Dog color patterns explained by modular promoters of ancient canid origin

A full list of authors and affiliations appears at the end of the article.

Abstract

Distinctive color patterns in dogs are an integral component of canine diversity. Color pattern differences are thought to have arisen from mutation and artificial selection during and after domestication from wolves but important gaps remain in understanding how these patterns evolved and are genetically controlled. In other mammals, variation at the *ASIP* gene controls both the temporal and spatial distribution of yellow and black pigments. Here we identify independent regulatory modules for ventral and hair cycle *ASIP* expression, and we characterize their action and evolutionary origin. Structural variants define multiple alleles for each regulatory module and are combined in different ways to explain five distinctive dog color patterns. Phylogenetic analysis reveals that the haplotype combination for one of these patterns is shared with arctic white wolves and that its hair cycle-specific module likely originated from an extinct canid that diverged from grey wolves more than 2 million years before present. Natural selection for a lighter coat during the Pleistocene provided the genetic framework for widespread color variation in dogs and wolves.

A central aspect of the amazing morphologic diversity among domestic dogs are their colors and color patterns. In many mammals, specific color patterns arise through differential regulation of *Agouti (ASIP)*, which encodes a paracrine signaling molecule and antagonist of the melanocortin 1 receptor (MC1R) that causes hair follicle melanocytes to switch from making eumelanin (black or brown pigment) to pheomelanin (yellow to nearly white pigment) [1–4]. In laboratory mice, *Asip* expression is controlled by alternative promoters in specific body regions, and at specific times during hair growth, and gives rise to the light-bellied agouti phenotype, with ventral hair that is yellow and dorsal hair that contains a mixture of black and yellow pigment [4]. Genetic variation in *ASIP* affects color pattern in many mammals; however, in dogs, the situation is still unresolved, in large part due to the complexity of different pattern types, epistatic relationships with variants at other loci,

[*]Corresponding Author.
[α,&]These authors contributed equally to this work

and challenges in distinguishing whether genetic association of one or more variants truly represents causal variation or just close linkage[5].

Here we investigate non-coding variation in *ASIP* regulatory modules and their effect on patterning phenotypes in domestic dogs. We expand our analysis to include modern and ancient wild canids and uncover an evolutionary history in which natural selection during the Pleistocene provided a molecular substrate for color pattern diversity today.

Results

Expression of *ASIP* promotes pheomelanin synthesis, therefore *ASIP* alleles associated with a yellow color are dominant to those associated with a black color. Although dominant yellow (DY) is common in dogs from diverse geographic locations, the most common coat pattern of modern wolves is agouti (AG)[6], in which the dorsum has banded hairs and the ventrum is light. Three additional color patterns are recognizable, but all have been described historically by different, inconsistent and sometimes overlapping names that predate genomic analysis; we refer to these as shaded yellow (SY), black saddle (BS), and black back (BB) (Fig. 1, Supplementary Table 1).

We analyzed skin RNA-seq data available from dogs of dominant yellow and black back patterns, and identified three alternative untranslated first exons for dog *ASIP* (Fig. 2a, Extended Data Fig. 1, Supplementary Table 2). As described below, two of the three transcripts vary in abundance between dominant yellow and black back dogs, and the corresponding 5'-flanking promoters have sequence variation associated with dog pattern phenotypes. The 5'-flanking promoter regions for these two transcripts are orthologous to the ventral promoter (VP) and hair cycle promoter (HCP) in the laboratory mouse [4]; however, our genetic analyses (Fig. 2) reveal that the dog VP and HCP give rise to more complex patterns than their mouse counterparts. Transcripts associated with the third promoter, which lies ~16 kb upstream of the VP (Fig. 2b) did not vary in abundance in our dataset.

To better understand the relationship between promoter usage and pattern phenotypes, we inspected whole genome sequence data from 77 dog and wolf samples with known color patterns (Supplementary Table 3). We used dogs that were homozygous at the *ASIP* locus to infer two VP haplotypes and five HCP haplotypes, consisting of multiple structural variants that lie within 1.5 kb of each transcriptional start site. VP1 contains a SINE element in reverse orientation relative to the transcription of *ASIP* and an A-rich expansion not found in VP2 (Fig. 2b, left, Supplementary Table 1); the five HCP haplotypes differ according to the number and identity of SINE elements, all in the same orientation as *ASIP*, as well as additional insertions and deletions (Fig. 2b, right, Supplementary Table 1). All structural variants were precisely delineated with Sanger sequencing.

These results were extended by developing PCR-based genotyping assays for the VP and HCP structural variants, examining their association with different pattern phenotypes in 352 dogs from 34 breeds, and comparing these results to previously published variants (Table 1, Extended Data Fig. 2, Supplementary Table 1, and Supplementary Table 4–7).

As depicted in Fig. 2c and Table 1, diplotype combinations of VP1 or VP2 with HCP1, 2, 3, 4, or 5 are correlated perfectly with variation in *ASIP* pattern phenotype. For example, homozygotes for VP1-HCP1, VP2-HCP1, VP2-HCP2 are dominant yellow, shaded yellow and agouti, respectively (Supplementary Table 4–7). Black saddle and black back dogs differ in their VP configuration, but all carry HCP3, 4, and/or 5 in homozygous or compound heterozygous configurations. Because the level of *ASIP* activity is directly related to the amount of yellow pigment production, these genetic association results suggest that VP1 has greater activity than VP2, HCP1 has greater activity than HCP2, and HCP3, 4, and 5 all represent loss-of-function, since the HCP4 haplotype includes a large deletion of the hair cycle first exon (Fig. 2b), and fails to complement HCP3 or HCP5 (Fig. 3, Supplementary Table 6). Importantly, increased activity from the ventral promoter (VP1 vs. VP2) correlates with dorsal expansion of yellow pigment in black saddle compared to black back phenotypes (Fig. 1, 2c), which indicates that the VP and HCP haplotypes function separately from each other.

The relationship between structural variation that delineates the different VP and HCP haplotypes and *ASIP* transcriptional activity was explored more directly in RNA-seq data from biopsies of dorsal and ventral dog skin (Supplementary Table 8, Extended Data Fig. 1). Read counts from the RNA-seq data were consistent with expectations from the genetic association results: VP1 has greater transcriptional activity and is spatially broadened relative to VP2 (which is only expressed ventrally), HCP1 has greater transcriptional activity relative to HCP2, and no reads are detected from HCP3 or HCP4 (Fig. 2b, Extended Data Fig. 1). Taken together, these results provide a molecular explanation for *ASIP* pattern variation in dogs in which the VP and HCP function independently and for which structural variants in close proximity to VP and HCP modulate promoter activity.

Genetic relationships between variant *ASIP* regulatory modules were examined by comparing haplotypes in 18 homozygous dogs (for the structural variants at the VP and HCP and coding sequences) to those from 10 contemporary grey wolves (Fig. 4a, Supplementary Table 9). Overall, agouti dog haplotypes were similar to those from grey wolves. However, dominant yellow and, to a lesser extent, shaded yellow dog haplotypes were similar to those from arctic grey wolves from Ellesmere Island and Greenland, where all wolves are white (Fig. 4a–c). Notably, white coat color in wolves represents pale pheomelanin, as in Kermode bears or snowshoe hares [7,8]. In the 64 kb segment that contains the VP, HCP, and coding sequence, the arctic grey wolf haplotypes are identical except for one polymorphic site (Fig. 4a, chr24: 23,337,523), and are distinguished from dog dominant yellow haplotypes by only 6 SNVs (Supplementary Table 10). Taken together, these observations suggest a common origin of dominant yellow in dogs and white coat color in wolves without recent genetic exchange.

The evolutionary origin of *ASIP* haplotypes was explored further by constructing maximum likelihood phylogenetic trees for dogs, wolves, and 8 additional canid species (Supplementary Table 9). Based on differences in SNV frequency, the 48 kb VP segment was considered separately from the 16 kb HCP-exon 2/3/4 segment (see Supplementary text, Fig. 4a). In the VP tree, all dogs and grey wolves form a single clade, consistent with known species relationships [9]. However, in the HCP tree, the dominant yellow and shaded yellow

dogs lie in a separate clade together with arctic grey wolves; remarkably, this clade is basal to the golden jackal and distinct from other canid species (Fig. 4b, Extended Data Fig. 3, 4).

The pattern of derived allele sharing provides additional insight (Fig. 4d and Extended Data Fig. 5). As depicted in Fig. 2c and 4d, HCP2 is characterized by three small repeat elements that are shared by all canids and is therefore the ancestral form. In the branch leading to core wolf-like canids (golden jackal, coyote, Ethiopian wolf, and grey wolf), there are nine derived SNV alleles within the HCP2-exon 2/3/4 segment (Extended Data Fig. 5, Supplementary Table 11), four of which flank the repeat elements close to HCP2 (Fig. 4d, Extended Data Fig. 5). None of the nine derived alleles are present in the dominant yellow HCP1-exon 2/3/4 segment haplotype which also carries an additional SINE close to HCP1; therefore, this haplotype must have originated prior to the last common ancestor of golden jackals and other wolf-like canids >2 Mybp [10]. Although the 16 kb HCP1-exon 2/3/4 segment haplotype could have originated on a branch leading to the core wolf-like canids, it would have had to persist via incomplete lineage sorting and absence of recombination for more than 2 million years and through three speciation events (Supplementary text). A more likely scenario is that HCP1 represents a ghost lineage from an extinct canid (Fig. 4d, 5b) that was introduced by hybridization with grey wolves during the Pleistocene (see below), as has been suggested for an ancestor of the grey wolf and coyote [9], and in high altitude Tibetan and Himalayan wolves [11].

We expanded our analysis of VP and HCP haplotypes to a total of 45 North American and 23 Eurasian wolves. The VP1-HCP1 haplotype combination is found mostly in the North American Arctic in a distribution parallel to that of white coat color (Extended Data Fig. 6a) [12] and is not observed in Eurasia. We also identified an ancestral HCP1 haplotype variant, referred to hereafter as HCP1^A, that does not extend to exons 2/3/4 and lacks the 24 bp insertion found in arctic grey wolves and dominant yellow dogs (Fig. 4d, Extended Data Fig. 7). A haplotype combination similar to shaded yellow, VP2-HCP1^A, was observed in seven light-colored wolves from Tibet or Inner Mongolia, representative of a distinct, high-altitude grey wolf population that is notably lighter than other Eurasian populations (Fig. 4d, Extended Data Fig. 6b) [13].

Additional insight into the demographic history of these haplotypes emerges from the analysis of ancient dog (n=5) and grey wolf (n=2) WGS data, dated 4,000 – 35,000 ybp (Supplementary text and Supplementary Table 12), in which both forms of the VP (VP1 and VP2), and four forms of the HCP (HCP1^A, HCP1, HCP2, HCP4) were observed in various combinations (Fig. 5a, Extended Data Fig. 7). Ancient wolves from the Lake Taimyr and Yana River areas of Arctic Siberia had at least one HCP1 haplotype, while ancient dogs from central Europe, Ireland, and Siberia carried HCP1^A, HCP1, and HCP4, respectively (Supplementary Table 12). Thus, diversity in *ASIP* regulatory sequences responsible for color variation today was apparent by 35,000 ybp in ancient wolves and by 9,500 ybp in ancient dogs.

Together with our phylogenetic results, comparative analysis of wolf and dog *ASIP* haplotypes suggests an evolutionary history in which multiple derivative haplotypes and associated color patterns arose by recombination and mutation from two ancestral

configurations corresponding to a white wolf (VP1-HCP1) and a grey wolf (VP2-HCP2), both present in the late Pleistocene (Fig. 5a, Extended Data Fig. 7). The distribution of derivative haplotypes explains color pattern diversity not only in dogs but also in modern wolf populations across the Holarctic, including white wolves in the North American Arctic (VP1-HCP1) and yellow wolves in the Tibetan highlands (VP2-HCP1[A]), and is consistent with natural selection for light coat color. A likely timeline for the origin of modules driving high levels of *ASIP* expression is depicted in Fig. 5b and indicates a dual origin. The HCP1 haplotype represents introgression into Pleistocene grey wolves from an extinct canid lineage that diverged from grey wolves more than 2 Mybp. This introgression as well as the mutation from VP2 to VP1 occurred prior to 33,500 ybp, based on direct observation from an ancient wolf sample (Fig. 5a).

Discussion

A relationship between *ASIP* and dog color pattern was recognized more than a century ago by Sewall Wright [14], and explored in depth by the work of C. C. Little in the decades that followed [15]. Previous studies have reported molecular variation in or around the *ASIP* region associated with some dog color patterns, including a 16 bp non-coding duplication associated with black saddle [16], a SINE insertion associated with black back and black saddle [16], and missense variants A82S and R83H associated with dominant yellow [17] (Fig. 2a, Extended Data Fig. 2). As shown here, availability of a broader dataset indicates that these previously reported associations represent linkage disequilibrium and/or breed structure rather than causal variation (Supplementary Table 7). Instead, our WGS-based comprehensive annotation of the region together with RNA expression data reveals a series of structural variants that define distinct haplotypes for each of two promoters that, in combination, explain five different pattern types (Fig. 2c).

In dogs, the key differences between VP1 and VP2 are a SINE element and a small insertion; similarly, the key differences between HCP1 and HCP2 are multiple SINE elements and a small insertion (Fig. 2b). In each case, we do not yet know if the transcriptional differences (VP1 > VP2, HCP1 > HCP2) are caused by the SINE element, the small insertion, or both. We note, however, that modularity of *ASIP* regulatory variation is a general theme in vertebrates, with non-coding changes driving adaptation in natural populations of deer mice [3], mountain hares [18], snowshoe hares [8], and several species of parulid warblers [19-22]. Likewise, artificial selection in goats [23], domestic rabbits [24,25] and laboratory mice [26] is associated with structural variation in *ASIP* regulatory regions that may lead to acquisition of novel promoters that modulate region-specific expression of *ASIP*.

ASIP color pattern diversification was likely an early event during dog domestication, since our analysis of ancient DNA data reveals several different VP and HCP haplotypes in Eurasia by 4,800 ybp. This is consistent with the wide distribution of dominant yellow across modern dog breeds from diverse locations, as well as the dingo (Supplementary Table 9), a feral domesticate, frequently dominant yellow, introduced to Australia at least 3,500 ybp [27]. Of particular interest is the Zhokov island dog from Siberia [28,29] Based on a haplotype combination of VP2-HCP4, this sled dog that lived 9,500 years ago exhibited a

black back color pattern, allowing it to be easily distinguished from white colored wolves in an arctic environment.

In wolves, natural selection for VP1 and HCP1 are a likely consequence of Pleistocene adaptation to arctic environments and genetic exchange in glacial refugia, driven by canid and megafaunal dispersal during interglacial periods. Modern grey wolves are thought to have arisen from a single source ~25,000 ybp close to the last glacial maximum [30,31]; during the North American glacial retreat that followed, the VP1-HCP1 haplotype combination was selected for in today's white-colored arctic wolves. Our results show how introgression, demographic history, and the genetic legacy of extinct canids played key roles in shaping diversity in dogs and modern grey wolves.

Methods

Ethics Statement

All animal experiments were done in accordance with the local regulations. Experiments were approved by the "Cantonal Committee For Animal Experiments" (Canton of Bern; permits 48/13, 75/16 and 71/19).

Skin biopsies and total RNA extraction

Skin biopsies (6 mm punch) were recovered from three dogs (a black back Miniature Pinscher and a dominant yellow Border Terrier and Irish Terrier) at necropsy and/or surgery for reasons unrelated to this study. Biopsies were recovered from the ventral abdomen and dorsal thorax, and are not matched for age or hair growth cycle. The biopsies were immediately put in RNAlater (Qiagen) for at least 24 h and then frozen at –20°C. Prior to RNA extraction, the skin biopsies were homogenized mechanically with the TissueLyser II device from Qiagen. Total RNA was extracted from the homogenized tissue using the RNeasy Fibrous Tissue Mini Kit (Qiagen) according to the manufacturer's instructions. RNA quality was assessed with a FragmentAnalyzer (Agilent) and the concentration was measured using a Qubit Fluorometer (ThermoFisher Scientific).

Whole transcriptome sequencing (RNA-seq)

From each sample, 1 μg of high quality total RNA (RIN >9) was used for library preparation with the Illumina TruSeq Stranded mRNA kit. The libraries were individually barcoded and pooled and sequenced on an S1 flow cell with 2x50 bp paired-end sequencing using an Illumina NovaSeq 6000 instrument. On average, 31.5 million paired-end reads per sample were collected. One publicly available Beagle sample was used (SRX1884098). All accession numbers and descriptive read statistics are given in Supplementary Table 8. All reads that passed quality control were mapped to the CanFam3.1 reference genome assembly using STAR aligner (version 2.6.0c) [34]

Transcript coordinates

The STAR-aligned bam files were visualized in the IGV browser [35]. Three different alternate untranslated first exons with splice junctions to the coding exons of *ASIP* were defined based on the visualizations of the read alignments in IGV based on the RNA-seq data

just described. These exact transcripts have not been documented in NCBI; however, the three transcripts of NCBI annotation release 105 are virtually identical except minor differences regarding the length of the 5'-UTRs (XM_014106843.2, transcription start sire (TSS) 22 nucelotides upstream compared to our annotation; NM_001007263.1, VP1-TSS98 bp downstream of our annotation; XM_022408819.1 HCP-TSS 36 bp downstream of our annotation). Our visually curated gene models are given in Supplementary Table 2.

Identification of genomic variants

WGS data from 71 dogs and 6 wolves was used for variant discovery (Supplementary Table 3). They included 15 agouti dogs and wolves, 25 black back dogs, 11 black saddle dogs, 14 dominant yellow dogs, 11 shaded yellow dogs and one white wolf. The genomes were either publicly available or sequenced as part of related projects in our group [36]. SNVs and small indels were called as described [36]. The IGV software [35] was used for visual inspection of the three promoter regions based on the transcripts identified in the RNA sequencing data. Structural variants were identified and association with coat color phenotypes was verified by visual inspection in IGV. The pattern of CNV at the third promoter did not associate with the coat patterns as defined in Fig. 1.

DNA samples for Sanger sequencing and genotyping

Samples for variant discovery included two dogs from each color phenotype and are designated with asterisks in Supplementary Table 3. Samples from dogs listed in Supplementary Table 4 were used for genotyping. The coat color phenotype of all animals (Supplementary Tables 3 and 4) was assigned based on breed-specific coat color standards or photographs or owner reporting. Genomic DNA was isolated from EDTA blood samples using the Maxwell RSC Whole Blood DNA kit (Promega).

Sequencing of promoter regions

Sanger sequencing of PCR amplicons was carried out to validate and characterize structural variants in the promoter regions at the sequence level. All primer sequences and polymerases used are listed in Supplementary Table 5. PCR products amplified using LA Taq polymerase (Takara) or Multiplex PCR Kit (Qiagen) were directly sequenced on an ABI 3730 capillary sequencer after treatment with exonuclease I and shrimp alkaline phosphatase. Sequence data were analyzed with Sequencher 5.1 (GeneCodes). Interspersed repeat insertions were classified with the RepeatMasker program [37]. Multiple copies of SINE elements from the same and different families were resolved this way. The CanFam3.1 reference genome assembly is derived from the Boxer Tasha, a dominant yellow dog, and represents a DY haplotype, VP1-HCP1, of the *ASIP* gene. Descriptions of the promoter variants and Genbank accession numbers for HCP2-5 are in Supplementary Table 1. Supplementary Table 1 lists the 7 combinations of VP and HCP regulatory modules observed in dogs. As HCP3, HCP4, and HCP5 all represent functionally equivalent loss-of-function alleles, the 7 listed combinations correspond to only 5 distinct phenotypes.

Genotyping assays

Five PCR-assays (ventral promoter assays 1, 2; hair cycle promoter assays 1, 2, 3) were required to unambiguously determine the VP and HCP haplotypes (Supplementary Table 5). The previously reported SINE insertion [32] was genotyped by fragment size analysis on an ABI 3730 capillary sequencer and analyzed with the GeneMapper 4.0 software (Applied Biosystems). The previously reported *ASIP* coding variants [17] were genotyped by Sanger sequencing of PCR products. The previously reported *RALY* intronic duplication [16] was genotyped by size differentiation of PCR products on a Fragment Analyzer (Agilent). Another primer pair was used for the amplification of the entire HCP (Supplementary Table 5). Genotyping results for all samples are shown in Supplementary Table 4. There is a perfect genotype-phenotype association in 352 dogs (see Table 1 and Supplementary Table 7). In the remaining 14 dogs, the presence of a eumelanistic mask prevented the reliable phenotypic differentiation of dominant yellow and shaded yellow dogs. Breeds and the different promoter haplotype combinations identified within each breed are indicated in Supplementary Table 6. In a few dogs that were heterozygous at both VP and HCP, the phasing of the VP and HCP haplotype combinations was performed based on haplotype frequency within the same breed as noted. A family of Chinooks were used to determine the segregation of extended haplotypes and the phenotypic equivalency of HCP3 and HCP5 (Figure 3). Summary of genotyping results and exclusion of previously associated variants is shown in Table 1 and Supplementary Table 7. This table lists the genotype-phenotype association in aggregated form. The table also contains the genotypes for variants that were previously reported to be associated with pattern phenotypes [16,17,32]

Comparison of promoter haplotype effects on transcripts

Transcript data was generated from a second set of samples. Sample descriptions and colors are shown in Supplementary Table 8 for all RNA experiments. Skin samples were collected from a male Swedish Elkhound (agouti), female German Pinscher (dominant yellow) and male Rottweiler (black back) after euthanasia that was conducted due to behavioral or health problems not related to skin. Samples were collected in RNAlater Stabilization Solution and stored at –80°C. RNA was extracted using the RNeasy Fibrous Tissue Mini Kit (Qiagen) according to the manufacturer's instructions. Integrity of RNA was evaluated with Agilent 2100 Bioanalyzer or TapeStation system (Agilent) and concentration measured with DeNovix DS-11 Spectrophotometer (DeNovix Inc.). The libraries for STRT (Single cell reverse tagged) RNA sequencing were prepared using the STRT method with unique molecular identifiers [38] and modifications including longer UMI's of 8 bp, addition of spike-in ERCC control RNA for normalization of expression, and the Globin lock method [39] with LNA-primers for the canine alpha- and betaglobin genes. The libraries were sequenced with an Illumina NextSeq 500. Reads were mapped to the CanFam3.1 genome build using HISAT1 mapper version 2.1.0 [40]

The alignment-free quantification method Kallisto (version 0.46.0) [41] was used to estimate the abundance and quantified as transcripts per million mapped reads (TPM) data based on an index built from CanFam3.1 Ensembl transcriptome (release 99). The curated *ASIP* transcript isoform models based upon alignment visualizations in the IGV browser [35] were

also included in the transcriptome. Results based on genotype of the promoter haplotypes are displayed in Extended Data Fig. 1 as TPM.

Haplotype construction

Haplotypes were constructed from two publicly available VCF files PRJEB32865 and PRJNA448733. The VCFs for selected dogs were merged using BCFtools merge tool (http://samtools.github.io/bcftools/) with the parameter --missing-to-ref, which assumed genotypes at missing sites are homozygous reference type 0/0. Only dogs homozygous for *ASIP* haplotypes (VP, HCP and coding exons) were used to visualize haplotypes (Supplementary Table 3). SNVs that had 100% call rate in these samples were color coded and displayed relative to the genome assembly and previously associated variants (Extended Data Fig. 2).

ASIP phylogenetic analysis in canids

Illumina whole genome sequence for 36 canids, including seven extant species and the dog, were downloaded from the NCBI short read archive as aligned (bam format) or unaligned (fastq format reads (Supplementary Table 9). Fastq data were aligned to the dog genome (CanFam3.1) using BWA (v.0.7.17) [42] after trimming with Trim Galore (v.0.6.4). SNVs within a 110 kb interval (chr24:23,300,000-23,410,000), which includes the *ASIP* transcriptional unit and regulatory sequences, were identified with Platypus (v.0.8.1) [43] and filtered with VCFtools (v.0.1.15) [44] to include 2008 biallelic SNVs. Phasing was inferred with BEAGLE (v.4.1) [45].

For phylogenetic analysis, the *ASIP* interval was partitioned in two regions, based on dog SNV density (Fig. 4a) and *ASIP* gene structure: a 48 kb region including the ventral first exon, extending to but excluding the hair cycle first exon (chr24:23,330,000-23,378,000), and a 16 kb region including the hair cycle first exon, extending to and including *ASIP* coding exons 2-4 (chr24:23,378,001-23,394,000). Consensus sequences of equal length were constructed for each inferred canid haplotype using BCFtools (v.1.9). Phylogenies were inferred using Maximum Likelihood method and Tamura-Nei model with 250 bootstrap replications, implemented in MEGAX [46,47], and including 34 canids (Fig. 4b, Extended Data Fig. 3, 4). For 34 of 36 individuals, consensus haplotype pairs were adjacent to each other or, in the case of a few wolf/dog haplotypes, were positioned in neighboring branches with weak bootstrap support. The exceptions were the African golden wolf, a species derived by recent hybridization of the grey wolf and Ethiopian wolf [9], and an eastern grey wolf from the Great Lakes region, which was also reported to have recent admixture with the coyote [48]. The African golden wolf and the eastern grey wolf were removed from the alignments, and a single haplotype for each individual was selected arbitrarily for tree building and display.

Haplotype analysis of *ASIP* locus in ancient dogs and wolves

Whole genome sequencing data from several recent studies [9,13,29,49–53], including five ancient dogs, two ancient grey wolves, and 68 modern grey wolves (Supplementary Table 12) were downloaded as aligned (bam format) or unaligned (fastq format) reads. Fastq data was aligned to the dog genome (CanFam3.1) using BWA-MEM (v.0.7.17) [42], after trimming with Trim Galore (v.0.6.4). Coverage depth for each sample ranged from 1

-78x (Supplementary Table 12). Genotypes at five structural variants and six SNVs were determined by visual inspection using the IGV browser (Supplementary Table 1 and 12). Variants in or near the ventral promoter (n=2), the hair cycle promoter (n=6), and the coding exons (n=3) distinguished ventral and hair cycle promoter haplotypes (Supplementary Table 12). SNV genotypes were determined by allele counts; structural variants were genotyped by split reads at breakpoint junctions. The base maps used for plotting the geographic distribution of haplotypes (Fig. 5, Supplementary Fig. 6) were generated in R (v. 4.0.3) with 'maps' and 'ggplot2' packages.

For 67 of 75 wolves (or ancient dogs), the phase of ventral and hair cycle promoter haplotypes was unambiguous. Seven wolves and one ancient dog were heterozygous with respect to both the ventral and hair cycle promoter haplotypes, and for these samples, haplotype phase was inferred based on the linkage disequilibrium in the 67 unambiguous individuals.
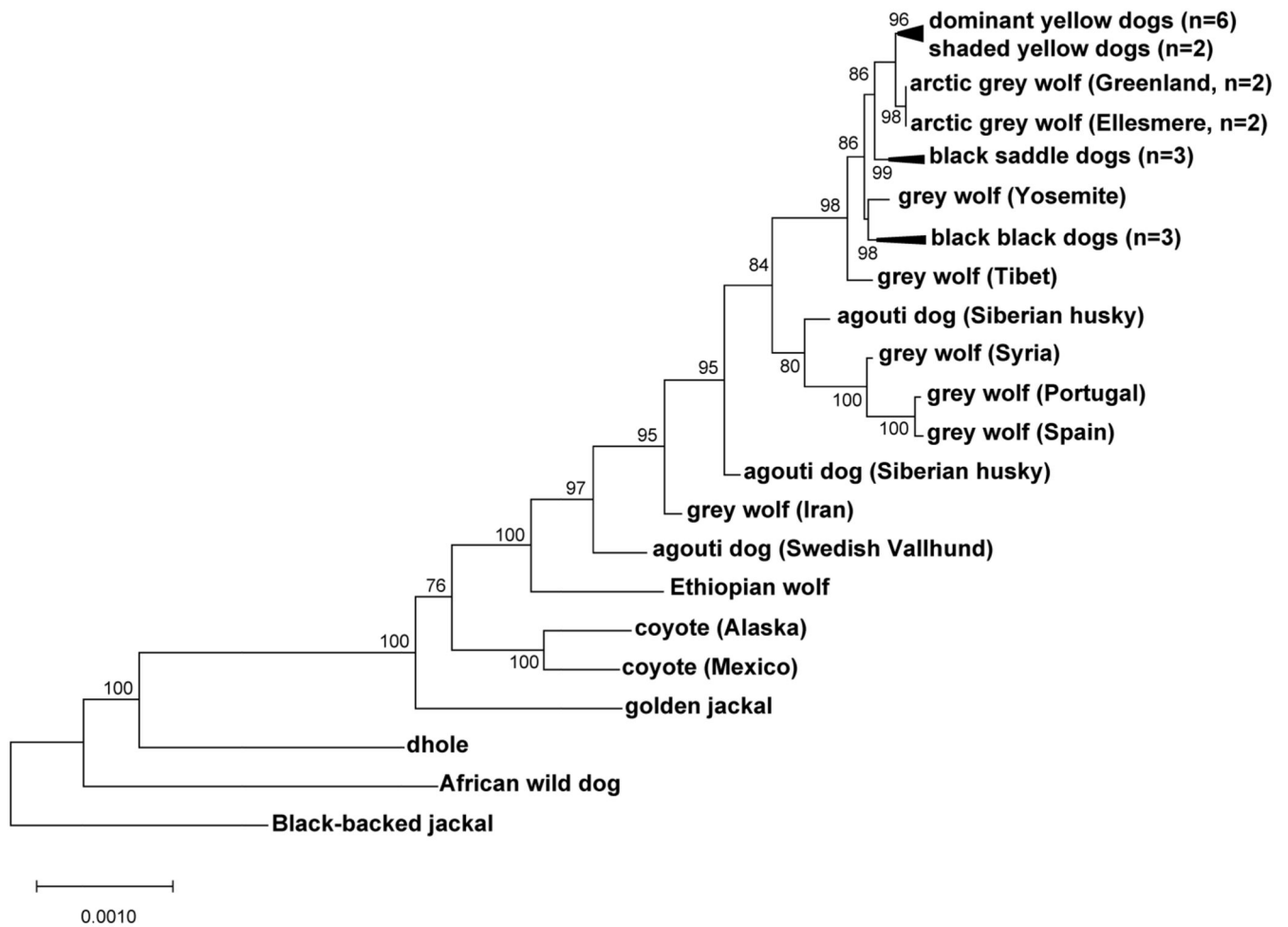
Extended Data
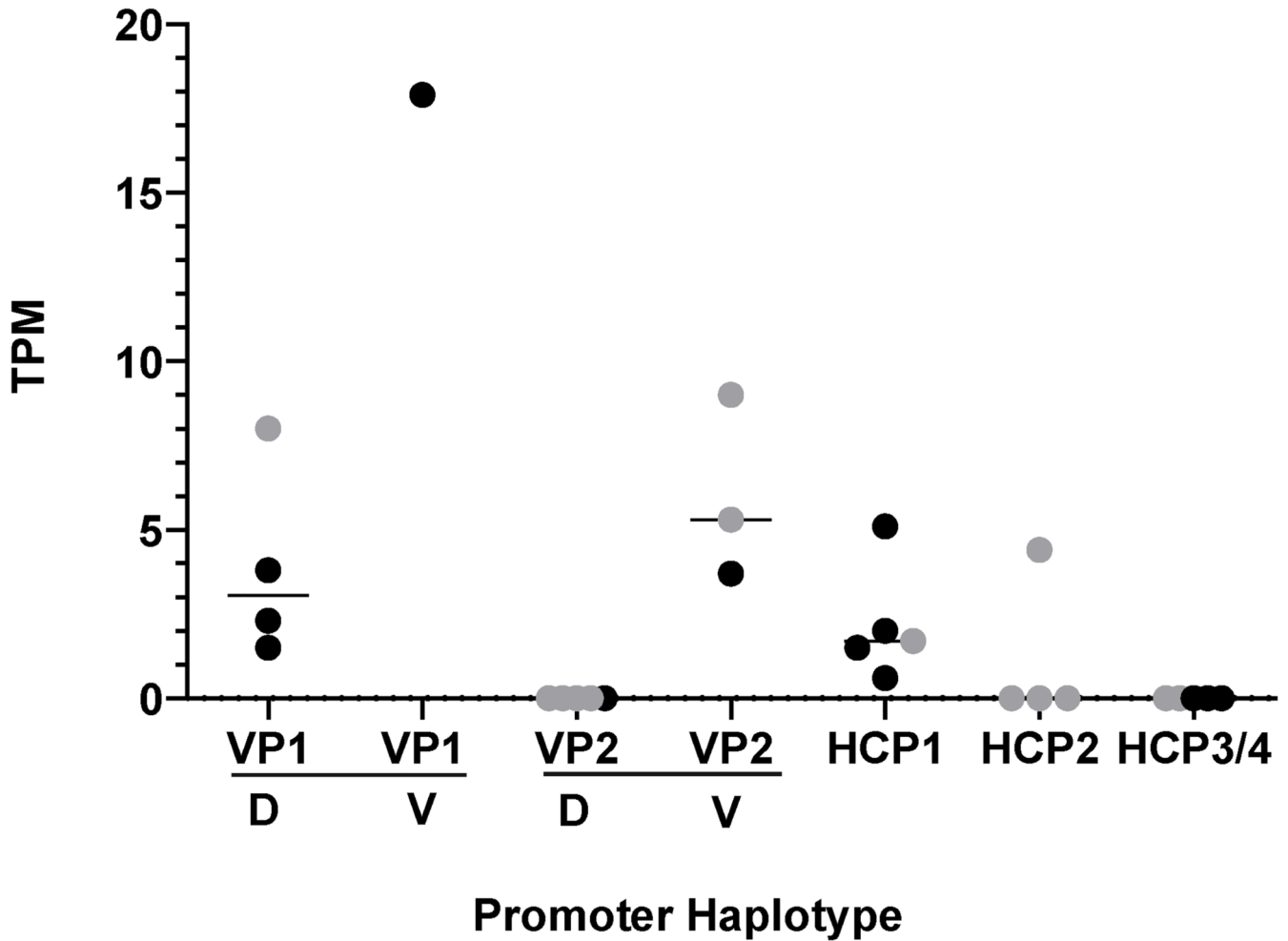
Extended Data Fig. 1. Relative transcription of promoter variants.

RNA-seq counts (transcripts per million reads, TPM) from dorsal (D) and ventral (V) regions of two dominant yellow, three black back and four agouti dogs obtained from skin biopsies as described in Methods and in Supplementary Table 8. HCP samples also included a black saddle dog. Black dots are from RNA-seq data and grey dots are from STRT RNA-seq data.
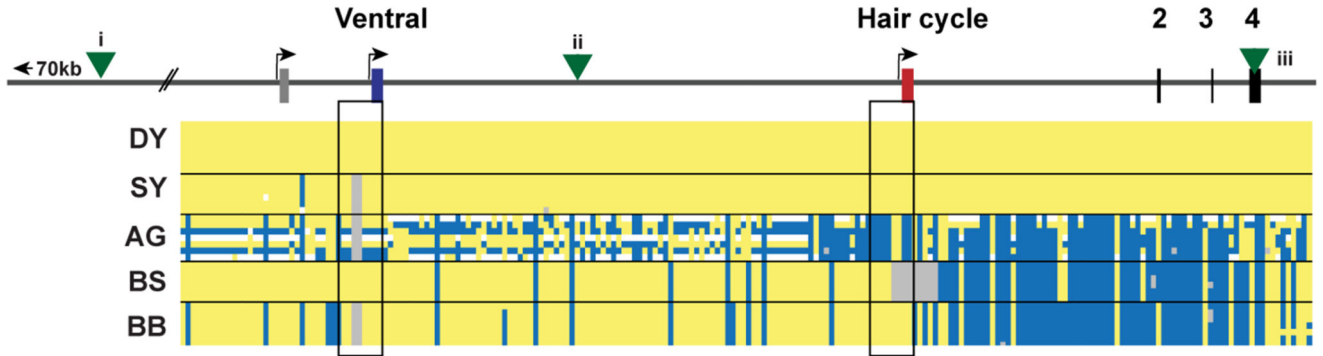

Extended Data Fig. 2. Dog haplotypes across the ASIP locus with comparison to previously associated variants for color patterns.

Dog coat pattern phenotypes are listed on the left. The genomic organization of the ASIP gene with its alternative promoters is illustrated at the top. Yellow indicates a homozygous match to the reference genome, grey deleted, white heterozygous and blue homozygous alternate allele. The black rectangles highlight the promoter regions. Green triangles represent the location of variants that were previously identified to distinguish different alleles for coat color patterns: (i) The previously identified intronic duplication, "RALY dup", associated with BS vs. BB haplotypes in some breeds, lies 86 kb to the left of the VP, but recombinants (Supplementary Table 7) exclude a causal role for ASIP pattern variation[16]. Similarly, (ii), a SINE insertion associated with BB and BS haplotypes in some breeds, and, (iii)[32], missense variants in exon 4 associated with DY haplotypes[17], are also excluded from a causal role in ASIP pattern variation by rare recombinants (Supplementary Table 7). In the samples presented here, the dominant yellow haplotype extends through the coding sequence where the missense variants associated with this haplotype were previously identified10. The results shown here will allow more accurate genetic testing in the future. Samples used are listed in Supplementary Table 3. Raw genotyping results are in Supplementary Table 4 and summary results comparing previously identified variants are in Table 1 and Supplementary Table 7.
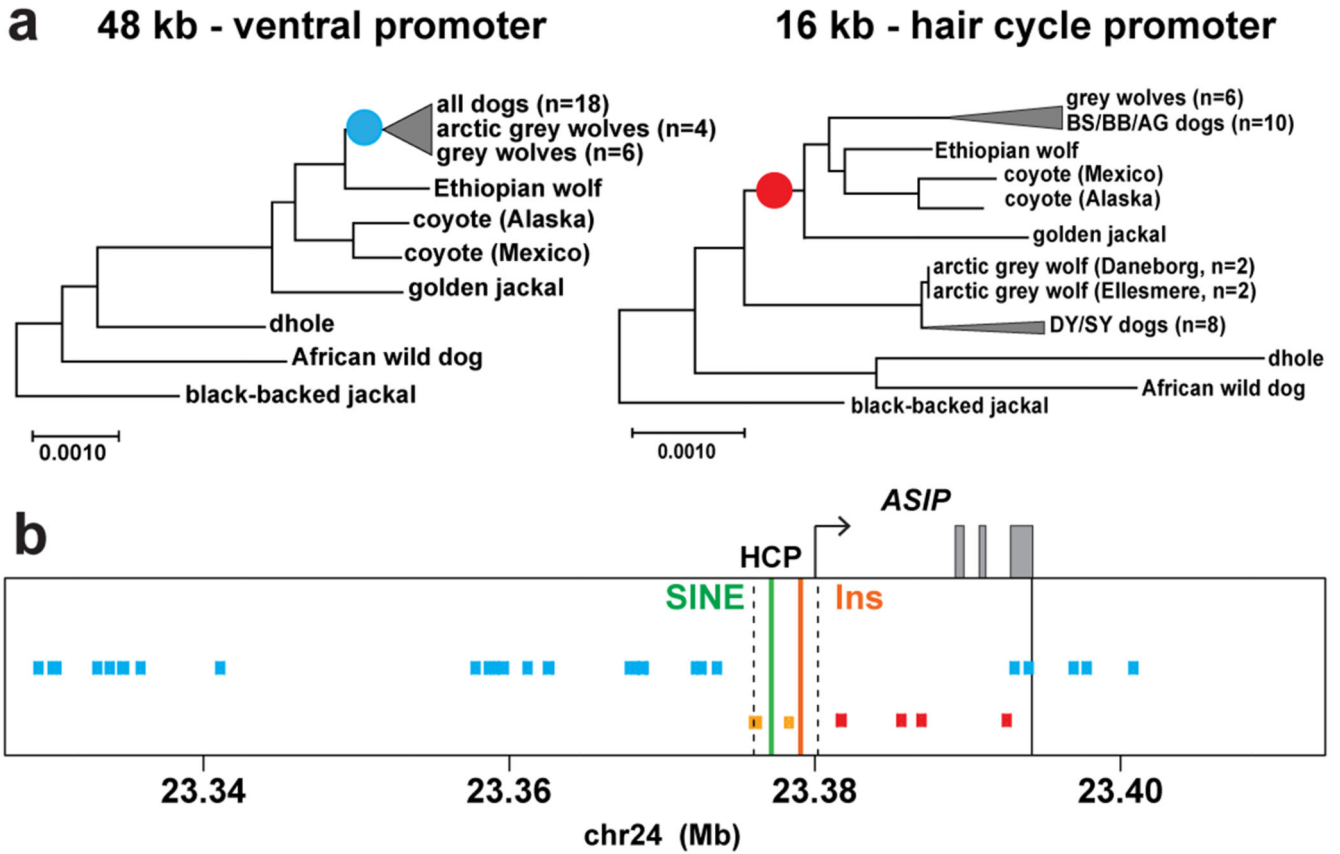
Extended Data Fig. 3. Expanded canid phylogenetic tree inferred from 48 kb region including the ventral promoter.

An expanded version of the maximum likelihood tree shown in Fig. 4b, with 34 canids, representing 7 of 9 extant species.
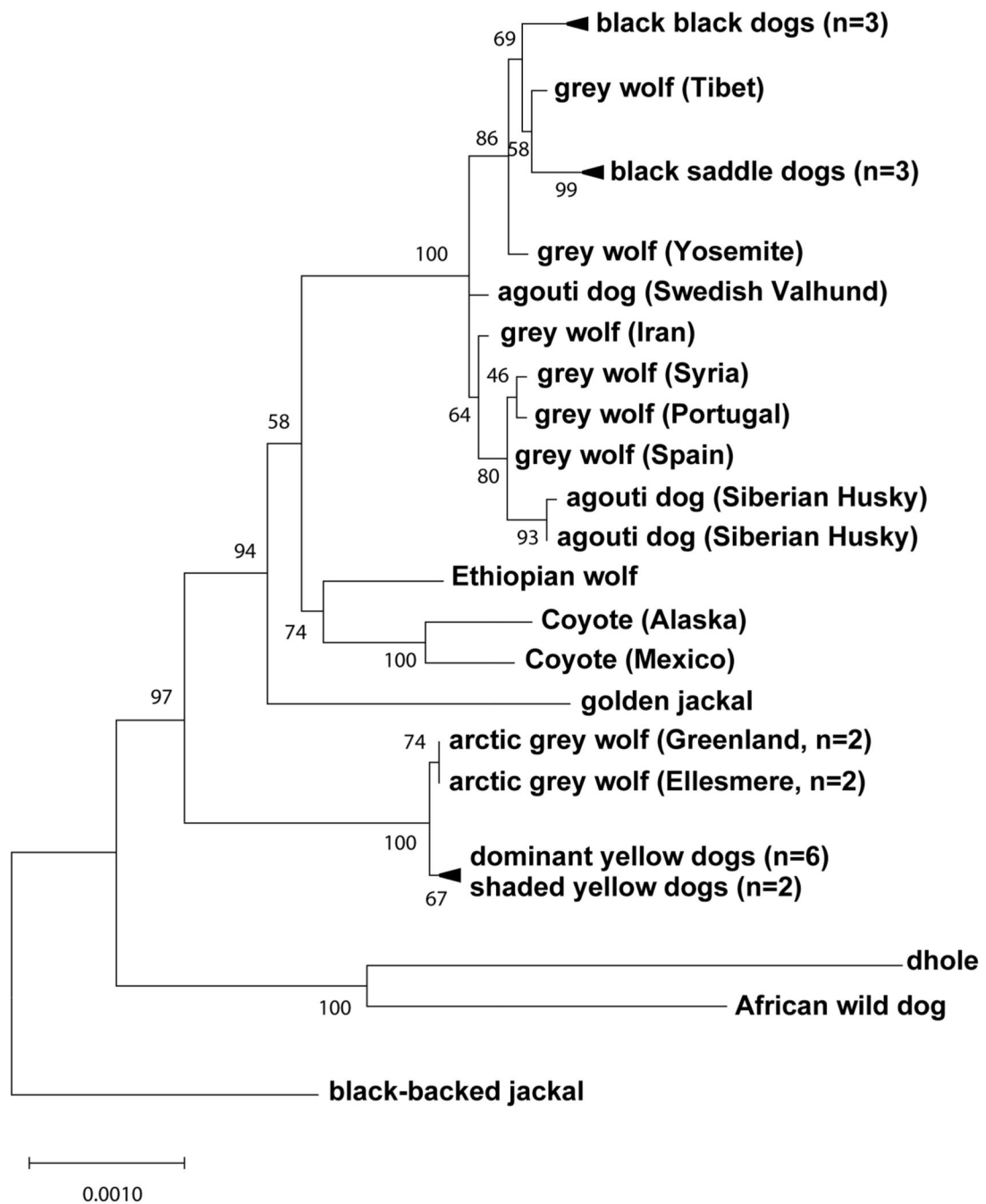
Extended Data Fig. 4. Expanded canid phylogenetic tree inferred from 16 kb region within and downstream of the hair cycle promoter.
An expanded version of the maximum likelihood tree shown in Fig. 4b, with 34 canids, representing 7 of 9 extant species.

Extended Data Fig. 5. Genomic distribution of derived substitutions across the ASIP locus.
(a) Canid phylogenies for the ventral (48 kb) and hair cycle (16 kb) promoter regions, with relevant internal branches marked by the occurrence of derived variants plotted in (b). (b) Derived substitutions shared by grey wolf and dogs (cyan). Ancestral alleles on DY/arctic wolf haplotypes (red) or BB and DY/arctic wolf haplotypes (orange) that correspond to derived substitutions among the core group of wolf-like canids (Supplementary Table 12). The broken lines demarcate the HCP region (chr24:23,375,800-23,380,000). The solid line signifies the downstream boundary for phylogenetic analysis. The solid green and orange lines indicate the positions of the SINE and 24 bp insertion, respectively, associated with the DY/arctic wolf haplotype.
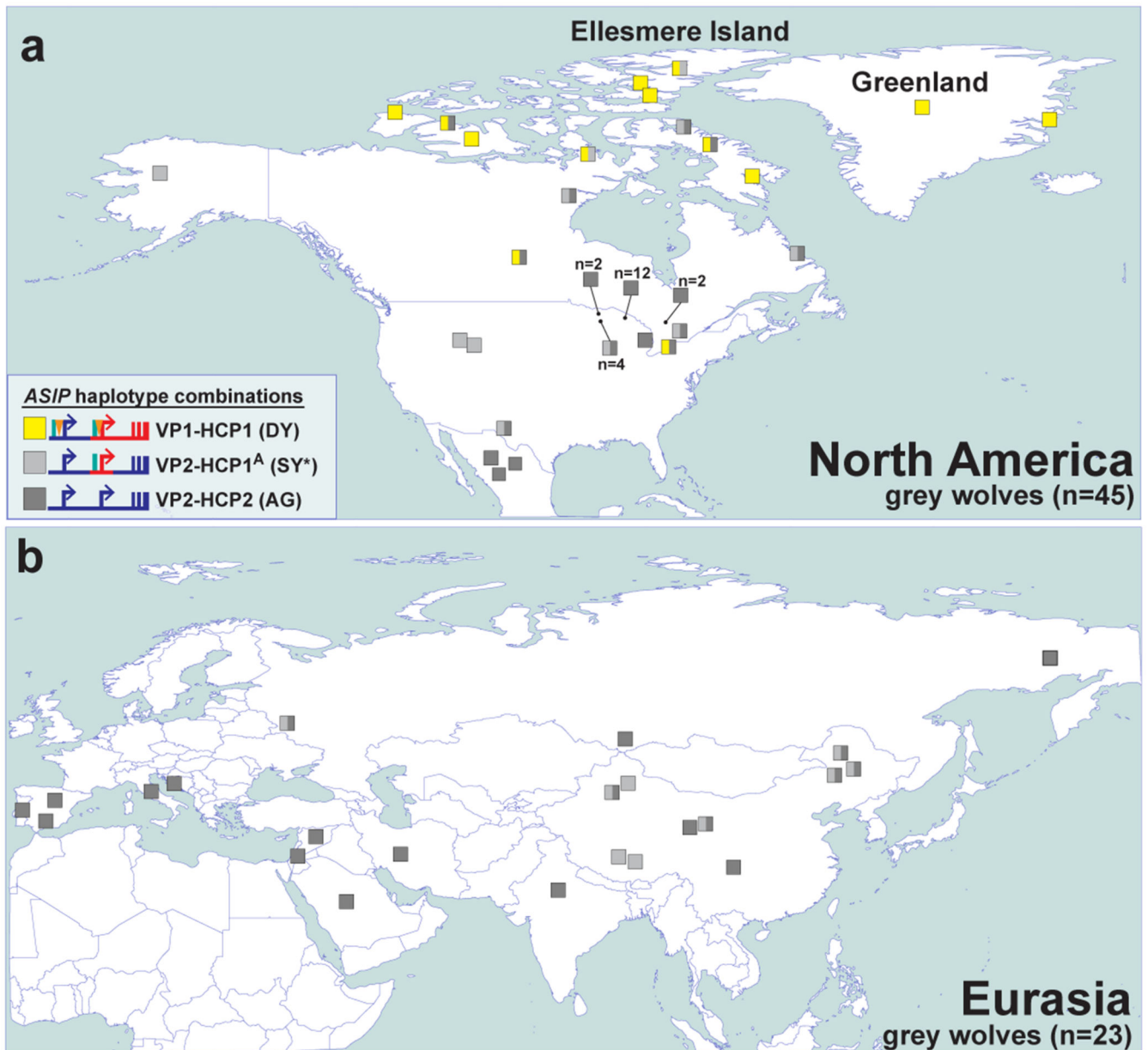
Extended Data Fig. 6. The distribution of ASIP haplotypes in modern grey wolves.
Modern grey wolves (squares) from (a) North America (n=45) or (b) Eurasia (n=23) were genotyped for 5 structural variants and 6 SNVs using whole genome sequencing data. Wolves are colored by inferred VP and HCP haplotypes, as indicated in the figure legend and in Supplementary Table 11. The asterisk indicates an SY-like haplotype without the HCP1 insertion.
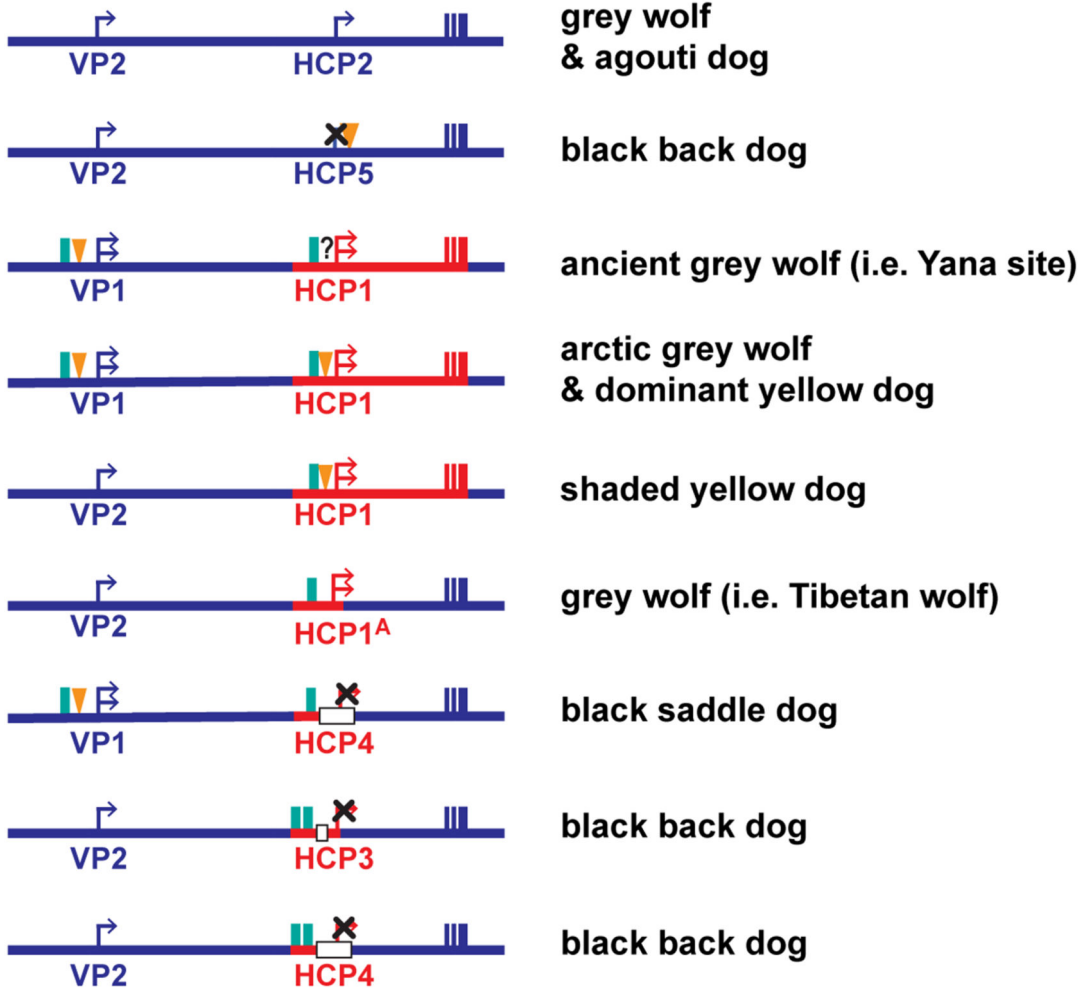
Inferred ancestral *ASIP* haplotypes

Observed *ASIP* haplotypes

Extended Data Fig. 7. Evolutionary diversification of ASIP haplotypes observed in grey wolves and dogs.

The color (red or blue) of ASIP haplotype segments indicates ancestral species of origin, inferred from phylogenetic analysis (Fig. 4b, Extended Data Fig. 3, 4). Relevant structural variants near the ventral (VP) and hair cycle (HCP) promoters are depicted as yellow triangles (polynucleotide expansions), green bars (SINE insertions), and white bars (deletions). Modified promoter activity is indicated by an X mark (no activity) or an additional arrow (elevated expression), based on RNA Seq (Extended Data Fig. 1) and/or inference from coat color (Fig. 1, 2c, 4c).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Danika L. Bannasch[*,α,1,2], Christopher B. Kaelin[α,3,4], Anna Letko[2,5], Robert Loechel[6], Petra Hug[2,5], Vidhya Jagannathan[2,5], Jan Henkel[2,5], Petra Roosje[5,7], Marjo K. Hytönen[8,9,10], Hannes Lohi[8,9,10], Meharji Arumilli[8,9,10], DoGA consortium, Hannes Lohi[1,2,3], Juha Kere[3,4,5], Carsten Daub[6,7], Marjo Hytönen[1,2,3], César L. Araujo[1,2,3], Ileana B. Quintero[1,2,3], Kaisa Kyöstilä[1,2,3], Maria Kaukonen[1,2,3], Meharji Arumilli[1,2,3], Milla Salonen[1,2,3], Riika Sarviaho[1,2,3], Julia Niskanen[1,2,3], Sruthi Hundi[1,2,3], Jenni Puurunen[1,2,3], Sini Sulkama[1,2,3], Sini Karjalainen[1,2,3], Antti Sukura[8], Pernilla Syrjä[8], Niina Airas[8], Henna Pekkarinen[8], Ilona Kareinen[8], Anna Knuuttila[8], Heli Nordgren[8], Karoliina Hagner[8], Tarja Pääkkönen[9], Antti Iivanainen[1], Kaarel Krjutskov[4], Sini Ezer[3,5], Auli Saarinen[3,5], Shintaro Katayama[3,4], Masahito Yoshihara[4], Auli Saarinen[3,5], Abdul Kadir Mukarram[6,7], Matthias Hörtenhuber[6,7], Rasha Fahad Aljelaify[6,7], Fiona Ross[6,7], Amitha Raman[6,7], Irene Stevens[6,7], Oleg Gusev[10], Danika Bannasch[11], Jeffrey J. Schoenebeck[12], Katie M. Minor[11], James R. Mickelson[11], Cord Drögemüller[2,5], Gregory S. Barsh[&,3,4], Tosso Leeb[&,2,5]

Affiliations

[1]Department of Population Health and Reproduction, School of Veterinary Medicine, University of California Davis, Davis, CA 95616 USA

[2]Institute of Genetics, Vetsuisse Faculty, University of Bern, 3001, Bern, Switzerland

[3]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

[4]Department of Genetics, Stanford University, Stanford, CA, USA

[5]Dermfocus, University of Bern, 3001 Bern, Switzerland

[6]VetGen, Ann Arbor, MI, 48108, USA

[7]Division of Clinical Dermatology, Department of Clinical Veterinary Medicine, Vetsuisse Faculty, University of Bern, 3001 Bern, Switzerland

[8]Department of Veterinary Biosciences, University of Helsinki, 00014 Helsinki, Finland

[9]Department of Medical and Clinical Genetics, University of Helsinki, 00014 Helsinki, Finland

[10]Folkhälsan Research Center, 00290 Helsinki, Finland

[11]Department of Veterinary and Biomedical Sciences, University of Minnesota, Saint Paul, MN 55108, USA

[1]Department of Veterinary Biosciences, University of Helsinki, 00014 Helsinki, Finland

[2]Department of Medical and Clinical Genetics, University of Helsinki, 00014 Helsinki, Finland

[3]Folkhälsan Research Center, 00290 Helsinki, Finland

[4]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden

[5]Stem Cells and Metabolism Research Program, University of Helsinki, Helsinki, Finland

[6]Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden

[7]Science for Life Laboratory, Karolinska Institutet, Stockholm, Sweden

[8]Department of Veterinary Pathology, Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland

[9]Department of Equine and Small Animal Medicine, University of Helsinki, Helsinki, Finland

[10]KFU-RIKEN "Translational genomics" Unit, RIKEN, Yokohama, 230-0045, Japan

[11]Department of Population Health and Reproduction, School of Veterinary Medicine, University of California Davis, US

[12]Roslin Institute, University of Edinburgh, Scotland

Acknowledgements

Data Availability Statement

All data generated or analyzed during this study are included in this published article (and its supplementary information files). GenBank accession numbers for promoter sequence variants are MT319114.1, MT319115.1, MT319116.1, MT319117.1.

References

1. Barsh G, Gunn T, He L, Schlossman S, Duke-Cohan J. Biochemical and genetic studies of pigment-type switching. Pigment Cell Res. 2000; 13 (Suppl 8) :48–53. DOI: 10.1034/j.1600-0749.13.s8.10.x [PubMed: 11041357]

2. Caro T, Mallarino R. Coloration in Mammals. Trends Ecol Evol. 2020; 35 :357–366. DOI: 10.1016/j.tree.2019.12.008 [PubMed: 31980234]

3. Linnen CR, et al. Adaptive evolution of multiple traits through multiple mutations at a single gene. Science. 2013; 339 :1312–1316. DOI: 10.1126/science.1233213 [PubMed: 23493712]

4. Vrieling H, Duhl DM, Millar SE, Miller KA, Barsh GS. Differences in dorsal and ventral pigmentation result from regional expression of the mouse agouti gene. Proc Natl Acad Sci U S A. 1994; 91 :5667–5671. DOI: 10.1073/pnas.91.12.5667 [PubMed: 8202545]

5. Dreger DL, et al. Atypical Genotypes for Canine Agouti Signaling Protein Suggest Novel Chromosomal Rearrangement. Genes (Basel). 2020; 11 doi: 10.3390/genes11070739

6. Kaelin CB, Barsh GS. Genetics of pigmentation in dogs and cats. Annu Rev Anim Biosci. 2013; 1 :125–156. DOI: 10.1146/annurev-animal-031412-103659 [PubMed: 25387014]

7. Ritland K, Newton C, Marshall HD. Inheritance and population structure of the white-phased "Kermode" black bear. Curr Biol. 2001; 11 :1468–1472. DOI: 10.1016/s0960-9822(01)00448-1 [PubMed: 11566108]

8. Jones MR, et al. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. Science. 2018; 360 :1355–1358. DOI: 10.1126/science.aar5273 [PubMed: 29930138]

9. Gopalakrishnan S, et al. Interspecific Gene Flow Shaped the Evolution of the Genus Canis. Curr Biol. 2018; 28 :3441–3449. e3445 doi: 10.1016/j.cub.2018.08.041 [PubMed: 30344120]

10. Koepfli KP, et al. Genome-wide Evidence Reveals that African and Eurasian Golden Jackals Are Distinct Species. Curr Biol. 2015; 25 :2158–2165. DOI: 10.1016/j.cub.2015.06.060 [PubMed: 26234211]

11. Wang MS, et al. Ancient hybridization with an unknown population facilitated high altitude adaptation of canids. Mol Biol Evol. 2020; doi: 10.1093/molbev/msaa113

12. Gipson PS, B EE, Bailey Theodore N, Boyd Diane K, Dean Cluff H, Smith Douglas W, Jiminez Michael D. Color Patterns among Wolves in Western North America. Wildlife Sociaety Bulletin. 2002; 30 :821–830.

13. Zhang W, et al. Hypoxia adaptations in the grey wolf (Canis lupus chanco) from Qinghai-Tibet Plateau. PLoS Genet. 2014; 10 e1004466 doi: 10.1371/journal.pgen.1004466 [PubMed: 25078401]

14. Wright S. Color Inheritance in Mammals. Journal Of Heredity. 1917; 8 :224–235.

15. Little, CC. The inheritance of coat color in dogs. Comstock Publishing Associates, Cornell University Press; Constable & Co. Ltd; London: 1957.

16. Dreger DL, Parker HG, Ostrander EA, Schmutz SM. Identification of a mutation that is associated with the saddle tan and black-and-tan phenotypes in Basset Hounds and Pembroke Welsh Corgis. J Hered. 2013; 104 :399–406. DOI: 10.1093/jhered/est012 [PubMed: 23519866]

17. Berryere TG, Kerns JA, Barsh GS, Schmutz SM. Association of an Agouti allele with fawn or sable coat color in domestic dogs. Mamm Genome. 2005; 16 :262–272. DOI: 10.1007/s00335-004-2445-6 [PubMed: 15965787]

18. Giska I, et al. Introgression drives repeated evolution of winter coat color polymorphism in hares. Proc Natl Acad Sci U S A. 2019; 116 :24150–24156. DOI: 10.1073/pnas.1910471116 [PubMed: 31712446]

19. Baiz MD, Wood AW, Brelsford A, Lovette IJ, Toews DPL. Pigmentation Genes Show Evidence of Repeated Divergence and Multiple Bouts of Introgression in Setophaga Warblers. Curr Biol. 2021; 31 :643–649. e643 doi: 10.1016/j.cub.2020.10.094 [PubMed: 33259789]

20. Kim KW, et al. Genetics and evidence for balancing selection of a sex-linked colour polymorphism in a songbird. Nat Commun. 2019; 10 1852 doi: 10.1038/s41467-019-09806-6 [PubMed: 31015412]

21. Toews DP, et al. Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers. Curr Biol. 2016; 26 :2313–2318. DOI: 10.1016/j.cub.2016.06.034 [PubMed: 27546575]

22. Wang S, et al. Selection on a small genomic region underpins differentiation in multiple color traits between two warbler species. Evol Lett. 2020; 4 :502–515. DOI: 10.1002/evl3.198 [PubMed: 33312686]

23. Henkel J, et al. Selection signatures in goats reveal copy number variants underlying breed-defining coat color phenotypes. PLoS Genet. 2019; 15 e1008536 doi: 10.1371/journal.pgen.1008536 [PubMed: 31841508]

24. Fontanesi L, et al. Characterization of the rabbit agouti signaling protein (ASIP) gene: transcripts and phylogenetic analyses and identification of the causative mutation of the nonagouti black coat colour. Genomics. 2010; 95 :166–175. DOI: 10.1016/j.ygeno.2009.11.003 [PubMed: 20004240]

25. Letko A, et al. A deletion spanning the promoter and first exon of the hair cycle-specific ASIP transcript isoform in black and tan rabbits. Anim Genet. 2020; 51 :137–140. DOI: 10.1111/age.12881 [PubMed: 31729778]

26. Duhl DM, Vrieling H, Miller KA, Wolff GL, Barsh GS. Neomorphic agouti mutations in obese yellow mice. Nat Genet. 1994; 8 :59–65. DOI: 10.1038/ng0994-59 [PubMed: 7987393]

27. Balme J, O'Connor S, Fallon S. New dates on dingo bones from Madura Cave provide oldest firm evidence for arrival of the species in Australia. Sci Rep. 2018; 8 9933 doi: 10.1038/s41598-018-28324-x [PubMed: 30026564]

28. Lee EJ, et al. Ancient DNA analysis of the oldest canid species from the Siberian Arctic and genetic contribution to the domestic dog. PLoS One. 2015; 10 e0125759 doi: 10.1371/journal.pone.0125759 [PubMed: 26018528]

29. Sinding M-HS, et al. Arctic-adapted dogs emerged at the Pleistocene–Holocene transition. Science. 2020; 368 :1495–1499. DOI: 10.1126/science.aaz8599 [PubMed: 32587022]

30. Fan Z, et al. Worldwide patterns of genomic variation and admixture in gray wolves. Genome Res. 2016; 26 :163–173. DOI: 10.1101/gr.197517.115 [PubMed: 26680994]

31. Loog L, et al. Ancient DNA suggests modern wolves trace their origin to a Late Pleistocene expansion from Beringia. Mol Ecol. 2019; doi: 10.1111/mec.15329

32. Dreger DL, Schmutz SM. A SINE insertion causes the black-and-tan and saddle tan phenotypes in domestic dogs. J Hered. 2011; 102 (Suppl 1) :S11–18. DOI: 10.1093/jhered/esr042 [PubMed: 21846741]

33. Freedman AH, Wayne RK. Deciphering the Origin of Dogs: From Fossils to Genomes. Annu Rev Anim Biosci. 2017; 5 :281–307. DOI: 10.1146/annurev-animal-022114-110937 [PubMed: 27912242]

34. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29 :15–21. DOI: 10.1093/bioinformatics/bts635 [PubMed: 23104886]

35. Robinson JT, et al. Integrative genomics viewer. Nat Biotechnol. 2011; 29 :24–26. DOI: 10.1038/nbt.1754 [PubMed: 21221095]

36. Dog Biomedical Variant Database, C. Jagannathan V, Drogemuller C, Leeb T. A comprehensive biomedical variant catalogue based on whole genome sequences of 582 dogs and eight wolves. Anim Genet. 2019; 50 :695–704. DOI: 10.1111/age.12834 [PubMed: 31486122]

37. RepeatMasker Open-4.0. 2013

38. Islam S, et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. 2011; 21 :1160–1167. DOI: 10.1101/gr.110882.110 [PubMed: 21543516]

39. Krjutskov K, et al. Globin mRNA reduction for whole-blood transcriptome sequencing. Sci Rep. 2016; 6 31584 doi: 10.1038/srep31584 [PubMed: 27515369]

40. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016; 11 :1650–1667. DOI: 10.1038/nprot.2016.095 [PubMed: 27560171]

41. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016; 34 :525–527. DOI: 10.1038/nbt.3519 [PubMed: 27043002]

42. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25 :1754–1760. DOI: 10.1093/bioinformatics/btp324 [PubMed: 19451168]

43. Rimmer A, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet. 2014; 46 :912–918. DOI: 10.1038/ng.3036 [PubMed: 25017105]

44. Danecek P, et al. The variant call format and VCFtools. Bioinformatics. 2011; 27 :2156–2158. DOI: 10.1093/bioinformatics/btr330 [PubMed: 21653522]

45. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007; 81 :1084–1097. DOI: 10.1086/521987 [PubMed: 17924348]

46. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018; 35 :1547–1549. DOI: 10.1093/molbev/msy096 [PubMed: 29722887]

47. Stecher G, Tamura K, Kumar S. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. Mol Biol Evol. 2020; 37 :1237–1239. DOI: 10.1093/molbev/msz312 [PubMed: 31904846]

48. vonHoldt BM, et al. A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. Genome Res. 2011; 21 :1294–1305. DOI: 10.1101/gr.116301.110 [PubMed: 21566151]

49. Botigué LR, et al. Ancient European dog genomes reveal continuity since the Early Neolithic. Nat Commun. 2017; 8 16082 doi: 10.1038/ncomms16082 [PubMed: 28719574]

50. Frantz LA, et al. Genomic and archaeological evidence suggest a dual origin of domestic dogs. Science. 2016; 352 :1228–1231. DOI: 10.1126/science.aaf3161 [PubMed: 27257259]

51. Ni Leathlobhair M, et al. The evolutionary history of dogs in the Americas. Science. 2018; 361 :81–85. DOI: 10.1126/science.aao4776 [PubMed: 29976825]

52. Skoglund P, Ersmark E, Palkopoulou E, Dalen L. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. Curr Biol. 2015; 25 :1515–1519. DOI: 10.1016/j.cub.2015.04.019 [PubMed: 26004765]

53. vonHoldt BM, et al. Identification of recent hybridization between gray wolves and domesticated dogs by SNP genotyping. Mamm Genome. 2013; 24 :80–88. DOI: 10.1007/s00335-012-9432-0 [PubMed: 23064780]

54. Pitulko VV, Kasparov AK. Archaeological dogs from the Early Holocene Zhokhov site in the Eastern Siberian Arctic. Journal of Archaeological Science: Reports. 2017; 13 :491–515. DOI: 10.1016/j.jasrep.2017.04.003

55. Huerta-Sanchez E, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 2014; 512 :194–197. DOI: 10.1038/nature13408 [PubMed: 25043035]

56. Miao B, Wang Z, Li Y. Genomic Analysis Reveals Hypoxia Adaptation in the Tibetan Mastiff by Introgression of the Gray Wolf from the Tibetan Plateau. Mol Biol Evol. 2017; 34 :734–743. DOI: 10.1093/molbev/msw274 [PubMed: 27927792]

Fig. 1. Coat patterns controlled by *ASIP*.

Drawings of the five pattern types caused by *ASIP* regulatory variation are shown on the left, with representative photographs shown on the right. A completely black coat caused by homozygosity for *ASIP* loss-of-function is not shown. Within each pattern type, there may be variation due to other factors including: (1) the position of the boundaries between pheomelanic and eumelanic areas, e.g. in black saddle or black back; (2) the shade of pheomelanin (red to nearly white); (3) presence of a black facial mask or white spotting caused by genes other than *ASIP*; and (4) length and/or curl of hair coat. Patterns are displayed in order of dominance.
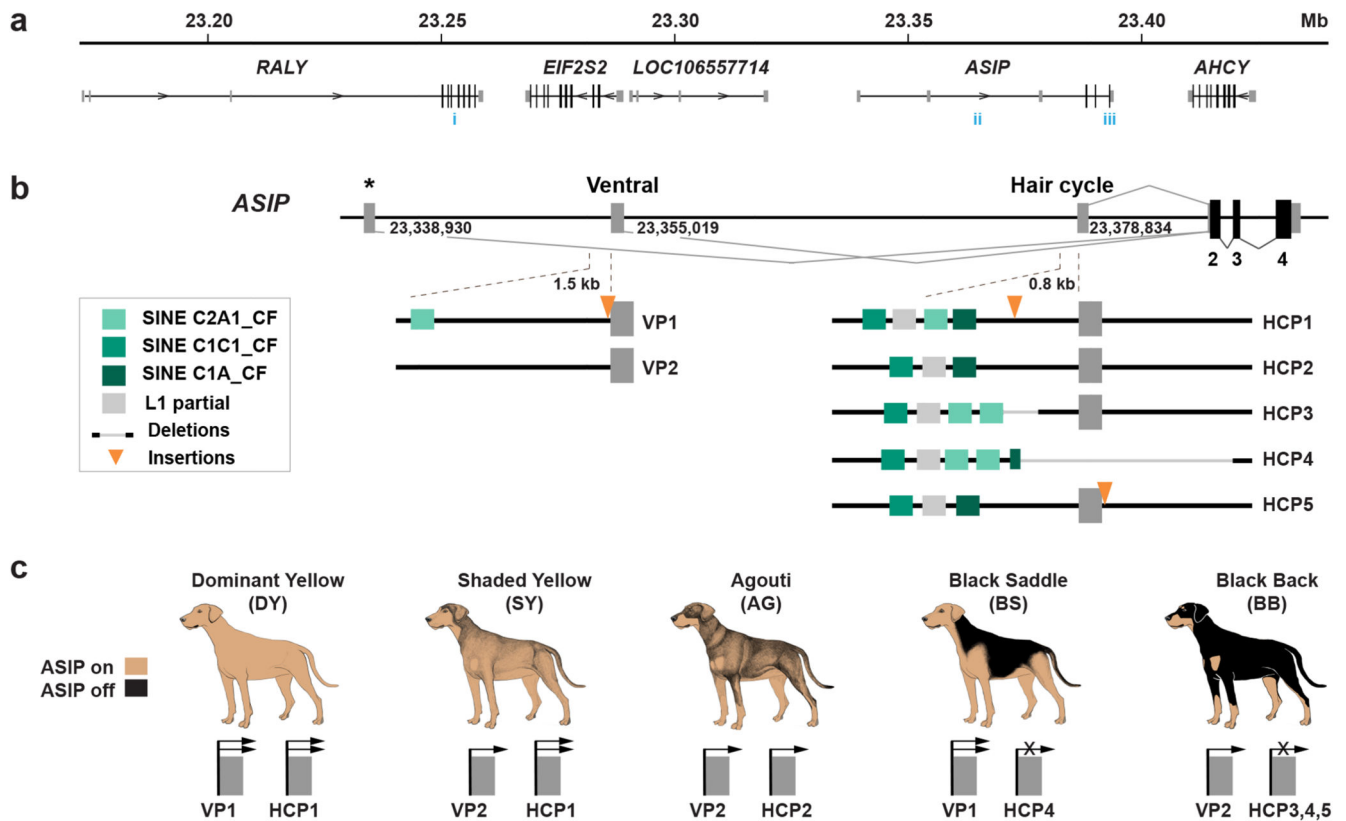
Fig. 2. Structural variation at the *ASIP* locus in domestic dogs with different color patterns
(a) Genomic context (NCBI annotation release 105, CanFam 3.1 assembly). Numbers in blue indicate previously reported variant associations (i [16], ii [32], iii [17]) referred to in the Discussion and in Extended Data Fig. 2. (b) The canine *ASIP* gene has three alternative promoters and 5'-non-coding exons (nucleotide coordinates denote their 3 '-ends. Structural variation within 1.5 kb sections of the ventral and hair cycle-specific promoter explains five different color patterns in dogs. Two different VP haplotypes and five different HCP haplotypes are schematically indicated. The asterisk represents the third promoter and non-coding exon that is not related to *ASIP* pattern variation as described in the text and Fig. 1. (c) Summary of how extended haplotype combinations are related to color pattern phenotypes. Semi-quantitative *ASIP* expression levels are depicted with one or two arrows or an X for no expression (Extended Data Fig. 1).
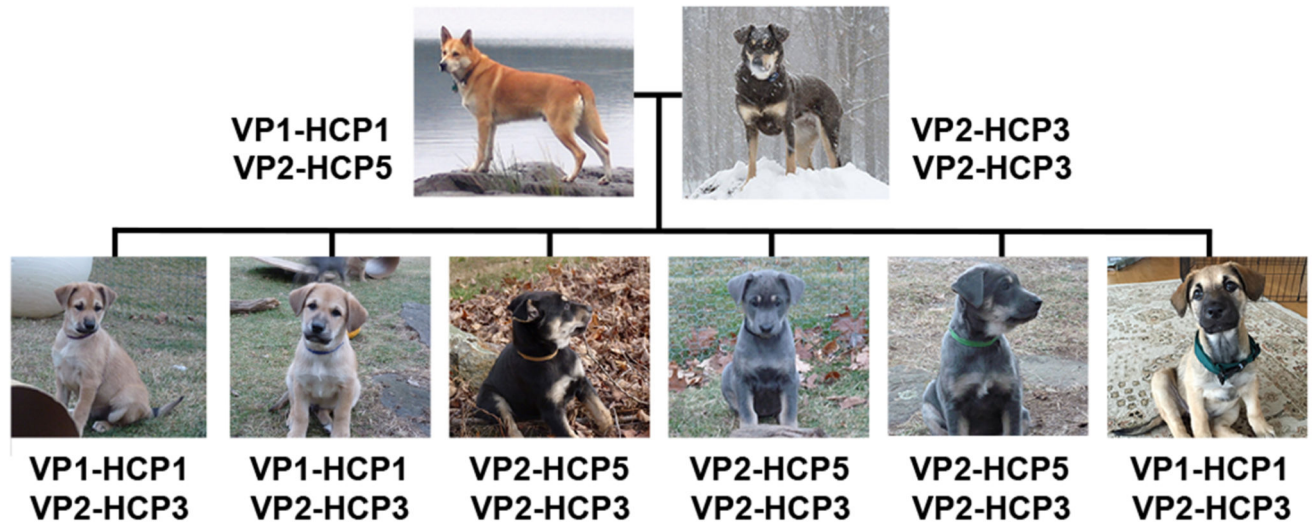
Fig. 3. A family of dogs segregating dominant yellow and two black back haplotypes.
Extended haplotype combinations in a family of Chinook dogs, illustrating, that in combination with VP2, HCP3 and HCP5 both confer a black back phenotype.
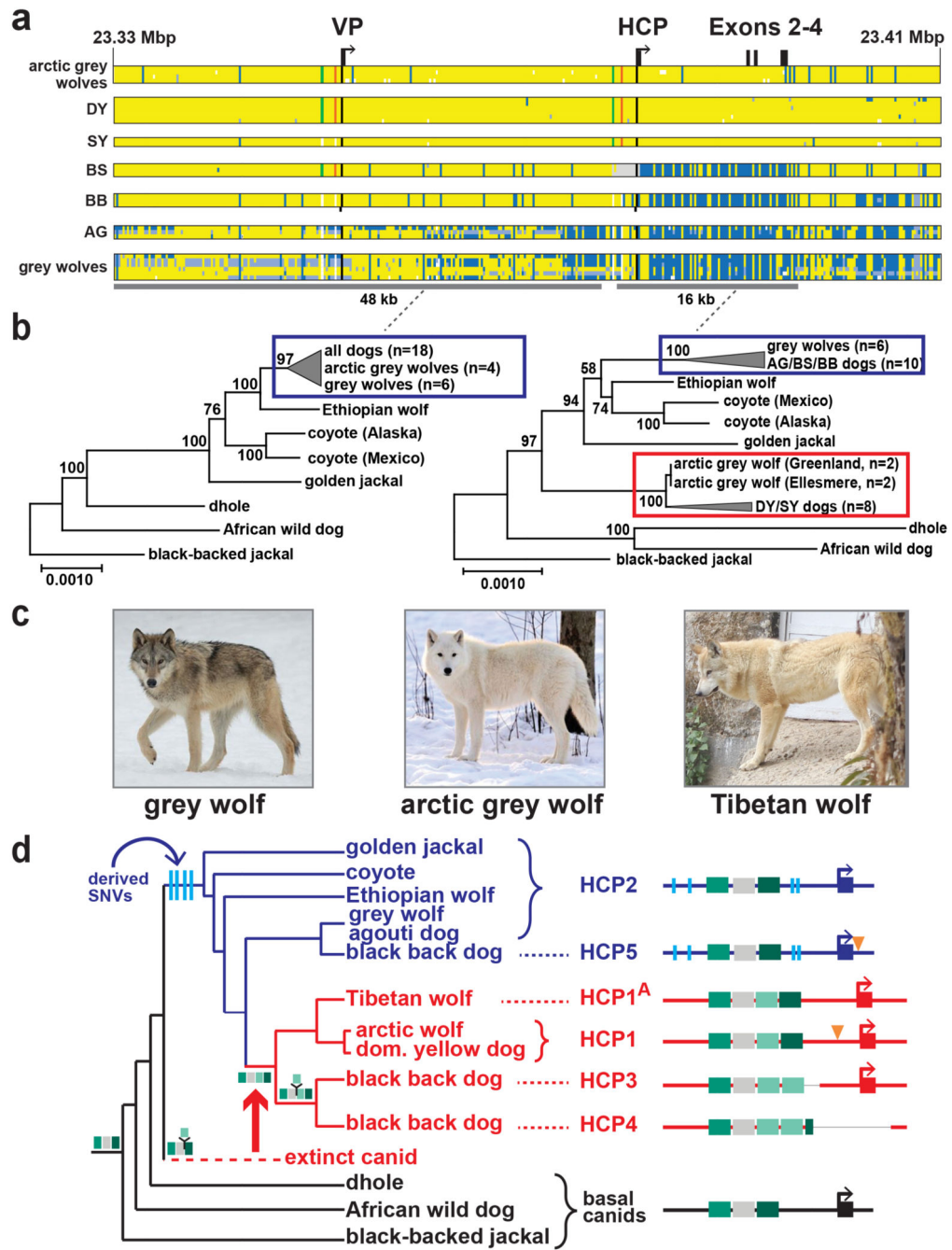
Fig. 4. Yellow dogs and white wolves share an ancient HCP haplotype.
(**a**) Genotypes at 377 SNVs (columns) at the *ASIP* locus in grey wolves and dogs (rows), coded for heterozygosity (light blue), homozygosity for the reference (yellow) or the alternate (dark blue) allele, or as missing genotypes (white). Alternate first exons (arrows) and nearby DY-associated structural variants (SINE insertions: green, polynucleotide expansions: orange) are included for reference. (**b**) Maximum likelihood phylogenies, including seven extant canid species and the dog, from 48 and 16 kb intervals upstream or downstream of the HCP, respectively. Grey wolf/dog phyletic clades are highlighted

with boxes to indicate relationships that are consistent (blue) or inconsistent (red) with genome-wide phylogenies. (**c**) Images of a grey wolf, arctic grey wolf, Tibetan wolf. (**d**) A phylogeny representing distinct HCP evolutionary histories inferred from genetic variation in extant canids. Structural variants (as represented in Fig. 2) and derived SNVs (cyan) distinguish wolf-like canid (blue), ghost lineage (red), and basal canid (black) haplotypes.
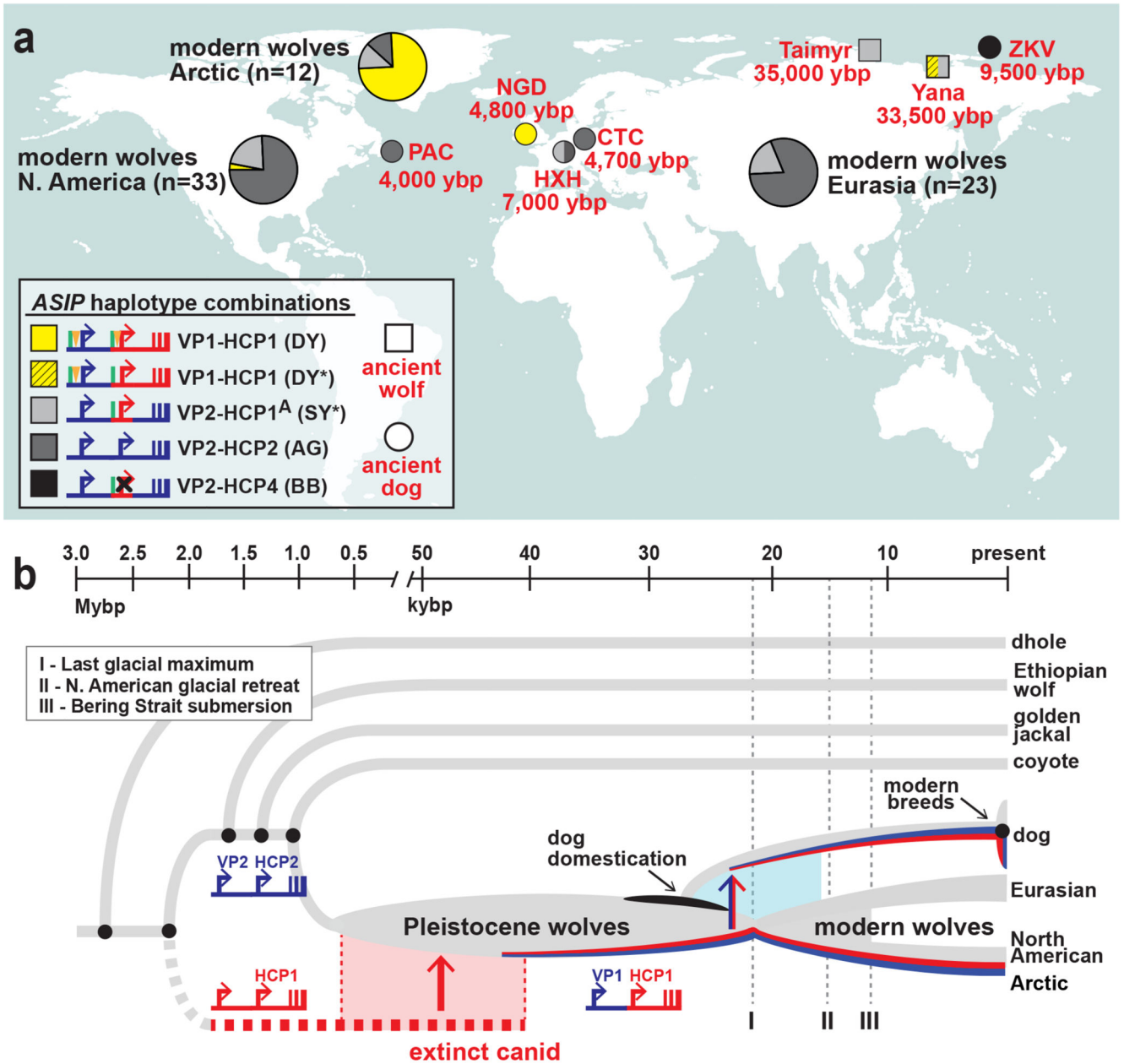
Fig. 5. Distribution of *ASIP* alleles in ancient dogs and wolves, and an evolutionary model for dominant yellow acquisition.

(**a**) *ASIP* haplotypes were inferred from whole genome sequencing of 5 ancient dog (circles), 2 ancient wolves (squares), and 68 modern wolves (pie charts) distributed across the Holarctic (see Extended Data Fig. 7 and Supplementary Table 12 for detailed haplotype representations). Asterisks indicate SY/DY haplotypes for which the HCP1 insertion is either absent (SY*) or not ascertainable (DY*). (**b**) A model for the origin of the dominant yellow haplotype and its transmission into dogs and arctic wolves, in which molecular alterations at modular promoters were acquired by introgression (red, HCP1) or by mutation

in the grey wolf (blue, VP1). The timeline for speciation events, dog domestication, and geological events affecting grey wolf dispersal are based on prior studies [10,33].

Table 1

***ASIP* diplotype association with pattern phenotype**

Pattern phenotype	Promoter diplotype	Concordant	Discordant
Dominant Yellow (n = 114)[a]	VP1-HCP1 / VP1,2-HCP1,3,4,5	113	1[b]
Shaded Yellow (n = 64)[a]	VP2-HCP1 / VP2-HCP1,3,5	52	12[b]
Agouti (n = 46)	VP2-HCP2 / VP2-HCP2,3,5	46	0
Black Saddle (n = 53)	VP1-HCP4 / VP1,2-HCP3,4	53	0
Black Back (n = 89)	VP2-HCP3,4,5 / VP2-HCP3,4,5	89	0

[a]Previous studies did not differentiate between the dominant yellow and the shaded yellow phenotype.

[b]These dogs had a eumelanistic masking pattern, which prevented reliable phenotype distinction between dominant yellow and shaded yellow.