

Assessing performance of single-sample molecular genetic methods to estimate effective population size: empirical evidence from the endangered *Gochu Asturcelta* pig breed

Juan Menéndez¹, Isabel Álvarez², Iván Fernández², Nuria A. Menéndez-Arias² & Félix Goyache²

¹ACGA, C/Párroco José Fernández Teral nº 5 A, Avilés, Asturias 33403, Spain

²SERIDA-Deva, Camino de Rioseco 1225, Gijón, Asturias 33394, Spain

Keywords

Coancestry, effective population size, linkage disequilibrium, pedigree, single sample, temporal sampling.

Correspondence

Félix Goyache, Área de Genética y Reproducción Animal, SERIDA-Deva, Camino de Rioseco, 1225, E-33394-Gijón, Asturias, Spain.

Tel: +34 984502010;

Fax: +34 984502012;

E-mail: fgoyache@serida.org

Funding Information

This work was partially funded by a specific contract between Government of Principado de Asturias and Asociación de Criadores de Gochu Asturcelta (ACGA; <http://www.gochuasturcelta.org/>). IA, IF, and FG are supported by grant FICYT GRUPIN14-113.

Received: 30 March 2016; Revised: 11 May 2016; Accepted: 16 May 2016

Ecology and Evolution 2016; 6(14): 4971–4980

doi: 10.1002/ece3.2240

Introduction

There is an increasing interest in estimating effective population size (N_e) using linkage disequilibrium (LD) information. Although these methodologies have been used basically in natural populations (Waples 1991; Barker 2011), they are of interest in livestock populations with shallow pedigrees in which no sound estimates of effective population size can be obtained using genealogies (Cervantes et al. 2011b).

The advantages of using LD information are clear in terms of time and operational costs: a single sample can

Abstract

Estimating effective population size (N_e) using linkage disequilibrium (LD) information ($N_{e(LD)}$) has the operational advantage of using a single sample. However, $N_{e(LD)}$ estimates assume discrete generations and its performance are constrained by demographic issues. However, such concerns have received little empirical attention so far. The pedigree of the endangered Gochu Asturcelta pig breed includes individuals classified into discrete filial generations and individuals with generations overlap. Up to 780 individuals were typed with a set of 17 microsatellites. Performance of $N_{e(LD)}$ was compared with N_e estimates obtained using genealogical information, molecular coancestry ($N_{e(M)}$) and a temporal (two-sample) method ($N_{e(JR)}$). Molecular-based estimates of N_e exceeded those obtained using pedigree data. Estimates of $N_{e(LD)}$ for filial generations F3 and F4 (17.0 and 17.3, respectively) were lower and steadier than those obtained using yearly or biannual samplings. $N_{e(LD)}$ estimated for samples including generations overlap could only be compared with those obtained for the discrete filial generations when sampling span approached a generation interval and demographic correction for bias was applied. Single-sample $N_{e(M)}$ estimates were lower than their $N_{e(LD)}$ counterparts. $N_{e(M)}$ estimates are likely to partially reflect the number of founders rather than population size. In any case, estimates of LD and molecular coancestry tend to covary and, therefore, $N_{e(M)}$ and $N_{e(LD)}$ can hardly be considered independent. Demographically adjusted estimates of $N_{e(JR)}$ and $N_{e(LD)}$ took comparable values when: (1) the two samples used for the former were separated by one equivalent to discrete generations in the pedigree and (2) sampling span used for the latter approached a generation interval. Overall, the empirical evidence given in this study suggested that the advantage of using single-sample methods to obtain molecular-based estimates of N_e is not clear in operational terms. Estimates of N_e obtained using methods based in molecular information should be interpreted with caution.

provide estimates of, probably, the most important evolutionary parameter for a given population. However, computation of effective population size using linkage disequilibrium ($N_{e(LD)}$) has well-known causes of severe bias, namely sample size, markers set size, and minor allele frequency (England et al. 2006; Luikart et al. 2010; Waples and Do 2010).

In any case, the major operational constraint for the estimation of $N_{e(LD)}$ is that this approach assumes discrete generations and only fit well to semelparous age-structured species. In iteroparous species, such as livestock, in

which generations overlapping is the rule, $N_{e(LD)}$ values are more likely estimates of the effective number of breeders (N_b ; effective number of adult individuals that give rise to a cohort) rather than the effective size for a generation (Waples 2006; Barker 2011; Goyache et al. 2011). Although N_e and N_b are closely related, there are large differences between them among species and within populations (Waples et al. 2013, 2014). In such scenario, the analysis of livestock populations with known mating policy, demographic structure, and pedigrees can shed light on the performance of $N_{e(LD)}$ in natural iteroparous populations.

The demography of a population evolving under generations overlap is likely to affect molecular-based estimates of N_e . When temporal (two-sample) methods for computing N_e are applied, it is necessary to assume that samples are far from being independent and that “temporal” estimates of N_e must be adjusted using life-traits data (Jorde and Ryman 1995, 1996). A similar rationale has been recently applied to estimates of $N_{e(LD)}$ obtained using single-cohort samples. Waples et al. (2014) suggested to adjust estimates of $N_{e(LD)}$ for demographic bias using the ratio N_b/N_e . This ratio can be calculated accurately using two key life-history traits (Waples et al. 2011, 2013).

The demographic concerns described above are not usually addressed even in research carried out in livestock (Corbin et al. 2010, 2012; Flury et al. 2010; Goyache et al. 2011). The endangered Gochu Asturcelta pig breed (Menéndez et al. 2016a) offers a unique scenario to deal with this task. A recovery program for the breed started in 2002 using six founders (three boars and three sows). The reproductive career of the founders and their direct descendants was prolonged as much as possible, and strict breeding policies avoiding matings between close relatives were applied in the population (Menéndez et al. 2016a). This allowed to identify, across years, a number of individuals which could be classified into discrete filial generations: F₁ (direct descendants of two founders), discrete generation F₂ (F₁ × F₁ crosses), and so on till discrete generation F₅ (F₄ × F₄ crosses). This unique scenario allows to compare, in the same population, estimates of $N_{e(LD)}$ obtained when discrete generations are considered with those obtained using yearly or biannual cohort samplings. Further, the effect of the correction for demographic bias using parameters obtained via direct observation of the pedigree can also be assessed.

The current research will analyze both the information registered in the herdbook of the Gochu Asturcelta pig breed from 2006 to 2010 and the genotypes obtained for paternity testing. This will allow to assess the performance of $N_{e(LD)}$ in the following scenarios: (1) samples obtained from discrete filial generations; (2) samples obtained from

yearly cohorts; and (3) samples drawn from a number of yearly cohorts equaling to or exceeding generation length. The effect of demographic adjustment of estimates will be assessed as well. For descriptive purposes, performance will be compared with estimates of N_e obtained using single-sample molecular coancestry, temporal (two-sample) methods, and genealogical information.

Materials and Methods

Data available and sampling

Pedigree data recently analyzed by Menéndez et al. (2016a) were available. Data consisted of 3156 records (including six founders), from 515 litters, with father and mother known, registered in the herdbook of the breeders association (ACGA) from its foundation to August 2014. A total of 109 boars and 309 sows had offspring in data. Genealogies were traced to identify 11 F₁ individuals (offspring of two founders), 47 F₂ individuals (F₁ × F₁), 216 F₃ individuals (F₂ × F₂), 147 F₄ individuals (F₃ × F₃), and seven F₅ individuals (F₄ × F₄).

Table 1 gives a detailed description of the data used. Analyses were limited to the period in which F₃ and F₄ individuals were born (from 2006 to 2010). This ensured that sample size and pedigree depth (at least three equivalents to complete generations; Gutiérrez et al. 2009) were enough to obtain reliable results. Finally, pedigree analyses involved a total of 2248 individuals, born between 2006 and 2010, including 363 F₃ or F₄ individuals and 1885 individuals with different pedigree depths due to generations overlap.

A set of 17 microsatellites (IGF1, S0002, S0026, S0071, S0101, S0155, S0225, S0226, S0227, SW240, SW632, SW857, SW911, SW936, SW951, S0005, and S0090) used in paternity testing and diversity analyses (Menéndez et al. 2015, 2016b) were typed in a representative sample of the available individuals. Most microsatellites used were included in the ISAG–FAO panel (<http://www-lgc.toulouse.inra.fr/pig/panel/panel2004.htm>). Primer sequences and Polymerase Chain Reaction (PCR) conditions can be found in Menéndez et al. (2016b). PCR was carried out in a GenAmp 9700 thermocycler (Applied Biosystems, Barcelona, Spain) and genotyping was performed on an ABI 3130 DNA-automated sequencer (Applied Biosystems).

Genotypes of a total of 780 individuals were available. They included: (1) 324 of 363 (89%) F₃ or F₄ individuals and (2) 456 of 1885 (24%) individuals with generations overlap in their pedigree born between 2006 and 2010. Altogether, yearly samples available varied from 83 (74-typed) individuals born in 2006 to 724 (225-typed) individuals born in 2010 (see Table 1).

Table 1. Description of samples used per year of birth. The number of litters and individuals involved in computations are detailed according to pedigree knowledge: (a) those individuals included into discrete generations (*F₃* or *F₄*) and (b) those having overlapped genealogies. Both the number of individuals registered in the herdbook (used for genealogical analyses) and the number of individuals typed (in brackets) are given.

Year of birth	Discrete generations				Overlapped generations		Totals	
	Litters	<i>F₃</i>	<i>F₄</i>	Subtotal	Litters	Individuals	Litters	Individuals
2006	6	39 (32)	0 (0)	39 (32)	5	44 (42)	11	83 (74)
2007	14	99 (98)	6 (6)	105 (104)	18	130 (32)	32	235 (136)
2008	21	50 (34)	109 (95)	159 (129)	51	404 (82)	72	563 (211)
2009	5	22 (21)	26 (26)	48 (47)	85	676 (178)	90	724 (225)
2010	3	6 (6)	6 (6)	12 (12)	79	631 (122)	82	643 (134)
Totals	49	216 (191)	147 (133)	363 (324)	238	1885 (456)	287	2248 (780)

According to the structure of data described above, analyses were sequentially carried out on: (1) discrete filial generations (*F₃* and *F₄*); (2) yearly cohorts from 2006 to 2010; and (3) sequential biannual samplings mimicking the average generation interval of 1.8 ± 0.03 years reported for the Gochu Astucelta breed by Menéndez et al. (2016a). As mating policy avoids crosses between close relatives (Menéndez et al. 2016a), a model of random mating was assumed when necessary.

Genealogical estimates of effective population size

The equivalent to complete generations traced (*t*), computed as the sum of $(1/2)^n$, where *n* is the number of generations separating the individual to each known ancestor (Maignel et al. 1996), was calculated for each individual in the pedigree born in the five-year period 2006–2010.

Effective population sizes (*N_e*) and their standard errors were estimated on the basis of individual increase in inbreeding ΔF_i (Gutiérrez et al. 2008, 2009) and coancestry ΔC_{ij} (Cervantes et al. 2011a) considering $\Delta F_i = 1 - \sqrt[t_i]{1 - F_i}$ and $\Delta C_{ij} = 1 - \sqrt[(t_i+t_j)/2]{1 - C_{ij}}$, where *F_i* is the inbreeding coefficient of individual *i*, *C_{ij}* is the coancestry coefficient between individuals *i* and *j* (the inbreeding of a descendent from both), and *t_i* and *t_j* are their respective equivalent to complete generations. Finally, effective sizes were computed by averaging the individual increase in inbreeding and the increase in pairwise coancestry for all pairs of individuals in a reference subpopulation using the following formulae: $N_e F_i = 1/(2\Delta F_i)$ and $N_e C_{ij} = 1/(2\Delta C_{ij})$. Finally, following Cervantes et al. (2011a), the ratio $N_e C_{ij}/N_e F_i$ was computed to ascertain the existence of a possible hidden structure in data.

Single-sample molecular estimates of effective population size

$N_{e(LD)}$ was estimated as $N_{e(LD)} = \frac{1}{3(\hat{r}^2 - 1/S)}$, where \hat{r}^2 is the estimate of the correlation among alleles and *S* is

the sample size, using the modification proposed by Waples (2006), which correct for biases resulting from the presence of rare alleles, and was empirically adapted to different sample sizes and mating systems (here large sample sizes, ≥ 30 , and random mating apply; see Waples and Do 2010). To check for the consistency of the results obtained, three separate analyses were performed via removing, respectively, alleles with frequencies (*P_{crit}*) lower than 0.05, 0.02, and 0.01. A jackknife procedure was used to construct 95% confidence intervals of the estimates.

For consistency with the genealogical methods, single-sample *N_e* was also estimated using the molecular coancestry method proposed by Nomura (2008) as $\hat{N}_{e(M)} = \frac{1}{2\hat{f}_i}$, where \hat{f}_i is the average over $n(n-1)/2$ pairs of individuals of the molecular coancestry between two individuals *i* and *j* over *L*-analyzed loci $\hat{f}_{ij,l} = \frac{1}{W} \sum_{l=1}^L w_l \frac{f_{ij,l} - \hat{s}_l}{1 - \hat{s}_l}$, where \hat{s}_l is the expected homozygosity at a locus *l*. Nomura (2008) followed the suggestion by Oliehoek et al. (2006) of: (1) removing from the computations those alleles alike-by-state and not identical-by-descent and (2) weighting the contributions over loci using $W = \sum_{l=1}^L w_l = \sum_{l=1}^L \frac{(1 - \hat{s}_l)^2}{\sum_{i=1}^{n_i} \hat{p}_i^2 (1 - \sum_{i=1}^{n_i} \hat{p}_i^2)}$, where *p_i* is the frequency of the allele *i* at a given locus, to increase the importance of loci with small *s_l* and balanced allele frequencies. This method uses alleles at any frequency for computations. A jackknife procedure was used to construct 95% confidence intervals of the estimates.

Two-sample molecular estimates of effective population size

To illustrate differences between single-sample and two-sample estimators of molecular-based *N_e*, the unbiased temporal method proposed by Jorde and Ryman (2007), which has been proved to give consistent estimates across cohort pairs in a livestock framework (Goyache et al. 2011), was performed as well. This

method is based on the computation of the estimator F_s , computed as

$$F_s = \frac{\sum_{i=1}^A (x_i - y_i)^2}{\sum_{i=1}^A z_i(1 - z_i)^2}$$

, where A is the number of alleles at the locus, x_i and y_i are the frequencies of the i th allele in the first and second samples, respectively, and z_i is the average frequency of the i th allele over samples. Computations were performed under a sampling plan I (Waples 1989) in which individuals are sampled nondestructively and subsequently returned to the population. Under this sampling plan, the Jorde and Ryman's (2007) estimator of N_e (here noted as $N_{e(JR)}$) is $N_{e(JR)} = \frac{F_s[1-1/(4\tilde{n})]-1/\tilde{n}+1/N}{(1+F_s/4)[1-1/(2n_y)]}$, where n_y is the number of individuals in the second sample, \tilde{n} is the harmonic mean of the sample sizes n_x and n_y , and N is the actual census size of the population at the time of first sampling. A jackknife procedure was used to construct 95% confidence intervals of the estimates.

Demographic adjustment for bias

Following Waples et al. (2014), $N_{e(LD)}$ estimates were corrected dividing them with the ratio N_b/N_e , where N_b and N_e are the effective number of breeders and effective population size, respectively. Ratio N_b/N_e was estimated from demographic information using a discrete-time, age-structured, and deterministic model and using age-specific survival rates (s_x) and birth rates (b_x) calculated separately for males and females (Table S1). The model assumes that: (1) reproduction occurs at intervals of exactly one time unit (here one year); (2) survival and fecundity are independent of events in previous time periods; (3) there is no upper bound to the number of offspring an individual can produce in one breeding cycle; and (4) individuals survive to their first birthday and, therefore, fecundities are scaled to result in a stable population that produces a fixed number (N_1) of individuals per cohort that survive to age 1.

Following Jorde and Ryman (1995, 1996), $N_{e(JR)}$ estimates were corrected multiplying them with the ratio C/G , where C is a correction factor obtained from life table data (see Table S2) and G is the generation interval. Factor C accounts for variance due to mortality as a cohort passes from one-year class to the next and for genetic covariance among cohorts (because individuals from multiple age classes are the parents of a given cohort). The model to compute factor C requires a basic life table with information on age-specific survival rates (l_i) and birth rates at each age class i (i.e., gametic contribution; b_i ; see Table S2).

In all cases, life table data were estimated directly from the Gochu Asturcelta pig pedigree limiting the age of the parents to 5 years old.

Software used

All demographic and genealogical analyses were computed using the program ENDOG v4.8 (Gutiérrez and Goyache 2005) freely available at http://www.ucm.es/info/prodanim/html/JP_Web.htm

Molecular-based estimates of N_e were computed in all cases using the program NeEstimator (Do et al. 2014) freely available at <http://www.molecularfisherieslaboratory.com.au/neestimator-software/>.

Ratio N_b/N_e was computed from life table data using the program AgeNe (Waples et al. 2011, 2013) freely available at <http://conserver.iugo-cafe.org/user/Robin%20Waples/AgeNe>.

The Jorde and Ryman's (1995, 1996) correction factor C was computed using a program kindly provided by Dr P. E. Jorde (<http://folk.uio.no/ejorde/software/factorc.zip>).

Results

Table 2 gives the estimates of N_e obtained using linkage disequilibrium ($N_{e(LD)}$), molecular coancestry ($N_{e(M)}$), and pedigree information. When discrete filial generations were considered, genealogical estimates of N_e were consistently the same varying from $N_e F_i = 5.0 \pm 0.8$ for F_3 to 5.6 ± 0.3 F_4 . Estimates of $N_e F_i$ and $N_e C_{ij}$ were comparable across yearly and biannual samplings with the lower estimates for the cohort sampled in 2006. Parameter $\overline{\Delta F_i}$ tended to have similar values across either yearly or biannual samplings. However, both $N_e F_i$ and $N_e C_{ij}$ tended to increase with pedigree depth (and size of the breeding stock) varying from $N_e F_i = 4.6 \pm 1.9$ for Cohort₂₀₀₆ to $N_e C_{ij} = 9.2 \pm 0.3$ for Cohort₂₀₁₀. Further, ratio $N_e C_{ij}/N_e F_i$ was roughly 1 for F_3 and F_4 . However, this ratio increased with years from 1.09 for Cohort₂₀₀₆ to 1.39 for Cohort₂₀₁₀, therefore suggesting the existence of a slight hidden structure in the Gochu Asturcelta pedigree (Table 2).

$N_{e(LD)}$ took values over 17.0 for both discrete filial generations F_3 and F_4 (Table 2). Estimates of $N_{e(LD)}$ obtained for yearly or biannual samplings were adjusted for generations overlap using the ratio N_b/N_e computed using demographic information. This ratio took a value of 0.667 corresponding to demographic estimates of N_b and N_e of 222.9 and 334.9, respectively. Estimates of $N_{e(LD)}$ were highly consistent no matter the P_{crit} used. Therefore, only estimates obtained using $P_{crit} = 0.05$ are given. When yearly samplings were considered, the corrected estimates were similar to those obtained for the discrete filial

Table 2. Number of individuals (N) involved and estimates of effective size for each discrete generation, yearly cohort, and biannual sampling analyzed in the Gochu Asturcelta pig breed population computed via molecular-based methods (linkage disequilibrium, $N_{e(LD)}$) and molecular coancestry, $N_{e(M)}$ and pedigree information (individual increase in inbreeding, $N_e F_i$ and individual increase in coancestry, $N_e C_{ij}$). In brackets, confidence intervals of the estimates on 95% (molecular-based methods) or standard errors of the estimates (genealogical methods) are provided. Additionally, the estimated correlation (r^2) and molecular coancestry (\hat{f}) among alleles are given for the molecular-based methods and mean inbreeding (F) and mean equivalent to discrete generations (t), and average individual increase in inbreeding ($\Delta \bar{F}_i$) are provided for pedigree data.

Sampling	Molecular estimates					Genealogical estimates						
	N^2	r^2	$N_{e(LD)}^3$	\hat{f}	$N_{e(M)}$	N^2	F	t	$\Delta \bar{F}_i$	$N_e F_i$	$N_e C_{ij}$	$N_e C_{ij}/N_e F_i$
Discrete generations												
Generation 3	191	0.00532	17.0 (14.4; 19.9)	0.0681	7.3 (5.2; 9.9)	216	0.20 ± 0.08	3	0.10 ± 0.04	5.0 ± 0.8	5.4 ± 0.5	1.08
Generation 4	133	0.00770	17.3 (13.5; 21.8)	0.0951	5.3 (2.3; 9.3)	147	0.25 ± 0.04	4	0.09 ± 0.01	5.5 ± 0.2	5.6 ± 0.3	1.02
Yearly sampling												
Cohort ₂₀₀₆	74	0.01410	6.3 ⁴ [9.4 ⁵ (7.7; 11.2)]	0.0925	5.4 (4.1; 6.8)	83	0.17 ± 0.15	2.7 ± 0.7	0.10 ± 0.10	4.6 ± 1.9	5.0 ± 0.6	1.09
Cohort ₂₀₀₇	136	0.00753	27.5 ⁴ [41.2 ⁵ (31.9; 53.7)]	0.0390	12.8 (3.1; 29.4)	235	0.20 ± 0.10	3.2 ± 0.4	0.10 ± 0.05	5.2 ± 1.0	6.2 ± 0.4	1.19
Cohort ₂₀₀₈	211	0.00481	16.9 ⁴ [25.3 ⁵ (20.5; 31.1)]	0.0666	7.5 (4.1; 12)	563	0.22 ± 0.05	3.8 ± 0.4	0.08 ± 0.02	6.0 ± 0.6	7.2 ± 0.4	1.20
Cohort ₂₀₀₉	225	0.00451	18.9 ⁴ [28.4 ⁵ (23.3; 34.4)]	0.0442	11.3 (6.3; 18.8)	724	0.22 ± 0.05	4.2 ± 0.5	0.08 ± 0.02	6.3 ± 0.8	8.3 ± 0.4	1.32
Cohort ₂₀₁₀	134	0.00764	13.2 ⁴ [19.8 ⁵ (16.4; 23.7)]	0.0062	8.4 (0.1; 40.4)	643	0.25 ± 0.07	4.7 ± 0.6	0.08 ± 0.02	6.6 ± 0.7	9.2 ± 0.3	1.39
Biannual sampling ¹												
Sampling ₂₀₀₆₋₂₀₀₇	210	0.00483	20.1 ⁴ [30.2 ⁵ (25.8; 38.3)]	0.0500	10.0 (6.3; 14.5)	318	0.19 ± 0.11	3.1 ± 0.5	0.10 ± 0.06	5.0 ± 1.4	5.7 ± 0.5	1.14
Sampling ₂₀₀₇₋₂₀₀₈	347	0.00291	24.2 ⁴ [36.3 ⁵ (29.3; 44.6)]	0.0356	14.1 (4.6; 28.8)	798	0.21 ± 0.07	3.6 ± 0.5	0.09 ± 0.03	5.7 ± 0.8	6.7 ± 0.4	1.18
Sampling ₂₀₀₈₋₂₀₀₉	436	0.00231	20.0 ⁴ [30.0 ⁵ (25.2; 35.5)]	0.0570	8.8 (5.7; 12.5)	1287	0.22 ± 0.05	4.0 ± 0.5	0.08 ± 0.02	6.2 ± 0.7	7.7 ± 0.4	1.24
Sampling ₂₀₀₉₋₂₀₁₀	359	0.00281	21.8 ⁴ [32.7 ⁵ (28.3; 37.7)]	0.0240	20.8 (7.2; 41.5)	1367	0.24 ± 0.06	4.4 ± 0.6	0.08 ± 0.03	6.5 ± 0.7	8.6 ± 0.4	1.32

¹Sampling mimicking the mean generation interval reported by Menéndez et al. (2016a) for the whole pedigree of the Gochu Asturcelta breed (1.8 ± 0.03 years).

²Number of individuals involved in the estimates.

³Values obtained removing alleles with frequencies (P_{crit}) lower than 0.05.

⁴Estimates of effective size after correction for bias due to age structure.

⁵Original estimates of effective size and confidence intervals.

generations when sample size was high (16.9 for Cohort₂₀₀₈ and 18.9 for Cohort₂₀₀₉). However, when sample size (Cohorts 2006, 2007, and 2010) was lower, estimates were clearly biased downward or upward. Using biannual samplings, mimicking the average generation interval as recommended by Waples et al. (2013, 2014), the corrected estimates were biased upward varying from 20.0 for Sampling_{2008–2009} to 24.2 for Sampling_{2007–2008}. The increase of sampling period to three years did not change the scenario described above (Table S3). In any case, it is worth mentioning that before demographic correction (using ratio N_b/N_e), $N_{e(LD)}$ estimates were always unacceptably biased upward (Table 2).

In general, estimates of $N_{e(M)}$ took lower values than their $N_{e(LD)}$ counterparts (Table 2). $N_{e(M)}$ for discrete filial generation F_4 (5.3) was significantly lower than that of F_3 (7.3) due to a noticeable increase in molecular coancestry (9.51% vs. 6.81% in F_3). When yearly or biannual samplings were considered, estimates of $N_{e(M)}$ followed a similar trend to those of $N_{e(LD)}$: the higher the $N_{e(LD)}$ values the higher the $N_{e(M)}$ estimates. Except for Sampling_{2008–2009}, decreases in \hat{r}^2 coincided with lower molecular coancestry values leading to estimates of $N_{e(M)}$ and $N_{e(LD)}$ highly biased upward (see Cohort₂₀₀₇ in Table 2). Again, the increase of sampling period to three years did not give any improvement in estimating N_e (Table S3).

Estimates of N_e were also obtained using a temporal method, previously tested in the livestock framework (Goyache et al. 2011), to gain more evidence on performance of single-sample methods to estimate N_e when samples are drawn from a number of yearly cohorts (Table 3). Estimates were corrected for overlapping generations by multiplying the original values with the ratio C/G (2.23) corresponding to a correction factor, C , computed following Jorde and Ryman (1995, 1996), of 4.01.

Although the program FactorC gave an estimate of generation interval, G , of 1.93 years, the “real” G of the population of 1.8 (± 0.03) years reported by Menéndez et al. (2016a) was used to the risk of slightly overestimate the $N_{e(JR)}$ values. When subsequent yearly samplings were considered, the estimates of $N_{e(JR)}$ had a noticeable variation fluctuating from 13.4 (Cohort₂₀₀₆ – Cohort₂₀₀₇) to 33.0 (Cohort₂₀₀₉ – Cohort₂₀₁₀). When the two samples used were separated by three years, the estimates obtained become more consistent varying from 23.6 (from Cohort₂₀₀₈ to Cohort₂₀₁₀) to 25.8 (from Cohort₂₀₀₇ to Cohort₂₀₀₉), therefore suggesting that drift signal was not strong enough in subsequent yearly samplings to give reliable estimates of N_e . Note that the estimates of $N_{e(JR)}$ for three-year samplings were slightly higher to the adjusted $N_{e(LD)}$ estimates obtained for biannual samplings (Table 2) and slightly lower to the adjusted and three-year sampling $N_{e(LD)}$ estimates (Table S3). In any case, these $N_{e(JR)}$ and $N_{e(LD)}$ estimates were fully comparable.

Discussion

The Gochu Asturcelta pig breed offers a very particular scenario useful to illustrate the performance of single-sample methods to estimate N_e in animal populations using molecular information. The breeding policy implemented by the breeders association allows to identify individuals that can be classified into discrete filial generations and, therefore, to compare the performance of different methods to estimate N_e under two different scenarios: generations overlap and discrete generations.

Genealogical estimates of effective size obtained using individual increase in inbreeding ($N_e F_i$) and individual increase in coancestry ($N_e C_{ij}$) kept consistency across reference populations (samples) and are in fully agreement with those recently reported by Menéndez et al. (2016a)

Table 3. Estimates of N_e obtained in the Gochu Asturcelta pig population using the temporal method of Jorde and Ryman (2007; $N_{e(JR)}$) with all possible combinations formed by subsequent and triennial samplings of the five yearly cohorts available. Both the original and the adjusted for overlapping generations estimates of $N_{e(JR)}$ are given. The 95% confidence intervals of the original estimates are in brackets. Sampling sizes for each sample regime are also provided.

Sample regime	Sample size	$N_{e(JR)}$ estimates		Confidence intervals
		Original	Adjusted	
Subsequent cohorts				
From Cohort ₂₀₀₆ to Cohort ₂₀₀₇	74–136	6.0	13.4	(4.0;11.9)
From Cohort ₂₀₀₇ to Cohort ₂₀₀₈	136–211	14.4	32.1	(11.3;19.7)
From Cohort ₂₀₀₈ to Cohort ₂₀₀₉	211–225	7.0	15.6	(4.8;13.6)
From Cohort ₂₀₀₉ to Cohort ₂₀₁₀	225–134	14.8	33.0	(11.4;20.9)
Triennial sampling				
From Cohort ₂₀₀₆ to Cohort ₂₀₀₈	74–211	11.4	25.4	(8.0;19.4)
From Cohort ₂₀₀₇ to Cohort ₂₀₀₉	136–225	11.6	25.8	(7.0;32.3)
From Cohort ₂₀₀₈ to Cohort ₂₀₁₀	211–134	10.6	23.6	(7.8;16.4)

for the most recent registered populations. Genealogical estimates are provided as a frame of reference for the understanding of the performance of the molecular-based methods to estimate N_e . Note that the genealogical methods applied correct for differences in pedigree depth and completeness of the individuals forming a reference population and, indirectly, account for the effects of mating policy, drift, overlap of generations, selection, and migration as a consequence of their reflection in the pedigree of each individual (Cervantes et al. 2008, 2009; Gutiérrez et al. 2008). Moreover, after the modification of the method suggested by Gutiérrez et al. (2009), and further applied for $N_e C_{ij}$ by Cervantes et al. (2011a), $N_e F_i$ accounts for the absence of self-fertilization allowing to obtain useful estimates of N_e using pedigrees with three equivalents to complete generations on average. In the current analysis, the lower estimates of $N_e F_i$ and $N_e C_{ij}$ were assessed for the yearly Cohort (2006) with mean pedigree depth ($t = 2.7 \pm 0.7$) on the limit of estimability (Gutiérrez et al. 2009).

It is not surprising that molecular-based estimates of N_e are higher than those obtained using genealogical data. Very recently, Silió et al. (2016), analyzing two experimental pig lines kept in herds closed for 24–28 generations and subject to a strict minimum coancestry mating policy, reported that molecular-based estimates of N_e based on either inbreeding or coancestry tended to exceed their genealogical counterparts. Unlike pedigree information, which refers to a virtually infinite number of loci, criteria based on observed molecular polymorphism refer to a finite number of loci. In any case, sampling sizes and number of loci used here can be considered enough to obtain reliable estimates of effective populations size even if the expected N_e were moderate or large (Antao et al. 2010).

Performance of the $N_{e(LD)}$ method

Even when discrete filial generations are considered, estimates of $N_{e(LD)}$ are at least threefold higher than the corresponding genealogical estimates (Table 2). However, estimates of $N_{e(LD)}$ for filial generations F_3 and F_4 were lower and steadier than those obtained using yearly or biannual sampling. The linkage disequilibrium method relies on the fact that, in a system where gametes are randomly distributed among a small number of zygotes, there will be departures from expected genotype frequencies and departures from expected gametic frequencies, both of which can be used to estimate N_e (Hill 1981; Waples 1991). These assumptions only fit well to samples obtained from age-structured populations. Moreover, in the case of overlapping generations, it is hard to assume that the available samples derive from a population with constant size. If population size changes, the

“background” LD from previous generations that has not broken down by recombination between loci and new LD generated by reproduction of a finite number of individuals reflect different effective sizes and, therefore, estimate of N_e based on \hat{r}^2 can be biased upward or downward for a few generations (Waples 2005; Waples et al. 2014).

In any case, estimates of \hat{r}^2 obtained from molecular information in the Gochu Asturcelta pig breed can be biased upward even when discrete filial generations are considered. Demographic information allows to estimate \hat{r}^2 assuming selective neutrality and constant population size as $\hat{r}^2 = \frac{1}{3H(N_e, N_b)}$, where $H(N_e, N_b)$ is the harmonic mean of N_e and N_b (see formula (5) in Waples et al. 2014). Demographic estimate of \hat{r}^2 would be here 0.00125 which underestimates the values of \hat{r}^2 obtained using molecular information whatever the sample considered (Table 2). Population studies ideally assume that LD is estimated using samples formed by unrelated individuals. This assumption is far from the pig population analyzed here and is not likely to occur in most livestock or natural animal populations therefore biasing upward the estimates of \hat{r}^2 . Even though breeding policy of the Gochu Asturcelta population is under strict control, some hidden structuring, characterized by the ratio $N_e C_{ij}/N_e F_i$ (Cervantes et al. 2011a), has appeared probably due to an excessive use for reproduction of the descendants of two founders (Menéndez et al. 2016a).

Our results confirm that estimates of $N_{e(LD)}$ obtained for filial generations F_3 and F_4 are more reliable than those assessed in scenarios with overlapping generations. Moreover, if no demographic adjustment is carried out, $N_{e(LD)}$ estimates for yearly or biannual sampling schemes were terribly wrong (Table 2). Yearly samplings appeared clearly insufficient to obtain sound estimates of $N_{e(LD)}$, probably due to small sample size (Cohorts 2006 and 2010) or sampling bias (Cohort 2007; see Table 1). Estimates can be substantially biased at small sample sizes unless the true N_e was smaller than the sample size used to estimate it (England et al. 2006; Waples 2006). Although biased upward, the current results confirm that $N_{e(LD)}$ estimates are more reliable when sampling span approaches a generation length (Waples et al. 2014). Such sampling span increases sample size, but also “homogenize” the actual number of breeders producing the sample across estimates. As theory suggests that $N_{e(LD)}$ estimates are function of the harmonic mean of N_e and N_b (Waples et al. 2014), $N_{e(LD)}$ should converge on true N_e when sampling span approaches the generation length.

Performance of other molecular-based methods to estimate N_e

The Nomura (2008) coancestry-based method gave lower estimates of N_e than $N_{e(LD)}$ (Tables 2 and S3) and nearer

to the “real” genealogical ones in the case of discrete filial generations. However, this may be due to the fact that $N_{e(M)}$ is more likely related to the number of founders represented in the samples rather than population size. Caballero and Toro (2002) reported that $1/2f$ (being f the average molecular coancestry of the analyzed population) is actually the founder genome equivalents (N_g). N_g is a key parameter to assess genetic losses due to drift which can be defined as the theoretically expected number of founders that would be required to provide the genetic diversity in the analyzed population if the founders were equally represented and had lost no alleles (Ballou and Lacy 1995). This definition is conceptually different to that of effective population size, the evolutionary analogous to census size, proposed by Wright (1931): 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 analyzed population. Even though the method by Nomura (2008) adjusts for the presence of alleles alike-in-state (Oliehoek et al. 2006), the main difference between his method and the Caballero and Toro’s (2002) approach is that self-coancestries (s_i), the diagonals in the between-individuals coancestry matrix (being $s_i = \frac{1+F_i}{2}$, where F_i is the homozygosity in a molecular context), are not included in the computations and, therefore, $\hat{f} < f$ and $N_{e(M)} > N_g$. Self-coancestries have a major importance in computing f : the lower the sample size the higher the weight of self-coancestries on f (Cervantes et al. 2011b). In any case, our results confirm the results by Miller et al. (2015) in bighorn sheep suggesting that \hat{f} tend to vary with \hat{r}^2 . Therefore, both estimates of N_e ($N_{e(M)}$ and $N_{e(LD)}$) cannot be considered independent.

Results obtained using the Jorde and Ryman’s (2007) approach illustrate that performance of temporal methods, when applied to data with overlapping generations, is highly dependent on sampling interval (Waples and Yokota 2007; Barker 2011) due to the particular age structure of the studied population. Here, $N_{e(JR)}$ estimates obtained using subsequent yearly samplings did not accumulate sufficient drift signal, therefore giving inconsistent N_e estimates. In turn, too long separation among samples gives estimates of N_e highly biased upward (Table S4). Note that, in our example, genealogical separation between subsequent yearly samples is $t \approx 0.5$ while four-year and five-year sampling plans (Table S4) are separated by 1.5 and 2 equivalent to discrete generations (Table 2). The Jorde and Ryman’s (2007) approach gave consistent estimates, comparable with adjusted $N_{e(LD)}$ estimates obtained for biannual (generation interval) sampling, under a three-year sampling plan. In our example, samples obtained under this sampling plan are separated by about $t = 1$. This scenario is consistent with the

performance of this method previously reported in horses (Goyache et al. 2011).

Conclusions

The current results confirm the fact that performance of $N_{e(LD)}$ can only be considered reliable in populations under generations overlapping when sampling span approaches a generation interval (Waples et al. 2014). Otherwise, sampling bias can affect the estimates of \hat{r}^2 , probably due to unaccounted variation in molecular coancestry among samples. This may be particularly important in scenarios in which samples are not likely to be formed by unrelated individuals. Furthermore, $N_{e(LD)}$ can only be considered useful if a correction of demographic bias is applied.

In such framework, even if no high variation of LD among yearly cohorts occur (Miller et al. 2015), the operational advantage of using single-sample methods to obtain molecular-based estimates of N_e is not clear: while two-sample methods may need a sample span exceeding a generation interval, single-sample methods (namely $N_{e(LD)}$) will need a representative sampling in each of the yearly cohorts included in that interval. These concerns particularly apply to natural and domestic populations with large generation intervals. As an example, it is worth mentioning that in domestic horses, generation interval usually exceeds 10 years (Cervantes et al. 2009). In such scenario, it is hard to assume that available samples are representative of a complete generation interval period (Corbin et al. 2010, 2012).

The current study has been performed using LD between unlinked loci. The availability of high-density SNP Chips offers the opportunity of estimating $N_{e(LD)}$ using LD between linked loci, therefore improving the performance of the method. However, the concerns about sampling span described above still apply. Actually, high-density SNP Chips have been used to ascertain the variation of N_e over time, expressed as generations in the past (Corbin et al. 2010, 2012; Flury et al. 2010). Even though some of these studies use complex models accounting for sources of variation such as sample size, mutation, phasing, or recombination rate together with thousands of linked SNP data (Corbin et al. 2012; Barbato et al. 2015), $N_{e(LD)}$ estimates at a given point of time are always function of both \hat{r}^2 and between-SNPs distance in Morgans (c). As fitting c is usually arbitrary, historical estimates of N_e mainly depend on \hat{r}^2 which, in turn, depends on sampling and demographic structure of the studied population.

Overall, the empirical evidence given in the current study confirms that estimates of N_e obtained using methods based in molecular information should be interpreted

with caution (Barker 2011; Goyache et al. 2011; Putman and Carbone 2014).

Acknowledgments

This work was partially funded by a specific contract between Government of Principado de Asturias and Asociación de Criadores de Gochu Asturcelta (ACGA; <http://www.gochuasturcelta.org/>). IA, IF, and FG are supported by grant FICYT GRUPIN14-113.

References

- Antao, T., A. Perez-Figueroa, and G. Luikart. 2010. Early detection of population declines: high power of genetic monitoring using effective population size estimators. *Evol. Appl.* 4:144–154.
- Ballou, J. D., and R. C. Lacy. 1995. Identifying genetically important individuals for management of genetic variation in pedigreed populations. Pp. 76–111 in J. D. Ballou, M. Gilpin, T. J. Foose, eds. *Population management for survival and recovery: analytical methods and strategies in small population management*. Columbia University Press, NY, USA.
- Barbato, M., P. Orozco-terWengel, M. Tapio, and M. W. Bruford. 2015. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front. Genet.* 6:109.
- Barker, J. S. F. 2011. Effective population size of natural populations of *Drosophila buzzatii*, with a comparative evaluation of nine methods of estimation. *Mol. Ecol.* 20:4452–4471.
- Caballero, A., and M. A. Toro. 2002. Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv. Genet.* 3:289–299.
- Cervantes, I., F. Goyache, A. Molina, M. Valera, and J. P. Gutiérrez. 2008. Application of individual increase in inbreeding to estimate effective sizes from real pedigrees. *J. Anim. Breed. Genet.* 125:301–310.
- Cervantes, I., J. P. Gutiérrez, A. Molina, F. Goyache, and M. Valera. 2009. Genealogical analyses in open populations: the case of three Arab-derived Spanish horse breeds. *J. Anim. Breed. Genet.* 126:335–347.
- Cervantes, I., F. Goyache, A. Molina, M. Valera, and J. P. Gutiérrez. 2011a. Estimation of effective population size from the rate of coancestry in pedigreed populations. *J. Anim. Breed. Genet.* 128:56–63.
- Cervantes, I., J. M. Pastor, J. P. Gutiérrez, F. Goyache, and A. Molina. 2011b. Effective population size as a measure of risk status in rare breeds: the case of three Spanish ruminant breeds. *Livest. Sci.* 138:202–206.
- Corbin, L. J., S. C. Blott, J. E. Swinburne, M. Vaudin, S. C. Bishop, and J. A. Woolliams. 2010. Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim. Genet.* 41:8–15.
- Corbin, L. J., A. Y. H. Liu, S. C. Bishop, and J. A. Woolliams. 2012. Estimation of historical effective population size using linkage disequilibria with marker data. *J. Anim. Breed. Genet.* 129:257–270.
- Do, C., R. S. Waples, D. Peel, G. M. Macbeth, B. J. Tillett, and J. R. Ovenden. 2014. NeEstimator V2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Mol. Ecol. Res.* 14:209–214.
- England, P. R., J.-M. Cornuet, P. Berthier, D. A. Tallmon, and G. Luikart. 2006. Estimating effective population size from linkage disequilibrium: severe bias in small samples. *Conserv. Genet.* 7:303–308.
- Flury, C., M. Tapio, T. Sonstegard, C. Drögemüller, T. Leeb, H. Simianer, et al. 2010. Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. *J. Anim. Breed. Genet.* 127:339–347.
- Goyache, F., I. Álvarez, I. Fernández, L. Pérez-Pardal, L. J. Royo, and L. Lorenzo. 2011. Usefulness of molecular-based methods for estimating effective population size in livestock assessed using data from the endangered black-coated Asturcón pony. *J. Anim. Sci.* 89:1251–1259.
- Gutiérrez, J. P., and F. Goyache. 2005. A note on ENDOG: a computer program for analysing pedigree information. *J. Anim. Breed. Genet.* 122:172–176.
- Gutiérrez, J. P., I. Cervantes, A. Molina, M. Valera, and F. Goyache. 2008. Individual increase in inbreeding allows estimating realised effective sizes from pedigrees. *Genet. Sel. Evol.* 40:359–378.
- Gutiérrez, J. P., I. Cervantes, and F. Goyache. 2009. Improving the estimation of realized effective population sizes in farm animals. *J. Anim. Breed. Genet.* 126:327–332.
- Hill, W. G. 1981. Estimation of effective population size from data on linkage disequilibrium. *Genet. Res.* 38:209–216.
- Jorde, P. E., and N. Ryman. 1995. Temporal allele frequency change and estimation of effective size in populations with overlapping generations. *Genetics* 139:1077–1090.
- Jorde, P. E., and N. Ryman. 1996. Demographic genetics of brown trout (*Salmo trutta*) and estimation of effective population size from temporal change of allele frequencies. *Genetics* 143:1369–1381.
- Jorde, P. E., and N. Ryman. 2007. Unbiased estimator for genetic drift and effective population size. *Genetics* 177:927–935.
- Luikart, G., N. Ryman, D. A. Tallmon, M. K. Schwartz, and F. W. Allendorf. 2010. Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conserv. Genet.* 11:355–373.
- Maignel, L., D. Boichard, and E. Verrier. 1996. Genetic variability of French dairy breeds estimated from pedigree information. *Interbull Bull.* 14:49–54.

- Menéndez, J., I. Álvarez, I. Fernández, B. de la Roza, and F. Goyache. 2015. Multiple paternity in domestic pig under equally probable natural matings. A case study in the endangered Gochu Asturcelta pig breed. *Arch. Anim. Breed.* 58:217–220.
- Menéndez, J., I. Álvarez, I. Fernández, and F. Goyache. 2016a. Genealogical analysis of the Gochu Asturcelta pig breed: insights for conservation. *Czech J. Anim. Sci.* 61:140–143.
- Menéndez, J., F. Goyache, A. Beja-Pereira, I. Fernández, N. A. Menéndez-Arias, R. Godinho, et al. 2016b. Genetic characterization of the endangered Gochu Asturcelta pig breed using microsatellite and mitochondrial markers: insights for the composition of the Iberian native pig stock. *Livest. Sci.* 187:162–167.
- Miller, J. M., J. Poissant, R. M. Malenfant, J. T. Hogg, and D. W. Coltman. 2015. Temporal dynamics of linkage disequilibrium in two populations of bighorn sheep. *Ecol. Evol.* 5:3401–3412.
- Nomura, T. 2008. Estimation of effective number of breeders from molecular coancestry of single cohort sample. *Evol. Appl.* 1:462–474.
- Oliehoek, P. A., J. J. Windig, J. A. M. van Arendonk, and P. Bijma. 2006. Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. *Genetics* 173:483–496.
- Putman, A. I., and I. Carbone. 2014. Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecol. Evol.* 4:4399–4428.
- Silió, L., C. Barragán, A. I. Fernández, J. García-Casco, and M. C. Rodríguez. 2016. Assessing effective population size, coancestry and inbreeding effects on litter size using the pedigree and SNP data in closed lines of the Iberian pig breed. *J. Anim. Breed. Genet.* 133:145–154.
- Waples, R. S. 1989. A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* 121:379–391.
- Waples, R. S. 1991. Genetic methods for estimating the effective size of Cetacean populations. *Rep. Int. Whal. Comm. (Special issue)* 13:279–300.
- Waples, R. S. 2005. Genetic estimates of contemporary effective population size: To what time periods do the estimates apply? *Mol. Ecol.* 14:3335–3352.
- Waples, R. S. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conserv. Genet.* 7:167–184.
- Waples, R. S., and C. Do. 2010. Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evol. Appl.* 3:244–262.
- Waples, R. S., and M. Yokota. 2007. Temporal estimates of effective population size in species with overlapping generations. *Genetics* 175:219–233.
- Waples, R. S., C. Do, and J. Chopelet. 2011. Calculating N_e and N_e/N in age-structured populations: a hybrid Felsenstein-Hill approach. *Ecology* 92:1513–1522.
- Waples, R. S., G. Luikart, J. R. Faulkner, and D. A. Tallmon. 2013. Simple life history traits explain key effective population size ratios across diverse taxa. *Proc. Biol. Sci.* 280:20131339.
- Waples, R. S., T. Antao, and G. Luikart. 2014. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. *Genetics* 197:769–780.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Life table used to calculate the ratio between effective number of breeders (N_b) and effective population size (N_e) using demographic data, as proposed by Waples et al. (2014, see references section).

Table S2. Life table used to calculate the correction factor (C) for overlapping generations proposed by Jorde and Ryman (1995, 1996, see references section).

Table S3. Number of individuals (N) involved and estimates of effective size for three-years sampling in the Gochu Asturcelta pig breed population computed via molecular-based methods (linkage disequilibrium, $N_{e(LD)}$, and molecular coancestry, $N_{e(M)}$) and pedigree information (individual increase in inbreeding, $N_e F_i$, and individual increase in coancestry, $N_e C_{ij}$).

Table S4. Estimates of N_e obtained in the Gochu Asturcelta pig population using the temporal method of Jorde and Ryman (2007; $N_{e(JR)}$) with all possible four-year and five-year sampling plans formed with combinations of the five yearly cohorts available.