

Tracking footprints of artificial selection in the dog genome

Joshua M. Akey^{a,1}, Alison L. Ruhe^b, Dayna T. Akey^a, Aaron K. Wong^b, Caitlin F. Connelly^a, Jennifer Madeoy^a, Thomas J. Nicholas^a, and Mark W. Neff^{b,c,d,1}

^aDepartment of Genome Sciences, University of Washington, Seattle, WA 98195; ^bCenter for Veterinary Genetics, University of California, Davis, CA 95616; ^cCenter for Canine Health and Performance, Translational Genomics Research Institute, Phoenix, AZ 85004; and ^dThe Van Andel Research Institute, Grand Rapids, MI 49503

Edited* by Jasper Rine, University of California, Berkeley, CA, and approved December 15, 2009 (received for review September 2, 2009)

The size, shape, and behavior of the modern domesticated dog has been sculpted by artificial selection for at least 14,000 years. The genetic substrates of selective breeding, however, remain largely unknown. Here, we describe a genome-wide scan for selection in 275 dogs from 10 phenotypically diverse breeds that were genotyped for over 21,000 autosomal SNPs. We identified 155 genomic regions that possess strong signatures of recent selection and contain candidate genes for phenotypes that vary most conspicuously among breeds, including size, coat color and texture, behavior, skeletal morphology, and physiology. In addition, we demonstrate a significant association between *HAS2* and skin wrinkling in the Shar-Pei, and provide evidence that regulatory evolution has played a prominent role in the phenotypic diversification of modern dog breeds. Our results provide a first-generation map of selection in the dog, illustrate how such maps can rapidly inform the genetic basis of canine phenotypic variation, and provide a framework for delineating the mechanistic basis of how artificial selection promotes rapid and pronounced phenotypic evolution.

Canis lupis | evolution

The modern domesticated dog (*Canis lupus familiaris*) represents one of the longest-running experiments in human history (1, 2). This experiment, still actively being conducted, has resulted in over 400 genetically distinct breeds that harbor considerable variation in behavioral, physiological, and morphological phenotypes (3). Although the domestication of dogs began over 14,000 years ago (4, 5), the spectacular phenotypic diversity exhibited among breeds is thought to have originated much more recently, largely through intense artificial selection and strict breeding practices to perpetuate desired characteristics. Thus, the canine genome, shaped by centuries of strong selection, likely contains many important lessons about the genetic architecture of phenotypic variation and the mechanistic basis of rapid short-term evolution. Indeed, dogs and other domesticated species played an important role in Darwin's *On the Origin of the Species* (6), as they provide vivid examples of descent with modification. However, relatively little progress has been made on systematically identifying which regions of the canine genome have been influenced by selective breeding during the natural history of the dog.

Most studies of artificial selection in dogs have focused on single-gene analyses arising from phenotype-driven studies. Notable examples include *IGF1* (7), an expressed *FGF4* retrogene (8), and three genes (*RSPO2*, *FGF5*, and *KRT71*) (9) that influence variation in size, limb length, and coat phenotypes, respectively. However, candidate gene approaches are not well suited to providing general insights into the frequency, location, and types of loci influenced by selection. Furthermore, disentangling the confounding effects of selection and demographic history on patterns of DNA sequence variation is notoriously difficult with single-locus analyses (10). To date, the only genome-wide analysis of selection in dogs has focused on a specific phenotype in a single breed, foreshortened limbs in

Dachshunds, using a relatively coarse panel of microsatellite markers (11).

Recent advances in canine genomics, including a high-quality reference sequence (12), the construction of a dense map of over 2.5 million SNPs (12), and the development of SNP genotyping arrays (13) have enabled systematic studies of canine genomic variation. Using these genomic resources, we performed the largest genome-wide scan to date for targets of selection in purebred dogs. By applying unique statistical methods to a map of over 21,000 SNPs genotyped in a phenotypically diverse panel of 10 breeds, we identified 155 regions of the canine genome that have likely been subject to strong artificial selection. Our results are unique in providing a detailed glimpse into the genetic legacy of centuries of breeding practices, suggest that regulatory evolution has played a prominent role in the rapid phenotypic diversification of breeds, and nominate numerous candidate genes for contributing to breed-specific differences in behavior, morphology, and physiology.

Results

SNP Characteristics and Data Quality. We genotyped $\approx 21,000$ autosomal SNPs with Illumina's Infinium CanineSNP20 Bead-Chip in a panel of 275 unrelated dogs from 10 phenotypically and genetically diverse breeds (Table 1). SNP markers were uniformly distributed throughout the genome, with a median SNP density of 103.5 ± 124.6 kb. Table 1 provides summary statistics of polymorphism for each breed. Note the average minor allele frequency was $\approx 25\%$ across breeds, which reflects the ascertainment bias toward common alleles based on the SNP discovery strategy (12). Relationships among breeds were investigated by principal components analysis, which demonstrated that the German Shepherd, Shar-Pei, Beagle, and Greyhound were particularly genetically distinct (Fig. S1).

We performed several analyses to assess SNP data quality. First, four individuals were genotyped in duplicate, and the concordance among genotype calls was $>99\%$ across all replicates. Second, for each breed, arrays were performed on a trio of samples and non-Mendelian transmission, indicative of genotyping errors or copy number variants, was assessed. In total, $\approx 0.4\%$ of markers exhibited Mendelian inconsistencies, consistent with the low genotyping error rate suggested by the replicate arrays. Finally, we assessed the genotyping call rate across

Author contributions: J.M.A., A.L.R., A.K.W., and M.W.N. designed research; J.M.A., A.L.R., D.T.A., C.F.C., J.M., and T.J.N. performed research; A.K.W. contributed new reagents/analytic tools; J.M.A., D.T.A., A.K.W., and M.W.N. analyzed data; and J.M.A. and M.W.N. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

¹To whom correspondence may be addressed. E-mail: akeyj@u.washington.edu or mark.neff@vai.org.

This article contains supporting information online at www.pnas.org/cgi/content/full/0909918107/DCSupplemental.

Table 1. Summary statistics of polymorphism in each breed

| Breed (abbreviation) | <i>n</i> | Average MAF (Var) | Average HE (Var) | Fraction monomorphic |
|----------------------------|----------|-------------------|------------------|----------------------|
| Beagle (BGL) | 26 | 0.238 (0.018) | 0.268 (0.035) | 0.138 |
| Border Collie (BC) | 44 | 0.237 (0.020) | 0.290 (0.034) | 0.097 |
| Brittany (BRT) | 27 | 0.242 (0.019) | 0.279 (0.033) | 0.103 |
| Dachshund (DSH) | 24 | 0.255 (0.018) | 0.279 (0.029) | 0.093 |
| German Shepherd (GSH) | 30 | 0.227 (0.018) | 0.237 (0.037) | 0.205 |
| Greyhound (GRY) | 21 | 0.243 (0.018) | 0.248 (0.032) | 0.171 |
| Jack Russell Terrier (JRT) | 24 | 0.263 (0.019) | 0.332 (0.029) | 0.045 |
| Labrador Retirevers (LBR) | 25 | 0.252 (0.018) | 0.302 (0.033) | 0.097 |
| Shar-Pei (SHP) | 27 | 0.242 (0.019) | 0.287 (0.034) | 0.104 |
| Standard Poodle (STP) | 27 | 0.253 (0.019) | 0.295 (0.032) | 0.099 |

HE, heterozygosity; MAF, minor allele frequency.

all individuals and found uniformly high call rates ($\geq 99\%$). Thus, these analyses suggest that the genotype data are of high quality.

Signatures of Selection in the Canine Genome. A large number of statistical tests have been developed to detect deviations from neutrality (10). We developed a population-genomics strategy based on levels of population differentiation, as it is well suited to detect lineage-specific selective events and is robust to whether selection acts on newly arisen or preexisting variation (14). Specifically, for each SNP we defined a statistic, d_i , which is a function of pairwise F_{ST} (15) between breed i and the remaining breeds. A formal description of d_i is provided in *Methods*, but in words, d_i measures the standardized locus-specific deviation in levels of population structure for breed i relative to the genome-wide average, summed across all pairwise combinations involving breed i . Large positive values indicate loci, with high levels of population structure relative to the genome-at-large. Thus, it is

particularly well suited for detecting selection specific to a particular breed, or subset of breeds, and isolating the direction of change, which is not possible when a single estimate of F_{ST} is calculated across all populations (16). To attenuate the stochastic variation inherent in single-locus estimates of population structure (17), we performed a sliding-window analysis in which d_i values were averaged in nonoverlapping 1-Mb windows throughout the genome.

The genome-wide distribution of d_i is shown in Fig. 1. We define candidate selection regions as outliers falling in the 99th percentile of the empirical distribution of d_i . In total, 155 out of the 1,933 windows met this criterion in one or more of the 10 breeds (Table S1). Several observations suggest that our set of outlier loci is enriched for targets of selection. First, all five genes that have been mapped to date through large-scale association studies of hallmark breed traits are among our list of most differentiated regions: *IGF1* in breeds of small size (7), a locus on

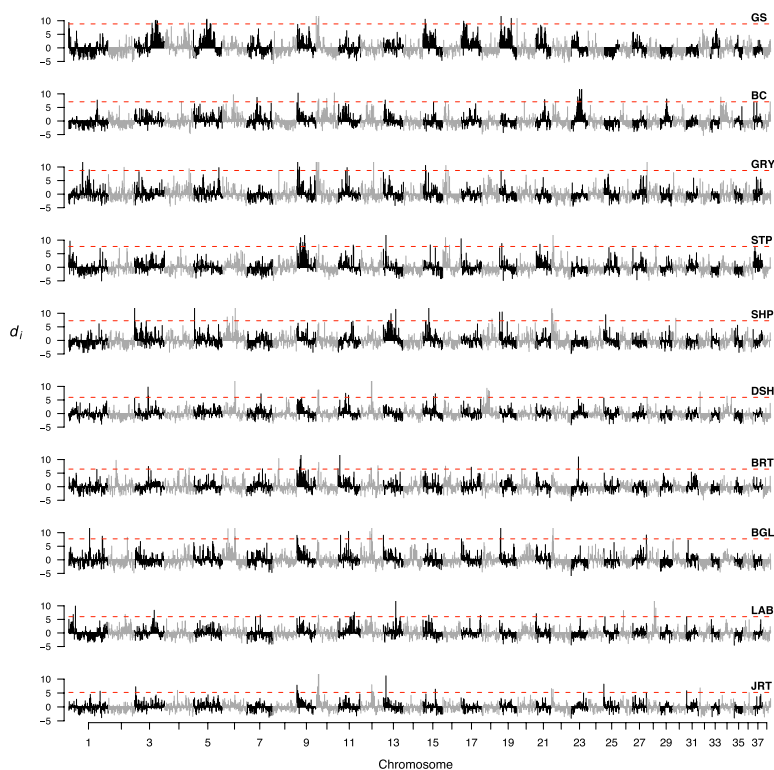


Fig. 1. Genomic distribution of population structure in 10 dog breeds. The distribution of d_i for each 1-Mb interval across all autosomes is shown for each breed. Alternating gray and black indicate values in d_i from adjacent chromosomes. The dashed red line denotes the 99th percentile for each breed. Breeds are abbreviated as described in Table 1.

CFA 18 that is responsible for the characteristic short-limb phenotype in Dachshunds and other breeds (8), and three genes (*RSPO2*, *FGF5*, and *KRT71*) that influence coat phenotypes in many breeds (9).

Second, we performed extensive coalescent simulations that take into account SNP ascertainment and major demographic features, such as population structure and breed-specific bottlenecks. The neutral coalescent model closely recapitulates many characteristics of the observed data, such as average pairwise F_{ST} , average number of markers per 1-Mb window and distribution of minor allele frequencies (Fig. S2). The observed data contains significantly more highly differentiated loci ($P = 1.3 \times 10^{-7}$) compared with the simulated data.

Third, we observed a significant enrichment of signatures of selection in or around genes relative to putatively neutrally evolving regions ($P = 2.95 \times 10^{-3}$), which is expected if adaptive variation is overrepresented in genic regions. Similar observations have also been made in genome-wide scans of selection in humans (18, 19). Collectively, these observations support the hypothesis that the most differentiated regions of the canine genome are enriched for targets of selection.

Shared Versus Unique Selective Events. To investigate how frequently selective events were unique or shared among breeds, we calculated the number of overlapping signatures of selection for each of the 155 significant 1-Mb windows (Fig. 2A). Approximately 103 of the 155 significant windows (~66%) were observed in just one or two breeds (Fig. 2A). These loci likely contain genes that confer breed-restricted phenotypes, such as skin wrinkling in the Shar-Pei (see below). Conversely, 16 of the 155 significant windows (~10%) exhibited signatures of selection in five or more breeds. Such pervasive differentiation at a single locus is consistent with the action of a gene that generally sorts individuals into phenotypic classes and breed groups. For example, one window with strong evidence of selection in multiple breeds is located on CFA15 (43.6–44.6 Mb) and contains

the *IGF1* gene, which governs the miniature size of breeds in the “toy” group (7). Interestingly, a region on CFA 3 (44.6–45.6 Mb) that includes the *IGF1R* gene also shows a strong signature of selection in the Dachshund and Brittany, suggesting that multiple steps in the insulin growth-factor signaling pathway have been substrates of artificial selection in dogs.

One of the most differentiated regions of the canine genome that shows evidence of selection in multiple breeds occurs in three contiguous windows on CFA 10 (Fig. 2B). Sliding-window analyses of pairwise F_{ST} across the 3-Mb interval suggests two or more independent selective events, reflected by two peaks of differentiation with distinct patterns of allele frequency divergence among breeds (Fig. 2B). The peak of differentiation observed from 11.2 to 11.3 Mb coincides with the *HMG2* gene, whose protein product is an integral component of enhancosomes and regulates gene expression (20). In mice, mutations in *HMG2* result in the pygmy phenotype (21), characterized by aberrations in adiposity and disrupted growth leading to dwarfism. In our data, the small-sized breeds (Dachshund, Beagle, Jack Russell Terrier, and Brittany) show high levels of differentiation at *HMG2* compared to the larger-sized breeds (Fig. 2B). At the most differentiated SNP near *HMG2*, allele frequency is significantly correlated with body weight (Pearson $r^2 = 0.68$, $P = 0.003$). Thus, *HMG2* is a strong candidate for mediating variation in size among dogs. The second peak of differentiation in the CFA 10 region (10.35–10.45 Mb), in which the German Shepherd, Jack Russell Terrier, Border Collie, and Greyhound are strongly differentiated from the Dachshund, Beagle, Brittany, and Shar-Pei (Fig. 2B), overlaps two genes, *GNS* and *RASSF3*. *GNS* is a particularly interesting candidate as its avian ortholog (QSulf1), regulates WNT signaling during embryogenesis in myogenic somite progenitors (22).

Overview of Candidate Selection Genes. The 155 candidate selection loci contain 1,630 known or predicted protein-coding genes. To obtain a broad overview into the molecular functions of these

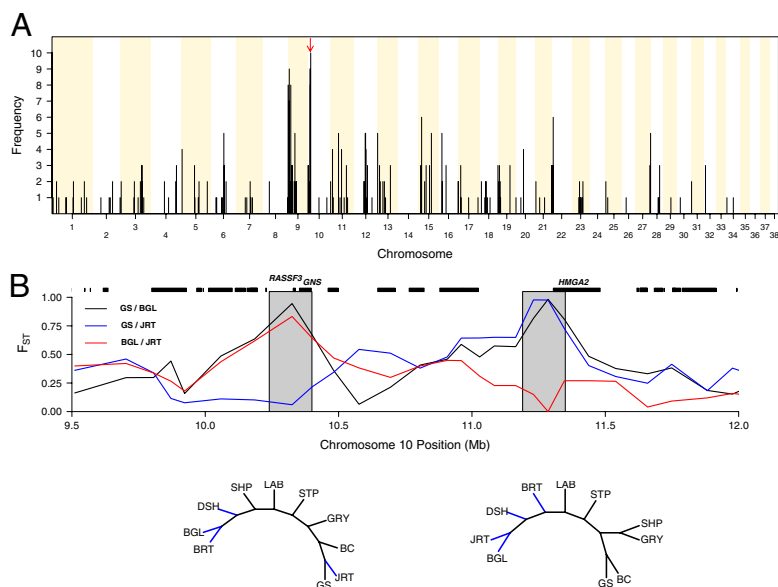


Fig. 2. Shared versus unique signatures of selection. (A) The number of overlapping signatures of selection in each 1-Mb window is shown. We define an overlapping signature of selection for each window if the empirical P value is ≤ 0.01 in one breed and ≤ 0.05 in another breed. Alternating white and vertical light yellow rectangles indicate adjacent chromosomes. The red arrow indicates the chromosomal region shown in *B*. (B) (Upper) Sliding-window analyses of pairwise F_{ST} among German Shepherds, Jack Russell Terriers, and Beagles. Gray boxes indicate two distinct peaks of differentiation (at ≈ 10.3 – 10.4 Mb and 11.3 – 11.4 Mb). Note the different patterns of pairwise F_{ST} in the two peaks of differentiation. Additional breed comparisons have been omitted for clarity. (Lower) Unrooted neighbor joining trees are shown for all breeds inferred from markers in each of the peaks of differentiation above. Note the distinct topology between the two trees (BGL, JRT, DSH, and BRT lineages are shown in blue).

genes and to test the hypothesis that particular functional classes are enriched in the most differentiated regions of the canine genome, we performed a gene ontology (GO) analysis. Table S2 summarizes GO molecular function and biological process terms that are significantly enriched among genes in candidate-selection regions. Similar to analyses of selection in natural populations (23), we find that genes involved in immunity and defense are also significantly overrepresented in the 155 candidate selection regions. This is somewhat surprising, as natural and artificial selection would not necessarily be expected a priori to act on similar classes of genes, and suggests that immune related genes are pervasive targets of selection because of their critical role in pathogen defense or propensity for pleiotropic effects (24).

The average number of genes in each of the 155 candidate selection regions was ≈ 11 . Thus, it is difficult to precisely identify the specific gene that has been influenced by selection. Nonetheless, the 155 most differentiated loci possess many strong candidate genes that influence phenotypes that vary conspicuously among breeds, such as size (*HMG2* and *IGF1R*), coat color and texture (*SILV* and *MITF*), behavior (*CDH9*, *DRD5*, and *HTR2A*), skeletal morphology (*SOX9*), and physiology (*FTO*, *SLC2A9*, and *SLC5A2*).

However, more definitive inferences can be made for eight regions, in which there is only a single protein-coding gene located within the interval (Table S3). Possible phenotypes that each gene may influence are listed in Table S3, and more detailed information is provided in the SI Text. Interestingly, three of the eight genes are transcription factors (*ZFX3*, *SOX9*, and *SATB1*). There has been considerable debate about the relative contribution of changes in gene regulation versus protein structure as mechanisms of evolutionary change (25, 26). Similar to analyses of artificial selection in other domesticated species (27–29), our data suggest that tinkering (30) with gene-expression networks may have played a prominent role in the rapid phenotypic diversification of modern dog breeds.

We note that the stringent threshold used to define candidate selection regions has likely excluded genuine substrates of selection. For example, regions on CFA 9 and 27 lie just beyond our

threshold of significance in Poodles (empirical P -values = 0.021 and 0.014, respectively). These regions contain numerous keratin gene family members, which are important structural proteins of the skin, nails, and hair (Fig. S3). Of particular interest are members of the type-I hair keratins on CFA 9 (*KRT25*, *KRT27*, *KRT28*, *KRT32*, *KRT35*, and *KRT36*) and type-II hair keratins on CFA 27 (*KRT71*, *KRT72*, *KRT73*, *KRT74*, *KRT82*, *KRT84*, and *KRT85*). Recently, variation in *KRT71* has been associated with curly coat phenotypes in several breeds (9), which validates our CFA 27 results. Our data suggest that additional keratin genes on CFA 9 are also strong candidates for contributing to the curly coat phenotype.

Regulatory Variation in *HAS2* Is Associated with Skin Wrinkling of Shar-Peis.

To characterize candidate selection genes in more detail, we focused on a region on CFA 13 with evidence of selection in the Shar-Pei (Fig. 3A) that contains three genes (*SNTB1*, *FTSJ1*, and *HAS2*). A distinguishing characteristic of the Shar-Pei is cutaneous mucinosis, or excessive skin wrinkling. The degree of skin folds correlates with high mucin content histologically and elevated levels of hyaluronic acid biochemically (31). *HAS2*, which is a hyaluronic acid synthase, was thus a strong candidate gene. In addition, rare mutations in human *HAS2* have been described that result in severe cutaneous mucinosis (32).

To test the hypothesis that genetic variation in *HAS2* contributes to skin wrinkling, we exploited the intrabreed phenotypic variation that exists in the degree of wrinkling within Shar-Pei (Fig. 3B). Specifically, we sequenced ≈ 3.7 kb of *HAS2* (including all exons, intron/exon boundaries, and untranslated regions) (Fig. 3C) from 32 wrinkled and 18 smooth-coated purebred Shar-Pei (Fig. 3B). In total, we discovered five polymorphisms, none of which are located in coding regions (Table S4). One of the upstream polymorphisms was nearly fixed in both wrinkled and smooth dogs and was not considered further (Table S4). Association mapping was performed with a permutation-based Cochran-Armitage trend test (33) on the four remaining SNPs, all of which demonstrated significant differences in genotype frequencies between wrinkled versus smooth Shar-Pei (Table S4).

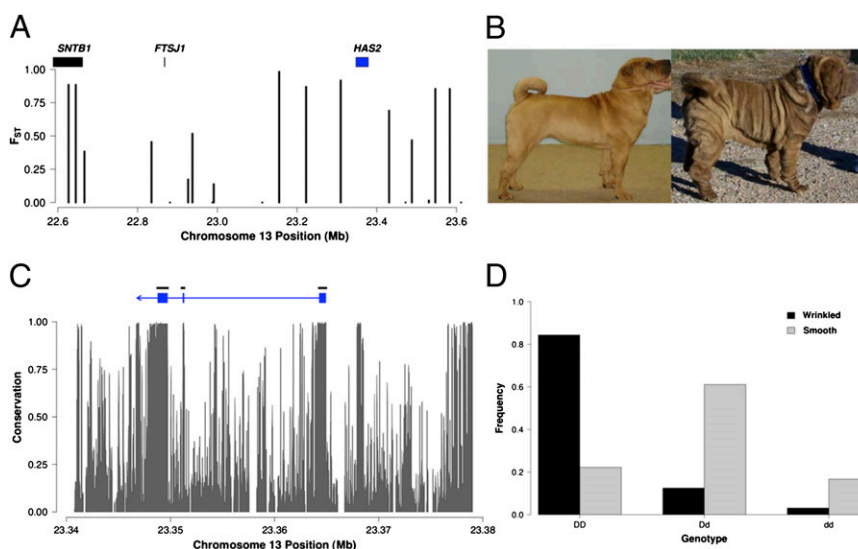


Fig. 3. Genetic variation in *HAS2* is associated with skin wrinkling in Shar-Pei. (A) Single locus estimates of F_{ST} between Shar-Pei and Dachshund across a 1-Mb window. Similar patterns were observed for Shar-Pei compared to other breeds, but have been omitted for clarity. The location of all protein-coding genes are shown as rectangular boxes. (B) An example of smooth (Left) and wrinkled (Right) Shar-Pei dogs. (C) Exon structure of *HAS2*. Conservation values obtained from the University of California Santa Cruz genome browser are shown below. Black horizontal lines indicate sequenced regions. (D) Genotype frequencies of the intron 2 indel (site 13805 in Table S4) in smooth and wrinkled Shar-Pei, which are significantly different ($P = 6.28 \times 10^{-5}$). Deletion and insertion alleles are denoted as "D" and "d," respectively.

We next sequenced all of the *HAS2* amplicons in a diverse panel of 94 dogs derived from 20 breeds (Table S5). The most differentiated SNP between the Shar-Pei and other breeds is a 2-bp indel ≈ 86 bp 3' of exon 2 (Table S5). The deletion allele is significantly associated with the wrinkling phenotype ($P = 6.28 \times 10^{-5}$; see site 13805 in Table S4 and Fig. 3D), where the frequency of the deletion allele is ≈ 0.91 and 0.53 in wrinkled and smooth Shar-Pei, respectively. The deletion allele is rare outside of the Shar-Pei ($\sim 1.6\%$) (Table S5) and no homozygous deletions were found in any of the 94 dogs.

Although experimental studies will ultimately be necessary to determine whether the polymorphisms described in Table S4 are functionally important, it seems unlikely that any of them are causally related to skin wrinkling in the Shar-Pei. The most strongly associated polymorphism (site -424) is common across breeds (Table S5). Even though the intron 2 polymorphism does possess patterns of variation between the Shar-Pei and other breeds expected for a causal polymorphism, it is not located in a region of high sequence conservation and is not embedded in any obvious regulatory elements. Therefore, we hypothesize that the polymorphisms in Table S4, and in particular the 2-bp intron 2 indel, are in linkage disequilibrium (LD) with the unidentified causative allele. As no variation was found in the *HAS2* coding region, the causal allele is likely a regulatory polymorphism. Consistent with this hypothesis, of the 50 Shar-Pei dogs in the resequencing panel, 22 overlap with the set of individuals used for Illumina SNP genotyping. As shown in Fig. S4, the strongest associations for this subset of individuals in the SNP data occur upstream of the *HAS2* gene, suggesting the causal polymorphism lies 5' to *HAS2*.

Discussion

The extensive phenotypic diversity that exists between dog breeds has long been recognized as a unique portal into the genetic architecture of phenotypes. However, much of this phenotypic variation has been refractory to traditional genetic mapping because traits of interest, such as morphology and behavior, largely vary between but not within breeds. This conundrum, referred to as the “segregation problem” (3), has only recently been addressed by genome-wide association mapping of phenotypes between breeds (7–9). Here, we describe a complementary approach to the segregation problem that is agnostic to phenotypes by identifying regions of the dog genome that exhibit signatures of artificial selection. In total, we identified 155 loci that possess strong signatures of recent selection, including all five genes previously identified by whole-genome association studies of hallmark breed traits (7–9). Our selected regions also contain many previously unconsidered candidate genes that contribute to phenotypic variation among breeds. Thus, the combination of genome-wide association mapping between breeds and hitchhiking mapping (34) such as has been pursued here is poised to rapidly dissect phenotypic variation in dogs.

Despite the insights gleaned from our data, it is important to note several limitations and challenges. Most importantly, simply possessing a pattern of variation that is unusual relative to the genome at large does not prove that a locus is under selection (10). Indeed, the stochastic variation in gene genealogies among dog breeds is expected to be large, given the dramatic demographic perturbations that canine populations have experienced. Ultimately, a denser map of polymorphisms in a wider collection of breeds will allow additional tests of neutrality to be performed (10), and positions of putatively selected variation to be refined.

In interpreting the signatures of selection that we identified, we have leveraged information about gene function from other species, particularly humans. For example, rare mutations in human *HAS2* have been described (32) as resulting in cutaneous mucinosis. As another example, there is a strong signature of selection that is coincident with the *FTO* gene in Beagles. A number of well-

replicated studies in humans have demonstrated that variation in *FTO* contributes to variation in body mass index and related metabolic traits (35), suggesting that this gene influences similar phenotypes in Beagles. However, the portability of genotype-phenotype correlations need not move exclusively from humans to dogs. Indeed, a motivating factor driving canine genomics is the potential to inform the genetic basis of human phenotypic variation and disease susceptibility (2, 12). Thus, delineating the phenotypic effects of selected variation in dogs holds considerable promise for providing unique insights into the genetic basis of heritable phenotypic variation in humans.

Similarly, fine-scale mapping signatures of selection in dogs may also facilitate the interpretation and resolution of genome-wide scans of selection in humans. Specifically, numerous genome-wide analyses of selection have been performed in humans that generally delimit broad genomic regions, leaving the precise target of selection ambiguous. We anticipate that in many cases it will be easier to localize substrates of selection in dogs, which can then be mapped to syntenic regions in humans. A selected gene in dogs that is located within a putatively selected locus in humans can engender testable hypotheses to fine-scale map-selected loci in humans. We note that as an initial foray into comparative selection mapping, of the 1,506 genes located in putatively selected regions in dogs, 169 overlap with genes located in well-supported selected regions in humans (10). Although this result should be interpreted with caution, as the specific targets of selection are generally not known with certainty in either dogs or humans, it does raise the intriguing possibility that recent selection has influenced common loci in both the human and dog lineages.

A better understanding of artificial selection in dogs will also provide important mechanistic insight into the molecular basis of rapid short-term evolution. Of particular interest will be to define the number of loci responsible for shaping the incredible diversity of form and function among the worlds >400 breeds, the types of genes and genetic variation therein that have responded to artificial selection, and whether adaptive alleles are dominant, recessive, or additive. Although our results do not provide definitive answers to these issues, they do afford some insight into the mechanistic basis of artificial selection. Specifically, there has been considerable debate into the relative contribution of protein versus regulatory variation in mediating evolutionary change (25, 26). Although both coding and noncoding alleles certainly contribute to canine phenotypic variation, the observation that several transcription factors (*ZFX3*, *SOX9*, and *SATB1*) were mapped to single-gene resolution in candidate-selection regions, and the functional *HAS2* allele for skin wrinkling in Shar-Pei is likely in a noncoding region, suggests that regulatory variation has been a sizeable target for artificial selection.

In summary, the continued maturation of dog genomics has created the opportunity to systematically identify loci that manifest signatures of selection, which will facilitate the genetic dissection of phenotypic variation. In particular, a canine genomic map of selection provides a roadmap to functional genetic variation that underlies breed-specific differences in behavior, morphology, physiology, and disease susceptibility. In addition, the resolution of selected loci into adaptive alleles will provide critical insights into the types of molecular variation that mediate rapid phenotypic diversification. Ultimately, a deeper understanding of artificial selection in dogs and other domesticated species may inform mechanisms of evolutionary change in natural populations, and illuminate the similarities and differences in how artificial and natural selection alter the evolutionary trajectory of populations.

Methods

DNA Samples and SNP Genotyping. Purebred dogs from the ten breeds described in Table 1 were sampled for large-scale SNP genotyping. Two trios were also collected per breed to verify Mendelian transmission of SNPs. For the *HAS2* association study in the Shar-Pei, phenotypic data were available

for 22 of the dogs used in large-scale SNP genotyping and 28 additional Shar-Pei samples were collected for a total sample size of 50. Furthermore, *HAS2* was resequenced in a panel of 94 diverse dogs from 20 breeds (Table S5). For all samples, DNA was prepared from blood or buccal swab samples using previously described methods (36, 37). Buccal swab samples were treated by whole genome amplification using GenomePlex for tissue (Sigma). All sample collections were approved by the Animal Care and Use Committee of the University of California, Davis (IACUC protocol 12682). DNA was genotyped at 22,362 SNP loci with the Infinium CanineSNP20 BeadChip. Genotyping was performed according manufacturer's instructions and data were collected with an Illumina BeadStation scanner. Genotypes were scored using BeadStudio.

Statistical and Bioinformatics Analyses. Although pedigree relationships could be verified for $\approx 74\%$ of all individuals to ensure they were unrelated by at least three generations, to be rigorous we also used the RELPAIR software (38) to infer putative relationships directly from genotype data in all samples. Of the initial 297 dogs genotyped, RELPAIR identified 22 pairs of presumptively related individuals. We randomly selected one individual from each pair, yielding the final set of 275 samples.

Exact tests of Hardy-Weinberg equilibrium were performed for each SNP and in each breed as previously described (39). SNPs that rejected the null of hypothesis of Hardy-Weinberg equilibrium at $P < 10^{-5}$ (0.05/22,000), possessed more than two alleles, exhibited Mendelian inconsistencies in the trio analysis, were located on the X-chromosome, or had $> 10\%$ missing data within breeds were excluded from further analysis. Our final data set consisted of 21,114 SNPs that passed these criteria in all 10 breeds.

We developed a simple summary statistic to measure the locus specific divergence in allele frequencies for each breed based on unbiased estimates of pairwise F_{ST} (15). In particular, for each SNP we calculated the statistic

$$d_i = \sum_{j \neq i} \frac{F_{ST}^{ij} - E[F_{ST}^{ij}]}{sd[F_{ST}^{ij}]}, \text{ where } E[F_{ST}^{ij}] \text{ and } sd[F_{ST}^{ij}] \text{ denote the expected value and}$$

standard deviation of F_{ST} between breeds i and j calculated from all 21,114 SNPs. For each breed, d_i was averaged over SNPs in nonoverlapping 1-Mb windows. The average number of SNPs per window was 9.5 and windows with fewer than four SNPs were discarded. We performed standard linear regression in R with the function "lm" to adjust window specific estimates of d_i for the number of SNP markers and average heterozygosity and found that it did not significantly affect the results ($P > 0.05$). Principal components analysis was performed in R with the "svd" function as previously described (40).

Coalescent Simulations. Coalescent simulations were performed with the software MS (41) using demographic parameters that were found to closely recapitulate features of the observed data such as average pairwise F_{ST} among breeds, average minor allele frequencies, and average number of SNPs per window. See Fig. S2 for more details.

HAS2 Resequencing and Association Mapping. Sequencing primers were designed from published dog sequence (NM_015120) with primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (primer sequences are available upon request). We used standard PCR-based sequencing reactions using Applied Biosystem's Big Dye sequencing protocol on an ABI 3130xl and analyzed the sequencing data as previously described (18). All polymorphic sites were manually verified. Association of *HAS2* variation with skin wrinkling was performed with permutation-based Cochran-Armitage trend test (33).

ACKNOWLEDGMENTS. This work was supported by research Grant 1R01GM076036-01A1 from the National Institutes of Health and a Sloan Fellowship in Computational Biology (to J.M.A.).

- American Kennel Club (1998) in *The Complete Dog Book*, eds American Kennel Club Staff (Howell Book House, Foster City).
- Sutter NB, Ostrander EA (2004) Dog star rising: the canine genetic system. *Nat Rev Genet* 5:900–910.
- Neff MW, Rine J (2006) A fetching model organism. *Cell* 124:229–231.
- Vilà C, et al. (1997) Multiple and ancient origins of the domestic dog. *Science* 276:1687–1689.
- Leonard JA, et al. (2002) Ancient DNA evidence for Old World origin of New World dogs. *Science* 298:1613–1616.
- Darwin C (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. (John Murray, London) 1st Ed.
- Sutter NB, et al. (2007) A single IGF1 allele is a major determinant of small size in dogs. *Science* 316:112–115.
- Parker HG, et al. (2009) An expressed *fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science* 325:995–998.
- Cadiou E, et al. (2009) Coat variation in the domestic dog is governed by variants in three genes. *Science* 326:150–153.
- Akey JM (2009) Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res* 19:711–722.
- Pollinger JP, et al. (2005) Selective sweep mapping of genes with large phenotypic effects. *Genome Res* 15:1809–1819.
- Lindblad-Toh K, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438:803–819.
- Karlsson EK, et al. (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat Genet* 39:1321–1328.
- Innan H, Kim Y (2008) Detecting local adaptation using the joint sampling of polymorphism data in the parental and derived populations. *Genetics* 179:1713–1720.
- Weir BS (1996) *Genetic Data Analysis II* (Sinauer Associates, Inc. Publishers, Sunderland).
- Shriver MD, et al. (2004) The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum Genomics* 1:274–286.
- Weir BS, Cardon LR, Anderson AD, Nielsen DM, Hill WG (2005) Measures of human population structure show heterogeneity among genomic regions. *Genome Res* 15:1468–1476.
- Kelley JL, Madeoy J, Calhoun JC, Swanson W, Akey JM (2006) Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Res* 16:980–989.
- Voight BF, Kudravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. *PLoS Biol* 4:e72.
- Groschedl R, Giese K, Pagel J (1994) HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. *Trends Genet* 10:94–100.
- Zhou X, Benson KF, Ashar HR, Chada K (1995) Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. *Nature* 377:771–774.
- Dhoot GK, et al. (2001) Regulation of Wnt signaling and embryo patterning by an extracellular sulfatase. *Science* 293:1663–1668.
- Kosiol C, et al. (2008) Patterns of positive selection in six mammalian genomes. *PLoS Genet* 4:e1000144.
- Ye YH, Chenoweth SF, McGraw EA (2009) Effective but costly, evolved mechanisms of defense against a virulent opportunistic pathogen in *Drosophila melanogaster*. *PLoS Pathog* 5:e1000385.
- Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61:995–1016.
- Wray GA (2007) The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 8:206–216.
- Wang RL, Stec A, Hey J, Lukens L, Doebley J (1999) The limits of selection during maize domestication. *Nature* 398:236–239.
- Cong B, Barrero LS, Tanksley SD (2008) Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. *Nat Genet* 40:800–804.
- Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell* 127:1309–1321.
- Jacob F (1977) Evolution and tinkering. *Science* 196:1161–1166.
- Zanna G, et al. (2008) Cutaneous mucinosis in Shar-Pei dogs is due to hyaluronic acid deposition and is associated with high levels of hyaluronic acid in serum. *Vet Dermatol* 19:314–318.
- Ramsden CA, et al. (2000) A new disorder of hyaluronan metabolism associated with generalized folding and thickening of the skin. *J Pediatr* 136:62–68.
- Agresti A (2002) *Categorical Data Analysis*, eds Agresti A, David HA (Wiley, New Jersey), 2nd Ed, pp 165–196.
- Harr B, Kauer M, Schlotterer C (2002) Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 99:12949–12954.
- Frayling TM, et al. (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316:889–894.
- Bell GI, Karam JH, Rutter WJ (1981) Polymorphic DNA region adjacent to the 5' end of the human insulin gene. *Proc Natl Acad Sci USA* 78:5759–5763.
- Oberbauer AM, et al. (2003) Alternatives to blood as a source of DNA for large-scale scanning studies of canine genome linkages. *Vet Res Commun* 27:27–38.
- Epstein MP, Duren WL, Boehnke M (2000) Improved inference of relationship for pairs of individuals. *Am J Hum Genet* 67:1219–1231.
- Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet* 76:887–893.
- Biswas S, Scheinfeldt LB, Akey JM (2009) Genome-wide insights into the patterns and determinants of fine-scale population structure in humans. *Am J Hum Genet* 84:641–650.
- Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.
- Guedj M, Wojcik J, Della-Chiesa E, Nuel G, Forner K (2006) A fast, unbiased and exact allelic test for case-control association studies. *Hum Hered* 61:210–221.