American Genetic Association

Journal of Heredity, 2020, 182–193
doi:10.1093/jhered/esz070
Original Article
Advance Access publication November 12, 2019

OXFORD

Original Article

MC1R and *KIT* Haplotypes Associate With Pigmentation Phenotypes of North American Yak (*Bos grunniens*)

Jessica L. Petersen, Theodore S. Kalbfleisch, Morgan Parris, Shauna M. Tietze, and Jenifer Cruickshank

From the Department of Animal Science, University of Nebraska-Lincoln, 3940 Fair St, ANSC A218g, Lincoln, NE 68583-0908 (Petersen, Parris, Tietze); Department of Veterinary Science, Gluck Equine Research Center, University of Kentucky, Lexington, KY 40546-0099 (Kalbfleisch); and Department of Animal and Rangeland Sciences, Oregon State University, Corvallis, OR 97330 (Cruickshank).

Address correspondence to J. L. Petersen at the address above, or e-mail: jessica.petersen@unl.edu.

Received March 27, 2019; First decision July 26, 2019; Accepted November 8, 2019.

Corresponding Editor: Eduardo Eizirik

Abstract

Small numbers of domestic yak (*Bos grunniens*) were imported to North America in the late 19th century indirectly from the Qinghai-Tibetan Plateau. Coat color of yak is of interest for fiber production, aesthetics, and as a potential indicator of recent hybridization with cattle. North American yak are classified into 3 major coat color patterns depending upon the presence and extent of white markings. They are further classified by nose pigmentation (black or gray). The aim of this study was to identify loci involved in white patterning and nose pigmentation of North American yak. Genotyping by mass spectrometry of markers identified through Sanger and whole-genome sequencing revealed a 388 kb haplotype of *KIT* associated in a semi-dominant manner with white coloration in this population of yak. This *KIT* haplotype is similar to both a haplotype found in white-faced Chinese yak and to haplotypes found in cattle but is divergent from other *Bos* species such as bison, gaur, and banteng. Melanocortin 1 receptor (*MC1R*) was implicated as a dominant determinant of black nose color with a single haplotype containing 2 missense mutations perfectly associated with the phenotype. The *MC1R* haplotype associated with black nose pigment is also similar to cattle haplotypes. No cattle studied, however, shared either of the 2 haplotypes associated with color in yak, suggesting these alleles were introgressed into yak before they were imported to North America. These results provide molecular insight into the history of North American yak and information from which breeders can determine possible color outcomes of matings.

Subject areas: Genomics and gene mapping, Molecular systematics and phylogenetics
Keywords: eumelanin, hybridization, introgression, *Poephagus grunniens*, white patterning

Near the turn of the 20th century, small numbers of domestic yak were imported into North America (White et al. 1946; Wiener et al. 2003) indirectly from the Qinghai-Tibetan Plateau (modern-day China, India, and Nepal), the region in which they were domesticated (Qiu et al. 2015). While a few records of yak housed in zoos exist in the literature (Mann 1930; Society 1931), the earliest documented reports of yak being kept as livestock in North America are associated with experimentation by both the Canadian and US governments in

crossing them with bison and domestic cattle (White et al. 1946). In the United States, the result of hybridization between cattle and yak is most readily observed in coat color variation, which often can be directly attributed to *PMEL* dilution alleles from Charolais (Kuhn and Weikard 2007) or Highland (Schmutz and Dreger 2013) cattle. While hybridization resulting in inclusion of these dilution alleles is relatively recent in the North American animals, there is significant evidence that historic introgression among *Bos* spp., including that between cattle and yak, has played a notable role in the evolution of the genus (Wu et al. 2018). Wu et al. (2018), suggested that the introgression of yak alleles into cattle benefited their environmental adaptability, whereas variation from cattle at loci implicated in docility and coat color were incorporated into the yak genome. Beyond being a visual indicator of potential hybridization, coat color is of importance aesthetically as well as for fiber production.

In Asian yak, coat color varies by breed and location (Zhang et al. 2014b; Qiu et al. 2015); however, with the exception of color variation attributed to recent hybridization with domestic cattle, North American yak only appear in 3 coat color patterns. Solid black animals are considered wild type. Yak with white only on the forehead and/or on the feet and tail tip are termed "trim," and those with significant white patterning are designated "royal" (Figure 1). Royal yak have a large blaze that extends from between the ears down the center of the face to the nose, 4 white legs and feet, and significant white markings on the rear half of the torso. Contrary to the markings of the rear, the front of a royal yak's torso is mostly pigmented, often forming a saddle covering the hump, extending to the sides of the neck and face. The color phenotype of North American yak is further classified by the pigment around the nose. Individuals either have a gray nose ("native") or a darkly pigmented, black nose ("imperial"; Figure 1).

Variations in the gene KIT proto-oncogene receptor tyrosine kinase (*KIT*) have been associated with white patterning across species (Giuffra et al. 2002; Brooks et al. 2007; Haase et al. 2009; David et al. 2014; Yan et al. 2014; Durig et al. 2017), including the Hereford pattern in cattle, which presents as a solid white face with white under the belly and on the lower legs and the tail (Grosz and MacNeil 1999; Fontanesi et al. 2010). These *KIT* variants range from single nucleotide polymorphisms (SNPs; Yan et al. 2014) and deletions (Durig et al. 2017) to structural variants such as a duplication 5′ of the gene associated with the Hereford pattern (Whitacre 2014) and 2 translocations resulting in color sidedness, or the "witrik" pattern (Durkin et al. 2012). Alternative to white patterning, the generation of dark (eumelanistic) pigmentation in many species is attributed to the function of melanocortin 1 receptor (*MC1R*); the disruption of this receptor can result in the production of red phaeomelanin as seen in horses (Marklund et al. 1996), dogs (Everts et al. 2000; Newton et al. 2000; Schmutz et al. 2003), and pigs (Kijas et al. 1998). In cattle, variation in *MC1R* plays a role in the determination of the base coat color with black pigment dominant to red (Klungland et al. 1995; Joerg et al. 1996).

The genetic determination of coat color has previously been investigated in Asian yak (Chen et al. 2009; Zhang et al. 2014b) and in yak × cattle hybrids (Xi et al. 2012b). Zhang et al. (2014b) associated variation in *PMEL* and *MC1R* with a brown coat, while

Figure 1. (Top) An imperial trim (front left) and native trim (front right) yak with varying degrees of white markings on the forehead, tail, and hind legs. (Bottom) A native, solid black cow with a royal cow and calf.

a haplotype in *KIT* was associated with a white-face phenotype; however, neither the brown coat color nor the white-face phenotype is common in North American yak. Further, Chen et al. (2009) found no role of *MC1R* in black versus white body coloration of Chinese yak. As the primary color patterns in North American yak are unique relative to those previously studied, the purpose of this investigation was to identify the loci involved in white patterning and in nose pigmentation of these yak. To address this goal, we utilized a custom genotyping assay informed by Sanger and whole-genome sequence of a subset of animals for statistical tests of association between variation in candidate genes and the animals' coat and nose color phenotypes. In addition to the potential to provide breeders a tool with which to identify matings to achieve particular coat color phenotypes, these data provide information on the origin of these heritable characteristics in the North American yak.

Materials and Methods

Samples

Samples of 180 yak were obtained from US breeders in the form of pulled hair, whole blood (EDTA), nasal or buccal swabs. All 3 base coat colors (black [n = 66], trim [n = 60], royal [n = 54]) were represented as well as both variations (native [n = 70] and imperial [n = 54]) of nose color (n = 56 unknown). Pedigree data were recorded when available as provided by the owners and/or through the International Yak Association (IYAK) registry. Yak registered in the United States are genotyped both to confirm parentage and assess the level of cattle introgression (Neogen GeneSeek, Lincoln, NE; Kalbfleisch, unpublished). Those data, available for 143 of the yak, report a mean estimate of cattle introgression of 0.38% with a maximum of 2.1%. In addition, DNA from 3 hybrids (*Bos grunniens* × *Bos taurus*) and 8 domestic cattle (4 Hereford, 4 Holstein) were included. Controls for the examination of color sidedness included DNA from semen of 2 Holstein bulls, one with the witrik (lineback/color-sided) pattern.

DNA Isolation

DNA was isolated from blood samples using the Gentra Puregene Blood Core Kit B (Qiagen) following the whole blood protocol. The same kit was utilized for DNA isolation from 8 to 10 hair roots cut 0.5 cm above the follicle. Those hair samples were lysed in 300 µL of Cell Lysis Buffer and 20 µL Proteinase K (20 mg/mL) at 55 °C. After cell lysis, 100 µL of Protein Precipitation Solution was added to the supernatant and the samples incubated on ice for 5 min before centrifuging (15 000 × g at 15 °C for 7 min). After pipetting supernatant into a new tube, 650 µL of isopropanol was added and incubated for 10 min at room temperature, followed by 10 min on ice. After a 15 min centrifugation step (15 000 × g at 4 °C), the supernatant was discarded. The remaining pellet was washed with 70% ethanol. After another 3-min centrifugation step (15 000 × g), the samples were allowed to dry for 5 min. The pellet was hydrated in 20 µL of DNA Hydration Solution (Qiagen) and incubated overnight at room temperature. DNA was isolated from nasal and buccal swabs using the QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's directions, and DNA was isolated from semen as described by Cruickshank et al. (2004). All isolated DNA was stored at –20 °C until use.

PCR and Sanger Sequencing—*KIT*

The 21 exons of *KIT* (ENSBTAT000000003498) were amplified via PCR utilizing 18 primer pairs (Supplementary Table 1), in 6 yak representing the 3 coat color patterns (2 each of black, trim, royal). Amplification was performed in a reaction consisting of 30.25 mM MgCl, 0.75 µL each primer (20µM), 0.5 µL dNTP (10mM), 1X buffer, 0.1 µL Faststart Taq (Roche), and 20 ng of DNA template with nuclease-free water added to a final volume of 12 µL. Thermocycling conditions included 4 min at 94° C for initial denaturation followed by 30 cycles of denaturation at 94° C for 30 s, 30 s at the primer-pair specific annealing temperature (Supplementary Table 1), and 72° C for 45 s; products were held at 72° C for 10 min for the final extension. PCR products were checked for amplification by electrophoresis on 1.2% agarose gels stained with GelRed DNA Stain (Biotium), and visualized on a GelDoc imaging system (BioRad). PCR products were prepared for sequencing using 0.75 µL Exosap-It (Affymetrix Inc, Santa Clara, CA) per 4 µL PCR product. The cleaned product was sent to the University of Nebraska Medical Center Genomics Core Facility (Omaha, NE) for Sanger sequencing in both directions. Variants were identified using Sequencher ver5.4.6 (Gene Codes Corp, Ann Arbor, MI). The exon containing variation segregating with coat color pattern (Exon 3; Chr6:71,871,914-71,872,195) was then sequenced in an additional 39 animals (17 black, 9 trim, 13 royal).

Translocation alleles involving the *KIT* locus that were previously associated with color sidedness in both cattle and yak were investigated to determine if they were associated with the trim and royal coat color patterns. Toward this goal, 3 PCR primer pairs specific to the junctions of the inserted chromosome 6 segments (Durkin et al. 2012; Supplementary Table 2) were used to identify the presence or absence of the translocation alleles. Amplification was performed in 10-µL reactions containing 1 µL (1X) buffer, 0.4 µL (2 mM) MgCl2, 0.08 µL (0.8 mM) dNTP, 0.2 µL (0.2 mM) each primer, 0.1 µL (0.05 U) Biolase Taq polymerase (Bioline), 0.4 µL (2 ng/ µL) DNA template, and 7.62 µL nuclease-free water. Thermocycling conditions were an initial denaturation at 94 °C for 3 min, then cycled 9 times in the following order: 94 °C for 30 s, 59 °C for 30 s with a decrease of 1 °C per cycle, 72 °C for 30 s. After the initial 9 cycles, 30 cycles with an annealing temperature of 53 °C were completed, finishing with a final extension step at 72 °C for 10 min. PCR products were visualized on 2% agarose gels stained with SYBR Safe (ThermoFisher) on a GelDoc imaging system (BioRad). The presence/absence of the translocation alleles was determined by the banding observed on the gel as described in Supplementary Figure 1 and Supplementary Table 2.

Whole-Genome Sequencing

Whole-genome sequence (WGS) of Queen Allante, a native trim female yak, mapped to bovine assembly UMD3.1 was available from prior work (Heaton et al. 2016). In addition, whole-genome sequence was available from a Chinese yak (Qiu et al. 2012) and, for this study, was generated from DNA isolated from blood from 2 additional yak: a royal cow and a solid imperial bull. Library preparation and 150 bp, paired-end sequencing of these 2 yaks was performed at the University at Buffalo Next-Generation Sequencing and Expression Analysis Core on an Illumina HiSeq 2500 platform. Sequence data from those 3 animals were mapped to the UMD3.1 assembly and single nucleotide variants along with short insertion and deletions were identified as previously described (Kalbfleisch and Heaton 2013). The predicted severity of the candidate nonsynonymous mutations was determined in SNPEff (Cingolani et al. 2012), and PROVEAN (Choi 2012; Choi et al. 2012). All genomic coordinates are reported according to the UMD3.1 genome build with the corresponding location in ARS-UCD1.2 identified using NCBI Genome Remapping Service (https://www.ncbi.nlm.nih.gov/genome/tools/remap).

Targeted Genotyping

Fifty-eight variants segregating in *KIT* identified by Sanger and/or NGS data were included in the design of a matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry (MALDI-TOF MS) genotyping assay. Thirty-seven additional markers in and around 7 candidate genes for coat color (*ASIP, CBD103, KITL, MC1R, MSH, PMEL*, and *TYRP1*), identified by WGS data from the 4 yak were also included, as were the 2 *PMEL* variants associated with color dilution in cattle (Kuhn and Weikard 2007; Schmutz and Dreger 2013). Samples from 178 yak (7 run in duplicate), 3 yak hybrids, and 8 cattle were genotyped for the 97 markers at Neogen Geneseek (Lincoln, NE; Supplementary Table 3). In addition to those animals, the genotyping company included a cattle control of unknown breed origin as well as a duplicate sample of Queen Allante. Resulting genotype data were filtered for a per locus genotyping rate >80%. Additionally, animals failing to genotype at 88% or better were excluded, as were duplicate samples.

Analysis—Nose Color

Genotypes from the targeted assay were computationally phased using fastPHASE (Scheet and Stephens 2006) under default parameters. Single marker (allelic) and haplotypic associations for nose coloration of yak for all loci except *KIT* were conducted using chi-square tests with significance determined after a Bonferroni correction for multiple testing.

Sanger sequencing for the entire coding region of *MC1R* was conducted on 33 yak (14 imperial, 8 native, 11 unknown) and 4 cattle utilizing the same PCR protocol as for *KIT* with the primers and annealing temperature found in Supplementary Table 1.

Haplotype Analysis

Genotypes across *KIT* and *MC1R* were obtained from 96 cattle representing 19 domestic breeds (Heaton et al. 2016), from WGS of gaur ($N = 2$), banteng ($N = 2$), eland ($N = 1$), plains bison ($N = 1$), water buffalo ($N = 1$), sheep ($N = 1$), and goat ($N = 1$) aligned to UMD3.1 (Kalbfleisch and Heaton 2013), and from Queen Allante (Heaton et al. 2016), and the Chinese yak (Qiu et al. 2012). The genotypes for these additional animals, derived from vcf data using FastaAlternateReferenceMaker in GATK (McKenna et al. 2010) including flag –useIUPAC, were phased in conjunction with the data from the yak using the default parameters of Beagle (Browning and Browning 2007). Evidence of mutation or recombination altering the length of the conserved haplotype was used to define haplotype boundaries. A TCS haplotype network (Clement 2002) was generated for both *KIT* and *MC1R* in PopART (Leigh and Bryant 2015). Also for *MC1R*, the full gene sequence across samples was aligned for building of a maximum likelihood tree using the IQ-TREE software (Nguyen et al. 2015) with the nucleotide substitution model determined using AIC as described in Kalyaanamoorthy et al. (2017), branch support approximated with 1000 bootstrap trees using UFBoot2 (Hoang et al. 2018), and sheep drawn as the root. The resulting tree was visualized in FigTree 1.4.3 (Rambaut 2016).

Results

A Haplotype of *KIT* is Associated With Trim and Royal Coat Color in North American Yak

The translocation alleles (Cs_{29}, Cs_{6}) previously associated with color sidedness in Chinese yak (Zhang et al. 2014b) were not present in any of the 14 North American yak in which they were evaluated, regardless of their coat color (Supplementary Figure 1).

Sanger sequencing of *KIT* in 6 yak, representing 2 of each coat color variation, revealed 51 variants across the locus (Table 1). All variants were synonymous or intronic with the exception of a missense variant (g.71872160T>C) predicted to alter the corresponding amino acid (c.584T>C; Met195Thr) with "moderate" or "neutral" impact according to snpEff and PROVEAN predictions, respectively. The 2 solid yak were homozygous across the locus (CC), the trim yak were heterozygous (TC), and the 2 royal yak homozygous (TT) for the allele alternative to that found in the solid yak.

Sequencing of Exon 3, which contained the missense variant (Met195Thr) and 3 proximal noncoding variants in 39 additional animals further supported the association of this locus with coat color pattern ($P < 0.001$); the *KIT* genotype of one of the 39 yak sequenced did not correspond to his registered coat color. Photos of the animal revealed minimal white markings on a hind foot, meaning that the animal was phenotypically trim but registered incorrectly as solid black. After correcting that phenotype, the genotype of all 39 animals at Met195Thr completely segregated with their phenotypic coat color ($P = 0$). With the exception of an intronic variant, g.71871905A>G, for which 2 solid animals were heterozygous, the variants identified and assayed via Sanger sequencing, including Met195Thr, were present in 2 conserved haplotypes spanning the gene; the 2 homozygous states were found in solid black (c.584T>C allele C) or royal (allele T) animals, with the heterozygous state segregating with the intermediate, trim phenotype (Table 1).

Whole-genome sequencing resulted in 120.2 and 84.6 million reads for the royal and imperial yak, respectively; the reads, when mapped to UMD3.1, resulted in an average coverage of 3.2X and 3.4X. With low coverage, and thus low confidence in genotyping calling, regions with few aligned reads were evaluated visually in the Integrative Genomics Viewer (Broad Institute; Cambridge, MA) to identify potential variants in *KIT* as well as in other candidate genes for color (*ASIP, CBD103, KITL, MC1R, MSH, PMEL*, and *TYRP1*). Including variants identified by both whole-genome and Sanger sequencing, 97 loci were assayed by MALDI-TOF, of which 36 were fixed across the animals genotyped, likely attributable to incorrect genotype calling in low coverage regions of WGS. Quality pruning of the MALDI-TOF genotypes resulted in data from 178 unique animals (167 yak, 3 yak hybrids, and 8 cattle) for analyses. The 3 hybrid yak were homozygous for the Charolais dilution (D_{ζ}; Kuhn and Weikard 2007) and were thus removed from the remaining analyses. In addition to the fixed loci, 5 were removed for genotyping failure rate and 2 for deviation from HWE, leaving 54 loci for analysis, including 37 loci in and proximal to *KIT*. At that locus, markers spanned from 38 kb upstream of the annotated 5′UTR to 557 kb downstream (Supplementary Table 3).

After quality filtering, in the 167 yak genotyped with the MALDI assay, the Met195Thr variant again segregated with coat color in a semi-dominant manner (Black=p.Met195Thr/p.Met195Thr, Trim=+/p.Met195Thr, Royal=+/+; $P<0.001$). With these additional genotype data, a 25-SNP, 388kb haplotype ("trim" haplotype; Table 2) surrounding this variant was identified, beginning in Exon 3 of the gene and extending approximately 20 kb 3′ of *KIT*. This trim haplotype was shared by all animals with white coat coloration. In 80.5% (128 of 159) of the trim chromosomes assayed, the haplotype extended 749.9 kb across all 37 SNPs genotyped in *KIT*. Chromosomes not spanning the entire region showed variation in the penultimate 5′ SNP (upstream of *KIT*), while evidence of recombination was identified 3′ of the gene. The alleles found by Sanger sequencing of *KIT* to be associated with black coat color were contained in a 10-SNP, 66 kb haplotype. This 10-SNP "black" haplotype

Table 1. Variants identified among 6 yak by Sanger sequencing of *KIT* exons (UMD3.1 Chromosome 6; Transcript ENSBTAT00000003498.5) with the genomic position with respect to ARS-UCD1.2 noted

Position (bp) UMD3.1	Position (bp) ARS-UCD1.2	Variant ID	Ref	Alt	Black	Trim	Royal	Variant position (predicted consequence) based on ENSBTAT00000003498.5
71796285	70166659	rs110394433	T	G	G/G	G/T	T/T	Upstream Variant
71796527	70166901	.	TT	T	T/T	T/TT	TT/TT	Intron 1 Variant
71868653	70205292	rs109344937	C	T	T/T	T/C	C/C	Synonymous Variant, Exon 2, c.177C>T, p.Thr59Thr
71868885	70205524	rs799246224	C	T	T/T	T/C	C/C	Intron 2 Variant
71871905	70208544	.	A	G	G/G[a]	G/A	A/A	Intron 2 Variant
71872059	70208698	rs109314357	A	G	G/G	G/A	A/A	Synonymous Variant, Exon 3, c.483A>G, p.Thr161Thr
71872160	70208799	.	T	C	C/C	C/T	T/T	Missense Variant (Moderate), Exon 3, c.584T>C, p.Met195Thr
71872250	70208889	rs467478061	A	ACTTCT	ACTTCT/ACTTCT	A/ACTTCT	A/A	Intron 3 Variant
71873480	70210119	rs109649112	C	G	G/G	G/C	C/C	Intron 4 Variant
71873703	70210342	.	A	G	G/G	G/A	A/A	Intron 4 Variant
71873749	70210388	rs1116496751	G	T	T/T	T/G	G/G	Intron 4 Variant
71877765	70214407	rs435078996	C	T	T/T	T/C	C/C	Splice Region and Intron 5 Variant
71877838	70214480	.	G	A	A/A	A/G	G/G	Intron 5 Variant
71882252	70218874	rs109236495	C	T	T/T	T/C	C/C	Intron 5 Variant
71882552	70219174	.	C	T	T/T	T/C	C/C	Intron 6 Variant
71882595	70219216	rs209891374	T	G	T/T	T/G	G/G	Intron 6 Variant
71884433	70221057	rs108989845	C	T	T/T	T/C	C/C	Intron 6 Variant
71884478	70221102	rs109078616	G	A	A/A	A/G	G/G	Intron 6 Variant
71884491	70221115	rs109862472	G	T	T/T	T/G	G/G	Intron 6 Variant
71884601	70221225	.	G	A	A/A	A/G	G/G	Intron 6 Variant
71900932	70237545	rs110595646	T	C	C/C	C/T	T/T	Intron 7 Variant
71901078	70237691	rs109745851	G	A	A/A	A/G	G/G	Intron 7 Variant
71901286	70237899	.	TG	T	T/T	TG/G	TG/TG	Intron 8 Variant
71901298	70237911	rs109723937	AT	A	A/A	AT/A	AT/AT	Intron 8 Variant
71902596	70239209	rs110871960	G	A	A/A	A/G	G/G	Intron 8 Variant
71902723	70239336	rs109921120	C	G	C/C	C/G	G/G	Intron 8 Variant
71904393	70241005	.	T	C	C/C	C/T	T/T	Synonymous Variant, Exon 10, c.1644T>C, p.Tyr548Tyr
71904433	70241045	.	G	A	A/A	A/G	G/G	Intron 10 Variant
71904863	70241475	rs207688993	T	C	C/C	C/T	T/T	Intron 11 Variant
71905095	70241707	rs378154728	C	A	A/A	A/C	C/C	Synonymous Variant, Exon 13, c.1899C>A, p.Thr633Thr
71906885	70243497	rs110171094	T	A	A/A	A/T	T/T	Intron 14 Variant
71908561	70245173	rs109891739	T	C	C/C	C/T	T/T	Intron 14 Variant
71908774	70245386	rs110798632	C	G	C/C	C/G	G/G	Intron 15 Variant
71908788	70245400	rs383702685	G	GTTC	GTTC/GTTC	G/GTTC	G/G	Intron 15 Variant
71909262	70245874	rs110818069	G	A	A/A	A/G	G/G	Synonymous Variant, Exon 16, c.2346G>A, p.Ala782Ala
71909906	70246518	rs110881216	T	C	T/T	T/C	C/C	Intron 16 Variant
71910036	70246648	rs110901406	A	AG	A/A	A/AG	AG/AG	Intron 16 Variant
71910091	70246704	.	G	A	A/A	A/G	G/G	Intron 16 Variant
71913208	70249821	.	G	A	A/A	A/G	G/G	Intron 17 Variant
71913504	70250117	.	C	G	G/G	G/C	C/C	Intron 18 Variant
71913721	70250334	rs1115157736	G	A	G/G	G/A	A/A	Intron 19 Variant
71914233	70250846	.	G	T	T/T	T/G	G/G	Intron 20 Variant
71914272	70250885	rs109479879	T	C	C/C	C/T	T/T	Intron 20 Variant
71914295	70250908	.	T	C	C/C	C/T	T/T	Intron 20 Variant
71914316	70250929	.	T	G	G/G	G/T	T/T	Intron 20 Variant
71915311	70251924	.	C	T	T/T	T/C	C/C	Synonymous Variant, Exon 21, c.2922C>T, p.His974His
71915332	70251945	.	G	A	A/A	A/G	G/G	3′ UTR Variant
71915337	70251950	rs799571173	T	C	C/C	C/T	T/T	3′ UTR Variant
71915355	70251968	.	C	T	T/T	T/C	C/C	3′ UTR Variant
71915529	70252142	rs801319071	A	T	T/T	T/A	A/A	3′ UTR Variant
71915774	70252387	.	G	A	A/A	A/G	G/G	3′ UTR Variant

Reference (Ref) and Alternative (Alt) alleles are defined with respect to UMD3.1. Variants without a variant ID are novel to this study. Observed genotypes of yak representing each coat color pattern ($N = 2$) are given with the exception of those marked in bold for which 33 animals were sequenced.

[a]Two black yak were heterozygous (G/A) at this locus.

Table 2. The 25-SNP *KIT* haplotype conserved in all yak with white markings (trim or royal) and corresponding haplotypes associated with solid black coat color

	Position (bp) Chromosome 6 (UMD3.1)																								
	71834186	71844222	71855663	71863243	71868653	71871905	71872160	71877765	71887903	71902723	71908561	71913504	71913721	71915529	71926526	71937845	71945528	71952915	71991742	71993435	72018821	72028441	72130781	72182635	72222821	Frequency (black haplotypes)
Trim	A	C	T	G	C	A	T	C	T	G	T	C	A	A	A	T	A	A	T	A	C	G	T	G	C	
Qiu et al. 2012	A	C	T	G	C	A	T	C	T	G	T	C	A	A	A	T	A	A	T	G	C	G	T	G	C	
Black 1	T	T	C	A	T	G	**C**	**T**	**C**	**C**	**C**	**G**	**G**	**T**	**G**	**G**	G	G	C	G	T	A	A	A	G	0.080
Black 2	T	T	C	A	T	G	**C**	**T**	**C**	**C**	**C**	**G**	**G**	**T**	**G**	**G**	G	G	C	G	T	G	A	A	G	0.006
Black 3	T	T	C	A	T	A	**C**	**T**	**C**	**C**	**C**	**G**	**G**	**T**	**G**	**G**	A	G	C	G	T	A	A	A	G	0.045
Black 4	T	T	C	A	T	G	**C**	**T**	**C**	**C**	**C**	**G**	**G**	**T**	**G**	**G**	A	G	C	G	T	G	A	A	G	0.006
Black 5	T	T	C	A	T	A	**C**	**T**	**C**	**C**	**C**	**G**	**G**	**T**	**G**	**G**	A	G	C	G	T	G	A	A	G	0.396
Black 6	T	T	C	A	T	G	**C**	**T**	**C**	**C**	**C**	**G**	**G**	**T**	**G**	**G**	A	G	C	G	T	A	A	A	G	0.466
Black 7	T	T	C	A	T	G	**C**	**T**	**C**	**C**	**C**	**G**	**G**	**T**	**G**	**G**	G	G	C	G	T	G	A	A	C	0.006
Black 8	A	T	C	A	T	G	**C**	**T**	**C**	**C**	**C**	**G**	**G**	**T**	**G**	**G**	G	G	C	G	T	A	A	A	G	0.006
	70170826	70180862	70192303	70199883	70205292	70208544	70208799	70214407	70224524	70239336	70245173	70250117	70250334	70252142	70263138	70274457	70282136	70289525	70328352	70330045	70355431	70365049	70467047	70518910	70557522	
	Position (bp) Chromosome 6 (ARS-UCD1.2)																									

The second haplotype for the presumed trim yak reported in Qiu et al. (2012) is identical to black Haplotype 8. The frequency that each 25-SNP haplotype associated with solid coat color was observed, is given and the boundaries of the 10-SNP conserved region are shown in bold.

Table 3. Count of imperial (black nose) and native (gray nose) yak having each of the 3-SNP genotypes observed across *MC1R*

	MC1R 3-SNP genotype		
	CC AA GG	CT AG GA	TT GG AA
Imperial	6	43	0
Native	0	0	66

The markers are on Chromosome 6 at 14757989, 14758485, and 14758950bp (UMD3.1), respectively.

was within the 25-SNP region associated with white patterning described above (Table 2). In the full, 388kb region, 8 haplotypes were identified containing the 10-SNP motif associated with solid black coat color. The 2 most common black haplotypes across the 388kb region accounted for 86% of all observations; these haplotypes differed by 2 base pairs: g.71871905A>G, the intronic SNP also found by Sanger sequencing to vary among solid individuals, and g.72028441G>A, downstream of *KIT* (Table 2).

MC1R Variants are Associated With the Imperial Nose Color

Nose color phenotypes were reported for 115 yak (49 imperial, 66 native) passing quality control measures for the MALDI-TOF assay. Single marker (chi-square) tests of the 17 SNP genotypes of these animals across the 7 candidate genes assayed via MALDI-TOF identified a significant association ($P < 0.001$) of the imperial phenotype with all 3 SNPs genotyped in *MC1R* (Table 3). No other loci remained significant after correction for multiple testing. As the 3 *MC1R* markers are in close proximity, the analysis was repeated utilizing haplotypic association across all loci, further supporting the statistical association of *MC1R* with nose color ($P < 0.001$). Upon inspection, a single 3-SNP haplotype of *MC1R* segregated completely with the phenotype (the "imperial" haplotype). These markers included one synonymous (g. 14757989C>T), one missense (rs135181132, c.871G>A, p.A291T), and one 3′UTR variant (g.14758950G>A; Table 4). The yak with a gray nose (native) had 1 of 3 haplotypes at *MC1R*, which differed from one another at a synonymous (rs442354353, g.14758256C>G) and 3′UTR (rs481591010, 14758691C>T) variant (Table 4).

Sanger sequencing of the complete *MC1R* locus in 33 yak (14 imperial, 8 native, 11 unknown nose color) identified 13 variable sites, with 2 additional variants identified in 4 domestic cattle (2 Holstein, 2 Hereford) representing the dominant black (E^D; ss974293047) and recessive red (*e*, rs110710422) alleles (Klungland et al. 1995; Table 5). Of the variants observed in yak, 2 were missense variants, 3 synonymous, and the remainder in the 5′ or 3′ UTR. Phasing of the data showed these 13 variants were found within 4 haplotypes. Considering only the 22 individuals with known nose color, a single haplotype containing the 3 loci associated with black nose color from the MALDI-TOF assay was found only in individuals with imperial nose color (Table 5). Two variants, p.Q114K and p.A291T, unique to the "imperial" haplotype were predicted by SNPEff (Cingolani et al. 2012) to have "moderate" impact; PROVEAN results suggest these variants are "neutral." These SNPs correspond to variants identified as comprising haplotype "Y1" in Chen et al. (2009).

Of the 49 yak in the genotyping panel that were reported to have a black (imperial) nose, 6 were homozygous and 43 heterozygous for the *MC1R*, imperial haplotype. The nose color phenotype (native vs. imperial) was not reported for 52 yak, which is common for royal animals that generally have white across their nose masking the trait. Eleven of these 52 yak without a recorded nose color phenotype were heterozygous for the *MC1R* imperial haplotype. The allele associated with imperial nose color, therefore, was present at a frequency of 0.198 in the population sampled.

Table 4. Variants identified by Sanger sequencing of *MC1R* and the haplotypes associated with imperial (black) and native (gray) nose phenotypes

Position (bp)					Observed haplotypes				
UMD 3.1	ARS-UCD 1.2	Variant ID	Ref	Alt	Imperial	Native 1	Native 2	Native 3	Variant position (predicted consequence), annotation
14757353	14705114	.	G	A	G	A	A	A	5′ UTR, c.-262G>A
14757486	14705247	rs525922202	T	C	T	C	C	C	5′ UTR, c.-129T>C
14757488	14705249	.	A	C	A	C	C	C	5′ UTR, c.-127A>C
14757509	14705270	rs460981838	C	T	C	T	T	T	5′ UTR, gain of start (Low), c.-106C>T
14757910[b]	14705671	ss974293047	T	C	T	T	T	T	Missense (Moderate), c.296T>C, p.Leu99Pro
14757924[b]	14705685	rs110710422	GG	G	GG	GG	GG	GG	Frameshift (High), c.311delG, p.Gly104fs
14757954	14705715	.	C	A	A	C	C	C	Missense (Moderate), c.340C>A, p.Gln114Lys
14757989[a]	14705750	.	C	T	C	T	T	T	Synonymous, c.375C>T
14758256	14706017	rs442354353	C	G	C	C	G	C	Synonymous, c.642C>G
14758277	14706038	rs525311468	T	C	T	C	C	C	Synonymous, c.663T>C
14758485[a]	14706246	rs135181132	G	A	A	G	G	G	Missense (Moderate), c.871G>A, p.Ala291Thr
14758691	14706452	rs481591010	C	T	C	T	C	C	3′ UTR, c.*123C>T
14758692	14706453	rs442584695	T	G	T	G	G	G	3′ UTR, c.*124T>G
14758896	14706657		TG	T	TG	T	T	T	3′ UTR, c.*334delG
14758950[a]	14706711		G	A	G	A	A	A	3′ UTR, c.*482A>G

Reference (Ref) and Alternative (Alt) alleles are defined with respect to UMD3.1. Variants (bp) shown in bold were described in the Y1 haplotype of Chen et al. (2009).
[a]Variants included in the MALDI genotyping assay.
[b]Variants found only in domestic cattle.

Table 5. Genotype information as determined by whole-genome sequence of yak representing 3 coat colors: native black (Queen Allante), trim (Chinese yak), and royal for the *KIT* variants associated with the white-face and wildtype haplotypes outlined in Zhang et al. (2014a)

	Position (bp)								
Chr	Btau 4.6.1	UMD 3.1	ARS-UCD1.2	Variant ID	Wildtype haplotype (S^{+})	White-face haplotype (S^{wf})	Queen Allante	Chinese Yak (Qiu et al. 2012)	Royal Yak
6	72744124	71868113	70204752	rs209595303	T	C	T/T	T/C	C/C
6	72744159	71868148	70204787	rs799436050	G	T	G/G	G/T	T/T
6	72744160	71868149	70204788	.	A	G	A/A	A/G	G/G
6	72744162	71868151	70204790	.	G	A	G/G	G/A	A/A
6	72744209	71868198	70204837	rs797660744	T	C	T/T	T/C	C/C
6	72744255	71868244	70204883	rs1116506610	G	A	G/G	G/A	A/A
6	72744262	71868251	70204890	rs211116638	G	A	G/G	G/A	A/A
6	72744283	71868272	70204911	.	G	A	G/G	G/A	A/A
6	72744334	71868323	70204962	rs799743560	T	C	T/T	T/C	C/C
6	72748070	71872059	70208698	rs109314357	G	A	G/G	G/A	A/A
6	72748262	71872251	70208890	rs109704112	CTTCTC	C	CTTCTC/CTTCTC	CTTCTC/C	C[a]
6	72777273	71901287	70237900	rs475037828	T	TG	T/T	T/TG	TG/TG
6	72777285	71901299	70237912	rs109723937	A	AT	A/A	A/AT	AT/AT
6	72784776	71908791	70245403	rs471701792	C	CTTC	CTTC/CTTC	C/CTTC	C[b]
6	72791146	71915160	70251773	rs449318084	G	T	G/G	G/T	T/T

The variants were described utilizing the Btau 4.6.1 reference genome; the corresponding coordinates in UMD3.1 and ARS-UCD1.2 were identified using the NCBI Genome Remapping Service. Sequence data from the imperial black yak bull sequenced for this study were not included due to low coverage in the region. Regions with low sequence coverage ([a]1 read, [b]2 reads) are noted.

KIT and *MC1R* Pigmentation Haplotypes in North American Yak are Most Proximal to Those From Domestic Cattle

Phasing the 25-SNP *KIT* genotype of the trim haplotype from the whole-genome sequence data of the yak reported in Qiu et al. (2012) revealed one haplotype of that yak shared 24 of the 25 alleles of the trim haplotype identified in our samples; the exception was an intergenic SNP at g.71993435G>A. The trim haplotype of the North American yak was also similar to that associated with the white-face phenotype, S^{wf}, of (Zhang et al. 2014b; Table 5). The other chromosome of the Chinese yak (Qiu et al. 2012) contained the same, 10-SNP haplotype associated with black coat color in North American yak (Table 2).

Comparing data across species, the 25-SNP trim haplotype differed from the most closely related haplotype by one variant in intron 19 (rs1115157736, g.71913721G>A); this haplotype was observed 14 times in domestic cattle including Hereford (6), Charolais (2), Tarentaise (2), Beefmaster (1), Maine-Anjou (1), Simmental (1), and Texas Longhorn (1). None of the domestic cattle genotyped shared a haplotype with any yak.

The TCS haplotype network of *KIT* positioned the trim and black yak haplotypes at opposing ends (Figure 2). The trim haplotype was most proximal to a cluster of haplotypes found in domestic cattle. Conversely, the haplotypes associated with black coat color in yak were more closely positioned to those found in Gaur, Banteng, and Plains Bison than they were to haplotypes found in domestic cattle (Figure 2). Similar results were found when evaluating only the 10 SNPs in the most conserved region implicated in yak coat color (Supplementary Figure 2).

Phasing genotypes across *MC1R* from Sanger sequence data of the 33 yak, WGS of the 96 domestic cattle, Queen Allante and the Chinese yak, and that of individuals of other species revealed 176 variable sites and 42 total haplotypes; 22 of these were unique to cattle. One additional "native" haplotype was identified in Queen Allante. A consensus maximum likelihood tree, built with the TVM+F+I+G4 model of substitution, placed the single haplotype associated with imperial nose color of yak in a clade with domestic cattle haplotypes, while those found in individuals with the native nose phenotype were positioned in a highly-supported clade that included the Plains Bison (Figure 3). Similar to the maximum likelihood tree, the relative relationships among *MC1R* haplotypes across the species was supported by the haplotype network (Supplementary Figure 3). The imperial haplotype was 3bp divergent from the most closely related haplotype found in domestic cattle; this haplotype was found in 58 animals representing 14 different cattle breeds.

Discussion

The analyses conducted support the role of *KIT* in the determination of white patterning in North American yak. Further, genotyping of several candidate genes resulted in the identification of *MC1R* as a dominant determinant of nose pigmentation. The associated haplotypes and their relationship to haplotypes reported in Asian yak support previous records of cattle introgression into domestic yak dating prior to their importation into North America. These data help to elucidate the genetic determination of these color phenotypes and contribute to building a better understanding of the history of domestic yak in North America.

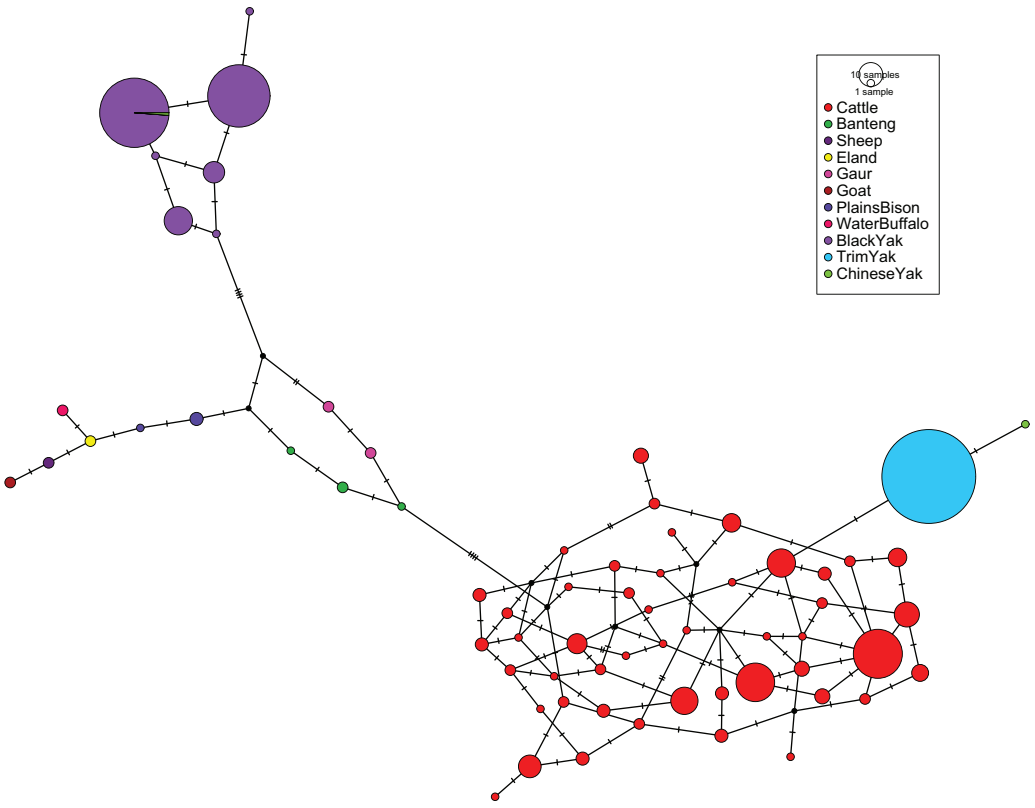

Figure 2. TCS haplotype network built from data across 25 SNPs across the *KIT* locus. The size of each circle is proportional to the number of animals observed with each haplotype. The yak haplotype carrying alleles associated with white coloration is considered the trim haplotype.

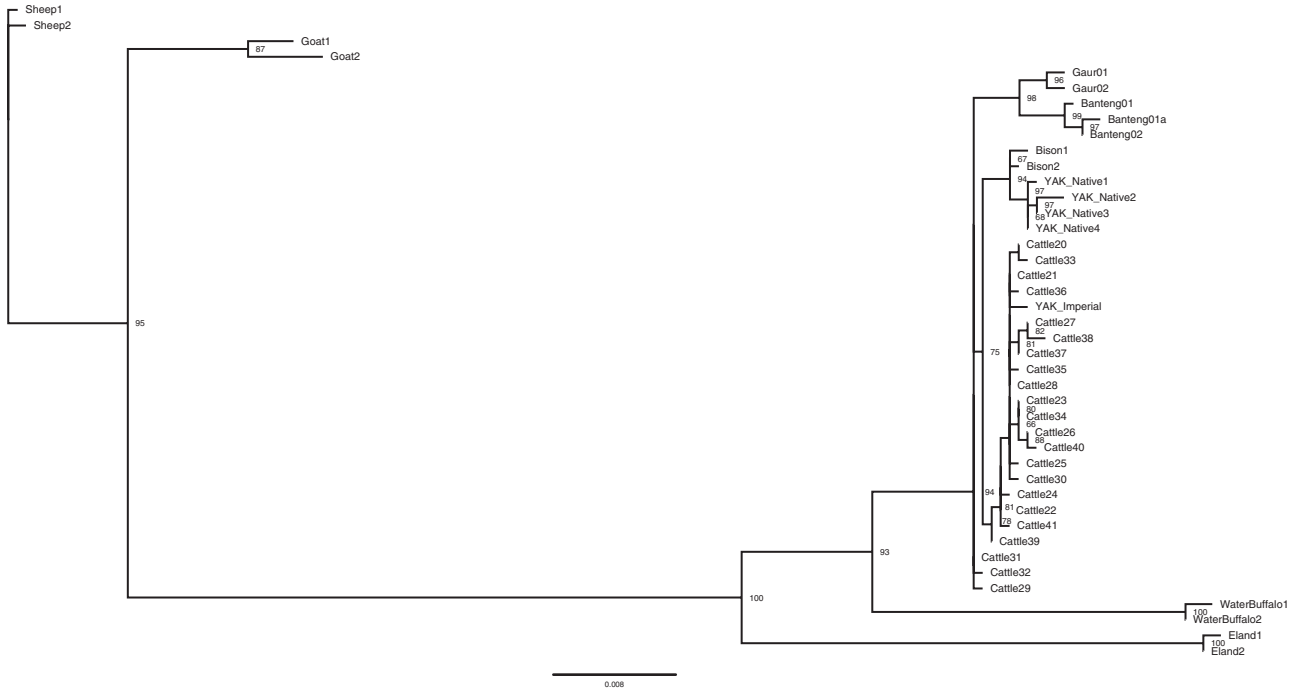

Figure 3. Consensus, maximum likelihood tree illustrating the relationships among *MC1R* haplotypes across species. Bootstrap support (percent of 1000 replicates) above 50 is shown. All haplotypes in yak associated with native nose color are positioned in the clade with bison. The haplotype associated with imperial nose color is positioned in the clade with cattle.

White Patterning and *KIT*

The translocation alleles, Cs_{29} and Cs_{6}, implicated in witrik patterning, and also found in all-white yak (Durkin et al. 2012; Zhang et al. 2014b) were not present in the samples we investigated. The absence of these alleles in our samples was not surprising given the phenotypic difference in distribution of white color on North American animals. The haplotype associated with trim and royal patterning, however, was nearly identical to the 15-variant haplotype identified by Zhang et al. (2014b) associated with the white-face phenotype of Asian yak. The white-face phenotype differs from trim yak in that trim animals have only a small white marking on their forehead rather than a full white mask. Additionally, the haplotype found in the North American animals is semi-dominant in expression while that associated with the white-face phenotype appears to be dominant (Zhang et al. 2014b). The MALDI genotyping assay did not target all 15 variants of the white-face haplotype nor did Zhang et al. (2014b) sequence the entire *KIT* locus in the identification of the white-face allele. The WGS of the 3 North American yak (2 solid black and one royal), however, allowed for the identification of what was described as white-face haplotype in the homozygous state in the royal yak but absent in the solid animals with the exception of an insertion (rs471701792, UMD3.1 g.71908791) in white-face animals. This insertion was absent from the royal yak sequenced although its absence cannot be confirmed due to low sequence coverage at this region. Even if this insertion were absent, the similarity between the haplotypes suggests the trim/royal and white face phenotypes have a common origin. This relationship between the haplotypes associated with white markings is notable as the trim haplotype otherwise is more similar to those found in domestic cattle than in the other *Bos* species studied. Markedly, the position of the trim haplotype in context with the other data finds it most distant from the alternative, black haplotype of *KIT* found in North American yak.

The missense variant in *KIT* (g.71872160T>C, p.M195T) is interesting as the trim haplotype contains the UMD3.1 reference allele while the alternative allele is found in the haplotype associated with black color. This variant, the only one with a predicted impact on amino acid sequence may contribute to white patterning. By itself, however, it was ruled out as causative of the trim or royal coat pattern because if that were the case, the Hereford reference animal, as well as the other domestic cattle that share the same variant should have trim or royal white patterning. Conversely, if this variant alone determined the Hereford pattern, the yak would instead demonstrate that variation in white markings. Yak, however, have a different distribution of white markings. Further, Provean prediction suggests the impact of this substitution is "neutral." Possibly narrowing the region hypothesized to contain the functional variation causative of the trim/royal coloring, the Chinese domestic yak (Qiu et al. 2012) had a truncated trim haplotype, deviating from the 25-SNP allele seen in the North American yak at g.71993435A>G (3′ of *KIT*). That said, *KIT* is a complex locus with known structural variation (Whitacre 2014). Therefore, while this work confidently shows that *KIT* is involved in white patterning in North American yak, it is possible that the haplotype identified is tagging additional variation in the locus that has not yet been characterized.

MC1R is Associated With Black Nose Color

Of the 7 candidate genes considered, *MC1R* was significantly associated with nose pigmentation, supporting pedigree observations of dominant inheritance of the imperial nose color. The *MC1R* haplotype of the imperial individuals is similar to the Y1 haplotype identified by Chen et al. (2009). In that work, the Y1 haplotype was common in yak found in 2 sampling locations, Tianzhu and Maiwa, and was also present in Jiulong yak. The haplotype associated with imperial nose color, therefore, is not unique to North American

animals. Further, as demonstrated with respect to *KIT*, the relationship of the imperial haplotype with those found in other species suggests it also was derived from introgression with cattle.

As noted in the prior work of Chen et al. (2009), the p.Q114K and p.A291T variants characteristic of this imperial *MC1R* haplotype are found in the first extracellular loop and seventh transmembrane domain of the receptor, respectively. Both PROVEAN and SnpEff predictions do not suggest a significant impact of either variant although Chen et al. (2009) reports PANTHER (Thomas et al. 2003) predicts the latter to have a greater functional impact than the former. It is uncertain if the 2 variants in concert may play a greater role than either is predicted to have alone. The p.A291T variant was also reported in Chinese cattle (Zhang et al. 2014a) as well as in the related gayal (*B. frontalis*; Xi et al. 2012a); however, neither study noted p.Q114K, suggesting it is contemporary to p.A291T, which appears to have been derived before the divergence of domestic cattle, gayal, and yak. Unfortunately, nose color of the animals studied in Zhang et al. (2014a) was not reported.

Similar to the current results, which place the trim and imperial haplotypes most proximal to those found in domestic cattle, Chen et al. (2009) suggested the *MC1R* Y1 haplotype in yak originated from cattle introgression. These data are consistent with the work of Wu et al. (2018), who show extensive introgression between yak and Tibetan cattle, as well as with whole-genome sequence analysis that estimates 1.3% of the yak genome is derived from cattle and which suggest cattle x yak hybridization has been nearly continuous over the past 1500 years (Medugorac et al. 2017). Prior and continued introgression of cattle into domestic yak populations is also supported by mitochondrial data (Lai et al. 2007), reports of admixed populations in the native range of yak (Kislovsky 1938; Phillips et al. 1946), and documents from 11th century China (Zhang 2000). Our data contribute further evidence of this complicated evolutionary relationship among the species. Intentional hybridization of yak and domestic cattle continues in some North American production systems for the purpose of enhancing meat production or to produce animals with solid white coat color; this was apparent in the 3 hybrid animals found to have the Charolais color dilution. Although we cannot completely rule out recent introgression with domestic cattle as the origin of these color alleles, neither the trim haplotype nor that conferring black nose color in the yak was shared with any domestic cattle studied; those cattle breeds represent both those commonly crossed with yak (e.g., Charolais and Highland) as well as popular breeds found in the United States (Heaton et al. 2016). Further, estimates of the proportion of cattle introgression (mean of 0.38%) available from nearly 80% of the samples studied support limited domestic cattle ancestry.

Utility of Markers for Testing

While the results confidently associate these variants with white patterning and nose pigment, a few discrepancies in the association of phenotype and genotype were present in these data. In 4 instances, the observed *KIT* genotype did not perfectly correlate with the reported phenotype; 2 of these cases are believed to be errors in sample handling or recording at the time hairs were collected (e.g., swapping of 2 samples). As noted above, in one instance an animal reported as solid black was found to have white on one foot; similarly, another yak reported to be black genotyped heterozygous across the *KIT* locus and upon further physical inspection was found to have white hairs on her forehead. These cases of mistakes in registration due to very subtle white markings supports the use of coat color genotype as one means to verify animal identity and for breeders who are interested in achieving specific coat color outcomes in their matings.

Supplementary Material

Supplementary data are available at *Journal of Heredity* online.

Supplementary Figure 1. PCR products from amplification with translocation allele primers. α-D identifies the Cs_{29} translocation allele, which confers color-sidedness. C-β identifies the Cs_{29} translocation allele and/or the Cs_{6} translocation allele, which both confer color-sidedness. α-β identifies wild-type chromosome 29. 1–14 are yaks where 1, 3, 7, 9 = native black; 4, 8 = imperial black; 5, 6, 13 = imperial trim; 11, 12, 13, 14 = native trim; 2, 10 = royal. W = witrik bull (positive control); H = "regular" Holstein bull; - = negative control; L = size standard ladder. Expected PCR product sizes are indicated in parentheses.

Supplementary Figure 2. TCS haplotype network across the 10 SNPs of the *KIT* region for which all chromosomes associated with black coat color in yak were conserved. The size of each circle is proportional to the number of individuals observed with each haplotype.

Supplementary Figure 3. TCS haplotype network for *MC1R* (1,764bp). The size of each circle is proportional to the number of individuals observed with each haplotype.

Supplementary Table 1. PCR primers, annealing temperature (°C), and expected product size (bp) used in Sanger sequencing of the exons of *KIT* and *MC1R*.

Supplementary Table 2. *KIT* translocation alleles PCR primers (from Durkin et al. 2012) and expected product sizes.

Supplementary Table 3. The 97 loci genotyped at 8 candidate loci for coat color. Noted is the predicted consequence of the variant and outcome (included vs. excluded from study) after genotyping on the MALDI-TOF platform. Those variants that were fixed across the samples were identified per visualization of-bam files in IGV from WGS of two yak that was of low coverage across some loci and likely are not true variants.

Funding

The University of Nebraska DNA Sequencing Core receives partial support from the National Institute for General Medical Science (NIGMS) INBRE - P20GM103427-14 and COBRE - 1P30GM110768-01 grants as well as The Fred & Pamela Buffett Cancer Center Support Grant - P30CA036727. This publication's contents are the sole responsibility of the authors and do not necessarily represent the official views of the NIH or NIGMS.

Acknowledgments

Students assisting in sample collection, DNA isolation, and sequencing: Michelle Gibbens, Ryan Hagenson, Sara Nilson, and Julia Yurco. Assay development: Veronica Basnayake and JR Tait (Neogen GeneSeek). Whole-genome sequence analyses were performed, in part, using the resources of the University of Nebraska's Holland Computing Center. Photos: Dr. Peter Hackett and Dianne Latona. Sample contributions and study support: International Yak Association and USYAKS. All protocols conformed to the policies set forth by the University of Nebraska Institutional Animal Care and Use Committee.

Data Availability

Next-generation sequence data have been deposited in the NCBI Sequence Read Archive (SRA) as BioProject PRJNA529217 and

novel variants deposited in the EMBL-EBI European Variation Archive (EVA). Sanger sequence data of individuals representing variation observed in *KIT* and *MC1R* are available in Genbank (Accession MN444722-MN444780).

References

Brooks SA, Lear TL, Adelson DL, Bailey E. 2007. A chromosome inversion near the KIT gene and the Tobiano spotting pattern in horses. *Cytogenet Genome Res*. 119:225–230.

Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*. 81:1084–1097.

Chen SY, Huang Y, Zhu Q, Fontanesi L, Yao YG, Liu YP. 2009. Sequence characterization of the MC1R gene in yak (*Poephagus grunniens*) breeds with different coat colors. *J Biomed Biotechnol*. 2009:861046.

Choi Y. 2012. A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein. *In Proceedings of the ACM Conference on Bioinformatics, Computational Biology, and Biomedicine*, p. 414–417, New York.

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012. Predicting the functional effect of amino acid substitutions and indels. *PLoS One*. 7:e46688.

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 6:80–92.

Clement MSQ, Walke P, Posada D, Crandall K. 2002. TCS: estimating gene genealogies. In: *16th International Parallel and Distributed Processing Symposium*, p. 184, Fort Lauderdale, FL.

Cruickshank J, Dentine MR, Berger PJ, Kirkpatrick. 2004. Evidence for quantitative trait loci affecting twinning rate in North American Holstein cattle. Anim Genet 35:206–212.

David VA, Menotti-Raymond M, Wallace AC, Roelke M, Kehler J, Leighty R, Eizirik E, Hannah SS, Nelson G, Schäffer AA, *et al.* 2014. Endogenous retrovirus insertion in the KIT oncogene determines white and white spotting in domestic cats. *G3 (Bethesda)*. 4:1881–1891.

Dürig N, Jude R, Holl H, Brooks SA, Lafayette C, Jagannathan V, Leeb T. 2017. Whole genome sequencing reveals a novel deletion variant in the KIT gene in horses with white spotted coat colour phenotypes. *Anim Genet*. 48:483–485.

Durkin K, Coppieters W, Drögemüller C, Ahariz N, Cambisano N, Druet T, Fasquelle C, Haile A, Horin P, Huang L, *et al.* 2012. Serial translocation by means of circular intermediates underlies colour sidedness in cattle. *Nature*. 482:81–84.

Everts RE, Rothuizen J, van Oost BA. 2000. Identification of a premature stop codon in the melanocyte-stimulating hormone receptor gene (MC1R) in Labrador and Golden retrievers with yellow coat colour. *Anim Genet*. 31:194–199.

Fontanesi L, Tazzoli M, Russo V, Beever J. 2010. Genetic heterogeneity at the bovine KIT gene in cattle breeds carrying different putative alleles at the spotting locus. *Anim Genet*. 41:295–303.

Giuffra E, Törnsten A, Marklund S, Bongcam-Rudloff E, Chardon P, Kijas JM, Anderson SI, Archibald AL, Andersson L. 2002. A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking KIT. *Mamm Genome*. 13:569–577.

Grosz MD, MacNeil MD. 1999. The "spotted" locus maps to bovine chromosome 6 in a Hereford-Cross population. *J Hered*. 90:233–236.

Haase B, Brooks SA, Tozaki T, Burger D, Poncet PA, Rieder S, Hasegawa T, Penedo C, Leeb T. 2009. Seven novel KIT mutations in horses with white coat colour phenotypes. *Anim Genet*. 40:623–629.

Heaton MP, Smith TPL, Carnahan JK, Basnayake V, Qiu J, Simpson B, Kalbfleisch TS. 2016. Using diverse US beef cattle genomes to identify missense mutations in EPAS1, a gene associated with high-altitude pulmonary hypertension. *J Anim Sci*. 94:161–162.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 35:518–522.

Joerg H, Fries HR, Meijerink E, Stranzinger GF. 1996. Red coat color in Holstein cattle is associated with a deletion in the MSHR gene. Mamm Genome. 7:317–318.

Kalbfleisch T, Heaton MP. 2013. Mapping whole genome shotgun sequence and variant calling in mammalian species without their reference genomes. *F1000Res*. 2:244.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 14:587–589.

Kijas JM, Wales R, Törnsten A, Chardon P, Moller M, Andersson L. 1998. Melanocortin receptor 1 (MC1R) mutations and coat color in pigs. *Genetics*. 150:1177–1185.

Kislovsky D. 1938. The domestic animals of Mongolia - A review. *J Hered*. 29:27–32.

Klungland H, Våge 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.

Kühn Ch, Weikard R. 2007. An investigation into the genetic background of coat colour dilution in a Charolais x German Holstein F2 resource population. *Anim Genet*. 38:109–113.

Lai SJ, Chen SY, Liu YP, Yao YG. 2007. Mitochondrial DNA sequence diversity and origin of Chinese domestic yak. *Anim Genet*. 38:77–80.

Leigh JW, Bryant D. 2015. POPART: full-feature software for haplotype network construction. *Methods Ecol Evol*. 6:1110–1116.

Mann WM. 1930. *Wild animals in and out of the Zoo*. New York: Smithsonian institution series.

Marklund L, Moller MJ, Sandberg K, Andersson L. 1996. A missense mutation in the gene for melanocyte-stimulating hormone receptor (MC1R) is associated with the chestnut coat color in horses. *Mamm Genome*. 7:895–899.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, *et al.* 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20:1297–1303.

Medugorac I, Graf A, Grohs C, Rothammer S, Zagdsuren Y, Gladyr E, Zinovieva N, Barbieri J, Seichter D, Russ I, *et al.* 2017. Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. *Nat Genet*. 49:470–475.

Newton JM, Wilkie AL, He L, Jordan SA, Metallinos DL, Holmes NG, Jackson IJ, Barsh GS. 2000. Melanocortin 1 receptor variation in the domestic dog. *Mamm Genome*. 11:24–30.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32:268–274.

Phillips RW, Tolstoy IA, Johnson RG. 1946. Yaks and yak-cattle hybrids in Asia. *J Hered*. 37:206; passim.

Qiu Q, Wang L, Wang K, Yang Y, Ma T, Wang Z, Zhang X, Ni Z, Hou F, Long R, *et al.* 2015. Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions. *Nat Commun*. 6:10283.

Qiu Q, Zhang G, Ma T, Qian W, Wang J, Ye Z, Cao C, Hu Q, Kim J, Larkin DM, *et al.* 2012. The yak genome and adaptation to life at high altitude. *Nat Genet*. 44:946–949.

Rambaut A. 2016. FigTree, v1.4.3. Available from: http://tree.bio.ed.ac.uk/software/figtree/.

Scheet P, Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet*. 78:629–644.

Schmutz SM, Berryere TG, Ellinwood NM, Kerns JA, Barsh GS. 2003. MC1R studies in dogs with melanistic mask or brindle patterns. *J Hered*. 94:69–73.

Schmutz SM, Dreger DL. 2013. Interaction of MC1R and PMEL alleles on solid coat colors in Highland cattle. *Anim Genet*. 44:9–13.

Society NYZ. 1931. *Thirty-fifth annual report of the New York zoological society*. New York: Clark and Fritts.

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res*. 13:2129–2141.

Whitacre L. 2014. Structural variation at the KIT locus is responsible for the piebald phenotype in Hereford and Simmental cattle. MS thesis, p. 70. University of Missouri, Columbia.

White WT, Phillips RW, Elting EC. 1946. Yaks and yak-cattle hybrids in Alaska. *J Hered*. 37:354–358.

Wiener G, Jianlin H, Ruijun L. 2003. *The Yak*. Bangkok (Thailand): Regional Office for Asia and the Pacific Food and Agriculture Organization of the United Nations.

Wu DD, Ding XD, Wang S, Wójcik JM, Zhang Y, Tokarska M, Li Y, Wang MS, Faruque O, Nielsen R, *et al.* 2018. Pervasive introgression facilitated domestication and adaptation in the Bos species complex. *Nat Ecol Evol*. 2:1139–1145.

Xi D, Liu Q, Huo Y, Sun Y, Leng J, Gou X, Mao H, Deng W. 2012a. Nucleotide diversity of the melanocortin 1 receptor gene (MC1R) in the gayal (Bos frontalis). *Mol Biol Rep*. 39:7293–7301.

Xi D, Wu M, Fan Y, Huo Y, Leng J, Gou X, Mao H, Deng W. 2012b. Isolation and characteristics of the melanocortin 1 receptor gene (MC1R) in the Chinese yakow (*Bos grunniens* × *Bos taurus*). *Gene*. 498:259–263.

Yan SQ, Hou JN, Bai CY, Jiang Y, Zhang XJ, Ren HL, Sun BX, Zhao ZH, Sun JH. 2014. A base substitution in the donor site of intron 12 of KIT gene is responsible for the dominant white coat colour of blue fox (*Alopex lagopus*). *Anim Genet*. 45:293–296.

Zhang RC. 2000. *Interspecies hybridization between yak, Bos taurus and Bos indicus and the reproduction of hybrids*. In: Zhao XX, Zhang RC, editors. Ithaca (NY): International Veterinary Information Service.

Zhang Y, Li Q, Ye S, Faruque MO, Yu Y, Sun D, Zhang S, Wang Y. 2014a. New variants in the melanocortin 1 receptor gene (MC1R) in Asian cattle. *Anim Genet*. 45:609–610.

Zhang MQ, Xu X, Luo SJ. 2014b. The genetics of brown coat color and white spotting in domestic yaks (*Bos grunniens*). *Anim Genet*. 45:652–659.