Supporting Material

Materials and Methods

Sample collection

DNA samples from Canadian (*S1*), Yellowstone (*S2*), and Italian (*S3, S4*) wolves were extracted from blood, pelts, or tissues. DNA samples from domestic dogs were collected by cheek swab from owners and breeders according to a Stanford IACUC-approved protocol; all samples were prepared, as described (*S5*). DNA samples from 61 gray and 2 black coyotes were provided through a generous collaboration with the United States Department of Agriculture (John Pingley and Bill Bonwell, USDA Wildlife Services, West Virginia; Chad Fox, Harold McDaniel Jr., and Eric Wilhelm, USDA Wildlife Services, Virginia). Tissue samples from an additional 4 black coyotes were provided by Ed Smallwood, Mike Laska, and Matthew Boggs.

Amplification and sequencing of *Agouti*, *Mc1r*, and *CBD103*

The wolf and coyote *Agouti* genes were analyzed in 4 amplicons as follows: 5'-CACCCAACACACTTCTGCG-3' and 5'-TACCATACCAAAACATCTGC-3' (283 bp, exon 1); 5'-AGGGCACAGCCTCTTATCAA-3' and 5 '-CAGGGCTTTTCCAAACCATA-3' (617 bp, exon 2); 5'-CACCCTGAGACTTCCTGGAG-3' and 5'-GAGGCCAAGAAGCCTTTAGA-3' (295 bp, exon 3); 5'-AAGTCCAGCGGACAGTCG-3' and 5'-CACACCTTGGAGCAGCCTA-3' (636 bp, exon 4). [Previous results (*S6, S7*) indicate the potential for alternative 5' untranslated exons in canid *Agouti*; the amplicon containing exon 1 described here corresponds to a cDNA isolated from a Doberman Pinscher and lies in a genomic location homologous to what has been described as the ventral-specific exons from the mouse *Agouti* gene (*S8*)]. Additional SNPs from intronic regions of *Agouti* are listed in Table S2; primer sequences are available on request. The wolf and coyote *Mc1r* genes were analyzed in 3 amplicons covering a single large coding exon: 5'-CACTTGTACAGACCGGGAGAG-3' and 5'-ACGTCAATGATGTCGTCCAG-3' (493 bp, first third); 5'-GTGACGAATGTGCTGGAGAC-3' and 5'-AAATGCCCAGCAGGATAGTG-3' (484 bp, central third); 5'-TCTTTGTAGCCATGCTGGTG-3' and 5'-ATCCACCACACCACAGATCA-3' (486 bp, last third). The wolf and coyote *CBD103* mature coding regions (the location of the K^B ΔG23 mutation) were analyzed in a single amplicon, 5'-TGTCTTCATCCCTGTGAGGT-3' and 5 '-CCAGGAGGCATTTTCACACT-3' (396 bp).

For PCR, a touchdown protocol was used with the following conditions: (1) 94°/90 sec; (2) 94°/30 sec; 65°/30 sec (-0.5°/cycle); (4) 72°/60 sec; (5) repeat steps 2 – 4 x 20; (6) 94°/30 sec; (7) 55°/30 sec; (8) 72°/60 sec; (9) repeat steps 6 – 8 x 20; (10) 72°/5 min. Direct sequencing of amplified product after primer hydrolysis with ExoSapIt (USB, Cleveland, OH) was carried out with standard fluorescent dye terminator technology on a capillary instrument.

Genomic and bioinformatic analysis of *CBD103* and flanking regions

Genome sequence coordinates used throughout this paper refer to the dog genome assembly CanFam2.0 (*S9*) as displayed and annotated in the UCSC genome browser,

http://genome.cse.ucsc.edu/ (*S10*). In this assembly, *CBD103* G23 lies at chr16:61,902,782-61,902,785, very close to the end of chromosome 16 (62,570,175); however, our previous studies suggest the last 5 Mb of the chromosome are mistakenly inverted in the CanFam2.0 assembly (*S5*).

For population genetic analysis of *CBD103* and surrounding regions, 8 single-copy noncoding segments distributed across a ~150 kb region centered on *CBD103* were analyzed with nested primers, whose position, sequence, and associated polymorphisms are given in Tables S2 and S3. Amplification with outer primers was carried out with the following conditions: (1) 92°/60 sec; (2) 92°/20 sec; (3) 65°/20 sec (-0.6°/cycle); (4) 72°/30 sec; (5) repeat steps 2 – 4 x 11; (6) 92°/20 sec; (7) 58°/20 sec; (8) 72°/30 sec; (9) repeat steps 6 – 8 x 20; (10) 72°/3 min. Then, amplification with inner primers was carried out with the following conditions: (1) 92°/60 sec; (2) 92°/20 sec; (3) 58°/20 sec (4) 72°/30 sec; (5) repeat steps 2- 4 x 29; (6) 72°/3 min. Other conditions for amplification were carried out as described (*S11*), which permits direct sequencing after amplification without primer hydrolysis.

All sequencing results were analyzed with CodonCode Aligner (CodonCode Corporation, Dedham, MA) or Sequencher 4.2 (Gene Codes Corporation, Ann Arbor, MI), and for every polymorphism between or within a species, the trace files were inspected visually for confirmation.

Nucleotide diversity and polymorphism statistics were determined with DnaSP 4.0 (*S12*), haplotypes for the region surrounding *CBD103* were inferred with PHASE 2.1 (*S13*), extended haplotype homozygosity statistics and haplotype bifurcation diagrams were calculated and visualized with Sweep (*S14*), LD values were calculated and visualized with Haploview (*S15*), phylogenetic analysis of the *CBD103* core region was carried out with MEGA (*S16*), and the mtDNA network (Fig. S3) was constructed with TCS (*S17*). The TMRCA molecular clock analysis was carried out, as described (*S18*); estimates from 5' and 3' polymorphisms were combined, averaged, and presented as a function of absolute distance (independent of direction) from *CBD103*.

Supporting text

Population genetics and evolutionary history of the K^B mutation

We estimated allele frequencies of 0.19 and 0.02 for K^B in forest and tundra wolves, using coat color frequencies reported for 11 different Arctic populations; 4 forest populations representing a total of 68 wolves, and 7 tundra or taiga populations representing a total of 336 wolves (Fig. 1A, *S1*). These estimates assume that 50% of the gray wolves carry K^B but are non-penetrant due to graying with age.

Our estimate for the age of the k^y to K^B mutation derives from dog resequencing data. In previous studies, we identified 4 K^B/K^B dogs (2 Great Danes, 1 Large Munsterlander, 1 Poodle) in which haplotype analysis delineated a 9.1 kb interval containing *CBD103* that lacked evidence of historical recombination (*S5*). Among these 8 K^B chromosomes, we identified 3 mutations from 63984 bp of high-quality sequence. Assuming a mutation rate of 1×10^{-9}/yr and a star-shaped genealogy, these data predict that K^B occurred 46,886 years ago, with a 95% confidence limit of 12,779 - 121,182 years. This is a minimal estimate because the dogs we analyzed may sample only a subset of all K^B chromosomes.

However, regardless of whether the original k^y to K^B mutation occurred prior, during, or after domestication of dogs from wolves, the haplotype data (Fig. 3A, Fig. S3), the TMRCA analysis (Fig. 3C), and the geographical distribution of K^B (which is much broader in dogs than in wolves) suggest that North American gray wolves acquired K^B from dogs and not vice versa.

Genetic relatedness among wolves and dogs: effect of the K^B mutation

The higher frequencies of K^B in forest compared to tundra wolves that we (*S1*) and others (*S19 – S21*) have observed suggest that melanism and/or K^B provide a selective advantage in forested habitats. An alternative hypothesis is that K^B was introduced recently into North American gray wolves, and that the differences in habitat-specific allele and phenotype frequencies are due to genetic drift. If so, one would expect black wolves to carry additional regions of the dog genome besides the K^B-bearing segment on chromosome 16, as a signature of residual hybridization with dogs.

To explore this possibility, we analyzed large-scale genotype data obtained with the Affymetrix v2 platform, which represents 49,663 SNPs distributed widely across the canine genome. These data are part of a larger study that will be described in detail elsewhere (R.K.W., C.D.B., E.A.O., B.M.V.). We examined a subset of the data selected for geographic origin and *K* locus genotype to investigate if K^B was associated with unlinked markers due to population structure, and to determine if there was cryptic relatedness between dogs and wolves carrying K^B.

We first analyzed genotypes from 10 Canadian wolves (5 K^B/k^y and 5 k^y/k^y) and 10 Yellowstone wolves (5 K^B/k^y and 5 k^y/k^y), and asked whether any SNPs other than those closely linked to *CBD103* were associated with K^B. From 28,739 SNPs with call rates > 80% and a minor allele frequency > 0.05, 32 lie within 2 Mb of *CBD103*, but outside the core region previously used to define *K* locus haplotypes (Figs. 2B, 3A, S2). In wolves, none of these 32 SNPs were strongly associated with K^B (Table S5). As a more sensitive test, we carried out a principal component analysis of the 20 wolves using progressively larger genomic regions: SNPs on the distal 10 Mb of dog chromosome 16 (144), all SNPs on the same chromosome as *CBD103* (739), and all SNPs (28,739). For each set of SNPs, one or more of the major principal components separated Canadian from Yellowstone wolves but did not distinguish K^B/k^y from k^y/k^y individuals (Fig.S3A, S3B), which indicates that geography is the major source of genetic differentiation, and suggests that the K^B-associated segment introgressed from dogs into black wolves was not accompanied by other large regions of the dog genome.

To probe potential relatedness between North American wolves and dogs, we considered data for the 20 wolves together with 20 dogs from different breeds and 5 wolf-dog hybrids, and applied principal component analysis to the full set of ~45,000 SNPs. For this dataset, the first principal component separated dogs and wolves, leaving the Italian dog-wolf hybrids at an intermediate position, while the second principal component separated the hybrids from both the dogs and the North American wolves (Fig. S3C). As with the previous analysis of wolves alone, K^B/k^y individuals were not separated from k^y/k^y individuals.

As a complementary approach, we used mitochondrial haplotypes previously obtained for 37 Canadian wolves (*S1*) to construct a phylogenetic network, and then projected

information about *K* locus genotype onto that network. As shown in Fig. S4, mitochondrial DNA haplotypes for K^B-carrying wolves are broadly distributed, and no more closely related to each other or to dog haplotypes than to haplotypes for non-K^B wolves. Taken together, these observations suggest that the geographic differences in K^B allele frequencies are explained by a positive selection for K^B in forested compared to open habitats, and not by genetic drift following dog-wolf hybridization.

Supplemental references

S1. M. Musiani *et al.*, *Mol Ecol* 16, 4149 (2007).

S2. B. M. Vonholdt *et al.*, *Mol Ecol* 17, 252 (2008).

S3. E. Randi, V. Lucchini, *Conservation Genetics* 3, 29 (2002).

S4. E. Fabbri *et al.*, *Mol Ecol* 16, 1661 (2007).

S5. S. I. Candille *et al.*, *Science* 318, 1418 (2007).

S6. D. I. Vage *et al.*, *Nat Genet* 15, 311 (1997).

S7. J. A. Kerns *et al.*, *Mamm Genome* 15, 798 (2004).

S8. H. Vrieling, D. M. Duhl, S. E. Millar, K. A. Miller, G. S. Barsh, *Proc Natl Acad Sci U S A* 91, 5667 (1994).

S9. K. Lindblad-Toh *et al.*, *Nature* 438, 803 (2005).

S10. D. Karolchik *et al.*, *Nucleic Acids Res* 31, 51 (2003).

S11. B. M. Smits *et al.*, *Pharmacogenet Genomics* 16, 159 (2006).

S12. J. Rozas, J. C. Sanchez-DelBarrio, X. Messeguer, R. Rozas, *Bioinformatics* 19, 2496 (2003).

S13. M. Stephens, N. J. Smith, P. Donnelly, *Am J Hum Genet* 68, 978 (2001).

S14. P. C. Sabeti *et al.*, *Nature* 419, 832 (2002).

S15. J. C. Barrett, B. Fry, J. Maller, M. J. Daly, *Bioinformatics* 21, 263 (2005).

S16. K. Tamura, J. Dudley, M. Nei, S. Kumar, *Mol Biol Evol* 24, 1596 (2007).

S17. M. Clement, D. Posada, K. A. Crandall, *Mol Ecol* 9, 1657 (2000).

S18. H. Tang, D. O. Siegmund, P. Shen, P. J. Oefner, M. W. Feldman, *Genetics* 161, 447 (2002).

S19. P. S. Gipson *et al.*, *Wildlife Society Bulletin* **30**, 821 (2007).

S20. L. D. Mech, L. G. Adams, T. J. Meier, J. W. Burch, B. W. Dale, *The Wolves of Denali*. (University of Minnesota Press, 1998).

S21. R. O. Peterson, J. D. Wollington, T. N. Bailey, *Wildlife Monographs* **88**, 3 (1984).

S22. C. Vila *et al.*, *Science* 276, 1687 (1997).

Table S1. *Agouti* and *Mc1r* variation in black and non-black wolves

Gene	Position[1]	Location[1]	SNP[2]	Effect[2]	No. black[3]			No. non-black[3]		
					R/R	R/V	V/V	R/R	R/V	V/V
Agouti	chr24:26,351,138	5' flank	G/A	No	4	0	0	5	1	0
Agouti	chr24:26,351,277	5' flank	G/A	No	0	0	4	0	1	5
Agouti	chr24:26,357,184	Intron 1	A/G	No	0	1	3	0	1	5
Agouti	chr24:26,357,895	Intron 1	T/G	No	0	0	3	0	1	5
Agouti	chr24:26,358,085	Intron 1	C/G	No	0	0	3	0	1	5
Agouti	chr24:26,359,694	Intron 1	G/A	No	3	0	0	4	1	0
Agouti	chr24:26,359,752	Intron 1	T/C	No	0	1	2	0	1	4
Agouti	chr24:26,360,917	Intron 1	G/A	No	0	1	3	0	1	5
Agouti	chr24:26,360,980	Intron 1	G/A	No	0	1	3	0	1	5
Agouti	chr24:26,361,818	Intron 2	T/A	No	0	1	3	0	1	2
Agouti	chr24:26,363,092	Intron 3	G/A	No	3	1	0	5	1	0
Agouti	missing	Coding	G/T	A82S	4	0	0	5	1	0
Agouti	chr24:26,365,984	Coding	G/A	No	4	0	0	5	1	0
Agouti	chr24:26,366,507	3' flank	C/T	No	4	0	0	4	2	0
Agouti	chr24:26,366,958	3' flank	G/C	No	3	1	0	3	3	0
Mc1r	chr5:66,693,084	Coding	A/G	S90G	10	0	0	16	1	0
Mc1r	chr5:66,692,869	Coding	G/A	No	10	0	0	16	1	0

[1]Location in genome from CanFam 2.0 (*S9*); position given relative to *Agouti* or *Mc1r* transcript. The region of the *Agouti* gene coding for codon 82 is missing from the dog assembly.

[2]SNP information given as reference sequence/variant sequence, with predicted effect, if any, on protein.

[3]No. of individuals according to genotype of Reference and Variant alleles. Wolves surveyed for variation in *Agouti* were from Canada. Wolves surveyed for variation in *Mc1r* were from Canada (21) or Yellowstone National Park (6); we also sequenced *Mc1r* in 2 Swedish wolves (not included in the Table) and did not observe any variation.

Table S2. *Agouti* and *Mc1r* variation in black and gray coyotes

Gene	Location[1]	Position[1]	SNP[2]	Effect[2]	No. black[3]			No. gray[3]		
					R/R	R/V	V/V	R/R	R/V	V/V
Agouti	chr24:26,326,144	5' flank	C/T	No	0	0	2	1	0	3
Agouti	chr24:26,327,291	5' flank	T/C	No	0	0	2	0	1	3
Agouti	chr24:26,327,318	5' flank	A/G	No	0	0	2	0	1	3
Agouti	chr24:26,327,427	5' UTR	T/G	No	2	0	0	3	1	0
Agouti	chr24:26,366,050	Coding	C/G	No	1	0	5	2	2	8
Agouti	chr24:26,366,076	3' UTR	G/C	No	3	3	0	6	0	6
Agouti	chr24:26,366,591	3' flank	G/C	No	2	0	0	2	1	0
Agouti	chr24:26,366,623	3' flank	C/A	No	2	0	0	2	1	0
Agouti	chr24:26,366,627	3' flank	G/T	No	2	0	0	2	1	0
Agouti	chr24:26,366,778	3' flank	T/C	No	2	0	0	2	0	1
Agouti	chr24:26,366,958	3' flank	G/C	No	2	0	0	2	1	0
Mc1r	chr5:66,693,444	5' flank	C/T	No	6	0	0	59	2	0
Mc1r	chr5:66,693,226	Coding	C/T	No	3	3	0	22	29	10
Mc1r	chr5:66,693,084	Coding	A/G	S90G	6	0	0	60	1	0
Mc1r	chr5:66,693,068	Coding	C/T	T95M	6	0	0	58	3	0
Mc1r	chr5:66,693,047	Coding	C/G	A102G	1	4	1	27	22	12
Mc1r	chr5:66,693,036	Coding	T/C	No	5	1	0	50	8	3
Mc1r	chr5:66,692,869	Coding	G/A	No	5	1	0	57	4	0
Mc1r	chr5:66,692,684	Coding	G/A	R223Q	6	0	0	60	1	0
Mc1r	chr5:66,692,562	Coding	G/A	M264V	0	0	6	0	4	53

[1]Location in genome from CanFam 2.0 (*S9*); position given relative to *Agouti* or *Mc1r* transcript.

[2]SNP information given as reference sequence/variant sequence, with predicted effect, if any, on protein.

[3]No. of individuals according to genotype of Reference and Variant alleles.

Table S3. Oligonucleotide primers for *CBD103* and surrounding region[1]

Name	Size	Location	Forward	Reverse
1O	731	61826079	GGACCATTGGAGGTGTCTGT	TGGTTTGTTTTTGCTGCTGT
1I	545	61826079	GGACCATTGGAGGTGTCTGT	AGGCTTTCCCAGATCCTCAT
2O	754	61853654	GGCTCCAGAAAGTGCAGAGT	AGATGCTCTTCGTCCACACA
2I	564	61853710	CCCAGCTTCAGGAAGTAGCA	GCTCCTTTTGGCATCACTGT
3O	910	61890274	TTCTCATTTTTATGAGATAGAGAGTCA	TTCTGGTTGTGTGGTGTTCC
3I	650	61890389	GTCTGAGACCCTGGTGGTGT	TTCCTGGGTAACAGGTGAATG
4O	653	61901112	TCCTGTCTGGACACATGCAC	TGCACTTCTGTGGAAACTGC
4I	612	61901134	CCTCCCTCAAGATCCATATCC	CCCTCCAACATGATGCAACT
5O	703	61905135	GGATGCAAGAGGGTGAGATG	TGCCTTTAAAATGCCTTCCA
5I	570	61905227	AAGGGGCATTTTGACAAGTG	ACCAAGAGTTCACCGTGGAG
6O	726	61914204	TTGAGCGAGCATCACAAAAC	ACTCATCAGAGCCAGGCATT
6I	601	61914263	TGGCTAACAGCTTGACCAGA	GACTTAGGGGGCTTTGGTTC
7O	407	61948285	GGTTGCTGCTCAATGGGTAT	TCTCATGCACACACACACAAA
7I	333	61948320	CAACATGGCTGTGTGTCAGA	TCTCATTGTCCACTTGCTCCT
8O	739	61968022	CGGGAAGCTCTTCAAGGATA	GCCAGCCTCAGAGTTTGTGT
8I	560	61968143	GGGAGGGGTTAATGGTATGC	CAGGGATTGCCCTGTAAGAA

[1]Each primer pair is numbered according to its relative position (from centromere-proximal to centromere-distal) and designated as Outer or Inner. Size of the amplicon is given along with the location on chromosome 16 of the forward primer from CanFam 2.0 (*S9*).

Table S4. Polymorphisms surrounding *CBD103*[1]

Location	Reference	Variant	W poly.	D poly.	C poly.
61826211	A	G	Yes	Yes	Yes
61826328	T	C	Yes	Yes	Yes
61826329	C	T	Yes	Yes	Yes
61826373	G	A	Yes	No	No
61826511	C	T	Yes	Yes	Yes
61853853	C	T	Yes	Yes	No
61853923	T	C	Yes	Yes	Yes
61853997	T	C	No	No	Yes
61854003	C	T	Yes	No	No
61854007	G	A	No	No	Yes
61854037	C	T	Yes	No	Yes
61854114	T	A	No	No	Yes
61854131	C	A	Yes	No	Yes
61854138	A	C	Yes	No	Yes
61854157	C	T	No	No	Yes
61854170	G	A	No	No	Yes
61854203	C	A	Yes	Yes	Yes
61890602	G	A	Yes	Yes	Yes
61890615	A	G	No	Yes	No
61890676	C	T	Yes	No	Yes
61901288	T	A	Yes	No	Yes
61901326	C	T	No	Yes	Yes
61901423	G	C	Yes	No	Yes
61901539	C	T	No	No	Yes
61901650	C	T	Yes	No	No
61901689	G	A	Yes	No	No
61905249	C	A	No	Yes	No
61905310	G	C	Yes	Yes	No
61905325	T	A	No	No	Yes
61905458	A	G	Yes	No	No
61905476	C	T	No	No	Yes
61905647	C	T	Yes	Yes	Yes
61905725	C	T	Yes	Yes	Yes
61914287	C	T	Yes	Yes	Yes
61914408	G	A	No	No	Yes
61914533	G	A	Yes	Yes	Yes
61914602	C	T	Yes	Yes	Yes
61914627	C	G	Yes	Yes	Yes
61914628	T	C	Yes	Yes	Yes
61914660	G	C	Yes	Yes	Yes
61914730	A	G	Yes	No	No
61914806	G	A	No	No	Yes
61914836	T	C	No	No	Yes
61948392	A	G	Yes	Yes	Yes
61948488	A	G	No	Yes	Yes
61948582	A	G	Yes	No	No
61968262	A	T	Yes	Yes	Yes
61968342	A	G	Yes	Yes	Yes
61968379	C	T	No	No	Yes
61968477	T	C	Yes	Yes	No
61968530	T	A	Yes	No	No
61968603	C	T	Yes	Yes	Yes

[1]Location in genome from CanFam 2.0 (*S9*) together with the reference and variant sequence, and an indication of whether a polymorphism is observed in Wolves, Dogs, and Coyotes.

Table S5. Genotypes for North American wolves on distal chromosome 16

	No. of genotypes in k^y/k^y animals[1]			No. of genotypes in K^B/k^y animals[1]		
	A/A	A/B	B/B	A/A	A/B	B/B
Coordinate[2]						
59903711	8	2	0	8	1	0
59961368	1	1	8	1	1	7
59968089	2	1	7	1	0	7
60009775	8	0	0	7	2	0
60009816	2	4	3	5	0	4
60084613	2	1	5	0	4	6
60296878	2	3	4	1	3	5
60344356	4	1	4	0	3	6
60472608	2	6	2	2	6	1
60483301	1	1	8	0	2	7
60583700	2	6	1	2	6	1
60805247	2	6	2	2	5	3
60820685	6	2	2	7	3	0
61164580	2	2	6	2	2	6
61165580	2	6	2	3	4	3
61166730	1	1	8	0	4	6
61195694	1	3	5	0	2	6
61210684	1	3	5	0	5	5
61265861	2	2	6	1	5	4
61298611	0	1	8	0	1	8
61323096	0	2	8	0	5	4
61352882	0	1	9	0	2	8
61434067	1	3	5	0	4	5
61436659	0	3	7	0	1	8
61504548	2	4	4	2	7	1
61620850	2	4	4	1	9	0
61643373	1	1	7	0	3	7
61690807	0	1	9	1	6	2
61718721	1	4	4	5	4	1
61735384	2	6	2	3	4	2
61836356	6	2	2	5	2	0
62422368	0	3	7	1	3	3

[1]Data are presented as No. of homozygotes (A/A or B/B) and heterozygotes (A/B) observed among 10 k^y/k^y and 10 K^B/k^y animals for 32 SNPs on distal chromosome 16 as described in Supporting Text. None of these SNPs exhibit a distribution similar to that observed for the *CBD103* core region (Figs. 2B, 2C) in which all K^B/k^y animals are heterozygotes and all k^y/k^y animals are homozygotes.

[2]Location in genome from CanFam 2.0 (*S9*); in this assembly, the K^B mutation is located at 61,902,782-61902785.

Supplemental Figures

Fig. S1. Nucleotide diversity (π, +/- sd) as a function of distance from *CBD103*. Dataset is analogous to that presented in Fig. 2A, and represents 22 K^B and 72 k^y haplotypes from 32 Canadian wolves and 15 Yellowstone National Park founder wolves.

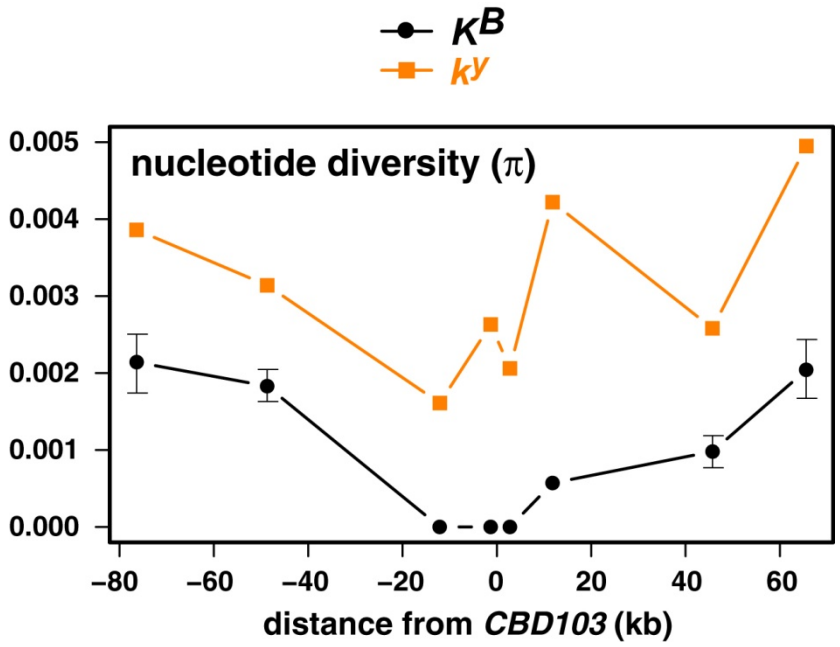

Fig. S2. Haplotypes surrounding *CBD103* in coyotes, dogs, and wolves. As in Fig. 2B (but with all canids rather than wolves alone), each row represents a K^B- or k^y-bearing chromosome, blue and yellow squares represent the major and minor alleles, respectively, gray squares represent missing data. Much greater diversity is apparent among k^y than K^B haplotypes.

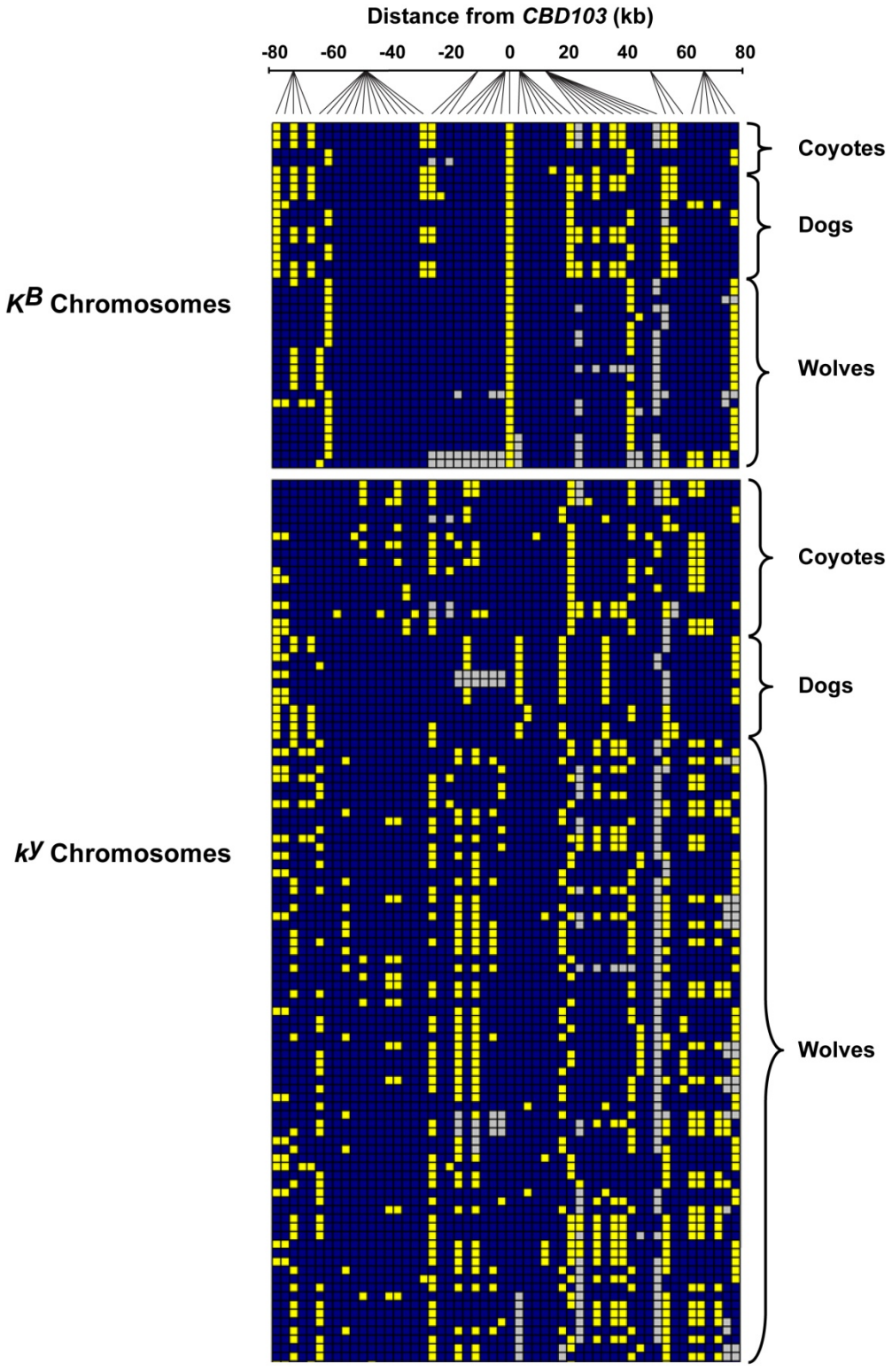
Distance from CBD103 (kb)
-80
-60
-40
-20
0
20
40
60
80
Coyotes
Dogs
Wolves
K^B Chromosomes
Coyotes
Dogs
Wolves
k^y Chromosomes

Fig. S3. Principal component analysis of wolf and dog genotypes. Genotypes for 20 wolves (5 K^B/k^y and 5 k^y/k^y from Canada, and 5 K^B/k^y and 5 k^y/k^y from Yellowstone) were analyzed by principal component analysis based on 3 progressively larger genomic regions; color (K^B/k^y vs. 5 k^y/k^y) and origin (Canada vs. Yellowstone) were then tested separately as distinguishing factors for each of the 5 first principal components. (A) p values, displayed on the ordinate as the negative logarithm, based on ANOVA; horizontal line corresponds to p = 0.05. (B) Scatterplot of individual wolves according to values for PC1 and PC2. (C) A second principal component analysis carried out on genotypes from the 20 wolves together with 20 dogs of different breeds and 5 wolf-dog hybrids from Italy; scatterplot of individuals according to values for PC1 and PC2.

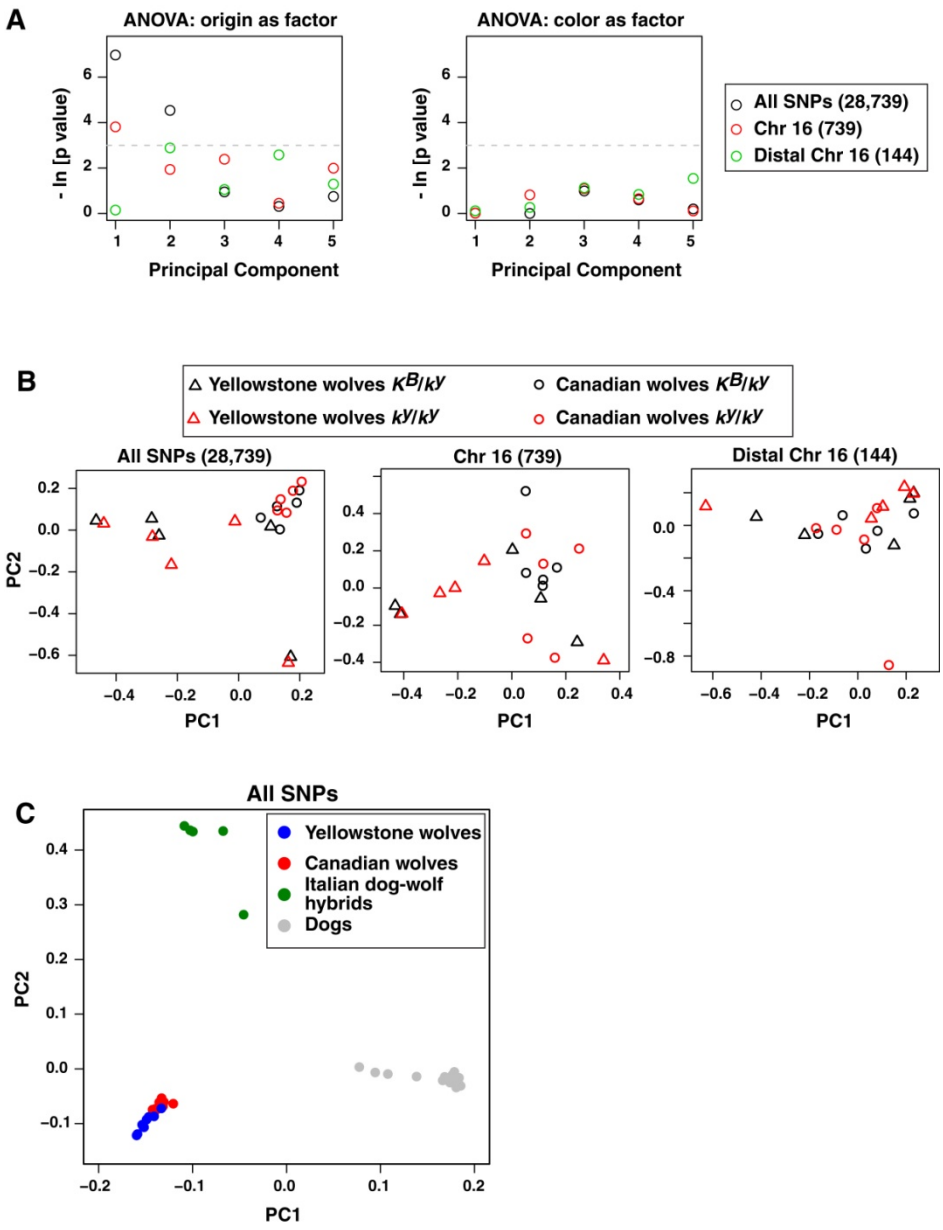
A
ANOVA: origin as factor
ANOVA: color as factor
- ln [p value)
Principal Component
All SNPs (28,739)
Chr 16 (739)
Distal Chr 16 (144)
B
Yellowstone wolves KB/kY
Yellowstone wolves kY/kY
Canadian wolves KB/kY
Canadian wolves kY/kY
All SNPs (28,739)
Chr 16 (739)
Distal Chr 16 (144)
PC1
PC2
C
All SNPs
Yellowstone wolves
Canadian wolves
Italian dog-wolf hybrids
Dogs
PC1
PC2

Fig. S4. Haplotype network for mitochondrial DNA control region from Canadian wolves and dogs. Wolf haplotypes are from (*S1*); dog haplotypes are from sequences representing the five major dog clades (Genbank AF531664, AF531729, AF531720, AF531738, and AF531741) (*S22*). Each polygon represents a haplotype whose area is proportional to its frequency; small circles are inferred haplotypes that represent individual "steps" connecting the network. The network was constructed from 37 haplotypes (14 black wolves, 18 non-black wolves, 5 dogs) with the number of black wolves indicated for each haplotype. Wolves carrying the K^B allele are not more closely related to each other than they are to wolves carrying only the k^y allele.