



Analysis of genetic diversity for the management of conserved subdivided populations

Armando Caballero^{1,*} & Miguel A. Toro²

¹Departamento de Bioquímica, Genética e Inmunología, Facultad de Ciencias, Universidad de Vigo, 36200 Vigo, Spain; ²Departamento de Mejora, Genética y Biotecnología, INIA, Carretera de la Coruña Km. 7, 28040 Madrid, Spain (*Author for correspondence: Phone: 986-812568/78; Fax: 986-812556; E-mail: armando@uvigo.es)

Received 9 April 2001; accepted 16 July 2001

Key words: conservation, effective size, genetic drift, heterozygosity, inbreeding

Abstract

Recent studies in the literature have applied phylogenetic methods based on genetic distances to set priorities for conservation of domestic animal breeds. While these methods may be appropriate for between-species conservation, they are clearly inappropriate for within-species breed conservation, because they ignore within-breed variation. In this paper we show the basic tools to analyse genetic diversity in subdivided populations within species, and illustrate the errors incurred by applying methods based exclusively on genetic distances. We also show that maximisation of genetic diversity (minimisation of coancestry or kinship) is equivalent to maximisation of effective population size, as in undivided populations, and derive a generalisation of previous equations for the prediction of effective size. Finally, we discuss the strategies for conservation in the light of the theory.

Introduction

Conservation of genetic diversity is one of the main current issues in the conservation biology literature (Frankham 1995). In recent years, a number of quantitative measures of taxonomic diversity have been proposed in the context of species conservation (see review by Krajewski 1994). All these measures are based, in one way or another, on the genetic distances between species. The basic idea from a conservation perspective is that the extinction of a species in the wild produces a loss in diversity because the family to which the species belongs is diminished. The loss of diversity is greater when the extinct species has low relation to the surviving species than when its relation is high.

A biological unit considered for conservation may be quite different depending on the scope of interest (Crandall et al. 2000). For wild populations the unit may be the species or subspecies, whereas for domesticated plants and animals the unit may be the breed or strain. Thus, the above idea has also been applied to the conservation of livestock breeds. For

example, Barker (1999) has indicated that phylogenetic diversity (based on microsatellite loci) will provide the best objective criterion for making initial conservation decisions for livestock breeds, i.e. breeds that are taxonomically distinct should be favoured for conservation. However, a problem with these phylogenetic measures of diversity, also mentioned by Barker (1999), is that genetic variation within groups is completely or partially ignored. While this may not be a key issue in the context of species conservation, it may be of great importance for the management of breeds within species.

Recent studies have considered analyses of genetic diversity in domestic animal species using phylogenetic methods based on genetic distances among breeds. For example, Thaon d'Arnoldi et al. (1998) emphasised the use of the Weitzman (1992) method to determine conservation strategies. This approach gives a measure of the diversity of a set of elements (species, subspecies, breeds, etc.) ignoring within-element diversity. The solution takes the form of a maximum likelihood estimated phylogeny conditional on the model, so that the contribution of an element to

group diversity is proportional to the reduction in tree length caused by the removal of the element from the group. Thaon d'Arnoldi et al. (1998) and Laval et al. (2000) applied this method to a set of 19 cattle breeds and 11 pig breeds, respectively, indicating possible conservation strategies based on their solutions.

In this paper we show that the conservation decisions using genetic distance methods can be completely misleading when applied to the breeds of a given species or, in general, to the subpopulations of a given metapopulation. The reason is that the global diversity of the metapopulation should also be considered, i.e. both within and between subpopulation variability should be included in conservation decisions. We derive the necessary tools to analyse the diversity of a metapopulation when pedigrees or neutral molecular markers are available, and apply these tools to an example in the literature to show the consequences of applying methods based on genetic distances. Finally, we show that minimising coancestry in a subdivided population is equivalent to maximising effective population size, as has been already shown for undivided populations (Caballero and Toro 2000). In doing this, we derive a generalisation for the case of subdivided populations of the predictive equation of effective size by Wray and Thompson (1990).

Throughout the paper we will exclusively consider the genetic diversity extracted from genealogical relations (pedigrees) or data on neutral molecular markers. Other aspects of conservation decisions, such as data on particular traits of economic value, specific adaptive features, allelic richness, presence of unique genes or genotypes, and local or regional importance of a breed (Oldenbroek 1999; Barker 2001), are specific to each particular case, and will not be discussed.

Basic genetic tools for the analysis of genetic diversity in subdivided populations

We will first summarise the tools necessary for the analysis of genetic diversity in subdivided populations. These will be referred, when convenient, to those for undivided populations described previously (Caballero and Toro 2000). Assume a metapopulation consisting of n subpopulations, subpopulation i with N_i breeding individuals. Let f_{ij} be the average pairwise coancestry (kinship) (Malécot 1948) between individuals of subpopulations i and j , including all

$N_i \times N_j$ pairs (Cockerham 1967). (Here we assume that all coancestries are known through genealogical information back to the base population, that is, correspond to identity by descent coefficients.) Thus, the average coancestry of subpopulation i is f_{ii} . Let s_i be the average self-coancestry of the N_i individuals of subpopulation i (this was called f_N by Caballero and Toro (2000) but here is denoted by s to avoid confusion). The average coefficient of inbreeding of subpopulation i is $F_i = 2s_i - 1$, and the average distance between individuals of subpopulations i and j is $D_{ij} = [(s_i + s_j)/2] - f_{ij}$.

The averages for the entire metapopulation for the mean coancestry, self-coancestry and inbreeding coefficient of subpopulations, and distances between individuals within subpopulations are, respectively,

$$\begin{aligned} \bar{f} &= \frac{\sum_{i=1}^n f_{ii} N_i}{N_T}, \quad \bar{s} = \frac{\sum_{i=1}^n s_i N_i}{N_T}, \quad \bar{F} = \frac{\sum_{i=1}^n F_i N_i}{N_T}, \\ \text{and } \bar{D} &= \bar{s} - \bar{f} = \frac{\sum_{i=1}^n D_{ii} N_i}{N_T}, \end{aligned} \quad (1)$$

where $N_T = \sum_{i=1}^n N_i$ is the total metapopulation size. For subpopulation i the following relationships are met:

$$\alpha_i = \frac{F_i - f_{ii}}{1 - f_{ii}}, \quad G_i = \frac{s_i - f_{ii}}{1 - f_{ii}} = \frac{D_{ii}}{1 - f_{ii}}, \quad \text{and } -1 = \frac{F_i - s_i}{1 - s_i}, \quad (2)$$

Here α_i is the deviation from Hardy-Weinberg proportions and G_i is the proportion of diversity between individuals for subpopulation i , so that $(1 - \alpha_i) = 2(1 - G_i)$ and $(1 - f_{ii}) = (1 - s_i) + D_{ii}$ (see Caballero and Toro 2000).

The genetic distance between subpopulations i and j (Nei's minimum distance; Nei 1987) is

$$\mathbf{D}_{ij} = D_{ij} - [(D_{ii} + D_{jj})/2] = [(f_{ii} + f_{jj})/2] - f_{ij}, \quad (3)$$

and its average over the entire metapopulation is

$$\bar{\mathbf{D}} = \frac{\sum_{i,j=1}^n \mathbf{D}_{ij} N_i N_j}{N_T^2} \quad (4)$$

Finally, the average coancestry over the entire metapopulation is

$$\begin{aligned} \bar{f} &= \frac{\sum_{i,j=1}^n f_{ij} N_i N_j}{N_T^2} = \frac{\sum_{i=1}^n f_{ii} N_i}{N_T} - \bar{\mathbf{D}} = \\ &= \sum_{i=1}^n \frac{N_i}{N_T} \left[f_{ii} - \frac{\sum_{j=1}^n \mathbf{D}_{ij} N_j}{N_T} \right]. \end{aligned} \quad (5)$$

The right hand side of Equation (5) shows how the average diversity (heterozygosity = $1 - \bar{f}$) depends on the within-subpopulation coancestry (first term in the

brackets) and the average distance among subpopulations (second term in the brackets). Wright's (1969) F -statistics are obtained as

$$F_{IS} = \frac{\bar{f} - \bar{f}}{1 - \bar{f}}, \quad F_{ST} = \frac{\bar{f} - \bar{f}}{1 - \bar{f}} = \frac{\bar{\mathbf{D}}}{1 - \bar{f}},$$

and $F_{IT} = \frac{\bar{f} - \bar{f}}{1 - \bar{f}}$ (6)

so that $(1 - F_{IT}) = (1 - F_{IS})(1 - F_{ST})$ and

$$(1 - \bar{f}) = (1 - \bar{f}) + \bar{\mathbf{D}} = 1 - \bar{s} + \bar{\mathbf{D}} + \bar{\mathbf{D}} =$$

$$(1 - \bar{s}) + (\bar{s} - \bar{f}) + (\bar{f} - \bar{f}). \quad (7)$$

(Note that, if all alleles are distinct in the metapopulation, $\bar{f} = 1/2N_T$). Expression (7) represents the partition of the total genetic diversity (heterozygosity), $GD_T = (1 - \bar{f})$, into three components: the genetic diversity within individuals, $GD_{WI} = (1 - \bar{s})$, the genetic diversity between individuals, $GD_{BI} = (\bar{s} - \bar{f})$ (the sum of these two elements gives the genetic diversity within subpopulations, $GD_{WS} = (1 - \bar{f})$, and the genetic diversity between subpopulations $GD_{BS} = (\bar{f} - \bar{f})$). It should also be noticed that

$$\frac{GD_{WI}}{GD_T} = (1 - G)(1 - F_{ST}), \quad \frac{GD_{BI}}{GD_T} = G(1 - F_{ST}),$$

$$\frac{GD_{WS}}{GD_T} = (1 - F_{ST}) \text{ and } \frac{GD_{BS}}{GD_T} = F_{ST}. \quad (8)$$

Finally, if the different subpopulations were imposed to give different contributions (c_i) to the next generation (see below) the genetic diversity could be obtained from (5) as

$$GD_T = 1 - \bar{f} = 1 - \sum_{i,j=1}^n f_{ij}c_i c_j =$$

$$1 - \sum_{i=1}^n c_i \left[f_{ii} - \sum_{j=1}^n \mathbf{D}_{ij}c_j \right]. \quad (9)$$

Molecular markers

With molecular markers the above parameters can also be calculated. The molecular coancestry between all pairs of individuals can be simply obtained by applying Malécot's (1948) definition but referring to identity in state instead of identity by descent, and then proceeding as before. Alternatively, the information on the frequencies of the markers can be used as follows. Thus, let $p_{k,l,i}$ be the gene frequency of allele k in individual l of subpopulation i (i.e., $p_{k,l,i} = 1$ if the individual is homozygote for allele k , 0.5 if heterozygote, and 0 if homozygote for another allele), and

$$\tilde{p}_{k,i} = \frac{\sum_{l=1}^{N_i} p_{k,l,i}}{N_i} \quad (10)$$

the average frequency of allele k in subpopulation i . Then, the average coancestry between subpopulations i and j is

$$f_{ij} = \sum_{k=1}^a \tilde{p}_{k,i} \tilde{p}_{k,j}, \quad (11)$$

where a is the number of alleles. The average self-coancestry of subpopulation i is

$$s_i = \sum_{k=1}^a \sum_{l=1}^{N_i} \frac{p_{k,l,i}^2}{N_i}, \quad (12)$$

and the average distance between individuals of subpopulations i and j is

$$D_{ij} = \sum_{k=1}^a \sum_{l=1}^{N_i} \sum_{m=1}^{N_j} \frac{(p_{k,l,i} - p_{k,l,j})^2}{2N_i N_j}. \quad (13)$$

With these basic parameters and the above expressions it is possible to complete all calculations. Nevertheless, for a more complete understanding of the different parameters we can also notice that Nei's minimum distance between subpopulations i and j is $\mathbf{D}_{ij} = \sum_{k=1}^a \frac{1}{2}(\tilde{p}_{k,i} - \tilde{p}_{k,j})^2$, and the total diversity $1 - \bar{f} = 1 - \sum_{k=1}^a \tilde{p}_k^2 = \sum_{k=1}^a \tilde{p}_k(1 - \tilde{p}_k)$, the overall variance of allele frequencies in the metapopulation, where $\tilde{p}_k = \sum_{i=1}^n \tilde{p}_{k,i} N_i / N_T$, can be partitioned into:

$$1 - \bar{s} = \frac{1}{N_T} \sum_{k=1}^a \sum_{l=1}^{N_i} \sum_{i=1}^n [p_{k,l,i}(1 - p_{k,l,i})],$$

the average variance of allele frequencies between alleles within individuals;

$$\bar{s} - \bar{f} = \bar{\mathbf{D}} = \sum_{k=1}^a \sum_{i=1}^n \frac{N_i}{N_T} \left[\sum_{l=1}^{N_i} \frac{p_{k,l,i}^2}{N_i} - \tilde{p}_{k,i}^2 \right],$$

the average variance of allele frequencies between individuals within subpopulations; and

$$\bar{f} - \bar{f} = \bar{\mathbf{D}} = \sum_{k=1}^a \left[\sum_{i=1}^n \frac{N_i}{N_T} \tilde{p}_{k,i}^2 - \tilde{p}_k^2 \right],$$

the variance of allele frequencies between subpopulations. All the above measures are calculated for each marker locus, and averaged over loci.

In all previous calculations we have assumed that we have information on all the individuals that constitute the population. If the studied individuals of a subpopulation are a sample of the subpopulation

and/or if the subpopulations are a sample of a larger set of subpopulations, some corrections should be included due to finite sample size (Nei and Chesser 1983; Pons and Chaouche 1995).

The results from molecular coancestry will give exactly the same results as the coancestry from pedigrees (previous section) if there is an infinite number of loci in the base population with all the alleles in each locus being different. In other cases it could be possible in principle to infer (relate) genealogical coancestry (\bar{f}) from molecular coancestry (\bar{f}_m) using the relationship $1 - \bar{f}_m = (1 - \sum_{k=1}^a p_k^2)(1 - \bar{f})$, where p_k is the allele k frequency in the base population (Caballero and Toro 2000). However, the use of this transformation is not without problems, as it has been shown by Lynch and Ritland (1999) and Toro et al. (2001).

Example of application

To illustrate the application of the above expressions, let us consider an imaginary example based on allele frequencies of neutral genetic markers that could resemble allozyme data. The metapopulation consists of four subpopulations with 12, 4, 8 and 6 individuals, respectively. We have information on three marker loci with 5, 3 and 3 alleles, respectively. The corresponding genotypes for each locus are given in Table 1.

Using Equations (10–13) and averaging over loci we obtain the expected coancestries and Nei's minimum distances among subpopulations. These are presented in Table 2a. The other parameters describing the metapopulation (Equations 1–8) are given in Table 2b. The last column of the table indicates the proportional contribution of each subpopulation to the global coancestry. This is given by the last equality of Equation (5) and includes two components: the average coancestry of the subpopulation (f_{ii} , the complement of the within-subpopulation diversity) minus its average distance with all the others (a measure of the between-subpopulation diversity). Thus, we can see, for example, that subpopulation 1 contributes most to the overall coancestry because it has the greatest contribution to within-subpopulation coancestry, and its distance from all the other subpopulations is not large.

A possible question to be solved in conservation decisions is to ascertain the loss/gain of diversity if one or several breeds are removed from the pool. This can

be answered by disregarding one (or more breeds) and recalculating the global average coancestry from the remaining pool (Petit et al. 1998). This type of calculation is shown in Table 3. Column 2 shows the global diversity when each of the subpopulations is removed. The proportional loss/gain (negative/positive sign, respectively) in genetic diversity is given in column 3. This is the result of two components, as given in Equation (5). For example, the removal of subpopulation 1 will reduce the overall diversity of the pool due to its internal diversity (−11.3%) and will increase it due to its mean distance (+4.7%), with a global negative balance of −6.6%. On the contrary, removal of subpopulation 2 will increase diversity because of its coancestry (+9.6%) but will decrease it because of its mean distance (−7.8%), with a global positive balance (+1.8%).

All the above calculations were made weighting all parameters by each subpopulation size, because these are assumed to be known. If this is not the case, we can assume a constant subpopulation size, $N = N_T/n$, for all i in Equations (1, 4, 5). For the calculations with marker frequencies (Equations (10, 12, 13)), N_i will simply be the number of sampled individuals. Under this assumption, the loss of genetic diversity by removing each subpopulation is given in column 4 of Table 3.

Column 5 gives the percentage loss of genetic diversity obtained by the Weitzman (1992) method using Reynolds' genetic distance, which is $D_{ij}/(1 - f_{ij})$ for subpopulations i and j (Reynolds et al. 1983). Results obtained using Nei's minimum distance (Equation 3) and Nei's standard distance ($-\ln(f_{ij}/\sqrt{f_{ii}f_{jj}})$; Nei 1987) were very similar and are not shown. Because the Weitzman method only uses genetic distances between subpopulations (not within-subpopulation variation), the removal of a subpopulation always implies a loss of diversity. The method also ignores the subpopulation sizes and, therefore, the comparison should be made with column 4. It is obvious from this example that the order of the subpopulations given by the Weitzman method is in complete disagreement with that of column 4. For example, using the Weitzman method, the highest loss in diversity (−65.5%) occurs when subpopulation 2 is removed. However, column 4 indicates that removal of subpopulation 2 produces, in fact, an increase of +6.5% in the overall diversity. The reason for this difference occurs because the Weitzman method (and, similarly, other phylogenetic methods) assumes that the largest loss in diversity

Table 1. Neutral marker genotypes for individuals belonging to an imaginary metapopulation with four subpopulations

Subpopulation 1			Subpopulation 2			Subpopulation 3			Subpopulation 4		
Locus 1	Locus 2	Locus 3	Locus 1	Locus 2	Locus 3	Locus 1	Locus 2	Locus 3	Locus 1	Locus 2	Locus 3
13	13	12	44	33	22	15	33	13	25	22	12
34	23	22	44	13	22	35	33	33	25	22	12
25	23	12	44	13	22	25	33	23	44	13	22
25	23	13	44	33	22	35	13	13	15	33	22
24	22	13				55	33	23	11	13	11
25	13	22				45	13	13	14	11	12
22	22	11				45	13	33			
13	13	22				15	23	23			
13	23	12									
22	23	22									
13	13	23									
15	22	11									

Table 2. Parameters describing the metapopulation in the example. (a) Diagonal and above: average coancestry among subpopulations i and j , f_{ij} . Below diagonal: Nei's minimum distance between subpopulations, \mathbf{D}_{ij} . (b) Other parameters defined in the text (Equations 1–8)

	(a)				(b)				
	Subpop. 1	Subpop. 2	Subpop. 3	Subpop. 4	F_i	s_i	α_i	G_i	Proportional contribution to \bar{f}
Subpop. 1	0.3391	0.3021	0.2535	0.3229	0.3056	0.6528	-0.0508	0.4746	0.1356 - 0.0352 = 0.1004
Subpop. 2	0.3050	0.8750	0.3073	0.3889	0.8383	0.9167	-0.3333	0.3333	0.1167 - 0.0356 = 0.0811
Subpop. 3	0.1544	0.3685	0.4766	0.2483	0.2917	0.6458	-0.3532	0.3234	0.1271 - 0.0389 = 0.0882
Subpop. 4	0.0318	0.2338	0.1752	0.3704	0.5000	0.7500	0.2059	0.6029	0.0741 - 0.0182 = 0.0559

$$\bar{f} = 0.4535, \bar{F} = 0.4111, \bar{s} = 0.7056, \bar{\mathbf{D}} = 0.1279, \bar{f} = 0.3256$$

$$F_{IS} = -0.0775, F_{ST} = 0.1897, F_{IT} = 0.1268$$

$$GD_T = 0.6744, GD_{WI} = 0.2944, GD_{BI} = 0.2521, GD_{BS} = 0.1279$$

occurs when the subpopulation with the largest overall distance with all the others is removed. This is the case of subpopulation 2 (note the second term, -12.2%, of column 4). However, in this example, subpopulation 2 also has a large average coancestry (see f_{22} in Table 2), that produces a large increase in global coancestry when removed (note the first term, +18.7%, of column 4 in Table 3).

Finally, let us consider a different question that may be useful for the conservation of genetic diversity either in a live conservation or in a cryoconservation programme. If we had to pool the different subpopulations to produce a single one (a synthetic population or a germplasm bank), what would be the contribution of each subpopulation to the pool in order to maintain the maximum possible genetic diversity? This question can be answered by obtaining the values of c_i in Equation (9) that maximise genetic diversity, with the restrictions $c_i \geq 0$ and $\sum_{i=1}^n c_i = 1$. The appropriate values of c_i can be obtained by integer

quadratic programming, and are given in the last column of Table 3. As expected, subpopulation 1 should contribute most, while subpopulation 2 should not contribute at all to the pool. With these contributions the average genetic diversity of the pool would be 0.6874, i.e. an increment of 1.92% over the previous one.

An application to real microsatellite data

To see the implications of the above theory in a real case, we will consider the analysis of genetic diversity carried out by Laval et al. (2000) for eleven European pig breeds. These authors obtained data from 18 microsatellite markers over 18 chromosomes and evaluated the marginal loss of diversity attached to each breed by the Weitzman method.

From the expected heterozygosities, and the Reynolds' and Nei's standard genetic distances among breeds (Tables III and V, respectively, of Laval et

Table 3. Total genetic diversity and loss(-)/gain(+) of diversity (in %) when each subpopulation is removed

Subpop. (<i>i</i>)	$GD_{T i}$ ¹	% Loss/gain ² GD	% Loss/gain ³ GD	Weitzman ⁴ (D_R)	Contributions for max GD ⁵
1	0.6296	-11.3 + 4.7 = -6.6	-9.1 + 2.3 = -6.8	-5.8	0.522
2	0.6869	+9.6 - 7.8 = +1.8	+18.7 - 12.2 = +6.5	-65.5	0.000
3	0.6481	-1.2 - 5.1 = -3.9	-1.9 - 5.1 = -7.0	-63.3	0.270
4	0.6690	-3.1 + 2.3 = -0.8	-7.7 + 4.1 = -3.6	-9.0	0.208

¹Total genetic diversity after removing subpopulation *i*

^{1,2}Averages of coancestries weighted by N_i

³Averages of coancestries unweighted

⁴Marginal losses in diversity (%) obtained by the Weitzman method (using Reynolds' distance)

⁵Contribution of each subpopulation to a pool with maximal genetic diversity

Table 4. Average coancestries (f_{ij}) among European pig breeds from Laval et al. (2000)

	BEPI	DKSO	FRBA	FRGA	FRLI	FRNO	DELR	DESH	NLLW	SELR	SEWP
BEPI	0.41	0.2781	0.3283	0.3582	0.3274	0.3170	0.2141	0.2245	0.3217	0.3566	0.2812
DKSO		0.45	0.2748	0.2691	0.3429	0.3257	0.2070	0.2128	0.2654	0.3258	0.2544
FRBA			0.65	0.4138	0.3088	0.3622	0.1618	0.2124	0.3082	0.3508	0.3277
FRGA				0.50	0.3378	0.3189	0.2052	0.2381	0.3243	0.3337	0.3002
FRLI					0.56	0.3308	0.2083	0.2368	0.3591	0.3453	0.2992
FRNO						0.50	0.2027	0.2303	0.2929	0.3525	0.3074
DELR							0.38	0.2770	0.2071	0.2233	0.1578
DESH								0.34	0.2115	0.2341	0.2126
NLLW									0.50	0.3420	0.2225
SELR										0.43	0.2853
SEWP											0.41

BEPI: Piétrain (Belgium). DKSO: Sortbroget (Denmark). FRBA: Basque (France). FRGA: Gascon (France). FRLI: Limousin (France). FRNO: Normand (France). DELR: German Landrace (Germany). DESH: Schwäbisch-Hällisches (Germany). NLLW: Great Yorkshire (The Netherlands). SELR: Swedish Landrace (Sweden). SEWP: European Wild Pig (Sweden).

al. 2000) we could obtain the pairwise coancestries among breeds. These are presented in Table 4.

Column 2 of Table 5 shows the marginal losses of diversity calculated by Laval et al. (2000, Table VI) with the Weitzman method, when each of the eleven breeds (column 1) is removed from the set. The phylogeny relating all breeds is shown by Laval et al. (2000, Figure 3). Column 3 of Table 5 gives the loss/gain of global genetic diversity when each of the breeds is removed, calculated as in the previous example. Again the first term of the sum refers to the loss/gain due to the average coancestry of the subpopulation, while the second term refers to the loss/gain due to its average distance with all the others.

According to Laval et al. (2000) (see column 2 of Table 5), the highest and lowest losses of diversity are incurred with the extinction of the French Basque (FRBA) and the Piétrain (BEPI) breeds, respectively. They also showed that the four French local breeds (FRBA, FRGA, FRLI and FRNO) altogether

account for half of the total diversity, supporting the potential value of preserving local endangered breeds in the maintenance of species diversity. However, our analysis of genetic diversity using the global coancestry when each breed is removed (column 3 of Table 5) gives quite different results. Removal of the FRBA breed will produce the largest increase in diversity over the remaining pool, while removal of the BEPI breed will produce a slight increase in diversity. In addition, removal of the four French breeds would produce a substantial increase in diversity ($7.02 - 3.81 = +3.21\%$) instead of a large decrease. Therefore, the conclusions that one can draw from the two analyses are very different and, in fact, can be opposite.

Column 4 gives the expected contributions from each breed, i.e. it gives the values of c_i that maximise genetic diversity in Equation 9 assuming we would like to generate a synthetic population with the maximal genetic diversity. As expected, the largest contribution (about 74%) to this synthetic population

Table 5. Reanalysis of genetic diversity with the data of Laval et al. (2000)

Breed	Weitzman ¹	Loss/gain ¹ GD_T	c_i for max GD_T	Effective size ²	Loss/gain ^{1,3} w. GD_T
BEPI	-3.8	-0.80 + 1.01 = +0.21	0.0005	32,686	-3.6
DKSO	-10.6	-0.23 - 0.22 = -0.45	0.1128	44	-0.0
FRBA	-15.2	+2.62 - 1.95 = +0.67	0.0228	13	-0.0
FRGA	-7.9	+0.48 + 0.12 = +0.60	0	28	-0.0
FRLI	-10.8	+1.34 - 0.66 = +0.68	0	13	+0.0
FRNO	-9.5	+0.48 - 0.05 = +0.43	0	33	-0.0
DELR	-11.6	-1.23 - 1.30 = -2.53	0.2832	1,837	-2.3
DESH	-5.2	-1.80 - 1.14 = -2.94	0.2019	128	-0.1
NLLW	-12.1	+0.48 - 0.58 = -0.10	0.1214	7,368	-0.1
SELR	-4.4	-0.52 + 1.16 = +0.64	0	-	-
SEWP	-9.4	-0.80 - 0.02 = -0.82	0.2573	-	-

¹Loss(-)/gain(+) of genetic diversity (in %) when the corresponding breed is removed.

²Effective sizes of breeds from Laval et al. (2000).

³Average coancestries weighted by the effective size of breeds.

would come from those breeds contributing more to the general diversity (DELR, DESH and SEWP; cf. columns 3 and 4).

Another potential problem of the method of Weitzman is that it does not account for differences in breed size. Table III of Laval et al. (2000) gives the estimated effective sizes of the different breeds, and these are repeated in column 5 of Table 5. The differences among breeds are enormous. While some breeds have an estimated effective size of thousands of individuals, others have values of the order of tens. Clearly, a weighting by the population size (or effective size in this case) might be desirable. It is intuitive that the conclusion reached by using the Weitzman method that removal of the BEPI breed gives the lowest loss in diversity to the whole set, is a dubious one when one considers its huge effective size compared to that of the other breeds. Column 6 of Table 5 presents the loss of diversity when the breed effective sizes are weighting the average global coancestries. The two last breeds are excluded from the analysis because there are no estimates of effective size. The results indicate that the breeds with the largest effective size (BEPI, DELR and NLLW) contribute most to the general diversity and, therefore, their removal produces the largest loss in diversity. The other breeds are proportionately so small in numbers (assuming effective size and census size are proportionately related) that their removal has a negligible effect on total diversity.

Minimising coancestry maximises effective population size

As reviewed by Caballero and Toro (2000), in a conservation programme for an unsubdivided population there are two decisions to be made: how to choose the breeding individuals and how to mate them. There is an optimal solution for the choice of the breeding individuals. The method should be based on minimising the average coancestry among the reproductive individuals weighted by their contributions to the next generation (Ballou and Lacy 1995). Analogously to the case of a single population, the minimisation of coancestry in a subdivided population is equivalent to the maximisation of the effective population size. To show this, consider the full pedigree of a metapopulation, maintained under a regular breeding system with discrete generations and any constant degree of migration among subpopulations. Generation 0 is the generation of unrelated founders, and each generation the population consists of N individuals in each of n subpopulations. The subpopulation sizes are constant every generation but the contribution of each subpopulation to the next generation may be variable. Following Caballero and Toro (2000; Equation 5), the average coancestry of the metapopulation at generation t is

$$\bar{f}_t = \frac{1}{2N^2n^2} \left[\sum_{s=1}^n \sum_{i=1}^N c_{is(0,t)}^2 \right] + \frac{1}{4N^2n^2} \left[\sum_{k=1}^t \sum_{s=1}^n \sum_{i=1}^N c_{is(k,t)}^2 (1 - F_{k-1}) \right],$$

where $c_{is}^2(k,t)$ is the square contribution of individual i of subpopulation s from generation k to descendants of generation t , and F_k is the average inbreeding coefficient of generation k . Noting that the variance of contributions from individuals within subpopulation s is $V_{WS(k,t)} = \sum_{i=1}^N \frac{c_{is}^2(k,t)}{N} - c_s^2(k,t)$, where $c_s(k,t)$ is the mean contribution from individuals of subpopulation s , and that the variance of $c_s(k,t)$ between subpopulations is $V_{BS(k,t)} = \sum_{i=1}^N \frac{c_s^2(k,t)}{n} - 1$, we obtain

$$\bar{f}_t = \frac{1}{Nn} \left\{ \frac{1}{2} [\bar{V}_{WS(0,t)} + V_{BS(0,t)} + 1] + \frac{1}{4} \left[\sum_{k=1}^t (\bar{V}_{WS(0,t)} + V_{BS(0,t)} + 1)(1 - F_{k-1}) \right] \right\},$$

where the bar denotes average over subpopulations. Assuming large Nn and using a linear approximation, the rate of increase in coancestry between generations 0 and t is $\Delta f_{0,t} \approx f_t \approx t/2N_e$, and the effective population size N_e is

$$N_e = \frac{2Nnt}{2(\bar{V}_{WS(0,t)} + V_{BS(0,t)}) + \sum_{k=1}^t (\bar{V}_{WS(0,t)} + V_{BS(0,t)} + 1)(1 - F_{k-1})}.$$

This is a generalisation, for the case of a subdivided population, of Equation (11) of Caballero and Toro (2000). After a number of generations ($t \rightarrow \infty$) the variance of contributions within and between subpopulations reaches an asymptotic value ($V_{WS(0,\infty)}$ and $V_{BS(0,\infty)}$) and

$$N_e \approx \frac{2Nn}{(\bar{V}_{WS(0,\infty)} + V_{BS(0,\infty)} + 1)(1 - F_{IT})}, \quad (14)$$

which generalises the equation of Wray and Thompson (1990) (see Equation 15 in Caballero and Toro 2000). As for single populations, the above equation can be expressed in the classical way as a function of the variances of offspring from parents within (S_W^2) and between (S_B^2) subpopulations. The necessary relations are

$$\bar{V}_{WS(0,\infty)} = \frac{S_W^2}{2} \frac{(1 + 3F_{IS})}{(1 - F_{IS})}$$

(see below Equation 17 of Caballero and Toro (2000) substituting α by F_{IS} and denoting $S_{(1)}^2$ by S_W^2) and, analogously,

$$V_{BS(0,\infty)} = \frac{(S_B^2/2)(1 + 3F_{IS})(1 - F_{ST}) + 2N(S_B^2/2)F_{ST}}{(1 - F_{IS})(1 - F_{ST})}.$$

The meaning of the above relations becomes clear when they are put together in a single expression,

$$\frac{\bar{V}_{WS(0,\infty)} + V_{BS(0,\infty)}}{(S_W^2 + S_B^2)(1 - F_{IS})(1 - F_{ST}) + (4S_W^2 + 4S_B^2)F_{IS}(1 - F_{ST}) + 2NS_B^2F_{ST}} =$$

The first term in the numerator indicates the variance within individuals. The second one indicates the component between full-sib couples, the factor 4 referring to the four gametes of each couple (for a monogamous species with self-fertilisation the corresponding factor is 2). Finally, the third term is the component between subpopulations, the factor $2N$ referring to the $2N$ gametes of each subpopulation. Substituting into (14) we obtain

$$N_e \approx \frac{Nn}{\frac{1}{4}(1 - F_{ST}) \left[(S_W^2 + S_B^2)(1 + 3F_{IS}) + 2(1 - F_{IS}) \right] + N(S_B^2/2)F_{ST}}, \quad (15)$$

which agrees with the expression (31) of Wang and Caballero (1999) with our notation, $S_k^2 = S_W^2$, $S_B^2 = 4V$, and assuming large n , so that $n/(n - 1) \approx 1$. For $S_W^2 = 2$ and $F_{IS} = 0$, it reduces to the original expression of Whitlock and Barton (1997). And when $F_{ST} = 0$ (undivided population of size Nn) the expression gives Equation (18) of Caballero (1994), as it should.

Discussion

The criteria to set priority to different breeds for conservation programmes are a controversial issue that has been recently reviewed by Ruane (1999). The final decisions should take into account several factors such as the adaptation to specific environments or diseases, the possession of specific traits of scientific or future economic value and the historical or cultural values of the breed.

Genetic information also plays an important role in conservation genetics. We have concentrated on developing tools for the genetic management of a conservation programme aimed to maintain the maximum overall genetic diversity, taking account of the subdivision of the population. The model we have in mind is one involving breeds in any domestic species, although the tools could also be applied to strains of plants, forest trees or even wild species. On the other hand, the measure of genetic diversity we have chosen is the expected heterozygosity, proposed by Nei (1973), that is defined as the probability that two alleles chosen at random from the population are different, and equals the proportion of heterozygotes within the population at Hardy-Weinberg equilibrium. Although this measure is usually applied to molecular markers and refers to identity in state, it

must be emphasised that it is equivalent to the classical Malécot (1948) coefficient of coancestry in a model where all the alleles in the base or reference population are assumed to be different. Thus, the analysis presented here can be applied either to molecular coancestry measured from markers or to genealogical coancestry coming from pedigree information.

Other criteria related to the rather vague concept of genetic uniqueness or distinctiveness have been put forward. Some of them are the number of distinct alleles (e.g., Petit et al. 1998), the maintenance of genetic diversity in special regions of the genome deemed important, such as the major histocompatibility complex, or to pay attention to the mitochondrial or Y chromosome lineages (Oldenbroek 1999). Perhaps, the most general genetic criterion apart from genetic diversity (expected heterozygosity) is allelic diversity (Petit et al. 1998; Barker 2001). As this latter is more sensitive to bottlenecks than average heterozygosity, it can be considered a better criterion to ascertain past fluctuations in the population size. However, because the "effective number of alleles" is, by definition, the inverse of mean coancestry (Crow and Kimura 1970, p. 324), minimisation of mean coancestry maximises the effective number of alleles. Therefore, it is expected that maximisation of genetic diversity will lead to maximum allelic richness in long-term conservation programmes, and will be appropriate as a general guide for conservation.

Closely related to the Nei's genetic diversity measure are the genetic distance measures, especially Nei's minimum and standard distances and Reynolds' distance. As pointed out by Ruane (1999) in recent years there has been an explosion of projects aiming to calculate genetic distances between domesticated breeds of animals mainly based on the use of microsatellite markers and aimed to set priorities for conservation. For example, Wimmers et al. (2000) analysed the genetic variability of several local chicken populations using 22 microsatellites and obtained genetic distances between them. They stated that the decision to preserve a certain population as a genetic resource should be drawn from the determination of the mean genetic distance of this population to all the others in the species. The value of such tools for helping in the maintenance of genetic diversity has scarcely been critically examined (but see Ponzoni 1997; Ruane 1999).

Matrix distances among populations are graphically represented by distance trees. The most widely

known methods are the Neighbour-Joining and the UPGMA (Takezaki and Nei 1996). Under a number of assumptions, namely isolation and independent evolution after divergence, the tree representation can be taken as a phylogeny. However, for domestic species the differences in effective size and the migration among breeds precludes such phylogenetic interpretation. From the genetic distance matrix, Weitzman (1992, 1993) has proposed an analytical framework to better rationalise conservation policies. The method provides a graphical representation in which the branch length of each breed can be read as approximately measuring its relative contribution to the diversity function. Thus, it is possible to calculate the contribution of each breed to the total diversity and the marginal diversity of each breed.

Thaon d'Arnoldi et al. (1998) discussed the properties of Weitzman analysis and concluded that it has some desirable properties such that the removal of an element always decreases the total variability. This is probably the case when thinking about the elimination of species, because it may be assumed that species are distinct and unchangeable entities. But, as we have shown in this paper, this property is at variance with the classical ideas about genetic diversity within a species, because in a population the rejection of some individuals could increase the genetic variability as long as these individuals are substituted by others more appropriate. The Weitzman method does not have a clear interpretation in terms of the most accepted measure of genetic diversity, Nei's (1973) expected heterozygosity, and ignores within breed variability, which constitutes a crucial part of the metapopulation variability. Besides, it introduces a method to study the marginal contribution of each breed to the total variability when there is no need for such a method, as there is a natural way of doing it in the classical Nei's setting. Finally, the Weitzman method does not account for different census or effective size of the breeds, which could be important in some settings. Therefore, we conclude that the Weitzman analysis does not add any insight to the more standard analysis of genetic diversity that follows from the seminal paper of Nei (1973). Furthermore, the application of the Weitzman method to give priorities for conservation within a species may lead to misleading conclusions, as we showed in this paper with the reanalysis of Laval et al. (2000) data.

In summary, the Weitzman method or any other phylogenetic method that ignores within-breed variability should not be used when applied to within-

species diversity. However, caution should also be used when within-breed variation is considered. It is expected that those breeds with the largest population size will be those having the highest internal diversity. Therefore, conservation policies accounting for within breed variation may tend to give priority to the largest breeds. In the case of the European breeds studies by Laval et al. (2000) there is no indication of such a problem. The breeds with the largest effective sizes (BEPI, DELR and NLLW; Table 5) do not have particularly lower within-breed coancestries (see diagonal of Table 4). Nevertheless, this issue has to be taken into account before applying any conservation decision. Chaiwong and Kinghorn (1999) gave five times more weight to the variation between breeds than to that within breeds in their approach to balance genetic diversity and genetic merit in animal conservation programs. Although this was arbitrary, heuristic solutions could be sought for each particular case depending on the levels of within-breed variation and breed size.

In the management of conservation programmes, decisions have to be directed towards the selection of breeding individuals and the system of mating. As discussed by Caballero and Toro (2000), the optimal choice of the breeding individuals requires minimisation of the average coancestry among the reproductive individuals weighted by their contributions to the next generation. In practice, this implies that after a small number of generations of regular breeding the contribution from each individual to the next generation is constant. For the case of a subdivided population, equal contributions from each subpopulation ($S_B^2 = 0$ in Equation 15) and equal contributions from each individual within subpopulations ($S_W^2 = 0$ in Equation 15) are required to minimise effective size (see Wang and Caballero 1999) and, therefore, to maximise the average coancestry of the metapopulation (Caballero and Toro 2000, this paper).

The choice of the mating system is apparently less simple because it depends on the time scale of interest. In the short term, avoiding mating between relatives in an undivided population (minimum coancestry matings) will reduce or delay inbreeding in the population but it will give higher inbreeding in the long term. However, because we can safely assume that inbreeding has negative consequences in terms of inbreeding depression and that there is a linear or concave relationship between inbreeding coefficient and inbreeding depression, it can be argued that the accumulated inbreeding (summed up to all genera-

tions) will be minimised when minimum coancestry matings are performed. Moreover, the interest for conservation will be more often in the short rather than in the long term. Therefore, avoidance of inbred matings seems more appropriate.

The subdivision of the population is the consequence of imposing a particular system of mating (e.g., Caballero 1994; Wang and Caballero 1999). Therefore, the effect of subdivision follows the same principles as those described before for the effect of non-random mating. If the population is subdivided permanently in groups (independent sublines with completely different pedigrees), the lines will become completely inbred but genetic drift will be minimised as different alleles will be fixed in the different groups. If population subdivision is not complete then inbreeding and drift will have the same final rate but the decline in inbreeding will precede the drift. Then, in absence of other factors, to practise minimum coancestry matings without any subdivision will lead to minimising the overall sum of inbreeding coefficients and, therefore, the total inbreeding depression. However, in practical situations, either in live conservation or cryoconservation programmes, other factors such as sanitary, genetic or cultural reasons could indicate maintenance of a subdivided population. In such cases additional constraints, such as the minimum levels of contributions of each breed, should be included in the mathematical or heuristic optimisation programme.

Acknowledgements

We are grateful to Jinliang Wang, Liz Boulding and two anonymous referees for helpful comments on the manuscript, to Jesús Fernández for the calculation of the contributions of minimum coancestry, and to Javier Cañón and David García for the software to calculate Weitzman distances. This work was supported by grant BOS2000-0896 (Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica, I+D+I) and Universidad de Vigo (64102C124) (A.C.) and by INIA Proyect SC98-083 (M.A.T.).

References

- Barker JSF (1999) Conservation of livestock breed diversity. *AGRI* 25, 33–43.

- Barker JSF (2001) Conservation and management of genetic diversity: A domestic animal perspective. *Can J For Res*, **31**, 588–595.
- Ballou JD, Lacy RC (1995) Identifying genetically important individuals for management of genetic diversity in captive populations. In: *Population Management for Survival and Recovery* (eds. Ballou JD, Gilpin M, Foose TJ), pp. 76–111. Columbia University Press, New-York.
- Caballero A (1994) Developments in the prediction of effective population size. *Heredity*, **73**, 657–679.
- 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.
- Chaiwong N, Kinghorn BP (1999) Use of genetic markers to aid conservation decisions for groups of rare domestic breeds. *Proc Assoc Advmt Anim Breed Genet*, **13**, 365–368.
- Cockerham CC (1967) Group inbreeding and coancestry. *Genetics*, **56**, 89–104.
- Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *TREE*, **15**, 290–295.
- Crow JF, Kimura M (1970) *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- Frankham R (1995) Conservation genetics. *Annu Rev Genet*, **29**, 305–327.
- Krajewski C (1994) Phylogenetic measures of bioiversity: A comparison and critique. *Biol Conserv*, **69**, 33–39.
- Laval G, Iannuccelli N, Legault C, Milan D, Groenen MAM, Giuffra E, Andersson L, Nissen PH, Jørgensen CB, Beeckmann P, Geldermann H, Foulley J-L, Chevalet C, Ollivier L (2000) Estimation of the coancestry coefficient: Basis for a short term genetic distance. *Genet Sel Evol*, **32**, 187–203.
- Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. *Genetics*, **152**, 1753–1766.
- Malécot G (1948) *Les Mathématiques de l'Hérédité*. Masson et Cie, Paris, France.
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci, USA*, **70**, 3321–3323.
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. *Ann Hum Genet*, **47**, 253–259.
- Oldenbroek JK (1999) *Genebanks and the Conservation of Farm Animal Genetic Resources*. DLO Institute for Animal Sciences and Health, Lelystad, The Netherlands.
- Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conserv Biol*, **12**, 844–855.
- Pons O, Chaouche K (1995) Estimation, variance and optimal sampling of gene diversity II. Diploid locus. *Theor Appl Genet*, **91**, 122–130.
- Ponzoni RW (1997) Genetic resources and conservation. In: *The Genetics of Sheep* (eds. Piper LR, Ruvinsky A), pp. 433–466. CAB International.
- Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: Basis for a short term genetic distance. *Genetics*, **105**, 767–779.
- Ruane J (1999) A critical review of the values of genetic distance studies in conservation of animal genetic resources. *J Anim Breed Genet*, **116**, 317–323.
- Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**, 389–399.
- Thaon d'Arnoldi C, Foulley J-L, Ollivier L (1998) An overview of the Weitzman approach to diversity. *Genet Sel Evol*, **30**, 149–161.
- Toro MA, Barragán C, Óvilo C, Rodríguez J, Rodríguez C, Silió L (2001) Estimation of coancestry in Iberian pigs using molecular markers. *Conserv Genet*, submitted.
- Wang J, Caballero A (1999) Developments in predicting the effective size of subdivided populations. *Heredity*, **82**, 212–226.
- Weitzman ML (1992) On diversity. *Quart J Econ*, **107**, 363–405.
- Weitzman ML (1993) What to preserve? An application of diversity theory to crane conservation. *Quart J Econ*, **108**, 157–183.
- Whitlock M, Barton NH (1997) The effective size of a subdivided population. *Genetics*, **146**, 427–441.
- Wimmers K, Ponsuksili S, Hardge T, Valle-Zarate A, Mathur PK, Horst P (2000) Genetic distinctness of African, Asian and South American local chickens. *Anim Genet*, **31**, 159–165.
- Wray NR, Thompson R (1990) Predictions of rates of inbreeding in selected populations. *Genet Res*, **55**, 41–54.
- Wright S (1969) *Evolution and the Genetics of Populations*. Vol. 2, *The Theory of Gene Frequencies*. The University of Chicago Press, Chicago, USA.

