

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
5 November 2009 (05.11.2009)(10) International Publication Number
WO 2009/134226 A1(51) International Patent Classification:
C12Q 1/68 (2006.01) *G01N 33/48* (2006.01)(21) International Application Number:
PCT/US2008/005666(22) International Filing Date:
1 May 2008 (01.05.2008)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): **THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY** [US/US]; 1705 El Camino Real, Palo Alto, California 94306-1106 (US).

(72) Inventors; and

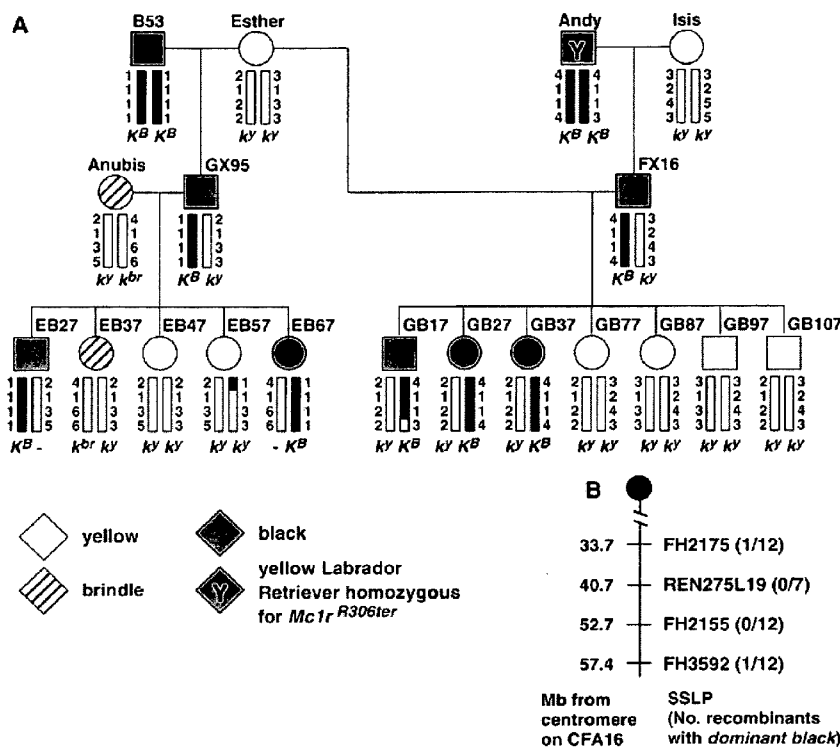
(75) Inventors/Applicants (for US only): **CANDILLE, Sophie Isabelle** [FR/US]; 300 Pasteur Drive, Department of Genetics, Alway Building, M-333, Stanford, California 94305-5120 (US). **KAELIN, Christopher Bryan**[US/US]; B275 Beckman Center, Mail Code: 5120, Stanford, California 94305-5120 (US). **KERNS, Julie** [US/US]; 1100 Fairview Avenue N, A1-162, Seattle, Washington 98109 (US). **BARSH, Gregory S.** [US/US]; B271A Beckman Center, Mail Code: 5323, Stanford, California 94305-5323 (US).(74) Agent: **SHERWOOD, Pamela J.**; Bozicevic, Field & Francis LLP, 1900 University Avenue, Suite 200, East Palo Alto, California 94303 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[Continued on next page]

(54) Title: CANINE COAT COLOR PREDICTION

FIGURE 1



[Continued on next page]



(84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI

(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (*Art. 21(3)*)

— with sequence listing part of description (*Rule 5.2(a)*)

(57) **Abstract:** Compositions and methods are provided for genotyping canines with respect to the genes that determine coat color. Interaction between alleles of the K locus with alleles at *Agouti* and *Mc1r* determine whether eumelanin or pheomelanin is produced in the coat. Three alleles are identified at the K locus, *yellow* (k^y), *brindle* (k^b), and *black* (K^B). Determination of the presence or absence of these alleles allows a prediction of canine coat color, and finds use in breeding programs for a desired color.

CANINE COAT COLOR PREDICTION

BACKGROUND OF THE INVENTION

- [0001] Morphologic variation among domestic dogs exemplifies the power of selective breeding to uncover a diversity of phenotypes from a relatively homogeneous founder population. The genetics of coat color variation in dogs with respect to eumelaninic vs. pheomelaninic coloration has pointed to a genetic system that is distinct from other known mammals.
- [0002] In all mammals that have been studied to date, hair follicle melanocytes synthesize red-yellow pheomelanin or black-brown eumelanin depending on the balance between two key genes, *Agouti* and *Melanocortin 1 receptor (Mc1r)* which encodes a seven transmembrane-spanning domain protein expressed on melanocytes. *Agouti* encodes a signaling molecule secreted from specialized cells in the dermis that acts as an inhibitory ligand for the *Mc1r* expressed on melanocytes. *Mc1r* activation causes exclusive production of eumelanin, while *Mc1r* inhibition causes exclusive production of pheomelanin, thus mutations that constitutively activate the *Mc1r* cause a uniform black appearance, generally inherited in a dominant manner, while mutations that inactivate the *Mc1r* cause a uniform red or yellow appearance, generally inherited in a recessive manner. Conversely, because *Agouti* protein inhibits *Mc1r* activity, gain-of-function mutations yield dominant inheritance of a yellow coat, while loss-of-function mutations yield recessive inheritance of a black coat.
- [0003] Most dogs with a uniform black appearance, e.g. the Newfoundland, the Flat-Coated Retriever, black Labrador Retrievers, or black Poodles, exhibit dominant transmission of the black color, consistent with mutations that constitutively activate the *Mc1r*. However, pedigree and segregation analyses indicated that dominant black was non-allelic with *recessive yellow*, leading to the suggestion that dominant black might be an unusual allele of the *Agouti* locus, *A^S*. Recently, a Labrador Retriever X Greyhound backcross was examined with molecular probes for *Agouti* and *Mc1r*, and it was concluded that neither gene could account for the Labrador Retriever-derived black variant, which was inherited in an apparent autosomal dominant manner.
- [0004] An additional aspect of coat color variation in domestic dogs that appears distinct from most other mammals is the phenotype known as brindle, in which stripes of red-yellow hair alternate with black-brown hair. Brindle stripes form an irregular pattern, typically with a "V" shape over the dorsum, and an "S" shape over flanks and ventrum. Brindle segregates as a single gene in a variety of dog breeds such as the Boxer, Greyhound, and French Bulldog, and has been thought by some authors to be caused by variation in *Agouti*, but by others to be caused by variation in *Mc1r*.

[0005] Dog breeders desire an educated approach to breeding, in which the outcome of a cross can be reasonably predicted. Genetic testing for color traits allows improved decision making in this respect.

SUMMARY OF THE INVENTION

[0006] Compositions and methods are provided for genotyping canines with respect to the genes that determine coat color. It is shown herein that interaction between alleles of the K locus, which may also be referred to as CBD103, with alleles at *Agouti* and *Mc1r* determine whether eumelanin or pheomelanin is produced in the coat. Three alleles are identified at the K locus, *yellow* (k^y), *brindle* (k^{br}), and *black* (K^B). Determination of the presence or absence of these alleles allows a prediction of canine coat color, and finds use in breeding programs for a desired color.

[0007] In one embodiment of the invention, a method is provided for canine coat color genotyping, the method comprising determining the genotype of a canine at the K locus. Various methods may be used for genotyping, for example by amplification of the K locus from a canine genetic sample, and determining the alleles that are present, e.g. by sizing the amplification product, hybridization, sequencing, and the like. Alternatively the genetic sample may be evaluated by hybridization, RFLP analysis, SNP analysis, etc. In some embodiments, the sample is also genotyped at one or both of the *Mc1r* locus, and the *Agouti* locus. In other embodiments, the sample is further genotyped at one or more additional coat color loci, including, without limitation, the loci for tyrosinase related protein 1, melanophilin, SILV (formerly PMEL17), and microphthalmia-associated transcription factor. The method for genotyping may further include an analysis of an animal or a breeding pair for probability of offspring having a coat color of interest.

[0008] In other embodiments of the invention, compositions are provided for determining the genotype of a canine at the K locus, which compositions may include, for example, primers for amplification of the CBD103 gene, primers for hybridization to CBD103, etc. Such compositions may include a polynucleotide having the sequence set forth in SEQ ID NO:3, or a fragment thereof of at least about 15, at least about 20, at least about 25 nucleotides or more comprising the $\Delta G23$ coding sequence, i.e. spanning at least SEQ ID NO:1 at the five G residues at positions 177-181, or the corresponding region of SEQ ID NO:3. The compositions may be provided in a kit format with instructions for use, including an analysis of an animal or a breeding pair for probability of offspring having a coat color of interest. Kits may further include genetic primers for determination of the genotype at the *Mc1r* locus, and the *Agouti* locus. In other embodiments, the kit further includes polynucleotide primers for genotyping canines at one or more additional coat color loci,

including, without limitation, the loci for tyrosinase related protein 1, melanophilin, SILV (formerly PMEL17), and microphthalmia-associated transcription factor.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] Figure 1. Segregation of CFA16 haplotypes in the EB and GB litters. A, Haplotypes based on the four SSLP markers shown in panel B are indicated with vertical bars, just above genotypes for *K* locus alleles. (As described in the text, brindle and yellow were considered in the same class, "non-black", for analysis of the genome scan; genotypes given here for k^{br} and k^y are based on information presented in Figure 3). Black-colored haplotypes originate from the Labrador Retriever grandparent carrying dominant black (B53 or Andy); white-colored haplotypes originate from the non-black Greyhound parent or grandparent (Esther or Isis). SSLP alleles are numbered arbitrarily according to increasing size for each marker. B, Physical location of SSLP markers used in panel A indicated in megabases (Mb) from the centromere. To the right of each marker name is given the number of animals recombinant between that marker and the *K* allele, over the total number of animals that were informative for that marker. Recombinant chromosomes are carried by EB57 and GB17, and define a critical region for *K* between FH2175 and FH3592. In addition to the *K* locus genotypes depicted in the figure, *Agouti* and *Mc1r* genotypes were determined for every dog.

[0010] Figure 2. Segregation of CFA16 haplotypes in the FB and HB litters. A, B, Symbols are as in Figure 1. Recombinant chromosomes are carried by FB27, FB67, and HB27, and indicate that *K* must lie centromere-distal to REN292N24. In addition to the *K* locus genotypes depicted in the figure, *Agouti* and *Mc1r* genotypes were determined for every dog. C, Diagram of the *K* critical region from REN292N24 to FH3592, indicating the location of RefSeq genes in the region (blue), and evolutionarily conserved regions in the human genome. Annotation is based on the CanFam1.0 dog genome assembly as displayed by UCSC Genome Browser using the "Human Net" comparative genomics track, in which red and yellow indicate sequence similarity to human chromosomes 8 and 4, respectively.

[0011] Figure 3. Segregation of FH2155 alleles in three kindreds with brindle and yellow. As described in the text, there is perfect cosegregation of FH2155 with brindle vs. yellow under a model of dominant inheritance with $K^B > k^{br} > k^y$, corresponding to a LOD score of 3.6.

[0012] Figure 4. Models for gene action at the *K* locus. Both models must account for the observations that (1) the dominance order of *Agouti* is opposite to that of *K*; (2) *Mc1r* alleles are epistatic to both *Agouti* and *K* locus alleles; (2) "black alleles" of *K* are epistatic to "yellow alleles" of *Agouti*; and (3) "black alleles" of *Agouti* are epistatic to "yellow alleles" of

K. These observations are consistent with a model in which (A) the K gene product functions to inhibit Agouti function, but are also consistent with a model in which (B) the K gene product acts directly at the Mc1r to stimulate melanocortin signaling and thereby oppose the action of Agouti protein indirectly.

[0013] Figure 5. Genetic mapping of the K locus. (A) Initial linkage studies (phase 1,) defined a 12 Mb critical region for KB; ascertainment and characterization of additional kindreds narrowed the interval to 3.8 Mb. Association analysis for 60 markers in brindle (n=12) vs. yellow (n=10) Boxers, and for 51 markers in black (n=9) vs. yellow (n=10) Great Danes was carried out as described in the text. (C,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). (B) Candidate genes in the 320 kb region of greatest association in Great Danes; this region includes 12 beta-defensin genes (shown in red). Annotation is based on the Non-dog RefSeq track in the UCSC Genome Browser, except for CBD102.

[0014] Figure 6. Resequencing and recombinant haplotype-based mutation analysis for K^B - vs. k^Y -bearing chromosomes. (A) A 20 kb region surrounding CBD103ΔG23 was resequenced (except for repetitive regions) in 10 dogs from 7 breeds, and haplotypes inferred for 28 high frequency biallelic polymorphisms. Blue and yellow squares represent the major and minor alleles in K^B - bearing chromosomes, respectively, and allow some haplotypes to be designated as " k^Y -parental", " K^B -parental" or "proximal recombinant", as indicated. White squares represent missing data. Genotypes for 5 Great Danes (*) were determined in a second resequencing round targeted specifically for distal recombinants as described in supporting online text. Within the 9.1 kb interval defined by recombinant haplotype analysis, 3 polymorphisms are completely associated with K^B vs. k^Y , as indicated in the upper part of the figure. (B) Exon structure of transcripts within the maximal candidate interval and alignment of selected CBD103 orthologs.

[0015] Figure 7. Expression of beta-defensin mRNA and protein in skin and in cultured keratinocytes. (A) Levels of Agouti, CBD1, or CBD103 mRNA from black or yellow dog skin, as indicated, determined relative to β -actin by quantitative RT-PCR, and expressed as percentage of mRNA present in the yellow samples. Results shown represent the mean $\pm$ s.e.m. of 4 different animals. (B) Expression of epitope (V5)-tagged CBD103 (+) or epitope (V5)-tagged CBD103ΔG23 (ΔG23) in cell layer and media after transfection of mouse keratinocytes as determined by Western blotting using antisera against the V5 epitope. Representative results are shown for 1 of 4 experiments; for each experiment, the 2 constructs were transfected in triplicate or quadruplicate.

[0016] Figure 8. Pigmentary effects of CBD103 in transgenic mice. Photographs of transgenic and non-transgenic littermates, representative of 2/2 independent founders for Tg.CBD103ΔG23 and 20/21 independent founders for Tg.CBD103.

[0017] Figure 9. Pharmacology of beta-defensin action on melanocortin receptors. (A) Ability of NDP-MSH or CBD103 to stimulate cAMP accumulation in cultured melanocytes. (B) Saturation binding of Eu-NDP-MSH to HEK293 cells transiently transfected with the dog Mc1r. (C – F) Competition binding assays in which varying amounts of unlabeled synthetic beta-defensins were added together with Eu-NDP-MSH tracer (at 1.8 nM – 3 nM) to HEK293 cells transiently transfected with dog (d), mouse (m), or human (h) melanocortin receptors, as indicated. In all panels, ligand concentration (either in nM, panel B) or log of the molarity (panels A, C, D, E, and F) is plotted on the abscissa; amount of cAMP formed (panel A) or Eu-NDP MSH bound (panels B-F), measured, respectively, as relative light units (RLUs) or relative fluorescence units (RFUs), is plotted on the ordinate. Each curve represents a single experiment carried out in triplicate; error bars represent +/- s.e.m.

[0018] Figure 10: Comparison of brindle (k^{br}), black (K^B) and yellow (k^y) bearing haplotypes. Haplotypes at 36 SNPs with a m.a.f. >0.1 and the ΔG23 mutation in *CBD103*, spanning 1 Mb interval around *CBD103*. Haplotypes are arranged by similarity and breed. Blue squares represent the allele most commonly found in k^{br} haplotypes. Yellow represents the other allele and white squares represent missing data. In black dogs, k^y and K^B haplotypes were distinguished by the genotype at *CBD103* or S54. In brindle Boxers, the clearly distinct pattern of k^y and k^{br} haplotypes allowed the unambiguous identification of the haplotypes. In brindle from other breeds than Boxers, dogs were chosen for which the genotype k^{br}/k^{br} or k^y/k^{br} was known from breeding records. For the four k^y/k^{br} dogs the haplotype which most closely resembled the k^y haplotypes was determined. The choice was obvious for a Bulldog and a Great Dane but uncertain for a Basenji and an Akita, and these dogs are indicated by a *. The three haplotypes containing a reversion from brindle to yellow are labeled $k^{br \rightarrow y}$ and are identical. The green rectangles indicate the extent of the brindle specific duplication. The gray and red outlines indicate the A and B haplotypes respectively mentioned in the text.

[0019] Figure 11. *CBD103*, *CBD1* and *Agouti* mRNA ratios in the black and yellow stripes of brindle skin. (A) Ratio of total cDNA in the black stripe compared to the yellow stripe of a brindle dog, D6, for *CBD1* (different primer pairs used for the quantitative RT-PCR), *CBD103* (5 different primer pairs) and *Agouti*. There is an average of 2.2 times more *CBD103* RNA in the black stripe than the yellow stripe whereas there is no difference in the RNA levels between stripes for the *CBD1* and *Agouti*. (B) *CBD103* cDNA sequence in the black and yellow stripe of D6. In the black stripe, both the *CBD103* and *CBD103*-ΔG23

alleles are expressed as shown by the double trace 5' of the G repeat in this reverse read. In the yellow stripe only the wild-type *CBD103* allele is expressed.

[0020] Figure 12 is a flow chart for coat color determination.

[0021] Figure 13. Brindle duplication breakpoint mapping using copy number quantification: the size of the brindle duplication was estimated by measuring PCR amplicon copy number at different loci flanking *CBD103*, based on the dog v1.0 genome assembly. Coordinates refer to the distance, in kilobase pairs, from the translational start position of *CBD103*. The positions of copy number markers are indicated by upward ticks, colored either blue (one copy/haploid genome) or red (two copies/haploid genome). Copy number determination is based on the amplification rate of the indicated amplicons relative to single copy amplicons located elsewhere in the genome, using genomic DNA samples from dogs known to be homozygous for the brindle mutation (based on breeding). Broken lines indicate the breakpoint intervals for the brindle duplication. Yellow boxes indicate low copy repeats, as annotated in the dog v1.0 genome build, and are likely to be misassembled in both builds of the dog genome. The positions of known genes in the region are displayed.

[0022] Figure 14. Model for the molecular evolution of the K locus. The wild-type ancestral *CBD103* allele allows coat color to be determined by variation in *Agouti*. (A) The $\Delta G23$ mutation is most likely to have arisen prior to breed formation in the large ancestral dog population. This mutation causes black coat color and is epistatic to variation at *Agouti*. (B) Brindle arose by a rearrangement, potentially through a non-allelic homologous recombination (NAHR) mechanism, between the K^B and k^Y chromosomes in a black dog, leading to a duplication of the defensin cluster at least 1000 years ago. (C) Reversions of the brindle allele to yellow can occur in brindle homozygotes most likely due to NAHR leading to the loss of the defensin cluster carrying the $\Delta G23$ mutation in *CBD103* and a yellow coat color. NAHR between defensin clusters may result in further increases in copy number of the defensin cluster.

DETAILED DESCRIPTION OF THE INVENTION

[0023] Compositions and methods are provided for genotyping canines with respect to the genes that determine coat color, particularly with respect to melanistic phenotype. Interaction of alleles of the K locus with alleles at *Agouti* and *Mc1r* loci determine whether eumelanin or pheomelanin is produced in the coat. Three alleles are identified at the K locus, *yellow* (k^Y), *brindle* (k^{br}), and *black* (K^B). Determination of the presence or absence of these alleles allows a prediction of canine coat color, and finds use in breeding programs for a desired color.

[0024] The present invention is directed to dog coat color prediction with regard to pigment type (eumelanin versus phaemelanin) and the striping pattern known as brindle, through elucidation of the relationship among alleles at three genetic loci – *Agouti*, *Mc1r*, and *CBD103*. The understanding of this relationship is provided in part by identification of the genetic alterations responsible for brindle and dominant black coat color in dogs described herein. Predictions regarding dog coat color can be made with certainty by testing for such genetic alterations, in addition to previously characterized polymorphisms at *Agouti* and *Mc1r*. The invention provides a general methodology for systematic determination of dog coat color phenotypes and molecular assays that can be employed for this purpose. Also provided are applications of a genetic test for dog coat color, such as predicting offspring coat color based on genotype information for one or both breeding parents.

Definitions

[0025] Unless defined otherwise, all technical and scientific terms used herein generally have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Generally, the nomenclature used herein and the laboratory procedures in cell culture, molecular genetics, organic chemistry and nucleic acid chemistry and hybridization described below are those well known and commonly employed in the art. Standard techniques are used for nucleic acid and peptide synthesis. Generally, enzymatic reactions and purification steps are performed according to the manufacturer's specifications. The techniques and procedures are generally performed according to conventional methods in the art and various general references (see generally, Sambrook et al. MOLECULAR CLONING: A LABORATORY MANUAL, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.), which are provided throughout this document. The nomenclature used herein and the laboratory procedures in analytical and organic chemistry described below are those well known and commonly employed in the art. Standard techniques, or modifications thereof, are used for chemical syntheses and chemical analyses.

[0026] *Melanin* is any of the polyacetylene, polyaniline, and polypyrrole "blacks" and "browns" or their mixed copolymers. The most common form of biological melanin is a polymer of either or both of two monomer molecules: indolequinone, and dihydroxyindole carboxylic acid. Because melanin is an aggregate of smaller component molecules, there are a number of different types of melanin with differing proportions and bonding patterns of these component molecules. Both pheomelanin and eumelanin are found in canine coat hair.

[0027] *Eumelanin* polymers have long been thought to comprise numerous cross-linked 5,6-dihydroxyindole (DHI) and 5,6-dihydroxyindole-2-carboxylic acid (DHICA) polymers;

recent research into the electrical properties of eumelanin, however, has indicated that it may consist of more basic oligomers adhering to one another by some other mechanism. The two primary types are black eumelanin and brown eumelanin. A eumelanistic animal produces eumelanin in areas that contain pigment cells. Eumelanistic dogs may be referred to as black; however the colors encompassed by the term include black, chocolate, brown, liver, blue, harlequin, mantle, bicolor (black and white), and spotted black dogs. In the presence of functional *Agouti* and *Mc1r* alleles, the K^B allele causes exclusive production of eumelanin in areas that contain pigment cells, i.e. a eumelanistic phenotype.

[0028] *Pheomelanin* differs from eumelanin in that its oligomer structure incorporates the amino acid L-cysteine, as well as DHI and DHICA units. It confers a yellow or red color. The term "yellow" may be used to refer to any pheomelanistic dog, however the colors encompassed by the term include yellow, fawn, red, apricot, cream, wheaten, tricolor (black, yellow and white), black-and-tan, tan point, and sable. The k^y allele is recessive, and permits production of pheomelanin in some or all body regions depending on the genotype of *Agouti* and the presence of pigment cells.

[0029] *Brindle*. In brindle dogs, stripes of yellow or red color are juxtaposed to stripes of black or brown color. The brindle dog stripes often form a V-shaped pattern at the dorsal midline and may form a swirling pattern on the face. The number of stripes and overall darkness of the coats varies between breeds, within breeds, and within litters. Brindle stripes are caused by melanocytes switching between eumelanin and pheomelanin pigment production in different parts of the coat. The brindle pattern requires the presence of pigment cells and the expression of *Agouti* to be observed. It may therefore be visible on the entire coat or be restricted to distinct regions of the coat, for example ventral areas in a dog carrying the a^l allele of *Agouti*.

[0030] *Beta-defensin* (CBD103, K locus). The "K locus" as used herein refers to the coding sequence of CBD103, provided herein as SEQ ID NO:1, and flanking sequences, including regulatory sequences such as promoter, enhancers, etc. The term also includes polymorphic variants of the locus, which are identified herein as variants in the coding sequence, and a duplication of the coding sequence. The locus may include genomic sequence flanking the coding sequence at the 3' and 5' ends. The brindle duplication extends between 126,819-129,965 bp upstream of CBD103 translational start position (3146 bp interval) and between 2621-9676 bp downstream of CBD103 translational start position (7055 bp interval), and includes a low copy repeat within the duplication.

[0031] The dominant allele at the K locus is K^b , or dominant black, which results in a eumelanin coat color in a permissive background. The recessive alleles at this locus are K^y , which results in a pheomelanistic coat; and K^{br} , which is recessive to K^b and dominant to k^y ,

and which provides for a brindle coat pattern. Brindle occurs over the entire body in dogs with an a^y allele but only on the ventral surfaces in dogs with an a^l/a^l genotype.

[0032] *Melanocortin 1 receptor*. *Melanocortin 1 receptor (MC1R)* was mapped to CFA5. The major alleles are the wild-type, sometimes designated $Mcr1R^E$; and $Mcr1R^e$, which is a loss-of-function mutation, 914C>T that causes an arginine to be replaced by a premature stop codon (R306ter), resulting in a pheomelanistic phenotype. A third allele E^M is caused by a single nucleotide substitution (799A>G) resulting in a M264V amino acid change. The melanistic mask caused by one copy of this allele is only visible on dogs that are fawn or brindle. Example of $Mcr1r$ primers to genotype R306ter are 5'-TCTTTGTAGCCATGCTGGTG-3' and 5'-ATCCACCACACCACAGATCA-3' (486 bp amplicon).

[0033] *Agouti*. *Agouti signal peptide (ASIP)* has several alleles that are involved in coat colour in dogs. *ASIP* has been mapped to CFA24 between *AHT 118* and *AHT 125*. There are four alleles present in dogs in the dominance hierarchy $a^y > a^w > a^l > a$. The wild type allele (a^w) causes some hairs to have a band of eumelanin, phaeomelanin, eumelanin pigment from base to tip. These banded hairs are typically along the dorsal region of the torso. The *ASIP* sequence of this allele has complete homology to the wolf sequence. There is a recessive allele which causes dogs to have a black coat, the R96C allele (a) (c.288C>T). The a^y allele, is dominant allele in this series. The allele has two amino acid changes in comparison to the wild type, A82S and R83H (c.246G>T and c.250G>A), and is associated with a dominant pheomelanistic phenotype. A fourth allele, known as a^l , is believed to exist in dogs that are black-and-tan. No differences have been observed in the coding sequence, and the allele likely differs in a regulatory regions.

[0034] Example of *Agouti* primers to genotype the R96C mutation (from Kerns et al Mammalian Genome 2004) are 5'-GATGTCTGGTCTGGAGCCTC -3' and 5'-CCTTCTCAAAGTCCCATCTC-3' (620 bp amplicon).

[0035] For the purposes of this invention, the "dominant yellow" *Agouti* allele, a^y , the rare wild-type a^w allele and the a^l allele may be grouped together and designated A , while the loss of function R96C allele is designated a .

[0036] *Tyrosinase related protein 1*. *Tyrosinase related protein 1 (TYRP1)* is the gene causing brown coat color in dogs. *TYRP1* was mapped to CFA 11 between microsatellites *C03109* and *FH2004*. The wild-type allele is dominant, and results in a black eumelanin. The recessive alleles produce a brown colored eumelanin, which alleles include b^s (Q331ter; c.991C>T); b^d (345delP, c.1033-6 deleted) and b^c (S41C; c.121T>A). The *TYRP1* alleles interact with the *MC1R* alleles.

[0037] *Melanophilin*. Alleles can cause a dilution in the coat color, where there is a modification of both eumelanin and phaeomelanin pigmented areas to a paler shade. This is apparently distinct from the progressive gray phenotype.

[0038] The terms "isolated", "purified" or "biologically pure" refer to material that is substantially or essentially free from components that normally accompany it as found in its native state. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein that is the predominant species present in a preparation is substantially purified. In particular, an isolated CBD103 nucleic acid may be separated from open reading frames that flank the gene and encode proteins other than CBD103. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

[0039] The terms "nucleic acid" and "polynucleotide" are used interchangeably herein to refer to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs).

[0040] Unless otherwise indicated, a particular nucleic acid sequence also encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

[0041] The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms also apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymer.

[0042] A "label" or "detectable label" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include fluorophores, radioisotopes (e.g., ^3H , ^{35}S , ^{32}P , ^{51}Cr , or ^{125}I), electron-dense reagents, enzymes (e.g., alkaline phosphatase, horseradish peroxidase, or others

commonly used in an ELISA), biotin, digoxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available.

[0043] An "amplification reaction" refers to any chemical reaction, including an enzymatic reaction, which results in increased copies of a template nucleic acid sequence. Amplification reactions include polymerase chain reaction (PCR) and ligase chain reaction (LCR) (see U.S. Pat. Nos. 4,683,195 and 4,683,202; PCR Protocols: A Guide to Methods and Applications (Innis et al., eds, 1990)), strand displacement amplification (SDA) (Walker, et al. Nucleic Acids Res. 20(7):1691 (1992); Walker PCR Methods Appl 3(1):1 (1993)), transcription-mediated amplification (Phyffer, et al., J. Clin. Microbiol. 34:834 (1996); Vuorinen, et al., J. Clin. Microbiol. 33:1856 (1995)), nucleic acid sequence-based amplification (NASBA) (Compton, Nature 350(6313):91 (1991), rolling circle amplification (RCA) (Lisby, Mol. Biotechnol. 12(1):75 (1999)); Hatch et al., Genet. Anal. 15(2):35 (1999)) and branched DNA signal amplification (bDNA) (see, e.g., Iqbal et al., Mol. Cell Probes 13(4):315 (1999)).

[0044] The term "primer" refers to a nucleic acid sequence that primes the synthesis of a polynucleotide in an amplification reaction. Typically a primer comprises fewer than about 100 nucleotides and preferably comprises fewer than about 30 nucleotides. Exemplary primers range from about 5 to about 25 nucleotides.

[0045] As used herein a "nucleic acid probe or oligonucleotide" is defined as a nucleic acid capable of binding to a target nucleic acid of complementary sequence through complementary base pairing. Probes may be labeled as with isotopes, chromophores, lumiphores, chromogens, or indirectly labeled such as with biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the target sequence.

[0046] The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, for example at least about 60% identity, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, or at least about 95% identity over a specified region when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially identical." This definition also refers to the complement of a test sequence. For sequence comparison, one sequence may act as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm

program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[0047] A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by manual alignment and visual inspection (see, e.g., Current Protocols in Molecular Biology (Ausubel et al., eds. 1995 supplement)).

[0048] The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 5-10° C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium).

[0049] "Genetic sample" as used herein is a sample of biological tissue or fluid that genomic DNA or mRNA of the individual from which the sample is obtained. These samples can be tested by the methods described herein and include whole blood, saliva, cell extracts, cheek swabs, hairs including cells from the base, cell cultures; fixed tissue specimens; and fixed cell specimens. A biological sample is obtained from any canine, usually a domestic dog, *Canis familiaris*. A biological sample may be suspended or dissolved in liquid materials such as buffers, extractants, solvents and the like. The term genetic sample may also be used to refer to the processed genetic material obtained from cells, e.g. purified and/or amplified DNA, mRNA, etc.

ANALYSIS OF CANINE COAT COLOR AND PHENOTYPE PREDICTION

[0050] In some embodiments of the invention, methods are provided for predicting the coat color of a dog based on the alleles present at the K locus, usually by determining the genotype at the locus, although for some purposes it may be sufficient to characterize the CBD103 protein. Such methods find various uses, including the design of a breeding program, where the coat color of the progeny of a candidate mating pair can be predicted based on the genotype and sorting of the parental alleles. For any such purpose, it is useful to genotype one or more individuals. As the K alleles interacts with other loci involved in canine coat color, in many cases it will be useful to obtain a genotype at multiple loci, particularly including the agouti and Mc1R loci. Where the dogs are predicted to be eumelanistic, it may be useful to genotype the *TYRP1* locus as well. As certain alleles are epistatic to other alleles, a flow chart for determination of an animal as pheomelanistic or eumelanistic may be utilized, for example as shown in Figure 12.

[0051] Eumelanistic and brindle phenotypes require a functional Mc1R to be present. The absence of a functional Mc1R is associated with the *Mc1R^e* allele, which is a loss-of-function mutation, 914C>T that causes an arginine to be replaced by a premature stop codon (R306ter). In phenotype prediction, a homozygote for *Mc1R^e* is unable to have a eumelanistic phenotype, regardless of the genotype at the *K* or *Agouti* loci. Genotypes that provide for a functional Mc1R are homozygous or heterozygous for the wild-type allele, *Mc1R⁺* and are permissive for eumelanistic, pheomelanistic and brindle phenotype. Methods of determining the genotype at the Mc1R locus are known in the art, for example as described by Newton *et al.* (2000) *Mammalian Genome* 11, 24–30; Schmutz *et al.* (2001) *Animal Genetics* 32, 43–4; Schmutz *et al.* (2003) *Journal of Heredity* 94, 69–73, herein specifically incorporated by reference. Presence of the *e* allele may be determined by hybridization of a primer specific for the mutation in a genetic sample, RFLP, *etc.*

[0052] Three alleles are described herein for the K locus, *K^B*, *K^Y* and *K^{br}*, which alleles may be determined as described in the following section. Where a functional Mc1R is present, the K locus can be epistatic to the Agouti locus. At the K locus, where the genotype provides for at least one copy of the “dominant black” allele, *K^B*, the animal will have a eumelanistic phenotype, which is dominant over Agouti.

[0053] In an animal that lacks at least one *K^B* allele, the phenotype is determined in part by the genotype of the Agouti locus. In an animal that is homozygous for the *K^Y* allele, it will be eumelanistic with an Agouti *a/a* genotype; and pheomelanistic in the presence of at least one dominant yellow or wild-type Agouti allele, i.e. *A/a* or *A/A*. Methods of determining the presence of the *a* Agouti allele are known in the art, for example as described by Kerns *et al.* (2004) *Mammalian Genome* 15, 798–808, herein specifically incorporated by reference.

- [0054] An animal that is homozygous for the brindle allele at the *K* locus, K^{br}/K^{br} , or that is heterozygous K^{br}/K^y will be eumelanistic with an Agouti *a/a* genotype; and brindle in the presence of at least one dominant yellow or wild-type Agouti allele, i.e. *A/a* or *A/A*.
- [0055] Eumelanistic coats may be further modified by the genotype of the animal at the *TYRP1* locus, where the wild-type black eumelanin is dominant, and the recessive alleles b^s , b^d and b^c provide for a brown color eumelanin. Both pheomelanistic and eumelanistic coats are apparently modified by the melanophilin locus.
- [0056] The methods of the present invention involve genotyping at least one canine genetic sample at the *K* locus, and determining the presence of at least one K^B allele. The methods may additionally comprise determining the presence of at least one K^{br} allele. Determining the presence of at least one K^y allele may also be performed. While the K^y allele genotype is not required for prediction of coat color in an animal (as the information from the K^B and K^{br} genotype is sufficient), it may be determined for predicting the outcome in mating crosses. These genotyping determinations may be generically referred to "determining the genotype at the *K* locus", and shall refer to determining the presence of one, two or three different alleles, where the presence of the allele may be determined, and additionally the copy number of the allele may be determined.
- [0057] Different breeds carry a subset of *K* alleles and this information can be used to help determine a testing strategy and interpret genotype results for the G23 mutation. For example, Boxers only carry k^{br} and k^y therefore testing for K^B is unnecessary in that breed. Poodles do not carry k^{br} therefore testing for the brindle duplication in that breed is unnecessary and a $\Delta G23/+$ genotype is interpretable as a K^B/k^y genotype.
- [0058] In some embodiments, the methods of the invention further involve genotyping at the *Mc1R* locus. The analysis may determine the presence of at least one wild-type *Mc1R* allele, and further involve determining the presence of one or two copies of the $Mc1R^o$ allele.
- [0059] In some embodiments, the methods of the invention further involve genotyping at the Agouti locus. The analysis may determine the presence of at least one *A* allele, and further involve determining the presence of one or two copies of the *a* allele.
- [0060] In some embodiments, the methods of the invention further involve genotyping at one or both of the melanophilin and *TYRP1* loci.
- [0061] The methods of the invention may further provide an analysis of progeny coat color that may be obtained from breeding the animal that is being genotyped, in combination with a mate of a known or unknown genotype. Such analysis may include the probability that progeny will have a phenotype of interest, e.g. based on the independent

assortment of alleles at the K, Mc1R and Agouti loci, and may further include additional loci involved in coat color determination.

[0062] The methods of the invention require genotyping at least a first canine genetic sample. In some embodiments the methods further comprise genotyping a second canine genetic sample. The first and the second genetic samples may be obtained from canines of different sexes, e.g. where a mating cross is contemplated. The methods may further provide an analysis of progeny coat color that may be obtained from breeding the first canine with the second canine. Such analysis may include the probability that progeny will have a phenotype of interest, e.g. based on the independent assortment of alleles at the K, Mc1R and Agouti loci, and may further include additional loci involved in coat color determination.

METHODS OF K LOCUS TYPING

[0063] The alleles present at the K locus in a canine sample are determined. As described above, three alleles are described herein. The K^Y allele has the sequence set forth in SEQ ID NO:1. The KB allele has the sequence set forth in SEQ ID NO:3, in which there is a three nucleotide deletion, resulting in an in-frame deletion in the amino acid sequence of glycine 23. The alleles may be distinguished genetically by identifying the sequence at the polymorphic residues, e.g. by hybridization, size of an amplification product, sequencing, etc. The protein products of the two alleles are different, and distinction may be made between the two alleles based on determination of the protein by a specific antibody, by affinity for Mc1R, etc.

[0064] The K^{br} allele is a duplication of the CBD103 gene. In brindle, one copy of the defensin cluster carries the wild-type *CBD103* allele and the other copy carries the *CBD103-ΔG23* allele. The brindle phenotype consists of adjacent stripes of yellow and black hair in a clonal pattern as a result of variegation of expression of the *CBD103-ΔG23* allele due to the presence of the duplication. The presence of the K^{br} allele can be determined by determining copy number of CBD103, where a copy number of 3 is indicative of a brindle allele heterozygote, and the presence of a copy number of 4 is indicative of a brindle allele homozygote. Alternatively, the presence of the K^{br} allele can be determined by size of the CBD103 region, e.g. by amplification with a primer that lies outside of the duplicated region, by sequencing of the region, e.g. in an amplification product, by hybridization to a duplication specific sequence, and the like.

[0065] Amplification techniques using primers can also be used to amplify and isolate CBD103 DNA or RNA. For example, nucleic acids encoding CBD103 or fragments thereof may be obtained by amplification of a canine DNA or reverse transcribed canine mRNA

from CBD103 expressing cells, e.g. a skin sample, hair sample, etc. using isolated nucleic acid primer pairs that flank the CBD103 Δ G23 polymorphism, e.g. as disclosed in the examples; using the primer pair (SEQ ID NO:5) TGTCTTCATCCCTGTGAGGT and (SEQ ID NO:6) CCAGGAGGCATTTTCACACT. In general, PCR and other methods of amplification use primers which anneal to DNA sequences flanking the region of interest. Such primers can be used, e.g., to amplify either the full length sequence or a region of one to several hundred nucleotides. The sequence of CBD103 at either side of the K^B mutation is as follows (SEQ ID NO:289):

GGAAGACATCAGACATCTGGGGAATTCCAAAAGCCTTGTACTCGGA[C/G]CAAAGCCTT
CCTAAAGCTTTCCGGCACGTTCTGTTTTTTTCTTTGCAGGAAATGGGGGAATTATAAAT
ACCCTGCAGAGGTATTATTGCAGAATAAGGAGCGGTGCGCCTTG[C/T]TGAGCT
GCCTGCCAAAGGAGGAGCAGATAGGCCGCTGTT, where SNP polymorphisms are depicted in brackets, and the region of the K^B mutation is underlined.

[0066] It will be understood by one of skill in the art that any number of primer pairs can be designed to amplify the CBD103 region, based on the known flanking sequences. The primary requirement is that the amplification encompasses at least the Δ G23 sequence if the detection is to be made by sequencing, sizing, etc. In some embodiments, an amplification primer is designed to hybridize to the Δ G23 sequence, such that only genetic material having the deletion, or only genetic material lacking the deletion, is amplified. In other embodiments all or part of the coding sequence is amplified, and determination of the allele is made by hybridization to a deletion specific probe; sequencing; etc.

[0067] Detection techniques for evaluating nucleic acids for the presence of a small deletion, duplication or single base change involve procedures well known in the field of molecular genetics. Ample guidance for performing the methods is provided in the art. Exemplary references include manuals such as PCR Technology: PRINCIPLES AND APPLICATIONS FOR DNA AMPLIFICATION (ed. H. A. Erlich, Freeman Press, NY, N.Y., 1992); PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS (eds. Innis, et al., Academic Press, San Diego, Calif., 1990); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, 1994-1999, including supplemental updates through April 2004; Sambrook & Russell, Molecular Cloning, A Laboratory Manual (3rd Ed, 2001).

[0068] Methods for detecting genetic polymorphisms well known in the art often entail one of several general protocols: hybridization using sequence-specific oligonucleotides, primer extension, sequence-specific ligation, sequencing, or electrophoretic separation techniques, e.g., singled-stranded conformational polymorphism (SSCP) and heteroduplex analysis. Exemplary assays include 5' nuclease assays, template-directed dye-terminator incorporation, molecular beacon allele-specific oligonucleotide assays, single-base extension assays, and SNP scoring by real-time pyrophosphate sequences. Analysis of

amplified sequences can be performed using various technologies such as microchips, fluorescence polarization assays, and matrix-assisted laser desorption ionization (MALDI) mass spectrometry. In addition to these frequently used methodologies for analysis of nucleic acid samples to detect single base changes, any method known in the art can be used to detect the presence of the CBD103 variants described herein.

[0069] Although such methods often employ PCR steps, other amplification protocols may also be used. Suitable amplification methods include ligase chain reaction (see, e.g., Wu & Wallace, *Genomics* 4:560-569, 1988); strand displacement assay (see, e.g., Walker et al., *Proc. Natl. Acad. Sci. USA* 89:392-396, 1992; U.S. Pat. No. 5,455,166); and transcription-based amplification systems, including the methods described in U.S. Pat. Nos. 5,437,990; 5,409,818; and 5,399,491; the transcription amplification system (TAS) (Kwoh et al., *Proc. Natl. Acad. Sci. USA* 86:1173-1177, 1989); and self-sustained sequence replication (3SR) (Guatelli et al., *Proc. Natl. Acad. Sci. USA* 87:1874-1878, 1990; WO 92/08800). Alternatively, methods that amplify the probe to detectable levels can be used, such as Q β -replicase amplification (Kramer & Lizardi, *Nature* 339:401-402, 1989; Lomeli et al., *Clin. Chem.* 35:1826-1831, 1989).

[0070] In some embodiments, the CBD103 allele(s) are detected using oligonucleotide primers and/or probes. Oligonucleotides can be prepared by any suitable method, including chemical synthesis. Oligonucleotides can be synthesized using commercially available reagents and instruments. Alternatively, they can be purchased through commercial sources. Methods of synthesizing oligonucleotides are well known in the art.

[0071] Polynucleotides that are used as probes may be labeled with radioisotopes, chemiluminescent moieties, or fluorescent moieties. Such labels are useful for the characterization and detection of genetic sequences, including amplification products using the methods and compositions of the present invention.

[0072] Amplified products can be detected using any means known in the art, including, e.g., hybridization to a probe of interest; restriction fragment length polymorphism (RFLP) analysis; denaturing gel electrophoresis, direct sequencing, and HPLC-based analysis.

[0073] Target polymorphic CBD103 sequences can be differentiated using single-strand conformation polymorphism analysis, which identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described, e.g., in Orita et al., *Proc. Nat. Acad. Sci.* 86, 2766-2770 (1989). Amplified PCR products can be generated using methods known in the art, and heated or otherwise denatured, to form single stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures which are partially dependent on the base sequence. The different

electrophoretic mobilities of single-stranded amplification products can be related to base-sequence difference between polymorphic CBD103 sequences.

[0074] A technique commonly referred to as allele specific oligonucleotide hybridization (ASO) (e.g., Stoneking et al., *Am. J. Hum. Genet.* 48:70-382, 1991; Saiki et al., *Nature* 324, 163-166, 1986) can be used to detect variant CBD103 genes. Two allelic DNA molecules differing are distinguished by hybridizing an oligonucleotide probe that is specific for one of the variants (e.g., wild type or mutant CBD103) to an amplified product obtained from amplifying the nucleic acid sample. The probes are designed to differentially hybridize to one variant versus another. Principles and guidance for designing such probes is available in the art (see, e.g., Jeffrys and Mays, *Genome Res.* 13(1): 2316-2324 (2003) and Howell et al., *Nature Biotech* 17(1): 87-88 (1999)). Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the sequences.

[0075] The presence of a CBD103 allele can be determined by measuring the amount of sequence-specific oligonucleotide that is hybridized to the sample. Typically, the oligonucleotide is labeled with a label such as a fluorescent label. For example, a mutant CBD103-specific oligonucleotide is applied to immobilized oligonucleotides representing CBD103 sequences. After stringent hybridization and washing conditions, fluorescence intensity is measured for each CBD103 oligonucleotide.

[0076] Suitable assay formats for detecting hybrids formed between probes and target nucleic acid sequences in a sample are known in the art and include the immobilized target (dot-blot) format and immobilized probe (array, microarray, reverse dot-blot or line-blot) assay formats. Examples of nucleic acid arrays are described by WO 95/11995. The same array or a different array can be used for analysis of characterized polymorphisms. WO 95/11995 also describes subarrays that are optimized for detection of variant forms of a pre-characterized polymorphism. Such a subarray can be used in detecting the presence of the mutant CBD103 gene described herein.

[0077] Polymorphisms are also commonly detected using sequence-specific amplification or primer extension methods. These reactions typically involve use of primers that are designed to specifically target a polymorphism via a mismatch at the 3' end of a primer. The presence of a mismatch affects the ability of a polymerase to extend a primer when the polymerase lacks error-correcting activity. For example, a primer complementary to the wild type or K^B CBD103 gene is designed such that the 3' terminal nucleotide hybridizes at the gly23 site. The presence of the K^B variant can be determined by the ability of the primer to initiate extension. If the 3' terminus is mismatched, the extension is impeded. Typically, the primer is used in conjunction with a second primer in an amplification reaction. The second

primer hybridizes at a site unrelated to the polymorphic position. Amplification proceeds from the two primers leading to a detectable product signifying the particular allelic form is present. Sequence-specific amplification- or extension-based methods are described in, for example, WO 93/22456; U.S. Pat. Nos. 5,137,806; 5,595,890; 5,639,611; and U.S. Pat. No. 4,851,331.

[0078] Using sequence-specific amplification-based genotyping, identification of the alleles requires only detection of the presence or absence of amplified target sequences. Methods for the detection of amplified target sequences are well known in the art. For example, gel electrophoresis and probe hybridization assays described are often used to detect the presence of nucleic acids.

[0079] As appreciated by one in the art, sequence-specific amplification methods can be performed in reactions that employ multiple sequence-specific primers to target particular alleles, e.g. a combination of primers specific for the K locus, and one or more of the Mc1R locus, the Agouti locus, *etc.* Primers for such multiplex applications are generally labeled with distinguishable labels or are selected such that the amplification products produced from the target sequences are distinguishable by size. Thus, for example, the presence of a K allele, and alleles from other loci involved in coat color determination in a single sample can be identified using a single amplification by gel analysis of the amplification product.

[0080] Genotyping can also be performed using a "TaqMan™" or "5'-nuclease assay", as described in U.S. Pat. Nos. 5,210,015; 5,487,972; and 5,804,375; and Holland et al., 1988, Proc. Natl. Acad. Sci. USA 88:7276-7280. In the TaqMan™ assay, labeled detection probes that hybridize within the amplified region are added during the amplification reaction. The probes are modified so as to prevent the probes from acting as primers for DNA synthesis. The amplification is performed using a DNA polymerase having 5' to 3' exonuclease activity. During each synthesis step of the amplification, any probe which hybridizes to the target nucleic acid downstream from the primer being extended is degraded by the 5' to 3' exonuclease activity of the DNA polymerase. Thus, the synthesis of a new target strand also results in the degradation of a probe, and the accumulation of degradation product provides a measure of the synthesis of target sequences. The hybridization probe can be a sequence-specific probe that discriminates between K locus alleles. Alternatively, the method can be performed using a sequence-specific primer and a labeled probe that binds to amplified product.

[0081] Any method suitable for detecting degradation product can be used in a 5' nuclease assay. Often, the detection probe is labeled with two fluorescent dyes, one of which is capable of quenching the fluorescence of the other dye. The dyes are attached to the probe, preferably one attached to the 5' terminus and the other is attached to an internal site, such that quenching occurs when the probe is in an unhybridized state and such that

cleavage of the probe by the 5' to 3' exonuclease activity of the DNA polymerase occurs in between the two dyes. Amplification results in cleavage of the probe between the dyes with a concomitant elimination of quenching and an increase in the fluorescence observable from the initially quenched dye. The accumulation of degradation product is monitored by measuring the increase in reaction fluorescence. U.S. Pat. Nos. 5,491,063 and 5,571,673 describe alternative methods for detecting the degradation of probe which occurs concomitant with amplification.

[0082] In some cases, mRNA can also be used for genotyping, provided that the mRNA is obtained from a cell that expresses CBD103, e.g. hair follicle cells. Such an analysis can be performed by first reverse-transcribing the target RNA from a biological sample from the dog using, for example, a viral reverse transcriptase, and then amplifying the resulting cDNA; or using a combined high-temperature reverse-transcription-polymerase chain reaction (RT-PCR), as described in U.S. Pat. Nos. 5,310,652; 5,322,770; 5,561,058; 5,641,864; and 5,693,517.

Kits

[0083] In some embodiment, kits are provided for practicing the methods of the invention. Such kits typically include oligonucleotide primers for specific amplification or specific hybridization to alleles of the K locus, and may further include reagents for detection of alleles at other coat color loci. Such kits may further comprise instructions for use. Kits may comprise positive genetic samples for each of the alleles of interest. Kits may comprise buffers, enzymes and the like for performance of the methods. Kits may include a polynucleotide having the sequence set forth in SEQ ID NO:3, or a fragment thereof of at least about 15, at least about 20, at least about 25 nucleotides or more that comprising the Δ G23 coding sequence, *i.e.* spanning at least SEQ ID NO:1 at the five G residues at positions 177-181, or the corresponding region of SEQ ID NO:3.

[0084] The invention also provides kits and solutions for carrying out the amplification methods of the invention. For example, the invention provides kits that include one or more reaction vessels that have aliquots of some or all of the reaction components of the invention in them. Aliquots can be in liquid or dried form. Reaction vessels can include sample processing cartridges or other vessels that allow for the containment, processing and/or amplification of samples in the same vessel. Such kits allow for ready detection of amplification products of the invention into standard or portable amplification devices.

[0085] Kits can include, for instance, amplification reagents comprising primers sufficient to amplify at least two different target sequences, a polynucleotide sequence comprising the sequences of the primers or subsequences of the primers as described herein; and at least

one probe for amplifying and detecting the polynucleotide sequence. In addition, the kit can include nucleotides (e.g., A, C, G and T), a DNA polymerase and appropriate buffers, salts and other reagents to facilitate amplification reactions.

EXAMPLES

[0086] Each publication cited in this specification is hereby incorporated by reference in its entirety for all purposes.

[0087] The following examples are offered to illustrate, but not to limit the claimed invention. It is to be understood that this invention is not limited to the particular methodology, protocols, cell lines, animal species or genera, and reagents described, as such may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention, which will be limited only by the appended claims.

[0088] As used herein the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells and reference to "the culture" includes reference to one or more cultures and equivalents thereof known to those skilled in the art, and so forth. All technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs unless clearly indicated otherwise.

Example 1

Linkage and segregation analysis of black and brindle coat color in domestic dogs.

MATERIALS AND METHODS

[0089] *DNA samples and pedigrees.* Genomic DNA from blood or cheek swab samples was isolated according to standard procedures. Pedigrees were established, and in all cases, pedigree relationships were verified by determining that multiple markers exhibited Mendelian segregation in accord with expectations.

[0090] *Genotyping, statistical analysis, and genomics.* Genotyping for the minimal screening set I panel (MSSI) of simple sequence length repeat (SSLP) markers was carried out using multiplex PCR as previously described Cargill et al. (2002) *Genomics* 80:250-253; Clark et al. (2004) *Genomics* 84:550-554). Fluorescently labeled PCR products were separated on an automated laser fluorescence DNA sequencer ABI377 (Perkin-Elmer), using GENESCAN (version 2.1) fragment analysis software, and alleles identified using the GENOTYPER program (version 2.0; Perkin Elmer).

[0091] Prior to linkage analysis, Mendelian error-checking was performed. The data were then analyzed under a model of autosomal dominant inheritance for black vs. non-black assuming complete penetrance. Two-point linkage analyses were carried out using the MLINK (to generate LOD scores at different theta values) and ILINK programs (to maximize LOD) from the LINKAGE 5.1 package. Given the small number of animals in the scan, further analysis of additional markers was done manually, to determine the critical region and to infer haplotypes, as depicted in Figures 1, 2, and 3.

[0092] To infer the epistasis relationships between *Agouti*, *Mc1r*, and *K* alleles, we determined the *Agouti* and *Mc1r* coding sequence by sequencing PCR-amplified fragments of genomic DNA. Primer sequences have been described previously. We determined *K* genotypes using linkage and pedigree analysis (as described below) and, in some cases, by using additional markers that are in linkage disequilibrium with *K* locus alleles and which will be described elsewhere. Genotypes for all three loci were determined (by resequencing *Agouti* and *Mc1r*, or by genotyping flanking markers for *K* as described above) for every individual depicted in Figures 1, 2, and 3. This included 35 animals from the Cornell Labrador Retriever x Greyhound cross, 10 Afghan Hounds, 8 Great Danes, and 10 Staffordshire Bull Terriers. At least 5 animals from each of 4 additional breeds: German Shepherd Dogs, French Bulldogs, Boxers, and Poodles were also genotyped for all three loci as described in Table 3.

[0093] The physical location of markers and consideration of candidate genes is based on the CanFam1.0 assembly of the dog whole genome sequence available through the UCSC Genome Browser.

RESULTS

[0094] *Nomenclature.* Historically, it was recognized that at least two different genes could give rise to a uniform pheomelanin coat, and the term "fawn" was used to refer to the phenotype caused by the a^y allele of *Agouti*, distinguished from a similar phenotype caused by a loss-of-function *Mc1r* allele, originally known as *recessive yellow* or *e*, and now known to represent *Mc1r*^{R306ter}. In some cases, differential gene action was inferred on the basis of the phenotype, with homozygosity for *Mc1r*^{R306ter} giving rise to a clear or diluted yellow color, and the a^y allele of *Agouti* referring to a deeper, often dark-tinged shade of red-yellow. In hindsight, the effects of *Agouti* and *Mc1r* alleles cannot always be distinguished by virtue of their phenotype; also the *K* locus genotype is equally important in determining the balance and distribution of pheomelanin vs. eumelanin. To reconcile the historical terms with both common usage and modern genetics, we propose that alleles of the *K* locus be designated as *yellow* (k^y), *brindle* (k^{br}), and *black* (K^B). A summary of this nomenclature that relates the historical terms to those used here and the underlying genetics is given in Table 1.

Table 1. Phenotype—genotype relationships for *Agouti*, *K*, and *Mc1r*¹

Common name	Phenotype and breed example	Historical names (symbols)	Possible genotypes based on current work ²		
			<i>Agouti</i>	<i>K</i> locus	<i>Mc1r</i>
Dominant black or "Self-colored"	Uniformly black, can be modified to brown or by white spotting: Newfoundland, black or brown Labrador Retriever	<i>dominant black, Agouti-Self (A^S)</i>	<i>a^y/a</i> <i>a^y/a^t</i> <i>a^y/a</i>	<i>K^B/K^B</i> <i>K^B/k^y</i> <i>K^B/k^{br}</i>	<i>+/+</i> <i>+/R306ter</i>
Recessive yellow	Uniformly red-yellow, can be modified to pale yellow or cream: yellow Labrador Retriever, Irish Setter, Samoyed	<i>recessive yellow, extension (e)</i>	<i>a^y, a^t, a</i> (all combinations)	<i>K^B, k^{br}, k^y</i> (all combinations)	<i>R306ter/R306ter</i>
Fawn	Red-tan, can be dark-tinged: Great Dane, fawn (i.e. yellow) Boxer	<i>dominant yellow, golden sable (a^y)</i>	<i>a^y/a</i> <i>a^y/a^t</i> <i>a^y/a</i>	<i>k^y/k^y</i>	<i>+/+</i> <i>+/R306ter</i>
Black-and-tan	Dobermann Pinscher	<i>tan points (a^t)</i>	<i>a^t/a</i> <i>a^t/a</i>	<i>k^y/k^y</i>	<i>+/+</i> <i>+/R306ter</i>
Brindle	Black- and yellow-colored stripes: brindle French Bulldog, brindle Boxer	<i>brindle, partial extension (e^{br})</i>	<i>a^t/a</i> <i>a^y/a^t</i> <i>a^y/a</i>	<i>k^{br}/k^{br}, k^{br}/k^y</i>	<i>+/+</i> <i>+/R306ter</i>
Recessive black	Uniformly black: black German Shepherd Dog	<i>recessive black (a)</i>	<i>a/a</i>	<i>K^B, k^{br}, k^y</i> (all combinations)	<i>+/+</i> <i>+/R306ter</i>

¹Explanations and references for names and symbols are given in the text.

²Possible genotypes according to epistasis relationships as described in the text and in Table 3. Only 3 *Agouti* alleles are considered for the sake of simplicity; the *a^w* allele would behave identically to the *a^t* allele. Also, the "+" allele at *Mc1r* is used to designate any *Mc1r* allele other than R306ter (also known as *recessive yellow* or *e*).

[0095] *Genome-wide linkage scan and fine mapping for dominant black.* In a Labrador Retriever x Greyhound cross that was established at Cornell University to study hip dysplasia, a subset of kindreds exhibit transmission of coat color variation in a pattern that is consistent with inheritance of dominant black. Black Labrador Retrievers crossed to yellow or brindle Greyhounds invariably yield black F1 offspring; when an F1 animal is backcrossed to the Greyhound parent, backcross progeny exhibit 1:1 segregation of black to non-black.

[0096] In previous studies of the EB and GB kindreds from this cross, we observed that variation at neither *Mc1r* nor *Agouti* could account for dominant black (as an allele of the putative *K* locus); therefore we carried out a genome-wide linkage scan of the same kindreds using a dense panel of highly informative SSLP markers. For the initial screen of

19 animals (Figure 1), 125 of 155 markers from the "minimal screening set" were informative. We analyzed the results by two point linkage analysis under a model of dominant inheritance with complete penetrance, and observed that there were three loci on chromosomes 4 and 16 (CFA4, CFA16) that exceeded a LOD score of 2 (Table 2).

Table 2. Two-point LOD scores for selected markers from genome wide linkage analysis

Marker Name	Chromosome	Theta			
		0.1	0.2	0.3	0.4
FH2309	1	-2.66	-1.163	-0.45	-0.106
FH2598	1	-6.48	-3.57	-1.93	-0.81
FH2294	1	-2.98	-1.58	-0.82	-0.33
CO2.342	2	-3.8717	-1.9691	-0.96843	-0.36165
CO2.864	2	-4.57067	-2.36704	-1.19028	-0.45856
CO2.608	2	-4.57	-2.37	-1.19	-0.459
FH2302	3	-2.98431	-1.58146	-0.81699	-0.32619
FH2107	3	-3.62	-1.76	-0.822	-0.282
FH2531	3	-2.66219	-1.16292	-0.45432	-0.10637
AHT128	4	2.11	1.85	1.39	0.774
FH2534	4	2.11	1.85	1.39	0.774
GLUT4	5	-4.57067	-2.36704	-1.19028	-0.45856
CPH18	5	-2.47	-1.17	-0.52	-0.168
TAT	5	-2.984	-1.58	-0.82	-0.326
FH2119	6	-3.61643	-1.76498	-0.8223	-0.28246
CPH3	6	-2.66219	-1.16292	-0.45432	-0.10637
FH2396	7	-2.66219	-1.16292	-0.45432	-0.10637
CO8.618	8	-3.93855	-2.18352	-1.18496	-0.50228
FH2138	8	-4.57067	-2.36704	-1.19028	-0.45856
FH2186	9	-2.66	-1.16	-0.454	-0.106
FH2537	10	-2.66	-1.16	-0.454	-0.106
FH2293	10	-2.00838	-0.85516	-0.35447	-0.11861
FH2422	10	-2.54061	-1.38764	-0.74127	-0.30846
FH2319	11	-2.66219	-1.16292	-0.45432	-0.10637
FH2096	11	-3.49485	-1.9897	-1.10924	-0.48455
AHT137	11	-4.57067	-2.36704	-1.19028	-0.45856
C12.852	12	-3.49	-1.99	-1.12	-0.484
CXX.391	13	-3.49	-1.2	-1.12	-0.485
FH2060	14	-5.08122	-2.77528	-1.48253	-0.61692
FH2547	14	-5.5	-2.7	-1.56	
CPH5	15	-2.21849	-0.9691	-0.3786	-0.08864
FH2321	15	-2.66219	-1.16292	-0.45432	-0.10637
COS15	15	-2.03006	-0.9794	-0.44901	-0.1501
AHT139	15	-4.2	-2.4	-1.33	-0.58
FH2171	15	-6.48	-3.6	-1.9	-0.81
FH2278	15	-2.00897	-0.86147	-0.38114	-0.17759
FH2175	16	2.109028	1.84738	1.38556	0.774084
FH2155	16	3.0632	2.449	1.754	0.950175
AHTK209	20	-2.98431	-1.58146	-0.81699	-0.32619
FH2312	21	-3.62	-1.76	-0.822	-0.282
FH2233	21	-2.98431	-1.58146	-0.81699	-0.32619
FH2538	22	-2.66219	-1.16292	-0.45432	-0.10637
REN49F22	22	-2.66	-1.16	-0.454	-0.11
FH2079	24	-2.96262	-1.45722	-0.72245	-0.2947
FH2261	24	-5.52491	-2.9691	-1.55826	-0.63465
C26.733	26	-2.54061	-1.38764	-0.74127	-0.30846
REN48E01	26	-2.54061	-1.38764	-0.74127	-0.30846
PEZ6	27	-2.03006	-0.9794	-0.44901	-0.1501
CXX.176	28	-2.47	-1.17	-0.525	-0.168

FH2208	28	-2.66219	-1.16292	-0.45432	-0.10637
FH2585	28	-2.66219	-1.16292	-0.45432	-0.10637
FH2305	30	-4.57067	-2.36704	-1.19028	-0.45856
FH2199	31	-4.57067	-2.36704	-1.19028	-0.45856
FH2239	31	-2.54061	-1.38764	-0.74127	-0.30846
FH2238	32	-3.61643	-1.76498	-0.8223	-0.28246
CPH2	32	-3.61643	-1.76498	-0.8223	-0.28246
REN41D20	32	-3.61643	-1.76498	-0.8223	-0.28246
AHT133	37	-2.98431	-1.58146	-0.81699	-0.32619
FH2532	37	-3.61643	-1.76498	-0.8223	-0.28246
FH2587	37	-2.66219	-1.16292	-0.45432	-0.10637

The table shows those markers for which two-point LOD scores between the marker and dominant black were either >2.0 or < -2.0 at $\theta=0.1$. The former category is shown in bold.

[0097] The strongest evidence for linkage was obtained on CFA16 with marker FH2155 ($Z_{\max}=3.6$ at $\theta=0$). We used the same marker, FH2155, to analyze four additional kindreds from the Cornell pedigree (Figure 2), and observed no recombinants between FH2155 and the *K* locus, yielding a LOD score of 6 at $\theta=0$.

[0098] To refine the map location, we examined three additional markers surrounding FH2155 that span a distance of ~24 Mb, FH3592, REN275L19, and FH2175. In the EB and GB kindreds, recombinant chromosomes in two animals, EB57 and GB17, define a critical region between FH2175 and FH3592 of 23.7 Mb (33.7 – 57.4). An additional marker that lies between FH2175 and FH2155, REN292N24, was only informative for the FB and HB kindreds (Figure 2), but exhibited the same segregation pattern as FH2175 (3/10 recombinants), and therefore narrows the critical region to a 12 Mb segment (45.4 – 57.4, Figure 2B).

[0099] This region of the dog genome contains two human homology segments, 4q34 – 4q35 and 8p12, and has been annotated with more than 250 genes, mostly from other mammalian genomes (Figure 2C). Notably, none of those genes has been previously implicated in pigmentation, i.e. as a cause of human albinism or a mouse coat color mutation. Thus, the dog *K* locus is likely to represent a previously unappreciated component of the Agouti—Melanocortin pathway.

[00100] *Allelism of yellow (k^y) and brindle.* As depicted in Figures 1 and 2, the FB, GB, and HB litters contain only black and yellow animals, whereas the EB litter and several of the parents in the Cornell cross are brindle. These observations are consistent either with brindle being an intermediate allele of the *K* locus, recessive to *black* (*K*) and dominant to *yellow* (k^y), or with brindle being caused by another gene whose effects are hypostatic to those of the *K* allele. Based on previous studies in which brindle x brindle crosses often yielded a mixture of brindle and yellow (but never black) progeny, many dog breeders and geneticists assumed that brindle is caused by an intermediate allele of the *extension* locus,

e^{br} , that is dominant to *recessive yellow* (e or $Mc1r^{R306ter}$) but hypostatic to *dominant black* (originally assigned to A^S).

[00101] To investigate whether allelism or epistasis was more likely to explain the relationship between brindle and the K locus, we ascertained several kindreds in which brindle and yellow were segregating, and asked whether FH2155, which cosegregated perfectly with black vs. yellow (Figures 1 and 2), might also cosegregate with brindle vs. yellow. As depicted in Figure 3, there was perfect cosegregation between FH2155 and brindle vs. yellow in 5 phase-known meioses (in an Afghan pedigree) and 14 phase-unknown meioses (across a Great Dane and a Staffordshire Bull Terrier pedigree), corresponding to a LOD score of 3.6. We also re-examined all kindreds in the entire Cornell cross and a Boxer pedigree, and found that, in every case, transmission of brindle was consistent with an intermediate allele of the K locus with the following dominance relationships: *dominant black* (K^B) > *brindle* (k^{br}) > *yellow* (k^y).

[00102] *Epistatic interactions between alleles of the Agouti, K, and Mc1r loci.* As indicated above, loss-of-function for $Mc1r$ (*recessive yellow*, e) causes a yellow coat color that may appear very similar or even indistinguishable from that caused by homozygosity for *yellow* (k^y). Likewise, loss-of-function for *Agouti* (*nonagouti*, a) causes a black coat color that is indistinguishable from that caused by heterozygosity for *black* (K^B). To investigate the epistatic relationships between K locus alleles and those of *Agouti* and $Mc1r$, we determined the genotype for all three loci in key animals from the pedigrees depicted in Figures 1 – 3, and additional animals described in previous studies. For *Agouti*, we used the predicted cDNA sequence to distinguish among the a^y , a^t , and a alleles; for $Mc1r$, we used the predicted cDNA sequence to distinguish between the $R306ter$ (e) allele and all others (referred to below as $Mc1r^+$).

[00103] The consequent genotype-phenotype relationships provide a coherent view of epistatic interactions. For example, the Labrador Retrievers Andy (Figure 1A) and A14 (Figure 2A) have a genotype of a^t/a^t ; K^B/K^B ; e/e ; both animals have a yellow coat demonstrating that loss-of-function for $Mc1r$ is epistatic to both the *black-and-tan* (a^t) allele of *Agouti*, and the *black* (K^B) allele of the K locus. In fact, Labrador Retrievers are fixed for the *black* (K^B) allele of the K locus and the *black-and-tan* (a^t) allele of the *Agouti* locus; thus, black Labrador Retrievers demonstrate that the ability of the K locus to produce black pigment is epistatic to that of the *Agouti* locus to produce yellow pigment (because a^t/a^t ; K^B/K^B animals are black rather than black-and-tan).

[00104] Observations for an additional two breeds are particularly demonstrative. Traditionally marked German Shepherd Dogs are fixed for the *yellow* (k^y) allele of the K locus and the + allele of $Mc1r$; the difference between black and black-and-tan German Shepherd Dogs is determined solely by the *nonagouti* (a) vs. the a^t allele of *Agouti*. Thus,

the ability of *Agouti* to prevent production of yellow pigment is epistatic to that of the *K* locus to allow yellow pigment. (Stated differently, the *yellow* allele (k^y) of the *K* locus can only give rise to yellow pigment in the presence of a functional *Agouti* allele). Finally, Afghan Hounds with a *K* genotype that would ordinarily yield brindle (k^{br}/k^{br} or k^{br}/k^y) may vary at both *Agouti* (a^t or a^y) and *Mc1r* (+ or *R306ter*). In all cases, the interactions between k^{br} and *Agouti* or *Mc1r* alleles can be predicted based on what happens for k^y and for K^B . In a^t/a^t ; k^{br}/k^{br} ; +/+ animals, brindling is restricted to the areas of the coat that would otherwise be tan ("black and brindle"); in e/e animals, brindling is not apparent because *Mc1r* is epistatic not only to K^B but also to k^{br} .

[00105] These relationships together with specific examples in which we have directly determined the genotype for *Agouti*, *Mc1r*, and *K* are summarized in Table 3, and their implications for understanding the underlying biochemical pathways are depicted in Figure 4. There are several key points. First, the relationship between *Mc1r* and *Agouti* in dogs is identical to that which occurs in other mammals where *Agouti* acts to antagonize melanocortin signaling in a manner that is completely dependent on a functional receptor. Second, *Mc1r* is epistatic to all *K* locus variation, and the *K* gene product behaves similar to *Agouti* protein in this way; each requires a functional *Mc1r* in order to modulate melanocortin signaling. Finally, the epistatic relationship between *Agouti* and *K* depends on the alleles being tested: "black alleles" of *K* are epistatic to "yellow alleles" of *Agouti*, but "black alleles" of *Agouti* are epistatic to "yellow alleles" of *K*. Thus, the relationship between *Agouti* and *K* is fundamentally different from the relationship between *Mc1r* and either *Agouti* or *K*.

Table 3. Epistasis relationships for *Agouti*, *K*, and *Mc1r*

Genotype ¹			Phenotype ²	Example
<i>Agouti</i>	<i>K</i> locus	<i>Mc1r</i>		
a^t/a^t	k/k	+/+	black-and-tan	German Shepherd Dog
a^t/a^t	k/k	e/e	yellow	Afghan Hound
a^t/a^t	k^{br}/k^{br}	+/+	black, brindle points	Staffordshire Bull Terrier
a^t/a^t	k^{br}/k^{br}	e/e	yellow	French Bulldog
a^t/a^t	K^B/K^B	+/+	black	black Labrador Retriever
a^t/a^t	K^B/K^B	e/e	yellow	yellow Labrador Retriever
a^y/a^y	k/k	+/+	yellow	Boxer
a^y/a^y	k/k	e/e	yellow	Afghan Hound
a^y/a^y	k^{br}/k^{br}	+/+	brindle	Boxer
a^y/a^y	k^{br}/k^{br}	e/e	yellow	Afghan Hound
a^y/a^y	K^B/K^B	+/+	black	Great Dane
a^y/a^y	K^B/K^B	e/e	yellow	Poodle

¹Nomenclature similar to Table 1, with the R306ter allele of *Mc1r* indicated as *e*. For each category, only homozygous genotypes are shown for the sake of simplicity; more genotypes are possible according to dominance relationships for each locus as indicated in Table 1.

²These designations refer only to the distribution of eumelanin and pheomelanin, and ignore the effects of modifiers that affect spotting and/or pigment quality. For example, black-and-tan in a Cocker Spaniel homozygous for the *b* allele of the *Tyrp1* locus would be modified to liver-and-tan; brindle in a French Bulldog carrying an *s* mutation would appear white with brindle spots.

[00106] These considerations provide two alternative models (Figure 4). The *K* gene product may lie genetically upstream of *Agouti* and inhibit its function, either as a negative regulator of *Agouti* mRNA expression, or as a post-translational inhibitor that reduces the levels of active *Agouti* protein at the *Mc1r* (Figure 4A). Alternatively, the *K* gene product may act directly at the *Mc1r* to stimulate melanocortin signaling and thereby oppose the action of *Agouti* protein indirectly (Figure 4B).

[00107] A general theme of pigmentary genetics for the last century is that patterns of Mendelian variation within one species frequently display apparent homology to those in other species. For example, similar segregation and dominance relationships are observed among mice, guinea pigs, rabbits, and cats for the phenotypic series full color > chinchilla > acromelanic > albino, leading to the suggestion that mutations in the same gene—now known as *Tyrosinase*—are responsible. These types of observations foreshadowed the field of comparative genomics. Indeed, comparison of genome sequences not only clarified the evolutionary relationships among mammals (and most other organisms), but also provided the tools to identify molecular alterations responsible for the *Tyrosinase* color series in mice, cats, cattle, and rabbits.

[00108] Dominant black and brindle in dogs have been curious and somewhat confusing exceptions to the aforementioned theme. Historically, the allelic relationships for the *Agouti* locus - to which dominant black was assigned as the A^S allele—were thought to be opposite to what pertains in other mammals, where "yellow alleles" are dominant to "black alleles". In the case of brindle, assigned to the *Mc1r* locus as e^{br} , epistasis relationships were confusing, with e^{br} epistatic to the a^y allele but not to the A^S allele (a^y/a^y ; e^{br}/e^{br} animals would be brindle but A^S/A^S ; e^{br}/e^{br} animals would be black).

[00109] The work described here resolves this confusion by demonstrating that both dominant black and brindling are due to alleles of a previously unappreciated pigmentation gene that we have named the *K* locus. Although the *K* locus is an apparent exception to the idea that the same set of molecular tools are used in all mammals, its recognition reinforces the general theme that genetic interactions and pathways for orthologous genes are conserved. Thus, interactions both within and between *Agouti* and *Mc1r* alleles in dogs are identical to those observed in other mammals: "black" *Mc1r* alleles are dominant to "yellow"

Mc1r alleles, "yellow" *Agouti* alleles are dominant to "black" *Agouti* alleles, and double mutants for *Mc1r* and *Agouti* always exhibit the phenotype of single *Mc1r* mutants.

[00110] Our original survey of *Mc1r* variation among domestic dogs was motivated by the idea that dominant black might be due to a gain-of-function *Mc1r* allele, as described in many other vertebrates. Although we and others have identified a number of *Mc1r* polymorphisms among domestic dogs, the only one for which there is an unequivocal effect on function is *R306ter*, responsible for the loss-of-function allele originally described as *recessive yellow* (*e*). Given the diversity of coat colors and patterns selected in modern breeds, it is perhaps surprising that an *Mc1r* mutation that causes dominant black has not been found in dogs. However, a likely explanation is that variation at the *K* locus is relatively old among the canid lineage, since preexisting polymorphism for *k* vs. *K* would make it less likely that a new dominant black mutation at *Mc1r* would be noticed.

[00111] Because the *black* (*K^B*) allele is epistatic to variation at *Agouti*, the *yellow* (*k^y*) allele probably represents the ancestral state, otherwise the *Agouti* phenotype (and other aspects of *Agouti*-induced variation such as white-bellied *Agouti* and black-and-tan) would have been cryptic in the ancestral population where variation at *K* first occurred. According to this hypothesis, wolf populations from which dogs were domesticated some 15,000 – 40,000 years ago were *Agouti*-colored or a gray modification of *Agouti*, similar to the appearance of modern wolves. Mutation from *k^y* to *K^B* is likely to have occurred prior to the origin of modern breeds several hundred years ago, and could even have been present in wolves as an adaptive polymorphism prior to domestication.

[00112] Superficially, the brindle phenotype in domestic dogs shares some features with tabby striping in domestic cats. Both involve patches or stripes of eumelanin vs. pheomelanin hairs, and both require the presence of a functional *Agouti* gene. However, the allelic system for tabby striping is probably opposite to brindle: the presence of black tabby stripes (*t^b*) is recessive to the absence of such stripes associated with the *Abyssinian* (*T^a*) allele in cats, while the presence of black brindle stripes (*k^{br}*) is dominant to the absence of such stripes associated with the *yellow* (*k^y*) allele in dogs. Equally important, the pattern of tabby striping is alternating and regular, consistent with an underlying pattern based on a Turing-like reaction-diffusion mechanism. By contrast, the pattern of brindle stripes is irregular and variegated, most consistent with an epigenetic mechanism. These considerations are consistent with the view based on phylogenetic distribution of color patterns that tabby striping and brindling have independent evolutionary histories.

[00113] The evolutionary history of variation at the *K* locus will also have an impact on strategies for its molecular identification. The 12 Mb region to which *K* has been mapped contains approximately 250 genes, many of which are plausibly involved in melanocortin signaling. Additional genotyping of SSLP and SNP markers within the 12 Mb interval

should reveal whether the K , k^{br} , and k alleles have specific haplotypes with which they are associated.

Example 2

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

[00114] *Linkage and association mapping of the K locus.* As described in Example 1, the K^B mutation maps to a 12 Mb interval on the distal end of dog chromosome 16, between markers REN292N24 and FH3592. We ascertained additional kindreds segregating K^B , k^{br} , and k^y to refine the map location, and defined overlapping critical regions of 3.8 and 7.6 Mb for the K^B and k^{br} mutations, respectively (Figure 5A).

[00115] We used an association-based strategy to narrow the critical region. Because most breeds were derived in the past 200 hundred years from small founding populations, 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 (Figures 1B, 1C, 1D). Sixteen genes have been annotated to the region of significant association in Great Danes, including a gene cluster that encodes 12 beta-defensins (Patil *et al.* (2005) *Physiol Genomics* 23, 5), small antimicrobial peptides that are secreted mainly by epithelial cells (Ganz (2003), *Nat Rev Immunol* 3, 710; Pazgier *et al.* (2006) *Cell Mol Life Sci* 63, 1294). We sequenced the mature protein-coding regions for 9 members of the beta-defensin cluster 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 ($\Delta G23$).

[00116] To evaluate the extent to which the $\Delta G23$ 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. Among 454 dogs, there were 13 cases where the $\Delta G23$ 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 (Table 4). These results indicate that K^B alleles in all breeds are probably identical by descent and suggest that the $\Delta G23$ polymorphism or a closely linked variant in complete linkage disequilibrium (LD) is the K^B mutation.

Table 4. Association of the coat color phenotype with *CBD103* genotype^a

Breed	Genotype of pheomelaninistic dogs			Genotype of eumelanistic dogs		
	+/+	$\Delta G/+$	$\Delta G/\Delta G$	+/+	$\Delta G/+$	$\Delta G/\Delta G$
Afghan Hound	1				3	1
Akita	34					
Australian Shepherd	5				2	
Basenji	26				7	
Border Collie	2				1	1
Borzois	2					
Boxer	41					
Bulldog	27					
Bullmastiff	23					
Cavalier King Charles Spaniel	2					
Chow Chow	3				3	
Curly-Coated Retriever						5
Dalmatian						6
Doberman Pinscher	13					
Field Spaniel						5
Flat-Coated Retriever					3	1
French Brittany					1	
French Bulldog	1					
German Shepherd	7					
German Shorthaired Pointer						5
German Wirehaired Pointer						5
Great Dane	68				17	40
Greyhound	4				3	1
Havanese	1					
Irish Wolfhound	1					
Labrador Retriever					1	9
Large Munsterlander						3
Mastiff	7					
Miniature Doberman Pinscher	1					
Miniature Schnauzer	1					1
Poodle					2	3
Portuguese Water Dog						3
Pug	2					1
Puli	2				2	1
Rottweiler	10					
Shar-Pei	3				4	1
Weimaraner	2				4	4
Whippet	3					
Total	292	0	0	0	53	96

^a Alleles are denoted as "+" for the reference sequence (since Tasha the Boxer is k^Y/k^Y) and " ΔG " for the 3 bp deletion (CFA16:55755801-55755803) referred to as S54 in Table S2, and which predicts the *CBD103* $\Delta G23$ deletion described in the text. Not shown in this table are the genotypes of 13 dogs whose genotypes at *CBD103* were discordant with their coat color phenotype due to epistasis with *Mc1r* or *Agouti* mutations. Five "black" animals (in the German Shepherd Dog or Weimaraner breeds) were homozygous for a loss-of-function *Agouti* allele, and 8 "yellow" animals (in the Afghan Hound, French Bulldog, German Shorthaired Pointer, Labrador Retriever, Puli, or Poodle breeds) were homozygous for a loss-of-function *Mc1r* allele.

[00117] 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 (Lindblad-Toh et al. (2005), Nature 438, 803). We identified 28 polymorphisms [22 single-nucleotide polymorphisms (SNPs) and 6 indels including the $\Delta G23$ 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 7 breeds. We observed 6 "parental" k^Y -bearing and 5 "parental" K^B -bearing chromosomes (depicted in yellow and blue, respectively, Figure 6A). We also identified 8 chromosomes that carried a single ancestral recombination event, which together defined a maximal interval for K^B of 9146 bp (Figure 6A). Complete resequencing of this interval (except for 3 homopolymer tracks) in 5 k^Y/k^Y animals, 1 K^B/k^Y animals, and 4 K^B/K^B animals from 7 breeds revealed 2 polymorphisms besides $\Delta G23$ that are perfectly concordant with *K* locus genotype (S104 and S105) (Figure 6A, Table 5).

Table 5:

SNPs and indel polymorphisms

ID	CFA16 CanFam1.0	(bp) Tasha's allele	Other allele	Broad SNP ID	Primer Forward	Primer Reverse
S1	53971555	T	C		TGCTGCTACACTGCGTTACC	ATGGCAGATGCCCTTTACAC
S2	53971607	T	C		TGCTGCTACACTGCGTTACC	ATGGCAGATGCCCTTTACAC
S3	53971621	T	C	BICF235J3047	TGCTGCTACACTGCGTTACC	ATGGCAGATGCCCTTTACAC
			T or			
S4	53971625	A	del		TGCTGCTACACTGCGTTACC	ATGGCAGATGCCCTTTACAC
S5	54258844	T	C		CTCAGGAGGCTGCTATCTGG	GCCATACACAGGTGCCTCTT
S6	54258880	T	A		CTCAGGAGGCTGCTATCTGG	GCCATACACAGGTGCCTCTT
S7	54258888	C	T		CTCAGGAGGCTGCTATCTGG	GCCATACACAGGTGCCTCTT
S8	54258951	C	T	BICF237J56729	CTCAGGAGGCTGCTATCTGG	GCCATACACAGGTGCCTCTT
S9	54258960	G	C		CTCAGGAGGCTGCTATCTGG	GCCATACACAGGTGCCTCTT
S10	54650550	G	C		CACCATTTCGTGGGAGAAAGT	GCTCAATGGGCATTTTGTCT
S11	54650563	G	C		CACCATTTCGTGGGAGAAAGT	GCTCAATGGGCATTTTGTCT
S12	54650581	G	A		CACCATTTCGTGGGAGAAAGT	GCTCAATGGGCATTTTGTCT
S13	54650600	A	G		CACCATTTCGTGGGAGAAAGT	GCTCAATGGGCATTTTGTCT
S14	54650616	A	G	BICF231J58300	CACCATTTCGTGGGAGAAAGT	GCTCAATGGGCATTTTGTCT
S16	55217908	G	A	BICF233J59359	ATGGAGATGGGTGAGAGTGG	GCATGGAACAGCAATCTGA
S17	55218721	C	T		TCTTGCTTTGAAATCTCTCTTCA	TGCACAGTCAGTTCATCCT
S18	55218747	C	A		TCTTGCTTTGAAATCTCTCTTCA	TGCACAGTCAGTTCATCCT
S19	55218887	T	C		TCTTGCTTTGAAATCTCTCTTCA	TGCACAGTCAGTTCATCCT
S20	55371177	A	C		GGGTGAAATCCACCAGATTG	CACAAATCAAGCTGCGAATC
S21	55371224	A	G	BICF237J39766	GGGTGAAATCCACCAGATTG	CACAAATCAAGCTGCGAATC
S22	55541817	G	C	BICF237J64796	TAGAAAAATGCCCCATTTGC	CTTTCTGAGCCGAGTCCTTG
S23	55541842	C	T		TAGAAAAATGCCCCATTTGC	CTTTCTGAGCCGAGTCCTTG
S24	55541843	T	G		TAGAAAAATGCCCCATTTGC	CTTTCTGAGCCGAGTCCTTG
S25	55541844	G	T		TAGAAAAATGCCCCATTTGC	CTTTCTGAGCCGAGTCCTTG
S26	55541858	G	T		TAGAAAAATGCCCCATTTGC	CTTTCTGAGCCGAGTCCTTG

S27	55548577	G	A		GAAAGTAAGGCCCTGTGTGC	AGGATCATGCCTTCCAACAG
S28	55548600	C	T	BICFPJ214189	GAAAGTAAGGCCCTGTGTGC	AGGATCATGCCTTCCAACAG
S29	55548608	C	G		GAAAGTAAGGCCCTGTGTGC	AGGATCATGCCTTCCAACAG
S30	55551795	C	G		GGACCTAATTCTCAGCCAGG	GGTGATCTCACCTCGCTTGT
S31	55551944	G	C		GGACCTAATTCTCAGCCAGG	GGTGATCTCACCTCGCTTGT
S32	55551952	C	T	BICFPJ547724	GGACCTAATTCTCAGCCAGG	GGTGATCTCACCTCGCTTGT
S33	55551998	T	G		GGACCTAATTCTCAGCCAGG	GGTGATCTCACCTCGCTTGT
S34	55581684	A	G		TCCTTCAGTCATTCTTACTCATTTG	TCAGACTTCACTGGGGTTTTC
S35	55581728	T	C		TCCTTCAGTCATTCTTACTCATTTG	TCAGACTTCACTGGGGTTTTC
						GGAAAACAAAAATCGCAAGA
S36	55585724	C	T		GACCCTGAGTCTGCACCTCT	A
						GGAAAACAAAAATCGCAAGA
S37	55585819	C	G	BICFPJ767999	GACCCTGAGTCTGCACCTCT	A
						GGAAAACAAAAATCGCAAGA
S38	55585840	T	C		GACCCTGAGTCTGCACCTCT	A
S39	55709569	T	G	BICF233J60510	ATCCACCAGCACCTTCTCAG	TAGTCAGGGTTGGGAACTGG
S40	55714953	C	T		TACGAAAGAAACCCGTGTCC	CAGCTCCTGCTAAGCCACTT
S41	55725863	A	T	BICFPJ945361	GTGGTGACCAGGGAAGGTAG	ACTCTGCTGTGAGCTGGGTT
S42	55726825	G	A		TCCATAAATGGGTAGTTGTGCTT	ATTTTTCCAAAGTTCGGGGT
S43	55726868	T	C	BICFPJ945365	TCCATAAATGGGTAGTTGTGCTT	ATTTTTCCAAAGTTCGGGGT
S44	55726885	G	C		TCCATAAATGGGTAGTTGTGCTT	ATTTTTCCAAAGTTCGGGGT
S45	55740429	A	G		AGCCTACAGAGGCAGGATGA	CTGGCTTAGGGACACACCTC
						TGGAGGACATATTTGTGTTTG
S46	55742802	C	T		AGCGCTATGAACACAAGGGT	TG
						TGGAGGACATATTTGTGTTTG
S47	55742803	C	A		AGCGCTATGAACACAAGGGT	TG
						TGGAGGACATATTTGTGTTTG
S48	55742826	C	T	BICFPJ70048	AGCGCTATGAACACAAGGGT	TG
S49	55754820	G	A		TCTGCCTTTCTCCATTCTTCA	AGGGACTGGTCACAGTGGTC
S50	55754893	T	C	BICF237J66607	TCTGCCTTTCTCCATTCTTCA	AGGGACTGGTCACAGTGGTC
S51	55754941	G	A		TCTGCCTTTCTCCATTCTTCA	AGGGACTGGTCACAGTGGTC
S52	55754981	T	A		TCTGCCTTTCTCCATTCTTCA	AGGGACTGGTCACAGTGGTC
S53	55755747	C	G		TGTCTTCATCCCTGTGAGGT	CCAGGAGGCATTTTCACACT
S54	55755801- 55755803	GGG	Del		TGTCTTCATCCCTGTGAGGT	CCAGGAGGCATTTTCACACT
S55	55755864	C	T		TGTCTTCATCCCTGTGAGGT	CCAGGAGGCATTTTCACACT
S56	55786080	G	A		CTGGGACGGTTTGCATAAGT	CCTTAAATCACCCCTGCCTCA
S57	55786273	G	A		CTGGGACGGTTTGCATAAGT	CCTTAAATCACCCCTGCCTCA
S58	55836084	A	C	BICFPJ950727	TTCTCAGAGGAGGTGATGGG	CACTCTTGGCTTCTGTTCCC
S59	55872144	T	C	BICF235J38779	TTGAGAACTGTTCCGACGTG	AGGCTCCCCAGGTGTTTAGT
S60	55872165	T	C		TTGAGAACTGTTCCGACGTG	AGGCTCCCCAGGTGTTTAGT
S61	55877317	C	T	BICFPJ1486217	AAAACCTGGTACACGGGGACA	CGATGTTCTCCAGCAGCATA
S62	55877321	C	T		AAAACCTGGTACACGGGGACA	CGATGTTCTCCAGCAGCATA
S63	55877353	T	C		AAAACCTGGTACACGGGGACA	CGATGTTCTCCAGCAGCATA
S64	55877420	C	T		AAAACCTGGTACACGGGGACA	CGATGTTCTCCAGCAGCATA
S65	55877468	T	C	BICFPJ1486213	GCGGATGCTTTAGTGACCTG	GAGCACATTCCCGCTCTTAG
S66	55885068	G	A		ACTCCGGAGAGAGGTTTCGAT	CTACCTGAGCCCCACCTAC
S67	55885113	T	C	BICFPJ1052677	ACTCCGGAGAGAGGTTTCGAT	CTACCTGAGCCCCACCTAC

S68	55885114	G	A		ACTCCGGAGAGAGGTTTCGAT	CTACCTGAGCCCCACCTAC
S69	55896178	T	C	BICF235J38779	AGCTGACTCAACATGCAGGA	CCGTTCTATCTTCAGGGCAG
S70	55896184	T	C		AGCTGACTCAACATGCAGGA	CCGTTCTATCTTCAGGGCAG
S71	55896216	C	T		AGCTGACTCAACATGCAGGA	CCGTTCTATCTTCAGGGCAG
S72	55896221	A	G		AGCTGACTCAACATGCAGGA	CCGTTCTATCTTCAGGGCAG
S73	55896301	A	G		AGCTGACTCAACATGCAGGA	CCGTTCTATCTTCAGGGCAG
S74	56250053	G	A	BICF237J21474	GCGTGCTTTCGAGACATAAA	CCCACTGGATCTTCAGGTGT
S75	56525724	T	G	BICF231J21146	AGCTTTTCCTGAACACGCAT	CCGAGAGTTTTACAGCCGAG
S76	56525736	T	C		AGCTTTTCCTGAACACGCAT	CCGAGAGTTTTACAGCCGAG
S77	56525745	C	T		AGCTTTTCCTGAACACGCAT	CCGAGAGTTTTACAGCCGAG
S78	56525821	T	A		AGCTTTTCCTGAACACGCAT	CCGAGAGTTTTACAGCCGAG
S79	55746083	del	T		GCATGTACATCCACTAGGGACA	CAGTGATGTGGTGCTGTGTG
S80	55746221	C	T		GCATGTACATCCACTAGGGACA	CAGTGATGTGGTGCTGTGTG
S81	55746536	G	A		CCTTCTTGATTTTGCCCTCA	TTCCTTGGCTCTTCTCTCCA
S82	55746684- 55746706	GA(...)T A	Del		CCTTCTTGATTTTGCCCTCA	TTCCTTGGCTCTTCTCTCCA
S83	55747087	A	G		GGACTACAGTGAAGGAGCCAAG	TGGAACCTCTCGGGAGACTA
S84	55747241	T	C		GGACTACAGTGAAGGAGCCAAG	TGGAACCTCTCGGGAGACTA
S85	55747266	A	C		GGACTACAGTGAAGGAGCCAAG	TGGAACCTCTCGGGAGACTA
S86	55747268	C	A		GGACTACAGTGAAGGAGCCAAG	TGGAACCTCTCGGGAGACTA
S87	55748067	G	A		GGGATCCTGGCTGGACTA	CTGTTGGGGAGGCCAACT
S88	55748115	G	Del		GGGATCCTGGCTGGACTA	CTGTTGGGGAGGCCAACT
S89	55749774	A	C		ACAGGCTCGTCTGCTTGACT	TCTCTGGGCTAGGAGAAGCA
S90	55749950	T	C		ACAGGCTCGTCTGCTTGACT	TCTCTGGGCTAGGAGAAGCA
S91	55752630- 55752632	TTT	Del		CCCCTCGTCAACCTTATCAA	TGAGCATCAGGTCACTGGAG
S92	55752862	G	A		TTGCCTTTAAATGCCTTCC	AGCACAGGGCTATGGAAACA
S93	55752940	G	A		TTGCCTTTAAATGCCTTCC	AGCACAGGGCTATGGAAACA
S94	55753338	G	T		TCTGAGGAAATTTTGCTTTGA	AAGGGGCTTAGGAAGGTACG
S95	55754820	G	A		GGGTGAAGGAATCGCAAATA	T
S96	55757261	G	A		TCCATTGTCTCCAGTCTCCTG	GCGCTACTACAAAGGCAAGC
S97	55764041	G	A		ATGGTCTGCAGCACTCAGG	AAAATCGCAGGACAGAGCAT
S98	55764169	G	A		ATGGTCTGCAGCACTCAGG	AAAATCGCAGGACAGAGCAT
S99	55764190	C	T		ATGGTCTGCAGCACTCAGG	AAAATCGCAGGACAGAGCAT
S100	55765110- 55765113	TATC	Del		AAGTCACAAAGGCAGGTTCC	CATGTCTCCTGAACCACAGG
S101	55765138	C	G		AAGTCACAAAGGCAGGTTCC	CATGTCTCCTGAACCACAGG
S102	55765233	C	T		AAGTCACAAAGGCAGGTTCC	CATGTCTCCTGAACCACAGG
S103	55765451	A	C		TGGCTACCAAATGGACACAG	GATCTTCTGCACCTTTTCATGG
S104	55751487	G	A		CCCCACCTGCTCCTTACTCT	CATAGTCGAAGGGTGGCCTA
S105	55751488	C	A		CCCCACCTGCTCCTTACTCT	CATAGTCGAAGGGTGGCCTA

[00118] The 9146 bp interval contains both exons of *CBD103*, the first exon of dog expressed sequence tag (EST) CX990240 and dog EST CO665262 (Figure 6-). However, several considerations indicate that *CBD103* is, indeed, the *K* locus and that the Δ G23 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* (Figure 7A). Finally, as discussed further below, *CBD103* is highly expressed in skin, the $\Delta G23$ deletion affects *CBD103* protein function, and pharmacologic studies reveal that *CBD103* can modulate melanocortin signaling.

[00119] 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 (Figure 6B), indicating that $\Delta G23$ and consequently the *K^B* mutation occurred specifically within the canid lineage.

[00120] *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 beta-defensin genes that are clustered on chromosomes 16 or 25 by reverse transcription polymerase chain reaction (RT-PCR). Expression was detectable only for 2 genes: *CBD1* and *CBD103*. We then used quantitative RT-PCR to measure levels of skin mRNA from 4 *K^B/k^y* samples and 4 *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 (Figure 7A).

[00121] 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 *CBD103 $\Delta G23V5$*) and studied their patterns and levels of protein expression after transfection of cultured mouse keratinocytes.

[00122] In cell extracts analyzed by Western blotting, antisera against the V5 epitope detects 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 (Figure 7B). 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 *CBD103 $\Delta G23V5$* as compared with that of *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.

- [00123] *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. We chose a genetic background that normally has Agouti banded hairs and observed that 2 transgenic founders generated with the *CBD103* Δ G23 cDNA (the K^B allele) displayed a predominantly black coat with small patches of banded hair (Figure 8). Unexpectedly, the normal *CBD103* cDNA (the k^Y allele) also produced transgenic mice with a black coat in 20 out of 21 founders (Figure 8). Furthermore, we observed that transgenic animals were smaller than their non-transgenic littermates. By two weeks of age, female transgenic animals were easily recognized by 2 weeks of age by their dark coat and small size; in adult mice, reduced body weight persists in both males and females.
- [00124] These considerations suggest three possible mechanisms by which *CBD103* might act to modulate melanocortin signaling: (i) by binding to and activating the Mc1r, (ii) by binding to the Mc1r and preventing its inhibition by Agouti protein; or (iii) by binding to Agouti protein leading to its sequestration and/or degradation. To distinguish among these ideas, we generated synthetic forms of *CBD103* and tested their ability to interact with the Mc1r and agouti signaling protein-YY (ASIP-YY), a synthetic version of the carboxy-terminus of Agouti protein that behaves as a competitive antagonist of α -MSH at the Mc1r and Mc4r.
- [00125] The *CBD103* k^Y allele predicts a mature peptide (after signal sequence cleavage) of 45 amino acids that contains 6 cysteine residues and begins with the glycine that is deleted in the K^B allele (Figure 6B). The predicted signal sequence cleavage site is supported by biochemical studies of the orthologous human protein [known as DEFB103 or human beta-defensin 3 (HBD3)] purified from human tissues. We synthesized the 45 residue k^Y form and 44 residue K^B form of the dog protein (hence referred to as *CBD103* and *CBD103* Δ G23, respectively), allowed oxidative refolding, and used mass spectrometry and high-performance liquid chromatography to confirm recovery of a single congener with 3 intrachain disulfide bonds.
- [00126] [Nle⁴, D-Phe⁷]- α -MSH, a potent derivative of α -MSH, stimulates robust accumulation of cyclic adenosine monophosphate (cAMP) in melanocytes. However, neither *CBD103* nor *CBD103* Δ G23 has any effect on cAMP accumulation (Figure 9A), indicating that *CBD103* is not a conventional Mc1r agonist. To test for receptor binding, we transiently transfected human embryonic kidney (HEK) 293 cells with Mc1r expression constructs and used europium-labeled NDP-MSH (Eu-NDP-MSH) as a fluorescent tracer. We used saturation binding assays to first calculate the affinity of Eu-NDP-MSH for the dog Mc1r (Figure 9B) and then carried out displacement assays with progressively increasing

concentrations of CBD103, CBD103ΔG23, or ASIP-YY. All three peptides exhibited qualitatively similar profiles characteristic for competitive binding to a single high-affinity site (Figure 9C, Table 1). Quantitatively, inhibition constant (K_i) values estimated from the data depicted in Figure 9C were 150.6 nM for CBD103 and 34.2 nM for CBD103ΔG23. These estimates varied according to experimental conditions, but CBD103ΔG23 consistently exhibited higher affinity for the dog Mc1r than did CBD103 (mean of fivefold across 4 paired experiments, Table 6).

Table 6: Affinity constants for melanocortin receptor ligands^a

Receptor	Eu-NDP- MSH	ASIP-YY	CBD103 ΔG23	CBD103	HBD1	HBD3
dog Mc1r	0.70	0.59 (0.51 – 0.95, n=3)	37.0 (16.4 – 61.1, n=5)	221.0 (150.6 – 398.7, n=4)	N.D.	N.D.
mouse Mc1r	1.38	2.30	15.1 (12.4 – 17.0, n=4)	9.7 (8.9 – 10.4, n=2)	N.D.	N.D.
human Mc1r	2.59	0.95	19.6 (12.3 – 26.8, n=2)	35.5	30.0	13.8
human Mc4r	3.17	0.82	104.5	N.D.	N.D.	N.D.

^aIn the column for Eu-NDP-MSH, saturation binding assays as depicted in Fig. 5B were used to derive dissociation constant (K_d) values (in nM) by fitting the data to a hyperbolic dose-response curve with the use of nonlinear regression. In the remaining columns, displacement binding assays as depicted in Figure 9 (C to F), were used to derive K_i values (in nM) by fitting the data to a sigmoidal dose-response curve with variable slope. For some ligand-receptor combinations, multiple experiments were carried out, in which case the mean K_i value is given followed by the range and number (in parentheses) of the separate experiments. ND, not determined.

[00127] In these same experiments, ASIP-YY exhibited K_i values of 0.51 to 0.95 nM for the dog Mc1r, in the same range as reported previously for ASIP-YY at the human Mc1r. Using quantitative RT-PCR, we found that the levels of *CBD103* mRNA in total skin were ~300-fold greater than that of *Agouti* mRNA; to the extent that this difference reflects protein abundance, the levels of CBD103ΔG23 in dog skin are likely to be much greater than those of Agouti protein, which is consistent with a model in which CBD103ΔG23 competitively inhibits the ability of Agouti protein to antagonize Mc1r signaling.

[00128] Finally, we investigated a potential interaction between Agouti protein and CBD103 by preincubating ASIP-YY with either CBD103 or CBD103ΔG23 before the binding assay, and found no evidence for a functional interaction. We also examined the effects of mixing the two peptides on two-dimensional nuclear magnetic resonance spectra and found no evidence for a structural interaction. Taken together, these observations indicate that CBD103 is a high-affinity ligand for the Mc1r and that the ΔG23 mutation may cause a black

coat color in dogs via two different mechanisms: increased affinity for the *Mc1r* and increased availability of the mature protein *in vivo*.

[00129] *Other mammals, other melanocortin receptors, and other defensins.* Dogs are the only mammals known for which dominant inheritance of black coat color is not caused by a *Mc1r* mutation. However, we observed that both CBD103 and CBD103ΔG23 are high-affinity ligands for the mouse (Figure 9D) and the human (Figure 9E) *Mc1r*, with estimated K_i values that were, unexpectedly, lower than that observed at the dog *Mc1r* (Table 6). Furthermore, unlike at the dog *Mc1r*, the affinity of CBD103 compared to that of CBD103ΔG23 was similar at both the mouse and the human *Mc1r* (Table 6), which may explain why expression of either allele caused a black coat in transgenic mice (Figures 9A, 9B).

[00130] Given the effects of the *CBD103* transgenes on body size, we investigated the potential for cross-talk between beta-defensins and other melanocortin receptors. We found that CBD103ΔG23 binds to the human *Mc4r* with an intermediate affinity ($K_i = 104.5$ nM) between that of CBD103ΔG23 and CBD103 for the dog *Mc1r* (Table 6). We also synthesized HBD3, the human ortholog of canine CBD103, and observed high-affinity binding ($K_i = 13.8$ nM) to the human *Mc1r* (Table 6). Finally, we tested human β -defensin 1 (HBD1), which lies in the same cluster as HBD3, but is more distantly related; HBD1 also exhibited high affinity binding to the human *Mc1r* ($K_i = 30$ nM) (Table 6).

Materials and Methods

[00131] *Genome sequence.* Throughout this paper the dog genome assembly used is CanFam1.0 as displayed and annotated in the UCSC genome browser. During the course of this work, a second assembly of the dog genome became available, CanFam2.0 (2), in which the last 5 Mb of CFA16, where the *K* locus maps, are inverted compared to CanFam1.0. However, the linkage map, two radiation hybrid maps, a FISH map and our own mapping data support the CanFam1.0 orientation. Therefore we conclude that the correct orientation is that of the CanFam1.0 sequence and all the map coordinates in this manuscript are presented according to the CanFam1.0 sequence.

[00132] *Dog DNA samples and skin biopsies collection.* Catch-All swabs (Epicenter) or Cytopak brushes (Medical Packaging) were used to collect cheek swab samples through the mail or at dog shows according to a Stanford IACUC-approved protocol. DNA from the swabs was prepared using the Qiamp DNA mini kit (Qiagen).

[00133] *Genotyping and sequencing.* Microsatellite markers are given in Table 7, and were identified before the dog genome assembly was released by blast-searching the dog trace archive at NCBI with repeat-masked human sequence from the regions of predicted homology on human chromosomes 4 and 8. SNP and indel markers were identified by

sequencing amplicons designed to validate SNPs from the dog genome variation database, or by resequencing.

Table 7: Microsatellite markers

Marker	CFA16 (bp) CanFam1.0	Primer F	Primer R	Size (bp)
REN292N24	45414007	GCCCTCAAGCTCAATCTACG	TGCAACAGTCAAGACATGGA	141
M1	49652784	TGTCTGAACATGAACTGCATCA	AGGGTTCAGGCTCACAAAAC	238
M2	49802408	CCATCCAAATTTCTGAGTGC	GCAGTTAAGGGTCTGCCTTG	209
FH2155-a	52719635	TCAATTCCTTTGAAGCCAGTTT	AACAAAAAGTTGATAGTACATTTTCCT	308
FH2155-b	52719635	TGTAGATGATGGAGACATTGGG	AGGCAAATATGCCAAGGATG	498
M3	52732889	ACCACAAGGGCACAGACACT	TGAGCTCCAAACCAACAGACT	249
M4	52736072	AGAGCAACATTCGCAGTCAG	TTTGAAATTGAAAGCAATCATCT	157
M5	54857328	CCTCTTGTGACAGGATCAGG	GCACATTAGCCTTCCCTTC	222
M6	54881106	GTTCTCCATTGTCCCTGAA	AGCATCTGGATTGCCAGAAA	211
M7	54910722	TCCTGCATCCTTATTTCCAGTT	CTCCAGGAAGCAAGCAAAAC	179
M8	55029560	AGCTCCATCCAGGTTGTTTG	GGAAGCAGCCAAGTATCCAT	211
M9	55089638	CTGGGATCGAGTTCCACATC	TCCAGGGACATCCTGTAGGT	226
FH2990	55596137	GGTTGGGAATTTCTTACTGG	CTTGCTCAGCTGGAGACC	455
REN96I08	55661755	CTTAAGTGAAATGGCAATAGG	AGGGAGCTGAATGGATGAT	211
M10	56504030	CCTCCTTGGGCAGACTTCTT	TACAGGCAGTGCCAGCTCTA	164
M11	56594012	GGGGGAAAAATAGCTTCCTG	ATGACCAAACGTTGCCTAGC	209
M12	56603029	TGGGAATGAGAGAAAAAGGAA	CTGGACTCCAAAGAGGATGC	247
FH3388	57317253	TCAGAAGGTGCTCATATTTTCC	TAATTGAACCCACAACCTTG	255
FH2764	57333658	ATATGCACGTCAGCACATTC	TGTCACAGAAGTGTTTTC	151
FH3592	57475457	AGGTGCTGAGCATGTTATCC	TGGTCTGTGTGTGAGGAGAAA	344

[00134] SNP genotyping and sequencing of PCR-amplified genomic DNA was carried out using standard fluorescent dye terminator technology on a capillary instrument, and analyzed using CodonCode Aligner 1.5.1c (CodonCode Corporation) or Sequencher 4.2 (Gene Codes Corporation). Microsatellite genotyping of PCR-amplified genomic DNA was carried out by incorporating a fluorescent primer in the PCR reaction or by using a fluorescently labeled universal primer, and calling genotypes with GeneMapper software (Applied Biosystems). Primer sequences for polymorphisms on CFA16 are given in Tables 7 and 8, including those used to genotype CBD103ΔG23, referred to as S54 in Table 8.

Genotyping of the R96C mutation in Agouti and the R306ter mutation in Mc1r has been described previously.

[00135] Genetic data analysis. Statistical tests of association and LD plots were carried out with Haploview. The markers do not significantly deviate ($p > 0.001$) from Hardy-Weinberg expectations (a requirement for accurate implementation of the allelic association test). However, the Bonferroni-corrected thresholds should be viewed as a general guide rather than absolute tests of significance for individual SNPs because many of the SNPs are correlated, and because of the potential for varying degrees of relatedness among some samples. Haplotypes were estimated using Phase v2.1. For estimation of haplotype-sharing and inferring the position of ancestral recombination events, missing genotypes and apparent double recombination events within less than 1 kb (which are most likely due to gene conversion) were ignored.

[00136] RT-PCR and qRT-PCR. Skin biopsies from euthanized dogs were immersed in RNAlater (Ambion) and stored at -80°C . After thawing, total RNA was isolated by homogenization with a rotor-stator and extraction using the RNeasy Fibrous Tissue Midi kit (Qiagen). Each RT or qRT-PCR reaction was carried out with 1 μg of total RNA. Primers were designed for each of the 16 beta-defensin genes on CFA16, the 4 beta-defensin genes on CFA25, Agouti, and Bactin as a reference.

[00137] Expression of CBD103 in transgenic mice and in keratinocytes. For transgenic mice, protein-coding sequences for CBD103 or CBD103 Δ 23 were PCR-amplified from genomic DNA, cloned into an EcoRI site in the pCAGGS expression vector, which drives transgenic expression under the regulation of a cytomegalovirus/beta-actin hybrid (CAGGS) promoter. During PCR amplification, we used primer sequences (5'-ACTGGAATTCGCCGCCACCATGAGGATCTATTACCTTCTCCTCC-3' and 5'-ACTGGAATTCCCATCCCATTTCTGGATTT-3') to incorporate an extended Kozak sequence to improve translation efficiency. The transgene was excised by digestion with Sall and BamHI, gel-purified, and subsequently microinjected into (FVB/N x C57BL6/J) F1 one cell embryos by the Stanford Transgenic Facility. Potential founder mice were genotyped by PCR using the following primers 5'-CGTGCTGGTTATTGTGCTGT-3' and 5'-CCTGCACCTGAGGAGTGAAT-3'.

[00138] For transfection of a mouse keratinocyte cell line, protein-coding sequences for CBD103 or CBD103 Δ 23 were PCR-amplified from the transgenic constructs using the following primers, 5'-ACTGCTCGAGGCCGCCACCATGAGGATCTATTACCTTCTCCTCC-3' and 5'-ACTGTTCTGAATTTCTTTCTTCGGCAGCATTT-3', then cloned as C-terminal fusion proteins into the pcDNA 3.1/V5-His A plasmid (Invitrogen) using XhoI and BstBI.

- [00139]** Keratinocytes were grown in 154CF medium supplemented with 0.2mM CaCl₂ and human keratinocyte growth supplement (Cascade Biologics), and transfected using the Eugene HD reagent (Roche Applied Science) in 6 well plates. 48 to 72 hours post-transfection, the proteins were extracted from both the cells and the media. Media was collected and centrifuged for 20 min (all centrifugations were at 4°C and 20,000g), and the supernatant precipitated using trichloroacetic acid. The cell layer was washed twice with cold PBS, scraped into 1M HCl and sonicated, then centrifuged for 10 min. The pellet was then reextracted with 5% acetic acid and spun down for 10 minutes. The HCl and acetic acid extracts were pooled and lyophilized. Proteins precipitated from the media, or lyophilized from the acid extract of the cell layer, were dissolved in 40 µl of Novex tricine SDS sample buffer (Invitrogen) prior to SDS-polyacrylamide gel electrophoresis under reducing conditions on a 16% Novex tricine gel (Invitrogen). After electrophoresis and transfer to nitrocellulose, the membrane was blocked in 5% milk and hybridized overnight at 4°C with a monoclonal anti-V5 antibody (Invitrogen) in 2.5% milk. (Equal protein loading was checked by staining the membrane using the Memcode stain (Pierce). An HRP-conjugated anti-mouse antibody (Biorad) was used as a secondary antibody and hybridized for 1 hour, then developed with the ECL substrate (GE Healthcare). The loading control, beta-actin, was subsequently detected using a monoclonal anti-β-actin antibody (Sigma), and the same secondary antibody.
- [00140]** This experiment was repeated 4 times; each experiment included 3 – 4 independent transfections of the CBD103 and CBD103GA23 constructs. Our conclusion that more CBD103GA23 than CBD103 accumulates in the media is based on the Cochran-Mantel-Haenszel Chi-Square test, comparing the relative levels of the two proteins across all 4 experiments.
- [00141]** Peptide synthesis. All peptides were produced on an Applied Biosystems 433A peptide synthesizer using standard Fmoc chemistry. Amino acids were purchased from NovaBiochem and assembled on Rink-amide-MBHA resin. Pre-activated Fmoc-Cys(Trt)-OPfp was used to avoid enantiomerization. Coupling reactions used four equivalents of each amino acid in N, N-Dimethyl Formamide (DMF), along with O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro-phosphate (HBTU) and N, N-Diisopropylethylamine (DIEA) except Pre-activated Fmoc-Cys(Trt)-OPfp which was only coupled in DMF solution. Fmoc protecting groups were removed using a 1% mixture of 1,8-diazabicyclo[5.4.0]undec-7-ene and Hexamethylenimine. Cleavage was performed in a mixture of Trifluoroacetic acid/Triisopropylsilane/1, 2-Ethanedithiol/phenol (90:4:4:2) for 90 min. Oxidative folding was achieved by dissolving peptides to a concentration of 0.1mg/mL in folding buffer (0.5-1.0 M Guanidine hydrochloride, 0.1 M Tris, 1.0 mM reduced Glutathione, 0.1 mM oxidized glutathione, pH 8.5) with stirring for 48 hours. Folding was

monitored by HPLC, and gave rise to one, or sometimes two, closely migrating but distinct major peaks. In cases where two major peaks were observed (for CBD103ΔG23 and HBD1), the pharmacologic behavior and mass of the material from the different peaks was identical, suggesting that the different peaks represented non-covalent conformers, as we have previously described for a synthetic version of Agouti protein. The folded products were purified by HPLC using a C18 column and identified as fully oxidized peptides by mass spectrometry (accurate to within 1.0 amu). Accurate concentrations were determined with amino acid analysis performed at the molecular structure facility at UC Davis.

[00142] cAMP and receptor-ligand binding assays. Melan-A cells were grown in RPMI supplemented with 200nM TPA at 37°C and 10% CO₂. Prior to each assay, cells were plated onto 96-well half-area plates (Corning #3885) at a density of 20,000 cells/well and grown for 24 hours. After washing the wells once with PBS, pH 7.4, peptides (diluted in PBS, pH7.4, containing 0.2% BSA and 0.5 mM IBMX) were added and the plate was incubated at 37°C and 10% CO₂ for 40 minutes. cAMP levels were measured using the Hithunter cAMP XS kit (DiscoverX) and a Centro LB 960 plate reader (Berthold Technologies). All assays were carried out in triplicate and analyzed using Graphpad (Prism).

[00143] All receptor-ligand binding assays were performed using DELFIA lanthanide-based detection system on intact 293T cells transiently transfected with melanocortin receptor expression constructs. The human Mc1r and human Mc4r constructs were generously provided by Dr. Ying-Kui Yang (Department of Pediatric Surgery, University of Alabama, Birmingham); the dog and mouse Mc1r constructs were made by cloning the protein-coding regions PCR-amplified from genomic DNA into the pcDNA3.1(-) vector (Invitrogen).

[00144] For binding assays, a 10 cm dish of 293T cells was transfected with 10 µg of a melanocortin receptor expression construct using calcium phosphate, with transfection efficiency monitored by parallel transfection with a GFP-tagged Mc4r construct. After 12 - 16 hours, the media was changed, and at 24 hours the cells were trypsinized, washed once in media, and resuspended in binding buffer, then added to a 96-well Acrowell filtration plate (Pall).

[00145] For saturation binding, 2/3 serial dilutions of Europium-labelled NDP-MSH (Eu-NDP MSH, Perkin-Elmer) were added to individual wells and incubated for 2 hours at 37°C. To measure non-specific binding, a similar dilution series of Eu-NDP MSH was added in the presence of 1.5 µM of unlabelled NDP-MSH. For displacement binding, a tracer level of Eu-NDP MSH (1.8 – 3.0 nM) was added together with varying concentrations of ASIP-YY or a defensin peptide. After the incubation, binding buffer was filtered by centrifugation, wells

were washed three times with 200 μ l ice-cold wash buffer, and 150 μ l of enhancement solution (Perkin-Elmer) was added to each well. The plates were then incubated at room temperature for 45 minutes and time-resolved fluorescence measured using a FluoStar Optima plate reader (BMG Labtech). The conditions for the binding assays with regard to the number of cells (20,000 – 70,000 per well) and the concentration of Eu-NDP MSH (1.8 – 3.0 nM) were based on pilot experiments estimating both B_{max} and K_i , and adjusted to maintain an appropriate dynamic range, and avoid depletion of free ligand by more than 10%. The time and temperature used for the binding assay were based on pilot experiments and recommendations from Perkin-Elmer regarding conditions required to reach binding equilibrium with Eu-NDP MSH. Data were fit with a sigmoidal dose response curve with variable slope and analyzed using Graphpad.

[00146] For studying a potential interaction between CBD103 and ASIP-YY, we preincubated varying concentrations of ASIP-YY with 50 nM CBD103, and used the mixture in a displacement binding assay. In this paradigm, if CBD103 functions solely as a competitive ligand for $Mc1r$, then a mixture of CBD103 and ASIP-YY will behave additively with regard to Eu-NDP MSH displacement, apparent as a reduction in the maximal Eu-NDP MSH binding, and a slight leftward shift in the binding curve. However, if CBD103 binds to and sequesters ASIP-YY, the mixture of CBD103 and ASIP-YY will cause a rightward shift in the displacement curve.

[00147] Linkage and association mapping. The 12 Mb interval to which we previously mapped the KB mutation is defined by markers REN292N24 and FH3592. We ascertained additional kindreds segregating KB, k^{br} , and k^y ; among a total of 151 meioses from 8 breeds, we identified 22 chromosomes recombinant between the K locus and markers located in the original 12 Mb interval that were obtained from the RH map or developed from the dog genome sequence (markers M1 – M12). Four of these recombinant chromosomes, 1 from a k^{br}/k^y parent, and 3 from a K^B/k^y parent, define the smallest interval by linkage analysis that contains the K locus.

[00148] For association mapping, we chose to first examine Boxers, because brindle vs. yellow is a simple dominant trait that involves only the K locus, and because Boxers exhibit relatively long haplotype blocks; both features increase the power to detect association. In this manuscript we use the term "yellow" to refer to any pheomelanic dog; however, most dog breeders distinguish between "yellow" vs. "fawn", referring to colors caused by $Mc1r^p$ vs. the combination of a^y and k^y . Although there is specificity for some breeds with regard to which alleles segregate (e.g. fawn Boxers and fawn Great Danes never carry $Mc1r^p$), the genotypes cannot be distinguished simply on the basis of phenotype. For the purposes of consistency, therefore, we use the term "yellow" to refer to a pheomelanic color in all

breeds, with the understanding that within the dog breeding community, the same animal might be described as fawn (e.g. a Boxer or Great Dane).

[00149] Within the 3.8 Mb interval defined by recombination, we identified 267 polymorphisms with a targeted resequencing strategy. A subset of these polymorphisms were genotyped on 10 yellow (k^y/k^y) vs. 12 brindle (k^{br}/k^{br} or k^{br}/k^y) Boxers, and revealed a broad peak of significant association (Bonferroni-corrected p value <0.05) across 1.9 Mb. We then tested 10 black (K^B/K^B , K^B/k^{br} , or K^B/k^y) vs. 9 yellow (k^y/k^y) Great Danes, and identified a more restricted peak of significant association across 320 kb. The difference in the size of the regions between breeds correlates with the differences in linkage disequilibrium (LD) structure; in Boxers, highly correlated SNP pairs ($r^2=1$) are found up to 1.1 Mb apart, compared to 60 kb apart in Great Danes.

[00150] The 320 kb region of significant association in Great Danes focused our attention on the beta-defensin cluster as candidate genes; sequencing the mature protein-coding regions for 9 members of the cluster in 3 k^y/k^y animals (2 Boxers, 1 Great Dane) and 2 K^B/K^B animals (Great Dane, Labrador Retriever) revealed several polymorphisms concordant with the K^B allele, including the $\Delta G23$ mutation in CBD103. We then genotyped $\Delta G23$, 2 additional SNPs in CBD103, and 30 SNPs in the surrounding 1 Mb in 9 yellow and 11 black dogs from 12 different breeds. The $\Delta G23$ polymorphism was again completely concordant with the K^B allele, and together with the 2 additional SNPs in CBD103, revealed a highly significant association ($p < 10^{-6}$) only at the 3 CBD103 polymorphisms.

[00151] For the large 38 breed study, we classified dogs into two categories with regard to their putative K locus genotype. In the presence of functional Agouti and Mc1r alleles, the K^B allele causes exclusive production of eumelanin in areas that contain pigment cells; therefore black, chocolate, brown, liver, blue, harlequin, mantle, bicolor (black and white), and spotted black dogs should be heterozygous or homozygous for K^B . Conversely, the k^y allele permits production of pheomelanin in some or all body regions depending on the genotype of Agouti; therefore yellow, fawn, red, apricot, cream, wheaten, tricolor (black, yellow and white), black-and-tan, tan point, and sable dogs should be homozygous for k^y . Among 454 dogs, there were 13 cases where the $\Delta G23$ polymorphism did not correlate with coat color phenotype (Table 8), but could be explained by known epistatic interactions: 5 "black" animals (in the German Shepherd Dog or Weimaraner breeds) were homozygous for a loss-of-function Agouti allele, and 8 "yellow" animals (in the Afghan Hound, French Bulldog, German Shorthaired Pointer, Labrador Retriever, Puli, or Poodle breeds) were homozygous for a loss-of-function Mc1r allele.

Table 8: Association of the coat color phenotype with *CBD103* genotype^a

Breed	Genotype of pheomelaninistic dogs			Genotype of eumelanistic dogs		
	+/+	$\Delta G/+$	$\Delta G/\Delta G$	+/+	$\Delta G/+$	$\Delta G/\Delta G$
Afghan Hound	1				3	1
Akita	34					
Australian Shepherd	5				2	
Basenji	26				7	
Border Collie	2				1	1
Borzois	2					
Boxer	41					
Bulldog	27					
Bullmastiff	23					
Cavalier King Charles Spaniel	2					
Chow Chow	3				3	
Curly-Coated Retriever						5
Dalmatian						6
Doberman Pinscher	13					
Field Spaniel						5
Flat-Coated Retriever					3	1
French Brittany					1	
French Bulldog	1					
German Shepherd	7					
German Shorthaired Pointer						5
German Wirehaired Pointer						5
Great Dane	68				17	40
Greyhound	4				3	1
Havanese	1					
Irish Wolfhound	1					
Labrador Retriever					1	9
Large Munsterlander						3
Mastiff	7					
Miniature Doberman Pinscher	1					
Miniature Schnauzer	1					1
Poodle					2	3
Portuguese Water Dog						3
Pug	2					1
Puli	2				2	1
Rottweiler	10					
Shar-Pei	3				4	1
Weimaraner	2				4	4
Whippet	3					
Total	292	0	0	0	53	96

^a Alleles are denoted as "+" for the reference sequence (since Tasha the Boxer is k^Y/k^Y) and " ΔG " for the 3 bp deletion (CFA16:55755801-55755803) referred to as S54 in Table S2, and which predicts the *CBD103* $\Delta G23$ deletion described in the text. Not shown in this table are the genotypes of 13 dogs whose genotypes at *CBD103* were discordant with their coat color phenotype due to epistasis with *Mc1r* or *Agouti* mutations. Five "black" animals (in the German Shepherd Dog or Weimaraner breeds) were homozygous for a loss-of-function *Agouti* allele, and 8 "yellow" animals (in the Afghan Hound, French Bulldog, German Shorthaired Pointer, Labrador Retriever, Puli, or Poodle breeds) were homozygous for a loss-of-function *Mc1r* allele.

[00152] Short-range resequencing and haplotype studies. We resequenced all non-repetitive regions in a 20 kb interval surrounding *CBD103* in 5 yellow and 5 black dogs

representing 7 breeds, and identified 28 polymorphisms (22 SNPs and 6 indels including the $\Delta G23$ polymorphism) at a frequency > 0.2 , that were then used to infer short-range haplotypes. Along with the 6 “parental” k^y -bearing and 5 “parental” K^B -bearing chromosomes, 3 chromosomes carrying a single ancestral recombinant event defined a proximal boundary for KB 7.69 kb 5' of the $\Delta G23$ mutation.

[00153] To define a distal boundary, we tested 44 additional K^B/k^y dogs representing 14 breeds for a SNP that lay 9.65 kb distal to the $\Delta G23$ mutation (S103, Table 7), and identified 5 dogs, all Great Danes, that were homozygous for the k^y -associated S103 variant. Further genotyping revealed that 4 of the recombinant chromosomes define a distal boundary for KB 1.46 kb 3' of the $\Delta G23$ mutation. Thus, the maximal interval for KB defined by haplotype analysis is 9146 bp, and contains both exons of CBD103, the first exon of dog EST CX990240 and dog EST C0665262.

[00154] Evolution of K locus haplotype structure and breed origins. The haplotype analysis presented in Figure 6 focuses on short-range haplotype structure to identify ancestral recombinants. In retrospect, knowing that the $\Delta G23$ mutation in CBD103 is the KB mutation allowed an evaluation of long-range haplotype structure in Boxers and Great Danes, which revealed some interesting correlations with breed history. Although most breeds were formally established late in the 19th century, there is wide variation in the time at which typical characteristics for specific breeds appear in historical records. In particular, dogs similar in morphology to Great Danes can be recognized in records from the Middle Ages and earlier, while dogs similar in morphology to Boxers were not evident until much later. Using genotype data from the association study (Figure 5), we inferred haplotypes within a 2.6 Mb region surrounding the K locus for 12 brindle (5 k^{br}/k^{br} and 7 k^{br}/k^y) and 9 yellow (k^y/k^y) Boxers, and for 10 (7 K^B/K^B and 3 K^B/k^y) black and 9 yellow (k^y/k^y) Great Danes.

[00155] In Boxers, all 17 k^{br} chromosomes share a single long haplotype, 618 – 991 kb, centered on CBD103, whereas the 25 k^y chromosomes have multiple haplotypes over the same region. Furthermore, haplotype diversity in the 2.6 Mb region surrounding CBD103 is greater in Boxer k^y chromosomes than in Boxer k^{br} chromosomes. Records from the first Deutscher Boxer Club show in 1896 show both brindle and yellow dogs; thus, both alleles were represented early in breed derivation. Taken together, these observations suggest that k^{br} is a recent mutation that occurred prior to breed formation, and predict that differences between k^{br} and k^y haplotypes in Boxers will be found in other breeds that segregate brindle.

[00156] In Great Danes, the single haplotype shared by 17 KB chromosomes is 13 – 93 kb, considerably shorter than the k^{br} -associated haplotype in Boxers. Furthermore, haplotype diversity is similar in K^B and k^y chromosomes in Great Danes with 1.5 and 1.9

chromosomes per haplotype respectively. Both observations are consistent with a longer history and/or less stringent population bottlenecks in Great Danes relative to Boxers. Overall our data indicate that the $\Delta G23$ mutation probably arose in the ancestral dog population, estimated to carry haplotype blocks on the order of 10 kb (2), which is consistent with the 9 kb haplotype shared by 53 KB chromosomes from 12 breeds around CBD103 (Figure 6). In fact, a black dog with harlequin markings—a black and white pattern that requires the KB allele—is drawn on the Beni Hasan tomb of Khnumhotep II, with an estimated age of 1800 – 2000 B.C.

Example 3

Brindle Coat Color in Dogs

[00157] *Genotypes of brindle dogs at the $\Delta G23$ mutation in CBD103.* The genotype for the $\Delta G23$ mutation in *CBD103* was obtained for 459 dogs from 13 breeds that segregate the brindle coat color. The non-brindle dogs were yellow, fawn, red, apricot, cream, tricolor (black, yellow and white), black-and-tan, dogs with tan points, wheaten in Irish Wolfhound, and “fawn with black overlay” in Akitas. We hypothesized that these dogs would have the k^y/k^y genotype in the presence of a functional *Mc1r* gene. For simplification purposes, they will be referred to as yellow in this chapter. The brindle dogs were brindle or black with brindle points. They will be referred to as brindle and they must have a k^{br}/k^y or k^{br}/k^{br} genotype. This set of dogs was chosen to segregate k^y and k^{br} but not K^B , which is dominant to k^y and k^{br} , and is recognized by a black color of the coat and complete absence of yellow hair.

[00158] By sequencing the PCR amplified exon 2 of *CBD103*, it was determined that all 240 yellow dogs, with the exception of 1 French Bulldog, were homozygous for the wild-type allele at *CBD103*. This result was expected since yellow, k^y , is recessive to k^{br} and K^B . The discrepant yellow French Bulldog ($\Delta G23/+$) was found to be homozygous for the recessive yellow allele, *e*, at *Mc1r*, which is epistatic to the *K* locus, and causes yellow coat color. Interestingly, all 219 brindle dogs, including those known to be homozygous for the brindle mutation based on pedigree information, were heterozygous for the $\Delta G23$ mutation.

[00159] *Genetic evidence for a brindle specific rearrangement.* Evidence for a brindle specific rearrangement was also obtained by examining the genotypes of brindle and yellow Boxers for polymorphisms around the *K* locus. The genotypes of 10 yellow and 12 brindle Boxers sampled in the USA and in Europe were determined for polymorphic sites in 25 amplicons located in a 2.6 Mb region around *CBD103* and including *CBD103* exon 2. Genotypes for a subset of 28 biallelic polymorphisms (26 of which are SNPs) were selected for their potential to be in high linkage disequilibrium (LD) with the brindle mutation. For biallelic polymorphisms in complete LD ($r^2=1$) with the brindle mutation, the minor allele

frequency (m.a.f.) should be high (>0.25), yellow dogs should be homozygous for one allele, and brindle dogs should be heterozygous or homozygous for the other allele. Therefore, the selection criteria used to identify potential tag polymorphism of the k^y and k^{br} bearing haplotypes was m.a.f. >0.1 and homozygosity in most yellow Boxers, more specifically all but one of the 10 yellow Boxers. Twenty eight polymorphisms out of 67 with a m.a.f. >0.1 meet this selection criteria. These tag polymorphisms span a 1.6 Mb region around *CBD103* and reveal the simple haplotype structure of the locus in Boxers where the k^y and k^{br} alleles are found on one major haplotype each and recombinant haplotypes can be readily identified. In fact, in Boxers, the haplotype diversity is sufficiently low to allow the k^y and k^{br} carrying chromosomes to be identified using this selected set of tag polymorphisms hundreds of kbs from *CBD103*.

[00160] Except for S45, a SNP with over 20% of missing data, all brindle dogs examined were heterozygous at 6 consecutive SNPs adjacent to the $\Delta G23$ mutation in *CBD103* (S54). Five brindle dogs, BC61, BC22, BC123, DS175, and DS164, carried genotypes associated with brindle homozygosity for SNPs flanking the region of heterozygosity, suggesting that they are of the k^{br}/k^{br} genotype and in fact BC22 was known from mating data to be k^{br}/k^{br} . It was deduced from these genotype data that brindle dogs most likely carry a duplication that encompasses *CBD103* and part of the beta-defensin cluster. This duplication leads to apparent heterozygosity in the genotyping assay (PCR amplification followed by sequencing) for polymorphisms within the duplication in both k^{br}/k^y and k^{br}/k^{br} dogs. The minimum duplicated locus is from S40 and S54 (chr16: 55714953-55755801 bp, 41 kb) and the maximum duplication is from S36 to S56 (chr16: 55585724-55786080 bp, 200 kb). The minimum interval contains the beta-defensin genes, *SPAG11e* and *CBD103*. The maximal interval for the duplication contains the canine orthologs of mouse *Pasma6* and *Dub2*, and 10 of the 15 canine defensins identified in this region: *CBD107*, *CBD105*, *CBD106*, *CBD104*, *SPAG11c*, *SPAG11e*, *CBD103*, *CBD102*, *CBD138* and *CBD139*. *CBD1* is outside of the duplication. The maximal interval for the duplication is included in the 320 kb peak of greatest association found for the black mutation (K^B allele of the *K* series) in Great Danes.

[00161] k^y bearing haplotypes display a pattern of polymorphisms clearly different from k^{br} bearing haplotypes and therefore can be identified in k^y/k^{br} Boxers. Over this 2.6 Mb region, there are only 6 brindle haplotypes out of 17 k^{br} chromosomes, two of which are found once. Ignoring two potential gene conversions (apparent double recombination within less than 1 kb), all brindle chromosomes share a 618 to 991 kb haplotype around *CBD103*. By comparison, the genetic diversity in the yellow Boxers is greater: there are 15 different haplotypes carrying the k^y alleles out of 27 chromosomes, 11 of which are present once, and the region shared by all yellow Boxers around *CBD103* is only at most 30 kb in length (S53 to S56).

[00162] *Genomic structure of the CBD103 locus and missassembly of the Boxer genome sequence.* *CBD103* is one of 15 beta-defensin genes found in a 345 kb cluster on dog chr. 16. The maximal interval for the duplication contains a region of poor quality assembly that lies within the defensin cluster on chr16 between *CBD108* and *CBD104*. In fact there is a 61 kb interval (chr16: 55629000-55690000 bp) where SNP density in the Broad Institute SNP collection within the Boxer sequence is extremely high: 1 SNP/72 bp, compared to 1 SNP/1600 bp on average in the Boxer genome sequence. The region of high SNP density contains the breakpoint for a 5 Mb inversion of telomeric sequences between the CanFam1.0 and CanFam2.0 genome assemblies and is labeled as “uncertified assembly” in CanFam2.0. It is also an evolutionary break point between dog and mouse and dog and human. Analysis of the dog sequence showed the presence of repetitive sequences within the region of high SNP density. This region contains repeats that may not have been correctly assembled in the dog genome sequence and that may mediate evolutionary breakpoint occurrence.

[00163] *Fine mapping of the brindle mutation by identification of two reversions from brindle to yellow.* One yellow Boxer (BC11) carries brindle specific alleles for polymorphisms flanking the rearrangement but is homozygous for the yellow specific alleles at all polymorphisms within the rearrangement. Haplotype estimation for BC11 shows that this dog carries a k^y specific haplotype and a k^{br} specific haplotype on either side of the rearrangement, and two k^y specific haplotypes over the duplicated defensin cluster. The observed BC11 genotypes and haplotypes may be the consequence of a reversion of the brindle allele ($k^{br \rightarrow y}$) due to the loss of the brindle specific duplication.

[00164] BC11's dam, BC98, a yellow Boxer, was genotyped and shown to also carry the reversion. The pedigree of BC11 was obtained and showed that BC11 descends from a popular homozygous brindle (k^{br}/k^{br}) sire, Kiwi. Kiwi produced only brindle pups in over 100 matings with one exception, Joker, a yellow Boxer from which BC11 and BC98 are descended. Therefore it is likely that the brindle to yellow reversion occurred in a meiosis of Kiwi and that BC11 and BC98 inherited the brindle reversion allele from Joker.

[00165] Further support for the hypothesis that BC11 carries a reversion of a brindle allele was obtained from molecular analysis of *CBD103* copy number: BC11 only carries two copies of *CBD103* per diploid genome whereas brindle dogs carry more than 3.

[00166] A second potential revertant, a yellow Boxer sired by a presumed brindle homozygote, was identified in an Australian kennel. Genotyping of this dog, BC163, at 36 SNPs spanning a 1 Mb interval around *CBD103* showed that BC163 was homozygous for the $k^{br \rightarrow y}$ revertant chromosome identified in BC11 (Figure 10). The dam side of BC163 pedigree included a dog imported from the same kennel in the UK that produced BC11. Therefore, BC163 likely carries a revertant chromosome identical by descent to BC11's $k^{br \rightarrow y}$

k^y chromosome and a new reversion that occurred during meiosis in his K^{br}/K^{br} sire. It is not surprising that an independent reversion would look identical to the one previously identified since it occurred on the most frequent brindle chromosome in Boxers.

[00167] The brindle mutation maps to a duplication of the chromosome 16 defensin cluster and only one of these defensins, *CBD103*, is expressed in dog skin. In brindle, one copy of the defensin cluster carries the wild-type *CBD103* allele and the other copy carries the *CBD103-ΔG23* allele. Because the brindle phenotype consists of adjacent stripes of yellow and black hair in an apparently clonal pattern and since loss of the duplication in revertants results in the loss of stripes and yellow coat color, variegation of expression of the *CBD103-ΔG23* allele due to the presence of a duplication most likely causes brindle.

[00168] *Comparison of the LD structure in Great Danes and Boxers.* In order to compare the structure and LD pattern of the *K* locus between Boxers and Great Danes and look for evidence that the duplication is also present in the brindle, Great Danes, 9 yellow, and 12 brindle were genotyped for the *CBD103 ΔG23* mutation and at SNPs, located within a 2.6 Mb region around *CBD103*. While significant association with brindle coat color is evident throughout the 2.6 Mb interval in Boxers, association in Great Danes is much weaker and is only significant for three polymorphisms covering 384 kb. Similarly, LD in Boxers is more extensive than in Great Danes. Perfectly concordant SNP pairs ($r^2=1$) are found up to 890 kb apart in Boxers but only up to 276 kb apart in Great Danes. Also, unlike the situation in Boxers, haplotype diversity is similar in brindle and yellow Great Danes. Nevertheless, brindle Great Danes most likely also carry a duplication since they are all heterozygotes at four of the five polymorphisms present within the duplication but not at any other polymorphisms.

[00169] *Origin of the brindle specific duplication.* In an effort to determine the origin of the brindle duplication, the haplotypes of brindle, yellow and black dogs were compared in 13 breeds. 15 amplicons were sequenced in 19 black, 17 brindle and 25 yellow dogs. Estimated haplotypes at the 36 SNPs identified with a m.a.f. >0.1 and the *ΔG23* mutation in *CBD103* are shown in Figure 10. All brindle dogs are heterozygous for all 10 SNPs within the duplication and polymorphic in brindle but not at any other SNPs in a 1 Mb interval around *CBD103*. The haplotype estimation compresses the duplicated defensin cluster into one cluster over which one finds two haplotypes from S39 to S54, that will be referred to as A and B. Homozygotes, k^{br}/k^{br} , carry both A and B. One can deduce that the actual haplotype over the brindle specific duplication is composed of the juxtaposition of haplotypes A and B. All brindle dogs from six breeds have the same juxtaposed A-B haplotypes. Haplotype A (from S39 to S54 mostly colored in yellow and outlined in gray in Figure 10) is only found in yellow k^y chromosomes and not in any black K^B chromosomes. It is the most frequent k^y haplotype (0.65 frequency) and is found in five out of six breeds

examined. Haplotype B (colored in blue and outlined in red on Figure 10) is found in one black Great Dane and is a K^B specific haplotype. A haplotype identical at 10 of the 11 SNPs is the most frequent haplotype in K^B chromosomes (frequency 0.5) and is present in 5 out of 10 breeds genotyped.

[00170] Therefore it can be concluded that the brindle chromosome is most likely to have arisen a single time by a duplication of part of the beta-defensin cluster. The segmental duplication must have occurred by a rearrangement between a common k^Y chromosome and a K^B chromosome.

[00171] *CBD103 copy number determination.* To confirm the genetic evidence for a genomic rearrangement in brindle dogs, we used quantitative PCR to measure the copy number of a genomic locus located 2.3 kb upstream of *CBD103* in black, brindle and yellow dogs from three breeds. As expected, three yellow Bulldogs, two yellow Boxers, three yellow Great Danes and seven black Great Danes have two copies of *CBD103* per diploid genome. Brindle dogs, however, had more than two copies: seven brindle Bulldogs, one brindle Boxer and three brindle Great Danes had an estimated three to five copies of the *CBD103* locus. Therefore, brindle dogs, but not black or yellow dogs, carry at least a duplication of the *CBD103* locus.

[00172] *Mechanism of stripe formation.* Because the $\Delta G23$ mutation in *CBD103* causes black and all brindle dogs carry at least a duplication of the locus that includes one copy of the *CBD103-ΔG23* mutation, we hypothesized that the genomic rearrangement found in brindle dogs leads to the variegated expression of the *CBD103-ΔG23* allele in the skin, thereby causing stripes. To test this hypothesis, *CBD103* mRNA levels were compared between skin biopsies underlying a yellow and a black stripe in a single brindle dog (D6). Quantitative RT-PCR using five different primer pairs showed approximately two fold more total *CBD103* RNA in the black stripe than the yellow stripe (Figure 4-5A). The mRNA levels of *CBD1*, the only other defensin on the CFA 16 defensin cluster that is expressed in the skin (Chapter 3), and *Agouti* were not different in the black and yellow stripe. Sequencing of the cDNA extracted from a black and a yellow stripe showed that both the *CBD103* and *CBD103-ΔG23* alleles are transcribed in the black stripe but only the wild-type *CBD103* allele is transcribed in the yellow stripe (Figure 11). Therefore, variegated expression of the *CBD103-ΔG23* allele is the most likely mechanism of brindle stripe formation.

[00173] *Mechanism for gene expression variegation.* Variegated expression of *CBD103* could be caused by either somatic rearrangements of the defensin locus or epigenetic silencing or activation of gene expression. We tested the somatic rearrangement using two different techniques: allele ratio and copy number quantification.

[00174] A somatic rearrangement in yellow stripes leading to the deletion of the *CBD103-ΔG23* alleles would lead to an altered ratio of wild-type *CBD103* and *CBD103-ΔG23* DNA

sequence. To look for evidence of somatic rearrangement, the ratio of the two alleles was determined in the yellow and black stripes of two brindle dogs, D6 and D16. To improve the detection of a somatic rearrangement, the dermis and epidermis were separated and the ratio of the alleles was determined in both. In these two brindle dogs there was no difference in allelic ratios in the dermis and epidermis. Since *CBD103* is presumed to be expressed in keratinocytes in dogs as in humans and since keratinocytes constitute most of the epidermis, a somatic rearrangement in keratinocyte clones should have been detected if present using this technique.

[00175] We also quantified copy number of the *CBD103* locus in the D6 sample and found no difference in copy number between the yellow and black stripe in either the dermis or the epidermis.

[00176] We conclude that the variegation in *CBD103* expression must not be caused by the somatic rearrangement of the *CBD103* locus but rather by an epigenetic mechanism. This is consistent with the fact that germline losses of the duplication appear to be rare compared to the extremely frequent clonal variation in keratinocytes as seen on the brindle coat. It is interesting to notice that one of the brindle samples, D6, showed a ratio of *CBD103* to *CBD103-ΔG23* allele of 0.65, when 1 was expected for a brindle homozygote dog (k^{br}/k^{br}) and 2 for a brindle heterozygote (k^{br}/k^y). Consistent with this observation, this sample was determined to have six copies of the locus per diploid genome by quantitative PCR, although that estimate is not precise due to limitations in the technique used. Based on the ratio and the total copy brindle duplication and one containing a triplication of the defensin cluster with one cluster that has the wild-type *CBD103* allele and two clusters that have the *CBD103-ΔG23* allele.

[00177] The identification of the molecular cause for the brindle mutation leads to a model for the evolution of the *K* locus. First the $\Delta G23$ mutation arose at least 4000 years ago as shown by historical records of black dogs on Egyptian tombs dated at around 2000 B.C. In fact because black dogs only share a 9 kb haplotype, which is in the range of the predicted size for haplotype blocks in the large ancestral dog population, the mutation is likely to have arisen even more than 4000 years ago. More recently the brindle mutation, k^{br} , arose by a rearrangement between a K^B and a k^y chromosome in a black dog and the same k^{br} chromosome causes the brindle phenotype in at least the six dog breeds we have studied. The haplotype shared between brindle chromosomes from 6 different breeds is long (200 kb) indicating that brindle arose much later than the black mutation. Historical records show that brindle greyhounds were present 1000 years ago in Britain so the brindle mutation has to be at least 1000 years old. Interestingly brindle dogs are found in many breeds that are well spread out in the current dog breed phylogeny and representative of the four major clusters of dog breeds identified by molecular analysis. The brindle color is found in breeds

of ancient origin, such as Akitas and Shiba Inu, and in breeds related by phylogeny to some herding types, such as Greyhound and Irish Wolfhound. Brindle is rarely found in the cluster of breeds of recent European origin and it is found in most breeds related to Mastiffs such as the Mastiff, Bullmastiff, Bulldog, Boxer, Bull Terrier and Great Dane.

[00178] Once segmental duplications are formed, they are prone to non-allelic homologous recombination (NAHR). NAHR between homologous chromosomes carrying the segmental duplication results in deletions or further increase in copy number of tandem duplications, and inversions of the sequence located between duplications in head to head orientation. We have documented two reversions from brindle to yellow due to the loss of one copy of the defensin cluster in the duplication and we have found one dog, D6, who appears to carry a triplication of the locus. These two occurrences suggest that the two defensin clusters in the brindle specific duplication are arranged in tandem orientation and that the duplication is evolving by rare NAHR between the two clusters.

[00179] Brindle stripes are caused by the variegated expression of the *CBD103-ΔG23* allele presumably induced by the duplication of the defensin cluster. *DEFB103*, the ortholog of *CBD103* in humans, is expressed in keratinocytes and the pattern of the brindle stripes resembles that of the lines of Blaschko seen in human conditions affecting keratinocytes clones. Therefore the brindle stripes are most likely keratinocyte clones that have silenced or activated the expression of the *CBD103-ΔG23* allele.

[00180] Segmental duplications are largely believed to cause disease or phenotypic variation by altering gene dosage of dosage-sensitive genes. Other mechanisms by which segmental duplications cause phenotypic variation involve the interruption or fusion of genes or the disruption of regulatory regions of genes that flank the duplication. The brindle mutation demonstrates a segmental duplication-induced epigenetic mechanism that is analogous to the epigenetic phenomenon of position-effect variegation studied primarily in *Drosophila* and yeast where a rearrangement locates a euchromatic gene close to a region of heterochromatin and leads to its variegated expression.

Materials and Methods

[00181] *Genome sequence.* The dog genome assembly used is CanFam1.0 as displayed and annotated in the UCSC genome browser. The CanFam2.0 assembly contains an inversion of the 5 Mb most telomeric on CFA16. The breakpoint of the inversion is the defensin cluster. This inversion is incompatible with the linkage map, two radiation hybrid maps, a FISH map and our own mapping data for the black and brindle mutations.

[00182] *Dog DNA samples and skin biopsies collection.* Catch-All swabs (Epicenter) or Cytopak brushes (Medical packaging) were used to collect cheek swab samples from the dogs of consenting owners. Swabs were collected through the mail or at dog shows. DNA from the swabs was prepared using the Qiamp DNA mini kit (Qiagen). DNA samples were

collected regardless of the relatedness of the dogs. For the association study, all dogs collected were genotyped regardless of their relatedness. For the haplotype mapping, only dogs not known to share parents and in general owned by different breeders were included. RNA from skin biopsies was stored in RNAlater (Ambion) was prepared using the Trizol reagent (Invitrogen).

[00183] *SNPs and indel genotyping.* All primers were designed using Primer3. 1 µl of cheek swab DNA preparation was PCR amplified in a 20 µl reaction. The PCR reaction was cleaned up with ExoSAP-IT (USB) and sequenced using BigDye Terminator v3.1 (Applied Biosystems). The sequencing reactions were purified using isopropanol precipitation and run on an ABI 3730 (Applied Biosystems). Sequences were analyzed using CodonCode Aligner 1.5.1c (CodonCode Corporation) or Sequencer 4.2 (Gene Codes Corporation). The list of polymorphisms, SNPs and indels, can be found in Table 5 (example 2)

[00184] *SNP discovery.* Putative SNPs were obtained from the Broad Institute collection and validated in two brindle and two fawn Boxers. Amplicons for the SNPs showing different alleles between brindle and fawn Boxers were selected and all polymorphisms within these amplicons were scored.

[00185] *Genotyping of the $\Delta G23$ mutation in *CBD103*, and the *R306ter* mutation in *Mc1r*.* All genotyping was done by PCR amplification and sequencing of the PCR products. The primers for $\Delta G23$ mutation genotyping were 5'-TGTCTTCATCCCTGTGAGGT-3' and 5'-CCAGGAGGCATTTTCACACT-3' (396 bp amplicon). The primers used to genotype the *R306ter* mutation in *Mc1r* were 5'-TCTTTGTAGCCATGCTGGTG-3' and 5'-ATCCACCACACCACAGATCA-3' (486 bp amplicon).

[00186] *Data analysis.* In general only SNPs for which at least 75% of genotypes were obtained were included in the analysis. Also to simplify the analysis and the display of haplotypes, only SNPs for which the minor allele frequency (m.a.f.) was greater than 0.1 in any dataset were considered. Haplotypes were estimated using Phase v2.1. Statistical tests of association were performed using Haploview. The Bonferroni correction was used to define the 0.05 level of significance for the association with the coat color phenotype. This correction is too conservative since the genotype of the loci are not independent from each other due to the presence of LD. Low copy repeats were identified using Miropeats.

[00187] *Copy number determination.* Multiple primers sets, designed to genomic sequences surrounding *CBD103* and *Pomc*, were compared by quantitative PCR, using a Lightcycler thermocycler (Roche, Indianapolis, IN), to identify a pair of amplicons with consistently similar PCR efficiencies. An amplicon pair meeting this criterion was then used to quantify defensin copy number in individual dogs. This amplicon was located on CFA16:

55752227-55752346, 3.5 kb proximal to the $\Delta G23$ mutation in *CBD103*. Threshold cycle values, generated by the lightcycler software, were determined for both amplicons at three dilutions of template (genomic DNA isolated from blood or cheek swabs). The difference in threshold cycle values was then used to calculate relative copy number per diploid genome. The primers used are as follows: *CBD103* 5'-TCCCCTCGTCAACCTTATCA-3' 5'-CTGGAACCTCGAGTCATGGA-3'; *Pomc* 5'-ATCACCGTGGGCAAGTGTA-3'; 5'-TGGAAATCATCCCAGAATGG-3'.

[00188] *RNA extractions, quantitative RT-PCR.* Total RNA was isolated from brindle dog skin biopsies containing either black or yellow hair using an RNeasy fibrous tissue RNA extraction kit (Qiagen, Valencia, CA). 1 μ g of total RNA was then used as template for oligo-dT primed reverse transcription, and expression levels of *CBD1*, *CBD103*, and *Agouti* were determined by quantitative PCR using *Bactin* expression as a reference for relative quantification. All analysis was done using the Lightcycler software package (Roche, Indianapolis, IN). The primers used are as follows: *CBD* 1F1 5'-GGCCTCTCTACTTGCTGCTG-3'; *CBD1R1* 5'-CTCCTTTCCTGGCACAGATG-3'; *CBD1F2* 5'-GGCCTCTCTACTTGCTGCTG-3'; *CBD1R2* 5'-AGGGGAGAAGTTGCAGGTG-3'; *CBD103F1* 5'-TGTTCCTTGATGCCTGTTCCA-3'; *CBD103R1* 5'-GGAAGAACAGCGGCCTATCT-3'; *CBD103F2* 5'-GCCTGTTGGTCATGAGGATCT-3'; *CBD103R2* 5'-GACCGCTCCTTATTCTGCAA-3'; *CBD103F3* 5'-TGTTCCTTGATGCCTGTTCCA-3'; *CBD103R3* 5'-GGAAGAACAGCGGCCTATCT-3'; *CBD103F4* 5'-AGATAGGCCGCTGTTCTTCC-3'; *CBD103R4* 5'-CCAGGAGGCATTTTCACACT-3'; *CBD103F5* 5'-TGCCAAAGGAGGAGCAGATA-3'; *CBD103R5* 5'-CCAGGAGGCATTTTCACACT-3'; *AgoutiF1* 5'-ACTCCTCTGTGAACCTTTTGG-3'; *AgoutiR1* 5'-ACGTTCTTCATCGAAGCCTTT-3'; *BactinF1* 5'-CGACAGGATGCAGAAGGAAA-3'; *BactinR1* 5'-ACATCTGCTGGAAGGTGGAC-3'.

[00189] *DNA copy number ratio determination.* Three biopsies from the yellow and black stripes of two brindle dogs were obtained (12 biopsies total). The dermis and epidermis of the skin biopsies were split by incubating the skin in 2M sodium bromide at 37°C for 4 hours. DNA was subsequently extracted and PCR amplified using standard protocols and the following primers: 5'-6FAM-ATCTGGGGAATTCCAAAAGC-3', 5'-GACCGCTCCTTATTCTGCAA-3'. These primers amplify different sized products for the *CBD103- $\Delta G23$* allele (133 bp) and the wild-type *CBD103* allele (136 bp) that can be separated on polyacrylamide gels and quantified. Four different dilutions of each PCR product were run on an ABI3730 (Applied Biosystems) and the dilutions giving the greatest fluorescent signal within the detection range for each biopsy were selected. The area under the peak of the *CBD103- $\Delta G23$* allele and the wild-type *CBD103* allele were determined

using the software GeneMapper 3.7 (Applied Biosystems). The ratio of the area under the peak is also the ratio of the alleles.

[00190] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, accession numbers, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

What is Claimed is:

1. A method for predicting the coat color phenotype of a first canine, the method comprising:
genotyping a first genetic sample from the first canine for the presence of at least one K locus allele.
2. The method of Claim 1, comprising determining the presence of a K^B allele.
3. The method of Claim 2, comprising further determining the presence of a K^{br} allele.
4. The method of Claim 3, further comprising determining the presence of a K^Y allele.
5. The method of Claim 4, further comprising determining the copy number of said K alleles.
6. The method of Claim 5, further comprising genotyping the first genetic sample for the presence of at least one Mc1R allele.
7. The method of Claim 6, comprising determining the presence of an Mc1R wild type allele.
8. The method of Claim 7, comprising determining the presence of an $Mc1R^o$ allele.
9. The method of Claim 8, further comprising genotyping the first genetic sample for the presence of at least one Agouti allele.
10. The method of Claim 9, comprising determining the presence of an Agouti A allele.
11. The method of Claim 10, comprising determining the presence of an Agouti a allele.
12. The method of any one of Claims 1-11, further comprising:

providing an analysis of the possible coat color phenotypes of progeny from said first canine.

13. The method of any one of Claims 1-11, further comprising genotyping a second genetic sample from a second canine for the K locus.

14. The method of Claim 13, further comprising genotyping the second genetic sample for at least one of the Agouti and Mc1R loci.

15. The method of Claim 14, further comprising:
providing an analysis of the possible coat color phenotypes of progeny from a cross of said first canine and said second canine.

16. The method according to any one of claims 1-15, further comprising determining the genotype the genetic sample at one or more additional loci involved in coat color determination.

17. The method according to any one of claims 1-16, wherein the genotyping comprises specifically amplifying a subsequence of a gene encoding CBD103.

18. The method according to any one of claims 1-17, wherein the presence of a K locus allele is determined by sequencing.

19. The method according to any one of claims 1-17, wherein the presence of a K locus allele is determined by hybridization with an allele specific probe.

20. The method according to any one of claims 1-17, wherein the presence of a K locus allele is determined by sizing an amplification product of the CBD103 gene.

21. A kit for predicting the coat color phenotype of a canine according to the methods of any one of Claims 1-20.

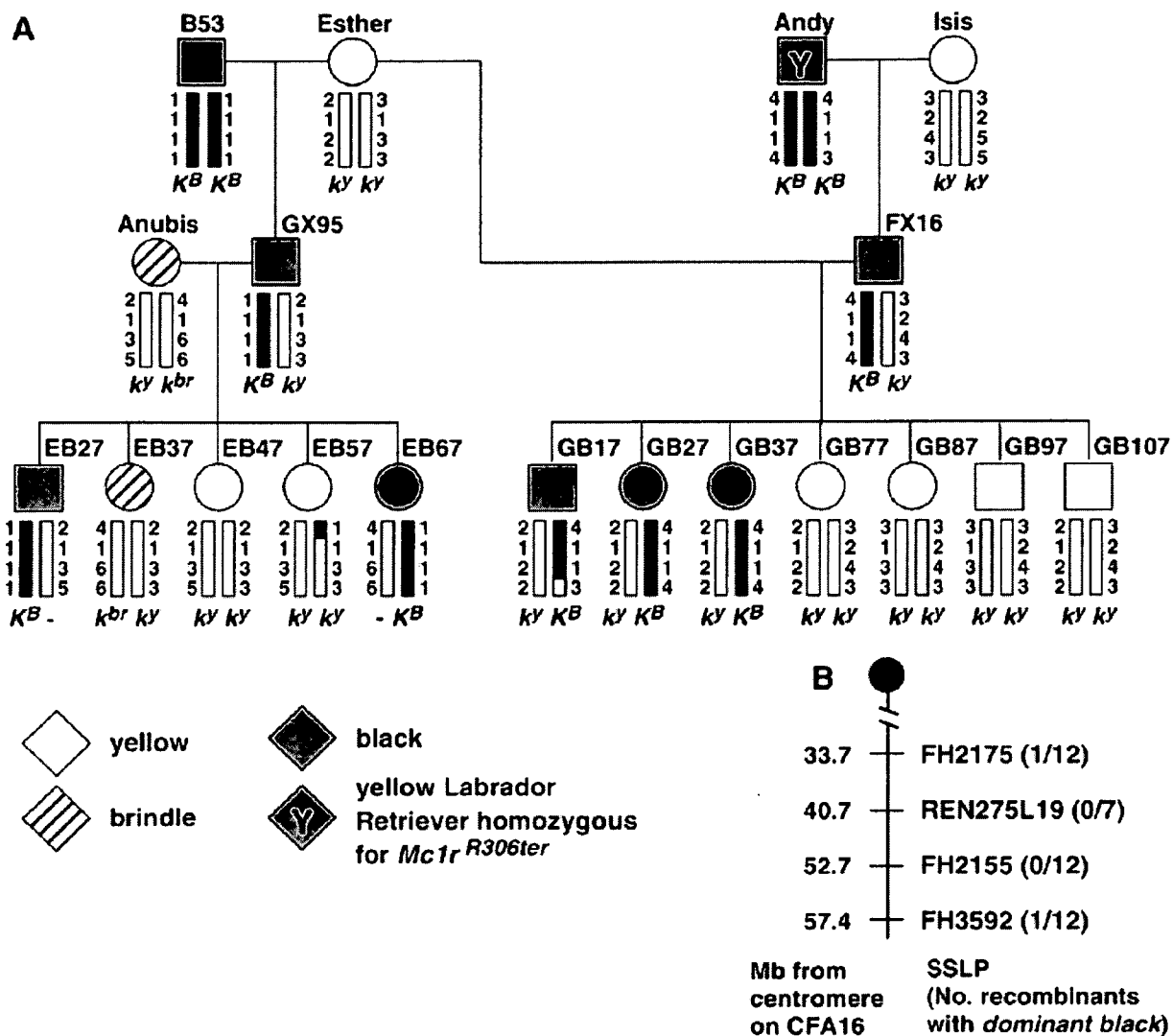
22. The kit according to Claim 21, comprising primers for amplification of the K locus.

23. The kit according to Claim 22, comprising at least one K locus allele specific probe.

24. The kit according to Claim 22, comprising primers for amplification of at least one of the Agouti and Mc1R loci.

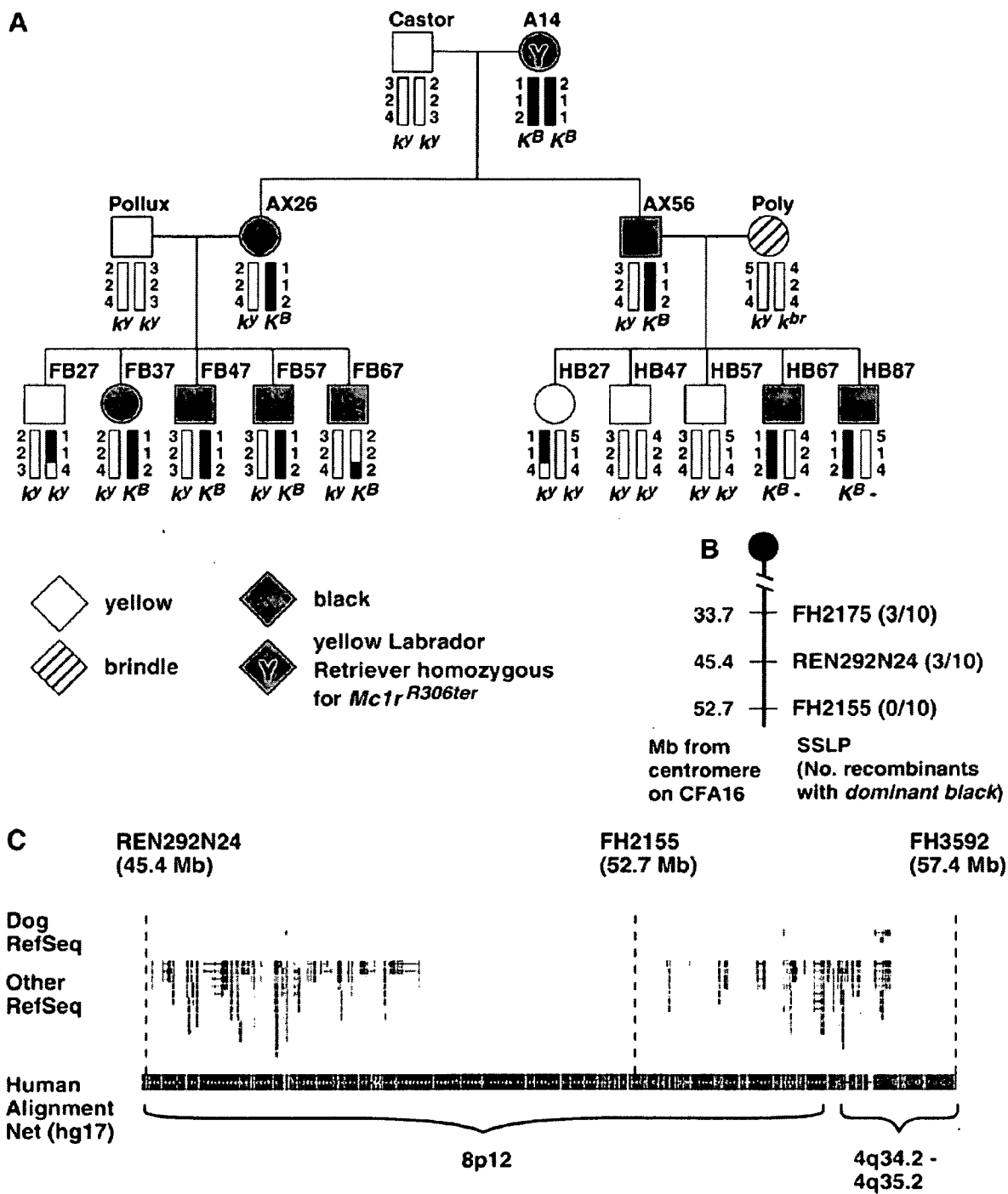
1/14

FIGURE 1



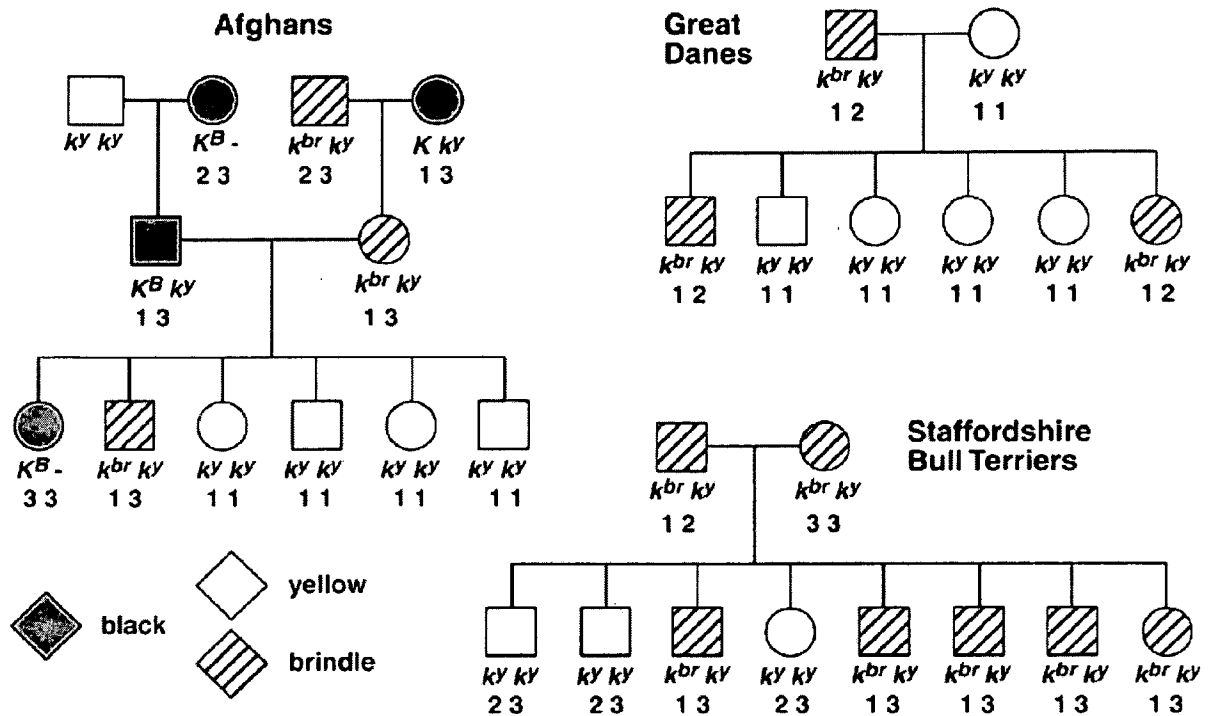
2/14

FIGURE 2



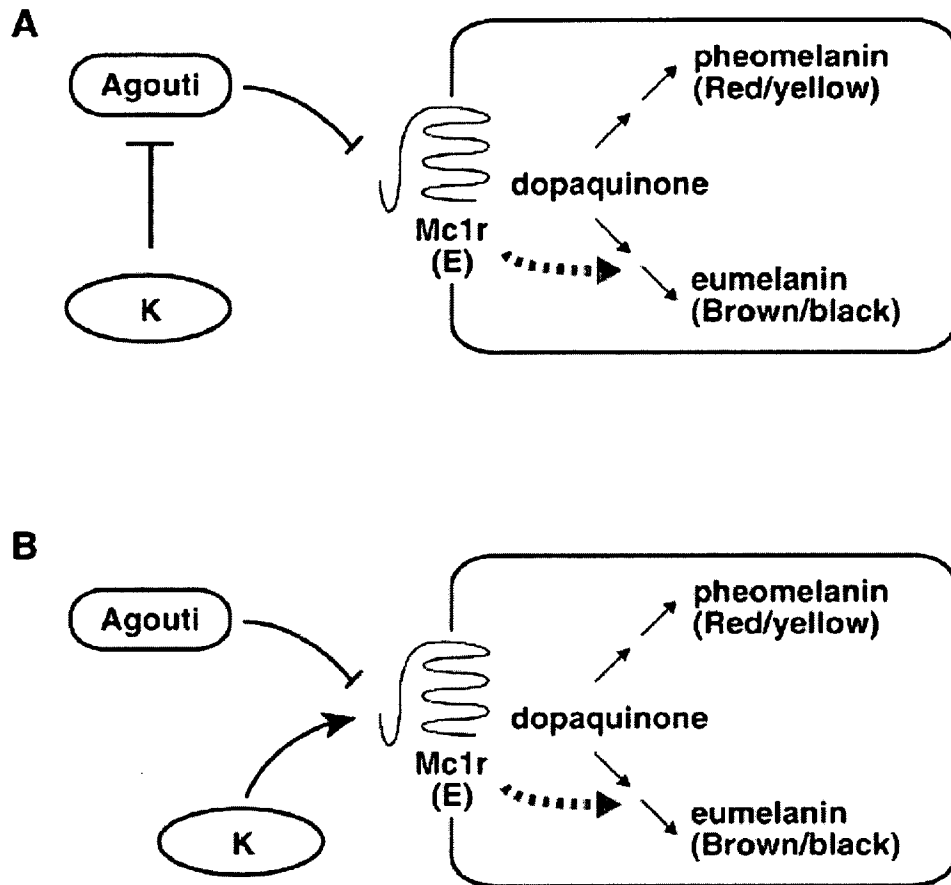
3/14

FIGURE 3



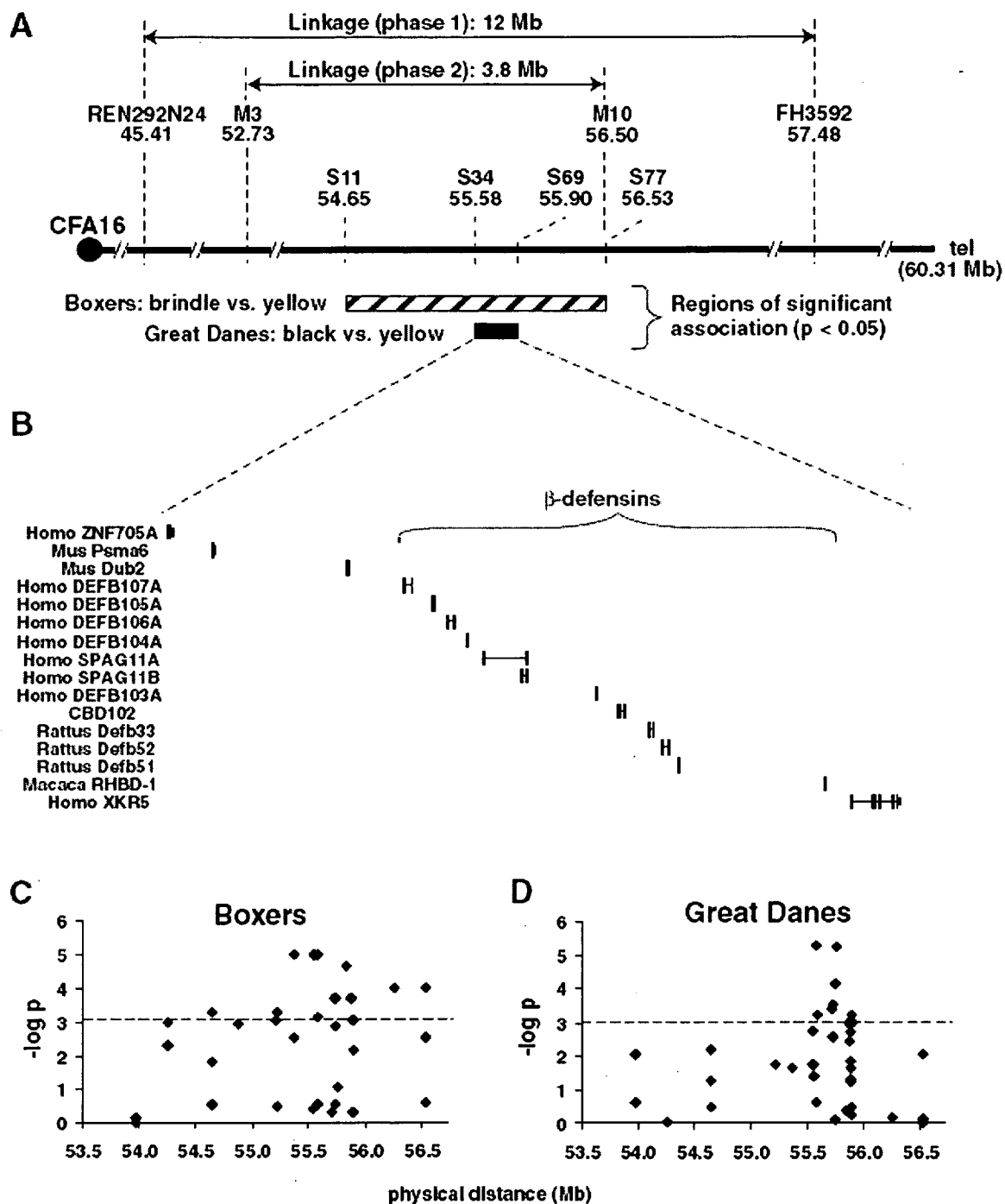
4/14

FIGURE 4

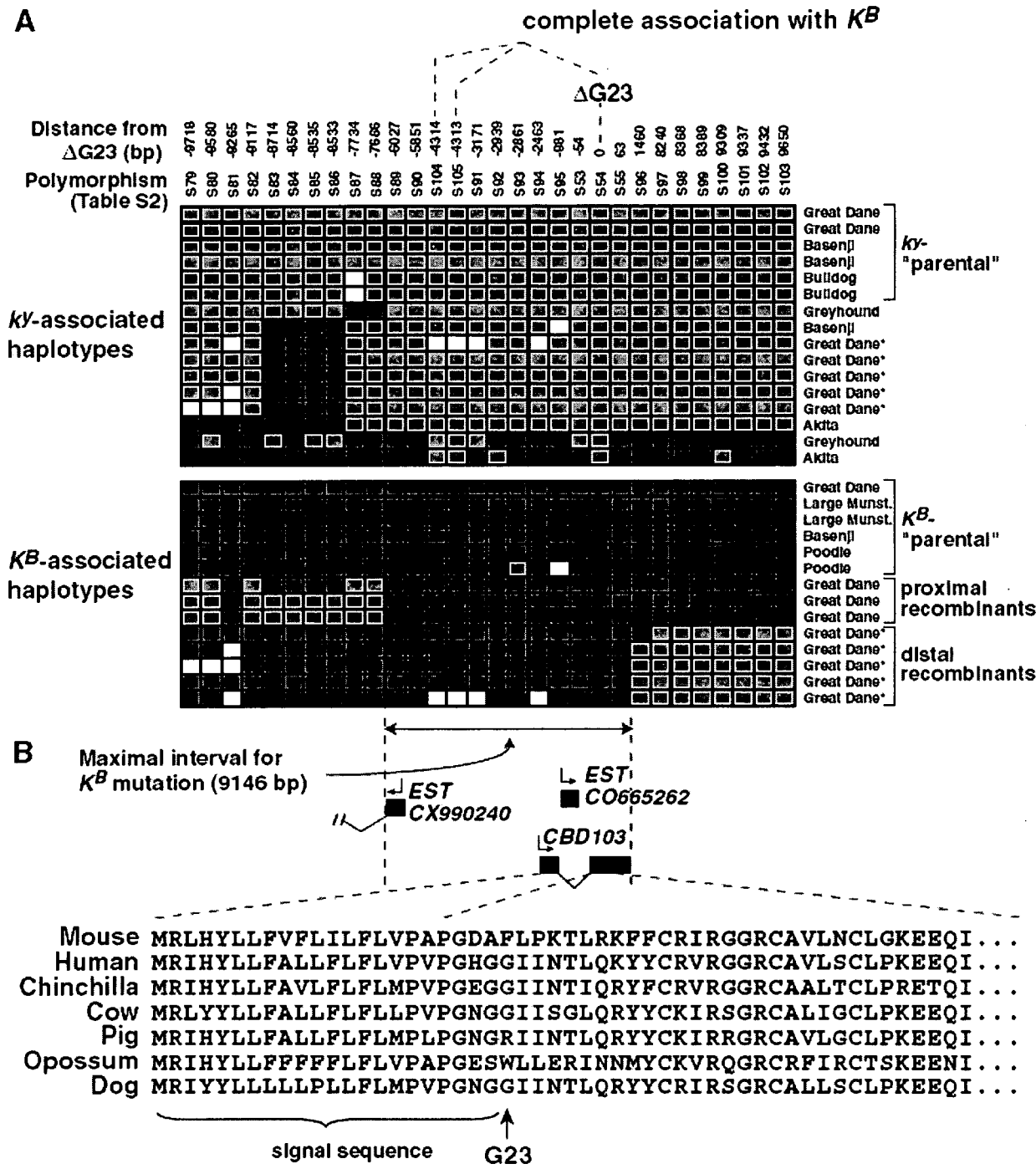


5/14

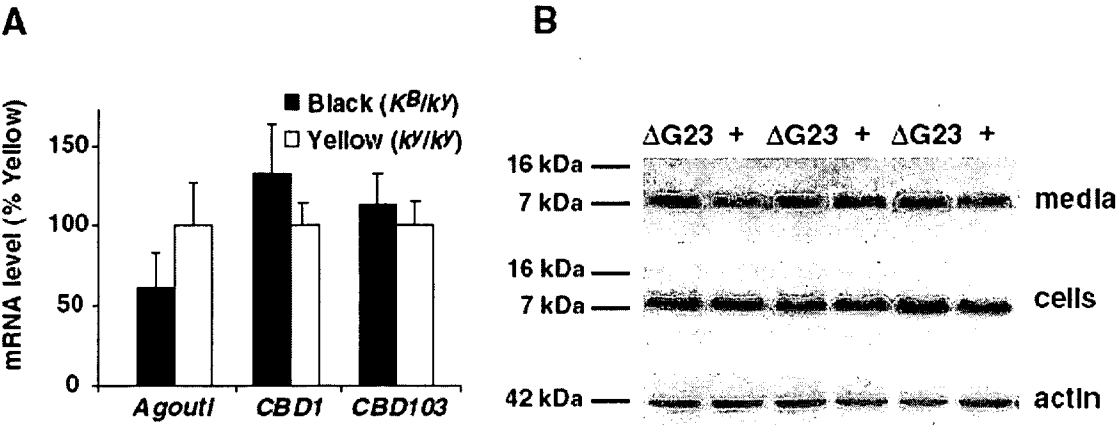
FIGURE 5



6/14
FIGURE 6

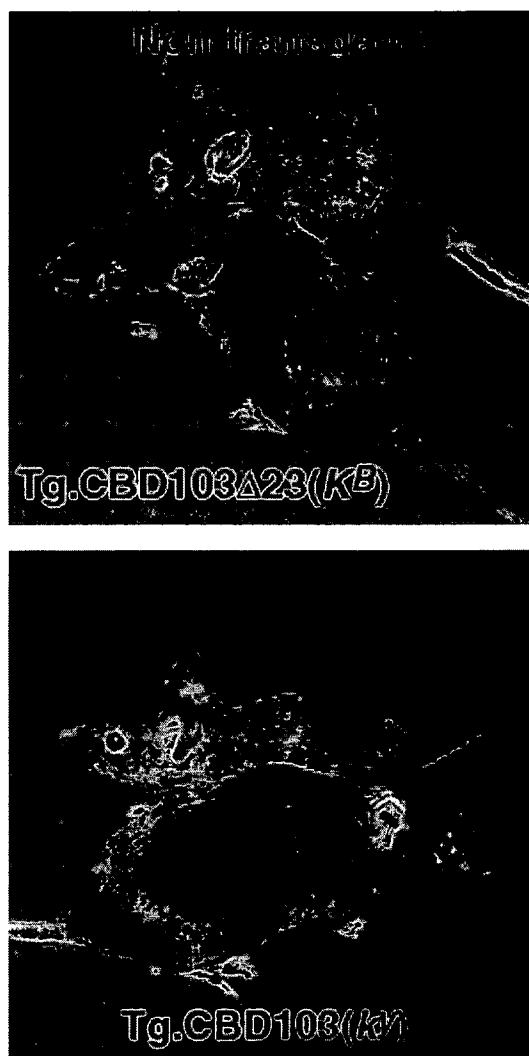


7/14
FIGURE 7



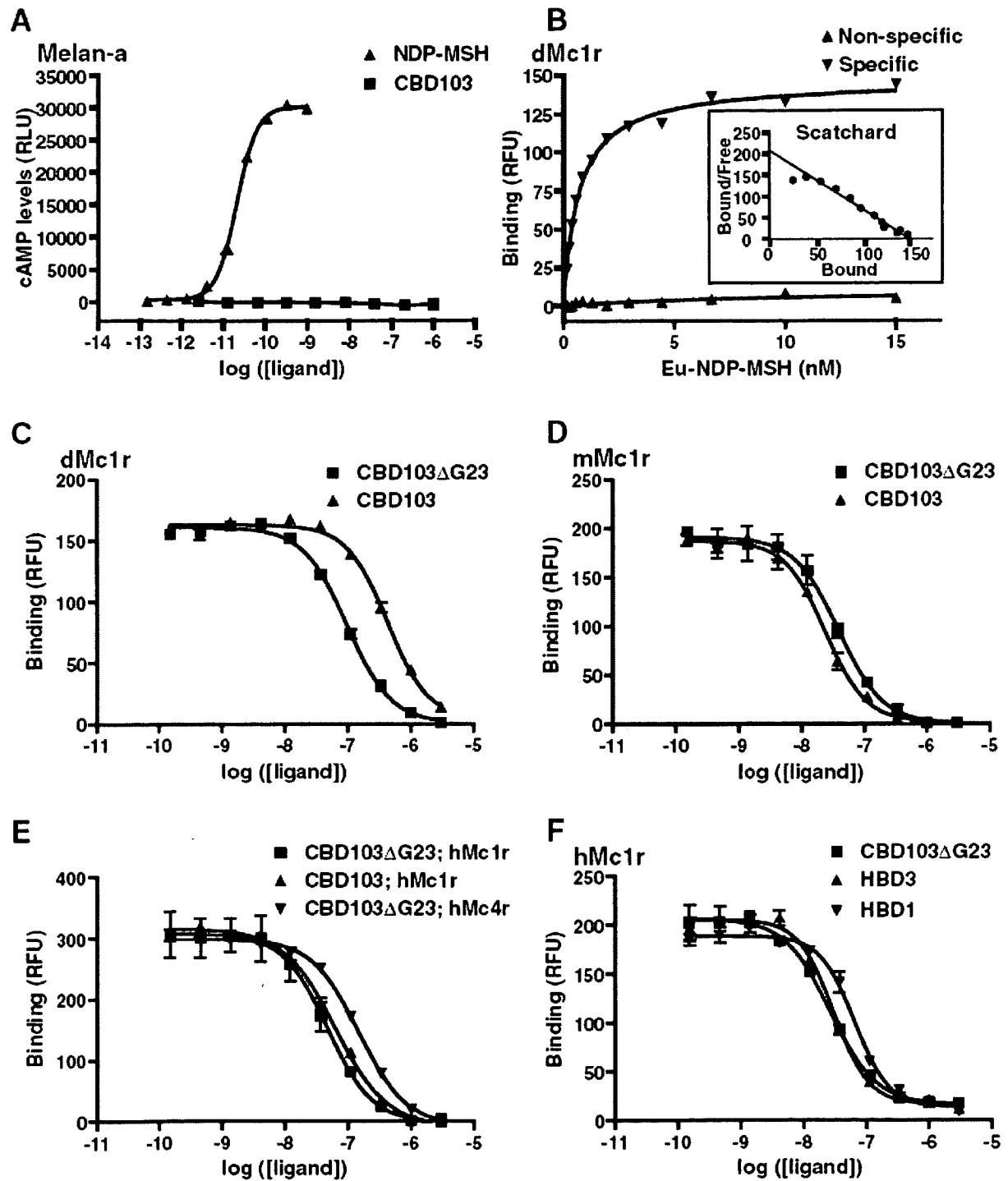
8/14

FIGURE 8



9/14

FIGURE 9



10/14
FIGURE 10

Multiple Breeds: haplotypes

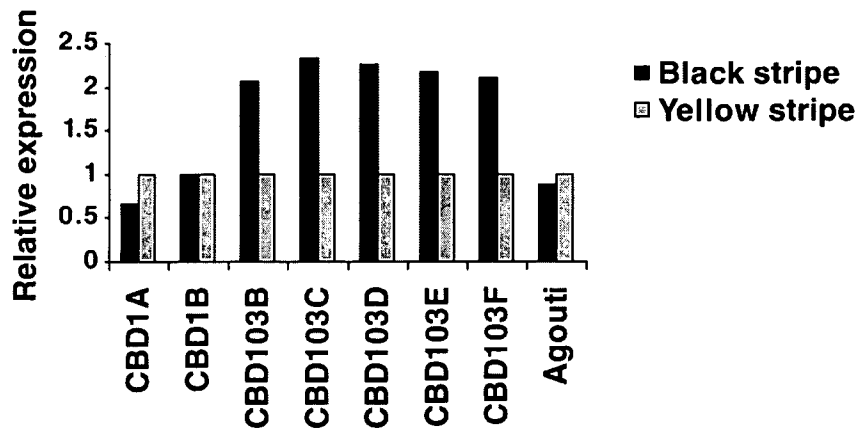
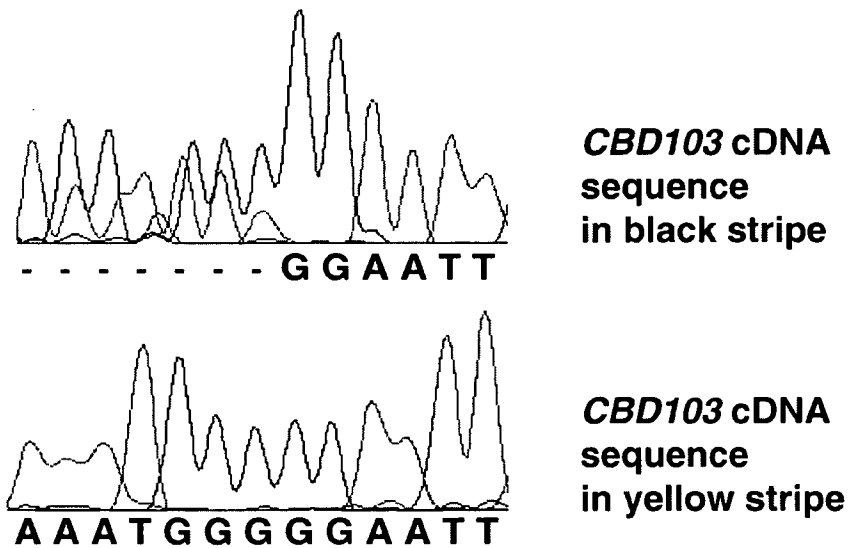
55.50 55.75 56.00 56.25 56.50 Mb

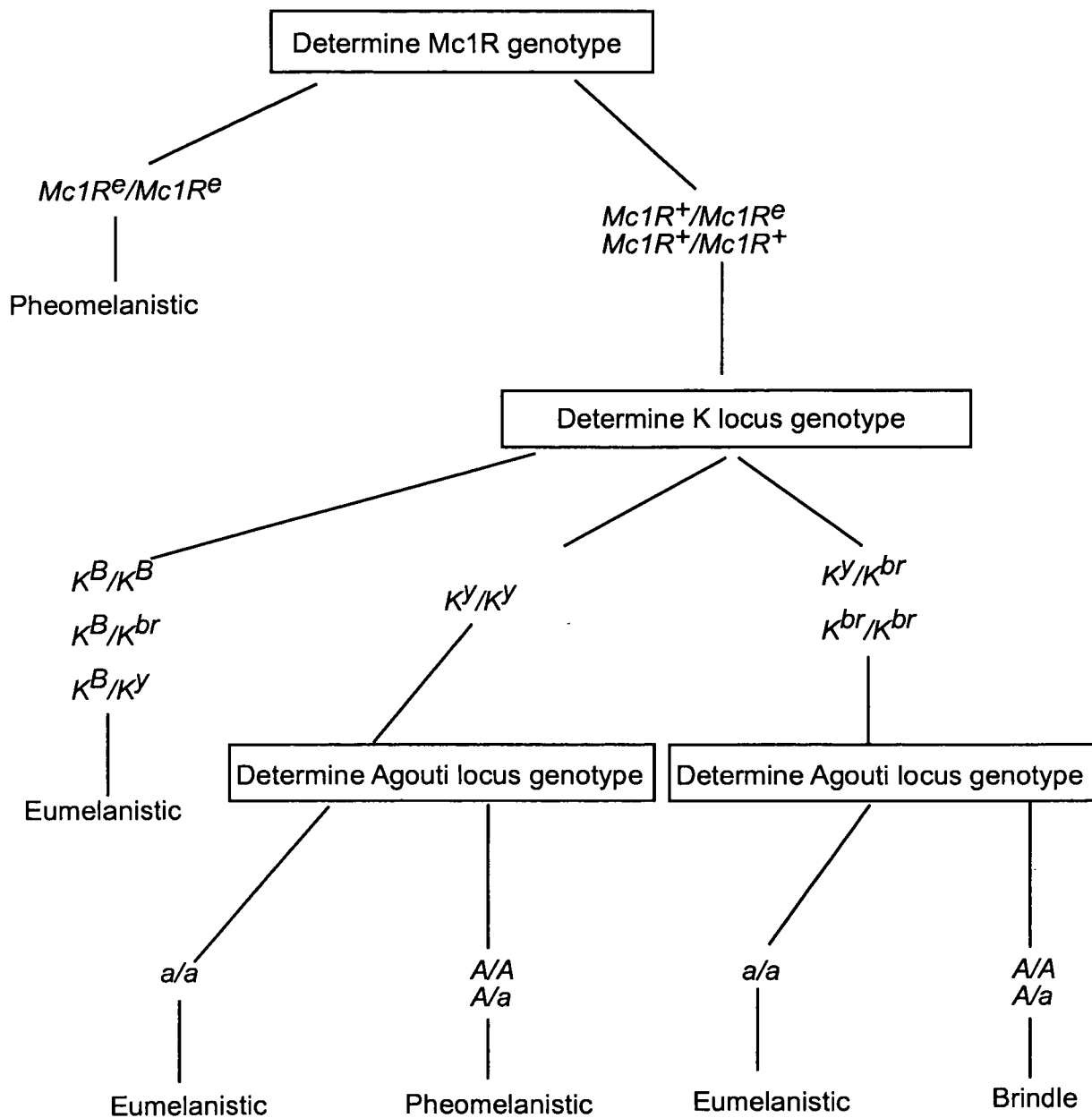
Polymorphism

 K^B
haplotypes k^{br}
haplotypes k^Y
haplotypes $k^{br} \rightarrow y$ 

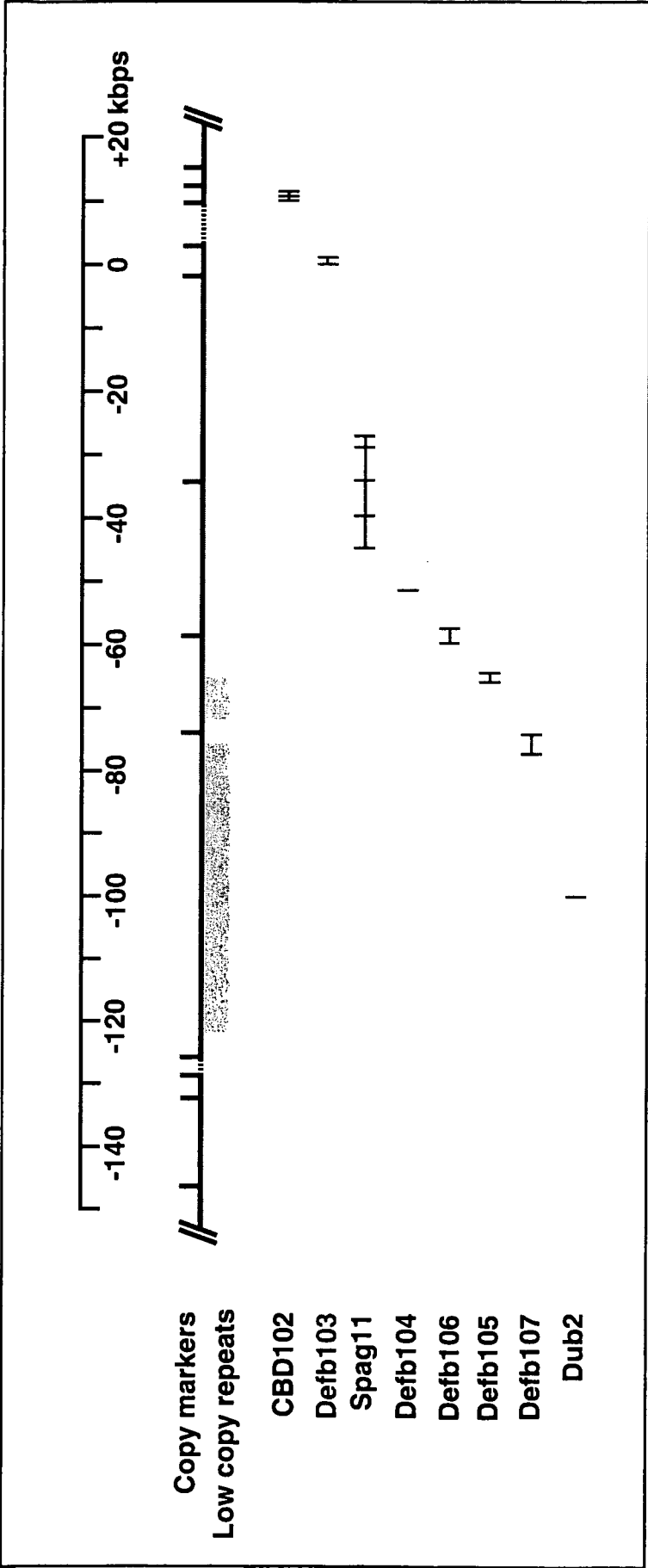
11/14

Figure 11

A Gene expression in brindle skin**B**

12/14
FIGURE 12

13/14
Figure 13



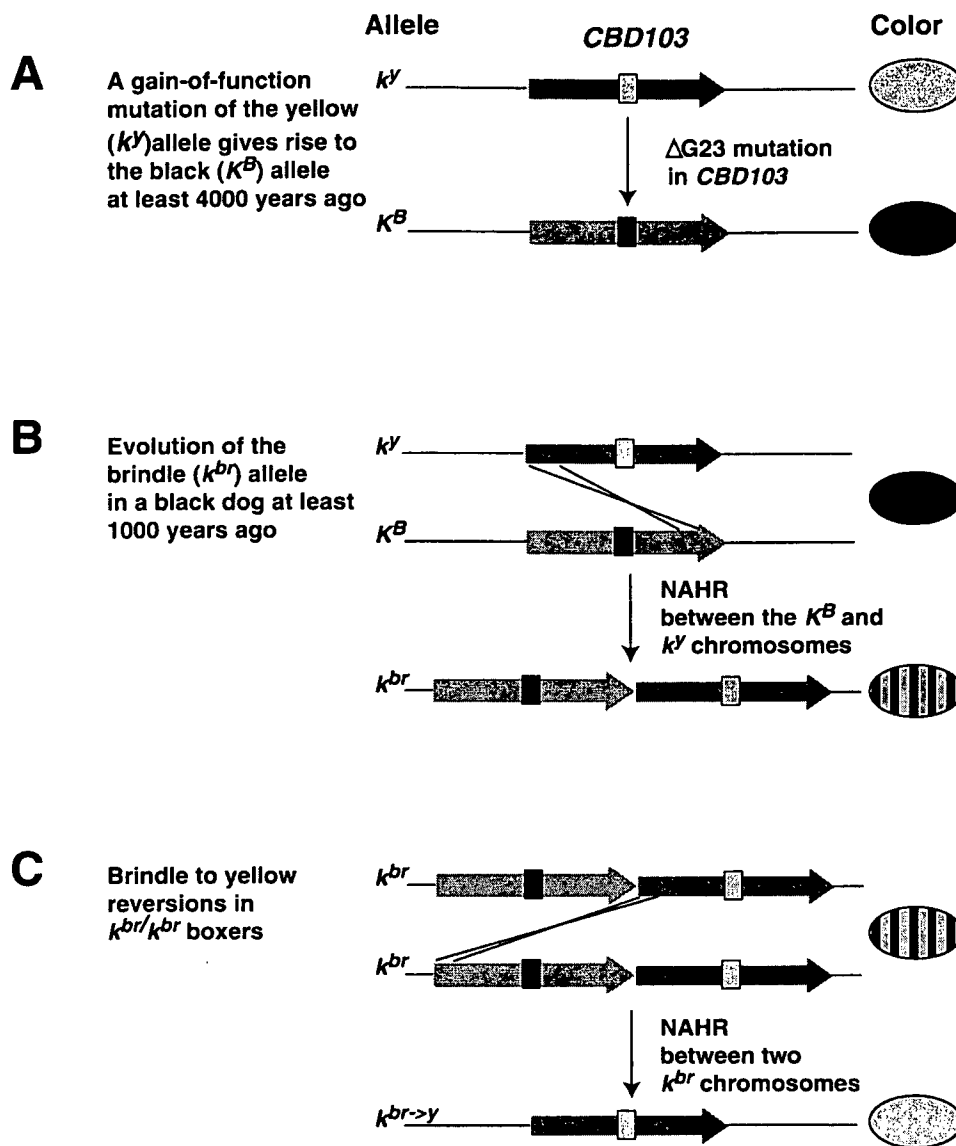
14/14
Figure 14

FIGURE 14

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 08/05666

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12Q 1/68; G01N 33/48 (2008.04)

USPC - 435/6; 702/20

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8) - C12Q 1/68; G01N 33/48 (2008.04)

USPC - 435/6; 702/20

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

USPC - 435/104

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PubWEST(USPT,PGPB,EPAB,JPAB); Google Patents; Google Scholar

Search terms: canine or dog, coat or fur or hair, color, gene, locus, allele, K, KB, Kbr, Ky, Mc1r or melanocortin receptor, Agouti, predict, copy number

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	KERNS et al., Linkage and segregation analysis of black and brindle coat color in domestic dogs, Genetics, July 2007, Vol 176, No 3, pp 1679-89 (p 1681, col 1, para 1; p 1681, Table 1; p 1681, col 1, para 2-col 2, para 1; p 1682, Fig. 1).	1-4 and 12-15 ----- 5-11
Y	CANDILLE et al., A Beta-Defensin Mutation Causes Black Coat Color in Domestic Dogs, Science, 30 November 2007, Vol 318, No 5855, pp 1418-1423 (p 1422, col 3, para 1).	5-11

☐ Further documents are listed in the continuation of Box C.


* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

19 August 2008 (19.08.2008)

Date of mailing of the international search report

08 SEP 2008

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 08/05666

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 16-24
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.