Single nucleotide polymorphism discovery in bovine liver using RNA-seq technology

Background
RNA-seq is a useful next-generation sequencing (NGS) technology that has been widely used to understand mammalian transcriptome architecture and function. In this study, a breed-specific RNA-seq experiment was utilized to detect putative single nucleotide polymorphisms (SNPs) in liver tissue of young bulls of the Polish Red, Polish Holstein-Friesian (HF) and Hereford breeds, and to understand the genomic variation in the three cattle breeds that may reflect differences in production traits.

Results
The RNA-seq experiment on bovine liver produced 107,114,4072 raw paired-end reads, with an average of approximately 60 million paired-end reads per library. Breed-wise, a total of 345.06, 290.04 and 436.03 million paired-end reads were obtained from the Polish Red, Polish HF, and Hereford breeds, respectively. Burrows-Wheeler Aligner (BWA) read alignments showed that 81.35%, 82.81% and 84.21% of the mapped sequencing reads were properly paired to the Polish Red, Polish HF, and Hereford breeds, respectively. This study identified 5,641,401 SNPs and insertion and deletion (indel) positions expressed in the bovine liver with an average of 313,411 SNPs and indel per young bull. Following the removal of the indel mutations, a total of 195,3804, 152,7120 and 205,3184 raw SNPs expressed in bovine liver were identified for the Polish Red, Polish HF, and Hereford breeds, respectively. Breed-wise, three highly reliable breed-specific SNP-databases (SNP-dbs) with 31,562, 24,945 and 28,194 SNP records were constructed for the Polish Red, Polish HF, and Hereford breeds, respectively. Using a combination of stringent parameters of a minimum depth of ≥10 mapping reads that support the polymorphic nucleotide base and 100% SNP ratio, 4,368, 3,780 and 3,800 SNP records were detected in the Polish Red, Polish HF, and Hereford breeds, respectively. The SNP detections using RNA-seq data were successfully validated by competitive allele-specific PCR (KASPTM) SNP genotyping assay. The comprehensive QTL/CG analysis of 110 QTL/CG with RNA-seq data identified 20 monomorphic SNP hit loci (CARTPT, GAD1, GDF5, GHRH, GHRL, GRB10, IGBFPL1, IGFL1, LEP, LHX4, MC4R, MSTN, NKAIf1, PLAG1, POUIF1, SDR16C5, SH2B2, TOX, UCP3 and WNT10B) in all three cattle breeds. However, six SNP loci (CCSER1, GHR, KCNIP4, MTSS1, EGFR and NSMCE2) were identified as highly polymorphic among the cattle breeds.

Conclusions
This study identified breed-specific SNPs with greater SNP ratio and excellent mapping coverage, as well as monomorphic and highly polymorphic putative SNP loci within QTL/CGs of bovine liver tissue. A breed-specific SNP-db constructed for bovine liver yielded nearly six million SNPs. In addition, a KASPTM SNP genotyping assay, as a reliable cost-effective method, successfully validated the breed-specific putative SNPs originating from the RNA-seq experiments.