Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information

J. E. Pryce,*†¹ B. J. Hayes,*†‡ and M. E. Goddard*†§

*Biosciences Research Division, Department of Primary Industries Victoria, 1 Park Drive, Bundoora 3083, Australia

†Dairy Futures Cooperative Research Centre, 1 Park Drive, Bundoora, Victoria 3083, Australia

‡La Trobe University, Bundoora, Victoria 3086, Australia

§Faculty of Land and Food Resources, University of Melbourne, Parkville 3010, Australia

ABSTRACT

In this study, 3 strategies for controlling progeny inbreeding in mating plans were compared. The strategies used information from pedigree inbreeding coefficients, genomic relationships, or shared runs of homozygosity. The strategies were compared for the reduction in genetic gain and progeny inbreeding that would be expected from selected matings, and for the decrease of homozygosity of deleterious recessive alleles. Using real pedigree, genotype [43,115 single nucleotide polymorphism (SNP) markers], and estimated breeding value data from Holstein cattle, mating plans were derived for herds of 300 cows with 20 sires available for mating, replicated 50 times. Each of the 300 individuals allocated as dams were matched to 1 of 20 sires to maximize genetic merit minus the penalty for estimated progeny inbreeding, and given the restriction that the sire could not be mated to more than 10% of the cows. The strategy that used a genomic relationship matrix (GRM) was the most effective in reducing average progeny inbreeding; this strategy also resulted in fewer homozygous SNP out of 1,000 low-frequency SNP compared with the strategy using pedigree information. In the future, large numbers of cattle may be genotyped for low-density SNP panels. A GRM constructed using 3,123 SNP produced results similar to a GRM constructed using the full 43,115 SNP. These results demonstrate that using GRM information, a 1% reduction in progeny inbreeding (valued at around \$5 per cow) can be made with very little compromise in the overall breeding objective. These results and the availability of low-cost, low-density genotyping make it attractive to apply mating plans that use genomic information in commercial dairy herds.

Key words: genomic, inbreeding, runs of homozygosity, mating plan

INTRODUCTION

In dairy cattle breeding, the ability to cost-effectively predict breeding values using dense genetic marker information through genomic selection has the potential to accelerate the rate of genetic gain above that achieved through progeny testing (Meuwissen et al., 2001; Schaeffer, 2006; Pryce et al., 2010b; Lillehammer et al., 2011). Genomic breeding values of young bulls can be accurately evaluated at birth, so these bulls can be widely used at sexual maturity rather than waiting until they have lactating daughters to provide progeny test results. Thus, generation intervals are likely to be much shorter under genomic selection than when using conventional progeny testing. However, even though the rate of inbreeding should be lower per generation from genomic selection compared with progeny testing, as Mendelian sampling can be estimated at the time of selection more accurately (Daetwyler et al., 2007), shorter generation intervals can lead to an increase in the rate of annual rate of inbreeding (e.g., Lillehammer et al., 2011). Furthermore, there is a risk that genomic selection could result in large homozygous segments of chromosome surrounding QTL in the selected population (Sonesson et al., 2010). For these reasons, strategies to control the rate of inbreeding using genomic selection schemes are important.

Inbreeding can be controlled at 2 levels. First, the rate of inbreeding in the population as a whole can be restricted to a desired level while maximizing the rate of genetic gain, by optimizing the long-term contributions of a selected number of breeding animals (e.g., Wray and Goddard, 1994; Meuwissen, 1997). Second, at an individual animal level, avoiding large inbreeding coefficients in progeny through control of matings is important to avoid reductions in fitness traits (Smith et al., 1998) and homozygous lethal recessive alleles. Control of inbreeding levels in progeny can be implemented using mate allocation (Kinghorn, 1998). Mate allocation can be considered independently of mate selection if all candidates are to be mated, which is often the case for females in many modern dairy herds because

Received February 9, 2011.

Accepted September 22, 2011.

¹Corresponding author: jennie.pryce@dpi.vic.gov.au

of poor fertility or compact calving patterns. For large herds in particular, mating plans could help resolve the choice of sires to mate to cows in an optimal way. Although mating plans are commonly used to control progeny inbreeding in livestock, they can also be used to make culling decisions in young animals and in situations where constraints, such as animal groups, exist (Kinghorn, 2011). They are also used when economic merit is nonlinear and when there is an intermediate optimum in the selection goal or to increase connectedness between herds and to utilize dominance effects (Toro and Varona, 2010).

Genomic information offers new possibilities to control the level of progeny inbreeding. Traditionally, mating plans constrain the inbreeding of predicted progeny by using a matrix of relationships among individuals constructed from pedigree information (**A**; e.g., Kinghorn, 1998, 2011). Elements of the **A** matrix are the expected proportion of the genome that is identical by descent given pedigree relationships between individuals and are double the predicted progeny inbreeding coefficient that would result from mating individuals. Using genetic markers, the realized proportion of the genome that is identical by descent (**IBD**) between individuals can be calculated (VanRaden, 2008; Hayes et al., 2009; Yang et al., 2010). If the genomic relationship matrix (**GRM**) is used instead of the **A** matrix for the prediction of EBV (i.e., genomic selection), the rate of genetic gain can be increased (Van Raden, 2008; Hayes et al., 2009). Using optimal contribution selection (Wray and Goddard, 1994; Meuwissen, 1997), the effect of genomic selection on rates of inbreeding was investigated by Sonesson et al. (2010). They found that both genomic relationships and pedigree relationships were successful strategies to control the rate of inbreeding under genomic selection. However, the genomic rate of inbreeding was around 3 times higher when using pedigree optimal contribution selection than in genomic optimal contribution selection. Pedigree optimal contribution selection seemed less able to restrict inbreeding at sites with QTL of large effect, whereas genomic optimal contribution selection resulted in a more evenly distributed increase in IBD across the genome. Therefore, using genomic relationships instead of pedigree relationships appears to be better at constraining genomic inbreeding under genomic selection.

Another way of assessing inbreeding is to look at the IBD of chromosome segments (Kirin et al., 2010). One measure is chromosome segment homozygosity, which is defined as the probability that 2 chromosome segments randomly drawn from the population are IBD (Hayes et al., 2003). However, instead of calculating chromosome segment homozygosity, an alternative is to observe chromosome segments that are identical by

state (**IBS**) in multiple, successive markers (MacLeod et al., 2009). Provided the number of markers in the haplotype is large, the probability of IBS and IBD will be very close. When the 2 chromosome segments occur in the same individual, this is an unbroken run of homozygosity (**ROH**), defined as whether 2 haplotypes share a given number of alleles at n adjacent markers (MacLeod et al., 2009). For both chromosome segment homozygosity and ROH, the chromosome segment is unbroken by a recombination event. Long shared ROH are most likely due to IBD chromosome segments and indicate a recent common ancestor. Conversely, if individuals share a shorter ROH this could be because of a long time span from their common ancestor, as the chromosomal segments will be broken down by repeated meioses. Individuals with similar genomic or pedigree relationships can have quite different proportions of identical ROH. Thus, ROH could be used to detect subtle variation in relationships at the genome level that are not detected using either pedigree or GRM. One of the underlying processes of inbreeding is that it increases the frequency of both favorable and deleterious homozygotes. If the occurrence of deleterious homozygotes is more likely to arise as a consequence of recent inbreeding (as opposed to more ancient inbreeding), then strategies to minimize ROH could be a way of reducing them.

In this paper, we compared mate allocation strategies to control the rate of progeny inbreeding in dairy herds. The aim was to maximize the breeding objective (the Australian Profit Ranking, **APR**; Pryce et al., 2009), while constraining the predicted average progeny inbreeding coefficient using a series of weights. Progeny inbreeding was quantified using the following measures calculated in potential parents of the next generation: (1) pedigree relationships; (2) a GRM calculated using 43,115 SNP or 3,123 SNP; and (3) avoiding ROH of varying lengths between individuals. The frequency of homozygous minor allele SNP out of 1,000 low-frequency SNP alleles (selected as surrogates for lethal mutations) was also investigated.

MATERIALS AND METHODS

Samples of DNA were obtained from 2,117 Holstein bulls and were genotyped using the Illumina BovineSNP50 BeadChip (Illumina, San Diego, CA; Matukumalli et al., 2009). Full details of the quality control methods for SNP editing are presented by Pryce et al. (2010a). Any SNP with no recorded heterozygotes were removed, as were animals with genotype-pedigree incompatibilities. After editing, 43,115 SNP were available for this study. The SNPs were ordered by chromosome position using Bovine Genome Build 4.0 (<http://>

www.ncbi.nlm.nih.gov/projects/genome/guide/cow/). The average spacing between SNP retained for this study was 0.067 Mbp. Missing genotypes were imputed using fastPhase (Scheet and Stephens, 2006). Haplotypes were phased using ChromoPhase, the long-range phasing program of Daetwyler et al. (2011). ChromoPhase was selected in preference to other programs (such as fastPhase; Scheet and Stephens, 2006) because it seeks out long stretches of chromosomal segments shared between relatives, rather than phasing segments of the chromosome and then rejoining segments. Accurate long-range phasing is required to calculate shared ROH; and ChromoPhase was designed to achieve this, whereas fastPhase was not. An example of phasing 2 genotyped full brothers using fastPhase and ChromoPhase is shown in Figure 1.

An important objective of this study was to evaluate the effect of the strategies tested on the homozygosity of deleterious recessives. To mimic this, 1,000 SNP from the 43,115 were chosen to be surrogate loci where the minor allele was assumed to be a deleterious mutation. In order to choose, as far as possible, loci that would be similar in frequency to deleterious alleles, 1,000 SNP with the lowest allele frequencies were selected. The maximum allele frequency of these 1,000 SNP was 3.3%, whereas the lowest allele frequency was close to zero. The average proportion of homozygous minor alleles among the 1,000 “surrogate loci” was then calculated for selected matings (described later). The 1,000 SNP were excluded from subsequent calculations of genomic relationships, so that the total number of SNP used to construct genomic relationship matrices was 42,115.

Figure 1. Percentage of genome in common in haplotypes of 2 full brothers calculated using ChromoPhase (Daetwyler et al., 2011) and FastPhase (Scheet and Stephens, 2006). Each haplotype finishes where a discordant SNP is identified. ROH = run of homozygosity. Color version available in the online PDF.

The correlation between a genomic relationship matrix constructed with and without these 1,000 SNP was >0.999.

Pedigree Relationships

For a given set of parents, the predicted inbreeding coefficient of progeny is half the additive genetic relationship between mating pairs. Therefore, the additive genetic relationship matrix, calculated using the R package “pedigree” (Coster, 2010), was divided by 2 so that it contained estimated inbreeding coefficients of progeny (**A**). The pedigree used in this study consisted of 6,019 animals, including the 2,117 Holstein bulls and their ancestors. The 2,117 bulls were born between 1955 and 2005, were sired by 316 different sires, and had 1,589 different dams. The sire of each genotyped bull was verified by comparing the bull’s genotype data to his sire’s genotypes using the parentage assignment computer algorithm of Hayes (2011). Four bulls had unknown sires and 19 had unknown dams. On average, the bulls had 5 generations of recorded pedigree, whereas 5.3% had more than 8 generations and 15.7% had 3 or fewer generations of ancestry.

Genomic Relationships

A GRM was constructed (VanRaden, 2009; Yang et al., 2010), except the off-diagonals of our genomic relationship matrix (**G_{ij}**) were half the usual calculation of GRM, so that they represented genomic estimates of progeny inbreeding coefficients that would result from matings:

$$\mathbf{G}_{ij} = \frac{1}{2N} \sum_m \frac{(x_{im} - 2p_m)(x_{jm} - 2p_m)}{2p_m(1 - p_m)},$$

where N is the number of SNP, p_m is the allele frequency of SNP m , and x_{im} and x_{jm} are the genotypes of sire i and dam j , respectively. Two GRM were calculated. The first (**G**) included 42,115 SNP, and the second (**G3k**) was a subset of these SNP, selected to be similar to the Illumina Bovine 3K chip (with 3,123 SNP extracted as a subset from the 42,115 SNP set described above).

The ROH were assessed for overlapping haplotype windows as follows:

$$ROH_{ij} = \frac{\sum_{i=1}^S x_{ij}}{4S - 4(n - 1)},$$

where n is the length of the haplotype in SNP (5, 10, 50, 100, 300, or 500), x is a haplotype that is in com-

Figure 2. An example of a run of homozygosity of 10 SNP (within dotted square) for a sire (S1) and dam (D1). The run of homozygosity finishes where a discordant SNP is identified.

mon between 2 individuals (i and j). The number of haplotypes that are in common was summed over S markers across the genome. ROH_{ij} is then the proportion of haplotypes of length n that are in common between individuals (i and j). An example of an ROH for 2 individuals (S1 and D1) is shown in Figure 2.

Phenotype Data

The breeding objective in this study was to maximize the APR, a total merit index of EBV of 9 traits weighted by their respective index weights and calculated by the Australian Dairy Herd Improvement Scheme (ADHIS; Pryce et al., 2009). The APR values for the bulls from August 2010 were retrieved from the ADHIS database. The APR can be interpreted as dollars net profit per cow per year. Therefore, 1 APR would be worth \$1 in merit passed on to the next generation. The standard deviation of APR in all (2,117) bulls in this data set was \$62.

Design of Mating Plans

A herd size of 300 cows (typical of Australia) was considered in all the scenarios tested. In each replicate, 320 bulls were randomly selected from the 2,117 bulls available for the analysis. The bulls were sorted according to their APR EBV and the 20 with the highest APR values were labeled as sires and the remaining 300 were allocated as “dams.” There were 50 replicates of 320 randomly selected individuals. These replicates

mimic different “herds” with different cows and bulls available for mating.

The following strategy was used to allocate matings for each of the 300 dams within a herd (replicate). Estimates of genetic merit minus the penalty for estimated progeny inbreeding of all 300×20 potential matings (i.e., sires and dams in a herd replicate; **B**) were calculated. The progeny inbreeding coefficient from the matings is estimated as half the genetic relationship between its sire and dam, as described previously. The genetic merit of potential calves was calculated as half of the sire’s APR plus half the dam’s APR; then, a weighted penalty on the estimated progeny inbreeding coefficient was subtracted (Kinghorn, 1998). Then,

$$B_{ij} = 1/2(APR_i + APR_j) - \lambda F_{ij},$$

where λ is a weighting factor on estimated progeny inbreeding and F_{ij} was estimated progeny inbreeding of a mating between sire i and dam j and obtained from the appropriate element of one of the following matrices of estimated progeny inbreeding: **A**, **G**, **G3k**, or **ROH** of length 5, 100, 300, or 500 SNP. A matrix of selected mates (mate allocation matrix; **M**) was constructed, where $M_{ij} = 1$ if the corresponding element, B_{ij} was the highest value in the column B_j ; that is, the maximum value of all feasible matings for dam j , all other elements were set to 0, and were rejected sire and dam combinations. The maximum number of matings allowable for each sire was set to 30. If B_{ij} corresponded to a sire that had already been assigned to 30 other matings, then the sire-dam combination with the next highest value in B_j was selected. If that bull also already had 30 matings allocated, then again, the next highest element in B_j was selected, and so on. As columns corresponded to dams and rows to sires, the sum of each column of **M** should be 1, whereas the sum of values in each row was less than or equal to 30. The average APR for a scenario was calculated by first constructing a matrix (**D**) of the estimated APR values of progeny calculated as:

$$D_{ij} = [1/2 (APR_i + APR_j)] M_{ij}.$$

Then, the average APR is $V'DV/300$, where **V** is a vector of 1s of length 300 (i.e., equivalent to the number of dams). The same method was applied to calculate the average estimated progeny inbreeding using the **F** matrix (from **A**, **G**, **G3k**, or **ROH**).

Penalty Applied to Progeny Inbreeding

Values used for the penalty weighting applied to progeny inbreeding (λ) ranged between 0 (no control)

and 20 per 1% increase in estimated progeny inbreeding. Although we considered a range of values of λ in our study, literature estimates cite approximate values of the cost of a 1% increase in inbreeding. Inbreeding affects profitability by adversely affecting traits related to fitness and production (Smith et al., 1998). Haile-Mariam et al. (2007) calculated the cost of inbreeding depression in Australian situations in 2 ways. Based on Australian data (Man, 2004) the cost of inbreeding was 3.1 APR units per 1% increase in inbreeding. Using data from the United States, Smith et al. (1998) reported that the cost of inbreeding over a lifetime was US\$24, which Haile-Mariam et al. (2007) converted to an equivalent cost of 8.5 APR units. Based on the study of Haile-Mariam et al. (2007), in addition to a range in values of λ (between 0 and 20), a conservative value of λ of \$5 was used as the economic value per 1% increase in inbreeding.

RESULTS

The correlation between off-diagonal elements of the **G** and **A** matrices was estimated using 50 replicates of 20 sires and 300 dams and was 0.76. When **G3k** was used instead of **G**, the correlation with **A** matrix co-ancestries was 0.71. The correlation between **G** and **A** is affected by the depth of the pedigree. When the number of generations of recorded ancestry was 2, 4, 6, and 8, the correlation between **G** and **A** was 0.67, 0.73, 0.84, and 0.87, respectively (Figure 3). When there were only 2 generations of recorded pedigree, the amount of information was insufficient to distinguish between close and more distant relationships. This could be one reason why more variation is found in the GRM relationships of 2 generations of recorded pedigree; that is, vertical spread in Figure 3a.

The correlations between co-ancestries estimated using **ROH** and **G** and **ROH** and **A** followed similar patterns (Figure 4). When **ROH** was 5 SNP long (approximately 0.34 Mbp) the correlation between co-ancestries was high (0.76 and 0.71 with **G** and **A**, respectively). The correlation decreased to 0.69 and 0.64 with **G** and **A**, respectively, when **ROH** was 50 SNP (approximately 3.35 Mbp). The second peak appears to be at around 250 SNP (16.8 Mbp). At very long **ROH** (i.e., >300 SNP and 20.1 Mbp), the relationship of **ROH** with both **G** and **A** was lower again.

The correlations between GRM and **ROH** of lengths 5, 50, 100, and 300 SNP are shown in a different way in Figure 5. When **ROH** was very short (5 SNP and approximately 0.34 Mbp), 23% of haplotypes, on average, were shared between a pair of individuals (Figure 6). In Figure 5, a dense region can be seen that corresponds to low genomic relationships and around 20

to 25% of haplotypes being shared. As the length of **ROH** increased, initially a rapid reduction occurred in the average proportion of haplotypes in common. When **ROH** was 500 SNP (around 33.5 Mbp), very few haplotypes were shared between all of the pairs of individuals (0.3%; Figure 6). This phenomenon results in a truncation of the shared **ROH** (Figure 5), which causes decay in the correlation between **GRM** and **ROH** at long lengths of **ROH** (Figure 4). At a length of 300 SNP, many instances occurred where individuals that have a GRM coefficient >0.1 actually have no shared runs of homozygosity of 300 or greater SNP (Figure 5d).

The main limitation of comparing methods to predict progeny inbreeding is that no best method of measuring inbreeding exists. Pedigree is flawed by errors and gaps and often the depth of pedigree, whereas genomic relationships calculated using SNP data sometimes have errors from incorrect identification of samples. For this reason, the results for controlling progeny inbreeding using **G**, **G3k**, and **A** are presented on both the **G** and **A** scales (Figure 7a and 7b). We also presented the results for **G**, **G3k**, and **A** on the **ROH** scale (Figure 8a and 8b). This is a more independent measure, but the correlations presented in Figure 4 between **ROH** of varying length and **G** and **A** indicate that this method is likely to favor **G** over **A**.

The most effective way of reducing the expected inbreeding of progeny was to use **G** (when progeny inbreeding was measured on the **G** scale), followed by **G3k**, and then the **A** matrix (Figure 7a). If average predicted progeny inbreeding was measured using **G3k**, then **G3k** was the most superior method (Figure 7b). Likewise, if **A** was used, then **A** was better than both **G** and **G3k** (Figure 7c).

When **G** was used to measure estimated progeny inbreeding, to reduce pedigree (**A**) inbreeding by 1%, then the average APR in the next generation would decrease by 0.5 APR compared with selection with no penalty on progeny inbreeding (Figure 7a). Using **G**, an equivalent reduction of 0.5 APR points would lead to a reduction in progeny inbreeding of 1.9%. That is, using information from the **G** matrix gives almost twice the reduction in inbreeding compared with using information from the **A** matrix, for the same reduction in genetic gain, provided that we accept that true inbreeding can be measured more accurately by **G** than **A**.

According to Haile-Mariam et al. (2007), a value of \$5 is a suitable (and conservative) economic value for a 1% increase in inbreeding in Australian conditions. Using this value for λ , the reduction in average predicted progeny inbreeding using **G** was 1.9% (on the **G** scale; Figure 7a) and the change in APR was -0.44 units (Figure 7a). When higher values of λ were used, there

Figure 3. Relationships between the pedigree relationship matrix (**A**) and the genomic relationship matrix (**G**) at (a) 2, (b) 4, (c) 6, and (d) 8 generations of recorded pedigree ancestry. Color version available in the online PDF.

was a greater trade-off in APR for a relatively small incremental reduction in inbreeding. But even a high trade-off in APR (3 APR units) is less than 0.05 of a genetic standard deviation of APR.

In Figure 8, where controlling inbreeding was achieved through using **G** or **A** and assessed on the same scale, namely **ROH** (of 10 or 100 SNs), it can be seen that the same pattern emerges as Figure 7a. For both lengths of **ROH**, **G** was most effective at reducing the proportion of runs of homozygosity in common in the next generation, followed by **G3k** and then pedigree; **ROH** of length 10 and 100 SNP were chosen, as they provide an interesting comparison. Their respective correlations with **G** were similar (Figure 4; 0.74 and 0.76 with **ROH** of lengths 10 and 100, respectively), whereas the frequency of **ROH** of length 10 was 11.5% and **ROH** of length 100 was much rarer (3%; Figure 6). Therefore, in view of the correlation between **G** and

ROH of length 10 and 100 being similar, perhaps the fact that fewer individuals shared **ROH** of length 100 was the reason why the performance of **G** was lower on **ROH** of length 100 than **ROH** of length 10.

The **ROH** itself can also be used as a tool to control inbreeding, as it may be desirable to avoid common runs of homozygosity. Then, progeny inbreeding was quantified using **G**. The most effective length of **ROH** to control inbreeding was 5 SNP (an average segment size of ~0.34 Mbp), whereas the least effective was 500 SNP (~33.5 Mbp). When the average progeny APR was decreased by 1 unit, **ROH** of 5 and 100 SNP reduced inbreeding by around 1.5 and 1.3%, respectively (Figure 9).

An independent way of comparing the effectiveness of reducing expected inbreeding in progeny is to investigate the proportion of homozygous minor allele SNP not in the marker panel used for calculating **G**.

Figure 4. Correlations of 50 replicates of co-ancestries calculated between 20 sires and 300 dams of runs of homozygosity of varying lengths (measured in SNP) with pedigree (**A**) and a genomic relationship matrix (GRM, **G**). Color version available in the online PDF.

In the potential offspring (i.e., all 4 allele combinations of each selected sire and dam), 0.032% of SNP were homozygous recessive when there was no control on progeny inbreeding. As λ was increased from 0 to 20, the percentage of homozygous recessive decreased when **A**, **GRM**, or **ROH** of length 5, 50, or 300 were used to control inbreeding. For **G**, the reduction in inbreeding appeared to plateau at around half its initial level at 0.018% (when λ was greater than 17; Figure 10). The proportion of minor allele homozygotes, when the **A** matrix was used instead of the **GRM** was 0.027%, and the plateau was reached with lower values of λ . That is, using information from **G** leads to half the frequency of minor allele homozygotes in progeny of **A**.

DISCUSSION

The results presented here show that using a GRM instead of pedigree in a mating plan is an effective way to reduce the expected inbreeding in progeny, with minimal effect on genetic gain for the breeding objective (here APR, a measure of net profit, was used). For example, by reducing pedigree (**A**) progeny inbreeding by 1%, the average APR in the next generation would decrease by \$0.50 compared with selection with no penalty on inbreeding in the progeny. In comparison, using information from **G**, an equivalent reduction of \$0.50 in APR would lead to an almost twice as great reduction in inbreeding of 1.9%. The reduction in inbreeding when using **G3k** was very similar.

The net benefit of controlling inbreeding can be calculated by subtracting the loss in genetic gain from APR from the value of reducing inbreeding. If the goal is to reduce inbreeding in progeny by between 1 and

2% using genomic technologies, then this is worth between \$4.96 and \$9.71 (Figure 7). In fact, the economic benefit of controlling inbreeding far exceeds the penalty in reducing the average APR in the next generation (which is less than \$0.70 when the desire is to reduce inbreeding by 1–2%). By itself, this economic benefit does not currently justify the investment in genotyping, if one considers that pedigree information is “free” and appears to do a reasonably good job of controlling inbreeding. However, it may be economically worthwhile for dairy farmers to invest in whole-herd low-density genotyping in conjunction with other uses of genotyping. Examples could include confirming parentage, selecting the best heifer calves to keep as herd replacements, managing genetic defects, and selling high-value pedigree stock. However, a full cost-benefit analysis of all the components that contribute to the value of genotyping is necessary. It seems likely, in the short term at least, that dairy herds will have a combination of genotyped and nongenotyped individuals. Therefore, a relationship matrix that includes a combination of elements from both the genomic and pedigree relationship matrices on the same scale will be required.

The reductions in inbreeding using **G**, **G3k**, and **A** were dependent on the way in which inbreeding was assessed (Figure 7). For example, the performance of **G** was superior when progeny inbreeding was measured using **G** (Figure 7a) and **A** was superior to **G** when progeny inbreeding was measured using **A** (Figure 7c). This shows that the method of measuring inbreeding is important when assessing different strategies to control inbreeding. The fact that the reduction in inbreeding using **G3k** (Figure 7b) was greater than that using either **G** (Figure 7a) or **A** (Figure 7b) could be misleading, because it implies that **G3k** was the most effective strategy. However, in none of these examples was the measurement scale independent of the method used to control inbreeding. We attempted to obtain an independent comparison by using the proportion of homozygous minor alleles of 1,000 SNP excluded from GRM calculations as a criterion. Although the reduction in homozygous alleles was small, the results presented in Figure 10 show that using the GRM to control inbreeding would reduce the frequency of homozygous minor alleles more effectively than either using pedigree or ROH. By inference, using the GRM would minimize the decrease in heterozygosity of SNP not under selection.

The availability of dense SNP data has made it possible to move beyond using pedigree data to estimate relationships and to use genotype data instead. Provided the SNP are dense enough, they will more accurately describe the co-ancestry between individuals than pedigree, which is based on expected related-

Figure 5. Scatterplots of runs of homozygosity (ROH) measured using SNP of length (a) 5, (b) 50, (c) 100, and (d) 300 against genomic relationships (G) between 320 Holstein bulls. Color version available in the online PDF.

ness. We have demonstrated that, for Holstein cattle at least, G and $G3k$ reduce inbreeding at comparable rates (Figure 7a and Figure 7b). This could mean that moving to denser genotype panels is not actually worthwhile for commercial on-farm mating plans in Holstein cattle. Furthermore, as the cost of genotyping depends on the number of SNP in the panel, it seems likely that farmers are more likely to invest in genotyping entire herds of females with low-density (3,000) SNP technology rather than with 50K (50,000).

The 2,117 bulls used in this analysis had reasonably deep and accurate pedigrees. For commercial farms, incomplete pedigrees and errors in parentage can be

more frequent. In fact, the advantage of using the genomic approach described here can be even greater for animals with few recorded generations of pedigree, as the correlation between G and A increased with the number of generations of recorded ancestry (Figure 3). However, even with very deep, accurate pedigrees, the GRM captures more information, or subtle relationships, than pedigree. For example, even though the pedigree relationships between a set of full or half sibs are equal, the genomic relationships between them can vary.

The bulls used in this study were born over many years and therefore had a greater range in APR val-

Figure 6. The average proportion of shared identical runs of homozygosity (ROH) in 50 replicates of 320 individuals when the length of ROH is varied from 5 to 500 SNP. Color version available in the online PDF.

ues and were more genetically diverse than might be expected for cows on a typical Australian dairy farm. In addition, as pseudo “males” were selected based on APR, they tended to be the younger bulls and were therefore genetically more similar to each other and genetically more different from the older pseudo “females.” On average, there were around 3 sons per sire. In commercial dairy farms, the number of daughters per sire ranges from 8 to 15, assuming a replacement rate of 25% for a 300-cow herd. In fact, the trade-off in the reduction of genomic inbreeding compared with the reduction in the APR will vary between herds because of sire choice and the closeness of relationships between those sires and cows currently in the herd. The reduction in inbreeding could be even greater in applying this methodology to commercial dairy cows.

Using runs of homozygosity to control the rate of inbreeding is a novel approach and, to the best of our knowledge, this is the first paper to report such a study. The correlation between **ROH** and pedigree inbreeding coefficients ranged between 0.64 and 0.71 (Figure 4). This is lower than correlations estimated in human populations ($r = 0.87$; Kirin et al., 2010) and could be due to depth of pedigree, SNP density, or fewer IBS versus IBD issues. Although **ROH** was not as effective as **G** in reducing the homozygosity of low frequency alleles (Figure 10), differences were found between the reduction of expected progeny inbreeding achieved when **ROH** of varying lengths were used (Figure 9). Out of the **ROH** tested, a length of 5 SNP appeared to be the most effective **ROH** matrix to control inbreeding in the next generation (Figure 9). This could have been because it was the shortest length of **ROH** evaluated and therefore most closely related to single SNP and the **G** matrix that was used to assess the **ROH** tested.

Figure 7. The effect on Australian Profit Ranking (APR) and average progeny inbreeding (F) when the reduction in F was achieved using a penalty weight ($\lambda = 0-20$) applied to a pedigree relationship matrix (**A**) and a genomic relationship matrix estimated with around 40,000 and 3,000 SNP (**G** and **G3k**, respectively) in 50 replicates of 320 animals. The unit of measurement for the reduction in F was estimated using (a) **G** (b) **G3k**, and (c) **A**. Color version available in the online PDF.

Runs of 5 SNP (~ 0.34 Mbp) were also very common in the dairy cattle population used in this study (Figure 7); on average, 0.23 were shared between individuals. Longer runs were rarer; for example, 50 SNP were shared with a frequency of, on average, 5%, whereas

Figure 8. The effect on Australian Profit Ranking (APR) and average progeny inbreeding measured using runs of homozygosity (ROH) of (a) 10 SNP and (b) 100 SNP when the reduction in co-ancestry (ROH) was achieved using a penalty weight ($\lambda = 0-20$) applied to a pedigree relationship matrix and a genomic relationship matrix estimated with 40,000 and 3,000 SNP (**G** and **G3k**, respectively) when selection was for APR in 50 replicates of 320 animals. Color version available in the online PDF.

500 SNP were shared with a frequency of 0.3%. This relationship partly explains the shape of the plot of correlations shown in Figure 5, where very short (5 SNP) and long (300 SNP) **ROH** had a strong relationship with **G** and pedigree calculated co-ancestries. The peak in correlation estimates that can be seen at around 20 Mbp (**ROH** of 300 SNP; Figure 4) appears to occur because of truncation (Figure 5d), due to long runs of shared homozygosity being rare (on average 0.8% are shared; Figure 6). Genotyping errors could result in lower than expected long shared ROH. It is also worth noting that to calculate ROH accurately, a method that phases using linkage information was important (e.g., ChromoPhase; Daetwyler et al., 2011).

The shape of the correlations between **G** and **ROH** (and between **A** and **ROH**) is unusual and worth exploring (Figure 4). Four of the correlations presented in Figure 4 are presented as scatterplots in Figure 5. The variation in the proportion of shared ROH is consistently largest when the **ROH** is 50 SNP in length.

Figure 9. The effect on Australian Profit Ranking (APR) and progeny inbreeding measured using a genomic relationship matrix (**G**) when the reduction in average progeny inbreeding (F) was achieved using a penalty weight ($\lambda = 0-20$) applied to shared runs of homozygosity of 5, 100, 300, and 500 SNP in length and selection was for APR in 50 replicates of 320 animals. Color version available in the online PDF.

For example, for a genomic relationship of 0.125 (i.e., elements of the **G** that equaled 0.125), the standard deviations of **ROH** were 0.036, 0.062, 0.052, and 0.031 for lengths 5, 50, 100, and 300. When **A** and **G3k** were used instead of **G**, the standard deviation was still largest when **ROH** was 50 SNP long. One might expect the divergence in relationships detected to increase with an increasing number of SNP in the **ROH**. However, this is not the case because of a truncation effect arising from few shared long runs of homozygosity even between closely related individuals. Therefore, the low correlation of **ROH** with a length of 50 SNP with **G** is because **ROH** detects subtle differences in relation-

Figure 10. The proportion of homozygous minor alleles out of 1,000 SNP of lowest allelic frequency in 50 replicates of 320 animals, when the reduction in co-ancestry was achieved using a penalty weight (λ , 0–20) applied to pedigree (**A**), genomic relationship matrix (**G**), genomic relationship matrix with 3,000 SNP (**G3k**), or runs of homozygosity of length 5 (**ROH5**), length 50 (**ROH50**), and length 300 (**ROH300**). Color version available in the online PDF.

ships that occur because of shared runs of homozygosity. Therefore, individuals that have the same genomic relationship; that is, a very similar **G** value but very different **ROH**, are likely to have different genetic backgrounds. For the same **G** value, a low **ROH** could arise from more ancient inbreeding, whereas a high **ROH** is indicative of more recent inbreeding. The fact that 2 subpopulations are visible in Figure 5 gives weight to this argument. Because the **ROH** is able to detect recent versus older inbreeding, it may be a promising tool for the control of inbreeding. One strategy would be to give larger **ROH** (i.e., recent inbreeding) a higher penalty than smaller **ROH**. However, **ROH** of length 50 were not as effective as the **G** methods in reducing the level of homozygous recessive alleles (Figure 10).

Mating plans such as the options we have described are deliberately designed to control the rate of inbreeding in a subset of individuals (such as a herd) in the next generation. If maximizing long-term genetic gain while controlling inbreeding in the entire population is the desired outcome, then strategies that select candidates including information on co-ancestry among the selection candidates should be used (e.g., Wray and Goddard, 1994; Meuwissen, 1997). Recently, Sonesson et al. (2010) extended these optimum contribution strategies to include genomic information. Here, these authors clearly demonstrated the importance of genomically controlling inbreeding under genomic selection instead of pedigree-controlled inbreeding. It would be reasonably straightforward to extend the methodology described here to include both mate allocation and optimal contributions in a mate selection strategy; that is, select a subset of animals from the candidates available if selection pressure is required in the cow to breed cow pathway. For example, the method proposed by Kinghorn (1998) uses an evolutionary algorithm to efficiently search through possible matings to find a solution (selected set of matings) of animals that is optimal not only for short-term inbreeding, but also for long-term inbreeding. Logistical constraints and running costs could also be included if desired.

CONCLUSIONS

The availability of high-density SNP data has revolutionized dairy cattle breeding, allowing restructuring of breeding schemes and greater rates of genetic gain. Given that greater rates of genetic gain from genomic selection could be associated with higher annual rates of inbreeding, methods to control inbreeding will become increasingly important. In commercial dairy herds, using mating plans should become more attractive as farmers aim to maximize returns on their investment

from genotyping. Based on our calculations, the value of genotyping to control inbreeding could be worth between \$5 and \$10 per cow.

ACKNOWLEDGMENTS

The authors acknowledge the Department of Primary Industries (Victoria, Australia) and Dairy Futures Cooperative Research Council (Melbourne, Australia) for funding this research. We thank Hans Daetwyler (DPI-Victoria, Australia) and Iona MacLeod (University of Melbourne, Australia) for useful discussions.

REFERENCES

- Coster, A. 2010. Package to deal with pedigree data. Accessed Jan. 5, 2011. <http://cran.r-project.org/web/packages/pedigree/pedigree.pdf>.
- Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. Inbreeding in genome-wide selection. *J. Anim. Breed. Genet.* 124:369–376.
- Daetwyler, H.D., G.R. Wiggans, B.J. Hayes, J.A. Woolliams, and M.E. Goddard. 2011. Imputation of genotypes from sparse to high density using long-range phasing. *Genetics* doi:10.1534/genetics.111.128082.
- Haile-Mariam, M. E., P. J. Bowman, and M. E. Goddard. 2007. A practical approach for minimising inbreeding and maximising genetic gain in dairy cattle. *Genet. Sel. Evol.* 39:369–389.
- Hayes, B. J. 2011. Technical note: Efficient parentage assignment and pedigree reconstruction with dense single nucleotide polymorphism data. *J. Dairy Sci.* 94:2114–2117.
- Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res. (Camb.)* 91:47–60.
- Hayes, B. J., P. M. Visscher, H. C. McPartlan, and M. E. Goddard. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res.* 13:635–643.
- Kinghorn, B. P. 1998. Managing genetic change under operational and cost constraints. Pages 9–16 in *Proc. 36th National Congress of the South African Association of Animal Science*, University of Stellenbosch.
- Kinghorn, B. P. 2011. An algorithm for efficient constrained mate selection. *Genet. Sel. Evol.* 43:4.
- Kirin, M., R. McQuillan, C. S. Franklin, H. Campbell, P. M. McKie, and J. F. Wilson. 2010. Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE* 5:e13996.
- Lillehammer, M., T. H. E. Meuwissen, and A. Sonesson. 2011. A comparison of dairy cattle breeding designs that use genomic selection. *J. Dairy Sci.* 94:493–500.
- MacLeod, I. M., T. H. E. Meuwissen, B. J. Hayes, and M. E. Goddard. 2009. A novel predictor of multilocus haplotype homozygosity: Comparison with existing predictors. *Genet. Res. (Camb.)* 91:413–426.
- Man, W. Y. N. 2004. Inbreeding in Australian Holstein Friesian cattle. PhD Thesis. University of Sydney, Australia.
- Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. L. Smith, T. S. Sonstegard, and C. P. Van Tassell. 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* 4:e5350.
- Meuwissen, T. H. E. 1997. Maximizing the response of selection with a predetermined rate of inbreeding. *J. Anim. Sci.* 75:934–940.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Pryce, J. E., S. Bolormaa, A. J. Chamberlain, P. J. Bowman, K. Savin, M. E. Goddard, and B. J. Hayes. 2010a. A validated genome wide

- association study in two dairy cattle breeds for variable length haplotypes. *J. Dairy Sci.* 93:3331–3345.
- Pryce, J. E., M. E. Goddard, H. W. Raadsma, and B. J. Hayes. 2010b. Deterministic models of breeding scheme designs that incorporate genomic selection. *J. Dairy Sci.* 93:5455–5466.
- Pryce, J. E., J. H. J. van der Werf, M. Haile-Mariam, B. Malcolm, and M. E. Goddard. 2009. Updated index weights for the Australian Profit Ranking in dairy cattle. *Proc. Assoc. Advmt. Anim. Genet.* 18:143–146.
- Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123:218–223.
- Scheet, P., and M. A. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78:629–644.
- Smith, L. A., B. G. Cassell, and R. E. Pearson. 1998. The effects of inbreeding on lifetime performance of dairy cattle. *J. Dairy Sci.* 81:2729–2737.
- Sonesson, A. K., J. A. Woolliams, and T. H. E. Meuwissen. 2010. Maximising genetic gain whilst controlling rates of genomic inbreeding using genomic optimum contribution selection. Abstract 0892 in *Proc. 9th World Congr. Genet. Appl. Livest. Prod.* German Society for Animal Science, Leipzig, Germany.
- Toro, M., and M. Varona. 2010. Genomic selection and mate allocation. Abstract 0268 in *Proc. 9th World Congr. Genet. Appl. Livest. Prod.* German Society for Animal Science, Leipzig, Germany.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423.
- Wray, N. R., and M. E. Goddard. 1994. Increasing long-term response to selection. *Genet. Sel. Evol.* 26:431–451.
- Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, and P. M. Visscher. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42:565–569.