How to implement genomic selection

abstract #66

Paul VanRaden

USDA, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD

Email: paul.vanraden@ars.usda.gov

Project web site: https://aipl.arsusda.gov/

Genomic selection requires:

- **DNA from many animals**

- **Genotypes for many variants per animal**

- **Phenotypes from many animals**

- **Quality control of the data**

- **Statistical methods to convert data into predictions**

- **Breeders convinced that the predictions will work**

- **Investment to continue collecting data**

USDA

Progress over the last decade

- **Progress in stability and reliability of predictions**

- **Expansion in number of variants tested**

- **Additional traits tested and new uses for genotyping**

- **Growth in numbers genotyped worldwide and in other species**

- **Genetic progress**

USDA

Sources of DNA for selected years

Source	2008	2010	2018	2008	2010	2018
	%	%	%	N	N	N
Tissue (ear)	0	<1	81	0	100	579,255
Hair	0	82	14	0	22,550	102,229
Blood	10	5	2	623	1,375	14,656
Embryos	0	0	1	0	0	6,188
Nasal swab	0	12	<1	0	3,300	656
Semen	90	<1	<1	5,285	200	319
Unknown	0	0	2	0	0	17,048

USDA

Potential for embryo genotyping

- **15,000** embryos already genotyped

- Genomic predictions sent only to breeder via nominator

- Change selection strategies if technology improves and costs decline
  - Genotype before instead of after pregnancy
  - Demand for embryo genotyping could be more than for calves
  - Could genotype **20 million** embryos/year instead of **4 million** calves

- May be difficult to automate **... requires switching to ET instead of AI**

USDA

Worldwide dairy genotyping (January 2009)

Countries	Animals
United States and Canada	22,344
France	8,500
Netherlands and New Zealand*	6,000
New Zealand and Ireland	4,500
Germany	3,000
Australia	2,000
Denmark, Finland, and Sweden	2,000

*Used customized Illumina 50K chip
 (different markers)

Source:
"Dairy Cattle Breeders Have Adopted Genomic Selection" (Gordon Conference on Quantitative Genetics and Genomics, February 2009)

USDA

Worldwide livestock genotyping (January 2019)

Species	Countries/Companies	Animals
Dairy	United States and Canada	3,020,000
	France	550,000
	Netherlands	465,000
	New Zealand	140,000
	Germany	785,000
Beef (Angus)	United States	550,000
Beef and dairy	Ireland	1,500,000
Poultry	Aviagen	1,000,000/year?

USDA

CDCB usable genotype counts/year by animal sex
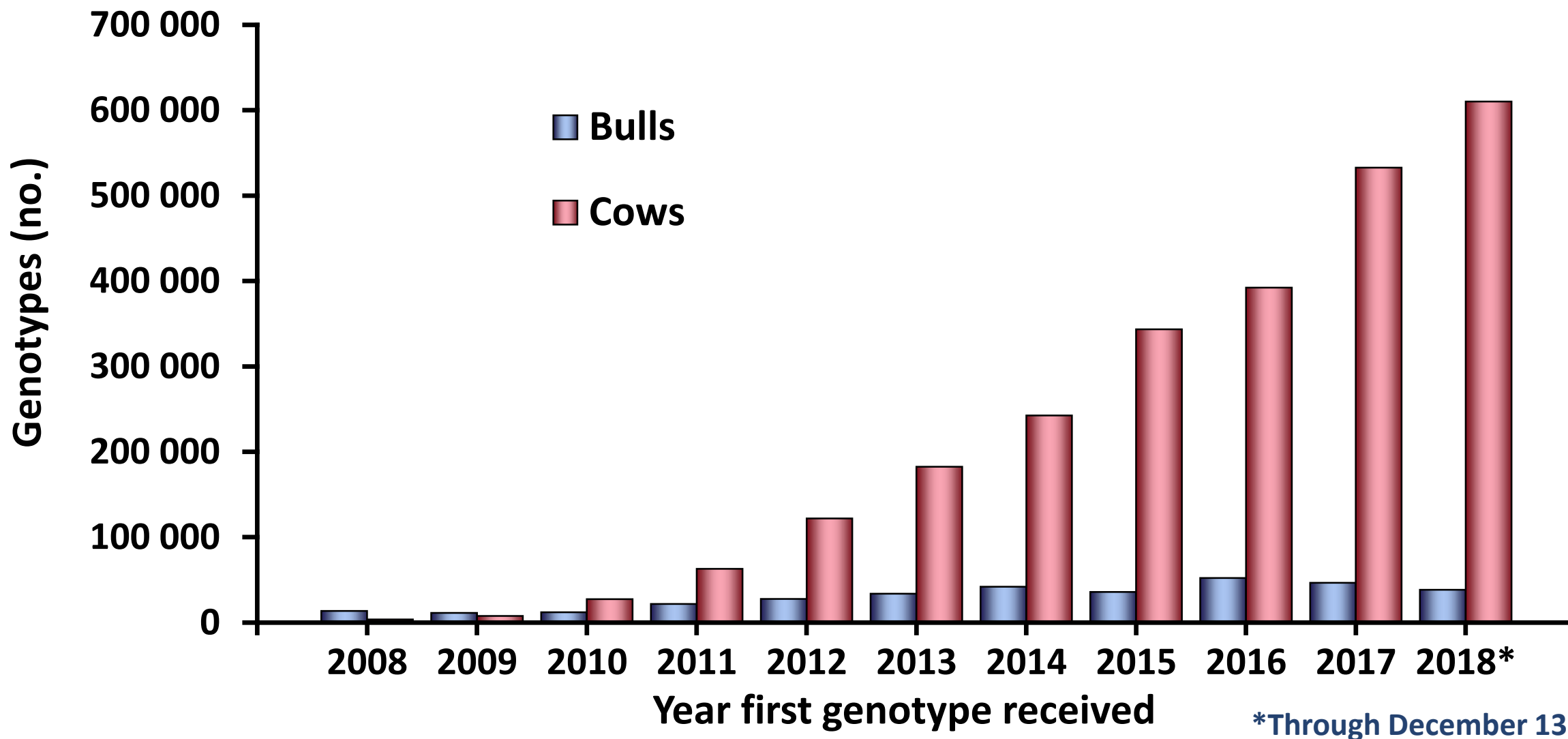
Formal international genotype exchanges

| Country exchanges with U.S.	Breed	Sex	Year
Canada	All	All	2008
United Kingdom, Italy	Holstein	Bulls	2011
All Europe via Interbull	Brown Swiss	Bulls	2012
Denmark (Viking)	Jersey	Bulls	2014
Japan, Switzerland, Germany	Holstein	Bulls	2016
United Kingdom	All	All	2017

Genotyped animals in database by region (2019)

2,585,255

200,002

7,749

23,408

27,356

1,943

18,513

USDA
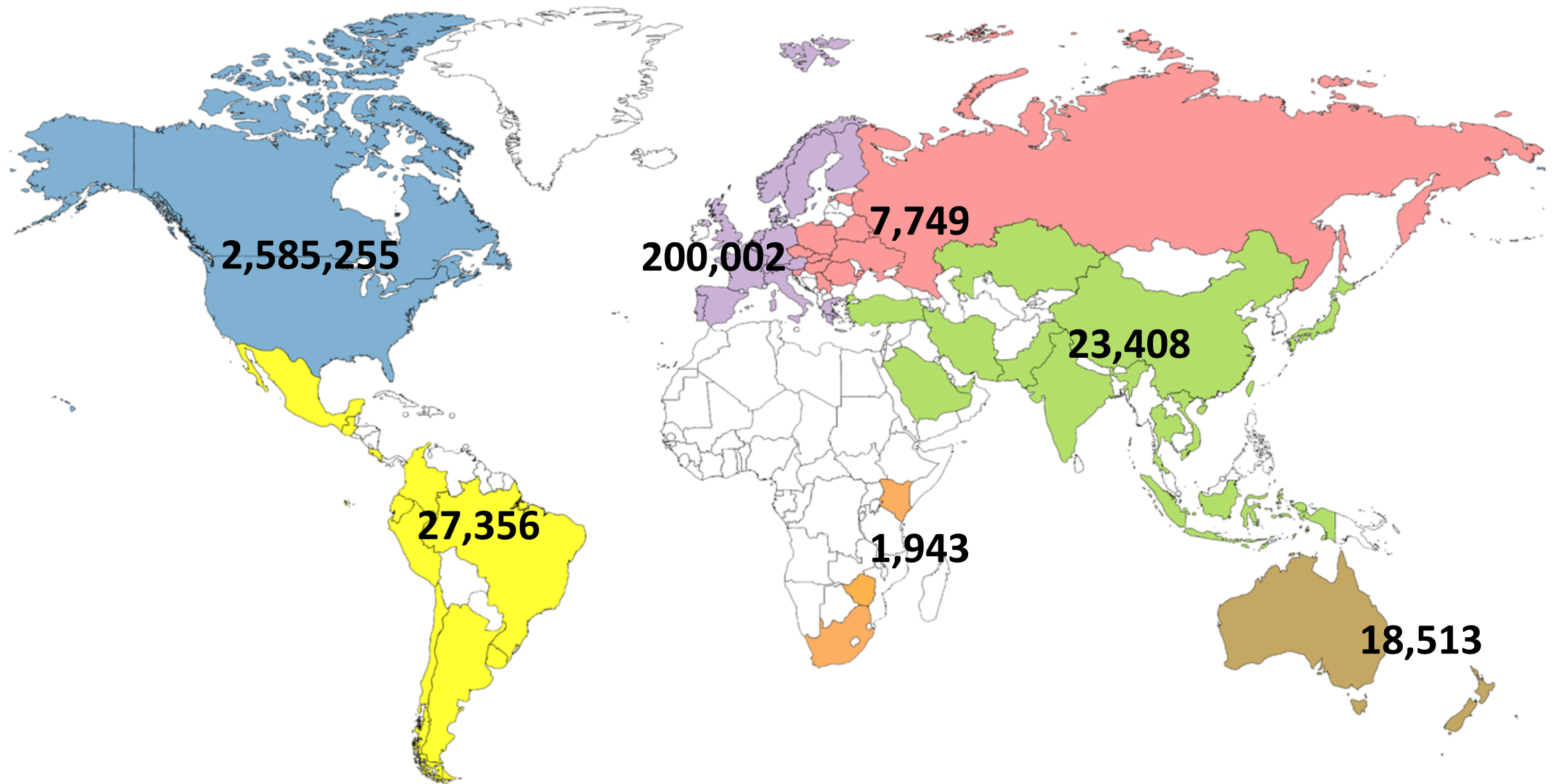
Top 10 NM$ genotyped Holsteins by country

Country	NM$	Country	NM$	Country	NM$	Country	NM$
USA	1114	JPN	854	CHE	763	NZL	595
CAN	1040	HUN	850	ZAF	724	RUS	555
GBR	1006	ESP	837	ROU	698	NOR	532
NLD	996	ARG	823	SWE	696	SVN	508
DEU	973	POL	799	FIN	684	URY	492
ITA	930	DNK	789	TUR	672	GRC	443
CZE	908	LUX	780	TWN	656	PKS	443
FRA	904	VNM	780	BRA	652	UKR	384
CHN	879	CHL	776	BLR	615	IDN	333
AUS	864	IRL	765	SAU	615		
BEL	858	MEX	763	PRT	614		

USDA

Numbers of variants tested in cattle

Year	Variants	Data source
<1995	0	Only pedigrees and phenotypes
2003	367	RFLP markers, not implemented
2008	50,000	50K SNP chip
2010	777,000	High-density chip
2015 2019	39,700,000 63,000,000	Whole genome sequencing (1000 Bull Genomes Project)
Future	New mutations (e.g., HCD)	Sequence each new AI bull
Future	Epigenetic	External control of gene function

USDA

Chip history (37 chips + sequence)
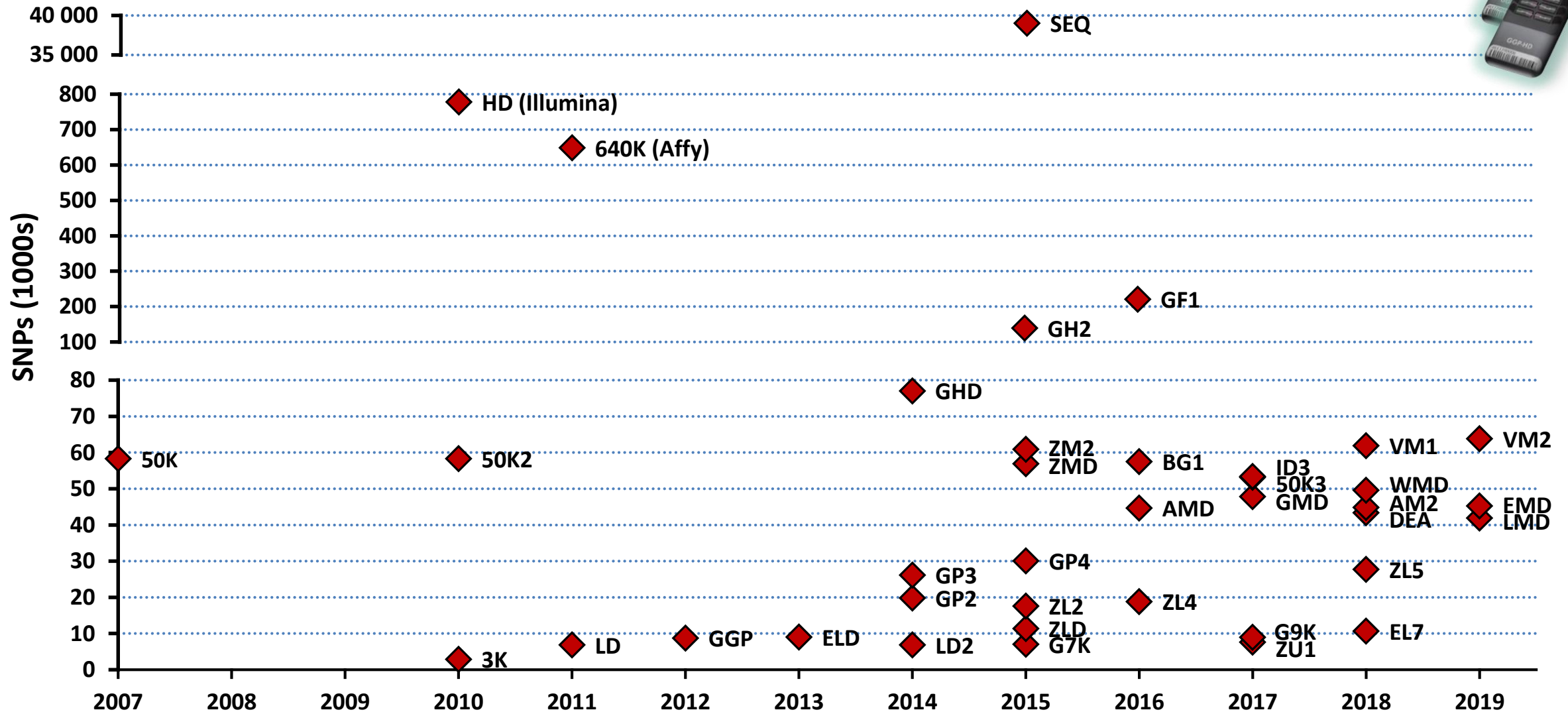
Human genotyping and sequencing (January 2019)

Technology	Database	People genotyped	Variants genotyped
High-density (HD) SNP chip	Ancestry.com	>10 million	700,000
	23andMe	Millions	700,000
	UK Biobank	483,000	826,000
Whole genome sequence	TopMed (U. Michigan)	135,000	500 million
	Illumina	600,000	500 million
HD imputed to sequence	U. Michigan	>25 million	500 million

Total raw data = 600,000 people × 40× coverage × 2.7 billion genome length) = 65 million GB

USDA

Detect new mutations by sequencing new AI bulls?

- **Normal DNA in both parents, but progeny has new mutation**

- **New dominant mutation examples**
  - NZL: **Half** of Halcyon daughters had thick hair and **gave no milk**
  - CAN: Rosabel had new dominant mutation for red hair color
  - Europe: **Bulldog** calves from Igale Masc **(1%)**, Captivo **(12%)**, Energy **(20%)**

- **New recessive mutation examples**
  - CAN: Maughlin Storm with **HCD** mutation causing calf death
  - USA: 118 animals with large new **chromosome deletions**; 252 with **XXY**

Statistical methods

- **Many models to choose from and several currently in use**

- **17 countries filled out Interbull Form GENO:**

  - **10 use multi-step GBLUP linear model**

  - **4 use multi-step Bayesian model (non-normal SNP effect distributions)**

  - **2 use single-step GBLUP linear model**

  - **1 uses linear haplotype model after choosing subset by Bayesian model**

  - **9 of the 17 countries include a polygenic term (non-SNP genetic variance)**

- **Where did these methods come from?**

USDA

Linear model using genomic relationships

- **Nejati-Javaremi, Smith, Gibson (1997) derived GBLUP methods**

- **Compared G matrix to A matrix**

  - **Used individual SNP genotypes (not haplotypes)**

  - **Simulated 1,000 phenotypes with 30% heritability**

  - **Selected directly on 100 QTLs (not markers, no polygenic effect)**

  - **Obtained 71% REL with G matrix vs. 42% REL with A matrix**

- **Research was from U. Guelph, Canada, but was not cited by Schaeffer (2006)**

- **First author is a professor at U. Tehran, Iran**

Bayesian haplotype models using Gibbs sampling

- **Meuwissen, Hayes, Goddard (2001) compared Bayesian to linear models**

  - **Used haplotypes (not individual SNP effects) from 1,000 markers**

  - **Simulated 2,200 phenotypes with 50% heritability and 100 QTLs with some very large effects**

  - **Obtained REL = 20% pedigree, 54% GBLUP, 64% BayesA, 72% BayesB**

  - **Did not cite Nejati-Javaremi et al (1997 JAS) in their 2001 or later papers**

- **Calus et al (2008) showed that individual SNP effects are better than haplotypes at higher SNP densities; polygenic effects should be included**

- **BayesA and BayesB had much smaller gains in later studies of real traits**

USDA

# Steps taken to reduce upward bias of young bulls

- **Highest parent averages and elite cows had been inflated for decades**

  - **Animal model rewarded cow families within herds**

- **Early genomic PTAs had similar biases, later corrected by:**

  - **Including a polygenic effect in the model (2010)**

  - **Adjusting the Mendelian sampling of top cows downward (2010)**

  - **Using pedigree index instead of parent average (mainly in Europe)**

  - **Reducing the weight on direct genomic value (2011)**

  - **Using lower heritability estimates for yield traits (2014)**
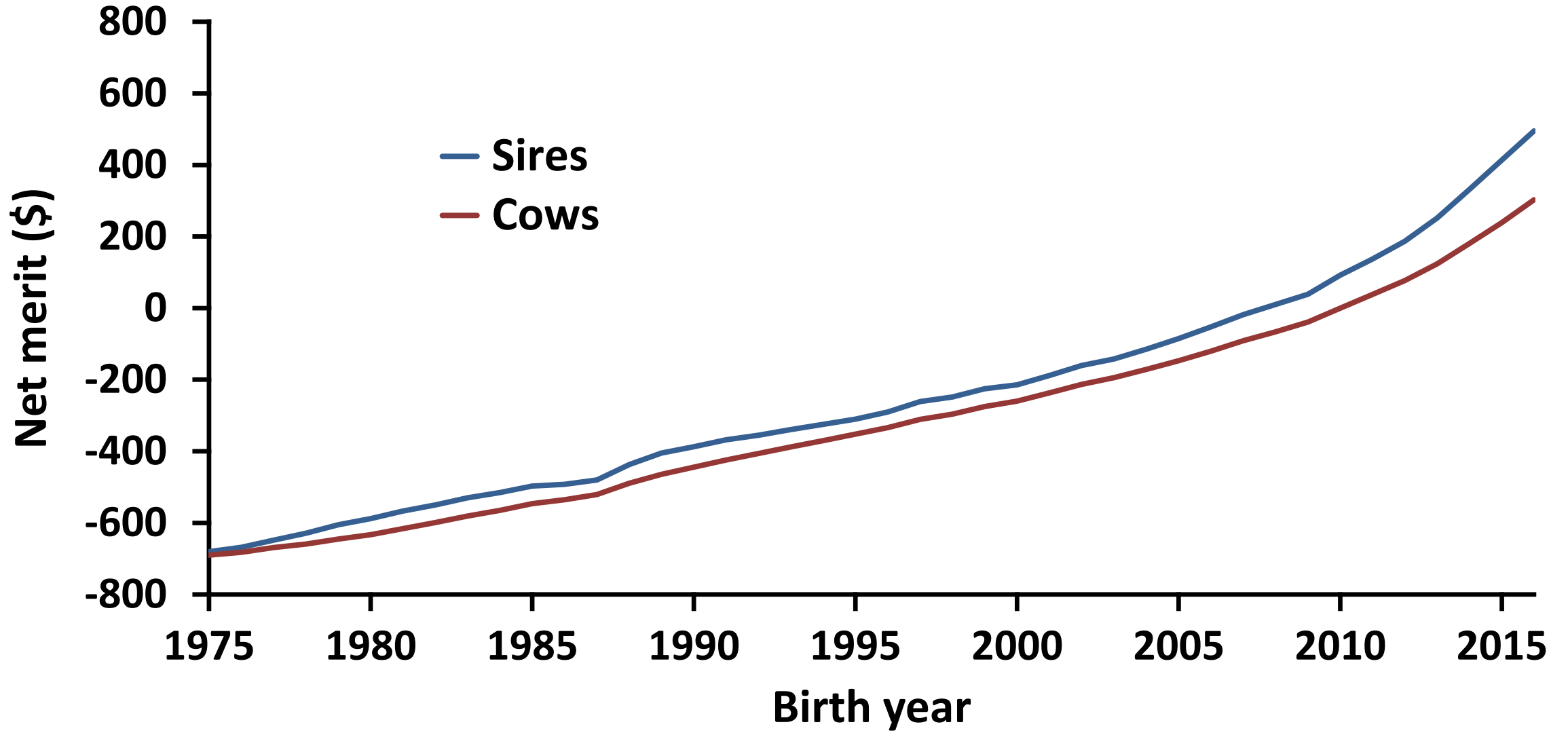
USDA

Proven vs. young bull merit and bias across time
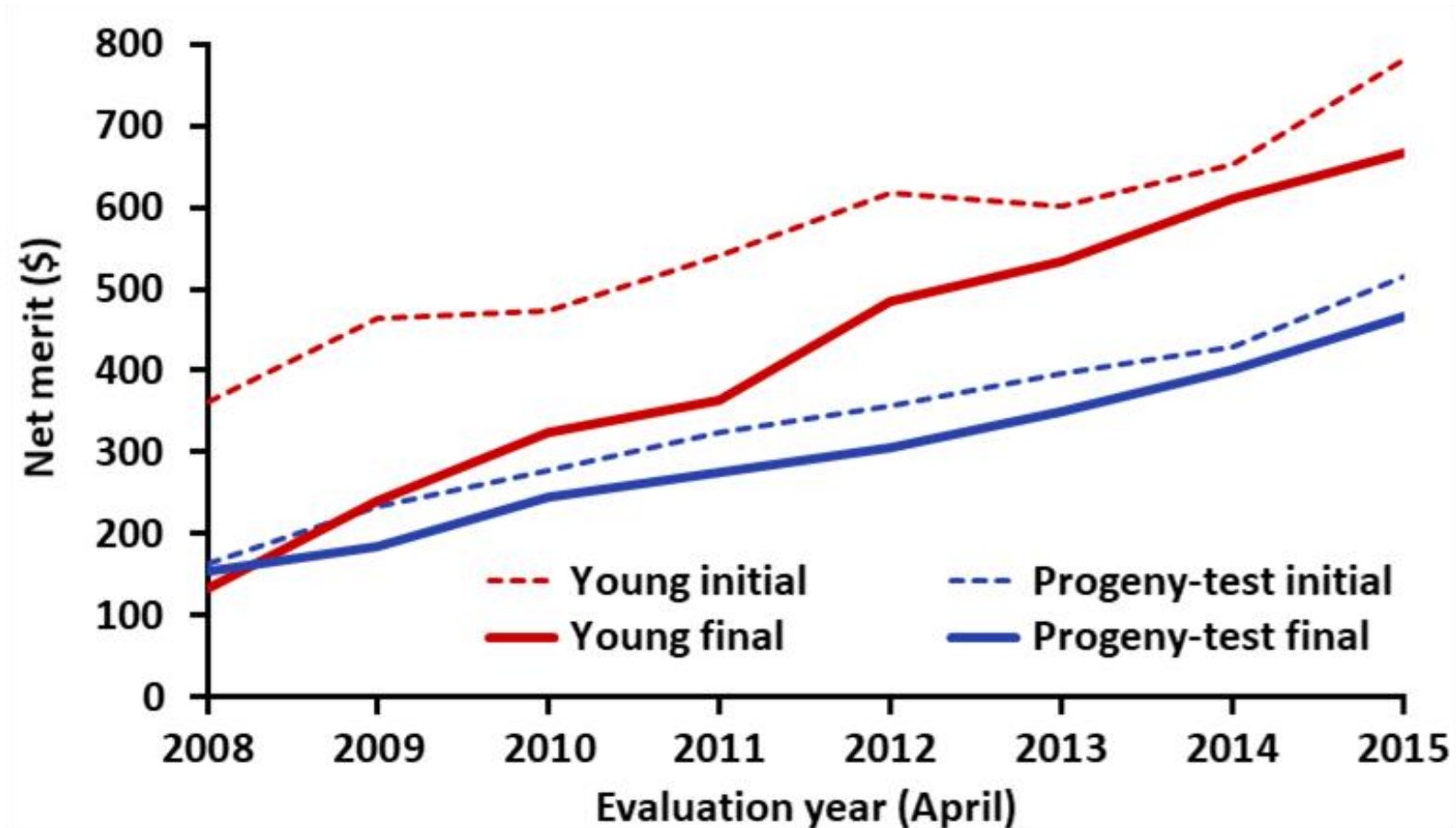
Initial and final (Apr 2019) PTA averages of the top 100 NM$ USA AI Holsteins.
Young: bulls with no milking daughters. Proven: bulls with ≥ 100 daughters.
NM$ adjusted to current base.

Generation interval – Holstein
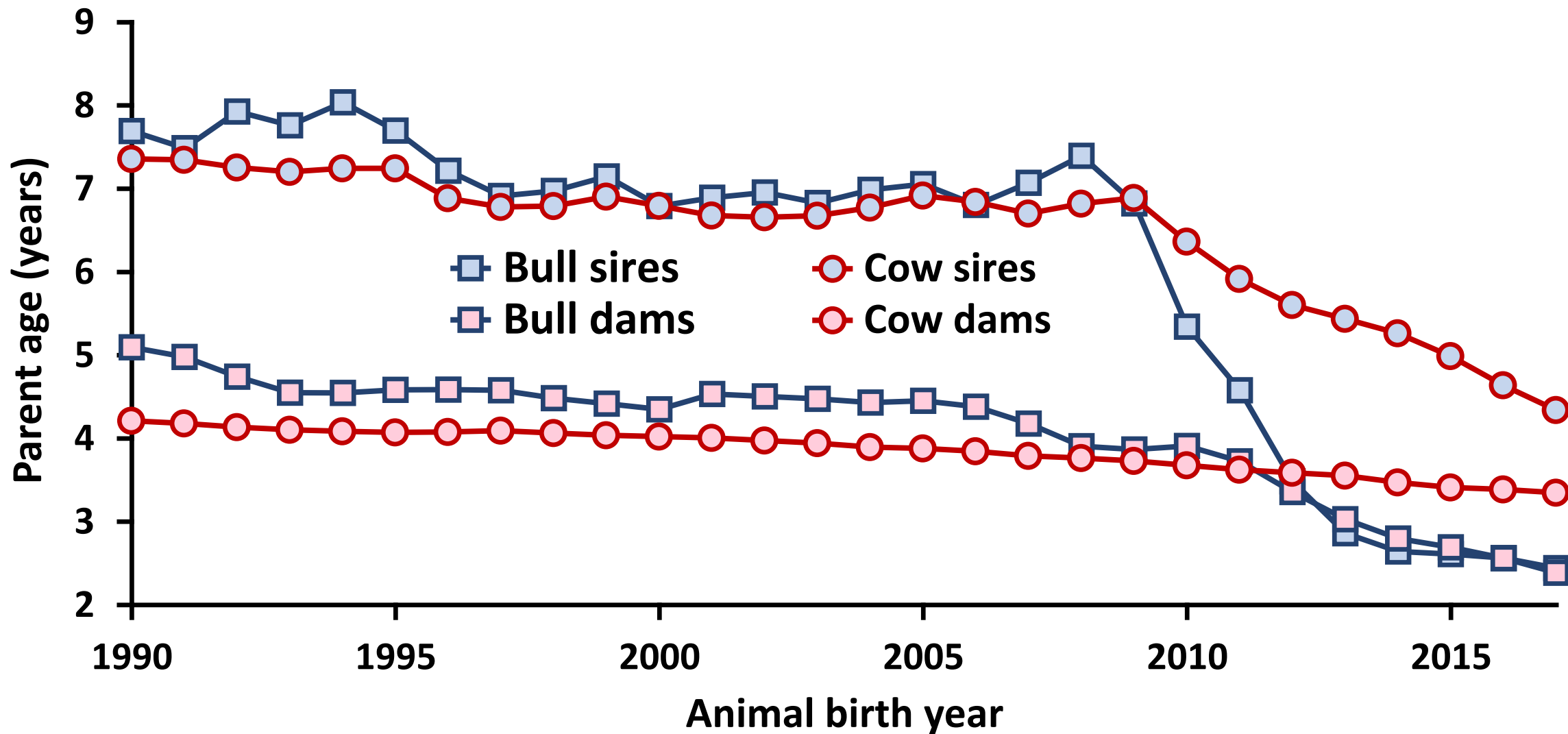
Genetic progress (Holstein)

Predict 50 years of genetic progress (Cole, 2018)

Trait	50-year progress	2017 mean	2067 mean
Milk (pounds)	+9,329	27,055	36,384
Fat (pounds)	+531	1,008	1,539
Protein (pounds)	+342	825	1,167
Productive life (months)	+65	26	91
Body size (pounds)	−230	1,500	1,230
Udder composite	+2.65	0	2.65
Daughter pregnancy rate (%)	+26	29	55
Cow conception rate (%)	+42	35	77
Heifer conception rate (%)	+21	57	78

USDA

Selection for more traits

- **26 traits had US genomic PTAs in 2009**

- **10 new traits added in 2010-2019, mostly with low h^2**

  - **Heifer and cow conception rates, cow livability, gestation length, mastitis, metritis, retained placenta, DA, ketosis, milk fever, and early first calving**

- **Several more future traits expected soon:**

  - **Feed intake, calf livability, heat tolerance**

- **High genomic reliability requires:**

  - **Large reference populations, historical databases, or high h^2**

USDA

# Genomic mating programs

- **Reduce inbreeding using genomic instead of pedigree relationships**
  - **Genomic relationship of each live female to each marketed bull**
  - **File contains 1 million females × 5,000 males**

- **Genomic mating increases heifer calf value by**
  - **+$33 compared with pedigree mating (1.3% lower inbreeding)**
  - **+$78 compared with random mating (3.1% lower inbreeding)**

- **Also improves conception rates by avoiding recessive carrier matings**

- **Also reduces inbreeding of bull calves, promoting faster growth**

Ancestor discovery

- **Accurate discovery if true ancestor is genotyped**

- **Accurate detection if reported pedigree is wrong**

- **Accuracy slightly reduced with lower density chips**

- **Methods published and used since 2013 (VanRaden et al.)**

Ancestor	Correctly discovered
Sire	~100%
MGS	97%
MGGS	92%

USDA

Discovering pedigrees (parents, grandparents, etc.)
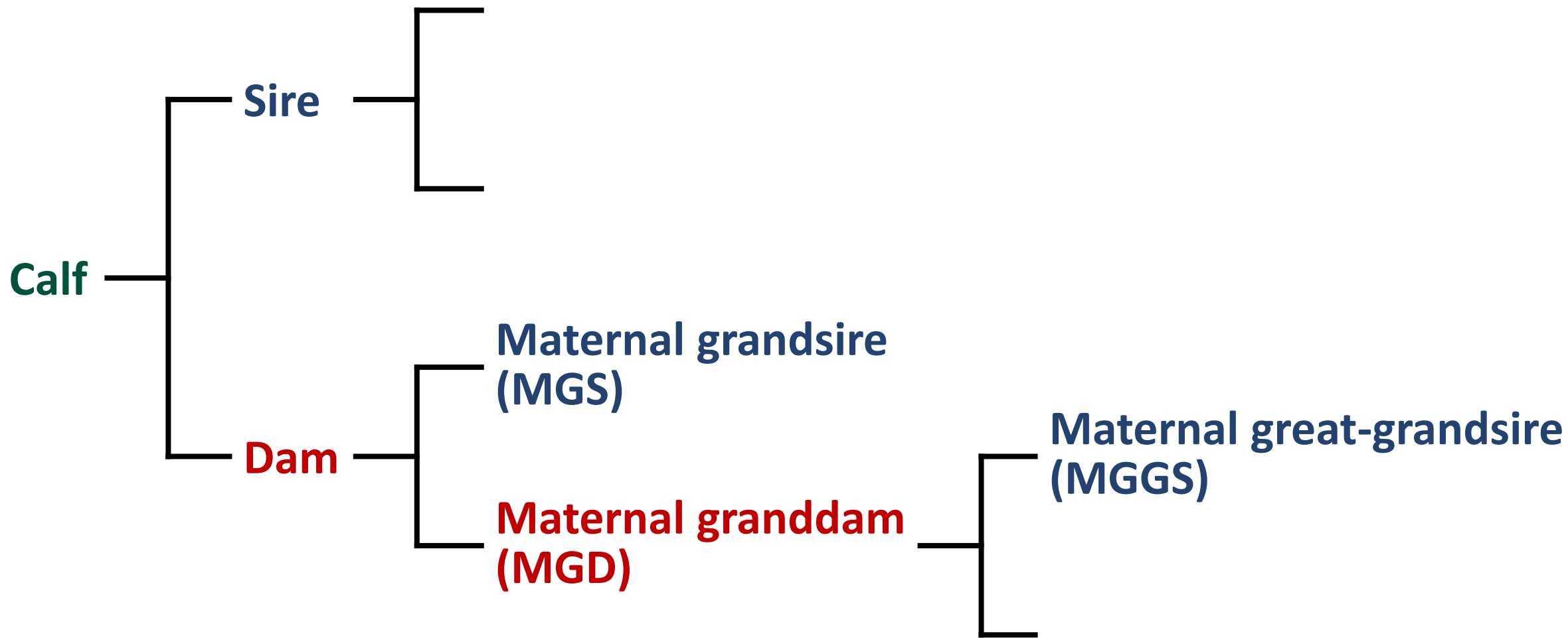
Filling in pedigrees

- **291,131 discovered MGS not yet used because no dam ID available**

- **153,909 discovered MGGS not yet used because no MGD ID available**

- **Procedures developed to fill missing IDs to link to ancestors**
  - **HO999DAM000000001, for example**
  - **Should predictions use and public see the discovered pedigrees?**

- **First check if true dam can be discovered in same herd (33,810 found)**
  - **Match birth and fresh dates (only 1 dam's pedigree matches calf's)**

- **Postdoc Juan Nani is documenting benefits from pedigree discovery**

Genetic and genomic milestones

Year	Researchers	Discovery/Development
1865	Mendel	Inherited traits of peas
1900	Bateson and others	Rediscovery of Mendel's laws
1905	Spillman (USDA)	Horned/polled (Mendelian trait in cattle)
1922	Wright (USDA)	Pedigree relationship matrix
1953	Watson, Crick	Structure of DNA
1974	Norman (USDA), Henderson (1976)	Use of pedigree relationships in PTAs
2001	Lander	Sequence of human genome
2007	Illumina, USDA, and others	Development of SNP chip for cattle
2009	Elsik (NIH and USDA funding)	Sequence of cattle genome
2009	USDA and many others	Genomic methods, official predictions

USDA

Genomic selection in other species: Example in peas

- **Genomic methods from dairy cattle are now used for many species**

- **Example: Roorkiwal et al. (2016) Front. Plant Sci. 7:1666**

  - **3,000 markers for 315 elite lines of chickpeas in India**

  - Tested G-BLUP, Bayes B, Bayes Cπ, Bayesian LASSO, and Random Forest

  - **"... models did not show any significant difference for estimating prediction accuracy within traits"**

  - **"... study establishes the necessary resources for deployment of genomic selection in chickpea breeding"**

Summary

- **The DNA predictions implemented 10 years ago have improved rapidly**
  - **Adjustments were needed to remove biases, give breeders confidence**
  - **Accuracy improved as datasets expanded, more traits were included**
  - **Many new uses such as pedigree discovery and genomic mating**
- **Future predictions will use more gene tests discovered from sequence data and more international genotypes**

- **Methods developed for dairy cattle are now used for beef, swine, poultry, humans, corn, alfalfa, soybeans, and even peas**

- **Mendel would be pleased**

Acknowledgments and Disclaimers

- **Council on Dairy Cattle Breeding (CDCB) and its industry suppliers for data**

- **Taxpayers for funding USDA-ARS-AGIL project 8042-31000-002-00, "Improving dairy animals by increasing accuracy of genomic prediction, evaluating new traits, and redefining selection goals"**

- **AGIL staff for doing this research**

- **Mention of trade names or commercial products in this presentation is solely for the purpose of providing specific information and does not imply recommendation or endorsement by USDA; USDA is an equal opportunity provider and employer**

USDA

What's the best cow we can make?
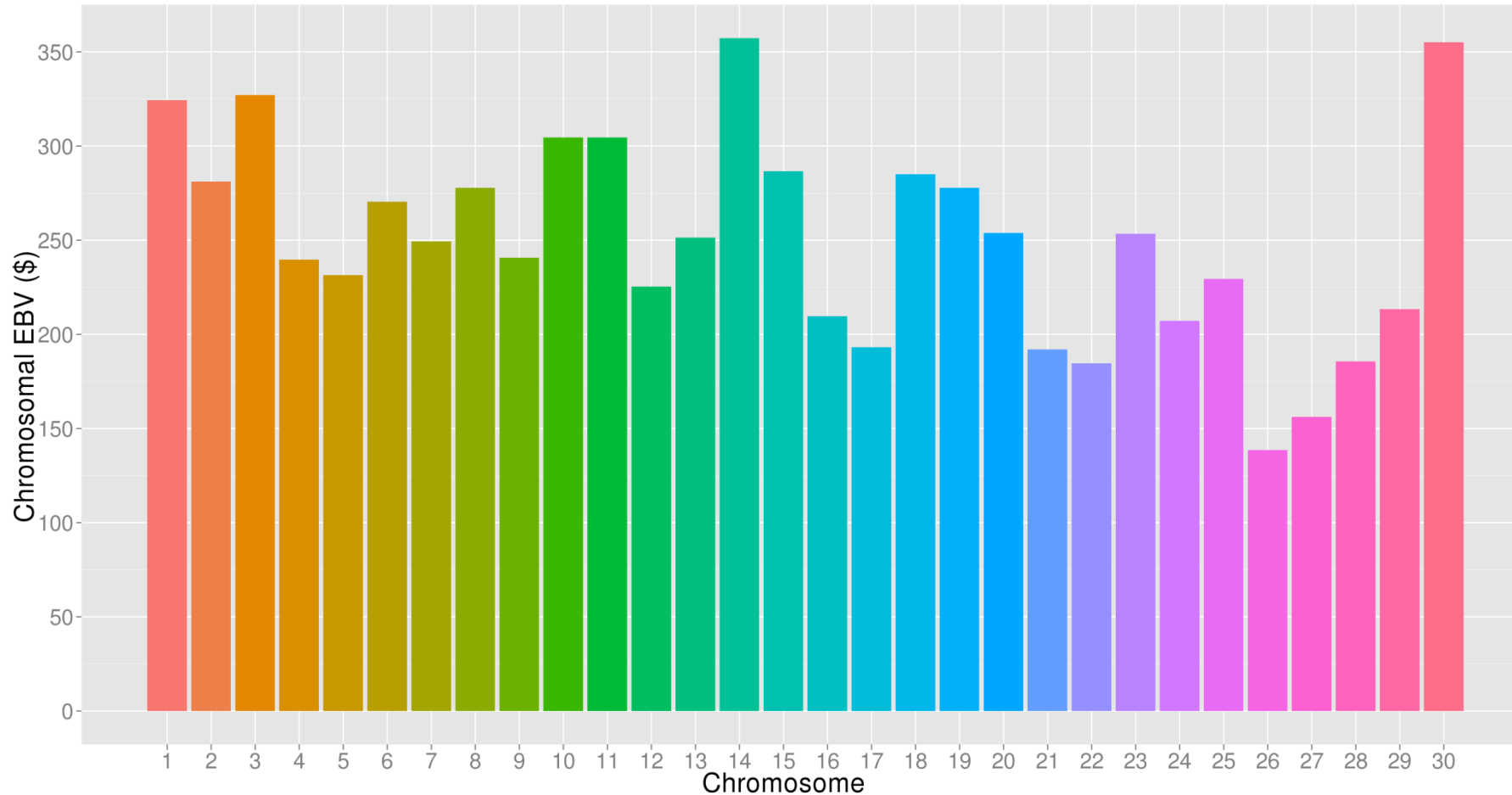
A "supercow" constructed from the best haplotypes in the Holstein population would have an EBV for NM$ of +$7,515 !