National Genetic & Genomic Evaluations
Present and Future

Dorian Garrick[1][2][3]

[1]National Beef Cattle Evaluation Consortium
[2]Theta Solutions LLC
[3]Department of Animal Science, Iowa State University

NBCEC Mission

- ***Develop and implement improved predictions so selection can enhance economic viability of US beef cattle producers***
- Goal to be able to provide science-based "*Genetic evaluation of pure- and crossbred animals for any economically relevant trait and management circumstances*"

www.nbcec.org

Prediction of Merit

- Philosophical concept embodied in the "model" that is the basis for prediction
- Statistical method used to estimate effects and perhaps other parameters in the model
- Computing algorithm(s) to implement the statistical method

Philosophical Concept

- A Model describes cause and effect - the underlying process believed to result in the observations

Performance = Breeding + Feeding

Phenotype = Genotype + Environment

- The model (or a simplification of the model) is the basis for prediction

Model Equation

$$y = Xb + Zu + e$$

Vector of phenotypes (performance)

Vector of non-genetic effects herd-year age of dam date-of-birth (fixed effects)

Vector of additive genetic (random) effects (EPD)

Vector of leftover parts we do not know how to model and cannot explain (residuals)

This represents a "mixed" model (as it contains fixed and random effects)

Computing Algorithm(s)

- Henderson invented an efficient strategy to predict EPD based on mixed model equations

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \lambda A^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Scalar Variance Ratio

$$\lambda = (1-h^2)/h^2$$

Inverse of pedigree-based relationship matrix

Single trait – readily extends to include multibreed, maternal effects and multiple traits

Pre-genomics approach using MME

- Set up and solve mixed model equations MME
  - Seen as a big job so only done 2-3x per year
- Use methods that approximate accuracy
- Approximate mixed model equations with separate interim calculations that can be run regularly in between major runs
  - Interims approximate EPD and accuracy

Parallel Developments

- Colorado State University
  - Bruce Golden developed ABTK in C
    - Animal Breeders Tool Kit – publicly available source
- Cornell University (Quaas, Pollak etc)
  - Developed multibreed Fortran code with ASA
- Iowa State University
  - Doyle Wilson developed Fortran code for AAA
- University of Georgia (+ Benyshek & Bertrand))
  - Ignacy Misztal develop Fortran code (BlupF90)
- University of New England/AGBU (Bruce Tier)
  - Developed Fortran code for Breedplan

National Beef Cattle Evaluation Consortium

- Co-ordinated research across US Universities
- Moved routine servicing – i.e. running evaluations from the 4 universities to breed associations
  - Led to some consolidation of approaches
- Tried to develop and fund a software development center at University of Georgia
  - To be funded by $1 per new animal registration
- All backed up with coordinated beef improvement extension and outreach programs

UGa/Misztal BlupF90 software

- American Angus Association contracted to use UGa/Misztal BlupF90 software in house through their subsidiary AGI
  - Ultimately moved to weekly runs (no interims)
  - Uses a 10 million animal pedigree +300k per year
  - Incudes fitting of about 10 different models for different subsets of traits
- American International Charolais Association
  - Contracted AGI to run their evaluations
  - Includes a pedigree a little over 1 million animals
  - Both Charolais and Charolais-cross data

Situation Today

Breedplan/AGBU/Bruce Tier

- American Hereford Association
  - Signed up with ABRI/Breedplan
    - merged separate Polled and Horned Herefords
  - Partnered with Canada, Argentina, and Uruguay to run Pan American Cattle Evaluation (PACE)
    - 13 trait growth, carcass & ultrasound evaluation plus a separate calving ease evaluation
    - Includes recent animals in a 6 million animal pedigree
    - Within-breed analysis

Situation Today

Cornell University/ASA Software

- Multibreed software moved in-house to Bozeman and run 2x per year
  - Attracted other breed associations for joint runs that included admixed and/or composite cattle
    - Red Angus, Limousin (LimFlex), Gelbvieh (Balancer)
    - Maine Anjou, Shorthorn
    - Plus many Canadian breed associations
  - Formed International Genetic Solutions (IGS)
  - Now uses a pedigree of 15 million animals +340k/yr
  - No remaining “inventor” support from Quaas/Pollak

Situation Today

Some Other Boutique Evaluations

- ABTK/Colorado State University
  - Continued to use ABTK to run certain analyses (eg stayability) for some breed associations
- Livestock Genetic Services/John Genho
  - Uses his own matlab code to run evaluations for Santa Gertrudis and Brangus

Situation Today

Developments – last 20 years

- Javaremi, Smith, Gibson (1997)
  - Showed how markers could be used to construct genomic instead of pedigree relationships (GBVM)
- Meuwissen, Hayes, Goddard (2001)
  - Described the EBV as sum of marker effects estimated by fitting marker effects (Marker Effects Models MEM)
    - BayesA, BayesB, RR-BLUP
  - Introduced Markov chain Monte Carlo (MCMC) to mainstream animal breeding applications
- Stranden & Garrick (2009) and others
  - Showed that EPDs were the same from GBVM & MEM

Developments – last 10 years

- Illumina SNP Chips
  - 50k and other density markers
- GeneSeek routine genotyping
  - Economies of scale

How do you react to new technology?

Sometimes see and grab new opportunities
But sometimes outweighed by new challenges

Different people see opportunities in different areas
(this is good!)

Reaction to New Opportunities

- University of Georgia (and collaborators) Misztal, Legarra, Aguilar etc
  - Worked to improve modeling of relationships
    - Had the advantage of leveraging existing software
    - Considerable development based on experiences with large national datasets for dairy, pigs and chickens

Single Step HBLUP

- Modelling covariance among relatives

$$H = var\begin{bmatrix} u_n \\ u_g \end{bmatrix}\sigma_a^{-2} = \begin{bmatrix} A_{nn} + A_{ng}A_{gg}^{-1}G_{gg}A_{gg}^{-1}A_{gn} & A_{ng}A_{gg}^{-1}G_{gg} \\ G_{gg}A_{gg}^{-1}A_{gn} & G_{gg} \end{bmatrix}$$

Legarra et al (2009)

- With inverse (for full rank G)

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G_{gg}^{-1} - A_{gg}^{-1} \end{bmatrix}$$

Aguilar et al (2010)

- The matrix **H**$^{-1}$ gets used in place of **A**$^{-1}$ in mixed model equations (eg in BlupF90)

Single Step GBLUP Mixed Model Equations

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{W} + \mathbf{H}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix}$$

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{pmatrix}$$

Minor modifications allow optimization of dense submatrix of H^{-1}

Yet to be routinely implemented in large national evaluations of beef cattle
– see Laurenco talk tomorrow re AAA prototype

Reaction to New Opportunities

- Iowa State University
  Fernando, Garrick, Dekkers, students & postdocs
  - Tried to understand & improve marker effects model
    - Developed BayesC, BayesCπ, BayesN, QTL Models etc
    - Required new software (GenSel) and learning about MCMC
    - Extended to categorical data
    - Extended to include dominance effects
    - Extended to fit haplotypes and QTL
    - Tested in a variety of species with > 600 global users
      - pigs, chickens, dairy, human, maize, barley, rice, trees, fish etc

Genotyped Animals

$$y_g = X_g b + Z_g u_g + e_g$$

Meuwissen, Hayes & Goddard (2001)

$$with\ u_g = M_g \alpha = \sum_{j=1}^{j=\#loci} m_j \alpha_j \delta_j$$

$$\alpha_j = substitution\ effect$$

$$\delta_j = (0,1)\ indicator\ variable$$

Marker Effects Models (MEM)

$$\begin{bmatrix} X'X & X'ZM \\ M'Z'X & M'Z'ZM + \phi \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} X'y \\ M'Z'y \end{bmatrix}$$

According to the choice of this diagonal matrix
and the use of variable selection
These equations represent RR-BLUP,
BayesA, BayesB, BayesC, BayesCπ, BayesN, BayesR etc

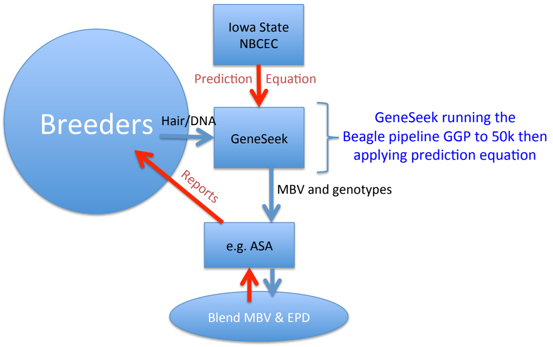
We have implemented these models in GenSel and/or in our Julia software (QTL.rocks)
GenSel developments were undertaken using USDA-NIFA funding BIGS and e-BIGS

Current two-step use of Genomics

- GenSel software was used to develop prediction equations to produce MBV for beef cattle marketed via
  - Merial/Igenity now owned by GeneSeek
  - Pfizer now rebranded as Zoetis
  - Dr. Mahdi Saatchi for Hereford, Simmental, Red Angus, Gelbvieh, Limousin, Brangus breed assocs etc
  - "Training" used deregressed EPDs as "data"
- AAA - Zoetis generates the MBV
- RAAA - Zoetis or GeneSeek generates MBV
- Other breeds GeneSeek generates the MBV

USDA competitive funding had a big impact on implementing genomic prediction in US beef cattle

Genomic Prediction Pipeline



Iowa State NBCEC

Prediction Equation

Breeders

Hair/DNA

GeneSeek

GeneSeek running the Beagle pipeline GGP to 50k then applying prediction equation

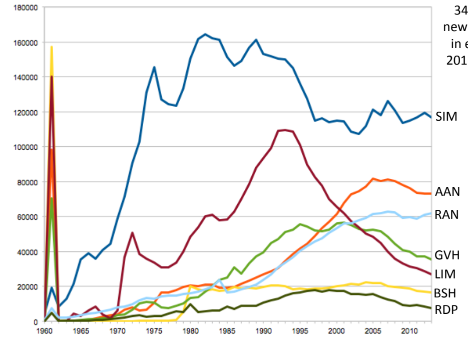
Reports

MBV and genotypes

e.g. ASA

Blend MBV & EPD

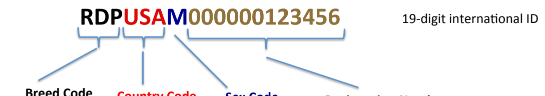
IGS Pedigree Counts by Birth Year



340,000 new animals in each of 2012-2013

Animal Identifiers

- Now use a variant of the Interbull ID system



Breed Code
AAN=Angus
BRG=Brangus
BSH=Shorthorn
CHA=Charolais
HER=Hereford
LIM=Limousin
NEL=Nellore
RAN=Red Angus
RDP=Maine-Anjou
SIM=Simmental

Country Code
ARG
AUS
CAN
URG
USA

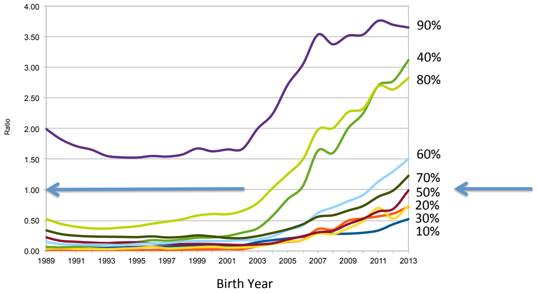
Sex Code
M=bull
F=cow
U=unknown

Registration Number
Left-padded with 0
Can include alphanumerics

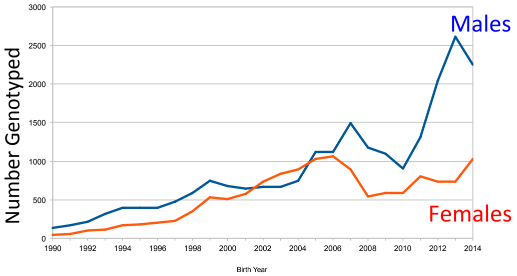
We use Breed Association rather than Breed
(unless animals are not registered)
Prefer to use country/breed of first registration

International IDs with SIM code

Fold numbers of crossbred to 100% Simmentals

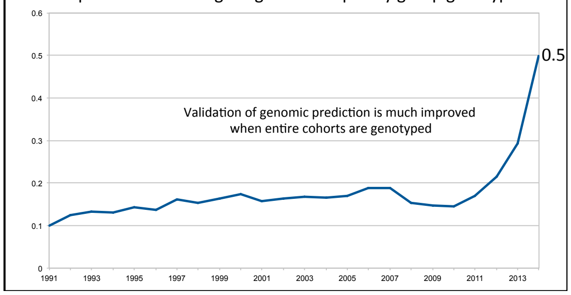


IGS - Most Recent Evaluation



IGS – most recent evaluation

Proportion of weaning weight contemporary group genotyped



Validation of genomic prediction is much improved when entire cohorts are genotyped

Incorporation of MBV in NCE

- AAA (first to introduce genomic predictions)
  - Use MBV as a correlated trait(s) in each of the multi-trait analyses for a class of traits
  - Now have 130,000 animals with 50k genotypes!
- AHA & IGS partners (ASA, RAAA, AGA, NALF)
  - Use selection index blending to pool information from pedigree analysis and MBV
  - About 15,000 HER and 35,000 SIM genotypes (IGS 55,000)
    - Wide variety of SNP chip densities – 50k, 700k, GGP-LD, GGP-HD
- Santa Gertrudis – small national evaluation
  - Uses Single Step HBLUP

Selection Index Blending Assumptions

Pb = g

$$var\begin{bmatrix}\widehat{u}\\ \widehat{m}\\ u\end{bmatrix} = \begin{bmatrix} r_p^2 & r_p^2 r_m^2 & r_p^2\\ r_p^2 r_m^2 & r_m^2 & r_m^2 \\ r_p^2 & r_m^2 & 1\end{bmatrix}\sigma_g^2$$

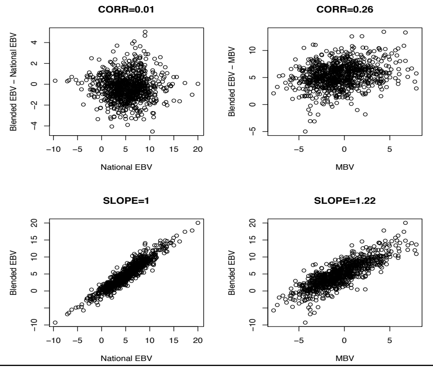
$$var\begin{bmatrix}u-\widehat{u}\\ m-\widehat{m}\end{bmatrix} = \begin{bmatrix} 1-r_p^2 & (1-r_p^2)(1-r_m^2)\\ (1-r_p^2)(1-r_m^2) & 1-r_m^2\end{bmatrix}$$

Kachman (unpublished)

Diagnostics of Good Behavior

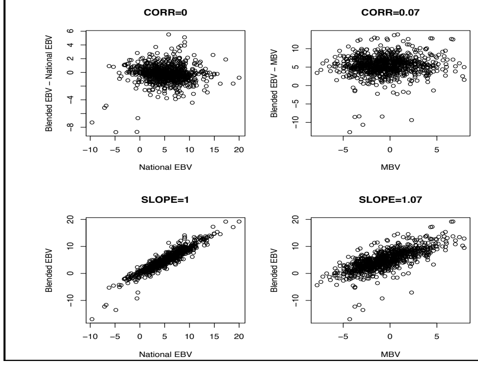
- Diagnostics will be provided routinely
- Regression of more accurate EPDs on less accurate EPDs should be 1
- Correlation of less accurate EPDs with change in EPDs (from less accurate to more accurate) should be zero

Validation of Birth Weight



Results from HER based on two step selection index blending

Inflation of EBV/MBV covariance



Results from HER based on two step selection index blending

For a variety of reasons everyone would prefer a single step approach combining pedigree, performance & genomics in one analysis

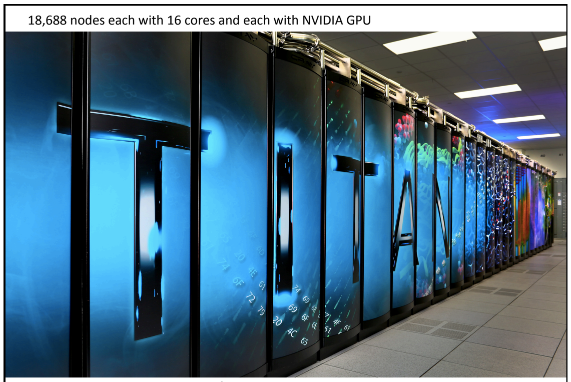
More New Technologies

- Multi-core CPU
  - Increasingly adopted over last 10 years
  - Reduces power demand and avoids overheating
- Graphics Cards
  - Spin off from gaming industry
  - Used for arithmetic calculations over last 5 years

Leverage technology built for computer gaming

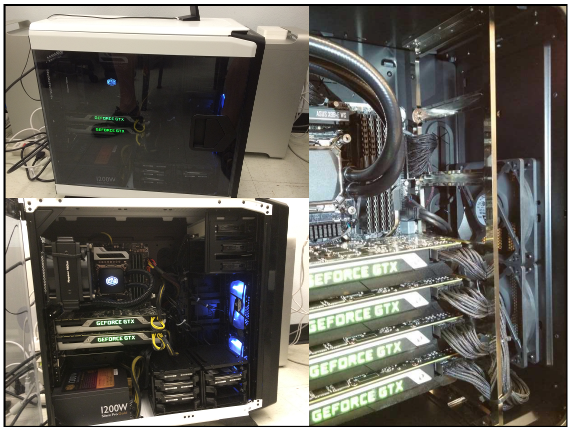
Computer gaming/animation is >$100 billion per year industry!!

18,688 nodes each with 16 cores and each with NVIDIA GPU



10x more powerful than predecessor Jaguar but uses same space and power

Cray supercomputer at Oak Ridge National Laboratories – fastest in US, 2nd fastest in world



More New Technologies

- Alternative computing strategy for single step based on the same model as single step HBLUP
  - But facilitates fitting other marker effects models….

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'ZM} & \mathbf{X'Z_1} \\ \mathbf{M'Z} & \mathbf{M'Z'ZM}+\phi & \mathbf{M_1'Z_1'Z_1} \\ \mathbf{Z_1'X} & \mathbf{Z_1'Z_1M_1} & \mathbf{Z_1'Z_1}+\mathbf{A}^{11}\lambda \end{bmatrix}\begin{bmatrix}\hat{\mathbf{b}}\\ \hat{\alpha}\\ \hat{\varepsilon}\end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{M'Z'y} \\ \mathbf{Z_1'y}\end{bmatrix}$$

1=non-genotyped

Fernando et al (2014) GSE

Implemented in GenSel prototype for testing practicality

Single Step Hybrid Model Mixed Model Equations

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'ZM} & \mathbf{X'Z_1} \\ \mathbf{M'Z} & \mathbf{M'Z'ZM}+\phi & \mathbf{M_1'Z_1'Z_1} \\ \mathbf{Z_1'X} & \mathbf{Z_1'Z_1M_1} & \mathbf{Z_1'Z_1}+\mathbf{A}^{11}\lambda \end{bmatrix}\begin{bmatrix}\hat{\mathbf{b}}\\ \hat{\alpha}\\ \hat{\varepsilon}\end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{M'Z'y} \\ \mathbf{Z_1'y}\end{bmatrix}$$

1=non-genotyped

First attempt at full-scale implementation

Shared in 2013 with Livestock Improvement Corporation - runs a large dairy evaluation and co-inventor of Aguilar et al (2010) single step HBLUP strategy

Didn't think it was computationally feasible

Fernando et al (2014) GSE

Single Step Hybrid Model Mixed Model Equations

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'ZM} & \mathbf{X'Z_1} \\ \mathbf{M'Z} & \mathbf{M'Z'ZM}+\phi & \mathbf{M_1'Z_1'Z_1} \\ \mathbf{Z_1'X} & \mathbf{Z_1'Z_1M_1} & \mathbf{Z_1'Z_1}+\mathbf{A}^{11}\lambda \end{bmatrix}\begin{bmatrix}\hat{\mathbf{b}}\\ \hat{\alpha}\\ \hat{\varepsilon}\end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{M'Z'y} \\ \mathbf{Z_1'y}\end{bmatrix}$$

1=non-genotyped

Second attempt at full-scale implementation

Applied for USDA-AFRI funds and were turned out

Fernando et al (2014) GSE

Single Step Hybrid Model
Mixed Model Equations

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}\mathbf{M} & \mathbf{X}'\mathbf{Z}_1 \\ \mathbf{M}'\mathbf{Z} & \mathbf{M}'\mathbf{Z}'\mathbf{Z}\mathbf{M} + \phi & \mathbf{M}_1'\mathbf{Z}_1'\mathbf{Z}_1 \\ \mathbf{Z}_1'\mathbf{X} & \mathbf{Z}_1'\mathbf{Z}_1\mathbf{M}_1 & \mathbf{Z}_1'\mathbf{Z}_1 + \mathbf{A}^{11}\lambda \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\alpha} \\ \hat{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{M}'\mathbf{Z}'\mathbf{y} \\ \mathbf{Z}_1'\mathbf{y} \end{bmatrix}$$

1=non-genotyped

Third attempt at full-scale implementation

Privately developed the software through Theta Solutions LLC

Gibbs Sampler
e.g. Bayes C (known π)

$$P(\theta \mid \mathbf{X})$$

Distribution of EPDs/data

Fernando et al (2014) GSE

Funding Model

- Annually licensed "BOLT" software with licensees' receiving ongoing updates
  - BOLT=Biometry Open Language Tools
  - Theta Solutions will produce efficient implementations for new methods and algorithms
- Day-to-day application of the software is the licensee's responsibility (not Theta Solutions)

Theta Solutions LLC

Single Step Hybrid Model

- Extended single step hybrid models to fit
  - multiple trait models
  - to allow many random factors per trait
    - including maternal & permanent environmental effects
  - to allow different marker effects models for different traits
- Using a portfolio of BOLT command line tools
- On inexpensive workstations using CUDA cards

Theta Solutions LLC

Single Site Gibbs Sampler

Solve x in Ax=b

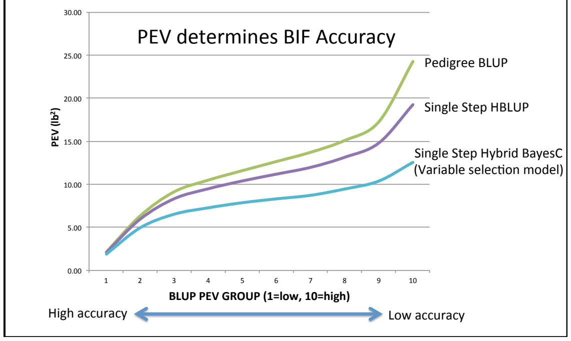
```
...
for( sample=0; sample<nSamples; sample++ )
{
    for( j=0; j<x.num_elem; j++ )
    {
        ax = rowDot( &A, &x, j ) - x.v[j] * diagA.v[j];
        xHat = ( b.v[j] - ax ) / diagA.v[j];
        x.v[j] = xHat + nrdGpu() * sqrtInvDiagA.v[j];
    }
}
...
```

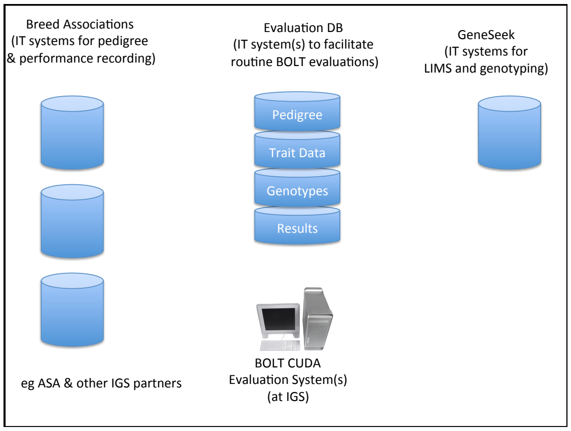
Markov chain Monte Carlo (MCMC) sampler for mixed model in national cattle evaluation

Example:
Simmental Birth Weight Analysis

N Animals total:	2,593,580
N Genotyped	13,867
N Imputed	2,579,713
N Observations	1,959,890

Variable Selection Models have more accurate EPD





Breed Associations
(IT systems for pedigree
& performance recording)

Evaluation DB
(IT system(s) to facilitate
routine BOLT evaluations)

GeneSeek
(IT systems for
LIMS and genotyping)

Pedigree

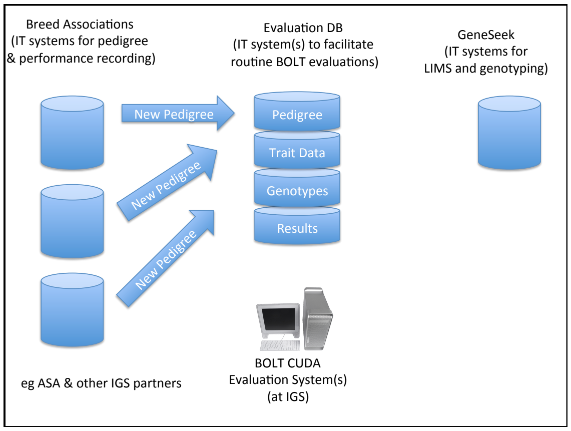
Trait Data

Genotypes

Results

eg ASA & other IGS partners

BOLT CUDA
Evaluation System(s)
(at IGS)



Breed Associations
(IT systems for pedigree
& performance recording)

Evaluation DB
(IT system(s) to facilitate
routine BOLT evaluations)

GeneSeek
(IT systems for
LIMS and genotyping)

New Pedigree

New Pedigree

New Pedigree

Pedigree

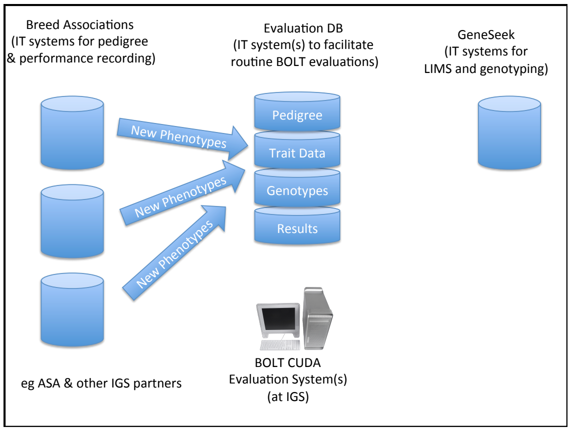
Trait Data

Genotypes

Results

eg ASA & other IGS partners

BOLT CUDA
Evaluation System(s)
(at IGS)



Breed Associations
(IT systems for pedigree
& performance recording)

Evaluation DB
(IT system(s) to facilitate
routine BOLT evaluations)

GeneSeek
(IT systems for
LIMS and genotyping)

New Phenotypes

New Phenotypes

New Phenotypes

Pedigree

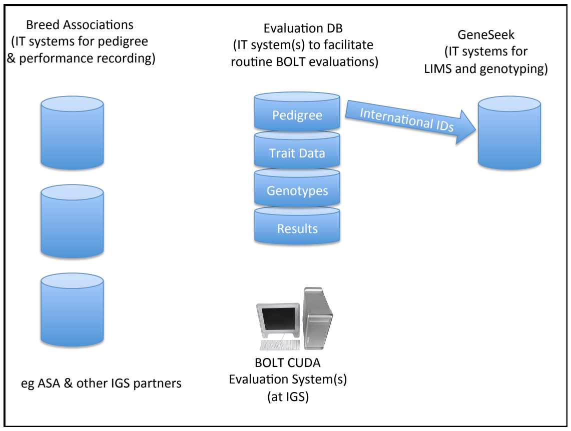
Trait Data

Genotypes

Results

eg ASA & other IGS partners

BOLT CUDA
Evaluation System(s)
(at IGS)



Breed Associations
(IT systems for pedigree
& performance recording)

Evaluation DB
(IT system(s) to facilitate
routine BOLT evaluations)

GeneSeek
(IT systems for
LIMS and genotyping)

Pedigree

International IDs

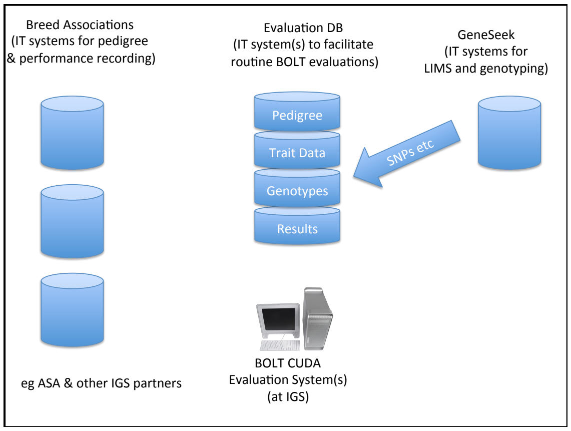
Trait Data

Genotypes

Results

eg ASA & other IGS partners

BOLT CUDA
Evaluation System(s)
(at IGS)



Breed Associations
(IT systems for pedigree
& performance recording)

Evaluation DB
(IT system(s) to facilitate
routine BOLT evaluations)

GeneSeek
(IT systems for
LIMS and genotyping)

Pedigree

Trait Data

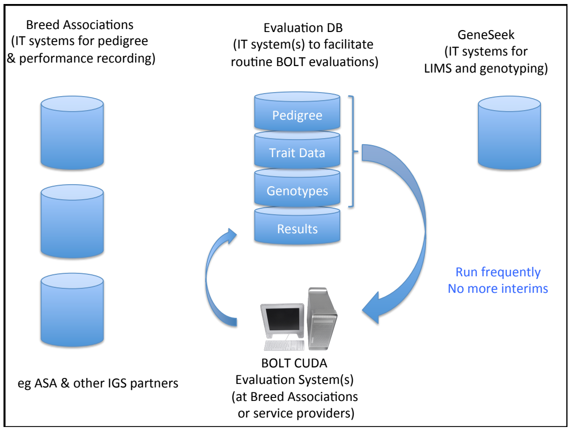
SNPs etc

Genotypes

Results

eg ASA & other IGS partners

BOLT CUDA
Evaluation System(s)
(at IGS)



Breed Associations
(IT systems for pedigree
& performance recording)

Evaluation DB
(IT system(s) to facilitate
routine BOLT evaluations)

GeneSeek
(IT systems for
LIMS and genotyping)

Pedigree

Trait Data

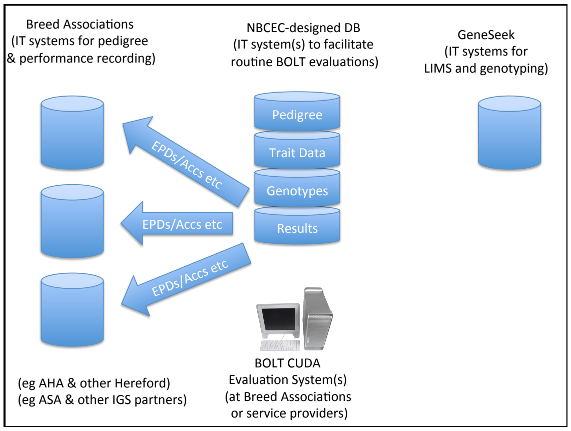
Genotypes

Results

Run frequently
No more interims

eg ASA & other IGS partners

BOLT CUDA
Evaluation System(s)
(at Breed Associations
or service providers)



Breed Associations
(IT systems for pedigree
& performance recording)

NBCEC-designed DB
(IT system(s) to facilitate
routine BOLT evaluations)

GeneSeek
(IT systems for
LIMS and genotyping)

Pedigree

Trait Data

Genotypes

Results

EPDs/Accs etc

EPDs/Accs etc

EPDs/Accs etc

(eg AHA & other Hereford)
(eg ASA & other IGS partners)

BOLT CUDA
Evaluation System(s)
(at Breed Associations
or service providers)

Haplotype model

- Expect a continuous migration in SNP chips
  - Gradually include causal mutations
  - Increase in SNP density (at same cost?)
- Expect to move towards fitting of haplotype effects rather than SNP effects
  - Haplotypes represent the SNP alleles inherited on one chromosome fragment that came from either the sire or the dam
  - Research being undertaken through a Zoetis PDF

Future National Evaluations

- Will run almost continuously
- Constantly improving genomic features
  - Markers to haplotypes to causals
  - Results from many researchers and projects
- Evolving models as genomic prediction matures
  - further refinements to single step, multibreed etc
  - Inventions throughout the world
- Strategies based on Markov chain Monte Carlo will become routine for all evaluations

Summary

- Breed Associations will benefit from
  - adopting international ID systems as an integral part of their databases
  - upgrading their IT systems to facilitate automated data extractions and imports to evaluation systems as these evolve

Summary

- Theta Solutions LLC is on track to deliver BOLT software for fitting single step hybrid models (and many other kinds of single step and pedigree models) by 1 January 2016
  - IGS and AHA are currently prototyping BOLT
- AAA is testing University of Georgia single step HBLUP and planning to upgrade hardware to implement single step in the next 12 months

Summary

- Although the NBCEC is no longer formally funded by USDA, the consortium of interested research and extension personnel are still working together to ***develop and implement improved predictions so selection can enhance economic viability of US beef cattle producers***