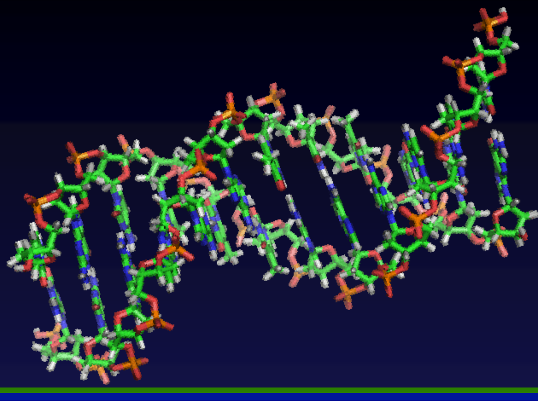
Walking the Cattle Continuum: Moving from the BovineSNP50 to Higher and Lower Density SNP Panels



C.P. Van Tassell[1]*, P.R. VanRaden[1], G.R. Wiggans[1], L.K. Matukumalli[2], S. Schroeder[1], J. O'Connell[1,3], R.D. Schnabel[4], J.F. Taylor[4], C.T. Lawley[5], D. Bailey[5], J. Downing[5], D. Lince[5], and T.S. Sonstegard[1].

[1]ARS, USDA, Beltsville, MD
[2]George Mason University, Manassas, VA,
[3]University of Maryland School of Medicine, Baltimore, MD,
[4]University of Missouri, Columbia, MO,
[5]Illumina, Inc., San Diego, CA

*curtvt@ars.usda.gov

Overview

- Introduction
- Genome selection
- Low-density SNP assay
    - Parentage
    - Relationships
- High-density SNP assay
- DNA sequencing

Traditional Selection Programs

- Estimate genetic merit for animals in a population

- Select superior animals as parents of future generations

Genetic Evaluations - Limitations

- Slow!
  - Progeny testing for production traits take 3 to 4 years from insemination
  - A bull will be at least 5 years old before his first evaluation is available
- Expensive!
  - Progeny testing costs $25,000 - $50,000 per bull
  - Only 1 in 8 to 10 bulls graduate from progeny test
  - At least $200,000 invested in each active bull!!

Bovine Genome

NATIONAL INSTITUTES OF HEALTH

National Institutes of Healt

Cowabunga!
Scientists to

NHGRI Approv

BETHESDA, Md., M
provisional go-ahe
widespread benefi

today gave its
to generate

exas have indicated
Medicine and Texas
tire project will cost
ney if the remaining
I Director Francis S.
r 2003 if the

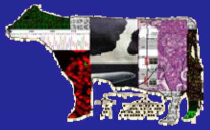
ick Perry announced

Cattle SNP Collaboration - iBMAC

- Develop 60,000 Bead Illumina iSelect® assay
  - USDA-ARS Beltsville Agricultural Research Center: Bovine Functional Genomics Laboratory and Animal Improvement Programs Laboratory
  - University of Missouri
  - University of Alberta
  - USDA-ARS US Meat Animal Research Center
- Starting 60,800 beads – expected 53,000 SNPs to result
- Planned to genotype ~30,000 animals for multiple projects

What is a Genomic PTA?

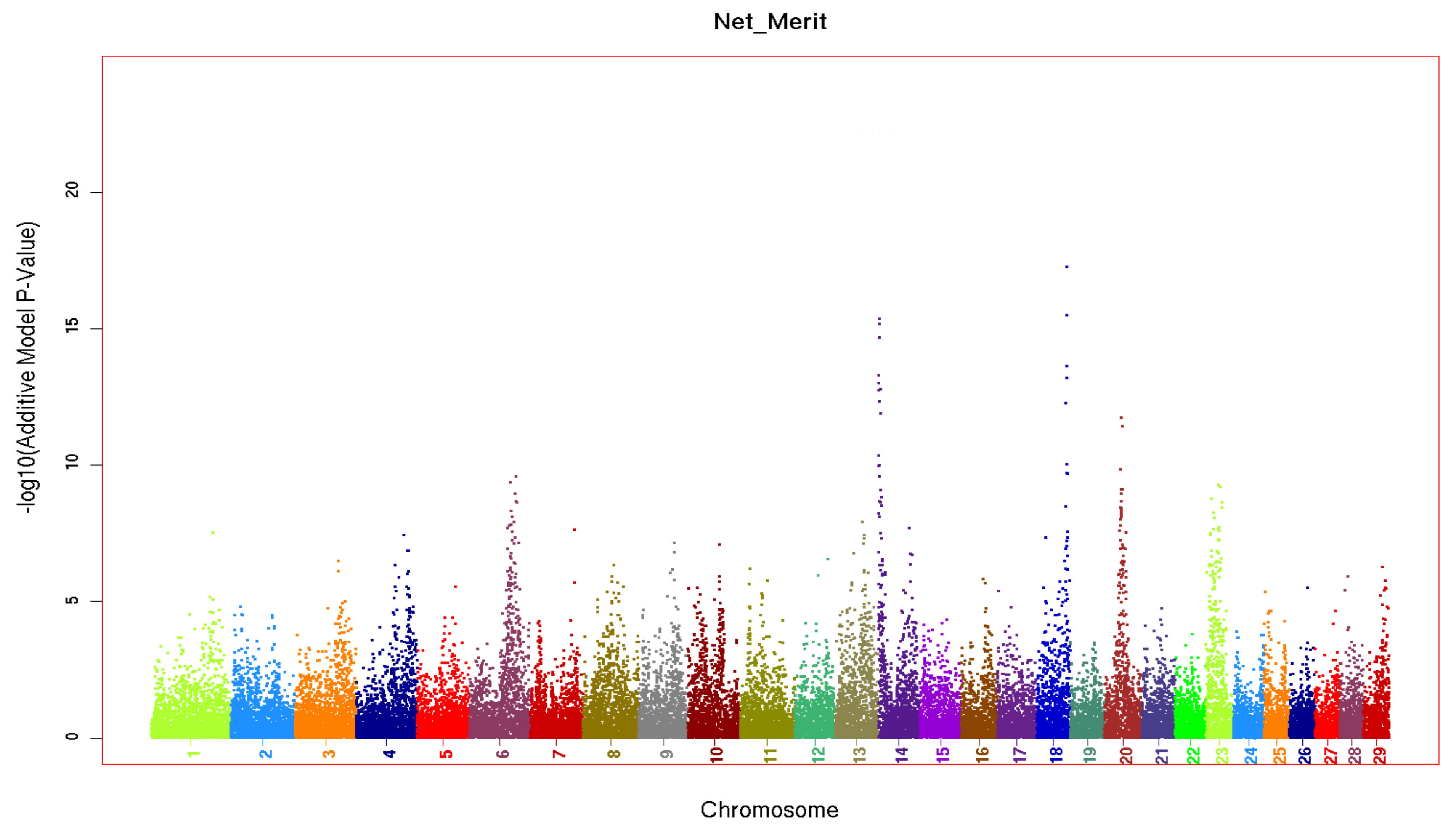
- “Train” system using phenotypic and genotypic data
    - Large regression system
- Approximately 40,000 genetic markers (SNPs) are evaluated
- For each SNP, the difference in predicted transmitting ability (PTA) between animals with 0, 1, or 2 copies of a specific allele is estimated
- Genomic evaluations combine SNP effect estimates with the existing parent average (PA) or PTA for each animal



Genome Selection

- This technology is revolutionizing dairy cattle breeding!
- Predict genetic merit at birth by combining pedigree merit and merit predicted from SNP
- Final genetic predictions are transparent to technology

Genomic Prediction



USDA-ARS Project

- Genotype 3,000 to 5,000 Holsteins
  - Built in validation - Test ability to predict "forward"
    - Historic group
      - 2000 bulls born in 1995-1997
      - 400 ancestor bulls born in 1950-1994
    - Prediction group
      - 800-1000 bulls born in 2001-2002
- Genotype 750 Jerseys
  - 1100 bulls
  - 200 cows

Cooperative Dairy DNA Repository

taurus-service inc.

Accelerated Genetics

ABS

SEMEX

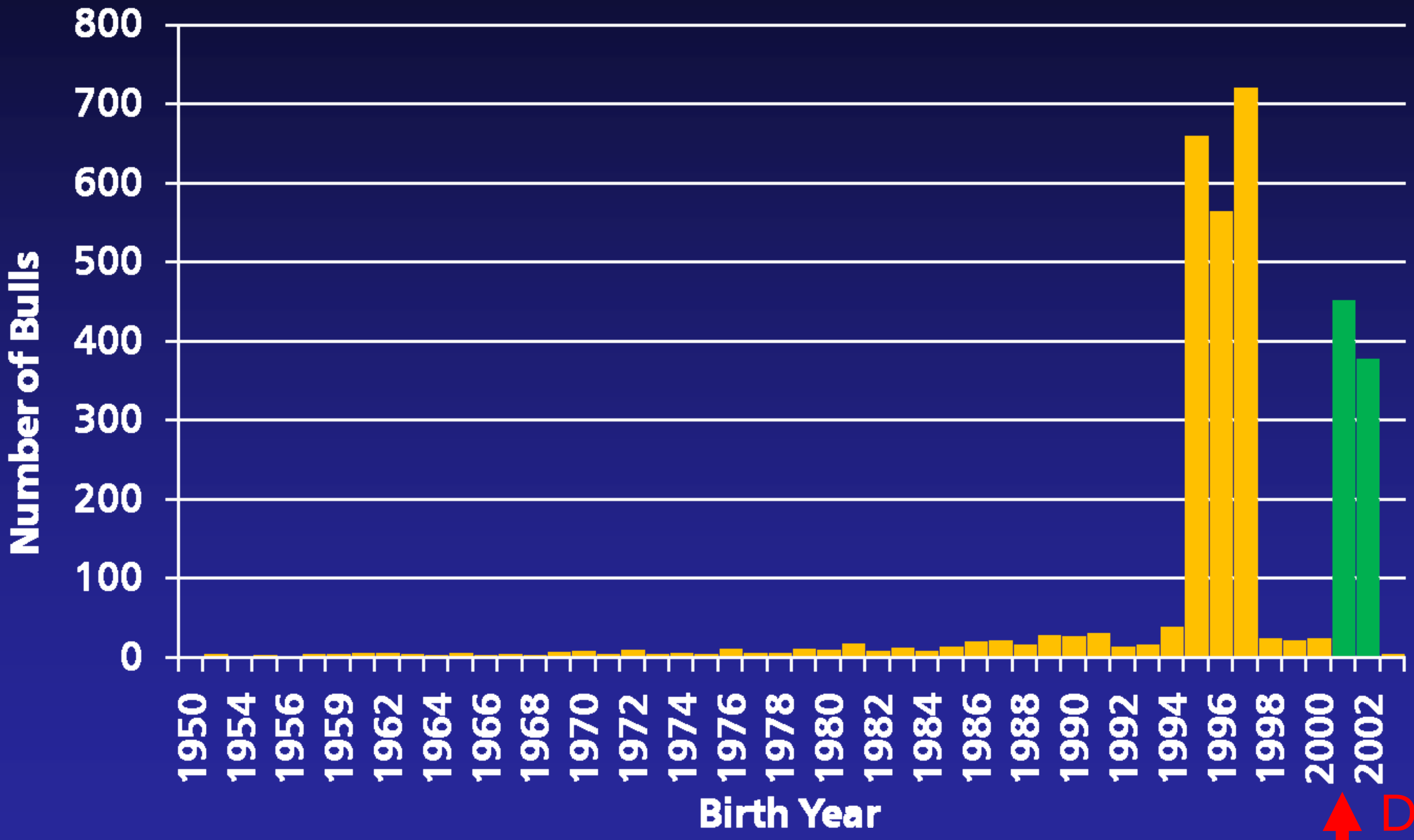
Alta Genetics

SELECT SIRES Inc.

Genex Cooperative, Inc.

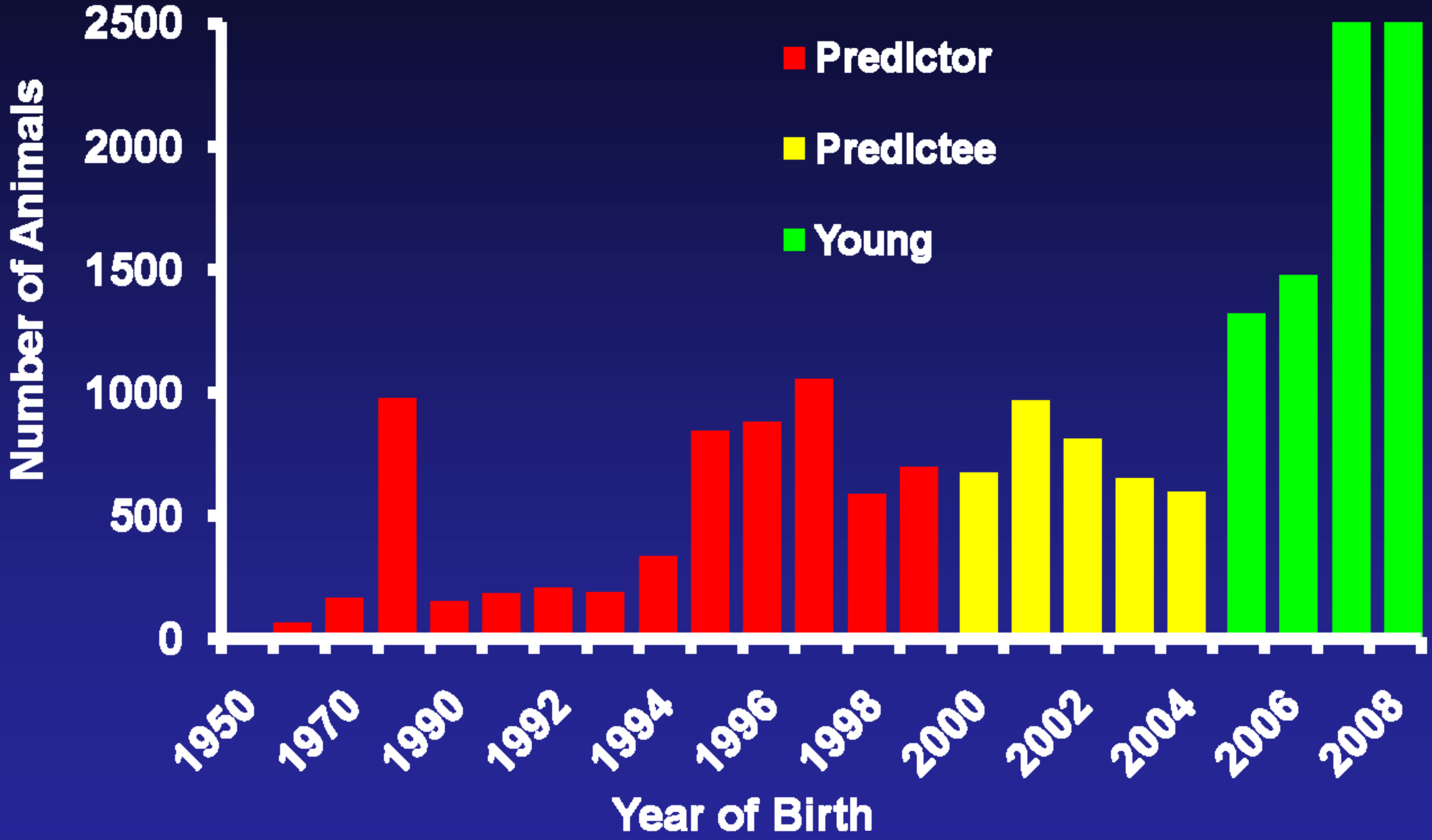
AgSource • Genex • Central Livestock CRI

Population Structure – Holstein



Data
Cutoff

Genotyped Animals (n=22,344*)



*In North America as of February 2009

Experimental Design - Update

Holstein, Jersey, and Brown Swiss breeds

	HOL	JER	BSW
Predictor:			
Bulls born <2000	4,422	1,149	225
Cows with data	947	212	
Total	5,369	1,361	225
Predicted:			
Bulls born >2000	2,035	388	118

Data from 2004 used to predict independent data from 2009

Reliability Gain[1] by Breed

Yield traits and NM$ of young bulls

Trait	HO	JE	BS
Net merit	24	8	3
Milk	26	6	0
Fat	32	11	5
Protein	24	2	1
Fat %	50	36	10
Protein %	38	29	5

[1] Gain above parent average reliability ~35%

Reliability Gain by Breed

Health and type traits of young bulls

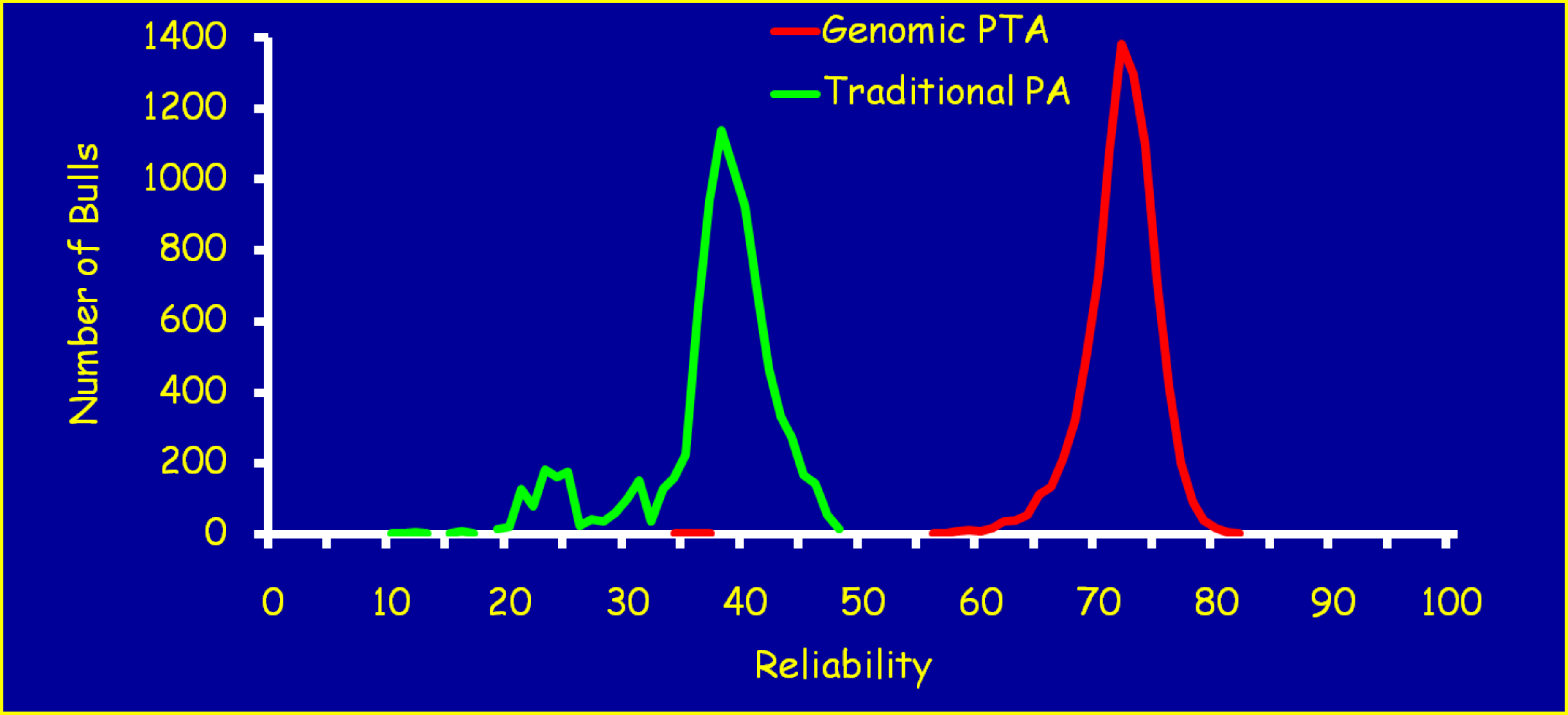
Trait	HO	JE	BS
Productive life	32	7	2
Somatic cell score	23	3	16
Dtr pregnancy rate	28	7	-
Final score	20	2	-
Udder depth	37	20	3
Foot angle	25	11	-
Trait average	29	13	N/A

Value of Genotyping More Animals

Actual and predicted gains for 27 traits and for Net Merit

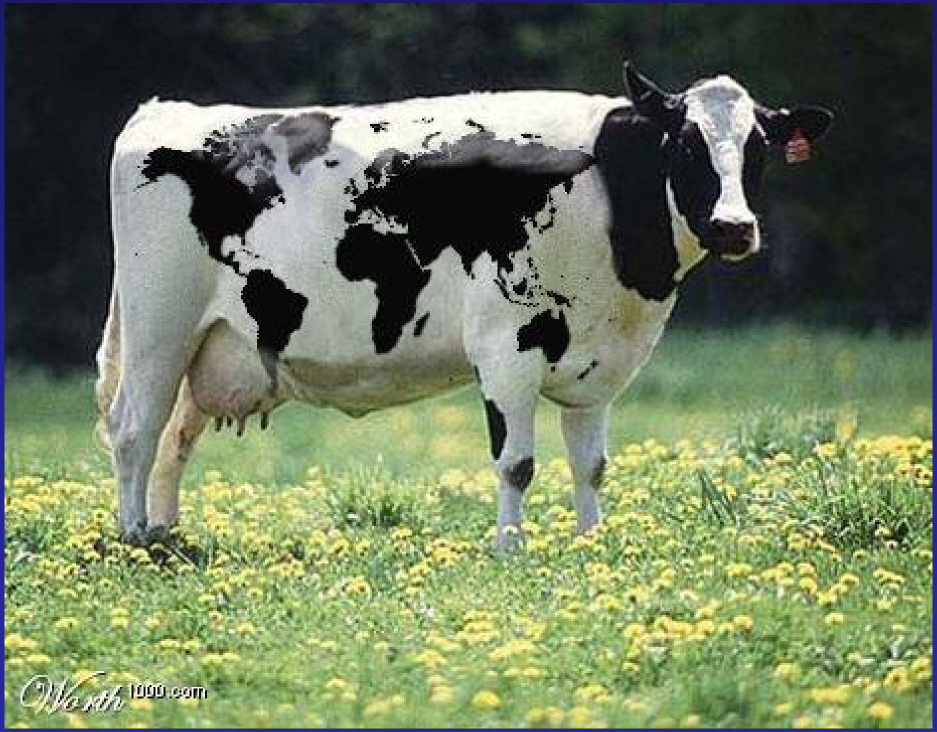
	Bulls		Reliability Gain	
Cows:	Predictor	Predicted	NM$	27 trait avg
	2130	261	13	17
	3576	1759	23	23
947	4422	2035	24	29
1916	6184	7330	31	30

Reliabilities for Young Bulls Parent Average vs. Genomic PTA



Now What?!

Where can we go next?



Low-Density Assay

- What?
  - 96, 384, 768, ….
- Why?
  - Parentage
    - 10 to 30% incorrect parentage
  - Traceability
    - Farm to fork
  - Genetic Prediction
    - Intermediate accuracy
  - Shortcut to pedigree data

Low-Density Assay

- What is low-density?
  - Today: 96, 384
  - Soon: 1,000-2,000
  - 1-2 years: 50K
- Density will depend on cost
- Technology is changing quickly

Low-Density Assay

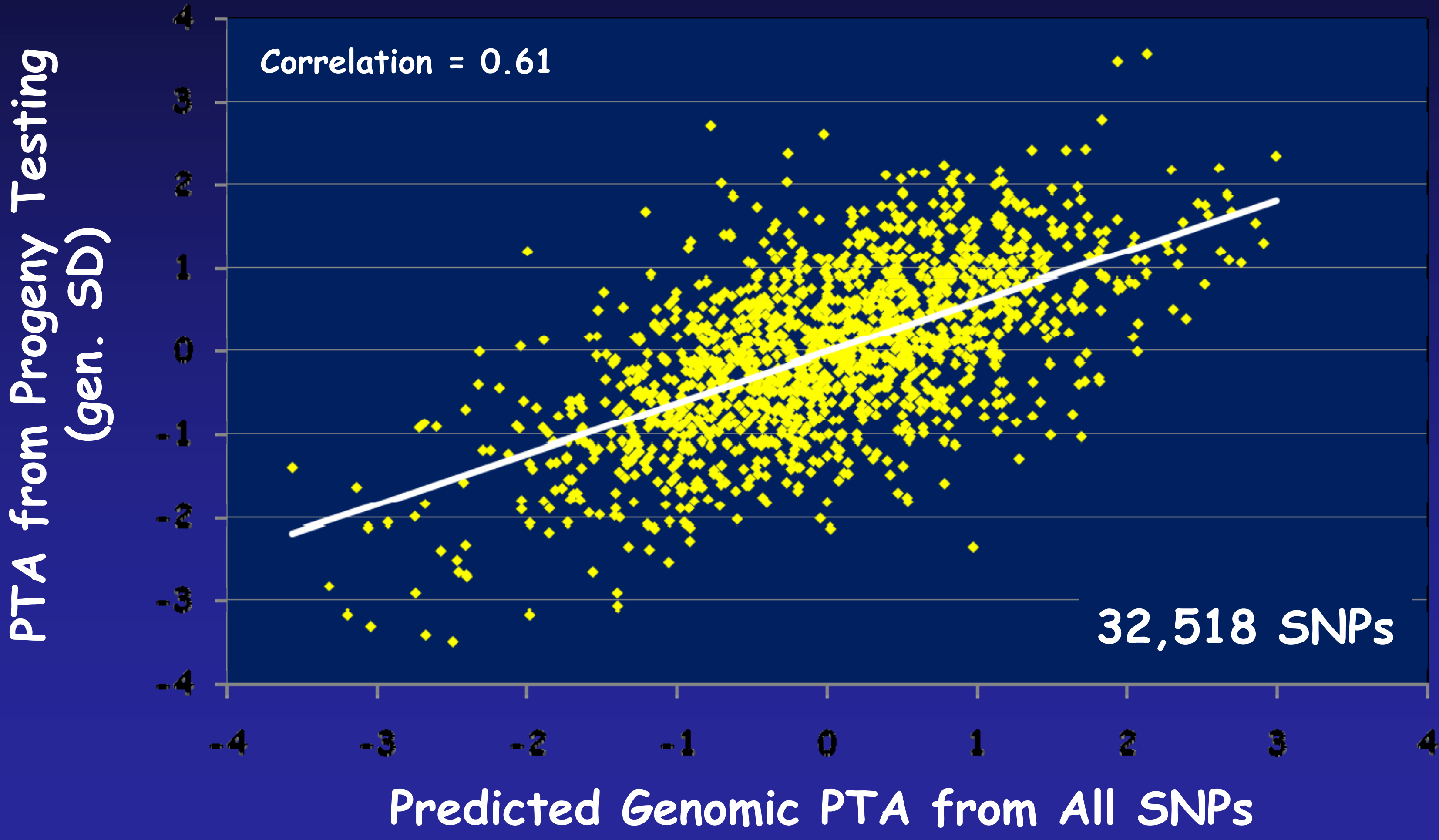
- Parentage and ID/traceability
  - 96 markers selected from BovineSNP50
  - Selected from SNP described by Heaton
  - Designed to be modular – incorporated into other assays
  - Hope to create *de facto* standard panel
    - Coordinated with ISAG and European research groups

Low-Density Assay

- Select 288 additional markers to predict Net Merit in Holsteins
- Accuracy intermediate between parent average and 50K
- Cost effective for commercial dairy cows
- Issues:
  - maximum predictive power
  - Multiple SNP (in LD) associated with a region

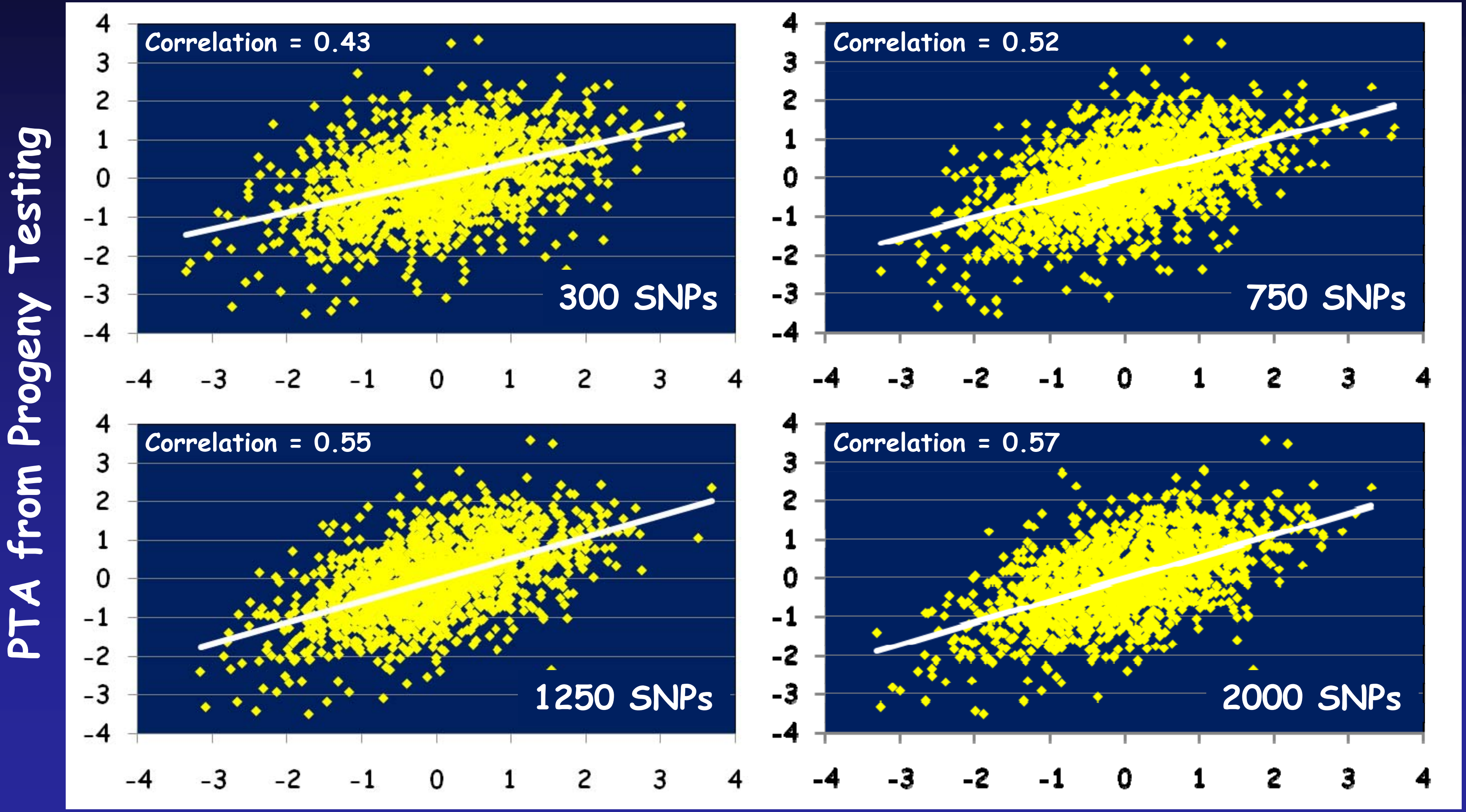
Predictive Ability for Net Merit

(Genomic PTA vs. Progeny Test PTA in Testing Set)



Predictive Ability for Net Merit

(Genomic PTA from SNPs vs. Progeny Test PTA in Testing Set)



Predicted Genomic PTA from Top ___ SNPs

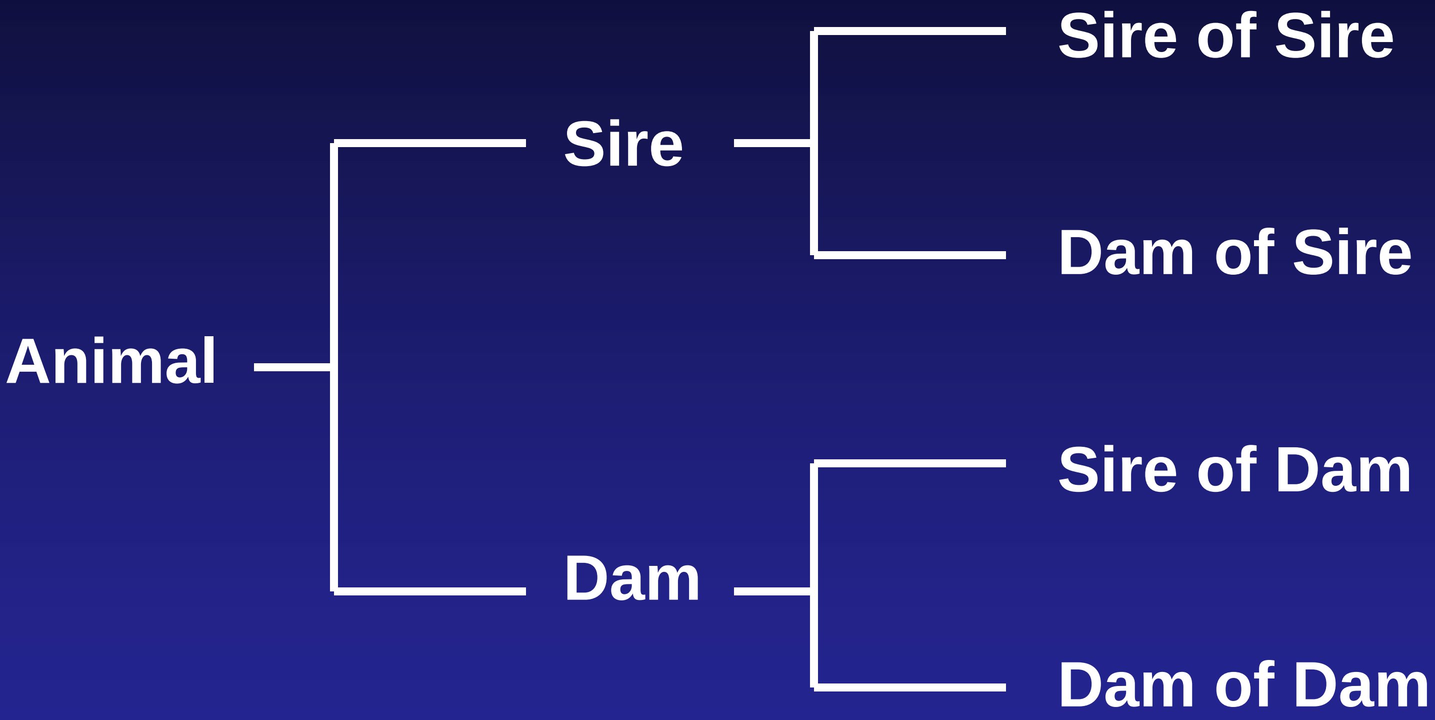
Low-Density Assay

- Enhance or replace pedigree
  - Low-budget genome selection
    - Account for Mendelian sampling from parents
  - Enable or improve genetic prediction where pedigree is unknown/incorrect
    - Developing world - Gates foundation
    - Extensive management conditions
    - Some dairy herds??

How Related are Relatives?

- Example: Full sibs
  - are expected to share 50% of their DNA on average
  - may actually share 45% or 55% of their DNA because each inherits a different mixture of chromosome segments from the two parents.
- Combine genotype and pedigree data to determine exact fractions

Traditional Pedigree



Animal

Sire

Dam

Sire of Sire

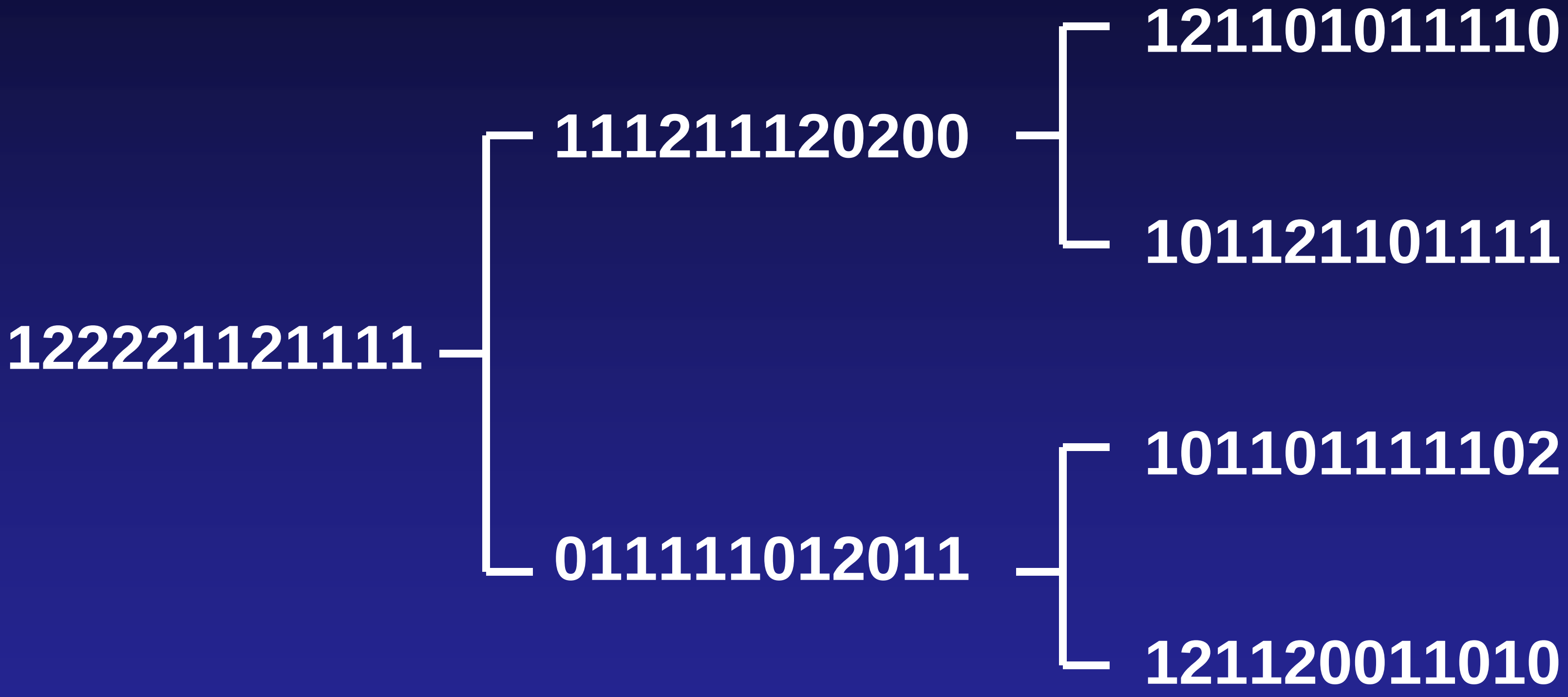
Dam of Sire

Sire of Dam

Dam of Dam

Genotype Pedigree

Count number of arbitrary allele



122221121111

111211120200

121101011110

101121101111

011111012011

101101111102

121120011010

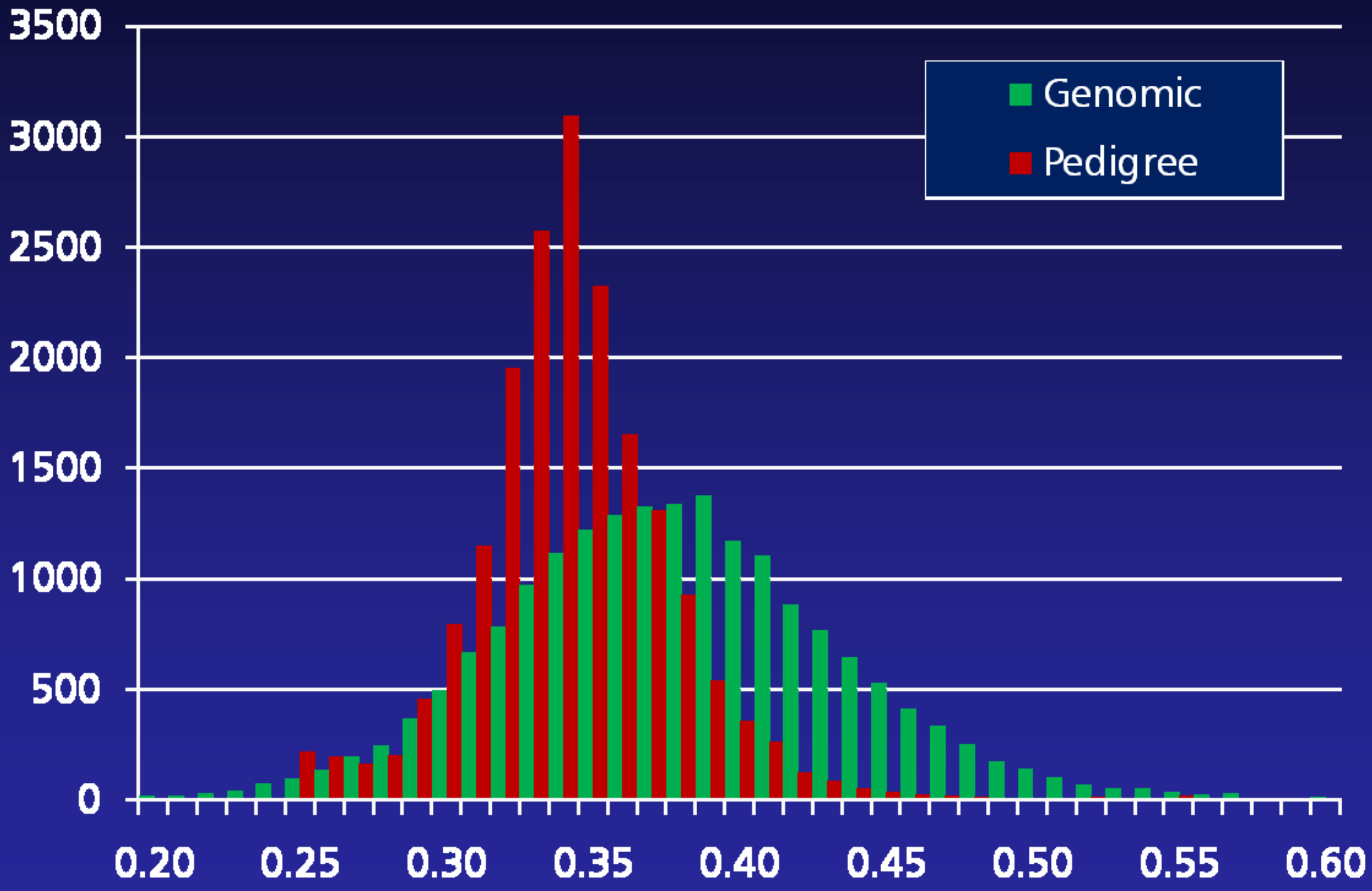
0 = homozygous for first allele
1 = heterozygous
2 = homozygous for second allele

Relationship Matrix

	DD	SD	DS	DD	Sire	Dam	Animal
DD	1.0				.5		.25
SD		1.0			.5		.25
DS			1.0			.5	.25
DD				1.0		.5	.25
Sire	.5	.5			1.0		.5
Dam			.5	.5		1.0	.5
Animal	.25	.25	.25	.25	.5	.5	1.0

Assumes grandparents are unrelated

Bull – MGS Relationships



Higher Density Assay

- What?
  - 500K +
- Why?
  - Across breed genome selection
    - At higher density, linkage disequilibrium may be sufficient for sharing haplotypes across breeds
  - Provide better information for Indicine cattle
  - Provide better within breed prediction of genetic merit?

Higher Density Assay

- Design issues
  - All of BovineSNP50 content
  - Taurine vs Indicine information content
    - Priority by breed?
  - Spacing
    - Uniform
    - Variable – more SNP in areas impact on important performance trait
  - Minor allele frequency
    - Want some rare SNP?

DNA Sequencing data

- Will be feasible to sequence AI bulls in 1-3 years
  - Harris Lewin sequenced a sire-son pair
    - Conservatively identified >600,000 SNP
- How do we use that data?
  - Paul VanRaden and Joel Weller proposed a strategy > 10 years ago
    - Manuscript rejected as being impractical and irrelevant!

DNA Sequence Data

- Haplotyping more practical with sequence data?
    - Rare SNP help identify phase
- Ultimate data are sequence derived from single molecule sequencing
    - Sequence each strand of DNA indivually and completely
    - On the horizon – e.g., Helicos

Summary

- SNP genotyping is transforming the dairy industry
  - Beef will follow… maybe… differently…
  - Other species to follow?
    - Lower density panels?
- Higher density should result in better (some?) ability to use data across breeds
- Sequencing of important animals is the ultimate endpoint

Funding

- USDA/NRI/CSREES
  - 2006-35616-16697
  - 2006-35205-16888
  - 2006-35205-16701
- USDA/ARS
  - 1265-31000-081D
  - 1265-31000-090D
  - 5438-31000-073D
- Merial
  - Stewart Bauck
- NAAB
  - Gordon Doak
  - ABS Global
  - Accelerated Genetics
  - Alta Genetics
  - CRI/Genex
  - Select Sires
  - Semex Alliance
  - Taurus Service

Teams…

- Missouri
  - Jerry Taylor
  - Bob Schnabel
  - Stephanie McKay
- USMARC
  - Tim Smith
  - Mark Allan
- Alberta
  - Steve Moore
- Illumina
  - Marylinn Munson
  - Cindy Lawley
  - Christian Haudenschild
  - Deb Bailey
  - Mike Thompson
- AIPL
  - Paul VanRaden
    - Mel Tooker
  - George Wiggans
  - Jeff O’Connell
  - John Cole
    - Dan Null
    - Jana Edwards
- BFGL
  - Curt Van Tassell
    - Steve Schroeder
  - Tad Sonstegard
    - Alicia Berteles
  - Lakshmi Matukumalli
- University of Wisconsin
  - Kent Weigel & Students