

**An Observatory of the Genetic Variability of Ruminants and Equids breeds**  
**C. Danchin-Burge\***, **E. Verrier†‡**, **D. Laloë‡**, **R. Saintilan§‡** and **G. Leroy†‡**

\* Institut de l'Élevage, 75012 Paris, France, †AgroParisTech, 75005 Paris, France,  
‡INRA, 78350 Jouy-en-Josas, France, § UNCEIA, 75012 Paris, France

**ABSTRACT:** The monitoring of livestock breed genetic variability is a major concern for sustainable agriculture. In this study, indicators of the within-population genetic variability - based on pedigree information - were computed for 79 ruminant and equid breeds. The feasibility of using SNP data to build a robust genetic variability indicator was tested in the Normande cattle breed and compared with the same indicator calculated by using pedigrees. The basic characteristics of the breeds are not the main factors of the evolution of the within-population genetic variability. Preliminary studies also demonstrate that pedigree and molecular data provide complementary information. Therefore, the French project to build a national observatory on genetic variability of all ruminant and equid breeds will be a useful tool for the managers of all breeds.

**Keywords:** Ruminants; Equids; Genetic Variability; Pedigree; SNP

### Introduction

Thanks to the application of quantitative genetic methods and the implementation of selection programs in many breeds, tremendous genetic progress has been achieved for various traits in livestock species, particularly in dairy breeds. One of the downsides of these achievements has been a sharp increase in the inbreeding level of the selected breeds, which leads to inbreeding depression and increased frequencies of genetic defects such as Brachyspina in the Holstein breed (Agerholm et al., 2006). Due to their small numbers, rare breeds are also prone to such side effects.

For these reasons, breeds managers need to be aware of the genetic “health” of their breeds in order to set up specific management plans if needed. In France, livestock officials decided in 2011 to set up an observatory of the genetic variability of all the ruminant and equid breeds that are under selection or conservation programs on French territory.

There is a basic consensus on the utility of using pedigree or molecular data to define genetic variability indicators. Thus, both types of data will be used to build the observatory. The indicators based on pedigrees will be published on a yearly basis starting in 2015, for all breeds with sufficient pedigree information (i.e. a pedigree depth of  $\geq 2.5$  generations). In addition, the feasibility of building an equivalent observatory based on SNP data has been studied for breeds undergoing genomic selection..

The objectives of the present study were (1) to estimate various pedigree-based parameters of genetic variability for French livestock breeds and underline any differences according to species and demography, and (2) to test the stability of comparable estimates obtained by using SNP data.

### Materials and Methods

**Breeds under study.** Pedigree indicators for 79 breeds were studied. These breeds were divided in seven different categories, depending on the species and the use: dairy and meat cattle, dairy and meat sheep, ponies, warm blood horses and draft horses. All these breeds are currently under selection or conservation in France.

**Data.** Pedigree data for the ruminant species were from the National Genetic Information System (French acronym: SNIG) and are stored on the Genetic Information Data Centre of INRA servers. For the equid species, the information was from the Equids Information System (SIRE) managed by the IFCE (Institut Français du Cheval et de l'Équitation). SNP data were generated from the genotypes used for genomic selection in the Normande dairy cattle breed.

**Genetic variability indicators based on pedigrees.** Numerous articles describe ways to calculate indicators of genetic variability based on pedigrees (e.g. Danchin-Burge et al. 2011; Leroy et al. 2013). For our study, we relied on the results provided by the PEDIG software (Boichard, 2002). For each breed, the indicators calculated were, first, simple demographic statistics such as the size (REF) of the analyzed populations (i.e. females born between 2004/2007 with known parents), the average generation interval (T), the numbers of sires (Ns) and dams (Nd) and the number of founders (f) of the analyzed populations. Other indicators were calculated based on the theory of probability of origin (Boichard et al., 1997) such as effective number of founders (Fe) and effective number of ancestors (Ae). Coefficients of inbreeding (F) and coancestry (C) within each population were also calculated. Finally, to compare results with the indicators based on SNP data, the effective population size (Ne) was calculated based on the Cervantes et al. (2011) approach. Various ratios were also included in the analysis, such as Fe/f, Ae/Fe, Ns/REF and Nd/REF.

**Genetic variability indicators based on SNPs.** Following Danchin et al. (2013), the effective population

size ( $N_e$ ) based on SNP data was computed by using the linkage disequilibrium (LD) between physically unlinked loci. Our previous studies, based on a limited number of genotypes, demonstrated the impact of the sample size and structure on the reliability of  $N_e$  estimates. This work was based on the Normande dairy cattle breed, which has a large number of animals and a good availability of SNP data. The 22,429 genotyped animals were split into six subsamples depending on their birth date (3 periods [2005-2007], [2008-2010] or [2011-2013]) and sex. The male population comprised AI candidates, while females included mostly bulls' relatives or bull-dam candidates.

Our objective was to check the stability of the  $N_e$  estimates depending on these two criteria (birth date and sex).  $N_e$  was estimated according to the following formula (Waples, 2006):

$$N_e = \frac{1}{3(r^2 - \frac{1}{n})}$$

where  $r^2$  is the LD, and  $n$  is the sample size.

## Results and Discussion

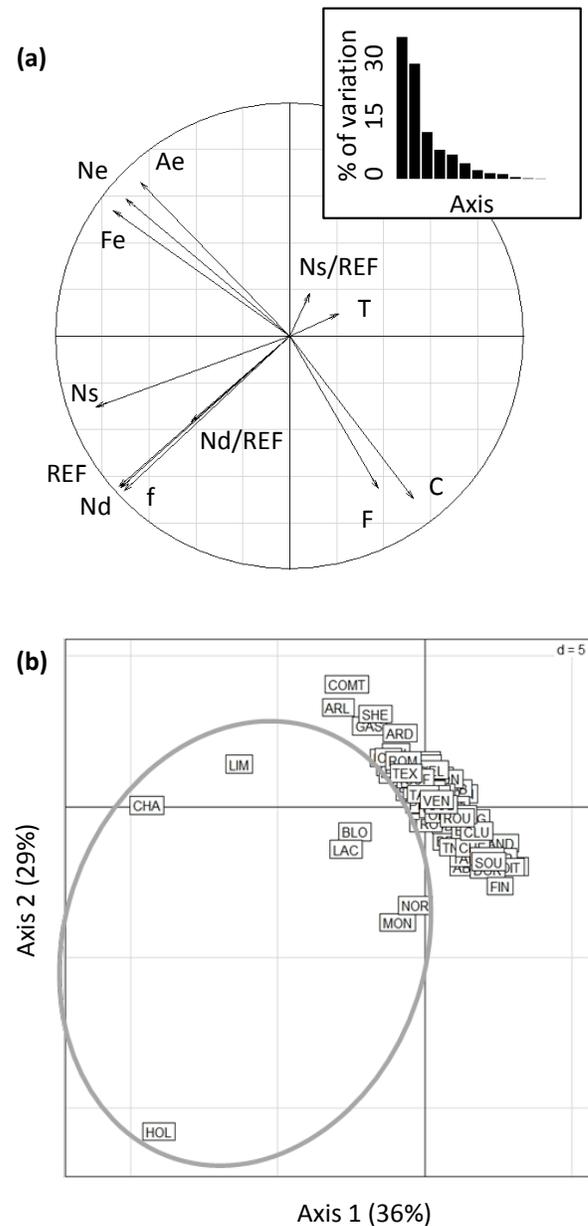
**Correlations between indicators based on pedigrees.** Only the main results are shown. The first two axes of the principal component analysis, which explain about 65% of the total variation (Figure 1), allowed the differentiation, on one hand, of demographic parameters (such as REF,  $N_s$ ,  $N_s$  and  $f$ , which show correlation coefficients among themselves larger than 73%), and, on the other hand, indicators related to genetic variability ( $C$ ,  $F$ ,  $A_e$ ,  $f_e$ ,  $N_e$ ). Among the demographic parameters, an interesting result was the significant positive correlation between REF and  $N_d/REF$  (0.41). It seems that the  $N_d/REF$  ratio increases faster than REF as the size of the population increases. As an interpretation, in large populations, dams have fewer offspring than in small ones, due to greater selection pressure leading to a shorter generation interval.

$f_e$  and  $f$  were found to not be significantly correlated (0.09) whereas the correlation between  $f$  and REF was close to 1. Our interpretation is that the genetic variability of a breed does not depend on its actual size but it is dependent on the severity of previous population bottlenecks. This result is illustrated even better with Figure 1: the main result found in our study is that there is limited correlation between demographic parameters (such as REF,  $N_s$  and  $N_d$ ) and genetic variability indicators ( $N_e$  included). Large populations are as likely to be a dangerous situation (from a genetic variability point of view) as small ones, and a small population may have more genetic variability than a much larger one if well managed.

On the other hand, there are strong correlations between all the genetic variability indicators: the smaller  $A_e$  and  $f_e$  are, the bigger is  $F$ . In other words, inbreeding and coancestry have a trend to be high in a population with a small number of ancestors.

PCA results (Figure 1) clearly differentiate breeds with a large population size, regardless of species (i.e. the Blonde d'Aquitaine: BLO, Charolais: CHA, Holstein: HOL, Limousine: LIM, Montbeliard: MON, Normande: NOR cattle breeds, and the Lacaune: LAC sheep breed).

Otherwise, it was not possible to find differences according to the species or the breeds' use.



**Figure 1: Principal Component Analysis on genetic variability indicators considering the first plane: (a) correlation circle, (b) breeds positions**

(a) REF: population size,  $N_d$ : number of dams,  $N_s$ : number of sires,  $f$ : number of founders,  $f_e$ : effective number of founders,  $A_e$ : effective number of ancestors,  $F$ : inbreeding coefficient,  $C$ : coancestry coefficient,  $N_e$ : effective population size. (b) Holstein (HOL), Charolais (CHA), Normande (NOR), Montbeliard (MON), Limousine (LIM), Blonde d'Aquitaine (BLO), and Lacaune (LAC) sheep breed. Other breeds do not show specific tendencies.

### Genetic variability indicators based on SNPs.

Our results (Table 1) showed that, for both sexes,  $N_e$  followed an upward trend, which was slightly less pronounced among females than males. The sample size is probably the explanation for this tendency, as the number of genotyped males has increased 500 times between the first period and the latter two. For the last two periods, the indicators were very similar for both sexes, showing that, in breeds where the number of breeding offspring per male is very high (thousands or even hundreds of thousands in some cases), the genetic variability of the analyzed population is equivalent across sexes. Preliminary results (not shown) in another dairy species - the Lacaune sheep breed, for which the number of offspring per male is much less than in dairy cattle - showed a clear sex effect on the  $N_e$ . These results need to be confirmed with more data; however initial information suggest that SNP data need to be grouped by time period and even sex to ensure reliable results for breeds under high selection pressure.

**Table 1 :  $N_e$  estimates based on SNP and pedigree data for the Normande breed, depending on the sex and birth period**

	SNP data				Pedigree data
	Sample size	Ne		Ne	
	Male	Female	Male	Female	Female
2005/07	466	3,619	103	132	66
2008/10	2,161	3,996	134	139	68
2011/13	2,251	5,078	151	153	73

### Comparison between pedigrees and SNP indicators.

For the Normande breed, the  $N_e$  based on pedigrees were calculated over the same three periods as for the SNPs data (see Table 1) for the female population. Two main outputs were apparent. First, we observed the same tendency as with the SNPs data, i.e. an improvement of the  $N_e$  value (from 66 for the first period to 73 for the last period). Secondly, the pedigree values were roughly half the values calculated from the SNP data. In contrast, a previous study of four dairy sheep breeds (Danchin et al., 2013) showed very similar results between pedigree and SNP  $N_e$  estimates. Several reasons might explain this difference. In the sheep breeds, the SNP data were pooled by sex and period. Also, in the last three decades, the Normande breed has been going through tremendous bottlenecks because only a limited number of bulls were used for breeding. Thanks to genomic selection, this tendency is decreasing, as shown by the increasing  $N_e$  values. Since  $N_e$  values based on pedigrees are based on a limited number of generations (8.5 generations known per animal on average for the Normande breed), this indicator is illustrating a recent trend. On the other hand, SNPs data are giving insight on the real genetic variability of the breed. Following Leroy et al. (2008) reasoning when comparing molecular and pedigree data for dog breeds, our hypothesis is that the Normande breed founder pool was highly variable (which follows the insight we have on how the breed was created) then drastically reduced recently.

### Conclusion

Our results showed that basic characteristics, such as the actual size of a population or the use of the breed, are not the main factors of the evolution of the within-population genetic variability. This evolution is mainly a matter of management. Some ancient events (e.g. bottlenecks) have long term consequences and only adapted indicators can provide a good view of both the past trends and the expected future evolutions.

Our preliminary results also show that SNP data can be used to build a reliable indicator of the effective population size. This indicator is quite complementary to the ones built with pedigrees since it gives an input on the founder pool that cannot be traced with pedigrees since these data are limited generation wise.

The French national observatory will be a useful resource to help breed managers to monitor the evolution of the genetic variability of livestock populations. The observatory will be complemented by guidelines and software for management rules allowing preserving the genetic variability for the long term.

### Acknowledgment

Our deep thanks to VALOGENE for providing us the genotypes. Funding provided by CASDAR (French Minister of Agriculture).

### Literature Cited

- Agerholm J.S., Mc Evoy F., Arnbjerg J. (2006). *J Vet Diagn Invest* 18:418–422.
- Boichard D. (2002). *Proc 7th WCGALPCD-Rom*, comm. n° 28-13.
- Boichard D., Maignel L., Verrier E. (1997). *Genet. Sel. Evol.* 29:5-23.
- Cervantes I., Goyache F., Molina A. et al. (2011), *J. AnimBreed. Genet.* 128 56-63.
- Danchin-Burge, C., Leroy, G., Brochard, M. et al. (2011). *J. Anim. Breed. Genet.*, doi:10.1111/j.1439-0388.2011.00967.x.
- Danchin-Burge C., François L., Laloë D. et al. (2013). *Interbull Bulletin* No. 47111-119.
- Leroy, G., Verrier, E., Meriaux, J. C. et al. (2009). *Animal Genetics*, 40: 323–332.
- Leroy, G., Vernet, E., Pautet, M.B. et al. (2014) *J. Anim Breed. Genet.* 131: 53–60
- Waples, R. (2006). *Cons. Genet.* 7:167-184.