Developing genomic resources for whole genome selection

J.C. McEWAN

AgResearch, Invermay, Mosgiel, New Zealand

ABSTRACT

After more than 20 years of research, genomic technology is now on the cusp of widespread adoption in animal industries. Industry adoption will initially involve improvement of existing breeding and traceability systems. DNA variants, commonly called markers, used in these systems can provide 4 pieces of information: identity, parentage, prediction of an animals genetic worth and its individual performance. The usability of the technology and its adoption is increasing as genotyping prices decline. The major challenges are: to decrease sample collection, processing and reporting costs, and integrate the systems so that all 4 separate pieces of information can be obtained from a single test. Underpinning future research will be availability of sequenced genomes for each of the production species, and large scale "SNP chips" derived from them. Currently, the technology provides the ability to identify and map and use variants affecting animal performance in commercial populations. These technologies use existing natural genetic variation already present in the population. The potential of these technologies is safer and more nutritious food produced with less environmental impact and greater animal welfare due to lower level of disease.

Keywords: SNP; genome sequence.

WHAT ARE DNA MARKERS?

Genomics is the study of DNA. DNA encodes the instructions on how an animal grows, reproduces and responds to its changing environment. This includes how it digests and stores nutrients and counters disease challenges. In farmed animals these instructions are encoded in chromosomes or segments of DNA. Each letter or nucleotide of DNA potentially stores an instruction and there are about 2.7 billion of them in every cell of a cow. In fact there is double that number, because each animal gets a copy of the instructions from its mother and another from its father. The instructions are not exactly the same. On average, if two segments of DNA are compared about one nucleotide in every 500 to 1000 differs. That is about 2.7 to 5.4 million differences! In fact, these differences are constantly generated and each offspring also has 20-50 new variants. These differences underlie the observed genetic variation. How these differences affect the performance of an animal is the key area of study in genomics. Most of the variants, perhaps as many as 93 percent, have little or no effect, another 6 percent have only minor effects, with only perhaps less than 1 percent contributing the majority of the observed genetic variation for all traits. However, that is still 50 to 100 thousand variants plus perhaps a new one naturally created each generation. For a given trait it has been estimated that typically 5-50 variants control more than 50% of the observed genetic variation.

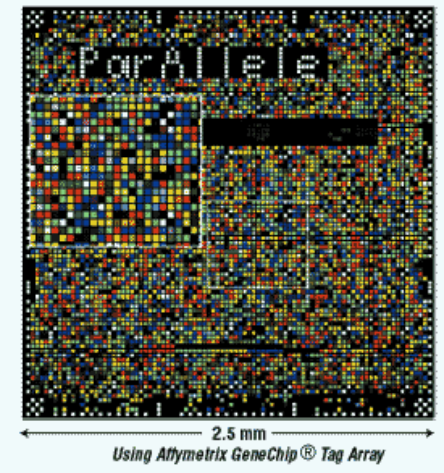
There are several different kinds of DNA variants, but only two kinds are commonly measured. The first are Single Nucleotide Polymorphisms (SNPs), where individual nucleotides are altered. The second are microsatellites where short repetitive segments of repeating bases, typically in units of 2 to 5 nucleotides long, vary in length. The latter typically have little effect on performance, but are highly variable and have historically been used for that reason. More recently SNPs are becoming more widely used for reasons detailed later.

SEQUENCING THE GENOME

Over the last 4 years perhaps the most dramatic advances in animal genomics has come via genome sequencing. Chicken and cattle have been sequenced (Kappes *et al*., 2006), pig is starting to be sequenced, and sheep is being physically mapped in preparation of being sequenced. Partial sequencing with a 3 times coverage using a whole genome shotgun approach with a Roche 454 FLX sequencer may commence this year (http://www.sheephapmap.org/). Costs of genome sequencing are dropping dramatically with new technology. A respectable skim genome sequence can now be done for between 5 and 10 million dollars. Sufficient sequence to identify several hundred thousand SNPs can now cost as little as 2 million dollars. However, even these costs are well beyond the capacity of most individual research groups and in fact most countries research infrastructure. This has led to the formation of international consortia to jointly fund and

undertake the work. The benefits of these consortia are considerable. Other than jointly providing funding and infrastructure, they constructively focus attention on critical basic research priorities for a species while at the same time formally adhering to prompt release of information to all researchers. It also formalizes previous informal collaborations across countries. These additional factors mean the benefits are magnified several fold more than the simple financial leveraging obtained by each group.

Figure 1: A "SNP Chip" each square measures the variants present at a particular genome location, a single chip can have tens of thousands of squares.



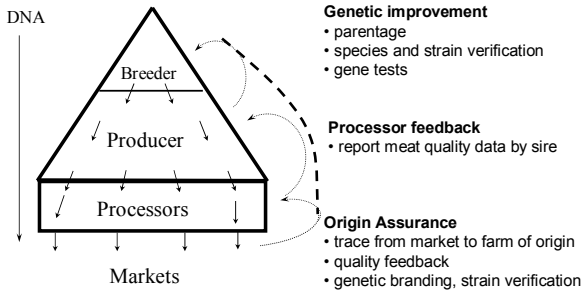
The major immediate outcome of having a genome sequence is that the process also identifies hundreds of thousands or millions of SNP markers and provided their genome location as well. Both pieces of information are essential for effective use. Previously markers had to be discovered and then mapped in a laborious and costly procedure. The variants position relative to individual protein encoding genes and associated control regions affecting production are also identified. SNP markers have the benefit that tens to hundreds of thousands can now be tested on a single animal at once and the measurement and scoring process is more amenable to automation. Currently, there are two major competing technologies, one from Affymetrix (http://www.affymetrix.com/index.affx) and another by Illumina (http://www.illumina.com). Commonly called "SNP Chips", see Figure 1, this new technology is already having dramatic impacts on cattle research even before the bovine genome has been fully completed. Based on advances in human research, prices have declined to less than 1 cent per SNP. This represents a 100 decrease over the last 5-10 years and nearly 1000 fold over the last 20 years. We expect these prices to continue to decline. It is for this reason that we expect SNP markers to largely replace alternatives over the next 5 years. As costs continue to decline additional applications become economically viable.

HOW CAN DNA MARKERS BE USED?

Figure 2 shows a typical animal industry. At the top is a nucleus breeding replacement males for the commercial tier. Typically, this unit is much less than 1% of the size of the commercial tier breeding progeny for slaughter and other products. Below that may be a fattening or feedlot tier, a processing and marketing tier and finally the end consumer. In all cases DNA flows out from the breeding tier and information returns from sales flows back upwards in a series of loops, some just between adjacent tiers and potentially also across several tiers. It is aiding the flow of information at all levels where DNA is of assistance. Some of this information can be made available by other means, but typically it cannot be easily audited and incurs substantial costs. We will look at how each component can be enhanced via use of DNA information.

Figure 2: DNA based information systems can aid information flow back up the production system at each stage, but its greatest value is obtained when the entire chain can be closed (dashed line), because breeding decisions can be optimized, and it provides full farm of origin assurance.

"Closing the loop"



TRACEABILTY AND IDENTITY

One use of DNA variants is to utilize them as a bar-code of an individual. This is the same technology as used in police work. It has the benefit that it is permanent, cannot be lost or modified, is present in every part of the animal and can be retrieved even when individual parts are not tracked or have been stored under less than optimal conditions. The problem has been the expense of collecting, storing and analyzing samples. The technology has been perceived as a cost to the

industry rather than a benefit. Sporadic industry implementation often only addresses one specific production tier or channel rather than a "farm of origin" or "paddock to plate" whole industry approach.

The ability to DNA audit a paper based traceability system is essential for "trace back" and "trace forward" when disease outbreaks, like BSE, or other events occur. This is because they allow the outbreak to be verifiably contained rather than put an entire countries production system at risk. Similar comments apply for breed or country of origin enquiries. Passing off products as originating from other countries is increasingly prevalent. In a recent case of label counterfeiting the only distinguishing characteristic was the detailed chemical composition of the plastic wrapping.

The trick is to make the costs negligible! This means only testing when required, and leveraging on samples collected for other uses of the technology. Perhaps the best approach is when all stud tier animals are already DNA tested for other reasons, and a sample of each commercial sire has then already been collected, stored and genotyped. If the sire location is also recorded then in theory any edible animal product from its progeny can be tracked back to the farm of origin in an auditable fashion by obtaining the DNA profile of the product, and rather than matching it to an individual, matching it to its sire. The commercial farming, processing and marketing tier incur no ongoing costs other than their existing paper based systems except when a DNA trace is actually required. In practice a two tier screening is required, because the combinations of markers commonly used today have insufficient power to always uniquely match a sample to an individual sire. Therefore a small number of additional tests are required in a second round of testing. However, as test costs continue to decline one stage testing will become a reality.

An example of an industry close to this situation is the New Zealand deer industry where over 30% of the stud tier is already fully tested for DNA parentage, and much of this component is vertically integrated. These portions of the industry are now in a position to implement this system.

Another area that has impeded progress has been tracing ground product, such as hamburger mince, sourced from many individuals. Recently techniques have been developed that allow DNA based techniques to undertake this task and allow "paddock to patty" traceability (Shackell, G.H. pers. comm.). However, in this case additional paper based records at the processing plant are needed. Audits in meat processing plants have found that paper based systems often have errors in the vicinity of 10% and again periodic DNA testing provides an independent audit of its efficiency (Heaton *et al*., 2005).

PARENTAGE

It is often overlooked, but using DNA to identify the parents of an animal is extremely valuable in many sire breeding contexts. Obviously manual processes are subject to error and typically estimated as 5-20% under commercial conditions. This reduces the rate of genetic gain. However, for extensive species other benefits are more important. It allows mob mating with consequent higher fertility, through elimination of sire failures, and closer spread of parturition dates. In the grazing context it reduces the number of mating mobs and thereby improves pasture management. It reduces disturbance at birth, thereby improving bonding of offspring with their mother. This is also usually a time of a labour shortage on extensive properties, so it reduces the need to hire extra staff and allows what is available to concentrate on activities that increase offspring survival. It also provides an opportunity to shift sire breeding operations to more extensive properties, freeing up intensive properties to fatten livestock. In the New Zealand context, the major benefit in deer is the reduced disruption at fawning, in sheep it is pasture management at mating, and in cattle it is animal handling. With the declining costs of DNA marker testing these benefits are now approaching the cost involved and this has meant a rapid adoption of the technology: in sheep from nearly nothing to 20% of the ram breeding industry in less than 5 years. Over a longer period deer have increased to 30%.

A key reason why industry adoption has occurred is because pregnancy, litter size and mating date information are now available in most flocks via ultrasound pregnancy diagnosis. This is an infrastructure that has also developed over the last 20 years.

DNA pedigrees and their use are now being fully integrated into genetic breeding systems (Dodds *et al*., 2005). Unique pedigrees are not required in all cases. In fact, little or no genetic progress is lost when using a mixture of unique and partial (*i.e*. progeny may have several probable parents) parentage records, if the majority can be uniquely assigned. However, genetic evaluation software has to be modified. The benefit is that marker testing costs can be reduced in the interim, while they contribute a significant fraction of the overall price.

STREAMLINING THE SYSTEM

The widespread adoption of these technologies highlights a critical aspect. Cost reduction and efficiencies depends as much on reducing sampling, data accessioning, DNA extraction, and reporting costs as on the DNA marker technology itself. This is especially important for lower value animals like sheep and countries with low cost production systems.

In practice DNA collection is linked to electronic tags, which are being implemented in the entire industry anyway as part of a national identification system. The DNA samplers (Figure 3) need to be low cost and when collected have to be able to be stored for years at room temperature. Freezers break down. They are also labelled with bar codes and this in turn offers the opportunity that all subsequent steps can be automated: including the provision of the results directly into the appropriate genetic evaluation databases. The storage of samplers is a critical aspect as it is required as part of any DNA auditing system and to allow for changes in genotyping technology.

MARKER ASSISTED SELECTION

In certain industries, such as dairy, DNA tests for deleterious recessive genetic traits have been available for many years (Crawford *et al*., 2007). More recently, markers associated with or actually causing modified performance have also become available (Morris *et al*., 2007). Usually, only one or two per trait are available and they typically only account for a minor proportion of the genetic variation. In the case where the variant has been identified selection is reasonably simple; the variant is tested and its status is included in the evaluation as a fixed effect. However, in many cases the variant has yet to be discovered, but a segment of DNA has had enough markers typed to identify a pattern of variability called a haplotype that is predictive. These patterns have to be confirmed in each new resource or breed that is used (Campbell & McLaren, 2007). The haplotypes are used in a similar manner to the known genes to predict performance as part of a genetic evaluation. Examples now include a number of variants affecting carcass muscling and fatness in beef cattle and sheep, lactation traits in dairy cattle, and even meat quality traits such as tenderness. Adoption has been driven by cost and perception of benefits.

While the rate of discovery and fine mapping of these variants has increased, this process takes a number of years and often requires specialized family structures. The process is expensive and can only be justified commercially for hard to measure traits that are sex limited, measured late in life or require the animal to be slaughtered. The variants also need to have reasonably large effects, because the discovery cost increases as effects get smaller.

We are now reaching the situation where most large obvious effects for key traits have at least been identified, if not yet progressed where they can be used commercially. Obviously an alternative is needed both to decrease the cost of discovery, and allow commercial industry animals to be used. This is the hope that the high density SNP genotyping previously mentioned offers. Put simply enough markers are used to allow segments of DNA to be traced over many generations, even in the absence of knowledge about pedigrees (Hayes & Goddard, 2007; Dodds *et al*., 2007).

Figure 3: DNA sampling onto DNA samplers that can be easily collected, stored at room temperature for extended periods, are coupled with barcode sample IDs, and linked to electronic animal tags. It is all about reducing cost and increasing accuracy.

These little signposts in turn allow us to determine if that region also has an association with traits of interest. Not only that, unlike traditional breeding schemes, it can also easily identify if certain variants located in different parts of the genome actually interact with each other. Typically, these combinations predict an individual's current or future performance, rather than their breeding worth, but complicate the prediction of the latter.

As there are so many variants detected by this method, the properties of them as a group become more important than their individual effects, so thresholds can be lowered. It matters little if a specific variant fails under some circumstances as long as the majority of the variants are predictive. This is the principle of whole genome selection.

The benefits are animals can be tested at a very young age and the genetic worth of the individual can be estimated with a level of accuracy that exceeds individual measurement and approaches progeny testing. How can this be a benefit? Well typically it can be used in three ways. For difficult to measure traits such as meat eating quality or disease resistance it can largely, but not entirely, replace measurement. These traits typically have little selection pressure placed on them currently, because of lack of measurement so genetic progress can be increased. Secondly, for sex limited traits or traits measured at older ages such as milk production and composition, it allows the potential to identify better candidates for progeny testing and/or allows selection at younger ages. In the first case the realized selection intensity increases and in the second the generation interval is reduced. In all three situations the rate of genetic progress for traits of interest can be increased.

Currently, the genotyping cost is very high so its potential is best used where the genetic gain can be disseminated widely and the existing measurements and progeny testing are expensive. Perhaps not surprisingly, dairy bull evaluation and selection is likely to be one of the first to examine this technologies potential. In other industries a layered strategy is likely to be applied, until costs decline, with high density SNP chips on key animals, coupled with perhaps 100 to 1000 selected SNPs measured on remaining animals in the stud tier.

One of the major challenges of the technology will be to cost effectively handle and process the information produced and optimize breeding structures to make best use of the technology. The mix of DNA pedigrees with some level of uncertainty, and many thousands of markers will require further development for robust industry use, as will pedigrees where only some of the animals are measured for performance traits and DNA markers. This is not an impossible barrier, but it will involve further research.

Finally it is important to not oversell the technology. Its benefits will be best captured in well structured industries already making near maximal genetic progress. Similarly, if DNA samples are already being collected for parentage determination and as part of a traceability auditing scheme, this technology can be introduced in a cost effective manner in staged steps as the technology develops. The links with the existing performance recording and genetic evaluation system will already be in place. It has taken 20 years development for animal genomics to get to this stage so it will be at least 5-10 years before marker assisted selection or whole genome selection is widely adopted in industry segments where it is cost effective. The major shift over the last 5 years is that it is now actually being trialled or implemented in industry.

In summary we have limited marker assisted selection being undertaken now, largely for major identified QTL with one or maybe several regions per trait, "bolted on" to existing genetic evaluations based on trait measurement. The future with extremely high density mapping will allow us to examine and select the genome as a whole for multiple traits simultaneously using industry breeding structures. Essentially, the animals breeding worth is the sum of its parts and for the first time we will be measuring the parts rather than merely measuring the total.

INTEGRATION OF THE TECHNOLOGY AND CONCLUSIONS

It is now obvious that as costs reduce and technology improves DNA technology will be increasingly used on an industry wide basis. It increases information flow back to the breeder thereby allowing better breeding decisions, provides parentage information, likely genetic worth of individuals and provides an independent method to audit traceability. These benefits already exceed costs in some industries. Experience has shown these systems need to be organized so that key animals are tested once with subsequent data analysis, and data transfers fully electronic and integrated into the genetic evaluation system. In future it will be the data processing, analysis, and electronic storage and transfer of results that will be as much the challenge for increasing industry adoption rather then the DNA marker measurement technology.

ACKNOWLEDGEMENTS

This paper is derived from discussions over an extended period of time with many people particularly those from the research group I am part of including: Allan Crawford, Rayna Anderson, Grant Shackell, Ken Dodds, Benoit Auvray, Jude Sise, Mike Tate, Sheryl Newman, Richard Hall and Mary McEwan, but also many others. I have merely provided an overview to set in context the other papers in this session.

Current status of QTL and association studies in New Zealand cattle, sheep and deer

C.A. MORRIS, A.W. CAMPBELL[2], N.G. CULLEN, G.H. DAVIS[2], J.M. EVERETT-HINCKS[2], R.J. HALL[2], H.M. HENRY[1], P.L. JOHNSON[2], J.C. McEWAN[2], S.H. PHUA[1] and T. WILSON[1]

AgResearch, Ruakura Research Centre, Hamilton
[1] AgResearch Molecular Biology Unit, Department of Biochemistry, University of Otago, Dunedin
[2] AgResearch, Invermay Agricultural Centre, Mosgiel

ABSTRACT

In the last 15 years, a series of experiments has been carried out to search for quantitative trait loci (QTL) in New Zealand cattle, sheep and deer. A QTL represents the position on a chromosome where there is a statistically significant genetic effect on a measured trait. This paper reviews data published so far from these QTL trials and from various association studies. Although the measurement of traits (or 'phenotypes') and the sampling of tissue or fluid for DNA may be complete, the genotyping of DNA is still incomplete in most studies; further work may be expected as new DNA genotyping techniques and new genomic data become available, and as funds allow. In cattle, we describe the current status of research for carcass composition, meat quality, pubertal traits, milk yield and milk composition traits, and resistance to facial eczema and bloat. In sheep, we include QTL searches for disease traits (resistance to nematode parasites, facial eczema, ryegrass staggers, and footrot), muscling and carcass composition, reproductive traits, wool traits, and lamb survival. In deer, QTL searches have been carried out for live weights, seasonality and pubertal traits, using measurements from an interspecific hybrid. Generally, significant results have been followed further by fine-mapping and independent validation, before release to industry. Some of the mapping techniques will be described, with examples. So far, QTLs under study in New Zealand have led to the identification and use of gene-tests or marker-tests for meat tenderness and carcass composition in beef cattle, milk yield and composition traits in dairy cattle, and meat yield %, muscling and litter size genes in sheep.

Keywords: cattle; sheep; deer; QTL.

INTRODUCTION

In the last 15 years, a series of experiments has been carried out to search for quantitative trait loci (QTL) in New Zealand cattle, sheep and deer. This paper describes the trial design of these major animal experiments, established to search for DNA markers linked to production traits, and ultimately for the causal gene variants. A QTL represents the position on a chromosome where there is a statistically significant genetic effect on a measured trait. We describe the approximate size of each study, and its time period and the trait groups recorded. Given the research climate in which we find ourselves in New Zealand, many unpublished or unpatented findings are retained by the research groups involved, so we cannot expect to be complete in reporting and summarising results at this stage. We have endeavoured to cite the published QTL and genes, and those citations give more detail of the trial design and methodologies used.

RESULTS AND DISCUSSION

Cattle

Carcass composition and meat quality

Table 1 (#1) shows a double-backcross experiment with beef cattle, being managed by AgResearch in collaboration with the University of Adelaide. Six first-cross Jersey x Limousin sires were generated, and were used to breed backcrosses from both breeds of straight-bred dams. The offspring were phenotyped primarily for carcass composition and meat quality traits, but many additional traits were also recorded,