

Genotypes are Useful for More Than Genomic Evaluation

Paul VanRaden, Chuanyu Sun, Tabatha Cooper, Dan Null, and John Cole

Animal Improvement Programs Laboratory
Agricultural Research Service, USDA
Beltsville, MD

paul.vanraden@ars.usda.gov



Uses for genotypes

- Ancestor discovery within breeds
- Inheritance tracking for chromosomes
- Mating programs
 - Genomic inbreeding, dominance
- Fertility defects – haplotypes and QTLs
- Breed composition of crossbreds

Sources of genotypes used



Continent	Phenotyped		Young		Totals
	Female	Male	Female	Male	
N. America	93,345	23,598	329,780	76,786	523,509
Europe	0	12,218	36,055	17,753	66,026
Oceania	0	338	3,232	1,733	5,303
S. America	0	3	2,720	333	3,056
Asia	0	0	284	35	319
Africa	0	0	281	3	284
Totals	93,345	36,157	372,352	96,643	598,497



Service applies to many chips

- Different companies and densities
 - ▶ Illumina, GeneSeek, Zoetis, European LD
 - ▶ 3K, 7K, 8K, 10K, 50K, 77K, and 777K chips
 - ▶ Imputed (nongenotyped) dams of ≥ 4 progeny
- All animals imputed to 61,013 markers
- Faster service without imputing also possible

Ancestor discovery tests

- Database also stores **initial** pedigree status
 - ▶ Farmers correct pedigrees after DNA test
 - ▶ 1 week for corrections before predictions
- Sire initial status summarized (**actual**)
- Maternal grandsire (**actual and simulation**)
 - ▶ 5% of MGS set to incorrect and 5% missing

Sire initial status and discovery

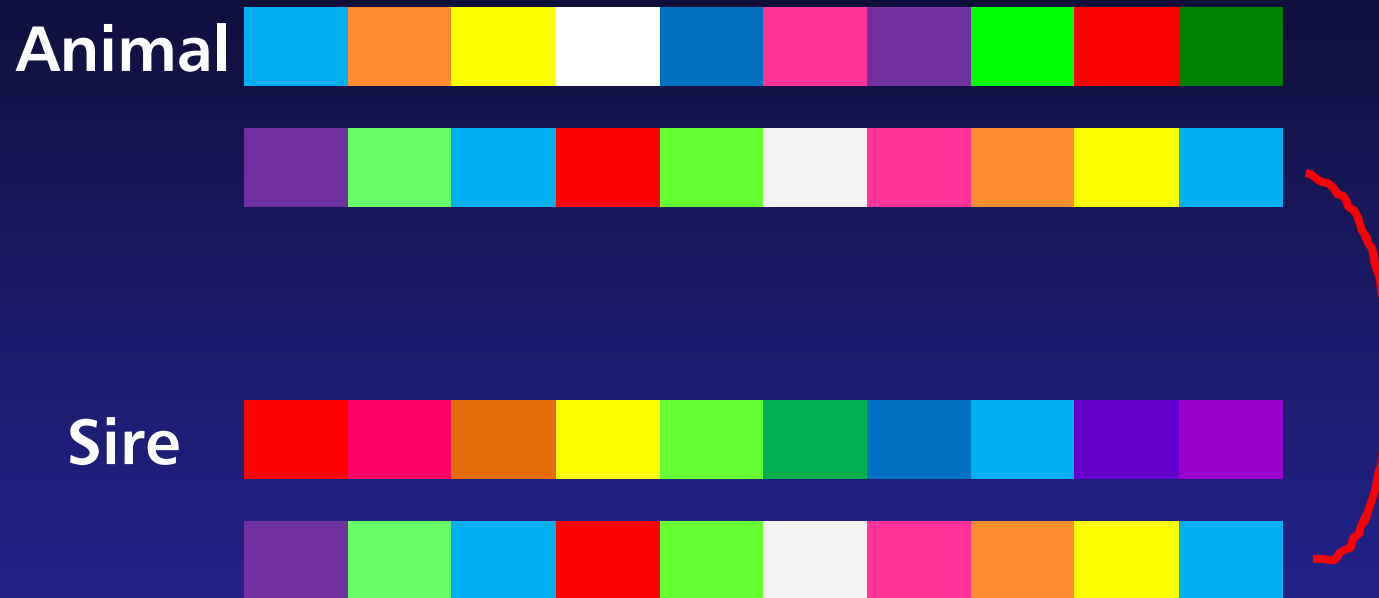
Sire status	Animals (N)		Animals (%)	
	Females	Males	Females	Males
Correct	195,770	37,416	68	85
Not genotyped	17,628	2,490	6	6
Incorrect	43,636	2,537	15	6
Missing	32,269	1,474	11	3
Total	289,390	43,931	100	100
Discovered	50,538	2,968	17	7



MGS status (after corrections)

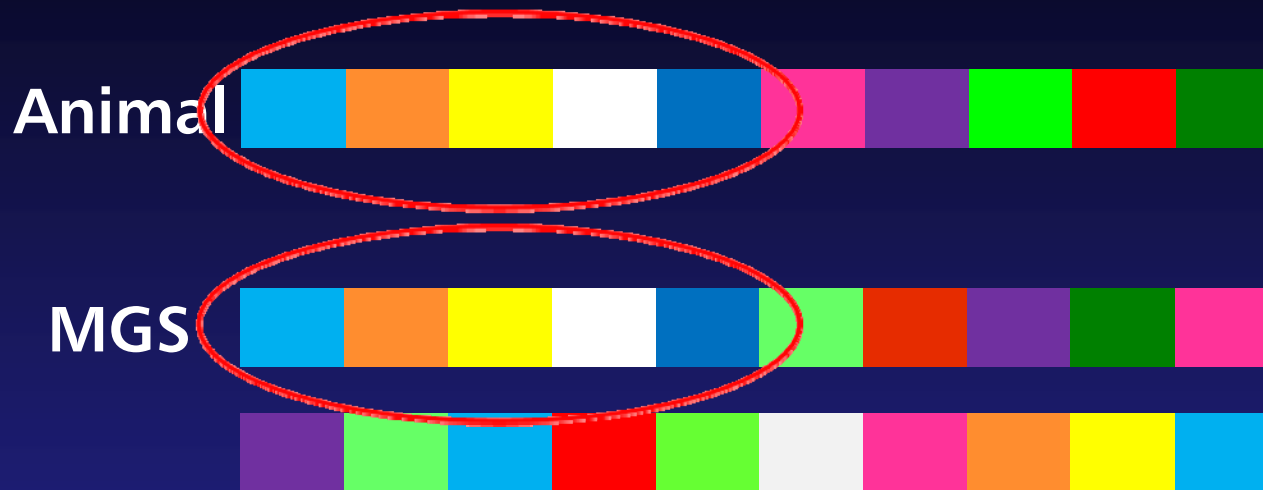
Sire status	Animals (N)		Animals (%)	
	Females	Males	Females	Males
Confirmed	141,963	38,270	49	87
Not genotyped or missing	119,160	3,874	41	9
Unlikely	28,267	1,787	10	4
Total	289,390	43,931	100	100

Separate paternal and maternal DNA



1 haplotype of animal matches 1 haplotype of sire
The animal's other haplotype must be from its dam

MGS discovery using haplotypes



Matches=5, haplotypes tested=10,
50% match vs. 45% expected (due to crossovers)
Thus, MGS is confirmed

Discovery of missing ancestors

Ancestor discovered (if genotyped)

	Sire	MGS	MGGS
Breed	% Correct*	% Correct	% Correct
Holstein	100	97	92
Jersey	100	95	95
Brown Swiss	100	97	85

* % Correct = Top ranked candidate ancestor matches the true ancestor.

MGS accuracy by chip

Chip	SNP Method		HAP Method	
	N	% Confirmed	N	% Confirmed
Bovine50K	3,620	97	3,197	98
Bovine3K	1,733	78	1,455	94
Imputed dams	--	--	106	92
BovineLD & GGP	7,690	96	--	--



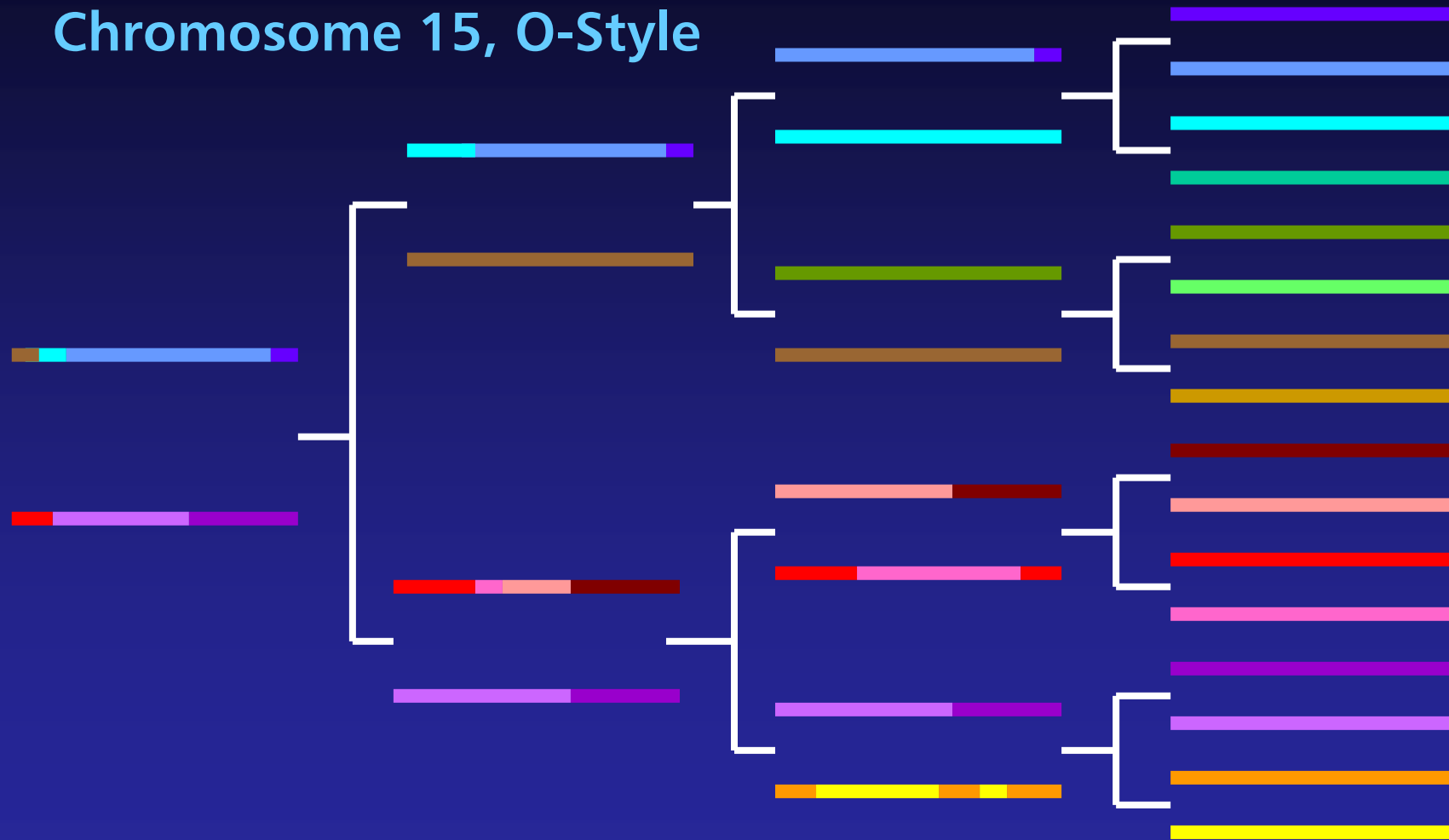
Genotyped ancestors, actual bull (HOUSA73431994)

			777K	777K
		50K	---	---
	50K		50K	50K
			50K	50K
		50K	777K	777K
			---	---
	50K		3K	777K
			50K	50K
		50K	777K	777K
			50K	Imputed
	50K		Imputed	50K
			50K	Imputed
		50K	50K	777K
			50K	50K
9K			50K	777K
			---	---
		50K	777K	777K
			Imputed	Imputed
	50K		50K	50K
			50K	777K
		50K	3K	777K
			50K	50K
		50K	50K	777K
			50K	50K
	50K		50K	777K
			777K	Imputed
		50K	777K	777K
			---	---
			Imputed	777K
			50K	50K

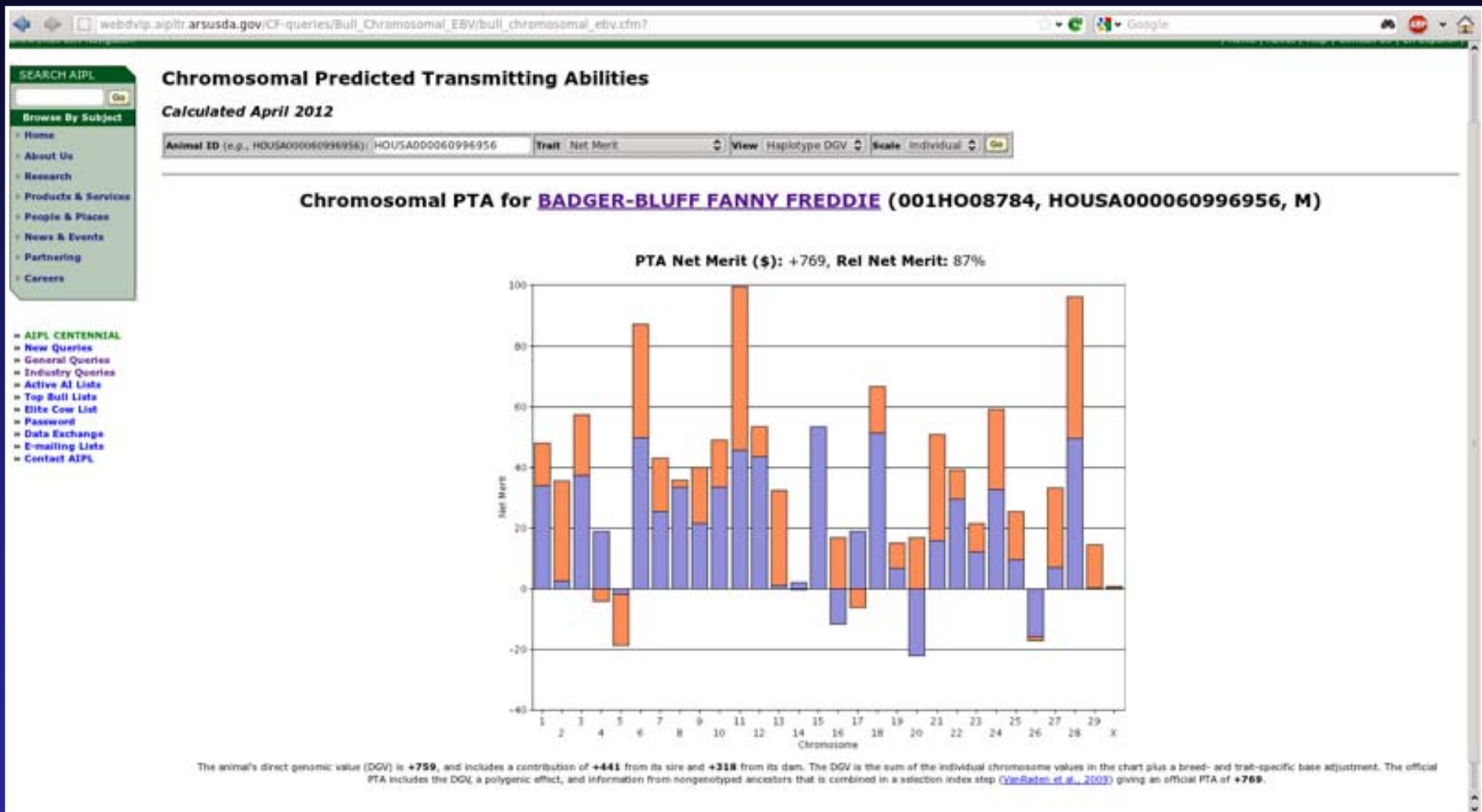


Haplotype pedigree

Chromosome 15, O-Style



Maternal / paternal haplotype values



https://www.cdcb.us/CF-queries/Bull_Chromosomal_EBV/bull_chromosomal_ebv.cfm

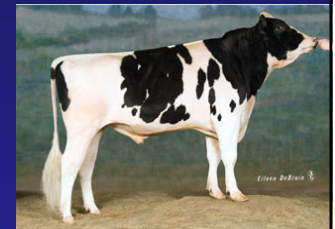
ICAR / Interbull annual meeting, Berlin, Germany May 20, 2014 (14)

Paul VanRaden



Pedigree or genomic mating?

- Computer mating programs have helped breeders identify potential mates with fewer **ancestors** in common to reduce **pedigree inbreeding**
- Such programs could instead help breeders to identify potential mates with fewer **alleles** in common to reduce **genomic inbreeding**
- This works best if both mates are genotyped



Pedigree
relationships



Genomic
relationships

Genomic mating programs

- Develop rapid methods to deliver genomic relationships (**G**) from central database to industry
 - ▶ Compute **G** for females with only marketed bulls
 - ▶ Compute **G** elements as needed by query
 - ▶ Compute **G** once, then retrieve elements as needed
- Compare methods to assign mates
 - ▶ Minimize pedigree or genomic relationships
 - Linear programming (LP)
 - Sequentially choose least-related mates ([Pryce et al., 2012](#))
 - Random mating
 - ▶ Include or exclude dominance effects of markers

Mating programs with dominance

- With dominant genes, progeny merit may not equal the average of parents' merit
- Predicting dominance effects was difficult from pedigrees, but is easier with genomics
- Dominance variance is smaller than additive
 - 4% dominance vs. 25% additive for yield
 - 1% dominance vs. 9% additive for SCS

Mating program results

- With linear programming, genomic inbreeding was:
 - ▶ 3% lower than with random mating
 - ▶ 1% lower than with sequential mate selection
- Genomic instead of pedigree relationships:
 - ▶ Added value was $\$32 * 184,693 \text{ calves} = \text{\$5.9 million / year}$ for Holstein females genotyped in 2013
- Extra benefits from predicting dominance were small
- Developed mating software is ready for service

Genomic mating computation

Times and disk storage required to compute **G** for all animals or recalculate elements of **G** as needed

Breed	G for cows and all proven bulls		G for cows and only marketed bulls		
	Time (h:min:s)	Disk Storage Gbytes	Animals (no.)	Computing time	
				Extraction (h:min:s)	Recalcu- lation (s)
Holstein	16:22:42	426	1,817	1:58:06	31
Jersey	00:17:11	7	585	0:01:46	6
Brown Swiss	00:00:13	0.03	338	0:00:01	4



Fertility and stillbirth defects

- Track new defects by haplotype or gene test
 - ▶ Holstein HH1, HH2, HH3, HH4 (Sebastien), HH5
 - ▶ Brown Swiss BH1, BH2 (Schwarzenbacher)
 - ▶ Jersey JH1, Fertility1 (LIC)
 - ▶ Montbeliarde MH1 and MH2 (Sebastien)
 - ▶ Ayrshire AH1
- Track previous defects by haplotype or test
 - ▶ BLAD, Brachyspina, CVM, DUMPS, Mulefoot, SDM, SMA, Weaver, etc.
 - ▶ Not included on early chips, can impute from markers

Economics of fertility defect **HH1**

- Pawnee Farm Arlinda Chief (born 1962)
 - ▶ Contributed 14% of global Holstein genes
 - ▶ \$25 billion value of increased milk yield
 - ▶ \$0.4 billion cost of **HH1** mid-term abortions
- How many more fertility defects are there?
 - ▶ Average 0.2 / animal based on inbreeding depression (VanRaden and Miller, 2006 JDS)

Haplotype tests, then lab tests

Genotypes	Frequency		Lab tests ¹	
	JH1	HH1	JH1	HH1
Normal	76.5	97.2	9,867	113,792
Carrier	21.3	2.4	2,750	2,793
Homozygous	0.0	0.0	0	0
No call	2.1	0.4	276	464
Total	100.0	100.0	12,893	117,049

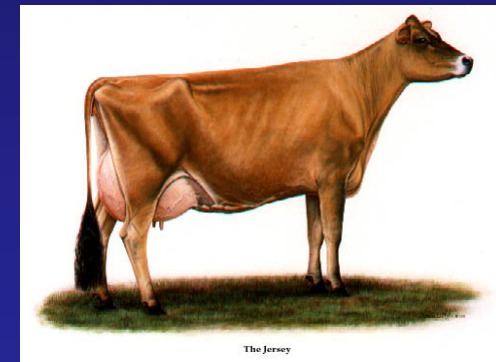
¹Data from the Geneseek Genomic Profiler (GGP) and GGP-HD for causative mutations (JH1 = CWC15 and HH1 = APAF1)

Jay Lush, 1948 (The Genetics of Populations)

- “The rapid rate at which genes have been found in each species, whenever people started to study it genetically, and the fact that in most of these species the rate of finding new genes actually seemed to increase with continued study until so many genes were known that interest in keeping stocks of each new one waned.” (p. 32)

Predict breed and breed composition

- Used purebred Holsteins, Jerseys, and Brown Swiss genotypes to develop equations
- Predict breed fractions for crossbred animals



Predict breed composition as a 'trait'

- Y- variable was breed of animal
 - ▶ A Holstein would receive a 1 in the Holstein analysis and a 0 in the Jersey and Brown Swiss analyses

Animal Breed	Holstein Analysis	Jersey Analysis	BSW Analysis
HOL	1	0	0
JER	0	1	0
BSW	0	0	1

Number of SNP to predict breed

- 3 different SNP sets were used for genomic prediction of breed composition (Bayes A)
 - ▶ The full 43,385 SNP set
 - ▶ A reduced 3K SNP set
 - ▶ The original 600 breed check SNPs
 - Each breed (HOL, JER, BSW) has ~200 SNP used for a quick check (not a genomic prediction)
- Included 22,679 males and 6,480 females



Accuracy of breed prediction

Means and standard deviations for predicting breed percentages for young, validation animals

Markers:	43 K	3 K	600
Breed:			
Holstein N = 14,794	100.0 ± 0.8	100.4 ± 3.1	100.2 ± 1.9
Jersey N = 919	99.6 ± 2.8	97.8 ± 6.3	98.9 ± 3.6
Brown Swiss N = 96	99.4 ± 2.1	98.9 ± 3.6	99.2 ± 5.1

Breed prediction example (crossbred)

- Animal pedigree = **87.5% HOL, 12.5% JER**
 - ▶ 43K prediction = **85.9% HOL, 13.3% JER**
 - ▶ 3K prediction = **84.4% HOL, 15.5% JER**
 - ▶ 600 SNP predict = **83.0% HOL, 16.6% JER**
- Accuracy is lower for very old or foreign animals with unusual pedigrees
- Genotypes for each pure breed are needed

Manage differently by genotype?

- Known as personalized medicine for humans
- Different management is costly in livestock
 - ▶ Early culling instead of veterinary treatment
 - ▶ Total mixed ration replaced individual rations
 - ▶ Several breeds and crossbreds grouped together
- Estimate herd management effects more accurately after subtracting genomic effects

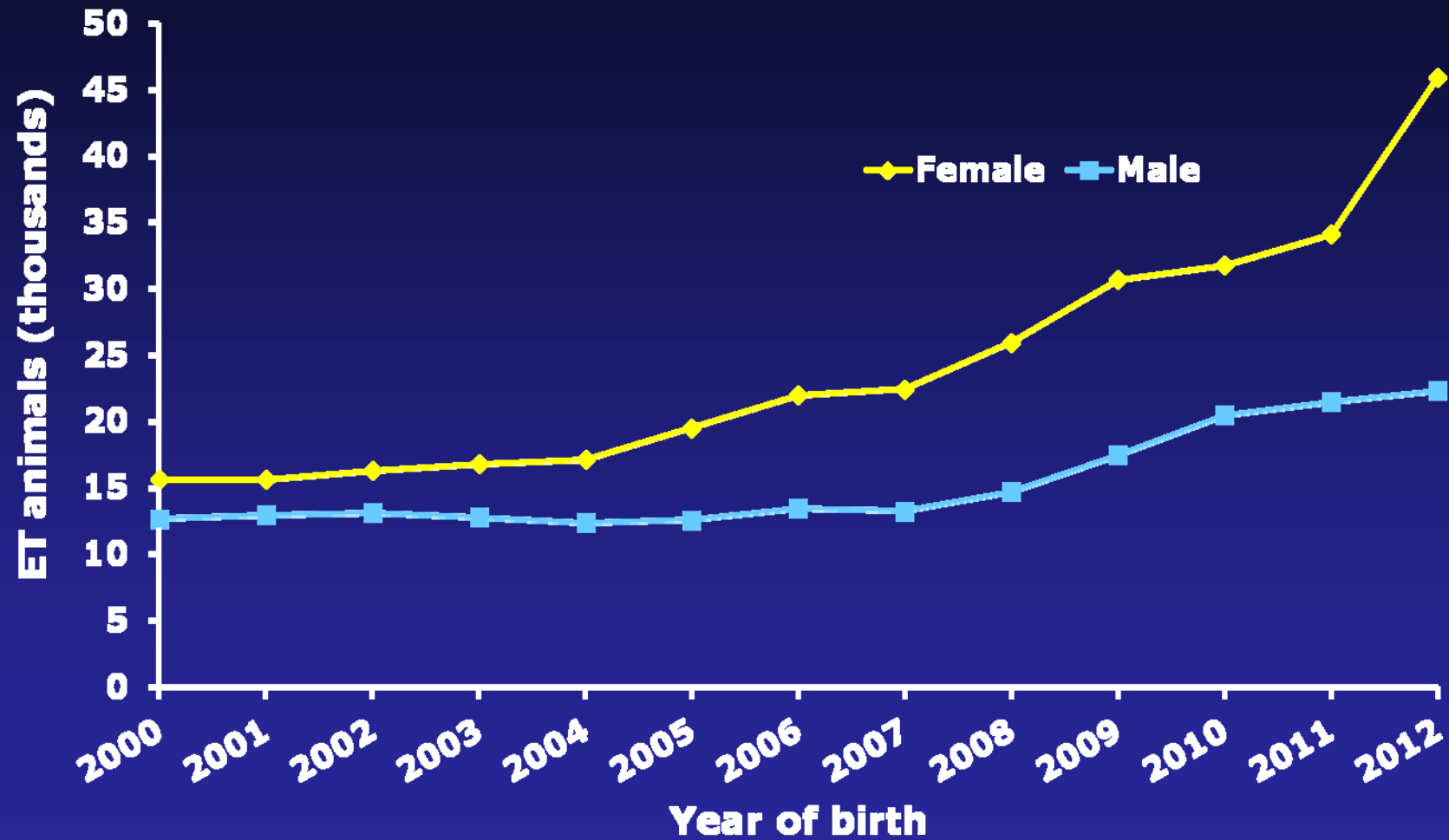
Top young bulls from April 2010



Bull	Net Merit			Daughters	
	2014	2010	PA 2010	2014	2010
Observer	646	848	552	2101	0
Robust	834	821	522	511	0
Twist	800	817	491	337	0
Edward	584	789	532	184	0
Erdman	821	778	529	349	0
Networth	619	771	566	218	0
Bookem	699	761	575	1482	0
Mauser	656	759	464	288	0
Top 8 Avg	707	793	529	684	0



Embryo transfer calves, by year



Conclusions

- 75,905 females had missing or incorrect sires; a true sire was suggested for 50,538 (67%)
- MGS and great grandsires can be discovered; pedigree corrections are more difficult
- Genomic mating programs will be profitable
- New defects are easy to discover and track
- Genotypes also useful for genomic evaluation

Acknowledgments

- Katie Olson, Mel Tooker, Jan Wright, Lillian Bacheller, George Wiggans, and Jeff O'Connell contributed to the computation and graphics
- Members of the Council on Dairy Cattle Breeding (CDCB) provided the data



Tear down this wall

