

# Single-Nucleotide-Polymorphism-Based Association Mapping of Dog Stereotypes

Paul Jones,\* Kevin Chase,<sup>†</sup> Alan Martin,\* Plus Davern,<sup>‡</sup> Elaine A. Ostrander<sup>§</sup> and Karl G. Lark<sup>†,1</sup>

\*The WALTHAM Centre for Pet Nutrition, Waltham on the Wolds, Leicestershire LE14 4RT, United Kingdom,

<sup>†</sup>Department of Biology, University of Utah, Salt Lake City, Utah 84112, <sup>‡</sup>Sundowners Kennels, Gilroy,

California 95020 and <sup>§</sup>National Human Genome Research Institute, National

Institutes of Health, Bethesda, Maryland 20892

Manuscript received February 5, 2008

Accepted for publication March 12, 2008

## ABSTRACT

Phenotypic stereotypes are traits, often polygenic, that have been stringently selected to conform to specific criteria. In dogs, *Canis familiaris*, stereotypes result from breed standards set for conformation, performance (behaviors), etc. As a consequence, phenotypic values measured on a few individuals are representative of the breed stereotype. We used DNA samples isolated from 148 dog breeds to associate SNP markers with breed stereotypes. Using size as a trait to test the method, we identified six significant quantitative trait loci (QTL) on five chromosomes that include candidate genes appropriate to regulation of size (*e.g.*, *IGF1*, *IGF2BP2*, *SMAD2*, etc.). Analysis of other morphological stereotypes, also under extreme selection, identified many additional significant loci. Less well-documented data for behavioral stereotypes tentatively identified loci for herding, pointing, boldness, and trainability. Four significant loci were identified for longevity, a breed characteristic not under direct selection, but inversely correlated with breed size. The strengths and limitations of the approach are discussed as well as its potential to identify loci regulating the within-breed incidence of specific polygenic diseases.

THE dog, “man’s best friend,” shares a large number of the complex phenotypes observed in human populations, including variation in morphology and behavior, as well as many types of polygenic disease. In the past decade, *Canis familiaris* has emerged as an excellent system for genetic analysis of complex phenotypes. Most of the advantages offered by the canine system over other mammalian systems derive from its population structure (OSTRANDER and KRUGLYAK 2000; SUTTER *et al.* 2004; PARKER and OSTRANDER 2005; GOLDSTEIN *et al.* 2006; KARLSSON *et al.* 2007; PARKER *et al.* 2007). There are >350 distinct breeds recognized in the world today, many of which are isolates that have been, for the most part, selected for morphology and behavior. Over hundreds of years humans and dogs have formed a multitude of mutualistic relationships harnessing the phenotypic flexibility of the dog genome. New dog breeds were often developed by crossing individuals of unique breeds bearing desired features, followed by strong selection for the desired phenotypes (hunting ability, coat color, skull shape, body size, etc.), thus increasing the frequency of selected genotypes in the modern-day population. To be a registered member of a breed, both parents of an individual must have been registered members of the same breed. As a result, genetic heterogeneity is reduced within breeds, but is high across breeds (PARKER *et al.*

2004; LINDBLAD-TOH *et al.* 2005). Consequently, many phenotypes are either fixed or close to fixation in a large number of populations.

Genetic isolates have provided the key analyses of complex polygenic disease (LINDBLAD-TOH *et al.* 2005; GOLDSTEIN *et al.* 2006; KARLSSON *et al.* 2007; PARKER *et al.* 2007) as well other phenotypes. However, the use of large numbers of such isolates has not, to date, been applied to allele trait association. The dog presents a unique opportunity to examine the power of this approach. Dog breeds, in which regions of the genome are “fixed,” can be treated in a manner similar to recombinant inbred populations: “Fixed” portions of a breed’s genome will remain invariant as long as the breeding population remains closed. These “fixed” aspects will continue to produce consistent phenotypes, and therefore the phenotype and genotype need not be measured on the same animal. Thus, both the allele frequency of a single nucleotide polymorphism (SNP) in fixed regions of the genome and the phenotype are characteristics of a breed. As a result, associating breed-specific genotypes with “fixed” phenotypes in multiple breeds (across-breed mapping) presents a powerful tool for identifying quantitative trait loci (QTL) that may form the genetic basis for the phenotypic diversity observed in dog breeds.

Similar approaches have been described using inbred mouse strains (GRUPE *et al.* 2001; LIAO *et al.* 2004; PLETCHER *et al.* 2004; WANG *et al.* 2005), and these have

<sup>1</sup>Corresponding author: Department of Biology, University of Utah, 257 S. 1400 East, Room 201, Salt Lake City, UT 84102.  
E-mail: lark@bioscience.utah.edu

been combined with classical QTL analysis (PARK *et al.* 2003; DIPETRILLO *et al.* 2004; WANG *et al.* 2004; CERVINO *et al.* 2005). However, the number of inbred mouse lines available are far fewer than the number of dog breeds, and the number of phenotypes offered by mice much fewer than what is offered by the nearly 300 breeds of domestic dog. Moreover, the genome structure of any inbred mouse line is far more restrictive than the genomes that characterize a dog breed. Genomes of dog breeds have far more heterozygosity and have survived for centuries in quite variable environments. In short, the selective environments experienced by any dog breed have been far less restricted than those used during the inbreeding procedures that give rise to an inbred mouse.

Ideally, two types of data are required for across-breed association analysis: (1) a common set of well-distributed, highly informative SNPs that characterize the entire genome for each of many breeds and (2) a careful quantitative evaluation of the fixed phenotypes associated with each breed. The phenotypes most amenable to this mapping strategy are those that have been under stringent selection, such as morphology and behavior. Here we analyze the genetic basis for size using across-breed mapping and then present examples of the technique applied to other classes of traits: additional morphological features, behavior, and the relationship between size and longevity.

## MATERIALS AND METHODS

A total of 148 domestic dog breeds were characterized for a variety of sex-averaged phenotypes: height, weight, other morphology characters, longevity, and behavior. Phenotypic values used for the different breeds are summarized in supplemental Table 1. Height at the withers and weight were obtained from the published American Kennel Club (AKC) breed standards (AMERICAN KENNEL CLUB 1998). The residuals from the regression of  $WT^{0.33}$  onto height were derived and used as a measure of shape (*e.g.*, breeds that are heavier or lighter than other breeds of the same height; see supplemental Figure 1). "Short coat" (WILCOX and WALKOWICZ 1995) was coded as a qualitative variable: 1 for all breeds with a very short coat as the standard and 0 for all others. "Ear bend" (WILCOX and WALKOWICZ 1995) was scored as the degree of bend in the ear on a scale from 1 (hanging low) to 4 (completely erect; cropped ears were not scored). "Tail curve" (WILCOX and WALKOWICZ 1995) was scored as the degree of curve in the tail on a scale of 1 (straight) to 5 (tightly curled). Additional phenotypes were measured from breed pictures (PALMER 1994; WILCOX and WALKOWICZ 1995; <http://images.google.com/>) using the metrics described in Figure 1. Because the pictures utilized were not standardized, only ratios of these metrics could be used. The following ratios were defined using the metrics in Figure 1: (1) snout:head [ $a/(a + b)$ ]; (2) snout height:head [ $c/(a + b)$ ]; (3) head:body [ $(a + b)/e$ ]; (4) leg:body [ $(h + i)/e$ ]; (5) tail:body [ $f/e$ ]; (6) neck:body [ $j/e$ ]; and (7) chest:body [ $g/e$ ].

Longevity data (supplemental Table 1) were compiled from a variety of sources (MICHELL 1999; <http://users.pullman.com/lostriver/longhome.htm>; KC/BSAVA 2004; EGENVALL *et al.* 2005). These represent data primarily from owner surveys.

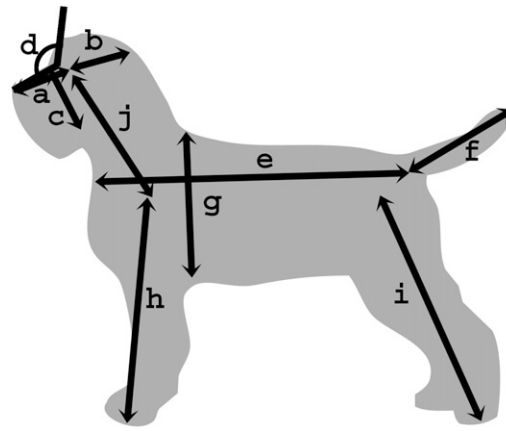


FIGURE 1.—Paths used to measure metrics of different breed characteristics. Shape components of morphology were scored referencing breed standards and pictures of purebred show dogs. The metrics shown above were measured using the "path" tool of Adobe Photoshop on side-view pictures: (a) tip of nose to eye; (b) eye to back of head; (c) top of snout to bottom of snout (perpendicular to the snout at the plane where the snout meets the face, adjusted for open mouths or long hair on the snout); (d) angle between the top of the snout and the forehead; (e) from breast bone to the base of the tail; (f) from the base of the tail to the tip of the tail, compensating for the tail curve; (g) from back to chest immediately behind the foreleg; (h) forefoot to shoulder socket; (i) from rear foot to hip joint; and (j) from eye to shoulder joint.

One of us, Pluis Davern, an experienced dog trainer and judge (<http://www.sundownerskennels.com/training.html>; <http://www.infodog.com/judges/17422/juddat.htm>; [http://www.akc.org/breeders/resp\\_breeding/Articles/truetoform.cfm](http://www.akc.org/breeders/resp_breeding/Articles/truetoform.cfm)), scored behavioral phenotypes as qualitative variables (0, 1, or NA). Four distinguishing patterns of dog behavior were scored: pointing, herding, boldness, and trainability. Additional behavioral data were taken from HART and MILLER (1985). Behavioral scores for the 148 breeds are in supplemental Table 1.

**DNA collection and isolation:** DNA samples were collected from dogs participating in AKC or otherwise sanctioned events, including dog shows, performance events, and obedience and behavior trials. Samples were collected as either whole blood or by cheek swab by registered veterinarians or licensed veterinary technicians after obtaining the owner's written consent. AKC or other registration numbers were collected on each dog, as was owner contact information, pedigree data, health history, and when possible, permission to recontact owners regarding future queries was also obtained. Wherever possible, care was taken to obtain samples from dogs unrelated at the grandparent level.

Blood samples were collected as whole blood in acid citrate dextrose or EDTA anticoagulation tubes. Buccal swabs were collected using standard protocols with Cytosoft cytology brushes (Medical Packaging, Camarillo, CA). DNA was extracted from the brushes using a QIAamp blood mini kit (QIAGEN, Valencia, CA) following the manufacturer's protocol. DNA was extracted from the blood samples using a standard phenol/chloroform extraction method (MANIATIS *et al.* 1982). Coded samples were aliquoted and stored for long-term use at  $-70^{\circ}$ . Information was entered into a MySQL custom database.

All procedures were performed in accordance with approvals from the Animal Care and Use Committees from the University of Utah, the National Human Genome Research Institute at the National Institutes of Health, and the Waltham Centre for Pet Nutrition, Mars.

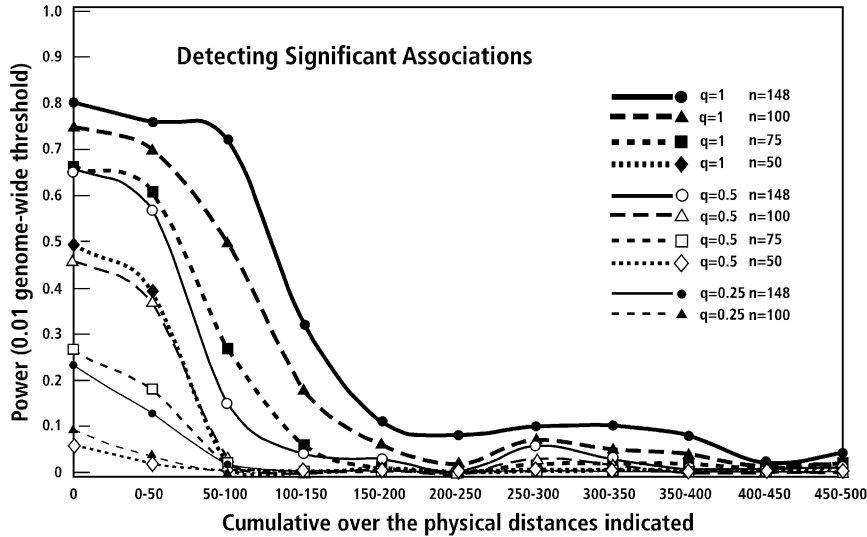


FIGURE 2.—Probability of detecting allele associations between two SNPs as a function of (1) the physical distance between the two markers ( $x$ -axis), (2) the number of breeds sampled ( $n = 148, 100, 75,$  and  $50$ ), and (3) the ratio of genotypic information to total variation of the allele frequency plus simulated noise ( $q = 1, 0.5,$  and  $0.25$ ). All SNP marker pairs within a physical distance of 500 kb of each other were tested using the weighted correlation described in the MATERIALS AND METHODS. Results were collected in bins of 50 kb. Power was defined as the fraction of trials within a bin that exceeded a  $\log P$ -value of 4 ( $\sim P \leq 0.01$ ). Trials with breed number  $< 148$  were averaged over five random subsamples of  $n$  breeds from the total. Ratios of  $q < 1$  were generated by adding the allele frequency for a SNP allele to one or three permutations of the frequencies for the same allele.

**Genotypes:** Multiple breeds were characterized using a common set of SNP markers. Variation in the informativeness of marker alleles is presented in supplemental Figure 2. SNPs were selected for use that met the following criteria: (i) SNPs with a  $q$  score  $> 45$  that have flanking sequence occurring only once in the genome sequence, (ii) SNPs that passed Illumina in-house suitability testing, (iii) SNPs where the minor allele was observed in  $\geq 2$  of 11 breeds tested; (iv) SNPs for which the minor allele was observed in  $\geq 1$  of 11 breeds as necessary, included to achieve complete coverage. The 25,073 SNPs resulting were filtered such that SNPs meeting all four criteria were added to the final data set sequentially if they were at least 380 Mb from all SNPs already in the data set. SNPs meeting criteria i, ii, and iv were then added, maintaining the minimal spacing. The resultant 4608 SNPs were submitted to Illumina to generate three oligo pools. DNA samples were submitted to Illumina for fast-track Golden Gate analysis (FAN *et al.* 2006).

For the experiments described, 2801 dogs representing 147 breeds were used. One hundred twenty-nine of these breeds were represented by  $\geq 10$  dogs (supplemental Figure 3, supplemental Table 3). DNA from each dog was genotyped using 1536 markers, of which 674 were spaced across the 38 canine autosomes. A total of 862 additional markers were concentrated in regions of interest that showed maximal variation in allele frequency between breeds. The focused selections were chosen to further characterize areas that allowed breeds to be easily distinguished and may be linked to traits of interest (*e.g.*, SUTTER *et al.* 2007). As a result, the median distance between markers was 409 kb although only  $\sim 26\%$  of the genome was within 250 kb of a marker (supplemental Table 2).

Details of SNP probe sequences associated with QTL and of the sequences in which these markers are imbedded are presented in supplemental Table 4 (see supplemental table legend). Relevant marker allele frequencies in different breeds are presented in supplemental Table 5.

**SNP association:** We tested for correlations between breed allele frequency ( $x_i$ ) and breed-characterized phenotypes ( $y_i$ ) using a weighted Pearson product correlation:

$$r_{xy} = \frac{\sum w_i (x_i - x_w) (y_i - y_w)}{\sqrt{\sum w_i (x_i - x_w)^2 \sum w_i (y_i - y_w)^2}}$$

where  $y_w = \sum w_i y_i / \sum w_i$ ,  $x_w = \sum w_i x_i / \sum w_i$  and  $w_i = \sqrt{n_i}$ , where  $n_i$  is the number of animals for breed  $i$ .

Two measures of significance were important: single SNP  $P$ -value and genomewide  $P$ -value (*e.g.*, the probability of a particular  $r_{xy}$  value in a single test and the multi-test correction when testing all SNPs across the genome, respectively).

We used permutation tests to establish the null distribution of the  $r_{xy}$  statistic for each SNP and for each phenotype. A generalized extreme value distribution was fit to the empirical “null” data using the `gevFit` function of the `fExtremes` package (WUERTZ 2006) for R (R DEVELOPMENT TEAM 2006). The Kolmogorof–Smirnov test (CONOVER 1971) of the R package (`ks.test`) was used to test the goodness of fit. Distributions with a `ks.test`  $P$ -value of  $\leq 0.01$  were considered poorly estimated and dropped from further analysis. The significance of  $r_{xy}$  values were estimated using the cumulative probability function (`pgev`) and  $-\log_{10}$  transformed for convenience (`logP`). For each permutation, the maximum score across all SNPs was recorded as the single genome-scan maximum. Genome-scan maximum values from 1000 permutations were used to estimate the null distribution of a genomewide scan. The 90, 95, and 99% percentiles of this distribution were used as the thresholds from genomewide significance of 0.1, 0.05, and 0.01, respectively.

**Power to detect association:** We estimated the power to detect association with a neighboring marker allele as a function of the number of breeds available. In Figure 2, it can be seen that the power to identify an association drops off rapidly as the number of breeds decreases. This loss of power becomes particularly relevant when phenotypes have been evaluated in only a small number of breeds.

Markers were considered informative if they had a wide range of allele frequencies across breeds. Conversely, a SNP for which both alleles displayed equal frequency across all breeds was uninformative. We estimated the power to detect an association as a function of allele-frequency variation among breeds. The significance (`logP`) of a single-marker test for differently modeled situations is graphed in Figure 3 ( $y$ -axis) as a function of the distance between the SNP markers ( $x$ -axis). Three patterns of variation in the SNP allele frequency among breeds were considered (Figure 3, insets): histograms representing the number of breeds ( $y$ -axis) in each allele-frequency bin ( $x$ -axis). The ability to detect QTL increases with increasing variation of its occurrence in different breeds.

**Regression analyses:** The “`lm`” function of R was used to perform a weighted multiple regression, with the square root

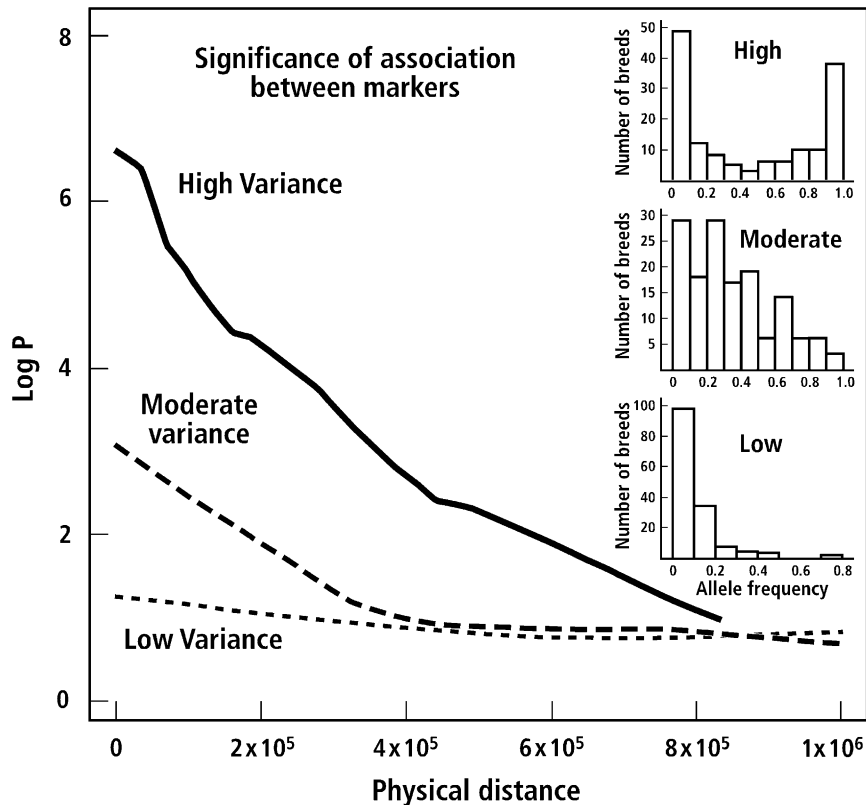


FIGURE 3.—Significance of association between linked markers as a function of physical distance and marker informativeness. LOWESS (CLEVELAND 1981) estimations of average significance are shown for markers in three groups: high, moderate, and low variance. Histograms representative of the three marker categories are shown as insets.

of breed count used for weights (CHAMBERS 1992). The “glm” function of R was used with the option family = “binomial” to carry out a logistic regression (HASTIE and PREGIBON 1992). The “regress” function was used to carry out a mixed-model analysis (CLIFFORD and McCULLAGH 2006) with allele counts as the fixed effects and the breed similarity matrix as the random effects. The variance matrix between breeds was calculated as the similarity between all pairs of breeds using markers separated by at least 500,000 bp. We defined the similarity between two breeds as one minus the average absolute difference in allele frequency across all markers (see supplemental Table 6 for all similarity values). Thus, breeds that are identical had a similarity score of 1 and breeds that were completely different had similarity scores of 0. A leave-one-out strategy was used to predict breed phenotypes with the mixed model. Coefficients estimated from the data with a breed left out were used to predict the phenotype of that breed (see supplemental Figure 4).

## RESULTS

**Morphology:** A number of genes regulating size or shape have been identified in different mammals (humans, mice, rats, or dogs). Several of these regulate relatively large amounts of phenotypic variation (*e.g.*, *IGF-1*, *IGF-2*). Identifying QTL containing such candidate genes provided evidence suggesting that the method proposed was robust.

Selected regions of the genome were examined using a SNP scan of 148 breeds. Using association analysis, several QTL were identified for size (WT) and shape (HT and residuals of WT<sup>0.33</sup> regressed on to height).

Table 1 presents the location and characterization of the loci for which the most evidence was accrued. Loci regulating both height-at-the-withers and body weight are located on *Canis familiaris* autosome (CFA) 7, 10, 15, and 34, whereas the locus on CFA 9 regulates only body weight. When Wt<sup>0.33</sup> is regressed onto height at the withers, a variation in shape can be distinguished that represents differences between breeds that range from dogs that are thin for their height (pursuit hounds such as the greyhound, Afghan hound, or whippet, as well as some smaller dogs such as the fox terrier) to ones that have a large body mass for their height (see supplemental Figure 3). The locus on CFA 6, associated with this phenotype, was not associated with either height or weight. In the Portuguese water dog (LARK *et al.* 2006), a highly significant locus on CFA 12 that regulates an inverse correlation between limb bone length and width was identified. This locus was not identified with genomewide significance in the present across-breeds WT<sup>0.33</sup> residual scan, but it was found in that scan at a significance that validated the pre-identified locus from the Portuguese water dog. Such instances of lowered significance may reflect a low frequency of breeds in which a locus has been fixed.

As can be seen in Table 1, many of the loci contain candidate genes that are associated with size, including *SMAD-2* and *NPR2* on CFA7; *HMG2* on CFA10; *IGF1* on CFA15, as well as a murine high-growth-regulating region containing *SOCS2*; and *IGF2BP2* on CFA34. Thus,

**TABLE 1**  
**Details of QTL for size-related traits**

Trait	Chromosome	Position (bp)	logP	Threshold ( $P < x$ )	No. of genes	Candidate genes
<u>HT</u>	<u>CFA 7</u>	<u>46,696,633</u>	<u>6.20</u>	<u>0.001</u>	<u>7</u>	
	CFA 10	7,033,361	4.94	0.01	1	
	CFA 10	11,465,975	4.36	0.01	5	
	<u>CFA 15</u>	<u>44,228,026</u>	<u>6.05</u>	<u>0.001</u>	<u>5</u>	
	CFA 34	21,414,695	3.66	0.05	9	
<u>WT</u>	<u>CFA 7</u>	<u>46,696,633</u>	<u>7.20</u>	<u>0.001</u>	<u>7</u>	<u>SMAD2, NPR2</u>
	CFA 9	46,401,136	4.99	0.01	19	
	<u>CFA 10</u>	<u>7,033,361</u>	<u>5.15</u>	<u>0.001</u>	<u>1</u>	
	CFA 10	11,465,975	3.63	0.05	5	<i>HMGA2</i>
	CFA 15	37,006,865	3.86	0.05	2	<i>SOCS2</i>
	CFA 15	44,228,026	4.59	0.01	5	<i>IGF1</i>
	CFA 34	21,414,695	3.36	0.1	9	<i>IGF2BP2</i>
	WT <sup>0.33</sup> resid	CFA 6	22,281,985	4.34	0.01	6

For details of traits, see text and this legend. Chromosomes on which these are located are indicated as well as the position in base pairs on each chromosome of the SNP at which significance was estimated. The logarithm of the  $P$ -value (logP) is given as well as the genomewide significance threshold that this  $P$ -value exceeds and the number of known genes in the LD interval (400 kb). Genomewide significance thresholds for logP varied between 3.26 and 3.29 for  $P \leq 0.1$ ; between 3.45 and 3.50 for  $P \leq 0.05$ ; and between 4.00 and 4.05 for  $P \leq 0.01$ . Extremely significant loci are underlined. For more details, see supplemental Tables 4 and 5. The number of genes within 200 kb of the SNP that were investigated for candidate genes and the names of the candidate genes are listed.

associating SNPs from multiple breeds with breed-specific metrics may facilitate association mapping of complex, polygenic phenotypes (across-breed mapping).

**Mapping breed characters:** In many breeds, a number of other desired morphological traits have been under stringent selection and thus should be fixed. Descriptions of these phenotypes are presented in the MATERIALS AND METHODS. Their distribution among breeds is presented in supplemental Table 1. We have used across-breed association mapping to identify putative QTL for many of these (Table 2). In all, 10 traits were associated with 26 loci distributed over 14 chromosomes at a significance better than  $P < 0.01$ . As expected, many of these QTL (10) were identified at high significance, exceeding a genomewide threshold of  $P \leq 0.001$ . QTL for two aspects of snout size or shape were associated with the same SNP on CFA 12; both the length of tail and the degree to which ears are erect were associated with a locus on CFA 15 that also is associated with overall size (see Table 1); similarly, size of snout and erectness of ears were associated with another size locus on CFA 34; and two closely linked loci on CFA 9 regulate variation in the size of the neck or head. Again, suggestive candidate genes were found associated with some of these QTL: *TNFRSF19* and *Fgf5* with short coat and *COL6A3* with the degree of tail curvature. As expected, this mapping technique appears to be very powerful for phenotypes that are very close to fixation and also are found in a large number of breeds, the optimal proportion approaching 50% of the breeds analyzed.

**Additional tests for significance and effects of breed structure:** QTL identified by single-marker tests may implicate causative regions of the genome, or they may represent false positives: shadow effects resulting from autocorrelations in the data. False-positive results may be caused by unequal sharing of genome regions between the breeds (breed structure), coselection of multiple unlinked regions, and/or codependence of unlinked genome regions (interactions). Multiple-regression analysis provides an estimate of the independence of the loci regulating a trait. QTL that deviate from the additive-independent model will not remain significant in a multiple regression and may represent false positives or more complex effects. QTL may appear less significant (or not significant) in a multiple regression if they were coselected with other loci, or if they are involved in interactions with other loci. Table 3 presents the results of multiple-regression analyses of those traits in Tables 1 and 2 that are associated with multiple loci. Several loci either were not significant or demonstrated marginal significance. In all but one instance, the sum of the significant single regression  $R^2$  values greatly exceeded the multiple  $R^2$  value, suggesting that some loci were not causative or that interactions and/or coselection were occurring. In the case of weight, there was an apparent interactive effect,  $P = 0.0009$ , between the major locus on CFA 15 (associated with SNP BICFPJ263341 at 44 Mbp) and the locus on CFA 10 (associated with SNP gnl.ti.360206886\_2 at 11.5 Mbp). This interaction remains significant in the multiple-regression (0.026) and in the mixed multiple-

**TABLE 2**  
**QTL associated with breed morphological characteristics**

Trait	Chromosome	Position (bp)	logP	Threshold ( $P < x$ )	No. of genes	Candidate genes
Short coat	CFA 25	17,862,111	3.88	0.01	5	<i>TNFRSF19</i>
	<u>CFA 32</u>	<u>7,806,734</u>	<u>5.43</u>	<u>0.001</u>	<u>1</u>	<u><i>Egf5</i></u>
Ear bend	CFA 10	11,915,402	4.70	0.01	2	
	CFA 15	44,137,464	4.10	0.01	5	
	CFA 32	14,508,914	4.12	0.01	9	
	<u>CFA 34</u>	<u>21,414,695</u>	<u>6.84</u>	<u>0.001</u>	<u>9</u>	
Tail curve	CFA 1	81,302,720	4.40	0.01	1	
	CFA 9	14,626,755	3.92	0.01	13	
	CFA 25	51,048,799	4.36	0.01	4	<i>COL6A3</i>
	CFA 38	6,614,004	3.94	0.01	0	
Snout angle	CFA 10	61,541,406	3.88	0.01	3	
	CFA 12	57,797,364	5.22	0.001	4	
<u>Snout ratio</u>	<u>CFA 1</u>	<u>97,045,173</u>	<u>5.43</u>	<u>0.001</u>	<u>4</u>	
	CFA 9	50,982,910	3.95	0.01	13	
	CFA 12	57,797,364	4.07	0.01	4	
	CFA 21	27,755,937	4.68	0.01	6	
	CFA 32	32,959,130	4.76	0.01	10	
Head ratio	CFA 9	25,422,459	4.01	0.01	16	<i>IGFBP4</i>
	CFA 22	10,294,335	4.84	0.01	2	
	<u>CFA 34</u>	<u>21,414,695</u>	<u>6.13</u>	<u>0.001</u>	<u>9</u>	
	CFA 38	24,931,616	4.05	0.01	14	
Leg ratio	CFA 3	64,678,450	4.27	0.01	8	<i>RNF4, MXD</i>
	CFA 6	22,280,330	4.06	0.01	6	
Tail ratio	CFA 15	44,239,862	4.45	0.01	5	
<u>Neck ratio</u>	<u>CFA 9</u>	<u>24,032,840</u>	<u>5.00</u>	<u>0.001</u>	<u>17</u>	<u><i>STAT3</i></u>

As in Table 1, traits (SEE MATERIALS AND METHODS) are presented together with the chromosome on which they are located, the position of the SNP with which they are associated, and the significance of the association (logP), number of known genes in the LD interval (400 kb), and genomewide  $P$ -value threshold exceeded. Genomewide significance threshold  $P$ -values for traits varied between 3.8 and 4.1 for  $P \leq 0.01$  and between 4.34 and 4.7 for  $P \leq 0.001$  (underlining). For more details, see supplemental Tables 4 and 5. The number of genes within 200 kb of the SNP that were investigated for candidate genes and the names of the candidate genes are listed.

regression model (0.003; see below). Coselection can mimic a significant interaction effect in this situation (see DISCUSSION). For one trait, the ratio of head to body metrics ("head ratio"), the sum of the three significant individual  $R^2$  values was only slightly greater than the multiple  $R^2$  value, suggesting that these loci might be acting independently.

Considerable population structure exists among dog breeds (PARKER *et al.* 2004). Using the popgen (MARCHINI 2004) package of R, we estimated measures of diversity among these breeds (NICHOLSON *et al.* 2002). The mean "c" (analogous to  $F_{ST}$ ) value is 0.25 with individual breed values ranging from 0.05 to 0.61. In an across-breed association analysis, noncausative (shadow) loci may result from effects of population substructure due to genetic relatedness among breeds. To test for this, we used a mixed-model analysis (see MATERIALS AND METHODS) to predict trait values of weight as well as head:body ratio (head ratio). We found that all of

the significant QTL for weight or head ratio (Table 3) remained significant in a mixed model correcting for genetic relatedness of breeds, with  $P$ -values ranging from  $10^{-2}$  to  $<10^{-5}$  for weight and  $<10^{-3}$  for the three significant head/rat loci.

**Examples illustrating the future potential of the mapping technique:** *Longevity and size:* In general, dogs representing breeds of small size (*e.g.*, Pekingese, toy poodle, terrier breeds) live appreciably longer than those from larger-sized breeds (*e.g.*, Great dane, St. Bernard, Irish wolfhound) (EGENVALL *et al.* 2005). We have mapped loci for longevity using multiple breeds spanning a comprehensive range of sizes. An analysis of breed longevity had been compiled by K. Cassidy (<http://users.pullman.com/lostriver/longhome.htm>), but many of the breeds for which we had genotypes were not included in that database. We therefore prepared a similar database for all breeds genotyped in our study using a variety of website resources (supplemental Table

**TABLE 3**  
**Single- and multiple-regression results for selected traits with multiple QTL**

	SNP	Chromosome	Position (bp)	Significance	Single $R^2$	Multiple $R^2$
WT	gnl.ti.390449323_1	7	46,696,633	**	0.34	
	BICFPJ1156983	9	46,401,136	**	0.10	
	BICF232J28587	10	7,033,361	NS	0.19	
	gnl.ti.360206886_2	10	11,465,975	**	0.12	
	gnl.ti.351411336_1	15	37,006,665	***	0.14	
	BICFPJ263341	15	44,228,026	***	0.48	
	BICFPJ1062878	34	21,414,695	**	0.20	
					$\Sigma = 1.6$ (1.4)	0.69
Snout rat	gnl.ti.355951851_2	1	97,045,173	***	0.15	
	BICF229J36361	9	50,982,910	**	0.11	
	BICF236J54123	12	57,797,364	**	0.11	
	gnl.ti.390310078_3	21	27,755,937	**	0.15	
	BICF229J63639	32	32,959,130	***	0.24	
					$\Sigma = 0.8$	0.44
HT	gnl.ti.390449323_1	7	46,696,633	***	0.35	
	BICF232J28587	10	7,033,361	NS	0.17	
	gnl.ti.360206886_2	10	11,465,975	***	0.16	
	BICFPJ263341	15	44,228,026	***	0.53	
	BICFPJ1062878	34	21,414,695	**	0.19	
					$\Sigma = 1.4$ (1.2)	0.65
head.rat	BICF229J19878	9	25,422,459	NS	0.08	
	gnl.ti.350815589_1	22	10,294,335	***	0.13	
	BICFPJ1062878	34	21,414,695	***	0.15	
	gnl.ti.390146013_1	38	24,931,616	***	0.13	
					$\Sigma = 0.5$ (0.4)	0.34

Trait, SNP, SNP chromosome location, and SNP base-pair position on the chromosome are indicated in the first four columns. Significance is noted as not significant (NS); (\*)  $0.01 < P < 0.05$ ; (\*\*)  $0.001 < P < 0.01$ ; (\*\*\*)  $P < 0.001$ . "Single  $R^2$ " presents the amount of variation explained by a single SNP in the single-regression model. "Multiple  $R^2$ " presents the amount of variation explained with all SNPs in the same model. The sum of SNP single  $R^2$  is presented in two forms: the total sum ( $\Sigma$ ) or the total minus the  $R^2$  of values that were not significant. Some traits were transformed to achieve a better fit to the normal distribution: Snout.rat was squared. Height was arcsine square root transformed; head.rat was log transformed.

1). Figure 4 compares longevity/size data between the two databases. The negative correlation between age of death (AOD) and size is obvious. The slope of the regression of size onto longevity is the same in both data sets, although the difference in intercepts indicates that the database that we developed yields an average age of death that is older. This may be due to the fact that Cassidy's data utilized information from both veterinarian records and owner responses to questionnaires, whereas our data were biased toward owner surveys, which typically prefer to reference longer-lived animals. Although this may produce an inflated mean value of AOD, it presents a more sensitive signal for genetic analysis. We therefore utilized our larger database, together with the genotyping used in Table 1, to identify QTL for breed-associated age of death (Table 4).

Included in Table 4 are data indicating the presence or absence of size loci associated with the same SNP. Seven loci were identified, three of which, CFA 7, 10, and 15 were associated with significant size (as in weight) loci. These were also the most significant loci for longevity. A fourth, on CFA 34, was associated with a less-significant weight locus. Loci on CFA 9, 23, and 25,

although quite significant for age of death, were not significant for size with the exception of the locus on CFA 9, which is linked to a very significant size locus (see Table 1). When these age-of-death loci were combined in a multiple regression, three on CFA 10, 25, and 34 were no longer significant and the multiple  $R^2$  was approximately half the value of the sum of the single  $R^2$  values.

*Behavior:* Two aspects of dog behavior that appear to be highly breed specific are herding and pointing. Pluis Davern, a nationally recognized dog trainer qualified to judge a large number of breeds (<http://www.infodog.com/judges/17422/juddat.htm>), scored the 148 genotyped breeds for two additional phenotypes: "boldness vs. timidity" and "trainability." Behavioral scores for the 148 breeds are presented in supplemental Table 3. Using these scores, we identified several loci of interest (Table 5). We identified one locus for pointing on CFA 8 with a genomewide significance threshold of  $0.01 < P < 0.05$ . Three loci were detected for herding; these were located on CFA 1 ( $P < 0.01$ ) and on CFA 4 and CFA15 ( $0.01 < P < 0.05$ ). While the boldness and trainability gestalts are subjective, and at best descriptive, we

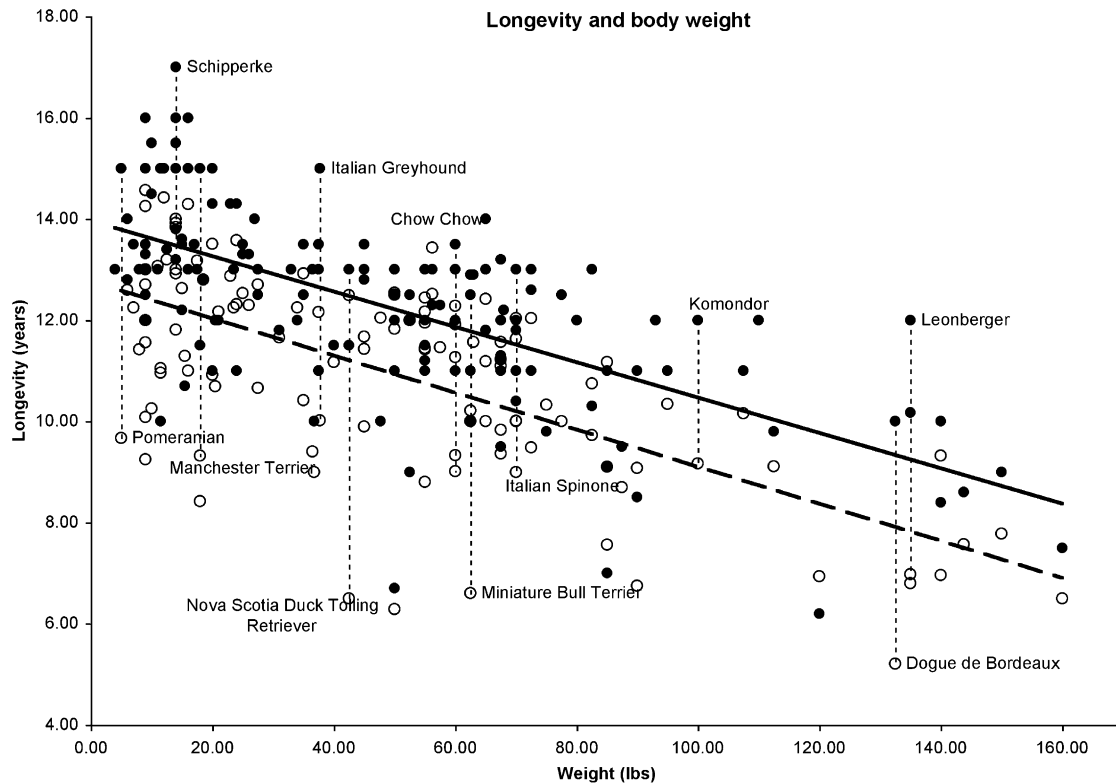


FIGURE 4.—Longevity or AOD as a function of body weight in pounds. For details, see text. Solid symbols represent the database created from websites (see supplemental Table 3). Open symbols are data from the database of Cassidy (<http://users.pullman.com/lostriver/longhome.htm>). A few dog breeds with extreme values are noted.

nevertheless found one significant ( $P < 0.01$ ) locus for trainability on CFA 10 as well as five for boldness on CFA 15 and 22 ( $P < 0.01$ ) and on CFA 1, 4, and 17 ( $0.01 < P < 0.05$ ). In a multiple regression, all of the loci for boldness remained significant. The locus on CFA 15 is interesting in that it does not appear to be related to size, as approximately equal numbers of large and small breeds were found to be bold (see supplemental Table 3), and boldness and size were not correlated ( $r = 0.18$ ;  $P = 0.3$ ). Possible candidate genes are listed in Table 5 for herding and pointing, along with two of the boldness QTL. Included in Table 5 are data for excitability (comprising 56 breeds) taken from HART and MILLER (1985). Two significant QTL were identified on CFA 7 and 15. Both coincided with major-size loci. Unlike the relationship between boldness and size, excitability was highly correlated with size ( $r = -0.8$ ;  $P < 10^{-12}$ , despite the small data set (56 breeds *vs.* the 148 used in the analysis of boldness).

## DISCUSSION

Three powerful genetic procedures are now available using a canine model:

1. Segregation in planned crosses or within a breed population can be used to identify loci for simple and

complex phenotypes. This approach takes advantage of the large LD distances that can be attributed to founder effects and bottlenecks (for example, MIGNOT *et al.* 1991; ACLAND *et al.* 1998, 1999; LINGAAS *et al.* 1998; VAN DE SLUIS *et al.* 1999; JONASDOTTIR *et al.* 2000; CHASE *et al.* 2005a, 2006; TODHUNTER *et al.* 2005).

2. LD mapping across breeds has been used to reduce haplotypes of simple and complex phenotypes to reasonably small DNA sequences and often to identify single genes (CLARK *et al.* 2006; GOLDSTEIN *et al.* 2006; KARLSSON *et al.* 2007; PARKER *et al.* 2007; SARGAN *et al.* 2007).
3. The across-breed mapping method described here, which combines association with multiple-breed LD mapping, thereby associating small regions of the genome with the phenotype, can be used.

The results presented here illustrate the power of across-breed mapping using a data set of  $>100$  breeds. Using morphological phenotypes, we have found an interaction between loci regulating weight on CFA 10 and CFA 15 and implicating a major locus for size on CFA 7. We have validated loci, first described in the Portuguese water dog: one, a major locus regulating shape (limb length *vs.* width) on CFA 12 (LARK *et al.* 2006) and two size loci on CFA 15 (44 and 37 Mb) identified in previous studies (CHASE *et al.* 2002, 2005b; SUTTER



TABLE 4

**QTL associated with AOD and the probability that size is also associated with that SNP**

Trait	Chromosome	Position (bp)	logP	Threshold $P < x$
<i>Age of death</i>	CFA 7	46,696,633	7.06	0.001
<i>Size</i>	CFA 7	46,696,633	7.73	0.001
Age of death	CFA 9	48,230,567	4.18	0.01
Size	CFA 9	48,230,567	2.82	>0.1
Age of death	CFA 10 <sup>a</sup>	7,033,361	4.46	0.01
Size	CFA 10 <sup>a</sup>	7,033,361	5.15	0.001
<i>Age of death</i>	CFA 15	44,228,026	8.94	0.001
<i>Size</i>	CFA 15	44,228,026	4.59	0.01
Age of death	CFA 23	35,509,334	4.12	0.01
Size	CFA 23	35,509,334	2.58	>0.1
Age of death	CFA 25 <sup>a</sup>	18,193,826	3.94	0.05
Size	CFA 25	18,193,826	2.55	>0.1
Age of death	CFA 34 <sup>a</sup>	21,414,695	3.73	0.05
Size	CFA 34	21,414,695	3.36	0.1

Trait, chromosome (CFA), logP, and significance (genome-wide  $P$ -value threshold) are as in Table 2. Genomewide significance thresholds for logP for association with AOD were 3.27 for  $P \leq 0.1$ , 3.45 for  $P \leq 0.05$ , and 3.95 for  $P \leq 0.01$ . Most significant are in italics. Thresholds for size were 3.26 for  $P \leq 0.1$ , 3.45 for  $P \leq 0.05$ , and 4.00 for  $P \leq 0.01$ .

<sup>a</sup>Loci no longer significant in a multiple-regression model (see text).

*et al.* 2007). In addition, we have found a number of loci affecting morphology, some of which may be independent regulators of the relation of the size of the skull to the post-cranial body. For these and other traits, the LD

distance associated with any SNP is  $\sim 500$  kb. This is much smaller than the LD distance associated with markers when mapping within a single breed (SUTTER *et al.* 2004). Nevertheless, as seen in Tables 1, 2, and 5, many genes remain to be explored in searching for alleles that regulate the phenotypes in question (in the data presented, this number ranges from 1 to 19, depending on the QTL).

Most often, across-breed mapping identifies markers that tend to be near or at fixation (homozygous) in breeds with the associated phenotype. Breeds in which the phenotype is still segregating will not contribute to the power of QTL identification. However, they will provide a resource in which the association can be validated using within-breed segregation analysis. Such breeds are readily identified from the across-breed SNP genotyping database. It should be possible now to validate the most significant ( $P \leq 0.001$ ) of the other loci in Table 2 using breeds in which the implicated SNPs are segregating [*e.g.*, the locus on CFA 32 for short coat (Table 2) was identified by segregation analysis using dachshunds or corgis (HOUSLEY and VENTA 2006)].

Limitations to across-breed mapping will always necessitate validation using within-breed segregation analysis. One limitation of the method is the potential for false positives that may arise from population structure, whereby causative regions of the genome displaying significant associative  $P$ -values cannot be distinguished from noncausative ones. Our simple association analysis has made the assumption that dog breeds are indepen-

TABLE 5

**QTL associated with behavior**

Trait	Chromosome	Position (bp)	logP	Threshold $P < x$	No. of genes	Candidate genes
<u>Herding</u>	<u>CFA1</u>	<u>27630805</u>	<u>7.20</u>	<u>0.001</u>	<u>4</u>	<u>MC2R, C18orf1</u>
Boldness	CFA1	67693978	4.26	0.05	7	
Herding	CFA4	42765963	4.83	0.05	6	
Boldness	CFA4	40782966	4.15	0.05	7	DRD1
Excitability <sup>a</sup>	CFA7	46696633	4.06	0.01	7	
Pointing	CFA8	33344686	5.33	0.05	6	CNIH
Trainability	CFA10	13396503	3.77	0.05	4	
Excitability <sup>a</sup>	CFA15	44228026	4.63	0.01	5	
Herding	CFA15	44229716	4.89	0.05	5	
<u>Boldness</u>	<u>CFA15</u>	<u>44137464</u>	<u>5.05</u>	<u>0.001</u>	<u>5</u>	
Boldness	CFA17	15478350	4.40	0.05	1	
<u>Boldness</u>	<u>CFA22</u>	<u>25446003</u>	<u>6.09</u>	<u>0.001</u>	<u>1</u>	<u>PCDH9</u>

The genomewide SNP scan (see Table 2) was used to associate SNP markers with several behavioral phenotypes: pointing, herding, boldness, and trainability. Scoring for these phenotypes is presented in supplemental Table 3. From left to right (top), columns list the trait, chromosome, nucleotide position on the chromosome, the logP value of the significance, the genomewide threshold of significance, the number of known genes in the LD interval (400 kb), and possible candidate genes. More significant loci are underlined. The genomewide significance thresholds for the four traits were herding,  $0.01 < P < 0.05 = 4.38$  and  $P \leq 0.01 = 5.04$ ; pointing,  $0.01 < P < 0.05 = 4.69$  and  $P \leq 0.01 = 5.69$ ; boldness,  $0.01 < P < 0.05 = 4.09$  and  $P \leq 0.01 = 4.81$ ; and trainability,  $0.01 < P < 0.05 = 3.48$  and  $P \leq 0.01 = 3.86$ .

<sup>a</sup>Two loci for excitability were identified using data by HART and MILLER (1985). The genomewide threshold  $P \leq 0.01$  for this trait was logP = 3.67. For more details, see supplemental Tables 4 and 5. The number of genes within 200 kb of the SNP that were investigated for candidate genes and the names of the candidate genes are listed.

dent of each other. However, this is rarely the case. Breed structure is the network of haplotype regions shared between breeds. For example, we would expect a high proportion of sharing between the genomes of the standard and the toy poodle, although we expect significant differences in regions with loci related to control of size. The mean  $F_{ST}$  between the breeds used in our study is 0.25 (SD = 0.11), indicating that they have not diverged greatly. Moreover, principal component (PC) analysis of the allele frequencies (data not presented) shows that the allele shared between breeds is not coherent (*e.g.*, the first PC explains only 4% of the total variation in allele frequency). Thus, different breeds share different parts of the genome.

Reviewing similar techniques applied to inbred mouse strains, PAYSEUR and PLACE (2007) have summarized the power and pitfalls of the technique (*e.g.*, they showed that unequal relatedness between strains can give rise to false-positive associations, since causative regions of the genome may be co-inherited with non-causative regions). Studies in the mouse suggest extensions to this technique in the dog as more robust SNP and phenotype data become available: (1) use of SNP haplotypes spanning a small physical distance (*e.g.*, 300 kb) instead of single SNP alleles (KARLSSON *et al.* 2007; SALMON HILLBERTZ *et al.* 2007; SUTTER *et al.* 2007), (2) correction for relatedness between breeds using mixed-model analysis, (3) balanced representation of breeds, and (4) correction for nonsystemic LD by testing multiple loci in the same model.

We have used an across-breed averaged correction for breed structure to correct for effects of breed structure on weight and head-to-body ratio and a multi-QTL regression model to rule out nonsystemic LD among loci that we have detected. Nevertheless, interactions and coselection can result in false positives and, as with mouse inbred strains, it will always be necessary to validate loci.

The current data set has several limitations. In Figure 2, we presented evidence that significance is limited to 250 kb on each side of a SNP. By this criterion, our database analyzes only 26% of the genome with the remainder not considered in the association mapping that identified the loci used in the multiple-regression model in Table 3. Therefore, within-breed validation of segregating loci will be required to completely rule out nonsystemic LD. Beyond shadow effects, there remain other complex effects, such as interactions between loci and/or coselection of loci during breed formation that may confound results. The data in Table 3 indicate that such effects may be present for most of the traits examined.

In the future, genotyping platforms should offer deeper coverage of the genome (~50,000 well-placed and informative SNPs), more robust and balanced breed representation, and more dogs per breed (30–50). Finally, improved phenotypic characterization of breed stereotypes is needed.

Phenotypes that have been under stringent selection are best suited to across-breed association mapping, and this is apparent in the data in Table 2 where highly significant values for several stringently selected morphological QTL were observed. Similarly, stringent selection for behavior may be responsible for the behavioral loci identified here. Candidate genes associated with these loci (Table 5) include ones that might be expected to play a major role in regulating behavior: MC2R on CFA 1 (27,381,939 bp) is a melanocortin receptor, and C18orf1 (27,572,327 bp) has been implicated in schizophrenia. DRD1, on CFA 4 (40,743,436 bp), encodes a dopamine subtype receptor. CNIH, on CFA 8 (33,396,000 bp), has been implicated in cranial nerve development. Finally, PCDH9, on CFA 22 (24 273 482 bp), encodes a protein localized to synaptic junctions and believed to be involved in specific neural connections and signal transduction. Although the behaviors involved are poorly defined, the presence of candidate genes appropriate to behavior is encouraging.

Despite the likely possibility of false positives, the across-breed mapping technique can focus attention on loci that may regulate genetic differences *between* breeds when these cannot be investigated using segregation *within* breeds. In an extensive study of within-breed longevity involving many different breeds, GALIS *et al.* (2007) were unable to find evidence for an inverse correlation between longevity and size and neither have we seen such an inverse correlation in Portuguese water dogs that display a range of sizes approaching threefold (our unpublished data). The peculiar inverse correlation between longevity and size seen in Figure 4 is strictly a between-breed phenomenon and provides an excellent example of a trait that can be approached with across-breed mapping. The data in Table 4 suggest that a subset of loci, which control body size, also contribute to longevity, with some playing a greater role in the aging process than others.

Across-breed mapping depends on variants of the genomic architecture that are relatively fixed in a large number of different breeds. Given accurate estimations of breed-disease frequency, this technique can be used to determine the impact of the breed-fixed genome regions on the disease. All of these breeds represent “successful” genome architectures. While some may be more or less prone to a disease, they are still functional productive genomes. It is not likely that a large number of breeds harbor a single deleterious mutation that can be detected in this fashion. Thus, it is likely that one of several functional genome variants will predispose to a disease state as, for example, one might encounter with size loci where particular alleles may predispose toward orthopedic diseases.

Because power in across-breed mapping derives from variation between breeds in the frequency of disease (as in the simulation in Figure 3), this approach functions

only if disease reporting is accurate. While databases of disease frequency exist, they are often based on breeder-directed health surveys and inherent biases exist. More useful may be the growing number of veterinary school databases spanning several years.

The quality of genotypic data is paramount as well. Ideally, large public databases providing comprehensive SNP data on ~50 independent lineages for most AKC-recognized dog breeds should be made available as the genotypic breed standard for future mapping studies. Such an effort, termed CanMap (<http://www.sciencemag.org/cgi/content/full/sci;317/5845/1668>), is currently underway in an effort initially involving investigators from Cornell, the University of California at Los Angeles, and the National Human Genome Research Institute (PENNISI 2000). The initial end point will be a public repository of dense SNP profiles of about a dozen dogs from each of nearly a hundred breeds, plus a set of wild canids, which together will be an invaluable resource for the genetic dissection of complex polygenic diseases, a large number of which are common to both dogs and humans.

In summary, across-breed mapping is another facet of the canine model that complements within-breed mapping and LD mapping. It implicates new regions of interest and can provide validation of previously identified loci.

We thank the thousands of pet owners who provided samples and data about their dogs for their participation and support of this work and the many dog show organizers who kindly allowed us to have collection stands to gather these samples for dog research. We thank John Fondon, III, and Heidi Parker for helpful comments regarding this manuscript. We gratefully acknowledge funding from the Judith Chiara Family Trust and National Institutes of Health GM063056 (K.G.L. and K.C.), the Intramural Program of the National Human Genome Research Institute (E.A.O.), and Mars (A.M. and P.J.).

#### LITERATURE CITED

- ACLAND, G. M., K. RAY, C. S. MELLERSH, W. GU, A. A. LANGSTON *et al.*, 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* **96**: 3048–3053.
- ACLAND, G. M., K. RAY, C. S. MELLERSH, W. GU, A. A. LANGSTON *et al.*, 1999 A novel retinal degeneration locus identified by linkage and comparative mapping of canine early retinal degeneration. *Genomics* **59**: 134–142.
- AMERICAN KENNEL CLUB, 1998 *The Complete Dog Book*. Howell Book House, New York.
- CERVINO, A. C., G. LI, S. EDWARDS, J. ZHU, C. LAURIE *et al.*, 2005 Integrating QTL and high-density SNP analyses in mice to identify *Insig2* as a susceptibility gene for plasma cholesterol levels. *Genomics* **86**: 505–517.
- CHAMBERS, J. M., 1992 Linear models, pp. 99–116 in *Statistical Models in S*, edited by J. M. CHAMBERS and T. J. HASTIE. Wadsworth & Brooks/Cole, Belmont, CA.
- CHASE, K., D. R. CARRIER, F. R. ADLER, T. JARVIK, E. A. OSTRANDER *et al.*, 2002 Genetic basis for systems of skeletal quantitative traits: principal component analysis of the canid skeleton. *Proc. Natl. Acad. Sci. USA* **99**: 9930–9935.
- CHASE, K., D. F. LAWLER, D. R. CARRIER and K. G. LARK, 2005a Genetic regulation of osteoarthritis: a QTL regulating cranial and caudal acetabular osteophyte formation in the hip joint of the dog (*Canis familiaris*). *Am. J. Med. Genet.* **135A**: 334–335.
- CHASE, K., D. CARRIER, F. ADLER, E. OSTRANDER and K. LARK, 2005b Interaction between the X chromosome and an autosome regulates size sexual dimorphism in Portuguese water dogs. *Genome Res.* **15**: 1820–1824.
- CHASE, K., D. SARGAN, K. MILLER, E. A. OSTRANDER and K. G. LARK, 2006 Understanding the genetics of autoimmune disease: two loci that regulate late onset Addison's disease in Portuguese water dogs. *Int. J. Immunogenet.* **33**: 179–184.
- CLARK, L. A., J. M. WAHL, C. A. REES and K. E. MURPHY, 2006 Retrotransposon insertion in *SILV* is responsible for merle patterning of the domestic dog. *Proc. Natl. Acad. Sci. USA* **103**: 1376–1381.
- CLEVELAND, W. S., 1981 LOWESS: a program for smoothing scatterplots by robust locally weighted regression. *Am. Stat.* **35**: 54.
- CLIFFORD, D., and P. MCCULLAGH, 2006 regress: Gaussian linear models with linear covariance structure. R package version 1.0–0. <http://galton.uchicago.edu/~clifford/>.
- CONOVER, W. J., 1971 *Practical Nonparametric Statistics*, pp. 295–301, 309–314. John Wiley & Sons, New York.
- DIPETRILLO, K., S. W. TSAIH, S. SHEEHAN, C. JOHNS, P. KELMENSEN *et al.*, 2004 Genetic analysis of blood pressure in C3H/HeJ and SWR/J mice. *Physiol. Genomics* **17**: 215–220.
- EGENVALL, A., B. N. BONNETT, A. HEDHAMMAR and P. OLSON, 2005 Mortality in over 350,000 insured Swedish dogs from 1995–2000: II. Breed-specific age and survival patterns and relative risk for causes of death. *Acta Vet. Scand.* **46**: 121–136.
- FAN, J. B., M. S. CHEE and K. L. GUNDERSON, 2006 Highly parallel genomic assays. *Nat. Rev. Genet.* **7**: 632–644.
- GALIS, F., I. VAN DER SLUIJS, T. J. VAN DOOREN, J. A. METZ and M. NUSSBAUMER, 2007 Do large dogs die young? *J. Exp. Zool. B Mol. Dev. Evol.* **308**: 119–126.
- GOLDSTEIN, O., B. ZANGERL, S. PEARCE-KELLING, D. SIDJANIN, J. KIJAS *et al.*, 2006 Linkage disequilibrium mapping in domestic dog breeds narrows the progressive rod-cone degeneration interval and identifies ancestral disease-transmitting chromosome. *Genomics* **88**: 541–550.
- GRUPE, A., S. GERMER, J. USUKA, D. AUD, J. K. BELKNAP *et al.*, 2001 In silico mapping of complex disease-related traits in mice. *Science*. **292**: 1915–1918.
- HART, B. L., and M. F. MILLER, 1985 Behavioral profiles of dog breeds. *J. Am. Vet. Med. Assoc.* **186**: 1175–1180.
- HASTIE, T. J., and D. PREGIBON, 1992 Generalized linear models, pp. 195–246 in *Statistical Models in S*, edited by J. M. CHAMBERS and T. J. HASTIE. Wadsworth & Brooks/Cole, Belmont, CA.
- HOUSLEY, D. J., and P. J. VENTA, 2006 The long and the short of it: evidence that *FGF5* is a major determinant of canine 'hair'itability. *Anim. Genet.* **37**: 309–315.
- JONASDOTTIR, T. J., C. S. MELLERSH, L. MOE, R. HEGGEBE, H. GAMLEM *et al.*, 2000 Genetic mapping of a naturally occurring hereditary renal cancer syndrome in dogs. *Proc. Natl. Acad. Sci. USA* **97**: 4132–4137.
- KARLSSON, E. K., I. BARANOWSKA, C. M. WADE, N. H. C. SALMON HILLBERTZ, M. C. ZODY *et al.*, 2007 Efficient mapping of Mendelian traits in dogs through genome-wide association. *Nat. Genet.* **39**: 1321–1328.
- KC/BSAVA, 2004 KC/BSAVA Purebred dog health survey. United Kingdom Kennel Club and British Small Animal Veterinary Association. <http://www.thekennelclub.org.uk/item/549>.
- LARK, K. G., K. CHASE and N. B. SUTTER, 2006 Genetic architecture of the dog: sexual size dimorphism and functional morphology. *Trends Genet.* **22**: 537–544.
- LIAO, G., J. WANG, J. GUO, J. ALLARD, J. CHENG *et al.*, 2004 In silico genetics: identification of a functional element regulating *H2-Ealpha* gene expression. *Science* **306**: 690–695.
- LINDBLAD-TOH, K., C. M. WADE, T. S. MIKKELSEN, E. K. KARLSSON, D. B. JAFFE *et al.*, 2005 Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**: 803–819.
- LINGAAS, F., T. AARSKAUG, M. SLETTEN, I. BJERKAS, U. GRIMHOLT *et al.*, 1998 Genetic markers linked to neuronal ceroid lipofuscinosis in English setter dogs. *Anim. Genet.* **29**: 371–376.
- MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

- MARCHINI, J. L., 2004 *popgen: Statistical and Population Genetics*. R package version 0.0–4. <http://www.stats.ox.ac.uk/~marchini/software.html>.
- MICHELL, A. R., 1999 Longevity of British breeds of dog and its relationships with sex, size, cardiovascular variables and disease. *Vet. Rec.* **145**: 625–629.
- MIGNOT, E., C. WANG, C. RATAZZI, C. GAISER, M. LOVETT *et al.*, 1991 Genetic linkage of autosomal recessive canine narcolepsy with a mu immunoglobulin heavy-chain switch-like segment. *Proc. Natl. Acad. Sci. USA* **88**: 3475–3478.
- NICHOLSON, G., A. V. SMITH, F. JÓNSSON, Ó. GÚSTAFSSON, K. STEFÁNSSON *et al.*, 2002 Assessing population differentiation and isolation from single-nucleotide polymorphism data. *JRSS(B)* **64**: 695–715.
- OSTRANDER, E. A., and L. KRUGLYAK, 2000 Unleashing the canine genome. *Genome Res.* **10**: 1271–1274.
- PALMER, J., 1994 *The Illustrated Encyclopedia of Dog Breeds*. Wellfleet Press, Edison, NJ.
- PARK, Y. G., R. CLIFFORD, K. H. BUETOW and K. W. HUNTER, 2003 Multiple cross and inbred strain haplotype mapping of complex-trait candidate genes. *Genome Res.* **13**: 118–121.
- PARKER, H. G., and E. A. OSTRANDER, 2005 Canine genomics and genetics: running with the pack. *PLoS Genet.* **1**: e58.
- PARKER, H. G., L. V. KIM, N. B. SUTTER, S. CARLSON, T. D. LORENTZEN *et al.*, 2004 Genetic structure of the purebred domestic dog. *Science* **304**: 1160–1164.
- PARKER, H. G., A. V. KUKKOVA, D. T. AKEY, O. GOLDSTEIN, E. F. KIRKNESS *et al.*, 2007 Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. *Genome Res.* **17**: 1562–1571.
- PAYSEUR, B. A., and M. PLACE, 2007 Prospects for association mapping in classical inbred mouse strains. *Genetics* **175**: 1999–2008.
- PENNISI, E., 2000 Human genome: finally, the book of life and instructions for navigating it. *Science* **288**: 2304–2307.
- PLETCHER, M. T., P. McCLURG, S. BATALOV, A. I. SU, S. W. BARNES *et al.*, 2004 Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS Biol.* **2**: e393.
- R DEVELOPMENT CORE TEAM, 2006 *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org>.
- SALMON HILLBERTZ, N. H., M. ISAKSSON, E. K. KARLSSON, E. HELLMÉN, G. R. PIELBERG *et al.*, 2007 Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. *Nat. Genet.* **39**(11): 1318–1320.
- SARGAN, D. R., D. WITHERS, L. PETTITT, M. SQUIRE, D. J. GOULD *et al.*, 2007 Mapping the mutation causing lens luxation in several terrier breeds. *J. Hered.* **98**: 534–538.
- SUTTER, N. B., M. A. EBERLE, H. G. PARKER, B. J. PULLAR, E. F. KIRKNESS *et al.*, 2004 Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Res.* **14**: 2388–2396.
- SUTTER, N. B., C. D. BUSTAMANTE, K. CHASE, M. M. GRAY, K. ZHAO *et al.*, 2007 A single IGF1 allele is a major determinant of small size in dogs. *Science* **316**: 112–115.
- TODHUNTER, R. J., R. MATEESCU, G. LUST, N. I. BURTON-WURSTER, N. L. DYKES *et al.*, 2005 Quantitative trait loci for hip dysplasia in a cross-breed canine pedigree. *Mamm. Genome* **16**: 720–730.
- VAN DE SLUIS, B. J., M. BREEN, M. NANJI, M. VAN WOLFEREN, P. DE JONG *et al.*, 1999 Genetic mapping of the copper toxicosis locus in Bedlington terriers to dog chromosome 10, in a region syntenic to human chromosome region 2p13-p16. *Hum. Mol. Genet.* **8**: 501–507.
- WANG, J., G. LIAO, J. USUKA and G. PELTZ, 2005 Computational genetics: From mouse to human? *Trends Genet.* **21**: 526–532.
- WANG, X., R. KORSTANJE, D. HIGGINS and B. PAIGEN, 2004 Haplotype analysis in multiple crosses to identify a QTL gene. *Genome Res.* **14**: 1767–1772.
- WILCOX, B., and C. WALKOWICZ, 1995 *The Atlas of Dog Breeds*. T. F. H. Publications, Neptune City, NJ.
- WUERTZ, D., 2006 *fExtremes: Rmetrics: Extreme Financial Market Data*, R package version 240.10068. <http://www.rmetrics.org>.

Communicating editor: M. JOHNSTON