

## Genomics in the U.S. Dairy Industry: Current and Future Challenges

*E.L. Nicolazzi<sup>1</sup>, J.W. Dürr<sup>1</sup> and G. R. Wiggans<sup>1</sup>*

<sup>1</sup> Council on Dairy Cattle Breeding, Bowie, MD 20716, USA

### Summary

The first U.S. dairy cattle genomic evaluation was released in April 2008. Since then, the number of animals genotyped per year has rapidly increased. The over 2 million genotypes currently stored, have increased by nearly 500,000 (90/10% ratio for female and male animals, respectively) added annually since 2016. Most animals are genotyped with chips that have between 18,000 and 30,000 single-nucleotide polymorphism (SNP) markers. In order to ensure the highest quality of SNP genotypes in the US system, only CDCB approved genotyping laboratories or other national genetic evaluation centers representing international exchange partners are allowed to submit genotypes to the U.S. cooperators' database. The SNP genotypes in each submission are evaluated for call rate, portion heterozygous, and parent-progeny consistency. Each animal genotype is checked against its parents and a grandparent if the parent has not been genotyped. Considering only animals with genotypes that are usable for genomic evaluation, 97% had sires and 38% had dams that were already genotyped. Each genotype is compared with all others to discover identical genotypes and parent-progeny relationships not in the pedigree and possible grandsires if they are unknown or unlikely. These checks will soon be revised due to the growth of the genomic database. Since October 2017, animals with an unlikely grandsire detected are excluded from evaluation. Currently, 60,671 SNPs are used in U.S. genomic evaluation of dairy cattle, and research is ongoing to increase that number to ~77,000. Various genetic abnormalities are also tested for, and these test results are included in the imputation to provide an imputed indication of carrier status for animals that are not tested. Haplotypes that affect fertility also are used to detect reproductive defects. To date, 12 such haplotypes have been detected, with the latest introduced in December 2017 (AH2, Ayrshire second haplotype affecting fertility).

Genomic evaluations for production, conformation and fertility/calving traits are released on a weekly basis for new animals. In 2017, 2 new traits (cow livability and gestation length) were introduced. The CDCB will start publishing 6 new health traits on Holsteins in April 2018. Reliance on genomic evaluations to select bulls has increased rapidly in the US. In 2016, 67% of breedings through AI were to bulls with no milking progeny. The age of parents at bull birth has dropped to just over 2 years, nearly the biological minimum. This reduction in generation interval has led to almost doubling the annual genetic improvement (over \$80/year for lifetime net merit, a genetic-economic index).

Other improvements planned for 2018 are to provide more standardized and documented feedback to customers and to perform a revision of the genotype process system and accommodate the increasing number of genotypes received.

**Key words:** dairy cattle, genomic evaluation, genomic selection, genomics

### Introduction

The first U.S. dairy cattle genomic evaluation was released in April 2008 (VanRaden *et al.*, 2011). Since then, the number of animals genotyped per year has rapidly increased. For over fifty years, the U.S. national evaluation was managed by ARS-USDA. The Council of Dairy Cattle (CDCB) started operating in 2013. The transition of genetic evaluation services was completed in December 2015, whereas the

ARS-USDA still provides the research and development to the U.S. dairy industry.

The CDCB manages the U.S. cooperators' database. Dairy producers are the data owners and the final users of the services delivered by CDCB. The U.S. cooperators' database is central to the U.S. dairy industry: it is the result of a collaborative effort to collect, validate and standardize the data received from all industry sources: i) CDCB approved genomic

nominators (a certified organization that provides and monitors genomic data that is going to the CDCB evaluation system in order to ensure quality of data that is used for CDCB evaluations); ii) CDCB approved genotyping laboratories; iii) dairy record providers (DHIs) through the dairy records processing centers (DRPCs); iv) the National Association of Animal Breeders (that includes all A.I. centers); v) Purebred Dairy Cattle Association; and vi) international partners. In order to ensure the highest quality of data possible, only CDCB approved partners (which are quality certified) are allowed to submit records to the database.

The data in the U.S. cooperators' database is used routinely to provide a wide range of services to the U.S. dairy industry. Genetic (i.e. traditional) and genomic evaluations and inbreeding values are only an example of services benefitting from this wealth of data.

Three traditional (April, August and December) and twelve (monthly) genomic official evaluations – plus weekly evaluations on newly genotyped animals - are released by the CDCB to the U.S. dairy industry and the general public. More than two million animals, five breeds (Holstein, Jersey, Brown Swiss, Ayrshire and Guernsey) and 35 traits are currently involved in the monthly genomic evaluations.

In this paper we will cover only aspects related to data and services associated with the genomic section of the database.

## Materials and Methods

A review of the current status of the CDCB collaborators' database was conducted using inhouse software and scripting. Most of the results shown here are part of the statistics regularly created and shared with the CDCB partners or publicly at [www.uscdcb.com](http://www.uscdcb.com).

## Results and Discussion

As of January 30, 2018, the CDCB collaborators' database contained a total of 2,372,681 genotypes. The CDCB calculates genomic evaluations for about 2.1 million of

these genotypes every month. Nearly 90% of the animals evaluated are female, a trend that has increased slightly over the last two years. Since 2015, the CDCB has received over a million genotypes (Table 1). The projected number of genotypes to be received in 2018 is between 600,000 and 700,000.

**Table 1.** Number of genotypes received by time period.

Year	Genotypes received	Comment
≤2015	1,125,495	Most of these genotypes were received by ARS-USDA.
2016	418,445	
2017	554,823	
2018	48,617	Only January data

Most of the animals included in the database were genotyped with densities that range between 18,000 and 30,000 single-nucleotide polymorphism (SNP) markers. Currently CDCB manages 30 different SNP arrays with (raw) densities ranging from ~2,700 to ~777,000.

Every genotype is evaluated for sex, call rate, portion heterozygous, parent-progeny consistency, grandparent likelihood, undeclared relationships, breed, among other characteristics. Considering that on a weekly basis, the CDCB receives between 10,000 and 20,000 new genotypes, ARS-USDA and the CDCB are actively testing methods to reduce the computation burden without impacting data quality significantly: i) genotype pre-processing (e.g. more stringent controls at reception); ii) a new methodology to identify maternal grandsire candidates; iii) a reduction in the SNP subset used to perform the edits; and iv) high performance computation techniques to reduce processing time. The latter three projects are covered in *Wiggans et al., 2018*.

Currently, 60,671 SNPs are used in U.S. genomic evaluation of dairy cattle, and research is ongoing to increase that number to ~77,000. This increase in the number of SNPs is due to the discovery of new informative SNPs. The preliminary results are encouraging, with an average increase in reliability of 1.4% across traits (*Wiggans et al., 2016*).

Various genetic abnormalities are also tested for, and these test results are included in the imputation to provide an imputed indication of carrier status for animals that are not tested. Haplotypes that affect fertility also are used to detect reproductive defects. To date, 12 such haplotypes have been detected, with the latest introduced in December 2017 (AH2, Ayrshire second haplotype affecting fertility).

Genomic evaluations for production, conformation and fertility/calving traits are released on a weekly basis for newly genotyped animals. In 2017, two new traits were introduced in the list of genomic traits: Cow Livability (avoidance of mortality) and Gestation Length. The CDCB will start publishing six new health traits on Holsteins in April 2018 (hypocalcemia/milk fever, displaced abomasum, ketosis, mastitis, metritis, retained placenta). These traits will not be immediately included in lifetime net merit formula (NM\$, a genetic-economic index). More information on these traits can be found in: <https://www.uscdcb.com/what-we-do/genetic-evaluations/> under “Trait Calculation Details / Health traits”. Another new trait, however still under research, is Feed Efficiency (Residual Feed Intake). The trait will benefit from a collaboration between the CDCB, USDA and an agreement with UW-Madison to increase the number of phenotypic records.

Finally, USDA in collaboration with the CDCB is performing research with the objective of evaluating the possibility of extending genomic evaluation to crossbreds. This research could lead to evaluations for the large number of crossbred genotyped animals that are currently not being included in the genomic evaluations..

Reliance on genomic evaluations to select bulls has increased rapidly. In 2016, 67% of breedings through AI were to bulls with no milking progeny. Evaluations of new animals are released weekly because of owner interest in making culling decisions as early as possible to minimize rearing costs. The AI organizations also rely heavily on genomic evaluations to generate the next generation of bulls. The age of parents at bull birth has dropped to just over

2 years, nearly the biological minimum. This reduction in generation interval has led to almost doubling the annual genetic improvement (over \$80/year for NM\$). As an example, all of the top 10 young (genomic) Holstein and Jersey bulls in 2014 have A.I. sons (up to 260 sons) in 2018. On the contrary, the same has not been observed for top proven bulls in 2014.

## Conclusions

The CDCB has been providing services to the U.S. dairy industry since 2013. A large number of activities are currently underway to improve the services provided, the infrastructure on which those services depend and the documentation provided. The CDCB is also involved in identifying new data pipelines to increase the number of informative traits to evaluate, and to keep providing the best service possible to the U.S. farmer. In this sense, the recent efforts in improving the communication with the U.S. industry is starting to prove effective and will continue to be a topic of interest for the CDCB.

## Acknowledgements

Authors wish to recognize the contribution to all the research presented in this paper done by our colleagues of ARS-USDA-AGIL.

## References

- Wiggans, G.R., Cooper, T.A., VanRaden, P.M., Van Tassel, C.P., Bickhart, D.M. & Sonstegard, T.S. 2016. Increasing the Number of Single Nucleotide Polymorphisms Used in Genomic Evaluation of Dairy Cattle. *Journal of Dairy Science* 99, 4504-4511.
- Wiggans, G.R., VanRaden, P.M. & Bacheller, L. 2018. Methods for Discovering and Validating Relationships Among Genotyped Animals. *Interbull Bulletin* 53, 27-30.
- VanRaden, P.M., O’Connell, J.R., Wiggans, G.R. and Weigel, K.A. 2011. Genomic Evaluations with Many More Genotypes. *Genetics Selection Evolution* 43:10, 11 pp.