



## Visualization of results from genomic evaluations

J. B. Cole<sup>1</sup> and P. M. VanRaden

Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

### ABSTRACT

Genomic predictions of estimated breeding values (EBV) for dairy cattle include effects of tens of thousands of markers distributed over 30 chromosomes for many traits. There are so many numbers that data are difficult to compare, levels of detail are obscured, and data cannot easily be tabulated. Well-designed graphics can present more information in a smaller area than text or tables and provide insight into the data. Subtle differences can be detected more easily among graphics than among data grids, allowing information to be presented with greater density. Genomic data can be visualized at several levels, such as the distribution of marker effects across the genome and relationships among markers on the same chromosome. All markers affecting a trait can be plotted on the same ordinate to visualize the distribution of marker effects across the genome, colors or textures can be used to differentiate between chromosomes, and stacked graphs can be constructed to compare interesting groups of traits. Chromosomal EBV can be presented as high-resolution graphics embedded in text to provide an overview of individual animals for comparison to potential mates. Small multiples of chromosomal genetic correlation matrices from which nonsignificant values have been excluded can be used to identify interesting patterns of association among traits, such as that on chromosome 18 associated with calving traits, conformation, and economic merit. Line plots of marker effects for recessive traits can be used to quickly locate chromosomal regions in which causative mutations are probably located, identifying areas of interest for further study. These graphics are easily produced automatically and added to online query systems, providing users with novel information at little cost.

**Key words:** genomics, graphics, visualization

### INTRODUCTION

The recent implementation of genomic evaluations in the United States (VanRaden, 2008; VanRaden et al., 2009) has been accompanied by a dramatic increase in the volume of data that are available. The August 2009 Holstein evaluation included data on 31 traits and 43,385 SNP. A total of 1,344,935 marker solutions were combined with genotypes of 28,047 animals to calculate individual genomic PTA. Several additional quantities can be derived from these solutions, such as 30 chromosomal PTA for each animal–trait combination. The resulting data are difficult to compare, levels of detail are obscured, and tabulation is not feasible. Alternative, high-resolution means of presentation are necessary.

There is a long tradition in animal breeding of using visual tools to present ideas. Wright's (1934) work on path analysis used arrow drawings to describe correlations among components of a system from which the path coefficients may be derived. Lush (1949) used illustrations to explain many points about animal breeding schemes, such as the accuracy of progeny testing schemes or programs for breed improvement through the exchange of genetic material. Animal breeding texts (e.g., Legates and Warwick, 1990) use bracket and arrow drawings of pedigrees to discuss inbreeding and relationships. Huang and Shanks (1995) described a method for plotting additive and dominance relationships and coefficients of inbreeding that used graduated circles to denote magnitude. Some properties of random regression test-day models were elegantly explained by Kachman (2004) using a series of response surfaces. Wickham et al. (2006) demonstrated how several graphical techniques could be applied to beef and dairy data. Recently, a graphical approach was used to identify bulls with bimodal patterns of inheritance for calf survival (Schlesser et al., 2009).

Pedigree visualization has received a substantial amount of attention, and there are many more software packages for that purpose than can be mentioned here (e.g., Garbe and Da, 2003). Some of those applications also provide tools for visualizing the sparsity and magnitude of numerator relationship matrices (Cole, 2007). The bioinformatics community has developed several tools for graphical presentation of data, includ-

---

Received September 22, 2009.

Accepted March 3, 2010.

<sup>1</sup>Corresponding author: john.cole@ars.usda.gov

ing metabolic networks (Carey et al., 2005), haplotype structure (e.g., Haploview; Barrett et al., 2005), and sequence structure and annotation (e.g., National Center for Biotechnology Information, 2009). Most of those tools focus on describing structure and function, rather than quantitative analysis.

There has been a substantial amount of research into statistical graphics and data visualization over the last 30 years (Tukey, 1977; Tufte, 1983; Cleveland, 1993). As Deming noted, “[G]raphical methods can retain the information in the data” (quoted in Cleveland, 1993), and are key to the complementarity of graphics and numerical techniques. Statistical procedures are “lossy,” meaning that information is discarded when calculations are performed, which is not the case with graphics. Visualizations do not have to be lossy but also cannot be compared in the same manner as statistical quantities, such as tests of hypotheses.

The objective of this paper is to present several approaches for visualizing high-dimensional numeric data. Emphasis will be placed on results from genomic evaluations.

## MATERIALS AND METHODS

### Data

Genomic data, phenotypic data, and edits were as reported in VanRaden et al. (2009). Genotypes for 43,385 SNP scored in 32,234 Brown Swiss, Holstein, and Jersey cattle were obtained using the Illumina Bovine SNP50 chip (Illumina, San Diego, CA; Matukumalli et al., 2009). Genomic predictions were computed using an infinitesimal model with a heavy-tailed prior as described in VanRaden (2008) and Cole et al. (2009). Predicted transmitting abilities were from the August 2009 US genetic evaluations calculated by the Animal Improvement Programs Laboratory (AIPL; USDA, Beltsville, MD).

### Tools

Most of the figures in this paper were produced using the Python programming language v. 2.6 (Langtangen, 2008) as distributed with SAGE v. 1.4.0 (Stein and Joyner, 2005). The data were processed with the NumPy module v. 1.3.0 (Oliphant, 2006) and visualization was performed with matplotlib v. 0.99.0 (Hunter, 2007). Sparklines were produced using code adapted from Gheorghiu’s (2006) sparkplot module (modified version available: <http://www.aipl.arsusda.gov/software/graphics/>; accessed September 9, 2009). Chromosomal PTA plots were produced using ColdFusion MX 7.0.1 (Adobe Inc., San Jose, CA). Calculations were

performed on servers and workstations running the Red Hat Enterprise Linux 5.0 (Red Hat Inc., Raleigh, NC) operating system.


### Sparklines

Sparklines are high-resolution graphics embedded in text (Tufte, 2006). They are intended for use within a body of text, rather than set apart, as is commonly the case with figures and illustrations, and may be used to represent many sorts of data. Most authors have focused on the application of sparklines to time-series data, such as levels of blood metabolites and wins-losses in sports, but Tufte (2006) discusses many applications of small graphics, including canonical work in the 17th century by the Italian astronomer Galileo Galilei. In animal breeding, obvious uses of such intertextual graphics include the display of genetic merit and illustration of genetic trends.

## RESULTS AND DISCUSSION

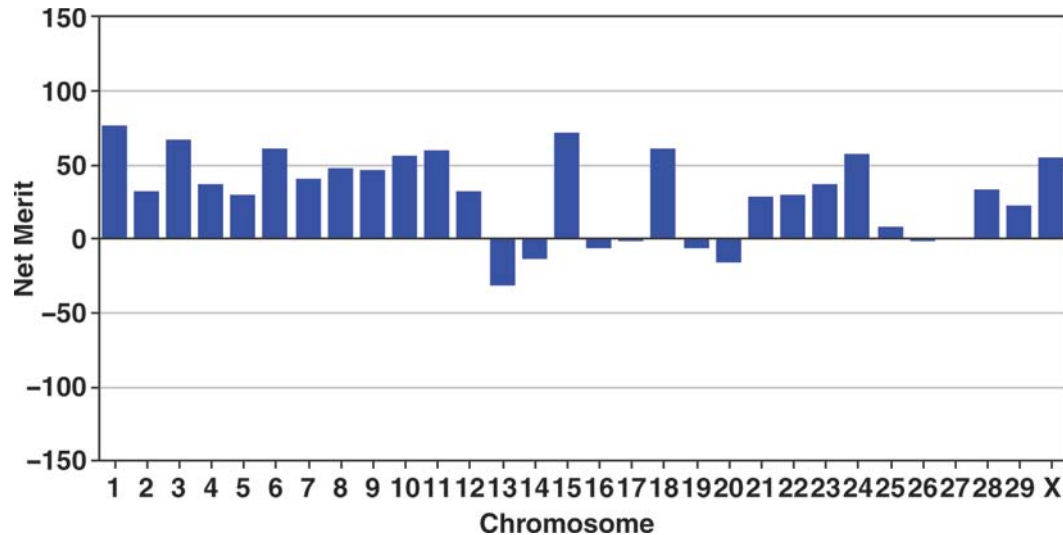
### Sparklines

**Individual Genetic Merit.** Predicted transmitting abilities can now be computed for each chromosome in the genome, and there is considerable variation among chromosomes. As an example, consider this excerpt from Cole et al. (2009) to which a sparkline showing the chromosomal PTA for the bull O-Bee Manfred Justice-ET (NAAB Code: 7HO6417) has been added:

“For example, cows with positive CEBV for chromosomes 13, 14, 16, 17, 19, or 20 might be selected for breeding to the bull O-Bee Manfred Justice-ET (7HO6417)  (Figure 1).”

The sparkline emphasizes the point being made by including the chromosomal PTA in the text. The reader is not forced to interrupt their reading to locate and study the figure. However, it is important to emphasize that these small plots are complements to, rather than replacements for, more detailed graphics. The range of values in the sparkline is unclear, and chromosomes 17, 26, and 27 appear to have values of 0, a problem that a larger figure showing more detail (Figure 1) does not have.

It is much easier to identify individual chromosomes, as well as the magnitude of their associated PTA, using the larger, more detailed graph shown in Figure 1. A query to display these data is available on the AIPL Web site ([http://aipl.arsusda.gov/CF-queries/Bull\\_Chromosomal\\_EBV/bull\\_chromosomal\\_ebv.cfm](http://aipl.arsusda.gov/CF-queries/Bull_Chromosomal_EBV/bull_chromosomal_ebv.cfm); accessed September 4, 2009) and includes all genotyped



**Figure 1.** Chromosomal PTA of lifetime net merit for the bull O-Bee Manfred Justice-ET (7HO6417). Color version available in the online PDF.

animals with published evaluations. In addition to a larger figure, the query also provides details about the animal's genomic and combined PTA and provides links to related information about the animal and the evaluation system.

Predicted transmitting abilities, such as those shown in the preceding sparklines and Figure 1, are calculated as half of the sum of the average effects of the genes carried by an individual. In the absence of an efficient procedure for haplotyping, animals are still assumed to transmit "average" chromosomes to their progeny, which are shown in the figures. For example, a bull with a PTA for lifetime net merit (NM\$) of +80 for BTA 1 may have one average chromosome (PTA NM\$ = 0) and one very good chromosome (PTA NM\$ = +160), not necessarily 2 chromosomes with similar PTA NM\$ of approximately +80. Abdel-Azim (2009) has developed a method for estimating this sampling variance using genomic data, which may be useful in planning matings.

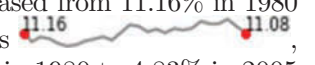

**Top Bull Lists.** Lists of bulls sorted by genetic merit for NM\$ are published by AIPL following each genetic evaluation and allow users to easily identify bulls that may be useful in their breeding programs. One of the interesting things that can immediately be noted when sparklines are included (Figure 2) is that there is considerable variation in high-value chromosomes among the top bulls. Clicking on a sparkline opens a new window containing the results from the chromosomal PTA query.

**Trends.** Genetic and phenotypic trends are often discussed, particularly when novel traits are being considered or when the efficacy of selection is under

consideration. In a discussion of phenotypic trends for stillbirth in US Holstein, Cole et al. (2007) wrote

"The incidence of SB increased from 11.2% in 1980 to 12.0% in 2004 for heifers, and decreased from 5.8% in 1980 to 5.6% in 2005 for multiparous cows."

This sentence succinctly presents their results, but does not tell the reader very much of value about the trends. Consider the same text with sparklines added to illustrate the heifer and cow trends:

"The incidence of SB decreased from 11.16% in 1980 to 11.08% in 2004 for heifers , and decreased from 5.29% in 1980 to 4.83% in 2005 for multiparous cows .

The first (1980) and last (2005) birth years are indicated by dots, and the stillbirth rate is included above the labels. When the data are shown in this manner, the reader can see that there is not a constant trend in stillbirth rates, which is implied by the text, but that stillbirth has increased over some intervals and decreased over others.

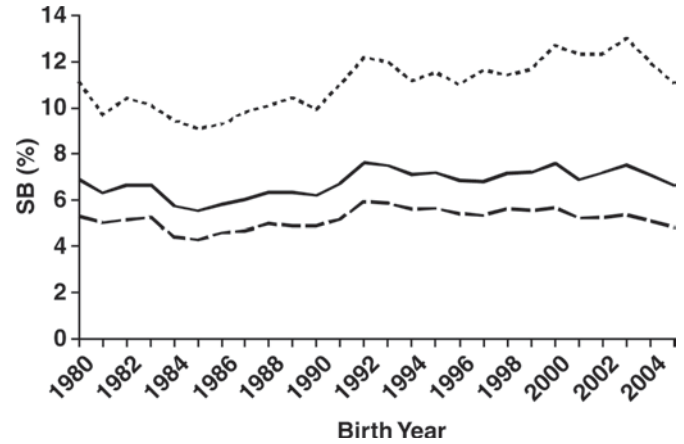
Figure 3 (based on Figure 2 of Cole et al., 2007) provides a more detailed view of the trends than do the sparklines. It is particularly easy to compare trends because results for primiparous, multiparous, and all cows are included in the graph. The sparklines provide the reader with an overview of the data without the need to leave the text, whereas the full-sized figure provides important details for in-depth study.

**USDA Evaluations for Available Holstein Sires  
Sorted by Net Merit  
August, 2009 (Example)**

ID	NAAB	Name	NMS	REL NMS	Chromosomal NMS	PTA Milk lbs	PTA Fat lbs	PTA Fat %	PTA Protein lbs	REL Yield	PTA SCS	PTA DPR	PTA PL	Herds	Daus
HOU5A000060989856	001HO08784	BAUGER-BLUFF FANNY FREDDIE	911	84	1549	73	0.06	0.04	57	89	2.72	3.5	8.9	35	60
HOU5A000139398125	007HO10072	BRIGGEN RUSSELL	889	70	1193	113	0.26	0.06	52	74	2.69	2.4	6.2	0	0
HOU5A000053557278	007HO10176	MR REGELCREEK SHOT AL-ET	870	71	1804	112	0.17	0.05	67	74	2.82	-1.4	6.0	0	0
HOU5A000139088251	007HO10219	DREAM-PRAIRIE SHADOW BOXER	837	71	1782	114	0.18	0.07	71	74	2.74	-1.4	5.0	0	0
HOU5A000095172932	001HO09208	PINE-TREE MARTHA SHAMPOO-ET	829	74	1772	97	0.12	0.03	60	77	2.64	0.6	6.6	0	0
HOU5A000064872951	014HO06090	LAESCHWAY JET BOWSER-ET	825	70	2060	58	-0.06	0.02	68	73	2.88	1.9	7.3	0	0
HOU5A000138738583	001HO02683	WA-DEL SEBASTIAN-ET	824	75	1360	71	0.08	0.04	51	78	2.88	2.8	7.3	0	0
HOU5A000053557280	007HO10059	MR REGELCREEK SHOT ALAN-ET	824	72	1928	97	0.10	0.05	72	75	2.69	-1.4	5.7	0	0
HOU5A000139057175	138HO04378	O-BEE INDIANA-ET	821	77	2180	87	0.03	-0.01	62	72	2.44	1.4	5.7	0	0
HOU5A000062253387	029HO13846	LARS-ACRES SHOT TRIGGER-ET	808	71	1881	74	0.02	0.02	62	74	2.76	0.6	6.5	0	0

[Return to the AIPL Home page](#)

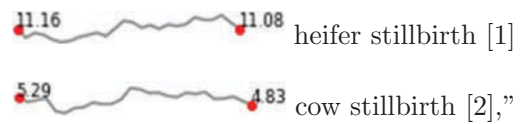
**Figure 2.** A web report showing the top 10 active AI and foreign bulls with semen distributed in the United States that are in or above the 80th percentile based on lifetime net merit (NMS) that includes sparklines showing chromosomal PTA. DPR = daughter pregnancy rate; PL = productive life. Color version available in the online PDF.



**Figure 3.** Phenotypic trend for stillbirths (SB) in primiparous heifers (dotted line), cows (broken line), and all animals (solid line) for calvings between 1980 and 2005.

Tufte (2006) provides an extremely thorough discussion of how sparklines should be designed and used, but overlooks the issue of cross-referencing, which is relevant to scientific publication. The use of sparklines could follow a system similar to that used for equations in *Journal of Dairy Science*, in which sparklines referred to multiple times in the text would be numbered for easy reference. The preceding example could be rewritten as

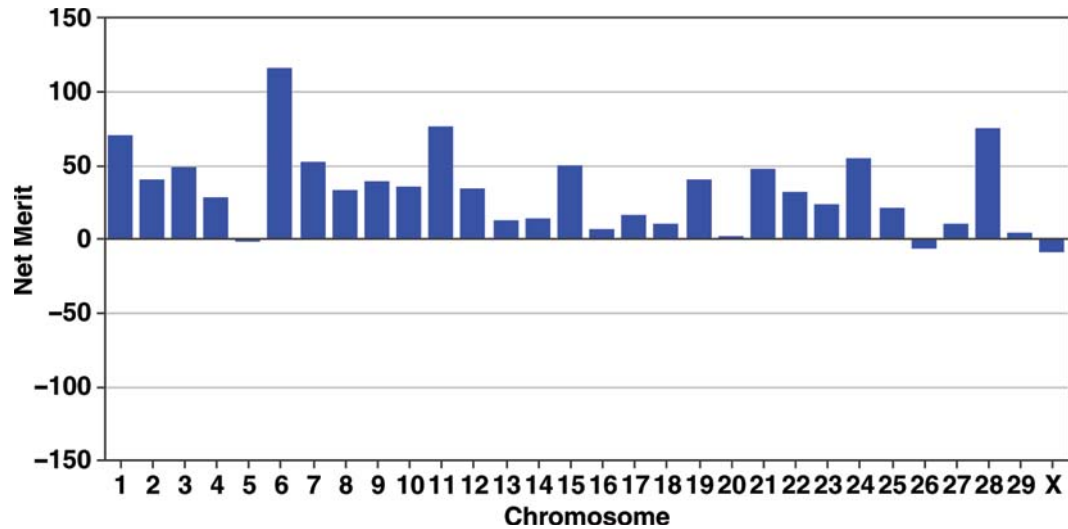
“The incidence of SB decreased from 11.16% in 1980 to 11.08% in 2004 for heifers (sparkline [1]), and decreased from 5.29% in 1980 to 4.83% in 2005 for multiparous cows (sparkline [2]):



which would permit unambiguous references to sparklines 1 and 2 when comparing them to Figure 3. If there is no need to explicitly reference the sparklines later in the discussion, then the original presentation is preferable.

**Optimal Matings**

Many tools for identifying optimal matings have been proposed, with the most common being index selection. Historically, most approaches have focused on bulls because much more accurate PTA are available for males than females. Chromosomal PTA are now available for 20,653 Holstein bulls and 7,402 cows, and the number of genotyped cows is expected to increase dramatically



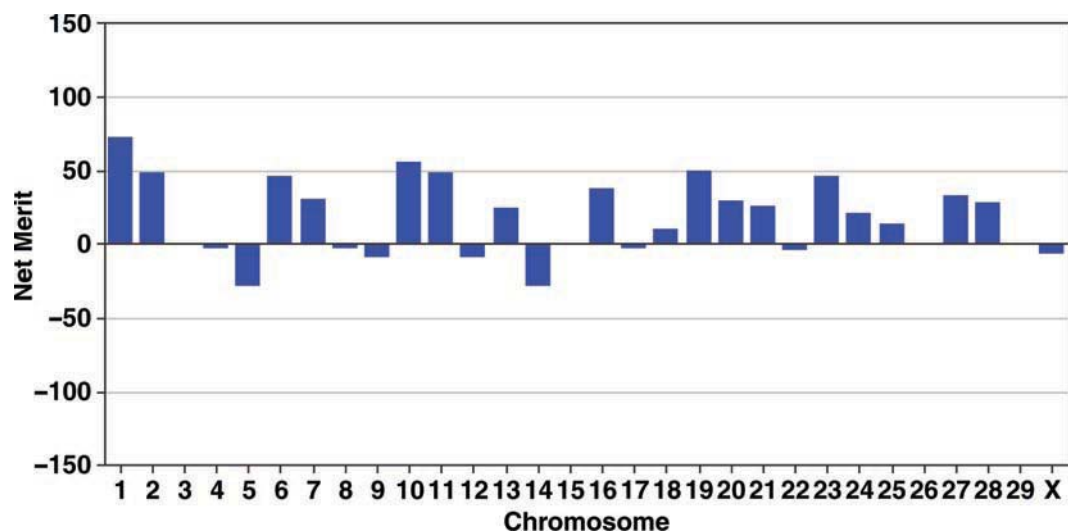
**Figure 4.** Chromosomal PTA of lifetime net merit for the cow Badger-Bluff Flo Fanny-TW (HOUSA000051854015). Color version available in the online PDF.

when a lower density, lower cost SNP chip becomes available. In this section, the application of graphical tools to selection decisions will be discussed.

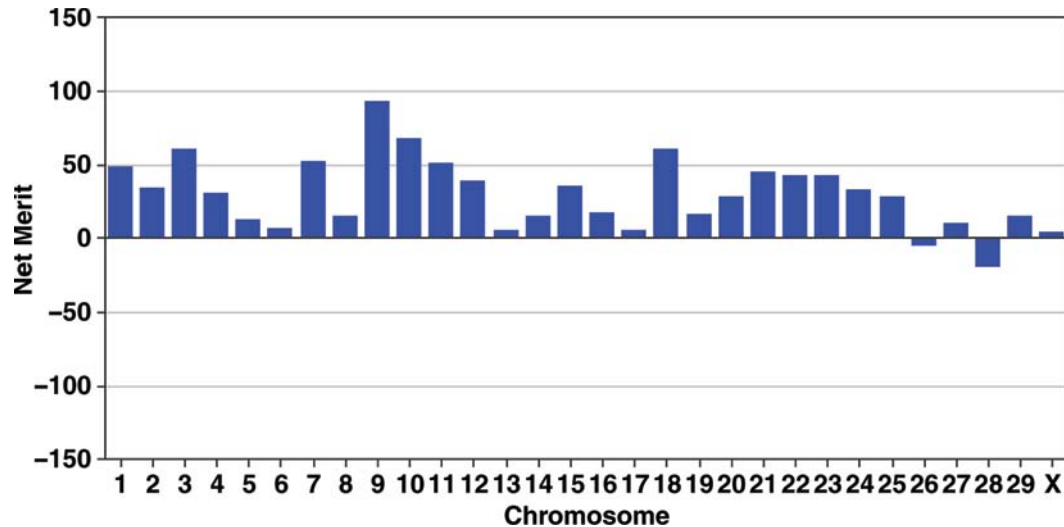
**Tracking Individual Chromosomes.** Results from genomic evaluations may be used to identify chromosomes with high genetic merit and track them through pedigrees. As of August 2009, the Holstein bull Badger-Bluff Fanny Freddie (Freddie, 001HO08784) was the highest net merit bull in the breed, with a PTA NM\$ of +911. Individual chromosomes can be tracked through Freddie's pedigree using the chromosomal PTA of his sire, O-Bee Manfred Justice-ET (O-Man, 007HO06417; Figure 1), and his dam, Badger-Bluff Flo Fanny-TW (Fanny, HOUSA000051854015; Figure 4).

Freddie has high chromosomal NM\$ for BTA1, BTA6, BTA11, BTA24, and BTA28 (Figure 5). O-Man was above average for BTA1 (+76) as was Fanny (+73), and Freddie had a PTA almost equal to the parent average (+71). Although both O-Man and Fanny had good PTA for BTA6 (+61 and +46, respectively), Freddie surpassed them substantially with a PTA of +116. Freddie's BTA11 (+77) was also slightly better than that of his sire (+59) and dam (+48). For BTA24, Freddie (+55) and O-Man (+58) had better PTA than Fanny (+21). Freddie had a much better PTA NM\$ for BTA28 (+77) than did O-Man (+33) or Fanny (+28).

The chromosomes for which Freddie had the poorest PTA NM\$ included BTA5 (-2), BTA14 (-12), BTA20




**Figure 5.** Chromosomal PTA of lifetime net merit for the bull Badger-Bluff Fanny Freddie (001HO08784). Color version available in the online PDF.



**Figure 6.** Chromosomal PTA of lifetime net merit for the bull Co-Op O-Style Oman Just-ET (001HO09167). Color version available in the online PDF.

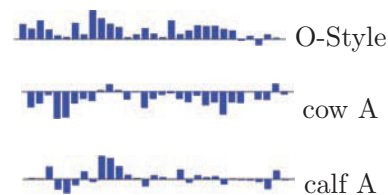
(+2), BTA26 (-7), and the X chromosome (-8). O-Man had slightly above average NM\$ for BTA5 (+29), whereas Fanny was below average (-28). Both O-Man and Fanny were poorer than average for BTA14, with PTA NM\$ of -13 and -28, respectively. O-Man was poor for BTA20 (-17), whereas Fanny was above average (+29). Freddie, O-Man, and Fanny had negative PTA NM\$ for BTA26 of -7, -2, and -1, respectively. Although O-Man had a good PTA NM\$ for the X chromosome (+55), both Freddie (-9) and Fanny (-7) were below average.

These results confirm that parents with desirable chromosomal PTA transmit good chromosomes to their offspring. Prospective mates can be genotyped to identify those with high chromosomal PTA to complement Freddie's strengths or improve on his weaknesses.

**Mate Selection.** In this section we will discuss the problems of using chromosomal PTA to select cows for mating to the bull Co-Op O-Style Oman Just-ET (O-Style, 001HO09167) , an O-Man son with a PTA NM\$ of +793 (Figure 6). Three cows, representing very high, average, and very low genetic merit animals, have been selected for ease of comparison. Cow A has a very poor PTA NM\$ of -757; cow B is slightly above average, with a PTA NM\$ of +96; and cow C has a very high PTA NM\$ of +823. In the following discussion, the chromosomal PTA for O-Style and a cow are presented, followed by the chromosomal parent averages (PA) of the offspring.

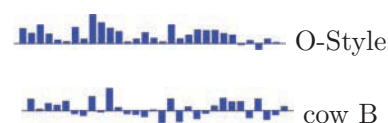
Calf A has a PA NM\$ of +12, making it an average Holstein. Cow A is not a genetically desirable animal, having only 4 chromosomes with PTA greater than

zero (BTA 9–11, 29). O-Style has positive chromosomal PTA for all but 2 chromosomes (BTA26 and 28).



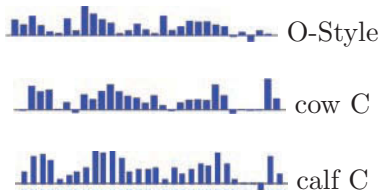
Examination of the sparklines for the parents shows which chromosomes of the calf are expected to be good (positive) or poor (negative) given the parental chromosomes. However, offspring inherit only one chromosome from each parent, or a blend if crossing-over occurs. Thus, a chromosomal PTA of 0 can be explained by 2 average chromosomes (e.g., 0 and 0), or one very good and one quite poor chromosome (e.g., +42 and -42). In the absence of haplotypes, sparklines provide a useful overview of an animal's genetic merit, but there may be considerable variation among offspring of the same parents.

Calf B has a PA NM\$ of +439, and positive chromosomal PA across most of the genome. Cow B is an above-average Holstein cow with particularly good BTA8 and 10, and might be an attractive mate for O-Style, who has an average BTA8:





Finally, calf C has a PA NM\$ of +803, placing it in the 99th percentile for all Holsteins. Cow C has very desirable PTA for most chromosomes and also ranks in the 99th percentile in the Holstein breed. O-Style appears to be an excellent mate for cow C, but there may be bulls that better complement her weaknesses, particularly BTA 24 through 28:



It is possible to overlay the chromosomal PTA of potential mates to produce a sparkline showing the chromosomal breeding values resulting from a proposed mating and the contributions of each parent.



Such stacked bar charts are not easy to produce with current software, and when they are embedded in text it can be difficult to differentiate among sire and dam contributions because of their small size. As haplotypes become available, it may be interesting to produce larger versions of this plot to see how the offspring's genotype is the sum of the parental contributions.

Sparklines are useful selection tools when a breeder wants to improve an individual chromosome, although the chromosomal PTA may differ from expectations because the sparklines do not distinguish between the aggregate breeding value and that of individual haplotypes. If the objective is to improve several chromosomes simultaneously, then a computational approach should be used to select mates.

**Chromosomal Selection.** Ideally, we would like to produce an animal whose genome includes all of the best chromosomes in the population. This could be done by identifying the animals with the best chromosomes and mating them in pairwise fashion to produce offspring that have the 2 best parental chromosomes. The offspring would then be mated in a similar fashion to produce individuals with 4 of the best chromosomes. In the best case, a series of 60 matings over 5 generations (approximately 25 yr) would be needed to produce such a "supercow," which could then be propagated by embryo transfer or cloning. This is akin to the production of consomic mouse strains, in which one chromosome in an inbred line of mice has been replaced by the homo-

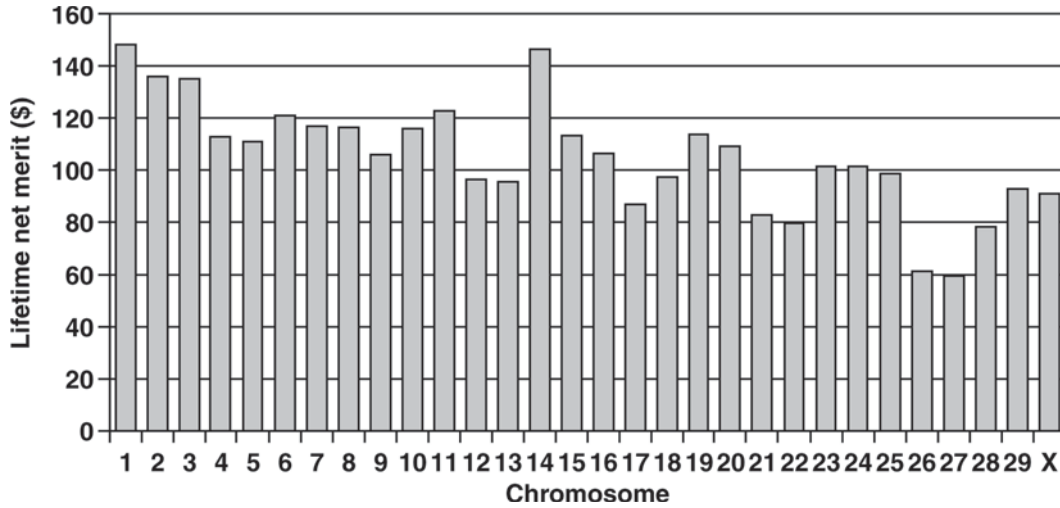
gous chromosome from another inbred strain through a series of backcrosses (e.g., Takada et al., 2008). A velogenetic approach as described by Georges and Massey (1991), which would combine advanced reproductive technologies with marker-assisted selection, could dramatically shorten the amount of time needed to produce 5 generations of animals but would be quite expensive. Chromosome selection provides many of the proposed benefits of velogenetics, such as the rapid introgression of desirable genes into a population (Odegård et al., 2009), without the costs of oocyte harvesting and embryo production. The generation interval will also decrease as use of young bull semen by dairy producers increases.

If the 30 best chromosomes in the US Holstein population were combined in a single animal, it would have a PTA NM\$ of +\$3,148 (Figure 7), about 3.5 times larger than Freddie (Figure 5), whose PTA NM\$ is +911. The current genetic trend for NM\$ is 0.25 SD/yr, and 1 SD = \$163 (VanRaden and Multi-State Project S-1008, 2006). Assuming that genetic trend for NM\$ remains constant, it would take approximately 77 yr (about 15 generations) to increase the population average to match the genetic merit of the "supercow."

The chromosomal PTA NM\$ for the 30 animals with the best individual chromosomes in the US Holstein population are shown in Figure 8. These animals all have higher than average PTA NM\$, which is the expected response to selection, but they tend to be outstanding only for one or a few chromosomes. The first 3 animals in the table have the largest PTA NM\$ for BTA1, BTA2, and BTA3, respectively, and much lower PTA for the other 29 chromosomes. Animal 26 has the highest PTA NM\$ for BTA26, but has an even higher value for BTA1, which is due, in part, to its inheritance of a superior chromosome from one of its parents, as well as the fact that BTA 1 (161,021,444 bp) is a much larger chromosome than BTA 26 (51,726,098 bp).

The potential value of chromosome selection lies not in producing a single genotype that will replace most other genotypes in the population, with a resulting rapid loss of genetic diversity, but in providing a tool that can be used to quickly propagate desirable haplotypes or eliminate undesirable ones. It may also be useful for the establishment of lines of cattle from which terminal crosses can be made to produce animals with genotypes optimized for different production systems, as is common in the swine and poultry industries.

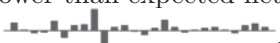
**Management of Genetic Heterozygosity.** Inbreeding increased rapidly following the introduction of animal model genetic evaluations in 1989 (Wiggans et al., 1988), and there is concern that genomic selection

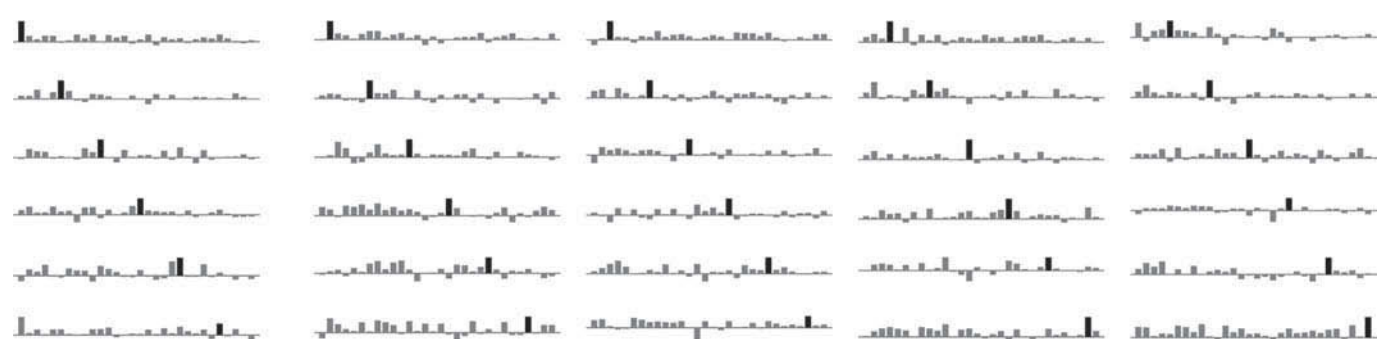


**Figure 7.** Chromosomal PTA of lifetime net merit for a hypothetical animal whose genotype consists of the best chromosomes in the current US Holstein population. The sum of the individual chromosome effects is \$3,148.

may make the problem worse. However, Daetwyler et al. (2007) have shown that the increased accuracy of genomic evaluations will allow for increased selection intensity while reducing inbreeding. This is due largely to a reduction in the magnitude of the between-family variance and increased emphasis on Mendelian sampling over BLUP selection, which favored the selection of close relatives. High levels of inbreeding clearly negatively affect animal performance, but as the cost of high-density SNP genotyping and even full-genome sequencing continue to decrease it may be that inbreeding as a measure of genetic health will cease to be important. Emphasis can be placed on regions where heterozygosity is associated with increased fitness, such as at the major histocompatibility complex, while still allowing fixation in regions where homozygosity is associated with increased profitability. However, there is a clear need to aggressively conserve existing germplasm using resources such as the National Center for Genetic

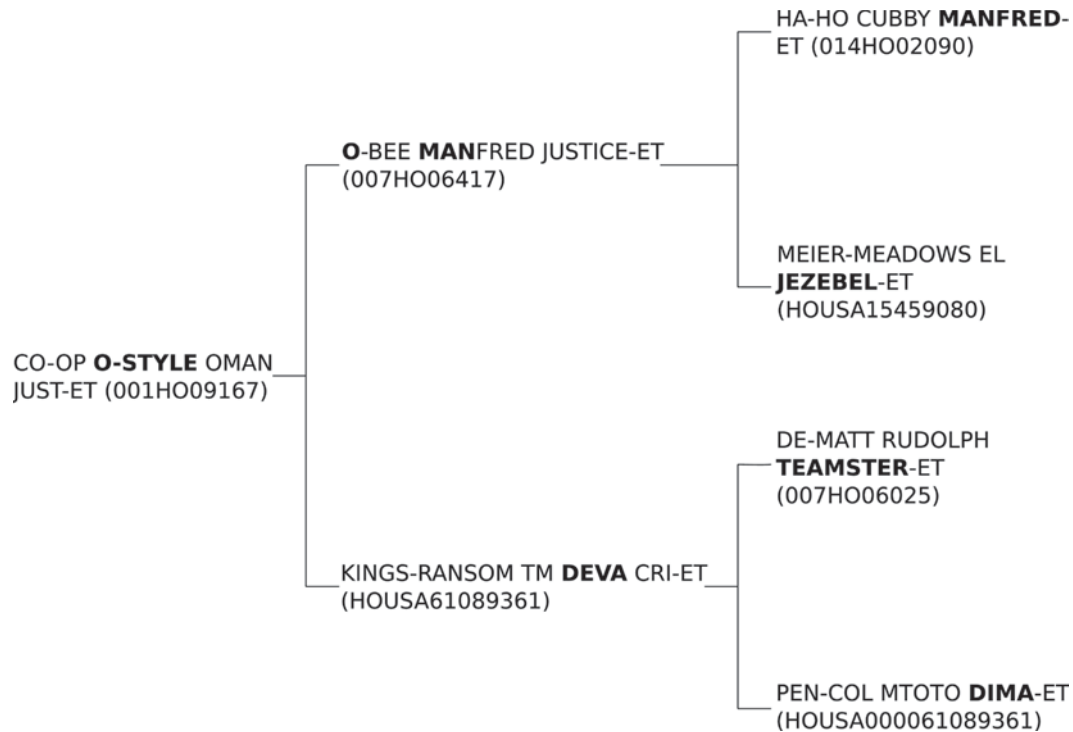
Resources Preservation (USDA-ARS, Fort Collins, CO).

In livestock breeding programs, maximizing profitability or some similar measure of economic performance is of general interest, but that does not have to be the case. Chromosome selection may be useful in conservation schemes in which the selection objective is some measure of genetic diversity rather than productivity. Although genomic evaluations cannot be computed for small populations because there are insufficient data to accurately estimate SNP effects, homozygosity still can be calculated. In that case, selection could be for flat sparklines with values near the axis, which would represent some measure of heterozygosity near 50%. For example, chromosomes with greater than expected heterozygosity can be plotted as positive values, and those with lower than expected heterozygosity as negative values: . More sophisticated schemes in which particular chromosomes are preserved



**Figure 8.** Sparklines showing the chromosomal PTA of lifetime net merit for the 30 animals in the US Holstein population with the best individual chromosomes. The best chromosome is indicated in black, and chromosome number increases across a row; the top left animal has the best chromosome 1; the animal in the bottom right has the best X chromosome.





**Figure 9.** Three-generation pedigree for the bull Co-Op O-Style Oman Just-ET (001HO09167); all of the animals in the pedigree have been genotyped.

can be imagined but may require further developments in haplotyping to achieve their full potential. The use of genomic relationships also may reduce the effect of incomplete and incorrect pedigrees on conservation programs (Oliehoek and Bijma, 2009).

### Relationship Matrices

**Three-Generation Pedigree with Complete Genotyping.** All of the animals in O-Style's 3-generation pedigree (Figure 9) have been genotyped, providing an opportunity to compare pedigree and genomic relationships. Coefficients of relationship and inbreeding were calculated from the national pedigree file using all ancestors back to 1960, as well as with SNP marker data (VanRaden, 2008). Differences between the pedigree and genomic relationships were calculated by subtracting the numerator relationship matrix from the genomic relationship matrix. For illustrative purposes, expected relationships were calculated assuming that O-Style's grandparents were unrelated. The resulting matrices were visualized as heatmaps (Figure 10), graphics in which relationships are represented as colors along a spectrum. Values near 0 are dark (blue) and values near 1 are light (red); the color version of this figure is available in the online version of the journal; <http://www.journalofdairyscience.org/>.

The expected relationship matrix is presented in Table 1, and contains only 4 values, which correspond to unrelated animals (0.0), parent–progeny pairs (0.5), grandparent–grandprogeny relationships (0.25), and animals with themselves (1.0). The contents of this matrix are visualized in the subplot labeled “Expected” in Figure 10. Unrelated animals are denoted by dark-colored squares on the off-diagonal, and light-colored squares indicate increasingly greater relationships. Individual animals are represented by squares on the diagonal. There is no inbreeding in this pedigree because relationships beyond the grandparents were not considered, and the upper left  $4 \times 4$  submatrix shows that none of the founders in this pedigree were related.

The pedigree relationships (Table 2) differ from the expected relationships because O-Style's grandparents shared ancestors, with coefficients of relationship ranging from about 0.05 to 0.11. Those relationships, labeled “Pedigree” in Figure 10, are shown in dark shades (blue in the color version). All of the animals shown in Figure 9 are inbred 4 to 6%, although that inbreeding was not reflected in the expected pedigree.

Relationships among grandparents calculated from the marker data (Table 3) were more variable than the pedigree estimates, ranging from about 0.02 to 0.13, with animals more and less related than suggested by the pedigree. Coefficients of inbreeding were much

**Table 1.** Expected numerator relationships in the 3-generation pedigree for Co-Op O-Style Oman Just-ET (001HO09167) assuming all grandparents are unrelated

	Manfred <sup>1</sup>	Jezebel <sup>1</sup>	Teamster <sup>2</sup>	Dima <sup>2</sup>	O-Man <sup>3</sup>	Deva <sup>3</sup>	O-Style
Manfred	1.00	0.00	0.00	0.00	0.50	0.00	0.25
Jezebel	0.00	1.00	0.00	0.00	0.50	0.00	0.25
Teamster	0.00	0.00	1.00	0.00	0.00	0.50	0.25
Dima	0.00	0.00	0.00	1.00	0.00	0.50	0.25
O-Man	0.50	0.50	0.00	0.00	1.00	0.00	0.50
Deva	0.00	0.00	0.50	0.50	0.00	1.00	0.50
O-Style	0.25	0.25	0.25	0.25	0.50	0.50	1.00

<sup>1</sup>Parents of O-Man, paternal grandparents of O-Style.

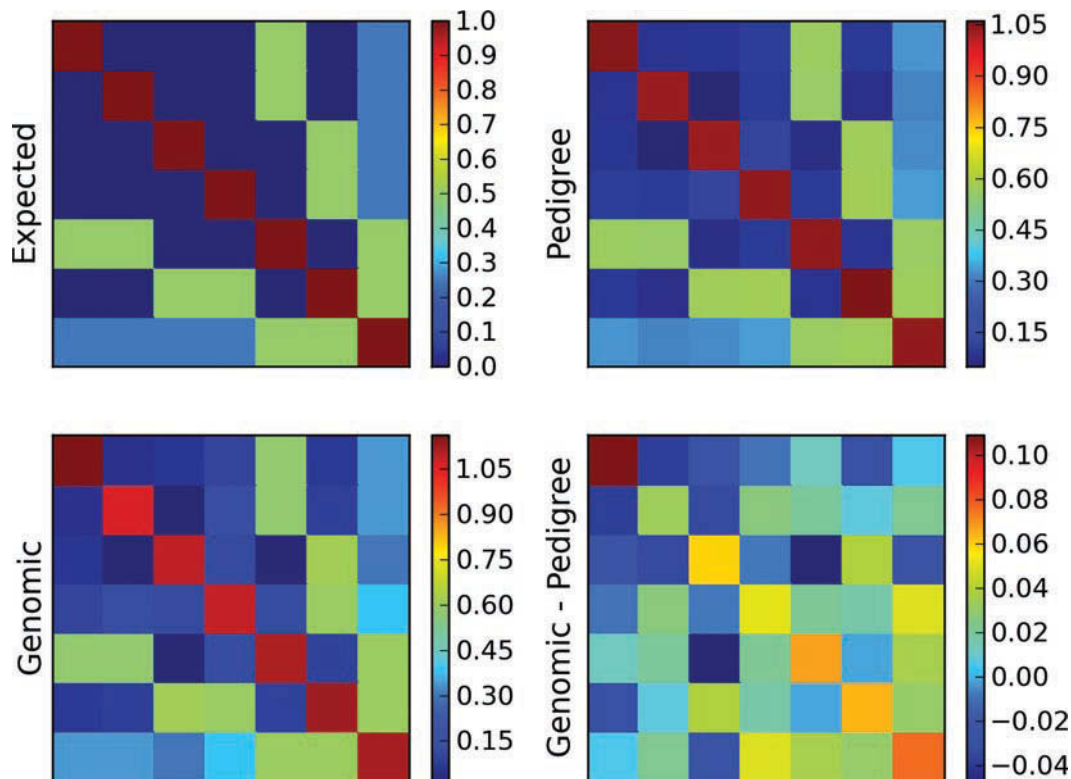
<sup>2</sup>Parents of Deva, maternal grandparents of O-Style.

<sup>3</sup>Parents of O-Style.

higher than in the pedigree data, ranging from 7 to 16%. These differences are quite obvious when the subplot labeled “Pedigree” in Figure 10 is compared with the expected and pedigree heatmaps.

The heatmap of differences between pedigree and genomic estimates (Figure 10) is labeled “Genomic – Pedigree,” and most values are near zero, with the notable exception of Ha-Ho Cubby Manfred-ET’s (014HO02090) coefficient of inbreeding, which was almost 11% higher in the genomic matrix, and the

relationship between De-Matt Rudolph Teamster-ET (HOUSA17367125) and Kings-Ransom TM Deva CRI-ET (HOUSA61089361), which was only one-third as large in the genomic matrix as the pedigree. The differences between the pedigree and genomic matrices (Table 4) reflect errors in pedigree files that do not influence the genomic relationships, as well as genes that are identical-in-state among animals that have unrelated pedigrees. Identity-in-state cannot be directly accounted for in the construction of numerator rela-



**Figure 10.** Colormaps showing coefficients of relationship (off-diagonals) and inbreeding (diagonals) for the bull Co-Op O-Style Oman Just-ET (001HO09167), his parents, and grandparents. Relationships were calculated assuming that all grandparents were unrelated (Expected), using the US pedigree file (Pedigree), and using single nucleotide polymorphism data (Genomic). Differences between the pedigree and genomic values also were visualized (Genomic – Pedigree). Color version available in the online PDF.

**Table 2.** Pedigree relationships in the 3-generation pedigree for Co-Op O-Style Oman Just-ET (001HO09167) including all ancestors back to 1960

	Manfred <sup>1</sup>	Jezebel <sup>1</sup>	Teamster <sup>2</sup>	Dima <sup>2</sup>	O-Man <sup>3</sup>	Deva <sup>3</sup>	O-Style
Manfred	1.053	0.090	0.091	0.105	0.571	0.098	0.335
Jezebel	0.090	1.037	0.051	0.099	0.563	0.075	0.319
Teamster	0.091	0.051	1.036	0.121	0.071	0.578	0.325
Dima	0.105	0.099	0.121	1.042	0.102	0.582	0.342
O-Man	0.571	0.563	0.071	0.102	1.045	0.087	0.566
Deva	0.098	0.075	0.578	0.582	0.087	1.061	0.574
O-Style	0.335	0.319	0.325	0.342	0.566	0.574	1.043

<sup>1</sup>Parents of O-Man, paternal grandparents of O-Style.

<sup>2</sup>Parents of Deva, maternal grandparents of O-Style.

<sup>3</sup>Parents of O-Style.

tionship matrices using only pedigree data, resulting in biased estimates of relationships and inbreeding using traditional methods.

**Large Populations.** Visualization of large relationship matrices may provide insight into population structure. Genomic coefficients of relationship and inbreeding for a group of 204 genotyped Holstein cows and bulls are shown in Figure 11 (color version available online; <http://www.journalofdairyscience.org/>). There is a block of largely unrelated cows in the upper, left corner of the matrix. The dark-colored rows and columns in that part of the matrix indicate that those cows are largely unrelated to the rest of the animals in the pedigree. Several lighter colored blocks of closely related animals can be seen moving from the top, left corner toward the bottom, right corner of the figure, and represent full- and half-sib families, including Freddie and many of his ancestors.

### Marker Effects

The SNP effects are updated with each genomic evaluation, as shown below, and an online query to display those data for each evaluated trait is available on the AIPL Web site ([http://aipl.arsusda.gov/Report\\_Data/Marker\\_Effects/marker\\_effects.cfm](http://aipl.arsusda.gov/Report_Data/Marker_Effects/marker_effects.cfm); accessed Sep. 10, 2009). Markers on the same chromosome are plotted in

the same color, and larger versions of each plot may be obtained by selecting an image or trait name (Figure 12). Marker effects are useful for identifying which SNP have large effects on a trait. For example, a marker on BTA18 is associated with dystocia and conformation (Cole et al., 2009).

The query originally presented the absolute values of the marker effects, but those units have no clear interpretation. The presence of markers with large effects on a chromosome also resulted in flattened plots with little apparent variation among markers. The query was revised in July 2009, and the marker solutions are now expressed in additive genetic standard deviations with an upper limit of 0.16 SD. The use of an upper limit on the SNP effects allows markers that explain lots of variation to stand out easily from the other markers and results in plots that more accurately show the variation among markers with small effects. When the actual value of 0.43 was used as the limit, the variation among the other marker effects was not visible in the graph, giving the mistaken impression that there was no variation among markers with small effects.

Although the query currently presents only results from the AIPL database, much more information could be provided. The National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>; accessed Sep. 10, 2009) provides tools for accessing its databases

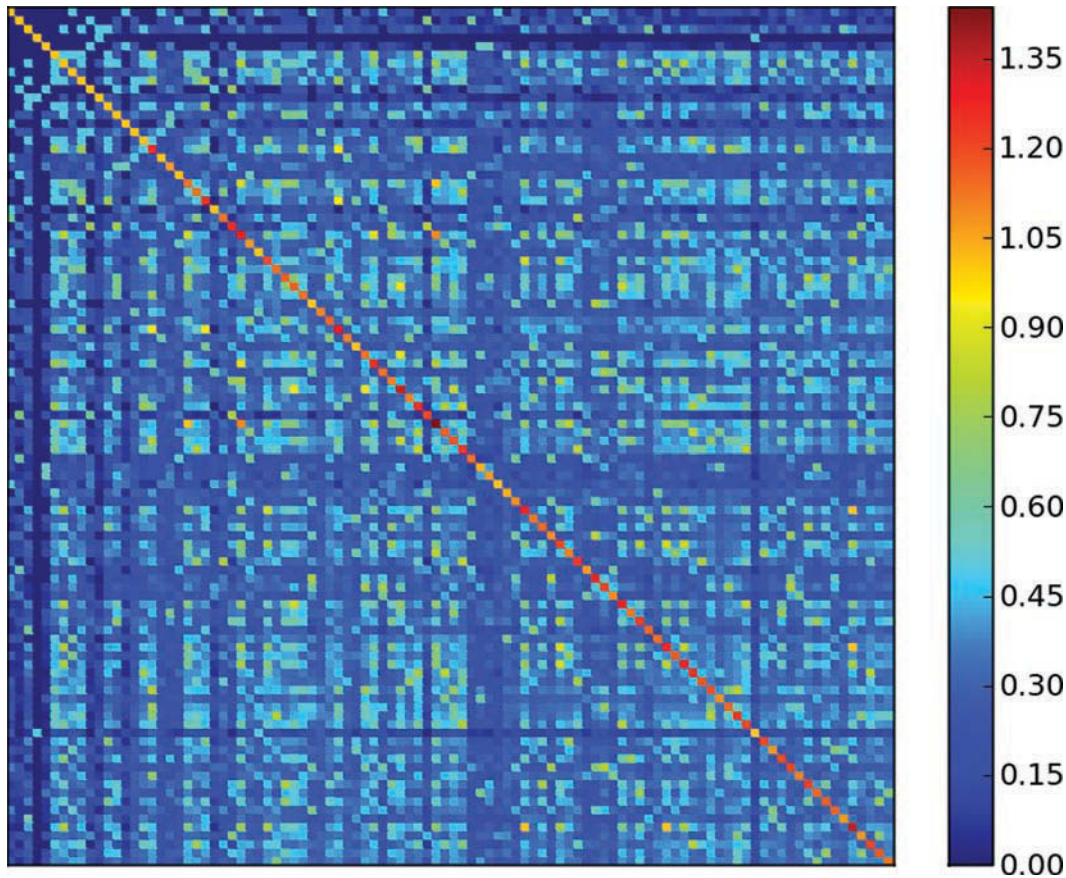
**Table 3.** Genomic relationships in the 3-generation pedigree for Co-Op O-Style Oman Just-ET (001HO09167)

	Manfred <sup>1</sup>	Jezebel <sup>1</sup>	Teamster <sup>2</sup>	Dima <sup>2</sup>	O-Man <sup>3</sup>	Deva <sup>3</sup>	O-Style
Manfred	1.161	0.052	0.066	0.096	0.584	0.069	0.341
Jezebel	0.052	1.070	0.018	0.127	0.584	0.084	0.342
Teamster	0.066	0.018	1.096	0.114	0.024	0.618	0.300
Dima	0.096	0.127	0.114	1.094	0.125	0.600	0.392
O-Man	0.584	0.584	0.024	0.125	1.113	0.086	0.603
Deva	0.069	0.084	0.610	0.600	0.086	1.126	0.605
O-Style	0.341	0.342	0.300	0.392	0.603	0.605	1.120

<sup>1</sup>Parents of O-Man, paternal grandparents of O-Style.

<sup>2</sup>Parents of Deva, maternal grandparents of O-Style.

<sup>3</sup>Parents of O-Style.



**Figure 11.** Genomic coefficients of relationship and inbreeding for a group of 204 genotyped Holstein cows and bulls. There is a block of largely unrelated cows in the upper left corner of the matrix. Several blocks representing full- and half-sib families, including the bull Badger-Bluff Fanny Freddie (001HO08784) and many of his ancestors, can be seen in the lower right portion of the matrix. Color version available in the online PDF.

over the Internet, as do many other institutions. A display can combine data from several sources to provide a comprehensive overview of the genomic context in which a marker of large effect is located. Such programming interfaces can be used to combine data in novel ways, demonstrating the value of making those data easily accessible. However, such queries may require

more effort to maintain because they depend on the correct operation of many systems.

## CONCLUSIONS

Well-designed graphics can present more information in a smaller area than text or tables and provide ad-

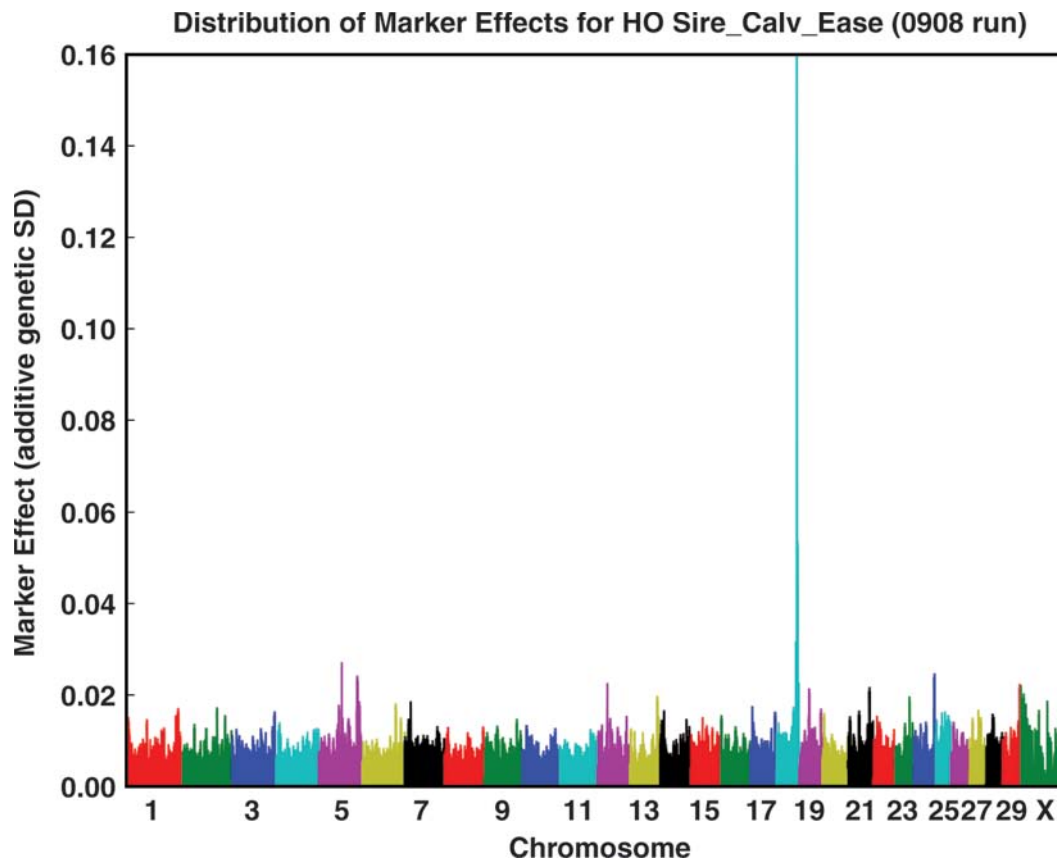
**Table 4.** Differences between the genomic and pedigree relationships in the 3-generation pedigree for Co-Op O-Style Oman Just-ET (001HO09167)

	Manfred <sup>1</sup>	Jezebel <sup>1</sup>	Teamster <sup>2</sup>	Dima <sup>2</sup>	O-Man <sup>3</sup>	Deva <sup>3</sup>	O-Style
Manfred	0.109	-0.038	-0.025	-0.009	0.013	-0.030	0.006
Jezebel	-0.038	0.034	-0.034	0.027	0.021	0.008	0.023
Teamster	-0.025	-0.034	0.060	-0.008	-0.047	0.039	-0.024
Dima	-0.009	0.027	-0.008	0.052	0.022	0.018	0.050
O-Man	0.013	0.021	-0.047	0.022	0.068	-0.001	0.037
Deva	-0.030	0.008	0.039	0.018	-0.001	0.065	0.031
O-Style	0.006	0.023	-0.024	0.050	0.037	0.031	0.077

<sup>1</sup>Parents of O-Man, paternal grandparents of O-Style.

<sup>2</sup>Parents of Deva, maternal grandparents of O-Style.

<sup>3</sup>Parents of O-Style.



**Figure 12.** The distribution of SNP effects for sire calving ease (Sire\_Calv\_Ease) in Holstein cattle from the August 2009 genomic evaluation. Marker effects are expressed in additive genetic standard deviations, and the marker on BTA 18 associated with dystocia and conformation (Cole et al., 2009) can clearly be seen. Color version available in the online PDF.

ditional insight into the data. Genomic data can be visualized at several levels, such as the distribution of marker effects across the genome, breeding values for individual chromosomes, and relationships among individuals in a population. Graphics can be produced at low cost in an automated manner and delivered through online query systems, providing users with novel information at low cost.

### ACKNOWLEDGMENTS

Tabatha Cooper of AIPL provided the data used in the chromosome selection example. Tom Lawlor of Holstein USA (Brattleboro, VT) graciously provided the pedigree for O-Style. Two anonymous reviewers are thanked for their valuable feedback on the manuscript.

### REFERENCES

Abdel-Azim, G. 2009. An approach to predict and manage Mendelian sampling variance based on dense SNP data. *J. Dairy Sci.* 92(E Suppl. 1):125. (Abstr.)

- Barrett, J. C., B. Fry, J. Maller, and M. J. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265.
- Carey, V. J., J. Gentry, E. Whalen, and R. Gentleman. 2005. Network structures and algorithms in Bioconductor. *Bioinformatics* 21:135–136.
- Cleveland, W. S. 1993. *Visualizing Data*. Summit Press, Hobart, NJ.
- Cole, J. B. 2007. PyPedal: A computer program for pedigree analysis. *Comput. Electron. Agric.* 57:107–113.
- 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. Wiggins. 2009. Distribution and location of genetic effects for dairy traits. *J. Dairy Sci.* 92:2931–2946.
- Cole, J. B., G. R. Wiggins, and P. M. VanRaden. 2007. Genetic evaluation of stillbirth in United States Holsteins using a sire-maternal grandsire threshold model. *J. Dairy Sci.* 90:2480–2488.
- Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. Inbreeding in genome-wide selection. *J. Anim. Breed. Genet.* 124:369–376.
- Garbe, J. R., and Y. Da. 2003. A software tool for the graphical visualization of large and complex populations. *Acta Genet. Sin.* 30:2293–2295.
- Georges, M., and J. M. Massey. 1991. Velogenetics, or the synergistic use of marker assisted selection and germ line manipulation. *Theriogenology* 35:151–159.
- Gheorghiu, G. 2006. Sparkplot: Creating sparklines in Python with matplotlib. <http://sparkplot.org/wiki/WikiStart> Accessed September 9, 2009.

- Huang, Y. C., and R. D. Shanks. 1995. Visualization of inheritance patterns from graphic representation of additive and dominance relationships between animals. *J. Dairy Sci.* 78:2877–2883.
- Hunter, J. D. 2007. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9:90–95.
- Kachman, S. D. 2004. Relationship between the choice of a random regression model and the possible shapes of the resulting variance functions. *J. Dairy Sci.* 87(Suppl. 1):243. (Abstr.)
- Langtangen, H. P. 2008. *Python Scripting for Computational Science*. 3rd ed. Springer-Verlag, Berlin, Germany.
- Legates, J. E., and E. J. Warwick. 1990. *Breeding and Improvement of Farm Animals*. 8th ed. McGraw-Hill Inc., New York, NY.
- Lush, J. L. 1949. *Animal Breeding Plans*. Iowa State College Press, Ames, IA.
- Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. R. O'Connell, S. S. Moore, T. P. L. Smith, T. S. Sonstegard, and C. P. Van Tassell. 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* 4:e5350.
- National Center for Biotechnology Information. 2009. National Center for Biotechnology Information Website. <http://www.ncbi.nlm.nih.gov/> Accessed August 25, 2009.
- Odegård, J., A. K. Sonesson, M. H. Yazdi, and T. H. E. Meuwissen. 2009. Introgression of a major QTL from an inferior into a superior population using genomic selection. *Genet. Sel. Evol.* 41:38–47.
- Oliehoek, P. A., and P. Bijma. 2009. Effects of pedigree errors on the efficiency of conservation decisions. *Genet. Sel. Evol.* 41:9–19.
- Oliphant, T. E. 2006. *Guide to NumPy*. Brigham Young University, Provo, UT.
- Schlessler, H. N., R. D. Shanks, P. J. Berger, and M. H. Healey. 2009. Graphical approach to evaluate genetic estimates of calf survival. *J. Dairy Sci.* 92:2166–2173.
- Stein, W., and D. Joyner. 2005. SAGE: System for Algebra and Geometry Experimentation. *Commun. Comput. Algebra* 39:61–64.
- Takada, T., A. Mita, A. Maeno, T. Sakai, H. Shitara, Y. Kikkawa, K. Moriwaki, H. Yonekawa, and T. Shiroishi. 2008. Mouse intersubspecific consomic strains for genetic dissection of quantitative complex traits. *Genome Res.* 18:500–508.
- Tufte, E. R. 1983. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT.
- Tufte, E. R. 2006. *Beautiful Evidence*. Graphics Press, Cheshire, CT.
- Tukey, J. W. 1977. *Exploratory Data Analysis*. Addison-Wesley Publishing Company, Reading, MA.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423.
- 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.
- VanRaden, P. M., and Multi-State Project S-1008. 2006. Net merit as a measure of lifetime profit: 2006 revision. <http://gov/reference/nmcalc-2006.htm> Accessed Sep.10, 2009.
- Wickham, H., A. Cromie, and D. Cook. 2006. Dynamic and interactive graphical methods for animal breeding. Commun. No. 03–15 in Proc. 8th World Congr. Genet. Appl. Livest. Prod., Belo Horizonte, Brazil.
- Wiggans, G. R., I. Misztal, and L. D. Van Vleck. 1988. Implementation of an animal model for genetic evaluation of dairy cattle in the United States. *J. Dairy Sci.* 71(Suppl. 2):54–69.
- Wright, S. 1934. The method of path coefficients. *Ann. Math. Stat.* 5:161–215.