Impact of genomic collaboration and the need to continually improve methods.

Paul M. VanRaden

USDA Animal Genomics and Improvement Laboratory

Beltsville, MD USA

Continuous improvement

Dairy cattle improve each generation as breeders apply new tools and additional data in genetic selection. Traits included in selection indexes have increased steadily and will further expand because genotypes multiply the value of data collection. An animal can be genotyped once to predict an unlimited number of traits. Producers desire flexible and inexpensive data collection methods, but uniformity is also a major goal. Dairy Herd Improvement organizations in the past had rigid rules to enforce centering dates and monthly testing plans, but then switched to more flexible statistical methods such as test interval method, best prediction, or test day models to include a wider variety of data. Similarly, genomic predictions began in 2008 with a standard chip containing 50,000 markers, but new data sources and computational methods have rapidly emerged, increasing flexibility and reducing costs to obtain genotypes.

Collaboration

Foreign selection of breeding stock has and will continue to greatly impact the domestic populations of most countries. Methods to rank foreign bulls progressed from simple conversion formulas to more advanced multitrait across country evaluation (MACE) and then to genotype exchanges and genomic MACE. All major AI companies in North America collaborated to provide a reference population large enough to estimate even small genetic effects. This collaboration began in 1992 with a common DNA repository for US and Canadian bulls. Some breeders in Europe asked if having so many companies work together was fair, but very quickly, several countries in Europe also merged their reference populations (Lund et al., 2011).

The North American collaboration was extended to include Holstein genotypes from Italy and the United Kingdom, Jersey genotypes from Denmark, and global Brown Swiss genotypes exchanged by Interbull (Intergenomics). Most national genomic evaluations treat domestic and foreign data as the same trait, but can be more accurate by treating MACE as a correlated trait if genetic correlations are < 1 (VanRaden et al., 2012). With the exception of New Zealand where genetic correlations to other countries are low, nearly all researchers found that accuracy of predictions improved when foreign reference bull genotypes were added.

Past surprises

Genotype quality was an immediate and pleasant surprise of genomics, with most markers having error rates < 0.1%. In 2009 when a more complete map of the bovine genome became available, new computer methods called imputation could accurately predict the missing markers from other chips containing subsets of markers. This led to a proliferation of new chips of both lower and higher density that gave nearly the same accuracy for less than half price, or genotyped 15 times more markers for only twice the price. After most ancestor bulls had been genotyped, the ease of correcting pedigrees and ability to discover unknown parents, grandparents, or even great grandparents were also pleasant surprises.

Discovery of many new lethal recessive genes was a pleasant surprise to researchers but perhaps not so pleasant to the owners of many popular bulls and cows now known to carry defects that cause embryo loss. Such defects were easy to detect by searching for haplotypes that never become homozygous in live animals or by using bioinformatics to search directly for damaged genes in DNA sequences. In the case of Holstein haplotype 1 (HH1), the defect in gene APAF1 was easy to detect as a cause of embryo loss in cattle, and deficiency of this same gene was already known to cause embryo loss in mice (Adams et al, 2012). Thus, bioinformatics has become a powerful tool to discover and confirm genetic effects across species. Total worldwide economic loss from the HH1 defect has been nearly a half billion US dollars since HH1 became prevalent around 1970.

Genomic selection is now widely practiced with fairly low cost in part because the basic concepts are free of intellectual property. This makes friendly collaboration possible even among competing organizations and countries. Much of this collaboration involves simply collecting and exchanging large data files or improving the data analysis methods. The U.S. Supreme Court (2013) decided that DNA tests could not be patented, but courts in some other countries still enforce gene test patents. Because of potential legal disputes, most genomic selection programs do not yet use direct tests for the QTLs with the largest effects such as DGAT1 and instead use tens of thousands of indirect markers. Most chips now contain QTLs and individual gene tests, and including those in selection is simple if legal.

The need for rapid processing (monthly and weekly genomic predictions) was a surprise. Breeding decisions can wait until 1 year of age, whereas DNA samples can be obtained at birth, so the urgency was not obvious. However, marketing and transfer of ownership happens quickly for top animals. Many herd owners now also use genomics as a culling tool, selling calves with low genomic predictions early in life instead of raising them for replacements. High labor costs for calves during their first month created a demand for weekly evaluations (Wiggans et al., 2015). Breeders can also use genotypes as a mating tool to reduce genomic inbreeding (Sun et al., 2013). Genomic relationships of genotyped females with marketed males are provided monthly to breeding organizations and owners.

Future surprises

Rapid growth of genomic databases can quickly cause previously tested statistical methods and validations to become outdated. Computer programs also must be constantly revised to keep up with database sizes that have doubled each year for the past 6 years to nearly 1 million genotyped animals today. When data sets were small, addition of genotypes from all bulls including very old bulls increased

reliability. More recent research found that the oldest bulls could be dropped with little loss of accuracy because genotypes for many more recent bulls are available that contain the same DNA. In fact, for the more heritable traits, dropping all of the 26,759 proven bull genotypes causes little loss because the reference population now also includes >100,000 genotyped cows with records. Early tests showed that adding highly selected females did not improve accuracy, but most females now are genotyped while heifers before their phenotypes arrive, reducing any selection bias.

Current statistical methods do not fully account for genomic pre-selection of progeny, genomic merit of mates and herdmates, lack of random sampling, embryo transfer, and extreme prices. In the past, use of highly selected foreign bulls sometimes caused similar biases, but with little effect on bulls tested in domestic sampling programs. Some new genomic biases might be reduced by using single-step instead of multi-step equations, but single-step algorithms cannot yet process very large datasets, and other biases are more difficult to estimate or solve.

Prediction accuracy is usually tested by predicting the latest 4 years of proven bulls from data with the last 4 years truncated (Mäntysaari et al., 2010). This tests the ability of genomic equations to predict 1 generation ahead, but many calves (or embryos) are now 2 or 3 generations away from the proven bulls used to compute the predictions. Also, only those bulls with the highest predicted merit now obtain daughters. This can greatly reduce the variance of true genetic merit and the correlations of predicted with true merit needed to estimate the reliability of the prediction system. The success of genomics has reduced our ability to test how well it works.

DNA sequence data will gradually allow the actual QTLs that cause genetic effects to replace the genetic markers now used in genomic selection (Daetwyler et al., 2014). Genotyping chips at first contained only evenly spaced, unselected genetic markers. Some recent chips also include selected markers from a higher density (777K) chip chosen from those with largest effects. Similarly, sequence variants predicted by bioinformatics or estimated from phenotypes to have large effects are being added to chips. Alternatively, sequencing could be used directly instead of chips if the technology continues to improve rapidly. Use of low density markers to impute the QTLs and sequence variants may continue to be a good option because for genotyping whole herds, costs per animal must be very low. Also, the QTLs must be imputed for all animals previously genotyped with chips that contained only markers.

Long term progress may be difficult to predict because genetic correlations can change across time and because economic functions are nonlinear, whereas selection indexes usually assume linear values. As selection proceeds or management systems and environmental conditions change, different factors can become limiting. The genetic correlation between milk yield and longevity was very positive at about 0.7 in the 1960s but declined quickly to near 0.1 currently (Powell and VanRaden, 2003) along with large genetic increases in milk yield and decreases in fertility. Phenotypic changes in culling decisions were less dramatic, but producers now put less emphasis on yield and more emphasis on all other traits (Norman et al., 2007). Somatic cell score is a good indicator of clinical mastitis, but mastitis researchers were worried when genetic selection on SCS began that eventually cows would simply stop fighting their infections and let the disease win in order to maintain low cell counts. The correlation between SCS and clinical mastitis has not declined yet, but could with intense selection for low SCS.

Genomic evaluations can counteract previous unintended declines in traits that were unfavorably correlated to the selection goal, but only if data are available and breeders act. Long ago, researchers at USDA were surprised when their experimental Guernsey line went extinct after several generations of intense inbreeding (Woodward and Graves, 1933). Researchers at Iowa State were surprised when their line of Tribolium castaneum (flour beetles) went extinct after 16 generations of selecting for a trait correlated by -0.43 with fertility (Berger, 1976). Breeders learned from this early research that they should always monitor trends in fertility, health, conformation, and management traits and invest in collecting new traits. Fortunately, new genomic tools allow faster positive progress that may pleasantly surprise us with how quickly the previous losses can be reversed. Dairy cattle will not become extinct and instead will improve rapidly, contributing more toward feeding the growing human population with each generation.

References

Adams, H.A., Sonstegard, T., VanRaden, P.M., Null, D.J., Van Tassell, C.P., and Lewin, H. Identification of a nonsense mutation in APAF1 that is causal for a decrease in reproductive efficiency in dairy cattle. Proc. Plant Anim. Genome XX Conf., San Diego, CA, Jan. 14–18, P0555. 2012.

Berger, P.J. 1976. Multiple-trait selection experiments: Current status, problem areas and experimental approaches. In: Proc. International Conf. Quantitative Genetics. Ames, IA.

Daetwyler, H D, et al. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nature Genetics 46:858–865.

Lund, M.S., de Roos, A.P.W., de Vries, A.G., Druet, T., Ducrocq, V., Fritz, S., Guillaume, F., Guldbrandtsen, B., Liu, Z., Reents, R., Schrooten, C., Seefried, F. & Su, G. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet. Sel. Evol. 43,* 43.

Mäntysaari, E., Z. Liu, and P. VanRaden. 2010. Interbull validation test for genomic evaluations. Interbull Bull. 41:5 pages.

Norman, H.D., Hutchison, J.L., Wright, J.R., Kuhn, M.T., and Lawlor, T.J. 2007. Selection on yield and fitness traits when culling Holsteins during the first three lactations. J. Dairy Sci. 90(2):1008–1020.

Powell, R.L., and VanRaden, P.M. 2003. Correlation of longevity evaluation with other trait evaluations from 14 countries. Interbull Bull. 30:15–19.

Sun, C., VanRaden, P.M., O'Connell, J.R., Weigel, K.A., and Gianola, D. 2013. Mating programs including genomic relationships and dominance effects. J. Dairy Sci. 96(12):8014–8023.

Supreme Court of the United States. 2013. Association for Molecular Pathology v. Myriad Genetics, Inc.

VanRaden, P.M., Olson, K.M., Null, D.J., Sargolzaei, M., Winters, M., and van Kaam, J.B.C.H.M. 2012. Reliability increases from combining 50,000- and 777,000- marker genotypes from four countries. Interbull Bull. 46:75–79.

Wiggans, G.R., VanRaden, P.M., and Cooper, T.A. *Technical note:* Rapid calculation of genomic evaluations for new animals. J. Dairy Sci. 98:(in press; published online Dec. 31, 2014)

Woodward, T.E. and R.R. Graves. 1933. Some results of inbreeding grade Guernsey and grade Holstein-Friesian cattle. USDA Tech. Bulletin No. 339. Summary