Genome-Wide Mapping Of Breed Differences – A Reciprocal Case-Control Design

I. Medugorac[*], S. Rothammer[*], I. Russ[†], M. Förster[*]

Introduction

Human and nature shaped present breeds and strains of farm animals by strong artificial and natural selection over thousands of generations. As a result, the genomes of domesticated animals contain hundreds of regions whose variability profiles indicate a recent positive natural and/or artificial selection. Most of the underlying genes and advantageous mutations are still unknown. The current most prominent surveys of selective events rely on three patterns of variation caused by a beneficial variant with subsequent rapid increase of prevalence in a population: (i) Long haplotypes associated with beneficial variant; (ii) Frequency of beneficial and neighboring alleles higher than expected under genetic drift; (iii) higher differentiated alleles between populations than anticipated for evolving alleles under neutrality (Grossman *et al.* (2010)). In farm animals with a divergent selection, these signals can be stronger than in humans (Andersson and Georges (2004)), and alleles under selection can be derived as well as ancestral. We introduced and tested here a reciprocal case control design between breeds in order to map the preferable strongest selection signals and to avoid foreseeable problems with the definition of derived alleles as well as the selection on standing variation.

Material and methods

Animals, breeds and markers. A set of 494 animals from ten cattle breeds was genotyped with the Illumina BeadChip (BovineSNP50). These ten breeds were prearranged into four selection trends: (i) Two virtually unselected breeds with neither recording system, nor common breeding goals (Anatolian Black (ABB), Illyrian Mountain Buša (ILB) (Medugorac *et al.* (2009))); (ii) Four typical dual-purpose cattle breeds of the alpine regions monitored by traditional breeding organizations with an advanced recording system (Murnau-Werdenfelser (MWF), Original Braunvieh (OBV), Franken Gelbvieh (FGV) and German Fleckvieh (DFV)); (iii) Two breeds highly specialized in dairy production (Braunvieh upgraded by Brown Swiss, Red Holstein (RH)); (iv) Two beef breeds with differing exterior and breeding goals (Blanc-Bleu Belge (BBB), Galloway (GLW)). We sampled and genotyped 47 to 55 possibly unrelated animals per breed and excluded animals showing a high realized relationship (genome-wide IBD>0.20) to one or more animals within the same breed. Afterwards, we used 23 to 50 animals per breed (call-rate >0.98) excluding any stratification within breeds (Table 1). The number of SNPs was reduced due to homozygosity and

[*] Chair of Animal Genetics and Husbandry, LMU Munich, Veterinaerstr. 13, 80539 Munich, Germany
[†] Tierzuchtforschung e.V. München, Senator-Gerauer-Straße 23. D-85586 Grub, Germany

plausibility control. In the end, 47,003 markers were retained with a marker call-rate >0.90, with at least two heterozygous animals over ten breeds and exclusion of paternity problems in families with confirmed paternity.

Table 1: Sample collection: Breed name and code, selection trend (Selection), country of sampling, native origin in parenthesis (origin) and sample size (N). The number of effectively unrelated animals is in parenthesis

Breed	Code	Selection	Origin	N
Anatolian Black	ABB	unselected	Turkey (Anatolia)	49 (43)
Illyrian Mountain Buša	IMB	unselected	Albania	45 (41)
Murnau-Werdenfelser	MWF	dairy-beef	Germany	53 (23)
Original Braunvieh	OBV	dairy-beef	Germany and Switzerland	48 (32) [a)]
Franken Gelbvieh	FGV	dairy-beef	Germany	50 (44)
Fleckvieh	DFV	dairy-beef	Germany	55 (50)
Braunvieh upgraded by Brown-Swiss	BBV	dairy	Germany and world-wide	50 (43)
Red Holstein	RH	dairy	Germany and world-wide	50 (46)
Blanc-Bleu Belge	BBB	beef	Belgium and Denmark	47 (28) [b)]
Galloway	GLW	beef	Germany (Scotland)	47 (30) [c)]

To avoid stratification problems, we only used unrelated animals for R-CCD originating from [a)] Switzerland, [b)] from Belgium and [c)] only black strain of Galloway.

Statistical analyses. Pair-wise genome-wide IBD were estimated by the program *PLINK* (Purcell *et al*. (2007)). Charlier *et al.* (2008) presented two different methods for homozygosity mapping of monogenic autosomal recessive disorders in a classical case control design named *ASSHOM* and *ASSIST*. Here, we adapted *ASSHOM* for mapping breed-specific chromosomal regions with fixation surrounded by long haplotypes. We used each breed *vice versa* as case and control and called this a Reciprocal Case-Control Design (R-CCD). To devise the chromosome and genome-wide significance threshold, we performed a permutation test. Unlike Charlier *et al.* (2008), we permutated complete markers along each of the 29 autosomes of case and control group, simultaneously. To receive an independent chain of test statistic values we only used each fourth position along chromosomes with permutated markers. In using a distinct distribution of test statistics for each reciprocal case-control pair, we accounted for the level of inbreeding and the marker quality of the case-control pairs and corrected for testing of 47,003 markers. The corresponding genome-wide *P* values were determined from all 290,000 permutations (10,000 for each the 29 autosomes). To extract the most determining breed differences we applied the following procedure of five steps: (i) Division of the bovine genome into 1,000 nearly equal segments (~2,540,000bp), obtaining an SNP-number per window comparable to Pickrell *et al.* (2009) (ii) For each breed pair and each segment the highest test statistic value was saved, (iii) 1,000 segments were sorted according to this, (iv) For each breed used as case the ten (0.01) most significant segments for each of the nine independent controls were selected (v) For each of the ten case breeds genome-wide significant segments were selected confirmed by at least three controls.

Results and discussion

Realized relationships (genome-wide IBD) over 0.20 were observed, although we collected representative samples for each of the ten breeds and avoided closely related animals. The successive removal of animals highly related to individuals of the same breed resulted into a substantial reduction of the sample sizes, especially in the endangered MWF cattle with a small effective population size (Table 1). The population structure analysis (results not shown) detected significant stratification in OBV, BBB and GLW. Thus, for the R-CCD study, we only used OBV animals from Switzerland and BBB animals from Belgium. We also divided GLW in black and belted strains and used the larger group of black GLW (N=30) for genome-wide tests and the smaller group of belted GLW (N=6) only exemplarily.

Within our study, there are three well defined and genetically verified phenotypes: The absence of horns (polled locus) selected in GLW, belt pattern selected in belted GLW, and double-muscling selected in BBB cattle breed. Using the black GLW as case group and nine different horned breeds as control groups, we detected a significant signal (P<0.0001 genome-wide) matching exactly with the polled locus on the proximal part of *Bos taurus* autosome 1 (BTA01; Drögemüller *et al.* (2005)). We mapped the polled gene in a region of 0.312 Mega base (Mb) (1.479-1.792Mp), even finer than currently published (Drögemüller *et al.* (2005)). Using only six belted GLW animals as case group and any other breed as control, we confirmed the position of the belted gene on BTA03 (Drögemüller *et al.* (2009)) as well as the position of the polled locus on BTA01. Using black GLW as control group led to finer mapping of the belted gene (125.160-125.516Mb) on BTA03 and to absence of the signal on BTA01. Finally, with BBB as case and other nine breeds as control, we recovered (P<0.0001) the position of the Myostatin gene, causal for double-muscling (Grobet *et al.* (1997)), on BTA02.

These three examples were highly significant at the genome-wide level ($P< 3.45 \times 10^{-6}$), i.e. none of 290,000 permutations resulted in such high test statistic values. But these were not the most significant signals, neither for the comparisons mentioned above nor for the remaining breed comparisons. With each of the ten breeds alternating as case and control group, we performed 90 independent genome-wide homozygosity mapping scans. For each reciprocal breed pair the complete genome-wide permutation test was done. Applying the very conservative procedure of five steps, we detected 68 chromosomal segments strongly selected in one or more breeds. The most significant breed differences concerned 24 autosomes, exceptions were BTA12, 15, 27, 28 and 29. Most divergent chomosomes were BTA04 (5 segments), BTA06 (5), BTA07 (6) and BTA08 (5). Two segments were selected in five breeds. The first one was on BTA06 at around 37Mb, selected in all five dual-purpose breeds, including BBV, all originating from the alpine region. The second one was located on BTA19 around 27Mb selected in MWF, OBV, BBV, RH and GLW in comparison with IMB and ABB (unselected breeds). One segment was selected in four, three in three breeds and twelve in two breeds. The remaining 50 genome-wide significant signals were detected in a single breed as case and confirmed by three or more independent controls. Most selection signals (76%) observed in breeds under strong selection were detected by using one or both unselected breeds as control group and confirmation by an additional breed. On average, there were 12 strongly selected segments in each of North-Western breeds (BBB, RH and GLW), 10 in each of alpine breeds (MWF, OBV, BBV, DFV, and FGV) and 6 in

each of two breeds without common breeding goals (ABB and IMB, *quasi* unselected). Interestingly, the strongest selection signal in IMB was on BTA23 within BoLA (Bovine lymphocyte antigen) complex at around 28Mb. This signal was confirmed by six independent control groups (OBV, BBV, DFV FGV, MWF and RH). Comparing ABB and IMB, we detected only two selection signals. These were selected in ABB and confirmed by four and six control groups including IMB.

Conclusion

It was shown, that the Reciprocal Case-Control Design developed here can be successfully used for mapping the most significant selection signals in divergent selected farm animal breeds. Only strict selected and fixed dissimilarities surrounded by long haplotypes were detected. The applied method does neither depend on any assumption regarding the mode of inheritance nor on derived and ancestral alleles nor selection on standing variation. Although the available SNP-Chip was designed to be highly informative in cosmopolitan breeds like Holstein (ascertainment bias), our results demonstrate the benefit of a combination of strongly selected breeds with highly diverse autochthonous breeds. Subsequent fine mapping and high throughput sequencing procedures will include the here mapped selection signals representing positional candidate genes under adaptive selection. Furthermore, combining different test statistics, as suggested by Grossman *et al.* (2010), will be pathbreaking for our further analyses.

Acknowledgements

The project was gratefully funded by Deutsche Forschungsgemeinschaft project ME3404/2-1. We thank all breeders and breeding associations who sent us samples free of charge to support this study. In particular, we thank Weilheimer Zuchtverbände eV (DE), the Bavarian Gene Reserves and LfL, Grub (DE), Rinderzuchtverband Würzburg (DE), Rinder-Union West eG, Münster (DE), Belgian Blue Group, Ciney (BE), Fleischrinderverband Bayern, Ansbach (DE) and Swissgenetics Zollikofen, (CH).

References

Andersson, L. and Georges, M. (2004). *Nat. Rev. Genet*., 5: 202-212

Charlier, C., Coppieters, W., Rollin, F. *et al.* (2008). *Nat. Genet*., 40: 449-454

Drögemüller, C., Wöhlke, A., Mömke, S. *et al.* (2005). *Mammalian Genome*, 16: 613-620

Drögemüller, C., Demmel, S., Engensteiner, M. *et al.* (2009). *Animal Genetics*.

Grobet, L., Martin, L. J., Poncelet, D. *et al*. (1997). *Nat. Genet*., 17: 71-74

Grossman, S. R., Shylakhter, I., Karlsson, E. K. *et al*. (2010). *Science*, 327: 883-886

Medugorac, I., Medugorac, A., Russ, I. *et al*. (2009). *Molecular Ecology*, 18: 3394-3410

Pickrell, J. K., Coop, G., Novembre, J. *et al*. (2009). *Genome Res*., 19: 826-837

Purcell, S., Neale, B., Todd-Brown, K. *et al*. (2007). *Am. J. Hum. Genet*., 81: 559-75

Sabeti, P.C., Reich, D. E., Higgins, J. M. et al. (2002), Natur, 419: 832-837