



ORIGINAL ARTICLE

Use of haplotypes to estimate Mendelian sampling effects and selection limits

J.B. Cole & P.M. VanRaden

Animal Improvement Programs Laboratory, ARS, USDA, Beltsville, MD, USA

Keywords

genetic gain; haplotypes; Mendelian sampling; selection limits.

Correspondence

J.B. Cole, Animal Improvement Programs Laboratory, ARS, USDA, Room 306, Bldg 005, BARC-West, 10300 Baltimore Avenue, Beltsville, MD 20705-2350, USA. Tel: 301-504-8334; Fax: 301-504-8092; E-mail: john.cole@ars.usda.gov

Received: 22 September 2010;
accepted: 13 February 2011

Summary

Limits to selection and Mendelian sampling (MS) terms can be calculated using haplotypes by summing the individual additive effects on each chromosome. Haplotypes were imputed for 43 382 single-nucleotide polymorphisms (SNP) in 1455 Brown Swiss, 40 351 Holstein and 4064 Jersey bulls and cows using the Fortran program findhap.f90, which combines population and pedigree haplotyping methods. Lower and upper bounds of MS variance were calculated for daughter pregnancy rate (a measure of fertility), milk yield, lifetime net merit (a measure of profitability) and protein yield assuming either no or complete linkage among SNP on the same chromosome. Calculated selection limits were greater than the largest direct genomic values observed in all breeds studied. The best chromosomal genotypes generally consisted of two copies of the same haplotype even after adjustment for inbreeding. Selection of animals rather than chromosomes may result in slower progress, but limits may be the same because most chromosomes will become homozygous with either strategy. Selection on functions of MS could be used to change variances in later generations.

Introduction

Mendelian sampling (MS) variance is generated by the process of randomly sampling parental chromosomes during meiotic division in gametogenesis and is commonly estimated from the difference between an individual's predicted transmitting ability (PTA) and its parent average (PA, the average of the sire and dam PTA). Individual PTA does not provide any information about the MS term for individual gametes or parents, and the within-family variance is not affected by selection (Bulmer 1971). However, genotypic information can provide early estimates of MS effects by allowing direct inspection of markers at the chromosomal level (Dekkers & Dentine 1991). Woolliams *et al.* (1999) showed that sustained genetic gain under selection depends on MS variance, and the increase in reliability of PTA observed

in genomic selection programmes is because of more precise estimation of MS effects (Hayes *et al.* 2009). Better estimates of MS also permit increased rates of genetic gain with lower increases in inbreeding than in traditional breeding programmes (Daetwyler *et al.* 2007).

Substantial benefits are not realized from genomic selection until there is a large enough pool of genotyped animals to provide accurate estimates of marker effects, which are essential for reliable prediction of MS terms. Marker-assisted selection (MAS) programmes have increased short-term selection response because the markers explain a portion of MS variance (Meuwissen & Van Arendonk 1992; Meuwissen & Goddard 1996), but in the long term, MAS results in decreased MS because the paternal and maternal genotypes become more similar as allele frequencies for the QTL near fixation when it

is assumed that populations are closed and there is no mutation.

The objective of this paper is to describe the MS variance present in the US Brown Swiss (BS), Holstein (HO), and Jersey (JE) populations using dense single-nucleotide polymorphism (SNP) genotypes, as well as to discuss selection limits based on haplotypes present in the genotyped population. Four traits representing a range of heritabilities and average reliabilities are included in the analysis.

Material and methods

Genotypes

Genotypes for 43 382 SNP in 1455 BS, 40 351 HO and 4064 JE bulls and cows were obtained using the Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA). Marker solutions from the June 2010 US genomic evaluation were used to calculate MS variance and selection limits for daughter pregnancy rate (DPR; a measure of female fertility) (VanRaden *et al.* 2004), milk yield, lifetime net merit (NM\$; a measure of lifetime profitability) (Cole *et al.* 2010) and protein yield. Haplotypes were imputed with the Fortran program findhap.f90 (VanRaden *et al.* 2011), which combines population and pedigree haplotyping methods. Calculations were performed with SAS 9.2 (SAS Institute Inc., Cary, NC, USA), and plots were produced with R 2.10.1 (R Development Core Team, 2010) and ggplot2 0.8.7 (Wickham 2009) on a workstation running 64-bit Red Hat Enterprise Linux 5 (Red Hat Inc., Raleigh, NC, USA).

Mendelian sampling variances

Estimated MS terms were computed for each trait assuming that loci on the same chromosome were in perfect linkage (MS_C), or that all loci in the genome were unlinked (MS_U), as:

$$MS_C = \sum_{c=1}^{30} \left(\sum_{m=1}^{n_c} s_m \alpha_m - \sum_{m=1}^{n_c} d_m \alpha_m \right)^2$$

and

$$MS_U = \sum_{m=1}^{43382} (s_m \alpha_m - d_m \alpha_m)^2$$

respectively, where m denotes a marker, s and d are the haplotypes for the m th marker inherited from the animal's sire and dam, respectively, α_m is the

estimated allele substitution effect for the m th marker, c is the c th chromosome, and n_c is the number of markers present on the c th chromosome. Marker effects were calculated using a Bayes A model as described in Cole *et al.* (2009). Calculations included markers from the pseudoautosomal region of the X chromosome, which contribute to MS, but not those located only on the X chromosome. For the purposes of comparison, expected MS was computed as half of the additive genetic variance (V_a) and inbreeding was ignored. It was assumed that there were no dominance or epistasis effects.

Allele substitution effects were estimated using an infinitesimal alleles model with a heavy-tailed prior (also known as a Bayes A model) in which smaller effects are regressed further towards 0 and markers with larger effects are regressed less to account for a non-normal prior distribution of marker effects (VanRaden 2007, 2008). Marker effects were randomly distributed with a heavy-tailed distribution generated by dividing a normal variable by $h^{|s-2|}$, where h determines departure from normality and s is the size of the estimated marker effect in standard deviations (VanRaden 2008). Marker effects are normally distributed with no additional weight in the tails when h is 1, and variance in the tails grows with increasing values of h ; a parameter of 1.12 is used in this study (Cole *et al.* 2009). Variances of estimated MS and marker effects are less than true effects in the same way that PTA has less variance than true transmitting abilities.

Selection limits

Marker values were summed for each genotyped animal to obtain chromosomal estimated breeding values (CEBV) for lifetime net merit, and the CEBV were summed to obtain the direct genomic values (DGV). Genomic estimated breeding values (GEBV), which include base adjustments, polygenic effects and information from non-genotyped relatives, were taken from the June 2010 genetic evaluation run. Empirical selection limits were calculated by combining the haplotypes with the best unadjusted or adjusted CEBV for DPR, milk, NM\$ and protein yield. These estimated limits represent progress that could be achieved with the current data. In the future, with more data and larger reference populations, true limits would be larger with more accurate SNP and haplotype estimates.

Lower bounds of selection limits (SL_C) were predicted by selecting the 30 best haplotypes for each trait, and upper bounds (SL_U) were calculated by

taking the allele at each marker locus with the most desirable value, as:

$$SL_C = \sum_{c=1}^{30} \max_H \left(\sum_{m=1}^{n_c} l_m \alpha_m \right)$$

and

$$SL_U = \sum_{m=1}^{43\,382} \max_L (l_m \alpha_m),$$

respectively, where c indicates a chromosome, m denotes a marker, α_m is the estimated allele substitution effect for the m th marker, H represents the set of all unique haplotypes in the genotyped population, n_c is the number of markers present on the c th chromosome, h_m represents the m th marker of an individual haplotype, L is the set of all marker loci in the genotyped population, and l_m represents the genotype of the m th marker locus.

The CEBV for NM\$ also were adjusted for inbreeding by subtracting 6% of an additive genetic standard deviation (\$11.88) per 1% increase in homozygosity above the breed average (Smith *et al.* 1998). Animals with above-average heterozygosity were credited in the same manner. Adjusted and unadjusted values were compared to determine the impact of such adjustments on GEBV. Homozygosity averaged 0.70 ± 0.01 in BS, 0.67 ± 0.01 in HO and 0.72 ± 0.02 in JE and was calculated as the average marker homozygosity of each pair of chromosomes in the genotyped animals.

Results

Mendelian sampling

Lower- and upper-bound estimates of MS are provided by MS_U and MS_C , respectively. In theory, the true MS variance should be calculated using individual linkage disequilibrium (LD) blocks or map distances rather than assuming that all markers on the same chromosome are a single linkage group, and MS_C may be overestimating the true variance. In a completely inbred population, all genotypes would be homozygous, and MS_U and MS_C both would be 0. In a heterozygous population in which all marker frequencies are 0.5, $MS_U \leq MS_C$, and both are proportional to the true MS variance.

The α_i used to compute MS_C and MS_U are estimates of marker effects rather than true marker effects and are therefore regressed towards the population mean. As a result, the calculated bounds on MS variance underestimate the true MS variance in

the population. New genotypes are continuously being collected, and the accuracy of the SNP effects will increase as the reference population used to calculate those effects increases in size. MS_C and MS_U are expected to increase asymptotically towards the true MS variance as the correlation between the true and predicted SNP effect approaches 1.

The SNP used for genotyping were selected to have high average minor allele frequencies, and most predicted allele substitution effects were near 0. If all loci are unlinked, then selection for a desirable allele has no effect on the frequency of other alleles, the frequency of other alleles does not change in response to selection, and the population average, which depends on allele frequency, remains close to 0. When loci are linked, however, selection for markers with positive effects generates LD blocks in which the sum of effects is >0 . Therefore, we expect that the sums of squared differences between chromosome haplotypes will be larger than the sum of squared differences between individual alleles, which was confirmed for all breeds and traits (Table 1). The range was largest for HO for all traits, reflecting the greater number of observed haplotypes in that breed than BS or JE. Results were generally similar for BS and JE, although in some cases, there was slightly more variation in JE than in BS. Ratios of MS_C to MS_U were generally smaller for HO and larger for BS and JE, ranging from 4.0 for JE milk to 17.4 for BS DPR. These results may reflect more

Table 1 Predicted upper and lower bounds and expectations of Mendelian sampling variance for daughter pregnancy rate (DPR), milk yield, lifetime net merit (NM\$) and protein yield for US Brown Swiss (BS), Holstein (HO) and Jersey (JE) cattle

Trait	Breed	Mendelian sampling variance		
		Lower bound	Expected ^{a,b}	Upper bound
DPR (%)	BS	0.09	1.45	1.57
	HO	0.57	1.45	4.02
	JE	0.09	0.98	1.27
Milk yield (kg)	BS	7264	44 238	104 255
	HO	46 879	53 736	219 939
	JE	30 855	42 238	123 813
NM\$ (USD)	BS	2539	19 602	40 458
	HO	16 601	19 602	87 449
	JE	3978	19 602	44 552
Protein yield (kg)	BS	6.40	37.29	91.11
	HO	35.95	37.29	145.25
	JE	10.33	33.47	92.35

^aExpected Mendelian sampling variances were calculated as $\frac{1}{2}V_a$ assuming no inbreeding.

^bThe same additive genetic variance is used for all breeds for NM\$.

precise estimation of MS variances for HO than BS or JE.

Expected MS variance was calculated for each breed and trait (assuming no inbreeding) as $\frac{1}{2}V_a$, and all estimates were bounded by MS_U and MS_C , as expected. This provides confirmation that MS_U and MS_C provide plausible estimates of MS variance. The expected HO variances were much closer to the lower bounds than those of BS and JE, which reflects the much larger number of HO haplotypes that have been sampled. As a greater number and more diverse groups of BS and JE animals are genotyped, the expected MS variances should increase. While the inbreeding of parents was not accounted for, relationships among mates would have needed to be very large to result in substantial reductions in estimated variances, and those kinds of close matings generally are avoided.

Bulmer (1971) showed that within-family variance should decrease as homozygosity increases, and it is well known that inbreeding levels have increased in dairy cattle over time (Young & Seykora 1996). Figures 1, 2 and 3 show the change in MS_C of NM\$ for genotyped BS, HO and JE cattle, respectively, born between 1990 and 2010 and representing approximately four generations of selection. Slopes were slightly negative for all breeds, and a decrease in MS variance was expected in all breeds based on the increased levels of pedigree inbreeding over that time (Figure 4), but only the HO slope differed from 0 ($p < 0.05$). The HO trend may reflect high statistical power because of a large sample size rather than a biologically meaningful decrease in

variance. These results suggest that while inbreeding in the population has increased over time, inbred matings have not been used to produce the genetically elite animals with genotypes in this study, or levels of inbreeding have not increased enough to result in a substantial loss of haplotypes. Changes over time may have been different for grade cows.

Correlation among genomic (F_G) and pedigree (F_P) inbreeding, MS_C and MS_U were calculated for each trait to confirm that MS decreases with inbreeding, which should result in a strong, negative correlation (Table 2). For DPR, correlations of F_G with MS_U ranged from -0.73 to -0.83 , and F_P with MS_U ranged from -0.38 to -0.53 . Pedigree inbreeding was expected to have lower correlations with MS than F_G because the incidence of pedigree errors has been shown to be approximately 10% in US Holsteins (Banos *et al.* 2001). However, correlations of F_G and F_P with MS_C were consistently near 0 across breeds and traits. This is probably because MS_C was calculated assuming that markers on the same chromosome were in perfect linkage, and the impact of a small number of loci becoming homozygous is small when blocks rather than individual alleles are selected. The observed range of genomic inbreeding was small, and there were no extremely inbred animals, in which you would expect to see whole LD blocks fixed, which also may contribute to the low correlations.

The correlations among MS_U for milk with inbreeding were near 0 for HO and JE, which was unexpected, as was the correlations of MS_U with F_G and F_P for HO NM\$. Holstein and JE differ from BS

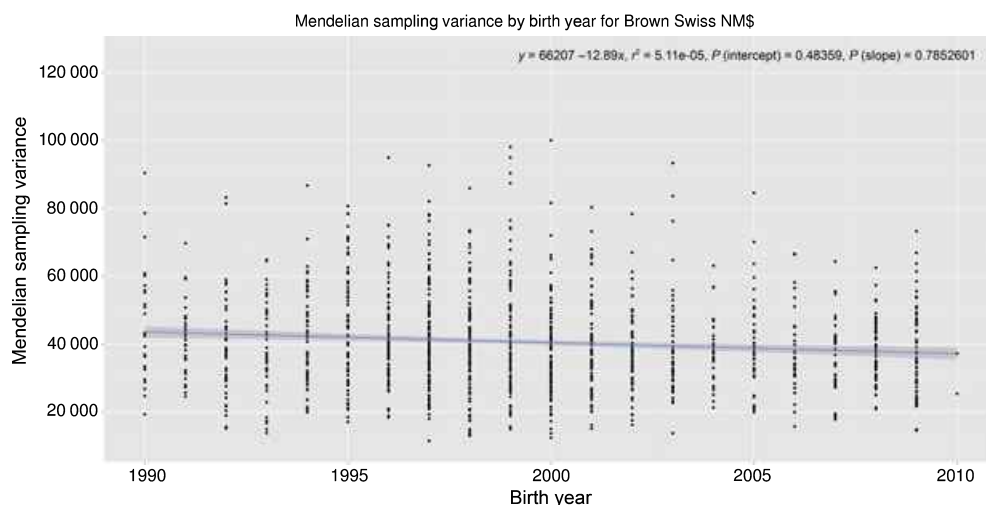


Figure 1 Changes in Mendelian sampling variance (upper bound) for lifetime net merit (NM\$) in US Brown Swiss cattle born between 1990 and 2010.

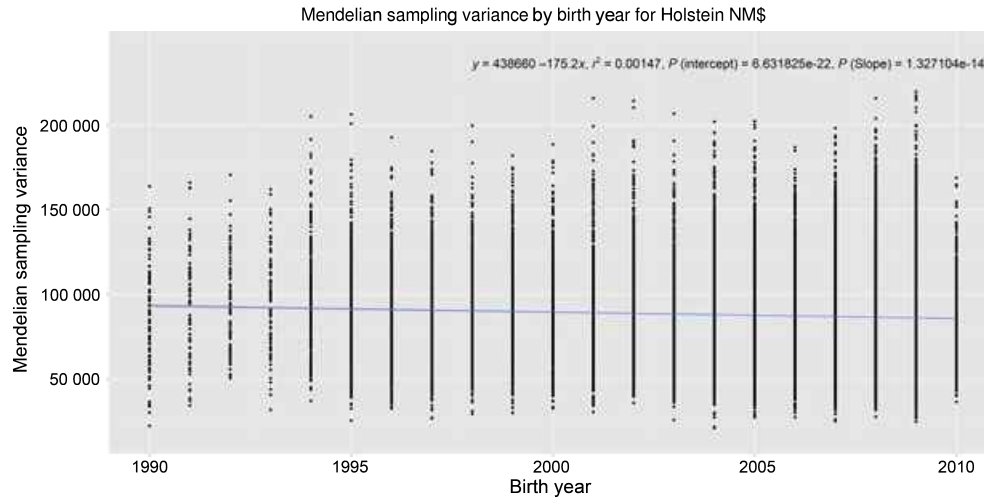


Figure 2 Changes in Mendelian sampling variance (upper bound) for lifetime net merit (NM\$) in US Holstein cattle born between 1990 and 2010.

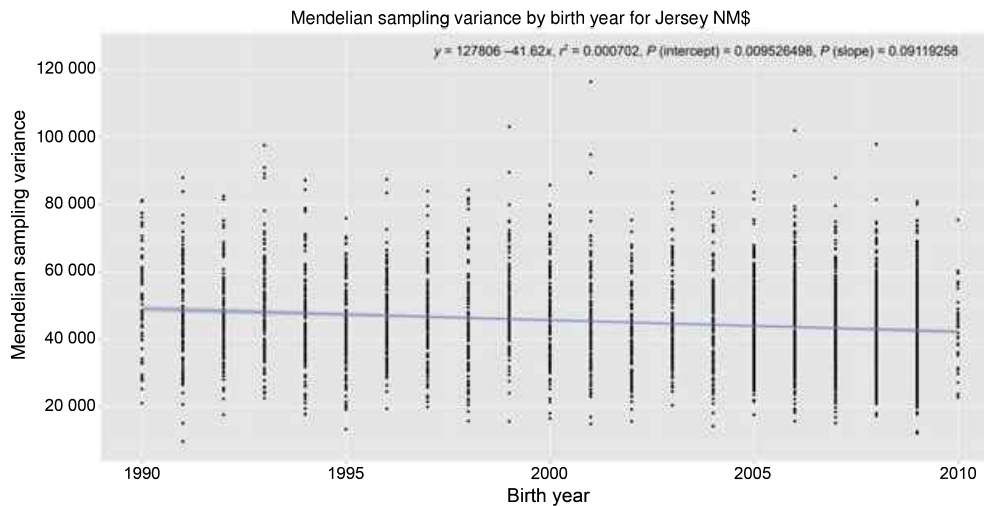


Figure 3 Changes in Mendelian sampling variance (upper bound) for lifetime net merit (NM\$) in US Jersey cattle born between 1990 and 2010.

in that the *DGATI* locus is not segregating in the latter population. Similarly, in addition to *DGATI*, there is a large QTL for NM\$ segregating on *Bos taurus* autosome 18 in HO (Cole *et al.* 2009). Individual QTL can have a large effect on the sampling variance but no effect on inbreeding because fixation at single locus has only a small effect on homozygosity. Note that in JE, in which there are no QTL for NM\$ segregating, the correlation of MS_U with inbreeding is similar to that of BS. Results for MS_U confirm that as inbreeding increases, sampling variance decreases.

Correlations of GEBV for NM\$ with MS_U and MS_C were calculated to determine whether animals with high GEBV also had greater MS variances. The GEBV were negatively correlated with MS_U and MS_C

in all breeds, ranging from -0.04 to -0.14 . This suggests that efforts to reduce the rate of the increase in inbreeding have been successful, although the animals with the most desirable GEBV still are more inbred than average animals.

Selection limits

Selection limits for the current population were estimated assuming that either whole chromosome haplotypes or individual alleles can be selected and combined at will to produce whole genomes, as described in Cole & VanRaden (2010). Lower and upper bounds for each trait, as well as the largest DGV observed in the genotyped population, are presented

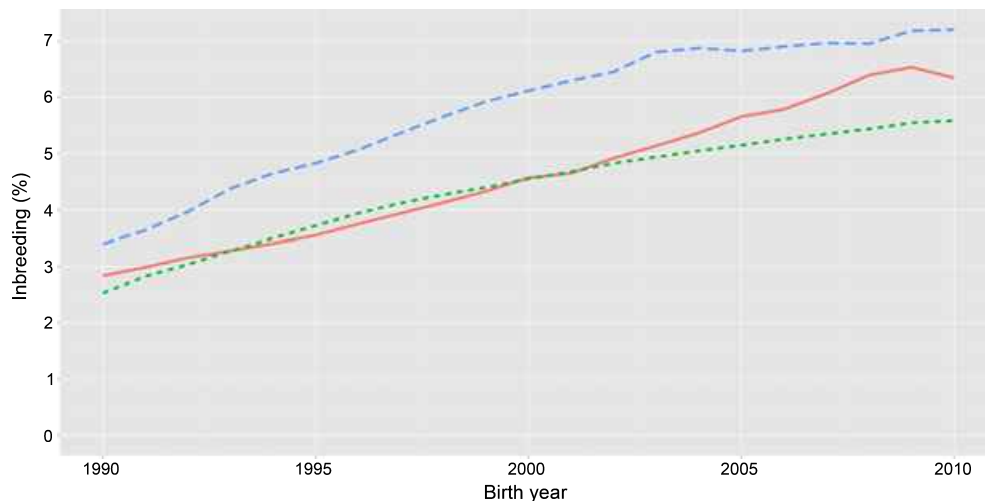


Figure 4 Changes in average inbreeding (%) between 1990 and 2010 for US Brown Swiss (solid line), Holstein (short-dashed line) and Jersey (long-dashed line) cattle.

Table 2 Correlations of lower and upper bounds of Mendelian sampling variance with genomic and pedigree inbreeding (F) for daughter pregnancy rate (DPR), milk yield, lifetime net merit (NM\$) and protein yield for US Brown Swiss (BS), Holstein (HO) and Jersey (JE) cattle

Trait	Breed	Lower bound		Upper bound	
		Genomic F	Pedigree F	Genomic F	Pedigree F
DPR (%)	BS	-0.73	-0.38	-0.02 ^a	0.09
	HO	-0.77	-0.40	-0.11	-0.03
	JE	-0.83	-0.53	-0.01 ^a	0.06
Milk yield (kg)	BS	-0.86	-0.55	-0.05	0.03 ^a
	HO	-0.12	-0.05	-0.10	-0.03
	JE	-0.01 ^a	0.03 ^a	-0.04	0.04
NM\$ (USD)	BS	-0.85	-0.49	0.03 ^a	0.13
	HO	-0.21	-0.12	-0.11	-0.03
	JE	-0.86	-0.53	-0.11	-0.02 ^a
Protein yield (kg)	BS	-0.86	-0.54	-0.06	0.00 ^a
	HO	-0.84	-0.47	-0.15	-0.08
	JE	-0.82	-0.54	-0.06	0.01 ^a

^aNot different from 0 (p > 0.05).

in Table 3. The lower bounds represent selection limits attainable by selection among haplotypes already in the population, while the upper bounds are limits attainable under the assumption that complete haplotypes can be constructed from individual alleles in the population. In all cases, SL_C and SL_U were largest for HO, reflecting the larger number of HO genotypes represented in the analysis. Limits were generally similar for BS and JE across traits.

Lifetime net merit

Lower selection limits for NM\$ with no adjustment for inbreeding were \$3857 (BS), \$7515 (HO) and \$4678 (JE). Adjusted values were slightly smaller and were \$3817 (BS), \$7494 (HO) and \$4606 (JE). Upper

bounds had values of \$9140 (BS), \$23 588 (HO) and \$11517 (JE) and were not adjusted for inbreeding because they were calculated from individual loci rather than complete haplotypes. The largest DGV among all genotyped animals in each breed were \$1102 (BS), \$2528 (HO) and \$1556 (JE). The top active bulls (AI and foreign bulls with semen distributed in the US that are in or above the 80th percentile, based on NM\$) in each breed following the August 2010 genetic evaluation had GEBV for NM\$ of +\$1094 (BS: 054BS00374), +\$1588 (HO: 001HO08784) and +\$1292 (JE: 236JE00146). Because DGV and GEBV include different information, and no reliability restriction was imposed, they are not directly comparable, but all DGV and GEBV were well below SL_C .

Table 3 Predicted upper and lower bounds of selection limits and largest observed direct genomic values (DGV) for daughter pregnancy rate (DPR), milk yield, lifetime net merit (NM\$) and protein yield for US Brown Swiss (BS), Holstein (HO) and Jersey (JE) cattle

Trait	Breed	Lower bound	Upper bound	Largest DGV ^a
DPR (%)	BS	20	53	8
	HO	40	139	8
	JE	19	53	5
Milk yield (kg)	BS	6461	15 465	2065
	HO	11 310	35 419	3634
	JE	7333	18 295	2554
NM\$ (USD)	BS	3857	9140	1102
	HO	7515	23 588	2528
	JE	4678	11 517	1556
Protein yield (kg)	BS	180	470	61
	HO	312	1138	114
	JE	218	568	79

^aDirect genomic values were calculated by summing the marker effects for each genotyped animal.

If two copies of each of the 30 best haplotypes in the US Holstein population were combined in a single animal (SL_C for NM\$), it would have a GEBV for NM\$ of +\$7515 (Figure 5), approximately five times larger than that of the current best Holstein bull in the US, whose GEBV for NM\$ are +1588. Cole & VanRaden (2010) presented a similar result based on CEBV that were averages of the actual parental haplotypes. When actual haplotypes are used rather than averages of haplotypes, there is an increase in SL_C of approximately 20%.

Correlations among the unadjusted and adjusted DGV ranged from 0.997 to 0.999 in BS and JE, and

all were >0.999 in HO. The best genotype after adjusting for inbreeding consisted of two copies of the same haplotype for 26 chromosomes in BS and HO and 22 in JE, although the differences between the first- and second-ranked haplotypes were usually very small (<\$10). Top unadjusted haplotype values ranged from \$82 for BTA 18 to \$192 for BTA 2 in BS, from \$71 for BTA 24 to \$309 for BTA 5 in JE and from \$143 for BTA 26 to \$375 for BTA 14 in HO. These values may seem large, but each of the top haplotypes was from a different animal in all three breeds. Differences between the best and poorest unadjusted haplotypes of a chromosome ranged from \$136 for BTA 26 to \$338 for BTA 1 in BS, from \$147 for BTA 24 to \$475 for BTA 5 in JE and from \$269 for BTA 26 to \$713 for BTA 14 in HO. The differences are larger for HO than BS and JE because many more haplotypes have been measured in that breed, and consequently, more haplotypes from each tail of the distribution have been identified. Results were similar for adjusted haplotypes, but the values were slightly smaller.

Daughter pregnancy rate, milk yield and protein yield

While individual values varied across traits, results for DPR, milk and protein yield were similar to those for NM\$ (Table 3). Selection limits were estimated to be lowest for BS, intermediate for JE and largest for HO, again reflecting differences in the number of genotyped animals in each breed. Direct genomic values were similar for BS and JE and larger for HO. The DGV and GEBV for all traits were well below SL_C , as was the case with NM\$.

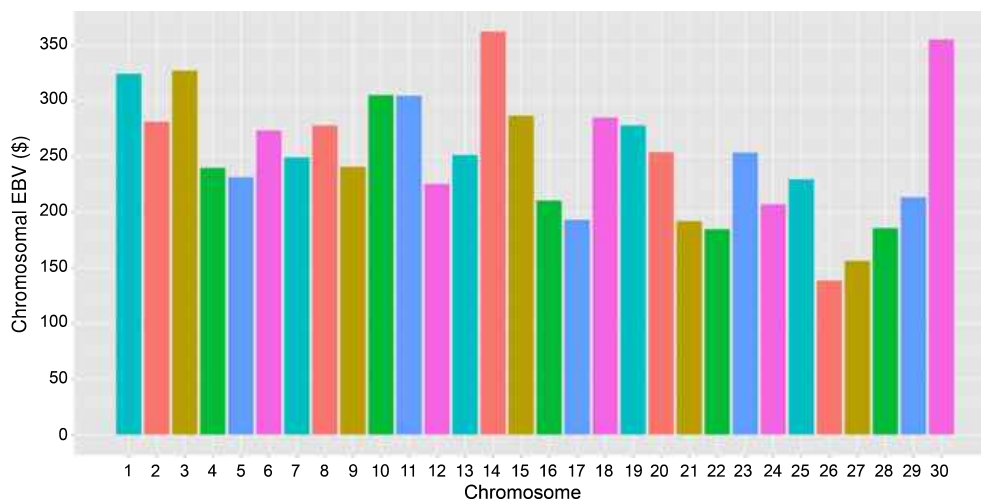


Figure 5 Chromosomal estimated breeding values (EBV) of lifetime net merit (NM\$) for a hypothetical animal whose genotype consists of two copies of each of the best haplotypes in the current US Holstein population. The sum of the individual chromosome effects is \$7515.

The top active bulls in each breed for DPR (%) had GEBV of +3.6 (BS: 001BS00553), +7.2 (HO: 001HO06360) and +4.6 (JE: 200JE00990), which are much smaller than the predicted lower selection limits. The upper limits are probably substantial overestimates of GEBV attainable in practice, particularly for a lowly heritable trait, but do show that DPR could be improved considerably if its economic value increases to the point that more weight in selection indices is warranted.

Selection for increased milk yield over the past 40 years was very successful (Hansen 2000), although milk volume has not received direct weight in the NM\$ index since 2003. The top active bulls in each breed had GEBV of +1451 (BS: 054BS00456), +2306 (HO: 014HO03831) and +1718 (JE: 001JE00604). In all cases, the GEBV were much smaller than the upper and lower bounds on the selection limit for each breed (Table 3). Despite the strong emphasis placed on milk yield in the past, it does not appear that the population is approaching a selection limit, and given the increasing emphasis on non-yield traits in NM\$ and breed association indices, it is possible that progress towards the limits will slow dramatically.

Protein yield now receives 16% of the emphasis in the 2010 revision of NM\$ (Cole *et al.* 2010) and also is an important selection objective in other countries (Miglior *et al.* 2005). The top active bulls in each breed had GEBV (kg) of +44 (BS: 054BS00456), +64 (HO: 014HO04929) and +40 (JE: 029JE03487). As was the case for DPR, milk yield and NM\$, the GEBV for the top animals in each breed are not near the selection limits. The increased weight placed on protein in NM\$ will result in faster rates of gain, but many generations of more intensive selection will be needed before the most extreme animals in the population near the selection limit.

Discussion

The objective of this study was to use genotypes from US BS, HO and JE cattle to estimate MS variances and predict selection limits for fertility, yield and economic merit. Lower and upper bounds for MS variance were calculated assuming either complete or no linkage among loci on the same chromosome. It is possible that those estimates are biased because of the shrinkage of the allele effect estimates in the genomic prediction model, and Goddard *et al.* (2009) provide an excellent discussion of sources of bias in genomic evaluation models and the magnitude of their importance. However, in all cases, the

expected MS variance calculated from population data falls between those upper and lower bounds, so the magnitude of any bias in the estimators likely is small and should not substantially affect our results.

Selection limits were calculated using the allele substitution effects and marker frequencies observed in the current BS, HO and JE populations in a manner that implies that those limits could be reached in one round of selection. That is useful to obtain initial estimates of limits to selection, but in reality, it would take many generations of selection for the same objective to reach those limits, and over such long periods of time epistasis (and even mutation) could prove to be important. The calculations also assume that the breeds are closed populations, but over very long periods of time, there almost certainly will be admixture with other groups. Current results are limited to four traits in three populations, and there are opportunities for future studies to provide limits for other traits of interest, as well as develop more sophisticated methodology.

Pong-Wong & Woolliams (1998) found that optimal index weights when selecting on MS variance depend on allele frequencies of the QTL and noted that there is a conflict between optimal short- and long-term selection responses. Goddard *et al.* (2009) and Hayes *et al.* (2009) have discussed weighting schemes for preserving low-frequency alleles in populations using genome-assisted selection programmes as a way of balancing selection response over time. Supporting this idea are the recent results of Jannink (2010), who showed that a simple weighting scheme can increase long-term selection gains with no appreciable loss of short-term gains, although accuracies are lower than in unweighted schemes. Cole & VanRaden (2010) recently suggested possible uses of marker data for mate selection, but noted that haplotypes were needed to make many schemes useful.

Now that haplotypes routinely are available for genotyped animals, repeated matings among parents of interest can be simulated and posterior distributions of resulting additive genetic values and MS variances computed. There are 2^{29} possible combinations of autosomes when haplotypes are sampled at random during gametogenesis (many more when recombination is considered) and haplotypes segregate independently, so there is no guaranteed way to produce animals with a specified set of haplotypes short of crossing completely inbred lines (if mutation is ignored). Matings can then be planned using various strategies, such as a factorial design in which potential sires and dams are cross-classified and simulated matings performed to identify the matings

most likely to produce the desired progeny genotypes. Offspring of matings with high expected additive genetic merit and low MS variance may be appealing to producers because differences between the expected and realized performance may be reduced. If embryos could be genotyped rapidly, cheaply and without adverse effects on viability, then such screening could increase the rate at which the MS variance is decreased.

Conversely, artificial insemination organizations may prefer matings that produce flushes of embryos with high expected additive genetic merit and high MS variance to maximize the probability of identifying individuals with extreme (high) genetic merit in the future. This represents a blending of traditional selection schemes that emphasize means gains at the expense of heterozygosity with optimal contribution systems (Sánchez *et al.* 2003) that constrain inbreeding by selection on MS with some loss of selection response. Such a scheme is easy to implement, should result in reduced rates of inbreeding with little or no loss in the rate of response to selection and will provide balance between short- and long-term gains.

Daughter pregnancy rate, milk yield and protein yield were investigated to determine whether there were differences in selection limits among traits of varying heritabilities and which had been subjected to differing amounts of selection pressure. Milk yield receives no direct weight in NM\$, but was an important selection criterion in the past, while fertility and protein yield account for 37% of the relative emphasis in NM\$. Most producers using artificial insemination in their herds are using indices rather than single-trait selection to choose bulls, so these results are hypothetical rather than representative of the real world. Even if long-term single-trait selection were common, there are antagonistic relationships among loci affecting many traits that will prevent GEBV from reaching the calculated selection limits. For example, Sonstegard *et al.* (2009) compared selected and unselected lines of Holstein cattle and found that several genomic regions had favourable effects on milk yield and unfavourable effects on DPR, suggesting an antagonistic mechanism underlying milk yield and fertility.

In the United States, estimated breeding values (EBV) are adjusted for inbreeding using a method very similar to that described above for adjusting for homozygosity. The animal model removes past inbreeding from EBV by regression and then adds back expected future inbreeding based on the current population (VanRaden 2005). In this paper, we took EBV that contain expected future inbreeding

and made an additional adjustment for inbreeding even farther in the future, when the chromosomes are projected to become even more homozygous. Adjustments for homozygosity were calculated assuming that the effect of inbreeding on lifetime performance described by Smith *et al.* (1998) was linear through the values observed in this study, which were much higher than typical pedigree inbreeding estimates. Such an assumption would probably not hold in the case of extremely inbred animals with homozygosities near 0, but the animals in this study were representative of their respective breeds. The effects on overall selection limits and DGV were small and may be attributed to the narrow range of homozygosities observed. Higher inbreeding penalties or the use of optimal contributions (Sánchez *et al.* 2003) in mate selection could be used to preserve haplotype diversity. Recombination and mutation will continue to generate novel haplotypes, and differences among the haplotypes with the most desirable CEBV are relatively small, so mate selection may be a more effective method of preserving MS variance than inbreeding adjustments.

Conclusion

Haplotypes provide managers of breeding programmes with new tools for managing heterozygosity in livestock populations. Significant progress for additive genetic merit can be made by selecting only the most desirable haplotypes, but this would lead to rapid decreases in MS variance and increases in homozygosity, producing a population that is vulnerable to rapid environmental changes or new deleterious recessives. This could be offset by carefully managing the MS variance in the population, which would result in lower rates of genetic gain. Selecting animals rather than chromosomes may result in slower progress, but limits may be the same because most chromosomes will become homozygous with either strategy.

Acknowledgements

The cooperation of the dairy records processing centres [AgriTech Analytics (Visalia, CA), AgSource Cooperative Services (Verona, WI), Dairy Records Management Systems (Raleigh, NC, and Ames, IA) and DHI Computing Services (Provo, UT)] in supplying lactation yield data is acknowledged. Two anonymous reviewers provided valuable feedback on the manuscript.

References

- Banos G., Wiggans G.R., Powell R.L. (2001) Impact of paternity errors in cow identification on genetic evaluations and international comparisons. *J. Dairy Sci.*, **84**, 2523–2529.
- Bulmer M.G. (1971) The effect of selection on genetic variability. *Am. Nat.*, **105**, 201–211.
- Cole J.B., VanRaden P.M. (2010) Visualization of results from genomic evaluations. *J. Dairy Sci.*, **93**, 2727–2740.
- Cole J.B., VanRaden P.M., O'Connell J.R., Van Tassell C.P., Sonstegard T.S., Schnabel R.D., Taylor J.F., Wiggans G.R. (2009) Distribution and location of genetic effects for dairy traits. *J. Dairy Sci.*, **92**, 2931–2946.
- Daetwyler H.D., Villanueva B., Bijma P., Woolliams J.A. (2007) Inbreeding in genome-wide selection. *J. Anim. Breed. Genet.*, **124**, 369–376.
- Dekkers J.C.M., Dentine M.R. (1991) Quantitative genetic variance associated with chromosomal markers in segregating populations. *Theor. Appl. Genet.*, **81**, 212–220.
- Goddard M.E., Wray N.R., Verbyla K., Visscher P.M. (2009) Estimating effects and making predictions from genome-wide marker data. *Stat. Sci.*, **24**, 517–529.
- Hansen L.B. (2000) Consequences of selection for milk yield from a geneticist's viewpoint. *J. Dairy Sci.*, **83**, 1145–1150.
- Hayes B.J., Visscher P.M., Goddard M.E. (2009) Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res. (Camb.)*, **91**, 47–60.
- Jannink J.-L. (2010) Dynamics of long-term genomic selection. *Genet. Sel. Evol.*, **42**, 35–68.
- Meuwissen T.H.E., Goddard M.E. (1996) The use of marker haplotypes in animal breeding. *Genet. Sel. Evol.*, **28**, 161–176.
- Meuwissen T.H.E., Van Arendonk J.A.M. (1992) Potential improvements in rate of genetic gain from marker-assisted selection in dairy cattle breeding schemes. *J. Dairy Sci.*, **75**, 1651–1659.
- Miglior F., Muir B.L., Van Doormaal B.J. (2005) Selection indices in Holstein cattle of various countries. *J. Dairy Sci.*, **88**, 1255–1263.
- Multi-State Project S-1040. (2010) Net merit as a measure of lifetime profit: 2010 revision (available at: <http://aipl.arsusda.gov/reference/nmcalc.htm>; last accessed 10 September 2010).
- Pong-Wong R., Woolliams J.A. (1998) Response to mass selection when an identified major gene is segregating. *Genet. Sel. Evol.*, **30**, 313–337.
- R Development Core Team (2010) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Sánchez L., Bijma P., Woolliams J.A. (2003) Minimizing inbreeding by managing genetic contributions across generations. *Genetics*, **164**, 1589–1595.
- Smith L.A., Cassell B., Pearson R.E. (1998) The effects of inbreeding on the lifetime performance of dairy cattle. *J. Dairy Sci.*, **81**, 2729–2737.
- Sonstegard T.S., Ma L., Cole J.B., Wiggans G.R., Crooker B.A., Van Tassell C.P., Mariani B.D., Da Y. (2009) Genome signature of artificial selection for milk yield in Holstein cattle. In: Plant and Animal Genome XVII, San Diego, USA, 10–14 January 2009, Poster 474.
- VanRaden P.M. (2005) Inbreeding adjustments and effect on genetic trend estimates. *Interbull Bull.*, **33**, 81–84.
- VanRaden P.M. (2007) Genomic measures of relationship and inbreeding. *Interbull Bull.*, **37**, 33–36.
- VanRaden P.M. (2008) Efficient methods to compute genomic predictions. *J. Dairy Sci.*, **91**, 4414–4423.
- VanRaden P.M., Sanders A.H., Tooker M.E., Miller R.H., Norman H.D., Kuhn M.T., Wiggans G.R. (2004) Development of a national genetic evaluation for cow fertility. *J. Dairy Sci.*, **87**, 2285–2292.
- VanRaden P.M., O'Connell J.R., Wiggans G.R., Weigel K.A. (2011) Genomic evaluations with many more genotypes. *Genet. Sel. Evol.*, **43**, 10.
- Wickham H. (2009) ggplot2: Elegant Graphics for Data Analysis. Springer, New York, NY, USA.
- Woolliams J.A., Bijma P., Villanueva B. (1999) Expected genetic contributions and their impact on gene flow and genetic gain. *Genetics*, **153**, 1009–1020.
- Young C.W., Seykora A.J. (1996) Estimates of inbreeding and relationship among registered Holstein Females in the United States. *J. Dairy Sci.*, **79**, 502–505.