View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Edinburgh Research Explorer

THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Alternative association analyses on boar taint using discordant sib pairs experimental design

Citation for published version:
Karacaoren, B, de Koning, DJ, Velander, I, Petersen, S, Haley, CS & Archibald, AL 2010, 'Alternative association analyses on boar taint using discordant sib pairs experimental design', Leipzig, 1/08/10 - 6/08/10, pp. paper 743.

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Early version, also known as pre-print

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. Apr. 2019

Alternative Association Analyses on Boar Taint using Discordant Sib Pairs Experimental Design

B. Karacaören[*], D.J. de Koning[*], I. Velander[+], S. Petersen[+], C.S. Haley[†*] and A.L. Archibald[*]

Introduction

Boar taint (BT) is an unwanted odour that may occur when cooking meat of male pigs. It has been shown that high levels of androstenone and skatole are the main cause for boar taint. Boar taint can be prevented through castration of male piglets but this also reduces feed efficiency. Furthermore, castration is currently being banned in many countries on welfare grounds. Using marker assisted selection following a genome-wide association study (GWAS) would be a sustainable way to reduce the occurrence of boar taint. In standard case-control GWAS studies, contingency tables are used to detect disease-marker associations. However, if the genetic relationships between individuals are not taken into account this may lead to false positives (Siegmund, D.O. and Yakir, B. (2007)). The main aim of this study was to search for SNP affecting BT using GWAS while taking the genetic structure into account using GRAMMAR (Aulchenko, S.Y., Ripke, S., Isaacs A. et al. (2007)) and discordant sib pairs test (DSP) (Boehnke, N. and Langefeld, C. (1998)).

Material and methods

Phenotypic and Genotypic Data. A Discordant Sib-Pair design was implemented for a commercial pig population. The population for the genome scan comprised 1,000 Danish Landrace boars selected from animals subjected to an in-line skatole test by Danske Slagterier. Five hundred boars with high skatole levels, together with 500 full sibs with low skatole levels were chosen from over 6,000 tested animals. We used 0.3 mg as the threshold to define cases for skatole levels. Animals were genotyped using the 7K porcine SNP chip. This is the first SNP chip in pigs and it contains 6,523 Illumina iSelect/Infinium II assays. During quality control, markers with a minor allele frequency (MAF) <0.01 or a call rate <93% were excluded. Also markers deviating from Hardy-Weinberg equilibrium ($p<1x10^{-10}$) using GenABEL (Aulchenko, S.Y., Ripke, S., Isaacs A, (2007)) were excluded. Animals with a call rate < 95% were omitted as well as potential genetic outliers on the basis of the genomic kinship matrix.

Statistical analyses. Trait scores for androstenone and skatole levels were normalised using log transformations to provide trait distributions that were close to normality. To estimate heritabilities and genetic correlations between skatole and androstenone, the following

[*] The Roslin Institute and R(D)SVS, University of Edinburgh, EH25 9PS, Roslin, United Kingdom.
[+] Danske Slagterier, Axeltorv 3, 1609 København V, DENMARK
[†] MRC Human Genetics Unit, Edinburgh, EH4 2XU, United Kingdom.

animal model was used in ASREML (Gilmour, A.R., Cullis, B.R., Welham, S.J. et al. (2002))

$$y = \mathbf{X}b + \mathbf{Z}a + e$$

where y represents the log transformed observations for skatole and androstenone, b is the vector of fixed effects, a is the additive genetic effect, matrices **X** and **Z** are incidence matrices, and **e** is a vector containing residuals. For improved fit and convergence we used the full data with 6178 records for skatole of which 1000 also had androstenone measurements. For the GWAS with pedigree information, we applied the GRAMMAR approach (Aulchenko, S.Y., de Koning, D.J., and Haley, C.S. (2007)): firstly, we obtained corrected residuals using polygenic and fixed effects for log transformed skatole and androstenone levels using Asreml. After that we used the residuals in the GenABEL as response variables for detecting possible associations. We used permutation tests to set empirical thresholds with 100,000 replicates. We also performed principal component analyses (Price, A.L., Patterson, N.J., Robert, M.P. et al. (2006)), to account for possible population stratification, as implemented in JMP Genomics 7 (SAS, 2007) using categorical case/control response variables without pedigree information. The number of PCP to include in the model was chosen from visual inspection of the scree plots. For the top 3 markers coming out of the preceding analyses, we performed further analyses using the comparison of allele counts as proposed for discordant sib pairs by Boehnke, N. and Langefeld, C. (1998). We used both the chi-square and Pearson homogeneity statistic to detect possible association. The Pearson homogeneity statistic could be estimated from a 2 x m table as follows:

$$T^1 = \sum_{j=1}^{m} \frac{(n_{1j} - n_{2j})^2}{n_{1j} + n_{2j}}$$

where n_{ij} stands for counted alleles among cases and controls, i=1,2 for cases and controls, respectively and $j=1\ldots m$ (number of markers). We used 100,000 randomizations where we switched or not switched the phenotype labels of each DSP independently with probability ½ as defined in Boehnke, N. and Langefeld, C. (1998) to asses the significance levels of T^1 under a null hypothesis of no association. Pigs without a genotype and their corresponding sib were removed from the analyses. To estimate the genotypic effects of significant SNPs, they were included in Model (1) as a fixed effect. These estimates were also used to determine the proportion of phenotypic variance explained by the SNPs.

Results and discussion

Visual inspection of the data showed that both androstenone and skatole were highly skewed and significantly deviate from normal distribution. Therefore log transformation was applied to obtain normality for these variables., The statistical model included meat percentage, slaughter weight, herd born, slaughter week and year born as covariates for androstenone (R^2 =9.94) and slaughter weight, meat percentage, weight at end of test and litter size (total born) for skatole (R^2 =3.74) chosen by stepwise regression (Minitab, 2006). The heritability for androstenone was estimated between 0.47 (0.10)-0.49(0.08). Our estimates of heritability of level of skatole from 0.54(0.04)-0.56(0.04) were lower compared with the estimates of Varona, L., Vidal, O., Quintanilla, R., et al. (2005) (0.74 to 0.89) but

corresponded well to estimates from Tajet, H., Andresen, O., and Meuwissen, T.E. (2006) (0.23 to 0.55). The estimated phenotypic correlations between skatole and androstenone were between 0.27(0.03)-0.29(0.03) while the genetic correlations were between 0.40(0.10)-0.41(0.10). Genetic correlations between androstenone and skatole were estimated as 0.40(0.10)-0.41(0.10) which is in agreement with Tajet, H., Andresen, O., and Meuwissen, T.E. (2006) who reported genetic correlations among skatole and androstenone as 0.36 for Landrace pigs.

A genome-wide association scan using 6523 SNPs was carried out using 993 individuals. Following QC, 1701 SNPs were excluded because their call rate was < 93%, 2037 SNPs were excluded with minor allele frequency <1% and a further 150 SNPs due to extreme deviation from Hardy-Weinberg equilibrium ($p < 1x10^{-10}$).On the animal side, 107 pigs were excluded with low call rates (<93%), 2 animals had too high IBS (>95%) and 7 animals had too high heterozygosity (FDR<1%). Another 48 animals were excluded because their genotypic kinship formed a cluster that deviated from the main population, suggesting they may be from a different breed. Following QC, we did association analyses with 2635 SNPs and 829 individuals. We applied the GRAMMAR approach to take the pedigree structure among sibships into account for detecting possible associations. The results show a significant effect of SNPs in the *CYP2E1* region on chromosome 14. We did not find any significant markers in association with androstenone levels. The phenotypic variance explained by *CYP2E_1*, *CYP2E_2*, *TXNL* and *STSG* was 4.79%, 5.19%, 4.11% and 5.97%, respectively, Using the methodology of Price, A.L., Patterson, N.J., Robert, M.P. et al
. (2006), as implemented in JMP Genomics 7, we found that the first 3 principal components explain most of the variation in genotype diversity. These principal components were included for the subsequent case-control analyses. The top ten markers include *CYP2E*, *TXNL* and *STSG* (p<0.056). For the most significant SNPs we also applied the DSP test to take the sampling strategy into account. Genotype counts are given in Table 1, based on affection status of the sibs. We obtained both chi-square and Pearson homogeneity statistics (Table 2) and applied 100,000 permutations to obtain significance levels for the top markers found from previous approaches.

This analysis confirmed the highly significant effects of *STSG*, *CYP2E* and *TXNL*. The different methodologies showed that markers in *CYP2E* region had significant effects on boar taint. These associations should now be confirmed in other populations. Previous linkage studies on QTL from similar population sizes, genotypes and measurements for skatole levels reported on different chromosomes from researchers in United Kingdom and France (Robic, A., Larzul, C., and Bonneau, M. (2008)). Three markers that we found significantly associated with BT from Chromosome 14 are consistent with the finding of a QTL affecting boar taint reported by Lee, G.J., Archibald, A.L., Law, A.S. et al. (2004) although they reported on markers that were not present in this study. Varona, L., Vidal, O., Quintanilla, R., et al. (2005) did not analyse SSC14 but they found a QTL on SSC6. Skinner, T.M., Doran, E., McGivan, J.D. et al. (2005) found evidence for association of *CYP2E1* and back fat skatole levels in commercial pigs but not in experimental crosses for the marker and skatole levels.

Conclusion

One of the assumptions of GWAS case-control studies is independence among and between cases and controls. This assumption will be violated in most animal breeding datasets. In this study we used various approaches to correct the BT data for pedigree effects treating skatole levels as both a quantitative and a categorical trait. Test statistics were higher for the DSP test than for GRAMMAR p values (1.0e-04) because the DSP test explicitly exploits the sib-pair nature of data. Although results showed that the *STSG*, *CYP2E* and *TXNL* were found associated with BT, results should be confirmed independently in other populations.

Table 1: Genotype counts for the CYP2E marker on pig chromosome 14.

	AFFECTED-SIB GENOTYPE		
UNAFFECTED-SIB GENOTYPE	AA	AB	BB
AA	125	40	3
AB	110	121	17
BB	9	22	16

Table 2: Test statistics for the top-3 markers from the GRAMMAR analyses using different tests in the discordant sib-pair analyses .

	All Alleles (*P* value)	DSP allele count (*P* value)	Genotype (*P* value)
CYP2E	18.71 (4e-05)	27.48 ($<10^{-6}$)	25.28 (3.23e-06)
TXNL	17.92 ($<10^{-6}$)	24.60 ($<10^{-6}$)	23.76 (6.91e-06)
STSG	18.80 ($< 10^{-6}$)	32.13 ($<10^{-6}$)	22.69 (1.17e-05)

References

Aulchenko, S.Y., Ripke, S., Isaacs A. *et al.* (2007). *Bioinformatics*, 23:1294-1296.

Aulchenko, S.Y., de Koning, D.J., and Haley, C.S. (2007). *Genetics*, 177: 577-585.

Siegmund, D.O. and Yakir, B. (2007). *The statistics of gene mapping*. Springer.

Boehnke, N. and Langefeld, C. (1998). *Am. J. of Hum. Gen.*, 62:950-961.

Gilmour, A.R., Cullis, B.R., Welham, S.J. *et al*. (2002). *ASReml user guide*.

Lee, G.J., Archibald, A.L., Law, A.S. *et al* (2004). *Anim. Genet.,* 36:14-22.

Minitab, Minitab Release 14 (2006). *Reference manual II*., Minitab Inc., State College.

Price, A.L., Patterson, N.J., Robert, M.P. *et al*. (2006). *Nat. Genet.*, 38: 904-909.

Robic, A., Larzul, C., and Bonneau, M. (2008). *Genet. Sel. Evol.,* 40: 129-143.

SAS Institute Inc. JMP User's Guide (Version 7). (2007). SAS Institute Inc. Cary, NC.

Skinner, T.M., Doran, E., McGivan, J.D. *et al*. (2005). *Anim. Genet.,* 36, 417–22.

Tajet, H., Andresen, O., and Meuwissen, T.E. (2006). *Acta Vete. Sca.*, 48:22–23.

Varona, L., Vidal, O., Quintanilla, R., *et al*. (2005). *J. of Ani. Sci.*, 83, 301-307.