A Four-step Approach for Selecting a Genetically Diverse Group of Animals from Pedigree Data using the Example of Endangered Austrian Goat Breeds

Michael KLAFFENBÖCK [1(✉)]
Beate BERGER [2]
Gábor MÉSZÁROS [1]
Johann SÖLKNER [1]

Summary

A four-step approach using pedigree data to create genetically diverse groups of animals for genotyping from a limited pool of available samples was developed and applied to a data set comprising five endangered Austrian goat breeds. Animals were selected according to their raw gene contribution, number of offspring, average relatedness to the living population and relatedness to other selected animals. Not all criteria could be applied to the same degree to all breeds due to small number of available samples. Methods to cope with this circumstance are presented in the paper. Inbreeding coefficient was lower in the selection groups than in the total populations in all breeds. Genotype data derived from the selected animals will be used to analyse diversity parameters and compare with available data from other Alpine goat breeds.

Key words

pedigree, goats, genetic diversity, selection method, genetic resources

[1] University of Natural Resources and Life Sciences,
Gregor-Mendel-Straße 33, 1180 Vienna, Austria
✉ e-mail: michael.klaffenboeck@boku.ac.at
[2] AREC Raumberg-Gumpenstein, Institute of Organic Farming and Biodiversity of Farm Animals, Austrasse 10, 4601 Thalheim, Austria

Received: April 25, 2017 | Accepted: July 25, 2017

ACKNOWLEDGEMENTS
This research was funded by the H2020-project IMAGE. Furthermore, we would like to thank Tatjana Walzer from Xenogenetik for providing information about their available samples, Michaela Umgeher from the Austrian Federal Sheep and Goat Breeding Association (ÖBSZ) for allowing us to use pedigrees of Austrian goat breeds and Christian Fürst from ZuchtData for providing said pedigrees to us.

Introduction

To enhance the management of genetic resources from endangered farm animal breeds, underlying factors like genetic diversity, inbreeding and breed distinctiveness need to be better understood. To investigate these parameters, five endangered should Austrian goat breeds will be SNP-chip genotyped, analysed and compared to available genomic data of other Alpine breeds. Genetic diversity studies on goats which use genomic data and a similar methodology as planned here often don't describe the selection process of the genotyped animals at all (Kijas et al., 2013; Brito et al., 2015), or simply state that minimally related animals were selected (Nicoloso et al., 2015; Visser et al., 2016). Therefore, an approach was developed combining various indicators based on pedigree information (Sölkner et al., 1997; Boichard et al., 1997; Binder et al., 2016), to find the geneticallly most diverse group of 30 animals per breed from a pool of available samples. Especially in combination with a method to confirm the desired degree of relatedness from genomic data, e.g. as indicated by Talenti et al. (2017), the approach presented here might improve efficiency of selection in genomic analyses of farm animals.

Material and methods

Five endangered Austrian goat breeds should be genotyped and therefore genetically diverse groups of 30 animals per breed had to be selected. Sperm or blood samples were available from two sources: the national gene bank of Austria and the bio bank of the private company Xenogenetik. The breeds considered were: Blobe goat (BLZ), Chamois-coloured Alpine goat (GG), Pinzgau goat (PZ), Styrian pied goat (SS) and Tauern pied goat (TA).

To select 30 animals from the pool of available samples for genotyping, a four step procedure was applied. In the first step, the program *prob_orig* (probability of gene origin) of the Fortran package PEDIG (Boichard, 2002) was used to calculate gene contributions for each ancestor of each breed. Raw rather than marginal gene contribution was considered to ensure that no influential available animals are missed because their gene contributions were accounted for by un-available ancestors. From those animals with a measurable raw gene contribution, a maximum of ten animals was selected.

The second step consisted of selecting animals due to their number of offspring. A simple in-house script in R was used to count the number of offspring for each available animal from the respective pedigree file. All available animals with less than ten offspring were excluded, and a maximum of ten animals with the highest number of offspring were selected for each breed.

In the third step, all remaining samples of each breed were separately grouped, as well as all animals marked as 'alive' in the pedigree. Average relatedness of animals for which samples are available to the living population of the respective breed was calculated using the *par2* program of the PEDIG package. Between ten and 30 animals with the lowest average relatedness to the living population of their breed were selected, filling up to the sample size of 30 selected animals per breed. Note that the number of 30 animals per breed was chosen according to the resources and possibilities of this study, and that the allocation of selection spots per step needs to be adjusted if a study allows for a different number of animals to be selected per breed.

The fourth step of checking relatedness with other already selected animals was associated with the other three steps, being executed after pre-selection of each animal. To exclude half-sib or closer relationships, an upper limit for the relationship coefficient between two selected animals was set at 0.15. Relationship coefficients were calculated using *parente* program in the PEDIG package. Pre-selected animals which showed a closer relationship to an already selected animal were excluded.

Further PEDIG programs used to describe the pedigree data set were *ped_util* for restructuring the pedigree, *ngen* for complete generation equivalents, *meuw* for inbreeding coefficients and *prob_orig* for number of founders and contribution of ten most important ancestors.

Results and discussion

The data set used is described in Table 1. Population sizes indicated by number of breeding females were between 197 in BLZ and 2,154 in GG. Pedigree sizes varied greatly between breeds with between 1,015 in BLZ and 24,368 entries in GG. The lowest number of available samples in the two bio banks combined was 63 for BLZ while the highest was 702 for TA. Average complete generation equivalent for the living population (and for the available animals in brackets) was lowest in BLZ with 1.97 (1.20) and highest in TA with 7.17 (6.73). Gene contribution of the ten most important ancestors, which was calculated

Table 1. Data structure

	BLZ	GG	PZ	SS	TA
Active female breeding goats (EFABIS AUSTRIA, 2016)	197	2,154	405	288	1,250
No. of animals in the pedigree	1,015	24,368	4,325	1,914	9,000
No. of samples available	63	541	240	150	702
Complete generation equivalent (living population)[1]	1.97	5.19	4.70	3.58	7.17
Complete generation equivalent (available animals)[1]	1.20	4.70	3.63	2.84	6.73
No. founders	152	2,552	467	176	231
Gene contribution of most important ten ancestors	0.32	0.17	0.30	0.42	0.62

[1] Average complete generation equivalent from last eight generations, only non-founders considered.

Table 2. Values for selection criteria and number of selected animals for respective criterion in brackets, as well as inbreeding coefficients F for full breeds and selection candidates

	Selection criteria			Inbreeding coefficient F	
	Raw gene contribution	Most offspring	Average relatedness	Living population	Selected animals
BLZ	0.028 (10)	- (0)	<0.001 (20)	0.007	0.000
GG	0.009 (10)	109 (10)	<0.001 (10)	0.014	0.006
PZ	0.020 (10)	30 (10)	0.005 (10)	0.023	0.001
SS	0.011 (9)	23 (10)	0.011 (11)	0.019	0.004
TA	0.019 (6)	40 (10)	0.087 (14)	0.100	0.061

as a pedigree indicator of genetic structure of breeds, ranged between 0.17 in GG and 0.62 in TA.

After applying the approach to the data set displayed in Table 1, a group of 30 animals was selected for each of the five endangered Austrian goat breeds. As indicated in Table 2, between 6 (TA) and 10 (BLZ,GG, PZ) animals were selected due to their raw gene contribution, with breed averages ranging from 0.009 in GG and 0.028 in BLZ. With the exception of BLZ, where no animal could be selected due to its number of offspring, 10 animals per breed were selected according to this criterion, 109 offspring in GG being the highest and 23 offspring in SS the lowest breed average among these candidates. Between 10 (GG, PZ) and 20 (BLZ) animals were selected due to their low average relatedness to animals classified as 'living' in the pedigree. Breed averages of average relatedness ranged from <0.001 in (BLZ, GG) to 0.087 (TA).

The fourth step of not allowing relationship coefficients >0.15 among selected animals proved the most difficult to execute strictly due to high levels of relatedness among available animals. This may partly be caused by the small population sizes at the time of the sample collection and the fact that the animals available in bio banks do not necessarily represent the current level of genetic diversity in the breed. To be able to select 30 animals per breed anyway, the relationship coefficient limit was raised to 0.25 for SS and TA. This sufficed to select 30 animals in SS, but for TA relatedness between all available samples was still too high. Therefore, the six animals that remained to be selected in TA were chosen by lowest average relatedness to the already selected 24 animals.

Inbreeding coefficients were calculated to compare the group of selected animals with the living population. With the exception of TA, low inbreeding coefficients were observed in all breeds, ranging from 0.007 (BLZ) to 0.023 (PZ) in the living population and from 0.000 (BLZ) to 0.006 (GG) in the selection groups. In TA, the inbreeding coefficient in the living population was 0.100 and 0.061 in the selection group. Inbreeding was therefore always lower in the selection group compared to the living population.

If this approach is applied to select animals for genotyping, an additional step using genomic data could be used to confirm that maximum levels of relatedness were not surpassed. Talenti et al. (2017) suggest a method to confirm unrelatedness by analysing the number of discordant homozygotes at each locus between all pairs of selected individuals.

Conclusion and recommendation

A four-step selection approach was used to select the most suitable candidates for genotyping based on their pedigree data. However, not all criteria (primarily relatedness to already selected animals) could be applied in all populations as hoped due to the small number of available samples and high levels of inbreeding, but methods to cope with this circumstance were developed and tested in this study. When genomic data is available from these samples, the usefulness of the presented approach to select diverse groups of animals can be put to the test as described above. Furthermore, the genotype data will be used to investigate genetic diversity, levels of inbreeding and distinctiveness of the breeds, and will be put into context by comparing it to freely available genotype data from other Alpine goat breeds.

References

Binder S., Berger B., Sölkner J. (2016). Few important ancestors contribute much of the genetic material: Pedigree analysis of three highly endangered cattle breeds in Austria. DAGENE meeting 27.

Boichard D., Maignel L., Verrier É. (1997). The value of using probabilities of gene origin to measure genetic variability in a population. Genet Sel Evol 29: 5-23.

Boichard D. (2002). PEDIG : a Fortran Package for Pedigree Analysis Suited for large populations. As available under https://qgsp.jouy.inra.fr/index.php?option=com_content&view=article&id=53&Itemid=57 (25.04.17, 15:50).

Brito L. F., Jafarikia M., Grossi D. A., Kijas J. W., Porto-Neto L. R., Ventura R. V., Salgorzaei M., Schenkel F. S. (2015). Characterization of linkage disequilibrium , consistency of gametic phase and admixture in Australian and Canadian goats. BMC Genet 16: 1-15.

EFABIS Austria. http://efabis.raumberg-gumpenstein.at/ (as by 16.02.2017 15:00).

Kijas J. W., Ortiz J. S., McCulloch R., James A., Brice B., Swain B., Tosser-Klopp G., International Goat Genome Consortium (2013). Genetic diversity and investigation of polledness in divergent goat populations using 52 088 SNPs. Anim Genet 44: 325-335.

Nicoloso L., Bomba L., Colli L., Negrini R., Milanesi M., Mazza R., Sechi T., Frattini S., Talenti A., Coizet B., Chessa S., Marletta D., D'Andrea M., Bordonaro S., Ptak G., Carta A. Pagnacco G., Valentini A., Pilla F., Ajmone-Marsan P., Crepaldi, P., International Goat Genome Consortium (2015). Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. Genet Sel Evol 47:62.

Sölkner, J., Filipcic L., Hampshire N. (1997). Genetic variability of populations and similarity of subpopulations in Austrian cattle breeds determined by analysis of pedigrees. Animal Science 67: 249-256.

Talenti A., Bertolini F., Pagnacco G., Pilla F., Ajmone-Marsan P., Rothschild M. F., Crepaldi P., The Italian Goat Consortium (2017). The Valdostana goat: a genome-wide investigation of the distinctiveness of its selective sweep regions. Mamm Genome 28:114.

Visser C., Lashmar S. F., van Marie-Köster E., Poli M. A., Allain D. (2016). Genetic Diversity and Population Structure in South African, French and Argentinian Angora Goats from Genome-Wide SNP Data. PLoS One 11:5.

---

acs82_18