Application of whole genome resequencing in the dissection of QTLs affecting boar taint

Rahul Agarwal[1,2], Maren Van Son[2,3], Matthew Peter Kent[1,2], Sigbjørn Lien[1,2], Eli Grindflek[3] ✉

[1]Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Ås, Norway
[2]Centre for Integrative Genetics (CIGENE), Norwegian University of Life Sciences, Ås, Norway
[3]NORSVIN (The Norwegian Pig Breeders Association), PO Box 504, 2304 Hamar, Norway

Motivation and objectives

Male piglets are usually castrated to remove urine like smell from the meat of some entire male pigs. The urine like smell also known as boar taint. However, castration also affects the reproductive and growth traits, which in turn causes the loss in revenue of pig industries. Andostenone and skatole are the major compounds behind the incidence of boar taint in intact boars. Hence, alternative method of boar taint selection in Norwegian pig breeding scheme requires the detection of functional genetic polymorphisms in association with boar taint without simultaneously interfering with the other traits especially male sex steroids responsible for reproduction and growth in male pigs. .

Methods

Fourty-seven Norwegian male pigs were resequenced using Illumina technology with mean coverage >10x. These animals are representing the whole Norwegian Duroc and Landrace population including the previously studied boar taint animals (Grindflek *et al.*, 2011). About 19.19 billion 100-bp paired end reads were obtained, which in turn filtered to remove duplicate reads, adaptor sequence from the reads, bad quality bases, and reads shorter than cutoff length using perl custom scripts. Remaining reads were mapped to pig build 10.2 reference sequence using the Bowtie2 mapping tool (Langmead *et al.*, 2012). Mapped reads were used for calling single nucleotide polymorphisms (SNPs) within the quantitative trait loci (QTL) regions on SSC13 across 23 Duroc samples and on SSC7 across total 47 samples using the Freebayes variant caller (Erik *et al.*,2012). To eliminate the bad quality SNPs, we applied some stringent filtering upon the detected SNPs. Filtered SNPs was classified to the different functional classes on the basis of the current annotation file available in the pig Ensembl. Of filtered SNPs, the subset of the SNPs will be genotyped using MassARRAY assays and then genotyped SNPs will be tested for associations to androstenone and other male sex steroids using AS-Reml. The BEAGLE version 3.3.1 software will be employed to deduce phase and impute missing genotypes. The estimation of linkage disequilibrium (LD) of the genotyped SNPs and then construction of haplotype blocks will be done with the Haploview software (Barrett, 2009).

Results and discussion

Recent mapping study for finding QTLs underlying boar taint compounds in Norwegian male pig breeds (Grindflek *et al.*, 2011) revealed the significant regions of interest on SSC7 which affect major boar taint compounds and on SSC13 which affect both androstenone and other three male specific sex steroids (testosterone, estradiol and estrone sulfate). Thus, fine mapping of these two regions and detection of underlying functional mutations would be attractive aims to pursue in this study. Two QTL regions on SSC7 are most likely located from 60.67 Mb to 68.40 Mb and 74.80 Mb to 80.53 Mb in build 10.2 with 77 and 46 protein coding genes respectively while region on SSC13 probably located from 25.2Mb to 25.9 Mb and having a total of 10 genes. Therefore, whole genome resequencing of 47 Norwegian Duroc pigs was performed and generated paired end reads were filtered and afterward filtered reads were mapped to latest pig reference genome with 77% mapping percentage. We detected a large number of SNPs within these three QTLs from mapped reads of 23 Duroc and 24 Landrace pigs simultaneously. Initially, many thousands of SNPs (with minimum support of 2 reads across samples) were called within QTLs on SSC7 and on SSC13 respectively. These SNPs was sorted out to remove those SNPs located in repeats and SNPs likely to be false. Finally, a sum of 8,591 and 1,231 SNPs within 2 QTLs on SSC7 and single QTL on SSC13 respectively were sorted out and annotated to exon, intron, 1000bp up-

stream and intergenic regions. A total of 59 non-synonymous SNPs (ns-SNP) , 93 synonymous SNPs (s-SNP), 6231 intergenic SNPs, 1995 intronic SNPs and 121 upstream SNPs were detected within two QTLs on SSC7 using SNPEff version 3.1. Similiarily, in case of SSC 13, a thousand of SNPs assigned under intronic and intergenic region, 60 exonic SNPs including both ns-SNP and s-SNP, and 12 SNPs located 1000bp upstream. The minor allele frequency (MAF) of the SNPs within QTLs on SSC7 and SSC13 was varying from 0.07-0.50 and the average values of total read depth comprising the sum of the read depth of reference allele and alternate allele were 250 and more than 400 in case of SSC13 and SSC7 respectively . As expected, the ratio between the number of transistions and transversions was estimated more than 2 within all the QTLs. So far, a potential candidate gene *ACVR2B* was identified within QTL on SSC13 and responsible for inhibitory activity of steroidogenic acute regulatory protein (*StAR*) and 3β-hydroxysteroid dehydrogenase (3β-*HSD*) which encode important enzymes for the biosynthesis of androstenone and other androgens. Seven SNPs were identified in this particular gene which might play important role in selection.

References

Grindflek E, Lien S, Hamland H, Hansen MH, Kent M et al. (2011) Large scale genome-wide association and LDLA mapping study identifies QTLs for boar taint and related sex steroids. *BMC Genomics* **12**: 362. doi: 10.1186/1471-2164-12-362.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359. doi: 10.1038/nmeth.1923.

Erik Garrison, Gabor Marth (2012) Haplotype-based variant detection from short-read sequencing. http://arxiv.org/abs/1207.3907 (Submitted on 17 Jul 2012 (v1), last revised 20 Jul 2012 (this version, v2)).

Barrett JC (2009) Haploview: Visualization and analysis of SNP genotype data. *Cold Spring Harbor Protocols* **2009**: pdb.ip71 doi:10.1101/pdb.ip71.