Estimating Rate of Inbreeding and Effective Population Size using Genomic Data in Norwegian Red Cattle

B. Hillestad[1], J.A. Woolliams[1,2], T. Meuwissen[1], D.I. Våge[1] and G. Klemetsdal[1]
[1]Norwegian University of Life Sciences, Norway, [2]University of Edinburgh, UK.

ABSTRACT: Traditionally, rates of inbreeding and effective population sizes have been estimated by the use of pedigree data. Here, inbreeding coefficients were estimated from runs of homozygosity in 322 Norwegian Red bulls born between 1982 and 2002. Further, inbreeding rates were estimated by regressing the natural logarithm of ($1-F_{ROH}$) on year of birth, resulting in an inbreeding rate per generation of 0.303 % and a corresponding effective population size of 165 individuals. This resembles the estimates made by the industry in 2011 based on pedigree information, giving an inbreeding rate of 0.26 % and an effective population size of 194. These results suggests that these two parameters can be estimated by the use of genomic data only, with possible application also to wild and/or endangered populations.
Keywords:
Effective population size (Ne)
Rate of inbreeding (ΔF)
Runs of Homozygosity (ROH)
Norwegian Red cattle

Introduction

Effective population size (Ne) and rate of inbreeding (ΔF) measure the loss of genetic variation, and can serve as tools for the management of inbreeding in livestock populations. In the advent of genomic data, attempts have been made to estimate historical Ne from a population view, based on genomic data only using linkage disequilibrium (LD) (MacLeod et al. (2009); Tenesa et al. (2007); Hayes et al. (2003)). One challenge is to convert genomic information from an individual F, to ΔF and Ne, without the use of pedigree information. These parameters could become useful in selection programs, in gene conservation projects, as well as in the management of wild populations or populations without a herdbook.

One method that is increasingly used to measure inbreeding (F) is runs of homozygosity (ROH), which are stretches of homozygous segments present in the genome caused by parents transmitting identical haplotypes to their offspring. If the same ancestral haplotype are inherited from both parents to the offspring, homozygosity occurs (Broman et al. (1999)).

Estimation of ΔF and Ne based on either molecular or pedigree data was compared by Saura et al. (2013) in Iberian pigs. Inbreeding rates were obtained by regressing the natural logarithm of (1-F) on year of birth, where F was estimated either from molecular or genealogical coancestry. Methods based upon LD show weaknesses in addressing the most recent generations when estimating effective population sizes (Hayes et al. (2003); Tenesa et al. (2007); Corbin et al. (2012); MacLeod (2009)).

The Norwegian Red cattle population possesses accurate pedigree and genotyping records, qualifying this population as suitable for the aim of this paper; to estimate ΔF and Ne based on genomic data from ROH, with the use of pedigree for comparison purposes only.

Materials and Methods

F_{ROH} was estimated from ROH in 381 Norwegian Red bull genotyped with an Illumina HD-chip containing 777,962 SNP-markers. Genotyping quality controls included: exclusion of markers on sex chromosomes, call rate per SNP > 90 % and deviation from Hardy-Weinberg ($P > 10^{-6}$), resulting in a density of 707,609 SNP.

ROH were identified with PLINK 1.07 (Purcell et al. (2007)) operating with sliding windows of 5,000 Kb. The following criteria ROH detection were set: (i) the minimum interval of adjacent homozygous SNP had to exceed 0.5 Mb; (ii) one heterozygous SNP was allowed within a window (assumed to be a genotyping error); (iii) the number of missing SNP allowed within a window were set to three; (iv) the maximum physical gap between adjacent SNP within a run were set to 250 Kb; and (v) the minimum SNP density within a run were one SNP every 50 Kb.

Individual F_{ROH} were estimated by the following equation:

$$F_{ROH} = \frac{\sum L_{ROH}}{\sum L_{AUTO}}$$

where $\sum L_{ROH}$ is the total length of ROH in one individual, and $\sum L_{AUTO}$ is the total length of the genome covered with SNP (McQuillan et al. (2008)).

In the present study, ΔF and Ne have been estimated on genomic data only, by regressing the natural logarithm of (1-F) on year of birth (Y):

$$y = \mu + \beta Y + e,$$

where $y=\ln(1-F_Y)$; $\mu= \ln(1-F_0)$; and $\beta= \ln(1-\Delta F_Y)$, utilizing data for those bulls that were born between 1982 to 2002 (322 bulls in total) of the 381 genotyped animals. From the estimate of the regression coefficient β we calculated by the following equation:

$$\Delta F_{ROH} = (1 - e^{\beta})L$$
$$Ne_{ROH} = 2\Delta F_{ROH}^{-1}$$

where ΔF_{ROH} is the inbreeding rate per generation and L is the generation interval estimated to 4.6 years by assuming the following:

$$L = (L_{SS} + L_{SD} + L_{DS} + L_{DD}) / 4$$
$$L = (6.5 + (0.4 \cdot 2.5 + 0.6 \cdot 6.5) + 3.5 + 3.5) / 4 = 4.6$$

where L_{SS}, L_{SD}, L_{DS} and L_{DD} is the generation interval between the sire and his sire, the sire and his dam, the dam and her sire and the dam and her dam, respectively.

Results

F_{Ped} ranged between 0-0.1289 with an average F_{Ped} equal 0.0202 and variance of 2.1E-04, while F_{ROH} ranged between 0.0389-0.2312 (0.1107, 7.49E-04) (Table 1). Even though F_{Ped} and F_{ROH} are rather different in average levels, Figure 1 illustrates clearly how these two parameters follow each other on a yearly basis.

ΔF_{ROH} and Ne_{ROH} gave quite similar results to those that *Geno* reports for the Norwegian Red population, based on pedigree data only (Geno (2011)). By the use of our method, ΔF_{ROH} was estimated to 0.00303 and Ne_{ROH} to 165 animals, while the breeding organization reports ΔF_{Ped} equal 0.0026 and Ne_{Ped} of 194 animals (Table 2).

Discussion

By knowing the recombination rate (*c*), different measures of linkage disequilibrium (LD) can be used to estimate effective population size. Sved et al. (1971) presented a method where LD was measured as the square correlation of allele frequencies at a pair of loci ($E[r^2]$). Another approach is measuring LD by chromosome segment homozygosity (CSH); defined as the probability of two chromosome segments of the same location and size, drawn at random from a population, to origin from the same ancestor without an intervening recombination (Hayes et al. (2003)). By analyzing the relationship between the length of CSH and past generations, assuming constant linear population growth, Ne and historical Ne can be estimated. By combining results of Sved (1971) and Hayes et al. (2003), past and recent Ne can be estimated (Tenesa et al. (2007). There seem though to be some difficulties estimating recent Ne by these methods due to lack of consistency (Corbin et al. (2012)). ROH and CSH have big similarities, but ROH have estimated ΔF and Ne from individual Fs instead of estimating ΔF and Ne directly from a population view.

In the current study there is only a small difference between Ne_{ROH} and Ne_{Ped}. The bias can be explained by the following: (i) Different approximations of the generation length have been used in the two methods. More exact results could be obtained by combining birthdate and pedigree data. (ii) Only sires have been genotyped. The genotyping of females may reduce bias, because the genotyped sires form a selected part of the population. (iii) Ne_{ROH} was calculated based on bulls born between 1982 and 2002, while the reference Ne_{Ped} was calculated in 2011. Some bias are likely to occur due to the two periods Ne was calculated.

The possibility of estimating Ne from an individual level based on genotyping, opens the opportunity to obtain Ne and ΔF in wild and endangered populations, without a pedigree, but with (approximately) known generation intervals. This could be a valuable tool to estimate the loss of genetic variation, evaluate the risk of endangerment and to control inbreeding also in wild populations.

Conclusions

The increase of F_{ROH} follows F_{Ped} closely over time, suggesting the possibility of estimating ΔF_{ROH} and Ne_{ROH} from an individual level based on genomic data only; 0.00303 and 165, compared with industry estimates of 0.0026 and 194, respectively. The method keeps the possibility of estimating Ne and ΔF in wild and/or endangered populations without a herdbook available, and as such with value also in gene conservation programs.

Acknowledgments

We would like to thank the Norwegian University of Life Sciences for founding this project, and the breeding organization for Norwegian Red cattle in Norway: *Geno*, for providing pedigree and genotype data. Also, a big thanks to BOKU and Solomon Antwi Boison for helping out with PLINK, genotyping quality control and ROH detection.

References

Broman, K. W. and Weber, J. L. (1999). Am. J. Hum. Genet. 65: 1493-1500.

Corbin, L. J., Liu, A. Y. H., Bishop, S. C. et al. (2012). J. Anim. Breed. Genet. 129: 257-270.

Geno (2011). Årsberetning og Regnskap 2011.

Hayes, B. J., Visscher, P. M., McPartlan, H. C. et al. (2003). Genome Res. 13: 635-643.

MacLeod, I. M., Meuwissen, T. H. E., Hayes, B. J. et al. (2009). Genet. Res. 91: 413-426.

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R. et al. (2008). Am. J. Hum. Genet. 83: 359 - 372.

Purcell, S., Neale, B., Todd-Brown, K. et al. (2007). Am. J. Hum. Genet., 81.

Saura, M., Fernández, A., Rodriguez , M. C. et al. (2013). Plos One 8.

Sved, J. A. (1971). Theor. Popul. Biol. 2: 125-141.

Tenesa, A., Navarro, P., Hayes, B. J. et al. (2007). Genome Res. 17: 520-526.

Table 1. Descriptive statistics of the inbreeding coefficients estimated by either pedigree or runs of homozygosity in Norwegian Red cattle.

Method	$\bar{F}$	σ^2	Range	
			Min.	Max.
Pedigree	0.0202	2.1E-04	0	0.1289
ROH	0.1107	7.49E-04	0.0389	0.2312

Table 2. Rates of inbreeding and effective population size in Norwegian Red cattle calculated by traditional pedigree data {Geno, 2011 #144} or from use of runs of homozygosity.

Estimated from pedigree	
ΔF_{Ped}	0.0026
Ne_{Ped}	194
Estimated from runs of homozygosity	
ΔF_{Y}	0.000659
ΔF_{ROH}	0.00303
Ne_{ROH}	165

Figure 1. Average inbreeding coefficients in Norwegian Red cattle by year of birth, estimated either by pedigree or from the use of runs of homozygosity.