Journal of
Animal Breeding and Genetics

J. Anim. Breed. Genet. ISSN 0931-2668

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

$$MS_U = \sum_{m=1}^{43\,382} (s_m \alpha_m - d_m \alpha_m)^2$$

respectively, where m denotes a marker, s and d are the haplotypes for the mth marker inherited from the animal's sire and dam, respectively, α_m is the estimated allele substitution effect for the mth marker, c is the cth chromosome, and n_c is the number of markers present on the cth 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 mth marker, H represents the set of all unique haplotypes in the genotyped population, n_c is the number of markers present on the cth chromosome, h_m represents the mth 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 mth 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

		Mendelian sampling variance		
Trait	Breed	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

[a]Expected Mendelian sampling variances were calculated as $\frac{1}{2}V_a$ assuming no inbreeding.
[b]The 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

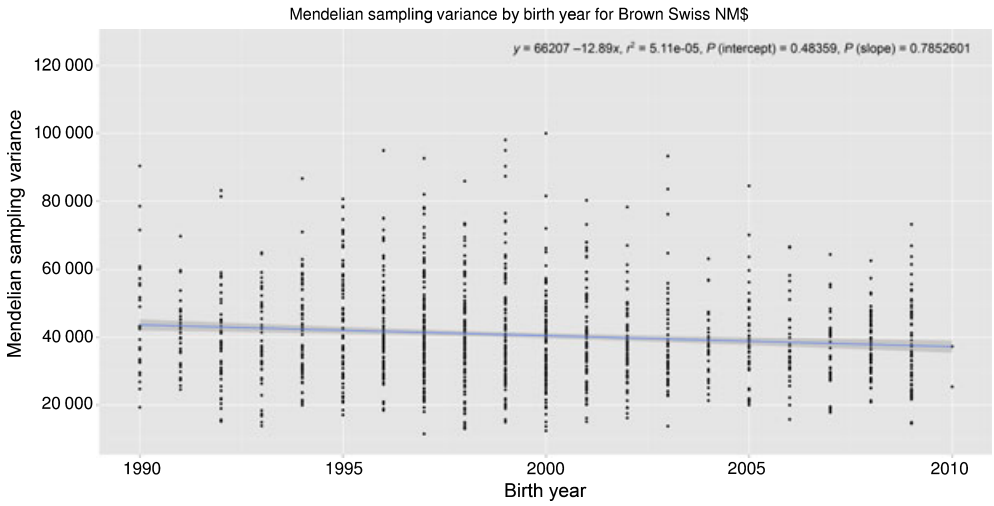

Figure 1 Changes in Mendelian sampling variance (upper bound) for lifetime net merit (NM$) in US Brown Swiss cattle born between 1990 and 2010.

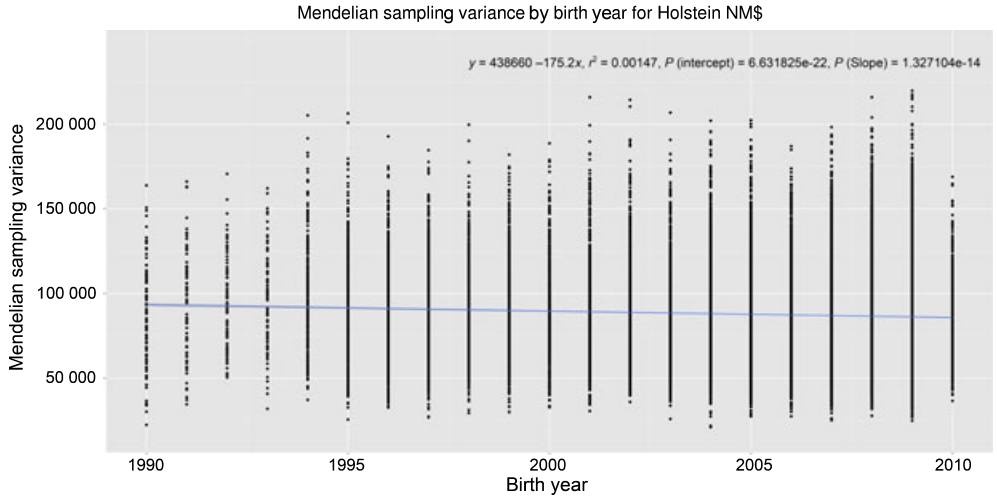

Figure 2 Changes in Mendelian sampling variance (upper bound) for lifetime net merit (NM$) in US Holstein cattle born between 1990 and 2010.

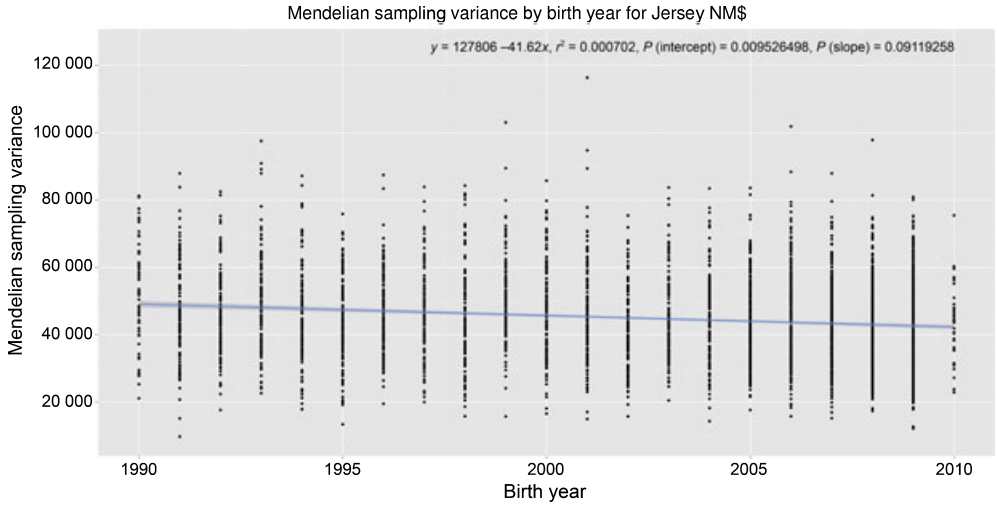

Figure 3 Changes in Mendelian sampling variance (upper bound) for lifetime net merit (NM$) in US Jersey cattle born between 1990 and 2010.

in that the *DGAT1* locus is not segregating in the latter population. Similarly, in addition to *DGAT1*, 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

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.