



Effect of two phantom parent grouping strategies on the genetic evaluation of growth traits in Mexican Braunvieh cattle



Luis Antonio Saavedra-Jiménez ^a

Rodolfo Ramírez-Valverde ^a

Rafael Núñez-Domínguez ^{a*}

Agustín Ruíz-Flores ^a

José Guadalupe García-Muñiz ^a

Mohammad Ali Nilforooshan ^b

^a Universidad Autónoma Chapingo. Departamento de Zootecnia, Posgrado en Producción Animal. Km. 38.5 Carretera México-Texcoco. 56230, Chapingo, Estado de México. México.

^b University of Otago, Department of Mathematics and Statistics, Dunedin, New Zealand.

*Corresponding author: rafael.nunez@correo.chapingo.mx

Abstract:

The study aimed to compare two grouping strategies for unknown parents or phantom parent groups (PPG) on the genetic evaluation of growth traits for Mexican Braunvieh cattle. Phenotypic data included birth (BW), weaning (WW) and yearling (YW) weights. Pedigree included 57,341 animals. The first strategy involved 12 PPG (G12) based on the birth year of the unknown parent's progeny and the sex of the unknown parent, while the second involved 24 PPG (G24) based on the birth year of the unknown parent's progeny and 4-selection pathways. The animal models included fixed effects and the random direct additive genetic effect; WW also included random maternal genetic and maternal permanent environmental effects. Product-moment correlations between EBV from G0 (no PPG) and G12 were 0.96, 0.77 and 0.69 for BW, WW and YW, respectively, and between EBV from G0 and G24 were 0.91, 0.54, and 0.53, respectively. Corresponding rank correlations between G0 and G12 were 0.94, 0.77, and 0.72, and between G0 and G24 were 0.89, 0.61, and 0.60. Genetic trends showed a base deviation from the genetic

trend of G0, except for BW of G12. The results did not support the use of the two grouping strategies on the studied population and traits, and further research is required. Introducing PPG to the model, enough phenotype contribution from descendants to PPG, and avoiding collinearity between PPG and fixed effects are important. Genetic groups should reflect changes in the genetic structure of the population to the unknown parents, including different sources of genetic materials, and changes made by selection over time.

Key words: Braunvieh cattle, EBV, Genetic groups, Rank correlation, Unknown parents.

Received: 27/11/2019

Accepted: 23/11/2020

Introduction

Mexican Braunvieh is a dual-purpose breed of cattle. Since June 2003, national genetic evaluations for growth traits have been undertaken for this breed in Mexico⁽¹⁾. Like in any livestock population, there are unknown parents in the pedigree. Unknown parents are assumed to be unrelated, non-inbred, and to have a single descendant. Unknown parents might correspond to base animals in the first generation or spread over generations. They affect genetic progress in several ways: (i) reducing selection intensity for animals with unknown parents, (ii) parentage uncertainty decreases the accuracy of genetic evaluations, (iii) miss-identification of parents yields both biased estimated breeding values (EBV) and heritability estimates⁽²⁾. Best linear unbiased prediction (BLUP) regresses genetic merit predictions of animals to unknown parents of mean zero. Depending on the genetic background, and the generation to which unknown parents belong to, their expected genetic merit could be different from zero. Quass⁽³⁾ established a methodology for considering phantom parent groups (PPG) or genetic groups in BLUP. Although PPG are not of interest *per se*, they are considered to facilitate modeling and computation⁽⁴⁾. Furthermore, along with statistical correction for non-random missing pedigree information, PPG enables direct estimation of quantitative genetic parameters⁽⁵⁾. Because there are no specific rules for determining PPG, its definition is mainly based on the researcher's criteria, but it usually includes a time component⁽⁶⁾. Other factors commonly considered in grouping strategies are the sex of the parent or selection intensity^(4,7,8). All descendants of an individual with PPG contribute to the estimation of the PPG effect⁽⁵⁾, so having PPG with an equal number of individuals is unlikely to affect the animal model's ability to estimate the PPG effect with acceptable precision. However, any strategy for assigning unknown parents to PPG should reflect the average genetic level of unknown parents⁽⁹⁾.

Due to the inclusion of PPG in the model, Theron *et al*⁽⁷⁾ observed a significant change and a reduction of bias in the genetic trend of milk yield for South African Holsteins. Similarly, a reduction in EBV bias was detected by including PPG in the genetic analyses for weaning, post-weaning and yearling weights, scrotal circumference, and muscling score in Nelore cattle⁽¹⁰⁾. The purpose of this study was to compare two strategies of grouping unknown parents to PPG on the genetic evaluation of growth traits in Mexican Braunvieh cattle.

Material and methods

Data

Pedigree and phenotypic records on Mexican Braunvieh cattle were obtained from *Asociación Mexicana de Criadores de Ganado Suizo de Registro* (Mexico City). Phenotypic records were birth (BW), weaning (WW), and yearling weights (YW) from animals born between 1985 and 2017, in 229 farms across Mexico. Weaning and yearling weights were adjusted to 240 d and 365 d of age, respectively, according to the procedure proposed by the Beef Improvement Federation⁽¹¹⁾. Records outside the mean \pm 3 SD range for the trait of interest were not included in the analyses. Also, WW and YW records outside 240 ± 45 d and 365 ± 45 d age were excluded from the analyses, respectively. The pedigree was extracted (parents), starting from animals with an available phenotype (for any of the three traits), and limited to animals born since 1970. Final pedigree included 57,341 individuals, 18,689 males, 38,652 females, 2,746 sires, and 27,015 dams.

Contemporary groups were formed considering herd, year, and season of birth (rainy or dry). Records from contemporary groups with less than four animals were excluded from the analyses. Table 1 shows the final number of records and descriptive statistics for each trait.

Table 1: Descriptive statistics for growth traits in the Mexican Braunvieh population

Trait	N	Minimum	Mean \pm SD	Maximum
Birth weight	31,654	23.00	38.11 \pm 4.84	53.00
Weaning weight	21,333	100.59	235.07 \pm 42.85	372.62
Yearling weight	14,439	146.66	324.07 \pm 56.07	504.84

Genetic analyses

The genetic analyses comprised estimation of genetic parameters and BLUP⁽¹²⁾ for the Mexican Braunvieh population, using the following single-trait models:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{u} + \mathbf{e}, \quad (1)$$

for BW and YW, and

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{u} + \mathbf{Z}_2\mathbf{m} + \mathbf{Z}_3\mathbf{mpe} + \mathbf{e}, \quad (2)$$

for WW, where \mathbf{y} , \mathbf{b} , \mathbf{u} , \mathbf{m} , \mathbf{mpe} , and \mathbf{e} are vectors of phenotypic records, fixed effects, direct additive genetic, maternal additive genetic, maternal permanent environmental, and residuals effects, respectively. \mathbf{X} , \mathbf{Z}_1 , \mathbf{Z}_2 , and \mathbf{Z}_3 are incidence matrices relating records to \mathbf{b} , \mathbf{u} , \mathbf{m} , and \mathbf{mpe} , respectively. The fixed effects were:

$\mathbf{b}_{\text{BW}} = [\text{sex, Braunvieh purity, age of dam, (age of dam)}^2, \text{birth contemporary group}]$

$\mathbf{b}_{\text{WW}} = [\text{sex, Braunvieh purity, age of dam, (age of dam)}^2, \text{pre-weaning contemporary group, milk feeding condition}]$

$\mathbf{b}_{\text{YW}} = [\text{sex, Braunvieh purity, post-weaning contemporary group, post-weaning feed}]$

There were 1,778, 1,450, and 1,038 birth contemporary groups, pre-weaning contemporary groups, and post-weaning contemporary groups, respectively. Milk feeding conditions were suckling without milking, suckling with additional milking, and feeding with a milk substitute. Post-weaning feed regimes were grazing, semi-confined, and total confinement. The sex ratios were close to 1. Age of dam at calving had a minimum, mean, SD, and maximum of 1.70, 6.64, 3.04, and 17.00 yr, respectively. Braunvieh purity had a minimum, mean, SD, and maximum of 0.88, 0.99, 0.01, and 1.00, respectively. It was used the official models for the evaluation of the studied traits in Mexican Braunvieh cattle. The (co)variance structures were:

$$\text{Var} \begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{A}\sigma_u^2 & 0 \\ 0 & \mathbf{I}_N\sigma_e^2 \end{pmatrix},$$

for BW and YW, and

$$\text{Var} \begin{pmatrix} \mathbf{u} \\ \mathbf{m} \\ \mathbf{pe} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{A}\sigma_u^2 & 0 & 0 & 0 \\ 0 & \mathbf{A}\sigma_m^2 & 0 & 0 \\ 0 & 0 & \mathbf{I}_{Na}\sigma_{mpe}^2 & 0 \\ 0 & 0 & 0 & \mathbf{I}_N\sigma_e^2 \end{pmatrix},$$

for WW, where \mathbf{A} is the pedigree-based additive genetic relationship matrix, \mathbf{I}_{Nd} and \mathbf{I}_N are identity matrices of order equal to the number of dams and observations. σ_u^2 , σ_m^2 , σ_{mpe}^2 , and σ_e^2 are the direct additive genetic, maternal additive genetic, maternal permanent environmental, and residual variances, respectively. Applying PPG, the term $\mathbf{Z}_1\mathbf{Qg}$ is added to the model (Eq. [1] and Eq. [2]), where \mathbf{g} is the vector of PPG effects, and \mathbf{Q} is

the matrix relating animals to PPG. Variance components were obtained without PPG, based on derivative-free REML, using the MTDFREML software⁽¹³⁾.

Genetic groups

Evaluation of the genetic grouping strategies was carried out through the comparison of EBV from BLUP with and without PPG. Criteria used to form unknown parents' groups were:

- 1) Year of birth: Year of birth of the unknown parent was five years before the year of birth of its progeny. Unknown parent birth years were grouped into six classes: 1965-69, 1970-74, 1975-79, 1980-84, 1985-89, and 1990-96.
- 2) Sex of the unknown parent.
- 3) Selection pathway (sire of sire, sire of dam, dam of sire, and dam of dam).

The two genetic grouping strategies were:

G12: Class of birth year (6 levels) × sex of the unknown parent (2 levels).

G24: Class of birth year (6 levels) × pathway of selection (4 levels).

Genetic groups based on criteria such as sex of missing ancestor or paths of selection allow evaluation of different genetic selection differentials⁽⁴⁾. Likewise, the inclusion of the year of birth category allows us to model the genetic improvement over time^(3,7). Table 2 shows the number of unknown parents in each PPG for each strategy.

Table 2: Criteria and frequency of unknown parents in phantom parent groups

Strategy ¹	Unknown parent	Year group ²					
		1965-1969	1970-1974	1975-1979	1980-1984	1985-1989	1990-1996
G12	Sire	540	513	820	941	678	433
	Dam	647	457	664	891	564	35
G24	Sire of sire	119	58	72	90	87	143
	Sire of dam	421	455	748	851	591	290
	Dam of sire	145	57	51	84	73	9
	Dam of dam	502	400	613	807	491	26

¹ Phantom parent group with 12 (G12) and 24 (G24) levels.

² Progeny's birth year – 5.

Without including PPG in the model, the mixed model equations were (for BW and YW):

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}, \quad (3)$$

where $\lambda = \sigma_e^2/\sigma_u^2$. Adding the effect of PPG to the model, the mixed model equations become⁽³⁾:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{Z}\mathbf{Q} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\lambda & \mathbf{Z}'\mathbf{Z}\mathbf{Q} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{X} & \mathbf{Q}'\mathbf{Z}'\mathbf{Z} & \mathbf{Q}'\mathbf{Z}'\mathbf{Z}\mathbf{Q} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{y} \end{bmatrix}. \quad (4)$$

Incorporating PPG effects into the genetic merit of animals (*i.e.*, $\mathbf{EBV} = \hat{\mathbf{u}} + \mathbf{Q}\hat{\mathbf{g}}$) can be made directly in the mixed model equations, using Quaas and Pollak⁽¹⁴⁾ transformation that involves absorption of PPG equations, which gives⁽³⁾:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{0} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\lambda & -\mathbf{A}^{-1}\mathbf{Q}\lambda \\ \mathbf{0} & -\mathbf{Q}'\mathbf{A}^{-1}\lambda & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q}\lambda \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} + \mathbf{Q}\hat{\mathbf{g}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{0} \end{bmatrix}. \quad (5)$$

This procedure avoids the extra step of calculating $\hat{\mathbf{u}} + \mathbf{Q}\hat{\mathbf{g}}$ after Eq. [4], and the need for re-creating matrix \mathbf{Q} , which is computationally expensive. Quaas and Pollak⁽¹⁴⁾ transformation is not implemented in MTDFREML software. Therefore, Eq. [3, 4] were applied for BLUP with and without PPG, respectively (additional terms of the maternal genetic and maternal permanent environmental effects were involved for WW). Estimated breeding values accounting for PPG ($\hat{\mathbf{u}} + \mathbf{Q}\hat{\mathbf{g}}$) were obtained using functions “qmat” and “Qgpu” from R package “gggroups”⁽¹⁵⁾, where the matrix of PPG contributions to individuals in a pedigree (\mathbf{Q}) was calculated, and PPG contributions ($\mathbf{Q}\hat{\mathbf{g}}$) were added to the genetic merit of animals ($\hat{\mathbf{u}}$), with $\hat{\mathbf{g}}$ and $\hat{\mathbf{u}}$ obtained from MTDFREML⁽¹³⁾.

Grouping strategy comparisons

Comparisons between grouping strategies were made by:

Pearson product-moment and Spearman rank correlations between EBV obtained with and without PPG.

Genetic trends obtained for each analysis, by averaging EBV per birth year.

Results and discussion

There were 3,925 animals with unknown sire, 3,258 animals with unknown dam, and 2,430 animals with both unknown sire and dam. Unknown parents were assigned to 12 or 24 PPG (G12 and G24; Table 2). Variance components obtained with and without PPG are shown in Table 3. Estimates of parameters for the studied traits under different scenarios were closely similar. Thus, the model choice should not interfere with the estimation of genetic parameters.

Table 3: Variance components for birth weight (BW), weaning weight (WW) and yearling weight (YW) for the Mexican Braunvieh population estimated with 12 (G12), 24 (G24) and without (G0) phantom parent groups

Strategy	Trait	σ_u^2	σ_m^2	σ_{mpe}^2	σ_e^2
G0	BW	2.69	-	-	8.54
	WW	87.76	8.80	23.12	435.85
	YW	86.27	-	-	692.96
G12	BW	2.69	-	-	8.53
	WW	83.14	8.43	23.06	436.58
	YW	81.30	-	-	695.01
G24	BW	2.71	-	-	8.52
	WW	90.27	10.09	21.37	435.37
	YW	85.72	-	-	692.52

σ_u^2 = additive genetic variance, σ_m^2 = maternal genetic variance, σ_{mpe}^2 = maternal permanent environmental variance, σ_e^2 = residual variance.

Theron *et al*⁽⁷⁾ reported that the inclusion of PPG has a minor influence on the estimation of co(variance) components, also, Shiotsuki *et al*⁽¹⁶⁾ showed that the use of a relationship matrix that includes genetic groups does not generate differences in the variance estimates contrasted with the use of a matrix without genetic groups. In some studies^(7,8,10) variance components were obtained considering a "control" model, which did not include genetic groups, and those variance components were used in predicting breeding values in a model including PPG, similar to the procedure applied in this research.

Descriptive statistics for EBV obtained with models including or excluding PPG are shown in Table 4. In a basic animal model, the existence of a single genetic group is assumed⁽¹⁷⁾. Given that breeding values are deviations from the genetic group mean, all values in the base population have an expectation of zero^(5,17). Genetic group methodology allows to assign genetic effects to multiple groups within the base population, which could have a different mean⁽⁵⁾. The EBV including genetic groups considers that each individual inherits the mean of the effects in the genetic group of their parents plus the mean of the genetic value of their parents; therefore, the expectation of EBV for the population is not zero^(5,17) because the assumption of breeding values distribution is not met. Assigning unknown parents to PPG with a possibly non-zero average of genetic merit would change their descendants' EBV. Consequently, the expected mean of the EBV obtained after considering genetic groups change to the product \mathbf{Qg} ⁽³⁾.

Table 4: Descriptive statistics of estimated breeding values for growth traits obtained with 12 (G12), 24 (G24) and without (G0) phantom parents groups for Mexican Braunvieh cattle

Strategy	Trait	Minimum	Mean \pm SD	Maximum
G0	BW	-5.36	0.03 \pm 0.79	5.13
	WW	-26.67	-0.17 \pm 3.87	25.32
	YW	-24.06	0.32 \pm 3.39	24.34
G12	BW	-5.24	0.31 \pm 0.83	5.37
	WW	-18.73	14.19 \pm 5.20	39.77
	YW	-50.60	-7.77 \pm 4.81	18.08
G24	BW	-3.67	2.06 \pm 0.88	7.57
	WW	-53.15	-10.75 \pm 8.17	32.23
	YW	-49.13	31.10 \pm 7.27	82.12

BW= Birth weight, WW= Weaning weight, YW= Yearling weight.

Pearson product-moment and Spearman rank correlations between EBV with and without PPG are shown in Table 5. Correlation coefficients between EBV without PPG (G0) and G12 were higher than those of G0 and G24, for all the traits and groups of animals (i.e., males and females, with and without phenotype). The correlations were lower for animals without phenotype than with phenotype, and lower for females than males. Generally, the correlations were higher for BW than for WW, and higher for WW than for YW (Table 5).

It has been proposed that correlation coefficients between EBV lower than 0.90 could change the ranking of animals for genetic evaluation⁽¹⁸⁾. Estimates of correlation coefficients obtained here suggest possible changes in the ranking mainly for WW and YW. Petrini *et al*⁽⁸⁾ also remarked changes in the rank for WW due to the inclusion of PPG (Pearson and Spearman correlation estimates ranged from 0.50 to 0.70). On the other hand, the inclusion of PPG resulted in small changes in the ranking for BW in this study. Results for BW agree with what was observed for milk production⁽⁷⁾, YW, and post-weaning weight gain⁽¹⁶⁾, scrotal circumference, or muscle score⁽⁸⁾. Rank changes are due to the shifts that genetic groups make in the EBV of their descendants.

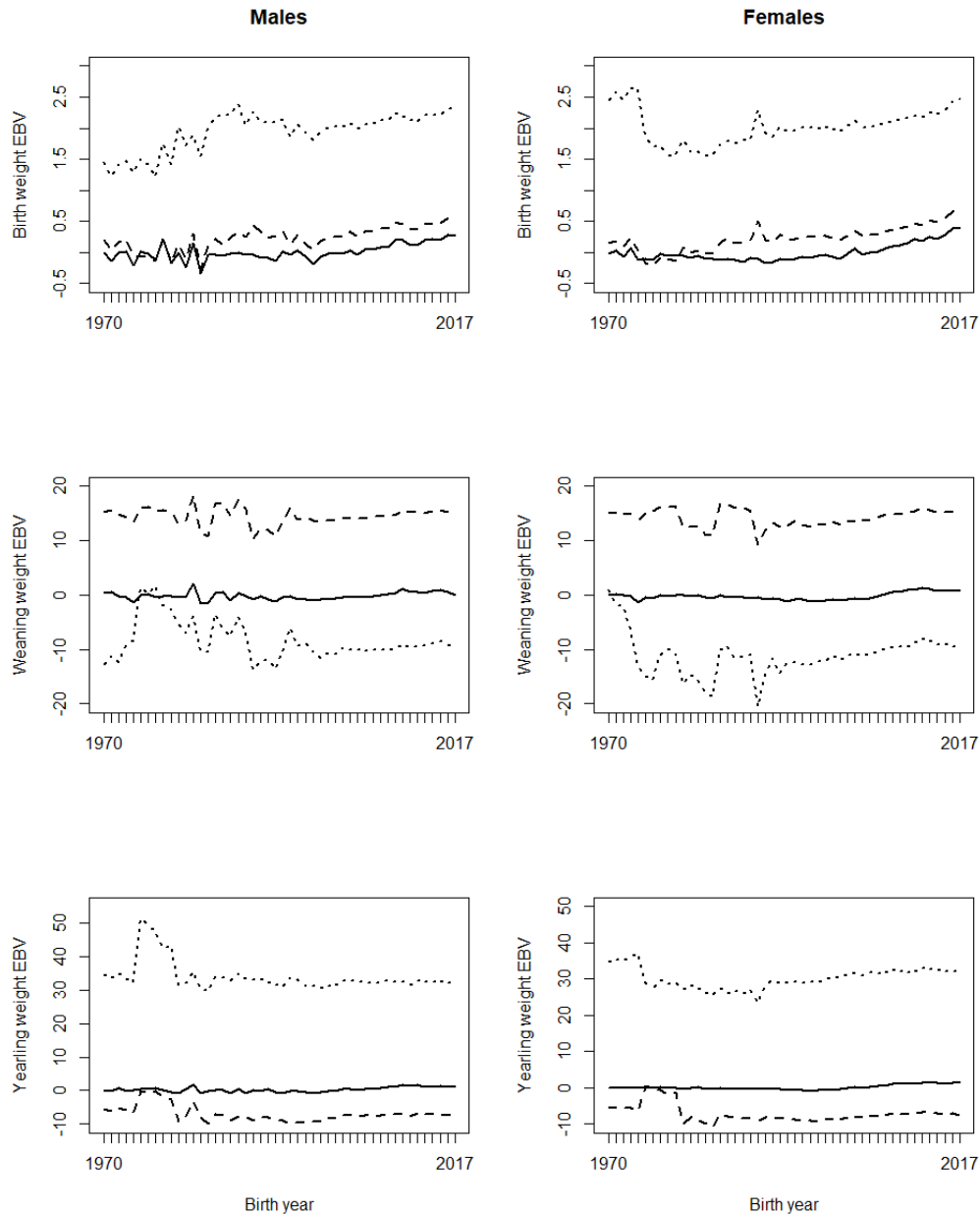
Figure 1 illustrates the effect of PPG on genetic trends. The trends were relatively similar for males and females. For BW, G12 increased the slope of the genetic trend, compared to G0; G24 also increased the slope of the genetic trend, but it showed a base deviation from G0. For WW, both G12 and G24 showed large fluctuations in the early years. Genetic trends from G12 and G24 showed positive and negative base differences with G0, respectively (Figure 1). For YW, the genetic trend of G24 showed a large base deviation from the G0 and G12 genetic trends (Figure 1). Generally, if a genetic trend does not pass over zero, it indicates a base problem for EBVs. Therefore, G24 is ruled out for BW and YW, and G12 is ruled out for WW, since G24 does not cross zero for BW and YW, and G12 does not cross zero for WW (Figure 1). On the other hand, a robust

grouping strategy is expected to perform well for different traits⁽¹⁶⁾, as a trait-specific grouping strategy would be a burden for routine genetic evaluations.

Table 5: Pearson (and Spearman) correlation coefficients between estimated breeding values without phantom parent groups (EBV_G0), with 12 phantom parent groups (EBV_G12), and 24 phantom parent groups (EBV_G24), in the Mexican Braunvieh population

Correlation type	Trait		
	Birth weight	Weaning weight	Yearling weight
Total animals	n = 57,341	n = 57,341	n = 57,341
r(EBV_G0, EBV_G12)	0.959 (0.942)	0.766 (0.766)	0.692 (0.717)
r(EBV_G0, EBV_G24)	0.912 (0.891)	0.538 (0.610)	0.535 (0.605)
Males with phenotype	n = 15,810	n = 10,748	n = 7,384
r(EBV_G0, EBV_G12)	0.988 (0.982)	0.914 (0.886)	0.853 (0.846)
r(EBV_G0, EBV_G24)	0.975 (0.964)	0.786 (0.763)	0.743 (0.737)
Males without phenotype	n = 2,879	n = 7,941	n = 11,305
r(EBV_G0, EBV_G12)	0.941 (0.923)	0.796 (0.797)	0.719 (0.760)
r(EBV_G0, EBV_G24)	0.861 (0.830)	0.606 (0.627)	0.587 (0.636)
Females with phenotype	n = 15,844	n = 10,585	n = 7,055
r(EBV_G0, EBV_G12)	0.986 (0.979)	0.901 (0.879)	0.844 (0.840)
r(EBV_G0, EBV_G24)	0.972 (0.960)	0.752 (0.749)	0.710 (0.743)
Females without phenotype	n = 22,808	n = 28,067	n = 31,597
r(EBV_G0, EBV_G12)	0.895 (0.877)	0.635 (0.659)	0.596 (0.634)
r(EBV_G0, EBV_G24)	0.799 (0.781)	0.407 (0.491)	0.445 (0.515)

Figure 1: Genetic trends of growth traits for BLUP (EBV (solid line)), BLUP with 12 phantom parent groups (EBV_G12 (dashed line)), and BLUP with 24 phantom parent groups (EBV_G24 (dotted line)), with six classes of birth year considered in phantom parent groups



However, in practice, a grouping strategy may perform well for a trait, but not well for another trait, especially if the fixed effects are different for the two traits⁽⁸⁾. Possible problems with PPG implementation are likely to be due to collinearity or confounding between PPG and fixed effect^(3,5). Considering PPG as random effect is a solution to this problem.

The effect of inclusion of PPG on the genetic trend has been variable. Theron *et al*⁽⁷⁾ showed that including PPG in genetic evaluation had a drastic effect on the genetic trend

for milk production traits, having a higher response (almost double) when PPG was included. Besides, Shiotsuki *et al*⁽¹⁶⁾ observed higher genetic trends for post-weaning weight and YW when the model included PPG. In contrast, PPG inclusion in genetic analyses for WW, scrotal circumference, and muscling score showed a lower genetic trend than when PPG was not included⁽⁸⁾. It could be concluded that the effectiveness of PPG on genetic evaluations depends on the population structure, studied traits and criteria adopted to define PPG. It has been proposed that the definition of PPG should balance between the complexity of genetic groups and the representation of genetic differences⁽⁸⁾. Also, genetic groups should consider the selection criteria adopted by breeders.

Two possible reasons are considered for the problems observed with the genetic trends: the shortage of progeny phenotypes supporting the inference of some PPG, and possible confounding or collinearity between PPG and the fixed effects in the model, especially contemporary groups. Figure 2 shows the frequency of animals, missing sires, and missing dams across birth years, and Figure 3 shows the frequency of phenotypes per birth year. It can be interpreted that there were not enough phenotypes supporting the prediction of PPG solutions in early years (unknown parents born before 1990, i.e., their progeny born before 1995). The phenotypes' contribution decreases as the number of generations between the PPG and the phenotyped descendant increases, lower with lower heritability.

Figure 2: Number of animals (solid line), unknown sires (dashed line) and unknown dams (dotted line) per year of birth

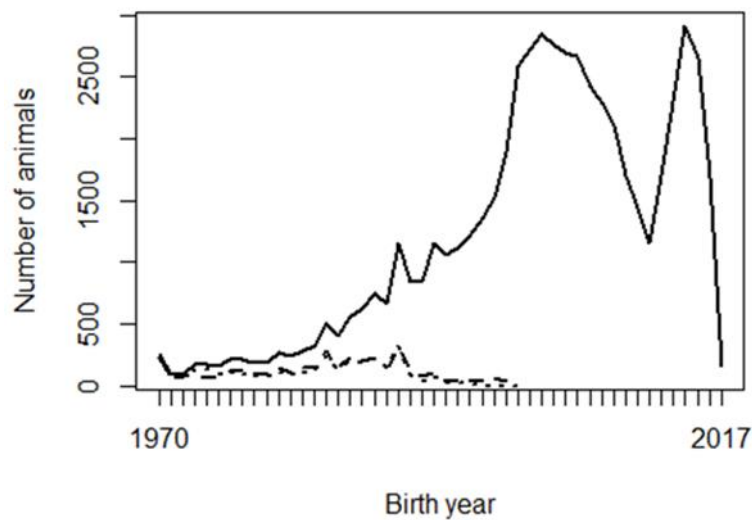


Figure 3: Number of birth weight (solid line), weaning weight (dashed line), and yearling weight (dotted line) phenotypes per year of birth

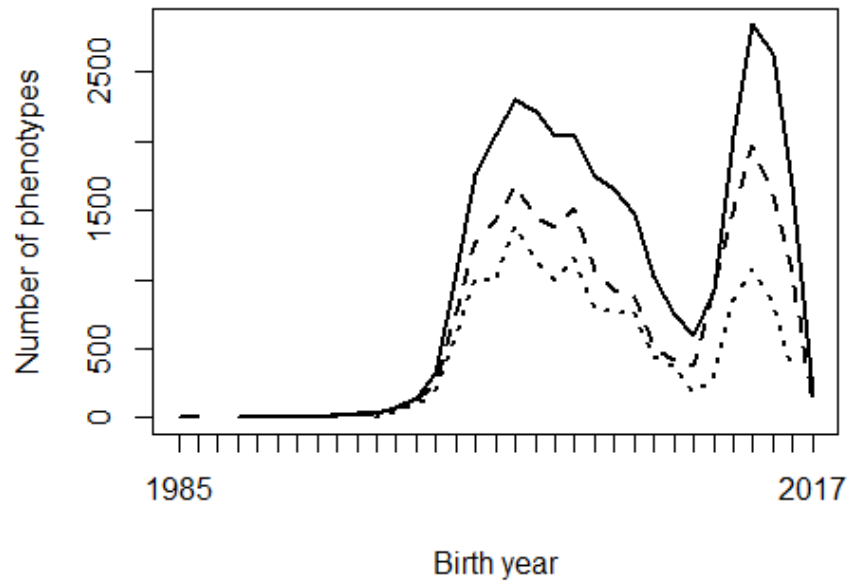


Figure 2 shows that the studied population was not in real need of PPG in the animal model, or PPG could be limited to only a few groups, so that phenotypes from distant generations of descendants could support the estimation of those few PPG. Animal models with PPG are more beneficial to populations with a higher and broader prevalence of missing pedigree information, especially if different genetic backgrounds (e.g., imported genetic materials) or different selection strategies/pressure are involved in different groups of animals (e.g., males vs. females or different selection pathways). The genetic trends (Figure 1) show that the population has not been under efficient selection, and there is an excellent opportunity for genetic improvement toward sustainable production in Mexican production systems and environments. Figures 2 and 3 also show data collection problems between years 2003 and 2014, and between 2014 and 2017. Data completeness and correctness are essential for accurate and reliable genetic evaluations. As mentioned in the Data subsection, pedigree (parentage) was extracted, starting from phenotyped animals. The number of animals and missing parents (Figure 2) were higher without this restriction. However, in that case, there were extra missing parents with no contribution from progeny performance; therefore, no information to make inferences upon them. It is recommended to extract pedigree from phenotyped animals, making decisions about forming genetic groups, then adding animals that had not been extracted and assigning their unknown parents to the existing PPG.

Ideally, there should be fixed group connectedness between different PPG (i.e., like the concept of genetic connectedness among fixed groups (levels of a fixed effect)). In other words, phenotypes from different groups of fixed effects should contribute information to different PPG. It has even been recommended to form (some) PPG composite of both sires and dams⁽⁸⁾. The R package “gggroups”⁽¹⁵⁾ allows to perform PPG from both sexes. Similar definitions of PPG and some fixed effects may cause collinearity. In that situation,

the number of fixed-effect groups contributing information to each PPG decreases. One way of checking the collinearity within and between fixed effects and PPG is checking the minimum eigenvalue of $[\mathbf{X} \ \mathbf{Q}_p][\mathbf{X} \ \mathbf{Q}_p]$, where \mathbf{Q}_p is \mathbf{Q} with rows limited to phenotyped animals. Estimability problems for PPG in the model are not limited to this study. Such problems are often observed due to confounded effects (collinearity) between PPG⁽¹⁹⁾. Even, reducing such confounding by changing the composition of PPG, there might be confounding between PPG and other fixed effects. Those estimability problems were removed and estimated breeding values look normal by considering PPG as random effects via adding σ_e^2/σ_u^2 to the diagonals of the PPG equations in the animal model⁽¹⁹⁾.

Conclusions and implications

Two strategies of grouping unknown parents to PPG (G12 and G24) were tested on BW, WW, and YW in Mexican Braunvieh cattle. The two strategies used the most common criteria for defining PPG (birth year of the progeny, sex of the unknown parent for G12, and selection pathway for G24). Genetic trends had an offset deviation from BLUP's genetic trend without PPG, except for BW of G12. Also, including PPG in the model may have caused collinearity between PPG and some fixed effects. The shortage of phenotypes supporting the solutions for some PPG effects was another reason for the lack of benefit from the two grouping strategies on the studied population and traits. It is recommended to define PPG based on a subset of pedigree, in which parents are connected to phenotyped descendants, then adding the rest of animals and assigning their unknown parents to the existing PPG, to avoid an excessive and unnecessary number of genetic groups. More important than the number of progeny per PPG or equal year intervals defining PPG, is the amount of phenotype contributions for predicting PPG effects. It is recommended to have less overlap between PPG definitions and fixed effects to reduce collinearity between them.

Acknowledgments

The authors thank *Consejo Nacional de Ciencia y Tecnología (CONACyT, México)* for the financial support to the first author during his Doctoral studies. The authors also thank to *Asociación Mexicana de Criadores de Ganado Suizo de Registro* for allowing the use of their data, and Dr. Dale Van Vleck for providing help and support with MTDFREML software.

Literature cited:

1. AMCGSR. Asociación Mexicana de Criadores de Ganado Suizo de Registro. Resumen de evaluaciones genéticas para ganado Suizo Americano 2017. Techn Bull. Mexico. 2018.

2. Van Vleck LD. 1970. Miss-identification in estimating the paternal sib correlation. *J Dairy Sci* 1970;53:1469-1474.
3. Quaas RL. Additive genetic model with groups and relationships. *J Dairy Sci* 1988;71:1338-1345.
4. Westell RA, Quaas RL, Van Vleck LD. Genetic groups in an Animal Model. *J Dairy Sci* 1988;71:1310-1318.
5. Wolak ME, Reid JM. Accounting for genetic differences among unknown parents in microevolutionary studies: How to include genetic groups in quantitative genetics animal models. *J Anim Ecol* 2017;86:7-20.
6. Fikse F. Fuzzy classification of phantom parent group in an animal model. *Genet Sel Evol* 2009;41:1-8.
7. Theron HE, Kanfer FHJ, Rautenbach L. The effect of phantom parent groups on genetic trend estimation. *South African J Anim Sci* 2002;32:130-135.
8. Petrini J, Pertile SF, Eler JP, Ferraz JB, Mattos EC, Figueredo LG, *et al.* Genetic grouping strategies in selection efficiency of composite beef cattle (*Bos taurus* x *Bos indicus*). *J Anim Sci* 2015;93:541-552.
9. Pollak EJ, Quaas RL. Definition of group effect in sire evaluation models. *J Dairy Sci* 1983;66:1503-1509.
10. Oliveira Junior GA, Eler JP, Ferraz JBS, Petrini J, Mattos EC, Mourão GB. Definição de grupos genéticos aditivos visando melhor predição de valores genéticos em bovinos de corte. *Rev Bras Saúde Prod Anim* 2013;14:277-286.
11. BIF. Beef Improvement Federation. Guidelines for Uniform Beef Improvement Programs. 9th ed. North Carolina State University, Raleigh, NC, USA. 2018.
12. Henderson CR. Best linear unbiased and prediction under a selection model. *Biometrics* 1975;31:423-447.
13. Boldman KG, Kriese LA, Van Vleck LD, Van Tassell CP, Kachman SD. A manual for use of MTDFREML, a set of programs to obtain estimates of variances and (co)variances. Washington, DC: USDA, ARS. 1995.
14. Quaas RL, Pollack EJ. Modified equations for sire models with groups. *J Dairy Sci* 1981;64:1868-1872.
15. Nilforooshan MA, Saavedra-Jiménez LA. ggroups: a R package for pedigree and genetic groups data. *Hereditas* 2020;157:17.
16. Shiotsuki L, Cardoso FF, Silva JAIV, Albuquerque LG. Comparison of a genetic group and unknown paternity models for growth traits in Nellore cattle. *J Anim Sci* 2013;91:5135-5143.

17. Van Vleck LD. Breeding value prediction with maternal genetic groups. *J Anim Sci* 1990;68:3998-4013.
18. Crews DH, Franke DE. Heterogeneity of variances for carcass traits by percentage Brahman inheritance. *J Anim Sci* 1998;73:1803-1908.
19. Schaeffer LR. Necessary changes to improve animal models. *J Anim Breed Genet* 2018;135:124-131.