

Combining microsatellite and pedigree data to estimate relationships among Skyros ponies

E. Bömcke^{1,2}, N. Gengler^{1,3}

¹Animal Science Unit, Gembloux Agricultural University, Passage des Déportés, Gembloux, Belgium

²F.R.I.A., Rue d'Egmont, Brussels, Belgium

³National Fund for Scientific Research, Rue d'Egmont, Brussels, Belgium

Abstract. Relationship coefficients are particularly useful to improve genetic management of endangered populations. These coefficients are traditionally based on pedigree data, but in case of incomplete or inexistent pedigrees they are replaced by coefficients calculated from molecular data. The main objective of this study was to develop a new method to estimate relationship coefficients by combining molecular with pedigree data, which is useful for specific situations where neither pedigree nor molecular data are complete. The developed method was applied to contribute to the conservation of the Skyros pony breed, which consists of less than 200 individuals, divided into 3 main herds or subpopulations. In this study, relationships between individuals were estimated using traditional estimators as well as the newly developed method. For this purpose, 99 Skyros ponies were genotyped at 16 microsatellite loci. It appeared that the limitation of the most common molecular-based estimators is the use of weights that assume relationships equal to 0. The results showed that, as a consequence of this limitation, negative relationship values can be obtained in small inbred populations, for example. By contrast, the combined estimator gave no negative values. Using principal component analysis, the combined estimator also enabled a better graphic differentiation between the 3 subpopulations defined previously. In conclusion, this new estimator can be a promising alternative to traditionally used estimators, especially in inbred populations, with both incomplete pedigree and molecular information.

Keywords: microsatellites, pedigree, relationships, Skyros pony.

Introduction

Numerous methods for computing inbreeding and relationship coefficients are available. The first type of methods is based on genealogical information. These methods can be classified into 2 categories (Boichard et al. 1996): (1) methods that consider the founders as unrelated and non-inbred (e.g. Wright 1922; Henderson 1976 or Meuwissen and Luo 1992), and (2) methods that consider the founders as potentially related and inbred (e.g. VanRaden 1992). These relationship coefficients are traditionally calculated on the basis of the pedigree of the breed. However, in many situations, complete pedigrees are not available for a

significant fraction of the individuals in the studied population or even for the whole population. Consequently, in recent years, scientists have become interested in the use of genetic markers to determine the genealogical relationships present in populations (Pamilo and Crozier 1982). Thus, the second type of methods is based on molecular markers, such as microsatellites. They also can be classified into 2 groups: (1) methods that use moment estimators to estimate the relatedness between pairs of individuals (e.g., Queller and Goodnight 1989; Li et al. 1993; Lynch and Ritland 1999), and (2) methods that use a likelihood approach to allocate pairs or groups of individuals to relationship categories, such as full-sibs, par-

Received: August 6, 2008. Revised: December 11, 2008. Accepted: December 17, 2008.

Correspondence: E. Bömcke, Gembloux Agricultural University, Animal Science Unit, Passage des Déportés 2, B-5030 Gembloux, Belgium; e-mail: bomcke.e@fsagx.ac.be

ent-offspring, etc. (e.g., Mousseau et al. 1998; Goodnight and Queller 1999; Thomas and Hill 2002).

Knowledge of relationships among animals is useful for both the study of wild populations and the genetic management of captive and/or threatened populations (Glaubitz et al. 2003). In wild populations, relationships can be used in studies of kin selection, social behavior and social organization (e.g. Morin et al. 1994). Furthermore, this knowledge is also useful in research concerning mating systems (e.g. Heg and van Treuren 1998; Engh et al. 2002), dispersal, isolation by distance, and special genetic structure (e.g. Goodisman and Crozier 2002), and for the estimation of quantitative genetic parameters, such as heritability (Mousseau et al. 1998; Ritland 2000). In captive populations, knowledge of relationships and shared ancestry enables the minimization of inbreeding by permitting matings only between the most distantly related individuals (e.g. Jones et al. 2002). The knowledge of the relationship coefficients is thus one of the principal tools used to optimize conservation strategies (Hedrick and Miller 1992; Rochambeau et al. 2000; Caballero and Toro 2002; Verrier et al. 2005).

In reality, both sources of information are often incomplete or limited. An example of this situation is the Skyros pony, a Greek indigenous horse breed, mainly found on the island of Skyros. The risk status of this breed is described as critical-maintained, according to the criteria established by the Food and Agriculture Organization (FAO 1998; DAD-IS 2007). The small size of this population, about 200 individuals concentrated in 3 main subpopulations (Skyros, Corfu and Thessaloniki), and the linked risk of demographic accidents, are major factors that explain this status. The studbook of the breed includes animals recorded for 10 to 20 years. The pedigree depth is thus not sufficient to estimate accurately the relationships, as many individuals have unknown parents. An accurate estimation of relationships among individuals from molecular data is also difficult. Indeed, the number of genotyped individuals and of markers is limited essentially for budgetary reasons. So combining molecular and pedigree data could permit a better estimation of the relationships among individuals than estimation based on only one type of available data.

Although molecular markers are available for most species for several years, it is interesting to note that so far no studies have tried to combine both sources of information into one single estimator. In our study, the idea was therefore to de-

velop a combined estimator. The choice to combine the 2 types of coefficient was made for 2 reasons. Firstly, if only the results of the DNA analysis are used to estimate the relatedness, individuals are related only for known markers. DNA analyses allow "telling the historical review of the breed". If only the pedigree data are used, it does not enable retracing all the history of the breed, but it is very informative for the close parents. Consequently, the simple replacement of the pedigree-based coefficients by molecular-based coefficients leads to a loss of information. Secondly, a limitation of the common molecular-based estimators was the use of weights that assume zero relationship. Thus including the known pedigree relationship in the estimation, allowed to correct the molecular coefficients for the relations existing in the studied population.

With regard to conservation breeding programs, it seems essential to use pedigree information whenever available, especially when the costs of genotyping for a high number of microsatellite markers are considered. But the situation might change if genotyping costs could be reduced in the near future. Meanwhile, DNA technologies are not suitable to replace a studbook (Baumung and Sölkner 2003).

In conclusion, the main objective of this study was to improve the calculation of relationship values by developing a method combining pedigree and marker data, using the endangered Skyros pony breed as a reference population.

Materials and methods

Source of information

A preliminary studbook for Skyros pony was established very recently and includes 395 animals, born between 1958 and 2006. A previous study showed that the Skyros population consists of 3 subpopulations: Skyros (about 100 individuals), Thessaloniki (50 individuals) and Corfu (30 individuals) (Bömcke 2007). The completeness of pedigree information was characterized by computing the number of generation-equivalents (*geq*). This parameter is often considered as the best criterion to characterize the quality of the pedigree information (Maignel et al. 1996; Baumung and Sölkner 2003). The *geq* was computed for each animal as the sum of $(1/2)^n$, where n is the number of generations separating the individual from each known ancestor (Huby et al. 2003).

DNA samples were collected from 99 ponies: 37 males and 62 females. The sampled animals

were born between 1982 and 2006 and represented the 3 herds: Skyros (50 individuals), Thessaloniki (25 individuals) and Corfu (24 individuals). The samples were tested for genetic variation at 16 loci on 12 different chromosomes with microsatellite markers, recommended for parentage testing by the International Society of Animal Genetics (ISAG) Equine Genetics Standing Committee. The microsatellites were the following: *VHL20*, *HTG04*, *AHT04*, *HMS07*, *HTG06*, *AHT05*, *HMS06*, *ASB23*, *ASB02*, *HTG07*, *HMS03*, *HMS02*, *ASB17*, *LEX03*, *HMS01* and *CA425* (Dimsoski 2003).

Definition of pedigree relationship coefficients

Basing on pedigree records, the construction of the additive relationship matrix was performed recursively, using the tabular method as described by Van Vleck et al. (1987). Additive relationships were sequentially established from oldest to youngest animals. Founders (i.e. animals without known parents) were considered as unrelated and non-inbred. Lets x and y denote a pair of individuals, p and q denote the parents of y , and a_{xy} denote the additive relationship coefficient among x and y . It is assumed that $a_{xy} = a_{yx} = 0.5(a_{xp} + a_{xq})$ for $x \neq y$ (Henderson 1976). The inbreeding coefficient of y is calculated as a half of the additive relationship between its parents (p and q). It follows that $F_y = 0.5a_{pq}$ and the additive relationship of y with itself is $a_{yy} = 1 + F_y$.

Table 1. Total number of alleles and polymorphism information content (PIC) of the 16 microsatellites

Locus	Total no. of alleles	PIC
VHL20	6	0.724
HTG04	4	0.166 [§]
AHT04	5	0.585
HMS07	7	0.767
HTG06	5	0.451 [§]
AHT05	6	0.712
HMS06	6	0.443 [§]
ASB23	6	0.734
ASB02	6	0.716
HTG07	4	0.585
HMS03	7	0.666
HMS02	5	0.691
ASB17	10	0.842
LEX03	6	0.670
HMS01	5	0.254 [§]
CA425	7	0.639
Mean	5.933	0.598

[§]Loci excluded from construction of matrix based on highly informative loci (PIC < 0.5)

The additive relationships (a_{xy}) were transformed into Wright et al. (1925) relationship coefficients (r_{xy}), according to the following equation: The obtained relationship, also called numerator relationship, is independent from the inbreeding coefficient of individuals x and y . This transformation allows the relationship coefficient to be a measure of the degree to which the genotypes of x and y are similar, rather than leave it in terms of the proportion of genes from a common source (Minvielle 1990).

Definition of molecular relationship coefficients

Basing on microsatellites, the total allelic relationship (TA) of the 2 alleles of an individual with the 2 alleles of the other individual was calculated for each locus (l): $TA_{xy,l} = 2 \times fM_{xy,l}$

The coefficient of 2 emphasizes that TA is twice the relationship coefficient (Malécot 1948) and is analogous to the numerator relationship (Wright 1922) calculated from the pedigree.

The fM is the molecular co-ancestry between 2 individuals (Caballero and Toro 2000; Eding and Meuwissen 2001). By definition, it is the probability that 2 alleles taken randomly (one from each individual) are identical-by-descent (IBD). This coefficient has the advantage of being defined in a similar way as the classical Malécot coefficient (Toro et al. 2002). The fM can be written as: $fM_{xy,l} = 1/4[S_{ac} + S_{ad} + S_{bc} + S_{bd}]$ where the subscript l indicates the locus, the subscripts a, b, c and d indicate allelic position 1 of l of individual x , allelic position 2 of l of x , allelic position 1 of l of individual y , and allelic position 2 of l of y , respectively, and $S_{.}$ refers to values depending on whether alleles at the allelic positions in the subscript are the same ($S_{.} = 1$) or not ($S_{.} = 0$).

The locus-specific relationship can be further averaged over L loci:

$$TA_{xy} = \frac{\sum_{l=1}^L TA_l}{L} \quad (\text{Nejati-Javaremi et al. 1997}).$$

Thereafter, TA_{xy} was transformed into $r_{mol,xy}$ by the following equation:

$$r_{mol,xy} = TA_{xy} / \sqrt{TA_{xx} \times TA_{yy}}$$

where TA_{xx} and TA_{yy} denote the total allelic identity of the individuals x and y with themselves. The obtained relationship matrix (R_{mol}) was thus comparable to the pedigree-based relationship matrix (R_{ped}).

Two different R_{mol} matrices were constructed: the first one based on all the 16 microsatellites of the marker set, and the second one based only on markers that were highly informative (PIC > 0.5), i.e. 12 microsatellites (Table 1).

Definition of a new combined relationship coefficient

In an ideal situation, the whole pedigree is known and relationship calculations are based on statistical assumptions about genes shared between individuals. Thus, the more information is available from the pedigree, the less molecular information is needed. The other extreme situation is when very dense marker maps are available. According to VanRaden (2007), these molecular relationships, called then genomic relationships, can replace pedigree information completely. In practice, and especially in conservation genetics, intermediate situations are observed, where both pedigree and marker data are incomplete. Many pedigree records are not very reliable due to random mating between individuals with incomplete or no recording of the births. To use all available data, we combine the 2 relationship coefficients $r_{ped,xy}$ and $r_{mol,xy}$ into a single parameter $r_{comb,xy}$. As the relative quality of information needs to be integrated, $r_{ped,xy}$ and $r_{mol,xy}$ should be weighted using a function that is proportional to the quality of information. Theoretical weights can be developed basing on the reciprocals of the error variances of relationship coefficient estimates. However, obtaining these error variances is not evident; in fact, no method exists. Therefore, we used an empirical function of relative pedigree deepness as weighting factor for the pedigree relationship coefficients. The following empirical weighting for the pedigree relationship coefficient between animals x and y was chosen and tested in this study:

$$w_{xy} = \sqrt{(1 + geq_x) \times (1 + geq_y) / (2 \times (1 + geq))}$$

where geq_i is the number of generation-equivalents for the individual i (x or y), and geq represents the average number of generation-equivalent for the analyzed population. For the animals that have the most complete pedigree, this function is close to 1, and for an animal with no pedigree record, this function is close to 0.

The molecular relationship coefficient was multiplied by the complementary value computed as $1 - w_{x,y}$ and by the average polymorphism information content (PIC) value. This last parameter was introduced by Botstein et al. (1980) and refers to the value of marker informativeness within a population, depending on the number of alleles and their frequencies.

Finally, the final formula of the new combined estimator is thus:

$$r_{comb,xy} = w_{x,y} \times r_{ped,xy} + (1 - w_{x,y}) \times PIC_{mean} \times r_{mol,xy}$$

For comparison, a classic molecular relationship estimator was also computed, as proposed by Lynch and Ritland (1999). Their single-locus relationship coefficient is:

$$\hat{r}_R(l) = \frac{p_a(S_{bc} + S_{bd}) + p_b(S_{ac} + S_{ad}) - 4p_a p_b}{(1 + S_{ab})(p_a + p_b) - 4p_a p_b}$$

where l indicates the locus, the subscripts a, b, c and d indicate allelic position 1 at locus l of individual x , allelic position 2 at locus l of x , allelic position 1 at locus l of individual y , and allelic position 2 at locus l of y , respectively, p_a and p_b denote the population allele frequency of the alleles at allelic positions a and b , and $S_{..}$ refers to values depending on whether alleles at the allelic positions in the subscript are identical ($S_{..}=1$) or not ($S_{..}=0$). Multilocus estimates can be obtained as the sum of the single-locus estimates weighted by the inverse of their sampling variance, assuming zero relatedness, which equals:

$$w_R(l) = \frac{(1 + S_{ab})(p_a + p_b) - 4p_a p_b}{2p_a p_b}$$

The Lynch and Ritland estimator was chosen because it performed best in the studies of Van de Castele et al. (2001) and Csilléry et al. (2006).

Example

Two genotyped individuals are chosen randomly: x and y .

The first individual (x) is a male coming from the Skyros herd. He was born in 1998. Its sire is unknown, its dam is known (Dx). For this animal, geq is equal to 0.5.

The second individual (y) is a female coming from the Corfu herd. She was born in 2004. Its sire is known (Sy), its dam is known (Dy). If we trace the complete pedigree of this animal, the correspondent geq is 1.938. For the genotyped population, the mean number of geq is 1.447.

Step 1: Calculation of the pedigree relationship coefficient:

Basing on the pedigree, there is no relation

(1) between the parents of individual y (Sy and Dy) and individual x , then:

$$a_{xy} = a_{yx} = 0.5 \times (a_{xDy} + a_{xSy}) = 0.5 \times (0 + 0) = 0$$

(2) between the parents of individual x (Sx and Dx , because Sx is unknown), and between the parents of individual y (Sy and Dy), then:

$$a_{xx} = 1 = 1 + 0.5a_{SxDx} = 1$$

$$a_{yy} = 1 + F_y = 1 + 0.5a_{SyDy} = 1 + 0.5 \times 0 = 1$$

It follows that:

$$r_{ped,xy} = a_{xy} / \sqrt{a_{xx} \times a_{yy}} = 0 / \sqrt{1 \times 1} = 0$$

Step 2: calculation of molecular relationship coefficient:

Based on the 2 microsatellite profiles (Table 2),
 $TA_{xy,VHL20} = 2 * fM_{xy,VHL20} = 1/2[S_{ac} + S_{ad} + S_{bc} + S_{bd}] = 1/2[0 + 0 + 1 + 0] = 0.5$

The same procedure is applied to all the microsatellites, TA is the average of the results for all microsatellites, therefore:

$$TA_{xy} = 0.7$$

$$TA_{xx} = 1.2$$

$$TA_{yy} = 1.125$$

It follows that:

$$r_{mol,xy} = TA_{xy} / \sqrt{TA_{xx} \times TA_{yy}} = 0.7 / \sqrt{1.2 \times 1.125} = 0.602$$

Step 3: Calculation of combined relationship coefficients

The weight is calculated from the following equation:

$$w_{x,y} = \sqrt{(1 + geq_x) \times (1 + geq_y)} / (2 \times (1 + geq))$$

$$= \sqrt{(1 + 0.5) \times (1 + 1.938)} / (2 \times (1 + 1.447)) = 0.430$$

The combined relationship coefficient between x and y is finally:

$$r_{comb,xy} = w_{x,y} \times r_{ped,xy} + (1 - w_{x,y}) \times PIC_{mean} \times r_{mol,xy}$$

$$= 0.430 \times 0 + (1 - 0.430) \times 0.598 \times 0.602 = 0.205$$

applied for this purpose the Factor procedure from SAS (1999).

Results and discussion

Pedigree deepness

Table 3 shows the number of genotyped individuals per class of known geq . More than 25% of the individuals had less than one known geq , i.e. one or both parents were unknown. In these cases, calculating a relationship coefficient is impossible. Indeed, a common ancestor is necessary to calculate a relationship value between 2 individuals. When the pedigree is missing, the relationship coefficient is considered to be zero. In this study, the mean number of geq is 1.447 and the maximum is 3.000, for an animal belonging to the Corfu subpopulation. These results showed well the necessity to include the collected molecular data in calculation of relationships.

Table 2. Results of the genotyping kit for 2 chosen individuals (size of the alleles in nucleotides)

°	VHL20		HTG4		AHT4		HMS7		HTG6		AHT5		HMS6		ASB23	
x	86 ^a	94 ^b	130	132	157	157	172	178	85	95	133	135	165	167	188	190
y	94 ^c	98 ^d	130	130	157	159	170	172	95	95	137	133	161	167	188	190
	ASB2		HTG7		HMS3		HMS2		ASB17		LEX3		HMS1		CA425	
x	249	249	117	123	md*	md	217	223	96	98	144	144	175	175	236	238
y	245	249	123	125	163	167	215	223	104	118	154	158	175	175	228	238

^{a,b,c} and ^d used for the calculation of the molecular co-ancestry (fM) between x and y ; *md = missing data

Analysis of combined and traditional relationships

An analysis was conducted to quantify the discriminating power of the various relationship coefficients, i.e. they were tested for their capacity to group the animals according to their herd of origin. The underlying idea is that relationship coefficients are one of the principal tools used to optimize conservation strategies; e.g. by permitting matings only between the most distantly related individuals and thus to minimize inbreeding. The geographic location of the 99 Skyros ponies was known and therefore could be used to test the discriminating power of the proposed measure of relationship. We conducted a principal component analysis (PCA) of 3 relationship matrices for the 99 animals. PCA makes it possible to reduce the number of dimensions, without much loss of information. In this study, PCA was used to present the results by scattered plots in 2-dimensional space, considering the first 2 principal components. We

Table 3. Percentage of genotyped individuals per class of known generation-equivalents (N=99)

Class of geq	% of individuals
0.00–0.49	5.05
0.50–0.99	25.25
1.00–1.49	16.16
1.50–1.99	28.28
2.00–2.49	15.15
≥2.50	10.10

Limitation of the traditional estimator

The relatedness values calculated with Lynch and Ritland estimator ranged from -0.985 to 1.000 , with a mean of -0.156 . We observed that 76.71% of the values were negative. The conclusion was that Lynch and Ritland estimator underestimates the genealogical coefficients, as it was also shown by Toro et al. (2002) and Oliehoek et al. (2006).

The quality of Lynch and Ritland estimator data depends on several factors. These include the number of loci, the allelic frequency and, especially, the degree of true relationship in the living population. One limitation of this estimator, and also of other common estimators of relatedness, is the use of weights that assume relationships in the population equal to zero. The best weight would be a function of the actual relationship (Lynch and Ritland 1999). A high level of relationship in the studied population could then explain the high percentage of negative values. Another limitation is that the allele frequencies are assumed to be the same as in the base population, which implies that there has been no change in gene frequencies. In Skyros ponies, it could not be assumed that the actual frequencies of marker alleles were identical to those of the base population due to the genetic drift accumulated over years. In our opinion, the main explanation for the under-estimation, and the high percentage of negative values, seems to be the relatively high degree of actual relationship in the studied population and the lack of information on the true allelic frequencies in the base population.

Discriminating capacity of various estimators

Table 4 shows the percentage of information explained by the first 3 principal components. The combined estimator showed the lowest total value, which is mainly due to the lower value for the first component, while the methods based on DNA alone led to higher values. TA_{xy} showed the highest total value, and Lynch and Ritland estimator reached the highest value for the first principal component.

As mentioned previously, all individuals from the reference population can be considered as belonging to 3 subpopulations. The repartition of genotyped individuals in these 3 subpopulations was known: a major group in Skyros (50 genotyped individuals) and 2 smaller groups in Thessaloniki (25 genotyped individuals) and Corfu (24 genotyped individuals). This pattern should be reflected in the plot of the first 2 principal components (axes).

Table 4. Percentage of information explained by the first 3 principal components

Principal component	TA_{xy}	Lynch & Ritland	Combined (16 markers)	Combined (12 markers)
1st	26.84%	29.06%	21.06%	21.37%
2nd	12.72%	10.18%	12.43%	11.62%
3rd	11.30%	9.22%	9.06%	8.82%
Total	50.86%	48.46%	42.55%	41.81%

Figure 1 shows the results of PCA of TA_{xy} coefficients. In spite of a high percentage of information explained by the first axis, the scattered plot showed no clear difference between the 3 groups of individuals.

Figure 2 shows the results of PCA of $r_{R,xy}$ coefficients. Although this method showed the highest percentage of information explained by the first axis, the scattered plot showed a better distinction between the 3 groups of individuals.

Figure 3 shows the results of PCA of $r_{comb,xy}$, based on 12 microsatellites. This method presented a lower percentage of information explained by the first axis than the 2 methods based on DNA alone. Nevertheless, the scattered plot showed a good distinction, with only few exceptions, between the 3 groups of individuals. Except for one individual, the first axis differentiated the individuals of the Skyros subpopulation from the individuals of the other 2 subpopulations. The second axis differentiated, with few exceptions, the individuals of the Thessaloniki subpopulation from the individuals of the Corfu subpopulations. In the case of $r_{comb,xy}$ based on 16 microsatellites, the scattered plots (not shown) indicated the same as shown for 12 microsatellites, except that the distinction between the individuals from Thessaloniki and the individuals from Corfu was not evident. This was probably linked to the fact that highly polymorphic microsatellites allow to improve the distinction between individuals. In the case of combined estimator, where the Skyros subpopulation was separated from both other subpopulations, one individual belonging to the Corfu group was placed in the Skyros group of individuals. The most plausible explanation is that this individual is the only one in the Corfu subpopulation, whose both parents were born in Skyros, and that this individual has no descendant. In the last case, the subpopulation of Thessaloniki was differentiated from the Corfu subpopulation with one exception. One individual belonging normally to the Corfu group was found in the Thessaloniki group. The explanation for this observation could be the same as described above, namely that both parents of this individual were born in Thessaloniki and that this individual has no descendant. In the other direction, 5 individuals belonging normally to the Thessaloniki group were considered as part of the Corfu group. A possible explanation for this observation is that these animals were sent to Corfu Island. All their descendants were thus belonging to the Corfu group and they were therefore regarded as being closer to this group than to their group of origin (Thessaloniki).

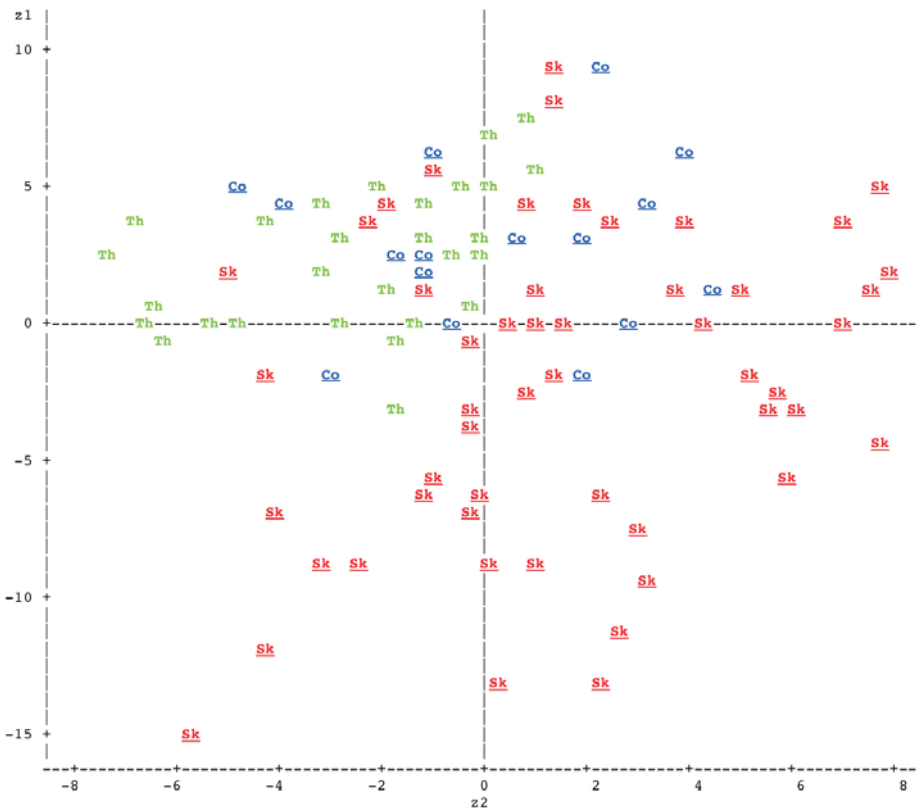


Figure 1. PCA representation of relationship coefficients obtained by calculation of TA_{xy} (Sk=Skyros, Th=Thessaloniki, Co=Corfu).

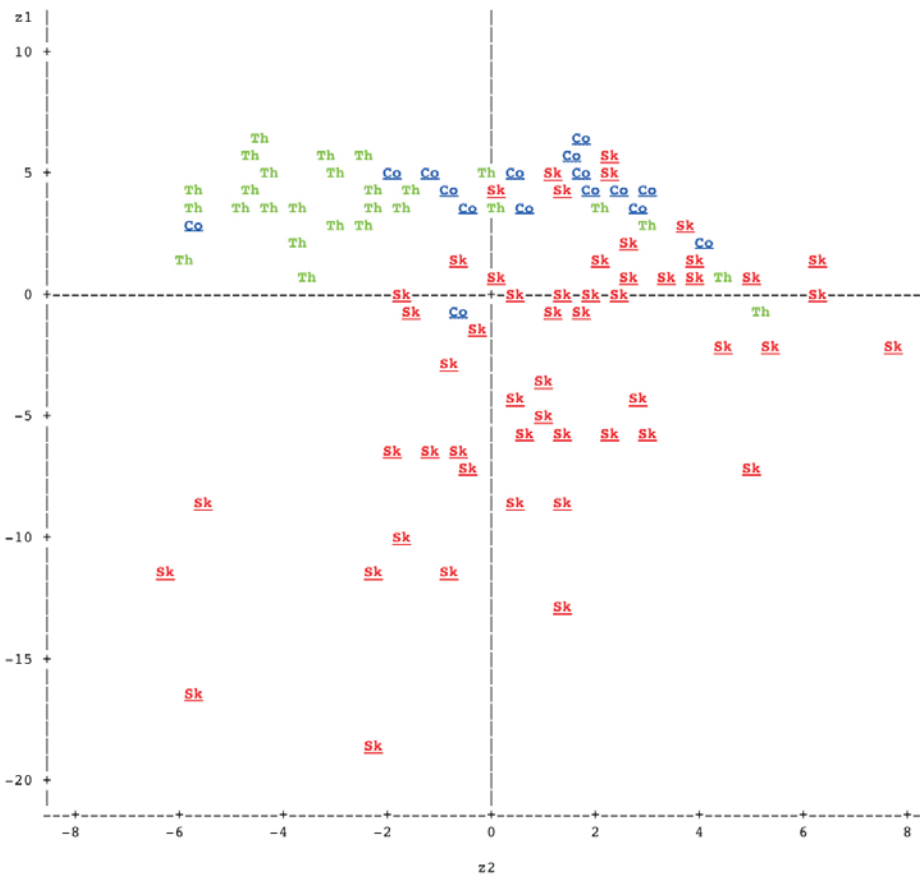


Figure 2. PCA representation of relationship coefficients obtained by calculation of Lynch & Ritland estimator.

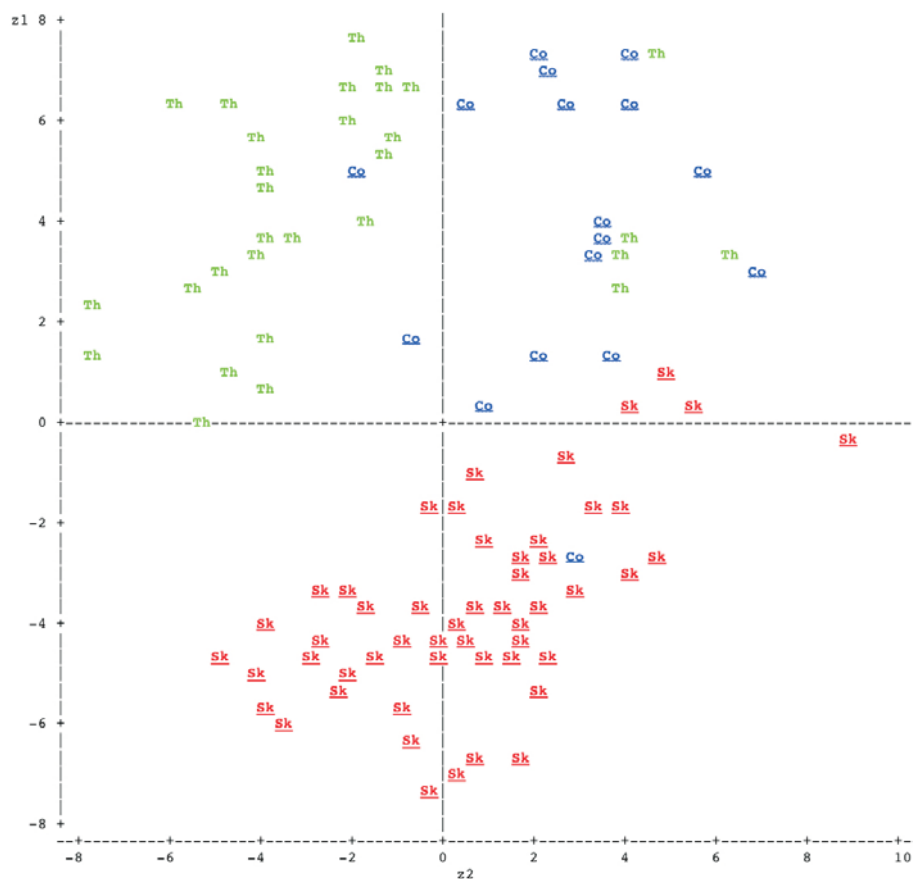


Figure 3. PCA representation of relationship coefficients obtained by calculation of combined estimator on 12 microsatellites.

Additionally, in the case of combined estimator, it was interesting to note that the estimations based on the 12 highly informative microsatellite markers ($PIC > 0.5$) differentiated the subpopulations better than those based on 16, even if the percentages of information explained by the 3 axes were similar. These results showed that the use of the PIC value in the weighting has the potential to correct the value for the informativeness of the markers. However, some improvements are necessary, e.g. the molecular relationship found for each locus should be multiplied by the corresponding PIC value instead of using only the average value.

In spite of this weakness, the combined method for the estimation of relationship showed promising results, as compared to the other estimators. Another problem is the fact that the relationship coefficients are overestimated if DNA information is highly favoured in comparison to the pedigree information, but this is less the case than with other estimators. The fact that the combined estimator gave the most differentiated groups is a quality indicator, but this statement needs to be confirmed in forthcoming studies.

Use of the coefficients

One aim of the study on Skyros pony was to help breeders to determine matings that minimize the increase in inbreeding within the population. For illustration, 6 females and 9 males were chosen: 2 females from each herd: one with the best-known pedigree and the other with the least-known pedigree of the herd, and 3 males of each herd: one with the best-known pedigree of the herd, one with the least-known, and one in between. Table 5 shows the relationship coefficients between those females and males.

Pedigree records alone (r_{ped}) are not sufficient to choose the best stallion, as more than half of the coefficients are equal to 0. For example, males 3, 8 and 9 could be recommended for all the females, and females 3 and 6 could be mated equally with all the males.

The $r_{mol,xy}$ coefficients show a very high level of relationship in the studied population and suggest that the population is already subject to inbreeding. This observation is consistent with the fact that it is an island population. However, if we consider these values, it seems impossible to limit

Table 5. Relationship coefficients, calculated using pedigree and other estimators, between 6 females (F) and 9 males (M) from the 3 different herds (classification of the males in exponent; first and last males of the classification in bold print)

Animal		F1		F2		F3		F4		F5		F6	
Herd		Co	Th	Sk	Co	Th	Sk	Co	Th	Sk	Co	Th	Sk
<i>egg</i>		3.000	2.719	1.875	1.625	1.000	0.000						
<i>r_{ped,xy}</i>													
M1	Th	2.875	0.241	⁷ 0.425	⁹ 0.000	¹ 0.115	⁵ 0.446	⁸ 0.000	¹	¹	¹	¹	¹
M2	Co	2.625	0.432	⁹ 0.225	⁶ 0.000	¹ 0.169	⁷ 0.185	⁶ 0.000	¹	¹	¹	¹	¹
M3	Sk	2.188	0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹
M4	Th	1.938	0.177	⁶ 0.291	⁸ 0.000	¹ 0.258	⁹ 0.016	⁴ 0.000	¹	¹	¹	¹	¹
M5	Co	1.625	0.057	⁴ 0.097	⁴ 0.016	⁹ 0.250	⁸ 0.125	⁵ 0.000	¹	¹	¹	¹	¹
M6	Co	1.500	0.170	⁵ 0.128	⁵ 0.000	¹ 0.063	⁴ 0.250	⁷ 0.000	¹	¹	¹	¹	¹
M7	Th	1.000	0.340	⁸ 0.256	⁷ 0.000	¹ 0.125	⁶ 0.500	⁹ 0.000	¹	¹	¹	¹	¹
M8	Sk	0.750	0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹
M9	Sk	0.000	0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹ 0.000	¹
<i>r_{mol,xy}</i>													
M1	Th	2.875	0.628	⁴ 0.783	⁹ 0.553	⁸ 0.652	⁸ 0.446	⁴ 0.413	¹	¹	¹	¹	¹
M2	Co	2.625	0.659	⁷ 0.593	⁷ 0.580	⁹ 0.636	⁴ 0.419	¹ 0.571	⁶	¹	¹	¹	¹
M3	Sk	2.188	0.569	³ 0.404	¹ 0.442	⁴ 0.637	⁵ 0.459	⁵ 0.511	⁴	¹	¹	¹	¹
M4	Th	1.938	0.557	² 0.688	⁸ 0.481	⁶ 0.646	⁷ 0.427	³ 0.625	⁹	¹	¹	¹	¹
M5	Co	1.625	0.628	⁵ 0.457	² 0.477	⁵ 0.652	⁸ 0.563	⁸ 0.609	⁸	¹	¹	¹	¹
M6	Co	1.500	0.683	⁸ 0.468	³ 0.393	² 0.637	⁵ 0.528	⁷ 0.468	³	¹	¹	¹	¹
M7	Th	1.000	0.721	⁹ 0.494	⁵ 0.519	⁷ 0.631	³ 0.750	⁹ 0.607	⁷	¹	¹	¹	¹
M8	Sk	0.750	0.479	¹ 0.495	⁶ 0.436	³ 0.434	¹ 0.477	⁶ 0.542	⁵	¹	¹	¹	¹
M9	Sk	0.000	0.628	⁶ 0.479	⁴ 0.377	¹ 0.539	² 0.423	² 0.413	¹	¹	¹	¹	¹
<i>r_{comb,xy}</i> (16 microsatellites)													
M1	Th	2.875	0.366	⁶ 0.549	⁹ 0.227	³ 0.342	³ 0.422	⁸ 0.218	¹	¹	¹	¹	¹
M2	Co	2.625	0.509	⁹ 0.361	⁶ 0.253	⁶ 0.379	⁵ 0.294	³ 0.319	³	¹	¹	¹	¹
M3	Sk	2.188	0.214	¹ 0.163	¹ 0.212	² 0.318	² 0.250	² 0.313	²	¹	¹	¹	¹
M4	Th	1.938	0.323	⁵ 0.456	⁸ 0.241	⁴ 0.460	⁸ 0.250	¹ 0.400	⁶	¹	¹	¹	¹
M5	Co	1.625	0.292	³ 0.256	³ 0.259	⁷ 0.471	⁹ 0.387	⁶ 0.411	⁷	¹	¹	¹	¹
M6	Co	1.500	0.381	⁷ 0.280	⁵ 0.211	¹ 0.384	⁶ 0.419	⁷ 0.322	⁴	¹	¹	¹	¹
M7	Th	1.000	0.492	⁸ 0.361	⁷ 0.299	⁹ 0.428	⁷ 0.664	⁹ 0.447	⁹	¹	¹	¹	¹
M8	Sk	0.750	0.224	² 0.249	² 0.259	⁸ 0.269	¹ 0.323	⁵ 0.413	⁸	¹	¹	¹	¹
M9	Sk	0.000	0.320	⁴ 0.262	⁴ 0.244	⁵ 0.364	⁴ 0.312	⁴ 0.342	⁵	¹	¹	¹	¹
<i>r_{comb,xy}</i> (12 microsatellites)													
M1	Th	2.875	0.346	⁷ 0.517	⁹ 0.184	⁵ 0.337	⁷ 0.387	⁸ 0.153	¹	¹	¹	¹	¹
M2	Co	Co	0.472	⁹ 0.309	⁶ 0.202	⁸ 0.328	⁶ 0.224	³ 0.243	⁵	¹	¹	¹	¹
M3	Sk	2.188	0.177	¹ 0.131	¹ 0.168	² 0.268	² 0.173	¹ 0.223	³	¹	¹	¹	¹
M4	Th	1.938	0.272	⁵ 0.419	⁸ 0.170	⁴ 0.412	⁹ 0.182	² 0.340	⁹	¹	¹	¹	¹
M5	Co	1.625	0.239	³ 0.206	³ 0.184	⁶ 0.390	⁸ 0.333	⁶ 0.305	⁷	¹	¹	¹	¹
M6	Co	1.500	0.338	⁶ 0.232	⁴ 0.116	¹ 0.302	³ 0.356	⁷ 0.172	²	¹	¹	¹	¹
M7	Th	1.000	0.459	⁸ 0.318	⁷ 0.186	⁷ 0.310	⁴ 0.632	⁹ 0.283	⁶	¹	¹	¹	¹
M8	Sk	0.750	0.179	² 0.247	⁵ 0.208	⁹ 0.180	¹ 0.284	⁵ 0.334	⁸	¹	¹	¹	¹
M9	Sk	0.000	0.271	⁴ 0.171	² 0.169	³ 0.318	⁵ 0.273	⁴ 0.241	⁴	¹	¹	¹	¹

Co = Corfu, Sk = Skyros, Th = Thessaloniki

the inbreeding increase in the population. We suspected that $r_{mol,xy}$ strongly over-estimated relationship. The explanation is that these estimators account not only for the IBD that arises during the population history, but also for the IBS (identical-by-state) present in the founder population (Oliehoek et al. 2006). The proposed coefficients were still high but smaller than the 2 previous measures. When using only 12 microsatellites, the combined estimator in-

creased the difference between values of first and last classified males.

The combination of both types of information allowed us to make a better distinction between close and distant relationship. Most of the time, the animals with the highest $r_{ped,xy}$ had also the highest $r_{comb,xy}$, except in case of female 3, where the highest $r_{ped,xy}$ was not significantly different from 0, and in case of female 6, which has no pedigree and therefore no known relation with the

other animals. So this method allowed to disadvantage the mating between animals that are related through close parents, as the better pedigree is limited to 3 *egg*. With the combined method, we saw that the best matings are always inter-herd matings, most of the time between individuals from Skyros and individuals from the other 2 herds. Finally, we can consider that combining both sources of information makes it possible also to decrease the relationship values to values that are easier to handle.

Conclusion

Combining the 2 types of coefficients is a promising strategy, because pedigree and molecular information are 2 complementary sources of information. Indeed, pedigrees are very informative for close relatives (parents, grand-parents), while DNA analysis makes it possible to retrace the history of the breed. Consequently, the simple replacement of the pedigree-based coefficients by molecular-based coefficients leads to a loss of information. Moreover, it relates individuals only for known markers. Another advantage of the combined estimator is that including the known pedigree relationship in the estimation allows correction of the molecular coefficients for the relations existing in the studied population, in opposition to the common molecular-based estimators that use weights that assume zero relationships in the studied population. As said in the introduction, it seems essential to use pedigree information whenever available, especially as long as genotyping for a high number of markers remains too expensive for breeders.

When DNA marker information is used, Lynch and Ritland (1999), Toro et al. (2002) and partially our results, showed that attempts to estimate relatedness with molecular markers can be greatly improved by using only highly polymorphic loci, with the highest gains in efficiency occurring with loci with a relatively even distribution of allele frequencies, than by using more loci.

However, we recommended to confirm the results obtained with the combined estimator by performing further investigations using different weighting, more polymorphic markers and/or other populations, simulated or not. In particular, the weighting needs to be improved, as in some cases (not encountered in the studied population) the developed strategy may not work. For example, when both geq_x and geq_y are higher than , then $w_{x,y}$ is higher than 1, which means that the weight-

ing for the molecular coefficients becomes negative.

In the future, this new estimator could be used in conservation genetics. Its use would allow to assess relationships of animals with no pedigree within a population (presenting a complete pedigree or not), without testing all the animals. This would make it possible to integrate more easily the animals of unknown origin in conservation programs.

Acknowledgements. We thank for financial support from the National Fund for Scientific Research (to N.G.) and F.R.I.A, Belgium (to E.B.).

REFERENCES

- Baumung R, Sölkner J, 2003. Pedigree and marker information requirements to monitor genetic variability. *Genet Sel Evol* 35: 369–383.
- Boichard D, Maignel L, Verrier E, 1996. Analyse généalogique des races bovines françaises. *INRA Prod Anim* 9: 323–335.
- Bömcke E, 2007. Conservation of animal genetic resources: Improvement of the relationship matrix. Master, Gembloux Agricultural University, Belgium.
- Botstein D, White RL, Skolnick M, Davis RW, 1980. Construction of a genetic linkage map in man using restriction fragment polymorphisms. *Am J Hum Genet* 32: 314–331.
- Caballero A, Toro MA, 2000. Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genet Res* 75: 331–343.
- Caballero A, Toro MA, 2002. Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv Genet* 3: 289–299.
- Csilléry K, Johnson T, Beraldi D, Cletton-Brock T, Coltman D, Hansson B, et al. 2006. Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics* 173: 2091–2101.
- DAD-IS, 2007. Domestic Animal Diversity Information System (DAD-IS), Food and Agriculture Organization of the United Nations. Retrieved October 07, 2007, from <http://dad.fao.org>
- Dimoski P, 2003. Development of a 17-plex microsatellite polymerase chain reaction kit for genotyping horses. *Croat Med J* 44: 332–335.
- Eding H, Meuwissen THE, 2001. Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *J Anim Breed Genet* 118: 141–159.
- Engh AL, Funk SM, van Horn RC, Scribner KT, Bruford MW, Libants S, et al. 2002. Reproductive skew among males in a female-dominated society. *Behav Ecol* 13: 193–200.
- FAO, 1998. Secondary Guidelines for Development of National Farm Animal Genetic Resources Manage-

- ment Plans. Management of Small Populations at Risk. Initiative for Domestic Animal Diversity. Food and Agriculture Organization of the United Nations, Rome.
- Glaubitz JC, Rhodes Jr OE, Dewoody JA, 2003. Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Mol Ecol* 12: 1039–1047.
- Goodisman MAD, Crozier RH, 2002. Population and colony genetic structure of the primitive termite *Mastotermes darwiniensis*. *Evolution* 56: 70–83.
- Goodnight KF, Queller DC, 1999. Computer software for performing likelihood test of pedigree relationship using genetic markers. *Mol Ecol* 8: 1231–1234.
- Hedrick PW, Miller PS, 1992. Conservation genetics: techniques and fundamentals. *Ecol Appl* 2: 30–46.
- Heg D, van Treuren R, 1998. Female-female cooperation in polygynous oystercatchers. *Nature* 391: 687–691.
- Henderson CR, 1976. A simple method for computing the inverse of a numerator relationship matrix use in prediction of breeding values. *Biometrics* 32: 69–83.
- Huby M, Griffon L, Moureaux S, De Rochambeau H, Danchin-Burge C, Verrier E, 2003. Genetic variability of six French meat sheep breeds in relation to their genetic management. *Genet Sel Evol* 35: 637–655.
- Jones KL, Glenn TC, Lacy RC, Pierce JR, Unruh N, Mirande CM, et al. 2002. Refining the whooping crane studbook by incorporating microsatellite DNA and leg-banding analyses. *Conserv Biol* 16: 789–799.
- Li CC, Weeks DE, Chakravarti A, 1993. Similarity of DNA fingerprints due to chance and relatedness. *Hum Hered* 43: 45–52.
- Lynch M, Ritland K, 1999. Estimation of pairwise relatedness with molecular markers. *Genetics* 152: 1753–1766.
- Maignel L, Boichard D, Verrier E, 1996. Genetic variability of French dairy breeds estimated from pedigree information. *Interbull Bull* 14: 49–54.
- Malécot G, 1948. *Les mathématiques de l'hérédité*. Paris: Masson et Cie. Cited by Van Vleck (1987) and Minvielle (1990).
- Meuwissen THE, Luo Z, 1992. Computing inbreeding coefficients in large populations. *Genet Sel Evol* 24: 305–313.
- Minvielle F, 1990. *Principes d'amélioration génétique des animaux domestiques*. Paris: INRA & Québec: PUL (Les presses Universitaires de Laval).
- Morin PA, Moore JJ, Chakraborty R, Jin L, Goodall J, Woodruff DS, 1994. Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science* 265: 1193–1201.
- Mousseau TA, Ritland K, Heath DD, 1998. A novel method for estimating heritability using molecular markers. *Heredity* 80: 218–224.
- Nejati-Javaremi A, Smith C, Gibson JP, 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. *J Anim Sci* 75: 1738–1745.
- Oliehoek PA, Windig JJ, van Arendonk JAM, Bijma P, 2006. Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. *Genetics* 173: 483–496.
- Pamilo P, Crozier RH, 1982. Measuring genetic relatedness in natural populations – methodology. *Theor Popul Biol* 21: 171–193.
- Queller DC, Goodnight KF, 1989. Estimating relatedness using genetic markers. *Evolution* 43: 258–275.
- Ritland K, 2000. Marker-inferred relatedness as a tool for detecting heritability in nature. *Mol Ecol* 9: 1195–1204.
- Rochambeau H (de), Fournet-Hanocq F, Vu Tien Khang J, 2000. Measuring and managing genetic variability in small populations. *Ann Zootech* 49: 77–93.
- SAS Institute, 1999. *SAS/STAT User's Guide: Version 8*. SAS Inst. Inc., Cary, NC.
- Thomas SC, Hill WG, 2002. Sibship reconstruction in hierarchical population structure using Markov chain Monte Carlo techniques. *Genet Res* 79: 227–234.
- Toro M, Barragán C, Óvilo C, Rodrigañez J, Rodríguez C, Silió L, 2002. Estimation of coancestry in Iberian pigs using molecular markers. *Conserv Genet* 3: 309–320.
- Van de Castele T, Galbusera P, Matthysen E, 2001. A comparison of microsatellite-based pairwise relatedness estimators. *Mol Ecol* 10: 1539–1549.
- VanRaden PM, 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. *J Dairy Sci* 75: 3136–3144.
- VanRaden PM, 2007. Genomic measures of relationship and inbreeding. *Interbull Annual Meeting Proceedings*. *Interbull Bull* 37: 33–36
- Van Vleck LD, Pollak EJ, Oltenacu EAB, 1987. *Genetics for the animal sciences*. New York: Freeman.
- Verrier E, Rognon X, de Rochambeau H, Laloe D, 2005. Les outils et méthodes de la génétique pour la caractérisation, le suivi et la gestion de la variabilité génétique des populations animales. *Ethnozootecnie* 76: 67–82.
- Wright S, 1922. Coefficients of inbreeding and relationship. *Am Nat* 56: 330. Cited by Van Vleck et al. (1987) and Henderson (1976).
- Wright S, McPhee HC, 1925. An approximate method of calculating coefficients of inbreeding and relationship from livestock pedigrees. *J Agr Res* 31: 377–383.