Use of Haplotypes to Estimate Mendelian Sampling Effects and Selection Limits

J.B. Cole[1] and P.M. VanRaden[1]

Introduction

Mendelian sampling (MS) is the difference between an individual's predicted transmitting ability (PTA) and its parent average (PA), which is the average of the sire and dam PTA. Genomic tools allow direct inspection of MS at the chromosomal level. 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 programs is due to 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 programs (Daetwyler et al., 2007). The objective of this paper is to describe the MS variation in the U.S. Jersey population, as well as discuss selection limits based on haplotypes present in the genotyped population.

Material and methods

Haplotyping. Genotypes for 43,385 single nucleotide polymorphisms in 3,689 Jersey bulls and cows were obtained using the Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA) and haplotypes were imputed with the Fortran program findhap.f90 (VanRaden et al., 2010), which combines population and pedigree haplotyping methods.

Mendelian sampling variances. Mendelian sampling was computed assuming that loci on the same chromosome were in perfect linkage (MS_C), and that all loci in the genome were unlinked (MS_U). MS_C and MS_U for each animal were calculated 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

$$MS_U = \sum_{m=1}^{43,385} (s_m \alpha_m - d_m \alpha_m)^2$$

respectively, where: m denotes a marker, s and d are the sire and dam genotypes for the m^{th} marker inherited from the animal's sire and dam, respectively, α_m is the 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. Details on the calculation of marker effects are provided in Cole et al. (2009).

[1]Animal Improvement Programs Laboratory, ARS, USDA, 10300 Baltimore Ave., Beltsville, MD 20705, USA

Selection limits. Marker values were summed for each genotyped animal to obtain chromosomal estimated breeding values (CEBV) for lifetime net merit (NM$). The CEBV were adjusted for inbreeding by subtracting 6% of an additive genetic standard deviation ($11.88) per 1% increase in homozygosity above the breed average (Smith et al., 1998). Animals with above-average heterozygosity were credited in the same manner. Mean homozygosity was 0.72 ± 0.02, ranging from 0.65 to 0.81, and was calculated as the average homozygosity of each pair of chromosomes in the genotyped animals. Empirical selection limits were calculated by combining the haplotypes with the best unadjusted or adjusted CEBV for NM$. The CEBV were summed to obtain the genomic EBV (GEBV). Matings of the 25 bulls with the best and poorest GEBV to each genotyped cow (n = 632) were simulated by randomly sampling haplotypes from each parent, and 5 matings were produced for each sire-dam pair to determine the impact of genomic selection on MS variance.

Results and discussion

Haplotyping. After haplotyping, 99.9% of Jersey genotypes were called correctly, and 141 ancestors without genotypes had sufficiently accurate imputed genotypes for inclusion in the genomic evaluation. Longer chromosomes had more distinct haplotypes, ranging from 7,287 for *Bos taurus* autosome 1 (BTA 1) to 2,460 for the X chromosome. This is expected because longer chromosomes undergo recombination more often than shorter ones, and because bulls carry only one copy of the X.

Mendelian sampling. MS_U provides an empirical estimate of the lower bound of the actual MS variance in the population, and MS_C is an approximate upper bound. 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, $MS_U = MS_C$ and will differ from the true MS variance ($0.25\Sigma\alpha_m^2$) by a constant.

The SNP used for genotyping were selected to have high average minor allele frequencies and most predicted allele substitution effects are 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 zero. When loci are linked, however, selection for markers with positive effects generates LD blocks in which the sum of effects are greater than 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 by the data. The average MS_U was much smaller (2,175 ± 123) than the average MS_C (49,290 ± 13,981).

The correlation between genomic inbreeding coefficients calculated using the genotypes and MS_U was strong and negative ($r = -0.95$; $P < 0.01$), and is expected because MS variance

decreases as homozygosity increases. However, the correlation between homozygosity and MS_C was near 0 (r = 0.06; P < 0.01). MS_C was calculated under the assumption that markers on the same chromosome were in perfect linkage. The impact of a small number of loci becoming homozygous is small when blocks of alleles are selected, as is the case in this population. The range of genomic inbreeding observed in these animals also is small – there are no extremely inbred animals in which you would expect to see whole LD blocks fixed – which may contribute to the low correlation.

Correlations of GEBV with MS_U and MS_C were calculated to determine if animals with high GEBV also had greater MS variances. The GEBV were positively correlated with MS_U (r = 0.08; P < 0.01) and negatively correlated with MS_C (r = -0.12; P < 0.01). When the 25 bulls with the best and poorest GEBV for NM$ were compared, the variance was slightly larger in the top group (58,149 ± 12,425) than the bottom group (51,556 ± 16,102).

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 and VanRaden (2010). The lower bound was produced by selecting the 30 best unadjusted haplotypes, resulting in an animal with a GEBV of +$5,243 for NM$. When inbreeding was accounted for a slightly lower GEBV of +$4,496 was obtained. The upper bound was approximated as a genotype with two copies of the marker with the largest allele substitution effect at each locus, and had a value of +$13,254. For comparison, the top Jersey bull, ALL LYNNS RESTORE VERNON-ET (29JE03647), has an EBV for NM$ of +$1,180 based on the January, 2010, genomic evaluation.

Correlations among the unadjusted CEBV and those adjusted for inbreeding ranged from 0.897 on BTA 12 to 0.998 on the X chromosome. The genotyped animals are mostly bulls and may all appear to be equally inbred on the X. For 11 chromosomes (BTA 4, 9, 13, 15, 20, 21, 22, 25, 26, 28, and X) the best genotype after adjusting for inbreeding consisted of two copies of the same haplotype, although the differences between the top- and second-ranked chromosomes was usually small. Higher inbreeding penalties or tools for calculating optimal contributions (Sánchez et al., 2003) could be used to preserve haplotype heterozygosity. Differences between the best and poorest haplotypes ranged from a maximum of $65 for BTA 1 to a minimum of $12 for BTA X.

The simulated matings provided an estimate of the impact of selection on GEBV on MS variance in the first generation of a genomic selection program. The MS variance from the matings to the top 25 bulls (n = 79,000) was 31,022 ± 8,784, which was about 63% of the MS_C estimate of 49,290 ± 13,981. This could reflect sampling bias among the genotyped cows, or relationships among bulls and/or cows, which would result in fewer distinct haplotypes. The MS variance from the matings to the bottom 25 bulls (n = 79,000) was 47,820 ± 23,210, only about 3% lower than MS_C. The genotyped bulls with the poorest GEBV also are the oldest bulls and may have haplotypes that have decreased in frequency over time, producing offspring with more sampling variation than those of the modern bulls.

Selection on GEBV assumes that average haplotype effects will be inherited, and while those effects can now be estimated there is no way to predict which haplotypes will be transmitted to the progeny. There are 2^{29} possible combinations of autosomes when haplotypes are sampled at random during gametogenesis and haplotypes segregate independently, so the only way to increase the frequency of desirable haplotypes is to select for both high overall genetic merit and reduced MS variance. The MS variance cannot be reduced to 0 due to deleterious inbreeding effects and because crossing-over continuously generates new haplotypes during meiosis. If embryos could be genotyped rapidly, cheaply, and without deleterious effects on viability then rapid-screening could increase the rate at which the MS variance is decreased.

Conclusion

Haplotypes provide managers of breeding programs 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.

References

Cole, J.B. and VanRaden, P.M. (2010). *J. Dairy Sci.*, 93:(accepted).

Cole, J.B., VanRaden, P.M., O'Connell, J.R. *et al.* (2009*). J. Dairy*. Sci., 92:2931–2946.

Daetwyler, H.D., Villanueva, B., Bijma, P. *et al.* (2007). *J. Anim. Breed. Genet.*, 124:369–376.

Hayes, B.J., Visscher, P.M., and Goddard, M.E. (2009). *Genet. Res. (Camb.)*, 91:47–60.

Sánchez, L., Bijma, P., and Woolliams, J.A. (2003). *Genetics* 164:1589–1595.

Smith, L.A., Cassell, B., and Pearson, R.E. (1998). *J. Dairy Sci.*, 81:2729–2737.

VanRaden, P.M., O'Connell, J.R., Wiggans, G.R. *et al.* (2010). J. Dairy Sci., 93(Abstr.): (accepted).

Woolliams, J.A., Bijma, P., and Villanueva, B. (1999). *Genetics*, 153:1009–1020.