CDCB Policy

ADDING GENOMICALLY DISCOVERED MATERNAL GRANDSIRES (MGS) AND MATERNAL GREAT GRANDSIRES (MGGS) TO PEDIGREES

This policy for regulating the addition of genomically discovered maternal grandsires and maternal great grandsires to the CDCB pedigree was approved by the CDCB Board of Directors on 08/25/2021. Updated on 11/10/2022.

0. REVISION HISTORY

08/25/2021: Original version approved by the CDCB Board of Directors.
11/10/2022: Updated version including more details on the constructed ID format and an updated implementation phase.

1. POLICY STATEMENT

The Council on Dairy Cattle Breeding (CDCB) recognizes the value of having complete pedigrees to increase the accuracy of the evaluations. Therefore, based on haplotypes matching, missing maternal male ancestors can be discovered and added to the pedigree. The purpose is to have a deeper pedigree in animals that previously had no link to their maternal ancestors. Implementing this on a routine basis will increase the accuracy of the genetic and genomic evaluations.

2. ADDING DISCOVERED MATERNAL GRANDSIRES (MGS) AND MATERNAL GREAT GRANDSIRES (MGGS), SIRES OF KNOWN DAMS OR MATERNAL GRANDDAM

Methodology – Based on haplotype matching, missing maternal male ancestors can be discovered. The CDCB has already added over 370,000 discovered MGS to dams with unknown sire where no pedigree data was submitted for the dam and has extended this to MGGS where no pedigree was submitted for the MGD. To extend this to cases where dam or MGD pedigree was submitted, but their sire was blank, the CDCB adds the MGS or MGGS to the CDCB database

and notifies the source of the addition. This is consistent with our policy of allowing a lower priority pedigree source to fill-in missing parents in a higher source pedigree. The CDCB stores the source of each parent separately from the source code of the animal record. Submitters have the option of rejecting the discovered MGS or MGGS (by submitting a format 1 record to blank the sire with a verification code indicating not to accept a discovered sire for the subject ID). The main goal is to obtain the improved accuracy from the more complete pedigree with the least imposition on the providers. In general, commercial dairy producers rely on genotyping to improve the accuracy of their pedigree but need the notification so they can reject cases where the proposed bull is incorrect.

3. ADDING MATERNAL GRANDSIRES (MGS) AND MATERNAL GREAT GRANDSIRES (MGGS) OF UNKNOWN DAMS OR MATERNAL GRANDDAM

Methodology – To add MGS or MGGS to the pedigree, where the dam or MGD is unknown, requires CDCB to create an ID for the dam or granddam. These constructed IDs must follow the same standards as an actual animal ID, with some added features to obtain a sustainable (constructed) identification system:

- **Unique:** must be unique in the pedigree, as they are only used to link source animals to their ancestors.
- **Traceable**: must be connected to the source animal from which it was derived (and the country that generated it).
- **Stable**: The connection between the constructed ID and its source animal must be perpetual – unless the true ancestor is found.
- **Recognizable**: must be easily recognizable as placeholders and never be considered as the ID of a true ancestor.

The constructed IDs consist of the breed of the discovered MGS or MGGS as the best guess of the breed on the unknown dam, a country code of USA, the letters 'DAM' or 'MGD' followed by the internal sequence number of the genotyped animal, for example HOUSADAM087654321 or JEUSAMGD012345678. In case of international IDs, the only difference will be the 3-letter breed code (HOL instead of HO) and the addition of the sex (F or M) after the country code. To further facilitate the constructed ID recognition, the name of each constructed animal will be "Dam of [ID of source animal]" or "MGD of [ID of source animal]".
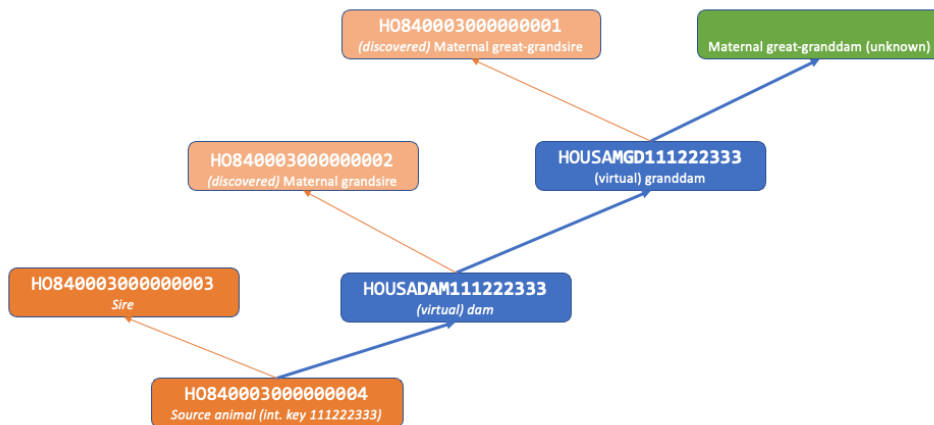
Figure 1. *The source animal HO840003000000004 (CDCB internal key **111222333**) is genotyped and has an incomplete pedigree on the maternal side. The haplotype discovery procedure identifies HO840003000000002 as its maternal grandsire and HO840003000000001 as its maternal great-grandsire. The dam is not known, not genotyped. Two constructed IDs are therefore created for this source animal. Please note: i) the breed of the source animal, ii) USA as country code; iii) DAM/MGD depending on the ancestor they're linking; iv) the numeric part corresponding to CDCB's internal key for the source animal (111222333).*

This design achieves all aforementioned requirements for this ID, by indicating clearly that these are constructed and unique IDs. To control the creation and usage of these IDs, they will ONLY be generated by the CDCB and will have source code "A". Other submitters can modify the record as usual, but with a limitation on constructed IDs. Submitters will only be allowed to replace constructed IDs with a true ancestor ID or delete the constructed ID if they reject the connection to the MGS or the MGGS. However, they will not be allowed to replace constructed IDs with alternative constructed IDs. Once identified, the system will not search for better ancestors unless the discovered ancestor no longer qualifies (for example, if the calf's genotype is reassigned, the – true – dam is identified and genotyped, etc..).

Although CDCB will use "A" source records to add constructed dams, which is a lower priority source, the fact that lower source records can fill missing parents of higher source pedigrees enables this plan to work for the intended purpose of improving the pedigree information.

For about 40,000 cases, a dam can be discovered by finding a cow whose sire is the discovered MGS and has a calving date in the herd of the genotyped animal that matches its birth date. In these cases, a notification will be provided to indicate if the cow was bred to the sire of the genotyped animal around 9 months before the birth. These dams will be added to the pedigree if there is only one such cow discovered, or if only one has a matching service sire. Otherwise, the CDCB will rely on the record submitted by the source, after confirmation by the owner or nominator, in cases where multiple cows qualify.

The notification will be a comma-separated values (csv) file in which the first column is a format 1 record containing a record for animal with the added constructed or discovered dam, or the dam or MGD that added the discovered MGS/MGGS followed by columns with the requester ID, ID of the genotyped animal, its sample ID, the relation of the format 1 animal to the genotyped animal (self, DAM, MGD) and other columns specifically for discovered dams including service sire confirmation and an indication of multiple suggestions

4. IMPLEMENTATION

In 2021 and 2022, CDCB staff successfully implemented the inclusion of discovered MGS and MGGS, sires of **known** dams and MGD (item 2 in this policy). **After** the December 2022 evaluation release, the CDCB will extend the procedure for **unknown** dams and MGD (item 3 in this policy). No constructed ID will be therefore distributed *during* the December 2022 evaluation. Weekly, monthly and triannual evaluations following the December 2022 release will be enhanced by the inclusion of discovered ancestors using the constructed ID methodology.

Test Run – To reassess the impact of the added pedigree on evaluation results, a test run has been performed in October 2022. Over one million animals benefitted from this procedure, which resulted in the creation of nearly 650,000 constructed dam IDs, and over 715,000 constructed granddam IDs. These numbers are approximately 3-fold higher compared to the ones reported by Nani *et al.* in 2019. However, test run results are in line with what was observed by Nani and collaborators. Animals with discovered ancestors (irrespectively if the dam was known or unknown) had the greatest impact: improved estimation of inbreeding and heterosis, more accurate parent averages, increased reliability. On a population level, the impact was extremely contained. AI bulls, which typically already have full pedigrees, were the least affected.

Testing Files – Small files have been provided to the industry before providing the > 1 million MGS and MGGS that have been discovered to service users.

Note regarding calf genotypes – Because calf genotypes often arrive at CDCB before calf pedigrees, industry and farm databases need to ensure that the constructed IDs do not replace the actual dam IDs.

Full database sweep – The inclusion of this methodology to the entire database will require a high level of coordination when the full genomic database pedigrees will be enhanced with constructed IDs. The test run performed in October 2022 revealed that nearly 800,000 animals will require reprocessing if the entire file is implemented in a single solution. Although a single implementation would be ideal from the practical point of view, such a large reprocessing of animals would not be compatible with the CDCB's processing and release schedule1. Therefore, starting January 2023, the constructed IDs will be implemented in batches to allow regular collaborator files being processed during the implementation phase. As a result, the full database sweep and implementation will require a few months to be completely implemented, as new genotypes during monthlies will need to be queued as well.

New genotypes (weeklies) – Starting with the December 10th weekly evaluation, weekly evaluations will start yielding constructed IDs for animals with discovered MGS/MGGS. Once the processing of the file is done after the weekly evaluation is finalized, animals will fully benefit from this enhancement in the following monthly evaluation(s). Since true instead of constructed IDs are preferred, animals from Italy, Great Britain, Germany, and Switzerland (CDDR exchange partners) will not automatically receive a constructed ID. A request for the correct pedigrees will be sent first, and only animals with unknown ancestors (or when no response is received) will

---

[1] A rough estimate by CDCB staff indicates that an 800,000 animals reprocessing would lock the CDCB database for any other "write" operation for approximately at least 15 days, and depending on how many levels of recursion would be necessary.

receive a constructed ID. This means that the modification of these pedigrees will be done a few days later than the rest of the weekly animals.

Monthly evaluations – Once the full database implementation is completed this procedure will also be implemented on monthly evaluations. Animals with assigned constructed IDs will be skipped, unless the discovered ancestor no longer qualifies (see section 2.3.1). In this case, the exchange of pedigrees with CDDR exchange partners happens in the week preceding the monthly calculation (pre-run week), therefore, all constructed IDs will be handled automatically. Similarly to what described in the weekly evaluation, animals will fully benefit from this enhancement in the next monthly evaluation after their pedigree has been modified

5. REFERENCES

Nani, J.P., Bacheller, L.R., Cole, J.B., and VanRaden, P.M. Discovering ancestors and connecting relatives in large genomic databases. J. Dairy Sci. 103(2):1729–1734. 2020.

VanRaden, P.M., Cooper, T.A., Wiggans, G.R., O'Connell, J.R., and Bacheller, L.R. Confirmation and discovery of maternal grandsires and great-grandsires in dairy cattle. J. Dairy Sci. 96(3):1874–1879. 2013.