



ORIGINAL ARTICLE

## Estimation of effective population size from the rate of coancestry in pedigreed populations

I. Cervantes<sup>1</sup>, F. Goyache<sup>2</sup>, A. Molina<sup>3</sup>, M. Valera<sup>4</sup> & J.P. Gutiérrez<sup>1</sup>

<sup>1</sup> Departamento de Producción Animal, Facultad de Veterinaria, Avda. Puerta de Hierro, s/n, E-28040-Madrid, Spain

<sup>2</sup> SERIDA-Deva, Camino de Rioseco, 1225, E-33394-Gijón, Spain

<sup>3</sup> Departamento de Genética. Universidad de Córdoba. Ctra. Madrid-Cádiz, km 396a, 14071-Córdoba, Spain

<sup>4</sup> Departamento de Ciencias Agro-Forestales, EUITA, Universidad de Sevilla, Ctra. Utrera km 1, E-41013 Sevilla, Spain

### Keywords

Effective population size; horses; increase in coancestry; increase in inbreeding.

### Correspondence

I. Cervantes, Departamento de Producción Animal, Facultad de Veterinaria, Avda. Puerta de Hierro s/n, E-28040-Madrid, Spain.  
Tel/Fax: +34 913 943773;  
E-mail: icervantes@vet.ucm.es

Received: 21 October 2009;  
accepted: 1 May 2010

### Summary

We introduce a simple method to estimate effective population size from increase in coancestry ( $\Delta c_{jk}$ ) for all pairs of individuals  $j$  and  $k$  in a reference subpopulation. An increase in pairwise coancestry for any pair of individuals  $j$  and  $k$  can be defined assuming that a hypothetical mating between them would give an individual with an inbreeding coefficient equal to  $c_{jk}$ , where  $c_{jk}$  is the coancestry coefficient between the individuals  $j$  and  $k$ . The equivalent measure to discrete generations value ( $g_{jk}$ ) corresponding to the individual  $jk$  can be computed by averaging discrete equivalent generations of its parents ( $g_j$  and  $g_k$ ). The mean increase in coancestry for all pairs of individuals in a reference subpopulation can be used to estimate a realized effective population size based on coancestries that would provide information on the effective size of a population under random mating. Performance of the new parameter was tested on simulated and empirical (horse) populations with different mating strategies and population structures. The routines needed to compute the introduced parameters have been included in a new version of the program ENDOG.

### Introduction

The effective population size ( $N_e$ ), usually computed from increase in inbreeding ( $\Delta F$ ), is a major criterion for classifying animal populations for degrees of endangerment (FAO 1998; Duchev *et al.* 2006). However, the interpretation of these parameters is not straightforward because their computation depends on completeness and depth of the pedigree of the studied population. And indirectly, for the effects of drift, mating policy, overlapping of generations, selection, migration or population subdivision. Overall, these constraints and others move real animal populations far from the theoretically idealized population referred to by genealogical tools (Falconer & Mackay 1996).

To overcome these problems, Gutiérrez *et al.* (2008, 2009) have recently proposed the estimation of a realized effective size ( $\bar{N}_e$ ; Cervantes *et al.* 2008) from the definition of the individual increase in inbreeding coefficients (González-Recio *et al.* 2007).

However, individual increase in inbreeding coefficients are still affected by population structures (Gutiérrez *et al.* 2008). The computed effective population size using  $\Delta F_i$  accounts for all historical pedigree of the individuals and the obtained  $N_e$  summarizes all the genealogical information of each individual. When structure exists (i.e. if local bottlenecks occur),  $\Delta F_i$  adjusts the recorded genealogies to account for this effect as soon as structure appears. This is a major issue in genealogical studies of real populations.

It is well known in population genetics that under a regular system of inbreeding, both the increase in inbreeding and the increase in coancestry reach an identical asymptotic value with the only exception of scenarios of permanent subdivision (Falconer & Mackay 1996). However, these theoretical assumptions do not work correctly in real populations because structuring, owing to geographical distance, selection or local founder effects, is the norm in real pedigrees with sufficient size and depth, particularly when the populations are under selective processes. This population structure is permanent but incomplete with intensity varying over time because of the success of new selected families, improvements in reproduction and transport facilities or changes in the structure of herds. This structuring is very often cryptic and in subpopulations (i.e. individuals sharing a major proportion of genes of a selected family) cannot be straightforwardly identified.

When dealing with structured populations,  $N_e$  can be estimated from increase in coancestry instead of increase in inbreeding (Caballero 1994). In contrast with increase in inbreeding, increase in coancestry is negligibly affected by population structure (Gutiérrez *et al.* 2008), which basically would reflect the effect of drift because of the finite size of the population (Caballero 1994). On the other hand, estimation of  $N_e$  from increase in inbreeding could be preferred when the focus is on inbreeding depression which depends on the inbreeding coefficient and not on the coancestry coefficient.

Here, we extend the methodology previously shown for the individual increase in inbreeding (Gutiérrez *et al.* 2008, 2009) to define an increase in pairwise coancestry ( $\Delta c_{jk}$ ) between individuals  $j$  and  $k$  with the aim of making it possible to estimate  $N_e$  from increase in coancestry in real populations with generations overlapping.

## Material and methods

### Theoretical background

The individual increase in inbreeding described by Gutiérrez *et al.* (2008) with the posterior modification to account for avoidance of self-fertilization (Gutiérrez *et al.* 2009) is defined as  $\Delta F_i = 1 - \frac{g_i}{\sqrt{1 - F_i}}$ , with  $g_i$  the discrete equivalent generation computed for the pedigree of each individual as the sum over all known ancestors of the term  $(1/2)^n$ , where  $n$  is the number of generations separating the individual from each known ancestor (Maignel *et al.* 1996; Boichard *et al.* 1997) and  $F_i$  the inbreeding

coefficient of an individual  $i$ . This method takes advantage of the parameter equivalent to discrete generations ( $g$ , Maignel *et al.* 1996; Boichard *et al.* 1997) to solve the problem that, in fact, overlapping generations are the normal scenario in real populations. The parameter equivalent to discrete generations was previously used to estimate effective population size by regression of inbreeding coefficients over generations instead of regressing them over year of birth (Gutiérrez *et al.* 2003). This overcomes the problem of differences in pedigree depth in individuals born on a given date. However, regression-based methods to compute  $N_e$  have been shown to be seriously affected by changes in the breeding policy, sometimes leading to nonsense negative estimates of effective population size (Cervantes *et al.* 2008; Gutiérrez *et al.* 2008).

The  $\Delta F_i$  values of the individuals belonging to a reference population can be averaged to give  $\overline{\Delta F}$  and, from this, a mean effective population size  $\overline{N_e}$ , called realized effective size by Cervantes *et al.* (2008), can be straightforwardly computed as  $\overline{N_e} = \frac{1}{2\overline{\Delta F}}$ , and a standard error of  $\overline{N_e}$  can be further computed from the standard deviation of the individual increase in inbreeding and the square root of the reference subpopulation size as  $\sigma_{\overline{N_e}} = 2\overline{N_e}^2 \sigma_{\Delta F} \frac{1}{\sqrt{\overline{N_e}}}$  (Cervantes *et al.* 2008) assuming  $\overline{N_e}$  independent individuals. The method by Gutiérrez *et al.* (2008, 2009) has been shown to indirectly account for most effects leading to increases in inbreeding higher than those expected for the idealized population because they affect the individual inbreeding coefficients (Cervantes *et al.* 2008).

In a similar way, an increase in pairwise coancestry for any pair of individuals  $j$  and  $k$ , belonging to different sexes, can be defined assuming that a hypothetical mating between them would give an individual with an inbreeding coefficient equal to  $c_{jk}$ , where  $c_{jk}$  is the coancestry coefficient between the individuals  $j$  and  $k$ . All coancestries including those between animals with the same sex can also be taken into account for a different use. The pedigree depth ( $g_{jk}$ ) corresponding to the  $jk$  offspring can be computed averaging the equivalent complete generations of its parents ( $g_j$  and  $g_k$ ). Therefore, increase in coancestry between any pair of individuals  $j$  and  $k$  can be computed as,  $\Delta c_{jk} = 1 - \frac{g_j + g_k}{2} \sqrt{1 - c_{jk}}$ , where  $c_{jk}$  is the inbreeding of a descendent from both, and  $g_j$  and  $g_k$  are the discrete equivalent generation for the parents. This also takes into account the absence of self-fertilization (Gutiérrez *et al.* 2009). By averaging the increase in coancestry for all pairs of individuals

in a reference subpopulation, we can estimate a realized effective population size based on coancestries as  $\bar{N}_{ec} = \frac{1}{2\Delta c}$  that would provide information on the effective size of a population under random mating.

As described by Cervantes *et al.* (2008), given that an increase in coancestry exists for each pair of animals in the reference subpopulation, a standard error of the realized  $\bar{N}_{ec}$  can be further computed from the standard deviation of these increases in coancestry ( $\sigma_{\Delta c}$ ), and the square root of the effective size of the effective number of pairwise coancestries in the reference subpopulation as  $\sigma_{\bar{N}_{ec}} = 2\bar{N}_{ec}^2 \sigma_{\Delta c} \frac{1}{\sqrt{\frac{N_{ec}(N_{ec}-1)}{2}}}$ . In this expression, it is

assumed that not all the coancestries are independent but only those computed in a population formed by  $\bar{N}_{ec}$  individuals. The reliability of this standard error, and also that for  $\bar{N}_e$ , is still not absolute given that it is assumed that increase in pairwise coancestries are assumed uncorrelated, but it can be useful as far as it provides an idea of the variability and a comparative reference.

We demonstrate this methodology on two *ad hoc* simulated populations and on several linked horse studbooks which can be studied separately and together as a single subdivided population.

### Simulated data

Three scenarios were created, in all of them the founder population was formed by the same number of individuals of both sexes. A total of 100 males and 100 females were born for each generation and acted as parents under random mating, with no differential viability or fertility. No restriction on the number and sex of offspring was established. The following assumptions were considered:

**Simulation (i)** The population consisted of a total of 100 males and 100 females born in each generation and acted as parents for the following discrete generations. After 10 generations, the population was split in 4 different subpopulations consisting of 25 males and 25 females evolving separately for 40 additional generations.

**Simulation (ii)** In this case, the initial population is subdivided from the foundation in 4 different subpopulations consisting of 25 males and 25 females each born in each generation and acted as parents for the following discrete generations, evolving separately until generation 10. After that, the four subpopulations were merged into a single population formed by 100 males and 100 females evolving

together under random mating over 40 additional generations.

**Simulation (iii)** The population consisted of a total of 100 males and 100 females born at each time period limiting the replacement of reproducing individuals to 50 males and 50 females at each time period and allowing the reproducing individuals to have offspring at two consecutive periods (overlapping generations). After 20 periods, the population was split into two subpopulations consisting of 50 males and 50 females evolving separately for 80 additional periods.

### Real data

Six breeds previously analyzed by Cervantes *et al.* (2008, 2009) were employed to demonstrate the methodology using data from real populations, including parental and derived populations. The data included information from the studbooks of three pure horse breeds [Spanish Arab (SA), Spanish Purebred (SPB) and Spanish Thoroughbred (TB)] and three composite breeds [Spanish Anglo-Arab (dAA), Hispano-Arab (dHA) and Spanish Sport horse breeds (dSSH)] and collected between the foundation of the studbook and 2004:

(i) The pedigree of the Spanish Purebred horse (SPB, Andalusian horse), totalling 140 629 individuals (Cervantes *et al.* 2009). We also included the Carthusian strain which is a subset of the pedigree of SPB with their own management structure, including the founder individuals classified as Carthusian because of their particular historical origin and the individuals resulting from matings between Carthusian parents (Valera *et al.* 2005).

(ii) The Spanish Thoroughbred (TB) pedigree, with a total of 33 463 individuals (Cervantes *et al.* 2009).

(iii) The Spanish Arab horse (SA) pedigree included 18 880 individuals (Cervantes *et al.* 2009).

(iv) The Anglo-Arab (dAA) breed that was formed by continuous crossing between Arab horses and Thoroughbred (TB) and totalled 8289 dAA animals (Cervantes *et al.* 2009).

(v) The Hispano-Arab (dHA) breed that was formed by continuous crossing between Spanish Purebred (SPB; Andalusian) horses. The total number of dHA animals included in the dataset was 3394 (Cervantes *et al.* 2009).

(vi) The Spanish Sport horse (dSSH), is a composite population that includes any individual participating in sports competitions that is not registered in any other studbook and, therefore, an individual belonging to SA, SPB, TB, dAA or dHA or any other breed

that can act as parent in the dSSH studbook (Cervantes *et al.* 2009). The total number of dSSH animals included in the dataset was 7099 (Cervantes *et al.* 2009).

Because of the significant genealogical connectedness among pedigrees analyzed by founder contribution and by coancestry matrix (for more details see Cervantes *et al.* 2009), the six horse studbooks were further merged and also analyzed as a single set of populations.

The horse pedigrees used here are expected to be useful to assess the ability of the introduced methodologies to identify cryptic structures within populations. From the history of the breeds, it is known that (i) within the SPB breed, the Carthusian population has a separate management, and within Carthusians, some studs such as that of 'Hierro del Bocado' are particularly important in size and as a source of genes (Valera *et al.* 2005); (ii) the SA breed has a deep pedigree which is still open for Arab horses from foreign countries and, within this breed, the Military Stud of Jerez de la Frontera has a particular breeding significance given that many farmers only select individuals from this origin to mate their animals (Maxwell 1995); (iii) the TB population has the shallowest pedigree of the parental horse populations studied (Cervantes *et al.* 2009); (iv) the level of integration of the parental breeds into the derived breed (dAA, dHA and dSSH) is different and, as a whole, the set of horse populations analyzed here has unbalanced contributions from the different individual horse breeds.

### Reference subpopulations

Changes in  $N_e$ , which are a consequence of variations in the population structure or mating policies, can be addressed by defining several reference subpopulations across time periods and plotting the obtained values by time. However, if only the animals born in the time period equivalent to the last equivalent generation interval are considered, the computed  $N_e$  is likely to correspond to the present reproductive subpopulation. Thus, in simulations i) and ii) we used each generation as reference subpopulations (RS) to analyze the effective size trends, while in the simulation iii) individuals of the last two generations were used to obtain values of a reproductive subpopulation for subdivision and overlapped generations combined scenario. In scenario i), the theoretical  $N_e$  per discrete generation in terms of coancestry, excluding self-fertilization, is the number of individuals +  $\frac{1}{2}$  (200.5), and in terms of inbreed-

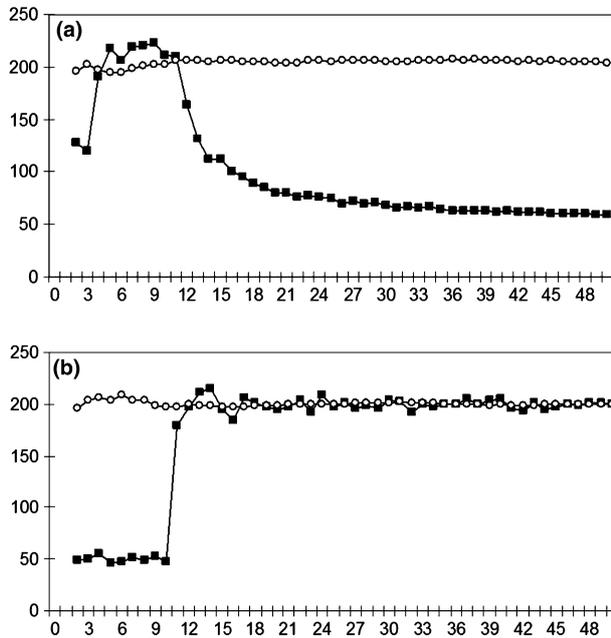
ing the same value before subdivision and 50 after that. Regarding scenario ii), the theoretical  $N_e$  per discrete generation at the time of subdivision is 200.5 in terms of coancestry and 50 in terms of inbreeding; after subdivision, theoretical  $N_e$  both in terms of inbreeding and in terms of coancestry have the same value (200.5). Finally, for scenario iii) the theoretical  $N_e$  in the last two generations in terms of coancestry can be derived from the expression  $N_e = \frac{8N_C}{V_{km} + V_{kf} + 4} L$  (Falconer & Mackay 1996), where  $N_C$  is the number of reproductive individuals included in the reference subpopulation (100),  $V_{km}$  and  $V_{kf}$  are, respectively, the variances of family sizes of reproducing males and females ( $V_{km} = V_{kf} = 2$  under random conditions), and  $L$  is the generation length in units of the specified time interval (1.5),  $N_{ec}$  equals to 150. And in terms of inbreeding can be derived from the expression for unequal numbers of animals breeding in successive generations  $\frac{1}{N_e} \approx \frac{1}{t} \sum_{i=1}^t \frac{1}{N_i}$  (Falconer & Mackay 1996), where  $t$  is the number of considered generations and  $N_i$  either 100 or 50 depending on whether before or after subdivision. In this case  $N_e$  equals to 83.3.

Regarding the real data, reference subpopulations were defined in real populations as the individuals born in the last 10 years from the available records for all the pedigrees. Pairwise coancestries between individuals were computed between all males and females in the RS except for SPB and the set of horse populations from which we selected a random sample of individuals because of computation time limitations.

All the analyses were performed using the ENDOG program (current version v4.8) (Gutiérrez & Goyache 2005).

### Results

Figure 1 shows the trend of realized effective population size estimated from  $\Delta F_i$  and from  $\Delta c_{jk}$  in the simulated datasets (i and ii). In Figure 1a, both the effective sizes, as expected, are very similar until generation 10 where the simulated population was split into 4 subpopulations. After that the ratio  $\frac{N_{ec}}{N_e}$  reached values over 3, and its trend showed that the real number of subpopulations can be reached asymptotically. In Figure 1b, the opposite scenario is shown: the ratio  $\frac{N_{ec}}{N_e}$  value was nearly four until generation 10 where population structure disappeared and both estimations lead to the same values. Common founders who had old relationships seem to have some weight on later generations (1a). Plot 1b



**Figure 1** Trends of realized effective population size ( $\bar{N}_e$ , squares) and effective size based on an increase in pairwise coancestries ( $\bar{N}_{cc}$ , circles) in the simulated example (i) (a; population split in 4 subpopulation after 10 generations) and in the simulated example (ii) (b; population subdivided in four subpopulations until generation 10). The x-axis shows generation number and the y-axis the effective sizes values.

shows that admixing unrelated populations immediately gives a ratio similar to that corresponding to a single population. Regarding simulation iii) and using the last two periods as reference population, the estimated  $\bar{N}_e$  was  $84.6 \pm 1.43$  and  $\bar{N}_{cc}$  was  $149.5 \pm 1.14$ , roughly the same values as the theoretical effective sizes (83.3 and 150, respectively). In this case, the ratio between the parameters was 1.8.

Table 1 gives the realized effective population sizes based on inbreeding and on coancestries for the RS defined in the real datasets analyzed. The values estimated for  $\bar{N}_e$  varied from 19.8 for the Carthusian horse strain to 332.1 for Spanish Thoroughbred, whereas the estimates for  $\bar{N}_{cc}$  were always higher, varying from 24.2 for the Carthusians to 819.1 for SSH. As expected, parental horse breeds reached lower estimates than composite breeds except for TB that had the shallower pedigree among the parental horse populations thus probably giving biased estimates of effective size. The estimates of  $\bar{N}_e$  had higher standard errors than  $\bar{N}_{cc}$  particularly for the composite horse populations, but remember that these are only standard errors.

The ratio  $\frac{\bar{N}_{cc}}{\bar{N}_e}$  in SA and SPB was similar and roughly 1.5, with the Carthusian horse value of 1.2

which is lower than that for SPB. In the Spanish Thoroughbred, the ratio was close to 1. Regarding the composite breeds, the lower ratio for dHA was 1.1. This is considered as an unstructured population. However, a noticeable ratio of 4.6 was assessed for the dSSH population. Finally, the set of horse populations, including all the available breeds, had a ratio of 1.9.

## Discussion

It is well known that many factors lead real populations to increases in inbreeding higher than those expected in the idealized population as a consequence of preferential matings (Falconer & Mackay 1996). However, increases in coancestry basically reflect the drift caused by the finite size of the population (Chesser *et al.* 1993; Caballero 1994; Wang & Caballero 1999). Thus, discrepancies between increases in inbreeding and in coancestry, except for minimum coancestry strategies, can be interpreted as cryptic population subdivision.

Population structuring is one of the major factors leading to increases in inbreeding higher than that expected under random mating regarding population size (Fernández *et al.* 2008). On the other hand, minimum coancestry mating systems lead to lower increases in inbreeding than in coancestry that would be expected under random mating. In most real populations, mating policies, breeding goals or, simply, geographical distance prevent optimal matings thus leading to population substructures. In such scenarios, non-random matings occur because reproduction is mainly carried out within subpopulations (Wang & Caballero 1999). Thus, substructuring greatly affects increase in inbreeding coefficients, while increase in coancestry remains approximately stable. When a permanent subdivision in a population is known, it is possible that this subdivision is taken into account and considered in the estimated parameters describing the population structure (Caballero 1994; Wang & Caballero 1999).

However, in real scenarios it is not always easy to establish the presence of population structuring. Thus, for instance, geographical distances (and, indeed, other sorts of 'distances' such as type differences) increase the probability that two individuals have common recent ancestors in the pedigree if the distance is short. Regarding herds, a population substructure leads to a situation in which some herds act as a nucleus, never receiving alleles from the rest of the population; these herds have higher increases in inbreeding than the others, while the mean

**Table 1** The number of individuals born in the last 10 years of the available records for all the pedigrees ( $N$ :total,  $N_f$ : females,  $N_m$ : males), realized effective population size ( $\overline{N_e}$ ) and the standard error (SE), effective size based on coancestries ( $\overline{N_{ec}}$ ) and the standard error (SE), the mean of equivalent complete generations ( $t$ ) and the standard error (SE), and number of computed coancestries ( $nc$ ) for individuals included in the reference sub-population for each breed

Population	$N$	$N_m$	$N_f$	$\overline{N_e} \pm SE$	$\overline{N_{ec}} \pm SE$	$t \pm SE$
Spanish Purebred (SPB)	4460 <sup>a</sup>	2209	2251	49.6 $\pm$ 4.75	72.9 $\pm$ 0.47	9.9 $\pm$ 0.6
Carthusian	1721	834	887	19.8 $\pm$ 1.51	24.2 $\pm$ 0.31	9.1 $\pm$ 0.7
Spanish Arab (SA)	6240	3032	3208	34.2 $\pm$ 3.93	51.3 $\pm$ 0.72	7.9 $\pm$ 1.3
Spanish Thoroughbred (TB)	3440	1910	1530	322.1 $\pm$ 74.19	368.3 $\pm$ 3.73	3.8 $\pm$ 0.7
Spanish Anglo-Arab (dAA)	2668	1297	1371	180.5 $\pm$ 43.40	291.2 $\pm$ 2.72	5.7 $\pm$ 1.1 (2.5) <sup>b</sup>
Hispano-Arab (dHA)	1783	925	858	128.5 $\pm$ 42.66	139.6 $\pm$ 1.50	7.7 $\pm$ 2.0 (1.4) <sup>b</sup>
Spanish Sport Horse (dSSH)	6055	2827	3228	176.7 $\pm$ 71.68	819.1 $\pm$ 8.93	3.4 $\pm$ 2.0 (1.2) <sup>b</sup>
Set of Horse populations	5472 <sup>c</sup>	2679	2793	52.6 $\pm$ 6.08	101.16 $\pm$ 1.03	9.09 $\pm$ 2.04

<sup>a</sup>Selected by random from a total of  $N = 89324$ .

<sup>b</sup>Average equivalent complete generations computed excluding pedigree information of parents from parental studbook.

<sup>c</sup>Selected by random from a total of  $N = 109510$ .

coancestry grows equally among all of them (Vassallo *et al.* 1986). Elite groups within populations would also lead to the same situation; the mating between elite and non-elite animals would produce offspring that are excluded from the elite group. All these examples and many others provide higher increases in inbreeding than in coancestry as a consequence of population substructure. Under a regular system of inbreeding, and with a very deep knowledge of pedigree, both the increase in inbreeding and the increase in coancestry will attain an identical asymptotic value except for the scenarios with permanent subdivision, but a long historical pedigree are not common in real livestock populations.

If the aim is to analyze the inbreeding depression, the increase in inbreeding will provide a sound value, but when the goal is to ascertain the real effective number of breeding individuals in a structured population, the estimation of the effective population size from the increase in inbreeding performs poorly whatever the method used (Falconer & MacKay 1996; Gutiérrez *et al.* 2008). Inbreeding coefficients are highly affected by population structuring and, in this case, the use of other methods such as that by Hill (1979), gathering the demography of a population, are more advisable (Cervantes *et al.* 2008). Some other approaches developed to predict effective population size of structured populations are based on F-statistics (Wright 1969) and the variances of offspring size of the identified subpopulations (see Wang & Caballero 1999 and Caballero & Toro 2002). These kinds of approaches, however, are not likely to be applied to real populations.

In our study, we propose a simple method to estimate effective population size, directly, from the

increase in coancestry between each pair of individuals  $j$  and  $k$  ( $\Delta c_{jk}$ ) in a given reference population. This method is an extension of the methodology previously used for individual increase in inbreeding by Gutiérrez *et al.* (2008, 2009) that overcomes the difficulty of assigning each individual in a given population to a generation and, therefore, applicable to real pedigrees to obtain reliable estimates of effective population size based on coancestry ( $\overline{N_{ec}}$ ). Moreover, a between-individuals coancestry matrix, standardized for the completeness of pedigree, can be built and analyzed using other statistical tools, such as that based on principal component analysis, to illustrate population structure (Jackson 1981).

Furthermore, it has been shown that the comparison between the new parameter and the individual increase in inbreeding gives information on the degree of population structure.  $\overline{N_{ec}}$  and its brother parameter  $\overline{N_e}$  (Gutiérrez *et al.* 2008) are assumed to be measures of the same cumulated drift processes from the foundation of the population to the present. As they would be asymptotically equivalent in an idealized population, the disagreement between them mainly falls on their ability to assess the effect of preferential matings. In other words, the comparison between  $\overline{N_{ec}}$  and  $\overline{N_e}$  would characterize the influence of preferential matings in the population. The method achieved reliable results in the three scenarios of simulated pedigrees giving the theoretical value for each simulation. In the case of simulations i) and ii) (Figure 1), both  $\overline{N_{ec}}$  and  $\overline{N_e}$  reflect the situation produced by the subdivision effect. Moreover, when the subdivision is combined with a scenario overlapping generations, the estimated

$\bar{N}_{ec}$  values approach the theoretical  $N_e$  values better than those obtained using unadjusted methodologies.

The horse example is useful to discuss the information provided by both effective sizes. The available Spanish horse populations can be studied separately or as a set of populations. Note that the  $\bar{N}_{ec}$  was computed in this case using coancestries between males and females to avoid computational limitations, but this procedure can also be useful to reflect a more realistic effective size from a predictive point of view if we are only interested in the next generation. The comparison between increases in inbreeding and in coancestry, allowed us to ascertain the presence of population structure within all the parental populations analyzed, except for TB. Consistently with the history of the populations, the ratio  $\bar{N}_{ec}/\bar{N}_e$  significantly deviates from 1 thus reflecting the presence of a particular management within the Carthusian, SPB and SA populations because of the existence of diffused subpopulations originated by famous studs and preferred by some breeders (see Valera *et al.* 2005). This is more evident in the case of SPB that includes in its studbook the Carthusian strain (Valera *et al.* 2005). The similar values for both the effective sizes computed for TB should be interpreted with caution because they are more likely to reflect a shallow pedigree (only 3.8 equivalent discrete generations) because of the continuous use of foreign TB stallions with unavailable genealogies (Cervantes *et al.* 2009).

The results from the composite horse breeds are more difficult to interpret. Both dAA and dHA had sufficiently deep pedigrees and assessing population structure from the last generation can give the idea of the unstructured breeds. The  $\bar{N}_{ec}$  was 1.6 higher than  $\bar{N}_e$  on the dAA pedigree, reflecting a similar structure to parental populations such as SPB or SA, but the standard error found was higher in this pedigree. In dHA that had the deepest pedigree of the derived breeds, both values were very similar. This would characterize the composite dHA breed as a population with good pedigree information and random matings. However, the low effective size estimated by Cervantes *et al.* (2009) in the dHA breed suggests that this is more likely to be a consequence of the small proportion of matings available between individuals of the composite breed. In Table 1, we have included the equivalent discrete generations computed when parents from pure breeds are considered as founders to show that pedigree within the composite breed is very shallow in this case.

Regarding the dSHH breed, we can assume that the recent foundation of its studbook and the lower pedigree depth (3.4 equivalent generations) would lead to increases in coancestry that are not translated into increases in inbreeding. However, within breed, individuals are selected for three different disciplines (Dressage, Show jumping and Eventing). These multiple selection policies probably lead to differential matings with specific stallions to improve the desired performance given an overall increase in inbreeding with no parallel increase in coancestry (with  $\bar{N}_{ec}$  almost 5 times higher than  $\bar{N}_e$ ). Therefore, original founder contributions seem to be kept by mating individuals always within a selection objective thus maintaining a high equivalent structuring value.

The results from the example of the set of horse populations gave a ratio between the effective sizes close to two. SPB individuals are nearly 90% of the total set. This extremely high SPB representation in the total set leads to a situation in which the other breeds appear as isolated individuals in a big and, more or less, connected population.

## Conclusions

The estimation of effective size from increases in pairwise coancestry introduced here has been shown to perform reasonably well in real populations. This can be compared with other methods to compute  $N_e$  from the increases in inbreeding, namely that from the individual increase in inbreeding to characterize the structuring degree. Population managers would prefer one or the other methods if descriptive or predictive purposes are involved, but both of them provide interesting information. Note that assuming some inbreeding, structuring would allow lower coancestry and consequently less drift. The comparison between the increases in coancestry and in inbreeding has performed reasonably well in both the simulated and the real examples used in this study. These enable us to consider the population as a whole to identify whether the mating policy carried out in the population tends to create substructure (lines) within it or not.

## Acknowledgements

The authors thank the 'Subdirección General de Medios de Producción Ganaderos' of the Spanish Ministry of Agriculture for their support. We also thank the 'Fondo de Explotación de los Servicios de Cría Caballar y Remonta' of the Spanish Ministry of

Defence and the Horse Breeders' Association for their collaboration.

## References

- Boichard D., Maignel L., Verrier E. (1997) The value of using probabilities of gene origin to measure genetic variability in a population. *Genet. Sel. Evol.*, **29**, 5–23.
- Caballero A. (1994) Developments in the prediction of effective population size. *Heredity*, **73**, 657–679.
- Caballero A., Toro M.A. (2002) Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv. Genet.*, **3**, 289–299.
- Cervantes I., Goyache F., Molina A., Valera M., Gutiérrez J.P. (2008) Application of individual increase in inbreeding to estimate realised effective sizes from real pedigrees. *J. Anim. Breed. Genet.*, **125**, 301–310.
- Cervantes I., Gutiérrez J.P., Molina A., Goyache F., Valera M. (2009) Genealogical analyses in open populations: the case of three Arab-derived Spanish horse breeds. *J. Anim. Breed. Genet.*, **126**, 335–347.
- Chesser R.K., Rhodes O.E. Jr, Sugg D.W., Schnabel A. (1993) Effective Sizes for Subdivided Populations. *Genetics*, **135**, 1221–1232.
- Duchev Z., Distl O., Groeneveld E. (2006) Early warning system for loss of diversity in European livestock breeds. *Archiv. Anim. Breed.*, **49**, 521–531.
- Falconer D.S., Mackay F.C. (1996) Introduction to quantitative genetics, 4th edn. Longman Group Ltd., England.
- FAO (1998) Secondary guidelines for the national farm animal genetic resources management plans: management of small populations at risk. FAO, Rome, Italy.
- Fernández J., Toro M.A., Caballero A. (2008) Management of Subdivided Populations in Conservation Programs: Development of a Novel Dynamic System. *Genetics*, **179**, 683–692.
- González-Recio O., López de Maturana E., Gutiérrez J.P. (2007) Inbreeding depression on female fertility and calving ease in Spanish dairy cattle. *J. Dairy Sci.*, **90**, 5744–5752.
- Gutiérrez J.P., Goyache F. (2005) A note on ENDOG: a computer program for analysing pedigree information. *J. Anim. Breed. Genet.*, **122**, 357–360.
- Gutiérrez J.P., Altarriba J., Díaz C., Quintanilla R., Cañón J., Piedrafita J. (2003) Pedigree analysis of eight Spanish beef cattle breeds. *Genet. Sel. Evol.*, **35**, 43–63.
- Gutiérrez J.P., Cervantes I., Molina A., Valera M., Goyache F. (2008) Individual increase in inbreeding allows estimating effective sizes from pedigrees. *Genet. Sel. Evol.*, **40**, 359–378.
- Gutiérrez J.P., Cervantes I., Goyache F. (2009) Improving the estimation of realised effective population sizes in farm animals. *J. Anim. Breed. Genet.*, **126**, 327–332.
- Hill W.G. (1979) A note on effective population size with overlapping generations. *Genetics*, **92**, 317–322.
- Jackson J.E. (1981) Principal components and factor analysis: part III. what is factor analysis? *J. Qual. Technol.*, **13**(2), 125–130.
- Maignel L., Boichard D., Verrier E. (1996) Genetic variability of French dairy breeds estimated from pedigree information. *Interbull Bulletin*, **14**, 49–54.
- Maxwell J. (1995) Spanish Arabian Horse Families. Alexander Heriot and Co Ltd, England.
- Valera M., Molina A., Gutiérrez J.P., Gómez J., Goyache F. (2005) Pedigree analysis in the Andalusian horse: population structure, genetic variability and influence of the Carthusian strain. *Livest. Prod. Sci.*, **95**, 57–66.
- Vassallo J.M., Díaz C., García-Medina J.R. (1986) A note on the population structure of the Avileña breed of cattle in Spain. *Livest. Prod. Sci.*, **15**, 285–288.
- Wang J., Caballero A. (1999) Developments in predicting the effective size of subdivided populations. *Heredity*, **82**, 212–226.
- Wright S. (1969) Evolution and the Genetics of Populations: The Theory of Gene Frequencies, vol. 2. University of Chicago Press, Chicago, USA.