GENOMIC TOOLS FOR USE IN THE NEW ZEALAND DEER INDUSTRY

K.G. Dodds, S-A.N. Newman, S.M. Clarke, R. Brauning, A.S. Hess, T.P. Bilton, J.F. Ward, A.J. Chappell, J.C. McEwan, T.C. Van Stijn, M. Bates and S.J. Rowe

AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand

SUMMARY

The New Zealand deer industry has recently adopted genotyping-by-sequencing (GBS) as a tool for parentage analysis. For cost reasons, sequencing is performed on many individuals at once with low sequencing read depth supporting the genotypes. It is important to account for the partial information provided by these low depth reads and to account for the high genetic diversity between breeds present in the population in any analysis. The genomic information provided by the more than 70,000 markers scored can also be used for additional purposes such as inbreeding and relatedness estimation, plus gender and breed prediction. The data provides a platform for genome wide association studies and genomic selection, which are being developed for this industry. These results provide evidence that GBS is a useful technique for genomic studies.

INTRODUCTION

The New Zealand (NZ) deer industry has been using DNA-marker testing since the early 1990s. This has been primarily for parentage assignment, as deer behaviour prevents manual recording of pedigree at birth. DNA markers have also been used to provide information about breed. The primary breeds are wapiti or elk (*Cervus canadensis*) and red deer (*Cervus elaphus*), which are regarded as distinct species, but there is also interest in estimating the components of differing European origin (Eastern or Western) in red deer. Initially a small panel of isozymes was used for breed discrimination. This was subsequently replaced by a microsatellite panel. Since 2017 genotyping-by-sequencing (GBS) has been used as the marker system. We show how recently developed methods for low-depth GBS data are being used in the New Zealand deer industry for parentage, breed prediction and gender assignment and consider how this GBS resource can be used for gene discovery and genomic selection.

MATERIALS AND METHODS

Animals. The Invermay, AgResearch deer herd is used to illustrate the use of genomics in the NZ deer industry. The 2018 cohort consisted of 554 genotyped calves, 621 potential dams and 46 potential sires. Industry-wide data used here refers to deer genotyped by GenomNZ (https://www.agresearch.co.nz/genomnz) using GBS, first used for the 2016 calf-drop and their parents. This industry GBS dataset currently contains ~80,000 animals.

GBS genotypes. The animals were genotyped by GBS using the methods described by Dodds *et al.* (2015). The resulting sequence reads from a set of animals likely to represent much of the genetic variation were adapter-trimmed and then UNEAK (Lu *et al.*, 2013) was used to detect variants (without the use of a reference genome). These variants were placed into a catalog which was used to report counts of reference and alternate alleles for each variant and sample (including any subsequently sequenced samples) using TagDigger (Clark and Sacks 2016). Each new set of GBS count data is compared against any previous results for the same sample, by comparing the relatedness, estimated taking into account the read depths (Dodds *et al.* 2015), between a pair of results for the same sample with the mean self-relatedness of those samples. Differences greater than 0.4 are reported for checking. Accepted results are then appended to the file of previous results. There is a corresponding comparison made during downstream analysis for any pairs of samples that have come from the same animal.

Population structure and breed prediction. The genetic structure of the population was portrayed as the principal components of the genomic relationship matrix (GRM), which in turn was calculated using the method of Dodds *et al.* (2015) which is based on VanRaden's (2008) first method, but accounts for the read depth in a genotype call. The GRM was calculated for a random sample of approximately 5000 deer from NZ commercial samples supplemented with NZ and overseas samples of reputedly pure 'breed' (wapiti/elk, English red and Eastern European red, denoted 'Wapiti', 'English' and "Eastern', respectively) standards. Breed prediction was undertaken by regressing the observed proportions of A alleles at each SNP for an animal on each breed's allele frequency (Kuehn *et al*. 2011). The breed allele frequencies were calculated from the breed standards.

Gender prediction. Gender is predicted using the method of Bilton *et al.* (2019) using the proportion of Y chromosome SNPs with reads and the heterozygosity of X chromosome SNPs. There were 15 SNPs located on the Y chromosome and 1006 SNPs located on the non-pseudoautosomal region of the X which passed the criteria given in Bilton *et al.* (2019).

Parentage analysis. Parentage assignment is based on the methods of Dodds *et al*. (2019) with the highest related potential sire and dam were assigned provided they achieved the chosen thresholds. The thresholds used for assigning parentage were 0.3 for estimated relatedness (from the GRM), 0.015 for parent-offspring excess (raw minus expected, where expected rate is calculated for the given read depths and offspring genotype) mismatch rate (EMM) and 0.03 for trio EMM.

RESULTS AND DISCUSSION

The GBS process resulted in a catalog of calls for 74,798 SNPs. After filtering SNPs for a Hardy-Weinberg disequilibrium coefficient (proportion of animals observed as homozygous for the A allele minus the squared A allele frequency; calculated using the Industry dataset) greater than -0.05 and a minor allele frequency (calculated for Invermay dataset) greater than 0.01, there were 66,824 SNPs remaining. These SNPs had a 76.9% call rate and mean read depth of 3.37 in the Invermay dataset.

Population structure and breed prediction. The first two principal components of an analysis of 6269 deer (109 breed standards, 1,211 Invermay herd deer, 4949 randomly chosen) is shown in Figure 1. The first component explains 80% of the variance and reflects the large genetic difference between wapiti and red deer which are at opposite ends of this axis. The second component explains 14% of the variance, and English and Eastern deer occur at the opposite ends of this axis. The Invermay deer mainly occur in a continuum between these two red deer types, with a few plotting part-way towards the Wapiti group, suggesting some wapiti ancestry in those animals.

The Invermay animals were predicted to be an average of 58% Eastern, 39% English and 3% Wapiti. The range in predicted breed percentages in the progeny were 4-91% Eastern, 7-96% English and 0-19% Wapiti. An estimated breed proportion could be used for a national across-breed genetic evaluation, but proportions estimated by different methods (marker systems and pedigree) need to be consistent.

Gender prediction. The results of the gender prediction for the Invermay herd are shown in Figure 2. The mean read depth of the sex chromosome SNPs in this herd was 2.97. The parents matched their recorded gender (apart from one uncertain), as expected. For the calves, three that were recorded as male were predicted as female, while one recorded as female was predicted as male. One of these recorded males was subsequently corrected to female, but the other inconsistencies could not be checked (sold or died). The gender test can be made at the same time as a parentage analysis and provides a check for assigned gender which can be difficult to assign in the field with 100% accuracy in young calves.

Parentage analysis. Both parents were assigned for 535 calves, 17 calves were assigned to a dam only, one was assigned a sire only and one was not assigned either parent (excluded based on

the trio EMM which was 0.05). A seemingly low threshold (0.3) was used for assigning parentage to accommodate variations in breed structure and the fact that the GRM used allele frequencies estimated from the same dataset which can depress relatedness estimates (Yang *et al.* 2010). Only four of the final sire or dam assignments were with estimated relatedness below 0.4. One assignment had relatedness to both parents less than 0.4 and in this case the sire was predominantly (81%) Eastern while the dam was predominantly (82%) English.

A GRM of the dam only progeny, visualised using the heatmap function in R (Figure 3), suggested that two or perhaps three different sires were involved. Such information could help to find additional sires to include in the analysis.

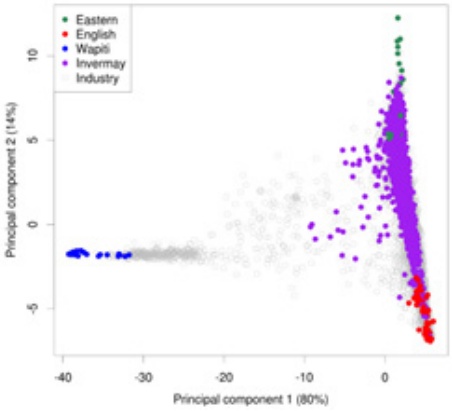

Figure 1. Principal components plot of the Invermay herd, breed standards and a random industry set of 5000 deer

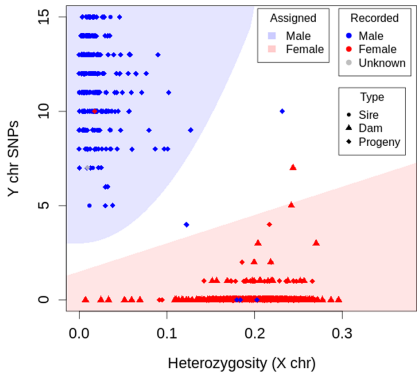

Figure 2. Gender plot with number of Y chromosome SNPs with reads plotted against heterozygosity of X chromosome SNPs. The lower and upper shaded areas are predicted as females and males, respectively

A by-product of calculating a GRM is that estimates of inbreeding are available (self-relatedness minus 1). The distribution of these estimates is shown in Figure 4. As is the case of most genomic estimators of inbreeding, values outside of [0,1] are possible. The progeny with both parents assigned with relatedness less than 0.4 had estimated inbreeding of -0.3, reflecting the high genetic separation between its parents. Reporting inbreeding estimates to breeders will alert them to issues with their breeding programme if high estimates are present, however care is needed to help breeders understand these values compared to pedigree-based calculations which are always within [0,1].

Future directions. The use of GBS in the NZ deer industry has enhanced the information available to breeders compared with that from marker systems previously used. Parentage, breed and estimated inbreeding results are returned to the breeders, but there is no systematic way of returning genomic relationships to breeders or service providers to allow enhanced breeding plan designs (e.g. optimal contributions). Further opportunities are available, such as the use of this genomic information for genome-wide association studies and genomic selection. Some initial investigations have been made by Rowe *et al.* (2017), including the consideration of calculating appropriate GRMs with GBS data in a multi-breed context. These or other methods need to be tested for their feasibility in the full industry dataset, and the data from non-genotyped animals included in the analysis. As the deer industry is much smaller than the dairy, beef and dual-purpose sheep industries, it will need to learn from the use of genomic information in those industries to allow affordable implementation.

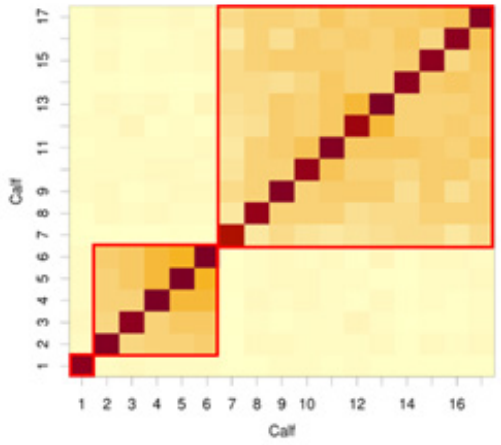

Figure 3. Heatmap of the relatedness between calves with only a dam assigned. The red boxes group potential sire groups

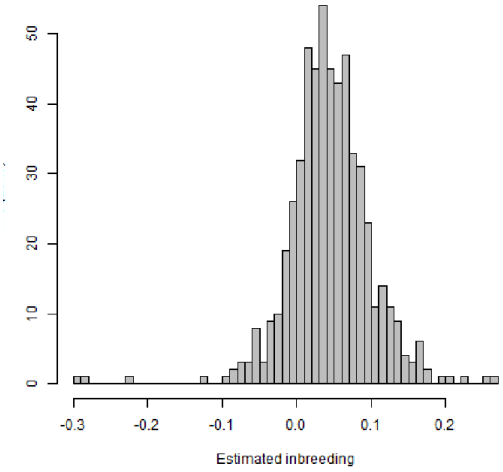

Figure 4. Distribution of estimated inbreeding in the Invermay progeny

ACKNOWLEDGEMENTS

This work was funded by DEEResearch and the Ministry of Business, Innovation and Employment (NZ), through the "Genomics for Production & Security in a Biological Economy" programme.

REFERENCES

Bilton T.P., Chappell A.J., Clarke S.M., Brauning R., Dodds K.G., McEwan J.C. and Rowe S.J. (2019) *Anim. Genet.* (in press).

Clark L.V. and Sacks E.J. (2016) *Source Code Biol. Med.* **11**:1.

Dodds K.G., McEwan J.C., Brauning R., Anderson R.A., Van Stijn T.C., Kristjánsson T. and Clarke S.M. (2015) *BMC Genomics* **16**:1047.

Dodds K.G., McEwan J.C., Brauning R., Van Stijn T.C., Rowe S.J., McEwan K.M. and Clarke S.M. (2019) *bioRxiv* 582585.

Kuehn L.A., Keele J.W., Bennett G.L., McDaneld T.G., Smith T.P.L., Snelling W.M., Sonstegard T.S. and Thallman R.M. (2011) *J. Anim Sci.* **89**:1742.

Lu F., Lipka A.E., Glaubitz J., Elshire R., Cherney J.H., Casler M.D., Buckler E.S. and Costich D.E. (2013) *PLoS Genet.* **9**:e1003215.

Rowe S.J., Dodds K.G., Ward J.F., Asher G.W., Brauning R., McEwan J.C. and Hess A. (2018) *Proc. 11th World Congress Genet. Appl. Livest. Prod.*:432

VanRaden P.M. (2008) *J. Dairy Sci.* **91**:4414.

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R., Madden P.A., Heath A.C., Martin N.G., Montgomery G.W., Goddard M.E. and Visscher P.M. (2010) *Nat. Genet.* **42**:565.