Mastering Mastitis: How Genetics Can Help & Where We Go From Here

Asha M. Miles[1] and Kristen L. Parker Gaddis[2]
[1]Animal Genomics & Improvement Laboratory, USDA-ARS, Beltsville, Maryland, USA
[2]Council on Dairy Cattle Breeding, Bowie, Maryland, USA

Introduction

The USDA's Animal Genomics and Improvement Laboratory (AGIL) has worn many names over the last century; for most of that time the lab calculated genetic evaluations and provided them freely so owners could select genetically superior animals in their breeding programs and improve their herds faster. In 2013, the newly reformulated Council on Dairy Cattle Breeding (CDCB) inherited responsibility for the data stewardship and administration of evaluations while AGIL focused on work as their research partner. National genetic evaluations were first calculated in the U.S. in 1994 for milk somatic cell score (SCS) as a proxy for mastitis resistance in response to the high prevalence of mastitis in our dairy cow population. Even with perfect management and housing, a cow can only perform as well as her genetics allow. This genetically imposed "ceiling" dictates the maximum heights that a cow can reach with regard to her milk quality, mastitis resistance, and a myriad of other economically important traits.

The U.S. Evaluation System

Genetic evaluations on a national scale are only possible through collaboration with key industry groups, and our story really begins in 1908, when the USDA's Bureau of Animal Industry determined to organize cow testing associations nationally (VanRaden and Miller 2008). This eventually led to the formation of the National Dairy Herd Information Association (NDHIA), which remains a key data provider facilitating accurate, credible, and uniform milk recording. Milk records are just one type of phenotypic data that flow to CDCB through Dairy Records Processing Centers (DRPCs). Phenotypes are an individual's observable traits and are usually considered to be the manifestation of that individual's genetic effects (or genotype) along with the animal's environment or management effects. There are several DRPCs across the country that collect and transmit lactation, reproduction, health, calving, test day, yearly average, and herd information data to CDCB, where it is stored in the National Cooperator Database. This database also houses pedigree information reported by breed associations and genomic nominators, which together with phenotypic data forms the basis of conventional genetic evaluations (expressed as Predicted Transmitting Ability (PTA)). The National Association of Animal Breeders (NAAB) provides artificial insemination (AI) codes to track which bulls are available from which companies and monitor sorted status and quality of semen. Cow genotypes for genomic evaluations (GPTA) may only be submitted through a genomic nominator who ensures that a unique animal ID has been assigned to both the actual genotypes and the associated phenotype records. The national U.S. evaluations are also included in international Multiple Across Country Evaluations (MACE) conducted by the Interbull Centre (Uppsala, Sweden). As part of MACE, Interbull evaluates sires of six breeds and seven trait groups comprising milk production, udder health (including SCS and clinical mastitis), conformation, longevity, calving, female fertility, and workability traits from countries throughout the world. Interbull provides sire (G)PTAs to each participating country on their own respective scale
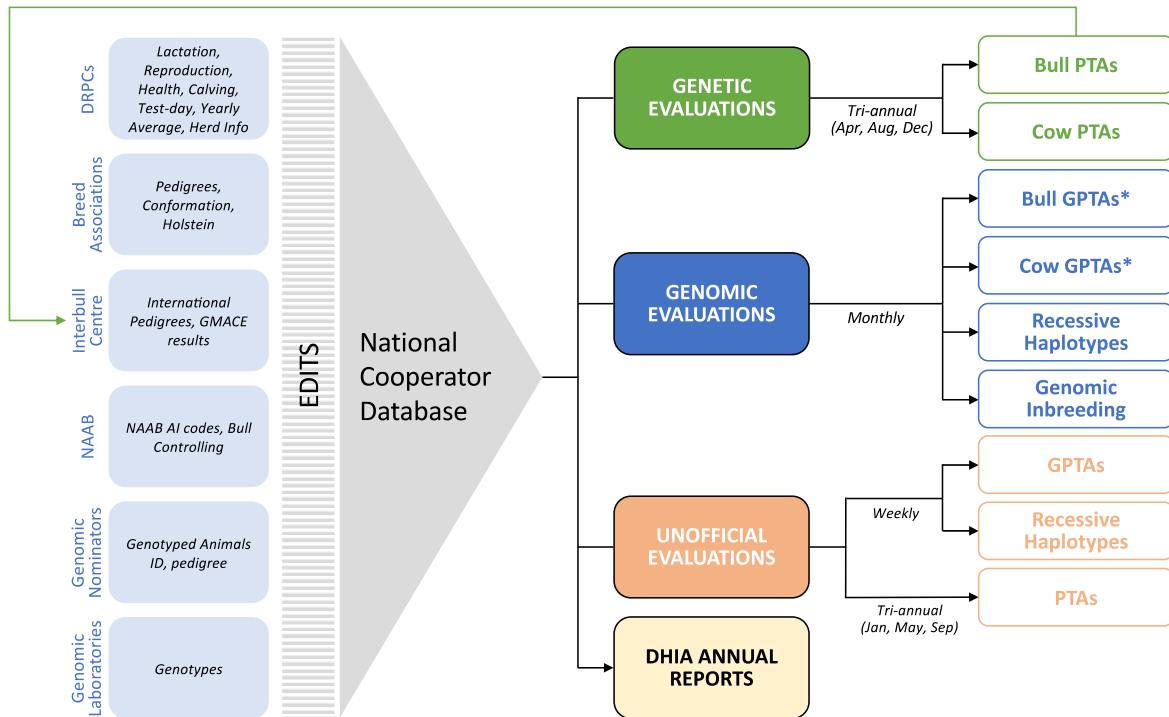
Figure 1. U.S. Genetic and Genomic Evaluation System Process. Data flows from various sources into the National Cooperator Database maintained by CDCB, where programming edits first perform quality control and data standardization checks. Conventional genetic evaluations using pedigrees and genomic evaluations using genotypes are released tri-annually computed with the most recent phenotype data available. Genomic evaluations are also calculated monthly, though do not always include the newest phenotypes as in the tri-annual release due to the time and computational cost of extracting and recalculating with all data. Unofficial evaluations for new animals are released weekly. *Monthly and tri-annual (Apr, Aug, Dec)

which accounts for genotype-by-environment interactions and cows performing better in certain countries (Interbull Centre, 2020). Since August 2019, the CDCB mastitis resistance genetic evaluations for bulls have been exchanged with Interbull participating countries (Parker Gaddis et al. 2020; Mota et al. 2021). This international exchange not only increases the size of the reference population for mastitis resistance evaluations, but also provides independent, external validation of the CDCB evaluation calculations. This system process is described in Figure 1.

Current Considerations

The Problem of Trait Definition

When research on developing genetic evaluations for mastitis resistance began, there were over 130 different acronyms being used in herd management software to describe mastitis. Standardizing extremely disparate health data was a primary obstacle, and it still warrants discussion today. Prior research resulted in the development of an extensive cross-reference

dictionary for the most commonly used abbreviations in herd management software (Cole et al., 2006; Parker Gaddis et al., 2012). This cross-reference dictionary was provided to each DRPC to aid in the development of their own dictionary based on the data provided by their customers. Standardization is performed at the DRPC level using this cross-reference information to assign the single abbreviation of "MAST" before submission to the National Cooperator Database. Although standardization to the acceptable MAST code is now performed for all clinical mastitis reporting, the exact criteria on which an appraiser arrives at that conclusion will likely vary by herd, and even within herd. A general diagnosis of MAST also lacks specificity and does not indicate different mastitis etiologies, whether it was propagated by environmental or contagious pathogens, or whether the pathogen was gram (+) or (-) which can inform the mechanism of host immune response. While a generalized phenotype limits our ability to decipher the functional genetic mechanisms underlying mastitis, it has been sufficient to achieve genetic progress from an animal breeding perspective. The current average rate of mastitis resistance among U.S. Holsteins is an estimated 89.8% (Council on Dairy Cattle Breeding, 2018). This is likely due to the well-documented redundancy built into the immune system (Casadevall and Pirofski 2003; Fischer and Rausell 2016; Miles and Huson 2021); the presence or absence of an overall robust immune system appears to be captured by measuring clinical mastitis incidence.

Even the SCS proxy, a much more objective, quantitative measurement, is not immune to confounding bias. Non-pathogenic causes of inflammation such as lactogenesis and uterine involution during early lactation could inflate mastitis incidence estimates, and different genetic patterns have been associated with SCS at different stages in lactation (Bionaz et al. 2007; Graugnard et al. 2012; Miles and Huson 2020). Alternative SCS traits such as area under the SCS curve, recoverability, standard deviation, and severity may refine SCS as a mastitis indicator trait. Evaluations for SCS are currently computed using the mean test day SCS across the first 305 d of lactation (Council on Dairy Cattle Breeding, 2020). This has been an effective approach as we observe favorable genetic trends for SCS (Figure 2, data from CDCB (2021)),
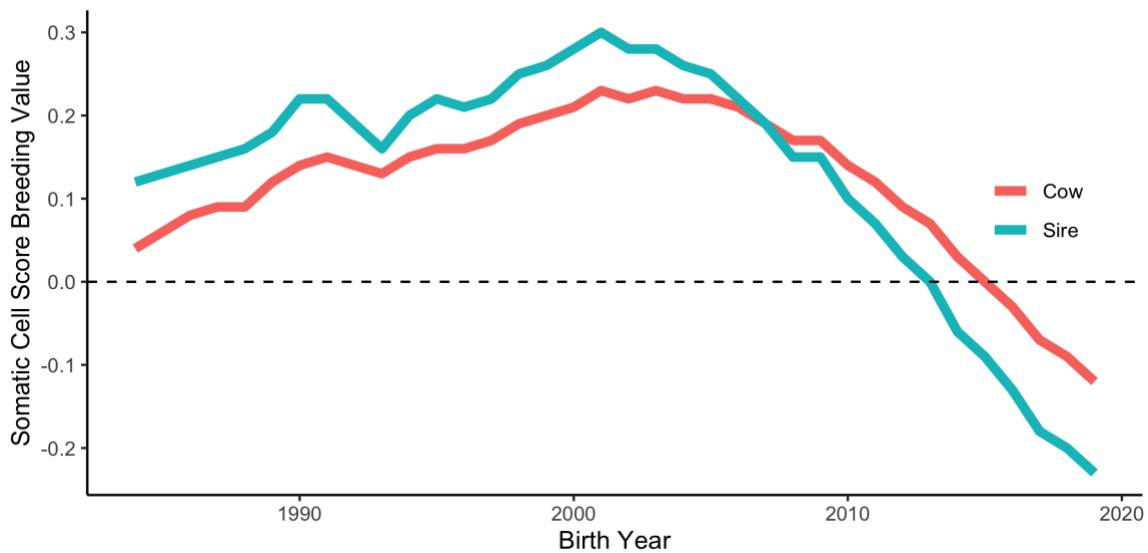

Figure 2. Genetic trends for SCS in Holstein and Red & White. These scores were calculated in December 2021 (CDCB, 2021) and are normalized to the breed average with higher scores being unfavorable (indicating higher SCS).

but there may be opportunity to refine SCS as described above and improve its current antagonistic correlation with milk yield.

Genetic v. Genomic Evaluations

In the pre-genomics era, genetic improvement was achieved from pedigree and phenotypic performance data. Traits like milk and fat production were significantly improved, thanks in part to the widespread use of AI with superior sires who were identified through the evaluation of their many daughters. Genetic selection goals have diversified as additional data become available and/or breeding objectives change, including the development of evaluations for SCS in 1994 and productive life in 2014 (VanRaden et al. 2021). But there are disadvantages to relying on traditional genetic evaluations, primarily time and financial costs. Production records cannot be collected from daughters until they have matured and produced their first calf. This problem compounds itself concerning low heritability traits, like mastitis resistance or other health traits, which are largely attributable to management and other environmental factors. These situations of low heritability require many more records to truly partition the genetic component of the trait from environmental factors, which is key to producing an accurate evaluation.

Following the genomics boom at the turn of the century, the Illumina Bovine Single Nucleotide Polymorphism (SNP)50 BeadChip was released for commercial use in 2007. The SNP50 remains the most popular bovine genotyping platform due to its affordability, contributing to the widespread availability of genomic data and revolutionizing the genetic improvement of dairy animals. Genomic evaluations were first introduced for the U.S. dairy industry shortly after the release of the chip, in 2009 (Wiggans et al. 2011). With genomic information now available at birth, the time to return on investment in genetic selection is significantly decreased. The accuracy of genetic prediction, or "reliability", is also improved with inclusion of genomic information compared to pedigree data alone. Reliabilities corresponding to each evaluation range from 0 to 100% and represent the degree of confidence that can be placed in the evaluation of an animal. They are impacted by the heritability of the trait and the amount of available phenotypic data. The inclusion of genomic information in dairy evaluations resulted in 2- to 3-fold increases in reliability across all traits (García-Ruiz et al. 2016), but was especially important for functional traits like mastitis resistance which have limited available data and a small genetic component compared to the much larger influence of environment. Gains in reliability for mastitis from the inclusion of genomic data average approximately 20 and 40 percentage points in proven and young bulls, respectively (Parker Gaddis et al. 2020).

Developing a Mastitis Resistance Evaluation

Significant hurdles for health traits in the U.S. included the lack of a centralized repository for health data and absence of standardization protocols. Research indicated that genetic selection was feasible through the utilization of health incidences recorded on-farm in management software, and so the National Cooperator Database was modified to accommodate health traits in 2008 (Zwald et al. 2004; Parker Gaddis et al. 2012). Data now flow from participating herds to their respective DRPC where they are standardized into a consistent format and transmitted to the National Cooperator Database at CDCB.

These data from commercial herds usually indicate a heritability of mastitis resistance ranging from 3-10% (Zwald et al. 2004; Koeck et al. 2012; Jamrozik et al. 2013; Vukasinovic et al. 2017; Parker Gaddis et al. 2020), though controlled studies often describe higher heritabilities. Low heritability suggests that most of the variation observed in mastitis resistance is due to non-additive genetic factors, namely management and the environment. But another likely explanation for low heritability is the inherent "noisiness" of the data. This noise may arise from the multiple types of undifferentiated mastitis included in a generalized MAST phenotype, as explained earlier, but some of the noise can be mitigated through multiple levels of strict editing constraints. Before upload to the National Cooperator Database, consistency among pedigree and lactation data is checked. For example, was the date of the MAST event after the cow's fresh date (or even her own birth date)? Do the sire and dam match what is currently known for this cow? Data are either accepted, modified when possible, or returned to the DRPC for correction. Prior to genetic evaluations, several more constraints are applied to include only the most consistent data. These include a minimum incidence constraint (at least 10% of current population incidence) to select herd-years that are reporting MAST events. Likewise, a maximum incidence constraint (not larger than 3 standard deviations greater than the mean population incidence) is imposed to remove data that primarily reflect management practices as opposed to biologically underlined mastitis. More details on the quality control measures ensuring the most reliable health data are used can be found in Parker Gaddis et al. (2020).

Mastitis phenotypes used in genetic evaluations consist of all available data that pass quality control. This currently includes data from 1992 through 2021 resulting in a total of 5.9 million records across all breeds as of December 2021. Healthy cows are defined as present in herd-years reporting MAST events but without any reported case of mastitis during that time. Current evaluations only consider the first report of a MAST event per cow-lactation. Factors such as parity, herd, and season, among others, are accounted for in the model. Mastitis is currently evaluated as a single trait but could be evaluated together with SCS in a bivariate analysis in the future. Several countries already include mastitis and SCS together as part of a multi-trait evaluation. Genomic evaluations incorporate the official set of 78,964 markers that are currently used for all U.S. national evaluations. Animal genotypes from the lower-density bovine SNP50 chip (53,218 markers) are imputed up, where missing markers are inferred based on the haplotypes (sets of markers inherited together) observed in the reference population (VanRaden et al. 2013). Official evaluations are currently only provided for Holstein and Jersey cows and bulls because there is not enough data yet to produce accurate evaluations for other breeds.

Selection Indexes & Mastitis Weighting

Mastitis is just one of many economically important traits that a producer needs to manage. A common strategy is to use a selection index which combines information on many traits into a single value that accounts for their heritability, reliability, economic importance, and intercorrelation. This is beneficial in cases like mastitis, because selecting on index values will identify animals that have both superior production and superior resistance to mastitis, despite the overall unfavorable correlation between yields and mastitis. Since its creation in 1994, the Net Merit (NM$) selection index published by CDCB has become the most widely used tool with which U.S. producers make breeding decisions and is regularly updated to reflect the most current economic values and new traits. Mastitis resistance is included in the NM$ health traits sub-index where it receives 33% of the total emphasis, more than any of the other six health

traits. The health traits sub-index then receives approximately 2% total emphasis in NM$, accounting for the direct costs of adverse health events. Other indirect costs related to mastitis events (e.g., loss of premiums) are accounted for in the considerations for SCS, which is in a separate category from the health sub-index and has a -2.8% emphasis (VanRaden et al. 2021).

Marker Validation & Functional Genetics

The Cattle Quantitative Trait Locus (QTL) Database is a repository documenting progress on the genetic investigation of complex traits. The most recent release (August 2021) reports 2,531 QTL associated with the mastitis phenotypes of raw somatic cell count (SCC), logarithmically transformed SCS, and clinical mastitis (Cattle QTLdb, 2021). The highly complex and polygenic nature of mastitis is emphasized by the significant association of these QTL across all 29 autosomes and the X chromosome. With the abundance of high-density genotype data, prevailing methods of marker identification have shifted from traditional linkage analysis and QTL mapping towards genome-wide association (GWA). GWA methods are comparatively less labor-intensive while offering high resolution, but they are sensitive to underlying population structure. This is particularly challenging in dairy cattle where the overuse of popular sires has contributed to high levels of genomic inbreeding and population structure that is replicated in multiple datasets, complicating independent validation studies. Positive associations in GWA studies may occur because the allele is causative, or if that allele is in linkage disequilibrium with the actual causative marker, or could occur as an artifact of population structure where alleles coincidentally appear in high frequency with traits of interest within a population. Methods to detect and remove spurious associations include imposing stringent multiple testing corrections, haplotype relative risk estimation, the inclusion of identity-by-state matrices in statistical models, and by ensuring repeatability in independent populations (e.g., across breeds) (Forutan et al. 2018; Miles and Huson 2021).

A crucial step in GWA studies is to investigate genes that are annotated in the regions surrounding the associated marker. This analysis will not identify causal variants but will support the theory that the associated marker is important to the trait of interest, especially if the marker is located within a biologically relevant gene. However, restricting gene investigation to biological relevance will mean ignoring most candidate genes and potentially losing a lot of information. Consider a controlled study of a cohort of ~500 U.S. Holsteins that were scored for udder and teat conformational risk factors for mastitis and with all test day SCC and clinical mastitis diagnosis data available (Miles et al. 2019a; Miles and Huson 2020; Miles et al. 2021). A combination of GWA and other population genetics approaches implicated 990 genes near significantly associated markers. A gene ontology analysis performed via the Protein ANalysis THrough Evolutionary Relationships (PANTHER) classification system revealed that only a fraction were protein-coding genes related to the immune response (Figure 3, adapted from Miles (2019b)) (Thomas et al. 2003; Mi et al. 2013). This disparity could be explained by the "omnigenic model" which posits that the genetic architecture of complex traits is produced by a massive regulatory network of genes, each with very small effect (Boyle et al. 2017). The "core" genes whose annotated functions are obviously related to mastitis have a minimal effect compared with the "peripheral" genes, which have non-disease-specific roles in regulating mastitis resistance. The central hypothesis is that almost all genes expressed in a cell type have weak effects on disease expression, and that this universal pleiotropy makes up the majority of the heritability of the trait. The implications are that to get a more complete picture of the genetic
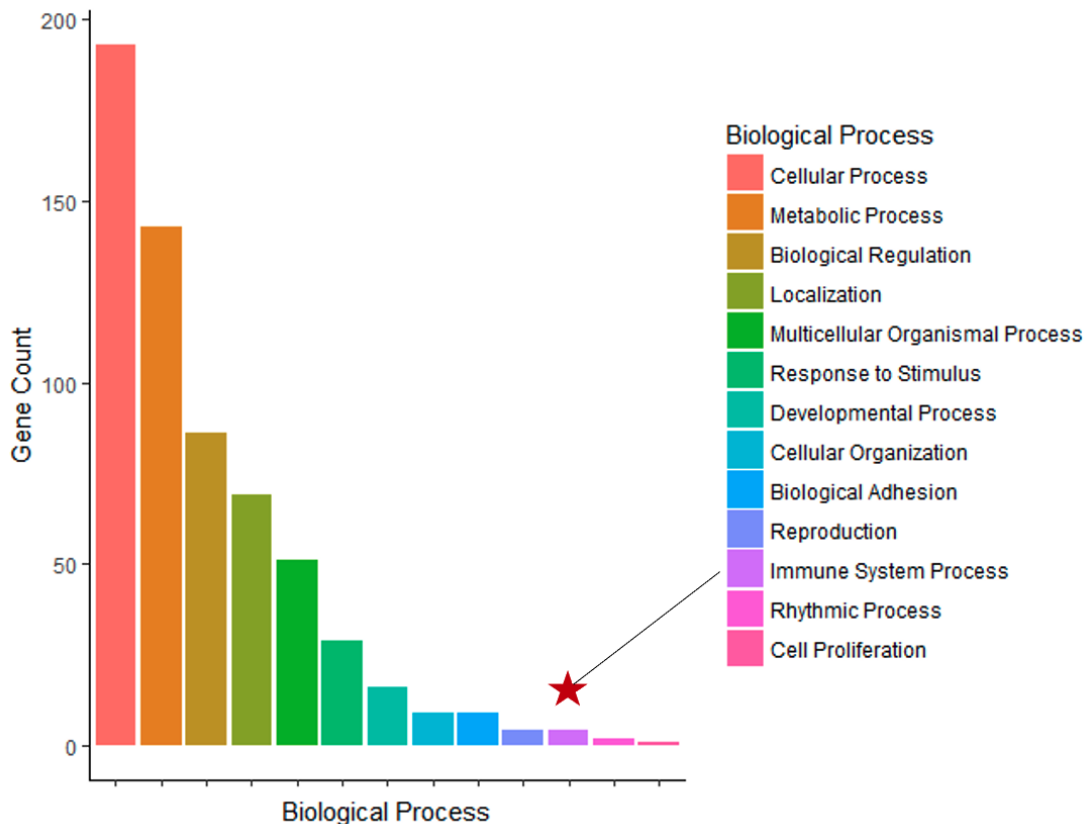
Figure 3. Gene Ontology Summary. Candidate genes implicated in multiple approaches to associating SNPs with SCS and clinical mastitis phenotypes in a controlled cohort study of Holstein cows. Genes were classified in PANTHER by "Biological Process" (the function of the protein in the context of a larger network which accomplishes a process at an organismal level). Figure adapted from (Miles, 2019b); red star indicates category of "core" genes.

mechanisms underlying mastitis we need a clearer understanding of cellular networks and regulatory function. This may feed back into genetic progress attributable to selection strategies by prioritizing SNP markers used in genomic selection based on biological insight and the downstream effects of marker variants. Generating this kind of information would require integrated -omics approaches which examine marker variants, gene expression, and protein function. The 78,964 SNP markers currently used in official U.S. evaluations were selected based on minor allele frequency, parent-progeny conflicts, and call rate, as described in Wiggans et al. (2016). This SNP list is updated as new information and markers become available.

Challenges & Opportunities

Computational Considerations

When the majority of available phenotypes are from cows who were "pre-selected" according to their genomic merit, biases and inaccuracies are introduced into evaluation models (Aguilar et al. 2020). Statistical methods called single-step Genomic Best Linear Unbiased Prediction (ssGBLUP) include all three data types (pedigree, phenotype, and genotype) simultaneously so genomic pre-selection is accounted for in predictions. These methods were developed a decade

ago but were too computationally demanding to apply to "big data". Refined models with more efficient algorithms are emerging that may enable the practical application of ssGBLUP on a national scale (Masuda et al. 2016; Mäntysaari et al. 2020). Computational limitations also apply to trait derivations calculated prior to genetic evaluation. Rigorous testing will be required to determine the feasibility of alternative traits derived from SCS, or using mixed linear models that incorporate repeated measures, or adding multiple categories to clinical mastitis trait definitions to capture pathogen-specific resistance or susceptibility. The challenge that arises is that all of these factors must be resolved while delivering these evaluations on the time schedule described in Figure 1.

Leveraging the Microbiome

A microbiome is the aggregate of all microbes (and their genetic material) that inhabits a unique habitat and has distinct properties. There has been much excitement about the various microbiomes of the human and their far-reaching impacts. The milk microbiome (and perhaps whether it even exists) is a rapidly expanding and rather contested field (Quigley et al. 2013; Rainard 2017; Taponen et al. 2019). There is a general lack of reproducibility of the results of microbiome studies, and consequently, we observe a growing impetus for complete data and methodology transparency in this field (Schloss and Ravel 2018; Moossavi et al. 2021). Nearly every stage of microbiome studies, from sample collection through bioinformatic analysis, is vulnerable to failing to produce repeatable results. This is particularly challenging for low-biomass samples like milk which have relatively small microbial populations. Environmental contamination from the farm or laboratory will be much more evident in milk than in fecal or rumen samples that are so rich in microbes that any contamination will appear in relatively low abundance. This problem can be tempered by including appropriate negative controls and statistical methods to identify and remove contaminant sequences during bioinformatic analysis have been developed (Davis et al. 2018). Other inconsistencies can be partially attributed to limitations in both sequencing and analysis methodology. Datasets generated by high-throughput sequencing have an arbitrary "total" imposed by the instrument. This total does not reflect the total microbial load of the sample, so computing changes in abundance or diversity across samples without accounting for the compositional nature of the data can introduce significant bias (Gloor et al. 2017). Methods have been developed to reduce bias in microbiome analysis, such as comparing the ratio of taxa (Morton et al. 2019) or addressing the sparsity issues common in microbiome data by only including non-zeros (Martino et al. 2019). These barriers to robust microbiome investigations are greatly reduced as guiding reviews detailing best practices are being published at an exponential rate (Knight et al. 2018; Bharti and Grimm 2019; Bokulich et al. 2020; Weinroth et al. 2021). As sequencing costs drop and microbiome studies become more accessible to fields of study outside genomics (e.g., physiology), the ability to understand host-microbe interactions will transform these disciplines. This moment is reminiscent of the genomics boom in the early 2000s.

As it stands, microbiome insights may be more practical for on-farm interventions than genetic selection because selecting for cows with favorable milk microbial profiles would require the mass phenotyping of microbiome-based traits. The goal of genomic selection research is to maximize the amount of information that can be predicted at birth from the same inexpensive DNA sample and set of genotypes; incorporating microbiome data for individual animals would significantly increase the costs of this application of genomic evaluations. Strategies like pooling

animals in contemporary groups prior to sequencing have been proposed, but this approach risks losing considerable resolution concerning host-microbiome interactions and our ability to describe the microbiome as a functional unit. Financial constraints aside, significant groundwork would be required to develop pipelines for high-quality, standardized microbiome data to flow into the evaluations system described earlier.

Some animal breeders are already exploring the integration of host and microbial genomics into their phenotype estimations. In addition to heritability (proportion of variability in trait explained by the host genome), researchers are observing "microbiability" and finding that variation in the trait can also be described by variation in the microbiome (Wen et al. 2019; Maltecca et al. 2020; Khanal et al. 2021). Microbial influence on their host has become such a widely recognized phenomenon that the term "holobiont" was coined to describe the host and all of its microbial symbionts. This theory suggests that together the host and microbiome make up a single unit of biological organization, which has significant implications for our understanding of host ecology and evolution (Bordenstein and Theis 2015). It also has been well-established that taxonomic composition may vary over the host's lifetime and physiological states (Ley et al. 2006; Greenhalgh et al. 2016; Dibner et al. 2021), but biochemical functions and their molecular signatures are more heavily conserved. Moreover, these collective signatures are what are responsible for observed functional effects, not the individual bugs. Much of the existing literature has focused on milk microbiome composition, not on its overall functionality. A more impactful approach to considering milk microbiome influence would be to select on molecular signatures, not individual taxa. As Doolittle and Booth (2017) wrote, "it's the song, not the singer". Coupling classical microbiome research with transcriptomics, metabolomics, etc. can provide new insight into the mechanistics of mastitis resistance and milk quality and funding this type of work would amplify opportunities for scientific innovation and real-world applicability.

Phenomics & Big Data

Genomic selection can only improve what we can measure, and as dairy production systems become increasingly complex, so do the requirements of accurately measuring animal performance. Many researchers are interested in working on traits likely to be economically important, but producers are not incentivized to contribute data because it will not directly impact farm profits (Cole et al. 2020). Enter precision agriculture. Precision livestock monitoring technologies primarily take the form of economically justifiable sensor systems which can be wearable (i.e., boluses; neck, leg, or ear tags) or mounted (i.e., a stationary camera, robotic milkers). Measurable traits include udder health, animal locomotion, heart rate, rumen pH, feed intake, and behavior, just to name a few (Halachmi et al. 2019). Autonomous, sensor-based phenotyping represents a huge opportunity to increase the number of traits that can be improved through genomic selection. These sensors generate vast amounts of information but translating this rapidly generated data into something that can be used in on-farm decision-making, and eventually genomic selection, will be a new challenge.

Mastitis data generated by automatic on-farm sensors for estimating milk composition are a good example of this problem. No standard practices for users or validation, maintenance, and calibration protocols have been established which creates both substantial system bias and individual sensor bias (Sievert, 2019). An important caveat is that these systems have been designed to provide information that prompts a management outcome, not to create a data

repository available for research use. Sensors that generate alerts from SCC measurements taken during initial letdown are very helpful for improving management and the quality of milk sold, but they are not representative of the entire milk flow and may not provide appropriate phenotypes on which to select. Furthermore, these automated systems rely on automatic animal ID detection during which there is ample opportunity for IDs to be mismatched with phenotypes. And because these systems are designed to function as a management tool, data are only stored for a short time before they are overwritten. Data storage, flow, quality control, and quality assurance standards need to be established before they can be used on a national scale, and even their on-farm use should be interpreted with caution.

No standards currently exist for sharing sensor-generated data which may limit their widespread use and potential benefits. Some companies appear to believe that because they own the technology, they will own the data. That philosophy is in direct contradiction to our current state of affairs where CDCB serves as a steward of herd data, but sole ownership and rights pertaining thereto remain with the producer. This issue of sensor data ownership is evocative of recent controversies surrounding John Deere and Tesla, backed by the "Right to Repair" movement. Extended Use License Agreements are usually complex and full of legalese that limits the use of equipment; these are signed at time of purchase and users often do not fully understand their ramifications. Similarly, frequent software and technology updates create problems for so-called legacy equipment as repairs and maintenance are not available for older versions. The big questions we face about sensor-based data are how we will standardize it and who can use it. Suffice to say, our biggest challenges in the coming years will likely be legal, not scientific.

Conclusions

TL;DR (Too Long; Didn't Read)

1) Genetic and genomic evaluations are delivered thanks to the cooperation of key industry partners through a complex system of data sharing and management
2) Standardizing definitions of mastitis across heterogeneous data sources and accurately describing the underlying biology of the trait are key challenges
3) Including genomic data has exponentially increased the rate of genetic gain compared to conventional genetic evaluations
4) Evaluations for SCS have been available since 1994 and for clinical mastitis from 2018, the latter of which required developing many new data quality control measures
5) Using a selection index like NM$ is an effective way to select for overall high performance and control for the unfavorable correlation of mastitis and milk yield
6) Examining the downstream functional effects of marker variants may help prioritize SNPs for inclusion in evaluations, but complex gene networks muddy the waters
7) ssGBLUP can mitigate biases in evaluations that arise from animals "pre-selected" on the basis of their genetic merit; faster computer algorithms are a must
8) Milk microbiome work focusing on collective molecular signatures and their effects can generate new insight into the mechanisms of mastitis resistance and milk quality
9) Among the challenges of using automatic sensor-based milk data are quality assurance within and across systems and potential legal controversy regarding data ownership and use

References

Aguilar, I., E. N. Fernandez, A. Blasco, O. Ravagnolo, and A. Legarra. 2020. 'Effects of ignoring inbreeding in model-based accuracy for BLUP and SSGBLUP', *Journal of Animal Breeding and Genetics*, 137: 356-64.

Bharti, R., and D. G. Grimm. 2019. 'Current challenges and best-practice protocols for microbiome analysis', *Briefings in Bioinformatics*, 22: 178-93.

Bionaz, M., E. Trevisi, L. Calamari, F. Librandi, A. Ferrari, and G. Bertoni. 2007. 'Plasma Paraoxonase, Health, Inflammatory Conditions, and Liver Function in Transition Dairy Cows', *Journal of Dairy Science*, 90: 1740-50.

Bokulich, N. A., M. Ziemski, M. S. Robeson, and B. D. Kaehler. 2020. 'Measuring the microbiome: Best practices for developing and benchmarking microbiomics methods', *Computational and Structural Biotechnology Journal*, 18: 4048-62.

Bordenstein, S. R., and K. R. Theis. 2015. 'Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes', *PLOS Biology*, 13: e1002226.

Boyle, E. A., Y. I. Li, and J. K. Pritchard. 2017. 'An Expanded View of Complex Traits: From Polygenic to Omnigenic', *Cell*, 169: 1177-86.

Casadevall, A., and L. Pirofski. 2003. 'Exploiting the Redundancy in the Immune System : Vaccines Can Mediate Protection by Eliciting 'Unnatural' Immunity', *Journal of Experimental Medicine*, 197: 1401-04.

Cattle QTLdb. 2021. QTL/associations by trait type. Accessed 12/14/2021. https://www.animalgenome.org/cgi-bin/QTLdb/BT/summary?summ=type&qtl=163,720&pub=1,069&trait=684&gene=185

Council on Dairy Cattle Breeding. 2018. Trait Reference Sheet: Resistance to Mastitis (MAST). Accessed 12/15/2021. https://www.uscdcb.com/wp-content/uploads/2018/03/CDCB-Reference-Sheet-MAST-03_2018.pdf

Council on Dairy Cattle Breeding. 2020. Description of National Genetic Evaluation Systems. Accessed 12/14/2021. https://www.uscdcb.com/wp-content/uploads/2020/04/USAudder.pdf

Council on Dairy Cattle Breeding. 2021. Trend in Somatic Cell Score for Holstein or Red & White. Accessed 12/14/2021. https://queries.uscdcb.com/eval/summary/trend.cfm?R_Menu=HO.s#StartBody

Cole, J. B., and P. M. VanRaden. 2006. Genetic Evaluation and Best Prediction of Lactation Persistency. *Journal of Dairy Science*, 89:2722-2728.

Cole, J. B., S. A. E. Eaglen, C. Maltecca, H. A. Mulder, and J. E. Pryce. 2020. 'The future of phenomics in dairy cattle breeding', *Animal frontiers: the review magazine of animal agriculture*, 10: 37-44.

Davis, N. M., D. M. Proctor, S. P. Holmes, D. A. Relman, and B. J. Callahan. 2018. 'Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data', *Microbiome*, 6: 226-26.

Dibner, R. R., A. M. Weaver, M. T. Brock, G. F. Custer, H. G. Morrison, L. Maignien, and C. Weinig. 2021. 'Time outweighs the effect of host developmental stage on microbial community composition', *FEMS Microbiology Ecology*, 97: fiab102.

Doolittle, W. F., and A.Booth. 2017. 'It's the song, not the singer: an exploration of holobiosis and evolutionary theory', *Biology & Philosophy*, 32: 5-24.

Fischer, A., and A. Rausell. 2016. 'Primary immunodeficiencies suggest redundancy within the human immune system', *Science Immunology*, 1: eaah5861.

Forutan, M., S. A. Mahyari, C. Baes, N. Melzer, F. S. Schenkel, and M. Sargolzaei. 2018. 'Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle', *BMC Genomics*, 19: 98.

García-Ruiz, A., J. B. Cole, P. M. VanRaden, G. R. Wiggans, F. J. Ruiz-López, and C. P. Van Tassell. 2016. 'Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection', *Proceedings of the National Academy of Sciences*, 113: E3995-E4004.

Gloor, G. B., J. M. Macklaim, V. Pawlowsky-Glahn, and J. J. Egozcue. 2017. 'Microbiome Datasets Are Compositional: And This Is Not Optional', *Front Microbiol*, 8: 2224.

Graugnard, D. E., M. Bionaz, E. Trevisi, K. M. Moyes, J. L. Salak-Johnson, R. L. Wallace, J. K. Drackley, G. Bertoni, and J. J. Loor. 2012. 'Blood immunometabolic indices and polymorphonuclear neutrophil function in peripartum dairy cows are altered by level of dietary energy prepartum1', *Journal of Dairy Science*, 95: 1749-58.

Greenhalgh, K., K. M. Meyer, K. M. Aagaard, and P. Wilmes. 2016. 'The human gut microbiome in health: establishment and resilience of microbiota over a lifetime', *Environmental Microbiology*, 18: 2103-16.

Halachmi, I., M. Guarino, J. Bewley, and M. Pastell. 2019. 'Smart Animal Agriculture: Application of Real-Time Sensors to Improve Animal Well-Being and Production', *Annu Rev Anim Biosci*, 7: 403-25.

Interbull Centre. 2020. Interbull Code of Practice. Accessed 7/1/2020. https://interbull.org/ib/codeofpractice.

Jamrozik, J., A. Koeck, F. Miglior, G. J. Kistemaker, F. Schenkel, D. F. Kelton, and B. Van Doormaal. 2013. 'Genetic and genomic evaluation of mastitis resistance in Canada', *Proceedings of the 2013 Interbull Meeting*, 47: 43-51.

Khanal, P., C. Maltecca, C. Schwab, J. Fix, and F. Tiezzi. 2021. 'Microbiability of meat quality and carcass composition traits in swine', *Journal of Animal Breeding and Genetics*, 138: 223-36.

Knight, R., A. Vrbanac, B. C. Taylor, A.Aksenov, C. Callewaert, J. Debelius, A. Gonzalez, T. Kosciolek, L. McCall, D. McDonald, A. V. Melnik, J. T. Morton, J. Navas, R. A. Quinn, J. G. Sanders, A. D. Swafford, L. R. Thompson, A. Tripathi, Z. Z. Xu, J. R. Zaneveld, Q. Zhu, J. G. Caporaso, and P. C. Dorrestein. 2018. 'Best practices for analysing microbiomes', *Nature Reviews Microbiology*, 16: 410-22.

Koeck, A., F. Miglior, D. F. Kelton, and F. S. Schenkel. 2012. 'Short communication: Genetic parameters for mastitis and its predictors in Canadian Holsteins', *Journal of Dairy Science*, 95: 7363-66.

Ley, R. E., D. A. Peterson, and J. I. Gordon. 2006. 'Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine', *Cell*, 124: 837-48.

Maltecca, C., M. Bergamaschi, and F. Tiezzi. 2020. 'The interaction between microbiome and pig efficiency: A review', *Journal of Animal Breeding and Genetics*, 137: 4-13.

Mäntysaari, E. A., M. Koivula, and I. Strandén. 2020. 'Symposium review: Single-step genomic evaluations in dairy cattle', *J Dairy Sci*, 103: 5314-26.

Martino, C., T., J. Morton, C. A. Marotz, L. R. Thompson, T. Anupriya, R. Knight, K. Zengler, and J. D. Neufeld. 2019. 'A Novel Sparse Compositional Technique Reveals Microbial Perturbations', *mSystems*, 4: e00016-19.

Masuda, Y., I. Misztal, S. Tsuruta, A. Legarra, I. Aguilar, D. A. L. Lourenco, B. O. Fragomeni, and T. J. Lawlor. 2016. 'Implementation of genomic recursions in single-step genomic best linear unbiased predictor for US Holsteins with a large number of genotyped animals', *Journal of Dairy Science*, 99: 1968-74.

Mi, H., A. Muruganujan, and P. D. Thomas. 2013. 'PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees', *Nucleic Acids Res*, 41: D377-86.

Miles, A. M., J. A. A. McArt, F. A. Leal Yepes, C. R. Stambuk, P. D. Virkler, and H. J. Huson. 2019a. 'Udder and teat conformational risk factors for elevated somatic cell count and clinical mastitis in New York Holsteins', *Prev Vet Med*, 163: 7-13.

Miles, A. M. 2019b. 'Understanding the Genetics Underlying Mastitis Using a Multi-Pronged Approach'. *Cornell Theses and Dissertations*. https://doi.org/10.7298/qq63-3f06.

Miles, A. M., C. J. Posbergh, and H. J. Huson. 2021. 'Direct Phenotyping and Principal Component Analysis of Type Traits Implicate Novel QTL in Bovine Mastitis through Genome-Wide Association', *Animals*, 11.

Miles, A. M., and H. J. Huson. 2020. 'Time- and population-dependent genetic patterns underlie bovine milk somatic cell count', *Journal of Dairy Science*, 103: 8292-304.

Miles, A.M., and H. J. Huson. 2021. 'Graduate Student Literature Review: Understanding the genetic mechanisms underlying mastitis*', *Journal of Dairy Science*, 104: 1183-91.

Moossavi, S., K. Fehr, E. Khafipour, and M. B. Azad. 2021. 'Repeatability and reproducibility assessment in a large-scale population-based microbiota study: case study on human milk microbiota', *Microbiome*, 9: 41.

Morton, J. T., C. Marotz, A. Washburne, J. Silverman, L. S. Zaramela, A. Edlund, K. Zengler, and R. Knight. 2019. 'Establishing microbial composition measurement standards with reference frames', *Nature Communications*, 10: 2719.

Mota, R. R., E. Nicolazzi, J. Megonigal, K. L. Parker Gaddis, J. W. Durr, and P. M. VanRaden. 2021. 'Use of international clinical mastitis data as independent trait in the US evaluation system', *Proceedings of the 2021 Interbull Meeting*, 56: 11-16.

Parker Gaddis, K. L., J. B. Cole, J. S. Clay, and C. Maltecca. 2012. 'Incidence validation and relationship analysis of producer-recorded health event data from on-farm computer systems in the United States', *J Dairy Sci*, 95: 5422-35.

Parker Gaddis, K. L., P. M. VanRaden, J. B. Cole, H. D. Norman, E. Nicolazzi, and J. W. Dürr. 2020. 'Symposium review: Development, implementation, and perspectives of health evaluations in the United States', *J Dairy Sci*, 103: 5354-65.

Quigley, L., O. O'Sullivan, C. Stanton, T. P. Beresford, R. P. Ross, G. F. Fitzgerald, and P. D. Cotter. 2013. 'The complex microbiota of raw milk', *FEMS Microbiology Reviews*, 37: 664-98.

Rainard, P. 2017. 'Mammary microbiota of dairy ruminants: fact or fiction?', *Veterinary Research*, 48: 25.

Schloss, P. D., and J. Ravel. 2018. 'Identifying and Overcoming Threats to Reproducibility, Replicability, Robustness, and Generalizability in Microbiome Research', *mBio*, 9: e00525-18.

Sievert, S. J. 2019. 'On-Farm Sensors for Estimation of Milk Composition - Where Will These Devices & Data Fit in DHI Programs in the Future?'. Accessed 12/14/2021. http://www.quality-certification.com/advcomm/2019%20LAC/Sievert%20-%20NALMA%20Milk%20Sensor%20Summary.pdf

Taponen, S., D. McGuinness, H. Hiitiö, H. Simojoki, R. Zadoks, and S. Pyörälä. 2019. 'Bovine milk microbiome: a more complex issue than expected', *Veterinary Research*, 50: 44.

Thomas, P. D., M. J. Campbell, A. Kejariwal, H. Mi, B. Karlak, R. Daverman, K. Diemer, A. Muruganujan, and A. Narechania. 2003. 'PANTHER: a library of protein families and subfamilies indexed by function', *Genome Res*, 13: 2129-41.

VanRaden, P. M., and R. Miller. 2008. "The USDA Animal Improvement Programs Laboratory: A century old and just getting started." In, edited by United States Department of Agriculture. AIPL Research Reports.

VanRaden, P. M., D. J. Null, M. Sargolzaei, G. R. Wiggans, M. E. Tooker, J. B. Cole, T. S. Sonstegard, E. E. Connor, M. Winters, J. B. C. H. M. van Kaam, A. Valentini, B. J. Van Doormaal, M. A. Faust, and G. A. Doak. 2013. 'Genomic imputation and evaluation using high-density Holstein genotypes', *Journal of Dairy Science*, 96: 668-78.

VanRaden, P. M., J. B. Cole, M. Neupane, S. Toghiani, K. L. Parker Gaddis, and R. J. Tempelman. 2021. "Net merit as a measure of lifetime profit: 2021 revision." In, edited by United States Department of Agricutlure. AIPL Research Reports.

Vukasinovic, N., N. Bacciu, C. A. Przybyla, P. Boddhireddy, and S. K. DeNise. 2017. 'Development of genetic and genomic evaluation for wellness traits in US Holstein cows', *Journal of Dairy Science*, 100: 428-38.

Wen, C. L., C. J. Sun, and N. Yang. 2019. 'The concepts and research progress: from heritability to microbiability', *Yi chuan = Hereditas*, 41: 1023-40.

Weinroth, M.D., Belk A.D., Dean C., Noyes N., Dittoe D.K., Rothrock M.J., Ricke S.C., Myer P.R., Henniger M.T., Ramirez G.A., Oakley B.B., Summers K.L., Miles A.M., Ault-Seay T.B., Yu Z., Metcalf J., and J. Wells. Considerations and best practices in animal science 16S rRNA gene sequencing microbiome studies. *Journal of Animal Science *in press**

Wiggans, G. R., P. M. Vanraden, and T. A. Cooper. 2011. 'The genomic evaluation system in the United States: past, present, future', *J Dairy Sci*, 94: 3202-11.

Wiggans, G. R., T. A. Cooper, P. M. VanRaden, C. P. Van Tassell, D. M. Bickhart, and T. S. Sonstegard. 2016. 'Increasing the number of single nucleotide polymorphisms used in genomic evaluation of dairy cattle', *J Dairy Sci*, 99: 4504-11.

Zwald, N. R., K. A. Weigel, Y. M. Chang, R. D. Welper, and J. S. Clay. 2004. 'Genetic selection for health traits using producer-recorded data. I. Incidence rates, heritability estimates, and sire breeding values', *J Dairy Sci*, 87: 4287-94.