A β-Defensin Mutation Causes Black Coat Color in Domestic Dogs

Sophie I. Candille,[1]* Christopher B. Kaelin,[1]* Bruce M. Cattanach,[2] Bin Yu,[3] Darren A. Thompson,[3] Matthew A. Nix,[3] Julie A. Kerns,[1]† Sheila M. Schmutz,[4] Glenn L. Millhauser,[3] Gregory S. Barsh[1]‡

AUTHORS' SUMMARY

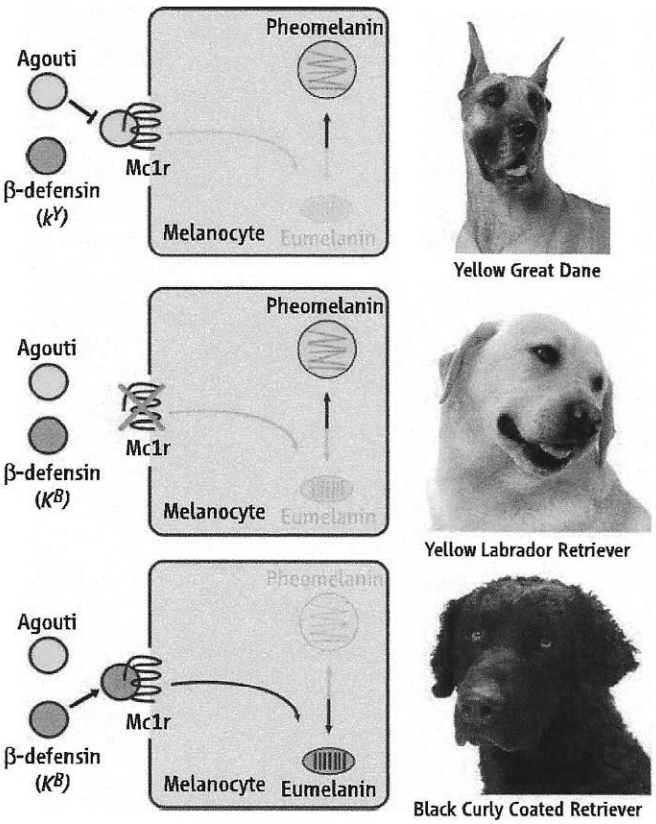
The marked spectrum of color and diversity of patterns that we see in mammals arises, unexpectedly, from variation in the quantity, quality, and regional distribution of just two types of pigment—black eumelanin and yellow pheomelanin. The appeal of unusual coat colors and patterns has motivated their selection in domestic animals, providing geneticists with a model for studying gene action and interaction that began a century ago and continues today. Most of the work has been carried out in laboratory mice, where studies of more than 100 different coat-color mutations have provided insight into stem cell biology (hair graying), biogenesis of intracellular organelles (pigmentary dilution), and hormone-receptor interactions (switching between the synthesis of eumelanin and pheomelanin).

The latter process—commonly known as pigment "type-switching"—is controlled primarily by the melanocortin system, in which a family of G protein–coupled receptors (identified by virtue of their response to α-melanocyte–stimulating hormone or adrenocorticotrophic hormone) has been implicated not only in pigmentation but also in cortisol production, body weight regulation, and exocrine gland secretion. In most mammals, pigment type-switching is controlled by two genes, the *Melanocortin 1 receptor* (*Mc1r*) and *Agouti*, which encode a seven transmembrane–domain receptor and its extracellular ligand, respectively. Indeed, our current understanding of melanocortin biology stems from the identification in laboratory mice of *Mc1r* mutations as the cause of *recessive yellow* and *Agouti* mutations as the cause of *lethal yellow*.

Clarence Cook Little, who developed many of the original laboratory mouse strains and founded The Jackson Laboratory, was also one of the first dog geneticists. He recognized that dominant inheritance of a black coat was mediated differently in dogs than in other animals (*1*). Using classical linkage analysis, we realized that the *dominant black* gene represented a previously unrecognized component of the melanocortin pathway (*2*). Unexpectedly, we found the responsible gene to encode a β-defensin, a secreted protein previously studied for its role in immunity.

The identification of *dominant black* (formally, an allele of the "*K* locus") relied on two major advances in dog genetics: the sequencing of the dog genome and recognition that the distinctive genetic structure of dog breeds allows for efficient gene mapping (*3*). Dogs were domesticated from wolves more than 15,000 years ago and expanded into a diverse population until the recent establishment of dog breeds. This population history is well-suited for high-resolution genetic mapping of old traits, like black coat color, that are found in multiple modern breeds. Using a combination of pedigree analysis and association studies within and among dog breeds, we identified a mutation in a β-defensin gene, *CBD103*, that correlates with black coat color in 38 different breeds. We confirmed the role of *CBD103* in pigment type-switching by demonstrating that the dog gene causes a black coat in transgenic mice. CBD103 is a member of a large family of secreted peptides with structures similar to that of Agouti and is highly expressed in dog skin.

We used biochemical and cell-based assays to show that CBD103, like Agouti, binds competitively to the Mc1r, leading to an updated model of the pigment type-switching pathway (see figure). Moreover, studies with another β-defensin and additional melanocortin receptors reveal the potential for extensive cross-talk between β-defensins and the melanocortin system. In humans and other animals, β-defensins are highly polymorphic in sequence and copy number. Current β-defensin research is focused primarily on the immune system. This stems from the early discovery of defensins in phagocytic cells and their antimicrobial properties in vitro, together with more recent work demonstrating that defensins can act as receptor-specific chemotactic agents. Our work indicates that β-defensins do more than defend and suggests that the marked molecular variation in this family supplies a diverse and rapidly evolving family of ligands for G protein–coupled receptors in many different biologic systems.



Production of yellow versus black pigment in dogs is controlled by three genes: *Mc1r*, *Agouti*, and *CBD103*. Dogs carrying wild-type alleles for all three genes have a yellow coat resulting from Agouti antagonism of Mc1r signaling in melanocytes (yellow Great Dane, top). Dogs carrying a loss-of-function mutation at *Mc1r* have a yellow coat, regardless of their genotype at *Agouti* or *CBD103* (yellow Labrador Retriever, middle). Dogs carrying wild-type alleles for *Mc1r* and *Agouti*, together with the *dominant black* allele of *CBD103* (K^B) have a black coat resulting from the interaction between a β-defensin and Mc1r (black Curly Coated Retriever, bottom).

Summary References

1. C. C. Little, *The Inheritance of Coat Color in Dogs* (Comstock, Ithaca, NY, 1957).
2. J. A. Kerns *et al.*, *Genetics* **176**, 1679 (2007).
3. K. Lindblad-Toh *et al.*, *Nature* **438**, 803 (2005).

FULL-LENGTH ARTICLE

Genetic analysis of mammalian color variation has provided fundamental insight into human biology and disease. In most vertebrates, two key genes, *Agouti* and *Melanocortin 1 receptor* (*Mc1r*), encode a ligand-receptor system that controls pigment type-switching, but in domestic dogs, a third gene is implicated, the *K* locus, whose genetic characteristics predict a previously unrecognized component of the melanocortin pathway. We identify the *K* locus as β-*defensin 103* (*CBD103*) and show that its protein product binds with high affinity to the Mc1r and has a simple and strong effect on pigment type-switching in domestic dogs and transgenic mice. These results expand the functional role of β-defensins, a protein family previously implicated in innate immunity, and identify an additional class of ligands for signaling through melanocortin receptors.

Genetic analysis of model systems in laboratory animals underlies much of what we know about major signaling pathways in multicellular organisms. In mammals, coat-color mutations have proven especially fruitful, because much of the molecular machinery used by the pigmentary system is either shared by, or homologous to, genes used for other physiologic pathways (*1*).

This approach has been particularly useful for pigment "type-switching," a phenomenon in which melanocytes synthesize eumelanin (a black or brown pigment) versus pheomelanin (a red or yellow pigment), depending on the phase of the hair growth cycle, position on the body, and the genotype of several key loci (*2*). In most mammals, two genes that control pigment type-switching have been recognized: the *Mc1r*, which encodes a seven transmembrane–spanning domain protein expressed on melanocytes (*3*–*5*), and *Agouti*, which encodes a ligand for the Mc1r that is secreted by specialized dermal cells and which inhibits Mc1r signaling (*6*–*9*). Mc1r activation causes exclusive production of eumelanin, whereas Mc1r inhibition causes exclusive production of pheomelanin (*5*, *10*). Thus, gain-of-function *Mc1r* mutations cause dominant inheritance of a black coat, whereas gain-of-function *Agouti* mutations cause dominant inheritance of a yellow coat. The Mc1r was first recognized by virtue of its ability to respond to peptides derived from proopiomelanocortin (POMC), such as α-melanocyte–stimulating hormone (α-MSH) (*3*, *4*); however, a null mutation of *Pomc* has no effect on pigmentation in C57BL/6 mice, leading to the suggestion that the Mc1r has high basal activity and/or additional stimulatory ligands (*11*).

[1]Departments of Genetics and Pediatrics, Stanford University, Stanford, CA, USA. [2]Medical Research Council (MRC) Mammalian Genetics Unit, Harwell, Oxfordshire, OX11 ORD, UK. [3]Departments of Chemistry and Biochemistry, University of California at Santa Cruz (UCSC), Santa Cruz, CA 95064, USA. [4]Department of Animal and Poultry Science, University of Saskatchewan, Saskatoon S7N 5A8, Canada.

*These authors contributed equally to the work.
†Present address: Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.
‡To whom correspondence should be addressed. E-mail: gbarsh@stanford.edu

In a classic series of papers in the early 1900s, Sewall Wright (*12*) concluded that genetic mechanisms for color variation were largely conserved across mammals. An exception, however, later became apparent from the work of Clarence Cook Little on domestic dogs (*13*), in which dominant inheritance of a black coat was shown to involve a locus distinct from *Mc1r*. At the time, Little posited that *dominant black* was caused by an unusual allele of *Agouti*; however, using molecular linkage analysis, we recently demonstrated the presence of a third gene in dogs that interacts with *Agouti* and *Mc1r*, which we named the *K* locus (*14*). We found that the *K* locus has three alleles with a simple dominance order [*Black* (K^B) > *brindle* (k^{br}) > *yellow* (k^y)], that the *K* locus and *Agouti* behave similarly in genetic interaction studies (*Mc1r* is epistatic to both *K* and *Agouti*), and that the genetic map position of *K* does not correspond to the predicted location of any previously known pigmentation gene. We use the distinctive evolutionary history of domestic dogs to show that the *K* locus encodes a previously unrecognized class of melanocortin receptor ligands.

Linkage and association mapping of the *K* locus. We showed previously that the K^B mutation mapped to a 12-Mb interval on the distal end of dog chromosome 16 (CFA16), between markers REN292N24 and FH3592 (*14*). We ascertained additional kindreds segregating K^B, k^{br}, and k^y to refine the map location [see supporting online material (SOM) text] and defined overlapping critical regions of 3.8 and 7.6 Mb for the K^B and k^{br} mutations, respectively (Fig. 1A and figs. S1 to S4).

We used an association-based strategy to narrow the critical region. Because most breeds were derived in the past 200 years from small founding populations (*15*), mutations within a breed are expected to be identical by descent and share extended haplotypes. In Boxers and Great Danes, we identified broad peaks of significant association (Bonferroni-corrected P value < 0.05) that extended over 1.9 Mb and 320 kb, respectively (Fig. 1, B to D). Sixteen genes have been annotated to the region of significant association in Great Danes, including a gene cluster that encodes 12 β-defensins (*16*): small antimicrobial peptides that are secreted mainly by epithelial cells (*17*, *18*). We sequenced the mature protein-coding regions for nine members of the β-defensin cluster (those known at the time) in dogs carrying K^B and/or k^y and identified several polymorphisms concordant with the K^B allele, including a 3–base pair (bp) deletion in the second exon of *CBD103*, the ortholog of human *DEFB103*, that predicts an in-frame glycine deletion (ΔG23).

To evaluate the extent to which the ΔG23 polymorphism distinguishes K^B versus k^y more broadly, we examined dogs from 38 breeds that could be classified into two categories with regard to their putative *K* locus genotype (SOM text). Among 454 dogs, there were 13 cases where the ΔG23 polymorphism did not correlate with coat-color phenotype. However, sequencing of *Agouti* and *Mc1r* revealed that each discordant case could be explained by known epistatic interactions (*19*, *20*) (table S3). These results indicate that K^B alleles in all breeds are probably identical by descent and suggest that the ΔG23 polymorphism or a closely linked variant in complete linkage disequilibrium (LD) is the K^B mutation.

Short-range haplotype and resequencing analysis of K^B-bearing chromosomes. By contrast to the pattern of LD within breeds, which affords a powerful approach for association mapping with megabase resolution, the pattern of LD across breeds is more fine-grained and therefore provides the opportunity for high-resolution haplotype mapping when mutations in different breeds are identical by descent (*21*). We identified 28 polymorphisms [22 single-nucleotide polymorphisms (SNPs) and six indels including the ΔG23 polymorphism] in a 20-kb interval surrounding *CBD103* that were then used to infer short-range haplotypes for 14 K^B-bearing and 16 k^y-bearing chromosomes selected from seven breeds (SOM text). We observed six "parental" k^y-bearing and five "parental" K^B-bearing chromosomes (depicted in yellow and blue, respectively, in Fig. 2A). We also identified eight chromosomes that carried a single ancestral recombination event, which together defined a maximal interval for K^B of 9146 bp (Fig. 2A). Complete resequencing of this interval (except for three homopolymer tracts) in five k^y/k^y animals, one K^B/k^y animal, and four K^B/K^B animals from seven breeds revealed two polymorphisms besides ΔG23 that are perfectly concordant with *K* locus genotype (S104 and S105) (Fig. 2A and table S2).

The 9146-bp interval contains both exons of *CBD103*, the first exon of dog expressed sequence tag (EST) CX990240, and dog EST CO665262 (Fig. 2). However, several considerations indicate that *CBD103* is, indeed, the

K locus and that the ΔG23 deletion in *CBD103* is the K^B mutation. First, the other two transcribed elements in the critical interval are represented in the database by single ESTs and are not known to encode proteins or to be expressed in the skin. Second, S104 and S105 lie in a long terminal repeat element that is 3 kb upstream of the first exon of *CBD103* and have no effect on mRNA levels of *CBD103* (Fig. 3A). Finally, as discussed further below, *CBD103* is highly expressed in skin, the ΔG23 deletion affects CBD103 protein function, and pharmacologic studies reveal that CBD103 can modulate melanocortin signaling.

The preceding discussion has referred to *K* locus variation as though k^y is ancestral, whereas K^B is derived: a hypothesis based on the comparative genetic distribution of coat-color phenotypes and inheritance patterns. Considerations based on sequence alignments confirm this hypothesis: Mammalian *CBD103* orthologs that we identified from the available genome sequence are each 67 amino acids in length, and the optimal sequence similarity alignment contains no gaps or insertions (Fig. 2B), indicating that ΔG23 and consequently the K^B mutation occurred specifically within the canid lineage.

Expression of dog defensins in skin and in transfected keratinocytes. We isolated RNA from the skin of a k^y/k^y Doberman Pinscher and a K^B/k^y mixed-breed dog and surveyed the expression of the 19 β-defensin genes that are clustered on chromosomes 16 or 25 by reverse transcription polymerase chain reaction (RT-PCR) (*16*). Expression was detectable only for two genes: *CBD1* and *CBD103* (fig. S6). We then used quantitative RT-PCR to measure levels of skin mRNA from four K^B/k^y samples and four k^y/k^y samples, which were all from mixed-breed dogs, and found no effect of *K* locus genotype on levels of *CBD1*, *CBD103*, or *Agouti* mRNA (Fig. 3A).

Available antisera against human DEFB103 are unable to detect the endogenous dog protein by Western blotting or immunohistochemistry; therefore, we generated epitope-tagged expression constructs for each allele (CBD103V5 and CBD103ΔG23V5) and studied their patterns and levels of protein expression after transfection of cultured mouse keratinocytes.

In cell extracts analyzed by Western blotting, antisera against the V5 epitope detect a single fragment whose size (about 8 kD) corresponds to the expected molecular mass of the tagged protein after signal peptide cleavage; in media, an additional slightly smaller band is present, which suggests additional processing (Fig. 3B). The relative ratios of the two bands are similar in media from keratinocytes transfected with either construct; however, the total amount of immunodetectable protein in media was significantly greater for CBD103ΔG23V5 as compared with that for CBD103V5 (P = 0.0021, Cochran-Mantel-Haenszel chi-square test). Thus, loss of the N-terminal glycine from CBD103 does not affect intracellular processing but allows more of the mature protein to accumulate in the media and/or extracellular space.

CBD103 activity in vivo and in vitro. To further explore the function of *CBD103* in an experimental genetic system, we generated transgenic mice in which a cDNA encoding either the K^B or the k^y allele was driven by a strong and widely expressed promoter (*22*). We chose a genetic background that normally has Agouti-banded hairs and observed that two transgenic founders generated with the *CBD103*ΔG23 cDNA (the K^B allele) displayed a predominantly black coat with small patches of banded hair (Fig. 4). Unexpectedly, the normal *CBD103* cDNA (the k^y allele) also produced transgenic mice with a black coat in 20 out of 21 founders (Fig. 4). Furthermore, we observed that transgenic animals were smaller than their nontransgenic littermates. By 2 weeks of age, female transgenic animals were easily recognized by their dark coat and small size; in adult mice, reduced body weight persists in both males and females (fig. S7).

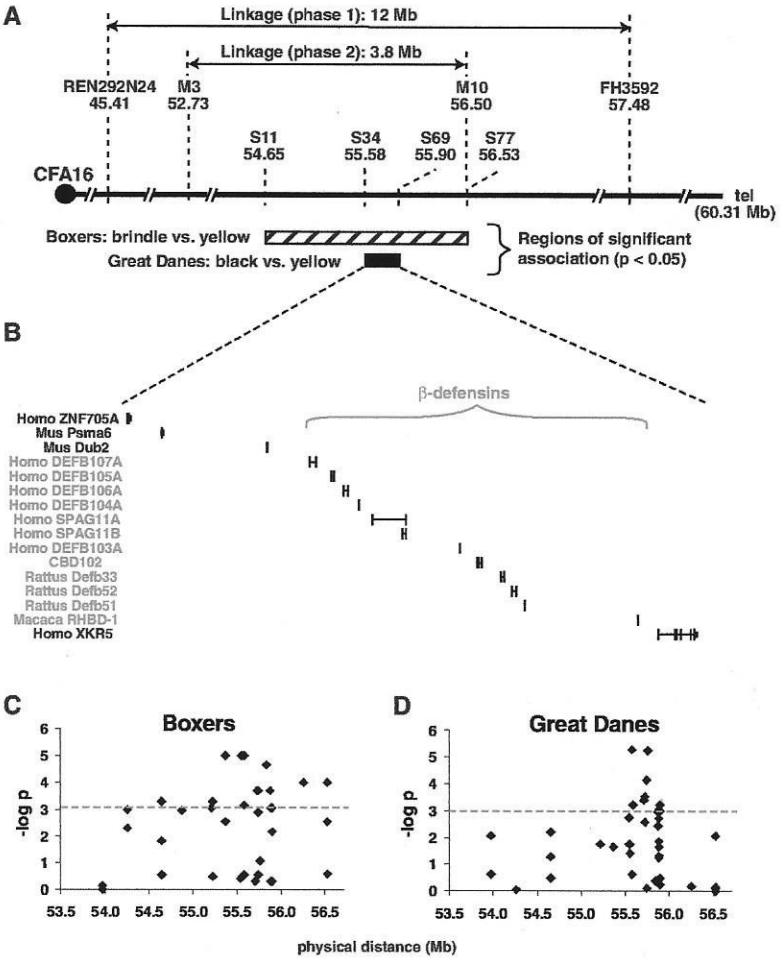


Fig. 1. Genetic mapping of the *K* locus. **(A)** Initial linkage studies [phase 1, (*14*)] defined a 12-Mb critical region for K^B; ascertainment and characterization of additional kindreds narrowed the interval to 3.8 Mb (phase 2, figs. S1 to S4). Association analysis for 60 markers in brindle (n = 12) versus yellow (n = 10) Boxers, and for 51 markers in black (n = 9) versus yellow (n = 10) Great Danes, was carried out as described in the text. **(B)** Candidate genes in the 320-kb region of greatest association in Great Danes; this region includes 12 β-defensin genes (shown in red). **(C** and **D)** Significance, plotted as −log of P values from a chi-square test of allele counts, is shown as a function of distance along CFA16 (only for SNPs present at greater than 10% frequency and genotyped in at least 75% of the samples). The dashed red line indicates a Bonferroni-corrected 5% significance level; these regions are indicated by hatched and black bars for Boxers and Great Danes, respectively, in (A). Annotation is based on the Non-dog RefSeq track in the UCSC Genome Browser, except for CBD102, identified by Patil *et al.* (*16*).