

# Combining Inter- and Intrapopulation Information with the Weitzman Approach to Diversity Conservation

D. GARCÍA, N. CORRAL, AND J. CAÑÓN

From the Laboratorio de Genética, Facultad de Veterinaria, Universidad Complutense de Madrid, Spain (García and Cañón); and Departamento de Estadística, Investigación Operativa y D.M., Universidad de Oviedo, Spain (Corral).

Address correspondence to Javier Cañón, Laboratorio de Genética, Facultad de Veterinaria, Universidad Complutense de Madrid, Avda. Puerta de Hierro s/n, 28040—Madrid, Spain, or e-mail: jcanon@vet.ucm.es.

---

## Abstract

This article introduces a new perspective on Weitzman's methodology for assessing the distribution of resources in genetic diversity conservation programs. Intrapopulation information is added to the procedure by diffusion process formulas to calculate genetic extinction probabilities, and therefore the marginal diversities and elasticities of diversity. The method was tested with a set of European cattle breeds from Spain and France and provided satisfactory results.

---

The efficient distribution of economic resources in biodiversity conservation programs is difficult since many economic, productive, morphological, genetic, social, and even affective and aesthetic factors are involved in decision making. In recent years, the preservation of biological diversity has been a headline topic of discussion [see Amos and Balmford (2001) for a short review]. In particular, the conservation of genetic diversity in livestock has received a great deal of attention (Barker 1999; Eding and Meuwissen 2001; Oldenbroek 1999; Ruane 2000). When developing a program for the conservation of genetic resources, special attention must be paid to how genetic diversity is measured, as it is well known that the rate of evolution in natural populations is limited by the degree of genetic variability (Fisher 1930). For small and endangered populations, it is important to avoid homozygosity as much as possible, because any increase leads to less genetic variation, and molecular estimates of inbreeding would suffer a parallel increase. This determines not only survival, but also adaptation to changing environments, including changes in consumer preferences and demands for animal-derived products (Frankham 1995; Hedrick and Kalinowski 2000; Lacy 1997). Considerable effort has been expended to minimize the long-term rate of inbreeding (Sonesson and Meuwissen 2001; Wang 1997; Wang and Hill 2000). Another point of particular interest is how to make decisions for investment policies to be applied to the different breeds or populations to maximize benefits in terms of diversity conservation.

Weitzman (1992) proposed a set of properties to be verified for the proper measurement of genetic diversity (see also Eding and Meuwissen 2001; Thaon d'Arnoldi et al. 1998). These properties make sense both intuitively and algebraically, but none of the usual measures of diversity verify them (Weitzman 1992). However, Weitzman developed a diversity function in which they all hold. This is ultimately based on pairwise distances between operational taxonomical units (OTUs), and it is easily implementable through a recursive algorithm. Since this is indirectly of the clustering type, it provides a graphic representation that has (in the strict sense) erroneously been called a "maximum-likelihood evolutionary tree." However, it supplies much useful information. Examples of its use can be found in Cañón et al. (2001), Laval et al. (2000), and Thaon d'Arnoldi et al. (1998). With this procedure, the contribution of each OTU, or groups of OTUs, to total diversity can easily be calculated, allowing conclusions on conservation policies to be drawn. However, as with any other methodology, there are some drawbacks (Caballero and Toro 2002; Eding and Meuwissen 2001). The most frequently mentioned is that it ignores within-population information, and being that assertion is not entirely true, since the method can be applied at any level—species, breed, population, and even individual—and therefore a within-population diversity can be computed, the fact is that the algorithm is a between-units one. A way of incorporating within-population information into the algorithm itself, involving computing Weitzman diversity within each population under study, is currently under

development. It is also commonly argued that it does not take into account different population sizes. The present work has filled these gaps by studying the development of gene frequencies within populations in order to compute extinction probabilities.

## Weitzman's Approach

Weitzman (1993) devised a natural way of studying the evolution of diversity over different generations and of weighing the contributions of each OTU more precisely in order to assess the distribution of resources for conservation of genetic diversity programs. Basically it consists of calculating the expected value of a certain diversity measure—the method does not depend on which is chosen—by assigning probabilities to all the possible values of diversity at a particular time. Obviously there are as many values as subsets of the whole set of OTUs, since at a particular time ( $t$ ) some OTUs may have become extinct and only the remainders count toward the final diversity. These values of probability depend on the probabilities of extinction of each particular OTU. Values are then obtained that measure how much the expected diversity at time  $t$  is affected by changes in each probability of extinction. Therefore a very logical and coherent, but conceptually simple way of weighing the OTUs with respect to their influence on expected diversity is constructed. This method has already been applied by Reist-Marti et al. (2003), but they do not escape the fact, already noted in Weitzman (1993), that the choice of these extinction probabilities is strongly subjective, and that is an important drawback of this approach. To circumvent this problem, the present investigation substitutes the concept of extinction with that of genetic extinction, in such a way that these probabilities can be calculated for any generation by using allelic frequencies and the theory of diffusion processes. In addition, this allows within-population diversity information to be brought into the analysis. The original procedure lacked this feature.

This method is a powerful tool for performing combined analyses integrating between- and within-population information. The former comes from pairwise distances, which can be genetic, and are studied by Weitzman's diversity measure. The latter come from allelic frequencies, and are studied in a second step using Weitzman's marginal diversity and elasticity of diversity (Weitzman 1993). These concepts are first introduced and later the theory of diffusion processes is applied to show how it accommodates Weitzman's model. Finally, an example of its use with a dataset from eight European cattle breeds is shown. All calculations were made using a FORTRAN program that implements the computation of Weitzman's diversity, including all the links, representatives, and percentages of diversity contributed by each OTU and each node of the tree. It also provides the "as if" ultra metric distance matrix, which, when used as an input in any tree-generating software, gives the unique hierarchical tree resulting from Weitzman's algorithm. Finally, for a given generation, the probabilities of genetic extinction, the marginal diversity, and the elasticity of the diversity of each OTU are derived.

## Expected Diversity, Marginal Diversity, and Elasticity of Diversity

When establishing a policy of conservation of genetic resources, it is important not only to be able to evaluate current diversity, but to broaden the time horizon and infer the behavior of that diversity a number of generations ahead, and to determine which breeds or species are more influential upon it.

Assuming a trustworthy measure of diversity already exists—whether it is Weitzman's or another—it is of interest to know how the diversity of a set of populations will change, and which of them are more important for the conservation of genetic diversity in the context of biodiversity management programs. It is with this in mind that the concepts of expected diversity, marginal diversity, and elasticity of diversity are proposed.

Let  $\mathcal{Q}$  be a set of  $N_p$  OTUs. Let also  $D: \wp(\mathcal{Q}) \rightarrow \mathbb{R}$  be a certain diversity function, and  $\{P_i\}_{i=1, \dots, N_p}$  be the probabilities of extinction in one generation for each of them. The theoretical development of the procedure that will be described does not depend on  $D$ . It is applicable to any diversity function, as long as for any subset  $S \subseteq \mathcal{Q}$ , the value of  $D(S)$  exists and is well defined. To exemplify the use of this methodology, the Weitzman diversity measure has been chosen. Its explicit definition will not be given here, since it can already be found and discussed in Weitzman (1992) and Thaon d'Arnoldi et al. (1998). If, for each subset of OTUs,  $S \in \mathcal{Q}$ , the probability  $P_t(S)$  of the OTUs in  $S$  having survived after  $t$  generations and those in  $\mathcal{Q} \setminus S$  having become extinct can be found, then  $D$  at time  $t$ ,  $D_t$ , can be thought of as a random variable whose expected value can easily be calculated. The expected diversity after  $t$  generations can therefore be defined as

$$ED_t = \sum_{S \in \wp(\mathcal{Q})} P_t(S) D_t(S). \quad (1)$$

This is not exactly the expression provided by Weitzman, since he defines the expected diversity as the cumulative sum of generations 1 to  $t$ , weighted by a "discount factor." The discount factor has been ignored here because, as shown below, the expected diversity is to be partially differentiated with respect to the probabilities of extinction. Therefore the discount factor provides no information for comparative purposes, and no advantage is to be found in accumulating the sum—besides, equation 1 is more coherent with the statistical concept of expected value.

Another important point that arises from the choice of Weitzman's diversity as the one to operate with is its very nature as a measure that uses pairwise distances as the input to calculate its value. If these distances do not vary with time, then  $D_t(S) = D(S)$ , the diversity at the present time. But  $D_t(S)$  might be different from  $D(S)$  if there is variation in the distances with time, as happens with most genetic distances. It is important therefore to assess the choice of the distance used to calculate the diversity, so that the effect of the chosen evolutionary model can be taken into account.

To calculate equation 1, it is assumed (Weitzman 1993) that the “lifetime” of population  $i$  follows an exponential distribution with parameter  $P_i$ . Thus  $P_i(S)$  is given by

$$P_i(S) = \prod_{i \in S} \exp(-P_i t) \prod_{i \in Q \setminus S} (1 - \exp(-P_i t)). \quad (2)$$

Observe that  $ED_t$  is a function of  $(P_1, \dots, P_{N_p})$ , and therefore can be partially differentiated with respect to each  $P_i$ . As a result, the marginal diversity for population  $i, i = 1, \dots, N_p$ , in generation  $t$ , is defined as

$$MD_t(i) = - \frac{\partial ED_t}{\partial P_i}(P_i). \quad (3)$$

With respect to the meaning of these expressions, the situation is as follows: a set of species, populations, or OTUs in general are under consideration; a measure of diversity has already been established for them; and the importance of each to conservation purposes can now be discussed. The future needs to be examined to see how that diversity might behave and how modifications to each population might affect that expected behavior. One of two things can happen to every single population: it may survive or it may become extinct. Therefore the diversity value at any future point in time will be that of the set of populations that have survived. Given the one-generation probabilities of extinction  $\{P_i\}_{i=1, \dots, N_p}$ , the probability of survival at time  $t$  of any possible subset of the original can easily be calculated, as already shown (equation 2). Now, for a set  $Q$  of  $N_p$  populations, there are  $2^{N_p}$  possible subsets of  $Q$ , that is, there are  $2^{N_p}$  possible patterns of survival/extinction, and hence  $2^{N_p}$  different values of diversity. Each of these values can be calculated along with its corresponding probability, so it makes sense to consider diversity at time  $t$  as a random variable and obtain its expected value. This is precisely the expected diversity.

As just shown, this expected diversity is ultimately a function of  $\{P_i\}_{i=1, \dots, N_p}$ , so studies can be made of how variations in these probabilities affect the expected value of diversity. The more the diversity is affected by changes in a population’s probability of extinction, the more that population should be considered a priority for action. But the concept of variation in this case is just that of a derivative, or more precisely, that of a partial derivative, with respect to each population’s probability of extinction. Note that the value of these derivatives must be negative, since the more a population is likely to become extinct, the lower the expected diversity value will be. The marginal diversity is therefore defined as the negative of the value of the partial derivative to handle positive numbers.

Since a reduction in the probability of extinction provokes an increase in the expected diversity, and this increase is greater for higher marginal diversity values, it is on those populations with the highest values that attention should be focused. This must be directed toward lowering their extinction probability.

Another interesting indicator is the elasticity of diversity or the conservation potential of the population  $i, i = 1, \dots, N_p$ , in generation  $t$ . Elasticity is a very popular and well-known concept in economy. In most cases it compares the relative

variation in the quantity of demand with respect to the relative variation in prices, thereby obtaining a measure of the responsiveness, or sensitivity, of the demand to changes in price. If, for example, the price of a particular good increases by 2% and the demand for that good decreases by 5%, then the elasticity of the demand is  $-2.5$ . Usually the elasticity of demand is negative, since an increase in prices has, as a consequence, a decrease in the quantity of demand. If the elasticity is less than  $-1$ , then the demand is said to be elastic with respect to prices, while if it is between  $-1$  and  $0$ , the demand is called inelastic. If the elasticity equals zero, the demand is completely inelastic, which means that the seller can raise the price of the product as much as he wants and the demand will not decrease at all. The other extreme is a perfectly elastic demand. In this situation, elasticity approaches  $-\infty$ , and the seller can sell virtually as much as he wants at the current market price. Raising the price would imply that demand would nearly disappear, and lowering it would be unprofitable, since demand is practically unlimited at the current price. Most real-life situations are, however, in the middle of these extremes, and sellers have to have adequate prices in order to optimize benefits.

The general formula for the price elasticity of demand is  $E = ((\Delta Q_D / Q_D) / (\Delta P / P))$ , where  $P$  is the current price of a certain good,  $Q_D$  is the quantity demanded at that price,  $\Delta P$  is a small change in the current price, and  $\Delta Q_D$  a small change in the quantity demanded. Note that this expression can be rearranged as  $E = (\Delta Q_D / \Delta P) (P / Q_D)$ . When the changes are not measured in a short time period, averages of the initial and final values are taken on  $Q_D$  and  $P$ . To approximate the elasticity at a particular point, short intervals can be taken around quantity and price values at that point to calculate differences. However, when differentiation is possible, exact computation is available, as will be shown in equation 4.

The concept of elasticity can be applied to many other economic variables, such as supply instead of demand, or in noneconomic contexts, as happens in this case. An analogy can be established here between quantity of demand and expected diversity value and between price and extinction probability, so the elasticity of diversity or conservation potential, as renamed in Weitzman (1993), can be defined as

$$CP_t(i) = - \frac{\partial ED_t}{\partial P_i}(P_i) \frac{P_i}{ED_t}, \quad (4)$$

where, as before, the minus sign is added to deal with positive numbers.

Elasticity is particularly relevant when the cost of reducing the probability of extinction of a population is directly proportional to the probability itself [see Weitzman (1993) for further discussion on the singularities of equations 3 and 4].

## Genetic Extinction

All the previous definitions are based on one-period probabilities of extinction of each population—which ought to be known. However, this is not as easy as it sounds. There is no objective way of determining this for a common species, population, or breed. Therefore the information obtained

depends on the subjectiveness of the probability values, which is not very desirable. Scientific, rigorous, and objective probabilities are required if the above methodology is to be correctly applied. Instead of the concept of physical extinction, it is proposed here that genetic extinction be used. The concept is even more appropriate if it is taken into account that it is genetic diversity that is being referred to in a framework of genetic resources conservation programs. In addition, this introduces within-breed information from allelic frequencies into the analysis.

The concept of genetic extinction is closely related to that of homozygosity, since the genetic extinction of a population is equivalent to a homozygosity rate of one. The more homozygous a population is, the more genetically endangered. But homozygosity is, in practice, nothing more than an average of a number of homozygous individuals taken across a series of representative loci. Therefore, by analogy, an average can be calculated for the allelic situation of the loci under study for each breed. There are two alternatives. One is to average the probability of fixation of the loci at time  $t$  and use that to calculate the marginal diversity and conservation potential. The other is to perform the analysis on each locus with its corresponding fixation probability, obtain a different value of marginal diversity and conservation potential for each, and average those values across the loci. The extinction probabilities can be loosely extended to the point in which allelic frequencies exceed a certain value  $\alpha$  instead of just equaling one.

For ease of notation, index  $i$  will be dropped in the following formulas, which is referred to a generic population, but all the calculations will be ulteriorly applied to each of the  $N_p$  populations to obtain the corresponding extinction probabilities.

Consider a set of  $L$  loci, and for locus  $l$ , let  $\{A'_1, \dots, A'_n\}$  be the alleles,  $l = 1, \dots, L$ . For each  $m = 1, \dots, n_l$  and  $l = 1, \dots, L$ , let  $p'_m(t)$  be the frequency of the allele  $A'_m$  in generation  $t$  in a certain population. For the first option,  $P_{ext}(t)$  is therefore calculated as

$$P_{ext}(t) = \frac{1}{L} \sum_{l=1}^L P \left[ \bigcup_{m=1}^{n_l} (p'_m(t) \geq \alpha) \right] = \frac{1}{L} \sum_{l=1}^L \sum_{m=1}^{n_l} P[p'_m(t) \geq \alpha], \quad (5)$$

the second equality holding because the events are disjointed (if one allele becomes fixed or has a frequency higher than  $\alpha$ , for  $\alpha > 0.5$ , no other allele can do the same). For the second option, a set of  $L$  probabilities  $\{P_{ext}(t, l)\}_{l=1, \dots, L}$  can be calculated:

$$P_{ext}(t, l) = P \left[ \bigcup_{m=1}^{n_l} (p'_m(t) \geq \alpha) \right] = \sum_{m=1}^{n_l} P[p'_m(t) \geq \alpha], \quad l = 1, \dots, L. \quad (6)$$

The change in gene frequencies over the generations is an intricate process that depends on very different factors.

However, diffusion processes allow explicit expressions for  $P[p'_m(t) \geq \alpha]$  in many cases. To illustrate the approach introduced in this article, the particular situation of populations solely under the effect of genetic drift will be explained. In this case, it can be proven that

$$P_{ext}(t) = \frac{1}{L} \sum_{l=1}^L \sum_{m=1}^{n_l} \left[ p'_m(0) + \sum_{i=1}^{\infty} (2i+1)p'_m(0) \times (1 - p'_m(0))F(i+2, 1-i, 2, p'_m(0)) \times [F(-i, i+1, 1, \alpha) - F(-i, i+1, 1, 1)] + (-1)^i \right] e^{-\frac{i(i+1)t}{4N}} \quad (7)$$

and

$$\{P_{ext}(t, l)\}_{l=1, \dots, L} = \left\{ \sum_{m=1}^{n_l} \left[ p'_m(0) + \sum_{i=1}^{\infty} (2i+1)p'_m(0) \times (1 - p'_m(0))F(i+2, 1-i, 2, p'_m(0)) \times [F(-i, i+1, 1, \alpha) - F(-i, i+1, 1, 1)] + (-1)^i \right] e^{-\frac{i(i+1)t}{4N}} \right\}_{l=1, \dots, L}, \quad (8)$$

where  $\{p'_m(0)\}_{m=1, \dots, n_l}^{l=1, \dots, L}$  are the present-time allelic frequencies of loci 1 to  $L$ , and  $F(\cdot, \cdot, \cdot, \cdot)$  is the hypergeometric function.

Equations 7 and 8 are calculated for a single population and require the allelic frequencies of all the marker loci involved in the study. For a whole set of populations, a set of probabilities exists,  $\{P_{ext}^i(t)\}_{i=1, \dots, N_p}$  (or  $\{P_{ext}^i(t, l)\}_{i=1, \dots, N_p}^{l=1, \dots, L}$ ), each of them calculated according to equation 7 or 8. Observe that these probabilities are not conceptually the same as those in  $\{P_i\}_{i=1, \dots, N_p}$ , since the latter are one-generation extinction probabilities, while the former are the direct probabilities of extinction at generation  $t$ .

Equations 7 and 8 allow that which is desired: to have at our disposal a series of objective probability values based on genetic data. Furthermore, with this approach, within-population information as well as information on population sizes is introduced into the analysis.

For ease of notation, only the first case, that of averaged probability values across loci will be analyzed, so probabilities,  $\{P_{ext}^i(t)\}_{i=1, \dots, N_p}$ , obtained for each population according to equation 7, will be used. Thus, for a subset  $S$  of the whole set of populations  $\mathcal{Q}$ , equation 2 is substituted by

$$P_t(S) = \prod_{i \in S} (1 - P_{ext}^i(t)) \prod_{i \in \mathcal{Q} \setminus S} P_{ext}^i(t) \quad (9)$$

in equation 1 to calculate the expected diversity. The definitions of marginal diversity and elasticity would be analogously modified, so the differentials would be made with respect to  $P_{ext}^i(t)$  instead of  $P_i$ .

Note that with this model,  $ED_t$  is a linear function of each  $P_{ext}^i(t)$  separately. This is, if for a given  $i$ , we consider

$P_{ext}^i(t)$  as variable and the rest,  $P_{ext}^j(t), i \neq j \in \{1, \dots, N_p\}$  as fixed, then

$$\begin{aligned}
 ED_t(i) &= \sum_{S \in \wp(Q)} D(S)P_t(S) \\
 &= \sum_{S \in \wp(Q)} D(S) \prod_{j \in S} (1 - P_{ext}^j(t)) \prod_{j \in Q \setminus S} P_{ext}^j(t) \\
 &= \begin{cases} \sum_{S \in \wp(Q)} D(S) (1 - P_{ext}^i(t)) \prod_{\substack{j \in S \\ j \neq i}} (1 - P_{ext}^j(t)) \\ \quad \times \prod_{\substack{j \in Q \setminus S \\ j \neq i}} P_{ext}^j(t), & \text{if } i \in S \\ \sum_{S \in \wp(Q)} D(S) P_{ext}^i(t) \prod_{\substack{j \in S \\ j \neq i}} (1 - P_{ext}^j(t)) \prod_{\substack{j \in Q \setminus S \\ j \neq i}} P_{ext}^j(t), \\ \text{if } i \notin S \end{cases} \tag{10}
 \end{aligned}$$

Therefore  $MD_t(i)$ , being the partial differential of  $ED_t$  with respect to  $P_{ext}^i(t)$ , is a constant value for each  $i = 1, \dots, N$ . In fact,

$$\begin{aligned}
 MD_t(i) &= \begin{cases} \sum_{S \in \wp(Q)} \frac{\partial}{\partial P_{ext}^i(t)} D(S) (1 - P_{ext}^i(t)) \prod_{\substack{j \in S \\ j \neq i}} (1 - P_{ext}^j(t)) \\ \quad \times \prod_{\substack{j \in Q \setminus S \\ j \neq i}} P_{ext}^j(t), & \text{if } i \in S \\ \sum_{S \in \wp(Q)} \frac{\partial}{\partial P_{ext}^i(t)} D(S) P_{ext}^i(t) \prod_{\substack{j \in S \\ j \neq i}} (1 - P_{ext}^j(t)) \\ \quad \times \prod_{\substack{j \in Q \setminus S \\ j \neq i}} P_{ext}^j(t), & \text{if } i \notin S \end{cases} \\
 &= \begin{cases} \sum_{S \in \wp(Q)} -D(S) \prod_{\substack{j \in S \\ j \neq i}} (1 - P_{ext}^j(t)) \prod_{\substack{j \in Q \setminus S \\ j \neq i}} P_{ext}^j(t), & \text{if } i \in S \\ \sum_{S \in \wp(Q)} D(S) \prod_{\substack{j \in S \\ j \neq i}} (1 - P_{ext}^j(t)) \prod_{\substack{j \in Q \setminus S \\ j \neq i}} P_{ext}^j(t), & \text{if } i \notin S \end{cases} \tag{11}
 \end{aligned}$$

so therefore

$$MD_t(i) = - \sum_{S \in \wp(Q)} D(S) P_t^*(S \setminus i) (-1)^{I_S(i)}, \tag{12}$$

where  $I_S(i)$  equals one when  $i \in S$  and zero when  $i \notin S$ , and  $P_t^*$  is defined over  $\wp(Q \setminus i)$  in the same way that  $P_t$  is over  $\wp(Q)$ , and consequently  $MD_t(i)$  does not actually depend on  $P_{ext}^i(t)$ . Had one-period extinction probabilities been employed, such as in Weitzman (1993), there would be a non-linear dependence of  $ED_t$  on each  $P_{ext}^i(t)$ , and  $MD_t(i)$  would indeed depend on  $P_{ext}^i(t)$ . Now, from equation 12,  $MD_t(i) = \sum_{\substack{S \in \wp(Q) \\ i \in S}} D(S) P_t^*(S \setminus i) - \sum_{\substack{S \in \wp(Q) \\ i \notin S}} D(S) P_t^*(S)$ , and, finally, equation 12 can be rewritten as

$$MD_t(i) = - \sum_{S \in \wp(Q \setminus i)} P_t^*(S) [D(S \cup i) - D(S)]. \tag{13}$$

**Table 1.** Breeds, countries of origin, and effective sizes

Breed	Country	Effective size
Alistana (ALI)	Spain	36
Asturiana de Montaña (ASM)	Spain	35
Asturiana de Valles (ASV)	Spain	89
Aubrac (AUB)	France	100
Gasconne (GAS)	France	100
Salers (SAL)	France	100
Sayaguesa (SAY)	Spain	21
Tudanca (TUD)	Spain	35

The expression in brackets recalls strongly the partial contribution of population  $i$ , calculated for every  $i = 1, \dots, N_p$  as  $PC_i = [D(Q) - D(Q \setminus i)]/D(Q)$ . In fact, if a random variable  $C_i: \wp(Q \setminus i) \rightarrow \mathbb{R}$  is defined such that for every  $S \subseteq Q \setminus i$ ,  $C_i(S) := D(S \cup i) - D(S)$  with probability  $P_t^*(S)$ , then the marginal diversity is the expected value of  $C_i$  at generation  $t$ , and, divided by the expected diversity at generation  $t$ , can be understood as an expected partial contribution at generation  $t$ .

### Application and Results

This procedure was tested with an example set of local French and Spanish cattle breeds from the European contract FAIR1 CT95 0702 project [see Cañón et al. (2001) for details]. Table 1 shows the names, origins, and effective sizes of the breeds.

Allelic frequencies from 16 microsatellite-type marker loci, obtained from a total of 50 animals per breed (Cañón et al. 2001), were used to calculate marginal diversities and elasticities. Goldstein et al. (1995) proposed a distance measure, the average squared distance, especially devised for markers with a high degree of polymorphism, such as microsatellites. Some of its properties include the fact that its expected value does not change over time under a genetic drift model, which allows for  $D_t(S) = D(S)$ . Although it is not a distance in the strict sense, since when measured from one population to itself it does not equal zero, the procedure can be applied without loss of generality. Thus this distance was applied to calculate the Weitzman diversity, and partial contributions of the breeds to the total current diversity,  $\{PC_i\}_{i=1, \dots, N_p}$  were obtained. Note that this distance is used just as an example, since the method presented here can be implemented with any distance, as long as its variation with time can be assessed effectively enough. Actually the distance measure does not necessarily have to be a genetic distance. It may involve morphological or other kinds of information, and it is up to the person in charge of putting the method into practice to decide which particular distance suits the situation under study.

Table 2 shows the partial contributions and the marginal diversities and conservation potentials obtained for 25, 35, and 50 generations, and an  $\alpha$  threshold of 0.90. Values were rescaled to 100. Other within-population variation statistics

**Table 2.** Mean expected (MEH) and observed (MOH) homozygosities, mean effective number of alleles (MENA), partial contributions (PC), and average probabilities of extinction ( $P_{ext}^i(t)$ ), marginal diversities ( $MD_i(t)$ ), and conservation potentials ( $CP_i(t)$ ) from averaged probabilities for  $t = 25, 35$ , and  $50$ , using the average squared distance

Breed	MEH	MOH	MENA	$P_{ext}^i(25)$	$P_{ext}^i(35)$	$P_{ext}^i(50)$	PC	$MD_i(25)$	$MD_i(35)$	$MD_i(50)$	$CP_i(25)$	$CP_i(35)$	$CP_i(50)$
SAL	0.369	0.420	4.35	0.030	0.050	0.084	23.778	23.830	23.846	23.817	15.038	14.182	14.100
AUB	0.389	0.431	4.68	0.029	0.054	0.093	19.009	19.054	19.088	19.123	11.723	12.208	12.532
GAS	0.292	0.370	3.89	0.005	0.014	0.034	25.686	25.673	25.652	25.565	2.780	4.256	6.154
ALI	0.319	0.371	4.01	0.079	0.151	0.267	5.817	5.818	5.827	5.856	9.653	10.469	11.025
ASM	0.295	0.348	4.53	0.076	0.146	0.263	6.634	6.566	6.528	6.495	10.522	11.351	12.028
ASV	0.317	0.344	6.49	0.025	0.043	0.075	5.025	5.051	5.087	5.206	2.630	2.615	2.755
SAY	0.293	0.346	3.88	0.194	0.326	0.504	8.184	8.165	8.149	8.128	33.137	31.618	28.853
TUD	0.349	0.404	3.74	0.119	0.192	0.306	5.868	5.842	5.823	5.809	14.517	13.301	12.553

are included, such as mean observed and expected homozygosity and mean effective number of alleles. The results from the two methods described above—average extinction probabilities and average marginal diversities and conservation potentials—are extremely similar (data not shown). Suffice it to say, the mean coefficient of variation of the marginal diversities and the conservation potentials were 0.004% and 0.09%, respectively. Therefore, for the sake of simplicity, only values for the first method (average extinction probability across loci) are shown. Table 2 also shows that the differences in the figures among the breeds for other traditional within-variation measures are too small to make clear distinctions.

According to the partial contributions, Gasconne and Salers are the two most contributing breeds, followed by Aubrac, and, by far, by the Sayaguesa breed. Alistana and Asturiana de Valles are the breeds with the lowest partial contributions.

Marginal diversities are almost identical to partial contributions. The ultimate reason is the very nature of the marginal diversity. As seen in equation 13, marginal diversity is a sort of expected partial contribution at the given generation. Now, in the weighted sum, the set  $Q \setminus i \in \emptyset(Q \setminus i)$  has a higher weight, since the probability of survival of all the populations in  $Q \setminus i$  is noticeably higher than the probability of any other combination of survival/extinction within the set of populations  $Q \setminus i$ . This happens especially in short to medium time intervals, because the fewer the number of generations, the lower the fixation probability for any allele and therefore the lower the extinction probability, making the product of the probabilities used to calculate each term of equation 13—except for  $Q \setminus i$ —much smaller. This fact may lead one to consider marginal diversity as a superfluous indicator once we already have partial contributions, but it may be seen so only in the short term, because as generations pass, the differences between partial contributions and marginal diversities increase. This increase is already suggested in Table 2, although it would be much more evident for a larger number of generations.

Conservation potentials offer a different prioritization scheme than marginal diversities. Because of its low effective size, Sayaguesa is now the highest scored breed, followed by Salers, due to its strong influence in the between-population variation. Salers and Aubrac, highly prioritized with the mar-

ginal diversity criterion, are also important with respect to the conservation potential. Excluding Sayaguesa, Salers and Aubrac would join Alistana, Asturiana de Montaña, and Tudanca in a middle group. Differences within this group tend to diminish notably as the time horizon increases. The figures for Gasconne are particularly noteworthy. Marginal diversity sets this breed as the first in the prioritization ranking, while its conservation potentials are remarkably low—only Asturiana de Valles shows lower values. This may possibly be due to the fact that Gasconne is uniformly the most distant breed, that is, for any certain breed, Gasconne has a higher pairwise distance with this breed than any other one (data not shown), so it scores high in partial contribution and marginal diversity, where between-breed information seems to have a greater weight. On the other hand, it has the lowest extinction probability of all the breeds (down to six times that of Salers and Aubrac; all three having the same effective size), and this pulls down its ranking in the conservation potential. Only Asturiana de Valles is below Gasconne. This might be due to the combination of having a relatively low extinction probability and being, as opposed to Gasconne, uniformly the closest breed, meaning that for any certain breed, Asturiana de Valles has a lower pairwise distance with this breed than any other one (data not shown). So, on one side, within-breed information says that its extinction probability is low, while on the other, between-breed information says that it is quite close to the rest of the breeds. Therefore it has the lowest conservation potential value.

## Discussion

Maintaining genetic variability is an important goal in the conservation of animal populations and Weitzman's procedure can be applied to achieve this. However, it has been criticized (Caballero and Toro 2002; Eding and Meuwissen 2001; Thaon d'Arnoldi et al. 1998) for not considering within-population variation. Only Ollivier and Foulley (2002) suggest the otherwise obvious, but always ignored, possibility of applying Weitzman's algorithm to obtain a within-diversity measure. To do this, the coefficient of molecular coancestry (Fabel et al. 2004) between individuals, which applies Malécot's (1948) definition, but uses the concept of identity

in state instead of identity by descent, could be used to build a matrix of genetic distances (actually  $1 - \text{Malécot's coefficient}$ ) between individuals within one population to compute the Weitzman diversity measure within each population. However, this way of calculating within-breed diversity is actually highly correlated ( $r = -0.985$  in the example presented in this article; data not shown) with within-breed heterozygosity. Other criteria have been proposed to estimate within-breed variation, such as allelic richness (Marshall and Brown 1975). Efficiency of heterozygosity or allelic richness depends on the interest in the short- or long-term selection response.

If no extinction probabilities, be they genetic based or not, and no future behavior of the diversity and each population's contributions to it are to be studied, other solutions to incorporate Weitzman's approach to a joint between- and within-population analysis have been suggested. Ollivier and Foulley (2002) proposed a linear combination of between-population diversity, calculated via the Weitzman approach, and within-population diversity, measured with the heterozygosity, as an aggregate diversity, similar in concept to the traditional aggregate genotype in animal breeding. Depending on the emphasis given to the different possible breeding objectives, for example, selection or crossbreeding, different weights can be justified. It can be shown (Ollivier and Foulley 2004) that when the weights of between-breed and within-breed contributions are proportional to  $F_{ST}$  (between-population differentiation index) and  $(1 - F_{ST})$ , breeds rank in a similar way when applying the method proposed by Petit et al. (1998) and Caballero and Toro (2002).

It is well known that more than 80% of the total variability in a domestic animal species is a consequence of the genetic differences among individuals within subpopulations, (e.g., breeds), so giving a relatively greater emphasis to this within-breed component means promoting the possibility of the selection response within breeds. However, the crossbreeding, to exploit the heterosis or the complementarity, is also an important genetic improvement strategy to be used in animal production, and their magnitudes are proportional to the genetic distances among breeds (Falconer and Mackay 1996:255). This should justify giving more emphasis to the between-breed component, as illustrated by Chaiwong and Kinghorn (1999).

Using extinction probabilities allows a different way of using Weitzman's methodology. Between- and within-genetic information are both included in the analysis, and the behavior of the diversity in future generations can be assessed. The two alternate definitions of extinction probabilities are computable without difficulty and the rationale behind them agrees with that of popular measures like homozygosity. The results from both procedures do not differ too much and allow clear differentiations among the breeds, something not possible with more classical measures of within-population variation, as can be deduced from Table 2. The close relationship between the rate of inbreeding and the probability of extinction defined here allows the application of Weitzman's methodology to classify populations by the effect on expected diversity when efforts (Sonesson and

Meuwissen 2001; Wang 1997; Wang and Hill 2000) are made to reduce their homozygosity increase per generation.

As discussed above, marginal diversities were similar to partial contributions in all breeds, which supports the concept of expected partial contributions. Nevertheless, there are upward trends in some of the breeds and downward trends in others, which reflects that although the most influential factor for the marginal diversity seems to be the weight of the breed in terms of diversity, extinction probabilities tend to force constant patterns of increase/decrease of the marginals. This perhaps excessive influence of the phylogenetic scene on marginal diversities might be overcome by introducing one-generation genetic extinction probabilities, such as in Weitzman's work (1993), and basing all the analysis at any generation on them. However, this is beyond the scope of this article.

Neither is it the aim of this article to decide or advise on which of the many available distance measures should be used to perform an analysis. On the contrary, the methodology presented here is introduced as a general procedure in which a distance measure is required, but its choice is not inherent to the procedure. However, the particulars of each individual application should be taken into account when choosing the distance to use, since different distances can actually lead to quite different conclusions. For example, since the effective size already has an important weight on the extinction probability, the more a distance weight drifts, the more the values for marginals and elasticities will correlate with drift-based measures such as heterozygosity. Different distances should be used depending on the kind of OTUs or taxas used. Some distances are better suited for species comparisons and others for close breed analysis, depending on the time since divergence and the drift-mutation model. The same applies to the solution of the Kolmogorov (1931) equation, used to obtain equations 7 and 8, which can be found for a number of different scenarios.

The example shows that the results for the partial contributions and marginal diversities in short- or medium-term applications are extremely similar. Since no within-population information is taken into account to obtain the partial contributions, it seems that the marginal diversity is not a very good indicator to account for both between- and within-population information in short- or medium-term applications. In the long term, the marginal diversity seems to depart from the value of partial contribution, so it should be interesting to consider it.

Regarding conservation potentials, figures show that effective size is an important factor, as proved by Sayaguesa's high values, but also that other important information is accounted for in the model, so the final result is a combination of all the genetic properties of the populations.

Also, as noted by Weitzman (1993), elasticity, or conservation potential, is the most trustworthy indicator when the mechanism of investment is directly related to extinction probability, since the marginal diversity is weighted by the extinction probability itself. Ideally an optimal study should include a functional relationship between costs and effects on extinction probability, so the problem could be examined in

terms of function optimization subject to a series of constraints. In the absence of this kind of information, a global view of the results must be taken to arrive at balanced advice. In addition, it must not be forgotten that many factors other than genetics are usually and necessarily taken into account when addressing the conservation of biological diversity, and the information provided by genetic diversity studies must not be considered the sole source of information for decision making. Nevertheless, the methodology introduced here gathers together different types of genetic data (between breed, within breed, and population size) and is a useful tool in diversity conservation assessment.

The kind of analysis presented is recommended when mid- or long-term diversity studies are proposed, since genes, on average, are not expected to suffer noticeable risk of fixation in short time periods, unless effective sizes are particularly small, in which case a lesser number of generations can be considered.

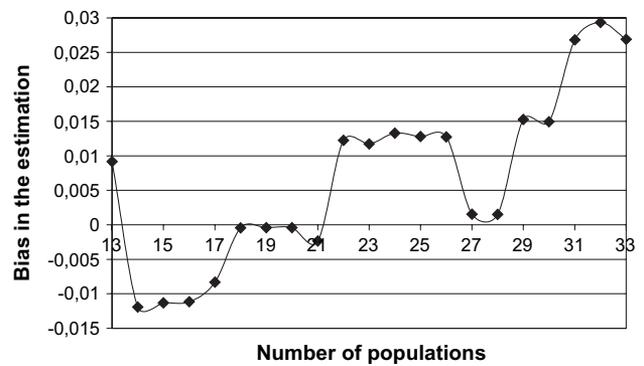
As a final consideration, the appendix details a new approximation of the Weitzman algorithm that allows for a greater number of populations to be introduced in the analysis and improves existing approximations.

## Appendix

The Weitzman algorithm (Weitzman 1992) is very computer time demanding, and computation time grows exponentially with the number of populations included in the analysis—log regression of the experimental computing time versus the number of populations provides an  $r^2$  of 0.998 (data not shown). For example, with an AMD processor at 2.4 GHz, computing only the diversity of 33 populations, without partial contributions or marginals, took almost 11 h, and time increases by an estimated factor of 2.14 (data not shown), so handling larger numbers of populations becomes infeasible. Therefore some kind of approximation of the exact algorithm must be used if the number of populations becomes larger than a few dozen.

Thaon d'Arnoldi et al. (1998) proposed an approximation consisting of randomly sampling trees among the  $2^{n-1}$  that the algorithm examines and taking as the approximated value of diversity the maximum of the values coming out of the sampled trees. This procedure works well with a relatively small number of populations—they tested it up to 29—but as the number of populations increases, so must the sample size. The problem is that while the exact algorithm can be implemented recursively, the approximated one cannot, and it comes to a point where, for reasonable precision in the estimation, the approximated algorithm actually takes more time than the exact one.

We propose a different approximation, and preliminary results seem to indicate that its behavior is rather satisfactory. In each step, the algorithm splits into two symmetrical recursive problems, each of which deals with a submatrix one dimension smaller than the input one. What we propose is, before inputting these submatrixes into the next step, checking their dispersion, and if all the values in one submatrix are



**Figure 1.** Plot of the bias between real diversity values and estimated ones, calculated as the estimated value minus the real value divided by the real value, for a limit of 0.25 in the coefficient of variation for the approximation.

close enough together around a mean value, to use this value, multiplied by the remaining number of steps, as the diversity value for that submatrix, and if not, then proceed with the exact computation in the usual way. The way to check for this “closeness” is to set a threshold to the variation coefficient (i.e., the standard deviation divided by the mean) of the values in the matrix. If this coefficient is lower than the threshold, then approximate, and if it is larger, then go on with the exact algorithm—which might eventually be truncated in some further step if the condition is met.

Figure 1 shows the bias we observed between the real and the approximated values of the diversity through increasing numbers of populations, from 13 to 33, and it can be seen that the bias never goes beyond 3% in absolute value. Calculations were made for a threshold on the variation coefficient of 0.25. Of course, the higher the threshold, the faster the computation, but with a higher bias attached, and vice versa. The behavior of the bias, as shown in Figure 1, is so irregular because the approximation is quite dependent on the data structure—the distance matrix in this case—but apparently it seems to provide good results. For 33 populations, the exact algorithm took, as mentioned above, almost 11 h to complete, while on the same computer the approximation took 2 min 40 s, with a bias of 2.7%. For a different dataset, 35 populations took 2 days minus 5 min until completion with the exact algorithm, and the approximation took less than 2 s with a threshold of 0.4, biasing the result less than 2%.

As promising as these results may seem, this procedure should be tested exhaustively before being applied to large datasets on the order of 50 or more populations, but that task is beyond the scope of this article.

## Acknowledgments

We thank A. Burton for providing language advice and the reviewers for their helpful comments on this article. This study received the financial support of FEDER project 2FD1997-1191 and European project CT-98-118.

## References

- Amos W and Balmford A, 2001. When does conservation genetics matter? *Heredity* 87:257–265.
- Barker JSF, 1999. Conservation of livestock breed diversity. *Anim Genet Resource Inform* 25:33–43.
- Caballero A and Toro M, 2002. Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv Genet* 3:289–299.
- Cañón J, Alexandrino P, Bessa I, Carleos C, Carretero Y, Dunner S, Ferran N, García D, Jordana J, Laloë D, Pereira A, Sanchez A, and Moazami-Goudarzi K, 2001. Genetic diversity measures of local European beef cattle breeds for conservation purposes. *Genet Sel Evol* 33:311–332.
- Chaiwong N and Kinghorn BP, 1999. Use of genetic markers to aid conservation decisions for groups of rare domestic breeds. *Proc Assoc Adv Anim Breed Genet* 13:365–368.
- Eding H and Meuwissen T, 2001. Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *J Anim Breed Genet* 118:141–159.
- Fabuel E, Barragán C, Silió L, Rodríguez MC, and Toro MA, 2004. Analysis of genetic diversity and conservation priorities in Iberian pigs based on microsatellite markers. *Heredity* 93:104–113.
- Falconer DS and Mackay TFC, 1996. An introduction to quantitative genetics, 4th ed. Harlow, UK: Longman.
- Fisher RA, 1930. The genetical theory of natural selection. Oxford: Clarendon Press.
- Frankham R, 1995. Inbreeding and conservation: a threshold effect. *Conserv Biol* 9:792–799.
- Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, and Feldman MW, 1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139:463–471.
- Hedrick PW and Kalinowski ST, 2000. Inbreeding depression in conservation biology. *Annu Rev Ecol Syst* 31:139–162.
- Kolmogorov A, 1931. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math Ann* 104:415–458.
- Lacy RC, 1997. Importance of genetic variation to the viability of mammalian populations. *J Mamm* 78:320–335.
- Laval G, Iannuccelli N, Legault C, Milan D, Groenen MAM, Giuffra, E, Andersson L, Nissen PH, Jorgensen CB, Beeckmann P, Geldermann H, Foulley JL, Chevalet C, and Ollivier L, 2000. Genetic diversity of eleven European pig breeds. *Genet Sel Evol* 32:187–203.
- Malécot G, 1948. Les Mathématiques de l'Hérédité. Paris: Masson et Cie.
- Marshall DR and Brown AHD, 1975. Optimum sampling strategies in genetic conservation. In: *Crop genetic resources for today and tomorrow* (Frankel OH and Hawkes JG, eds). Cambridge: Cambridge University Press; 53–80.
- Oldenbroek JK, 1999. Genebanks and the conservation of farm animal genetic resources. Lelystad, The Netherlands: DLO Institute for Animal Science.
- Ollivier L and Foulley JL, 2002. Some suggestions on how to preserve both within- and between-breed genetic diversity. In: *Book of Abstracts of the 53rd Annual Meeting of the European Association for Animal Production*, Cairo, Egypt, 1–4 September 2002 (van der Honing Y, ed). Wageningen, The Netherlands: Wageningen Academic Publishers.
- Ollivier L and Foulley JL, 2004. Objectives in livestock diversity preservation: the European pig example. In: *Wissenschaftliches Kolloquium "Nutztierzüchtung im Wandel der Zeit."* Göttingen, Germany: Cuvillier Verlag; 87–106.
- Petit RJ, El Mousadik A, and Pons O, 1998. Identifying populations for conservation on the basis of genetic markers. *Conserv Biol* 12:844–855.
- Reist-Marti SB, Gibson J, Rege JEO, Simianer H, and Hanotte O, 2003. Weitzman's approach and conservation of breed diversity: an application to African cattle breeds. *Conserv Biol* 17:1299–1311.
- Ruane J, 2000. A framework for prioritizing domestic animal breeds for conservation purposes at the national level: a Norwegian case study. *Conserv Biol* 14:1385–1393.
- Sonesson AK and Meuwissen THE, 2001. Minimization of rate of inbreeding for small populations with overlapping generations. *Genet Res* 77: 285–292.
- Thaon d'Arnoldi C, Foulley JL, and Ollivier L, 1998. An overview of the Weitzman approach to diversity. *Genet Sel Evol* 30:149–161.
- Wang J, 1997. Effective size and F-statistics of subdivided populations. 2. Dioecious species. *Genetics* 146:1465–1474.
- Wang J and Hill WG, 2000. Marker assisted selection to increase effective population size by reducing Mendelian segregation variance. *Genetics* 154:475–489.
- Weitzman ML, 1992. On diversity. *Q J Econ* 107:363–405.
- Weitzman M, 1993. What to preserve? An application of diversity theory to crane conservation. *Q J Econ* 108:157–183.

Received June 23, 2004

Accepted June 23, 2005

Corresponding Editor: Leif Andersson