Proc. Assoc. Advmt. Anim. Breed. Genet. 18:187-190

A GENOME-WIDE ASSOCIATION ANALYSIS IDENTIFYING SNPS FOR PRRS TOLERANCE ON A COMMERCIAL PIG FARM

C.R.G. Lewis[1,3], M. Torremorell[2], L Galina-Pantoja[2], N Deeb[2], M.A. Mellencamp[2,4], A.L. Archibald[1] and S.C. Bishop[1]

[1]The Roslin Institute & R(D)SVS, University of Edinburgh, Roslin, Midlothian EH25 9PS, UK
[2]Genus/PIC, 100 Bluegrass Commons Blvd, suite 2200, Hendersonville, TN 37075, USA
[3]Current Address: Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351, Australia
[4]Current Address: Ralco Nutrition, Inc., 1600 Hahn Road, Marshall, MN 56258, USA

SUMMARY

Porcine reproductive and respiratory syndrome (PRRS) is currently the most economically important viral disease affecting pig production outside Australia. This study utilised commercial data to perform a genome wide association study looking for single nucleotide polymorphism (SNP) markers associated with PRRS resistance or tolerance, as assessed indirectly from reproductive traits. In total, phenotypes were measured on 1,545 sows, with the data split according to whether the trait was measured during a healthy PRRS-free phase on the farm (4,378 litters from 1,019 sows) or a diseased phase (1,977 litters from 1,526 sows). All animals were genotyped using the Illumina porcine 7k SNP chip. Associations between each individual SNP and reproductive outcomes were assessed using the residual values from a linear mixed (animal) model analysis of each trait in a series of single SNP analyses. Significant SNP associations were only observed for reproductive traits recorded during the disease phase, implying specificity of identified SNPs to a PRRS active phase. After correction for false positives, six significant SNP markers were identified for piglets born alive, piglets born dead and mummified piglets born per litter. SNP effects were then re-estimated from linear mixed model analyses of the data in which the significant SNPs were fitted as additional fixed effects. The SNPs are generally additive in their mode of action and on average the significant SNPs explain 4.7% of the genetic variation for born alive, 1.6% for born dead and 2.2% for born mummified. After further validation and research, these SNPs may allow breeding of pigs that are more robust in the face of PRRSV infection.

INTRODUCTION

Porcine reproductive and respiratory syndrome (PRRS) is by far the most economically important viral problem to the global pig industry. Although vaccines and control measures exist, the disease is persistently endemic and causes problems in both the growing and breeding herds of infected farms. The associated losses from PRRS virus (PRRSV) infection are considerable due to the impact of the virus on the respiratory system, thus affecting overall productivity in the grower/finisher pigs. It also has effects on the reproductive success of the host. Between-breed genetic variation in response to PRRSV infection has been previously demonstrated several times, however it is only recently that evidence for within-breed host genetic variation has been identified, as assessed by impacts on reproductive performance traits (Lewis *et al.* 2009).

This study seeks to dissect the host genetic variation highlighted by Lewis *et al.* (2009) and identify single nucleotide polymorphisms (SNPs) associated with PRRSV tolerance. This study was done using a porcine 7k SNP chip that was developed by the Roslin Institute, the University of Aarhus (Denmark) and the Sanger Centre (UK) on an Illumina platform. The aim was to

identify significant SNPs for traits of relevance to PRRS, which could potentially be utilized in a breeding program to make commercial lines more robust to PRRSV infection.

MATERIALS AND METHODS

The data were collected from a multi-line multiplication herd that had continual recording of the main herd descriptors and production traits. This dataset has been previously described by Lewis *et al*. (2009). DNA was extracted by PIC/Genus, and all SNP genotyping was done at the Wellcome Trust Clinical Research Facility of the Western General Hospital, University of Edinburgh. Individual litters in the data were then assigned to two groups (baseline vs. diseased) using the methodology described by Lewis *et al*. (2009). There were 821 sows that had records common to both data sets.

The SNP genotype data were subjected to quality control (QC) measures. Genotyping was done in two batches of 864 sows and 768 sows respectively for 6,523 SNP markers. The first level of QC removed all of the SNPs that failed to call (1,072), then the SNP markers with a minor allele frequency (MAF) less than 0.01 were removed. Individual sows that had more than 40% of their SNPs uncalled were also removed from the data (~40 individuals). This left 1,545 individual sows recorded for reproductive performance, genotyped with 4,595 SNPs.

The SNP association analyses were done separately on the baseline and disease data for all traits: number of services, total services to conception, gestation length, total piglets *in-utero*, born alive, born dead (mummified + stillborn), mummified piglets (log+1 transformed as it was a non-normally distributed trait), stillborn piglets, total weaned piglets and lactation length. The approach used was that described by Aulchenko *et al.* (2007), i.e. regressing residuals obtained from a mixed model analysis of each trait on the SNP genotypes, performing the regression one SNP at a time. The residuals were obtained from the estimation of the genetic parameters presented in Lewis *et al.* (2009) fitting an animal model including all pedigree and the fixed effects of line (11 levels including European e.g. Large White, Landrace etc. and Asian breeds e.g. Meishan) and parity (seasonal effects were explored and found non-significant). The SNP association analysis was done utilizing the GenABEL package in R. To correct for multiple testing, p-values were corrected using permutation (10,000 iterations) to identify the genome-wide significant SNPs (corrected p-value < 0.05). SNPs that were significant from the genome-wide association analysis were further explored in mixed model single and multi-SNP analyses. These analyses were used to estimate additive and dominance effects for each SNP, to test for SNP interactions (pair-wise combinations to test independence) and also SNP by line interactions. The significant SNPs were then added individually to the model to determine their direction and size of effect. Finally, the proportion of genetic variation explained by each significant SNP was calculated. The additive genetic variance due to the SNP was calculated as $2pq[a+d(q-p)]^2$ (Falconer and MacKay, 1996) and the total additive genetic variance was obtained from the animal model analyses of the same trait, in a model ignoring SNP effects.

RESULTS

A full description of the data is found in Lewis *et al*. (2009). Briefly, the mean and phenotypic standard deviation values of the baseline data for all traits were well within the bounds of what is expected on a commercial farm. The impact of disease on the farm was marked. The PRRS outbreak had a significant ($P<0.001$) impact on the numbers of piglets born alive (10.3 in baseline vs 9.0 during disease), piglets mummified (0.04 in baseline vs 1.13 during disease), and piglets born dead (0.59 in baseline vs 2.15 during disease). The losses highlighted due to PRRSV infection hint at the costs associated with the disease. Indeed, piglets weaned per litter is a common measure of whole farm performance and this also decreased significantly ($P<0.05$) in the

presence of the disease (9.57 in baseline vs 8.39 in the disease data). Further, variability (i.e. the standard deviation) increased in most traits in the presence of the disease.

Details of the significant SNPs (from the single SNP analysis) are shown in Table 1. After permutation testing there were no significant SNPs for any trait measured during the baseline phase; all significant SNPs were for traits affected by PRRS. In total six significant SNP were found from genome wide association with four SNPs affecting more than one trait. Significant SNPs were for: born alive (1 SNP), born dead (6 SNPs) and mummified piglets (4 SNPs), all during the disease phase. There were no significant SNP effects for all other traits investigated.

Upon examination of the two-way SNP interactions, in each case the SNP effects were statistically significant, i.e. they had independent effects on the trait of interest, and none of the interactions between SNP and sow lines were significant. The results indicate that the fixed effects of line were successfully removed as there were no interactions between the SNP and line. The proportion of additive genetic variance explained by each SNP for each trait is also shown in Table 1. The SNPs identified in this study explain a relatively small proportion of the additive genetic variance (across lines). It should be noted that the size of effect and variance explained by the significant SNPs may decrease in validation populations.

Table 1. The estimated effects of significant SNP markers identified in the disease phase data, along with the proportion of genetic variation explained by each of the six significant correction for genome wide significance using permutation

Trait - Total Born:	SNP	P-value raw	P-value*	Minor Allele Freq.	Additive effect	Dominance effect	Proportion of genetic variance explained
Alive	3393	$3.30E^{-05}$	0.08	0.332	0.285	0.575	0.047
Dead	3393	$1.16E^{-10}$	<.001	0.332	-0.055	0.005	0.012
	297	$3.10E^{-07}$	<.001	0.026	-0.210	0.050	0.018
	2589	$6.82E^{-06}$	0.01	0.031	-0.115	0.055	0.006
	870	$9.67E^{-06}$	0.02	0.229	-0.080	-0.010	0.029
	382	$2.13E^{-05}$	0.04	0.039	0.025	-0.015	0.001
	1479	$4.78E^{-05}$	0.09	0.228	-0.050	-0.070	0.030
Mummies	3393	$9.71E^{-07}$	0.001	0.332	-0.070	-0.050	0.029
	382	$1.33E^{-06}$	0.001	0.039	-0.200	-0.030	0.009
	2589	$4.75E^{-06}$	0.009	0.031	-0.150	0.000	0.022
	297	$4.31E^{-06}$	0.08	0.026	0.085	-0.055	0.027

* P-values after permutation for genome wide significance, using 10,000 iterations.

DISCUSSION

This study is the first genome wide SNP association study for PRRS resistance or tolerance in pigs. The study found that significant SNPs can be identified that are associated with disease tolerance using data collected from an outbreak in a commercial herd. The quantity of data needed for a SNP association study of this type, or indeed to estimate genetic parameters for any type of disease resistance or tolerance is considerable. Large numbers of animals with detailed phenotypic records and well managed DNA collections are all needed for a successful study. Indeed, studies of this type are still in their infancy and a major recommendation from this study is that high quality phenotypic data is needed to dissect the traits of interest. The difficulty of detecting SNPs that meet genome-wide significant thresholds is highlighted by the finding that no significant

SNPs were detected for performance traits measured during the PRRS-free period. The data quality issue for this and other studies also extends to the level of the SNP chip. The 7k SNP chip was utilized within this study as at the time it was the best available, however, one could argue that this chip is inadequate as there are large areas of the genome that are not covered and it is likely that there are mutations contributing to the trait variation that are not in LD with any of the SNPs on the chip. A denser chip would be required for a true genome-wide association analysis and likewise for genomic selection.

The choice of dependant variable, i.e. using the residual, was proposed by Aulchenko *et al*. (2007) and it was utilized because all fixed effects (in this case line and parity) are removed, the background polygenic variation is removed (by fitting the pedigree), and genetic covariances between related individuals are accounted for (thus reducing false positives due to population structure). Effectively, associations between the SNP and the within-family Mendelian segregation term are being investigated. This enabled a robust association scan for SNPs associated with the traits of interest across the whole population removing, wherever possible, SNPs associated within specific lines or groups of animals. Another major advantage from utilizing this methodology is that it is very fast computationally, thus saving considerable time in the analysis of each trait. A major issue with the data presented here is that we did not find any SNPs associated with any of the underlying traits (in the absence of disease).

SNP 3393 had significant effects on piglets born alive, born dead, and mummies, but was not significant for total born. Since there is no evidence of PRRSV affecting ovulation rate then the SNP effects for born alive and born dead must be opposite for the SNP effect on total born to be non significant. This can be observed to be the case from the results in Table 1, providing some support for the estimated SNP effects to be biologically meaningful, given that PRRS infected sows have abnormally high numbers of mummies and therefore lower born alive.

Now that SNP markers for PRRS tolerance have been identified in this data, the next step is to validate these in another dataset. Ideally, this would be a PRRS outbreak on another commercial farm. It would also be beneficial to repeat this process using a denser SNP array. Once any significant SNPs are validated, the effects of these SNPs on other traits important to production (e.g. FCR or growth rate) need to be examined to determine whether selection for these markers would have detrimental effects on performance traits. If all of these safeguards are cleared successfully then marker assisted selection could be used to create lines that are more robust to PRRSV infection. Further research is needed; however identification of causal genes may also shed light on the mechanisms of host resistance to PRRS.

ACKNOWLEDGEMENTS

This work was financially supported by the BBSRC (grant EGA16307) and Genus plc. The authors would also like to thank Dr. Kim Bunter for critical revision and constructive comments.

REFERENCES

Aulchenko, Y.S., De Koning, D.J. and Haley, C.S. (2007) *Genet*. **177**:577.

Falconer, D.S. and MacKay, T.C. (1996) "Introduction to Quantitative Genetics" 4th ed. Pearson Education Ltd., Essex, UK.

Lewis, C.R.G., Torremorell, M., Galina-Pantoja, L. and Bishop, S.C. (2009) *J. Anim. Sci*. **87**:876.