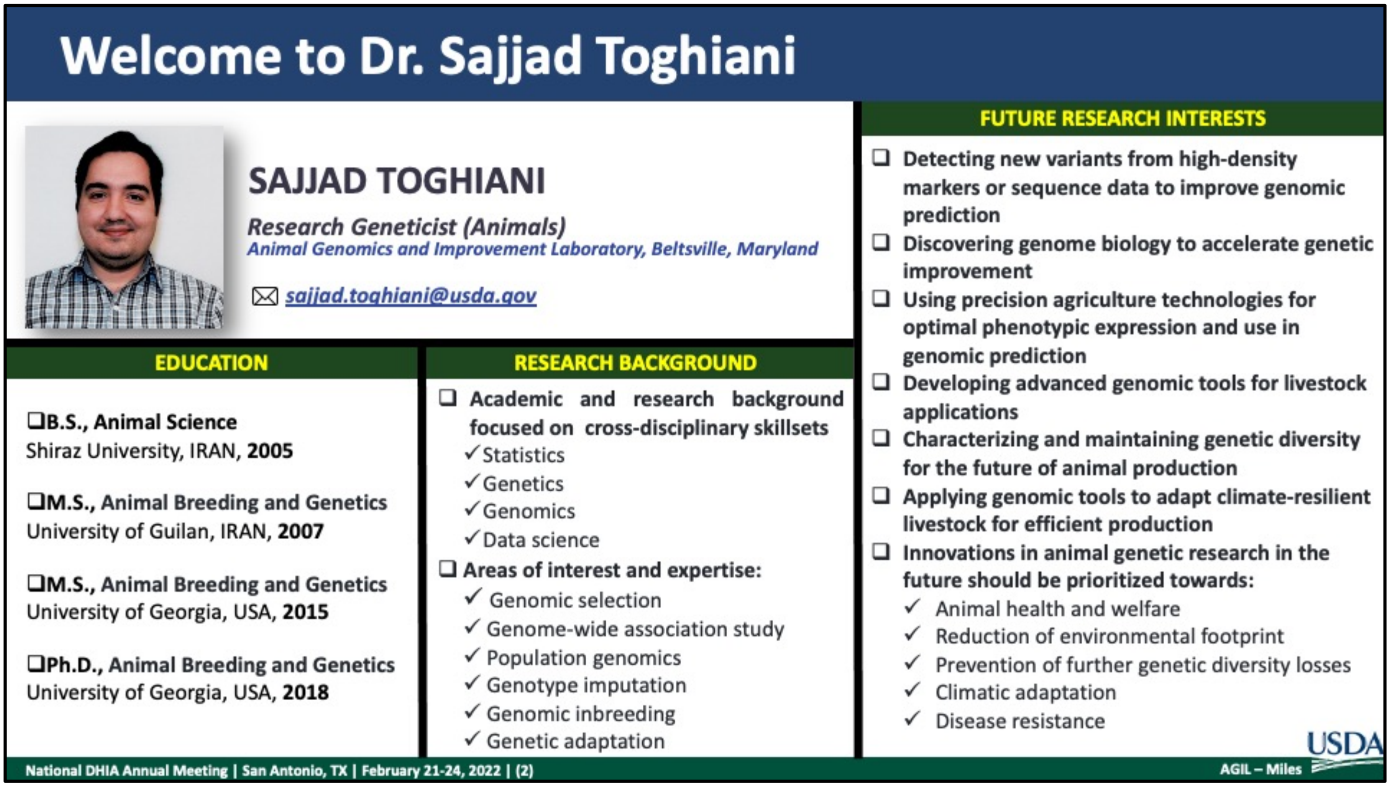
Frontiers in Animal Breeding & Genetics

Asha M. Miles, PhD | Research Geneticist

Animal Genomics & Improvement Laboratory (AGIL)
USDA Agricultural Research Service
Beltsville, MD 20705
asha.miles@usda.gov



Welcome to Dr. Sajjad Toghiani

SAJJAD TOGHIANI

Research Geneticist (Animals)

Animal Genomics and Improvement Laboratory, Beltsville, Maryland

✉ sajjad.toghiani@usda.gov

EDUCATION

☐ **B.S., Animal Science**
Shiraz University, IRAN, **2005**

☐ **M.S., Animal Breeding and Genetics**
University of Guilan, IRAN, **2007**

☐ **M.S., Animal Breeding and Genetics**
University of Georgia, USA, **2015**

☐ **Ph.D., Animal Breeding and Genetics**
University of Georgia, USA, **2018**

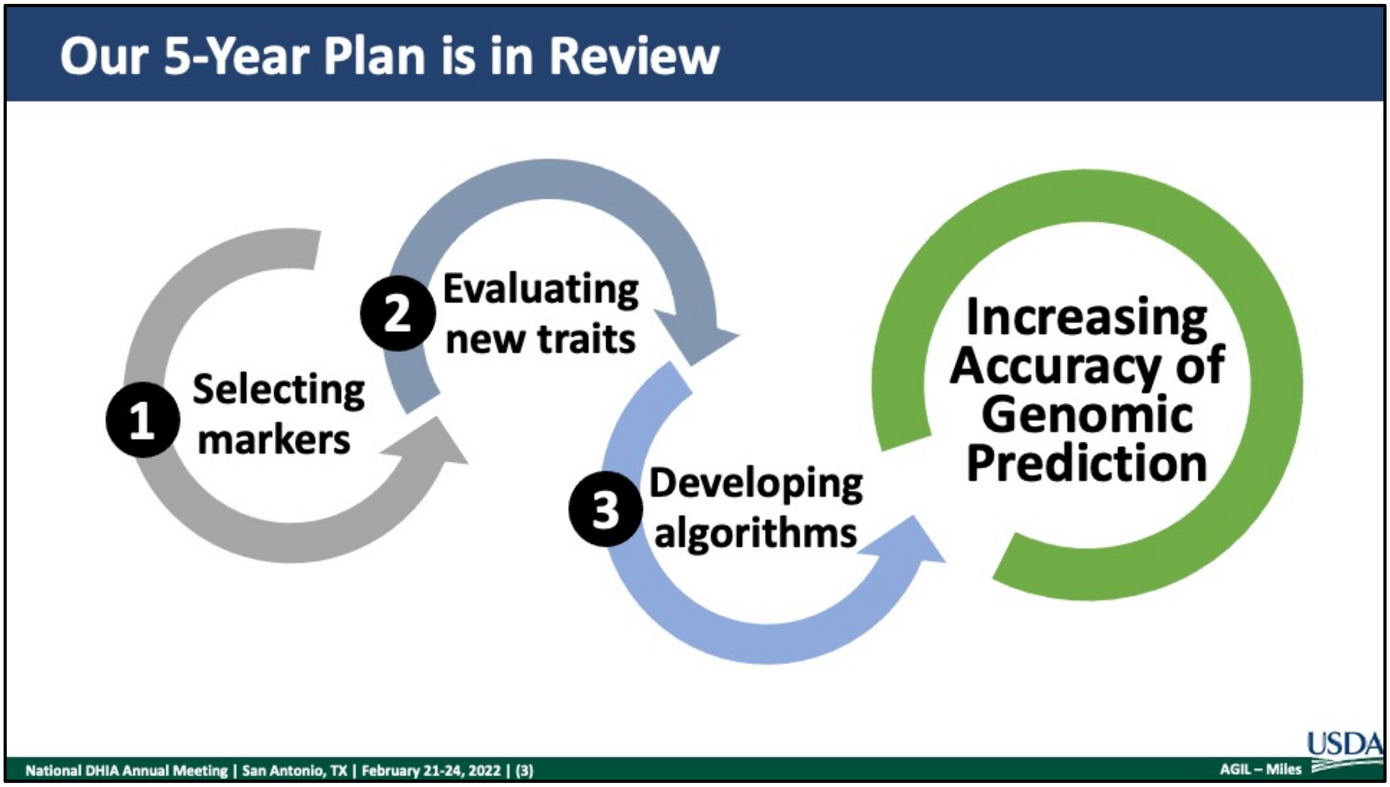
RESEARCH BACKGROUND

- ☐ **Academic and research background focused on cross-disciplinary skillsets**
  - ✓ Statistics
  - ✓ Genetics
  - ✓ Genomics
  - ✓ Data science
- ☐ **Areas of interest and expertise:**
  - ✓ Genomic selection
  - ✓ Genome-wide association study
  - ✓ Population genomics
  - ✓ Genotype imputation
  - ✓ Genomic inbreeding
  - ✓ Genetic adaptation

FUTURE RESEARCH INTERESTS

- ☐ **Detecting new variants from high-density markers or sequence data to improve genomic prediction**
- ☐ **Discovering genome biology to accelerate genetic improvement**
- ☐ **Using precision agriculture technologies for optimal phenotypic expression and use in genomic prediction**
- ☐ **Developing advanced genomic tools for livestock applications**
- ☐ **Characterizing and maintaining genetic diversity for the future of animal production**
- ☐ **Applying genomic tools to adapt climate-resilient livestock for efficient production**
- ☐ **Innovations in animal genetic research in the future should be prioritized towards:**
  - ✓ Animal health and welfare
  - ✓ Reduction of environmental footprint
  - ✓ Prevention of further genetic diversity losses
  - ✓ Climatic adaptation
  - ✓ Disease resistance

Firstly, I would like to welcome and introduce you to Dr. Sajjad Toghiani, who accepted a permanent position on our team of scientists just last month. He completed is PhD in Animal Breeding and Genetics @ U of Georgia and has been with USDA for 4 years, the last 2 of which he worked for AGIL as a support scientist. We are very excited about his enthusiasm and expertise and look forward to sharing his work with you in the future.



Our 5-Year Plan is in Review

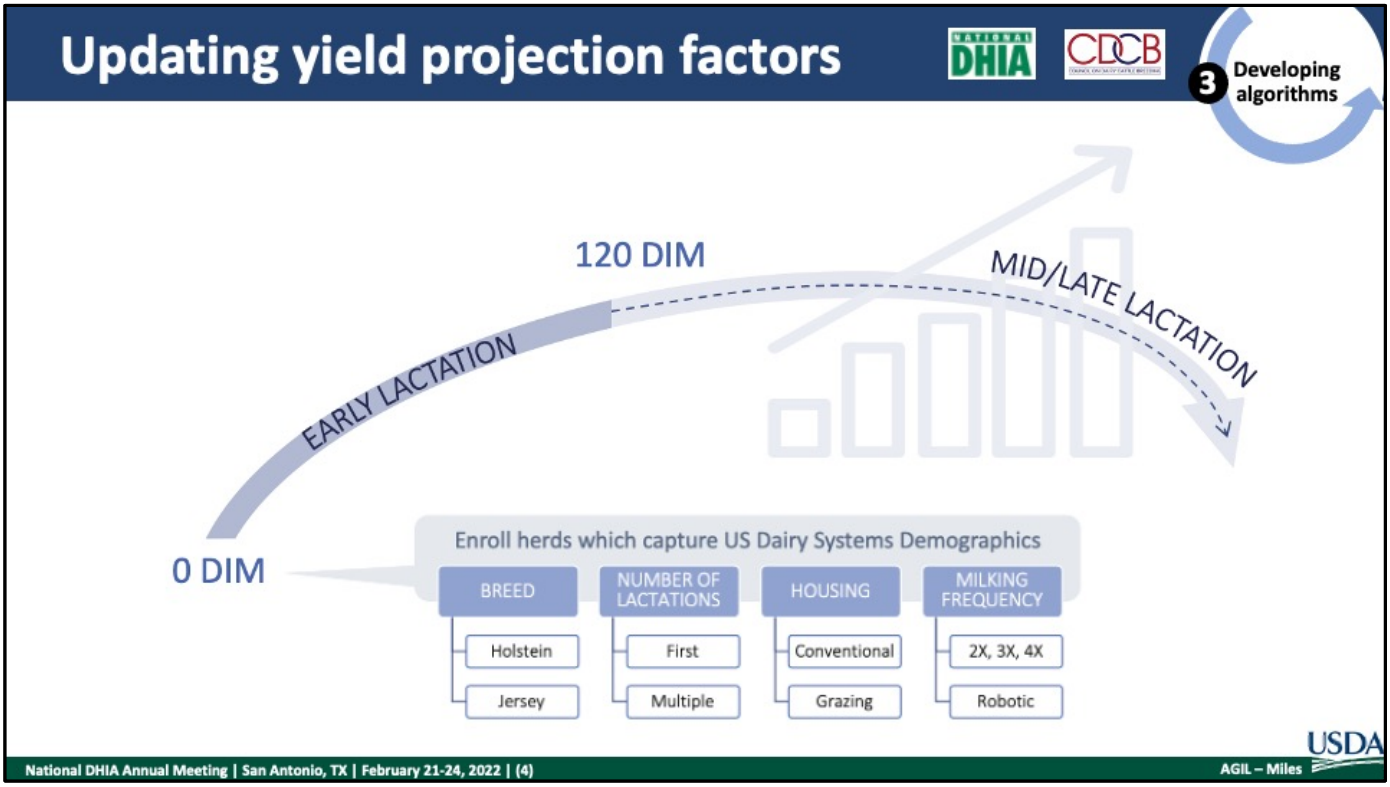
1 Selecting markers

2 Evaluating new traits

3 Developing algorithms

Increasing Accuracy of Genomic Prediction

Every 5 years, laboratories in the ag research service submit a research plan that goes through a peer-review process similar to a grant panel and is how we receive our funding allocation from Congress. Last year we collected feedback through various stakeholder meetings and have outlined research priorities that reflect the needs of the US dairy industry. The plan is structured by three objectives all in the pursuit of increasing the accuracy of genomic prediction. This talk will cover some highlights from the proposed plan.



Updating yield projection factors

3 Developing algorithms

120 DIM

EARLY LACTATION

MID/LATE LACTATION

0 DIM

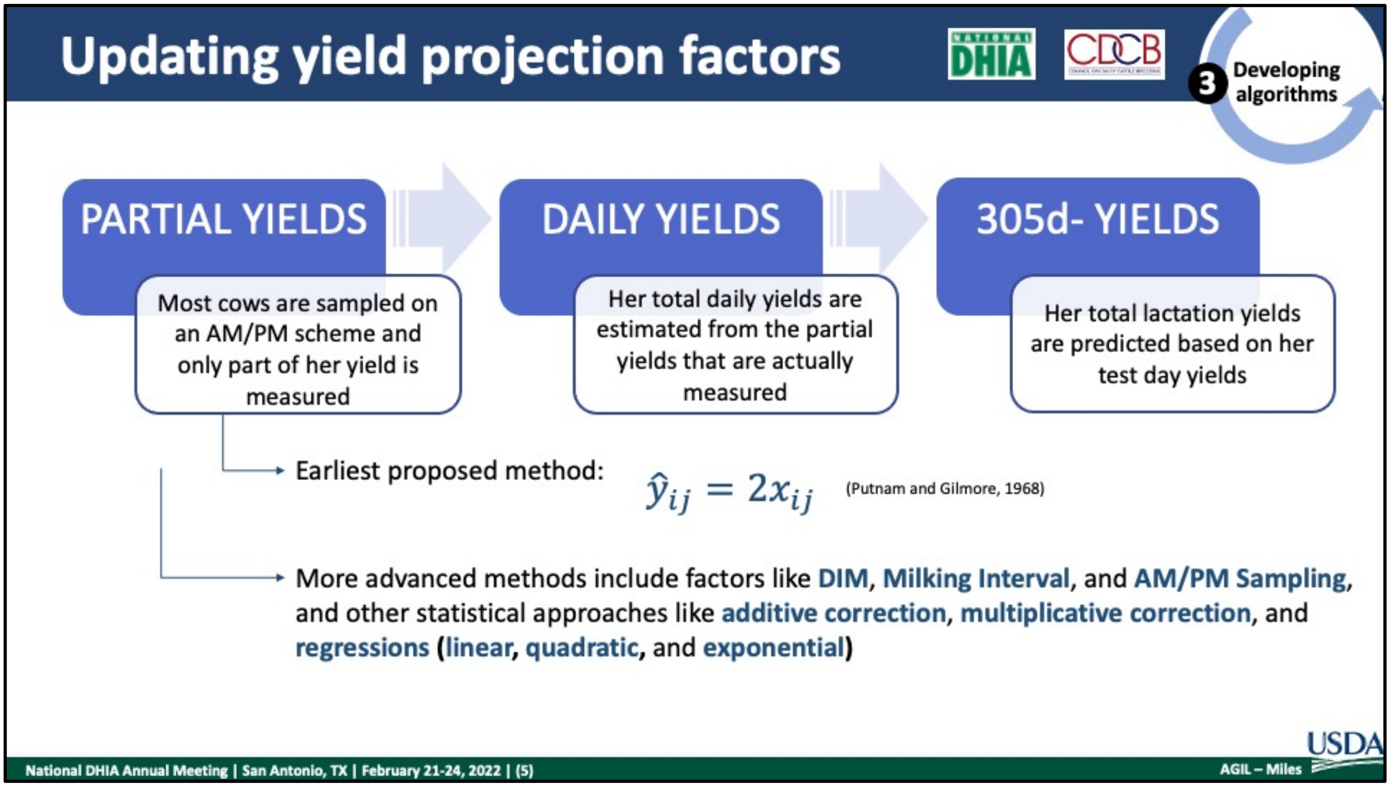
Enroll herds which capture US Dairy Systems Demographics

- BREED
  - Holstein
  - Jersey
- NUMBER OF LACTATIONS
  - First
  - Multiple
- HOUSING
  - Conventional
  - Grazing
- MILKING FREQUENCY
  - 2X, 3X, 4X
  - Robotic

National DHIA Annual Meeting | San Antonio, TX | February 21-24, 2022 | (4)

AGIL – Miles

We are delighted to be collaborating with National DHIA and CDCB on a joint venture to update our predictions of milk, fat, and protein yields for the modern dairy cow. Current factors were developed 30 years ago and following decades of genetic selection and evolving management, especially during the transition period, modern cows are quite different from their ancestors. The goal of this project is to evaluate projection factors to update our yield trait predictions. This will involve enrolling dairy herds across the US, considering factors like most common breeds, number of lactations, and management styles to capture US dairy systems demographics. This graphic details our initial approach and later phases will expand to include other breeds. A subset of cows will be chosen from each herd and sampled weekly for milk & component yields & milking interval. Intensive sampling has been designed for the first 120 DIM when key economic decisions are being made and we want as much resolution as is realistic to understand yield trends. After 120 DIM, this subset will be sampled according to their regular DHIA schedule along with the rest of the herd. This approach will give us a curated, high-quality dataset which has not existed before and will allow us to both characterize the modern cow and explore different methodologies to identify the best approach to accurately predicting yields.



Updating yield projection factors

PARTIAL YIELDS

Most cows are sampled on an AM/PM scheme and only part of her yield is measured

DAILY YIELDS

Her total daily yields are estimated from the partial yields that are actually measured

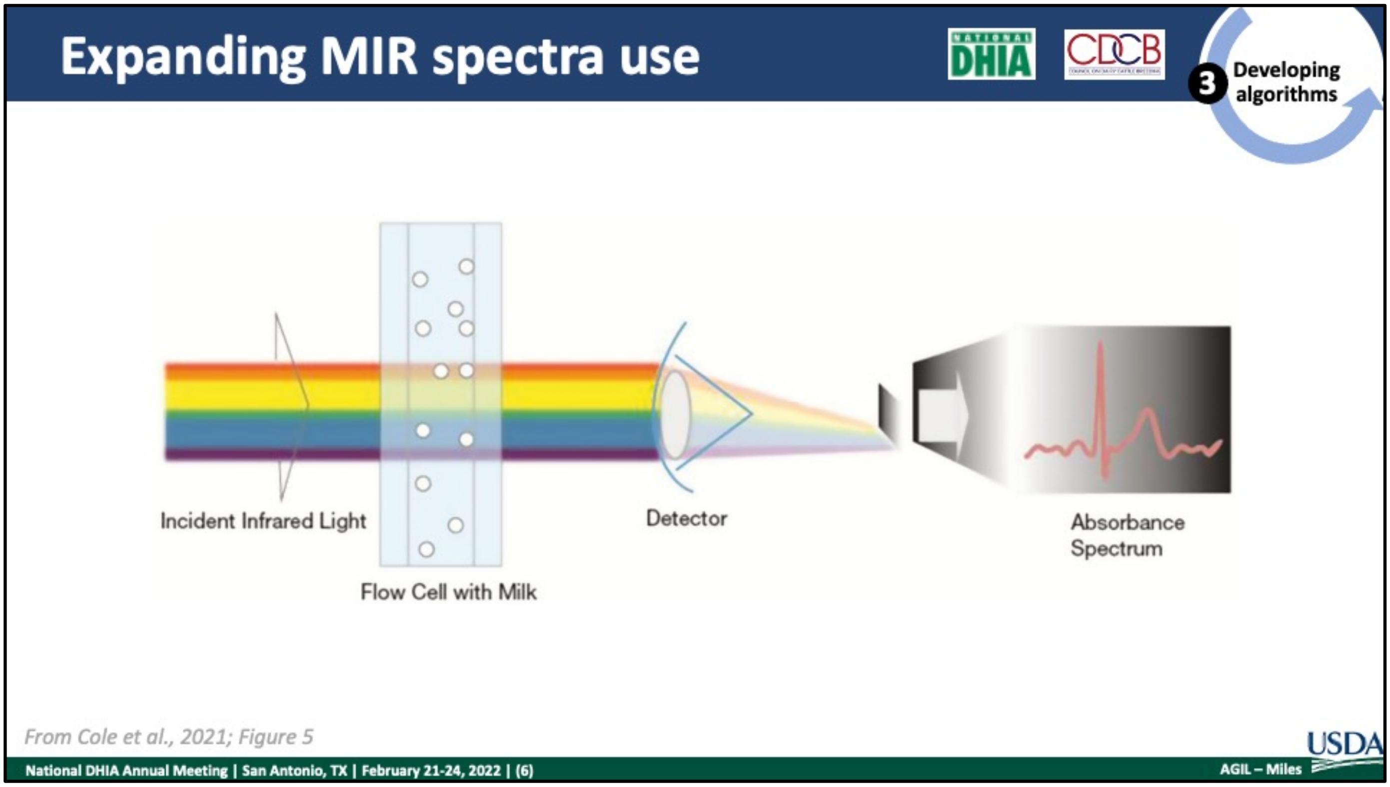
305d- YIELDS

Her total lactation yields are predicted based on her test day yields

Earliest proposed method: $\hat{y}_{ij} = 2x_{ij}$ (Putnam and Gilmore, 1968)

More advanced methods include factors like **DIM**, **Milking Interval**, and **AM/PM Sampling**, and other statistical approaches like **additive correction**, **multiplicative correction**, and **regressions** (**linear**, **quadratic**, and **exponential**)

Our research priorities are a response to the producer's need to accurately assess their cow. These needs are related to culling and breeding decisions, but also reflect more dynamic, daily concerns like estimating the amount of milk that will be shipped or monitoring changes in a cow's production level. As it relates to this project, There are 3 levels to accurately assessing cow performance **graphic**. At the current stage, research is focused on accurately estimating daily yields from partial yields. The earliest proposed method for estimating total daily yield from partial yields was by simply multiplying by 2 (for a 2X herd). This makes intuitive sense, but it was quickly discovered that estimating yields is much more complicated than that, with milk secretion being affected by factors like daylight and milking interval. Since then, many more advanced methods of estimating daily yields have been proposed considering confounding factors and different statistical methodologies. Dr. Nick Wu @ CDCB has been working on a comprehensive review of the methodology where he compares their performance using simulation data (in prep) to ensure we move forward with the most robust and biologically appropriate models.

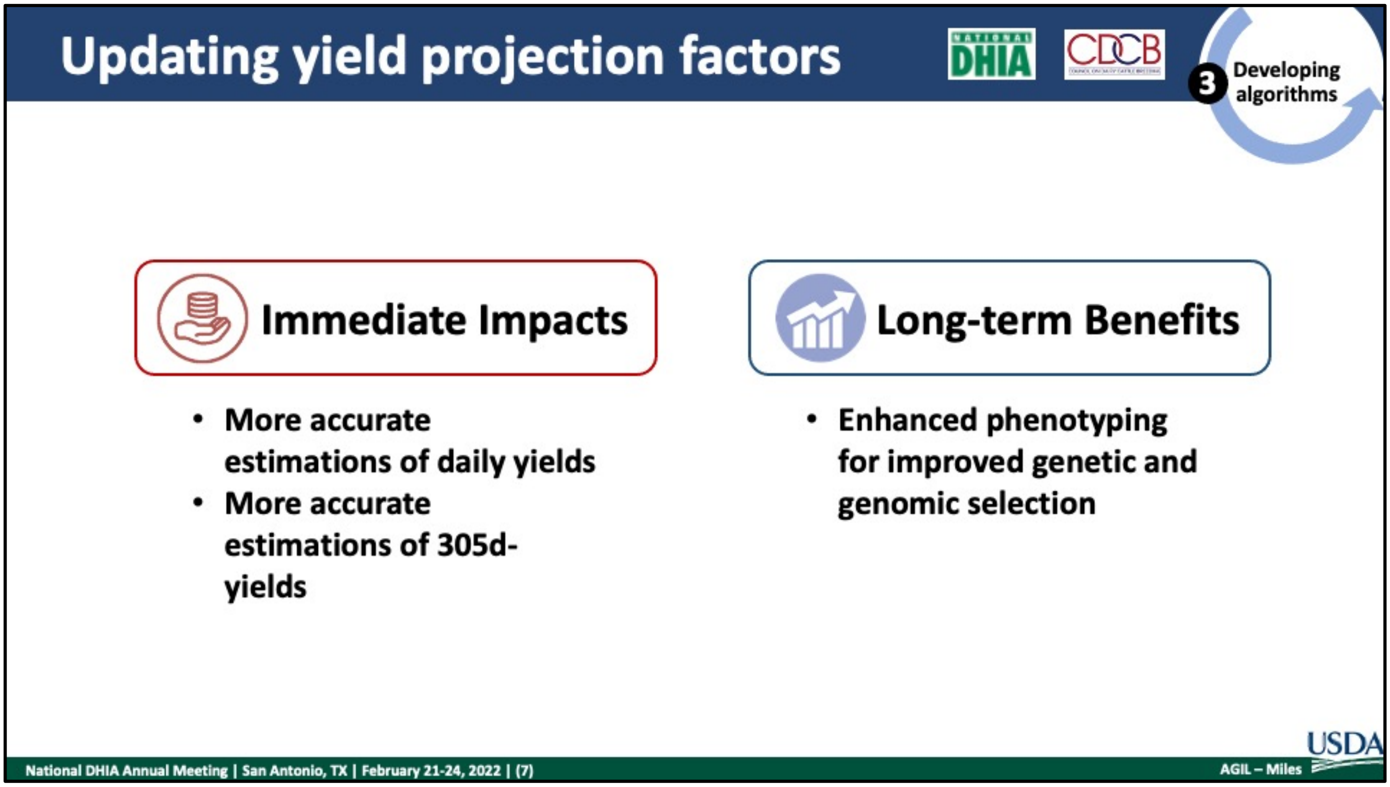


Expanding MIR spectra use

3 Developing algorithms

From Cole et al., 2021; Figure 5

Another element of this project is related to expanding the use of mid-infrared spectra data and exploring their possibility as a convenient and inexpensive proxy to harder to collect phenotypes, like metabolic profiles and methane emissions. These data can be used in a similar manner to genomic prediction, where phenotypes collected from a reference population are regressed against the absorbance spectrum, and those resulting values used to estimate values in the larger population.



Updating yield projection factors

3 Developing algorithms

Immediate Impacts

- More accurate estimations of daily yields
- More accurate estimations of 305d-yields

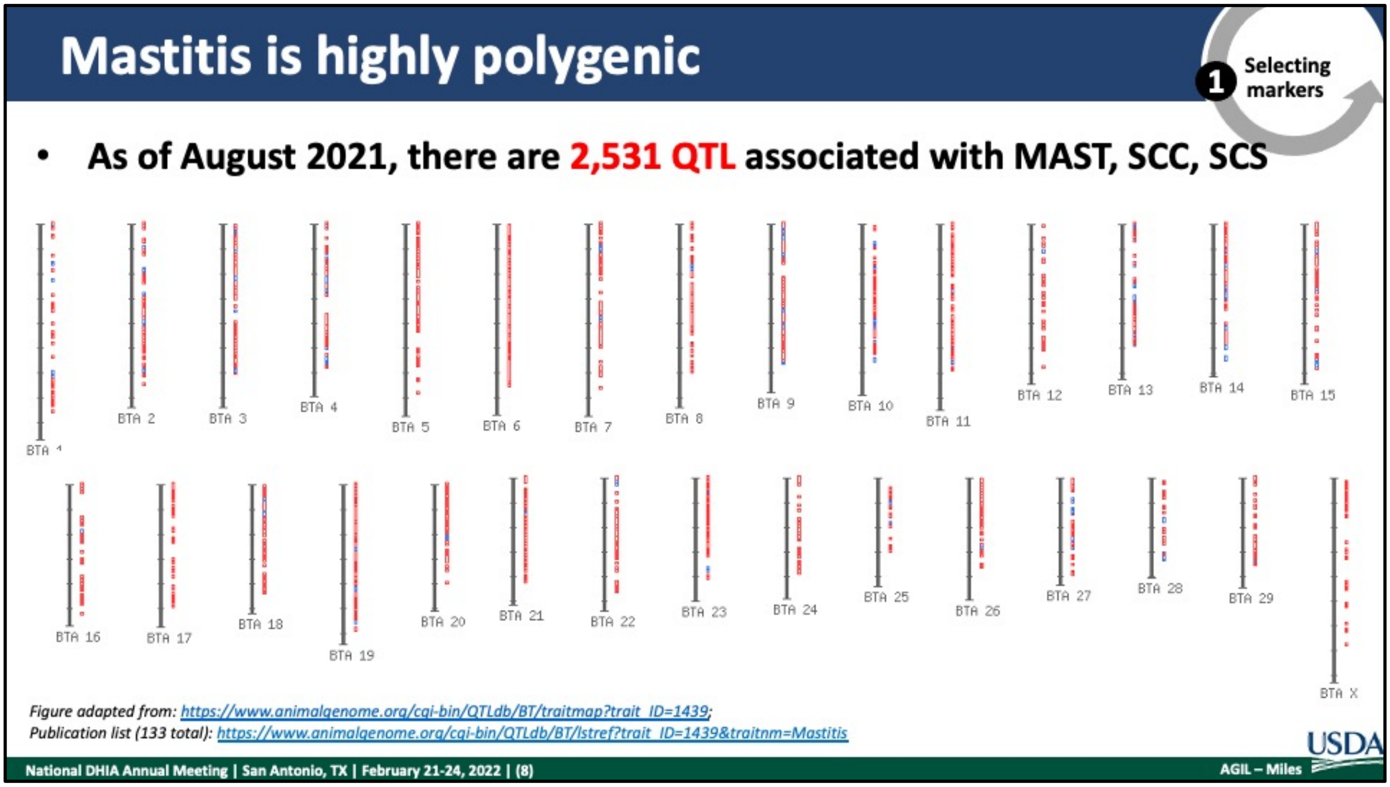
Long-term Benefits

- Enhanced phenotyping for improved genetic and genomic selection

National DHIA Annual Meeting | San Antonio, TX | February 21-24, 2022 | (7)

AGIL – Miles

The benefits of this project cannot be overstated. Key economic decisions are made during the first 120 DIM and improving the early lactation yield projections will benefit all dairy herds. The long-term impacts of this work include enhanced genetic selection, but more significantly, an immediate return on investment by providing herd managers with better performance predictions for their cows. The primary impact of this work will be in facilitating informed management affecting the daily operations of the farm.



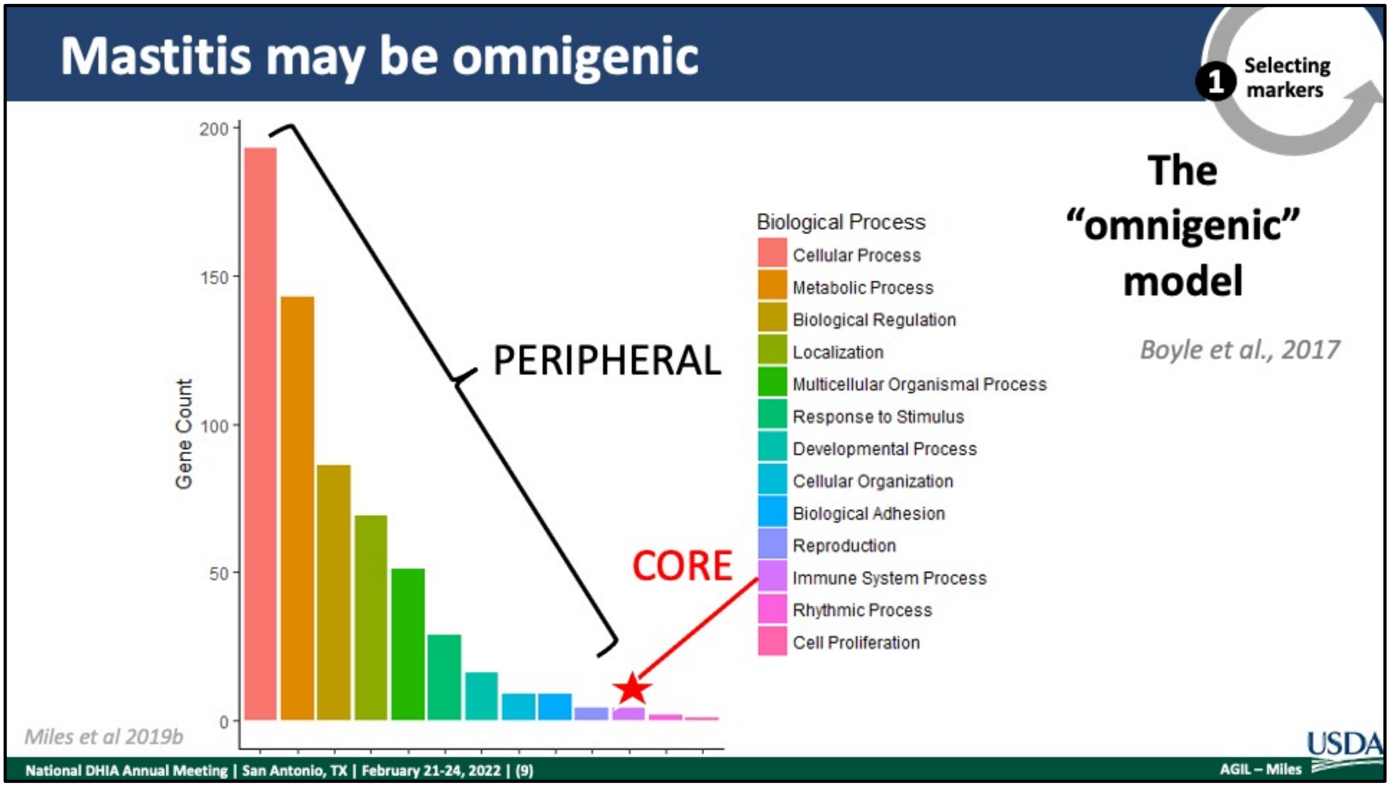
Mastitis is highly polygenic

1 Selecting markers

- As of August 2021, there are **2,531 QTL** associated with MAST, SCC, SCS

Figure adapted from: https://www.animalgenome.org/cgi-bin/QTLdb/BT/traitmap?trait_ID=1439;
Publication list (133 total): https://www.animalgenome.org/cgi-bin/QTLdb/BT/lstref?trait_ID=1439&traitnm=Mastitis

Related to our objective of selecting new markers, I want to take you through an example using mastitis. These black lines represent all 29 autosomes and the X chromosome, and all of these red and blue boxes next to them are showing QTL that are significantly associated with mastitis. You can see they span the entire genome, and this drives home the highly complex and polygenic nature of mastitis.



Mastitis may be omnigenic

❶ Selecting markers

The “omnigenic” model

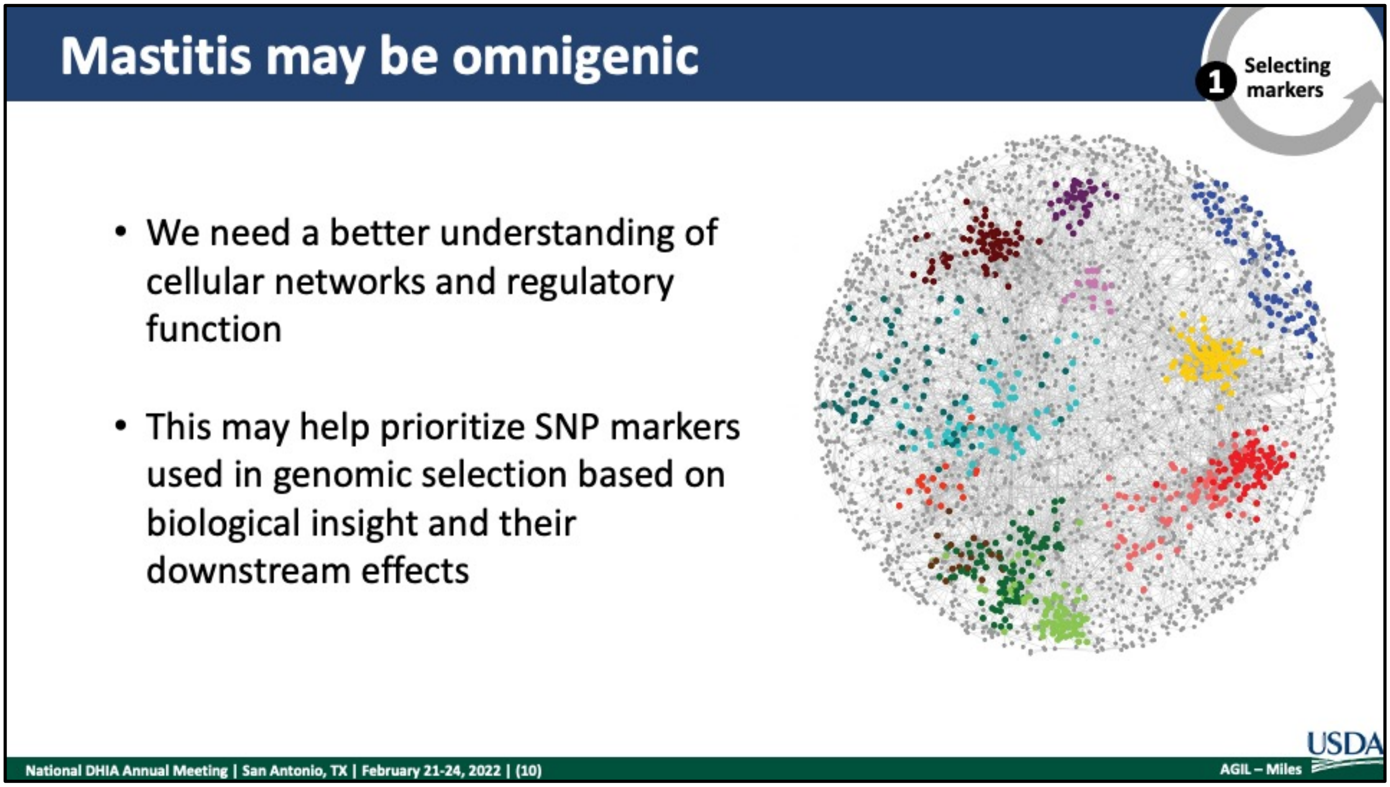
Boyle et al., 2017

Miles et al 2019b

National DHIA Annual Meeting | San Antonio, TX | February 21-24, 2022 | (9)

AGIL – Miles

But mastitis may also be omnigenic. To illustrate this point, consider smaller scale example. This graph is showing gene ontology analysis from some of my past work on mastitis, where 990 genes implicated in various mastitis phenotypes are categorized by the biological process they are thought to be involved in: only a fraction were protein-coding genes related to the immune response. This surprising observation could be explained by the “omnigenic model” which proposes that the genetic architecture of complex traits (like MAST) is produced by a massive regulatory network of genes, each with very small effect. 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.

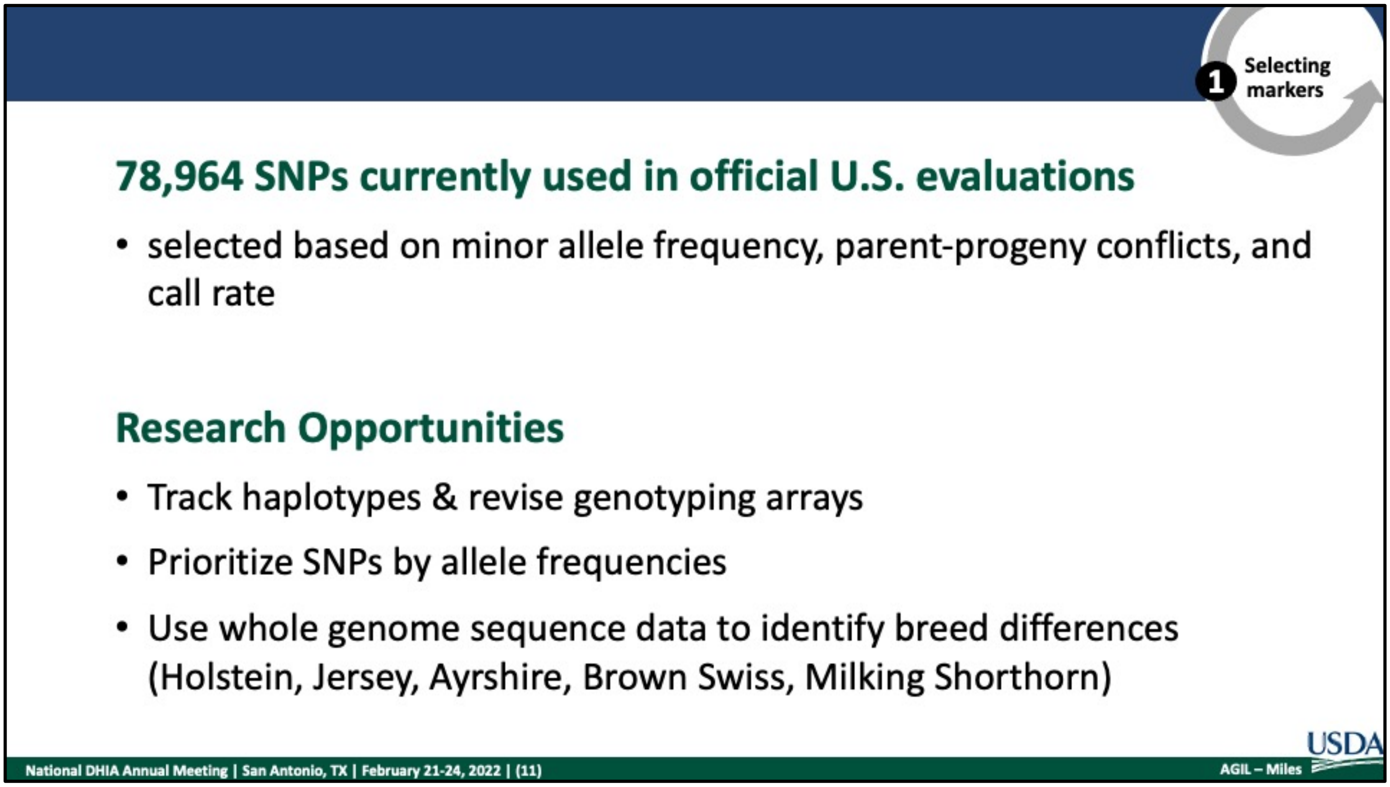


Mastitis may be omnigenic

1 Selecting markers

- We need a better understanding of cellular networks and regulatory function
- This may help prioritize SNP markers used in genomic selection based on biological insight and their downstream effects

The implications are that to get a more complete picture of the genetics of complex traits we need a clearer understanding of cellular networks and regulatory function. This may feed back into genetic progress attributable to selection strategies by prioritizing markers used in genomic selection based on biological insight and the downstream effects of marker variants.



1 Selecting markers

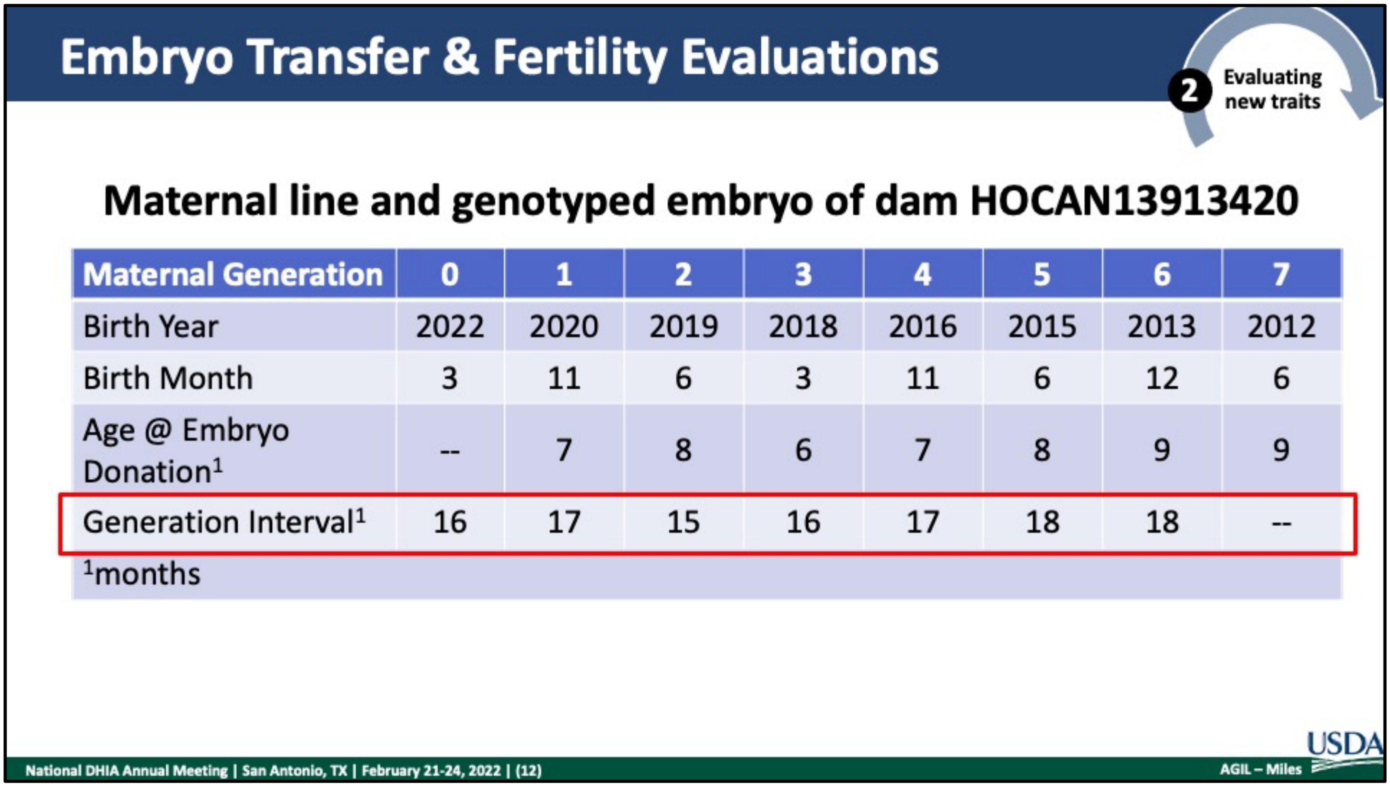
78,964 SNPs currently used in official U.S. evaluations

- selected based on minor allele frequency, parent-progeny conflicts, and call rate

Research Opportunities

- Track haplotypes & revise genotyping arrays
- Prioritize SNPs by allele frequencies
- Use whole genome sequence data to identify breed differences (Holstein, Jersey, Ayrshire, Brown Swiss, Milking Shorthorn)

The nearly 79k SNP markers currently used in official U.S. evaluations were selected based on minor allele frequency, parent-progeny conflicts, and call rate, and One of our opportunities moving forward will be fortifying the delivery of practical and affordable selection strategies with improved biological understanding through this type of functional work. This is just one of many approaches to selecting markers, and AGIL will also be tracking haplotypes to more precisely locate QTL associated with important dairy cattle trait which may result in revised genotyping arrays, and will explore using allele frequencies in the population to prioritize SNPs, an area Dr. Toghiani has expertise in.



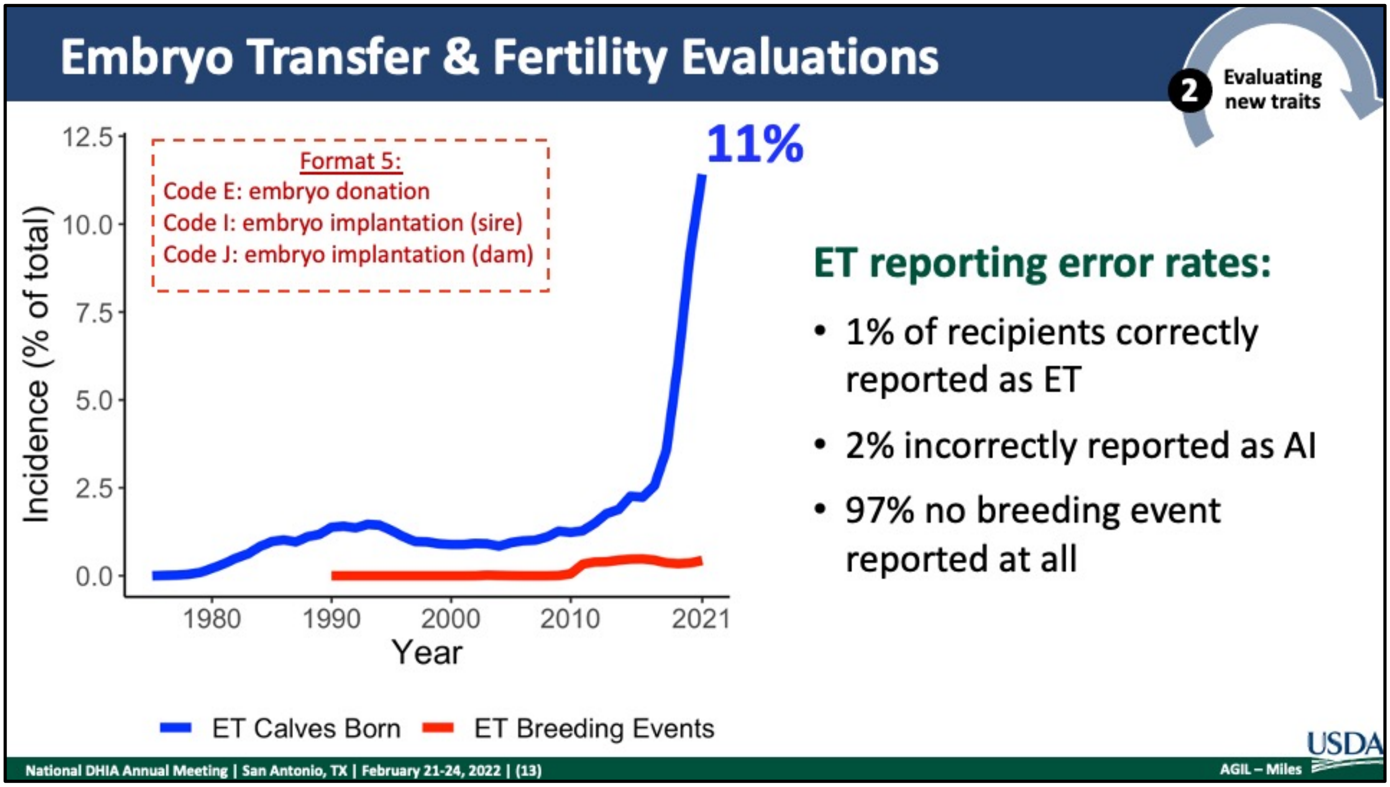
Embryo Transfer & Fertility Evaluations

Maternal line and genotyped embryo of dam HOCAN13913420

Maternal Generation	0	1	2	3	4	5	6	7
Birth Year	2022	2020	2019	2018	2016	2015	2013	2012
Birth Month	3	11	6	3	11	6	12	6
Age @ Embryo Donation[1]	--	7	8	6	7	8	9	9
Generation Interval[1]	16	17	15	16	17	18	18	--

[1] months

A major goal of this plan is to identify new traits or important modifications to existing ones by analyzing changing industry trends. A great example of that is ET Part of the attraction of ET is in reduced generation intervals and if we consider a top genetic merit heifer whose embryos were collected starting at 7 months, we see the maternal line continued this rapid turnover resulting in a pretty remarkable 7 generations in 10 years *(average generation interval 16.7 months)*.. ET is becoming more affordable and is likely here to stay.



Embryo Transfer & Fertility Evaluations

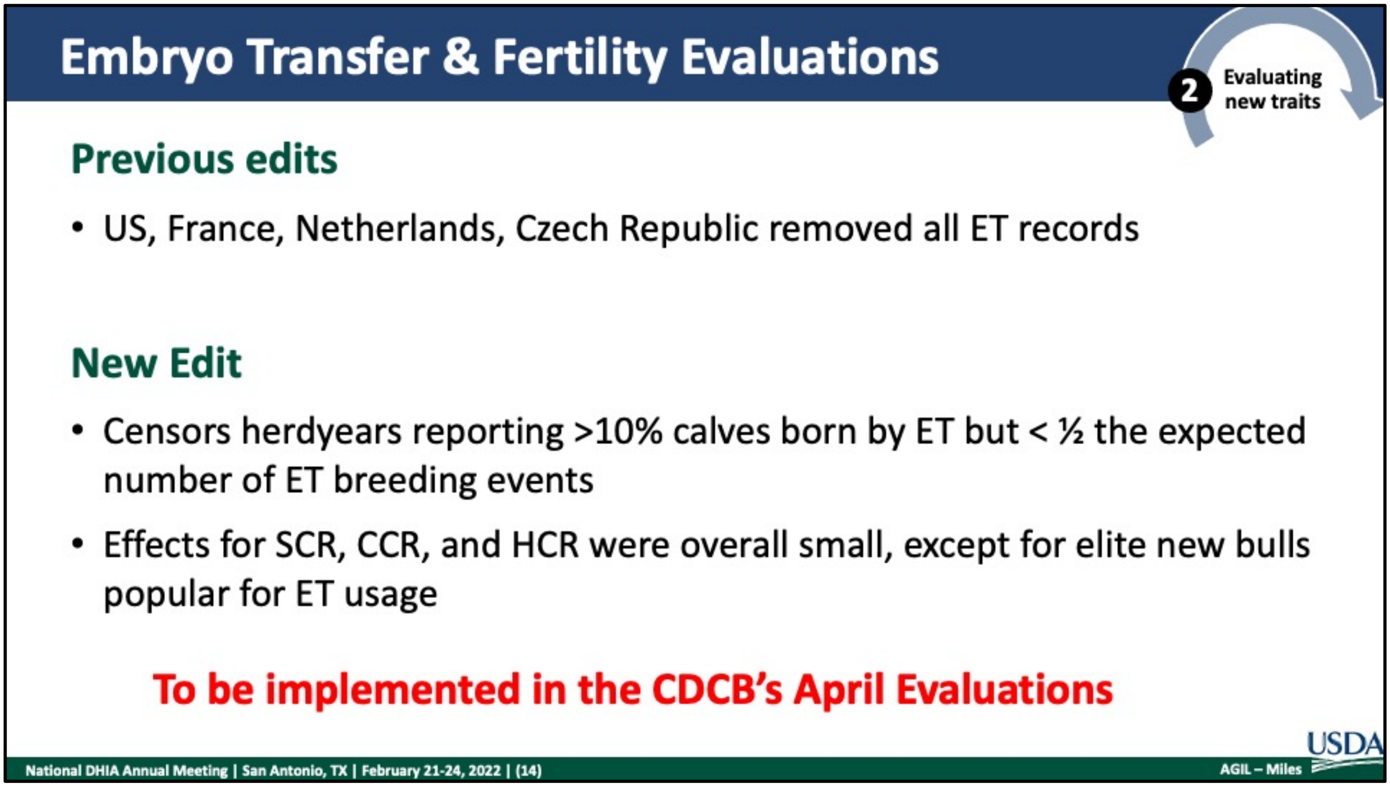
Format 5:
Code E: embryo donation
Code I: embryo implantation (sire)
Code J: embryo implantation (dam)

ET reporting error rates:

- 1% of recipients correctly reported as ET
- 2% incorrectly reported as AI
- 97% no breeding event reported at all

USDA

Unfortunately, ET breeding event reporting is not matching its popularity. 11% of calves born in the US last year were attributable to ET, but only a fraction of the expected breeding events were reported. Reported mating types were matched with recorded birth types to assess error rates and for the most part ET calves have no associated breeding events at all.



Embryo Transfer & Fertility Evaluations

Previous edits

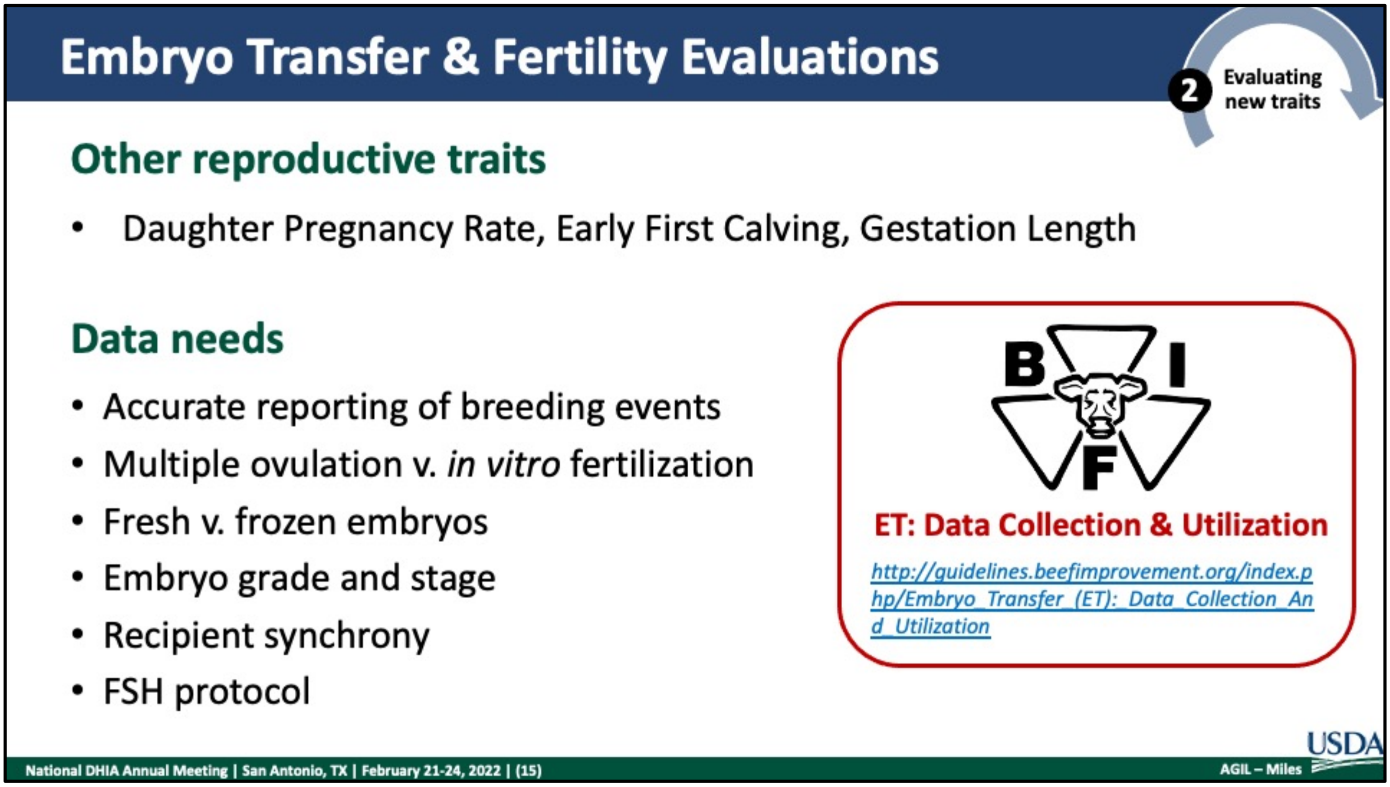
- US, France, Netherlands, Czech Republic removed all ET records

New Edit

- Censors herdyears reporting >10% calves born by ET but < ½ the expected number of ET breeding events
- Effects for SCR, CCR, and HCR were overall small, except for elite new bulls popular for ET usage

To be implemented in the CDCB's April Evaluations

Previously the US and 3 other Interbull-participating countries accounted for ET by censoring any ET-related records, but this isn't necessarily the best approach as these are representing our most elite animals and becoming a larger and larger portion of the dairy cattle population. We have proposed a new edit which maximizes data retention while removing records most likely to introduce confounding bias. Currently these edits have been applied for SCR, CCR, and HCR with overall small effects, though the changes were largest for elite new bulls popular for ET use. These edits will debut in CDCBs April evaluations, but there is a lot more that can be done here.



Embryo Transfer & Fertility Evaluations

2 Evaluating new traits

Other reproductive traits

- Daughter Pregnancy Rate, Early First Calving, Gestation Length

Data needs

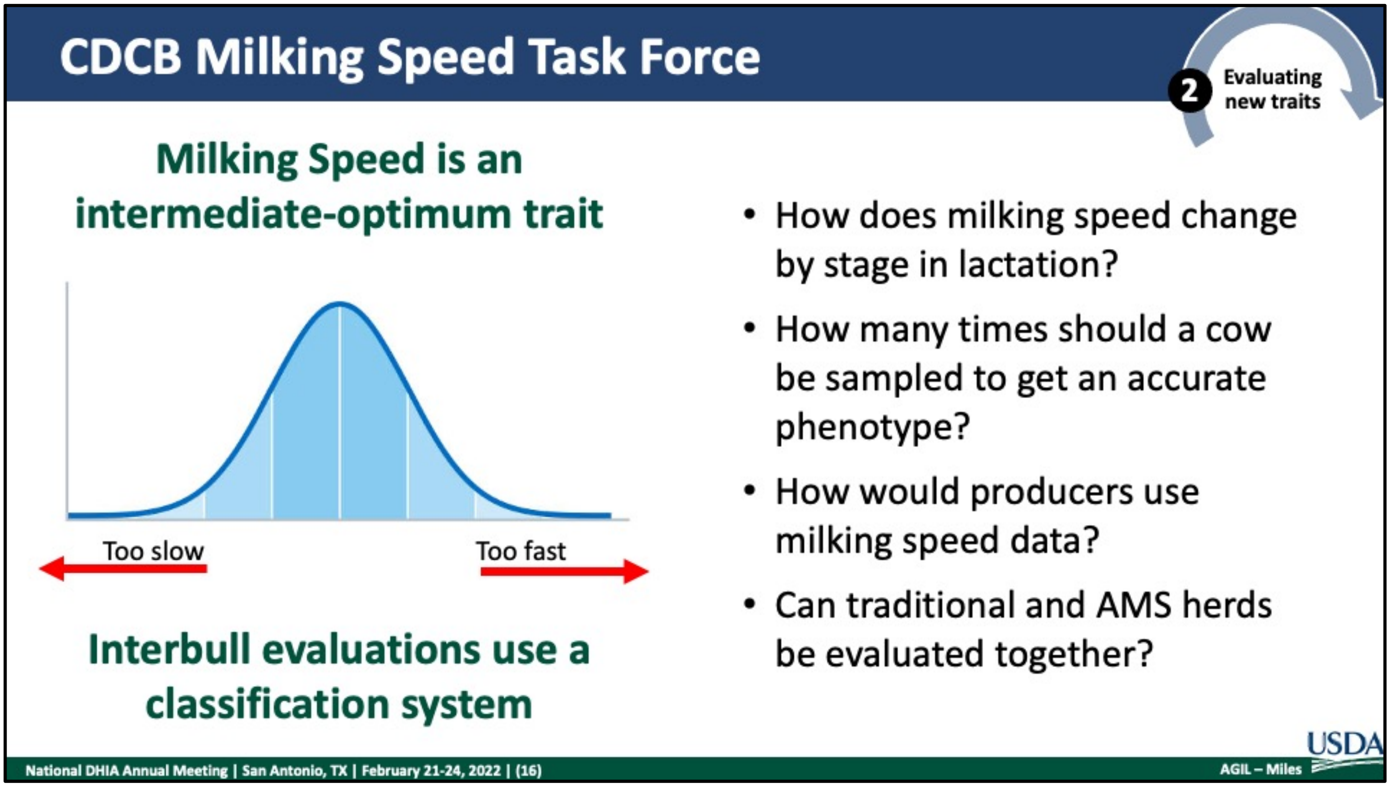
- Accurate reporting of breeding events
- Multiple ovulation v. *in vitro* fertilization
- Fresh v. frozen embryos
- Embryo grade and stage
- Recipient synchrony
- FSH protocol

B I F

ET: Data Collection & Utilization

http://guidelines.beefimprovement.org/index.php/Embryo_Transfer_(ET):_Data_Collection_And_Utilization

This edit may also be needed for other fertility-related traits, but a more urgent need is better data on ET use: Firstly, better reporting of breeding events which should be a more straight-forward fix, but even more useful will be acquiring high-quality ET records (comprising details like multiple ovulation ET versus in vitro fertilization, fresh versus frozen, embryo grade and stage, recipient synchrony, follicle stimulating hormone protocol, numbers of degenerate embryos, etc.). This can help partition genetic effects among fertility, conception rate, and pregnancy rate. The Beef Improvement Federation recently approved guidelines for using ET records and necessary considerations in evaluation models, such as accounting for the effects of both the donor (e.g., calf genetics) and recipient dams (e.g., birth weight). These principles could also be adopted by the dairy industry with the establishment of new data pipelines. Currently, unstandardized ET reporting is having minimal effect, but that is not a guarantee if the ET calf birth rate continues to rise without a parallel increase in accurate ET breeding event reporting.



CDCB Milking Speed Task Force

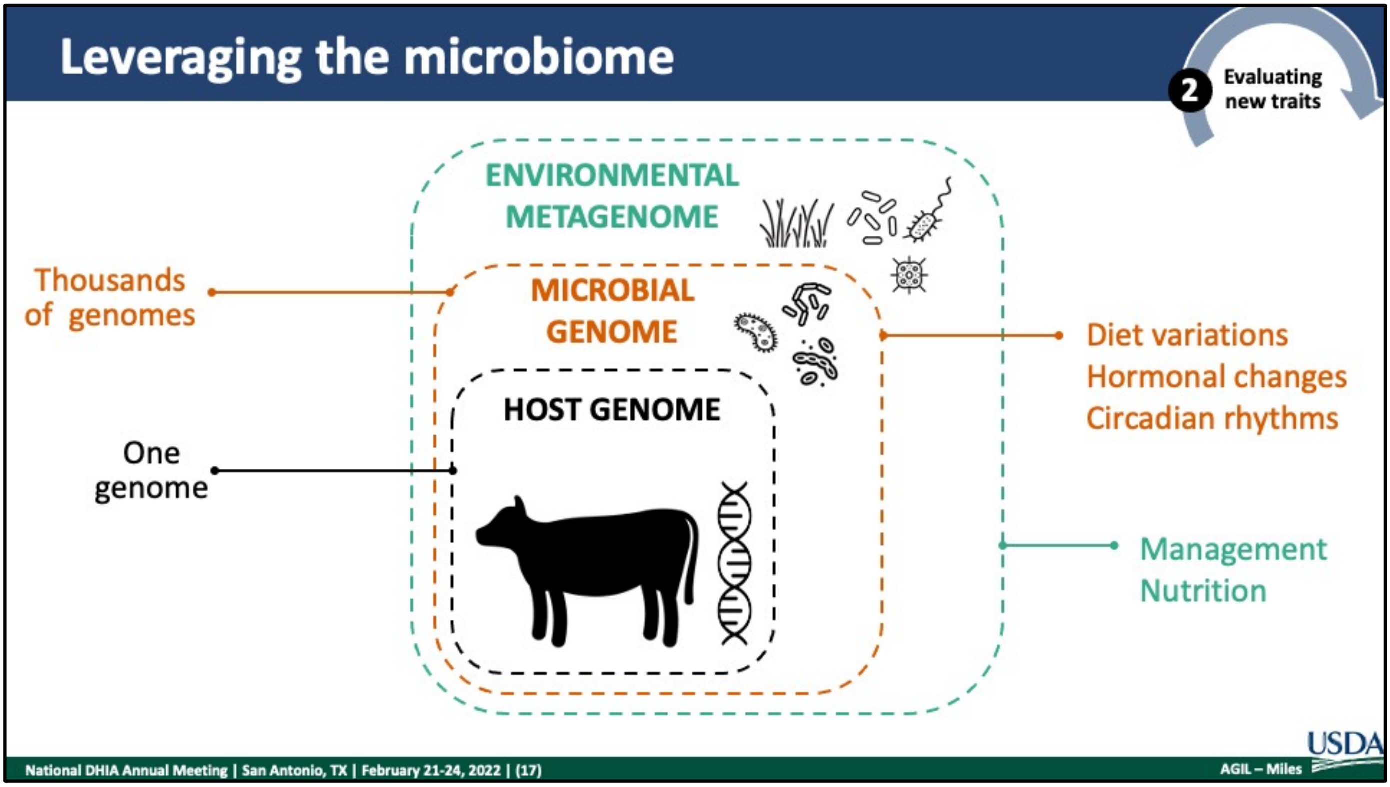
2 Evaluating new traits

Milking Speed is an intermediate-optimum trait

Interbull evaluations use a classification system

- How does milking speed change by stage in lactation?
- How many times should a cow be sampled to get an accurate phenotype?
- How would producers use milking speed data?
- Can traditional and AMS herds be evaluated together?

Our mandate is to review existing data types and data availability, develop a clear trait definition, suggest quality standards for milking speed data, and identify requirements of a data pipeline into the National Cooperator Database. We know MS is intermediate-optimum trait, & most countries which include MS in their evaluations use a classification system where a trained appraiser travels to the farm and works with the herd manager to assign subjective scores on a scale of slow to fast. This is unlikely to be practical here with larger average herd size, so we are exploring the use of quantitative data coming off milk meters. We are using some preliminary data to answer key questions about MS like how milking speed changes by stage in lactation. MS is especially important during peak production when she would spend the longest in the parlor. We also have observed that MS is slower and more variable during the first 10 DIM, which makes sense with the clearance of colostrum and inflammation associated with parturition. This also raises the question of how often a cow should be sampled to accurately construct her MS phenotype. And we want to consider how this information will be useful to producers, and that they may use it differently. For example, we know conventional herds use MS to choose their milking groups, but herds with AMS may also use this information to decide how many robots they need to buy to milk the number of cows they have. There's an opportunity here to create a new trait based on hard numbers that would have both management value and be suitable for genetic selection.



Leveraging the microbiome

2 Evaluating new traits

ENVIRONMENTAL METAGENOME

MICROBIAL GENOME

HOST GENOME

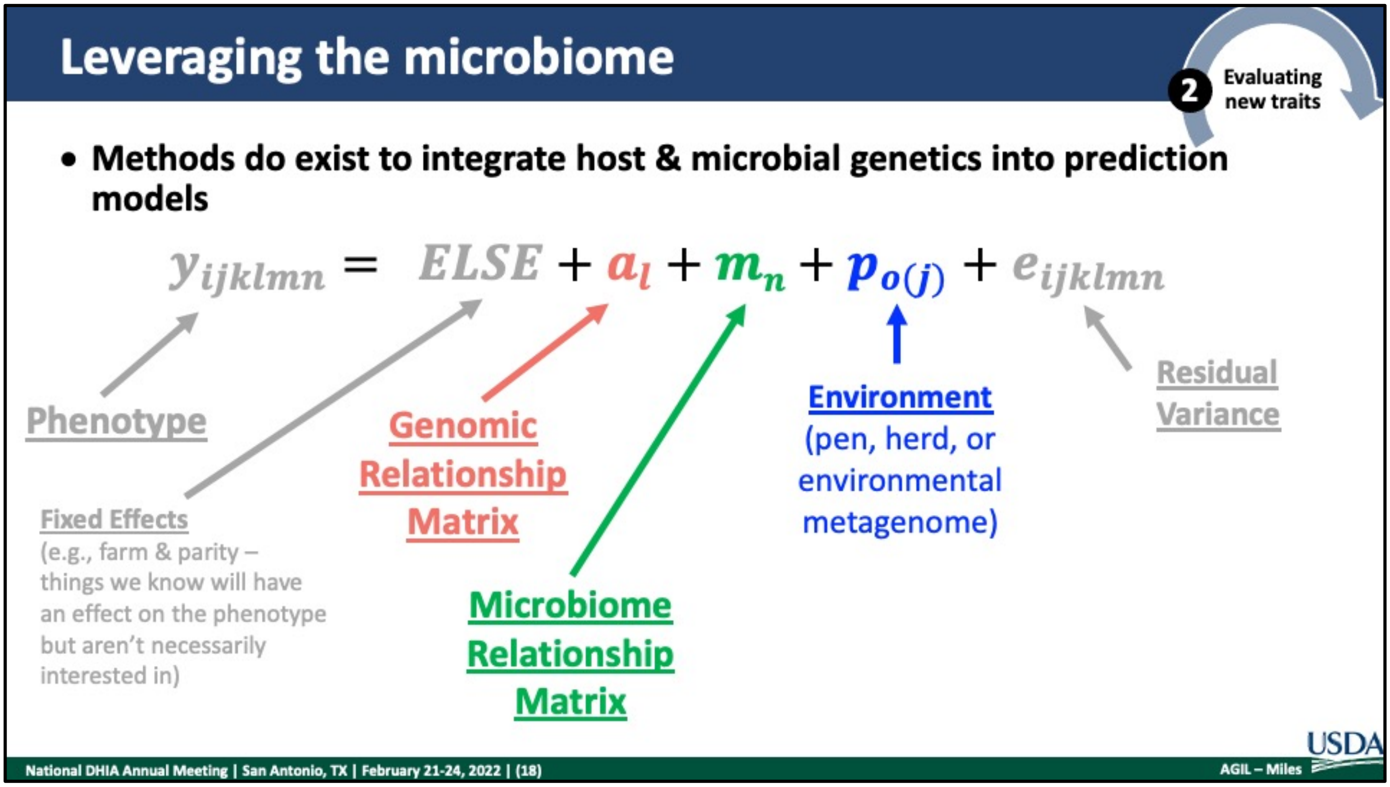
Thousands of genomes

One genome

Diet variations
Hormonal changes
Circadian rhythms

Management
Nutrition

When we talk genetic improvement traditionally, we are trying to understand how the cow genome is influencing her observable traits. But the host genome is not the only thing in play and there has been a rapid uptick of animal microbiome research in recent years and so this paradigm is starting to change. Microbes are in constant flux with environment, so when we talk about genetic influences on phenotypes, these include the host genome, the microbial genome, and the env metagenome. But we have our work cut out for us here because the microbiome is a tricky beast; it can be considered both a phenotype and a source of genetic material. Cows have one genome; a microbiome has thousands of genomes and none of them are stable. They can vary by age, diet, and physiological state – not to mention the fact that the environmental metagenome can also be influenced by management factors. Accounting for the role of the microbiome in cow health and production requires both sophisticated computational tools (to handle the magnitude of data) and a deep knowledge of biology (to tease true insights out of the noise).



Leveraging the microbiome

- **Methods do exist to integrate host & microbial genetics into prediction models**

$$y_{ijklmn} = ELSE + a_l + m_n + p_{o(j)} + e_{ijklmn}$$

Phenotype

Fixed Effects
(e.g., farm & parity – things we know will have an effect on the phenotype but aren't necessarily interested in)

Genomic Relationship Matrix

Microbiome Relationship Matrix

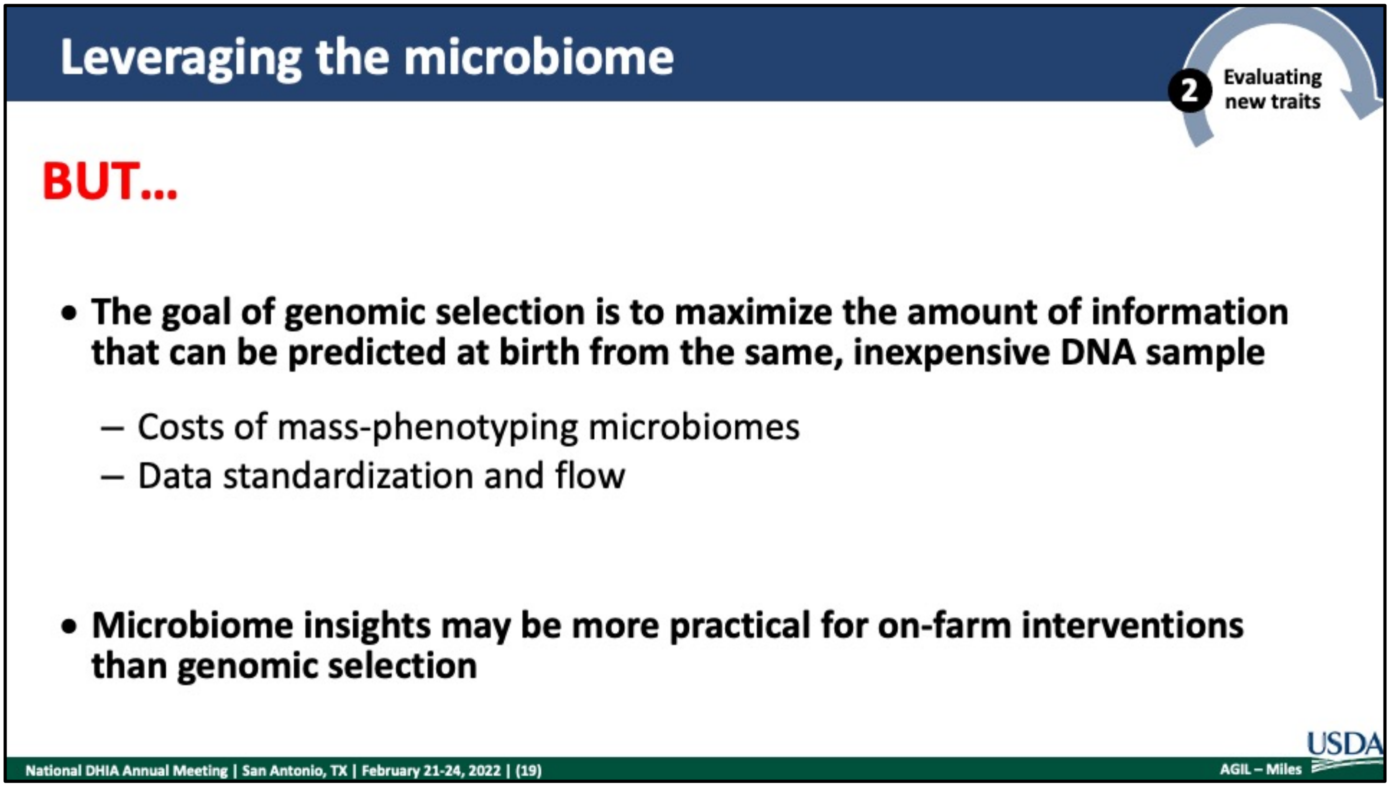
Environment
(pen, herd, or environmental metagenome)

Residual Variance

National DHIA Annual Meeting | San Antonio, TX | February 21-24, 2022 | (18)

AGIL – Miles

Methods do exist to account for the microbiome in genomic prediction where in addition to host genomic information, we can add a term for the microbial metagenome and even the environmental metagenome. We are probably all familiar with “heritability” (the amount of phenotype variation that can be explained by genetic variation). But by examining the variance component of the microbiome we can calculate “microbiability” or the amount of variance in the phenotype attributable to variance in the microbiome.

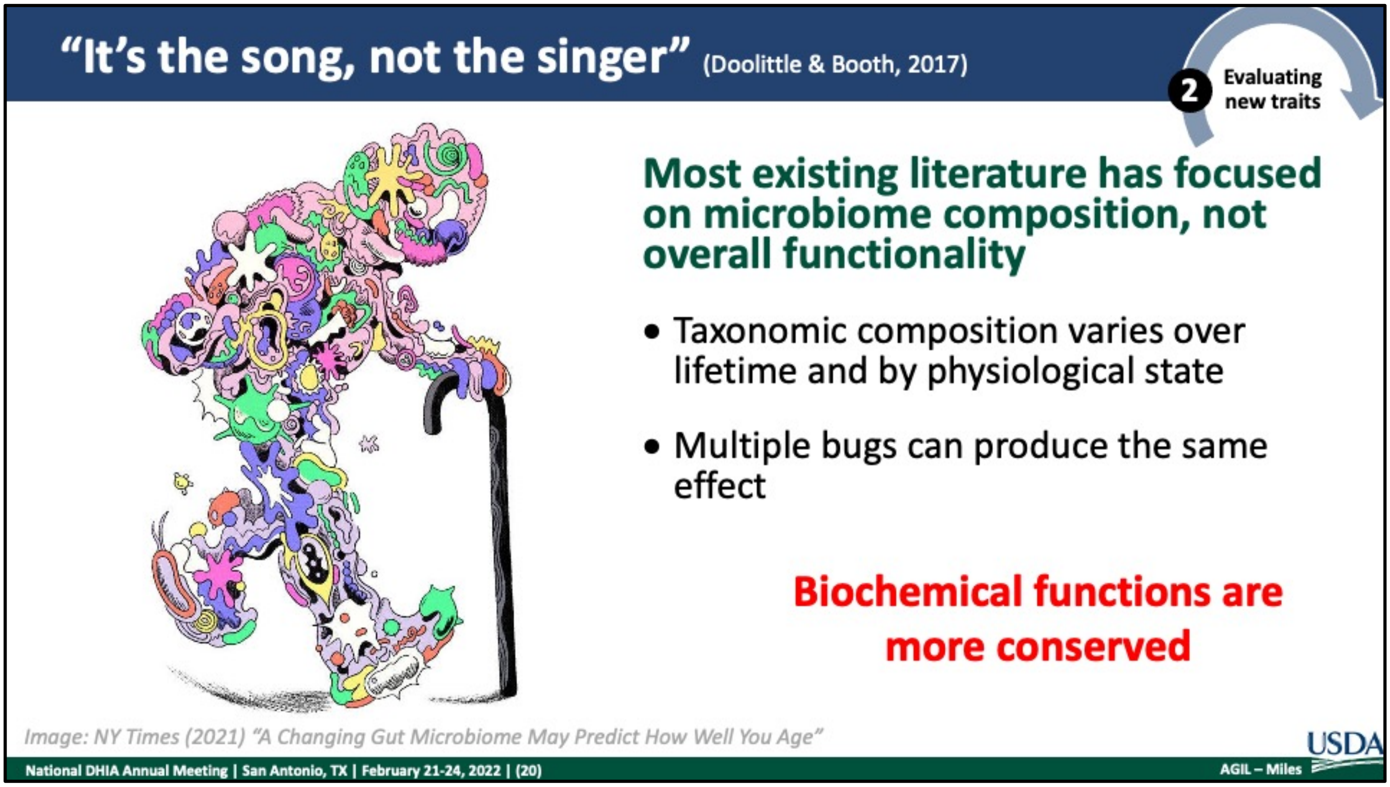


Leveraging the microbiome

BUT…

- **The goal of genomic selection is to maximize the amount of information that can be predicted at birth from the same, inexpensive DNA sample**
  - Costs of mass-phenotyping microbiomes
  - Data standardization and flow
- **Microbiome insights may be more practical for on-farm interventions than genomic selection**

The benefits of microbiome research become a little less clear when we think about applying them. The goal of genomic selection is to maximize the information that can be predicted at birth from the same inexpensive DNA sample; 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. Repeatability of results among microbiome samples and studies is already a big challenge, especially for low biomass samples like milk, and financial constraints aside, significant groundwork would be required to develop pipelines for high-quality, standardized microbiome data to flow into the National Cooperator Database.



"It's the song, not the singer" (Doolittle & Booth, 2017)

2 Evaluating new traits

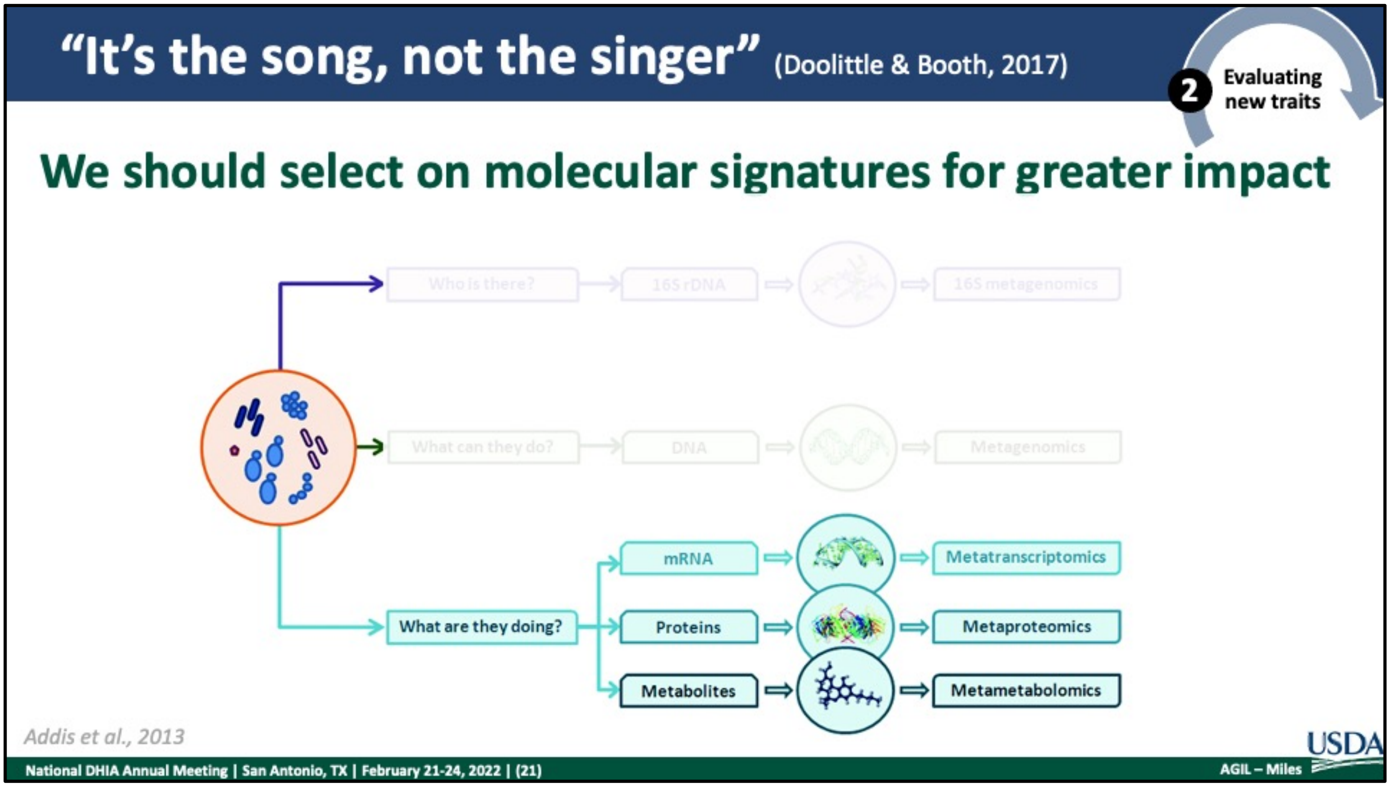
Most existing literature has focused on microbiome composition, not overall functionality

- Taxonomic composition varies over lifetime and by physiological state
- Multiple bugs can produce the same effect

Biochemical functions are more conserved

Image: NY Times (2021) "A Changing Gut Microbiome May Predict How Well You Age"

Most research has focused on microbiome composition, not overall functionality, and this is probably due to the relative affordability of 16s rRNA gene profiling compared to more intensive characterization methods. It has been well-established that taxonomic composition may vary over the host's lifetime and physiological states, and often multiples species of microbes use and produce the same metabolites (VFA production in the rumen is a good example of this) , and so while taxonomy can change, biochemical functions and their molecular signatures are more heavily conserved, and it's these collective signatures that are responsible for observed functional effects



"It's the song, not the singer" (Doolittle & Booth, 2017)

2 Evaluating new traits

We should select on molecular signatures for greater impact

Who is there? → 16S rDNA ⇒ ⇒ 16S metagenomics

What can they do? → DNA ⇒ ⇒ Metagenomics

What are they doing? → mRNA ⇒ ⇒ Metatranscriptomics

What are they doing? → Proteins ⇒ ⇒ Metaproteomics

What are they doing? → Metabolites ⇒ ⇒ Metametabolomics

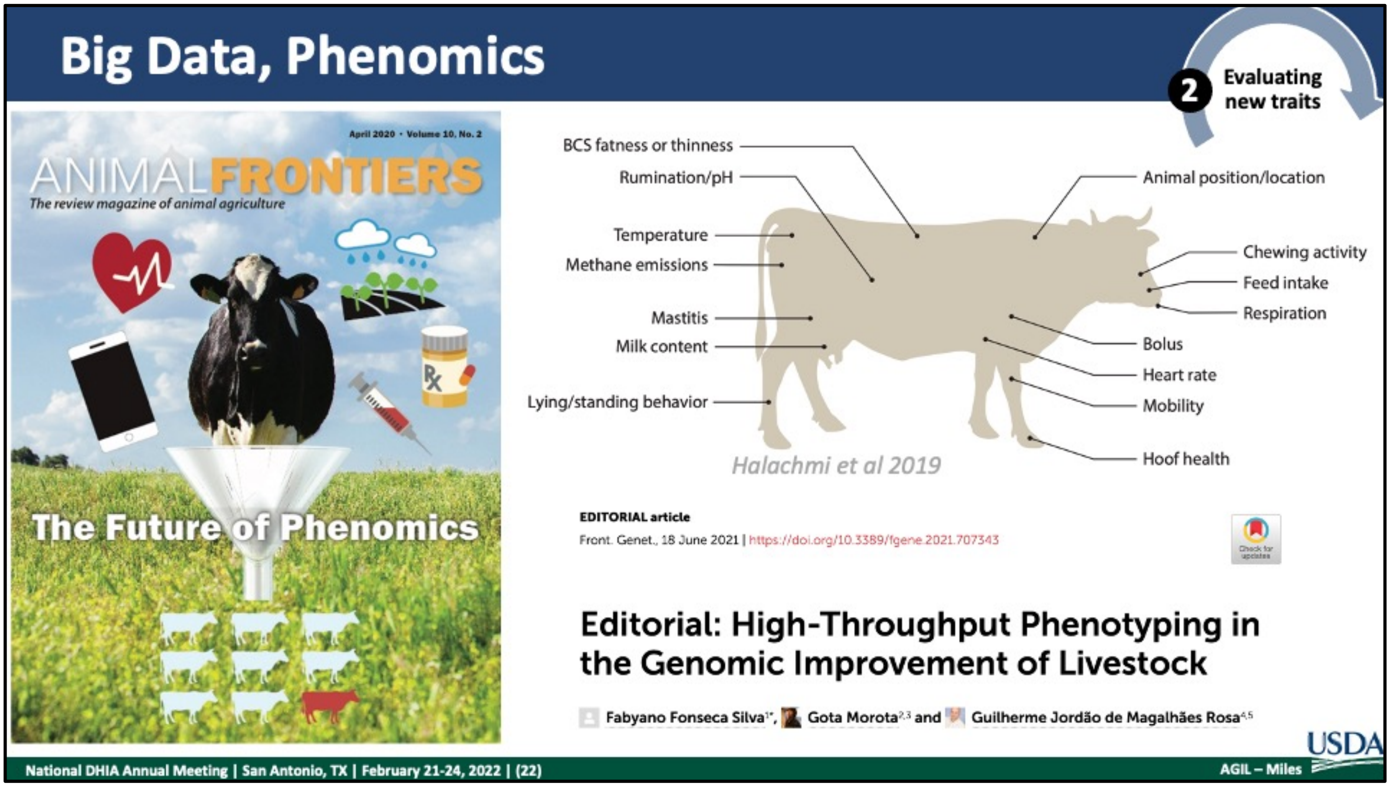
Addis et al., 2013

National DHIA Annual Meeting | San Antonio, TX | February 21-24, 2022 | (21)

USDA

AGIL – Miles

A more impactful approach would be to select on molecular signatures, not individual taxa. Funding this type of work would amplify opportunities for scientific innovation and real-world applicability. As sequencing costs drop and more robust analysis methods are developed, microbiome studies become more accessible to fields outside genomics for the first time, our ability to understand host-microbe interactions is really going to transform these disciplines. I find this moment very exciting and reminiscent of the genomics boom of the early 2000s.



Big Data, Phenomics

Halachmi et al 2019

EDITORIAL article

Front. Genet., 18 June 2021 | https://doi.org/10.3389/fgene.2021.707343

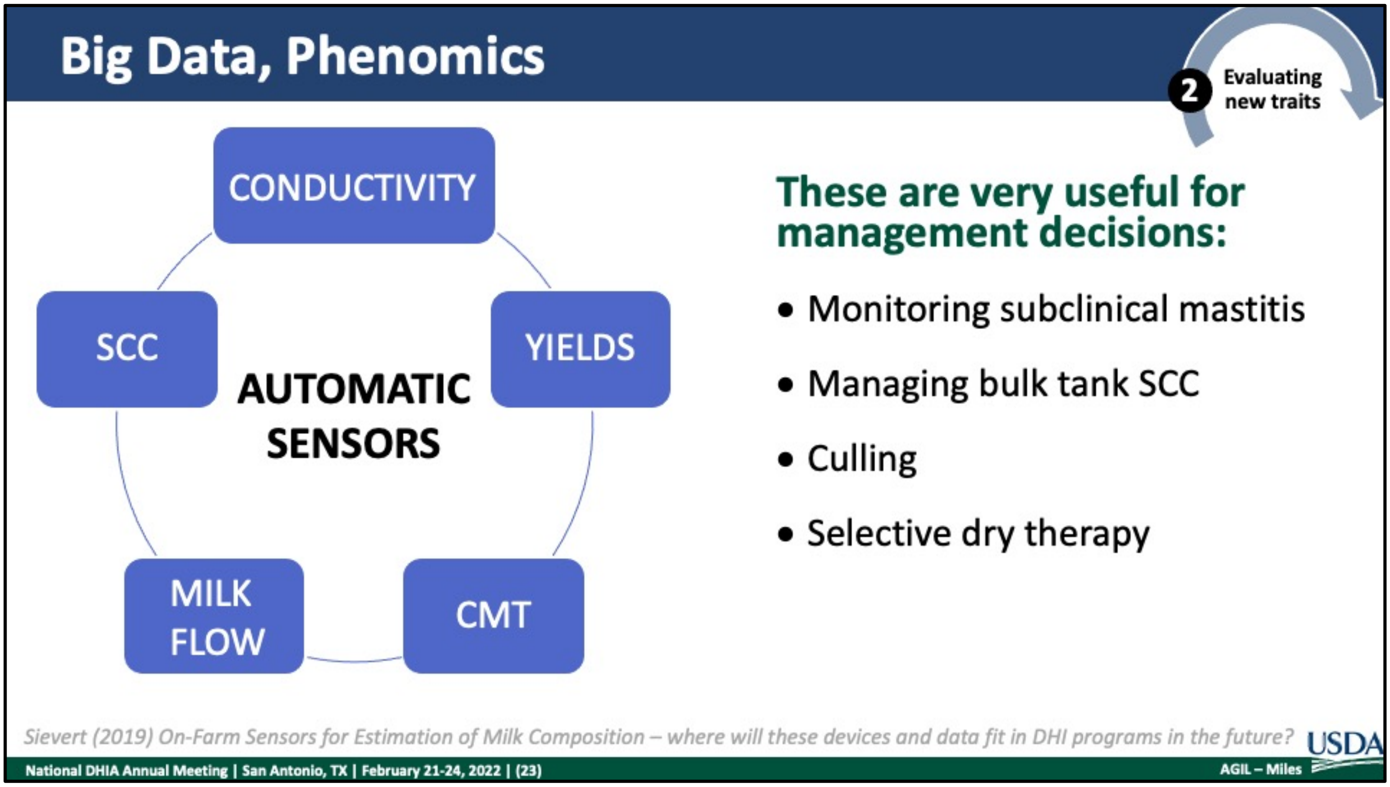
Editorial: High-Throughput Phenotyping in the Genomic Improvement of Livestock

Fabyano Fonseca Silva[1*], Gota Morota[2,3] and Guilherme Jordão de Magalhães Rosa[4,5]

National DHIA Annual Meeting | San Antonio, TX | February 21-24, 2022 | (22)

AGIL – Miles

Genomic selection can only improve what we can measure, and as dairy production systems become more complex, so do the requirements of accurately measuring performance. There has been a lot of excitement around high-throughput phenotyping – which can measure all aspects of performance across a lifetime, and provide information useful for both management decisions and genetic improvement



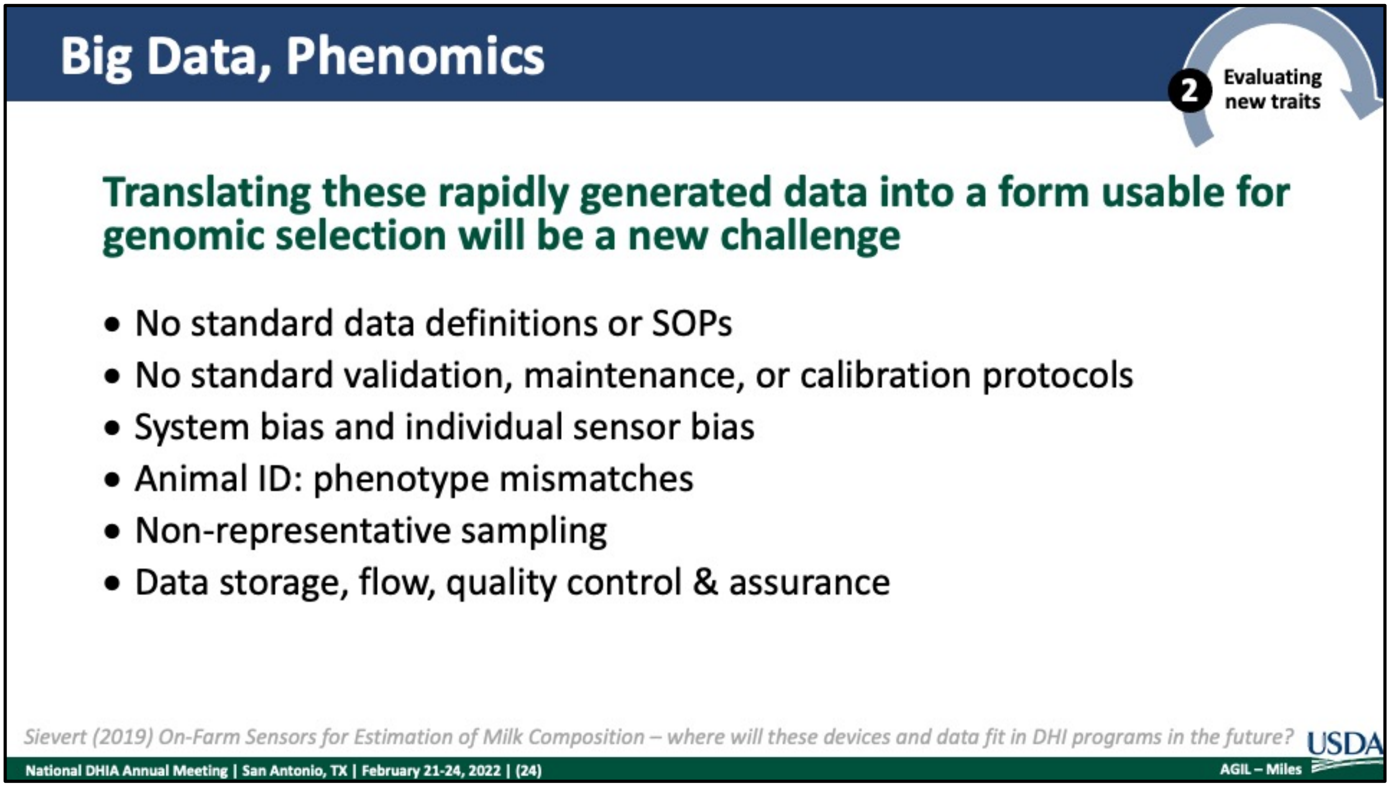
Big Data, Phenomics

These are very useful for management decisions:

- Monitoring subclinical mastitis
- Managing bulk tank SCC
- Culling
- Selective dry therapy

Sievert (2019) On-Farm Sensors for Estimation of Milk Composition – where will these devices and data fit in DHI programs in the future?

Just one example would be in automatic on-farm sensors for estimating milk composition. There are a variety of products on the market which can offer things like SCC estimation, milk flow rates, and even milk, fat and protein yield estimates. An important caveat to their use is that these systems have been designed to provide information that prompts a management outcome, and they are quite good at that, but not designed to create a data repository available for research use or even appropriate for research use.



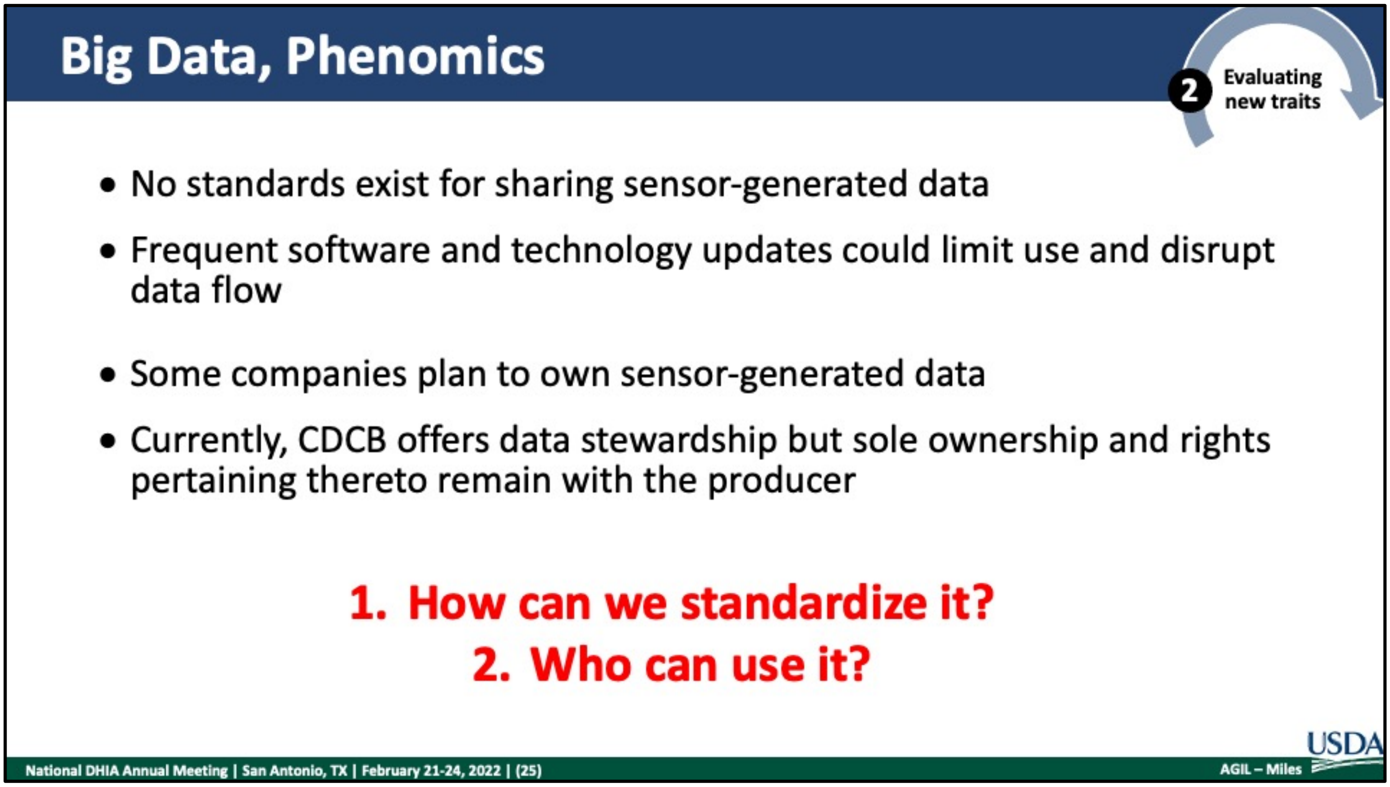
Big Data, Phenomics

Translating these rapidly generated data into a form usable for genomic selection will be a new challenge

- No standard data definitions or SOPs
- No standard validation, maintenance, or calibration protocols
- System bias and individual sensor bias
- Animal ID: phenotype mismatches
- Non-representative sampling
- Data storage, flow, quality control & assurance

Sievert (2019) On-Farm Sensors for Estimation of Milk Composition – where will these devices and data fit in DHI programs in the future?

No standard practices for users or validation, maintenance, and calibration protocols have been established which can creates both system bias and individual sensor bias. These autonomous systems rely on automatic animal ID detection during which there is ample opportunity for IDs to be mismatched with phenotypes. Anyone who's worked in a parlor can see how this can happen, with misbehaving cows skipping a stall or taking a joy-ride around the rotary. And sensors generate somatic cell alerts (for example) based on measurements taken during initial letdown – they 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. 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.

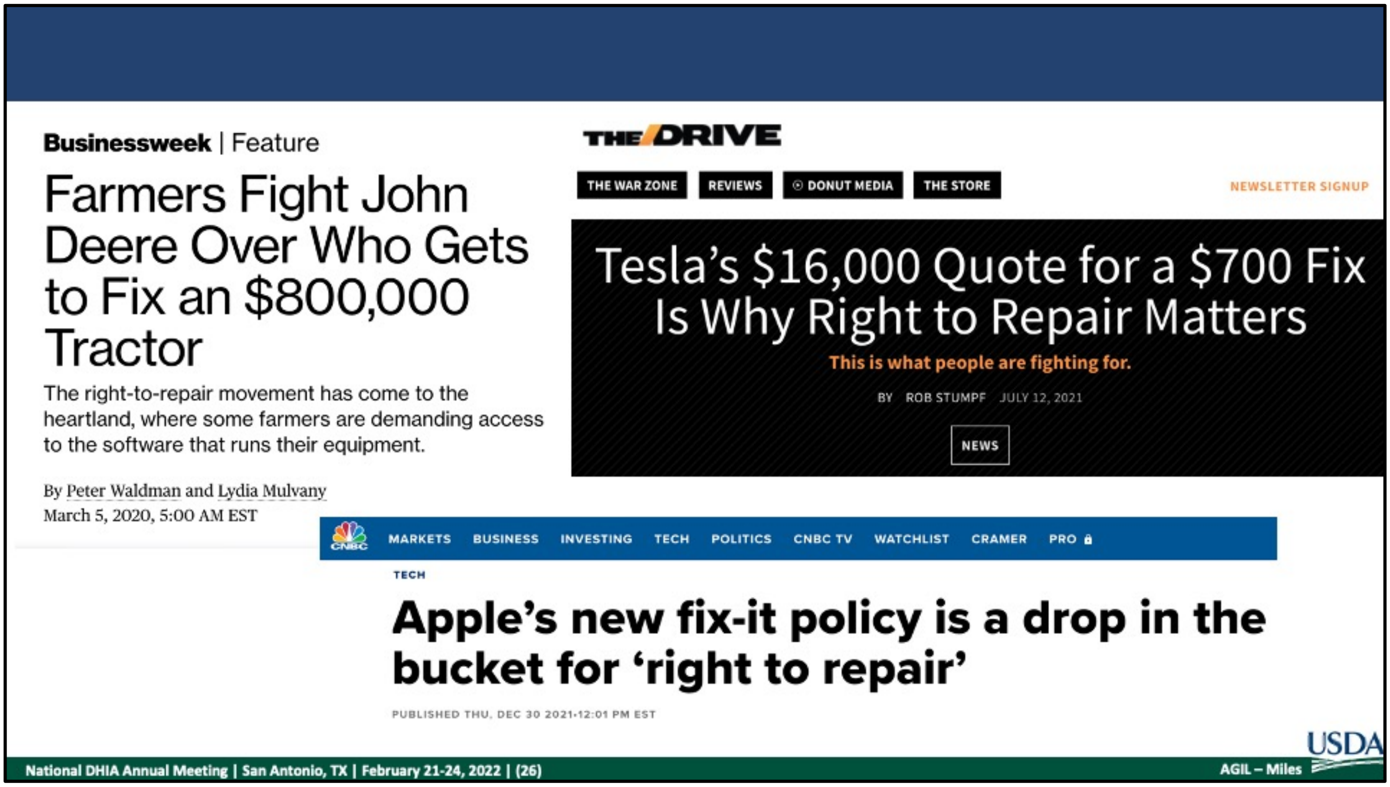


Big Data, Phenomics

- No standards exist for sharing sensor-generated data
- Frequent software and technology updates could limit use and disrupt data flow
- Some companies plan to own sensor-generated data
- Currently, CDCB offers data stewardship but sole ownership and rights pertaining thereto remain with the producer

1. How can we standardize it?
2. Who can use it?

And when we talk about a national scale, other questions arise. No standards currently exist for sharing sensor-generated data which may limit their widespread use and potential benefits. Frequent software and technology updates create problems for legacy equipment as repairs and maintenance may not be available for older versions which are otherwise suitable for use. This could also seriously impair the data pipelines that would have to be developed to get these data into the National Cooperator Database. Frequent updates can result in broken code and incompatibilities which are disruptive to data flow. Some companies are indicating that because they own the technology, they will own the data. That philosophy is in 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. It’s possible this new type of data will not be as usable as we think because the costs will become prohibitive. None of these challenges are insurmountable but they do need to be talked about. Two big questions to emerge in this new data-driven age are: how can we standardize it, and who can use it?



Businessweek | Feature

Farmers Fight John Deere Over Who Gets to Fix an $800,000 Tractor

The right-to-repair movement has come to the heartland, where some farmers are demanding access to the software that runs their equipment.

By Peter Waldman and Lydia Mulvany
March 5, 2020, 5:00 AM EST

Tesla's $16,000 Quote for a $700 Fix Is Why Right to Repair Matters

This is what people are fighting for.

BY ROB STUMPF JULY 12, 2021

NEWS

TECH

Apple's new fix-it policy is a drop in the bucket for 'right to repair'

PUBLISHED THU, DEC 30 2021·12:01 PM EST

We don't want to end up here. These questions of data ownership, proprietary software, and equipment usability are not unfamiliar… Healthy relationships are all founded in open communication and aligned expectations: if we anticipate our status quo will change, let's start this dialogue now so that we move into the future together in the full spirit of cooperation.

Thank you

Contact: asha.miles@usda.gov

https://www.ars.usda.gov/northeast-area/beltsville-md-barc/beltsville-agricultural-research-center/agil/