55th EAAP meeting
September 3-8 2004
Bled, Slovenia

Commission on Animal Genetics
Session III: Management of genetic variation
September 6th

SAUVAGE, a software to manage a population with few pedigrees

J. Raoul[1,2], C. Danchin-Burge[2*], H. de Rochambeau[3], E. Verrier[1],

[1]*INA Paris-Grignon / INRA, UMR Génétique et Diversité Animales, 16 rue Claude Bernard, 75331 Paris Cedex 05, France*
[2]*Institut de l'Elevage, Genetics Department, 149 rue de Bercy, 75595 Paris Cedex 12, France,*
[3]*INRA SAGA, Auzeville BP 27, 31326 Castanet Tolosan cedex, France*

Abstract - There are numerous ways to manage genetic variability within endangered breeds of small ruminants species, but a main difficulty is to find one that the breeders and/or the endangered breeds managers will actually comply with. The purpose of this study was to realize a software to help farmers to choose their most appropriate males (from a genetic variability point of view) among all the available ones.

The starting point for this work was a program developed by H. de Rochambeau (INRA) for two endangered breeds of small ruminants. This tool's main strong point is that it works whatever the level of pedigree completeness in the breed, including the case when no pedigree information exists. However, this program is appropriate only for breeds without a selection program because it doesn't take into account any selection criteria (such as EBVs) to dispatch the males. The software was tested successfully in three different breeds. It is now fully operational but it needs to be transferred on a more use- friendly platform before it can be transmitted to endangered breeds managers.

Management of genetic variability / endangered breeds / software

1 INTRODUCTION

In the last sixty years, the consequences of the intensification and modernization of agriculture are breeds' specialization and an intensification of their selection. A few breeds, well adapted to the new conditions of production, were developed whereas a majority of other breeds were declining. In order to stop this regression, endangered breeds conservation programs were initiated as early as the beginning of the seventies.

Due to their limited size, endangered breeds are particularly exposed to genetic drift and to fast increase in inbreeding. The management of their genetic variability (that we will call here "genetic program") is thus a crucial stake. For small ruminants such as goat and sheep, sire-dam individual mating plan is often not very appropriate, as breeders generally use several males for a large number of females and paternity control rarely exists. Methods of management such as the "rotational scheme" already proved their effectiveness both from theoretical arguments (Chevalet and Rochambeau, 1985) or by analysis of real populations (Huby et al., 2003). These methods, however, are often implying more work for the breeders and major changes in their daily management practices : as a result, lot of breeders eventually give up the genetic program because it is too constraining and not flexible enough on a practical level.

Thanks to a calling proposal from the BRG (Bureau des Ressources Génétiques), a method was developed by the INRA in collaboration with the breeders association of the Bizet sheep. The goal was to help the genetic management of a population of farm animals split into several flocks and that had no pedigree information (Rochambeau, 2000). This method, called method of the flocks' genetic contributions, was applied to the Bizet and to the Pyrenees goat breed. It provided for each flock recommendations for the attributions of the young replacement males.

In order to broaden this method to other endangered breeds and also to breeds with more pedigree information, it was decided to develop a project specification for a computer program called SAUVAGE (Safeguard of Genetic Variability). This tool calculates an indicator called measure of similarity which helps the endangered breeds managers to organize the males exchanges between the breeders. It also calculates a series of indicators which allows to have a simple follow-up on how effective is the management of the genetic variability in the population. Selection criteria are not taken into account in the design of the software, as the most important objective for these small populations is to have a better and efficient way to manage their genetic variability.

---

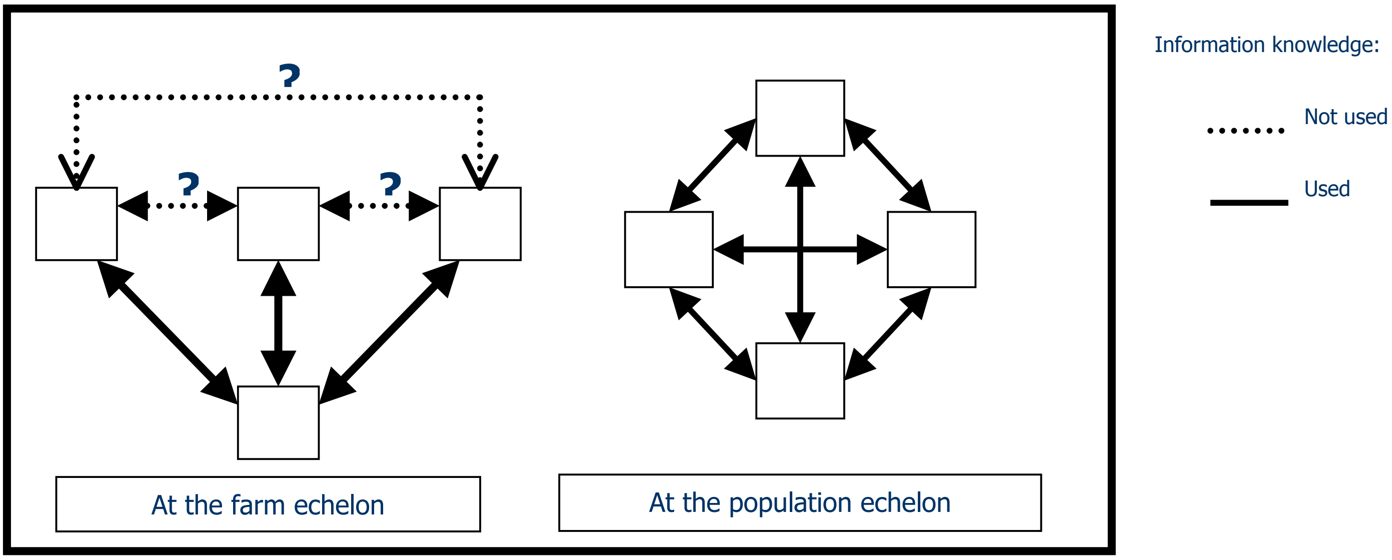
[*] corresponding author: coralie.danchin@inst-elevage.asso.fr

2 PROGRAM DESIGNING

The design of this tool, a data-processing program in FORTRAN language, is based on the following points: a breed has to be managed as a whole and its genetic program should be more than simply avoiding mating between related individuals. This last point is a necessary rule but is not sufficient by itself.

As we just said, the most important point is to reason the genetic program on the whole population. If most breeders are careful to vary the origins of their replacement males in order to avoid the harmful effects of inbreeding, within-flock management has a limited effect at the population level. Breeders usually know where their replacement males are coming from but, as figure 1 shows, they don't know the exchanges that occurred between the other breeders. If we want to take into account all the available information, it is necessary to consider all the genes transfers that are taking place in the population.

Figure 1: Level of information knowledge according to the echelon where the genetic program is based



2.1 Calculation principles

2.1.1 Vector of contribution

SAUVAGE is designed to manage breeds with limited pedigree information. The program is based on a modeling of each flock. The individuals in a flock are characterized by a vector of contribution which includes N terms (with N the number of flocks in the population). The value of each N term corresponds to the probability for a gene taken at random for an individual to be coming from one of the N original pools (which are the flocks at an initial time, cf. following paragraph).

This vector of contribution is calculated for each group of males Km (i.e. all the males born the same year, in the same farm and, if information on the sire exists, with the same father) and each cohort of females Kf (i.e. all the females born the same year and from the same flock), and contains N terms.

The main hypothesis is that at an initial time each flock represents an independent gene pool. Then we take into account all the males exchanges that occurred between the initial time and a time *t*, which is in general the year of study of the population. The contribution of each gene pool (described at the initial time) to the formation of a flock (at a time *t*) is calculated by the method of probability of gene origin.

From a practical point of view the set up of the initial time depends on the breed structure: for a breed with a small number of flocks, one generation length is enough. In this case, it was shown that if you take a longer period of time the probabilities of origin are the same for all flocks and the method is not discriminating any more. On the other hand, for breeds with higher number of flocks, the initial time can be chosen up to two or three generation lengths as the probability of origin remains discriminating in this case.

The population structure at a given year *t* is formalized by the matrix P^t. This matrix contains [Km + Kf]*N lines, which correspond to each cohort in a given flock at a time *t*, and N columns which correspond to all the flocks at a time *t*.

The flocks' structural evolution between *t* and *t+1* depends on demographic parameters which can be variable or identical from one flock to another (cf. § 2.2).

To determine the new gene structure of a flock at a time *t+1*, a transition matrix C^t is built as :

$$P^{t+1} = C^t * P^t$$

The elements $p_{jk,j'}$ of the matrix P^{t+1} represent the contribution from the flock cohort *jk* at *t* to the constitution of the flock cohort *j'q* at *t+1* : they are the probability that a gene taken at random in the flock cohort *jk* comes from the flock cohort *j'* at a time *t*.

The calculation of the elements $c_{jq,j'k}$ of the transition matrix depends on the cohorts *k* and *q* considered and as a result they change according to demographic parameters. Each year the replacement males are parameterized in the program by indicating for each of them where they were born and where they were bought and used for mating. Also, for each male cohort, the program uses an average culling rate, parameterized by the programmer: there is no individual management of the culled animals. On the other hand, pedigree information (on the sire) can be input for a male if it is known. As for the females cohorts' contributions, female exchanges between flocks cohort can be parameterized, but not at an individual level. It is an average "flow" of female between two flocks that is input in the program.

As we stated at the beginning of paragraph 1.1, the female population and the young males cohort for each flock are characterized by a vector of contributions. Two matrices Y_m and Y_f can therefore be built as :

- $Y_m \{y^m_{jj'}\}$: this matrix represents the contribution of each flock *j'* to the new cohorts of males *jq*. Therefore Y_m has N lines and N columns. The lines of the matrix correspond to the lines *jq* (q=1) of the matrix P^{t+1}. Thus, there is only one same vector for all the young males of the same cohort unless their sire is known. In this case, there is a common vector for all the sons (born the same year) that are sharing the same sire.

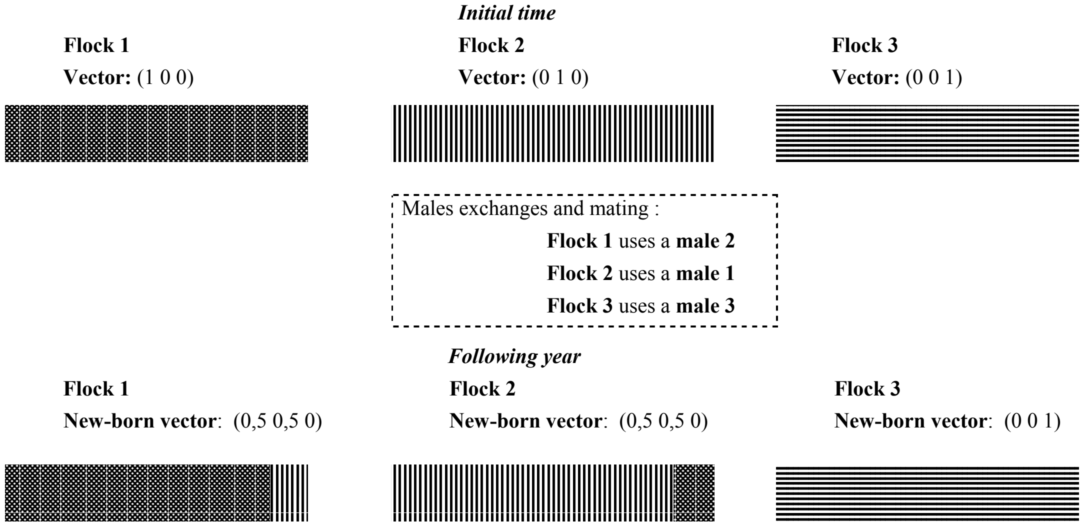
- $Y_f \{y^f_{jj''}\}$: this matrix represents the contribution of each flock *j'* to the entire female population from a flock *j*. Y_f also has N lines and N columns. Its elements are computed as :

$$y^f_{jj'} = \left(\sum_{q=Km+1}^{Km+Kf} p_{jq,j'} \right) / Kf$$

The contributions are calculated with the assumption that there is an equal distribution of the females in each year class. Therefore, the exact number of female by cohorts is not taken into account and unlike what is done for the male, there isn't a specific culling rate by cohort.

For better understanding, let us take an example (figure 2). At an initial time each flock represents an independent gene pool. The flocks are then exchanging males: for example the breeder 2 buys a male to the breeder 1. SAUVAGE simulates the exchange of genes that are occurring after mating the animals in each flock and gives the new contribution of the initial gene pools to a given flock. The oldest year-class is culled while the origins of the new one are computed. In our example we can see that now, in the flock 2, the initial gene pool 1 contributes for 1/8th of its genetic structure. This was computed by saying that it is the paternal contribution of the genes pool 1 to the young cohort X (1/2) X the demographic weight of the young cohort (for example 1/4 if it is the culling rate entered by the programmer).

Figure 2: Example of a simulation carried out by SAUVAGE



2.1.2 Measure of similarity

By using the computed vectors of contribution, the program calculates a measure of similarity $d_{mj,fj''}$ between a male of the population and the females of a given flock. This measure varies between 0 and 1. It represents the proximity between the new males cohort of a given flock *j* and the females of the flock *j''*. Its value varies according to the level of relationship between the two populations: if its is close to 1, it means that the two population have close common origins.

The measure of similarity $d_{mj,fj''}$ is computed as :

$$d_{mj,fj''} = \sum_{j'=1}^{N} \min\left(y^{m}_{jj'}, ,y^{f}_{j''j'}\right)$$

with : *mj* : the young male cohort in the flock *j*

fj'': the entire female population of the flock *j'*

It is equivalent to the probability of drawing a gene in the male group and in the female group that is coming from the same initial gene pool. This measure gives an indication on the male exchanges between the flocks from the initial time till the time of computation.

2.2 Program parametering

To simulate all the mating that are taking place in each flock, the breed's manager must configure the program by entering the breed's demographic parameters such as the total number of flocks participating in the program, the male's cohort demographic weights etc.

He also needs to enter every year the list of the males used in each participating flock, by indicating for each of them its identification (its official identification number or a simple name or denomination in populations without generalized identification), its flock of origin and its age, and, if known, its sire. If some females have also been exchanged, the breed's manager needs to enter their flock of origin and the number of females exchanged.

SAUVAGE simulates mating in each flock by integrating these demographic parameters and calculates new vectors of contribution: the vector of the new males groups (there is only one group by flock if there is no paternity control), the vector of the young female cohorts, and then, by using demographic weighting, the average females vector for a given flock. These vectors are then used for the calculation of measure of similarity between the new males to be sold and each flocks' females.

3 PROGRAM OUTPUT

3.1 Double-entry table

By using the information on the males used in each flock, SAUVAGE computes measures of similarity which give the breeds managers explicit indicators on how well the population is managed from a genetic variability point of view.

The first output of this program is a double-entry table (cf. figure 3) where all the measures of similarity are gathered. The columns represent the purchasers (i.e. the female population) and the lines represent the males to be sold. Genetic mixing is encouraged by advising the breeders to buy the males with the smallest measure of similarity with their flock.

For example, the breed manager will advise the breeder "BR" not to take the male number 4 and to prefer the male number 3. Yet, the breed manager needs to keep in mind the global picture of the breed: it is necessary to be careful to have an optimal ventilation of the males within the flocks. As an example, we can see that the breeder BL should pick between the males number 1, 3 or 4. But, as we know that the choice of BR is restrictive and is directed towards male number 3, BL should be advised to buy males number 1 or 4.

Figure 3: simulation of a SAUVAGE output

Males' flock of origin		**Buyers names**			
	Male number	**BR**	**BL**	**BN**	**BC**
BR	**1**	0,19	0,01	0,05	0,07
	2	0,24	0,08	0,00	0,01
BN	**3**	0,04	0,00	0,38	0,03
BC	**4**	0,10	0,02	0,06	0,20

Measure of similarity

<= 0,05	Recommended
Between 0,05 et 0,10	Not advised
> 0,10	Forbidden

The figure represents the measure of similarity between a male and a flock. The higher the value is the more related the male and the flock are.

The use of the measure of similarity is just an helping tool to find solutions which will maximize genetic mixing while avoiding mating between related animals. As a result, this double-entry table has a mainly a teaching purpose: the main purpose of this tool is to let the breeders have enough choice. Thus it is necessary to keep in mind the breeders satisfaction by letting them choose their replacement males among lot of animals. For example, if 20 males are available, it sounds reasonable if 5 of them are forbidden and 5 others recommended.

3.2 Other indicators of genetic variability

Another output from SAUVAGE is a series of indicators (recapitulated in table 1) which give a better understanding and a better follow-up of the genetic variability both in the flocks and in the population.

If the breed has a collective rearing station for the replacement male, we recommend to carry out this calculation before the new males are entered: the results can help the breed manager in his choice among all the replacement males candidates. As an example, with these indicators it is easy to check which sires are the most used in the population and to reduce accordingly the number of their sons in the rearing station.

Table 1: Summary of the indicators calculated by SAUVAGE

Rule	Indicator	Goal	Where the indicator is used ?
Genes exchanges	Auto-replacement %	Should be as low as possible	Flock and population
	Categories of males in the program : • Recommended • Unadvised • Without indication	Maximizing the number of males recommended	Flock and population
	Males % coming from the collective rearing station	As high as possible	Population
	Nb of ♂ born in a flock / Nb of females in the flock	The values obtained must be similar in all flocks	Population
Equilibrium of family size	Number of males in the conservation program • By sire • By grand-sire	Ideal=1 It is necessary that this number is similar for all sires and grand-sires. This indicator also helps to check if each male has a replacement progeny.	Population
Increase the number of males and their turnover	Females/males ratio	As low as possible	Flock and population
	Males population pyramid	The oldest males must be culled	Population
	Number of year of service by male : • Average by flock • Population average	Limit the length of their reproductive life. For each male, a maximum of two years seems appropriate, and it should be reduced to one year in the smallest flocks.	Flock Population

4 DISCUSSION :

SAUVAGE was successfully tested in three different populations, two breeds of sheep (Bizet and Roussin) and one breed of goat (the Pyrenees goat). This test gave some relevant information on how to use this program in an appropriate way on a practical level. As we saw previously, it is necessary to be efficient to realize the genetic program in a breed at the population level.

Rules such as equilibrium of progeny sizes by sire, low sex-ratio and small length of reproductive life (mostly for the males) need to be considered to preserve the genetic variability of the population. For example, it is necessary to avoid a traditional structuring of the population known as "pyramidal", where a small number of breeders are providing almost all the males to the rest of the population. In this type of population, the breeders that are at the top of the pyramid are replacing their males by using sires from their own flock. Because of the limited size of their flock, a limited number of alleles will be randomly fixed. These alleles fixed at the top of the pyramid are then slowly propagated to the rest of the population. For this reason, in a pyramidal scheme, there are high risks that some unfavorable genes may spread in the population (e.g. an allele of susceptibility to scrapie disease) and it also favors a fast increase in inbreeding (see, for instance, Huby et al., 2003). On the contrary, in a system where all the breeders are selling and buying males, the risk to lose or fix an allele is much weaker: the probability that the loss occurs simultaneously in all the flock is much weaker than the probability of loss in a single flock of limited size.

For this reason, a minimum of collective organization of the breed is necessary: a maximum number of breeders must be part of the genetic program. As a result, the use of SAUVAGE must go along with writing down the genetic program's technical rules, which are going to define what are the different constraint levels. These technical rules constitute an essential point to control the management of the genetic variability in the breed. They must be discussed, approved and respected by all the breeders that are part of the program.

It is obvious that by following the indicators calculated by SAUVAGE it will help to preserve a good genetic variability in a breed, but is even more interesting to set up the whole " SAUVAGE package", i.e. to include an agreement among breeders on technical rules for a program to preserve genetic variability. This implies setting up a collective organization of the breeders, which is anyway a key point to create a successful conservation program.

The indicators provided by SAUVAGE can be very useful to develop the breed's technical rules. They give a fast characterization on how well the population is managed from a genetic point of view and they show which are the criteria that should be improved for the maintenance of the population genetic variability. Once the genetic program is set up, these indicators can be used to check if the breeders are actually following the technical rules that were set up. Finally, the indicators can be used to manage males exchanges and also which are the best males candidates for a collective rearing station (for example males with rare sires or males issued from specific recommended sires etc).

Yet, it appears that for populations characterized by frequent transfers of small number of females, or when the real origins of the sire-dams are atypical compared with the rest of the flocks, the measure of similarity calculated leads to erroneous results. For these populations, a considerable error can be made if we do not take into account the maternal pedigree: unfortunately, they are not always available. Although the general indicators remain always useful in these populations, it is necessary, in these cases, to use other methods to avoid mating between related animals. Also, this program is not meant for population where it is common that each dam is covered with a different sire (by using artificial insemination for example) and where good pedigree information is available. For these populations, management programs based on pedigree information are more useful and accurate to maintain a good genetic variability.

5 CONCLUSION

There are numerous ways of managing the genetic variability of small populations (Rochambeau 1990, Verrier, 1992). However, on a practical level, these solutions are often too constraining for the breeders (limited choice for the attribution of replacement males etc.) and lot of them eventually give up the genetic program.

In France, there was an acute need to find a practical way to manage the genetic variability of breeds with poor pedigree information and where a limited number of males are used to service a large number of females - which is often the case in small ruminants. To answer this need, it was decided to design a computer program called SAUVAGE.

SAUVAGE is a program based on elementary rules for the management of endangered breeds genetic variability, such as equilibrium of progeny size, high numbers of males and fast turnover, genetic mixing maximization etc. SAUVAGE is particularly well adapted for the management of small populations with few or without pedigrees and where flocks are mostly exchanging males. It was tested to manage endangered breeds of sheep and goat, but it can also be used in any farm species as long as their structure is close from what we just described. On a practical level, this program should be used only in breeds with a collective management to be really efficient.

Two other important principles directed the development of SAUVAGE: genetic management has to be thought at population level, not at flock level, and it can not be summarized at simple advises on how to avoid mating between related individuals.

SAUVAGE calculates a measure of similarity between each male and each females flock. It also gives a series of indicators to check if the elementary rules of genetic management are followed. By using this program, the person in charge of the genetic management of the breed can pick the best males for its collective rearing station (upstream part of the genetic program) and then advise farmers in their choice for replacement males (downstream part of the genetic program). However, these rules are just helping the breed manager on how to choose the best males from a genetic variability point of view, as no selection parameters were included in the program designing.

Also, SAUVAGE by itself is not a genetic variability management program: it just gives some help to determine and check how well breeders are following their genetic program. Its use must go along with the setting up and then the application of technical rules fixing the program's levels of constraints. These rules can formalize the compromise that the breeders are ready to make between preservation of genetic variability and selection.

This work must be continued by programming SAUVAGE under a more convivial platform before it can be transferred to endangered breeds organizations. The entry of the initial parameters in the program needs also to simplified. When the software will be fully operational, the priority will be to transfer it to endangered breeds organization. However, it is likely that selected population might be interested by using this software in order to set up a program to maintain their genetic variability, or just to have a tool to monitor their genetic variability.

REFERENCES

Chevalet C., Rochambeau H. de, 1985. Livest. Prod. Sci. 13, 207-218.

Huby M., Griffon L., Moureaux S., Rochambeau H. de, Danchin-Burge C., Verrier E., 2003. Genet. Sel. Evol. 35, 637-655.

Rochambeau H. de, Chevalet C., 1990. 4th World Cong. Genet. Appl. Liv. Prod. 13, 434-442.

Rochambeau H. de, Perrin M. 2000. -Rencontres Recherches Ruminants, 7, 157-160.

Verrier E., 1992. INRA Prod. Anim., HS, 265-271.