Adding and reporting genomically discovered ancestors in US evaluations

J. Carrillo[1], P.M. VanRaden[2], E.O.O. Ogwo[2], R.R. Mota[1], G.R. Wiggans[1], L.R. Bacheller[1], G.C. Fok[2], and D.J. Null[2]

[1] Council on Dairy Cattle Breeding, Bowie, MD 20716, USA

[2] U.S. Department of Agriculture, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD 20705-2350, USA

Abstract

Unknown maternal grandsires (MGS) and great grandsires (MGGS) can be discovered accurately based upon haplotype matching. The Council on Dairy Cattle Breeding (CDCB) has already added about 370,000 discovered MGS to dams with unknown sire where no pedigree was submitted for the dam and 30,000 discovered MGGS where no pedigree was submitted for the maternal granddam (MGD). To add MGS or MGGS to the pedigree, where the dam or MGD is unknown, requires creating an ID for the dam or granddam. These constructed IDs consist of the breed of the discovered MGS or MGGS as the best guess of the unknown dam breed, the 'USA' code, the letters 'DAM' or 'MGD' followed by the genotyped animal internal sequence number. For about 30,000 cases, a calf's non-genotyped dam can be discovered by finding a cow in the same herd whose sire is the discovered MGS and has a calving date that matches the calf's birth date. CDCB plans to add > 1 million more discovered MGS and MGGS linked to their genotyped descendants by constructed IDs to the national pedigree in 2022. Accuracy was tested from 2021 data by randomly removing genotypes and pedigrees for dams and MGD with confirmed sires to determine how often the correct MGS and MGGS could be discovered. Within each breed about 92% of true grandsires were automatically filled correctly, about 5% of true grandsires were suggested but not filled, and < 2% of the added ancestors were incorrect. Updated edits further improved accuracy. Discovered ancestors thought to be incorrect can be set back to missing by animal owners. More complete pedigrees will help breeders to avoid inbreeding and improve evaluation accuracy.

Key words: pedigrees, grandsires, genomics, haplotypes

Introduction

For a long time, phenotypic data and pedigrees were all that was required for breeders to assess the breeding values for each cow. As a result, the estimated breeding values were determined using the animal's and its relatives' records. This method required a multitude of records and data eventually lead to more precise estimates of an animal's genetic potential (Nani et al., 2020). The method was ultimately not sustainable, which started the rise of genetic selection and genotyping. However, because genotypes for many animals were not readily available, researchers and dairy breeders were faced with developing statistical methodologies, simulations, and selection processes (VanRaden, 2020). Later, researchers and dairy breeders wanted to verify if the various genomic-based methodologies would hold up under the scrutiny of real-world data.

Therefore, it is critical that dairy breeders and geneticists work together to improve each generation of cows. In order to do so, dairy breeders and geneticists need to have accurate and completed pedigrees especially from the maternal parentage. Missing MGS and MGGS can often be discovered using percentages of haplotypes shared after removing paternal haplotypes in each generation. Accuracy of ancestor discovery was originally tested with 2011 data (VanRaden et al., 2013) and was retested here with 2021 data to account for more markers, different chips, revised imputation, faster generation intervals, and many more candidate grandsires. The detection of MGS and MGGS is pedigree completion technique that aids in making better mating decisions. Finally, improved mating decisions prevent mating close relatives and can help to decrease inbreeding depression and improve fertility and performance among the various cow breeds.

Materials and Methods

Accuracy of ancestor discovery was determined from a random sample of 88,995 calves whose dams, granddams (MGD), MGS, and MGGS were genotyped and already confirmed to have correct parent-progeny relationships. The genotypes and pedigrees of the dam and MGD were removed to determine how often the correct MGS and MGGS could be discovered within each breed: Holstein, Jersey, Brown Swiss, Ayrshire, and Guernsey.

Discovery was further improved by adjusting the birth year and haplotype sharing limits to accept and add more of the first-place candidates because most were correct. The finding of ancestors enabled the correction and completion of pedigrees, increasing the amount of data available for evaluations.

The objective of this model was to find MGS and MGGS for pedigrees with missing dam information and assess their accuracy in five different cattle breeds. Because most historical bulls are genotyped, finding male ancestors in dairy cattle was easier than finding female relatives. The accuracy of ancestry finding tests was improved, which resulted in less genetic prediction bias for animals from unknown parent groups. The ability to link these new animals to the remainder of the pedigree enhanced genetic and genomic forecasts.

Finally, lowering the haplotype limit resulted in a greater number of MGS and MGGS candidates who were more likely to be the correct ancestor (s).

Results & Discussion

In all 5 breeds about 92% of true grandsires were automatically filled correctly, about 5% of true grandsires were suggested but not filled, and < 2% of the added ancestors were incorrect. Most incorrect ancestors were the dam’s MGS instead of her sire or the MGD’s MGS instead of her sire. Counts across all test animals were 78,492 MGS and MGGS correctly added, 7,576 in 1st place but not added, 402 incorrectly added, and 2,525 other cases. The other cases were mostly where the true grandsire was in second place or was less than 2 years older than the dam and no grandsire was automatically added.

The previous version of fixped (fixped 4change) underwent modification by reducing the maxmatch limit from .15 to .13 to add more MGGS and decrease candidates within the “No” and “Bad” columns. Upon further modification to fixped 4change which resulted in fixped 5change, there were improvements that are seen in the net differences. The changes to fixped demonstrated a significant difference when attempting to appropriately locate MGS and MGGS.

Data Tables

Table 1. - The number of MGS candidates found for each breed in its respective categories.

Accuracy of MGS discovery (original Fixped code)										
	Genotyped	**# of Sample**								
Breed	**Animals**	**Animals**	**Yes**	**1st**	**2nd**	**3rd**	**Bad**	**No**	**Miss**	**Extras**
HO	**4,605,698**	**70,417**	**63,990**	**4,501**	**770**	**63**	**193**	**847**	**53**	**22,449**
JE	**553,246**	**7,398**	**6,739**	**451**	**68**	**3**	**26**	**107**	**4**	**1,790**
BS	**56,706**	**218**	**200**	**7**	**1**	**2**	**2**	**5**	**1**	**28**
AY	**12,382**	**135**	**127**	**6**	**1**	**0**	**0**	**1**	**0**	**17**
GU	**6,359**	**66**	**65**	**1**	**0**	**0**	**0**	**0**	**0**	**154**

Table 2. - The number of MGGS candidates found for each breed in its respective categories.

Accuracy of MGGS discovery (original Fixped code)										
	Genotyped	**# of Sample**								
Breed	**Animals**	**Animals**	**Yes**	**1st**	**2nd**	**3rd**	**Bad**	**No**	**Miss**	**Extras**
HO	**4,605,698**	**10,102**	**6,750**	**2,594**	**400**	**58**	**83**	**211**	**6**	**123,128**
JE	**553,246**	**1,041**	**721**	**262**	**26**	**5**	**9**	**18**	**0**	**9,608**
BS	**56,706**	**24**	**16**	**7**	**1**	**0**	**0**	**0**	**0**	**251**
AY	**12,382**	**17**	**10**	**6**	**1**	**0**	**0**	**0**	**0**	**113**
GU	**6,359**	**6**	**5**	**1**	**0**	**0**	**0**	**0**	**0**	**283**

Table 3. - The net difference of MGS candidates found for each breed in its respective categories.

Net differences on MGS discovery from improved Fixped							
Breed	**Yes**	**1st**	**2nd**	**3rd**	**Bad**	**No**	**Extras**
HO	**1,461**	**-701**	**15**	**4**	**-63**	**-716**	**0**
JE	**133**	**-32**	**0**	**0**	**-12**	**-89**	**0**
BS	**5**	**1**	**0**	**0**	**-1**	**-5**	**0**
AY	**3**	**-2**	**0**	**0**	**0**	**-1**	**0**
GU	**2**	**0**	**0**	**0**	**-1**	**-1**	**0**

Table 4. - The number of MGGS candidates found for each breed in its respective categories.

Net differences on MGGS discovery from improved Fixped							
Breed	**Yes**	**1st**	**2nd**	**3rd**	**Bad**	**No**	**Extras**
HO	**300**	**-297**	**-10**	**-5**	**20**	**-7**	**51,302**
JE	**32**	**-32**	**0**	**0**	**4**	**-4**	**3,320**
BS	**1**	**-1**	**0**	**0**	**0**	**0**	**54**
AY	**2**	**-2**	**0**	**0**	**0**	**0**	**55**
GU	**0**	**0**	**0**	**0**	**0**	**0**	**7**

Conclusions

The detection of MGS and MGGS was subsequently improved by changing the fixped program's birth year edit and the findhap program's shared haplotype limit. As a result, the program accepted and recommended the admission of more MGS and MGGS first-place candidates. This automated system has already added hundreds of thousands of MGS for known dams and MGGS for known MGD. It will add > 1.3 million more ancestors for animals with unknown dams or MGD using virtual dam IDs to connect calves to their MGS and/or MGGS. Any discovered ancestors thought to be incorrect can be set back to missing by animal owners. Farmers who breed these cows will gain immensely from the discoveries, as it will help them avoid pairing close relatives, reduce inbreeding depression, and improve performance and fertility in all breeds. Full implementation is expected in 2022.

Acknowledgments

The Council on Dairy Cattle Breeding (Bowie, MD) provided the data for this research under USDA Agricultural Research Service (ARS) Material Transfer Research Agreement 58-8042-8-007. The authors thank S.M. Hubbard (Animal Genomics and Improvement Laboratory, ARS, USDA, Beltsville, MD) for technical review. Funding for G.C. Fok, D.J. Null, E.O.O. Ogwo, and P.M. VanRaden was from USDA-ARS appropriated project 8042-31000-002-00-D, “Improving Dairy Animals by Increasing Accuracy of Genomic Prediction, Evaluating New Traits, and Redefining Selection Goals.” The Council on Dairy Cattle Breeding and USDA are equal opportunity providers and employers.

References

Nani, J., Bacheller, L., Cole, J., & VanRaden, P. (2020). Discovering ancestors and connecting relatives in large genomic databases. *Journal of Dairy Science*, *103*(2), 1729–1734. https://doi.org/10.3168/jds.2019-17580

VanRaden, P. M., Null, D. J., Sargolzaei, M., Wiggans, G. R., Tooker, M. E., Cole, J. B., Sonstegard, T. S., Connor, E. E., Winters, M., van Kaam, J. B., Valentini, A., Van Doormaal, B. J., Faust, M. A., & Doak, G. A. (2013). Genomic imputation and evaluation using high-density Holstein genotypes. *Journal of Dairy Science*, *96*(1), 668–678. https://doi.org/10.3168/jds.2012-5702

VanRaden, P. M. (2020). Symposium review: How to implement genomic selection. *Journal of Dairy Science*, *103*(6), 5291–5301. https://doi.org/10.3168/jds.2019-17684