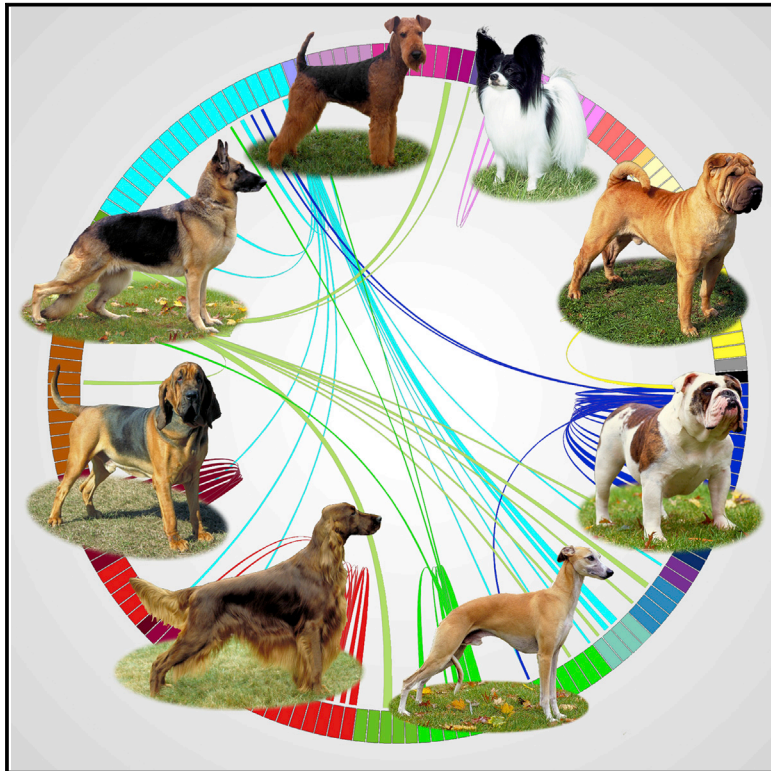


# Cell Reports

## Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development

### Graphical Abstract



### Authors

Heidi G. Parker, Dayna L. Dreger, Maud Rimbault, Brian W. Davis, Alexandra B. Mullen, Gretchen Carpintero-Ramirez, Elaine A. Ostrander

### Correspondence

eostrand@mail.nih.gov

### In Brief

The domestic dog is divided into hundreds of island-like populations called breeds. Parker et al. examine 161 breeds and show that they were developed through division and admixture. The analyses define clades, estimate admixture dates, distinguish geographically diverse populations, and help determine the source of shared mutations among diverse populations.

### Highlights

- Neighbor joining cladogram of 161 breeds establishes 23 supported clades
- Crossing between diverse clades was done recently to add new traits
- Migration of a breed to a new region alters both immigrant and indigenous breeds
- Tracking recent crosses can identify the source of mutations in multiple breeds

### Accession Numbers

GSE90441

GSE83160

GSE70454

GSE96736



# Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development

Heidi G. Parker,<sup>1</sup> Dayna L. Dreger,<sup>1</sup> Maud Rimbault,<sup>1</sup> Brian W. Davis,<sup>1</sup> Alexandra B. Mullen,<sup>1</sup> Gretchen Carpintero-Ramirez,<sup>1</sup> and Elaine A. Ostrander<sup>1,2,\*</sup>

<sup>1</sup>Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

<sup>2</sup>Lead Contact

\*Correspondence: [eostrand@mail.nih.gov](mailto:eostrand@mail.nih.gov)

<http://dx.doi.org/10.1016/j.celrep.2017.03.079>

## SUMMARY

There are nearly 400 modern domestic dog breeds with a unique histories and genetic profiles. To track the genetic signatures of breed development, we have assembled the most diverse dataset of dog breeds, reflecting their extensive phenotypic variation and heritage. Combining genetic distance, migration, and genome-wide haplotype sharing analyses, we uncover geographic patterns of development and independent origins of common traits. Our analyses reveal the hybrid history of breeds and elucidate the effects of immigration, revealing for the first time a suggestion of New World dog within some modern breeds. Finally, we used cladistics and haplotype sharing to show that some common traits have arisen more than once in the history of the dog. These analyses characterize the complexities of breed development, resolving longstanding questions regarding individual breed origination, the effect of migration on geographically distinct breeds, and, by inference, transfer of trait and disease alleles among dog breeds.

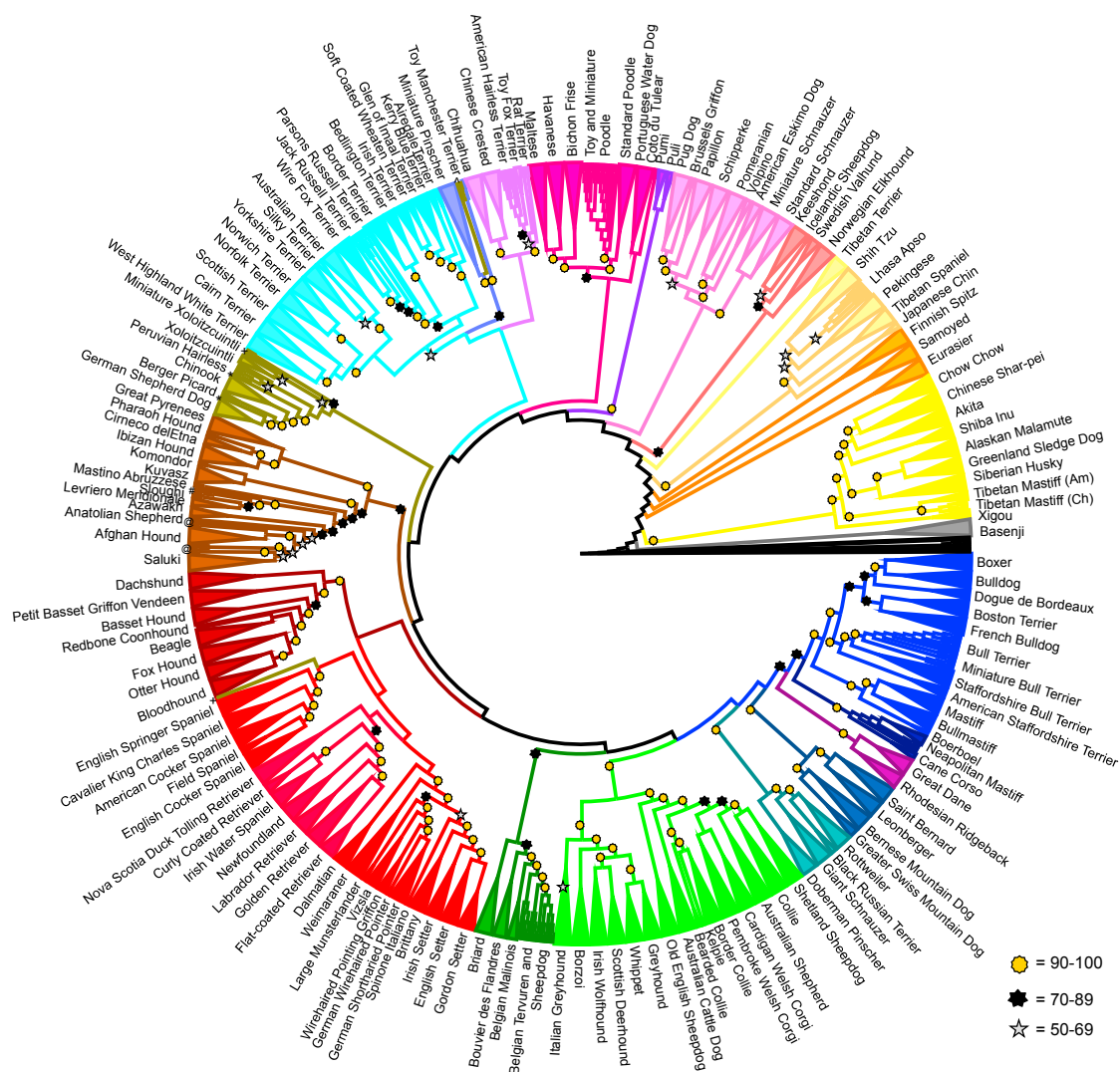
## INTRODUCTION

The dog, *Canis familiaris*, is the first domesticated animal earning a place within nearly every society across the globe for thousands of years (Druzhkova et al., 2013; Thalmann et al., 2013; Vilà et al., 1997, 1999). Over the millennia, dogs have assisted humans with hunting and livestock management, guarding house and field, and played crucial roles in major wars (Moody et al., 2006). Providing a range of services from companionship to production of fur and meat (Wilcox and Walkowicz, 1995), the diversity of talents and phenotypes combined with an unequalled emotional connection between dogs and humans has led to the creation of more than 350 distinct breeds, each of which is a closed breeding population that reflects a collage of defining traits (<http://www.akc.org>).

Previous studies have addressed the genomic makeup of a limited number of breeds, demonstrating that dogs from the same breed share common alleles and can be grouped using measures of population structure (Irion et al., 2003; Koskinen, 2003; Parker et al., 2004), and breeds that possess similar form and function often share similar allelic patterns (Parker et al., 2004, 2007; Vonholdt et al., 2010). However, none of these studies have effectively accounted for the variety of mechanisms through which modern breeds may have developed, such as geographic separation and immigration; the role of hybridization in the history of the breeds; and the timeline of the formation of breeds. In this study, we overcome these barriers by presenting an expansive dataset, including pure breeds sampled from multiple sections of the globe and genotyped on a dense scale. By applying both phylogenetic methods and a genome-wide analysis of recent haplotype sharing, we have unraveled common population confounders for many breeds, leading us to propose a two-step process of breed creation beginning with ancient separation by functional employment followed by recent selection for physical attributes. These data and analyses provide a basis for understanding which and why numerous, sometimes deleterious mutations are shared across seemingly unrelated breeds.

## RESULTS

We examined genomic data from the largest and most diverse group of breeds studied to date, amassing a dataset of 1,346 dogs representing 161 breeds. Included are populations with vastly different breed histories, originating from all continents except Antarctica, and sampled from North America, Europe, Africa, and Asia. We have specifically included breeds that represent the full range of phenotypic variation present among modern dogs, as well as three breeds sampled from both the United States and their country of origin. Samples from 938 dogs representing 127 breeds and nine wild canids were genotyped using the Illumina CanineHD bead array following standard protocols. Data were combined with publically available information from 405 dogs genotyped using the same chip (Hayward et al., 2016; Vaysse et al., 2011). For three dogs from one breed, genotypes were retrieved from publically available sequence



**Figure 1. Cladogram of 161 Domestic Dog Breeds**

Breeds that form unique clades supported by 100% of bootstraps are combined into triangles. For all other branches, a gold star indicates 90% or better, black star 70%–89%, and silver star 50%–69% bootstrap support. Breeds are listed on the perimeter of the circle. A small number of dogs do not cluster with the rest of their breed, indicated as follows: \*cane paratore, +Peruvian hairless dog, #sloughi, @country-of-origin salukis, and ^miniature xoloitzcuintle.

files, and all were merged into a single dataset (Table S1). After pruning for low quality or genotyping rate, 150,067 informative SNPs were retained.

Ascertainment bias has been shown to skew population genetic calculations that require estimation of allele frequencies and diversity measures (Lachance and Tishkoff, 2013). It has also been shown that ascertainment based on a single individual provides less bias than a mixed group (Patterson et al., 2012). The SNPs used in this study were identified primarily within the boxer or from boxer compared to another genome (Vaysse et al., 2011), which has exaggerated the boxer minor allele frequency (MAF; 0.351 in boxer compared to 0.260 overall) but has little effect the other breeds (MAF range, 0.247–0.284). To minimize the effect this might have, we have chosen to use distance measures based on allele sharing rather than frequency

and to enhance these analyses with unbiased haplotype sharing for a robust assessment of canine population structure.

A bootstrapped cladogram was obtained using an identity-by-state distance matrix and a neighbor-joining tree algorithm (Supplemental Experimental Procedures). After 100 bootstraps, 91% of breeds (146/161) formed single, breed-specific nodes with 100% bootstrap support (Figure 1). Of the 15 breeds that did not meet these criteria, seven (Belgian tervuren, Belgian sheepdog, cane corso, bull terrier, miniature bull terrier, rat terrier, and American hairless terrier) were part of two- or three-breed clades that were supported at 98% or greater, and two breeds (Lhasa apso and saluki) formed single-breed clades that were supported at 50% and 78%, respectively. Four breeds (redbone coonhound, sloughi, cane paratore, and Jack Russell terrier) were split within single multi-breed clades, and the last





**Figure 2. Representatives from Each of the 23 Clades of Breeds**

Breeds and clades are listed for each picture from left to right, top to bottom.

- (A) Akita/Asian spitz.
- (B) Shih tzu/Asian toy (by Mary Bloom).
- (C) Icelandic sheepdog/Nordic spitz (by Veronica Druk).
- (D) Miniature schnauzer/schnauzer.
- (E) Pomeranian/small spitz.
- (F) Brussels griffon/toy spitz (by Mary Bloom).
- (G) Puli/Hungarian.
- (H) Standard poodle/poodle.
- (I) Chihuahua/American toy.
- (J) Rat terrier/American terrier (by Stacy Zimmerman).
- (K) Miniature pinscher/pinscher.
- (L) Irish terrier/terrier.
- (M) German shepherd dog/New World (by Mary Bloom).
- (N) Saluki/Mediterranean (by Mary Bloom).
- (O) Basset hound/scent hound (by Mary Bloom).
- (P) American cocker spaniel/spaniel (by Mary Bloom).

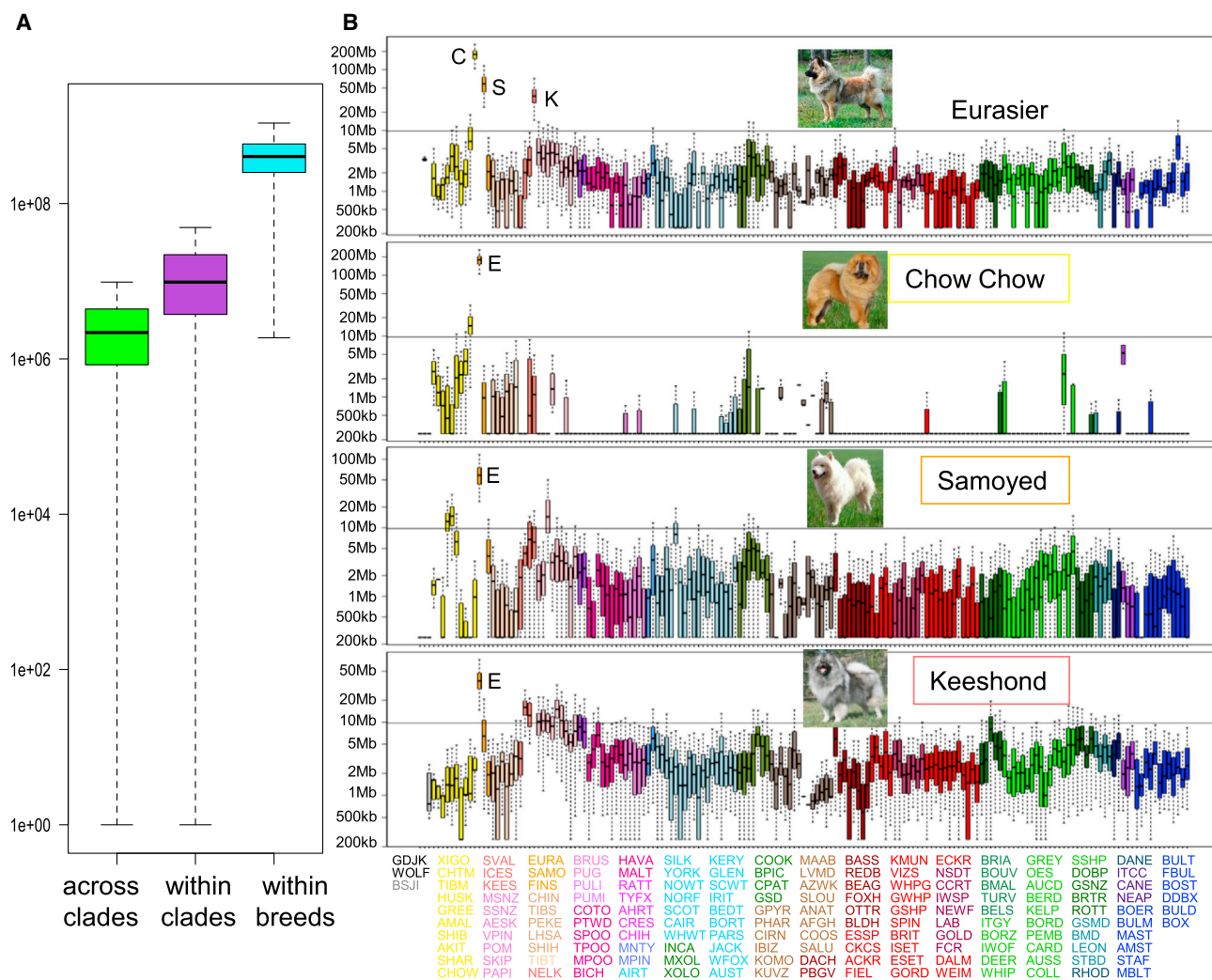
- (Q) Golden retriever/retriever (by Mary Bloom).
- (R) German shorthaired pointer/pointer setter (by Mary Bloom).
- (S) Briard/continental herder (by Mary Bloom).
- (T) Shetland sheepdog/UK rural.
- (U) Rottweiler/drover
- (V) Saint Bernard/alpine.
- (W) English mastiff/European mastiff (by Mary Bloom).

two breeds (xoloitzcuintli and Peruvian hairless dog) were split between divergent clades. Nine of the breeds that were not monophyletic were either newly recognized by the American Kennel Club (AKC) or not recognized at the time of sample collection and likely represent a breed under development. Two other non-monophyletic breeds are composed of dogs collected in two countries; the cane corsos collected in Italy form a fully supported, single clade, as do the salukis collected in the United States. However, the cane corsos collected in the United States form a paraphyletic clade near the Neapolitan mastiffs, and the salukis collected in the Middle East form multiple paraphyletic groups within a clade that includes the US salukis and Afghan hounds.

Not including those that are breed specific, this study defined 105 phylogenetic nodes supported by  $\geq 90\%$  of bootstrap replicates, 133 by  $\geq 70\%$ , and 150 supported by  $\geq 50\%$  of replicates. We identify 29 multi-breed clades that are supported at  $\geq 90\%$ . Each of these clades includes 2–16 breeds and together account for 78% of breeds in the dataset. 150 breeds, or 93% of the dataset, can be divided into 23 clades of 2–18 breeds each, supported at  $>50\%$ . These multi-breed clades reflect common behaviors, physical appearance, and/or related geographic origin (Figure 2).

Eleven breeds did not group with significance to any other breeds. Five breeds form independent clades and six others are paraphyletic to established clades with  $<50\%$  bootstrap support (Table S2). The lack of grouping may indicate that we have not sampled the closest relatives of these breeds or that these breeds comprise outcrossings that are not shared by similar breeds.

To assess hybridization across the clades, identical-by-descent (IBD) haplotype sharing was calculated between all pairs of dogs from the 161 breeds. Haplotypes were phased using the program Beagle (Browning and Browning, 2013) in 100-SNP windows, resulting in a minimum haplotype size of 232 kb, well above the shared background level established in previous studies (Lindblad-Toh et al., 2005; Sutter et al., 2004). The large haplotypes specifically target admixture resulting from breed formation rather than domestication, which previous studies have not addressed. The total length of the shared haplotypes was summed for each pair of dogs. Individuals from within the same breed clade share nearly four times more of their genome within large IBD haplotype blocks than dogs in different breed clades (median shared haplotype lengths of 9,742,000 bp and 2,184,000 bp, respectively;  $p$  [Kolmogorov-Smirnov (K-S) and Wilcoxon]  $< 2.2e^{-16}$ ; Figure 3A). Only 5% of the across-breed pairings have a median greater than 9,744,974 bp. These exceptions argue for recent admixture events between breeds, as evidenced by the example of the Eurasier breed, created in the 1970s by mixing chow chow with other spitz-type breeds (Fogle, 2000) (Figure 3B). The data reveal not only the components of the breed but also the explanation for its placement on the cladogram. The Eurasier (unclustered) shows significant haplotype sharing with the samoyed (unclustered), keeshond (Nordic spitz), and chow chow (Asian spitz) (Figure 3B). Because all three breeds are located in different clades, unrelated to each other, the Eurasier falls between the component breeds and forms its own single-breed clade. Haplotype-sharing bar graphs for each of 161 breeds, including 152 AKC breeds, are available in Data S1. This provides a long-term resource for identifying



**Figure 3. Gross Haplotype Sharing across Breeds**

(A) Boxplot of total haplotype sharing between all pairs of dogs from breeds within the same clade, across different clades, and within the same breed. The difference between the distributions is highly significant ( $p < 2e-16$ ).

(B) Example of haplotype sharing between three breeds (samoyed, chow chow, and keeshond) and a fourth (Eurasier) that was created as a composite of the other three. Combined haplotype length is displayed on the y axis, and 169 breeds and populations are listed on the x axis in the order they appear on the cladogram, starting with the jackal and continuing counterclockwise. Haplotype sharing of zero is set at 250,000 for graphing, a value just below what is detected in this analysis. Breeds are colored by clade. 95% significance level is indicated by the horizontal line. Breed abbreviations are listed under the graph in the order they appear and colored by clade. Definitions of the breed abbreviations can be found in Table S1.

populations that likely share rare and common traits that will be invaluable for mapping the origins of deleterious and beneficial mutations.

Strong evidence of admixture across the clades was found in 117 breeds (Figure 4). A small number of these were identified in previous studies using migration analysis (Pickrell and Pritchard, 2012; Shannon et al., 2015) 30% of these breeds share with only one breed outside their clade. Therefore, more than half (54%) of the breeds that make up the 23 established clades share large haplotypes with one or zero breeds outside their clade, indicating breed creation by selection based on the initial founder population rather than recent admixture. Only 6 of the 161 breeds share extensive haplotypes with many (>8) different

groups, suggesting recent creation of these breeds from multiple others or that they provide a popular modern breed component. The overall low level of sharing across diverse breeds suggests that interclade crosses are done thoughtfully and for specific reasons, such as the introduction of a new trait or the immigration of a breed to a new geographic region.

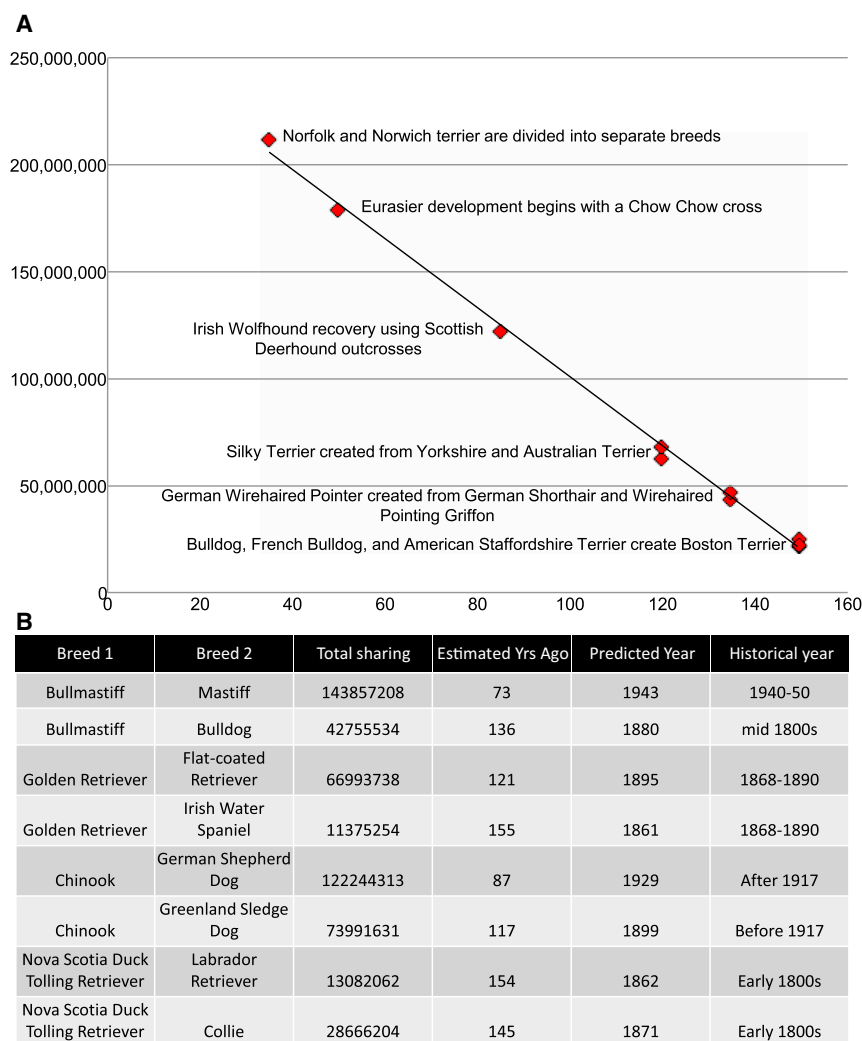
As importation and establishment in a new country has been shown to have a measurable effect on breed structure (Quignon et al., 2007), we chose three breeds, the Tibetan mastiff, saluki, and cane corso, for inclusion in the study, with each collected in the country of origin as well as from established populations in the United States. In each case, there is division of the breed based on collection location. The split between the US and





Chinese Tibetan mastiffs is likely due to independent lineage formation stemming from an importation bottleneck, as is evident from estimations of inbreeding coefficients (Chinese Tibetan mastiffs average  $F = 0.07$ , and US Tibetan mastiffs average  $F = 0.15$ ). Similarly, the average inbreeding coefficient of salukis collected in the United States is twice as high as those sampled from the countries of origin ( $F = 0.21$  and  $0.10$ , respectively). Since the US salukis form a more strongly bootstrapped clade than the country-of-origin dogs, we suggest that there is a less diverse gene pool in the United States. In comparison, the cane corsos from Italy form a single clade, while the cane corsos from the United States cluster with the Neapolitan mastiffs, also collected in the United States. Significant shared haplotypes are observed between the US cane corsos and the rottweiler that are

Our analyses were designed to detect recent admixture; therefore, we were able to identify hybridization events that are described in written breed histories and stud-book records. Using the most reliably dated crosses that produced modern breeds, we established a linear relationship between the total length of haplotype sharing and the age of an admixture event, occurring between 35 and 160 years before present (ybp) (Figure 5A). Applying this equation to the total shared haplotypes calculated from the genotyping data, we have validated this relationship on a second set of recently created breeds arriving at



**Figure 5. Total Haplotype Sharing Is Inversely Correlated with the Time of Hybridization between Breeds that Have Developed within the Last 200 Years**

(A) The time of hybridization in years before present is graphed on the x axis and the median total haplotype sharing on the y axis for six breeds of dog with reliable recent histories of admixture in breed formation or recovery. The trendline shows a linear correlation with  $r^2 = 1$ .

(B) The slope and intercept of the trendline from A was applied to the median haplotype sharing values from the data for four additional breeds with reliable breed creation dates to establish accuracy of estimated hybridization dates.

likelihood score over a zero migration tree (Figure 6A). Only 2 of the 18 clades, New World and Asian toy (Figures 6B and 6C), showed evidence of extensive hybridization between the breeds. Thus, the modern breeds were likely created through selection for unique traits within an ancient breed type with possible admixture from unrelated breeds to enhance the trait.

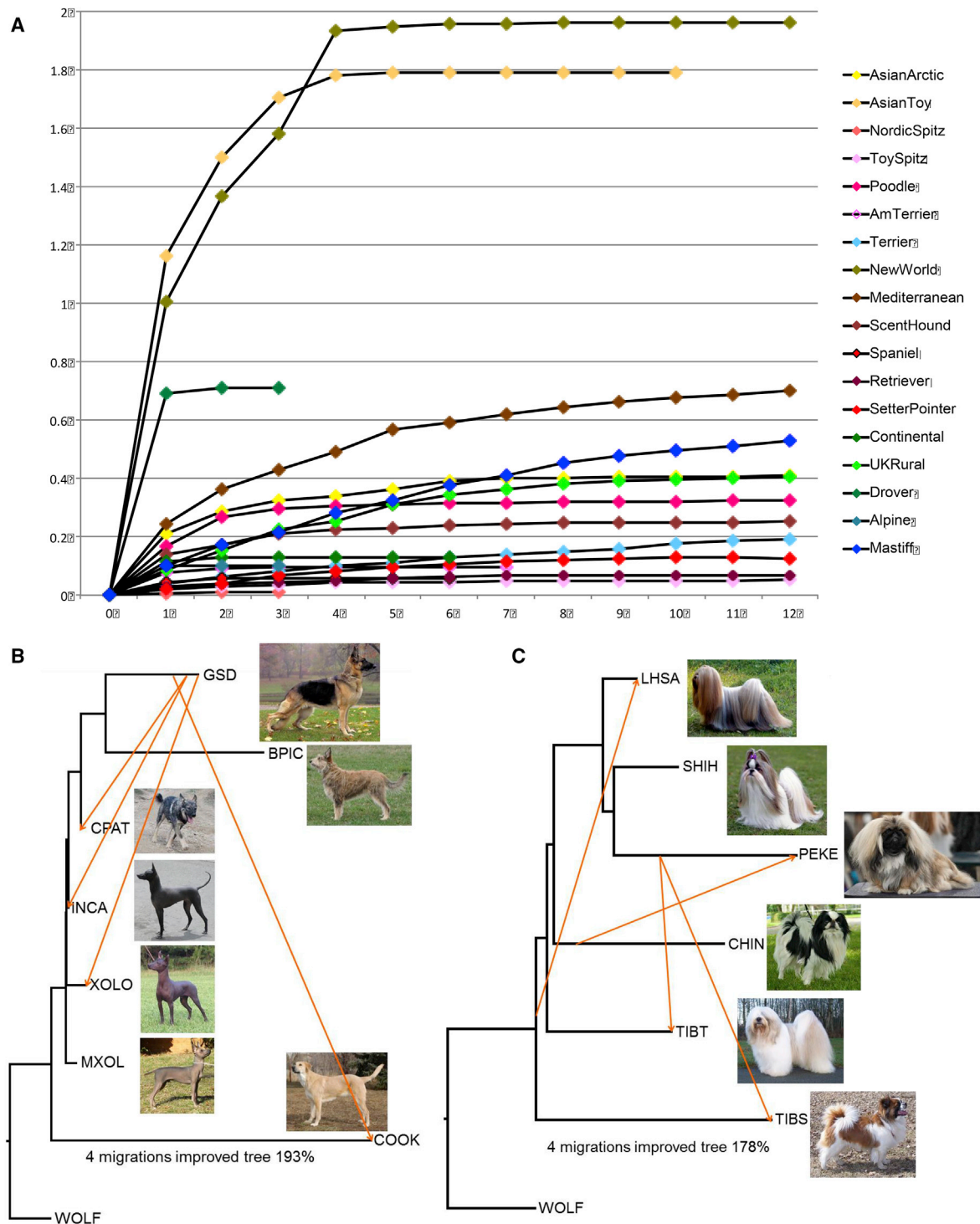
Our hybridization analysis reveals evidence for disease sharing across the clades. For instance, collie eye anomaly (CEA) is a disease that affects the development of the choroid in several herding breeds, including the collie, Border collie, Shetland sheepdog, and Australian shepherd, all members of the UK Rural clade (Lowe et al., 2003). The mutation and haplotype pattern displayed IBD across all affected breeds, and we speculated that all share a common obviously affected ancestor (Parker et al., 2007).

We were unable to explain, however, the presence of the disease in the Nova Scotia duck tolling retriever, a sporting dog developed in Canada from an unknown mixture of local breeds, which also shares the same haplotype. This perplexing observation can now be explained, as this analysis shows that collie and/or Shetland sheepdog were strong, undocumented, contributors to the formation of the Nova Scotia duck tolling retriever and, therefore, the likely source of the CEA mutation within that breed (Figure 7A).

Similarly, a mutation in the *MDR1* gene (multi-drug resistance 1), which causes life-threatening reactions to multiple drugs in many of the UK Rural breeds, has been reported in 10% of German shepherd dogs (Mealey and Meurs, 2008). These data display a link between the German shepherd dog and UK Rural breeds through the Australian shepherd, highlighting the unexpected role the Australian shepherd or its predecessor played in the development of the modern German shepherd dog (Figure 6B). Earlier this year, the *MDR1* mutation was identified in the chinook at a frequency of 15% (Donner et al., 2016). Our analysis reveals recent admixture between this breed and the German shepherd dog as

historically accurate time estimations (Figure 5B). Using the relationship equation  $y = -1,613,084.67x + 262,137,843.89$ , where  $y$  is the total shared haplotype length and  $x$  is the number of years, we can estimate the time at which undocumented crosses or divisions from older breeds took place. For example, based on a median haplotype sharing value of 66,993,738, the golden retriever was separated from the flat-coated retriever in 1895, and the written history of the golden retriever dates to crosses between multiple breeds taking place between 1868 and 1890 (Figure 5B), a near-perfect match.

To determine if the multi-breed clades are formed through recent admixture rather than through common ancestral sources, we examined migration in 18 groups of four or more breeds. These include 16 of the clades established on the tree, including nearby unclustered breeds, and two groups of small clades (American terrier/American toy and small spitz/toy spitz/schnauzer) that are monophyletic, but not well supported. Using the program Treemix (Pickrell and Pritchard, 2012), and allowing 0–12 predicted migration events, we determined the effect of admixture on clade formation by calculating the increase in maximum



**Figure 6. Assessment of Migration between Breeds within Clades**

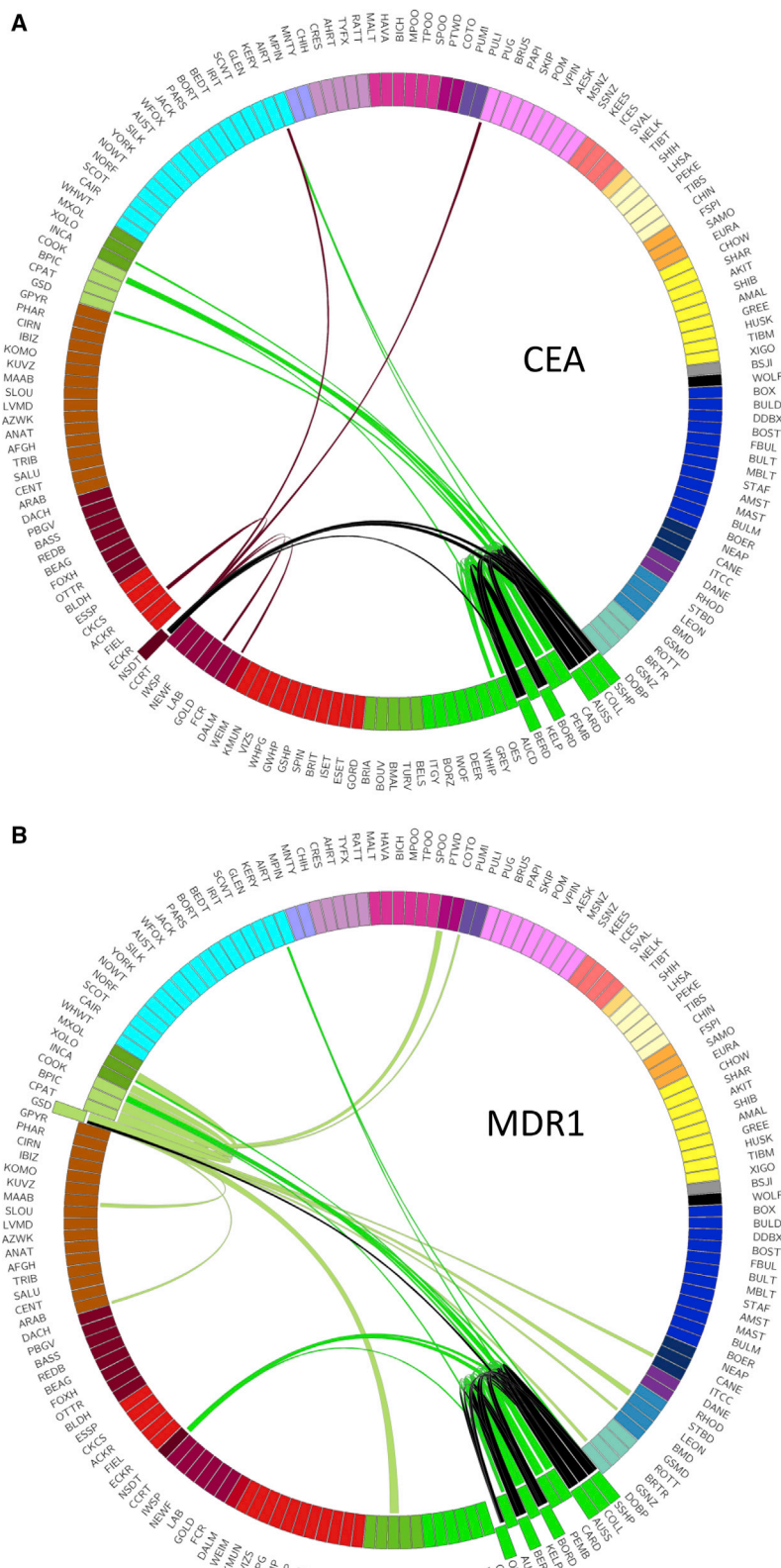
Admixture was measured in Treemix for 18 groups of breeds representing clades or combinations of small clades.

(A) Improvement to the maximum likelihood tree of each group as the result of admixture. The y axis shows fold improvement over the zero admixture tree.

(B) Cladogram of the New World breeds with European herders allowing four migration events. Arrows show estimated migration between breeds colored by weight (yellow to red = 0–0.5).

(C) Cladogram showing migration within the Asian toy clade, including a neighboring breed, the Tibetan terrier. Pictures by Terri Gueck (TIBT), Yuri Hooker (INCA), Mary Bloom (GSD and SHIH), Maurizio Marziali (CPAT), Mary Malkiel (COOK), and John and Debbie Caponetto (large and small XOLO/MXOL).





**Figure 7. Haplotypes Shared with Breeds that Carry Known Deleterious Mutations**

Breeds are connected if the median shared haplotype size exceeds the 95% threshold for interclade sharing. Sharing between breeds that are known to carry the mutation is colored black, and sharing with other breeds is colored according to the breed that carries the mutation.

(A) Collie eye anomaly is found in a number of herding breeds developed in the United Kingdom and some sporting breeds developed in the United States.

(B) Multi-drug resistance 1 mutation is carried by many UK herding breeds as well as the German shepherd.

well as previously unknown addition of Collie, both carriers of the *MDR1* mutation. Haplotype sharing with the same affected breeds is found in the xoloitzcuintli, which allows us to predict that this rare breed may also carry the deleterious allele but has yet to be tested.

## DISCUSSION

Phylogenetic analyses have often been applied to determine the relationships between dog breeds with the understanding that a tree structure cannot fully explain the development of breeds. Prior studies have shown that single mutations produce recognizable traits that are shared across breeds from diverse clades, suggesting that admixture across clades is a notable source of morphologic diversity (Cadieu et al., 2009; Parker et al., 2009; Sutter et al., 2007). Studies of linkage disequilibrium and haplotype sharing suggest further that within regions of ~10–15 kb, there exist a small number of haplotypes that are shared by the majority of breeds, while breed specificity is revealed only in large haplotypes (Lindblad-Toh et al., 2005; Sutter et al., 2004).

In this study, we observe that the majority of dog breeds either do not share large haplotypes outside their clade or share with only one remote breed. The small number of breeds that share excessively outside their assigned clade could be recently created from multiple diverse breeds or may have been popular contributors to other breeds. For example, the pug dog groups closely with the European toy breed, Brussels griffon (Figure 2F), in the toy spitz clade but also shares extensive haplotypes with the Asian toy breeds (Figure 2B) as well as many small dog breeds from multiple other clades. This likely indicates the pug's early exportation from Asia and subsequent contribution to many small breeds (Watson, 1906). Consider also the extensive cross-clade haplotype sharing in the chinook, a recently created breed with multiple ancestors from different breeds. Our data both recapitulates and enhances the written history of this breed (<http://www.chinook.org/history.html>) (Data S1). Extreme examples such as these underscore the complications implicit in relying on phylogeny alone to describe breed relationships. Overall, our data show that admixture has played an important role in the development of many breeds and, as new hybrids are added to phylogenetic analyses, the topology of the cladogram will likely rearrange to accommodate.

The ability to determine a time of hybridization for recent admixture events can refine sparse historical accounts of breed formation. For example, when dog fighting was a popular form of entertainment, many combinations of terriers and mastiff or bully-type breeds were crossed to create dogs that would excel in that sport. In this analysis, all of the bull and terrier crosses map to the terriers of Ireland and date to 1860–1870. This coincides perfectly with the historical descriptions that, though they do not clearly identify all breeds involved, report the popularity of dog contests in Ireland and the lack of stud book veracity, hence undocumented crosses, during this era of breed creation (Lee, 1894).

The dates estimated from these data are approximations, as selection for or against traits that accompanied each cross, as well as the size of the population at the time of the cross, would have affected retention of the haplotypes within the genome.

Based on these estimates, the excess haplotype sharing that we have identified represents the creation of breeds since the Victorian era breed explosion. Most breeds within each clade share haplotypes at this level (<200 ybp); however, the lack of sharing across the clades, outside of very specific crosses, suggests the clades were developed much earlier than the breed registries. Dividing the data by clade, the median haplotype sharing is lowest in the Asian spitz (median = 0) and the Mediterranean clades (median = 516,900) (median range across all clades = 0–3,459,000), indicating that these clades are most divergent and possibly older than the rest. This fits well with previous studies that suggest the earliest dogs came from Central and East Asia (Pang et al., 2009; Shannon et al., 2015). Interestingly, the mean haplotype sharing is slightly higher in the Asian spitz clade than it is in the Mediterranean clade (mean = 1,596,000 and 1,317,000, respectively) (Figure S1), implying that the Asian spitz breeds have been used in recent crosses while the Mediterranean breeds are currently more segregated. These data describe a staggered pattern of dog breed creation starting with separation by type based on required function and the form necessary to carry out that function. This would have taken place as the need arose during early human progression from hunter-gather to pastoral, agricultural, and finally urban lifestyles. During the last 200 years, these breed types were refined into very specific breeds by dividing the original functional dog into morphotypes based on small changes in appearance and with occasional outcrosses to enhance appearance or alter behavior (e.g., reduce aggression, increase biddability).

Though most breeds within a clade appear to be the result of descent from a common ancestor, the New World dogs and the Asian toys showed nearly 200% improvement in the maximum likelihood score by allowing for admixture between the breeds within the clade. Based on this analysis, the Asian toy dogs were likely not considered separate breeds when first exported from their country of origin resulting in multiple admixture events (Figure 6C). Unexpectedly, the New World clade admixture events center exclusively on the German shepherd dog, which informs both the development of this breed as well as immigration of dog breeds to the New World (Figure 6B). The inclusion of German shepherd dog with cane paratore, an Italian working farm dog, likely indicates a recent common ancestor among these breeds, as the German shepherd dog was derived from a herding dog of unknown ancestry in the late 1800s (<http://gsdca.org>). However, the hybridization of the German shepherd dog with the Peruvian hairless dog and the xoloitzcuintli, also a hairless breed, is unexpected and could be the result of recent admixture to enhance the larger varieties of these breeds or could indicate admixture of generic herding dogs from Southern Europe into South America during the Columbian Exchange.

Dogs have been in the Americas for more than 10,000 years, likely traveling from East Asia with the first humans (Wang et al., 2016). However, studies of mitochondrial DNA suggest that the original New World dogs were almost entirely replaced through European contact (Castroviejo-Fisher et al., 2011; Wayne and Ostrander, 1999; Witt et al., 2014) and additional Asian migrations (Brown et al., 2015). As colonists came to the Americas from the 16<sup>th</sup> to the 19<sup>th</sup> centuries, they brought Old

World livestock, and therefore the dogs required to manage and tend the livestock, to the New World (Crosby, 1972). Many of the newly introduced animals outcompeted the native animals, which may explain the surprising and very strong herding dog signature in the native hairless breeds of South and Central America that were not developed to herd. In this analysis, we observe that the ancient hairless breeds show extensive hybridization with herding dogs from Europe and, to a lesser extent, with each other. We also identify two additional clades of New World breeds, the American terriers and the American toys (Figures 2I and 2J), two monophyletic clades of small-sized breeds from North/Central America, which include a set of related terriers, and the Chihuahua and Chinese crested. Written records state that the terriers trace their ancestry to the feists, a North American landrace dog bred for hunting (<http://www.americanfreeingfeist.com,akc.org>). The Chihuahua and Chinese crested are both believed to have originated in Central America (American Kennel Club, 1998; Parker et al., 2017), despite the nomenclature of the latter, which implies Asian ancestry. In contrast, most new breeds developed in the Americas were created from crosses of European breeds and cluster accordingly (i.e., Boston terrier [European mastiff], Nova Scotia duck tolling retriever [retriever], and Australian shepherd [UK Rural]). The separation of the older American breeds on the cladogram, despite recent European admixture, suggests that both clades may retain the aboriginal New World dog genomic signatures intermixed with the European breed haplotypes, similar to the admixture among European, African, and Native American genomes that can be found in modern South American human populations (Mathias et al., 2016; Ruiz-Linares et al., 2014). This is the first indication that the New World dog signature may not be entirely extinct in modern dog breeds, as has been previously suggested (Leonard et al., 2002).

In addition to the effects on the native population, our analysis of geographically distinct subsets of the same breeds shows that some degree of admixture also occurs in the imported breeds when first introduced into a new country. These data suggest two outcomes of breed immigration that mirror human immigration into a new region: the immigrant population is less diverse than the founding population, and there is often admixture with the native population in early generations (Baharian et al., 2016; Zhai et al., 2016).

We observe further evidence of the role geography plays in the distribution of breeds within the clades. For instance, both the UK Rural and the Mediterranean clades include both sighthound and working dog breeds, two highly divergent groups in terms of physical and behavioral phenotype. Sighthounds are lithe and leggy hunters, built to run fast, and have a strong prey drive. Working dogs include both the tall and heavy flock guards that are bred to live among herds without human interaction, preventing predator attacks, and mid-sized herders (Figure 2T), which are agile and bred to work closely with humans to control the movement of the flock without harming them. Despite the opposing phenotypes under selection, both breed types form single clades stemming from distinct geographical regions. Haplotype analysis shows no recent admixture between the geographically distinct clades, suggesting that these groups arose independently (Figures S2A and S2B). Archeological

depictions show sighthound-type hunting dogs that date back 4,000 ybp (Alderton, 2002; Fogle, 2000), and one of the earliest known writings regarding segregation of dogs based by type clearly delineates hunting dogs from working dogs (Columella, 1954). The new cladogram presented herein suggests that the switch from hunting to agricultural pursuits may have initiated early breed formation and that this occurred in multiple regions. These data show that geographical region can define a foundational canid population within which selection for universally relevant behaviors occurred independently, separating the regional groups also by function long ago.

The lack of admixture across clades that appear to share a common trait suggests that these traits may have arisen independently, multiple times. For example, these data show no recent haplotype sharing between the giant flock guards of the Mediterranean and the European mastiffs (Figure S2D). These breed types required large size for guarding; however, each used that size in a different way, a fact that was recognized at least 2,000 years ago (Columella, 1954). The flock guards use their size to defeat animal predators, while the mastiffs use their size to keep human predators at bay, often through fierce countenance rather than action. The phylogenetic placement of these breeds and lack of recent admixture suggests that giant size developed independently in the different clades and that it may have been one of the earliest traits by which breeds were segregated thousands of years ago.

The cladogram of 161 breeds presented here represents the most diverse dataset of domestic dog breeds analyzed to date, displaying 23 well-supported clades of breeds representing breed types that existed before the advent of breed clubs and registries. While the addition of more rare or niche breeds will produce a denser tree, the results here address many unanswered questions regarding the origins of breeds. We show that many traits such as herding, coursing, and intimidating size, which are associated with specific canine occupations, have likely been developed more than once in different geographical locales during the history of modern dog. These data also show that extensive haplotype sharing across clades is a likely indicator of recent admixture that took place in the time since the advent of breed registries, thus leading to the creation of most of the modern breeds. However, the primary breed types were developed well before this time, indicating selection and segregation of dog populations in the absence of formal breed recognition. Breed prototypes have been forming through selective pressures since ancient times depending on the job they were most required to perform. A second round of hybridization and selection has been applied within the last 200 years to create the many unique combinations of traits that modern breeds display. By combining genetic distance relationships with patterns of haplotype sharing, we can now elucidate the complex makeup of modern dogs breeds and guide the search for genetic variants important to canine breed development, morphology, behavior, and disease.

## EXPERIMENTAL PROCEDURES

Further details and an outline of resources used in this work can be found in the [Supplemental Experimental Procedures](#).



## ACCESSION NUMBERS

The accession numbers for the raw data files for the SNP genotype arrays reported in this paper are GEO: GSE90441, GSE83160, GSE70454, and GSE96736.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, two tables, and two data files and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2017.03.079>.

## AUTHOR CONTRIBUTIONS

H.G.P. conceived of project, performed analyses, created figures, and prepared the manuscript. D.L.D. created figures and assisted in manuscript preparation. M.R. ran SNP chips and worked on early analysis. B.W.D. and A.B. performed experiments. G.C.-R. performed sample collection and DNA isolation. E.A.O. organized and directed the study and contributed to manuscript preparation.

## ACKNOWLEDGMENTS

We gratefully acknowledge support from the Intramural Program of the National Human Genome Research Institute. We thank Sir Terence Clark for collecting DNA samples from multiple breeds of sighthounds from their countries of origin in Africa and Asia; Mauricio Lima, Flavio Bruno, and Robert Gennari for collecting samples from native Italian breeds; and Lei Song for collecting samples from native Tibetan mastiffs.

Received: January 3, 2017

Revised: February 10, 2017

Accepted: March 28, 2017

Published: April 25, 2017

## REFERENCES

- Alderton, D. (2002). *Dogs* (Dorling Kindersley, Ltd.).
- American Kennel Club (1998). *The Complete Dog Book*, 19th Edition Revised (Howell Book House).
- Baharian, S., Barakatt, M., Gignoux, C.R., Shringarpure, S., Errington, J., Blot, W.J., Bustamante, C.D., Kenny, E.E., Williams, S.M., Aldrich, M.C., and Gravel, S. (2016). The great migration and African-American genomic diversity. *PLoS Genet.* 12, e1006059.
- Brown, S.K., Darwent, C.M., Wictum, E.J., and Sacks, B.N. (2015). Using multiple markers to elucidate the ancient, historical and modern relationships among North American Arctic dog breeds. *Heredity* (Edinb.) 115, 488–495.
- Browning, B.L., and Browning, S.R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* 194, 459–471.
- Cadiou, E., Neff, M.W., Quignon, P., Walsh, K., Chase, K., Parker, H.G., Vonholdt, B.M., Rhue, A., Boyko, A., Byers, A., et al. (2009). Coat variation in the domestic dog is governed by variants in three genes. *Science* 326, 150–153.
- Castroviejo-Fisher, S., Skoglund, P., Valadez, R., Vilà, C., and Leonard, J.A. (2011). Vanishing native American dog lineages. *BMC Evol. Biol.* 11, 73.
- Columella, L.J.M. (1954). *On Agriculture* (De Re Rustica), Vol. Books 5–9 (E.S. Forster and E.H. Heffner, Trans.) (Harvard University Press).
- Crosby, A.W., Jr. (1972). *The Columbian Exchange* (Greenwood Publishing Company).
- Donner, J., Kaukonen, M., Anderson, H., Möller, F., Kyöstiä, K., Sankari, S., Hytönen, M., Giger, U., and Lohi, H. (2016). Genetic panel screening of nearly 100 mutations reveals new insights into the breed distribution of risk variants for canine hereditary disorders. *PLoS ONE* 11, e0161005.
- Druzhkova, A.S., Thalmann, O., Trifonov, V.A., Leonard, J.A., Vorobieva, N.V., Ovodov, N.D., Graphodatsky, A.S., and Wayne, R.K. (2013). Ancient DNA analysis affirms the canid from Altai as a primitive dog. *PLoS ONE* 8, e57754.
- Fogle, B. (2000). *The New Encyclopedia of the Dog*, Second Edition (Dorling Kindersley Publishing, Inc.).
- Hayward, J.J., Castelano, M.G., Oliveira, K.C., Corey, E., Balkman, C., Baxter, T.L., Casal, M.L., Center, S.A., Fang, M., Garrison, S.J., et al. (2016). Complex disease and phenotype mapping in the domestic dog. *Nat. Commun.* 7, 10460.
- Irion, D.N., Schaffer, A.L., Famula, T.R., Eggleston, M.L., Hughes, S.S., and Pedersen, N.C. (2003). Analysis of genetic variation in 28 dog breed populations with 100 microsatellite markers. *J. Hered.* 94, 81–87.
- Koskinen, M.T. (2003). Individual assignment using microsatellite DNA reveals unambiguous breed identification in the domestic dog. *Anim. Genet.* 34, 297–301.
- Lachance, J., and Tishkoff, S.A. (2013). SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. *BioEssays* 35, 780–786.
- Lee, R.B. (1894). *A History and Description of the Modern Dogs of Great Britain and Ireland* (Horace Cox).
- Leonard, J.A., Wayne, R.K., Wheeler, J., Valadez, R., Guillén, S., and Vilà, C. (2002). Ancient DNA evidence for Old World origin of New World dogs. *Science* 298, 1613–1616.
- Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., 3rd, Zody, M.C., et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438, 803–819.
- Lowe, J.K., Kukekova, A.V., Kirkness, E.F., Langlois, M.C., Aguirre, G.D., Acland, G.M., and Ostrander, E.A. (2003). Linkage mapping of the primary disease locus for collie eye anomaly. *Genomics* 82, 86–95.
- Mathias, R.A., Taub, M.A., Gignoux, C.R., Fu, W., Musharoff, S., O'Connor, T.D., Vergara, C., Torgerson, D.G., Pino-Yanes, M., Shringarpure, S.S., et al.; CAAPA (2016). A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. *Nat. Commun.* 7, 12522.
- Mealey, K.L., and Meurs, K.M. (2008). Breed distribution of the ABCB1-1Delta (multidrug sensitivity) polymorphism among dogs undergoing ABCB1 genotyping. *J. Am. Vet. Med. Assoc.* 233, 921–924.
- Moody, J.A., Clark, L.A., and Murphy, K.E. (2006). Canine history and breed clubs. In *The Dog and Its Genome*, E.A. Ostrander, U. Giger, and K. Lindblad-Toh, eds. (Cold Spring Harbor Laboratory Press), pp. 1–18.
- Pang, J.F., Kluetsch, C., Zou, X.J., Zhang, A.B., Luo, L.Y., Angleby, H., Ardaian, A., Ekström, C., Sköllerö, A., Lundeberg, J., et al. (2009). mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Mol. Biol. Evol.* 26, 2849–2864.
- Parker, H.G., Kim, L.V., Sutter, N.B., Carlson, S., Lorentzen, T.D., Malek, T.B., Johnson, G.S., DeFrance, H.B., Ostrander, E.A., and Kruglyak, L. (2004). Genetic structure of the purebred domestic dog. *Science* 304, 1160–1164.
- Parker, H.G., Kukekova, A.V., Akey, D.T., Goldstein, O., Kirkness, E.F., Baysac, K.C., Mosher, D.S., Aguirre, G.D., Acland, G.M., and Ostrander, E.A. (2007). Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. *Genome Res.* 17, 1562–1571.
- Parker, H.G., Vonholdt, B.M., Quignon, P., Margulies, E.H., Shao, S., Mosher, D.S., Spady, T.C., Elkahoul, A., Cargill, M., Jones, P.G., et al. (2009). An expressed *fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science* 325, 995–998.
- Parker, H.G., Harris, A., Dreger, D.L., Davis, B.W., and Ostrander, E.A. (2017). The bald and the beautiful: hairlessness in domestic dog breeds. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372, 372.
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093.

- Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8, e1002967.
- Quignon, P., Herbin, L., Cadieu, E., Kirkness, E.F., Hédan, B., Mosher, D.S., Galibert, F., André, C., Ostrander, E.A., and Hitte, C. (2007). Canine population structure: assessment and impact of intra-breed stratification on SNP-based association studies. *PLoS ONE* 2, e1324.
- Ruiz-Linares, A., Adhikari, K., Acuña-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., Fuentes, M., Pizarro, M., Everardo, P., de Avila, F., et al. (2014). Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet.* 10, e1004572.
- Shannon, L.M., Boyko, R.H., Castelano, M., Corey, E., Hayward, J.J., McLean, C., White, M.E., Abi Said, M., Anita, B.A., Bondjengo, N.I., et al. (2015). Genetic structure in village dogs reveals a Central Asian domestication origin. *Proc. Natl. Acad. Sci. USA* 112, 13639–13644.
- Sutter, N.B., Eberle, M.A., Parker, H.G., Pullar, B.J., Kirkness, E.F., Kruglyak, L., and Ostrander, E.A. (2004). Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Res.* 14, 2388–2396.
- Sutter, N.B., Bustamante, C.D., Chase, K., Gray, M.M., Zhao, K., Zhu, L., Padhukasahasram, B., Karlins, E., Davis, S., Jones, P.G., et al. (2007). A single IGF1 allele is a major determinant of small size in dogs. *Science* 316, 112–115.
- Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V.J., Sawyer, S.K., Greenfield, D.L., Germonpré, M.B., Sablin, M.V., López-Giráldez, F., Domingo-Roura, X., et al. (2013). Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* 342, 871–874.
- Vaysse, A., Ratnakumar, A., Derrien, T., Axelsson, E., Rosengren Pielberg, G., Sigurdsson, S., Fall, T., Seppälä, E.H., Hansen, M.S., Lawley, C.T., et al.; LUPA Consortium (2011). Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet.* 7, e1002316.
- Vilà, C., Savolainen, P., Maldonado, J.E., Amorim, I.R., Rice, J.E., Honeycutt, R.L., Crandall, K.A., Lundeberg, J., and Wayne, R.K. (1997). Multiple and ancient origins of the domestic dog. *Science* 276, 1687–1689.
- Vilà, C., Maldonado, J.E., and Wayne, R.K. (1999). Phylogenetic relationships, evolution, and genetic diversity of the domestic dog. *J. Hered.* 90, 71–77.
- Vonholdt, B.M., Pollinger, J.P., Lohmueller, K.E., Han, E., Parker, H.G., Quignon, P., Degenhardt, J.D., Boyko, A.R., Earl, D.A., Auton, A., et al. (2010). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464, 898–902.
- Wang, G.D., Zhai, W., Yang, H.C., Wang, L., Zhong, L., Liu, Y.H., Fan, R.X., Yin, T.T., Zhu, C.L., Poyarkov, A.D., et al. (2016). Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res.* 26, 21–33.
- Watson, J. (1906). *The Dog Book, Volume II* (Doubleday, Page & Company).
- Wayne, R.K., and Ostrander, E.A. (1999). Origin, genetic diversity, and genome structure of the domestic dog. *BioEssays* 21, 247–257.
- Wilcox, B., and Walkowicz, C. (1995). *Atlas of Dog Breeds of the World*, Fifth Edition (T.F.H. Publications).
- Witt, K.E., Judd, K., Kitchen, A., Grier, C., Kohler, T.A., Ortman, S.G., Kemp, B.M., and Malhi, R.S. (2014). DNA analysis of ancient dogs of the Americas: identifying possible founding haplotypes and reconstructing population histories. *J. Hum. Evol.* 79, 105–118.
- Zhai, G., Zhou, J., Woods, M.O., Green, J.S., Parfrey, P., Rahman, P., and Green, R.C. (2016). Genetic structure of the Newfoundland and Labrador population: founder effects modulate variability. *Eur. J. Hum. Genet.* 24, 1063–1070.