

The Use Of SNP Data For The Estimation Of Effective Population Size In Cattle Breeds

C. Flury^{*}, M. Tapio^{†‡}, T. Sonstegard[§], C. Drögemüller^{**}, T. Leeb^{**}, O. Hanotte^{†††}, H. Simianer^{‡‡} and S. Rieder^{*}

Introduction

Knowledge on effective population size (N_e) and its trend contain relevant information for the monitoring of genetic diversity in livestock populations. However, in extensive production systems the relevant information to estimate effective population size is often lacking. In the recent past densely spaced genome-wide SNP data became available for cattle and other livestock species. LD between such markers contains information about effective population size (Hayes et al. 2003). In this study, we investigate the use of genome-wide SNP-data for marker based estimation of effective population size and we present estimates for recombination rates for two taurine cattle breeds from Africa (N'Dama and Sheko) and two local cattle breeds from Switzerland (Eringer and Original Brown Cattle). The two local cattle breeds from Switzerland are understood as reference breeds, as pedigree-information is available for the evaluation of marker based estimates of recent effective population size.

Material and methods

Data

In total 330 individuals were genotyped for the Illumina BovineSNP50 beadchip. 51 of the genotyped individuals belong to the breed Sheko (SHK) and 38 of the genotyped individuals to the breed N'Dama – both taurine cattle breeds of Africa. The two local Swiss breeds were represented by 128 genotypes (Eringer breed; ERI) and 113 genotypes (Original Brown Cattle; OB), respectively. SNP alignment was done with the bovine genome assembly Btau_4.0. For SHK, NDA and ERI 53'903 SNPs had known physical locations (in bp) on the 29 *Bos taurus* autosomal chromosomes. The number of known physical locations was less for OB (51'000 SNPs). OB was genotyped in a different lab than the other three samples. Therefore the names from 2'903 SNPs were different and the assignment of the correct position was not possible.

Derivation of LD and recombination rates

After filtering Haploview (Barrett, 2005) was used to estimate pair-wise r^2 -values for the remaining SNPs. For the estimation of recombination rates, PHASE 2.1 (Crawford et al.,

^{*} Swiss College of Agriculture, 3052 Zollikofen, Switzerland

[†] MTT, 31600 Jokioinen, Finland

[‡] International Livestock Research Institute ILRI, Nairobi, Kenya

[§] USDA, ARS Beltsville Agricultural Research Centre, Beltsville, MD 20705, USA

^{**} Institute of Genetics, Vetsuisse Faculty, University of Berne, Berne, Switzerland

^{††} The School of Biology, University of Nottingham, NG7 2RD Nottingham, UK

^{‡‡} Department of Animal Sciences, Georg-August-University, Goettingen, Germany

2004; Li & Stephens, 2003) was applied using the settings to model recombination based on a subset of 50 individuals. For further description see Flury et al. (2010).

Estimation of effective population size

N_e was estimated based on the known relationship between r^2 , N_e and the inter-marker genetic map distance c between two loci under consideration. In the absence of mutation, the expectation of r^2 is:

$$E(r^2) = \frac{1}{(1 + 4N_e c)} \quad [1] \text{ Sved (1971).}$$

Two different values for inter-marker genetic map distance c were used: a) the approximation of c by using physical map distances directly and b) estimating c as genetic map distances based on estimated recombination rates between markers. This way the effect of the approximation on estimated N_e could be analysed. The mean r^2 -values and the two different values for c were applied in formula 1. For the 128 genotyped ERI individuals and the 113 genotyped OB individuals pedigree information was kindly provided by the Swiss Eringer Breeders' Federation and the Swiss Brown Cattle Breeders' Federation, respectively. The total pedigree of ERI and OB consisted of 4,798 individuals and 4,112 individuals, respectively. Pedigree-based N_e estimates were obtained as described in Flury et al. (2010).

Results and discussion

The fraction of informative SNPs from the total number of SNPs with known map position differed between breeds: 66% OB; 63% ERI, 52% SHK and 45% NDA, respectively. For the two African breeds the fraction of monomorphic SNPs (NDA 40%; SHK 26%) was higher compared to the two Swiss breeds (ERI and OB 15% each). Beside others, the fraction of monomorphic SNPs is seen as a result of the breeds considered for the definition of the chip. It is expected, that for the definition of the 50k Illumina beadchip the focus was mainly on global cattle breeds.

Derivation of LD and recombination rates

For all four breeds the r^2 -values rapidly decline with increasing distance. For small distances the highest r^2 -values were found for N'Dama and the lowest for Sheko, respectively. The high r^2 -values for N'Dama are in line with the results of the bovine Hapmap Consortium (2009). The variation in r^2 -values between different autosomes for a given distance is larger for the two local breeds of Switzerland than for the two African breeds (not shown). This variation might be a result of the more intense selection on specific traits in the cattle breeding programmes of Switzerland.

The estimated recombination rates along chromosome 10 are given for each breed in figure 1. Generally, the recombination rates vary along different chromosomal regions. Based on microsatellites Thomson et al. (2001) found for four genomic regions differences in recombination rates between breeds. Considering figure 1 no large differences in recombination rates become obvious between the four breeds. If this finding holds for the remaining 28 autosomes as well, it seems appropriate to estimate recombination rates for the breed with the highest number of informative markers. These estimated genetic map distances could then be used for all other breeds with genotypes from the same SNP-chip. More detailed analysis of the estimated recombination rates along different regions of the 29 autosomes and between breeds are planned for the near future.

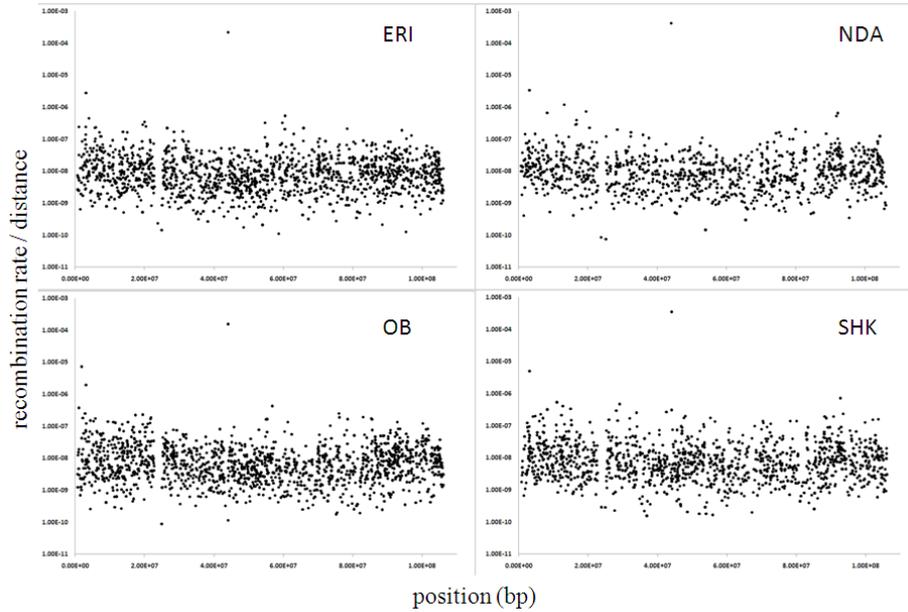


Figure 1: Estimated recombination rates scaled according physical map distance of BTA10 for ERI, OB, NDA and SHK, respectively.

Estimation of effective population size

The results for the N_e -estimates of the four breeds by using the approximation for genetic map distance c are given in figure 2a). In figure 2b) estimated N_e using derived genetic map distances in Morgan units for NDA, SHK and ERI are presented. At the actual state, estimation of recombination rates was not finished for all autosomes of the OB-sample and therefore no estimates for OB could be presented in figure 2b). Variation in recombination rates levels out over larger distances and the effect of varying recombination rates can be ignored for the estimation of recent N_e from 0 – 50 generations or inter-marker distances above 1 Mb (Flury et al. 2010). Based on this finding, it seems appropriate to avoid time intensive estimation of recombination rates for the monitoring of actual within breed diversity. Recent effective population size (2 generations ago) for the four breeds was found in the range from 57 (NDA) to 73 (OB) to 83 (SHK) to 149 (ERI). This result is surprising as the N'Dama is the breed with the largest actual population size. Further the results for actual population size of NDA and SHK differ from the results from Bovine Hapmap (2009). The definition of “recent” effective population size (i.e. number of generations ago) might influence these estimates and has to be considered for comparison. Further the SNP-density of the chip used might lead to differences in the extent of observed LD. For the two local breeds of Switzerland the estimates for recent effective population sizes are in the range of the pedigree based estimates (not shown).

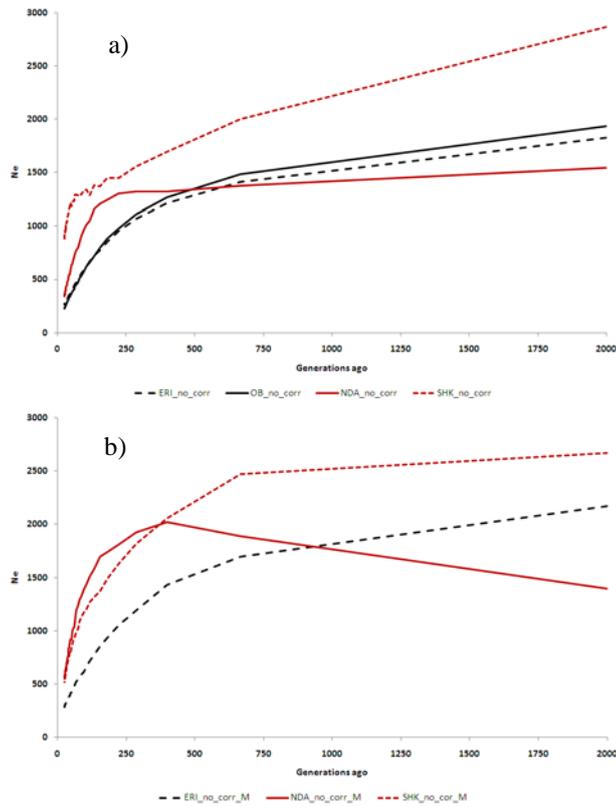


Figure 2: Estimated effective population size using physical map distances a) and estimated genetic map distances b).

Conclusion

Recombination rates along BTA10 and estimates of N_e were presented for four cattle breeds. The drastic drop in actual N_e must be very recent, as for all of the four breeds N_e was >300 25 generations ago. From conservationist's point of view immediate action is required to better understand reasons for this drop and for accurate action on population level.

References

- Barrett, J.C., *et al.* (2005). *Bioinformatics*, 21(2):263–265.
- Crawford, D.C., *et al.* (2004). *Nat. Genet.*, 36:700–706.
- Flury, C., *et al.* (2010). *J. Anim. Breed. Genet.*, (in press).
- Hayes, B.J., *et al.* (2003). *Gen. Res.*, 13(4):635–643.
- Sved, J.A. (1971). *Theor. Pop. Biol.*, 2(2):125–141.
- The Bovine HapMap Consortium. (2009). *Science*, 324(5926):528–532.
- Thomsen, H., *et al.* (2001). *Mamm. Genome*, 12:724–728