final report

Project code: B.BSC.0092
Prepared by: Michael Goddard
Date published: October 2015

PUBLISHED BY
Meat & Livestock Australia Limited
PO BOX 1961
NORTH SYDNEY NSW 2059

Genome sequencing for SNP research tool development and use

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government to support the research and development detailed in this publication.

PROGRAM 7
UNDERPINNING SCIENCE AND GENE DISCOVERY

Take home message

The CRC has developed methods to predict breeding value for economically important traits from DNA samples. These DNA tests are commercially available to cattle breeders and the results will be delivered to them as BREEDPLAN EBVs that include the DNA information and traditional performance data on the animal and its relatives. In this way cattle breeders will be able to select more accurately for most important traits.

Overview

The Underpinning Science Program coordinates gene discovery research across the CRC and provides services to support the R&D being conducted by the other research programs. Services provided include maintenance of the CRC's database of phenotypic records, maintenance of the DNA collection, design and analysis of gene discovery and gene expression experiments, and bioinformatics support.

The aim of the gene discovery research is to provide DNA based tests that predict the breeding values of cattle for economically important traits. Some important traits are controlled by a single gene such as red vs black coat colour. Polledness is predominantly due to a single gene, at least in European cattle, and consequently a single DNA marker could potentially diagnose cattle that are carrying the gene for horns. (The CRC's research has discovered a DNA test for polled and this is reported under Program 3.) However, most traits are quantitative or complex traits controlled by many genes and by environmental factors. In the past the genes controlling variation in these traits have been largely unknown and so DNA tests were of little use in selecting for traits such as growth rate and fertility. Instead the performance of each animal and his or her relatives was used to estimate the animal's breeding value. The breeding value is the sum of the effects of all the genes the animal carries. Selection on the basis of these estimated breeding values (EBVs) works very well for traits, such as yearling weight, that can be scored on each animal before the age when it can be used for breeding. However, many traits are difficult to select for by traditional performance recording because they can only be observed in cows (eg calving rate) or after slaughter (eg tenderness) or are too expensive to measure (eg feed conversion efficiency). DNA tests that predict the breeding value of cattle for these traits would allow cattle breeders to select for them more accurately and so make more progress and therefore these traits were the main target of the CRC research program.

At the time the CRC started in 2005, it was believed that 5 to 10 genes would explain most of the variation in each trait. Therefore, the CRC searched for a small number of markers that would predict breeding value for our target traits. This strategy met with only limited success. In the case of meat tenderness, 4 markers explain 8% of the genetic variance in Brahman cattle and a commercial test for these markers is available through Pfizer Animal Genetics. The purpose of DNA tests is to estimate the breeding value of each animal for the trait and so the best way to deliver the DNA information to cattle breeders is in the form of an EBV for each animal. This combines the information from the different markers and, in addition, the DNA information can be combined with other data such as the tenderness of progeny of a bull. EBVs for tenderness that combine the DNA markers and traditional performance data are available from BREEDPLAN.

However, the research by the CRC and other scientists around the world discovered that there are hundreds or thousands of genes for most complex traits. Each gene has a very small effect, so a DNA test based on a small panel of markers was unlikely to be useful for most traits. Therefore the CRC changed its research strategy to one that uses a large number of DNA markers covering the whole genome. An equation combining all these markers is used to estimate the breeding value of each animal. This DNA based prediction is combined with traditional performance information to calculate an EBV, delivered through BREEDPLAN, so that cattle breeders can select for carcase and meat quality, feed conversion efficiency and fertility more accurately than in the past. The accuracy of the DNA prediction of breeding value for each trait is given in table 1.

Table 1. The correlations between the DNA prediction equations and true breeding value for 14 traits

Trait	Accuracy
Feed conversion efficiency	0.44
Fat depth (P8 site)	0.27
Tenderness	0.38
Retail beef yield	0.13
Marbling	0.30
Eye muscle area	0.16
Carcase weight	0.30
Hip height	0.36
Yearling weight	0.36
Age at puberty in cows	0.28
Post-partum anoestrous interval	0.21
Cow fertility when lactating	0.36
Age at puberty in bulls	0.28
Percent normal sperm	0.31

The accuracies listed in table 1 should be continually improved as more cattle are added to the database and as better methods are developed.

More detail on the projects within program 7 is given below.

Project 7.1 The CRC database

This CRC and previous Beef CRCs have collected performance information on tens of thousands of cattle and matching DNA genotypes on many of them. This database is an important resources for the cattle industry and will be maintained after the end of the CRC by the Animal Genetics and Breeding Unit at the University of New England for use by BREEDPLAN and by researchers. It contains data from 3 CRCs devoted to the Beef industry (Table 2).

Table 2. Summary of data in Beef CRC database

	Number of animals	Number of records	
		Live	Carcase
CRC I	9,677	579,640	1,034,388
CRC II steers	2,221	142,510	114,888
			Ovarian scans
CRC III cows	2,185	776,888	351,046
CRC III progeny	9,259	76,276	
			Semen data
CRC III bulls	4,278		544,920

Project 7.2 The CRC DNA collection

The database, maintained in project 7.1, is made even more valuable because there are DNA samples available for many of the cattle. Therefore further research can use these samples to test future DNA based tests. The DNA collection is maintained by the Animal Genetics Laboratory of the University of Queensland. The AGL currently holds in storage for the Beef CRC 16,509 DNA samples from CRCI-CRCIII cattle, 1,745 DNA samples from the Industry Sires project (Hans Graser), 350 DNA samples from the Maternal Productivity project (Wayne Pitchford) and ~3000 DNA and hair samples from the recent Young Animal Genotyping project (Hans Graser).

Project 7.3 Coordination of gene discovery

When the CRC begun the research to discover DNA markers for our target traits was located within Programs 1-4. However, when it became clear that all 4 programs needed to rely on the same panel of genome wide markers, the research for all traits was brought together in this project.

Genotype data

As the CRC has progressed we used the latest technology available in DNA markers. This uses markers called single nucleotide polymorphisms (SNPs) that can assayed using a ‘SNP chip’ which genotypes thousands of SNPs at a time. The different SNP chips used and the number of animals tested are given in table 3.

Table 3 Number of cattle genotyped by different SNP panels

SNP panel	Number genotyped
50k version 1	2607
50k version 2	7601
700k	1698
7k	2313
10k	317
Total	14536

The SNP genotypes can be used to investigate the relationship between breeds. Figure 1 shows that first 2 principle components of the relationship matrix among 8747 animals. The first principle component separates the animals on Bos Indicus content with Brahmans at the left and Bos Taurus breeds at the right. The Santa Gertrudis, Droughtmaster, Pastoral company composite breeds and Belmont Reds are all in between these extremes indicating that they all have 3/8 to ½ Brahman. The other dimension separates B.Taurus breeds with Angus at the top to Shorthorn at the bottom. Interestingly, the Belmont Red breed is genetically similar to the tropical composite breeds bred by the large pastoral companies in northern Australia.

Using the SNP genotypes we can also recognise whether a piece of a chromosome in one animal comes from a Bos Taurus or B. Indicus ancestor. By doing this we discovered that 10% of Brahman genes are of B. Taurus origin, presumably as a result of grading up to Brahman from B. Taurus cows. The proportion of genes in Brahmans of B. Taurus origin is not constant across the genome. At some places there is a high proportion of Brahmans carrying Taurus genes probably because, at these sites, Brahman breeders have inadvertently selected for the allele from B. Taurus. One such site contains a gene (PLAG1) with effects on growth, carcase and fertility traits and this is described below.

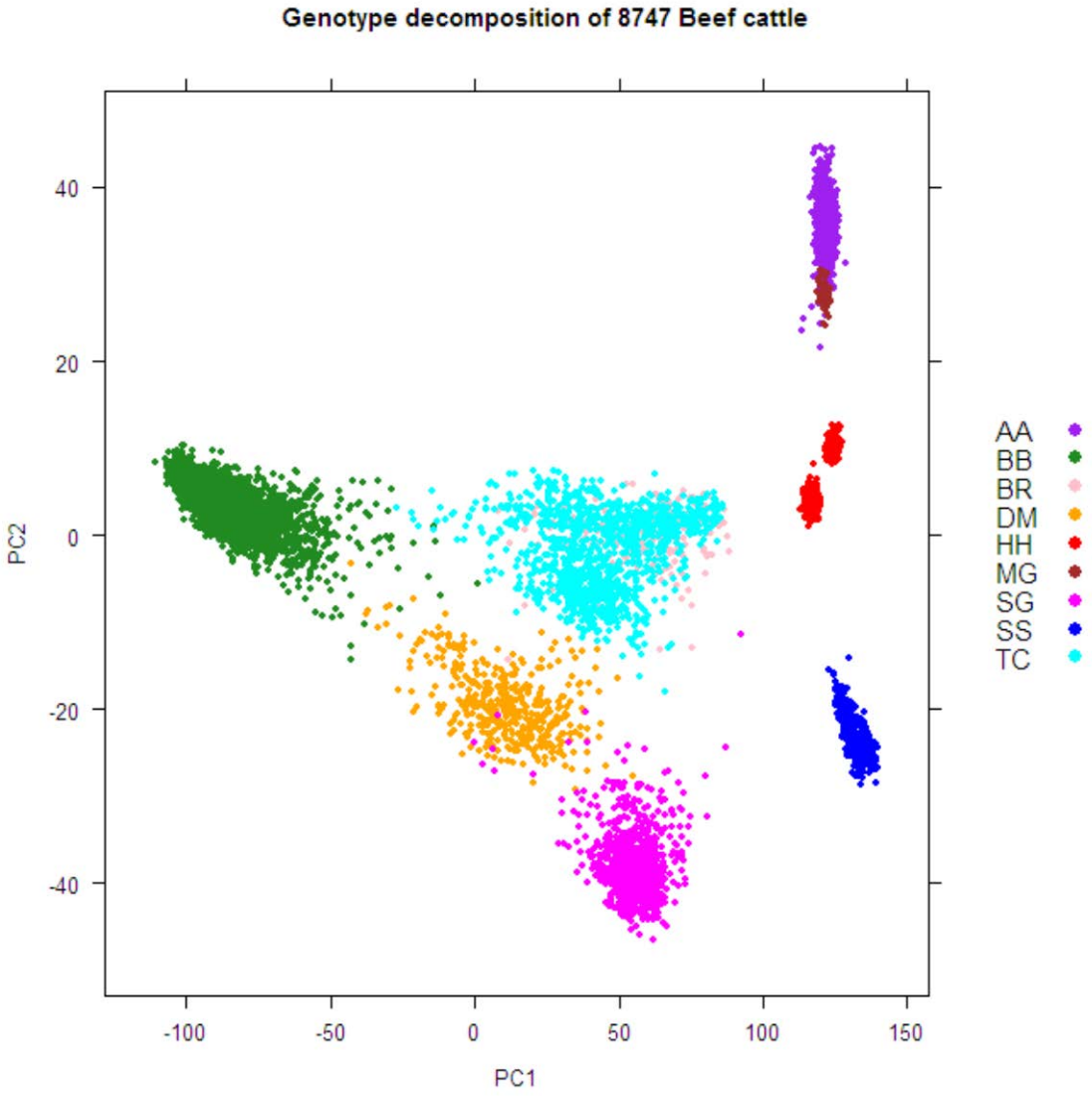

Figure 1 Relationships between cattle breeds based on 50,000 SNPs
(AA= Angus, BB= Brahman, BR = Belmont Red, DM= Droughtmaster, HH = Hereford, MG = Murray Grey, SG = Santa Gertrudis, SS = Shorthorn, TC = Pastoral company composites)

Within a breed, linkage disequilibria (LD) between SNPs on the same chromosome is high. This means that if some a sample of animals within a breed have been genotyped for a panel of dense SNPs, it is possible to impute or infer all missing SNPs on animals of the same breed that have only been genotyped for a panel of sparse SNPs. Therefore, we were able to impute 700,000 SNPs on all genotyped animals regardless of the SNP chip with which they had been genotyped because we had genotyped a sample of each breed with the 700k SNP chip. This is important because we do not expect that a prediction equation based on sparse SNPs will work in multiple breeds. The accuracy of imputation is shown in table 4.

Table 4. Accuracy of imputing dense genotypes from sparse SNPs

Data source	Imputation	Accuracy (r2)
7k	7k to 50k	0.76
7k	50k to 700k	0.88
50k	50k to 50k	0.97
50k	50k to 700k	0.90
10k	10k to 50k	0.88
10k	50k to 700k	0.96

Phenotype data

Of course not all cattle are measured for all traits. For instance, cows that were recorded for lifetime reproduction cannot also be measured for meat quality. The number of animals that were measured for each trait and genotyped is given in Table 5.

Table 5. Number of animals with phenotypes and genotypes

Trait	Number of cattle measured	h2	mean	sd
Net feed intake	4559	0.35	-1.4	2.1
Tenderness (LDPF)	6151	0.22	4.5	1.1
Scanned fat depth at feedlot exit	5162	0.49	11.2	4.5
Retail beef yield	3255	0.38	67.0	3.4
Marbling (cimf)	6267	0.31	3.6	1.9
Scanned eye muscle area	4905	0.14	68.0	10.9
Carcase weight	6811	0.39	282.7	56.8
Hip height post-weaning	6668	0.55	120.3	8.1
Liveweight post weaning	10690	0.39	238.2	55.4
Age at puberty in heifer	2115	0.53	700	110
Post partum anoestrous interval	1492	0.40	160	60
Fertility when lactating	4287	0.09	0.69	0.46
Scrotal circumference at 12 months	1112	0.60	21	2.7
Percentage normal sperm	964	0.22	73	22

Genome wide association studies

The SNPs that are assayed using SNP chips are not expected to cause variation in complex traits. However, we expect that causal mutations anywhere in the genome will be close enough to at least one of the SNPs that they will be inherited with the SNP most of the time and hence in LD with it. This will generate an association between the SNP and the trait that the causal mutation affects. By using SNPs covering the whole genome we can detect and map the causal mutations for our target traits. Such an experiment is called a genome wide association study or

GWAS.

For all traits studied, the GWAS found that there are many associated SNPs scattered throughout the genome indicating many mutations that cause phenotypic variation. Using a Bayesian analysis, we estimate over 2000 SNPs are independently associated with each trait. This means that the apparent effect of most SNPs is very small and constantly hard to detect and to estimate accurately. Nevertheless, some SNPs have larger effects which are highly significant. The number of significant SNPs is shown in Table 6.

Table 6. Genome wide association studies

Trait	Number of SNPs significant (p<0.0001)	FDR (%)
Net feed intake	580	12
Fat depth	599	12
Tenderness	997	7
Retail beef yield	193	36
Marbling	360	19
Eye muscle area	165	42
Carcase weight	1472	5
Hip height	2823	2
Live weight	1666	4
Scrotal circumference	97	5
Percent normal sperm	69	6

FDR = false discovery rate

An example of GWAS results is shown in figure 2. This shows the strength of association between SNPs on chromosome 5 and hip height. There is a highly significant association ($P< 10^{-15}$) at about 47 Mb. This is near the position of the gene HMGA2. HMGA2 also contains SNPs associated with height in humans. Although this position contains the most significant association with height on chromosome 5, there are many other less significant associations visible in figure 2. Thus even one chromosome contains many genes affecting each trait. SNPs in the same position on chromosome 5 are associated with a range of traits including weight and feed conversion efficiency. Thus, the SNPs appear to map a mutation(s) with widespread effects on traits related to growth.

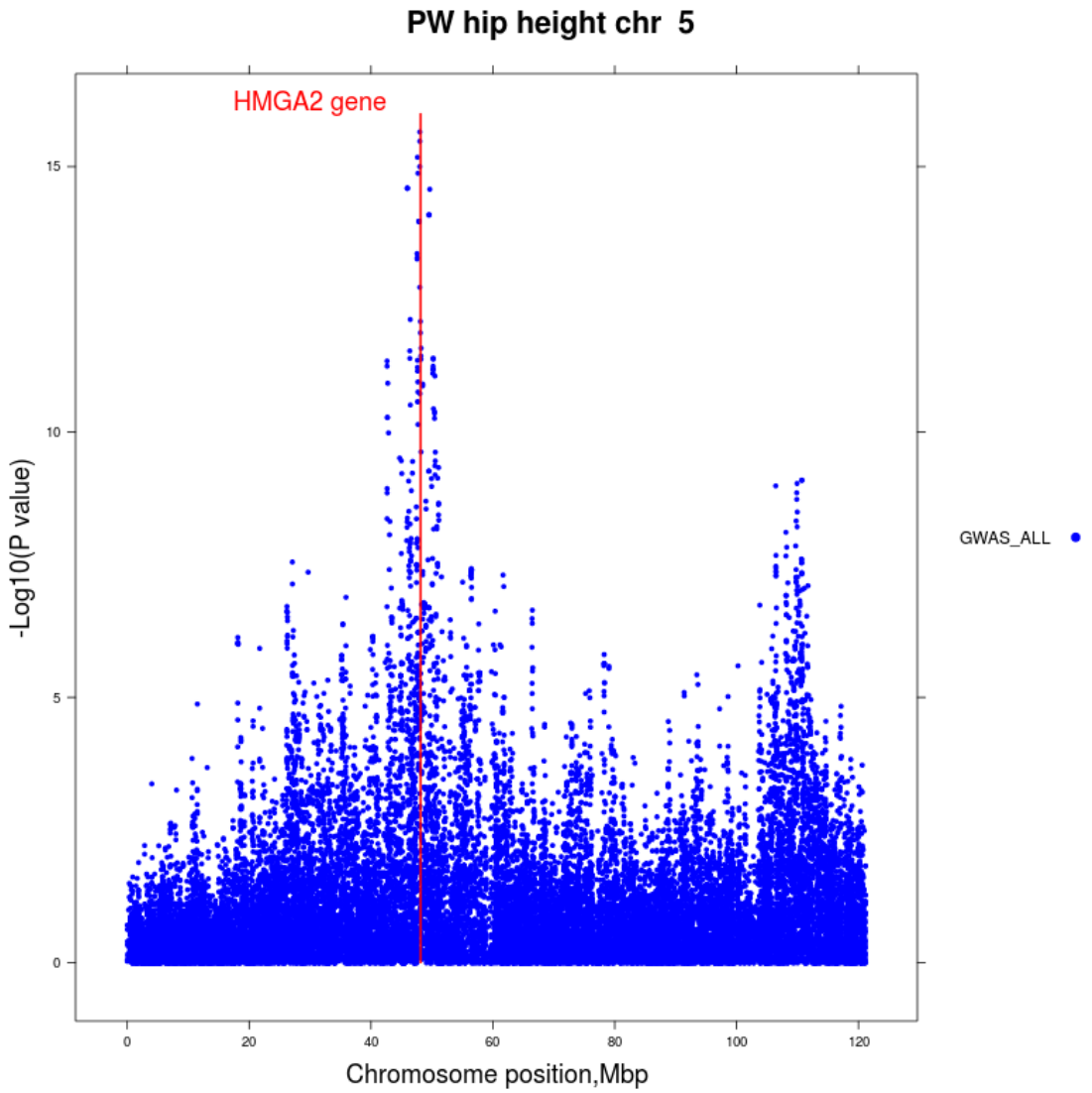

Figure 2. Association between SNPs on chromosome 5 and hip height. Each dot is one SNP.

We were able to use our GWAS results to answer the question “Are the same genes important in genetic variation in humans and in cattle”? Large GWAS have been done for height in humans and we had data on hip height in cattle which showed that there was a significant overlap between the genes discovered in humans and in cattle. This is useful information: It shows that the genes discovered in cattle are not false positives, it shows that there are physiological pathways controlling these complex traits and that the vast literature on human genetics is of value to cattle.

Prediction of breeding value from SNP genotypes

Although SNPs in the genes HMGA2 shown in figure 2 are associated with height, they only explain a small fraction of the genetic variance in the trait. Consequently, it is necessary to use SNPs from all over the genome to predict the breeding value of an individual. Research has shown that the best way to do this is to use a method called “genomic selection”. The research strategy was as follow. A large sample of cattle were genotyped for the genetic markers and measured for the traits of interest. From this data an equation that predicts the trait from the genetic markers was developed. This prediction equation was then tested in another sample of cattle to determine its accuracy. There are many breeds of cattle used in Australia and the CRC’s policy has been to sample a range of breeds and to develop a prediction

equation that can be used within any breed.

The accuracy of the prediction equations developed from 700,000 SNPs are shown in Table 1.

The ability to distinguish genes that come from B. taurus from genes that originate in B.indicus in tropical composite breeds could be used to select cattle that carry the desired gene (taurus or indicus) at each position in the genome. Surprisingly we found few genes where there was a big difference between an average taurus allele and an average indicus allele. Thus the difference between B. taurus and B. indicus cattle seems to be due to genes of small effect and perhaps just to difference in the allele frequency of these genes between the sub-species of cattle.

Plag1

Genome wide association studies discovered associations between many traits and SNPs near the gene PLAG1 on chromosome 14. Others had previously reported a mutation in PLAG1 that possibly caused an increase in the height and weight of cattle. We genotyped this mutation on some of our experimental cattle and imputed the genotype in other cattle that had been genotyped with one of the SNP chips. Analysis of this data showed that the PLAG1 mutation was associated with increased age at puberty in heifers and bulls, increased time to commence cycling after calving in cows, increased height and weight, decreased fatness and marbling, increased feed conversion efficiency and decreased IGF1 concentration in blood.

We were surprised to discover that this mutation occurs in both Brahmans and B. taurus breeds. In our B. taurus breeds, the mutation has a frequency of 0.9 but in Brahmans it is 0.5. Investigation of the SNPs in the region surrounding PLAG1 on chromosome 14 showed that the mutant allele in Brahmans had come from B.taurus. Presumably, during grading up of Brahmans from B. taurus cows the mutation had been introduced and inadvertently selected by Brahman breeders. Unfortunately, as well as increasing mature size it also decreases fertility. In future Brahman breeders can decide either to select for the mutation or against it depending on their breeding objectives.

Feed conversion efficiency, carcase and meat quality

Most of the genetic variation for carcase and meat quality traits is due to a large number of genes each with a small effect but there are some exceptions. Tenderness, measured by the force needed to cut through a sample of cooked meat from the longismus lumborum muscle show evidence in the GWAS for three genes of moderate effect. Two were already known (calpastatin and calpain 1) but a new finding is the effect located near the gene corin. These genes and all others tracked by the 700,000 SNPs are included in the genomic estimated breeding values calculated for these traits.

As well as the genome wide association studies, we also investigated candidate genes whose known function suggested a role in these traits. Polymorphisms were discovered in these genes and tested for an association with traits measured in the Jersey x Limousin project at University of Adelaide. Some genes were significant and these are being tested in the wider CRC data.

Cow and bull fertility

Numerous genes contain SNPs associated with male or female fertility. Amongst those with larger than usual effects are SNPs in the gene inhibin alpha (INHA) associated with the concentration of inhibin in the blood and SNPs in the gene IGF1R are associated with age at puberty in heifers. These associations are likely to reflect causal polymorphisms in those genes (INHA and IGF1R).

Project 7.4 Analytical Support

This project assisted staff in programs 1, 2, 3 4, and 7 with the design and analysis of major experiments

Project 7.5 Bioinformatics

This project provided all CRC scientists with bioinformatic support.

Appendices

Report on the accuracy of genomic selection and GWAS for feedlot, growth and carcase traits in Beef CRC cattle

Genome wide association results from the CRC for Beef Genetics Technologies

Candidate gene SNP associations

Reproduction final report