

The Cooperative Dairy DNA Repository (CDDR)

How it has helped the AI Industry

Kent A. Weigel

Department of Dairy Science

University of Wisconsin – Madison

Why Form a DNA Repository?

Collections of biological materials, such as seeds, blood, semen, or other plant and animal tissues, can be created for a variety of reasons. Genetic conservation and future research are the commonly stated goals, but such collections have little value unless the stored samples have been characterized in a manner that will allow rapid and precise identification of samples that can be withdrawn to address a specific need or opportunity. An example of a specific need would be the emergence of a new infectious disease for which living members of the species lack the ability to resist the infection or combat the physical symptoms. An example of a specific opportunity would be the discovery of a new invention that represents a technological leap from the status quo. The Cooperative Dairy DNA Repository (CDDR) was formed roughly 20 years ago to address the latter objective – that is, to facilitate future research in the area of marker-assisted (or DNA-based) selection. At the time, the term “genomics” was not yet in our collective vocabulary, and genetic marker research tended to focus on simplistic approaches that had been used successfully in identifying genetic mutations that caused recessively inherited lethal defects. Genotyping with restriction fragment length polymorphisms or microsatellite markers was costly and time-consuming, yet hundreds of studies sought to identify the gene that caused high milk production, or the gene that caused poor female fertility. Successes of this approach in food animal species can be counted on one hand, and even those have had a minimal impact on genetic selection programs. Nonetheless, forward-thinking leaders of the bovine artificial insemination (AI) industry decided to invest in development of the CDDR collection, with the hope of stimulating DNA-based research that would someday lead to methods for enhancing genetic progress in dairy cattle. Prior to establishment of the CDDR, semen from bulls that failed progeny testing, produced limited quantities, or died prematurely was often very difficult to procure for research studies. The decision by competing AI studs to embark on this venture collectively, rather than individually, stemmed from the notion that the most effective studies would require samples from a very large population of bulls, and that such studies would benefit from large half-sib families created by widespread use of influential AI sires (and, to a lesser extent, full-sib families created by embryo transfer (ET) of elite cows) for which sons and grandsons were spread across competing organizations.

How does the CDDR Work?

The basic premise of the CDDR is that 10 units of semen from every bull that enters a participating AI stud (i.e., CDDR Contributor) is sent to a central archive that is physically located at the USDA-ARS Beltsville Agricultural Research Center (Beltsville, MD), and that biological samples from this collection, or bull-specific information derived from these samples, will be provided to scientific researchers (i.e., CDDR Collaborators) following approval of the

researcher's request by the Board of Managers. The Board of Managers is comprised of one staff member from each of the Contributors, which include: ABS Global (DeForest, WI), Accelerated Genetics (Baraboo, WI), Alta Genetics (Watertown, WI), Genex Cooperative (Shawano, WI), Select Sires (Plain City, OH), Semex Alliance (Guelph, ON), and Taurus Service (Mehoopany, PA). Collaborators have been numerous, both domestic and international, but the majority have been academic researchers at land-grant universities throughout the US or government scientists at various USDA-ARS facilities. Although the goal of both Contributors and Collaborators is easy, timely, and minimally restrictive access to specific CDDR materials that will maximize the chance of success for a given research project, this has proven challenging in practice. By definition, every research project is unique, and therefore it isn't possible to develop a standard set of biological materials or portfolio of DNA-based information that meets the needs of every Collaborator. Furthermore, prospective Collaborators come from many different universities and research institutes, and each of these has different policies, restrictions, and interpretations as regards the development of mutually agreeable material transfer agreements, particularly with respect to intellectual property (IP). Lastly, and particularly in the case of international Collaborators, the Contributors must be cognizant of the fact that the CDDR collection has very significant value, and it could be used by another company or organization to gain a competitive advantage over the CDDR Contributors.

What Types of Studies have been Facilitated?

The CDDR collection has become massive over time, and at present it contains biological samples from 32,999 bulls of 9 breeds, including 29,304 Holsteins, 2,958 Jerseys, 741 Brown Swiss, 284 Guernseys, 212 Ayrshires, and a handful of samples from other breeds. Despite the aforementioned logistical challenges, dozens of research studies have benefitted from the use of CDDR materials. The objectives and approaches of these studies have varied widely, but it is safe to say that the "typical" study initially envisioned by the Contributors, and the most common type of request in practice, is a genome-wide association study (GWAS) aiming to identify mutations or polymorphisms in the coding or regulatory regions of specific genes that affect traits such as milk yield, milk composition, health, fertility, longevity, or physical conformation. Examples of this type of study, such as a GWAS for susceptibility to paratuberculosis in Holstein cattle (Kirkpatrick et al., 2011), fine-mapping of polymorphisms in the bovine osteopontin gene that are associated with variation in milk yield (Schnabel et al., 2005), or a GWAS for differences in gestation length in Holstein and Brown Swiss cattle (Maltecca et al., 2011), are numerous. Other studies have utilized new technologies of the post-genome era, such as exome sequence analysis and targeted genotyping of single nucleotide polymorphisms (SNPs) associated with failed conception or early embryonic loss (McClure et al., 2014), whereas others have focused on methodological developments, such as studying the effects of population stratification on GWAS results in dairy cattle (Ma et al., 2012), and others have led to improvements in existing practices, such as imputation of microsatellite alleles from dense SNP genotypes for parentage verification (McClure et al., 2012).

How did it Impact Whole Genome (Genomic) Selection?

Without question, the greatest impact of the CDDR collection has been in the development and implementation of genomic selection in dairy cattle using dense, whole-genome SNP markers (VanRaden et al., 2009; Aguilar et al., 2010). The availability of relatively inexpensive high-throughput arrays for SNP genotyping, such as the BovineSNP50 BeadChip (Illumina, San Diego, CA), coupled with 19,000 CDDR samples of AI bulls that had passed or failed progeny testing in the past two decades, and a national database with more than 6 million performance records of the daughters of these bulls, was nothing short of a “perfect storm” for transforming dairy cattle selection programs. Within 13 months of the launch of a commercial bovine SNP array, genomic selection was implemented in practice as a replacement for the traditional pedigree-based genetic evaluation system that had served our industry for nearly a half-century. Predicted transmitting abilities (PTAs) of bulls, cows, heifers, and calves gave way to genomic predicted transmitting abilities (GPTAs), where the latter represent a blending of DNA-based information from genotyped animals with pedigree-based information from non-genotyped animals. The existence of the CDDR provided USDA-ARS scientists and their collaborators with ready access to a large population of reference animals with highly characterized merit for traits of economic performance – the key ingredients to development of a successful genomic selection program. Scientists in several other leading dairy countries were able to develop effective genomic selection programs as well, albeit more slowly, while in other food animal species that lacked DNA repositories much time was spent collecting blood, hair, or tissue samples from phenotypically characterized animals with the hope of building useful reference populations. The US and Canada worked hand-in-hand in development of genomic evaluation systems for North American dairy cattle, largely due to the partnership that was established nearly two decades earlier during formation of the CDDR collection. Although the full impact of genomic selection has not yet been realized, it is clear that genetic trends for many important traits will increase dramatically, due in large part to a halving of generation intervals in the males-to-produce-males, males-to-produce-females, and females-to-produce-males selection pathways (Hutchison et al., 2014). More than 750,000 North American dairy cattle have been genotyped to date using one of nearly two dozen commercially available SNP arrays, and the resulting information flows nearly seamlessly into the US and Canadian national genetic evaluation systems and back to farmers, for use in selection, mating, and culling decisions.

What does the Future Hold?

The CDDR collection has contributed to the success of dozens of projects in its first two decades, most notably the development and implementation of genomic selection programs for North American dairy cattle. Current research efforts are focused on topics such as improvement of the accuracy of genomic predictions using high-density SNP arrays, use of whole-genome sequencing to identify causative mutations affecting fertility and other important traits, and development of genome-based mate selection algorithms. Lastly, CDDR Contributors may wish to identify other situations in which working together collectively can lead to achievements far greater than each individual AI stud can realize independently.

References

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. *Hot topic*: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* 93:743-752.

Hutchison, J. L., J. B. Cole, and D. M. Bickhart. 2014. *Short communication*: Use of young bulls in the United States. *Journal of Dairy Science* 97:3213-3220.

Kirkpatrick, B. W., X. Shi, G. E. Shook, and M. T. Collins. 2011. Whole genome association analysis of susceptibility to paratuberculosis in Holstein cattle. *Animal Genetics* 42:149-160.

Ma, L., G. R. Wiggans, S. Wang, T. S. Sonstegard, J. Yang, B. A. Crooker, J. B. Cole, C. P. Van Tassell, T. J. Lawlor, and Y. Da. 2012. Effect of sample stratification on dairy GWAS results. *BMC Genomics* 13:536.

Maltecca, C. K. A. Gray, K. A. Weigel, J. P. Cassidy, and M. Ashwell. 2011. A genome-wide association study of direct gestation length in US Holstein and Italian Brown populations. *Animal Genetics* 42:585-591.

McClure, M. C., D. Bickhart, D. Null, P. VanRaden, L. Xu, G. Wiggans, G. Liu, S. Schroeder, J. Glasscock, J. Armstrong, J. B. Cole, C. P. Van Tassell, and T. S. Sonstegard. 2014. Bovine exome sequence analysis and targeted SNP genotyping of recessive fertility defects BH1, HH2, and HH3 reveal a putative causative mutation in SMC2 for HH3. *PLOS One* 9:e92769.

McClure, M., T. Sonstegard, G. Wiggans, and C. P. Van Tassell. 2012. Imputation of microsatellite alleles from dense SNP genotypes for parental verification. *Frontiers in Genetics* 3:140.

Schnabel, R. D., J.-J. Kim, M. S. Ashwell, T. S. Sonstegard, C. P. Van Tassell, E. E. Connor, and J. F. Taylor. 2005. Fine-mapping milk production quantitative trait loci on BTA6: Analysis of the bovine osteopontin gene. *Proceedings of the National Academy of Science* 102:6896-6901

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. *Journal of Dairy Science* 92:16-24.