doi:10.1111/j.1365-2052.2008.01769.x

Characterization of a novel microdeletion polymorphism on BTA5 in cattle

G. E. Liu*, R. W. Li*, T. S. Sonstegard*, L. K. Matukumalli*,†, M. V. G. B. Silva*,‡ and C. P. Van Tassell*

*Bovine Functional Genomics Laboratory, USDA-ARS, ANRI, BARC-East, Beltsville, MD 20705, USA. †Bioinformatics and Computational Biology, George Mason University, Manassas, VA 20110, USA. ‡Embrapa Dairy Cattle, Juiz de Fora, MG 36038-330, Brazil

Summary

We present a detailed breakpoint mapping and population frequency analysis of a 214-kb microdeletion that removes multiple olfactory receptor genes. Using progressive rounds of PCR assays, we mapped the upstream and downstream breakpoints of this microdeletion event to ~1 and 12 kb genomic regions, respectively. We developed PCR-based genotyping assays, characterized a dairy cattle panel of 96 samples and found that the frequency of the deletion allele was over 51%. Our results indicated that this microdeletion is an ancient event occurring in one of the earlier founders, and that it has been stably inherited across generations in the North American dairy cattle population.

Keywords breakpoint, cattle genome, deletion, population frequency, structural variants.

Several recent studies have brought about increasing interest in genomic structural variation such as large-scale insertions/microdeletions (>1 kb) or copy number variation (CNV) in humans and rodents (Sebat *et al.* 2004; Redon *et al.* 2006; Graubert *et al.* 2007; Korbel *et al.* 2007). Earlier cattle studies only identified a couple of such local disease-causing microdeletions ranging from 2 kb to over 100 kb (Ohba *et al.* 2000; Drogemuller *et al.* 2001). We initiated a systematic study of cattle CNV using array comparative genomic hybridization (array CGH) followed by real-time PCR confirmation (Liu *et al.* 2008). Our initial data demonstrated that significant amounts of CNV exist in cattle; many CNVs are common both across diverse cattle breeds and among individuals within a breed. Although these CNVs were presumably associated with individual health and disease, the structure and population characteristics of these variants remained largely unexplored. One of the deletions was discovered within BTA5 at positions 57 103 490–57 454 046 (Btau_3.1) by array CGH using multiple probes. This 350-kb region in the olfactory receptor (OR) gene cluster produced significantly lower ratios compared with the flanking regions. We present here a PCR-based breakpoint fine-mapping analysis and a population frequency survey of this microdeletion.

Address for correspondence

G. E. Liu, Bovine Functional Genomics Laboratory, USDA-ARS, ANRI, Building 200, Room 124B, BARC-East, Beltsville, MD 20705, USA.
E-mail: george.liu@ars.usda.gov

Accepted for publication 18 May 2008

In the Beltsville Agricultural Research Center (BARC) Dairy Cattle Panel (BDCP 1.0), semen DNA samples were collected from 96 sires used for artificial insemination (Sonstegard & Van Tassell 2004; T.S. Sonstegard, personal communication). From this dairy cattle panel, three Holstein bulls (anonymized as Hs1, Hs2 and Hs3), representing major branches of the Holstein breed, were screened by array CGH as described (Graubert *et al.* 2007). Genomic DNA was purified from sperm, whole blood and ear notch as described (Sonstegard *et al.* 2000). DNA samples from the sequenced cow, Hereford L1 Dominette 01449 (whole blood: Dt), or its sire, L1 Domino 99375 (semen: Do) were used as the reference.

PCR-based assays were used to detect quantitative differences of genomic segments associated with deletions and insertions. The copy numbers were determined by real-time PCR using the TaqMan and/or SYBR Green detection chemistries as recommended by manufacturers. The cycle thresholds (CTs) from multiple randomly selected genome-wide loci from L1 Domino 99375 were averaged (mean ± standard deviation were 22.3 ± 1.3, $n = 61$) and used to infer copy number. Triplicate CT values were averaged and then normalized against the control primer (C3 in Fig. 1 and Table 1). Assuming that there were two copies of DNA in the control regions, the relative copy number for each test region was calculated as $2^{(1+\mathrm{ddCT})}$, where $\mathrm{ddCT} = (\mathrm{CT}_{\mathrm{reference,T}} - \mathrm{CT}_{\mathrm{reference,C3}}) - (\mathrm{CT}_{\mathrm{sample,T}} - \mathrm{CT}_{\mathrm{sample,C3}})$. Using this method, the microdeletion events in Hs3 and Domino (+/+), Hs2 (+/−) and Hs1 (−/−) were characterized. Therefore, these four DNA samples were used as calibration controls. The corrected CTs

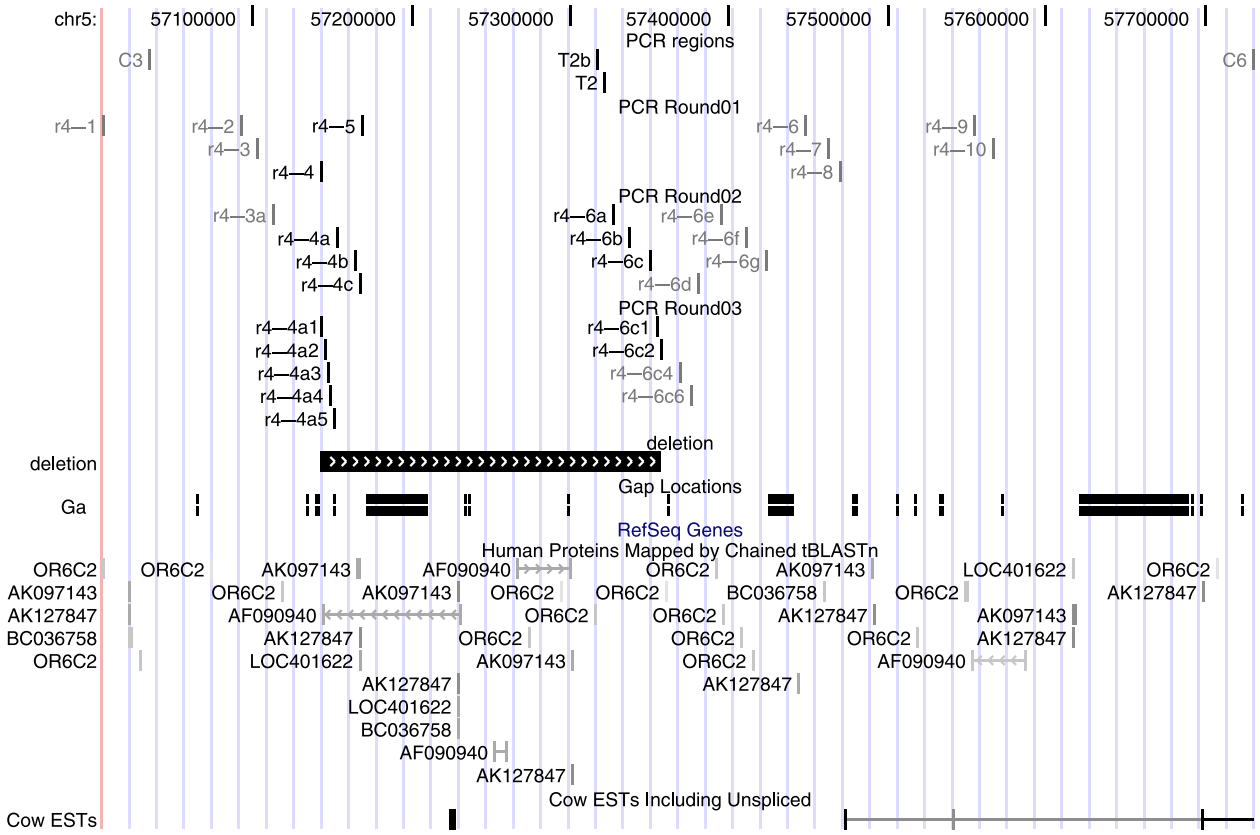


Figure 1 Breakpoint mapping and annotation tracks. A deletion event in BTA5 at 57 Mb that had been identified by whole-genome array CGH was confirmed by real-time PCR. Progressive rounds of real-time PCR mapped upstream and downstream breakpoints with the following intervals: the upstream breakpoint was in an interval of 709 bp between r4-4 and r4-4a1 (BTA5: 57 142 127–57 142 836) and the downstream breakpoint was in an interval of 11631 bp between r4-6c2 and r4-6c4 (BTA5: 57 356 666–57 368 297). Gap, gene and EST annotation tracks are displayed.

from tested loci in all samples tended to form three discrete clusters in the distribution histogram: i.e. ddCT of 0.0, −1.2 and −13.8, corresponded to diploid (+/+, two copies), hemizygous (+/−, one copy) and homozygous deletion (−/−, null alleles) respectively. One-way ANOVA using the *F*-test comparing these three genotypes produced a *P*-value < 0.0001. Assuming a conversion factor of 6.6 pg of DNA per diploid genome, the input amount of 20 ng would correspond to 6000 copies of haploid genome. The positive control cycles were normally around cycle 22; therefore, if the test cycle was more than 36, the DNA segment was probably not present because a ddCT of −13.8 suggested the input DNA was less than one copy (i.e. $6000/2^{13.8} = 6000/14263$). The PCR products were subjected to electrophoresis on 1.0% agarose gels and then visualized by the GelDoc System (BIO-RAD). The deletion locus with various annotation tracks was generated using the UCSC Genome Browser.

To identify the breakpoints of this microdeletion, several rounds of progressive real-time PCR experiments were performed on genomic DNA from Domino and Hs1, using primers designed from the Btau_3.1 reference genome sequence (Fig. 1 and Table 1). In Hs1, the breakpoints were mapped to the following genomic regions: the upstream breakpoint was in an interval of 709 bp between r4-4 and r4-4a1 (57 142 127–57 142 836 on BTA5) and the downstream breakpoint was in an interval of 11631 bp between r4-6c2 and r4-6c4 (57 356 666–57 368 297 on BTA5). Although the downstream breakpoint was less well defined, the minimal deleted region in Hs1 extended ~214 kb (57 142 836–57 356 666 on BTA5) (Fig. 1 and Table 1). The genome mapping information was also confirmed on Btau_4.0. Based on BLAT mapping results, except for the offset of 6096739 bp, there was no detectable change in the relative position of any primer within this OR region between Btau_3.1 and Btau_4.0.

The BDCP 1.0 comprises the most influential Holstein (84), Jersey (8), Brown Swiss (2) and Guernsey (2) sires in North America. PCR primers (T2b) in the middle of the microdeletion were initially designed to survey its frequency in these 96 samples. Using the ddCT distribution calibrated by Hs3 and Domino (+/+), Hs2 (+/−) and Hs1 (−/−), these 96 samples were categorized into one of the three classes (+/+, +/− or −/−). Their counts and percentages within breeds are summarized in Fig. 2. In 84 Holsteins, 21 samples were classified as +/+, 40 as +/− and 23 as −/−. As these samples represent the North American Holstein

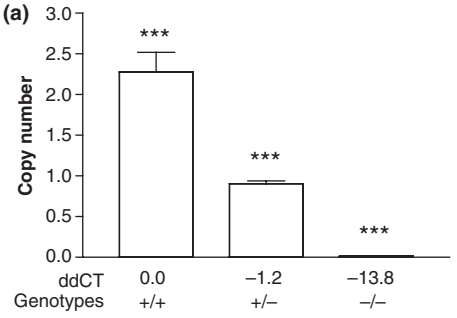
Table 1 Real-time PCR primer locations and results.

	Forward		Reverse			
Region	Begin	End	Begin	End	Domino	Hs1
r4-1	57 005 940	57 005 963	57 006 113	57 006 133	+	+
C3	57 034 288	57 034 311	57 034 486	57 034 508	+	+
r4-2	57 092 908	57 092 930	57 093 115	57 093 138	+	+
r4-3	57 101 762	57 101 785	57 101 981	57 102 002	+	+
r4-3a	57 111 720	57 111 745	57 112 139	57 112 160	+	+
r4-4	57 141 908	57 141 927	57 142 104	57 142 127	+	+
r4-4a1	57 142 836	57 142 861	57 143 056	57 143 080	+	−
r4-4a2	57 145 066	57 145 091	57 145 169	57 145 193	+	−
r4-4a3	57 146 865	57 146 891	57 147 055	57 147 084	+	−
r4-4a4	57 148 290	57 148 315	57 148 328	57 148 357	+	−
r4-4a5	57 150 828	57 150 856	57 150 984	57 151 013	+	−
r4-4a	57 152 756	57 152 775	57 152 912	57 152 933	+	−
r4-4b	57 163 374	57 163 397	57 163 569	57 163 593	+	−
r4-4c	57 167 304	57 167 329	57 167 527	57 167 550	+	−
r4-5	57 168 576	57 168 597	57 168 712	57 168 734	+	−
T2b	57 316 239	57 316 262	57 316 425	57 316 449	+	−
T2	57 321 747	57 321 768	57 321 906	57 321 929	+	−
r4-6a	57 326 181	57 326 205	57 326 402	57 326 425	+	−
r4-6b	57 336 221	57 336 248	57 336 392	57 336 418	+	−
r4-6c	57 350 119	57 350 143	57 350 319	57 350 343	+	−
r4-6c1	57 353 603	57 353 628	57 353 746	57 353 767	+	−
r4-6c2	57 356 425	57 356 453	57 356 634	57 356 666	+	−
r4-6c4	57 368 297	57 368 321	57 368 589	57 368 619	+	+
r4-6c6	57 376 190	57 376 216	57 376 333	57 376 355	+	+
r4-6d	57 379 861	57 379 886	57 380 050	57 380 077	+	+
r4-6e	57 393 672	57 393 699	57 393 873	57 393 900	+	+
r4-6	57 409 847	57 409 871	57 409 999	57 410 023	+	+
r4-6g	57 423 719	57 423 742	57 423 848	57 423 877	+	+
r4-6	57 447 903	57 447 926	57 413 723	57 413 744	+	+
r4-7	57 461 407	57 461 427	57 461 616	57 461 639	+	+
r4-8	57 468 876	57 468 899	57 469 009	57 469 032	+	+
r4-9	57 554 585	57 554 604	57 554 769	57 554 792	+	+
r4-10	57 565 924	57 565 947	57 566 119	57 566 142	+	+
C6	57 730 198	57 730 217	57 730 321	57 730 344	+	+

All genome mapping information was based on Btau_3.1 and was confirmed on Btau_4.0.

population, the frequency of this microdeletion is ~51% in the North America Holsteins. Despite the relatively small sample size, the results from the other three breeds (Jersey, Brown Swiss and Guernsey), when combined with those from Holsteins, began to reveal that this microdeletion exists in various dairy breeds. Furthermore, a larger ongoing survey of more beef breeds and more individuals confirmed this observation (G. E. Liu *et al.*, unpublished results). To define whether these deletions are the same event, 16 selected animals with null alleles (−/−; 12 Holsteins, three Jerseys, one Brown Swiss) were subjected to PCR assays with multiple primers flanking both upstream and downstream breakpoints. The 12 Holstein animals included Hs1, its grand sire, its great grand sire and two paternally related individuals. The other seven Holstein samples were not paternally related to Hs1, and included two sire-offspring pairs. Interestingly, all animals produced the same PCR results as Hs1, which strongly indicates these events may derive from a single ancient event and that was stably inherited across generations. Those paternally unrelated individuals may have obtained this ancient event maternally. The microdeletion events were also overlaid with the known Holstein pedigree information. All inheritance patterns within the pedigree followed the Mendelian laws of inheritance with only one exception, which may be an error in the pedigree tree and deserves further investigation.

In the 214-kb locus, multiple sequence features were detected, including genes, assembly gaps and ESTs (Fig. 1). As the bovine gene annotation is still in its early phase, this analysis was performed on human orthologous genes. In addition to several cDNA clones of unknown function, the human protein OR6C2 mapped *in silico* to this locus four times. OR genes are known to be involved in human and mouse CNVs. The OR family is a class of G protein-coupled



(b)

Genotypes	+/+	+/−	−/−	Subtotal
Holstein	21 (25.0%)	40 (47.6%)	23 (27.4%)	84
Jersey	0 (0.0%)	5 (62.5%)	3 (37.5%)	8
Guernsey	1 (50.0%)	1 (50.0%)	0 (0.0%)	2
Brown Swiss	1 (50.0%)	0 (0.0%)	1 (50.0%)	2
Total	23 (24.0%)	46 (47.9%)	27 (28.1%)	96

Figure 2 Microdeletion genotyping and population frequencies in dairy cattle. (a) PCR-based assays were used to genotype 96 dairy cattle samples. Three discrete clusters in the distribution histogram of ddCT were detected at 0.0, −1.2 and −13.8, corresponding to the diploid (2.28 ± 0.24, +/+, two copies), hemizygous (0.90 ± 0.03, +/−, one copy) and homozygous deletion (0.00 ± 0.00, −/−, null alleles) genotypes respectively. The copy numbers and their standard errors of the means were plotted. One-way ANOVA using the *F*-test comparing copy numbers in these three genotypes produced a *P*-value < 0.0001 (labelled by ***). (b) The counts and percentages of three genotypes within four dairy cattle breeds were calculated.

receptors mediating signal transduction to OR neurons. The OR family is the largest mammalian gene family and may comprise 3% of the mammalian genome. Humans were estimated to have ~400 functional genes and ~600 pseudogenes; mice were estimated to have ~1200 functional genes and ~300 pseudogenes (Young *et al.* 2002) and rats were estimated to have ~1400 functional genes and ~600 pseudogenes (Gibbs *et al.* 2004). The involvement of OR genes in CNVs may relate to the fitness of the host, probably through generating new genes (Eichler 2001). The large number of OR genes and their redundancy in function provides the host with a system to recognize an almost unlimited number of molecules. At this locus, the largest gap (57 171 598–57 210 393 on BTA5) is a 'bridged' gap of 38796 bp, i.e. the relative order and orientation of the contigs on either side of the gap is defined by read-pair data. Multiple cattle ESTs and production QTL were also detected in this locus (data not shown).

In conclusion, our results indicate that this microdeletion is an ancient event that occurred in one of the earlier founders, and that it has been stably inherited across generations in the dairy cattle population. There is a remote possibility that the nature of this microdeletion is more complicated than we currently perceive; multiple recurrent deletion events may have occurred independently within these two breakpoint hotspots. It is also possible that other smaller deletion events in this locus were overlooked. Despite these possibilities, it is unlikely that the microdeletion we detected is an isolated event that occurred within a subpopulation of sperm during spermatogenesis. In summary, our data confirmed that a deletion polymorphism exists in cattle and that it is widespread across dairy cattle breeds and among individuals within a breed. We are investigating the impact of this deletion, its association with neighbouring SNPs and its possible biological functions. Annotation of CNVs in the cattle genome combined with sequence- and population-based analyses provide an important resource that will help define the genetic basis of complex traits.

Acknowledgements

We thank G. Wei, H. Preston, J. Boone and F. Koroma for their technical support. This work was financially supported by National Research Initiative (NRI) Grant No. 2007-35205-17869 from USDA-CSREES and Project 1265-31000-090-00D from USDA-ARS.

References

Drogemuller C., Distl O. & Leeb T. (2001) Partial deletion of the bovine *ED1* gene causes anhidrotic ectodermal dysplasia in cattle. *Genome Research* **11,** 1699–705.

Eichler E.E. (2001) Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends in Genetics* **17,** 661–9.

Gibbs R.A., Weinstock G.M., Metzker M.L. *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428,** 493–521.

Graubert T.A., Cahan P., Edwin D. *et al.* (2007) A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genetics* **3,** e3.

Korbel J.O., Urban A.E., Affourtit J.P. *et al.* (2007) Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318,** 420–6.

Liu G.E., Van Tassell C.P., Sonstegard T.S. *et al.* (2008) Detection of germline and somatic copy number variations in cattle. *Developments in Biologicals* **132,** 231–7.

Ohba Y., Kitagawa H., Kitoh K. *et al.* (2000) A deletion of the *paracellin-1* gene is responsible for renal tubular dysplasia in cattle. *Genomics* **68,** 229–36.

Redon R., Ishikawa S., Fitch K.R. *et al.* (2006) Global variation in copy number in the human genome. *Nature* **444,** 444–54.

Sebat J., Lakshmi B., Troge J. *et al.* (2004) Large-scale copy number polymorphism in the human genome. *Science* **305,** 525–8.

Sonstegard T.S. & Van Tassell C.P. (2004) Bovine genomics update: making a cow jump over the moon. *Genetical Research* **84,** 3–9.

Sonstegard T.S., Garrett W.M., Ashwell M.S. *et al.* (2000) Comparative map alignment of BTA27 and HSA4 and 8 to identify conserved segments of genome containing fat deposition QTL. *Mammalian Genome* **11,** 682–8.

Young J.M., Friedman C., Williams E.M. *et al.* (2002) Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Human Molecular Genetics* **11,** 535–46.