

RESEARCH ARTICLE

Whole-genome sequence, SNP chips and pedigree structure: building demographic profiles in domestic dog breeds to optimize genetic-trait mapping

Dayna L. Dreger¹, Maud Rimbault^{1,2}, Brian W. Davis¹, Adrienne Bhatnagar^{1,3}, Heidi G. Parker¹ and Elaine A. Ostrander^{1,*}

ABSTRACT

In the decade following publication of the draft genome sequence of the domestic dog, extraordinary advances with application to several fields have been credited to the canine genetic system. Taking advantage of closed breeding populations and the subsequent selection for aesthetic and behavioral characteristics, researchers have leveraged the dog as an effective natural model for the study of complex traits, such as disease susceptibility, behavior and morphology, generating unique contributions to human health and biology. When designing genetic studies using purebred dogs, it is essential to consider the unique demography of each population, including estimation of effective population size and timing of population bottlenecks. The analytical design approach for genome-wide association studies (GWAS) and analysis of whole-genome sequence (WGS) experiments are inextricable from demographic data. We have performed a comprehensive study of genomic homozygosity, using high-depth WGS data for 90 individuals, and Illumina HD SNP data from 800 individuals representing 80 breeds. These data were coupled with extensive pedigree data analyses for 11 breeds that, together, allowed us to compute breed structure, demography, and molecular measures of genome diversity. Our comparative analyses characterize the extent, formation and implication of breed-specific diversity as it relates to population structure. These data demonstrate the relationship between breed-specific genome dynamics and population architecture, and provide important considerations influencing the technological and cohort design of association and other genomic studies.

KEY WORDS: Population, Homozygosity, Canine, Inbreeding

INTRODUCTION

Early mapping studies utilized a combination of pedigree and linkage analyses to find genes important in disease susceptibility in dogs (i.e. Acland et al., 1998; Cattanaach et al., 2015; Jónasdóttir et al., 2000; Yuzbasiyan-Gurkan et al., 1997). The construction of a canine genetic map, based on a large number of informative

families, was key to the success of these experiments (Mellersh et al., 1997; Neff et al., 1999; Wong and Neff, 2009; Wong et al., 2010). Early studies aimed at mapping disease traits similarly relied on the ability to collect samples from large informative pedigrees. Indeed, one of the unique advantages of domestic dogs for mapping traits has been the availability of large families, with single stud dogs often producing dozens of litters (Benson et al., 2003; Jónasdóttir et al., 2000). Such resources have been used to generate estimates of genetic heritability for many disorders and behaviors (i.e. Cooper et al., 2014; Lappalainen et al., 2015; Todhunter et al., 2005; Persson et al., 2015). Similarly, the effect of genetic bottlenecks and inbreeding on disease (i.e. Pedersen et al., 2015; Reist-Marti et al., 2012; Wilson et al., 2013), together with overall trends in genetic diversity among purebred dogs (Hayward et al., 2016; Lewis et al., 2015), have all been investigated with single-nucleotide polymorphism (SNP) or microsatellite studies.

Genome-wide association studies (GWAS) carried out using SNP markers on chips allow for the identification of loci potentially associated with causation, using populations rather than large families. Although published reports describe loci identified from GWAS associated with morphologic traits (i.e. Cadieu et al., 2009; Drögemüller et al., 2008; Hayward et al., 2016; Karlsson et al., 2007; Parker et al., 2009; Schoenebeck et al., 2012; Vaysse et al., 2011; Wolf et al., 2014), disease susceptibility (reviewed in Lequarré et al., 2011; Parker et al., 2010; Schoenebeck and Ostrander, 2014) and even behavior (Dodman et al., 2010; Tiira et al., 2012; Våge et al., 2010), only a few such studies have been able to pinpoint precise mutations. The current standard for canine SNP assays – the Illumina HD chip, with 173,662 potential data points – is of limited utility due to low SNP density in many genomic regions, differential probe affinity, SNP ascertainment bias and the ability to generate only biallelic single-nucleotide variant (SNV) data. Long stretches of linkage disequilibrium (LD) characterize the dog genome, further reducing the utility of GWAS (Lindblad-Toh et al., 2005; Sutter et al., 2004). Finally, because each breed has a unique history, and therefore unique patterns of genomic diversity, GWAS studies are of varying success in correctly identifying loci in different breeds or unique lineages (Björnerfeldt et al., 2008; Boyko et al., 2010; Quignon et al., 2007; vonHoldt et al., 2010).

Canine researchers are increasingly turning to whole-genome sequencing (WGS) to supplement the limitations associated with the less data-dense methods of pedigree analysis and SNP chip mapping. Success stories include studies of domestication (Axelsson et al., 2013; Freedman et al., 2014; Wang et al., 2013), genome architecture (Auton et al., 2013), trait selection and adaption (Gou et al., 2014; Marsden et al., 2016; Wayne and vonHoldt, 2012), and disease susceptibility (i.e. Decker et al., 2015;

¹Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA. ²Institut de Génétique et Développement de Rennes, Rennes 35043, France. ³PIC North America, Hendersonville, TN 37075, USA.

*Author for correspondence (eostrand@mail.nih.gov)

 E.A.O., 0000-0001-6075-9738

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Drögemüller et al., 2008), among others. Variation within dog WGS data, combined with the unparalleled diversity of phenotypes in the dog, provides a unique lens for how genomic underpinnings influence organismal variation. WGS data, once obtained, can also be utilized beyond the initial study for which it was generated, including any hypothesis-driven analyses in which genomic signatures can inform biological questions.

Studies of how breed structure and history can be integrated with industry-standard use of SNP chip analysis and WGS to design the most successful canine genetic studies have yet to be explored. In this paper we consider all of the above in the context of many breeds of differing population substructure, demonstrating that, although a combination of approaches is optimal, population traits can dramatically impact how each should be applied. We define metrics through which population structure can be compared between breeds and determine how that structure should be interpreted with regard to study design and cohort assembly.

RESULTS

Genomic atlas for representation of dog breed diversity

Selection of breeds for pedigree, SNP and WGS analysis focused both on the availability of data as well as representation of diverse breeds as defined by breed type, history, geographical origin and modern popularity. To avoid selection bias, considerable effort was given to equal representation of breed type, function and size. The pedigree data reflect the current status of each breed in the United States (US), as well as the impact of historical changes such as global importation, breed registry recognition and trends in physical type or variety. There are seven historical American Kennel Club (AKC) breed groupings (toy, sporting, terrier, hound, herding, working and non-sporting) that categorize recognized breeds by their traditional function, type or purpose (Club, 2006). The pedigree analysis represents six of these groups, whereas the SNP and WGS analyses include breeds from all seven groups. The non-sporting group has the lowest representation overall, with seven breeds for both the SNP and WGS datasets, and no breeds in the pedigree data. The working group has the highest representation in the SNP and WGS datasets, with 19 and 16 breeds, respectively.

Combined, the WGS, SNP and pedigree datasets represent 112 dog breeds. Genetic variation at 1,510,327 LD-pruned loci

aggregated from WGS of 90 dogs representing 80 breeds was analyzed, providing an impressive breadth of diversity by which to compare breeds and individuals. Complementing this, ten dogs from each of 80 breeds were genotyped using the current industry-standard canine HD SNP array with 173,662 potential sites of variation, providing definition within breeds, but potentially lacking in private variation. The 11 breeds utilized in the pedigree analysis were selected to reflect a range of breed population structures resulting from influences of time, geography and human intervention.

Assessment of inbreeding coefficients by data type

The breed average inbreeding coefficient (F) was calculated three ways: (1) pedigree analysis from five generations, ten generations or the entire breed reference pedigree; (2) from SNP genotype homozygosity analysis averaged over multiple dogs per breed; and (3) from WGS homozygosity analysis of one dog per breed. Calculating F from five- and ten-generation pedigrees evaluates recent trends in inbreeding that occurred at some point after initial breed formation. Additionally, it accounts for differences in pedigree depth when compared to the complete reference population, because there was a large range in the number of effective generations (g_e) for pedigree breeds. All pedigree breeds, however, did have g_e values greater than ten (Table 1).

Pedigree-based inbreeding coefficients ranged from 0.059 (Papillon) to 0.267 (Norwich Terrier) for whole-pedigree data, 0.051 (Papillon) to 0.251 (Nova Scotia Duck Tolling Retriever) when considering ten-generation pedigrees, and 0.022 (Bernese Mountain Dog) to 0.064 (Belgian Sheepdog) for five-generation pedigrees (Table 2). F -value calculations from both SNP genotypes and WGS data are higher than the pedigree analysis across all but the youngest breed (Nova Scotia Duck Tolling Retriever), with the SNP analysis showing a range of 0.179 (Papillon) to 0.536 (Basenji), and WGS F -values ranging from 0.118 (Portuguese Water Dog) to 0.571 (Basenji) (Table 2). Comparing F -values from subsets of the pedigrees, we observed the largest inbreeding coefficients when using the entire reference pedigree and the smallest inbreeding coefficients when examining only the most recent five generations in all breeds. Using only the most recent five generations reduces the across-breed range of F -values to only a

Table 1. Pedigree database population demographics and breed history

Breed ¹	Total pedigree (no. dogs)	Reference population (no. dogs) ²	Reference pedigree (no. dogs) ³	g_e ⁴	Country of origin	AKC ⁵ approval (year)	AKC rank (2014)
ACD	63,203	9037	16,456	11.5	Australia	1980	54
BELS	19,723	2298	8251	22.8	Belgium	1912	117
BMD	122,347	45,961	63,820	22.2	Switzerland	1937	31
BORZ	57,259	2805	14,165	24.8	Russia	1891	99
BSJI	89,894	15,555	27,630	18.2	Central Africa ⁶	1944	84
GOLD	311,260	90,314	204,893	24.8	United Kingdom	1925	3
LAB	86,994	24,021	37,827	13.6	Canada	1917	1
NOWT	12,962	4453	8803	20.2	United Kingdom	1936	94
NSDT	31,734	11,594	14,266	12.9	Canada	2003	96
PAPI	64,144	8064	21,761	13.1	France	1915	42
PTWD	45,355	16,032	19,314	11.7	Portugal	1983	51

¹ACD, Australian Cattle Dog; BELS, Belgian Sheepdog; BMD, Bernese Mountain Dog; BORZ, Borzoi; BSJI, Basenji; GOLD, Golden Retriever; LAB, Labrador Retriever; NOWT, Norwich Terrier; NSDT, Nova Scotia Duck Tolling Retriever; PAPI, Papillon; PTWD, Portuguese Water Dog.

²Dogs born between 2005 and 2015.

³Pedigree created based only on individuals in reference population.

⁴Equivalent number of known generations, measure of pedigree completeness.

⁵AKC, American Kennel Club.

⁶Foundational imports of the modern day Basenji come from present day Democratic Republic of the Congo, Republic of the Congo, Central African Republic and Sudan.

Table 2. Average inbreeding, F , calculated from pedigree data for the entire breed pedigree, ten generations or five generations, and from SNP chip and whole genome sequence (WGS) heterozygosity

Breed ¹	Effective population size (no. dogs)	Average inbreeding				WGS ³
		Whole pedigree	Ten-generation pedigree	Five-generation pedigree	SNP chip ²	
ACD	22.7	0.067	0.064	0.038	N/A ⁴	0.185
BELS	37.0	0.193	0.126	0.064	0.300	0.286
BMD	165.1	0.197	0.061	0.022	0.350	0.314
BORZ	71.8	0.128	0.086	0.054	0.311	0.265
BSJI	16.9	0.221	0.118	0.059	0.536	0.571
GOLD	6.5	0.160	0.079	0.027	0.284	0.218
LAB	58.0	0.082	0.073	0.026	0.217	0.211
NOWT	48.6	0.267	0.167	0.057	0.408	N/A ⁴
NSDT	44.2	0.266	0.251	0.034	N/A ⁴	0.205
PAPI	182.3	0.059	0.051	0.031	0.179	N/A ⁴
PTWD	54.5	0.176	0.162	0.052	0.270	0.118

¹ACD, Australian Cattle Dog; BELS, Belgian Sheepdog; BMD, Bernese Mountain Dog; BORZ, Borzoi; BSJI, Basenji; GOLD, Golden Retriever; LAB, Labrador Retriever; NOWT, Norwich Terrier; NSDT, Nova Scotia Duck Tolling Retriever; PAPI, Papillon; PTWD, Portuguese Water Dog.

²Ten dogs per breed genotyped on the Illumina HD 170K SNP chip.

³WGS of one dog per breed.

⁴Insufficient data was available for calculation of inbreeding values for ACD and NSDT SNP chips, and NOWT and PAPI WGS.

span of 0.042 points compared to 0.208 and 0.200 points in the whole-pedigree and ten-generation calculations, respectively. This flattening of the values indicates that short-range pedigrees cannot account for the relationships between the earlier ancestors and therefore are no longer representative of the breed.

Comparing only the WGS and SNP data across 50 breeds, we observed a range of F -values calculated from the WGS of 0.488, from a minimum of 0.084 (Beagle) to a maximum of 0.571 (Basenji), and a range in SNP-based F -values of 0.423, from 0.113 (Chihuahua) to 0.536 (Basenji). The full list of F -values is shown in Table S1. Breed rankings of SNP- and WGS-calculated F -values showed positive significant correlation ($t=6.179$, $P=1.24 \times 10^{-7}$); however, neither the SNP nor the WGS F -values correlate with pedigree-based inbreeding coefficients (Table 3).

Effective population size (N_e), is the number of individuals in a population who contribute to offspring in the next generation, or the number of breeding individuals that would be required to explain the diversity apparent in a given generation. We hypothesized that this would vary strongly between breeds, because many breeds have undergone unique bottlenecks. In this case, N_e is measured as the change in the inbreeding value of a reference population with that of their parents' generation, and ranged from 6.5 (Golden Retriever) to 182.3 (Papillon) when measured from pedigree data. Using SNP data, the N_e was calculated for each breed over a time span of 13 to 995 generations prior to the acquisition date of the samples. The most recent N_e values, dated 13 generations ago, ranged from 53 (Bull Terrier) to 230 (Chihuahua). That is, at a reference point of 13 generations ago, the Bull Terrier had a reference population size of 53 dogs and the Chihuahua had an effective population size of 230 dogs.

For each breed, N_e values decrease in an approximately exponential rate from distant past to present (Fig. 1A). Although the average slope of each breed-specific N_e curve ranges from 1.52 to 3.92, the breeds are each characterized by a unique N_e value at any given generation point. The data was normalized relative to the breed age, as determined by the AKC date of breed recognition, and a generation interval of 3.76 years (Windig and Oldenbroek, 2015) (Fig. 1B). The rate of change for N_e was calculated for time points from generation 13 to the year of registration, and from the year of registration to an earlier time point equivalent to the amount of time between generation 13 and the recognition year. The N_e at the time of AKC breed recognition ranged from 75 (Norwich Terrier) to 430

(Chihuahua). The difference between the slopes in N_e pre-AKC recognition and post-AKC recognition range from -1.77 (Basset Hound) to 5.16 (Chihuahua). The Basset Hound had a pre-AKC recognition slope indicating a loss of 4.21 breeding dogs per generation, whereas the post-AKC recognition slope showed a loss of 2.44 breeding dogs per generation. The Chihuahua had a pre-AKC recognition slope indicating a loss of 6.44 breeding dogs per generation, and a post-AKC recognition slope indicating a loss of 11.60 breeding dogs per generation. This can further be interpreted as the Basset Hound experiencing greater reduction in effective population size prior to breed recognition, and a lesser rate of reduction of breeding individuals after breed recognition. Conversely, the Chihuahua demonstrates the opposite scenario, with a larger generational decrease in effective population size after breed recognition, compared to before breed recognition. There is no significant correlation between the reference population pedigree N_e and the SNP-based generation-13 N_e ($P=0.166$), with the pedigree N_e values calculated as consistently lower than is revealed by SNP analysis.

Population dynamics from purebred pedigree analysis

The earliest documented relatives that contributed genetically to the reference population for each breed (EDR_e) was used as a means to estimate the original diversity of the breed at the earliest recorded point in breed history. Calculated from the most recent generation, EDR_e ranged from 5.2 (Nova Scotia Duck Tolling Retriever) to 113.1 (Papillon) (Fig. 2). When expanding the potential influencing relatives to include any ancestors, dependent on their marginal contribution to the reference population, the number of effective ancestors (EDR_a) was shown to range from 4.9 (Nova Scotia Duck Tolling Retriever) to 51.4 (Papillon). According to the metrics of EDR_e and EDR_a , and the number of effective genomes (EDR_g) ranging from 2.2 (Norwich Terrier and Nova Scotia Duck Tolling Retriever) to 16.1 (Papillon), the Nova Scotia Duck Tolling Retriever displays the lowest amount of genetic diversity of all the pedigree breeds, whereas the Papillon shows the highest. The ranked order of the least to most diverse breeds remains nearly consistent whether considering using either the EDR_e or EDR_a metric, except with regard to the Australian Cattle Dog and Borzoi, which switch order in the rankings when considering non-founder ancestor contribution (EDR_a). This suggests that, although the Australian

Table 3. Pearson correlations for genetic parameters

WGS	SNP		Pedigree										
	F	RoH	LnH	LnH	F _{5-gen}	F _{10-gen}	F _{all-gen}	EDR _g /EDR _a	EDR _g	EDR _a	EDR _g		
WGS	LnH	70 kb	F	5.363 (2.20 ^{e-6})	5.592 (9.90 ^{e-7})	5.363 (2.20 ^{e-6})	1.036 (0.335)	-0.454 (0.664)	-0.838 (0.430)	-0.843 (0.427)	-0.645 (0.539)		
				-2.683 (0.010)	-2.376 (0.022)	-2.202 (0.031)	-1.145 (0.290)	-0.117 (0.910)	-0.775 (0.464)	0.584 (0.578)	0.512 (0.625)	0.994 (0.354)	0.688 (0.514)
	1000 kb	10 Het	F	2.106 (0.040)	2.050 (0.046)	2.106 (0.040)	0.890 (0.403)	-0.460 (0.660)	-0.537 (0.608)	-0.497 (0.635)	-0.302 (0.772)		
				4.616 (2.94 ^{e-5})	4.340 (7.30 ^{e-5})	4.579 (2.53 ^{e-5})	1.060 (0.324)	-0.196 (0.850)	1.223 (0.261)	-0.378 (0.717)	-1.101 (0.307)	-0.808 (0.446)	-0.639 (0.543)
	1000 kb	10 Het	F	3.247 (0.002)	3.029 (0.004)	3.247 (0.002)	0.821 (0.439)	-0.123 (0.906)	-1.257 (0.249)	-0.626 (0.551)	-0.548 (0.601)		
				3.083 (0.003)	3.029 (0.004)	3.247 (0.002)	1.060 (0.324)	-0.196 (0.850)	1.223 (0.261)	-0.378 (0.717)	-1.101 (0.307)	-0.808 (0.446)	-0.639 (0.543)
	Pedigree	LnH	70 kb	F	3.220 (0.002)	2.947 (0.005)	3.220 (0.002)	1.157 (0.285)	-0.267 (0.798)	-1.056 (0.326)	-0.743 (0.482)	-0.611 (0.560)	
					2.056 (0.044)	2.056 (0.044)	2.056 (0.044)	-0.143 (0.890)	-0.101 (0.922)	0.677 (0.520)	0.092 (0.930)	-0.953 (0.373)	-0.083 (0.936)
		1000 kb	10 Het	F	1.531 (0.132)	1.476 (0.145)	1.531 (0.132)	0.514 (0.623)	0.850 (0.424)	0.542 (0.605)	-0.446 (0.669)	0.007 (0.995)	-0.251 (0.809)
					3.517 (9.65 ^{e-04})	2.575 (0.013)	2.727 (0.009)	0.924 (0.386)	0.033 (0.975)	0.681 (0.518)	-0.436 (0.676)	-0.357 (0.732)	-0.807 (0.446)
1000 kb		10 Het	F	3.682 (5.85 ^{e-4})	3.989 (1.88 ^{e-4})	3.682 (5.85 ^{e-4})	-1.214 (0.264)	0.080 (0.939)	-1.590 (0.156)	0.566 (0.589)	1.117 (0.301)	0.912 (0.392)	
				3.899 (2.99 ^{e-4})	4.624 (2.15 ^{e-5})	3.956 (2.51 ^{e-4})	1.145 (0.290)	-0.093 (0.929)	0.640 (0.543)	-0.500 (0.633)	-0.360 (0.730)	-0.742 (0.483)	-0.561 (0.592)
1000 kb		10 Het	F	2.931 (0.005)	2.959 (0.018)	2.931 (0.005)	0.835 (0.407)	0.008 (0.994)	0.479 (0.646)	-0.270 (0.795)	-0.815 (0.442)	-0.567 (0.589)	
				0.453 (0.653)	0.453 (0.653)	0.453 (0.653)	0.855 (0.421)	0.008 (0.994)	0.479 (0.646)	-0.384 (0.712)	-0.150 (0.885)	-0.639 (0.543)	-0.416 (0.690)
1000 kb		10 Het	F	3.956 (2.51 ^{e-4})	3.956 (2.51 ^{e-4})	3.956 (2.51 ^{e-4})	1.068 (0.321)	-0.359 (0.730)	0.349 (0.738)	-0.280 (0.780)	-0.081 (0.938)	-0.434 (0.677)	
				0.453 (0.653)	0.453 (0.653)	0.453 (0.653)	0.121 (0.907)	-1.023 (0.340)	0.546 (0.602)	0.259 (0.803)	-0.819 (0.440)	-0.087 (0.933)	-0.110 (0.916)
1000 kb	10 Het	F	3.956 (2.51 ^{e-4})	3.956 (2.51 ^{e-4})	3.956 (2.51 ^{e-4})	0.565 (0.590)	0.665 (0.528)	0.593 (0.572)	-0.519 (0.620)	-0.542 (0.605)	-0.404 (0.699)		
			0.453 (0.653)	0.453 (0.653)	0.453 (0.653)	-0.850 (0.423)	-1.702 (0.133)	-3.270 (0.014)	3.498 (8.10 ^{e-3})	6.827 (1.34 ^{e-3})	7.600 (6.31 ^{e-3})	-0.404 (0.699)	
1000 kb	10 Het	F	2.217 (0.057)	2.511 (0.036)	2.217 (0.057)	-0.633 (0.547)	-2.377 (0.049)	-7.818 (1.06 ^{e-4})	-0.122 (0.876)	11.705 (2.59 ^{e-6})	-0.404 (0.699)		
			2.238 (0.056)	2.525 (0.036)	2.238 (0.056)	-0.696 (0.509)	-1.631 (0.147)	-5.905 (5.97 ^{e-4})	1.635 (0.141)			-0.404 (0.699)	
1000 kb	10 Het	F	2.619 (0.031)	2.959 (0.018)	2.619 (0.031)	-0.985 (0.358)	-1.700 (0.133)	-2.081 (0.076)			-0.404 (0.699)		
			0.456 (0.661)	0.456 (0.661)	0.456 (0.661)	1.229 (0.250)	3.663 (5.21 ^{e-3})					-0.404 (0.699)	
1000 kb	10 Het	F	1.486 (0.176)	1.383 (0.204)	1.486 (0.176)	1.148 (0.281)					-0.404 (0.699)		
			1.525 (0.166)	1.383 (0.204)	1.486 (0.176)							-0.404 (0.699)	
1000 kb	10 Het	F	68.761 (2.20 ^{e-16})	68.761 (2.20 ^{e-16})	68.761 (2.20 ^{e-16})						-0.404 (0.699)		
			17.025 (2.20 ^{e-16})	17.025 (2.20 ^{e-16})	17.025 (2.20 ^{e-16})							-0.404 (0.699)	

T-scores for inbreeding coefficients (F), regions of homozygosity (RoH), and total length of homozygosity (LnH) from pedigree, SNP chip and whole genome sequence (WGS) analysis. Correlations are assumed significant when P<0.05, P-values in parentheses. Significant correlations are indicated in bold text.

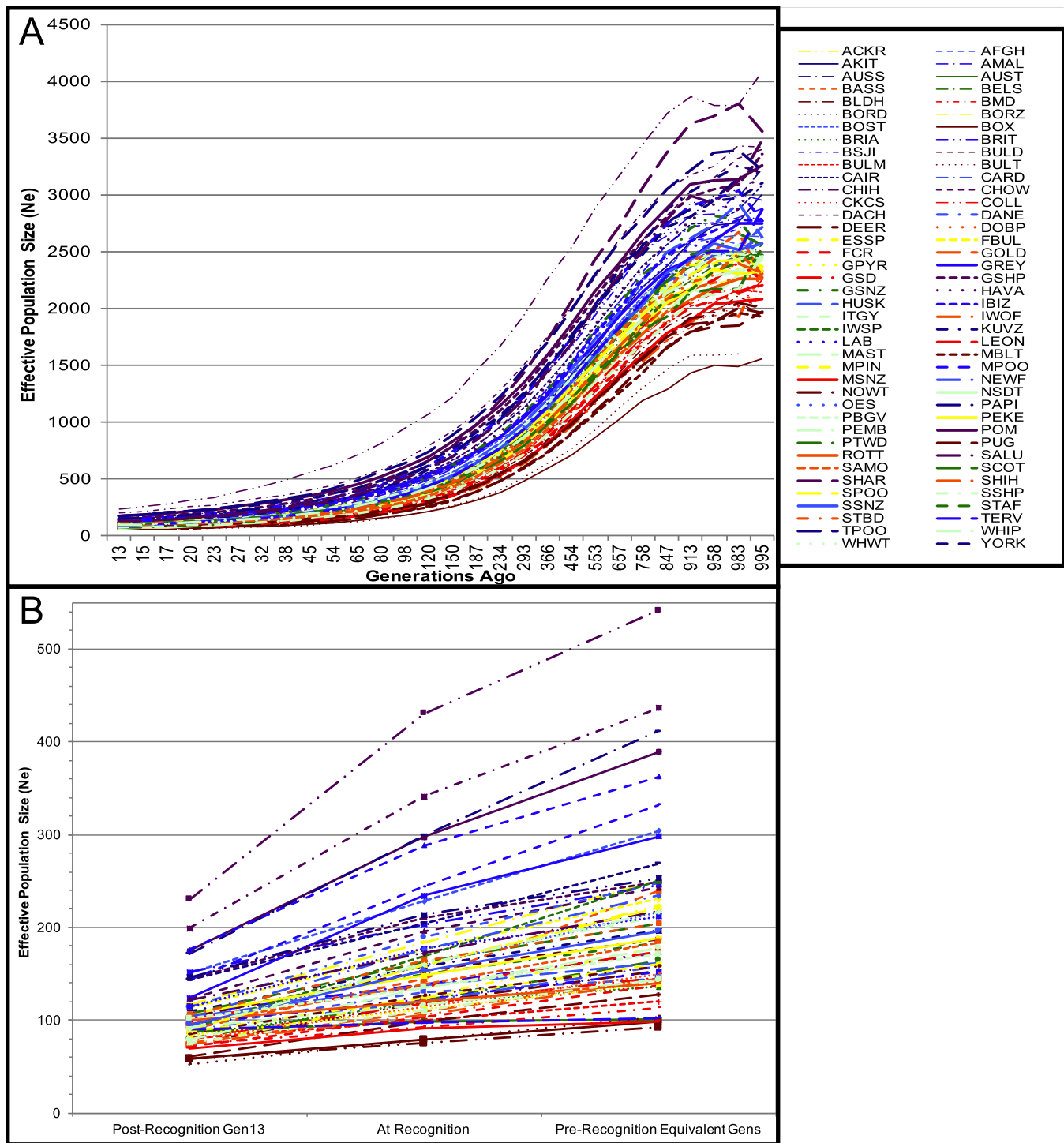


Fig. 1. Effective population size (N_e) calculated from SNP genotypes. (A) Change in N_e (number of dogs) per breed from 13 to 995 generations past. (B) Normalized values of N_e for each breed in the year that they were recognized by the American Kennel Club (AKC), the most recent calculated generation, and the time point prior to breed recognition equal to the span of time from breed recognition to present. The abbreviations used for each breed can be found in Table S1.

Cattle Dog is less genetically diverse than the Borzoi when considering EDR_e , there is notable influence of non-founder ancestors in the Borzoi breed that result in their EDR_a value dropping below that of the Australian Cattle Dog. A ratio of EDR_e/EDR_a of greater than one is indicative of a bottleneck event in the history of the breed. Although minimal in the Nova Scotia Duck Tolling Retriever (1.06), all breeds analyzed showed some

indication of a bottleneck event, with the strongest event occurring in the Labrador Retriever (3.67).

Genomic analysis of homozygosity

To estimate the level of breed homozygosity, it was necessary to filter out the regions of private homozygosity, i.e. those regions that are homozygous in an individual dog but might be heterozygous in

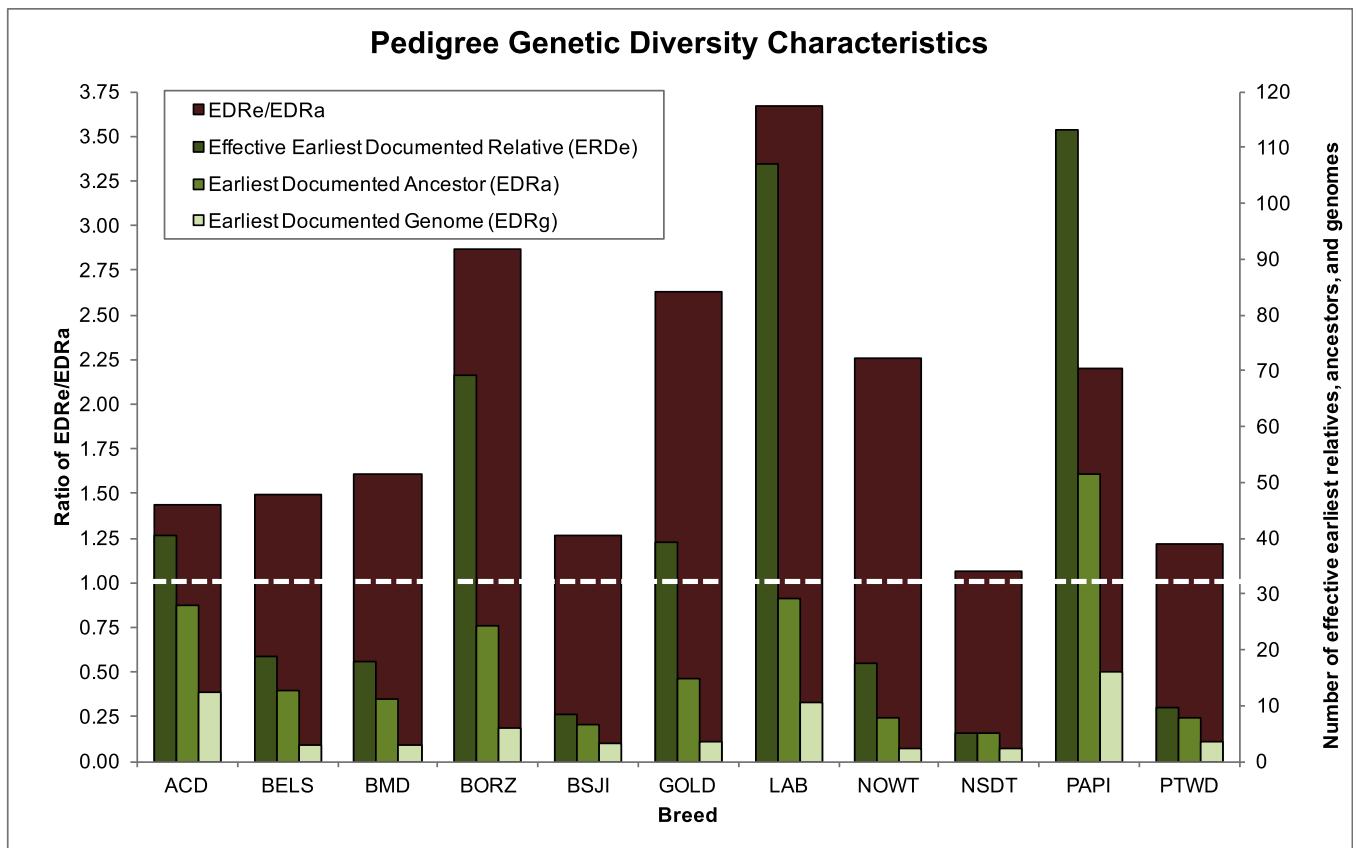


Fig. 2. Genetic diversity parameters reflected in the pedigree databases. The ratio of effective to actual ancestors (f_e/f_a) is measured on the primary axis and suggests the occurrence of a bottleneck event when over one, which is indicated with a dashed white line. The abbreviations used for each breed can be found in Table S1.

other dogs of the same breed, and thus not an indicator of breed-specific homozygosity. For the purpose of this study, homozygous regions present in all sampled individuals of a given breed are denoted 'shared'. Shared regions of homozygosity (RoH) and length of homozygosity (LnH) are therefore common across all representatives of a breed. These were calculated incrementally for each SNP-genotyped breed by randomly adding individuals and recalculating shared homozygosity until all members of the breed, to a maximum of ten, were included (Fig. 3; Fig. S2). In 23 of the 80 SNP-genotyped breeds, shared RoH temporarily increased with the addition of the second dog but shared LnH decreased, indicating that large private runs of homozygosity, present in the initial dog, were broken into smaller pieces by the addition of a second dog. At three through ten dogs, both the RoH and the LnH values decreased by exponentially lesser extents with each new additional dog, such that the tenth dog reduced the first dog's private LnH by between 0.28% (Miniature Poodle) and 7.18% (Shetland Sheepdog) (Fig. 3).

Although the same general pattern was observed in each breed, the rate at which each breed decreased in terms of shared LnH varies. The rate of decay for shared LnH, i.e. the proportion by which shared LnH decreased from the first-dog private LnH with each additional dog, ranged from 0.1996 (English Springer Spaniel) to 0.6065 (Miniature Poodle), with a mean of 0.4098 and s.d. of 0.085 (Table 4).

Because WGS is often available for only a single individual of a given dog breed due to cost considerations, we compared the relative value and utility of SNP chip genotyping on multiple dogs versus WGS analysis of a single dog. The WGS data was first pruned to

remove the SNPs in LD with one another. Because the average spacing of SNPs on the Illumina Canine HD SNP chip is approximately 14 kb and the WGS variants are, on average, only 306 bp apart, homozygosity in the WGS was calculated based on length of region rather than number of SNPs. Additionally, SNP chip analysis might call a region as homozygous despite the potential for heterozygosity between genotyped SNPs, whereas, in WGS, essentially every SNP is genotyped, leaving no missed heterozygosity. In order to compare these two disparate datasets, multiple parameters (see Materials and Methods) were used to calculate homozygosity from the WGS data of single-dog breeds varying both the window size and the allowed heterozygosity within the window. Using the metrics of a 70 kb window and zero heterozygotes, approximately equivalent to a five SNP window from the chip data calculations, the single-dog WGS predicted greater LnH than the shared SNP chip LnH values across all breeds. The shared values of LnH from the SNP chip analysis most closely resemble the single-dog WGS LnH values when calculated with parameters of 1000 kb minimal length and zero allowed heterozygotes, although the breed pattern is not correlated ($P=0.899$).

In addition to single breed representatives, the WGS collection included six breeds for which two dogs were sequenced and two breeds for which three dogs were sequenced. For these breeds, shared LnH and RoH were calculated in the same manner used to assess shared values in the SNP-genotyped breeds. Across all eight breeds, the LnH of the first dog was lower using WGS data than for the SNP analyses. However, the relationship was reversed with the addition of the second dog from each breed such that the shared LnH

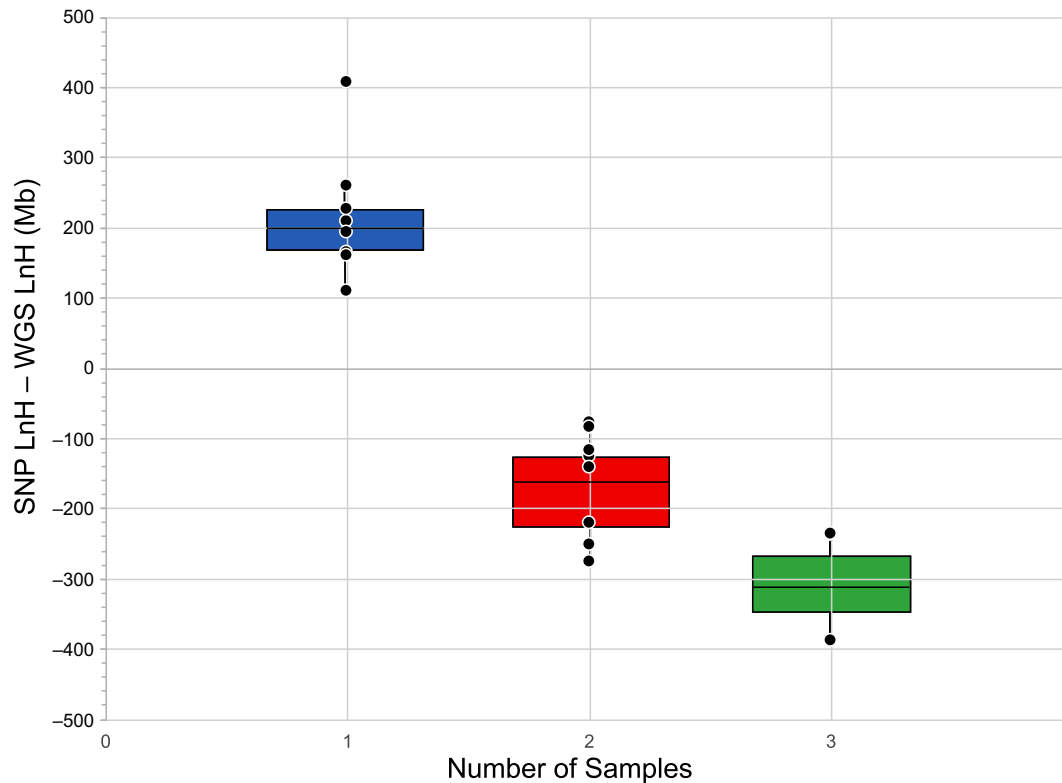


Fig. 4. Difference between single-nucleotide polymorphism (SNP) and whole-genome sequence (WGS) shared length of homozygosity (LnH). Based on one dog, LnH is greater for SNP data than for WGS data. For two or three dogs, however, shared LnH from SNP data is less than for the same breeds from WGS data.

breeds, respectively. Both shared SNP-based measures of homozygosity were positively correlated with whole-pedigree inbreeding coefficients ($P_{\text{RoH}}=0.018$, $P_{\text{LnH}}=0.031$) and WGS homozygosity values when parameters dictated small minimal lengths (70 kb) of homozygosity, and allowed for zero to five heterozygotes. The SNP-based calculation of N_e showed significant negative correlation with SNP RoH ($P=3.41 \times 10^{-11}$), LnH ($P=9.22 \times 10^{-10}$) and F ($P=4.51 \times 10^{-8}$). However, no significant correlation was observed between the SNP-based N_e and pedigree-based N_e values ($P=0.166$). The pedigree analyses did not correlate with any of the WGS RoH or LnH calculations. Shared LnH calculated from SNP chips correlates most closely with the LnH from WGS analyses using a 70 kb window, allowing for a heterozygote at only one locus. Shared RoH calculated from SNP chips correlates most closely with WGS using a 70 kb window and no heterozygotes (Table 3). Although the same patterns of correlation were seen for LnH and RoH when considering WGS data and either single-dog SNP data or multi-dog shared SNP data, the correlations were of highest significance between WGS and single-dog SNP values. However, considering the observed variation between individuals, the shared SNP homozygosity values better represent the homozygosity status of an entire breed.

DISCUSSION

Although the diversity of the dog is increasingly prized for its contribution to human health and mammalian biology, we, like others, observe that the source of this diversity, namely breed structure, presents barriers and complications (Lindblad-Toh et al., 2005; Marsden et al., 2016; Schlamp et al., 2016; Wijnrocx et al., 2016). Modern domestic dog breeds exist because humans have

carefully selected traits of desired purpose. Importantly, they have been, and continue to be, influenced by geographic, cultural and societal forces (Björnerfeldt et al., 2008; Quignon et al., 2007). The work presented here examines variables of inbreeding and homozygosity in a large and comprehensive set of dog breeds through parallel use of pedigree data, genome-wide SNP genotyping and WGS. Specifically, we compare data from extended pedigree analysis, genotyping with an SNP chip of 173,622 potential data points, and WGS with an average depth of 27.79X.

We found that each dog breed has a unique profile of genome diversity, varying by amount of total homozygosity as well as number and size of homozygous regions. Likewise, although we observe variation between members of the same breed, multiple individuals from a single breed can be combined to obtain an accurate reflection of breed-specific homozygosity and knowledge regarding fluidity of variation within breed confines. This allows us to define metrics that inform the design of canine genetic studies while also allowing us to develop an understanding of the intricate complexity of the diversity of dog breeds. Individual diversity metrics are provided for over 100 breeds as a resource for investigators in the field (Table 4; Tables S1, S2).

Population characteristics reflect breed history

As expected, pedigree records for all breeds were erratic and often incomplete prior to breed establishment in their country of origin. In spite of these inconsistencies, considerable across-breed patterns, as well as breed- or situation-specific fluctuations, were identified through pedigree analysis (Fig. 5). Breed recognition by a kennel club registry both requires and facilitates pedigree tracking, thus

Table 4. Single-nucleotide polymorphism (SNP) length of homozygosity (LnH) for breeds with ten representatives, sorted by rate of decay from the lowest to highest

Breed	1st dog LnH (Mb)	10-dog shared LnH (Mb)	Rate of decay	$t_{(1\%)}^1$	$t_{(1\text{ nt})}^2$	Shared LnH (% of genome) ³
ESSP	1595.02	139.64	0.1996	18.03	100.66	6.19
SSHP	1602.12	306.87	0.2070	18.52	97.64	13.61
SCOT	1512.98	216.72	0.2150	17.40	93.57	9.59
BRIA	1594.61	186.37	0.2169	18.90	94.79	8.26
BSJI	1708.18	483.26	0.2520	13.96	78.72	21.43
BLDH	1654.75	492.23	0.2526	14.72	79.12	21.82
IWSP	1462.51	178.89	0.2824	14.18	72.14	7.95
SHIH	1314.09	188.80	0.2964	13.15	67.93	8.37
IWOF	1590.19	443.54	0.3013	12.85	66.80	19.67
COLL	1622.73	567.46	0.3026	12.80	66.24	25.16
MSNZ	1657.82	422.33	0.3049	12.03	65.59	18.71
GPYR	1576.78	206.32	0.3133	12.38	64.83	9.14
BULT	1814.29	769.78	0.3159	11.69	62.85	34.14
BULD	1738.53	278.32	0.3338	11.76	61.18	12.34
CHIH	1293.84	48.09	0.3350	12.63	61.40	2.12
BRIT	1423.88	138.37	0.3370	11.51	60.09	6.13
CKCS	1610.50	455.18	0.3443	11.46	58.69	20.18
DOBP	1632.68	446.11	0.3471	10.84	57.77	19.78
PUG	1665.29	467.65	0.3523	10.74	57.01	20.74
GOLD	1441.33	225.52	0.3572	10.86	56.53	9.98
GSD	1554.77	301.93	0.3647	10.16	54.97	13.39
MBLT	1678.29	614.22	0.3664	10.44	54.60	27.24
WHWT	1464.72	282.65	0.3700	10.66	54.67	12.53
BULM	1565.06	286.97	0.3782	10.26	53.52	12.71
CAIR	1098.37	101.95	0.3804	10.11	52.47	4.50
BMD	1561.30	314.05	0.3816	10.37	53.19	13.90
PEKE	1457.81	285.65	0.3928	9.92	51.35	12.66
LEON	1518.50	241.84	0.3961	9.65	50.96	10.72
AUST	1421.47	167.06	0.3988	10.03	51.01	7.38
FCR	1503.72	309.93	0.4052	9.58	49.80	13.73
MAST	1583.41	252.14	0.4060	9.81	50.21	11.14
BELS	1508.99	223.60	0.4092	9.53	49.54	9.90
BASS	1468.13	287.28	0.4108	9.64	49.29	12.73
NEWF	1532.87	182.96	0.4114	9.76	49.67	8.11
SAMO	1446.18	185.27	0.4140	9.68	49.17	8.21
NOWT	1675.17	382.20	0.4151	9.21	48.66	16.93
AFGH	1492.44	187.90	0.4167	9.36	48.68	8.31
BORD	1395.50	104.84	0.4173	9.81	49.05	4.62
CHOW	1557.54	255.05	0.4183	9.45	48.62	11.29
AMAL	1430.41	163.11	0.4238	9.44	48.03	7.22
AKIT	1504.58	219.23	0.4243	9.38	47.96	9.71
GSHP	1346.27	85.97	0.4274	9.13	47.38	3.81
DEER	1613.92	612.01	0.4286	9.08	46.69	27.14
BOX	1546.11	346.90	0.4343	8.79	46.32	15.39
BORZ	1625.82	223.81	0.4352	9.18	46.99	9.91
ACKR	1604.19	294.41	0.4382	9.11	46.51	13.04
HUSK	1414.91	190.54	0.4390	9.31	46.48	8.41
IBIZ	1411.14	139.81	0.4394	9.28	46.51	6.19
CARD	1453.12	131.53	0.4408	9.01	46.21	5.82
FBUL	1462.48	172.23	0.4414	9.05	46.14	7.64
OES	1386.78	158.78	0.4450	8.68	45.36	7.02
ITGY	1570.53	191.81	0.4467	8.89	45.69	8.50
SSNZ	1330.45	133.41	0.4483	9.00	45.35	5.91
PTWD	1359.92	181.62	0.4523	8.52	44.52	8.04
PEMB	1486.21	217.00	0.4532	8.77	44.86	9.61
LAB	1356.22	92.49	0.4586	8.86	44.51	4.09
WHIP	1627.76	209.14	0.4604	8.39	44.15	9.25
GREY	1498.32	187.64	0.4610	8.68	44.23	8.32
TPOO	1495.93	60.17	0.4642	8.82	44.32	2.67
PBGV	1421.03	156.58	0.4657	8.56	43.67	6.92
GSNZ	1416.27	108.95	0.4660	8.55	43.71	4.43
DANE	1599.95	149.12	0.4722	8.81	43.73	6.61

Continued

Table 4. Continued

Breed	1st dog LnH (Mb)	10-dog shared LnH (Mb)	Rate of decay	$t_{(1\%)}^1$	$t_{(1\text{ nt})}^2$	Shared LnH (% of genome) ³
ROTT	1707.31	303.63	0.4728	8.46	43.27	14.59
KUVZ	1344.86	81.68	0.4731	8.64	43.20	3.62
BOST	1477.08	153.86	0.4741	8.51	43.10	6.81
STAF	1423.94	195.37	0.4806	8.20	42.16	8.67
STBD	1528.69	249.20	0.4901	8.39	41.78	11.04
SPOO	1469.28	90.46	0.4945	8.15	41.40	3.99
TERV	1422.95	162.29	0.5028	7.95	40.46	7.18
YORK	1330.56	102.37	0.5046	7.91	40.26	4.54
SALU	1378.66	126.93	0.5049	8.01	40.38	5.35
HAVA	1291.60	103.82	0.5183	7.73	39.16	4.60
MPIN	1464.77	174.71	0.5201	7.57	39.05	7.74
SHAR	1327.35	107.67	0.5204	7.79	39.14	4.74
POM	1404.18	74.48	0.5225	7.55	38.95	3.30
PAPI	1328.76	102.62	0.5226	7.77	39.00	4.55
AUSS	1341.70	60.82	0.5407	7.26	37.52	2.67
DACH	1368.74	69.16	0.5412	7.71	37.98	3.06
MPOO	1346.35	86.97	0.6065	6.90	33.86	3.86

The exponential rate of decay calculated over ten dogs per breed, based on the shared LnH lost with subsequent addition of individual dogs, relative to the total LnH of the first dog of that breed. Assuming the same rate of decay with >ten dogs, the number of dogs (t) required to bring the rate of shared LnH loss to 1% of the first-dog LnH, and to one nucleotide.

¹ $t_{(1\%)}$, number of dogs such that the t^{th} dog reduces the amount of shared LnH by 1% of the first-dog LnH.

² $t_{(1\text{ nt})}$, number of dogs such that the t^{th} dog reduces the amount of shared LnH by one nucleotide.

³Dog genome length of 2410.98 Mb.

improving records for each breed relative to the point at which the origin registry granted breed status. Concurrent with establishment of a breed, we observe that levels of inbreeding increase steadily and often immediately. Additional increases are noted when the breed achieves registry recognition. This implies that organization of a breed reduces the available gene pool, first by closing the registration database or 'studbook' to the introduction of non-breed-associated genetic variation, imposing an artificial population bottleneck. In addition, it provides a merit system through organized competitions in the sense that a small number of individuals 'win' at dog shows and the genetic contribution of those popular dogs are overrepresented in subsequent generations of the breed, recognized as the 'popular sire effect', thus decreasing variability in the breed pool still further. This is displayed clearly in the ten-generation and all-generation inbreeding graphs of the Australian Cattle Dog, Golden Retriever, Norwich Terrier, Bernese Mountain Dog, Borzoi and Basenji (Fig. S1). This trend is likewise reflected in pedigree-calculated N_e values, where the Golden Retriever, despite being ranked among the top five most popular dog breeds by the AKC from 1993 to 2015 (www.akc.org), has an N_e of 6.5, and the Papillion, with popularity ranking of 48th for 2015, records a substantially higher N_e of 182.3. The large range of N_e values does not necessarily coincide with the total number of breed individuals registered per year or the breed's relative popularity. Rather, it speaks to the general within-breed diversity of the breeds and the contribution of popular dams and sires to the subsequent generation.

In some cases, we observed that formation of the original breed club and creation of an official standard seemed to briefly increase the breed pool, thus transiently decreasing inbreeding values, perhaps by legitimizing and unifying previously distinct lines. This diversity is misleading, however, because dogs that seem unrelated

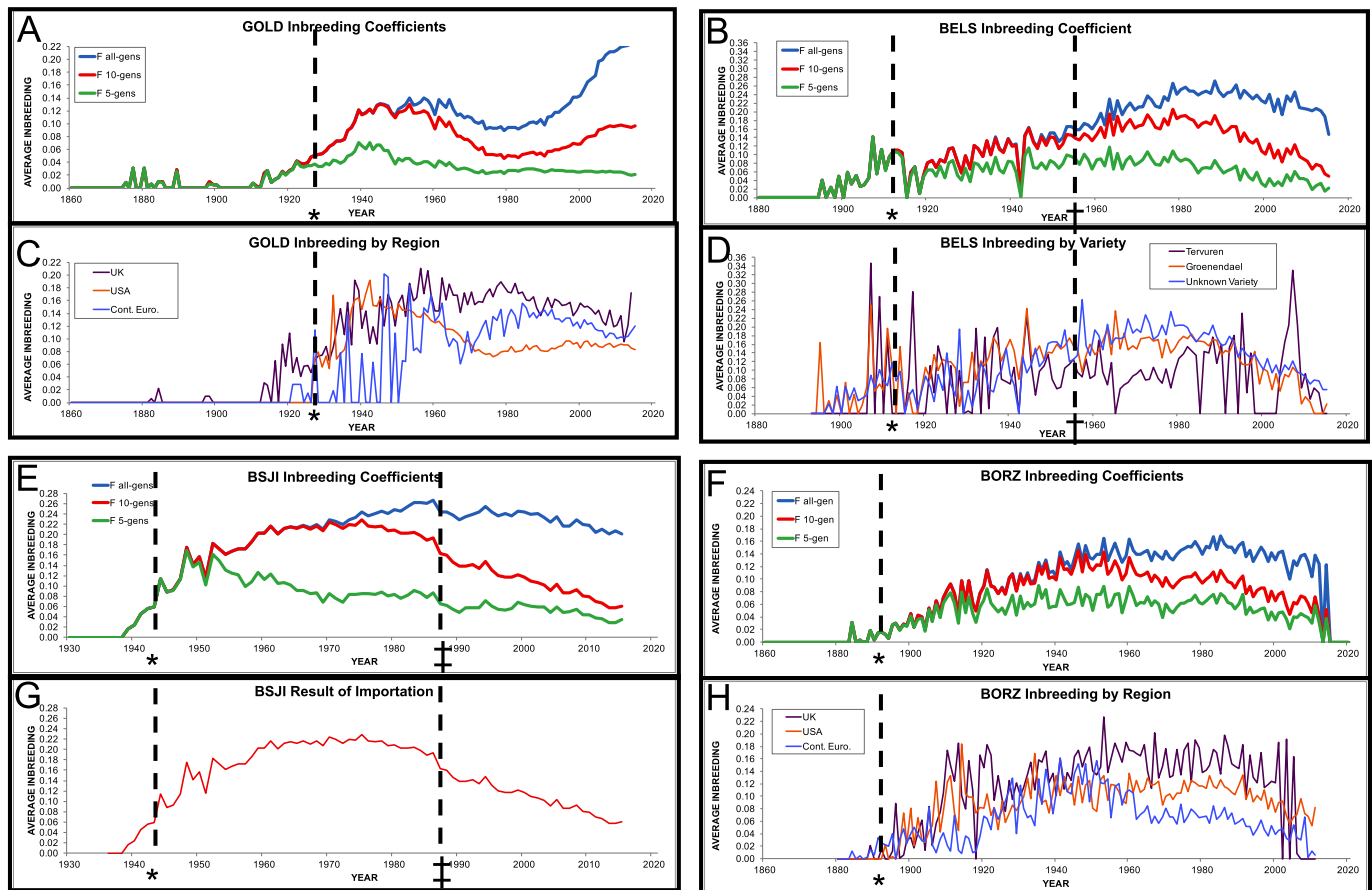


Fig. 5. Graphical representation of inbreeding coefficients. (A,C) Golden Retriever (GOLD); (B,D) Belgian Sheepdog (BELS); (E,G) Basenji (BSJI); (F,H) Borzoi (BORZ). Inbreeding coefficients are calculated over the entire reference pedigree, and including only ten generations or five generations (A,B,E,F). Populations were split by geographic region (C,H) or by breed variety (D) to demonstrate influences on inbreeding values. *Year of breed recognition by the American Kennel Club (AKC). †Year at which the AKC classified the Belgian Tervuren and Belgian Sheepdog as separate breeds. ‡Year at which the Basenji studbook allowed for limited new imports of dogs from Africa to the USA.

at the five- to ten-generation level ultimately trace back to the same small number of founders. This is supported by the correlation between full-pedigree F -values and whole-genome homozygosity measures ($P_{\text{SNP-LnH}}=0.031$; $P_{\text{SNP-RoH}}=0.018$), concurrent with a lack of correlation with shorter-generation F -values (Table 3). The breed ranking of five-generation F -values compared to the larger pedigree analysis or to molecular measures of homozygosity suggests that examination of short-range pedigrees for information about the relationships between individuals in a breed will likely provide misleading information as to the status of the breed as a whole. With the exception of the youngest breed included (the Nova Scotia Duck Tolling Retriever, which seems to have reached a plateau), all analyzed pedigrees showed a peak in inbreeding post-AKC recognition, although the time required to reach that peak ranges from 9 (Labrador Retriever) to 69 (Papillon) years.

When ranking breeds by level of inbreeding and ancestor contributions, the Nova Scotia Duck Tolling Retriever shows the lowest diversity in a majority of categories (ten-generation pedigree inbreeding, EDR_c , EDR_a , EDR_g), whereas the Papillon demonstrates the highest level of diversity, as calculated from the same metrics, plus whole-pedigree inbreeding measures. These values reflect the historical account of breed formation for these breeds. Whereas the Papillon has a diffuse record of origin, spanning much of Western Europe and staking claim to the small spaniel-like dogs represented in artistic renderings from the 16th

century (<https://www.papillonclub.org/History/Welcome.html>), the Nova Scotia Duck Tolling Retriever is the result of a concerted breeding effort to produce dogs that display a very specific game-luring behavior to aid hunters, centered in the Maritimes of Canada during the 19th century (<http://nsdtrc-usa.org/breed/history/>).

An $\text{EDR}_c/\text{EDR}_a$ ratio greater than one indicates the presence of a bottleneck event in the history of a population. Although all breeds achieved a ratio score greater than one, it was most prominent (greater than two) in Borzoi, Golden Retriever, Labrador Retriever and Norwich Terrier (Fig. 2). By evaluating breed history and years of birth of influential ancestors, we can speculate as to the cause of most bottlenecks. We find that, in each case, 10-29% of the most influential ancestors in a breed could be traced to a 5- to 10-year period. These time periods coincide with import/export events in the case of the Borzoi, the recognition of the breed or breed club for the Golden Retriever and Norwich Terrier, and a drastic increase in popularity and population size in the Labrador Retriever.

The databases used for this study were largely US-centric, hence bottleneck events at breed formation or coincident with importation of breeds to the US would be expected, but were not necessarily observed for all breeds. This likely represents events that predate the earliest available records. For example, the first Basenjis were imported to the US around 1941, and our data show that a single male, born in 1939, accounts for 30.3% of the genetic diversity observed in the current reference population. The importation of

dogs from central Africa to the US would almost certainly have resulted in a bottleneck event, but this cannot be confirmed without accurate pedigree data for the African Basenji. We can, however, recognize a distinctive drop in inbreeding values in the Basenji during the late 1980s, corresponding with the time at which new African imports of Basenjies were allowed for registration with the AKC. The intensity of the bottleneck does not correlate with any of the molecular measures of population structure. Independently, however, the number of RoH measured by SNP chip showed significant negative correlation with all measures of earliest documented relatives: EDR_c ($t=-2.717$, $P=0.026$), EDR_a ($t=-2.511$, $P=0.036$) and EDR_g ($t=-2.525$, $P=0.036$). Therefore, a decreased number of effective earliest documented relatives, ancestors and genomes correlated with an increase in shared RoH as measured by SNP chip.

Implications of homozygosity decay

When shared homozygosity measures are calculated within a breed by sequentially increasing the number of dogs in the dataset from one to ten individuals, the LnH that is lost with each additional dog is assumed to be private to the individual or not fixed within that breed. Conversely, there is a point at which the homozygosity identified as shared across a pool of individuals is likely shared within the entire breed (Fig. 3). By utilizing the extent to which the LnH decreases with each additional dog added to the shared LnH calculation, an exponential rate of decay can be calculated, which can in turn be used to predict the number of individuals required to represent the entire range of variation within a given breed as inferred from the current SNP array. When these calculations were performed on 80 breeds using data from SNP genotyping, the rate of decay ranged from 0.1996 (English Springer Spaniel) to 0.6056 (Miniature Poodle). Thus, as each dog is sequentially added to the data set, the shared LnH decreases by 19.96% in the English Springer Spaniel or 60.56% in the Miniature Poodle (Table 4).

With a normal distribution across 80 breeds, and a s.d. of 0.085, 54 breeds have a rate of decay within 1 s.d. of the all-breed mean. The breeds located at the extremes of 3 s.d. from the mean include the Miniature Poodle (0.6056) at the high end and the English Springer Spaniel (0.1996), Shetland Sheepdog (0.2070), Scottish Terrier (0.2150) and Briard (0.2169) at the low end (Table 4). When designing studies, such as GWAS, that utilize SNP data, these decay rates can be implemented to predict the number of dogs in a breed that would be required to represent the full range of variation within that breed. The breeds in the extreme left tail of the decay rate distribution (English Springer Spaniel, Shetland Sheepdog, Scottish Terrier, Briard) would require 17 to 19 individuals to bring the per-dog loss of LnH down to 1% of the initial-dog LnH, or 93 to 100 individuals to bring the per-dog loss of LnH down to only one nucleotide. Conversely, the Miniature Poodle on the extreme right tail of the decay rate distribution requires only seven individuals to reach 1% of initial-dog homozygosity or 34 individuals to reach one nucleotide of homozygosity reduction (Table 4).

Caveats to the above, however, include the fact that decay rate does not speak to the overall amount of homozygosity contained within each breed, only the amount of variation between individuals within that breed. For example, the breed with the largest shared LnH is the Bull Terrier (769.78 Mb), whereas the breed with the smallest shared LnH is the Chihuahua (48.09 Mb) (Fig. 3). Comparatively, the decay rate difference between the Chihuahua (0.3350) and the Bull Terrier (0.3159) is only 1.91%, dictating that both breeds require approximately 12 dogs to represent 99% of the genomic variation within the breed, despite the difference in shared

homozygosity of 721.69 Mb. This concept is important and, as a result, we provide breed-specific measurements of homozygosity as well as unique signatures of decay representing variation within 80 breeds (Table 4) as a utility to researchers in the field. Both of these variables are necessary to guide genome-based experimental design in pure-breed dogs because our data suggest that the degree to which individual dogs of the same breed vary in homozygosity in relationship to each other, reflected by the decay rate, is of at least equal importance to the overall shared homozygosity of a breed for determining required sampling size for mapping studies.

In order to test the capabilities of breed-specific homozygosity decay to predict appropriate cohort size, a small-scale proof-of-concept experiment was conducted. Genetic variation of the *RSPO2* gene has previously been associated with the wire or furnished coat type, a phenotype that is fixed and easily recognizable in many dog breeds (Cadieu et al., 2009). The Miniature Poodle and Scottish Terrier are breeds that are near fixation for the furnishing *RSPO2* variant, and have high (0.6065) and low (0.2150) homozygosity decay rates, respectively. Likewise, the Papillon (high decay rate=0.5226) and Shetland Sheepdog (low decay rate=0.2070) are wild-type at the furnishings locus. The high decay rates suggest that between seven and eight dogs would be necessary to reduce the shared LnH to 1% of the initial, whereas the low decay rates indicate 17 to 19 dogs would be required to produce the same results. Using the ten dogs of each breed genotyped on the Illumina HD SNP array, a standard non-adjusted association analysis was conducted with PLINK software (Purcell et al., 2007) using high-decay-rate Papillons as controls and high-decay-rate Miniature Poodles as cases, and separately, with low-decay-rate Shetland Sheepdog controls and low-decay-rate Scottish Terrier cases (data not shown). With ten cases and ten controls, the comparison of Papillon with Miniature Poodle assigned highest significance to the region surrounding *RSPO2* on CFA13. The equivalent analysis of low-decay-rate Shetland Sheepdog and Scottish Terrier breeds did not indicate significant association with *RSPO2*. Subsequently, with sequential reduction of one random dog per breed from the high-decay-rate analysis, the *RSPO2* region maintained the highest relative significance from $n=10$ through $n=7$, although the P -value did drop below genome-wide significance. Therefore, even under loose constraints, there is evidence for homozygosity decay rates to serve as predictive for trait-mapping ability.

Observation of pedigree structure by molecular means

Population traits calculated through pedigree analysis did not correlate with genomic measures unless the entire reference pedigree was used. Pedigree-based inbreeding coefficients calculated over the entire reference pedigrees correlated significantly with the RoH ($t=2.959$, $P=0.018$) and LnH ($t=2.619$, $P=0.031$) calculated using SNP genotypes shared across ten dogs of each breed, but did not correlate at all with homozygosity measures of F , LnH or RoH derived from individual WGS. This would suggest that deep reference pedigree structure can be captured by SNP analysis of multiple breed individuals, but is not apparent in WGS analysis of only one individual. Likewise, five- or ten-generation pedigree analysis is not sufficient to elucidate the larger breed-specific population structure apparent in the entire pedigree reference populations.

Comparison of WGS and SNP measures of homozygosity

In this study we compared the genome-wide estimates of breed-expected homozygosity based on SNP genotyping of several individuals versus WGS of one individual, two similarly priced

methods currently in use among canine geneticists. Genome-wide and breed-specific estimates of homozygosity dynamics can be useful for designing association studies by aiding in determination of the number of individuals that will be necessary to identify the locus of interest.

When comparing breeds in which shared homozygosity was calculated from SNP data, a significant positive correlation was observed between F -values and shared LnH ($t=16.261$, $P=2.20\times 10^{-16}$) and shared RoH ($t=17.025$, $P=2.20\times 10^{-16}$) (Table 3), supporting the expectation that higher inbreeding coefficients result from increased homozygosity. Within that SNP-based data, increasing F in breeds with similar shared LnH (<1000 kb difference) increased RoH in 53.8% of breed pairs. However, increasing shared LnH in breeds with similar F -values (<0.1% difference) resulted in an increased RoH in 84% of the breed pairs. This suggests that increased inbreeding produces a greater number of homozygous regions, rather than increased length of pre-existing homozygous regions. This finding corroborates the previously discussed negative correlation between lower pedigree-based values of EDR_c , EDR_a and EDR_g and increased SNP-based RoH.

Although correlation was observed between shared SNP LnH and single-dog WGS LnH over short regions, the significance of that measure varied greatly depending on the number of allowable heterozygotes or minimum length of the WGS regions assessed. For instance, the correlation value for SNP LnH with WGS LnH with a minimum length of 70 kb and one allowed heterozygote was $t=4.579$, $P=2.53\times 10^{-5}$, whereas the correlation value for SNP LnH with WGS LnH of the same minimum length but with five allowed heterozygotes was $t=2.106$, $P=0.040$ (Table 3). Recognizing that calculating shared homozygosity over multiple individuals will reduce the effects of private homozygosity in SNP-based data, we applied the same process to small numbers of dogs for which WGS was obtained. As such, we could directly compare the difference in single-dog WGS homozygosity and two- or three-dog WGS shared homozygosity, and determine whether the patterns were reflective of those seen in comparable SNP-based calculations.

SNP-based homozygosity calculations rely on the expectation of long regions of LD to assume homozygosity between markers. This can lead to artificially reduced numbers and increased size of RoH if chip SNPs are homozygous while intervening regions contain variation, or are not reflective of heterozygosity in certain breeds, creating artificially inflated measures of LnH. Although WGS LnH for one dog per breed yielded lower values than observed from SNP-calculated shared LnH, the inclusion of two WGS dogs per breed resulted in shared LnH values greater than observed in the respective SNP genotyped breeds (Fig. 4). Because WGS has the potential to provide a more accurate representation of genomic metrics, these results indicate that the ~170K SNP chip overestimates individual LnH and underestimates shared LnH, relative to WGS values. Owing to cost, WGS is usually applied to fewer individuals than SNP chip technology; however, the inclusion of two WGS dogs can reduce the LnH from one dog by 15.5% to 25.5%. In the two breeds for which we used three WGS dogs to calculate shared LnH, the third dog reduced the initial LnH by an additional 8% in each case. These values might, therefore, support a cost/benefit argument for the inclusion of WGS from two dogs per breed, instead of only one, when considering study design.

Despite the two genomic methods occasionally producing discordant values for breed-specific RoH and LnH, the significant correlation between the two distributions across all 50 breeds for

which both SNP and WGS data was analyzed suggests that either SNP chip analysis or WGS of a small number of dogs would be sufficient to estimate degree of homozygosity when considered relative to other breeds. Although WGS from multiple individuals will produce continually more accurate data, through comparison to the database of breed homozygosity measurements provided here, single-dog sequencing homozygosity measurements can be sufficient to predict breed-wide genomic structure.

For the majority of dog breeds for which WGS data is publicly available thus far, generally one individual per breed has been sequenced (<http://www.ncbi.nlm.nih.gov/sra>). If multiple dogs have undergone WGS, it often reflects different geographic origins of the lineage and the utility of that data for designing a genomic study in one geographic region might be limited (Quignon et al., 2007). The data employed in this study are restricted in large degree to purebred dogs sampled within the US. However, similar studies undertaken in European populations are likely to produce equally informative results, both for understanding breed genomic architecture among European breeds and for the design of association studies.

In this study, we leveraged data from three sources to study genome structure in dog breeds and, in doing so, have characterized the unique and extensive breed-specific population structure present in modern dog breeds so as to predict optimal study designs for genetic experiments. By combining comprehensive genotype and relationship data from these three separate technologies, we have produced a series of metrics that are directly applicable to studies of domestic dog breed structure and suggest how those metrics should be applied to the design of canine genetic studies. In summary, we propose that each study design be considered independently in terms of not only the expected inheritance pattern displayed, but also the breed in which the trait occurs. Cohort sizes for SNP analysis should reflect the breed-specific decay rates and levels of homozygosity displayed in Table 4. When considering WGS analysis, greater benefit is achieved by obtaining data from two individuals than is gained with the inclusion of a third individual of the same breed. The relative degree of inbreeding, effective population size and homozygosity from SNPs correlates with the inbreeding and short-range (70 kb) homozygosity rankings from WGS. However, pedigree-based metrics do not necessarily provide an accurate representation of the genetic population measurements. Although mode of inheritance and effects of penetrance, epistasis and pleiotropy all play a role in the outcome of an association study, we show that accurate assessment of breed-specific genomic topography is exquisitely valuable for determining the most effective sample size. We believe that these efforts to capture the distinctive and dynamic homozygosity landscape of over 100 pure dog breeds will serve as a platform upon which future research initiatives can be modeled.

MATERIALS AND METHODS

Pedigree datasets

Data used for pedigree analysis in this study were provided by private breed databases for the following 11 breeds: Australian Cattle Dog, Belgian Sheepdog, Bernese Mountain Dog, Borzoi, Basenji, Golden Retriever, Labrador Retriever, Norwich Terrier, Nova Scotia Duck Tolling Retriever, Papillon and Portuguese Water Dog. Due to breed history, the Belgian Sheepdog database also includes several Belgian Tervurens and Belgian Malinois, and the Golden Retriever database includes early contributing breeds including several Flat-coated Retrievers, Irish Setters and Labrador Retrievers. The abbreviations used for each breed throughout this study can be found in Table S1. Breeds for pedigree analysis were selected based on the availability of high-quality database information. To be entered into this study, the breed database had to include at least 10,000 individuals over >10

effective generations (g_e), calculated as described in the Pedigree analysis section, and reflect the modern breeding population of the breed by including at least 10% of the AKC-registered dogs for that breed, determined by registration data for 2014. Exceptions were made for the Labrador Retriever and Golden Retriever breeds, which, due to their immense popularity, reflected only 0.47% and 6.55% of AKC registrations for those breeds, respectively. Despite this, the database reference population sizes for Golden Retriever and Labrador Retriever ranked first and third highest of included pedigree breeds.

Breed databases were predominantly compiled by volunteer enthusiasts using registries, studbooks, historical records, dog-show entries and personal breeder records. Initial pedigree analysis required an internal error check whereby improbable relationships are flagged for manual correction. Breeds were removed from study inclusion when the pedigree database contained insurmountable inconsistencies. When data regarding country or year of birth was missing from the raw database, estimated values were extrapolated based on equivalent data from siblings, offspring, dam and sire, and assuming a standard generation interval of 2 years. Demographics from the datasets are provided in Table 1, demonstrating that these breeds reflect the variation present in American populations of modern dog breeds with regard to country of origin, population size, popularity and history. Pedigree databases ranged in size from 12,962 (Norwich Terrier) to 311,260 (Golden Retriever) dogs.

The AKC began accepting breeds for registration in the 1880s and the earliest recognized pedigree breed in this study is the Borzoi from 1891 (<http://www.akc.org/press-center/facts-stats/page-3/>). The databases used in these analyses, however, record parentage as far back as the 1830s (Labrador Retriever), 1840s (Golden Retriever) and 1860s (Borzoi). Of the breeds included in the study, the Nova Scotia Duck Tolling Retriever was the most recent to gain recognition by the AKC, obtaining breed status in 2003. Pedigree data for this breed, however, dates back to 1930. North America is listed as the predominant region of birth or initial registration for all 11 breeds, representing between 37.04% (Norwich Terrier) and 76.29% (Portuguese Water Dog) of dogs within the pedigrees of the modern dogs. The dogs not listed as being born in North America derive primarily from Europe, Asia and Oceania.

All dogs born between 2005 and 2015 make up the reference population for each breed. The direct ancestors of those dogs were traced back through successive parent-offspring relationships to the earliest dogs within the respective database. Importantly, use of a reference population to represent the modern breed removes any individuals present within the database that do not genetically contribute to the development of the breed as it currently exists. Reference pedigrees, comprising the reference populations and their ancestors, ranged in size from 8251 (Belgian Sheepdog) to 204,893 (Golden Retriever) and covered between 11.5 (Australian Cattle Dog) and 24.8 (Golden Retriever and Borzoi) effective generations. All subsequent calculations were based on the reference pedigree.

Pedigree analysis

Pedigree completeness, inbreeding coefficients, and effective numbers of founders, ancestors and founder genomes were calculated using PEDIG software (Boichard, 2002). On the basis of these calculations, population founders are construed to be individuals with no parental data, assuming that this indicates the creation period of that breed. However, because some of the pedigrees are lacking complete data for individuals at time points post-breed-formation, this is not completely applicable in our data set. To avoid confusion in terminology, we will refer to the PEDIG founder calculations instead in terms of 'earliest documented relative' (EDR). Thus, we account for individuals that might be founders in the traditional sense, as well as those individuals who have contributed genetically to the breed in more recent years but lack documented pedigree data. Pedigree completeness for the reference population related to each breed was evaluated by first calculating the proportion of known ancestors at each generation and then summing these proportions over all generations, resulting in the number of equivalent complete generations (g_e).

Inbreeding coefficients (F) were calculated for each dog in the reference pedigree by computing relationship matrices for the individual and their ancestors (Wiggans et al., 1995), using subsets of all ancestors, ten

generations of ancestors, or five generations of ancestors, using the 'vanrad' function of PEDIG (Boichard, 2002). The individual F -values were averaged across all dogs of a breed that were born within a given year to obtain breed-specific inbreeding coefficients over time. Effective population size (N_e) was calculated as the difference in inbreeding (ΔF) between the reference population and the parents of those individuals:

$$N_e = \left(\frac{1}{2}\right) \Delta F. \quad (1)$$

Because the EDR include founder animals as well as recent genetic contributors with missing parentage data, the effective EDR (EDR_e), defined as the number of equally contributing EDRs expected to produce the amount of genetic diversity observed in the reference population, was calculated as:

$$EDR_e = \frac{1}{\sum(p_i^2)}, \quad (2)$$

where p_i is the proportional contribution of the EDR over all descendants in a pedigree (Lacy, 1989). To account for loss of genetic diversity after the foundation of a breed, effective number of ancestors (EDR_a) was calculated as

$$EDR_a = \frac{1}{\sum(p_k^2)}, \quad (3)$$

where p_k is the marginal contribution of each ancestor in the relevant reference pedigree (Boichard et al., 1997). EDR_a is the minimum number of ancestors that explain the amount of genetic diversity observed in the reference population. A list of the 100 most influential ancestors and their contributions were also identified for each breed. Because ancestors do not necessarily need to be founder animals, and can be related to one another, the marginal contribution of specific ancestors, i.e. that contribution not yet explained by other ancestors, was considered when evaluating the effect of influential individuals. The 'prob_orig' function of PEDIG (Boichard, 2002) was utilized to calculate the EDR_e , EDR_a and ancestor contributions. Additionally, the ratio of EDR_e/EDR_a was calculated because it indicates the occurrence of a bottleneck event when greater than one. Finally, the effective number of EDR genomes (EDR_g), which measures the probability that an EDR haplotype is still present at a given locus in the reference population and which accounts for all random loss of alleles during segregation and due to genetic drift, was calculated with the 'segreg' function of PEDIG (Boichard, 2002) as:

$$EDR_g = \frac{1}{\sum(r_i^2/r_i)}, \quad (4)$$

where r_i is the fraction of an EDR's alleles presumed to have been retained in the population (Lacy, 1989).

Sample collection and genotyping

Blood samples from purebred dogs were collected from private owners as described previously (Parker et al., 2009) and DNA was isolated using standard phenol-chloroform methods (Sambrook et al., 1989), aliquoted and stored at -80°C . The majority of dogs were registered with the AKC, and AKC registration numbers were used to verify breed affiliation and relatedness. Those that were not registered were pedigree-verified as eligible purebreds. Owners signed an informed consent prior to sample collection in accordance with a National Human Genome Research Institute (NHGRI) Animal Care and Use Committee. Ten dogs from each of 80 breeds, determined to be unrelated within three generations, were genotyped by the Ostrander lab using the Illumina Canine HD SNP chip (Illumina, San Diego, CA). Genotypes were called using Illumina Genome Studio, retaining SNPs with >90% call rate, heterozygous excess of -0.7 to 0.5 , and GenTrain score of >0.4 .

SNP-based analysis of population metrics

A homozygous region was calculated as five or more consecutive SNPs, predicted to span at least 70 kb when considering the array average SNP spacing of 14 kb, which were homozygous in an individual dog. These

homozygosity parameters would therefore include regions of homozygosity less than the expected minimum length of canine LD of 20 kb (Gray et al., 2009). The total number of such regions is termed regions of homozygosity (RoH). The combined length of all homozygous regions is termed the length of homozygosity (LnH). The RoH and LnH were first calculated for each individual dog, and then combined in sets increasing in size from two to the maximum number of dogs of the same breed. Dogs of the same breed were added one at a time by random selection and RoH and LnH were recalculated as regions of five or more consecutive SNPs that were homozygous in all individuals. In this way, the regions and length of homozygosity common across all individuals of the same breed are termed the shared RoH and shared LnH, respectively. Calculation of shared LnH and shared RoH allowed for a genotype missingness rate of <20% across included dogs with retained complete homozygosity in the remaining genotyped individuals.

The LnH of the first randomly selected dog of each breed was decreased with each additional same-breed dog added to the sequential shared LnH calculations, such that the difference between one-dog private LnH and two-dog shared LnH is exponentially greater than the difference between nine-dog and ten-dog shared LnH. For each breed with ten representatives, exponential rate of decay (k) was calculated as:

$$k = \frac{\ln(y_t/a)}{t}, \quad (5)$$

where t is the number of additional dogs (i.e. 9), y_t is the difference in shared LnH between $t-1$ and t , and a is the amount of LnH lost by the addition of the second dog. Therefore, the number of dogs required to reduce the shared LnH by 1% of the first dog's LnH was calculated with the equation:

$$y_t = a(e^{kt}), \quad (6)$$

where y_t is 1% of the first-dog LnH. Likewise, $y_t=1$ was used to determine the number of dogs required to reduce the shared LnH by only one nucleotide, estimating the point at which the first-dog private LnH will not be decreased further by the addition of more dogs.

The program SNePv1.1 (Barbato et al., 2015) was used to calculate effective population size (N_e) with the SNP genotypes from each breed. Predicted N_e values for 13 to 995 generations prior to sample age were calculated and values were interpolated using a generation interval of 3.76 years (Windig and Oldenbroek, 2015).

Whole-genome sequencing data generation

WGSs were compiled using data from 90 purebred dogs representing 80 distinct breeds. Seventy-two breeds were represented by one sequenced dog each, two dogs were sequenced for each of six breeds (Chow Chow, Bernese Mountain Dog, Greyhound, Rottweiler, Scottish Terrier, West Highland White Terrier), and three dogs were sequenced for each of two breeds (Flat-coated Retriever, Irish Water Spaniel). Data were obtained via the Short Read Archive (ncbi.nlm.nih.gov/sra) from previously published studies, or sequenced for this study by the National Institutes of Health (NIH) Intramural Sequencing Center (NISC) using the Illumina TruSeq DNA PCR-Free Protocol (Cat.# FC-121-3001) from DNA samples provided by the Ostrander laboratory (Table S2). Previously unpublished data from 27 sequenced dogs are deposited in the Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra>). Libraries were sequenced on the Illumina HiSeq 2000 platform with 100 bp paired-end fragments of 300-500 bp. Paired data was aligned to the CanFam3.1 reference genome (<http://genome.ucsc.edu/cgi-bin/hgGateway?db=canFam3>) with the BWA 0.7.10 MEM algorithm (Li and Durbin, 2009), sorted with SAMtools 0.1.10 (Li et al., 2009), and screened for putative PCR duplicate reads with PicardTools 1.119 (<https://github.com/broadinstitute/picard>).

Sequences were locally realigned based on documented and novel insertions-deletions (Axelsson et al., 2013) using GATK 3.2-2 (DePristo et al., 2011), and training sets of dbSNP and Illumina Canine HD chip positions were used for base quality recalibration. HaplotypeCaller was used in 'gVCF' mode (Van der Auwera et al., 2013) to call SNVs for each individual dog, and then jointly across all 90 dogs. GATK best practices and default parameters, together with the initial alignment training sets, were

used for variant quality score recalibration of SNVs. Joint-called and compiled variant call files (VCFs) were filtered for CpG islands, gaps and repeats, as annotated in the CanFam3.1 reference dog genome assembly (<http://genome.ucsc.edu/cgi-bin/hgGateway?db=canFam3>). The resulting set of SNVs was used in subsequent analyses.

WGS-based analysis of homozygosity

The list of 7,095,427 SNVs from the 90 WGS dogs was pruned for excessive linkage using the 'indep' function of PLINK v1.07 (Purcell et al., 2007), with a window size of 50 SNPs, a window step of five SNPs and a variance inflation factor of 2. The pruned variant set of 1,510,327 SNPs was used to calculate RoHs and LnH for each WGS using the 'homozyg' function of PLINK (Purcell et al., 2007). To determine optimal conditions for which SNP and single-dog WGS homozygosity measurements are most comparable, homozygosity was calculated per dog from the WGS data with parameters set at a window size of 10 kb, 70 kb, 100 kb and 1000 kb, each with allowed heterozygosity of zero, one, five, or ten heterozygotes per window, equaling 16 total homozygosity conditions for each WGS. The scenario with 70 kb minimum LnH with zero allowed heterozygotes was designed to most closely mimic the parameters set for the SNP chip homozygosity analysis. Shared LnH and shared RoH were calculated for the breeds for which WGS was obtained from two dogs (Chow Chow, Bernese Mountain Dog, Greyhound, Rottweiler, Scottish Terrier, West Highland White Terrier) and three dogs (Flat-coated Retriever, Irish Water Spaniel) utilizing the same methodology and criteria implemented with the SNP data and applied to the pruned WGS.

Inbreeding coefficients from SNP and WGS data

Inbreeding coefficients were calculated for each of the dogs across 154,230 SNPs from the chip genotyping dataset and 1,510,327 SNPs from the pruned WGS data using the 'heterozygosity' function of PLINK v1.07 (Purcell et al., 2007). The within-breed means of the individual dog inbreeding coefficients were used to represent breed-specific inbreeding coefficients for each of the SNP analysis breeds, as well as for the WGS breeds for which more than one dog of a given breed was sequenced (Bernese Mountain Dog, Chow Chow, Flat-coated Retriever, Greyhound, Irish Water Spaniel, Rottweiler, Scottish Terrier, West Highland White Terrier).

Statistical analysis

The `cor.test` function of the Hmisc R package was used to calculate Pearson correlation statistics and significance values between SNP-based inbreeding coefficients, RoH and LnH, WGS-based inbreeding coefficients, RoH and LnH, and pedigree-based inbreeding coefficients. Correlation analyses utilized all breeds shared between each pair of data acquisition methods. Because F -values were calculated per individual within the WGS and SNP chip analyses, breed values were determined by averaging all contributing dogs of the same breed for each genotyping method. As such, a total of 11 breeds were represented in the pedigree data, 80 breeds in the SNP chip data, and 80 breeds in the WGS data. Fifty breeds were common to the WGS and SNP data, nine breeds to the WGS and pedigree data, and 11 breeds between SNP and pedigree data.

Acknowledgements

We thank Cord Drögemüller, Vidhya Jagannathan and Tosso Leeb for providing select dog genome sequences. We also thank the many dog owners for contributing DNA samples for research. Thanks to the following groups and individuals for providing the extensive parentage data utilized in the pedigree analysis: Nicholas at <http://www.papillonpedigrees.org>, Karen Berggren and the Portuguese Water Dog Study Group at <http://www.pwdinfo.com/psgdb/>, Eric Johnson and Tollerdata at <http://www.toller-l.org/tollerdata>, Blair Kelly at <http://www.shakspornorwich.net/pedigreedb/>, Amy Raby at <http://www.k9data.com>, Claire and Jim Trethewey at <http://www.alfirin.net/cdl/search.html>, Katherine Branson at <http://www.acdpedigree.com>, Gary Gulanas and Bernergarde at <http://www.bernergarde.org>, and Bonnie Dalzell at <http://www.borzoi-pedigree.batw.net/zoipedindex.htm>.

Competing interests

The authors declare no competing or financial interests.

Author contributions

D.L.D.: conceptualization, data curation, formal analysis, investigation, visualization, writing – original draft. M.R.: formal analysis, investigation, data curation, software. B.W.D.: data curation, methodology, software, writing – original draft. A.B.: conceptualization, data curation, formal analysis, investigation. H.G.P.: methodology, writing – review and editing, software. E.A.O.: funding acquisition, project administration, supervision, writing – review and editing.

Funding

This work was funded by the Intramural Program of the National Human Genome Research Institute.

Data availability

Whole-genome sequence data produced for this project has been submitted to Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra>) under BioProject PRJNA318762. Accession numbers for all genome sequences used in this study are available in Table S2. Raw SNP array data is available on Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) under BioProjects GSE83160 and GSE90441.

Supplementary information

Supplementary information available online at <http://dmm.biologists.org/lookup/doi/10.1242/dmm.027037.supplemental>

References

- Acland, G. M., Ray, K., Mellersh, C. S., Gu, W., Langston, A. A., Rine, J., Ostrander, E. A. and Aguirre, G. D. (1998). Linkage analysis and comparative mapping of canine progressive rod-cone degeneration (prcd) establishes potential locus homology with retinitis pigmentosa (RP17) in humans. *Proc. Natl. Acad. Sci. USA* **95**, 3048-3053.
- Auton, A., Rui Li, Y., Kidd, J., Oliveira, K., Nadel, J., Holloway, J. K., Hayward, J. J., Cohen, P. E., Grealis, J. M., Wang, J. et al. (2013). Genetic recombination is targeted towards gene promoter regions in dogs. *PLoS Genet.* **9**, e1003984.
- Axelsson, E., Ratnakumar, A., Arendt, M.-L., Maqbool, K., Webster, M. T., Perloski, M., Liberg, O., Arnemo, J. M., Hedhammar, Å. and Lindblad-Toh, K. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**, 360-364.
- Barbato, M., Orozco-terWengel, P., Tapio, M. and Bruford, M. W. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front. Genet.* **6**, 109.
- Benson, K. F., Li, F.-Q., Person, R. E., Albani, D., Duan, Z., Wechsler, J., Meade-White, K., Williams, K., Acland, G. M., Niemeyer, G. et al. (2003). Mutations associated with neutropenia in dogs and humans disrupt intracellular transport of neutrophil elastase. *Nat. Genet.* **35**, 90-96.
- Björnerfeldt, S., Hailer, F., Nord, M. and Vilà, C. (2008). Assortative mating and fragmentation within dog breeds. *BMC Evol. Biol.* **8**, 28.
- Boichard, D. (2002). PEDIG: a fortran package for pedigree analysis suited for large populations. In 7th World Congress on Genetics Applied to Livestock Production: Montpellier, France.
- Boichard, D., Maignel, L. and Verrier, E. (1997). The value of using probabilities of gene origin to measure genetic variability in a population. *Genet. Sel. Evol.* **29**, 5-23.
- Boyko, A. R., Quignon, P., Li, L., Schoenebeck, J. J., Degenhardt, J. D., Lohmueller, K. E., Zhao, K., Brisbin, A., Parker, H. G., vonHoldt, B. M. et al. (2010). A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol.* **8**, e1000451.
- 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.
- Cattanach, B. M., Dukes-McEwan, J., Wotton, P. R., Stephenson, H. M. and Hamilton, R. M. (2015). A pedigree-based genetic appraisal of Boxer ARVC and the role of the Striatin mutation. *Vet. Rec.* **176**, 492.
- Club, T. A. K. (2006). *The Complete Dog Book*. New York: Ballantine Books.
- Cooper, A. E., Ahonen, S., Rowlan, J. S., Duncan, A., Seppälä, E. H., Vanhapelto, P., Lohi, H. and Komáromy, A. M. (2014). A novel form of progressive retinal atrophy in Swedish vallhund dogs. *PLoS ONE* **9**, e106610.
- Decker, B., Davis, B. W., Rimbault, M., Long, A. H., Karlins, E., Jagannathan, V., Reiman, R., Parker, H. G., Drögemüller, C., Corneveaux, J. J. et al. (2015). Comparison against 186 canid whole-genome sequences reveals survival strategies of an ancient clonally transmissible canine tumor. *Genome Res.* **25**, 1646-1655.
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M. et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491-498.
- Dodman, N. H., Karlsson, E. K., Moon-Fanelli, A., Galdzicka, M., Perloski, M., Shuster, L., Lindblad-Toh, K. and Ginns, E. I. (2010). A canine chromosome 7 locus confers compulsive disorder susceptibility. *Mol. Psychiatry* **15**, 8-10.
- Drögemüller, C., Karlsson, E. K., Hytönen, M. K., Perloski, M., Dolf, G., Sainio, K., Lohi, H., Lindblad-Toh, K. and Leeb, T. (2008). A mutation in hairless dogs implicates FOXI3 in ectodermal development. *Science* **321**, 1462.
- Freedman, A. H., Gronau, I., Schweizer, R. M., Ortega-Del Vecchyo, D., Han, E., Silva, P. M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B. et al. (2014). Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet.* **10**, e1004016.
- Gou, X., Wang, Z., Li, N., Qiu, F., Xu, Z., Yan, D., Yang, S., Jia, J., Kong, X., Wei, Z. et al. (2014). Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. *Genome Res.* **24**, 1308-1315.
- Gray, M. M., Granka, J. M., Bustamante, C. D., Sutter, N. B., Boyko, A. R., Zhu, L., Ostrander, E. A. and Wayne, R. K. (2009). Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics* **181**, 1493-1505.
- Hayward, J. J., Castelhamo, 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.
- Jónasdóttir, T. J., Mellersh, C. S., Moe, L., Heggebo, R., Gamlem, H., Ostrander, E. A. and Lingaas, F. (2000). Genetic mapping of a naturally occurring hereditary renal cancer syndrome in dogs. *Proc. Natl. Acad. Sci. USA* **97**, 4132-4137.
- Karlsson, E. K., Baranowska, I., Wade, C. M., Salmon Hillbertz, N. H., Zody, M. C., Anderson, N., Biagi, T. M., Patterson, N., Pielberg, G. R., Kulbokas, E. J. III et al. (2007). Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat. Genet.* **39**, 1321-1328.
- Lacy, R. C. (1989). Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalents. *Zoo. Biol.* **8**, 111-123.
- Lappalainen, A. K., Mäki, K. and Laitinen-Vapaavuori, O. (2015). Estimate of heritability and genetic trend of intervertebral disc calcification in Dachshunds in Finland. *Acta Vet. Scand.* **57**, 78.
- Lequarré, A.-S., Andersson, L., André, C., Fredholm, M., Hitte, C., Leeb, T., Lohi, H., Lindblad-Toh, K. and Georges, M. (2011). LUPA: a European initiative taking advantage of the canine genome architecture for unravelling complex disorders in both human and dogs. *Vet. J.* **189**, 155-159.
- Lewis, T. W., Abhayaratne, B. M. and Blott, S. C. (2015). Trends in genetic diversity for all Kennel Club registered pedigree dog breeds. *Canine Genet. Epidemiol.* **2**, 13.
- Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079.
- 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., III, Zody, M. C. et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803-819.
- Marsden, C. D., Ortega-Del Vecchyo, D., O'Brien, D. P., Taylor, J. F., Ramirez, O., Vilà, C., Marques-Bonet, T., Schnabel, R. D., Wayne, R. K. and Lohmueller, K. E. (2016). Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl. Acad. Sci. USA* **113**, 152-157.
- Mellersh, C. S., Langston, A. A., Acland, G. M., Fleming, M. A., Ray, K., Wiegand, N. A., Francisco, L. V., Gibbs, M., Aguirre, G. D. and Ostrander, E. A. (1997). A linkage map of the canine genome. *Genomics* **46**, 326-336.
- Neff, M. W., Broman, K. W., Mellersh, C. S., Ray, K., Acland, G. M., Aguirre, G. D., Ziegler, J. S., Ostrander, E. A. and Rine, J. (1999). A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. *Genetics* **151**, 803-820.
- Parker, H. G., VonHoldt, B. M., Quignon, P., Margulies, E. H., Shao, S., Mosher, D. S., Spady, T. C., Elkhahloun, 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., Shearin, A. L. and Ostrander, E. A. (2010). Man's best friend becomes biology's best in show: genome analyses in the domestic dog. *Annu. Rev. Genet.* **44**, 309-336.
- Pedersen, N. C., Brucker, L., Tessier, N. G., Liu, H., Penedo, M. C. T., Hughes, S., Oberbauer, A. and Sacks, B. (2015). The effect of genetic bottlenecks and inbreeding on the incidence of two major autoimmune diseases in standard poodles, sebaceous adenitis and Addison's disease. *Canine Genet. Epidemiol.* **2**, 14.
- Persson, M. E., Roth, L. S. V., Johnsson, M., Wright, D. and Jensen, P. (2015). Human-directed social behaviour in dogs shows significant heritability. *Genes Brain Behav.* **14**, 337-344.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J. et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575.

- 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.
- Reist-Marti, S. B., Dolf, G., Leeb, T., Kottmann, S., Kietzmann, S., Butenhoff, K. and Rieder, S. (2012). Genetic evidence of subaortic stenosis in the Newfoundland dog. *Vet. Rec.* **170**, 597.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular cloning: A laboratory manual*. NY: Cold Spring Harbor Laboratory Press.
- Schlamp, F., van der Made, J., Stambler, R., Chesebrough, L., Boyko, A. R. and Messer, P. W. (2016). Evaluating the performance of selection scans to detect selective sweeps in domestic dogs. *Mol. Ecol.* **25**, 342-356.
- Schoenebeck, J. J. and Ostrander, E. A. (2014). Insights into morphology and disease from the dog genome project. *Annu. Rev. Cell Dev. Biol.* **30**, 535-560.
- Schoenebeck, J. J., Hutchinson, S. A., Byers, A., Beale, H. C., Carrington, B., Faden, D. L., Rimbault, M., Decker, B., Kidd, J. M., Sood, R. et al. (2012). Variation of BMP3 contributes to dog breed skull diversity. *PLoS Genet.* **8**, e1002849.
- 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.
- Tiira, K., Hakosalo, O., Kareinen, L., Thomas, A., Hielm-Björkman, A., Escriou, C., Arnold, P. and Lohi, H. (2012). Environmental effects on compulsive tail chasing in dogs. *PLoS ONE* **7**, e41684.
- Todhunter, R. J., Mateescu, R., Lust, G., Burton-Wurster, N. I., Dykes, N. L., Bliss, S. P., Williams, A. J., Vernier-Singer, M., Corey, E., Harjes, C. et al. (2005). Quantitative trait loci for hip dysplasia in a cross-breed canine pedigree. *Mamm. Genome* **16**, 720-730.
- Våge, J., Wade, C., Biagi, T., Fatjó, J., Amat, M., Lindblad-Toh, K. and Lingaas, F. (2010). Association of dopamine- and serotonin-related genes with canine aggression. *Genes Brain Behav.* **9**, 372-378.
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J. et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11 10 1-33.
- Vaysse, A., Ratnakumar, A., Derrien, T., Axelsson, E., Rosengren Pielberg, G., Sigurdsson, S., Fall, T., Seppälä, E. H., Hansen, M. S. T., Lawley, C. T. et al. (2011). Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet.* **7**, e1002316.
- 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., Fan, R.-X., Cao, X., Zhong, L., Wang, L., Liu, F., Wu, H. and Cheng, L.-G. (2013). The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat. Commun.* **4**, 1860.
- Wayne, R. K. and vonHoldt, B. M. (2012). Evolutionary genomics of dog domestication. *Mamm. Genome* **23**, 3-18.
- Wiggans, G. R., VanRaden, P. M. and Zuurbier, J. (1995). Calculation and use of inbreeding coefficients for genetic evaluation of United States dairy cattle. *J. Dairy Sci.* **78**, 1584-1590.
- Wijnrocx, K., Francois, L., Stinckens, A., Janssens, S. and Buys, N. (2016). Half of 23 Belgian dog breeds has a compromised genetic diversity, as revealed by genealogical and molecular data analysis. *J. Anim. Breed Genet.* **133**, 375-383.
- Wilson, B. J., Nicholas, F. W., James, J. W., Wade, C. M. and Thomson, P. C. (2013). Estimated breeding values for canine hip dysplasia radiographic traits in a cohort of Australian German Shepherd dogs. *PLoS ONE* **8**, e77470.
- Windig, J. J. and Oldenbroek, K. (2015). Genetic management of Dutch golden retriever dogs with a simulation tool. *J. Anim. Breed Genet.* **132**, 428-440.
- Wolf, Z. T., Leslie, E. J., Arzi, B., Jayashankar, K., Karmi, N., Jia, Z., Rowland, D. J., Young, A., Safra, N., Sliskovic, S. et al. (2014). A LINE-1 insertion in DLX6 is responsible for cleft palate and mandibular abnormalities in a canine model of Pierre Robin sequence. *PLoS Genet.* **10**, e1004257.
- Wong, A. K. and Neff, M. W. (2009). DOGSET: pre-designed primer sets for fine-scale mapping and DNA sequence interrogation in the dog. *Anim. Genet.* **40**, 569-571.
- Wong, A. K., Ruhe, A. L., Dumont, B. L., Robertson, K. R., Guerrero, G., Shull, S. M., Ziegler, J. S., Millon, L. V., Broman, K. W., Payseur, B. A. et al. (2010). A comprehensive linkage map of the dog genome. *Genetics* **184**, 595-605.
- Yuzbasiyan-Gurkan, V., Blanton, S. H., Cao, V., Ferguson, P., Li, J., Venta, P. J. and Brewer, G. J. (1997). Linkage of a microsatellite marker to the canine copper toxicosis locus in Bedlington terriers. *Am. J. Vet. Res.* **58**, 23-27.

Table S1: Inbreeding coefficients calculated from SNP data, whole genome sequencing (WGS), and pedigree analysis of entire reference pedigrees, 10-generation pedigrees, and 5-generation pedigrees. Percent identity to the Boxer reference is the mean of all SNP-genotyped dogs of each breed.

Breed	Abbr.	SNP	WGS	PED _{all}	PED _{10-gen}	PED _{5-gen}
Afghan Hound	AFGH	0.332				
Airedale Terrier	AIRT		0.467			
Akita	AKIT	0.347				
Alaskan Malamute	AMAL	0.313	0.199			
American Cocker Spaniel	ACKR	0.378	0.378			
American Rat Terrier	ARTR		0.383			
Australian Cattle Dog	ACD		0.185	0.067	0.064	0.038
Australian Shepherd	AUSS	0.207				
Australian Terrier	AUST	0.260				
Basenji	BSJI	0.536	0.571	0.221	0.118	0.059
Basset Hound	BASS	0.357				
Beagle	BEAG	0.244	0.084			
Bearded Collie	BRDC		0.261			
Belgian Malinois	BMAL		0.265			
Belgian Shepherd	BELS	0.301	0.286	0.193	0.126	0.064
Belgian Tervuren	TERV	0.264	0.357			
Bernese Mountain Dog	BMD	0.350	0.314	0.197	0.061	0.022
Black Russian Terrier	BRTR		0.263			
Bloodhound	BLDH	0.431	0.316			
Border Collie	BORD	0.210	0.254			
Border Terrier	BORT		0.379			
Borzoi	BORZ	0.311	0.265	0.128	0.086	0.054
Boston Terrier	BOST	0.225	0.316			
Bouvier des Flandres	BOUV		0.314			
Boxer	BOX	0.357				
Briard	BRIA	0.298				
Brittany	BRIT	0.263	0.331			
Bull Mastiff	BULM	0.345				
Bull Terrier	BULT	0.579				
Bulldog	BULD	0.373	0.454			
Cairn Terrier	CAIR	0.193				
Cardigan Welsh Corgi	CARD	0.250				
Cavalier King Charles Spaniel	CKCS	0.421				
Chihuahua	CHIH	0.113	0.099			
Chinese Crested	CRES		0.374			
Chinese Shar Pei	SHAR	0.236				
Chow Chow	CHOW	0.370	0.374			
Clumber Spaniel	CLSP		0.475			
Collie	COLL	0.481				
Dachshund	DACH	0.203	0.422			
Doberman Pinscher	DOBP	0.431	0.385			

English Cocker Spaniel*	ECKR	0.390	0.494			
English Springer Spaniel	ESSP	0.312	0.352			
Eurasier	EURA		0.254			
Flat Coated Retriever	FCR	0.354	0.324			
French Bulldog	FBUL	0.281	0.287			
German Shepherd Dog	GSD	0.373	0.479			
German Shorthaired Pointer	GSHP	0.176				
German Wirehaired Pointer	GWHP		0.223			
Giant Schnauzer	GSNZ	0.252				
Glen of Imaal Terrier*	GLEN	0.277				
Golden Retriever	GOLD	0.284	0.218	0.160	0.079	0.027
Great Dane	DANE	0.272	0.377			
Great Pyrenees	GPYR	0.350	0.381			
Greater Swiss Mountain Dog	GSMD		0.420			
Greyhound	GREY	0.300	0.314			
Havanese	HAVA	0.178				
Ibizan Hound	IBIZ	0.289				
Irish Setter	ISET		0.319			
Irish Terrier	IRTR		0.322			
Irish Water Spaniel	IWSP	0.267	0.323			
Irish Wolfhound	IWOF	0.426	0.399			
Italian Greyhound	ITGY	0.292	0.338			
Jack Russell Terrier	JRT		0.117			
Kerry Blue Terrier	KERY		0.425			
Komondor	KOMO		0.316			
Kromforlander	KROM		0.468			
Kuvasz	KUVZ	0.171				
Labrador Retriever	LAB	0.217	0.211	0.082	0.073	0.026
Lagotto Romagnolo	LAGO		0.378			
Leonberger	LEON	0.316				
Mastiff	MAST	0.335	0.270			
Miniature Bull Terrier	MBLT	0.507				
Miniature Pinscher	MPIN	0.307				
Miniature Poodle	MPOO	0.237	0.255			
Miniature Schnauzer	MSNZ	0.452	0.564			
Newfoundland	NEWF	0.288				
Norwegian Lundehund	NLUN		0.868			
Norwich Terrier	NOWT	0.408		0.267	0.167	0.057
Nova Scotia Duck Tolling Retriever	NSDT		0.205	0.266	0.251	0.034
Old English Sheepdog	OES	0.276				
Papillon	PAPI	0.179		0.059	0.051	0.031
Pekingese	PEKE	0.356	0.379			

Pembroke Welsh Corgi	PEMB	0.301	0.329			
Petit Basset Griffon Vendeen	PBGV	0.236				
Pointer	PNTR		0.255			
Pomeranian	POM	0.180				
Portuguese Podengo	POPO		0.166			
Portuguese Water Dog	PTWD	0.270	0.118	0.176	0.162	0.052
Pug	PUG	0.442	0.383			
Rhodesian Ridgeback	RHOD		0.331			
Rottweiler	ROTT	0.337	0.280			
Saint Bernard	STBD	0.296	0.214			
Saluki	SALU	0.234	0.159			
Samoyed	SAMO	0.301	0.281			
Scottish Deerhound	DEER	0.459	0.420			
Scottish Terrier	SCOT	0.346	0.322			
Shetland Sheepdog	SSHP	0.386	0.354			
Shih Tzu	SHIH	0.325				
Siberian Husky	HUSK	0.327	0.100			
Sloughi	SLOU		0.065			
Soft Coated Wheaten Terrier	SCWT		0.474			
Spanish Water Dog	SPWD		0.117			
Staffordshire Bull Terrier	STAF	0.299				
Standard Poodle	SPOO	0.181	0.249			
Standard Schnauzer	SSNZ	0.255	0.283			
Tibetan Mastiff	TIBM		0.118			
Tibetan Terrier	TIBT		0.140			
Toy Poodle	TPOO	0.182	0.117			
West Highland White Terrier	WHWT	0.357	0.419			
Whippet	WHIP	0.324				
Yorkshire Terrier	YORK	0.254	0.147			

* SNP F-values for ECKR and GLEN were calculated from 9 individuals instead of 1

Table S2: Whole genome sequence quality metrics and sources

Name	Variants	Average Depth	Source
ACKR01	5,054,866	33.713	Decker et al. 2015
ARTR01	5,099,158	40.002	PRJNA318762
AIRT02	5,204,116	28.623	PRJNA263947
AMAL01	4,736,533	37.749	PRJNA318762
ACD01	4,935,757	17.836	PRJEB13468
BEAG01	4,797,005	40.359	Decker et al. 2015
BELS01	4,973,385	35.936	Decker et al. 2015
BLDH01	4,999,169	33.431	Decker et al. 2015
BMAL01	5,136,968	9.438	Wang et al. 2013
BMD01	4,979,710	36.058	Decker et al. 2015
BMD02	5,065,175	39.096	PRJNA318762
BORD02	5,031,466	16.033	Decker et al. 2015
BORT01	5,113,993	21.11	PRJNA263947
BORZ01	4,962,089	40.516	PRJNA318762
BOST01	5,151,014	41.779	PRJNA318762
BOUV01	5,027,893	46.821	PRJNA318762
BRDC01	5,016,518	17.328	PRJEB13468
BRIT01	5,243,695	8.478	Decker et al. 2015
BRTR01	5,026,403	29.47	PRJNA263947
BSJI01	5,386,656	5.201	Freedman et al. 2014
BULD01	5,327,606	39.587	Decker et al. 2015
CHIH01	4,796,797	41.391	Decker et al. 2015
CHOW01	5,489,944	4.457	Decker et al. 2015
CHOW02	4,733,596	46.627	Decker et al. 2015
CLSP01	4,830,777	15.489	Decker et al. 2015
CRES03	4,696,633	19.026	PRJNA261736
DACH01	5,173,689	15.057	Decker et al. 2015
DANE01	5,093,550	43.995	Decker et al. 2015
DEER01	5,133,707	22.681	PRJNA263947
DOBP01	4,690,581	26.809	Decker et al. 2015
ECKR02	5,527,883	6.285	PRJNA263947
ESSP01	5,084,780	25.765	PRJNA263947
EURA01	5,002,113	10.735	PRJEB6079
FBUL01	5,173,531	18.228	PRJEB13468
FCR01	4,999,924	47.452	Decker et al. 2015
FCR02	5,038,664	36.299	Decker et al. 2015
FCR03	5,063,014	36.002	PRJNA318762
GOLD01	4,977,578	30.888	Decker et al. 2015
GPYR01	5,146,474	16.122	Decker et al. 2015
GREY07	5,045,975	36.383	PRJNA247491
GREY01	5,221,305	14.221	Decker et al. 2015

GSD01	5,354,584	8.636	Wang et al. 2013
GSMD01	5,117,279	36.148	PRJNA318762
GWHP01	4,954,916	20.896	PRJEB13468
HUSK01	4,648,407	46.636	Decker et al. 2015
IRTR01	5,094,366	18.957	PRJEB13468
ISET01	5,007,794	38.254	PRJNA318762
ITGY01	5,091,056	20.088	Decker et al. 2015
IWOF01	5,111,487	34.234	Decker et al. 2015
IWSP01	4,965,143	42.841	PRJNA318762
IWSP02	5,005,180	42.469	PRJNA318762
IWSP03	5,109,996	41.355	PRJNA318762
JRT01	4,921,017	24.849	PRJNA263947
KERY01	5,216,770	22.316	PRJNA263947
KOMO01	4,998,899	39.872	PRJNA318762
KROM01	5,200,404	22.237	PRJEB6076
LAB04	5,000,603	28.266	PRJNA263947
LAGO01	5,479,537	6.462	PRJEB13468
MAST01	5,041,589	50.977	Decker et al. 2015
MPOO01	4,946,955	41.789	PRJNA318762
MSNZ01	5,708,610	5.265	PRJNA263947
NLUN01	5,546,324	20.074	PRJNA186960
NSDT01	4,929,427	21.895	PRJNA263947
PEKE01	4,995,474	44.938	Decker et al. 2015
PEMB02	5,086,521	21.515	PRJNA263947
PNTR01	5,036,900	18.147	Decker et al. 2015
POPO01	4,909,232	23.586	PRJNA263947
PTWD01	4,759,535	36.337	PRJNA318762
PUG05	5,251,865	12.06	Decker et al. 2015
RHOD01	5,086,725	18.091	Decker et al. 2015
ROTT01	4,968,608	33.229	Decker et al. 2015
ROTT04	5,021,172	40.855	PRJNA318762
SALU01	4,750,584	46.554	Decker et al. 2015
SAMO01	4,892,384	40.68	PRJNA318762
SCOT01	5,097,777	39.698	Decker et al. 2015
SCOT02	4,914,841	37.802	PRJNA318762
SCWT01	5,350,393	14.183	PRJNA263947
SLOU01	4,740,780	19.944	PRJNA318762
SPOO01	5,090,679	11.104	Decker et al. 2015
SPWD01	4,888,534	14.514	PRJEB7903
SSHP01	5,028,406	36.223	Decker et al. 2015
SSNZ01	5,070,274	14.846	PRJNA263947
STBD01	4,938,360	34.199	Decker et al. 2015
TERV01	5,053,248	32.35	Decker et al. 2015

TIBM01	4,849,992	10.493	Wang et al. 2013
TIBT01	4,818,133	24.428	PRJNA263947
TPOO01	4,825,194	41.563	Decker et al. 2015
WHWT01	5,278,603	9.701	Decker et al. 2015
WHWT02	5,122,584	37.589	PRJNA318762
YORK01	4,924,101	19.177	PRJNA318762

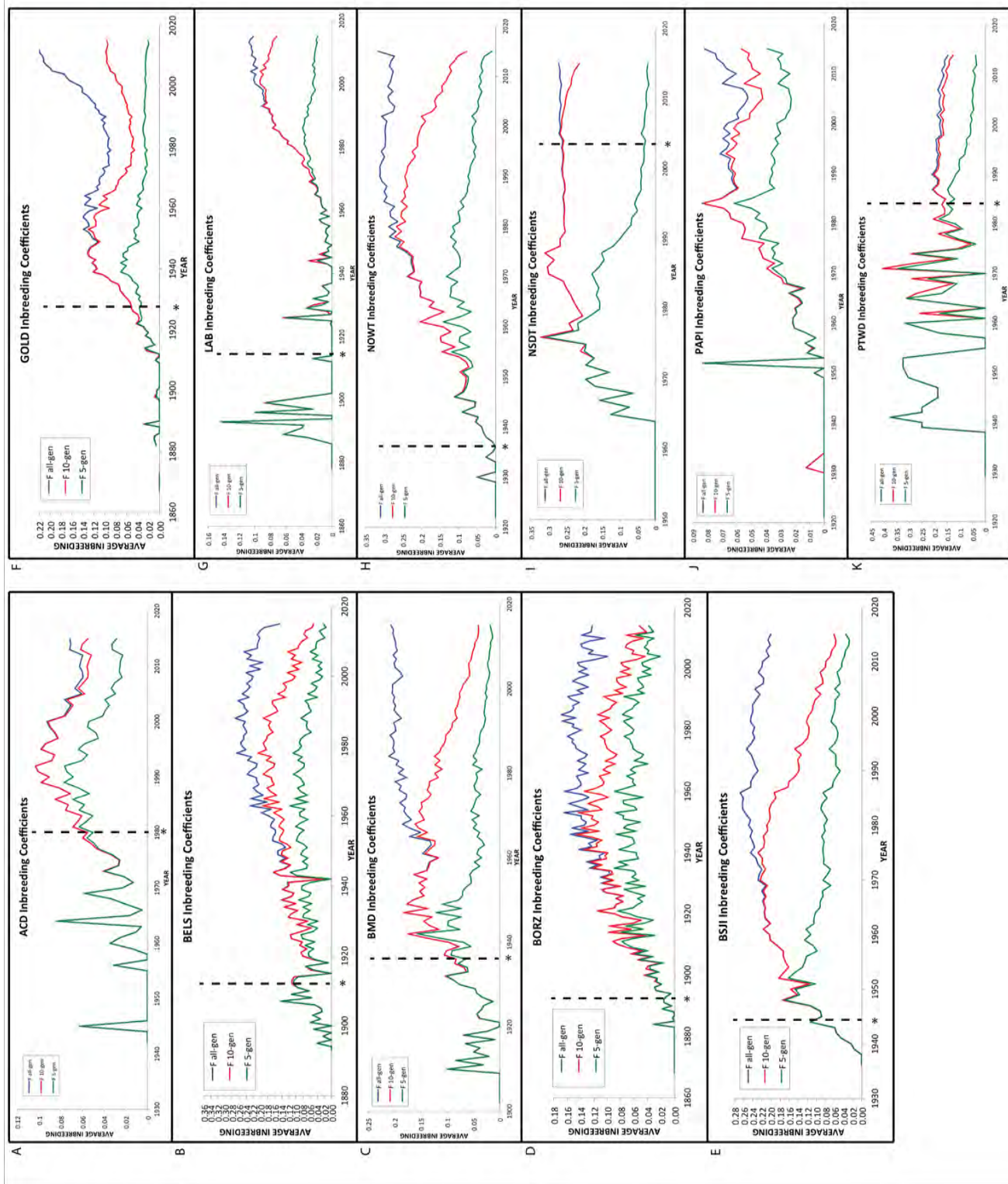


Figure S1: Inbreeding coefficients over time for 11 pedigree breeds. Calculations of F completed using the entire reference pedigree, or the most recent 10 or 5 generations.

* indicate the year at which the breed was recognized by the AKC. Australian Cattle Dog (ACD) = 1980, Belgian Sheepdog (BELS) = 1912, Bernese Mountain Dog (BMD) = 1937, Borzoi (BORZ) = 1891, Basenji (BSJI) = 1944, Golden Retriever (GOLD) = 1920, Labrador Retriever (LAB) = 1917, Norwich Terrier (NOWT) = 1936, Nova Scotia Duck Tolling Retriever (NSDT) = 2003, Papillon (PAPI) = 1915, Portuguese Water Dog (PTWD) = 1983.

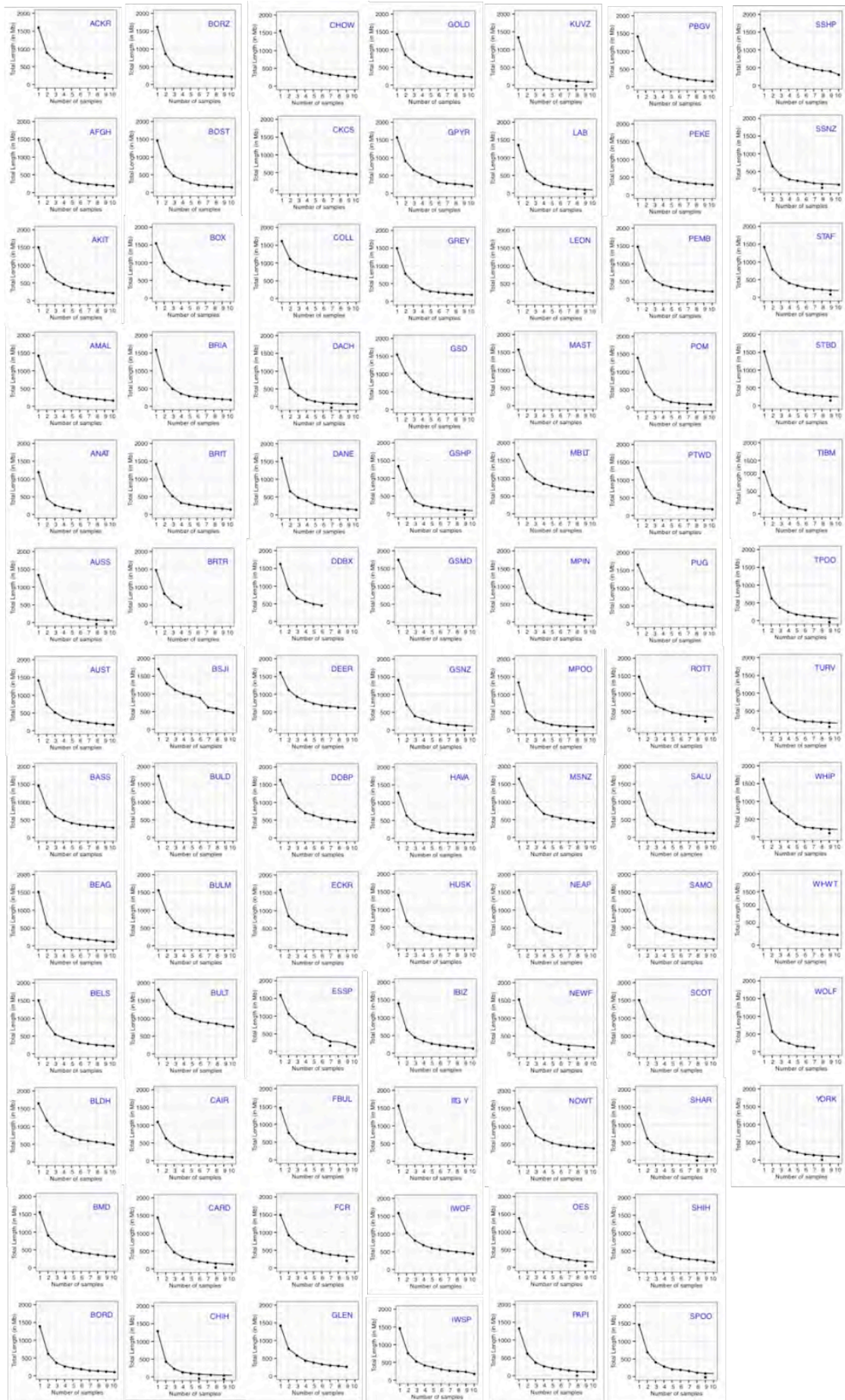


Figure S2: Single breed graphs displaying the decrease in shared LNH with SNP chip data from an additional same-breed dog.