



Discovery, characterization and validation of single nucleotide polymorphisms within 206 bovine genes that may be considered as candidate genes for beef production and quality

J. L. Williams^{*†}, S. Dunner[‡], A. Valentini[§], R. Mazza[¶], V. Amarger^{#, **}, M. L. Checa[‡], A. Crisà[§], N. Razzaq[†], D. Delourme^{#, **, ††}, F. Grandjean^{#, **, ††}, C. Marchitelli[§], D. García[‡], R. Pérez Gomez[‡], R. Negrini^{††}, P. Ajmone Marsan^{††} and H. Levéziel^{#, **}

^{*}Parco Tecnologico Padano, Via Einstein, Polo Universitario, 26900 Lodi, Italy. [†]Division of Genetics and Genomics, Roslin Institute, Roslin, EH25 9PS, Scotland, UK. [‡]Dpto de Producción Animal Facultad de Veterinaria, 28040 Madrid, Spain. [§]Dipartimento di Produzioni Animali, Università della Tuscia, via De Lellis, 01100 Viterbo, Italy. [¶]Latemar Unit, Istituto di Zootecnica, Università Cattolica del S. Cuore, via E. Parmense 84, 29100 Piacenza, Italy. [#]INRA, UMR 1061, F-87000 Limoges, France. ^{**}Université de Limoges, UMR 1061, F-87000 Limoges, France. ^{††}Istituto di Zootecnica, Università Cattolica del S. Cuore, via E. Parmense 84, 29100 Piacenza, Italy

Summary

A large number of putative single nucleotide polymorphisms (SNPs) have been identified from the bovine genome-sequencing project. However, few of these have been validated and many will turn out to be sequencing artefacts or have low minor allele frequencies. In addition, there is little information available on SNPs within coding regions, which are likely to be responsible for phenotypic variation. Therefore, additional SNP discovery is necessary to identify and validate polymorphisms both in specific genes and genome-wide. Sequence-tagged sites within 286 genes were resequenced from a panel of animals representing a wide range of European cattle breeds. For 80 genes, no polymorphisms were identified, and 672 putative SNPs were identified within 206 genes. Fifteen European cattle breeds (436 individuals plus available parents) were genotyped with these putative SNPs, and 389 SNPs were confirmed to have minor allele frequencies above 10%. The genes containing SNPs were localized on chromosomes by radiation hybrid mapping and on the bovine genome sequence by BLAST. Flanking microsatellite loci were identified, to facilitate the alignment of the genes containing the SNPs in relation to mapped quantitative trait loci. Of the 672 putative SNPs discovered in this work, only 11 were found among the validated SNPs and 100 were found among the approximately 2.3 million putative SNPs currently in dbSNP. The genes studied in this work could be considered as candidates for traits associated with beef production and the SNPs reported will help to assess the role of the genes in the genetic control of muscle development and meat quality. The allele frequency data presented allows the general utility of the SNPs to be assessed.

Keywords cattle, meat, polymorphism, SNP.

Introduction

Microsatellite markers selected from genome-wide linkage maps (Ihara *et al.* 2004; Snelling *et al.* 2005) have been used to localize quantitative trait loci (QTL) with effects on

several economically important traits in cattle (Boichard *et al.* 2003; Casas *et al.* 2003; Ashwell *et al.* 2004; Hu *et al.* 2007). However, genotyping microsatellite markers is labour-intensive and allele calls are laboratory-specific. In addition, these anonymous markers provide no information on the genes underlying QTL. Single nucleotide polymorphisms (SNPs) are more abundant than microsatellite loci, occurring at a frequency of about one SNP per kb in humans and about one SNP per 500 base pairs (bp) in mice (Lindblad-Toh *et al.* 2000) and cattle (Heaton *et al.* 2001). Despite being bi-allelic and so having a lower information

Address for correspondence

J. L. Williams, Parco Tecnologico Padano, Via Einstein, Polo Universitario, 26900 Lodi, Italy.
E-mail: john.williams@tecnoparco.org

Accepted for publication 28 December 2008

content than microsatellite markers, the availability of high-throughput SNP genotyping platforms makes it feasible to undertake high-density scans using large numbers of SNP markers either distributed across the whole genome or focussed in specific regions. SNPs within genes may also be the causative mutations responsible for variations in the phenotypes.

Mining the bovine genome sequence data has identified over 2.3 million putative SNPs (<http://www.ncbi.nlm.nih.gov/SNP/>); however, many of these are likely to be sequencing artefacts and only around 123 000 were identified from regions with deep sequence coverage, and so can be considered as being of 'high quality' and hence have a higher probability of being validated. This lack of validated cattle SNPs contrasts with the situation for humans, where 3.9 million SNPs have been validated (Jorgenson & Witte 2006). For genome-wide association studies, the number of SNP markers required depends on the structure of populations under investigation, in particular the extent of linkage disequilibrium (LD). The human HAPMAP project (<http://www.genome.gov/10001688/>) is developing a panel of 600 K SNP markers to tag LD-blocks. Breeding strategies adopted in cattle, particularly in recent times where selection has favoured the use of a small number of breeding sires, means that blocks of LD in the bovine genome extend over larger distances than in human (McKay *et al.* 2007) and so fewer markers will be necessary. Even so, for genetic association studies, a much larger number of confirmed SNP markers is required than is currently available. Recently, the lack of validated SNPs has been addressed either by mining additional existing bovine sequence data such as bacterial artificial chromosome (BAC) end sequences (Charlier *et al.* 2008) or by extensive shotgun resequencing of reduced representation genomic libraries (Van Tassell *et al.* 2008).

Up to now, genome-wide association studies have used panels of SNP markers selected to give a uniform distribution across the genome. However, to identify the causative mutations within genes or regulatory sequences, recent work has suggested that SNPs within genes provide a higher power for association analysis, even when the SNPs themselves do not produce functional variations, compared with using a SNP set with uniform genome-wide distribution (Jorgenson & Witte 2006). This conclusion is based on the observation that, in general, LD in regions with a high density of coding sequences is shorter than in gene-sparse regions of the genome. As a consequence, SNPs within genes are more likely to be in LD with the causative variations than SNPs in flanking non-coding regions. Therefore, additional SNP discovery may be most useful if targeted within coding sequences.

The objective of this work was to identify SNPs within genes that are predominantly involved in muscle development, metabolism and structure. Such genes potentially affect meat quality and could be considered as candidate genes for meat production related traits. SNPs within candidate genes may either result in functional variation or

may be very closely linked to such functional variants, and so be used as markers to test putative trait genes for an effect on phenotypic variation. Using a resequencing approach, 672 putative SNPs were identified in 206 genes which were examined in 15 breeds of European cattle.

Materials and methods

Animals

A panel of 436 largely unrelated bulls belonging to 15 European cattle breeds was used to identify, validate and measure allele frequencies of SNPs within the selected candidate genes. The panel of animals consisted of 30 Aberdeen Angus, 29 Highland, 31 Jersey and 27 South Devon, from the UK; 29 Danish Red Cattle, 29 Holstein and 20 Simmental, from Denmark; 30 Asturiana de los Valles, 30 Aleviana-Negra Ibérica, 31 Casina and 31 Pirenaica from Spain; 28 Marchigiana and 30 Piedmontese from Italy; and 30 Charolais and 31 Limousin from France. Blood samples were collected and DNA prepared from these animals and their parents, where available. In total, DNA was available from 975 individuals.

Selection of genes

Following extensive searches of the literature, a total of 504 candidate genes (Table S1) were identified based on knowledge of muscle physiology and their likely involvement in muscle growth, differentiation and development, in controlling muscular composition and metabolic pathways, and in influencing the maturation of muscle into meat. Suitable sequence information for designing resequencing primers was available in the public databases (GenBank, <http://www.ncbi.nlm.nih.gov/>) for 286 of these genes.

SNP identification

Gene sequence was used to design PCR primers to amplify 308 fragments from coding regions and from 5' regions likely to contain transcription regulatory elements in the 286 genes that had suitable sequence information. Amplicons were produced for each primer pair from a minimum of 15 animals that were selected to include all breeds in the panel. The amplicons were sequenced using ABI sequencers and Big Dye labelled nucleotides in conjunction with the original PCR amplification primers. Putative SNPs were identified by alignment and inspection of the sequences produced.

Genotyping

The putative SNPs identified from resequencing were initially genotyped on 96 animals selected to be representative of the 15 breeds. SNPs that were not confirmed, or had a

minor allele frequency below 0.1 in all breeds, were discarded. The remaining SNPs were genotyped in the 436 animals from the 15 breeds plus the available parents: in total 975 individuals. Genotyping was performed by Kbioscience using the proprietary Kaspar© methodology.

Radiation hybrid mapping of genes

PCR sequencing primers were tested for amplification of hamster DNA, and those which either did not amplify a hamster product, or gave a product that could be distinguished from the bovine product, were selected for mapping on the 3K Roslin bovine radiation hybrid (RH) panel (Williams *et al.* 2002). All hybrid cells were typed in duplicate and presence or absence of a PCR product assessed by two independent observers. The genes were then assigned to chromosomes by two-point analysis against markers included in the latest version of the RH map for this panel (Jann *et al.* 2006) using the RH MAPPER software (Slonim *et al.* 1997). Multi-point maps were constructed for Lod 6 linkage groups using CARTHAGENE software (de Givry *et al.* 2005). Linkage groups were placed on the chromosomes by building whole chromosome maps; however, in many cases the position or orientation of linkage groups was not well supported, and where possible were verified by reference to independent information, such as *in situ* mapping data. Positions of genes in relation to the nearest microsatellite marker on the bovine RH map (Jann *et al.* 2006) were determined by inspection.

Position on the bovine sequence and sequence-derived SNPs

Gene fragments and SNP position were localized in the BTAU_4.0 bovine genome sequence (ftp://ftp.ensembl.org/pub/current_fasta/bos_taurus/dna/) using BLASTN (<http://blast.wustl.edu/>; Altschul *et al.* 1990) with the amplicon sequences.

The positions obtained were compared with positions of SNPs present in dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>) to determine whether identical SNPs had been identified. All SNPs present in dbSNP (about 2.3 million SNP) have been mapped on the BTAU_4.0 bovine genome sequence by the Ensembl team, and their position was retrieved using the biomart tool (<http://www.ensembl.org/biomart/martview/>). Among these, the validated SNPs (about 14 500 SNPs) were identified in the RefSNP files (ftp://ftp.ncbi.nih.gov/snp/organisms/cow_9913/ASN1_flat/).

Results

A total of 504 genes were selected on the basis of their biological roles, which might influence muscle development, structure, metabolism or meat maturation. In addition, hormones, growth factors and transcription factors were

included in the initial list of target genes (see Table S1). These genes could be considered as candidates that may affect meat quality parameters. Available annotation and sequence information for many of these genes was incomplete with, for example, intron/exon structures of the genes being unknown. Where sufficient information was available (286 genes), primers were designed to amplify 2–3 fragments in different parts of the gene, e.g. 3' UTR, coding and promoter regions. Overall, 670 primer pairs were designed to amplify fragments from, on average, 2.5 regions per gene (primers for gene fragments containing SNPs are given in Table S2). Amplified fragment lengths ranged from 123 to 1747 bp, with an average of 510 bp. Sequencing of PCR fragments from 15 unrelated animals drawn at random from different breeds revealed that the majority of SNPs were present at high frequency in the study population. Sequencing the fragments from additional animals only identified further SNPs with very low minor allele frequencies. SNPs were discovered in 308 fragments corresponding to 206 different genes, giving a total of 672 SNPs. The SNP data have been submitted to dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>; see Table S2 for SNP and NCBI identifiers). For 80 of the 286 genes examined, no polymorphism was identified in sequences from at least 15 individuals, and these were not considered further. For fragments containing putative polymorphic sites, the number of SNPs identified ranged from 1 to 16 per fragment (2.2 per fragment on average). Interestingly, there was no correlation between the fragment length and the number of SNPs identified (Fig. 1). In this data set, for the 308 amplicons that contained a polymorphic site, SNPs occurred, on average, every 245 bp (ranging from 1 in 11 bp to 1 in 1620 bp). The number of SNPs detected per gene ranged between 1 and 27 (3.2 on average) with most having a low number of SNPs.

The putative SNPs were mostly bi-allelic (671), with only one having additional alleles, (ss77832407; four alleles). There were 24 deletions or insertions (four deletion/insertions of more than one base). The SNPs resulted in 180 transversions (27%) and 466 transitions (68%), the remaining polymorphisms being indels or complex alleles.

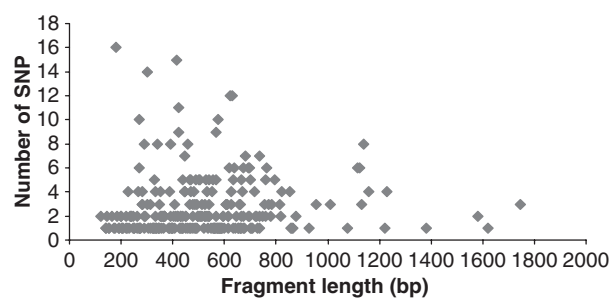


Figure 1 Relationship between the length of each fragment sequenced and the number of SNPs identified.

About 32% of the SNPs were located in coding regions of exons, approximately 12% in 3' or 5' UTR regions and approximately 56% in introns. Among the putative SNPs located in exons, 33% were non-synonymous and are predicted to result in amino acid substitutions in the corresponding proteins.

Genomic location of the SNPs

To facilitate the alignment of the candidate meat quality genes containing polymorphic sites with other information, e.g. QTL data, the genes were localized on bovine chromosomes by RH mapping and also by *in silico* mapping on the bovine genome sequence. It was not possible to map all genes by one approach. A total of 188 were successfully incorporated into RH maps and the flanking microsatellite markers included in the RH map of Jann *et al.* (2006) were identified (Table S3). All polymorphic genes were localized by BLAST on the Btau_4.0 bovine genome sequence. Positions of 662 of the 672 SNPs could be reliably determined by BLAST. Among these, 39 were localized on sequences not previously assigned to chromosomes and 15 were localized to mitochondrial DNA. The 10 SNPs for which a position could not be reliably determined were localized to seven different genes. There were one or more genes containing a SNP on each chromosome, with the exception of BTA23. The number of genes per chromosome containing a SNP ranged from 1, for chromosome 27 and X, to 18 for chromosome 19, not taking into account SNPs on sequences not assigned to chromosomes.

The 672 putative SNPs were tested in a panel of 96 animals randomly selected from the 15 test breeds. Markers for which an assay could not be developed or which had a minor allele frequency below 0.1 in this panel or which proved to be monomorphic were rejected (283 SNPs). The remaining 389 SNP markers were genotyped across 436 animals, to allow the calculation of allele frequencies. When available, parents (539) were also typed to confirm Mendelian inheritance of markers. In total, 975 animals were genotyped and allele frequencies across 15 breeds of cattle estimated (see Table S4).

Discussion

There has been a rapid increase in genomic information available for cattle, including new gene sequences from expressed sequence tags (ESTs) (Snelling *et al.* 2005), and recently the release of the bovine genome sequence (BTAU_4.0). The genome-sequencing project has produced about sevenfold genome sequence coverage: about sixfold coverage from shotgun sequencing of a Hereford cow and about onefold shotgun coverage from BAC clones representing a minimum tiling path across the genome of a Hereford bull. Mining this sequence information has identified over 2.3 million putative SNPs. However, most of

these SNPs have not been validated and many may be sequencing artefacts, or are present at such low frequency that they will not be useful in population studies. For humans, around 40% of SNPs discovered have a minor allele frequency of 0.025 or less (Schneider *et al.* 2003). Additional shotgun sequencing was carried out on individuals from 10 cattle breeds (Van Tassell & The International Bovine Hapmap Consortium 2006) to facilitate SNP discovery. A more thorough analysis of publicly available bovine SNP data has identified about 123K SNPs where deep sequence coverage gives higher confidence in the presence of the SNP. The international bovine HapMap project used these data to select a panel of about 30K SNP markers spanning the whole bovine genome. This panel of SNPs has been validated across 18 breeds of cattle selected from across the world, including both *Bos indicus* and *Bos taurus* types of cattle. SNP validation and allele frequencies have been determined for this HapMap SNP set (Van Tassell and The International Bovine Hapmap Consortium 2006). However, for most of the available SNP markers, allele frequencies across breeds are unknown, and in addition there are very few validated SNPs within coding regions. This conclusion is clearly illustrated by the fact that only 11 of the SNPs identified in this study are among the validated SNPs currently in the dbSNP database. Where SNPs are required within specific genes or chromosomal regions, the most effective strategy for their discovery remains direct resequencing and validation in target populations.

In this study, we report 672 putative SNP variations within genes selected for their potential functional importance in development, muscle structure, metabolism and meat production. The SNPs were tested in 15 European cattle breeds representing popular commercial breeds and local breeds. Overall, 389 SNPs with allele frequencies over 0.1 were confirmed. This conversion rate of putative SNPs to confirmed SNPs (56%), following testing in an extensive panel of animals, is better than that obtained in some other studies in cattle. Hawken *et al.* (2004) confirmed about 50% of putative SNPs from 55 genes, while Lee *et al.* (2006) found only 29% of 920 non-synonymous putative SNPs to have a minor allele frequency of >10%. Both these studies used *in silico* analyses of database sequences to identify the putative SNPs. Van Tassell *et al.* (2008) resequenced over 1.3 billion bp to discover 62 042 putative SNPs from which 23 257 were successfully genotyped, giving an apparent conversion rate of 37.6%, although the criteria for selecting the SNPs for genotyping is not discussed.

Surprisingly, in this study, a similar frequency of SNPs was observed in exons as in introns. In the data reported here, a density of one SNP for every 297 bp in exons was observed, with a similar density of one SNP for every 259 bp in introns. This finding was unexpected, as it is generally assumed that coding regions are more conserved than non-coding regions. SNP frequencies reported for human genes suggest a slightly lower frequency in exons, at about 1 per

800 bp, compared with introns at 1 in 500 bp (Zhao *et al.* 2003). However, these rates vary across the genome, e.g. estimates for HSA21 suggest SNPs in exons occur at 1 per 530 bp and at 1 per 350 bp in introns, while for HSA22 the densities are 1 per 520 bp in introns and a similar density of 1 per 530 bp for exons (Balasubramanian *et al.* 2002). Schnider *et al.* (2003) in a study of 1630 human genes reported a similar SNP density to that reported in this study for cattle, with a SNP every 234 bp in introns and an SNP for every 220 bp in exons. However, they found higher SNP densities in 3' UTR (1 per 163 bp) and 5' UTR regions (1 per 175 bp) than in introns. Unfortunately, insufficient data are available in this study for 3' and 5' regions for cattle to compare with these findings.

The proportion of transversions (27%) and transitions (69%) identified in this study was similar to that obtained in other studies in cattle, e.g. Hawken *et al.* (2004) observed 67% transitions, and also to that seen in the human genome, where transitions represent around 71% of variations (Schnider *et al.* 2003; Freudenberg-Hua *et al.* 2003). The higher frequency of transitions would be expected, as these cause fewer amino acid substitutions at the protein level than transversions. Among the SNPs identified in this study, 35% were C/T and a further 34% were A/G, while A/T was 6%, and A/C, C/G and T/G were at all 7% frequency. In humans, data from about 40K SNPs within genes identified that 35% were G/A or A/G and a further 35% were C/T or T/C (Schnider *et al.* 2003). In this study, for all SNPs involving an A or T with a C or G allele, the C or G allele was at higher frequency (about 70%). None of the polymorphisms found resulted in a premature stop codon. Among the polymorphisms found, 22 were indels (3.2%); of these, 18 were single base insertions or deletions and only four involved two or more bases. This is lower than the 20% frequency of indels found in the analysis of human resequencing data reported by Mills *et al.* (2006), although the data these researchers used included all sequences, not just genes, and among their data, 37% of the indels were within known genes.

Some genes show many more polymorphisms than others (range 1–27 SNPs per gene); however, the strategy adopted in this study will have biased the SNP discovery rate. Where there was limited sequence information available, this restricted the regions surveyed within particular genes. In addition, the strategy adopted would not have identified very low frequency SNPs. This was deliberate, as low-frequency SNPs are less informative in association studies (the most common use of SNPs in cattle). Nevertheless, analysis of the data indicated that there was no relationship between the number of SNPs identified in a given gene and either the number of fragments examined for that gene, the length of the fragments or the total length sequenced for that gene, suggesting that some genes are generally more polymorphic than others. SNPs were identified in 206 genes among the 286 examined, indicating that a relatively large number of the genes have some variation within coding regions.

From the 672 putative SNPs identified in this study, only 11 were found among the validated SNPs currently present in dbSNP. Searches of all 2.3 million putative bovine SNP in the dbSNP dataset revealed only 100 of the 672 SNPs characterized in this study. The bovine HapMap project validated about 30K SNPs and determined allele frequencies across a number of breeds (Van Tassell and The International Bovine Hapmap Consortium 2006). Other recent work used existing sequence to identify additional SNPs to design a 60K bovine SNP panel (Charlier *et al.* 2008), while resequencing from reduced representation libraries has identified a further 21K confirmed SNPs (Van Tassell *et al.* 2008). Nevertheless, the majority of confirmed SNPs available are in non-coding regions. In this report, we describe 389 fully validated SNPs, for which allele frequencies have been estimated in a wide range of breeds, many of which are important in commercial terms worldwide. This represents a significant increase in available SNP information. More importantly, these SNPs are within gene sequences, and so in some cases the polymorphism identified may prove to be functional, and hence could be directly involved in the control of variation in commercially important traits. Alternatively, if SNPs are simply to serve as markers for association studies, it is now suggested that SNPs within genes are preferable to randomly selected SNPs that are uniformly distributed across the genome, because the former are more likely to be linked to the functional variations than random markers in intergenic regions (Jorgenson & Witte, 2006). In selecting such markers, the allele frequency data on the SNP markers presented in this study for a wide range of breeds from different genetic backgrounds will be essential to evaluate the value of particular SNPs for genetic studies.

Acknowledgement

This work was supported by an EC grant QLK5 – CT2000-0147.

References

- Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.
- Ashwell M.S., Heyen D.W., Sonstegard T.S., Van Tassell C.P., Da Y., VanRaden P.M., Ron M., Weller J.I. & Lewin H.A. (2004) Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. *Journal of Dairy Science* **87**, 468–475.
- Balasubramanian S., Harrison P., Hegyi H., Bertone P., Luscombe N., Echols N., McGarvey P., Zhang Z. & Gerstein M. (2002) SNPs on human chromosomes 21 and 22 – analysis in terms of protein features and pseudogenes. *Pharmacogenomics* **3**, 393–402.
- Boichard D., Grohs C., Bourgeois F., Cerqueira F., Faugeras R., Neau A., Rupp R., Amigues Y., Boscher M.Y. & Levéziel H. (2003)

- Detection of genes influencing economic traits in three French dairy cattle breeds. *Genetics Selection Evolution* **35**, 77–101.
- Casas E., Shackelford S.D., Keele J.W., Koohmaraie M., Smith T.P. & Stone R.T. (2003) Detection of quantitative trait loci for growth and carcass composition in cattle. *Journal of Animal Science* **81**, 2976–83.
- Charlier C., Coppieters W., Rollin F., Desmecht D., Agerholm J.S., Cambisano N., Carta E., Dardano S., Dive M., Fasquelle C., Frennet J.C., Hanset R., Hubin X., Jorgensen C., Karim L., Kent M., Harvey K., Pearce B.R., Simon P., Tama N., Nie H., Vandeputte S., Lien S., Longeri M., Fredholm M., Harvey R.J. & Georges M. (2008) Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nature Genetics* **40**, 449–54.
- de Givry S., Bouchez M., Chabrier P., Milan D. & Schiex T. (2005) CARHTA GENE: multipopulation integrated genetic and radiation hybrid mapping. *Bioinformatics* **21**, 1703–1704.
- Freudenberg-Hua Y., Freudenberg J., Kluck N., Cichon S., Propping P. & Nöthen M.M. (2003) Single nucleotide variation analysis in 65 candidate genes for CNS disorders in a representative sample of the European population. *Genome Research* **13**, 2271–2276.
- Hawken R.J., Barris W.C., McWilliam S.M. & Dalrymple B.P. (2004) An interactive bovine *in silico* SNP database (IBISS). *Mammalian Genome* **15**, 819–827.
- Heaton M.P., Grosse W.M., Kappes S.M., Keele J.W., Chitko-McKown C.G., Cundiff L.V., Braun A., Little D.P. & Laegreid W.W. (2001) Estimation of DNA sequence diversity in bovine cytokine genes. *Mammalian Genome* **12**, 32–37.
- Hu Z., Fritz E.R. & Reecy J.M. (2007) Animal QTLdb: a livestock QTL database tool for positional QTL information mining and beyond. *Nucleic Acids Research* **35**, D604–D609.
- Ihara N., Takasuga A., Mizoshita K., Takeda H., Sugimoto M., Mizoguchi Y., Hirano T., Itoh T., Watanabe T., Reed K.M., Snelling W.M., Kappes S.M., Beattie C.W., Bennett G.L. & Sugimoto Y. (2004) A comprehensive genetic map of the cattle genome based on 3802 microsatellites. *Genome Research* **14**, 1987–1998.
- Jann O.C., Aerts J., Jones M., Hastings N., Law A., McKay S., Marques E., Prasad A., Yu J., Moore S.S., Floriot S., Mahé M.F., Eggen A., Silveri L., Negrini R., Milanese E., Ajmone-Marsan P., Valentini A., Marchitelli C., Savarese M., Janitz M., Herwig R., Hennig S., Gorni C., Connor E., Sonstergard T., Drögemüller C. & Williams J.L. (2006) A second generation radiation hybrid map to aid the assembly of the bovine genome sequence. *BMC Genomics* **7**, 283.
- Jorgenson E. & Witte J.S. (2006) A gene-centric approach to genome-wide association studies. *Nature Reviews Genetics* **7**, 885–91.
- Lee M.A., Keane O.M., Glass B.C., Manley T.R., Cullen N.G., Dodds K.G., McCulloch A.F., Morris C.A., Schreiber M., Warren J., Zadissa A., Wilson T. & McEwan J.C. (2006) Establishment of a pipeline to analyse non-synonymous SNPs in *Bos taurus*. *BMC Genomics* **7**, 298.
- Lindblad-Toh K., Winchester E., Daly M.J., Wang D.G., Hirshhorn J.N., Lavolette J.P., Ardlie K., Reich D.E., Robinson E., Sklar P., Shah N., Thomas D., Fan J.B., Gingeras T., Warrington J., Patil N., Hudson T.J. & Lander E.S. (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature Genetics* **24**, 381–386.
- McKay S.D., Schnabel R.D., Murdoch B.M., Matukumalli L.K., Aerts J., Coppieters W., Crews D., Neto E.D., Gill C.A., Gao C., Mannen H., Wang Z., Van Tassell C.P., Williams J.L., Taylor J.F. & Moore S.S. (2007) Whole genome linkage disequilibrium maps in cattle. *BMC Genetics* **8**, 74.
- Mills R.E., Luttig C.T., Larkins C.E., Beauchamp A., Tsui C., Pittard W.S. & Devine S.E. (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Research* **16**, 1182–1190.
- Schnider J.A., Pungliya M.S., Choi J.Y., Jaing R., Sun J., Salisbury B.A. & Stephens J.C. (2003) DNA variability in human genes. *Mechanism of Ageing and Development* **124**, 17–25.
- Slonim D., Kruglyak L., Stein L. & Lander E. (1997) Building human genome maps with radiation hybrids. *Journal of Computational Biology* **4**, 487–504.
- Snelling W., Casas E., Stone R.T., Keele J.W., Harhay G.P., Bennet G.L. & Smith T.P.L. (2005) Linkage mapping bovine EST-based SNP. *BMC Genomics* **6**, 74.
- Van Tassell C. & The International Bovine Hapmap Consortium (2006) An overview of the Bovine HapMap Project. In: *Proceedings of the 30th International Conference on Animal Genetics ISAG, 20–25 August 2006*, Vol. 60 (Ed. by the International Society for Animal Genetics), B533 p. Porto Seguro, BA, Brazil.
- Van Tassell C.P., Smith T.P., Matukumalli L.K., Taylor J.F., Schnabel R.D., Lawley C.T., Haudenschild C.D., Moore S.S., Warren W.C. & Sonstergard T.S. (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods* **5**, 247–252.
- Williams J.L., Eggen A., Ferretti L., Farr C., Gautier G., Amati G., Ball G., Caramori T., Critcher R., Costa S., Hextall P., Hills D., Jeulin A., Kiguwa S.L., Ross O., Smith A.L., Saunier K.L., Urquhart B.G.D. & Waddington D. (2002) A bovine whole genome radiation hybrid panel and outline map. *Mammalian Genome* **13**, 469–474.
- Zhao Z., Fu Y.-X., Hewett-Emmett D. & Boerwinkle E. (2003) Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene* **312**, 207–213.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1 Details of the complete list of genes considered for SNP analysis.

Table S2 Complete list of SNPs identified in the study.

Table S3 RH mapping data.

Table S4 Allele frequencies.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors.