Proc. Assoc. Advmt. Anim. Breed. Genet. 22:437-440

SNP-PANEL DESIGN FOR DAIRY PROPORTION ESTIMATION AND PARENTAGE TESTING

E.M. Strucken[1], C. Esquivelzeta-Rabell[2], H.A. Al-Mamun[1], C. Gondro[1], O.A. Mwai[3] and J.P. Gibson[1]

[1] School of Environmental and Rural Science, University of New England, NSW, Australia
[2] Pic Improvement Company (PIC), Genetic Services, Hendersonville, USA
[3] International Livestock Research Institute (ILRI), Nairobi, Kenya

SUMMARY

The selection of small numbers of SNPs to analyse population features is an important task in the livestock industry. Populations differ in their genetic architecture, which often requires the selection of population specific SNPs. Different tasks, such as breed proportion prediction or parentage testing, also require specific panels. We tested which selection methods are best for breed proportion estimation and parentage testing in a crossbred dairy population from East Africa. We selected SNPs from a 735k SNP panel (Illumina) based on several methods: **a)** high minor allele frequencies; **b)** high allele frequency differences between ancestral populations; **c)** at random; **d)** with a differential evolution algorithm. Estimates of breed proportions in the subsets were tested against *true* breed proportions based on all 770k SNP obtained from ADMIXTURE. Parentage assignments was based on opposing homozygotes. Panels selected for largest allele frequency differences in ancestral populations gave best results for breed proportion predictions and panels selected for highest minor allele frequency gave best parentage resolution.

INTRODUCTION

The selection of small numbers of SNPs to carry out a variety of genomic test is at the forefront of the livestock industry. Challenges for small SNP panels are the accurate prediction of breed proportions and the assignment of parentages. Knowledge about breed proportions is important to the livestock sector for quality trait marks (*e.g.* Wagyu) but also for breeding decisions, especially in crossbred population. Whilst pure breeds such as Holstein, Jersey, or Wagyu are mostly used in industrialised settings, crossbreds find their application in developing countries where one animal must fulfil multiple purposes (*i.e.* milk and meat). To improve crossbreds, their breed proportion must be determined to choose the best breed or animal for mating. Similarly, assigning parentages is important in the livestock industry, as the pedigree determines factors such as inbreeding, breeding value estimation, or a tracking of agricultural goods. Again, in industrialized settings, record keeping of pedigrees is common practice whilst in developing countries accurate pedigrees are often missing.

Both breed proportion prediction and parentage assignment can be carried out on the basis of genomic information. In theory, however, to accurately predict breed proportions in a crossbred animal, a prior knowledge based on trading history and breeding preferences is required to determine the most likely ancestral breeds. The genomic information of these ancestral breeds is then traced within the crossbred animals. To distinguish the different genomic footprints of the ancestral breeds, it is favourable if the ancestral breeds are genomically different from each other. This should lead to a large allele frequency range of selected markers in the crossbred population.

For parentage assignment, most tests rely on the likelihood that a parent-offspring pair shows the same genotype. Simpler tests only consider homozygous genotypes, especially if only one parent is known. Opposing homozygotes describe the occurrence of a parent displaying one homozygous genotype whilst the offspring displays the other homozygous genotype (Hayes 2011). The more opposing homozygotes are found between two animals, the less likely it is that they are

a parent-offspring pair. The highest likelihood, according to Hardy-Weinberg, to observe opposing homozygotes in a population is given for markers with high minor allele frequency. Thus, both breed proportion prediction and parentage assignment depend on different qualities of SNPs.

In this study, we used different selection methods to choose small panels of SNPs (100 to 1500 SNPs) from a 735k panel to determine breed proportions and parentages in a crossbred dairy population of East Africa. Based on the crossbreeding history in Kenya and Uganda (Rege and Tawan 1999; Hanotte *et al.* 2000), an African *Bos taurus* and a *Bos indicus* reference breed as well as 5 European dairy breeds were chosen to determine breed proportions.

MATERIALS AND METHODS

A total of 1,933 crossbred dairy cows from Kenya and Uganda and local indigenous breeds of Ankole (n=43), Nganda (n=17), and Small East African Zebu (Zebu; n=58) were sampled (Dairy Genetics East Africa, DGEA1, project)**.** Additionally, genotypic datasets for N'Dama (as the reference African *Bos taurus* breed; n=20), Nelore (as the reference *Bos indicus* breed; n=20), Guernsey (n=20), Holstein (n=20), and Jersey (n=20) were sourced from the International Bovine HapMap consortium. Further, British Friesian (n=25) from the SRUC in Scotland and Canadian Ayrshire (n=20) from the Canadian Dairy Network (CDN) were used as reference breeds.

All animals were genotyped with the 770k BovineHD Beadchip array (Illumina Inc., San Diego, CA, USA). Genotypes were filtered using *SNPQC* (Gondro *et al.* 2014) with a sample-wise call rate of 90%; a median GC score <0.6; and a GC score<0.6 in at least 10% of the samples. Only markers contained on the 29 autosomal chromosomes were included in the analysis. The cleaned population datasets were merged and included 735,293 SNPs. Markers that were excluded due to quality control criteria in one breed but not in another were set to NA in the breed for which they were excluded.

True breed proportions of the crossbreds were estimated using the full quality controlled data in the ADMIXTURE 1.23 program (Alexander *et al.* 2009). The analysis was supervised with N'Dama, Nelore, Ayrshire, Friesians, Guernseys, Holstein, and Jerseys as assumed ancestral populations.

The pedigree of the crossbreds was reconstructed based on presence or absence of opposing homozygotes (Hayes 2011) and contained 171 cows with 189 offspring, of which 15 cows had two and one cow had three offspring. Parentage testing was based on opposing homozygotes and panel resolution determined based on the separation value (Strucken *et al.* 2014).

Subsets of SNPs ranging from 100 to 1,500 markers were selected based on **a)** highest minor allele frequency in the crossbreds, **b)** absolute allele frequency difference between the ancestral breeds (European dairy breeds vs. a combination of Nelore and N'Dama), at **c)** random (results were averaged across 10 random samples), and **d)** a differential evolution algorithm (Gondro *et al.* 2013, Esquivelzeta *et al.* 2015). Accuracies of dairy proportion prediction were assessed with the coefficient of determination (r^2) between the *true* proportions and the estimated proportions from the subsets. Parentage assignment was assessed with the *separation value* which is based on opposing homozygotes (Strucken *et al.* 2014, 2016).

RESULTS AND DISCUSSION

Allele frequencies showed relatively large interquartile ranges for all selection methods (0.35-0.65) apart for highest MAF (Figure 1). Allele frequencies of SNP subsets were assumed to play a major role for their performance in breed proportion prediction and parentage assignment. Markers with largest allele frequency differences between ancestral breeds should be able to distinguish breed proportions in crossbred animals. Therefore, allele frequencies were expected to show a larger variation in the crossbreds. Markers with a high minor allele frequency, i.e. both alleles occur equally often, have the highest probability to show opposing homozygotes between two

unrelated individuals, therefore should work best for parentage assignment.

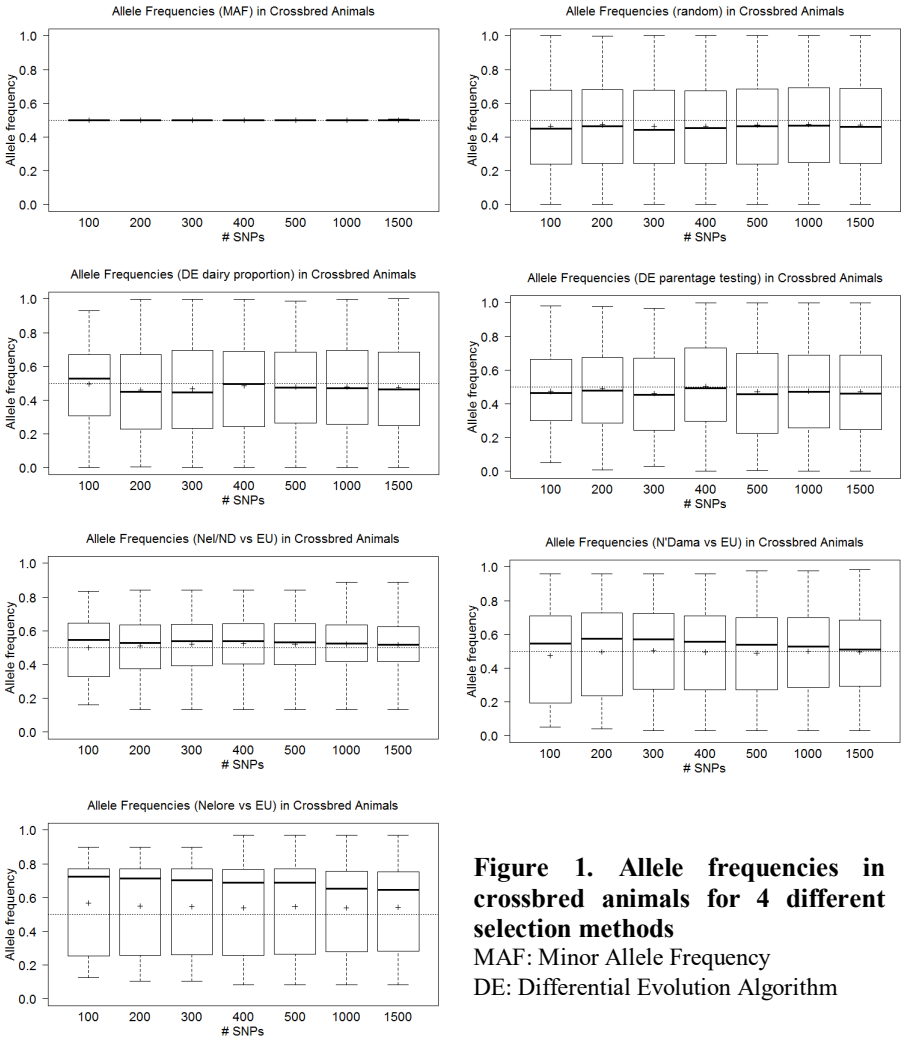

Figure 1. Allele frequencies in crossbred animals for 4 different selection methods
MAF: Minor Allele Frequency
DE: Differential Evolution Algorithm

Individual breed proportion estimates of the ancestral breeds proved to be highly variable depending on the number of assumed ancestral breeds. Therefore, we used the total proportion of European dairy breeds as a more reliable contrast to the African N'Dama and indicine Nelore. Dairy breed proportions of the crossbred animals were on average 0.7 (SD 0.21).

The various panels predicted total dairy breed proportions with an r^2 of 0.694-0.950 (SE 0.005-0.013) for the smallest subsets of 100 markers (Figure 2a). The best results for all panel sizes was achieved with SNPs selected for largest absolute allele frequency difference between the ancestral breeds.

Lowest numbers of opposing homozygotes were found for panels selected for high minor allele frequency (Figure 2b), thus should perform best for parentage assignments. With 100 markers, however, none of the selection methods resulted in a panel that was able to assign all parentages correctly, as this requires a separation value >0.

All panels that were selected based on the Kenyan and Ugandan crossbred animals were validated in independent crossbred populations of Ethiopia and Tanzania (N=545, N=462). The

selection methods rank similarly in the validation populations with the panels selected for largest allele frequency differences between Nelore/N'Dama and the EU dairy breeds performing best for breed proportion prediction and the panels selected for highest minor allele frequency in the Kenyan and Ugandan crossbreds performing best for parentage assignment.

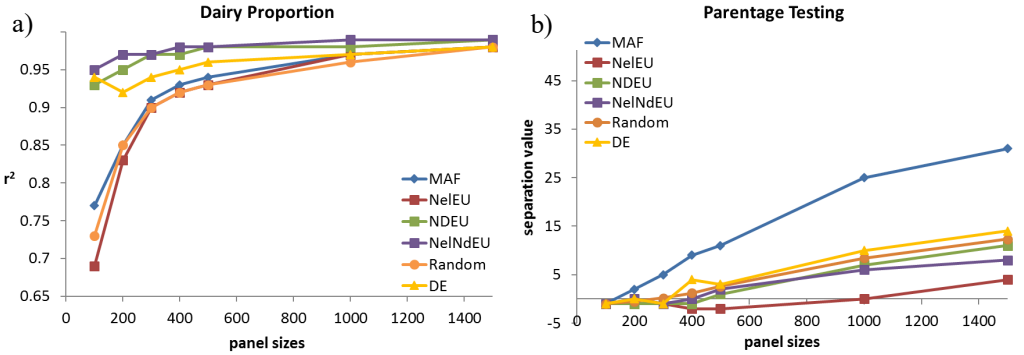

Figure 2.a) Accuracy (r^2) of dairy proportion prediction and b) parentage resolution (separation value) of SNP subsets from 4 different selection methods

MAF: Minor Allele Frequency; NelEU: Nelore vs EU; NDEU: N'Dama vs. EU; NelNdEU: combined Nelore and N'Dama vs EU; DE: Differential Evolution Algorithm

A combination of panels performing best for breed proportion prediction and parentage assignment performed poorer than the individual panels with same number of SNPs. Further, it showed that breed proportion prediction mainly depends on allele frequencies, i.e. the difference between allele frequencies in ancestral breeds, and the ability to assign parentages is mainly limited by the number of markers (Strucken *et al.* 2016).

ACKNOWLEDGEMENT

We acknowledge the Bill and Melinda Gates foundation. HAM and CG acknowledge the Next-Generation BioGreen 21 Program (PJ01134906), Rural Development Administration, Republic of Korea, and the Australian Research Council (DP130100542).

REFERENCES

Alexander D.H., Novembre J. and Lange K. (2009) *Genome Res.* **19**, 1655.

Esquivelzeta-Rabell C., Al-Mamun H.A., Lee S.H., Lee H.K., Song K.D. and Gondro C. (2015) *Association for the Advancement of Animal Breeding and Genetics Conference (AAABG)*, Lorne, Australia.

Gondro C., Porto-Neto L.R. and Lee S.H. (2014) *Anim. Genet.* **45**, 758-61.

Gondro C. and Kwan P. (2013) In '*Bioinformatics: Concepts, Methodologies, Tools, and Applications'*, pp. 105, IGI Global.

Hayes B.J. (2011) *J. Dairy Sci.* **94**, 2114.

Hanotte O., Tawah C.L., Bradley D.G., Okomo M., Verjee Y., Ochieng J. & Rege J.E. (2000) *Mol. Ecol.* **9**, 387.

Rege J.E.O. & Tawan C.L. (1999) *Anim. Genet. Resources Information Bulletin* **26**, 1.

Strucken E.M., Gudex B., Ferdosi M.H., Lee H.K., Song K.D., Gibson J.P., Kelly M., Piper E.K., Porto-Neto L.R., Lee S.H. and Gondro C. (2014) *Anim. Genet.* **45**, 572.

Strucken E.M., Lee S.H., Lee H.K., Song K.D., Gibson J.P. & Gondro C. (2016) *J. Anim. Breed Genet.* **133**, 13.