X AND Y CHROMOSOME SNPS AS INDICATORS OF SEX IN QUALITY ASSURANCE CHECKS FOR GENOMIC ANALYSIS IN WAGYU

C. Teseling[1] and D.P. Garrick[2,3,4]

[1]Australian Wagyu Association, Armidale, NSW, 2350 Australia
[2]The Helical Company Limited, Rotorua, BOP, New Zealand
[3]Theta Solutions LLC, Lacey, WA, USA
[4]A.L. Rae Centre for Genetics and Breeding, Massey University, New Zealand

SUMMARY

In 2018 genomic information was incorporated in the Australian Wagyu BREEDPLAN analysis. This development necessitated the implementation of genotype quality assurance (QA) checks to ensure the genotypes which were included in the genomic relationship matrix (GRM) were from samples collected from the same registered animals as the phenotypes in the analysis. One of the quality assurance checks is to compare the recorded sex of the registered animal to that of the predicted sex from the genotype. The Australian Wagyu population developed from a relatively small number of founder animals which limited the genetic diversity of the breed. Lower levels of genetic diversity reduce the usefulness of X chromosome heterozygosity to predict the sex of the animal from which the sample was collected. Modern SNP chips include Y chromosome data, the use of which in sex prediction can greatly increase the accuracy of sex-based quality control checks.

INTRODUCTION

The Wagyu breed in Australia was established from 221 Fullblood foundation animals mostly exported from Japan between 1990 and 1997. In comparison to other breeds, this is a relatively small number of foundation animals which results in the risk of increased subsequent inbreeding and lower levels of genetic diversity (Ferdosi *et al.* 2019). In populations where the effective population size is relatively small, the diversity on the X chromosome could be expected to be half of that of the autosomes (Schaffner 2004), while Mészárosová *et al*. (2022) found that X chromosome heterozygosity could vary significantly in the same population.

In 2018 genomic information was incorporated in the Australian Wagyu BREEDPLAN analysis. An important component of utilising genomic information is to ensure the genotype is associated with the correct registered animal and corresponding phenotypes. The Animal Genetics and Breeding Unit (AGBU) developed a data pipeline which incorporates a range of QA checks to ensure genotype integrity (Connors *et al*. 2017).

An important genotype QA check which should be implemented by genetic analysis service providers is to compare the recorded sex of the registered animal to the sex predicted from the animal's genotype (Connors *et al*. 2017, ICAR Guidelines 2022, McClure *et al*. 2018). Not every commercial chip includes chromosome Y SNPs, however, they typically contain chromosome X markers. Both the X and Y chromosomes contain a pseudo-autosomal region (PAR) and it is important to ensure only SNPs from the non-PAR (nPAR) region are used in sex prediction analysis.

The sex of the genotype can be predicted by evaluating the nPAR X chromosome heterozygosity and/or the presence or absence of calls on nPAR Y chromosome SNPs. Normally females have two copies of the X chromosome while males have one X chromosome and one Y chromosome.

Combining the X and Y chromosome results sometimes lead to conflicting sex prediction outcomes. In the very rare occurrence where an animal has Turner (X0) or Klinefelter's syndrome (XXY), the X and Y chromosome results will conflict. Also, when a semen sample only report X chromosome SNPs, it could indicate a female sex selected semen sample was submitted for genotyping.

When the early bovine chips were manufactured, the sequence information on sex chromosomes were not well assembled, and no Y chromosome SNPs were on the chips. These issues, as well as a low genetic diversity on the X chromosome can result in registered females having low X chromosome heterozygosity and their genotypes incorrectly excluded from the genetic analysis. This paper investigates the use of nPAR X and Y chromosome data in estimating genotype sex and considers results where the X and/or Y predicted sex of the genotype and the recorded sex of the registered animal may conflict. The aim is to determine population specific thresholds for Australian Wagyu to improve accuracy of sex prediction and reduce incorrect exclusion of valid female genotypes from genomic analyses.

MATERIALS AND METHODS

The Australian Wagyu Association (AWA) has more than 325,000 animals genotyped, from more than 3,600 different chips or manifests. These data are stored in their genotype database hosted by the Helical Company (Garrick *et al*. 2023). More than 50% of these genotypes are from research or commercial animals which are not registered in the AWA's registration database hosted by the Australian Business and Research Institute.

The samples can be categorised based on their number of SNPs. Table 1 shows the different categories of chips, the number of animals genotyped for each category, their number of SNPs as well as the numbers of nPAR X and Y chromosome SNPs present.

Table 1. Number of genotypes in each chip category with number of SNPs, number of nPAR X chromosome SNPs and number of nPAR Y chromosome SNP.

Chip Category	#Genotypes	#SNPs	#X Chrom	#Y Chrom
Parentage	59,620	180 to 641	0	0
10K	7,821	6,900 to 10,000	218 to 271	0 to 7
30K	5,444	19,000 to 35,000	634 to 919	0 to 7
50K	86,313	35,000 to 49,000	291 to 1,893	0 to 189
70K	17,730	50,000 to 77,000	288 to 1,561	6 to 239
100K	148,006	93,000 to 96,000	2,091	269
140K	272	137K to 140K	2,015	25
770K	187	777,963	2,821	267

Genotypes were extracted and analysed to compare the accuracy of predicting the sex of the animals from which the sample was collected. Animals genotyped using chips with no nPAR X chromosome SNPs or less than six nPAR Y chromosome SNPs or a call rate lower than 95% were excluded from the analysis. The first analysis included all animals that had nPAR X and at least six nPAR Y chromosome SNPs present on the chip which resulted in a total of 247,057 genotyped animals. The second analysis excluded all genotypes of animals not registered in the AWA Herdbook, which reduced the total to 104,860 genotypes. An animal must be parent verified to both its parents to be registered in the AWA's Herdbook, which ensure a high level of quality assurance.

Traditionally the focus has been on using the nPAR X chromosome to predict the sex of the animal from which the sample was collected. The increasing number of Y chromosome SNPs on more recently developed chips make it possible to consider the number of called nPAR Y chromosome SNPs to improve the accuracy of the prediction. If the genotype has a high proportion of nPAR Y chromosome SNPs reported, it could indicate that the sample was collected from a male while a very low proportion or no called nPAR Y chromosome SNPs would be suggestive the animal was a female.

The sex prediction accuracies of three different methods were assessed by comparing the predicted sex of the genotype with the sex of the associated registered animal. The three methods

are presented in Table 2 where Method 1 only used the nPAR X chromosome heterozygosity, Method 2 use both the nPAR X chromosome heterozygosity (Method 1) as well as the number of called nPAR Y chromosome SNPs (McClure et al. 2018, Garrick 2019) and Method 3 use nPAR X and Y chromosome SNPs in a stepwise approach where the proportion of nPAR Y chromosome SNPs are first used and only if that doesn't predict a clear result, the nPAR X chromosome heterozygosity of Method 1 is considered.

Table 2. The criteria used to determine the sex prediction accuracy of three different methods

	Predicted Male	**Predicted Female**	**Ambig.**
1	nPAR X ≤ 5% Heterozygosity	nPAR X > 5% Heterozygosity	
2	Method 1 + nPAR Y > 5	Method 1 + nPAR Y < 2	X ≠ Y
3	nPAR Y < 0.4 or nPAR Y = 0.4 to 0.6 + Method 1	nPAR Y > 0.6 or nPAR Y = 0.4 to 0.6 + Method 1	

RESULTS AND DISCUSSION

Table 3 shows the distributions of the proportion of nPAR Y called SNPsfor all genotypes and registered animals. The results show there is a wide continuum of proportions observed across all the samples tested with no clear cut-off between what may be expected as male vs. female samples for the nPAR Y SNPs used. To determine if earlier chips which tended to have fewer Y SNPs available are disproportionally contributing to this observed variation, genotypes with less than 100 Y SNPs on the chip were removed from the analysis containing all genotypes (96,348 genotypes removed). The results still indicate no clear break in the observed proportions.

Table 3. Distribution of the numbers of animals with various proportions of called Y chromosome SNPs when all genotyped animals (All Genos), only chips with more than 100 Y SNPs (> 100 Y), or only registered animals (Registered) were analysed

	Proportion of called Y chromosome SNPs									
	<0.1	<0.2	<0.3	<0.4	<0.5	<0.6	<0.7	<0.8	<0.9	<1.0
All Genos	141,762	885	315	141	43	92	1,686	101	13,009	89,023
> 100 Y	85,045	601	194	79	24	7	2	10	68	64,679
Registered	78,266	489	189	96	30	13	286	33	4,018	21,440

Figure 1 shows the distribution of the proportion of genotypes after genotypes with nPAR X chromosome heterozygosity of zero (expected males based on X chromosome only) were excluded, resulting in 144,308 and 79,147 from all and registered animal genotypes respectively. The graph on the left in Figure 1 displays the distribution of the number of animals genotyped relative to the heterozygous proportion of nPAR X chromosome. The graph on the right includes only that subset of animals with low X chromosome heterozygosity excluding animals with the proportion of genotypes with nPAR Y > 0.5 (1,725 which are expected to be male). The same reduction was observed when genotypes from registered animals were analysed.

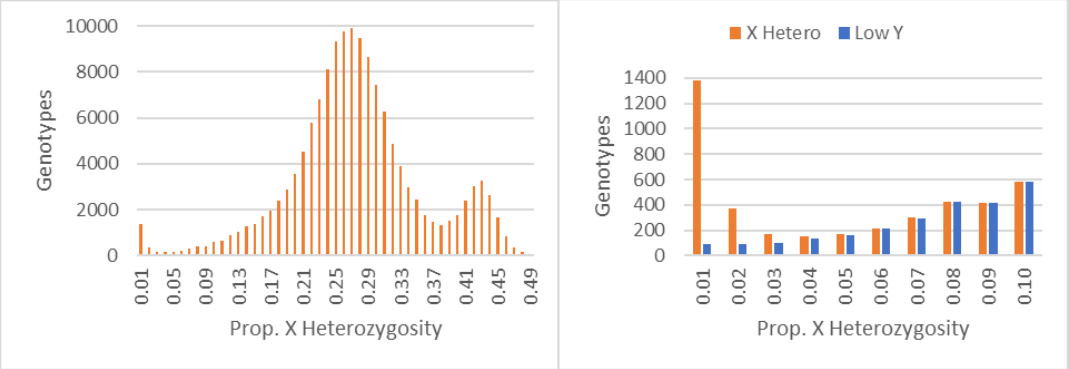

Figure 1. Number of genotyped animals exhibiting X chromosome heterozygosity (left) and animals with <0.1 proportion X chromosome heterozygosity after genotypes with more than 0.5 proportion of called Y chromosomes were excluded (right)

Using the three methods presented in Table 2 to predict the sex of the 104,860 genotypes of registered animals (78,906 females and 25,689 males) found that Method 1 incorrectly predicted 975 (1.23%) females to be males and 165 (0.65%) males to be females. Method 2 predicted 2,955 (3.75%) females to be ambiguous and 513 (0.65%) females to be males while 312 (1.21%) and 147 (0.57%) of males were predicted to be ambiguous and females respectively. Method 3 incorrectly predicted 508 (0.64%) females to be males and 433 (1.6%) males to be female.

Inspection of the "problem" genotypes suggests that some batches of samples exhibited little or no variation in the number of called Y chromosome SNPs, perhaps due to problems with the cluster files used for SNP calling at those loci in those batches.

CONCLUSIONS

Reduced genetic diversity negatively impacts the usefulness of X chromosome heterozygosity as the only criteria to predict animal sex from called SNPs.

Combined use of X and Y chromosome SNPs reduces the number of animals with incorrectly predicted sex. However, consideration of the chip content and careful scrutinization of variation and distribution of results will be required as additional criteria to reduce the percentage of incorrectly predicted sex from genotype calls to less than 0.5% of all animals.

REFERENCES

Connors N., Cook J., Girard C., Tier B., Gore K., Johnston D. and Ferdosi M. (2017) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **22:** 317.

Ferdosi M.H., Connors N.K. and Khansefid M. (2019) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **23:** 118.

Garrick D.P. (2019) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **23:** 23.

Garrick D.P., Howse, J. E. and Garrick, D.J. (2023) Retrieved from https://www.helicalco.com/

ICAR Guidelines (2022) Available online at: https://www.icar.org/Guidelines/04-DNA-Technology.pdf (Accessed February15, 2023).

McClure M.C., McCarthy J., Flynn P., McClure J.C., Dair E., O'Connell D.K. and Kearney, J.F. (2018). *Frontiers in Genetics* **9**(84).

Mészárosová M., Moravčíková N., Mészáros G., Pavlík I., Margetín M."and Kasarda R. (2022) https://www.wageningenacademic.com/pb-assets/wagen/WCGALP2022/34_011.pdf

Schaffner (2004) *Nat. Rev. Genet*. **5:** 43.