Visualization of the transmission of direct genomic values for paternal and maternal chromosomes for 15 traits in US Brown Swiss, Holstein, and Jersey cattle

J. B. Cole[1] and D. J. Null
Animal Improvement Programs Laboratory, Agricultural Research Service, US Department of Agriculture, Beltsville, MD 20705-2350

ABSTRACT

Haplotypes are available for 220,671 Brown Swiss, Holstein, and Jersey bulls and cows that received genomic evaluations in August 2012. Differences in least squares means of direct genomic values (DGV) for paternal and maternal haplotypes of *Bos taurus* autosomes 1, 6, 14, and 18 for lifetime net merit were significant in all but one case. Those chromosomes were chosen to represent cases with and without known quantitative trait loci, and other chromosomes may differ as well. Paternal haplotypes had higher DGV than maternal haplotypes in most cases, and differences were larger when quantitative trait loci were present. Longer chromosomes generally accounted for more variance than shorter chromosomes, and differences among breeds were consistent with known mutations of large effect. For example, *Bos taurus* autosome 18 accounted for 2.5, 7, and 2.6% of the variance in lifetime net merit for Brown Swiss (BS), Holsteins, and Jerseys, respectively. Distributions of the number of positive DGV inherited from sires and dams were negatively skewed in all breeds, and modes were slightly higher for paternally than maternally derived haplotypes in Holsteins and BS (22 vs. 20 and 22 vs. 21, respectively) and slightly lower in BS (17 vs. 19). Graphical representations of DGV are available to all users through a query on the Animal Improvement Programs Laboratory (ARS, USDA, Beltsville, MD) web site. Query results were also used to illustrate several quantitative genetic principles using genotype information from real animals. For example, offspring DGV can be compared with parental DGV to demonstrate that a parent transmits the average value of its 2 chromosomes to its progeny. The frequency of DGV with positive and negative values in animals of different ages can be used to show how selection affects allele frequencies. The effect of selection for alleles with large effects versus those with small effects is demonstrated using an animal with undesirable alleles for a marker with a large effect but many desirable alleles for markers with small effects. Strategies for the use of those data in selection programs are being studied, and work is underway to add data on conformation traits to the system.

Key words: direct genomic value, genomic selection, haplotype, visualization

INTRODUCTION

The success of genomic selection programs in dairy cattle populations has resulted in a growing demand for new tools to help understand the resulting data. Genotypes for more than 200,000 animals are available in the national dairy database (Beltsville, MD) and reliable haplotypes are available for most of those animals as well (VanRaden et al., 2011b). Several papers have addressed the effect of genomic selection in cattle improvement and conservation programs (Engelsma et al., 2011; Bastiaansen et al., 2012; Börner and Reinsch, 2012), and some studies have discussed the use of SNP genotypes for mate selection (Nishio et al., 2010; Toro and Varona, 2010; Cole and VanRaden, 2011; Pryce et al., 2012). Cole and VanRaden (2011) suggested that the greatest selection response possible on an overall economic index in a population could be achieved by constructing genotypes from desirable haplotypes, creating an animal whose genotype consisted of 2 copies each of the best haplotype for each chromosome. When the direct genomic values (**DGV**) of each haplotype were adjusted to account for inbreeding, the best genotypes consisted of 2 copies each of the same haplotype for 22 to 26 chromosomes, and differences between the best and next-best haplotypes generally were small (<$10). Kemper et al. (2012) recently have confirmed that such a strategy produces the largest long-term selection gains under some conditions, but they showed that selection on genomic breeding values with constraints on coancestry provided similar long-term gains, faster short-term gains, and greater flexibility. However, strategies for using haplotype data for mate selection still have not been described in detail in the literature.

Cole and VanRaden (2010) discussed several approaches for presenting data visually, including the display of chromosomal PTA. In August 2009, a publicly accessible query was added to the Animal Improvement Programs Laboratory's (Agricultural Research Service, USDA, Beltsville, MD) web site, which produced plots of chromosomal PTA for genotyped animals. However, those PTA were calculated as half the sum of the average effects of the genes carried by an individual, and suggested that animals transmit average chromosomes to their progeny rather than distinct haplotypes, which can vary dramatically. Due to growing demand from dairy farmers for genomic tools to assist in mate selection, the original query was modified to display DGV for maternal and paternal haplotypes, provide greater control over plot options, and include new traits. The purpose of this paper is to describe those new tools, and to discuss challenges related to the application of haplotypes to mating decisions.

MATERIALS AND METHODS

Overview

The flow of data through the US national genomic evaluation system is presented schematically in Figure 1. Inputs to the system include phenotypes, pedigrees, and genotypes that are provided by dairy farmers, data collection centers, artificial insemination firms, and breed associations. Pedigree and phenotypic data are combined using an animal model to produce traditional PTA that do not include genomic information. Independently of this process, the program findhap.f90 (VanRaden et al., 2011b) uses pedigree and genotypic data to construct haplotypes. The densemap.f90 program uses the traditional PTA and genotypes are then used to compute SNP effects, genomic PTA (**gPTA**), and DGV for paternal and maternal haplotypes. Finally, the DGV for paternal and maternal chromosomes are combined with animal identification and stored in a relational database following each official genetic evaluation. The query is implemented using a web application development framework and will run in all major web browsers.

Modifications to the Database

The DGV and associated information are stored in a relational database table (DB2 V9.5, IBM Corporation, Armonk, NY) for access by web query. As part of the routine data processing for each official genetic evaluation, an SAS program (SAS 9.3, SAS Institute Inc., Cary, NC) is used to drop the existing table, create an empty table in the database, and load the data.

This ensures that only current data are included in the query, and that those data reflect the most recent information available for each animal. The addition of the haplotype data and inclusion of results for 2 more traits (heifer and cow conception rates) increased the size of the table by approximately a factor of 3.

The database table includes 3 rows for every genotyped animal-trait combination: 1 row each for paternal and maternal haplotypes, and a third row for the sum of the haplotypes. The overall DGV may be computed as the sum of the parental DGV, but storing it as a separate row simplifies the display logic for some queries. Rows are indexed by animal ID, trait, and DGV type (paternal DGV, maternal DGV, or total DGV).
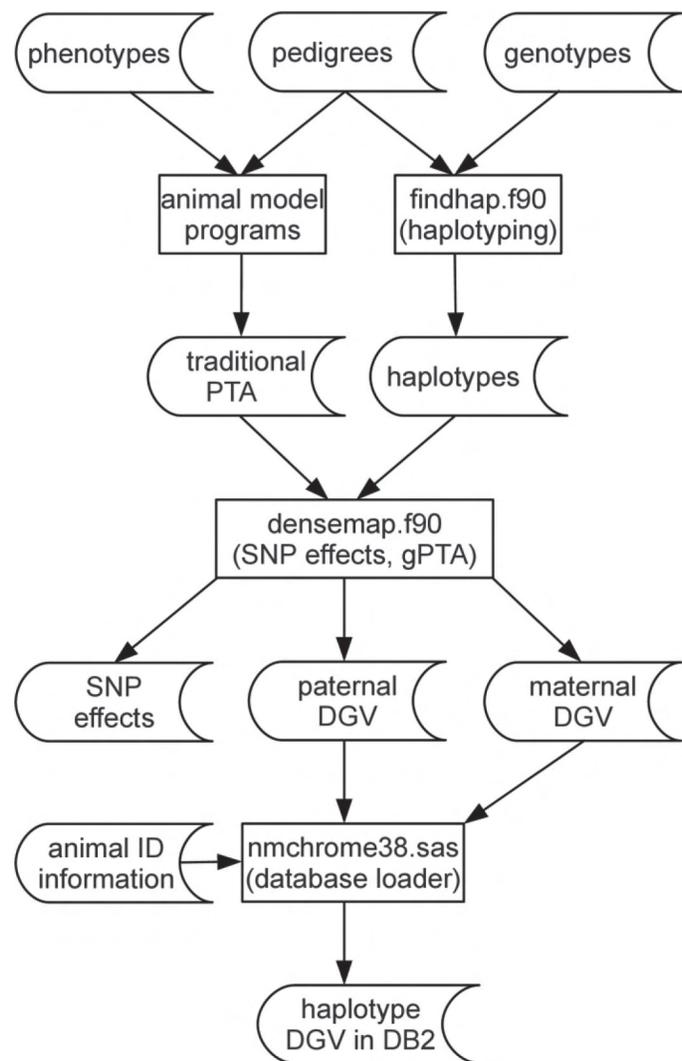

Figure 1. Flow of data and programs used to calculate direct genomic values (DGV) using traditional PTA and SNP. The resulting DGV for paternal and maternal haplotypes for each trait are combined with animal identification and stored in a table in a relational database.

The table was loaded following the completion of the August 2012 genetic evaluation, and currently includes 9,108,648 rows, representing 220,671 animals [Brown Swiss (**BS**): 6,129 bulls, 1,238 cows; Holsteins (**HO**): 40,990 bulls, 148,386 cows; Jersey (**JE**): 4,462 bulls, 19,466 cows). Whereas performance and storage are currently not limiting factors, and do not adversely affect query performance, they could become a problem as the number of genotyped animals increases and new traits are added to the evaluation system. The number of genotyped animals is rapidly growing, and 2,006,409 rows were added to the table between the April 2012 and August 2012 evaluations—an increase of 28%. The number of traits varies by breed, as shown in Table 1, ranging from 11 to 15 traits for each animal. Each HO animal has 45 rows in the table (15 traits × 3 DGV types), whereas BS and JE animals have 39 and 33 rows, respectively.

Calculation and Summary of Haplotype Effects

Haplotypes for all genotyped animals were formed using version 2 of the Fortran program findhap.f90 (VanRaden et al., 2011b). The findhap.f90 program divides chromosomes into smaller (typically 75-SNP) blocks, and then uses population and pedigree haplotyping strategies to compute paternal and maternal haplotypes for each chromosome segment. The chromosome segments are then combined to produce a haplotype for each chromosome. The use of variable-length segments improves haplotype quality compared with fixed-length segments, resulting in more accurate prediction of chromosomal DGV. In this paper, the term haplotype refers to haploid chromosomes, and should not be confused with the shorter chromosome segments used by findhap.f90.

The program densemap.f90, which is used to estimate allele substitution effects and compute gPTA, was modified to optionally compute DGV for paternal and maternal chromosomes by summing allele substitution effects for each trait. The result of these calculations is a gPTA for each chromosome (paternal and maternal) for each animal and trait, which are written to plaintext files that are loaded into a relational database for access by the online query system. Allele substitution effects were estimated using an infinitesimal alleles model with a heavy-tailed prior, in which smaller effects are regressed further toward 0 and markers with larger effects are regressed less to account for a nonnormal distribution of marker effects (VanRaden, 2008; VanRaden et al., 2011a). 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). When h is 1, marker effects are normally distributed with no additional weight in the tails, and variance in the tails grows with increasing values of h. Values for h of 1.2 for HO and JE, and 1.1 for BS, were used in this study. The additional calculations require minimal computational resources, but the output files are large. For example, the August 2012 Holstein files were approximately 2 gigabytes (3.2 million lines) of plaintext.

Haplotype DGV Query

The haplotype DGV for all genotyped animals are available using a publicly accessible query on the Animal Improvement Programs Laboratory web site (http://aipl.arsusda.gov/CF-queries/Bull_Chromosomal_EBV/bull_chromosomal_ebv.cfm). The chromosomal

Table 1. Traits for which direct genomic values of maternal and paternal haplotypes are available[1] by breed

Trait	Description	Brown Swiss	Holstein	Jersey
Milk yield	Milk production	X	X	X
Fat yield	Fat production	X	X	X
Protein yield	Protein production	X	X	X
Fat percent	Fat concentration in milk	X	X	X
Protein percent	Protein concentration in milk	X	X	X
Productive life	Cow longevity	X	X	X
Somatic cell score	Udder health	X	X	X
Lifetime net merit	Lifetime profitability	X	X	X
Daughter pregnancy rate	Cow fertility	X	X	X
Sire calving ease	Direct dystocia	X	X	
Daughter calving ease	Maternal dystocia	X	X	
Sire stillbirth	Direct stillbirth		X	
Daughter stillbirth	Maternal stillbirth		X	
Heifer conception rate	Ability to conceive as a heifer	X	X	X
Cow conception rate	Ability to conceive as a cow	X	X	X

[1]An empty cell indicates that a particular trait is not evaluated for that breed.

Figure 2. The initial query screen showing the animal ID entry field and the pulldown menus of plot controls. Color version available in the online PDF.

PTA query uses ColdFusion MX 7.0.1 (Adobe Inc., San Jose, CA) to accept user input and return results from the database. The only information the user must provide is a 17-byte animal ID number. Pulldown menus control other aspects of the plot, including the trait(s) to be shown, the values to be plotted, and the manner in which the axes for the graph should be displayed (Figure 2).

Summary Statistics

Summary statistics for paternal and maternal haplotypes in each breed were computed using the DGV stored in the database. The expected proportion of additive genetic variance (**EXPV**) for lifetime net merit (**NM$**) was computed as

$$EXPV_c = \frac{length_c}{length_g},$$

where $EXPV_c$ is the expected proportion of variance associated with chromosome c, and $length_c$ and $length_g$ are the lengths of the cth chromosome and the genome, respectively, in base pairs.

The actual proportion of additive genetic variance (**ACTV**) for NM$ associated with each chromosome in the 3 breeds was calculated similarly to Cole et al. (2009b):

$$ACTV_{bc} = \frac{\sum_{i=1}^{n_c} 2p_i\left(1 - p_i\right)\alpha_i^2}{\sum_{g=1}^{n_g} 2p_g\left(1 - p_g\right)\alpha_g^2 \times EXPV_c},$$

where $ACTV_{bc}$ = the ratio of the observed to the expected sums of allele substitution effects for the cth

chromosome in the bth breed; n_c = the number of alleles on chromosome c; p_i = the base population frequency of the ith allele on the cth chromosome for breed b; α_i = the effect of the ith allele on the cth chromosome for breed b; n_g = the number of alleles in the genome; p_g = the base population frequency of the gth allele in the genome for breed b; and α_g = the effect of the gth allele in the genome for breed b. The denominator differs from Cole et al. (2009b) in that the paternal and maternal chromosomes are used in the current study, rather than a single chromosome that is the average of the parental chromosomes. The allele effects are estimated within breed and are equivalent to average allele substitution effects (Falconer and MacKay, 1996), and the model is assumed to be purely additive (dominance and epistatic variances are 0).

Visualization Tools

Figure 1 was drawn with LibreOffice Draw 3.5.3.2 (The Document Foundation, 2012) running on Kubuntu Linux 12.04 (Canonical Ltd., 2012). Figures 2 and 5–10 are screenshots of output from ColdFusion MX 7.0.1, and were prepared with Safari 6.0.2 on a MacBook Pro running Mac OS X 10.6 (Apple Inc., Cupertino, CA). Figures 3, 4, and 11 were created using ggplot2 0.9.2 (Wickham, 2009) and R 2.15.1 (R Development Core Team, 2010) on an IBM xSeries 3850 server (IBM Corp., Armonk, NY) running Red Hat Enterprise Linux 5.0 (Red Hat Inc., Raleigh, NC).

RESULTS AND DISCUSSION

Summary of Haplotypes in the Population

Differences in least squares means of DGV for paternal and maternal chromosomes of BTA 1, 6, 14, and

18 for NM$ are shown in Table 2. Those chromosomes were chosen to represent cases including the presence and absence of mutations with large effects. No known QTL affecting NM$ on BTA 1 have been noted; the *ABCG2* gene (Cohen-Zinder et al., 2006) affecting protein percentage in Holsteins is located on BTA 6; the *DGAT1* gene (Grisart et al., 2004) affecting fat percentage is located on BTA 14; and a QTL affecting calving traits, conformation, NM$, and longevity in Holsteins (Cole et al., 2009b) is found on BTA 18. Differences were significant in all but one case: paternal and maternal chromosomes did not differ ($P = 0.1497$) for BTA 6 in BS. With the exception of BTA 18 in BS, paternal chromosomes had higher average DGV than maternal chromosomes, which is consistent with most selection pressure being applied through the sire side of the pedigree, although intensity varies among breeds. Large differences between paternal and maternal DGV for BTA 1 are expected because it is the longest chromosome. Differences were greater when QTL are present, such as BTA 6, 14, and 18 in HO.

Expected and actual proportions of additive genetic variance for NM$ explained by each chromosome in the 3 breeds are presented in Figure 3 (Supplemental Table S1, available online at http://www.journalofdairy-science.org/). Results are similar to those of Cole et al. (2009b), who reported ACTV (denoted as VWB in that paper) for more traits but fewer chromosomes. As expected, longer chromosomes generally accounted for more variance than shorter chromosomes. Differences among breeds were consistent with the segregation of known mutations of large effect. For example, BTA 18 accounted for 2.5, 7, and 2.6% of the variance in NM$ for BS, HO, and JE, respectively. With the exception of a few well-known loci, such as *DGAT1* on BTA 14 and the calving traits complex on BTA 18 in HO, NM$ is not under the control of mutations with large effects and most SNP explain only a small amount of variance for NM$, producing results which are consistent with previous findings (Cole et al., 2009b; Hayes et al., 2010).

The distribution of the number of paternal and maternal chromosomes with positive DGV for NM$ are shown in Figure 4. Left-skewed distributions are expected because elite animals are more likely than average animals to be genotyped, and selection intensity is greater through the sires-of-bulls and dams-of-bulls paths (Van Tassell and Van Vleck, 1991). Results were similar for HO and JE, but the distributions were less skewed for BS, probably because far fewer genotyped BS animals were represented in the data. The results for BS in Figure 4 are consistent with those shown in Table 2 (e.g., no paternal-maternal differences on BTA6 were noted). In general, genotyped animals have more chromosomes with positive (desirable) DGV than negative (undesirable) DGV. Modes were slightly higher for paternally than maternally derived haplotypes in HO and BS (22 vs. 20 and 22 vs. 21, respectively), and slightly lower in BS (17 vs. 19). The abundance of haplotypes with desirable DGV is expected in a population under selection, and the skewness in the HO and JE populations may be attributable to higher rates of genotyping among elite animals and larger population sizes. The BS distributions are much less skewed and, because their population is much smaller in the United States than HO and JE, the genotyped BS animals may more closely represent the whole population rather than the genetically elite males and females. In smaller populations, fewer opportunities to produce animals with extremely high genetic merit exist because fewer samples are drawn from the pool of available gametes. If the cost of genotyping continues to decrease, and more dairy producers begin to make culling decisions based on those results (De Vries et al., 2011; Lawlor, 2011), the skewness of the HO and JE distributions may decrease. One HO animal received 30 chromosomes with positive DGV from its sire, but no animal in any breed received only chromosomes with desirable DGV from both parents.

Chromosomal DGV Query

Operating the Query. The query result screen (Figure 5) includes identification information that is linked to an animal's official genetic evaluation, the of-

Table 2. Differences in LSM of direct genomic values for paternal and maternal haplotypes for lifetime net merit in Brown Swiss (BS), Holstein (HO), and Jersey (JE) cattle for chromosomes 1, 6, 14, and 18

Chromosome[1]	BS		HO		JE	
	Difference	*P*-value	Difference	*P*-value	Difference	*P*-value
1	2.93	<0.0001	2.69	<0.0001	3.05	<0.0001
6	0.60	0.1497	3.99	<0.0001	2.12	<0.0001
14	1.21	0.0167	2.62	<0.0001	1.92	<0.0001
18	−1.14	0.0005	6.13	<0.0001	1.07	<0.0001

[1]No known QTL affecting lifetime net merit on chromosome 1 were noted. The *ABCG2* gene affecting protein percentage is located on chromosome 6; the *DGAT1* gene affecting fat percentage in BS and HO is located on chromosome 14; and a QTL affecting calving traits, conformation, lifetime net merit, and longevity in HO is found on chromosome 18.
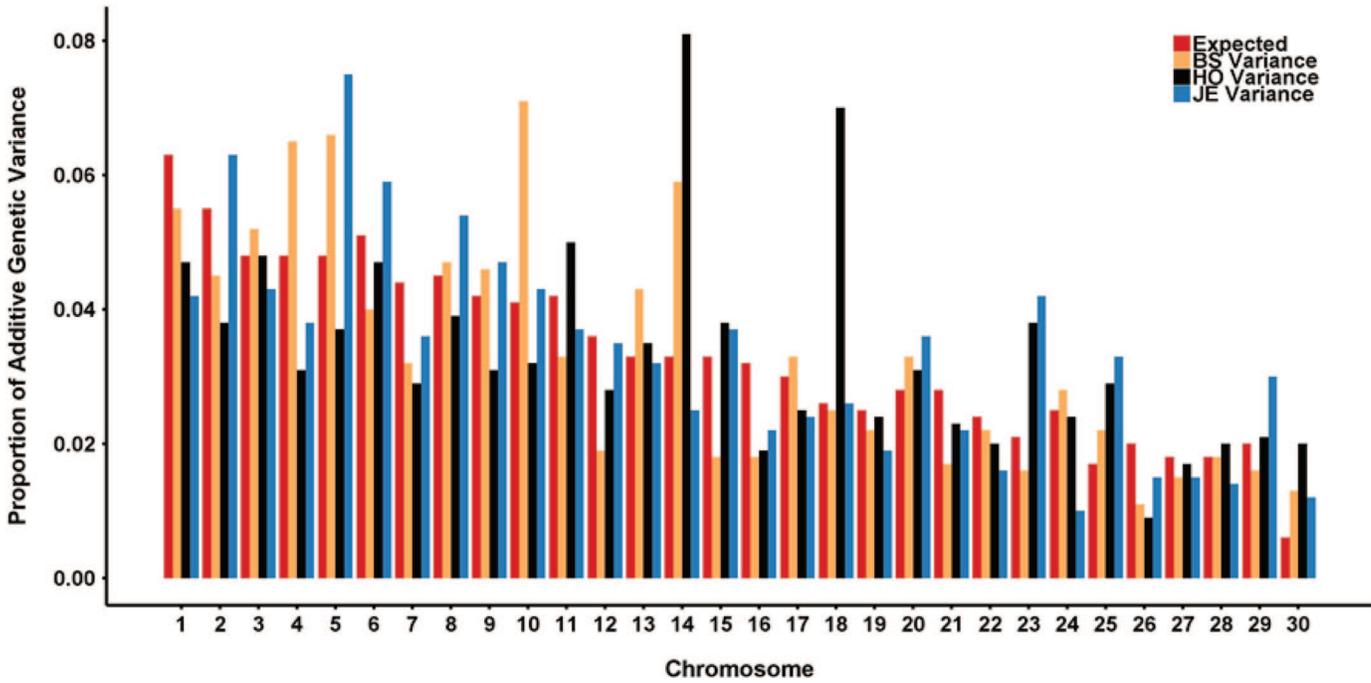
Figure 3. Expected (red bars) and actual proportion of additive genetic variance for lifetime net merit explained by genetic markers on each chromosome in Brown Swiss (BS; orange bars), Holstein (HO; black bars), and Jersey (JE; blue bars) cattle. Color version available in the online PDF.
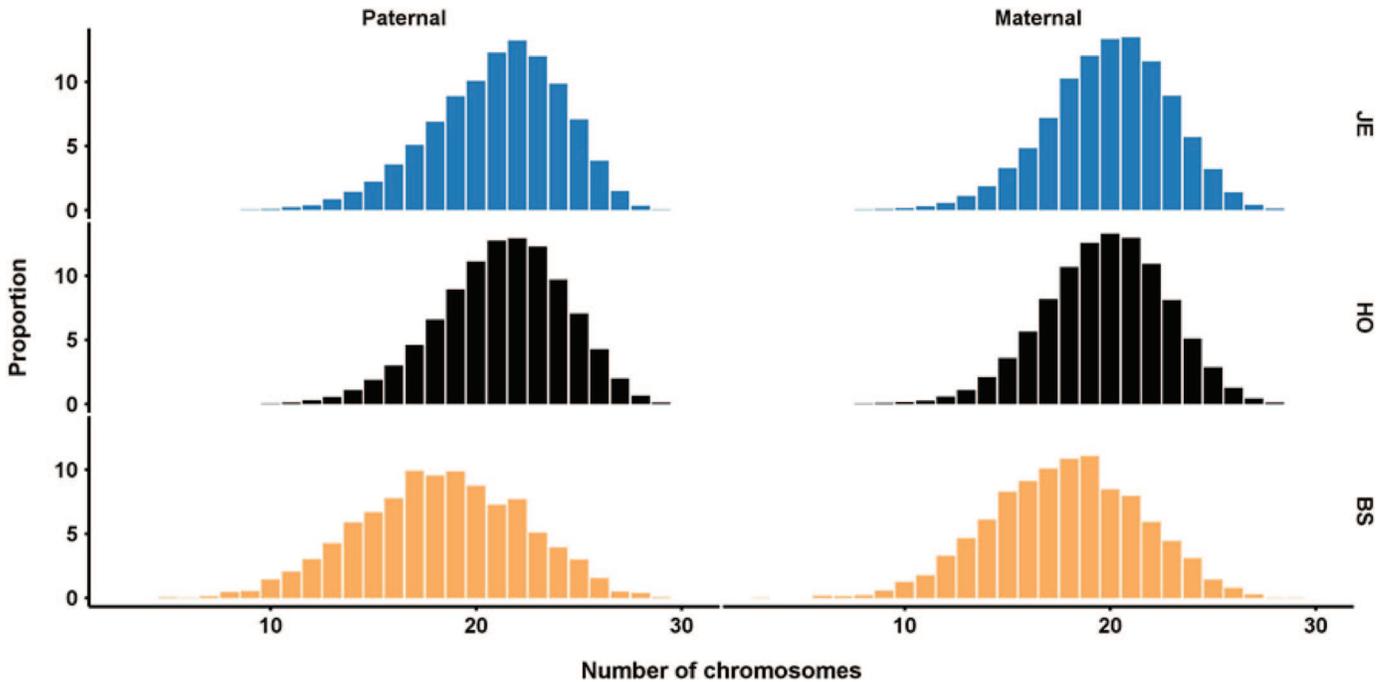

Figure 4. The distribution of the number of paternal and maternal chromosomes having positive direct genomic values for lifetime net merit in Brown Swiss (BS; orange bars), Holsteins (HO; black bars), and Jerseys (JE; blue bars). Color version available in the online PDF.

ficial PTA, a prominent link to a document explaining how to use the query and interpret its output (Cole and Null, 2012), a plot showing the paternal and maternal DGV for each of the 30 chromosomes in the bovine genome, and information about the calculation of the DGV. If both chromosomes have a positive DGV, their sum can be read as the y-axis value associated with the upper limit of the top bar. If both are negative, the sum of the DGV can be read as the y-axis value corresponding to the lower limit of the bottom bar. If one DGV is positive and one negative, the y-axis value corresponding to their sum is the midway point between the upper limit of the top bar and the lower limit of the bottom bar. The DGV are expressed in the same units as the PTA for the trait (e.g., pounds for milk and dollars for NM$).

The trait menu allows the user to select the trait in which they are interested from a list. Results will be returned for NM$ (Cole et al., 2009a) unless a different trait is selected. The all traits option returns a table of bar graphs including results for all traits available for the breed to which the animal belongs (Figure 6).

Clicking on the title for a trait opens a new browser window containing the relevant single-trait view. The trait menu includes all available traits regardless of breed, so it is possible to select a trait that is not evaluated for a given breed (e.g., stillbirth in JE). In those cases, an error message is returned indicating those results are not available.

Different view options are available in the query: haplotype DGV and summed DGV. The default view is haplotype DGV, which produces a chart showing the DGV of the maternal and paternal haplotypes for each chromosome. The height of the bar for each chromosome is determined by adding up the effects of all of the SNP in each haplotype. The other view menu option is summed DGV, which produces a chart with the sum of all the marker effects on both the maternal and paternal chromosomes (Figure 7). The summed DGV plot indicates how good or poor an animal is for each chromosome, but does not describe the value of individual haplotypes.

The scale menu option controls the range of values used in setting plot axes. The default value of indi-
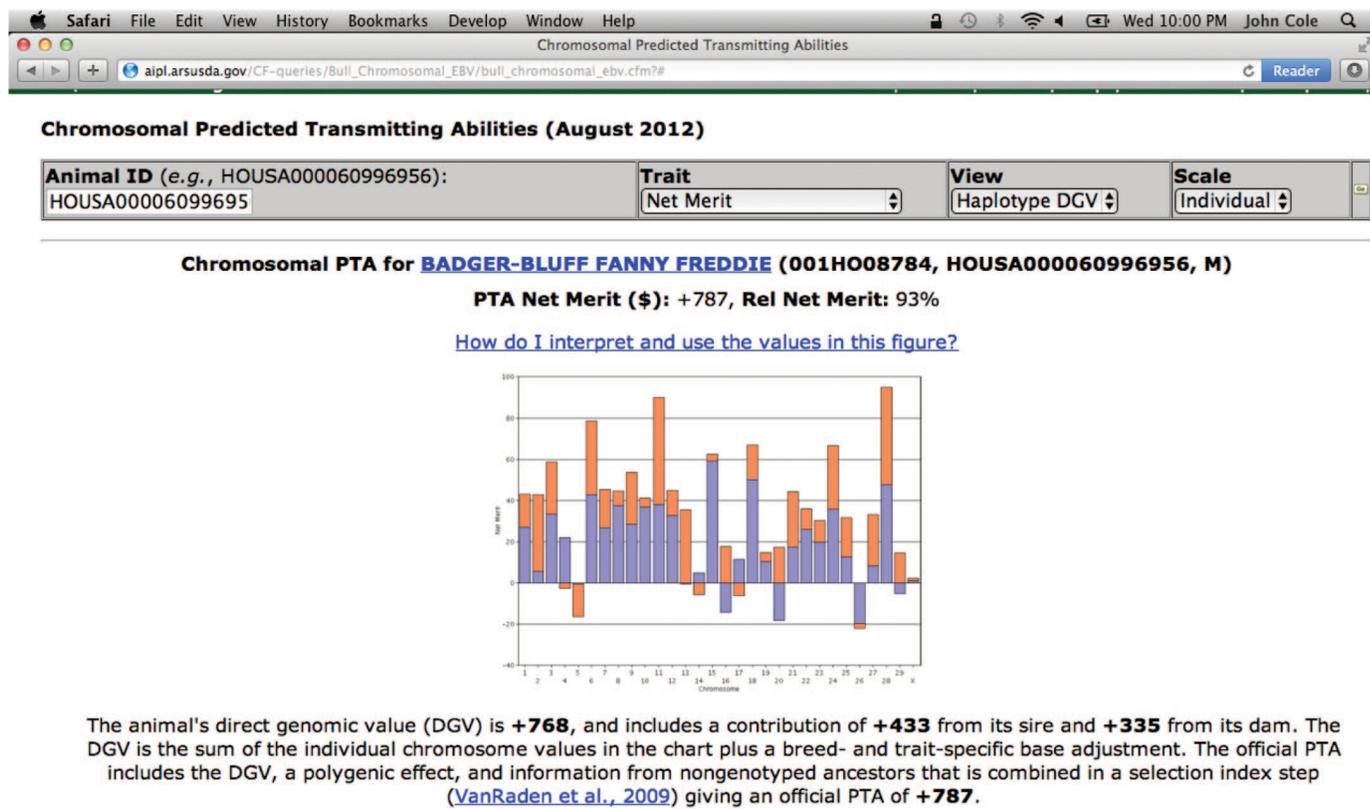

Figure 5. The lifetime net merit view of haplotype direct genomic values (DGV) for the Holstein bull BADGER-BLUFF FANNY FREDDIE (HOUSA000060996956). Color version available in the online PDF.
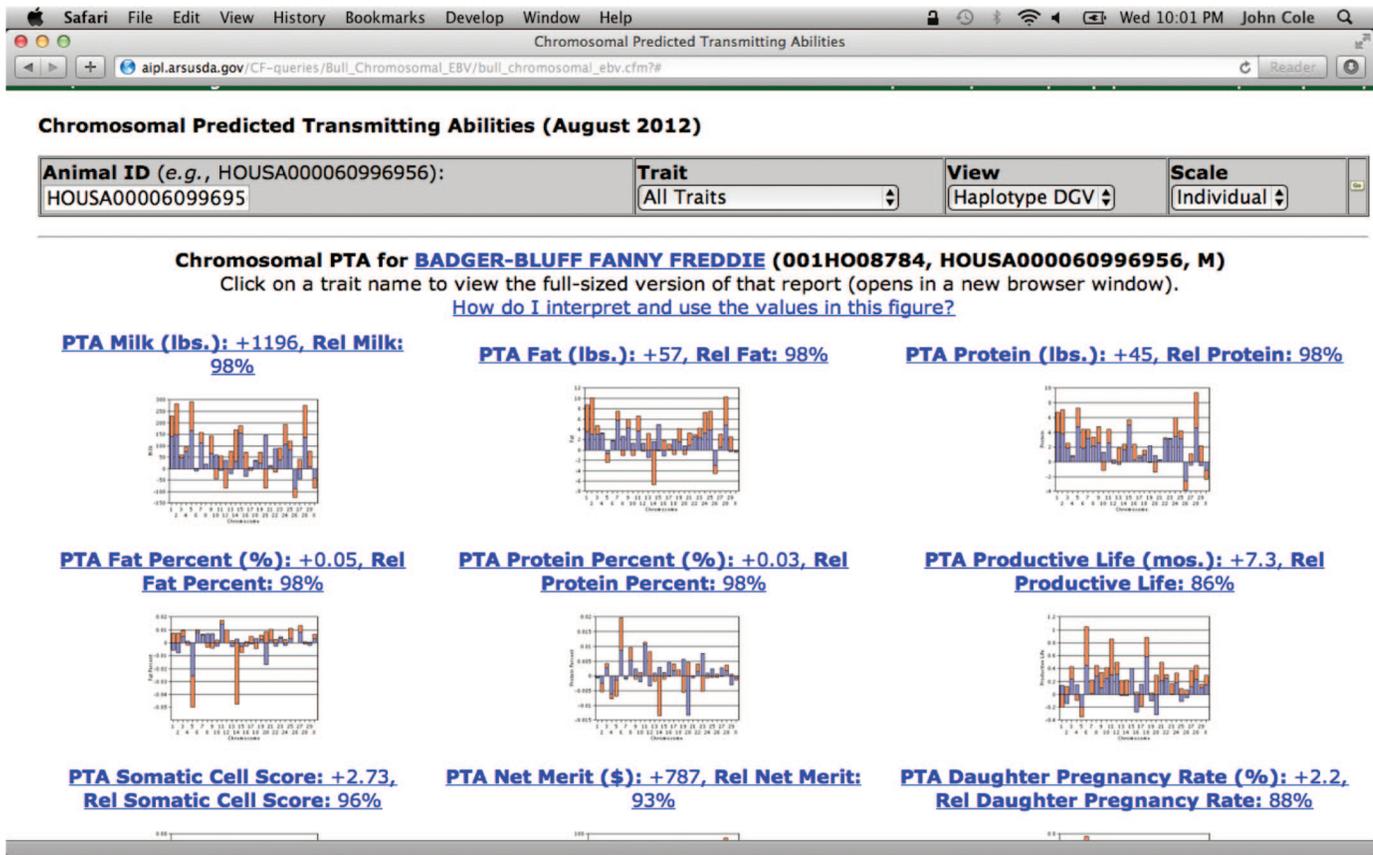
Figure 6. The All Traits view of haplotype direct genomic values (DGV) for the Holstein bull BADGER-BLUFF FANNY FREDDIE (HOUSA000060996956). Color version available in the online PDF.

vidual scale sets the largest and smallest values on the y-axis using only information from the animal and trait specified in the query. The resulting plot uses all of the available area for presenting the data. However, comparing query results for different animals is difficult with the individual scale because each animal's plot is on a slightly different scale. The shared scale option results in the use of a common y-axis scale for a given trait. This makes it easier to compare one animal with another because 2 bars of the same height represent the same value when shared scales are used. The original query used a complex SQL statement that included several UNION operations to determine the y-axis scales, and was very slow. When the script used to create and populate the database table was revised, the minimum and maximum values were calculated for each trait and stored in the database. The minimum and maximum values for the shared scales are now retrieved when fetching the haplotype or chromosomal DGV, dramatically reducing the time needed to return query results.

Relationships Among Values in the Query Results. When the query page is loaded, the ID for

the HO bull Badger-Bluff Fanny Freddie (Freddie; HOUSA000060996956) is provided as an example for users. The query result screen for Freddie (Figure 5) includes several different values, such as the DGV for each chromosome, paternal and maternal DGV, and his official gPTA. Freddie's official PTA for NM$ is +787 and has a reliability of 93%. The official PTA includes the DGV, a polygenic effect to account for genetic variance not explained by the SNP, and information from nongenotyped ancestors that is combined in a selection index step (VanRaden et al., 2009). In this case, the DGV is +768, which includes a contribution of +433 from the paternal chromosomes and +335 from the maternal chromosomes; the paternal and maternal chromosome values, which are displayed when the mouse pointer passes over a bar in the chart, sum to +1,060. However, this total does not include the intercept for NM$ in HO, which was −292 in August 2012, or the SNP effects from the pseudoautosomal region of the X chromosome, which is −30 for Freddie. When those values are added to the DGV the result is +765, which differs from the official DGV of +768 by
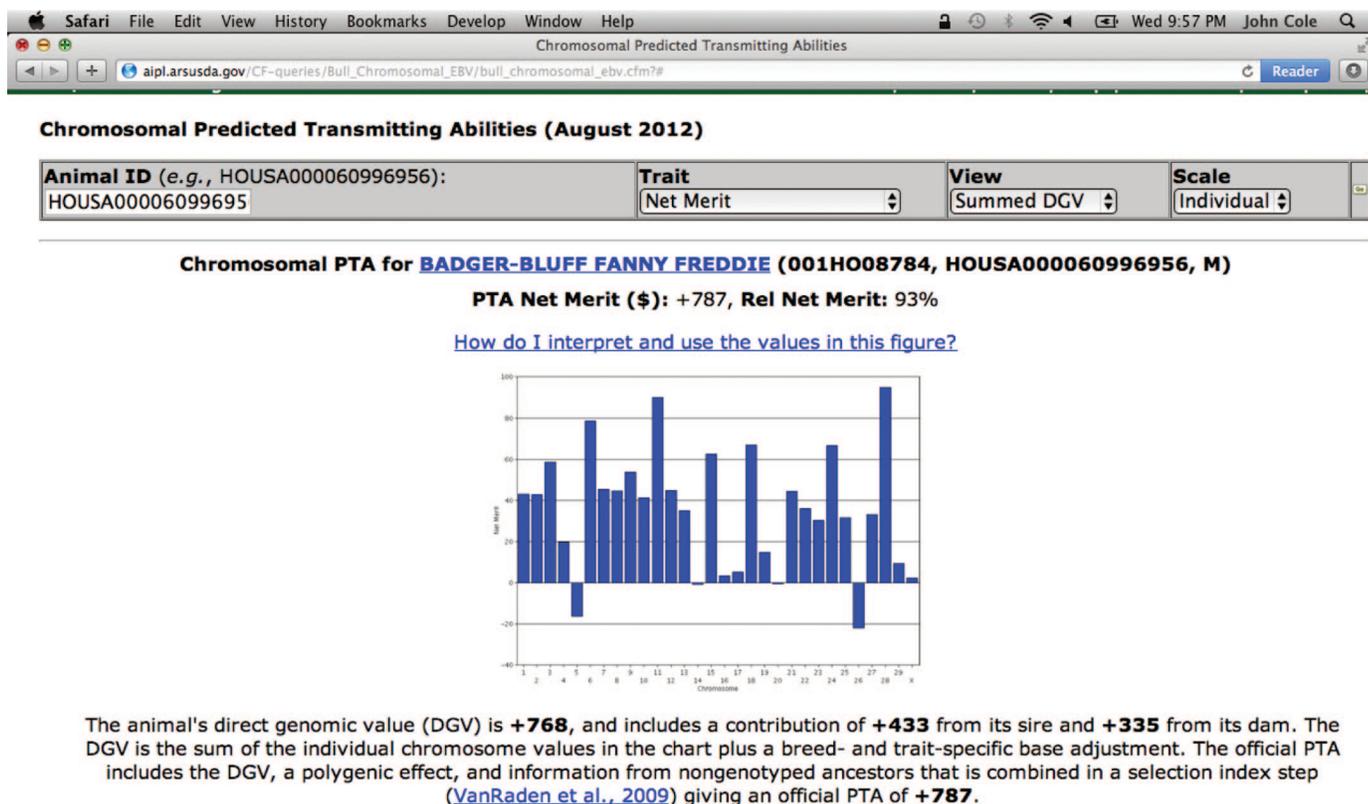
Figure 7. The lifetime net merit view of summed direct genomic values (DGV) for the Holstein bull BADGER-BLUFF FANNY FREDDIE (HOUSA000060996956). Color version available in the online PDF.

an amount attributable to rounding. The DGV for the pseudoautosomal region are not currently displayed by the query because of abnormally large estimates of SNP effects in this region for some traits, which indicates a discrepancy between evaluations of males and females (Wiggans et al., 2011).

Use of Haplotype DGV to Illustrate Quantitative Genetic Principles

Crossing-Over and Independent Assortment. The processes of crossing-over during prophase I of meiosis, followed by independent assortment during metaphase I of meiosis (Snustad and Simmons, 2008), often are assumed to be well-understood. However, the query provides a way to clearly demonstrate these principles as they occur in real animals. Consider the 10 genomic-tested offspring resulting from the mating of the bull LADYS-MANOR PL SHAMROCK-ET (HOUSA000068977120) to the cow DE-SU OMAN 6121-ET (HOUSA000061681442). The dam's average DGV for BTA 1 is +31, which is similar to the calves'

average maternal chromosome DGV of +35, which supports the principle that a parent transmits the average value of its 2 chromosomes to its progeny. The maternal DGV for each of 10 calves also are slightly different, indicating that at least one crossover event occurred for each chromosome, as expected.

At present, variation of chromosomal DGV among individuals is a novel area of research. The standard deviations of the paternal and maternal DGV of BTA 1 for these 10 offspring were 14.1 and 10.6, respectively. Mean progeny DGV for BTA 1 differed from parental mean DGV by an average of 7.3 and 3.3 for paternally and maternally inherited chromosomes, respectively. These results underscore the difficulty in predicting the results of an individual mating, as well as the challenge of trying to assemble all of the top chromosomes in a single individual.

Genetic Trend in Populations Under Selection. In populations undergoing selection, the frequency of alleles favorably associated with the selection objective will increase over time and the population mean will increase as well (Lush, 1945). This should lead
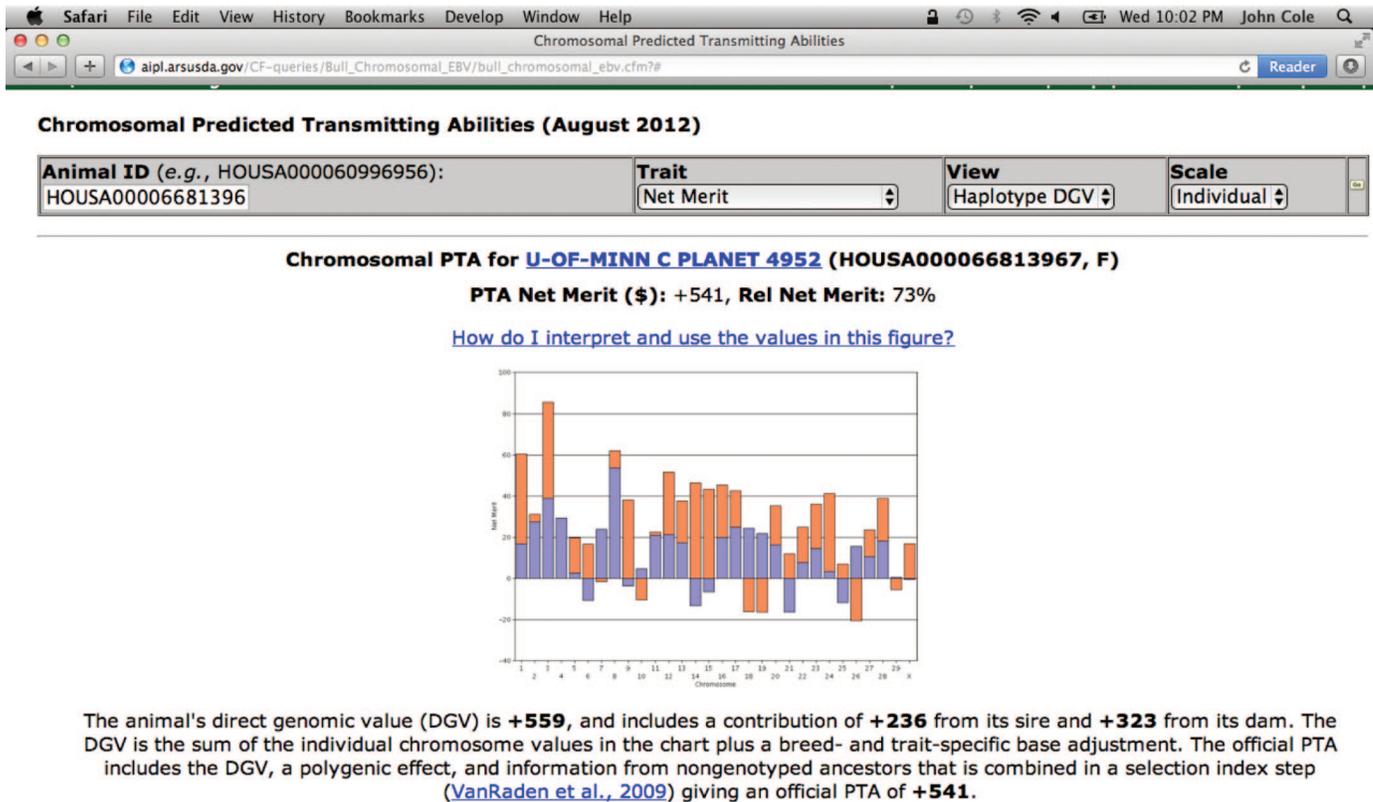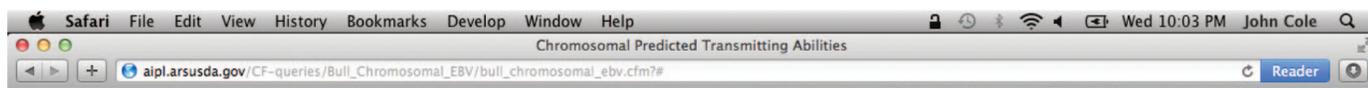
Figure 8. The lifetime net merit view of haplotype direct genomic values (DGV) for the University of Minnesota selection line Holstein cow U-OF-MINN C PLANET 4952 (HOUSA000066813967). Color version available in the online PDF.

to an increase over time in the frequency of chromosomes with positive DGV and a decrease in those with negative DGV; this can be illustrated by comparing the number of chromosomes with positive and negative DGV. Consider the case of 2 cows from the University of Minnesota's selection and control lines (Young, 1977; Hansen, 2000). Cows in the selection line were mated to bulls with high genetic merit for production, whereas control line cows were mated to sires with average production for the 1960s. The chromosomes in the control line should be representative of the population in 1964, and comparison with those in the selection line will illustrate how the population has changed in response to almost 50 yr (10 generations) of selection for increased yield.

The cow U-OF-MINN C PLANET 4952 (HOUSA000066813967) was born in 2009 and has a gPTA for NM\$ of +541 (Figure 8), whereas the cow U-OF-MINN W CADFARCH JANUS (HOUSA000015334728) was born in 1993 and has a gPTA for NM\$ of −1,343 (Figure 9). The former cow is from the selection line and has desirable (positive) DGV for 47 of 60 individual chromosomes. The control line cow is a mirror im-

age of this, with undesirable (negative) DGV for 45 of 60 chromosomes. These examples demonstrate that selection in the US dairy population has resulted in improvement across the entire genome. The distribution of the number of positive chromosomes in Figure 4 shows that more positive DGV are paternal in origin, which reflects the greater selection pressure achieved through the sires of sons and sires of cows' paths (Van Tassell and Van Vleck, 1991). Traits uncorrelated with the selection objective are expected to have genetic trend of 0, average chromosomal DGV of 0, and normal distributions of chromosome with positive DGV in both bulls and cows.

Alleles with Small Versus Large Effects. Whereas attention often is paid to alleles with large effects (Cole et al., 2009b), most traits of economic importance in dairy cattle are controlled by a large number of alleles with small effects. The prominent Holstein bull O-BEE MANFRED JUSTICE-ET (O-Man; HOUSA000122358313; Figure 10) has a large negative-summed DGV for fat percentage on BTA 14 of −0.02, but an overall DGV of +0.11 and a gPTA of +0.13. Whereas O-Man is heterozygous for the undesir-
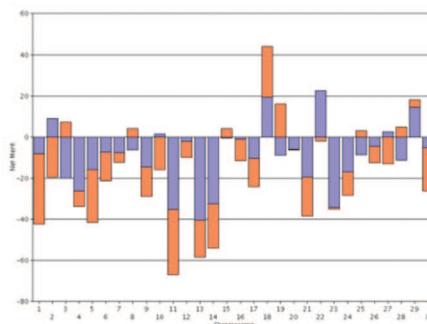
Figure 9. The lifetime net merit view of haplotype direct genomic values (DGV) for the University of Minnesota control line Holstein cow U-OF-MINN W CADFARCH JANUS (HOUSA000015334728). Color version available in the online PDF.

able allele of the *DGAT1* gene (Grisart et al., 2004), which has a large, negative effect on fat percentage, he also has many chromosomes with favorable (positive) DGV. This shows that alleles with large effects can be important, but they are not necessarily more important than a large number of alleles with small effects. The difference between bulls and cows for DGV of NM$ shown in Table 2 is evidence that selection for higher lifetime profitability, and differences in intensity among the 4 paths of selection, has resulted in differences between the sexes in the frequency of the desirable form of *DGAT1*. These results also are consistent with the greater-than-expected proportion of variance in NM$ explained by alleles on BTA 14, relative to its length. Genetically elite animals receive a superior sample of genes from both parents, rather than desirable alleles for a few alleles with large effects.

Use of Haplotype DGV in Mating Programs

The best (above x-axis; green bars) and worst (below x-axis; red bars) haplotypes in the HO breed as of Au-

gust 2012 are shown in Figure 11. The most desirable HO haplotype is on BTA 14, where the *DGAT1* gene is located (Grisart et al., 2004). The least desirable HO haplotype is on BTA 18, which carries a QTL that is associated with NM$, calving traits, and conformation (Cole et al., 2009b). The results for the best haplotypes are similar to those presented in Cole and VanRaden (2011), with the notable exception of the X chromosome, which is due to corrections now used to account for bias in cow evaluations (Wiggans et al., 2011). The values plotted in Figure 11 are the DGV of the best individual haplotypes for each chromosome and are interpretable in a manner analogous to transmitting abilities, whereas the results shown in Figure 5 in Cole and VanRaden (2011) are on a breeding value basis. The variation between the best and the worst haplotypes in the population indicates that substantial genetic gain is possible using haplotypes already present in the population.

To produce bull dams or young sires, a minimum value for NM$ PTA should be established and mates selected from animals with PTA at or above that threshold with
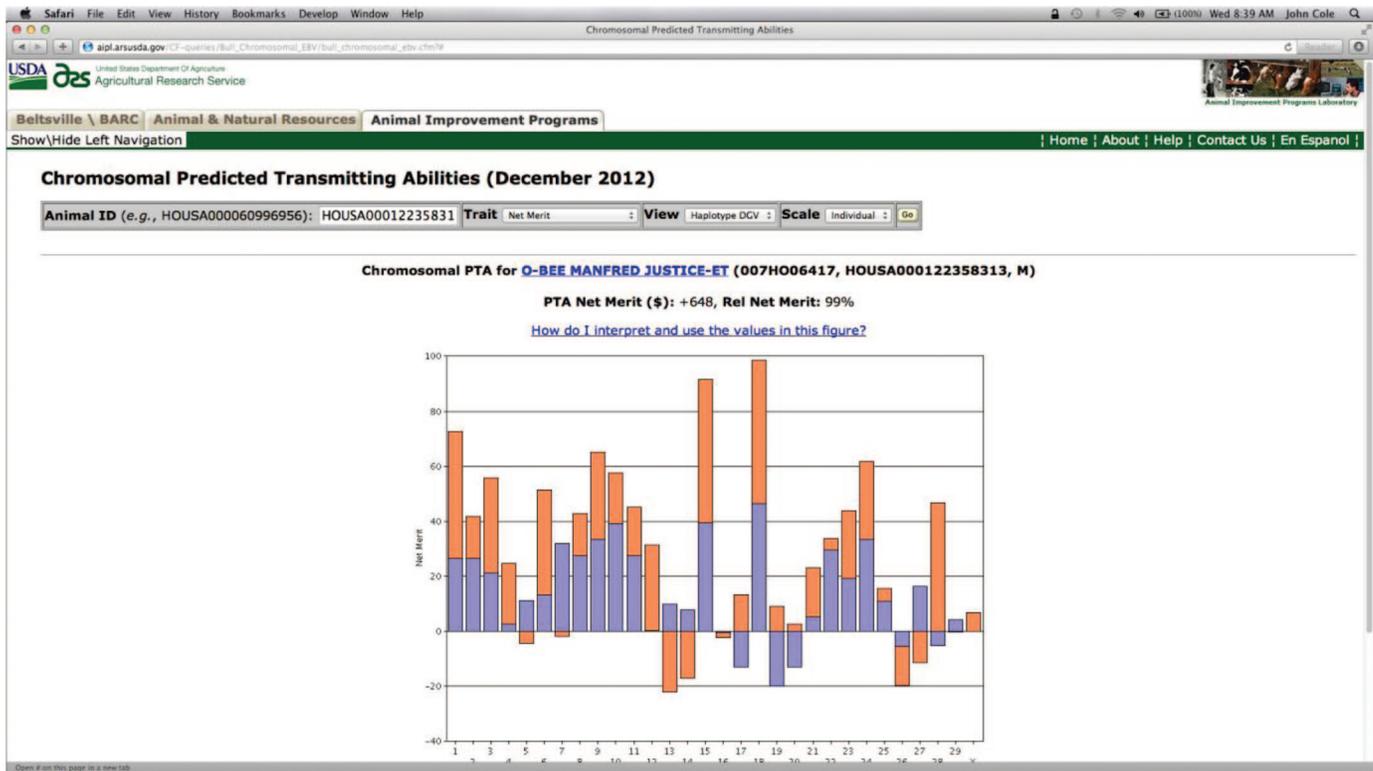
Figure 10. The lifetime net merit view of haplotype direct genomic values (DGV) for the Holstein bull O-BEE MANFRED JUSTICE (HOUSA000122358313). Color version available in the online PDF.

high heterogeneity between their paternal and maternal DGV, which may increase the probability of producing an extreme outlier. Preliminary research (Cole, 2012) suggests that identifying animals with large differences between chromosomes across the genome is easier than finding animals with very high DGV for all (or most) chromosomes. However, crossing-over tends to equalize Mendelian sampling effects by swapping material between chromosomes that are very different from one another. A potential mate may have large differences between its paternal and maternal chromosomes, but those differences may be averaged out due to crossing-over.

If the selection objective is a uniform crop of calves for producing milk or milk solids, a minimum value for NM$ PTA should be established and mates selected from the animals with PTA at or above that threshold and having low heterogeneity between the paternal and maternal chromosomes. Animals with low heterogeneity between the paternal and maternal chromosomes have chromosomes that are similar, which will result in similar outcomes regardless of which chromosomes the offspring receives. Selection candidates should not be eliminated for carrying some chromosomes with unfavorable DGV. No animal in the national dairy database had only positive DGV for all chromosomes.

In most cases, genomic inbreeding should be minimized for individual matings to avoid rapid increases in homozygosity. Such decisions require that breeders know if mates have the same haplotypes, in addition to their DGV. Cole and VanRaden (2011) found that adjustments for future inbreeding had little effect on selection limits and DGV for lifetime economic merit, but those results may not accurately reflect actual inbreeding depression at extremely high levels of homozygosity. More recently, Kemper et al. (2012) showed by simulation that selection on gPTA for economic merit with a constraint on coancestry can result in greater short-term genetic responses than a strategy pairing the 2 best ancestral haplotypes, although long-term selection responses are slightly greater under some conditions using the latter strategy. Recent simulation results using haplotyped US JE cattle found correlations of genomic relationships among mates and mean DGV of offspring of 0.09, and correlations of 0.11 with the standard deviation of DGV (J. B. Cole, unpublished data). This result is counterintuitive because we expect the variance among offspring to decrease as inbreeding increases. The best animals in the population may be more related than the average animals and tend to have more offspring, which would increase the standard deviation among the offspring.
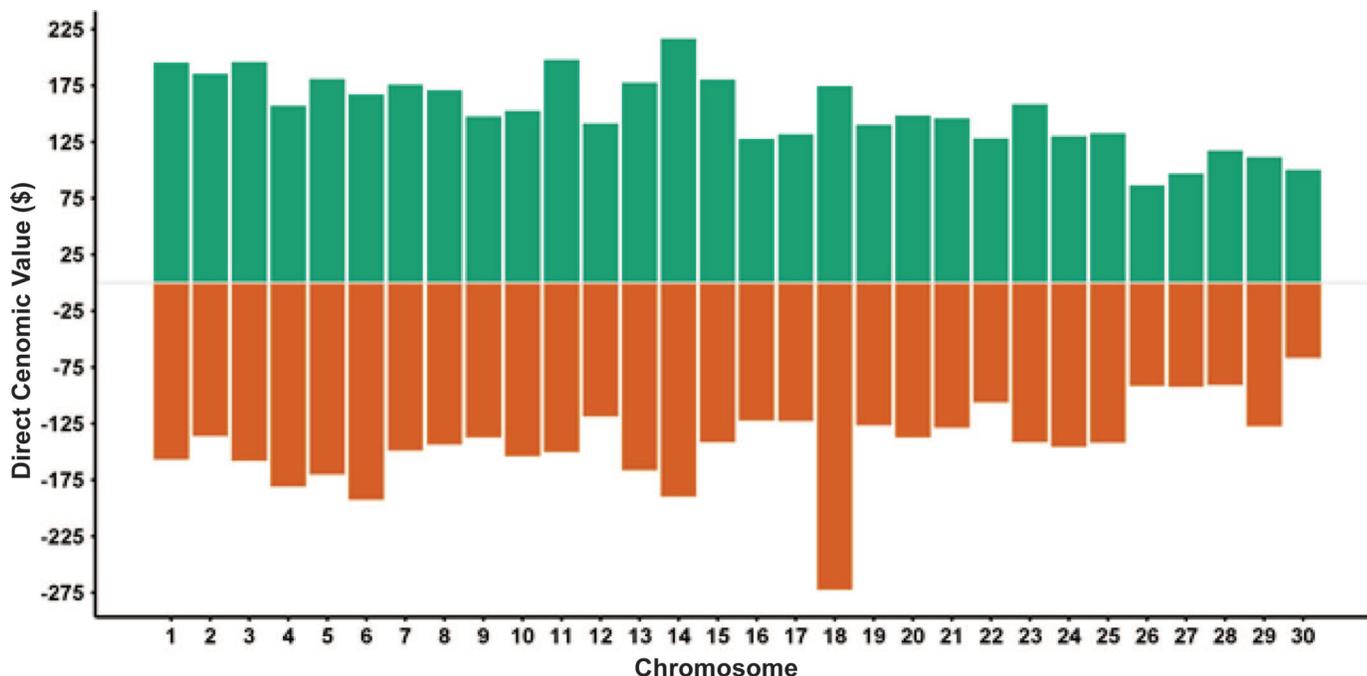
Figure 11. The best (green, above x-axis) and worst (red, below x-axis) direct genomic values for lifetime net merit in the national dairy database. Color version available in the online PDF.

Because independent assortment and crossing-over are random processes, the outcome of a particular mating cannot be guaranteed. Haplotypes may be useful to help manage risk in a genetic program by identifying animals with acceptable variation among the paternal and maternal genomic contributions, but use of haplotypes will not guarantee a desired outcome. Those data should be used carefully as part of an ongoing selection program. Research is currently underway to develop better tools for using haplotypes in mating programs, including the estimation of dominance effects using the many cow genotypes now available.

Ongoing Research

Several additional features have been proposed for this tool, including addition of conformation traits, representation of breed composition, illustration of inheritance from grandparents, visualization of the heterogeneity of paternal and maternal DGV from different animals, and comparison of multiple animals on a single screen. The easiest of these, and perhaps most valuable, is some measure of heterozygosity of chromosome pairs. In principle, it would be easy to compare SNP at each locus when the paternal and maternal DGV are computed for each chromosome and compute SNP heterozygosity. The resulting values could be stored in a new table in the database and displayed when the chromosomal DGV are plotted. It is not obvious how best to present those data, however, and additional research is needed to ensure that the query results remain easily interpretable. Discussions with the purebred cattle associations about the addition of their conformation data to the database also are underway.

REFERENCES

Bastiaansen, J. W. M., A. Coster, M. P. L. Calus, J. A. M. van Arendonk, and H. Bovenhuis. 2012. Long-term response to genomic

selection: Effects of estimation method and reference population structure for different genetic architectures. Genet. Sel. Evol. 44:3.

Börner, V., and N. Reinsch. 2012. Optimising multistage dairy cattle breeding schemes including genomic selection using decorrelated or optimum selection indices. Genet. Sel. Evol. 44:1.

Canonical Ltd. 2012. Kubuntu 12.04 LTS (Precise Pangolin). Accessed Jun. 14, 2012. http://www.kubuntu.org/.

Cohen-Zinder, M., E. Seroussi, D. M. Larkin, J. J. Loor, A. Evertsvan der Wind, J. H. Lee, J. K. Drackley, M. R. Band, A. G. Hernandez, M. Shani, H. A. Lewin, J. I. Weller, and M. Ron. 2005. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res. 15:936–944.

Cole, J. B. 2012. Applications of haplotypes in dairy farm management. 62nd Europ. Assoc. Anim. Prod. Mtg., Bratislava, Slovak Republic. 18:189. (Abstr.) 10.3920/978-90-8686-761-5.

Cole, J. B., and D. J. Null. 2012. AIPL Research Report GENOMICS2: Use of chromosomal predicted transmitting abilities. Accessed Jun. 13, 2012. http://aipl.arsusda.gov/reference/chromosomal_pta_query.html.

Cole, J. B., and P. M. VanRaden. 2010. Visualization of results from genomic evaluations. J. Dairy Sci. 93:2727–2740.

Cole, J. B., and P. M. VanRaden. 2011. Use of haplotypes to estimate Mendelian sampling effects and selection limits. J. Anim. Breed. Genet. 128:446–455.

Cole, J. B., P. M. VanRaden, J. R. O'Connell, C. P. Van Tassell, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and G. R. Wiggans. 2009b. Distribution and location of genetic effects for dairy traits. J. Dairy Sci. 92:2931–2946.

Cole, J. B., P. M. VanRaden, and Multi-State Project S-1040. 2009a. AIPL Research Report NM$4: Net merit as a measure of lifetime profit: 2010 revision. Accessed Jun. 13, 2012. http://aipl.arsusda.gov/reference/nmcalc.htm.

De Vries, A., D. T. Galligan, and J. B. Cole. 2011. Some ideas on the use and economic value of the 3K SNP genomic test for calves on dairy farms. Dairy Update 11:5–9.

Engelsma, K. A., R. F. Veerkamp, M. P. L. Calus, and J. J. Windig. 2011. Consequences for diversity when prioritizing animals for conservation with pedigree or genomic information. J. Anim. Breed. Genet. 128:473–481.

Falconer, D. S., and T. F. C. MacKay. 1996. Introduction to Quantitative Genetics. 4th ed. John Wiley and Sons, New York, NY.

Grisart, B., F. Farnir, L. Karim, N. Cambisano, J. J. Kim, A. Kvasz, M. Mni, P. Simon, J. M. Frere, W. Coppieters, and M. Georges. 2004. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. Proc. Natl. Acad. Sci. USA 101:2398–2403.

Hansen, L. B. 2000. Consequences of selection for milk yield from a geneticist's viewpoint. J. Dairy Sci. 83:1145–1150.

Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and M. E. Goddard. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: Coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genet. 6:e1001139. http://dx.doi.org/10.1371/journal.pgen.1001139.

Kemper, K. E., P. J. Bowman, J. E. Pryce, B. J. Hayes, and M. E. Goddard. 2012. Long-term selection strategies for complex traits using high-density genetic markers. J. Dairy Sci. 95:4646–4656.

Lawlor, T. J. 2011. What's your genomic testing plan? Holstein Pulse. Winter 2011:8–9.

Lush, J. L. 1945. Animal Breeding Plans. Iowa State College Press, Ames, IA.

Nishio, M., A. K. Kahi, and H. Hirooka. 2010. Optimization of mate selection based on genotypic information with overlapping generations. J. Anim. Breed. Genet. 127:34–41.

Pryce, J. E., B. J. Hayes, and M. E. Goddard. 2012. Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. J. Dairy Sci. 95:377–388.

R Development Core Team. 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Snustad, D. P., and M. J. Simmons. 2008. Principles of Genetics. 5th ed. John Wiley & Sons Inc., Hoboken, NJ.

The Document Foundation. 2012. LibreOffice. Accessed Jun. 14, 2012. http://www.libreoffice.org/.

Toro, M. A., and L. Varona. 2010. A note on mate allocation for dominance handling in genomic selection. Genet. Sel. Evol. 42:33.

Van Tassell, C. P., and L. D. Van Vleck. 1991. Estimates of genetic selection differentials and generation intervals for four paths of selection. J. Dairy Sci. 74:1078–1086.

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

VanRaden, P. M., J. R. O'Connell, G. R. Wiggans, and K. A. Weigel. 2011b. Genomic evaluations with many more genotypes. Genet. Sel. Evol. 43:10.

VanRaden, P. M., K. M. Olson, G. R. Wiggans, J. B. Cole, and M. E. Tooker. 2011a. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. J. Dairy Sci. 94:5673–5682.

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

Wickham, H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer, New York, NY.

Wiggans, G. R., T. A. Cooper, P. M. VanRaden, and J. B. Cole. 2011. Technical note: Adjustment of traditional cow evaluations to improve accuracy of genomic predictions. J. Dairy Sci. 94:6188–6193.

Young, C. W. 1977. Review of regional project NC-2. J. Dairy Sci. 60:493–498.