Study of Relationships Between Production and Fertility Traits in Dairy Cattle Using Genomic Data

E.C.G. Pimentel[*], M. Tietze[*], H. Simianer[*], F. Reinhardt[†], J. Tetens[‡], G. Thaller[‡], S. Bauersachs[§], E. Wolf[§] and S. König[*]

Introduction

Selection strategies focusing on milk yield in dairy cattle have generally led to a decline in fertility (Pryce *et al.* 2004). This fact is due to an antagonistic relationship between production and reproduction, which has been proven via the estimation of genetic correlations in quantitative genetic studies (e.g. König *et al.* 2009). Novel molecular techniques can be used to get a deeper insight into fertility mechanisms and a detailed explanation of relationships between fertility and production on the gene level. In our studies of gene expression changes in bovine endometrium during the estrous cycle (Bauersachs *et al.* 2005; Mitko *et al.* 2008) we have found a number of genes up-regulated in the luteal phase compared to the ovulatory phase. This up-regulation suggests a potential role of these genes for proper development of the early embryo and for successful pregnancy establishment and maintenance since the steroid hormone progesterone (P4) plays a key role in these reproductive events in the endometrium (reviewed in Bauersachs *et al.* 2008). Importantly, these candidate genes have been recovered in the target species and gender.

The objective of this work was to integrate these findings from functional genomics into association analyses for fertility and production traits in dairy cattle. Hence, this work can make an additional contribution to the evidence of antagonistic relationships between production and reproduction based on a novel concept. Detection of (unexpected) beneficial SNP-genotypes for both production and reproduction within candidate genes would provide the potential to improve production without decreasing fertility.

Material and methods

Data. Genotypic, phenotypic and pedigree information were collected on 2294 Holstein-Friesian bulls. Phenotypes were estimated breeding values (EBV) for non-return rate to 56 d (NRh) and interval from first to successful insemination in heifers (FLh), non-return rate to 56 d (NRc) and interval from first to successful insemination in cows (FLc), interval from calving to first insemination (CFc), days open (DOc), milk yield (Mkg), fat yield (Fkg), protein yield (Pkg), fat percentage (Fpr), protein percentage (Ppr) and somatic cell score (SCS). EBVs for SCS and fertility traits are standardized to a mean of 100 and a standard

[*] Department of Animal Sciences, Georg-August University, 37075 Göttingen, Germany
† Vereinigte Informationssysteme Tierhaltung w.V.. 27283 Verden / Aller, Germany
‡ Institute of Animal Breeding and Animal Husbandry, Christian-Albrechts University, 24098 Kiel, Germany
§ Chair for Molecular Animal Breeding and Biotechnology, and Laboratory for Functional Genome Analysis (LAFUGA), LMU Munich, München, Germany

deviation of 12 points, with higher EBVs being in the desired direction of selection. Samples were successfully (< 3% missing calls) genotyped with the Illumina BovineSNP50 BeadChip. Markers located in unassigned contigs (unknown chromosome and/or position in the BTAU4.0 assembly) or with > 5% missing calls or with minor allele frequency < 0.05 were excluded from the data set. The final number of markers in the data set after quality control was 39'557. 151 genes found to be up-regulated in the endometrium during the luteal phase compared to the ovulatory phase were used as candidate genes. Positions of markers from the panel of SNPs were checked in order to identify the ones that were within physical location of the candidate genes.

Statistical analyses. Association analyses were done using the model $\mathbf{y = Xb + Zu + e}$, where **y** is the vector of EBVs; **b** is the vector of fixed effects, including an overall mean and the marker genotype; **u** is the vector of random polygenic effects and **e** is a random residual term. Marker effects were coded as the number of copies of one of the alleles. Assumed distributions of polygenic effects and the error term were: $\mathbf{u \sim N(0, A\sigma^2_a)}$ and $\mathbf{e \sim N(0, R\sigma^2_e)}$ where **A** is a symmetric matrix which accounts for population structure via relatedness among individuals, **R** is a diagonal matrix of residual variances which were assumed to be inversely proportional to the accuracies of the corresponding EBV, σ^2_a is the additive genetic variance and σ^2_e is the residual variance. Matrix **A** was derived in two alternative ways: i) using pedigree information to calculate numerator relationship coefficients; and ii) using genotype information on the full set of 39'557 markers to compute kinship coefficients, as in Hayes and Goddard (2008). In the former case, matrix **A** included all relevant animals (21,646) in the pedigree, tracing back to 1906.
Association between a marker and a given trait was tested with a Student t-test, contrasting a full model including the marker effect with a reduced model with just the polygenic term. The problem of testing multiple hypotheses was addressed by deriving a significance threshold considering a 5% false discovery rate (FDR, Benjamini and Hochberg 1995).

Results and discussion

From the markers that passed the filtering process, 59 were located within chromosomal segments covered by genes in the list of candidates and were therefore used in the association analyses. As in the study by Hayes and Goddard (2008), marker and pedigree-derived relationship coefficients were highly associated (correlation of 0.78). The two ways of computing matrix **A** provided two different sets of results, with one of the sets almost contained in the other. The following results are from the analysis performed with the kinship matrix built from marker data, which was found to be more conservative. From the 59 SNPs located within candidate genes, 38 were found to be significantly associated to at least one of the six production traits but to none of the fertility traits (results not shown). Pronounced antagonistic relationships between SNP effects for Mkg and percentage traits (Fpr and Epr) were observed, confirming previous findings on the quantitative genetic scale. Hence, breeding strategies to improve both yield and percentage traits simultaneously remain a challenge for animal breeding programs. Special breeding goals with focus either on yield or on percentage traits require genetically different groups of sires.
Fertility traits can be distinguished in those being relevant for the start of the first cycle after calving (CFc), those being relevant for conception (NRh, NRc, FLh, and FLc), and those

combining both aspects (DOc). Seventeen SNPs were found to be significantly associated to at least one of the fertility plus one of the production traits. Estimated allele substitution effects of these SNPs are presented in Table 1.

Table 1: Estimated effects (standard errors) of SNPs that were significantly associated (at a 5% FDR) to at least one fertility plus one production trait

Chr	Gene symbol	Trait											
		Mkg	Fkg	Pkg	Fpr	Ppr	SCS	NRh	FLh	CFc	NRc	FLc	DOc
2	LRP2	41.49	1.47	1.46						-0.90			-1.02
2	IFIH1	-36.23			1.63	0.75						-0.96	
3	HFM1					0.63						-1.33	
3	UGT1A6		-1.35		-1.94				0.73				
4	CDCA7L				0.83			0.63					
5	MGP	52.05	-3.51		-6.30	-1.56	1.24			0.61			
7	COL5A3		-0.98	-0.62	-1.23	-0.89	-0.57		-0.59				
8	SLC1A1	-25.71								0.62	-0.69		
13	PSMF1		0.80		1.80	0.75				-0.54			
17	KSR2		1.00	0.63	0.70	0.28			-0.79				
18	NDRG4		-0.83	-0.75						0.95			0.87
19	BAIAP2	31.82			-1.14	-0.71		-1.37			-0.98		
20	CCNB1	-34.44		-1.24	1.28					-1.07	1.16		-0.78
23	BOLA		1.83	1.24									-0.98
24	NPC1		-1.81	-1.05	-1.61	-0.61					1.30		
28	GALNT2		-0.79		-1.19	-0.41						-0.64	-0.61
29	EED		-1.86	-1.07	-1.05	-0.37		0.83					

We found several SNP alleles having favourable effects on yield traits (Mkg, Fkg, and Pkg), as well as on at least one fertility trait. The favourable SNP allele located within the *IFIH1* gene increased Mkg, and also increased the EBV for FLc, which means a reduction in the interval from first to successful insemination. Also for the SNP located in the *SLC1A1* gene, the same allele has favourable effects on production (Mkg) and conception in cows (NRc). Bauersachs *et al.* (2008) reported that elevated mRNA levels of *SLC1A1* have been detected in human endometrium during the luteal phase, as well as in human placenta.

Antagonistic relationships between NRh and Mkg, as well as between Mkg and NRc, were found for the SNP located in the *BAIAP2* gene, which has an impact on the insulin like growth factor system, and for the SNP located in the cell cycle related gene *CCNB1* (Bauersachs *et al.* 2005). This provides evidence that high milk yield in first lactation is genetically negatively associated with non-return rates in heifers, as well as in cows. Days open combines both aspects of the start of a cycle after calving and conception and can be considered as an ultimate breeding goal for fertility (Pasman *et al.* 2006). Genetic correlations between milk yield in first lactation and days open in first lactation were moderately high and unfavourable in most of conducted quantitative genetic studies (e.g. Strandberg and Danell 1989). Following the results from our study, differentiated breeding strategies by using relevant SNP within candidate genes seem to be possible. Breeding on the same allele located in the *CCNB1* gene improves significantly both Mkg and DOc. The same positive direction was found for the traits Fkg and DOc for the SNP located in the *GALNT2* gene, coding for an N-acetylgalactosaminyltransferase. The beneficial SNPs to be used in breeding would have the combined effect of increasing Mkg or Fkg, and reducing days open.

A similar strategy was discussed by Moe *et al.* (2009) for the simultaneous improvement of antagonistic fertility traits in swine. Remaining estimated effects of SNPs that were significantly associated to at least one yield trait (Mkg, Fkg, or Pkg) plus DOc were unfavourable for practical breeding in dairy cattle. These SNPs were located in the genes *LRP2* and *NDRG4*, and within the segment covered by *BOLA*. As shown by König *et al.* (2009), genome wide selection increases the risk to widen the gap between production and functionality. Sophisticated approaches as suggested in the present study, i.e. combining knowledge from quantitative genetics, genome wide association studies and gene expression studies, have the potential to overcome some of these obstacles. Other studies exist which focused on SNPs located in specific regions of the genome, but mostly applying whole-genome scans for QTL mapping. Targeting relevant SNPs within genes derived from gene expression studies is a novel concept in the context of animal breeding strategies. This strategy has been successfully applied in human genetics for functional selection of SNPs associated with prostate cancer (Xu and Taylor 2009).

Conclusion

While most of quantitative genetic studies proved a genetic antagonism between yield traits and functional traits (in our study fertility and SCS), improvements in both production and functionality may be possible when focusing on a few relevant SNPs. Investigations combining input from quantitative genetics and functional genomics with association analysis may be applied for the identification of such SNPs.

Acknowledgements

This study is part of the projects FUGATO-plus GenoTrack and COMPENDIUM and was financially supported by the German Ministry of Education and Research, BMBF, the Förderverein Biotechnologieforschung e.V. (FBF), Bonn, and Lohmann Tierzucht GmbH, Cuxhaven. The transcriptome studies revealing candidate genes for fertility were performed with financial support from the Deutsche Forschungsgemeinschaft (Research Unit "Mechanisms of embryo-maternal communication") and the BMBF (FUGATO Fertilink).

References

Bauersachs, S. *et al.* (2005). *J. Mol. Endocrinol.*, 34:889–908.

Bauersachs, S. *et al.* (2008). *Exp. Clin. Endocrinol. Diabetes.*, 116:371–384.

Benjamini, Y. and Hochberg, Y. (1995). *J. R. Stat. Soc. B*, 57:289–300.

Hayes, B.J. and Goddard, M.E. (2008). *J. Anim. Sci.*, 86:2089–2092.

König, S. *et al.* (2009). *J. Dairy Sci.*, 92:382–391.

Mitko, K. *et al.* (2008). *Reproduction*, 135:225-240.

Moe, M. *et al.* (2009). *BMC Genetics*, 10:32.

Pasman, E. *et al.* (2006). *Interbull Bull.*, 34:34–37.

Pryce, J.E. *et al.* (2004). *Livest. Prod. Sci.*, 86:125–135.

Strandberg, E. and Danell, B. (1989). *Acta Agric. Scand.*, 39:203–215.

Xu, Z. and Taylor, J.A. (2009). *Nucleic Acids Res.*, 37:W600–W605.