A GENETIC EXPLORATION OF AUSTRALIAN LARGE WHITE PIGS

K.P. Gore, R.G. Banks, V. Boerner and S. Hermesch

Animal Genetics and Breeding Unit[*], University of New England, Armidale, NSW 2351

SUMMARY

Limited information is available about the genetic structure of Australian pigs. The genetic relationships of 20 Large White and 2 Duroc pigs sampled from one herd were explored and their genetic distance to European pigs of the same and different breeds was estimated using SNP data. On average, 36% of SNPs were heterozygous for Large White pigs. Mean correlations between Australian Large White and European breeds were highest for European Large White (0.35) and Middle White (0.34), and lowest for European Duroc (0.20) and Meishan (0.09). The analysis of breed percentages based on constrained genomic regression showed highest similarity of Australian Large White with European Large White (36.8%) followed by Middle White (15.9%) and Welsh (12.6%). Based on this small sample, the Australian pigs sampled retained significant heterozygosity and can be regarded as a distinct population to the sampled European breeds.

INTRODUCTION

Importation of porcine genetic material into Australia has been prohibited for over 2 decades. The domestic industry has had to breed for productivity while controlling inbreeding without the ability to introduce external genetics. With the number of producers shrinking by approximately two thirds during that time (Australian Pork Limited 2013), sourcing diverse off-farm genetics has become more difficult.

Genomic information being accessible on-line offers opportunities to examine relationships and genetic diversity between populations. The availability of this information vastly reduces the cost to any given individual researcher, and enables initial exploration of the genetic (genomic) structure of the Australian population to be undertaken.

MATERIAL AND METHODS

Hair samples were obtained from 22 pigs (21 boars and 1 sow) in the herd at the University of Queensland, Gatton Australia which has used boars from other Australian herds. The samples were from 20 Large White and 2 Duroc. The 2 Duroc pigs were not discussed in detail in the results presented here due to the small numbers.

Samples were genotyped using the GeneSeek Genomic Profiler HD chip. The chip originated from the original Illumina Porcine60k chip with approximately 12,000 SNPs of low informativeness in major commercial breeds removed and an extra 20,000 added to fill gaps in the chromosomes (J. Walker, personal communication, April 1 2017). Quality control consisting of minimum 85% call rate was applied resulting in 1 sample being rejected. One duplicate was also detected. SNPs were removed where there was at least 1 missing call resulting in 44,749 SNPs. The 2 possible heterozygote calls were not considered as different in any calculation.

The publically available European SNP data was already subject to quality controls from its original publication (Wilkinson *et al*. 2013a,b). The Australian genotypes were merged with the European SNP data by SNP name and SNPs were removed if there was at least 1 missing call resulting in 24,564 SNPs that were available in both SNP data sets. Pearson correlation in R (R Core Team 2015) was used to generate the correlation matrix between the genotypes.

---

[*]AGBU is a joint venture of NSW Department of Primary Industries and the University of New England

Breed percentage was calculated for the Australian Large White pigs using the European data as the reference set with constrained genomic regression (Boerner 2017). This was repeated for the European Large White pigs with the Australian Large White pigs in the reference set for comparison. Principal component analysis on the Middle White, European Large White and Australian Large White breeds was done by generating a genetic relationship matrix (Yang *et al.* 2010) followed by singular value decomposition using the SVD function from the NumPy package (van der Walt *et al.* 2011) for python. The R functions kmean and dist (R Core Team 2015) were used to assign individuals to clusters and calculate the distance between the cluster centres.

RESULTS AND DISCUSSION
Within herd comparison. The mean correlation between the genotypes was 0.41 with a range from 0.33 to 0.62 and a standard deviation of 0.04. The maximum value of 0.62 was confirmed by pedigree records to be a parent-progeny pair.

The percentage of heterozygous SNPs for each pig ranged from 34 to 38 with a standard deviation of 1. The low standard deviation is likely to be the result of considering a single breed. Including the 2 Duroc pigs increased the standard deviation to 2 because of their lower percentage of heterozygous SNPs (31 and 32). Duroc is the smaller breed in comparison to Large White in Australia which was reflected in higher inbreeding levels and smaller effective population size for Duroc in comparison to Large White based on pedigree information (D'Augustin *et al.* 2017). The results of this study based on genomic information corresponded to the findings based on pedigree data despite the small sample size.

The percentage of heterozygous SNPs in the European and Australian breeds is shown in Table 1. The Australian Large White pigs (AULW) were the highest of all with a mean of 35.5. These means were much lower than those of Zhang and Plastow (2011), which may be the result of only considering SNPs that were called for all pigs. Imputation of these sporadic uncalled SNPs may allow more of the data to be used. Li *et al.* (2006) showed that reasonable accuracy of imputed SNPs can be achieved with as few as 90 individuals which could be achieved for this sample of pigs with additional genotyping.

European comparison. The heat map (Figure 1) of the correlation matrix indicated that the Australian animals could be considered a separate breed to the European Large Whites. The squares along the diagonal show the groups of animals of the same breed. The order of breeds from the top left to the bottom right corner was Meishan (MS), Gloucestershire Old Spots (GL), Berkshire (BK), Wild boar (WB), Large Black (LB), British Saddleback (BS), Tamworth (TA), Hampshire (HA), Mangalica (MA), Australian Duroc, Duroc (DU), Landrace(LR), Welsh (WE), Pietrain (PI), Middle White (MW), European Large White (LW), Australian Large White (AULW). The highest mean correlations to Australian Large White were European Large White (0.35) and Middle White (0.34) as shown in Table 2. The lowest correlation was Meishan(0.09).

Table 1. Mean percentage of heterozygous SNPs for breeds

AULW	MS	GL	BK	WB	LB	BS	TA	HA	MA	DU	LR	WE	PI	MW	LW
35.5	16.3	23.3	23.8	20.7	26.7	30.2	21.7	24.3	15.2	26.3	32.3	33.9	34.2	29.6	33.9

Principal component analysis of the White breeds (Middle White, European Large White and Australian Large White) showed that each breed formed separate clusters (Figure 2). The Middle White and European Large White breeds were focused on due to the higher relationship shown in the correlation heat map. The distance between the cluster centers was 14.0 from the Middle White

to the European Large White, 12.6 from the Middle White to the Australian Large White and 9.0 between the European and Australian Large Whites.

There were significant proportions for European Large White, Middle White, Pietrain and Welsh breeds. Minimal proportions of Landrace, Duroc, British Saddleback, Gloucestershire Old Spots, Meishan, Tamworth, Hampshire and Large Black (Table 3). Breeds where the mean was below 1% are not shown (Berkshire 0.6%, Mangalica 0.2%, Wild Boar 0.8%).

Table 2. Mean correlations between Australian Large White and European breeds (*100)

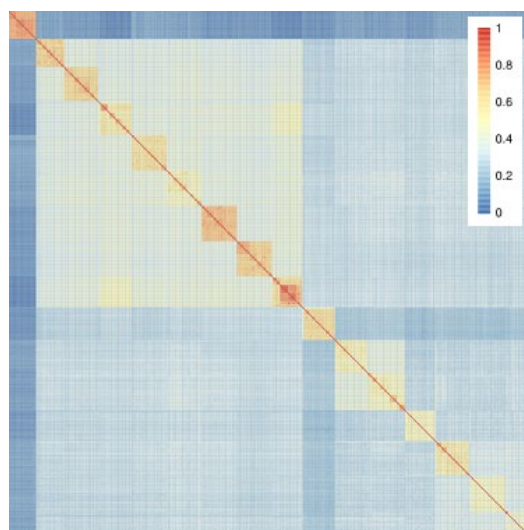|	MS	GL	BK	WB	LB	BS	TA	HA	MA	DU	LR	WE	PI	MW	LW
Mean	9	28	28	28	29	31	30	28	29	20	29	30	29	34	35
SD	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2

Figure 1. Heatmap of individual correlations between Australian and European Breeds
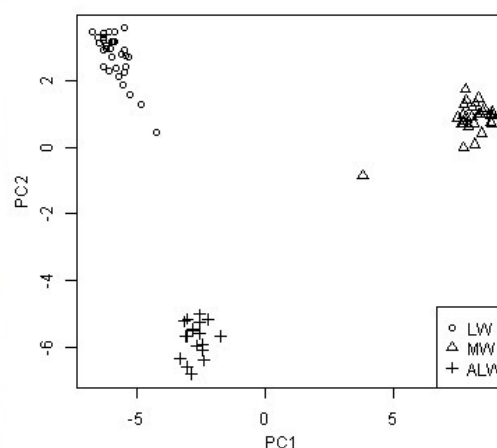

Figure 2. PC1 vs PC2 for European Large White, Middle White and Australian Large White

Table 3. European breed percentages of Australian Large White pigs

	MS	GL	LB	BS	TA	HA	DU	LR	WE	PI	MW	LW
Mean	2.3	2.4	1.1	2.6	2.0	1.9	3.1	5.8	12.6	12.5	15.9	36.8
Min	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	2.6	10.3	21.1
Max	4.1	5.8	4.6	7.5	4.8	4.5	5.9	18.2	26.0	18.3	22.3	44.5
SD	1.1	2.0	1.5	2.6	1.4	1.2	1.6	4.7	5.3	3.8	3.4	5.2

The calculated breed percentages of European Large White pigs is shown in Table 4. Breeds where the mean was below 1% are not shown (Duroc 0.2%, Gloucestershire Old Spots 0.8%, Hampshire 0.3%, Mangalica 0.6%). The highest percentage was the Australian Large White at

45%. The Australian Large White pigs showed nearly 10 times the percentage of the Welsh breed (12.6%) than the European Large White pigs (1.3%).

Table 4. Breed percentages of European Large White pigs

	MS	BK	WB	LB	BS	TA	LR	WE	PI	MW	AULW
Mean	3.9	2.3	3.5	2.3	5.5	1.3	6.8	1.3	10.2	15.9	45
Min	1.4	0	0	0	0	0	1.5	0	3.1	8.2	31.3
Max	6.8	6.8	8.2	7.1	10.2	4.1	11.6	10.3	38.6	21.8	52.8
SD	1.3	1.6	2.5	1.8	2.8	1.2	2.9	2.3	6	3.1	5.2

CONCLUSIONS

This Australian Large White population has different from the European Large White population. The between-breed correlation matrix showed a higher relationship between the White breeds when compared to the other breeds but the principal component analysis showed that this sample of Australian Large White pigs was distinctly different from the European White breeds. Although the Australian genotypes originated from just 1 herd, there was a similar level of genetic diversity within this one herd as within the European Large White population, suggesting that this herd at least is maintaining diversity. Both the Australian and European Large White populations retained genetic contributions from other breeds, presumably reflecting introductions over time. This study is based on a small sample and caution should be exercised in concluding that the diversity estimated within this herd is an accurate estimate of that in the whole Australian population. Further investigation of the genomic structure of a larger sample of Australian pigs is required in order to obtain more detailed knowledge of the genetic diversity of Australian pigs.

REFERENCES
Australian Pork Limited (2013) Australian Pig Annual 2012-2013
Boerner V. (2017) Proc. Assoc. Adv. Anim. Breed. Genet. **22**: In prep.
D'Augustin O., Boerner V. Hermesch S. (2017) Proc. Assoc. Adv. Anim. Breed. Genet. **22**:Subm.
Li Y., Willer C, Sanna S, Abecasis G (2006) *Annu Rev Genomics Hum Genet* **10**:387
R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Van der Walt S, Colbert S Chris, Gael Varoquaux (2011) *Computing in Science & Engineering* **13**:22-30
Wilkinson S, Lu ZH, Megens H, Archibald AL, Haley C, Jackson IJ, Groenen MAM, Crooijmans RPMA, Ogden R and Wiener P (2013) *PLoS Genetics* **9**: e1003453. doi:10.1371/journal.pgen.1003453
Wilkinson S, Lu ZH, Megens H, Archibald AL, Haley C, Jackson IJ, Groenen MAM, Crooijmans RPMA, Ogden R, Wiener P (2013) Data from: Signatures of diversifying selection in European pig breeds. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.c2124
Yang J, Benyamin B, McEvoy B, Gordon S, Henders A, Nyholt D, Madden P, Heath A, Martin N, Montgomery G, Goddard M, Visscher P (2010) *Nat Genet* **42**:565
Zhang C, Plastow G (2011) *Curr Genomics* **12**:138