Identifying rare genetic variants of milk genes.

Henriette Pasgaard Bertelsen, Pernille Kræmer Andersen, Rasmus Ory Nielsen, Frank Panitz and Christian Bendixen.

Dept. Of Genetics and Biotechnology, Faculty of Agricultural Sciences, University of Aarhus, Tjele, Denmark

Introduction:

In dairy cattle only a few specific genetic variants are known to affect the quantity and/or quality of milk and DNA-based selection for/against several of these variants has been successfully implemented. The gene variants known at present typically have large effects and are widespread across the cattle population. Rare genetic variants are likely to persist undiscovered even if they have large, either beneficial or detrimental effects and thereby affect the variation in the technological qualities of milk. The recent advances of the next generation sequencing technologies now makes it possible to identify and exploit this particular type of variation. Here we aim to develop an experimental design that enables us to detect rare genetic variation in the beta-lactoglobulin (\(\beta\)-LG) gene from pools of DNA.

Results and discussion:

Several criteria were used in the detection of true SNPs: the minor allele frequency should be above 2%, the read coverage should be above 200, the SNP should not be located within a homopolymer tract and the rare allele should be covered by bidirectional sequencing.

A total of 27648 animals (1152 pools) from 19 different breeds have been sequenced on the next generation sequencing platform GS-FLX provided by Roche. Analysis of the sequence reads revealed a number of putative SNPs of which 20 have been validated by Sanger sequencing. The validated SNPs are positioned in exon 1-5 and in the intronic regions flanking the CDS (Fig. 2).

With our experimental design we have been able to detect both common and rare SNPs in the \(\beta\)-LG gene. Furthermore we were able to calculate the minor allele frequencies for each SNP in each individual breed (Table 1).

Methods:

• In the pooling strategy used, purified DNA from 24 individual samples of similar breed was pooled and separate amplicons covering the first six exons of bovine \(\beta\)-LG as well as some flanking regions were amplified by PCR. Finally the different amplicons obtained from each pool were mixed (Fig. 1).

• Amplicons from each pool were tagged with a 10mer-oligo according to a coded tag ligation protocol allowing for multiplex sequencing.

• All sequence reads were screened for genetic variation using the software GS Reference Mapper provided by Roche. All reads were separately aligned to the reference Bos Taurus chromosome 11 (Btau 4.0). Alignments were required to have a minimum overlap identity of 95 and a minimum overlap length of 90.

• Selected SNPs were validated by Sanger sequencing on individual animals.

Table 1 Minor allele frequencies