

Large-Scale SNP Genotyping with Canine Buccal Swab DNA

MELANIE LEE CHANG, REBECCA LEE TERRILL, MARIA M. BAUTISTA, ELAINE J. CARLSON, DONNA J. DYER, KAREN L. OVERALL, AND STEVEN P. HAMILTON

From the Department of Psychiatry and Institute for Human Genetics, Langley Porter Psychiatric Institute, University of California San Francisco, Box 0984-NGL, San Francisco, CA 94143-0984 (Chang, Terrill, Bautista, and Hamilton); the Genomics Core Facility, University of California San Francisco, San Francisco, CA 94143-0984 (Carlson); and the Center for Neurobiology and Behavior, Psychiatry Department, University of Pennsylvania, Philadelphia, PA 19104-3403 (Dyer and Overall).

Address correspondence to S. P. Hamilton, MD, PhD, at the address above, or e-mail: stevch@ppi.ucsf.edu.

Abstract

The dog is an attractive model for genetic studies of complex disease. With drafts of the canine genome complete, a large number of single-nucleotide polymorphisms (SNPs) that are potentially useful for gene-mapping studies and empirical estimations of canine diversity and linkage disequilibrium (LD) are now available. Unfortunately, most canine SNPs remain uncharacterized, and the amount and quality of DNA available from population-based samples are limited. We assessed how these real-world challenges influence automated SNP genotyping methods such as Illumina's GoldenGate assay. We examined 384 SNPs on canine chromosome 9 and successfully genotyped a minimum of 217 and a maximum of 275 SNPs using buccal swab samples for 181 dogs (86 beagles, 76 border collies, and 15 Australian shepherds). Call rates per SNP and sample averaged 97%, with reproducibility within and between analyses averaging 98%. The majority of these SNPs were polymorphic across all 3 breeds. We observed extensive LD, albeit less than reported for surveys using fewer dogs, consistent between breeds. Analyses of population substructure indicated that beagles are distinct from border collies and Australian shepherds. These results demonstrate the suitability of amplified canine buccal samples for high-throughput multiplex genotyping and confirm extensive LD in the dog.

Over the past decade, a number of researchers have advocated the use of the dog as a model system for understanding the genetic basis of disease, morphology, and behavior (Barsoum et al. 2000; Ostrander and Kruglyak 2000; Ponder et al. 2002; Galibert et al. 2004; Sutter and Ostrander 2004). Due to its history of domestication (Savolainen et al. 2002), the dog offers a number of advantages for gene-mapping studies. Most extant breeds were developed through selection for specific types of tasks or work (Clutton-Brock 1999; Koskinen and Bredbacka 2000) and are less than 150 years old, resulting in reduced heterogeneity within breeds and increased heterogeneity between breeds (Parker et al. 2004). As breeds represent canalization of genetic variation, different breeds may vary with respect to genes responsible for biological processes, including growth, cognitive development, and liability for the development of behavioral or disease pathologies.

Gene-mapping studies require adequate numbers of informative markers (Ostrander and Kruglyak 2000; Brooks and Sargan 2001). Large numbers of short tandem repeat

markers (STRs) have been described for the dog (Guyon et al. 2003), with a subset defined for linkage studies (Clark et al. 2004). With draft sequences of the canine genome complete (Kirkness et al. 2003; Lindblad-Toh et al. 2005), a large number of single-nucleotide polymorphisms (SNPs) are now available for the domestic dog. A small subset of these have already been used to develop empirical estimations of diversity and linkage disequilibrium (LD) in the canine genome (Sutter et al. 2004; Lindblad-Toh et al. 2005), and panels of SNPs for linkage analysis and whole-genome association studies are being developed to facilitate the identification of traits of interest. However, almost all known canine SNPs remain uncharacterized, and what would constitute an adequate set of SNPs for LD-mapping efforts remains unknown. Although recently developed platforms for genotyping large numbers of SNPs (i.e., Illumina, Inc., San Diego, CA; Affymetrix, Inc., Santa Clara, CA) can facilitate efficient data collection, the feasibility of large-scale genetic analyses in dogs is complicated by a number of challenges; the potential for low net yield of canine DNA from population-based

samples (characteristically extracted from buccal swabs) is particularly daunting. The amount of target DNA that can be retrieved from canine buccal swab samples is low, and the total DNA extracted from buccal swabs is invariably contaminated by large amounts of microbial DNA.

For the present study, we assessed how real-world challenges inherent to genetic investigations in the dog influence automated SNP genotyping methods. We chose 384 SNPs on canine chromosome 9 from the Broad Institute's set of approximately 2.55 million uncharacterized canine SNPs (Lindblad-Toh et al. 2005), and evaluated their performance on the Illumina BeadArray genotyping platform using whole-genome–amplified DNA samples extracted from buccal swabs. In the process, we were able to use the data produced for an assessment of LD and population structure in a sample of 86 purebred beagles, 76 purebred border collies, and 15 purebred Australian shepherds. Our results indicate that amplified buccal samples yield enough canine DNA for use in standard high-throughput multiplex genotyping assays and confirm that LD is extensive in the dog. To our knowledge, no other group has successfully genotyped from canine buccal swab DNA samples using the Illumina platform.

Methods

DNA Sample Collection and Storage

We included a sex-balanced sample of 90 beagles, 76 border collies, and 15 Australian shepherds in these analyses, choosing these 3 breeds for analysis due to existing research priorities that have resulted in the accumulation of large in-house samples of each breed. Canine subjects came from the United States and Canada and were selected from a larger sample of more than 3000 dogs collected for a survey of overall diversity within and between breeds as part of the Canine Behavioral Genetics Project, a study exploring the genetic background of anxiety-related behaviors in domestic dogs (<http://psych.ucsf.edu/K9BehavioralGenetics>). Participating owners are recruited from the community and asked to submit buccal swabs and behavioral questionnaires for each dog (Overall et al. 2006). Therefore, as in typical population-based surveys, the DNA samples used in these analyses were collected by diverse owners and technicians. Owners were asked to sample dogs at least 25 min after eating, to cleanse dogs' mouths with water or encourage them to drink water before swabbing, and to allow swabs to air-dry before packaging them for return to minimize the potential for contamination and bacterial growth. Canine buccal swab samples (Cytosoft; Medical Packaging Corporation, Camarillo, CA) were received from owners and technicians via post and stored at room temperature in 2.0-ml microcentrifuge tubes prior to extraction.

DNA Extraction, Amplification, and Quantification

Genomic DNA was extracted from 3 buccal swabs per dog using a modified protocol designed specifically for buccal swabs, with the 3 swabs extracted together in a single reaction (QIAamp DNA Mini Kit; QIAGEN, Inc., Valencia, CA).

Table 1. Breeds and samples used in analyses

Breed	Number of individuals	Type of sample
Beagle	90	Buccal
Border collie	76	Buccal
Border collie ^a	2	Blood
Papillon	1	Blood
Mixed breed	1	Blood
Unknown purebred ^b	2	Blood

^a Two border collies included as buccal swab samples were replicated with blood-drawn DNA samples.

^b Samples were from a multibreed group of military working dogs that included German shepherd dogs, Belgian shepherds, and Labrador retrievers.

Typical yields from 3 buccal swabs were 1–2 µg per 100 µl, compared with a typical yield of 30 µg of DNA per 1 ml of whole blood (Lench et al. 1988). Genomic DNA was extracted from whole-blood samples (PureGene; QIAGEN, Inc., Valencia, CA) from 2 of the border collies (replicates) and 4 unrelated dogs of different breeds (unique samples) and included in analyses for comparison (see Table 1).

Buccal swab samples (1 µl of a 100-µl extraction) were subjected to whole-genome amplification (WGA), using a method that employs bacteriophage Φ29 DNA polymerase enzyme multiple displacement amplification (MDA) (GenomiPhi; GE Healthcare, Buckinghamshire, England), in triplicate. DNA from blood, buccal, and whole-genome–amplified buccal swab samples has been shown to be comparable in quality using microsatellites and SNPs (Barker et al. 2004; Short et al. 2005; Thompson et al. 2005), and the specific utility of MDA WGA product for Illumina BeadArray genotyping has also been demonstrated (Pask et al. 2004). In earlier experiments, we routinely obtained high-quality DNA sequence from polymerase chain reaction (PCR) using buccal samples, as well as microsatellite analyses from PCR templates using buccal samples. We have also found that amplifying samples in triplicate followed by pooling of the samples leads to 100% concordance between amplified and source genomic DNA microsatellite genotypes (Bravo O, Hamilton SP, unpublished data).

Whole-genome–amplified samples were quantified using a standard PicoGreen analysis (Invitrogen Corp., Carlsbad, CA) to determine total DNA concentration. All liquid handling was performed using TECAN Genesis 150 8-probe and Hamilton Microlab 4200 12-probe robotic liquid handlers. Additionally, a PCR amplicon predicted to occur as a single copy in the canine genome was amplified and subjected to quantitative PCR using the double-stranded DNA intercalating dye SYBR Green (Applied Biosystems, Foster City, CA), allowing specific quantification of the amount of canine DNA in the extracted buccal swab sample. This step was necessary because canine buccal swab DNA samples are invariably contaminated by significant amounts of microbial DNA. For this assay, we designed a single pair of primers (5'-TCCCACTGTTGACAGAAGTGAA-3' and 5'-TGCTTCAAGTTCTGGGTTATGG-3') for an amplicon on canine chromosome 9 that was verified as a single copy in the dog genome by sequence alignment on the

May 2005 assembly of the canine genome using the University of California Santa Cruz Genome Browser's BLAT function (<http://genome.ucsc.edu>). PCR amplification was performed in a 10- μ l reaction containing 1 μ l WGA DNA sample, 1 \times Power SYBR Green PCR Master Mix (Applied Biosystems), and the above primers at 150 nM final concentration. Samples were cycled at 50 °C for 2 min, 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s, and 60 °C for 1 min on an ABI 7900HT DNA analyzer. We included a dissociation step (95 °C for 15 s, 60 °C for 15 s, ramping at a 2% ramp rate to 95 °C, and holding for 15 s) to ensure that only the specific product was amplified. Final DNA amounts were highly variable between samples. Canine DNA typically represents 3–15% of the total genomic DNA, the balance presumably being from oral microbial flora. This proportion is much lower than published estimates (Min et al. 2006) and our own in-house data with human buccal swabs. The reason for this lower yield is currently obscure.

Marker Selection

Because our primary objective was to assess the utility of buccal swab samples on the Illumina genotyping platform, our criteria for selecting SNPs were not governed by chromosomal coverage, representativeness of chromosomal regions, or SNP density. We chose to use SNPs from a single canine chromosome and planned ultimately to survey approximately 384 SNPs equally spaced across this chromosome, a number dictated by the design constraints of the Illumina genotyping platform. The first assembly of the boxer genome (CanFam 1.0, July 2004) included 55 235 SNPs assigned to chromosome 9. The Broad Institute used 9 additional breeds for SNP discovery, generating 100 000 whole-genome shotgun reads for each breed and 25 000 reads each for 4 gray wolves and 1 coyote (Lindblad-Toh et al. 2005).

Canine chromosome 9 was chosen for this assay because it contains a region syntenic with the human serotonin transporter gene (5-HTT) and is therefore of particular interest for studies of behavioral genetics in the dog. We downloaded all chromosome 9 SNPs (CanFam 1.0) from the Broad Institute Web site (<http://www.broad.mit.edu/mammals/dog/>) and then filtered this set for SNPs observed in beagles, the original target breed for this experiment. Preference was given to SNPs that were polymorphic in any additional dog breeds or wild canids. As the Broad Institute originally identified SNPs using shotgun sequence of relatively few representatives of each breed, and we then selected SNPs from this set that are seen across breeds, the canine SNPs we employed in these analyses are not likely to be rare. We selected ~1100 SNPs for Illumina's GoldenGate design, and submitted the flanking genomic region and polymorphic bases for each SNP to Illumina for a design score. Approximately 700 SNPs with design scores >0.5 were then sorted to allow relatively equal spacing between markers, resulting in the identification of 384 markers. These markers span 52.1 Mb of a 64.4 Mb chromosome (based on map distances from the latest assembly of the dog genome, CanFam 2.0, May 2005), with no repeats in flanking regions and no masked sequence (no repeats closer

than 20 bp to flanking regions). The map positions for these markers changed considerably between genome builds, with an additional 16 Mb of sequence added to the end of the chromosome closest to the centromere, moving the most proximal SNPs from 3 to 19 Mb in position. The ratio of transitions to transversions was 3:1. Average intermarker distance was $139\,276 \pm 10\,720$ bp (median intermarker distance: 67 604 bp), with a maximum distance between markers of 1 688 210 bp and a minimum of 1084 bp. National Center for Biotechnology Information RS identification numbers and map positions (CanFam 2.0) for all genotyped SNPs are available in Supplementary Table.

Data Generation and Analysis

Genotyping

DNA samples, including those extracted from buccal swabs, along with several control DNAs isolated from blood, were analyzed using Illumina's GoldenGate genotyping assay (Steemers and Gunderson 2005; Fan et al. 2006) as per manufacturer instructions. Protocols for this assay recommend a minimum of 250 ng of target DNA per sample. To compensate for sample quality, we assayed DNA samples quantified at 1 μ g total DNA by PicoGreen.

We analyzed a total of 183 samples derived from buccal swabs, including 90 beagles, 76 border collies (with 2 replicates for a total of 78 buccal swab samples from border collies), and 15 Australian shepherds. Several internal laboratory control DNA samples, derived from blood samples, were also used. Three samples were included in duplicate or triplicate to verify genotype reproducibility.

Data were output in the form of intensity files and analyzed using Illumina's BEADSTUDIO software suite, which offers automated genotype clustering and calling, and allows data to be visualized for further analysis. We removed samples with poor quality scores from the analysis (p10gc score < 0.40) and then reclustered the SNPs excluding these samples. Next, we dropped SNPs with <60% call rate. The remaining SNPs were evaluated by cluster separation score and then visually evaluated for call integrity. Representative acceptable and dropped SNPs are shown in Figure 1.

To include almost 200 dogs of 3 different breeds, we split our samples into 2 analyses using separate Sentrix Array Matrices (SAMs), which accommodate 96 samples per SAM. The first analysis included buccal DNA samples from 90 purebred beagles and blood DNA samples from 6 additional unrelated dogs (2 border collies, 2 unknown purebreds, 1 papillon, and 1 mixed breed dog). The second analysis included buccal samples from 76 purebred border collies and 15 Australian shepherds and blood samples from 2 of the border collies (replicates) and 1 mixed breed dog. The 2 border collies represented by both buccal and blood DNA samples were unrelated in-house dogs (owned by an author). These border collies and the mixed breed dog were represented on both SAMs.

Genotype data for a total of 86 beagle samples (after dropping underperforming samples; see above) and 217 quality-controlled SNPs (see below) were prepared from the results

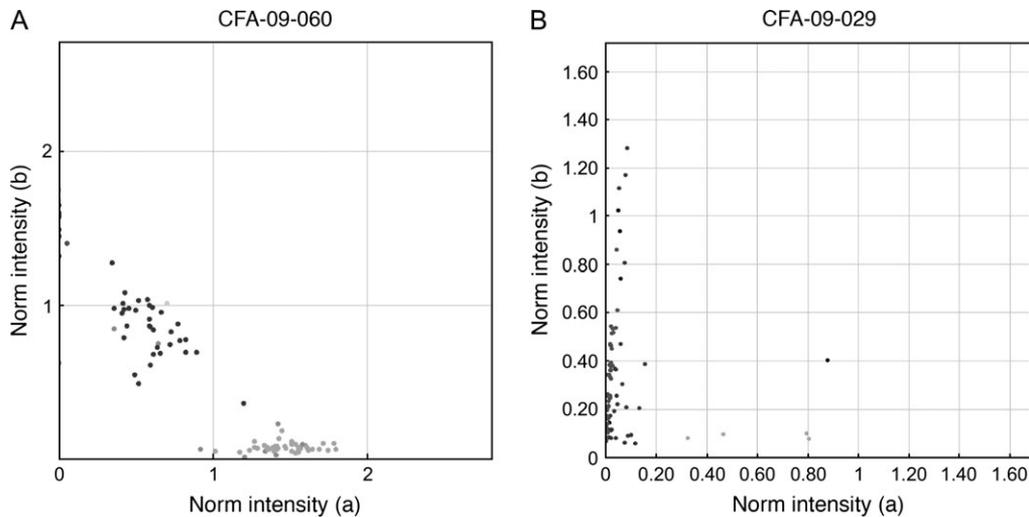


Figure 1. Representative acceptable and excluded SNPs. (A) Clustering typical of an acceptable SNP. (B) Clustering typical of a rejected SNP.

of the first analysis, whereas 275 SNPs for the border collie/Australian shepherd sample were genotyped in 91 purebred dogs (this SAM included triplicates for 2 border collies, a purebred papillon, and a mixed breed control dog; no samples were dropped) in the second analysis. A total of 188 SNPs were successfully assayed in all 3 breeds between analyses.

Analyses of LD and Population Structure

Genotypes were imported into HAPLOVIEW (Barrett et al. 2005). Due to unresolvable discrepancies in map position between CanFam 1.0 and CanFam 2.0, we omitted 2 markers from our analyses of LD in beagles and 6 markers from our analyses of LD in border collies and Australian shepherds. We performed pairwise LD comparisons between markers within 10 Mb of one another and excluded markers from analysis that had call rates $<75\%$ and Hardy–Weinberg equilibrium deviation ($P < 0.0001$). Graphical representations of LD were generated with HAPLOVIEW. We identified tagging SNPs using the TAGGER routine (de Bakker et al. 2005) implemented in HAPLOVIEW.

Finally, to determine whether or not the SNP data we obtained would allow us to explore population structure in the 3 breeds examined, we ran simulations in STRUCTURE (Pritchard et al. 2000), which uses a model-based algorithm for clustering samples. We examined 176 individuals and 188 loci and incorporated information about population identity (beagle, border collie, or Australian shepherd) in these analyses. Each simulation incorporated 20 000 burn-ins, followed by 100 000 repetitions, optimizing over $K = 1$ through $K = 6$, and generated estimated proportion of membership for each predefined population. Five simulations were completed for each K , and the summary statistics averaged over all 5 runs. The estimated posterior probabilities for the data at each K , $\Pr(X|K)$ (“LnPD” in STRUCTURE output), and ΔK (Evanno et al. 2005) were used to identify the best model fit.

Results

Hypotheses

We sought to address 3 interlocking methodological hypotheses. First, we examined the suitability of amplified canine buccal swab DNA for high-throughput genotyping. Second, we tested whether an uncharacterized set of SNPs would be useful for studies of genetic diversity in dogs. Third, we evaluated the hypothesis that these SNPs would be useful for measuring diversity across breeds as well as within breeds, despite being ascertained from a single breed.

SNP Performance and Quality

Beagles

We identified 384 SNPs based on sequence differences between shotgun reads of the beagle and the boxer genome. Of the 384 SNPs subjected to genotyping via the Illumina GoldenGate assay, we examined a total of 217 SNPs (57%) after excluding SNPs with call rates $<60\%$ and poor cluster separation, leaving 167 SNPs that could not be converted to a readable assay. Of the 217 markers, 34 had minor allele frequencies <0.01 , of which 30 were monomorphic. Remarkably, 107 (49.3%) of the markers had allele frequencies >0.2 . See Table 2 for characteristics of the 217 SNPs successfully assayed in beagles.

Border Collies and Australian Shepherds

As we selected our original set of 384 SNPs based primarily on polymorphism in beagles, the question of how these SNPs would perform in other breeds remained open, so we selected 2 additional breeds for analysis to explore this question. Border collies and Australian shepherds are assigned, roughly, to either the same (border collie) or overlapping (Australian shepherd) genetic clusters as beagles (Parker and Ostrander 2005), according to a classification based on microsatellite

Table 2. Characteristics of 217 SNPs on canine chromosome 9 that were successfully assayed in a sample of 86 purebred beagles

Mean call rate per SNP ^a	0.97 ± 0.05
Mean call rate per sample ^b	0.97 ± 0.002
Average minor allele frequency	0.203 (median = 0.198)
Average intermarker distance	243 kb
Markers within 500 kb of nearest marker	185 (86.0%)
Markers within 100 kb of nearest marker	90 (41.8%)
Markers within 10 kb of nearest marker	34 (15.8%)

^a Median 1.0, range 0.64–1.0.^b Median 0.98, range 0.88–1.0.

markers (Parker et al. 2004). The origins of these breeds are obscure, but all 3 are either known or thought to trace back (at least in part, in the case of the Australian shepherd) to Great Britain. The border collie and Australian shepherd are both used, either currently or historically, as livestock working dogs and classified as herding breeds by most major multibreed registries. The beagle, by contrast, is a hound, developed and often still used as a pack hunting dog for pursuit of small game.

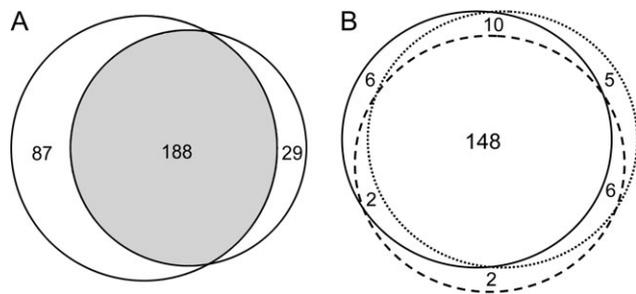
We genotyped a total of 76 border collies and 15 Australian shepherds. Of the 384 SNPs assayed, we were able to successfully genotype the same set of markers in both breeds using the exclusion criteria described above for a total of 275 successful SNPs (72%), leaving 109 SNPs that could not be converted to a readable assay for these breeds. Among the 275 SNPs assessed in border collies and Australian shepherds, 28 SNPs had minor allele frequencies <0.01, 39 (14%) were monomorphic in either Australian shepherds or border collies, and 18 (64%) were monomorphic in both breeds. Of 275 SNPs, 130 (47%) of the markers had allele frequencies >0.2. See Table 3 for characteristics of the 275 SNPs successfully assayed in border collies and Australian shepherds.

Combined Data

We noted differences between our first genotyping array analysis and the second. Although call rates were similar between experiments, we were able to score 217 SNPs on the first array and 275 on the second. Of these SNPs, 188 were called in common between SAMs, whereas 29 and 87 SNPs were called only in the beagle and border collie/Australian shepherd experiments, respectively (Figure 2A). For the 188 SNPs successfully assayed for both arrays, 179 were polymorphic in at least one of the 3 breeds. Of these 179 SNPs, 148 (82.7%) were polymorphic in all 3 breeds (Figure 2B). Alleles for only 1–3% of these SNPs were fixed (monomorphic) in any one of the 3 breeds.

Genotyping Quality

We included blood samples for 3 dogs (2 unrelated border collies and a mixed breed dog) in both analyses. For 2 of these

**Figure 2.** (A) Convergence of SNPs called for 2 analyses, including beagles and border collies/Australian shepherds. (B) Polymorphic SNPs in 3 breeds. Solid line, beagles; dotted line, border collies; dashed line, Australian shepherds. The numbers indicate number of polymorphic SNPs shared by breeds.

dogs (one border collie and the mixed breed dog), we found 100% concordance of genotypes for the 188 SNPs in common between arrays. For the third dog, 3 SNPs out of 188 did not have concordant calls (98.4% concordance between SAMs).

We also included triplicates of the 2 border collie samples on our second SAM, with 2 samples from each dog extracted from buccal swabs, and the third sample extracted from whole blood. Genotype calls between buccal swab samples were concordant for 185 out of 188 SNPs (98.4%) for the first dog and 187 out of 188 (99.5%) for the second dog. Concordance between buccal and blood samples on the second SAM were more than 98% for both of these dogs.

Results of LD Analyses

Beagles

Pairwise comparisons were performed between markers within 10 Mb. For these markers, the mean $r^2 = 0.06$, median = 0.01. For D' , the mean = 0.54, median = 0.46. When measuring LD between 183 markers having a minor allele frequency of 0.001 and that are ≤ 1 Mb apart, the average $D' = 0.61$, with an average pairwise r^2 of 0.16 (see Table 4).

At this coarse resolution, we noted 11 haplotype blocks comprising 52.1 Mb scattered across 64.4 Mb of chromosome

Table 3. Characteristics of 275 SNPs on canine chromosome 9 that were successfully assayed in a sample of 76 purebred border collies and 15 purebred Australian shepherds

Mean call rate per SNP ^a	0.97 ± 0.06
Mean call rate per sample ^b	0.97 ± 0.03
Average minor allele frequency	0.20 (median = 0.18)
Average intermarker distance	194 kb
Markers within 500 kb of nearest marker	237 (86.2%)
Markers within 100 kb of nearest marker	137 (49.8%)
Markers within 10 kb of nearest marker	51 (18.5%)

^a Median 0.99, range 0.65–1.0.^b Median 0.98, range 0.81–1.0.

Table 4. Average D' and r^2 for 3 breeds. LD statistics were calculated using 215 SNPs for beagles and 269 SNPs in both border collies and Australian shepherds.

Breed	<5 Mb		<1 Mb		<500 kb		<100 kb		<50 kb		<10 kb	
	D'	r^2	D'	r^2	D'	r^2	D'	r^2	D'	r^2	D'	r^2
Beagle ($n = 86$)	0.59	0.09	0.65	0.15	0.67	0.16	0.72	0.24	0.76	0.30	0.86	0.45
Border collie ($n = 76$)	0.63	0.10	0.71	0.15	0.73	0.16	0.81	0.25	0.86	0.32	0.90	0.41
Australian shepherd ($n = 15$)	0.67	0.14	0.70	0.17	0.71	0.19	0.80	0.28	0.84	0.37	0.89	0.46

9 using the block definition described by Gabriel et al. (2002), with one major haplotype per block (allele frequency ≥ 0.5) and the largest block = 980 Kb. One region showed extensive LD (Figure 3). The 3 blocks in this region averaged 913 kb, whereas the other 8 blocks averaged 5.3 kb in size. The average number of inferred haplotypes in these 11 blocks was 3.4 per block (range, 2.0–5.0), and the average frequency of the most common inferred haplotype was 0.61 ± 0.10 . The 2 or 3 most common haplotypes accounted for $>80\%$ of the inferred haplotypes, and the 2 most common haplotypes within each block accounted for an average of 85% of the haplotypes within a block.

To determine the ability of these SNPs to tag the genetic diversity in the tested region, we used the TAGGER routine as implemented in HAPLOVIEW to determine how many of the 183 markers would be sufficient to represent, or tag, the others with an $r^2 \geq 0.8$ (Barrett et al. 2005). A total of 155 SNPs out of 183 were required (85%), suggesting a higher

density of markers would be needed to more exhaustively represent the genetic diversity in this breed on chromosome 9.

Border Collies

For markers within 10 Mb of each other, the mean $r^2 = 0.07$, median = 0.01. For D' , the mean = 0.57, median = 0.53. When measuring LD between 269 markers having a minor allele frequency ≥ 0.001 and that are ≤ 1 Mb apart, the average $D' = 0.71$, with an average pairwise r^2 of 0.15 (see Table 4.). Haplotype block structure was more fragmented than in beagles, with 18 smaller blocks identified and the largest block 2.48 Mb in size. The average number of inferred haplotypes in these 18 blocks was 3.2 per block (range, 2.0–8.0), and the average frequency of the most common inferred haplotype was 0.68 ± 0.18 . As with the beagles, the 2 or 3 most common haplotypes accounted for $>80\%$ of the inferred haplotypes.

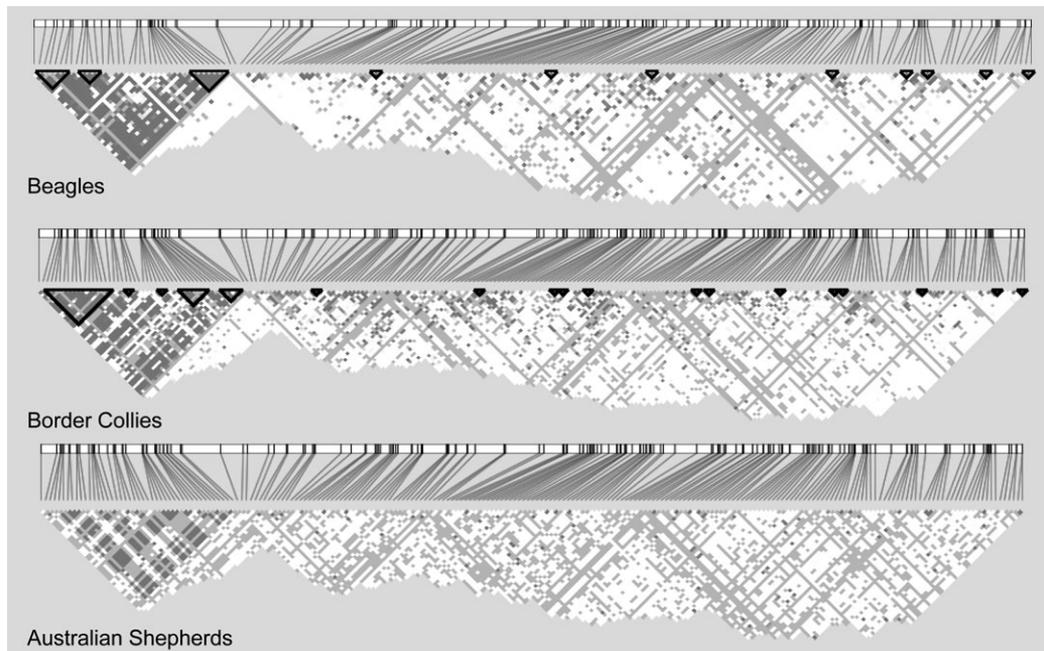


Figure 3. HAPLOVIEW plots of LD for 3 breeds. Upper panel, beagles ($n = 86$); middle panel, border collies ($n = 76$); and lower panel, Australian shepherds ($n = 15$). SNPs distributed more than ~ 52 Mb (centromere to left, distal telomere to right) are shown as tick marks. Pairwise D' relationships are shown for markers ≤ 10 Mb apart, with red areas representing higher levels of LD. Blue areas represent LD comparisons with low confidence of estimation. Dark triangles represent haplotype blocks as defined by Gabriel et al. (2002).

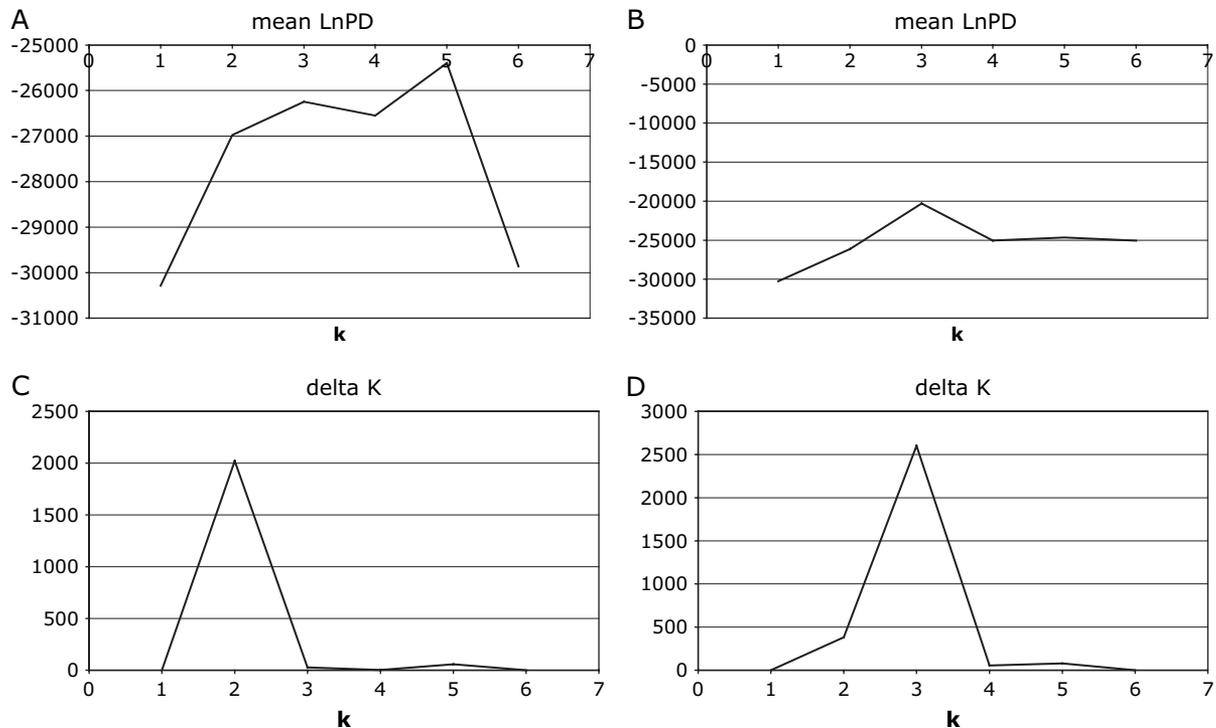


Figure 4. Identification of most probable estimate of K calculated from 188 SNPs in 3 breeds. **(A)** Estimated posterior probability of data for $K = 1$ through $K = 6$ under a model assuming no admixture. **(B)** Estimated posterior probability of data for $K = 1$ through $K = 6$ under a model assuming admixture. **(C)** Calculation of ΔK for $K = 1$ through $K = 6$ under a model assuming no admixture. **(D)** Calculation of ΔK for $K = 1$ through $K = 6$ under a model assuming admixture.

Australian Shepherds

For markers within 10 Mb of each other, the mean $r^2 = 0.11$, median = 0.05. For D' , the mean = 0.64, median = 0.71. When measuring LD between 249 markers having a minor allele frequency of 0.001 and that are ≤ 1 Mb apart, the average $D' = 0.70$, with an average pairwise r^2 of 0.17 (Table 4). There were no discernible haplotype blocks in Australian shepherds, possibly due to small sample size.

Population Structure

We assessed population structure and substructure across breeds using the 188 markers readable in common between beagles, border collies, and Australian shepherds. The results of these analyses are represented graphically in Figure 4. Under a model assuming no admixture, our results estimate 2 populations in these data according to the ΔK statistic (Evanno et al. 2005), which is based on the rate of change in the log probability of the data. Under a model assuming admixture, they recover 3 populations.

Under a model assuming 2 populations ($K = 2$) and no admixture, we found that beagles were distinct from the herding breeds (Figure 5). The estimated proportion of ancestry derived from the 2 assumed populations varied between the 2 extremes (0.996/0.004, beagles; 0.012/0.988, border collie) with an intermediate value for Australian shepherds (0.20/0.80). Applying a model that allows admixture

led to reduced precision in assigning ancestry (Figure 5), but distinctions were still observed between the 3 breeds. Under a model assuming 3 populations ($K = 3$), we observed clear distinctions between all 3 breeds using these SNPs. The results of simulations assuming more than 3 populations were highly unstable.

Conclusions

The results described here were unexpected in several ways. Although amplified human buccal swab DNA samples have proved amenable to high-throughput genotyping (Park et al. 2005), it was unknown whether canine DNA from buccal swabs would be of sufficient quality for use in the highly sensitive GoldenGate assay. Canine buccal swab samples are also typically much lower in total DNA recovered than human buccal samples. Although the conversion rate of designed assays to usable SNPs was not overwhelming, we were able to use the majority of SNPs for our analyses.

Our conversion rates increased from 57% to 71% between trials with the Illumina genotyping platform. This suggests that we have not yet optimized our ability to extract useful marker data from the designed SNPs. Beyond our increasing familiarity with the technology, there are other possible explanations for the differences in success of the assay between our first SAM analysis and our second. Experience from our initial experiments led us to further optimize our

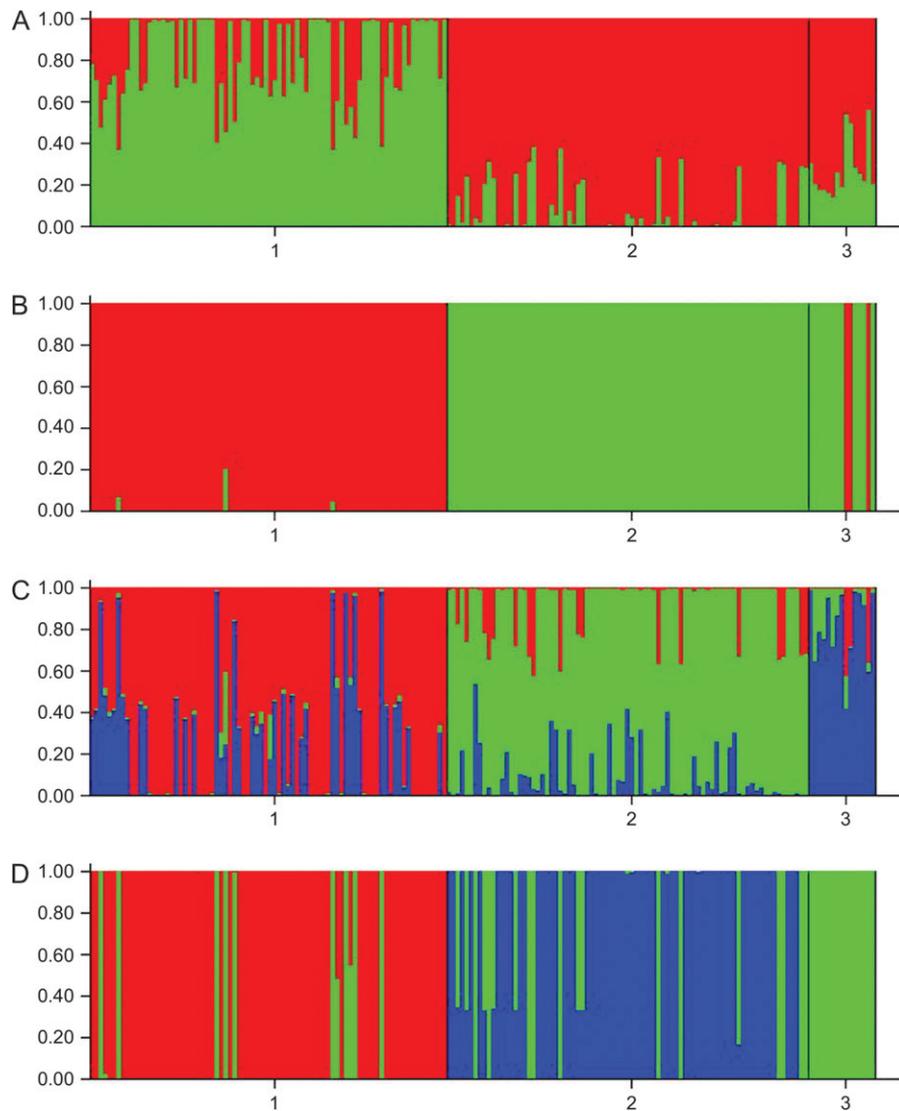


Figure 5. Population structure in 3 breeds. **(A)** Under a model assuming $K = 2$ and allowing admixture. **(B)** Under a model assuming $K = 2$ and no admixture. **(C)** Under a model assuming $K = 3$ and allowing admixture. **(D)** Under a model assuming $K = 3$ and no admixture. Breeds are denoted below each figure as population 1 = beagles, 2 = border collies, 3 = Australian shepherds. The y axis represents proportion of ancestry.

DNA quantification methodology, leading to a more uniform application of samples to the arrays. To compensate for the low amount of target template and the complex mixture of genomes found in amplified buccal swab DNA, we also increased the total amount of DNA analyzed to a much larger amount than recommended by Illumina.

Another unexpected finding was the extent of polymorphism in the SNPs selected for study. This result may not be entirely surprising, given that a single beagle was used to score sequence differences with the boxer, thus enriching the resulting SNP set for more common variants, although we would have expected a number of rare variants as well. It might be argued that the algorithm we used to choose SNPs led to a final evaluation pool biased toward elevated allele

frequencies. This cannot entirely explain our results as the SNPs that could not be adequately clustered were not necessarily of low frequency (see Figure 1B). We found it remarkable that the mean and median allele frequencies for usable SNPs were 0.2 for all 3 breeds. Similarly, we found that the allele frequency spectra for each breed were similar, with more than half of the SNPs having minor allele frequencies ≥ 0.25 . This suggests that a substantial number of SNPs discovered during the sequencing of the dog genome will be useful markers for gene-mapping studies.

Finally, we have demonstrated the portability of markers selected for polymorphism in beagles to 2 other breeds, the border collie and the Australian shepherd. Although these 3 breeds have been suggested to belong to the same

or overlapping genetic clusters (Parker et al. 2004), we expected that SNPs chosen for polymorphism in beagles would be enriched for breed-specific markers. We were somewhat surprised that only 6 SNPs were polymorphic in beagles but not in border collies or Australian shepherds. This finding accords with the results described by Lindblad-Toh et al. (2005), where approximately 73% of 1283 SNPs chosen from 373 383 SNPs found by sequence variation between 9 breeds (including the beagle) and the boxer were found to be polymorphic in samples of 20 dogs from each of 10 additional breeds. This suggests that the current canine SNP database is enriched for SNPs that are shared among breeds, which would be highly desirable for developing a mapping set of SNPs with portability across breeds.

We observed significant differences in allele frequency between breeds. For example, when comparing SNPs scored in both beagles and border collies, we found an average allele frequency difference of 31%, with 43% of the SNPs having opposite minor alleles between the breeds. Because of these differences, our relatively small marker set from a single chromosome facilitated fairly accurate identification of population substructure. The utility of a working collection of canine SNPs is enhanced by interbreed allele frequency differences, reflecting differences in underlying patterns of LD.

Our estimates of canine LD were not significantly different from those found in previous surveys. We observed that LD extended over long distances in the canine genome and that LD was quite variable across regions of chromosome 9 in all 3 breeds examined. One region several megabases in size showed substantial patterns of LD within each breed. Lacking comparable analyses of chromosome 9 across a wide sample of breeds, we cannot interpret the biological significance of this finding at this time. Within all 3 breeds, the level of LD was at the lower end of estimates reported by Lindblad-Toh et al. (2005). These 3 breeds show an average r^2 per genomic interval greater than the Labrador retriever, roughly equivalent to the golden retriever and English springer spaniel, and less than the Akita, pug, bullmastiff, Irish wolfhound, and rottweiler. The relatively low levels of LD we found may result from the use of larger sample sizes in our study compared with prior surveys as small sample size often leads to inflated estimates of LD based on sampling bias. Indeed, our highest estimates of average r^2 per genomic interval were in the Australian shepherd (Table 4), for which our sample was only 15 individuals, <20% of the size of the other breed samples.

Although pedigree analysis in canine gene finding experiments is still very useful (Cargill et al. 2005; Lohi et al. 2005), there are early indications that an LD-mapping approach in case-control samples is also feasible, demonstrated by successful efforts to separately map the merle and dermatomyositis loci in unrelated Shetland sheepdogs (Clark et al. 2005, 2006). Indeed, theoretical considerations based on assessments of canine LD suggest that relatively small case-control samples genotyped with tens of thousands of SNPs may be useful for trait-mapping studies (Lindblad-Toh et al. 2005). Thus, it is likely that many investigators will rely on SNP-based genotyping approaches to carry out LD-mapping stud-

ies in the dog. Because buccal swab samples are used in so many canine genetics studies, it is reassuring that technologies for SNP genotyping will allow utilization of such samples.

In summary, we determined that available SNP catalogs include markers useful for studies of genetic diversity and trait mapping. We also found that DNA samples extracted from buccal swabs are amenable to high-throughput genotyping analysis. Although we have not yet attempted genotyping using higher density oligonucleotide array platforms, our success using the GoldenGate assay suggests that those experiments may be fruitful. The low overall conversion rate, likely due to assay sensitivity to low DNA template concentrations, suggests that many redundant markers may be required to assure desired SNP coverage using buccal swab samples.

Funding

McKnight Foundation Neuroscience of Brain Disorders Award; a Hellman Family Faculty Fund award; a University of California San Francisco Innovations in Basic Science award. National Institutes of Health (T32 MH019552-13) to M.L.C.

Supplementary Material

Supplementary Table can be found at <http://www.jhered.oxfordjournals.org/>.

Acknowledgments

The authors would like to extend special thanks to Jonathan Woo and Wendy Chan of the University of California San Francisco Genomics Core Facility for technical assistance with SNP genotyping. We would like to thank all participating dog owners for their cooperation. The authors would also like to acknowledge Elaine Ostrander for helpful discussions.

References

- Barker DL, Hansen MST, Faruqi AF, Giannola D, Irsula OR, Lasken RS, Latterich M, Makarov V, Oliphant A, Pinter JH, et al. 2004. Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. *Genome Res.* 14:901–907.
- Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 21:263–265.
- Barsoum SC, Callahan HM, Robinson K, Chang PL. 2000. Canine models for human genetic neurodegenerative diseases. *Prog Neuropsychopharmacol Biol Psychiatry.* 24:811.
- Brooks M, Sargan DR. 2001. Genetic aspects of disease in dogs. In: Ruvinsky A, Sampson J, editors. *The genetics of the dog*. New York: CABI Publishing, p. 191–266.
- Cargill E, Famula T, Schnabel R, Strain G, Murphy K. 2005. The color of a dalmatian's spots: linkage evidence to support the TYRP1 gene. *BMC Vet Res.* 1:1.
- Clark LA, Credille KM, Murphy KE, Rees CA. 2005. Linkage of dermatomyositis in the Shetland sheepdog to chromosome 35. *Vet Dermatol.* 16:392–394.

- Clark LA, Tsai KL, Steiner JM, Williams DA, Guerra T, Ostrander EA, Galibert F, Murphy KE. 2004. Chromosome-specific microsatellite multiple sets for linkage studies in the domestic dog. *Genomics*. 84:550.
- Clark LA, Wahl JM, Rees CA, Murphy KE. 2006. From the cover: retrotransposon insertion in SILV is responsible for merle patterning of the domestic dog. *PNAS*. 103:1376–1381.
- Clutton-Brock J. 1999. A natural history of domesticated mammals. Cambridge (UK): Cambridge University Press.
- de Bakker PIW, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. 2005. Efficiency and power in genetic association studies. *Nat Genet*. 37:1217–1223.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol*. 14:2611–2620.
- Fan J-B, Chee MS, Gunderson KL. 2006. Highly parallel genomic assays. *Nat Rev Genet*. 7:632.
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, et al. 2002. The structure of haplotype blocks in the human genome. *Science*. 296:2225–2229.
- Galibert F, Andre C, Hitte C. 2004. Le chien, un modele pour la genetikue des mammiferes. *Med Sci*. 20:761–766.
- Guyon R, Kirkness E, Lorentzen TD, Hitte C, Comstock KE, Quignon P, Derrien T, Andre C, Fraser CM, Galibert F, et al. 2003. Building comparative maps using 1.5x sequence coverage: human chromosome 1p and the canine genome. *Cold Spring Harb Symp Quant Biol*. 68:171–177.
- Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, Delcher AL, Pop M, Wang W, Fraser CM, et al. 2003. The dog genome: survey sequencing and comparative analysis. *Science*. 301:1898–1903.
- Koskinen MT, Bredbacka P. 2000. Assessment of the population structure of five Finnish dog breeds with microsatellites. *Anim Genet*. 31:310–317.
- Lench N, Stanier P, Williamson R. 1988. Simple non-invasive method to obtain DNA for gene analysis. *Lancet*. 1:1356–1358.
- Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ, Zody MC, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*. 438:803.
- Lohi H, Young EJ, Fitzmaurice SN, Rusbridge C, Chan EM, Vervoort M, Turnbull J, Zhao X-C, Ianzano L, Paterson AD, et al. 2005. Expanded repeat in canine epilepsy. *Science*. 307:81.
- Min JL, Lakenberg N, Bakker-Verweij M, Suchiman E, Boomsma DI, Slagboom PE, Meulenbelt I. 2006. High microsatellite and SNP genotyping success rates established in a large number of genomic DNA samples extracted from mouth swabs and genotypes. *Twin Res Hum Genet*. 9:501–506.
- Ostrander EA, Kruglyak L. 2000. Unleashing the canine genome. *Genome Res*. 10:1271–1274.
- Overall KL, Hamilton SP, Chang ML. Forthcoming 2006. Understanding the genetic basis of canine anxiety: phenotyping dogs for behavioral, neurochemical, and genetic assessment. *J Vet Behav*. 1:124–141.
- Park JW, Beaty TH, Boyce P, Scott AF, McIntosh I. 2005. Comparing whole-genome amplification methods and sources of biological samples for single-nucleotide polymorphism genotyping. *Clin Chem Lab Med*. 51:1520–1523.
- Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB, Johnson GS, DeFrance HB, Ostrander EA, Kruglyak L. 2004. Genetic structure of the purebred domestic dog. *Science*. 304:1160–1164.
- Parker HG, Ostrander EA. 2005. Canine genomics and genetics: running with the pack. *PLoS Genet*. 1:e58.
- Pask R, Rance H, Barratt B, Nutland S, Smyth D, Sebastian M, Twells R, Smith A, Lam A, Smink L, et al. 2004. Investigating the utility of combining Phi29 whole genome amplification and highly multiplexed single nucleotide polymorphism BeadArray™ genotyping. *BMC Biotechnol*. 4:15.
- Ponder KP, Melniczek JR, Xu L, Weil MA, O'Malley TM, O'Donnell PA, Knox VW, Aguirre GD, Mazrier H, Ellinwood NM, et al. 2002. From the cover: therapeutic neonatal hepatic gene therapy in mucopolysaccharidosis VII dogs. *Proc Natl Acad Sci USA*. 99:13102–13107.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics*. 155:945–959.
- Savolainen P, Zhang Y-P, Luo J, Lundeberg J, Leitner T. 2002. Genetic evidence for an east asian origin of domestic dogs. *Science*. 298:1610–1613.
- Short AD, Kennedy LJ, Forman O, Barnes A, Fretwell N, Wiggall R, Thomson W, Ollier WER. 2005. Canine DNA subjected to whole genome amplification is suitable for a wide range of molecular applications. *J Hered*. 96:829–835.
- Stemers FJ, Gunderson KL. 2005. Illumina, Inc. *Pharmacogenomics*. 6:777–782.
- Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, Ostrander EA. 2004. Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Res*. 14:2388–2396.
- Sutter NB, Ostrander EA. 2004. Dog star rising: the canine genetic system. *Nat Rev Genet*. 5:900.
- Thompson MD, Bowen RAR, Wong BYL, Antal J, Liu Z, Yu H, Siminovitch K, Kreiger N, Rohan TE, Cole DEC. 2005. Whole genome amplification of buccal cell DNA: genotyping concordance before and after multiple displacement amplification. *Clin Chem Lab Med*. 43:157–162.

This work was presented in poster form at the Third International Conference on Advances in Canine and Feline Genomics and Inherited Diseases, Davis, CA, August 2006.

Corresponding Editor: Elaine Ostrander