Adding and reporting genomically discovered ancestors in US evaluations

José A. Carrillo[1], Rodrigo Mota[1], George Wiggans[1], Lillian Bacheller[1], Gary Fok[2], Dan Null[2], Emmanuella Ogwo[2], and Paul VanRaden[2]

[1]CDCB - Council on Dairy Cattle Breeding, Bowie, MD, USA
[2]Animal Genomics and Improvement Laboratory, ARS, USDA, Beltsville, MD, USA

Interbull Meeting - Montreal, Canada

May 31, 2022

Discovering pedigrees in the US database

- Each animal entering the database is checked for dam and sire (counting opposite homozygotes)

- Mendelian conflicts resolved

- Imputation to 80K using **Findhap.f90**

- Proportions of **validated sires** and **dams** are different:

  - 2,868,531 genotypes (88.1% cows)

  - 78% sires confirmed - 33% dams confirmed

  - 80% calves with no pedigree information received a Sire suggestion

  - 18% calves with no pedigree information received a Dam suggestion

Discovering pedigrees

Fixped.f90 uses haplotypes to discover distant relationships as MGS and MGGS when
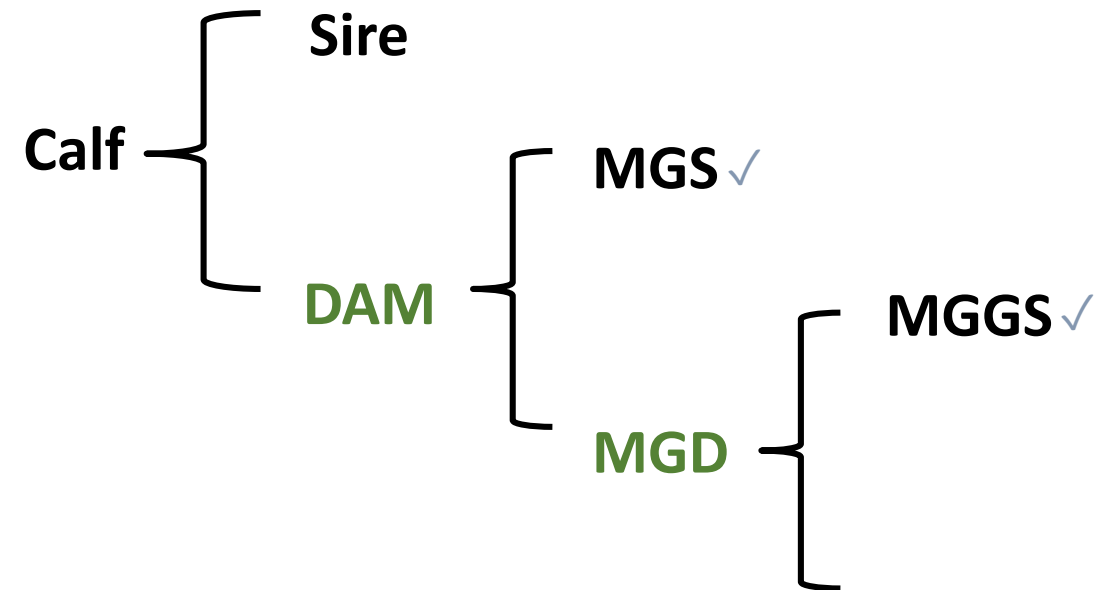
DAM is missing or not reported (VanRaden et al., 2013)

Sire

Calf

MGS

Dam

MGGS

MGD
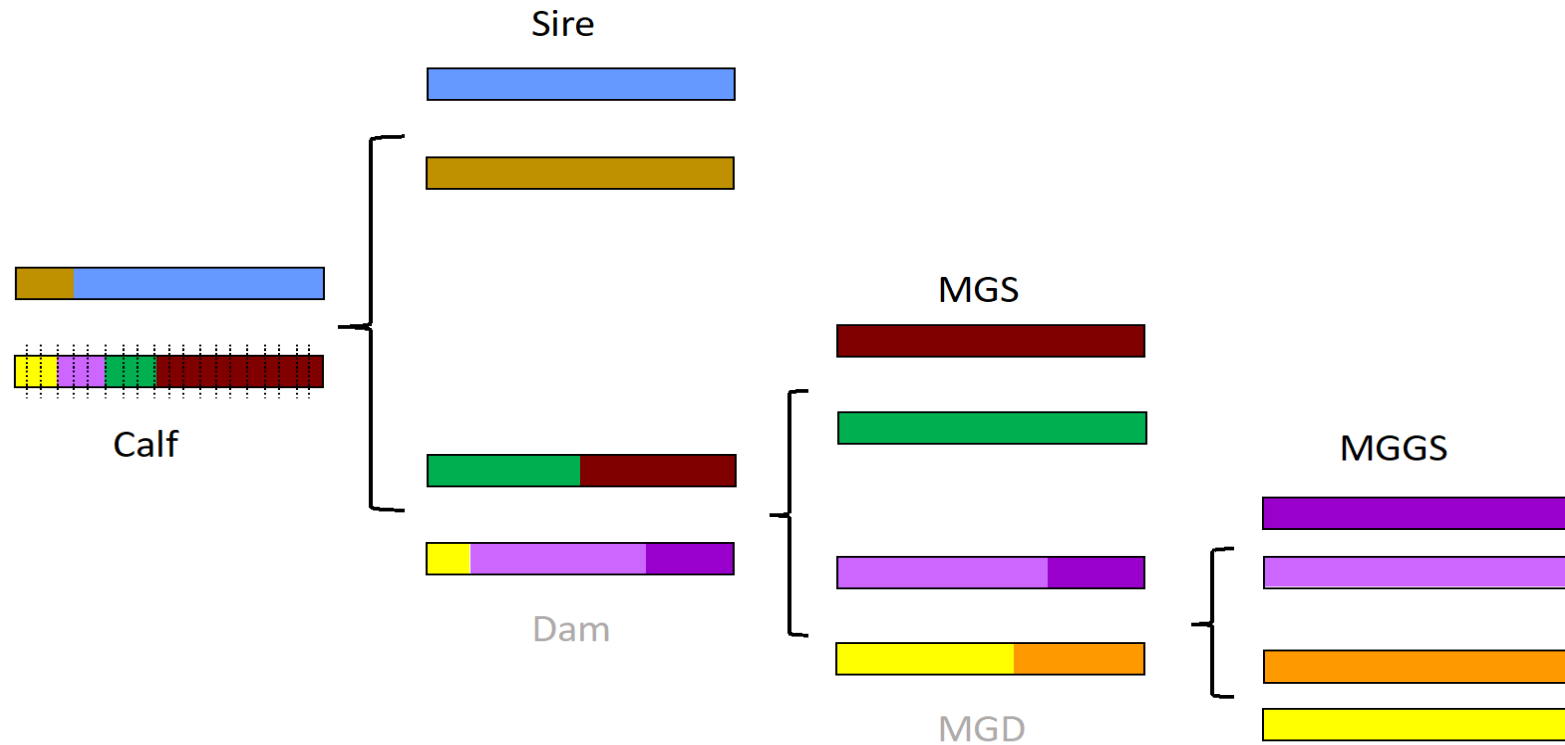
Linking pedigree information with FindDam

Finddam creates the **constructed dam** and/or **MGD ID** to link calves to MGS and MGGS in the pedigree

➢ ~ 1 million animals with no dam ID can be linked to their discovered MGS and MGGS by creating a constructed dam or maternal granddam (MGD) ID

Filling in pedigrees

- **Constructed ID**: **HO USA DAM (MGD) calf ID** numeric
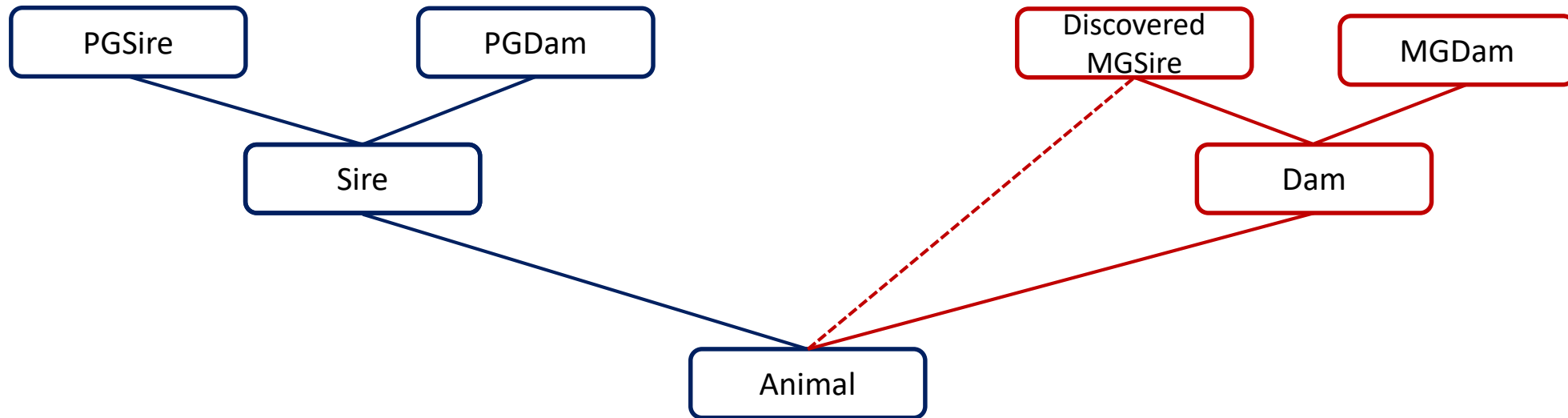
    HOUSA**DAM**004235395

    HOUSA**MGD**004235395

➢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 (only non-ambiguous cases are filled)

> Already discovered and added 400,000 MGS and MGGS whose dam ID and MGD ID were already reported by the owners

MGD known, discovered MGGS added as her sire
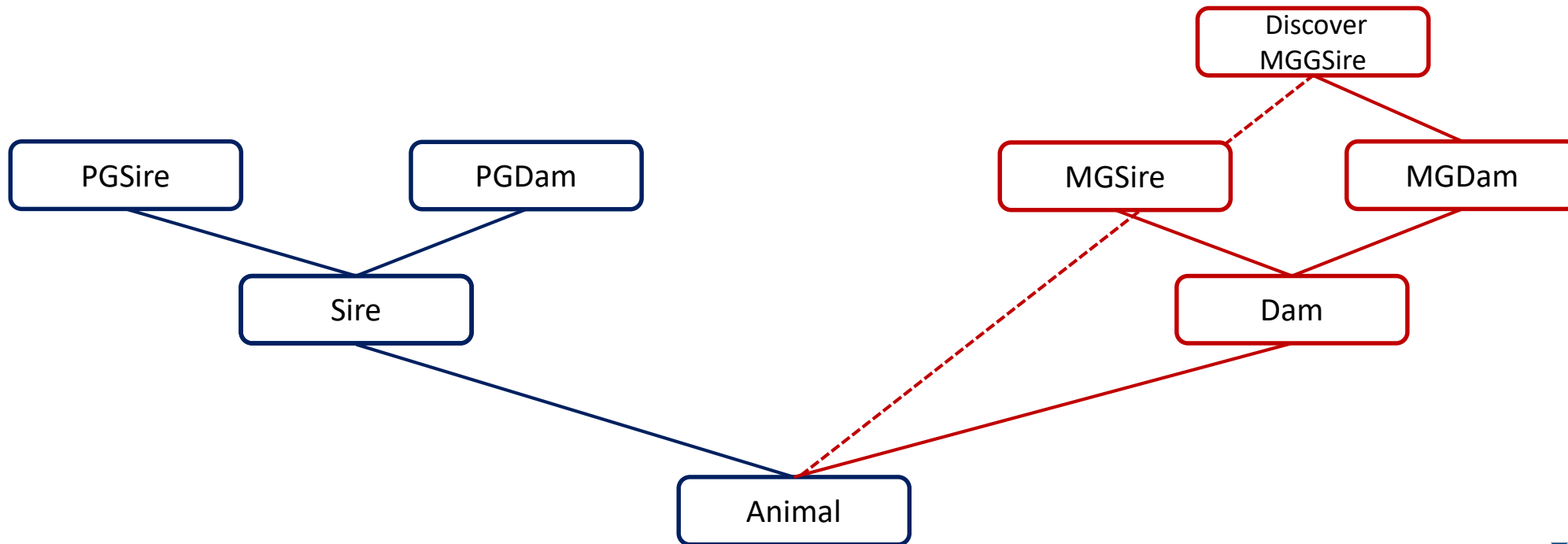
Dam unknown; Dam suggested based on herd, sire, and calving date

HO840000013897236
HO840000013897237
HO840000013897238
HO840000013897239
HO840000013897240

Dam unknown; Dam ID constructed, and pedigree created with discovered MGS as sire

HO840000013897236
HO840000013897237
HO840000013897238
HO840000013897239
HO840000013897240

PGSire

PGDam

MGSire

Sire

Animal

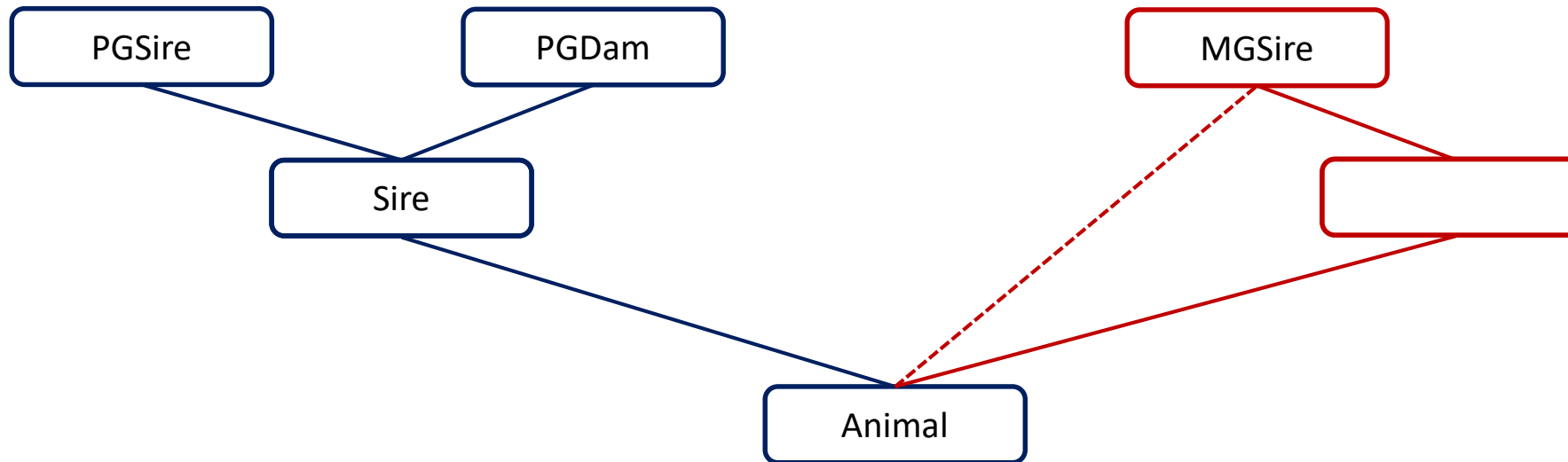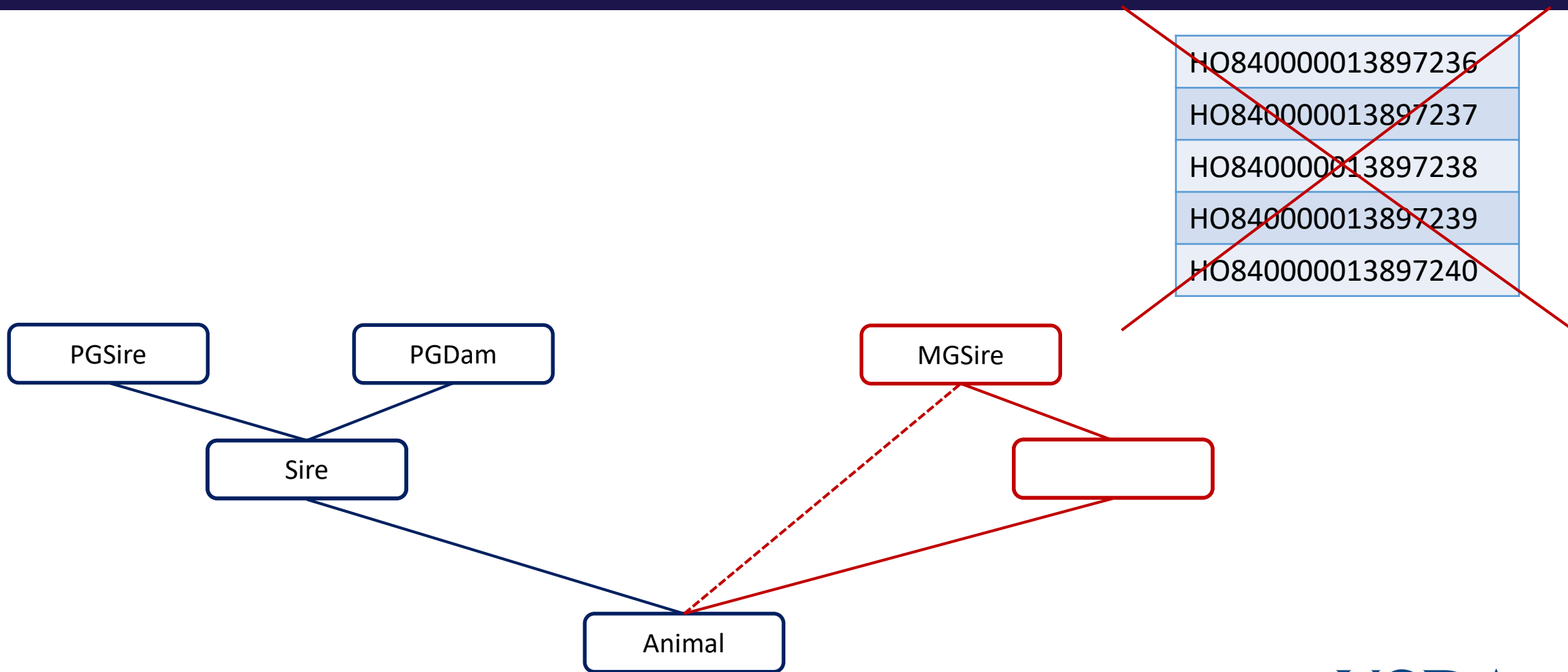# Filled pedigree results

Discovering ancestors and connecting relatives in large genomic databases

J. P. Nani,[1,2]* L. R. Bacheller,[3] J. B. Cole,[1] and P. M. VanRaden[1]
[1]USDA, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD 20705-2350
[2]Estación Experimental Agropecuaria Rafaela, Instituto Nacional de Tecnología Agropecuaria, 22-2300, Rafaela, SF, Argentina
[3]Council on Dairy Cattle Breeding, Bowie, MD 20716

Table 2. Traditional and genomic EBV value means, SD, and reliabilities for yield traits of 295,136 animals with newly found ancestors, before and after pedigree completion

EBV	Trait	Incomplete pedigree			Complete pedigree		
		Mean (kg)	SD (kg)	Reliability (%)	Mean (kg)	SD (kg)	Reliability (%)
Traditional	Milk	1,948	720	26.6	2,064	811	32.6
	Fat	72.2	15.1	25.0	76.7	18.2	32.0
	Protein	59.4	14.6	26.9	63.0	17.0	32.9
Genomic	Milk	2,186	492	76.2	2,258	513	77.1
	Fat	74.9	19.7	76.0	77.6	20.5	76.9
	Protein	63.6	13.4	76.3	65.7	14.1	77.3

Testing accuracy of ancestor discovery

- Genotypes and pedigrees of the dam and MGD were removed to determine how often the correct MGS and MGGS could be discovered within each of 5 breeds

- Haplotypes were reimputed from the reduced data using **Findhap.f90** and the ancestry information was rediscovered using **Fixped.f90** (hide and seek)

- The discovered ancestors were compared to the true ancestors to estimate frequencies of correct detection

- Ancestor discovery was further improved by adjusting the birth year and haplotype sharing limits in **Fixped.f90**

2021 test to delete and rediscover known MGS

Breeds	Genotyped animals	# of samples	Original Fixped MGS Candidates							
			Yes	1st	2nd	3rd	No	Bad	Miss	Extra
Holstein	4605698	70417	63990	4501	770	63	847	193	53	22449
Jersey	553246	7398	6739	451	68	3	107	26	4	1790
Brown Swiss	56706	218	200	7	1	2	5	2	1	28
Ayrshire	12382	135	127	6	1	0	1	0	0	17
Guernsey	6359	66	65	1	0	0	0	0	0	154

Yes	Fixped found the correct ancestor and added it to the pedigree
1st	Fixped found the correct ancestor but did not add it (due to fewer haplotypes shared)
2nd,3rd	The true ancestor was in second or third place and thus not added
No	No ancestor was discovered for the calf
Bad	Fixped added a wrong ancestor to the pedigree (usually a more remote ancestor of the dam)
Miss	An animal was missing from the report
Extra	Refers to other animals not part of the study that changed, primarily from differing imputation

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

USDA

12

2021 test to delete and rediscover known MGSS

Breeds	Genotyped animals	# of samples	Original Fixped MGS Candidates							
			Yes	1st	2nd	3rd	No	Bad	Miss	Extra
Holstein	4605698	10102	6750	2594	400	58	211	83	6	123128
Jersey	553246	1041	721	262	26	5	18	9	0	9608
Brown Swiss	56706	24	16	7	1	0	0	0	0	251
Ayrshire	12382	17	10	6	1	0	0	0	0	113
Guernsey	6359	6	5	1	0	0	0	0	0	283

Yes	Fixped found the correct ancestor and added it to the pedigree
1st	Fixped found the correct ancestor but did not add it (due to fewer haplotypes shared)
2nd,3rd	The true ancestor was in second or third place and thus not added
No	No ancestor was discovered for the calf
Bad	Fixped added a wrong ancestor to the pedigree (usually a more remote ancestor of the dam)
Miss	An animal was missing from the report
Extra	Refers to other animals not part of the study that changed, primarily from differing imputation

2021 Net differences on rediscovering known MGS

MGS Net Differences

Breed	Yes	1st	2nd	3rd	No	Bad	Extra
Guernsey	1	0	0	0	-1	0	0
Ayrshire	3	0	0	0	-3	0	0
Brown Swiss	4	0	0	0	-4	0	0
Jersey	137	14	0	0	-134	-17	0
Holstein	1663	232	17	4	-1657	-259	0

Yes	Fixped found the correct ancestor and added it to the pedigree
1st	Fixped found the correct ancestor but did not add it (due to fewer haplotypes shared)
2nd,3rd	The true ancestor was in second or third place and thus not added
No	No ancestor was discovered for the calf
Bad	Fixped added a wrong ancestor to the pedigree (usually a more remote ancestor of the dam)
Miss	An animal was missing from the report
Extra	Refers to other animals not part of the study that changed, primarily from differing imputation

MGS Net Differences							
Breed	**Yes**	**1st**	**2nd**	**3rd**	**No**	**Bad**	**Extra**
Guernsey	0	0	0	0	0	0	0
Ayrshire	2	0	0	0	-2	0	56
Brown Swiss	1	0	0	0	-1	0	76
Jersey	45	0	0	0	-48	3	3315
Holstein	464	0	0	0	-503	39	51177

Yes	Fixped found the correct ancestor and added it to the pedigree
1st	Fixped found the correct ancestor but did not add it (due to fewer haplotypes shared)
2nd,3rd	The true ancestor was in second or third place and thus not added
No	No ancestor was discovered for the calf
Bad	Fixped added a wrong ancestor to the pedigree (usually a more remote ancestor of the dam)
Miss	An animal was missing from the report
Extra	Refers to other animals not part of the study that changed, primarily from differing imputation

2022 research to further improve Fixped.f90

- **Fixped edit limits revised to increase the pedigree discovery rate**

  ➢ Lower haplotype sharing limits to add more 1st place ancestors

  ➢ Birth year edits revised to allow discovering younger MGS and MGGS

- **Results from revised Fixped code**

  ➢ About 100,000 more total MGS and MGGS were added

  ➢ 2,127 more test animal ancestors were correctly added

  ➢ 240 fewer test animal ancestors were incorrectly added

Summary and implementation

- **Linking disconnected animals improved genetic and genomic estimations**

- **Identifying pedigree errors can help to make better mating decisions**

- Complete pedigrees can assure **accuracy** and **consistency** of genomic and pedigree data

- **Real dam IDs** will be preferred over constructed IDs unless it does not match the calf and grandsire

- Pedigree providers will have an option to remove discovered relationships

- Potential implementation in August 2022

Acknowledgements and disclaimers

- Participating **dairy producers** for supplying data

- **DHI** organizations and **DRPCs** for processing and relaying the information to the Council on Dairy Cattle Breeding (CDCB)

- **Purebred breed associations** for providing pedigree data

- Mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by CDCB

- CDCB is an equal opportunity provider and employer

Thank you!