

Tailoring Single-Step Evaluations for the U.S. System

Taylor McWhorter, Simone Savoia,
Andres Legarra, and John B. Cole



Take-home messages

- Single-step evaluations will make the best possible use of all available information – pedigrees, phenotypes, and genotypes
- There will be some re-ranking of young genotyped animals because there will be less bias in the new evaluations
- Single-step is a refinement of existing tools, not a revolutionary change
- As we transition to single-step, our systems are being overhauled to make it easier for us to meet future industry needs

Topics to discuss

- Predicted transmitting abilities (PTA)
- Statistical methods
 - Legacy “multi-step”
 - New “single-step”
- Why switch methods?
- Implementation philosophy at CDCB
- What we can and cannot expect



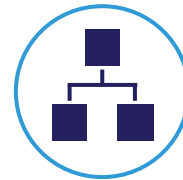
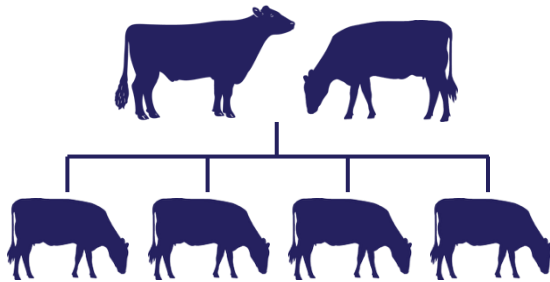
PTA: *before* genomics



Predicted Transmitting Ability (PTA):

Genetic merit that a parent can pass on to progeny

Relationships are very important to estimating PTA



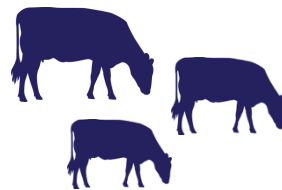
Can measure expected relationship through a pedigree



“We expect a sire to pass on 50% of his genes”



Observe animals'
traits



Make connections
to family members

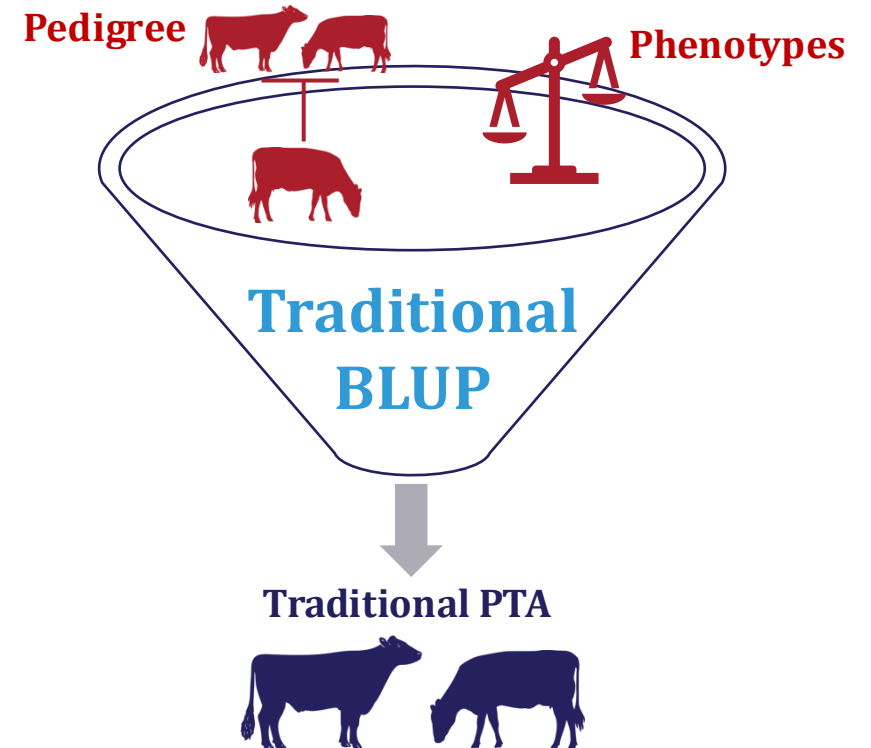


And their
performance

PTA calculation *before* genomics

Method: **BLUP** or **traditional BLUP** or **pedigree BLUP (pBLUP)**

- **BLUP** = *best linear unbiased prediction*
- Output: traditional **PTA**
 - AKA PTA without genomic information



PTA: *after* genomics

Measure more accurate relationships with genomic information



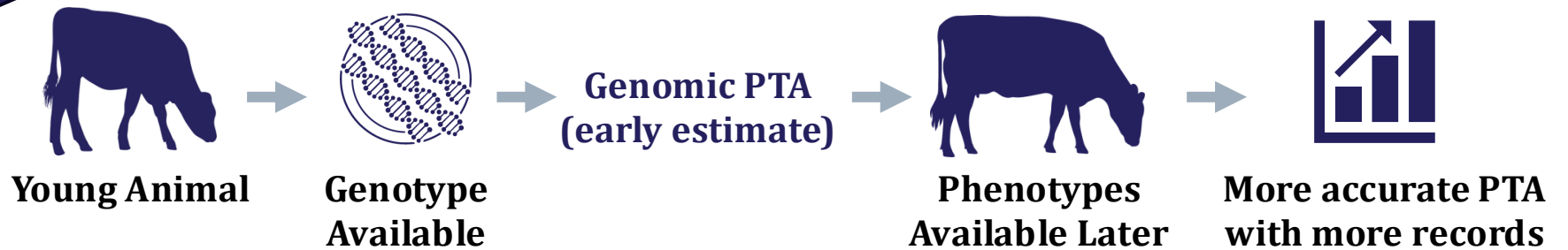
More information
= more accurate

Performance Information

Genomic Information

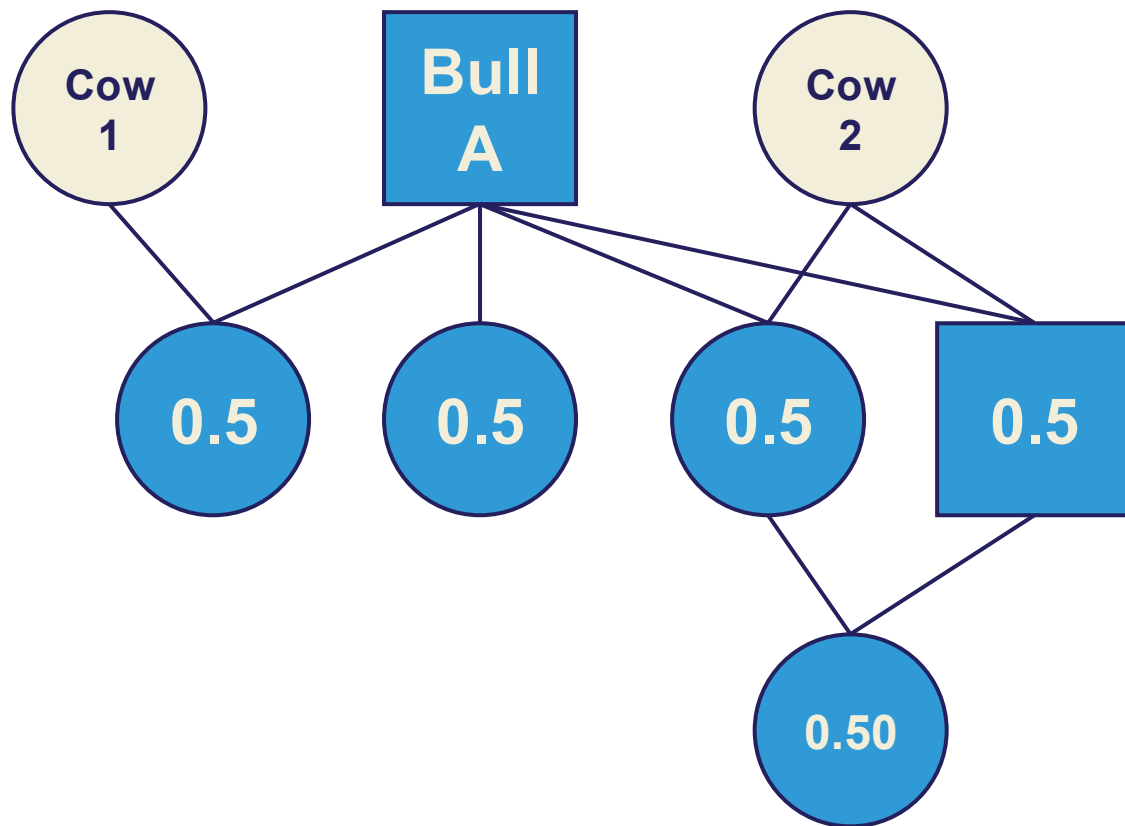
Higher Accuracy

Earlier estimations

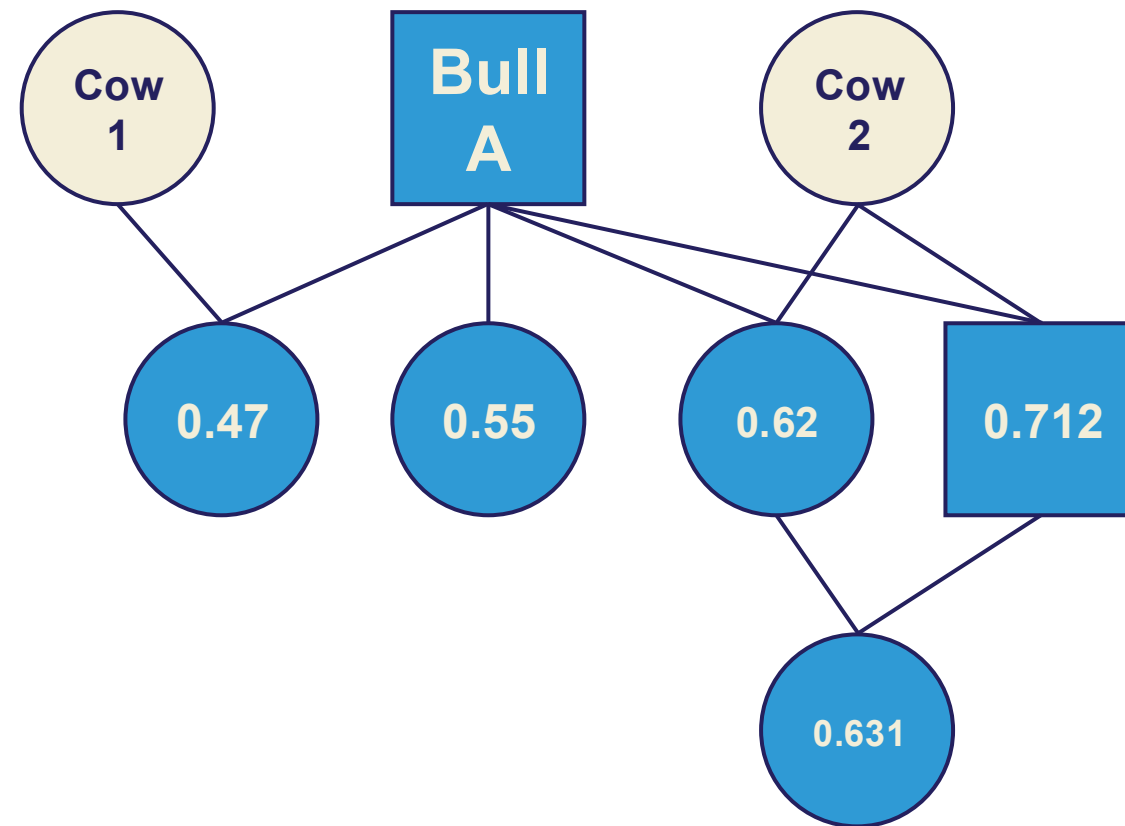


Needed a new statistical method to include genomics

DNA measures true relationships

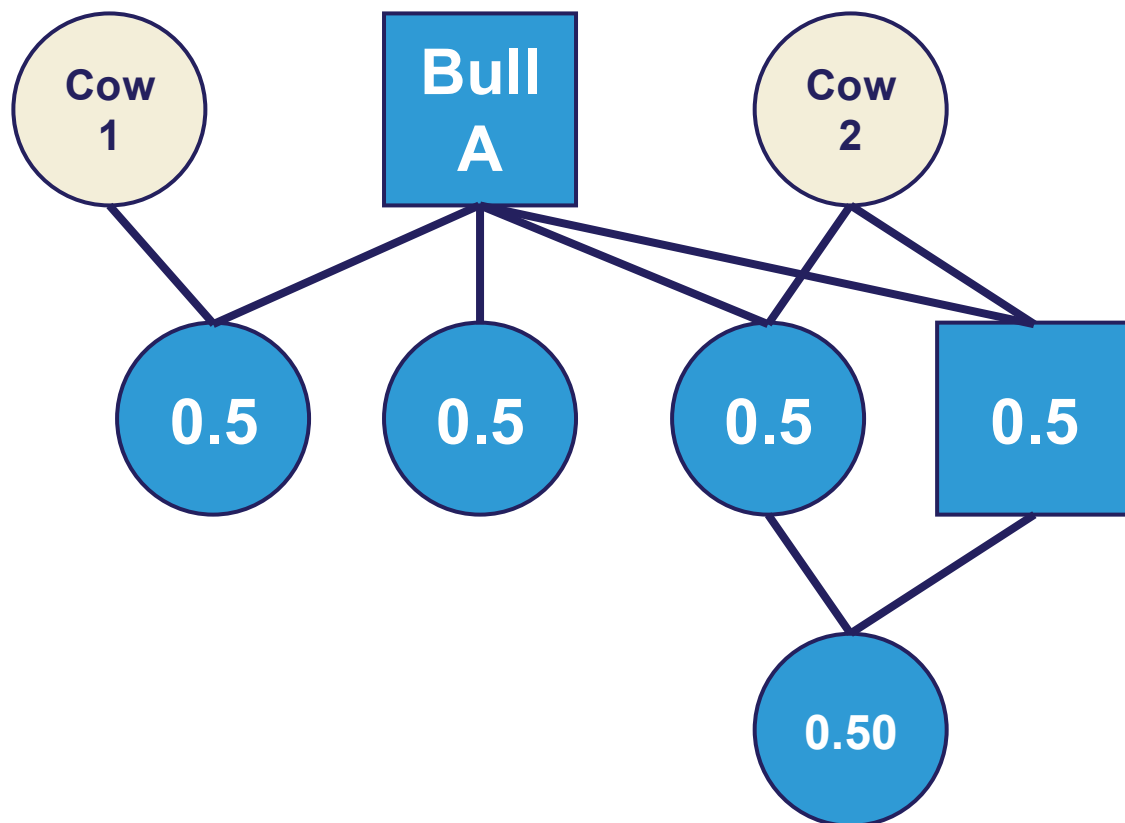


Pedigree relationship of Bull A to descendants

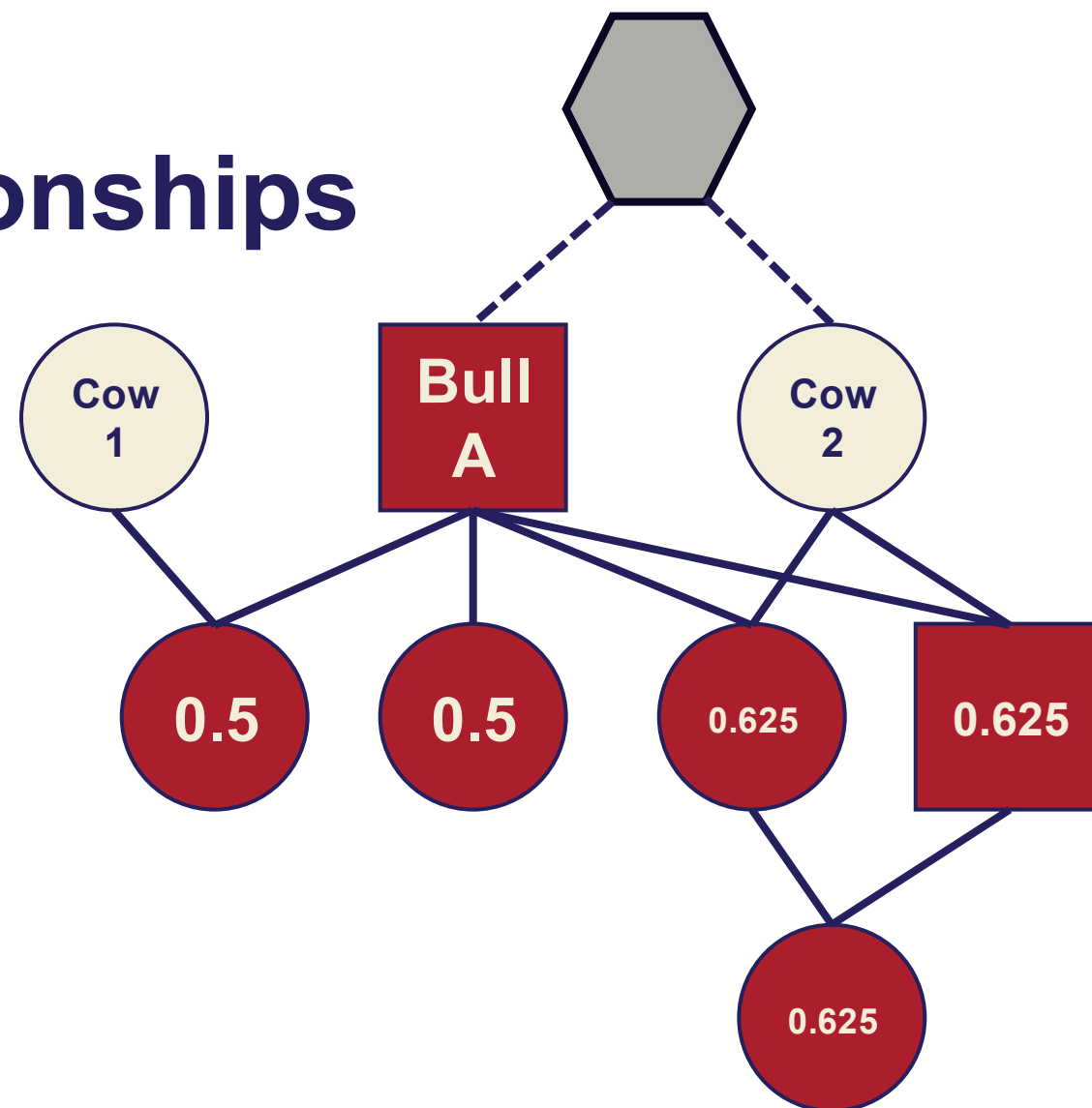


Genomic relationship of Bull A to descendants

DNA identifies all relationships



Pedigree says Bull A and Cow 2 are unrelated



Genomics tell us they're related after all

Statistical methods *after* genomics

Multi-step



Developers:
Paul VanRaden
and USDA AGIL



Year Published:
2008

Single-step



Developers:
Ignacy Misztal, Andres Legarra,
Ignacio Aguilar, and a large
team at University of Georgia



Year Published:
2009



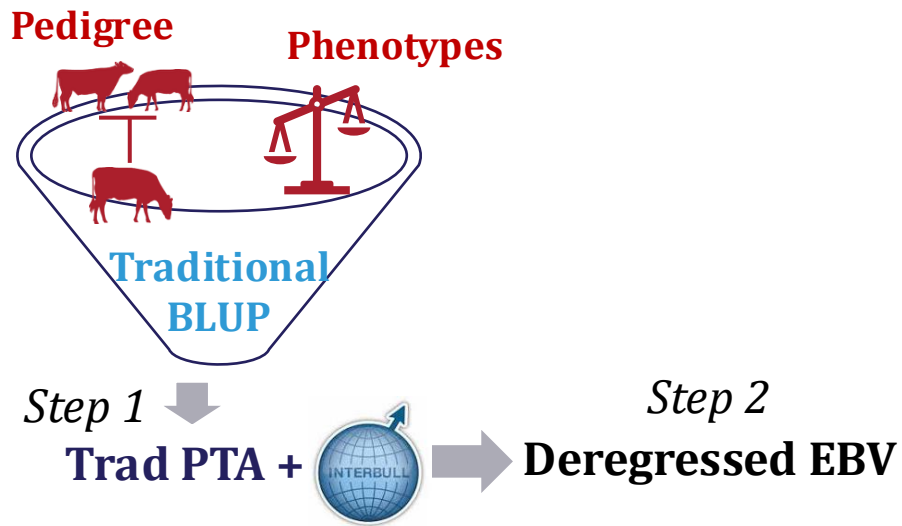
**Many developments
since**

SUITE OF PROGRAMS:

BLUPF90+
BLUPF90IOD3
GIBBSF90
ETC.

PTA calculation *after* genomics

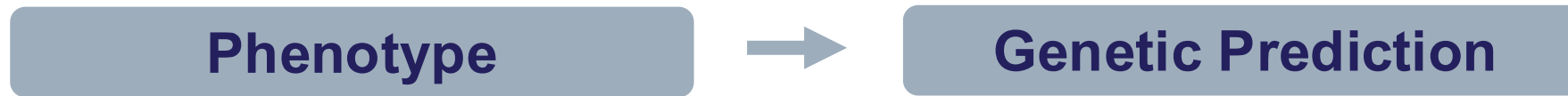
Legacy multi-step method



1. Use **phenotypes + pedigrees** first → traditional PTA
2. Traditional evaluations for a subset of reference animals are deregressed
 - Obtain *pseudo* phenotypes for genomic evaluation

Deregression or “backsolving”

- Instead of going from:

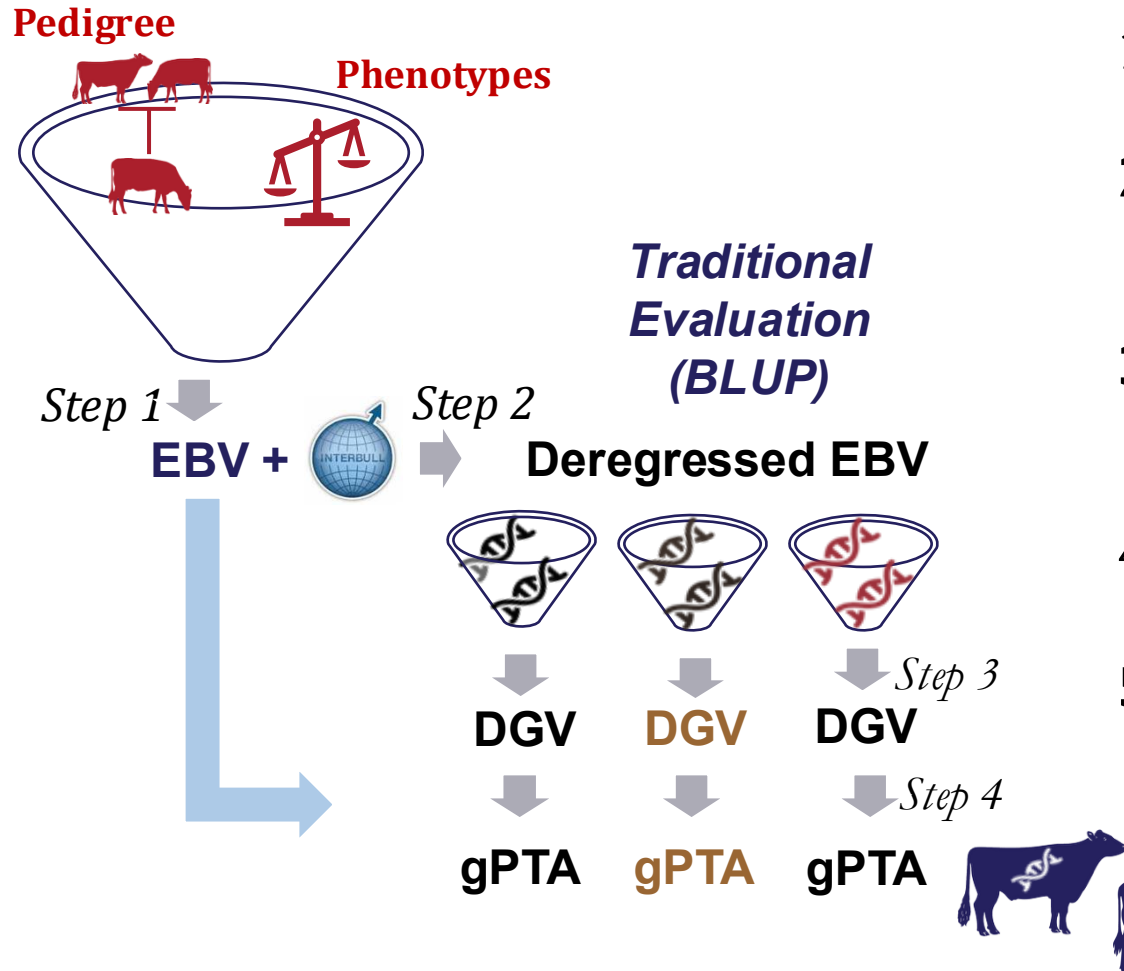


- We use traditional PTA to work backwards



- We mathematically estimate the reference animal’s underlying performance
 - WHY NOT USE ACTUAL PHENOTYPE OF GENOTYPED ANIMALS?
 - Psuedo-phenotype has much more info coming from traditional eval
 - Actual phenotype of one genotyped animal is often not the best summary of everything we know about that animal genetically (for ex. a bull doesn’t have a milk phenotype)

Legacy multi-step method

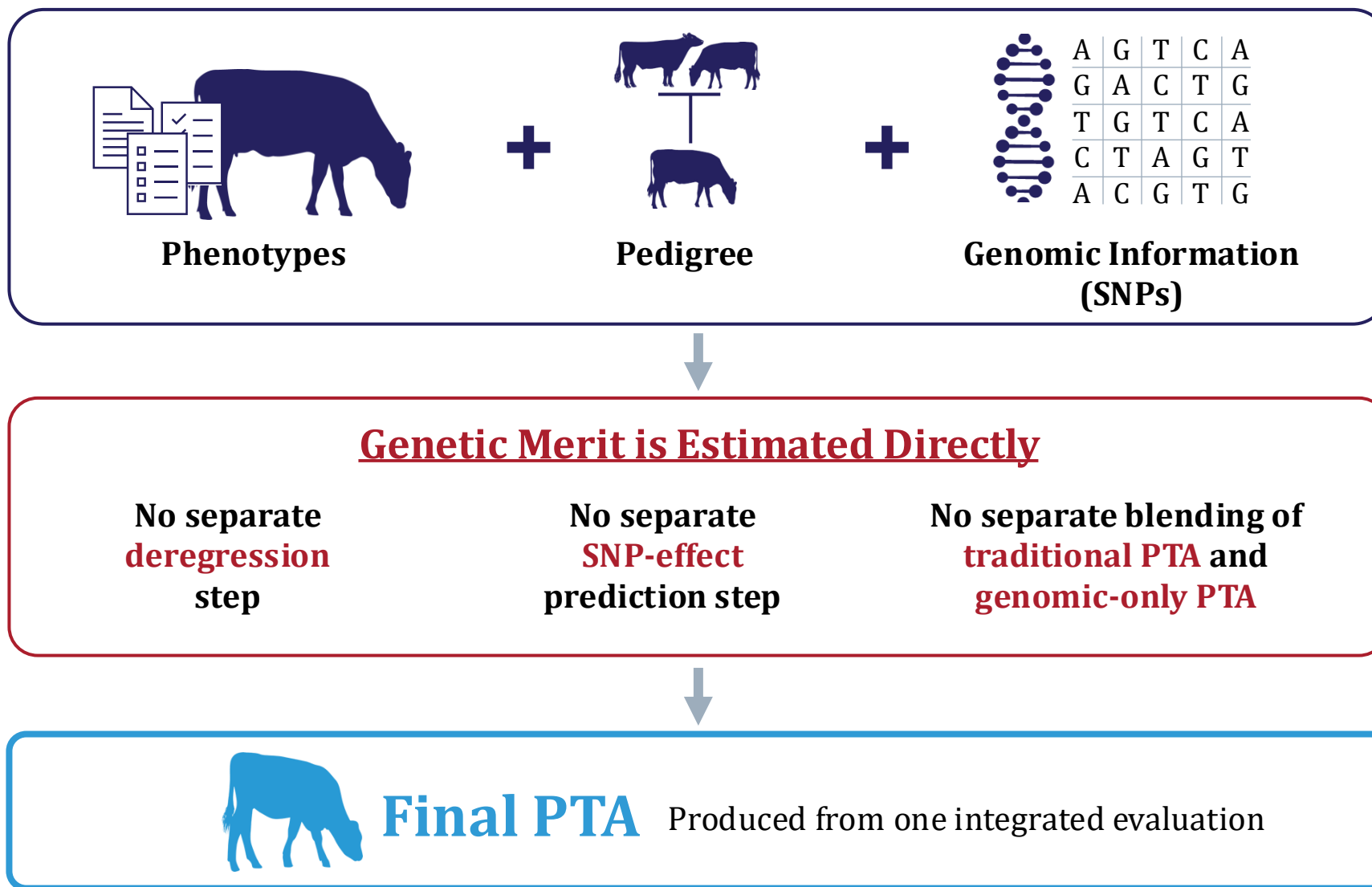


1. Phenotypes + Pedigree first
 - Traditional PTAs
2. Traditional evaluations deregressed
 - Pseudo-phenotypes for genomic prediction
3. Genotypes incorporated afterward
 - SNP effects estimated from genotyped animals with evaluation info
4. Genomic predictions are calculated
 - From those SNPs
5. Genomic and traditional information are blended
 - Final genomic PTAs

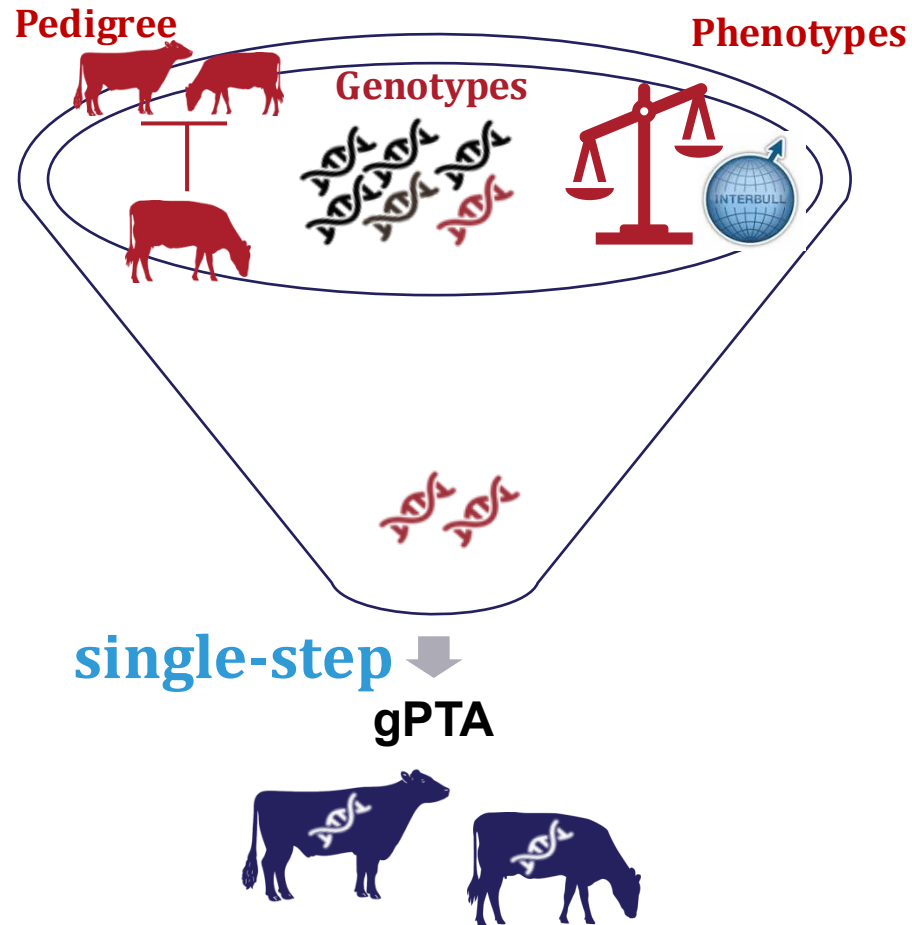
Why change? The scientific rationale

- **STEP 2:** the deregression process isn't perfect
 - Can introduce “*noise*” → variation in an evaluation that comes from the process, rather than from a real genetic difference between animals.
- Genetic trend estimates can be slightly biased due to genomic pre-selection of young bulls
- Accuracy for traits with fewer records and for smaller breeds can improve

Along comes Single-step...



The “future” single-step GBLUP



UGA Team



Why aren't we already using single-step?

- CDCB has much more data than other users, and the software needed to be adapted to accommodate that
 - Performance is substantially improved
 - Many bugs were fixed throughout the programs
 - Threshold models – used for calving traits – work much better now
- Changes to data preparation and data delivery systems were needed
- Reliability updates were needed for some special cases
- Methods for including Interbull data had to be developed
- Cloud tools had to be created and tested

The implementation philosophy

- Things that will remain the same after implementation of single-step:
 - Pedigree → still used
 - Phenotypes extraction
 - Usage of foreign information through Interbull MACE
 - Genotypes quality control and imputation
 - Models' factors and parameters (except for calving traits)
 - Trait-group independent single-step evaluations, all-breed, with final PTAs expressed within breed
- Using the software “engine” developed by University of Georgia

How genotypes are used in each method

Multi-step

- Computationally infeasible to use all 13M genotypes to estimate SNP effects
- Uses ~2.5M reference animals to estimate SNP effects
- Other ~11.5M genotyped animals get genomic estimate from estimated SNP effects
- Use genotypes within breed to estimate SNP effects

Single-step

- Computationally infeasible and unnecessary to use all 13M genotypes to measure more accurate relationships
 - Common ancestors = shared DNA
- Algorithm for Proven and Young (APY)
 - “CORE”: subset of the most informative genotyped animals used for genomic relationship matrix algebra
 - “NON-CORE”: remaining genotyped animals’ relationships handled recursively
- Use all genotypes together (not within breed)

Computational efficiency in the cloud

- Increases speed of computations → can scale up data/traits or frequency
 - Improved performance
 - **Enables monthly full evaluations**
(both phenotypes and genotypes updated)
- High security

What can we expect?

Transitioning to single-step

- Young genomic animals' trends will change
- Proven bulls with many daughters should not change
- Young genomic bulls should remain at the top with some reranking

After the transition

- Greater stability for young genomic bulls run-to-run
- Higher reliabilities for traits with few records and smaller breeds
- Unexpected cases may arise in this new system due to the scale and complexity of the system

What single-step doesn't do

- Magically fix all problems 
- Increase reliability of non-genotyped animals *“as if they were genotyped”*
 - Non-genotyped animals' information is more correctly allocated to their genotyped relatives
- Remove need for imputation to fill in missing SNP information
- Impact haplotype or genetic conditions

What's happening behind the scenes?

- All software and data pipelines are being reviewed
 - Programs are being rewritten to consolidate similar functions
 - Older tools and languages are being replaced with more modern ones
 - Processes are being comprehensively documented
- These changes will allow CDCB to move more quickly while ensuring robustness

Evaluation methods change over time

Daughter-dam comparison	1960	Contemporary comparison
Contemporary comparison	1989	Animal model BLUP
Animal model BLUP	2008	Multi-step GBLUP
Multi-step GBLUP	2027	Single-step GBLUP

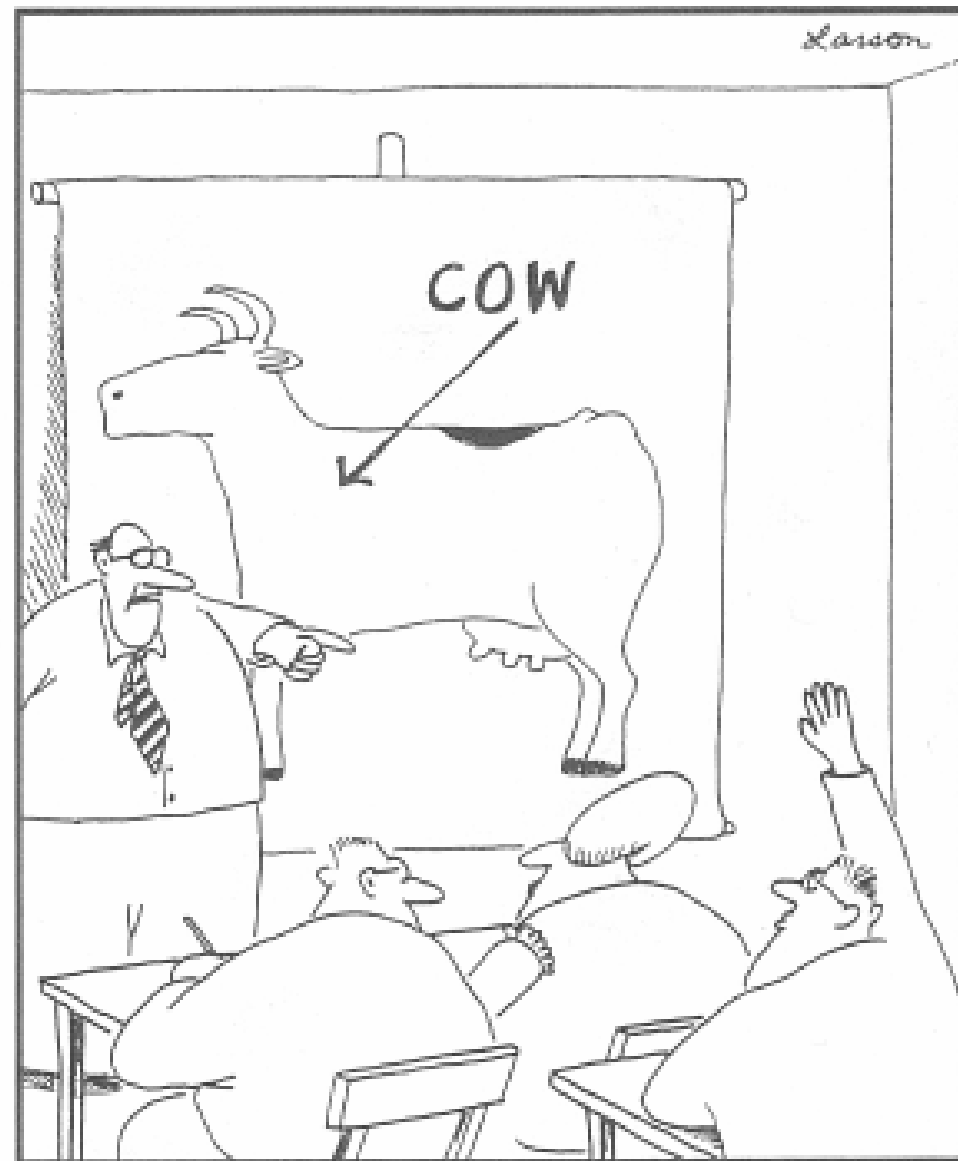
Closing thoughts

- Single-step provides a unified framework that integrates pedigree, performance, and genotype data
- It's likely to eliminate some sources of bias that are present in the current evaluations
- Single-step is a refinement of existing tools, not a revolutionary change

Questions?

Thank you for your attention!

john.cole@uscdcb.com



"Yes ... I believe there's a question there in the back."