J. Dairy Sci. 94:5673–5682
doi:10.3168/jds.2011-4500

Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss

P. M. VanRaden,*[1] K. M. Olson,† G. R. Wiggans,* J. B. Cole,* and M. E. Tooker*
*Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350
†National Association of Animal Breeders, Columbia, MO 66205-1033

ABSTRACT

Genomic measures of relationship and inbreeding within and across breeds were compared with pedigree measures using genotypes for 43,385 loci of 25,219 Holsteins, 3,068 Jerseys, and 872 Brown Swiss. Adjustment factors allow genomic and pedigree relationships to match more closely within breeds and in multibreed populations and were estimated using means and regressions of genomic on pedigree relationships and allele frequencies in base populations. Correlations of genomic relationships with pedigree inbreeding were higher within each breed when an allele frequency of 0.5, rather than base population frequencies, was used, whereas correlations of average genomic relationships with average pedigree relationships and also reliabilities of genomic evaluations were higher using base population frequencies. Allele frequencies differed in the 3 breeds and were correlated by 0.65 to 0.67 when estimated from genotyped animals compared with 0.72 to 0.74 when estimated from breed base populations. The largest difference in allele frequency was between Holstein and the other breeds on chromosome *Bos taurus* autosome 4 near a gene affecting appearance of white skin patches (vitiligo) in humans. Each animal's breed composition was predicted very accurately with a standard deviation of <3% using regressions on genotypes at all loci or less accurately with a standard deviation of <6% using subsets of loci. Genomic future inbreeding (half an animal's mean genomic relationship to current animals of the same breed) was correlated by 0.75 to 0.94 with expected future inbreeding (half the average pedigree relationship). Correlations of both were slightly higher with parent averages than with genomic evaluations for net merit of young Holstein bulls. Thus, rates of increase in genomic and pedigree inbreeding per generation should be slightly reduced with genomic selection, in agreement with previous simulations. Genomic inbreeding and future inbreeding have been provided with individual genomic predictions since 2008. New methods to adjust pedigree and genomic relationship matrices so that they match may provide an improved basis for multibreed genomic evaluation. Positive definite matrices can be obtained by adjusting pedigree relationships for covariances among base animals within breed, whereas adjusting genomic relationships to match pedigree relationships can introduce negative eigenvalues. Pedigree relationship matrices ignore common ancestry shared by base animals within breed and may not approximate genomic relationships well in multibreed populations.
Key words: inbreeding, breed relationship, genotype, pedigree

INTRODUCTION

Genomic relationship matrices (**G**) measure actual allele sharing, whereas pedigree relationship matrices (**A**) estimate fractions of alleles expected to be identical by descent. Pedigree relationships were the foundation of animal breeding and genetic selection, but genomic relationships are replacing pedigree relationships in many national evaluation systems (Loberg and Dürr, 2009). Genomic evaluations are calculated within breed for all countries except New Zealand, where predictions for crossbred animals are also computed (Harris and Johnson, 2010). Toosi et al. (2010) concluded from simulated data that crossbred predictions should be accurate if parent breeds are included in multibreed evaluation. de Roos et al. (2008) and Villa-Angulo et al. (2009) estimated that 50,000 markers are sufficient for estimating within-breed effects but that 300,000 to 600,000 markers may be needed to predict effects across breeds. The latter study examined genetic distances among 19 breeds including Holstein, Jersey, and Brown Swiss using genotypes of 31,857 markers for 487 animals that were grouped into parent-progeny trios within each breed.

Methods to combine **G** with **A** are needed because genotypes are not available for many animals (Legarra et al., 2009). Harris and Johnson (2010) examined methods to adjust multibreed **G** to more closely match

Received April 29, 2011.
Accepted July 20, 2011.
[1]Corresponding author: paul.vanraden@ars.usda.gov

A. An alternative is to adjust multibreed **A** to more closely match **G** and express inbreeding and relationships relative to the ancestral population that existed before breeds diverged (VanRaden, 1992). Three formulas to compute within-breed **G** were compared using simulated data (VanRaden, 2008), but tests with actual within-breed or multibreed data are needed. In the simulation, use of allele frequencies estimated from the base population rather than simple estimates resulted in much higher correlations of genomic inbreeding with pedigree inbreeding, but correlations were highest using either true frequencies in the base population or setting all frequencies to 0.5. Different goals such as computing genomic evaluations, relationships, or inbreeding could results in different choices of allele frequencies.

Selection on genomic evaluations should reduce inbreeding per generation because the best individual animals rather than families with the best genes are selected (Daetwyler et al., 2007). However, use of younger animals counteracts that benefit by increasing the number of generations per unit of time (Pedersen et al., 2010; de Roos et al., 2011). Simulation results are convincing, but few studies have examined effects of genomic selection with actual data, and few have routinely provided genomic inbreeding statistics along with evaluations. Breeders can now quantify actual allele sharing by directly examining DNA instead of only probabilities of identity by descent from pedigrees, but guidance is needed on interpreting genomic measures of inbreeding and relationship.

Breeders are interested in estimating breed composition from genotypes (Sölkner et al., 2010; Kuehn et al., 2011; Wilkinson et al., 2011). The GoldenGate Bovine3K Genotyping BeadChip (Illumina, San Diego, CA) may be applied to many animals that do not have parentage or even a breed assigned when they are genotyped. To ensure that a genotyped animal is associated with the correct breed, a current quality control step for the USDA genotype database for dairy cattle uses 622 SNP (~200 monomorphic SNP for each breed) and counts how many breed-specific SNP conflict with an animal's breed (Wiggans et al., 2010). Allele frequency differences at all loci or a genomic relationship matrix across breeds can be used to predict breed composition from genotypes rather than only verifying the breed identity of DNA samples.

Objectives of the current research were to (1) compare current and base allele frequencies within Holstein, Jersey, and Brown Swiss breeds; (2) use allele frequency differences to predict breed identity or composition; (3) develop and compare formulas to convert genomic and pedigree relationship and inbreeding coefficients to a common scale within or across breeds; and (4) explore the effect of within-breed selection on genomic and pedigree inbreeding.

MATERIALS AND METHODS

Genotypes and Allele Frequencies

Genotyped animals included 25,219 Holsteins, 3,068 Jerseys, and 872 Brown Swiss, of which 22,679 were males and 6,480 were females. The pedigree file with all known ancestors of those animals included a total of 128,425 cows and bulls. The base or founding population was defined as the unknown sires and dams of the earliest known generation, which was often 10 generations or 50 yr before the current generation. Earliest birth year was 1957 for genotyped animals and 1930 for ancestors in the pedigree file. Genotypes for crossbred animals were not available. Marker genotypes were obtained using the BovineSNP50 BeadChip (Illumina). Edits reduced the SNP used in genomic relationships to 43,385 loci for all 3 breeds (Wiggans et al., 2009).

Allele frequencies in the base (founding) population were estimated within each breed by the algorithm of Gengler et al. (2007), which uses a pedigree relationship matrix and linear mixed model equations to account for selection and drift in allele frequencies across time. Gene content of nongenotyped animals was estimated from their genotyped relatives, separately for each locus. Allele frequency of a more distant, ancestral population that existed before the breeds diverged was estimated as a simple average of the 3 breed base frequencies and used in computing relationships across breeds. Largest differences between allele frequencies in the 3 breeds were examined from both current and base frequencies, and comparative maps of other species were used to investigate potential causative genes.

Breed Identity

To predict breed identity, a dependent variable $\mathbf{y}_i$ was created for each of the 3 breeds, where i is breed (1 = Holstein, 2 = Jersey, and 3 = Brown Swiss); for example, to predict fraction of Holstein genes, Holsteins received a 1, and Jerseys and Brown Swiss received a 0 in $\mathbf{y}_1$. A few animals known to have some ancestry from another breed were excluded from the data vectors. A genomic evaluation was then computed to obtain SNP estimates and genomic PTA (**GPTA**) for each breed using the programs of VanRaden (2008). Genomic evaluations were done separately to predict the 3 breed fractions using the following model:

$$\mathbf{y}_i = \mathbf{Z}\mathbf{g}_i + \mathbf{1}'\mu_i + \mathbf{e}_i,$$

where **Z** contains genotypes for each animal (row) and marker (column) centered by base allele frequencies, $\mathbf{g}_i$ is the marker regression to predict breed i, μ_i is an intercept, and $\mathbf{e}_i$ is error. Observations are usually weighted by reciprocal of error variance, but a heritability of 99% was assumed for the reported breeds. A heavy-tailed prior was assumed for the marker effects in $\mathbf{g}_i$.

To test predictions of breed identity, data were divided into training and validation data sets. The training data set comprised bulls and cows that were proven (had daughter or their own information based on milk traits) as of July 2009 and totaled 11,053 Holsteins, 2,208 Jerseys, and 778 Brown Swiss. The validation data set had 14,794 Holsteins, 919 Jerseys, and 96 Brown Swiss bulls and heifers that were unproven as of July 2009.

Three different SNP sets were tested: the full set of 43,385 SNP, a smaller set of 3,209 SNP that included the final 2,900 SNP available on the Bovine3K chip, and the breed-specific set of 622 SNP used for quality control tests (Wiggans et al., 2010). By design, the Bovine3K chip includes 82 of the 622 breed-specific SNP. To compare different density predictions of breed composition, SNP estimates for each breed from the 3 different SNP sets were applied to the validation data.

Within-Breed Relationships

Within-breed genomic relationship and inbreeding coefficients were computed with a formula that used either counts of alleles shared (which was equivalent to using 0.5 allele frequency) or allele frequencies estimated from the base population. Simple estimates of frequency gave poor results with simulated genotypes (VanRaden, 2008) and with real genotypes; therefore, only the 2 better options for allele frequencies are reported. With p containing either 0.5 or allele frequencies and with **Z** containing values of $0 - 2p$ for homozygotes, $1 - 2p$ for heterozygotes, or $2 - 2p$ for opposite homozygotes of each animal (row) and each marker (column), **G** was computed as $\mathbf{G} = \mathbf{ZZ}'/\Sigma 2p(1 - p)$.

Coefficients of **G** were examined either unadjusted or after adjusting for regression of **G** on **A** as in VanRaden (2008). When all 3 breeds were included in the same relationship matrix, the average of the base allele frequencies for Holsteins, Jerseys, and Brown Swiss was used.

Traditional inbreeding coefficients from pedigrees ($\boldsymbol{F_A}$) were computed within breed by tabular method and compared with genomic inbreeding coefficients ($\boldsymbol{F_G}$). For young animals that had no phenotypic information or progeny data, diagonal elements F_A and F_G were obtained but not relationships among the young animals because those off-diagonals were not needed in genomic evaluation. Correlations of F_A with F_G were calculated using bulls born after 1990 because pedigrees before that time included only a few generations. Higher correlations were an easy way to check if F_G were accurate, but some other test could be better because F_A does not equal true inbreeding.

Genomic estimates of future inbreeding (**GFI**) have been provided to the dairy industry since October 2008 for each genotyped animal and are analogous to the expected future inbreeding (**EFI**) of VanRaden and Smith (1999). Both GFI and EFI are defined as half the average relationship of an animal to the current population. For GFI, the reference population is all genotyped animals born in the last 10 yr. For EFI, the reference population is a sample of females born in the last 5 yr. Thus, published GFI and EFI should have similar interpretations, but EFI examined in the remainder of the paper was computed using the same reference population as GFI to make comparisons more precise.

Correlations of official genomic evaluations and parent averages (**PA**) from April 2009 with all 4 inbreeding statistics (F_A, F_G, EFI, and GFI) were calculated for 12,296 young Holstein bulls without daughter records. The test used genomic inbreeding computed with allele frequencies estimated within each breed's base population. The only trait investigated was net merit as an indicator of routinely practiced selection. Means of the 4 statistics for all young bulls versus the top 100 for net merit were examined to quantify effects of selection.

Genomic EBV (**GEBV**) are computed officially by adjusting genotypes for base population frequencies in marker regression equations. Using the same data and nonlinear iteration method of VanRaden et al. (2009), an alternative allele frequency of 0.5 was also tested for 7 traits, and predictions were compared using squared correlations for 1,759 Holstein bulls that had daughter yield deviations in 2008 but no daughters in 2004. The purpose of this test was to determine if the allele frequencies that make elements of **G** and **A** more similar also make GEBV more accurate.

Across-Breed Relationships

Across-breed relationship and inbreeding coefficients were constructed from genomic or pedigree data by extending the concepts of VanRaden (1992). Base animals within each breed are related more to each other than to base animals of another breed, and purebred animals are more inbred than crossbred animals. Relationships among base animals within breed were determined directly from genotypes instead of the previous proposal of VanRaden (1992) to use ratios of heterosis to inbreeding depression, which can vary by trait. However,

selection of SNP can also affect genomic relationships because a particular breed may appear to have greater than its actual heterozygosity if most SNP discovery was conducted in that breed. Average heterozygosity of QTL may be less than for SNP because SNP are selected for high minor allele frequency.

Matrix **G** was computed directly using a mean of allele frequencies estimated from the base population for each breed or indirectly using 0.5 instead of base allele frequencies. Then, **G** was adjusted to $\mathbf{G}_0$ so that the 2 least-related breeds had a mean genomic relationship of 0, the upper range for $\mathbf{G}_0$ was 2, and F_G was 1 for completely homozygous animals:

$$\mathbf{G}_0 = 2(\mathbf{G} - g_{min})/(2 - g_{min}),$$

where g_{min} is the minimum mean relationship between any 2 breeds from **G**. Hayes and Goddard (2008) used a similar adjustment within breed but set the minimum of $\mathbf{G}_0$ to 0 for the least-related pair instead of group of animals. Use of a group allows pedigree relationships to act as the expected value of genomic relationships (VanRaden, 2008) and also makes elements of $\mathbf{G}_0$ much less dependent on inclusion of 1 less-related animal, whereas use of the least-related pair of individuals may prevent negative eigenvalues. The best choice may differ if the intended use of the comparison of $\mathbf{G}_0$ to **A** on a multibreed scale ($\mathbf{A}_m$) is for genomic evaluation, for display to the public, or discovery of pedigree errors. Results comparing multibreed to single-breed genomic evaluations based on the data and methods of this study were reported by Olson and VanRaden (2010).

Pedigree and genomic relationships can be compared across breeds if within-breed **A** is converted to $\mathbf{A}_m$. Harris and Johnson (2010) instead removed covariances shared within breed to produce a **G** with expected value equal to **A**. The tabular method can be modified to propagate within- and across-breed relationships among ancestors in the base population to their purebred and crossbred descendants (VanRaden, 1992). That modification required a 3 × 3 matrix containing within- and across-breed genomic relationships in the base populations (**B**), a 3 × 3 matrix containing average pedigree relationships among genotyped animals (**P**), and averages of $1 + F_A$ and $1 + F_G$ within each breed. Diagonal elements of **P** (p_{ii}) were averages of off-diagonal elements from **A** within breed i, and off-diagonals of **P** were 0 because animals of different pure breeds had no common ancestors in the pedigree file. Averages of the diagonals of **A** and **G** for breed i were labeled a_i and g_i, respectively.

Off-diagonal elements of **B** (b_{ij}) were averages of elements of **G** for animals of different breeds i and j. Diagonal elements of **B** (b_{ii}) were computed from averages of off-diagonal elements of **G** for animals within the same breed $\left(b^*_{ii}\right)$ with an adjustment for average pedigree relationship among genotyped animals in breed i. Diagonals (but not off-diagonals) of **B** are adjusted because inbreeding and relationships accumulate within, but not between, distinct breeds. Accumulated pedigree relationships were scaled using differences between diagonals and off-diagonals of **G** and $\mathbf{A}_m$ and then subtracted from current genomic relationships within breed:

$$b_{ii} = b^*_{ii} - p_{ii}(g_i - b^*_{ii}) / (a_i - p_{ii}).$$

Within each breed, individual elements of **A** can be converted to elements of $\mathbf{A}_m$ using

$$\mathbf{A}_m = b_{ii} + \mathbf{A}(g_i - b^*_{ii}) / (a_i - p_{ii}).$$

For purebred animals of different breeds, elements of **A** are 0 and elements of $\mathbf{A}_m$ are set to b_{ij}, the breed relationship. The tabular method can be used to obtain elements of $\mathbf{A}_m$ for crossbred animals or for all animals and results in positive definite $\mathbf{A}_m$ if **B** is positive definite. Adjusting elements in **A** to match those in **G** seems more sensible than adjusting **G** to match **A**, because removal of covariance from **G** can create negative eigenvalues (Misztal et al., 2010). This problem was confirmed with Brown Swiss and Jersey data from this study. One negative eigenvalue about 2% of the size of the largest positive eigenvalue occurred within each of the 2 breeds. Holstein and all-breed matrices were not examined because of their size. Use of $\mathbf{A}_m$ and **G** more closely model actual gene sharing in mixed breed populations and account for past drift of allele frequencies within isolated breed populations. Wright's (1922) inbreeding and relationship coefficients in **A** assume just one randomly mating base population.

RESULTS AND DISCUSSION

Genotypes and Allele Frequencies

Long segments of homozygous chromosome indicate that an animal's parents share recent common ancestors. The largest number of consecutive homozygous SNP was 1,995, and 38 animals had long segments of >1,500 consecutive homozygous loci. Those segments represent chromosome pairs that are completely or almost completely homozygous, because a chromosome contained 1,446 markers on average, and the longest (chromosome 1) contained 2,748 markers. Four-generation pedigrees were examined for the 10 animals with the longest homozygous regions. In 9 pedigrees, both

parents had ≥1 parent or grandparent in common. In the remaining pedigree, a famous bull appeared 3 times as a great-grandsire of the parents. Thus, genomic evidence of inbreeding accurately detected pedigree inbreeding.

Current allele frequencies in the 3 breeds were correlated by 0.65 to 0.67, whereas allele frequencies estimated from the base populations of the 3 breeds were correlated by 0.72 to 0.74. Thus, the estimation process removed recent drift within the populations and revealed that breeds had been more similar >10 generations ago than they are now, as expected from genetic drift of frequencies across time.

The largest frequency difference was observed for marker HapMap53144-ss46525999, which is located on BTA4 at 79,550,928 bp on Btau_4.2 and at 77,555,681 bp on UMD 3.1 genome assemblies (National Center for Biotechnology Information, 2011). Currently, 1 allele is nearly fixed for Holsteins with a frequency >0.99, but frequencies are <0.02 for both Jerseys and Brown Swiss, for which the opposite allele is nearly fixed. Two nearby markers have almost the same distortion for Jerseys and intermediate allele frequencies for Brown Swiss. Frequency differences among base populations were also >0.95, which indicated that selection occurred long ago. Prior to development of effective statistical tools for genetic selection, selection operated largely on simple phenotypic characteristics, such as coat color and pattern. The nearby calcium-dependent protein kinase II β (*CAMK2B*) gene that begins at 79,687,054 bp on Btau_4.2 and at 77,693,040 bp on UMD 3.1 may affect production of melanin pigment (melanogenesis) in several species. Markers indicate that this gene has a role in causing white patches to appear on human skin (Smith and Sturm, 2010), a condition known as vitiligo. Thus, *CAMK2B* may also be partially responsible for the coat color pattern of Holsteins.

Jersey and Brown Swiss allele frequencies differed by >0.99 for marker HapMap33628-BTC-041023 on BTA6 at 38,407,127 bp on Btau_4.2 and at 38,939,012 bp on UMD 3.1 genome assemblies (National Center for Biotechnology Information, 2011). Five other markers within 1 million bp of this marker also had frequency differences >0.96 between the 2 breeds. Holstein frequencies ranged from 0.20 to 0.72 for those markers. Known genes in this range include ATP-binding cassette, sub-family G (WHITE), member 2 (*ABCG2*) and secreted phosphoprotein 1 (*SPP1*), but those genes may not be an obvious explanation for the breed differences. The v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (*KIT*) gene at a different location on BTA6 was previously thought to control the presence or absence of spots, but Fontanesi et al. (2010) concluded that *KIT* instead affects the size of spots.

Marker BTB-00557585 on BTA14 with current frequency 99% in Holstein, 8% in Jersey, and 41% in Brown Swiss was also included in the 622 SNP breed identity panel at 22,803,366 bp on Btau_4.2 and 25,352,790 on UMD 3.1 at the same location as a large QTL affecting stature in Holstein by Jersey crossbreds (Karim et al., 2011). This also demonstrates how different selection within breeds may have changed the allele frequencies near QTL.

Breed Identity

Predicted breed compositions for the validation data are in Table 1. For the 43,385 SNP set, means ranged from 0.994 to 1.000, and standard deviations (**SD**) were 0.008 for Holsteins, 0.028 for Jerseys, and 0.021 for Brown Swiss within their respective breeds. The 3,209 SNP set had larger SD of 0.031 for Holsteins, 0.063 for Jerseys, and 0.036 for Brown Swiss. For the 622 SNP set, 5 of the 9 SD were actually slightly smaller than corresponding SD for the 3,209 SNP set. This may indicate that a prior distribution across 3,209 was not as effective as a prior with all emphasis on the preselected 622 known to have large frequency differences.

Breed predictions in this study were more accurate than those of Kuehn et al. (2011) and Wilkinson et al. (2011), probably because their frequencies for many breeds were each estimated using few animals per breed, whereas the frequencies for 3 dairy breeds in this study were each estimated using hundreds or thousands of animals. The validation methods also differed: predictions were tested on crossbreds for beef breeds versus on purebreds for dairy breeds.

Breed predictions were very accurate for most animals, as indicated by breed fractions very close to 1 or 0 and small SD of <3% when all markers were used. Predictions for a few foreign animals were less accurate, with estimated breed fractions of <1 for the reported breed. Other animals that were least like their identified breed were born in the early 1960s for the Holstein and Brown Swiss breeds. Jersey animals identified as least like Jerseys were descendants from 1 Jersey cow that was determined to have been sired by a bull with some Holstein ancestry. The SD was largest for Jerseys partially because of that common ancestor for 83 of the animals in the Jersey validation data set. When those animals were removed from the validation data set, SD decreased to 0.054 for the 3,209 SNP data set and to 0.015 for the 43,385 set. The alternate breed was correctly identified as being Holstein for the removed animals. In March 2010, the American Jersey Cattle Association (Reynoldsburg, OH) revised its registration, recovery, and expansion programs to accommodate this information from DNA (American Jersey Cattle Association, 2010).

Table 1. Means for estimated breed fractions of validation animals using different numbers of SNP

SNP (no.)	Reported breed of validation animals[1]	Estimated breed fraction		
		Holstein	Jersey	Brown Swiss
43,385	Holstein	1.000 ± 0.008	<0.001 ± 0.006	<0.001 ± 0.006
	Jersey	0.004 ± 0.027	0.996 ± 0.028	<0.001 ± 0.009
	Brown Swiss	0.003 ± 0.015	0.003 ± 0.025	0.994 ± 0.021
3,209	Holstein	1.004 ± 0.031	<0.001 ± 0.022	−0.005 ± 0.024
	Jersey	0.016 ± 0.055	0.978 ± 0.063	0.003 ± 0.015
	Brown Swiss	0.007 ± 0.032	0.004 ± 0.025	0.989 ± 0.036
622	Holstein	1.002 ± 0.019	−0.001 ± 0.017	−0.001 ± 0.016
	Jersey	0.007 ± 0.049	0.989 ± 0.047	0.004 ± 0.026
	Brown Swiss	0.006 ± 0.051	0.002 ± 0.028	0.992 ± 0.051

[1]14,794 Holsteins, 919 Jerseys, and 96 Brown Swiss.

The genotyped animals included no first-generation crossbreds; therefore, validation based on crossbred information is not available. However, the pedigree file indicated that one animal was seven-eighths Holstein and one-eighth Jersey from a Jersey great-grandparent, and the 43,385 SNP analysis correctly identified that cow as 13% Jersey and 86% Holstein. Percentages from each breed do not always sum to exactly 100% because the 3 equations are solved independently and even purebreds may be more or less related to their breed.

An animal with a great-grandparent from a different breed can also be detected with the smaller 3,209 SNP set, but ancestors from another breed further back in the pedigree cannot be detected as accurately as with the 43,385 SNP set. Bovine3K genotypes can be used as a first step in identification and then if results are questionable and breed is of great concern, BovineSNP50 genotypes could be used.

Within-Breed Relationships

Intercepts and coefficients for regression of genomic on pedigree relationships within breed are in Table 2. When computed using allele frequencies estimated from the base population, elements of **G** are similar to elements of **A** for each breed. Coefficients for regression of **G** on **A** were less than the desired 1.0 and ranged from 0.89 to 0.97; intercepts were slightly less than the desired value of 0 and ranged from −0.03 to −0.05. When computed with an allele frequency of 0.5, regression coefficients ranged from 0.632 to 0.692, and intercepts ranged from 0.559 to 0.775. Thus, scaling is needed to make **G** and **A** more similar if 0.5 allele frequencies are used.

Breed means and SD of genomic and pedigree inbreeding statistics GFI, EFI, F_G, and F_A for animals born since 1990 are in Tables 3 and 4. Using base allele frequencies, mean F_G were lower than F_A and slightly negative before adjusting for the regressions of **G** on **A** in Table 2 but higher than F_A after adjustment. Mean F_G before adjustment were also slightly negative in simulation when estimated allele frequencies from the base population were used (VanRaden, 2008). For all 3 breeds, mean F_G computed using 0.5 allele frequencies were very high initially but close to F_A after adjusting for the regressions. The bull with the highest F_G also had a high F_A of 15.2 because his parents were half siblings.

The SD for F_G were larger than for F_A because genomic inbreeding measures each individual's homozygosity instead of the mean homozygosity expected from common ancestors. Means of F_G could be larger than F_A because selection causes favorable alleles to be transmitted to later generations at greater than the assumed probability of 50%. Thus, true inbreeding may exceed pedigree inbreeding (Pedersen et al., 2010).

Table 2. Within-breed regressions of genomic on pedigree relationships computed using an allele frequency estimated from the base population or 0.5 in genomic relationships

Breed	Allele frequency			
	Base population		0.5	
	Intercept	Regression coefficient	Intercept	Regression coefficient
Holstein	−0.05	0.89	0.559	0.692
Jersey	−0.03	0.97	0.775	0.632
Brown Swiss	−0.05	0.95	0.729	0.644

Table 3. Within-breed means for animals born since 1990 for expected future inbreeding (EFI) and genomic future inbreeding (GFI) computed using allele frequencies estimated from the base population or 0.5 and adjusting for regression of genomic on pedigree relationships

		Inbreeding (%)		
Statistic	Allele frequency	Holstein	Jersey	Brown Swiss
EFI	—	6.0 ± 0.7	8.2 ± 1.7	7.0 ± 1.4
GFI	Base population	5.7 ± 0.5	7.9 ± 1.8	7.1 ± 1.5
	0.5	6.1 ± 0.8	8.0 ± 2.1	7.0 ± 1.6

Multiple regressions of F_G on F_A, birth year, and sex effect (male minus female) for animals born since 1990 are in Table 5 by breed. Genomic inbreeding was higher for males than for females because the X chromosome is coded as homozygous in males. The alternative strategy of reporting only autosomal inbreeding was not used because the X chromosome does contribute to inbreeding in females and because Mendelian sampling within offspring of the same sex is lower from the male parent, which agrees with their higher reported F_G.

Correlations of F_G with F_A and GFI with EFI are also in Table 5 by breed. Only bulls born since 1990 were included so that the sex difference in mean F_G and the less-complete pedigrees of earlier bulls would not affect the correlations. Correlations of F_G with F_A ranged from 0.50 to 0.56 when base allele frequencies were used and from 0.59 to 0.68 when an allele frequency of 0.5 was used. A slightly smaller advantage of 0.06 was observed with simulated genotypes (VanRaden, 2008), and correlations using true instead of estimated base frequencies equaled those using 0.5 frequencies. Correlations in the simulation were slightly higher for both Holstein and Jersey bulls. Hayes and Goddard (2008) also obtained a higher correlation of 0.69 using 9,323 actual markers and 0.5 frequencies for Australian Angus bulls. Estimation of true frequencies in the base population is difficult, and ignoring these estimates when computing **G** may be better than using them.

Correlation results for future inbreeding were opposite of those for individual inbreeding: correlations of GFI with EFI with F_A were higher within each breed when base allele frequencies rather than 0.5 frequencies were used. Means and SD of GFI and EFI in Table 3 were similar whether base frequencies or 0.5 frequencies were used, and means were also close to F_A. Correlations of GFI with EFI are higher than those of F_G with F_A because sampling of chromosomes can cause large deviations of F_G from F_A, whereas deviations between GFI and EFI are reduced because both are averages across many individuals. Computation of F_G and GFI used base frequencies for USDA genetic evaluations beginning in April 2008 but then switched to 0.5 frequencies in January 2010 for better F_G.

Marker regressions gave more accurate predictions for all 7 traits tested when base population rather than 0.5 allele frequencies were subtracted from marker genotype values. Squared correlations with daughter yield deviations averaged 0.403 using base frequencies compared with 0.385 using 0.5 frequencies. Thus, the procedure of subtracting base frequencies in genomic regression to compute GEBV was continued for USDA evaluations. For Canadian evaluations using the same genomic data and base frequency, slight gains in accuracy resulted if **G** was not adjusted by regression of genomic on pedigree relationships (Gerrit Kistemaker, Canadian Dairy Network, Guelph, Ontario, Canada, personal communication). This could be caused by a negative eigenvalue introduced by the regression adjustment.

The PA for net merit of young Holstein bulls had correlations of 0.10 with F_A, −0.09 with F_G, 0.43 with EFI, and 0.30 with GFI. The GEBV for those same

Table 4. Within-breed means for animals born since 1990 for pedigree inbreeding coefficients (F_A) and genomic inbreeding coefficients (F_G) computed using allele frequencies estimated from the base population or 0.5, with or without adjustment for regression of genomic on pedigree relationships

			Inbreeding (%)		
Statistic	Allele frequency	Regression adjustment	Holstein	Jersey	Brown Swiss
F_A	—	—	5.5 ± 1.9	6.5 ± 3.1	5.1 ± 2.5
F_G	Base population	No	−3.2 ± 3.0	−1.5 ± 3.9	−1.2 ± 3.6
		Yes	20.8 ± 3.5	16.1 ± 4.4	9.2 ± 3.8
	0.5	No	32.7 ± 2.2	43.6 ± 2.9	40.7 ± 2.5
		Yes	11.0 ± 3.2	4.6 ± 4.6	5.4 ± 3.9

Table 5. Within-breed multiple regressions of genomic inbreeding coefficient (F_G) computed using allele frequencies estimated from the base population or 0.5 on pedigree inbreeding coefficient (F_A), birth year, and sex effect (male minus female) for animals born since 1990 and correlations of F_G with F_A and genomic future inbreeding (GFI) with expected future inbreeding (EFI) for bulls born since 1990

		Multiple regression of F_G (%) on:				Correlations	
Allele frequency	Breed	1990 Intercept	F_A	Birth year	Sex	F_G, F_A	GFI, EFI
Base population	Holstein	15.2	0.90	−0.06	1.8	0.50	0.82
	Jersey	8.8	0.76	0.09	1.9	0.56	0.94
	Brown Swiss	2.9	0.82	0.02	2.0	0.56	0.94
0.5	Holstein	4.0	0.97	0.02	1.8	0.59	0.75
	Jersey	−3.4	1.01	−0.06	2.4	0.68	0.88
	Brown Swiss	−0.8	0.92	0.05	1.2	0.61	0.87

bulls had correlations of 0.07 with F_A, −0.07 with F_G, 0.33 with EFI, and 0.28 with GFI. The correlations near zero with individual inbreeding are not a concern. However, the higher correlations for PA than for GEBV with EFI and GFI indicate that PA selection causes closer relationships among the selected animals, which agrees with results from simulation (Daetwyler et al., 2007; Pedersen et al., 2010; de Roos et al., 2011). The difference in mean relationships from genomic or PA selection is smaller when measured by GFI than by EFI because GFI reflects actual alleles selected, whereas EFI assumes that all alleles have 50% transmission probability.

Means of F_A, F_G, EFI, and GFI were all 0.1 to 0.2% higher for the top 100 young bulls for GEBV net merit than for all young bulls, indicating only small increases in inbreeding from genomic selection. However, EFI and GFI quantify relationships of the selected animals to the existing population rather than to the other selected animals. Average relationships among the selected young animals may be higher but are not routinely available.

Across-Breed Relationships

Average pedigree relationships (Table 6) ranged from 0.110 to 0.135 within breed but were exactly zero for animals of different breeds as expected. Genomic relationships (Table 6) computed using an average of the 3 base population frequencies had much higher means within breed (0.442 to 0.621) but slightly negative means for animals of different breeds (−0.084 to −0.116). Average elements of **G** computed with 0.5 frequencies ranged from 0.639 to 0.860 within breed and 0.485 to 0.549 across breeds before adjusting minimum breed covariance to 0. Elements of $\mathbf{G}_0$ were much lower after adjustment (0.204 to 0.495 within breed and 0.000 to 0.083 across breeds).

Genomic relationships within each breed's base population are also compared in Table 6. Elements of **B** ranged from 0.086 to 0.368 within breed and were lower than the observed genomic relationships because accumulated pedigree relationships were removed. Jerseys were more homozygous than Holsteins, in agreement with Harris and Johnson (2010), as were Brown

Table 6. Mean numerator relationships from a 3-breed matrix using pedigree alone (**A**), pedigree adjusted for breed relationships ($\mathbf{A}_m$), 0.5 allele frequencies without (**G**) or with ($\mathbf{G}_0$) adjustment of minimum breed covariance to 0, breed base relationships (**B**), and allele frequencies estimated from the base population

		Mean numerator relationships					
		Within breed			Across breeds		
Allele frequency	Matrix	Holstein	Jersey	Brown Swiss	Holstein, Jersey	Holstein, Brown Swiss	Jersey, Brown Swiss
—[1]	**A**	0.110	0.135	0.124	0.000	0.000	0.000
—	$\mathbf{A}_m$	0.204	0.495	0.433	0.000	0.005	0.083
0.5	$\mathbf{G}_0$	0.204	0.495	0.433	0.000	0.005	0.083
0.5	**G**	0.639	0.860	0.650	0.485	0.485	0.549
0.5	**B**	0.086	0.368	0.327	0.000	0.005	0.083
Base population[2]	**G**	0.442	0.621	0.555	−0.103	−0.084	−0.116

[1]Relationships calculated from pedigree (no genomic data).

[2]Simple average of Holstein, Jersey, and Brown Swiss base frequencies.

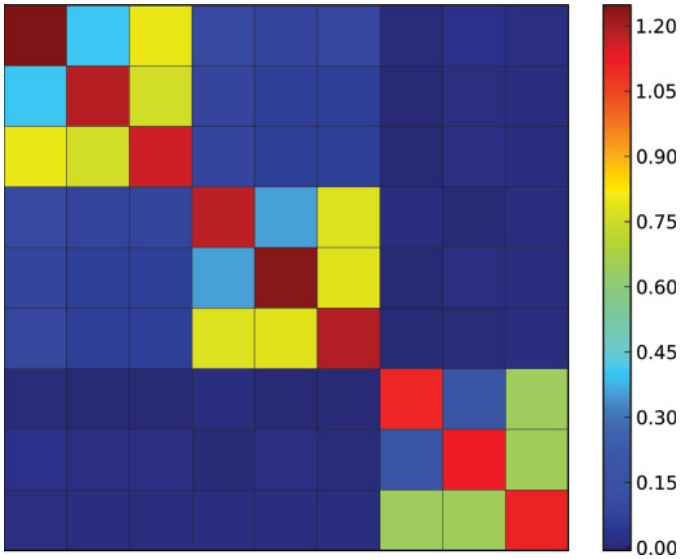

Figure 1. Genomic relationships among 3 Brown Swiss (upper left), Jersey (center), and Holstein (bottom right) trios consisting of sire, dam, and 1 progeny. Color version available in the online PDF.

Swiss. Homozygosity and breed relationships may have been affected by more SNP discovery and selection in Holsteins. Jerseys and Brown Swiss were slightly more related to each other than to Holsteins, whereas Villa-Angulo et al. (2009) reported Holsteins and Brown Swiss as less related and Jerseys intermediate between them. When elements of **A** were adjusted for the breed base relationships, the average pedigree relationships within and across breeds in $\mathbf{A}_m$ by design exactly equal average genomic relationships in $\mathbf{G}_0$.

Inbreeding coefficients computed using pedigree alone, pedigree adjusted for breed relationships, 0.5 allele frequencies without or with adjustment of minimum breed covariance to 0, or base allele frequencies are shown in Table 7. Means within breed ranged from 32.8 to 43.3% from **G** compared with 11.2 to 25.1% from $\mathbf{G}_0$ after minimum breed covariance was adjusted to zero. Averages from **G** using base population frequency ranged from 7.0 to 16.4% and were expected to be much higher than those from **A** but were not. After adjustment for breed relationships, pedigree inbreeding coefficients from $\mathbf{A}_m$ matched those from $\mathbf{G}_0$ on average by design. If 0.5 frequencies are used, F_G within each breed do not change rank when converted to multibreed scale, whereas if average base allele frequency across the 3 breeds is used instead of within each breed, the F_G do change rank and were correlated by 0.83 to 0.87 on within- versus across-breed scale.

Genomic relationships among 3 Brown Swiss (upper left), Jersey (center), and Holstein (bottom right) trios (sire, dam, and 1 progeny) are shown in Figure 1. The relationships among each sire and dam are easily seen, as are relationships among parents and progeny. The pattern of $\mathbf{G}_0$ also revealed relationships among those animals that were not accounted for by pedigree, which was expected because breeds of cattle have diverged only recently in an evolutionary sense. A similar figure that included many more purebred and crossbred Jerseys and Holsteins was presented by Harris and Johnson (2010) and also demonstrated that individuals of the same breed are more related to each other than to those of a different breed. The breed covariances can be included in $\mathbf{A}_m$ instead of removed from **G**.

CONCLUSIONS

Genomic relationship matrices were computed for 29,159 animals of 3 breeds to quantify fractions of alleles identical-in-state among animals of the same breed or of different breeds. Jerseys were the most homozygous, and Brown Swiss and Jersey were the most related of the 3 breeds. Large differences in allele frequency, such as the opposite alleles on BTA4 and BTA6 that are fixed in different breeds, identify regions of the genome that diverged during breed formation. Breed composition was predicted more accurately by regressing on all SNP than on subsets or by counting only monomorphic breed-specific SNP. Methods were developed to compare genomic and pedigree relationships on the same scale by modifying pedigree relationship matrices to account

Table 7. Mean inbreeding coefficients from pedigree alone (**A**), pedigree adjusted for breed relationships ($\mathbf{A}_m$), 0.5 allele frequencies without (**G**) or with ($\mathbf{G}_0$) adjustment of minimum breed covariance to 0, and allele frequencies estimated from the base population

Allele frequency	Matrix	Holstein	Jersey	Brown Swiss
—[1]	**A**	5.6	6.1	4.9
—	$\mathbf{A}_m$	11.2	25.1	13.0
0.5	$\mathbf{G}_0$	11.2	25.1	13.0
0.5	**G**	32.8	43.3	34.1
Base population[2]	**G**	16.4	8.1	7.0

[1]Relationships calculated from pedigree (no genomic data).
[2]Simple average of Holstein, Jersey, and Brown Swiss base frequencies.

for allele sharing within each breed's founding population. Adjustment of pedigree relationships to match genomic rather than vice versa preserves positive definite matrices and the biologic fact that animals are more related within than across breeds. Genomic inbreeding and future inbreeding are currently reported within breed for all genotyped animals to measure true rather than expected homozygosity. In 2008, F_G and GFI were reported using within-breed base allele frequencies but were revised in 2010 to use 0.5 frequencies because means of F_G were more similar to F_A and correlations of F_G with F_A were higher. However, correlations of GFI with EFI were higher when base allele frequencies were used, and genomic evaluations continue to subtract base frequencies from genotypes because this made GEBV more accurate. Use of 0.5 frequencies is useful in across-breed relationship matrices with adjustment to set average relationship of the 2 least-related breeds to zero. Selection on GEBV instead of PA reduced average relationships and inbreeding in the selected population slightly, which agrees with previous simulation studies. Further development and refinement of methods may be needed to include crossbreds in genomic evaluations.

ACKNOWLEDGMENTS

The authors thank two anonymous reviewers for helpful suggestions and Suzanne Hubbard, Animal Improvement Programs Laboratory, ARS, USDA (Beltsville, MD), for technical editing of the manuscript.

REFERENCES

American Jersey Cattle Association. 2010. Animal recording program expanded, action on "Gratitude." Jersey J. 57:19.

Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. Inbreeding in genome-wide selection. J. Anim. Breed. Genet. 124:369–376.

de Roos, A. P. W., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics 179:1503–1512.

de Roos, A. P. W., C. Schrooten, R. F. Veerkamp, and J. A. M. van Arendonk. 2011. Effects of genomic selection on genetic improvement, inbreeding, and merit of young versus proven bulls. J. Dairy Sci. 94:1559–1567.

Fontanesi, L., E. Scotti, and V. Russo. 2010. Analysis of SNPs in the *KIT* gene of cattle with different coat colour patterns and perspectives to use these markers for breed traceability and authentication of beef and dairy products. Ital. J. Anim. Sci. 9:e42.

Gengler, N., P. Mayeres, and M. Szydlowski. 2007. A simple method to approximate gene content in large pedigree populations: Application to the myostatin gene in dual-purpose Belgian Blue cattle. Animal 1:21–28.

Harris, B. L., and D. L. Johnson. 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. J. Dairy Sci. 93:1243–1252.

Hayes, B. J., and M. E. Goddard. 2008. Technical note: Prediction of breeding values using marker-derived relationship matrices. J. Anim. Sci. 86:2089–2092.

Karim, L., H. Takeda, L. Lin, T. Druet, J. A. C. Arias, D. Baurain, N. Cambisano, S. R. Davis, F. Farnir, B. Grisart, B. L. Harris, M. D. Keehan, M. D. Littlejohn, R. J. Spelman, M. Georges, and W. Coppieters. 2011. Variants modulating the expression of a chromosome domain encompassing *PLAG1* influence bovine stature. Nat. Genet. 43:405–413.

Kuehn, L. A., J. W. Keele, G. L. Bennett, T. G. McDaneld, T. P. L. Smith, W. M. Snelling, T. S. Sonstegard, and R. M. Thallman. 2011. Predicting breed composition using breed frequencies of 50,000 markers from the U.S. Meat Animal Research Center 2,000 bull project. J. Anim. Sci. 89:1742–1750.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663.

Loberg, A., and J. W. Dürr. 2009. Interbull survey on the use of genomic information. Interbull Bull. 39:3–13.

Misztal, I., I. Aguilar, A. Legarra, S. Tsuruta, D. L. Johnson, and T. J. Lawlor. 2010. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation. Commun. No. 50 in Proc. 9th World Congr. Genet. Appl. Livest. Prod., Leipzig, Germany. Gesellschaft für Tierzuchtwissenschaften e.V., Gießen, Germany.

National Center for Biotechnology Information. 2011. Reference sequence (RefSeq). Accessed April 21, 2011. http://www.ncbi.nlm.nih.gov/RefSeq/.

Olson, K. M., and P. M. VanRaden. 2010. Multibreed genomic evaluation of dairy cattle. J. Dairy Sci. 93(E-Suppl. 1):471. (Abstr.)

Pedersen, L. D., A. C. Sørenson, and P. Berg. 2010. Marker-assisted selection reduces expected inbreeding but can result in large effects of hitchhiking. J. Anim. Breed. Genet. 127:189–198.

Smith, A. G., and R. A. Sturm. 2010. Multiple genes and locus interactions in susceptibility to vitiligo. J. Invest. Dermatol. 130:643–645.

Sölkner, J., A. Frkonja, H. W. R. Raadsma, E. Jonas, G. Thaller, E. Gootwine, E. Seroussi, C. Fuerst, C. Egger-Danner, and B. Gredler. 2010. Estimation of individual levels of admixture in crossbred populations from SNP chip data: Examples with sheep and cattle populations. Preliminary proceedings of 2010 Interbull meeting, Riga, Latvia. Accessed April 6, 2011. http://www.interbull.org/images/stories/Solkner_Interbull10_submitted.pdf.

Toosi, A., R. L. Fernando, and J. C. M. Dekkers. 2010. Genomic selection in admixed and crossbred populations. J. Anim. Sci. 88:32–46.

VanRaden, P. M. 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. J. Dairy Sci. 75:3136–3144.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423.

VanRaden, P. M., and L. A. Smith. 1999. Selection and mating considering expected inbreeding of future progeny. J. Dairy Sci. 82:2771–2778.

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24.

Villa-Angulo, R., L. K. Matukumalli, C. A. Gill, J. Choi, C. P. Van Tassell, and J. J. Grefenstette. 2009. High-resolution haplotype block structure in the cattle genome. BMC Genet. 10:19 doi:10.1186/1471-2156-10-19.

Wiggans, G. R., T. S. Sonstegard, P. M. VanRaden, L. K. Matukumalli, R. D. Schnabel, J. F. Taylor, F. S. Schenkel, and C. P. Van Tassell. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. J. Dairy Sci. 92:3431–3436.

Wiggans, G. R., P. M. VanRaden, L. R. Bacheller, M. E. Tooker, J. L. Hutchison, T. A. Cooper, and T. S. Sonstegard. 2010. Selection and management of DNA markers for use in genomic evaluation. J. Dairy Sci. 93:2287–2292.

Wilkinson, S., P. Weiner, A. L. Archibald, A. Law, R. D. Schnabel, S. D. McKay, J. F. Taylor, and R. Ogden. 2011. Evaluation of approaches for identifying population informative markers from high density SNP chips. BMC Genet. 12:45.

Wright, S. 1922. Coefficients of inbreeding and relationship. Am. Nat. 56:330–338.