Long-term selection strategies for complex traits using high-density genetic markers

K. E. Kemper,*¹ P. J. Bowman,†† J. E. Pryce,†† B. J. Hayes,††§ and M. E. Goddard*††

*Melbourne School of Land and Environment, University of Melbourne, Parkville, Victoria 3010, Australia

†Biosciences Research Division, Department of Primary Industries, Bundoora, Victoria 3083, Australia

‡Dairy Futures Cooperative Research Centre, Bundoora, Victoria 3083, Australia

§La Trobe University, Bundoora, Victoria 3086, Australia

ABSTRACT

Selection of animals for breeding ranked on estimated breeding value maximizes genetic gain in the next generation but does not necessarily maximize long-term response. An alternative method, as practiced by plant breeders, is to build a desired genotype by selection on specific loci. Maximal long-term response in animal breeding requires selection on estimated breeding values with constraints on coancestry. In this paper, we compared long-term genetic response using either a genotype building or a genomic estimated breeding value (GEBV) strategy for the Australian Selection Index (ASI), a measure of profit. First, we used real marker effects from the Australian Dairy Herd Improvement Scheme to estimate breeding values for chromosome segments (approximately 25 cM long) for 2,650 Holstein bulls. Second, we selected 16 animals to be founders for a simulated breeding program where, between them, founders contain the best possible combination of 2 segments from 2 animals at each position in the genome. Third, we mated founder animals and their descendants over 30 generations with 2 breeding objectives: (1) to create a population with the “ideal genotype,” where the best 2 segments from the founders segregate at each position, or (2) obtain the highest possible response in ASI with coancestry lower than that achieved under breeding objective 1. Results show that genotype building achieved the ideal genotype for breeding objective 1 and obtained a large gain in ASI over the current population (+A\$864.99). However, selection on overall GEBV had greater short-term response and almost as much long-term gain (+A\$820.42). When coancestry was lowered under breeding objective 2, selection on overall GEBV achieved a higher response in ASI than the genotype building strategy. Selection on overall

GEBV seems more flexible in its selection decisions and was therefore better able to precisely control coancestry while maximizing ASI. We conclude that selection on overall GEBV while minimizing average coancestry is the more practical strategy for dairy cattle where selection is for highly polygenic traits, the reproductive rate is relatively low, and there is low tolerance of coancestry. The outcome may be different for traits controlled by few loci of relatively large effects or for different species. In contrast to other simulations, our results indicate that response to selection on overall GEBV may continue for several generations. This is because long-term genetic change in complex traits requires favorable changes to allele frequencies for many loci located throughout the genome.

Key words: long-term response, genomic selection, genotype building, optimal contribution

INTRODUCTION

In genomic selection (Meuwissen et al., 2001), the genomic estimated breeding value (GEBV) of an animal is obtained by adding the estimated effects of all marker alleles that an animal carries. Animals with the highest overall GEBV are then selected for breeding. However, plant breeders often use a different, more targeted, strategy that builds a predefined genotype. They may use techniques such as backcrossing and gene stacking or pyramiding to introgress one or more desirable loci into an elite variety (Zamir, 2001; Dekkers and Hospital, 2002; Hospital, 2003). Building of a desired genotype using introgression increases genetic variation because the frequencies of rare favorable alleles are increased, potentially improving long-term selection response (Hill, 2000). Selection based on overall GEBV gives the highest response in the next generation but it is not guaranteed to maximize gains in the long term. Strategies for optimal long-term response require a trade-off between short-term improvement in GEBV and coancestry (Wray and Goddard, 1994; Meuwissen, 1997). Our aim here was to determine whether geno-

Received December 22, 2011.

Accepted March 19, 2012.

¹Corresponding author: kathryn.kemper@dpi.vic.gov.au

type building is an appropriate strategy for long-term selection response for a complex trait in dairy cattle.

Several factors prevent a genotype-building strategy from being implemented in animal breeding as it is in plants. The first barrier is genetic architecture, where complex traits in dairy cattle are highly polygenic and therefore require thousands of markers to accurately predict phenotype (Cole et al., 2009; Hayes et al., 2010). It is not practical, in a moderate number of generations, to combine the thousands of desired alleles into one genotype because the reproductive and recombination rates in cattle are not high enough. Even if the desired genotype could be constructed, the phenotype may be less desirable than the EBV because the effect of individual markers is small and the estimates may have large sampling errors (Goddard, 2009). However, the effect of chromosome segments is more accurate than that of the individual SNP markers. This is because compensating errors can occur in the estimates of individual SNP that are in linkage disequilibrium (Goddard, 2009).

The second barrier preventing animal breeders from adopting a genotype-building strategy is the high degree of inbreeding necessary for plant breeding designs. For example, gene stacking in plants often aims to create a homozygous (or “true breeding”) variety at specified loci (Hospital, 2003). However, in dairy cattle, creating an “ideal genotype” that was homozygous throughout the genome would be risky because of recessive lethal alleles and inbreeding depression (Von Krosigk and Lush, 1958; Thompson et al., 2000; VanRaden et al., 2011). Therefore, a more useful definition of the ideal genotype for cattle would be a population of individuals with 2 very good alleles segregating at each position in the genome. A heterozygous individual would be best but it is impossible to maintain a heterozygous random-mating population because of allele segregation when forming gametes. If 2 alleles were segregating at each position, the coancestry (f) in the final population would be approximately 0.5 because each animal would be homozygous for half the genome. However, in dairy cattle, this level of coancestry is probably undesirably high. Lower coancestry could be obtained with several desirable alleles segregating at each position in the genome so that homozygosity in the population is reduced. One way to obtain such a population is to specify a target frequency for each allele in the final population. The ideal number of alleles and their target frequencies could be calculated using the optimum contribution theory of Wray and Goddard (1994). Achieving a population with lower coancestry than the ideal genotype would provide an appropriate trade-off between maximizing mean breeding value and minimizing coancestry (or inbreeding) in the population.

In this study, we compared the long-term selection response for a complex trait in dairy cattle using either a genotype building strategy or an approach that selects an overall GEBV with a penalty for coancestry. We assumed a simplified model of the bovine genome consisting of 25-cM chromosome segments to minimize errors associated with individual marker effects. Thus, the objective was to build a genome of high breeding value using the best chromosome segments available in the current Australian Holstein population. We implemented the genotype-building strategy using 2 algorithms based on Goddard (2009) to achieve a target frequency for each segment at each position in the genome. The second approach, based on the overall GEBV, applied the method of Wray and Goddard (1994) and Meuwissen (1997) to maximize genetic gain subject to constraints on marker-derived coancestry (Sonesson et al., 2010). The aim was to determine which of these strategies is better for long-term response to selection.

Several steps were needed to simulate the breeding program and compare the 2 strategies. First, we present an algorithm to select 16 founder animals that, between them, contain the best combination of 2 chromosome segments from 2 founders at all positions in the genome. Next, we used these founder animals to simulate a breeding program that aimed to create either (1) a population with the “ideal genotype,” where the 2 best segments identified from the founders segregate with equal frequency, or (2) a population with the highest possible GEBV and coancestry lower than the ideal genotype (i.e., $f < 0.5$), implying that more than 2 segments segregate at each position.

MATERIALS AND METHODS

Data

The SNP marker effects and genotypes for Holstein bulls ($n = 2,650$) were obtained from the Australian Dairy Herd Improvement Scheme (ADHIS) in April 2011 (www.adhis.com.au). Marker effects were for protein (kg), fat (kg), and milk (L) production per lactation. Genotypes consisted of 45,358 SNP distributed across the 29 bovine autosomes.

The trait under selection was the Australian Selection Index (ASI). This index is calculated by ADHIS and is a measure of profit, expressed in dollars per lactation, reflecting the income received for milk minus feed costs. Three steps were taken to calculate the chromosome segment effects for each animal. First, the ASI marker effect for each SNP was calculated using the ADHIS weightings for protein, fat, and milk. This calculation was

$$\hat{b}_{ASI} = 7.096\hat{b}_{protein} + 1.434\hat{b}_{fat} - 0.084\hat{b}_{milk},$$

where $\hat{b}$ is the marker effect for ASI, protein, fat, and milk, respectively. Second, animal genotypes were phased into chromosomes using Beagle 3.3.0 (Browning and Browning, 2007), and then divided into approximately 25-cM lengths using map positions from UMD3.0 (Zimin et al., 2009). In total, 101 segment positions (2–6 segments per chromosome) had an average length of 24.6 cM (range: 20.6 to 30.5 cM). Lengths of 25 cM were chosen to minimize possible compensating errors between adjacent markers while permitting recombination between segments. Thus, at each position in the genome there were 2 segments for each animal. Third, the ASI effect for each segment was calculated as the sum of the marker allele effects; that is, $\mathbf{X}\hat{\mathbf{b}}_{ASI}$, where $\mathbf{X}$ is an incidence matrix of SNP alleles (0 or 1) for each segment. The overall GEBV for an individual was the sum of all segment effects.

Selection of Founders for the Breeding Program

The aim was to select 16 founders such that the sum of the best chromosome segments from 2 animals at each position was as high as possible. Thus, an animal could be selected as a founder, even if they were not one of the best animals on overall GEBV, if they carried an individual chromosome segment or segments of high value. Segments were required to be from 2 different animals to minimize the number of homozygous positions. A genetic algorithm was developed to select the ancestors. The algorithm aimed to maximize ASI considering the constraints described above; that is, if x_{ij} is the breeding value of the better segment at position j from animal i , the objective was to maximize

$$\sum_{i=1}^{16} \sum_{j=1}^{101} a_{ij} x_{ij}, [1],$$

where a_{ij} is 0 (not a target segment) or

1 (a target segment) and $\sum_{i=1}^{16} a_{ij} = 2$ at all j . Thus, the 16 founders selected by the algorithm contained 2 segments of high breeding value for all 101 positions making up the genome. Further details of the genetic algorithm are given in the Supplementary Material (available online at <http://www.journalofdairyscience.org/>).

Breeding Objectives

Two breeding objectives were investigated. The aim of the first was to obtain the “ideal genotype,” where a population segregates in equal proportions for the 2 best segments identified from the founder animals at each position. Most positions had target frequencies of

0.5 but, for some positions, the 2 segments, although selected from different founders, were actually identical. For these homozygous positions, the target frequency was 1.

The second objective aimed to maximize response in ASI with coancestry lower than that assumed by the ideal genotype under breeding objective 1. This objective tested the flexibility of the genotype-building strategy because, to lower coancestry, more than 2 segments must segregate at each position in the genome. To do this for the genotype-building strategy, we defined target frequencies for segments by using the optimal contribution methodology of Wray and Goddard (1994). That is, $\mathbf{t}'\mathbf{x} - k_1\mathbf{t}'\mathbf{t}$ was maximized at each position, where $\mathbf{t}$ is a vector of target segment frequencies (maximum length of 32; i.e., 16 ancestors with 2 segments per position = 32 segments), $\mathbf{x}$ is a vector of ASI effects corresponding to the segments in $\mathbf{t}$, k_1 is the segment diversity penalty, and $\mathbf{t}'\mathbf{1} = 1$ (i.e., the sum of the segment frequencies at each position must equal 1). Using the LaGrange multiplier (λ), it can be shown that

$$\frac{\partial}{\partial \mathbf{t}} = \mathbf{t}' - k_1\mathbf{t}'\mathbf{t} - \lambda(\mathbf{t}'\mathbf{1} - 1) = 0.$$

Solving for $\mathbf{t}$ and λ ,

$$\mathbf{t} = \frac{\mathbf{x} - \lambda \mathbf{1}}{2k_1} \text{ and } \lambda = \frac{-2k_1 + \mathbf{x}'\mathbf{1}}{N},$$

where N is the number of

desirable segments. Beginning with $N = 1$, substituting for λ and realizing that no value of $\mathbf{t}$ can be negative, the criterion for including the $N + 1$ next-best segment

becomes $\mathbf{x}_{N+1} < \bar{x} - \frac{2k_1}{N}$, where $\bar{x}$ is the mean of the N segments previously selected. When the condition is no longer satisfied, the target frequency is calculated as $\mathbf{t} = \frac{\mathbf{x} - \bar{x}}{2k_1} - \frac{1}{N}$. A range of values for k_1 was chosen for successive simulations to generate the response curve for ASI and coancestry.

Breeding Scheme

Matings of the founder animals and their descendants were performed using stochastic simulation over 30 generations. Each segment was inherited independently; that is, free recombination occurred between segments but no recombination occurred within segments. Each generation consisted of 8 full-sib families with 200 progeny per family. Two progeny from each family were selected as parents for the next generation.

Selection Strategies

The 2 selection strategies (genotype building and GEBV selection) were compared for their ability to achieve the breeding objectives. For the genotype-

building strategy, we used 2 algorithms: logarithmic (**LOG**) and arcsin (**ARCSIN**; Table 1) based on the theory of Goddard (2009). These algorithms both attempt to minimize the time taken for segments to reach their target frequency in the population. The target frequency for breeding objective 1 was either 0.5 (2 desirable segments per position) or 1 (for homozygous positions), whereas for the second breeding objective, the target frequency was determined by the optimal contribution vector (**t**). Algorithms used either a logarithmic or arcsin transformation of segment frequencies, penalized the loss of desirable segments, and placed stronger selection pressure on favorable segments that were rare than on those that were common. The steps and further details of these algorithms are described in Table 1.

Selection on overall GEBV was implemented through the optimal contributions (**OPT**) algorithm (Table 1), based on the method of Wray and Goddard (1994). This strategy aimed to maximize overall GEBV while restricting coancestry. In contrast to ARCSIN and LOG, OPT does not consider the frequency or effect of individual segments. For breeding objective 1, a fair comparison of OPT to the genotype-building algorithms requires a similar level of coancestry. Therefore, the coancestry penalty for the OPT algorithm (k_2) was

chosen for breeding objective 1 such that the coancestry was equal to the expected coancestry of the “ideal genotype” [i.e., $f = 0.5 \times (1 + \text{proportion of homozygous segments in the ideal genotype})$]. Similarly, for breeding objective 2, k_2 was varied so that results from OPT could be compared with LOG and ARCSIN at similar coancestry. Simulations were the average of 100 (breeding objective 1) or 20 (breeding objective 2) replicate simulations for all algorithms.

Measures of Response

The strategies were compared in each generation using 4 measures. First, the distance to the ideal genotype was calculated in the selected individuals as $\sum_{j=1}^N \left[\arcsin(\sqrt{t_j}) - \arcsin(\sqrt{p_j}) \right]$, (Goddard, 2009), where N is the total number of target segments over all positions (i.e., length = $101 \times$ the average number of segments at each position), t_j is the target frequency of segment j , and p_j is the frequency of j . Second, the coancestry (f_{ik}) between parents was calculated as $0.5G_{ik}$, where G_{ik} is the relationship between individuals i and k , as calculated from identical segment effects (Nejati-Javaremi et al., 1997). That is,

Table 1. Selection algorithms used in the breeding programs

Algorithm/step	Process
Genotype-building strategy aiming to minimize the distance to the target genotype (ARCSIN)	
Step 1	Calculate $V_1 = \sum_{j=1}^M \beta$ for selection candidates, where $\beta = -100$ for $p_j = 0$ or $\beta = \left[\arcsin(\sqrt{t_j}) - \arcsin(\sqrt{p_j}) \right]^2$ for $p_j > 0$, and p_j is the frequency of the segment if the candidate is selected, M is the total number of target segments, and t_j is the target frequency of segment j . Select the candidate with the lowest V_1 .
Step 2	Repeat step 1 until 16 animals are selected. Update p_j at each step according to the current animals selected and the proposed candidate.
Step 3	Return each selected animal to the selection candidates and repeat step 1. Continue until all animals in the selected group have been returned to the selection candidates at least once.
Genotype-building strategy using a logarithmic index (LOG)	
Step 1	Calculate $V_2 = \sum_{j=1}^M t_j \beta$ for selection candidates, where t_j is the target frequency of the desirable segment j , $\beta = -100$ for $p_j = 0$ or $\beta = \log(p_j)$ for $p_j > 0$. Select the candidate with the highest V_2 .
Step 2-3	As for ARCSIN but selecting to maximize V_2 .
Selection on overall GEBV ¹ with a constraint on coancestry (OPT)	
Step 1	Select the animal with the highest mean GEBV
Step 2	Calculate $V_3 = \mathbf{c}'\mathbf{EBV} - k_2\mathbf{c}'\mathbf{Gc}$, for the remaining candidates, where $\mathbf{c}$ is a $n + 1$ vector of the expected contributions from each animal; that is, $1/(n + 1)$, where n is the number of animals already selected; $\mathbf{EBV}$ is a $n + 1$ vector of breeding values for the n animals selected and the selection candidate; k_2 is the coancestry penalty, and $\mathbf{G}$ is the relationship matrix of rank $n + 1$ for selected animals and the new candidate. Select the candidate with the highest V_3 .
Step 3	Repeat step 2 until 16 animals are selected.

¹GEBV = genomic EBV.

Figure 1. Boxplots of chromosome segment effects for the Australian Selection Index at each of 101 positions across the bovine autosomes. Segments are ordered along chromosomes with the boxes of segments from alternate chromosomes shaded in dark gray (blue) and light gray. Color version available in the online PDF.

$G_{ik} = \frac{1}{101} \sum_j \left[\frac{1}{2} \sum_{m=1}^2 \sum_{n=1}^2 I_{mn} \right]$, where I_{mn} is the identity of the m th segment for individual i and the n th segment for individual j and takes the value of 0 (nonidentical segments) or 1 (identical segments). Third, the response in ASI was measured as the average overall GEBV of the selected individuals. Finally, the variance of GEBV in the selection candidates was calculated at each generation.

RESULTS

Distribution of ASI

Among the 2,650 bulls, the mean ASI was 21.37 and the standard deviation (SD) was 66.2. This is a slightly higher mean and slightly lower SD than the average Australian breeding value for the current Holstein population in Australia (mean ASI 16.6 and SD 69.0; ADHIS, 2011).

Distribution of ASI Segment Effects

Segment effects showed an approximate normal distribution, with a mean of 0.11 and SD of 2.2, and

an even distribution across the autosomes (Figure 1). There were between 1,762 and 3,650 different segments at each position (estimated to 17 decimal places). Detailed inspection of the segment effects for 2 animals, one with high and one with low overall ASI breeding value, showed the accumulation of above-average segments in the high ASI animal (Figure 2). Both animals had a similar range in segment values and, notably, the high ASI animal still contained some very poor (below average) segments.

Selection of the Founder Animals

The maximum ASI achieved by selecting 16 animals from the current population of dairy bulls and combining the 2 best segments from 2 animals was \$881.65. Segments from the founder animals showed an approximate normal distribution (Supplementary Figure S1; available online at <http://www.journalofdairyscience.org/>) with a slightly higher mean (0.57) and similar SD (2.3) compared with all animals. The mean overall ASI for individual founder animals was \$115.71, with each founder contributing between 5 and 21 segments to the “ideal genotype.” Assuming that the fat and protein percentages of milk remain at the current level for milk in Australia (3.9 and 3.3%, respectively), an ASI of

Figure 2. Chromosome segment effects for the Australian Selection Index (ASI, \$) at each position for 2 animals with high and low overall ASI.

\$882 equates to an additional 4,280 L of milk, 169 kg of fat, and 141 kg of protein per lactation. The ideal genotype was homozygous at 17 (16.8%) positions, indicating that some ancestors were close relatives and contributed identical segments to the target genotype.

Notably, the first iteration of the genetic algorithm showed that combining the best segments from any 16 animals would result in an average ASI of \$727 (Supplementary Figure S2; <http://www.journalofdairyscience.org/>). This indicates high segment variation within and between animals and suggests that good segment combinations can be found in most subsets of 16 animals. If the 16 animals with the highest overall GEBV (group mean overall ASI = \$183.79) were chosen as founders, the potential ASI by combining their best segments

would be \$775.00. This is higher than the average ASI for random sets of 16 animals but lower than the maximum obtained from the genetic algorithm. Thus, the animals selected using the genetic algorithm were not necessarily the 16 with the highest overall ASI. The animals selected as founders contributed good segments to the ideal genotype at particular positions, despite having lower overall ASI.

Breeding Objective 1

The aim of objective 1 was to create a population with the 2 best segments identified in the founder animals segregating at each position in the genome. For this objective, the genotype-building strategies

Figure 3. Response under breeding objective 1 when aiming to obtain the “ideal genotype” using 3 selection algorithms (see Table 1): LOG (light gray), ARCSIN (dark gray/blue), and OPT (black). Shown is the mean overall Australian Selection Index genomic EBV (GEBV; A), mean coancestry (B) and distance to the ideal genotype (C) in the parents, and the standard deviation of GEBV in the selection candidates (D). Color version available in the online PDF.

(ARCSIN and LOG) were most efficient. These strategies obtained parental groups with segment frequencies close to the ideal genotype after 30 generations (Figure 3). The mean overall GEBV for ASI was \$880.02 (± 0.127 , SE) and \$881.59 (± 0.053) for ARCSIN and LOG, respectively. One target segment was lost in 3% of replicates for ARCSIN and 6% of replicates for LOG. Selection on overall GEBV using the OPT algorithm ($k_2 = 120$) achieved a lower overall ASI of \$837.02 (± 1.08) for approximately the same level of coancestry (ARCSIN: $0.584 \pm 1.15 \times 10^{-4}$, LOG: $0.586 \pm 1.80 \times 10^{-4}$, OPT: $0.585 \pm 1.63 \times 10^{-3}$). The OPT strategy retained only 68.1% of target segments at each position and, consequently, was much further from the ideal genotype than the genotype-building strategies.

The LOG algorithm achieved its selection limit more rapidly than ARCSIN (Figure 3A). The differences between the strategies was most evident in the GEBV SD change for selection candidates (Figure 3D), where both ARCSIN and LOG had a similar SD in generations 1 to 10, but the SD of LOG decreased more rapidly than that of ARCSIN during generations 15 to 22. During this period, LOG had a higher GEBV response compared with ARCSIN but, interestingly, it also showed a larger deviation from the target genotype. This suggests that, during generations 15 to 22, LOG permitted target segments to attain frequencies higher than their ideal. However, by generation 23, LOG was more efficient at obtaining frequencies closer to the ideal because segments were more abundant at each

position. In contrast, ARCSIN optimized the frequency of the target segments at all times and therefore other (undesirable) segments persisted in the population. This persistence reduces short-term overall GEBV response and slows the attainment of the ideal genotype. It appears that LOG was better suited as the selection criterion under the simulated conditions, even though ARCSIN was more closely aligned with the breeding objective. The OPT algorithm had a greater short-term response in mean overall GEBV, but it lost desirable segments as selection progressed and reached a lower selection limit compared with LOG and ARCSIN. The maximum rate of genetic gain for all selection strategies was approximately equal (i.e., equal slope for the response curves in Figure 3A), probably reflecting an equal selection intensity for all breeding programs.

Breeding Objective 2

The aim of the second breeding objective was to test the flexibility of the genotype-building and GEBV selection methods under lower coancestry ($f < 0.5$) than that imposed by the ideal genotype in breeding objective 1. This objective therefore required strategies to maintain more than 2 segments at each position in the genome. Final responses show a clear trade-off between GEBV response and coancestry for all algorithms (Figure 4). Thus, the highest response in overall ASI was achieved by LOG when coancestry was high ($f > 0.3$), but OPT performed better when coancestry was lower ($f < 0.3$).

First, consider the results for breeding objective 2 at low coancestry (Figure 4; $f < 0.3$). Here, the flexibility of OPT to select the best available segments seemed to provide an advantage over the genotype-building strategies of LOG and ARCSIN. For example, suppose that the genotype building strategies had 6 target segments at a position. It is very difficult for algorithms to maintain all these segments in the small population of parents. Thus, some desirable segments are inevitably lost. Upon losing segments, ARCSIN simply tries to obtain the remaining segments at the target frequency and accepts any other segments to “replace” the lost segments. In contrast, LOG aims to increase the frequency of all remaining targeted segments, with the relative intensity of selection on each segment determined by its desired frequency in the final population. This distinction, either optimizing (ARCSIN) or increasing (LOG) segment frequency, explains why LOG achieves a higher ASI than ARCSIN when coancestry is low. However, the tendency for LOG to increase the frequency of desirable segments also prevents it from reducing coancestry. In contrast, OPT accepts any seg-

ments and simply selects the best overall GEBV with a restriction on coancestry. The flexibility of OPT was superior to the precision of segment selection when the objective dictates a low coancestry.

When coancestry was not constrained, then the long-term advantage of genotype building was more apparent (Figure 4). The maximum response for the genotype-building strategies was about \$920. This limit was higher than that achieved by breeding objective 1 (~\$882) because coancestry was higher. The response using LOG was generally higher than that using ARCSIN because, as stated above, LOG increases the frequency of the target segments rather than optimizing frequencies as for ARCSIN. However, the differences between the 2 algorithms diminished as coancestry increases because fewer target segments are lost. At high coancestry, OPT achieved an overall mean GEBV for ASI of \$862.83. This is lower than ARCSIN and LOG because when coancestry was not constrained OPT favored homozygous individuals, resulting in the loss of good segments in early generations and the fixation of the remaining best segments at the limit.

DISCUSSION

In this study, we used a simplified model of the bovine genome, consisting of 25-cM segments with no internal recombination but free recombination between

Figure 4. Response curves for overall mean Australian Selection Index (A\$) and coancestry after 30 generations of selection under breeding objective 2 using 3 selection algorithms (see Table 1): LOG (light gray), ARCSIN (dark gray/blue), and OPT (black). Color version available in the online PDF.

segments, to evaluate long-term selection strategies for using high-density DNA markers. We found that genotype-building strategies (LOG and ARCSIN) were successful in achieving an “ideal genotype” and had a very large long-term selection response over 30 generations of intense selection. However, selection on overall GEBV while constraining coancestry (OPT) achieved greater short-term response and almost as much long-term gain. The OPT strategy also showed the greatest flexibility to precisely control coancestry while achieving a high selection response because it chose the best segments within its coancestry constraints. In contrast, genotype-building strategies found it difficult to control coancestry and maintain high ASI because they ignored nontarget segments even when some target segments were lost (due to drift). We conclude that in practice, given the coancestry constraints, OPT is a more flexible strategy for long-term selection response in dairy cattle.

We do not believe our simplifications led to qualitatively misleading results; in fact, they may give more realistic results than some alternative methods. In reality, recombination will not occur at exactly 25-cM intervals, but the effects of the actual segments could be estimated from SNP effects with similar accuracy to the 25-cM segment we used. Over many generations, recombination will yield smaller segments and generate new genetic variance by separating favorable and unfavorable alleles previously occurring on the same chromosome segment (e.g., Habier et al., 2009). However, the accuracy with which we could predict the breeding value of these new segments is uncertain. Therefore, our method of maintaining chromosome segments without recombination is conservative because internal recombination may generate additional genetic variance with uncertain properties. In practice, marker effects could be re-estimated to maintain accuracy but the properties of these estimates are unlikely to be substantially different from those we have assumed. Our method has the advantage of using SNP effects estimated from real cattle. An alternative would be to simulate QTL and SNP but then the results might depend on the exact properties of the simulation (e.g., Muir, 2007; Habier et al., 2009; Jannink, 2010).

The average rate of gain in the breeding program at generation 10 was about 0.72 SD per generation (σ/gen) for the OPT strategy (from Figure 3). This is lower than the expected gains under an efficient progeny-testing scheme (1.32 σ/gen) or a breeding scheme incorporating selection on overall GEBV (i.e., genomic selection, 0.99 σ/gen ; Pryce et al., 2010). However, the SD of the breeding values in selection candidates for our simulations was comparatively low. For example, in

generations 1 to 10, OPT captured about 10% of the genetic variance in the current population of Australian Holsteins (i.e., $22^2/69^2$, where 22 is the average SD of GEBV for the selection candidates), whereas Pryce et al. (2010) assumes this figure is much higher (60% of the genetic variance). The differences between Pryce et al. (2010) and the current results arise because of the reliability of GEBV and population size. Thus, Pryce et al. (2010) had a higher response to selection on GEBV compared with our results because many more animals were available for selection and reliability of GEBV was higher.

Figure 1 shows that chromosome segments at all positions across the genome showed variance in GEBV for ASI. This relatively even distribution is expected if a large number of genes influence the trait because each segment contains multiple QTL. Although the best genotype constructed from the segments present in the current population represents a large gain in ASI over the mean of today's population, the results from the genetic algorithm suggest that combining the best segments from almost any subset of 16 ancestors would achieve such an outcome. A response of magnitude many times the genetic SD in the current population is not unusual for long-term selection experiments (Laurie et al., 2004; Johansson et al., 2010) and might be expected when a large number of genes causes variation in a trait. This is because selection is spread across the whole genome and there is a slow accumulation of favorable alleles in selected animals.

Our results indicate that response to genomic selection may continue for many years as, for example, a consistent response was maintained for several generations with the OPT algorithm (Figure 3). This contrasts to the results of some simulation studies (Muir, 2007). The reason for this could be that if many QTL underlie the variation in ASI, then changes in allele frequency occur slowly and so genetic variance captured by markers will slowly decline. In effect, we modeled 101 multi-allelic QTL positions where the distribution of effects was approximately normal across all loci. The disadvantage of a large number of QTL is that simple pyramiding or genotype-building strategy to breed a line of cattle carrying all the desirable segments at all positions is not possible in a small number of generations. However, an advantage is that the next-best segment is only slightly inferior to the most desirable segment. A relatively even distribution of segment effects across the genome is a significant factor leading to the advantages of the OPT strategy in our simulations.

Past simulations for long-term selection response with major loci (i.e., with effect of 0.5 phenotypic SD) suggest a reduced long-term selection response com-

pared with an approach using phenotypic information and EBV (Gibson, 1994). Results in this paper contrast with that finding because the long-term response was maximized by targeting an ideal genotype but the advantage compared with selection on overall GEBV was relatively small. These differences probably occur because loci in our study had a normal distribution of effects. Therefore, it was possible for OPT to achieve a high GEBV without specifically combining the best segments from the ideal genotype. A genotype-building strategy might be more applicable if there were few target loci at rarer frequencies rather than loci across the entire genome. However, several authors have proposed other optimal methods for selection based on relatively few loci (e.g., Dekkers and van Arendonk, 1998; Sánchez et al., 2006).

We compared 2 methods of maximizing long-term response from dense genomic markers using the ARCSIN and LOG algorithms. Although Goddard (2009) describes analytically a method for finding the optimum long-term response, we found this difficult to implement in a convenient algorithm. The method of Goddard (2009) was modified to account for the existence of a possible 32 alleles per locus instead of 2 and aims to minimize the time required to reach target allele frequencies. A feature of the LOG and ARCSIN methods is that they apply additional selection pressure to favorable segments that were initially at low frequency. However, the simulated conditions may not have favored the algorithms because segments started with a minimum frequency of 1/32. In situations where some favorable alleles start at moderate or high frequency and some start at low frequency, strategies that place additional selection pressure on rare favorable segments might be more beneficial than the simulations reported here. The genotype-building algorithm could be improved to reduce the effect of lost loci (through genetic drift) either by breeding program structure or by reevaluating the target genotype when loci are lost. The LOG and ARCSIN methods contrast with the optimal methods to select on major loci (Dekkers and van Arendonk, 1998), as previous methods account for the variance contributed by the loci and therefore consider both its frequency and effect.

The similarity of outcomes between selection on overall GEBV and the genotype-building strategy parallels the equivalence of genomic selection based on estimates of SNP alleles and that based on relationships estimated from the SNP genotypes (Goddard, 2009). Thus, a method that controls coancestry, such as OPT, indirectly controls allele frequencies. Several authors have demonstrated this by using approaches similar to OPT. For example, Sonesson et al. (2010) found that the increase in allele frequencies is slower near favor-

able loci when coancestry penalties were from genomic compared with pedigree relationships. Similarly, Pryce et al. (2012) demonstrates that penalties on genomic coancestries are more effective for reducing coancestry than average (pedigree-estimated) relationships.

Our simulations assumed an unrealistically high reproductive rate of 200 offspring per cow. This is unlikely in practice because Ax et al. (2005), for example, reported fewer than 10 viable embryos per superovulation. If we had assumed a lower reproductive rate, similar results would have been obtained but after a greater number of generations. However, the results illustrate a potential synergy between genomic selection and reproductive technology, resulting in high rates of genetic improvement. In the future, it may be possible to obtain much higher reproductive rates in cattle. If larger numbers of embryos could be produced, early genotyping could allow only the best to be implanted. Developments in molecular techniques such as single-cell genotyping (Agrawala et al., 1992; Fan et al., 2011) may be beneficial for further improving these methods. Ideally, dense genomic markers would be combined with in vitro reproductive techniques (for example, “whizzo” genetics; Haley and Visscher, 1998), but this has significant technological barriers before implementation in cattle. A genotype-building strategy would be best suited for a situation in which large numbers of embryos could be produced, screened, and propagated in vitro. In the immediate future, genotype building might be more applicable in situations where there are higher reproductive rates and tolerance of coancestry, such as in field crops or for aquaculture species.

CONCLUSIONS

In summary, we used real genotypes and marker effects to investigate strategies to maximize the long-term selection response for a complex trait in dairy cattle. We investigated methods to create an “ideal genotype” and those selecting on overall GEBV. Results indicate that an ideal genotype is difficult to achieve, even under simulation conditions, because of the large number of genes and target loci affecting complex traits. We conclude that selection on overall GEBV with a constraint on coancestry is the most flexible selection strategy, particularly for dairy cattle with low reproductive rates and low tolerance of coancestry. This work highlights the benefits of genome-wide markers to distinguish between closely related individuals and the potential for genetic improvement with only a limited number of founders for the population. High reproductive efficiencies are required to implement an intensive breeding program as presented in this paper.

ACKNOWLEDGMENTS

The authors thank the Australian Dairy Herd Improvement Scheme (Melbourne, Australia) for the provision of genotypes and marker effects for this work. This research was supported by the Dairy Futures CRC (Melbourne, Australia) and the Australian Research Council's Discovery Projects funding scheme (project DP1093502). We acknowledge Michel Georges (University of Liège, Liège, Belgium) for discussions initiating this research.

REFERENCES

- ADHIS. 2011. Ranges and means for bull ABVs, April 2011. Vol. 2011. Australian Dairy Herd Improvement Scheme, Melbourne, Australia.
- Agrawala, P. L., V. A. Wagner, and H. Geldermann. 1992. Sex determination and milk protein genotyping of preimplantation stage bovine embryos using multiplex PCR. *Theriogenology* 38:969–978.
- Ax, R. L., S. Armbrust, R. Tappan, G. Gilbert, J. N. Oyarzo, M. E. Bellin, D. Selner, and T. C. McCauley. 2005. Superovulation and embryo recovery from peripubertal Holstein heifers. *Anim. Reprod. Sci.* 85:71–80.
- Browning, S. R., and B. L. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81:1084–1097.
- Cole, J. B., P. M. van Raden, J. R. O'Connell, C. P. van Tassell, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and G. R. Wiggins. 2009. Distribution and location of genetic effects for dairy traits. *J. Dairy Sci.* 92:2931–2946.
- Dekkers, J. C. M., and F. Hospital. 2002. The use of molecular genetics in the improvement of agricultural populations. *Nat. Rev. Genet.* 3:22–32.
- Dekkers, J. C. M., and J. A. M. van Arendonk. 1998. Optimizing selection for quantitative traits with information on an identified locus in outbred populations. *Genet. Res.* 71:257–275.
- Fan, H. C., J. B. Wang, A. Potanina, and S. R. Quake. 2011. Whole-genome molecular haplotyping of single cells. *Nat. Biotechnol.* 29:51–57.
- Gibson, J. P. 1994. Short-term gains at the expense of long-term response with selection of identified loci. Pages 201–204 in *Proc. 5th World Congr. Genet. Appl. Livest. Prod.*, Guelph, ON, Canada.
- Goddard, M. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica* 136:245–257.
- Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2009. Genomic selection using low-density marker panels. *Genetics* 182:343–353.
- Haley, C. S., and P. M. Visscher. 1998. Strategies to utilize marker-quantitative trait loci associations. *J. Dairy Sci.* 81(Suppl. 2):85–97.
- Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and M. E. Goddard. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: Coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genet.* 6:e1001139.
- Hill, W. G. 2000. Maintenance of quantitative genetic variation in animal breeding programmes. *Livest. Prod. Sci.* 63:99–109.
- Hospital, F. 2003. Marker-assisted breeding. Pages 30–56 in *Plant Molecular Breeding*. H. J. Newbury, ed. Blackwell Scientific Publishers, London, UK.
- Jannink, J.-L. 2010. Dynamics of long-term genomic selection. *Genet. Sel. Evol.* 42:35.
- Johansson, A. M., M. E. Pettersson, P. B. Siegel, and Ö. Carlborg. 2010. Genome-wide effects of long-term divergent selection. *PLoS Genet.* 6:e1001188.
- Laurie, C. C., S. D. Chasalow, J. R. LeDeaux, R. McCarroll, D. Bush, B. Hauge, C. Q. Lai, D. Clark, T. R. Rocheford, and J. W. Dudley. 2004. The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics* 168:2141–2155.
- Meuwissen, T. H. 1997. Maximizing the response of selection with a predefined 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.
- Muir, W. M. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J. Anim. Breed. Genet.* 124:342–355.
- Nejati-Javaremi, A., C. Smith, and J. P. Gibson. 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. *J. Anim. Sci.* 75:1738–1745.
- Pryce, J. E., M. E. Goddard, H. W. Raadsma, and B. J. Hayes. 2010. Deterministic models of breeding scheme designs that incorporate genomic selection. *J. Dairy Sci.* 93:5455–5466.
- Pryce, J. E., B. J. Hayes, and M. E. Goddard. 2012. Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. *J. Dairy Sci.* 95:377–388.
- Sánchez, L., A. Caballero, and E. Santiago. 2006. Palliating the impact of fixation of a major gene on the genetic variation of artificially selected polygenes. *Genet. Res.* 88:105–118.
- 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 no. 892 in *Proc. 9th World Congr. Genet. Appl. Livest. Prod.*, Leipzig, Germany.
- Thompson, J. R., R. W. Everett, and N. L. Hammerschmidt. 2000. Effects of inbreeding on production and survival in Holsteins. *J. Dairy Sci.* 83:1856–1864.
- VanRaden, P. M., K. M. Olson, D. J. Null, and J. L. Hutchison. 2011. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J. Dairy Sci.* 94:6153–6161.
- Von Krosigk, C. M., and J. L. Lush. 1958. Effect of inbreeding on production in Holsteins. *J. Dairy Sci.* 41:105–113.
- Wray, N., and M. Goddard. 1994. Increasing long-term response to selection. *Genet. Sel. Evol.* 26:431–451.
- Zamir, D. 2001. Improving plant breeding with exotic genetic libraries. *Nat. Rev. Genet.* 2:983–989.
- Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. van Tasse, T. S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, and S. L. Salzberg. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10:R42.