Proc. Assoc. Advmt. Anim. Breed. Genet. 19:79-82

EFFECTS OF ASCERTAINMENT BIAS WHEN TESTING FOR POLLED IN AUSTRALIAN BEEF CATTLE

J.M. Henshall[1], E. Piper[2] and C.A. McDonald[3]

Cooperative Research Centre for Beef Genetic Technologies
[1]CSIRO Livestock Industries, FD McMaster Laboratory Chiswick, Armidale, NSW, 2350
[2]Animal Genetics Laboratory, University of Queensland, Gatton, Qld, 4343
[3]Australian Limousin Breeders Society Ltd, Armidale, NSW, 2350

SUMMARY

During pre-commercialisation validation of a marker test for polledness in beef cattle, hundreds of animals have been tested. Breeders selected which animals to test, and in the Limousin breed, animals were predominantly polled. This ascertainment bias affects the estimates of polled haplotype frequencies obtained using these data. In Limousin, an allele that appears to be commonly associated with horns in the wider population appears to be frequently associated with polled in the animals submitted for testing. This allele is also common in Angus, and the use of Angus as base cows in the grading up process may have resulted in polled alleles segregating that are of Angus origin, in addition to polled alleles originating in purebred Limousin.

INTRODUCTION

Breeding for polledness in beef cattle has been of interest in recent years due to increasing concern about the animal welfare implications of dehorning. The fastest way to increase the proportion of polled calves is to exclusively use homozygous polled bulls. However, in many breeds, homozygous polled bulls cannot be distinguished from heterozygous bulls by phenotype alone. Consequently, there is a need for molecular tests that allow bulls to be identified and marketed as homozygous polled, without the requirement for a progeny test. Although the general location on BTA1 of the locus responsible for most variation in horn phenotype is well known (Georges *et al*. 1993; Brenneman *et al*. 1996; Drogemuller *et al*. 2005), to date, the causal mutation has not been identified. Current tests for the polled genotype therefore utilise linked markers, and while these tests may have high accuracy within breeds or breed types, no single marker test is available that has been validated to perform well across all breeds.

The association between the microsatellite *CSAFG29* and the polled phenotype was discovered in the Brahman breed (Prayaga *et al*. 2009; Mariasegaram *et al*. 201X). However, the utility of the marker in other breeds was also evaluated. Most recently, as part of a pre-commercialisation trial, cattle breeders were invited to submit DNA for testing. Unlike a designed experiment, there is potential for ascertainment bias in such a trial. Breeders are unlikely to submit samples from horned animals as they are most likely to be homozygous horned, or scurred animals as they are unlikely to produce a homozygous polled genotype, or animals that are known to be homozygous polled by pedigree, unless for marketing purposes. The phenotypes submitted by breeders may also be less reliable than those from experimental datasets, as breeders may deliberately provide an incorrect phenotype as part of their own evaluation of the marker test. However, the samples are representative of those that would be received from industry using a commercial product, and so provide an important evaluation of the test as it would be applied in the marketplace.

In this study we examine the effects of ascertainment bias when testing for polled. In particular we report some results from the Limousin breed, where polled and horned are at intermediate frequencies. Based on the molecular marker data and knowledge of the history of Limousin in Australia, we explore the estimates of polled haplotype frequencies and predictions of genotype obtained from the pre-commercialisation trial.

MATERIAL AND METHODS

DNA and phenotypes were available on 143 pedigreed Limousin cattle (referred to here as the Limousin commercial population) submitted for testing by Australian breeders. One animal was phenotypically scurred, all of the others were phenotypically polled. All animals were genotyped for the *CSAFG29* microsatellite. Phenotypes and *CSAFG29* marker test results were also available for the 52 Limousin cattle and 91 Angus cattle from the validation study reported by Mariasegaram *et al*. (201X). Here, these animals are referred to as the experimental populations. They were chosen from herds judged to be representative of the diversity of genetics available to industry, and contained both polled (n = 29), scurred (n = 1) and horned (n = 22) animals in the case of Limousin, while all Angus were polled. In Prayaga *et al*. (2009) one *CSAFG29* allele was identified that in Brahman cattle did not occur in horned animals. Here we refer to that allele as allele zero (A0). Twelve other alleles segregate in the Limousin and Angus populations described above and we refer to them here as alleles A1 to A12.

Alleles at *CSAFG29* were summarised by phenotype and these summaries were used to estimate the linkage between each marker allele and alleles at the causal mutation for polled (coded P and H). In Angus, all *CSAFG29* alleles were assumed to be in complete linkage disequilibrium with the P allele at the causal mutation, forming haplotypes A0P to A12P. For Limousin, haplotype frequencies and penetrance probabilities (i.e., P(phenotype|genotype)) were estimated from the data as follows. Given a matrix of penetrance probabilities and a vector of frequencies for polled haplotypes A0P to A12P, the vector of phenotype probabilities can be calculated for each marker test genotype. We found the penetrance probabilities and haplotype frequencies that minimised the sum of squares obtained from the difference between this vector of phenotype probabilities and the observed phenotype vector for all animals, subject to the constraints that haplotype frequencies were $\geq$ zero and $\leq$ 1.0, and that penetrance probabilities were $\geq$ 0 and summed to 1.0 within genotypes. The calculations were carried out using the solver function in Microsoft Excel. This estimation was conducted 3 times: using the Limousin experimental population, the Limousin commercial population, and using both Limousin datasets combined.

RESULTS AND DISCUSSION

Table 1 contains penetrance probabilities estimated using the Limousin experimental, Limousin commercial, and Limousin combined datasets. The probabilities estimated using the experimental dataset are reasonably similar to those estimated using the combined dataset, but the estimates obtained using the commercial dataset are very different. There is no power to estimate penetrance probabilities from the commercial dataset as there are no phenotypically horned animals, and only 1 scurred animal. Consequently, for all but 1 arbitrary genotype the probability of a polled phenotype can equal 1.0, the other genotype (PP in this case) having a small probability of producing scurs.

Table 1. Penetrance probabilities - probability of phenotype (P or H) given genotype at the causal mutation (PP, PH or HH) - estimated using the experimental, commercial or combined Limousin datasets

	Experimental			Commercial			Combined		
Genotype	PP	PH	HH	PP	PH	HH	PP	PH	HH
Phenotype									
P	1.00	0.87	0.00	0.92	1.00	1.00	1.00	0.96	0.00
S	0.00	0.07	0.00	0.08	0.00	0.00	0.00	0.03	0.00
H	0.00	0.06	1.00	0.00	0.00	0.00	0.00	0.01	1.00

In Table 2 allele frequencies and polled haplotype frequencies are displayed for the Angus, Limousin experimental, Limousin commercial and Limousin combined datasets. These were estimated concurrently with the penetrance probabilities in Table 1, except for Limousin commercial, where there was no power to estimate penetrance probabilities. For the Limousin commercial dataset penetrance probabilities were fixed to be those estimated from the Limousin experimental dataset.

Table 2. Allele frequencies and frequencies of the allele forming a polled haplotype, estimated from the Angus and Limousin datasets. The polled haplotype frequencies were estimated concurrently with the penetrance probabilities given in Table 1, except for Limousin commercial, where penetrance probabilities from Limousin experimental were used

	Allele Frequency				Polled Haplotype Frequency		
	Angus	Limousin			Limousin		
Allele		experimental	commercial	combined	experimental	commercial	combined
A0	0.30	0.36	0.55	0.50	1.00	1.00	1.00
A1	0.58	0.33	0.34	0.33	0.15	0.89	0.43
A2	0.00	0.11	0.00	0.03	0.00	-	0.00
A3	0.06	0.06	0.02	0.03	0.21	1.00	0.31
A4	0.00	0.05	0.02	0.03	0.00	1.00	0.20
A5	0.00	0.07	0.01	0.02	0.00	1.00	0.00
A6	0.04	0.01	0.01	0.01	0.00	1.00	0.00
A7	0.00	0.01	0.02	0.02	0.00	1.00	0.44
A8	0.00	0.00	0.01	0.01	-	1.00	1.00
A9	0.01	0.00	0.01	0.01	-	1.00	1.00
A10	0.00	0.00	0.00	0.00	-	1.00	0.00
A11	0.00	0.00	0.00	0.00	-	1.00	1.00
A12	0.01	0.00	0.00	0.00	-	-	-

As expected, the A0 allele was more frequent in the Limousin commercial dataset than in the Limousin experimental dataset. This increase is at the expense of alleles A2 to A6, all of which have low frequencies of polled haplotypes in the experimental dataset but are almost missing from the commercial dataset. The frequency of the A1 allele is similar across all Limousin datasets, but the frequency with which it forms a polled haplotype is not. In the experimental dataset, allele A1 forms a polled haplotype only 15% of the time, while in the combined dataset the frequency was 43%. The very high polled haplotype frequency for all alleles in the Limousin commercial dataset is an artefact of the lack of horned and scurred animals in this dataset. Even if penetrance probabilities are assumed known, polled haplotype frequencies cannot be estimated at all from datasets that contain no scurred or horned animals, and are likely to be biased if only a few scurred or horned animals are present. That the single scurred animal carried an A1 allele is evident from allele A1 being the only one for which the polled haplotype frequency is less than 1.0. In the Angus dataset, where all alleles are assumed to form polled haplotypes, allele A1 was at the highest frequency, and almost 90% of alleles are A0 or A1.

The most common marker genotype in the Limousin commercial dataset, carried by 44% of animals, was heterozygous A0-A1. In Table 3, genotype estimates for heterozygous A0-A1 animals are provided, estimated using the experimental dataset, the commercial dataset, or both Limousin datasets. The question is- which one should we use when reporting results for new samples from polled animals? Clearly, not the estimates derived using haplotype frequencies estimated from the commercial dataset. However, although biased, the commercial dataset does

suggest that the A1P haplotype may be at higher frequency in breeder submitted samples than in the experimental dataset. The difference could be due to the small sample size in the experimental dataset, or due to real differences between the frequency of the A1P haplotype in polled animals likely to be submitted for testing, and the breed frequency of the A1P haplotype. So haplotype frequency estimates derived from the experimental dataset may not be appropriate for calculating genotype probabilities for commercial samples. The estimates from the combined datasets, while appearing reasonable, will be totally dependent on the relative numbers of individuals in the experimental and commercial datasets.

Table 3. Probabilities of polled genotype (PP, PH or HH) for heterozygous A0-A1 animals given polled haplotype frequencies estimated from the 3 Limousin datasets

Dataset Genotype	Experimental	commercial	combined
PP	0.15	0.89	0.43
PH	0.85	0.11	0.57
HH	0.00	0.00	0.00

One possible explanation for the between dataset differences in A1P haplotype frequency is suggested by noting that allele A1 is the most common allele in Angus, where it is assumed to form a polled haplotype. Examination of the pedigree of the Limousin animals in the commercial dataset revealed that all have a component of Angus in their ancestry, obtained during the grading up process in Australia. The haplotype formed by allele A1 may depend on the origin of the allele: horned if from French pure Limousin, or polled if from Angus.

CONCLUSIONS

Estimates of polled haplotype probabilities are required when predicting polled genotype from marker genotype. Ideally these estimates would be specific for the populations being submitted for testing. However, samples submitted for a commercial test are likely to have considerable ascertainment bias: potentially only samples from polled individuals might be submitted. This makes them unsuitable for estimating polled haplotype probabilities, so validation studies will always be required that use data that does not originate from commercial genotyping operations. In the case of Australian beef cattle a study that meets this criterion is underway.

ACKNOWLEDGEMENTS

This project is supported by Meat and Livestock Australia. We acknowledge the use of DNA and phenotypes provided by Limousin breeders.

REFERENCES

Brenneman R.A., Davis S.K., Sanders J.O., Burns B.M., Wheeler T.C., Turner J.W. and Taylor, J.F. (1996) *J. Hered.* **87**:156

Drogemuller C., Wohlke A., Momke S. and Distl O. (2005) *Mamm. Genome* **16**:613

Georges M., Drinkwater R., King T., Mishra A., Moore S.S., Nielsen D., Sargeant L.S., Sorensen A., Steele M.R., Zhao X.Y., Womack J.E. and Hetzel D.J.S. (1993) *Nat. Genet*. **4:**206

Mariasegaram M., Harrison B.E., Bolton J.A., Tier B., Henshall J.M., Barendse W. and Prayaga K.C. (201X) *Submitted for publication*

Prayaga K.C., Mariasegaram M., Harrison B., Tier B., Henshall J.M. and Barendse W. (2009) *Proc. Ass. Adv. Anim. Breed. Genet.* **18**:92