

Genomic selection: Achievements in the last decade and plans for the next



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/>

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
1976	Henderson (also Norman, USDA)	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	Genomic methods, official predictions

Progress in past decade and next decades

- **Progress in 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**

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

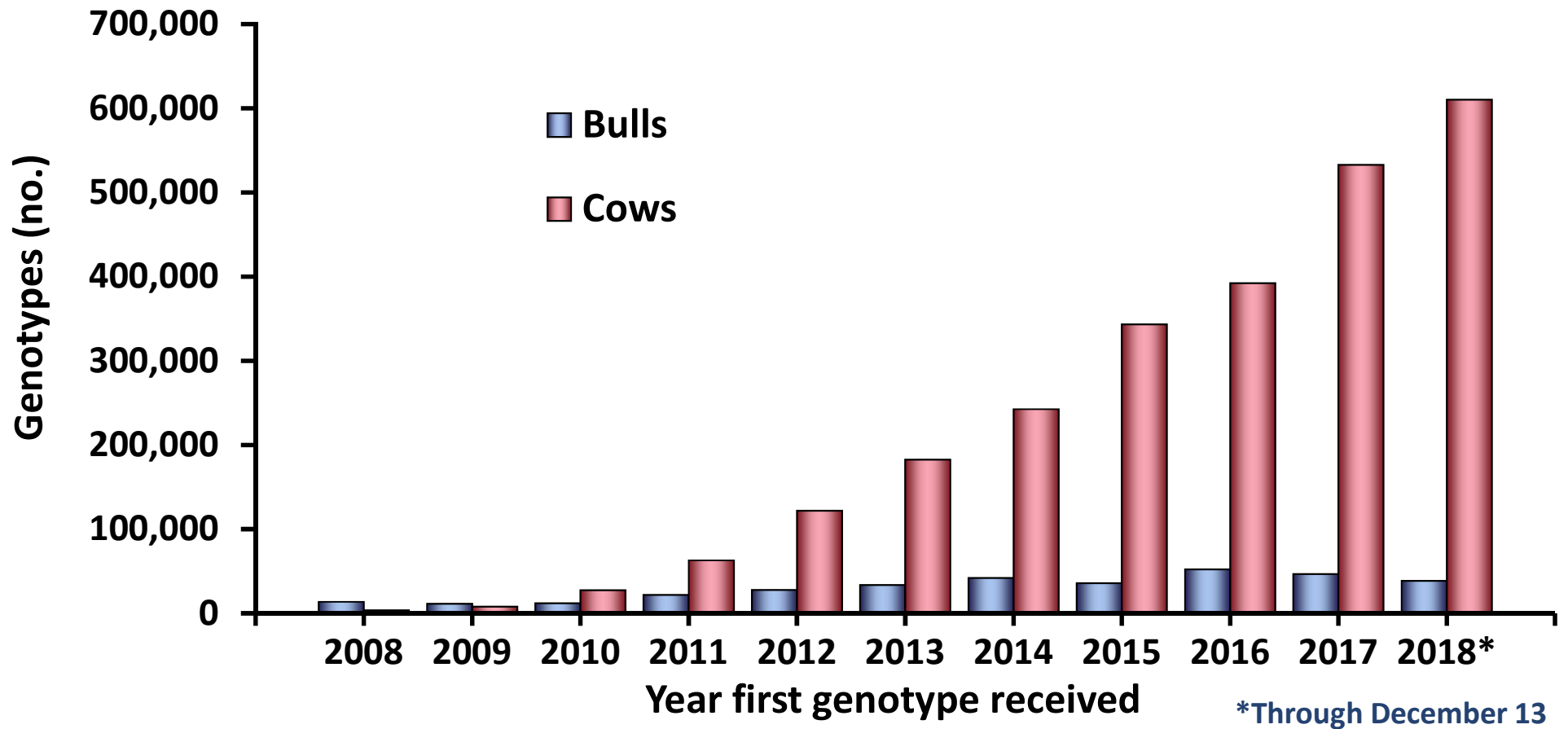
*Used customized Illumina 50K chip
(different markers)

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

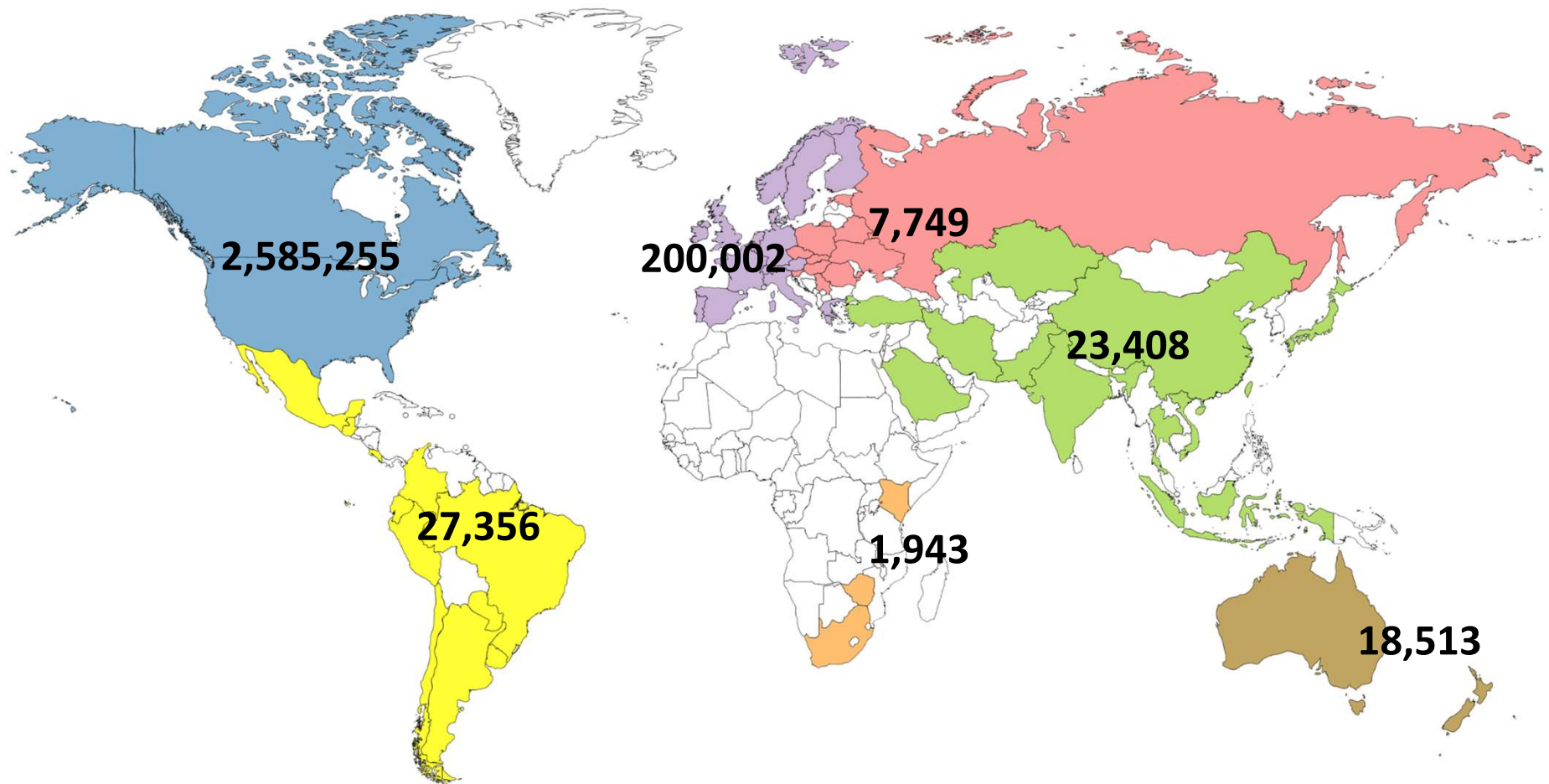
Worldwide livestock genotyping (January 2019)

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

CDCB usable genotype counts/year by animal sex



Genotyped animals in database by region (2018)



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		

Formal international genotype exchanges

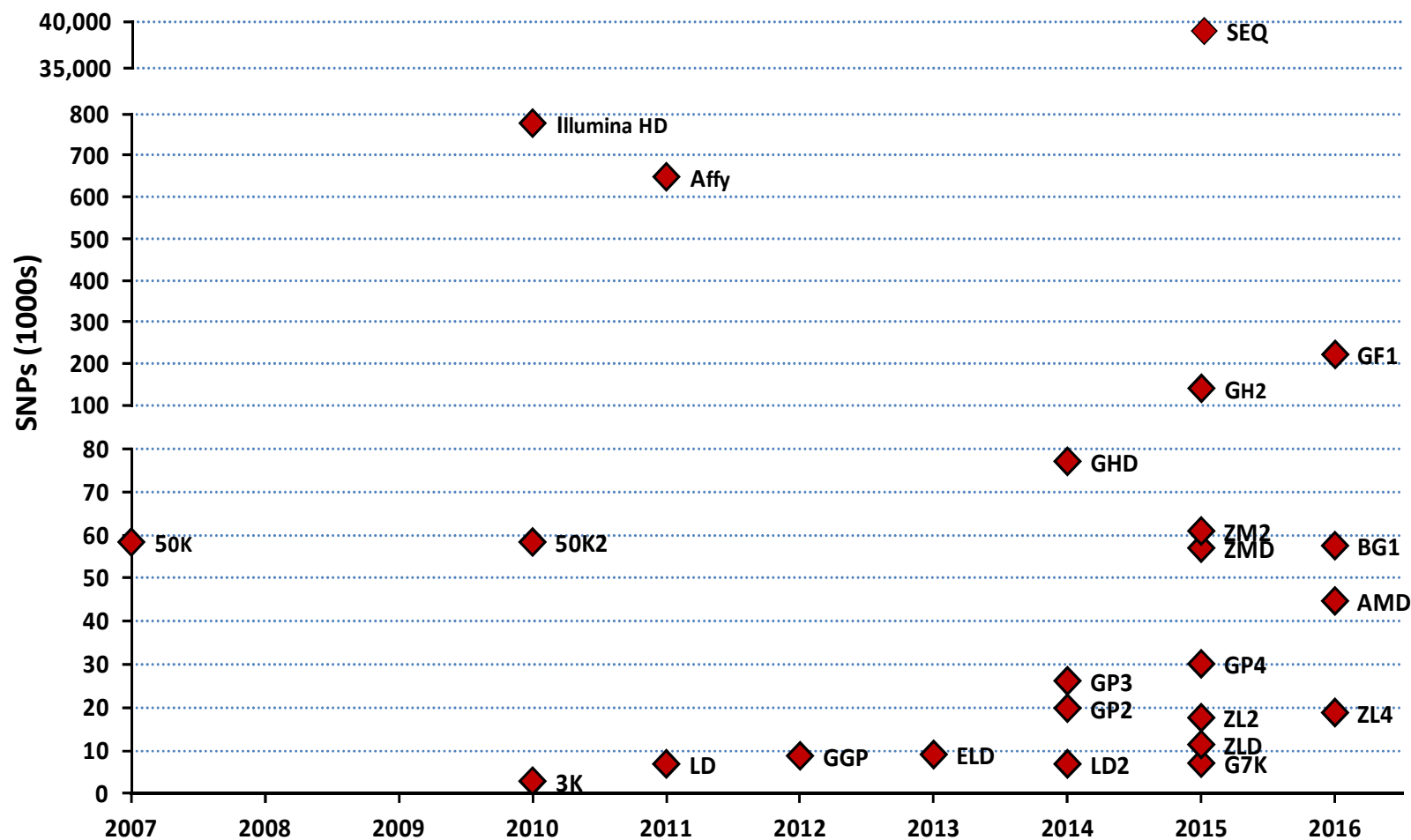
Country exchanges with U.S.	Breed	Sex	Since
Canada	All	All	2009
United Kingdom, Italy	Holstein	Bulls	2011
All Europe via Interbull	Brown Swiss	Bulls	2012
Denmark	Jersey	Bulls	2014
United Kingdom	Guernsey	All	2016
Japan, Switzerland, Germany	Holstein	Bulls	2016



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	39,700,000	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

Many new chips have been added



Predictions now use more SNPs

- New list of **79,239** instead of **60,671** SNPs
 - SNPs from high density and sequence data with biggest NM\$ effects
 - More gene tests (**DGAT1, ABCG2, casein, leptin**) from recent chips
 - More gene tests (**HH0, HH3, HH4, BH2, AH1, AH2**) in fertility haplotypes
 - Improved SNP editing system and improved prediction reliability (**1–3%**)
- Implemented by CDCB in **December 2018**
- Most other countries still use **50,000** SNPs from **2009**

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

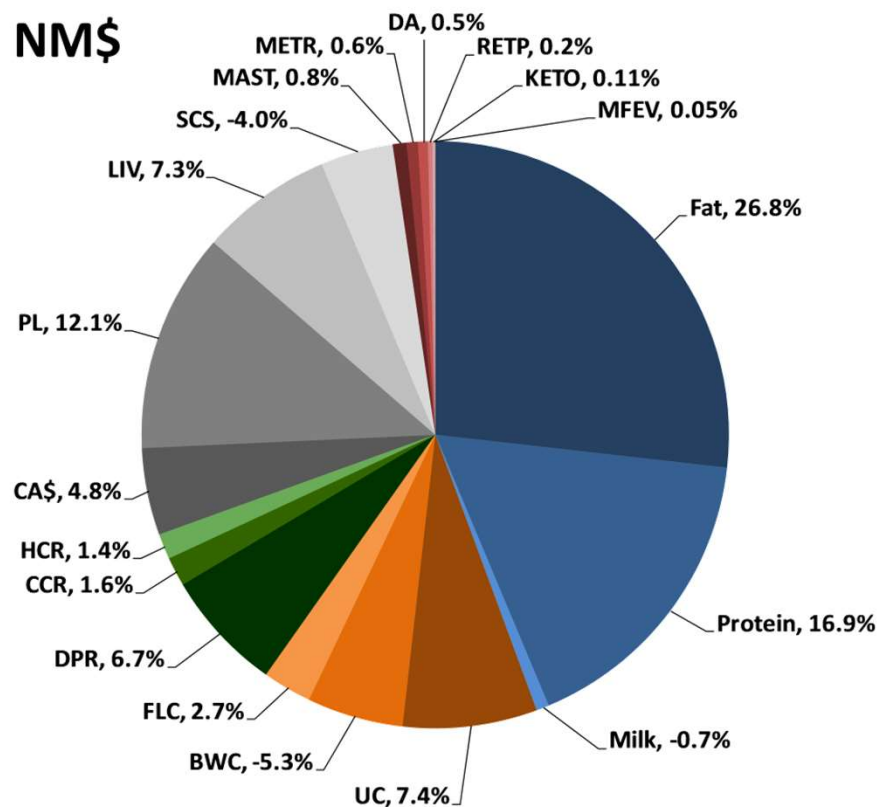
Total raw data = 600,000 people × 40× coverage × 2.7 billion genome length) = 65 million GB

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**



Relative weighting in 2018 net merit (NM\$)



UC	Udder composite
BWC	Body weight composite
FLC	Feet-legs composite
DPR	Daughter pregnancy rate
CCR	Cow conception rate
HCR	Heifer conception rate
CA\$	Calving ability (calving ease & stillbirth rate)
PL	Productive life
LIV	Livability
SCS	Somatic cell score
MAST	Mastitis
METR	Metritis
DA	Displaced abomasum
RETP	Retained placenta
KETO	Ketosis
MFEV	Milk fever

New traits and evaluations expected

- April 2019
 - Genomic predictions for **crossbreds**
 - **Earlier first calving** age (heifer maturity rate)
- Future: More traits already tested in USDA research
 - Individual **feed intakes** measured for 5,000 cows
 - **Heat stress**: Cows adapted to southern vs. northern climate
 - **Persistency**: Productivity in late vs. peak lactation
- Other new traits require local data be combined with national data

Earlier first calving age (EFC)

- Heifers eat feed but produce no milk until calving
 - Earlier calving is worth **\$2.50** per day
 - Economic emphasis could be **3%** of NM\$
 - Removes some emphasis from heifer conception rate
- Large database (**23 million records**) available for EFC
 - Heritability of **2.7%**
 - PTA standard deviation only about **3** days
 - Reliability of genomic predictions of **66%** for Holsteins



Genomic predictions for crossbreds

- Genomic predictions computed within-breed since 2009
- More crossbreds will receive predictions in April 2019
 - Weighted combination of purebred predictions
 - Breed base representation (BBR): Genetic contribution of each breed
- Currently about 60,000 animals with **>10%** BBR from another breed
- About 17,000 animals with **6–10%** BBR from another breed
 - Will no longer contribute to purebred reference
 - Will be predicted using only purebred data



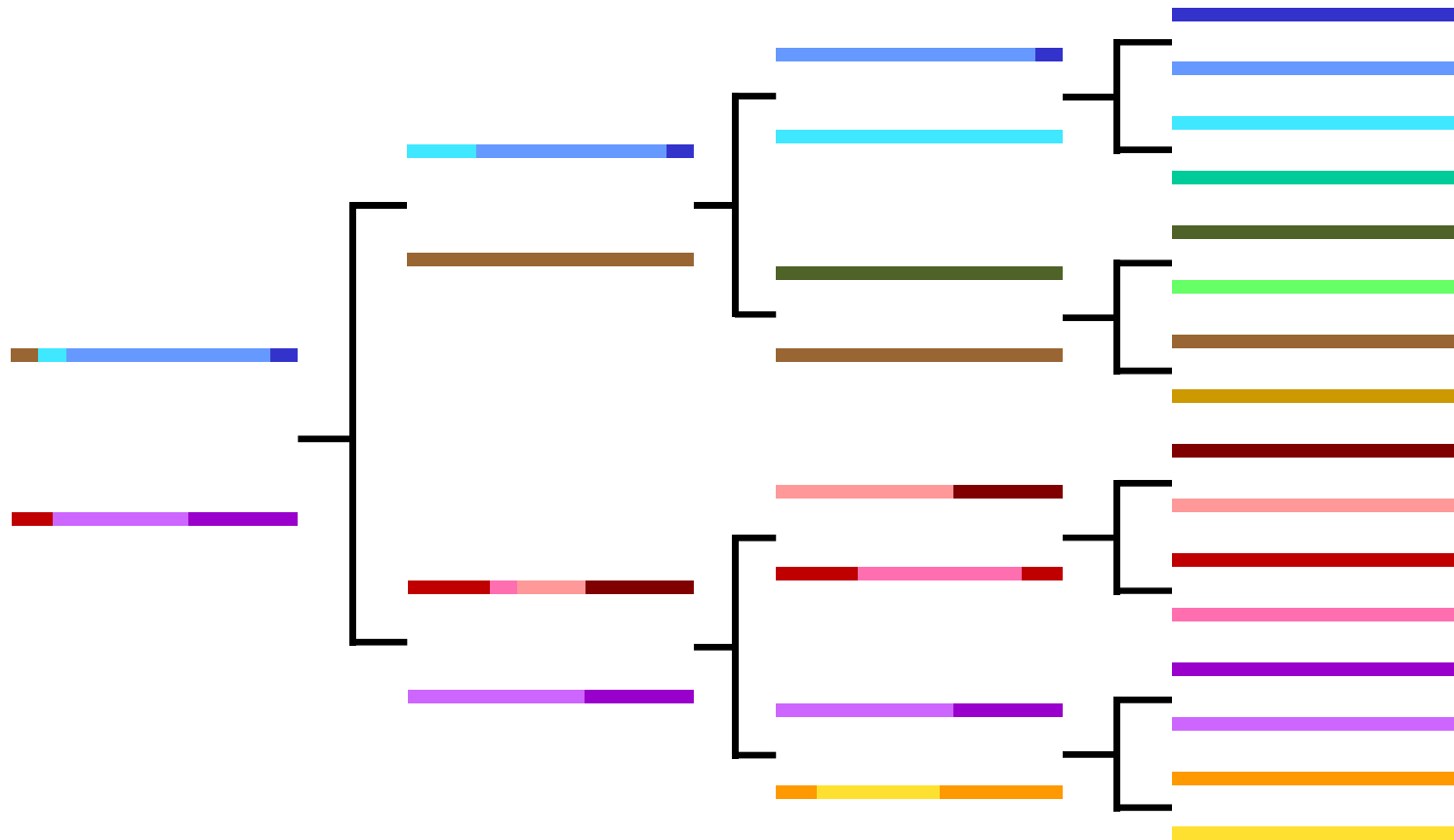
Genomic mating programs



- Reduce inbreeding using genomic instead of pedigree relationships
 - Genomic relationship of each live female to each marketed bull
 - File contains 1 million females \times 5,000 males
- Genomic mating increases heifer calf value by
 - **+\$84** compared with pedigree mating (1.3% lower inbreeding)
 - **+\$214** compared with random mating (2.9% lower inbreeding)
- Also improves conception rates by avoiding recessive carrier matings
- Also reduces inbreeding of bull calves, promoting faster growth

Bull's chromosome pedigree (O-Style)

Chromosome 15

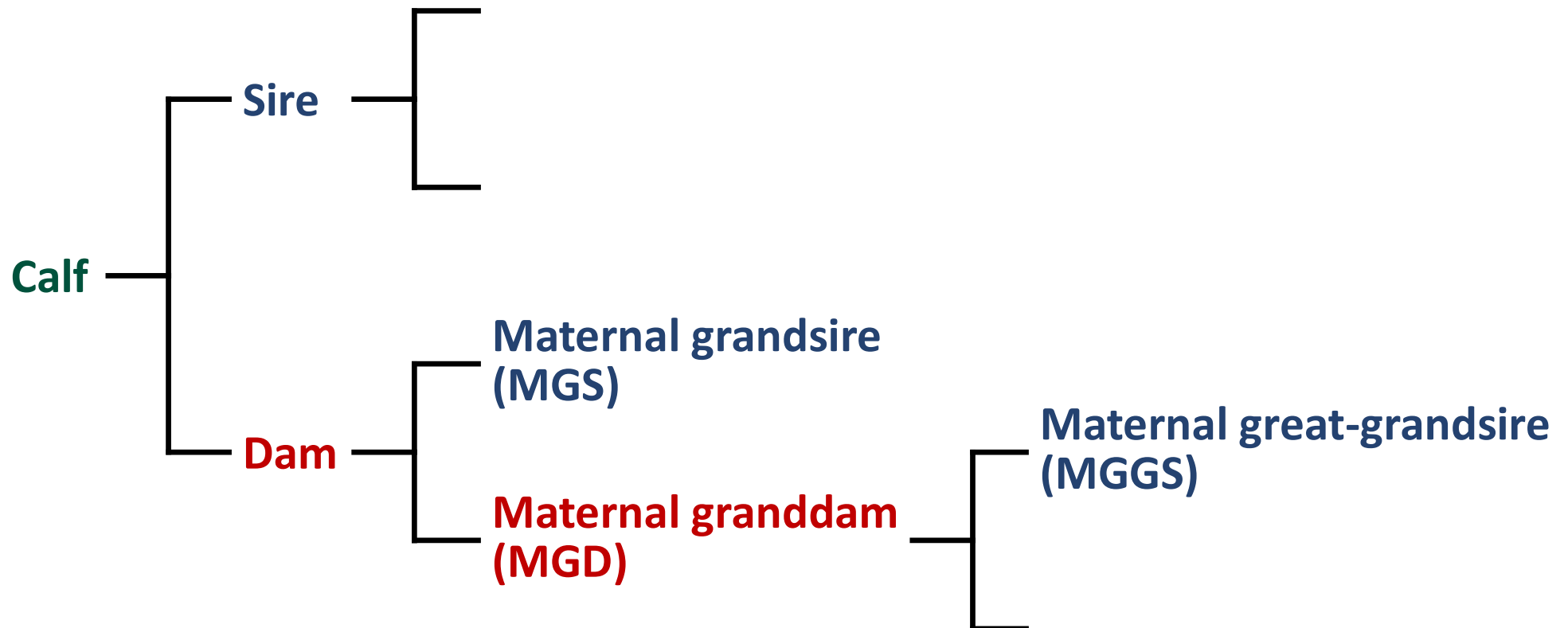


Ancestor discovery

- If true ancestor genotyped, accuracy of discovery very high
- If reported pedigree wrong, accuracy of detection also high
- With lower density chips, accuracy slightly reduced
- Methods published and used since 2013 (VanRaden et al.)

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

Discovering pedigrees (parents, grandparents, etc.)



Filling in pedigrees

- >200,000 discovered **MGS** not yet used because no **dam** ID available
- >200,000 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 (**20,000 found**)
 - Match birth and fresh dates (**only 1 dam's pedigree matches calf's**)

Genotyping is easy

- DNA source sent to genotyping lab (2018)

Source	Samples (no.)	Samples (%)
Tissue (ear)	579,255	81
Hair	102,229	14
Blood	14,656	2
Nasal swab	656	<1
Semen	319	<1
Unknown	17,048	2

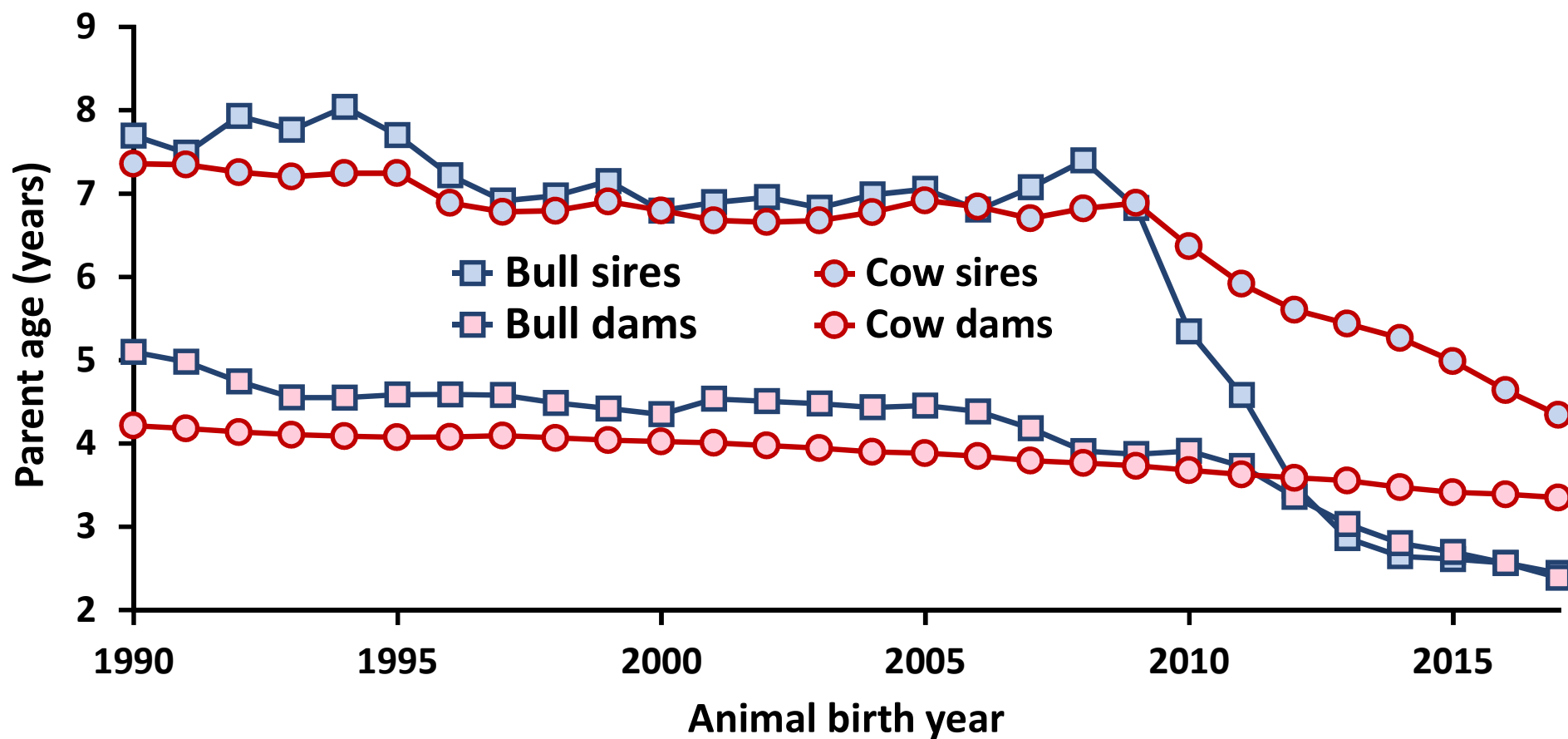


Photo source: Zoetis

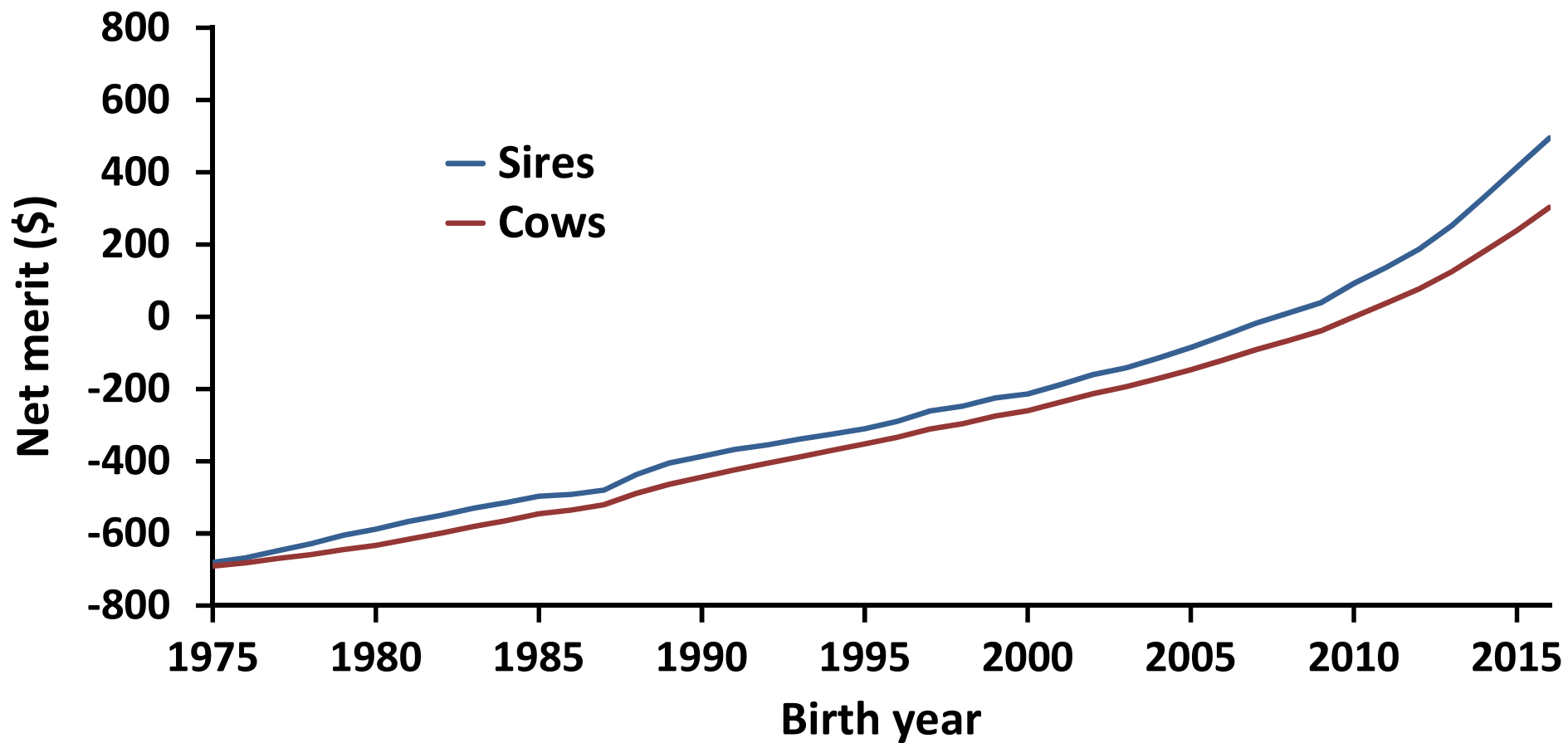
Potential for embryo genotyping

- **13,000** 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

Generation interval – Holstein



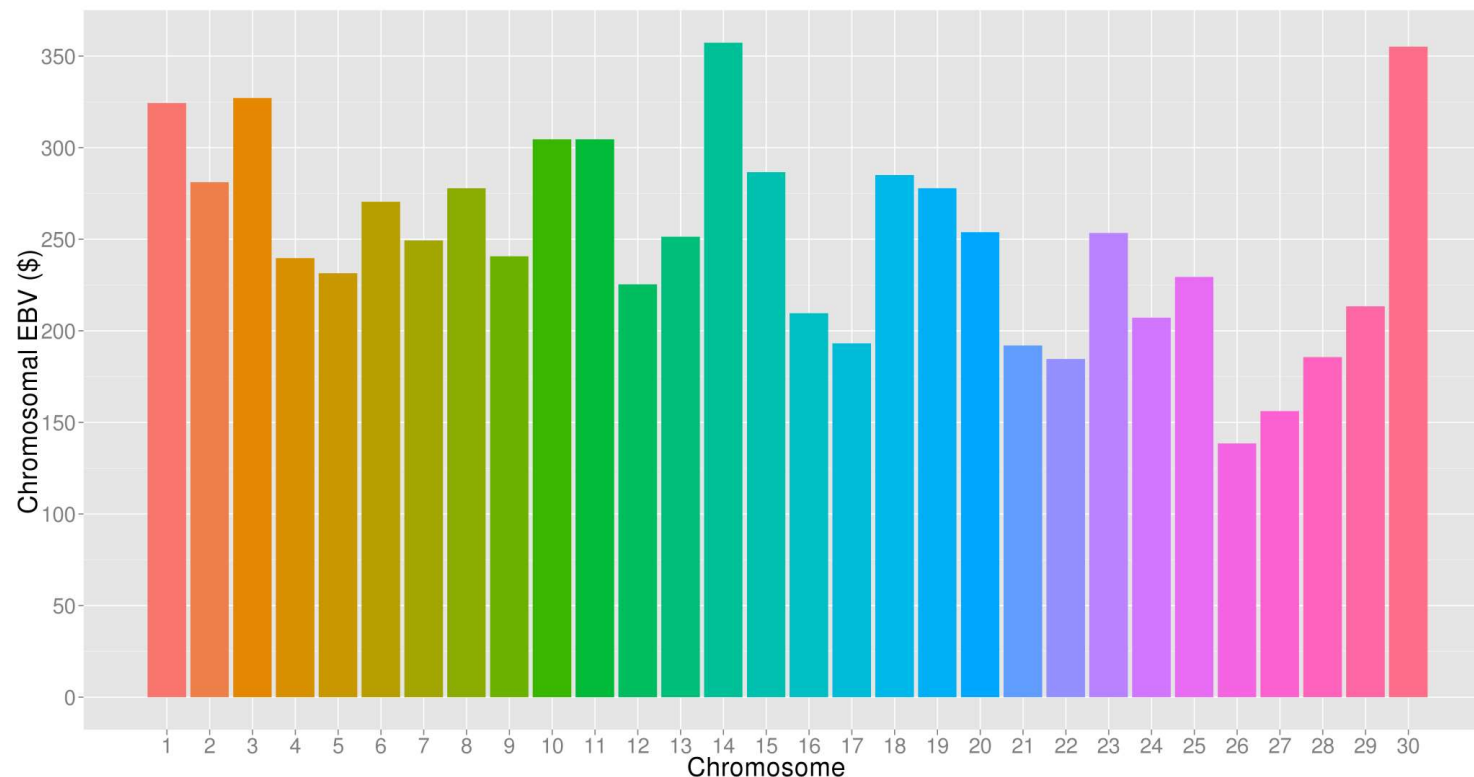
Genetic progress (Holstein)



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

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 !**

USDA-AGIL 5-year plan

- **Objective 1: Expand genomic data** used in prediction by selecting new variants that more precisely track the true gene mutations that cause phenotypic differences
- **Objective 2: Evaluate new traits** that can all be predicted at birth from the same inexpensive DNA sample
- **Objective 3: Improve efficiency of genomic prediction** and computation by developing faster algorithms, testing new adjustments and models, and accounting for genomic pre-selection in evaluation
- Full research plan: <https://aipl.arsusda.gov/AboutUs/AboutUs.htm>

Summary

- The DNA predictions adopted 10 years ago have improved rapidly
 - Accuracy improves as datasets expand and more traits are included
 - Many new uses such as pedigree discovery and genomic mating
 - New predictions for crossbreds and early first calving coming in April
- 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
- Genotyping labs for adding many important SNPs to their chips
- 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”
- 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

Genetic-economic index changes across time

Trait	Relative emphasis on traits (%)								
	PD\$ 1971	MFP\$ 1976	NM\$ 1994	NM\$ 2000	NM\$ 2003	NM\$ 2006	NM\$ 2014	NM\$ 2017	NM\$ 2018
Milk	52	27	6	5	0	0	-1	-1	-1
Fat	48	46	25	21	22	23	22	24	27
Protein	...	27	43	36	33	23	20	18	17
PL	20	14	11	17	19	13	12
SCS	-6	-9	-9	-9	-7	-7	-4
BWC	-4	-3	-4	-5	-6	-5
UC	7	7	6	8	7	7
FLC	4	4	3	3	3	3
DPR	7	9	7	7	7
CA\$	6	5	5	5
HCR	1	1	1
CCR	2	2	2
LIV	7	7
Health\$	2

New cattle genome reference map



- New **2018** ARS-UCD map replaced previous **2009** UMD map in December
 - Both about 2.7 billion bases of DNA across 30 chromosomes
- **ACGTGTCATCGATCCTAGGCTTAGCTACGATCCGATCGATTTCGATCGATCGAATCG**
- The reference map
 - Lets researchers track genetic differences using a common language
 - Shows **where genes are** and how DNA encodes proteins
 - **Corrects locations** of many SNPs
 - Improves **imputation** of missing SNPs from low-density chips