CSIRO PUBLISHING

Animal Production Science, 2012, **52**, 92–99
http://dx.doi.org/10.1071/AN11120

Invited Perspective

Genomics and the global beef cattle industry[1]

E. J. Pollak[A,B], *G. L. Bennett*[A], *W. M. Snelling*[A], *R. M. Thallman*[A] *and L. A. Kuehn*[A]

[A]USDA/ARS[2]-US Meat Animal Research Center, PO Box 166, Clay Center, NE 68933, USA.
[B]Corresponding author. Email: e.john.pollak@ars.usda.gov

Abstract. After two decades of developing DNA-based tools for selection, we are at an interesting juncture. Genomic technology has essentially eliminated the potentially large negative impact of spontaneous single-mutation genetic defects as the management of recent examples in beef cattle have demonstrated. We have the ability to perform more accurate selection based on molecular breeding values (MBVs) for animals closely related to the discovery population. Yet the amount of genetic variation explained falls short of expectations held for the technology. Tests are less effective in distant relatives within a breed and are not robust enough for across-breed use. It is hypothesised that 'larger single-nucleotide polymorphism (SNP) panels' will help extend the effective use of tests to more distantly related animals and across breeds. Sequencing and imputing sequences across individuals will enable us to discover causative mutations or SNPs in perfect harmony with the mutation. However, the investment to revisit discovery populations will be large. We can ill afford to duplicate genotyping or sequencing activities for prominent individuals. Hence, a global strategy for genotyping and sequencing becomes an attractive proposition as many of our livestock populations are related. As we learned more of the complexities of the genome, the number of animals in discovery populations necessary to achieve high levels of predictability has grown dramatically. No one organisation has the resources to assemble the animals needed, especially for novel, expensive or hard to measure phenotypes. This scenario is fertile ground for increased international collaboration in all livestock species.

Received 21 June 2011, accepted 12 January 2012, published online 20 February 2012

Introduction

Global communities are becoming increasingly interdependent. Current events are shared in real time, economies are global and communications are instant. Scientific endeavour has long enjoyed global exchange of information through literature, scientific colloquia and exchange programs. In agriculture, many of our animal populations are now related to some degree. It is intuitively appealing to consider greater collaborations especially with increasing complexity of problems and pressures on agricultural sciences to deliver technologies to address the impending food crisis projected as a result of human population growth.

Background

The applications of DNA tools for selection and management hold great promise to enhance the response to selection and to improve efficiencies of management. The former can be viewed as an opportunity to increase the accuracy of genetic assessment of replacement candidates at critical selection points and the latter as optimally categorising individuals for targeted management strategies.

Selection

For marker-assisted selection (MAS), two categories of traits are considered. The first includes traits for which routine genetic evaluations are obtained from phenotypic and pedigree information and the second comprises traits without routine evaluations. The development of DNA tests for traits with information already available can enhance the accuracy of those genetic evaluations if the DNA information is seamlessly integrated into existing genetic prediction infrastructures and used to augment other sources of information. The increase in accuracy will be most pronounced in young animals with no recorded progeny, and, hence, has high value for selection of replacement animals. The magnitude of the increase in accuracy depends on available records on relatives, heritability and portion of heritable variation accounted for by the tests, (Lande and Thompson 1990; Thallman *et al.* 2009). For traits not routinely recorded or evaluated, the benefits are obvious and potentially allow for selection pressure to be applied to novel, yet economically relevant traits such as feed efficiency or susceptibility to certain complex diseases. Unfortunately, for predictive tests to be developed that achieve sufficient reliability, large discovery populations need to be constructed

[1]Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA.
[2]USDA is an equal opportunity provider and employer.

Journal compilation © CSIRO 2012

that measure these novel phenotypes. Consequently, we have witnessed tests being deployed that are based on limited information and that describe little phenotypic variation (Van Eenennaam *et al.* 2009).

Management

Marker-assisted management (MAM) is the process of making management decisions on the basis of the prediction of an animal's performance from marker panels. Examples include developing individualised implanting strategies to complement genetic potential, placing feeder cattle into pens on the basis of the risk of disease to provide preventative treatment, or on the basis of predictions of time to finish with the goal of optimising feeding programs. The DNA prediction equations for MAM would ideally account for both additive and non-additive genetic effects. Steps to developing DNA tools to support MAS and MAM include discovery of associations of markers with quantitative trait loci (QTL) explaining variation in the trait of interest and then, ideally, replicating discovery results in independent populations.

Discovery

The strategy of discovery in beef cattle has evolved from experiments using 'informative families' created from divergent parent lines (Casas *et al*. 2003; Kim *et al*. 2003) or large half-sib groups (Casas *et al*. 2000, 2001) to the current process of using large populations phenotyped for multiple traits of interest and genotyped with high-density panels (Miller *et al*. 2010; Snelling *et al*. 2010, 2011; Bolormaa *et al*. 2011*a*, 2011*b*). Allelic effects are estimated and combined in prediction equations for molecular breeding values (MBVs). Statistical approaches for simultaneous estimation of multiple marker effects are evolving (Meuwissen *et al*. 2001; Cleveland *et al*. 2010; Habier *et al*. 2011) and bioinformatics tools for routine analysis are being developed (e.g. GenSel software at http://bigs.ansci.iastate.edu, verified 17 January 2012). With these methods, the focus has shifted from a single polymorphism as the cause of genetic variation to multiple markers covering several genomic regions.

Fine mapping follows broader searches to further study interesting regions in hope of identifying markers in greater linkage disequilibrium with causative variation than those on the discovery panel and candidate genes within those regions that then can be sequenced to identify polymorphisms in the hope of determining the causative mutations. The efficacy of this approach has been demonstrated several times with simply inherited traits such as recent genetic defects in beef cattle (Charlier *et al*. 2008; Meyers *et al*. 2010). The complexity is obviously increased for polygenic quantitative traits; therefore, the probability of success is lower.

Replication

The process of replicating scientific findings has always been important. Early application of this process for commercially available DNA tests initiated by the National Beef Cattle Evaluation Consortium (NBCEC) for the US beef industry was referred to as 'validation.' Results are available on the NBCEC website (www.NBCEC.org, verified 17 January 2012) and in Van Eenennaam *et al*. (2007, 2009). It is interesting to note that international collaborations on validation were carried out for several of the commercial panels going through the process, capitalising on the existence of international discovery populations for the purpose of validation. In Australia, the Beef CRC and the Animal Breeding and Genetics Unit (AGBU), a joint unit of the University of New England (UNE) and New South Wales Department of Primary Industries, have conducted genomic calibrations in coordination with the NBCEC (see http://agbu.une.edu.au/genomic_calibrations.php, verified 17 January 2012). Early DNA tests consisted of relatively few markers and the process of validation determined significance of the marker effects in populations independent of the discovery population. As more markers were added to the commercial panels, this became increasingly difficult to do because the populations needed to be large enough to ensure adequate representation of all genotypes. As such, the process evolved to assessing the significance of the regression of phenotypic performance of animals on the MBVs obtained for those animals in the independent population. This approach in turn evolved to the current strategy of estimating the proportion of genetic variation the MBVs account for in the independent population (Thallman *et al*. 2009). Application of the later approach can be found at http://www.beefcrc.com.au/Aus-Beef-DNA-results (verified 17 January 2012) for commercially available DNA tests and in MacNeil *et al*. (2010) using methods proposed by Kachman (2008).

Application

DNA tests have been mostly used as an independent source of information for selection, even for traits for which routine genetic evaluations are available. Some applications of augmenting genetic evaluations and molecular information have been implemented. Aguilar *et al*. (2010) proposed the inclusion of the genomic relationship matrix for animals with marker information, along with the additive relationship matrix for those that do not have genotypes, as a one step process. This assumes the availability of marker genotypes to providers of genetic evaluations. For cases where marker genotypes are not available due to intellectual property issues, Kachman (2008) proposed treating the MBV as a correlated trait in the genetic evaluation system. Recently, the American Angus Association adopted the latter strategy by incorporating MBVs into their genetic evaluations of carcass traits (MacNeil *et al*. 2010) and has since adopted this strategy for numerous other traits (http://www.angus.org/AGI/GenomicChoice070811.pdf, verified 17 January 2012). A blending approach is being used by BREEDPLAN to incorporate MBVs into the genetic evaluation for Angus Australia and Angus New Zealand (http://www.angusaustralia.com.au/Breedplan/BP_SS_Intro.pdf, verified 17 January 2012). BREEDPLAN has incorporated information from several SNPs into Australian Brahman estimated breeding values for shear force since 2008 (http://agbu.une.edu.au/brahman%20tenderness%20EBVs.pdf).

Need for international collaboration

There has been substantial progress in the evolution of DNA technology and, to a lesser extent, transfer of tools from that

technology to the industry for use in MAS. However, furthering the effort will require circumventing several challenges which include:

- continuing development of discovery populations (and populations for replication) for novel traits,
- the need for replication of results in other populations to establish the consistency of marker associations,
- understanding the interactions between genotypes and environments,
- assessing the relationship of novel traits with routinely recorded traits to position them correctly in multiple-trait selection to improve economically relevant traits, and
- developing decision support tools to aid in making selection and or management decisions from all sources of information.

Underlying many of the challenges is the need to increase the number of phenotyped (genotyped) animals available for study.

Animal populations

Discovery in livestock has been disappointing if judged by the amount of genetic variation accounted for by the DNA predictors (Van Eenennaam *et al.* 2009). This is especially true for novel traits where discovery populations consist of individual animals phenotyped for those traits, as opposed to traits where genetic predictions are available on highly proven animals to use as the 'phenotypic' data, Garrick *et al.* (2009). Goddard (2009) derived the number of phenotyped animals needed to achieve an accuracy of prediction of either 0.5 or 0.7 given the assumptions of an effective population size equal to 100 and every QTL in perfect linkage disequilibrium with a SNP. For a trait with an h^2 of 0.3, 4000 animals are needed to achieve an accuracy of 0.5. The equivalent animal numbers for a trait with h^2 of 0.1 or 0.4 are approximately 12 and 3000 animals, respectively. If the goal is to achieve a higher accuracy of 0.7, the comparable numbers for h^2 of 0.1, 0.3 and 0.4 are ~38 000, 12 500 and 9000, respectively. These are staggering numbers when one considers the costs of data collection and of genotyping the animals.

Table 1 shows the number of beef animals reported to be represented in the discovery populations of organisations in three countries when an international collaboration was proposed between these countries in 2008. The countries and organisations represented are as follows: Australia, Beef Cooperative Research Center (CRC); Canada, Universities of Alberta and Guelph; the United States, USA Meat Animal Research Center (MARC).

From Table 1, several things are immediately apparent. First, these are large resource populations that on the surface appear to meet requirements of animal number needs. But further inspection shows that these datasets represent a plethora of breeds (and composites) and are a mixture of heifer, steer, bull and cow data (the latter being quite limited at the time of compiling the table). Not apparent from the table is the fact that not all animals are measured for all traits, for example, feed intake was measured on only ~7800 animals. Still, the numbers are formidable. However, given the cost of genotyping animals with the high-density panels with >750 000 SNPs, only a fraction of these animals have been genotyped. A concerted effort is underway in all the organisations to increase genotype information. This will be carried out by using the new high-density panels (and targeted sequencing), in conjunction with lower-density panels to allow for imputation. This will be an expensive venture and yet, in the end, individual organisations are still likely to be short of the numbers of phenotyped animals to account for large proportions of genetic variation. Collectively, however, the number of animals that will be represented across organisations will be quite substantial. The problem is compounded by the current assumption that genomic predictions must be breed-specific and estimated from single-breed populations. There is hope that the requirement for predictions to be breed-specific may be relaxed with increasing marker density and/or individual animal sequence and that improved statistical models may provide for sharing of information among populations, but this hope is yet to be realised.

Each organisation has reported research results on association studies from their respective populations. The next logical step was to compare findings from the independent populations. We did this for the Australian results reported by Bolormaa *et al.* (2011*a*) for feedlot growth and efficiency traits using results reported by Snelling *et al.* (2011). The traits included were residual feed intake, average daily gain and mid-test bodyweight. Bolormaa *et al.* (2011*a*) reported 25 1-Mbp intervals where significant ($P < 0.05$) associations were found for all three traits (Table 2). The MARC analysis also identified SNPs significant ($P < 0.05$) for all three traits in 7 of these 25 intervals. There were four instances where there were no significant results found for any of the three traits in the MARC analysis. In the remaining 15 intervals, one or two traits were found to be associated with SNPs in that region in the MARC dataset.

Significance of individual SNP effects is obviously not the final criterion for assessment. Table 3 shows a comparison of preliminary analyses conducted in the three countries (*Bos Taurus* data only) for growth traits relative to total and discordant matches among the 2500 SNPs, with the greatest significance values for yearling weight or mid-test weight, daily gain on test, feed intake, and residual feed intake from each collaborator. It is interesting that there is a higher degree of concordance between the MARC and Canadian results for these traits than there is concordance of results from either MARC or the Canadian study with those results from the Australian study. This probably reflects the greater degree of relatedness between the USA and Canadian beef populations than of either with the Australian population.

The reasons for discordance in particular comparisons need to be understood, especially if datasets were to be combined for joint estimation of allelic effects. On one hand, there will be a fraction of discordant results stemming from spurious associations. On the other hand, the results could be real and the discordance could be due to differences in phase between markers and QTL in the different populations or interactions with the environment or different genetic backgrounds. Although strong evidence of reversals of marker and QTL phase are not common in the literature, there are cases of markers being in phase with QTL in one population and not another. One instance of this is a TG5 marker being in phase with marbling score in Wagyu-derived

Table 1. Project attributes and number of observations represented in discovery populations of the Beef Co-operative Research Center (CRC), US Meat Animal Research Center (MARC) and the Universities of Alberta and Guelph

Breed designations: AN = Angus, BB = Belgian Blue, BM = Beefmaster, BN = Brangus, Bo = Boron, BON = Bonsmara, BR = Brahman, BRd = Belmont Red, BV = Braunvieh, CH = Charolais, CI = Chianina, Comp = Bos Taurus Composite, Fr = Friesian, GV = Gelbvieh, HE = Hereford, LM = Limousin, MA = Maine Anjou, MG = Murray Grey, MII = MARC II Composite, MIII = MARC III Composite, NR = Norwegian Red, Pd = Piedmontese, PI = Pinzgaur, RA = Red Angus, Ro = Romosinuano, SA = Salers, SG = Santa Gertrudis, SH = Shorthorn, SM = Simmental, SR = Swedish Red and White, TComp = Tropical Composite, Tm = Terminal, Tu = Tuli, UNK = unknown, Wa = Wagyu. Sex designations: S = steer, H = heifer, B = bull and C = cow. GPE F_1s = Germ Plasm Evaluation cycles V-2, VI-2, VII-2, VIII-2; GPE advanced generation = GPE V-3, VI-3, VII-3, VIII-3, VII-4, VII-6. IMF% = percentage intramuscular fat. Ultrasound = ultrasound measurement of rib-eye area, fat thickness, and intramuscular fat. Blood traits = IGF-I, cortisol and hormone assays in Australia; they are plasma glucose, plasma urea nitrogen, T3, and T4 at MARC, with serum samples available to add additional assays. Cooler data = fat thickness, rib-eye area, carcass weight, and internal fat. Cow weight = cow weight, height, and body condition. Mating performance = multisire natural-service mating performance, as determined by DNA-based paternity testing. Worms = worms, coat color, coat score, sheath or navel score. Ticks = ticks, buffalo fly, temperature

Parameter	Beef CRC1	Beef CRC	Beef CRC	MARC	MARC	MARC	Univ. of Guelph	Univ. of Guelph
Project	Straight-bred	Female fertility	Male fertility	GPE F_1s	GPE advanced generation	GPE continuous	Performance	Cow performance
Total number	7800	4400	6000	4204	8244	713	2000	500
Breed	AN, HE, SH, MG, BR, SG, BRd	BR, TComp	BR, TComp	21 Brds[A]	21 Brds[A]	16 Brds[B]	AN, SM, Tm	AN, SM Comp
Sex	S, H	S, H	B	S, H, B	S, H	S, H, B	B, H and S	H
Calving ease	0	0		4204	8244	713		
Gestation length	0	0	0	4204	133	713		
Birth weight	3000	500	6000	4204	8244	713	1800	500
Weaning weight	7500	4400	6000	4134	7909	670	1800	500
Average daily gain	7500	4400	3000	3930	7523	461	1800	500
Ultrasound	7500	4400	3000	0	0			
Feed intake	1500	1500	0	0	1885		1000	
Temperament	4000	4400	6000	0	1885			
Blood traits	1600	4400	6000	0	1300			
Cooler data	4000	2200	0	1270	5203	430	1800	
pH	7800	2200	0	0	0			
Meat colour	7800	2200	0	1270	1283			
%Cutout	3000	800	0	1174	1250		1800	
IMF%	7800	2200	0	0	0		1000	
Shear force	7800	2200	0	1176	1250		1800	
Taste panel	4000	0	0	1176	0			
Fatty acids	1600	0	0	606	0			
Puberty age		2200		2293	711			
Conception date		2200		2724	993			
Postpartum interval				1954	1028			
Teat and udder		2200		936	365			
Cow weight	0	2200	0	2406	993			500
Cow maintenance				200	0			
Scrotal size			3000	75	0	96		
Semen quality			3000	70	0	96		
Mating performance				70	0			
Worms		4400		0	0			
Ticks		2200		0	0			
Disease treatment				4204	8244	713		

[A]AN, HE, SM, CH, LM, RA, GV, BR, Bo, Tu, BB, Pd, Fr, SR, NR, Wa, BN, BM, Ro, Bon, MIII.
[B]AN, HE, SM, CH, LM, RA, GV, BN, BM, SH, MA, BR, SG, CI, SA, BV.

animals but not in animals derived from other sources (Casas *et al*. 2007).

To expand the utility of discovery populations, Saatchi *et al*. (2011) promoted combining the discovery and validation steps in a systematic process called cross-validation. This process involves dividing the discovery population into *n* subsets, using *n*–1 subsets in a discovery analysis and assessing the results against the excluded subset. This process is then repeated with all possible combinations of the *n* subsets used in discovery, such that, in the end, every subset contributes both to discovery and validation. Results are then summarised across 'experiments.' How the discovery population is divided can, in some instances, be by categories (e.g. breed, location, or year of birth), while in other cases, it could be based on the degree of

Table 2. Comparison of results from USA Meat Animal Research Center (MARC; Snelling *et al.* 2011) association studies for residual feed intake (RFI), average daily gain (ADG) and mid-test bodyweight (mMWT) to regions of interest containing significant associations with all three traits in Australian Cooperative Research Center (CRC) studies (Phase I, Bolormaa *et al.* 2011*a*)

Position: Btau4.0. SNP, single-nucleotide polymorphism. Bold indicates those locations where all three traits were found to be significantly associated with SNPs in the region in both the USMARC data and the Australian study

BTA	Position (Mbp)	Dataset	Number of significant ($P < 0.05$) SNPs			Minimum P		
			RFI	ADG	mMWT	RFI	ADG	mMWT
2	106–107	CRC1	2	3	1	0.0001	0.0019	0.0009
		MARC	0	0	0	0.1155	0.0589	0.0579
3	**105–106**	**CRC1**	**7**	**1**	**1**	**0.0000**	**0.0445**	**0.0131**
		MARC	**4**	**5**	**2**	**0.0094**	**0.0009**	**0.0041**
3	51–52	CRC1	6	1	4	0.0004	0.0199	0.0124
		MARC	0	0	0	0.0715	0.0927	0.0506
3	84–85	CRC1	4	2	3	0.0002	0.0075	0.0356
		MARC	1	0	7	0.0100	0.0847	0.0050
4	46–47	CRC1	1	2	2	0.0258	0.0000	0.0004
		MARC	1	0	2	0.0471	0.0778	0.0222
4	**91–92**	**CRC1**	**2**	**4**	**1**	**0.0002**	**0.0114**	**0.0239**
		MARC	**2**	**6**	**1**	**0.0384**	**0.0088**	**0.0333**
6	111–112	CRC1	3	3	3	0.0007	0.0095	0.0066
		MARC	0	0	0	0.0540	0.1311	0.1473
6	**41–42**	**CRC1**	**3**	**2**	**5**	**0.0002**	**0.0132**	**0.0067**
		MARC	**1**	**7**	**11**	**0.0163**	**0.0006**	**0.0000**
8	104–105	CRC1	4	2	2	0.0009	0.0253	0.0187
		MARC	0	1	0	0.1729	0.0133	0.1159
8	**86–87**	**CRC1**	**6**	**2**	**2**	**0.0000**	**0.0047**	**0.0155**
		MARC	**4**	**1**	**1**	**0.0142**	**0.0212**	**0.0331**
8	87–88	CRC1	5	2	3	0.0001	0.014	0.0121
		MARC	0	2	0	0.0532	0.0164	0.1062
8	88–89	CRC1	7	1	2	0.0009	0.0382	0.0301
		MARC	0	1	4	0.0608	0.0272	0.0113
8	89–90	CRC1	7	3	2	0.0006	0.0104	0.0053
		MARC	2	0	0	0.0039	0.0857	0.1456
9	78–79	CRC1	4	3	1	0.0003	0.0037	0.0450
		MARC	1	0	1	0.0097	0.0556	0.0441
10	**18–19**	**CRC1**	**2**	**1**	**2**	**0.0006**	**0.0432**	**0.0103**
		MARC	**2**	**3**	**2**	**0.0238**	**0.0004**	**0.0037**
11	46–47	CRC1	1	2	10	0.0483	0.0002	0.0002
		MARC	0	0	4	0.0725	0.1568	0.0097
14	**17–18**	**CRC1**	**6**	**5**	**5**	**0.0005**	**0.0008**	**0.0081**
		MARC	**1**	**1**	**1**	**0.0345**	**0.0322**	**0.0268**
16	25–26	CRC1	1	4	1	0.0417	0.0006	0.0005
		MARC	1	0	0	0.0380	0.0519	0.1565
17	10–11	CRC1	3	3	2	0.0008	0.0358	0.0194
		MARC	0	0	0	0.0692	0.2861	0.1420
17	37–38	CRC1	2	1	2	0.0005	0.0492	0.0051
		MARC	1	0	1	0.0370	0.1747	0.0165
19	38–39	CRC1	2	1	2	0.0003	0.0417	0.0346
		MARC	5	1	0	0.0001	0.0409	0.1047
20	30–31	CRC1	1	2	2	0.0001	0.0085	0.0047
		MARC	0	1	0	0.1362	0.0393	0.0839
22	45–46	CRC1	0	1	2	0.0003	0.0356	0.0243
		MARC	0	0	2	0.1889	0.0523	0.0046
23	18–19	CRC1	0	1	1	0.0432	0.0001	0.0009
		MARC	0	3	1	0.0718	0.0072	0.0321
23	**49–50**	**CRC1**	**5**	**3**	**1**	**0.0002**	**0.0274**	**0.0448**
		MARC	**1**	**2**	**1**	**0.0433**	**0.0327**	**0.0478**

Table 3. Number of pair-wise matching single-nucleotide polymorphisms (SNPs) among 2500 highest-probability SNPs from each of three collaborators

Number of SNPs among the 2500 highest probabilities within each collaborator that match between pairs of collaborators. Numbers are counts of concordant (CON: *t*-values with the same signs) matches and discordant (DIS: *t*-values with different signs) matches. MARC = USA Meat Animal Research Center. UGUA = Canadian Universities of Guelph and Alberta. CRCB = Australian CRC for Beef Genetic Technologies

Trait	MARC–UGUA		MARC–CRCB		CRCB–UGUA	
	CON	DIS	CON	DIS	CON	DIS
Yearling or mid-test weight	113	35	77	67	71	47
Daily gain on test	72	55	52	74	66	66
Residual feed intake	98	58	85	71	69	71
Feed intake	108	48	73	65	75	50

relationship from closely to distantly related subsets. This process could be quite useful for combining datasets from collaborating countries.

However, in doing so, the issues of concordance are not addressed. That is, when combining datasets, estimated allelic effects for discordant QTL will be some weighted average of the information. This would not be appropriate if the discordance was reflecting a different phase of marker association with a real QTL across the populations or if there was a genotype by environment (G × E) interaction associated with the QTL being marked. Whereas, if discordance really is due to a different phase of marker association, the combined dataset may allow discovery of markers in higher linkage disequilibrium with the causative variation in the same region, through fine-mapping or higher-density genotyping. Furthermore, the breed-specific nature of current genomic predictions tends to limit international collaboration to those breeds that are present in the respective countries. Regardless, it seems the advantage to combining data relative to the increased power of discovery in an international collaboration would be beneficial, and in doing so, does enhance the opportunity for exploration of multi-breed predictions.

There are obvious challenges to combining datasets from collaborating countries. These include differences in the definitions of traits, differences in breed representation and differences in environments. These challenges do not disappear in the absence of collaboration because the commercialisation of DNA tests is global; DNA tests developed in Australia have been commercialised in the USA and Canada, and *vice versa*.

Recently, purebred resource populations have gained popularity because it has become clear that current predictions are not portable across breeds, or in some cases, across subpopulations within breeds. However, most commercial cattle are crossbred, and therefore, crossbred resource populations will continue to contribute essential information. Mapping heterosis with respect to the genome will require crossbred resource populations. As more individual animal sequence data becomes available, it seems likely that crossbred resource populations will become increasingly valuable. For example, estimation of the difference in effects of haplotypes that exist only in different breeds will be much more efficient in populations with parents that are crosses of the respective breeds.

Genotypes

Genotyping costs have been dramatically reduced, even for the higher-density panels. Cost of sequencing has decreased even faster. Nevertheless, the cost of genotypes, sequencing and phenotypes required to reach the number of animals needed is exorbitant and difficult to fund individually. In addition to phenotyped animal populations, genotyping of important industry animals has and will continue to occur. For example, MARC researchers have genotyped over 2000 prominent artificial insemination (AI) beef bulls representing multiple breeds in the USA and will be genotyping (770K) and sequencing at low coverage (2X) AI beef bulls that are prominent ancestors of the discovery population pedigree at MARC. Sequence of highly influential animals will be imputed to other animals in the population using lower density (e.g. 50K and/or 3K) marker sets. Sequencing will make available the causative SNPs, many of which are likely to have allele frequencies too low to even be considered for inclusion on general-purpose SNP chips. Having access to the genotypes of causative SNPs does not make it easy to discern them from the rest (as the total number of SNPs to be considered will increase dramatically), but it does make it possible to discern them.

The national herds of beef and dairy across countries are related. Prominent males appear in the pedigree of animals in these national herds (which has enhanced international genetic evaluations). Hence, to avoid the duplication of genotyping, the concept of a global strategy for genotyping and sequencing becomes an attractive proposition and this strategy would be made plausible by developing an international database of the animal identifications for those males that have been genotyped. As an example, the above mentioned organisations (in USA, Australia and Canada) that have ventured into previous collaborations for beef genomics have successfully partnered on a grant to Genome Canada, with Project Leaders Drs Stephen Moore and Stephen Miller, entitled ‘Whole Genome Selection through Genome Wide Imputation’. This collaboration has an objective of sharing sequence information on industry animals among the three countries combining to generate a total of 960X coverage of the bovine genome from some combination of X coverage per animal by *n* animals.

Genome-enabled genetic predictions

For traits in which national genetic evaluations exist, integration of the information from genomic tools is particularly appealing. As mentioned previously, efforts in this area consider indexing MBVs with genetic predictions in a two-step process of calculating MBVs and genetic predictions and indexing (blending) the results from each, fitting genomic relationships or estimating SNP effects and fitting the MBV as a correlated trait.

The latter being a concession to the possibility that issues with intellectual property would prevent sharing of the raw SNP genotypes. There are international collaborations in genetic evaluations and, as such, collaborating on genotyping industry animals and sharing discovery information for the genetic evaluations would enhance the efforts towards integration. International evaluations still face the issue of discordance of SNP effects resulting from differences in phase and G × E interactions when fitting SNP or the genetic variance–covariance matrix built on genomic relationships.

Future opportunities

One wonders just how much of the genetic variation could be explained by the summation of single effects of markers following simple Mendelian inheritance based on our current tools and perhaps, more importantly, our gross measures of phenotypes. Many of the economically relevant traits of interest are complex in nature and breaking those complex traits down to simpler forms may lead to more effective discovery models. These simpler forms, termed physiological indicator traits by Thallman *et al*. (2008), allow examination of genomic influences on pathways that ultimately influence our ERTs even when the expression of the ERT is not observable (e.g. female reproduction in bulls or disease incidence with limited pathogen exposure). Thallman *et al*. (2008, p. 329) addressed this concept for disease resistance stating the following:

> *Physiological indicator traits* (*PIT*) *are those that are expected to be closely related to physiological processes that are components of disease resistance*. *In most cases*, *it should be useful to measure them in all animals in a population*, *whether sick or not. Ideally*, *it should be useful to measure them regardless of the level of natural exposure to disease. Because they are related to components*, *they are expected to have higher heritability than disease resistance itself. Because of higher expected heritability and greater effective numbers of observations than for disease incidence* (*especially when depending on natural exposure*), *QTL detection for PIT is likely to be considerably more successful than QTL detection for disease incidence directly.*

Addressing the concept of phenomics (generating an extensive set of phenotypes at the biological level that contribute to the architecture of our suite of complex traits measured across time and environments) will add additional complexity to the data-collection strategies employed for future discovery work to enable successful MAS.

Finally, understanding the complex biological intermediaries from the genome to the phenotype and interactions with the proteome and metabolome will be important for the prediction of phenotypes for the application of MAM. Accurate phenotypic predictions will greatly enhance future strategies for activities such as optimising intervention for disease or predicting responses to vaccination or stimulation by hormonal growth promotants.

References

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010) A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* **93**, 743–752. doi:10.3168/jds.2009-2730

Bolormaa S, Hayes BJ, Savin K, Hawken R, Barendse W, Arthur PF, Herd RM, Goddard ME (2011*a*) Genome-wide association studies for feedlot and growth traits in cattle. *Journal of Animal Science* **89**, 1684–1697. doi:10.2527/jas.2010-3079

Bolormaa S, Porto Neto LR, Zhang YD, Bunch RJ, Harrison BE, Goddard ME, Barendse W (2011*b*) A genome-wide association study of meat and carcass traits in Australian cattle. *Journal of Animal Science* **89**, 2297–2309. doi:10.2527/jas.2010-3138

Casas E, Shackelford SD, Keele JW, Stone RT, Kappes SM, Koohmaraie M (2000) Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate forms of myostatin. *Journal of Animal Science* **78**, 560–569.

Casas E, Stone RT, Keele JW, Shackelford SD, Kappes SM, Koohmaraie M (2001) A comprehensive search for QTL affecting growth and carcass composition of cattle segregating alternative forms of the myostatin gene. *Journal of Animal Science* **79**, 854–860.

Casas E, Shackelford SD, Keele JW, Koohmaraie M, Smith TPL, Stone RT (2003) Detection of quantitative trait loci for growth and carcass composition in cattle. *Journal of Animal Science* **81**, 2976–2983.

Casas E, White SN, Shackelford SD, Wheeler TL, Koohmaraie M, Bennett GL, Smith TPL (2007) Assessing the association of single nucleotide polymorphisms at the thyroglobulin gene with carcass traits in beef cattle. *Journal of Animal Science* **85**, 2807–2814. doi:10.2527/jas.2007-0179

Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, Cambisano N, Carta E, Dardano S, Dive M, Fasquelle C, Frennet J-C, Hanset R, Hubin X, Jorgensen C, Karim L, Kent M, Harvey K, Pearce BR, Simon P, Tama N, Nie H, Vandeputte S, Lien S, Longeri M, Fredholm M, Harvey RJ, Georges M (2008) Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nature Genetics* **40**, 449. doi:10.1038/ng.96

Cleveland MA, Forni S, Deeb N, Maltecca C (2010) Genomic breeding value prediction using three Bayesian methods and application to reduced density marker panels. *BioMed Central Proceedings* **4**(Suppl.1), S6

Garrick DJ, Taylor JF, Fernando RL (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics, Selection, Evolution.* **41**, 55. doi:10.1186/1297-9686-41-55

Goddard ME (2009) How can we best use DNA data in selection of cattle? In 'Proceedings of Beef Improvement Federation 41st annual research symposium', 30 April–3 May 2009, Sacramento, CA. pp. 81–91.

Habier D, Fernando RL, Kizikaya K, Garrick DJ (2011) Extension of the Bayesian alphabet for genomic selection. *BioMed Central Bioinformatics* **12**, 186.

Kachman SD (2008) Parameters needed to add genomics to genetic prediction. In 'Proceedings of the 9th genetic prediction workshop', Kansas City, MO, pp. 92–98. (Beef Improvement Federation: Raleigh, NC)

Kim JJ, Farnir F, Savell J, Taylor JF (2003) Detection of quantitative trait loci for growth and beef carcass fatness traits in a cross between *Bos taurus* (Angus) and *Bos indicus* (Brahman) cattle. *Journal of Animal Science* **81**, 1933–1942.

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**, 743.

MacNeil MD, Nkrumah JD, Woodward BW, Northcutt SL (2010) Genetic evaluation of Angus cattle for carcass marbling using ultrasound and genomic indicators. *Journal of Animal Science* **88**, 517–522. doi:10.2527/jas.2009-2022

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **15**, 1819–1829.

Meyers SN, McDaneld TG, Swist SL, Marron BM, Steffen DJ, O'Toole D, O'Connell JR, Beever JE, Sonstegard TS, Smith TP (2010) A deletion mutation in bovine SLC4A2 is associated with osteopetrosis in Red Angus cattle. *BioMed Central Genomics* **11**, 337.

Miller S, Lu D, Vander Voort G, Sargolzaei M, Caldwell T, Wang Z, Mah J, Plastow G, Moore S (2010) Beef tenderness QTL on BTA25 from a whole genome scan with the BovineSNP50 Beadchip. In 'Proceedings of the 9th world congress on genetics applied to livestock production', 1–6 August 2010, Leipzig, Germany. CD-ROM Communication 0675.

Saatchi M, McClure MC, McKay SD, Rolf MM, Kim JW, Decker JE, Taxis TM, Chapple RH, Ramey HR, Northcutt SL, Bauck S, Woodward B, Dekkers JCM, Fernando RL, Schnabel RD, Garrick DJ, Taylor JF (2011) Accuracies of genomic breeding values in American Angus beef cattle using k-means clustering for cross-validation. *Genetics, Selection, Evolution.* **43**, 40. doi:10.1186/1297-9686-43-40

Snelling WM, Allan MF, Keele JW, Kuehn LA, McDaneld T, Smith TPL, Sonstegard TS, Thallman RM, Bennett GL (2010) Genome-wide association study of growth in crossbred beef cattle. *Journal of Animal Science* **88**, 837–848. doi:10.2527/jas.2009-2257

Snelling WM, Allan MF, Keele JW, Kuehn LA, Thallman RM, Bennett GL, Ferrell CL, Jenkins TG, Freetly HC, Nielsen MK, Rolfe KM (2011) Partial-genome evaluation of postweaning feed intake and efficiency of crossbred beef cattle. *Journal of Animal Science* **89**, 1731–1741. doi:10.2527/jas.2010-3526

Thallman RM, Kuehn LA, Allan MF, Bennett GL, Koohmaraie M (2008) Opportunities for collaborative phenotyping for disease resistance traits in a large beef cattle resource population. *Developments in Biologicals* **132**, 327–330. doi:10.1159/000317178

Thallman RM, Hanford KJ, Quaas RL, Kachman SD, Templeman RJ, Fernando RL, Kuehn LA, Pollak EJ (2009) Estimation of the proportion of genetic variation accounted for by DNA tests. In 'Proceedings of the Beef Improvement Federation 41st annual research symposium and annual meeting', 30 April–3 May 2009, Sacramento, CA. pp. 184–209.

Van Eenennaam AL, Li J, Thallman RM, Quaas RL, Dikeman ME, Gill CA, Franke DE, Thomas MG (2007) Validation of commercial DNA tests for quantitative beef quality traits. *Journal of Animal Science* **85**, 891–900. doi:10.2527/jas.2006-512

Van Eenennaam AL, Thallman RM, Quaas RL, Hanford K, Pollak EJ (2009) Validation and estimation of additive genetic variation associated with DNA tests for quantitative beef cattle traits. *Proceedings of the Association for Advancement of Animal Breeding and Genetics* **18**, 129–132.