

Genomic Database/ Information System

H. Blackburn, M. Thomas, M. Enns and
T. Manahan

USDA-ARS-National Animal Germplasm Program and Colorado
State University, Department of Animal Science

Information Systems Have Lagged Behind Genotyping Automation

- ◉ General inability to effectively store and utilize large amounts of data
- ◉ Awareness that data may have value beyond original function

Genomic Data Concerns – Private Sector (Breed Associations)

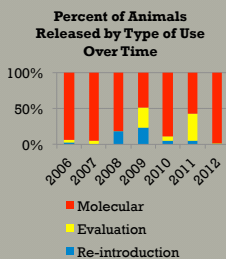
- ◉ Long term maintenance and use of their genomic data
- ◉ Maintenance of samples associated with the data
- ◉ Duplication of genotyping

Genomic Data Concerns – Public Sector Issues

- ◉ Long term data storage
- ◉ Long term sample storage
- ◉ Compliance with granting agencies
- ◉ Leveraging investment

Genomic Data Concerns – National Animal Germplasm Program

- ◉ Storage of genomic information associated with samples
- ◉ Increase the utility of the collection
- ◉ A high proportion of the 16,244 samples released for genomic studies
- ◉ Repatriation & storage of genomic results for future use



Why NAGP and CSU?

- ◉ NAGP
 - Mission to maintain and conserve animal genetic resources
 - Documenting phenotypes, management and genotypes are a necessary element
- ◉ CSU
 - Long term involvement with genetic prediction
 - Long term storage of samples and genotypes



Long Term Goal

- Develop a genomics component to Animal-GRIN to facilitate research and industry moving forward with genomic data utilization and collection utilization.

Existing Database Animal-GRIN V2

(<http://nrcc.ars.usda.gov/A-GRIN>)

- Database includes:
 - Samples – semen, blood, various tissues
 - Phenotypes &/or EPDs,
 - Management system
 - Environmental descriptors
 - Genotypes on Mendelian traits
- Existing capacities to handle multiple IDs per animal
- Public webpages allowing searches
- Secure physical location for data storage

Example of Animal with Samples in Repository

National Animal Genepool Program (NAGP)

Animal Details

Registration Number: 20713
 Registration Date: 2/1/2010
 Species: Cow
 Breed: Brown
 Sex: Female
 Birth Date: 2/1/2010
 Origin: United States, South Dakota
 Registration Number: 20713
 Registration Date: 2/1/2010

Molecular Observations

Descriptor Name	Value	Date Observed
SNP Array	Applied	
Genotype	Applied	
Genotype	Applied	
Genotype	Applied	
Genotype	Applied	
Genotype	Applied	

Phenotype Quantitative Observations

Descriptor Name	Value	Date Observed
Body weight	121.0 kg	
Adjusted 205-day weight	348.4 kg	
Adjusted 305-day weight	388.4 kg	
Milk yield	1775.8 kg	
Milk yield (kg)	1742.1 kg	
Milk yield (lb)	3827.0 lb	

Example of Animal with Samples in Repository

Circled/Shaded animal id's are animals have samples in repository

Scope of the Genomics Component

- Database is not a mechanism for genomic analysis
- It is constructed for long term storage of genomic data
- Data will not be stored in tabular structure of Animal-GRIN
- Data will be stored as separate files on a different system
- Sufficient information describing the data will be available from Animal-GRIN's public webpages

Infrastructure

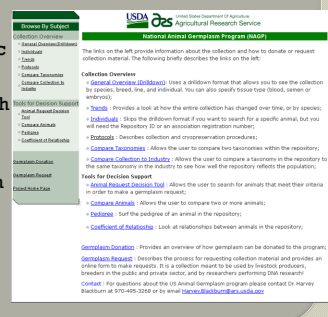
- Internet access at ~ 10 gigabytes per second
- Initial genomics storage system 780 terabytes with expansion capacity
- Existing development team: Agriculture and Agri-Foods Canada, EMBRAPA – Brazil Animal Science Department/CSU & ARS/NAGP

Conceptual Operation – Data Input

1. SNP data file received from donor
2. Data uploaded and format validated during the process
3. Necessary links between animal id and summary information in/with existing Animal-CRIN components
4. File placed in an external file separate from the existing Animal-CRIN database
5. File remains in the external file until requested (either in part or whole)

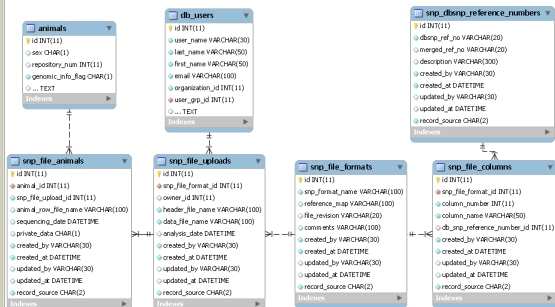
Conceptual Operation – Data Request

1. Donor requests all or part of data on specific animal(s)
2. Data file is created with the requested information
3. Requested information shipped to requestor



Five New Database Tables

Animals and db_users tables already exist

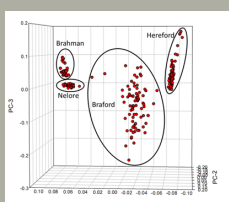
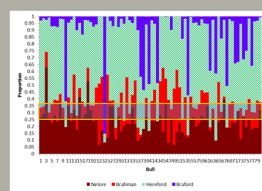


How Might Database be Used

- Question – Do composite breeds maintain expected composition of founder breeds in the US and Brazil?
- 770K SNP data stored for Braford, Hereford, Brahman & Nelore
- Researchers determined they only wanted to use a predetermined panel of 60K SNPs
- Request from the database specific SNPs for the breeds of interest

Hypothetical Example – US & Brazil Braford

• Results suggest Braford maintain 3/8's 5/8's composition but there is variability.

Program Development and Use

Currently 3-4 months ahead of schedule



Final Points

- ⊙ Genomics component will address a number of public and private sector concerns:
 - Long term storage of data
 - Leveraging previous investments
 - Ability to access data for multiple projects
 - Linkage between physical samples, phenotypes, management systems
- ⊙ Provides a long term stable platform for both public and private sector use