CORRECTING SAMPLING BIAS IN MICROSATELLITE MARKER TESTING FOR POLLEDNESS

N.K. Connors, R.G. Banks and D.J. Johnston

Animal Genetics & Breeding Unit*, University of New England, Armidale, NSW, 2351 Australia

SUMMARY

The poll microsatellite test has been available to Australia's beef industry for approximately 7 years and in that time, the bias in polled phenotyped animals submitted for testing from industry has influenced the accuracy of polled probability assignment to observed haplotypes. This article describes examples of observed mis-assigned haplotypes and their respective phenotypic observations, and the steps taken to correct the poll probabilities and resulting genotype estimations.

INTRODUCTION

The costs associated with carcase defects are largely attributed to damage from horned animals (Prayaga 2007). While dehorning is common practice to address these issues, questions remain regarding the animal's welfare, and breeding naturally polled animals provides a long term solution. The microsatellite DNA marker test for polledness was developed by the Beef Cooperative Research Centre and CSIRO (Henshall *et al.* 2011), and has been available to Australia's beef industry for approximately 7 years. In that time, samples submitted from industry have been biased towards polled submissions, due to a logical disinterest in testing horned animals. Prior to this study, the vast majority of phenotypes submitted to the test were unknown (>60%), over a quarter polled (27%) and the least horned (5%) and scurred (5%) (Connors *et al.* 2018). Given the number of potential haplotypes possible, there is no realistic option of a large enough reference population. As such the test uses all available industry data to estimate genotypes and an appropriate representation of different phenotypes should be present in the data so that microsatellite haplotypes can be assigned the appropriate poll probability based on the observed phenotypes. The bias in polled phenotypes has influenced the accuracy of the genotype estimations, such that haplotypes which should be assigned as horned, have been mis-assigned as polled due to only polled phenotypes being observed with this haplotype. Recently additional horned phenotypes were sourced for inclusion into the test to correct this sampling bias and to demonstrate the effect that these additional phenotypes have on haplotype poll probability assignment. This paper describes a number of haplotypes with mis-assigned poll probabilities, the resulting genotype estimations, and the effect of including additional horned phenotypes on the haplotypes' assignments.

MATERIALS AND METHODS

The microsatellite test estimates an animal's genotype as homozygous polled (PP), heterozygous polled (PH), or homozygous horned (HH), and detailed methodology has been discussed previously (Piper *et al.* 2014; Connors and Tier 2016; Connors *et al.* 2016). Briefly, samples submitted for testing are accompanied with a phenotype (i.e. horned, polled, scurred, or unknown). The test uses ten microsatellite markers to form haplotype pairs for each sample, where each haplotype is labelled with a unique number. Haplotypes are assigned as either horned or polled, providing each haplotype with a polled probability based on the following criteria: (i) observed in polled animals with homozygous haplotypes; (ii) observed within progeny-tested animals (i.e. phenotyped progeny); (iii) observed in horned animals; (iv) observed in polled or scurred animals, where the other haplotype is horned. If the

* A joint venture of NSW Department of Primary Industries and the University of New England

haplotypes are not observed in any of these situations, then they cannot be assigned as horned or polled.

Samples from 278 animals from four different breeds (Angus, Santa Gertrudis, Brahman, and Droughtmaster) were phenotyped at dehorning (with photographic records) and microsatellite genotyped. Angus samples were sourced from breeder/s, and all others originated from the Repronomics™ project (Johnston *et al.* 2017). Genotypes were compared with the phenotypes and agreement or mismatch was quantified. Where a mismatch between the genotype and phenotype occurred, the haplotypes were investigated for potential bias in phenotype observations.

RESULTS AND DISCUSSION

Of the 278 samples sent for genotyping, 45 samples had incomplete microsatellite results (less than 10 markers). Microsatellite genotypes were obtained from 231 animals, consisting of 5 scurred, 14 polled, and 212 horned animals. Genotype estimations from the poll test had complete concordance with 221 phenotypes, such that:

- 5 samples with ≥90% PP microsatellite call matching polled phenotype;
- 6 samples with ≥90% PH microsatellite call matching polled and scurred phenotypes;
- 177 samples with ≥90% HH microsatellite call matching horned phenotype;
- 33 samples with 70-90% HH microsatellite call matching horned phenotype;

Six samples had a mismatch with the phenotype result (shaded orange in Table 1), and another four had low probability genotype estimations (i.e. <70%) due to haplotype uncertainty (shaded blue in Table 1). Haplotypes with poll probability of 0.01 are high likelihood of being horned, and are associated with high number of horned phenotypes. Those with a poll probability of 0.99 are high likelihood of being polled, and are associated with high number of polled phenotypes. Deviation from either end towards the centre (i.e. 0.5) represents a level of uncertainty in the assignment of polled or horned, and is most often due to variation in phenotype observations. Haplotypes suspected of mis-assignment/low certainty are shaded grey in Table 1. Phenotypes associated with these haplotypes are counted, shown in Table 2.

Table 1. Microsatellite poll results from mis-assigned/uncertain haplotypes. Orange shading indicates mismatching genotypes and probability (e.g. 96% PH); blue shading indicates low probability genotypes; grey shading indicates mis-assigned/uncertain haplotype

Breed	Phenotype	Haplotype 1	Haplotype 2	Haplotype 1 poll probability	Haplotype 2 poll probability	PP	PH	HH
Santa Gertrudis	horned	19	660	0.01	0.97	0.01	0.96	0.03
Santa Gertrudis	horned	22	660	0.01	0.97	0.01	0.96	0.03
Droughtmaster	horned	87	1655	0.01	0.85	0.01	0.84	0.15
Santa Gertrudis	horned	3	463	0.01	0.69	0.01	0.68	0.31
Angus	scurred	8	166	0.99	0.99	0.98	0.02	0
Angus	scurred	6	999	0.99	0.92	0.91	0.09	0
Droughtmaster	horned	254	383	0.2	0.38	0.07	0.43	0.5
Brahman	horned	135	771	0.01	0.38	0	0.39	0.61
Santa Gertrudis	horned	3	745	0.01	0.3	0	0.31	0.69
Droughtmaster	horned	254	1587	0.2	0.15	0.03	0.29	0.69

The phenotype counts for haplotypes driving mismatched genotypes (orange in Table 2) overwhelmingly show a bias towards polled phenotypes along with a significant number of unknown phenotypes which are uninformative. These phenotype counts explain the haplotype's high poll probability assignment of each mismatching genotype highlighted orange in Table 1.

The phenotype counts of less certain haplotypes are shown in blue in Table 2. Each such haplotype has a mix of contradicting phenotypes. The inclusion of varied and contradicting phenotypes leads to probability uncertainty and thus, low probability genotype estimations.

Table 2. Phenotype counts for haplotypes with mis-assigned poll probabilities (before additional samples submission). Orange shading indicates mis-assigned haplotypes; blue shading indicates uncertain probability haplotypes

Haplotype	Unknown	Scurred	Horned	Polled	Total
660	24	1	0	9	36
1655	1	1	1	3	8
463	4	2	0	1	8
166	8	0	2	7	17
999	2	0	0	5	7
254	17	5	1	5	28
771	3	2	3	0	8
383	40	1	9	17	68
745	10	1	3	8	22

Inclusion of more consistent phenotype observations will improve the certainty of the haplotype probabilities. An example of this is shown in Figure 1. Haplotype 383 had a poll probability of 0.38 due to contradicting horned and polled phenotypes (Table 2). Incremental inclusion of over 20 horned phenotypes saw the poll probability drop to approximately 0.01.

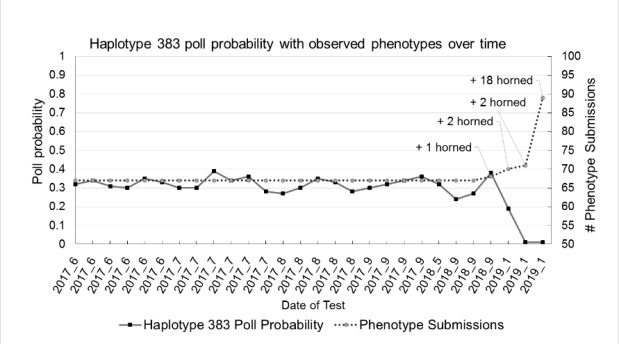

Figure. 1. Effect of phenotype submission over time on poll probability of haplotype 383

Table 3. Poll probability changes for less certain haplotypes (after additional horned submissions)

Haplotype	Poll probability before	Poll probability after	Poll probability change	No. horned additions
383	0.38	0.01	-0.37	22
771	0.38	0.31	-0.07	1
745	0.3	0.03	-0.27	5
254	0.2	0.1	-0.1	5

As a result of inclusion of more than 212 horned phenotypes, the less certain haplotypes from Table 2 have shifted poll probabilities significantly, shown in Table 3. These shifts towards zero poll probability are a direct result of the inclusion of horned phenotypes associated with these haplotypes. Unfortunately, further horned samples possessing haplotypes causing the mismatches in Table 1 could not be sourced; inclusion of further samples would be needed to adequately shift the poll probabilities of these haplotypes.

A shift in poll probability of some haplotypes may have happened historically at any point, and is a direct reflection of the reference data of the test. Reliable horned phenotypes are the most informative data as they exclude the possibility of being genetically polled. Submission of horned phenotypes is challenged in two major aspects. Firstly, it is difficult to ensure animals' phenotypes are accurate when (i) horns may be labelled as scurs and vice versa; (ii) horns may develop after the phenotyping time; and (iii) animals may be dehorned and mis-phenotyped polled. Secondly, data submission under commercial conditions makes submission of horned animals extremely unlikely; the cost of receiving a horned genotype result, when the horned phenotype is already known is unnecessary. Each of these aspects have likely impacted the observed sampling bias of the poll microsatellite test. As a result, some historical genotype predictions may be incorrect. This will likely become apparent using newly available technology, such as the commercial poll SNP test, which is now offered to the beef industry, where the microsatellite test will run in parallel. It is possible that the SNP test will provide a SNP result contradicting the microsatellite result, where the microsatellite haplotypes have been mis-assigned due to phenotype observations. It is impossible to know how many haplotypes have been affected, though reassuringly in this dataset, the microsatellite test had approximately 96% accuracy in relation to known phenotypes recorded.

CONCLUSIONS

This paper describes the effect of phenotyping bias on haplotype poll probabilities and resulting genotype estimations for the poll microsatellite test. This dataset had 96% genotype to phenotype concordance. The remaining four percent was demonstrated to be a result of haplotype mis-assignment due to associated phenotype observations, which can be corrected with additional horned phenotype submissions.

ACKNOWLEDGEMENTS

This research includes data generated by the BeefCRC. This research is supported by Meat and Livestock Australia (MLA) projects L.GEN.1803 and L.GEN.0174.

REFERENCES

Connors N.K. and Tier B. (2016) In 'Breeding Focus 2016 - Improving Welfare', pp. 59. editors S Hermesch, S Dominik. Animal Genetics and Breeding Unit, University of New England, Armidale, NSW.

Connors N.K., Tier B. and Johnson D.J. (2016) *Proc. Northern Beef Research Update Conference.* 196.

Connors N.K., Tier B. and Johnston D.J. (2018) *Proc. World Congr. Genet. App. Livest. Breed.* **11:** 344.

Henshall J.M., Piper E.K. and McDonald A. (2011) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **19:** 79.

Johnston D.J., Grant T.P., Schatz T.J., Burns B.M., Fordyce G. and Lyons R.E. (2017) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **22:** 385.

Piper E.K., Tier B. and Henshall J.M. (2014) *Proc. World Congr. Genet. App. Livest. Breed.* **10:** 237.

Prayaga K.C. (2007) *Aust. J. Agric. Res.* **58:** 1.