

# An overview of the Weitzman approach to diversity

Caroline Thaon d'Arnoldi, Jean-Louis Foulley\*, Louis Ollivier

Station de génétique quantitative et appliquée,  
Institut national de la recherche agronomique,  
78352 Jouy-en-Josas cedex, France

(Received 10 July 1997; accepted 30 January 1998)

**Abstract** – The diversity of a set of breeds or species is defined in the Weitzman approach by a recursion formula using the pairwise genetic distances between the elements of the set. The algorithm for computing the diversity function of Weitzman is described. It also provides a taxonomy of the set which is interpreted as the maximum likelihood phylogeny. The theory is illustrated by an application to 19 European cattle breeds. The possible uses of the method for defining optimal conservation strategies are briefly discussed.

© Inra/Elsevier, Paris

**diversity / taxonomy / conservation / phylogeny / genetic distance**

**Résumé** – Un aperçu sur l'approche de la diversité selon Weitzman. La diversité d'un ensemble d'espèces, ou de races, est définie par Weitzman de façon récursive ; les données de départ sont les distances génétiques entre les éléments de l'ensemble pris deux à deux. L'algorithme de calcul de la diversité fournit, comme résultat intermédiaire, un arbre de classement des espèces en présence, qui est interprété comme une phylogénie du maximum de vraisemblance. La théorie est illustrée par un exemple d'application à 19 races bovines européennes, et les utilisations possibles de la méthode pour définir des stratégies optimales de conservation sont discutées brièvement. © Inra/Elsevier, Paris

**diversité / taxonomie / conservation / phylogénie / distance génétique**

## 1. INTRODUCTION

The question of preserving biological diversity is currently attracting a great deal of attention. Choices are necessary when it comes to deciding which endangered species must be protected and which not. Conserving breeds of farm animals, or domestic animal diversity, presents strong analogies with the more general question of preserving biological diversity. In both cases, owing to the limited resources

---

\* Correspondence and reprints  
E-mail: foulley@jouy.inra.fr

which can be devoted to conservation, the central question is 'what to preserve' [6]. The choices are difficult and it would be much easier if an operational theoretical framework based on this concept of 'diversity' were available. As noted by Solow et al. [5], this concept of diversity itself appears to have not so far been precisely defined, apart from a few attempts which can be traced back to May [3].

An analytical framework able to guide actual conservation policy in a diversity-improving direction through the use of a diversity function has been provided by Weitzman, an economist, who has given an example of application to the problem of crane species conservation [8–10]. Since his theory is recent and almost unknown to animal geneticists (see, however, Cunningham [1] and Ollivier [4]), and as it has not yet been used in the context of livestock breed diversity, we found it useful to describe it briefly and, as an illustration, to apply it to a set of cattle breeds.

## 2. THEORY

The method applies to 'elements' which may represent species, breeds, subspecies or any other operational taxonomic unit. Pairwise distances between elements are given, presenting basic properties of positivity, symmetry and nil distance of an element to itself. It is concerned with diversity between units; the theory ignores diversity due to variation within units.

### 2.1. Computing diversity

Computing diversities is straightforward if one knows how much the addition of one element, say  $j$ , increases the diversity of a given set  $Q$ . Intuitively, the magnitude of the gain should be related to how different the new element is from the set  $Q$ ; the more different  $j$  is from  $Q$ , the greater the gain. This difference is measured by the distance  $d(j, Q)$ . Here, the distance from a point  $j$  to a set  $Q$  is defined, as usual in set theory, by  $\min_{i \in Q} d(i, j)$ , in other words, the distance between  $j$  and its closest neighbour in  $Q$ .

More precisely, the intuitive property of the diversity function (which will be called  $V$  from now on) is the 'monotonicity in species': the gain of one element increases the diversity by at least  $d(j, Q)$

$$V(Q \cup j) \geq V(Q) + d(j, Q) \quad (1)$$

However, this is too loose a property to define a unique function. In fact, we will consider (1) as general conditions to satisfy for any member  $i$  withdrawn from the whole set  $S$ , i.e.

$$V(S) \geq V(S \setminus i) + d(i, S \setminus i), \quad \forall i \in S \quad (2)$$

where  $\setminus$  is the complement set symbol, i.e. here  $S \setminus i$  stands for  $S$  without  $i$ .

Let  $V'_i$  be defined as  $V'_i = V(S \setminus i) + d(i, S \setminus i)$ . For a given set  $S$ , the value of  $V'$  will depend on the element  $i$  chosen so that  $V(S)$  should verify:

$$V(S) \geq V'_1, V(S) \geq V'_2, \dots, V(S) \geq V'_i$$

If such a condition holds for the largest  $V'_i$ , it will also be true for all the other ones since:

$$V(S) \geq \max_{i \in S} V'_i \implies V(S) \geq V'_i, \quad \forall i \in S \quad (3)$$

According to (2), all the functions having larger values than  $V'_i$  also meet the criterion; to make the definition of  $V(S)$  unique, it will be restricted to the lowest one (minimum of  $V$ ), i.e. precisely to that equal to  $V'_i$ . This leads to the recursive definition of the Weitzman diversity function as:

$$V(S) = \max_{i \in S} [V(S \setminus i) + d(i, S \setminus i)] \quad (4)$$

with the initial conditions

$$V(i) = K$$

The value of  $K$  is taken by Weitzman [8, 9] as a normalizing constant which computationally can be set to zero.

Equation (4) provides a unique function having some interesting properties:

- the ‘twin property’: the addition of an element which is identical to an element of  $S$  does not increase  $V$ ;
- the monotonicity in species [see (1)];
- the continuity in distances: if the pairwise distances in set  $S$  are slightly modified, the modification of diversity is slight too;
- the monotonicity in distances: if every pairwise distance in set  $S$  is increased, the diversity of  $S$  increases too.

These properties are fundamental. They have the merit to remove ambiguity and to lay down the definition of diversity on simple and rigorous principles. In particular, the property of continuity in distances is of critical importance for any utilization of the results, given that there is some uncertainty on the real values of the pairwise distances.

## 2.2. The fundamental representation theorem

The dynamic programming recursion of equation (4) involves  $n!$  calculations,  $n$  being the number of elements. Fortunately, the following property allows us to reduce this computation to  $2^n$  calculations. The dynamic programming recursion produces, as a secondary result, a graphical representation of the relations between the elements.

### 2.2.1. Link property

By definition, and as shown previously, there exists an element  $i$  in any set  $S$  for which the maximum of equation (4) is achieved:

$$V(S) = V(S \setminus i) + d(i, S \setminus i) \quad (5)$$

Weitzman has shown that the element  $i$  in  $d(i, S \setminus i)$  is one of the two closest neighbours in  $S$ , i.e.  $d(i, S \setminus i) = \min_{u, v \in S} d(u, v)$ . In other words, there exists an

element  $i$  in  $S$  the loss of which involves a minimal reduction of diversity equal to  $d(i, S \setminus i)$ . This element is called the link.

### 2.2.2. Theorem

Having identified such a pair  $(i, j)$ , how will we know which one is the link? Remember from (3) that  $V(S) = \max(V'_i, V'_j)$ . Now  $V'_i = d(i, j) + V(S \setminus i)$ , and  $V'_j = d(i, j) + V(S \setminus j)$  so that the link is the element satisfying  $\max\{V(S \setminus i), V(S \setminus j)\}$ .

The dynamic programming recursion becomes:

$$V(S) = d[g(S), h(S)] + \max\{V[S \setminus g(S)], V[S \setminus h(S)]\} \quad (6)$$

where, using Weitzman's notations, the element  $g(S)$ , satisfying  $\max[V(S \setminus g), V(S \setminus h)]$  is called the link, the other one,  $h(S)$ , is the representative.

A proof of the theorem can easily be written by mathematical induction with respect to the size of the set  $S$ .

### 2.2.3. Algorithm and graphical representation by a taxonomic tree

Applying equation (6) recursively generates a rooted directed tree whose twig-tips are the elements of the set  $S$  and the nodes are the unknown 'ancestors'.

The different steps of the algorithm to be applied recursively are (beginning with the value of diversity set to zero):

- i) find the two closest neighbours  $i$  and  $j$  among the elements of  $S$  and add  $d(i, j)$  to diversity;
- ii) determine the link  $g$  and the representative  $h$  by using the property:

$$g = \operatorname{argmax}\{V(S \setminus g), V(S \setminus h)\};$$

- iii) given  $V(S) = d(g, h) + V(S \setminus g)$ , consider a new set without the link  $g$ , i.e.  $S \setminus g$ ;

- iv) return to i) until the size of the current set reaches 1; then add the constant  $K$  defined in (4) to diversity and stop.

While drawing the tree, it is useful to place the link  $g$  between the representative  $h$  and the closest neighbour of  $h$  in  $Q \setminus g$ ,  $Q$  being the subset whose diversity is computed at this step. Intuitively, it means that the loss of the link is less consequential for the diversity than the loss of any other element. It presents the advantage of allowing only one symmetry through the possible representations for the tree, while most hierarchical clustering methods result in a number of possible representations by rotation of the branches. The diversity of the set  $S$  can be read on the tree as the sum of the branch lengths, or the sum of the ancestor ordinates.

Weitzman also showed that the particular tree generated by the dynamic recursion algorithm in (6) and steps i-iv can be interpreted as the tree maximizing the probability that all of elements of  $S$  exist at the current time (see Appendix).

An APL2 program has been written to run the computations on Unix and Microsoft platforms. It is available upon request from the authors.

### 2.2.4. Example

Let us consider a set of four primate species. Pairwise distances are given in the following matrix (data are provided by Weitzman [9]):

	Go	Or	HyL	HyS
Go	0	357	532	498
Or	357	0	477	488
HyL	532	477	0	126
HyS	498	488	126	0

The closest neighbours to be found in the set  $\{Go, Or, HyL, HyS\}$  are HyL and HyS.

$$V\{Go, Or, HyL, HyS\} = \max[V\{Go, Or, HyL\}, V\{Go, Or, HyS\}] + d(HyS, HyL)$$

Now we need to know which element is the link in the couple (HyL, HyS).

The following matrices contain pairwise distances for the subsets  $\{Go, Or, HyL\}$  and  $\{Go, Or, HyS\}$ :

	Go	Or	HyL		Go	Or	HyS	
Go	0	357	532		Go	0	357	498
Or	357	0	477		Or	357	0	488
HyL	532	477	0		HyS	498	488	0

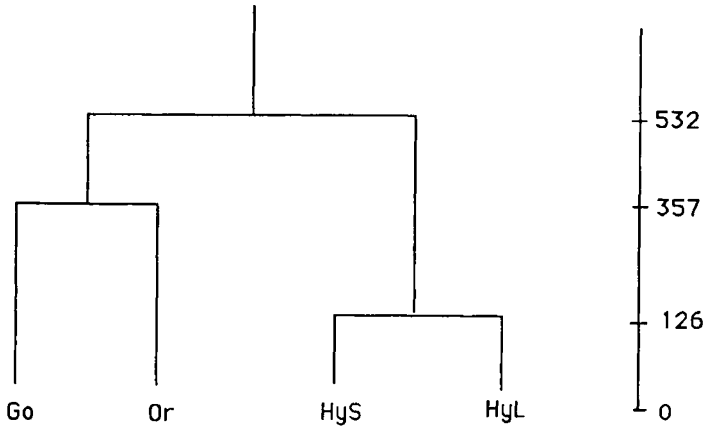
$$\begin{aligned} V\{Go, Or, HyL\} &= d(Go, Or) + \max[V\{Go, HyL\}, V\{Or, HyL\}] \\ &= d(Go, Or) + d(Go, HyL) \quad (\text{so Or is the link element in} \\ &\hspace{15em} \{Go, Or, HyL\}) \\ &= 889 \end{aligned}$$

$$\begin{aligned} V\{Go, Or, HyS\} &= d(Go, Or) + \max[V\{Or, HyS\}, V\{Go, HyS\}] \\ &= d(Go, Or) + d(Go, HyS) \quad (\text{so Or is the link element in} \\ &\hspace{15em} \{Go, Or, HyS\}) \\ &= 855 \end{aligned}$$

$V\{Go, Or, HyL\} > V\{Go, Or, HyS\}$ , thus we have determined that the link element in the couple (HyL, HyS) is HyS, and consequently the representative is HyL. Considering the remaining set after the suppression of the link element, i.e.  $\{Go, Or, HyL\}$  we found that the closest neighbours are (Go, Or), with Or as the

link element. This information then makes it possible to compute the total diversity, which is worth  $1015 = d(\text{Go}, \text{HyL}) + d(\text{Go}, \text{Or}) + d(\text{HyL}, \text{HyS})$ , and to draw the corresponding taxonomic tree (*figure 1*).

The link HyS in  $\{\text{Go}, \text{Or}, \text{HyL}, \text{HyS}\}$  is placed between the representative HyL and the closest neighbour Or of HyL in  $\{\text{Go}, \text{Or}, \text{HyL}\}$ . The link Or in  $\{\text{Go}, \text{Or}, \text{HyL}\}$  is then placed between the representative Go and the closest neighbour HyL of Go in  $\{\text{Go}, \text{HyL}\}$ , resulting in a final order of Go, Or, HyS, HyL.



**Figure 1.** Dendrogram of relationships among four species of primates.

### 3. APPLICATION: EXAMPLE OF EUROPEAN CATTLE BREEDS

#### 3.1. Evaluation of diversity

The Weitzman method has been applied to data collected by F. Grosclaude [2] on biochemical polymorphisms (11 blood group loci and the locus of blood serum transferrin and that of beta-casein) of 19 European cattle breeds, including 18 French breeds and the British Shorthorn. This latter was included because of its Durham ancestor that has been introduced in some French regions during the last century. The authors calculated the Nei standard distances considering the 13 polymorphic loci (*table I*). Results of the different steps of the computations of diversity are shown in *table II*.

The graphical representation of the result is shown in *figure 2*. A clear discrimination is observed between two groups i.e. i) a first group made of Northern dairy breeds (Frisonne, Flamande, Maine Anjou, Shorthorn) and ii) another group involving beef and hardy breeds of the Center and West part of France (Salers, Aubrac, Limousine, Charolais, Ferrandaise, Blonde d'Aquitaine) as well as Western and Eastern dual purpose breeds (e.g. Pie Rouge, Abondance, Tarentaise, Brune des Alpes, Bretonne Pie-Noire, Montbéliarde and Parthenaise); the original location of the Normande breed between those two groups as already mentioned by Grosclaude et al. [2] should also be noted.

Table I. Nei standard<sup>a</sup> distances among cattle breeds (from Grosclaude et al. [2]).

	NO	CH	PA	BR	FE	MA	MO	VO	PR	TA	AU	AB	SA	BU	FL	FR	BL	LI	SH
NO	0	824	455	449	863	720	487	589	717	585	1003	751	972	441	617	714	706	801	1115
CH	824	0	376	481	175	739	291	445	289	361	284	336	379	472	439	788	166	309	1100
PA	455	376	0	254	306	606	223	440	362	240	514	312	436	308	298	733	213	346	960
BR	449	481	254	0	389	530	244	302	313	172	518	233	691	141	195	356	348	466	1040
FE	863	175	306	389	0	849	280	535	247	298	171	254	216	404	459	864	124	157	1290
MA	720	739	606	530	849	0	648	401	519	662	1045	906	1254	525	273	226	826	838	255
MO	487	291	223	244	280	648	0	374	251	254	365	241	439	244	335	616	229	289	1066
VO	589	445	440	302	535	401	374	0	331	362	658	342	771	366	285	315	485	573	834
PR	717	289	362	313	247	519	251	331	0	267	381	331	450	282	255	488	284	283	971
TA	585	361	240	172	298	662	254	362	267	0	338	178	374	154	273	588	263	395	1275
AU	1003	284	514	518	171	1045	365	658	381	338	0	307	183	533	623	1062	304	194	1631
AB	751	336	312	233	254	906	241	342	331	178	307	0	377	222	372	705	268	404	1533
SA	972	379	436	691	216	1254	439	771	450	374	183	377	0	632	790	1373	304	226	1772
BU	441	472	308	141	404	525	244	366	282	154	533	222	632	0	191	338	364	522	1103
FL	617	439	298	195	459	273	335	285	255	273	623	372	790	191	0	178	440	495	680
FR	714	788	733	356	864	226	616	315	488	588	1062	705	1373	338	178	0	862	954	640
BL	706	166	213	348	124	826	229	485	284	263	304	268	304	364	440	862	0	289	1201
LI	801	309	346	466	157	838	289	573	283	395	194	404	226	522	495	954	289	0	1255
SH	1115	1100	960	1040	1290	255	1066	834	971	1275	1631	1533	1772	1103	680	640	1201	1255	0

<sup>a</sup> Values multiplied by 10 000.

Table II. Ordinates of the nodes.

Representative ( <i>h</i> )	Link ( <i>g</i> )	$d(g, h)$	Diversity ( <i>V</i> )
BL	FE	124	124
BR	BU	141	265
CH	BL	166	431
BR	TA	172	603
FR	FL	178	781
SA	AU	183	964
PA	MO	223	1187
FR	MA	226	1413
SA	LI	226	1639
AB	BR	233	1872
CH	PR	289	2161
PA	AB	312	2473
FR	VO	315	2788
PA	CH	376	3164
SA	PA	436	3600
SH	FR	640	4240
SA	NO	972	5212
SA	SH	1772	6984

## Abbreviations:

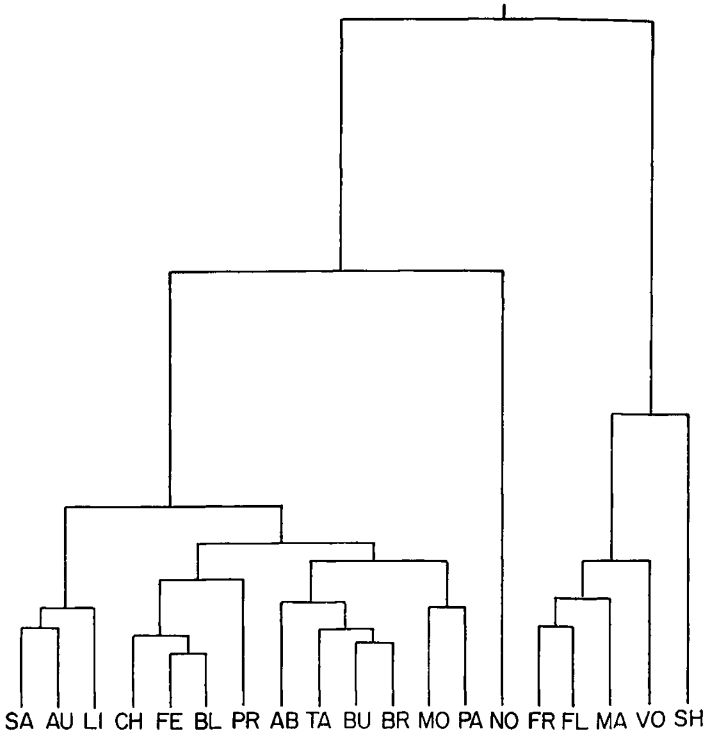
Normande	NO	Tarine	TA
Charolaise	CH	Aubrac	AU
Parthenaise	PA	Abondance	AB
Bretonne Pie Noire	BR	Salers	SA
Ferrandaïse	FE	Brune	BU
Maine Anjou	MA	Flamande	FL
Montbéliarde	MO	Frisonne Pie Noire	FR
Vosgienne	VO	Blonde d'Aquitaine	BL
Pie Rouge de l'Est	PR	Limousine	LI
		Shorthorn	SH

Current population sizes in some of those breeds are so restricted that they are said to be endangered: e.g. Bretonne Pie Noire, Ferrandaïse, Vosgienne or the Shorthorn.

The Weitzman method allows us to quantify the loss of diversity caused by the extinction of any subset among the 19 original breeds. By looking at the tree it is evident that the extinction of the Shorthorn causes a much greater loss of diversity than the extinction of the Flamande, whose distance from its closest neighbour, the Frisonne Pie Noire, is quite small.

By computing the diversities of the initial set of breeds and the set minus the Flamande, or the Shorthorn, or both the Flamande and the Shorthorn, one finds that the loss of the set Flamande + Shorthorn induces a reduction of diversity equal to the sum of the reductions caused by the loss of each of these breeds. This property of additivity is related to the degree of 'independence' between the two breeds. On the other hand, if the extinctions of the Montbéliarde and the Parthenaise were in





**Figure 2.** Dendograms of relationships among 19 cattle breeds (Weitzman's method).

question, the sum of the two would not exactly equal the magnitude of this joint loss. These breeds cannot be considered as independent from each other.

Diversity of the initial set	6 984	
Loss of Shorthorn + Flamande	5 509	(-1 475)
Loss of Shorthorn	5 687	(-1 297)
Loss of Flamande	6 806	(-178)
Loss of Montbéliarde + Parthenaise	6 418	(-566)
Loss of Montbéliarde	6 761	(-223)
Loss of Parthenaise	6 659	(-325)

The loss of diversity caused by the extinction of a set of breeds can be estimated by the sum of the ordinates of the nodes that would disappear from the tree if the extinct breeds were to be removed, without any other change. Thus, just by looking at the tree, it is obvious that the loss of the Normande would decrease the diversity eight or nine times more than the loss of the Blonde d'Aquitaine, and even more than the loss of a set including Charolaise, Ferrandaise and Blonde d'Aquitaine.

### 3.2. Further considerations on conservation strategies

The algorithm may be applied to evaluate the relative merit of breeds with small or medium population sizes regarding diversity. Let us consider the whole set (say  $Q$ ) of the 18 French cattle breeds analysed in this study, and that (say  $L$ ) of the six largest dairy (Française Frisonne, Montbéliarde and Normande) and beef breeds (Blonde d'Aquitaine, Charolaise and Limousine). The relative loss due to keeping those six breeds only is 57.2 %. Now one may ask which is the most interesting breed to select among the rest if any of them has to be preserved. This can be evaluated by considering the relative loss of diversity between  $Q$  and  $L$  plus each of those 12 breeds. Results based on Nei and (Cavalli-Sforza) distances are the following:

PA	52.3 (56.5)	MA	53.3 (54.9)	TA	50.9 (55.6)	SA	43.6 (45.8)
BR	51.8 (56.0)	VO	51.7 (55.1)	AU	51.0 (52.3)	BU	52.1 (57.1)
FE	55.1 (57.8)	PR	52.8 (54.7)	AB	51.0 (57.2)	FL	54.1 (56.1)

The breed providing the lowest loss of diversity is the Salers breed followed by the Aubrac. The ranking is consistent across the two distances used. Although this is only an illustration which would deserve further analysis including additional markers, this example is a significant one as those breeds have been recognized as key hardy breeds for a long time [7].

## 4. DISCUSSION AND CONCLUSION

The method presented provides several results with different degrees of robustness and different potential applications.

As indicated above, the value of diversity possesses a useful property of continuity in distances. The results may be considered as relevant to support decisions affecting the breeds or species to be preserved. The choice would be based only on objective computations, without relying on such subjective characteristics as beauty, interest for future or present generations or any other intrinsic criterium. Experience has shown that it is difficult to base priorities on such criteria.

The Weitzman approach to diversity allows further developments. Weitzman [10] suggests defining a diversity expected after a given period of time, based on the extinction probability of each element of the set considered. If  $n$  elements are endangered,  $2^n$  survival-extinction patterns may occur with given probabilities, and for each pattern the resulting diversity may be calculated. Weitzman then defines a 'marginal diversity' of each element, obtained as the partial derivative of the expected diversity with respect to the extinction probability of this element. The marginal diversity of breed  $i$  measures the relative gain in expected diversity (after 50 years say) from improving the survival probability of breed  $i$ . In a similar fashion, one could assume that the extinction of a breed can be completely avoided by using cryopreservation and calculate the gain in expected diversity obtained by cryopreserving each endangered breed. Knowing the pairwise genetic distances

and the risk status of a given set of endangered breeds as expressed through their respective probabilities of extinction, an order of priority for a cryopreservation programme could thus be established.

Because diversity is computed recursively, it involves very long calculations when the size  $n$  of the set is larger than 25. The approximation proposed in this study relies on a random choice of the link at each stage of the recursive algorithm, i.e. on sampling trees among the  $2^{n-1}$  possible trees. The procedure can be applied as follows: i) compute  $V$  among the elements of  $S$  by choosing at each step the link not from the formula in (6), but at random out of the pair of closest neighbours, ii) repeat i)  $m$  times such as to generate  $m$  different values of  $V$ , iii) take as the estimated value of  $V(S)$  the maximum value of  $V$  over all values computed. This can be performed by choosing at random  $m$  integers smaller than  $2^{n-1}$ , convert them into their binary expression and use the convention that the link will be the first element if the value is 0 and the second if it is 1. This procedure was tested on a set of 29 cattle breeds using data from Moazami-Goudarzi (pers. comm.). For  $m = 10\,000$ , the estimated value of  $V$  was at least of 13 200 as compared to a real value of 13 722, i.e. bias lower than 4 %. This approximation is quite good regarding the time of computation required by this estimation (20 min) while the complete algorithm needed more than 8 days.

On the other hand, the graphical representation might be sensitive to slight modifications of the distance matrix if the values of diversity are close for certain subsets. Simulation procedures to evaluate the robustness of clades have been proposed by Weitzman [8]. Although the clustering power looks satisfying on the examples we considered, any phylogenetic interpretation of the results should be used with caution. It should also be emphasized that the use made of genetic distances in this approach differs from their use in deriving genealogical trees. Though trees are useful geometric representations of diversity – the diversity function defined above is indeed equal to the total branch length of the corresponding tree – they must be considered as telling the evolutionary story that best fits the diversity observed, but not necessarily as telling the 'true' story. In fact, as emphasized by Weitzman [9], there is no need for the elements to have been generated by any real evolutionary phylogeny. This has to be kept in mind particularly when sets of domestic breeds are considered. Given the exchanges known to have occurred in their past histories, domestic breeds are indeed not likely to have resulted from a strict tree-like branching process. Whereas taxonomists are essentially interested in finding the evolutionary story behind a given observed diversity, conservationists, especially breed conservationists, do not need that type of information as they are more concerned with the future evolution of diversity.

The main use of the Weitzman method is to determine preservation strategies. It supposes, however, that the elements of the set considered are and remain distinct. If this constraint can be removed, it may be suggested that certain endangered breeds be amalgamated with other ones. The population size would increase, no additional costs would be engaged, and the direct loss of alleles that results from an extinction could be avoided. Of course, this implies that the breed standards should be relaxed for a while, but it is a dynamic conception of preservation that may offer interesting solutions in some cases.

Despite the criticisms which can be raised against the Weitzman approach, including that it ignores the differences in within unit variation, it should be kept in mind that it does satisfy certain basic properties which do not always hold with traditional criteria. The principle (1) of 'monotonicity in species' means that the change in diversity  $V(S \setminus i) - V(S)$  due to the loss of some population  $i$  is always negative or nil (for  $i$  being a twin element). In contrast, this property does not apply to variance, for it can be easily shown that the total variance of a mixture of populations can increase after some of them are deleted.

## ACKNOWLEDGMENTS

This work was conducted while Caroline Thaon was on a '*stage de fin d'études*' at the Station de génétique quantitative et appliquée (SGQA), Inra, Jouy-en-Josas as a student from the Ecole Polytechnique, Palaiseau. She greatly acknowledges the support of both institutions in making this stay feasible. Special thanks are expressed to F. Grosclaude and K. Moazami-Goudarzi (Laboratoire de génétique biochimique, Jouy-en-Josas) for providing the data on cattle analysed in this study. We are also grateful to C. Dillmann and P. Dubreuil (Inra, Station de génétique végétale, Le Moulon) and S. Lemarié (Inra-SERD, Grenoble) for having provided additional test-examples, and to Bruce Southey and an anonymous referee for their valuable comments which helped to improve the manuscript. E. Thompson is also thanked for her English revision of the text.

## REFERENCES

- [1] Cunningham P., Genetic diversity in domestic animals: strategies for conservation and development, in: Miller R.H., Pursel V.G., Norman H.D. (Eds.), XX Biotechnology's Role in the Genetic Improvement of Farm Animals, American Society of Animal Science, Savoy, IL, USA, 1996, pp. 13–23.
- [2] Grosclaude F., Aupetit R.Y., Lefebvre J., Mériaux J.C., Essai d'analyse des relations génétiques entre les races bovines françaises à l'aide du polymorphisme biochimique, *Genet. Sel. Evol.* 22 (1990) 317–338.
- [3] May R.M., Taxonomy as destiny, *Nature* 347 (1990) 129–130.
- [4] Ollivier L., Génétique et conservation animales, in: Matassino D., Boyazoglu J., Capuccio A. (Eds.), International Symposium on Mediterranean Animal Germplasm and Future Human Challenges, EAAP publication no. 85, Wageningen Pers, Wageningen, 1997, pp. 211–219.
- [5] Solow A., Polasky S., Broadus J., On the measurement of biological diversity, *J. Environ. Econom. Manag.* 24 (1993) 60–68.
- [6] Vane-Wright R.I., Humphries C.J., Williams P.H., What to protect? Systematics and the agony of choice, *Biol. Cons.* 55 (1991) 235–254.
- [7] Vissac B., Etude génétique de la race d'Aubrac, in: L'Aubrac, CNRS, Paris, I, 1970, pp. 29–102.
- [8] Weitzman M., A reduced form approach to maximum likelihood estimation of evolutionary trees, Harvard Institute of Economic Research, Paper No. 1569, 1991.
- [9] Weitzman M., On diversity, *Quart. J. Econ.* 107 (1992) 363–405.
- [10] Weitzman M., What to preserve? An application of diversity theory to crane conservation, *Quarter. J. Econ.* 108 (1993) 157–183.

**APPENDIX: the maximum likelihood tree**

Weitzman [8] provides the following phylogenetic interpretation. Let us note  $p(i, j)$  the conditional probability  $P(i|j)$  that a species  $i$  exists given that a species  $j$  exists. Assume that this probability is a function of the genetic distance between  $i$  and  $j$ . The hypothesis underlying this assumption is that the distance  $d(i, j)$  between two species  $i$  and  $j$  measures the time since their separation. More precisely, we will suppose that  $p(i, j) = \exp[-\lambda d(i, j)]$  where  $\lambda$  is a 'universal extinction rate'.

The maximum likelihood tree is the evolution scheme (i.e. the set of unknown ancestors) which maximizes the probability that every element of  $S$  exists at the current time.

Let  $P(j|i)$  be the conditional probability that species  $j$  exists given  $i$  exists. Assuming that the evolution scheme is known, it can be shown that, for any subset  $Q \in S$ , and  $J \in S \setminus Q$ , the conditional probability  $P(j|Q)$  that  $j$  survived given  $Q$  exists satisfies

$$P(j|Q) \leq \max_{i \in Q} P(j|i) \quad (\text{A.1})$$

Note  $p(j, Q) = \max_{i \in Q} P(j|i)$ . Now, from basic probability theory,  $P(j|Q) = P(Q \cup j)/P(Q)$ , and combining this with (A.1) leads to:

$$P(Q \cup j) \leq P(Q) \times p(j, Q) \quad (\text{A.2})$$

Let us note  $\Pi(S)$ , the largest probability that  $S$  exists, i.e. the probability of existence under the most favourable evolution scheme. Equation (A.2) applied for  $Q = S \setminus i$ , and  $j \equiv i$  implies

$$\Pi(S) \leq \min_{i \in S} [\Pi(S \setminus i) p(i, S \setminus i)] = \Pi^* \quad (\text{A.3})$$

Any evolution scheme that would induce a value of  $P(S) = \Pi^*$  would be identified as the scheme under which the probability that  $S$  exists is maximal, i.e. the maximum likelihood tree. Taking the logarithm of equation (A.3) and normalizing  $\lambda$  to 1, it becomes:

$$W(S) = -\ln \Pi^* \quad (\text{A.4})$$

$$W(S) = \max_{i \in S} \{W(S \setminus i) + d(i, S \setminus i)\} \quad (\text{A.5})$$

Since (A.5) has been studied above and solved by algorithm (6), we are able to exhibit such an evolution scheme. The tree generated by the Weitzman method can be interpreted as the maximum likelihood tree, i.e. the tree that maximizes the likelihood of the current survival pattern of the species.