Savoia, S., Nicolazzi, E., Legarra, A. (2026) Replacing unknown parent groups with metafounders in traditional pedigree-based U.S. dairy genetic evaluations, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 23589: https://doi.org/10.31274/wcgalp.23589 ©2026 Savoia, et al. CC BY-NC-ND

Replacing unknown parent groups with metafounders in traditional pedigree-based U.S. dairy genetic evaluations

S. Savoia[1*], E. Nicolazzi[1], and A. Legarra[1, 2]

[1]*Council on Dairy Cattle Breeding, Bowie, USA;* [2]*University of Georgia, Athens, USA;*
[*]*Corresponding author:* simone.savoia@uscdcb.com

Introduction

Routine dairy cattle genomic evaluations increasingly combine information streams developed under different base assumptions, such as pedigrees and genomic data. As evaluation systems move toward single-step genomic evaluations (Legarra *et al*., 2014), base compatibility becomes a practical concern: the pedigree and genomic relationship structures must be aligned to avoid unintended bias in breeding values and ambiguous interpretation of genetic trends (Bermann *et al*., 2020).

Unknown parent groups (UPG) are widely used in pedigree-based BLUP traditional genetic evaluations to mitigate bias from missing pedigree by assigning animals with unknown parents to groups (based on e.g., birth year, breed, and parent sex) (Quaas, 1988). While effective operationally, the standard UPG approach treats groups as non-inbred and unrelated assuming no relationships among founder populations. This limitation matters when genotypes enter single-step genomic evaluation because the pedigree and genomic bases must be reconciled under a coherent base definition.

Metafounders (MF) extend the UPG concept by treating groups as founders that are related through a covariance matrix Γ, providing a principled route to align pedigree and genomic bases (Legarra *et al*., 2015). However, adopting MF in a production environment raises operational questions: how to define robust groups automatically, how to estimate Γ using routinely available data, and what changes to expect in evaluation results.

The objective of this study was therefore to assess the operational impact of redefining UPG as MF in U.S. pedigree-based traditional genetic BLUP evaluations performed by the Council on Dairy Cattle Breeding (CDCB), motivated by the need for future smoother transitions when incorporating genotypes into single-step genomic evaluation.

Materials & Methods

Pedigree and phenotypic records from the CDCB August 2024 U.S. domestic pedigree-based evaluation were used as the common input dataset across both UPG and MF scenarios. Traits included yield traits (YLD; milk, fat, protein), somatic cell score (SCS), productive life (PL), cow and heifer livability (LIV, HLV), health traits (HTH; hypocalcaemia, displaced abomasum, ketosis, mastitis, metritis, and retained placenta), and gestation length (GL), evaluated in Holstein (HO), Jersey (JE), Brown Swiss (BS), Guernsey (GU), and Ayrshire (AY). Official CDCB edits and trait-specific model specifications were retained.

Official UPG strategy. In the official CDCB system, unknown parents are assigned to UPGs defined by birth year, breed, and pedigree pathway (dam of cow, sire of cow, and parents of sire) in the pedigree-construction step. Adjacent birth years are aggregated when support is limited; year blocks are split only when both the accumulated group and the remaining years exceed 5,000 offspring, restricting splits to birth years ≥1960 to avoid weakly supported groups. HLV and HTH use a reduced number of UPG levels. UPGs are fitted as i.i.d. random effects with variance equal to the additive genetic variance. EBVs are estimated with

1step.f90 software developed by Paul VanRaden (Animal Genomics and Improvement Laboratory, USDA).

New MF strategy. For each trait group, the pedigree was subset to include animals with phenotypic records and all their ancestors without pruning. UPG assignments from the CDCB pipeline were then refined using automated decision rules that quantify group informativeness and merge small adjacent groups (Legarra and Aguilar, 2025), resulting in trait group-specific number of MF levels. Relationships among groups (Γ) were estimated from base allele frequencies inferred during CDCB's routine imputation workflow, using an "increase of relationships" adjustment to better represent stronger relationships within breeds and increasing coancestry in more recent groups (Legarra *et al.*, 2024). EBVs were therefore obtained with blup90iod3 (Lourenco *et al.*, 2022).

EBV comparison. For each trait group and breed combination, EBVs from the CDCB August 2024 U.S. domestic pedigree-based run (UPG) were compared against EBVs obtained from the MF system. Differences in EBVs standard deviation and genetic trends were investigated for bulls (males with known parents and first-lactation daughters in more than 10 herds) and cows (females with known parents and phenotype). Agreement between systems was also assessed by across-animal Pearson correlation. To ensure comparability EBVs from both systems were set to a common genetic base defined by cows born in 2015.

Results & Discussion

Table 1 shows the number of UPGs used in the CDCB August 2024 U.S. domestic pedigree-based run and the corresponding number of levels after redefining those groups as MF in the new pipeline, by trait group.

Table 1. Number of unknown-parent group levels in the August 2024 traditional domestic CDCB evaluations (UPG) and corresponding metafounder levels (MF) after automated refinement, by trait group.

	Trait group[1]						
	YLD	SCS	PL	LIV	HLV	HTH	GL
UPG	335	335	335	335	1	5	335
MF	309	176	308	311	46	96	12

[1] YLD=yield, SCS=somatic cell score, PL=productive life, LIV=cow livability, HLV=heifer livability, HTH=health, GL=gestation length.

Table 1 indicates that redefining UPG as MF did not induce a uniform change across trait groups. In most cases, the automated refinement reduced the number of MF levels relative to the official UPG structure by merging weakly supported adjacent groups. HLV and HTH resulted in a larger number of MF levels, consistent with trait group-specific pedigree subsetting and information available to support distinct groups. Table 2 summarizes EBVs standard deviation changes between evaluation systems, by breed and category.

Across trait groups and breeds, changes in dispersion were small, with ΔSD predominantly negative, indicating slightly smaller EBV standard deviations under MF than under UPG (i.e., greater "shrinkage" under MF). Differences between bulls and cows were broadly similar, suggesting that the observed shifts primarily reflect evaluation base effects rather than category-specific artifacts.

Visual inspection of genetic trends indicated consistent patterns across evaluation systems. For most of the traits, the MF and UPG trends diverged mainly for cohorts born before year 2000 (consistent with greater sensitivity of older cohorts to base assumptions), while trends

were essentially identical for animals born more recently in both bulls and cows. For SCS, trends overlapped almost completely across the full range of birth years. The most notable differences were observed for HTH and HLV, aligning with these traits showing the largest increases in MF group effects relative to UPGs. For HTH, differences were moderate and breed-dependent, consistent with variation in health traits phenotype availability across breeds and time. HLV exhibited the largest differences among major breeds in both categories, indicating higher sensitivity of heifer livability evaluations to how unknown ancestry is modelled.

Table 2. Differences in all-breed EBV standard deviations (ΔSD) between the metafounder (MF) and August 2024 traditional domestic CDCB (UPG) evaluations system, by trait group, category, and breed. Values are expressed in units of genetic SD. For multi-trait groups (yield and health trait groups), ΔSD were computed for each trait and then averaged.

		Breed[3]				
Trait group[1]	Cat.[2]	HO	JE	BS	GU	AY
YLD	Bulls	-0.10	-0.13	-0.12	-0.14	-0.11
	Cows	-0.12	-0.14	-0.13	-0.12	-0.13
SCS	Bulls	-0.02	-0.02	-0.04	-0.04	-0.02
	Cows	-0.02	-0.02	-0.02	-0.04	-0.04
PL	Bulls	-0.04	-0.07	-0.05	-0.08	-0.07
	Cows	-0.05	-0.07	-0.06	-0.07	-0.08
LIV	Bulls	-0.02	-0.03	-0.04	-0.17	-0.04
	Cows	-0.01	0.00	-0.02	-0.16	-0.01
HLV	Bulls	-0.20	-0.07	-0.10	-0.16	-0.19
	Cows	-0.22	-0.05	-0.09	-0.13	-0.07
HTH	Bulls	-0.10	-0.10	-0.16	-	-
	Cows	-0.06	-0.07	-0.10	-	-
GL	Bulls	-0.01	-0.03	-0.04	-0.08	-0.04
	Cows	-0.01	-0.03	-0.05	-0.03	-0.04

[1] YLD=yield (milk, fat, and protein), SCS=somatic cell score, PL=productive life, LIV=cow livability, HLV=heifer livability, HTH=health (hypocalcaemia, displaced abomasum, ketosis, mastitis, metritis, and retained placenta), GL=gestation length; [2] Category: Bulls=males with known parents and first-lactation daughters in more than 10 herds; Cows=females with known parents and phenotypes; [3] Breed: HO=Holstein, JE=Jersey, BS=Brown Swiss, GU=Guernsey, AY=Ayrshire.

Correlations between EBVs from the two systems were essentially 1.00 for YLD, SCS, PL, LIV, and GL across breeds and for both categories, indicating that replacing UPG with MF largely preserved animal rankings in these traits. For the remaining trait groups, correlations remained high but were lower overall, with HLV ranging from 0.89 to 0.98 and HTH from 0.87 to 1.00. The very low heritability of these traits (HLV < 0.01; HTH 0.01–0.03), together with the relatively recent availability of phenotypes for HLV (since 2009), likely contributed to their greater sensitivity to the definition of unknown-parent structure under MF.

Figure 1. Genetic trends for yield traits in Holsteins (MILK = milk yield, FAT = fat yield, PROT = protein yield) under the UPG and MF models. Bulls were defined as males with known parents and first-lactation daughters in more than 10 herds, and cows as females with known parents and phenotypic records. Cows born in 2015 were used as genetic base. UPG = Unknown Parent Groups; MF = metafounders.

Conclusions

Replacing UPG with MF proved to be a practical and methodologically coherent option for CDCB's U.S. domestic pedigree-based evaluations. The results indicate that MF can be implemented without disruption while maintaining continuity with current genetic interpretation and strengthening the foundation for transition to single-step genomic evaluation. Differences were most evident for the HLV and HTH trait groups, where higher MF levels relative to UPG coincided with greater sensitivity in genetic trends and cohort-based investigations. Accordingly, additional work is warranted for low-heritability traits to better characterize the definition of MF and their impact on trend estimation and evaluation diagnostics.

Acknowledgements

The authors acknowledge participating dairy producers for data supply; Dairy Herd Improvement organizations and Dairy Records Processing Centers for processing and relaying information to CDCB; purebred breed associations for providing pedigree data. The CDCB is an equal opportunity provider and employer.

References

Bermann M., Lourenco D., and Misztal I. (2020) J. Dairy Sci. 104:2027-2031. doi.org:10.3168/jds.2020-18969

Lourenco D., Tsuruta S., Aguilar I., *et al*. (2022) Recent updates in the BLUPF90 software suite. Proc. 12th WCGALP, Rotterdam, Netherlands.

Legarra A., Christensen O.L., Aguilar I., *et al*. (2014) J. Liv. Sci. 166:54-65. doi.org:10.1016/j.livsci.2014.04.029

Legarra A., Christensen O.F., Vitezica Z.G., *et al*. (2015) Genetics 2:455–468. doi.org:10.1534/genetics.115.177014

Legarra A., Bermann M., Mei Q., *et al*. (2024) Genet. Sel. Evol. 56:35. doi.org:10.1186/s12711-024-00892-9

Legarra A., and Aguilar I. (2025) J. Dairy Sci. 108:1747-1756. doi.org:10.3168/jds.2024-25408

Quaas R.L. (1988) J. Dairy Sci. 71:1338-1345. doi.org:10.3168/jds.S0022-0302(88)79691-5