Supplemental Note A: Phasing, Haplotype Analysis, and Tree Reconstruction

We describe the method of phasing SNP genotypes and defining windows for haplotype analyses. From phased data, we reconstructed neighbor-joining trees for analysis of dog domestication.

Haplotype diversity and sharing analysis: We inferred haplotype phase using the program `fastPHASE` version 1.4.0[1] for both datasets. All dogs were phased together in a single analysis, but we designated breeds as different subpopulations. This procedure was shown to yield optimal results when phasing human data[2]. We specified the number of haplotype clusters (K) to be equal to 40. Through preliminary analyses using subsets of the data, we found that the genotype imputation error rate (estimated from masking and imputing known genotypes) continues to decrease as K increases (up to K = 100), albeit, quite slowly. This suggests that higher values of K may yield more accurate results. However, since the practical advantages of using higher values of K were marginal, we assessed the sensitivity of the number of haplotypes per breed to the value of K used. We found that the value of K had little impact on the overall results, and thus chose K = 40 as a compromise between the true number of "haplotype clusters" in the sample and computational efficiency. We included 44,156 SNPs in the phased haplotypes that had MAF ≥1% and <10% missing data in 912 dog samples.

We divided the genome into 500kb windows to be used for the haplotype analyses. Since the number of SNPs within each window is a complex function of the mutation rate, genetic drift, and the ascertainment process, and the number of SNPs within a window can influence haplotype diversity, we fixed the number of SNPs within a window. Specifically, we divided the genome into 500kb windows and from those windows with ≥15 SNPs, we selected a random subset of 15 SNPs. Similarly, for windows with <15 SNPs, but at least 5 SNPs, we selected 5 SNPs at random. Windows with fewer than 5 SNPs were excluded from the analysis. The same randomly-selected SNPs were used for all individuals. Since the number of haplotypes is influenced by the sample size, we selected a random subset of nine dogs from each group for analysis. Using this approach, 2,634 windows of 500kb were defined that contained 5 SNPs and 944 windows that contained 15 SNPs.

Haplotype diversity: We chose to summarize haplotype diversity within each group as the number of distinct haplotypes within each window across the genome. We chose this statistic because it has been shown to be informative about population history through simulations and empirical analyses[3,4]. For this analysis, we only included breeds with $n_{individuals} \geq 6$, and took a random sample of 6 individuals if there were more dogs per breed. We counted the number of distinct haplotypes within each breed for each window using the inferred haplotypes from `fastPHASE`[1].

doi: 10.1038/nature08837

Haplotype Sharing: Using the defined haplotype windows, we calculated the number of total unique haplotypes and the proportion of sharing these haplotypes for each dog breed and wolf population (Middle East, Europe, and China). Each of these wolf populations has been suggested as a potential ancestral population[5-9]. We also tabulated sharing with North American wolves, as they have not been considered directly ancestral to dogs[10]. This analysis focused on well-sampled breeds ($n_{\text{individuals}} \geq 9$ per breed, $n_{\text{breeds}} = 64$). For breeds with more than 9 individuals, we used a random subset of 9 individuals. Specifically, we tabulated the number of haplotypes within a dog breed that were present in only one of the four wolf populations. Specifically, let ME_i denote the number of haplotypes present in the dog breed and Middle Eastern wolves (and absent from China, Europe, and North America) at window *i*; CN_i denote the number of haplotypes present in the dog breed and Chinese wolves (and absent from Middle East, Europe, and North America) at window *i*; NA_i denote the number of haplotypes present in the dog breed and North American wolves (and absent from Middle East, Europe, and China) at window *i*; and EA_i denote the number of haplotypes present in the dog breed and European wolves (and absent from Middle East, North America, and China) at window *i*. Let p_{ME} denote the proportion of haplotypes across the genome present in the dog breed and Middle Eastern wolves (to the exclusion of the other wolf populations). Then $p_{ME} = \frac{\sum_{\text{all } i} ME_i}{\sum_{\text{all } i} ME_i + NA_i + CN_i + EA_i}$. The other proportions (p_{CN}, p_{NA}, p_{EA}) can be found in a similar manner.

We also preformed two permutation tests using the haplotype windows. The first test determined whether for a given dog breed, significantly more haplotypes are shared with Middle Eastern or Chinese wolves. Essentially, this is a two-sided test testing the hypothesis $p_{CN} = p_{ME}$ vs. $p_{CN} \neq p_{ME}$. The second test assessed whether any one of four wolf populations had excess haplotype-sharing with a dog breed if haplotypes were equally represented among all wolf populations. This tests whether $\max(p_{CN}, p_{NA}, p_{EA}, p_{ME})$ is larger than expected. Test 1 only compares Chinese and Middle Eastern wolves to dogs and significant results for test 1 do not indicate if the proportion of European haplotypes is larger than expected. Permutation test 2 determines if haplotype sharing is larger than expected and includes all wolf populations.

In our permutation strategy, we randomly assigned each of 36 wolves to one of four arbitrary groups, keeping dog assignments fixed within their breed. For each permutation, we then calculated p_1, p_2, p_3, p_4 the same way we calculated p_{ME} in the observed data. Note, p_1 is simply the proportion of haplotypes in the dog breed that are present in the first group of permuted wolves, but absent from groups 1-3. We then record $p_1 - p_2$ and $\max(p_1, p_2, p_3, p_4)$. The p-value for test 1 is calculated as the proportion of permutations where $|p_{ME} - p_{CN}| > |p_1 - p_2|$. The p-value for test 2 is calculated as the proportion of permutations where $\max(p_{CN}, p_{NA}, p_{EA}, p_{ME}) > \max(p_1, p_2, p_3, p_4)$. We analyzed the 5 and 15-SNP windows separately and conducted 1,000 permutations for each.

Test 1 shows that for 6.3% (4/64) and 27% (17/64) of breeds (using 5 and >15-SNP windows, respectively), the proportion of haplotypes shared with Middle Eastern and Chinese wolves was significantly different. In all of these cases, there was more sharing with Middle Eastern than Chinese wolves.

Additionally, to include a larger number of breeds in the haplotype sharing analysis, we also performed the permutation tests using breeds with at least six individuals (Supplemental Fig. 14). For breeds with more than six individuals, we took a random sample of six individuals. The testing follows as stated above. Overall, as before, we see the most significant p-values for sharing with ME wolves for both test 1 and 2. Some of the Asian breeds (such as Dingo and Chow-chow in the 15-SNP windows) show the highest sharing with CN wolves and the proportion shared with CN wolves is significantly higher than expected based on test 2. Test 1 also suggests that there is significantly more sharing with CN wolves than with ME wolves for both of these breeds. There is still the highest sharing with CN wolves for these breeds in the windows with 5 SNPs, however the results are not significant. There are fewer breeds sharing the most haplotypes with EA wolves, all being non-significant. We also find more breeds that significantly share unique haplotypes with ME wolves, compared to the analysis using nine individuals per breed shown in Figure 2. The difference is likely due to using a different and smaller sample of dogs and wolves (six individuals here as compared to nine individuals in Figure 2), resulting in a loss of subtle signatures. The sample of six individuals used here may not contain as many haplotypes shared between EA wolves and certain dog breeds, as did the samples of nine individuals.

Overall, the results from permutation tests 1 (described above) and 2 (Figure 2d, Supplemental Table 3; Supplemental Fig. 14) suggest multiple wolf populations contributed to the genome of dogs due to the fact that, for certain breeds, we find significant levels of haplotype sharing with multiple wolf populations (e.g. Middle Eastern and European wolves). This result is similar to histories of other domestic species[11,12]. However, we find the greatest fraction of significant results using test 2 (100% for 5-SNP windows and 75% for 15-SNP windows, both from the analysis using 9 individuals per breed), which supports the notion that Middle Eastern wolves have uniformly contributed a greater proportion of ancestry to dogs than other wolf populations (Supplemental Table 3).

In summary, the Middle East is supported as the mostly likely center for dog origination, although the heterogeneity in haplotype sharing suggests multiple wolf populations have contributed to the dog genome early in the history of dog domestication. Moreover, European wolves may have been a greater contributor to haplotype diversity in dogs than wolves from East Asia. Finally, the Dingo, Chinese Shar-Pei, Chow-chow and Basenji may represent the extant breeds retaining the most genetic similarity to ancestral wolf populations with the former three breeds derived from wolves that inhabited East Asia and the latter, the Middle East.

Tree reconstruction: For tree reconstruction, we used two datasets: 574 dogs and Old World wolves; and 530 dogs and Old World wolves. The 574 dataset consisted of six individuals from 75 dog breeds where six or more individuals were typed, and five breeds with less than six individuals typed, for a total of 490 dogs. From the available sample set of Old World wolves, we removed all identified dog-wolf hybrids (n = 40 as described in Supplemental Note B). We also removed 13 closely related individuals from six populations identified by IBS analysis: Israel (n = 4); India (n = 2); Saudi Arabia (n = 3); Iran (n = 1); Oman (n = 2); and Sweden (n = 1). In total, 84 Old World wolves from China, Central Asia, the Middle East, and Europe, including the Italian and Spanish population samples, were used. The dataset included one coyote from California for rooting purposes. The 530 dataset was created for the population-level and haplotype-sharing distance-based analyses and used a subset of 530 dogs and Old World wolves. This dataset was chosen to provide near equal numbers of individuals from each breed or population and consisted of 79 dog breeds with six individuals each and Chinese (n = 6), Middle Eastern (n = 7), Central Asian (n = 6), Italian (n = 6), Spanish (n = 7) and other European wolves (n = 18; n_{total} = 50). Six coyotes from California, Washington state and Alaska were used for rooting purposes.

We generated neighbor-joining (NJ) trees based on allele-sharing distances among the subset of 574 representative canids using the pruned 43,954 SNPs and haplotype data partitioned into 5-SNP and 10-SNP haplotype windows (see below). The allele-sharing distance used was one minus the proportion of alleles shared, as calculated using the program `microsat` (denoted as 1-p(s) in `microsat`)[13]. For computing allele sharing using haplotypes, each haplotype window was considered as a locus and each unique haplotype within the window was considered as a unique allele. One thousand bootstrap replicates were generated using `microsat`. Note that the bootstrapping resamples over SNP loci, and thus only represents the sampling variance associated with sampling a finite number of loci.

The resulting pairwise matrices of allele sharing distance were input to `Neighbor` from the `PHYLIP` package and then consensus trees were generated using the majority rule option in the program `consense` from the `PHYLIP` package[14]. The resulting trees were visualized using `Dendroscope`[15]. For population-level analyses, an identical procedure of running `microsat`, `neighbor`, and `consense` was followed, where allele-sharing distances were instead calculated between populations.

For population-level analyses, breeds and wild canid populations containing fewer than six individuals were excluded, and the remaining populations were subsampled to obtain six individuals each. The resulting set contains 78 breeds (see Supplemental Table 1 for a list of breeds) and seven wild canid populations consistently defined geographically with `STRUCTURE` analyses (Eastern and Northern Europe, Spain, Italy, China, Central Asia, and Middle East wolves, plus the coyote; see Supplemental Methods; results not shown).

To prepare the haplotype data, we ran `fastPHASE` version 1.4.0[1] with 40 specified haplotype clusters (K; see above). Because we were concerned with maximizing the number of informative loci rather than comparative estimates of haplotype diversity (see above), we used windows of 5 and 10 contiguous SNPs rather than a region of defined size. This greatly increased the number of loci and resulted in 9,576 and 4,788 loci for the 5 and 10-SNP windows, respectively. For haplotype analyses, breeds and wild canid populations containing fewer than six individuals were excluded, and the remaining populations were subsampled randomly to obtain six individuals each. The resulting set contains 78 breeds (see Supplemental Table 1 for a list of breeds) and seven wild canid populations (See Supplementary Figs. 6-11).

Literature Cited

1. Scheet, P & Stephens, M. A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypeic Phase. *Am. J. Hum. Genet.* **78**, 629-644 (2006).

2. Conrad, D, Jakobsson, M, Coop, G, Wen, X, Rosenberg, N, et al. A worldwide survey of haplotypes variation and linkage disequilibrium in the human genome. *Nat. Genet.* **38** (11), 1251-1260 (2006).

3. Auton, A, Bryc, K, Boyko, A, Lohmueller, K, Novembre, J, et al. Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res.* **19** (5), 795-803 (2009).

4. Lohmueller, K, Bustamante, C & Clark, A. Methods for human demographic inference using haplotype patterns from genomewide single-nucleotide polymorphism data. *Genetics* **182** (1), 217-231 (2009).

5. Savolainen, P, Zhang, YP, Luo, J, Lundeberg, J, & Leitner, T. Genetic Evidence for an East Asian Origin of Domestic Dogs. *Science* **298** (5598), 1610-1613 (2002).

6. Sablin, M & Khlopachev, G. The Earliest Ice Age Dogs: Evidence from Eliseevichi I. *Curr. Anthropol.* **45** (5), 795-819 (2002).

7. Germonpre, M, Sablin, M, Rhiannon, E, Hedges R, Hofreiter, M, et al. Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes. *J. Archaeol. Sci.* **36**, 473-490 (2009).

8. Dayan T. Early Domesticated Dogs of the Near East. *J. Archaeol. Sci.* **21**, 633-640 (1999).

9. Zeder, M, Emshwiller, E, Smith, B & Bradley, D. Documenting domestication: the intersection of genetics and archaeology. *Trends Genet.* **22** (3), 139-155.

10. Leonard, J, Wayne, R, Wheeler, J, Valadez, R, Guillén, S, et al. Ancient DNA Evidence for Old World Origin of New World Dogs. *Science* **298** (5598), 1613-1616 (22 November 2002).

11. Bruford, M, Bradley, G & Luikart, G. DNA markers reveal the complexity of livestock domestication. *Nat. Rev. Genet.* **4** (11) 900-910 (2003).

12. Larson, G, Dobney, K, Albarella, U, Fang, M, Matisoo-Smith, E, et al. Worldwide Phylogeography of Wild Boar Reveals Multiple Centers of Pig Domestication. *Science* **307** (5715), 1618-1621 (2005).

13. Minch, E, Ruiz-Linares, A, Goldstein, D, Feldman, M, Cavalli-Sfroza, L, et al. MICROSAT: A Computer Program for Calculating Various Statistics on Microsatellite Allele Data. Palo Alto, CA: Stanford University (1996).

14. Felsenstein, J. PHYLIP (Phylogency Inference Package) version 3.5c. Seattle, WA: Department of Genetics, University of Washington (1993).

15. Huson, D, Richter, D, Rausch, C, Dezulian, T, Franz, M, et al. Dendroscrope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* **8**, 460-465 (2007).

Supplemental Note B: Principal Component Analysis

We describe the methodology in detail for principal component analysis (PCA) of domestic dogs and gray wolves. We also identify SNP loci diagnostic for dogs and wolves and support our assertion in the text that modern dog breeds and gray wolves are genetically distinct and only rarely admixed.

Principal Component Analysis: We used the `smartpca` program distributed in the `Eigensoft` package[1]. We initially explored the effect of various sample sizes of wild and domestic canids and found that when all dog samples are included (n = 912), PC1 is primarily a dog-wolf axis and PC2 is dominated by a contrast between mastiff-like breeds (including Boxer) and all other canids (data not shown). To reduce the impact of the large numbers of dogs relative to wolves (which leads the PCA to resolve dog diversity), we reduced the sample size of dogs to two individuals per breed for our principal component analyses (Supplemental Figs. 1 and 2). The reduction in the domestic dog sample permits resolution of the early ancestry of domestic dogs rather than partitioning individual breeds. We included only Old World wolf populations because they alone are hypothesized as direct ancestors of domestic dogs and we included individuals having pairwise genetic similarity[2] below the threshold IBS < 0.8 (see Supplemental Materials). The wolf populations included China (n = 9), Central Asia (n = 3), the Middle East (n = 7), and Europe (n = 43). We excluded wolves from highly inbred populations (Italy, Spain, Sweden)[3] to avoid their influence in the cluster analysis. We also excluded putative dog-wolf hybrids from the wild wolf population (n = 40) identified with the `Eigensoft` package[1]. We performed PCA for SNPs discovered from different ascertainment panels (see Supplemental Fig. 3 for top 5 components).

The first principal component (PC1; 11% of variation) is predominantly a wolf-dog axis with modern breeds having low values, ancient breeds intermediate, and gray wolves demonstrating high values and tight clustering (Supplemental Fig. 1). Dingoes and New Guinea Singing dogs are among the oldest known dog populations and are closest to wolves on PC1[4], followed by breeds such as Chow-chow, Basenji, Akita, Chinese Shar-Pei, Siberian Husky and Alaskan Malamute. Other axes primarily distinguish individual breeds and further identify the Basenji as divergent (Supplemental Figs. 2 and 3).

To identify a set of SNPs for distinguishing between dogs and wolves (Supplemental Fig. 4), we ranked the SNPs on PC1 in order of decreasing magnitude of SNP weights. The top 20 SNPs with the highest loadings on PC1 were used for an additional `STRUCTURE`[5] analysis using all dog and wolf samples at K=2 (2,000 burn-in iterations and 5,000 MCMC iterations, for three repetitions; see Supplemental Methods) to obtain the joint probability of species assignments for dogs and wolves. This `STRUCTURE` analysis identifies all sampled modern dogs and wolves correctly and with high confidence (K=2, assignment probabilities >0.999). With the exception of a few ancient breeds, these results show that although backcrossing between dogs and wolves is known to occur[6], extensive admixture in the modern dog genome is not evident. Further, because the breeds showing evidence of admixture are commonly thought to have diverged early from all other dogs during the history of domestication, their genetic similarity to wolves may reflect admixture in the first stages of domestication.

Literature Cited

1. Price, A, Patterson, N, Plenge, R, Weinblatt, M, Shadick, N, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38** (8), 904-909 (2006).

2. Purcell, S, Neale, B, Todd-Brown, K, Thomas, L, Ferreira, M, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics* **81** (3), 559-575 (2007).

3. Gray, M, Granka, J, Bustamante, C, Sutter, N, Boyko, A, et al. Linkage Disequilibrium and Demographic History of Wild and Domestic Canids. *Genetics* **181**, 1493-1505 (2009).

4. Savolainen, P, Leitner, T, Wilton A, Matisoo-Smith, E & Lundeberg, J. A detailed picture of the origin of the Australian dingo, obtained from the study of mitochondrial DNA. *P. Natl. Acad. Sci. USA* **101** (33), 12387-12390 (2004).

5. Pritchard, J, Stephens, M & Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155** (2), 945-959 (2000).

6. Vilà C, Seddon J & Ellegren H. Genes of domestic mammals augmented by backcrossing with wild ancestors. *Trends Genet.* **21** (4), 214-218 (2005).

Supplemental Note C: Detecting candidate loci positively selected during domestication

We describe the methodology in detail for identifying genome regions and candidate genes under positive selection early during dog domestication.

Detecting Positive Selection: To identify loci that may have undergone adaptive evolution during early dog domestication, we focused on comparing patterns of differentiation between gray wolves (n = 92, excluding related wolves and potential hybrids) and dogs from all modern breeds (n = 701). We purposefully exclude ancient breeds because of observed admixture with wolves that might dilute any signatures of differentiation. We computed the summary statistics F_{ST} and cross population extended haplotype homozygosity (XP-EHH) using a subset of 43,452 autosomal SNPs. The SNP set represents less than the complete set because we focused on autosomal SNPs and excluded SNPs that did not have assigned ascertainment panels or had complex ascertainment schemes (see below).

The degree of population differentiation between wolves and modern dogs was measured at each SNP by F_{ST}[1] using scripts written by J. Novembre. J. Pickrell[2] kindly provided the script to compute XP-EHH. The default parameters of the script were modified to allow for a larger spacing between SNPs (1Mb as the threshold gap between SNPs when computing the intermediate statistic, EHH, and 4Mb as the threshold gap between SNPs when searching for the stop position for integration when computing the intermediate statistic, iHH).

To account for the variable ascertainment strategies used, we normalized F_{ST} and XP-EHH to have a mean of zero and standard deviation of 1 within ascertainment categories. We then computed the empirical percentile of each SNP for the normalized F_{ST} and XP-EHH values associated with each SNP. We used the product of the F_{ST} percentile and XP-EHH percentile to obtain a single percentile summarizing the strength of the two signatures (the “bi-variate percentile score”). To rank gene regions with regards to evidence for selection, we collapsed multiple extreme SNPs in a region into "clusters". Specifically, if two or more SNPs were in the 95th percentile of the bi-variate percentile score and were spaced less than 300kb apart, they were joined into a single cluster. We then ranked clusters by the number of SNPs they contain, and for all clusters with the same number of SNPs, we sorted them by the bi-variate percentile score of the central SNP.

We emphasize that for various reasons these outlier regions should only be considered candidates for having undergone adaptive evolution (e.g. background selection can lead to enhanced levels of differentiation at regions under purifying rather than positive selection[3]; the general limitations of outlier approaches[4]; and the potential for genotyping artifacts due to CNV regions[5,6]). For confirmation, we suggest these regions are worth characterizing using re-sequencing-based and functional approaches. In addition, besides concerns regarding potential false positives, many sites involved in adaptive evolution in dogs have likely gone undetected. Specifically, following a complete sweep, the region around the beneficial substitution is expected to transiently show low heterozygosity and an excess of low-frequency alleles[7]. These loci will preferentially go undetected by the ascertainment scheme used for the canine SNP array. We speculate there are additional sweep regions to be found with denser SNP panels and/or re-sequencing.

Test for genic region enrichment in F_{ST} and XP-EHH outliers: The ENSEMBL Perl API was used to query the genomic context of each SNP. The SNP was defined as "genic" if a portion of any gene was found within a fixed length of the SNP. Otherwise, the SNP was defined as "non-genic." Different fixed lengths were tested, ranging from 10kb to 60kb. We used the empirical distribution to identify SNPs with extreme patterns of differentiation ("empirical outliers"). We considered three definitions of empirical outliers: 1) SNPs having extreme values of F_{ST}; 2) SNPs having an extreme value of XP-EHH consistent with a selective sweep on the dog lineage (i.e., SNPs with strongly positive values of XP-EHH); and 3) SNPs with extreme values of both F_{ST} and XP-EHH. A one-sided conditional exact test[8] was performed to test whether the genic SNPs were enriched in outliers conditional on the ascertainment bias panel. Different thresholds for defining empirical outliers were tested and significant results ($p<0.05$) were found for the 1% tail of F_{ST} over a range of values (10kb - 40kb), and the 5% tail of XP-EHH at the 10kb and 20kb scales.

Discussion of best hits. The highest ranked signal falls in a region containing an unknown gene in EntrezGene with high amino acid sequence similarity to a neurotrimin gene (NTM, OMIM: 607938) and an opioid receptor (OPCML, OMIM: 600632; Supplemental Fig. 17). The former is a cell adhesion gene involved in neurite formation and the latter binds opioid alkaloids in the presence of acidic lipids, is generally highly conserved, and is important in stress response. The next strongest signals (Supplemental Fig. 17a-c) are near ryanodine receptor 3 (RYR3, OMIM: 180903), associated with acquired memory, and adenylate cyclase 8 (ADCY8, OMIM: 103070)[9,10], which is implicated in sensitization to pain in mice and memory formation in humans. Our fourth and fifth strongest signatures are near a cluster of interleukin family 1 genes (Supplemental Fig. 17d) and a region containing two genes from the carnosinase dipeptidase family (CNDP1, OMIM 609064 and CNDP2, OMIM 169800; Supplemental Fig. 18). CNDP1 is a neurotransmitter expressed in the brain, which degrades carnosine, a dipeptide primarily found in muscle tissue, while CNDP2 is a non-specific peptidase expressed predominantly in the kidney and liver.

Examining F_{ST} alone, we found 12 consecutive SNPs in the top 5^{th} percentile for normalized F_{ST} values located at the SLC24A4 gene, a gene whose polymorphisms in humans are associated with hair and eye color[11]. We observe a single SNP with a high F_{ST} value in the WBSCR17 gene. The deletion of this gene and neighboring genes gives rise to Williams-Beuren syndrome in humans (OMIM: 194050; Supplemental Fig. 16). While outliers of the genome are not necessarily the result of adaptive evolution, we propose the gene regions mentioned above are interesting candidates for loci involved in the phenotypic evolution of dogs from their wolf ancestors.

Literature Cited

1. Weir, B & Cockerham, C. Estimating F-statistics for the analysis of population structure. *Evolution* **38** (6), 1358-1370 (1984).

2. Pickrell, J, Coop, G, Novembre, J, Kudaravalli, S, Li, J, et al. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* **19** (5), 826-837 (2009).

3. Charlesworth, B, Nordborg, M & Charlesworth, D. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* **70** (2), 155-174 (1997).

4. Teshima, K, Coop, G & Przeworski, M. How reliable are empirical genomic scans for selective sweeps? *Genome Res.* **16** (6), 702-712 (2006).

5. Nicholas, T, Cheng, Z, Ventura, M, Mealey, K, Eichler E, et al. The genomic architecture of segmental duplications and associated copy number variants in dogs. *Genome Res.* **19**, 491-499 (2009).

6. Chen, W, Swartz, J, Rush, A & Alvarez, C. Mapping DNA structural variation in dogs. *Genome Res.* **19**, 500-509 (2009).

7. Nielsen, R, Hellmann, I, Hubisz, M, Bustamante, C & Clark, A. Recent and ongoing selection in the human genome. *Na.t Rev. Genet.* **8** (11), 857-68 (2007).

8. Agresti, A. Dealing with discreteness: making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact. *Stat. Methods Med. Res.* **12**, 3 (2003).

9. Wei, F, Qiu, C, Kim, S, Muglia, L, Maas, J, et al. Genetic elimination of behavioral sensitization in mice lacking calmodulin-stimulated adenylyl cyclases. *NEURON* **36**, 713-716 (2002).

10. de Quervain, D & Papassotiropoulos, A. Identification of a genetic cluster influencing memory performance and hippocampal activity in humans. *P. Natl. Acad. Sci. USA* **103**, 4270-4274.

11. Sulem, R, Gudbjartsson, D, Stacey, S, Helgason, A, Rafnar, T, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* **39** (12), 1443-1452 (December 2007).

Supplemental Table 1. Wild and domestic canids genotyped on the dog genome-wide SNP array. Breeds are grouped according to geographic location (* as defined by reference [3]), or as modern and ancient breeds[4], and phenotypic/functional groups[8,9]. (Geographic abbreviations: East Asia = E Asia; North America = N America ; Southeast Asia = SE Asia; Southwest Asia = SW Asia).

Species	Common Name (reference number for Fig. 1)	Geographic Origin[3,8,9]	Parker Group[4]	Dog breed group[8,9]	Sample size
Canis familiaris	Afghan Hound	SW Asia*	Ancient-Asia	Ancient-Spitz	12
	Africanis	Africa	--	--	3
	Akita	E Asia*	Ancient-Asia	Ancient-Spitz	12
	Alaskan Malamute	N America	Ancient-Asia	Ancient-Spitz	11
	American Cocker Spaniel	N America	Hunting	Spaniel	12
	American Eskimo	N America*	--	Ancient-Spitz	7
	Australian Shepherd	N America	Herding-Sight hound	Herding	12
	Australian Terrier	Europe	Herding-Sight hound	Small terriers	12
	Basenji	Africa*	Ancient-Asia	Ancient-Spitz	13
	Basset Hound	Europe	Hunting	Scent hound	11
	Beagle	Europe*	Hunting	Scent hound	10
	Bernese Mountain Dog	Europe	Mountain	Mastiff-like	11
	Bloodhound	Europe	Hunting	Scent hound	9
	Border Collie	Europe*	Herding-Sight hound	Herding	12
	Borzoi	Europe	Herding-Sight hound	Sight hound	12
	Boston Terrier	America	Mastiff-Terrier	Mastiff-like	6
	Boxer	Europe*	Mastiff-Terrier	Mastiff-like	12
	Briard (14)	Europe	--	Mastiff-like	12
	Brittany Spaniel	Europe	Hunting	Spaniel	12
	Brussels Griffon (1)	Europe	Hunting	Toy	7
	Bullmastiff	Europe	Mastiff-Terrier	Mastiff-like	12
	Bull Terrier	Europe	Mastiff-Terrier	Mastiff-like	3
	Bulldog	Europe	Mastiff-Terrier	Mastiff-like	11
	Cairn Terrier	Europe	Hunting	Small terriers	12
	Canaan Dog	Middle East	--	--	3
	Cardigan Welsh Corgi	Europe	Herding-Sight hound	Herding	12
	Cavalier King Charles Spaniel	Europe*	Hunting	Spaniel	12
	Chihuahua (9)	N America	Hunting	Toy	9
	Chinese Shar-Pei	E Asia*	Ancient-Asia	Ancient-Spitz	12
	Chow-chow	E Asia*	Ancient-Asia	Ancient-Spitz	11
	Collie	Europe*	Herding-Sight hound	Herding	12
	Dachshund (16)	Europe*	Hunting	Scent hound	12
	Dingo	SE Asia	--	Ancient-Spitz	12
	Doberman Pinscher (6)	Europe*	Hunting	Working dog	6
	English Cocker Spaniel	Europe	Mountain	Spaniel	12
	English Springer Spaniel	Europe	--	Spaniel	6
	Flat-coated Retriever	Europe*	Hunting	Retriever	12
	French Bulldog	Europe	Mastiff-Terrier	Mastiff-like	12
	German Shepherd Dog (13)	Europe*	Mountain	Working dog	12
	German Short-haired Pointer	Europe*	Hunting	Spaniel	12
	Giant Schnauzer (17)	Europe	Hunting	Working dog	11
	Glen of Imaal Terrier (12)	Europe	Mastiff-Terrier	Mastiff-like	12
	Golden Retriever	Europe*	Hunting	Retriever	12
	Great Dane	Europe	Herding-Sight hound	Mastiff-like	12
	Greyhound	SW Asia	Herding-Sight hound	Sight hound	12
	Havanese	Europe	--	Working dog	12
	Ibizan Hound (8)	Europe	Hunting	Ancient-Spitz	11
	Irish Water Spaniel	Europe	Hunting	Spaniel	11
	Irish Wolfhound	Europe*	Herding-Sight hound	Sight hound	12
	Italian Greyhound	Europe	--	Sight hound	13
	Jack Russell Terrier (15)	Europe	Hunting	Small terriers	12
	Kuvasz (7)	Europe*	Herding-Sight hound	Pastoral	12
	Labrador Retriever	Europe*	Mastiff-Terrier	Retriever	12
	Mastiff	Europe	Mastiff-Terrier	Mastiff-like	12
	Miniature Bull Terrier	Europe	Mastiff-Terrier	Mastiff-like	12
	Miniature Pinscher (5)	Europe	--	Toy	12
	New Guinea Singing Dog	SE Asia	--	Ancient-Spitz	12
	Newfoundland	N America	Mastiff-Terrier	Retriever	3
	Norwich Terrier	Europe	Mastiff-Terrier	Small terriers	12
	Old English Sheepdog	Europe	Herding-Sight hound	Herding	10
	Papillion (11)	Europe*	--	Toy	12
	Pekingese (2)	E Asia*	Hunting	Toy	12
	Pembroke Welsh Corgi	Europe	Mastiff-Terrier	Herding	11
	Petit Basset Griffon Vendeen (PBGV)	Europe	Hunting	Scent hound	12
	Pomeranian (10)	Europe	Hunting	Toy	12
	Portuguese Water Dog	Europe	Hunting	Working dog	12

	Pug (3)	Europe*	Hunting	Toy	12
	Rhodesian Ridgeback	Africa*	Hunting	Scent hound	12
	Rottweiler	Europe*	--	Mastiff-like	3
	Saint Bernard	Europe*	Mountain	Mastiff-like	12
	Saluki	Middle East	Ancient-Asia	Ancient-Spitz	12
	Samoyed	Siberia*	Ancient-Asia	Ancient-Spitz	12
	Scottish Deerhound	Europe*	Herding-Sight hound	Sight hound	6
	Scottish Terrier	Europe	Hunting	Small terriers	12
	Shetland Sheepdog	Europe	Herding-Sight hound	Herding	12
	Shih Tzu (4)	E Asia	Hunting	Toy	10
	Siberian Husky	Siberia	Ancient-Asia	Ancient-Spitz	12
	Staffordshire Bull Terrier	Europe*	Mastiff-Terrier	Mastiff-like	12
	Standard Poodle	Europe	--	Working dog	12
	Standard Schnauzer (18)	Europe	Hunting	Working dog	12
	Sussex Spaniel	Europe*	--	Spaniel	5
	Toy Poodle	Europe*	Hunting	Working dog	12
	West Highland White Terrier	Europe*	Hunting	Small terriers	12
	Whippet	Europe*	Herding-Sight hound	Sight hound	12
	Yorkshire Terrier	Europe	--	Small terriers	8
Dog-wolf hybrid	Dog-wolf hybrid; Europe	--	Hybrid	--	17
Canis aureus	Golden Jackal	--	--	--	2
Canis mesomelas	Black-backed Jackal	--	--	--	6
Canis adustus	Side-striped Jackal	--	--	--	1
Canis simensis	Ethiopian wolf	--	--	--	4
Canis rufus	Red wolf	--	--	--	12
Canis latrans	Coyote	--	Coyote	--	60
Canis lupus	Gray wolf, North America	--	N America	--	62
	Gray wolf, Great Lakes	--	--	--	22
	Gray wolf, Europe	--	Europe	--	87
	Gray wolf, India	--	Central Asia	--	3
	Gray wolf, Iran	--	Central Asia	--	2
	Gray wolf, Israel	--	Middle East	--	8
	Gray wolf, Oman	--	Middle East	--	3
	Gray wolf, Saudi Arabia	--	Middle East	--	5
	Gray wolf, Turkey	--	Central Asia	--	1
	Gray wolf, Middle East	--	Middle East	--	22
	Gray wolf, China	--	China	--	10
	Gray wolf, Mexican	--	--	--	10

Supplemental Table 2. Eigenvalues and allele frequencies of the 20 SNPs with the highest magnitude loadings on PC1 (Supplemental Fig. 1).

SNP Ranking	Magnitude of PC1 SNP Loadings	Dog (n = 914)		Gray wolf (n = 155)	
		f(A)	f(B)	f(A)	f(B)
1	3.320	0.0195	0.9805	0.8137	0.1863
2	3.251	0.0556	0.9444	0.9247	0.0753
3	3.205	0.0868	0.9132	0.9771	0.0229
4	3.179	0.0590	0.9410	0.9567	0.0433
5	3.154	0.0829	0.9171	0.9933	0.0067
6	3.151	0.0690	0.9310	0.9757	0.0243
7	3.138	0.0833	0.9167	0.9500	0.0500
8	3.125	0.1213	0.8787	0.8517	0.1483
9	3.114	0.0962	0.9038	0.9733	0.0267
10	3.105	0.0615	0.9385	0.9833	0.0167
11	3.087	0.0626	0.9374	0.9225	0.0775
12	3.077	0.0874	0.9126	0.9667	0.0333
13	3.072	0.0669	0.9331	0.8562	0.1438
14	3.071	0.0438	0.9562	0.9430	0.0570
15	3.061	0.0618	0.9382	0.9615	0.0385
16	3.056	0.0877	0.9123	0.9228	0.0772
17	3.038	0.0745	0.9255	0.8212	0.1788
18	3.030	0.0408	0.9592	0.8660	0.1340
19	3.019	0.0456	0.9544	0.9205	0.0795
20	2.994	0.0457	0.9543	0.7384	0.2616

Supplemental Table 3. Haplotype sharing and permutation test results for two SNP densities in 500kb windows (5-15 SNPs and >15 SNPs). The highest haplotype sharing is shown for each breed with one of four wolf populations. Permutation test 1 determined if there is significantly more haplotype sharing with Middle Eastern or East Asian wolves. Permutation test 2 assessed whether any one of the four test wolf populations had excess haplotype-sharing with a dog breed assuming haplotypes are equally represented among all wolf populations. All breeds are represented by nine individuals (bold values indicate p-value<0.05; PBGV: Petit Basset Griffon Vendeen). The percent of haplotypes explained is obtained by $\max(p_{ME}, p_{CN}, p_{EA}, p_{NA})$, where ME: Middle East; EU, Europe; CN, China; NA, North America; see also Supplemental Methods).

	5-15 SNPs					≥15 SNPs				
Breed Name	Wolf group with highest sharing	Percent Haplotypes Explained	$p_{ME} - p_{CN}$	P test1	P test2	Wolf group with highest sharing	Percent Haplotypes Explained	$p_{ME} - p_{CN}$	P test1	P test2
Afghan Hound	Middle East	0.327	0.080	0.083	**0.022**	Middle East	0.306	0.089	0.056	0.104
Akita	Middle East	0.295	0.025	0.527	0.177	China	0.284	-0.032	0.413	0.293
Alaskan Malamute	Middle East	0.283	0.023	0.572	0.417	Middle East	0.283	0.028	0.440	0.304
Aust. Terr.	Middle East	0.332	0.105	**0.038**	**0.028**	Europe	0.324	0.072	0.203	0.088
Basset Hound	Middle East	0.307	0.060	0.243	0.206	Middle East	0.342	0.113	0.054	**0.026**
Beagle	Middle East	0.306	0.057	0.250	0.193	Europe	0.320	0.082	0.164	0.139
Bernese Mtn. Dog	Middle East	0.317	0.068	0.178	0.101	Europe	0.338	0.113	0.142	0.168
Borzoi	Middle East	0.327	0.086	0.074	**0.039**	Middle East	0.341	0.106	**0.060**	**0.023**
Boxer	Middle East	0.335	0.111	**0.046**	**0.042**	Europe	0.359	0.114	0.139	0.079
Briard	Middle East	0.317	0.082	0.086	0.089	Middle East	0.306	0.093	0.115	0.308
Basenji	Middle East	0.373	0.134	**0.019**	**0.001**	Middle East	0.368	0.156	**0.001**	**0.002**
Bull mastiff	Middle East	0.298	0.050	0.314	0.290	Europe	0.341	0.138	**0.043**	0.085
Cairn Terr.	Middle East	0.320	0.083	0.098	0.079	Middle East	0.309	0.089	0.123	0.228
Cardigan Corgi	Middle East	0.308	0.056	0.278	0.175	Middle East	0.327	0.076	0.218	0.119
Chihuahua	Middle East	0.314	0.067	0.140	0.088	Middle East	0.326	0.093	**0.046**	**0.026**
Chow-chow	Middle East	0.283	0.011	0.794	0.313	China	0.291	-0.044	0.188	0.115
Cavalier King Charles Sp.	Middle East	0.324	0.081	0.136	0.101	Middle East	0.328	0.092	0.113	0.082
Collie	Middle East	0.313	0.073	0.149	0.137	Middle East	0.307	0.069	0.285	0.347
Dachshund	Middle East	0.309	0.065	0.168	0.122	Middle East	0.324	0.107	0.069	0.108
Grt. Dane	Middle East	0.312	0.068	0.165	0.127	Middle East	0.324	0.067	0.282	0.132
Dob. Pin.	Middle East	0.302	0.058	0.255	0.230	Europe	0.341	0.109	0.102	0.079
Eng. Springer Sp.	Middle East	0.304	0.056	0.254	0.174	Middle East	0.340	0.123	**0.036**	**0.042**
French Bulldog	Middle East	0.313	0.069	0.173	0.150	Europe	0.340	0.109	0.120	0.123
Flat-coated Ret.	Middle East	0.311	0.062	0.218	0.155	Middle East	0.339	0.126	**0.036**	**0.041**
Glen of Imaal	Middle East	0.320	0.072	0.151	0.081	Europe	0.329	0.111	0.081	0.125
Golden Ret.	Middle East	0.325	0.083	0.096	0.069	Europe	0.351	0.144	0.067	0.110
Greyhound	Middle East	0.312	0.068	0.180	0.142	Europe	0.347	0.058	0.339	**0.035**
German Shep. Dog	Middle East	0.313	0.086	0.105	0.186	N America	0.000	0.086	0.220	0.973
German Short-haired Ptr.	Middle East	0.311	0.065	0.164	0.097	Middle East	0.325	0.109	**0.040**	0.098
Gt. Schnauzer	Middle East	0.306	0.055	0.276	0.210	Europe	0.326	0.094	0.189	0.228
Havanese	Middle East	0.307	0.063	0.179	0.153	Europe	0.296	0.036	0.517	0.321
Sib. Husky	Middle East	0.294	0.031	0.495	0.215	Middle East	0.305	0.060	0.082	**0.016**
Iibizan Hound	Middle East	0.311	0.070	0.161	0.146	Middle East	0.321	0.084	0.102	0.057
It. Greyhound	Middle East	0.318	0.075	0.144	0.117	Middle East	0.308	0.078	0.179	0.248
Irish Wolfhound	Middle East	0.310	0.058	0.294	0.205	Europe	0.342	0.044	0.507	0.074
Irish Water Sp.	Middle	0.307	0.063	0.175	0.156	Middle	0.353	0.152	**0.006**	**0.007**

	East					East				
Jack Russell	Middle East	0.307	0.071	0.130	0.142	Europe	0.312	0.105	0.057	0.206
Kuvasz	Middle East	0.311	0.062	0.193	0.107	Middle East	0.348	0.120	**0.015**	**0.003**
Labrador Ret.	Middle East	0.314	0.075	0.130	0.131	Middle East	0.337	0.136	**0.032**	0.096
Mastiff	Middle East	0.312	0.066	0.197	0.140	Europe	0.335	0.100	0.118	0.108
Mini. Bull Terr.	Middle East	0.330	0.095	0.088	0.079	Europe	0.337	0.120	0.096	0.125
Mini. Pin.	Middle East	0.320	0.069	0.169	0.087	Europe	0.334	0.096	0.068	**0.031**
Newfoundland	Middle East	0.320	0.082	0.095	0.080	Middle East	0.324	0.087	0.140	0.111
Norwich Terr.	Middle East	0.305	0.068	0.204	0.233	Middle East	0.315	0.082	0.195	0.241
Old Eng. Sheep Dog	Middle East	0.313	0.065	0.183	0.106	Middle East	0.316	0.057	0.321	0.155
Papillon	Middle East	0.316	0.079	0.100	0.101	Middle East	0.311	0.093	0.091	0.198
PBGV	Middle East	0.314	0.059	0.240	0.111	Europe	0.333	0.108	0.065	0.083
Pekingnese	Middle East	0.316	0.063	0.199	0.085	Middle East	0.338	0.110	**0.030**	**0.013**
Pembroke Corgi	Middle East	0.318	0.062	0.184	0.059	Europe	0.297	0.041	0.510	0.477
Pomeranian	Middle East	0.303	0.060	0.205	0.188	Middle East	0.326	0.121	**0.040**	0.122
Portuguese Water Dog	Middle East	0.320	0.084	0.101	0.099	N America	0.000	0.135	**0.039**	0.976
Pug	Middle East	0.312	0.063	0.259	0.208	Middle East	0.362	0.150	**0.012**	**0.007**
Rottweiler	Middle East	0.316	0.080	0.137	0.144	Middle East	0.348	0.136	**0.037**	0.064
Saluki	Middle East	0.334	0.096	**0.041**	**0.013**	Middle East	0.320	0.102	**0.045**	0.055
Scottish Terr.	Middle East	0.313	0.078	0.135	0.164	Middle East	0.323	0.118	0.073	0.183
Shih-Tzu	Middle East	0.318	0.069	0.137	0.054	Middle East	0.281	0.022	0.654	0.515
Std. Poodle	Middle East	0.307	0.064	0.163	0.129	Middle East	0.363	0.174	**0.003**	**0.023**
Shetland Sheep Dog	Middle East	0.312	0.064	0.199	0.131	Middle East	0.361	0.149	**0.049**	0.052
Std. Schnauzer	Middle East	0.322	0.091	0.071	0.071	Europe	0.342	0.085	0.207	0.082
Staff. Bull Terr.	Middle East	0.317	0.075	0.135	0.103	Europe	0.381	0.020	0.774	**0.010**
St. Bernard	Middle East	0.310	0.062	0.229	0.172	Middle East	0.331	0.102	0.103	0.093
Toy Poodle	Middle East	0.306	0.065	0.165	0.170	Europe	0.317	0.102	0.077	0.181
Whippet	Middle East	0.318	0.083	0.107	0.100	Europe	0.353	0.117	0.068	**0.038**
West Highland Terr.	Middle East	0.320	0.083	0.121	0.112	Europe	0.321	0.110	0.061	0.148

Supplemental Table 4. History of breeds with discordant phenotypic/functional and genetic group assignments[8,9].

Breed Name	Genetic Cluster (see Figure 1)	Phenotypic/ Functional Group	Breed Information[8,9]	Concordance with Historical Evidence
Briard	Small Terrier	Herding	Possible East Asian origin from crosses with local dogs to create a new breed used for flock guarding	No historical evidence for breed admixture between small terriers and herding dogs
Brussels	Toy	Terriers	European origins from crosses with Affenpinscher (terrier) and Toy breeds (i.e. English Toy Spaniels, Yorkshire Terriers, Pekingese, or Pug) to miniaturize the breed	Evidence for breed admixture between toy and terrier breeds
Chihuahua	Toy	Ancient	Probable Chinese origins with introduction to Mexico from Spanish traders returning from East Asia	Evidence for breed admixture between East Asian Ancient and toy breeds
German Shep. Dog	Gun	Herding	European breed with recent origins	Inconclusive
Gt. Schnauzer	Gun	Herding	European origins likely from crosses with smooth-haired dogs and possibly Great Danes	Inconclusive
Glen of Imaal	Mastiff-like	Terriers	European origins from crosses of Bullterriers, Staffordshire terriers (Mastiff-like breeds) and other fighting dogs; Glen of Imaal is an aggressive hunter (e.g. badgers, rats)	Evidence for admixture between Mastiff-like and terrier breeds
Mini. Pin.	Toy	Terriers	European origins from crosses of German Pinscher (terrier) and Dachshunds or Italian greyhounds	Evidence for admixture between toy and terrier breeds
Newfoundland	Retrievers	Mastiff-like	North American origins with possible crosses to Mastiff or Portuguese Water dog; considered an ancestor of the modern Labrador Retriever	Evidence for Retriever and Mastiff-like breed admixture
Papillon	Toy	Spaniels	European origins from crosses of Spaniels and Bichon-type (toy) breeds	Evidence for admixture of toy and spaniel breeds
Pekingnese	Toy	Herding	Chinese origins; considered a dwarfed Tibetan terrier or Pug (toy)	Evidence of admixture of toy and other breeds
Pomeranian	Toy	Spitz	European origins from crossing European herding and spitz-type breeds	Inconclusive
Portuguese Water Dog	Gun	Spaniels	European origins; bred to be a water dog	Inconclusive
Pug	Toy	Mastiff-like	Chinese origins; considered a “mini-mastiff”, likely from miniaturizing the Affenpinscher (Terrier) or the English Bulldog and crossing with the Tibetan Mastiff (Mastiff-like breeds)	Evidence for breed admixture of Mastiff-like and toy breeds
Shih Tzu	Toy	Herding	Tibet/Chinese origins; considered a dwarf of Tibetan terriers or Lhasa Apsos (herding breeds)	Evidence for admixture of Toy and herding breeds
Std. Schnauzer	Gun	Herding	European origins from crossing the Standard Pinscher, Poodles, “Wolfspitzs", or Shepherds	Inconclusive

Supplemental Table 5. Analysis of molecular variance for groupings of dogs and wolves as follows: 1) groups in Fig. 1; 2) geographic dog breed groups (Fig. 2; Supplemental Table 1); and 3) wolves and dogs as separate populations (df = degrees of freedom, SS = sum of squares; all comparisons have p<0.001).

Analysis	Grouping tested	df	SS	Variance component	Percent (%) of variation
1. Breed Groups	Among dog breed groups [Φ_{ct}]	9	11.49	0.006 Va	3.8
	Among dog breeds within dog breed groups [Φ_{sc}]	67	47.50	0.053 Vb	31.1
	Within dog breeds [Φ_{st}]	794	88.10	0.111 Vc	65.1
2. Geographic groups	Among geographic dog breed groups [Φ_{ct}]	6	6.66	0.007 Va	4.3
	Among dog breeds within geographic dog breed groups [Φ_{sc}]	77	54.78	0.056 Vb	31.9
	Within dog breeds [Φ_{st}]	818	90.87	0.111 Vc	63.8
3. Wolves and Dogs	Among dog-wolf group [Φ_{ct}]	1	11.37	0.041 Va	19.9
	Among dog breeds and wolf populations [Φ_{sc}]	105	67.90	0.054 Vb	26.5
	Within dog breeds and wolf populations [Φ_{st}]	960	104.85	0.109 Vc	53.6

Supplemental Table 6. Top-ranked “clusters” of empirical outliers for the XP-EHH and FST statistics. If two or more SNPs were in the 95th percentile of the bi-variate percentile score and were spaced less than 300kb apart, they were joined into a single cluster. We then ranked clusters by the number of SNPs they contain, and for all clusters with the same number of SNPs, we sorted them by the bi-variate percentile score of the central SNP. For each cluster, we show the position of the central SNP, the size in base pairs, the number of SNPs, the maximum of the bivariate percentile score, and then include several rows with the names of genes overlapping the cluster (if known) and their putative functions based on online resources (principally OMIM). Gene names in parentheses represent genes that were labeled as unknown in ENSEMBL but had been annotated in other resources (EntrezGene). In cases where there were genes of unknown function, we denote them on the last row of information for each cluster.

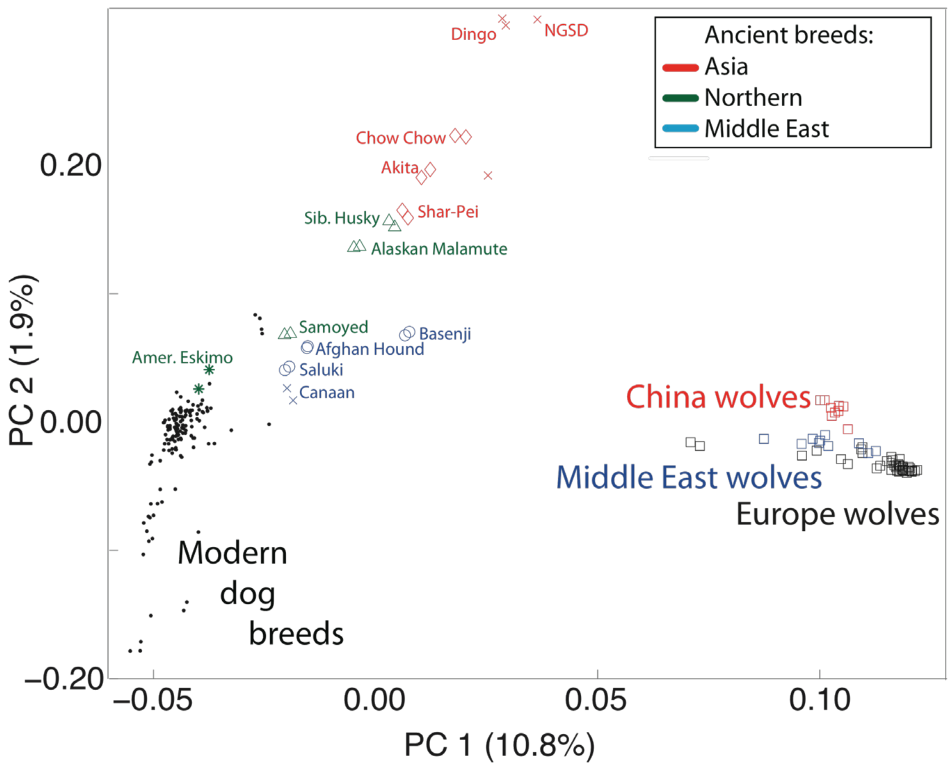
Chr	Position	Cluster Size (bp)	# of SNPs	Max Joint percentile	Gene Name	Putative Function
5	5,323,685	623,946	11	0.997	(OPRM1)	Inhibits neurotransmitter release by reducing calcium ion currents and increasing potassium ion conductance
					(hNT)	Neural cell adhesion molecule
30	4,357,124	1,239,195	11	0.992	LOC607369	Uncharacterized protein C15orf29
					LOC478242	UPF0480 protein C15orf24 Precursor
					CHRM5	Cholinergic receptor mediating various cellular responses, including inhibition of adenylate cyclase, breakdown of phosphoinositides and modulation of potassium channels through the action of G proteins
					AVEN	Protects against apoptosis mediated by Apaf-1
					RYR3	Intracellular calcium ion release channels responsible for the release of calcium from intracellular stores following transduction of many different extracellular stimuli
13	30,806,609	543,841	8	0.997	ADCY8	Catalyses the formation of camp from ATP
17	40,389,688	275,966	5	0.982	IL1F5, IL1F8, IL1F10	Participates in a network of interleukin 1 family members to regulate adapted and innate immune responses
					IL1RN	Inhibits the activity of IL-1 by binding to its receptor
					(PSD4)	Contains pleckstrin domain (intracellular signaling or cytoskeleton) and Sec7 domain (guanine nucleotide exchange)
					Pax8	Transcription factor for the thyroid-specific expression of the genes exclusively expressed in the thyroid cell type
					+3 genes of unknown function, 1 snoRNA, and 1 pseudogene	
1	7,888,709	446,711	5	0.964	ZNF407	May be involved in transcriptional regulation
					CNDP1 CNDP2	Carnosinase and peptidase A, associated with diabetes
					+5 genes of unknown function	
12	42,565,679	161,093	4	0.987	N/A	
36	20,651,032	387,059	4	0.961	(MRK)	Signal transduction during cell cycle arrest and checkpoint regulation
					Unknown	
1	20,841,571	212,973	3	0.996	NEDD4L	E3 ubiquitin-protein ligase
24	20,236,943	435,197	3	0.990	PRND, PRNP	Mutations in the repeat region as well as elsewhere in this gene have been associated with Creutzfeldt-Jakob disease, fatal familial insomnia, Gerstmann-Straussler disease, Huntington disease-like 1, and kuru.
					LOC485786	
					SMOX	Involved in the regulation of polyamine intracellular concentration and has the potential to act as a determinant of cellular sensitivity to antitumor polyamine analogs.
					SAMD12	sterile alpha motif domain containing 12
					snRNA	
13	20,988,744	277,010	3	0.988	TNFRSF11B	Acts as decoy receptor for RANKL and thereby neutralizes its function in osteoclastogenesis. Inhibits the activation of osteoclasts and promotes osteoclast apoptosis in vitro.
14	63,313,543	164,185	3	0.988	CADPS2	Calcium-binding protein involved in exocytosis of vesicles filled with neurotransmitters and neuropeptides.
2	85,645,949	146,038	3	0.985	n/a	
16	19,679,560	172,721	3	0.983	Pseudogene	
1	111,507,791	682,061	3	0.979	MEIS3	Myeloid ecotropic viral integration site 1 homolog 3
					GPR77	Receptor for the chemotactic and inflammatory peptide anaphylatoxin C5a, C4a and C3a
					C5AR1	Receptor for the chemotactic and inflammatory peptide anaphylatoxin C5a.
					+3 genes of unknown function	

“Chr” and “Position” denote the position of the SNP with the maximum joint percentile for F_{ST} and XP-EHH, “Cluster size” is the size of the cluster in basepairs, “number of SNPs” is the number of extreme SNPs found in the cluster, “Max joint percentile” is the maximum joint percentile of F_{ST} and XP-EHH. The gene names are derived from the EntrezGene annotations or if in parentheses, ENSEMBL. Functions are from searches of OMIM and EntrezGene.

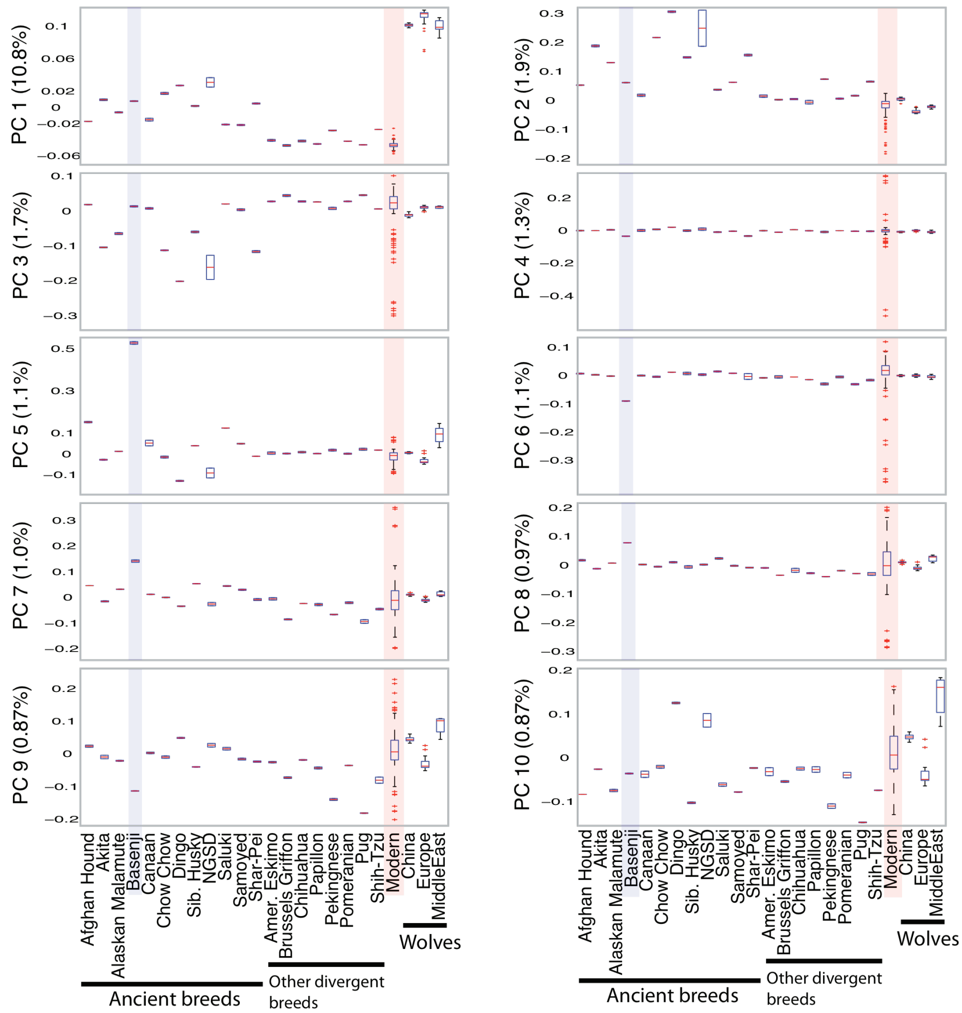
Supplemental Table 7. Number of SNPs for each ascertainment category from 48,036 SNPs used in this study. Ascertainment panel indicates the breed or species to which the Boxer genome sequence was compared (wolf populations: Alaska, China, India, and Spain; Coyote, California[38]).

Ascertainment Panel	SNPs (~48K)
Boxer x Boxer	13,318
Boxer x Dog	5653
Boxer x Coyote x Dog	6
Boxer x Wolf x Dog	68
Boxer x Wolf x Coyote x Dog	1
Boxer x Poodle	27,742
Boxer x Poodle x Dog	614
Boxer x Wolf x Poodle x Dog	0
Boxer x Wolf x Coyote x Poodle x Dog	0
Boxer x Coyote x Poodle	6
Boxer x Wolf x Poodle	0
Boxer x Coyote	146
Boxer x Wolf	480
Boxer x Wolf x Coyote	2
Total	48,036

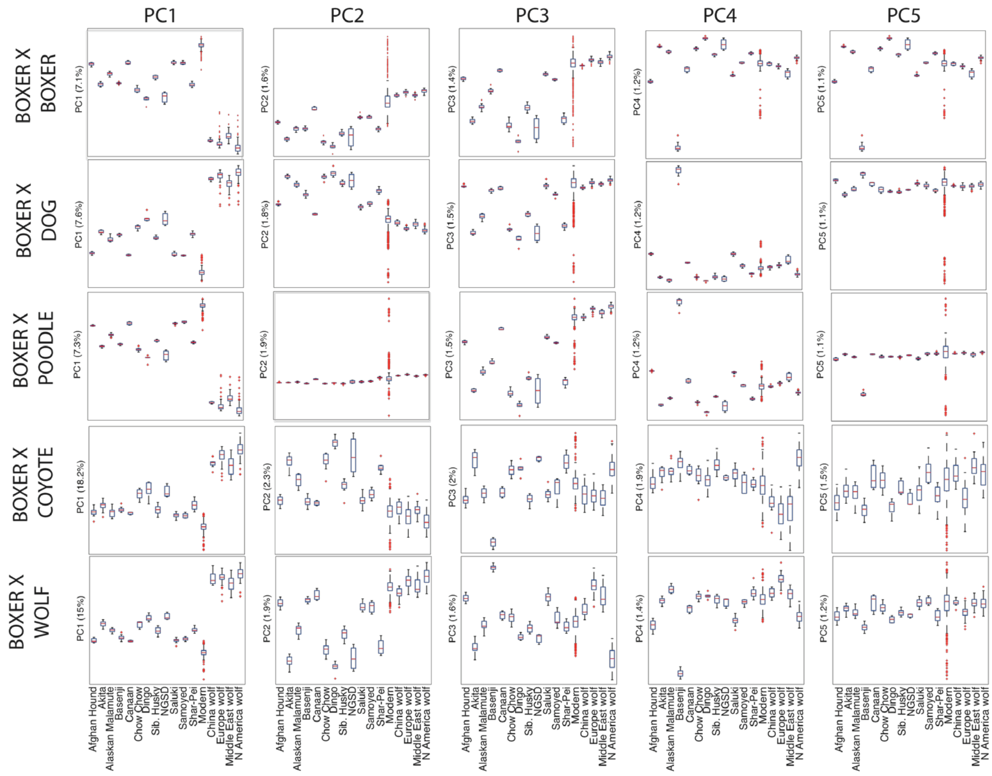
Supplemental Figure 1. Principal component analysis (PCA) of 48,036 SNPs for two representative dogs per breed (n = 171) and Eurasian wolves (n = 58). As domestication is generally believed to have taken place in Eurasia[19], we excluded North American wolf populations from the analysis. NGSD is the New Guinea Singing Dog.



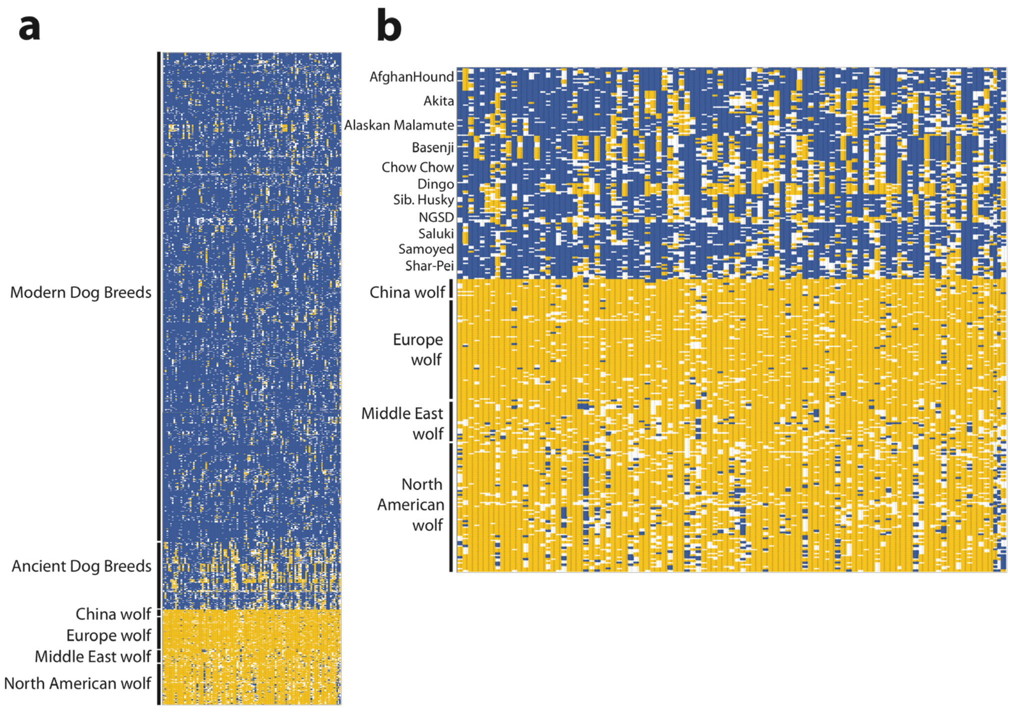
Supplemental Figure 2. The first five principal components of a dog-wolf PCA. Ancient lineages are compared to modern breeds (red color) and gray wolf populations (right). The Basenji is indicated by a purple line. Percent of variation explained by each component is indicated in parenthesis on the y-axis. NGSD is the New Guinea Singing Dog.



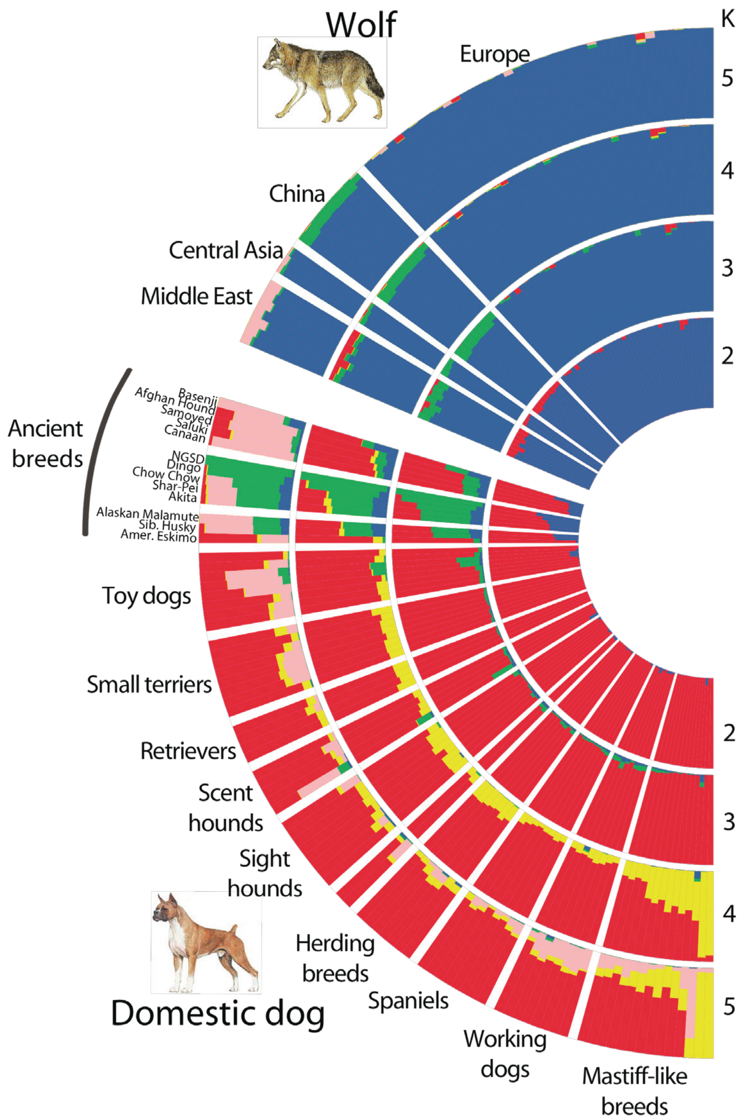
Supplemental Figure 3. PCA of each ancient dog breed and modern breeds (n = 912) and wolves (n = 155) with subsets of SNPs based on ascertainment method (boxer, n = 13,318 SNPs; dog, n = 5,632 SNPs; poodle, n = 27,671 SNPs; coyote, n = 146 SNPs; wolf, n = 480 SNPs). Similar clustering trends are observed across ascertainment panels. NGSD is the New Guinea Singing Dog.



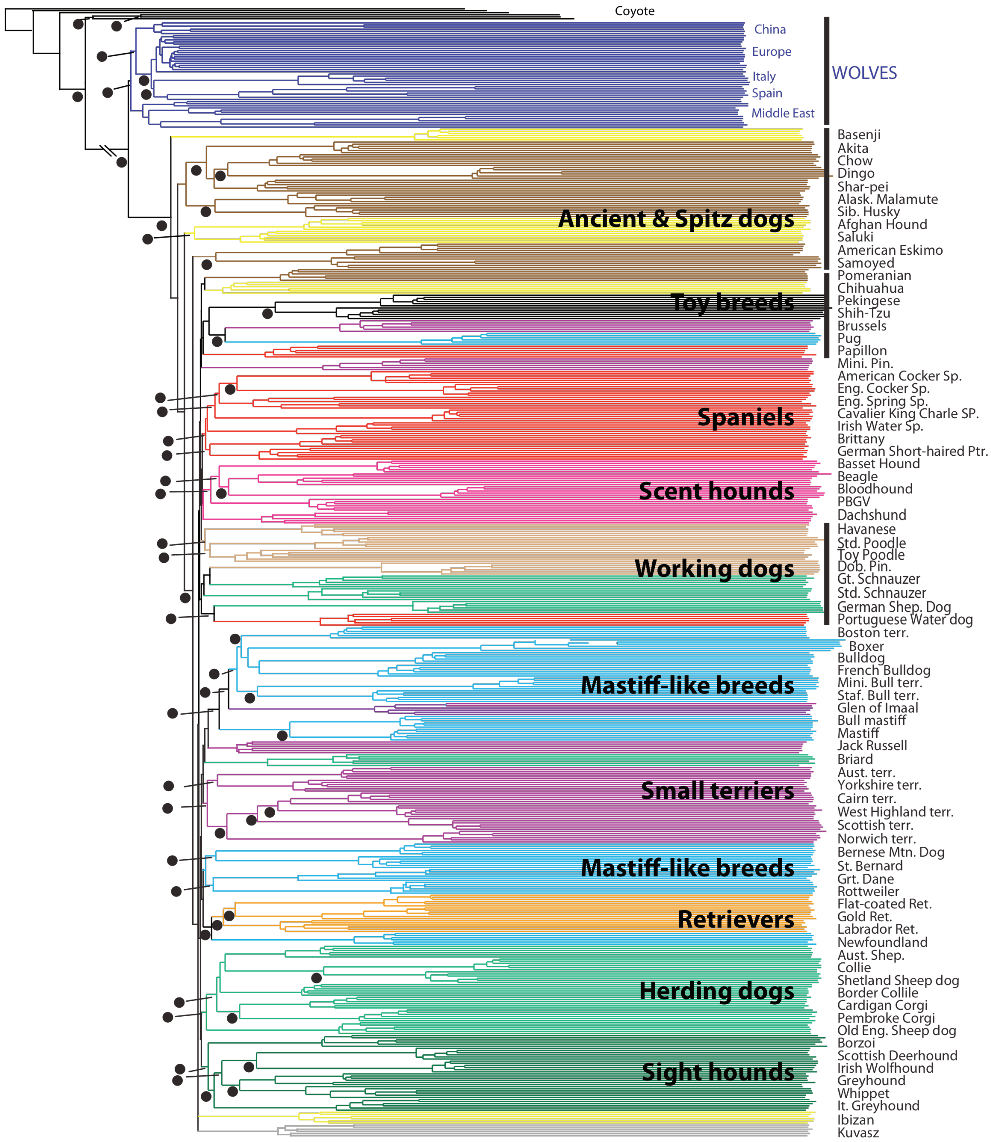
Supplemental Figure 4. Genotypes of the 100 SNPs with highest loadings on PC1 (Supplemental Fig. 1). **A.** SNP genotypes for the entire wild and domestic canid sample. **B.** Subset of genotypes from **A.** highlighting just ancient dog lineages and gray wolf populations. SNPs are ranked with the left being the top-ranking SNP in descending order towards the right (blue indicates the major allele in dogs; yellow indicates the major allele in wolves). NGSD is the New Guinea Singing Dog.



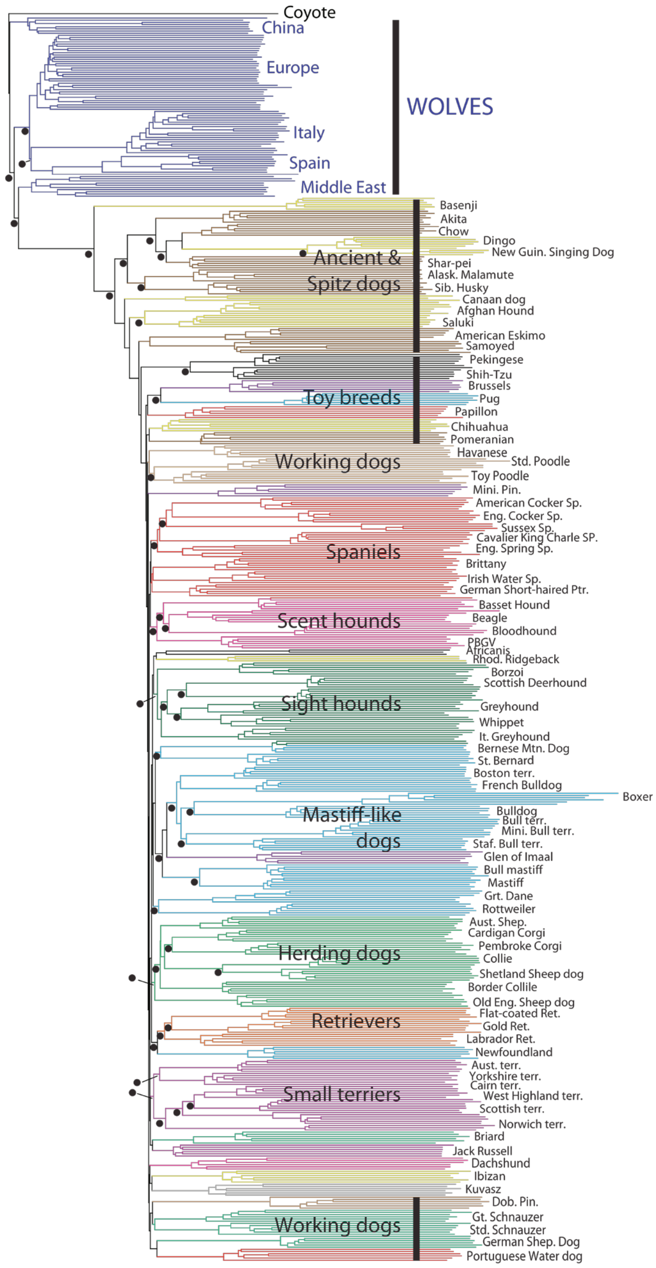
Supplemental Figure 5. Ancestry analysis of dog breeds (n = 85; one dog per breed) and Eurasian gray wolves (China, n = 9; Middle East, n = 9; Europe, n = 43) using the program STRUCTURE[35] for 43,954 pruned SNPs (LD pruned: $r^2<0.5$). As domestication is generally believed to have taken place in Eurasia[3], we excluded North American wolf populations from the analysis. We varied the number of ancestral populations (K) from 2 to 5. The composition of each individual genome is reflected by colors. The absence of the blue wolf component in modern dog breeds at K=2 suggests an absence of admixture between them and gray wolves. NGSD is the New Guinea Singing Dog.



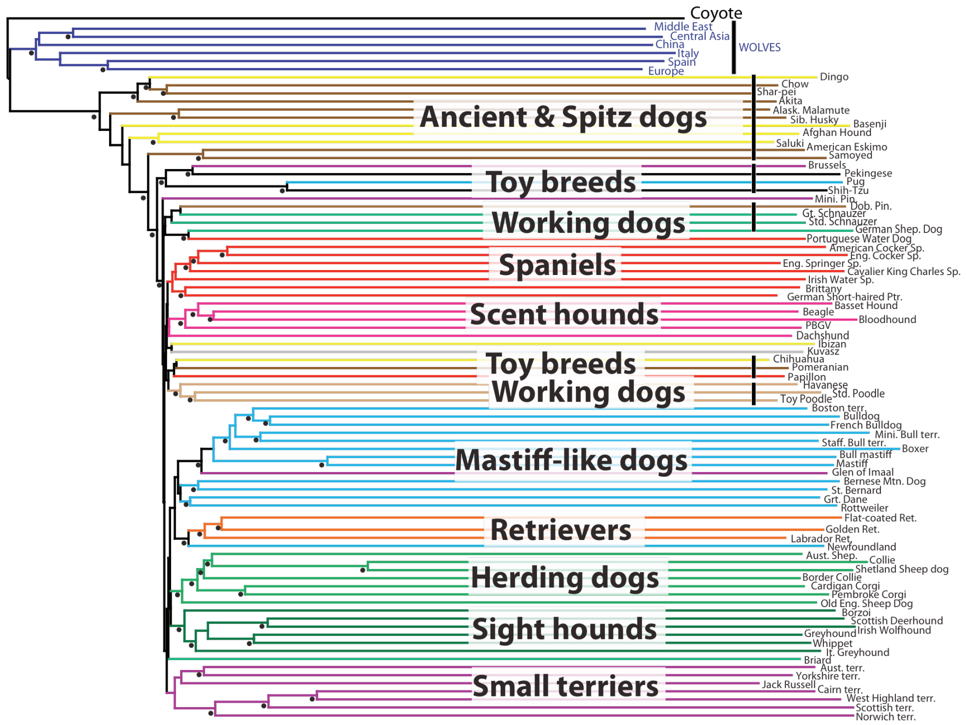
Supplemental Figure 6. Consensus haplotype-sharing neighbor-joining phylogram for phased SNP data for non-overlapping 10-SNP windows (n = 6 for all breeds and wolf populations and breeds with n<5 excluded). A dot indicates >95% bootstrap support from 1,000 replications. See Fig. 1a for corresponding cladogram.



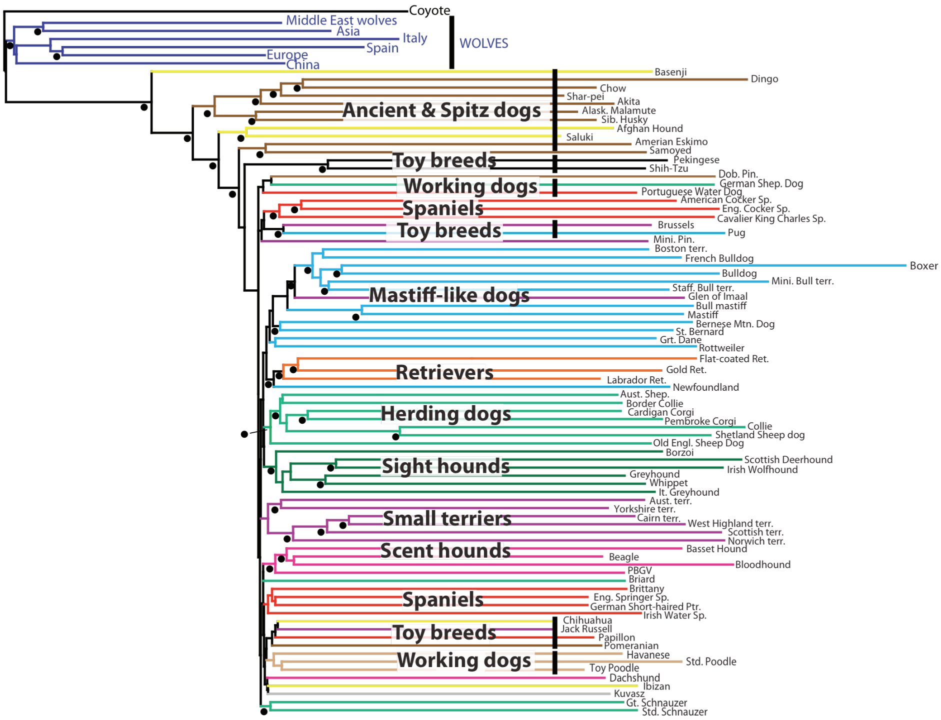
Supplemental Figure 7. A neighbor-joining phylogram of individuals constructed from allele-sharing distances. This consensus tree was generated from 1,000 bootstrap replications and rooted with coyote SNP data (a dot at a node indicates >95% bootstrap support). See Fig. 1b for corresponding cladogram.



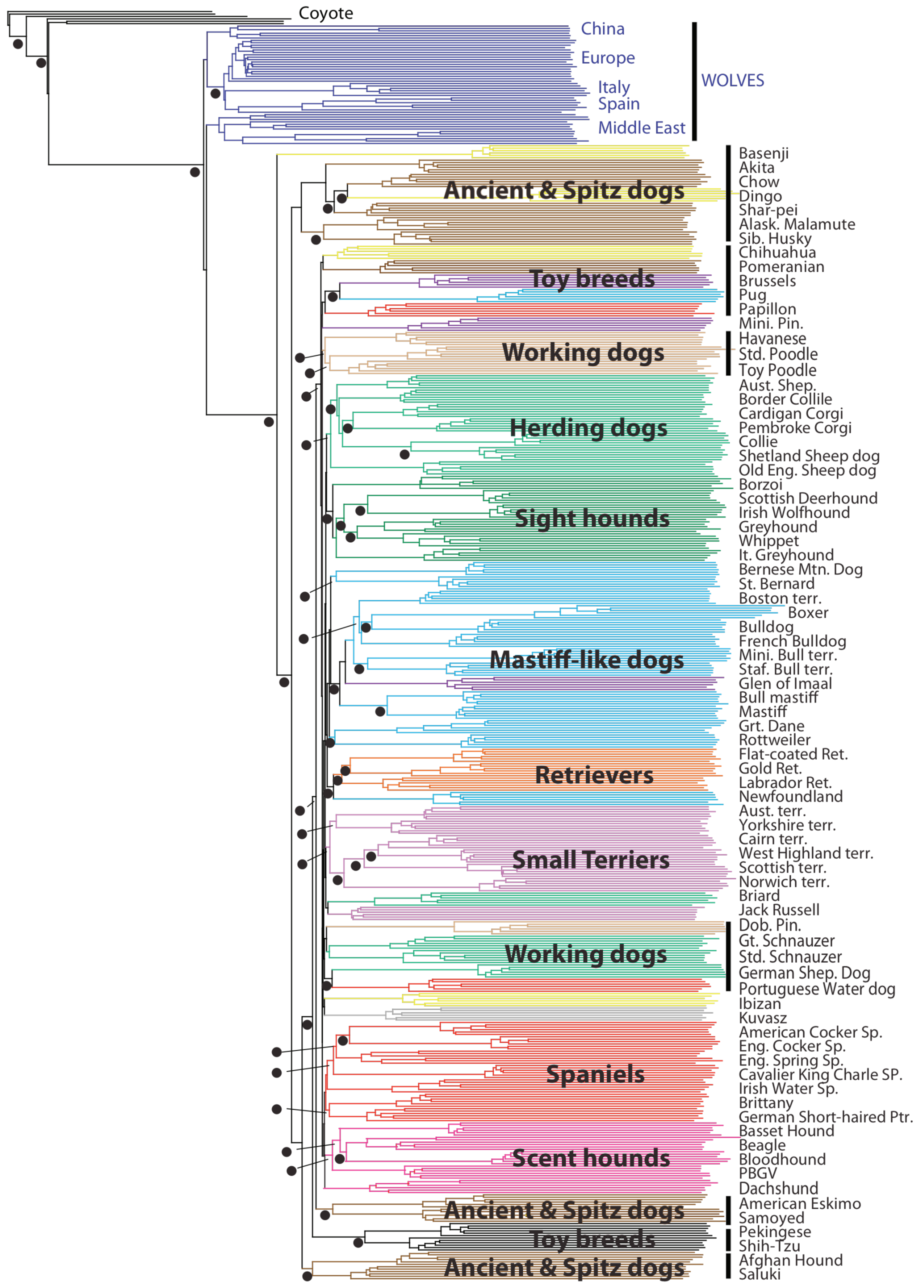
Supplemental Figure 8. A neighbor-joining phylogram of dog breeds and wolf populations based on haplotype-sharing distances between populations for 10-SNP windows. The consensus tree was generated with 1,000 bootstrap replications and rooted with coyote SNP data (a dot at a node indicates >95% bootstrap support).



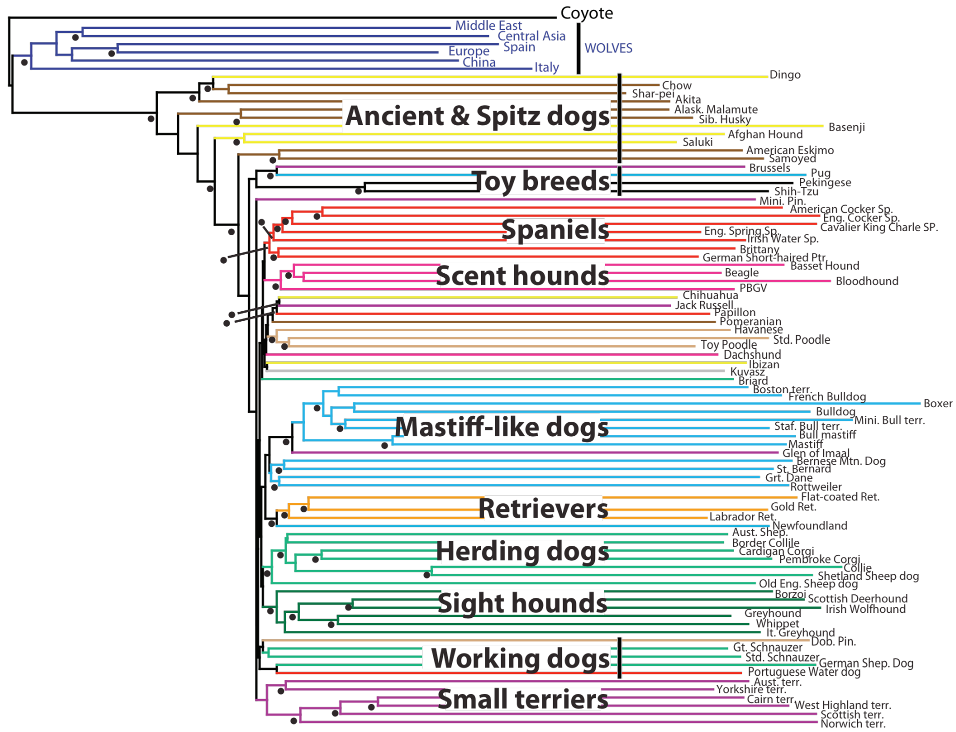
Supplemental Figure 9. A neighbor-joining phylogram based on allele-sharing distances among dog breeds and wolf populations. The consensus tree was generated with 1,000 bootstrap replications and rooted with coyote SNP data (a dot at a node indicates >95% bootstrap support).



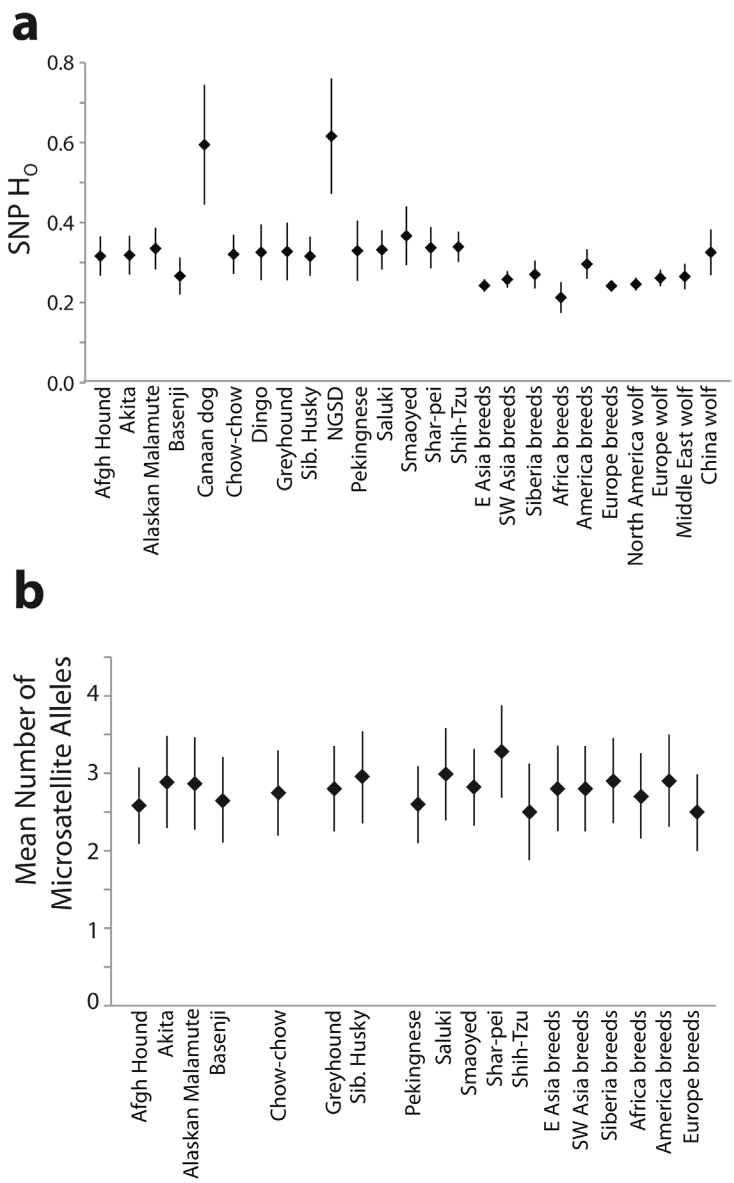
Supplemental Figure 10. A neighbor-joining phylogram of dog breeds and wolf populations based on haplotype-sharing distances between individuals for 5-SNP windows. The consensus tree was generated with 1,000 bootstrap replications and rooted with coyote SNP data (a dot at a node indicates >95% bootstrap support).



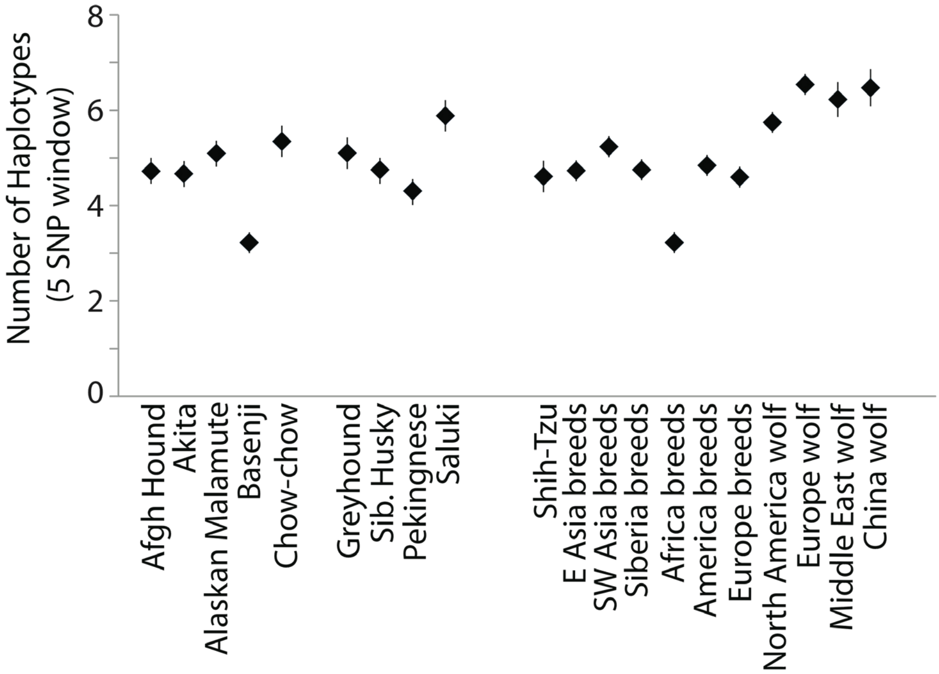
Supplemental Figure 11. A neighbor-joining phylogram of dog breeds and wolf populations based on haplotype-sharing distances between populations for 5-SNP windows. The consensus tree was generated with 1,000 bootstrap replications and rooted with coyote SNP data (a dot at a node indicates >95% bootstrap support).



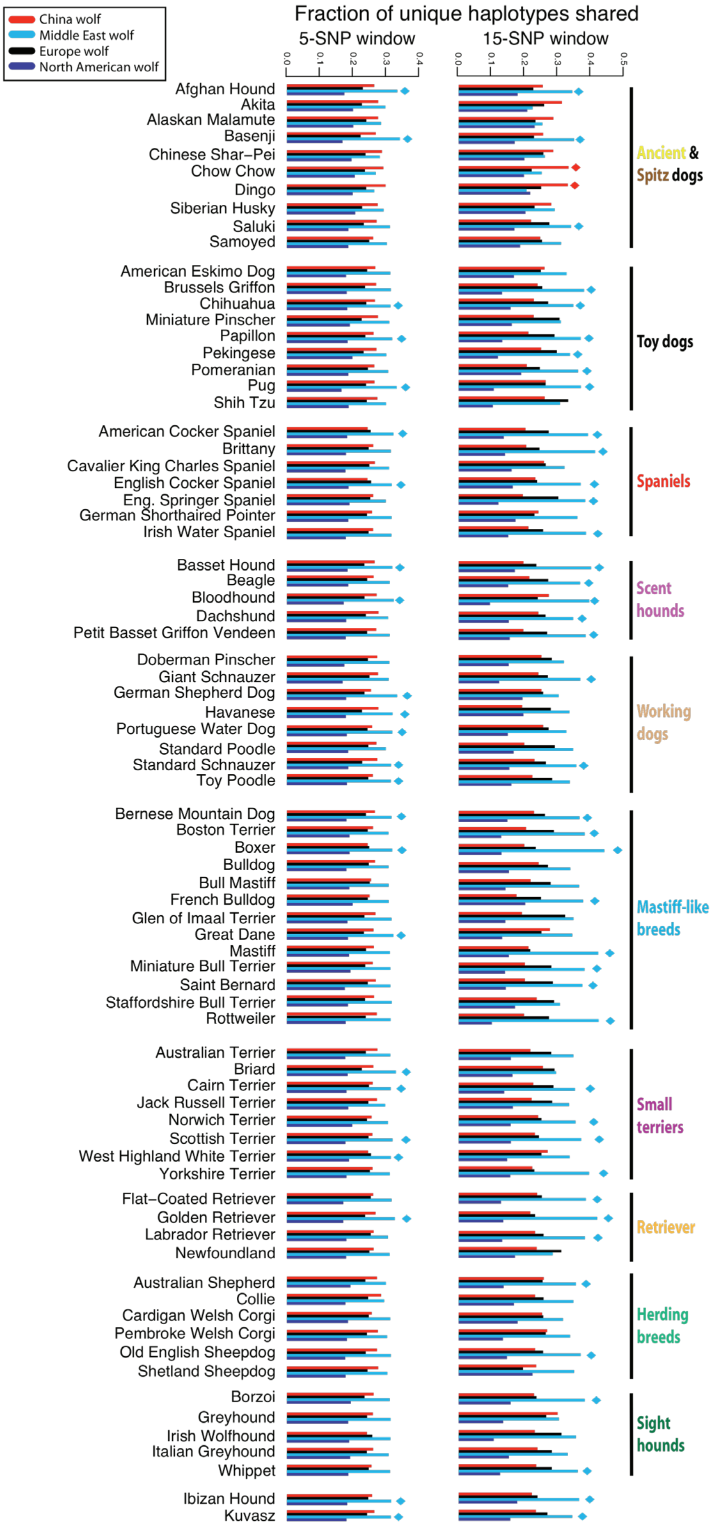
Supplemental Figure 12. Estimates of **A.** SNP-based average observed heterozygosity (± s.e.m bars) for 546 SNPs ascertained from dog-wolf genome comparisons; and **B.** average number of alleles (± s.e.m bars) from a previous microsatellite survey[7]. NGSD is the New Guinea Singing Dog.



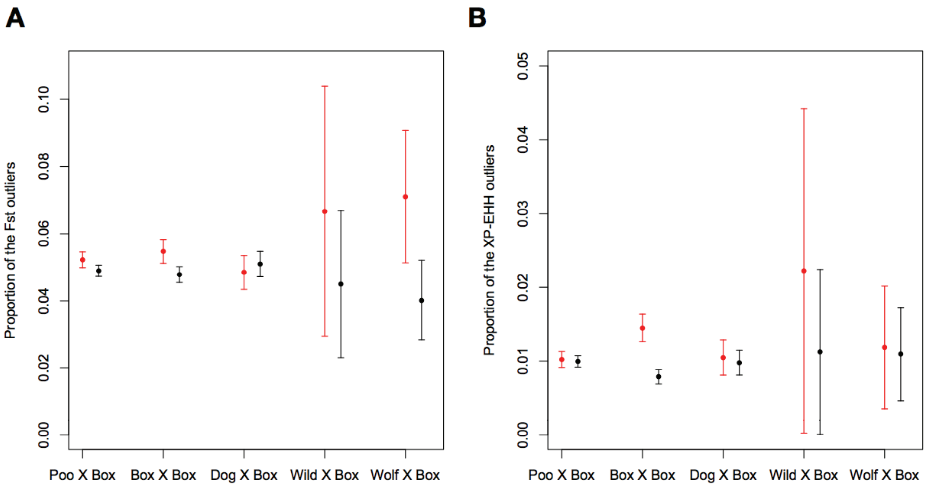
Supplemental Figure 13. Average number of haplotypes (± s.e.m bars) per breed or breed group for phased SNP loci across 5-SNP windows. Note the higher diversity in gray wolves as predicted because haplotype data are expected to show less ascertainment bias. NGSD is the New Guinea Singing Dog.



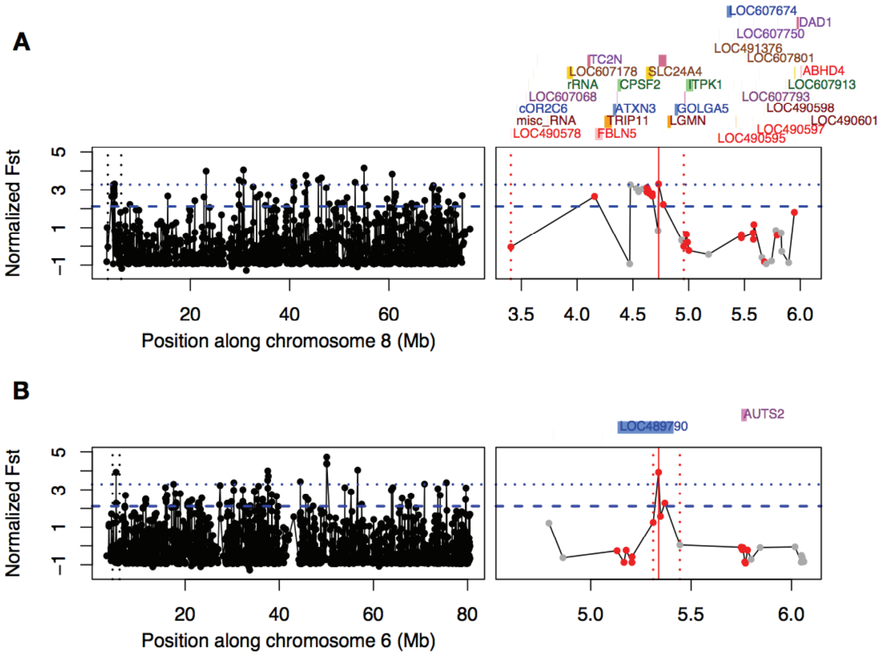
Supplemental Figure 14. The fraction of unique haplotypes shared between 77 dog breeds and each of four wolf populations (China, Europe, North America, and Middle East) for 5 (left panel) and 15-SNP (right panel) haplotype windows. Six individuals represent each breed and wolf population; consequently, breeds with fewer individuals are not included in this analysis. The diamond to the right of histogram bars indicates significantly higher sharing (p<0.05) using permutation test 2 (Supplementary Note A) and the color of the diamond indicates the wolf population having the highest sharing.



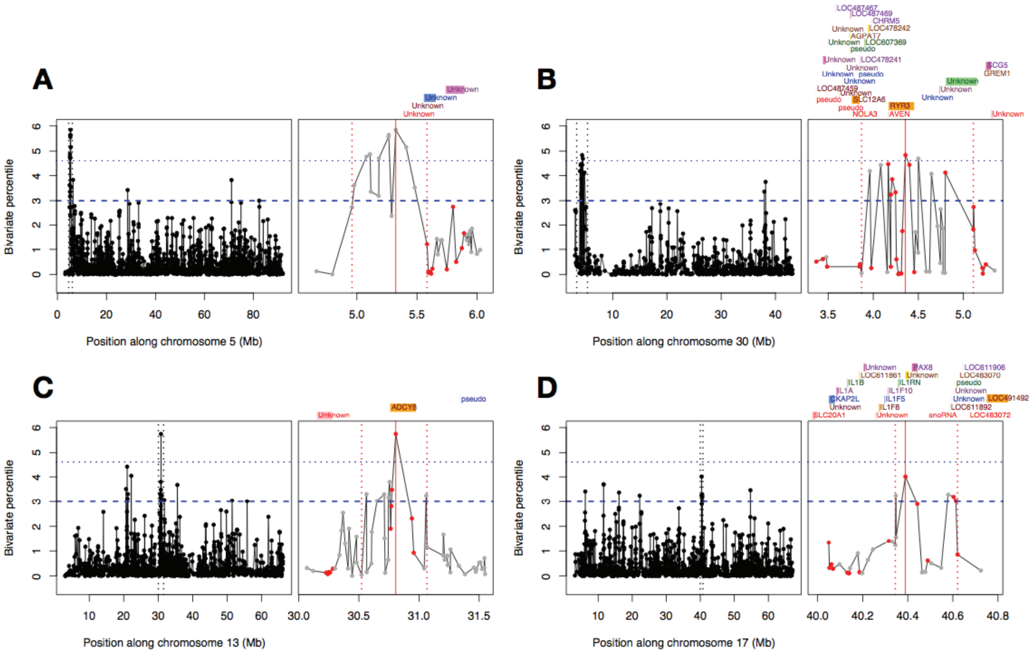
Supplemental Figure 15. Enrichment of genic regions for F_{ST} and XP-EHH outliers. The figure shows for each ascertainment panel the proportion of SNPs that are genic (red) or non-genic (black) for those SNPs whose **A.** F_{ST} values fall into the upper 5% tail of normalized across panels or whose **B.** XP-EHH values fall into the into the upper 1% tail of normalized values. There is variation across panels in the strength of the effect, but overall the enrichment of genic regions is significant across all panels (p=0.04 for F_{ST}, p=0.02 for XP-EHH, one-sided exact conditional test, controlling for the ascertainment panel).



Supplemental Figure 16. Putative signature of positive selection in the genomic region near **A.** SLC24A4 and **B.** WBSCR17 (described as LOC489790). The plots are arranged as in Supplemental Figure 17, but the vertical axis only shows the normalized F_{ST} scores (i.e. F_{ST} normalized to have mean zero and standard deviation one within each ascertainment bias class; only known genes are indicated in the figure; unknown and pseudo genes are not included).



Supplemental Figure 17. Putative signatures of positive selection in dogs. The bivariate percentile score measures jointly how extreme the F_{ST} and XP-EHH scores are for a given SNP relative to the empirical distribution for all SNPs [i.e. -log(F_{ST} empirical p-value) x (XP-EHH empirical p-value)]. Each left sub-panel shows a view at the scale of the complete chromosome and each right sub-panel shows a focal region with a cluster of SNPs that showed extreme values. Dashed and dotted horizontal lines represent the 95th percentile and 99th percentile of the scores, respectively. Vertical dotted lines represent the extent of each cluster, with the solid red line demarking the central SNP of the cluster. EntrezGene annotations are plotted above the right sub-panel (known SNPs are plotted as short vertical lines; for further descriptions of **A – D**, see Supplemental Note C).



Supplemental Figure 18. Putative signature of positive selection in the CNDP1/CNDP2 region. See Supplemental Figure 17 for description of sub-panels and axes (only known genes are indicated in the figure; unknown and pseudo genes are not included). Discussion in Supplemental Note C.