

Supporting Online Material

Methods

Sample Collection and DNA Extraction. DNA samples from 513 American Kennel Club (AKC) registered dogs representing 103 breeds were obtained by collecting buccal (cheek) swabs and/or blood samples from AKC-sanctioned dog shows, specialty events, and mail-in donations. SNP analysis used two unrelated dogs from each of 60 breeds, and microsatellite analysis used 4-5 unrelated dogs from each of 85 breeds. Particular efforts were made to sample the most popular breeds. AKC registration number and detailed pedigree information were requested for all dogs, as participation was limited to unrelated dogs which did not share grandparents. Pedigree information was collected for 84% of sampled individuals. In many cases, five-generation pedigrees were obtained, and while dogs sometimes appear redundantly at the great-grandparent level or higher, inspection of the complete lineage indicates a high degree of unrelatedness among dogs of the same breed. For those individuals for whom a pedigree was not available, unrelatedness was verified by breed club representatives. Buccal swab samples were collected using standard protocols, as suggested by the AKC (<http://www.akc.org/>) using cytology brushes (Medical Packaging Corp., Camarillo, CA). DNA was extracted from buccal swabs using QiaAmp DNA extraction kits following manufacturer's protocol (Qiagen, Valencia, CA). DNA was extracted from blood samples using a phenol/chloroform protocol as described(1).

SNP Discovery and Genotyping. Fifty canine BACs were chosen at random from the canine radiation hybrid map(2). Primer3 (<http://www-genome.wi.mit.edu/cgi->

bin/primer/primer3_www.cgi) was used to design primers from each BAC end sequence. The resulting amplicons averaged 334 base pairs in length. Primers were used to amplify genomic DNA, and the resulting PCR products were sequenced using standard methods on an ABI 3700 capillary sequencer with standard ABI dye terminator chemistry (ABI, Foster City, CA). Sequence reads were aligned and viewed using Phred, Phrap and Consed(3-5). Polyphred was used to identify polymorphisms(6). All allele calls were confirmed through visual inspection of the traces. A complete list of BACs, primer sequences and SNP loci is available at

http://www.fhrc.org/science/dog_genome/dog.html.

Microsatellite Genotyping. One hundred dinucleotide microsatellite markers were chosen from the 1596 microsatellites currently localized on the 3300 marker map of the dog (2). Markers were selected based on informativeness, calculated as a PIC value, and distribution across all autosomes. Markers selected had an average PIC value of 0.65 (range 36%-86%) and an average spacing of 29.5 Mb (range 21.5 - 50.9 Mb). Forward primers were redesigned to include the 19 base M13 sequence on the 5'-end. A complete list of markers, primer sequences and map locations is provided in Table S3 and at http://www.fhrc.org/science/dog_genome/dog.html. DNA samples were arrayed in four 96-well plates. A positive control was included on each plate to ensure consistent allele binning. PCR was carried out using one ng of genomic DNA and standard protocols (http://www.fhrc.org/science/dog_genome/dog.html). Samples were labelled by the addition of 0.25 pmol of an M13 primer tagged with either 6FAM, VIC, NED or PET dyes (ABI, Foster City, CA) to each reaction. Four samples labelled with different dyes were multiplexed following completion of PCR. Samples were denatured in Hi-Di

formamide with 15 pmol of GeneScan-500LIZ size standard (ABI, Foster City, CA) according to manufacturer's protocols. All samples were loaded on an ABI 3730 capillary electrophoresis instrument for allele separation. Genotypes were called using GeneMapper 3.0 (ABI, Foster City, CA). All calls were checked manually and each plate was scanned for the appearance of new alleles outside existing bins. Four markers failed to amplify consistently and were discarded.

Statistical Analysis. Expected heterozygosity for each breed was calculated from allele frequencies using Tajima's unbiased estimator (7). The results are shown in Table S4 (microsatellites) and Table S5 (SNPs). Analysis of molecular variance (AMOVA) was performed with GDA(8) under assumption of Hardy-Weinberg equilibrium. Similar results were obtained for the fraction of genetic variation among breeds when inbreeding was allowed for in the analysis. The assignment test was carried out with the *Doh* assignment test calculator available from J. Brzustowski (<http://www2.biology.ualberta.ca/jbrzusto/Doh.php>). All dogs were designated with their known breed except for the one dog to be tested, which was then assigned by the program to the breed with the highest probability of generating the test dog's genotype. The program repeats this procedure with each dog as the test dog. The Belgian Sheepdog and Belgian Tervuren breeds were combined into one designation for this analysis; when they are treated as separate breeds the individual dogs are assigned to one or the other essentially at random.

We ran *structure*(9, 10) for 100,000 iterations of the Gibbs sampler after a burn-in of 20,000 iterations. The correlated allele frequency model was used with asymmetric admixture allowed. The similarity coefficient across runs of *structure* was computed as

described (11). When we ran the program on a partial data set of 68 breeds, we noted that at values of K above 40 the program created clusters to which no individuals were assigned, and the clusters were unstable from run to run. This is most likely because the algorithm, which was initially designed to separate 2-3 populations, is unable to handle such large numbers of populations simultaneously. Because *structure* has previously been shown to reliably separate 20 populations (12), we divided the entire data set into 8 subsets of 10 to 11 breeds each and analyzed all possible pairs of these subsets. We retained historically related or morphologically similar breeds in the same subset. Three runs of *structure* were performed with K equal to the total number of breeds in each subset. Figure 1 is a composite of representative runs showing the most consistent breed pairs. The Belgian Sheepdog and Belgian Tervuren grouped together in all runs. The Collie and Shetland Sheepdog formed a breed pair in 95% of runs, while the Mastiff and Bullmastiff and Bernese Mountain Dog and Greater Swiss Mountain Dog formed pairs in 90% of the runs. The Greyhound and Whippet and the Siberian Husky and Alaskan Malamute formed pairs in 62% and 67% of the runs, respectively. In 95% of all runs the Cairn Terrier appeared to share a portion of its genome with the West Highland White Terrier. Also of note was the intermixing of clusters involving the Lhasa Apso, Pekingese and Shih Tzu, at least two of which shared a cluster or were divided between the same clusters in 57% of the runs.

We then applied *structure* to the entire data set at K=2 to K=10, with fifteen runs at each K (see Figure S1 for all runs at K=2 to 5). As K is increased, *structure* first separates the most divergent groups into clusters, followed by separation of more closely related groups (11). In our analysis, the likelihood increased with increasing values of K,

reflecting additional structure found at each K, but multiple different clustering solutions were found for $K > 4$, and we therefore used $K = 2$ to 4 to describe the global breed structure, with phylogenetic analysis and cluster analysis of subgroups used to define constellations of closely related breeds. *Structure* runs at $K = 2-5$ were repeated under the no admixture model with similar results. In a separate analysis, eight wolves were added to the structure run at $K = 2$. The wolves were sampled from eight countries: China, Oman, Iran, Italy, Sweden, Mexico, Canada (Ontario) and the United States (Alaska). All wolves clustered together with the first (yellow) cluster of dog breeds.

Each breed was assigned to one of the four groups based on breed average majority and *structure* was run on each group at $K = 2-4$. No additional consistent patterns were observed within the individual groups apart from the reported breed pairs and trio.

We carried out outlier analysis using the software package *fdist2* available at <http://www.rubic.rdg.ac.uk/~mab/software.html>. Eleven markers were identified as potential “outliers” with F_{st} values above the 95th percentile achieved by simulation under the infinite allele model with 85 populations assumed and an average of 10 haploid genotypes per population (13). Assignment and *structure* analysis performed with these markers removed did not result in significant changes.

For the phylogenetic tree analysis, individual dogs and wolves were assigned to one of 86 populations based on breed or species. Distances between the populations were computed using the program *Microsat* (14) with the chord distance measure. 500 bootstrap replicates were generated. This program can be downloaded from the website <http://hpgl.stanford.edu/projects/microsat/microsat.html>. Neighbor-joining trees were

constructed for each replicate using the program Neighbor, and the program Consense was used to create a majority-rule consensus tree. Both of these programs are part of the Phylip package (15) available at <http://evolution.genetics.washington.edu/phylip.html>. The wolf population was designated as the outgroup in order to root the tree. Wolves from eight different countries were combined into one population for simplicity on the tree shown in figure 2. When taken as individuals, all wolves split off from a single branch, which falls in the same place as the root (16). The splitting order in the phylogenetic analysis was not correlated with heterozygosity (Table S4), and the twelve breeds that split off first closely mirrored the first cluster identified by *structure*. These observations argue that the analysis identified a distinct subgroup of genetically related breeds, rather than splitting off idiosyncratic breeds that are unusually inbred or that recently mixed with wild canids.

Supporting References

1. K. E. Comstock *et al.*, *Mol Ecol* **11**, 2489 (2002).
2. R. Guyon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5296 (2003).
3. B. Ewing, P. Green, *Genome Res* **8**, 186 (1998).
4. B. Ewing, L. Hillier, M. C. Wendl, P. Green, *Genome Res* **8**, 175 (1998).
5. D. Gordon, C. Abajian, P. Green, *Genome Res* **8**, 195 (1998).
6. D. A. Nickerson, V. O. Tobe, S. L. Taylor, *Nucleic Acids Res* **25**, 2745 (1997).
7. F. Tajima, *Genetics* **123**, 585 (November, 1989).
8. P. O. Lewis, and D. Zaykin. 2001.,
<http://lewis.eeb.uconn.edu/lewishome/software.html> (2001).
9. J. K. Pritchard, M. Stephens, N. A. Rosenberg, P. Donnelly, *Am J Hum Genet* **67**, 170 (2000).
10. D. Falush, M. Stephens, J. K. Pritchard, *Genetics* **164**, 1567 (August, 2003).
11. N. A. Rosenberg *et al.*, *Science* **298**, 2381 (2002).
12. N. A. Rosenberg *et al.*, *Genetics* **159**, 699 (2001).
13. M. A. Beaumont, R. A. Nichols, *Proceedings: Biological Sciences* **263**, 1619 (December 22, 1996).
14. E. Minch, A. Ruiz-Linares, D. Goldstein, M. Feldman, L. L. Cavalli-Sforza. (1995, 1996).
15. J. Felsenstein, *Cladistics* **5**, 164 (1989).
16. H. G. Parker. *unpublished data*.

Supplementary Table 1. 75 SNPs found in 120 dogs representing 60 breeds. SNP name includes the name of the BAC from which sequence was obtained, the end of the BAC (either S or T) and the base position from the 5 prime end of the forward primer. BAC end sequences and primers are listed online at

http://www.fhrc.org/science/dog_genome/dog.html.

SNP	major allele	minor allele	minor allele frequency	heterozygosity
372c5t_82	C	T	0.004	0.009
372c15t_285	G	A	0.013	0.025
372e2s_271	G	T	0.029	0.057
372e2s_257	C	T	0.071	0.132
372e2s_128	C	G	0.046	0.087
372e2s_93	C	G	0.021	0.041
372e13t_57	T	C	0.004	0.008
372e15t_258	C	T	0.009	0.018
372e18t_165	G	C	0.254	0.379
372g17t_66	T	A	0.134	0.232
372i23s_384	A	G	0.312	0.429
372m6t_138	C	A	0.275	0.399
372m6t_88	T	C	0.004	0.009
372m9t_108	A	T	0.368	0.465
372m9t_58	G	C	0.362	0.462
372m18t_129	G	A	0.159	0.267
372m23t_76	C	T	0.017	0.034

372m23t_108	G	A	0.081	0.149
372m23t_229	G	A	0.078	0.143
372m23t_238	T	C	0.078	0.143
372m23t_263	A	G	0.157	0.265
372o13s_212	T	C	0.316	0.433
373a10s_274	T	C	0.131	0.228
373a15t_112	G	A	0.004	0.008
373a17t_136	A	G	0.394	0.477
373a21s_89	C	T	0.017	0.034
373c13s_93	C	T	0.028	0.054
373c15t_242	C	T	0.209	0.331
373c15t_202	C	T	0.174	0.288
373e1t_50	T	C	0.009	0.019
373e1t_130	G	A	0.010	0.020
373e21t_282	A	G	0.049	0.093
373e21t_116	C	T	0.215	0.338
373g7t_243	C	T	0.014	0.028
373g19t_246	G	A	0.004	0.008
373g19t_378	A	C	0.082	0.150
373i8s_199	A	C	0.073	0.136
373i8s_224	G	A	0.004	0.009
373i16s_312	A	G	0.078	0.144
373i16s_254	G	A	0.240	0.365

373i16s_250	C	T	0.079	0.146
373i16s_249	C	T	0.031	0.060
373k8s_181	C	T	0.005	0.009
372c5s_112	A	G	0.357	0.459
372c5s_168	A	G	0.010	0.020
372c15s_121	T	C	0.017	0.034
372c15s_196	G	A	0.004	0.009
372e15s_67	A	G	0.186	0.303
372e15s_71	A	C	0.013	0.026
372e15s_165	G	A	0.105	0.188
372e15s_221	C	A	0.189	0.307
372i23t_97	A	G	0.119	0.210
372m6s_67	A	G	0.323	0.437
372m6s_73	A	C	0.042	0.081
372m6s_100	T	C	0.042	0.081
372m6s_147	T	G	0.349	0.454
372m6s_186	A	G	0.008	0.017
372m7t_100	C	A	0.101	0.181
372m7t_273	A	G	0.051	0.097
372m18s_131	T	C	0.339	0.448
373a14t_290	T	C	0.224	0.347
373a14t_197	C	T	0.225	0.349
373a14t_160	A	T	0.441	0.493

373e21s_136	C	T	0.332	0.443
373e21s_175	C	T	0.332	0.443
373e21s_191	G	C	0.330	0.442
373a21t_93	A	G	0.008	0.017
373i16t_47	G	G	0.457	0.496
373i16t_210	G	A	0.147	0.250
373i16t_302	C	T	0.476	0.499
373i16t_319	C	A	0.381	0.472
373g7s_263	C	T	0.204	0.325
373g7s_266	T	C	0.201	0.321
373k10t_261	A	C	0.353	0.457
373k10t_264	T	C	0.008	0.017

Supplementary Table 2: Average membership coefficient for each breed from the K=4 cluster results shown in figure 3.

Breed	Number of Individuals	Inferred Clusters			
		Yellow	Blue	Green	Red
Shiba Inu	5	0.974	0.007	0.010	0.009
Chow Chow	5	0.983	0.006	0.005	0.006
Akita	5	0.977	0.005	0.013	0.006
Alaskan Malamute	5	0.884	0.029	0.023	0.064
Basenji	5	0.925	0.030	0.012	0.033
Chinese Shar-Pei	5	0.894	0.050	0.029	0.027
Siberian Husky	5	0.828	0.021	0.071	0.080
Afghan Hound	5	0.634	0.041	0.068	0.256
Saluki	5	0.392	0.041	0.058	0.509
Tibetan Terrier	5	0.368	0.120	0.141	0.371
Lhasa Apso	5	0.402	0.030	0.444	0.125
Samoyed	5	0.404	0.017	0.501	0.078
Pekingese	5	0.210	0.026	0.603	0.161
Shih Tzu	5	0.199	0.026	0.616	0.159
Irish Wolfhound	5	0.011	0.165	0.650	0.173
Saint Bernard	5	0.016	0.201	0.557	0.226
Greyhound	5	0.017	0.091	0.740	0.152
Belgian Sheepdog	5	0.013	0.009	0.962	0.016
Belgian Tervuren	4	0.018	0.022	0.856	0.103
Borzoi	5	0.041	0.024	0.720	0.215
Collie	5	0.007	0.019	0.766	0.208
Shetland Sheepdog	5	0.017	0.105	0.684	0.193
Pug Dog	5	0.022	0.017	0.466	0.494
Komondor	5	0.039	0.101	0.206	0.653
Whippet	5	0.007	0.087	0.480	0.426
Standard Poodle	5	0.032	0.144	0.370	0.454
Bichon Frise	4	0.074	0.087	0.362	0.477
Keeshond	5	0.016	0.043	0.479	0.462
Manchester Terrier, Toy	4	0.024	0.161	0.303	0.513
Norwegian Elkhound	5	0.104	0.090	0.329	0.477
Kuvasz	5	0.077	0.043	0.378	0.502
Great Dane	5	0.067	0.085	0.240	0.608
Welsh Springer Spaniel	5	0.007	0.083	0.255	0.654
Doberman Pinscher	5	0.015	0.103	0.194	0.688

Standard Schnauzer	5	0.006	0.149	0.165	0.681
Italian Greyhound	5	0.074	0.068	0.096	0.762
Old English Sheepdog	5	0.024	0.086	0.122	0.768
American Water Spaniel	5	0.023	0.127	0.131	0.719
Miniature Schnauzer	5	0.009	0.136	0.129	0.726
Australian Terrier	5	0.022	0.107	0.104	0.767
English Cocker Spaniel	5	0.004	0.088	0.182	0.725
Irish Setter	5	0.005	0.074	0.117	0.804
West Highland White Terrier	5	0.019	0.079	0.058	0.844
Pointer	5	0.019	0.067	0.105	0.809
Basset Hound	4	0.020	0.086	0.077	0.818
Cavalier King Charles Spaniel	5	0.013	0.078	0.122	0.787
Giant Schnauzer	5	0.106	0.082	0.060	0.752
Pharaoh Hound	4	0.102	0.081	0.025	0.792
Golden Retriever	5	0.009	0.184	0.019	0.789
Beagle	5	0.016	0.175	0.058	0.751
Bloodhound	5	0.009	0.203	0.014	0.775
Airedale Terrier	4	0.016	0.127	0.109	0.748
American Cocker Spaniel	5	0.010	0.103	0.053	0.834
American Hairless Rat Terrier	5	0.009	0.149	0.064	0.778
Chesapeake Bay Retriever	5	0.019	0.173	0.032	0.776
Cairn Terrier	5	0.015	0.123	0.073	0.790
Portuguese Water Dog	5	0.007	0.134	0.139	0.720
German Shorthaired Pointer	5	0.015	0.172	0.094	0.719
Border Collie	5	0.037	0.116	0.101	0.746
Bedlington Terrier	4	0.010	0.233	0.145	0.613
Clumber Spaniel	5	0.005	0.355	0.066	0.573
Ibizan Hound	5	0.015	0.149	0.120	0.716
Rhodesian Ridgeback	5	0.010	0.215	0.150	0.625
Dachshund	5	0.015	0.315	0.192	0.479
Australian Shepherd	5	0.068	0.221	0.170	0.540
Chihuahua	5	0.028	0.229	0.161	0.582
Kerry Blue Terrier	5	0.008	0.257	0.147	0.588
Schipperke	4	0.011	0.195	0.078	0.717
Irish Terrier	4	0.009	0.277	0.070	0.644
Flat-coated Retriever	5	0.005	0.207	0.084	0.704
Soft Coated Wheaten Terrier	4	0.035	0.329	0.163	0.473
Pomeranian	5	0.055	0.340	0.203	0.402
Labrador Retriever	5	0.033	0.488	0.075	0.404
Presa Canario	5	0.036	0.762	0.044	0.158

Rottweiler	5	0.006	0.798	0.098	0.098
Bullmastiff	5	0.008	0.873	0.032	0.087
Newfoundland	5	0.020	0.923	0.018	0.040
German Shepherd Dog	5	0.006	0.858	0.090	0.046
French Bulldog	4	0.009	0.945	0.012	0.034
Miniature Bull Terrier	5	0.013	0.921	0.020	0.047
Bulldog	5	0.008	0.962	0.019	0.011
Boxer	5	0.003	0.923	0.065	0.008
Mastiff	5	0.010	0.934	0.032	0.024
Bernese Mountain Dog	5	0.006	0.708	0.229	0.057
Greater Swiss Mountain Dog	5	0.015	0.488	0.373	0.124

Supplementary Table 3: 96 microsatellite markers on 38 canine autosomes

Marker Number	Marker Name	Repeat number	Chromosome	Forward Primer	Reverse Primer	Product Size	Annealing Temperature	Published PIC	References
1	REN285G14	2	1	AACATGGTGACGAGAAGGCT	CCATTGCGCCACAAGTAGGTT	134	55	NA	Breen et al., 2001 (Genome Res. 11, 1784-1795)
2	C01.673	2	1	CCACCACTTACATAAACAATGGC	TTCCAGACTGCTGCCCTCC	162	58	0.36	Breen et al., 2001 (Genome Res. 11, 1784-1795)
3	REN112I02	2	1	ATAGCCCATGAAAATCCA	CCCAAATACATCCCTACAT	250	58	0.76	Breen et al., 2001 (Genome Res. 11, 1784-1795)
4	REN172C02	2	1	CCCTTAGGAGGAGGCAAGAC	CGTGTCTTTGTTATGGCTTGA	344	55	0.48	Breen et al., 2001 (Genome Res. 11, 1784-1795)
5	FH2793	2	1	CTATGTGCACGCTGAGAGAG	TACCCATAAAGTTGGGCTTG	208	58	0.76	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
6	REN143K19	2	1	GTCCAGAGCCCCAGAGACAAG	CTTTCTCCGCCACTACCTG	138	55	0.5	Breen et al., 2001 (Genome Res. 11, 1784-1795)
7	FH2890	2	2	CCAGATTAACCAGGATGAGG	AATGGCAAGGATGCTACTCC	198	55	0.59	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
8	C02.466	2	2	TCTGGATTGTGGTCACAACC	ACTGGACACTTCTTTTCAGACG	147	58	0.55	Breen et al., 2001 (Genome Res. 11, 1784-1795)
9	C02.894	2	2	TCAGCATCTAGAAAATTAGGT	ACTCATTTTCTCTTATTCTGCAG	141	58	0.72	Breen et al., 2001 (Genome Res. 11, 1784-1795)
11	FH2895	2	3	TGGGATACACTTAGAAAACAGG	AAAACCCACAACAAGTGC	378	58	0.7	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
12	REN157C08	2	3	GCAAAATGGCAAGATTTTCGTT	CAAGGGTTAGCACCTGGGTA	239	55	0.72	Breen et al., 2001 (Genome Res. 11, 1784-1795)
13	C03.445	2	3	TC1TTGGTAAAGTCTCCATGGG	GGGTGTGAAAACAGCCAACT	175	58	0.6	Breen et al., 2001 (Genome Res. 11, 1784-1795)
14	FH2732	2	4	GAAAGGGTGTGGATAAAGAGC	TGGCATATTAACAATAATGTC	215	58	0.84	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
15	FH2776	2	4	GGAACAGATGAGAAGCATGG	CTGGGTGGTTTCAGTAGTTGG	226	58	0.49	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
16	REN160J02	2	4	ACCGTGCACAAAGTCAGTCAG	AAGCTGAGCCATTCTTTTCC	248	58	0.82	Breen et al., 2001 (Genome Res. 11, 1784-1795)
17	REN262N08	2	4	AGGATTTTGTGGTGATTGGG	TGCGAGCCTATTGTGGACTTT	136	55	NA	Breen et al., 2001 (Genome Res. 11, 1784-1795)
18	REN92G21	2	5	TACTCGTCTGTGCAATTCT	CACACCAGCTCGTCCCTATAA	188	58	0.66	Breen et al., 2001 (Genome Res. 11, 1784-1795)
19	REN285I23	2	5	CAGCTGTCCGGGGATATAAA	GGCTGTGGTTTGTCTCTGTT	244	55	0.58	Breen et al., 2001 (Genome Res. 11, 1784-1795)
20	C05.414	2	5	GCCCTATTTTTTTTTTTCC	TTGGGTTTTCACACTCAGCAG	156	58	0.47	Breen et al., 2001 (Genome Res. 11, 1784-1795)
21	FH2752	2	6	TTTAGCATTGGAGGAGGAG	GGCACAGAATCCAACCTTGAG	201	58	0.38	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
22	REN210I14	2	6	CTGCTCTCTCCCAAACCTTA	CAGGGCAATTGGTCTAGAAA	267	55	0.66	Breen et al., 2001 (Genome Res. 11, 1784-1795)
23	REN37H09	2	6	ATTCCCTTGTATTGTCTCA	CCCCAAAAATCCAACCA	214	58	0.67	Jouquand*, Priat* et al., 2000 (Animal Genetics 31, 266-272)
24	REN97M11	2	7	TGAGGCTCCGTGGGTATGTG	AGGCCAAGGGAATGATGCTC	195	55	NA	Breen et al., 2001 (Genome Res. 11, 1784-1795)
25	REN286L19	2	7	CAGTAATGTTTTGTGGCCTCTG	GGGAGTGGGGGAAATAAATG	214	58	0.66	Breen et al., 2001 (Genome Res. 11, 1784-1795)
26	FH2860	2	7	TGCCCTCAACAATAATGTC	ATCATCTAGCACTCAGAAGG	171	55	0.62	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
27	REN204K13	2	8	TCGGGATGTTTCTCTCCAC	CTGTGTAATTTCTCCAGCG	249	55	0.48	Breen et al., 2001 (Genome Res. 11, 1784-1795)
28	C08.373	2	8	TATTTAAAAATCCCAGGCACA	AGCATCAATTAGATGTCAGCG	114	58	0.68	Breen et al., 2001 (Genome Res. 11, 1784-1795)
29	C08.618	2	8	CAACCCAGGTGGGAAGC	TAGCAAGAAAATGTGCCCA	199	55	0.82	Breen et al., 2001 (Genome Res. 11, 1784-1795)
30	C09.173	2	9	ATCCAGGCTGGAATACCCC	TCCTTTGAATTAGCACTTGGC	110	58	0.78	Breen et al., 2001 (Genome Res. 11, 1784-1795)
31	C09.474	2	9	TTAAGCCTTATTTGTGTGGG	TCCAGGAAGTGTCTGCAGG	109	55	0.78	Breen et al., 2001 (Genome Res. 11, 1784-1795)
32	FH2885	2	9	CTTTTAGGGTGCCTTCAACC	TGGATTATTAAGGGGAATTTAGC	209	55	0.74	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
33	C10.781	2	10	ACCTCCAAGATGGCTCTTGA	ACGTCGAGCTCTGGCAT	190	55	0.62	Breen et al., 2001 (Genome Res. 11, 1784-1795)
34	REN73F08	2	10	ATTCCCCAGCGTACFACCA	AACAGCAATTCAGACAGAGG	200	55	0.54	Breen et al., 2001 (Genome Res. 11, 1784-1795)
35	REN154G10	2	10	TGGAGCTAAGCCTGACTTT	AAACGGAGTTCCATCTCTGG	247	55	0.71	Breen et al., 2001 (Genome Res. 11, 1784-1795)
36	REN164B05	2	11	TGGGATGTGTGCATGTGTG	CCTATGCAGGGTAGGCACAT	232	55	0.5	Breen et al., 2001 (Genome Res. 11, 1784-1795)
37	FH2874	2	11	AACTTCTGGCTTTCATACC	TCTTCAAAGAAAACCAACAGG	178	55	NA	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
38	C11.873	2	11	CTGGCAGATTACAGGTAGC	GTCTCCAAAGCACTCAT	133	58	0.81	Breen et al., 2001 (Genome Res. 11, 1784-1795)
40	REN213F01	2	12	CATTCTGGGCAGGTTTCAIT	GGGGACAATAATCCACTCTC	265	55	0.82	Breen et al., 2001 (Genome Res. 11, 1784-1795)
41	REN208M20	2	12	TGCATCATTGTGGGTGTTT	ATGGAGCATGGGTGAGAAAT	318	58	0.64	Breen et al., 2001 (Genome Res. 11, 1784-1795)
42	REN94K11	2	12	ATACCACTGTCCCTCCTCTTA	ATCCTGGTATCAAATCTATCA	123	55	0.56	Breen et al., 2001 (Genome Res. 11, 1784-1795)
44	REN286P03	2	13	GCACATTCACAAAAGTGGTG	GCAATGGAAAAGAGGATGGAA	331	58	0.78	Breen et al., 2001 (Genome Res. 11, 1784-1795)
45	C13.758	2	13	AAGCATCCAGAATCCCTGG	GTGTATTGGGAGATAATCCACA	228	55	0.75	Breen et al., 2001 (Genome Res. 11, 1784-1795)
46	C14.866	2	14	TGTCATAATAGTTGGAATGAC	TTAGAGCTTACTCATGATATCTG	242	55	0.74	Breen et al., 2001 (Genome Res. 11, 1784-1795)
47	FH3072	2	14	CCTGGATTAACAGTTGTCTGG	CCTGGATTATAAGCATGAGAGC	197	55	0.63	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
48	FH3802	2	15	TTTTACCAGCTCTGAGATAGC	TTTCTGAGCCACTTTCCATAG	265	55	0.44	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
49	REN06C11	2	15	TGCAGGGCAGAGGCTGGAGG	GGGGTGTGCGGTGGAGTTCT	89	58	0.79	Jouquand*, Priat* et al., 2000 (Animal Genetics 31, 266-272)
50	REN144M10	2	15	TCTCATGCAAATCTTCGTTC	GCCACCTCATTCCAAAAGA	211	58	0.66	Breen et al., 2001 (Genome Res. 11, 1784-1795)
51	REN85N14	2	16	AAGGCAGGAGGAGGCAC	TATGGAGATGGAGGGCACAC	239	58	0.78	Breen et al., 2001 (Genome Res. 11, 1784-1795)
52	FH3096	2	16	TCAGCTCAGGGAGTGATCC	ATGAGGAGGTGCAACTATCC	209	55	0.79	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
53	C17.402	2	17	AAATGGGTAATTCATCCAGTGC	CAGGCTTTGTTGAGGTGTCA	175	58	0.75	Breen et al., 2001 (Genome Res. 11, 1784-1795)
54	REN50B03	2	17	ATGTGGGGTAAAATAAAGG	AACACTGACATGCATCCAC	193	58	0.74	Jouquand*, Priat* et al., 2000 (Animal Genetics 31, 266-272)
55	REN112G10	2	17	GAGGGATGGCTGTCAAGA	TGGTGGATAAATAGATAAGGA	172	55	0.7	Breen et al., 2001 (Genome Res. 11, 1784-1795)
56	REN186N13	2	18	AGGCTCTCCGAGGGTAAGAC	ATTGCTTGATAAAGAGGGGG	369	58	0.66	Breen et al., 2001 (Genome Res. 11, 1784-1795)
57	FH2795	2	18	GGCTGGACTTTTGTCAATTTG	TGTGAGTAGGGTAGGGCAAG	204	58	0.71	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
58	C18.460	2	18	CTTCCCAATATAGCCCTGCC	GGTGTCCAGAAAATGAGACCA	128	58	0.53	Breen et al., 2001 (Genome Res. 11, 1784-1795)
59	FH2783	2	19	TCCCTCCGACACTGCTTTA	AGGCTGTGTTTCTCTCT	197	55	NA	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
60	REN911I4	2	19	CCCTTTTGCACACTACTGTGAG	TCCCTTTTGTGGCTGAA	149	58	0.72	Breen et al., 2001 (Genome Res. 11, 1784-1795)

61	REN274F18	2	19	TTCTCGTGACCCCTAAAGGA	GGCATAAATTGTCTTTGCC	196	58	0.66	Breen et al., 2001 (Genome Res. 11, 1784-1795)
62	FH2887	2	20	AAACAAGACAGTAGGAAGAGAGG	CAAGATTCAAACAAGCAACC	268	55	0.77	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
63	FH3109	2	20	TGGGAGCTTCATACAAATGC	CCATGGAGAGTGGTTATTGC	206	58	0.62	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
64	REN293N22	2	20	AATTACTCGGCCCTCTCTGGG	AATGACATTGAGCCTGGGAA	206	58	0.48	Breen et al., 2001 (Genome Res. 11, 1784-1795)
65	FH2914	2	21	GTGATCCACTTGTGTATCC	ATAGCCTTGGGAATTTTTC	199	55	0.61	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
66	FH3069	2	21	CATGCCTGACTCAACTGATG	TATGGACCTTCGTTTCAGAGG	161	55	0.53	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
67	REN49F22	2	22	GGGGCTCTGTATTAGGTG	TCATAAGGCAAAGAAAACC	157	55	0.66	Jouquand*, Priat* et al., 2000 (Animal Genetics 31, 266-272)
68	REN107H05	2	22	TGACTTACTTAGCCACTTTT	TTGTCCCTTGATAACTGATG	120	55	0.86	Breen et al., 2001 (Genome Res. 11, 1784-1795)
69	REN78I16	2	22	TTACATTTAGGGGCTCCAGT	TTAGACAAAATAGGCTTCAA	184	55	0.63	Breen et al., 2001 (Genome Res. 11, 1784-1795)
70	FH3078	2	23	GCCTTCTTTGGAAAAACACC	TTCAGGGAATTCCTTTCTGG	199	55	0.67	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
71	C23.277	2	23	ACACATTTGTGTCTTGTCTTG	TTGTATGGAGGTGGGAGAG	145	55	0.54	Breen et al., 2001 (Genome Res. 11, 1784-1795)
72	REN181K04	2	23	ACAAGCCGACTCTAGCGAAA	AGATGGGGCCTAACCAAAGT	216	58	0.64	Breen et al., 2001 (Genome Res. 11, 1784-1795)
73	REN106I06	2	24	GGTCCCATCTCAAAATCCTC	CCCTTCTGGCCTCCTACACA	250	55	0.58	Breen et al., 2001 (Genome Res. 11, 1784-1795)
74	FH3083	2	24	ATTTGCCAGGTACCATTCC	CAGGTTATTCTGGGCTATGG	216	55	0.61	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
75	REN54E19	2	25	GCAGACGAGCACCCGAA	TATCCACATCATTACC	173	55	0.54	Breen et al., 2001 (Genome Res. 11, 1784-1795)
77	REN87O21	2	26	TGGCTGTGGCTAAGGCTTTGT	CCCCITCCAGCTTCGGTGTAG	240	55	0.62	Breen et al., 2001 (Genome Res. 11, 1784-1795)
78	C26.733	2	26	CCCTCTACTTATGTCTCGGCC	GAGAGGAGAAAACAACCAACC	255	55	0.61	Breen et al., 2001 (Genome Res. 11, 1784-1795)
79	C27.442	2	27	CCAAGAACAGCCTAAGCTGG	ACACATACAGCCCAATTCA	164	55	0.74	Breen et al., 2001 (Genome Res. 11, 1784-1795)
80	C27.436	2	27	CAGCTGGATTGGGGACTC	CATCTTGTCTCTCAAC	190	55	0.51	Breen et al., 2001 (Genome Res. 11, 1784-1795)
81	REN72K15	2	27	CCGATGCCTGTCTTTGA	ACCTGGCCTCTCTGTGTCT	250	55	0.66	Breen et al., 2001 (Genome Res. 11, 1784-1795)
82	FH2759	2	28	AGTACTTGAGGCTTGGAGTCAG	CAAGCTGAGAGCCATGTAGG	197	55	0.71	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
83	FH2785	2	28	ATGGCAGGTCAAGAGTATGG	GATAGATCCAAGCCAACACC	316	55	0.46	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
84	REN239K24	2	29	TTCAITGGCTGGTGACTTTG	CCCAGGACCATTGTAGA	307	55	0.78	Breen et al., 2001 (Genome Res. 11, 1784-1795)
85	FH3082	2	29	AGTAAAGGGTCTCAAGTGTGC	CATTTGTCAITGTGGAAAACC	216	55	0.54	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
86	REN51C16	2	30	CAGTTTATCCTTCCCCTCTC	GTGCTAGTCTGGCTGTGCTCA	258	55	0.8	Mellersh et al., 2000 (Mammalian Genome 11, 120-130)
87	FH3053	2	30	GATTAAGGGCAAGCAACC	TCATCTCCAGCTTTTCATGG	193	55	0.74	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
88	REN43H24	2	31	CAGTGAGCAAAGCAAAATGAA	ATGTGAACCCCGCCAATA	181	55	0.66	Jouquand*, Priat* et al., 2000 (Animal Genetics 31, 266-272)
89	FH2712	2	31	AAGGTAGTCCCACGATCCTC	GAGCCCTGTTCTCAGGTTG	186	55	0.67	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
90	FH2875	2	32	TGATACCCATTAAGTCCATCC	CAATACCCTGATACCAAAACC	209	55	0.6	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
91	FH2790	2	33	CCAATATTGTTAAGAAGTTCAAGC	AGGCCTTCTCTGTCTCTTG	216	55	0.58	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
92	REN291M20	2	33	CCTCTAGATCCATCCATATTGTCA	TGTCCACCCACAGATGAATG	155	58	0.76	Breen et al., 2001 (Genome Res. 11, 1784-1795)
93	REN160M18	2	34	AATGGTGGTGATATTCACAGAGA	GGATGCTTGGGAATCTTGAA	101	58	0.76	Breen et al., 2001 (Genome Res. 11, 1784-1795)
94	FH3060	2	34	AGGCAGAGCTAAACCTGAGC	TACAGGCACCTCCTCCTACG	193	55	0.4	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
95	REN314H10	2	34	GAAATGTTCATTTCTGTGACAT	AGGTTTGGGCTCCTCATCT	172	55	0.54	Breen et al., 2001 (Genome Res. 11, 1784-1795)
96	REN01G01	2	35	TACATCTCCACATCTACTGA	GAGACTTAACACAGTATTTG	224	55	0.54	Jouquand*, Priat* et al., 2000 (Animal Genetics 31, 266-272)
97	REN112C08	2	35	ATGGCCCACCGATACACA	TCGGGGACATACTTGAACC	222	55	0.42	Breen et al., 2001 (Genome Res. 11, 1784-1795)
98	REN106I07	2	36	TTCCCAAGCCACACC	AACCACTATCCAACCTTAT	211	55	0.78	Breen et al., 2001 (Genome Res. 11, 1784-1795)
99	FH2708	2	37	CAGCAATTGGACAAGAAAAAG	GAAAGAGGATGAAGGGTGTG	204	55	0.63	Guyon et al., 2003 (PNAS 100{9}, 5296-5301)
100	REN86G15	2	38	CAGCGAATTTGGGCACTAA	CAACTAAGGCAGAGAATACCA	157	55	0.76	Breen et al., 2001 (Genome Res. 11, 1784-1795)

Supplementary Table 4: Heterozygosity of 85 dog breeds. The expected heterozygosity of each breed was averaged across 96 microsatellites using the Tajima's unbiased estimator (7).

Population	Heterozygosity
Bedlington Terrier	0.312842
Miniature Bull Terrier	0.321619
Boxer	0.343151
Clumber Spaniel	0.363595
Greater Swiss Mountain Dog	0.364943
Airedale Terrier	0.372793
Soft Coated Wheaten Terrier	0.37376
Collie	0.383453
Doberman Pinscher	0.383763
Irish Terrier	0.390427
Bloodhound	0.391559
German Shepherd Dog	0.397957
Pug Dog	0.398442
Bernese Mountain Dog	0.399599
Flat-coated Retriever	0.402832
Miniature Schnauzer	0.414528
Irish Wolfhound	0.418039
Pharaoh Hound	0.420188
Cavalier King Charles Spaniel	0.427633
Shetland Sheepdog	0.43244
Manchester Terrier - Toy	0.432937
French Bulldog	0.439855
Basset Hound	0.441171
American Cocker Spaniel	0.443841
Schipperke	0.445437
Irish Setter	0.446656
Basenji	0.447739
Bulldog	0.449549
Standard Schnauzer	0.450041
Whippet	0.450959
American Hairless Terrier	0.454113
Mastiff	0.455126
Rottweiler	0.45651
Pekingese	0.459983
English Cocker Spaniel	0.46565
Saint Bernard	0.465724
Italian Greyhound	0.468797
Afghan Hound	0.468924
Pointer	0.469444
Shih Tzu	0.472193
Welsh Springer Spaniel	0.473917
Kerry Blue Terrier	0.477836

Dachshund	0.483817
Borzoi	0.487909
Great Dane	0.488697
Alaskan Malamute	0.489877
Newfoundland	0.490617
West Highland White Terrier	0.493936
Belgian Sheepdog	0.495114
Australian Terrier	0.499343
Ibizan Hound	0.503981
Keeshond	0.505126
Bullmastiff	0.509243
Akita	0.510396
Greyhound	0.513409
Chesapeake Bay Retriever	0.514166
Golden Retriever	0.517779
Tibetan Terrier	0.519535
Chow Chow	0.52043
Rhodesian Ridgeback	0.520493
Siberian Husky	0.527344
Bichon Frise	0.528271
Standard Poodle	0.529948
Old English sheepdog	0.530192
Norwegian Elkhound	0.532854
German Shorthaired Pointer	0.538761
American Water Spaniel	0.540183
Lhasa Apso	0.541245
Samoyed	0.542932
Pomeranian	0.546007
Beagle	0.549119
Border Collie	0.549583
Belgian Tervuren	0.551091
Kuvasz	0.553538
Shiba Inu	0.560543
Labrador Retriever	0.56059
Giant Schnauzer	0.56131
Saluki	0.563037
Portuguese Water Dog	0.568882
Komondor	0.57321
Cairn Terrier	0.575823
Chinese Shar-Pei	0.584412
Perro de Presa Canario	0.589397
Chihuahua	0.592353
Australian Shepherd	0.609668

Supplementary Table 5. Expected heterozygosity of 60 breeds based on allele frequencies at 75 SNP loci. Each breed is represented by two dogs. Heterozygosity calculated as described in Supplementary Table 4.

Breed	Heterozygosity ($\times 10^{-4}$)
Scottish Deerhound	2.0683
Field Spaniel	2.3165
Flat-coated Retriever	2.6474
Bernese Mountain Dog	2.8129
Standard Schnauzer	2.8129
Boxer	3.0611
Collie	3.0611
Bearded Collie	3.1438
Miniature Bull Terrier	3.2266
Perro de Presa Canario	3.392
Bull Terrier	3.8057
Mastiff	3.8057
Petite Basset Griffon Vendéen	3.8884
Bedlington Terrier	3.9712
Saluki	4.1366
Standard Poodle	4.1366
Cavalier King Charles Spaniel	4.2194
Sussex Spaniel	4.2194
American Water Spaniel	4.5503
Ibizan Hound	4.7158
Beagle	4.7985
Boston Terrier	4.7985
German Pinscher	4.8812
Basset Hound	4.964
Bichon Frise	4.964
Rottweiler	4.964
Bullmastiff	5.1294
English Springer Spaniel	5.1294
Greater Swiss Mountain Dog	5.3776
Pug Dog	5.3776
Boykin Spaniel	5.5431
Italian Greyhound	5.5431
Newfoundland	5.5431
American Hairless Terrier	5.7086
Borzoi	5.7913
German Shepherd Dog	5.7913

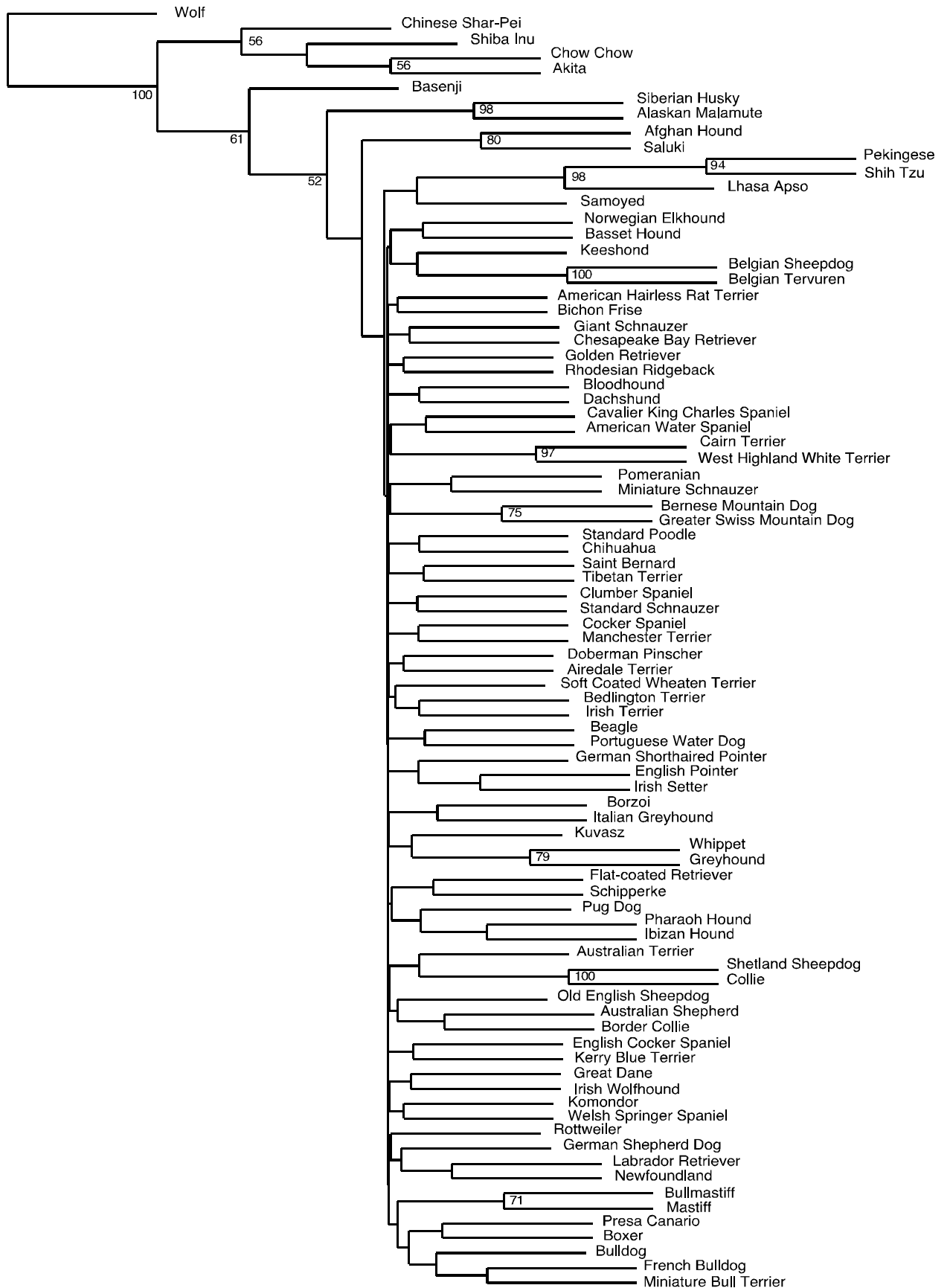
Saint Bernard	5.7913
Dachshund	5.874
Akita	5.9568
Cocker Spaniel	6.0395
French Bulldog	6.0395
Greyhound	6.0395
Irish Water Spaniel	6.0395
Shetland Sheepdog	6.205
Papillon	6.2877
Foxhound (English)	6.3704
Tibetan Terrier	6.4532
Welsh Springer Spaniel	6.4532
German Shorthaired Pointer	6.6186
Welsh Terrier	6.6186
Dalmatian	6.7014
Irish Setter	6.7014
Alaskan Malamute	6.8668
Golden Retriever	7.0323
Portugese Water Dog	7.115
Weimaraner	7.6942
Labrador Retriever	8.4388
Spinoni Italiano	8.9352
Chesapeake Bay Retriever	9.1006
English Shepherd	9.2661

Supporting Figure Legends

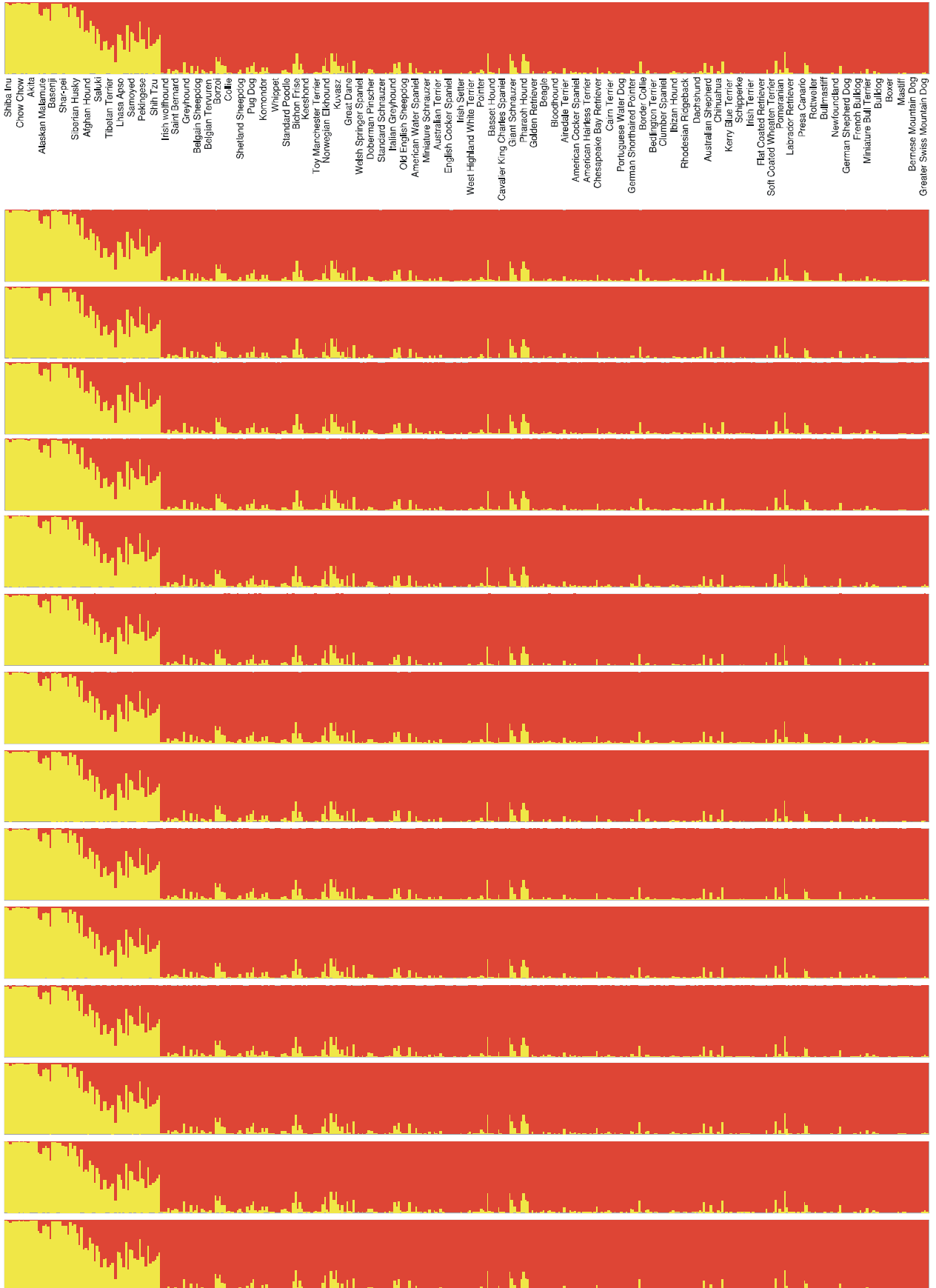
Supplementary Figure 1. Consensus neighbor-joining tree of 85 dog breeds and the gray wolf. The trees that formed the consensus are based on the chord distance measure. 500 bootstrap replicates of the data were carried out, and the fraction of bootstraps supporting each branch is indicated at the corresponding node as a percentage. Only those branches supported in over 50% of the replicates are labeled. The wolf population at the root of the tree consists of 8 individuals, one from each of the following countries: China, Oman, Iran, Sweden, Italy, Mexico, Canada and the United States. Branch lengths are proportional to bootstrap values.

Supplementary Figure 2: Fifteen independent runs of structure at each value of K from 2 to 5. Breeds are ordered as in figure 3 and are listed below the first run. All subsequent runs are ordered the same. Individual dogs are represented by a thin line divided into 2(A), 3(B), 4(C), or 5(D) colors representing the fraction of their genome assigned to each cluster. Fifteen runs at K=2 (A) show almost identical results. At K=3 (B) the only differences between runs involve three breeds: Bloodhound, Clumber Spaniel and Dachshund. At K=4 (C) variation between runs increases as suggested by the drop in the similarity coefficient though the core breeds in each group remain consistent. At K=5 (D) each run depicts a different clustering solution with only the first two subpopulations (yellow and blue) remaining consistent.

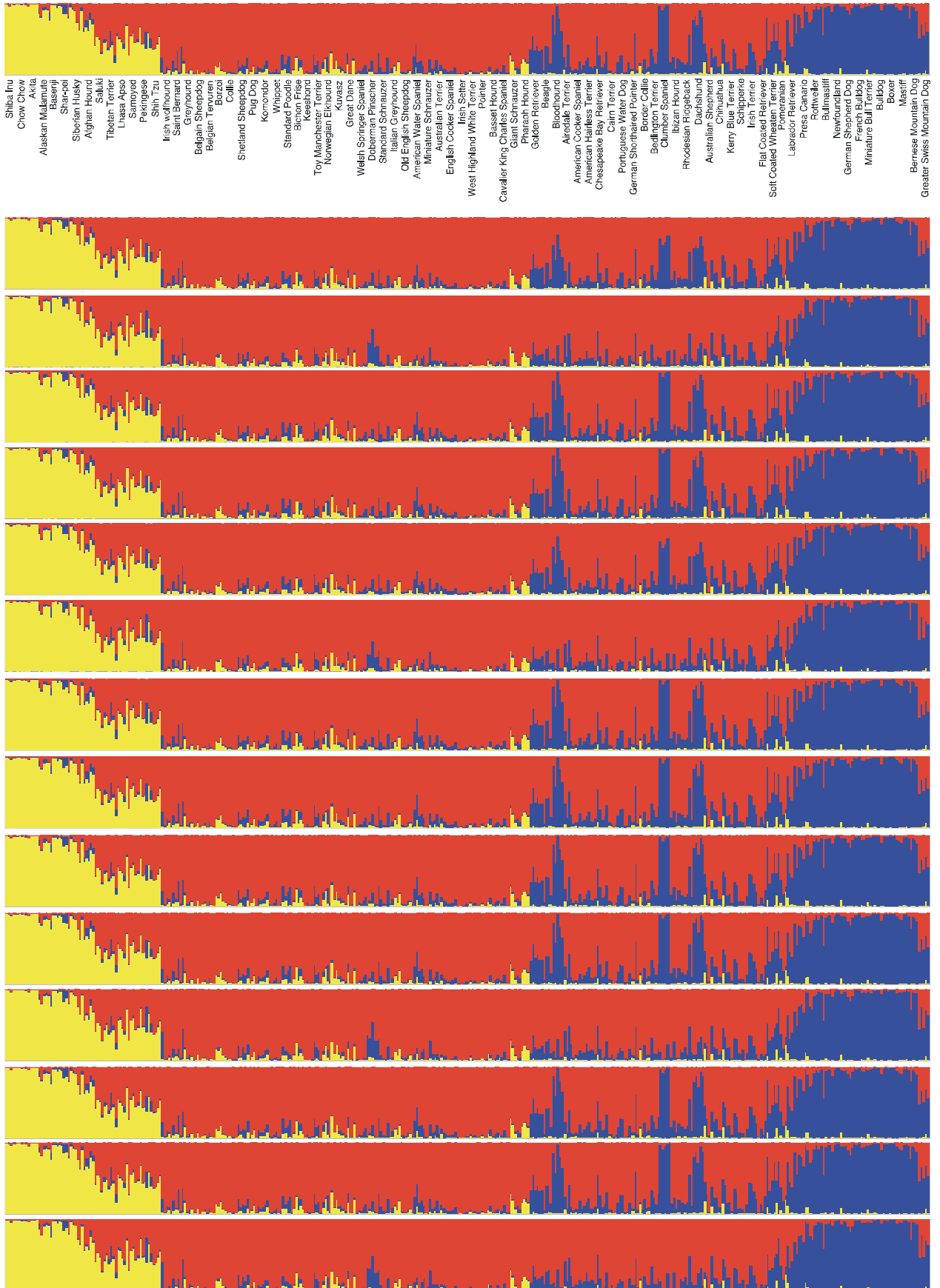
Supplementary Figure 1.



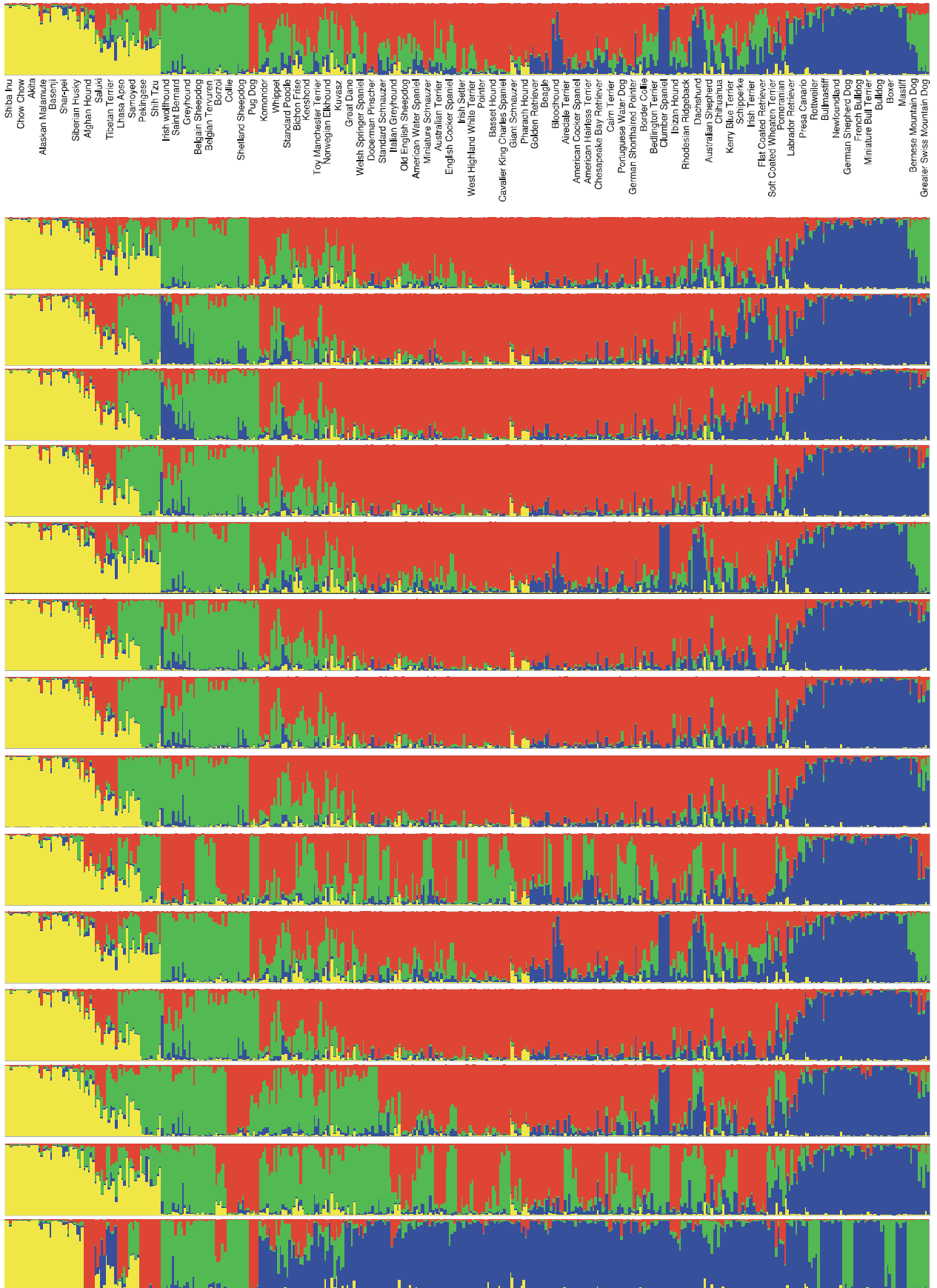
Supplementary Figure 2A



Supplementary Figure 2B



Supplementary Figure 2C



Supplementary Figure 2D

