Estimation of Individual Levels of Admixture in Crossbred Populations from SNP Chip Data: Examples with Sheep and Cattle Populations

J. Sölkner[1], A. Frkonja[1], H.W. Raadsma[2], E. Jonas[2], G. Thaller[3], E. Gootwine[4], E. Seroussi[4], C. Fuerst[5], C. Egger-Danner[5] and B. Gredler[1]

[1] University of Natural Resources and Applied Life Sciences Vienna, Department of Sustainable Agricultural Systems, Division of Livestock Sciences, Gregor Mendel Str. 33, A-1180 Vienna, Austria.
[2] University of Sydney, Faculty of Veterinary Science, Camden, NSW 2570, Australia.
[3] Universität Kiel, Institut für Tierzucht und Tierhaltung, Olshausenstr. 40, D-24098 Kiel, Germany.
[4] The Volcani Center, Institute of Animal Science, POB 6, Bet Dagan, 50250, Israel
[5] ZuchtData EDV-Dienstleistungen GmbH, Dresdner Str. 89/19, A-1200 Vienna, Austria.

Abstract

Individual levels of admixture in crossbred livestock populations are traditionally derived as expected values from pedigree information. SNP chip data provide the opportunity of estimating crossbreeding levels more accurately. Also, breed composition may be estimated for animals without knowledge of known ancestors. Scientists working in human genetics have developed tools for deriving levels of admixture for large groups of individuals when part of them can be assigned to parental (local) populations. The aims of this study were to evaluate the accuracy of results using one of these programs, STRUCTURE (Hubisz et al., 2009) in an Awassi-Merino sheep resource population and to apply the method to estimate levels of Holstein Friesian admixture in ~2500 Austrian and German Fleckvieh bulls.

The Awassi-Merino resource population was developed at the University of Sydney by mating Awassi sires to Merino dams and consecutive mating of the resulting F1 crossbred animals to Merino or Merino-Awassi ewes for the development of backcross, double backcross and intercross generations. A total of 1247 animals including Awassi and Merino founders and crossbred animals were genotyped using the Illumina ovine SNP chip, only genotypes of chromosomes 1 and 26 were used further. Identity by descent (IBD) contributions of the Awassi founders, analysed using the program MERLIN enabled a very accurate estimation of the proportion of Awassi contributions. While pedigree based Awassi admixture levels ranged from 0.25 to 0.4375, the IBD derived levels ranged from 0.09 to 0.60 for a set of 60 animals. The correlation of pedigree- and IBD derived levels of crossbreeding was 0.57. Estimates derived using the STRUCTURE software, including genotypes of 23 improved Awassi of Israeli origin and 75 Australian Industry Merino, were similar to the IBD-derived values (r=0.94), the correlation of STRUCTURE based and pedigree based levels of crossbreeding was 0.66. Estimation of Holstein Friesian admixture with STRUCTURE for 2046 Fleckvieh bulls with reference populations of 417 "superpure" Fleckvieh and 500 Holstein Friesian animals resulted in an average of 0.0331, similar to the 0.0359 based on pedigree information. The correlation of these two types of estimate was 0.62.

We conclude that, for the populations analysed, individual levels of breed composition based on panels of purebred reference animals - but ignoring pedigree - may be estimated more accurately than with pedigree information alone. Combining pedigree information with incomplete sets of genotyped ancestors is a challenge to be yet met.

Keywords: crossbreeding, admixture, SNP chip, sheep, cattle

1. Introduction

Crossbreeding is a widely used strategy in livestock improvement. In pedigreed populations, levels of crossbreeding are usually derived from the pedigree charts of animals, assuming strict halving of the contributions of individuals across generations of progeny. The four grand parents of an individual are considered to have contributed a quarter each of the genome of an animal, great grand parents contribute one eighth each according to this method of calculation. Recombination of parental chromosomes in the process of meiosis leads to deviation from these expected values in the actual genomes of individuals. SNP chip data may be used to trace ancestral haplotypes in each chromosome and identity by descent (IBD) states provide a more accurate representation of ancestral contributions. For fully genotyped crossbred populations, software like MERLIN (multi-point engine for rapid likelihood inference, Abecasis et al., 2002) may be used for deriving crossbreeding levels via IBD states of purebred animals in the crossbreds. Accurate estimates of crossbreeding levels are also interesting in non-pedigreed populations. Farmers in Uganda have developed crossbreeding schemes of local Ankole cattle with Holstein Friesian (Mulindwa et al., 2009). In a study, a milk recording system was implemented in 16 large herds. Levels of crossbreeding were assigned to the animals based on farmer memory of the pedigrees. As the "Friesian" bulls used in these herds are never pure Holstein Friesian but rather high level crosses, information on the true level of admixture would have helped when analysing the relative performance of animals with different levels of crossbreeding.

Methods of estimating levels of admixture for samples of admixed individuals relative to samples of parental populations have been developed by scientists working in human genetics. Model based clustering methods implementing hidden Markov models are most frequently applied. STRUCTURE (Pritchard et al., 2000, with extensions by Hubisz et al., 2009) is the most frequently used software package.

In this study we tested the accuracy of the method as implemented in STRUCTURE in comparison to IBD derived breed composition using animals of the University of Sydney Awassi x Merino resource population. We also applied STRUCTURE to infer levels of Holstein Friesian admixture in Austrian and German Fleckvieh cattle and compared these values with pedigree based admixture values.

2. Material and Methods

2.1 Data

At the University of Sydney a resource population of a complex crossbreeding design of Awassi (A) and Merino (M) sheep (Raadsma et al., 2009) was established to exploit the extreme differences between these two types of sheep in a range of production characteristics. The Awassi type used was designated Improved Awassi which has been developed in Israel by selection for high milk production (Gootwine and Pollott, 2000). The Merino used in the experiment is of much smaller body size compared to the Awassi but with greatly superior wool charcateristics compared to the carpet wool of the Awassi. The design involved four Awassi founder rams, which were mated to Merino ewes to produce four F1 AM rams. These rams were further mated to Merino ewes to generate AMM backcrosses with expected 0.25 Awassi contribution. These backcrosses were again mated to the AM rams to produce AM_AMM intercrosses (0.375 A) and in later generations AM_AM_AMM (0.4375 A) and AM_AMM_AM_AMM (0.375 A) intercross animals were generated.

Genotypes from the Illumina ovine SNP50 BeadChip were available for 1456 animals of this ovine resource population. As only four Awassi founders contributed to the population, additional genotypes from 47 Awassi were provided by the Volcani Center, Israel. Among these, 23 animals are from the same Improved Awassi strain as the founders of the Australian resource population. Additional 24 animals were of the non-improved local Awassi type.

To complement the relatively small number of purebred Merino in the pool of genotypes, the International Sheep Genomics Consortium provided genotypes of 293 animals comprising various types and strains of Merino. A total of 5347 SNP on chromosome 1 and 782 SNP on chromosome 26 were used for the current analysis.

For the analysis of Holstein Friesian admixture in Fleckvieh bulls in Austria and Germany, Illumina bovine SNP50 BeadChip genotypes of 2463 Fleckvieh animals were provided by ZuchtData GmbH while genotypes of 2402 Holstein Friesian bulls were provided by the German Genotrack Project (Thaller *et al.,* 2009). For the analysis presented here, 3343 SNP on chromosome 1 and 987 SNP on chromosome 25 were used.

2.2 Methods

The first part of this study allows a derivation of IBD states connecting the crossbred animals to the four Awassi founders of the resource population. MERLIN was used to estimate probabilities of two individuals sharing zero (IBD_0), one (IBD_1) or two (IBD_2) alleles for each SNP on the basis of identity by descent (http://www.sph.umich.edu/csg/abecasis/merlin/tour/ibd.html). The IBD probabilities $(0.5*p(IBD\_1)+p(IBD\_2))$ for each SNP of a crossbred individual were merged with any ancestor in its pedigree that was an Awassi founder. The average of this value from all SNPs was considered as the proportion of Awassi ancestry in this animal. MERLIN was performed for the pedigrees of 20 animals of the fourth generation with pedigree breed composition of 0.4375 A, giving breed composition estimates for 60 crossbred animals beyond F1 (20 with 0.25 A, 20 with 0.375 A and 20 with 0.4375 A, based on pedigree information).

For the STRUCTURE population analysis ignoring pedigree information we added genotypes of 23 Improved Awassi from Israel and 75 Australian Industry Merino from the ISGC pool to the 4 Awassi and 6 Merino founders from the resource population. In the process, we had applied STRUCTURE and confirmed that the Improved Awassi were close to the Awassi animals of the resource population and the Australian Industry Merino were close to the Merino founders of the resource population. The Bayesian clustering algorithm was performed using 1000 burn in and 1000 evaluation runs, prior population information considered the 27 Awassi and 81 Merino individuals to estimate admixture levels of 64 individuals, 4 F1 and the 60 described above.

For the STRUCTURE analysis of the bovine samples, 417 Fleckvieh bulls were considered "superpure". To be included in this set, bulls were required to have very long and complete pedigrees and no Friesian ancestor in the pedigree. Only bulls with all known ancestors of at least four complete generations having exclusively Fleckvieh ancestors according to the national herdbook database were considered For the Holstein Friesian reference population, 500 bulls were randomly sampled from the available pool of potentially admixed individuals (N=2046 bulls). STRUCTURE run specifications were the same as for the ovine analysis.

3. Results

IBD analysis revealed that the range of individual Awassi ancestry for the crossbred sheep under consideration was much larger than the levels of 0.25 to 0.4375 derived from pedigree information (Table 1). Considering only chromosomes 1 and 26, the range was 0.091 to 0.648.

Figure 1. Admixture plot presenting STRUCTURE results for 27 Awassi (1), 81 Merino (2) and 64 crossbred (0) animals.
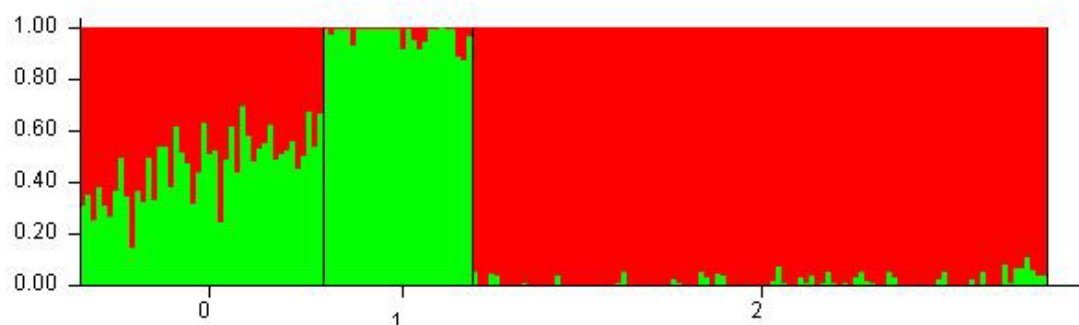

Table 1. Awassi admixture levels (means and indicators of variability) for 60 crossbred animals.

	Mean	S.D.	Min	Max
Pedigree	0.354	0.079	0.250	0.4375
IBD-derived	0.371	0.116	0.091	0.648
STRUCTURE	0.341	0.111	0.094	0.534

Table 2. Correlations of individual Awassi admixture levels estimated from pedigree, levels of IBD and cluster analysis.

	Pedigree	IBD-derived	STRUCTURE
Pedigree	1	0.57	0.66
IBD-derived		1	0.94
STRUCTURE			1

Results of STRUCTURE provided similar results in terms of mean and variation of admixture levels (Table 1 and Figure 1). STRUCTURE was run in a mode that did not fix breed composition at 100% for purebred animals because of the uncertain ancestry of the Australian Awassi. Other settings of STRUCTURE gave very similar levels of admixture for crossbred animals (r=0.999). Admixture levels derived from IBD and via the STRUCTURE algorithm were very highly correlated (0.94) using the animals of the ovine resource population. Pedigree admixture was slightly higher correlated with STRUCTURE (0.66) admixture than IBD admixture (0.57). Regarding the bovine samples, the average pedigree based Holstein Friesian admixture level in the Fleckvieh population was 0.0359 for the 2046 bulls considered. The average admixture based on STRUCTURE results was 0.0331. The range of admixture levels was higher for pedigree estimates (S.D of 0.065, maximum of 0.320) than for STRUCTURE results (S.D. of 0.035, maximum of 0.226). The correlation between the two estimates was 0.62.

4. Discussion and Conclusions

The results with a correlation of 0.94 of IBD derived levels of admixture and levels based on clustering algorithms ignoring pedigree information are very encouraging, considering situations where pedigree information is not available or ancestors of individuals of interest may not be genotyped. The analyses presented here will be expanded by involving a much larger sample of crossbred animals for which IBD derived levels of admixture will be

evaluated. Alternative methods of deriving admixture for population samples (e.g., LAMP, Panasiuc *et al.,* 2009) will be tested and compared. To test the robustness of the approach, the founder individuals of the crossbred population will be excluded and more remote groups of "purebred" animals (e.g. the local Awassi instead of the improved Awassi) will be included.

Results can also be expanded to determine a minimum subset of ancestry informative SNP that can be used to accurately determine individual admixture in populations with both known and unknown ancestry.

Individual levels of crossbreeding that are more accurate than those derived from pedigree information provide interesting information to members of the Interbull community. It is worth while to investigate the approaches presented here as well as others that make use of the genomic relationship matrix as implemented in genomic selection procedures.

5. References

Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. 2002. Merlin -rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet. 30,* 97-101.

Gootwine, E. & Pollott, G.E. 2000. Factors affecting milk production in Improved Awassi dairy ewes. *Anim. Sci. 71,* 607-615.

Hubisz, M.J., Falush, D., Stephens, M. & Pritchard, J.K. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources 9,* 1322-1332.

Mulindwa, H., Galukande, E., Wurzinger, M., Okeyo, M.A. & Sölkner, J. 2009. Modelling of long term pasture production and estimation of carrying capacity of Ankole pastoral production system in South Western Uganda. *Livestock Research for Rural Development 21,* art. no. 151.

Pasaniuc, B., Sankararaman, S., Kimmel, G. & Halperin, E. 2009. Inference of locus-specific ancestry in closely related populations. *Bioinformatics 25,* 213-221.

Pritchard, J.K., Stephens, M. & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics 155,* 945-959.

Raadsma, H.W., Thomson, P.C., Zenger, K.R., Cavanagh, C., Lam, M.K., Jonas, E., Jones, M., Attard, G., Palmer, D. & Nicholas, F.W. 2009. Mapping quantitative trait loci (QTL) in sheep. I. A new male framework linkage map and QTL for growth rate and body weight. *Genetics Selection Evolution 41 ,* art. no. 34.