Original article

Individual increase in inbreeding allows estimating effective sizes from pedigrees

Juan Pablo Gutiérrez[1]*, Isabel Cervantes[1], Antonio Molina[2],
Mercedes Valera[3], Félix Goyache[4]

[1]Departamento de Producción Animal, Facultad de Veterinaria, Avda. Puerta de Hierro s/n, 28040 Madrid, Spain
[2]Departamento de Genética, Universidad de Córdoba, Ctra. Madrid-Cádiz, km 396ª, 14071 Córdoba, Spain
[3]Departamento de Ciencias Agro-Forestales, EUITA, Universidad de Sevilla, Ctra. Utrera km 1, 41013 Sevilla, Spain
[4]SERIDA-Somió, C/ Camino de los Claveles 604, 33203 Gijón (Asturias), Spain

Abstract – We present here a simple approach to obtain reliable estimates of the effective population size in real world populations *via* the computation of the increase in inbreeding for each individual (delta F_i) in a given population. The values of delta F_i are computed as t-root of $1 - (1 - F_i)$ where F_i is the inbreeding coefficient and t is the equivalent complete generations for each individual. The values of delta F computed for a pre-defined reference subset can be averaged and used to estimate effective size. A standard error of this estimate of N_e can be further computed from the standard deviation of the individual increase in inbreeding. The methodology is demonstrated by applying it to several simulated examples and to a real pedigree in which other methodologies fail when considering reference subpopulations. The main characteristics of the approach and its possible use are discussed both for predictive purposes and for analyzing genealogies.

effective size / increase in inbreeding / overlapped generation / genetic contribution

1. INTRODUCTION

The effective population size (N_e), defined as 'the size of an idealized population which would give rise to the rate of inbreeding, or the rate of change in variance of gene frequencies observed in the population under consideration' [27], is a key parameter in conservation and population genetics because of its direct relationship with the level of inbreeding, fitness and the amount of genetic variation loss due to random genetic drift [5,7]. As a consequence, N_e

---

*Corresponding author: gutgar@vet.ucm.es

is usually considered as a useful criterion for classifying the livestock breeds according to the degree of endangerment [6,8].

When genealogies are available, the effective population size can be estimated from the increase in inbreeding (ΔF) between two discrete generations as in $N_e = \frac{1}{2\Delta F}$, with $\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}}$, where F_t and F_{t-1} are the average inbreeding at t and $t-1$ generations [7]. The increase in inbreeding is constant for an ideal population of constant size with no migration, no mutation and no selection over discrete generations. However, in real populations with overlapping generations, the number of males and females is usually different and non-random mating is the rule, making ΔF a difficult parameter to deal with [7]. In most cases the definition of a 'previous' generation is quite difficult to establish. In fact, taking the average inbreeding of a pre-defined reference subpopulation and referring it to the founder population in which inbreeding is null by definition, fits poorly in any given real population and is only acceptable in small populations with shallow pedigree files [1,10,12] leading to the risk of overestimating the actual effective population size.

Some attempts have been proposed to overcome these challenges in the real world, namely the computation of N_e from the variances of family sizes of males and females [7,13,14] or the use of the regression coefficient of the individual inbreeding coefficients on the number of generations known for each animal as an estimate of ΔF [12]. In a scenario of overlapping generations, computation of N_e based on family variances unrealistically ignores population subdivision and several other causes of variation of the parameter, such as mating between relatives, migration, or different representation of founders. Most methodologies applied to compute N_e under overlapping generations are also affected by the difficulties in fitting individuals to generations because data over time usually appear as registered by year regardless of when the renewal of the population is done at a generation interval. On the contrary, the computation of regression coefficients with the aim of approximating $\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}}$, also has the difficulty of defining the 'previous' generation with respect to the identified reference subpopulation. The estimation of effective size could be approximated by using $1 - F_t = \left(1 - \frac{1}{2N_e}\right)^t$ to derive its value from a log regression of $(1 - F)$ over a generation number [20], thus avoiding the need to define a previous generation. When the value of t is difficult to establish, this can be estimated by considering the year of birth as t and further correcting for the length of the generation interval [20]. However, variations in the breeding policy, such as planning mating to minimize coancestry after a period in which mating between close relatives was preferred, can lead to a temporal decrease in average inbreeding. When the animals of interest are those born in the period in which

the inbreeding decreased, methods based on assessing the increase in inbreeding would lead to negative values of N_e.

Moreover, in real populations in which selection is likely to occur, an increase in inbreeding is not a consequence of the sole accumulative change of gene frequency of a neutral gene over generations but of the long-term genetic contributions made by the ancestors [25,26]. In fact, the average inbreeding coefficient of a current reference subpopulation depends on both the number of generations separating this reference subpopulation from the founder population, and how rapidly the inbreeding accumulates.

The concept of effective size can therefore be interpreted not only as a useful parameter to predict inbreeding, but also as a tool to analyse genealogies [5]. Many attempts have been made to deal with the different real world scenarios in order to obtain reliable estimates of the effective population size [4,5]. However, there is no standard method for general application to obtain the effective population size. Here we present a straightforward approach to deal with this task by the computation of the increase in inbreeding for each individual (ΔF_i) in a given population. The values of ΔF_i are useful to obtain reliable estimates of N_e. The N_e estimated this way roughly describes the history of the pedigrees in the population of interest. The approach directly accounts for differences in pedigree knowledge and completeness at the individual level but also, indirectly, for the effects of mating policy, drift, overlap of generations, selection, migration and different contributions from a different number of ancestors, as a consequence of their reflection in the pedigree of each individual in the analyzed population. This approach, which is based on the computation of individual increase in inbreeding, also makes it possible to obtain confidence intervals for the estimates of N_e.

2. MATERIALS AND METHODS

2.1. Individual increase in inbreeding

We will start from a population with a size of N individuals bred under conditions of the idealized population [7]. Under these conditions the inbreeding at a hypothetical generation t can be obtained by [7]:

$$F_t = 1 - (1 - \Delta F)^t. \tag{1}$$

The idea presented here is to calculate inbreeding values and a measure of equivalent discrete generations for each animal belonging to a subgroup of animals of interest (the so called reference subpopulation) in a scenario with overlapping generations. Then, from (1), and equating the individual inbreeding

coefficient to that for a hypothetical population with all individuals having the same pedigree structure ($F_t = F_i$), an individual increase in inbreeding (ΔF_i) can be defined as

$$\Delta F_i = 1 - \sqrt[t]{1 - F_i}, \tag{2}$$

where t is the 'equivalent complete generations' [3,18] calculated for the pedigree of the individual 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. Notice that, on average, for a given reference subpopulation, t is equivalent to the 'discrete generation equivalents' proposed by Woolliams and Mäntysaari [24], thus characterizing the amount of pedigree information in datasets with overlapping generations. Parameter t has been widely used to characterize pedigree depths both in real [1,9,21] and simulated datasets [2].

The set of ΔF_i values computed for a number of individuals belonging to the reference subpopulation can be used to estimate the N_e regardless of the presence of individuals which would be assigned to different discrete generations according to their pedigree depth. The ΔF_i values of the individuals belonging to the reference population can be averaged to give $\overline{\Delta F}$. From this, a mean effective population size $\overline{N_e}$ can be straightforwardly computed as $\overline{N_e} = \frac{1}{2\overline{\Delta F}}$. Notice that this way of computing effective population size is not dependent on the whole reference subpopulation mating policy but on the mating carried out throughout the pedigree of each individual.

Moreover, since we are assuming a different individual increase in inbreeding for each individual i in the reference subpopulation, ascertaining the confidence on the estimate of $\overline{\Delta F}$ is also feasible, and the corresponding standard error can be easily computed. Kempen and Vliet [17] described how the variance of the ratio of the mean of two variables x and y can be approximated using a Taylor series expansion. Assigning in our case $x = 1$, and $y = 2\Delta F$, we can obtain the standard error of $\overline{N_e}$ as $\sigma_{\overline{N_e}} = \frac{2}{\sqrt{N}}\overline{N_e}^2\sigma_{\Delta F}$, with N being the number of individuals in the reference subpopulation, $\sigma_{\Delta F}$ the standard deviation of ΔF and $\sigma_{\overline{N_e}}$ the standard error of $\overline{N_e}$. It can also be easily shown that this is equivalent to assuming that N_e has the same coefficient of variation as ΔF.

2.2. Other methods to estimate N_e using pedigree information

Various additional approaches have been used to compare estimates of N_e obtained from individual increase in inbreeding. First, N_e was estimated from the rate of inbreeding (ΔF) or the rate of coancestry (Δf) observed between two discrete generations as, respectively, $N_e = \frac{1}{2\Delta F}$ and $N_e = \frac{1}{2\Delta f}$, with

$\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}}$ and $\Delta f = \frac{f_t - f_{t-1}}{1 - f_{t-1}}$, where F_t and F_{t-1} and f_t and f_{t-1} are the average inbreeding and the average coancestry at the t and $t-1$ generations. Moreover, N_e was estimated from the variances of family sizes as [13]

$$\frac{1}{N_e} = \frac{1}{16ML}\left[2 + \sigma_{mm}^2 + 2\left(\frac{M}{F}\right)\text{cov}(mm, mf) + \left(\frac{M}{F}\right)^2 \sigma_{mf}^2\right]$$
$$+ \frac{1}{16FL}\left[2 + \left(\frac{F}{M}\right)^2 \sigma_{fm}^2 + 2\left(\frac{F}{M}\right)\text{cov}(fm, ff) + \sigma_{ff}^2\right], \tag{3}$$

where M and F are the number of male and female individuals born or sampled for breeding at each time period, L the average generation interval σ_{mm}^2 and σ_{mf}^2 are the variances of the male and female offspring of a male, σ_{fm}^2 and σ_{ff}^2 are the variances of the male and female offspring of a female, and cov(mm, mf) and cov(fm, ff) the respective covariances. Note that the family size of a parent (male or female) consists of its number of sons and daughters kept for reproduction [14]. The three approaches described above were applied to the simulated pedigree files with the data structured in discrete generations.

When datasets with no discrete generations were analyzed, N_e was estimated from the variances of family sizes but also from ΔF using three different approaches: first, following Gutiérrez *et al.* [12], the increase in inbreeding between two generations ($F_t - F_{t-1}$) was obtained from the regression coefficient (b) of the average inbreeding over the year of birth obtained in the reference subpopulation, and considering the average generation interval (l) as follows:

$$F_t - F_{t-1} = l \times b$$

with F_{t-1} computed from the mean inbreeding in the reference subpopulation (F_t) as

$$F_{t-1} = F_t - l \times b.$$

Second, in a similar way N_e was obtained using t directly instead of considering the generations through generation intervals. By using this approach, N_e was computed from the regression coefficient (b) of the individual inbreeding values over the individual equivalent complete generations approximating t. In this case

$$\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}} \approx \frac{b}{1 - (F_t - b)} \tag{4}$$

with F_t being the average F of the reference subpopulation.

Finally, we applied the approach developed by Pérez-Enciso [20] to estimate N_e *via* a log regression of $(1 - F)$ (obtained from (1) as $1 - F_t = \left(1 - \frac{1}{2N_e}\right)^t$) on generation number. When datasets with no discrete generations were analyzed, N_e was estimated by a log regression of $(1 - F)$ on the date of birth and then divided by the generation interval [20].

2.3. Examples

The methodology is demonstrated by applying it to four simulated examples embracing a wide range of typical theoretical scenarios.

The simulated datasets evolved during 50 generations (200 periods of time in the third example under overlapping) from a founder population consisting of 200 individuals under the following assumptions:

(i) The founder population is formed by the same number of individuals of two different sexes. A total of 100 males and 100 females are born in each generation and act as parents of the following generation under random mating with the individuals of the other sex and no differential viability or fertility. The theoretical N_e excluding self-fertilization is the number of individuals $+ \frac{1}{2}$ (200.5) [7].

(ii) Like the simulated population (i) but splitting the populations in four different subpopulations consisting of 25 males and 25 females evolving separately after generation 25. The theoretical N_e is as (i) before subdivision. After that, the theoretical N_e for each subpopulation is 50.5.

(iii) Like the simulated population (i) but limiting the renewal of reproductive individuals to 25 males and 25 females each period of time and allowing the reproductive individuals to have offspring during four consecutive periods. Under overlapping generations, the expected N_e can be derived from the expression $N_e = \frac{8N_C}{V_{km} + V_{kf} + 4} L$ [7,13], where N_C is the number of reproductive individuals included in the reference subpopulation (50), V_{km} and V_{kf} are, respectively, the variances of family sizes of reproductive males and females ($V_{km} = V_{kf} = 2$ under random conditions), and L is the generation length in units of the specified time interval (2.5). Here N_e equals to 125.

(iv) Like the simulated population (i) but all parents having two offspring in the next generation. This is a case where mating is random but the variance of family sizes does not follow a Poisson distribution. The expected value of N_e computed from the expression: $N_e = \frac{8N_C}{V_{km} + V_{kf} + 4} L$ [7,13], after equalling $V_{km} = V_{kf} = 0$, is 400.
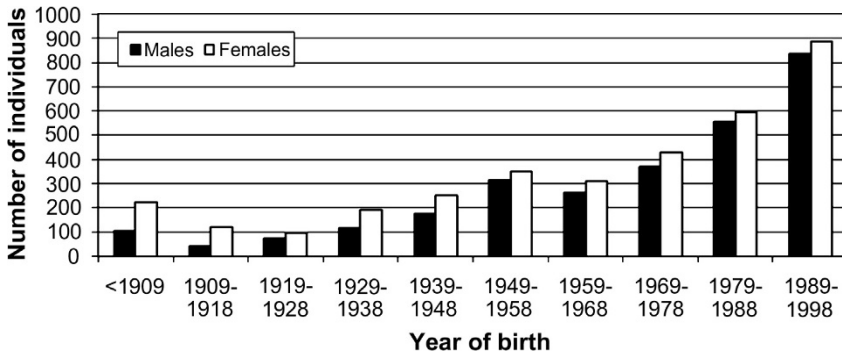
Figure 1. Evolution of registered individuals per year of birth in the Carthusian subpopulation.

The simulated pedigree files (i) to (iv) listed above are expected to characterize classical theoretical scenarios of populations evolving randomly with two sexes (i), population subdivision (ii), overlapping generations (iii), and non-Poisson variance of family sizes (iv). Within each pedigree file, a reference subset (RS) was defined as the last 400 animals born.

Additionally, the pedigree file of the Carthusian strain of the Spanish Pure-bred horse was used to demonstrate the methodology on a real example. It is a subpopulation of the pedigree file of the Andalusian horse (SPB, Spanish Pure-bred horse) [22] and included a total of 6 318 individuals since the foundation of the studbook. This population is expanding with 45% of the registered individuals born over the last 20 years. This period of time is roughly the last two generations (Fig. 1) [22]. The pedigree knowledge is reasonably high: 95% of ancestors tracing back seven generations were known and the mean equivalent complete generations for the animals born in the last decade was 9.1.

The Carthusian strain was chosen as a real example of an inbred population, because it had been subjected to a planned mating strategy using the minimum coancestry approach beginning in the 1980's [22]. Due to this mating policy, a decrease in the mean inbreeding coefficients along the period involving the last generation was also found [22]. This enables testing for the possible influence of a particular supervened breeding policy on N_e. Two RSs were defined in the Carthusian pedigree file: the individuals born in the last 10 years of available records (RS$_{10}$), and the individuals born in a given period of years allowing their use for reproduction (1977–1989; RS$_{77-89}$). The pedigree files of the fitted RSs were also edited to include only individuals with four equivalent generations or more, and eight equivalent generations or more. The main parameters describing the Carthusian pedigree file are given in Table I.

Table I. Number of individuals (N), average number of equivalent generations and standard deviation ($t \pm$ s.d.), maximum number of equivalent generations (Max. t), average inbreeding (F, in percent), number of male and female reproductive individuals and average family size for males and females (in brackets), and variances of family sizes for reproductive males (V_m) and females (V_f) for the whole Carthusian pedigree file (WP) and their reference (RS) subset used as an example in the present analyses.

	N	$t \pm$ s.d.	Max. t	F(%)	Stallions	Mares	V_m	V_f
WP	6318	6.6 ± 2.75	10.9	13.0	424 (4.9)	933 (2.3)	46.96	2.85
RS$_{10}$	1721	9.1 ± 0.68	10.9	18.6	1[a] (1.0)	5[b] (1.2)	0	0.2
RS$_{77-89}$	1464	8.2 ± 0.64	9.8	17.5	97[a] (4.1)	16[b] (2.1)	21.98	1.71

RS$_{10}$: Animals born in the last decade.

RS$_{77-89}$: Animals born between the years 1977 and 1989.

[a]Individuals born in the defined period that acted sequentially as stallions.

[b]Individuals born in the defined period that acted sequentially as mares.

2.4. Program used

The analyses were performed using the ENDOG program (current version v4.4) [11], which can be freely downloaded from the World Wide Web at http://www.ucm.es/info/prodanim/html/JP_Web.htm.

3. RESULTS

The results from the analyses carried out on the simulated pedigree files are summarized in Figure 2. A discontinuous line was drawn for the theoretical effective size as reference under the different scenarios. In the case of subdivision (ii) the theoretical effective population size was also computed as the harmonic mean over generations, which expresses the expected N_e under descriptive rather than predictive purposes.

Note the erratic behavior over generations, in Figure 2, of N_e computed using the rate of inbreeding in the idealized (plot i) and non-Poisson offspring size variance (plot iv) populations. N_e tended to fit better in the case of population subdivision (plot ii) and could not be used under a scenario with overlapping generations (plot iii). This erratic behavior was caused by the use of a single replicate in the simulation and could be overcome by using the harmonic mean of N_e by generations. Estimations of N_e based on an increase in coancestry, are, however, more precise because they are computed using much more data (all pairs of individuals rather than the number of individuals), and is almost exact in the case of all animals having identical offspring size. N_e values computed using Δf and those based on variance of family size, tended to fit well in the idealized population and in the case of overlapping generations, but it failed when considering the case of population subdivision because the method ignores that such a subdivision exists. After about eight generations, performance of the individual increase in inbreeding tended to fit better than those based on Δf and variance of family sizes in the idealized population. In the case of a population subdivision, the N_e computed from an individual increase in inbreeding fits very closely to the N_e computed as the harmonic mean of the number of animals over generations for descriptive purposes and the N_e using rate of inbreeding tended to approximate the theoretical N_e for predictive purposes. The computed effective population size using ΔF_i accounts for all historical pedigree of the individuals and the obtained N_e summarizes all the genealogical information of each individual. Therefore, the genealogies recorded before subdivision weigh much more at the time closer to the population fission but their weight decreases with the accumulation of generations. If the estimation of N_e from the generations after fission is carried out for predictive purposes the harmonic mean of N_e throughout
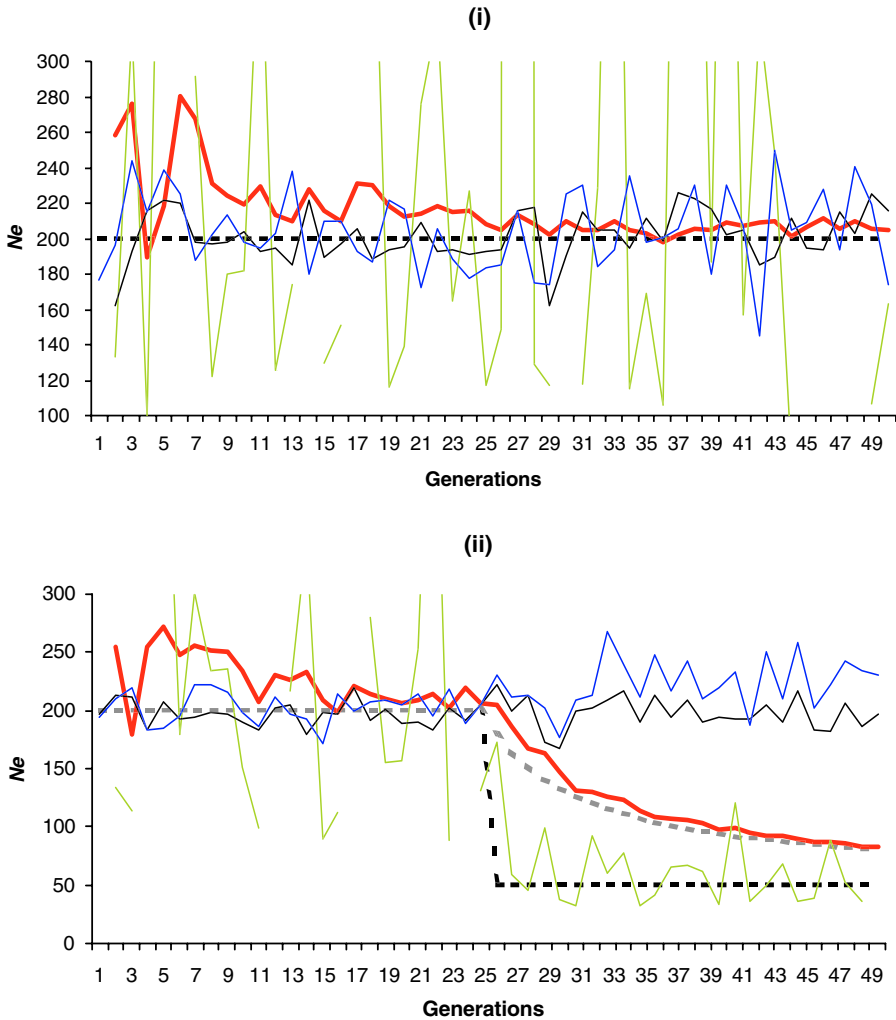
Figure 2. Variation over time of the estimates of N_e in four simulated populations. (i) Ideal population assuming two sexes; (ii) population subdivision; (iii) overlapping generations; and (iv) non-Poisson variance of family sizes.

- - - Theoretical N_e, ··· theoretical N_e by harmonic mean, — N_e from rate of inbreeding (ΔF), — N_e from the rate of increase in coancestry (Δf), — N_e from the variance in the family sizes, — N_e from individual increase in inbreeding.

generations would be preferred rather than the N_e based on individual increase in inbreeding since this converges much slower towards the 'theoretical' N_e. However, the latter better addresses the history of the population if the estimation of N_e is carried out for descriptive purposes. In the case of overlapping generations
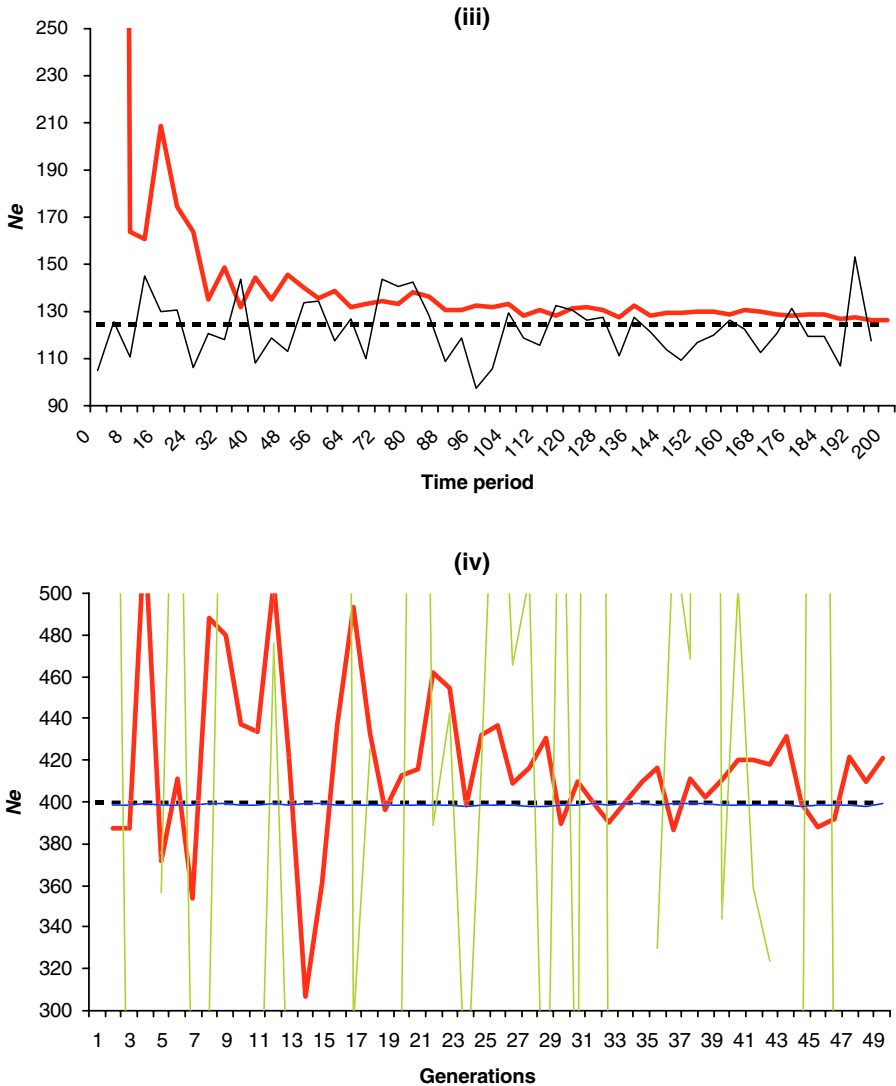
Figure 2. Continued.

(plot iii), $\overline{N}_e$ had a better performance than that based on variance of family size, which resulted in a more variable N_e. In the non-Poisson case in which all individuals have an offspring size of two (plot iv), all the methodologies involved tended to give the correct value. Regarding dispersion, the estimates of N_e based on individual increase in inbreeding were intermediate between those from the rate of inbreeding (ΔF) and from the rate of coancestry (Δf). Obviously the method based on variance of family size was, in this case, exact.
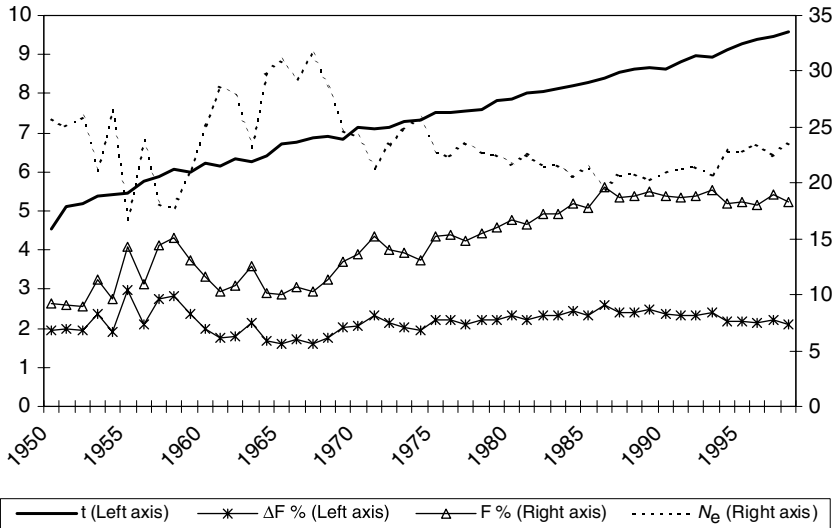
Figure 3. Evolution of the mean inbreeding, mean individual increase in inbreeding, mean equivalent generations, and $\overline{N}_e$ across year of birth.

Table I gives the main parameters describing the real pedigree file. The pedigree size was 6318 and the size for the fitted RS was 1464 individuals for RS_{77-89} and 1721 for RS_{10}. The mean equivalent generation ± standard deviation was the following: 6.6 ± 2.75 for the whole pedigree (WP), 9.1 ± 0.68 for RS_{10} and 8.2 ± 0.64 for RS_{77-89}.

Figure 3 shows the evolution of the mean inbreeding, mean individual increase in inbreeding, mean equivalent generations, and $\overline{N}_e$ across years of birth for the whole Carthusian pedigree file. It can be noted that mean inbreeding became approximately stable in the last generation interval, whilst mean equivalent generations increased leading to a reduction in the mean ΔF_i during this period. The flat or slightly negative trend of inbreeding coefficients would lead to illogical estimates of N_e when using methods based on regression of inbreeding on either generations or year of birth. However, N_e obtained from individual increase in inbreeding remained approximately stable since pedigree knowledge achieves about five equivalent generations.

Table II gives the estimates of N_e obtained using regression of the individual coefficients of inbreeding on equivalent generations, variance of family sizes and individual increase in inbreeding (ΔF_i) in the whole pedigree file and the defined RS of the Carthusian horse. The estimates obtained using variances in family sizes are quite larger than those obtained using both regression on equivalent generations and ΔF_i.

Table II. Estimates of N_e using the whole pedigree file (WP) and RSs fitted using the individuals born in the last 10 years (RS_{10}) or in a given period of years (RS_{77-89}) obtained using regression of the individual coefficients of inbreeding on equivalent generations [$b(t)$], variance of family sizes (V) and individual increase in inbreeding ($\overline{N_e}$).

Population	No pedigree depth restriction			$t \geq 4$			$t \geq 8$		
	N_e [$b(t)$]	N_e (V)	$\overline{N_e} \pm$ s.d.	N_e [$b(t)$]	$\overline{N_e} \pm$ s.d.	N	N_e [$b(t)$]	$\overline{N_e} \pm$ s.d.	N
WP	19.5	35.9	26.4 ± 0.26	20.4	22.4 ± 0.17	5172	273.5	21.4 ± 0.14	2518
RS_{10}	33.4		22.2 ± 0.18	40.0	22.2 ± 0.18	1719	-142.2	22.0 ± 0.18	1671
RS_{77-89}	12.2	88.1	21.4 ± 0.21	11.6	21.3 ± 0.21	1463	21.9	20.3 ± 0.22	924

Analyses were carried out without pedigree depth restrictions, and using only individuals with four equivalent generations (t) or more and eight equivalent generations (t) or more.

RS_{10}: Animals born in the last decade.

RS_{77-89}: Animals born between the years 1977 and 1989.

Table III. Effective population size computed in four simulated examples and a real population for the whole pedigree file and the last 400 individuals born (simulated examples) or the Carthusian reference subsets (RSs, see text).

Pedigrees	Theoretical	N	Whole pedigree file			Reference subsets					
			$b(t)$	b(date)	log b(date)	Var(offs)	$b(t)$	b(date)	log b(date)	$\overline{N}_e$	s.d.($\overline{N}_e$)
Simulated											
(i) Random – two sexes	200.5	400	214.0	202.3	201.7	216.1	163.7	162.9	163.0	204.9	1.8466
(ii) Subdivision	200.5/50.5	400	94.6	86.1	83.1	196.6	89.9	89.8	93.9	83.3	0.5917
(iii) Overlapping	125	400	145.6	126.0	125.1	130.7	129.7	142.7	148.7	126.0	0.7420
(iv) Non-Poisson	400	400	412.6	400.8	400.2	400.0	−11 671.6	−11 308.7	−3749.1	401.6	6.4152
Carthusian											
Whole pedigree	–	6318	–	–	–	77.2	19.5	19.3	19.4	26.4	0.2563
RS$_{79-89}$	–	1464	–	–	–	–	12.2	11.7	11.2	21.4	0.2067
RS$_{10}$	–	1721	–	–	–	–	33.4	−49.8	−46.4	22.2	0.1817

Effective size obtained $b(t)$ – *via* regression of inbreeding over equivalent generations, b(date) – *via* regression on birth date, log b(date) – *via* log regression on birth date, Var(offs) – *via* variance of family size, and $\overline{N}_e$ – *via* individual increase in inbreeding. s.d.($\overline{N}_e$) is the standard error of $\overline{N}_e$.

The values for $\overline{N}_e$ were stable regardless of any restrictions on the pedigree depth of the individuals included in the pedigree file or the corresponding RS, whilst the N_e obtained from regression of F on equivalent generations showed a non-consistent behavior. It is possible to find a noticeable increase in N_e (40.0) when RS_{10} includes individuals with four or more equivalent generations in the pedigree file but also negative values of N_e (-142.2) when RS_{10} includes individuals with eight or more equivalent generations in the pedigree file. Note again that a negative estimation of N_e can be obtained when the increase in inbreeding is obtained by regression of the inbreeding coefficient on the year of birth and younger individuals are less inbred than older individuals. Thus, the N_e obtained depends partially, on the effect of the changes in the mating policy. In the Carthusian population, the criterion of minimal coancestry has recently been used to define the mating policy of this population [22]. However, the N_e obtained by individual increase in inbreeding shows a stable value of about 22.

The overall behavior of $\overline{N}_e$ is summarized in Table III. In this table a RS as the last 400 individuals born was defined, which involves the last two generations in the simulated examples under discrete generations. The real whole pedigree file and the RS defined in the Carthusian pedigree file were also included together with the simulated examples. Different approaches to obtain N_e were gathered: $b(t)$ as described in (4), $b(date)$ as described in Gutiérrez *et al.* [12], log $b(date)$ as described in Pérez-Enciso [20], Var(offs) as described in (3), and $\overline{N}_e$ based on the individual increase in inbreeding together with its standard error. The theoretical N_e is also addressed in the table when possible. The method using the variances of family size does not result in accurate N_e estimates in examples reflecting scenarios closer to reality (such as population subdivision) because they only consider one source of variation of N_e. The N_e computed from regression approaches (regardless of whether it is done on generations or birth date) tend to perform poorly when RSs are defined in a pedigree file, particularly in extreme cases in which the subset of interest has lower mean F than the other individuals included in the pedigree file. Interestingly enough, $\overline{N}_e$ is quite stable and precise regardless of the pedigree file (real or simulated) analyzed and the particular conditions of the whole and the RSs defined.

4. DISCUSSION

The simple methodology presented here to assess N_e in real populations accounts for pedigree knowledge of each individual in a population in order to obtain individual increase in inbreeding values (ΔF_i). The increase in inbreeding is not treated here as a single value but as a variable with an associated

mean ($\overline{\Delta F}$) that can easily be used to compute N_e (in fact $\overline{N_e}$) for a given RS as $\overline{N_e} = \frac{1}{2\overline{\Delta F}}$. The current methodology addresses N_e directly from ΔF which, theoretically, becomes constant in a population with stable size and breeding policy. This is contrary to F, which increases from one generation to another [7]. Thus, ΔF is completely independent of pedigree depth in the idealized population [7]. The use of ΔF_is additionally contributes to overcome the problem of using the F_i coefficients because the latter are non-linearly dependent on the pedigree depth of each individual. When trying to assess ΔF, after averaging F_i coefficients by generation, differences in absolute mean values from one generation to another must still be divided by one minus the mean inbreeding in the previous generation. This is not easy to carry out in real populations which usually have overlapping. In such a scenario, it is unrealistic to work under assumptions such as no inbreeding in previous generations, or linear trend of inbreeding by generations. However, individual increase in inbreeding is "free" from these effects since it is also adjusted for the generation number of each animal. Of course, ΔF_is are still dependent on the completeness of the analyzed pedigree file and need a few generations to become constant at the population level (Fig. 3). Notice also that in real-diploid populations in which self-fertilization is not possible and where there are two different sexes, ΔF (but also $\overline{\Delta F}$), will need more generations to reach asymptotic values than those expected in the idealized population. In this respect, the simple plot of ΔF_i over equivalent generations helps with the correct definition of the RSs and on its usefulness to obtain reliable estimates of N_e (Fig. 4).

Of course, as with other methodologies, the $\overline{N_e}$ obtained is still dependent on the way the RS is defined. However, the present approach can be used to assess the confidence interval of $\overline{N_e}$ from the standard error of $\overline{\Delta F}$. Indeed, no standard error can be computed directly for $\overline{N_e}$ because the presence in our RS of individuals with $\Delta F_i = 0$ would lead to individual estimates of $N_e = \infty$. Following the methodology presented here, values of $\Delta F_i = 0$ would affect $\overline{N_e}$ but an estimation of the parameter should still be possible.

When compared with the other methods to obtain N_e assayed here, the N_e obtained from $\overline{\Delta F}$ was more stable regardless of the particular situation of the analyzed pedigree files. This was shown to be true regardless of whether we included individuals with shallower or deeper pedigree in the RSs (Tab. II). In this respect, the use of the regression coefficient of the individual inbreeding values over the individual equivalent complete generations to compute N_e, gave estimates that were highly dependent on the mating policy carried out in the RS. This last concern was particularly noticeable when, after having mated relatives, a plan of mating based on low coancestry was implemented, leading to a decrease in the mean inbreeding thus providing negative N_e values.
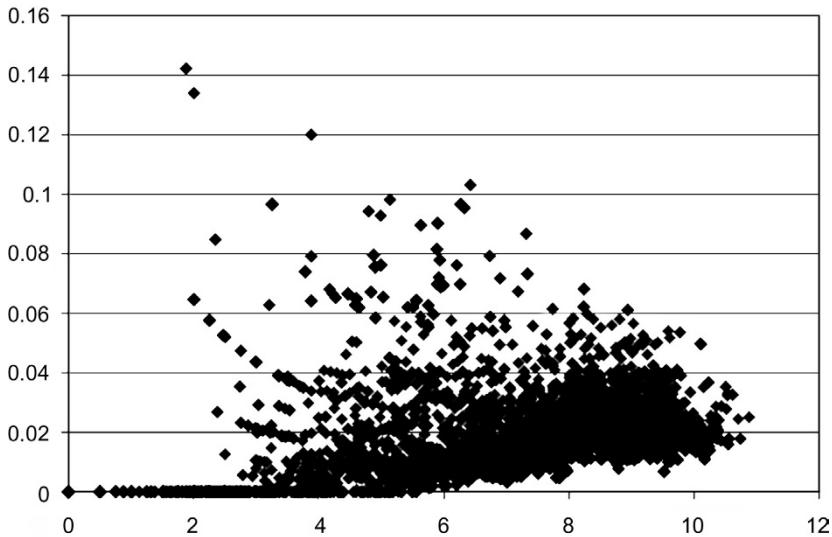
Figure 4. Plot summarizing the dispersal of the individual increase in inbreeding (ΔF_i; on the Y-axis) per individual number of equivalent generations (on the X-axis) in the whole Carthusian pedigree file.

As expected, the N_e computed from variances of family sizes was not useful to characterize the 'real' effective size, as shown in the Carthusian population analyzed here. This was also accurately reflected by the results obtained from the simulated example (ii) involving subdivision. If population subdivision is known, one would not treat a pedigree file as a single population, but each sub-population separately. Then, the number of breeding animals would be smaller than when ignoring subdivision. The simulated example is only an unrealistic example of extreme subdivision but almost all real populations have some degree of uncontrolled subdivision, reflected by the fact that inbreeding is usually higher than coancestry. However, the method based on variance of family size reflects a temporary mating policy and can be useful when the pedigree knowledge is limited and/or subdivision has not yet occurred.

The methodology shown here does not consider phenomena such as population subdivision and selection. However, since these phenomena are reflected in the pedigree of individuals belonging to the analysed RS, the effective size computed from ΔF_is is expected to capture these effects. The concept of effective size usually has an asymptotic meaning in a regular system and can be used for predictive purposes rather than for analyzing genealogies [12]. The effective size obtained here is also related to pedigree tools used for describing genetic diversity in real populations, such as (effective) number of founders

or ancestors [3,15]. The parameters computed here ($\overline{\Delta F}$ and $\overline{N_e}$) are therefore related to those computed using the concept of long-term genetic contributions of ancestors [25,26] even though the current approach does not need complex or iterative computations. The relationship between inbreeding – actually the partial inbreeding coefficients of Lacy *et al.* [16] – and founder contributions were used by Man *et al.* [19] to predict the frequency of carriers of an autosomal-deleterious gene when the ancestral source(s) of the gene is known. There is an inverse relationship between partial inbreeding coefficients coming from founders or ancestors and the effective number of founders or ancestors. In fact, under a random mating scheme, these partial inbreeding coefficients should be the sum of their squared contribution, which is precisely the denominator in the computation of the effective number of founders or ancestors. The relationship between F_is and ΔF_is is straightforward and the inverse relationship between increase in inbreeding and N_e is given by definition.

However, the unbalance of the contributions from founders or ancestors is not the sole cause of the increases in inbreeding. They are also dependent on many other circumstances such as population structure, mating policy, changes in population size, *etc*. The estimates of N_e based on individual increase in inbreeding would accurately reflect the genetic history of the populations, namely the size of their founder population, their mating policy or bottlenecks due to abusive use of reproductive individuals. All these phenomena influence the pedigree of the individual and are therefore reflected in the individual increase in inbreeding.

Inbreeding coefficients are widely used to calculate the rate of inbreeding and consequently N_e [7]. However, N_e can also be computed from the average coancestry of a RS [4]. In regular non-structured populations, average coancestry and inbreeding coefficients are analogous and the N_e obtained using either approach should be the same. However, it has been reported that this is not always the case [23] and our results from the simulated population (ii) would confirm this fact. The effective size is defined here from the individual rate of inbreeding. However, the method can be easily extended to apply the same methodology to coancestry coefficients. To do this, all the coancestries among individuals in the RS would be assigned to the inbreeding coefficient of a hypothetical offspring from each couple of individuals with an equivalent discrete generation number of half the sum of the parents plus one, by simply applying (3).

To conclude, we show here a simple new approach to estimate N_e in real populations that give stable estimates for N_e. This methodology is derived from the definition of the individual increase in inbreeding. It is treated as a variable, which can be computed for each individual and can be useful for other purposes. Moreover, the possibility of estimating N_e using this approach, and after assuming that it is the true N_e value, provides a chance to estimate other useful

parameters by just making use of the multiple existent expressions to predict N_e in different situations [4,7], as for example, rates of selection, migration and Hardy-Weinberg deviations. Although this parameter is still a novelty, its usefulness makes it worthwhile to be investigated further.

ACKNOWLEDGEMENTS

REFERENCES

[1] Álvarez I., Royo L.J., Gutiérrez J.P., Fernández I., Arranz J.J., Goyache F., Relationship between genealogical and microsatellite information characterising losses of genetic variability: empirical evidence from the rare Xalda sheep breed, Livest. Sci. 115 (2008) 80–88.

[2] Baumung R., Sölkner J., Pedigree and marker information requirements to monitor genetic variability, Genet. Sel. Evol. 35 (2003) 369–383.

[3] Boichard D., Maignel L., Verrier E., The value of using probabilities of gene origin to measure genetic variability in a population, Genet. Sel. Evol. 29 (1997) 5–23.

[4] Caballero A., Developments in the prediction of effective population size, Heredity 73 (1994) 657–679.

[5] Caballero A., Toro M.A., Interrelations between effective population size and other pedigree tools for the management of conserved populations, Genet. Res. 75 (2000) 331–343.

[6] Duchev Z., Distl O., Groeneveld E., Early warning system for loss of diversity in European livestock breeds, Archiv. Anim. Breed. 49 (2006) 521–531.

[7] Falconer D.S., Mackay F.C., Introduction to Quantitative Genetics, 4th edn., Longman Group Ltd, England, 1996.

[8] FAO, Secondary Guidelines for the National Farm Animal Genetic Resources Management Plans: Management of Small Populations at Risk, FAO, Rome, Italy, 1998.

[9] González-Recio O., López de Maturana E., Gutiérrez J.P., Inbreeding depression on female fertility and calving ease in Spanish dairy cattle, J. Dairy Sci. 90 (2007) 5744–5752.

[10] Goyache F., Gutiérrez J.P., Fernández I., Gómez E., Álvarez I., Díez J., Royo L.J., Monitoring pedigree information to conserve the genetic variability in

endangered populations: the Xalda sheep breed of Asturias as an example, J. Anim. Breed. Genet. 120 (2003) 95–103.

[11] Gutiérrez J.P., Goyache F., A note on ENDOG: a computer program for analysing pedigree information, J. Anim. Breed. Genet 122 (2005) 357–360.

[12] Gutiérrez J.P., Altarriba J., Díaz C., Quintanilla R., Cañón J., Piedrafita J., Pedigree analysis of eight Spanish beef cattle breeds, Genet. Sel. Evol. 35 (2003) 43–64.

[13] Hill W.G., A note on effective population size with overlapping generations, Genetics 92 (1979) 317–322.

[14] James J.W., A note on selection differentials and generation length when generations overlap, Anim. Prod. 24 (1977) 109–112.

[15] Lacy R.C., Analysis of founder representation in pedigrees: founder equivalent and founder genome equivalents, Zoo Biol. 8 (1989) 111–123.

[16] Lacy R.C., Alaks G., Walsh A., Hierarchical analysis of inbreeding depression in *Peromyscus polionotus*, Evolution 50 (1996) 2187–2200.

[17] Kempen G.M.P., van Vliet L.J., Mean and variance of ratio estimators used in fluorescence ratio imaging, Cytometry 39 (2000) 300–305.

[18] Maignel L., Boichard D., Verrier E., Genetic variability of French dairy breeds estimated from pedigree information, Interbull Bull. 14 (1996) 49–54.

[19] Man W.Y.N., Nicholas F.W., James J.W., A pedigree-analysis approach to the descriptive epidemiology of autosomal-recessive disorders, Prevent. Vet. Med. 78 (2007) 262–273.

[20] Pérez-Enciso M., Use of the uncertain relationship matrix to compute effective population size, J. Anim. Breed. Genet. 112 (1995) 327–332.

[21] Royo L.J., Álvarez I., Gutiérrez J.P., Fernández I., Goyache F., Genetic variability in the endangered Asturcón pony assessed using genealogical and molecular information, Livest. Sci. 107 (2007) 162–169.

[22] Valera M., Molina A., Gutiérrez J.P., Gómez J., Goyache F., Pedigree analysis in the Andalusian horse: population structure, genetic variability and influence of the Carthusian strain, Livest. Prod. Sci. 95 (2005) 57–66.

[23] Wang J., Effective size and F-statistics of subdivided populations. II. Dioecious species, Genetics 146 (1997) 1465–1474.

[24] Woolliams J.A., Mäntysaari E.A., Genetic contributions of Finnish Ayrshire bulls over four generations, Anim. Sci. 61 (1995) 177–187.

[25] Woolliams J.A., Wray N.R., Thompson R., Prediction of long-term contributions and inbreeding in populations undergoing mass selection, Genet. Res. 62 (1993) 231–242.

[26] Wray N.R., Thompson R., Prediction of rates of inbreeding in selected populations, Genet. Res. 55 (1990) 41–54.

[27] Wright S., Evolution and the Genetics of Populations: The Theory of Gene Frequencies, Vol. II, University of Chicago Press, Chicago, USA, 1969.