Animal (2011), **5:3**, pp 337–346 © The Animal Consortium 2010
doi:10.1017/S1751731110002144

animal

A definition of unknown parent groups based on bull usage patterns across herds

A. Bouquet[1,2†], G. Renand[1] and F. Phocas[1]

[1]*INRA, UMR1313 Génétique Animale et Biologie Intégrative, F-78352 Jouy-en-Josas, France;* [2]*AgroParisTech, UMR1313 Génétique Animale et Biologie Intégrative, F-75231 Paris Cedex 05, France*

(Received 26 February 2010; Accepted 27 July 2010; First published online 29 October 2010)

In genetic evaluations, the definition of unknown parent groups (UPG) is usually based on time periods, selection path and flows of foreign founders. The definition of UPG may be more complex for populations presenting genetic heterogeneity due to both, large national expansion and coexistence of artificial insemination (AI) and natural service (NS). A UPG definition method accounting for beef bull flows was proposed and applied to the French Charolais cattle population. It assumed that, at a given time period, unknown parents belonged to the same UPG when their progeny were bred in herds that used bulls with similar origins (birth region and reproduction way). Thus, the birth period, region and AI rate of a herd were pointed out to be the three criteria reflecting genetic disparities at the national level in a beef cattle population. To deal with regional genetic disparities, 14 regions were identified using a factorial approach combining principal component analysis and Ward clustering. The selection nucleus of the French cattle population was dispersed over three main breeding areas. Flows of NS bulls were mainly carried out within each breeding area. On the contrary, the use and the selection of AI bulls were based on a national pool of candidates. Within a time period, herds of different regions were clustered together when they used bulls coming from the same origin and with an estimated difference of genetic level lower than 20% of genetic standard deviation (σ_g) for calf muscle and skeleton scores (SS) at weaning. This led to the definition of 16 UPG of sires, which were validated as robust and relevant in a sire model, meaning numerically stable and corresponding to distinct genetic subpopulations. The UPG genetic levels were estimated for muscle and SS under sire and animal models. Whatever the trait, differences between bull UPG estimates within a time period could reach 0.5 σ_g across regions. For a given time period, bull UPG estimates for muscle and SS were generally larger by 0.30 to 0.75 σ_g than those of cows. Including genetic groups in the evaluation model increased the estimated genetic trends by 20% to 30%. It also provoked re-ranking in favor of bulls and cows without pedigree.

Keywords: beef cattle, unknown parents, genetic evaluation, genetic groups, genetic trend

Implications

In genetic evaluation models, genetic groups are used to better account for differences in the genetic level between animals without pedigree, also called founders. Usual definition criteria for genetic groups are birth year, country of origin and sex of founders. In the case of a heterogeneous beef cattle population, we showed that a finer group definition, which also accounts for the artificial insemination rate and region of the birth herd of founders, is important to assess genetic disparities in the founder population. This improves both the quality of adjustment to data of the evaluation model and the ranking of candidates for selection.

† E-mail: alban.bouquet@gmail.com

Introduction

In beef cattle populations, breeding objectives can differ according to the large variety of production systems across and within breeds. In France, regions are characterized by different beef cattle production systems, which correspond to different breeding objectives. The relative use of artificial insemination (AI) and natural service (NS) may also greatly vary between herds and regions. These different selection and mating practices may generate genetic heterogeneity in the population of any beef cattle population. Ignoring such heterogeneities in the base population may lead to biased prediction of breeding values (BV) and suboptimal choice of breeding stock. The best linear unbiased prediction (BLUP) of BV is truly unbiased when pedigree and data information are exhaustive to trace back to an unselected and

https://doi.org/10.1017/S1751731110002144 Published online by Cambridge University Press

unbred population. Only under such an assumption, it makes sense to use a genetic evaluation model with a unique base population whose BV are normally distributed with a zero mean and a variance equal to twice the Mendelian variance (Henderson, 1973). However, this assumption is violated in practical evaluation of a population undergoing selection for decades. To account for the genetic trend and heterogeneity in the base population, unknown parent groups (UPG) may be defined. As suggested by Westell *et al.* (1988) and Banos *et al.* (1991), group definition is often based on birth year, country of origin and selection path. However, for large beef cattle populations, those criteria may not be fully sufficient to define UPG. As far as we know, no methodology has been proposed to define relevant UPG at the national level for large heterogeneous populations. This study presents a procedure to define UPG for beef cattle genetic evaluations. The method is based on the characterization of bull usage patterns across herds.

Material and methods

Data

The French Charolais cattle were analyzed as a case study in order to define genetic groups in a large beef cattle population. This breed has had great expansion in France since the 1960s, potentially creating genetic heterogeneous subpopulations across French regions. Data were extracted from the French national cattle database, which is used for on-farm genetic evaluations of weaning traits. The analyzed population included 2 936 732 calves sired by 98 848 bulls. Animals were born between 1972 and 2006 in 8157 herds having at least five calves recorded at weaning per year.

Two traits were analyzed: muscle score at weaning (MS, mean = 63.0 points, s.d. = 11.9 points) and skeleton score at weaning (SS, mean = 64.3 points, s.d. = 9.7 points). Both of them are scored from the visual appraisal over a 100-point scale. The MS is a sum of scores for five different items in order to assess the calf muscularity (muscle thickness and width) of its shoulders, back and rump. The SS is a sum of scores for five other items to assess the calf skeletal frame (bone thickness and length) of its back, pelvis and canon bone. Statistical analyses were carried out with version 9.1 of the SAS software (SAS Institute Inc., Cary, NC, USA, 2004). The ASREML software (Gilmour *et al.*, 2002) was used to estimate variance components and log-likelihoods of models. The Genekit software (V. Ducrocq, personal communication) was used to perform BLUP evaluations.

Identification of genetic heterogeneity sources in the population

Three factors, which could cause genetic heterogeneity in the population, were identified: time, AI rate and region of herds.

Characterization of time periods. Since the beginning of the French beef cattle recording system in 1972, important political or technical events have led to strong changes in French beef cattle breeding schemes over time. Four periods were considered to analyze the data. The first one occurred from 1972 to the setting up of milk quotas in 1983. This latter event boosted the countrywide expansion of the Charolais breed during the second period running from 1983 to 1991. The third time period (1992 to 1998) started with the setting up of the BLUP animal model methodology for on-farm genetic evaluations. The last time period (1999 to 2006) was characterized by an increase in the unknown parent rate. Indeed, quite a lot of new herds entered the performance recording system at this period and more stringent validation procedures of genealogies were set up in 1999 (Journaux *et al.*, 2006). These phenomena led to an increase in the annual rate of unknown parents from about 2% to 5%.

Characterization of the relative use of AI and NS. In France, about one-third of Charolais calves are procreated by AI in herds with performance recording (Bouquet *et al.*, 2009). To account for genetic differences between AI and NS bulls due to different selection intensities, herds in each region were separated according to the rate of calves sired by AI bulls: herds with an AI rate below 50% (NS herds) and above or equal to 50% (AI herds).

Characterization of regional subpopulations. Given the great variety of production systems in France, there was evidence of regional genetic heterogeneity in the Charolais population (Bougler *et al.*, 1973). Unfortunately, no information about production systems was recorded. Several regions were therefore identified using data available for genetic evaluations. An approach combining a principal component analysis (PCA) and a clustering algorithm was chosen to automate the classification of administrative departments into production regions.

Normalized PCA was carried out to synthesize 22 parameters that were means of diverse Charolais herd characteristics per administrative department. Parameters are listed in Table 1 and are related to herd performance at birth and weaning of calves, herd demography, breeding and calving practices. Departments were then aggregated in regions presenting common features using a Ward hierarchical ascendant clustering (HAC) algorithm on the PCA results. This approach combining PCA and Ward HAC was applied to the four time periods, described in the previous paragraph, in order to detect temporal evolutions of regional beef cattle production and breeding practices. Within a time period, analyses were carried out only on departments, which counted at least five herds and 1000 animals to avoid inconsistent classification due to insufficient number of records. For instance, the last period included 59 departments, representing 1 274 304 animals born in 4514 herds. The automatic clustering of departments allowed identifying four large regions with similar calf performance and breeding practices. To describe more exhaustively all potential regional genetic heterogeneities in the population, a finer definition of regions had been adopted on the basis of three *ad hoc* division criteria. The first criterion was to form regions

https://doi.org/10.1017/S1751731110002144 Published online by Cambridge University Press

Table 1 *Means, standard deviations and extrema of 22 Charolais herd parameters used in the PCA of average Charolais herd records from 59 French departments in the period 1999 to 2006*

Specification	Abbreviation		Mean	s.d.	Minimum	Maximum
Performance	BW_mean	Mean birth weight (kg)	46.4	0.92	44.2	48.0
	WW_mean	Mean weaning weight (kg)	277.9	8.80	253.2	300.5
	MS_mean	Mean muscle score points	60.3	2.14	55.6	66.0
	SS_mean	Mean skeleton score points	62.0	1.77	57.0	67.2
Demography	Herd_size	Mean herd size	36.5	10.8	18.41	78.3
	HB_rate	Rate of herds subscribing to the herd book	0.24	0.19	0.03	0.76
	UP_rate	Rate of unknown parents	0.05	0.03	0.00	0.15
Breeding	AI_rate	Rate of calves bred by AI sires	0.38	0.14	0.12	0.71
	No_AI_herd	Rate of herds with an AI rate equal to zero	0.12	0.09	0.00	0.38
	High_AI_herd	Rate of herds with an AI rate over 70%	0.22	0.17	0.00	0.65
	SL_AI_rate	AI rate with AI bulls progeny tested for slaughter value	0.48	0.07	0.24	0.64
	BR_AI_rate	AI rate with AI bulls progeny tested for breeding value	0.38	0.06	0.21	0.66
	EM_AI_rate	AI rate with AI bulls progeny tested for early muscularity	0.01	0.02	0.00	0.11
	Private AI_rate	AI rate with AI bulls not progeny tested	0.08	0.08	0.00	0.49
Calving	Parity_mean	Mean parity of cows	3.43	0.21	3.04	4.07
	Early_calv	Rate of heifers calving before 30 months	0.06	0.05	0.01	0.23
	Standard_calv	Rate of heifers calving between 34 to 39 months	0.72	0.09	0.47	0.86
	Late_calv	Rate of heifers calving after 40 months	0.05	0.03	0.02	0.15
	Fall_calv	Rate of fall calving (October to December)	0.33	0.15	0.06	0.69
	Winter_calv	Rate of winter calving (January to March)	0.44	0.11	0.24	0.69
	Spring_calv	Rate of spring calving (April to June)	0.18	0.07	0.05	0.35
	Summer_calv	Rate of summer calving (July to September)	0.05	0.05	0.00	0.22

PCA = principal component analysis; AI = artificial insemination.

with at least 100 000 registered animals. The second one was to consider additional information on the selection history of the French Charolais breed related by Bougler *et al.* (1973) and experts of the Charolais breed association. The third one was to constitute regions composed of close departments to conserve a geographical consistency.

Definition of UPG

The definition of UPG aims at capturing genetic differences existing between genetically homogenous subpopulations of unknown parents, which were used in the same period (Robinson, 1986; Westell *et al.*, 1988). However, a too fine definition of UPG is not desirable because a sufficient number of animals in each group is necessary to get precise estimates of UPG effects and to avoid confounding with other fixed effects in the model (Quaas, 1988). In practice, Interbull guidelines for international evaluation of dairy traits recommend a minimum of 10 to 20 bulls per group (Interbull, 2001). Although the introduction of UPG in a genetic evaluation model may reduce bias, it also increases prediction error variance of BV due to the higher number of parameters in the statistical model. If differences between UPG estimates are not large enough, Kennedy (1981) proved that a model without UPG could be preferred to a model with groups for minimizing the mean square error.

To reduce the number of UPG defined for the beef cattle evaluation in a relevant way, we postulated that two herds with a similar supply of identified bulls should have used unknown parents with similar genetic levels. These unknown parents were therefore supposed to be drawn for the same genetic group. The supplying strategy in cows was not considered to define UPG because cow flows across herds were negligible in comparison to bull flows. Moreover, differences between BV of cows were expected to be smaller than the ones between BV of bulls, due to the higher selection intensity on the sire selection path.

Within each time period, founders born in herds from different combinations of 'AI rate × region' were assigned to a same UPG when their birth herds had used at least 70% of bulls with similar genetic levels and origins. The origin of a NS bull was defined by the region and the degree of specialization in breeding of the birth herd. Breeding herds were split into two categories: specialized herds, which sold at least five males for breeding each year and occasional breeders, which sold less than five bulls per year to other herds carrying out performance recording. A third category was defined for AI bulls because they were submitted to higher selection intensity than NS bulls.

Hence, within each time period, sire origin effects were estimated for each trait (MS and SS) as fixed effects in a linear model including also the birth region, herd, birth year and sex of the recorded calves with known sires. These effects provided an estimation of the mean genetic level of AI and NS bulls according to their origin. Then, for each trait, the mean genetic level of sires used in herds of each combination of 'AI rate × region' was calculated. To do so, estimates of sire origin were weighted by their contributions to the total number of progeny of those herds. In practice, differences in estimated genetic

https://doi.org/10.1017/S1751731110002144 Published online by Cambridge University Press

level of identified sires had to be lower than 20% of genetic standard deviation (σ_g) for both traits to consider that founders born in different herd categories belonged to the same UPG. This arbitrary value was fixed as an *ad hoc* limit corresponding to the minimum genetic difference below which Phocas and Laloë (2004a) considered that including groups in the genetic evaluation of a progeny testing scheme will be counterproductive in terms of selection response and average mean square error of BV.

In the construction of the genealogy file, three categories of animals had to be considered to define UPG: bulls, cows and calves not yet used for reproduction. All calves belonging to the evaluated population and not yet used for reproduction had systematically their dam, birth date and herd known. When a bull or a cow had an unknown parent, this parent was replaced with the UPG corresponding to the criteria relative to the year, the region and the AI rate of the herd in which it was used. When one of the criteria was missing, it was deducted according to the following rules. In case of a founder bull, it was determined if it was used by AI. Such information was available in a database, which listed all bulls used by AI since the 1960s. The birth year of a founder was derived from the average birth year of its offspring and a generation interval of 8 years for AI bulls, 4 years for NS bulls and 5 years for cows. As birth dates of a founder's offspring could be unknown, an iterative procedure was used to predict the birth year of all ancestors from the controlled calf population to the founders. Genetic groups were cross-classified by sex to account for bull and cow selection paths. The same UPG definition criteria were adopted to define bull and cow groups. A cross-classification with the sex of the unknown parent would have accounted for the usual four selection paths, but it would have induced confounding effects between sire and dam UPG because both parents were unknown in 92% of cases of incomplete pedigree. Eventually, when the birth herd of a founder was unknown, the region was derived from the national identification number of the founder, which contains the number of the birth department, and the AI rate was assumed to be below 50% because NS is the main reproductive mode in the French Charolais cattle population (Bouquet *et al*., 2009).

Estimation of UPG effects and BV

We proceeded to a genetic evaluation without and with genetic groups to evaluate the impact of introducing such groups in the model. Without UPG, the following mixed linear model was considered for each trait:

$$y = \alpha a + Xb + Zu + Wc + e \qquad (1)$$

where y is the vector of observations, a is the age at scoring taken into account as a covariate, α is the corresponding regression coefficient, b is a vector of fixed effects, u, c and e are random vectors representing BV, permanent environmental (dam) and residual effects, respectively. X, Z and W are the corresponding incidence matrices.

The b vector considered the fixed effects of status at scoring, particular treatments before weaning, birth season, age of dam and contemporary group. Status at scoring counted three levels: weaned, just weaned and not weaned calves. Particular treatments before weaning also corresponded to three levels: no treatment, favorable and unfavorable treatments. The 10 classes of the birth season effect corresponded to each single month from October to May and to the association of June and July on the one hand and August and September on the other hand. Classes of dam age had 10 levels as follows: classes 1, 2, 3, 4, 5 corresponded to primiparous dams with age at first calving, respectively, strictly lower than 27 months, between 27 and 30 months, 31 and 33 months, 34 and 39 months and, finally, over 39 months. Classes 6, 7, 8, 9 and 10 corresponded to multiparous cows whose age belonged, respectively, to the interval 34 to 39 months, 40 to 52 months, 53 to 65 months, 66 to 126 months and over 126 months. Contemporary groups were defined as combinations of birth herd, birth year, management group and sex of the calves. The random vector of BV included either sire or animal genetic effects assumed to be normally distributed in a sire and animal model, respectively. Genetic relationships among sires, or among animals, took into account up to three generations of ancestors, including 195 238 and 3 310 940 animals, respectively. Dam permanent environmental effects were taken into account but no maternal genetic effects were considered due to small maternal heritabilities for SS and MS (Phocas and Laloë, 2004b).

In the model with UPG, genetic groups were estimated as fixed effects according to Westell *et al*. (1988). The vector of total BV $\hat{\mathbf{u}}$ is predicted as $\hat{\mathbf{u}} = Q\hat{\mathbf{g}} + \hat{u}^*$ where $\hat{\mathbf{g}}$ is a vector containing the UPG effects, Q is a matrix whose rows represent the gene proportion of the recorded animal originating from each UPG and $\hat{\mathbf{u}}^*$ is the vector of BV within UPG effects. Keeping the same matrix notations, the mixed model for y can be written as:

$$y = \alpha a + Xb + ZQg + Zu^* + Wc + e \qquad (2)$$

In a sire model, the BV of a sire ($\hat{s}$) is equal to half of its additive genetic value, which would be predicted in an animal model ($\hat{s} = 0.5\hat{a}$). The UPG estimates in a sire model are therefore equal to half of UPG effects estimated in an animal model.

Variance components were estimated under each sire model (1) and (2) on 10 data samples of 300 herds. Whatever the model, average phenotypic variances were estimated to 62.8 points^2 and 97.4 points^2, respectively, for SS and MS. The corresponding estimates of heritability ($\mathbf{h}^2$) were 0.19 and 0.26. The ratio ($\mathbf{c}^2$) of permanent environmental variance to phenotypic variance also remained equal to 0.05 and 0.11, respectively, for SS and MS. Therefore, variance components used in the evaluation models of the whole data set were fixed at the herein values.

Validation of the UPG definition

The definition of large and homogenous genetic groups is necessary to ensure that UPG estimates will be accurate and

https://doi.org/10.1017/S1751731110002144 Published online by Cambridge University Press

relatively constant between two evaluation runs, avoiding inappropriate re-ranking of animals (De Jong, 2003; Fikse, 2003). The UPG definition will be declared relevant if it is fine enough to model correctly genetic heterogeneity among the founder population but not too fine to get precise and robust UPG estimates. To limit computation requirements, a sire model was used to validate the UPG definition in terms of goodness-of-fit and predictive ability of the model but also significance and robustness of the UPG estimates.

Goodness-of-fit of a model without and with UPG. The goodness-of-fit of models with and without groups was evaluated for genetic evaluation of MS and SS on representative samples of 300 herds (10 replicates). It was not calculated with the Genekit software on the whole population because this software did not display the log-likelihood of the model. Nested models were compared in order to assess the interest of the full model including UPG defined according to the three criteria (time period, herd AI rate and region) to a model corresponding to current industry standard only accounting for the genetic trend in UPG, or to an intermediate model with a UPG definition based on time period and AI rate of herds to account for the AI *v.* NS sire selection paths. As we were interested in the predictive ability of the models, we chose to compare the goodness-of-fit of models using the Akaike information criteria (AIC; Akaike, 1973). Indeed, the AIC appears to be more appropriate than the likelihood ratio test or the bayesian information criterion to select the best model in a prediction purpose (Ward, 2008).

Significance of UPG estimates. To test the significance of differences between pairs of UPG estimates, we used approximated standard errors. They were calculated by directly inverting the coefficient matrix of the left-hand side of BLUP equations for the whole population under a simplified sire model. This simplified model only took into account the sire and the contemporary group effects. In a 2000 herd sample, we checked that approximated standard errors were underestimated only by 2% to 6% compared to standard errors obtained by directly inverting the coefficient matrix of the left-hand side of BLUP equations for the complete model. As standard errors of UPG estimates were slightly underestimated, we considered a stricter threshold significance level (1%) than usual (5%) to compare UPG estimates.

Robustness of the UPG effects. To test the homogeneity and numerical stability of the N genetic groups, UPG were split into two genetic subgroups of equal size and founders of each group were randomly attributed to each subgroup. Genetic evaluation was then carried out with a model accounting for 2N genetic groups in order to estimate all the subgroup effects. This operation was replicated 10 times and means and standard deviations of subgroup effects were derived from the estimates of the 10 replicates.

Results and discussion

Characterization of regional subpopulations with factorial methods

Identification of regions. Results of the factorial analyses used to build up regions are presented only for the last time period (1999 to 2006) because they are consistent with those of the preceding time periods. The first principal component of the PCA accounted for about 30% of the total variance existing between French departments. It discriminated departments in the center of France, the originating region of the Charolais breed, from the rest of the departments (Figure 1). The second component of the PCA summarized 18% of the variance and mainly discriminated North Eastern departments from Western and Southern ones. The main characteristics of each region are given in Figure 1, which shows the correlation between each variable and the first two components of the PCA.

The Ward HAC enabled to automate the aggregation of the 59 departments into four geographically consistent regions. They corresponded to (i) the departments of the center of France, (ii) the Western departments, (iii) the Southern region and eventually (iv) a large set of departments in the North Eastern part of France. These four clusters explained 42% of the total variance, which was rather limited. However, a more detailed automatic clustering would have formed regions with unbalanced population sizes or made up of geographically distant departments. To have a deeper insight into regional heterogeneities, these four clusters were further split into 14 regions (Figure 2).

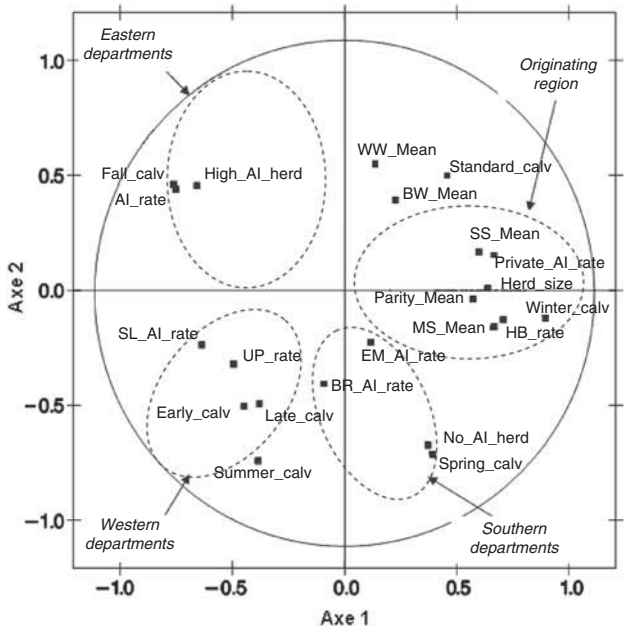


Figure 1 Correlation circle of the first two principal components of the principal component analysis with the 22 parameters describing Charolais herds across French departments for the 1999 to 2006 period. Circles in dotted line and italic names indicate the localization of departments on a map of France. Abbreviations of variables are listed in Table 1.

https://doi.org/10.1017/S1751731110002144 Published online by Cambridge University Press

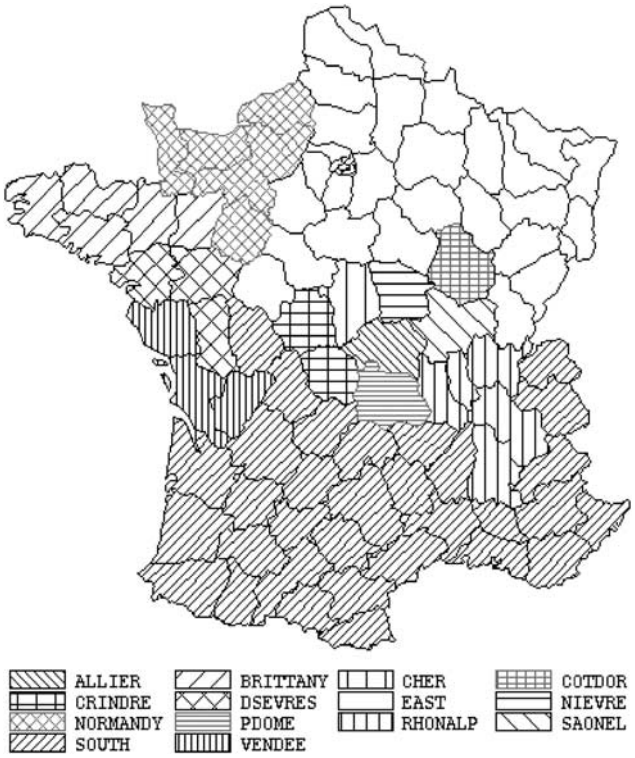


Figure 2 Representation of the 14 French Charolais breeding regions.

Flows of bulls between regions. Since 1983, flows of NS bulls remained very similar across regions and appeared to be carried out mostly on short distances. Whatever the region considered, more than 50% of the NS bulls were used in their native region. The remaining NS bulls came essentially from herds of the Charolais selection nucleus, which are mainly located in three breeding regions: Allier, Vendée and the association of Saône-et-Loire and Nièvre departments. It was observed that those breeding regions specifically supplied in NS bulls herds located in three distinct areas. Consequently, NS bull flows across the 14 regions could be encompassed within three main breeding areas: the Center Southern breeding area around Allier, the Western breeding area around Vendée and the North Eastern breeding area around Saône-et-Loire and Nièvre. There were very limited (below 13%) exchanges of NS bulls between those breeding areas.

Definition of UPG. Taking into account the four time periods, the 14 regions and the two AI classes could have led to the definition of 112 UPG. To reduce the number of defined UPG for each time period, sire origin effects were estimated in order to gather in the same UPG founders from combinations of 'AI rate × Region' exhibiting similar genetic levels (results not shown). Large differences in sire effects were estimated between AI and NS bulls, especially for SS. For this latter trait, average genetic differences between AI and NS bulls reached more than one genetic standard deviation. This highlighted the high selection intensity on AI sires to produce large framed cows with wide pelvises. These results

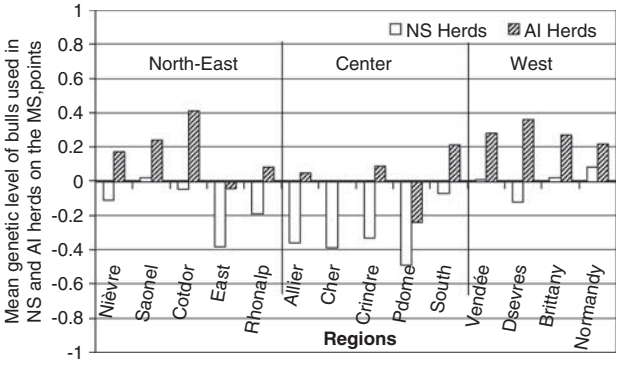


Figure 3 Mean genetic level on the muscle score (MS) of bulls having sired calves born in natural service (NS) herds and in artificial insemination (AI) herds belonging to the 14 French regions between 1999 and 2006.

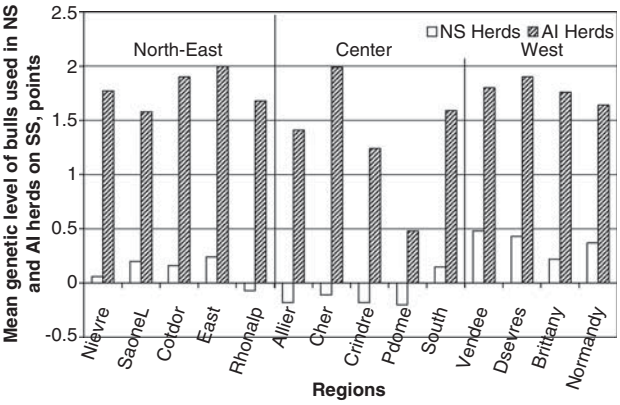


Figure 4 Mean genetic level on the skeleton score (SS) of bulls having sired calves born in natural service (NS) herds and artificial insemination (AI) herds belonging to the 14 French regions between 1999 and 2006.

validated the partition of herds according to their AI rate as a good indicator of the selection intensity on the various sire selection paths.

For NS bulls, sire origin estimates were relatively homogenous across birth regions within a breeding area, but quite different across areas. The low genetic mixing between areas contributed to maintaining three distinct subpopulations in the French Charolais cattle population.

The average genetic levels of sires used in each region are presented for MS and SS in Figures 3 and 4, respectively. They are expressed in contrast to the sire origin estimates for NS bulls bred by specialized breeders from the Nièvre. For both traits, the mean genetic levels of bulls used in NS herds were quite heterogeneous between breeding areas. On the contrary, differences in the mean sire level of NS herds located in regions of a same breeding area were lower than 20% σ_g. Consequently, within each time period, three regional UPG (West, Center South and North East) were defined for unknown parents used in NS herds. They correspond to the three breeding areas that encompassed the initial 14 regions. As expected, within a region, the genetic level of sires used in the AI herds was higher than the one of

https://doi.org/10.1017/S1751731110002144 Published online by Cambridge University Press

sires used in the NS herds. Some genetic differences were detected between AI herds of various regions, but they were lower than 20% σ_g on MS and SS, except for the PDome region. As very few bulls without pedigree were used in AI herds of PDome, hardly one bull per year, and the choice of AI sires was mainly done among a national pool of bulls, a single UPG was defined for AI herds within a given time period, whatever their location in France.

Between 1972 and 1982, since the number of AI herds was negligible, only three regional groups were defined whatever the AI rate of herds. Before the start of the recording system in 1972, a single UPG was defined because most of the animals were located in the originating region of the breed (Bougler *et al.*, 1973). Thus, 16 genetic groups were defined groups, as reported in Table 2.

Validation of the UPG definition

Goodness-of-fit of models. For each of the 10 samples of 300 herd records, the goodness-of-fit estimated in sire models for MS and SS with the 16 UPG herein defined were compared to two other models with more parsimonious UPG definitions and also to a model without genetic groups. Whatever the sample and the trait, the AIC was always lower in the current industry standard model with a group definition based only on time periods (five groups) than in the model without a group. This indicated that the model with genetic groups should be preferred in terms of goodness-of-fit and predictive ability for the complete data set (Table 3). The goodness-of-fit of a model with eight UPG based on time periods and AI rate of herds was compared to the one of the model with the time-based group definition (five groups).

Table 2 *Summary of the defined UPG and the criteria used for their definition*

Herd AI rate	Geographical location	Abbreviation	Calf birth period	Number of bulls without pedigree
NS herds (AI rate <50%)	France	Founders	before 1972	6156
NS herds	Center	CENTER	1972 to 1982	815
			1983 to 1991	1326
			1992 to 1998	715
			1999 to 2006	221
NS herds	West	WEST	1972 to 1982	292
			1983 to 1991	679
			1992 to 1998	461
			1999 to 2006	114
NS herds	North-east	NORTH-EAST	1972 to 1982	1324
			1983 to 1991	2221
			1992 to 1998	1580
			1999 to 2006	398
AI herds	France	AI HERD	1983 to 1991	209
(AI rate ⩾ 50%)			1992 to 1998	298
			1999 to 2006	127

UPG = unknown parent groups; AI = artificial insemination; NS = natural service.

Table 3 *AIC calculated on MS and SS for models without group and with a UPG definition based on time periods (GPI1), on time periods and AI rate (GPI2) and on time periods, AI rate and regions (GPI3)*

	SS				MS			
	Model definition				Model definition			
Sample	Without UPG	GPI1	GPI2	GPI3	Without UPG	GPI1	GPI2	GPI3
1	0	−4	−20	−46	0	−16	−62	−84
2	0	−4	−12	−28	0	−4	−12	−30
3	0	−4	−12	−24	0	−16	−26	−46
4	0	−4	−14	−28	0	−6	−14	−28
5	0	−2	0	−28	0	−2	0	−44
6	0	−6	−24	−44	0	−10	−36	−62
7	0	−6	−16	−32	0	−14	−20	−38
8	0	−8	−20	−36	0	−4	−12	−34
9	0	−2	−12	−40	0	−14	−22	−60
10	0	−6	−50	−32	0	−6	−52	−42

AIC = Akaike information criteria; MS = muscle score; SS = skeleton score; UPG = unknown parent groups; AI = artificial insemination.
All criteria are expressed in contrast to the AIC of the model without group.

https://doi.org/10.1017/S1751731110002144 Published online by Cambridge University Press

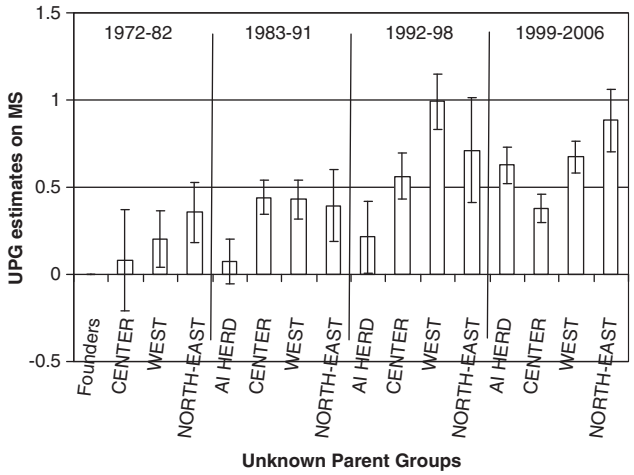


Figure 5 Sire unknown parent groups (UPG) estimates for muscle score (MS), expressed in contrast to the founder UPG estimate and in units of genetic standard deviation and their 99% confidence intervals.

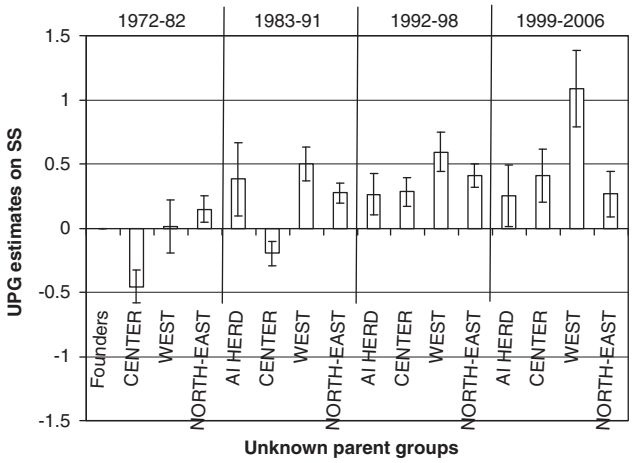


Figure 6 Sire unknown parent groups (UPG) estimates for the skeleton score (SS) expressed in contrast to the founder UPG estimate and in units of genetic standard deviation and their 99% confidence intervals.

For both traits, in nine out of the 10 samples, the AIC clearly indicated that the AI rate of the herds had to be considered in the UPG definition. The full model with 16 UPG was finally compared to a group definition, based on the criteria 'time period' and 'herd AI rate'. Lower AIC were found in nine out of the 10 samples on SS and MS. The regional group definition was also necessary to improve the goodness-of-fit to the data. Thus, it can be concluded that the time period, the AI rate and the breeding area of the herds in which unknown parents were used had to be accounted for in the UPG definition to improve the model fit to data of the French Charolais cattle population.

Estimation of the unknown sire group effects. Figures 5 and 6 present the UPG effects estimated in a sire model for MS and SS, respectively. They were converted to an animal model equivalent ($\hat{a} = 2\hat{s}$) and then expressed in σ_g units. Whatever the trait, UPG estimates within a breeding area generally increased over time due to the genetic trend in the unknown parent subpopulations. Great regional disparities within a time period were also marked, up to 50% σ_g on the last decade, for both traits. This confirmed the AIC results that both the genetic trend and the regionalized breeding strategies were important sources of genetic heterogeneity to take into account for the definition of the genetic groups.

Although significant differences were estimated between UPG at a 1% level, those differences could be insufficient in magnitude to effectively recommend their introduction in the evaluation model. Indeed, Phocas and Laloë (2004a) showed that a model without UPG could be preferred to a model with groups in terms of mean square error and selection efficiency when considering the likely genetic trend and the amount of information available for progeny test evaluation of beef cattle.

Robustness of the UPG. Simple rules consisting of randomly dividing UPG into two subgroups of equal size were used to validate the robustness and numerical stability of the defined UPG. The 10 replicate means of subgroup effects of a same UPG were not significantly different from each other and were close to those estimated for the whole UPG (results not shown). Until the 1980s, the 10 replicate standard deviations of subgroup estimates was moderate ($<$20% σ_g), showing that UPG estimates were stable and relied on a homogenous bull population that diffused a lot across herds. Those standard deviations were greater for the following periods, especially for UPG of small size. About 60 bulls directly linked to the smallest subgroups seemed to be insufficient to get stable estimation of UPG effects. By partitioning UPG, we noticed that at least 100 bulls per subgroup were necessary to get stable estimates of UPG, meaning their sampling standard errors were below 20% σ_g. This minimal number of bulls is much greater than the Interbull recommendations, probably because Charolais bulls without pedigree have 14 progeny on average for the estimation of their BV, which is much fewer than in dairy cattle populations.

Application of the UPG definition for the animal model used in the French beef evaluation

The final aim was to implement a new animal model including relevant UPG for the French on-farm genetic evaluations. A genetic evaluation was therefore carried out under an animal model accounting for UPG using the same definition criteria for bull and cow genetic groups. Bull and cow UPG estimates are presented in Figures 7 and 8 for MS and SS, respectively. They are expressed in contrast to the UPG estimate of the bull founder group. With regard to MS, bull UPG effects were close when comparing estimates from the sire and the animal models. With regard to SS, bull UPG estimates were systematically higher (up to 70% σ_g) in the animal model than in the sire model. Indeed, cows kept for breeding are generally highly selected on SS. Their genetic level was observed to be about 30% σ_g higher than the one of contemporary animals born in the same year. Therefore, ignoring dams in the sire model tended to increase the

https://doi.org/10.1017/S1751731110002144 Published online by Cambridge University Press

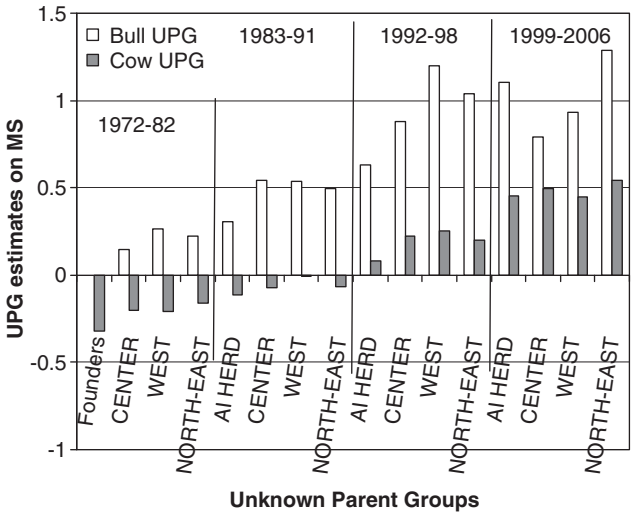


Figure 7 Unknown parent groups (UPG) estimates in the animal model for muscle score (MS), expressed in contrast to the bull founder UPG estimate in units of genetic standard deviation.

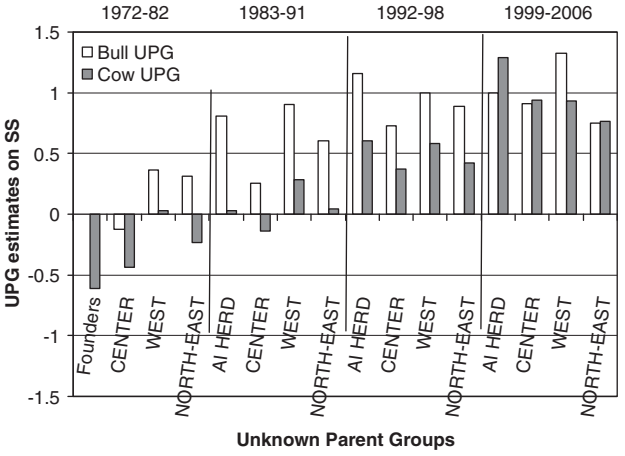


Figure 8 Unknown parent groups (UPG) estimates in the animal model for skeleton score (SS), expressed in contrast to the bull founder UPG estimate in units of genetic standard deviation.

contemporary group estimates to the detriment of genetic group effects.

Cow UPG effects estimated for SS were also lower by 30% to 50% σ_g than bull estimates until the 1990s (Figure 8). However, over the last decade, cow UPG estimates became equivalent or even greater than bull UPG estimates. With regard to MS, cow UPG estimates were lower by about 50% σ_g than the ones of bulls because of lower selection intensity for cows (Figure 7). They were also homogenous across breeding areas, whereas bull UPG estimates exhibited differences across areas up to 50% σ_g.

Genetic trends in MS and SS are presented in Figures 9 and 10, respectively, for the whole calf population and for bulls without and with pedigree in models without and with UPG. Genetic trends estimated in the model with groups were expressed in contrast to the effect of the bull founder group. To gain clarity, genetic trends estimated in the model without group were centered so that the mean genetic level of the calf population born in 1972 is the same in both models. As the definition of genetic groups in the evaluation model allowed accounting for differences in genetic levels among unknown parents, a greater genetic trend was estimated both for MS (Figure 9) and SS (Figure 10). For instance, between 1992 and 2004, genetic trends of 2% and 7% σ_g per year were estimated in the model without UPG on MS and SS, respectively. These trends reached 3% and 8% σ_g per year on MS and SS when calculated in the model with groups. Between 1972 and 1991, genetic trends were lower but the difference in genetic trend between both models remained comparable to the one calculated between 1992 and 2004. As expected, animals with the most affected BV when including UPG in the evaluation model were the ones without pedigree. Comparison of the BV ranking across models indicated that there would be little re-ranking for bulls, especially for elite bulls whose genealogy was well known. On the contrary, large re-ranking would happen for

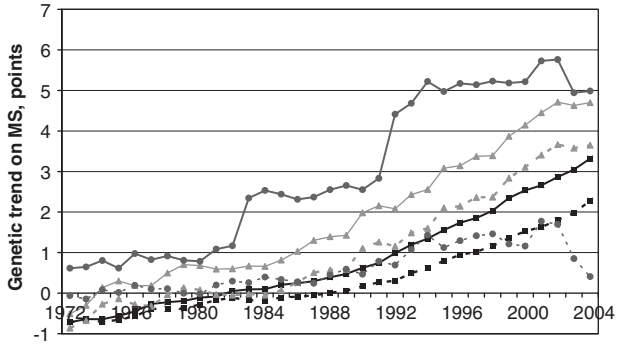


Figure 9 Genetic trends estimated for the muscle score (MS, $\sigma_g = 5.0$ points) in an animal model with genetic groups (plain line) and without groups (dotted line) according to calves (■), bulls with known pedigree (▲) and bulls without pedigree (●).

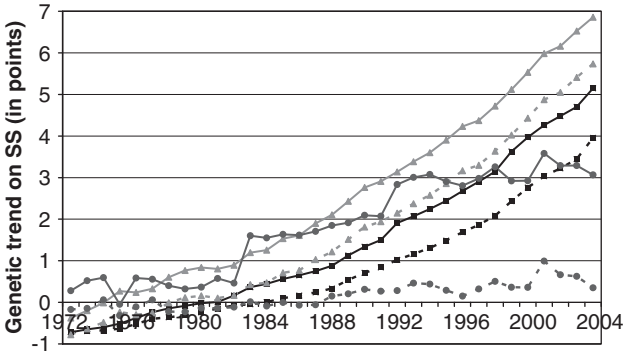


Figure 10 Genetic trends estimated for the skeleton score (SS, $\sigma_g = 3.5$ points) in an animal model with genetic groups (plain line) and without groups (dotted line) according to calves (■), bulls with known pedigree (▲) and bulls without pedigree (●).

cows without pedigree between the model with UPG and the current model without group.

As genealogies of AI sires were extensively known, bulls without pedigree were NS bulls, even when they were used

https://doi.org/10.1017/S1751731110002144 Published online by Cambridge University Press

in AI herds. As shown in Figures 9 and 10, very different genetic levels were estimated between bulls with and without pedigree. Therefore, it seems that bulls without pedigree are not selected for the same objectives as bulls with pedigree. However, it is consistent that farmers who have a similar supplying strategy in bulls would use bulls without pedigree for the same grounds. In the French Charolais population, evidence supported that bulls without pedigree were mostly 'back-up' bulls used to produce calves for slaughtering (high MS wanted) rather than for replacement (high SS wanted). In practice, the postulate led to the definition of bull UPG, which exhibited significant genetic differences between them. The postulate is also clearly relevant to define genetic groups for cows. Actually, a frequent motive of absence of pedigree is the joining of new herds to the recording system. In those herds, parents of animals born before subscription are generally not recognized. As flows of cows across herds are rather limited, the genetic level of cows without pedigree reflects quite well the genetic level of their native herd. Over the last decade, this was confirmed on SS by both a higher mean genetic level of bulls used in AI herds than in NS herds (Figure 4) and a 35% σ_g higher cow genetic group estimate in AI herd than in regional NS herd UPG (Figure 8). Similarly, differences in genetic level estimated on MS were small between bulls used in AI and NS herds across regions (Figure 3), which was consistent with the homogenous cow UPG effects estimated within a given time period (Figure 7).

To conclude, bull usage patterns observed across herds are a relevant source of information to define UPG for a beef cattle genetic evaluation. The method presented in this study permitted to identify relevant UPG that exhibited large genetic differences across time periods and regions. This methodology, as well as the three criteria – time, AI rate and breeding area – proposed to define UPG for the French Charolais population, may be of general interest for any large beef cattle population undergoing selection.

Acknowledgements

Authors would like to acknowledge French beef cattle breeding associations and companies for their collaboration and financial support to this work.

References

Akaike H 1973. Information theory as an extension of the maximum likelihood principle. In Proceedings of the 2nd International Symposium on Information Theory (ed. BN Petrov and F Csaki), pp. 267–281. Akademiai Kiado, Budapest, Hungary.

Banos G, Schaeffer LR and Burnside EB 1991. Genetic relationships and linear model comparisons between United States and Canadian Ayrshire and Jersey bull populations. Journal of Dairy Science 74, 1060–1068.

Bougler J, Le Liboux H, Hocde H and Vissac B 1973. Performances de production. In La race Charolaise – resultats français (ed. J Bougler, B Coudurier, JM Duplan, H Hocde, H Le Liboux, R Tondu and B Vissac B), pp. 265–339. INRA, Jouy-en-Josas, France.

Bouquet A, Renand G and Phocas F 2009. Evolution de la diversité génétique des populations françaises de bovins allaitants spécialisés de 1979 à 2008. Productions Animales 22, 317–330.

De Jong G 2003. MACE – Options for improvement. Interbull Bulletin 30, 112–116.

Fikse F 2003. Approaches to incorporate genetic groups in MACE. Conference at the Interbull Workshop, Beltsville, MD, USA. Retrieved October 8, 2008, from http://www-interbull.slu.se/bulletins/framesida-pub.htm

Gilmour AR, Gogel BJ, Cullis BR, Welham SJ and Thompson R 2002. ASREML user guide, release 1.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Henderson CR 1973. Sire evaluation and genetic trends. In Proceedings of the animal breeding and genetics symposium in honor of Dr. Jay L. Lush. Journal of Animal Science 36 (Symposium), 10–41.

Interbull 2001. Use of phantom parent groups. In Interbull guidelines for national and international genetic evaluation systems in dairy cattle with focus on production traits (ed. H Jorjani, J Philipsson and JC Moquot), pp. 16–17. Interbull, Uppsala, Sweden.

Journaux J, Berrechet P and Ledos H 2006. Etude d'un dispositif déclaratif des filiations en bovins et ses conséquences sur la fiabilité des parentés et l'organisation collective. Technical Bulletin 010678012. Institut de l'Elevage, Paris, France, 56 pp. Retrieved December 13, 2006, from www.inst-elevage.asso.fr/html1/IMG/pdf/QUAPI_CR_Final_n010678012-1.pdf

Kennedy BW 1981. Bias and mean standard error from ignoring genetic groups in mixed model sire evaluation. Journal of Dairy Science 64, 689–697.

Phocas F and Laloë D 2004a. Should genetic groups be fitted in BLUP evaluation? Practical answer for the French AI beef sire evaluation. Genetics Selection Evolution 36, 325–345.

Phocas F and Laloë D 2004b. Genetic parameters for birth and weaning traits in French specialized beef cattle breeds. Livestock Production Science 89, 121–128.

Quaas RL 1988. Additive genetic model with groups and relationships. Journal of Dairy Science 71, 1338–1345.

Robinson GK 1986. Group effects and computing strategies for models for estimating breeding values. Journal of Dairy Science 69, 3106–3111.

Statistical Analysis Systems Institute Inc. 2004. Base SAS 9.1 procedures guide. SAS Institute Inc., Cary, NC, USA.

Ward EJ 2008. A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools. Ecological Modelling 211, 1–10.

Westell RA, Quaas RL and Van Vleck LD 1988. Genetic groups in an animal model. Journal of Dairy Science 71, 1310–1318.

https://doi.org/10.1017/S1751731110002144 Published online by Cambridge University Press