

Estimation of ancient and recent effective population size from linkage disequilibrium in a closed herd of Iberian pigs

M. Saura^{*}, *J. A. Woolliams*[†], *A. Tenesa*[†], *A. Fernández*^{*}, *B. Villanueva*^{*}

^{*}Departamento de Mejora Genética Animal, INIA, Madrid, Spain, [†]The Roslin Institute and R(D)SVS, University of Edinburgh, Midlothian, UK

ABSTRACT: Effective population size (N_e) is an important parameter in conservation genetics as it can help to explain patterns of genetic variation in damaged populations. N_e has been traditionally estimated from demographic data or from molecular markers based on the temporal variance of the allele frequencies that provide pointwise or average estimates over a time period. Here we have inferred both ancestral and current N_e in an ancient closed population of Iberian pigs that is currently under serious danger of extinction, using information from linkage disequilibrium and recombination rate. Our results show that N_e has suffered a progressive decline through time in this population, with a critical estimate of current N_e of 36 individuals. The availability of genome-wide genotyping platforms allows us now to study populations from a more detailed perspective, providing information on the genetic status of the populations and on its evolution across time.

Keywords: effective population size; linkage disequilibrium; high dense SNPs; conservation genetics

Introduction

The effective size for a population (N_e) is defined as the number of individuals that with true random selection and random mating would give rise to the same rate of inbreeding as observed in the actual breeding population (Wright 1931). It is an important parameter in conservation genetics, as it can help to explain contemporary patterns of genetic variation in damaged populations, and by definition describes the rate of inbreeding accumulation and loss of genetic variation. N_e has been traditionally estimated from demographic data (not always available) or from molecular markers. Commonly used genetic estimates of N_e are based on the temporal variance in allele frequencies between two samples separated in time (Waples (1989)) or on the levels of linkage disequilibrium (LD) in the population (Laurie-Ahlberg and Weir (1979)). While the temporal method estimates the average N_e over the period separating the samples (Waples 2005), the strength of LD at different genetic distances between loci can be used to infer both current and ancestral N_e (Waples and Do (2010)). Genetic models describing the relationship between N_e and LD appear in the literature from the 1960s onwards. In particular, Sved (1971) derived a formula for the expectation of r^2 ($E[r^2]$) between a particular pair of markers as a function of N_e and recombination frequency (c). Hill (1981) explored the use of such equation in inferring N_e from LD, suggest-

ing that LD at closely linked markers would reflect ancient population history, whilst LD between markers farther apart would reflect more recent events. The main advantage of this method is the possibility of investigating the change of N_e over time, as LD between loci at a specific recombination distance reflects the ancestral N_e $1/2c$ generations ago (Hayes 2003) if linear growth is assumed. The advent of new high-density genotyping platforms has allowed us to estimate N_e from LD measures when information on a large number of closely linked markers is available (Tenesa et al. (2007); Qanbari et al. (2010); Corbin et al. (2010); (2012)).

In this study, we have estimated ancient and recent N_e from measures of LD in a closed herd of Iberian pigs that is currently under serious danger of extinction. Genotypes for more than 35,000 SNPs were obtained using the Illumina Porcine SNP60 BeadChip v1.

Materials and Methods

Samples and SNP genotypes. A herd of the strain Guadyerbas of Iberian pigs was established in 1944 from 24 founders (4 males and 20 females) and has been maintained isolated to date under a genetic conservation program that has focused on avoiding matings between relatives. The number of samples that were available for the study included 86 males and 141 females born between 1992 and 2011 and expanded about 6 generations (generation interval in this breed is about 3 years). The Illumina Porcine SNP60 BeadChip v1 comprises 62,163 probes that are distributed along 18 autosomal and two sex chromosomes, according to the map version *Sscrofa* 10.2. DNA was extracted from blood samples. Quality control was performed on SNPs and animal samples following the criteria described in Saura et al. (2013). After the quality control filters were applied, the final number of Guadyerbas samples and autosomal SNPs available for the analyses were 219 and 35,519, respectively.

LD and c estimation. For each autosome, pairwise r^2 , the squared correlation of allele frequencies at a pair of loci (Hill and Robertson (1968)), was calculated for all syntenic marker pairs in sliding windows of 50Mb, which translated in 2.2 million pairs of SNPs. This distance ensured the coverage of all SNP pairs that can potentially be in disequilibrium and saved computation time. Marker pairs < 100 bp further apart were excluded to avoid gene conversion. SNPs with minor allele frequency < 0.05 were

also discarded. In order to avoid the dependence between linkage disequilibrium and recombination rate estimated from the same data, we used independent estimates of c that were reported by Tourtereau et al. (2012). They described a high density recombination map for the pig using the Porcine SNP60 BeadChip. This was developed from four independent pedigrees, including Berkshire, Duroc, Meishan, Yorkshire and Landrace breeds.

N_e estimation. All our analyses were based on the known approximate relationship between LD (measured as r^2) and N_e . In particular, we used:

$$E(r^2) \approx 1/(\alpha + 4N_e c) + 1/n \quad [1]$$

where α was set to 2 to account for mutation and n is the chromosome experimental sample size, this term included in order to adjust for sampling effects (see Corbin et al. (2012)). Recombination rates were transformed to recombination distances (in Morgans) following Tourtereau et al. (2012), and using their chromosome average estimates of c . The mean distance and the mean r^2 between marker pairs for different size bins were computed and substituted into equation [1]. These different size bins allowed to obtain different estimates of N_e at different time points ($N_{e(t)}$) by calculating the number of generations (t) in the past as $1/2c$ (Hayes et al. (2003)). This process was carried out for markers pooled across chromosomes simultaneously, as suggested by Hayes et al. (2003) to reduce the variability of estimates of $N_{e(t)}$ caused by finite population size. In order to study the evolution of N_e since the foundation of the herd (26 generations ago) until present, we used size bins between 0.5 and 2.7 Mb.

Results and Discussion

Linkage disequilibrium. In these pigs the average r^2 between adjacent SNP in the Porcine SNP60 BeadChip was 0.53 (± 0.42) and the mean distance between adjacent SNPs was 0.068 Kb. Table 1 shows the average LD for each autosome for distances of 0.5, 5, 10 and 50 Mb. The average across chromosomes ranged from 0.08 (50 Mb) to 0.42 (0.05 Mb bins). Figure 1 shows, as expected, a decline in r^2 with increasing distance between syntenic marker pairs. Average r^2 for SNPs at distances between 0 and 0.05Mb (i.e., representing the intercept) was 0.61. The most rapid decline was observed at marker distances lower than 0.9Mb where mean r^2 decreased by more than half. In order to obtain an approximation of the LD that can be expected by chance in this population, we calculated r^2 between non-syntenic markers taken at random and found that at distances greater than 15Mb, average r^2 between syntenic SNPs was reduced to non-syntenic levels, achieving nearly an asymptotic value. Both the amount and the extension of LD were much higher than those observed in other pig breeds (Harmegnies et al. (2006)); Uimari and Tapio (2011)) and this may be explained by the high inbreeding rate detected in the Guadyerbas population (Saura et al. (2013)).

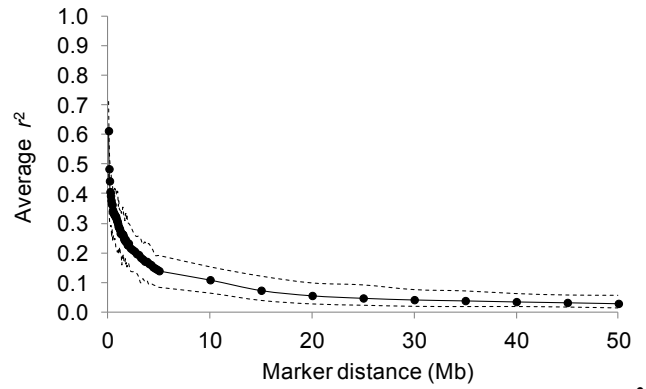


Figure 1. Average linkage disequilibrium measured as r^2 (solid line) and 5th and 95th percentiles (dashed lines) plotted against distance.

Average r^2 is represented according to distance into bin sizes of: (i) 0.05Mb bins for distance between 0 to 2Mb; (ii) 0.2Mb bins for distance between 2 to 5Mb; and (iii) 5Mb bins for distance over 5 Mb (up to 50 Mb).

Table 1. Average r^2 for each autosome (and averaged over the whole genome) for distances of 0.5, 1, 5, 10 and 50 Mb.

	N SNP	0.5Mb	1Mb	5Mb	10Mb	50Mb
SSC1	3953	0.24	0.18	0.12	0.10	0.06
SSC2	2462	0.42	0.37	0.24	0.18	0.09
SSC3	1868	0.46	0.41	0.28	0.22	0.11
SSC4	2337	0.49	0.44	0.28	0.20	0.07
SSC5	1567	0.40	0.36	0.22	0.16	0.07
SSC6	2352	0.48	0.42	0.28	0.21	0.08
SSC7	2295	0.38	0.33	0.21	0.16	0.07
SSC8	2064	0.50	0.45	0.28	0.21	0.11
SSC9	2308	0.41	0.37	0.24	0.18	0.08
SSC10	1217	0.32	0.28	0.17	0.12	0.05
SSC11	1410	0.37	0.31	0.19	0.14	0.06
SSC12	1022	0.38	0.33	0.18	0.13	0.06
SSC13	2802	0.52	0.47	0.31	0.25	0.16
SSC14	2530	0.47	0.42	0.28	0.21	0.10
SSC15	2099	0.44	0.40	0.26	0.20	0.10
SSC16	1134	0.42	0.36	0.22	0.18	0.08
SSC17	1054	0.47	0.41	0.26	0.20	0.09
SSC18	1045	0.39	0.34	0.20	0.14	0.06
Average	1973	0.42	0.37	0.23	0.18	0.08
SD	769	0.07	0.07	0.05	0.04	0.03

The number of SNP markers (N SNP) within each autosome is also indicated.

Estimates of N_e . Figure 2 plots estimated N_e at generations in the past. Based on these results a progressive decrease of N_e across generations was observed. The pointwise estimate of N_e when the herd was founded, 26 generations ago, was 42 animals. This value of N_e larger than N (24 founders) makes sense if the variance in family size was lower than the mean when the population was established (see Caballero (1994)), which we ignore but that would be consistent with the establishment of the conservation program when the herd was founded. However, the discrepancy between N_e and N observed here could be explained if genetic distances between SNPs varied across

populations (i.e. between our population and that from the study of Tourtereau et al. (2012)). This would misestimate the number of generations back the estimate of N_e is from. The current estimate of N_e was close to 36 individuals, a low number considering the critical size of 50 recommended by FAO. In those cases where applicable, using LD measures to estimate N_e has the remarkable advantage (over other methods) of giving insight on the evolutionary history of a population, which is crucial for conservation purposes.

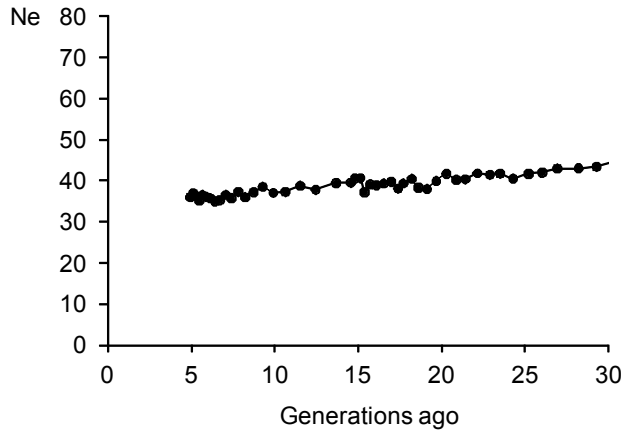


Figure 2. Ancestral and current N_e estimated from bins of size ranging from 0.5 to 15Mb across all autosomes.

For each autosome and for each SNP distance bin, N_e was estimated from the mean r^2 obtained here and the mean c from Tourtereau et al (2012), using equation [1]. The number of generations from the foundation of the herd and the present was calculated as $1/2c$.

Conclusion

We have inferred both current and ancestral effective population size in an ancient closed population of Iberian pigs based on the observed amount of LD. Our results show that Guadyerbas effective population size has decreased since the foundation of the herd, achieving a critical value according to the current estimate of N_e . The availability of high dense genotyping platforms allows now the study of in danger populations providing a more detailed picture of the genetic diversity and its evolution across time.

Literature Cited

- Caballero, A. (1994). *Heredity* 73:657-679. Corbin, L., Blott, S. C., Swinburne, J. E. et al. (2010). *Anim. Genet.* 41:8-15. Corbin, L., Liu, A. Y. H., Bishop, S. C., J. B. et al. (2012). *J. Anim. Breed. Genet.* 129:257-270. Harmegnies, N., Farnir, F., Davin, D. et al. (2006). *Anim. Genet.* 37:225-231. Hill, W. G. (1981). *Genet. Res. Camb.* 38:209-216. Hill, W. G. and Robertson, A. (1968). *Theor. Appl. Genet.* 38:226-231. Laurie-Ahlberg, C. C. and Weir, B. S. (1979). *Genetics* 92:1295-1314. Robinson, J. D. and Moyer, G. R. (2012). *Evol. Appl.* 6:290-302. Saura, M., Fernández, A., Rodríguez, M. C. et al. (2013). *PLoS ONE* 8: e78314. Sved, J. A. (1971). *Theor. Pop. Biol.* 2:125-141. Tenesa, A., Navarro, P., Hayes, B. et al. (2007). *Genome Res.* 17:520-526. Tourtereau, F., Servin, B., Frantz, L. et al. (2012). *BMC Genomics* 13:586. Uimari, P. and Tapio, M. (2011). *J. Anim. Sci.* 89:609-614. Waples, R. S. (1989). *Genetics* 121:379-391. Waples, R. S. and Do, C. (2010). *Evol. Appl.* 3:244-262. Wright, S. (1931). *Genetics* 16:97-169.