Proc. Assoc. Advmt. Anim. Breed. Genet. 22:421-424

A SIMPLE METHOD FOR EVALUATING THE GENOTYPE QUALITY OF THE SIRE X CHROMOSOME USING HALF-SIB FAMILIES

M.H. Ferdosi[1], D. Johnston[1], H. A. Al-Mamun[2] and C. Gondro[2]

[1] Animal Genetics and Breeding Unit, University of New England, Australia
[2] School of Environmental and Rural Science, University of New England, Australia

SUMMARY

Recent studies have shown that the bovine X chromosome contains more than a thousand genes, some of which may be economically important. However, since males are heterogametic for the X chromosome, it has, to date, largely been ignored for genomic prediction and its information content has not been fully explored. The genotyping quality of the X chromosome is the first question that must be addressed. In this study, we suggest a simple method to impute the X chromosome of the sire using a half-sib family in order to check genotyping accuracy. The results showed that the suggested method allows for a robust imputation of the X chromosome in ungenotyped sires and is useful for the routine quality control of the genomic data.

INTRODUCTION

Chromosome X contains more than a thousand genes and it is the second largest chromosome in the bovine genome (Su *et al.* 2014). In most genomic prediction applications, the X chromosome is ignored as it requires different algorithms and methods to become useful (Sargolzaei *et al.* 2014; Su *et al.* 2014). Recent studies have shown that there are some genes in the X chromosome that may be economically important (Richardson 2016). The first step in genomic prediction is to evaluate the quality of genotyping. Therefore, it is important to check the genotyping quality of the X chromosome before any further analyses. Previous studies (Ferdosi *et al.* 2014a; Ferdosi *et al.* 2014b) have shown that the sire imputation accuracy from half-sib family genotype data is very high and that the imputed sire can be used to measure genotyping quality. However, we require a different method of sire imputation for the X chromosome.

The sex chromosomes in bovine males consist of an X and Y chromosome with a small region of homology at the telomere called the pseudo-autosomal region (Das *et al.* 2009). Thus, the sire X chromosome can be treated as a mostly haploid chromosome with a small diploid region. The haploid region should not have any heterozygosity and this fact can be used to identify the cut-off between the haploid and diploid regions. In addition, it can be used to identify the animals' gender, i.e. the males should not have any heterozygosity in this region except for genotyping errors and is therefore another way of checking the quality of genotypes. Once the pseudo-autosomal region has been identified, the same method used to impute the autosomal regions of sires (Ferdosi *et al.* 2014a) can be used to impute their genotypes in the pseudo-autosomal region. In this study, we discuss a very simple method to impute the remaining haploid region and illustrate its use for evaluating the genotyping quality of the X chromosome.

METHODS

Genotype data. Female offspring receive the X chromosome from their sire and only half-sib families that included at least one female were used for the sire imputation. The dataset included 8453 Angus (379 half-sib - HS), 4710 Brahman (323 HS), 53 Droughtmaster (9 HS), 1550 Hereford (37 HS), 527 Santa Gertrudis (131 HS), 1325 Wagyu (40HS) and 3411 Hanwoo. The Hanwoo half-sibs were only males and not used for sire imputation. They were only used to identify genotyping errors (heterozygous SNPs) on the X chromosome. The number of female half-sibs in each family

was between 1 and 103. The genotyping density was varied (20k, 50k, 80k and 800k) but 150k SNP that covered the 20k, 50k and 80k panels was used as a consensus panel.

Identification of the pseudo-autosomal region. 1218 male individuals with 800k markers were used to identify the pseudo-autosomal region. This was the only region in the X chromosome where, aside from genotyping errors, there can be heterozygous sites. Therefore, the region at the end of X chromosome with a clear heterozygous block was identified as the pseudo-autosomal region in the X chromosome.

Sire imputation of the haploid and pseudo-autosomal regions. The haploid region of X chromosome was imputed by using the homozygous loci that were not in the pseudo-autosomal region in the female offspring. These homozygous loci can be used to infer the sire allele directly since it only had one X chromosome. This makes sire imputation very simple and errors in imputation were only due to genotyping errors. To resolve these genotyping inconsistencies, the rounded average of the homozygous regions in the female offspring were recorded as the sire's allele; i.e. it was sufficient to simply identify the most common homozygous sites in the female offspring. This function is available in the new version of hsphase (Ferdosi *et al*. 2014b). A similar method to hsphase (Ferdosi *et al.* 2014b) but based on the log-likelihood was used to impute the pseudo-autosomal regions of the sires. Finally, the Sire imputation accuracy was calculated as the number of correctly imputed markers that must be common with genotyped X chromosome divided by the total number of markers that were available for both imputed and real sire genotypes.

RESULTS AND DISCUSSION

The pseudo-autosomal region based on the appearance of heterozygous sites was located at the end of chromosome X around position 86.2 Mb in assembly Btau4.6.1. This position is in agreement with the region previously reported by (Das *et al*. 2009). When we aligned chromosome Y with chromosome X using BLAST, we failed to find the expected very large contiguous matching block between the two chromosomes. This could be due to the quality of the assembly of the X and Y chromosomes (Tellam *et al.* 2009). We noticed a lot of missing nucleotides in both of the chromosomes; however, the largest matching block (about 10 kb) on chromosome X was still found around the 86.3 MB region.

Figure 1 shows the boxplot of proportion of heterozygous sites in the haploid region of X chromosome in males. The results showed that the majority of individuals had less than one percent genotyping errors.

The sire imputation accuracy of the X chromosome (haploid region) for 6 cattle breeds is shown in figure 2. Generally, as the number of half-sibs in the families increased, the sire imputation accuracy increased but even small family sizes have high accuracy of imputation. The accuracy was not dependent on the SNP panel of the sire nor the breed. However, the number of SNPs that can be imputed varied according to breed and panel and the number of half-sibs in a family is more important (Figure 3). The 800k and 80k panels were suitable for genotyping the X chromosome in the Brahman and Santa Gertrudis breeds which have an indicine background (Figure 3).

The sire imputation accuracy in the pseudo-autosomal region was 0.93±0.07. This accuracy was lower than for the autosomal chromosomes reported in (Ferdosi *et al.* 2016). In that report, sire imputation accuracy in the autosomal chromosomes was more than 94% for the half-sib families with more than 7 individuals. The small number of markers and mapping errors in the diploid region may be the reason for the lower accuracy of sire imputation. However, in the haploid region only genotyping errors can decrease the sire imputation accuracy as the order of markers is not relevant.

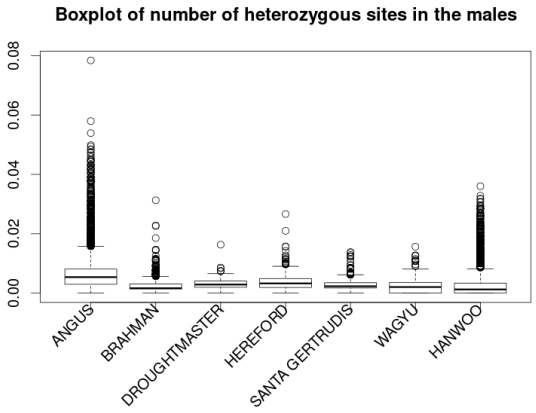

Figure 1. The proportion of number of heterozygous markers in the haploid region of the X chromosome in male animals

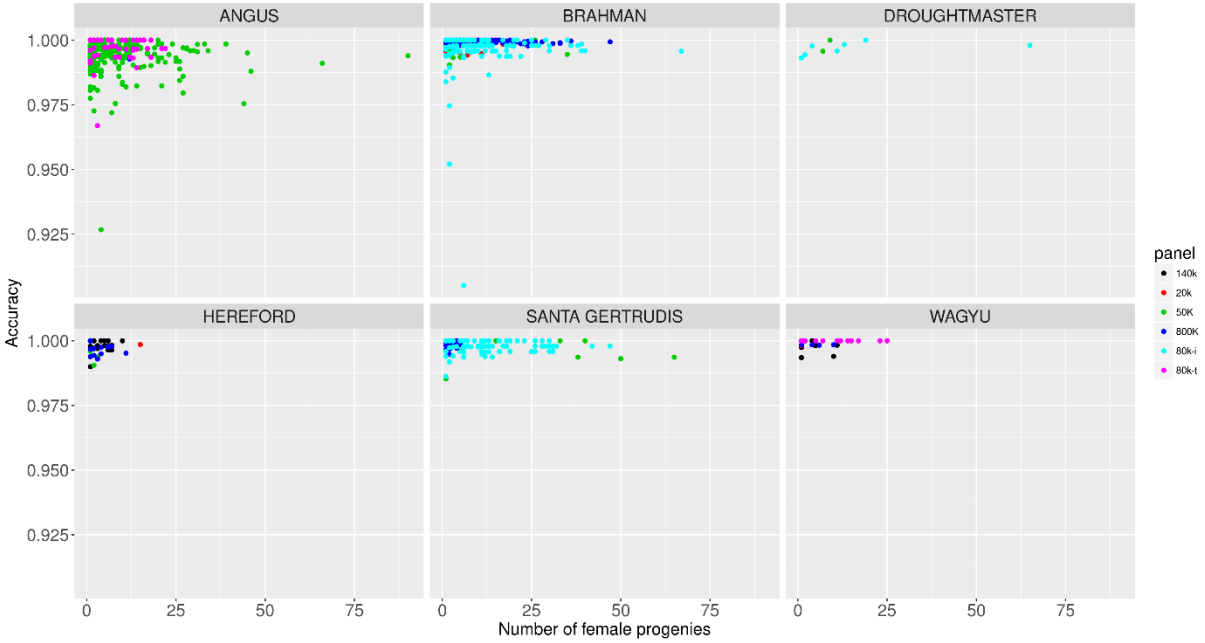

Figure 2. Sire imputation accuracy for six breeds using different panels – haploid region (i: indicus, t: taurus)

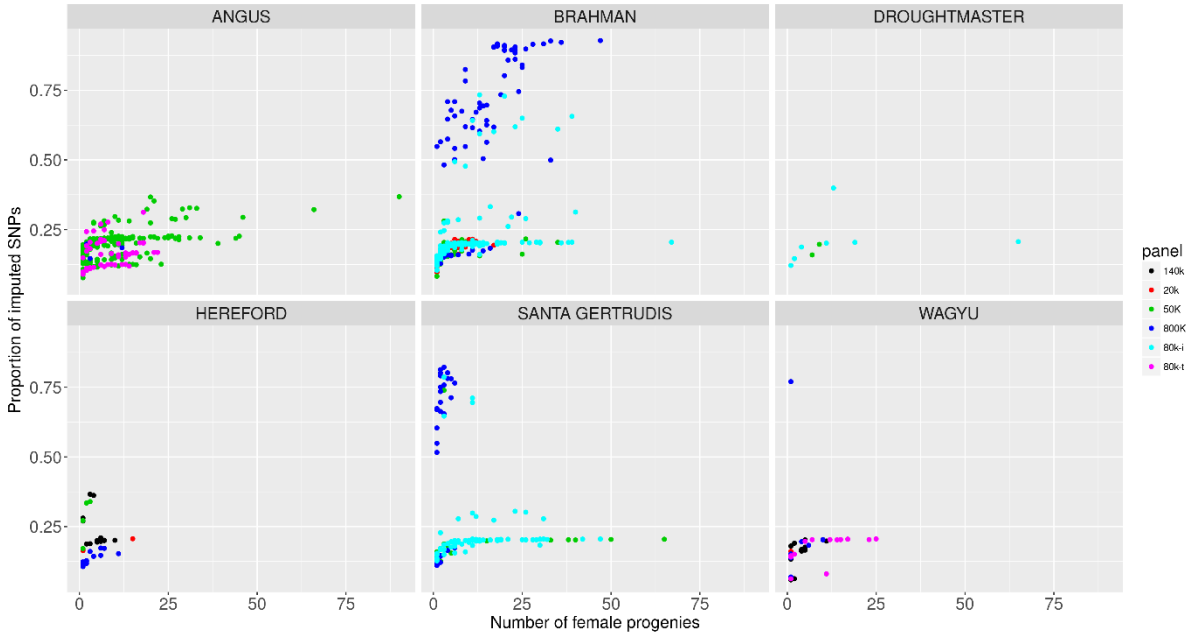

Figure 3. The proportion of SNP that can be imputed – haploid region (i: indicus, t: taurus)

The simple method detailed in this work allows for robust imputation of the X chromosome in ungenotyped sires and is useful for routine checking of the quality of genotyping. We expect that future extensions to genomic prediction methodology will make better use of the information in the sex chromosomes and this work provides an easy framework for routine imputation of the X chromosome.

ACKNOWLEDGEMENTS

MHF and DJ were supported by Meat and Livestock Australia. HAM and CG were supported by a grant from the Next-Generation BioGreen 21 Program PJ01134906 and PJ012611, Rural Development Administration, Republic of Korea and Australian Research Council (DP130100542).

REFERENCES

Das P.J., Chowdhary B.P., Raudsepp T. (2009) *Cytogenetic and genome research* **126:** 139.

Ferdosi M., Börner V., Tier B. (2016) Identifying recombination events and haplotypes in beef cattle from half-sib families. In 'International Conference on Quantitative Genetics. Madison, Wisconsin, USA'.

Ferdosi M.H., Kinghorn B.P., van der Werf J.H., Gondro C. (2014a) *Genet Sel Evol* **46:** 11.

Ferdosi M.H., Kinghorn B.P., van der Werf J.H., Lee S.H., Gondro C. (2014b) *Bmc Bioinformatics* **15:** 172.

Richardson L.A. (2016) *Plos Biol* **14:**

Sargolzaei M., Chesnais J.P., Schenkel F.S. (2014) *Bmc Genomics* **15:**

Su G.S., Guldbrandtsen B., Aamand G.P., Stranden I., Lund M.S. (2014) *Genet Sel Evol* **46:**

Tellam R.L., Lemay D.G., Van Tassell C.P., Lewin H.A., Worley K.C., Elsik C.G. (2009) *BMC Genomics* **10:** 193.