justification often used is that the goal is to be able to use the test on animals with unknown genetics, but this is faulty logic. In reality, we want the association to be due to linkage disequilibrium, not to some nebulous population stratification.
- In beef cattle, the most common abusive situation is tests on “feedlot cattle.” Pens of feedlot cattle usually fall into at least one of the following situations: 1) they are of heterogeneous breed composition and it is not recorded or 2) they are ranch-raised pens with paternal half-sib groups of substantial size.
- In dairy cattle, a similar, yet different situation occurs. It is essentially impossible to have large groups of “unrelated commercial Holstein cows,” but association studies are sometimes performed on populations in which such claims are made in order to avoid accounting for pedigree in the analysis. Similar, even if less extreme, examples of ignoring relevant pedigree information can be found in beef cattle studies.
- It is not unreasonably difficult to find populations of cattle with known breed composition and pedigree. In fact, most cattle that have extensive phenotypic information meet that criteria and evaluation of DNA tests is most effective when conducted on cattle with phenotypes for a number of traits.
- The cost per animal of DNA testing is no greater for properly designed and analyzed populations than for poor ones, but the information generated can be much greater and free of bias.
- Evaluating DNA tests in such “real world situations” makes for impressive sound bytes, but it is not the most productive use of resources. A population with known breed composition and pedigree can easily be analyzed ignoring those factors (for the real world application), but those factors can not be added to the analysis of a population in which they are unknown.
- If a polymorphism has a real effect that is functional or due to tight linkage disequilibrium, it should be present even if genetic effects are ignored. In fact, failure to find an effect in a “real world” population with sufficient numbers of each genotype would imply that a test does not work, although the statistical power would be better (because residual variance would be lower) if genetic effects were included. However, the converse is not true; if a significant association is detected in the absence of breed and pedigree effects, it is difficult or impossible to distinguish whether the effect is due to population stratification (which is not very useful) or to either linkage disequilibrium or a functional effect (which are useful).
- It can be argued legitimately that associations that are due to population stratification could be useful for predicting the performance of commercial animals with unknown genetics, for example, in a feedlot. The basis for the

argument is that the polymorphisms differ in frequency between breeds and, hence, are useful as predictors of breed composition. However, it seems obvious that a large number of such markers would be required to account for a substantial fraction of phenotypic variance and that the same fraction of variance could be accounted for by a much smaller number of markers that are, or are in linkage disequilibrium with, a functional polymorphism. The latter polymorphisms may also be useful in estimating breed composition. The number of markers is especially important in the feedlot application because of the current laboratory cost per polymorphism relative to the limited potential value per animal of genetic information in the feedlot. Furthermore, markers with associations due only to population stratification have no value for testing seedstock for selection purposes. It seems logical to focus on tests that can be used in both markets.

- In spite of these comments, when used properly linkage disequilibrium is a powerful tool and association tests will certainly contribute to beef cattle improvement. They just need to be evaluated in appropriate populations with appropriate models.

Recommendations for Associations Discovered Through Linkage Disequilibrium

- Validation should include checking for segregation in large families in which flanking microsatellite (or SNP) markers have been scored in the progeny.
- The progeny should segregate according to the prediction of the association test for the parents(s) (Transmission Disequilibrium Test).
- Should be a routine part of the protocol, just like testing DNA markers for Mendelian segregation.

Prioritization of Resources in Developing New Tests and Refining Existing Tests

- Need high-throughput systems for converting QTL into association tests
- Need a new generation of resource populations
- Microsatellite markers will continue to be an important resource, especially in developing robust association and/or functional tests based on SNPs.
- Converting QTL to association tests should be a high priority.
- Refining existing tests should be done primarily when opportunities for improvement are obvious. In the future, this will become a higher priority.
- In general, spending considerable resources to convert association tests into functional tests is likely to be less productive than the above approaches.

Analysis of Multiple-SNP DNA Test Data

- Fitting haplotype effects seems more reasonable genetically than fitting effects of multilocus genotypes.
- In many situations, it may be more reasonable to fit haplotypes as random effects.
- Multiple-trait haplotype models may be useful.

Which DNA Tests Should Be Included in NCE?

- A number of questions about a DNA test should be answered before deciding whether to use it or not, but the most important question is “Has it been independently validated?”

Conclusions

- DNA testing is quite challenging and not as simple as it first appears.
- DNA tests should be viewed as very fluid systems.
- The potential benefits are enormous.
- It will eventually become widespread in cattle breeding.
- It will probably eventually become widespread in cattle management.

EXAMPLES OF MARKER-ASSISTED SELECTION IN SHEEP AND CATTLE IMPROVEMENT IN NEW ZEALAND

Dorian J. Garrick^{1,2} and Patricia L. Johnson²

**¹Department of Animal Sciences, Colorado State University,
Fort Collins, CO 80523.**

**²Institute of Veterinary, Animal & Biomedical Sciences, Massey University,
Palmerston North, New Zealand.**

Genetic markers

A genetic marker can be thought of as a single gene trait that can be used to follow the transmission of chromosomes or other traits from parents to offspring. Early genetic markers were blood groups or restriction enzyme sites whereas modern markers are more likely to be microsatellite markers or single nucleotide polymorphisms (SNPs). One use for such markers is in parentage determination which although now widely used in sheep and cattle improvement in New Zealand will not be considered further in this paper.

The transmission of a genetic marker from parent to offspring can provide information about the inheritance of a chromosome region encompassing the marker. The exact size of each chromosome region transmitted intact from parent to offspring cannot be determined as there will usually be crossover events upstream and downstream from the marker. Adding further markers up- and downstream on the same chromosome can however provide boundaries on the likely size of the region.

Knowledge of marker inheritance provides some information about the likely genetic merit of the offspring and therefore provides an opportunity for so-called marker-assisted selection (MAS). From a theoretical basis, it can be helpful to distinguish three different circumstances whereby markers can be used to assist breeding decisions (de Koning, Dekkers and Haley, 2003). These circumstances are best described by example (in the following section) whereby the marker locus has two alleles (M and m) and a linked major gene locus has two alleles (Q and q) of interest for selection.

Classes of marker-assisted selection

The poorest circumstance for marker-assisted selection occurs when the marker is in linkage equilibrium with the gene that has a significant influence on a trait of interest (LE-MAS). If the marker and the gene are each in Hardy-Weinberg equilibrium, then from a population perspective there will be individuals with marker genotypes MM, Mm and mm (according to the frequency of the marker allele M) and there will be individuals with genotypes for the major gene of QQ, Qq and qq (according to the frequency of Q). If the marker and gene locus are also in linkage equilibrium, then all three marker genotypes will be present in combination with all three genotypes for the major genes with probabilities that can be determined from the products of the genotype probabilities for each separate locus. In practice, this means that from a population perspective the

existence of say the M allele, tells us nothing about the existence of the major gene allele Q. However, from the perspective of a particular individual, if we can determine the haplotype or phase of the marker and major gene, we will be able to predict the major gene allele inherited by the offspring on the basis of knowledge of the marker allele. For example, if a parent carries the M and the Q alleles on one chromosome, inheritance of the M allele will likely indicate inheritance of the Q allele, with the likelihood depending upon the recombination frequency or distance between the marker and major gene locus. Different haplotypes, such as Mq, mQ or MQ will be represented in other parents so marker-assisted selection cannot be effective unless the parental haplotype is first ascertained to some level of certainty. In circumstances where a marker and a nearby gene have been segregating in a population for many generations and were not subject to selection, it is likely they will be in or near linkage equilibrium. Linkage equilibrium is more likely the further the marker locus is located from the major gene locus.

A better circumstance for marker-assisted selection occurs when the marker and the gene are in linkage disequilibrium (LD-MAS). This is more likely to be the case if the marker and gene are close together, if the gene has been subject to selection and/or if the gene has been relatively recently “introduced” to the population by mutation or migration. For example, suppose a sire is introduced into a population bringing with him a new marker allele M and a new major gene allele Q, in phase so his haplotype is MQ. Suppose the sire and his offspring are widely used. After a few generations all three marker genotypes and all three major gene genotypes are likely to exist and each will be in Hardy-Weinberg equilibrium. However, if the marker and major gene are tightly linked (i.e. close together) there may be many generations before a crossover event occurs between the marker and major gene so all animals with the major gene Q allele will also have the M allele and vice versa. This allows the M allele to be used as a reliable proxy for the Q allele without having to determine the haplotype or phase of individual parents before utilizing MAS. However, at some stage, a recombination event will break the linkage and begin to erode the extent of disequilibrium.

The best circumstance for marker-assisted selection occurs when the causal polymorphism is known so that the gene allele itself can be used as the marker. This is sometimes known as gene-assisted selection (GAS).

There are some problems with making effective use of these three MAS categories (LE-MAS, LD-MAS, GAS) in practice. First, when a chromosome region (quantitative trait locus or QTL) is discovered and shown to influence a trait, we often do not know whether we are in a situation of linkage equilibrium or disequilibrium. As the process of positional cloning or fine-mapping is begun, a number of candidate genes may be identified. It is not a trivial matter to prove beyond reasonable doubt that a particular polymorphism is the causal effect underlying the observed phenotypic expression of the major gene. Thus, we may progress through the three stages at a somewhat unknown rate as further research is undertaken.

Regardless of the approach (LE-MAS, LD-MAS or GAS), there are some other critical animal breeding issues related to the application of MAS within the context of a livestock industry. First, there is the manner in which the genotype for the major gene effect is identified. This influences the ease and cost of determining genotype. Second, there is the issue of the relationship between the nature of the major gene effect, the breeding objective and the selection criteria. Among other factors, this can influence the extent to which the major gene is deliberately or inadvertently “weighted” in selection decisions in relation to polygenic effects (for the same or other traits in the objective). Third there is the issue of the tier or selection pathway to which the MAS is applied. These second and third factors dictate the number of years and the number of animals for which marker information will be required. These latter two factors will now be considered in more detail.

Major genes and the breeding objective

A well-defined breeding program begins with a clear definition of the breeding goal. Given the goal, a breeding objective can be defined, involving two parts: a list of traits that influence the goal; and the relative (usually economic) emphasis of each of the traits in the list. The collection of pedigree and performance data on animals of interest for selection and/or their relatives allows the calculation of expected progeny differences (EPDs). Such EPDs have proven to be useful tools in achieving selection advances. Ideally, there should be an EPD corresponding to each of the traits in the objective. However, there are often traits in the ideal objective for which EPDs are not available, perhaps because measurement of the trait is too expensive, is not currently technically possible, occurs too late in life for selection, or cannot be obtained without compromising the health of the animal.

Some examples of MAS apply to a characteristic (such as medullation, a fleece attribute or a polymorphism with functional food qualities) that completely changes the nature of some animal product, beyond the usual scope considered in a breeding objective. Other examples of MAS apply to traits that are in the “ideal” objective, but not part of the “working” objective for which EPDs are routinely available (e.g., some disease traits). Remaining MAS examples apply to traits for which EPDs are already routinely available. However, in this context it is worth distinguishing situations where the EPDs have high reliability (e.g., for lactation traits of dairy bulls with progeny test daughters) or low reliability (e.g., carcass or fertility traits on young live animals).

This issue of the “place” of the QTL in relation to the breeding objective tends to influence the partitioning of selection pressure applied to the major gene in relation to other polygenic effects that influence the breeding objective.

Industry structure and marker-assisted selection

Developed livestock industries are structured, typically comprising a sire-breeding sector and a sire-buying or so-called commercial sector. In the sheep industry, the rate of genetic improvement in the sire breeding sector is primarily determined by two

pathways of selection – the selection of replacement sires and the selection of replacement dams. MAS can be used for either (or both) of these pathways. In the dairy industry and to some extent in the beef cattle industry, genetic progress is determined by selection in four pathways. These are the choice of bull fathers, bull mothers, cow fathers and cow mothers. MAS can be applied in any one or more of these pathways.

There is a genetic lag between the sire-breeding and sire-buying sector that is determined by three factors. First, the rate of genetic gain in the sire-breeding sector. Second, the generation interval in the sire-buying sector. Third, the extent to which above-average sires can be purchased from the sire-breeding sector for use in the sire-buying sector. Ignoring this third factor, the genetic lag would be twice the genetic gain in the sire-breeding sector in product with the generation interval in the sire-buying sector. MAS could be used to reduce this genetic lag by identifying above-average sires, regardless of whether or not MAS is used to influence the rate of gain in the sire-breeding sector.

Some examples of marker-assisted selection

There are many examples of traits influenced by major genes that have been subjected to MAS in sheep and cattle improvement in New Zealand. Some of the examples will be common to other countries but others may not. Some superficial details regarding the nature and discovery of these genes have been included. The examples demonstrate the diverse nature of MAS applications with respect to marker knowledge (LE-MAS, LD-MAS and GAS), trait nature and industry structure.

The first markers used in the sheep and cattle industries

The sire-buying or commercial sector of most industries has the opportunity of utilizing a wide variety of management practices in order to convert natural resources such as sunshine, rainfall or feed into consumer products such as meat, milk or fiber. Animals vary markedly in the efficiency with which they can carry out such conversions. Some of these differences in efficiency have a genetic contribution while others are due to so-called non-genetic or environmental factors such as herd, year, date of birth, rearing rank, sex or age of dam. We typically remove these factors as fixed effects in the process of genetic evaluation. However, one of these factors, sex, is in fact nothing more than a genetic marker which has enormous impact on productivity. This genetic marker is typically of little interest in the sire-breeding sector but has been incredibly valuable in segregating animals with respect to future management practices. For example, castrate male sheep (known as wethers) have been segregated from females for fine-wool production. Male dairy calves have been left intact for specialized bull beef production whereas surplus females have been used for veal production. Meat production from sheep and beef cattle has exploited different finishing strategies to account for differences in mature weight, voluntary intake and composition of growth that exist between the sexes.

Farmers have long recognized the advantage of crossbred animals in certain production circumstances. The use of dam lines selected for maternal attributes and sire lines selected for carcass attributes can be crossed to produce animals with superior carcass attributes in comparison to animals of straight maternal lines. A problem with such a production strategy is in having to partition the reproductive capacity of the maternal lines in order to simultaneously produce maternal line replacements while maximizing the creation of crossbreds. In this circumstance, coat color attributes have long been exploited as a genetic marker. A classic example would be the use of black-face markings of Suffolk- or Southdown-sired lambs to distinguish these offspring from straight Romney lambs in a mixed breed cohort.

Sex and coat color are two easily overlooked examples of GAS applied in the commercial tier of the industry.

Favorable genes associated with functional attributes not directly in the objective

Drysdale. In the 1930's and 40's Dr. Dry demonstrated the existence of a dominant mutation in Romney sheep that caused wool to be medullated (hairy) and had a pleiotropic effect in creating horns (Dry, 1955). A subsequent dual-purpose breed was established, known as the Drysdale, producing wool with desirable attributes for carpet production. This breed was for many years represented in New Zealand by some 5-600,000 breeding ewes. Creating a true-breeding line of sheep from a dominant mutation can be problematic without GAS, but Dr Dry noticed a subtle difference in the midside of the lamb birth coat that distinguished hairy heterozygotes from homozygotes. This marker was used to rapidly introgress and fix the so-called N gene to create ram-breeding and commercial flocks of Drysdale sheep. The chromosomal region that contains the causal polymorphism has not been established yet this is an example of GAS, in which the selection concurrently occurred in sire-breeding and sire-buying sectors, with the objective of rapidly creating a new pure breed with novel wool characteristics for a specialized end use.

B variant β -lactoglobulin. The B-variant of the milk protein β -lactoglobulin is associated with a DNA polymorphism that results in decreased synthesis of this protein variant in milk, leading to a decrease in whey protein concentration and a 7-12% increase in casein concentration. This in turn results in an increase in cheese yield as cheese is composed primarily of casein and milk fat. Research trials involving milk manufactured from cows with alternative variants demonstrated a 2-3% increase in cheese yield from BB milk. There were no differences in cheese quality or taste. Tests have long been available for milk protein variants, originally based on testing for the presence of each variant in the milk protein and more recently from direct DNA tests. From 1995 onwards 28 suppliers (representing 8,500 cows) that were a small cheese manufacturing cooperative bred predominantly to homozygous BB β -lactoglobulin bulls. These Holstein-Friesian and Jersey BB bulls were obtained from among the highest index merit bulls available to industry each year. In 1998 the company changed its payment system to reward suppliers on the basis of casein rather than protein productivity with an adjustment for fat and a penalty for volume (Boland et al., 2000). The cooperative

merged in 2001 with larger companies to form the company Fonterra and its milk is now used for many products. Accordingly, its unique payment system and the BB selection program have been terminated. This was another example of GAS, with the selection occurring in a small segment of the bull to breed cow pathway. The industry rate of gain would be unaffected by the adoption of this single gene in this small sector of the industry, for as long as high merit BB bulls could be readily accessed.

A2 variant β casein. A recent focus on “functional foods” has led to the recognition that certain food components might be associated with favorable health benefits. Accordingly, some other food components may be associated with unfavorable effects. There are some studies that have found associations between the consumption of the A1 variant of β casein and the onset of insulin-dependent diabetes. Some subsequent studies have been unable to confirm this association and experiments to remove the A1 variant from the diet have not influenced the incidence of insulin dependent diabetes. Some epidemiological evidence exists to suggest there was a relationship between the consumption of milk (and therefore the A1 variant of β casein) and coronary heart disease but this relationship no longer exists. Hill et al., (2002) reviews some of these studies. The studies showing a benefit to A2 milk have led to the formation of a NZ company (A2 Corporation) which aims to market milk obtained from homozygous A2 variant cows, presumably with claims as to their health benefits with regard to diabetes and heart disease. The A1 and A2 variants are at intermediate frequency (varying with breed) so it is a relatively straightforward process to genotype cows and collect together herds of homozygous A2 variant animals. Five supermarket chains are (in 2003) trialing sales of this product which is currently available in twenty-two retail locations in New Zealand. A recent agreement has been reached between A2 Corporation and US Corporation IdeaSphere who intends to test some 100,000 cows in the US, leading to the retailing of A2 milk in 5,000 US health food outlets. This strategy demonstrates GAS primarily to segregate commercial cows to create pure A2 herds. At this stage it is not known how the Artificial Breeding (AB) companies will react, but it is possible that they will impose A2 selection on the bull to breed cow pathway prior to progeny testing and perhaps the bull to bull and bull to cow pathways if they view long-term advantage in converting the national herd to homozygous A2 status.

Unfavorable genes associated with functional attributes not directly in the objective

Mannosidosis. This is a lysosomal storage disease that, until a test was introduced, affected several thousand Angus cattle born in New Zealand each year in the 1960's and early 70's. Heterozygotes are normal but homozygous recessive individuals lack a functional copy of the α -mannosidase enzyme. Mannose cannot be metabolized by these individuals and accumulates in their cells, resulting in death, typically by about one year of age. Research at Massey University (Jolly et al., 1973) led to an enzyme assay that enabled carriers to be distinguished from homozygous “normals” on the basis of plasma enzyme concentrations. The Breed Association prohibited the registration of Angus bulls without undergoing testing. This led to a rapid decline in the incidence of the disease in New Zealand.

DUMPS (deficiency of uridine monophosphate synthetase). This recessive genetic disease in cattle results from insufficient activity of an enzyme that is involved in the synthesis of pyrimidine nucleotides, which, as constituents of DNA and RNA, are essential for normal growth and development. The disease results in embryonic death. The condition was first described by Robinson (1983). The first marker was physiologically-based by assaying UMP synthase activity from blood. The enzyme activity in heterozygotes was about half the level of homozygous normal individuals. A point mutation responsible for the condition was identified some 10 years later, leading to a direct DNA test for carriers (Schwenger et al., 1993). It was readily confirmed that there was a greater incidence of failed pregnancy when daughters of carrier bulls were mated to a carrier bull. The immediate response by AI companies was to try and eliminate the use of all carrier bulls as bull fathers and bull mothers. The discovery of DUMPS was quickly followed by the identification of some other recessive defects, primarily in dairy cattle.

Citrullinemia is a recessive genetic defect that leads to a buildup of citrulline and ammonia, due to a fault in the urea cycle. The disease was discovered in Australia (Harper et al., 1986). Afflicted animals are normal at birth, but show depression within a few hours. They usually die within a week of birth. A DNA test was soon developed to identify carrier animals.

BLAD (bovine leukocyte adhesion deficiency). This recessive genetic disease in cattle results from a polymorphism in a glycoprotein (MAC-1) responsible for transporting white blood cells from the blood stream into infected tissue (Schuster et al., 1992). The defective gene is present in Holstein cattle and results in the death of homozygous recessive calves, typically within a few days of birth. A direct test for the defective polymorphism enabled carrier bulls to be readily identified.

Given that every animal likely carries a few deleterious genes, the culling of carrier bulls is unlikely to be an effective solution. Unpublished simulations showed that there was net industry benefit from continuing to use existing proven carrier bulls with conditions such as BLAD and DUMPS if their aggregate genetic merit was sufficiently high, but semen allocation was managed to minimize the subsequent mating of carrier bulls to daughters of carrier bulls. Future carrier males were discarded prior to progeny testing in favor of non-carrier half-sibs. This strategy has led to a simultaneous improvement in aggregate merit and erosion of the frequency of carrier bulls.

CVM (complex vertebral malformations). This condition is the most recent genetic recessive disorder to be discovered (Agerholm et al., 2000, Wouda et al., 2000) and has been found in a number of Holstein populations including New Zealand. Affected calves have a misshapen backbone and may be aborted or born prematurely. Some are stillborn and a few born alive. Marker-based tests have been developed in Denmark and The Netherlands and are used to identify carriers. To date there is no published information on the gene/pathway involved, nor the causal polymorphism.

Spider syndrome. Hereditary chondrodysplasia was introduced into New Zealand when, in 1992, a Suffolk ram of US origin was imported from Australia. The disease is a recessive skeletal disorder represented by abnormally long bent limbs and curvature of the spine. The gene was shown to occur on chromosome 6 and the causal mutation was soon discovered (Beever et al., 1998). In 1993 the suspected heterozygous ram import was confirmed as a carrier by progeny test mating to his daughters. Thereafter rams could be progeny tested using daughters of the first carrier ram. Once a DNA test became available, most potential carriers were tested over the course of the next 2-3 years.

The above are all examples of GAS applied to distinguish carriers from homozygous normal individuals, albeit mannosidosis used an enzyme phenotype rather than a molecular assay to establish genotype. In general, the philosophy has shifted from trying to eliminate the defect as rapidly as possible to one of a controlled reduction in gene frequency at population level, in concert with the avoidance (where practical) of matings that will bring together heterozygote sires with daughters of heterozygote sires.

Genes associated with fecundity (and fertility) in sheep

Most New Zealand sheep are dual-purpose (with both wool and meat contributing significantly to income). Only a small fraction of the national flock is dedicated to wool production (e.g., super-fine merinos) or to meat production (e.g., terminal sire breeds). Increasing prolificacy is therefore an important trait in the breeding objective. It is historically assessed from counting the numbers of lambs born or reared. More recently ultrasonic diagnosis of litter size has been a useful selection criterion.

Booroola. The Booroola is a highly prolific strain of Merino sheep that has an average litter size more than twice that typically found in Merino ewes. It was shown simultaneously in Australia and NZ that this prolificacy often segregated in offspring of Booroola sheep (Piper et al., 1982, Davis et al., 1982). The effect of the so-called FecB gene is additive in respect to ovulation rate and litter size (adding about 1.5 ova or 0.7 offspring per favorable allele). Introgression into a variety of non-Merino breeds was achieved by identifying sire genotype from laparoscopic identification of small groups of daughters measured prior to one year of age. The gene was subsequently mapped to chromosome 6 (Montgomery et al., 1993) allowing its detection with flanking markers. This facilitated rapid introgression of the gene into new breeds. Introgression would have occurred rapidly enough to allow LD-MAS to identify homozygous animals. The causal polymorphism was recently identified (Wilson et al., 2001). The commercial use of the gene has been somewhat limited by the fact that litter size was often too extreme in homozygotes leading to a marked reduction in lamb survival. The gene likely has more practical application in commercial dam lines comprising heterozygous females. Continued creation of heterozygotes requires ongoing outcross of homozygous males or a cheap and rapid test to discriminate among potential replacement females on the basis of their FecB genotype.

Inverdale. Selection for ovulation rate in a screened prolific flock, identified in the mid 1980's a family of prolific Romneys descended from one ewe. Progeny tests of male descendants indicated a major gene situated on the X chromosome that increased ovulation rate by about 1.0 and litter size by about 0.6 (Davis et al., 1991). Subsequent studies showed that homozygous females have small non-functional 'streaky' ovaries and are infertile. The X-linked nature of the gene in combination with the infertility of homozygous females means that only half of the offspring of prolific females inherit the favorable allele. Sons carrying the gene pass it on to all their daughters and none of their sons. This mode of inheritance is ideal for producing Inverdale sons for use as sires of a terminal dam line. The first commercial exploitation of this system relied on the progeny testing of ram lambs to identify carrier sons. Six to eight ewe lamb offspring from Inverdale ewes mated to a ram lamb would be examined laparoscopically or at slaughter for the presence of a streaky ovary – which would indicate the ram carried the Inverdale gene. The causal polymorphism was subsequently identified (Galloway et al., 2002) leading to an exact DNA test.

The Inverdale gene has since been introgressed into breeds other than the Romney. The continued exploitation of this gene might best be achieved at the level of the multiplier – a tier of the sire-breeding sector sometimes introduced below the nucleus. This is because a loss of selection intensity would occur if the gene is introduced to the nucleus flock as a result of half the daughters being infertile when carrier rams are mated to carrier ewes. A multiplication tier can rapidly generate Inverdale sons by mating carrier ewes to elite (non-carrier) rams sourced from the nucleus. Half of the sons would carry the gene and if readily identified could be sold as maternal sires for outcrossing. The carrier daughters born in the multiplication tier would be retained as replacements in that tier.

Woodlands. The screened mixed breed prolific flock formed in the 1980's (that led to the discovery of the Inverdale gene) also identified an interesting family of Coopworths. This family produced male offspring carrying a putative gene that increased ovulation rate by 0.4 in daughters but produced non-carrier sons. It has since been shown that the gene is imprinted (Davis et al., 2001). That is, the so-called Woodlands gene is only expressed upon paternal inheritance from carrier males that were the progeny of non-expressing carrier females. The gene is not expressed when inherited from carrier females (expressing or non expressing) nor from carrier males that were the progeny of expressing carrier females. The causal mutation has not yet been identified.

These fertility genes are somewhat problematic to exploit because the favorable alleles have economic advantages yet fixation (i.e. homozygotes) are undesirable. Managing heterozygotes requires careful mating plans and ongoing gene testing. These require different (more complex) industry structures from those that have been historically employed. The prolificacy EPDs calculated from pedigree information are typically of low accuracy because fecundity is lowly heritable and sex-limited and the major genes are not subject to much selection pressure via EPDs. This creates a perfect situation for benefiting from GAS.

Genes associated with carcass attributes

Myostatin in cattle. The muscle-hypertrophy phenotype in certain cattle breeds such as the Belgian Blue has long been documented. This phenotype results in increased muscling particularly in the hind quarter of affected animals. Discovering the underlying gene involved in the phenotype was aided by mouse models, with it shown that mutations to the Growth-Differentiation-Factor-8 (GDF8) or myostatin gene can cause increased muscling (McPherron et al., 1997). Smith et al., (1997) went on to show that the myostatin gene maps to a region where a QTL for muscling had been mapped in cattle. Myostatin is a negative-regulator of muscle growth, and any mutations to the gene cause its regulatory ability to be impaired leading to increased muscle production. Subsequent sequencing of the gene in a number of cattle breeds has shown a variety of mutations to the gene. In the Belgian Blue, the gene is not functional leading to the extreme muscle hypertrophy seen in that breed (Kambadur et al., 1997). Other European breeds, such as the Piedmontese (Kambadur et al., 1997), Charolais and Maine-Anjou (Grobet et al., 1998), have also been shown to have mutations which impair rather than knock out the gene leading to milder increases in muscle hypertrophy. Work is being carried out to assess the frequency of the different myostatin mutants in various NZ beef populations, with the aim of allowing individual breeders to increase the frequency of desirable genes while minimizing associated problems of dystocia.

Carwell. An increased eye muscle dimension phenotype in a line of Australian Poll Dorsets on the Carwell Stud was first reported by Banks et al., (1997). Landcorp farms, a New Zealand government owned farming corporation imported semen from these animals, and bred for this phenotype in the resulting offspring. In sheep carrying the mutation, there is an 8% increase in longissimus weight. Early attempts to identify associated markers showed that variations at markers in a region of ovine Chr18 could explain the variations in phenotypes seen. This locus is referred to as the Rib Eye Muscling (REM) locus, with the Carwell allele associated with the improved phenotype (Jopson et al., 2001) residing at this locus. This locus is in a similar region to the Callipyge locus (responsible for increased muscling in a line of American Dorsets). Recent work has shown that the two loci are distinct (McClaren et al., 2003). Landcorp has continued to select for the genotype using flanking marker haplotypes (LD-MAS) developing the Landmark Carwell line of sheep which exhibit the phenotype.

Mu-calpain. Collaborative research between New Zealand and American research facilities (AgResearch and USDA) has shown that a mutation to the Calpain-I gene which maps to bovine chromosome 29 is associated with meat tenderness (Page et al., 2002). This mutation has been identified in two experiments comprising a Piedmontese-Angus cross and a Limousin-Jersey cross. Two single nucleotide polymorphisms were genotyped in this region (Page et al., 2002). One specifically results in a change in amino acid (from alanine to glycine) with the alanine variant having more tender meat relative to the glycine allele. This QTL explains over 30% of the residual variation in meat tenderness (Cullen et al., 2003). Work is being carried out

to assess the frequency of the variant SNP in NZ beef populations. Once the frequencies are determined, it is hoped that GAS will be used to improve tenderness.

Myostatin in sheep. The Texel breed of sheep is known for increased muscling relative to other breeds. Three groups around the world have attempted to identify the underlying genetic cause of this increased muscling, using myostatin as a candidate gene (the gene responsible for the double muscling phenotype in cattle). The first report by Marcq et al., (1998) showed no differences in the coding sequence of the myostatin gene between Texels and Romanov controls, although an F2 population provided evidence for a QTL segregating in the myostatin region. Further studies in New Zealand and Scotland (Broad et al., 2000; Walling et al., 2001) have shown associations between markers surrounding myostatin and ultrasound measurements of composition, although the evidence for a QTL was not conclusive. In follow up work Marcq et al., (2002) reported stronger evidence based on carcass rather than ultrasonic traits. A similar study involving dissection of the leg and subsequent meat quality analysis is near completion in New Zealand. It is hoped that a marker haplotype can be established on which LD-MAS can take place. As Texels were only introduced into New Zealand recently, the effect is likely to be in linkage disequilibrium, particularly given that all known heterozygotes carrying the favorable allele are closely related. Texels are routinely used as terminal sires as well as a component breed in newly created composite dam lines.

Genes associated with lactational performance

In collaboration with University of Liege, a grandsire model for QTL detection was applied to NZ and Dutch families in a project that began in 1994. This led to the identification and verification of six QTL for lactation traits. For each QTL a number of heterozygous sires were identified. These QTL are at various stages of positional cloning with what are believed to be causal polymorphisms having been identified for two of these (DGAT1 and GHR, below).

Investigations were concurrently undertaken into the manner in which QTL could be exploited in the context of the NZ dairy industry structure. Results of those simulation studies demonstrated that limiting the selection of bull fathers or bull mothers to animals with particular QTL alleles could significantly erode selection differentials for polygenic effects in those pathways (Spelman & Garrick, 1997). In the long-term this could sometimes result in the merit of the national herd being worse off as a result of GAS being applied for short term gains. Further work elucidated the nature of this long-term loss. It resulted from linkage disequilibrium forming between the QTL and polygenic effects. Such linkage disequilibrium can develop in the absence of physical linkage between polygenes and the chromosome region containing the QTL. In that circumstance, the gametes with the best polygenic effects tend to be those with the unfavorable QTL allele. The best overall gametes tend to be those carrying the favorable QTL allele and sourced from heterozygous parents rather than from animals homozygous for the favorable QTL allele (Garrick, 1997). Long-term loss could readily be avoided by using GAS that does not favor QQ animals but harnesses the Q-carrying

gamete from Qq parents. This is most effectively achieved by selection among full-sibs. This would require multiple ovulation and embryo transfer being applied to those families where one or other parent is heterozygous for the QTL.

A feature of the grandsire model used for QTL detection is that findings are limited to traits for which young bulls are typically progeny tested. Accordingly, the resulting QTL are for lactation traits. Existing progeny test schemes have been optimized for cost-effective genetic improvement of these particular traits. The progeny test involves 80-100 daughters per bull which gives a very reliable test regardless of the presence or absence of segregating QTL. Nevertheless, consider three bulls with identical EPD, two of which are homozygous for alternate forms of the QTL alleles (QQ or qq) and the other is heterozygous (Qq). In the case of the homozygous sires, genetic markers will not be particularly helpful other than confirming whether the bull is QQ or qq. In practice, this can likely be inferred from ancestry information. In contrast, the heterozygous sire will produce two offspring classes – those that inherit Q and those that inherit q. These offspring classes will differ in performance according to the size of the QTL. MAS can provide useful information to segregate offspring into one of these two classes, prior to the collection of any phenotypic information.

Risk is a particular concern when an AB company adopts a new breeding strategy. Market share is determined to a large extent by genetic merit and the consequences of a poor implementation strategy could be devastating. In the early stages of QTL validation there is a conflict between the desire to use the results early and the need to be sure that the suggestive QTL really does exist. Given the fact that the progeny test already does a good job of predicting EPD (accounting for both polygenic and QTL effects) the obvious place for applying MAS is in preselection of young bulls prior to the progeny test. Such young bulls are typically the offspring of a small number of bull fathers.

Livestock Improvement Corporation progeny tests bulls in dedicated progeny testing herds. Those bulls that are satisfactory in all respects and outperform the existing cow father team are selected for immediate service. The top few graduates are also used immediately as bull fathers. These bulls do not directly compete with the bull fathers from the previous year as those bulls already have a crop of young sons working their way through the progeny test system. Only these few bull fathers that had just graduated from their progeny test were subjected to marker-assisted selection.

The strategy adopted involved two considerations. First, was a particular bull father segregating a particular QTL (i.e. was the bull heterozygous)? Second, what was the phase or marker haplotype associated with the favorable allele? The answers to these questions were obtained from the analysis of the performance and marker genotypes of their progeny test daughters. Without convincing knowledge to the contrary, it was assumed that LE-MAS was the sensible approach. The approach assumed that marker genotypes alone provided no information as to whether or not a sire was heterozygous (and therefore segregating the QTL), nor the phase relationship between the marker alleles and the favorable QTL allele. This leads to a within-family approach known as

“bottom up” marker-assisted selection (Mackinnon and Georges 1998) that selected for QTL alleles among young bull full sibs prior to progeny testing. There was no selection on bull sires or bull dams on the basis of QTL genotype, minimizing the chance that the rate of short or long-term genetic gain could be less than what would have been achieved using conventional procedures. If a bull was incorrectly determined to be heterozygous, the selection among full-sib sons would incur costs but would not compromise genetic gain. If the bull was correctly identified as heterozygous, the worst problem that could occur was to incorrectly infer the marker phase and accordingly select the unfavorable QTL allele. The chance of so doing was minimized by determining a critical value for the test-statistic to contrast daughter haplotypes using prior knowledge of the likely size of the QTL effect.

Flanking markers were therefore used to implement LE-MAS on DGAT1 and HGR in 1998 and 1999.

DGAT1 (Diacyl glycerol acyl transferase). The most promising of the six QTL was on chromosome 14. Its allelic effect was principally additive (some 6 kg fat, -2 to -3 kg protein and -120 to -130 litres milk per lactation). The current milk payment system in NZ favors the allele that increases fat and decreases protein and volume. An increase in the relative price of protein compared to fat could reverse the favorable allele from an economic standpoint. Since its first use in MAS, the position of this QTL was fine-mapped within a few cM and the causal polymorphism subsequently identified (Grisart et al., 2002) from a candidate gene identified through comparative work with mice.

GHR (Growth hormone receptor). A number of studies had detected a QTL influencing milk production on chromosome 20. This QTL was segregating in some New Zealand families. The magnitude of the allelic effects are smaller than for DGAT1 and are different in Jerseys and Holstein-Friesians but the favorable allele increases fat and protein together and decreases volume – changes in accord with the NZ payment system. Recent work has uncovered a causative mutation (Blott et al., 2003). Interestingly, retrospective analysis has demonstrated that although about 10% Jersey and 30% Holstein-Friesian bulls entering progeny test were heterozygous for this QTL, only 3% Jerseys and 15% Holstein-Friesians that went on for widespread use were heterozygotes. No bulls homozygous for the unfavorable allele were selected for use after the progeny test. The QTL would be an effective way of prescreening prior to the progeny test if it weren't for the fact that the favorable allele has a gene frequency of about 0.9 in the young bull population.

After screening the bull fathers to test for heterozygosity, it turned out (in both years) that there were only two bulls segregating a QTL (one Jersey and one Holstein-Friesian). This limited the opportunity for MAS and reflected the impact of traditional selection approaches on increasing the frequency of favorable QTL influencing traits that are already included as selection criteria and have reliable EPDs.

The bull dams that were to be mated to the two segregating bulls were treated to encourage multiple ovulation prior to embryo transfer (MOET) by trans vaginal recovery

(TVR). It was envisaged that only those sons that had inherited the favorable QTL would enter the progeny test. In 1998 and 1999 some 24 and 28 donor cows were subjected to MOET. In 1998 only half of the donors produced multiple bull calves with 7 producing 3 bull calves. In 1999 only 8 cows produced multiple bull calves. These levels of success are inadequate for routine application of even a single QTL. Ideally, one would want to select full-sibs that have inherited a favorable combination of a number of segregating alleles. This will require a quantum improvement in reproductive performance.

More recently, the discovery of the causal polymorphisms for these genes allowed GAS to be adopted with bull mothers being routinely genotyped for these effects.

Genes associated with disease

Footrot in sheep. Infections by *Dichelobacter* (Bacterioides) *nododus* bacterium act with other bacterial agents to infect and destroy foot tissue between the living tissue and the outer hoof horn, causing separation of the horn. Infected animals have suppressed feed intakes leading to reduced liveweight, fleeceweight and reproductive performance. The characteristic is moderately heritable but most farmers use management techniques to keep infection levels as low as possible. Ram breeders are unlikely to want to compromise the performance of sale rams by challenging them with footrot. A progeny test study involving a number of rams compared their average progeny performance with alleles at the DQA1 & DQA2 loci within the major histocompatibility complex (MHC). A test is now offered to ram breeders based on that study assuming the results are applicable for LD-MAS. However, as most users of the test do not allow the disease to develop they are unlikely to be able to verify the adequacy of the test. Previous Australian research had identified the MHC as a candidate region for footrot resistance but further research in that country dismissed the region for MAS.

Facial Eczema (FE). Facial eczema is a hepatogenous photosensitization disease of ruminant animals, particularly sheep. Sporidesmin is the mycotoxin produced by the fungus *Pithomyces chartarum* that is responsible for the liver damage. The fungal toxicity varies widely as does susceptibility to the toxin across lines of sheep. A number of breeders in New Zealand have made progress in creating lines resistant to the toxin by selecting animals on the basis of plasma concentration of the liver enzyme gamma-glutamyl transferase (GGT) following natural or artificial exposure to sporidesmin. GGT is a liver enzyme that leaks into the blood in greater concentrations when the liver is damaged. Catalase was chosen as a candidate gene to explain differences in susceptibility, as it is thought that the toxicity of sporidesmin is due to its ability to generate 'active oxygen' species, and catalase is an enzyme with antioxidant functions (Phua et al., 1999). A QTL search using markers surrounding this gene failed to find any association. However, differences in allele frequency for two markers were shown between experimental selection lines of susceptible and resistant sheep. A genome wide scan is being conducted in an attempt to find QTL associated with FE. No results have been published to date but a collaborating breeder has indicated that sale rams

will be available early in 2004 that have been identified on the basis of genetic markers. It is not known whether these markers reflect an application of LE-MAS or LD-MAS.

Internal Parasites. Internal parasites (such as *Haemonchus*, *Trichostrongylus* and *Nematodirus* sp.) suppress the growth rate of lambs and lead to pasture contamination with faecal eggs. The less resistant the sheep, the greater the faecal egg output and the greater the pasture contamination. Higher intakes of faecal eggs are associated with greater reductions in performance. Lambs can be selected for resistance based on faecal egg counts (FEC). Crosses between lines of sheep divergently selected for resistance to internal parasites (i.e. FEC) have identified three QTL that are currently being used in industry validation studies (J.C. McEwan, personal communication). One of these QTL is gamma interferon. Any confirmed QTL are targeted for industry use in early 2004.

Disease genes are desirable targets for MAS because breeders do not want to compromise the performance of their animals by directly challenging them with the disease, both for ethical reasons and because it may impact animal saleability. This provides challenges for researchers as routine field measures of phenotypic performance are less readily available than is the case with other measures of production and reproduction.

Summary

There is a large portfolio of QTL known to exist, although much of this knowledge is yet to be communicated as researchers try first to discover and patent the causal polymorphisms. Where QTL knowledge has been published, applications of MAS are rapidly progressing from LE-MAS to LD-MAS and GAS in line with scientific discovery of genes. Most causal polymorphisms have been found through the application of comparative genomics to published findings in other species such as mice and humans. Applications of MAS fall broadly into three categories. These include: gene elimination whereby an unfavorable (typically deleterious) allele is selected against; introgression where a favorable allele is introduced and selection applied to increase its gene frequency; or an "index" type approach where selection is applied simultaneously to one or more QTL and to polygenic effects influencing the same or other traits that contribute to the breeding objective.

Despite the number of QTL known, for most livestock production circumstances there are relatively few major genes that have reached the commercialization stage. Accordingly, few MAS applications need to simultaneously consider the selection of animals segregating multiple QTL. The discovery and commercialization of new genes will lead to opportunities for simultaneously exploiting many QTL. Optimizing the use of these QTL in nucleus schemes will require better success in routine application of reproductive technologies than can be achieved at present.

A reduction in the cost of genotyping will facilitate the use of segregation technologies to separate commercial animals in order to apply different management strategies that are

in concert with their genetic predisposition. This provides the opportunity to exploit major genes without necessarily imposing selection to change gene frequencies in the population.

Bibliography

- Agerholm, J.S., C. Bendixen, O. Andersen and J. Arnbjerg. 2000. Report no 564. National committee on Danish Cattle Husbandry, Aarhus, Denmark.
- Banks, R. 1997: The meat elite project: establishment and achievements of an elite meat sheep nucleus. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 12: 598-601.
- Beever, J.E., T.L. Shay, J. Albretson, A. Maciulis, T.D. Bunch, G.R. Holyoak and N.E. Cockett. 1996. Identification of the causative mutation in ovine hereditary chondrodysplasia [abstract]. Plant & Animal Genome VI: abstract W21.
- Blott, S., J.J. Kim, S. Moiso, A. Schmidt-Kuntzel, A. Cornet, P. Berzi, N. Cambisano, C. Ford, B. Grisart, D. Johnson, L. Karim, P. Simon, R. Snell, R. Spelman, J. Wong, J. Vilkki, M. Georges, F. Farnir and W. Coppieters. 2003. Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genetics 163:253-266.
- Boland, M., J. Hill and P. O'Connor. 2000. Changing the milk supply to increase cheese yield: the Kaikoura experience. In: Milk Composition, British Society of Animal Science Occasional Publication 25:313-316.
- Broad, T. E., B.C. Glass, G.J. Greer, T.M. Robertson, W.E. Bain, E.A. Lord and J.C. McEwan. 2000: Search for a locus near to myostatin that increases muscling in Texel sheep in New Zealand. New Zealand Society of Animal Production 60: 110-112.
- Cullen, N. G., B.T. Page, C.A. Morris, S.M. Hickey and J.L. Dobbie. 2003: A single-nucleotide polymorphism on Calpain-1 is associated with meat tenderness in cattle. New Zealand Society of Animal Production 63: 53-56.
- de Koning, D.J., J.C.M. Dekkers and C.S. Haley. 2003. Designs for QTL detection in livestock and their implications for MAS. Proceedings of an Electronic forum on biotechnology in Food and Agriculture. <http://www.fao.org/biotech/Torino.htm>
- Davis, G.H., J.C. McEwan, P.F. Fennessy, K.G. Dodds and P.A. Farquhar. 1991. Evidence for the presence of a major gene influencing ovulation rate on the X-chromosome of sheep. Biology of Reproduction 44:620-624.

- Davis, G.H. K.G. Dodds, R. Wheeler and N.P. Jay. 2001. Evidence that an imprinted gene on the X chromosome increases ovulation rate in sheep. *Biology of Reproduction* 64:216-221.
- Dry, F.W. 1955. The dominant N-gene in New Zealand Romney sheep. *Australian Journal of Agricultural Research*, 6:725-769.
- Galloway, S.M., K.P. McNatty, L.M. Cambridge, M.P.E. Laitinen, J.L. Juengel, T.S. Jokiranta, R.J. McLaren, K. Luiro, K.G. Dodds, G.W. Montgomery, A.E. Beattie, G.H. Davis, O. Ritvos. 2000. Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nature Genetics* 25:279-283.
- Garrick, D.J. 1997. Improving a trait influenced by a major gene (QTL) and many genes with small effects. *Proceedings of the Association for the Advancement of Animal Breeding & Genetics* 12:202-207.
- Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research* 12:222-231.
- Grobet, L., D. Poncelet, L. Jose Royo, B. Brouwers, D. Pirottin, C. Michaux, F. Menissier, M. Zanotti, S. Dunner and M. Georges. 1998. Molecular definition of an allelic series of mutations disrupting the myostatin function and causing double-muscling in cattle. *Mammalian Genome* 9: 210-213.
- Harper, P.A.W., P.J. Healy, J.A. Dennis, J.J. O'Brien and D.G. Hayward. 1986. Citrullinemia as a cause of neurological disease in neonatal Friesian calves. *Australian Veterinary Journal* 63:378-380.
- Hill, J.P., R.A. Crawford and M.J. Boland. 2002. Milk and consumer health: a review of the evidence for a relationship between the consumption of beta-casein A1 with heart disease and insulin-dependent diabetes mellitus. *Proceedings of the New Zealand Society of Animal Production* 62:111-114.
- Jopson, N. B., G.B. Nicoll, J.M. Stevenson-Barry, S. Duncan, G.J. Greer, W.E. Bain, E.M. Gerrard, B.C. Glass, T.E. Broad and J.C. McEwan. 2001. Mode of inheritance and effects on meat quality of the rib-eye muscling (REM) QTL in sheep. *Proceedings of the Association for the Advancement of Animal Breeding and Genetics* 14: 111-114.
- Kambadur, R., M. Sharma, T.P.L. Smith and J.J. Bass. 1997: Mutations in myostatin (GDF8) in double-muscléd Belgian Blue and Piedmontese cattle. *Genome Research* 7: 910-815.

- Mackinnon, M.J. and Georges. M. 1998. Marker-assisted preselection of young dairy sires prior to progeny testing. *Livestock Production Science* 54:229-250.
- Marcq, F., J.-M. Elsen, S. El Barkouki, J. Bouix, F. Eychenne, L. Grobet, L. Karim, E. Laville, C. Nezer, L. Royo, T. Sayd, B. Bibe, P.L. Le Roy and M. Georges. 1998. Investigating the role of myostatin in the determinism of double muscling. *Animal Genetics* 29: 52.
- Marcq, F., C. Larzul, V. Marot, J. Bouix, F. Eychenne, E. Laville, B. Bibe, P.L. Leroy, M. Georges and J.-M. Elsen. 2002. Preliminary results of a whole-genome scan targeting QTL for carcass traits in a Texel x Romanov intercross. 7th World Congress on Genetics Applied to Livestock Production: 2 - 14.
- McClaren, R., J.C. McEwan, R. For, B.C. Glass, T.E. Broad, G.J. Greer and G.B. Nicoll. 2003. Recombination breakpoint mapping of the Carwell locus for rib-eye muscling in sheep. *Proceedings of the 11th International Congress of Genetics* 11: 103.
- McPherron, A. C., A.M. Lawler and S.-J. Lee. 1997. Regulation of skeletal muscle mass in mice by a new TGF- β superfamily member. *Nature* 387: 83-90.
- Montgomery, G.W., A.M. Crawford, J.M. Penty, K.G. Dodds, A.J. Ede, H.M. Henry, C.A. Pierson, E.A. Lord, S.M. Galloway, A.E. Schmack, J.A. Sise, P.A. Swarbick, V. Hanrahan, F.C. Buchanan, D.F. Hill. 1993. The ovine booroola fecundity gene (FecB) is linked to markers from a region of human chromosome 4q. *Nature Genetics* 4:410-414.
- Page, B. T., E. Casas, M.P. Heaton, N.G. Cullen, D.L. Hyndman, C.A. Morris, A.M. Crawford, T.L. Wheeler, M. Koohmaraie, J.W. Keele and T.P. Smith. 2002. Evaluation of single-nucleotide polymorphisms in CAPN1 for association with meat tenderness in cattle. *J Anim Sci* 80: 3077-3085.
- Phua, S. H., K.G. Dodds, C.A. Morris, K. A. Paterson, J.C. McEwan, H.G. Garmonsway, N.R. Towers and A.M. Crawford. 1999. Catalase gene is associated with facial eczema disease resistance in sheep. *Anim Genet* 30: 286-295.
- Robinson, J.L., M.R. Drabik, D.B. Dombrowski and J.H. Clark. 1993. Consequences of UMP synthase deficiency in cattle. *Proceedings of the National Academy of Sciences (USA)* 80:321-323.
- Schuster, D.E., M.E. Kehrli, M.R. Ackermann and R.O. Gilbert. 1992. Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. *Proceedings of the National Academy of Sciences (USA)* 89:9225-9229.

- Schwenger, B, S. Schober and D. Simon. 1993. DUMPS cattle carry a point mutation in the uridine monophosphate synthase gene. *Genomics* 16:241-244.
- Smith, T. P. L., N.L. Lopez-Corrales, S.M. Kappes and T.S. Sonstegard. 1997. Myostatin maps to the interval containing the bovine mh locus. *Mammalian Genome* 8: 742-744.
- Spelman, R., and D.J. Garrick. 1997. Utilization of marker assisted selection in a commercial dairy cow population. *Livestock Production Science* 47:139-147.
- Walling, G. A., P.M. Visscher, G. Simm and S.C. Bishop. 2001. Confirmed linkage for QTLs affecting muscling in Texel sheep on chromosomes 2 and 18. 52nd Annual Meeting of the European Association for Animal Production 52: G5.6.
- Wilson, T. X.Y. Wu, J.L. Juengel, I.K. Ross, J.M. Lumsden, E.A. Lord, K.G. Dodds, G.A. Walling, J.C. McEwan, A.R. O'Connell, K.P. McNatty and G.W. Montgomery. 2001. Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of bone morphogenetic protein IB receptor (ALK-6) that is expressed in both oocytes and granulosa cells. *Biology of Reproduction* 64:1225-1235.
- Wouda, W., I.J. Visser, G.H. Borst, J.H. Vos, A.A. Zeeuwen and N.H. Peperkamp. 2000. Developmental abnormalities in aborted and stillborn calves in The Netherlands. *Veterinary Record* 147:612.

Commercial gene testing for major genes discussed in this paper are offered in New Zealand by

- Equine Blood Typing and Research Centre (http://ivabs.massey.ac.nz/centres/centre_blood.asp) – a centre linked to the Institute of Veterinary, Animal & Biomedical Sciences at Massey University. Primarily equine tests but also spider syndrome and parrot sexing.
- Genemark(http://www.lic.co.nz/main.cfm?menuid=3&sub_menuid=45) – a subsidiary of Livestock Improvement Corporation. Exclusively bovine testing.
- GenomNZ (<http://www.agresearch.co.nz/genomnz/default.htm>) – a subsidiary of AgResearch. Primarily sheep and deer DNA tests.
- Lincoln University (<http://www.lincoln.ac.nz/afs/research/anires.htm#footrot>). Footrot testing using MHC markers.
- Signagen (www.signagen.co.nz) – a subsidiary the Forest Research Institute. Bovine and ovine parentage and gene tests.

VALIDATING GENETIC TESTS FOR QUANTITATIVE TRAITS

Richard L. Quaas
Morrison Hall, Cornell University,
Ithaca, NY 14853 USA

After years of research and development, the first DNA-based genetic tests for quantitative traits in beef cattle have reached the U.S. market. No longer are we in the “what if” stage of thinking about the use genetic markers. The first few are here, and producers ask what the benefits are beyond printing a favorable genotype in a sale catalog. Concurrent with the commercial advent of genetic tests, and here I will consider only those affecting quantitative traits, was the formation of the National Beef Cattle Evaluation Consortium (NBCEC). This is a federally-funded consortium of universities that carry out routine national cattle evaluations (NCE), i.e., compute EPD. Among the consortium’s several functions, there are several project teams, one of which is the “QTL team.”

The QTL team (Quaas, Cornell; Thallman, Meat Animal Research Center; Fernando, Iowa State) has two objectives:

1. to develop protocol for validation (or not) of gene marker or test results with independent (field) data, and
2. to develop and test methods for incorporating markers into national cattle evaluation.

At this time no protocol has been written and objective 2 has been discussed only in general. In spite of that, however, we have examined the effects of several DNA markers on phenotypes from independent field data. Thus the objectives of this paper are not to present formal NBCEC results but share my experiences with the kinds of data – genotypic and phenotypic – and *a priori* information that might be available for validation and (or) incorporation into NCE. Emphasis will be on the validation objective. I will present only a few results emphasizing instead how the kinds of situations encountered have altered my notion of “validation.” Though Thallman and Fernando have offered advice, the responsibility for any of the analyses done thus far is mine.

QTL Tests – available, near or maybe

- Thyroglobulin (TG5)
 - Genetic Solutions
 - GeneStar Marbling
- Calpastatin
 - Genetic Solutions
 - GeneSTAR Tenderness
- Leptin
 - Merial – Quantum Genetics
 - ingenity L

- mu-Calpain (also called Calpain 1)
 - Frontier Beef
 - TenderGENE
 - codon316 (exon 9 SNP) + codon530 (exon 14 SNP)
 - Genetic Solutions
 - GeneSTAR 2
 - Calpain exon 9 SNP + calpastatin
- DGAT1
 - K232A dinucleotide substitution
 - Marbling ??
- Carcass Merit Project microsatellite markers
 - MMI-NBCA-???
 - Tenderness

All these markers are for carcass trait QTL. We have analyzed five (carcass) genotypes for GeneSTAR Marbling, Leptin, mu-Calpain (SNP316 & SNP 530) and DGAT1 for effects on either marbling or Warner-Bratzler shear force (WBSF).

Data Sets

NBCEC Taurus data. The source has been carcass phenotypes including WBSF and genotypes on Simmental-sired cattle (mostly commercial Angus dams). (We have data on progeny of a few Angus and Red Angus sires, but these have not been included in any analysis.) For some markers, there are genotypes on sires. The data came primarily from the Carcass Merit Project (CMP) with additional data from the American Simmental Association's progeny testing program. The data came from ASA because that is the only association that collects substantial numbers of WBSF data in addition to the CMP.

NBCEC Indicus-influenced data. Thus far the only Indicus-influenced data analyzed have been a set of 330 King Ranch Santa Gertrudis cattle. Carcass phenotypes included WBSF. These were genotyped by both Frontier Beef (GeneSeek) and Genetic Solutions for the MARC mu-Calpain single nucleotide polymorphisms (SNP). In the near future, ~300 CMP Simbrah-sired carcasses will be genotyped for calpain.

Analyses

Mixed Model. $y = \text{Genotype} + \text{CG} + \text{Sire}(\text{random}) + e$, where 'Genotype' includes category for ungenotyped contemporaries. Fitting the model included computing REML estimates of sire and residual components of variance. A fixed model, without a sire effect, was also fit.

Generally there were no qualitative differences between genotype effects whether or not sire was included in the model. Emphasis was placed on comparing estimates with *a priori* information about the putative effects. This fit with my original (admittedly limited) view of validation in the following scenario:

1. Genomic company ready to market DNA test with data on test "published" in some form.

2. Validation was to check company's results/claims with independent field data
 - a. Genotype effects
 - b. Mode of action

As will be seen, this was not the situation for most of the tests/markers examined, but it was pretty much the case for the first one examined – GeneSTAR Marbling.

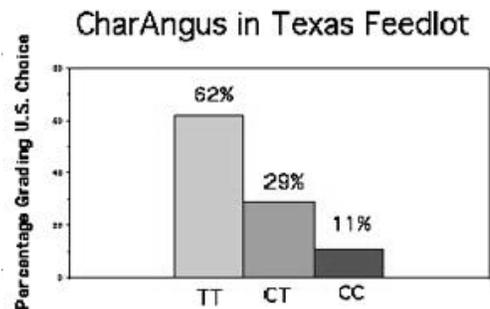
GeneSTAR Marbling. Genetic Solutions had results from several studies in the U.S. and Australia. Genetic Solutions provided genotypes on 273 Simmental-sired progeny from the CMP. The Genetic Solutions trial chosen to be most similar to these cattle were a group of U.S. calf-fed cattle. In this case the estimated difference between homozygotes was very similar. The mode of action was less clear.

Genotype	Genetic Solutions (calf-fed)	NBCEC (SimAngus)
2 star	370 ¹	527 ²
1 star	369	517
0 star	358	517
2 vs. 0 star	12	11

¹ Small⁰⁰ = 400 ² Small⁰⁰ = 500

mu-Calpain. In this case the markers examined were two of the SNP discovered at MARC in the mu-Calpain gene and studied in detail in progeny of two F1 bulls, one at MARC and one in New Zealand. Results for SimAngus carcasses were compared to unpublished MARC results from Germ Plasm Evaluation-cycle 7 cattle (seven sire breeds). Genotypes were provided on SimAngus cattle (assayed at GeneSeek). In this case there was no commercial test at the time of the ‘validation.’ Subsequently Frontier Beef and Genetic Solutions began to market a Calpain test based on two or one of the SNP, respectively.

Leptin. The marker was a “cytosine (C) to thymine (T) transition that encoded an amino acid change ... (in) exon 2 of the leptin gene.” (Buchanan, *et al.* 2002; GSE 34:105). Genotypes of SimmAngus cattle were assayed by GeneSeek. In this case there were few ‘hard’ estimates of the effects of the alternate genotypes in the literature or even which traits in beef cattle would be affected. Information could be found on the web such as in the figure (<http://sask.usask.ca/~schmutz/meat.html>). Shortly after the genotyping was done, the following appeared:



Press Release Merial and Quantum Genetics, Inc. Sign Global Pact July 23. SASKATOON, Sask., Canada; DULUTH, Ga., ... a Global Marketing Agreement ... provides Merial with exclusive rights to market Quantum's new patent-pending **DNA test to determine an animal's leptin genotype** {emphasis added}, as well as the

*application of this knowledge. The leptin protein is ... a major determinant of **feed intake and energy balance** {emphasis added} in beef and dairy cattle. The test helps to determine an animal's genetic propensity to "**marble**", ... helps to determine an animal's genetic propensity for **increased milk production** {emphasis added}.*

Marbling score was chosen as the trait of interest.

DGAT1. “The ApA (Adenine p Adenine) to GpC (Guanine p Cytosine) dinucleotide substitution in exon VIII (which causes a lysine to alanine amino acid substitution) of DGAT1” (Spellman, et al., J. Dairy Sci. 85:3514) was assayed by GeneSeek on SimAngus carcasses. The effects of this mutation are well documented for dairy cattle production traits but much less so in beef cattle. Thaller *et al.* (2003; Animal Genetics 34:354) reported the polymorphism has a significant effect on intramuscular fat %. The lysine polymorphism increases IMF in the semitendinosus based on 28 German Holsteins and 27 Charolais.

All the previous tests (except for DGAT1, which is a dinucleotide substitution) are SNP in or close to a gene; they are probably not causal mutations, i.e., in Thallman’s nomenclature, are “association markers.” Another type which in some sense has been validated are the CMP microsatellite markers (see Thallman’s contribution to these proceedings.)

Discussion

The two major differences among the various markers examined were the stage of commercialization and the amount of *a priori* information publicly available about the magnitude of allele effects. The latter is critical for my original view of validation – “*the act of finding or testing the truth of something*” (dictionary.com). If (semi-) hard estimates are available either from a genomics company’s trials or from a research organization, then it is feasible to check these estimates with independent field data. This was the situation for TG5 (Genetic Solutions trials) and Calpain (MARC GPE7 results). Sometimes the *a priori* information is less quantitative, e.g., report of significant effects, qualitative statements of effects, or possible effects inferred from a different type of animal. In these cases, at most we can follow-up such indications.

Real Validation. A real validation needs, first and foremost, statistical power, a design which would allow a valid validation. This requires large numbers of animals. Ideally it would include multiple breeds with enough animals in each to estimate allele frequencies and to test for differences among breeds in genotype effects. We would have data on most important traits to check for unexpected adverse epistatic effects. We would be able to determine the mode of inheritance clearly. Ideally we would assay duplicate samples for the repeatability of test. The preceding list is not exhaustive but suggests the resources that would be required.

Real World. In reality there are few resources available for validation. In our attempts, phenotypes and DNA were provided by breed associations or breeders, and data

sources were limited because much emphasis was on tenderness requiring the need for WBSF. (This is not likely to be unique to tenderness. Common wisdom is that the traits most benefiting from marker- assisted selection are those for which data are difficult to come by.) Genotyping was provided by genomics companies. Only ~300-400 animals with phenotypes were genotyped, and not many breeds were represented. The studies did not have sufficient statistical power to accomplish what we would like. The numbers of animals do not indicate the modest power of the sample, which is greatly exacerbated by unequal allele frequencies. DGAT1 was the worst: 95% of the alleles in the sample were of the alanine (q) variant – 248 qq, 24 Qq and 1 QQ – no genotype differences came close to significance.

The relatively small sample is less critical when trying to replicate quantitative results. The null hypothesis is not that there are no effects but that the effects published, claimed, etc., are reasonable. It is critical when there is virtually no quantitative information on magnitude of effects and for many traits there is not even a qualitative assessment. Thallman (pers. comm.) suggests testing with very stringent criterion; my concern is that any 'positive' result will appear in an advertisement.

Conclusion

Validation studies of the strict FDA life form are unlikely. Most definitely the NBCEC cannot validate – “*To mark with an indication of official sanction.*” (dictionary.com) – a genetic test. It is possible, however, to provide the beef industry with useful independent information concerning genetic markers. We just need a less loaded word than “validation” to describe what can be provided.

Appendix

Following this Workshop, an expanded QTL Committee of the NBCEC spent an afternoon discussing a 'protocol.' The following are some of the points raised.

- Test in 2 of 3 types of populations (sire pedigreed), sired by:
 - British, Continental or Indicus-influenced sires
 - each at least 300 head, steers or heifers
 - crossbreds allowed
- Populations with phenotypes and stored biological material will be identified by NBCEC
 - breed association, breeders, experiment stations, etc.
 - phenotypes will be maintained in an NBCEC database
 - phenotypes will not be provided to genotyping company
- Genotyping company
 - will have (some) choice as to reference family
 - will provide genotypes at no cost
 - sire genotypes
 - progeny genotypes

- return genotypes to NBCEC QTL committee (or other designated group) for analysis
- Analysis (minimum)
 - $y = CG + \text{genotype} + \text{sire}(\text{random}) + e$
 - NBCEC will report results NOT put a stamp of approval on a test
- Information dissemination
 - results distributed to
 - Genotyping company
 - owner of reference family
 - results published
 - on NBCEC web site
 - scientific journal or scientific/educational meetings

GENESTAR MARKERS – DELIVERING PRODUCTIVITY IMPROVEMENT TO THE BEEF INDUSTRY

Dr. Jay Hetzel

**Genetic Solutions Pty. Ltd., PO Box 145 Albion Brisbane, Qld Australia 4010
Jay.Hetzel@geneticsolutions.com.au**

Introduction

Despite impressive research achievements over the past 15 years, only a small number of DNA markers have found application in beef industry breeding and management programs. The majority of these markers have been for simple genetic traits such as genetic disorders (e.g. Pompe's disease), coat color and double muscling. However most economically important traits are complex in nature, being controlled by multiple genes that interact with each other as well as with environmental factors. Success in identifying robust markers for the major production traits has been limited. But there have been some successes. The current GeneSTAR[®] markers for marbling and tenderness are the market leaders, both in terms of industry uptake and demonstrated value. There is also a pipeline of new GeneSTAR markers for both meat quality and productivity traits which bodes well for the future.

What are we looking for?

DNA markers can increase the accuracy of selection, especially at young ages, thereby reducing generation interval. The end result is to increase the rate of genetic gain. The maximum impact of DNA markers will be on traits which are currently difficult to breed for, either because phenotypic measurements are inaccurate, expensive, only possible on one sex or cannot be carried out at a young age. Most carcass and meat quality traits fall into this category; but so do feed efficiency, female reproduction and disease susceptibility traits. For complex traits controlled by many genes, multiple markers will increase accuracy. Clearly, markers for traits in the breeding objective as distinct from selection traits will have greatest economic value. For example, markers for Quality Grade are more valuable than markers that only affect marble score or Intra muscular fat % (IMF).

How do we evaluate new markers?

We believe it is important to fully evaluate new markers in order to determine economic value as well as under what circumstances the markers have value. Some markers may only have utility in certain breed types or specific production systems. Accordingly, Genetic Solutions is committed to an evaluation process which determines:

- the consistency and repeatability of effects in key breed types and production systems;
- effects in commercially relevant groups;
- correlated effect on other production traits;

- interactions between different markers for the same trait;
- genotype profiles for the key breed types.

Although comprehensive evaluation adds to the cost and delay in developing and delivering new markers, we believe it serves the best interests of the beef industry. Put simply, the information provides greater predictability of the value of individual markers.

Current GeneSTAR Markers

GeneSTAR Marbling

The GeneSTAR[®] marbling test is based on a marker in the Thyroglobulin gene and has been available for over three years (see www.geneticsolutions.com.au). Originally discovered by Dr Bill Barendse from CSIRO, Australia, it has now been evaluated on more than 3500 cattle covering a range of breed types and feeding systems. Frequency of the high marbling allele is highest in the Wagyu breed (Fig 1), intermediate in *Bos Taurus* and lowest in *Bos indicus*. Effects on marble score in lot fed cattle have shown a difference between alternative homozygotes of 3.5 to 11%. In long fed Wagyu cattle, the effects have ranged from 14 to 20%. Differences in Quality Grade have been remarkably consistent, ranging from 16-19% more USDA Choice (rather than Select) grade. The numbers of premium marbling score carcasses is doubled. Finally, no statistically significant correlated effects on other carcass traits, either in a positive or negative direction have been detected.

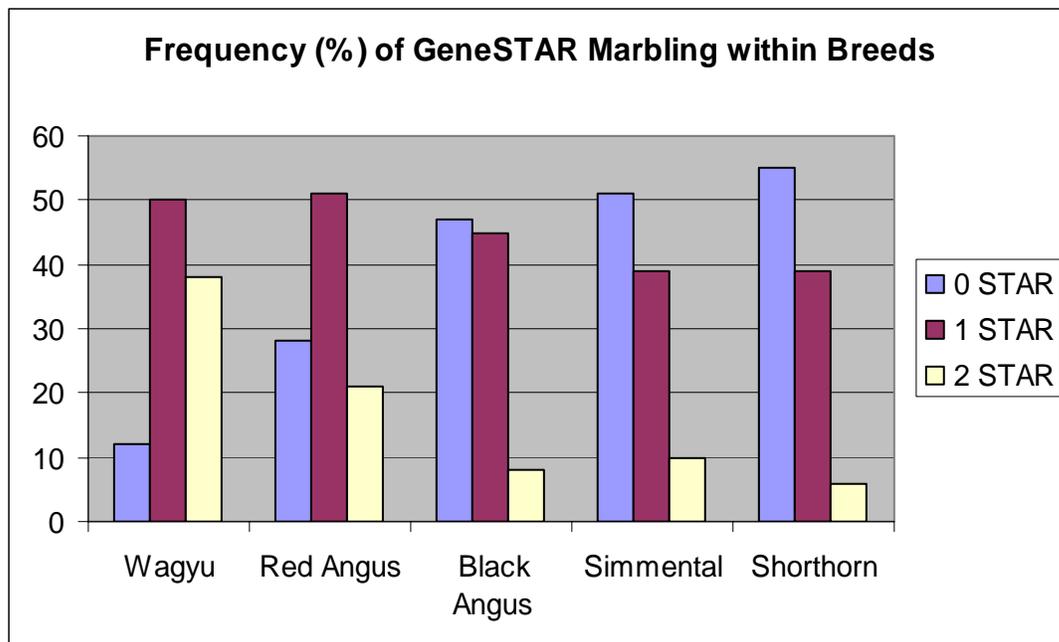


Figure 1. Genotype frequencies for GeneSTAR marbling in different breeds. (2 STAR signifies homozygous for the high marbling allele, zero STAR is homozygous for the low marbling allele and one STAR is heterozygous).

An independent evaluation of GeneSTAR marbling by the USA National Beef Cattle Evaluation Consortium (NBCEC) has confirmed the effects on marbling and Quality Grade in Simmental cattle.

GeneSTAR Tenderness 2

The first DNA marker test for tenderness was released in November 2002. It was based on research carried out by CSIRO and the Cattle and Beef Quality Cooperative Research Centre in Australia, which detected a marker in the bovine calpastatin gene. In a trial on over 5000 cattle from seven breed types, the marker was associated with a difference of 0.8 lb on the Warner-Bratzler shear force scale (see www.geneticsolutions.com.au). This difference is predicted to reduce the proportion of unacceptably tough carcasses, as rated by consumers, from 21 to 8%. The tender allele is at highest frequency in *Bos taurus* cattle breeds and lowest in *Bos indicus* breeds (Fig 2).

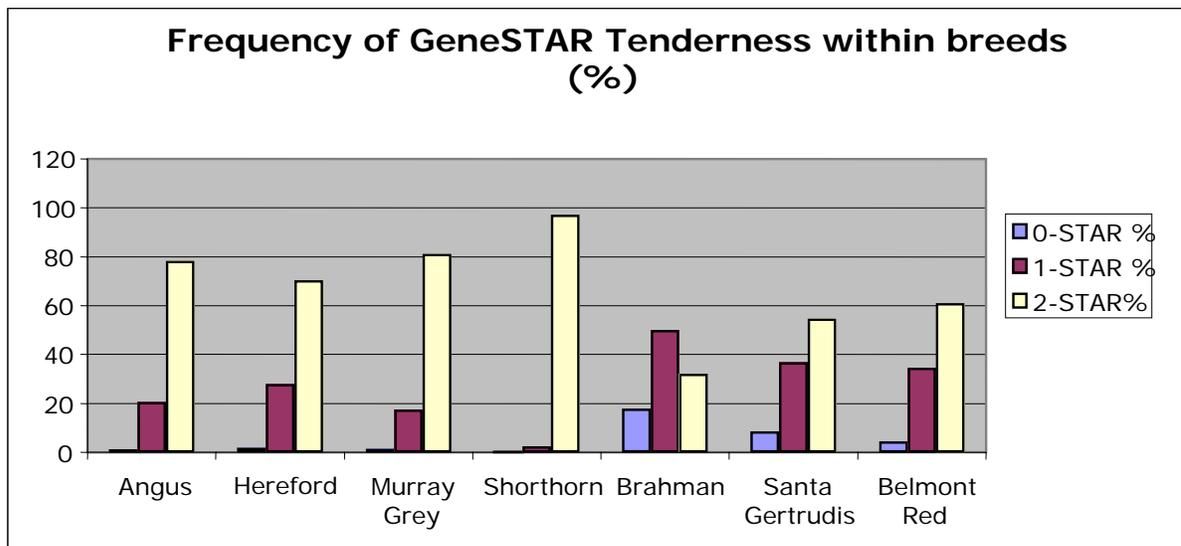


Figure 2. Genotype frequencies for GeneSTAR tenderness in different breeds. (2 STAR signifies homozygous for the tender allele, zero STAR is homozygous for the tough allele and one STAR is heterozygous).

DNA marker technology took a giant step forward in November 2003 when GeneSTAR Tenderness 2, a two gene test for tenderness was released (see. DNA markers in the second gene (Calpain 1) had been discovered by scientists at the US Meat Animal Research Centre (MARC). Research conducted by Genetic Solutions in Angus and Santa Gertrudis cattle showed that the SNP316 marker is consistently linked to tenderness regardless of breed type and also explains most of the Calpain 1 effect (see www.geneticsolutions.com.au). GeneSTAR Tenderness 2 tests for the Calpain 1 (SNP316) DNA marker in addition to the Calpastatin DNA marker. Results for each marker are added together, thereby doubling the measurement scale.

Data collected on Angus (N=742) and Santa Gertrudis (N=329) cattle show that the Angus breed has relatively high frequencies of 3 and 4 STAR (tender) animals compared with indicus-derived Santa Gertrudis cattle (Fig 3). Around 50% of Angus cattle are expected to be 3 or 4 STAR.

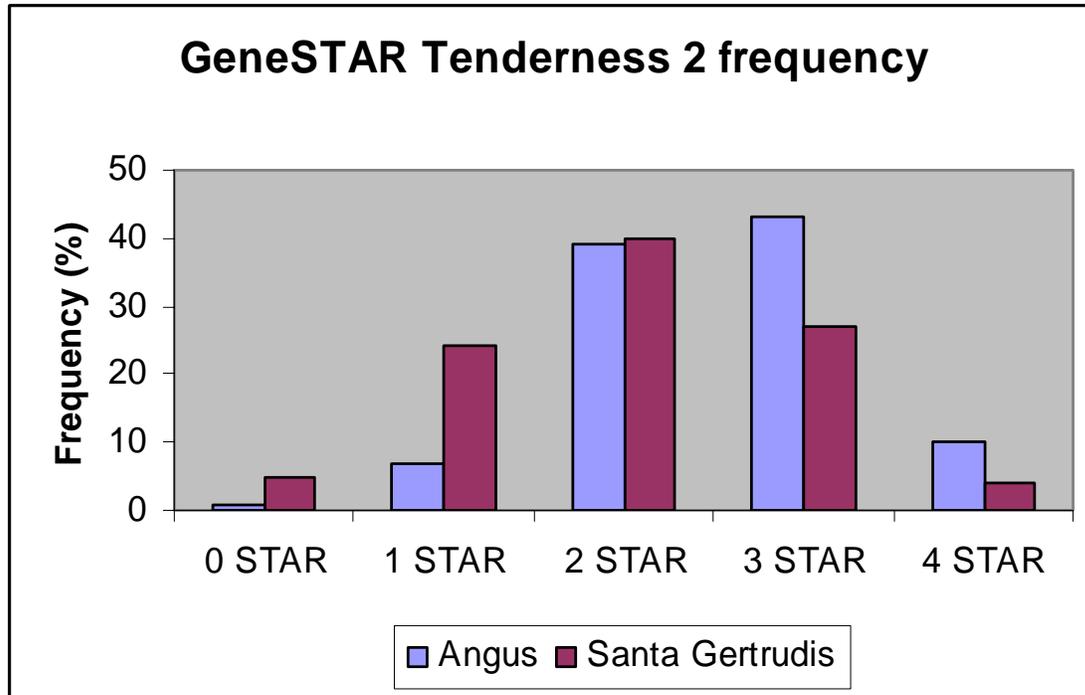


Figure 3. Frequency (%) of GeneSTAR® Tenderness 2 results in straightbred Angus and Santa Gertrudis cattle

In the Angus trial, less than 2% of 4-STAR carcasses graded tough (WBS>11 lbs) compared to 12% for animals with 1 or 2- STAR of either gene. In the Santa Gertrudis trial, there was a 3 to 4-fold reduction in the percentage of tough animals between 4 STAR and 0, 1 or 2 STAR animals. Therefore the GeneSTAR® Tenderness 2 test can accurately identify animals that will produce genetically tender cuts and by selecting for 3 & 4-STAR animals will reduce herd toughness due to genetics. To date, no correlated effects on other traits have been observed.

What does the future hold?

The ultimate impact of gene marker technology will be dependent on the number of available gene markers, size of associated direct and correlated effects and the allele frequency distribution. At this time, the number of validated markers is extremely limited, with markers only being available for two production traits i.e. marbling and tenderness. However, this situation is likely to change in the future as researchers continue to analyze the major economic traits such as tenderness, marbling, retail product yield, feed conversion efficiency and parasite and disease resistance using improved tools and genomic databases. Therefore the likelihood of multiple DNA markers being developed for a range of traits is high.

Gene markers for complex traits will be best applied as a component of an integrated genetic improvement system. The reasons for this are multiple. Firstly, markers on their own are unlikely to ever describe the total genetic variation in such traits and thus analyzing the total genetic variation expressed as an index of genetic merit such as EPD will always be desirable. Secondly, it is probable that a selection policy based solely on gene markers will yield genetic improvement at a rate less than that possible through utilizing all available genetic information. And finally, while the use of gene markers could reduce the need to measure animals for costly to measure traits, there will usually be a requirement for some measurement and this is best conducted as part of an ongoing genetic evaluation program. Whilst general methodologies exist for combining marker information, phenotypic and pedigree records, robust strategies and associated software is needed for the range of field situations.

Finally, the beef cattle industry has been eagerly awaiting the arrival of gene markers as new breeding and management tools. However, it is clear that all sectors, including scientists will benefit from a deeper understanding of the strengths and weaknesses of the technology and the opportunities it affords. Both the public and private sectors will need to invest in education and training programs if the industry is to optimize the returns from DNA markers.

FRONTIER BEEF SYSTEMS – YOUR DNA TECHNOLOGY PARTNER

**Dr. Jim Gibb
Frontier Beef Systems
Louisville, Co**

**and
Geneseek
Lincoln, Ne**

GeneSeek is a premier, global biotechnology company dedicated to creating unique business advantages to their partners by providing quality DNA-based products and services.

- Swine
- Beef
- Dairy
- Sheep

Cooperators

GeneSeek

- Genotyping
- Validation

American Simmental Association

- Full partner
- Phenotype database (NCMP)
- Validation

NBCEC

- Validation

Products

- *ParentMATCH*
- *DoubleBLACK*
- *TenderGENE*

ParentMATCH

- Parental Validation
 - Verify one or both parents as per breeder records
 - Usually to fulfill breed association requirements.
 - Listed parent either qualifies or is excluded.
 - Typically use about 12 micro-satellite markers of which 9 are ISAG markers.

ParentMATCH

- Multi-sire testing
 - “Most likely sire” based on screening potential sires and selecting the sire that has the most genes in common with each calf.

ParentMATCH

- Multi-sire testing
 - Calving difficulty
 - Identify extremes (top and bottom 20%)
 - Cleanup after A.I.
 - Bull dominance
 - Enhance pasture management
 - Enable EPD calculation
 - Genetic/product/process verification

Most Likely Sire Report							
Two Cleanup Bulls							
Animal ID Number	FBS Sample Collector Number	Sex M/F	Breed	Most Likely Sire	Number of Exclusions	Probability	Comments
032N	1000001393		CA	802L	1	95	
035N	1000001392		CA	802L	0	99	
031N	1000001391		CA	802L	1	97	
029N	1000001390		CA	802L	0	100	
025N	1000001389		CA	729L	0	100	
165L	1000001388		CA				Both possible sires excluded
021N	1000001387		CA	729L	0	100	

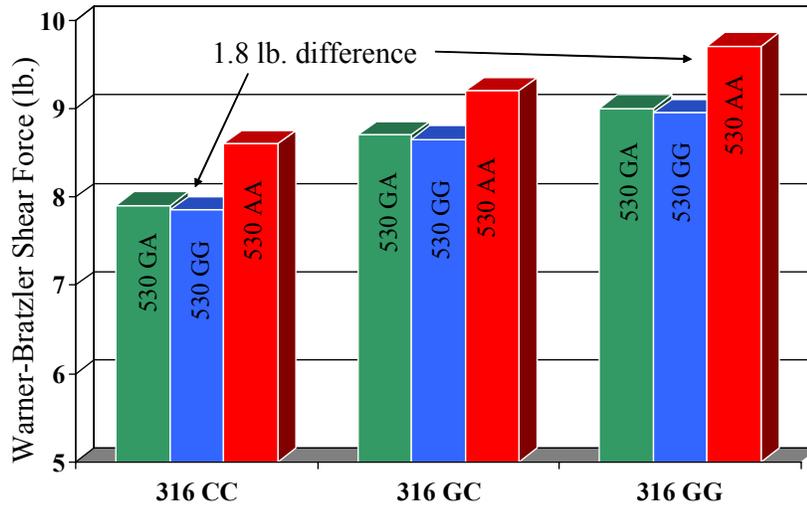
<i>DoubleBLACK</i>					
Animal ID Number	Sex M/F	FBS Sample Collector Number	Breed	Genotype	Homozygous Black
26N	F	3931098	SM	EDe	No
41N	F	3931099	SM	EDED	Yes
24M	F	3931100	SM	EDe	No
483M	M	3931101	SM	EDED	Yes
107M	F	3931102	SM	EDe	No
4N	F	3931103	SM	EDe	No

Calpain

Developed and Validated at the US Meat Animal Research Center and independently validated by FBS, GeneSeek and the National Beef Cattle Evaluation Consortium.

Two Additive Genotypes

Simmental- and Angus-sired calf-feds (270 head)



R.L. Quaas, 8/22/03

*Tender*GENE

Genotype Rank

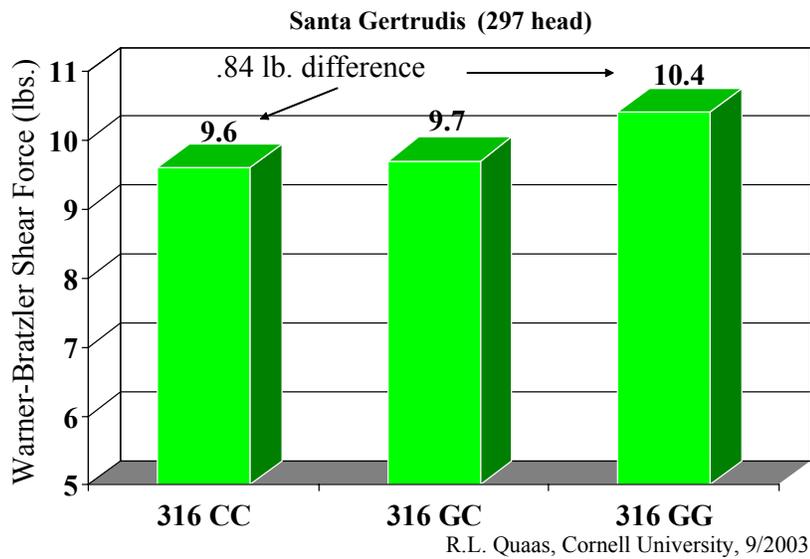
Based on current research. Subject to change with additional data.

Rank	SNP316	SNP530	Genotype Score
1.	CC	GG	5
1.	CC	GA	4
3.	CC	AA	3
3.	GC	GG	3
5.	GC	GA	3
5.	GG	GG	3
7.	GG	GA	2
7.	GC	AA	2
9.	GG	AA	1

*Tender*GENE Report

Animal ID Number	Sex M/F	Breed	Breed Reg. Number	FBS Sample Collector Number	Genotypes		*Score 1-5
					SNP316	SNP530	
277	F	AN	XXXXXXXX	4054265	CC	GG	5
335	M	AN	XXXXXXXX	4054266	GG	GG	3
330	M	AN	XXXXXXXX	4054267	CC	GG	5
329	M	AN	XXXXXXXX	4054272	CC	GG	5
280	F	AN	XXXXXXXX	4054274	CC	GG	5

Animals with the CC genotype were significantly more tender than those with the GG genotype.



Future

- EPDs “GPDs”
- Evaluate new markers.
- Market-driven traceability
 - COOL?
- Producer Education

IGENITY-L™ : GENOTYPING OF THE OBESE (LEPTIN) GENE FOR TRAITS OF ECONOMIC SIGNIFICANCE

**Dr. Stewart Bauck
Merial Ltd., 3239 Satellite Blvd.,
Duluth, GA 30096-4640**

The burgeoning field of genomics represents a tremendous opportunity for advancement in the animal production sector. For those that wish to commercialize the technology – in one form or another – it is a challenge to define your specific role and business model.

Merial Limited has recently undertaken an initiative to venture into the field of functional genomics. The fundamental business model was to enter into a collaboration with Quantum Genetics, who in turn are affiliated with the University of Saskatchewan, to commercialize a test that evaluated an animal's genotype based upon a polymorphism within the obese gene. The missense mutation in the bovine obese (leptin) gene has been shown to be stable and associated with increased fat deposition in beef cattle and increased milk yield in dairy cattle (1).

On a level, the understanding of single nucleotide polymorphisms as a fundamental basis for genetic variation is an interesting one from the perspective of a company. Because polymorphisms are essentially by definition aa, ab or bb, it makes them easy to work with and good candidates to commercialize. To quote a recent publication in Mammalian Genome – “SNPs are the fundamental unit of genetic variation and attractive as markers because they are abundant in cattle (Heaton et al. 2001b), genetically stable in mammals (Markovtsova et al. 2000, Nielsen 2000, Thomson et al. 2000) and amenable to high throughput automated analysis” (2). At its essence – simpler is better.

Simply having SNPs may not be the total answer. There are, after all, many thousands of SNPs in the genes that produce proteins coding for all manner of life's functions. One consideration for the development of a successful business model for functional genomics is to follow closely what is being done on the human side with obesity control, for example (3). In this instance, the strategy seems to be to target pathways for important processes in the body, identify the hormones or receptors important in those pathways, and characterize the genes - and polymorphisms - that exist in the animal population controlling those processes. The prospects may be good therefore, to identify a series of proteins or receptors (perhaps no more than 3 or 4) in the cascade or pathway, that result in a significant amount of the animal to animal variation in the traits controlled by the proteins in the pathway. In the case of Merial, one of the attractions to the obese gene/leptin protein was an understanding that it had an impact on something that seemed important in agriculture, namely energy balance.

Moving beyond SNPs and even moving beyond pathways, there are other important considerations. One is to try to find proteins, genes and polymorphisms that have

application in as many species or in as many production systems within a species as you can possibly find. After all, if you can target something that is useful in both beef and dairy, then you have at least doubled your target audience for the technology. Furthermore, you must have a frequency of the genotype within the commonly used breeds which is sufficient to be able to characterize the effect of the genotype, and potentially influence the genotype distribution within the population.

What are some of the observations from early explorations in commercialization? On the positive side, producers accept the technology. Their experience with genomics in the crop business leads cattle producers to accept that it is real – they believe you should be able to identify the superior animal by examining the actual genome. With a good target, and a reasonable data set, it should be possible to develop a value proposition and begin to explain to the producer exactly what is in it for him to test. Many innovators intuitively seek tools to help differentiate their animals or their breed, but mass adoption of this new technology will require that you be able to clearly demonstrate the cost-benefit ratio for implementation. Proper selection of the target gene heightens the likelihood of a success in that one gene may frequently have an impact on many traits. Today, we are still challenged that selection of animals based on genotype is “single trait selection”. In fact, the correct statement is that this may be “single gene” selection, but a good candidate gene will influence several economically important traits. For any particular gene, the influence may not always be unidirectional, however. Selection for a particular genotype may be positive for some traits, while selection for a different genotype may be positive for other traits. Farmers understand this and do not expect something for nothing – they simply want to know what happens if they choose one genotype over another and manage the animals accordingly.

There are challenges with adoption of this new technology to be sure. Farmers do not always appreciate the elegance of the science. Their frame of reference is frequently inconsistent with the cost or sophistication of the test. For example, they often equate genotyping with ultrasound for backfat. In their minds, they are both tools for genetic selection with the same effect and hence they should have the same price tag. Whether this is true or not, it is where they start from and it is up to us to try to demonstrate the difference in value for the different technologies.

The goal for companies then is to develop a solid value proposition for the technology. A genetic test does not equal a value proposition. In some instances, the test measures a gene which impacts a trait that the farmer does not get paid for. Likewise, the variation in genotype may be responsible for only a small part of the total animal to animal variation and the difference is insufficient to justify the apparent cost of the test. Compounding this is the apparent dichotomy which exists in the scientific community concerning the value of genomics. Many would argue that our current, tried and true techniques of quantitative genetics have served us exceptionally well over the years. New molecular techniques are not required. We have all we need with the advances in computing technology – bigger computers, better number crunching, better data. The classic example is milk production recording in the dairy industry. All that is really

needed is for farmers to keep better and more accurate records, and for researchers to have better tools to collect and analyze that information. That can be characterized as the “quantitative versus molecular” debate. Groups such as the Beef Improvement Federation, and this meeting, are important forums for these issues to be debated and for progress to be made. In the absence of some measure of consensus on this, we risk doing a disservice to the scientists working on the genome, and to the farmers that want to consider to use the technology. They may be scared off unnecessarily by those that are reluctant to embrace the change.

Even if one can see their way through all of these various hurdles, and a convincing argument can be made that a business is possible within this arena, there are still a variety of stumbling blocks. The future is “one stop shopping” for genomics – micro-array technology holds out the promise that a single sample of DNA will yield results for a variety of genotypes, and the farmer will check them off the order form like items on a menu. Unfortunately for those that are attempting to commercialize this, discovery is a bit fragmented at this time. Many different researchers, in many different labs, are all busy working on one or two genes or SNPs. To make a business of this, they may need to be collected into one basket to be marketed. Secondly, progress on the genome is very much a “public venture”. The internet holds out great promise for rapid progress on sequencing through sharing of information. However, that is not necessarily consistent with businesses that trade in secrets and rely on 20 years of patent protection. The willingness to invest in commercialization of the technology is limited if a relatively secure patent position cannot be found. It is also a rapidly evolving field and a challenge for anyone to really stay on top of. Success may well come to those that are fleet of foot.

Finally, as anyone enters into this arena, the first thing they need to decide is what you are. Are you a SNP identifier – in essence a sequencer, or are you a SNP quantifier – in essence you have a database of phenotypic information, or are you a SNP tester – in essence a high throughput lab, or are you a SNP marketer – in essence selling and marketing the value proposition to the farmer. Each of these positions is slightly different and it is doubtful that any one organization can be all things to all people. Correctly identifying your corporate mission, developing good business models and sticking to them while confirming the value proposition for the farmer will allow you to participate in what can clearly be described as transforming technology.

References

Buchanan FC, et al. Association of a missense mutation in the bovine leptin gene with carcass fat content and leptin mRNA levels. *Genet Sel Evol* 34 (2002): 105-116.

Heaton MP, et al. Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. *Mammalian Genome* 13 (2002): 272-281.

Marx J. Cellular warriors at the Battle of the Bulge. *Science* 299 (2003):846-849.

PRACTICAL CONSIDERATIONS OF DNA TESTING FOR QUANTITATIVE TRAIT LOCI

John Pollak
Department of Animal Science, Cornell University
Ithaca, NY 14853 USA

In this presentation I will attempt to enumerate some of the practical considerations regarding DNA testing for quantitative trait loci (QTL) and in doing so also wrap up the first day of this 8th Genetic Prediction Workshop. First, however, I will reflect on the purpose of these workshops. The idea of Genetic Prediction Workshops was conceived in the early 1980s as a method to address new technologies in a forum that included academics, beef industry support organizations, and producers. At these workshops, we are charged with discussing new topics with the aim of providing the industry (and our parent organization, the Beef Improvement Federation) with recommendations on issues and programs relating to genetic predictions and selection. Hence, these workshops are not simply another meeting for presenting research results but rather are an exploration of new topics with the goal of setting direction.

So, with this objective in mind, what is new in the industry since our last workshop? As seen from previous presentations today, we now have tools available for testing beef cattle for QTL that influence beef carcass characteristics, namely marbling and tenderness. We have heard presentations by representative of several DNA testing companies and will hear a report on the beef checkoff funded Carcass Merit Project (CMP).

I will start with considerations that are global in nature. This will include a discussion on the types of tests available and on the concept of validation of commercially available tests. The tests currently available are for marbling (a polymorphism in the thyroglobulin gene), tenderness as measured by shear force (polymorphisms in the mu-calpain and calpastatin genes) and fat deposition (a polymorphism in the leptin gene). All of these tests are biallelic in nature. Mark Thallman suggested that additional QTL alleles may exist in these genes, which by the nature of the tests are not identified. The study of myostatin polymorphisms for double muscling have taught us that multiple alleles at QTL very well may be segregating in our beef populations. Mark suggested that more research into the allelic forms of the gene tests now being commercialized needs to be done. Currently such research is not occurring. We may find that we lose important allelic forms of these genes with selection for the most beneficial allele that is currently identified by each test.

The current DNA tests were referred to by Mark Thallman as “association tests.” It is unlikely that the polymorphism screened for by each test is the causative mutation but rather a closely linked marker to the true causative mutation. Are these tests then universal for all breeds? Results from mu-calpain studies (Jay Hetzel and Richard Quaas) suggest that one of the two single nucleotide polymorphisms (SNP) (referred to

as SNP 530) identified by researchers at USDA Meat Animal Research Center (MARC) works in opposite directions in comparing results from a *Bos taurus* population (mostly Simmental-sired F1 Simmental - Angus calves) and a *Bos indicus*-based population (Santa Gertrudis). This raises a question about the level of our knowledge of the test being used when applied to all of our cattle breeds.

There are potentially several more QTL identified in the CMP for tenderness. These QTL were discovered in the Texas A&M research program using a family based on Angus – Brahman crosses and targeted for validation in the CMP. If commercialized now, these tests would be the more classical marker tests and would be applicable only to the sire families studied in the CMP. This raises the question as to whether these kinds of marker tests can be successfully marketed. A poll of the attendees of this workshop shows we are uncertain and split in our opinion on this question. For successful commercialization, there first has to be a population of cattle large enough to warrant the marketing efforts. In the CMP, breed associations were asked to nominate bulls for the DNA component of that study that they felt would be legacy bulls in their breed. A look at two such Simmental bulls shows one now has 685 sons and 1809 daughters registered to that breed and the other has 2485 sons and 5210 daughters. For QTL for which these two bulls are heterozygous, this represents the testable population. We could go one generation further with these tests using flanking markers, recognizing the potential for mistaken associations as those associations break down with recombination. **Table 1 shows the total number of progeny (males and females) and grandprogeny for the legacy bulls used by several breeds.** Is this a large enough population to motivate commercialization?

Breed	Sons	Daughters	Grandprogeny
Hereford	3287	3324	13593
Red Angus	7384	7561	31702
Simmental			
Angus			

Effective use of marker tests will require continued collection of phenotypes to discover segregation in new sire families and to document the continued efficacy of the tests within sire families where we seek to test generations beyond just their progeny. For traits like tenderness (shear force), is the industry prepared to continue doing so?

A topic revisited by many speakers and emphasized by several questions from the audience today relates to the concept of validation of DNA tests. Richard Quaas reviewed the status of validations being done by the National Beef Cattle Evaluation Consortium (NBCEC). First, why might validation be an important step in adopting a particular DNA test? There are two compelling reasons, in my opinion. First, validation is an attempt to replicate results, independent of the population used for discovery. That is simply good science, as we have always relied on replication as an integral part of accepting results. Second, there is very little information in scientific literature on some of the tests available (in fact, none at all in some cases). For those to whom the industry turns for an unbiased assessment of any new discovery, this fact leaves us in a

quandary. Recommendations on the use of the tests in selection programs or in genetic evaluation systems require evidence that can be scrutinized by the academic community. With the lack of such evidence, recommendations are tenuous at best.

Having argued above for validation, let me turn now to what constitutes validation. In my opinion, the best approach to validation is for the commercial company to blindly DNA test populations that are available for validation. Richard Quaas discussed two such populations that are currently being used by the NBCEC. There are populations at MARC and from the CMP resources that could be used as well. I believe in the blind testing approach using an unbiased third party to circumvent any potential criticism by the ultimate end user regarding the suitability of validations done by the commercial company.

Throughout this first day of the workshop, the definition of validation was challenged. Quaas discussed validations done by the NBCEC as having looked simply to see if the claims made by the commercial company could be replicated in another population. For example, with the validation of the GeneStar Marbling test done in conjunction with Genetic Solution, Inc., analyses were successfully done for marbling and for yield grade. But this was in one population. I have already alluded to the question of whether the test is viable across all breeds. Questions and comments from the audience today seemed to suggest that the effect of the QTL being tested needs to be across the gamut of traits having economic relevance to the industry. We should make sure there are no adverse **pleiotropic** effects of the gene on these traits. Do we have the resources available to do so? Are we developing databases of phenotypes on animals that could be used to do so? It was also mentioned that the interaction of the gene being tested with the environment and/or different management strategies needs to be investigated. For example, will the available tests work the same on feedlot cattle that are in an aggressive program using metabolic **modifiers** as they do in programs that do not. These are all good questions, and I agree they need to be evaluated. But at the same time, the tests are valuable tools and we need to **imply apply?** them. I suggest for the short term, that we define validation as replicating results to show the efficacy of the test and do so in two or three different populations. Then we can assess through subsequent research how the QTL affect other traits as the resources necessary to do so become available.

I will now turn to some genetic considerations regarding the DNA tests. Even if the DNA tests work on many breeds, we need to know the magnitude of the effect of the genes on phenotypes to assess the economic benefits of testing. It is unlikely that the gene effect will be the same in all breeds. We do not have this information available for most breeds. Obviously the mode of inheritance will play a role in how the tests impact the economy of a ranch. The two tests for tenderness each appear to be additive. Quaas showed results for the GeneStar Marbling test, and when sires are included as a random effect in the model, the beneficial allele appears recessive for marbling but additive for yield grade. The frequency of the beneficial allele is also important. For mu-calpain and thyroglobulin, the frequencies by breed range from low to moderate. For calpastatin, the frequency is high. A final consideration is the interaction between

genes. Jay Hetzel showed that mu-calpain and calpastatin seem to work together in an additive fashion. These are the only results available at this time for these two tests. Are we prepared to test gene interactions in our validation?

I believe that the feeling of attendees in general is that the effective use of the tests in selection will come from combining DNA test results with genetic evaluations. This would either be done by including results in genetic evaluation models or, as suggested by Dorian Garrick, perhaps by indexing test results with estimated progeny differences (EPD). The NBCEC is currently working on a pilot project for inclusion of DNA information in genetic evaluations for shear force under the direction of Richard Quaas.

At this time, I think a comment on economics is important. In our enthusiasm at this workshop for the concept of DNA testing and our focus on the genetics of available tests, we did not address the economics of testing. At current costs, how does a producer recover the costs of DNA testing? Seedstock producers under any model save one of benevolence must capture a return on investment either among themselves or from the commercial industry and so on down the cattle pipeline. Very little has been done to date to address economic models that generate a return on investment from selection response.

The next area of concern to me is, for the lack of a better term, the accuracy of test results. Under the current system of testing, any producer can send any animal's sample to a lab for testing. Results are returned directly to the producer. My concern is twofold. I do not see where the checks and balances are for ensuring the results of the test nor do I see how, under this model, we capture the results for use in genetic programs for a breed.

Regarding my first concern, right now there are no checks and balances. DNA testing will be subject to the same problems as any field data recording process to include misidentified animals, misidentified samples, etc. Under the current system, there is no random retesting of animals to confirm their genotype by a breed association or any other organization as is done with parentage. Quite simply this means to me that there will be a significant proportion of animals for whom the genotype is wrong (misidentified sample) or not consistent with that animal's pedigree (misidentified animals).

Regarding the second concern, an informal polling of the breeds represented at this workshop revealed that there is no mechanism for routinely capturing tests results. Hence, even if we wanted to include DNA results in evaluations or wanted to set up a program that could be used to confirm genotypes, we are not prepared to do this at this time. Both these issues need to be resolved if there is any hope of using DNA testing for genetic selection in a effective manner.

The final consideration in this presentation has to do with testing at the ranch level. At the current pricing of DNA tests, there is no question in my mind that testing will be done selectively. I would like to see us develop a decision aid program that helps producers decide on which animals to test. In many cases there are some obvious

candidates for testing, such as herd sires and potential herd sires. However, beyond these individuals, selective testing should be done with the objective of maximizing information on the cow herd. I will use an example to demonstrate this concept. In Figure One, there are 33 animals in this pedigree. The maximum information on a particular gene in this pedigree would be obtained by testing all 33 animals. But the economic benefits of testing all 33 are questionable. Assume the first animal to be genotyped, simply because he is a live herd bull, is the one circled in Figure Two. If he is genotyped as homozygous for any of the QTL alleles, we learn quite a bit about the pedigree. First we know all of his progeny are heterozygous at the minimum for that allele. We also know that his sire is at least a heterozygote and that his siblings all have a 50% chance of having received the same allele as is homozygous in the tested bull. If he is a heterozygote, we learned very little about the rest of this pedigree. If we are selectively testing to maximize some objective function such as the proportion of total information (defined as that amount of information achieved when all animals are genotyped), then which animal should be genotyped next depends on the information generated by knowing the genotype on the first tested animal. This process would be repeated as each test result is received. These kinds of decision tools and systematic approaches to using testing will enhance the acceptability for the tests and add some logic to the process of testing.

If we reflect again on the objectives of this Genetic Prediction Workshop as being to identify concerns and issues and develop strategies for implementation of new technologies, then I think this first day for the workshop has been quite successful

Figure 1.

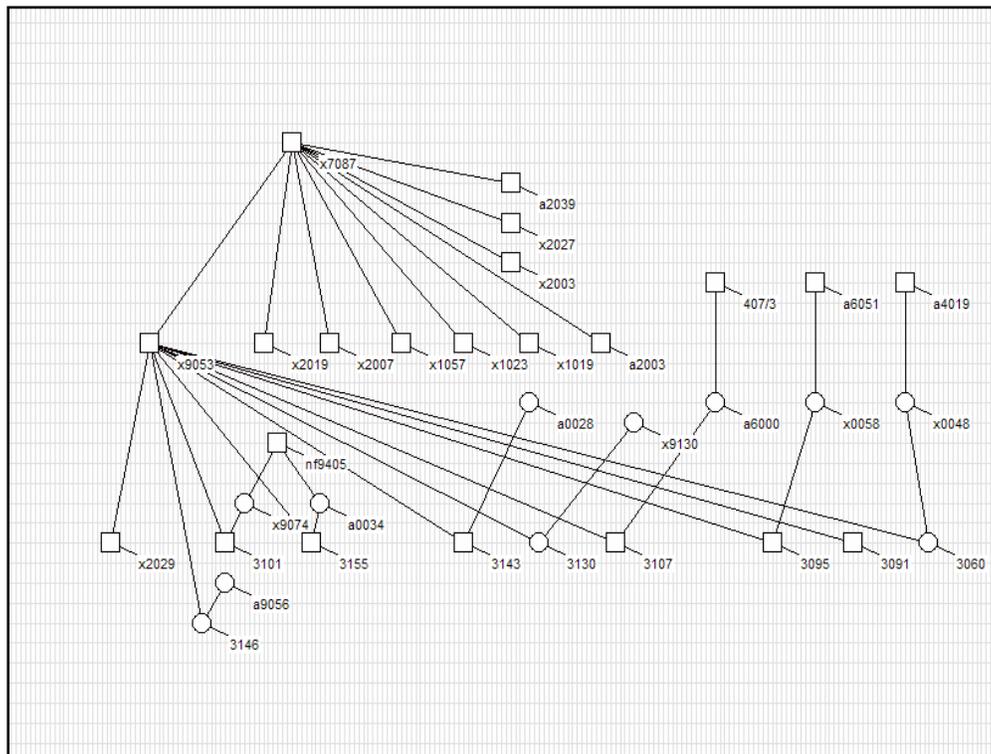
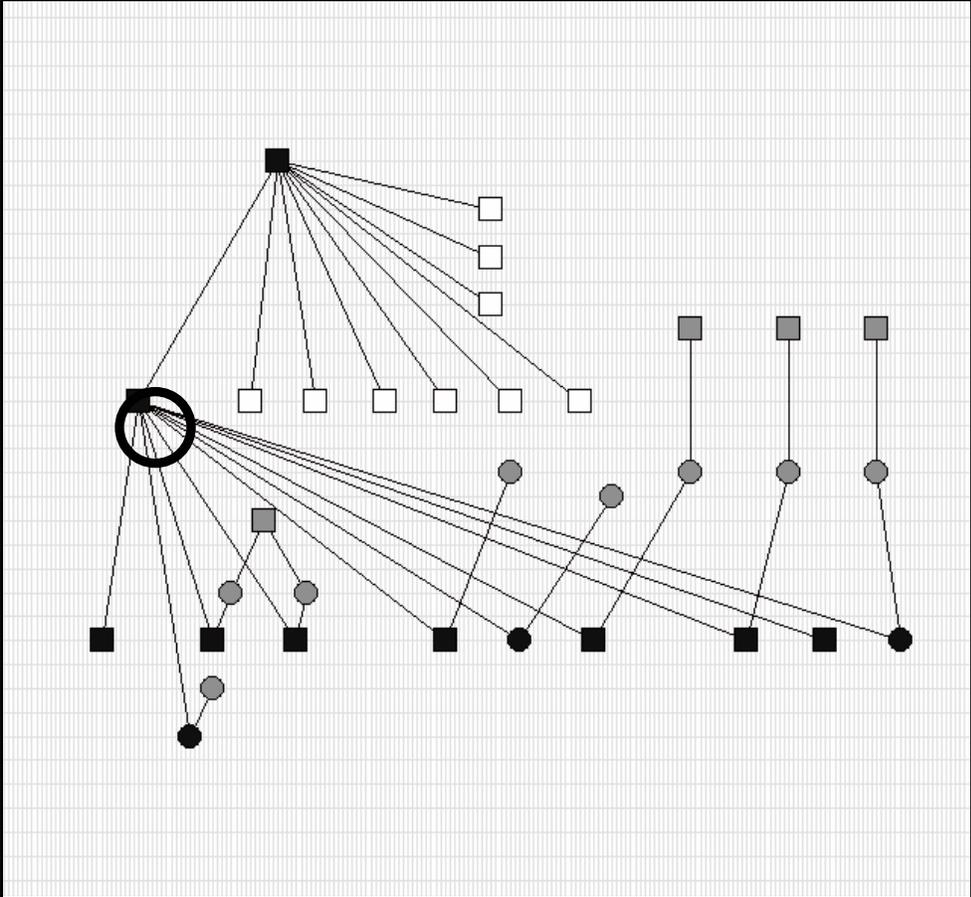


Figure 2.



MULTIPLE SIRE MATING: AN EXPERIMENTAL PERSPECTIVE^{1, 2, 3}

M. D. MacNeil
USDA Agricultural Research Service,
243 Fort Keogh Road, Miles City, MT 59301
Phone: (406) 232-8213
E-mail: mike@larrl.ars.usda.gov

¹This research was conducted under a cooperative agreement between USDA - ARS and the Montana Agric. Exp. Sta. USDA, Agricultural Research Service, Northern Plains Area, is an equal opportunity/affirmative action employer and all agency services are available without discrimination.

² Mention of a proprietary product does not constitute a guarantee or warranty of the product by USDA, Montana Agric. Exp. Sta. or the authors and does not imply its approval to the exclusion of other products that may also be suitable.

³ Extensive laboratory work of Mr. Larry French in conducting the genotyping necessary to establish and verify parentage of Line 1 Hereford cattle and the contribution of Ms. Vicki Reisenauer in summarizing the data presented herein are greatly appreciated.

Introduction

Maximizing the probability that each cow exposed produces a calf is often an implied goal of both researchers seeking to improve beef production efficiency and seedstock producers. In the contexts of research and seedstock production, reproductive management options may be limited by the need to maintain accurate records of paternity. The advent of paternity testing based on DNA markers opens up the possibility of multiple-sire mating. Previous research may be interpreted to suggest potential to increase reproductive performance through multiple-sire mating.

Bulls used in pairs each served more heifers than the same bulls did when used singly (Godfrey and Lunstra, 1989). In addition, slight increases in pregnancy rates have been observed in breeding pastures with two bulls relative to breeding pastures with one bull (Ferin et al., 1982; Neville et al., 1987). Heterospermic insemination, which is commonplace when two or more bulls are present in a pasture, may also increase the probability of conception (Ferin et al., 1982; Nelson et al., 1975). Thus, with the advent of molecular genetic technology for determining paternity, we hypothesized that using two bulls in each breeding herd would increase the probability of each female being pregnant following a 60-d breeding season. Objectives of this research were to test this hypothesis and to examine the consequences of implementing the practice of using two bulls in each breeding herd on the *in situ* genetic conservation of Line 1 Hereford germplasm.

Materials and Methods

The work described herein is not a classical experiment, but a comparison of two periods in time over which the use of one or two bulls per breeding herd is confounded

with years. The data reported here result from implementing the practice of using two bulls in each breeding herd of the Line 1 Hereford population maintained at Miles City, Montana beginning in 1997. Data from the preceding 7 yr are used for the purpose of comparison. The targeted inventory of breeding females was approximately 240 per year, except 1995 to 1997 when approximately half the available Line 1 females were bred to crossbred bulls.

Sires were selected largely from within paternal half-sib families and no dam produced more than two bulls that become sires (MacNeil et al., 1998). The selection criteria used in choosing bulls that became sires emphasized light birth weight and greater growth to 1-yr of age. Bulls used for breeding also passed a breeding soundness exam before the breeding season each year.

In all years, inbreeding coefficients were calculated for all potential progenies of the females available and the bulls that had been selected for use. Females were stratified by age and within strata assigned to mates such that the level of inbreeding was restricted by a pre-selected maximum threshold with approximately equal numbers females being assigned to each breeding herd. All bulls were either 15-17 mo. or 27-29 mo. of age at the beginning of the breeding season.

During the first 7-yr period (1990 to 1996), Line 1 Hereford females were exposed to a single sire for approximately 60 d. Average number of bulls used each year was 9.7 and the bull:cow ratio was approximately 1:18.

During the second 6-yr period (1997 to 2002) females were exposed in breeding herds of two bulls for approximately 60-d. Average number of bulls used each year was 10.8 and the bull:cow ratio was approximately 1:21. Pairs of bulls used in each breeding herd were of similar ages and had been together at least 2 mo before the breeding season began. Differences in semen characteristics were not considered in pairing bulls.

When females were exposed to more than one bull, paternity of their offspring was determined using highly polymorphic microsatellite DNA markers. To facilitate the identification of sires for all progeny, genomic DNA from each of the bulls was purified from a blood sample. Genotypes of all bulls used were determined for a panel of approximately 120 microsatellite markers. For each pair of potential sires, microsatellites were identified where each bull had different alleles. A blood sample was collected from each calf at the time of branding 10 to 70 days after birth and genomic DNA was harvested. Calves from dams exposed to those bulls were genotyped for the markers that had been pre-selected by based on their differing alleles at the locus. If neither of the calf's two alleles were present in one of bulls to which its dam was exposed, that bull was excluded from paternity and the other bull recorded as the sire of the calf. This process is known as single marker exclusion. With rare exceptions, all calves were genotyped for as many microsatellites as needed to exclude one or the other of the bulls to which their dam was exposed. On occasion, breeding herds became mixed for short periods of time as a result of less than perfect fences and every

other day observation by herdsman. For calves born in the window of time 272 to 292 d from the period of mixing, the four bulls assigned to the mixed herd were considered potential sires, necessitating the examination of additional markers to conclusively identify their sires. When paternity could not be conclusively established by single marker exclusion, dams of the calves were also genotyped.

Results and Discussion

Average annual pregnancy rates were 81.9 and 84.3 % for the periods 1990 to 1996 and 1997 to 2002, respectively. Assuming 240 breeding females exposed annually, this difference in pregnancy rate resulted in nearly 6 more calves being born annually as a result of having two bulls in each breeding herd rather than one. While this difference is not significant ($P > 0.10$), it is of a magnitude to be of economic importance and more than offsets the cost of genotyping in our laboratory. However, it is unlikely to be sufficient to offset the commercial cost of genotyping. Risk, in terms of variation in pregnancy rates, was virtually unchanged in the two time periods.

In closed experimental populations and in other applications where limited population size is a concern, limiting the rate of increase in inbreeding may also be an objective. Exposing each bull to a fixed and constant number of females limits the potential number of progeny he can sire, whereas placing two bulls in a pasture with twice as many females may introduce the potential for more variable family sizes. In these data, the average intra-year variance in number of progeny per sire was 12.5 with single-sire mating and 64.3 with two sire breeding herds. However, sixty-five percent of the sires produced progeny in two years and the variance in paternal half-sib family size was 54.3 and 172.0 for the periods 1990 to 1996 and 1997 to 2002, respectively. As a consequence, inbreeding is expected to accumulate 2.8 times more rapidly when using two bulls per breeding herd than if single-sire breeding herds are used.

Competition among bulls can increase their sexual response and other interactions among bulls in multi-sire groups are well recognized (Chenoweth, 2000). Social dominance may affect numbers of progeny with dominant bulls siring the majority of calves in multiple sire groups (Chenoweth, 1999). Additionally, efficiency of breeding, in terms of services per conception, is compromised by multiple sire mating (Ferin et al., 1982). With multiple-sire mating, more efficient breeding occurs if the bulls are less than 3 yr old, of similar size and age, and had been raised together (Chenoweth, 1999). These conditions were met in assigning bulls to Line 1 breeding herds during the period 1997 to 2002.

Literature Cited

Chenoweth, P. J. 1999. Assessment and management of breeding bulls. Rev. Fac. Agron. 16:677-689.

- Chenoweth, P. J. 2000. Bull sex drive and reproductive behavior. In: P. J. Chenoweth (ed.) Topics in Bull Fertility. International Veterinary Service Organization, www.ivis.org. Doc. No. A0509.0600 accessed 11/21/2003.
- Ferin, P. W., P. J. Chenoweth, E. R. Mateos, and J. E. Paxton. 1982. Bulls mated to estrus synchronized heifers: single vs multiple sire breeding groups. *Theriogenology* 17:365-372.
- Godfrey, R. W. and D. D. Lunstra. 1989. Influence of single or multiple sires and serving capacity on mating behavior of beef bulls. *J. Anim. Sci.* 67:2897-2903.
- MacNeil, M. D., J. J. Urick, and W. M. Snelling. Comparison of selection by independent culling levels for below average birth weight and high yearling weight with mass selection for high yearling weight in Line 1 Hereford cattle. *J. Anim. Sci.* 76:458-467.
- Nelson, L. D., B. W. Pickett, and G. E. Seidel, Jr. 1975. Effect of heterospermic insemination on fertility of cattle. *J. Anim. Sci.* 40:1124-1129.
- Neville, W. E., K. L. Richardson, and P. R. Utley. 1987. Breeding performance of bulls assigned either to 40 cows per bull or 80 cows per two bulls during the breeding period. *J. Anim. Sci.* 65:872-876.

DNA PARENTAGE TESTING—RANCH SETTING

**Robert L. Weaber
Cornell University, B-47 Morrison Hall,
Ithaca, NY 14853**

Although there have been considerable advances in beef cattle molecular genetics for quantitative and qualitative trait selection in the past few years, DNA parent validation and parent testing remains the most frequent and broadly applied of the molecular genetic tools in the beef industry. DNA parent validation and testing has been used primarily by the beef seedstock sector, but with costs decreasing considerably the technology is seeing some use in the commercial sector as well.

Validation versus Testing

It is important to recognize the subtle distinction between DNA parent validation and DNA parentage testing as one considers their potential usefulness or application. DNA parent validation is an exclusion based methodology whereby the genotypes of alleged sire and dam are compared against a given progeny genotype to determine if the parents qualify as such. DNA parent validation is commonly used by breed associations to assure pedigree integrity. For example, most breed organizations require that bulls who will be used to service cows via AI be parent verified. In this case, DNA samples are collected from the bull as well as his sire and dam. Typically genotypes for each animal are produced using microsatellite markers. Our expectation is that the bull inherited one allele at each locus from his sire and one from his dam. Operating under this assumption, the genotypes are analyzed for 'exclusions.' An exclusion is an allele at a specific progeny locus that could not have been inherited from either of the alleged parents. If a sire or dam has no exclusion then that parent is said to 'qualify' as the parent. Since DNA parentage testing is an exclusionary-based methodology, we cannot prove an animal is a parent. We can prove those animals that are NOT a parent. Therefore, it is critical that marker panels used to genotype animals are sufficiently robust to exclude a random animal as a parent.

DNA parentage testing is a considerably different procedure designed to assign paternity (or maternity) with some degree of probability to one or several of the alleged parents given the progeny genotype. One or both of the alleged parents may be unknown. Seedstock and commercial producers use DNA parentage testing. This procedure is sometimes used to make paternity assignments when multiple-sire breeding pastures have been used to mate cows. For instance, a rancher may have 200 cows and 10 bulls with no way to separate the animals into single sire mating pastures for the breeding season. For each progeny that result from this mating scheme, it is known that the animal's sire is one of the 10 bulls but which of the ten is unknown. By genotyping the sires and progeny and then analyzing the genotypes for exclusions, it is possible to disqualify many of the bulls as a calf's sire. Hopefully this leaves a single bull qualified as the sire.

DNA Parentage Testing

Motivation

The use of DNA parentage testing is motivated by a number of desired outcomes. One of these may be the assembly of a putative pedigree structure for sires and progeny or sires, dams and progeny. A common use of the pedigree structure would be to track the inheritance of desirable or undesirable genes. A pedigree structure is also useful in genetic improvement or selection programs where phenotypes are also collected and used in the generation of EPDs. This is the situation in which researchers at Cornell University found themselves as they worked to develop the Commercial Ranch Project (CRP). A pilot of this is underway at the Bell Ranch in New Mexico. In the pilot, a large scale progeny test program is underway where yearling bulls from the integrated seedstock unit are randomly mated to cows from the commercial division in large multiple-sire breeding pastures. The goal is to select progeny tested sires for use in the commercial herd based on economically important EPDs. In the project it is important to assign paternity probabilities to sires based on progeny genotypes, construct a pedigree and collect phenotypes for genetic evaluation.

Challenges

Unfortunately, a number of complications arise during the process of DNA parentage testing. The process of collecting and identifying DNA samples from bulls and calves is not a small task in a large operation. Hair roots from the animals' switch proved to be the most convenient DNA source. Commercially available barcoded hair collection cards were used in the CRP pilot. Electronic (RFID) ear tags were installed into animals for identification. A palm computer system equipped with a barcode reader, electronic ear tag reader and customized database was developed to cross-reference the ear tags and DNA samples as well as collect other phenotypic data such as sex and horn status.

Following the collection of the DNA and subsequent genotyping, analysis of the data for paternity assignment presents several challenges. A common problem is the inclusion or qualification of several bulls as the sire of a given calf. It is possible for bulls, particularly those that are half or full siblings to have similar genotypes passing similar haplotypes to their progeny. As a result it may be difficult to disqualify one of them as a parent. Genotyping of the known dam of a calf partially mitigates the problem of inclusion of multiple sires by eliminating some of the calf's alleles from consideration when making paternity assignments. Genotyping of dams may be feasible in seedstock operations where calving cows are closely observed and calves are identified at or near time of birth. In this case, cost of genotyping may still prohibit dam genotyping. In large commercial operations dam genotyping is not feasible as most cows, except maybe first calving females, are not observed closely at calving time and calves are not identified at birth. In this case, the dam portion of a calf's pedigree is considered unknown. Cost of genotyping these dams is a major consideration for the commercial producer.

While genotyping of dams may not be feasible or cost effective, sorting sires with similar genotypes into separate breeding pastures is an effective method to help minimize multiple sire inclusions. When there are more than just a very few sires, the task of sorting bulls by genotype is best left to a computer. A visiting student to the Cornell Animal Breeding group developed a software program to sort sires into groups based on their genotypes that maximizes the probability of unique identification of a sire's progeny.

The possibility of a stray or unknown sire is an important consideration as one contemplates the computation of paternity probabilities or assignment of paternity. It is likely in large extensively managed operations that a bull from an adjacent pasture may jump the fence and breed cows involved in your test. This bull may be one that is genotyped and being tested or it could be the neighbor's stray bull. In any case, it is important to recognize the possibility that a calf may not have any bull(s) in the mating group to which its dam was assigned identified as the sire.

Genotyping errors present another distinct possibility for the reason a calf does not have a bull included as a sire. Likewise, genotyping errors may result in a bull being included as a sire for some calves that without the error he would be excluded. Dr. R. L. Quass (2003) has suggested marker genotypes are observations of an animal's true genotype. It is the molecular and informatics tools we employ that provide the observation. It is clear that while these tools are extremely useful they are not without error. To help minimize this problem, sires in the CRP are genotyped three times using independent DNA extraction, PCR amplification and automated genotype scoring. Nonetheless, the possibility of a genotyping error, especially involving a calf, should be considered when computing paternity probabilities.

Molecular Tools

Currently DNA microsatellite markers are used most commonly for both DNA parentage testing and validation. Fluorescent multiplexes of microsatellite marker panels have been developed for semi-automated genotyping on high throughput DNA sequencing equipment. The relative ease and low cost of microsatellite markers makes them good candidates for genotype paternity testing. Use of a standardized panel of markers such as the International Society of Animal Genetics (ISAG) bovine panel is common. Most commercially available genotyping services offer a panel of markers that include the ISAG panel. Use of a standard panel may be advantageous if one plans to use several laboratories or service providers for genotyping or anticipates a future change in laboratory or provider. However, a number of other panels are available for parentage testing.

These alternate panels have been developed to improve lab efficiency, facilitate multiplexing and minimization of the number of PCR reactions needed for genotyping, or to improve the informativeness of the marker panel. It is important to consider which panel you will use for parentage testing since sires and progeny must be tested using the same panels.

A number of key marker panel features should be considered. The marker panel's exclusion probability is primary. The exclusion probability is the panel's ability to exclude a randomly chosen animal from the population as a parent of a given animal. Vankan and Faddy (1999) suggest that a high reliability of paternity assignment when missing sires are considered requires the use of marker panels with high exclusion probability. A panel's exclusion probability is a function of the number of markers and alleles in the panel as well as the allelic frequencies in the testing population (Jamieson and Taylor 1997). Table 1 illustrates that greater than 70% of the calves sired by a group of 10 bulls will have a single bull included as its sire when a marker panel with an exclusion probability of 0.97 is used for genotyping and dams are not genotyped. The probabilities of single and multiple sire inclusion resulting from the use of marker panel with exclusion probabilities between 0.90 and 0.9950 are illustrated in Table 2. A high degree of the allelic heterozygosity for the marker alleles is also desirable.

In addition to a marker panel's statistical power, one should be considerate of the panel's laboratory features. Clearly, this list of panel features will be of most concern to molecular geneticists and lab technicians. It is beneficial for users of the data to be casually aware of a panel's behavior in the lab. It is desirable that the marker panel consists of markers that are reliably amplified by PCR. A high frequency of incompletely amplified fragments will result in incorrect fragment length and allele scoring. Reliable amplification and properly calibrated allele scoring software is essential for proper genotyping, especially for homozygotes and heterozygotes that differ by only two base pairs.

Finally, the cost per animal for genotyping and paternity assignment is often the critical point of differential when comparing marker panels. The cost of genotyping can vary considerably depending on the marker panel used. Key variables in determining genotyping cost are the number of markers in the panel, the number of PCR reactions required and technician labor. Genotyping was contracted independently of paternity assignment and probability services for the CRP. Researchers at Cornell analyze sire and progeny genotypes to make paternity assignments. These methods are described in the following section. Uncoupled genotyping and analysis will likely be unfeasible for the average user of DNA parentage testing tools. Fortunately, commercial services that bundle genotyping and paternity assignment are available.

Informatics Tools

As mentioned above the Animal Breeding Group at Cornell University has developed a suite of software applications for genotype analysis and paternity assignment (Pollak, 2003). The suite consists of three applications, each addressing specific needs that have arisen in the CRP. The three applications are briefly described below to provide some perspective of the capabilities and features that have been useful when using genotypes for paternity assignment.

Sire Match

Sire match is the first application in the suite. It imports formatted genotype files, which contain the genotypes of bulls, calves and dams (if available). If dam genotypes are present then dams relation to their respective calf is validated. The software utilizes a likelihood-based algorithm developed by R.L. Quaas to assign paternity probabilities. The software utilizes breeding group information in its testing procedures but also tests against all sires to allow for the occasional 'fence jumper'. Additionally it considers the possibility of a missing (ungenotyped) sire and genotyping errors.

Sire Management

The second application in the suite is Sire Management. The data source for Sire Management is an output file created by Sire Match. Sire Management allows the user to set varying acceptance levels for paternity probability and exclusion when making paternity assignments. The software can recommend second panel genotyping to resolve multiple sire inclusions as well as recommendations for regenotyping if a user defined number of alleles are missing in an animal's genotype.

Sire Diagnostics

Sire Diagnostics is the final application in the suite. This software provides drill down capability to allow for easy viewing and user analysis of the results of the Sire Management software. Records may be sorted in a variety of ways. The software is particularly useful for analyzing genotypes of calves that have a high probability of being sired by a specific bull but may have an exclusion(s). In some cases this exclusion maybe the result of a genotyping error involving the sire.

Tips for DNA Parentage Testing Success

1. Use a marker panel with a sufficiently high exclusion probability. A panel with 12 highly polymorphic microsatellite markers is typically sufficient. An even balanced allele frequency within each marker is most desirable. Avoid markers with few alleles or ones with an allele that appears in high frequency in the testing population.
2. Minimize the number of bulls per breeding pasture when possible. Testing calves against a smaller number of alleged sires is more powerful than testing against a very large number of alleged sires. A reasonable guideline is ten bulls or fewer in a breeding pasture.
3. Keep breeding exposure records. Knowing which group of bulls sired a particular group of calves aides in the assignment of paternity probabilities.
4. Sort similar genotype bulls into different breeding groups to enhance the probability of have a high percentage of single sire inclusions. This task is best accomplished through a software application. However, if no such application is available, sort half- or full-sib bulls into different groups.

5. Dam genotypes are helpful in assigning paternity. Use them if they are available and it's cost effective to obtain them.
6. Minimize the possibility of a sire's genotype containing an error by obtaining three independent genotypes. Several labs provide this as a routine procedure when genotypes are going to be used for paternity testing. An error in genotyping an alleged sire will result in many false exclusions of progeny, whereas a genotyping error involving a calf will only effect that calf's paternity assignment.

Table 1.

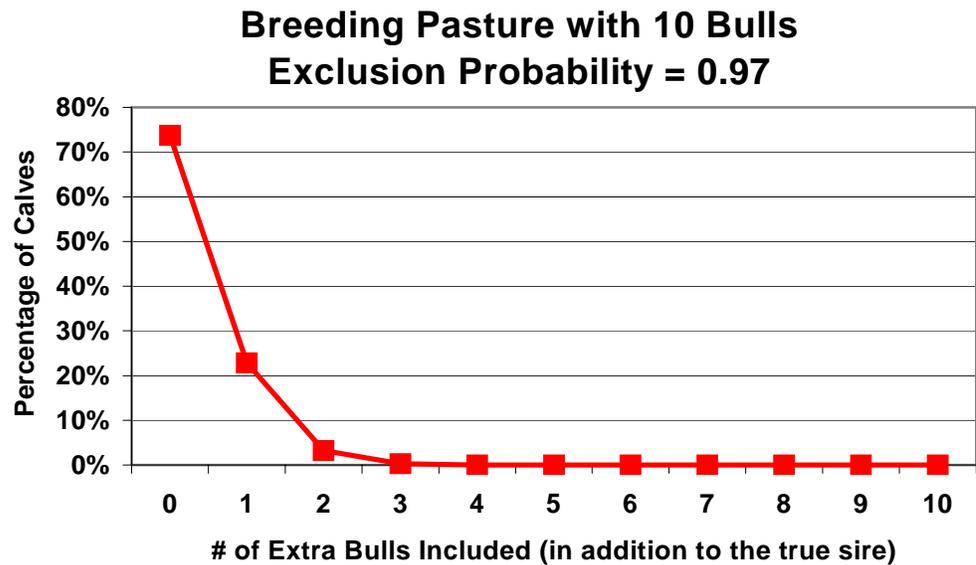
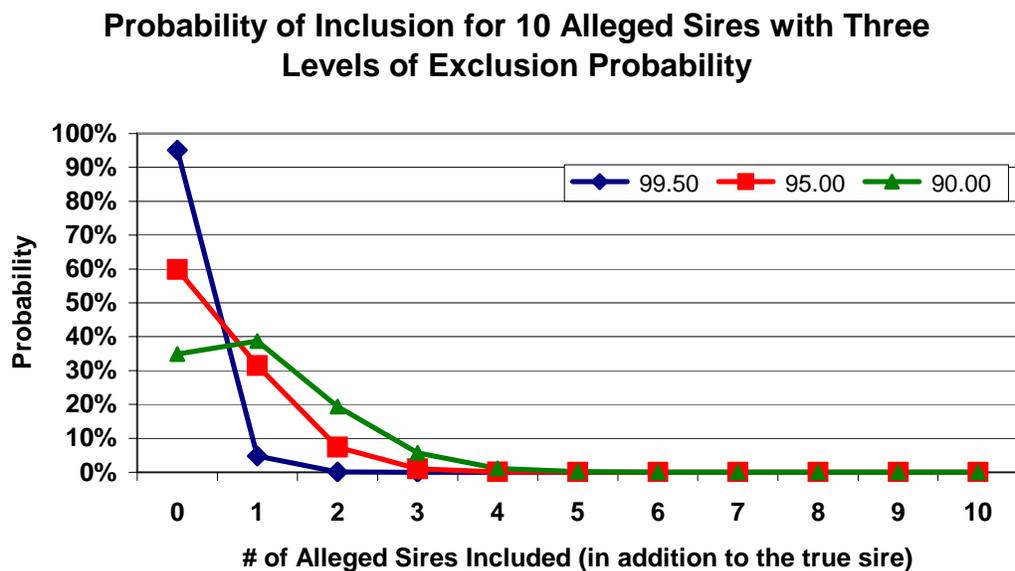


Table 2.



Selected References

Heyen D.W., Beeve J.E., Evert R.E. et al. (1997) Exclusion probabilities of 22 bovine microsatellite markers in fluorescent multiplexes for semiautomated parentage testing. *Animal Genetics* 28, 21-27.

Jamieson A. & Taylor St. C.S. (1997) Comparisons of three probability formulae for parentage exclusion. *Animal Genetics* 28, 397-400.

Pollak, E.J., Personal communication, December 2003, Cornell University, Ithaca, NY.

Quaas, R.L., Personal communication, December 2003, Cornell University, Ithaca, NY.

Vankan D. M. & Faddy M. J. (1999) Estimations of the efficacy and reliability of paternity assignments from DNA microsatellite analysis of multiple-sire matings. *Animal Genetics* 30, 355-361.

CARCASS MERIT PROJECT: DNA MARKER VALIDATION

R. Mark Thallman, U.S. Meat Animal Research Center^{1,2}
Dan W. Moser, Kansas State University
Elizabeth W. Dressler, National Cattlemen's Beef Association
L. Radu Totir, Iowa State University
Rohan L. Fernando, Iowa State University
Stephen D. Kachman, University of Nebraska
Janice M. Rumph, Montana State University
Michael E. Dikeman, Kansas State University
E. John Pollak, Cornell University

The Carcass Merit Project (CMP) was initiated in 1998 stemming from concern over the frequency of unsatisfactory eating experiences due to inadequate tenderness. The project was funded by America's beef producers through the \$1 per head checkoff, by the participating breed associations, and by MMI Genomics, Inc.

The primary goal of the project was to provide the tools and mechanisms to genetically identify superior animals in the U.S. beef cattle population that will produce progeny with the greatest potential for meeting the demands of consumers. The participating breed associations generated individual databases that allowed the development of EPDs for important carcass traits, including Warner-Bratzler shear force (WBSF) and sensory attributes. In addition, genetic markers for economically important carcass and consumer satisfaction traits were validated in the general U.S. beef cattle population. The markers evaluated resulted from previous checkoff-funded research at Texas A&M University (TAMU).

At completion, carcass data have been collected on over 8,200 progeny of project sires. DNA marker analysis was completed on progeny of 70 sires representing 13 breeds.

Four universities, the USDA Agricultural Research Service, and 13 breed associations cooperated with the National Cattlemen's Beef Association (NCBA) on the Carcass Merit Project. Shear force and sensory panel data was collected at Kansas State University. MMI Genomics performed the primary DNA laboratory work for the study. Dr. Dan Moser of Kansas State University acted as the facilitator and liaison to the breed associations.

Texas A&M University conducted the quality control testing for the project and performed the DNA marker statistical analysis on an individual sire basis. An independent validation of the TAMU analysis, as well as a breed-wide and project-wide analysis of marker data was performed by the U. S. Meat Animal Research Center. A secure database containing all the relevant data for the project has been maintained by Cornell University. Colorado State University economists have estimated economic returns to producers using carcass EPDs for cattle selection.

¹ Agricultural Research Service, U.S. Department of Agriculture, Clay Center, NE 68933

² Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Project Objectives

- Generate data from which genetic evaluations for tenderness and other sensory traits could be computed.
- Develop methodology and procedures for collection of information necessary for further development of EPDs for carcass traits.
- Validate DNA markers discovered in previous checkoff-funded research at TAMU for use in industry-wide marker-assisted selection programs for improvement of carcass traits.
- Measure costs and returns of implementing EPDs for carcass traits for the alternative genetic selection programs and combinations of management x genetic improvement of carcass traits.
- Breed comparison was strictly precluded from being an objective.

Project Design

The logistics for the project are described in Figure 1. All US beef breed associations were invited to participate. The following breeds participated in the project:

- Angus
- Brahman
- Brangus
- Charolais
- Gelbvieh
- Hereford
- Limousin
- Maine-Anjou
- Red Angus
- Salers
- Shorthorn
- Simmental
- Simbrah
- South Devon

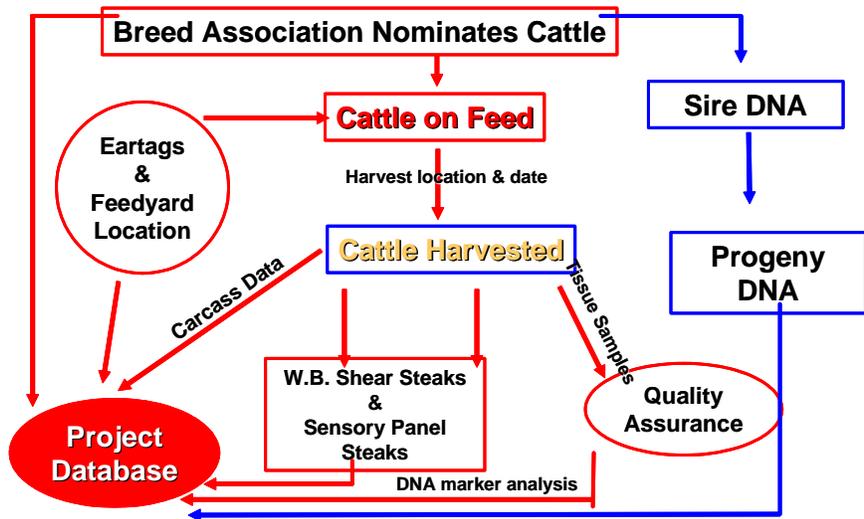


Figure 1. Carcass Merit Project Flowchart

Commercial cows were inseminated to several of the most widely used AI sires of each of the breed associations cooperating and supporting the research project. It was the responsibility of each breed association to select the sires and provide the leadership and all costs associated with nominating cattle for the study. The associations were also responsible for the semen, AI, collection of feedlot performance data, blood collection, shipping of blood samples, and the development of EPDs for their respective breeds. Breed identity was coded to prevent breed associations and/or breeders from comparing breeds.

Up to ten bulls from each breed were designated as “DNA sires.” Fifty progeny of each of these sires were used for DNA analysis and shear force measurements. Five of the

ten DNA sires for each breed were designated “sensory sires.” Sensory panel data were collected on all fifty progeny of each of the sensory sires. Additional bulls were allocated by breed based on registration numbers for EPD analysis. Twenty-five progeny of each of these “EPD sires” were allotted for shear force measures.

Progeny were fed at several locations and ultimately harvested at several cooperating processors. Age at which cattle were started on feed and other appropriate information were collected by each participating breed association. Breed associations were encouraged to minimize the number of contemporary groups and to harvest each group in a single day, whenever possible.

Carcass data, including carcass weight, ribeye area, fat thickness, marbling score and percentage of internal fat, were collected. In addition, researchers obtained one steak from each progeny of every sire and two steaks from each progeny of the DNA sires designated for the sensory panel component of the project. Steaks were shipped to Kansas State University to collect WBSF values and for trained sensory panel evaluation. Steaks measured for shear force were cooked fresh at 14 days post-mortem, whereas sensory panel steaks were frozen and later thawed for trained sensory panel evaluations. Overall tenderness (OT) is a linear function of two sensory panel traits, myofibrillar tenderness (MT) and connective tissue tenderness (CT). Higher scores for all sensory traits, which also include flavor and juiciness, are more desirable. Because WBSF is the force required to shear through cooked meat, higher values are less desirable.

The project was not designed to provide comparisons among breeds and consequently, no valid breed comparisons can be drawn. A breed’s average relative to the overall project average is due to management as much as genetics. Furthermore, some breed associations bred their sires to cows of the same breed, while other breed associations used cows of breeds known to be above average for tenderness. There is not a reasonable statistical approach to adjust for these differences in genetics of the cows.

Phenotypic Results

Analysis of the phenotypic data showed significant variation among all breeds for shear force. Ranges of average shear force values for sires within breeds were from 1.90 lb. to 6.62 lb (Dikeman *et al.*, 2003), indicating that every breed has significant variation in tenderness, and opportunity to improve this trait.

Greater than 8,200 progeny of over 300 sires representing 14 breeds were harvested for collection of carcass and meat quality data. The analysis excluded data from 883 progeny because of incorrect animal or carcass identification. There were 7,319 progeny used in carcass and WBSF analyses and 2,422 progeny with sensory panel data.

Carcass traits of the project cattle were representative of the beef industry with average hot carcass weight of 771 lb, fat thickness of 0.48 in, ribeye area of 13.2 in², yield grade of 2.8, and marbling score of Small²⁰.

Although the cattle were young, mostly from AI sires, and managed optimally, 26% of the steaks had WBSF values > 11.0 lb (considered tough) and 19.4% had sensory panel tenderness scores of < 5.0 (5 = slightly tender; 4 = slightly tough).

Heritabilities and Genetic Correlations

Data from 2,615 progeny of 70 sires were used to estimate heritabilities and genetic and phenotypic correlations (Table 1) using an animal model with relationships among sires (dams were assumed unrelated) in a series of 4-trait analyses. The identity and paternity of these carcasses were verified by DNA marker data. The genetic correlations between WBSF and the sensory panel tenderness scores are highly negative (favorable) and therefore WBSF is a useful measure of tenderness. The genetic correlations between marbling and sensory tenderness are much closer to zero. Furthermore, WBSF is a heritable trait, and hence, it will respond to selection. Therefore, EPDs for WBSF can be computed for all sires in the CMP and can be generated on an ongoing basis if new phenotypic information is generated. Four breeds (Simmental, Simbrah, Shorthorn, and Hereford) have published shear force EPDs mostly based on data collected in this project. The CMP data has augmented the carcass EPDs of many breeds and has allowed one breed, (Maine-Anjou) to publish its first carcass EPDs of any kind.

Table 1. Heritabilities and genetic and phenotypic correlations.

Trait Name	Trt	WBSF	MT	CT	CL	FL	JC	MB	FT	KPH	HCW	REA
Shear Force	WBSF	0.43	-0.99	-0.79	0.41	-0.65	-0.65	-0.56	-0.28	0.47	0.02	0.23
Myofib Tnd	MT	-0.68	0.29	0.92	-0.26	0.79	0.74	0.38	0.14	-0.86	0.20	-0.51
Cn Tiss Tnd	CT	-0.63	0.82	0.25	-0.24	0.74	0.62	0.19	0.22	-0.82	0.38	-0.53
Cooking Loss	CL	0.27	-0.06	-0.04	0.14	-0.16	-0.21	-0.72	-0.16	-0.02	0.16	0.28
Flavor	FL	-0.14	0.24	0.23	-0.03	0.18	0.98	0.35	-0.23	-0.61	-0.17	-0.63
Juiciness	JC	-0.05	0.26	0.14	0.02	0.43	0.29	0.56	-0.11	-0.39	-0.14	-0.66
Marbling	MB	-0.23	0.21	0.13	-0.14	0.14	0.20	0.76	0.20	-0.19	-0.27	-0.36
Fat Thick	FT	-0.10	0.07	0.09	-0.08	0.05	0.02	0.22	0.24	0.37	0.27	-0.19
Internal Fat	KPH	0.01	-0.03	-0.02	-0.04	-0.06	-0.08	0.03	0.13	0.42	0.76	0.37
Hot Carc Wt	HCW	-0.07	0.12	0.15	0.02	0.03	-0.01	0.10	0.31	0.19	0.24	0.24
Ribeye Area	REA	0.07	-0.06	-0.04	0.08	-0.05	-0.08	-0.04	-0.11	0.06	0.41	0.30

Heritabilities are on the diagonals in bold black, genetic correlations are above the diagonals in black, and phenotypic correlations are below the diagonals in blue.

DNA Component of the CMP

The objectives of the DNA component of the CMP were to validate and characterize 11 quantitative trait loci (QTL) for carcass and meat quality traits that were discovered in previous checkoff-funded research at TAMU (the Angleton Project). The Angleton Project used a resource population (Figure 2) comprised of greater than 600 progeny in large, full-sib families (produced by embryo transfer) of a double, reciprocal backcross design between Angus and Brahman.

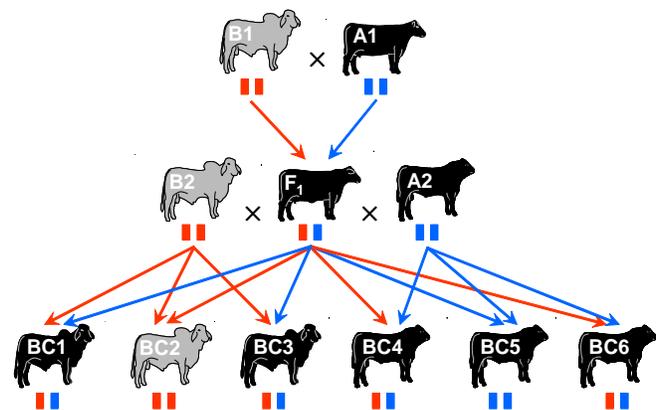


Figure 2. Design of Angleton Project

Validation of QTL discovery projects is necessary because of the substantial risk of false positive results, even in large, well-designed projects. However, failure to validate a QTL does not necessarily imply that the QTL was a false positive; it may simply mean that the QTL was segregating in the resource population used for discovery, but not in the population used for validation.

Characterization of QTL involves determining which QTL are segregating in each breed, how many sires per breed appear to be segregating for each QTL, and which traits are affected by each QTL. In other words, characterization seeks to determine the potential utility of the QTL in genetic improvement programs.

Segregation of QTL in the CMP occurs within paternal half-sib families (Figure 3). Some sires segregate QTL, but many are homozygous at the QTL. The QTL analysis involved 70 sires with 2,615 progeny with DNA marker data and phenotypes in 210 contemporary groups. There were 1,458 progeny with sensory data and DNA marker data.

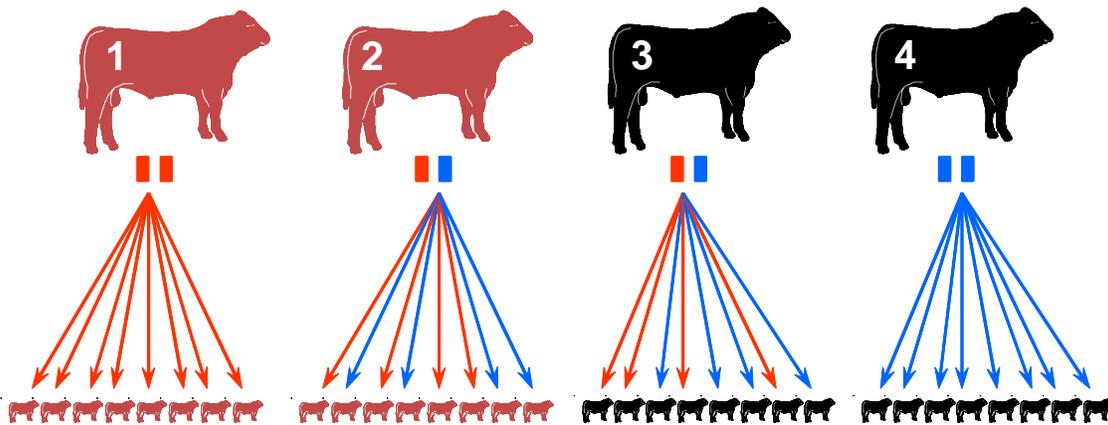


Figure 3. Segregation of QTL in Paternal Half-Sib Families

Individual Sire Analysis

As marker data were collected on the progeny of each sire, an analysis of the sire's progeny was performed and reported to the respective breed association. Figure 4 shows an example of such a report. Each QTL was evaluated for the trait for which it showed the greatest association in the Angleton project. The model used was:

$$y_{ij} = X_j\beta_j + Q_{ij}\alpha_{ij} + e_{ij}$$

where y_{ij} is a vector of observations on the progeny of sire i for trait j , X_j relates observations to contemporary groups, β_j is a vector of fixed contemporary group effects, Q_{ij} is a vector of probabilities that each progeny of i inherited QTL allele A of sire i minus the probability it inherited QTL allele B from its sire, α_{ij} is the fixed within-sire effect of QTL allele A minus the effect of QTL allele B on trait j , and e_{ij} is a vector of residuals.



NCBA CARCASS MERIT PROJECT
PRELIMINARY RESULTS



Sire:
Breed:

Date: 7/10/02

Number of progeny entered: 45
Progeny excluded due to mis-paternity: 3
Progeny excluded due to mis-identity: 4

Total progeny analyzed: 38
Progeny with sensory data: 34

RESULTS OF SEGREGATION ANALYSIS

QTL	TRAIT	P VALUE
1	<u>Wamer-Eratzler</u> Shear Force	P<0.05*
2	<u>Wamer-Eratzler</u> Shear Force	NS
3	<u>Wamer-Eratzler</u> Shear Force	NS
4	<u>Wamer-Eratzler</u> Shear Force	NS
5	<u>Wamer-Eratzler</u> Shear Force	NS
6	<u>Wamer-Eratzler</u> Shear Force	NS
7	Overall tenderness	NS
8	Marbling	NS
9	Marbling	NS
10	Marbling	NS
11	<u>Ribeye</u> Area	NS

Figure 4. Example of an Individual Sire Report

Overall Analysis – Single Trait, Fixed QTL Model

Upon completion of data collection, the entire dataset was analyzed together to determine which QTL were segregating and which traits they influenced. The first model used was:

$$y_j = X_j\beta_j + Z_{sj}s_j + Q_j\alpha_j + e_j$$

where y_j is a vector of observations on trait j , Z_{sj} relates observations to sires, s_j is a random vector of residual polygenic breeding values of sires, Q_j is a block diagonal matrix of the Q_{ij} 's, α_j is a vector of the α_{ij} 's (one element per sire), e_j is a vector of residuals and the remaining terms are as previously defined. Sires were considered unrelated and Q was computed at the QTL position estimated in the TAMU Angleton project. The Q matrices were computed by an extended version of the GenoProb software (Thallman *et al.*, 2001a,b; Thallman *et al.*, 2002).

Table 2 contains significance levels from the fixed QTL analysis. It addresses the question “Does the QTL have an effect on the trait and is it segregating in the set of sires that were sampled in this breed?” A value of + indicates weak evidence ($P < 0.10$) that the QTL affects the trait. A value of ++ indicates moderate evidence ($P < 0.05$) and +++ indicates strong evidence ($P < 0.01$) that the QTL influences the trait. An empty cell indicates that we do not have sufficient evidence to conclude that the QTL influences the trait, but it does not imply that we have evidence that the QTL does not influence the trait (it could be that we simply do not have enough evidence to decide). No QTL should be expected to influence every trait, but we should expect that some QTL will influence several traits. Therefore all QTL were tested for effects on all of the traits.

Table 2. Significance levels from Fixed QTL analysis

Level of Significance	Quantitative Trait Locus										
	1	2	3	4	5	6	7	8	9	10	11
Shear Force						+++		++			
Overall Tend.						++		++			
Myofibrillar Tend.						+		++			
Conn. Tiss. Tend.						+++		+			
Juiciness				+		+++	++			+	
Flavor									+	+	
Marbling											+++
Ribeye Area			+			++	++		+		
Fat Thickness				+++		++					
Hot Carc. Wt.							+++	+++			
KPH Fat		+				++					
Yield Grade	++					+					

+ ($P < 0.10$) ++ ($P < 0.05$) +++ ($P < 0.01$)

Table 3 presents the number of sires with highly significant QTL effects for each trait by QTL combination. It addresses the question “Is a specific sire segregating at the QTL?” and then counts the affirmative answers for each trait. It is the number of sires in the analysis that have strong evidence ($P < 0.01$) of the QTL segregating for the indicated trait. A significant overall result can occur from either a few sires with strong evidence or a larger number of sires with weaker evidence. Therefore, it is useful to look at the results from both perspectives.

Table 3. Number of Highly Significant Sires

No. of Highly Sig. Sires	Quantitative Trait Locus										
	1	2	3	4	5	6	7	8	9	10	11
Shear Force	2	1	1	1	1	6	1	3	1		
Overall Tend.	1	1			1	1	1	2		3	
Myofibrillar Tend.		1			1	1	2	2		1	
Conn. Tiss. Tend.	1	1			1	3	2	1		2	
Juiciness	1	1	2	2		3	1	1	1	2	
Flavor			1		1		1				1
Marbling		1	1	1	2		1	1	1	2	1
Ribeye Area	1		3		1	3	3		1	1	
Fat Thickness	1			3	1	1	1	2	1	2	2
Hot Carcass Wt.	1		2		1		4	4	1	1	3
KPH Fat	1	2		2	2	3				2	1
Yield Grade	2			1	1	2				1	1

Analysis of the marker data with a single trait model fitting QTL as fixed effects validates that at least some of the previous QTL have significant effects, and are good candidates both for marker-assisted selection, and for further study. Most QTL seem to have pleiotropic effects, where the QTL influence two or more traits.

Breed Analysis

Marker data were also analyzed by breed to validate which markers were segregating in each breed, as well as in the overall population. Tables similar to the previous two were constructed for each breed and distributed to the respective breed associations. As expected, significance of marker effects varies widely across breeds, indicating the degree of heterozygosity for each QTL also varies by breed. However, because there were only between one and ten sires per breed, the within-breed analyses are not generally very powerful. More emphasis should be placed on the combined analyses for evaluating the effects of the QTL.

Multiple Trait, Random Regression QTL Model

The single trait analyses presented above involve a very large number of statistical tests. With so many tests, it is expected that a number of false positive results would occur. Because the number of significant results exceeds the expected number of false positives substantially, it appears likely that at least several of the QTL evaluated are segregating in the CMP families. However, it is not clear which ones are real and which are spurious.

There are several sources of information that are not considered by the single trait analysis.

- Most QTL probably influence a number of traits, to different degrees. In fact, due to chance, a real QTL would probably show significance in the single trait analysis for some, but not all, of the traits that it influences. Obviously, those traits that it has the greatest influence on are more likely to show significance.
- If a sire is segregating at a real QTL, then his progeny that inherit QTL allele A instead of allele B should differ for all traits that the QTL influences. The magnitudes of those differences should be in proportion to the degree of influence that the QTL has on each of those traits. Furthermore, if allele A is defined as the allele with favorable effect on the trait most influenced by the QTL, then the directions of differences between alleles A and B for secondary traits should tend to be consistent across sires.
- If a QTL is not real, then in a single trait analysis, we should not be surprised if one set of sires contributes evidence of segregation for one trait and a different set of sires contributes evidence of segregation for another trait. Furthermore, the directions and magnitudes of differences between alleles A and B for secondary traits would likely be inconsistent across sires, although residual correlations will prevent complete independence of these effects.

The single trait analysis cannot take the above factors into account, but these factors are considered by a multiple trait analysis in which the QTL effects are fit as random. Such a model was implemented to help discern the real QTL from the false positives.

Unfortunately, the multiple trait, random QTL model increases the computational requirements by several orders of magnitude and software to implement it is more specialized than that required for the single trait analysis.

Given the simple structure of the CMP pedigree, a random regression model similar to the fixed effects model described previously is a logical choice. The random regression model has essentially the same terms as the fixed effects model:

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{Z}_{sj}\mathbf{s}_j + \mathbf{Q}_j\mathbf{a}_j + \mathbf{e}_j$$

where \mathbf{a}_j is a vector of the random effects of sire QTL allele A minus allele B, nested within sire, and the remaining terms are the same as previously defined. However, in this model, records for t different traits are analyzed simultaneously. The (co)variance of sire effects can be represented as $\boldsymbol{\Sigma}_s \otimes \mathbf{I}$, where $\boldsymbol{\Sigma}_s$ is a matrix of sire (co)variance parameters among traits and the (co)variance of QTL effects, \mathbf{a} , can be represented as $\boldsymbol{\Sigma}_q \otimes \mathbf{I}$, where $\boldsymbol{\Sigma}_q$ is a matrix of QTL (co)variance parameters among traits and \otimes represents the operator for the Kronecker product of two matrices. The (co)variances of the residuals of a progeny with all traits measured are represented as $\boldsymbol{\Sigma}_e$. Sires were considered unrelated.

The size of the mixed model equations (MME) is the same per trait as for the model in which the QTL were considered fixed (210 equations for contemporary groups, 70 for sire breeding values, and 70 for QTL segregation effects, for a total of 350 equations per trait). Because the computations required to invert a matrix are roughly proportional to the cube of its size, two, three, and four trait analyses require roughly 8, 27, and 64, respectively, times as much computation as a single trait analysis. Furthermore, the single trait, fixed QTL analysis required estimation of only two variances, whereas two, three, and four trait analyses require estimation of 9, 18, and 30 (co)variance parameters, respectively. Consequently, the number of iterations required for variance component estimation is considerably greater with more traits.

This model was initially attempted with the Mixed Procedure of SAS. However, it would run only if the residual correlations were constrained to zero ($\boldsymbol{\Sigma}_e$ was forced to be diagonal). This compromise was accepted because those correlations were not of primary interest and were thought to only help to reduce the noise in the analysis. However, this led to results that looked too good to be true being presented at the workshop. It was discovered, after the December 2003 BIF Genetic Prediction Workshop, that covariances among QTL were able to account for some of the covariance among residuals and this biased upward the estimates of QTL variances and the significance of QTL effects. Subsequently, all analyses have included residual correlations.

After the workshop, a group consisting of Steve Kachman, Janice Rumph, Dick Quaas, Rohan Fernando, Dale Van Vleck, Kathy Hanford, Gerhard Moser (Genetic Solutions, Australia), John Pollak, Dan Moser, Elizabeth Dressler, and Mark Thallman met to discuss statistical and computational options for proceeding with the analysis. Steve

Kachman tested MATVEC (Kachman and Fernando, 2002; Wang *et al.*, 2003) and Janice Rumph tested ASREML for analysis of the CMP data with the random regression model. The ASREML software worked well for some trait combinations, but did not provide estimates when marbling was included in the analysis. The MATVEC software (developed jointly by Tianlin Wang, Rohan Fernando, and Steve Kachman) rapidly provided estimates that agreed with estimates Mark Thallman had obtained by doing a Cholesky transformation of the data to account for residual correlations with SAS (an approach that was slow and that complicated the interpretation of the results).

For each QTL, multiple trait analyses were run at one or two cM intervals throughout the region spanned by the marker data to estimate the position of each QTL from the CMP data. All analyses described subsequently are conditional on those estimated positions.

One of the primary questions to be answered in the CMP project was “which of the 11 QTL could be shown to segregate in the U. S. beef cattle population?” When QTL are fit as random effects, the typical approach to hypothesis testing is to use the likelihood ratio test (LRT) to test whether the variance due to the QTL is zero. However, with multiple traits, there were a variety of hypotheses that could have been tested and it was not clear which of them would be most powerful (have the greatest likelihood of detecting a QTL that is real). Furthermore, testing all possible hypotheses would probably decrease the power. The group agreed on sets of two or three biologically related traits that spanned the economically important traits that were measured.

Another challenge in testing hypotheses with multiple trait models is that the test statistics may depart substantially from the textbook distributions, which are based on asymptotic theory. Permutation testing (Churchill and Doerge, 1994) is one way to determine the sampling distribution of the test statistic, but it requires repeating the analysis thousands of times. Therefore, it is only feasible for analyses that can be conducted very rapidly.

Radu Totir (a postdoctoral researcher working with Rohan Fernando at Iowa State) performed multiple trait hypothesis tests and permutation tests using the random regression model in MATVEC, which is feasible for such analyses. The significance levels resulting from those tests are reported in Tables 4 and 5. Significance levels are the probabilities that the variation accounted for by the QTL is due to chance (probability of obtaining a spurious result). Therefore, smaller numbers indicate stronger evidence supporting the effect of the QTL on a trait.

Several QTL showed significant effects for two or more traits. QTL 6 had significant effects on shear force, overall tenderness, and ribeye area. QTL 7 was significant for ribeye area, hot carcass weight, and juiciness. QTL 8 had significant effects on shear force, overall tenderness, ribeye area, hot carcass weight, and flavor. QTL 10 had significant effects on overall tenderness and juiciness. In addition, QTL 4 and 5 had significant effects on fat thickness, and QTL 11 had noteworthy effects on marbling.

Table 4. χ^2 significance levels from multiple trait hypothesis tests^{a,b}.

Trait	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
Shear Force	0.50	0.50	0.50	0.32	0.25	0.0003	0.50	0.02	0.16	0.49	0.50
Overall Tnd	0.50	0.50	0.50	0.37	0.44	0.0007	0.50	0.02	0.49	0.003	0.50
Fat Thick	0.10	0.50	0.48	0.08	0.05	0.12	0.20	0.50	0.10	0.34	0.33
Marbling	0.15	0.40	0.44	0.31	0.32	0.11	0.12	0.50	0.06	0.37	0.007
Fat Thick	0.19	0.34	0.50	0.15	0.12	0.23	0.32	0.50	0.17	0.38	0.40
Marbling	0.17	0.41	0.48	0.31	0.21	0.16	0.11	0.50	0.11	0.45	0.01
Internal Fat	0.50	0.08	0.50	0.38	0.16	0.47	0.49	0.50	0.44	0.43	0.44
Ribeye Area	0.50	0.50	0.18	0.44	0.32	0.05	0.009	0.05	0.50	0.50	0.50
Hot Carc Wt	0.50	0.50	0.19	0.47	0.50	0.47	0.01	0.003	0.50	0.50	0.49
Juiciness	0.50	0.23	0.40	0.18	0.24	0.10	0.10	0.50	0.50	0.12	0.38
Flavor	0.50	0.34	0.47	0.26	0.22	0.31	0.28	0.02	0.50	0.10	0.34
Overall Tnd	0.50	0.38	0.42	0.41	0.49	0.02	0.48	0.03	0.50	0.24	0.49
Flavor	0.50	0.45	0.50	0.50	0.28	0.30	0.34	0.03	0.32	0.11	0.33
Overall Tnd	0.50	0.29	0.50	0.50	0.45	0.01	0.35	0.02	0.50	0.24	0.50
Juiciness	0.50	0.33	0.46	0.11	0.50	0.11	0.05	0.49	0.50	0.06	0.41
Overall Tnd	0.50	0.37	0.49	0.31	0.50	0.01	0.45	0.09	0.50	0.14	0.46

^aTwo or three traits were analyzed together as indicated by the groupings in the table. For each cell, the hypothesis tested was that the variance in the indicated trait (and associated covariances) due to the corresponding QTL was zero (with the remaining QTL variances and covariances included in the model).

^bOne additional analysis was run for QTL 8: Shear Force (P<0.03), Overall Tnd (P<0.01), and Hot Carc Wt (P<0.0005).

Table 5. Significance levels from multiple trait permutation tests^a.

Trait	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
Shear Force						0.008		0.040		n.t.	
Overall Tnd						0.001		0.030		0.004	
Fat Thick				0.043	0.030	0.180			0.200		n.t.
Marbling				n.t.	n.t.	0.110			0.079		0.002
Ribeye Area						0.011	0.008	0.037			
Hot Carc Wt						n.t.	0.006	0.002			
Flavor								0.022			
Overall Tnd								0.015			
Juiciness						0.107	0.024			0.050	
Overall Tnd						0.021	n.t.			n.t.	

^aTwo traits were analyzed together as indicated by the groupings in the table. Only selected hypotheses were tested by permutation. Cells containing "n.t." indicate that a QTL effect for that trait was not included in the analysis; the hypothesis for the other trait is that the QTL influences that trait (1 df).

The analyses were performed in groups of two or three traits, as indicated in Table 4. The hypotheses tested were that the QTL variance (and the associated covariances) for a specific trait were zero. Thus, the full model included QTL (co)variance parameters for all traits in the model, and the null hypothesis model included (co)variance parameters for all traits, except the one being tested. The significance levels in Table 4 are “textbook” values, computed from a χ^2 distribution with degrees of freedom equal to the number of parameters constrained (number of traits) and the resulting probability divided by two to account for the null hypothesis being on a boundary of the parameter space (Self and Liang, 1987; Littell et al., 1996).

As is evident from Table 4, significance levels can differ considerably, depending on the set of traits included in the model. For example, overall tenderness is included in three different two-trait analyses in Table 4. In the cases of QTL 6 and QTL 8, overall tenderness was most significant when paired with another trait with a significant effect on that trait. However, in the case of QTL 10, overall tenderness was most significant when paired with the trait (shear force) that showed the least evidence of being affected by QTL 10.

Table 4 also demonstrates that significance levels can differ substantially when a third trait is added to a two-trait analysis, depending on whether the QTL affects the third trait, or not. In the case of fat thickness and marbling, adding internal fat to the model decreased levels of significance for the first two traits rather uniformly. However, when hot carcass weight was added to the analysis of shear force and overall tenderness for QTL 8 (results not shown), the p-value for shear force increased slightly to 0.03, for overall tenderness decreased slightly to 0.01, and for hot carcass weight decreased substantially to 0.0005.

Tests were also performed for single trait analyses using the random regression model (results not shown). They followed the same general pattern as the single trait fixed QTL tests presented in Table 2.

Permutation tests were performed to obtain more reliable significance levels where the values in Table 4 appeared to at least approach significance. Table 5 contains significance levels from 2,000 permutations for selected hypotheses. A small proportion of the permutations failed to converge and were discarded. In most cases, the values in Table 4 were more significant for a set of two-trait analyses than for the three-trait analyses. Consequently, all of the permutation tests were for two-trait analyses. Where values are presented for both traits in a pair, the hypothesis being tested is the same as was described for Table 4.

However, when one trait in a pair is unaffected by the QTL, the power of detecting the QTL for the second trait is improved by dropping the QTL effect for the first trait from both the alternative and null hypotheses, resulting in a one degree of freedom test for the second trait. In this case, the effect of the QTL on the first trait is not tested. Where this approach was used in Table 5, a significance level is given for the second trait and the first trait is indicated by n.t. (not tested).

More impressive significance levels could undoubtedly have been obtained by testing additional trait combinations, especially trait combinations suggested by preliminary analyses. However, such significance levels would need to be adjusted for the multiple (and perhaps selective) tests of hypotheses that would not be independent. It was not feasible to use permutation testing to obtain appropriate significance thresholds for such an approach.

Nonetheless, an approach that considers the correlated effects of a QTL on a subset (determined by the data) of the traits measured seems to offer the potential for increased power, provided that appropriate significance thresholds can be established. Much work remains to be done on QTL analyses with true multiple trait models.

Multiple Trait, Random, Gametic QTL Model

Another of the primary results of a random QTL analysis is the amount of variance in phenotypes that can be explained by the QTL. This can be most easily interpreted as the proportion of phenotypic variance explained by the QTL, just as heritability is the proportion of phenotypic variance explained by breeding value. In a sire model, the variance due to breeding value is estimated by multiplying the estimate of sire variance by four because every calf inherits exactly half of its breeding value from its sire.

Unfortunately, with the random regression model, the transformation from the QTL variance parameter to the phenotypic variance explained by the QTL is not so simple, because the values in \mathbf{Q} for each calf differ, depending on the amount of information contributed by the DNA markers. Calves with informative markers on both sides of and close to the QTL position have \mathbf{Q}_{ijk} very close to either +1 or -1 (are essentially fully informative, but calves with no marker information have $\mathbf{Q}_{ijk} = 0$ (are completely uninformative) and yet other calves have an informative marker on only one side of, and at varying distances from, the QTL position. The latter calves have intermediate values of \mathbf{Q}_{ijk} and are partially informative. While every calf in the DNA analysis had marker data at most of the QTL, very few of them were completely informative at every QTL.

A fully informative calf has a residual that is “truly residual” to the QTL model, but a completely uninformative calf has a residual variance that consists of the “true residual variance” plus variance due to segregation at the QTL (expected to be half of the phenotypic variance accounted for by the QTL). Partially informative calves have residual variances somewhere in between. Therefore, calves differ in the proportion of phenotypic variance accounted for by the QTL when the random regression model is used.

Furthermore, the relationship between the QTL variance parameter and phenotypic variance explained by the QTL changes with the position of the QTL. This is most evident when there is marker data on only one side of the QTL. The estimated QTL variance increases as the putative distance between the QTL and the markers increases, but the phenotypic variance explained by the QTL must remain the same. A similar phenomenon is well known in the analysis of QTL as fixed effects: as the putative distance between a QTL and markers to one side increases, the magnitude of

the QTL effect estimate increases. It is not possible to distinguish between a modest QTL close to the markers and a large one far away.

Consequently, the random regression model is not well-suited to estimation of the phenotypic variance accounted for by the QTL. However, the gametic model proposed by Fernando and Grossman (1989) is well-suited to this problem because the QTL effect for each progeny is considered in the model, instead of only the sire's expected contribution to each of his progeny at the QTL, as in the random regression model. The multiple trait gametic model is:

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{Z}_{uj}\mathbf{u}_j + \mathbf{Z}_{uj}\mathbf{v}_{mj} + \mathbf{Z}_{uj}\mathbf{v}_{pj} + \mathbf{e}_j$$

where \mathbf{Z}_{uj} relates observations to individuals, \mathbf{u}_j is a random vector of residual polygenic breeding values of the individual, \mathbf{v}_{mj} and \mathbf{v}_{pj} are random vectors of maternal and paternal, respectively, gametic values of the individual at the QTL, and the remaining terms are the same as previously defined. The (co)variance of QTL effects, can be represented as $\boldsymbol{\Sigma}_q \otimes \mathbf{G}$, where $\boldsymbol{\Sigma}_q$ is a matrix of QTL (co)variances among traits. Sires were considered unrelated.

Because the segregation of QTL alleles from sire to progeny is accounted for in \mathbf{v}_{pj} instead of in \mathbf{e}_j , a clean estimate of the phenotypic variance accounted for by the QTL is available. It is equal to twice the appropriate diagonal element of $\boldsymbol{\Sigma}_q$.

The size of the mixed model equations (MME) is considerably greater for the gametic model. There are 2,615 progeny, 2,615 dams, and 70 sires (5,300 individuals) in each of \mathbf{u}_j , \mathbf{v}_{mj} , and \mathbf{v}_{pj} , plus 210 contemporary group equations, for a total of 16,110 equations per trait.

These data were analyzed in a set of four-trait analyses using an extended version of the MTDFREML software package (Boldman *et al.*, 1995). Because MTDFREML uses sparse matrix techniques, the computational requirements do not increase as rapidly with size of the MME as was indicated previously, but the large number of (co)variance parameters to be estimated (30) requires that a large number of iterations be performed. Additional analyses were performed with modified starting values for subsets of parameters to ensure that convergence was achieved. The computational demands of this approach currently preclude its use in permutation testing.

Table 6 contains the estimated percentages of phenotypic variance accounted for by each of the 11 QTL for each trait. The amount of variance explained indicates the magnitude and practical significance of the QTL effect. As is the case for all analyses reported herein, each QTL was analyzed separately; no multiple QTL analyses have been performed. Values of at least 5% are indicated in blue. We should expect much of the variation in truly quantitative traits to be accounted for by a number of QTL that each account for a relatively small proportion of the variation, especially when evaluating data that are pooled over 13 breeds.

As expected, most of the QTL with significant evidence of segregation account for some of the variance in a number of traits. Typically, one or a few closely related traits will be most influenced by a QTL and a number of other traits will be influenced to a lesser extent.

Table 6. Percentage of phenotypic variance accounted for by each QTL.

Trait	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
Shear Force	1	0	1	3	3	12	0	6	3	1	0
Overall Tnd	0	1	1	1	0	10	2	8	0	4	0
Myofib Tnd	1	2	0	1	0	9	2	8	0	4	0
Cn Tiss Tnd	3	1	0	0	0	12	4	8	0	2	0
Cooking Loss	4	2	5	0	3	1	0	0	2	2	1
Flavor	1	1	0	3	2	1	3	3	2	5	3
Juiciness	0	4	1	6	0	6	7	0	0	5	3
Marbling	1	2	2	2	1	4	4	1	4	1	8
Fat Thick	3	1	3	5	6	4	3	2	2	2	2
Internal Fat	0	7	0	3	5	2	0	1	2	1	1
Hot Carc Wt	1	0	2	0	0	2	6	10	0	1	3
Ribeye Area	0	0	4	2	3	7	7	3	0	1	1

Correlations Among QTL Effects on Multiple Traits

Whenever QTL appear to have effects on multiple traits, it is useful to know whether the allele that is favorable for one trait is favorable or unfavorable for others, or whether there are antagonistic relationships that decrease the difference in net merit between the genotypes of the QTL. To answer this question, correlations among the effects of QTL 6, 7, and 8 are reported in Tables 7, 8, and 9, respectively. These correlations were obtained from a series of 19 four-trait analyses for each of the QTL and each correlation represents the average of from 1 to 6 analyses in which that correlation was estimated. The proportions of phenotypic variance accounted for by the QTL in these tables are each averages of five to seven four-trait analyses.

If a QTL has pleiotropic effects on multiple traits, then it should be expected to influence all of those traits in the progeny of sires that are heterozygous for the QTL and no traits in progeny of sires that are homozygous for the QTL. Furthermore, assuming pleiotropy, the directions of QTL effects should be consistent across heterozygous sires and the magnitudes of effects should at least be proportional, with respect to traits, across sires. If these conditions are met, then the expected values of the correlations of QTL effects among the traits will be either 1 or -1.

Not all genetic correlations are due to pleiotropy; some certainly are due to multiple, linked genes that affect different traits. Therefore, we should not expect all correlations among QTL effects to be 1 or -1, but it does seem reasonable to expect such correlations to tend to be of large absolute value. The effect of the QTL correlations is taken into account in multiple trait hypothesis tests and, thus, the correlations themselves do not directly provide additional information from which to decide which

QTL are most likely to be real. However, the QTL correlations may help to explain why multiple trait tests differ from each other or from the single trait tests.

Table 7. Correlations among effects of QTL 6 (proportion of phenotypic variance accounted for by QTL 6 on the diagonals).

Trait	Trt	WBSF	OT	MT	CT	CL	FL	JC	MB	FT	KPH	HCW	REA
Shear Force	WBSF	0.12											
Overall Tnd	OT	-0.96	0.10										
Myofib Tnd	MT	-0.99	1.00	0.09									
Cn Tiss Tnd	CT	-0.89	0.95	0.96	0.12								
Cooking Loss	CL	0.67	-0.86	-0.84	-0.76	0.01							
Flavor	FL	0.31	0.62	0.32	0.57	-0.57	0.01						
Juiciness	JC	0.11	0.04	0.04	0.27	-0.74	0.79	0.06					
Marbling	MB	0.23	-0.57	-0.41	-0.62	0.90	-0.85	-0.24	0.04				
Fat Thick	FT	-0.40	0.43	0.28	0.03	0.18	0.85	0.70	0.84	0.04			
Internal Fat	KPH	-0.79	0.93	0.99	0.70	-0.90	-0.43	-0.66	-0.19	0.41	0.02		
Hot Carc Wt	HCW	-0.23	0.18	0.02	-0.16	-0.51	-0.35	-0.45	0.87	0.54	1.00	0.02	
Ribeye Area	REA	0.17	0.09	0.14	0.06	0.47	-0.57	-0.77	0.35	0.91	0.21	-0.02	0.07

In the case of QTL 6, all of the correlations among WBSF and sensory tenderness are strong and in the favorable directions. This means the allele that increases shear force decreases tenderness score, where higher tenderness scores indicate greater tenderness. Therefore, selecting for the favorable allele at QTL 6 for shear force will also improve overall tenderness. This, together with the proportion of phenotypic variance (9-12%) accounted for by QTL 6, supports the conclusion that the effects of QTL 6 on tenderness are real, as indicated by the significance levels reported in Tables 4 and 5. This also suggests that all of these effects are likely due to pleiotropic effects of the same polymorphism(s) in a single gene.

Correlations of the effect of QTL 6 on flavor (FL) and juiciness (JC) with tenderness traits were weak, but in the favorable direction. Correlations of internal fat (KPH), fat thickness (FT), and marbling (MB) with tenderness were generally antagonistic.

The estimate that QTL 6 accounts for 7% of the phenotypic variance ($P = 0.011$) in ribeye area (REA) is large enough to be quite interesting. However, the correlations of the effect of QTL 6 on REA with its effects on the tenderness traits are very weak. This

Table 8. Correlations among effects of QTL 7 (proportion of phenotypic variance accounted for by QTL 7 on the diagonals).

Trait	Trt	WBSF	OT	MT	CT	CL	FL	JC	MB	FT	KPH	HCW	REA
Shear Force	WBSF	0.00											
Overall Tnd	OT	-0.99	0.02										
Myofib Tnd	MT	-0.99	1.00	0.02									
Cn Tiss Tnd	CT	-1.00	1.00	0.99	0.04								
Cooking Loss	CL	0.41	-0.33	-0.27	-0.98	0.00							
Flavor	FL	0.98	0.83	0.76	0.84	0.99	0.03						
Juiciness	JC	0.99	0.57	0.60	0.58	0.99	0.97	0.07					
Marbling	MB	-0.84	0.75	0.50	0.70	0.96	0.73	0.55	0.04				
Fat Thick	FT	0.95	1.00	0.92	1.00	0.99	1.00	1.00	0.81	0.03			
Internal Fat	KPH	0.95	1.00	1.00	1.00	0.99	0.99	1.00	0.54	0.98	0.00		
Hot Carc Wt	HCW	-0.98	0.48	0.24	0.60	-0.99	-0.23	-0.23	0.70	0.43	0.20	0.06	
Ribeye Area	REA	-0.99	0.20	-0.09	0.34	-0.99	-0.52	-0.52	0.56	0.12	-0.13	0.96	0.07

would suggest that there may be another gene, in the same general region of the genome as, but in linkage equilibrium with, the gene influencing tenderness. If this is the case, it should be possible to select for favorable effects of both genes. The lack of correlation between the QTL effect on REA and hot carcass weight (HCW) is puzzling, especially in light of the positive remaining correlations among these traits, FT and KPH.

For QTL 7, the significant effects of increased REA and HCW were the result of the same allele, but the allele that increased those traits tended to decrease JC and FL. This allele tended to improve MB and tenderness, although the proportion of variance in tenderness traits accounted for by QTL 7 was low.

Table 9. Correlations among effects of QTL 8 (proportion of phenotypic variance accounted for by QTL 8 on the diagonals).

Trait	Trt	WBSF	OT	MT	CT	CL	FL	JC	MB	FT	KPH	HCW	REA
Shear Force	WBSF	0.06											
Overall Tnd	OT	-0.97	0.08										
Myofib Tnd	MT	-1.00	1.00	0.08									
Cn Tiss Tnd	CT	-0.81	0.97	0.99	0.08								
Cooking Loss	CL	0.90	-0.14	0.00	0.46	0.00							
Flavor	FL	-1.00	1.00	1.00	0.93	-0.99	0.03						
Juiciness	JC	-0.99	0.93	0.99	0.67	-0.99	0.83	0.00					
Marbling	MB	-0.99	0.95	0.87	0.67	-0.92	0.95	0.99	0.01				
Fat Thick	FT	-0.35	0.67	0.15	0.67	-0.58	1.00	0.99	0.78	0.02			
Internal Fat	KPH	0.69	-0.44	0.24	-0.63	0.85	-1.00	-0.99	-0.47	-0.87	0.01		
Hot Carc Wt	HCW	-0.49	0.66	0.61	0.70	-0.14	0.61	-0.60	0.97	0.18	0.01	0.10	
Ribeye Area	REA	-0.81	0.86	0.88	0.91	-0.52	0.55	-0.65	0.98	-0.30	0.48	0.93	0.03

It was especially encouraging that the correlations among effects of QTL 8 were all favorable. The allele that decreased shear force improved overall tenderness, as expected, but was also associated with increased flavor, increased carcass weight, and increased ribeye area.

The effect of QTL 10 on JC is favorably correlated (results not shown) with effects on FL (0.73), MB (0.73), CT (0.86), OT (0.71), and MT (0.61). QTL 10 accounts for 2% or less of phenotypic variances in the remaining traits. The effect of QTL 11 on MB is favorably correlated (0.63) with its effect on JC. The effects of QTL 11 on remaining traits are either weakly correlated with its effect on MB or account for 1% or less of phenotypic variance. The effect of QTL 5 on FT is uncorrelated (-0.05) with its effect on KPH.

There are inconsistencies in the correlations in Tables 7-9 due to the combinations of four-trait analyses that were used, along with the variable number of records per trait. A single 12-trait analysis for each QTL would have provided more consistent correlations, but assuring convergence of four-trait analyses (each with 30 (co)variance parameters) is sufficiently challenging for today's software.

The amount of residual variance associated with phenotypes for economically important traits is one of the most limiting factors in determining which locations in the genome influence those traits. Considering observations for a large number of traits simultaneously, especially when some of the traits are inexpensive to measure, is likely to be an effective way to reduce the impact of this limitation. Therefore, it seems that

statistical developments, both computational and theoretical, in the area of multiple trait analysis of QTL should be an area of emphasis for the next several years.

Interpretation of the results presented here should be made in the context of both the amount of statistical evidence obtained for each QTL on the various traits and the amount of evidence in the literature from other studies.

How Can Cattle Breeders Use the Results?

The most direct and immediate way is for breed associations to compute and publish EPDs for shear force and sensory traits from the data generated by the CMP.

Use of the DNA results is contingent on a partner commercializing tests based on the QTL. This could be done either in the form of direct tests or linked markers.

The existing linked markers could be used to select among progeny and grandprogeny of the 70 legacy bulls that were evaluated in the DNA component of the CMP. While this may seem to be a small number of bulls, these 70 bulls were very influential in their respective breeds and have produced hundreds of thousands of progeny and even larger numbers of grandprogeny.

Linked markers could be commercialized quickly with relatively little development cost and could be used to improve accuracy of selection among progeny of the CMP sires. The technology would probably be used effectively by only a small proportion of the breeders in any breed, but the improved selection response in those herds would likely benefit the entire breed. Some additional development of statistical/computational methods would be required to include marker information in national cattle evaluation.

This approach would also require continued collection of phenotypes and marker data on progeny groups for the approach to be sustainable long-term. However, fewer phenotypes would be required than without the markers and accurate genetic evaluations could be obtained earlier in life (prior to breeding decisions).

The linked markers from the CMP could be used effectively in intensive breeding programs for tenderness, as shown in Figure 5. Young bulls would be progeny test-

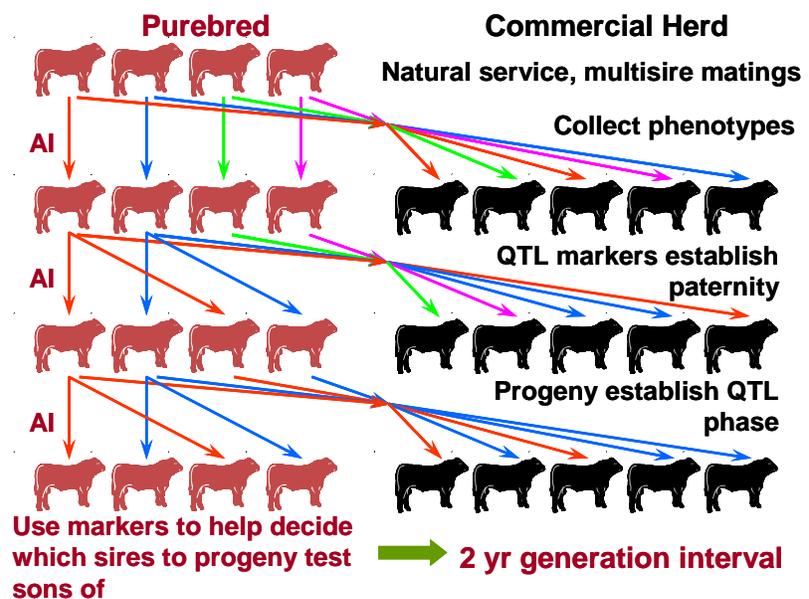


Figure 5. Progeny Testing Scheme

ed in multiple sire matings to commercial cows and, at the same time, would be mated to seedstock cows to produce the next generation of herd sire candidates. Some of the CMP markers would be used to determine paternity of the multiple sired calves. This would not necessarily be any more expensive than a paternity test based on anonymous DNA markers and should be cost-competitive with progeny testing by artificial insemination (AI) matings. The QTL effects of the sires for the markers used in paternity testing could be estimated at no additional cost and those estimates could be used with marker data on seedstock progeny of the tested sires to select the next generation of bulls to be progeny tested. Provided that phenotypes could be collected by about 14 months of age, this approach would allow marker assisted progeny testing at a two-year generation interval with the existing markers.

The Next Steps for Application of the CMP Results

Although there are several scenarios under which the CMP QTL could be used as linked markers, most commercial interest is in association (linkage disequilibrium) or functional tests. Therefore, the most promising QTL should be converted into association tests based on SNP. As illustrated in Figure 6, the CMP population and DNA samples could be an important resource to aid in converting the QTL (using Q6 as an example) into association tests as follows:

1. Identify positional candidate genes under QTL6.
 2. Rank sires by evidence of segregation at QTL6.
 3. Sequence portions of positional candidate gene(s) in top 8 sires.
 4. The best SNPs are heterozygous in the greatest number of top sires.
 5. Score those in remaining sires and test concordance with QTL6.
- The SNP with greatest concordance could then be used in commercial DNA tests following validation in other populations.

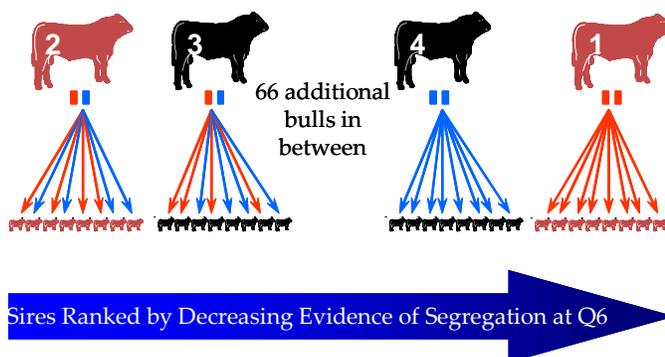


Figure 6. CMP population is a valuable resource for conversion of QTL into Association Tests

Conclusions

The primary objectives of the NCBA Carcass Merit Project were to collect data for carcass merit EPDs, including tenderness, and to attempt to validate previously discovered QTL for carcass merit in the U.S. cattle population. Both of those objectives were accomplished, but much work remains to be done in this area.

Five of the QTL (6, 7, 8, 10, and 11) show solid evidence of segregation in relevant beef industry populations. Two additional QTL (4 and 5) appear worthy of further investigation. The remaining QTL (1, 2, 3, and 9) did not show significant effects on any of the traits evaluated in the study. However, it is possible that the latter QTL have

unique alleles in Brahman and non-Brahman populations. Unfortunately, only six Brangus or Simbrah sires were analyzed in CMP, with a total of 161 progeny.

Besides the stated objectives, several other benefits have resulted from the Carcass Merit Project, both tangible and intangible. The project perhaps represents the greatest cooperative effort ever among U.S. beef breed associations. Experiences gained and goodwill generated in this project will allow further cooperative research by breeds, which will benefit the entire beef industry. The project has also raised the visibility of marker-assisted selection and genomics in the beef industry. The considerable publicity received and educational efforts undertaken by the Carcass Merit Project have moved the industry closer to embracing selection aided by DNA tests, and have improved the understanding of issues with these technologies. In addition, the project has revealed the considerable cost and coordination required for industry-wide tenderness data collection. Furthermore, the CMP has resulted in greater understanding of, and development of methods to address, statistical issues in the validation of quantitative trait loci in populations representative of the U.S. beef industry.

The most significant result of the Carcass Merit Project is the sizeable database of phenotypic information and DNA samples collected from a wide cross section of U.S. beef germplasm. Already, data and samples stored by breed associations are being used to validate DNA tests marketed to U.S. cattle producers. These resources could be extremely valuable tools for converting QTL (both those developed in the Angleton project and in public research) into more easily used association or functional tests. Having a large unbiased resource population, representative of the U.S. beef cattle population, justifies the industry's investment in this project, and stands to be the project's greatest legacy.

Literature Cited

- Churchill, G. A. and R. W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* 138:963-971.
- Boldman, K. G., L. A. Kriese, L. D. Van Vleck, C. P. Van Tassell, and S. D. Kachman. 1995. *A Manual for Use of MTDFREML, a Set of Programs to Obtain Estimates of Variances and Covariances [Draft]*. USDA, ARS, Clay Center, NE.
- Dikeman, M. E., E. J. Pollak, J. Taylor, S. Davis, T. Holm, S. Koontz, C. Gill, D. Moser, and E. W. Dressler. 2003. Warner-Bratzler shear force values and ranges of steaks from cattle of known sires. *Kansas Agr. Exp. Sta. Report of Progress* 908:24-28.
- Fernando, R. L., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. *Genet. Sel. Evol.* 21:467-477.
- Kachman, S. D. and R. L. Fernando. 2002. Analysis of generalized linear mixed models with MATVEC. *Proc. 7th World Congress on Genetics Applied to Livestock Production*, Montpellier, France, August 19-23, 2002, CDROM communication no. 28-04.
- Littell, R. C., G. A. Milliken, W. W. Stroup, and R. D. Wolfinger. 1996. SAS System for Mixed Models. p. 44.

- Self, S. G. and K. Y. Liang. 1987. Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *J. Am. Stat. Assn.* 82:605-610.
- Thallman, R. M., G. L. Bennett, J. W. Keele, and S. M. Kappes. 2001a. Efficient computation of genotype probabilities for loci with many alleles: I. Allelic peeling. *J. Anim. Sci.* 79:26-33.
- Thallman, R. M., G. L. Bennett, J. W. Keele, and S. M. Kappes. 2001b. Efficient computation of genotype probabilities for loci with many alleles. II. Iterative method for large, complex pedigrees. *J. Anim. Sci.* 79:34-44.
- Thallman, R. M., D. B. Cooke, and G. L. Bennett. 2002. GenoProb: Computation of genotype and grandparental origin probabilities in complex pedigrees with missing marker data. *Proc. 7th World Congress on Genetics Applied to Livestock Production*, Montpellier, France, August 19-23, 2002, CD-Rom Communication No. 28-09.
- Wang, T., Fernando, R. L., and Kachman, S. D. (2003) Matvec Users' Guide Version 1.03.

MARKER- AND GENE-ASSISTED SELECTION IN LIVESTOCK

Jack C. M. Dekkers

**Department of Animal Science and Center for Integrated Animal Genomics
Iowa State University, Ames, IA, 50011**

Introduction

Over the past decades, substantial advances have been made in the identification of loci and chromosomal regions that contain loci that affect traits of importance in livestock production. This has enabled programs for marker-assisted selection and gene introgression. Objectives of this paper are to give examples of genes and markers used in livestock breeding programs and to review and discuss strategies and opportunities for the use of genes or markers in genetic improvement.

Types of genetic markers

Application of molecular genetics for genetic improvement relies on the ability to genotype individuals for specific genetic loci. For these purposes, three types of observable genetic loci can be distinguished:

- 1) Direct genes: loci for which the causative polymorphism can be genotyped.
- 2) LD markers: in population-wide linkage disequilibrium with the causative mutation.
- 3) LE markers: in population-wide linkage equilibrium with the causative mutation.

As described by Anderson (2001), LE markers can be readily detected on a genome-wide basis using breed crosses or analysis of large half-sib families, requiring only sparse marker maps (20 cM spacing). LD markers are by necessity close to the functional mutation (within 1 to 5 cM) and their identification requires candidate gene (Rothschild and Soller, 1997) or fine-mapping approaches (Anderson, 2001). Functional mutations are most difficult to detect and few examples are available (Anderson, 2001).

The three types of loci differ in methods of detection and their application in selection; whereas direct and LD markers allow selection on genotype across the population, use of LE markers must allow for different marker-QTL linkage phases from family to family. Thus, their ease and ability for use in selection is opposite to their ease of detection. In what follows, selection on markers will be referred to as gene assisted selection (GAS), LD markers assisted selection (LD-MAS), and LE marker assisted selection (LE-MAS).

Traits with MAS application

Molecular markers have been used to identify loci or chromosomal regions that affect single gene traits and quantitative traits. Single gene traits include genetic defects, disorders and appearance. Quantitative traits can be categorized into a) routinely

recorded traits, which be further subdivided into traits that are recorded on both sexes, sex-limited traits, and traits that are recorded late in life, b) difficult to record traits (feed intake, product quality), and c) unrecorded traits (disease resistance). The ability to detect QTL decreases in the order a), b), c) because of the availability of phenotypic data. For a similar reason, genome scans, which require more phenotypic data, are often used to detect QTL for traits in category a), whereas candidate gene approaches are more often used to identify QTL for traits that are not routinely recorded (b and c). Potential extra genetic gains from MAS or GAS are greatest for traits in category c) and lowest for traits in category a), in particular for traits that are routinely recorded on both sexes prior to selection, in inverse proportion to the ability to make genetic progress using conventional methods (e.g. Meuwissen and Goddard, 1996).

Examples of genes and markers in, or available for, commercial use

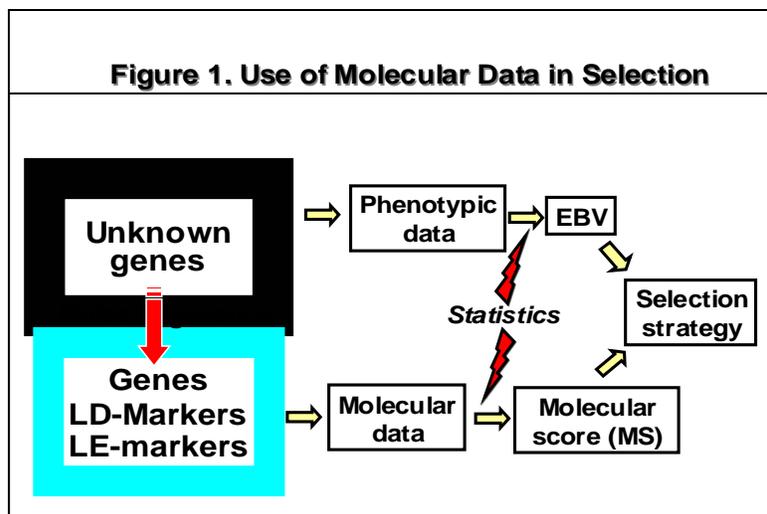
Although there is a large number of scientific reports on detection of QTL and genes of importance to livestock, most of these were identified in experimental populations and on an experimental basis. Published reports on the use of genes or markers in commercial livestock programs are lacking. Nevertheless, a substantial number of gene or marker tests are currently available on a commercial basis. A non-exhaustive summary is given in Table 1, with tests categorized by the type of trait and the type of application. Several marker tests are used for within-house selection only (e.g. the PICmarqTM markers), whereas others are available through commercial genotyping services. To date, the majority of publicly available tests are for GAS or LD-MAS. Use of LE-MAS is primarily limited to dairy cattle, although the extent of use of LE-MAS in this or other species is unclear because the main use is on an in-house basis. The use of LE-MAS in dairy cattle stems from the extensive use of the grand-daughter design for QTL detection in dairy cattle. In contrast to the breed cross designs that are used in other species, this design focuses on detection of QTL that segregate within a breed and is facilitated by the large progeny-testing programs in dairy cattle.

Strategies for the use of genes and markers in selection

Once genes or markers linked to QTL have been identified, their effects can be estimated based on associations between phenotype and genotype (Fernando, 2003). Resulting estimates can be used to assign a 'molecular score' to each selection animal, which can be used to predict the genetic value of the individual and used for selection (Figure 1). The constitution and method of quantification of the molecular score depends on type of LD that is used and on how the marker will be used in selection. In addition to a molecular score, individuals can also obtain a regular estimate of the breeding value for the collective effect of all the other genes (polygenes) on the trait.

Table 1. Examples of gene tests used in commercial selection programs by type of trait and test.

Trait category	Direct gene	LD-marker	LE-marker
Congenital defects	BLAD (dairy) DUMPS (dairy) CVM (dairy) RZR (pig)		
Coat color	cKIT (pig) MC1R (pig, cattle) MGF (beef) MGH (cattle)		
Horns			Polled (cattle)
Milk quality	Caseins (dairy) β -lac (dairy)		
Meat quality	RZR (pig) RN (pig)	FABP (pig) RN (pig) CAST (pig, cattle) >15 PICmarq™ (pig) THYR (cattle) Leptin (cattle)	
Feed intake	MC4R (pig)		
Disease	Prp (sheep) F18 (pig) K88 (pig)	B blood (poultry)	
Reproduction	Booroola (sheep) Inverdale (sheep)	ESR (pig) PRLR (pig)	
Growth & composition	MC4R (pig) IGF-2 (pig) Mst (beef) Cpg (sheep)	CAST (pig) IGF-2 (pig) Calpain (beef) Carwell (sheep)	[QTL (pig)] [QTL (beef)]
Milk yield&composition	DGAT (dairy) GRH (dairy) Caseins (dairy) β -lac (dairy)	PRL (dairy)	QTL (dairy)



The following four selection strategies can then be distinguished:

- 1) Select on the molecular score alone.
- 2) Tandem selection, with selection on molecular score, followed by selection on a phenotype-based EBV.
- 3) Selection on an index of the molecular score and EBV.
- 4) Pre-selection on molecular score (or an index of molecular score and phenotype-based EBV) at an early age, followed by selection on a phenotype-based EBV at a later age.

Selection on molecular score alone ignores information that is available on all the other genes (polygenes) that affect the trait and is expected to result in the lowest response to selection. This strategy does, however, not require additional phenotypes, other than those that are needed to estimate marker-effects, and can be attractive when phenotype is difficult or expensive to record (e.g. disease traits, meat quality, etc.). If both phenotypic and molecular information is available on selection candidates, index selection is expected to result in greater response to selection than tandem selection. The reason is similar to why two-trait selection using independent culling levels is expected to give lower multiple-trait response than index selection; two-stage selection does not select individuals for which a low molecular score may be compensated by a high phenotype-based EBV.

The above comparison applies not only to QTL for quantitative traits but also to selection on genes or markers associated with single gene traits. Selection on such loci will reduce selection pressure on and response in other economic traits. Thus, genotypes for single gene traits should also be incorporated into a total merit index for selection, which can be accomplished by assigning them an economic value relative to other economic traits.

Molecular data Although several genetic tests are available, the extent to which they are used in commercial applications is unclear, as is the manner in which they are used and whether their use leads to greater response to selection. In general, the following four strategies for MAS or GAS can be distinguished:

- I) Tandem selection, with selection on genotype followed by selection on EBV.
- II) Index selection on a combination of genotype and EBV: $I = b_1 \text{ genotype} + b_2 \text{ EBV}$
- III) Pre-selection on genotype (or an index of genotype and EBV) at a young age, followed by selection on EBV at a later age.

Although strategy I results in the most rapid fixation of the gene of interest, it results in the greatest loss in response for polygenes or other traits and, may, therefore, result in lost overall (QTL and polygenes) response to selection, in particular over multiple generations. In theory, strategy II results in the greatest overall response to selection for a given selection stage, in particular if weights are optimized (e.g. Dekkers and van Arendonk, 1998). However, if selection is over multiple stages, the impact on response

for polygenes and other traits can be minimized if genotype information can be used at an early age when limited or no phenotypic information is available to distinguish selection candidates. An example is pre-selection among full-sib dairy bulls for entry into progeny testing programs (Kashi et al., 1990). The choice between these strategies (and other alternatives) also depends on other factors, such as market and cost considerations. It is unclear to which extent each of these strategies is applied in commercial MAS and GAS.

Success of MAS also depends on the consistency of QTL effects across populations and environments. Results from introgression programs in plants have found that effects tend to be consistent for major genes for simple traits but not for QTL for complex traits (e.g. yield) (Hospital, 2002). Inconsistent effects have also been observed for some well-studied genes in livestock. For example, for the ESR gene for litter size in pigs some studies have found no effect and interactions with line and environment have been identified (Rothschild and Plastow, 2002). Reasons for inconsistent results across studies and populations include statistical anomalies such as false positive or negative results (small sample sizes) and overestimation of significant QTL effects, as well as true effects, such as inconsistent marker-QTL linkage phases across populations for LD-markers, genotype by environment interactions, and epistatic effects. This points to the need to continuously evaluate and monitor gene or QTL effects in the target population and environment, which requires continuous emphasis on phenotypic recording in both nucleus and field populations.

Whereas the success of GAS and LD-MAS can be evaluated at the level of changes in gene frequencies, success of LE-MAS can only be evaluated on the basis of extra genetic gain. LE-MAS has primarily been applied to dairy cattle for pre-selection of young bulls for entry into progeny testing programs (strategy III) (e.g. American Breeders Service (Dennis Funk, pc), Accelerated Genetics (Mike Cowan, pc), Livestock Improvement Corporation (Spelman, 2002), Holland Genetics (Erik Mullaard, pc), France (Boichard, 2002), Germany (Bennewitz et al. 2003)). The success of these applications is unclear and depends on the ability to integrate several technologies, including genetic evaluation and reproductive technology to produce full-sib families with selection space. Strategies and challenges for the implementation of LE-MAS for dairy cattle have been described by several (Spelman, 2002, Boichard et al., 2002).

GAS vs. LD-MAS vs. LE-MAS

An important consideration for the use of molecular genetics in breeding programs is whether to work toward the application of LE-MAS, LD-MAS, or GAS. Table 2 summarizes the relative requirements and opportunities for these three strategies.

Requirements for detection are least for LE markers and greatest for identification of functional mutations. However, once a functional mutation is identified, requirements for estimation and confirmation of effects in other populations are much lower than for LE-markers because the latter requires phenotypes and genotypes on pedigreed

populations versus a random sample. Requirements for integration of genotype data in routine genetic evaluation procedures are also much greater for LE-MAS than for LD-MAS and GAS, both with regard to requirements of individuals that must be phenotyped and genotyped and with regard to methods of analysis. Genetic evaluation requirements are slightly greater for LD-MAS than GAS because LD-MAS requires identification and analysis of marker haplotypes and confirmation of marker-QTL linkage phases.

Table 2. Requirements and opportunities for the implementation of LE-MAS vs. LD-MAS vs. GAS.

QTL detection requirements	LE-MAS < LD-MAS << GAS
Within-line confirmation requirements	LE-MAS >> LD-MAS > GAS
Routine genetic evaluation requirements	LE-MAS >> LD-MAS > GAS
Phenotyping – relatives (LE) vs. sample (LD/GAS)	LE-MAS >> LD-MAS > GAS
Genotyping – candidate + relatives (LE) vs. candidate only	LE-MAS >> LD-MAS > GAS
Analysis – MA-BLUP (LE) vs. fixed effect (LD/GAS)	LE-MAS >> LD-MAS > GAS
Genome-wide analysis opportunities	LE-MAS >> LD-MAS >> GAS
Implementation logistics	LE-MAS >> LD-MAS > GAS
Genetic gain opportunities (for given QTL)	LE-MAS < LD-MAS ~< GAS
Marketing opportunities (patents, product differentiation)	LE-MAS << LD-MAS < GAS

Whereas the previous refer to requirements for a give QTL, LE-MAS allows for genome-wide analysis and evaluation of QTL with a limited number of markers. This is also possible for LD-MAS with high-density genotyping. Meuwissen et al. (2001) demonstrated that EBV of high accuracy could be obtained from a Bayesian mixed model analysis of marker haplotypes with high-density genotyping.

Opportunities for increases in genetic gain from a given QTL are lowest from LE-MAS because of the limited information that is available to estimate effects on a within-family basis (Pong-Wong et al., 2002), while for both LD-MAS and GAS, effects are estimated from data across families. Accuracy of estimates may be slightly lower for LD-MAS than GAS as a result of incomplete marker-QTL disequilibrium and a greater number of effects (marker haplotypes versus QTL genotypes) (Hayes et al., 2002). Opportunities for intellectual property protection and product differentiation are greatest for GAS but limited for LE-MAS.

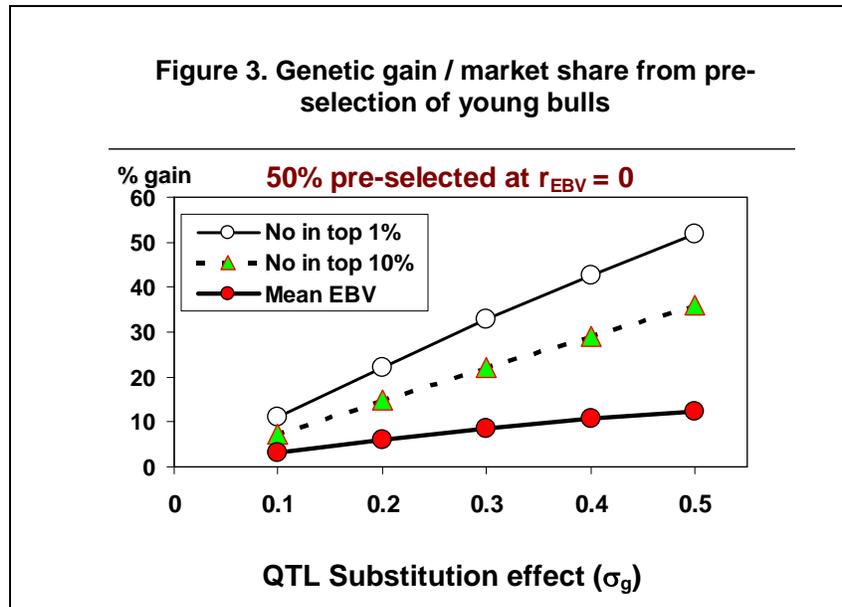
Integration of MAS in breeding programs

Whereas initial applications may have been on an ad-hoc basis, it is clear that successful implementation of a MAS program requires a comprehensive integrated approach that is closely aligned with business goals and markets. In practice, all three types of markers are available for the categories of traits described previously and a comprehensive approach is needed to collect, integrate, and analyze data on

phenotypes for multiple traits, LE markers, LD markers, and genes and to develop selection strategies that meet business goals.

Commercial application of MAS also requires careful consideration of economic aspects and business risks. Economic analysis of MAS requires a comprehensive approach that aims to evaluate the economic feasibility and optimal implementation of MAS. An excellent example of such an analysis is in Hayes and Goddard (2003), who conducted a comprehensive economic analysis of the implementation of LE-MAS in the nucleus breeding program of an integrated pig production enterprise. QTL detection and MAS on identified QTL regions for a multi-trait breeding goal and associated genotyping costs and extra returns from the production phase of the integrated enterprise were considered in the economic assessment. They concluded that implementation of LE-MAS was feasible for the assumed cost and price parameters. They also found that, in particular if QTL detection was based on small sample sizes, stringent thresholds should be set during the QTL detection phase such that genotyping costs during the implementation phase are reduced and selection of false positives is minimized.

Whereas Hayes and Goddard (2003) evaluated economic returns from MAS from increased profit at the production level, which is proportional to extra genetic gain, most commercial breeding programs derive profit from increased market share of breeding stock or germ plasm. In general, implementation of MAS will have a greater impact on market share than on genetic gain. An example is in Figure 3, which evaluates the impact of pre-selection of young dairy bulls in a competitive market. For a QTL with a substitution effect of 0.3 genetic standard deviations, pre-selection increased genetic gain of selected (top 10%) progeny-tested bulls increased by 7% but the number of bulls in the top 10 and 1% increased by 20 and 30%. This does not imply that the economic feasibility of MAS is greater in a competitive market because that also depends on absolute returns associated with a % increase in genetic gain versus market share; Brascamp et al. (1993) showed, in fact, that economic returns from increased market share were less than from increased production for a pre-selection situation similar to that considered here. Nevertheless, it is important that economic analysis is conducted in relation to business and market realities and goals.



Optimal implementation of MAS also involves careful consideration of alternative selection strategies, business goals, and integration of molecular with other technologies (e.g. reproductive technologies following Georges and Massey, 1991). Opportunities also exist to implement LD-MAS in synthetic lines, capitalizing on the extensive disequilibrium that exists in crosses and their power to detect QTL (Zhang and Smith, 1992). In addition, strategies must be developed to estimate gene effects at the commercial level for nucleus breeding programs, in particular if they involve crossbreeding. This also opens opportunities to use markers to capitalize on non-additive effects and assignment of specific matings. Finally, markers can be used to control inbreeding, parental verification, and product tracing.

Discussion and Conclusions

Marker-assisted selection is used in the livestock breeding industry, primarily through GAS and LD-MAS. Use of LE-MAS has been limited and hampered by implementation issues. Success of commercial application of MAS is unclear and undocumented and will depend on the ability to integrate marker information in selection and breeding programs. The lack of success stories does, however, not imply that commercial applications of MAS have been unsuccessful. The impact of MAS on commercial breeding programs is difficult to evaluate, similar to the difficulty in demonstrating that the numerous improvements in genetic evaluation procedures that have been implemented in recent decades have led to greater rates of improvement in the industry, since controls are lacking and genetic improvement is affected by many factors.

Opportunities for the application of MAS in industry programs exist, in particular for GAS and LD-MAS and, to a lesser degree for LE-MAS because of greater implementation requirements. Regardless of the strategy used, successful application of MAS requires

a comprehensive integrated approach with continued emphasis on phenotypic recording programs to enable QTL detection, estimation and confirmation of effects, and utilization of estimates in selection. Whereas initial expectations for the use of MAS were high, the current attitude is one of cautious optimism.

Acknowledgements

Information was provided by the following people and organizations: Dennis Funk (ABS), John McEwan (AgResearch), Jay Hetzel (Genetic Solutions), Mike Cowan (Genetic Visions), Nadine Buys (Gentec), Erik Mullaard (Holland Genetics), Janet Fulton and Jim Arthur (Hy-Line Int.), Richard Spelman (LIC), Gert Nieuwhof (MLC), Mike Lohuis and Jeff Veenhuizen (Monsanto), Graham Plastow, Guy Pral, and Olwen Southwood (Sygen/PIC), Egbert Knol (TOPIGS), Didier Boichard, Kim Bunter, Rohan Fernando, John Gibson, Ben Hayes, Frederic Hospital, Ricardo Pong-Wong, Max Rothschild, Sheila Schmutz, Mark Thallman, Kent Weigel. Financial support from the State of Iowa, Hatch and Multi-state Research Funds.

References

- Andersson, L., 2001. *Nature Reviews Genetics* 2:130-138 (2001)
- Bennewitz, J., N. Reinsch, J. Szyda, F. Reinhardt, C. Kuhn, M. Schwerin, G. Erhardt, C. Weimann, and E. Kalm. 2003. *Eur. Assoc. Anim. Prod.* Rome, Italy.
- Boichard, D., S. Fritz, M.N. Rossignol, M.Y. Boscher, A. Malafosse, and J.J. Colleau. 2002. *Proc. 7th Wld. Congr. Genet. Appl. Livest. Prod.* 22-03.
- Brascamp, E.W, J.A.M. van Arendonk, and A.F. Groen, 1993. *J. Dairy Sci.* 76:1204-1214.
- Dekkers, J.C.M., and F. Hospital, 2002. *Nature Reviews: Genetics* 3:22-32.
- Dekkers, J.C.M., and J.A.M. van Arendonk, 1998. *Genet. Res.* 71: 257-275.
- Fernando, R.L. 2003. These proceedings.
- Georges, M., and J.M. Massey, 1991. *Theriogenology* 25:151-159.
- Hayes, B., P.J. Bowman, and M.E. Goddard. 2001. *Proc. Assoc. Advmt. Anim. Breed. Genet.* 14: 269-272.
- Hayes, B., and M.E. Goddard. 2003. *Livest. Prod. Sci.*.
- Hospital, F. 2002. *Proc. 7th Wld. Congr. Genet. Appl. Livest. Prod.* 22-05.

Kashi, Y., E. Hallerman, and M. Soller, 1990. *Anim. Prod.* 51:63-74.

Meuwissen, T.H.E., and M.E. Goddard, 1996. *Genet. Sel. Evol.* 28:161-176.

Meuwissen, Hayes, and Goddard. 2001. *Genetics* 157: 1819-1829.

Rothschild, M. F. and M. Soller. 1997. *Probe* 8:13.

Spelman, R.J., 2002. *Proc. 7th Wld. Congr. Genet. Appl. Livest. Prod.* 22-02

Zhang, W. ., Smith, C., 1992. *Theor. Appl. Genet.* 83, 813-820.

STATISTICAL ISSUES IN MARKER ASSISTED SELECTION

R. L. Fernando
Iowa State University, Department of Animal Science, 225 Kildee Hall,
Ames, IA 50011-3150

INTRODUCTION

Marker assisted selection refers to the use of marker genotypes together with trait phenotypes for making selection and mating decisions. A key component of marker assisted selection is the prediction of genotypic values using marker genotypes in addition to trait phenotypes. Methodology has been developed for genetic evaluation by Best Linear Unbiased Prediction (BLUP) using genotypic and phenotypic data (2, 5, 13, 15, 14, 4, 1). This paper presents how these methods can be used for genetic evaluation by BLUP in beef cattle. In the following, BLUP using marker genotypes and trait phenotypes will be referred to as marker assisted BLUP (MABLUP).

GENOTYPIC DATA

A locus that has a direct effect on a quantitative trait is referred to as a quantitative trait locus (QTL). Genotypes at a QTL can be included in BLUP by treating the effects of these genotypes as fixed effects. In practice, however, genotypes may not be available on all the animals. Thus, as discussed later, even in this situation, BLUP is not straightforward.

Most of the currently available genotypes do not have a direct effect on any trait. But, these genotypes may provide useful information if they are closely linked to QTL. Two loci are said to be in gametic phase equilibrium (or linkage equilibrium) if in a randomly sampled gamete the alleles at these two loci are independently distributed. Suppose genotypes are available at a marker locus that is closely linked to a QTL, which will be referred to as the marked QTL (MQTL). If the marker locus and the MQTL are in equilibrium, then the observed marker genotype of an animal does not provide any information about the MQTL genotype of that animal. Thus, in this case, marker genotypes do not provide any information to model the mean or variance of the trait. However, if two relatives receive the same marker allele from a common ancestor, then it is likely that these two relatives also receive the same allele at a closely linked MQTL from this common ancestor. Thus, genotypes at a linked marker do provide information to model the genetic covariances between relatives, even when the marker and MQTL are in equilibrium.

When disequilibrium between the marker and MQTL is partial, the genotype at the marker provides some information about the genotype at the MQTL. For example, if a randomly chosen haplotype contains marker allele M_1 , then at the linked MQTL, the allele may be Q_1 with probability 0.8 or Q_2 with probability 0.2; if the randomly chosen haplotype contains marker allele M_2 , then at the linked MQTL, the allele may be Q_1 with probability 0.1 or Q_2 with probability 0.9. In this case, given M_2 at the marker locus, the allele is Q_2 with high probability at the MQTL. But, because disequilibrium is partial and not complete there is some uncertainty about the MQTL allele.

In the following presentation, we consider using genotypes at linked marker loci that have an arbitrary level of disequilibrium with a single MQTL. Using genotypes at the MQTL is a special case of this arbitrary disequilibrium case, where the recombination rate between the marker and the MQTL is zero and the disequilibrium is complete. Also, using genotypes at markers that are in equilibrium with the MQTL is a special case of the arbitrary disequilibrium case where the level of disequilibrium is zero between the markers and the MQTL.

MODEL

We assume that the genotypic value is completely additive and consider genetic evaluation for a single trait. The phenotypic values for this trait can be modeled as

$$y = X\beta + ZQ\mu + Zu + e, \quad (1)$$

where y is the vector of phenotypic values, β is a vector of fixed effects, X and Z are known incidence matrices, μ is a vector of fixed effects for genotypes at the MQTL, Q is the unobservable random incidence matrix for genotypes at the MQTL, u is the vector of additive effects at all the other QTL, and e is a vector of residuals. BLUP methodology cannot be directly applied to the model (1) because Q is an unobservable random incidence matrix. However, we can write

$$ZQ\mu = Zg, \quad (2)$$

where $g = Q\mu$ is the unobservable, random vector of genotypic values at the MQTL. When the level of disequilibrium between the markers and the MQTL is not zero, the conditional expectation of g given the marker information (M) will not be null.

Thus, let

$$a = g - E(g|M).$$

As shown later, under additive gene action, $E(g|M)$ can be written as

$$E(g|M) = X_g \alpha,$$

where X_g is a matrix with one column of probabilities that can be computed recursively, and α has a single element equal to half the difference between the two homozygous

genotypes at the MQTL. Now, the model can be written as

$$y = X\beta + ZX_g \alpha + Za + Zu + e, \quad (3)$$

where all the incidence matrices are observed, β and α are fixed effects, and a, u , and e are random effects with null means. To construct the mixed model equations, we need to obtain the inverse of the covariance matrix for a and for u . The covariance matrix for u is proportional to the additive relationship matrix, and its inverse can be obtained efficiently (6, 10). Unfortunately, the inverse of the covariance matrix for a is not sparse and its inverse cannot be obtained efficiently. Under additive gene action, however, a_i can be written as

$$a_i = v_i^m + v_i^p,$$

where v_i^m and v_i^p are the maternal and paternal gametic values for the MQTL of animal i . The covariance matrix for the vector of gametic values, v , can be constructed using a simple tabular algorithm, and thus, it can be inverted efficiently (2, 13, 15). Now, BLUP of v and of u can be obtained by solving Henderson's mixed model equations (7) corresponding to the mixed linear model

$$y = X\beta + ZX_g\alpha + Wv + Zu + e, \quad (4)$$

where W is a known incidence matrix relating the elements of v to the elements of y . As shown in the following sections, if we assume that only two alleles are segregating at the MQTL, modeling the mean and the covariances of the MQTL values is considerably simplified.

EXPECTED GENOTYPIC VALUES OF MQTL

The two MQTL alleles are denoted Q_1 and Q_2 , and $2a$ denotes the difference between genotypes Q_2Q_2 and Q_1Q_1 . Then, given additive gene action, the expected genotypic value for the MQTL of animal i is

$$E(g_i|M) = (p_i^m + p_i^p)a, \quad (5)$$

where $p_i^m = \Pr(Q_i^m = Q_2|M)$ and $p_i^p = \Pr(Q_i^p = Q_2|M)$. So, the i^{th} row of the single column of X_g has $(p_i^m + p_i^p)$ and the single element of α has a . The probabilities p_i^m and p_i^p can be computed recursively as follows. If animal i is not a founder, then Q_i^m can be traced either to its dam's maternal allele (Q_d^m) or paternal allele (Q_d^p). So,

$$p_i^m = \Pr(Q_i^m \leftarrow Q_d^m|M)p_d^m + \Pr(Q_i^m \leftarrow Q_d^p|M)p_d^p, \quad (6)$$

where $\Pr(Q_i^m \leftarrow Q_d^m|M)$ is the conditional probability that the maternal MQTL allele of animal i is its dam's maternal allele and $\Pr(Q_i^m \leftarrow Q_d^p|M)$ is the conditional probability that it is the dam's paternal allele. These conditional segregation probabilities are also used to model variances and covariances of the MQTL alleles, and their computation will be discussed later. If i is a founder,

$$p_i^m = \sum_j \Pr(H_i^m = H_j) \pi_j, \quad (7)$$

where H_i^m is the maternal marker haplotype of animal i , H_j is the j th sample state for a marker haplotype, and

$$\pi_j = \Pr(Q_i^m = Q_j | H_i^m = H_j) \quad (8)$$

Similarly,

$$p_i^p = \Pr(Q_i^p \leftarrow Q_s^m | M) p_s^m + \Pr(Q_i^p \leftarrow Q_s^p | M) p_s^p, \quad (9)$$

where $\Pr(Q_i^p \leftarrow Q_s^m | M)$ and $\Pr(Q_i^p \leftarrow Q_s^p | M)$ are the conditional segregation probabilities for the paternal MQTL allele of i . And, if i is a founder,

$$p_i^p = \sum_j \Pr(H_i^p = H_j) \pi_j. \quad (10)$$

VARIANCE AND COVARIANCE OF GAMETIC VALUES OF MQTL

Given only two alleles at the MQTL, it is easy to see that

$$\text{Var}(v_i^m) = a^2 p_i^m (1 - p_i^m), \quad (11)$$

and

$$\text{Var}(v_i^p) = a^2 p_i^p (1 - p_i^p). \quad (12)$$

Consider computing the covariance between two gametic values v_i^m and v_j^p . This covariance can be constructed recursively as follows. Given any pair of animals i and j , one is not a direct descendant of the other. So, without loss of generality, suppose j is not a direct descendant of i . Then,

$$\begin{aligned} \text{Cov}(v_i^m, v_j^p | M) &= \Pr(Q_i^m \leftarrow Q_d^m | M) \text{Cov}(v_d^m, v_j^p | M) \\ &\quad + \Pr(Q_i^m \leftarrow Q_d^p | M) \text{Cov}(v_d^p, v_j^p | M). \end{aligned} \quad (13)$$

Use of (13) to compute the covariance matrix Σ_v of v can be expressed in matrix notation as follows. The rows and columns of Σ_v are ordered such that those for ancestors precede those for

descendants. Suppose Σ_s is the gametic covariance matrix for animals 1, 2, . . . , $i - 1$. This matrix can be expanded to include the covariances with v_i^m , for example, as

$$\Sigma_{s+1} = \begin{bmatrix} \Sigma_s & \Sigma_s q \\ q' \Sigma_s & \text{Var}(v_i^m) \end{bmatrix} \quad (14)$$

where q is a $2(i-1) \times 1$ vector with the maternal and paternal segregation probabilities for Q_i^m at the positions corresponding to v_d^m and v_d^p and zero at all the other positions, and $\text{Var}(v_i^m)$ is computed using (11). The inverse of Σ_v can be obtained recursively as follows (13, 15). Suppose that Σ_s^{-1} is the inverse of the sub matrix Σ_s defined previously, then the inverse of Σ_{s+1} is

$$\Sigma_{s+1}^{-1} = \begin{bmatrix} \Sigma_s^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -q \\ 1 \end{bmatrix} v^{ii} [-q' \quad 1], \quad (15)$$

where

$$v^{ii} = [\text{Var}(v_i^m) - q' \Sigma_s q]^{-1}. \quad (16)$$

SEGREGATION PROBABILITIES

In computing the expected value, variance or covariance of gametic values at the MQTL, observable marker information is used to trace the unobservable MQTL alleles of an animal to its maternal or paternal grandparent. This tracing of MQTL alleles back to the grandparents is mathematically accomplished through the use of conditional segregation probabilities. Here, we will briefly discuss how these conditional probabilities can be computed given the observed marker information. When the marker information consists of genotypes at a single marker locus and all genotypes are observed, simple rules have been developed to compute the segregation probabilities (15). In most situations, however, more than one marker will be used to trace the alleles at a single MQTL. In this case, when the linkage phase for the markers is known, segregation probabilities can be computed easily (5). Suppose an animal i has genotype $\frac{A_1 Q_i^m B_3}{A_1 Q_i^p B_1}$, where the alleles above the line are those inherited from dam and those below from the

sire. Suppose the dam, d , of this animal has genotype $\frac{A_3 Q_d^m B_1}{A_1 Q_d^p B_3}$. Here, we can observe that animal i received its dam's paternal alleles at both the A and B loci. Thus, ignoring double recombination, $\Pr(Q_i^m \leftarrow Q_d^p | M) = 1$. Consider another animal j with genotype $\frac{A_1 Q_j^m B_3}{A_1 Q_j^p B_2}$ and its

sire, s , with genotype $\frac{A_1 Q_s^m B_1}{A_2 Q_s^p B_2}$. In this case, animal j received its sire's maternal allele at the A locus and paternal allele at the B locus. Thus, assuming the MQTL is at the midpoint between A and B, $\Pr(Q_j^p \leftarrow Q_s^m | M) = \Pr(Q_j^p \leftarrow Q_s^p | M) = 0.5$. In practice, however, marker genotypes will not be observed on all animals. Further, even when marker genotypes are observed the linkage phase between markers may not be known. To accommodate these problems, simple methods have been used to approximate segregation probabilities (12). Unfortunately, using these approximations can reduce the advantage of MABLUP (12).

When marker information is not complete, i.e., when the marker genotypes are missing or linkage phase between markers is unknown in parents, the recursive equation (13) used to compute the covariance between gametic values is only approximate (15). Thus, there are two components that contribute to the observed reductions in the advantage of MABLUP. The first is that due to approximations in the segregation probabilities, and the second is that due to use of equation (13) to compute the covariance between gametic values, which is approximate when marker information is not complete. In order to study the magnitude of each of these components, a Markov Chain Monte Carlo (MCMC) sampler was used to get estimates of the segregation probabilities for the pedigree used by Totir et al. (12). When these probabilities are used in equation (13), the reduction in the advantage of MABLUP is entirely due to use of the recursive formula (13) to compute covariances between gametic values. In the worst case, when approximate segregation probabilities were used the advantage of MABLUP was about 50% of the potential maximum. When segregation probabilities were computed using MCMC the advantage of MABLUP was about 80% of the maximum value (9). When segregation probabilities were obtained by MCMC using four flanking markers the advantage of MABLUP was almost 100% of the potential maximum that could be achieved using two flanking markers (9).

ELIMINATING MQTL EQUATIONS

When the segregation probability for an MQTL allele is 1.0, this allele can be traced without any uncertainty to an allele in a parent. If both this allele and the parental allele are included in the model, the gametic covariance matrix will be singular, and such alleles should not be included in the model. Further, if inbreeding of an individual at the MQTL is 1.0, only one MQTL allele should be included in the model. Also, if the MQTL has only two alleles, only gametic values for a single trait should be included in the model. The vector of gametic values for any other trait is proportional to that included in the model (3). Further, gametic values of ungenotyped terminal offspring and of ancestors that do not connect genotyped offspring can be eliminated from the model (8).

DISCUSSION AND CONCLUSIONS

As described in this paper, theory and algorithms are available for including marker genotypes in BLUP. If genotypes are available for the QTL itself, the effects of these can be included as fixed effects in the model. When the genotype for an individual is missing, equation (5) can be used to model the mean of the MQTL. Further, for these individuals random gametic values need to be

included in the model. The covariance between these random effects are modeled using equations (11), (12), and (13).

In this paper we only discussed MABLUP with a single MQTL. The same theory can be used to include more MQTL in the model. For each MQTL, gametic values of animals need to be included in the model. However, as described here, gametic values are not included in the model when an MQTL allele can be traced exactly to its ancestral allele. If a large number of tightly linked markers are used for each MQTL, most of the gametic values can be eliminated from the model. Tightly linked markers, however, are known to cause problems with MCMC methods (11). Further, in multiple-trait MABLUP, only gametic values for a single trait are included in the model. This not only reduces the number of mixed model equations, but the number of covariance parameters that need to be estimated is greatly reduced.

To implement MABLUP as described here requires two types of probabilities: 1) segregation probabilities for use in recursive equations (6), (9), and (13); and 2) haplotype probabilities for use in equations (7) and (10). MCMC methods can be used to get good estimates of these probabilities (11, 9). If the pedigree is very large, the sampler can be applied to a subset of the animals closest to the animal for which the probability is being computed. Calculation of these probabilities is critical for implementation of MABLUP, and research is needed to develop more efficient strategies to accommodate large complex pedigrees and tightly linked markers.

REFERENCES

1. Gamal Abdel-Azim and Albert E. Freeman. A rapid method for computing the inverse of the gametic covariance matrix between relatives for a marked quantitative trait locus. *Genet. Sel. Evol.*, 33:153–173, 2001.
2. R. L. Fernando and M. Grossman. Marker assisted selection using best linear unbiased prediction. *Genet. Sel. Evol.*, 21:467–477, 1989.
3. R. L. Fernando and L. R. Totir. Advances in genetic and statistical models to predict breeding values. In 7th World Congress Genet. Appl. Livest. Prod. CD-ROM Communication No 20-01, 2002.
4. R. L. Fernando and L. R. Totir. Incorporating molecular information in breeding programs: methodology. In W. M. Muir and S. E. Aggrey, editors, *Poultry Breeding and Biotechnology*. CABI Publishing, Cambridge, 2003.
5. M. E. Goddard. A mixed model for analysis of data on multiple genetic markers. *Theor. Appl. Genet.*, 83:878–886, 1992.
6. C. R. Henderson. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics*, 32:69–83, 1976.
7. C. R. Henderson. *Applications of Linear Models in Animal Breeding*. Univ. Guelph, Guelph, Ontario, Canada, 1984.

8. Hoeschele. Elimination of quantitative trait loci equations in an animal model incorporating genetic marker data. *J. Dairy Sci.*, 76:1693–1713, 1993.
9. F. V. C. Pita, R. L. Fernando, L. R. Totir, and P. Lopes. An improved approximation of the gametic covariance matrix for marker assisted genetic evaluation by BLUP. *Genet. Sel. Evol.*, (submitted) 2004.
10. R. L. Quaas. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics*, 32:949–953, 1976.
11. C. Stricker, M. Schelling, F. Du, I. Hoeschele, S. A. Fernandez, and R. L. Fernando. A comparison of efficient genotype samplers for complex pedigrees and multiple linked loci. In 7th World Congress Genet. Appl. Livest. Prod. CD-ROM Communication No 21-12, 2002.
12. L. R. Totir, R. L. Fernando, J. C. M. Dekkers, S. A. Fernandez, and B. Gulbrandsen. Effect of using approximate gametic variance covariance matrices on marker assisted selection by BLUP. *Genet. Sel. Evol.*, 36 (in press), 2004.
13. J. A. M. van Arendonk, B. Tier, and B.P. Kinghorn. Use of multiple genetic markers in prediction of breeding values. *Genetics*, 137:319–329, 1994.
14. T. Wang, R. L. Fernando, and M. Grossman. Genetic evaluation by BLUP using marker and trait information in a multibreed population. *Genetics*, 148:507–515, 1998.
15. T. Wang, R. L. Fernando, S. van der Beek, M. Grossman, and J. A. M. van Arendonk. Covariance between relatives for a marked quantitative trait locus. *Genet. Sel. Evol.*, 27:251–274, 1995.

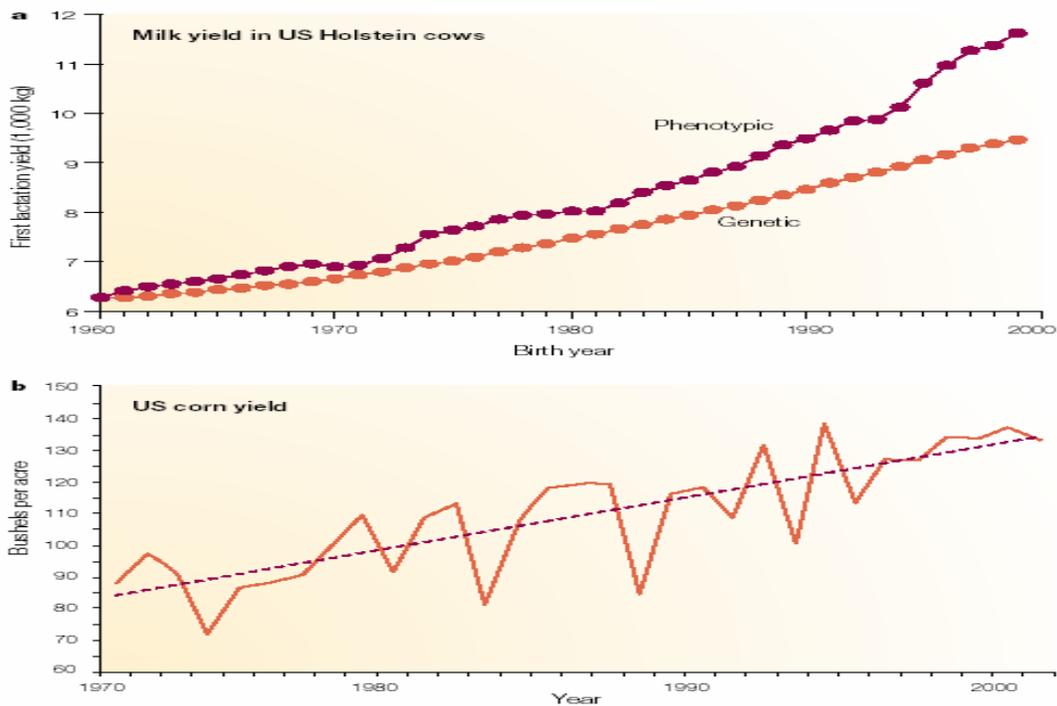
ON MARKER ASSISTED IMPROVEMENT OF LIVESTOCK IN THE POST-GENOMICS ERA

Daniel Gianola
Department of Animal Sciences
Department of Biostatistics and Medical Informatics
Department of Dairy Science, University of Wisconsin-Madison

OUTLINE

- On molecular markers
- On post-genomic data
- On the economics of the beast

Animal and plant breeding work!
Dekkers and Hospital (2002), *Nature Reviews Genetics* 3, 22-32



THE MOLECULAR MARKERS ERA

Terms

- QTL (Quantitative Trait Loci)
 - Locus or group of linked loci influencing a trait
- Genetic marker
 - Genetic entities used to trace Mendelian inheritance
- MAS (Marker Assisted Selection)

- Use marker information in animal breeding scheme
 - to “improve” something
 - Main problem is the “something”!!



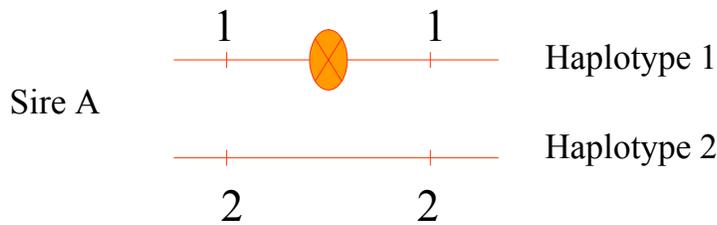
Assumptions

- QTL-markers association identified and confirmed
- Animals carrying the favorable QTL allele have desirable phenotype



MAS breeding schemes

- Select animals based on genetic marker or marker haplotype known to be in association with the favorable QTL allele



Genetic evaluation

- Fernando & Grossman, 1989:

$$y = X\beta + Zu + ZQq + e$$

- Q = incidence matrix linking QTL allelic effects to animals
- q = vector of allelic effects for the QTL
- other terms are as in a standard model
- Marker genotyping needed

Problems?

- Recombination between marker and QTL
- Homozygosity at the marker locus
- Tracing marker alleles from sire to offspring
- Inaccurate QTL-parameter estimates
 - Overestimating QTL variance
 - Location error

HUMAN TRANSCRIPTOME MAP: REGIONS OF CO-EXPRESSION Caron et al. (2001), Science 291, 1289-1292

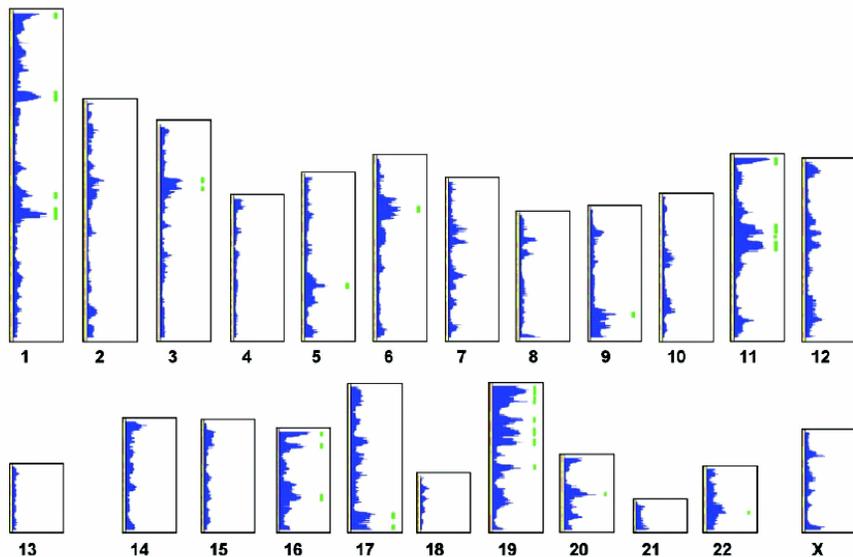


Fig. 3. Regional expression profiles for 23 human chromosomes show a clustering of highly expressed genes in RIDGES. Expression levels are shown as a moving median with a window size of 39 genes. There are 74 regions

with one or more consecutive moving medians that have a lower limit of four times the genomic median; 27 of them have a length of at least 10 consecutive moving medians (indicated by green bars).

MARKER ASSISTED PREDICTION OF BREEDING VALUE

- QTL cartography is not needed for use of marker information
- Markers can be incorporated in a statistical model as covariates
- Hierarchical modelling (Gianola, Perez-Enciso and Toro, 2003, Genetics 163, 347-365)

Model structure

- Data given polygenes
- Polygenes given marker effects
- Marker effects given chromosomes
- Spatial structure within chromosomes (co-expression)

- “Stuff” model with family structures
- Use REML-BLUP tandem or Bayesian implementation

Application to pigs (Varona et al., unpublished)

- 3 boars from Iberian line mated to 31 Landrace sows
- F₁ : 6 boars X 73 sows
- F₂ : 577 pigs
- 321 recorded (58 full-sib families) for carcass traits
- Genotyping for 92 markers

Model

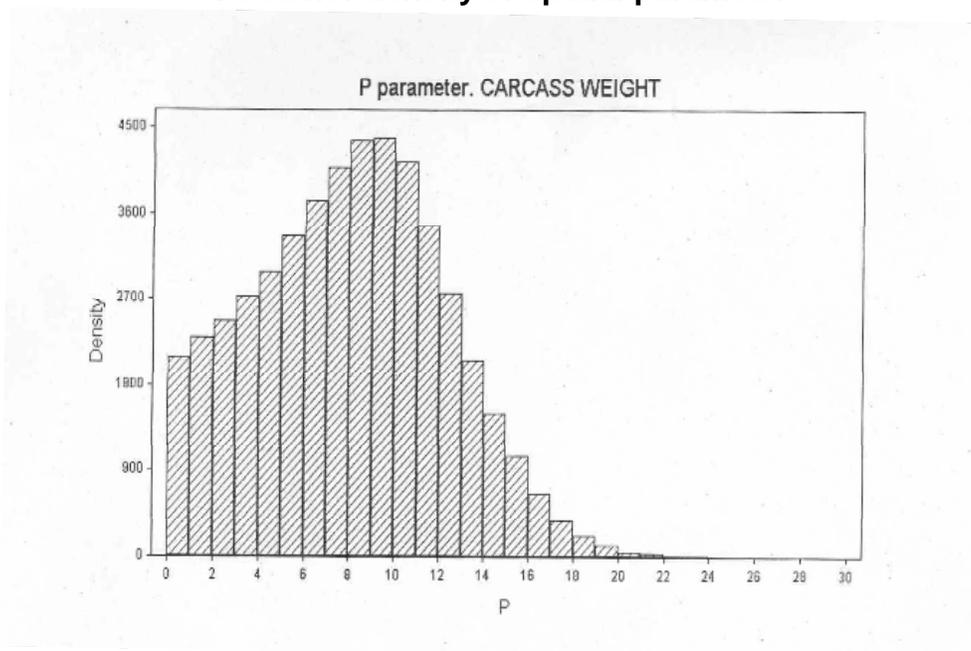
- Sex, family and additive effects of 92 markers
- Marker effects linked to data via stochastic matrix
- Marker effects in a linkage group correlated with a Gaussian decay model

Distribution of markers by chromosome

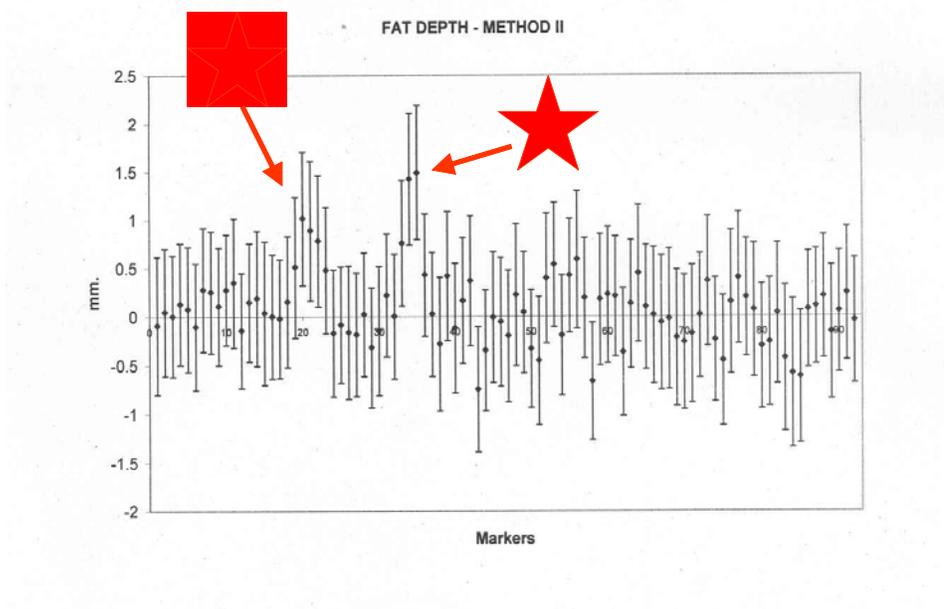
Table 1. Marker positions (Pos) and information content (IC) at marker position sorted by chromosome (Chr.).

Chr	Marker	Pos	IC	Chr	Marker	Pos	IC	Chr	Marker	Pos	IC
1	SW1515	0.0	0.53	6	S0035	0.0	0.65	12	S0143	0.0	0.78
	CGA	30.1	0.99		SW1057	44.3	0.96		GH	31.4	0.71
	S0113	46.2	0.57		S0087	57.7	1.00		SW874	48.6	0.98
	S0155	55.0	0.78		SW316	81.2	0.90		S0106	81.7	0.79
2	SW1828	85.0	0.85	7	S0228	96.0	0.48	13	S0219	0.0	0.57
	IGF2	0.0	0.70		SW1881	108.7	0.82		SW935	30.9	0.54
	S0141	30.3	0.93		SW2419	145.3	0.96		SWR100	64.0	0.97
	SW240	41.8	0.98		S0025	0.0	0.73		SW398	81.4	0.96
	SW395	64.7	0.95		S0064	40.1	0.76		SW1056	91.2	0.41
	S0226	72.4	0.99		TNFB	68.9	0.93		SW769	121.5	0.58
	S0378	87.0	0.93		S0066	87.8	0.99		SW857	0.0	0.89
3	SWR308	130.1	1.00	8	SW632	111.9	0.94	14	SW1125	18.8	0.92
	SW72	0.0	1.00		S0101	137.7	0.92		SW210	42.2	0.82
	S0206	25.6	0.51		SW764	160.3	0.94		S0007	55.8	0.87
	S0216	55.1	1.00		SW2410	0.0	0.98		SW1557	90.8	0.43
	S0002	77.5	0.72		SW905	26.0	0.60		SW2515	114.0	0.75
4	Sw349	86.0	0.99	9	SWR110	44.7	1.00	15	SW919	0.0	1.00
	SW2404	0.0	0.80		S0017	66.5	0.98		SW1111	16.3	0.56
	S0301	40.8	0.85		S0225	86.1	0.83		S0149	38.3	1.00
	S0001	59.5	0.88		SW61	109.1	1.00		SW936	56.0	0.85
	SW839	72.8	1.00		SW983	0.0	0.99		SW1119	79.9	0.45
	DECR2	78.8	0.18		SW911	31.1	0.71		SW742	0.0	0.99
	S0214	95.1	1.00		SW2571	79.5	0.73		S0298	18.4	0.44
5	SW445	116.8	1.00	10	SW2093	109.2	0.99	16	SW2517	35.9	1.00
	S0097	134.4	0.84		SW1349	160.9	0.76		S0061	69.4	0.37
	SW413	0.0	0.98		S0038	0.0	0.74		SW24	0.0	1.00
	SW2425	66.1	0.50		S0070	45.5	0.82		SW1920	28.3	0.95
	S0005	81.8	1.00		SW1626	100.5	0.97		SW2431	72.2	0.51
	IGF1	113.8	0.91		S0385	0.0	0.87		SW1023	0.0	0.85
SWR111	130.9	0.95	11	S0071	43.1	0.75	17	SW787	21.5	0.84	
				SW703	72.3	1.00		S0120	35.1	1.00	

Posterior density of spatial parameter

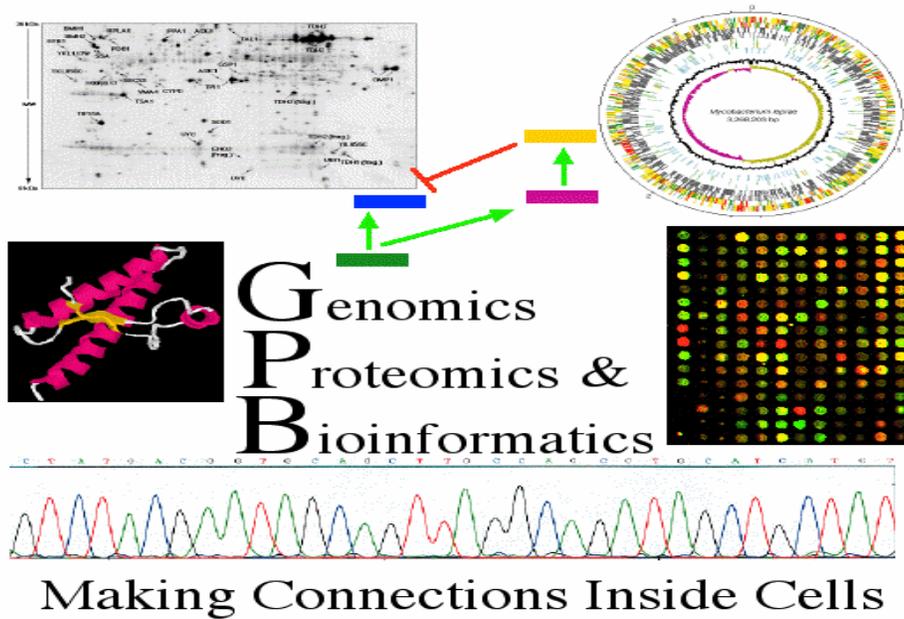


“Known” QTL confirmed, more detected than with Haley et al (1994)



THE POST-GENOMICS ERA
a scientist's point of view

New Data, New Problems, New Paradigms?



The post-genomics era

- Measures of *mRNA* abundance
 - Gene expression: ability to produce a biologically active protein
- Proteomics: outputs from mass spectrometry
- Metabolomics: genetic control of metabolic pathways → statics, kinetics
- Energetics:

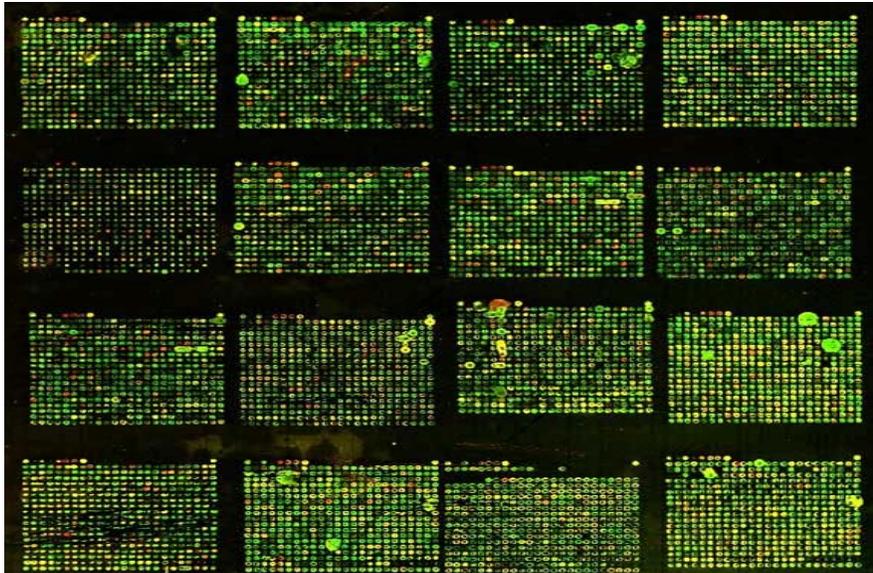
$$\Delta Gibbs = \Delta Energy - T\Delta Entropy$$

The post-genomics era - What is a c DNA microarray?

- Suppose have “genes” or pieces of known *DNA* coding for known protein
- Put each *DNA*, single stranded, in known spot within a non-porous slide. Many, many such spots...
- Extract *mRNA* from target tissue. Make complementary *DNA* (*cDNA*)
- Pour *cDNA* onto slide. If “gene” is in tissue and in some spot → hybridization
- Detect hybridization via fluorescence. Interpret intensity as measure of *mRNA* abundance.

→ **Massively Parallel Hybridization Experiment!!**

See what you get!



A genomics company point of view...

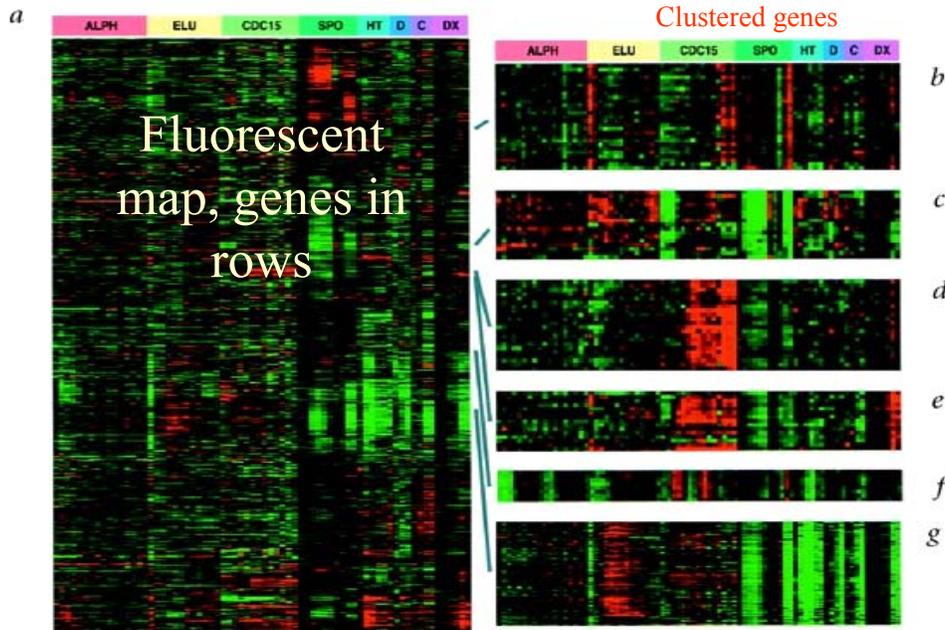


Example of what can be done

(Brown and Botstein, 1999)

- Fluorescent map of yeast genome
- 2467 rows (genes with known function)
- 8 columns (“treatments”: mitotic cell cycles, sporulation, heat and cold shocks)
- Map compressed via cluster analysis

Fluorescent map, genes in rows



STATISTICALLY, THINGS GET EVEN MORE COMPLICATED

NEED TO MODEL DISTRIBUTION OF SPOT READINGS

- Density of joint distribution of readings:

$$p(R, G|\theta_R, \theta_G, a) = p(R|\theta_R, a)p(G|\theta_G, a) \\ = \frac{(\theta_R \theta_G)^a}{\Gamma^2(a)} R^{a-1} G^{a-1} \exp[-(R\theta_R + G\theta_G)],$$

- Target fold-change in expression:

$$\rho = \frac{\mu_R}{\mu_G} = \frac{\theta_G}{\theta_R}$$

- Change variables to $T=R/G$ and $V=G$. Integrate over V :

$$p(T|\theta_R, \theta_G, a) = \frac{\Gamma(2a) \left(\frac{T}{\rho}\right)^{a-1}}{\Gamma^2(a) \rho \left(1 + \frac{T}{\rho}\right)^{2a}}$$

- Can show that T is not independent of RG , so it does not use additional information about $mRNA$ abundance

- Marginal likelihood at spot G :

$$\begin{aligned}
& l_g(a, a_0, v) \\
&= \iint p(R_g | \theta_{R_g}, a) p(G_g | \theta_{G_g}, a) p(\theta_{R_g} | a_0, v) p(\theta_{G_g} | a_0, v) d\theta_{R_g} d\theta_{G_g} \\
&= \int p(R_g | \theta_{R_g}, a) p(\theta_{R_g} | a_0, v) d\theta_{R_g} \int p(G_g | \theta_{G_g}, a) p(\theta_{G_g} | a_0, v) d\theta_{G_g} \\
&= \int \frac{\theta_{R_g}^a}{\Gamma(a)} R_g^{a-1} \exp(-R_g \theta_{R_g}) \frac{v^{a_0}}{\Gamma(a_0)} \theta_{R_g}^{a_0-1} \exp(-\theta_{R_g} v) d\theta_{R_g} \times \\
&\quad \int \frac{\theta_{G_g}^a}{\Gamma(a)} G_g^{a-1} \exp(-G_g \theta_{R_g}) \frac{v^{a_0}}{\Gamma(a_0)} \theta_{G_g}^{a_0-1} \exp(-\theta_{G_g} v) d\theta_{G_g},
\end{aligned}$$

- Analytical integration yields:

$$l_g(a, a_0, v | R_g, G_g) = \left[\frac{\Gamma(a+a_0)}{\Gamma(a)\Gamma(a_0)} \right]^2 \frac{v^{2a_0} (R_g G_g)^{a-1}}{[(R+v)(G+v)]^{a+a_0}}$$

- Log-likelihood over all genes:

$$L(a, a_0, v | \text{all readings}) = \sum_g l_g(a, a_0, v)$$

Gamma-Gamma-Bernoulli Model (Newton et al., 2001)

- A reading may reflect (or not) differential expression
- Typical mixture problem
- Introduce binary variable Z_i to denote differential expression (variable=1) or absence of differential expression (variable=0)
- The distribution of readings under expression was deduced to be:

$$l_g(a, a_0, v | R_g, G_g) = \left[\frac{\Gamma(a+a_0)}{\Gamma(a)\Gamma(a_0)} \right]^2 \frac{v^{2a_0} (R_g G_g)^{a-1}}{[(R+v)(G+v)]^{a+a_0}}$$

- Distribution under no DEX arrived at by making no distinction between *RED* and *GREEN* parameters, yielding:

$$\begin{aligned}
& l_{0,g}(a, a_0, v) \\
&= \int p(R_g | \theta, a) p(G_g | \theta, a) p(\theta | a_0, v) d\theta \\
&= \frac{\Gamma(2a + a_0)}{\Gamma^2(a)\Gamma(a_0)} \frac{v^{a_0} (RG)^{a-1}}{(R + G + v)^{2+a_0}}.
\end{aligned}$$

NEED TO INCORPORATE MIXTURE STRUCTURE

- Let p = proportion of differentially expressed spots
- Binary variable can be treated as “missing data”
- For spot g the joint probability of “observing” the **red** and **green** measurements and the binary variable is:

$$\begin{aligned} & q(R_g, G_g, Z_g | a, a_0, v, p) \\ &= s(R_g, G_g | Z_g, a, a_0, v) \Pr(Z_g | p) \\ &= [l_{1,g}(a, a_0, v | R_g, G_g) p]^{Z_g} [l_{0,g}(a, a_0, v) (1 - p)]^{1 - Z_g}. \end{aligned}$$

- The complete data likelihood is:

$$l_{\text{Complete}}(a, a_0, v, p) = \prod_g [l_{1,g}(a, a_0, v | R_g, G_g) p]^{Z_g} [l_{0,g}(a, a_0, v) (1 - p)]^{1 - Z_g}$$

- Imputation for spot g

$$\begin{aligned} \hat{Z}_g(a, a_0, v, p) &= E(Z_g | a, a_0, v, \text{all } R_g, G_g) \\ &= \Pr(Z_g = 1 | a, a_0, v, p, \text{all } R_g, G_g) \\ &= \frac{l_{1,g}(a, a_0, v | R_g, G_g) p}{l_{1,g}(a, a_0, v | R_g, G_g) p + l_{0,g}(a, a_0, v) (1 - p)} \end{aligned}$$

Use in quantitative genetics?

- Genetic evaluation of livestock
- Detection of Quantitative Trait Loci
- Mere covariates for predictive discrimination?
- Still not obvious how to incorporate in a context of transmission genetics
- Transcription versus protein genetics
- Uncharted waters...

Computational feasibility?

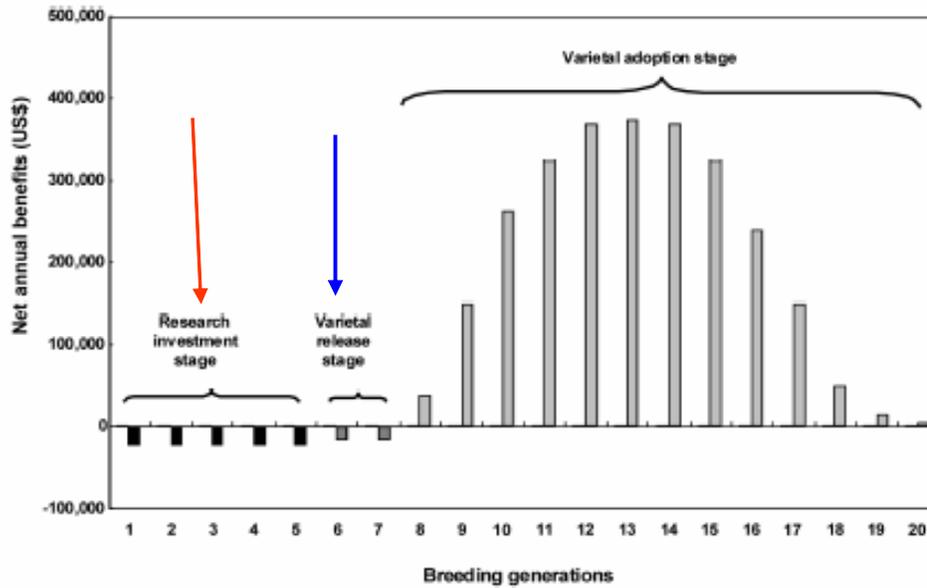
- Q does 30 trillion operations/second
- **Everyone on the planet needs to do 5000 calculations/second to keep up**
- Celera assembles human genome in 150 days
- **Q does it in 4**
- Q: 33 terabytes of memory
- Q: one day what takes 60 years in a high-end personal computer

THE BOTTOM LINE

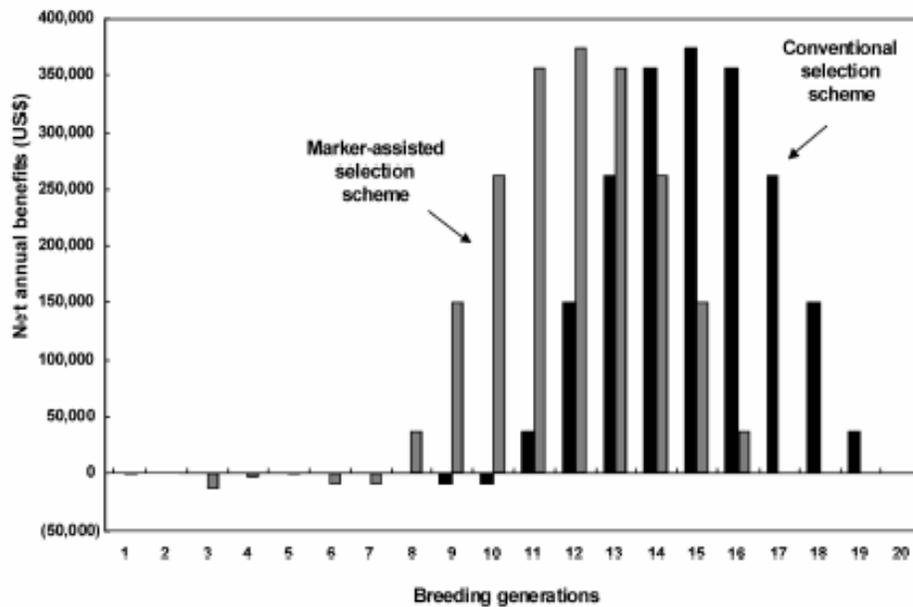
The economics of animal and plant breeding...Tremendous scarcity of studies!

STYLIZED ECONOMIC MODEL: BENEFITS DURING THE INVESTMENT, RELEASE AND ADOPTION STAGES

Morris et al. (2003): Molecular Breeding 11, 235-247



Distribution of benefits in the course of time



**Economic analysis: net present value and rates of return of
Conventional versus MAS improvement**

	Conventional	MAS-A	MAS-B
Net present value (NPV)	\$364,150	\$497,773	\$503,167
NPV-based rank	3	2	1 
Internal rate of return (IRR)	131%	74%	98%
IRR based rank	1 	3	2

CONCLUSION

- Statistically, can cope with MARP (Marker assisted prediction)
- Can ignore or take into account QTL cartography
- Unprecedented speed of changes in genomics knowledge and technology
- Economics of MAS not beyond reasonable doubt (current FAO electronic conference)

STATISTICAL CHALLENGES

**Stephen D. Kachman
Department of Statistics, University of Nebraska
Lincoln, NE**

Summary

Incorporating marker information into genetic evaluations poses a number of statistical challenges. These challenges arise from the effects that an individual gene have on traits of economic importance. Typically the impacts of these genes tend to account for only a small percentage of the phenotypic variability and have effects that vary depending on the genetic and environmental backgrounds. Statistical models will therefore need to be developed to account for this variability. If selection will be placing heavy pressure on quantitative trait loci, it will be important to understand the effect of the associated genes on growth, reproduction, and health. It will therefore be important to eventually develop functional models.

Outline of Presentation

Introduction

- Utilizing marker information
- Shift in focus from the average effect of many genes
 - Smooth out many of the rough edges
- to focusing on the effect of individual genes
 - Rough edges start to come into focus

Some of the rough edges

- Relatively small (Few percent of the residual variability)
- Influenced by genetic and environmental background
- Influence multiple traits
- Traits are influenced by many QTL
 - Epistatic effects

Some choices

- Statistical model
 - Work with an individual QTL
 - Uncertainty modeled using random effects
 - * Uncertainty about the genotype
 - * Uncertainty about the effect
- Functional model
 - Identify the gene and its effects
 - * Shotgun type approaches (Pairing QTL positions and expression arrays)
 - * Information from other species
 - * Database mining

Statistical Challenges

- Considerable effort has been placed on developing statistical models
- Building a functional model is very expensive
- Pairing of QTL positions and expression arrays
 - QTL scan is represented by a single point
 - How do we quantify our uncertainty?
 - How do we combine the information?
 - * Different tissues
 - * Different points in time

Which way?

- Statistical model
 - Initial widespread application
 - The more pressure that is placed on a QTL
 - The more important it is to understand the consequences
- Functional model
 - Reduce the uncertainty
 - Some uncertainty will remain

CHALLENGES OF APPLICATION OF MARKER ASSISTED SELECTION

Ignacy Misztal
University of Georgia

Marker assisted selection in animal breeding

- Detection of markers associated with QTL
- Use of markers in marker-assisted selection

- Possibility:
 - Evaluate genotypes without phenotypes
 - Success with sex limited and low heritability traits

- Optimism and enthusiasm
- Large funding (public and private)
- Large projects

- Many scientists trained in new techniques
- New scientific insights

Quote

- M. Georges (1991): In a few years there will be no need for BLUP

MAS, cont.

- **Commercialization**
 - Promises
 - Markers found almost for any trait
 - Problems with validation
 - Bankruptcies and closings

Should commercialization start before research is reasonably advanced?

Attitude towards MAS/QTLs

- **Optimism in academia proportional to funding**
 - correlation <1!
 - Lack of faith among many practitioners

- **Skepticism about MAS in industry**
 - Large expenditures while important problems not solved
Some unsolved problems of the industry

- U.S. Holsteins (T. Lawlor)
 - Reputation of poor and declining fertility
 - Threat of importation of other breeds and crossbreeding
- Pigs (M. Culbertson)
 - Little or no improvement at the commercial level over last 10 years
 - Mortality challenges (some farms up to 30% birth to harvest)
 - Reduced meat quality
 - Increased challenges with diseases

Quote

- **B. Hill and R. Thompson (1993):**

If 100 QTLs are responsible for a trait and they act additively, animal breeders will soon be doing simple counting

but assume interactions, and the QTL model becomes intractable



Latest Results

- WCGALP 2002 on MAS/QTL
 - General
 - Simulation
 - New Zealand (Spelman)
 - Marker for increased fat, reduced protein yield (DGAT1)
 - Losses due to testing fewer bulls
 - France (Boichard et al.)– no results
 - Plants (Hospital et al.)– selection in wrong direction
- Beef (Quaas et al, 2002) – uncertain mode of inheritance
- Pigs (Noguera et al, 2003): ESR (fertility) gene downgraded to ESR marker with effect close to 0 in European populations

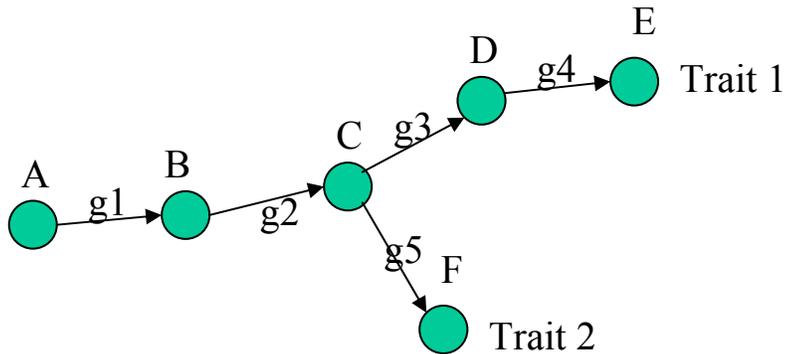
Problem with QTLs Details or basic assumptions?

2003 ADSA/ASAS Symposium Molecular Genetics: Lessons from Past / New Directions

Speakers and Program

- Current status in QTLs/MAS - Jack Dekkers
 - Many markers
 - No markers for low heritability traits
 - Benefits from commercial applications hard to assess
- QTLs in mice - Daniel Pomp
 - Many markers found but no genes
 - Underlying genetic model much more complicated than Mm/Qq
 - Microarray studies useful – long time horizon
- Future - Bruce Walsh
 - A long time before regular selection obsolete
 - Microarray data potentially useful
 - Caution because:
 - Expression tissue dependent
 - Biochemical pathways involve genes in complex way, involving branches and loops

Example of a gene network



A-E: products in the pathway g1-g5: genes or enzymes

Other Issues

- Some traits not well defined (e.g., fertility)
- Traits change over time
- Genetic correlations among traits change over time
- GxE
- If selection is an optimization, then if something is gained then what is lost?

Perils of single-trait selection are well known (Dudley, 2003)

Conclusions

- Some tools from molecular genetics already useful
- Path from gene to phenotype not simple
- Future: complex and nonlinear biological model in 10-50 years....
- Industry:
 - Constant need to solve new problems
 - Majority of gains through traditional selection
 - Strong reliance on molecular genetics risky to bottom line