

ORIGINAL ARTICLE

The genetic basis of recessive self-colour pattern in a wild sheep population

J Gratten¹, JG Pilkington², EA Brown¹, D Beraldi², JM Pemberton² and J Slate¹¹Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK and ²Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK

Bridging the genotype–phenotype gap for traits of ecological and evolutionary importance in natural populations can provide a novel insight into the origin and maintenance of variation. Here, we identify the gene and putative causal mutations underlying a recessive colour pattern phenotype ('self' or uniform colour) in a wild population of primitive Soay sheep. We targeted the *agouti signalling protein (ASIP)* gene, a positional candidate based on previous study that mapped the *Coat pattern* locus to a presumptive region on chromosome 13. We found evidence for three recessive mutations, including two functional changes in the coding sequence and a putative third *cis*-regulatory mutation that inactivates the promoter. These mutations define up to five haplotypes in Soays, which collectively explained the coat pattern in all but one member of a complex multi-generational

pedigree containing 621 genotyped individuals. The functional mutations are in strong linkage disequilibrium in the study population, and are identical to those known to underlie the self phenotype in domestic sheep. This is indicative of a recent (and simultaneous) origin in Soay sheep, possibly as a consequence of past interbreeding with modern domestic breeds. This is only the second study in which *ASIP* has been linked to variation in pigmentation in a natural population. Knowledge of the genetic basis of self-colour pattern in Soay sheep, and the recognition that several mutations are segregating in the population, will aid future studies investigating the role of selection in the maintenance of the polymorphism.

Heredity (2010) **104**, 206–214; doi:10.1038/hdy.2009.105; published online 12 August 2009

Keywords: pigmentation; *ASIP*; wild population; *Ovis aries*; genetic heterogeneity; domestication

Introduction

Bridging the genotype–phenotype gap for traits of ecological and evolutionary importance in natural populations is a fundamental goal of evolutionary genetics (Feder and Mitchell-Olds, 2003; Stinchcombe and Hoekstra, 2008). Establishing such links can shed light on the origin and maintenance of phenotypic variation in the wild (Hanski and Saccheri, 2006; Ellegren and Sheldon, 2008; Gratten *et al.*, 2008; Anderson *et al.*, 2009), and for traits that are shared by wild and domestic relatives of the same species, can provide insights into the history of domestication (Fang *et al.*, 2009; Ludwig *et al.*, 2009). A valuable model system for addressing this challenge is pigmentation in mammals, because it is highly variable, particularly in domestic species (Hoekstra, 2006; Mundy, 2007), and the gene networks involved in melanin synthesis and transport have been characterized in domestic mice (Bennett and Lamoreux, 2003). Another reason is that mutations in a few key genes, such as the *melanocortin 1 receptor (MC1R)*, have been documented in many wild (Vage *et al.*, 1997; Ritland *et al.*, 2001; Nachman *et al.*, 2003) and domestic (Klungland *et al.*, 1995; Kijas *et al.*, 1998; Vage *et al.*, 1999; Newton *et al.*, 2000) mammals, supporting a candidate gene approach to gene discovery.

In a wild population of Soay sheep on St Kilda, a majority of individuals (~95%) have a typical wild-type mammalian pigmentation pattern (dark upper body and pale belly), and the remainder are uniformly, or 'self' coloured (Clutton-Brock and Pemberton, 2004) (Figure 1). This phenotype also occurs in domestic sheep, in which it is understood to be a product of positive artificial selection for novel colour variants during domestication (Sponenberg, 1997), as has been described in pigs (Fang *et al.*, 2009) and horses (Ludwig *et al.*, 2009). In contrast, Soay sheep are among the most primitive of all surviving breeds (Chessa *et al.*, 2009) and have experienced a long history of natural selection (Clutton-Brock and Pemberton, 2004). Therefore, there is considerable interest in the genetic basis, evolutionary origin and maintenance of the polymorphism in this population. Previous studies in Soay sheep indicate that inheritance of recessive self-colour pattern is consistent with the action of a single-locus *agouti*, at which the wild-type allele (A^+) is dominant to self (A^a) (Clutton-Brock and Pemberton, 2004). We recently mapped this locus to a region of chromosome 13 that is likely to contain the candidate gene, *agouti signalling protein (ASIP)* (Beraldi *et al.*, 2006).

The *ASIP* gene encodes for a signalling molecule that inhibits the binding of *MC1R* to its ligand, α -melanocyte stimulating hormone, inducing a switch from production of eumelanin (black or brown pigment) to pheomelanin (yellow or red pigment) in melanocytes (Jackson, 1994). Consequently, recessive mutations that either impair *agouti* protein function or abrogate *ASIP* expression result in more darkly pigmented self phenotypes.

Correspondence: Dr J Gratten, Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK.

E-mail: j.gratten@sheffield.ac.uk

Received 14 May 2009; revised 30 June 2009; accepted 13 July 2009; published online 12 August 2009



Figure 1 Wild (right) and self-type (left) coat pattern morphs in Soay sheep (Photograph: Arpat Ozgul).

Conversely, dominant mutations arising from deregulated expression of *ASIP* result in lighter phenotypes. *ASIP* has been linked with natural mammalian colour variation in only one instance (Steiner *et al.*, 2007), but numerous functional and regulatory mutations have been identified in domestic species, including mice (Bultman *et al.*, 1992), cattle (Girardot *et al.*, 2006), dogs (Kerns *et al.*, 2004), horses (Rieder *et al.*, 2001), pigs (Drogemuller *et al.*, 2006) and sheep (Smit *et al.*, 2002; Norris and Whan, 2008; Royo *et al.*, 2008).

In domestic sheep, mutations at *ASIP* account for two of the three pigmentation phenotypes that have been genetically characterized to date. Dominant white colour is caused by deregulated expression of the *agouti* protein, resulting from a 190-kbp genomic duplication that places a functional *ASIP*-coding sequence under the control of a duplicated promoter from the neighbouring itchy homolog E3 ubiquitin protein ligase (*ITCH*) locus (Norris and Whan, 2008). Recessive self-colour pattern is caused by an absence of *ASIP* expression, presumably arising from a *cis*-regulatory mutation that inactivates the promoter (Norris and Whan, 2008; Royo *et al.*, 2008), but may also result from the expression of functional mutations in exon 2 (Smit *et al.*, 2002; Royo *et al.*, 2008) and exon 4 (Norris and Whan, 2008). The third genetically characterized pigmentation phenotype in sheep is dominant black, which is caused by functional mutations in the *MC1R*-coding sequence that generate a constitutively activated receptor (Vage *et al.*, 1999). A further pigmentation phenotype has been characterized in wild Soay sheep; this is a dilution of the ancestral dark brown coat and is caused by a single non-synonymous mutation in the *TYRP1* gene, such that wild and self-type sheep can be either dark or light brown (Gratten *et al.*, 2007). This recessive colour polymorphism segregates independently of self-colour pattern (Coltman and Pemberton, 2004) and is therefore not the focus of this paper.

Neither the dominant white nor dominant black phenotypes have been documented in Soay sheep, whereas the mutations underlying recessive self-colour pattern in domestic sheep may be the same, or distinct

from those present in Soays. Addressing this question is of interest for several reasons. If novel mutations are present in Soay sheep, then this would represent evidence for independent evolution of a convergent phenotype. Conversely, if the self-colour pattern mutations are shared, then genetic variation in Soays may be ancestral or more recently derived because of interbreeding with modern domestic breeds; either way this is crucial to our understanding of the maintenance of the polymorphism. In this study, we combined an investigation of the *ASIP*-coding region with gene expression analyses, association studies and linkage mapping to elucidate the genetic basis of recessive self-colour pattern in Soay sheep. This knowledge will facilitate future studies on the role of selection in the maintenance of the polymorphism.

Materials and methods

Study population

The Soay sheep is a primitive feral breed that is thought to resemble the earliest domestic sheep that migrated into Europe during the Neolithic agricultural revolution (Clutton-Brock and Pemberton, 2004; Chessa *et al.*, 2009). The breed is endemic to the island of Soay in the St Kilda archipelago (Scotland, UK, 57°49'N, 8°34'W), but a free-living population was established on the neighbouring island of Hirta in 1932. The population in Village Bay, Hirta, has been studied on an individual basis since 1985. Each spring, at least 95% of lambs born in the 230 Ha study area have been caught, ear tagged (to enable individual identification), sampled for genetic analysis and phenotyped for coat pattern (Clutton-Brock and Pemberton, 2004).

Sequence diversity in the Soay sheep *ASIP* gene

We characterized sequence diversity in the *ASIP*-coding region in Soay sheep by amplifying and sequencing each of the three coding exons (2–4) in six wild and six self-type sheep. Genomic DNA was extracted from ear punches using the QIAamp DNA Mini kit (QIAGEN, Dusseldorf, Germany), following the standard protocol for animal tissues. A consensus sequence of the entire ovine *ASIP* gene was assembled from publicly available 454 sequence data using the CAP3 program (Huang and Madan, 1999). Using this reference sequence, PCR primers were designed using Primer3 (Rozen and Skaletsky, 2000) within intronic regions flanking exon 2 (*ASIP_F2*, *ASIP_R2*), exon 3 (*ASIP_F3*, *ASIP_R3*) and exon 4 (*ASIP_F4*, *ASIP_R4*; Supplementary Table S1). PCRs were carried out in 50- μ l reaction mixture using ~50 ng genomic DNA, 0.5 U Taq polymerase (Bioline, London, UK), and a final concentration of 1 \times PCR buffer (Bioline), 2.0 mM MgCl₂, 0.2 mM dNTPs and 0.5 mM of each primer. The PCR reagent mix for exon 4 included QIAGEN's Q-solution at a final concentration of 1 \times , to account for high GC base content in this region. PCR reactions were carried out on an MJ DNA Engine Dyad thermal cycler (MJ Research, Waltham, MA, USA) using a two-stage touchdown thermal cycling program. In the first stage, the annealing temperature (T_a) was decreased by 1 °C per cycle from 60 to 51 °C (a total of 10 cycles). In the second stage, T_a was set to 50 °C for an additional 25 cycles. In each case, one thermal cycle comprised 94 °C for 30 s, T_a for 30 s and 72 °C for 1 min.

Initial denaturation of genomic DNA was carried out at 94 °C for 3 min, and we included a final extension at 72 °C for 5 min. PCR products were gel-purified using the Wizard gel extraction kit (Promega, Madison, WI, USA) and then sequenced in both directions using standard protocols (BigDye Terminator mix v. 1.1, Applied Biosystems, Foster City, CA, USA). DNA sequence extension products were resolved by capillary electrophoresis on a 3730 instrument (Applied Biosystems). Chromatograms were visualized and edited using SEQSCAPE v2.1 (Applied Biosystems) and aligned in CLUSTAL X (Thompson *et al.*, 1997).

Association between coat pattern and *ASIP*

We quantified the association between coat pattern and *ASIP* by genotyping two putative functional mutations in 691 members of a complex multi-generational pedigree comprising 882 sheep. The pedigree used for this analysis and for the linkage mapping study (below) is described in Beraldi *et al.* (2006). Methods for isolation of genomic DNA have been described previously (Gratten *et al.*, 2007). Genotyping was carried out using the multiplex SNP Scoring by Colour And Length Exclusion (SNP-SCALE) method (Kenta *et al.*, 2008). Tailed allele-specific oligonucleotides (ASOs) and common reverse oligonucleotides for each assay were designed using Primer3 (Rozen and Skaletsky, 2000). Universal fluorescent oligonucleotides (UFOs), identical in sequence to the 5' tails attached to the ASOs, were selected from those previously described by Kenta *et al.* (2008). Multiplex PCR was carried out in 2- μ l reaction mixture containing ~20 ng DNA (dried), 1 μ l QIAGEN multiplex PCR master mix, 0.02 μ M of each ASO (APS077_ASO_N + T2, APS077_ASO_D + T4, APS078_ASO_A + T1, APS078_ASO_T + T2), 0.2 μ M of each CRO (APS077_CRO2, APS078_CRO), 0.2 μ M of UFO_T2_FAM and 0.1 μ M of UFO_T1_VIC and UFO_T4_NED (Supplementary Table S1). The PCR thermal cycling program was the same as that described above, with the following exceptions: genomic DNA was denatured initially for 15 min at 94 °C, the final extension at 72 °C was for 10 min, and each thermal cycle comprised 94 °C for 30 s, T_a for 90 s and 72 °C for 90 s. PCR products were visualized on a 3730 DNA Analyzer (Applied Biosystems) and sized relative to LIZ500 size standard (Applied Biosystems) using the GeneMapper 3.7 software package (Applied Biosystems).

We used PedCheck v1.1 (O'Connell and Weeks, 1998) to screen for Mendelian inconsistencies and either resolved or removed the identified mismatches. We tested for non-random association of coat pattern and genotype at each of the two *ASIP* mutations using Fisher's exact test. The assumption of independence between data points is not strictly adhered to in our analysis because individuals are related, but analyses based solely on the mapping panel founders ($N=44$) give qualitatively the same result. Calculations were carried out in R v2.8.0 (R Development Core Team). We inferred maternal and paternal two-locus haplotypes for each individual in the pedigree using default run parameters in Simwalk 2.91 (Sobel and Lange, 1996). Owing to the large size of the pedigree and the presence of inbreeding loops, we split the pedigree into 15 approximately independent sub-pedigrees (up to 200 individuals and three generations) using the CRIGEN function in a version of the CRIMAP v2.4 software (Green *et al.*, 1990) modified by Xuelu Liu (Monsanto Company, St Louis,

MO, USA). Simwalk input files and linkage disequilibrium (LD) estimates (D' , Δ^2 , Cramer's V) were obtained using the GOLD software package (Abecasis and Cookson, 2000).

Genetic linkage between coat pattern and *ASIP*

We added two putative functional *ASIP* mutations to chromosome 13 of the Soay sheep linkage map described in Beraldi *et al.* (2006), using Xuelu Liu's (Monsanto) version of CRIMAP v2.4 (Green *et al.*, 1990). We used the TWOPOINT command to detect linkage with existing chromosome 13 markers and the BUILD and FLIPS commands to identify optimal marker order, using the published chromosome 13 Soay sheep map as a starting point. Finally, we used the CHROMPIC command to identify and remove potentially spurious double or triple recombinants indicative of genotyping error.

Expression of *ASIP* in Soay sheep

We tested for *ASIP* expression in skin collected from wild ($N=6$) and self-type ($N=6$) Soay lambs at between 2 and 4 days of age. Isolation of total RNA and synthesis of cDNA was carried out as described in Gratten *et al.* (2007). Primers were designed to amplify an 82-bp cDNA product spanning exons 2 and 3 (qASIP_F1 and qASIP_R1, Supplementary Table S1). End-point reverse transcriptase-PCR was carried out in 10- μ l reaction mixture using ~200 ng cDNA, 0.5 U Taq polymerase (Bioline) and a final concentration of 1 \times PCR buffer (Bioline), 1 \times Q-solution (QIAGEN), 2.0 mM MgCl₂, 0.2 mM dNTPs and 0.5 mM of each primer. PCR reactions were carried out on an MJ DNA Engine using a two-stage touchdown thermal cycling program, identical to that described above for amplification of the coding exons. In each case, one thermal cycle comprised 94 °C for 30 s, T_a for 30 s and 72 °C for 30 s. PCR products were visualized by agarose gel electrophoresis.

Results

Sequence diversity in the Soay sheep *ASIP* gene

We determined the complete 402-bp coding sequence of the *ASIP* gene in six wild and six self-type Soay sheep and confirmed sequence identity using BLAST (GenBank accession no. FJ889180, FJ889181). Three polymorphisms were identified, all of which have been previously reported in domestic sheep (Smit *et al.*, 2002; Norris and Whan, 2008). A 5-bp deletion, g.100-104delAGGAA (hereafter referred to as D₅), was present in exon 2. This is a frame-shift mutation that introduces a premature stop codon at amino acid position 64, resulting in a mature peptide that lacks the functionally important cysteine-signalling domain. There were also two single nucleotide polymorphisms (SNPs) in exon 4, one of which (g.5172T \rightarrow A) is a non-synonymous mutation that is predicted to cause a cysteine to serine substitution at codon 126 within the signalling domain of the protein. All seven cysteine residues at the 3' terminus of the *ASIP* amino acid sequence are highly conserved across mammals (data not shown), suggesting strong functional constraint on loss-or-gain of cysteine mutations. Thus, D₅ and g.5172T \rightarrow A are both likely to influence the function of the *Agouti* protein, as has been previously proposed by Norris and Whan (2008). The second SNP (g.5051G \rightarrow C) is synonymous and would not be expected to cause a

functional change. All six self-type sheep were homozygous for a putatively non-functional haplotype comprising D₅, g.5051C and g.5172A, whereas wild-type sheep were either homozygous for a functional haplotype containing the non-deleted allele at g.100–104 (referred to as N₅), g.5051G and g.5172T ($N = 4$), or were heterozygous at all three positions ($N = 2$).

Association between coat pattern and *ASIP*

We obtained high-quality genotypes for D₅ and g.5172T → A in 650 and 624 sheep (out of a total of 691), respectively. Genotype was strongly associated with coat pattern at each mutation (Fisher's exact test: $P < 2.2e^{-16}$, for both D₅ and g.5172T → A; Table 1). We used pedigree information to infer unambiguous maternal and paternal haplotypes in a total of 621 sheep that were phenotyped for coat pattern (93.88% wild, 6.12% self; Table 2). The two putative loss of function mutations (D₅ and g.5172A) were almost always inherited together, as were the two functional alleles (N₅ and g.5172T; haplotype frequencies: $f(N_5T) = 0.7810$, $f(N_5A) = 0.0177$, $f(D_5T) = 0.0008$, $f(D_5A) = 0.2005$), suggesting strong LD between the two polymorphisms (LD based on 44 haplotyped founders: $D' = 1.000$, $\Delta^2 = 0.889$, Cramer's V = 0.943).

The majority of self-type sheep (28/38, 73.7%) were homozygous for the putatively non-functional D₅A haplotype, but a substantial proportion (6/38, 15.8%) had a N₅A/D₅A haplotype combination (referred to as a diplotype; Table 2). The remaining four self-type sheep had either an N₅T-N₅T or N₅T-D₅A diplotype, both of which would predict a wild phenotype. Sequencing of the entire *ASIP*-coding region in each of these individuals failed to identify any novel mutations, suggesting that a third recessive mutation may exist in a *cis*-regulatory region.

Most wild-type sheep were homozygous for N₅T (382/583, 65.5%), and a substantial proportion had a N₅T-D₅A diplotype (183/583, 31.4%). Of the remainder, a small proportion were heterozygous for the N₅A haplotype (N₅T-N₅A, 16/583, 2.7%), one sheep was found to carry the D₅T haplotype (N₅T-D₅T, 0.2%) and a single individual was homozygous for D₅A, which would predict a self phenotype. Four of the six diplotypes that would predict a self phenotype were not observed in any sheep (Table 2). This is unsurprising given the extent of LD between mutations and the size of the pedigreed population. In total, 99.2% of sheep (616/621) had phenotypes that were consistent with the morph predicted by their coding-region diplotype.

Genetic linkage between coat pattern and *ASIP*

The coat pattern locus in Soay sheep has been previously mapped to a position on chromosome 13 containing the microsatellite CTBJ12 (Beraldi *et al.* 2006). In this study, we added the *ASIP* gene to the chromosome 13 linkage map using genotype data for the D₅ and g.5172T → A coding mutations in the mapping pedigree (Figure 2). We found no evidence for recombination between CTBJ12 and either D₅ or g.5172T → A (two-point logarithm of odds (LOD) = 2.41 and 2.71, respectively), whereas there was evidence for recombination between *ASIP* and all other chromosome 13 markers ($N = 8$). *ASIP* therefore maps to the same location as the *Coat pattern* locus.

Table 1 Association between coat pattern and genotype at two putative functional mutations in the *ASIP* gene, a 5-bp deletion in exon 2 (g.100–104delAGGAA) and a non-synonymous SNP in exon 4 (g.5172T → A)

<i>ASIP</i> mutation	Genotype	Coat pattern		Total
		Wild	Self	
Exon 2, 5-bp deletion	N ₅ N ₅	412	2	414
	N ₅ D ₅	199	8	207
	D ₅ D ₅	1	28	29
	Total	612	38	650
Exon 4, T → A SNP	TT	375	2	377
	TA	210	2	212
	AA	1	34	35
	Total	586	38	624

Abbreviations: *ASIP*, *agouti signalling protein*; SNP, single nucleotide polymorphism.

The two alleles at the 5-bp deletion are denoted by N₅ (AGGAA) and D₅ (deletion). Genotype was strongly associated with coat pattern at each mutation (Fisher's exact test: $P < 2.2e^{-16}$ for both D₅ and g.5172T → A).

Table 2 Association between coat pattern and *ASIP*-coding region diplotype in Soay sheep

<i>ASIP</i> diplotype	Coat pattern		Total
	Wild	Self	
N ₅ T-N ₅ T	382	2	384
N ₅ T-N ₅ A	16	0	16
N ₅ T-D ₅ T	1	0	1
N ₅ T-D ₅ A	183	2	185
N ₅ A-D ₅ A	0	6	6
D ₅ A-D ₅ A	1	28	29
Total	583	38	621

Abbreviation: *ASIP*, *agouti signalling protein*.

The two putative functional mutations in *ASIP* (g.100–104delAGGAA and g.5172T → A) define four coding region haplotypes (N₅T, N₅A, D₅T and D₅A) that can be combined into 10 possible diplotypes. Six of these diplotypes were observed in the population and four were not identified in any sheep (N₅A-N₅A, N₅A-D₅T, D₅T-D₅T and D₅T-D₅A). Maternal and paternal haplotypes for each animal were inferred from genotypes within a complex multi-generational pedigree. Alleles for the exon 2 deletion are defined as in Table 1.

Expression of *ASIP* in Soay sheep

Gene expression studies were used to try to resolve the five apparent mismatches between phenotype and coding-region diplotype, in particular, whether there was evidence of a third *cis*-regulatory mutation affecting coat pattern. We could not directly quantify *ASIP* expression in skin tissue from the mismatching sheep as they are deceased. As an alternative, we carried out end-point reverse transcriptase-PCR of *ASIP* expression in skin tissue collected from wild ($N = 6$) and self-type ($N = 6$) lambs. This analysis revealed *ASIP* mRNA transcripts in all wild-type lambs, each of which carried one or more copies of an intact N₅T-coding region haplotype, whereas five of the six self-type lambs showed no evidence for *ASIP* transcription (Supplementary Figure S1). The self-type individual that was an exception had a N₅A-D₅A diplotype, whereas the remaining selfs were homozygous for D₅A. This suggests that haplotype D₅A, and in particular the 5-bp deletion in exon 2, is in LD with a *cis*-regulatory mutation that

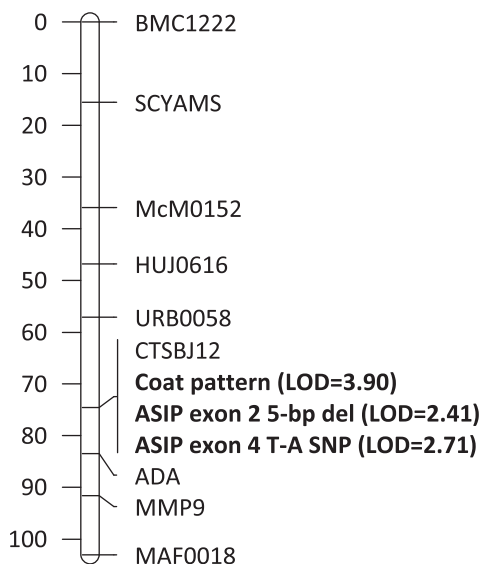


Figure 2 Genetic linkage map of Soay sheep chromosome 13, showing the location of *agouti signalling protein* (*ASIP*) functional mutations in exon 2 (g.100-104delAGGAA) and exon 4 (g.5172T→A) and the *Coat pattern* locus (latter from Beraldi *et al.* 2006). Two-point LOD scores for linkage with marker CTSBJ12 are provided in brackets. Map distances are in Kosambi centimorgans. SNP, single nucleotide polymorphism.

inactivates the promoter, whereas the N₅ allele is normally associated with an intact promoter. It follows that the four mismatching self-type individuals (that is, those with one or more intact N₅T-coding haplotypes) may be carrying recombinant N₅T haplotypes in which the promoter is silent. On the other hand, the presence of a *cis*-regulatory mutation cannot explain the single wild-type mismatch.

Discussion

We have previously shown that the *Coat pattern* locus in Soay sheep maps to a region on chromosome 13 that is likely to contain a single candidate gene, *ASIP* (Beraldi *et al.*, 2006). Here, our investigation of this gene in wild and self-type Soay sheep, including analyses of coding sequence variation and gene expression, revealed evidence for three recessive mutations, two functional changes in the coding sequence (a 5-bp deletion at position g.100-104 in exon 2 and a non-synonymous T→A SNP at position g.5172 in exon 4) and a third putative *cis*-regulatory mutation that inactivates the promoter. These mutations, all of which have been identified or inferred in domestic sheep (Smit *et al.*, 2002; Norris and Whan, 2008; Royo *et al.*, 2008), collectively explained coat pattern in all but one member of a complex multi-generational pedigree of 621 Soay sheep. We also confirmed that the *Coat pattern* locus and *ASIP* are tightly linked in Soay sheep by mapping the two putative coding-region mutations. Below, we describe a model for recessive self-colour pattern in Soay sheep, and consider points of similarity and difference between this model and that previously proposed for modern domestic breeds (Norris and Whan, 2008). We then speculate on the origin of the self phenotype in Soays, and consider the implications for future studies that aim

to investigate the role of selection in the maintenance of the polymorphism.

The genetic basis of recessive self-colour pattern in Soay sheep

We propose a model in which recessive self-colour pattern in Soay sheep can be caused by homozygosity of any of three separate mutations in the *ASIP* gene (Figure 3). These mutations can, in theory, define up to eight haplotypes, of which seven would be predicted to generate a self phenotype. We found evidence for five of these putative haplotypes in Soays, including four recessive haplotypes. The two coding-region mutations are in strong LD in the study population, and it is clear from the reverse transcriptase-PCR results that this is likely to extend to encompass the third (putative) *cis*-regulatory mutation in the promoter region. Subsequently, two haplotypes account for the majority of chromosomes in the population, one has an intact promoter and a functional N₅T-coding sequence (haplotype 1) and the other has a silent promoter and a non-functional D₅A-coding sequence (haplotype 2). Haplotype 1 results in a functional *agouti* protein, such that sheep with one or more copies of this haplotype have a wild-type coat pattern. This haplotype is common in the study population (estimated frequency = 0.770; see below) and is similar to that described in ancient Barbary sheep (Norris and Whan, 2008), suggesting that it is ancestral in Soays. Haplotype 2 is a recessive haplotype that accounts for almost three quarters of self phenotypes and ~20% of all chromosomes in the population.

In addition to these common haplotypes, we also found support for the existence of a number of rare haplotypes in which functional (N₅T) and non-functional (D₅A, D₅T and N₅A) coding sequences are associated with active and/or inactivated promoter regions; these novel haplotypes must be the product of recombination between haplotypes 1 and 2 (Figure 3). There was evidence, in the form of four self-type sheep with coding diplotypes that would predict a wild-type coat (N₅T-N₅T, N₅T-D₅A), that recombination between the putative *cis*-regulatory mutation and downstream coding exons has generated a novel haplotype in which a functional coding sequence (N₅T) is associated with an inactivated promoter (haplotype 3). Although we were unable to directly quantify *ASIP* expression in these individuals (because they are deceased), we are confident that these instances of self-colour pattern are due to the presence of a recombinant N₅T haplotype that is not expressed, as we found no evidence for novel coding mutations in any of the four sheep. Alternative explanations for the mismatches, including genotyping error, phenotyping error and sample mislabelling, can also be dismissed because results were confirmed by repeat genotyping, coat pattern was repeatedly verified for all four sheep during their lifetime and there were no mismatches between these individuals and their parents or offspring at 250 microsatellites used to construct the Soay sheep genetic linkage map (data not shown). On the basis of the observed proportion of animals that were heterozygous (2/185) for N₅T, but which had a self-type coat pattern, we estimate that 1.08% (95% confidence interval, 0.30–3.86%) of N₅T chromosomes in Soays have an inactivated promoter region. Thus, the estimated

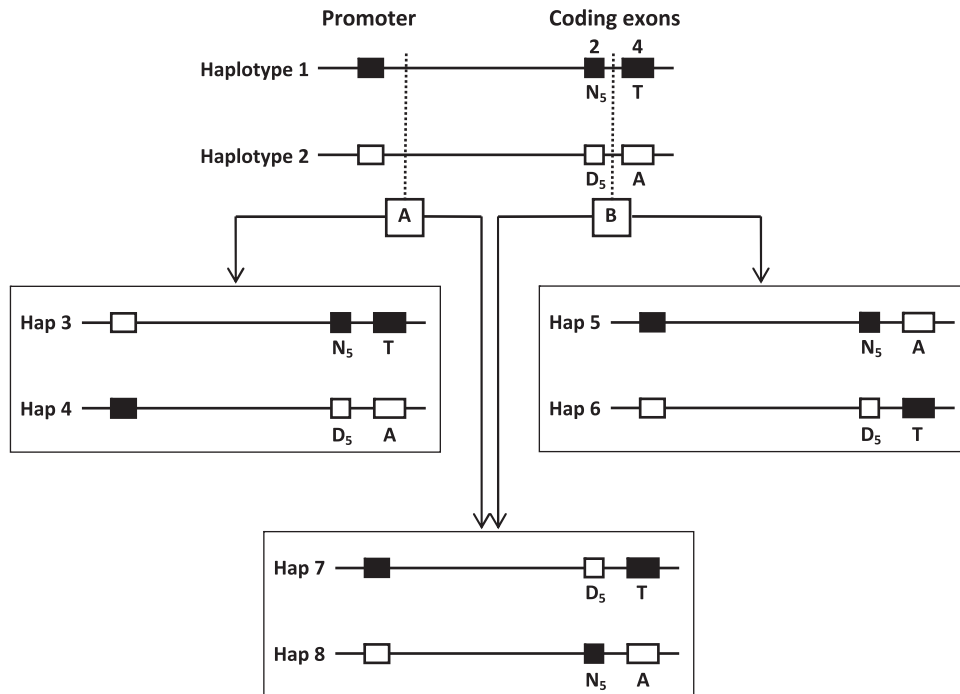


Figure 3 Schematic of a putative model for the role of *agouti signalling protein* (*ASIP*) in recessive self-colour pattern in Soay sheep. Functional coding and promoter sequences are represented by solid black boxes, and putative non-functional coding sequences and inactivated promoter regions by open boxes. The two alleles at the 5-bp deletion in exon 2 (g.100-104delAGGAA) are denoted by N_5 (AGGAA) and D_5 (deletion). T and A are the two alleles at the non-synonymous single nucleotide polymorphism in exon 4 (g.5172T→A). Exon 3 is not shown because no functional mutations are present. Haplotype 1 is the ancestral wild-type haplotype in Soay sheep, with intact promoter and functional N_5T -coding sequence. Haplotype 2 is a recessive self haplotype with inactivated promoter and putative non-functional D_5A -coding sequence, similar to that in domestic sheep (Norris and Whan 2008), and likely to resemble the first self haplotype to exist in Soays. Recombination at position A, between the promoter region and coding exons, would generate two novel haplotypes, one in which a functional N_5T -coding sequence is not expressed (haplotype 3), and another in which a non-functional D_5A -coding sequence is associated with an intact promoter (haplotype 4). Recombination at position B, between the putative functional mutations in coding exons 2 and 4, would also generate two novel haplotypes, one in which an intact promoter is associated with a recombinant N_5A -coding sequence (haplotype 5) and another in which an inactivated promoter is associated with a recombinant D_5T -coding sequence (haplotype 6). Recombination at both A and B is unlikely, but would generate two additional haplotypes, one in which the D_5T -coding sequence is expressed (haplotype 7) and another in which the N_5A haplotype is not expressed (haplotype 8). Haplotypes 2–8 are all predicted to be non-functional, and thus haplotype 1 is the only functional haplotype. Not drawn to scale.

frequency of the expressed N_5T haplotype in the population is 0.770 (0.781–0.011).

Recombination between the putative *cis*-regulatory mutation and coding exons would also generate another novel haplotype in which the putatively non-functional D_5A -coding sequence is associated with an active promoter region (haplotype 4). We found no direct evidence for this haplotype, as expressed and inactivated D_5A -coding sequences would be expected to generate the same (that is, self) coat pattern. However, unless this putative haplotype has been lost due to drift or selection, it is likely that the D_5 frame-shift mutation is expressed in Soay sheep.

There was also evidence for recombination in Soay sheep between the two putative functional mutations in exons 2 and 4, generating two additional haplotypes, one in which a recombinant N_5A -coding sequence is expressed (haplotype 5), and another in which a D_5T -coding sequence is associated with an inactivated promoter (haplotype 6). The reverse transcriptase-PCR results provided direct evidence for an expressed N_5A haplotype in Soay sheep, as *ASIP* mRNA transcripts were detected in the single self-type individual with an N_5A - D_5A diplotype. This finding also shows that the g.5172A mutation does cause a loss of function in the

agouti protein, because if it did not this individual would have had a wild-type pigmentation pattern. In total, six sheep were identified with an N_5A - D_5A diplotype and all were self. It is possible that some of these sheep carry an inactivated N_5A haplotype, but this explanation is less parsimonious because such a haplotype could only arise given more than one recombination event (Figure 3). One individual was identified with coding haplotype D_5T , which is likely to represent haplotype 6 with silent promoter.

In summary, we found evidence for four of the seven putative recessive haplotypes (2, 3, 5 and 6), and one other (haplotype 4) may be present at low frequency. The remaining two recessive haplotypes (7 and 8) would only exist if there have been multiple recombination events within *ASIP* (Figure 3), which is unlikely (although they may still arise because of future recombination events). A single wild-type sheep with a D_5A - D_5A -coding region diplotype was the only individual whose phenotype was not explained by our model. This mismatch may be due to sampling error, although we cannot exclude the possibility of a rare mutation at another locus that antagonizes *MC1R* in the absence of a functional *agouti* protein. Haplotype 2, and in particular the putative *cis*-regulatory mutation, explains most instances of the self-

colour pattern phenotype, but several other recessive haplotypes, including haplotype 5 in which the g.5172T→A loss of cysteine mutation is expressed, contribute to at least 15% of all self phenotypes (Table 2). Coat pattern is therefore genetically heterogeneous in free-living Soay sheep, as three independent mutations at the same genetic locus can all determine the self phenotype.

Same genotype-different phenotype

One interesting consequence of the self phenotype being recessive and the D₅ and g.5172T→A mutations both being functionally important is that doubly heterozygous individuals could, in theory, have either phenotype depending on allelic phase. The majority of double heterozygotes would be expected to have a wild-type coat pattern and carry the N₅T-D₅A diplotype, as these two-coding region haplotypes account for more than 97% of all chromosomes; 183 out of a total of 185 double heterozygotes fulfilled this expectation. However, there are two ways in which doubly heterozygous sheep may express a recessive self phenotype. The first is the case in which the N₅T haplotype has an inactivated promoter, such that the functional coding sequence is not expressed (Figure 3); we identified two self-type sheep with this putative diplotype. Alternatively, individuals carrying one copy of each of the recombinant haplotypes, N₅A and D₅T, would also be self because both result in loss of function at *ASIP* (irrespective of the third *cis*-regulatory mutation). It is unsurprising that we failed to observe any self-type sheep in this diplotype category, because LD between the two sites is very extensive.

Points of similarity and difference between Soay sheep and modern domestic breeds

There are several important distinctions between models for colour pattern variation in Soay sheep and modern domestic breeds (Smit *et al.*, 2002; Norris and Whan, 2008; Royo *et al.*, 2008). Perhaps the most striking is that the wild-type *agouti* allele, which is rare in domestic sheep but typical of their wild ancestors (Sponenberg, 1997), is common in Soays. This is consistent with the fact that Soay sheep are one of a small group of surviving primitive breeds that migrated into Europe earlier than the immediate predecessors of most modern breeds (Chessa *et al.*, 2009). Another difference is the absence in Soays of the dominant black (Vage *et al.*, 1999) and dominant white (Norris and Whan, 2008) phenotypes described in domestic sheep, and in particular of the large genomic duplication of *ASIP* that is responsible for deregulated expression of *ASIP* resulting in dominant white colour. A final difference is that the *TYRP1* mutation underlying dark versus light brown coat colour in Soays is yet to be identified in any domestic breed (although to our knowledge few, if any domestic breeds have been surveyed at this locus).

In spite of these distinctions there is much in common between Soays and domestic sheep in relation to the genetic basis of recessive self-colour pattern. Both of the putative functional mutations in the Soay sheep *ASIP*-coding region (D₅ and g.5172T→A), as well as a third synonymous exon 4 SNP (g.5051G→C), have been described in modern domestic breeds (Norris and Whan, 2008). A *cis*-regulatory mutation is also common

to Soays and a number of different domestic breeds (Norris and Whan, 2008; Royo *et al.*, 2008), although we cannot confirm the homology of this change in the absence of a molecular genetic understanding of the ovine *ASIP* promoter. Interestingly, not all mutations that have been identified in domestic sheep are segregating in Soays, including a 9-bp deletion in *ASIP* exon 2 that was described in Romanov sheep by Norris and Whan (2008).

The origin of self-colour pattern in Soay sheep

It is unlikely that the *ASIP* mutations shared between Soay sheep and modern domestic breeds have arisen independently in each lineage, and thus the diversity is either ancestral, or has been introduced to or from Soays by interbreeding with modern domestic sheep. We propose that the self-mutations originate in domestic sheep, for several reasons. First, Norris and Whan (2008) proposed that recessive self-colour pattern allele(s), which are single copy, were generated by 'nonallelic pairing and recombination and/or gene conversion events' following the genomic duplication responsible for the dominant white phenotype that is present in domestic sheep. As this duplication does not exist in Soays, and the self allele arose after the duplication event, it seems unlikely that the self allele was present in the ancestral Soay population.

Second, the two functional mutations are in strong LD in Soays. This is indicative of a recent and simultaneous origin in the population, and inconsistent with the idea that self haplotypes in Soays are the predecessors of those identified in modern breeds, although we cannot yet rule out alternative explanations, including a past selective sweep, or epistasis favouring particular combinations of alleles. Future studies that aim to quantify the relationship between *ASIP* diplotype and individual fitness in the study population should help to address this question.

Third, there is historical and genetic evidence for past admixture between Soay sheep and domestic sheep. The historical evidence comes in the form of anecdotal reports of the introduction of the now extinct Scottish short wool or Dunface sheep on to the Island of Soay before 1930 (Elwes, 1912). The genetic evidence comes in the form of extensive LD around the coat colour gene *TYRP1* on chromosome 2 (Gratten *et al.*, 2008), and in the vicinity of the *interferon-gamma* (*IFN(γ)*) locus on chromosome 3 (J Gratten *et al.*, unpublished data). In combination with this study, this suggests that extended LD may be a common feature of the Soay sheep genome. Such a pattern is likely to have arisen from demographic events, such as admixture or drift, although we cannot exclude the possibility of a series of local selective sweeps (Nordborg and Tavaré, 2002).

Finally, Soays are among the most primitive surviving domestic sheep, lacking many of the improved production traits found in modern breeds, such as retention of fleece (Clutton-Brock and Pemberton, 2004; Chessa *et al.*, 2009). This suggests that the ancestors of contemporary Soay sheep did not experience (prolonged) human selection for products such as wool that are secondary to meat, which was the original focus of human domestication (Chessa *et al.*, 2009). As the accumulation of colour variants in domestic sheep is likely to have coincided with strong artificial selection for various wool

traits, it is plausible that Soays diverged from the predecessors of modern breeds before the emergence of much of the colour variation present today.

These observations collectively suggest that a recessive self haplotype, similar to haplotype 2, was first introduced into the Soay population as a consequence of admixture with modern domestic sheep before 1930. Confirmation of this hypothesis will require examination of genome-wide patterns of LD in Soays, in addition to genome surveys for long-range haplotypes in potential source breeds that are shared with Soays.

Summary

In this study we show that recessive self-colour pattern in a free-living population of Soay sheep, a trait previously thought to exhibit single-locus Mendelian inheritance based on analysis of phenotype data (Doney *et al.*, 1974; Clutton-Brock and Pemberton, 2004), is genetically heterogeneous. The three identified recessive mutations are likely to have originated in domestic sheep, although further analysis will be required to confirm this hypothesis. This is only the second study in which *ASIP* has been linked to variation in pigmentation in a natural population (Mundy, 2007; Steiner *et al.*, 2007), and is among the first studies in which genetic heterogeneity for pigmentation has been fully characterized in the wild (Nachman *et al.*, 2003). This knowledge adds to our general understanding of the diversification of sheep during domestication. More specifically, an understanding of the genetic basis of recessive self-colour pattern in Soay sheep, and the recognition that several mutations are segregating in the population, will enable the differentiation of phenotypically identical individuals with different underlying genotypes (for example, homozygous and heterozygous wild-type sheep). This knowledge and the understanding that the self phenotype is likely to have a recent origin in the population will enhance future studies that investigate the role of selection in the maintenance of the polymorphism, not least because genetic effects on fitness can be masked at the phenotypic level, for instance by dominance (Gratten *et al.*, 2008).

Acknowledgements

We thank the many Soay sheep project members and volunteers who have collected phenotypic data and genetic samples during the course of the long-term study on St Kilda. We thank the National Trust of Scotland for granting permission to work on St Kilda, and QinetiQ for logistical support. The long-term data collection on St Kilda has been supported by grants from the Natural Environment Research Council (NERC) and the Wellcome Trust to TH Clutton-Brock, BT Grenfell, LEB Kruuk, MJ Crawley and JMP. This study was funded by NERC through its Environmental Genomics thematic program (grant no. NER/T/S/2002/00189). Alastair Wilson, Jay Storz and one anonymous reviewer made constructive comments that improved the paper.

References

Abecasis GR, Cookson WOC (2000). GOLD—graphical overview of linkage disequilibrium. *Bioinformatics* **16**: 182–183.

- Anderson TM, vonHoldt BM, Candille SI, Musiani M, Greco C, Stahler DR *et al.* (2009). Molecular and evolutionary history of melanism in North American gray wolves. *Science* **323**: 1339–1343.
- Bennett DC, Lamoreux ML (2003). The color loci of mice—a genetic century. *Pigment Cell Res* **16**: 333–344.
- Beraldi D, McRae AF, Gratten J, Slate J, Visscher PM, Pemberton JM (2006). Development of a linkage map and mapping of phenotypic polymorphisms in a free-living population of Soay sheep (*Ovis aries*). *Genetics* **173**: 1521–1537.
- Bultman SJ, Michaud EJ, Woychik RP (1992). Molecular characterization of the mouse agouti locus. *Cell* **71**: 1195–1204.
- Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, Mainland I *et al.* (2009). Revealing the history of sheep domestication using retrovirus integrations. *Science* **324**: 532–536.
- Clutton-Brock TH, Pemberton JM (2004). *Soay Sheep: Dynamics and Selection in an Island Population*. Cambridge University Press: Cambridge.
- Coltman DW, Pemberton JM (2004). Inheritance of coat colour and horn type in Hirta Soay sheep. In: Clutton-Brock TH, Pemberton JM (eds). *Soay Sheep: Dynamics and Selection in an Island Population*. Cambridge University Press: Cambridge, pp 321–327.
- Doney JM, Ryder ML, Gunn RG, Grubb P (1974). Colour, conformation, affinities, fleece, patterns of inheritance of the Soay sheep. In: Jewell PA, Milner C and Boyd JM (eds). *Island Survivors: The Ecology of the Soay Sheep of St Kilda*. Athlone Press: London, pp 88–125.
- Drogemuller C, Giese A, Martins-Wess F, Wiedemann S, Andersson L, Brenig B *et al.* (2006). The mutation causing the black-and-tan pigmentation phenotype of Mangalitzta pigs maps to the porcine *ASIP* locus but does not affect its coding sequence. *Mamm Genome* **17**: 58–66.
- Ellegren H, Sheldon BC (2008). Genetic basis of fitness differences in natural populations. *Nature* **452**: 169–175.
- Elwes HJ (1912). Notes on the primitive breeds of sheep in Scotland. *The Scottish Naturalist* **1**: 25–32.
- Fang M, Larson G, Soares Ribeiro H, Li N, Andersson L (2009). Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *PLoS Genet* **5**: e1000341.
- Feder ME, Mitchell-Olds T (2003). Evolutionary and ecological functional genomics. *Nature Genet* **4**: 649–655.
- Girardot M, Guibert S, Laforet MP, Gallard Y, Larroque H, Oulmouden A (2006). The insertion of a full-length *Bos taurus* LINE element is responsible for a transcriptional deregulation of the Normande Agouti gene. *Pigment Cell Res* **19**: 346–355.
- Gratten J, Beraldi D, Lowder BV, McRae AF, Visscher PM, Pemberton JM *et al.* (2007). Compelling evidence that a single nucleotide substitution in *TYRP1* is responsible for coat-colour polymorphism in a free-living population of Soay sheep. *Proc R Soc B Biol Sci* **274**: 619–626.
- Gratten J, Wilson AJ, McRae AF, Beraldi D, Visscher PM, Pemberton JM *et al.* (2008). A localized negative genetic correlation constrains microevolution of coat color in wild sheep. *Science* **319**: 318–320.
- Green P, Falls K, Crooks S (1990): *Documentation for Cri-Map*. Washington University: St Louis.
- Hanski I, Saccheri I (2006). Molecular-level variation affects population growth in a butterfly metapopulation. *PLoS Biol* **4**: e129.
- Hoekstra HE (2006). Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* **97**: 222–234.
- Huang XQ, Madan A (1999). CAP3: a DNA sequence assembly program. *Genome Res* **9**: 868–877.
- Jackson IJ (1994). Molecular and developmental genetics of mouse coat colour. *Annu Rev Genet* **28**: 187–217.
- Kenta T, Gratten J, Haigh NS, Hinten GN, Slate J, Butlin RK *et al.* (2008). Multiplex SNP-SCALE: a cost-effective medium-

- throughput single nucleotide polymorphism genotyping method. *Mol Ecol Resour* **8**: 1230–1238.
- Kerns JA, Newton J, Berryere TG, Rubin EM, Cheng JF, Schmutz SM *et al.* (2004). Characterization of the dog Agouti gene and a nonagouti mutation in German Shepherd Dogs. *Mamm Genome* **15**: 798–808.
- Kijas JMH, Wales R, Tornsten A, Chardon P, Moller M, Andersson L (1998). Melanocortin receptor 1 (MC1R) mutations and coat color in pigs. *Genetics* **150**: 1177–1185.
- Klungland H, Vage DI, Gomez-Raya L, Adalsteinsson S, Lien S (1995). The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. *Mamm Genome* **6**: 636–639.
- Ludwig A, Pruvost M, Reissmann M, Benecke N, Brockmann GA, Castanos P *et al.* (2009). Coat color variation at the beginning of horse domestication. *Science* **324**: 485.
- Mundy NI (2007). Coloration and the genetics of adaptation. *PLoS Biol* **5**: 1842–1843.
- Nachman MW, Hoekstra HE, D'Agostino SL (2003). The genetic basis of adaptive melanism in pocket mice. *Proc Natl Acad Sci USA* **100**: 5268–5273.
- Newton JM, Wilkie AL, He L, Jordan SA, Metallinos DL, Holmes NG *et al.* (2000). Melanocortin 1 receptor variation in the domestic dog. *Mamm Genome* **11**: 24–30.
- Nordborg M, Tavaré S (2002). Linkage disequilibrium: what history has to tell us. *Trends Genet* **18**: 83–90.
- Norris BJ, Whan VA (2008). A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Res* **18**: 1282–1293.
- O'Connell JR, Weeks DE (1998). PedCheck: a program for the identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* **63**: 259–266.
- Rieder S, Taourit S, Mariat D, Langlois B, Guerin G (2001). Mutations in the agouti (ASIP), the extension (MC1R), and the brown (TYRP1) loci and their association to coat color phenotypes in horses (*Equus caballus*). *Mamm Genome* **12**: 450–455.
- Ritland K, Newton C, Marshall HD (2001). Inheritance and population structure of the white-phased 'Kermode' black bear. *Curr Biol* **11**: 1468–1472.
- Royo LJ, Alvarez I, Arranz JJ, Fernandez I, Rodriguez A, Perez-Pardal L *et al.* (2008). Differences in the expression of the ASIP gene are involved in the recessive black coat colour pattern in sheep: evidence from the rare Xalda sheep breed. *Anim Genet* **39**: 290–293.
- Rozen S, Skaletsky HJ (2000). Primer 3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds). *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press: Totowa, NJ, pp 365–386.
- Smit MA, Shay TL, Beever JE, Notter DR, Cockett NE (2002). Identification of an agouti-like locus in sheep. *Anim Genet* **33**: 383–385.
- Sobel E, Lange K (1996). Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker sharing statistics. *Am J Hum Genet* **58**: 1323–1337.
- Sponenberg DP (1997). Genetics of colour and hair texture. In: Piper L, Ruvinsky A (eds). *The Genetics of Sheep*. CAB International: Wallingford, pp 51–85.
- Steiner CC, Weber JN, Hoekstra HE (2007). Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol* **5**: 1880–1889.
- Stinchcombe JR, Hoekstra HE (2008). Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**: 158–170.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **24**: 4876–4882.
- Vage DI, Klungland H, Lu D, Cone RD (1999). Molecular and pharmacological characterization of dominant black coat color in sheep. *Mamm Genome* **10**: 39–43.
- Vage DI, Lu DS, Klungland H, Lien S, Adalsteinsson S, Cone RD (1997). A non-epistatic interaction of agouti and extension in the fox, *Vulpes vulpes*. *Nature Genet* **15**: 311–315.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)