

# Genetic Structure of the Purebred Domestic Dog

by Heidi G. Parker et.al.

---

Heidi G. Parker,<sup>1,2,3</sup> Lisa V. Kim,<sup>1,2,4</sup> Nathan B. Sutter,<sup>1,2</sup> Scott Carlson,<sup>1</sup> Travis D. Lorentzen,<sup>1,2</sup> Tiffany B. Malek,<sup>1,3</sup> Gary S. Johnson,<sup>5</sup> Hawkins B. DeFrance,<sup>1,2</sup> Elaine A. Ostrander,<sup>1,2,3,4\*</sup> Leonid Kruglyak<sup>1,3,4,6</sup>

We used molecular markers to study genetic relationships in a diverse collection of 85 domestic dog breeds. Differences among breeds accounted for 30% of genetic variation. Microsatellite genotypes were used to correctly assign 99% of individual dogs to breeds. Phylogenetic analysis separated several breeds with ancient origins from the remaining breeds with modern European origins. We identified four genetic clusters, which predominantly contained breeds with similar geographic origin, morphology, or role in human activities. These results provide a genetic classification of dog breeds and will aid studies of the genetics of phenotypic breed differences.

The domestic dog is a genetic enterprise unique in human history. No other mammal has enjoyed such a close association with humans over so many centuries, nor been so substantially shaped as a result. A variety of dog morphologies have existed for millennia, and reproductive isolation between them was formalized with the advent of breed clubs and breed standards in the mid-19th century. Since that time, the promulgation of the "breed barrier" rule - no dog may become a registered member of a breed unless both its dam and sire are registered members - has ensured a relatively closed genetic pool among dogs of each breed. At present, there are more than 400 described breeds, 152 of which are recognized by the American Kennel Club (AKC) in the United States. (1). Over 350 inherited disorders have been described in the purebred dog population (2). Many of these mimic common human disorders and are restricted to particular breeds or groups of breeds as a result of aggressive in-breeding programs used to generate specific morphologies. We have previously argued that mapping genes associated with common diseases, including cancer, heart disease, epilepsy, blindness, and deafness, as well as genes underlying the striking diversity among breeds in morphology and behavior, will best be accomplished through elucidating and taking advantage of the population structure of modern breeds. (3). Understanding the genetic relationships among breeds will also provide insight into the directed evolution of our closest animal companions.

Mitochondrial DNA analyses have been used to elucidate the relationship between the domestic dog and the wolf (4-6), but the evolution of mitochondrial DNA is too slow to allow inferences about relationships among modern dog breeds, most of which have existed for fewer than 400 years (1, 7, 8). One previous study showed that nuclear microsatellite loci could be used to assigning dogs from five breeds to their breed of origin, demonstrating large genetic distances among these breeds (9). Another study used microsatellites to detect the relatedness of two breed pairs in a collection of 28 breeds but could not establish broader phylogenetic relationships among the breeds (10). The failure to find such relationships could reflect the properties of microsatellite loci (10), the limited number of breeds examined, or the analytical methods used in the study. Alternatively, it may reflect the complex structure in purebred dog populations, resulting from the recent origin of most breeds and the mixing of ancestral types in their creation. Here, we show that microsatellite typing of a diverse collection of 85 breeds, combined with phylogenetic analysis and modern genetic clustering methods (11, 12), allows the definition of related groups of breeds and that genetic relatedness among breeds often correlates with morphological similarity and shared geographic origin.

To assess the amount of sequence variation in purebred dogs, we first resequenced 19,867 base pairs of noncontiguous genomic sequence in 120 dogs representing 60 breeds. We identified 75 single nucleotide polymorphisms (SNPs), with monor allele frequencies ranging from 0.4 to 48% (table S1). Fourteen of the SNPs were breed specific. When all dogs were considered as a single population, the observed nucleotide heterozygosity (13) was  $8 \times 10^{-4}$  (to the minus 4), essentially the same as that found in the human population (14, 15).

To further characterize genetic variation within and among breeds, we genotyped 96 microsatellite loci in 414 purebred dogs representing 85 breeds (five unrelated dogs that lacked any common grandparents

were sampled from most breeds; table S2). We predicted that, because of the existence of breed barriers, dogs from the same breed would be more similar genetically than dogs from different breeds. To test this prediction, we estimated the proportion of genetic variation among individual dogs that could be attributed to breed membership. An analysis of molecular variance (16) in the microsatellite data showed that variation among breeds accounts for more than 27% of total genetic variation. Similarly, the average genetic distance between breeds calculated from the SNP data is  $F_{st} = 0.33$ . These observations are consistent with previous reports that analyzed fewer dog breeds (9, 10), confirming the prediction that breed barriers have led to strong genetic isolation among breeds, and are in marked contrast to the much lower genetic differentiation (typically in the range of 5 to 10%) found among human populations (17, 18). Variation among breeds in dogs is on the high end of the range reported for domestic livestock populations (19, 20).

Strong genetic differentiation among dog breeds suggests that breed membership could be determined from individual dog genotypes (9). To test this hypothesis, we first applied a Bayesian model-based clustering algorithm, implemented in the program *structure* (11, 12, 21), to the microsatellite data. The algorithm attempts to identify genetically distinct sub-populations on the basis of patterns of allele frequencies. We applied *structure* to overlapping subsets of 20 to 22 breeds at a time (22) and observed that most breeds formed distinct clusters consisting solely of all the dogs from that breed (Fig. 1A). Dogs in only four breeds failed to consistently cluster with others of the same breed: Perro de Presa Canario, German Shorthaired Pointer, Australian Shepherd, and Chihuahua. In addition, six pairs of breeds clustered together in the majority of runs. These pairings - Alaskan Malamute and Siberian Husky, Belgian Sheepdog and Belgian Tervuren, Collie and Shetland Sheepdog, Greyhound and Whippet, Bernese Mountain Dog and Greater Swiss Mountain Dog, and Bullmastiff and Mastiff - are all expected on the basis of known breed history. To test whether these closely related breed pairs were nonetheless genetically distinct, we applied *structure* to each of these clusters. In all but one case, the clusters separated into two populations corresponding to the individual breeds (Fig. 1B). The single exception was the cluster containing Belgian Sheepdogs and Belgian Tervurens. The European and Japanese Kennel Clubs classify these as coat color and length varieties a single breed (23, 24), and although the AKC recognizes them as distinct breeds, the breed barrier is apparently too recent or insufficiently strict to have resulted in genetic differentiation.

To further characterize genetic variation within and among breeds, we genotyped 96 microsatellite loci in 414 purebred dogs representing 85 breeds (five unrelated dogs that lacked any common grandparents were sampled from most breeds; table S2). We predicted that, because of the existence of breed barriers, dogs from the same breed would be more similar genetically than dogs from different breeds. To test this prediction, we estimated the proportion of genetic variation

among individual dogs that could be attributed to breed membership. An analysis of molecular variance (16) in the microsatellite data showed that variation among breeds accounts for more than 27% of total genetic variation. Similarly, the average genetic distance between breeds calculated from the SNP data is  $F_{ST} = 0.33$ . These observations are consistent with previous reports that analyzed fewer dog breeds (9, 10), confirming the prediction that breed barriers have led to strong genetic isolation among breeds, and are in marked contrast to the much lower genetic differentiation (typically in the range of 5 to 10%) found among human populations (17, 18). Variation among breeds in dogs is on the high end of the range reported for domestic livestock populations (19, 20). Strong genetic differentiation among dog breeds suggests that breed membership could be determined from individual dog genotypes (9). To test this hypothesis, we first applied a Bayesian model-based clustering algorithm, implemented in the program *structure* (11, 12, 21), to the microsatellite data. The algorithm attempts to identify genetically distinct subpopulations on the basis of patterns of allele frequencies. We applied *structure* to overlapping subsets of 20 to 22 breeds at a time (22) and observed that most breeds formed distinct clusters consisting solely of all the dogs from

that breed (Fig. 1A). Dogs in only four breeds failed to consistently cluster with others of the same breed: Perro de Presa Canario, German Shorthaired Pointer, Australian Shepherd, and Chihuahua. In addition, six pairs of breeds clustered together in the majority of runs. These pairings—Alaskan Malamute and Siberian Husky, Belgian Sheepdog and Belgian Tervuren, Collie and Shetland Sheepdog, Greyhound and Whippet, Bernese Mountain Dog and Greater Swiss Mountain Dog, and Bullmastiff and Mastiff—are all expected on the basis of known breed history. To test whether these closely related breed pairs were nonetheless genetically distinct, we applied *structure* to each of these clusters. In all but one case, the clusters separated into two populations corresponding to the individual breeds (Fig. 1B). The single exception was the cluster containing Belgian Sheepdogs and Belgian Tervurens. The European and Japanese Kennel Clubs classify these as coat color and length varieties of a single breed (23, 24), and although the AKC recognizes them as distinct breeds, the breed barrier is apparently too recent or insufficiently strict to have resulted in genetic differentiation.

We next examined whether a dog could be assigned to its breed on the basis of genotype data alone. Using the direct assignment method (25) with a leave-one-out analysis, we were able to assign 99% of individual dogs to the

correct breed. Only 4 dogs out of 414 were assigned incorrectly: one Beagle as a Perro de Presa Canario, one Chihuahua as a Cairn Terrier, and two German Shorthaired Pointers as a Kuvasz and a Standard Poodle. All four errors involved breeds that did not form single-breed clusters in the *structure* analysis.

Having demonstrated that modern dog breeds are distinct genetic units, we next sought to define broader genetic relationships among the breeds. We first used standard neighbor-joining methods to build a majorityrule consensus tree of breeds (Fig. 2), with distances calculated using the chord distance measure (26), which does not assume a particular mutation model and is thought to perform well for closely related taxa (27). The tree was rooted using wolf samples. The deepest split in the tree separated four Asian spitz-type breeds, and within this branch the Shar-Pei split first, followed by the Shiba Inu, with the Akita and Chow Chow grouping together. The second split separated the Basenji, an ancient African breed. The third split separated two Arctic spitz-type breeds, the Alaskan Malamute and Siberian Husky, and the fourth split separated two Middle Eastern sight hounds, the Afghan and Saluki, from the remaining breeds.

The first four splits exceeded the majorityrule criterion, appearing in more than half of the

bootstrap replicates. In contrast, the remaining breeds showed few consistent phylogenetic relationships, except for close groupings of five breed pairs that also clustered together in the *structure* analysis, one new pairing of the closely related West Highland White Terrier and Cairn Terrier, and the significant grouping of three Asian companion breeds of similar appearance, the Lhasa Apso, Shih Tzu, and Pekingese (fig. S1). A close relationship among these three breeds was also observed in the *structure* analysis, with at least two of the three clustering together in a majority of runs. The flat topology of the tree likely reflects a largely common founder stock and occurrence of extensive gene flow between phenotypically dissimilar dogs before the advent of breed clubs and breed barrier rules. In addition, it probably reflects the fact that some historically older breeds that died out during the famines, depressions, and wars of the 19th and 20th centuries have been recreated with

1

**Fig.1.** Clustering assignment of 85 dog breeds. **(A)** Seventy-four breeds are represented by five unrelated dogs each, and the remaining 11 breeds are represented by four unrelated dogs each. Each individual dog is represented on the graph by a vertical line divided into colored segments corresponding to different genetic clusters. The length of each colored segment is equal to the estimated proportion of the individual's membership in the cluster of corresponding color (designated on the y axis as a percentage). Breeds are labeled below the figure. **(B)** Six clusters

containing two breeds each are subdivided at  $K = 2$ , with colors representing the estimated proportion of individual membership in only two possible clusters. Black lines separate individual dogs and the two breeds are labeled below the figures.

2

**Fig.2.** Consensus neighbor-joining tree of 85 dog breeds and the gray wolf. Nine breeds that form branches with statistical support are shown. The remaining 76 breeds show little phylogenetic structure and have been combined into one branch labeled “All other breeds” for simplification. The entire tree is shown in Fig. S1. The trees that formed the consensus are based on the chord distance measure. Five hundred 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 for those branches supported in more than 50% of the replicates. The wolf population at the root of the tree consists of eight 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.

continues today). Second, methods based on genetic distance matrices lose information by collapsing all genotype data for pairs of breeds into a single number. The clustering algorithm implemented in *structure* was explicitly designed to overcome these limitations (11, 12, 28) and has been applied to infer the genetic structure of several species (17, 28, 29). We therefore ran *structure* on the entire data set using increasing values of  $K$  (the number of subpopulations the program attempts to find) to identify ancestral source populations. In this analysis, a modern breed could closely mirror a

single ancestral population or represent a mixture of two or more ancestral types.

At  $K = 2$ , one cluster was anchored by the first seven breeds to split in the phylogenetic analysis, whereas the other cluster contained the large number of breeds with a flat phylogenetic topology (Fig. 3A). Five runs of the program produced nearly identical results, with a similarity coefficient ( $I7$ ) of 0.99 across runs. Seven other breeds share a sizeable fraction of their ancestry with the first cluster. These fourteen breeds all date to antiquity and trace their ancestry to Asia or Africa. When a diverse set of wolves from eight different countries was included in the

3

**Fig.3. (A)** Population structure of 85 domestic dog breeds. Each individual dog is represented by a single vertical line divided into  $K$  colors, where  $K$  is the number of clusters assumed. Each color represents one cluster, and the length of the colored segment shows the individual's estimated proportion of membership in that cluster. Black lines separate the breeds that are labeled below the figure. Representative breeds pictured above the graph from left to right: Akita, Pekingese, Belgian Sheepdog, Collie, Doberman Pinscher, Basset Hound, American Cocker Spaniel, Bedlington Terrier, Flat-Coated Retriever, Newfoundland, and Mastiff. Results shown are averages over 15 *structure* runs at each value of  $K$ . **(B)** Population structure, as in (A), but with gray wolves included. Graph shown is



averaged over *ve structure* runs at  $K = 2$ .

analysis, they fell entirely within this cluster (Fig. 3B). The branch leading to the wolf outgroup also fell within this group of breeds in the phylogenetic analysis (Fig. 2).

At  $K = 3$ , additional structure was detected that was not readily apparent from the phylogenetic tree. The new third cluster consisted primarily of breeds related in heritage and appearance to the Mastiff and is anchored by the Mastiff, Bulldog, and Boxer, along with their close relatives, the Bullmastiff, French Bulldog, Miniature Bull Terrier, and Perro de Presa Canario. Also included in the cluster are the Rottweiler, Newfoundland, and Bernese Mountain Dog, large breeds that are reported to have gained their size from ancient Mastiff-type ancestors.

Less expected is the inclusion of the German Shepherd Dog. The exact origins of this breed are unknown, but our results suggest that the years spent as a military and police dog in the presence of working dog types, such as the Boxer, are responsible for shaping the genetic background of this popular breed. Three other breeds showed partial and inconsistent membership in this cluster across *structure* runs (fig. S2), which lowered the similarity coefficient to 0.84.

At  $K = 4$ , a fourth cluster was observed, which included several breeds used as herding

dogs: Belgian Sheepdog, Belgian Tervuren, Collie, and Shetland Sheepdog. The Irish Wolfhound, Greyhound, Borzoi, and Saint Bernard were also frequently assigned to this cluster. Although historical records do not suggest that these dogs were ever used to herd livestock, our results suggest that these breeds are either progenitors to or descendants of herding types. The breeds in the remaining cluster are primarily of relatively recent European origins and are mainly different types of hunting dogs: scent hounds, terriers, spaniels, pointers, and retrievers.

Clustering at  $K = 4$  showed a similarity coefficient of 0.61, reflecting similar cluster membership assignments for most breeds but variable assignments for other breeds across runs (fig. S2). At  $K = 5$ , the similarity coefficient dropped to 0.26 and no additional consistent subpopulations were inferred, suggesting a lack of additional high-level substructure in the sampled purebred dog population.

Our results paint the following picture of the relationships among domestic dog breeds. Different breeds are genetically distinct, and individuals can be readily assigned to breeds on the basis of their genotypes. This level of divergence is surprising given the short time since the origin of most breeds from mixed ancestral stocks and supports strong reproductive isolation within each breed as a result of the breed barrier rule. Our results support at least four distinct breed

groupings representing separate “adaptive radiations.”

A subset of breeds with ancient Asian and African origins splits off from the rest of the breeds and shows shared patterns of allele frequencies.

At first glance, it is surprising that a single genetic cluster includes breeds from Central Africa (Basenji), the Middle East (Saluki and Afghan), Tibet (Tibetan Terrier and Lhasa Apso), China (Chow Chow, Pekingese, Shar-Pei, and Shi Tzu), Japan (Akita and Shiba Inu), and the Arctic (Alaskan Malamute, Siberian Husky, and Samoyed). However, several researchers

have hypothesized that early pariah dogs originated in Asia and migrated with nomadic human groups both south to Africa and north to the Arctic, with subsequent migrations occurring throughout Asia (5, 6, 30). This cluster includes Nordic breeds that phenotypically resemble the wolf, such as the Alaskan Malamute and Siberian Husky, and shows the closest genetic relationship to the wolf, which is the direct

ancestor of domestic dogs. Thus, dogs from these breeds may be the best living representatives of the ancestral dog gene pool. It is notable that several breeds commonly believed to be of ancient origin, such as the Pharaoh Hound and Ibizan Hound, are not included in this group. These are often thought to be the oldest of all dog breeds, descending directly from the ancient

Egyptian dogs drawn on tomb walls more than 5000 years ago. Our results indicate, however, that these two breeds have been recreated in more recent times from combinations of other breeds. Thus, although their appearance matches the ancient Egyptian sight hounds, their genomes do not. Similar conclusions apply to the Norwegian Elkhound, which clusters with modern European breeds rather than with the other Arctic dogs, despite reports of direct descent from Scandinavian origins more than 5000 years ago (*1, 24*).

The large majority of breeds appears to represent a more recent radiation from shared European stock. Although the individual breeds are genetically differentiated, they appear to have diverged at essentially the same time. This radiation probably reflects the proliferation of distinct breeds from less codified phenotypic varieties after the introduction of the breed concept and the creation of breed clubs in Europe in the 1800s. A more sensitive cluster analysis was able to discern additional genetic structure of three subpopulations within this group. One contains Mastiff-like breeds and appears to reflect shared morphology derived from a common ancestor.

Another includes Shetland Sheepdog, the two Belgian Sheepdogs, and Collie, and may reflect shared ancestral herding behavior. The remaining population is dominated by a proliferation of breeds dedicated to various aspects of

the hunt. For these breeds, historical and breed club records suggest highly intertwined bloodlines, consistent with our results.

Dog breeds have traditionally been grouped on the basis of their roles in human activities, physical phenotypes, and historical records. Here, we defined an independent classification based on patterns of genetic variation. This classification supports a subset of traditional groupings and also reveals previously unrecognized connections among breeds. An accurate understanding of the genetic relationships among breeds lays the foundation for studies aimed at uncovering the complex genetic basis of breed differences in morphology, behavior, and disease susceptibility.

#### References and Notes

1. J. Crowley, B. Adelman, Eds., *The Complete Dog Book; Official Publication of the American Kennel Club* (Howell Book House, New York, ed. 19, 1998).
2. D. F. Patterson *et al.*, *J. Am. Vet. Med. Assoc.* **193**, 1131 (1988).
3. E. A. Ostrander, L. Kruglyak, *Genome Res.* **10**, 1271 (2000).
4. C. Vila *et al.*, *Science* **276**, 1687 (1997).
5. P. Savolainen, Y. P. Zhang, J. Luo, J. Lundeberg, T. Leitner, *Science* **298**, 1610 (2002).
6. J. A. Leonard *et al.*, *Science* **298**, 1613 (2002).
7. C. A. Rogers, A. H. Brace, *The International Encyclopedia of Dogs* (Howell Book House, New York, ed. 1, 1995).
8. B. Fogel, *The Encyclopedia of the Dog* (DK Publishing, New York, 1995).
9. M. T. Koskinen, *Anim. Genet.* **34**, 297 (2003).
10. D. N. Irion *et al.*, *J. Hered.* **94**, 81 (2003).

11. J. K. Pritchard, M. Stephens, P. Donnelly, *Genetics* **155**, 945 (2000).
12. D. Falush, M. Stephens, J. K. Pritchard, *Genetics* **164**, 1567 (2003).
13. F. Tajima, M. Nei, *Mol. Biol. Evol.* **1**, 269 (1984).
14. R. Sachidanandam *et al.*, *Nature* **409**, 928 (2001).
15. J. C. Venter *et al.*, *Science* **291**, 1304 (2001).
16. L. Excoffier, P. E. Smouse, J. M. Quattro, *Genetics* **131**, 479 (1992).
17. N. A. Rosenberg *et al.*, *Science* **298**, 2381 (2002).
18. L. L. Cavalli-Sforza, P. Menozzi, A. Piazza, *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, NJ, 1994).
19. D. E. MacHugh, R. T. Loftus, P. Cunningham, D. G. Bradley, *Anim. Genet.* **29**, 333 (1998).
20. G. Laval *et al.*, *Genet. Sel. Evol.* **32**, 187 (2000).
21. J. K. Pritchard, M. Stephens, N. A. Rosenberg, P. Donnelly, *Am. J. Hum. Genet.* **67**, 170 (2000).
22. Materials and methods are available as supporting material on *Science Online*.
23. T. Yamazaki, K. Yamazaki, *Legacy of the Dog: The Ultimate Illustrated Guide to Over 200 Breeds* (Chronicle Books, San Francisco, CA, 1995).
24. B. Wilcox, C. Walkowicz, *Atlas of Dog Breeds of the World* (TFH Publications, Neptune City, NJ, ed. 5, 1995).
25. D. Paetkau, W. Calvert, I. Stirling, C. Strobeck, *Mol. Ecol.* **4**, 347 (1995).
26. L. L. Cavalli-Sforza, A. W. Edwards, *Evolution* **32**, 550 (1967).
27. D. B. Goldstein, A. R. Linares, L. L. Cavalli-Sforza, M. W. Feldman, *Genetics* **139**, 463 (1995).
28. N. A. Rosenberg *et al.*, *Genetics* **159**, 699 (2001).
29. D. Falush *et al.*, *Science* **299**, 1582 (2003).
30. M. V. Sablin, G. A. Khlopachev, *Curr. Anthropol.* **43**, 795 (2002).
31. Supported by the Burroughs Wellcome Innovation Award (E.A.O. and L.K.), grants from the AKC–Canine

Health Foundation (G.S.J.), NIH training grant T32 HG00035 (H.G.P.), and a postdoctoral fellowship from the Waltham Foundation (N.B.S.). E.A.O. also acknowledges support from K05 CA90754. L.K. is a James S. McDonnell Centennial Fellow. We thank the many dog owners, researchers, and breeders who provided DNA samples for this work, especially G. Brewer, C. Gaiser, and K. Murphy; D. Lynch, A. Ziska, C. Ramirez, M. Langlois, and D. Akey for assistance with sample collection; M. Stephens for advice and assistance with the program *structure*; K. Markianos and J. Akey for helpful discussions; M. Eberle for computing assistance; H. Coller, E. Giniger, R. Wayne, and three anonymous reviewers for comments on the manuscript; and the AKC and C. Jierski for use of the canine artwork included in Fig. 3.

#### **Supporting Online Material**

[www.sciencemag.org/cgi/content/full/304/5674/1160/](http://www.sciencemag.org/cgi/content/full/304/5674/1160/)

DC1

Materials and Methods

Figs. S1 and S2

Tables S1 to S5

References

2 March 2004; accepted 21 April 2004

<sup>1</sup>Division of Human Biology, <sup>2</sup>Division of Clinical Research,

Fred Hutchinson Cancer Research Center, Post

Office Box 19024, 1100 Fairview Avenue North, D4-

100, Seattle, WA 98109–1024, USA. <sup>3</sup>Molecular and

Cellular Biology Program, University of Washington, Box

357275, Seattle, WA 98195-7275, USA. ¶Department of

Genome Sciences, University of Washington, Box

351800, Seattle, WA 98195-7275, USA. §Department of

Veterinary Pathobiology, College of Veterinary Medicine,

University of Missouri, Columbia, MO 65211, USA.

¶Howard Hughes Medical Institute, 1100 Fairview Avenue

North, D4-100, Seattle, WA 98109-1024, USA.

\*To whom correspondence should be addressed. Email: [eostrand@fhcrc.org](mailto:eostrand@fhcrc.org)

[eostrand@fhcrc.org](mailto:eostrand@fhcrc.org)

[www.sciencemag.org](http://www.sciencemag.org) SCIENCE VOL 304 21 MAY 2004