A gene duplication affecting expression of the ovine *ASIP* gene is responsible for white and black sheep

Belinda J Norris and Vicki A Whan

Genome Res. published online May 20, 2008
Access the most recent version at doi:10.1101/gr.072090.107

A gene duplication affecting expression of the ovine *ASIP* gene is responsible for white and black sheep

Belinda J Norris* and Vicki A Whan

CSIRO Livestock Industries, Queensland Bioscience Precinct, 306 Carmody Rd., St Lucia 4067, QLD, Australia.

Running Title: copy number variation at ovine *ASIP* locus affects coat color

Keywords: agouti, itch, gene duplication, Merino, dominant white, pigmentation

*Corresponding author:

Dr Belinda J Norris
Senior Research Scientist
CSIRO Livestock Industries
Queensland Bioscience Precinct
306 Carmody Road, St. Lucia, QLD 4067, Australia
Phone: +61 7 3214 2282
Fax: +61 7 3214 2900
e-mail: belinda.norris@csiro.au

ABSTRACT

Agouti signalling protein (ASIP) functions to regulate pigmentation in mice while its role in many other animals and in humans has not been fully determined. In this study, we identify a 190 kb tandem duplication encompassing the ovine *ASIP* and *AHCY* coding regions and the *ITCH* promoter region as the genetic cause of white coat colour of dominant white/tan (A^{Wt}) *agouti* sheep. The duplication 5' breakpoint is located upstream of the *ASIP* coding sequence. Ubiquitous expression of a second copy of the *ASIP* coding sequence regulated by a duplicated copy of the nearby *ITCH* promoter causes the white sheep phenotype. A single copy *ASIP* gene with a silenced *ASIP* promoter occurs in recessive black sheep. In contrast, a single copy functional wild-type (A^{+}) *ASIP* is responsible for the ancient Barbary sheep coat colour phenotype. The gene duplication was facilitated by homologous recombination between two non-LTR SINE sequences flanking the duplicated segment. This is the first sheep trait attributable to gene duplication and shows non-allelic homologous recombination and gene conversion events at the ovine *ASIP* locus could have an important role in the evolution of sheep pigmentation.

Supplemental material is available online at www.genome.org. The sequence data from this study have been submitted to GenBank under accession nos. EU185093-EU185100 and EU420022-EU420031.

The genomic organisation of the agouti signalling protein gene (*ASIP*) is generally highly conserved in mammalian species including the mouse (Bultman et al. 1992), human (Kwon et al. 1994), horse (Rieder et al. 2001), pig (Leeb et al. 2000), cow (Girardot et al. 2005), and dog (Kerns et al. 2004). In the mouse genome, various insertions and deletions affecting *ASIP* and nearby loci have resulted in deregulated *ASIP* expression causing yellow pigmentation, adult onset obesity, diabetes, tumor growth and embryonic lethality (Wolff et al. 1999). Other mice mutations in both the coding and regulatory regions affect *ASIP* expression and function and subsequent coat pigmentation patterns (Bennett et al. 2003; Eppig et al. 2005). *ASIP* alleles causing coat colour variation have also been characterised in domestic dogs (Kerns et al. 2004), cats (Eizirik et al. 2003), pigs

(Drogemuller et al. 2006), horses (Rieder et al. 2001), rats (Kuramoto et al. 2001), and foxes (Vage et al. 1997). Although *ASIP* transcripts are present in various human tissues including liver, kidney, heart and adipose tissue, the role of *ASIP* in humans is not clear but may include both pigmentation (Bonilla et al. 2005; Kanetsky et al. 2002; Voisey et al. 2006) and energy homeostasis (Voisey et al. 2002).

In mammalian species, coat colour is an important form of camouflage and can be an integral part of social communication and recognition (Sponenberg 1997). The standard wild-type sheep coat colour is generally dark-bodied with a pale belly, similar to other mammalian wild-type coat colour patterns (Sponenberg 1997). However, this wild-type coat colour pattern is much rarer in domestic sheep where coat colour is an important breed characteristic and production trait. In domestic breeds, unlike in their wild ancestors, the lack of natural selection allows coat colour variants to arise and segregate. As a result of artificial selection for white fibres the white coat phenotype has reached a high frequency in certain breeds and shows autosomal dominant inheritance. In mice the dark dorsal and pale ventral wild-type coat colour pattern has been shown to be caused by the spatial expression of different transcripts from a single *ASIP* gene (Vrieling et al. 1994), presumably controlled by different regulatory elements.

The contribution of *ASIP* to coat colour patterns of domestic sheep has been investigated via classical genetics since the early 1920's. The dominant white or tan (A^{Wt}) *ASIP* allele is responsible for the phaeomelanic (yellow/red) phenotype in modern sheep breeds while the most recessive allele, non-agouti (A^{a}) results in eumelanic (black/brown) phenotypes (Brooker et al. 1969a; Adalsteinsson 1970). Another *ASIP* allele, badgerface (A^{b}) is characterised by a pale dorsal phaeomelanic and darker ventral eumelanic pattern; it is recessive to A^{Wt} and dominant to A^{a} (Brooker et al. 1969a). Independent loci controlling colour directly and structural features of the hair/wool coat also modify the phaeomelanic colours such that wool areas are typically white (Sponenberg 1997). Despite strong selection for white wool colour, the frequency of the recessive non-agouti (also known as self-colour black) A^{a} allele in the Australian Merino flock was estimated at 0.03 (Hayman et al. 1965). While strong comparative and classical genetic evidence for *ASIP*

involvement in sheep dominant white and recessively coloured coat phenotypes exist, the molecular genetic cause has remained unresolved (Parsons et al. 1997; Parsons et al. 1999a; Parsons et al. 1999b).

To identify the molecular genetic cause of coat color variation in domestic sheep we have used sequence analysis of genomic DNA, BAC clones, and RT-PCR products from Romanov, Texel, Merino, and ancient Barbary sheep to characterize the genomic structure of the *ASIP* locus. We report a tandem duplication of a 190 kb portion of the ovine genome is responsible for the dominant white coat colour ($A^{w/t}$) allele of domestic sheep. An asymmetric competitive PCR assay was developed and used to show that the multiple copy *ASIP* alleles consistently segregate with the dominant white coat colour in Merino sheep. RT-PCR experiments were used to demonstrate alternative splicing of *ASIP* transcripts and show expression in multiple tissues of white sheep and lack of expression of the single copy alleles in black sheep. 5'RACE experiments determined that unlike the silent single copy *ASIP* of recessive black Merino sheep, a single copy Barbary sheep *ASIP* with a functional promoter is responsible for the wild-type pale belly phenotype.

RESULTS

The ovine *ASIP* gene intron and exon structure

By sequencing the 5379 bp ovine *ASIP* gene from Merino sheep, (GenBank accession no. EU420022) we determined the genomic organisation of the ovine *ASIP* coding exons 2, 3, and 4 and introns 2 and 3 to be similar to that reported for the bovine, human and mouse genes (Fig. 1). Each coding exon is flanked by a consensus splice donor and acceptor site. BLASTN of the *ASIP* sequence against GenBank databases showed greatest identity to the bovine *ASIP* sequence (95.2%). The ovine gene encodes a putative 133 amino acid protein which is 98%, 76% and 74% identical to the bovine (133aa), mouse (131aa) and human (132aa) proteins respectively.

Mutation screening

To identify possible recessive mutations causing pigmentation, we sequenced the entire *ASIP* coding and non-coding sequence of Romanov, Texel, and Merino sheep and identified four coding sequence mutations compared to the published ovine *ASIP* sequence (Parsons et al. 1999b) (Fig. 1). A previously reported 5 bp deletion, g.100-105delAGGAA (denoted D_5) was present in exon 2 (Smit et al. 2002). A novel 9 bp deletion, g.10-19delAGCCGCCTC (denoted D_9) was also found to be present in exon 2 10bp from the ATG start codon, and two SNPs were found in exon 4 (g.5051G>C and g.5172T>A). Both the D_5 deletion and the g.5172T>A SNP in exon 4 would be predicted to independently cause functional changes to the agouti protein. The D_5 deletion would result in a frame shift followed by a premature stop codon 63 amino acids downstream of the start site, truncating the agouti protein before the functionally important cysteine signalling domain (amino acids 91-130). The g.5172T>A SNP would predict a change of cysteine (amino acid 123) to serine which would disrupt the highly conserved signalling region of the protein. The D_9 deletion would result in the loss of a tripeptide (SRL), which may affect the function of the ASIP transport leader sequence but not the remainder of the protein, while the synonymous g.5051G>C SNP in exon 4 would not be expected to disrupt ASIP function.

Association analysis of the D_9 and D_5 deletions and A allele with recessive black Merino phenotypes

A coat colour panel comprising 373 DNA samples from white (n = 183), self-colour black (n = 142) and badgerface (n = 48) Australian Merino sheep was genotyped for the D_5 and D_9 indels and the g.5172T>A SNP. The self-colour black Merinos displayed a typical agouti symmetrical pigmentation pattern of a dark body with a white blaze on the head and neck (Fig 2A). A characteristic badgerface Merino with dark ventral and lighter dorsal symmetrical agouti pigmentation pattern is shown in Figure 2B. None of the white Merinos were homozygous for the D_9, D_5 or A alleles. Furthermore, 78 self-colour black Merinos that were homozygous D_5, were also homozygous T at the g.5172T>A SNP and only 11% of the recessive black sheep were homozygous A. Approximately 16% of the white sheep and <2% of the black sheep had the D_9 allele and none of the animals had an allele with both of these deletions. During genotyping we also deduced that 8% of the

Merino sheep had all alleles — non-deleted alleles at the D_9 and D_5 positions (denoted N_9 and N_5) and the D_9 and D_5 alleles, confirming the previous report of another *agouti*-like locus (Smit et al. 2002). As none of the white Merinos were homozygous for the D_9, D_5 or A alleles, the presence of the functional N_9, N_5, and T haplotype thus seems necessary for a Merino sheep to be white. However, not all black animals (self-colour and badgerface) in our panel could be explained by homozygosity of a recessive non-functional D_5 or A allele. PCR products spanning *ASIP* exons 2, 3, and 4 in four self-colour black Merino sheep homozygous for the N_9, N_5, and T alleles were amplified and sequenced directly. No other coding mutations that might predict a non-functional mature agouti protein were present.

ASIP expression in Merino sheep

Allele specific expression of functional *ASIP* transcripts in white sheep skin

The expression of alleles with the protein altering D_5 mutation and the D_9 mutation were investigated in nine white Merino sheep. PCR and polyacrylamide electrophoresis was used to examine the expression of the variant alleles (Fig. 3). Seven white sheep with the N_9N_5 and N_9D_5 alleles and one sheep (Lane 5) with all three alleles (N_9N_5, N_9D_5, and D_9N_5) showed that the non-deleted N_9N_5 and deleted D_9N_5 alleles were preferentially expressed in Merino skin as visualised by the band intensities. The N_9D_5 allele was not amplified by competitive RT-PCR (arrow Fig. 3 Lanes 1-9) and was considered comparatively lowly expressed or not expressed at all.

ASIP expression in white sheep is controlled by an *ITCH* promoter

Using gene-specific primers designed from *ASIP* ESTs (data not shown), RT-PCR analysis of skin and internal tissues from white and recessive black Merinos indicated that the dominant white phenotype of Merino sheep was caused by high levels of deregulated expression of the *agouti* gene from an itchy homolog E3 ubiquitin protein ligase (*ITCH*) gene promoter. The same alternatively spliced forms of ovine *ASIP* were expressed in liver, kidney, heart, spleen, and skin tissues tested from white Merinos but were not amplified from any tissues of a self-colour black Merino (Fig. 4 A-F). *ASIP* ubiquitous expression has also been reported in humans (Voisey et al. 2002) and cattle

(Girardot et al. 2005) and represents a departure from what is seen in mice where expression is generally confined to the skin (Bultman et al. 1992). RT-PCR (Fig. 4 A,B) and 3' RACE (Fig. 4C) experiments also indicated all the *ASIP* transcripts started with two non-coding exons, designated It and It' and the same transcripts were present in all the white Merino sheep tissues examined. The expression of ovine *ITCH*, however, was not affected with expression of the same order of magnitude in both the black and white sheep tissues (Fig. 4 D,F). The 'normal' expression of *ITCH* explains why sheep do not experience the spectrum of immunological conditions evident in the 'itchy' mouse in which both *ITCH* and *ASIP* expression is disrupted due to a genomic inversion affecting both loci (Perry et al. 1998). *ASIP* and *ITCH* expression levels in the skin and other tissues was confirmed by real time quantitative PCR analysis of three biological replicates each of Merino white and black sheep tissue (data not shown).

Alternatively spliced *ASIP* transcripts

RT-PCR identified several *ASIP* transcripts from the skin of white Merino sheep. Primer Agt9 (positioned in the first non-coding exon It) and primer Agt6 (in the 3' UTR of *ASIP*) (Supplemental Table S2 online) were used to amplify cDNA derived from the skin of white Merino sheep. These products were cloned and sequenced and seven alternative transcripts (GenBank accession nos. EU420024-EU420030) obtained (Fig. 5A). Significant differences were evident for the sequences of the ovine skin non-coding exons compared to those reported for cow, mouse and pig (Girardot et al. 2005; Vrieling et al. 1994; Leeb et al. 2000). Every ovine skin transcript had the two non-coding exons which we have labelled It and It' in addition to one or two other non-coding exons designated IA to IE (Fig. 5A). A terminal G nucleotide on the sequence of *ASIP* transcripts derived by 5' RACE determined exon It to be the most 5' non-coding exon. The non-coding exons It and It' were also part of the 5' region of the downstream *ITCH* gene (Fig. 5A) and were 67% identical to the non-coding exons in *ASIP* transcripts from the sienna yellow (Asy) mouse (Duhl et al. 1994).

The expression and RT-PCR results indicated that the dominant white phenotype of Merino sheep is caused by high levels of expression of functional *ASIP* transcripts driven

by an *ITCH* promoter region. Comparative analysis of the human, dog and mouse genomes (http://genome.ucsc.edu/) places the *ITCH* gene downstream of *ASIP* in these species. The *ITCH-ASIP* hybrid transcripts therefore suggested the presence of a tandem duplication or rearrangement and deletion in this region of the ovine genome.

Identification of a duplication encompassing the *ASIP*, *AHCY* and *ITCH* regions

As the ovine genome sequence was not yet determined, to characterise the ovine *ASIP-ITCH* genomic region we obtained and sequenced three Romanov sheep and five Texel sheep BAC clones. Sequencing of these ovine BAC clones identified a large (~190 kb) tandem genomic duplication (Fig. 5A; Supplemental Figure S1 online). Alignment of *ASIP* non-coding exons It (77 bp), It' (78 bp) and IA (54 bp) sequences identified from alternatively spliced transcripts from the skin of white Merino sheep to BAC clone INRA-164H8 and INRA-229C6 sequences from a Romanov sheep positioned them approximately 119, 123, and 131 kb, respectively, 3' of the *ASIP* start codon (Fig. 5A). The ID (56 bp) and IE (88 bp) non-coding exons were located approximately 6.6 kb and 298 bp, respectively, 5' of the *ASIP* start codon (Fig. 5A). The IB (178 bp) and IC (97 bp) non-coding exons were not present in INRA-164H8 sequence, which ended ~10 kb 5' of the *ASIP* start codon. Comparative alignment of the non-coding exon IB and IC sequences to the dog, human and cow genome sequences placed these exons further 5' of the *ASIP* start codon in these species (data not shown).

The Merino *ASIP* intron and exon sequence data and the INRA BAC clone sequence data were used to search the CHORI Texel sheep library BAC-end sequences to identify BAC clones in the region of the ovine *ASIP* gene. In addition, the mapping of the full set of sheep BACs to the cow, dog and human genomes (Dalrymple et al. 2007) was used to identify BACs potentially located across the putative duplication and flanking genomic regions. Twenty-nine clones that spanned approximately 2 Mb of sequence encompassing the *ASIP*, S-adenosylhomocysteine hydrolase (*AHCY*), *ITCH*, and flanking genomic regions of ovine chromosome 13 were obtained for analysis. The BAC clones were then ordered across the region and five BACs predicted to span the complete region including the junction point, and 5' and 3' breakpoints were sequenced. We obtained approximately

500 kb of contiguous sequence that contained a 190 kb tandem duplication including the complete *ASIP* and *AHCY* coding regions and the *ITCH* promoter and non-coding exon sequences It, It' and IA (Fig. 5A). Sequence data from clones CH243-455O4, CH243-234K21 and CH243-489F15 spanned the junction between the tandem duplicated copies. BAC clone CH243-160L8 spanned the 3' breakpoint and clone CH243-373J16 spanned the 5' breakpoint.

Romanov, Texel and Merino sheep Haplotypes

Sequencing of the Romanov and Texel Sheep BAC clones and Merino sheep PCR products from genomic DNA identified seven ovine *ASIP* haplotypes (Supplemental Table S1). Haplotypes 1 and 2 were determined by cloning and sequencing PCR products from the genomic DNA of a white (A^{Wt}) and a self-colour black (A^{a}) Merino sheep. The white animal contained haplotypes 1 and 2 and the self-colour black animal contained only haplotype 2. Ovine *ASIP* haplotypes 3 and 4 were determined by sequencing of INRA BAC clones INRA-218G7, INRA-229C6 and INRA-164H8 from Romanov sheep. Ovine *ASIP* haplotypes 5, 6 and 7 were identified from the Texel sheep BAC clone sequences. Haplotype 5 was designated *ASIP* copy 1 as it was positioned 5' to haplotypes 6 and 7, both designated *ASIP* copy 2 (Fig. 5A). Haplotypes 6 and 7 (*ASIP* copy 2) were 99.9% identical to the functional haplotypes 1 and 3 from Merino and Romanov sheep while haplotype 5 (*ASIP* copy 1) showed greatest identity (99.7%) to the non-functional A allele containing haplotype 4 from Romanov sheep (see also Fig. 6). The position of the sheep functional haplotypes 6 and 7 at *ASIP* copy 2 places them 3' to and therefore under the regulation of the duplicated *ITCH* promoter (Fig. 5A).

Haplotypes 1, 6 and 7 all contained the proposed functional $N_{9,}$ $N_{5,}$ and T alleles. Haplotype 2 contained the predicted non-functional D_5, and functional T alleles and was otherwise very similar to Haplotype 1. Haplotype 3 contained the $D_{9,}$ $N_{5,}$ and T alleles and was also otherwise very similar to the functional haplotypes 1, 6 and 7. Haplotypes 4 and 5 both contained the predicted non-functional A allele and were quite different to all other haplotypes. Using haplotype 1 as the reference sequence, haplotypes 4 and 5 had 35 and 40 nucleotide substitutions respectively. Transcripts from haplotypes 1, 3, 6 and 7

would be predicted to produce functional agouti protein while transcripts from haplotypes 2, 4 and 5 would be predicted to produce non-functional agouti protein.

Non-allelic pairing, crossover and gene conversion events arguably occur more frequently between almost identical tandem duplications (Myers et al. 2006; Lindsay et al. 2006) and thus, the tandem duplications of the *ASIP* gene are likely candidates for these events. Indeed, the sequences of the *ASIP* copy 1 and 2 haplotypes suggest mutations, recombination and gene conversion events possibly have occurred between them. It is conceivable that Haplotype 5 (*ASIP* copy 1) could be derived from recombination and/or gene conversion between Texel sheep equivalents of haplotypes 1 and 4. In addition Texel sheep haplotype 6 (*ASIP* copy 2) that was otherwise identical to haplotype 7 (*ASIP* copy 2) shared two nucleotides (at positions 4124 bp and 5051 bp) with haplotype 5 (*ASIP* copy 1), which are likely a result of mutation or gene conversion events. Haplotype 5 also had an A nucleotide at position 3096 bp not present in any other haplotypes that is likely a result of a mutation.

Analysis of the duplication junction & breakpoint sequences

We analysed the duplication junction and breakpoint sequences for regions that could facilitate duplication events. By comparing the DNA sequence of each of the BAC clones containing the junction, 5' and 3' breakpoints, a region of sequence similarity of approximately 143 bp was identified at all three sites (Fig. 5B). These sequences were identified by Repbase (Kohany et al. 2006) as having 82-88% identity to Bovidae non-LTR BOV2 and BDDF2 repetitive SINE sequence regions. The duplication may have been facilitated by homologous recombination between these repetitive elements.

Analysis of *ASIP* copy number variation in domestic Merino sheep.

Sequence flanking the shared 143 bp sequence at the junction and 5' breakpoint was used to develop an asymmetric competitive PCR copy number assay (Supplemental Figure S2). The junction point and 5' breakpoint PCR products (see Materials and Methods) were used initially to assess the presence and absence of duplicated copy alleles in the genomes of white and recessive black Merino sheep. All white Merinos successfully

assayed (n = 177) amplified both a junction point and 5' breakpoint PCR product and therefore had duplicated *ASIP* alleles (Table 1). By contrast all of the successfully assayed recessive black Merinos (n = 180) amplified a 5' breakpoint product only and thus had only single copy silent (refer expression data) *ASIP* alleles.

To calibrate the ABI3130xl copy number assay, PCR products of the junction point and 5' breakpoint were cloned into pCR®2.1-TOPO® and a standard curve was established using a dilution series of plasmid DNA containing the junction point mixed with plasmid DNA containing the 5' breakpoint from 0% to 100%. The resulting standard curve was linear (Supplemental Figure S3). The genomic DNA of 16 white Merino sheep (8 that were confirmed as carriers of a single copy recessive black-allele by pedigree analysis), was next analysed (in quadruplicate) in an initial evaluation of the ABI3130xl copy number assay. Three of the confirmed carriers assayed contained a single junction point allele as would be predicted for a carrier animal (Supplemental Figure S4, Lanes 2, 3 and 8). However, the 5 other carriers contained 2 junction points, which indicated triplicated alleles and suggested even greater diversity in copy number could occur at the ovine *ASIP* locus. The ABI3130xl copy number test was used to assess *ASIP* copy number variation in the white sheep of the Merino sheep coat colour panel. The copy number genotypes of the assayed carrier and random white sheep indicated that white Merino sheep contain multiple copy *ASIP* alleles (3 to 6 copies) per genome (Table 1).

Comparative analysis of *ASIP* copy number variation and expression in Barbary sheep (*Ammotragus lervia*), an ancient Caprinae species.

Barbary sheep (Fig.2) have a wild-type agouti coat colour pattern - tan body and pale belly - and are assumed to be homozygous for the wild-type *ASIP* allele (A^+). PCR analysis (n = 40) amplified only the 5' breakpoint product indicating single copy *ASIP* alleles (Table 1). The 5'RACE experiment determined that unlike the silent single copy *ASIP* of recessive black Merino sheep, the single copy Barbary sheep *ASIP* has a functional promoter and is responsible for the pale belly phenotype. The partial *ASIP* transcript (GenBank accession no. EU420031) from Barbary sheep ventral skin had a single non-coding 1A-like exon and coding exons 2, 3, and partial exon 4 sequence (Fig

5A) with high similarity to the homologous ventral-specific mouse (Bultman et al. 1992; Vrieling et al. 1994) and pig 1A (Drogemuller et al. 2006) transcripts. Analysis of the Texel sheep BAC clone genomic sequence positioned the Barbary sheep non-coding 1A-like exon sequence approximately 69 kb 5' of *ASIP* copy 1 exon 2 in domestic sheep; 27 kb further upstream of the 5' breakpoint (Fig. 5A). Our data supports a functional *ASIP* promoter driving expression of a single *ASIP* gene determining coat colour patterns in this ancient species of the Caprinae (goat-antelope) subfamily.

DISCUSSION

It has taken more than a century of mouse classical genetics and two-decades of molecular genetic analysis to begin to understand the molecular mechanisms regulating yellow coat colour and the associated pleiotropic effects from deregulated murine *ASIP* expression (Wolff 2003). The molecular regulation of coat colour of other mammalian species at the *ASIP* locus however still remains comparatively unknown. The expression of the mouse *ASIP* gene ordinarily is tightly spatially and temporally regulated and restricted primarily to the skin (Bultman et al. 1992). In humans, (Voisey et al. 2002) cattle (Girardot et al. 2005; Girardot et al. 2006) and, as we have now shown in domestic sheep, *ASIP* is more widely expressed with various transcripts identified from several tissues. While there are no apparent negative effects of this ubiquitous *ASIP* expression in sheep, the possible consequences of the wider tissue expression is still to be investigated. Interestingly, in Icelandic sheep, pleiotropic effects of the dominant white or tan (A^{Wt}) allele result in reduced fecundity and greater seasonality in reproduction (Adalsteinsson 1975). However, in mice ubiquitous *ASIP* expression results in adult onset obesity, suggesting that the signalling pathways that normally regulate reproduction and body weight may be different in these mammalian species.

It was also recently shown that the *ASIP* coding sequence has been completely deleted from the gibbon genome (Nakayama and Ishida 2006) with no apparent detrimental effect, while it remains intact in all other investigated primate genomes (Mundy et al. 2006). Nakayama and Ishida (2006) speculate that the gibbon *ASIP* deletion may affect its energy homeostasis and so contribute to the smaller body mass of the gibbon, relative

to other primates. Studies in humans and animal models have elucidated a role for components of the melanocortin pathway in immunity, energy homeostasis and reproduction (Carroll et al. 2005; Henry 2003). *ASIP* regulation of lipid metabolism in adipose tissue has been examined in mouse and humans (Voisey et al. 2002) but to date, no studies have investigated the potential impact of increased or deregulated *ASIP* expression on ovine energy balance or fertility. However, lean sheep fed *ad libitum* can become obese and have proven to be a useful animal model for human obesity (McCann et al. 1991; Henry 2003). Investigating the impact of *ASIP* copy number variation on key production traits of leanness, body fat and fertility could provide new insights into the physiology of energy metabolism and reproduction of livestock and other mammalian species.

Mammalian species thus far investigated have had a single copy of the *ASIP* gene identified within their genomes. The genetic cause of coat colour patterns in domestic sheep at the *ASIP* locus has been difficult to determine due in part to the prospect of a second *ASIP*-like locus, which was proposed by Smit *et al.* (Smit et al. 2002). Further, with the exception of the mouse (Bultman et al. 1992; Bultman et al. 1994) and the pig (Drogemuller et al. 2005), mutations in the coding sequence which disrupt functional ASIP protein were associated with the recessive black phenotypes of these species (Eizirik et al. 2003; Rieder et al. 2001; Kuramoto et al. 2001; Vage et al. 1997; Kerns et al. 2004). We have shown that the regulation of the recessive black coat colour phenotypes by the ovine *ASIP* gene is not attributable to simple coding mutations as described for other mammals and that the dominant white phenotype involves variation in gene copy number and deregulated expression.

Analysis of the ovine *ASIP* coding regions of recessive self-colour black and badgerface Merino sheep identified four mutations, two of which (a 5bp deletion and a g.5172T >A SNP) would be predicted to disrupt the functional protein. Both these coding sequence mutations however failed to completely associate with the recessive black Merino phenotypes. Of the recessive black Merinos, only ~60% and 11% were homozygous for either the D5 or A allele respectively, and none were homozygous for both. All white

Merinos investigated were either homozygous for the normal alleles or heterozygous for one or the other of these mutations. Additionally all white Merinos had at least one duplicated *ASIP* allele while all of the recessive black Merinos contained only single-copy alleles.

ASIP expression was detected in all tissues examined from white sheep with multiple transcripts identified from both the skin and internal tissues. Expression was not detected from any tissues of recessive self-colour black sheep. Further, all transcripts identified from white sheep skin began with the *ITCH* non-coding exons It and It' and none contained the 5bp or g.5172T>A predicted non-functional mutations. This data indicated that expression in the white sheep was driven from the duplicated copy of an *ITCH* promoter positioned upstream of *ASIP* and that the identified functional mutations are not present in these expressed *ASIP* copies. The data also indicated that in domestic sheep the progenitor *ASIP* promoter, unlike the promoter of the single copy *ASIP* of ancient Barbary sheep, was not functional. Because *ASIP* expression was not detected in the single copy alleles of the recessive black sheep, the likely causative mutation of the recessive black phenotype is an as yet unidentified regulatory mutation of the progenitor gene promoter region.

In mice the molecular events associated with an unusually high reversion rate of *agouti* recessive black coat colour pattern mutations to more dominant alleles were identified to be as a result of insertions and deletions mediated by homologous recombination between repetitive elements (Bultman et al. 1994). In the gibbon the removal of a 100 kb region including the *ASIP* coding sequence was mediated by Alu repeat elements (Nakayama and Ishida 2006). A large duplication involving the *KIT* locus causing dominant white skin colour in pigs has originated by homologous recombination between LINE elements (Marklund et al. 1998; Giuffra et al. 2002). The sequences at the borders of the duplicated ovine *ASIP* gene segments contain repeat elements identified as BOV2 and BDDF2 non-LTR SINE sequences (Kohany et al. 2006). Thus, it seems likely that the ovine dominant white and recessive black phenotypes are mediated at least in part by homologous recombination involving these repeat elements. The sequencing and subsequent

characterization of the human genome revealed that 5% of the human genome consists of long highly similar duplicated sequences known as low copy repeats (LCRs) or segmental duplications (Lindsay et al. 2006). It is now emerging that such segmental duplications are hotspots for allelic and nonallelic homologous recombination events that can result in genomic disorders causing clinical diseases (Bischof et al. 2006).

The exchange of sequence variations between duplicated loci known as gene conversion events mediated through allelic and nonallelic recombination (Myers et al. 2006) can lead to a high degree of similarity between duplicated genes (Lindsay et al. 2006). A high degree of nucleotide sequence similarity (>99%) occurred between the 5379 bp of the multiple *ASIP* gene copies of each of the BAC clone sequences from the start of exon 2 to the end of the 3' untranslated region (Supplemental Table S1). Gene conversion events together with allelic and nonallelic homologous recombination between repetitive sequences are likely to be contributing to variation and evolution of ovine *ASIP* sequences.

Based on our results we propose a model for the evolution of the ovine *ASIP* locus (Fig. 6). The model proposes that a duplication, mediated by homologous sequences in SINES flanking the *ASIP* coding and *ITCH* promoter regions, occurred in an ancestor of domestic sheep positioning a second copy of the *ITCH* promoter upstream of a second *ASIP* coding sequence, deregulating its expression. Subsequently, possibly with the creation of the duplicated segment, or by other successive mutation events, the progenitor *ASIP* promoter was inactivated. Sheep with genotypes that have at least one duplicated *ASIP* allele are always white as strong (artificial) selection pressure for white coat colour has maintained the functional ($N_{5,}$ T) *ASIP* alleles at the expressed (copy 2) position (Fig 6A). With the expressed (copy 2) *ASIP* functional coding sequence under strong selection pressure, the two *ASIP* regions subsequently diverged in sequence generating functional and non-functional haplotype clusters (Supplemental Table S1 and Fig. 6B). The most likely major route of generation of single copy loci was nonallelic pairing and recombination between duplicated alleles (Fig 6C). The single copy recessive black causing alleles, with functional or non-functional coding sequence haplotypes,

subsequently resulted from nonalleic recombination and/or gene conversion events (Fig. 6C). Our data from asymmetric competitive PCR copy number assays of recessive black and white Merinos shows that while recessive black Merinos always have a single *ASIP* copy, white Merinos can have 2 to 4 and possibly even 5 *ASIP* copy alleles, further supporting this model.

Tandem gene duplication represents an under-investigated source of molecular variation in livestock species. The duplication/deletion of the agouti gene is the first characterised example in sheep of the involvement of gene duplication in the creation of a genetic variation that contributes to a major breed and production trait- coat colour phenotypes. This gene copy-number polymorphism introduces variation into the sheep genome and causes a phenotypic trait to which natural or artificial selection may then apply. The ovine chromosome 13 tandem duplication, encompassing the *ASIP*, *AHCY* genes and promoter region of the *ITCH* gene is an ideal locus for the investigation of the consequences of duplication on sequence variation and evolution via natural and artificial selection in commercially important livestock species.

Classical genetic analysis first proposed alleles of the *agouti* gene as a major determinant of coat colour phenotypes in sheep more than a half a century ago (Rendel 1957; Brooker and Dolling 1969). With a major goal of identifying the molecular cause of unwanted recessive black sheep phenotypes, the molecular genetic investigation of the ovine *agouti* gene in the Australian Merino sheep was begun a decade ago (Parsons et al. 1997; Parsons et al. 1999a; Parsons et al. 1999b). Until now, without the benefit of an ovine genome sequence, progress in understanding the molecular mechanisms regulating coat colour in domestic and wild sheep populations at the *ASIP* locus has been elusive. The characterization of a large tandem duplication encompassing the *ASIP*, *AHCY* and *ITCH* genes provides an explanation for this difficulty. In the future, further comparative investigation of the *agouti* architecture in domestic sheep breeds and extant wild sheep populations will, in evolutionary terms, determine when this dominant *agouti* mutation first occurred and its impact on and evolution since the domestication of sheep.

METHODS

Panel of sheep coat colour phenotypes. Samples of 6ml blood (EDTA tubes) and 0.9 mm skin biopsies were collected from Australian Merino sheep. All experiments were approved by and performed according to the guidelines of the CSIRO Chiswick NSW Animal Ethics Committee (AEC approval numbers (01/19, 02/03, 03/62, 04/30). Colour photographs displaying the dorsal, ventral, sides and face were taken of each pigmented animal for determination of the self-colour black or badgerface Merino phenotypes. Sheep were sourced from a wide geographic distribution throughout Australia and animals from different producers were presumed to be unrelated. The Merino DNA panel consisted of samples from 94 baldy self-colour black sheep (from 28 different producers); 48 spotted self-colour black sheep (from 16 producers); 48 badgerface animals (from 12 producers); 87 white carrier animals (57 confirmed to be carriers by commercial pedigree testing and 30 suspected carriers) from a total of 20 different producers; and 96 random white animals (of unknown carrier status) from 22 producers. Barbary sheep (*A. lervia*) blood and ventral skin samples were sourced from the Western Plains Zoo, Dubbo, NSW Australia.

Preparation of gDNA and total RNA. QiAmp DNA Mini kit (Qiagen) was used to extract genomic DNA from whole blood. Skin biopsies were collected in RNA Later (Ambion) and stored at -80 °C. Total RNA was prepared from the skin of white and recessive black Merino sheep and from the ventral skin of Barbary sheep using Trizol (Invitrogen) in accordance with the manufacturer's recommendation. In each extraction, 5 ml of Trizol was used to extract 200-500 mg of skin. To ensure removal of contaminating DNA, total RNA was treated with Dnase1 (Ambion DNA-free Austin, TX, USA). RNA quality was visually assessed using agarose gel electrophoresis and a UV transilluminator and quantified by spectrophotometry.

***ASIP* PCR, cloning and sequencing.** PCR primers (Supplemental Table S2 online) for the amplification of intron sequences from genomic DNA templates were designed from sheep *ASIP* EST sequence. PCR primers were designed to amplify intron 1 regions between non-coding exon 1E and coding exon 2, intron 2 regions between exon 2 and

exon 3 and intron 3 regions between exons 3 and 4. PCR reactions (20 µl) contained 50 ng gDNA, 200 µM dNTP's, 10 pmole of each primer, 1 x Q solution, 1.5 mM $MgCl_2$ 0.5 units Taq polymerase (Qiagen) and 1 x reaction buffer (Qiagen). PCR conditions were 94 °C for 3 min; (94 °C for 30 s, 57 °C for 1 min and 72 °C for 2 min) for 35 cycles; and 72 °C for 5 min. PCR products were amplified from a white merino ram and a self-colour black merino lamb. PCR products were cloned using a TOPO TA Cloning Kit (Invitrogen) and sequenced using Big Dye Terminator 3.1 reaction mix. PCR products spanning exons 2, 3 and 4 were generated (see Supplemental Table S2 primer sequences) and sequenced directly to identify coding mutations in the DNA of selected animals from the panel of Merino coat colour phenotypes. For direct sequencing, PCR products were processed with ExoSAP-IT® (USB Corporation) according to the manufacturer's recommendations. All sequencing was done using ABI Prism Big Dye Terminator 3.1 chemistry and either an ABI 377 Prism DNA autosequencer or ABI 3130xl Genetic Analyser (Applied Biosystems, Australia). Sequence data was imported into Sequencher v4.2 (Gene Codes Corp., Ann Arbor, MI) for analysis. Sequences from the white and black Merino identified as Haplotypes 1 and 2 in Supplemental Table S1 have been deposited in GenBank (accession nos. EU420022 and EU420023).

Genotyping polymorphisms. The presence of D_9 and D_5 polymorphisms in exon 2 of the *ASIP* coding region was analysed by PCR spanning both indels (see Supplemental Table S2 primer sequences) in a 20 µl PCR reaction volume with Qiagen Taq performed according to the manufacturer's recommendations. Alleles were visualised by autoradiography and/or fluorescence on an ABI 377 Prism DNA autosequencer or ABI 3130xl Genetic Analyser (Applied Biosystems, Australia). A TaqMan MGB genotyping assay was developed for the g.5172T>A SNP in exon 4 of ovine *ASIP* (Assay by Design - Applied Biosystems). Genotypes of the panel of Merino coat colour phenotypes were automatically assigned with a quality score of 90, using the Applied Biosystems 7900HT sequence detector

Characterisation of *Ovis aries* BAC library clones. The three INRA Romanov Sheep BAC library clones (Vaiman et al. 1999) were obtained from the BAC-YAC Resource centre of the Animal Genetics Department of INRA. (http://dga.jouy.inra.fr/grafra/INRA_libraries_database_simplified.htm.). Shotgun plasmid libraries were prepared by the Australian Genome Research Facility (AGRF, http://www.agrf.org.au/) from each of three INRA ovine BAC clones INRA-164H8, INRA-218G7 and INRA-229C6 DNA which had tested positive for *ASIP* coding sequence by PCR. Clones INRA-164H8 and INRA-218G7 were sequenced to 8x coverage and clone INRA-229C6 to 2x coverage by the AGRF (GenBank accession nos. EU185098-EU185100). Sequence data was imported into Sequencher v4.2 (Gene Codes) for analysis. Clones INRA-218G7 (~75kb) and INRA-229C6 (~150kb) partially shared the same sequence and were aligned to the longer clone INRA-164H8 (~150kb). Twenty nine Texel Sheep BAC clones (Dalrymple et al. 2007) predicted to span the region within two Megabase pairs of *ASIP* were ordered from the Children's Hospital and Research Center at Oakland (CHRCO) (http://bacpac.chori.org/home.htm). The 29 BAC clones were PCR analysed for the presence or absence of both coding and non-coding regions of *ASIP*, *AHCY*, and *ITCH* genes. Primers used in this analysis are listed in Supplemental Table S2 online. BAC clone INRA-164H8 was used as the ovine reference sequence. BAC-end sequences from clones CH243-455O4 and CH243-489F15 that aligned to the reference sequence in a 'head-to-head' orientation (with 5' ends closest) were predicted to span a junction between duplicated DNA copies. BAC clones with end sequences that aligned to the reference sequence in a 'tail-to-tail' orientation (with 3' ends closest) were predicted to be fully contained within the reference sequence. Clone CH243-373J16 was predicted to provide sequence data further 5' and clone CH243-160L8 further 3' of the INRA-164H8 sequence. Clones CH243-234K21 and CH243-373J16 contained the putative non-functional 'A' *ASIP* allele. These five clones of the 29 CHORI Texel sheep BAC clones - CH243-455O4, CH243-489F15, CH243-373J16, CH243-160L8, CH243-234K21- were selected for complete sequencing to 6x coverage (Macrogen, Korea) (http://www.macrogen.com/eng/macrogen/macrogen_main2.jsp). Sequences were analysed using Sequencher software v4.2 (GeneCodes). Sequences have been deposited in Genbank (GenBank accession nos. EU185093-EU185097).

Haplotypes phylogenetic analysis

The ovine *ASIP* haplotype sequences from this study were aligned and phylogenetic analyses performed using Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0 (Tamura et al. 2007). The p-distance was used to estimate genetic distances and Neighbour -Joining used to construct the phylogeny. Testing of inferred phylogeny was by bootstrap with 500 replications.

RT-PCR analysis. First strand oligo$(dT)_{15}$ primed cDNA was generated using SuperScript™ III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. PCR products were amplified, labelled with ^{33}P-dCTP, and separated by electrophoresis on a 5% polacrylamide gel and analysed by autoradiography or visualised with ethidium bromide staining following electrophoresis on a 1.0% agarose gel.

5' and 3' RACE. For 5' and 3' RACE experiments respectively, oligo$(dT)_{18}$ primed or UAP primed cDNA was prepared from 1-2 µg of total RNA using SuperScript™ III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. oligo$(dT)_{18}$ primed cDNA was poly A-tailed with dATP using Terminal Transferase (NEB) and excess primer and dNTPs removed using Qiaquick PCR Purification columns (Qiagen). 5' RACE products were amplified in touchdown PCR reactions (Tm 68-63 ^{o}C) with primers GSP1 5' and UAP (Supplemental Table S2) and Qiagen Taq according to the manufacturer's instructions, followed by a nested PCR with primer GSP2 5' and UAP. Nested PCR was also used to amplify 3' RACE products. Primers GSP1 3' and UAP were used in the first PCR reaction followed by primers GSP2 3' and UAP in the second PCR reaction. PCR products were visualised by electrophoresis and ethidium bromide staining on 1.5% agarose gels. Products were gel purified and cloned for sequencing using TOPO TA Cloning (Invitrogen).

Junction point and 5' breakpoint assay.

To estimate ovine *ASIP* copy numbers an asymmetric competitive PCR protocol was developed following modification to the PCR method of Pielberg et al. (2003). Both the

junction between duplicated DNA copies and the 5' breakpoint sequences were amplified in the one reaction tube. Primers Agt16 and Agt18 (Supplemental Table S2) spanning the junction between the duplicated copies produced a unique 242 bp product, while Agt16 and Agt17 spanning the 5' breakpoint sequence produced a 238 bp product (Supplemental Figure S2). The products were amplified by asymmetric competitive PCR in a 5 µl reaction volume in 384-well plates containing 5-50 ng gDNA, 500 µM dNTP's, 3 mM $MgCl_2$, 1.25 x reaction buffer (Qiagen), 5 pmole of specific forward primers (Agt17 and Agt18), 0.05 pmole of shared reverse primer (Agt16) and 0.15 units of HotStar Taq polymerase (Qiagen) PCR conditions were 95 °C for 15 min; (95 °C for 30 s, 60 °C for 30s and 72 °C for 1 min) for 40 cycles and 72 °C for 5 min. A 1 µl aliquot of a 1:10 dilution of PCR products was analysed on an ABI 3130xl Genetic Analyser and the data analysed using GeneMapper software (Liz 600 as size standard). The number of junction points (and *ASIP* copies) is estimated from the standard curve (Supplemental Figure S3) using the ratio of the area under the 242bp (junction point) peak compared to the total area represented by both peaks (junction point and breakpoint). To deduce discrete copy number categories for the data, peak areas were analysed using the mean and standard deviation for each expected category (1, 2, 3 or 4 copies) and the normdist statistical function.

ACKNOWLEDGEMENTS

The assistance of S McWilliam in the identification of the CHORI-243 BACs and management of BAC clone sequences and B Dalrymple for demonstrating how to use the virtual sheep genome and related resources to identify sheep BACs and for his advice during the progress of the project is gratefully acknowledged. We thank D Maxwell for sheep industry advice, assistance with industry liaison, the Merino photograph and together with G Uphill for assistance with organization and collection of Merino Sheep industry animal samples. We thank the Australian sheep industry producers who provided sheep samples. Sincere thanks to T Portas and the staff of Western Plains Zoo Dubbo, NSW Australia for the Barbary sheep samples and photograph. We thank J McEwan for the Texel and J Kirts for the Romanov sheep photographs. Thank you to I Franklin and Y LI for stimulating discussions and suggestions on genetics, agouti and sheep

pigmentation. Thank you to G Davies, R Forage, W Barendse and J van der Werf for their support. We are grateful to SheepGenomics, an initiative of Australian Wool Innovation Limited and Meat and Livestock Australia for bioinformatics support. Thank you to G Wijffels, B Dalrymple, W Barendse and F Nicholas for comments on the manuscript. We gratefully acknowledge the BAC-YAC Resource centre of the Animal Genetics Department of INRA in providing the INRA Romanov sheep BAC clones and the Children's Hospital & Research Centre at Oakland for the Texel sheep BAC clones. This work was a project of the Co-operative Research Centre for the Australian Sheep Industry, Armidale, Australia.

FIGURE LEGENDS

Figure 1. Schematic of the structure of the ovine *ASIP* gene showing the sizes of the 3 coding exons (black) and the intervening intron sequences (white). The exon-intron organisation of the ovine gene is similar to that reported for the bovine, human, and mouse genes. Coding exons 2, 3, and 4 are separated by 1312 and 3481 bp intron sequences respectively. The nucleotide positions of various described recessive black 'non-agouti' mutations for the mouse, rat, horse, cat, fox, and dog are shown. The positions of 4 mutations identified from sheep in this study are shown.

Figure 2. Illustration of sheep coat colour patterns. Three coat colours in the Australian Merino; (A) the dominant white, (B) the non-agouti, also known as recessive self-colour-black, and (C) the badgerface pattern. The dominant white coat colour in Texel is displayed in D. Romanov sheep (E) appear to have a pattern similar to recessive self-colour-black and are proposed as homozygous for the most recessive *ASIP* allele, A^a. (F) displays the Barbary sheep wild-type coat colour. Hypothesized genotypes for *ASIP* are indicated.

Figure 3. Allele specific expression of *ASIP* transcripts in nine white Merino sheep. Autoradiographs of PCR products from gDNA and cDNA from the skin of nine white sheep are shown. The size difference between the 189 (226)- and 184 (221)-, and the 189

(226)- and 180 (217)-bp fragments is due to the corresponding D_5 and D_9 deletions in the latter fragments. Seven sheep had both the N_9N_5 and N_9D_5 alleles. In one sheep (Lane 5) all three alleles N_9N_5, N_9D_5, and D_9N_5 were present. One sheep had only the non-deleted N_9N_5 alleles (Lane 6). The nine white animals showed allele specific expression of the functional N_9N_5 and D_9N_5 alleles. Expression of the non-functional N_9D_5 allele (arrow) was not detected. The different band intensities in the gDNA autoradiograph likely reflect the different allele copy numbers present in each animal.

Figure 4. RT-PCR end-point gene expression analysis of ovine *ASIP* and *ITCH* in five tissues of a white and a self-colour black Merino sheep. The position of primers used to amplify each product is shown in a schematic to the right of the corresponding 1.5% agarose gel image. It, It' and IE are non-coding exons. Open boxes Ex2, Ex3 Ex4 (*A*, *B* and *C*) and Ex17, Ex18, Ex1 and Ex2 (*D*, *E* and *F*) represent coding exons of the *ASIP* or *ITCH* genes respectively. Other alternatively spliced transcripts of the agouti gene have also amplified as evident from the bands in panel *B*. The most abundant *ASIP* transcripts are those with exon IE (~368 bp as shown in the panel *B* schematic). It and It' were both present in *ITCH* transcripts and generated bands of the expected size (based on similarity to the bovine sequence) but no alternatively spliced forms were evident. 3' RACE (panel *C*) showed *ASIP* transcripts in all tissues studied to have the same polyadenylation site.

Figure 5. Schematic showing the sequenced ovine BAC clones below the structure of the ovine genomic tandem duplication and an alignment of the breakpoint and junction point sequences. (*A*) Non-coding exons are shown as open boxes. Protein coding exons are shown as solid black boxes. Arrows above or below the genes indicate the direction of transcription. Clones INRA-164H8, INRA-218G7 and INRA-229C6 were from the INRA Romanov Sheep BAC library (Vaiman et al. 1999) and clones CH243-160L8, CH243-234K21, CH243-455O4, CH243-373J16, CH243-489F15 from the CHORI-243 Texel sheep BAC library (Dalrymple et al. 2007). The three INRA BAC clones did not span the Junction, 5' or 3' breakpoints. Clones CH243-455O4, CH243-489F15 and CH243-234K21 spanned the junction between copies; clone CH243-160L8 spanned the 3' breakpoint and clone CH243-373J16 the 5' breakpoint. Seven *ASIP* transcripts identified

from the skin of a white Merino sheep (*Ovis aries*) and one transcript identified from the ventral skin of a Barbary sheep (*Ammotragus lervia*) are shown in the lower section. The coding exons 2, 3, and 4 and Barbary sheep non-coding exon 1A-like are numbered according to the nomenclature of Bultman et al. (1992) and Vrieling et al (1994). All other non-coding exons are named alphabetically, IA to IE. Exons It and It' are non-coding exons of both *ITCH* and *ASIP* Merino transcripts (see also Fig. 4). The positions of the 5' and 3' breakpoints are located 5' of the *ASIP* and *ITCH* coding sequence regions respectively. The *ITCH* promoter including non-coding exons It, It' and IA was duplicated and positioned upstream of the duplicated *ASIP* non-coding exons IB to IE creating a new ovine hybrid *ITCH/ASIP* promoter. The complete *ASIP* and *AHCY* coding exons were also within the approximately 190 kb duplicated segment. Not drawn to scale. (*B*) DNA sequences from five Texel sheep BAC clones comprising the regions spanning the 5' and 3' breakpoints and the junction point between the gene duplication. Sequence identity to the master sequence is shown with a dash. The boundaries of the 143 bp of sequence similarity are marked with vertical arrows. Flanking sequences unique to the 5' and 3' breakpoint regions respectively are in bold. The highlighted (grey) regions of clone CH243-160L8 sequence of the 3' Breakpoint was identified by Repbase (Kohany et al. 2006) as having 82% identity to a region of Bovidae non-LTR/RTE BDDF2 repetitive SINE sequence. The highlighted (grey) regions of clones CH243-373J16 (of the 5' breakpoint) and clones CH243-489F15, CH243-234K21 and CH243-455O4 (all of the junction point) were identified as having 87-88% identity to a region of Bovidae non-LTR BOV2 repetitive SINE sequence.

Figure 6. A model for the organisation and evolution of the ovine *ASIP* locus in dominant white and recessive black domestic sheep. The proposed positions (*A*) and clustering (*B*) of copy 1 (inactive) and copy 2 (active) haplotypes identified from Texel, Romanov and Merino sheep are shown. Solid arrows indicate high levels of expression of haplotypes at copy 2. A dashed arrow indicates expression of a haplotype 3-like D_9 allele detected in Merino skin. The two haplotypes identified in the Romanov BAC library may not be derived from a duplicated *ASIP-ITCH* region as pure Romanov sheep have a recessive black-like phenotype. However, haplotype 4 clusters with haplotype 5 in copy 1

of the Texel and haplotype 3 clusters with haplotype 6 and 7 from the Texel and haplotype 1 from the Merino (*B*). This clustering suggests that the Romanov haplotypes are derived from the two different copies of the putative original gene duplication. The low frequency presence of a haplotype 4/5-like copy 1 region in the Merino sheep population is deduced from the identification of N_9N_5A genotypes in the black Merinos and is consistent with the organisation of the region in the Texel sheep CHORI-243 library BACs. (*C*) Schematic showing resolution of nonallelic pairing between duplicated *ASIP-ITCH* copies with crossover products showing reciprocal deletion and triplication. The 190 kb segments of copy 1 and copy 2 of a white Merino are shown as light and dark grey boxes respectively. Black boxes indicate the similar SINE sequence regions at the junction, 5' and 3' breakpoints. In this example, the *ASIP* genotype shown represents a white animal heterozygous at copy 1 for the non-functional alleles (N_9D_5T/ N_9N_5A) and homozygous at copy 2 for the functional alleles (N_9N_5T/ N_9N_5T). Arrows below the *ASIP* genes indicate the direction of transcription driven by the *ITCH* promoter ($ITCH^P$). Nonallelic pairing and crossover between the junction and 3' breakpoints (dark dashed line) would result in creation of a single copy 1 non-functional (N_9N_5A) allele, as shown. The resulting single *ASIP* copy is not expressed as the ancestral *ASIP* promoter ($ASIP^P$) is silent. A crossover point (grey cross) that would result in the positioning of a non-functional (N_9N_5A) allele under the regulation of the duplicated *ITCH* promoter is unlikely to occur as expression of 'A' alleles was not detected in Merino sheep. Variable positioning of functional haplotypes at copy 1 and 2 combined with nonallelic pairing and crossover explains the different single copy alleles identified in black Merinos. Mutations and gene conversion events could also contribute to the diversity at the locus.

FIGURES

Figure 1.

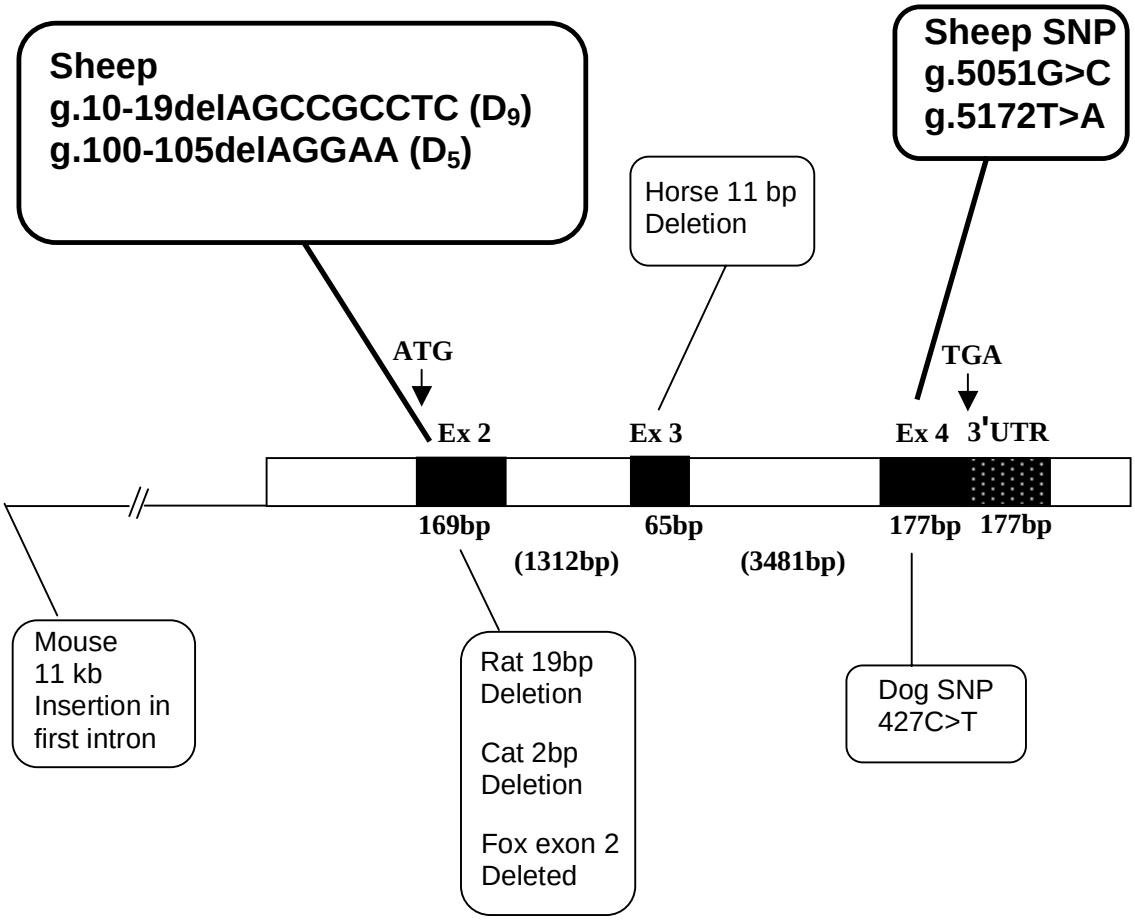


Figure 2.

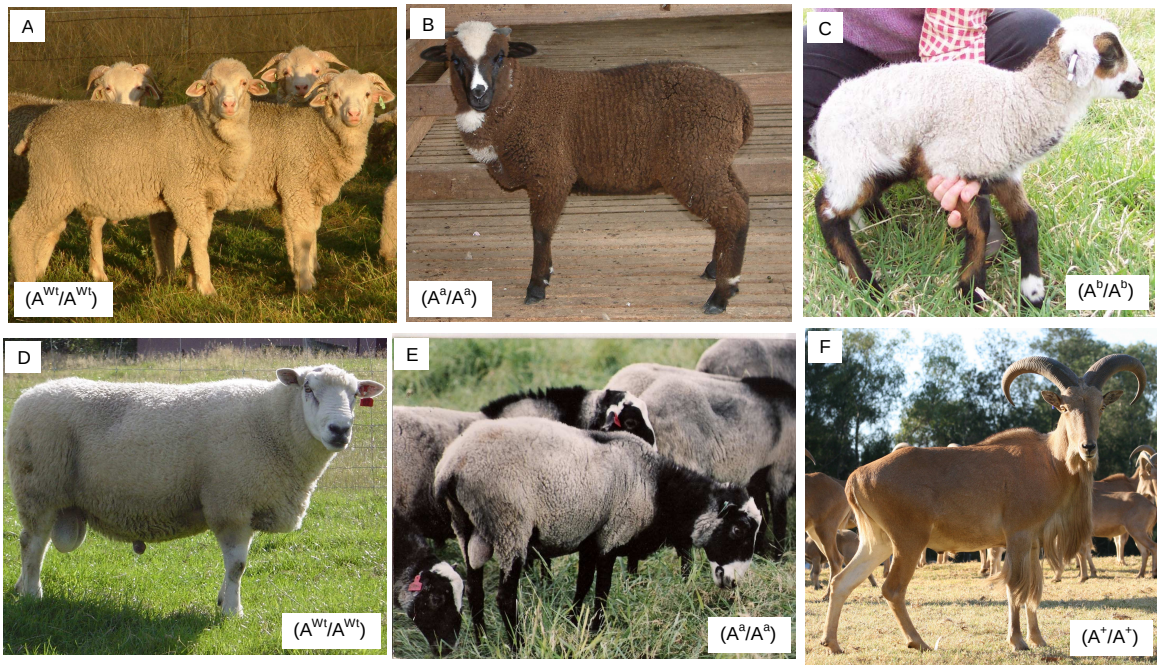


Figure 3.

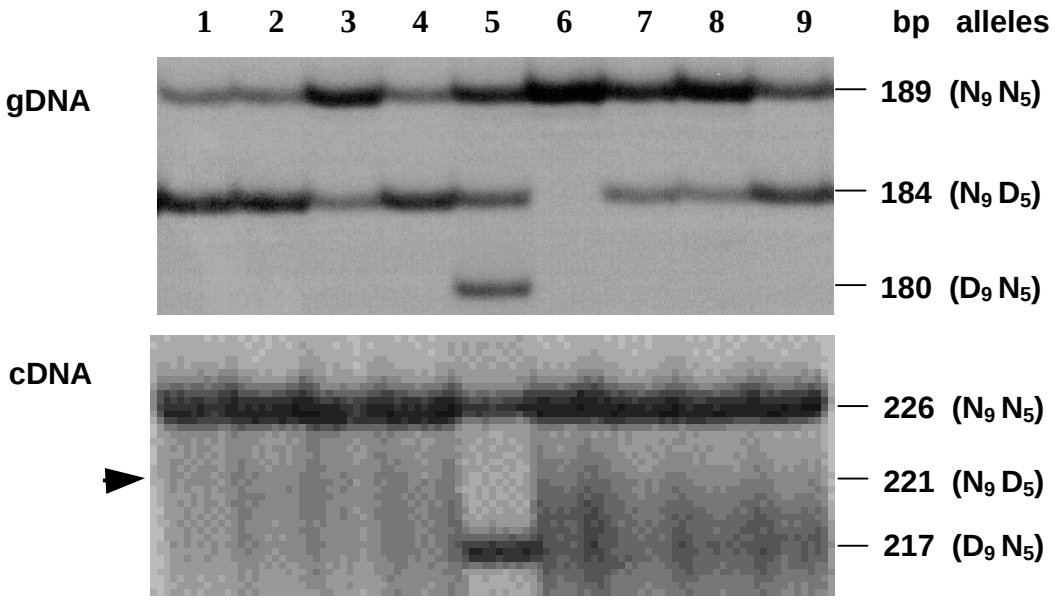


Figure 4

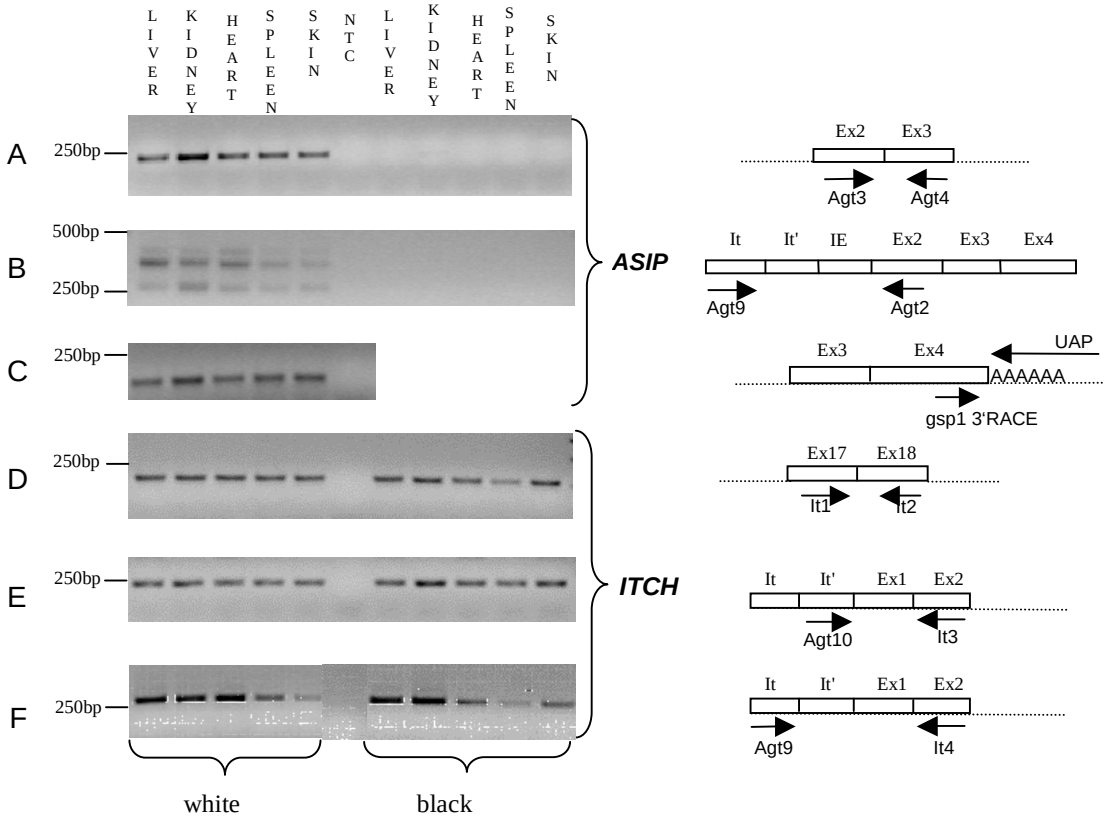
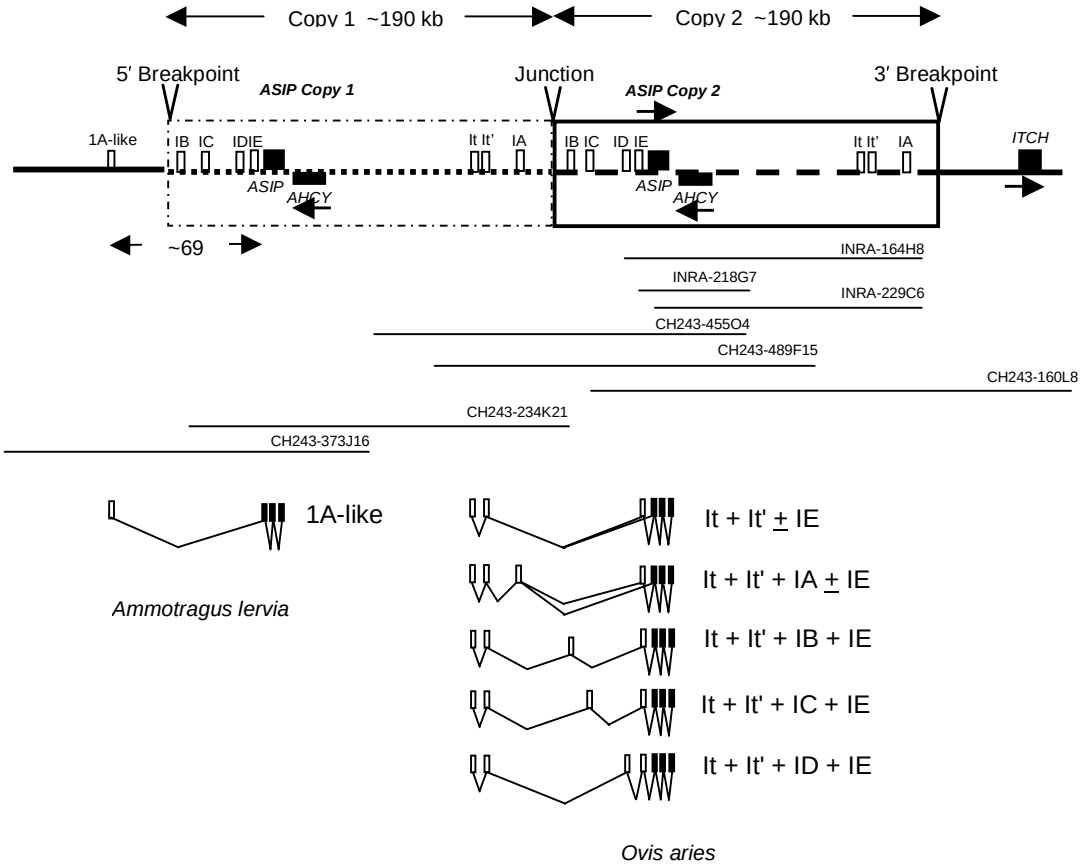
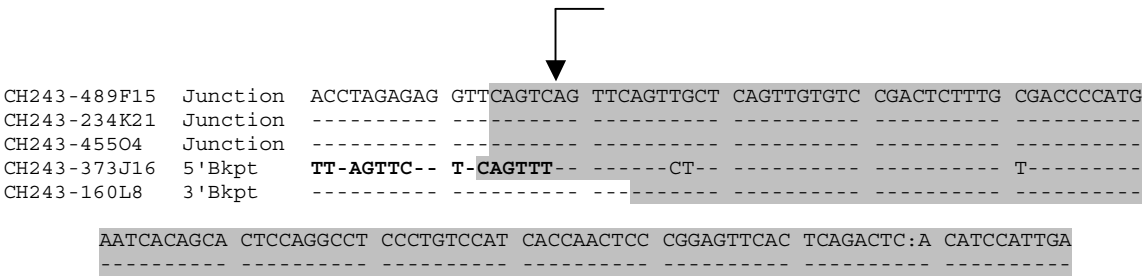


Figure 5

A.



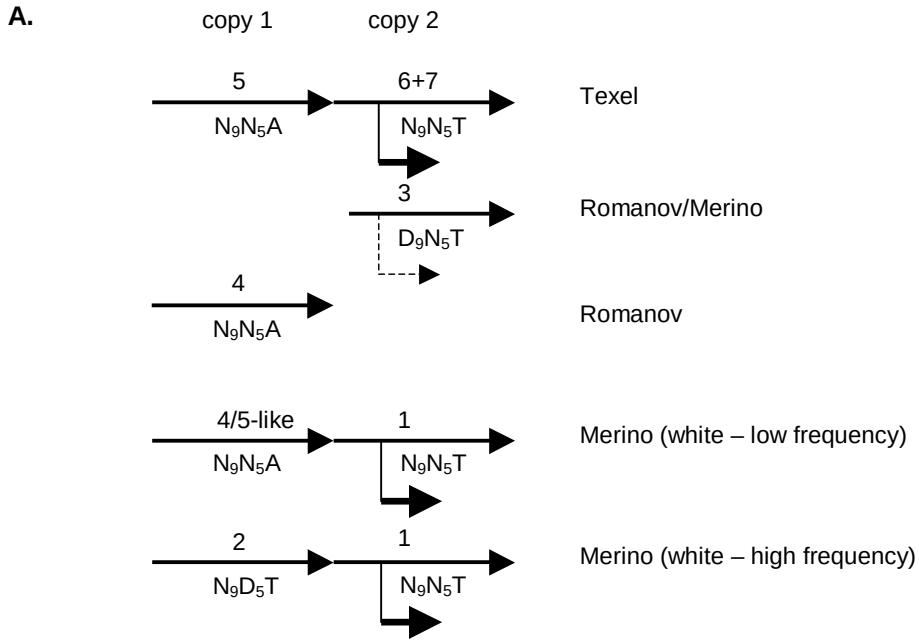
B.

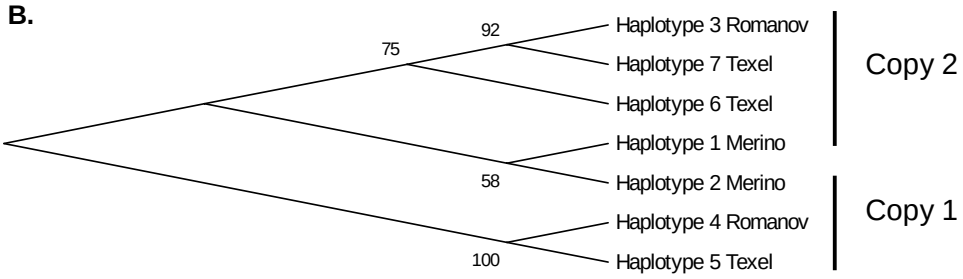


```
---------- ---------- ---------- ---------- ---------- ---------- ----------
---------- -A-------- ---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- T-AA------ T--------- ---C----A- -G-----C--

GTCAGTGATA CCATCCAGCC ATCTCATCCT CCTTCCCCCA AAACTCACGT CCTCATTGCT
---------- ---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- --------   ---------- ----------
---C-----G ---------- ---------- -TG--GT--C CTT---CTCC TGC-CCCAG-
```

Figure 6

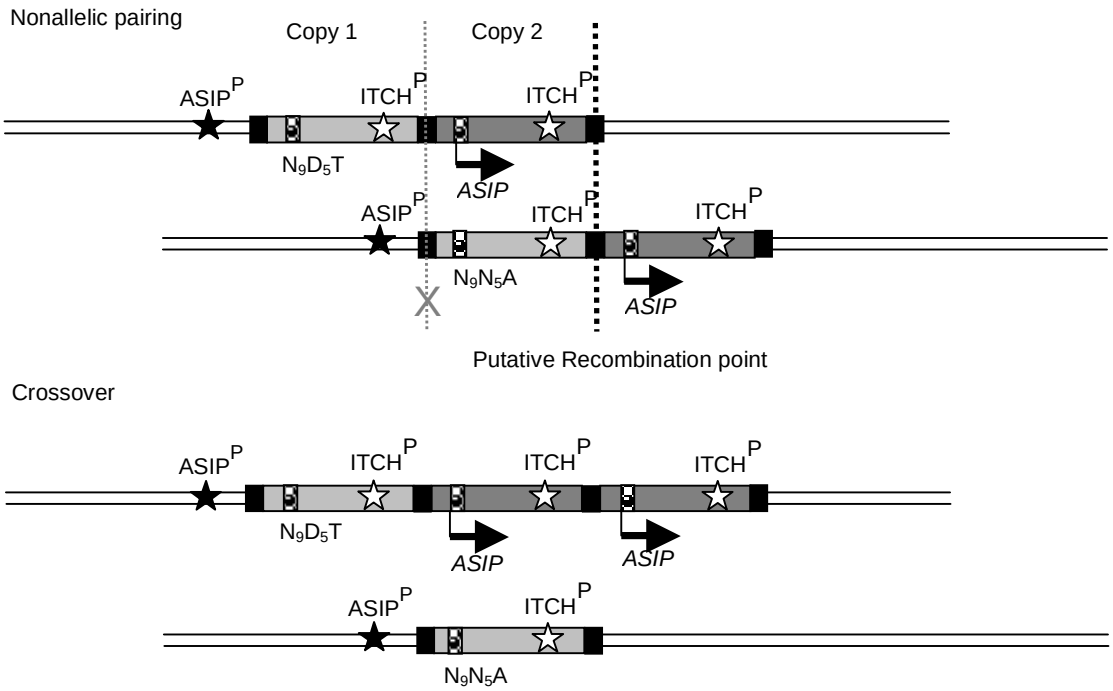




891

892

C.



893
894

895

896

897 **Table 1**. Presence of junction points and estimated *ASIP* copy numbers associated with
898 dominant white coat colour in different sheep populations

899

Population	Proposed Genotypes	Total no. of Junction Points (*ASIP* copies)					Total
		0 (2)	1* (3)	2 (4)	3 (5)	4 (6)	
Recessive Black Merino	A^a/A^a or A^b/A^b	180	-	-	-	-	180
White Carrier Merino	A^{Wt}/A^a	0	47	29	9	1	86
Random White Merino	A^{Wt}/A^{Wt} or A^{Wt}/A^a	0	8	67	16	0	91
Barbary Sheep	A^+	40	-	-	-	-	40

*the single junction point indicates these white animals have a duplicated allele and a single copy black allele

REFERENCES

Adalsteinsson, S. 1970 Colour inheritance in Icelandic sheep and relation between colour, fertility and fertilisation. . *J. Agric. Res. in Iceland* **2:** 3-135.

Adalsteinsson, S. 1975. Depressed fertility in Icelandic sheep caused by a single colour gene. *Ann. Genet. Sel. Anim.* **7:** 445-447.

Bennett, D.C. and Lamoreux, M.L. 2003. The colour loci of mice-a genetic century. *Pigment Cell Res.* **16:** 333-44.

Bischof, J.M., Chiang, A.R., Scheetz, T.E., Stone, E.M., Casavant, T.L., Sheffield, V.C., and Braun, T.A. 2006. Genome-wide identification of pseudogenes capable of disease-causing gene conversion. *Hum. Mutat.* **27:** 545-552.

Bonilla, C., Boxill, L.A., Donald, S.A., Williams, T., Sylvester, N., Parra, E.J., Dios, S., Norton, H.L., Shriver, M.D., and Kittles, R.A. 2005. The 8818G allele of the agouti signaling protein (ASIP) gene is ancestral and is associated with darker skin colour in African Americans. *Hum. Genet.* **116:** 402-6.

Brooker, M.G. and Dolling, C.H.S. 1969. Pigmentation of Sheep. II. The inheritance of coloured patterns in black Merino. *Aust. J. of Agric. Res*. **20:** 387-394.

Bultman, S.J., Klebig, M.L., Michaud, E.J., Sweet, H.O., Davisson, M.T., and Woychik, R.P. 1994. Molecular analysis of reverse mutations from nonagouti (a) to black-and-tan (a^t) and white-bellied agouti (A^w) reveals alternative forms of agouti transcripts. *Genes Dev.* **8:** 481-90.

Bultman, S.J., Michaud, E.J., and Woychik, R.P. 1992. Molecular characterization of the mouse agouti locus. *Cell* **71:** 1195-204.

Carroll, L., Voisey, J., and van Daal, A. 2005. Gene polymorphisms and their effects in the melanocortin system. *Peptides* **26**:1871–1885

Dalrymple, B.P, Kirkness, E.F., Nefedov, M., McWilliam, S., Ratnakumar, A., Barris, W., Zhao, S., Shetty, J., Maddox, J.F., O Grady, M., et al. 2007. Using comparative genomics to reorder the human genome sequence into a virtual sheep genome.

Genome Biol. **8:** R152 doi:10.1186/gb-2007-8-7-r152

Drogemuller, C., Giese, A., Martins-Wess, F., Wiedemann, S., Andersson, L., Brenig, B., Fries, R., and Leeb, T. 2006. The mutation causing the black-and-tan phenotype of Mangalitza pigs maps to the porcine *ASIP* locus but does not affect its coding sequence. *Mamm. Genome.* **17:** 58-66

Duhl, D.M., Vrieling, H., Miller, K.A., Wolff, G.L., and Barsh, G.S. 1994. Neomorphic agouti mutations in obese yellow mice. *Nat. Genet.* **8:** 59-65.

Eizirik, E., Yuhki, N., Johnson, W.E., Menotti-raymond, M., Hannah, S.S., and Obrien, S.J. 2003. Molecular genetics and evolution of melanism in the cat family. *Curr. Biol.* **13:** 448-453.

Eppig, J.T., Bult, C.J., Kadin, J.A., Richardson, J.E., Blake, J.A., Anagnostopoulos, A., Baldarelli, R.M., Baya, M., Beal, J.S., Bello, S.M., Boddy, W.J., et al. 2005. The Mouse Genome Database (MGD): from genes to mice-a community resource for mouse biology. *Nucleic Acids Res.* **33:** D471-D475.

Girardot, M., Guibertm, S., Laforet, M.P., Gallard, Y., Larroque, H., and Oulmouden, A. 2006. The insertion of a full-length Bos taurus line element is responsible for a transcriptional deregulation of the Normande Agouti gene. *Pigm. Cell Res.* **19:** 346-355.

Girardot, M., Martin, J. Guibert, S., Leveziel, H., Julien, R., and Oulmouden, A. 2005. Widespread expression of the bovine Agouti gene results from at least three allternative promoters. *Pigm. Cell Res.* **18:** 34-41.

Giuffra, E., Tornsten, A., Marklund, S., Bongcam-Rudloff, E., Chardon, P., Kijas, J.M., Anderson, S.I., Archibald, A.L., and Andersson, L. 2002. A large duplication associated with dominant white colour in pigs originated by homologous recombination between LINE elements flanking KIT. *Mamm. Genome* **13:** 569-577.

Hayman, R.H. and Cooper, D.W. 1965. The frequency of pigmented sheep in the Australian Merino. *Wool Technology and Sheep Breeding* **12:** 81-85.

Henry, B.A. 2003 Links between the appetite regulating systems and the neuroendocrine hypothalamus:lessons from the sheep. *J. Neuroendocrinol.* **15**: 697-709

Kanetsky, P.A., Swoyer, J., Panossian, S., Holmes, R., Guerry, D., and Rebbeck, T.R. 2002. A polymorphism in the agouti signalling protein gene is associated with human

963 pigmentation. *Am. J. Hum. Genet.* **70:** 770-775.
964 Kerns, J.A., Newton, J., Berryere, T.G., Rubin, E.M., Cheng, J.F., Schmutz, S.M., and
965 Barsh, G.S. 2004. Characterization of the dog Agouti gene and a nonagouti mutation
966 in German Shepherd dogs. *Mamm. Genome* **15:** 798-808.
967 Kohany, O., Gentles, A., Hankus, L., and Jurka, J. 2006. Annotation, submission and
968 screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC*
969 *Bioinformatics* **7:** 474.
970 Kuramoto, T., Nomoto, T., Sugimura, T., and Ushijima, T., 2001. Cloning of the rat
971 Agouti gene and identification of the rat nonagouti mutation. *Mamm. Genome* **12:**
972 469-71.
973 Kwon, H.Y., Bultman, S.J., Loffler, C., Chen, W.J., Furdon, P.J., Powell, J.G., Usala,
974 A.L., Wilkison, W., Hansmann, I., and Woychik, R.P. 1994. Molecular structure and
975 chromosomal mapping of the human homolog of the Agouti gene. *Proc. Natl. Acad.*
976 *Sci.* **91:** 9760-9764.
977 Leeb, T., Deppe, A., Kriegesmann, B., and Brenig, B. 2000. Genomic structure and
978 nucleotide polymorphisms of the porcine Agouti signalling protein gene (ASIP).
979 *Anim. Genet.* 31: 335-336.
980 Lindsay, S.J., Khajavi, M., Lupski, J.R., and Hurles, M.E. 2006. A Chromosomal
981 rearrangement hotspot can be identified from population genetic variation and is
982 coincident with a hotspot for allelic recombination. *Am. J. Hum. Genet.* **79:** 890-902.
983 Marklund, S., Kijas, J., Rodriguez-Martinez, H., Ronnstrand, L., Funa, K., Moller, M.,
984 Lange, D., Edfors-Lilja, I., and Andersson, L. 1998. Molecular basis for the dominant
985 white phenotype in the domestic pig. *Genome Res.* **8:** 826-33.
986 McCann, J. P., Bergman, E.N., and Beermann, D.H. 1991. Dynamic and Static Phases of
987 Severe Dietary Obesity in Sheep: Food Intakes, Endocrinology and Carcass and
988 Organ Chemical Composition. *J. Nutr.* **122**: 496-505
989 Mundy, N.I. and Kelly, J., 2006. Investigation of the role of the Agouti signalling protein
990 gene (ASIP) in coat colour evolution in primates. *Mamm. Genome* **17:** 1205-1213.
991 Myers, S.R. and McCarroll, S.A. 2006. New insights into the biological basis of genomic
992 disorders. *Nat. Genet.* **38:** 1363-1364.
993 Nakayama, K. and Ishida, T. 2006. Alu-Mediated 100-kb deletion in the primate genome:

994 the loss of the Agouti signalling protein gene in the lesser Apes. *Genome Res.* **16:**
995 485-490.
996 Parsons, Y.M., Fleet, M.R., and Cooper, D.W. 1997. Investigation of the gene
997 responsible for recessive pigmentation in Australian Merino sheep. *Proc. Assoc.*
998 *Advanc. Anim. Breed. Genet.* **12:** 447-451,
999 Parsons, Y.M., Fleet, M.R., and Cooper, D.W. 1999a. The Agouti gene: a positional
1000 candidate for recessive self-colour black pigmentation in Australian Merino sheep.
1001 *Aust. J. Agric. Res.* **50:** 1099-1103
1002 Parsons, Y.M., Fleet, M.R., and Cooper, D.W. 1999b. Isolation of the ovine Agouti
1003 coding sequence. *Pigm. Cell Res.* **12:** 394-397
1004 Perry, W.L., Hustad, C.M., Swing, D.A., O'Sullivan, T.N., Jenkins, N.A., and Copeland,
1005 N.G., 1998. The itchy locus encodes a novel ubiquitin protein ligase that is disrupted
1006 in a^{18H} mice. *Nat. Genet.* **18:** 143-146.
1007 Rendel, J. 1957. Heredity of colour and white designs in farm animals. *Kunglishe*
1008 *Hogskole Landbruksadamie Tidskrift* **96:** 207-263.
1009 Rieder, S., Taourit, S., Mariat, D., Langlois, B., and Guerin, G. 2001. Mutations in the
1010 agouti (ASIP), the extension (MC1R), and the brown (TYRP1) loci and their
1011 association to coat colour phenotypes in horses (*Equus caballus*). *Mamm. Genome*
1012 **12:** 450-455.
1013 Smit, M.A., Shay, T.L., Beever, J.E., Notter, D.R., and Cockett, N E. 2002. Identification
1014 of an agouti -like locus in sheep. *Anim. Genet.* **33:** 383-385.
1015 Sponenberg, D.P, 1997. Genetics of colour and hair texture. In *The genetics of sheep*
1016 (eds. L.R. Piper and A. Ruvinsky), pp51-86. C.A.B. International, Oxon, UK; New
1017 York, USA.
1018 Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. MEGA4: Molecular Evolutionary
1019 Genetics Analysis (MEGA) Software Version 4.0 *Mol. Biol. Evol.* **24**:1596–1599.
1020 Vage, D.I, Lu, D., Klungland, H., Lien, S., Adalsteinsson, S., and Cone, R.D. 1997. A
1021 non-epistatic interaction of agouti and extension in the fox, *Vulpes vulpes*. *Nat. Genet.*
1022 **15:** 311-315.
1023 Vaiman D, Billault A, Tabet-Aoul K, Schibler L, Vilette D, Oustry-Vaiman A, Soravito
1024 C, Cribiu E.P. 1999. Construction and characterization of a sheep BAC library of

three genome equivalents. *Mamm. Genome* **10:** 585-7.

Voisey, J., Gomez-Cabrera, M.D.C., Smit, D.J., Leonard, J.H., Sturm, R.A., and van Daal, A. 2006. A polymorphism in the Agouti signalling protein (ASIP) is associated with decreased levels of mRNA. *Pigm. Cell Res.* **19:** 226-231.

Voisey, J., Imbeault, P., Hutley, L., Prins, J.B. and van Daal A. 2002. Body Mass Index-Related Human Adipocyte agouti Expression Is Sex-Specific but Not Depot-Specific. *Obes Res* **10**: 447-452.

Voisey, J. and van Daal, A. 2002. Agouti: from mouse to man, from skin to fat. *Pigm. Cell Res.* **15:** 10-18.

Vrieling, H., Duhl, D.M., Millar, S.E., Miller, K.A., and Barsh, G.S. 1994. Differences in dorsal and ventral pigmentation result from regional expression of the mouse Agouti gene. *Proc. Natl. Acad. Sci.* **91:** 5667-71.

Wolff, G.L. 2003. Regulation of yellow pigment formation in mice: a historical perspective. *Pigm. Cell Res.* **16:** 2-15.

Wolff, G.L., Roberts, D.W., and Mountjoy, K.G. 1999. Physiological consequences of ectopic agouti gene expression: the yellow obese mouse syndrome. *Physiol. Genomics* **1:** 151-63.

1056