A Genomic Information Nucleus to Accelerate Rates of Genetic Improvement in Sheep

J.H.J. van der Werf*† and R.G. Banks*‡

Introduction

Any breeding program has to find the right balance between investment in information and obtained rates of genetic improvement in the wider population. It was already pointed out by Robertson and Rendel (1950) that “the apparent lack of connection between an animals’ breeding value and its phenotype led many to advocate more complicated methods for judging genotype”. This led to their proposal of a progeny testing scheme for dairy, as it is particularly hard to select dairy bulls directly based on phenotype. The use of artificial insemination (AI) along with the fact that the main traits are sex limited has led to progeny testing becoming prevalent in dairy breeding systems. It is interesting to note that with the advent of genomic selection tools, it was quickly pointed out that these dairy breeding programs could also greatly benefit from this new technology (Schaeffer, 2006). Indeed, whether testing progeny or genomes, these are examples of ‘more complicated methods’ to increase selection efficiency, and breeding programs have been and will have to be designed to exploit such tools optimally.

Another important breakthrough in breeding program design was the proposal of closed nucleus breeding schemes where selection of young animals based on less accurate breeding values could give accelerated rates of improvement (Nicholas and Smith, 1983). In their case, a new design was proposed to use more optimally the emerging technology of artificially increased female reproductive rate, through multiple ovulation and embryo transfer (MOET). The importance of the paper was that it highlighted the importance of the balance between generation interval and selection accuracy and that this should be optimized, possibly leading to selection of young animals based on less accurate information. Although MOET nucleus schemes seem to achieve the opposite effect to progeny testing schemes in terms of accuracy (r) and generation interval (L), - they reduce both r and L - both developments have demonstrated that the ratio r/L needs to be optimized and that the optimum can be critically affected by new breeding technologies. It is easy to see how selection based on DNA information will affect this ratio as it provides scope for obtaining more information about breeding value early in life, thereby changing the optimal balance toward shorter generation intervals. Economically, the important feature of nucleus schemes is that investment in measuring a limited number of animals can greatly benefit genetic improvement of a much larger commercial population. Bichard (1971) referred to this as the “selection-multiplication overhead”.

* Coop. Research Centre for Sheep Industry Innovation, Armidale, NSW 2351, Australia
† School of Environm. & Rural Science, University of New England, Armidale, NSW, 2351
‡ Meat & Livestock Australia, c/o. Animal Science, UNE Armidale, NSW 2351, Australia
A third balance that has to be found in breeding programs is the balance between traits. Typically, selection index methods are proposed for multiple trait selection (Hazel, 1943) but it is well known that the method optimizes selection conditional on the information that is available to select on. It does not work out which information should be collected to maximize response. In practice, especially where most recording is done by small-medium businesses, as is the case in sheep and beef cattle breeding, breeders tend only to record the cheaper and easy-to-measure traits, and those for which they are most likely to receive remuneration via premiums for bulls or rams superior for those traits. The result is that selection responses are obtained for traits with a lot of information, whereas little response is obtained for other traits not typically measured, unless they have favourable correlations with traits selected for. In the worst case, negative responses are obtained where correlations are unfavourable. For example it is hard to prevent a decline in fertility in dairy cattle without accurate breeding values for fertility if there is strong selection for milk production. Investment in hard-to-measure traits can be worthwhile to overcome this problem.

Sheep production systems are predominantly pastoral based and extensive in nature. Most sheep breeding programmes have a relatively flat structure, meaning that compared with intensive livestock industries many operations participate in genetic improvement, thereby forming a wide base for the nucleus breeding sector. Although the nucleus tier uses AI widely, further dissemination is mainly by natural mating resulting in low multiplication factors and the number of commercial expressions resulting from most nucleus rams is relatively low. This makes it difficult to invest in information, making it difficult to implement breeding programs that focus on quality traits (both on product side and animal side) as this would require information about traits not normally measured.

This paper discusses methods for genetic improvement in sheep where the need for more measurement on traits is considered. The ‘information nucleus’ concept, earlier introduced by Banks et al. (2006), is further developed here and more formally evaluated. The scheme proposes a progeny testing facility where sires are sampled strategically, and phenotypic measurements are collected on their progeny especially for traits not commonly measured in commercial studs. Sampling of sires can be optimized to maximize the information generated for the overall breeding program. The potential value of this information is assessed. Genomic selection potentially gives a powerful additional dimension to this concept, as more information is available about the genomes of a broader range of selection candidates. We discuss how such a design can be implemented to address the need for improved measurement of breeding objective traits, and we estimate the potential benefits.

**Methods**

**The Information Nucleus concept**

The Australian sheep industry jointly with R&D funding from the Australian government invests in an *Information Nucleus Flock (INF)* (Banks et al, 2006; Fogarty *et al.*, 2007) to obtain data on new traits and for genome wide association studies. Each year, about ~5000 ewes are mated to ~100 new sires and progeny are born and raised in 8 flocks across the range of environments in Australia. The objectives of this information nucleus project are 1) to estimate accurately genetic parameters for new traits, and their correlations with existing
traits, 2) estimate the extent of genotype by environment interaction, 3) to generate information for Australian breeding values (ASBV’s) for new traits for breeding animals and 4) to undertake a whole genome association study to explore the potential for genomic selection. The current CRC information nucleus design allows for 5 matings over a 5 year period, i.e. a total of nearly 22,000 matings generating about 20,000 progeny. A significant research question concerns both the value and optimal form of an ongoing form of an INF (likely based to a greater extent on industry investment). Objectives 1 and 2 are mainly related to parameter estimation which does not require a large scale intense ongoing measurement effort. However, objectives 3 and 4 are contributions to a breeding program on an ongoing basis and the value of such an investment needs to be explored.

The new traits considered broadly cover the areas ‘meat’, ‘wool’ and ‘easy care sheep’. New traits are characteristics normally not measured in commercial studs, but they may potentially have economic relevance either as objective traits or as selection criterion traits. For example, most of the new traits in relation to meat sheep efficiency are traits of the carcass, meat quality or meat nutritional content. For wool we measure staple strength, colour and a number of visual traits, including some relating to breech wrinkle, which could be useful selection criterion trait in a breeding program for sheep that require no mulesing. ‘Easy care’ traits relate to lamb survival (incl. maternal effects), and ewe resilience, e.g. parasite resistance, reproductive ability in multiple lambing, ewe efficiency (incl. feed efficiency and methane production efficiency) and longevity, measured in a range of environments. A proportion of the progeny is slaughtered with detailed measurements on carcass and meat, whereas a smaller proportion is kept for wool measurement (whether and females) and reproduction and ewe efficiency measurements.

The Australian sheep industry consists of sheep from a wide range of sheep breeds, or breed crosses. Wool production is largely due to purebred (fine wool) merinos. Merinos are often crossed to maternal breeds, mainly Border Leicester, to produce first cross ewes with improved maternal ability. These “first-cross ewes” are often crossed with terminal sire breeds, mainly Poll Dorset and White Suffolk and with some minor breeds like Texel and Suffolk, to create prime lambs. Due to the decline of wool prices and increase in meat, a number of new maternal or dual purpose breeds is rapidly gaining popularity in the industry, e.g. South African Meat Merino (SAMM), Dohnes, Coopworth, and the proportion of merino ewes joined direct to meat breed sires is rising. This variety of breeds creates somewhat a dilemma, as genetic parameters and genomic selection parameters are best estimated on a large number of individuals within a breed, but accurate estimation of breed specific genetic and genomic parameters is hard to achieve with a large number of breeds. For the moment, this issue is being tackled by focussing on parameters specific for each breed type (terminal sire, dual-purpose or maternal, and merino). The current CRC design is reflective of breed proportion uses in industry: about 40% of sires are from terminal breeds, 40% are merinos and 20% are maternal breeds. The dams are pure merinos (80%) or Border Leicester x merino crosses (20%). The number of progeny targeted per sire is about 40. Numbers of progeny measured per trait group and per breed type are given in Table 1.
Table 1. Number of animals per breed type as measured in current information nucleus used by Australian Sheep CRC (targeting 4000 lambs weaned per year).

<table>
<thead>
<tr>
<th>Sire Breed Type</th>
<th>Dam Breed type</th>
<th># Lambs</th>
<th># Slaughtered</th>
<th># Wool records</th>
<th># Wool and Reproduction records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminal (40%)</td>
<td>Merino (50%)</td>
<td>619</td>
<td>619</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bl-Merino (50%)</td>
<td>866</td>
<td>866</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal (20%)</td>
<td>Merino</td>
<td>1088</td>
<td>544</td>
<td>544</td>
<td></td>
</tr>
<tr>
<td>Merino (40%)</td>
<td>Merino</td>
<td>1425</td>
<td>356</td>
<td>1069</td>
<td>713</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>4,000</td>
<td>2,385</td>
<td>1,069</td>
<td>1,257</td>
</tr>
</tbody>
</table>

Selection of INF sires
INF sires are young sires with high genetic merit (within breed) such that they have a high likelihood of contributing to the future gene pool of the population. This is a first selection step, choosing about five times the number of sires that is ultimately selected. Of the highest ranking sires we select sires for use in the INF based on relationship and diversity within traits. Optimal selection is achieved with a genetic algorithm using the following objective:

\[ \sum_{i} \sum_{j} (D_{ij} - \bar{D}) (A_{ij} - \bar{A}) \]

where D is the deviation from the trait mean and A is the coefficient of the numerator relationship. Application of this criterion samples males that differ little in trait values but are distantly related, so that their similarity is less likely to be due to polygenes – giving a higher effective population size in that trait region, whereas animals that are distant from each other should be closely related, so that their difference is less likely to be due to polygenes, and more likely to be due to single gene effects (as for sib-pair designs) (Brian Kinghorn, pers. communication). The distance criterion was applied only for traits usually under selection (i.e. those that contribute most to the merit index) to attempt ‘de-selection’ among the candidate young sires, such that overall a higher genetic diversity as well as a high trait diversity is achieved. In addition, the algorithm tries to balance trait-relationship diversity with a high number or paternal grandsires. With genomic selection potentially becoming an important tool, selection of sires could be further optimized if selection candidates had ordered genotypic data available. Sire selection could then attempt to maximize the number of haplotypes sampled from the population (across genomic regions).

The effect of an INF on genetic gain without genomic selection.
In this section we calculate the effect of an INF on the accuracy of selection in the breeding nucleus. In general, the information provided by the INF to the breeding program will be a function of the correlation between animals tested in the INF and the breeding values of selection candidates. In absence of genomic selection, this correlation will be purely determined by the additive genetic relationship between the animals tested, and selection candidates. Usually in breeding programs we test selection candidates directly, as this gives the closest relationship. However, in case of sex limited characters, or traits requiring slaughter, the options are to test sibs or progeny. Testing sibs has the advantage of the
information being available earlier in life, whereas progeny testing provides very accurate predictions of breeding value but at the expense of generation interval.

Consider a population of 200,000 stud ewes delivering 80,000 rams per year to the commercial population. Basically this is the multiplication tier, and a small proportion of these ewes can be considered a breeding nucleus. The multiplier factor in terminal sire breeding programs is relatively low, as beyond the nucleus most reproduction is by natural mating. We assume a multiplication rate of 40, based on 50 matings per ram, 0.4 saleable rams per breeding ewe and a replacement rate of rams of 50%. Hence, 200,000 multiplier ewes could serve a population of 8 million ewes and the size of the breeding nucleus would be 5000, annually mated to 100 rams. Currently such a clear tiered structure does not exist in most sheep breeding programs, as the stud sector operates as an open nucleus scheme. However, implementation of a progeny based performance testing scheme would require, or perhaps automatically generate, a more pyramidal structure.

Currently indexes for terminal sires in Australia are based on traits measured on live animals, such as weights and ultrasound scans for muscle and fat (e.g. Carcase Plus Index). A more appropriate terminal sire index, which is not (yet) used by the industry, would contain carcass traits in the objective and growth and ultrasound scan measures for fat and muscle as selection criteria traits. Such an index was derived with SHEEP OBJECT (Andrew Swan, personal communication) with three carcass traits having all a similar dollar value per genetic standard deviation (dressing percentage, muscle conformation, percent saleable meat yield and carcass fat), except that the value for carcass fat is negative and about one third as large. Young breeding animals have measurements for none of these traits at first selection.

Animals in the stud sector are measured for live traits and young animals will be based on own performance and information on sire, dam and 40 half sibs. If 100 sires (for one breed) are tested in an INF each year, each with 25 progeny and sires chosen such that each INF sire is chosen from each nucleus sire family, all young nucleus-born animals will then have a half sib with a progeny test for all objective traits. Progeny of half sibs have an additive genetic relationship of 1/8 with the nucleus born selection candidates. Alternatively we could test progeny of all 100 nucleus sires themselves, giving 25 HS tested for each selection candidate.

**The effect of an INF on genetic gain with genomic selection**

We used selection index theory to model genomic selection, following Lande and Thompson (1990), where the genomic information is modelled as an additional information source known at birth, and explaining a percentage of the additive genetic variation (\(V_{GTL}\)). This percentage will be larger if more animals are used in deriving prediction equations and when the heritability is higher (Goddard, 2007). Selection accuracy increases with age, and can vary between sexes. The age structure of a breeding program was optimized by truncation selection across age classes for each sex. Therefore, if genomic information was available, it is more likely that younger animals had high enough accuracies to be selected, and the optimal generation interval would be lower. Increased selection accuracy and decreased generation intervals were therefore modelled as contributing to the benefit of GS.
Two GS scenarios were compared: under GS1 it was assumed that $V_{QTL}$ is equal to the trait heritability ($h^2$). This is not a functional relationship, but a reasonable approximation of expected predicted accuracy of GS when phenotypic information on about 2500 animals is used for deriving GS prediction equations. Alternatively, $V_{QTL} = \frac{1}{2} h^2$, which would be achievable in an experiment with about 1000 phenotypic measurements.

**Results**

The accuracy of selection for an index based on carcass traits for animals that have only measurements on live traits is 0.521. With an information nucleus using young sires, nucleus born animals would have a slight increase in selection accuracy to 0.542 (4.1% increase) due to additionally having 25 offspring of a half sib measured for the full range of traits. Ranking based on this index would favour carcass traits considerably more, e.g. the predicted response in dressing percent, %saleable meat yield and carcass fat would change by 20%, 14% and -3%, respectively. The accuracy of EBV of the INF tested sires would be 0.882. Some of these sires would immediately be in a strong position to become nucleus sires. However, there would a small selection differential based on within family selection of INF candidate sires. If nucleus sires themselves are tested in the INF, each nucleus born would have 25 half sibs with the full range of traits measured. This would give an accuracy of 0.581 the predicted response in dressing percent, %saleable meat yield and carcass fat would change by 55%, 40% and -10%, respectively. Again, selection intensities would be limited since this additional information is based on family information.

**Table 2. Average accuracy across age classes and annual response to selection for meat sheep and fine wool merino indexes in scenarios with no (no GS), less accurate (GS1) and more accurate (GS2) genomic selection.**

<table>
<thead>
<tr>
<th>Trait</th>
<th>Pheno</th>
<th>$h^2$</th>
<th>Accuracy</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>no GS</td>
<td>GS1</td>
</tr>
<tr>
<td>Muscle conformation (mm)</td>
<td>0</td>
<td>0.25</td>
<td>0.66</td>
<td>0.69</td>
</tr>
<tr>
<td>Dressing %</td>
<td>0</td>
<td>0.30</td>
<td>0.26</td>
<td>0.46</td>
</tr>
<tr>
<td>Saleable meat yield %</td>
<td>0</td>
<td>0.30</td>
<td>0.33</td>
<td>0.48</td>
</tr>
<tr>
<td>Carcass fat depth (mm)</td>
<td>0</td>
<td>0.30</td>
<td>0.46</td>
<td>0.57</td>
</tr>
<tr>
<td>Post weaning weight</td>
<td>1</td>
<td>0.25</td>
<td>0.71</td>
<td>0.73</td>
</tr>
<tr>
<td>PW fat depth (US)</td>
<td>1</td>
<td>0.25</td>
<td>0.71</td>
<td>0.73</td>
</tr>
<tr>
<td>PW eye muscle depth (US)</td>
<td>1</td>
<td>0.30</td>
<td>0.76</td>
<td>0.80</td>
</tr>
<tr>
<td>Overall Merit ($\text{Index}$)</td>
<td></td>
<td></td>
<td>0.58</td>
<td>0.64</td>
</tr>
</tbody>
</table>

*Variance explained by genotypes as proportion of additive genetic variance is equal to the heritability ($h^2$) for GS2 and $\frac{1}{2} h^2$ for GS1. Pheno indicates whether or not a trait is measured.*

Results in Table 2 show for terminal sire breeds a 21% increase in overall response with accurate genomic selection (GS2) and a 11% increase with less accurate genomic selection (GS1). Obviously, individual carcass traits benefit greatly from GS as these traits are not measured on breeding animals. As a consequence, the GS response for traits measured post weaning is lower than with no GS. Results for a selection index for fine wool merino show a 38% increase in overall response with accurate genomic selection (GS2) and a 19% increase
with GS1 (Van der Werf, 2009). Although fine wool traits are highly heritable and can be measured, the adult expression of these traits is usually not recorded. Moreover, number of lambs weaned has limited measurability and is much more improved whereas staple strength, being an unmeasured trait, declines much less under genomic selection.

**Discussion**

The Australian Sheep CRC has greatly benefited from an investment in an INF for the purpose of estimation of genetic parameters for a large range of new traits. The benefits of the INF for industry also included an improved linkage between flocks or various research datasets, to the benefit of genetic evaluation of all sheep in Australia. We have not modeled these linkage benefits, but assume they might become less relevant if more genetic improvement is concentrated in a small breeding nucleus.

Whereas some accuracy might be gained in measuring INF progeny either as half sibs or as progeny of half sibs, it might hard to create sufficient selection space to exploit this information in a breeding program unless the INF is very large. More design alternatives could be explored, and predicted progress would need to account for the Bulmer effect (reduced variation among families after selection), which would further reduce the benefits from the small increase in selection accuracy. Hence, an information nucleus maybe of limited benefit in absence of genomic selection. It is clear that the additive genetic relationship between animals phenotyped and selection candidates should be as high as possible. This relationship could be one for traits that are easy to measure, i.e. the INF would be the same as the breeding nucleus, but for traits requiring sacrifice, or those measured late in life, information on sibs or progeny is required.

Genomic selection allows selection of candidates without being directly related to animals tested, assuming that is mainly based on predicted QTL effects in linkage disequilibrium with SNPs. More insight is still needed with regard to the information exploited in genomic selection. If individual QTL contribute greatly to genetic variation, an INF would be more focusing on QTL detection and evaluation, i.e. test individuals with contrasting genotypes. However, there is increasing evidence that genomic selection is largely based on improved estimates of relationship (e.g. Habier et al, 2009) and that most traits follow basically the infinitesimal model of inheritance, which implies that animals in a training set need to have some relationship to those selected, and that a reasonably large training set needs to be continuously measured for objective traits.

Banks and van der Werf (2009) evaluated cost-benefit of a number of scenarios for meat sheep genetic improvement in Australia at industry level, incorporating various combinations of additional performance recording via an Information Nucleus and implementation of Genomic Selection (GS). Based on this perspective, they also concluded that implementation of either IN and or GS will both require but also stimulate evolution of industry structure towards a more clearly defined nucleus:multiplier:commercial base model, and will further evolution of co-investment, from a mix of stud breeders, collective industry and taxpayer funds, to models likely to involve investment from processors, retailers and potential from genotyping companies.
Acknowledgement

Andrew Swan from AGBU is acknowledged for assisting with genetic and economic parameters.

References

Hazel, L.N. (1943) Genetics, 28:476-490.