

ORIGINAL ARTICLE

Improving the estimation of realized effective population sizes in farm animals

J.P. Gutiérrez¹, I. Cervantes¹ & F. Goyache²

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

² SERIDA-Somió, C/Camino de los Claveles 604, E-33203 Gijón (Asturias), Spain

Keywords

Effective size; increase in inbreeding; overlapped generations; real populations.

Correspondence

J.P. Gutiérrez, Departamento de Producción Animal, Facultad de Veterinaria, Avda. Puerta de Hierro s/n, E-28040-Madrid, Spain.

Tel: +34913943767;

Fax: +34913943767;

E-mail: gutgar@vet.ucm.es

Received: 5 January 2009;

accepted: 22 March 2009

Summary

Computation of inbreeding rate (ΔF) must consider that inbreeding is delayed with one generation with respect to the idealized population when addressed using individual inbreeding coefficients. The expression relating inbreeding in generation t with inbreeding rate $F_t = 1 - (1-\Delta F)^t$ should be more correctly written in real animal populations as $F_t = 1 - (1-\Delta F)^{t-1}$, as changes in allele frequencies occur in the equivalent co-ancestries in the previous generation. This simple approach is tested on simulated and real pedigrees thus demonstrating that: (i) the adjusted individual increase in inbreeding becomes stable in populations under random mating while the unadjusted parameter does not; (ii) regression of the unadjusted parameter over generations in pedigrees under random mating is highly significant while after correction it is not significant; and (iii) the variance of the adjusted parameter is reduced with the generations.

Introduction

Effective population size (N_e), the number of individuals of an equivalent idealized population that would give rise to the rate of inbreeding, or the rate of change in variance of gene frequencies observed in the population under consideration (Wright 1969), is one important research topic in population genetics, given its usefulness as a measure of the long-term performance of the population regarding both diversity and inbreeding (Fernández *et al.* 2005) and, therefore, for characterizing the risk status of livestock breeds (FAO 1998; Duchev *et al.* 2006).

Real populations are usually bred under conditions far from that of the idealized population to which genealogical tools are referred (Falconer & Mackay 1996), namely, generation overlapping. Therefore, differences in pedigree depth between contemporary

individuals and difficult assignment of an individual to a given generation are a norm. In such a scenario, interpretation of N_e is not straightforward. To overcome such concerns, the researchers have assayed estimating inbreeding rate from mean individual inbreeding coefficients and their evolution by time or generation, usually via regression (Pérez-Enciso 1995; Gutiérrez *et al.* 2003). Recently, the individual increase in inbreeding (González-Recio *et al.* 2007) has been proposed as a measure of standardized inbreeding rate per generation. Gutiérrez *et al.* (2008) have used this parameter to give an estimate of N_e (called 'realized' N_e by Cervantes *et al.* 2008) on real pedigrees.

When self-fertilization is excluded, effective population size can be approached as $N_e = 1/2 + N$ (Wright 1931), with N being the number of breeding individuals. However, this expression works only asymptotically and thus, because pedigrees in real populations

are usually shallow and do not usually follow an asymptotic pattern, this approximation is of little value. Moreover, in real diploid populations self-fertilization is not possible and, even when assuming a constant random mating policy, if addressed from individual inbreeding coefficients, asymptotic values of the inbreeding rate will be reached after more generations than those expected in the idealized population.

Moreover, from all the exceptions to the idealized population in mammals (such as, e.g. avoidance of sib-matings, see Falconer & Mackay 1996), self-fertilization is the only one that occurs independently of the mating policy and must be considered as a general property of farm populations. Therefore, while considering any exception to the idealized population, which comes from mating policies in genealogical studies, is optional and dependent on the scenario of a particular population, considering that the exception of self-fertilization can only be mandatory in mammals.

Some approximations have been proposed to deal with populations with separate sexes based on the variance of family sizes given a deviation from Hardy–Weinberg equilibrium (Hill 1972, 1979; Crow & Denniston 1988; Caballero & Hill 1992), or using an inbreeding approach under a subdivision context assuming an equal size of subpopulations with unequal contributions to the next generation (Nunney 1999). However, these methods work under assumptions that poorly fit with real populations.

Here, we suggest modifying the computation of increases in inbreeding to obtain sound estimates of realized effective sizes in real populations (Cervantes *et al.* 2008; Gutiérrez *et al.* 2008), in which inbreeding coefficients are erroneously assigned to the generation of the animal instead of the generation of their parents. Inbreeding coefficients must be interpreted as a result of the co-ancestry between the individuals mated in the previous generation. Therefore, it is well known that inbreeding appears one generation later than co-ancestry. This delay is particularly important when pedigrees are shallow. In such a scenario, computation of inbreeding rates must consider that inbreeding is delayed one generation with respect the co-ancestry between parents. This little but important modification to the previous approach is tested on the recently proposed parameter individual increase in inbreeding, which has been shown to work reasonably well in most scenarios (Cervantes *et al.* 2008; Gutiérrez *et al.* 2008). Performance of the suggested modification is demonstrated on some simulated and real examples (Cervantes

et al. 2008; Gutiérrez *et al.* 2008) previously used to describe the original, now modified, parameter.

Materials and methods

Background

General population genetic theory predicts non-zero inbreeding after one generation. However, when self-fertilization is avoided, inbreeding equals zero after one generation, as mating between relatives cannot occur. Thus, measuring drift fails when it is measured by the inbreeding coefficient of the individuals. In fact, drift actually increases because of divergence of gene frequency from the base population, which occurs immediately, but its effect on inbreeding coefficients appears one generation later. Given that the inbreeding coefficient of an individual equals co-ancestry between its parents, co-ancestry coefficients are not affected by this delay.

Increase in inbreeding cannot be computed in real animal populations in the same manner as in the idealized population as self-fertilization is not possible. However, accounting for a delay in one generation should be enough to correct such a limitation. Basically, the expression $F_t = \Delta F + (1 - \Delta F)F_{t-1}$ (Falconer & Mackay 1996) is valid for $t > 1$ but not for $t = 0$ because inbreeding coefficients in the first generation are zero ($F_1 = 0$). Consequently, panmictic index P (the complement of the inbreeding coefficient: $1 - F$) in generation t (P_t ; Falconer & Mackay 1996) is equal to 1 in the first generation. P_t can thus be computed as: $P_t = (1 - \Delta F)^{t-1}$; $P_1 = (1 - \Delta F)^{t-1}$, which leads to the expression $F_t = 1 - (1 - \Delta F)^t$ to be more correctly written as: $F_t = 1 - (1 - \Delta F)^{t-1}$ in animal populations without self-fertilization.

The original individual increase in inbreeding described by González-Recio *et al.* (2007) was defined as: $\Delta F_i = 1 - \sqrt[t_i]{1 - F_i}$, with t_i being the equivalent of the discrete generations (Maignel *et al.* 1996; Boichard *et al.* 1997) and F_i the inbreeding coefficient of an individual i . The ΔF_i values of the individuals belonging to the reference population could be averaged to give $\overline{\Delta F}$.

The parameter t is computed as the sum over all known ancestors of the term of $(\frac{1}{2})^n$, where n is the number of generations separating the individual from each known ancestor (Boichard *et al.* 1997; Gutiérrez *et al.* 2008).

From this, a mean effective population size $\overline{N_e}$ (Gutiérrez *et al.* 2008), called realized effective size by Cervantes *et al.* (2008), can be straightforwardly

computed as: $\overline{N_e} = \frac{1}{2\Delta F}$. Under a scenario without self-fertilization, the methodology should be modified and the adjusted individual increase in inbreeding should be written as: $\Delta F_i^* = 1 - \sqrt{1 - F_i}$ and averaged as: $\overline{\Delta F^*}$ to give an adjusted realized effective population size as: $\overline{N_e^*} = \frac{1}{2\overline{\Delta F^*}}$. Note that individuals with less than two discrete generations have to be dropped from the sample used to fit a reference population given their insufficient pedigree depth to obtain a not null inbreeding coefficient.

Expressions for the standard errors of the different parameters (Cervantes *et al.* 2008; Gutiérrez *et al.* 2008) can be applied as such in those adjusted expressions, given that they are inferred from their own parameters and variability.

Examples

A simulated example already analysed in Gutiérrez *et al.* (2008), has been used to test the performance of the adjusted parameter. In the simulation, the founder population consisted of 100 males and 100 females. Mating at random was undertaken for 50 generations. A total of 100 males and 100 females were born in each generation and acted as parents of the following generation under random mating with the individuals of the other sex and no differential viability or fertility. Ten replicates with this structure were analysed here and the results for each replicate were averaged by generation. The theoretical N_e excluding self-fertilization is the number of breeding individuals plus one half (Wright 1931) or, in our case, 200.5.

The evolution and dispersion of the new adjusted parameters are also demonstrated on a real example fully described and previously analysed under the original methodology by Cervantes *et al.* (2008). Briefly, we used the pedigree information available from the beginning of the twentieth century in the Fighting Bull herd of Juan Pedro Domecq, including a total of 33 330 individuals. This is a closed population of constant size with a permanent mating policy avoiding common ancestors in the three previous generations of each individual thus resembling a random mating policy. Such a mating policy is expected to give stable increases in inbreeding throughout the time periods and generations.

Program used

The analyses were performed using the ENDOG program (available for Windows environment, current version v4.6; Gutiérrez & Goyache 2005), which can

be freely downloaded from the World Wide Web at http://www.ucm.es/info/prodanim/html/JP_Web.htm.

Results and discussion

Figure 1 shows the dispersal of the adjusted individual increase in inbreeding of the individuals generated in the simulated data, starting from the second generation and the linear trends of the original and the adjusted parameter. Theoretically, a constant mating policy should lead to constant rates of inbreeding throughout generations and thus, even without looking at the distribution of individual increases in inbreeding, a simple regression of that parameter on generation number should be zero. Interestingly enough, the estimated regression coefficient of ΔF_i over equivalent discrete generation was significant ($p < 0.000001$), but that of ΔF_i^* over equivalent discrete generation was not significant ($p = 0.74426$).

Figure 2 shows the evolution of the effective size estimated from both original (black line) and adjusted (grey line) increases in inbreeding averaged by the 10 replicates of the simulated examples. Interestingly enough, effective size computed from the adjusted parameter approaches the real N_e value without the delay shown in the original parameter.

Figure 3 shows the dispersal of the original [plot (a)] and adjusted [plot (b)] individual increase in inbreeding by generation in the Fighting Bull

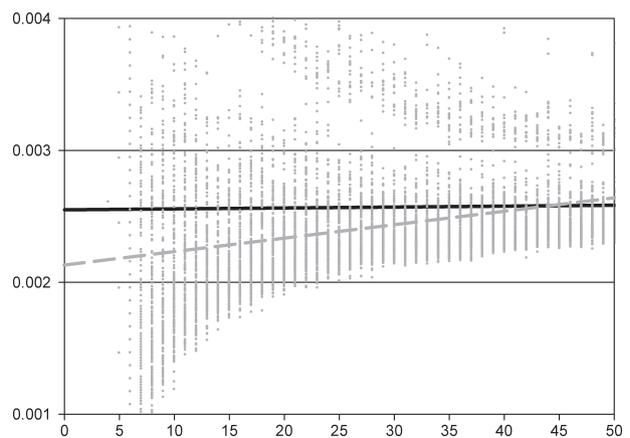


Figure 1 Dispersal of the adjusted individual increase in inbreeding from the second generation of one random simulated example. Although a reduction in the variability of the parameter over generations was statistically evident with respect to the non-adjusted individual increase in inbreeding, the dispersion of the latter parameter is not shown as no visual differences were noticeable between clouds of points. In any case, linear trends for the original (broken grey line) and adjusted (continuous dark line) parameters are shown.

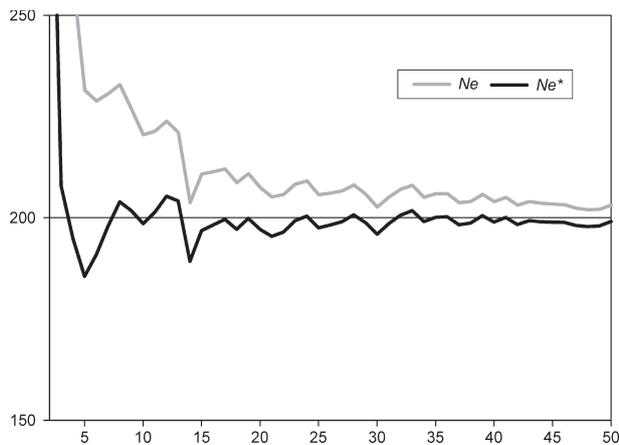


Figure 2 Evolution in the simulated examples of the mean of the effective sizes computed, from the second generation, using both original (grey line) and adjusted (dark line).

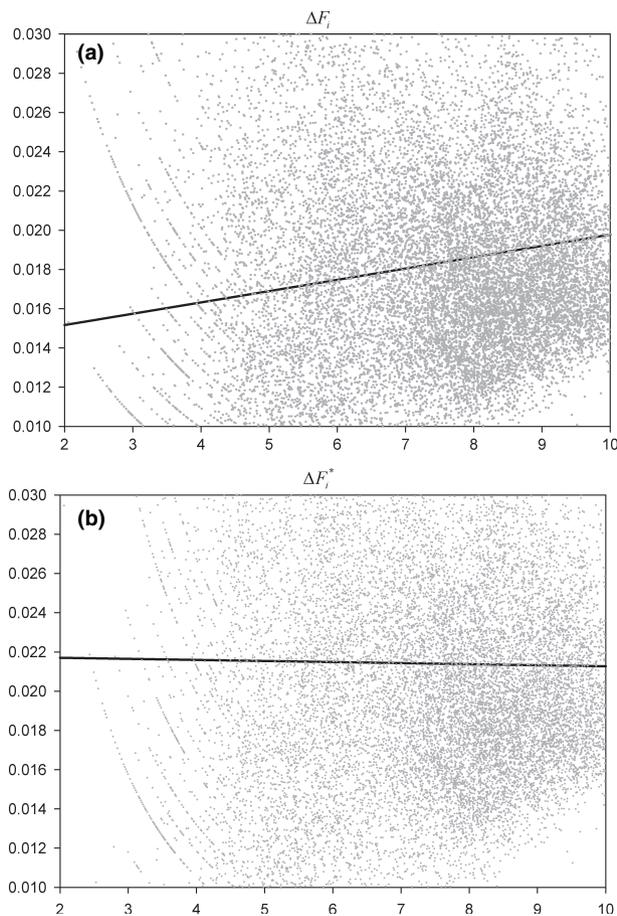


Figure 3 Dispersal of the original [plot (a)] and adjusted [plot (b)] individual increase in inbreeding from the second equivalent to discrete generation of the Fighting Bull population used as an example. Linear trends for both parameters are also shown.

population as an example of the real population with a stable mating policy in which the parameter is expected to take a constant value. The linear trends of the parameters are also drawn showing that the adjusted individual increase in inbreeding gave the expected behaviour while the original one did not. Again, regression over equivalent to discrete generations was significant for the unadjusted parameter ($p < 0.000001$) while it was not significant for the adjusted parameter ($p = 0.215891$) as expected under the constant mating policy.

The estimation of N_e from individual increase in inbreeding has been shown to fit reasonably well with most animal population scenarios using simulated (Gutiérrez *et al.* 2008) and real (Cervantes *et al.* 2008) examples. These authors stated that, although this methodology addresses N_e directly from ΔF , which theoretically becomes constant in an ideal population. A few generations were needed to reach a stable value that asymptotically leads to an estimate of effective population size. However, asymptotic properties are not useful in livestock populations with finite pedigree knowledge. Here, we present a slight but important modification to the methodology that overcomes this concern. It takes into account that mating between animals with the same gender is impossible in mammals. In this situation, drift arising in a generation only can be measured by means of the co-ancestry between the mated animals, which equals the inbreeding coefficients in their offspring. Therefore, inbreeding coefficients must be considered as having been reached in $t-1$ generations instead of t .

Given that the mating in farm animals is not possible between individuals of the same sex, the modified parameter must be considered to be of general use. Note that this is the only break of the assumptions on idealized population applying in all mammal populations, unlike others that are dependent on the mating policy. Thus, the next break that is usually considered (Falconer & Mackay 1996) is the one in which mating between sibs are avoided. Such a scenario is highly interesting from a theoretical point of view but, although it is claimed to be used, in fact it is usually only roughly applied. For instance, the proportion of mating between sibs included in the examples analysed in Cervantes *et al.* (2008) varied from 1.57% to 3.48%. This was so even in the Fighting Bull population analysed, where the mating policy aimed at avoiding mating between closely related individuals. This practice is usual when breeders try (although occasionally) to fix the phenotype of some traits.

As stated in the *Background* section, the underlying idea is that, with respect to the ideal population, the inbreeding coefficient of an individual belonging to a population with two different sexes, in which self-fertilization is excluded, will carry a delay of one generation. This simple modification leads to the adjusted individual increase in inbreeding to become stable in populations under random mating (Figures 1 and 3) while the original increase in inbreeding does not. In a scenario near the idealized population (Figure 1), except for the exclusion of self-fertilization, the linear trend of the parameter over generations changes from a highly significant positive regression coefficient to a clear null regression coefficient after the correction. Along with this effect, the variance of the adjusted parameter is reduced with the generations, the larger the number of generations the larger the reduction of the variance, thus improving the good ability of the non-adjusted parameter to estimate the effective size.

Cervantes *et al.* (2008) reported that the trend of the ΔF_i values tended to stabilize with the increase in t owing to the correction for pedigree depth of the individuals. However, the ΔF_i values were still dependent on the shallowness of the analysed pedigree and they needed a few generations to become constant at the population level (approximately five equivalents to discrete generations). Thus, breeders might wait for few more generations to make proper use of the parameter. Note for instance that horses have a generation interval of approximately 10–12 years; a delay in two to four generations to obtain sound estimates of the parameter would represent 20–40 years of a breeder's life.

If the small, but significant, change we presented in our manuscript is considered, breeders would obtain useful estimates immediately. Cervantes *et al.* (2008) showed that individual increase in inbreeding was not stable in the initial generations as a consequence of ignoring the delay in the inbreeding in diploid species. In contrast, Figure 3 shows how the new parameter holds a noticeable stability of the increases in inbreeding. As a practical consequence, breeders working under ΔF^* will gather generations that are enough to compute the realized effective size earlier than working under ΔF . This is especially true for populations approaching the random mating policy (Figure 3). In any case, the higher the number of generations in the pedigree, the better is the performance of the adjusted individual increase.

When few generations are considered, the skewness in the distribution of ΔF_i^* is important as a consequence of few animals presenting high inbreeding

coefficients (and also high increase in inbreeding coefficients) while coefficient for many others is null; the higher the effective size of the population, the higher and more persistent is the skewness throughout the generations. Together with this skewness, the parameter would show a higher variance thus affecting the precision of the estimation of effective size.

Figure 1 visually shows the effect of increasing the number of generations on ΔF_i , improving the quality of all related parameters. Moreover, it can be suggested that genealogical studies of real diploid animal populations should not include animals with a pedigree depth lower than two generations, as they have not had the chance of being inbred.

As a conclusion, we introduced a new formula to calculate individual increase in inbreeding. The new formula was used in a method to estimate effective population size in a diploid population without self-fertilization. This new formula did not overestimate the effective population size unlike its predecessor. In addition, a good estimate by the new formula required fewer generations in the pedigree than the older one.

Acknowledgements

The authors thank J.P. Domecq and the staff of the Fighting Bull Breeder Association (UCTL) for providing access to their pedigree data.

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., Hill W.G. (1992) Effective size of non random mating populations. *Genetics*, **130**, 909–916.
- 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.
- Crow J.F., Denniston C. (1988) Inbreeding and variance effective population numbers. *Evolution*, **42**, 482–495.
- 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*. 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.B., Villanueva B., Pong-Wong R., Toro M.A. (2005) Efficiency of the use of pedigree and molecular

- marker information in conservation programs. *Genetics*, **170**, 1313–1321.
- 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**, 172–176.
- Gutiérrez J.P., Altarriba J., Díaz C., Quintanilla R., Cañón J., Piedrafita J. (2003) Genetic analysis of eight Spanish beef cattle breeds. *Genet. Sel. Evol.*, **35**, 43–64.
- 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.
- Hill W.G. (1972) Effective size of population with overlapping generations. *Theor. Popul. Biol.*, **3**, 278–289.
- Hill W.G. (1979) A note on effective population size with overlapping generations. *Genetics*, **92**, 317–322.
- Maignel L., Boichard D., Verrier E. (1996) Genetic variability of French dairy breeds estimated from pedigree information. *Interbull. Bull.*, **14**, 49–54.
- Nunney L. (1999) The effective size of a hierarchically structured population. *Evolution*, **53**, 1–10.
- Pérez-Enciso M. (1995) Use of the uncertain relationship matrix to compute effective population size. *J. Anim. Breed. Genet.*, **112**, 333–340.
- Wright S. (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.
- Wright S. (1969) Evolution and the Genetics of Populations: The Theory of Gene Frequencies, Vol. 2. University of Chicago Press, Chicago, IL.