BRIEF COMMUNICATION: Developing genomic tools in the New Zealand Deer Industry

SJ Rowe*, SM Clarke, TC Van Stijn, DL Hyndman, JF Ward, KM McEwan, KG Dodds, JC McEwan, S-AN Newman and GW Asher.

AgResearch, Invermay Agricultural Centre, Private Bag 50034, Mosgiel, New Zealand.

**Corresponding author.* Email: suzanne.rowe@agresearch.co.nz

Keywords: deer; parentage; breed assignment; genomics.

Introduction

The New Zealand deer industry currently uses genomic technology via microsatellite markers to assign parentage and breed. In many other species, microsatellites have been superseded by SNP markers, which are more abundant in the genome and technically easier to use. SNP assays or 'chips' that record thousands to hundreds of thousands of markers have now been developed in many species including humans, pigs and cattle.

A SNP genotyping assay was developed for deer (*Cervus* genus) by comparing genome sequence data from red deer & wapiti (sub species of *Cervus elaphus*) and sika (*Cervus nippon*) (Brauning et al. 2015). The assay contains 54 000 loci expected to differ or be 'polymorphic' in Scottish and NZ reds; 2,250 are expected to have different fixed alleles between red and sika and 2,250 expected to have different fixed alleles between red and wapiti, giving a total of 58,500 loci. Using this new platform, a deer genomics project was initiated to collate samples required to identify markers that could be used in the New Zealand deer industry for i) parentage and ii) to collect breed standards that could be used to differentiate primarily wapiti and red genomics within New Zealand and iii) for traceability purposes to differentiate amongst sub-species of deer that might be used for product substitution in the global venison market. Because the current chip was developed in red, wapiti and sika deer, it is unknown whether the markers will be polymorphic in other sub-species/genera of interest. To mitigate this, the project also looked at using newly developed SNP discovery methods to find additional or complementary markers.

Materials and methods

DNA was extracted from blood, tissue, semen, velvet, bone and antler. Processed samples were hybridised to the Illumina 50K CervusSNP50, 24-sample Bead Chips (Illumina, San Diego). The fluorescent-stained alleles were detected by an Illumina iScan reader and the data collected in Genome Studio Analysis Software.

Four hundred and thirty-two mixed species male and female samples were genotyped including controls. Samples were removed for contamination, low call rates (<96%) and duplication. Some samples in the *Cervus* genus (axis, fallow, mule, rusa, sambar and white tail) were not removed if call rates were >90% because the low call rate was deemed a result of species divergence, not failure of the sample. Quality control of genotypes was carried out using the GenABEL package in R software (Aulchencko, 2007). After exclusion of duplicates, low call rates (a single SNP called in less than 95% of the population), non-autosomal and non-Mendelian inheritance patterns, the final data set contained 44,448 SNP genotypes on 396 individuals.

For parentage, the aim was to identify a minimum of 100 SNPs highly polymorphic in the New Zealand deer (*Cervus genus*) industry. Only markers with a minor allele frequency >0.4, and a call rate >0.98, in red deer were selected resulting in a subset of 8,690 SNPs. These criteria were applied in turn to the Canadian wapiti, NZ wapiti (a red wapiti composite) and wapiti x red deer populations.

To determine SNPs informative for breed assignment, each individual was first assigned to a genetic group using all SNPs. Genetic groups were based on genomic relationship, i.e., proportion of SNP genotypes shared in common. The number of genetic groups in the data was modelled using multidimensional scaling and the mclust clustering package in R software (Fraley, 2012). SNPs were ranked according to information content using the ORCA algorithm (Rosenberg et al., 2003). Using multi-dimensional scaling and the software package STRUCTURE (Pritchard, 2000), subsets of SNPs were evaluated to determine the minimum number required to correctly assign genetic group or breed.

Results

After applying the described criteria, 136 high-quality and highly polymorphic SNPs were identified for use as a deer parentage panel. Table 1 gives marker statistics for these SNPs in all individuals and in each group of individuals. Heterozygosity is extremely high in the main groups underpinning the New Zealand industry. Each of the parent offspring pairs (n=166) or trios (n=18) were tested for mismatches using all SNPs and the 136 SNPs identified. For one pair there were less than 0.5% mismatches. For two animals there were greater than 7% mismatches, suggesting incorrect parentage assignment and this was reported back to the DNA testing laboratory. After eliminating these individuals, one SNP in the putative 136 parentage SNPs had a single mismatch and another SNP had nine mismatches. Both of these SNPs were subsequently excluded from the panel. The markers were also evaluated for proximity and two further SNPs were discarded as they were very close to other SNPs in the set when the positions were based on the bovine reference genome build UMD 3 (Figure 2) and were, therefore, extremely likely to be inherited together. The final putative parentage panel consisted of 132 SNPs.

Table 1 Descriptive statistics for 136 SNPs selected for parentage in deer. Proportion of SNPs grouped by minor allele frequency(Maf), call rates and mean heterozygosity are given for all animals and for each of 9 genetic sub-groups.

Genetic group	n	Maf >0.1	Maf >0.2	Maf >0.3	Maf >0.4	Maf >0.45	Call rate>0.99	Het (SNP)	het (ind)
All individuals	396	1	1	1	0.75	0.46	0.87	0.49	0.42
red deer	202	1	1	1	1	0.43	0.99	0.49	0.48
Canadian wapiti	21	1	1	0.99	0.99	0.55	0.96	0.49	0.47
NZ wapiti	11	1	1	1	1	0.65	1	0.49	0.51
wapiti x red	62	1	1	1	1	0.57	0.99	0.49	0.51
Iberian	39	0.94	0.79	0.56	0.31	0.13	0.99	0.39	0.40
maral	4	98	90	70	0.28	0.28	1	0.43	0.52
sika	11	0.11	0.01	0	0	0	0.96	0.06	0.04
Père David	6	0.86	0.68	0.47	0.16	0.05	0.99	0.36	0.47
white-tail/ sambar/ mule	40	0.16	0.10	0.08	0.02	0	0.73	0.09	0.04

Maf is the minor allele frequency in the population, i.e., the proportion of time that the less-common allele is identified. Call rate is the proportion of animals with a call rate greater than 0.99. Het is heterozygosity, i.e., the proportion of genotypes that are heterozygous or show both alleles. For a marker, heterozygosity is calculated over all individuals and for an individual, heterozygosity is estimated over all markers.

Figure 1 Genetic grouping when clustered by multidimensional scaling of genomic relationship matrix and coloured by recorded breed. Legend gives the major recorded species or sub-species present in each group, White-tailed includes sambar and mule deer. PD are Père David's deer, and red are red deer from the UK, Europe and New Zealand.

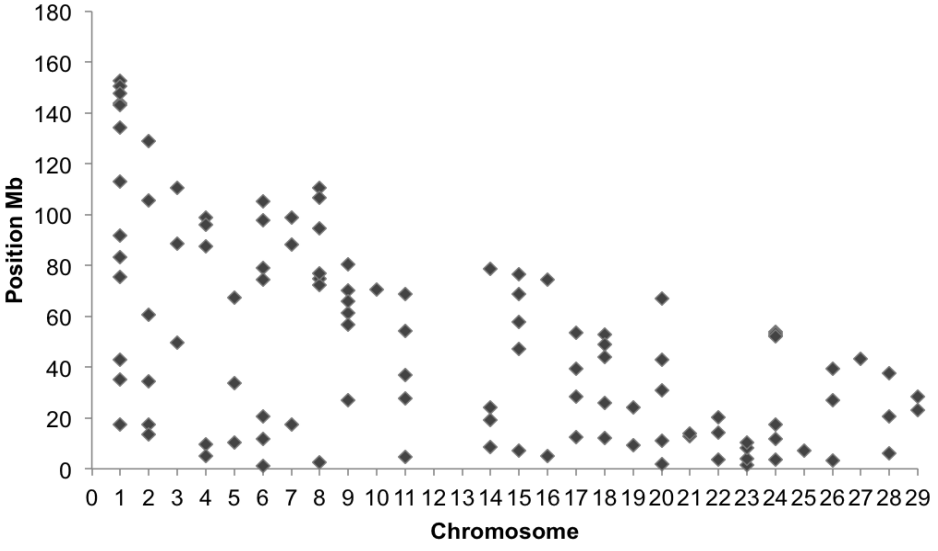
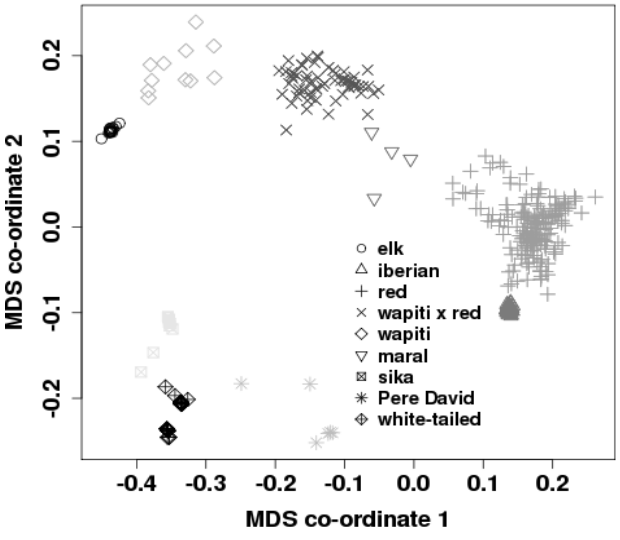


Figure 2 Genomic position of putative parentage panel showing that markers are spaced across the genome. Marker positions are based on where they map to the bovine reference genome build UMD 3.0.



For species and sub-species assignment, the best statistical fit for the data based on genomic relationship was nine clusters or genetic groups (Figure 1). Grouping by genotype alone agreed well with recorded species. The DPT animals that were NZ wapiti x red all clustered half way between the red and the NZ wapiti group. NZ wapiti clustered close to but separately from, imported samples of Canadian wapiti. Iberian red clustered tightly and apart from NZ red. The eastern european animals also tended to cluster separately. Sub-species evolutionary distinct from the *Cervus elaphus* subspecies tended to cluster together, because there was less information from the SNPs to differentiate them. Assigning samples to genetic groups or sub-species was successful with 1000 randomly selected SNPs. Ranking SNPs for information content did not appear to significantly outperform selection of SNPs at random and requires further investigation.

Discussion

The Illumina cervine SNP assay, developed as an international collaboration has clear utility in the New Zealand deer population. The identification of markers segregating in the population offers new opportunity to the deer industry. A parentage panel was identified for final validation containing 132 SNP markers. Markers selected for parentage, had call rates >0.99 and were highly polymorphic (maf>0.4) in NZ wapiti and red deer, i.e. segregated within the breeds that underpin the

New Zealand deer industry. This compares well with other such panels. The current parentage panel for sheep consists of 84 markers, was evaluated in 79 breeds, and has 60% of SNPs with a Maf >0.4 (Clarke et al. 2014).

For deer, markers were used to segregate the main genetically and geographically distinct groups. Assigning samples to genetic groups was successful with 1000 randomly selected SNPs, however, these are unlikely to perform as well for animals genetically distinct from the *Cervus elaphus* subspecies. This is not an issue within the New Zealand population but additional markers may be required for international use. Low-coverage sequencing methods could provide future opportunities to identify further markers. Opportunities include parentage, breed assignment to improve accuracy of breeding value estimation, sample traceability, genomic tests for traits of economic importance, and genomic breeding values.

Acknowledgements

We are very grateful to Josephine Pemberton from the University of Edinburgh, UK for supplying the cervine SNP chips; breeders who supplied samples and information; international collaborators – Dr Tomás Landete-Castillejos, Vice-President of Federation of European Deer Farmers Associations, Helen Schwantje, Ministry of Forests, Lands and Natural Resources Operations, BC, Canada, David Coltman, University of Alberta, Murray Woodbury, University of Saskatchewan, and Colin Coros, Delta Genomics. We would like to thank the deer progeny test for providing trios and families. The genotyping by sequencing work was done in collaboration with FarmIQ and a current MBIE program for Food production and security.

References

Aulchenko YS, Ripke S, Isaacs A, Van Duijn CM 2007. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, 23:1294–1296.

Brauning R, Fisher PJ, McCulloch AF, Smithies RJ, Ward JF, Bixley, MJ, Lawley CT, Rowe SJ, McEwan JC 2015. Utilisation of high throughput genome sequencing technology for large scale single nucleotide polymorphism discovery in red deer and Canadian elk G3:Genes Genomes Genetics.

Clarke SM, Henry HM, Dodds KG, Jowett TWD, Manley TR, Anderson RA, McEwan JC 2014 A high throughput single nucleotide polymorphism multiplex assay for parentage assignment in New Zealand sheep. PloS ONE 9(4).

Fraley C, Raftery A, Murphy B 2012. MCLUST Version 4 for R: Normal mixture modelling for model-based clustering, classification, and density estimation, technical report no. 597, Department of Statistics, University of Washington, June 2012.

Pritchard JK, Stephens M, Donnelly P 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Rosenberg N, Li L, Ward R, Pritchard J 2003. Informativeness of genetic markers for inference of ancestry. American Journal of Human Genetics 73: 1402-1422.